APUNTES PARA EL CURSO DE ECONOMETRÍA

Enero – Mayo 2015

DAVID RUELAS RODRÍGUEZ

Estos apuntes fueron revisados y corregidos para cubrir por completo el temario de los cursos de Econometría I (EST 11103) y Econometría (EST 11104) del Departamento de Estadística del Instituto Tecnológico Autónomo de México (ITAM) para el semestre Enero – Mayo 2015.

El documento no es un libro de texto, sino una síntesis de las principales definiciones, teorema, conceptos y métodos de un primer curso en Econometría de nivel intermedio, que en algunos programas académicos también se le denomina Análisis de Regresión o Pronóstico de Negocios. Esta compilación puede tenerse como referencia en el estudio de otras materias que dependen fundamentalmente de los Modelos de Regresión; por ejemplo, Microeconometría, Macroeconometría, Econometría Financiera, Análisis de Series de Tiempo, Modelos Lineales Generalizados, Modelo Econométricos Dinámicos, etc.

Cada sección fue escrita con la finalidad de revisar rápidamente la teoría y dedicar la mayor parte del tiempo de clase a: (i) la demostración y análisis de los principales resultados; y (ii) a la resolución de ejercicios que ilustran el uso de los Modelos de Regresión con aplicaciones en Economía, Finanzas y Administración.

Los apuntes de la Sección 3 sobre el Modelo de Regresión Lineal Múltiple son más largos que los demás, pues incluyen un repaso de los principales conceptos y resultados de Álgebra Matricial, que facilitan la notación y demostración de algunos teoremas.

Al inicio del documento se incorporó un índice para facilitar la búsqueda de algún tema en particular. Al final del documento se presenta como anexo el *Material de Apoyo para el Curso de Econometría* repartido durante las clases de este semestre, incluyendo las preguntas de los 58 ejercicios revisados en clase.

Esta es la cuarta versión de estos apuntes completos e incluyen las correcciones que gracias a las preguntas y comentarios de mis alumnos de este semestre fuimos detectando durante las clases. Sé que todavía hay errores por corregir y explicaciones por mejorar, así que cualquier comentario es bienvenido al correo electrónico davidruelas@hotmail.com.

David Ruelas Rodríguez

Mayo 2015

Índice

1. Introducción a la Econometría

1.1. Propósito y definición

Figura 1.1

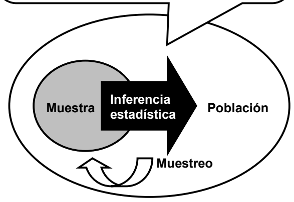
PROBABILIDAD, INFERENCIA ESTADÍSTICA Y ECONOMETRÍA

Probabilidad

- Rama de las Matemáticas y de la Estadística que se encarga del estudio formal de las reglas de la incertidumbre que permiten modelar fenómenos aleatorios
- Cálculo de probabilidades: enfoques (clásico, frecuentista y subjetivo) y desarrollo axiomático
- Probabilidad condicional e independencia estadística (RPT y TB)
- Variables aleatorias y distribuciones de probabilidad (univariadas y multivariadas)
- Modelos paramétricos: Binomial, Poisson, Uniforme, Gamma, Normal, etc.

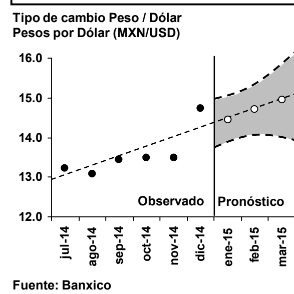
Inferencia Estadística (Estadística Matemática)

- Proceso inductivo (de lo particular a lo general) que permite obtener conclusiones sobre la población a partir del estudio de una muestra
- Siempre existe posibilidad de error y se mide en términos de probabilidad



- Distribuciones de Muestreo de estadísticas y estimadores
- Estimación puntual y por intervalo de parámetros
- Pruebas de hipótesis paramétricas

Econometría (Regresión, Pronóstico de Negocios, Series de Tiempo)



- ¿Cuál es el pronóstico para el tipo de cambio Peso / Dólar? ¿Qué intervalos garantizan cierto nivel de confianza?
- Con base en un conjunto de precios y cantidades observadas, ¿cuál es la función de demanda estimada?

La Econometría es la rama de la Economía que utiliza las técnicas estadísticas del Análisis de Regresión para analizar los fenómenos económicos

Figura 1.2

ECONOMETRÍA

Definición y propósito

- Etimológicamente econometría significa “medición económica”
- La Econometría es una disciplina que combina:
  - Teoría económica
  - Economía matemática (expresión matemática de la teoría mediante ecuaciones)
  - Estadística económica (recopilación, procesamiento y análisis exploratorio de datos económicos)
  - Estadística matemática (inferencia estadística para modelar la incertidumbre y hacer pronósticos)
- Su propósito es validar y dar contenido empírico a gran parte de la teoría económica.

Clasificación

- Categorías
  - Econometría Teórica: elaboración de métodos estadísticos para medir las relaciones económicas
  - Econometría Aplicada: uso de las herramientas de la econometría teórica para estudiar campos específicos de la economía y los negocios
- Enfoques (de cada categoría)
  - Econometría Clásica: se basa en métodos de la Estadística Frecuentista
  - Econometría Bayesiana: se basa en métodos de la Estadística Bayesiana

Figura 1.3

ANÁLISIS DE REGRESIÓN

Análisis de Regresión

- Estudio de la dependencia de una variable de interés respecto de una o más variables explicativas
- Su objetivo es estimar o predecir la media poblacional de la variable de interés

- La variable de interés (Y) es variable aleatoria
- Las variables explicativas ($X_1, X_2,...., X_k$) se supondrán conocidas
- Se busca estimar $E[Y \mid X_1, X_2,..., X_k]$

Ejemplos:

Variable dependiente	Variable(s) explicativas
• Dinero / Ingreso	• Tasa de inflación
• Consumo personal	• Ingreso personal neto disponible
• Tasa de cambio de los salarios	• Tasa de desempleo
• Producción	• Trabajo • Capital
• Mortalidad infantil	• PIB per cápita • Tasa de alfabetización de las mujeres

Origen histórico

- El científico y humanista inglés Francis Galton (1822-1911) estableció la "Ley de regresión universal"
- Los padres de estatura alta (o baja) tienden a procrear hijos altos (o bajos), sin embargo, la estatura promedio de cada grupo tiende a "regresar" a la estatura promedio de la población total ("regresión a la mediocridad")
- Este hecho fue confirmado por el matemático inglés Karl Pearson (1857-1936) mediante una muestrea de más de mil miembros de grupos familiares

1.2. Modelos Econométricos

Figura 1.4

RELACIONES DETERMINISTAS VS ESTADÍSTICAS

Relación determinista

- Relaciones funcionales propias de la Física (v.gr., Segunda Ley de Newton: F = m*a)
- La variable dependiente queda perfectamente descrita una vez que las variables explicativas son conocidas
- Mismos niveles o valores de las variables explicativas conducen siempre al mismo valor de la variable dependiente

Relación estadística

- Relaciones aleatorias o estocásticas, es decir, variables con distribuciones de probabilidad
- La variable dependiente tiene una variabilidad intrínseca o aleatoria que no puede explicarse en su totalidad por las variables explicativas
- Mismos niveles o valores de las variables explicativas no necesariamente conducen al mismo valor de la variable dependiente

Variable dependiente

Variables explicativas

Modelo Matemático: $Y = \beta_0 + \beta_1 X_1 + \beta_2 \ln(X_2)$

Parámetros

Modelo Econométrico: $Y = \beta_0 + \beta_1 X_1 + \beta_2 \ln(X_2) + u$

Perturbación Estocástica

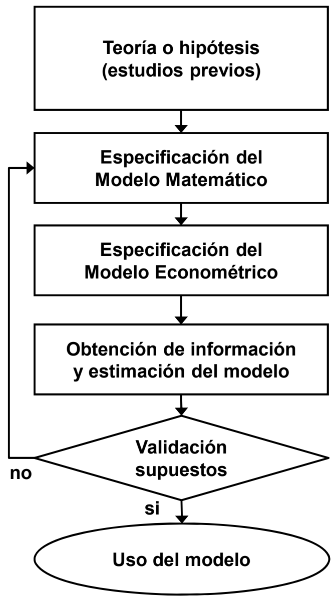
- Sustituto de las variables que se omiten en el modelo
- Hace que la variable dependiente sea variable aleatoria

1. Vaguedad de la teoría
2. Falta de disponibilidad de datos
3. Variables periféricas
4. Aleatoriedad intrínseca
5. Variables representantes (proxy) inadecuadas
6. Principio de parsimonia (lo más simple posible)
7. Forma funcional incorrecta

Figura 1.5

METODOLOGÍA DE LA ECONOMETRÍA

Metodología



Ejemplo

- Las personas incrementan su consumo (Y) al incrementar su ingreso (X) (Keynes, 1936)
- La tasa de cambio del consumo por unidad de ingreso adicional (β_1) es la propensión marginal a consumir (PMC)

$Y = \beta_0 + \beta_1 X$

- β_0 es el nivel de consumo cuando no hay ingreso (no necesariamente es cero)

$Y = \beta_0 + \beta_1 X + u$

- u es el término de perturbación o de error (media cero y varianza constante σ^2)

$\hat{Y} = -0.619 + 0.758\,X$

$\hat{\sigma}^2 = 0.007$

- Estimación de β_0, β_1 y σ^2 (Método de Mínimos Cuadrados Ordinarios)

- Significancia de los parámetros (v.gr., $H_0: \beta_1 = 0$)
- Análisis de residuos (v.gr., $H_0: \sigma^2 = 0.005$)

- Interpretación del modelo (v.gr., PMC ≈ 0.758)
- Pronóstico o predicción (v.gr., $\hat{Y}(11.5) = 8.1$)
- Análisis con fines de control o de políticas

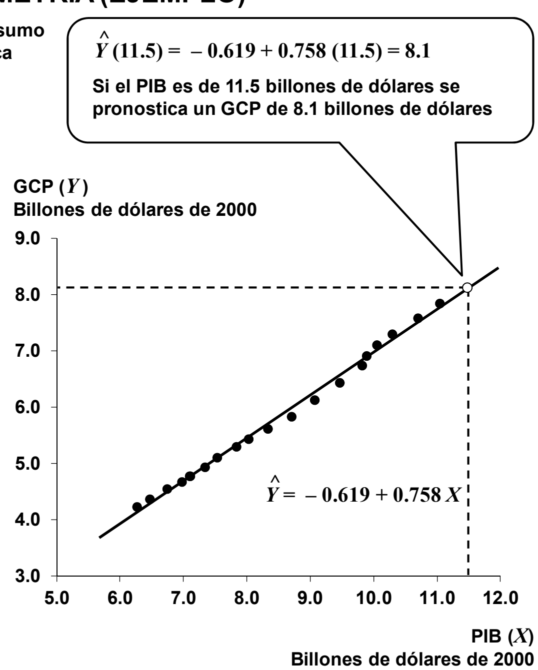
Figura 1.6

METODOLOGÍA DE LA ECONOMETRÍA (EJEMPLO)

Producto Interno Bruto (PIB) y Gasto en Consumo Personal (GCP) en Estados Unidos de América
Cifras en billones de dólares de 2000

Año	PIB (X)	GCP (Y)
1986	6.3	4.2
1987	6.5	4.4
1988	6.7	4.5
1989	7.0	4.7
1990	7.1	4.8
1991	7.1	4.8
1992	7.3	4.9
1993	7.5	5.1
1994	7.8	5.3
1995	8.0	5.4
1996	8.3	5.6
1997	8.7	5.8
1998	9.1	6.1
1999	9.5	6.4
2000	9.8	6.7
2001	9.9	6.9
2002	10.0	7.1
2003	10.3	7.3
2004	10.7	7.6
2005	11.0	7.8

Fuente: Gujarati (2009), *Econometría*, 5ª Edición. Extracto de la Tabla I.1 (p. 6).



1.3. Tipo de datos

Figura 1.7

CLASIFICACIÓN DE LOS DATOS

Clasificación de los datos		Características
Por su tipo	**Cualitativos** • Cualidades o atributos de los elementos • Se pueden codificar	• Se pueden clasificar en un número finito de categorías • Las categorías deben ser mutuamente excluyentes (cada dato pertenece a una sola clase) y exhaustivas (abarcan todos los valores potenciales de los datos)
	Cuantitativos • Valores de los elementos con significado numérico	• Discretos: provienen de un proceso de conteo. • Continuos: provienen de un proceso de medición.
Por el tiempo de recolección y observa-ción	**Transversales**	• Se reúnen en el mismo tiempo o en el mismo período • Corresponden a un conjunto de unidades experimentales
	Series de tiempo	• Se reúnen a lo largo de varios intervalos en el tiempo • Regularmente son observaciones de una misma variable en distintos momentos del tiempo
	Datos en panel	• Combina datos transversales y de series de tiempo • Se reúnen a lo largo de varios intervalos en el tiempo para el mismo conjunto transversal

Figura 1.8

ESCALAS DE MEDICIÓN

Datos	Escala	Características	Ejemplo
Cualitativos	Nominal	• Nivel más bajo de medición • Utiliza la operación más sencilla y básica: la clasificación • No es posible establecer una relación de orden	• Preferencia sobre partido político (v.gr., PAN, PRD, PRI)
	Ordinal	• Establece una relación de orden de acuerdo con la posesión de cierto atributo • No permite realizar operaciones aritméticas	• Calificación de un servicio (v.gr., malo, regular, bueno)
Cuantitativos	De intervalo	• Los datos se clasifican con base en el grado de posesión del atributo • El cero es relativo (no representa la ausencia del atributo) • Es posible realizar operaciones de suma o resta	• Temperatura en grados Centígrados (v.gr., 0°C no es ausencia de temperatura)
	De razón	• Los datos se clasifican con base en la medición exacta del grado de posesión del atributo • El cero es absoluto (ausencia del atributo) • Es posible realizar operaciones de suma, resta, multiplicación y división	• Estatura, peso, distancias o tiempo (v.gr.,1.75 metros, 72 kg, 5 km, 3 horas)

Conversión de datos

1.4. Variables

Figura 1.9

TERMINOLOGÍA

Variable dependiente	Variable explicativa
• Variable explicada	• Variable independiente
• Predicha	• Predictora
• Regresada	• Regresora
• Respuesta	• Estímulo
• Endógena	• Exógena
• Resultado	• Covariante o covariable
• Variable controlada	• Variable de control

Figura 1.10

NOTACIÓN

Caso	Variable dependiente	Variable(s) independiente(s)	Datos
Univariado o Simple $Y = g(X)$ $g: \mathbf{R} \to \mathbf{R}$	Y	X	$Y_1, Y_2, \ldots, Y_n$ $X_1, X_2, \ldots, X_n$ *n* observaciones (X_i, Y_i)
Multivariado o Múltiple $Y = g(X_1, X_2, \ldots, X_k)$ $g: \mathbf{R}^n \to \mathbf{R}$	Y	$X_1, X_2, \ldots, X_k$	$Y_1 \quad X_{11}, X_{21}, \ldots, X_{k1}$ $Y_2 \quad X_{12}, X_{22}, \ldots, X_{k2}$ $\vdots \quad \vdots \quad \vdots \quad \vdots$ $Y_n \quad X_{1n}, X_{2n}, \ldots, X_{kn}$ *k* variables *n* obs.

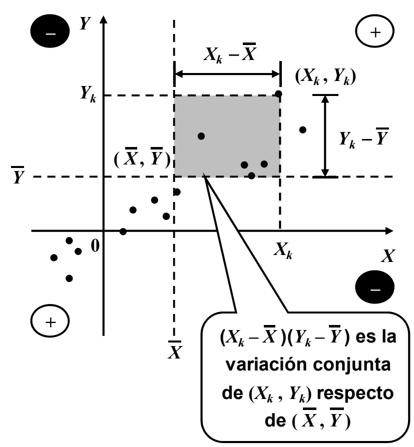
Estadísticos univariados (variable dependiente o variables independientes)

	Definición	Proposición	
Media muestral:	$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$	$\sum_{i=1}^{n} Y_i = n\overline{Y}$	• Estimador insesgado de la media poblacional μ
Varianza muestral:	$S_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2$	$S_Y^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2 \right)$	• Estimador insesgado de la varianza poblacional σ^2

Figura 1.11

ESTADÍSTICOS BIVARIADOS: COVARIANZA Y CORRELACIÓN

Interpretación geométrica



- Puede ser positiva o negativa
- Se expresa en unidades cuadradas

Def. Covarianza Muestral

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$$

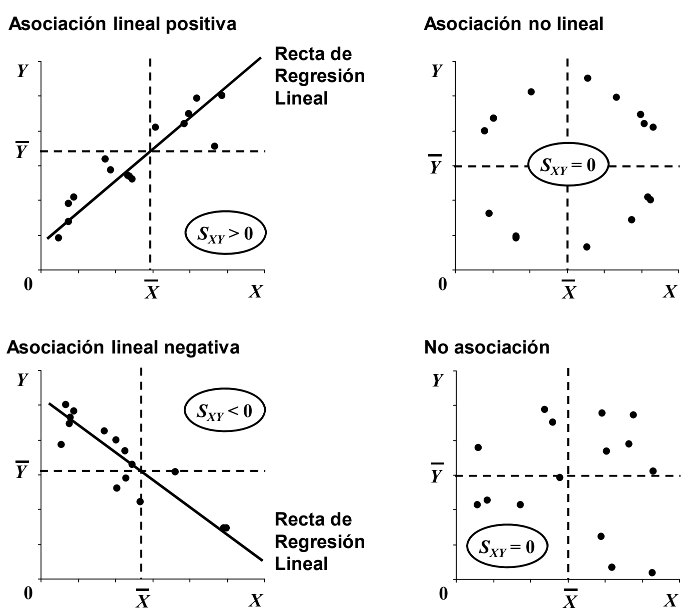
- Es el promedio de la variación conjunta
- Identifica la asociación lineal entre dos variables
- Puede ser positiva o negativa
- Es simétrica: $S_{XY} = S_{YX}$
- La covarianza de X con X es la varianza de X, es decir, $S_{XX} = S_X^2$

Proposición

$$S_{XY} = \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i Y_i - n \overline{X}\,\overline{Y} \right)$$

Figura 1.12

INTERPRETACIÓN DE LA COVARIANZA



- La covarianza solo identifica la asociación lineal entre dos variables
- Puede ser cualquier valor real, es decir, $S_{XY} \in (-\infty, \infty)$
- Para cuantificar el grado de asociación lineal se define al coeficiente de correlación lineal

Figura 1.13

COEFICIENTE DE CORRELACIÓN LINEAL

Definición

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

- **Mide el grado de asociación lineal entre dos variables**
- **Es adimensional e invariante ante transformaciones lineales**
- **Es simétrico: $r_{XY} = r_{XY}$**

Proposición

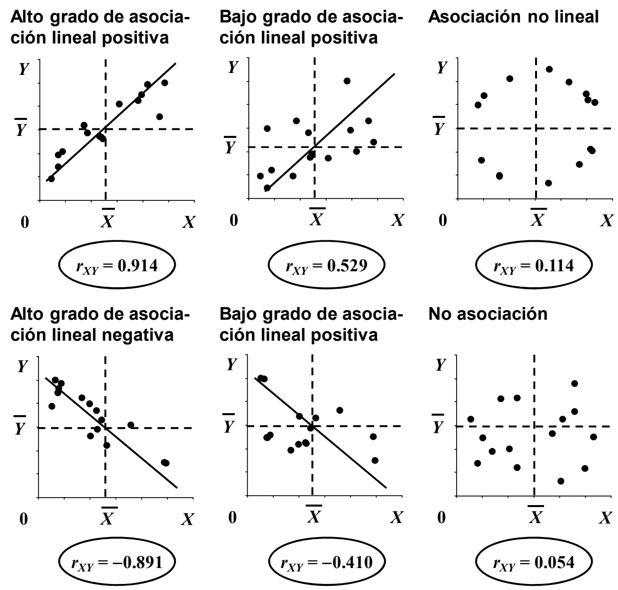
$$-1 \le r_{XY} \le 1$$

- $r_{XY} \to -1 \Rightarrow$ **Alto grado de asociación lineal negativa**
- $r_{XY} \to 1 \Rightarrow$ **Alto grado de asociación lineal positiva**
- $r_{XY} \to 0 \Rightarrow$ **No asociación o asociación no lineal entre las variables X y Y**

Correlación Perfecta

$|\, r_{XY} \,| = 1 \Leftrightarrow \exists\ \beta_0, \beta_1 \neq 0,$ **tales que** $Y = \beta_0 + \beta_1 X$

Ejemplos de interpretación



1.5. Usos y abusos de la Econometría

Figura 1.14

REGRESIÓN VS CORRELACIÓN: NO CAUSALIDAD

Análisis	Objetivo	Supuestos sobre las variables
Correlación	• Mide el grado de asociación lineal entre dos variables • Genera un escalar (Coeficiente de Correlación Lineal)	• Tratamiento simétrico • Las variables son aleatorias • Las variables, dependiente e independiente, son indistintas ($r_{XY} = r_{YX}$)
Regresión	• Estima o predice el valor promedio de una variable con base en los valores fijos de otras • Genera una ecuación (v.gr., Recta de Regresión Lineal)	• Tratamiento asimétrico • La variable dependiente es aleatoria pero las variables independientes toman valores fijos (en muestras repetidas)

- **El Análisis de Correlación y el Análisis de Regresión estudian estadísticamente la relación que hay entre dos o más variables**
- **Una relación estadística, por más fuerte y sugerente que sea, nunca podrá establecer una conexión causal (ni correlación ni regresión implican causalidad)**
- **Incluso la correlación perfecta ($|\, r_{XY}| = 1$) significa que todos los datos (X_i, Y_i) se ubican sobre la recta de regresión (colinealidad), sin embargo, ni en esos casos se puede concluir que haya causalidad**
- **La causalidad sólo se puede establecer a partir de consideraciones teóricas, y se pueden probar estadísticamente mediante Diseño de Experimentos (se controlan las variables independientes y se observa a la variable dependiente)**

Ejercicios $\boxed{\text{E1}}$ y $\boxed{\text{E2}}$.

2. Modelo de Regresión Lineal Simple

2.1. Modelo de regresión y esperanza condicional.

Def. Análisis de Regresión.

Es el estudio de la dependencia de una variable de interés respecto de una o más variables explicativas. Su objetivo es estimar o predecir la media poblacional de la variable de interés.

El **caso univariado o simple** se caracteriza por considerar únicamente una variable explicativa (X). Si se denota por Y a la variable de interés o variable dependiente, entonces el objetivo del Análisis de Regresión será estimar la media de Y suponiendo conocidos los valores de X, es decir, la esperanza condicional $E[Y|X]$.

Hay que recordar que si X y Y son variables aleatorias discretas (o continuas), $f_{X,Y}(x,y)$ es su función de masa (o densidad) de probabilidad conjunta y $f_X(x)$ la función de masa (o densidad) de probabilidad marginal de X, entonces la función de masa (o densidad) de probabilidad condicional de $\{Y|X=x\}$ está dada por $f_{Y|X}(y|x)=\dfrac{f_{X,Y}(x,y)}{f_X(x)}$, $f_X(x)\neq 0$.

Proposición

i) $E[Y|X=x]=\sum_y y\, f_{Y|X}(y|x)$ si X y Y son variables aleatorias discretas.

ii) $E[Y|X=x]=\int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)dy$ si X y Y son variables aleatorias continuas.

Para valores fijos de X ($X = x$), la esperanza condicional de Y es una **función de *x***, es decir, $E[Y|X=x]=g(x)$.

Ejercicios E3 y E4 .

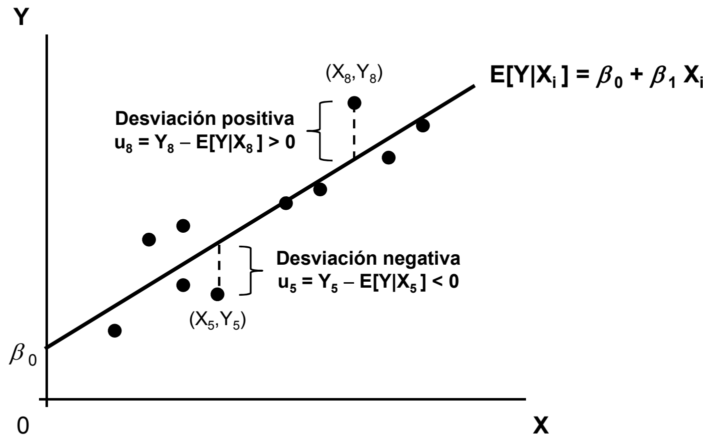
Notación: En al Análisis de Regresión Univariado se considerarán las parejas de datos (X_i, Y_i), i = 1, 2,…, n, de modo que los valores fijos de X serán X_1, X_2,…, X_n. Para simplificar la notación se denotará a la esperanza condicional por $E[Y|X_i]=g(X_i)$.

Def. Función de Regresión Poblacional (FRP). Es el lugar geométrico de las medias condicionales de la variable dependiente para los valores fijos de la variable explicativa, es decir, $E[Y|X_i]=g(X_i)$.

La **forma de la FRP** depende del comportamiento aleatorio de X y Y. El análisis de los datos (X_i, Y_i), i = 1, 2,…, n, y la teoría pueden ayudar empíricamente a determinar la forma de $g(X_i) = E[Y|X_i]$.

Por ejemplo, para los datos (X_1, Y_1), (X_2, Y_2),…, (X_{10}, Y_{10}) de la figura 2.1 podría pensarse que la FRP es "lineal", es decir, $E[Y|X_i] = \beta_0 + \beta_1 X_i$, donde $\boldsymbol{\beta_0}$ **y** $\boldsymbol{\beta_1}$ se denominan **coeficientes de regresión**. Geométricamente se trata de una línea recta con ordenada al origen β_0 (también llamada *intercepto*) y pendiente β_1.

Figura 2.1



Def. Linealidad en variables. Se dice que la FRP es lineal en variables si en $g(X_i) = E[Y|X_i]$ sólo aparece X_i elevada a la potencia 1 acompañada de constantes. Por ejemplo, $E[Y|X_i] = \beta_0 + \beta_1 X_i$, $E[Y|X_i] = \beta_1^2 X_i$ y $E[Y|X_i] = \dfrac{\alpha_0 - \alpha_1^2 X_i}{\ln(\alpha_3)}$ con $\alpha_3 > 0$ son FRPs lineales en variables.

Def. Linealidad en parámetros. Se dice que la FRP es lineal en parámetros si en $g(X_i) = E[Y|X_i]$ sólo aparecen parámetros individuales elevados a la potencia 1, sin importar la forma en que aparece X_i. Por ejemplo $E[Y|X_i] = \beta_0 + \beta_1 X_i$, $E[Y|X_i] = \beta_0$ y $E[Y|X_i] = \beta_0 + \beta_1 \ln(X_i) + \beta_2 X_1^2$ son FRPs lineales en parámetros, pero $E[Y|X_i] = \beta_1^2 X_i$, $E[Y|X_i] = \beta_0 + \beta_0 \beta_1 X_i$ y $E[Y|X_i] = \dfrac{\alpha_0 - \alpha_1^2 X_i}{\ln(\alpha_3)}$, $\alpha_3 > 0$, no lo son.

Ejercicios $\boxed{\text{E5}}$ y $\boxed{\text{E6}}$.

En la figura 2.1 también se puede observar cómo la esperanza condicional $E[Y|X_i]$ no necesariamente coincide con los datos Y_i para todas las X. La *desviación* $Y_i - E[Y|X_i]$ (dato Y_i correspondiente a X_i menos su respectiva media condicional) puede ser positiva o negativa.

Def. Perturbación estocástica o término de error (u_i).

Son las desviaciones $u_i = Y_i - E[Y|X_i]$, y son variables aleatorias no observables que sustituyen o representan a todas las variables omitidas o ignoradas que, en conjunto, pueden afectar a Y.

Existen muchas razones por las cuáles no es necesario o no es posible considerar algunas variables en los modelos econométricos, por ejemplo:

- Vaguedad de la teoría
- Falta de disponibilidad de datos
- Variables periféricas (cuyo efecto conjunto es marginal)
- Aleatoriedad intrínseca (comportamiento humano)
- Variables representantes (o *proxy*) inadecuadas
- Principio de parsimonia (lo más simple posible)
- Forma funcional incorrecta

Despejando Y_i de la definición de perturbación estocástica se obtiene el Modelo de Regresión Simple, compuesto de una parte determinista (parámetros y valores conocidos X_i) y una parte aleatoria.

Def. Modelo de Regresión Simple.

$$Y_i = \underbrace{E[Y|X_i]}_{\substack{\text{Componente}\\ \text{sistemático o}\\ \text{determinista}}} + \underbrace{u_i}_{\substack{\text{Componente}\\ \text{no sistemático}\\ \text{o aleatorio}}}$$

En particular, se estudiarán los Modelos de Regresión cuyo componente determinista es lineal en los parámetros, partiendo de aquel que además es lineal en las variables.

2.2. Modelo de Regresión Lineal Simple (MRLS).

2.2.1. Elementos del MRLS.

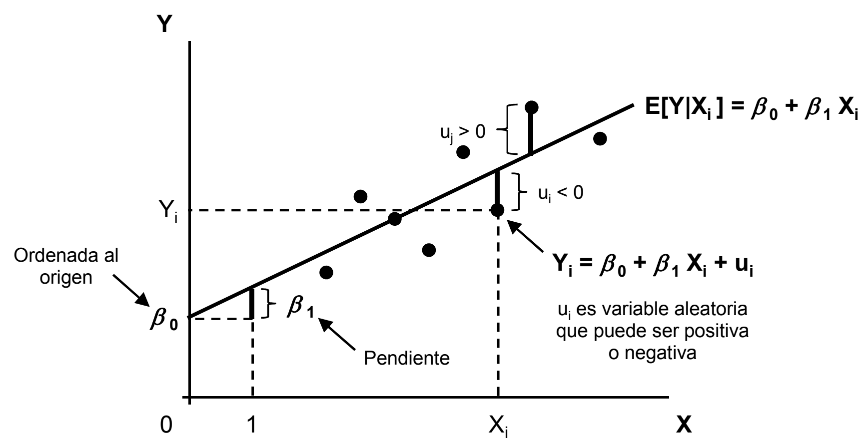
Si $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ son n observaciones de la variable dependiente Y y la variable explicativa X, cuya FRP es la recta $E[Y|X_i] = \beta_0 + \beta_1 X_i$, entonces el **Modelo de Regresión Lineal Simple (MRLS)** establece que valores Y_i pueden ser modelados como la suma de esta media condicional y un término de error u_i de la siguiente manera:

$$\boxed{Y_i = \beta_0 + \beta_1 X_i + u_i}$$

REPRESENTACIÓN GEOMÉTRICA DEL MRLS

Geométricamente, el MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$ se presenta en la figura 2.2. Para cada dato X_i su correspondiente Y_i es el valor de la media condicional $E[Y|X_i] = \beta_0 + \beta_1 X_i$ más un término de error u_i que puede ser negativo o positivo. La media condiciona es la FRP y para el MRLS es una recta con **pendiente β_1** y **ordenada al origen β_0**.

Figura 2.2



INTERPRETACIÓN DE LOS PARÁMETROS DEL MRLS

Los parámetros del MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$ son β_0 y β_1 y se interpretan de la siguiente manera:

- **La ordenada al origen o intercepto β_0** corresponde al valor esperado de la variable dependiente Y en ausencia de la variable independiente X, es decir, $\beta_0 = E[Y|X = 0]$.
- **La pendiente β_1** representa el cambio (positivo o negativo) de la media condicional $E[Y|X_i]$ al aumentar la variable independiente X en una unidad.

En algunas ocasiones β_0 carece de interpretación, por ejemplo, cuando la variable independiente es el tiempo (*t*), $t = 0$ no significa ausencia de tiempo y el valor β_0 sólo indica el nivel de inicio de una serie de tiempo.

En general, para analizar la FRP $E[Y|X_i] = g(X_i)$ frecuentemente se consideran algunas **medidas de sensibilidad** como *efectos parciales*, *elasticidades* o *semi-elasticidades*.

Def. Efecto parcial ($\Delta E[Y|X_i]$). Es el cambio marginal de $E[Y|X_i]$ cuando X_i crece en una cantidad pequeña ΔX_i, es decir, $\Delta E[Y|X_i] \approx \left(\frac{d}{dX_i} E[Y|X_i]\right)\Delta X_i$.

Def. Elasticidad (η). Es el cambio porcentual aproximado en $E[Y|X_i]$ cuando X_i aumenta *un punto porcentual*, es decir, $\eta = \left(\frac{d}{dX_i} E[Y|X_i]\right)\frac{X_i}{E[Y|X_i]} = \frac{d\ln(E[Y|X_i])}{d\ln(X_i)}$.

Def. Semi-elasticidad ($\overline{\eta}$). Es el cambio porcentual aproximado en $E[Y|X_i]$ cuando X_i aumenta en *una unidad*, es decir, $\overline{\eta} = 100\left(\frac{d}{dX_i} E[Y|X_i]\right)\frac{1}{E[Y|X_i]}$.

En Economía y otras disciplinas estas medidas también se acostumbra definirlas para la relación determinista $Y = g(X)$. Por ejemplo, en el contexto del MRLS, si $Y = \beta_0 + \beta_1 X$, y las medidas de sensibilidad dependen del parámetro β_1 como se muestra a continuación.

Proposición.

Si se tiene la relación determinista $Y = \beta_0 + \beta_1 X$ entonces:

i) Efecto parcial: $\Delta Y \approx \beta_1 \Delta X$

ii) Elasticidad: $\eta = \beta_1\left(\frac{X}{Y}\right)$

iii) Semi-elasticidad: $\overline{\eta} = \frac{100\beta_1}{Y}$

Demostración: Si $Y = \beta_0 + \beta_1 X$ entonces $\frac{dY}{dX} = \frac{d}{dX}(\beta_0 + \beta_1 X) = \beta_1$, y aplicando las definiciones: (i) $\Delta Y \approx \beta_1 \Delta X_i$, (ii) $\eta = \beta_1\left(\frac{X}{Y}\right)$ y (iii) $\overline{\eta} = 100(\beta_1)\frac{1}{Y} = \frac{100\beta_1}{Y}$.

TRANSFORMACIÓN DE MODELOS NO LINEALES

La estimación del MRLS permite estimar fácilmente modelos no lineales que es posible convertir en modelos lineales en parámetros. Por ejemplo, $Y_i = \dfrac{1}{\beta_0 + \beta_1 X_i + u_i}$ no es modelo lineal en variables, pero tomando recíprocos se obtiene $\dfrac{1}{Y_i} = \beta_0 + \beta_1 X_i + u_i$. Si en lugar de considerar los datos (X_i, Y_i) se consideran los datos (X_i, W_i), con $W_i = \dfrac{1}{Y_i}$, el modelo a estimar es $W_i = \beta_0 + \beta_1 X_i + u_i$, que es MRLS. La figura 2.3 muestra algunos de estos modelos.

Figura 2.3

Modelo	**Forma no lineal**	**Forma lineal (en parámetros)**
Logarítmico o multiplicativo	$Y_i = \alpha_0 X_i^{\beta_1} e^{u_i}$	$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$, donde $\beta_0 = \ln(\alpha_0)$
Semilogarítmico o Lin-Log	$e^{Y_i} = \alpha_0 X_i^{\beta_1} e^{u_i}$	$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$, donde $\beta_0 = \ln(\alpha_0)$
Semilogarítmico inverso o Log-Lin	$Y_i = \alpha_0 \alpha_1^{X_i} e^{u_i}$	$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$, donde $\beta_0 = \ln(\alpha_0)$ y $\beta_1 = \ln(\alpha_1)$
Recíproco	$(Y_i - \beta_0 - u_i) X_i = \beta_1$	$Y_i = \beta_0 + \beta_1 \left(\dfrac{1}{X_i}\right) + u_i$
Logarítmico recíproco	$Y_i = \alpha_0 (\alpha_1)^{\frac{1}{X_i}} e^{u_i}$	$\ln(Y_i) = \beta_0 + \beta_1 \left(\dfrac{1}{X_i}\right) + u_i$, donde $\beta_0 = \ln(\alpha_0)$ y $\beta_1 = \ln(\alpha_1)$

Ejercicio $\boxed{\text{E7}}$.

2.2.2. Supuestos del modelo.

En el Modelo de Regresión Lineal Simple, $Y_i = \beta_0 + \beta_1 X_i + u_i$, el término de error $\boldsymbol{u_i}$ **es variable aleatoria** y consecuentemente $\boldsymbol{Y_i}$ **también es variable aleatoria**. Para poder hacer inferencia estadística sobre los parámetros β_0 y β_1 es necesario hacer algunos supuestos acerca de u_i y de su relación con X_i.

Independencia vs no correlación. Es importante recordar que: (i) las variables aleatorias $Z_1, Z_2, \ldots, Z_n$ son independientes si y sólo si $f_{Z_1, Z_2, \ldots, Z_n}(z_1, z_2, \ldots z_n) = \prod_{i=1}^{n} f_{Z_i}(z_i)$; (ii) si $Z_1, Z_2, \ldots, Z_n$ son independientes entonces también son no correlacionadas, es decir, $Cov[Z_i, Z_j] = 0$ para toda $i \neq j$; pero (iii) no correlación no implica independencia, salvo algunas excepciones (v.gr., variables aleatorias normales).

Supuestos del Modelo de Regresión Lineal Simple

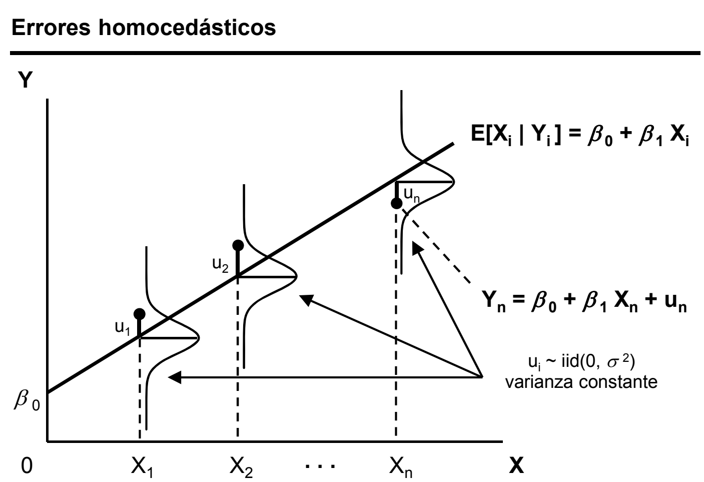
i)	La relación entre X_i y Y_i es lineal en parámetros (no necesariamente en variables).
ii)	Los valores X_i son fijos en muestras repetidas (variable no estocástica): $Cov[u_i, X_i \| X_i] = 0$. O bien, X_i y u_i son independientes: $Cov[u_i, X_i] = 0$ (si X_i es estocástica).
iii)	La media del error es cero: $E[u_i] = 0$ o $E[u_i \| X_i] = 0$ (si X_i es estocástica).
iv)	La varianza del error es constante: $Var[u_i] = \sigma^2$ o $Var[u_i \| X_i] = \sigma^2$ (si X_i es estocástica).
v)	Los errores son independientes y por tanto no están correlacionados: $Cov[u_i, u_j] = 0$ o $Cov[u_i, u_j \| X_i, X_j] = 0$ (si X_i, X_j son estocásticas), para toda $i \neq j$.
vi)	El número de observaciones (n) debe ser mayor al número de parámetros a estimar: $n > 2$.
vii)	Los valores X_1, X_2,… X_n de una muestra no pueden ser todos iguales ni existen observaciones atípicas.

El supuesto (ii) establece la independencia estadística entre X_i y u_i (sin importar si X_i es variable aleatoria o no), de modo que la distribución condicional de $\{u_i | X_i\}$ y la distribución incondicional de u_i son la misma. Para simplificar la notación se considerará la distribución incondicional de u_i.

Los supuestos (iii), (iv) y (v) pueden resumirse estableciendo que los errores u_1, u_2,…, u_n deben ser independientes e idénticamente distribuidos con media cero y varianza constante σ^2, es decir, $\boldsymbol{u_i \sim iid(0, \sigma^2)}$.

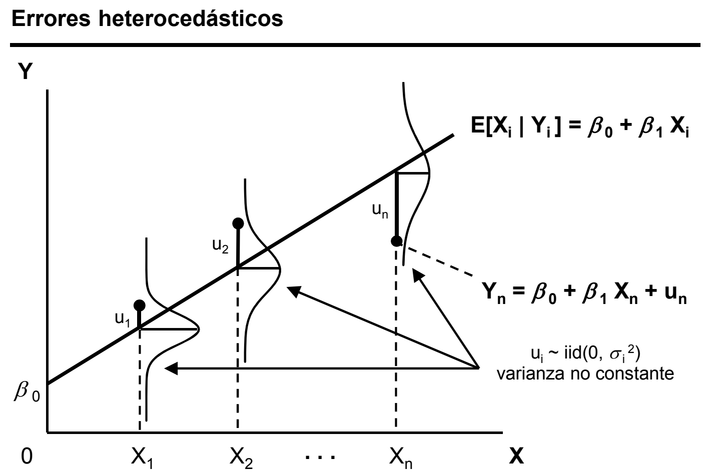
El supuesto (iv) de **varianza constante** se conoce también como supuesto de **homocedasticidad** (ver figura 2.4).

Figura 2.4



Si la **varianza no es constante** se dice que hay **heterocedasticidad**, es decir, $Var[u_i|X_i] = \sigma_i^2$, donde σ_i^2 va cambiando dependiendo del nivel X_i (ver figura 2.5).

Figura 2.5



De los supuestos (ii) y (iii) se puede deducir fácilmente que $E[u_i^2] = \sigma^2$, pues como $E[u_i] = 0$ y $Var[u_i] = \sigma^2$, entonces $Var[u_i] = E[u_i^2] - (E[u_i])^2 = E[u_i^2] - 0^2 = E[u_i^2] = \sigma^2$.

Para determinar esperanza, varianza y covarianza condicionales de $Y_1, Y_2, \ldots, Y_n$, dados $X_1, X_2, \ldots, X_n$, es útil recordar las principales propiedades de estos operadores.

Teorema

Si $X_1, X_2, \ldots, X_m$ y $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias, $c_0, c_1, \ldots, c_m$ y $d_0, d_1, \ldots, d_n$ son constantes, entonces:

i) $E\left[\sum_{i=1}^{n} d_i Y_i\right] = \sum_{i=1}^{n} d_i E[Y_i]$ (Operador Lineal);

y en particular $E[d_0 + d_1 Y_1] = d_0 + d_1 E[Y_1]$

ii) $Var\left[\sum_{i=1}^{n} d_i Y_i\right] = \sum_{i=1}^{n} d_i^2 Var[Y_i] + 2\sum_{i<j} d_i d_j Cov[Y_i, Y_j] = \sum_{i=1}^{n}\sum_{j=1}^{m} d_i d_j Cov[Y_i, Y_j]$;

y en particular $Var[d_0 + d_1 Y_1] = d_1^2 Var[Y_1]$

iii) $Cov\left[\sum_{i=1}^{m} c_i X_i, \sum_{j=1}^{n} d_j Y_j\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} c_i d_j Cov[X_i, Y_j]$ (Operador Bilineal);

y en particular $Cov[c_0 + c_1 X_1, d_0 + d_1 Y_1] = c_1 d_1 Cov[X_1, Y_1]$

Proposición

Bajo los supuestos del MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$:

i) $E[Y_i | X_i] = \beta_0 + \beta_1 X_i$

ii) $Var[Y_i | X_i] = \sigma^2$

iii) $Cov[Y_i, Y_j | X_i, X_j] = 0$ para toda $i \neq j$

Ejercicio $\boxed{\text{E8}}$.

Para determinar esperanza, varianza y covarianza incondicionales de Y_1, Y_2,…, Y_n es necesario recordar las *Leyes de Valor Esperado Iterado*.

Teorema. Leyes de Valor Esperado Iterado.

Si X, Y y Z son variables aleatorias, entonces:

i) $E[Y] = E[E[Y|X]]$

ii) $Var[Y] = E[Var[Y|X]] + Var[E[Y|X]]$

iii) $Cov[Y, Z] = E[Cov[Y, Z|X]] + Cov[E[Y|X], E[Z|X]]$

Proposición

Bajo los supuestos del MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$, considerando que X_i es estocástica:

i) $E[Y_i] = \beta_0 + \beta_1 E[X_i]$

ii) $Var[Y_i] = \sigma^2 + \beta_1^2 Var[X_i]$

iii) $Cov[Y_i, Y_j] = \beta_1^2 Cov[X_i, X_j]$ para toda $i \neq j$

2.2.3. Definición del MRLS.

Def. Modelo de Regresión Lineal Simple (MRLS).

Si (X_1, Y_1), (X_2, Y_2),…, (X_n, Y_n) son n observaciones de la variable dependiente Y y la variable explicativa X ($n > 2$ y no todas las X_i's toman el mismo valor), cuya FRP es la recta $E[Y|X_i] = \beta_0 + \beta_1 X_i$, entonces el Modelo de Regresión Lineal Simple (MRLS) supone que valores Y_i pueden ser modelados como la suma de esta media condicional y un término de error u_i de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 X_i + u_i,$$

donde los valores X_i son fijos y $u_i \sim iid(0, \sigma^2)$.

2.3. Estimadores de Mínimos Cuadrados.

El problema de **estimación puntual** parte del supuesto de que se tiene una muestra aleatoria $Y_1, Y_2,\ldots, Y_n$ de una población caracterizada por el parámetro θ y el objetivo es encontrar una estadística $\hat{\theta} = g(Y_1, Y_2, \ldots, Y_n)$ llamada **estimador**, que para una muestra observada $y_1, y_2,\ldots, y_n$, permita obtener una **estimación** o aproximación del verdadero valor de θ. A su vez, $\hat{\theta}$ es una variable aleatoria (su valor cambia de muestra en muestra) y su distribución de probabilidad se denomina **distribución de muestreo**.

Los estimadores más populares en Inferencia Estadística son los *Estimadores de Máxima Verosimilitud* (EMV), sin embargo, para poder determinarlos es indispensable suponer un modelo (o distribución) de probabilidad para Y_i que dependa de θ, $f_{Y_i}(y_i|\theta)$, que permita calcular y maximizar la función de verosimilitud $L(\theta) = \prod_{i=1}^{n} f_{Y_i}(y_i|\theta)$.

En el MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$ **los parámetros a estimar son β_0 y β_1**, sin embargo, ninguno de los supuestos del MRLS establece una distribución ni para los términos de error u_i, ni para la variable dependiente Y_i. Adicionalmente, como en el MRLS se supone que $u_i \sim iid(0, \sigma^2)$, **también el parámetro σ^2 se debe estimar**.

Ante esta dificultad es posible aplicar otro método de estimación de la Inferencia Estadística denominado *Mínimos Cuadrados Ordinarios* (MCO), que fue formulado a finales del siglo XVIII por Karl Friedrich Gauss (1777-1855) y por Adrien-Marie Legendre (1752-1833).

2.3.1. Método de Mínimos Cuadrados Ordinarios (MCO).

Def. Función de Regresión Muestral (FRM). Es el lugar geométrico de las *estimaciones* de las medias condicionales de la variable dependiente para los valores fijos de la variable explicativa, es decir, $\hat{Y}_i = \hat{E}[Y|X_i]$.

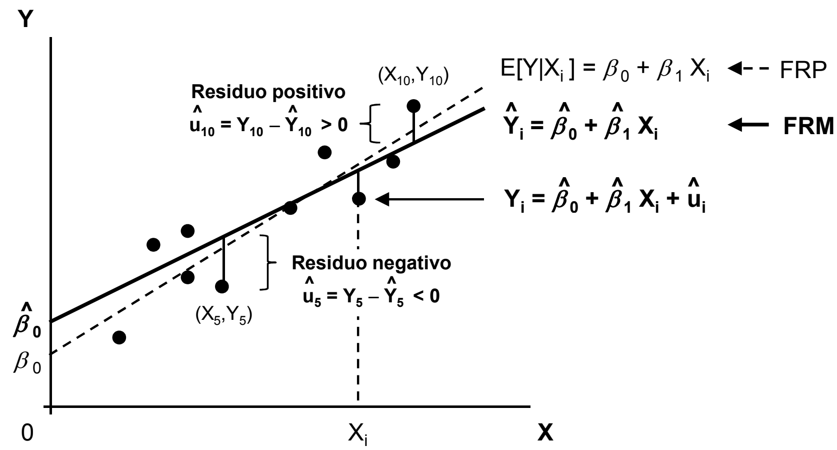
En el caso de la FRP lineal $E[Y|X_i] = \beta_0 + \beta_1 X_i$, su respectiva **FRM lineal** es $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$, donde $\hat{\beta}_0$ y $\hat{\beta}_1$ son los estimadores de β_0 y β_1, respectivamente.

Importante: En el contexto del MRLS la FRP $E[Y|X_i] = \beta_0 + \beta_1 X_i$ es una recta teórica no observable, pero a partir del conjunto de datos $(X_1, Y_1), (X_2, Y_2),\ldots, (X_n, Y_n)$ se buscarán $\hat{\beta}_0$ y $\hat{\beta}_1$ que permitan determinar la FRM $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. Esta recta estimada cambiará con cada conjunto de datos observados $(x_1, y_1), (x_2, y_2),\ldots, (x_n, y_n)$.

La formulación teórica del MRLS supone que Y_i se puede expresar como la suma de la FRP $E[Y|X_i] = \beta_0 + \beta_1 X_i$ y el término de error u_i, sin embargo, estos componentes son no observables. Para estimar los parámetros del MRLS se remplazará cada componente, por su estimador para expresar a Y_i como la suma de la FRM $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ y del *residuo* $\hat{u}_i$ como se muestra a continuación (ver figura 2.6):

$$Y_i = \hat{Y}_i + \hat{u}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i$$

Figura 2.6



Def. Residuos.

$$\hat{u}_i = Y_i - \hat{Y}_i$$

En general, en el Análisis de Regresión, los residuos $\hat{u}_i$ son los estimadores de los términos de error u_i.

Como la variación $Y_i - \hat{Y}_i$ puede ser positiva o negativa, en lugar de tomar su valor absoluto se prefiere elevarla al cuadrado y trabajar con los cuadrados de los residuos $\hat{u}_i^2 = (Y_i - \hat{Y}_i)^2$. Los Métodos de Mínimos Cuadrados buscan minimizar de manera conjunta la suma de los cuadrados de los residuos (Mínimos Cuadrados Ordinarios) o un promedio ponderado de ellos (Mínimos Cuadrados Generalizados).

Método de Mínimos Cuadrados Ordinarios (MCO).

Consiste en minimizar la suma de cuadrados de los residuos $SCR = \sum_{i=1}^{n} \hat{u}_i^2$

Para el MRLS los residuos son de la forma $\hat{u}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$, de modo que la suma de cuadrados de residuos es $SCR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2 = g\left(\hat{\beta}_0, \hat{\beta}_1\right)$.

Teorema. Estimadores MCO del MRLS

Para el MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde $u_i \sim iid(0, \sigma^2)$:

i) $\hat{\beta}_1 = \dfrac{S_{XY}}{S_X^2}$ y $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$ son los estimadores MCO de β_1 y β_0, respectivamente.

ii) $\hat{\sigma}^2 = \dfrac{1}{n-2} \sum_{i=1}^{n} \hat{u}_i^2$ es estimador insesgado de σ^2.

Los estimadores MCO de $\hat{\beta}_0$ y $\hat{\beta}_1$ surgen directamente en el proceso de minimización de la suma de cuadrados de residuos del MRLS. El estimador $\hat{\sigma}^2 = \dfrac{1}{n-2} \sum_{i=1}^{n} \hat{u}_i^2$ se elige así por conveniencia para poder hacer inferencia estadística al suponer normalidad en los términos de error u_i.

Al aplicar el método de MCO para estimar el MRLS aparecen 2 ecuaciones, llamadas *ecuaciones normales*, que son útiles al analizar las propiedades de $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\sigma}^2$.

Corolario. Ecuaciones Normales del MRLS

i) $\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i$

ii) $\sum_{i=1}^{n} X_i Y_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2$

EXCEL. *Estimadores MCO de* β_0 y β_1. La funciones =INTERSECCION.LINEAL(rangoY,rangoX) y =PENDIENTE(rangoY,rangoX) calculan $\hat{\beta}_0$ y $\hat{\beta}_1$, respectivamente, si los datos de la variable dependiente Y están en el "rangoY" y los datos de la variable explicativa X están en el "rangoX".

Ejercicios $\boxed{\text{E9}}$, $\boxed{\text{E10}}$ y $\boxed{\text{E11}}$.

2.3.2. Modelo de regresión ajustado, valores ajustados y residuos.

Si se considera el MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$, para un conjunto de datos (X_i, Y_i) es fácil calcular los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$. A partir de esos valores es posible definir un modelo determinista para modelar Y al considerar distintos valores de X, comparar los valores reales contra los valores ajustados y analizar esas discrepancias.

Def. MRLS ajustado. Es la relación determinista $Y = \hat{\beta}_0 + \hat{\beta}_1 X = g(X)$, también llamada recta de regresión lineal.

Def. Valores ajustados. Son los valores estimados de Y_i, es decir, $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$.

A partir de los valores reales (Y_i) y de los valores ajustados ($\hat{Y}_i$) es fácil calcular los residuos $\hat{u}_i = Y_i - \hat{Y}_i$.

Propiedades del MRLS ajustado por MCO.

Si el MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$ se estima por MCO, entonces:

i) La recta de regresión lineal siempre pasa por $(\overline{X}, \overline{Y})$, es decir, $\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X}$.

ii) La media de los valores ajustados coincide con la media de los valores reales, es decir, $\overline{\hat{Y}} = \overline{Y}$.

Demostración: (i) Evaluando $g(X) = \hat{\beta}_0 + \hat{\beta}_1 X$ en $\overline{X}$ y sustituyendo el estimador MCO de $\hat{\beta}_0$ se obtiene $g(\overline{X}) = \hat{\beta}_0 + \hat{\beta}_1 \overline{X} = (\overline{Y} - \hat{\beta}_1 \overline{X}) + \hat{\beta}_1 \overline{X} = \overline{Y}$, por lo tanto $\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X}$; y (ii) $\overline{\hat{Y}} = \frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 X_i) = \frac{1}{n}(n\hat{\beta}_0) + \hat{\beta}_1\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \hat{\beta}_0 + \hat{\beta}_1 \overline{X}$, y de nuevo sustituyendo el estimador MCO de $\hat{\beta}_0$ se obtiene $\overline{\hat{Y}} = (\overline{Y} - \hat{\beta}_1 \overline{X}) + \hat{\beta}_1 \overline{X} = \overline{Y}$.

Propiedades de los residuos.

Si el MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$ se estima por MCO y $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$:

i) La suma de los residuos es cero: $\sum_{i=1}^{n} \hat{u}_i = 0$.

ii) Los residuos y los regresores no están correlacionados: $\sum_{i=1}^{n} \hat{u}_i X_i = 0$

iii) Los residuos y los valores ajustados no están correlacionados: $\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = 0$.

Ejercicios $\boxed{\text{E12}}$ y $\boxed{\text{E13}}$.

2.4. Propiedades de los estimadores MCO.

2.4.1. Teorema de Gauss-Markov.

Def. Estimador Lineal

Se dice que $\hat{\theta} = g(Y_1, Y_2, \ldots, Y_n)$ es estimador lineal del parámetro θ si g es función lineal, es decir, $\hat{\theta} = c_1 Y_1 + c_2 Y_2 + \cdots + c_n Y_n = \sum_{i=1}^{n} c_i Y_i$, donde $c_1, c_2, \ldots, c_n \in \mathbf{R}$.

Por ejemplo, la media muestral $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n}\left(\frac{1}{n}\right) Y_i$ es un estimador lineal, donde $c_i = \frac{1}{n}$ para $i = 1, 2, \ldots n$. Los estimadores MCO del MRLS también lo son.

Lema

Los estimadores MCO del MRLS $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores lineales:

i) $\hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i$ con $c_i = \dfrac{X_i - \overline{X}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$

ii) $\hat{\beta}_0 = \sum_{i=1}^{n} d_i Y_i$ con $d_i = \frac{1}{n} - \overline{X} c_i$

Demostración: (i) Se sabe que $\hat{\beta}_1 = \dfrac{S_{XY}}{S_X^2}$, donde $S_X^2 = \dfrac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$ y

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i - \overline{X})Y_i - \sum_{i=1}^{n}(X_i - \overline{X})\overline{Y}\right] = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})Y_i$$

pues $\sum_{i=1}^{n}(X_i - \overline{X}) = \sum_{i=1}^{n} X_i - n\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = 0$. Entonces $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})Y_i}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$, es decir, $\hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i$ con $c_i = \dfrac{X_i - \overline{X}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$. (ii) Sustituyendo $\hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i$ en $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$ se obtiene $\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} Y_i - \overline{X}\sum_{i=1}^{n} c_i Y_i = \sum_{i=1}^{n}\left(\frac{1}{n} - \overline{X} c_i\right) Y_i = \sum_{i=1}^{n} d_i Y_i$ con $d_i = \frac{1}{n} - \overline{X} c_i$.

Teorema de Gauss-Markov

Bajo los supuestos del MRLS, los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$ son los Mejores Estimadores Lineales Insesgados (MELI)

De este teorema hay que aclarar lo siguiente:

- $\hat{\beta}_0$ y $\hat{\beta}_1$ efectivamente son **estimadores lineales** como se demostró previamente.
- $\hat{\beta}_0$ y $\hat{\beta}_1$ son los **"mejores"** estimadores lineales, pues dentro de su tipo, son los que tienen **menor varianza**, es decir, $Var\left[\hat{\beta}_j\right] \leq Var\left[\hat{\beta}_j^*\right]$ para cualquier otro estimador lineal $\hat{\beta}_j^*$ de β_j, $j = 0, 1$.
- $\hat{\beta}_0$ y $\hat{\beta}_1$ **insesgados** significa que $E\left[\hat{\beta}_j\right] = \beta_j$, $j = 0, 1$.

Teorema

Bajo los supuestos del MRLS, los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$ son consistentes.

Hay que recordar que $\hat{\theta} = g(Y_1, Y_2, \ldots, Y_n)$ es **estimador consistente** si $\hat{\theta} \xrightarrow{p} \theta$, es decir, si $\lim_{n\to\infty} P\left[\left|\hat{\theta}_n - \theta\right| \geq \varepsilon\right] = 0$ o bien si $\lim_{n\to\infty} P\left[\left|\hat{\theta}_n - \theta\right| < \varepsilon\right] = 1$ para todo $\varepsilon > 0$.

2.4.2. Matriz de Varianzas y Covarianzas.

Teorema

Si $\hat{\beta}_0$ y $\hat{\beta}_1$ son los estimadores MCO del MRLS, entonces:

$$Var\left[\hat{\beta}_0\right] = \frac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sigma^2, \quad Var\left[\hat{\beta}_1\right] = \frac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} \quad \text{y} \quad Cov\left[\hat{\beta}_0, \hat{\beta}_1\right] = -\frac{\overline{X}\sigma^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

Note cómo $Var\left[\hat{\beta}_0\right]$, $Var\left[\hat{\beta}_1\right]$ y $Cov\left[\hat{\beta}_0, \hat{\beta}_1\right]$ dependen de σ^2. Sus respectivos estimadores, denotados por $S_{\hat{\beta}_0}^2$, $S_{\hat{\beta}_1}^2$ y $S_{\hat{\beta}_0, \hat{\beta}_1}$, respectivamente, se definen del mismo modo remplazando a σ^2 por $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$, que es **estimador insesgado de $\boldsymbol{\sigma^2}$**.

Ejercicios $\boxed{\text{E14}}$, $\boxed{\text{E15}}$ y $\boxed{\text{E16}}$.

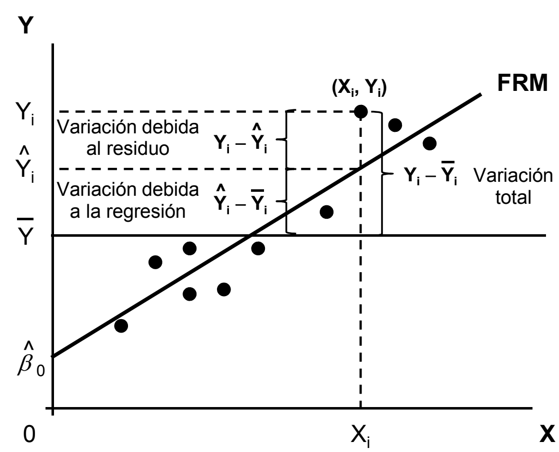
2.5. Bondad de ajuste.

En el contexto del MRLS la **bondad de ajuste** quiere decir qué tan bien se ajusta la recta de regresión lineal al conjunto de datos observados. La bondad de ajuste será mejor en la medida en que todos los residuos $\hat{u}_i = Y_i - \hat{Y}_i$ sean pequeños. Para medir dicha bondad de ajuste se define el *coeficiente de determinación* (R^2).

2.5.1. Coeficiente de determinación (R^2).

Si para la variable de interés Y se tienen los datos $Y_1, Y_2, \ldots, Y_n$, se sabe que el mejor estimador de la media poblacional es la media muestral $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. En el contexto del MRLS, adicionalmente se cuenta con los datos $X_1, X_2, \ldots, X_n$ de la variable explicativa X, que permiten modelar $Y_i = \hat{Y}_i + \hat{u}_i$ donde $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. De esta manera la *variación total* $Y_i - \overline{Y}$ se puede expresar como suma de la *variación debida a la regresión* (o variación explicada por la regresión) $\hat{Y}_i - \overline{Y}$, más la *variación debida al residuo* $Y_i - \hat{Y}_i$, como se muestra en la figura 2.7. Esta misma relación se mantiene al considerar las variaciones al cuadrado.

Figura 2.7



Proposición

Si el MRLS se estima por MCO, entonces:

i) $Y_i - \overline{Y} = \left(\hat{Y}_i - \overline{Y}\right) + \left(Y_i - \hat{Y}_i\right)$

ii) $\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2 + \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$

Demostración: (i) Se obtiene directamente sumando y restando $\hat{Y}_i$. (ii) Por propiedades de la estimación MCO del MRLS $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i$, $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$, $\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X}$ y $\overline{\hat{Y}} = \overline{Y}$, de modo que $\hat{Y}_i - \overline{\hat{Y}}_i = \hat{Y}_i - \overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_i - \left(\hat{\beta}_0 + \hat{\beta}_1 \overline{X}\right) = \hat{\beta}_1 \left(X_i - \overline{X}\right)$ entonces elevando al cuadrado y sumando para *i* = 1, 2,…, *n*, $\sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2 = \hat{\beta}_1^2 \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$. Por su parte, $Y_i - \overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i - \left(\hat{\beta}_0 + \hat{\beta}_1 \overline{X}\right) = \hat{\beta}_1 \left(X_i - \overline{X}\right) + \hat{u}_i$, entonces elevando al cuadrado $\left(Y_i - \overline{Y}\right)^2 = \hat{\beta}_1{}^2 \left(X_i - \overline{X}\right)^2 + 2\hat{\beta}_1 \left(X_i - \overline{X}\right)\hat{u}_i + \hat{u}_i^2$ y sumando para *i* = 1, 2,…, *n*, se tiene que $\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \hat{\beta}_1{}^2 \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 + 2\hat{\beta}_1 \sum_{i=1}^{n}\left(X_i - \overline{X}\right)\hat{u}_i + \sum_{i=1}^{n}\hat{u}_i^2$. Por propiedades de los residuos $\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\hat{u}_i = \sum_{i=1}^{n} X_i \hat{u}_i - \overline{X}\sum_{i=1}^{n}\hat{u}_i = 0 - \overline{X}(0) = 0$, por lo que se concluye que $\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2 + \sum_{i=1}^{n}\hat{u}_i^2$.

Esta última expresión establece que la **suma de cuadrados totales** $SCT = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$ es igual a la **suma de cuadrados explicada** $SCE = \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2$ más la **suma de cuadrados de los residuos** $SCR = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\hat{u}_i^2$, es decir:

$$\boxed{SCT = SCE + SCR}$$

Dividiendo esta última expresión entre *SCT* se obtiene $1 = \frac{SCE}{SCT} + \frac{SCR}{SCT}$, es decir, el 100% de la variación total se puede descomponer como la proporción o porcentaje explicado por la regresión $\frac{SCE}{SCT}$ más la proporción o porcentaje explicado por los residuos $\frac{SCR}{SCT}$.

Def. Coeficiente de Determinación (R^2)

$$\boxed{R^2 = \frac{SCE}{SCT} = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}$$

El **coeficiente de determinación** es una **medida de bondad de ajuste** que indica la proporción o porcentaje de la variación cuadrática total que es explicada por la regresión ($0 \leq R^2 \leq 1$). Si $R^2 \to 1$ se tiene un buen ajuste (residuos pequeños), en cambio, si $R^2 \to 0$ se tiene un mal ajuste (residuos grandes).

Propiedades del Coeficiente de Determinación

i) $$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum_{i=1}^{n} \hat{u}_i^2}{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2}$$

ii) $$R^2 = \hat{\beta}_1^2 \left(\frac{S_X^2}{S_Y^2} \right)$$

iii) $$R^2 = \left(r_{XY}\right)^2$$

2.5.2. Análisis de Varianza (ANOVA).

En el Análisis de Regresión a la descomposición $SCT = SCE + SCR$ se le conoce como **Análisis de Varianza** (ANOVA por sus siglas en inglés) y consiste en presentar estos componentes en una tabla como se muestra en la figura 2.8.

Figura 2.8

Fuente de variación	**Suma de cuadrados**	**g.l.**	**Suma de cuadrados medios**
Regresión (explicada)	$SCE = \sum_{i=1}^{n} \left(\hat{Y}_i - \overline{Y}\right)^2 = \hat{\beta}_1^2 \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$	1	$CME = \frac{SCE}{1} = \hat{\beta}_1^2 \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$
Residuos	$SCR = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} \hat{u}_i^2$	$n-2$	$CMR = \frac{SCR}{n-2} = \frac{1}{n-2} \sum_{i=1}^{n} \hat{u}_i^2 = \hat{\sigma}^2$
Total	$SCT = \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$	$n-1$	$CMT = \frac{SCT}{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2 = S_Y^2$

A cada suma de cuadrados corresponden ciertos **grados de libertad (g.l.)**, que son el número de observaciones independientes requeridas para calcular dichas estadísticas. *SCT* tiene $n - 1$ grados de libertad pues pierde un grado de libertad en el cálculo de $\overline{Y}$, *SCE* tiene 1 grado de libertad pues sólo depende de $\hat{\beta}_1$ (las X_i's son conocidas), y *SCR* tiene $n - 2$ grados de libertad pues pierde dos grados de libertad en el cálculo de $\hat{\beta}_0$ y $\hat{\beta}_1$. La **suma de cuadrados medios** es simplemente la suma de cuadrados totales entre sus respectivos grados de libertad.

Ejercicios $\boxed{\text{E17}}$ y $\boxed{\text{E18}}$.

2.6. Estimadores de Máxima Verosimilitud (EMV).

Para estimar el MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$ bajo el Método de Máxima Verosimilitud es necesario suponer una distribución de probabilidad para cada término de error u_i. Para cumplir con el resto de los supuestos del MRLS la distribución de u_i debe tener media cero y varianza constante σ^2. Por practicidad y conveniencia, para fines de inferencia estadística se considerará a la **Distribución Normal**, es decir, $\boldsymbol{u_i \sim N(0, \sigma^2)}$.

Teorema. Transformaciones Lineales de Distribuciones Normales

Si $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias independientes con $Y_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$, y $c_0, c_1, \ldots, c_n$ son constantes entonces $\sum_{i=1}^{n} c_i Y_i \sim N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right)$, y en particular, $c_0 + c_1 Y_1 \sim N(c_0 + c_1\mu_1, c_1^2 \sigma_1^2)$

Def. Modelo de Regresión Lineal Simple Normal (MRLSN).

Si $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ son n observaciones de la variable dependiente Y y la variable explicativa X ($n > 2$ y no todas las X_i's toman el mismo valor), cuya FRP es la recta $E[Y|X_i] = \beta_0 + \beta_1 X_i$, entonces el Modelo de Regresión Lineal Simple Normal (MRLSN) supone que valores Y_i pueden ser modelados como la suma de esta media condicional y un término de error normal u_i de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 X_i + u_i\,,$$

donde los valores X_i son fijos y $u_i \sim iid\ N(0, \sigma^2)$.

Proposición

Bajo los supuestos del MRLSN $\{Y|X_i\} \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$

Demostración: Bajo los supuestos del MRLSN los valores X_i son fijos e independientes de los errores u_i, de modo que $E[u_i|X_i] = E[u_i] = 0$ y $Var[u_i|X_i] = Var[u_i] = \sigma^2$. Como $Y_i = \beta_0 + \beta_1 X_i + u_i$ es transformación lineal de una variable aleatoria normal, entonces Y_i también es normal con media $E[Y|X_i] = E[\beta_0 + \beta_1 X_i + u_i|X_i] = \beta_0 + \beta_1 X_i$ y varianza $Var[Y|X_i] = Var[\beta_0 + \beta_1 X_i + u_i|X_i] = \sigma^2$, por lo tanto, $\{Y|X_i\} \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$.

Teorema. EMV del MRLSN

Para el MRLSN $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde $u_i \sim iid\ N(0,\ \sigma^2)$:

i) $\hat{\beta}_1 = \dfrac{S_{XY}}{S_X^2}$ y $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$ son los EMV de β_1 y β_0, respectivamente; y

ii) $\hat{\sigma}_*^2 = \dfrac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2$ es el EMV de σ^2.

Ejercicio $\boxed{\text{E19}}$.

Este teorema establece que **los EMV y los estimadores MCO de β_1 y β_0 son los mismos**. De modo que suponer normalidad en los errores u_i garantiza que $\hat{\beta}_0$ y $\hat{\beta}_1$ posean la propiedad de **invarianza**, así como las **propiedades asintóticas de los EMV**, particularmente **eficiencia y consistencia**. Desafortunadamente $\hat{\sigma}_*^2 = \dfrac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2$, el EMV de σ^2, no es insesgado, por eso se preferirá estimar a σ^2 mediante el estimador insesgado $\hat{\sigma}^2 = \dfrac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2$.

Hay que recordar que $\hat{\theta} = g(Y_1, Y_2, \ldots, Y_n)$ es **estimador eficiente** si (i) es insesgado; y (ii) es de mínima varianza, es decir, su varianza alcanza la Cota Inferior de Crámer-Rao: $Var[\hat{\theta}] = CICR = \dfrac{1}{nI(\theta)}$, donde $I(\theta) = E\left[\left(\dfrac{\partial}{\partial\theta} \ln f_Y(Y|\theta)\right)^2\right] = -E\left[\dfrac{\partial^2}{\partial\theta^2} \ln f_Y(Y|\theta)\right]$ es la Información de Fisher.

Distribuciones de muestreo del MRLSN (σ^2 conocida)

Si $\hat{\beta}_0$ y $\hat{\beta}_1$ son los EMV del MRLSN y $\hat{\sigma}^2 = \dfrac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2$, entonces:

i) $Z_0 = \dfrac{\hat{\beta}_0 - \beta_0}{\sigma_{\hat{\beta}_0}} \sim N(0,1)$, donde $\sigma_{\hat{\beta}_0}^2 = \dfrac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n}(X_i - \overline{X})^2}\sigma^2$

ii) $Z_1 = \dfrac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim N(0,1)$, donde $\sigma_{\hat{\beta}_1}^2 = \dfrac{\sigma^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$

iii) $J = \dfrac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$

Estas distribuciones de muestreo son fáciles de deducir al recordar que $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores lineales, es decir, existen $c_1, c_2,\ldots, c_n$ y $d_1, d_2,\ldots, d_n$ tales que $\hat{\beta}_0 = \sum_{i=1}^{n} d_i Y_i$ y $\hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i$. Como Y_i tiene una distribución normal, entonces $\hat{\beta}_0$ y $\hat{\beta}_1$ también tienen una distribución normal, de hecho, $\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2_{\hat{\beta}_0}\right)$ y $\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2_{\hat{\beta}_1}\right)$.

Deducir la distribución de muestreo asociada a $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$ no es tan directo, pero la lógica es la misma que se utiliza para demostrar que si $Y_1, Y_2,\ldots, Y_n$ es muestra aleatoria con $Y_i \sim N\left(\mu_Y, \sigma_Y^2\right)$, $i = 1, 2,\ldots, n$, entonces $\frac{(n-1)S_Y^2}{\sigma_Y^2} \sim \chi^2(n-1)$.

El cálculo de Z_0 y Z_1 depende de $\sigma^2_{\hat{\beta}_0}$ y $\sigma^2_{\hat{\beta}_1}$, que a su vez dependen de σ^2. En realidad la varianza de los errores σ^2 no se conoce y sólo se aproxima mediante $\hat{\sigma}^2$, modificando la distribución de muestreo asociada a $\hat{\beta}_0$ y $\hat{\beta}_1$.

Distribuciones de muestreo del MRLSN (σ^2 desconocida)

Si $\hat{\beta}_0$ y $\hat{\beta}_1$ son los EMV del MRLSN y $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$, entonces:

i) $$T_0 = \frac{\hat{\beta}_0 - \beta_0}{S_{\hat{\beta}_0}} \sim t(n-2), \text{ donde } S^2_{\hat{\beta}_0} = \frac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\hat{\sigma}^2$$

ii) $$T_0 = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t(n-2), \text{ donde } S^2_{\hat{\beta}_1} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

2.7. Estimación por intervalo y pruebas de hipótesis.

Es importante notar que en el contexto del MRLSN las estadísticas Z_0, Z_1, J, T_0 y T_1 son **cantidades pivotales** (estadísticas que involucran a los parámetros de interés β_0, β_1 y σ^2, pero cuya distribución de probabilidad no depende de dichos parámetros), que permiten construir fácilmente **intervalos de confianza** mediante el Método Pivotal y que permiten ser consideradas como **estadísticas de prueba** en la formulación de **pruebas de hipótesis**.

2.7.1. Intervalos de confianza sobre los parámetros del MRLSN.

La figura 2.9 muestra el resumen de los intervalos simétricos (misma área en cada cola) con nivel de significancia 1 – α para estimar por intervalo los parámetros β_0, β_1 y σ^2, bajo los supuestos del MRLSN.

Figura 2.9

Parámetro	**Pivote**	**Intervalo de confianza simétrico con nivel de confianza 1 – α**
β_0	$T_0 = \dfrac{\hat{\beta}_0 - \beta_0}{S_{\hat{\beta}_0}} \sim t(n-2)$	$\hat{\beta}_0 \pm t_{n-2,\frac{\alpha}{2}} S_{\hat{\beta}_0}$ donde $S^2_{\hat{\beta}_0} = \dfrac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\hat{\sigma}^2$
β_1	$T_1 = \dfrac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t(n-2)$	$\hat{\beta}_1 \pm t_{n-2,\frac{\alpha}{2}} S_{\hat{\beta}_1}$ donde $S^2_{\hat{\beta}_1} = \dfrac{\hat{\sigma}^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$
σ^2	$J = \dfrac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$	$\left[\dfrac{(n-2)\hat{\sigma}^2}{\chi^2_{n-2,\frac{\alpha}{2}}}, \dfrac{(n-2)\hat{\sigma}^2}{\chi^2_{n-2,1-\frac{\alpha}{2}}}\right]$

Es útil recordar que la distribución t de Student se puede aproximar mediante la distribución Normal Estándar si sus grados de libertad tienden a infinito, es decir, para muestras grandes los cuantiles $t_{n-2,\frac{\alpha}{2}}$ pueden ser remplazados por $z_{n-2,\frac{\alpha}{2}}$.

Ejercicios $\boxed{\text{E20}}$ y $\boxed{\text{E21}}$.

2.7.2. Pruebas de hipótesis acerca de los parámetros del MRLSN.

De la relación entre los intervalos de confianza y las regiones de rechazo de las pruebas de hipótesis es fácil inferir las regiones de rechazo de las pruebas de hipótesis para β_0, β_1 y σ^2, bajo los supuestos del MRLSN.

Tomando en cuenta que β_1 es relevante en la descripción de la relación entre la variable de dependiente Y y la variable explicativa X (v.gr., efecto parcial, elasticidad o semi-elasticidad dependen de β_1), es útil probar H_0: $\beta_1 = \beta_{1,0}$ vs H_1: $\beta_1 \neq \beta_{1,0}$, donde $\beta_{1,0}$ es un valor particular del espacio paramétrico de β_1.

La figura 2.10 muestra estadísticos de prueba y regiones de rechazo (RR) para pruebas de tamaño α (de una y dos colas) para β_0, β_1 y σ^2 bajo los supuestos del MRLSN.

Figura 6.6

H_0:	vs. H_1:	$RR_\alpha =$	Estadístico de prueba
$\beta_0 = \beta_{0,0}$	$\begin{cases}\beta_0 < \beta_{0,0}\\ \beta_0 > \beta_{0,0}\\ \beta_0 \neq \beta_{0,0}\end{cases}$	$\begin{cases}t_0 < -t_{n-2,\alpha}\\ t_0 > t_{n-2,\alpha}\\ \lvert t_0\rvert > t_{n-2,\frac{\alpha}{2}}\end{cases}$	$T_0 = \dfrac{\hat{\beta}_0 - \beta_{0,0}}{S_{\hat{\beta}_0}} \sim t(n-2)$
$\beta_1 = \beta_{1,0}$	$\begin{cases}\beta_1 < \beta_{1,0}\\ \beta_1 > \beta_{1,0}\\ \beta_1 \neq \beta_{1,0}\end{cases}$	$\begin{cases}t_1 < -t_{n-2,\alpha}\\ t_1 > t_{n-2,\alpha}\\ \lvert t_1\rvert > t_{n-2,\frac{\alpha}{2}}\end{cases}$	$T_1 = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{S_{\hat{\beta}_1}} \sim t(n-2)$
$\sigma^2 = \sigma_0^2$	$\begin{cases}\sigma^2 < \sigma_0^2\\ \sigma^2 > \sigma_0^2\\ \sigma^2 \neq \sigma_0^2\end{cases}$	$\begin{cases}j < \chi^2_{n-2,1-\alpha}\\ j > \chi^2_{n-2,\alpha}\\ j < \chi^2_{n-2,1-\frac{\alpha}{2}}\ ,\ j > \chi^2_{n-2,\frac{\alpha}{2}}\end{cases}$	$J = \dfrac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$

Las pruebas para β_0 y β_1 se denominan **Pruebas T**.

La mayoría de los paquetes estadísticos calculan la estadística de prueba bajo la hipótesis nula y la acompañan de su respectivo **valor-p**. Sobre el valor-p hay que recordar que:

- Se calcula como la probabilidad de obtener un valor de la estadística de prueba tan extremo como el observado al evaluar la región de rechazo dado H_0 verdadero.
- Se rechaza H_0 siempre que $\alpha >$ valor-p.
- Si valor-p $\to 0$ entonces H_0 siempre se rechaza.

Desde el punto de vista del modelaje Econométrico es relevante:

- **Rechazar H_0: $\beta_0 = 0$**, pues de lo contrario es necesario evaluar la necesidad de hacer la estimación del MRLS sin intercepto $Y_i = \beta_1 X_i + u_i$, con la posibilidad de que los residuos no sumen cero o que el coeficiente de determinación deje de tener interpretación como porcentaje (ver problema 11 de la Tarea 1); y
- **Rechazar H_0: $\beta_1 = 0$**, pues de lo contrario se debería hacer la estimación de $Y_i = \beta_0 + u_i$, es decir, la variable X en realidad no explica a la variable de interés Y.

Si se rechaza H_0: $\beta_1 = 0$ entonces hay evidencia estadística de que la variable X realmente explica a la variable de interés Y. Una manera alternativa de probar H_0: $\beta_1 = 0$ es considerar la suma de cuadrados medios de la tabla ANOVA.

Proposición

Bajo los supuestos del MRLSN, si $\beta_1 = 0$, entonces $F = \dfrac{CME}{CMR} = \dfrac{(n-2)SCE}{SCR} \sim F(1, n-2)$

Teorema. Prueba *F* (ANOVA).

Para el MRLSN $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde $u_i \sim iid\ N(0, \sigma^2)$, si se quiere probar

H_0: $\beta_1 = 0$ vs. H_1: $\begin{cases} \beta_1 > 0 & \text{(una cola)} \\ \beta_1 \neq 0 & \text{(dos cola)} \end{cases}$, entonces $RR_\alpha = \begin{cases} f > f_{1,n-2,\alpha} & \text{(una cola)} \\ f > f_{1,n-2,\frac{\alpha}{2}} & \text{(dos cola)} \end{cases}$,

donde $F = \dfrac{CME}{CMR} = \dfrac{(n-2)SCE}{SCR} \sim F(1, n-2)$.

Ejercicios $\boxed{\text{E22}}$ y $\boxed{\text{E23}}$.

2.8. Predicción en el MRLS.

Si el MRLS estimado $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ es significativo, uno de sus principales usos es el **pronóstico o predicción** de la *media condicional* $E[Y_0 | X_0]$ o del *valor individual* Y_0 para un valor X_0 de la variable explicativa *X*. A estas predicciones también se les suele llamar **interpolación** si X_0 se ubica dentro del conjunto original de datos X_i, o **extrapolación** si X_0 queda fuera de dicho conjunto.

2.8.1. Predicción de la media condicional.

Dado que la variable explicativa *X* toma el valor X_0, el valor de la media poblacional (ubicado sobre la FRP pero no observable) es $E[Y_0 | X_0] = \beta_0 + \beta_1 X_0$ y para estimarla basta sustituir los estimadores MCO de $\hat{\beta}_0$ y $\hat{\beta}_1$.

Teorema. Predicción puntual de la media condicional

Bajo los supuestos del MRLS, el estimador puntual de $E[Y_0 | X_0] = \beta_0 + \beta_1 X_0$ es $\hat{E}[Y_0 | X_0] = \hat{Y}_0^* = \hat{\beta}_0 + \hat{\beta}_1 X_0$, donde:

i) $\hat{Y}_0^*$ es MELI.

ii) $Var\left[\hat{Y}_0^*\right] = \sigma^2 \left[\dfrac{1}{n} + \dfrac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \right]$

Para calcular $S^2_{\hat{Y}_0^*}$, el estimador de $Var\left[\hat{Y}_0^*\right]$, basta sustituir σ^2 por $\hat{\sigma}^2 = \dfrac{1}{n-2} \sum_{i=1}^{n} \hat{u}_i^2$.

Bajo el supuesto de normalidad en los errores es fácil ver que $\hat{Y}_0^*$ sigue una distribución normal, de modo que $\dfrac{\hat{Y}_0^* - (\beta_0 + \beta_1 X_0)}{\sqrt{Var[\hat{Y}_0^*]}} \sim N(0,1)$, y sustituyendo $Var[\hat{Y}_0^*]$ por $S^2_{\hat{Y}_0^*}$ se obtiene el pivote $\dfrac{\hat{Y}_0^* - (\beta_0 + \beta_1 X_0)}{S_{\hat{Y}_0^*}} \sim t(n-2)$.

Teorema. Predicción por intervalo de la media condicional

Bajo los supuestos del MRLSN, el intervalo de confianza simétrico al 100(1 – α)% para $E[Y_0|X_0] = \beta_0 + \beta_1 X_0$ es $\hat{\beta}_0 + \hat{\beta}_1 X_0 \pm t_{n-2,\frac{\alpha}{2}} S_{\hat{Y}_0^*}$, donde $S^2_{\hat{Y}_0^*} = \hat{\sigma}^2 \left[\dfrac{1}{n} + \dfrac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \right]$

Ejercicio $\boxed{\text{E24}}$.

2.8.2. Predicción individual.

Dado que la variable explicativa X toma el valor X_0, su respectivo valor individual Y_0 está dado por $Y_0 = \beta_0 + \beta_1 X_0 + u_0$. Como el término de error u_0 tiene media cero lo más razonable es eliminarlo en la estimación ($\hat{u}_0 = 0$) y simplemente sustituir los estimadores MCO de $\hat{\beta}_0$ y $\hat{\beta}_1$.

Teorema. Predicción puntual de valores individuales

Bajo los supuestos del MRLS, el estimador puntual de $Y_0 = \beta_0 + \beta_1 X_0 + u_0$ es $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0$, donde:

i) $\hat{Y}_0$ es MELI.

ii) $Var[Y_0 - \hat{Y}_0] = \sigma^2 \left[1 + \dfrac{1}{n} + \dfrac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \right]$

A la diferencia $e_0 = Y_0 - \hat{Y}_0$ se le denomina **error de predicción**. Para estimar su varianza, $S^2_{Y_0 - \hat{Y}_0}$, basta sustituir σ^2 por $\hat{\sigma}^2 = \dfrac{1}{n-2} \sum_{i=1}^{n} \hat{u}_i^2$.

Importante: Al pronosticar la media condicional $E[Y_0|X_0] = \beta_0 + \beta_1 X_0$ se está estimando un parámetro y en ese contexto es posible hablar de intervalos de confianza. Al pronosticar un **valor individual** $Y_0 = \beta_0 + \beta_1 X_0 + u_0$ no se está pronosticando un parámetro sino una **variable aleatoria** y por lo tanto no es posible hablar de intervalos de confianza sino de **intervalos de predicción**.

Bajo el supuesto de normalidad en los errores es fácil ver que $e_0 = Y_0 - \hat{Y}_0$ tiene una distribución normal con media cero y varianza $Var[Y_0 - \hat{Y}_0]$, de modo que $\frac{Y_0 - \hat{Y}_0}{\sqrt{Var[Y_0 - \hat{Y}_0]}} \sim N(0,1)$, y sustituyendo $Y_0 - \hat{Y}_0$ por $S^2_{Y_0 - \hat{Y}_0}$ se obtiene la estadística $\frac{Y_0 - \hat{Y}_0}{S_{Y_0 - \hat{Y}_0}} \sim t(n-2)$.

Teorema. Predicción por intervalo de valores individuales

Bajo los supuestos del MRLSN, el intervalo de predicción simétrico al 100(1 – α)% para

$Y_0 = \beta_0 + \beta_1 X_0 + u_0$ es $\hat{\beta}_0 + \hat{\beta}_1 X_0 \pm t_{n-2,\frac{\alpha}{2}} S_{Y_0 - \hat{Y}_0}$, donde $S^2_{Y_0 - \hat{Y}_0} = \hat{\sigma}^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \right]$

Sobre la **predicción por intervalo** hay que notar lo siguiente:

- La longitud del intervalo de predicción del valor individual Y_0 siempre es mayor que la longitud del intervalo de confianza de la media condicional $E[Y_0|X_0] = \beta_0 + \beta_1 X_0$.
- Al lugar geométrico de los extremos de los intervalos de predicción para todos los posibles valores de X_0 se denominan **bandas de predicción**, y su amplitud es más pequeña para valores de X_0 cercanos a $\overline{X}$.

Antes de finalizar con el Análisis de Regresión Simple es importante notar que **la estimación, la bondad del ajuste y la predicción serán válidas únicamente si se cumplen los 8 supuestos del MRLNSN**. Las técnicas para hacer la validación de supuestos y, en su caso, corregirlos se estudiará más adelante.

En la práctica la, estimación, validación de supuestos, bondad de ajuste y predicción del MRSL (y del MRLSN) se realiza mediante paquetes de cómputo (v.gr., Excel, E-Views, R, Stata, Minitab, Statgraphics).

EXCEL. *Análisis de Regresión*. Instalar el complemento “Herramienta para Análisis” e iniciarlo mediante *Datos* > *Análisis de datos* > *Regresión*.

Ejercicios E25 , E26 y E27 .

3. Modelo de Regresión Lineal Múltiple

3.1. Modelo de Regresión Lineal Múltiple (MRLM).

El **Análisis de Regresión Multivariado o Múltiple** estudia la dependencia de una variable de interés (Y) respecto de k variables explicativas ($X_1, X_2, \ldots, X_k$). Su objetivo es estimar o predecir la media poblacional de Y suponiendo conocidos los valores de X_1, $X_2, \ldots, X_k$, es decir, la FRP $E\left[Y \middle| X_1, X_2, \ldots, X_k\right] = g(X_1, X_2, \ldots, X_k)$.

Para la Regresión Múltiple se considerarán arreglos ordenados de datos correspondientes a una muestra de n observaciones, es decir, $(X_{1i}, X_{2i}, \ldots, X_{ki}, Y_i)$, $i = 1, 2, \ldots, n$; donde X_{j1}, $X_{j2}, \ldots, X_{jn}$, son n valores fijos de la j-ésima variable explicativa, $j = 1, 2, \ldots, k$.

En forma similar al MRLS, para el caso multivariado se analizará el caso en que la FRP toma una forma lineal, es decir, $E\left[Y \middle| X_{1i}, X_{2i}, \ldots, X_{ki}\right] = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$.

El **Modelo de Regresión Lineal Múltiple (MRLM)** establece que los valores Y_i pueden ser modelados como la suma de esta media condicional lineal y un término de error u_i de la siguiente manera:

$$\boxed{Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \text{, donde } u_i \sim iid(0, \sigma^2).}$$

Con fines de estimación se asumirán los siguientes supuestos.

Supuestos del Modelo de Regresión Lineal Múltiple

- *i*) El modelo es lineal en parámetros (no necesariamente en variables).
- *ii*) Las variables X_j son no estocásticas e independientes de u_i : $Cov\left[u_i, X_{ji}\right] = 0$.
- *iii*) La media del error es cero: $E\left[u_i\right] = 0$.
- *iv*) La varianza del error es constante: $Var\left[u_i\right] = \sigma^2$ (homocedasticidad).
- *v*) Los errores son independientes y no correlacionados: $Cov\left[u_i, u_j\right] = 0$, $i \neq j$.
- *vi*) El número de observaciones (n) debe ser mayor al número de parámetros a estimar: $n > k + 1$.
- *vii*) Los valores de cada variable X_j no pueden ser todos iguales y no deben existir observaciones atípicas.
- *viii*) No debe haber colinealidad entre las variables X_j.
- *ix*) No debe haber sesgo de especificación.

Los supuestos (i) a (vii) son los mismos que se consideraron en el MRLS.

Al supuesto (viii) se le denomina de **no colinealidad** y significa que ningún regresor puede expresarse como combinación lineal de algunos otros regresores. Si existen constantes $c_1,\ldots, c_k$, tales que $X_j = \sum_{h \neq j} c_h X_h$ para alguna $c_h \neq 0$, entonces se dice que hay problema de **multicolinealidad**.

El **sesgo de especificación** del supuesto (ix) significa que los regresores considerados en el modelo son consistentes con el planteamiento teórico que hay detrás de un modelo. Por ejemplo, en Microeconomía la Teoría del Productor establece que la *función de producción Cobb-Douglas* en su forma econométrica está dada por $P_i = \alpha_0 L_i^{\alpha_1} K_i^{\alpha_2} e^{u_i}$, donde P_i es la producción, L_i el insumo trabajo, K_i el insumo capital y u_i es perturbación estocástica. Su respectivo MRLM es $\ln(P_i) = \beta_0 + \beta_1 \ln(L_i) + \beta_2 \ln(K_i) + u_i$, donde $\beta_0 = \ln(\alpha_0)$, $\beta_1 = \alpha_1$ y $\beta_2 = \alpha_2$. Si se estima el modelo $\ln(P_i) = \beta_0 + \beta_1 \ln(L_i) + u_i$ entonces se estaría incurriendo en un sesgo de especificación pues al omitir el regresor $\ln(K_i)$ no se estaría considerado el "modelo verdadero".

Notación: Para el análisis del MRLM es útil considerar datos, parámetros y errores en forma de vectores (letras minúsculas negritas) y matrices (letras mayúsculas negritas).

De este modo las n observaciones de la variable de interés, los $k + 1$ parámetros y los n términos de error se pueden representar mediante los **vectores**:

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \qquad \text{y} \qquad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.$$

Por su parte, los valores de las n variables explicativas se pueden acomodar en una **matriz** de dimensión $n \times (k + 1)$, con un renglón por cada observación, una columna por cada variable y una columna auxiliar de unos, de la siguiente manera:

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix}.$$

A la matriz **X** se le conoce como **matriz de diseño**, y permite expresar simultáneamente las n observaciones del MRLM de mantera compacta: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$. Note que la FRP es simplemente $E[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$. Agregando los supuestos se puede definir formalmente el MRLM de la siguiente manera.

Def. Modelo de Regresión Lineal Múltiple (MRLM).

Si (X_{1i}, X_{2i},…, X_{ki}, Y_i), i = 1, 2,…, n, j = 1, 2,…, k, son n observaciones de la variable dependiente Y acomodadas en el vector **y**, y n observaciones de las variables explicativas X_1, X_2,…, X_k ($n > k + 1$ y no todas las X_{ji}'s toman el mismo valor) acomodadas en la matriz de diseño **X**, cuya FRP es $E[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$, entonces el Modelo de Regresión Lineal Múltiple supone que **y** puede ser modelado como la suma de esta media condicional y un vector de errores **u** de la siguiente manera:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

donde **X** es no estocástica, $\mathbf{X}'\mathbf{X}$ es no singular y $\mathbf{u} \sim (\mathbf{0}, \sigma^2\mathbf{I})$.

Note que $\mathbf{0}$ es un *vector de n ceros* (o *vector nulo*), $\mathbf{I}$ es una *matriz identidad* de dimensión $n \times n$ y σ^2 una constante positiva (o un escalar positivo). En esta definición se incluyen todos los supuestos establecidos previamente, excepto el (ix) de no sesgo de especificación. Si $\mathbf{X}'\mathbf{X}$ es *no singular* significa: (i) que sus columnas son linealmente independientes, lo que garantiza la no colinealidad; y (ii) que su inversa $(\mathbf{X}'\mathbf{X})^{-1}$ existe.

Def. Esperanza y Varianza de Vectores Aleatorios

Si **w** es el vector aleatorio que contiene a las n variables aleatorias W_1, W_2,…, W_n:

i) $$E[\mathbf{w}] = \begin{bmatrix} E[W_1] \\ E[W_2] \\ \vdots \\ E[W_n] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \boldsymbol{\mu}$$

ii) $$\begin{aligned} Var[\mathbf{w}] &= E\left[(\mathbf{w}-\boldsymbol{\mu})(\mathbf{w}-\boldsymbol{\mu})'\right] \\ &= \begin{bmatrix} E\left[(W_1-\mu_1)^2\right] & E[(W_1-\mu_1)(W_2-\mu_2)] & \cdots & E[(W_1-\mu_1)(W_n-\mu_n)] \\ E[(W_2-\mu_2)(W_1-\mu_1)] & E\left[(W_2-\mu_2)^2\right] & \cdots & E[(W_2-\mu_2)(W_n-\mu_n)] \\ \vdots & \vdots & & \vdots \\ E[(W_n-\mu_n)(W_1-\mu_1)] & E[(W_n-\mu_n)(W_2-\mu_2)] & \cdots & E\left[(W_n-\mu_n)^2\right] \end{bmatrix} \\ &= \begin{bmatrix} Var[W_1] & Cov[W_1,W_2] & \cdots & Cov[W_1,W_n] \\ Cov[W_1,W_2] & Var[W_2] & \cdots & Cov[W_2,W_n] \\ \vdots & \vdots & & \vdots \\ Cov[W_1,W_n] & Cov[W_n,W_2] & \cdots & Var[W_n] \end{bmatrix} = \boldsymbol{\Sigma}_{\mathbf{w}} \end{aligned}$$

Note que $Var[\mathbf{w}] = \boldsymbol{\Sigma}_{\mathbf{w}}$ es la **matriz de varianzas y covarianzas** y por propiedades de la covarianza es una *matriz simétrica*, es decir, $\boldsymbol{\Sigma}'_{\mathbf{w}} = \boldsymbol{\Sigma}_{\mathbf{w}}$.

La expresión $\mathbf{u} \sim \left(\mathbf{0}, \sigma^2\mathbf{I}\right)$ es una manera compacta de establecer los supuestos (iii), (iv) y (v) acerca del término de error pues $E[\mathbf{u}] = \mathbf{0} \Leftrightarrow E[u_i] = 0$ para todo i = 1, 2,…, n; y $Var[\mathbf{u}] = \sigma^2\mathbf{I} \Leftrightarrow Var[u_i] = \sigma^2$ para todo i = 1, 2,.., n, y $Cov[u_i, u_j] = 0$ para todo $i \neq j$.

Antes de estimar el MRLM es necesario recordar algunos conceptos de Álgebra Lineal útiles en el Análisis de Regresión.

3.2. Álgebra matricial del MRLM.

3.2.1. Operaciones con Matrices.

Def. Matriz Transpuesta. Si **A** es matriz de dimensión $m \times n$, con elementos a_{ij}, entonces $\mathbf{A}'$ es la matriz transpuesta de **A** y es una matriz de dimensión $n \times m$ con elementos a_{ji}.

EXCEL. *Matriz transpuesta*. La función matricial =TRANSPONER(rango), permite calcular la transpuesta de una matriz (o vector) ubicada en un “rango”. Por ser función matricial primero hay que seleccionar las celdas en donde quedará la correspondiente matriz (o vector) transpuesta y después de escribir la función presionar simultáneamente Ctrl + Shift + Enter.

Def. Producto Matricial. Si **A** es matriz de dimensión $m \times n$ con elementos a_{ij}, y **B** es matriz de dimensión $n \times p$ con elementos b_{ij}, entonces **C** = **AB** es una matriz de dimensión $m \times p$ con elementos $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$. Este producto no es conmutativo.

EXCEL. *Producto Matricial*. La función matricial =MMULT(rango1,rango2), permite calcular el producto matricial de las matrices ubicadas en el “rango1” y el “rango2”. Por ser función matricial primero hay que seleccionar las celdas en donde quedará la correspondiente matriz y después de escribir la función presionar simultáneamente Ctrl + Shift + Enter. **Nota**: aunque no haya coincidencia entre columnas del “rango1” y los renglones del “rango2”, esta función generará un resultado que no corresponde al producto deseado.

Def. Producto Interior y Producto Exterior. El *producto interior* (también llamado *producto punto* o *producto escalar*) de los vectores **a** y **c** de dimensión n se define como $\mathbf{a}'\mathbf{c} = a_1c_1 + a_2c_2 + \cdots + a_nc_n = \sum_{i=1}^{n} a_i c_i$, es decir, produce un escalar. En cambio, el *producto exterior* $\mathbf{ac}'$ produce una matriz.

Def. Determinante de una Matriz. El determinante de la matriz cuadrada **A**, denotado por |**A**| o det(**A**), es un escalar que se puede calcular de manera recursiva por la *expansión de Laplace* o *expansión de cofactores*. En particular para n =2:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

Y para n = 3, considerando el primer renglón de la *matriz de signos* $\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}$

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \end{aligned}$$

EXCEL. *Determinantes*. La función matricial =MDETERM(rango), permite calcular el determinante de una matriz cuadrada ubicada en un "rango". Por ser función matricial primero hay que seleccionar la celda en donde quedará el escalar correspondiente y después de escribir la función presionar simultáneamente Ctrl + Shift + Enter.

Def. Matriz Inversa. Si **A** es matriz cuadrada de dimensión $n \times n$ se dice que $\mathbf{A}^{-1}$ es su matriz inversa si $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$, donde **I** es la *matriz identidad* de dimensión $n \times n$.

Existen distintos métodos para calcular la inversa de una matriz:

- *Eliminación Gaussiana*: Parte de la matriz aumentada [**A** | **I**] y consiste en hacer operaciones elementales por renglón hasta obtener [**I** | $\mathbf{A}^{-1}$].
- *Solución Analítica*: $\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\mathrm{Adj}(\mathbf{A})$, donde $\mathrm{Adj}(\mathbf{A}) = \mathbf{C}'$ es la *matriz adjunta* de **A** y es la transpuesta de la *matriz de cofactores* **C**. Los elementos de la matriz **C** son de la forma $C_{ij} = (-1)^{i+j} M_{ij}$, donde M_{ij} es el *menor* de la matriz **A** y es el determinante de la submatriz que se genera al eliminar el renglón i y la columna j de la matriz **A**.

En particular para n =2:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

EXCEL. *Matriz Inversa*. La función matricial =MINVERSA(rango), permite calcular la inversa de una matriz cuadrada ubicada en un "rango". Por ser función matricial primero hay que seleccionar el rango donde quedará la matriz inversa y después de escribir la función presionar simultáneamente Ctrl + Shift + Enter.

Def. Traza. Si **A** es matriz cuadrada de dimensión $n \times n$ se dice su *traza* es la suma de los elementos de su diagonal principal, es decir, $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$.

Propiedades de las Operaciones con Matrices

Si **a** y **b** son vectores de dimensión n, y **A**, **B** y **C** son matrices de dimensión $n \times n$:

i)	$\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a}$	*viii*)	$\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}$
ii)	$\mathbf{a}'\mathbf{a} = \sum_{i=1}^{n} a_i^2 \geq 0$	*ix*)	$\lvert\mathbf{AB}\rvert = \lvert\mathbf{A}\rvert\lvert\mathbf{B}\rvert$
iii)	$\mathbf{ab}' = (\mathbf{ba}')'$	*x*)	$\lvert\mathbf{A}^{-1}\rvert = \frac{1}{\lvert\mathbf{A}\rvert}$
iv)	$(\mathbf{A}')' = \mathbf{A}$	*xi*)	$\lvert c\mathbf{A}\rvert = c^n \lvert\mathbf{A}\rvert$
v)	$(\mathbf{A}+\mathbf{B})' = \mathbf{A}' + \mathbf{B}'$	*xii*)	$\mathrm{tr}(c\mathbf{A}) = c\,\mathrm{tr}(\mathbf{A})$
vi)	$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$	*xiii*)	$\mathrm{tr}(\mathbf{A}+\mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$
vii)	$\left(\mathbf{A}^{-1}\right)' = (\mathbf{A}')^{-1}$	*xiv*)	$\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{CBA})$ y cualquier otra permutación posible.

3.2.2. Otros conceptos y propiedades relevantes del álgebra matricial.

Def. Matriz Simétrica. Se dice que la matriz cuadrada **A** es *simétrica* si $\mathbf{A}' = \mathbf{A}$, es decir, $a_{ij} = a_{ji}$.

Def. Matriz Idempotente. Se dice que la matriz cuadrada A es *idempotente* si se cumple que $\mathbf{A} = \mathbf{A}^2 = \mathbf{A}^3 = \cdots$, donde $\mathbf{A}^2 = \mathbf{AA}$, $\mathbf{A}^3 = \mathbf{AAA}$, etc.

Def. Vectores Linealmente Independientes. Se dice que los vectores $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ son *linealmente independientes* si ninguno de ellos puede ser escrito como combinación lineal de los otros, es decir, si la única solución de $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_k\mathbf{a}_k = \mathbf{0}$ es $c_1 = c_2 = \cdots = c_k = 0$.

Def. Rango. El *rango* de la matriz $\mathbf{A}$, denotado por Rango($\mathbf{A}$), es el número de columnas o renglones linealmente independientes. Si $\mathbf{A}$ es matriz cuadrada de dimensión $n \times n$ y su rango es n, entonces se dice que es *matriz de rango completo*.

Def. Forma Cuadrática. Se dice que el escalar $q = \mathbf{c}'\mathbf{A}\mathbf{c}$ es una *forma cuadrática* si $\mathbf{A}$ es una matriz simétrica y $\mathbf{c}$ es un vector no nulo.

Def. Matriz Positiva Definida. Se dice que la matriz cuadrada $\mathbf{A}$ de dimensión $n \times n$ es *positiva definida* si $\mathbf{c}'\mathbf{A}\mathbf{c} > 0$ para todo vector no nulo $\mathbf{c}$ de dimensión n. Si $\mathbf{c}'\mathbf{A}\mathbf{c} \geq 0$ se dice que $\mathbf{A}$ es matriz es *positiva semidefinida*.

Def. Matriz Negativa Definida. Se dice que la matriz cuadrada $\mathbf{A}$ de dimensión $n \times n$ es *negativa definida* si $\mathbf{c}'\mathbf{A}\mathbf{c} < 0$ para todo vector no nulo $\mathbf{c}$ de dimensión n. Si $\mathbf{c}'\mathbf{A}\mathbf{c} \leq 0$ se dice que $\mathbf{A}$ es matriz es *negativa semidefinida*.

Def. Matriz Diagonal. Se dice que la matriz cuadrada $\mathbf{D}$ es *matriz diagonal* si $d_{ij} = 0$ para toda $i \neq j$, es decir, si los elementos fuera de la diagonal principal son ceros.

Def. Matriz Ortogonal. Se dice que $\mathbf{Q}$ es matriz ortogonal si su inversa es simplemente su transpuesta, es decir, $\mathbf{Q}^{-1} = \mathbf{Q}'$.

Propiedades Matriciales

Si $\mathbf{A}$ es matriz de dimensión $n \times n$, $\mathbf{B}$ es matriz de dimensión $n \times m$, $\mathbf{c}$ es un vector de dimensión n, y $\mathbf{D}$ es matriz diagonal de dimensión $n \times n$, entonces:

i) Si $\mathbf{A}$ es de rango completo entonces $|\mathbf{A}| \neq 0$

ii) Si $\mathbf{A}$ es de rango completo entonces $\mathbf{A}^{-1}$ siempre existe.

iii) $\mathbf{A}'\mathbf{A}$ siempre es matriz simétrica.

iv) Si $\mathbf{A}$ es de rango completo $\mathbf{A}'\mathbf{A}$ siempre es positiva definida.

v) Si $\mathbf{A}$ es idempotente entonces Rango($\mathbf{A}$) = tr($\mathbf{A}$).

vi) Si $\mathbf{A}$ es positiva definida y Rango($\mathbf{B}$) = m, entonces $\mathbf{B}'\mathbf{A}\mathbf{B}$ es positiva definida.

vii) Si $\mathbf{A}$ es simétrica, entonces $(\mathbf{c}'\mathbf{A}\mathbf{c})' = \mathbf{c}'\mathbf{A}\mathbf{c}$.

viii) $|\mathbf{D}| = \prod_{i=1}^{n} d_{ii}$

3.2.3. Diferenciación Matricial.

Def. Derivada Matricial

Si **c** es vector de dimensión n y $z = g(\mathbf{c})$ donde g: $\mathbf{R}^n \to \mathbf{R}$, entonces:

$$\frac{\partial z}{\partial \mathbf{c}} = \frac{\partial}{\partial \mathbf{c}} g(\mathbf{c}) = \begin{bmatrix} \frac{\partial}{\partial c_1} g(\mathbf{c}) \\ \frac{\partial}{\partial c_2} g(\mathbf{c}) \\ \vdots \\ \frac{\partial}{\partial c_n} g(\mathbf{c}) \end{bmatrix} \quad \text{y} \quad \frac{\partial^2 z}{\partial \mathbf{c} \partial \mathbf{c}'} = \frac{\partial^2}{\partial \mathbf{c} \partial \mathbf{c}'} g(\mathbf{c}) = \begin{bmatrix} \frac{\partial^2}{\partial c_1^2} g(\mathbf{c}) & \frac{\partial^2}{\partial c_1 \partial c_2} g(\mathbf{c}) & \cdots & \frac{\partial^2}{\partial c_1 \partial c_n} g(\mathbf{c}) \\ \frac{\partial^2}{\partial c_2 \partial c_1} g(\mathbf{c}) & \frac{\partial^2}{\partial c_2^2} g(\mathbf{c}) & \cdots & \frac{\partial^2}{\partial c_2 \partial c_n} g(\mathbf{c}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2}{\partial c_n \partial c_1} g(\mathbf{c}) & \frac{\partial^2}{\partial c_n \partial c_2} g(\mathbf{c}) & \cdots & \frac{\partial^2}{\partial c_n^2} g(\mathbf{c}) \end{bmatrix}$$

Proposición

Si **a** y **c** son vectores de dimensión n, **0** es el vector nulo de dimensión n, **1** es el vector de unos de dimensión n, **A** es matriz simétrica de dimensión $n \times n$, g y h son funciones (g, h: $\mathbf{R}^n \to \mathbf{R}$) y k es un escalar, entonces:

i) $\frac{\partial}{\partial \mathbf{c}}(\mathbf{a}'\mathbf{a}) = \mathbf{0}$

ii) $\frac{\partial}{\partial \mathbf{c}}(\mathbf{c}'\mathbf{1}) = \mathbf{1}$

iii) $\frac{\partial}{\partial \mathbf{c}}(kg(\mathbf{c})) = k\frac{\partial}{\partial \mathbf{c}} g(\mathbf{c})$

iv) $\frac{\partial}{\partial \mathbf{c}}(g(\mathbf{c}) \pm h(\mathbf{c})) = \frac{\partial}{\partial \mathbf{c}} g(\mathbf{c}) \pm \frac{\partial}{\partial \mathbf{c}} h(\mathbf{c})$

v) $\frac{\partial}{\partial \mathbf{c}}(\mathbf{a}'\mathbf{c}) = \frac{\partial}{\partial \mathbf{c}}(\mathbf{c}'\mathbf{a}) = \mathbf{a}$

vi) $\frac{\partial}{\partial \mathbf{c}}(\mathbf{c}'\mathbf{A}\mathbf{c}) = 2\mathbf{A}\mathbf{c}$

vii) $\frac{\partial^2}{\partial \mathbf{c} \partial \mathbf{c}'}(\mathbf{a}'\mathbf{c}) = \mathbf{O}_{n \times n}$ y $\frac{\partial^2}{\partial \mathbf{c} \partial \mathbf{c}'}(\mathbf{c}'\mathbf{A}\mathbf{c}) = 2\mathbf{A}$

Demostración: (i) a (iv) son triviales. (v) Recuerde que $\mathbf{a}'\mathbf{c} = \mathbf{c}'\mathbf{a} = \sum_{i=1}^{n} a_i c_i$ de donde se observa que $\frac{\partial}{\partial c_i}(\mathbf{a}'\mathbf{c}) = \frac{\partial}{\partial c_i}(\mathbf{c}'\mathbf{a}) = a_i$ para $i = 1, 2, \ldots, n$, de modo que el vector con todos estos elementos es simplemente el vector **a**, es decir, $\frac{\partial}{\partial \mathbf{c}}(\mathbf{a}'\mathbf{c}) = \frac{\partial}{\partial \mathbf{c}}(\mathbf{c}'\mathbf{a}) = \mathbf{a}$. (vi) Se considerará el caso en que $n = 3$:

$$\begin{aligned}
\mathbf{c}'\mathbf{A}\mathbf{c} &= \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} c_1 a_{11} + c_2 a_{12} + c_3 a_{13} \\ c_1 a_{12} + c_2 a_{22} + c_3 a_{23} \\ c_1 a_{13} + c_2 a_{23} + c_3 a_{33} \end{bmatrix} \\
&= c_1(c_1 a_{11} + c_2 a_{12} + c_3 a_{13}) + c_2(c_1 a_{12} + c_2 a_{22} + c_3 a_{23}) + c_3(c_1 a_{13} + c_2 a_{23} + c_3 a_{33}) \\
&= c_1^2 a_{11} + c_2^2 a_{22} + c_3^2 a_{33} + 2c_1 c_2 a_{12} + 2c_1 c_3 a_{13} + 2c_2 c_3 a_{23}
\end{aligned}$$

Entonces $\frac{\partial}{\partial c_1}(\mathbf{c}'\mathbf{A}\mathbf{c}) = 2c_1 a_{11} + 2c_2 a_{12} + 2c_3 a_{13}$, $\frac{\partial}{\partial c_2}(\mathbf{c}'\mathbf{A}\mathbf{c}) = 2c_1 a_{12} + 2c_2 a_{22} + 2c_3 a_{23}$ y $\frac{\partial}{\partial c_3}(\mathbf{c}'\mathbf{A}\mathbf{c}) = 2c_1 a_{13} + 2c_2 a_{23} + 2c_3 a_{33}$, y factorizando el 2 de cada componente se obtienen los renglones de la matriz $\mathbf{Ac}$, por lo tanto, $\frac{\partial}{\partial \mathbf{c}}(\mathbf{c}'\mathbf{A}\mathbf{c}) = 2\mathbf{Ac}$.

Ejercicio $\boxed{\text{E28}}$.

3.3. Estimación por MCO del MRLM.

En forma análoga al caso simple, en el MRLM se busca estimar la FRP $E[Y|X_1, X_{2i}, \ldots, X_{ki}] = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$ mediante la Función de Regresión Muestral (FRM) $\hat{E}[Y|X_{1i}, X_{2i}, \ldots, X_{ki}] = \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}$, $i = 1, 2, \ldots, n$, donde $\hat{\beta}_0$, $\hat{\beta}_1, \ldots, \hat{\beta}_k$ son los estimadores de los parámetros $\beta_0, \beta_1, \ldots, \beta_k$.

Si $\hat{\mathbf{y}}$ denota el vector de dimensión n con los valores ajustados $\hat{Y}_1$, $\hat{Y}_2, \ldots, \hat{Y}_n$, y $\hat{\boldsymbol{\beta}}$ denota el vector de dimensión $k + 1$ con los estimadores $\hat{\beta}_0$, $\hat{\beta}_1, \ldots, \hat{\beta}_k$, entonces la FRM se puede escribir como $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, y es el estimador de la FRP $E[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$.

Además, si $\hat{\mathbf{u}}$ denota el vector de dimensión n con los residuos $\hat{u}_1$, $\hat{u}_2, \ldots, \hat{u}_n$, entonces el vector de observaciones $\mathbf{y}$ se puede expresar como suma de la FRM $\hat{\mathbf{y}} = \hat{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\hat{\boldsymbol{\beta}}$ y el vector de residuos $\hat{\mathbf{u}}$ como se muestra a continuación:

$$\boxed{\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{u}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}}$$

Es fácil ver que el vector de residuos se puede expresar de la siguiente manera:

$$\boxed{\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}}$$

Al aplicar el Método de Mínimos Cuadrados Ordinarios, la Suma de Cuadrados de Residuos se puede expresar como $SCR = \sum_{i=1}^{n} \hat{u}_i^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}} = \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = g\left(\hat{\boldsymbol{\beta}}\right)$, donde g: $\mathbf{R}^n \to \mathbf{R}$ es la función a minimizar.

Teorema. Estimadores MCO del MRLM

Para el MRLM $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde $\mathbf{u} \sim \left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$ y $\mathbf{X}'\mathbf{X}$ es no singular:

i) $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$ es el estimador MCO de $\boldsymbol{\beta}$.

ii) $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$ es estimador insesgado de σ^2.

El vector de estimadores MCO $\hat{\boldsymbol{\beta}}$ surge directamente en el proceso de minimización de la suma de cuadrados de residuos del MRLM. El estimador $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$ se elige así por ser estimador insesgado.

En forma análoga al caso simple, al plantear las condiciones de primer orden en la aplicación del Método MCO surgen las llamadas *ecuaciones normales*.

Corolario. Ecuaciones Normales del MRLM

$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$

Ejercicios $\boxed{\text{E29}}$ y $\boxed{\text{E30}}$.

A partir de los estimadores MCO es posible definir el **MRLM ajustado**, que es simplemente la relación determinista $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k$. A partir de los valores reales $\mathbf{y}$ y de los valores ajustados $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ es fácil calcular los residuos $\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}$, que satisfacen las siguientes propiedades:

Propiedades de los residuos.

Si el MRLM $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ se estima por MCO y $\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$:

i) Los residuos y los regresores no están correlacionados: $\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}$

ii) La suma de los residuos es cero: $\sum_{i=1}^{n} \hat{u}_i = 0$.

iii) Los residuos y los valores ajustados no están correlacionados: $\hat{\mathbf{y}}'\hat{\mathbf{u}} = 0$.

Ejercicios $\boxed{\text{E31}}$ y $\boxed{\text{E32}}$.

3.4. Propiedades estadísticas de los estimadores MCO.

3.4.1. Esperanza y Varianza de los estimadores MCO.

Teorema

Si $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$ y $\hat{\sigma}^2 = \dfrac{\mathbf{u}'\mathbf{u}}{n-(k+1)}$ son los estimadores MCO del MRLN $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde $\mathbf{u} \sim (\mathbf{0}, \sigma^2\mathbf{I})$ y $\mathbf{X}'\mathbf{X}$ es no singular, entonces:

i) $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$ es estimador insesgado de $\boldsymbol{\beta}$.

ii) $Var[\hat{\boldsymbol{\beta}}] = \boldsymbol{\Sigma}_{\hat{\beta}} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$ es la matriz de varianzas-covarianzas de $\hat{\boldsymbol{\beta}}$.

iii) $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$ es estimador insesgado de σ^2.

Para estimar la matriz de varianzas-covarianzas de $\hat{\boldsymbol{\beta}}$ basta remplazar a σ^2 por $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$, de modo que $\hat{\boldsymbol{\Sigma}}_{\hat{\beta}} = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$.

Al estimar el MRLM por MCO es posible escribir el vector de residuos como sigue:

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{X}\left[(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})\right] = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{M}\mathbf{y}$$

A la matriz $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ se le conoce como **matriz gorro** (*hat matrix*), es de dimensión $n \times n$, y geométricamente es la proyección ortogonal sobre el espacio de columnas de la matriz de diseño **X**; de hecho, tr(**H**) = Rango(**X**). La matriz $\mathbf{M} = \mathbf{I} - \mathbf{H}$ tiene las siguientes propiedades: (i) es simétrica e idempotente; (ii) $\mathbf{M}\mathbf{X} = \mathbf{O}_{n\times(k+1)}$; y (iii) $\mathbf{M}\hat{\mathbf{u}} = \hat{\mathbf{u}}$ (ver Tarea 2).

La matriz **M** es útil para demostrar que $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$ es insesgado y para calcular la matriz de varianzas-covarianzas del vector de residuos. Bajo los supuestos del MRLM se puede demostrar que $Var[\hat{\mathbf{u}}] = \sigma^2\mathbf{M}$.

3.4.2. Teorema de Gauss-Markov.

Teorema de Gauss-Markov

Bajo los supuestos del MRLM, el estimador MCO $\hat{\boldsymbol{\beta}}$ es MELI.

Ejercicios E33 y E34 .

3.5. R^2 y R^2 ajustada.

En el contexto del MRLM la **bondad de ajuste** se enfoca en determinar qué tan bien se ajusta el modelo lineal en parámetros al conjunto de datos observados. En forma completamente análoga al MRLS, en el MRLM se considerará el *coeficiente de determinación* (R^2), pero adicionalmente se definirá el *coeficiente de determinación ajustado* ($\overline{R}^2$) que permite comparar modelos con distinto número de regresores.

Proposición. Descomposición de la SCT

Si el MRLM $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ se estima por MCO, entonces $SCT = SCE + SCR$, donde $SCT = \mathbf{y}'\mathbf{y} - n\overline{Y}^2$, $SCE = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - n\overline{Y}^2$ y $SCR = \hat{\mathbf{u}}'\hat{\mathbf{u}} = \sum_{i=1}^{n} \hat{u}_i^2$.

Demostración: Al estimar por MCO, los valores observados se pueden expresar como los valores ajustados más los residuos, es decir, $\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{u}}$. Entonces la suma de cuadrados de las Y_i's es $\mathbf{y}'\mathbf{y} = (\hat{\mathbf{y}} + \hat{\mathbf{u}})'(\hat{\mathbf{y}} + \hat{\mathbf{u}}) = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\mathbf{y}}'\hat{\mathbf{u}} + \hat{\mathbf{u}}'\hat{\mathbf{y}} + \hat{\mathbf{u}}'\hat{\mathbf{u}} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\mathbf{u}}'\hat{\mathbf{u}}$ pues los residuos y los valores ajustados no están correlacionados ($\hat{\mathbf{y}}'\hat{\mathbf{u}} = \hat{\mathbf{u}}'\hat{\mathbf{y}} = 0$). Se sabe que la **suma de cuadrados totales** es $SCT = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2 = \mathbf{y}'\mathbf{y} - n\overline{Y}^2$ entonces $\mathbf{y}'\mathbf{y} - n\overline{Y}^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\overline{Y}^2 + \hat{\mathbf{u}}'\hat{\mathbf{u}}$. Finalmente, sustituyendo $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ y $\hat{\mathbf{y}}' = (\mathbf{X}\hat{\boldsymbol{\beta}})' = \hat{\boldsymbol{\beta}}'\mathbf{X}'$:

$$\underbrace{\mathbf{y}'\mathbf{y} - n\overline{Y}^2}_{SCT} = \underbrace{\left(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - n\overline{Y}^2\right)}_{SCE} + \underbrace{\hat{\mathbf{u}}'\hat{\mathbf{u}}}_{SCR}$$

De esta última expresión se sabe que $1 = \frac{SCE}{SCT} + \frac{SCR}{SCT}$, es decir, el 100% de la variación cuadrática total se puede descomponer como el porcentaje explicado por la regresión $\frac{SCE}{SCT}$ más el porcentaje explicado por los residuos $\frac{SCR}{SCT}$. De este modo la definición del coeficiente de determinación del MRLM es la misma que en el MRLS.

Def. Coeficiente de Determinación (*R*²)

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\mathbf{y}'\mathbf{y} - n\overline{Y}^2}$$

Es decir, el **coeficiente de determinación** mide la proporción o porcentaje de la variación cuadrática total que es explicada por la regresión ($0 \le R^2 \le 1$). Si $R^2 \to 1$ se tiene un buen ajuste (residuos pequeños), en cambio, si $R^2 \to 0$ se tiene un mal ajuste (residuos grandes).

A diferencia del MRLS, en el MRLM no se puede establecer que $R^2 = (r_{XY})^2$ pues no hay un solo regresor X, sino un conjunto de k regresores $X_1, X_2, \ldots, X_k$.

Desafortunadamente en el contexto multivariado R^2 crece conforme aumenta el número de regresores, de manera que para evitar esa distorsión y poder comparar una medida de bondad de ajuste de modelos con distinto número de regresores se define al *coeficiente de determinación ajustado*.

Def. Coeficiente de Determinación Ajustado ($\overline{R}^2$)

$$\overline{R}^2 = 1 - \frac{\dfrac{SCR}{n-(k+1)}}{\dfrac{SCT}{n-1}} = 1 - \frac{(n-1)SCR}{[n-(k+1)]SCT}$$

Proposición

$$\overline{R}^2 = \frac{n-1}{n-(k+1)}R^2 - \frac{k}{n-(k+1)}$$

Demostración: $\overline{R}^2 = 1 - \dfrac{(n-1)SCR}{[n-(k+1)]SCT} = 1 - \dfrac{n-1}{n-(k+1)}\left(1 - \dfrac{SCE}{SCT}\right)$, donde $\dfrac{SCE}{SCT} = R^2$, entonces

$$\overline{R}^2 = 1 - \frac{n-1}{n-(k+1)}\left(1-R^2\right) = \frac{n-1}{n-(k+1)}R^2 + \frac{n-k-1-n+1}{n-(k+1)} = \frac{n-1}{n-(k+1)}R^2 - \frac{k}{n-(k+1)}.$$

Importante: La comparación de R^2 o $\overline{R}^2$ de distintos modelos es válida únicamente si: (i) la variable dependiente es la misma; y (ii) incluyen mismo número de observaciones.

3.6. Análisis de varianza.

En el MRLM también se acostumbra presentar la descomposición $SCT = SCE + SCR$ mediante la Tabla ANOVA como se muestra en la figura 3.1.

Figura 3.1

Fuente de variación	**Suma de cuadrados**	**g.l.**	**Suma de cuadrados medios**
Regresión (explicada)	$SCE = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\overline{Y}^2 = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - n\overline{Y}^2$	k	$CME = \dfrac{SCE}{k} = \dfrac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - n\overline{Y}^2}{k}$
Residuos	$SCR = (\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}}) = \hat{\mathbf{u}}'\hat{\mathbf{u}} = \sum_{i=1}^{n}\hat{u}_i^2$	$n-(k+1)$	$CMR = \dfrac{SCR}{n-(k+1)} = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)} = \hat{\sigma}^2$
Total	$SCT = \mathbf{y}'\mathbf{y} - n\overline{Y}^2$	$n-1$	$CMT = \dfrac{SCT}{n-1} = \dfrac{\mathbf{y}'\mathbf{y} - n\overline{Y}^2}{n-1} = S_Y^2$

A cada suma de cuadrados corresponden ciertos **grados de libertad**, que son el número de observaciones independientes requeridas para calcular dichas estadísticas. *SCT* tiene $n-1$ grados de libertad pues pierde un grado de libertad en el cálculo de $\overline{Y}$, *SCE* tiene k grado de libertad (correspondientes a $\hat{\beta}_1$, $\hat{\beta}_2,\ldots,$ $\hat{\beta}_k$ asociados a las k variables explicativas $X_1, X_2,\ldots, X_k$), y *SCR* tiene $n-(k+1)$ grados de libertad pues pierde $k+1$ grados de libertad en el cálculo de $\hat{\beta}_0$, $\hat{\beta}_1,\ldots,$ $\hat{\beta}_k$. La **suma de cuadrados medios** es simplemente la suma de cuadrados totales entre sus respectivos grados de libertad.

Ejercicio $\boxed{\text{E35}}$.

3.7. Estimadores de Máxima Verosimilitud (EMV).

Para estimar el MRLM $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u}$ bajo el Método de Máxima Verosimilitud es necesario suponer una distribución de probabilidad para el vector de errores $\mathbf{u}$. Para cumplir con el resto de los supuestos del MRLM la distribución multivariada de $\mathbf{u}$ debe tener vector de medias cero, $E[\mathbf{u}]=\mathbf{0}$, y matriz de varianzas-covarianzas de la forma $Var[\mathbf{u}]=\sigma^2\mathbf{I}$, es decir, con varianzas constantes σ^2 y no correlacionados. Por practicidad y conveniencia, para fines de inferencia estadística se considerará a la **Distribución Normal Multivariada**, es decir, $\mathbf{u}\sim N_n\left(\mathbf{0},\sigma^2\mathbf{I}\right)$.

Def. Modelo de Regresión Lineal Múltiple Normal (MRLMN).

Si $(X_{1i}, X_{2i},\ldots, X_{ki}, Y_i)$, $i = 1, 2,\ldots, n$, $j = 1, 2,\ldots, k$, son n observaciones de la variable dependiente Y acomodadas en el vector $\mathbf{y}$, y n observaciones de las variables explicativas $X_1, X_2,\ldots, X_k$ ($n > k + 1$ y no todas las X_{ji}'s toman el mismo valor) acomodadas en la matriz de diseño $\mathbf{X}$, cuya FRP es $E[\mathbf{y}|\mathbf{X}]=\mathbf{X}\boldsymbol{\beta}$, entonces el Modelo de Regresión Lineal Múltiple Normal supone que y puede ser modelado como la suma de esta media condicional y un vector de n errores $\mathbf{u}$ de la siguiente manera:

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u},$$

donde $\mathbf{X}$ es no estocástica, $\mathbf{X}'\mathbf{X}$ es no singular y $\mathbf{u}\sim N_n\left(\mathbf{0},\sigma^2\mathbf{I}\right)$.

Def. Distribución Normal Multivariada

Si $\mathbf{w}$ es el vector aleatorio que contiene a las n variables aleatorias $W_1, W_2,\ldots, W_n$, se dice que $\mathbf{w}$ tiene una Distribución Normal Multivariada de dimensión n con vector de medias $\boldsymbol{\mu}$ y matriz de varianzas $\boldsymbol{\Sigma}$, denotado por $\mathbf{w}\sim N_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$, si su función de densidad de probabilidad conjunta está dada por:

$$f_{\mathbf{w}}(\mathbf{w})=\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\right\},\ \mathbf{w}\in\mathbf{R}^n$$

En particular, para $\mathbf{u} \sim N_n\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$ se tiene que $\boldsymbol{\mu} = \mathbf{0}$ y $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$. Como $\sigma^2\mathbf{I}$ es matriz diagonal, propiedades de los determinantes y de la inversa $|\boldsymbol{\Sigma}| = \left|\sigma^2\mathbf{I}\right| = \left(\sigma^2\right)^n|\mathbf{I}| = \sigma^{2n}$ y $\boldsymbol{\Sigma}^{-1} = \left(\sigma^2\mathbf{I}\right)^{-1} = \frac{1}{\sigma^2}\mathbf{I}$, por lo tanto, $f_{\mathbf{u}}(\mathbf{u}) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}}\exp\left\{-\frac{\mathbf{u}'\mathbf{u}}{2\sigma^2}\right\}$, $\mathbf{u} \in \mathbf{R}^n$.

Aunque en general no correlación no implica independencia, es importante recordar que en el caso particular de la Distribución Normal Multivariada la no correlación sí implica independencia.

Suponer $u_i \sim iid\ N(0,\ \sigma^2)$ es equivalente a suponer $\mathbf{u} \sim N_n\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$ pues el producto de las densidades Normales marginales independientes coincide con la conjunta de la Normal Multivariada:

$$\prod_{i=1}^{n} f_{u_i}(u_i) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{u_i^2}{2\sigma^2}} = \left(2\pi\sigma^2\right)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}u_i^2\right\} = f_{\mathbf{u}}(\mathbf{u}).$$

La propiedad más importante de la Distribución Normal Multivariada es que **cualquier transformación lineal** que se le aplique **continua siendo Normal Multivariada**, como se establece en el siguiente teorema.

Teorema. Transformaciones Lineales de la Distribución Normal Multivariada

Si $\mathbf{w} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{A}$ es matriz de dimensión $m \times n$, $\mathbf{c}$ es vector de dimensión m, entonces $\mathbf{Aw} + \mathbf{c} \sim N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{c}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$

Note que por propiedades de valor esperado y varianza:

- $E[\mathbf{Aw} + \mathbf{c}] = \mathbf{A}E[\mathbf{w}] + \mathbf{c} = \mathbf{A}\boldsymbol{\mu} + \mathbf{c}$ (operador lineal); y
- $Var[\mathbf{Aw} + \mathbf{c}] = E\left[(\mathbf{Aw} + \mathbf{c} - \mathbf{A}\boldsymbol{\mu} - \mathbf{c})(\mathbf{Aw} + \mathbf{c} - \mathbf{A}\boldsymbol{\mu} - \mathbf{c})'\right] = E\left[\mathbf{A}(\mathbf{w} - \boldsymbol{\mu})(\mathbf{w} - \boldsymbol{\mu})'\mathbf{A}'\right]$

$$= \mathbf{A}E\left[(\mathbf{w} - \boldsymbol{\mu})(\mathbf{w} - \boldsymbol{\mu})'\right]\mathbf{A}' = \mathbf{A}Var[\mathbf{w}]\mathbf{A}' = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$$

Proposición

Bajo los supuestos del MRLMN $\mathbf{y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$.

Demostración: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} = \mathbf{u} + \mathbf{X}\boldsymbol{\beta}$ es transformación lineal de $\mathbf{u} \sim N_n\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$; aplicando el teorema anterior con $\mathbf{A} = \mathbf{I}$ y $\mathbf{c} = \mathbf{X}\boldsymbol{\beta}$ se concluye que $\mathbf{y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$.

Teorema. EMV del MRLMN

Para el MRLMN $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde $\mathbf{u} \sim N_n\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$:

i) $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{y}\right)$ es el EMV de $\boldsymbol{\beta}$; y

ii) $\hat{\sigma}_*^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n}$ es el EMV de σ^2.

Ejercicio $\boxed{\text{E36}}$.

Es decir, bajo los supuestos del MRLMN **los EMV y los estimadores MCO de $\boldsymbol{\beta}$ son los mismos**, por lo tanto $\hat{\beta}_0$, $\hat{\beta}_1, \ldots, \hat{\beta}_k$ poseen la propiedad de **invarianza**, así como las **propiedades asintóticas de los EMV**, particularmente **eficiencia y consistencia**. Desafortunadamente el EMV de σ^2, $\hat{\sigma}_*^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n}$, no es insesgado, por eso se preferirá estimar a σ^2 mediante el estimador insesgado $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$.

Se sabe que si $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{y}\right)$ entonces $E\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$ y $Var\left[\hat{\boldsymbol{\beta}}\right] = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$. Bajo los supuestos del MRLMN $\mathbf{y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$ y $\hat{\boldsymbol{\beta}}$ es transformación lineal de **y**, por lo tanto su distribución de muestreo es Normal, es decir, $\hat{\boldsymbol{\beta}} \sim N_{k+1}\left(\boldsymbol{\beta}, \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$.

Además, como $\hat{\boldsymbol{\beta}}$ tiene una distribución conjunta Normal Multivariada, sus respectivas distribuciones marginales son Normales Univariadas, es decir, $\hat{\beta}_j \sim N\left(\beta_j, \sigma^2_{\hat{\beta}_j}\right)$, donde $\sigma^2_{\hat{\beta}_j} = Var\left[\hat{\beta}_j\right]$ es el *j*-ésimo elemento de la diagonal principal de la matriz de varianzas-covarianzas $\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$, j = 0, 1,…, k. A partir de estas distribuciones es posible construir estadísticos Z_j con distribución Normal Estándar.

Las varianzas en realidad nunca son conocidas pero sus respectivas estimaciones, $S^2_{\hat{\beta}_j} = \hat{\sigma}^2_{\hat{\beta}_j}$, se pueden obtener sustituyendo a σ^2 por $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$ en $\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ para construir estadísticos T_j con distribución *t*-Student.

Por su parte $\mathbf{u} \sim N_n\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$, es decir, **u** sigue una distribución Normal Multivariada con componentes no correlacionados (e independientes), y sus distribuciones marginales también son Normales Univariadas con $u_i \sim N(0, \sigma^2)$. Estandarizando y elevando al cuadrado se obtiene $\dfrac{u_i^2}{\sigma^2} \sim \chi(1)$, por lo tanto $\dfrac{u_1^2 + u_2^2 + \cdots + u_n^2}{\sigma^2} \sim \chi(n)$, es decir, la distribución de muestreo asociada a $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$ tiene una distribución Ji-Cuadrada.

Distribuciones de muestreo del MRLMN

Bajo los supuestos del MRLMN $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde $\mathbf{u} \sim N_n\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$, si $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{y}\right)$ y $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$, entonces:

i) $Z_j = \dfrac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \sim N(0,1)$, donde $\sigma^2_{\hat{\beta}_j}$ se obtiene de $\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$, j = 0, 1,…, k

ii) $T_j = \dfrac{\hat{\beta}_j - \beta_j}{S_{\hat{\beta}_j}} \sim t(n-(k+1))$, donde $S^2_{\hat{\beta}_j}$ se obtiene de $\hat{\sigma}^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$, j = 0, 1,…, k

iii) $J = \dfrac{[n-(k+1)]\hat{\sigma}^2}{\sigma^2} = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sigma^2} \sim \chi^2(n-(k+1))$

Estas estadísticas son **cantidades pivotales** que permiten fácilmente construir **intervalos de confianza** y hacer **pruebas de hipótesis** sobre los parámetros β_0, β_1,…, β_k y σ^2.

3.8. Estimación por intervalo y pruebas de hipótesis.

3.8.1. Intervalos de confianza sobre los parámetros del MRLMN.

La figura 3.2 muestra el resumen de los intervalos simétricos (misma área en cada cola) con nivel de confianza 1 – α para estimar por intervalo los parámetros β_0, β_1,…, β_k y σ^2, bajo los supuestos del MRLMN.

Figura 3.2

Parámetro	**Pivote**	**Intervalo de confianza simétrico con nivel de confianza 1 – α**
β_j	$T_j = \dfrac{\hat{\beta}_j - \beta_j}{S_{\hat{\beta}_j}} \sim t(n-(k+1))$	$\hat{\beta}_j \pm t_{n-(k+1),\frac{\alpha}{2}} S_{\hat{\beta}_j}$ donde $S^2_{\hat{\beta}_0}$ se obtiene de $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} = \hat{\sigma}^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$
σ^2	$J = \dfrac{[n-(k+1)]\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-(k+1))$	$\left[\dfrac{[n-(k+1)]\hat{\sigma}^2}{\chi^2_{n-(k+1),\frac{\alpha}{2}}}, \dfrac{[n-(k+1)]\hat{\sigma}^2}{\chi^2_{n-(k+1),1-\frac{\alpha}{2}}}\right]$

En el contexto multivariado es posible hablar de **conjuntos de confianza** o de **regiones de confianza** para $\boldsymbol{\beta}$ (versión multivariada de los intervalos de confianza), pero en la práctica estos conjuntos no se utilizan.

Ejercicio $\boxed{\text{E37}}$.

3.8.2. Pruebas de hipótesis acerca de los parámetros del MRLMN.

Tomando en cuenta que en el contexto del MRLMN puede haber muchos más parámetros que en el MRLSN y que la inferencia se puede de manera individual sobre cada parámetro o de manera conjunta sobre todos o algunos de ellos, se considerarán 3 tripos de pruebas de hipótesis:

- Individuales para $\beta_0, \beta_1, \ldots, \beta_k$ y σ^2.
- Conjunta para $\boldsymbol{\beta}$.
- Conjunta para alguna transformación lineal de $\boldsymbol{\beta}$.

PRUEBAS DE HIPÓTESIS INDIVIDUALES

Con base en la dualidad entre intervalos de confianza y regiones de rechazo de las pruebas de hipótesis, la figura 3.3 muestra regiones de rechazo (RR) y estadísticos de prueba para Pruebas de Hipótesis individuales de tamaño α (de una y dos colas) los parámetros $\beta_0, \beta_1, \ldots, \beta_k$ y σ^2, bajo los supuestos del MRLMN.

Figura 3.3

H_0:	vs. H_1:	$RR_\alpha =$	Estadístico de prueba
$\beta_j = \beta_{j,0}$	$\begin{cases}\beta_j < \beta_{j,0}\\ \beta_j > \beta_{j,0}\\ \beta_j \neq \beta_{j,0}\end{cases}$	$\begin{cases}t_j < -t_{n-(k+1),\alpha}\\ t_j > t_{n-(k+1),\alpha}\\ \lvert t_j\rvert > t_{n-(k+1),\frac{\alpha}{2}}\end{cases}$	$T_j = \dfrac{\hat{\beta}_j - \beta_{j,0}}{S_{\hat{\beta}_j}} \sim t(n-(k+1))$
$\sigma^2 = \sigma_0^2$	$\begin{cases}\sigma^2 < \sigma_0^2\\ \sigma^2 > \sigma_0^2\\ \sigma^2 \neq \sigma_0^2\end{cases}$	$\begin{cases}j < \chi^2_{n-(k+1),1-\alpha}\\ j > \chi^2_{n-(k+1),\alpha}\\ j < \chi^2_{n-(k+1),1-\frac{\alpha}{2}},\ j > \chi^2_{n-(k+1),\frac{\alpha}{2}}\end{cases}$	$J = \dfrac{[n-(k+1)]\hat{\sigma}^2}{\sigma_0^2} \sim \chi^2(n-(k+1))$

Las pruebas para $\beta_0, \beta_1, \ldots, \beta_k$, se denominan **Pruebas *T***. Es relevante:

- Rechazar H_0: $\beta_0 = 0$, pues de lo contrario es necesario evaluar la necesidad de hacer la estimación del MRLMN sin intercepto, con la posibilidad de que los residuos no sumen cero o que R^2 deje de tener interpretación como porcentaje; y
- Rechazar H_0: $\beta_j = 0$, $j = 1, 2, \ldots, k$, pues de lo contrario se debería hacer la estimación sin la variable explicativa X_j pues individualmente (manteniendo las demás variables explicativas) no explica a la variable de interés Y.

La mayor parte de los paquetes estadísticos estiman el MRLMN por MCO y presentan los estimados $\hat{\beta}_0$, $\hat{\beta}_1, \ldots, \hat{\beta}_k$ junto con sus errores estándar estimados $S_{\hat{\beta}_0}$, $S_{\hat{\beta}_1}, \ldots, S_{\hat{\beta}_k}$, y sus respectivos valores-p de las Pruebas T (H_0: $\beta_j = 0$, $j = 0, 1, \ldots, k$).

Importante: Que cada Prueba T resulte significativa no implica que el modelo completo resulte significativo. Para probar esto último hay que hacer el análisis conjunto.

PRUEBA DE HIPÓTESIS CONJUNTA DE LA REGRESIÓN

Proposición

Bajo los supuestos del MRLMN, si $\boldsymbol{\beta} = \mathbf{0}$, $F = \dfrac{CME}{CMR} = \dfrac{[n-(k+1)]SCE}{k\,SCR} \sim F(k, n-(k+1))$

Teorema. Prueba *F* (ANOVA).

Para el MRLMN $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde $\mathbf{u} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$, si se quiere probar $H_0: \beta_0 = \beta_1 = \cdots = \beta_k = 0$ vs. H_1: alguna $\beta_j \neq 0$, entonces $RR_\alpha = \{ f > f_{k,n-(k+1),\alpha} \}$, donde $F = \dfrac{CME}{CMR} = \dfrac{[n-(k+1)]SCE}{k\,SCR} \sim F(k, n-(k+1))$.

Note que la hipótesis nula se puede expresar en forma compacta como H_0: $\boldsymbol{\beta} = \mathbf{0}$.

PRUEBAS DE HIPÓTESIS CONJUNTAS PARA TRANSFORMACIONES LINEALES

En el contexto multivariado, en algunas aplicaciones puede ser relevante hacer inferencia acerca de **transformaciones lineales de algunos de los parámetros** del MRLMN, por ejemplo, $\beta_j = 0$ para alguna j = 0, 1,…, k, $\beta_1 + \beta_2 + \cdots + \beta_k = 1$, $\beta_1 = \beta_2 = \cdots = \beta_k = 0$, $\beta_0 = \beta_1$ y $\beta_k = \beta_{k,0}$, etc.

Todas estas transformaciones se pueden expresar mediante una matriz **R** de dimensión $q \times (k + 1)$, donde q es el número de restricciones sobre los parámetros involucrados en la prueba de hipótesis, de modo que bajo H_0 se quiere probar $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, donde **r** es un vector de dimensión q que contiene los valores fijos $r_i = R_i(\boldsymbol{\beta})$, y donde R_i: $\mathbf{R}^{k+1} \to \mathbf{R}$. La figura 3.4 muestra la forma que tomaría **R** para modelar las transformaciones lineales de parámetros del MRLMN mencionadas previamente.

Figura 3.4

Hipótesis nula en la forma H_0: $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$	q	Estructura de **R** de dimensión $q \times (k + 1)$	Estructura de **r** de dimensión q
$\beta_j = 0$	1	$\mathbf{R} = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]$ con 1 en la columna $j + 1$	$\mathbf{r} = 0$
$\beta_1 + \beta_2 + \cdots + \beta_k = 1$	1	$\mathbf{R} = [0 \ 1 \ \cdots \ 1]$	$\mathbf{r} = 1$
$\beta_1 = 0$ $\beta_2 = 0$ $\vdots$ $\beta_k = 0$	k	$\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$	$\mathbf{r} = \mathbf{0}_k$
$\beta_0 - \beta_1 = 0$ $\beta_k = \beta_{k,0}$	2	$\mathbf{R} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$	$\mathbf{r} = \begin{bmatrix} 0 \\ \beta_{k,0} \end{bmatrix}$

Proposición

Bajo los supuestos del MRLMN, si $\mathbf{R}$ es matriz de dimensión $q \times (k+1)$, $\mathbf{r}$ es vector de dimensión q, $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ (o bien $\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}_q$) y $\boldsymbol{\beta}$ se estima por MCO, entonces:

$$F = \frac{[n-(k+1)]\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right)'\left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right)}{q\,\hat{\mathbf{u}}'\hat{\mathbf{u}}} \sim F(q, n-(k+1))$$

En la estadística F de la proposición anterior el término $\frac{n-(k+1)}{\hat{\mathbf{u}}'\hat{\mathbf{u}}} = \frac{1}{\hat{\sigma}^2} = \left(\hat{\sigma}^2\right)^{-1}$ podría incorporarse en la matriz inversa del numerador para completar la matriz de varianzas-covarianzas estimada de $\hat{\boldsymbol{\beta}}$, es decir, $\hat{\boldsymbol{\Sigma}}_{\hat{\beta}} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$.

Teorema. Prueba *F* (Transformación Lineal General).

Para el MRLMN $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde $\mathbf{u} \sim N_n\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$, si $\mathbf{R}$ es la matriz de q restricciones sobre los parámetros $\boldsymbol{\beta}$, $\mathbf{r}$ es el vector de valores de las q restricciones y se quiere probar $H_0 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}_q$ vs. $H_0 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r} \neq \mathbf{0}_q$, entonces $RR_\alpha = \left\{ f > f_{q,n-(k+1),\alpha} \right\}$,
donde $F = \frac{1}{q}\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right)'\left[\mathbf{R}\hat{\boldsymbol{\Sigma}}_{\hat{\beta}}\mathbf{R}'\right]^{-1}\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \sim F(q, n-(k+1))$.

Una forma alternativa de probar relaciones de igualdad entre distintos parámetros del vector $\boldsymbol{\beta}$ es a través del Método de *Mínimos Cuadrados Restringidos* (MCR). Este método considera que hay q restricciones lineales bajo H_0.

Método de Mínimos Cuadrados Restringidos (MCR)

1) Modificar el MRLMN para incorporar las q restricciones.
2) Estimar el modelo restringido por MCO.
3) Calcular la suma de cuadrados del modelo restringido (SCR_R).
4) Aplicar la Prueba F de Mínimos Cuadrados Restringidos.

Teorema. Prueba *F* (MCR).

H_0: Se cumplen las q restricciones sobre $\boldsymbol{\beta}$ vs. H_1: Alguna restricción no se cumple, $RR_\alpha = \left\{ f > f_{q,n-(k+1),\alpha} \right\}$, donde $F = \frac{[n-(k+1)](SCR_R - SCR_{NR})}{q\,SCR_{NR}} \sim F(q, n-(k+1))$ y los subíndices indican R = Restringido y NR = No restringido.

Ejercicios $\boxed{\text{E38}}$ y $\boxed{\text{E39}}$.

3.9. Predicción en el MRLM.

En forma análoga al MRLS, en el MRLM se considerará el **pronóstico o predicción** de la *media condicional* $E\left[Y_0 \middle| X_{1,0}, X_{2,0}, \ldots, X_{k,0}\right]$ o del *valor individual* Y_0 para un conjunto de valores fijos $\mathbf{x}_0 = \left[1, X_{1,0}, X_{2,0}, \ldots, X_{k,0}\right]'$ de las variables explicativas X_1, X_2,…, X_k. Note que el vector $\mathbf{x}_0$ es de dimensión $k + 1$ porque además de los valores de las k variables, su primer elemento es un 1.

Por practicidad se denotará a la **media condicional** por $E\left[Y_0 \middle| \mathbf{x}_0\right] = \mathbf{x}_0'\boldsymbol{\beta}$ y al **valor individual** por $Y_0 = \mathbf{x}_0'\boldsymbol{\beta} + u_0$.

3.9.1. Predicción de la media condicional.

Teorema. Predicción puntual de la media condicional

Bajo los supuestos del MRLM, el estimador puntual de la media condicional $E\left[Y_0 \middle| \mathbf{x}_0\right] = \mathbf{x}_0'\boldsymbol{\beta}$ es $\hat{E}\left[Y_0 \middle| \mathbf{x}_0\right] = \hat{Y}_0^* = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$, donde:

i) $\hat{Y}_0^*$ es MELI.

ii) $Var\left[\hat{Y}_0^*\right] = \sigma^2\left[\mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0\right]$

Para calcular $S^2_{\hat{Y}_0^*}$, el estimador de $Var\left[\hat{Y}_0^*\right]$, basta sustituir σ^2 por $\hat{\sigma}^2 = \dfrac{\mathbf{u}'\mathbf{u}}{n-(k+1)}$.

Bajo el supuesto de normalidad en los errores es fácil ver que $\hat{Y}_0^*$ sigue una distribución Normal, de modo que $\dfrac{\hat{Y}_0^* - \mathbf{x}_0'\boldsymbol{\beta}}{\sqrt{Var\left[\hat{Y}_0^*\right]}} \sim N(0,1)$, y sustituyendo $Var\left[\hat{Y}_0^*\right]$ por $S^2_{\hat{Y}_0^*}$ se obtiene el pivote $\dfrac{\hat{Y}_0^* - \mathbf{x}_0'\boldsymbol{\beta}}{S_{\hat{Y}_0^*}} \sim t\left(n-(k+1)\right)$.

Teorema. Predicción por intervalo de la media condicional

Bajo los supuestos del MRLMN, el intervalo de confianza simétrico al 100(1 – α)% para el parámetro $E\left[Y_0 \middle| \mathbf{x}_0\right] = \mathbf{x}_0'\boldsymbol{\beta}$ es $\mathbf{x}_0'\hat{\boldsymbol{\beta}} \pm t_{n-(k+1),\frac{\alpha}{2}} S_{\hat{Y}_0^*}$, donde el error estándar estimado se calcula a partir de $S^2_{\hat{Y}_0^*} = \hat{\sigma}^2\mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0 = \mathbf{x}_0'\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\mathbf{x}_0$

3.9.2. Predicción individual.

Dado que la variables explicativas X_1, X_2,…, X_k toman los valores fijos de $\mathbf{x}_0$ el respectivo **valor individual** Y_0 está dado por $Y_0 = \mathbf{x}_0'\boldsymbol{\beta} + u_0$. Como los errores tienen media cero, lo más razonable es eliminarlo en la estimación ($\hat{u}_0 = 0$) y simplemente sustituir los estimadores MCO de $\boldsymbol{\beta}$.

Teorema. Predicción puntual de valores individuales

> Bajo los supuestos del MRLM, el estimador puntual del valor individual $Y_0 = \mathbf{x}_0'\boldsymbol{\beta} + u_0$ es $\hat{Y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$, donde:
>
> *i*) $\hat{Y}_0$ es MELI.
>
> *ii*) $Var\left[Y_0 - \hat{Y}_0\right] = \sigma^2\left[1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\right]$

A la diferencia $e_0 = Y_0 - \hat{Y}_0$ se le denomina **error de predicción**. Para estimar su varianza, $S^2_{Y_0-\hat{Y}_0}$, basta sustituir σ^2 por $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$.

Importante: Al pronosticar la media condicional $E\left[Y_0 | \mathbf{x}_0\right] = \mathbf{x}_0'\boldsymbol{\beta}$ se está estimando un parámetro y en ese contexto es posible hablar de intervalos de confianza. Al pronosticar un **valor individual** $Y_0 = \mathbf{x}_0'\boldsymbol{\beta} + u_0$ no se está pronosticando un parámetro sino una **variable aleatoria** y por lo tanto no es posible hablar de intervalos de confianza sino de **intervalos de predicción**.

Bajo el supuesto de normalidad es fácil ver que $e_0 = Y_0 - \hat{Y}_0$ tiene una distribución Normal con media cero y varianza $Var\left[Y_0 - \hat{Y}_0\right]$, de modo que $\dfrac{Y_0 - \hat{Y}_0}{\sqrt{Var\left[Y_0 - \hat{Y}_0\right]}} \sim N(0,1)$, y sustituyendo $Var\left[Y_0 - \hat{Y}_0\right]$ por $S^2_{Y_0-\hat{Y}_0}$ se obtiene la estadística $\dfrac{Y_0 - \hat{Y}_0}{S_{Y_0-\hat{Y}_0}} \sim t(n-(k+1))$.

Teorema. Predicción por intervalo de valores individuales

> Bajo los supuestos del MRLMN, el intervalo de predicción simétrico al 100(1 – α)% para $Y_0 = \mathbf{x}_0'\boldsymbol{\beta} + u_0$ es $\mathbf{x}_0'\hat{\boldsymbol{\beta}} \pm t_{n-(k+1),\frac{\alpha}{2}} S_{Y_0-\hat{Y}_0}$, donde el error estándar estimado se calcula a partir de $S^2_{Y_0-\hat{Y}_0} = \hat{\sigma}^2\left[1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\right] = \hat{\sigma}^2 + \mathbf{x}_0'\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\mathbf{x}_0$

Es importante recalcar que la longitud del intervalo de predicción del valor individual Y_0 siempre es mayor que la longitud del intervalo de confianza de la media condicional $E[Y_0|\mathbf{x}_0] = \mathbf{x}_0'\boldsymbol{\beta}$.

Al considerar más de una variable explicativa no es práctico considerar conjuntos de predicción (generalización de las bandas de predicción).

Ejercicios $\boxed{\text{E40}}$ y $\boxed{\text{E41}}$.

3.10. Modelos de Regresión Polinomial.

Def. Modelo de Regresión Polinomial (MRP).

Si la variable dependiente Y se explica mediante la variable explicativa X y su FRP es de la forma $E[Y_i|X_i] = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k$, entonces el Modelo de Regresión Polinomial supone que los valores Y_i pueden ser modelados por esta media condicional y un término de error de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + u_i,$$

donde X_i es no estocástica y $u_i \sim iid(0,\ \sigma^2)$.

Para expresar el MRP en forma matricial es necesario considerar siguiente matriz de diseño (*Matriz de Vandermonde*):

$$\mathbf{X}_P = \begin{bmatrix} 1 & X_1 & X_1^2 & \cdots & X_1^k \\ 1 & X_2 & X_2^2 & \cdots & X_2^k \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_n & X_n^2 & \cdots & X_n^k \end{bmatrix}$$

Entonces el MRP se puede expresar como $\mathbf{y} = \mathbf{X}_P\boldsymbol{\beta} + \mathbf{u}$, donde $\mathbf{u} \sim iid(\mathbf{0}, \sigma^2\mathbf{I})$, es decir, un MRLM pero con una sola variable explicativa.

A pesar de que el MRP es una extensión del MRLS (una sola variable explicativa), su estimación se realiza por MCO del respectivo MRLM.

Para este tipo de modelos sí es posible construir bandas de confianza, siempre y cuando haya normalidad en los errores, es decir, $\mathbf{u} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$.

Ejercicio $\boxed{\text{E42}}$.

3.11. Contribuciones incrementales de una o varias variables.

3.11.1. Análisis de los parámetros del MRLM.

SIGNIFICADO DE LOS COEFICIENTES DE REGRESIÓN

La relación determinista del MRLM es $Y = \beta_0 + \beta_1 X_1 + \beta_1 X_2 + \cdots + \beta_k X_k$, al parámetro β_0 se le denomina **intercepto**, mientras que a los parámetros β_1, β_2,…, β_k se les denomina **coeficientes de regresión parcial** y es fácil ver que corresponden con la derivada parcial de Y respecto a cada regresor, es decir, $\frac{\partial Y}{\partial X_j} = \beta_j$, j = 1, 2,…, k.

MEDIDAS DE SENSIBILIDAD

Para el **análisis marginal** de $Y = g(X_1, X_2, \ldots, X_k)$ se consideran las medidas de sensibilidad estudiadas previamente (efecto parcial, elasticidad y semi-elasticidad). Sus definiciones son muy similares a las del caso univariado pero cambian la notación de derivada a derivada parcial, pues ahora hay k variables explicativas:

- *Efecto parcial*: $\Delta Y_j \approx \left(\frac{\partial Y}{\partial X_j}\right)\Delta X_j$, es el cambio aproximado en Y asociado a un cambio pequeño en X_j, considerando todo lo demás constante (*caeteris paribus*).
- *Elasticidad*: $\eta_j = \left(\frac{\partial Y}{\partial X_j}\right)\frac{X_j}{Y} = \frac{\partial \ln(Y)}{\partial \ln(X_j)}$ es el cambio porcentual aproximado en Y cuando X_j aumenta *un punto porcentual*, considerando todo lo demás constante.
- *Semi-elasticidad*: $\overline{\eta}_j = 100\left(\frac{\partial Y}{\partial X_j}\right)\frac{1}{Y}$ es el cambio porcentual aproximado en Y cuando X_j aumenta en *una unidad*, considerando todo lo demás constante.

Para el **análisis conjunto** de $Y = \beta_0 + \beta_1 X_1 + \beta_1 X_2 + \cdots + \beta_k X_k$ se define el *efecto total* como se muestra a continuación.

Def. Efecto total. En el MRLM el efecto total, denotado por ΔY, es el cambio aproximado en Y asociado a un pequeño cambio en cada X_j, j = 1, 2,…, k, es decir, $\Delta Y \approx \left(\frac{\partial Y}{\partial X_1}\right)\Delta X_1 + \left(\frac{\partial Y}{\partial X_2}\right)\Delta X_2 + \cdots + \left(\frac{\partial Y}{\partial X_k}\right)\Delta X_k$.

Si se calcula el efecto total considerando un cambio unitario en cada variable explicativa entonces $\Delta Y \approx \beta_1 + \beta_2 + \cdots + \beta_k$.

3.11.2. Análisis de los parámetros estimados del MRLM.

En las secciones anteriores se analizaron las técnicas estadísticas para estimar el MRLM, sin embargo, es importante vincular el resultado de la estimación con el significado de los parámetros.

Analizando el MRLM con 2 regresores. Al estimar $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$ por MCO se obtienen directamente $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\beta}_2$. En particular $\hat{\beta}_1$ es la estimación del coeficiente de regresión parcial β_1, es decir, cuánto cambia en términos absolutos Y ante un aumento de X_1 en una unidad. Es posible calcular $\hat{\beta}_1$ de manera indirecta mediante el siguiente método.

Método de estimación indirecta

Para estimar en forma indirecta a β_1 en $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$:

1) Estimar por MCO los modelos $Y_i = \alpha_0 + \alpha_1 X_{2i} + v_{Y,i}$ y $X_{1i} = \gamma_0 + \gamma_1 X_{2i} + v_{X_1,i}$ para aislar el efecto de X_2.
2) Calcular los residuos $\hat{v}_{Y,i}$ y $\hat{v}_{X_1,i}$, pues contienen la parte que X_2 no pudo explicar de Y y de X_1, respectivamente.
3) Estimar por MCO el modelo $\hat{v}_{Y,i} = \delta \hat{v}_{X_1,i} + w_i$, sin considerar intercepto pues $E[v_{Y,i}] = 0$.
4) En la regresión de residuos del punto anterior, δ es la parte que X_1 explica de Y sin tomar en cuenta el efecto de X_2, por lo tanto $\hat{\beta}_1 = \hat{\delta}$.

Ejercicio $\boxed{\text{E43}}$.

3.11.3. Coeficiente de Correlación Parcial.

En el contexto multivariado la **correlación** juega un papel importante pues permite analizar qué tanto contribuyen los regresores X_1, X_2,…, X_k en la explicación de la variable dependiente Y, ya sea en forma **conjunta o en forma individual**.

Al coeficiente de correlación lineal muestral entre las variables Y y X_j, $r_{YX_j} = \dfrac{S_{YX_j}}{S_Y S_{X_j}}$, se le denomina **coeficiente de correlación simple**, y por practicidad se simplificará su notación a r_{Yj}, de modo que es posible calcular k coeficientes de correlación simples de este tipo.

En el MRLS es fácil demostrar que $R^2 = (r_{XY})^2$, es decir, que el coeficiente de determinación es simplemente el coeficiente de correlación elevado al cuadrado. En el contexto del MRLM existen k posibles coeficientes de correlación simples y sus cuadrados no necesariamente coinciden con el coeficiente de determinación. Sin embargo, por analogía, a la raíz cuadrada del coeficiente de determinación del MRLM se le conoce como *coeficiente de correlación múltiple*.

Def. Coeficiente de Correlación Múltiple

$$R = \sqrt{R^2}$$

R es una medida del grado de asociación entre Y y todas las variables explicativas en conjunto. R siempre es positivo por lo que en la práctica tiene poca importancia y se prefieren R^2 y $\overline{R}^2$ como medidas de bondad de ajuste.

Para analizar la correlación entre Y y cada variable explicativa de manera individual, aislando el efecto de todas las demás variables (*caeteris paribus*), se define *coeficiente de correlación parcial*.

Coeficiente de correlación parcial con 2 regresores. Para el MRLM con 2 regresores $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$ es posible definir 2 coeficientes de regresión parcial:

- El coeficiente de correlación parcial entre Y y X_1, manteniendo constante X_2 o aislando la influencia de X_2, denotado por $r_{Y1.2}$; y
- El coeficiente de correlación parcial entre Y y X_2, manteniendo constante X_1 o aislando la influencia de X_1, denotado por $r_{Y2.1}$.

Siguiendo la lógica del método de estimación indirecta, para calcular $r_{Y1.2}$ hay que aislar la influencia de X_2 mediante los MRLS $Y_i = \alpha_0 + \alpha_1 X_{2i} + v_{Y.2,i}$ y $X_{1i} = \gamma_0 + \gamma_1 X_{2i} + v_{1.2,i}$. Al estimar estos modelos por MCO se obtienen los vectores de residuos $\hat{\mathbf{v}}_{Y.2}$ y $\hat{\mathbf{v}}_{1.2}$, que contienen las partes de Y y X_1 que X_2 no pudo explicar. Finalmente el coeficiente de correlación parcial $r_{Y1.2}$ es el coeficiente de correlación lineal muestral entre estos dos conjuntos de residuos (cuyas medias muestrales son cero).

Def. Coeficiente de Correlación Parcial (con 2 regresores)

Para el MRLM con 2 regresores $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$:

i) $$r_{Y1.2} = r_{\hat{v}_{Y.2},\hat{v}_{1.2}} = \frac{\hat{\mathbf{v}}'_{Y.2}\hat{\mathbf{v}}_{1.2}}{\sqrt{(\hat{\mathbf{v}}'_{Y.2}\hat{\mathbf{v}}_{Y.2})(\hat{\mathbf{v}}'_{1.2}\hat{\mathbf{v}}_{1.2})}}$$

ii) $$r_{Y2.1} = r_{\hat{v}_{Y.1},\hat{v}_{2.1}} = \frac{\hat{\mathbf{v}}'_{Y.1}\hat{\mathbf{v}}_{2.1}}{\sqrt{(\hat{\mathbf{v}}'_{Y.1}\hat{\mathbf{v}}_{Y.1})(\hat{\mathbf{v}}'_{2.1}\hat{\mathbf{v}}_{2.1})}}$$

Los coeficientes de correlación parcial se pueden calcular a partir de los coeficientes de correlación simples mediante las fórmulas de la siguiente proposición.

Proposición.

Para el MRLM con 2 regresores $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$:

i) $$r_{Y1.2} = \frac{r_{Y1} - r_{Y2} r_{12}}{\sqrt{\left(1 - r_{Y2}^2\right)\left(1 - r_{12}^2\right)}}$$

ii) $$r_{Y2.1} = \frac{r_{Y2} - r_{Y1} r_{21}}{\sqrt{\left(1 - r_{Y1}^2\right)\left(1 - r_{21}^2\right)}}$$

La **interpretación** del coeficiente de correlación parcial es muy similar a la del coeficiente de correlación simple: $r_{Y1.2}$ indica qué tan fuerte es la asociación lineal entre la variable de interés *Y* y el regresor X_1, manteniendo constante o aislado el efecto de X_2. Igual que en el caso simple, $-1 \le r_{Y1.2} \le 1$, de modo que si $r_{Y1.2} \to -1$ indica un alto grado de asociación lineal negativo, $r_{Y1.2} \to 1$ indica un alto grado de asociación lineal positivo y $r_{Y1.2} \to 0$ indica que no hay asociación o que la asociación no es de tipo lineal. Si existe alguna *correlación espuria* (sin lógica o sin fundamento) entre *Y* y X_2 o entre X_1 y X_2 el coeficiente de correlación parcial la elimina.

En este contexto al coeficiente de determinación R^2 también se le denota por $R_{Y.12}^2$ y puede ser calculado mediante los coeficientes de correlación simples y parciales.

Proposición.

Para el MRLM con 2 regresores $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$:

i) $R^2 = R_{Y.12}^2 = r_{Y1}^2 + r_{Y2.1}^2\left(1 - r_{Y1}^2\right)$; o

ii) $R^2 = R_{Y.12}^2 = r_{Y2}^2 + r_{Y1.2}^2\left(1 - r_{Y2}^2\right)$

Def. Coeficiente de determinación parcial. Al coeficiente de correlación parcial al cuadrado $r_{Y2.1}^2$ también se le denomina *coeficiente de determinación parcial* y se interpreta como la proporción de la variación cuadrática de *Y* no explicada por la variable X_1 que se explica por la variable X_2.

Análisis de contribución incremental. La principal aplicación de los coeficientes de correlación parciales es el *análisis de contribución incremental o marginal* que tienen los regresores en el MRLM. Este análisis consiste en desglosar la *SCE* de la Tabla ANOVA. La **Tabla ANOVA de contribución incremental** se muestra en la figura 3.5

Figura 3.5

Fuente de variación	**Suma de cuadrados**		**Fuente de variación**	**Suma de cuadrados**
X_1	$SCE_1 = r_{Y1}^2\left(\mathbf{y}'\mathbf{y} - n\overline{Y}^2\right)$		X_2	$SCE_2 = r_{Y2}^2\left(\mathbf{y}'\mathbf{y} - n\overline{Y}^2\right)$
Incremento debido a X_2	$SCE_{2.1} = r_{Y2.1}^2\left(1 - r_{Y1}^2\right)\left(\mathbf{y}'\mathbf{y} - n\overline{Y}^2\right)$		Incremento debido a X_1	$SCE_{1.2} = r_{Y1.2}^2\left(1 - r_{Y2}^2\right)\left(\mathbf{y}'\mathbf{y} - n\overline{Y}^2\right)$
X_1 y X_2	$SCE = R_{Y.12}^2\left(\mathbf{y}'\mathbf{y} - n\overline{Y}^2\right)$		X_1 y X_2	$SCE = R_{Y.12}^2\left(\mathbf{y}'\mathbf{y} - n\overline{Y}^2\right)$
Residuos	$SCR = \left(1 - R_{Y.12}^2\right)\left(\mathbf{y}'\mathbf{y} - n\overline{Y}^2\right)$		Residuos	$SCR = \left(1 - R_{Y.12}^2\right)\left(\mathbf{y}'\mathbf{y} - n\overline{Y}^2\right)$
Total	$SCT = \mathbf{y}'\mathbf{y} - n\overline{Y}^2$		Total	$SCT = \mathbf{y}'\mathbf{y} - n\overline{Y}^2$

A esta tabla se le pueden agregar columnas con los grados de libertad (1 para los dos primeros renglones, 2 para la contribución de X_1 y X_2, n – 3 para los residuos y n – 1, para el total), cuadrados medios y porcentaje de contribución.

Para validar la significancia del incremento debido a X_2 es posible aplicar la **Prueba *F* (ANOVA incremental)** con estadístico de prueba $F = \dfrac{(n-3)SCE_{2.1}}{SCR} \sim F(1, n-3)$; y para probar la significancia del incremento debido a X_1 $F = \dfrac{(n-3)SCE_{1.2}}{SCR} \sim F(1, n-3)$.

El análisis de contribución incremental se puede extender a más de 2 regresores mediante coeficientes de correlación parcial de orden superior.

Ejercicio $\boxed{\text{E44}}$.

Se dice que $r_{Y1.2}$ y $r_{Y1.1}$ son *coeficientes de correlación parcial de **primer orden***, pues sólo se está aislando el efecto de un regresor. Si se considera el MRLM con 3 regresores entonces es posible hablar de los *coeficientes de correlación parcial de **segundo orden*** $r_{Y1.23}$, $r_{Y2.13}$ y $r_{Y3.12}$; y así sucesivamente. A los *coeficientes de correlación simple* también se les denomina *coeficientes de correlación parcial de **orden cero***.

Coeficiente de correlación parcial con *k* regresores. Para el MRLM con k regresores $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ es posible definir k coeficientes de regresión parcial de orden k – 1: $r_{Y1.23\cdots k}$, $r_{Y2.13\cdots k}, \ldots, r_{Yk.12\cdots(k-1)}$. Para calcularlos es necesario definir las siguientes matrices:

- $\mathbf{X}_j$, que es una matriz de dimensión $k \times n$, y corresponde a la matriz de diseño $\mathbf{X}$ pero reducida al eliminar la columna $\mathbf{x}_j$ correspondiente al regresor X_j; y
- $\mathbf{M}_j = \mathbf{I} - \mathbf{H}_j = \mathbf{I} - \mathbf{X}_j\left(\mathbf{X}_j'\mathbf{X}_j\right)^{-1}\mathbf{X}_j'$, que es una matriz de dimensión $n \times n$, y se construye a partir de la *matriz gorro reducida* $\mathbf{H}_j$.

Def. Coeficiente de Correlación Parcial de Orden *k* – 1

Para el MRLM con k regresores $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, el coeficiente de correlación parcial de entre Y y X_j de orden $k - 1$ está dado por $r_{Yj.1\cdots(j-1)(j+1)\cdots k} = \dfrac{\mathbf{x}_j'\mathbf{M}_j\mathbf{y}}{\sqrt{(\mathbf{x}_j'\mathbf{M}_j\mathbf{x}_j)(\mathbf{y}'\mathbf{M}_j\mathbf{y})}}$.

3.12. Criterios para la selección de modelos.

3.12.1. Criterios de selección

El principal problema de la Econometría Aplicada al considerar el MRLM es determinar el conjunto de regresores $X_1, X_2, \ldots, X_k$ que expliquen a la variable de interés Y.

En la selección de regresores se debe tener presente que el MRLM debe satisfacer los siguientes criterios:

- **Adecuado para los datos**. Sus predicciones deben ser lógicamente posibles.
- **Consistente con la Teoría**. Si un regresor no cabe dentro de la teoría que sustenta el modelo no debe incluirse (aunque tenga una contribución incremental significativa).
- **Regresores exógenos débiles**. Las variables explicativas no deben estar correlacionadas con los errores.
- **Consistencia paramétrica**. Los parámetros deben ser estables para que el pronóstico fuera del conjunto muestral tenga validez (*Prueba de Chow*).
- **Coherente en los datos**. Los residuos deben ser *ruido blanco*: $u_i \sim iid(0, \sigma^2)$.
- **Inclusivo**. El modelo debe abarcar o incluir todo modelo contendiente, es decir, no puede haber mejor modelo que el elegido.

Estos criterios deben prevalecer siempre al construir un Modelo Econométrico.

Por regla general, si hay una teoría sólida que define a las variables explicativas de un modelo, no deben agregarse regresores adicionales.

Sin embargo, en muchas ocasiones no existe una teoría que ayude a definir los regresores y en esos casos lo que se acostumbra es analizar la relación entre la variable de interés y algunas otras variables que tentativamente podrían explicarla (ya sea por lógica o por algún conocimiento específico del ámbito de aplicación). Estas técnicas de "ensayo y error" forman parte de lo que actualmente se denomina **Minería de Datos (*data mining*)**.

3.12.2. Criterios de Información de Akaike, Schwarz y Hannan-Quinn

Se sabe que al aumentar el número de regresores el coeficiente de determinación (R^2) aumenta, revelando mayor bondad de ajuste (menor *SCR*). Pero por otro lado, siempre es preferible un modelo lo más simple posible o con el menor número de regresores posibles (*Principio de Parsimonia*).

Una forma de resolver este conflicto es mediante el coeficiente de determinación ajustado ($\overline{R}^2$), pues toma en cuenta el número de regresores considerados en el modelo y permite comparar la bondad de ajuste de distintos modelos con la misma variable de interés y muestras del mismo tamaño.

Otra alternativa es considerar los llamados **Criterios de Información**, que permiten analizar un adecuado balance entre bondad de ajuste y simpleza del modelo.

Los *criterios de información de Akaike, Schwarz y Hannan-Quinn* consideran estadísticas basadas en la *SCR* pero imponen un castigo por incluir un número creciente de regresores en el MRLM. Estos criterios establecen que **se debe elegir aquel modelo cuya estadística sea mínima**.

Def. Criterio de Información de Akaike (CIA)

$$CIA = \ln\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n}\right) + \frac{2(k+1)}{n}$$

Def. Criterio de Información de Schwarz (CIS)

$$CIS = \ln\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n}\right) + \frac{(k+1)\ln(n)}{n}$$

Def. Criterio de Información de Hannan-Quinn (CIHQ)

$$CIHQ = \ln\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n}\right) + \frac{2(k+1)\ln(\ln(n))}{n}$$

Nota: Algunos paquetes estadísticos al estimar un MRLM calculan automáticamente estas cantidades, o alguna variante de ellas. En particular, EViews reporta aproximadamente el doble de estas cantidades.

3.12.3. Métodos para la Construcción de Modelos Econométricos

La construcción de Modelos Econométricos requiere: (i) determinar qué variables explicativas potencialmente podrían formar parte del modelo; (ii) la forma funcional en que dichas variables se podrían incorporar al MRLM; y (iii) un método que permita decidir cuales variables deben permanecer en el modelo.

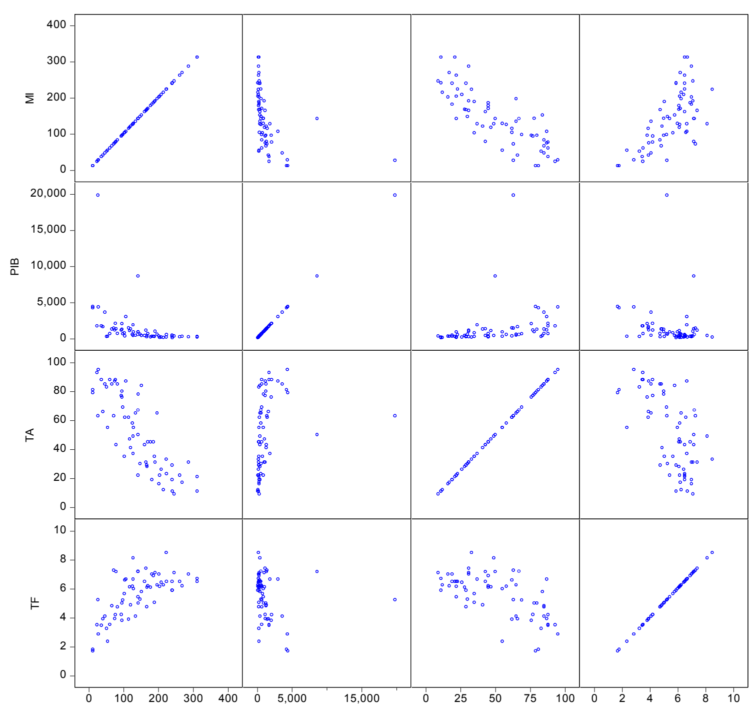
IDENTIFICACIÓN DE VARIABLES EXPLICATIVAS POTENCIALES

A falta de una teoría sólida que defina las variables explicativas de un modelo, se acostumbra analizar un conjunto de variables que lógicamente estén asociadas o vinculadas con la variable de interés. En este proceso se corre el riesgo de identificar *correlaciones espurias*, es decir, variables que están correlacionadas pero que carecen de una conexión lógica.

Un método gráfico que facilita la determinación de variables explicativas potenciales es el **Diagrama de Dispersión Múltiple**, que consiste en una matriz cuadrada que en cada uno de sus elementos (i, j) muestra el diagrama de dispersión entre la variable del renglón i y la variable de la columna j. La mayoría de los paquetes estadísticos generan este tipo de gráfico.

Por ejemplo, la figura 3.6 muestra el Diagrama de Dispersión Múltiple de EViews entre mortalidad infantil (MI), PIB *per cápita* (PIB), tasa de alfabetización femenina (TA) y tasa de fecundidad (TF) para una muestra de 64 países.

Figura 3.6



SELECCIÓN DE LA FORMA FUNCIONAL

El Diagrama de Dispersión Múltiple también permite identificar empíricamente la forma funcional en la que los regresores pueden incorporarse al MRLM.

Existen pruebas estadísticas para determinar si la forma funcional de un MRLM es la correcta. Por ejemplo, la *Prueba MWD* (MacKinnon, White y Davidson) permite determinar si un Modelo Log-Lin es preferido sobre un Modelo Lineal (aplicable a la función de producción Cobb-Douglas).

MÉTODOS DE SELECCIÓN DE VARIABLES

Estos métodos iterativos consisten en ir analizando la significancia estadística (individual y conjunta) del MRLM con la entrada o salida de variables explicativas.

Método ascendente (*stepwise regression*)

1) Formular el MRLS con el regresor más correlacionado. Si no es significativo (Prueba *T*) el modelo correcto es una constante. Si resulta significativo continuar.
2) Incorporar el regresor con correlación parcial de primer orden más alto, aislando el efecto del primer regresor. Si no es significativo (Prueba *F* incremental) el modelo correcto es un MRLS. Si resulta significativo continuar incluir y continuar.
3) Incorporar el regresor con correlación parcial de segundo orden más alto, aislando el efecto de los dos primeros regresores. Si no es significativo (Prueba *F* incremental) el modelo correcto es un MRLM con $k = 2$. Si resulta significativo incluir y continuar en forma similar hasta agotar el análisis de los regresores disponibles.

Método descendente (*backward elimination*)

1) Formular el MRLM con los k* los regresores disponibles.
2) Analizar la significancia incremental (Prueba *F* incremental) de cada uno de los k* regresores considerando a cada regresor como la última variable en haberse incorporado al MRLM.
3) El estadístico de prueba más bajo, f_L, se compara con un valor crítico preseleccionado f_0. Comúnmente $f_0 = f_{1,n-k^*,0.05}$. Si $f_L < f_0$ eliminar el regresor correspondiente y volver a aplicar estos tres pasos. Si $f_L > f_0$ adoptar el MRLM calculado en el primer paso de la última iteración.

Ejercicios E45 y E46 .

4. Variables Dicótomas

4.1. Naturaleza de las variables dicótomas.

4.1.1. Variables Dicótomas y Variables Cualitativas

En las secciones previas se han analizado los Modelos de Regresión en donde la variable de interés Y y las variables explicativas X_1, X_2,…, X_k son variables cuantitativas; sin embargo, también es posible utilizar **variables cualitativas (o variables categóricas) como variables explicativas** mediante el uso de las llamadas *variables dicótomas*.

Def. Variable Dicótoma.

Se dice que la variable D es dicótoma si sólo puede tomar dos valores: $D_i = 1$ si la i-ésima observación posee cierto atributo y $D_i = 0$ si no lo posee. A este tipo de variables también se les denomina variables instrumentales o variables *dummy*.

Por ejemplo, si se quiere explicar el salario mensual de las personas (Y) considerando si están titulados o no, entonces se puede plantear el modelo econométrico:

$$Y_i = \beta_0 + \beta_1 D_i + u_i\,,\ u_i \sim iid(0,\ \sigma^2),$$

donde: $D_i = 1$ si la i-ésima persona está titulada; y
$D_i = 0$ en caso contrario.

Con base en esta definición se presentan 2 posibilidades para Y:

$$Y_i = \begin{cases} \beta_0 + \beta_1 + u_i & si\ \ D_i = 1 \\ \beta_0 + u_i & si\ \ D_i = 0 \end{cases}$$

Este modelo econométrico es un MRLS que se puede estimar por MCO para obtener $\hat{\beta}_0$ y $\hat{\beta}_1$; y a partir de estos valores concluir que el salario mensual estimado para los titulados es de $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1$, mientras que para los no titulados es de $\hat{Y} = \hat{\beta}_0$.

Si la variable cualitativa presenta **más de 2 categorías** es necesario incorporar más variables dicótomas. Por ejemplo, si se quiere explicar el ingreso mensual de las personas (Y) considerando su situación laboral (activo, retirado o desempleado), entonces el modelo econométrico se puede expresar como:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i\,,\ u_i \sim iid(0,\ \sigma^2),$$

donde: $D_{1i} = 1$ si la i-ésima persona es activa y $D_{1i} = 0$ en caso contrario; y
$D_{2i} = 1$ si la i-ésima persona es retirada y $D_{2i} = 0$ en caso contrario.

Importante: A pesar de que hay 3 posibles categorías no es necesario agregar una cuarta variable dicótoma, pues las personas desempleadas quedarían descritas por la situación en que $D_1 = 0$ y $D_2 = 0$. Lo anterior atiende a que las *categorías* de una variable cualitativa deben ser *exhaustivas*.

Con base en esta definición se presentan 3 posibilidades:

$$Y_i = \begin{cases} \beta_0 + \beta_1 + u_i & si\ D_{1i} = 1\ y\ D_{2i} = 0 \\ \beta_0 + \beta_2 + u_i & si\ D_{1i} = 0\ y\ D_{2i} = 1 \\ \beta_0 + u_i & si\ D_{1i} = 0\ y\ D_{2i} = 0 \end{cases}$$

Como las *categorías* de una variable cualitativa deben ser *mutuamente excluyentes* no puede ocurrir que simultáneamente se tengan $D_{1i} = 1$ y $D_{2i} = 1$.

Este modelo econométrico es un MRLM que se puede estimar por MCO para obtener $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\beta}_2$; y a partir de estos valores concluir que el ingreso mensual estimado para los activos es $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1$, el ingreso mensual estimado de los retirados es $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_2$ y que el ingreso mensual estimado para los desempleados es $\hat{Y} = \hat{\beta}_0$.

Note que en los MRLM con variables dicótomas la matriz de diseño **X** conserva el vector de unos como primera columna (correspondientes al intercepto β_0), y el resto de sus columnas están formadas sólo por ceros y unos.

Proposición

Si una variable cualitativa tiene m categorías, hay que incluir $m - 1$ variables dicótomas.

Trampa de la variable dicótoma. Ocurre si por error se agregan las m variables dicótomas y se deja el intercepto, ya que esto conduce al problema de *multicolinealidad*, pues para un conjunto de n observaciones, la suma de las columnas de las m variables dicótomas forzosamente conducen a una columna de unos. En estos casos la matriz de diseño **X** deja de ser de rango completo y hace que $\mathbf{X}'\mathbf{X}$ sea singular. Una forma de resolver la "trampa de la variable dicótoma" es hacer la regresión sin intercepto.

Nota: No existe una forma única para incorporar las variables dicótomas en un Modelo de Regresión.

Significancia estadística de las variables dicótomas. Bajo el supuesto de *normalidad* en los errores es posible probar la significancia individual de las categorías (Pruebas T) o de la variable cualitativa en su conjunto (Prueba F).

4.1.2. Múltiples Variables Cualitativas e Interacciones

Es posible incorporar **varias variables cualitativas** en los Modelos de Regresión.

Por ejemplo, si para explicar el ingreso mensual de las personas (Y) además de considerar su situación laboral (activo, retirado o desempleado) se considera su género (hombre o mujer), se puede formular el modelo econométrico:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 C_i + u_i\,,\ u_i \sim iid(0,\ \sigma^2),$$

donde: $D_{1i} = 1$ si la i-ésima persona es activa y $D_{1i} = 0$ en caso contrario; y
$D_{2i} = 1$ si la i-ésima persona es retirada y $D_{2i} = 0$ en caso contrario;
$C_i = 1$ si la i-ésima persona es hombre y $C_i = 0$ si es mujer.

Con base en esta definición se presentan 6 posibilidades para Y:

$$Y_i = \begin{cases} \beta_0 + \beta_1 + u_i & si\ D_{1i} = 1,\ D_{2i} = 0,\ C_i = 0 \\ \beta_0 + \beta_2 + u_i & si\ D_{1i} = 0,\ D_{2i} = 1,\ C_i = 0 \\ \beta_0 + u_i & si\ D_{1i} = 0,\ D_{2i} = 0,\ C_i = 0 \\ \beta_0 + \beta_1 + \beta_3 + u_i & si\ D_{1i} = 1,\ D_{2i} = 0,\ C_i = 1 \\ \beta_0 + \beta_2 + \beta_3 + u_i & si\ D_{1i} = 0,\ D_{2i} = 1,\ C_i = 1 \\ \beta_0 + \beta_3 + u_i & si\ D_{1i} = 0,\ D_{2i} = 0,\ C_i = 1 \end{cases}$$

Al estimar el MRLM por MCO se obtienen $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ y $\hat{\beta}_3$, de modo que los ingresos mensuales estimados se pueden presentar en una tabla como se muestra en la figura 4.1.

Figura 4.1
Ingreso mensual estimado

Situación laboral	**Género**	
	Mujer	**Hombre**
Activo	$\hat{\beta}_0 + \hat{\beta}_1$	$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_3$
Retirado	$\hat{\beta}_0 + \hat{\beta}_2$	$\hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_3$
Desempleado	$\hat{\beta}_0$	$\hat{\beta}_0 + \hat{\beta}_3$

A los MRLM cuyos regresores son exclusivamente variables cualitativas se les denomina **Modelos de Análisis de Varianza** o **Modelos ANOVA**, pues permiten analizar el comportamiento de la variable cuantitativa Y a través de las distintas categorías (o combinaciones de categorías) de las variables explicativas.

De la figura 4.1 se observa cómo en los Modelos ANOVA es importante poder analizar **transformaciones lineales de los parámetros**. Bajo el supuesto de *normalidad* en los errores es posible aplicar la Prueba F (TLG) para validar si, por ejemplo, el ingreso mensual de los hombres activos es el salario mínimo (H_0: $\beta_0 + \beta_1 + \beta_3 = w_{\min}$).

En los ejemplos previos se ha considerado que las variables dicótomas tienen un **efecto aditivo** en la variable explicativo; sin embargo, también es posible considerar que las variables dicótomas tengan un **efecto multiplicativo**.

Por ejemplo, considerando nuevamente el ingreso mensual de las personas (Y) explicado por su situación laboral (activo, retirado o desempleado) y su género (hombre o mujer), se puede formular el siguiente modelo econométrico alternativo:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 C_i + \boxed{\beta_4 D_{1i} C_i} + u_i \text{, } u_i \sim iid(0,\ \sigma^2),$$

donde: $D_{1i} = 1$ si la i-ésima persona es activa y $D_{1i} = 0$ en caso contrario; y
$D_{2i} = 1$ si la i-ésima persona es retirada y $D_{2i} = 0$ en caso contrario;
$C_i = 1$ si la i-ésima persona es hombre y $C_i = 0$ si es mujer.

En este modelo al término $\beta_4 D_{1i} C_i$ se le denomina **término de interacción**. El parámetro β_4 es un efecto multiplicativo adicional al efecto aditivo de β_1 de los activos y β_3 de los hombres, de modo que en este modelo la estimación del ingreso mensual de los hombres activos es $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_3 + \hat{\beta}_4$.

Ejercicio $\boxed{\text{E47}}$.

4.2. MRLM con Variables Cuantitativas y Dicótomas

A los MRLM cuyos regresores combinan variables cuantitativas y variables cualitativas se les denomina **Modelos de Análisis de Covarianza** o **Modelos ANCOVA**, pues permiten explicar el comportamiento de la variable cuantitativa Y mediante regresores cuantitativos pero considerando además las distintas categorías (o combinaciones de categorías) asociadas a los regresores cualitativos.

La incorporación de las variables dicótomas a los Modelos ANCOVA puede ser tanto aditiva como multiplicativa, y con frecuencia se consideran **términos de interacción entre variables cualitativas y cuantitativas**.

El Modelo ANCOVA más simple es el que incluye un regresor cuantitativo (X) y una variable dicótoma (D). Mediante este modelo se pueden estimar:

- Cambio de intercepto (ver Figura 4.2, rectas paralelas):

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + u_i = \begin{cases} \beta_0 + \beta_2 X_i + u_i & si \;\; D_i = 0 \\ (\beta_0 + \beta_1) + \beta_2 X_i + u_i & si \;\; D_i = 1 \end{cases} \text{; o}$$

- Cambio de pendiente (ver Figura 4.3, rectas concurrentes):

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i X_i + u_i = \begin{cases} \beta_0 + \beta_1 X_i + u_i & si \;\; D_i = 0 \\ \beta_0 + (\beta_1 + \beta_2) X_i + u_i & si \;\; D_i = 1 \end{cases}$$

Figura 4.2
Cambio de intercepto (rectas paralelas)

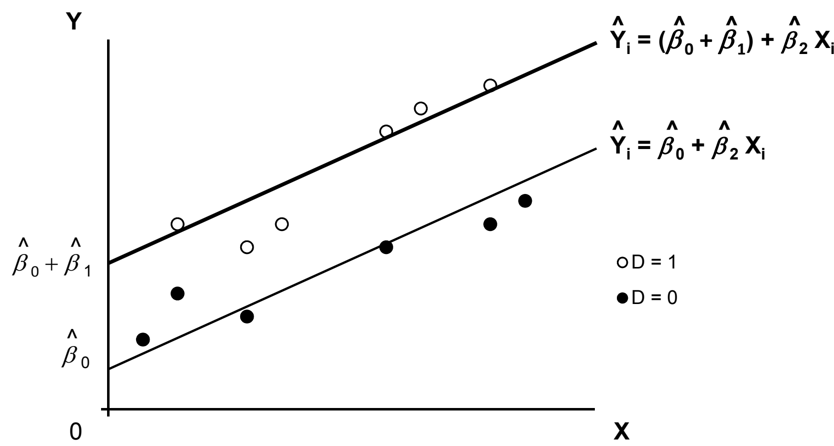
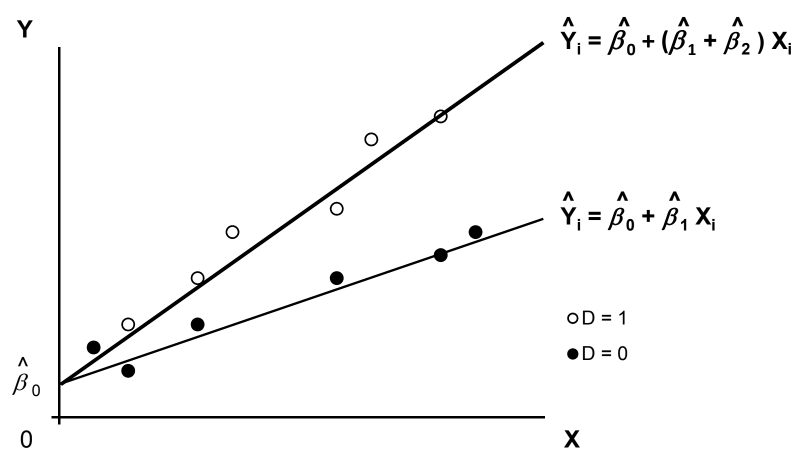


Figura 4.3
Cambio de ordenada (rectas concurrentes)



Si se desean modelar simultáneamente cambio de intercepto y cambio de pendiente el modelo econométrico es:

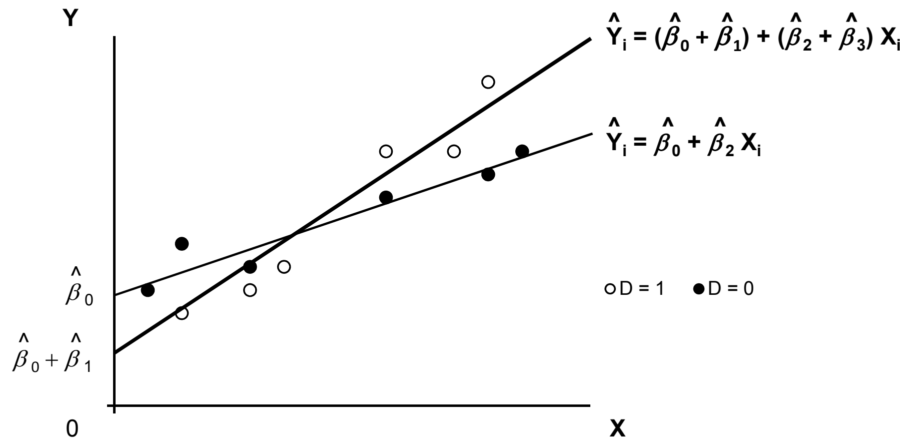
$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 D_i X_i + u_i \text{, } u_i \sim iid(0,\ \sigma^2),$$

a partir del cual se generan 2 posibles regresiones:

- Recta de comparación: $Y_i = \beta_0 + \beta_2 X_i + u_i$, si $D_i = 0$; y
- Recta disímbola: $Y_i = (\beta_0 + \beta_1) + (\beta_2 + \beta_3) X_i + u_i$ si $D_i = 1$.

La figura 4.4 muestra las gráficas de las estimaciones de la recta de comparación y de la recta disímbola.

Figura 4.4
Cambio de intercepto y de pendiente (rectas de comparación y disímbola)



Los Modelos ANCOVA se pueden generalizar para cualquier número de regresores cuantitativos y de variables dicótomas (aditivas o multiplicativas). Por ejemplo, si se consideran k regresores cuantitativos $X_1, X_2, \ldots, X_k$, y k variables dicótomas (aditivas y multiplicativas) $D_1, D_2, \ldots, D_k$, se puede formular el modelo econométrico:

$$Y_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j D_{ji} + \sum_{j=1}^{k} \beta_j X_{ji} + \sum_{j=1}^{k} \gamma_j D_{ji} X_{ji} + u_i \text{, } u_i \sim iid(0,\ \sigma^2),$$

en el cual aparecen $1 + k + k + k + 1 = 3k + 2$ parámetros y para poder estimarse por MCO se requieren $n > 3k + 2$ observaciones.

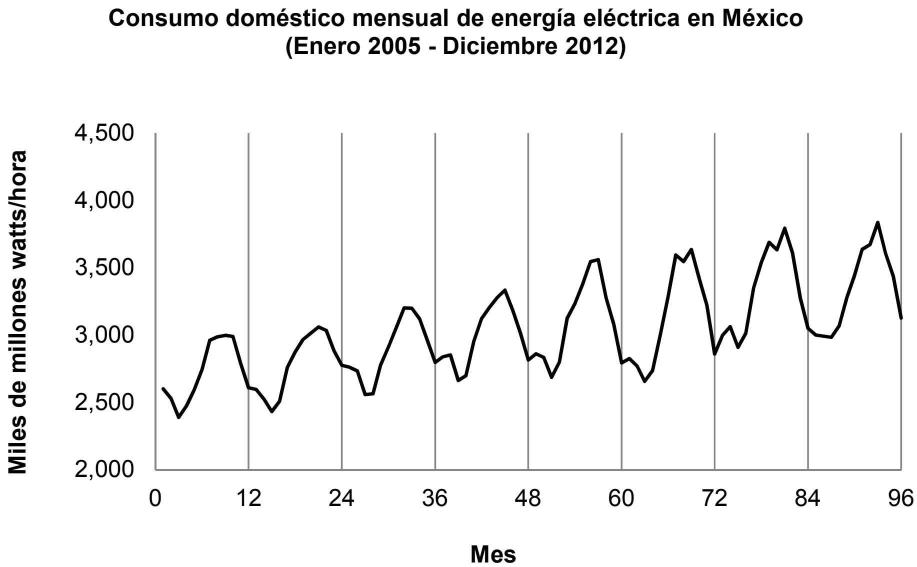
Ejercicio $\boxed{\text{E48}}$.

4.3. Análisis estacional

En el análisis de series de tiempo, donde la variable de interés (Y_t) se explica mediante la variable tiempo (t), es común encontrar cierto patrón que se repite a lo largo del tiempo sobre intervalos de igual longitud (típicamente un año calendario). A ese patrón se le denomina **estacionalidad** y se puede modelar mediante variables dicótomas.

Por ejemplo, en la figura 4.5 se muestra el consumo mensual doméstico de energía eléctrica en México (Y_t), de enero de 2005 a diciembre de 2012 (t = 1, 2,…, 96). Esta serie presenta estacionalidad pues año con año el patrón es el mismo: inicia disminuyendo alcanzando su mínimo en marzo, luego crece de manera sostenida hasta septiembre, y finalmente cierra en diciembre con una tendencia a la baja.

Figura 4.5



Fuente: www.inegi.gob.mx

En este caso particular, además de estacionalidad, existe una tendencia lineal creciente, de modo que se puede considerar el modelo econométrico:

$$Y_t = \beta_0 + \sum_{j=1}^{11} \beta_j M_{jt} + \beta_{12} t + u_i \text{ , } u_i \sim iid(0,\ \sigma^2),$$

donde: M_{jt} = 1 en el mes j, j = 1, 2,…, 11; y
M_{jt} = 0 en caso contrario.

Note que no se incluye una variable dicótoma para el mes 12 a fin de evitar la "trampa de la variable dicótoma" (problema de multicolinealidad).

Ejercicio $\boxed{\text{E49}}$.

Con la finalidad de analizar con mayor profundidad a las series de tiempo (metodología de Box-Jenkins, 1970) y poder hacer mejores pronósticos, se acostumbra eliminar la estacionalidad de la serie. A este proceso se le conoce como **desestacionalización** de la serie. Es común que las fuentes de información oficial de los países, como el INEGI o BANXICO en México, presenten series de tiempo macroeconómica desestacionalizadas.

A continuación se presenta un método básico para desestacionalizar series de tiempo, sin embargo, en la práctica existen muchos otros métodos más sofisticados y precisos para desestacionalizar series de tiempo como el ARIMA X-12 (desarrollado por el *United States Census Bureau*) o TRAMO (desarrollado por el Banco Central de España).

Método de Desestacionalización con Variables Dicótomas

1) Estimar la serie de tiempo por MCO considerando intercepto y utilizando como regresores sólo las variables dicótomas que modelan la estacionalidad. Los coeficientes de este modelo permiten calcular los *factores de estacionalidad*.
2) Construir la serie de tiempo desestacionalizada sumando los residuos del modelo anterior y la media muestral de la serie de tiempo original.

Los **factores de estacionalidad** son los aumentos o disminuciones periódicos (v.gr., mensuales, trimestrales) que se presentan a lo largo del tiempo y se calculan como el intercepto más el coeficiente de cada variable dicótoma menos la media muestral de la serie de tiempo original. La suma de los factores de estacionalidad siempre debe ser cero.

Al eliminar la estacionalidad de una serie de tiempo es más fácil identificar si existe algún tipo de **tendencia**.

Para hacer **pronósticos** de una serie de tiempo a partir de su respectiva serie desestacionalizada es necesario incorporar nuevamente la estacionalidad sumando sus respectivos *factores de estacionalidad*.

Ejercicio $\boxed{\text{E50}}$.

4.4. Cambio estructural y regresión lineal por segmentos.

4.4.1. Cambio estructural.

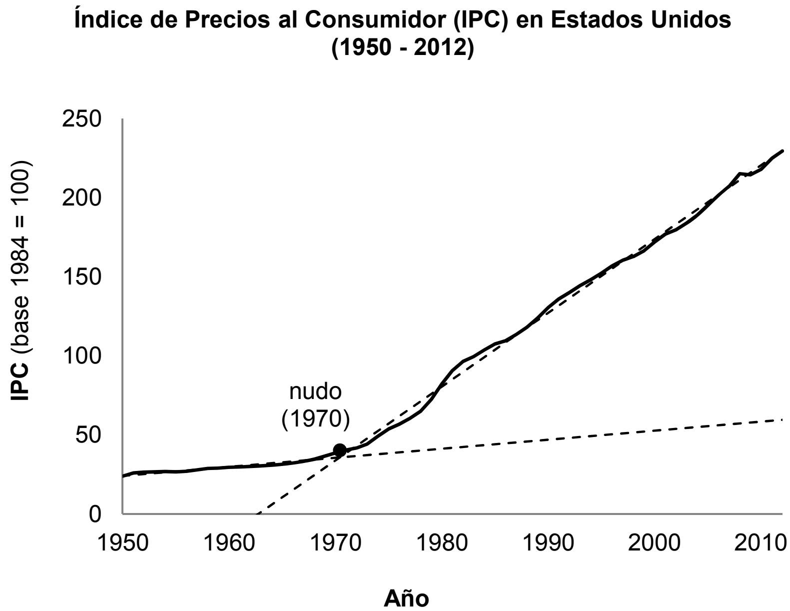
Otra de las aplicaciones de las variables dicótomas en el análisis de series de tiempo es el **Modelo de Cambio Estructural**, que permite modelar cambios en la tendencia de la serie en distintos intervalos de tiempo. La principal característica es que cada segmento de tendencia tiene sus propios parámetros pero hay *continuidad en el ajuste*.

Por ejemplo, en la figura 4.6 se muestra la serie de tiempo del índice de precios al consumidor (IPC_t) en Estados Unidos de 1950 a 2012, t = 1950, 1951,…, 2012, (base 1984 = 100). En esta gráfica es fácil identificar que el comportamiento de 1950 a 1970 es lineal y que de 1971 a 2012 el comportamiento también es lineal pero con intercepto y pendiente distintos. Al punto del tiempo en el que se presenta el cambio de tendencia de le denomina **nudo**, y en este caso corresponde al tiempo t = 1970. En este caso se puede considerar el modelo econométrico:

$$IPC_t = \beta_0 + \beta_1 t + \beta_2 (t - 1970) CE_t + u_t,\ u_t \sim iid(0,\ \sigma^2),$$

donde: $CE_t = 1$ si t = 1970, 1971,…, 2012; y
$CE_t = 0$ en caso contrario.

Figura 4.6



Definición del nudo. La definición del nudo no es única. Existen distintos criterios para determinarlo; uno de ellos consiste en elegir como nudo aquel valor para el cuál se maximizan simultáneamente la R^2 de ambas tendencias lineales.

Importante: A diferencia del Modelo ANCOVA de un regresor cuantitativo y una variable dicótoma con cambio de intercepto y de ordenada (4 parámetros), en el Modelo de Cambio Estructural sólo es necesario estimar 3 parámetros pues el valor del nudo es conocido de antemano, lo que garantiza la *continuidad en el ajuste*.

4.4.2. Regresión Lineal por Segmentos

Al igual que el Modelo de Cambio Estructural, la **Regresión Lineal por Segmentos** consiste en un MRLS que experimenta un cambio de intercepto y/o de pendiente a partir de cierto nivel del regresor (no necesariamente el tiempo) denominado **umbral**.

En el contexto de la Regresión Lineal por Segmentos al **umbral** se le denota por X^* y el modelo econométrico puede formularse de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 \left(X_i - X^*\right) D_i + u_i \text{, } u_t \sim iid(0,\ \sigma^2),$$

donde: $D_i = 1$ si $X_i \geq X^*$; y
$D_i = 0$ si $X_i < X^*$.

Ejercicio $\boxed{\text{E51}}$.

5. Análisis de Supuestos del MRLM

La validez de los resultados que se pueden obtener mediante la aplicación del MRLM recae en gran medida en que se cumplan los supuestos planteados para su estimación, inferencia estadística y predicción.

Para definir al MRLM $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ se hicieron **nueve supuestos**: (i) linealidad en los parámetros; (ii) regresores no estocásticos e independientes de los errores; (iii) errores con media cero; (iv) errores con varianza constante (homocedasticidad); (v) errores independientes y no correlacionados; (vi) número de observaciones mayor al número de parámetros; (vii) valores diferenciados en los regresores y sin valores atípicos; (viii) no colinealidad entre los regresores; y (ix) no debe haber sesgo de especificación.

Estos supuestos se deben cumplir para que el MRLM se pueda estimar por MCO, que $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$ sea MELI y poder hacer predicción puntual.

Adicionalmente se formuló un **décimo supuesto**: (x) normalidad en los errores. Este supuesto se debe cumplir para poder afirmar que el EMV de $\boldsymbol{\beta}$ coincide con su respectivo estimador MCO y que en consecuencia $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$, además de ser MELI, posee las propiedades de los EMV: invarianza, consistencia y eficiencia asintótica.

El supuesto de normalidad es indispensable para realizar estimación por intervalo (intervalos de confianza para los parámetros), pruebas de hipótesis (particularmente las Pruebas *T* y *F*) y predicción por intervalo (intervalos de confianza para las medias e intervalos de predicción para los valores).

Algunos de estos supuestos se validan por la **especificación del modelo**: (i) y (ix). Otros se validan mediante el análisis exploratorio de los datos que integran la **muestra disponible** para la estimación: (ii), (vi) y (vii).

El supuesto (iii) de errores con **media cero** ($\mu_u = 0$) es fácil de validar mediante la gráfica de los residuos vs. la variable de interés o de los residuos vs. algún regresor. Si el MRLM se estima con intercepto se sabe que $\sum \hat{u}_i = 0$, lo cual garantiza el cumplimiento de este supuesto pues $\hat{\mu}_u = \overline{\hat{u}_i} = 0$.

Los supuestos (iv), (v) y (viii) son más complejos de validar, y su incumplimiento conduce a los siguientes **problemas**:

- *Multicolinealidad*, que surge por incumplimiento del supuesto (viii);
- *Heterocedasticidad*, que surge por incumplimiento del supuesto (iv); y
- *Autocorrelación*, que surge por incumplimiento del supuesto (v).

Para identificar, analizar y corregir algunos de estos problemas es útil el supuesto de normalidad, así que iniciaremos este tema presentando algunas pruebas estadísticas que permiten validar el supuesto (x) de **normalidad**.

5.1. Normalidad

La **normalidad de los errores del MRLM** se puede validar analizando sus respectivos residuos mediante métodos gráficos o mediante pruebas estadísticas más formales como se resume en la figura 5.1

Figura 5.1

Métodos gráficos	**Pruebas de bondad de ajuste**
• Histogramas • Gráfica de Probabilidad Normal • Gráfica Cuantil-Cuantil (*Q-Q Plot*).	• Jarque-Bera (1980) • Lilliefors (1967) • Cramér-Von Mises (1928-1930) • Watson (1961) • Anderson-Darling (1952)

Si un conjunto de residuos tiene una distribución normal, entonces su histograma debe ser simétrico y acampanado, y en sus respectivas gráficas de probabilidad normal y gráfica cuantil-cuantil los cuantiles muestrales deben ubicarse sobre la recta (o muy cercanos a ella).

En las pruebas de bondad de ajuste se plantea que los errores $u_1, u_2, \ldots, u_n$ tienen una distribución Normal caracterizada por $F_0 \sim N\left(0, \sigma_u^2\right)$, es decir,

$$H_0\text{: } F_u = F_0 \quad \text{vs.} \quad H_1\text{: } F_u \neq F_0 .$$

La Prueba Jarque-Bera utiliza un estadístico de prueba basado en los coeficientes de asimetría y curtosis muestrales:

$$CA = \frac{m_3}{s^3}, \quad CC = \frac{m_4}{s^4} - 3, \quad \text{donde} \quad s = \sqrt{m_2} \quad \text{y} \quad m_k = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^k .$$

Desafortunadamente esta prueba sólo es válida para muestras muy grandes ($n > 2{,}000$).

El resto de estas pruebas son modificaciones más potentes de la **Prueba Kolmogorov-Smirnov** (1933-1948), válidas para muestras de cualquier tamaño, que comparan las distancias (en valor absoluto) entre la función de distribución acumulada teórica $F_0\left(\hat{u}_i\right)$ y la función de distribución acumulada empírica $F_n\left(\hat{u}_i\right) = \frac{1}{n} \sum_{\hat{u}_j \leq \hat{u}_i} I_{(-\infty, \hat{u}_i]}\left(\hat{u}_j\right)$. La figura 5.2 resume las regiones de rechazo y estadísticas de prueba en cada caso:

Figura 5.2

Prueba	Región de rechazo	Estadístico de prueba
Jarque-Bera	$RR_\alpha = \{JB > \chi^2_{2,\alpha}\}$	$JB = \frac{n}{6}\left[(CA)^2 + \frac{(CC)^2}{4}\right] \xrightarrow{d} \chi^2(2)$
Lilliefors	$RR_\alpha = \{D > l_\alpha\}$, donde l_α se obtiene numéricamente mediante simulación Monte Carlo	$D = \sup_i \left\{ \left\| F_n(\hat{u}_{(i)}) - F_0(\hat{u}_{(i)}) \right\| \right\}$
Cramér-Von Mises	$RR_\alpha = \{\omega^2 > c_{n,\alpha}\}$, donde $c_{n,\alpha}$ aparece en tablas para esta prueba.	$\omega^2 = \int_{-\infty}^{\infty} [F_n(u) - F_0(u)]^2 f_0(u)du$
Watson	$RR_\alpha = \{U^2 > u^2_{n,\alpha}\}$, donde $u^2_{n,\alpha}$ aparece en tablas para esta prueba.	$U^2 = T - n\left(\frac{1}{n}\sum_{i=1}^{n} F_0(\hat{u}_i) - \frac{1}{2}\right)^2$, donde $T = \frac{1}{12n} + \sum_{i=1}^{n}\left[\frac{2i-1}{2n} - F_0(u_i)\right]$
Anderson Darling	$RR_\alpha = \{A^2 > a^2_\alpha\}$, donde a^2_α aparece en tablas para esta prueba.	$A^2 = -n - S$ donde $S = \sum_{i=1}^{n} \frac{2i-1}{n}\left[\ln(F_0(\hat{u}_{(i)})) + \ln(1 - F_0(\hat{u}_{(n+1-i)}))\right]$

Como en todos estos casos bajo H_0 se plantea la distribución teórica Normal, **el supuesto de normalidad se valida si el valor-p es grande**.

La mayor parte de los paquetes Estadísticos y Econométricos presentan algunas de estas pruebas de bondad de ajuste para analizar la normalidad de los errores. En particular EViews presenta todas estas pruebas.

Ejercicio $\boxed{\text{E52}}$.

Si para un MRLM no se valida la normalidad de sus residuos, es posible transformar la variable dependiente Y_i en busca de que su respectivo MRLM transformado tenga residuos normales. Una transformación utilizada con frecuencia para este fin es la transformación potencia, que depende de $\lambda \in \mathbf{R}$, que es una parámetro adicional.

Def. Transformación Potencia.

$$Y_i(\lambda) = \begin{cases} Y_i^\lambda & si \;\; \lambda \neq 0 \\ \ln(Y_i) & si \;\; \lambda = 0 \end{cases}$$

Algunos paquetes de Estadísticos y Econométricos estiman simultáneamente λ y los parámetros del MRLM por máxima verosimilitud. Una forma simple de determinar el valor de λ que conduzca a la normalidad de los errores, es estimar por MCO los modelos $Y_i(\lambda) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i(\lambda)$ para algunos valores de λ (por ejemplo, 0, ±0.5, ±1, ±1.5 y ±2) y elegir aquél con menor SCR.

5.2. Multicolinealidad

5.2.1. Definición y Causas de la Multicolinealidad

Def. Multicolinealidad.

Se dice que el MRLM $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$ presenta el problema de multicolinealidad si algún regresor (X_j) se puede expresar como combinación lineal de algunos otros regresores, es decir, si existen constantes $c_1, \ldots, c_k$, tales que $X_j = \sum_{h \neq j} c_h X_h$ para alguna $c_h \neq 0$.

Esta definición establece que hay una *relación exacta* entre los regresores, por lo que también se le denomina **multicolinealidad perfecta**. Desafortunadamente, incluso si existe una *relación aproximada* entre los regresores el problema de multicolinealidad persiste y se denomina **multicolinealidad alta pero imperfecta**.

Consecuencias de la multicolinealidad:

- Si la multicolinealidad es perfecta, el MRLM $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ no se puede estimar por MCO, pues la matriz de diseño **X** no es de rango completo (Rango(**X**) < k + 1), $\mathbf{X}'\mathbf{X}$ es singular y por lo tanto $(\mathbf{X}'\mathbf{X})^{-1}$ no existe.
- Si la multicolinealidad es alta pero imperfecta, el MRLM $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ sí se puede estimar por MCO pero los errores estándar de $\hat{\beta}_j$, $j = 0, 1, \ldots, k$, son muy grandes, lo que implica que las estimaciones puntuales no son muy precisas (intervalos de confianza amplios y pruebas de significancia poco significativas). En estas situaciones hay que tener en cuenta que el simple hecho de tener relaciones aproximadas entre los regresores aumenta el valor de la R^2.

Causas de la multicolinealidad:

- Es común que al considerar series de tiempo que comparten una tendencia se presente el problema de multicolinealidad. Por ejemplo, si se quiere explicar el consumo mediante el ingreso y la riqueza, existe altas correlaciones entre el ingreso y la riqueza a lo largo del tiempo que conducen a una multicolinealidad alta pero imperfecta.
- Al incorporar variables dicótomas en el MRLM se puede incurrir en la "trampa de la variable dicótoma", que conduce de inmediato a una multicolinealidad perfecta.
- En general, la multicolinealidad es un problema de tipo muestral, la variable de interés se observa para intervalos limitados de los regresores. Ante una multicolinealidad imperfecta lo relevante es saber qué tan fuerte es dicha colinealidad para poder corregirla.

5.2.2. Identificación de la Multicolinealidad

No existe un método único que permita identificar la multicolinealidad, pero existen algunas reglas prácticas:

- $\mathbf{X'X}$ **singular**. En estas situaciones los paquetes estadísticos no pueden calcular $(\mathbf{X'X})^{-1}$ al tratar de estimar el MRLM por MCO e indican al usuario la existencia de multicolinealidad perfecta.
- $\boldsymbol{R^2}$ **alta pero Pruebas *T* poco significativas**. Esto es un indicativo de la posible existencia de multicolinealidad imperfecta.
- **Altas correlaciones entre parejas de regresores**. La matriz de correlaciones y el diagrama de dispersión múltiple son útiles para realizar este análisis. Altas correlaciones pueden ocasionar multicolinealidad alta pero imperfecta. Correlaciones perfectas (±1) provocan multicolinealidad perfecta.

La multicolinealidad perfecta siempre de debe corregir. La multicolinealidad imperfecta es un problema con el que se puede vivir. Una regla práctica para determinar si la multicolinealidad imperfecta es "grave" es la *Regla de Klein*.

Regla de Klein

> La multicolinealidad imperfecta es un problema "grave" únicamente si la R^2 de una regresión auxiliar es mayor a la R^2 global.

En esta regla la $\boldsymbol{R^2}$ **global** es la R^2 de la regresión en donde se explica a Y mediante los regresores X_1, X_2,… X_k; y $\boldsymbol{R^2}$ **de una regresión auxiliar** corresponde a la R^2 de la regresión en donde se explica a uno de los regresores (X_j) mediante el resto de los regresores ($X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_k$) y sin incluir a la variable de interés.

5.2.3. Corrección de la Multicolinealidad

Si un MRLM presenta multicolinalidad perfecta es indispensable corregirla, sin embargo, si hay multicolinealidad imperfecta es posible vivir con ella y simplemente saber que las estimaciones que se puedan hacer mediante ese MRLM son poco precisas.

A continuación se expondrán algunas técnicas que permiten corregir la multicolinealidad:

- Eliminación de variables.
- Incorporación de información *a priori*.
- Transformación de variables.
- Métodos multivariados.

Eliminación de variables. La solución más simple para resolver la multicolinealidad es eliminar del modelo las variables que la provocan. Si $X_j = \sum_{h \neq j} c_h X_h$ quiere decir que se posible eliminar a X_j pues toda la información relevante está en los regresores X_h's; o que es posible eliminar a las X_h's y dejar sólo a X_j, ya que ésta resume la información de dichos regresores.

Por ejemplo, si se incurre en la "trampa de la variable dicótoma" se puede resolver mediante esta técnica eliminando la columna de unos de la matriz de diseño (regresión sin intercepto) o bien, eliminando la variable dicótoma de alguna de las categorías del regresor cualitativo.

Esta forma de proceder parece trivial pero conceptualmente es relevante pues puede conducir a un sesgo de especificación. Si la Teoría Económica establece cuáles son los regresores que hay que utilizar en un modelo y se presenta el problema de multicolinealidad lo que se tiene es un problema de muestreo (con los datos disponibles) y no deben eliminarse los regresores del modelo pues se incurre en sesgo de especificación. Es necesario aplicar alguna otra técnica de corrección.

Incorporación de información *a priori*. Si se tienen identificados los regresores que ocasionan la multicolinealidad (v.gr., aquellos que están altamente correlacionados) es posible incorporar información *a priori* proveniente de la teoría del área específica de estudio que permita establecer una relación determinista entre dichos regresores y posteriormente incluir dicha relación como un nuevo regresor.

Transformación de variables. Al considerar series de tiempo es común que algunos regresores se muevan en la misma dirección a lo largo del tiempo provocando multicolinealidad alta pero imperfecta. Una forma de eliminar la multicolinealidad es transformar dicha variables. Las transformaciones más comunes son las de *primeras diferencias* y de *razón*.

Por ejemplo, si el modelo con alta multicolinealidad es $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$, entonces sus transformaciones son:

- *Primeras diferencias*: $Y_t - Y_{t-1} = \beta_1 \left(X_{1,t} - X_{1,t-1} \right) + \beta_2 \left(X_{2,t} - X_{2,t-1} \right) + v_t$
- *Razón*: $\dfrac{Y_t}{X_{2t}} = \beta_0 \left(\dfrac{1}{X_{2t}} \right) + \beta_1 \left(\dfrac{X_{1t}}{X_{2t}} \right) + \beta_2 + w_t$, donde $w_t = \dfrac{u_t}{X_{2t}}$

Hay que tener cuidado con la *transformación de razón* pues al dividir los errores del modelo original entre uno de los regresores es posible incurrir en heterocedasticidad.

Métodos multivariados. En Estadística Multivariada se estudian a detalle las posibles relaciones entre los regresores y pueden utilizarse algunas de sus técnicas como el *Análisis Factorial* o el *Análisis de Componentes Principales* para resolver el problema de multicolinealidad.

Ejercicio $\boxed{\text{E53}}$.

5.3. Heterocedasticidad

5.3.1. Definición y Causas de la Heterocedasticidad

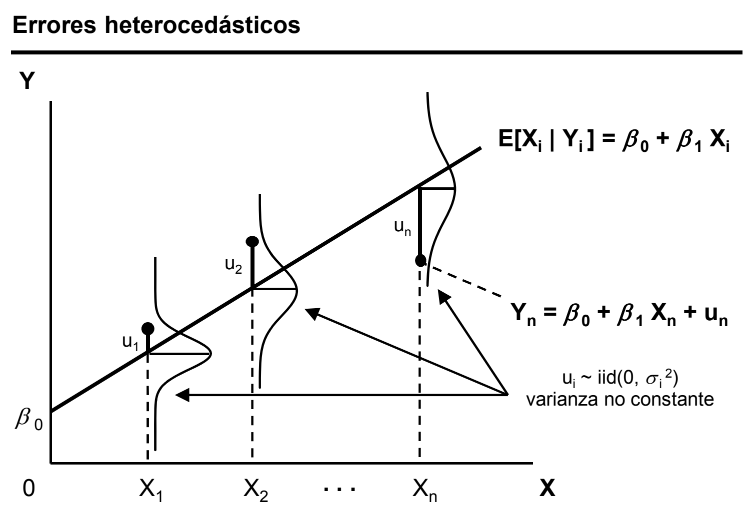
El supuesto (iv) del MRLM establece que los errores deben tener varianza constante (homocedasticidad), es decir, $Var[u_i] = \sigma^2$ para toda i = 1, 2,…, n. El problema que surge por la violación de este supuesto se denomina heterocedasticidad.

Def. Heterocedasticidad.

Se dice que el MRLM $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$ presenta el problema de heterocedasticidad si $Var[u_i] = \sigma_i^2$, i = 1, 2,…, n, donde $\sigma_i^2 \neq \sigma_j^2$ para alguna $i \neq j$.

Geométricamente los errores heterocedásticos se pueden representar para el MRLS como se muestra en la figura 5.3.

Figura 5.3



Para el MRLM con heterocedasticidad (y sin autocorrelación) la matriz de varianzas y covarianzas es una matriz diagonal de la forma:

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix},$$

de modo que $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, $\mathbf{u} \sim (\mathbf{0}, \mathbf{V})$.

Consecuencias de la heterocedasticidad:

- Los estimadores MCO son lineales, insesgados y consistentes, pero dejan de ser de mínima varianza (no son MELI).
- Si la matriz de varianzas-covarianzas de los estimadores MCO se estima como $\hat{\Sigma}_{\hat{\beta}} = \hat{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$ donde $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$, $\hat{\sigma}_i^2$ (el i-ésimo elemento de la diagonal principal) sobrestima el verdadero valor de σ_i^2, invalidando las Pruebas T y F, a pesar de que el supuesto de normalidad se valide.

Causas de la heterocedasticidad:

- *Modelos de aprendizaje*. La variable dependiente puede tener menos variabilidad conforme alguno de los regresores aumentan (o disminuyen) su valor. Por ejemplo, el número de errores de un capturista va presentando menor variabilidad (menor error) conforme aumentan sus horas dedicadas a la captura de información.
- Presencia de observaciones atípicas (*outliers*).
- Asimetría en la distribución de uno o varios de los regresores. Por ejemplo, la distribución del ingreso y riqueza típicamente presentan un sesgo derecho.

5.3.2. Identificación de la Heterocedasticidad

No existe un método único que permita identificar la heterocedasticidad, pero existen algunos métodos (informales y formales) para su detección.

MÉTODOS INFORMALES

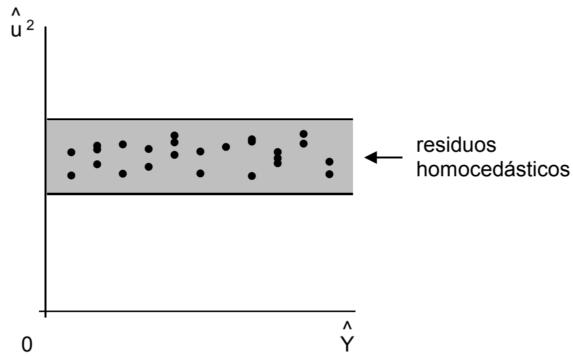
Los métodos informales más simples son:

- Naturaleza del problema.
- Método gráfico.

Naturaleza del problema. El investigador sabe de antemano que la dispersión de la variable explicada es mayor o menor ante distintos niveles de sus regresores. Por ejemplo, modelos de aprendizaje.

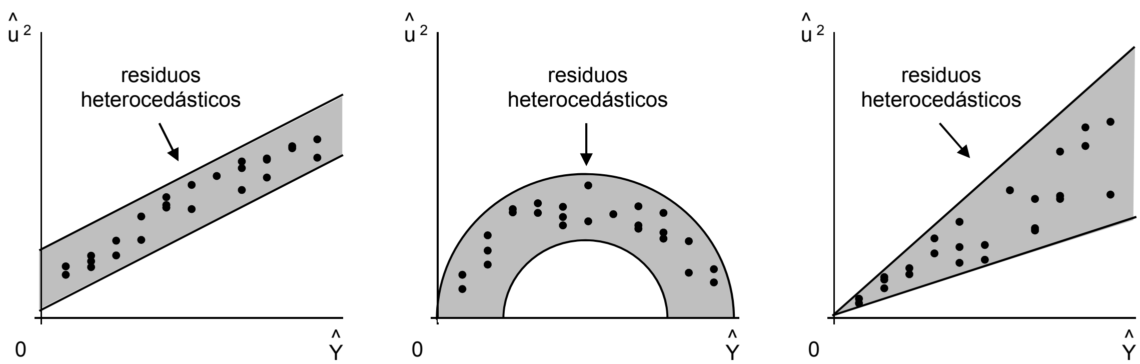
Método gráfico. Consiste en estimar el MRLM por MCO y analizar el comportamiento de los residuos. El estimador MCO de σ^2 es $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)} = \dfrac{1}{n-(k+1)}\sum_{i=1}^{n}\hat{u}_i^2$, de modo que bajo el supuesto de homocedasticidad, el comportamiento de $\hat{u}_1^2$, $\hat{u}_2^2, \ldots,$ $\hat{u}_n^2$, debe ser bastante uniforme. Para verificarlo la gráfica de los residuos al cuadrado vs. los valores ajustados deben revelar un patrón como el de la figura 5.4.

Figura 5.4



Si el supuesto de homocedasticidad no se cumple, entonces la gráfica de los residuos al cuadrado vs. los valores ajustados deberán mostrar algún tipo de patrón no uniforme, como los que se muestran en la figura 5.5, revelando la heterocedasticidad.

Figura 5.5



MÉTODOS FORMALES

Se expondrán las siguientes pruebas estadísticas que son fáciles de implementar en EViews para el diagnóstico de los residuos:

- Breusch-Pagan-Godfrey.
- Glejser.
- White.
- Igualdad de varianzas por grupos.

Otro método formal de fácil implementación es la Prueba Goldfeld-Quandt, sin embargo, no se expondrá aquí pues no está programada en EViews.

Prueba de Breusch-Pagan-Godfrey

Si $Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i$, $Var[u_i] = \sigma_i^2 = \alpha_0 + \alpha_1 Z_{1i} + \cdots + \alpha_m Z_{mi} + v_i$ y Z_i son todas o algunas de las X_j, $m \leq k$, entonces el MRLM tiene errores homocedásticos sí y sólo si $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$. Entonces para la prueba H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$ vs H_1: alguna $\alpha_i \neq 0$, la región de rechazo es $RR_\alpha = \{j > \chi^2_{m,\alpha}\}$ y el estadístico de prueba es $J = nR_v^2 \xrightarrow{d} \chi^2(m)$, donde R_v^2 es el coeficiente de determinación de la regresión auxiliar $\hat{u}_i^2 = \alpha_0 + \alpha_1 Z_{1i} + \cdots + \alpha_m Z_{mi} + v_i$.

El **rechazo de la hipótesis nula** en la Prueba de Breusch-Pagan-Godfrey confirma la **heterocedasticidad**. La distribución asintótica Ji-Cuadrada de esta prueba se fundamenta en el Teorema Central del Límite. Para que la distribución de $J = nR_v^2$ sea Ji-Cuadrada en muestras chicas, es necesario validar primero el supuesto de normalidad en los errores u_i.

Una variante de esta prueba es considerar la región de rechazo $RR_\alpha = \{scre > \chi^2_{m,\alpha}\}$, donde el estadístico de prueba es $SCRE = \frac{1}{2}\sum_{i=1}^{n} \hat{v}_{*i}^2 \xrightarrow{d} \chi^2(m)$ y los residuos se toman de la regresión auxiliar $\frac{\hat{u}_i^2}{\hat{\sigma}_*^2} = \alpha_0 + \alpha_1 Z_{1i} + \cdots + \alpha_m Z_{mi} + v_{*i}$, $\hat{\sigma}_*^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n} = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2$.

Prueba de Glejser

Si $Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i$, $\sigma[u_i] = \sigma_i = \alpha_0 + \alpha_1 Z_{1i} + \cdots + \alpha_m Z_{mi} + v_i$ y Z_i son todas o algunas de las X_j, $m \leq k$, entonces el MRLM tiene errores homocedásticos sí y sólo si $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$. Entonces para la prueba H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$ vs H_1: alguna $\alpha_i \neq 0$, la región de rechazo es $RR_\alpha = \{f_v > f_{m,n-(m+1),\alpha}\}$ y el estadístico de prueba es $F_v = \frac{[n-(m+1)]SCE_v}{mSCR_v} \sim F(m, n-(m+1))$, de la Prueba F (ANOVA) de la regresión auxiliar $|\hat{u}_i| = \alpha_0 + \alpha_1 Z_{1i} + \cdots + \alpha_m Z_{mi} + v_i$, $v_i \sim N(0, \sigma_v^2)$.

El **rechazo de la hipótesis nula** en la Prueba de Glejser confirma la existencia de **heterocedasticidad** en los errores.

Una variante de esta prueba es la **Prueba de Park**, que supone que las varianzas se pueden expresar mediante el modelo no lineal $Var[u_i] = \sigma_i^2 = \sigma^2 Z_{1i}^{\alpha_1} \cdots Z_{mi}^{\alpha_m} e^{w_i}$, de modo que $\ln(\sigma_i^2) = \ln(\sigma^2) + \alpha_1 \ln(Z_{1i}) + \cdots + \alpha_m \ln(Z_{mi}) + w_i$. En este caso la región de rechazo es $RR_\alpha = \{f_w > f_{m,n-(m+1),\alpha}\}$, donde $F_w = \dfrac{[n-(m+1)]SCE_w}{mSCR_w} \sim F(m, n-(m+1))$ es el estadístico de prueba de la Prueba F (ANOVA) de la regresión auxiliar $\ln(\hat{u}_i^2) = \alpha_0 + \alpha_1 Z_{1i} + \cdots + \alpha_m Z_{mi} + w_i$, $\alpha_0 = \ln(\sigma^2)$ y $w_i \sim N(0,\ \sigma_w^2)$.

Prueba de White

Si $Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i$ y la varianza de los errores se determinan mediante la forma cuadrática $Var[u_i] = \sigma_i^2 = \alpha_0 + \sum_{j=1}^{k} \alpha_j X_{ji} + \sum_{j=k+1}^{2k} \alpha_j X_{ji}^2 + \sum_{\{j:l<h\}} \alpha_j X_l X_h + v_i$, los errores son homocedásticos sí y sólo si $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$, donde $m = 2k + \binom{k}{2}$.
Entonces para la prueba H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$ vs H_1: alguna $\alpha_i \neq 0$, la región de rechazo es $RR_\alpha = \{J > \chi^2_{m,\alpha}\}$ y el estadístico de prueba es $J = nR_v^2 \sim \chi^2(m)$, donde R_v^2 es el coeficiente de determinación de la regresión auxiliar $\hat{u}_i^2 = \alpha_0 + \sum_{j=1}^{k} \alpha_j X_{ji} + \sum_{j=k+1}^{2k} \alpha_j X_{ji}^2 + \sum_{\{j:l<h\}} \alpha_j X_l X_h + v_i$

El **rechazo de la hipótesis nula** en la Prueba de White confirma la existencia de **heterocedasticidad** en los errores. A diferencia de las Pruebas de Breusch-Pagan-Godfrey y de Glejser que requieren forzosamente del supuesto de normalidad, **la prueba de White no requiere del supuesto de normalidad**. La regresión auxiliar de esta prueba involucra muchos parámetros y su estimación requiere muchos datos. Esta dificultad se puede superar eliminando los términos lineales y las interacciones.

Ejercicio $\boxed{\text{E54}}$.

Pruebas de igualdad de varianzas por grupos. Las Pruebas F, Bartlett, Levene y Brown-Forsythe, comparten la siguiente lógica: $Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i$, donde $Var[u_i] = \sigma_i^2 = g(X_{ji})$, es decir, la varianza del error está en función de alguno de los regresores (X_j) de modo que σ_i^2 toma distintos valores para distintos niveles X_{ji}. Si se eligen m grupos, definidos por m intervalos mutuamente excluyentes y exhaustivos de los posibles valores de X_j, entonces se busca probar H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2$ vs. H_1: $\sigma_i^2 \neq \sigma_j^2$ para alguna $i \neq j$. **El rechazo de H_0 confirma la existencia de heterocedasticidad**.

Si los n residuos $\hat{u}_1$, $\hat{u}_2, \ldots,$ $\hat{u}_n$ se dividen en m grupos de tamaños n_1, $n_2, \ldots, n_m$, tales que $\sum_{h=1}^{m} n_h = n$, para cada grupo es posible calcular varianza muestral $S_{\hat{u}h}^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\left(\hat{u}_i - \bar{\hat{u}}_i\right)^2$, desviación estándar muestral $S_{\hat{u}h} = \sqrt{S_{\hat{u}h}^2}$, desviación media respecto a la media $dm_h\left(\bar{\hat{u}}\right) = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\left|\hat{u}_i - \bar{\hat{u}}_i\right|$ y desviación media respecto a la mediana $dm_h\left(\tilde{\hat{u}}\right) = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\left|\hat{u}_i - \tilde{\hat{u}}_i\right|$. En la tabla de la figura 5.6 se muestran las regiones de rechazo y las distribuciones de muestreo de los estadísticos de prueba (no se presentan los estadísticos *per se*, pues son fórmulas muy extensas y los paquetes estadísticos las calculan automáticamente).

Figura 5.6

Prueba y supuestos	**Región de Rechazo**	**Estadístico de Prueba**
F (Normalidad, m = 2)	$RR_\alpha = \left\{ f > f_{n_1-1, n_2-1, \alpha} \right\}$	$F = \frac{S_1^2}{S_2^2} \sim F\left(n_1 - 1, n_2 - 1\right)$
Bartlett (Normalidad)	$RR_\alpha = \left\{ b > \chi_{m-1,\alpha}^2 \right\}$	$B = g\left(S_{\hat{u}1}, S_{\hat{u}2}, \ldots, S_{\hat{u}m}\right) \sim \chi^2\left(m-1\right)$
Levene	$RR_\alpha = \left\{ l > f_{m-1, n-m, \alpha} \right\}$	$L = g\left(dm_1\left(\bar{\hat{u}}\right), dm_2\left(\bar{\hat{u}}\right), \ldots, dm_m\left(\bar{\hat{u}}\right)\right) \sim F\left(m-1, n-m\right)$
Brown-Forsythe	$RR_\alpha = \left\{ bf > f_{m-1, n-m, \alpha} \right\}$	$BF = g\left(dm_1\left(\tilde{\hat{u}}\right), dm_2\left(\tilde{\hat{u}}\right), \ldots, dm_m\left(\tilde{\hat{u}}\right)\right) \sim F\left(m-1, n-m\right)$

Ejercicio $\boxed{\text{E55}}$.

5.3.3. Corrección de la Heterocedasticidad

Si un MRLM presenta heterocedasticidad se puede corregir con distintas técnicas, dependiendo si las varianzas σ_i^2, son conocidas o no.

CORRECCIÓN DE LA HETEROCEDASTICIDAD CON VARIANZAS CONOCIDAS

Si las varianzas σ_i^2, i = 1, 2,…, n, son conocidas, la estimación del Modelo de Regresión Lineal Múltiple Heterocedástico (MRLMH) se debe hacer mediante *Mínimos Cuadrados Generalizados* (MCG).

Teorema. Estimadores MCG del MRLMH

Para el MRLMH $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde $\mathbf{u} \sim (\mathbf{0}, \mathbf{V})$ y $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ es no singular:

i) $\hat{\boldsymbol{\beta}}^* = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)$ es el estimador MCG de $\boldsymbol{\beta}$.

ii) $Var\left[\hat{\boldsymbol{\beta}}^*\right] = \boldsymbol{\Sigma}_{\hat{\beta}^*} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$.

iii) $\hat{\boldsymbol{\beta}}^*$ es MELI

La demostración de este Teorema es muy similar a la demostración de MCO, basta minimizar la suma de cuadrados de residuos $SCR = g\left(\hat{\boldsymbol{\beta}}^*\right) = \hat{\mathbf{u}}'\hat{\mathbf{u}}$ y calcular esperanza y varianza de $\hat{\boldsymbol{\beta}}^* = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)$.

Para el MRLSH $Y_i = \beta_0 + \beta_1 X_i + u_i$, $u_i \sim \left(0, \sigma_i^2\right)$, es posible dividir ambos lados del modelo entre su desviación estándar para obtener $\frac{Y_i}{\sigma_i} = \beta_0\left(\frac{1}{\sigma_i}\right) + \beta_1\left(\frac{X_i}{\sigma_i}\right) + \frac{u_i}{\sigma_i}$, donde:

$$E\left[\frac{u_i}{\sigma_i}\right] = \frac{1}{\sigma_i}E[u_i] = 0 \quad \text{y} \quad Var\left[\frac{u_i}{\sigma_i}\right] = \frac{1}{\sigma_i^2}Var[u_i] = \frac{1}{\sigma_i^2}\left(\sigma_i^2\right) = 1,$$

es decir, si $Y_i^* = \frac{Y_i}{\sigma_i}$, $X_{0i}^* = \frac{1}{\sigma_i}$, $X_i^* = \frac{X_i}{\sigma_i}$ y $u_i^* = \frac{u_i}{\sigma_i}$ entonces $Y_i^* = \beta_0^* X_{0i}^* + \beta_1^* X_i^* + u_i^*$ es MRLS con $u_i^* \sim (0,1)$, y se puede estimar por MCO por tener errores homocedásticos.

Al calcular la suma de cuadrados de residuos para este modelo se obtiene:

$$SCR = \sum_{i=1}^{n}\left(Y_i^* - \hat{\beta}_0^* X_{0i}^* - \hat{\beta}_1^* X_i^*\right)^2 = \sum_{i=1}^{n}\left[\left(\frac{1}{\sigma_i}\right)\left(Y_i - \hat{\beta}_0^* - \hat{\beta}_1^* X_i\right)\right]^2 = \sum_{i=1}^{n} w_i\left(Y_i^* - \hat{\beta}_0^* - \hat{\beta}_1^* X_i\right)^2,$$

es decir, una suma de cuadrados de residuos ponderada por $w_i = \frac{1}{\sigma_i^2}$.

Debido a lo anterior, a los estimadores MCG con errores heterocedásticos también se les denominan estimadores *Mínimos Cuadrados Ponderados* (MCP).

Teorema. Estimadores MCP del MRLSH

Para el MRLSH $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde $u_i \sim \left(0, \sigma_i^2\right)$ y $w_i = \dfrac{1}{\sigma_i^2}$:

i) $\hat{\beta}_1^* = \dfrac{\sum w_i \sum w_i X_i Y_i - \sum w_i X_i \sum w_i Y_i}{\sum w_i \sum w_i X_i^2 - \left(\sum w_i X_i\right)^2}$ y $\hat{\beta}_0^* = \overline{Y}_w - \hat{\beta}_1^* \overline{X}_w$ son los estimadores MCP de β_1 y β_0, respectivamente; $\overline{Y}_w = \dfrac{\sum w_i Y_i}{\sum w_i}$ y $\overline{X}_w = \dfrac{\sum w_i X_i}{\sum w_i}$.

ii) $Var\left[\hat{\beta}_0^*\right] = \dfrac{\frac{1}{n}\sum w_i \sum w_i X_i^2}{\sum w_i \sum w_i X_i^2 - \left(\sum w_i X_i\right)^2}$, $Var\left[\hat{\beta}_1^*\right] = \dfrac{\sum w_i}{\sum w_i \sum w_i X_i^2 - \left(\sum w_i X_i\right)^2}$ y

$Cov\left[\hat{\beta}_0^*, \hat{\beta}_1^*\right] = -\dfrac{\overline{X}_w \sum w_i}{\sum w_i \sum w_i X_i^2 - \left(\sum w_i X_i\right)^2}$

Desde el punto de vista computacional, para calcular los estimadores MCP es mucho más práctico hacer las transformaciones $Y_i^* = \dfrac{Y_i}{\sigma_i}$, $X_{0i}^* = \dfrac{1}{\sigma_i}$, $X_1^* = \dfrac{X_i}{\sigma_i}$ y estimar por MCO el MRLM sin intercepto $Y_i^* = \beta_0^* X_{0i}^* + \beta_1^* X_{1i}^* + u_i^*$.

La corrección de heterocedasticidad mediante MCG (o MCP) genera estimadores puntuales $\hat{\boldsymbol{\beta}}^*$ similares a los de MCO $\hat{\boldsymbol{\beta}}$ (ambos estimadores son insesgados y consistentes). Sin embargo, en presencia de heterocedasticidad, MCO sobrestiman las varianzas, por lo que generalmente $Var\left[\hat{\beta}_j\right] > Var\left[\hat{\beta}_j^*\right]$ (es decir, $\hat{\boldsymbol{\beta}}$ no es MELI pero $\hat{\boldsymbol{\beta}}^*$ si lo es).

Ejercicio $\boxed{\text{E56}}$.

CORRECCIÓN DE LA HETEROCEDASTICIDAD CON VARIANZAS DESCONOCIDAS

Si las varianzas σ_i^2, i = 1, 2,…, n, son desconocidas, la corrección de la heterocedasticidad se puede realizar mediante los siguientes métodos:

- Procedimiento de White.
- Transformaciones de estabilización de varianza.

Procedimiento de White. Consiste en estimar el MRLMH mediante MCO pero ajustar la varianza de los estimadores para no sobrestimarla. La estimación de la matriz de varianzas-covarianzas propuesta por White considera los residuos $\hat{u}_1$, $\hat{u}_2$,…, $\hat{u}_{1n}$, de la estimación MCO :

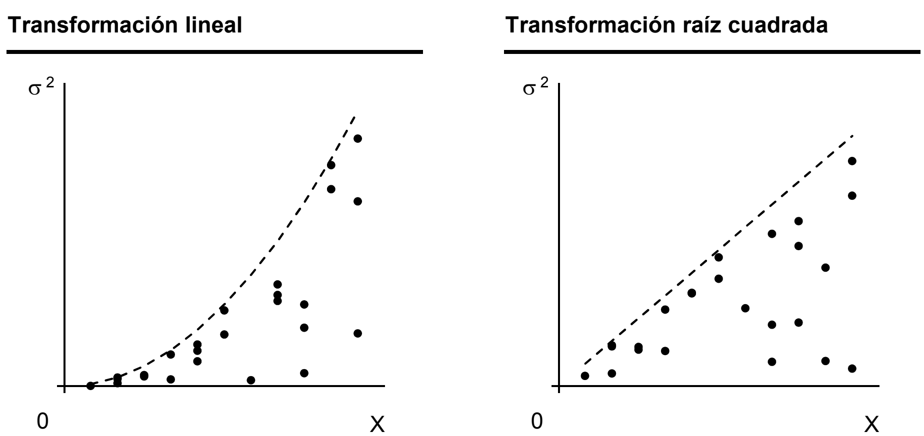
$$\hat{\mathbf{\Sigma}}_{\hat{\beta}^*} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{V}}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}, \quad \text{donde} \quad \hat{\mathbf{V}} = \begin{bmatrix} \hat{u}_1^2 & 0 & \cdots & 0 \\ 0 & \hat{u}_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \hat{u}_n^2 \end{bmatrix}.$$

Algunos paquetes estadísticos permiten calcular directamente la matriz de varianzas-covarianzas del procedimiento de White e incluso reportan sus respectivos errores estándar (raíz cuadrada de los elementos de la diagonal principal de $\hat{\mathbf{\Sigma}}_{\hat{\beta}^*}$), a los cuales se les denomina **errores estándar robustos**.

Transformaciones de estabilización de varianza. Consisten en transformar un MRLMH para que sea homocedástico. Parten del supuesto de que el patrón que sigue la varianza de los errores en función del regresor, $\sigma_i^2 = g(X_i)$, es conocido.

La figura 5.7 muestra el caso en que la varianza del error es proporcional al cuadrado de los regresores $\sigma_i^2 = \sigma^2 X_i^2$ (*transformación lineal*, lado izquierdo) y el caso en que la varianza del error es directamente proporcional a los regresores $\sigma_i^2 = \sigma^2 X_i$ (*transformación raíz cuadrada*, lado derecho). En ambos casos σ^2 es una constante (que por conveniencia denotamos de esta manera).

Figura 5.7



Por ejemplo, si $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde $u_i \sim \left(0, \sigma_i^2\right)$ y se sospecha que $\sigma_i^2 = \sigma^2 X_i^2$, entonces es posible aplicar la **transformación lineal** dividiendo ambos lados del MRLS entre X_i obteniendo $\frac{Y_i}{X_i} = \beta_0 \left(\frac{1}{X_i}\right) + \beta_1 + \frac{u_i}{X_i}$, donde:

$$E\left[\frac{u_i}{X_i}\right] = \frac{1}{X_i} E[u_i] = 0 \quad \text{y} \quad Var\left[\frac{u_i}{X_i}\right] = \frac{1}{X_i^2} Var[u_i] = \frac{1}{X_i^2}\left(\sigma^2 X_i^2\right) = \sigma^2,$$

es decir, el modelo transformado es homocedástico, puede ser estimado por MCO y sus estimadores son MELI.

Importante: Las transformaciones de estabilización de varianza eliminan la heterocedasticidad pero no estiman a Y_i directamente, sino una transformación de Y_i. Si se desean hacer estimaciones o predicciones sobre Y_i hay que **aplicar la transformación inversa**.

Existen otras transformaciones que buscan estabilizar la varianza (eliminar la heterocedasticidad) pero que no requiere suponer un patrón para la varianza de los errores. Entre estas transformaciones están:

- La *transformación logarítmica*, que comprime las escalas de X y Y disminuyendo el problema de heterocedasticidad; y
- La *transformación potencia*, que elimina la heterocedasticidad al considerar el tiempo como regresor y se utiliza con frecuencia en el Análisis de Series de Tiempo.

Ejercicio $\boxed{\text{E57}}$.

5.4. Autocorrelación

5.4.1. Definición y Causas de la Autocorrelación

El supuesto (v) del MRLM establece que los errores deben ser independientes, lo que implica que sean no correlacionados, es decir, $Cov[u_i, u_j] = 0$ para toda $i \neq j$. El problema que surge por la violación de este supuesto se denomina autocorrelación.

Existen 2 situaciones en las cuales los errores del MRLM pueden estar correlacionados:

- En *datos transversales*, la correlación puede presentarse en unidades muestrales contiguas. Por ejemplo, en estudios de consumo los errores de hogares en la misma colonia pueden estar correlacionados. A este tipo de correlación se le denomina **correlación espacial**; y
- En *series de tiempo*, la correlación puede presentarse en observaciones consecutivas en el tiempo (en uno o varios períodos). A este tipo de correlación se le denomina **correlación serial** o **autocorrelación**.

En esta sección se analizará la autocorrelación que surge al explicar la serie de tiempo Y_t, mediante las series de tiempo X_{1t}, X_{2t},…, X_{kt}, t = 1, 2,…, n; es decir, todas las variables estarán indizadas por el tiempo t.

Def. Autocorrelación.

Se dice que el MRLM $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t$ presenta el problema de autocorrelación si $Cov[u_t, u_{t-k}] \neq 0$, para alguna k = 1, 2,…

Bajo el supuesto de homocedasticidad, $Cov[u_t, u_t] = Var[u_t] = \sigma^2$. Si $Cov[u_t, u_{t-1}] \neq 0$ se dice que hay **autocorrelación de primer orden**, si $Cov[u_t, u_{t-2}] \neq 0$ se dice que hay **autocorrelación de segundo orden**, y así sucesivamente.

Consecuencias de la autocorrelación:

- Los estimadores MCO son lineales, insesgados y consistentes, pero dejan de tener mínima varianza (no son MELI).
- Es posible que el estimador de la varianza $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$ subestime el verdadero valor de σ^2 y sobrestime R^2, invalidando las Pruebas T y F, a pesar de que el supuesto de normalidad se valide.

Causas de la autocorrelación:

- *Inercia o pasividad*. Surge por la existencia de un impulso ascendente o descendente en el comportamiento de una serie de tiempo.
- *Sesgo de especificación*. Omisión de alguna variable que teóricamente debe explicar el comportamiento de una serie de tiempo. Por ejemplo, si existe una tendencia cuadrática omitir el término t^2 provoca autocorrelación.
- *Rezagos*. Ocurre cuando la variable explicativa depende de rezagos (de la propia variable o de algún regresor) y éstos se omiten en el modelo.

5.4.2. Identificación de la Autocorrelación

No existe un método único que permita identificar la autocorrelación, pero existen métodos gráficos y pruebas estadísticas para su detección.

MÉTODOS GRÁFICOS

Serie de tiempo de los residuos. Consiste en graficar los residuos $\hat{u}_1$, $\hat{u}_2$,…, $\hat{u}_n$, en forma ordenada a lo largo del tiempo t. La existencia de algún tipo de patrón es indica que posiblemente haya autocorrelación.

Residuos vs residuos rezagados. Consiste en construir el diagrama de dispersión de los residuos $\hat{u}_t$ vs. los residuos rezagados $\hat{u}_{t-1}$, t = 2, 3,…, n. Una asociación (positiva o negativa) indica la existencia de autocorrelación de primer orden. Para analizar la existencia de autocorrelación de orden superior basta graficar $\hat{u}_t$ vs. $\hat{u}_{t-k}$, k = 2, 3,…

Estos gráficos permiten identificar 2 tipos de autocorrelaciones:

- **Autocorrelación positiva**, si se identifican grupos de residuos positivos y/o grupos de residuos negativos y las parejas ($\hat{u}_{t-1}$, $\hat{u}_t$), t = 2, 3,…, n, presentan asociación positiva (ver figura 5.8); y
- **Autocorrelación negativa**, si se identifican brincos alternados de residuos positivos y negativos, y las parejas ($\hat{u}_{t-1}$, $\hat{u}_t$), t = 2, 3,…, n, presentan asociación negativa (ver figura 5.9).

Figura 5.8

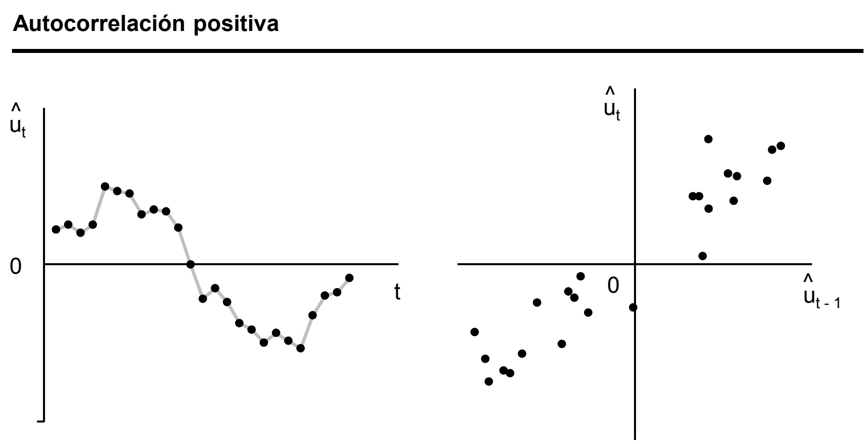
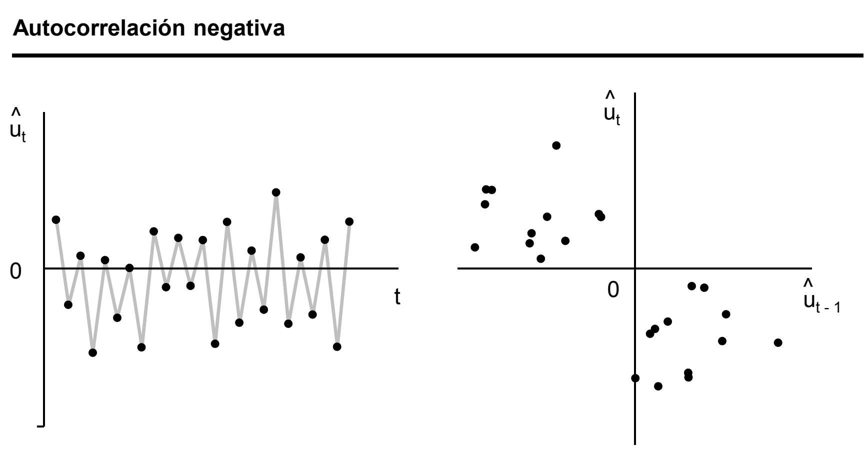


Figura 5.9



PRUEBAS ESTADÍSTICAS

Prueba Durbin-Watson. Esta prueba permite identificar **autocorrelación de primer orden**, pues suponte que el error (u_t) sigue un *proceso autorregresivo de primer orden*, AR(1), definido por el MRLS sin intercepto:

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid\left(0, \sigma_\varepsilon^2\right),$$

donde ρ es el *coeficiente de autocovarianza* y los errores ε_t con media cero, varianza constante y no correlacionados se denominan *ruido blanco*.

Si $\rho = 0$, entonces no hay autocorrelación, pero si $\rho > 0$ hay autocorrelación positiva y si $\rho < 0$ hay autocorrelación negativa. Es decir, existen 2 posibles hipótesis alternativas, por lo que es más fácil establecer 2 hipótesis nulas negando las posibles alternativas:

- H_0^+: No hay autocorrelación positiva vs H_1^+: Hay autocorrelación positiva; y
- H_0^-: No hay autocorrelación negativa vs. H_1^-: Hay autocorrelación negativa.

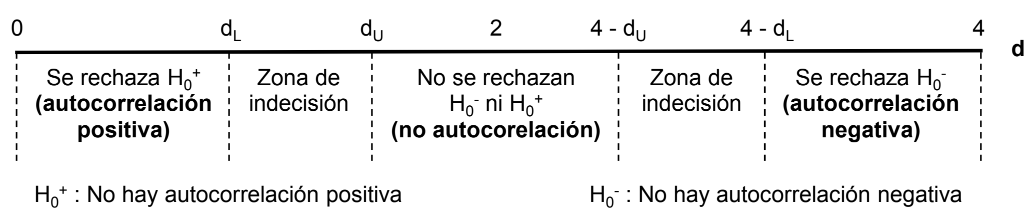
El estadístico de prueba Durbin-Watson es:

$$d = \frac{\sum_{t=2}^{n}\left(\hat{u}_t - \hat{u}_{t-1}\right)^2}{\sum_{t=1}^{n}\hat{u}_t^2},$$

y se compara contra **valores críticos d_L y d_U** calculados para distintos tamaños de muestra n y para distintos niveles de confianza α que aparecen en la **Tabla Durbin-Watson** (ver Formulario, p.42-45, en donde se consideran n = 6, 7,…, 40, 45,…, 100, 150 y 200, y α = 0.01 y 0.05).

La figura 5.10 muestra las **regiones de rechazo** de las hipótesis nulas.

Figura 5.10



La Prueba Durbin-Watson **puede ser no concluyente** si el estadístico de prueba d se ubica en las *zonas de indecisión* de la figura 5.10. La mayoría de los paquetes estadísticos reportan automáticamente el cálculo del estadístico de prueba d.

5.4.3. Corrección de la Autocorrelación

Si un MRLM presenta **autocorrelación de primer orden** se puede corregir con distintas técnicas, dependiendo si el coeficiente de autocovarianza ρ es conocido o no.

CORRECCIÓN DE LA AUTOCORRELACIÓN DE PRIMER ORDEN CON ρ CONOCIDA

Si el coeficiente de autocovarianza ρ es conocido, es posible hacer la estimación MCG aplicando la *transformación de primeras diferencias*, para obtener un modelo no correlacionado y estimarlo por MCO.

Por ejemplo, para el MRLS $Y_t = \beta_0 + \beta_1 X_t + u_t$ con $u_t = \rho u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid\left(0, \sigma_\varepsilon^2\right)$, es posible considerar el modelo al tiempo $t - 1$ y multiplicarlo por ρ:

$$\rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 X_{t-1} + \rho u_{t-1},$$

y restando esta última expresión del modelo original se obtiene el MRLS con intercepto:

$$Y_t - \rho Y_{t-1} = \beta_0\left(1-\rho\right) + \beta_1\left(X_t - \rho X_{t-1}\right) + \varepsilon_t .$$

Desde el punto de vista computacional, para calcular los estimadores MCG es más práctico hacer las transformaciones $Y_t^* = Y_t - \rho Y_{t-1}$, $\beta_0^* = \beta_0\left(1-\rho\right)$, $X_t^* = X_t - \rho X_{t-1}$ y estimar por MCO el modelo con intercepto $Y_t^* = \beta_0^* + \beta_1^* X_t^* + \varepsilon_t$.

CORRECCIÓN DE LA AUTOCORRELACIÓN DE PRIMER ORDEN CON ρ DESCONOCIDA

Si ρ no es conocida, se puede estimar su valor y a partir de éste corregir la autocorrelación, o simplemente estimar por MCO y corregir su respectiva matriz de varianzas-covarianzas.

Estimación de ρ. Existen métodos iterativos como el de Cochrane-Orcutt para estimar a ρ. En EViews también es posible estimarla fácilmente incorporando el proceso $u_t = \rho u_{t-1} + \varepsilon_t$ directamente en el MRLM mediante el regresor AR(1). Su respectivo coeficiente de MCO corresponde al valor de $\hat{\rho}$.

Método Newey-West. Consiste en estimar el MRLM con autocorrelación mediante MCO pero ajustar la varianza de los estimadores. La estimación de la matriz de varianzas-covarianzas propuesta por Newey-West es compleja pues ajusta simultáneamente autocorrelación y heterocedasticidad, sin embargo, algunos paquetes estadísticos la incluyen. A los errores estándar que genera este método se le denominan **errores estándar CHA** (consistentes con heterocedasticidad y autocorrelación).

Ejercicio $\boxed{\text{E58}}$.

Material de Apoyo para el Curso de Econometría

David Ruelas Rodríguez, Enero – Mayo 2015

E1. Para las variables cuantitativas X y Y se tienen los datos $(X_1, Y_1), (X_2, Y_2),\ldots, (X_n, Y_n)$.

a) A partir de las definiciones de varianza muestral y covarianza muestral demuestre que $S_Y^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right)$ y $S_{XY} = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}\right)$.

b) Demuestre que $r_{XY} = \dfrac{n\sum_{i=1}^{n} X_i Y_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{\sqrt{\left[n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right]\left[n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2\right]}}$.

E2. Algunos Economistas apoyan la teoría de que un mayor "ingreso mensual" (Y) está asociado a un mayor número de "años de estudios de posgrado" (X). A continuación se presentan estas variables y algunas otras para una muestra aleatoria de 22 personas con estudios de posgrado:

Años de estudios de posgrado	**Ingreso mensual** (miles de dólares)	**Tipo de ingreso** (1)	**Tipo de actividad**	**Estudios** (2)
2.0	4.0	Medio bajo	Otra	1
5.0	4.0	Medio bajo	Academia	1
3.0	5.2	Medio alto	Otra	1
2.0	7.5	Alto	Negocios	2
3.5	2.3	Medio bajo	Academia	1
2.0	4.5	Medio alto	Otra	2
4.0	2.7	Medio bajo	Academia	2
0.5	4.5	Medio alto	Negocios	2
5.0	4.8	Medio alto	Academia	2
4.0	6.2	Alto	Otra	1
4.0	5.8	Medio alto	Otra	1
2.0	7.5	Alto	Negocios	2
5.5	5.8	Medio alto	Academia	1
3.0	2.1	Medio bajo	Academia	2
1.0	5.5	Medio alto	Negocios	1
2.0	4.0	Medio bajo	Otra	2
2.5	4.3	Medio alto	Otra	2
1.5	6.5	Alto	Negocios	2
6.0	7.5	Alto	Academia	1
3.0	4.8	Medio alto	Otra	1
3.5	5.5	Medio alto	Otra	1
4.5	3.3	Medio bajo	Academia	2

(1) Clasificado como Bajo en [0, 2], Medio bajo en (2, 4], Medio alto en (4, 6] y Alto mayor a 6 mil dólares.

(2) 1 = en universidades de Estados Unidos, 2 = en universidades fuera de Estados Unidos.

Fuente: Ilustrativo

Adicionalmente para las variables X y Y se calcularon las siguientes estadísticas:

$\sum X_i = 69.5$	$\sum X_i^2 = 265.75$	$\sum X_i Y_i = 341.4$
$\sum Y_i = 108.3$	$\sum Y_i^2 = 585.21$	$n = 22$

a) Calcule el coeficiente de correlación entre las variables "años de estudios de posgrado" e "ingreso mensual". ¿Este resultado es consistente con la teoría de algunos Economistas?

b) Clasifique cada una de las variables e indique su escala de medición.

c) Construya el diagrama de dispersión de las variables "años de estudios de posgrado" e "ingreso mensual", distinguiendo el "tipo de actividad". ¿Qué se puede decir acerca de la variable "tipo de actividad"?

d) Calcule por separado el coeficiente de correlación entre las variables "años de estudios de posgrado" e "ingreso mensual" para cada categoría de la variable "tipo de actividad". ¿Estos resultados son consistentes con la teoría de algunos Economistas?

e) Utilice Excel para determinar las ecuaciones de las rectas de regresión para explicar el "ingreso mensual" mediante los "años de estudio de posgrado". ¿Qué puede concluir en cada caso?

f) Utilice Excel para determinar la ecuación de la regresión exponencial para explicar el "ingreso mensual" mediante los "años de estudio de posgrado" de las personas con posgrado que se dedican a la academia.

g) Para las personas con posgrado que se dedican a la academia construya el diagrama de dispersión de las variables "años de estudios de posgrado" y el "log-ingreso mensual". Calcule en este caso el coeficiente de correlación y compárelo con el resultado del inciso d). Determine en este caso también la ecuación de la recta de regresión y compárela con la ecuación de la regresión exponencial del inciso f).

E3. Una inmobiliaria ha determinado que si X es el número de habitaciones de los departamentos que maneja y Y el número de lugares de estacionamiento, entonces su distribución conjunta se muestra en la siguiente tabla:

Número de habitaciones	**Lugares de estacionamiento**					**Total**
	0	1	2	3	4	
1	0.06	0.12	0.02	0	0	0.20
2	0.03	0.18	0.21	0	0	0.42
3	0.01	0.09	0.11	0.07	0	0.28
4	0	0.02	0.05	0.02	0.01	0.10
Total	0.10	0.41	0.39	0.09	0.01	1

a) ¿Son X y Y variables independientes?

b) Calcule $E[Y|X=x]$ y $Var[Y|X=x]$ para $x = 1, 2, 3, 4$.

c) Grafique $g(x) = E[Y|X=x]$ para $x = 1, 2, 3, 4$. ¿Qué puede concluir?

E4. Las empresas petroleras integradas típicamente pagan impuestos compuestos por una regalía (*royalty*) y una serie de impuestos y derechos adicionales. Si X denota la tasa de la regalía y Y la tasa del impuesto total que paga una empresa petrolera, entonces Y puede ser explicada por el nivel de X. Si $f_{X,Y}(x,y) = 2xI_{(x,1)}(y)I_{(0,1)}(x)$, calcule y grafique $E[Y|X=x]$.

E5. Para cada una de las siguientes Funciones de Regresión Poblacionales indique si son lineales en variables, en parámetros, en ambos o en ninguna de las dos:

a) $E[Y|X_i] = \beta_0 + \beta_1^2 X_i$

b) $E[Y|X_i] = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k$

c) $E[Y|X_i] = e^{\beta_0 + \beta_1 X_i}$

d) $E[Y|X_i] = \dfrac{X_i}{\beta_0 + \beta_1 X_i}$

E6. Se sabe que si X y Y tienen distribución conjunta Normal Bivariada con medias μ_X, μ_Y, varianzas σ_X^2 y σ_Y^2, y coeficiente de correlación ρ, entonces su función de densidad de probabilidad conjunta es $f_{X,Y}(x,y) = \dfrac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\dfrac{1}{2(1-\rho^2)}Q(x,y)\right\}$ para x, $y \in \mathbf{R}$, donde $Q(x,y) = \left(\dfrac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\dfrac{x-\mu_X}{\sigma_X}\right)\left(\dfrac{y-\mu_Y}{\sigma_Y}\right) + \left(\dfrac{y-\mu_Y}{\sigma_Y}\right)^2$. Además se sabe que $\{Y|X=x\} \sim N\left(\mu_Y + \rho\dfrac{\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y^2(1-\rho^2)\right)$. Si se quiere explicar a Y mediante el valor fijo X_i …

a) ¿Se puede decir que su FRP es lineal en variables o lineal en parámetros?

b) Grafique $E[Y|X_i]$ y $Var[Y|X_i]$.

c) Si se quiere expresar la esperanza condicional $E[Y|X_i]$ en la forma $\beta_0 + \beta_1 X_i$, ¿qué valores deben tomar β_0 y β_1?

d) ¿Qué pasa con $f_{X,Y}(x,y)$ si X y Y están perfectamente correlacionadas, es decir, si $\rho = 1$? Explique.

E7. Considere el Modelo Multiplicativo $Y_i = \alpha_0 X_i^{\beta_1} e^{u_i}$, donde u_i es perturbación estocástica:

a) Demuestre que este modelo puede ser transformado en Modelo de Regresión Lineal Simple.

b) Calcule efecto parcial, elasticidad y semi-elasticidad para este modelo expresado como relación determinista.

E8. Bajo los supuestos del Modelo de Regresión Lineal Simple $Y_i = \beta_0 + \beta_1 X_i + u_i$, demuestre que:

a) $E\left[Y_i \middle| X_i\right] = \beta_0 + \beta_1 X_i$

b) $Var\left[Y_i \middle| X_i\right] = \sigma^2$

c) $Cov\left[Y_i, Y_j \middle| X_i, X_j\right] = 0$ para toda $i \neq j$

E9. **Estimación por Mínimos Cuadrados Ordinarios (MCO) del MRLS.** Para el MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde $u_i \sim iid(0, \sigma^2)$:

a) Obtenga las ecuaciones normales:

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i \quad \text{y} \quad \sum_{i=1}^{n} X_i Y_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2$$

b) Demuestre que $\hat{\beta}_1 = \dfrac{S_{XY}}{S_X^2}$ y $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$

E10. El desarrollo de las telecomunicaciones en los últimos 20 años ha provocado que cada vez más personas destinen parte de su ingreso a la adquisición de teléfonos celulares. A continuación se presentan el consumo de teléfonos celulares (por cada 100 habitantes) y el ingreso *per cápita* anual de 2003 para una muestra de 34 países.

País	**Teléfono celular** (por cada 100 personas)	**Ingreso *per cápita*** (Miles de dólares)	**País**	**Teléfono celular** (por cada 100 personas)	**Ingreso *per cápita*** (Miles de dólares)
Alemania	78.5	27.6	Holanda	76.8	28.6
Arabia Saudita	32.1	13.2	Hungría	76.9	13.8
Argentina	17.8	11.4	India	2.5	2.9
Australia	72.0	28.8	Indonesia	8.7	3.2
Bélgica	79.3	28.9	Inglaterra	91.2	27.7
Brasil	26.4	7.5	Italia	101.8	26.8
Bulgaria	46.6	0.1	Japón	67.9	28.5
Canadá	41.9	30.0	México	29.5	9.0
China	21.5	5.0	Pakistán	1.8	2.0
Colombia	14.1	6.4	Polonia	45.1	11.2
Ecuador	18.9	3.9	República Checa	96.5	15.6
Egipto	8.5	3.9	Rusia	24.9	9.0
España	91.6	22.2	Sudáfrica	36.4	10.1
Estados Unidos	54.6	37.8	Suecia	98.1	26.7
Francia	69.6	27.6	Suiza	84.3	32.2
Grecia	90.2	19.9	Tailandia	39.4	7.5
Guatemala	13.2	4.1	Venezuela	27.3	4.8

Fuente: Gujarati, Econometría, quinta edición, 2009. Tabla 3.3 Statistical Abstract of the United States, 2006.

Si X denota el ingreso per cápita (en miles de dólares) y Y el consumo de teléfonos celulares (por cada 100 habitantes) entonces:

$\sum\left(X_i - \overline{X}\right)^2 = 4{,}160.3$	$\sum\left(Y_i - \overline{Y}\right)^2 = 34{,}004.3$
$\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = 9{,}232.5$	$\sum Y_i^2 = 117{,}600.2$

Además, la media muestral del ingreso per cápita es de 15.8 miles de dólares. Considere el MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde $u_i \sim iid(0, \sigma^2)$. Estime por Mínimos Cuadrados Ordinarios (MCO) los parámetros β_0 y β_1.

E11. Continuando con el MRLS del ejercicio E10…

a) Construya el diagrama de dispersión que permita ver si el consumo de teléfonos celulares (Y) puede ser explicado por el ingreso per cápita (X).

b) Utilice Excel para verificar la estimación MCO de β_0 y β_1.

c) Calcule los residuos $\hat{u}_i = Y_i - \hat{Y}_i$ (valor real menos valor ajustado) y grafíquelos para cada valor observado X_i. A partir de esta gráfica y el diagrama de dispersión del inciso a) analice empíricamente si se cumplen los 8 supuestos del MRLS.

d) Estime la varianza de los términos de error (σ^2).

E12. Demuestre que, si el MRLS $Y_i = \beta_0 + \beta_1 X_i + u_i$ se estima por MCO entonces:

a) La suma de los residuos es cero: $\sum_{i=1}^{n} \hat{u}_i = 0$.

b) Los residuos y los regresores no están correlacionados: $\sum_{i=1}^{n} \hat{u}_i X_i = 0$.

c) Los residuos y los valores ajustados no están correlacionados: $\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = 0$.

E13. Continuando con los ejercicios E10 y E11 verifique que: (i) $\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X}$, (ii) $\overline{\hat{Y}} = \overline{Y}$, (iii) $\sum_{i=1}^{n} \hat{u}_i = 0$, (iv) $\sum_{i=1}^{n} \hat{u}_i X_i = 0$, y (v) $\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = 0$. Interprete el resultado en cada caso.

E14. Si $\hat{\beta}_0$ y $\hat{\beta}_1$ son los estimadores MCO del MRLS, demuestre que:

a) $\hat{\beta}_1$ es insesgado.

b) $Var\left[\hat{\beta}_1\right] = \dfrac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$ y $Cov\left[\hat{\beta}_0, \hat{\beta}_1\right] = -\dfrac{\overline{X}\sigma^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$

E15. **Teorema de Gauss-Markov**. Bajo los supuestos del MRLS demuestre que los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$ son los Mejores Estimadores Lineales Insesgados (MELI).

E16. Continuando con los ejercicios E10, E11 y E13, estime la matriz de varianzas y covarianzas de los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$.

E17. Demuestre que al estimar el MRLS mediante MCO, el coeficiente de determinación es igual al coeficiente de correlación lineal muestral al cuadrado, es decir, $R^2 = (r_{XY})^2$.

E18. Continuando con los ejercicios E10, E11, E13 y E16…

a) Verifique que la suma de cuadrados totales (*SCT*) es igual a la suma de la suma de cuadrados explicada (*SCE*) más la suma de cuadrados de los residuos (*SCR*), es decir, $SCT = SCE + SCR$.

b) Calcule e interprete el coeficiente de determinación.

c) Construya la tabla ANOVA.

E19. **Estimación por Máxima Verosimilitud del MRLSN**. Considere el MRLSN $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde $u_i \sim iid\ N(0,\ \sigma^2)$, y demuestre que los EMV de β_0, β_1 y σ^2 son $\hat{\beta}_1 = \dfrac{S_{XY}}{S_X^2}$, $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$ y $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2$, respectivamente.

E20. Considere el MRLSN. Utilice el Método Pivotal para demostrar que el intervalo simétrico al $100(1-\alpha)\%$ de confianza para β_0 es $\hat{\beta}_0 \pm t_{n-2,\frac{\alpha}{2}} S_{\hat{\beta}_0}$. ¿Cuál es el respectivo intervalo de confianza para β_1?

E21. Continuando con los ejercicios E10, E11, E13, E16 y E18, se sabe que $\hat{\sigma}^2 = 422.4$ y que la matriz de varianzas-covarianzas estimada de $\hat{\beta}_0$ y $\hat{\beta}_1$ es $\begin{bmatrix} 37.833 & -1.606 \\ -1.606 & 0.102 \end{bmatrix}$. Si se supone que $u_i \sim iid\ N(0,\ \sigma^2)$…

a) Calcule los intervalos de confianza al 95% para β_0 y β_1. ¿Podría pensarse que alguno de estos parámetros es cero?

b) Calcule el intervalo de confianza al 95% para la desviación estándar de los términos de error u_i.

c) ¿Es razonable el supuesto de normalidad en los errores?

E22. Demuestre que, bajo los supuestos del MRLSN, si $\beta_1 = 0$, entonces la estadística $F = \dfrac{(n-2)SCE}{SCR} \sim F(1, n-2)$.

E23. Continuando con los ejercicios E10, E11, E13, E16, E18 y E21, si $u_i \sim iid\ N(0,\ \sigma^2)$…

a) Si se quiere probar H_0: $\beta_0 = 0$ vs. H_1: $\beta_0 > 0$, calcule su respectivo valor-p. ¿Qué se puede concluir?

b) ¿Se puede afirmar con un error tipo I de 5% que el ingreso per cápita permite explica el consumo de teléfonos celulares? Realice las pruebas *T* y *F*. Calcule el valor-p en cada caso.

E24. Bajo los supuestos del MRLS, considere a $\hat{E}\left[Y_0 \middle| X_0\right] = \hat{Y}_0^* = \hat{\beta}_0 + \hat{\beta}_1 X_0$ como estimador puntual de $E\left[Y_0 \middle| X_0\right] = \beta_0 + \beta_1 X_0$.

a) Demuestre que $\hat{Y}_0^*$ es predictor insesgado.

b) Demuestre que $Var\left[\hat{Y}_0^*\right] = \sigma^2 \left[\frac{1}{n} + \frac{\left(X_0 - \overline{X}\right)^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} \right]$.

E25. Continuando con los ejercicios E10, E11, E13, E16, E18, E21 y E23 se sabe que $n = 34$, $\hat{\beta}_0 = 14.48$, $\hat{\beta}_1 = 2.22$, $\hat{\sigma}^2 = 422.4$, $\overline{X} = 15.8$, y $\sum\left(X_i - \overline{X}\right)^2 = 4{,}160.3$. Si se supone que $u_i \sim iid\ N(0, \sigma^2)$…

a) Estime la media del consumo de teléfonos celulares (por cada 100 habitantes) para una economía con ingreso per cápita de 40 mil dólares. Proporcione el pronóstico puntual y su respectivo intervalo de confianza al 95% de confianza.

b) Estime el nivel de consumo de teléfonos celulares (por cada 100 habitantes) para una economía con ingreso per cápita de 40 mil dólares. Proporcione el pronóstico puntual y su respectivo intervalo de predicción con 95% de probabilidad.

E26. Continuando con los ejercicios E10, E11, E13, E16, E18, E21, E23 y E25, considerando válido el supuesto de normalidad, es decir, si $u_i \sim iid\ N(0, \sigma^2)$…

a) Calcule y grafique las bandas de predicción al 95% de la media condicional y de valores individuales del consumo de teléfonos celulares (Y) considerando los valores observados del ingreso per cápita (X).

b) ¿Para qué valor del ingreso per cápita la amplitud de estas bandas del inciso anterior se minimiza?

E27. Los *Modelos de Gasto de Engel*, en honor del Estadístico Alemán Ernst Engel (1821 – 1896) postulan que "el gasto total en un bien tiende a incrementarse en progresión aritmética, mientras que el gasto total aumenta en progresión geométrica".

En la siguiente tabla se presentan los datos sobre el gasto de consumo personal total (*GT*), el gasto en bienes duraderos (*GD*, incluyendo vehículos automotores y refacciones, muebles y equipo doméstico), el gasto en bienes perecederos (*GP*, incluyendo comida, ropa, gasolina, aceite, combustibles de petróleo y carbón mineral) y el gasto en servicios (*GS*, incluyendo vivienda, electricidad, gas, transporte y atención médica), todos medidos en miles de millones de dólares de 2000, del primer trimestre de 2003 al cuarto trimestre de 2006 en Estados Unidos.

Gasto trimestral en Estados Unidos 2003-2006
Cifras en miles de millones de dólares de 2000

Año y trimestre	Gasto de Consumo Personal Total (GT)	Gasto en Bienes Duraderos (GD)	Gasto en Bienes Perecederos (GP)	Gasto en Servicios (GS)
2003-I	7,184.9	971.4	2,072.5	4,143.3
2003-II	7,249.3	1,009.8	2,084.2	4,161.3
2003-III	7,352.9	1,049.6	2,123.0	4,190.7
2003-IV	7,394.3	1,051.4	2,132.5	4,220.2
2004-I	7,479.8	1,067.0	2,155.3	4,268.2
2004-II	7,534.4	1,071.4	2,164.3	4,308.4
2004-III	7,607.1	1,093.9	2,184.0	4,341.5
2004-IV	7,687.1	1,110.3	2,213.1	4,377.4
2005-I	7,739.4	1,116.8	2,241.5	4,395.3
2005-II	7,819.8	1,150.8	2,268.4	4,420.0
2005-III	7,895.3	1,175.9	2,287.6	4,454.5
2005-IV	7,910.2	1,137.9	2,309.6	4,476.7
2006-I	8,003.8	1,190.5	2,342.8	4,494.5
2006-II	8,055.0	1,190.3	2,351.1	4,535.4
2006-III	8,111.2	1,208.8	2,360.1	4,566.6

Fuente: Gujarati, *Econometría,* quinta edición, 2009. Adaptación de la Tabla 6.3, p. 161

a) Estime puntualmente el gasto de consumo personal total para el cuarto trimestre de 2006. ¿Cambia su estimación dependiendo de la forma de etiquetar el tiempo?

b) Analice gráficamente si es razonable pensar que el gasto de consumo personal total explica el gasto en bienes duraderos.

c) Ajuste los siguientes modelos:

Modelo 1: $GD_i = \alpha_0 + \alpha_1 GT_i + u_i$

Modelo 2: $\ln(GD_i) = \beta_0 + \beta_1 GT_i + v_i$

Modelo 3: $\ln(GD_i) = \gamma_0 + \gamma_1 \ln(GT_i) + w_i$

¿Son comparables sus coeficientes de determinación? ¿Qué modelo considera que proporciona el mejor ajuste? Justifique su respuesta.

d) Utilice los Modelos 1 y 2 del inciso anterior para estimar puntualmente y por intervalo al 95% el valor del gasto en bienes duraderos si $GT_0 = 2{,}800$.

E28. Considere el MRLM $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde $\mathbf{u} \sim (\mathbf{0}, \sigma^2\mathbf{I})$.

a) Calcule $\mathbf{X}'\mathbf{X}$ para $k = 2$.

b) Calcule $(\mathbf{X}'\mathbf{X})^{-1}$ para $k = 1$. ¿Qué puede concluir?

c) Calcule esperanza y varianza condicionales de $\mathbf{y}$ dado $\mathbf{X}$.

d) Calcule $\dfrac{\partial}{\partial \boldsymbol{\beta}}\left(2(\mathbf{X}\boldsymbol{\beta})'\mathbf{y}\right)$ y $\dfrac{\partial}{\partial \boldsymbol{\beta}}\left((\mathbf{X}\boldsymbol{\beta})'\mathbf{X}\boldsymbol{\beta}\right)$. Obtenga las matrices de segundas derivadas en cada caso.

E29. **Estimación MCO del MRLM.** Para el MRLM $\mathbf{y} = \boldsymbol{\beta}\mathbf{X} + \mathbf{u}$ donde $\mathbf{X}'\mathbf{X}$ es no estocástica y no singular, y donde $\mathbf{u} \sim \left(\mathbf{0}, \sigma^2\mathbf{I}\right)$:

a) Obtenga las ecuaciones normales: $\left(\mathbf{X}'\mathbf{X}\right)\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$.

b) Demuestre que el vector de estimadores MCO de $\boldsymbol{\beta}$ es $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{y}\right)$.

E30. Estudios demográficos indican que la mortalidad infantil puede ser explicada por la educación de las mujeres de la población, su tasa de fecundidad y la riqueza de sus habitantes. En la siguiente tabla se presenta la *mortalidad infantil* (*MI*, medida en número de defunciones de menores de 5 años por cada 1,000 nacidos vivos), la *tasa de alfabetismo femenina* (*TA*, en porcentaje), el *producto interno bruto per cápita* (*PIB*, en dólares correspondiente a 1980) y la *tasa de fecundidad* (*TF*, promedio de hijos por mujer con información entre 1980 y 1985) para una muestra de 64 países:

Mortalidad infantil

Obs.	MI	TA	PIB	TF	Obs.	MI	TA	PIB	TF
1	128	37	1,870	6.66	33	142	50	8,640	7.17
2	204	22	130	6.15	34	104	62	350	6.60
3	202	16	310	7.00	35	287	31	230	7.00
4	197	65	570	6.25	36	41	66	1,620	3.91
5	96	76	2,050	3.81	37	312	11	190	6.70
6	209	26	200	6.44	38	77	88	2,090	4.20
7	170	45	670	6.19	39	142	22	900	5.43
8	240	29	300	5.89	40	262	22	230	6.50
9	241	11	120	5.89	41	215	12	140	6.25
10	55	55	290	2.36	42	246	9	330	7.10
11	75	87	1,180	3.93	43	191	31	1,010	7.10
12	129	55	900	5.99	44	182	19	300	7.00
13	24	93	1,730	3.50	45	37	88	1,730	3.46
14	165	31	1,150	7.41	46	103	35	780	5.66
15	94	77	1,160	4.21	47	67	85	1,300	4.82
16	96	80	1,270	5.00	48	143	78	930	5.00
17	148	30	580	5.27	49	83	85	690	4.74
18	98	69	660	5.21	50	223	33	200	8.49
19	161	43	420	6.50	51	240	19	450	6.50
20	118	47	1,080	6.12	52	312	21	280	6.50
21	269	17	290	6.19	53	12	79	4,430	1.69
22	189	35	270	5.05	54	52	83	270	3.25
23	126	58	560	6.16	55	79	43	1,340	7.17
24	12	81	4,240	1.80	56	61	88	670	3.52
25	167	29	240	4.75	57	168	28	410	6.09
26	135	65	430	4.10	58	28	95	4,370	2.86
27	107	87	3,020	6.66	59	121	41	1,310	4.88
28	72	63	1,420	7.28	60	115	62	1,470	3.89
29	128	49	420	8.12	61	186	45	300	6.90
30	27	63	19,830	5.23	62	47	85	3,630	4.10
31	152	84	420	5.79	63	178	45	220	6.09
32	224	23	530	6.50	64	142	67	560	7.20

Fuente: Gujarati, *Econometría,* quinta edición, 2009. Adaptación de la Tabla 6.6, p. 168.

a) Estime el modelo $MI_i = \beta_0 + \beta_1 TA_i + \beta_2 PIB_i + \beta_3 TF_i + u_i$, donde $u_i \sim iid(0,\ \sigma^2)$, por MCO.

b) Utilice la función matricial de Excel =ESTIMACION.LINEAL(rangoY,rangoX) para estimar el modelo del inciso anterior.

E31. Considere el MRLM $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ estimado por MCO.

a) Demuestre que $\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}$, es decir, que los residuos y los regresores no están correlacionados.

b) Demuestre que $\sum_{i=1}^{n} \hat{u}_i = 0$, es decir, que la suma de los residuos es cero.

c) Demuestre que $\hat{\mathbf{y}}'\hat{\mathbf{u}} = 0$, es decir, que los residuos y los valores ajustados no están correlacionados.

E32. Continuando con el ejercicio E30 …

a) Estime σ^2 por MCO.

b) Verifique que: $\sum_{i=1}^{n} \hat{u}_i = 0$, $\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}$ y $\mathbf{y}'\hat{\mathbf{u}} = 0$.

E33. Se sabe que $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$ y $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$ son los estimadores MCO del MRLN $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde $\mathbf{u} \sim (\mathbf{0}, \sigma^2\mathbf{I})$ y $\mathbf{X}'\mathbf{X}$ es no singular.

a) Demuestre que $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$ es estimador insesgado de $\boldsymbol{\beta}$.

b) Demuestre que $Var\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\Sigma}_{\hat{\beta}} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$.

c) Demuestre que $\hat{\sigma}^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-(k+1)}$ es estimador insesgado de σ^2.

E34. **Teorema de Gauss-Markov**. Considere el MRLM $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ donde $\mathbf{u} \sim (\mathbf{0}, \sigma^2\mathbf{I})$ y $\mathbf{X}'\mathbf{X}$ es no singular. Demuestre que el estimador MCO $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$ es MELI.

E35. Continuando con los ejercicios E30 y E32…

a) Construya la Tabla ANOVA del modelo $MI_i = \beta_0 + \beta_1 TA_i + \beta_2 PIB_i + \beta_3 TF_i + u_i$ y calcule sus respectivos coeficientes de determinación (simple y ajustado).

b) Estime por MCO el modelo $MI_i = \gamma_0 + \gamma_1 TA_i + \gamma_2 PIB_i + w_i$, $w_i \sim iid(0,\ \sigma_w^2)$. Calcule sus respectivas R^2 y $\overline{R}^2$. ¿Son comparables estos coeficientes de determinación con los del inciso anterior? ¿Qué modelo elegiría bajo este criterio?

E36. **Estimación por Máxima Verosimilitud del MRLMN**. Considere el MRLMN $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde $\mathbf{u} \sim N_n\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$, y demuestre que los EMV de $\boldsymbol{\beta}$ y σ^2 son $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{y}\right)$ y $\hat{\sigma}_*^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n}$, respectivamente.

E37. Continuando con los ejercicios E30, E32 y E35, considere el modelo con 3 variables explicativas $MI_i = \beta_0 + \beta_1 TA_i + \beta_2 PIB_i + \beta_3 TF_i + u_i$, donde $u_i \sim iid(0,\ \sigma^2)$.

a) Estime la matriz de varianzas-covarianzas de $\hat{\boldsymbol{\beta}}$.

b) Calcule intervalos de confianza individuales al 95% para cada uno de los parámetros del modelo. ¿Se puede afirmar que todas las β's son significativas?

E38. Continuando con los ejercicios E30, E32, E35 y E37, considerando el MRLM con 3 variables explicativas se sabe que:

Modelo estimado: $\hat{MI}_i = \underset{(32.892)}{168.307} - \underset{(0.248)}{1.768} TA_i - \underset{(0.002)}{0.006} PIB_i + \underset{(4.191)}{12.869} TF_i$ $\quad R^2 = 0.7474$

(errores estándar) $\quad \hat{\sigma} = 39.13$

a) Pruebe individualmente las siguientes hipótesis: $\beta_0 < 165$, $\beta_2 = 0$ y $\sigma = 40$.

b) ¿El modelo con su conjunto es significativo? Pruebe $\boldsymbol{\beta} = \mathbf{0}$ para justificar su respuesta.

c) Utilice Excel para determinar si las β's del modelo son significativas (en forma individual y de manera conjunta).

d) ¿Se pueden sustentar estadísticamente la validez de las pruebas anteriores?

E39. Continuando con los ejercicios E30, E32, E35, E37 y E38, considerando el modelo $MI_i = \beta_0 + \beta_1 TA_i + \beta_2 PIB_i + \beta_3 TF_i + u_i$, donde $u_i \sim iid(0,\ \sigma^2)$…

a) Utilice la Prueba F de Transformación Lineal General para probar la hipótesis conjunta $\beta_0 = 165$ y $\beta_3 = -7\beta_1$.

b) Utilice la Prueba F de Mínimos Cuadrados Restringidos para probar la hipótesis conjunta del inciso anterior.

E40. Continuando con los ejercicios E30, E32, E35, E37 y E38, considere una economía con un PIB *per cápita* de 7,700 dólares, una tasa de alfabetización femenina de 85% y una tasa de fecundidad de 2.2 hijos por mujer para calcular lo siguiente:

a) Intervalo de confianza al 95% para la mortalidad infantil promedio.

b) Intervalo de predicción al 95% para el nivel de mortalidad infantil.

E41. En Microeconomía la Teoría del Productor establece que la *función de producción Cobb-Douglas* en su forma econométrica está dada por $P_i = \alpha_0 L_i^{\alpha_1} K_i^{\alpha_2} e^{u_i}$, donde P_i es la producción, L_i el insumo trabajo, K_i el insumo capital y u_i es perturbación estocástica.

Se sabe que para este modelo la suma de los parámetros α_1 y α_2 proporciona información sobre los rendimientos a escala, es decir, la respuesta de la producción a un cambio proporcional en los insumos: (i) si la suma es 1, existen *rendimientos constantes a escala*, es decir, duplicar los insumos duplica la producción, triplicar los insumos triplica la producción y así sucesivamente; (ii) si la suma es menor que 1 existen *rendimientos decrecientes a escala*, es decir, al duplicar los insumos la producción crece en menos del doble; y (iii) si la suma es mayor que 1 hay *rendimientos crecientes a escala*, pues duplicar los insumos conducen a un aumento en la producción de más del doble.

La siguiente tabla muestra el *valor agregado* (P_i en miles de dólares), las *horas de trabajo* (L_i en miles de horas) y la *inversión de capital* (K_i en miles de dólares) para los 50 estados y el Distrito de Columbia (Washington, D.C.) de Estados Unidos de Norteamérica referentes al sector manufacturero durante 2005.

Entidad	**Valor agregado** (miles de dólares)	**Horas de trabajo** (miles de horas)	**Inversión de capital** (miles de dólares)	**Entidad**	**Valor agregado** (miles de dólares)	**Horas de trabajo** (miles de horas)	**Inversión de capital** (miles de dólares)
Alabama	38,372,840	424,471	2,689,076	Montana	2,644,567	24,167	334,008
Alaska	1,805,427	19,895	57,997	Nebraska	14,650,080	163,637	627,806
Arizona	23,736,129	206,893	2,308,272	Nevada	7,290,360	59,737	522,335
Arkansas	26,981,983	304,055	1,376,235	New Hampshire	9,188,322	96,106	507,488
California	217,546,032	1,809,756	13,554,116	New Jersey	51,298,516	407,076	3,295,056
Colorado	19,462,751	180,366	1,790,751	New Mexico	20,401,410	43,079	404,749
Connecticut	28,972,772	224,267	1,210,229	New York	87,756,129	727,177	4,260,353
Delaware	14,313,157	54,455	421,064	North Carolina	101,268,432	820,013	4,086,558
DC	159,921	2,029	7,188	North Dakota	3,556,025	34,723	184,700
Florida	47,289,846	471,211	2,761,281	Ohio	124,986,166	1,174,540	6,301,421
Georgia	63,015,125	659,379	3,540,475	Oklahoma	20,451,196	201,284	1,327,353
Hawaii	1,809,052	17,528	146,371	Oregon	34,808,109	257,820	1,456,683
Idaho	10,511,786	75,414	848,220	Pennsylvania	104,858,322	944,998	5,896,392
Illinois	105,324,866	963,156	5,870,409	Rhode Island	6,541,356	68,987	297,618
Indiana	90,120,459	835,083	5,832,503	South Carolina	37,668,126	400,317	2,500,071
Iowa	39,079,550	336,159	1,795,976	South Dakota	4,988,905	56,524	311,251
Kansas	22,826,760	246,144	1,595,118	Tennessee	62,828,100	582,241	4,126,465
Kentucky	38,686,340	384,484	2,503,693	Texas	172,960,157	1,120,382	11,588,283
Louisiana	69,910,555	216,149	4,726,625	Utah	15,702,637	150,030	762,671
Maine	7,856,947	82,021	415,131	Vermont	5,418,786	48,134	276,293
Maryland	21,352,966	174,855	1,729,116	Virginia	49,166,991	425,346	2,731,669
Massachusetts	46,044,292	355,701	2,706,065	Washington	46,164,427	313,279	1,945,860
Michigan	92,335,528	943,298	5,294,356	West Virginia	9,185,967	89,639	685,587
Minnesota	48,304,274	456,553	2,833,525	Wisconsin	66,964,978	694,628	3,902,823
Mississippi	17,207,903	267,806	1,212,281	Wyoming	2,979,475	15,221	361,536
Missouri	47,340,157	439,427	2,404,122				

Fuente: Gujarati, *Econometría,* quinta edición, 2009. Tabla 7.3, p. 208.

a) Exprese la función de producción Cobb-Douglas econométrica como Modelo de Regresión Lineal Múltiple. ¿Qué supuestos debe cumplir u_i?

b) Estime la función de producción Cobb-Douglas utilizando Mínimos Cuadrados Ordinarios. Calcule su respectiva R^2, R^2 ajustada y tabla ANOVA.

c) ¿Se puede afirmar que en esta economía los rendimientos a escala son constantes? Justifique estadísticamente su respuesta.

d) Si el estado de Colorado planea duplicar los insumos trabajo y capital, ¿cuál será el nivel de producción estimado? ¿Dicha estimación se ubica dentro del conjunto de datos muestrales?

E42. A continuación se presentan los datos sobre la producción de un bien (X en unidades producidas) y su costo total (Y en pesos).

Producción (X en unidades producidas)	**Costo total** (Y en pesos)
1	193
2	226
3	240
4	244
5	257
6	260
7	274
8	297
9	350
10	420

Fuente: Gujarati, *Econometría,* quinta edición, 2009. Tabla 7.4, p. 211.

a) Analice la relación entre la producción y el costo total mediante un diagrama de dispersión. ¿Es posible explicar el costo total mediante la producción utilizando un Modelo de Regresión Polinomial (MRP)? ¿De qué grado?

b) Utilice Excel para ajustar un MRP de grado 3 utilizando MCO. ¿Es el modelo significativo? Calcule su respectiva Tabla ANOVA.

c) Utilice EViews para ajustar nuevamente el modelo del inciso b). Obtenga la matriz de varianzas y covarianzas estimada de los estimadores MCO.

d) Utilice EViews para ajustar MRP's de grados 4 y 5. ¿Son estos modelos significativos? Compare sus respectivas $\overline{R}^2$. ¿Qué puede concluir?

e) Construya y grafique las bandas de predicción al 95% para la media y el nivel del costo total.

E43. Continuando con el ejercicio E41 sobre la producción manufacturera en Estados Unidos durante 2005, se sabe que la estimación directa por MCO del modelo Log-Lin de la función de producción Cobb-Douglas es $\widehat{\ln}(P_i) = 3.888 + 0.468\ln(L_i) + 0.521\ln(K_i)$.

a) Utilice el método de estimación indirecta para obtener $\hat{\beta}_1$.

b) Utilice el método de estimación indirecta para obtener $\hat{\beta}_2$.

E44. Continuando con los ejercicio E41 y E43 sobre la producción manufacturera en Estados Unidos durante 2005, considerando el modelo Log-Lin de la función de producción Cobb-Douglas $\ln(P_i) = \beta_0 + \beta_1 \ln(L_i) + \beta_2 \ln(K_i) + u_i$, $u_i \sim iid(0,\ \sigma^2)$…

a) Calcule los coeficientes de correlación parciales a partir de la definición (utilizando los residuos de los MRLS) y compárelos vs. sus respectivos coeficientes de correlación simples.

b) Calcule nuevamente los coeficientes de correlación parcial pero ahora a partir de los coeficientes de correlación simples.

c) Calcule el coeficiente de determinación a partir de los coeficientes de correlación simples y parciales. Reporte e interprete el coeficiente de correlación múltiple.

d) Construya la Tabla ANOVA de contribución incremental y pruebe la significancia incremental de los insumos trabajo y capital. ¿Qué puede concluir?

E45. De la teoría microeconómica, se sabe que la demanda de un bien suele depender del ingreso real del consumidor, del precio real del bien y de los precios reales de los bienes complementarios o que compiten con él (bienes sustitutos). A continuación se muestran datos sobre la demanda de carne de pollo en Estados Unidos de 1960 a 1982:

Año	**Consumo *per cápita* de carne de pollo** (Z en libras)	**Ingreso *per cápita* real disponible** (W_1 en dólares)	**Precio real de la carne de pollo** (W_2 en centavos de dólar por libra)	**Precio real de la carne de res** (W_3 en centavos de dólar por libra)	**Precio real de la carne de puerco** (W_4 en centavos de dólar por libra)	**Precio real de sustitutos del pollo *** (W_5 en centavos de dólar por libra)
1960	27.8	397.5	42.2	78.3	50.7	65.8
1961	29.9	413.3	38.1	79.2	52.0	66.9
1962	29.8	439.2	40.3	79.2	54.0	67.8
1963	30.8	459.7	39.5	79.2	55.3	69.6
1964	31.2	492.9	37.3	77.4	54.7	68.7
1965	33.3	528.6	38.1	80.2	63.7	73.6
1966	35.6	560.3	39.3	80.4	69.8	76.3
1967	36.4	624.6	37.8	83.9	65.9	77.2
1968	36.7	666.4	38.4	85.5	64.5	78.1
1969	38.4	717.8	40.1	93.7	70.0	84.7
1970	40.4	768.2	38.6	106.1	73.2	93.3
1971	40.3	843.3	39.8	104.8	67.8	89.7
1972	41.8	911.6	39.7	114.0	79.1	100.7
1973	40.4	931.1	52.1	124.1	95.4	113.5
1974	40.7	1,021.5	48.9	127.6	94.2	115.3
1975	40.1	1,165.9	58.3	142.9	123.5	136.7
1976	42.7	1,349.6	57.9	143.6	129.9	139.2
1977	44.1	1,449.4	56.5	139.2	117.6	132.0
1978	46.7	1,575.5	63.7	165.5	130.9	132.1
1979	50.6	1,759.1	61.6	203.3	129.8	154.4
1980	50.1	1,994.2	58.9	219.6	128.0	174.9
1981	51.7	2,258.1	66.4	221.6	141.0	180.8
1982	52.9	2,478.7	70.4	232.6	168.2	189.4

* Promedio ponderado de los precios reales al menudeo por libra de carne de puerco y res. Ponderadores basados en el consumo relativo.

Fuente: Gujarati, D., Porter, D. (2009). *Econometría*. 5a edición, p. 221. Se modificó el orden de los regresores.

A partir de esta información analice la forma funcional que mejor convenga para construir un MRLM aplicando el método ascendente (*stepwise regression*) respondiendo a las siguientes preguntas:

a) Calcule el diagrama de dispersión múltiple de estas variables. ¿Es razonable pensar que la forma funcional entre Z y los regresores es lineal?

b) Calcule el diagrama de dispersión múltiple del logaritmo natural de todas las variables. ¿Es razonable pensar que la forma funcional entre $Y = \ln(Z)$ y el logaritmo natural del resto de los regresores, $X_j = \ln(W_j)$, j = 1, 2, 3, 4 y 5, es lineal? Especifique el modelo econométrico en su forma lineal y no lineal considerando todos los regresores potenciales.

c) Calcule la matriz de correlaciones considerando los logaritmos de todas las variables. ¿Cuál es el regresor más correlacionado con $Y = \ln(Z)$? Interprete su resultado.

d) Considere el MRLS incluyendo sólo a X_1 y estímelo por MCO. ¿Es significativo el modelo? ¿Qué puede concluir?

e) Calcule los coeficientes de correlación parcial entre Y y X_j, j = 2, 3, 4 y 5, aislando el efecto de X_1. ¿Qué regresor está más correlacionado con Y aislando el efecto de X_1? Interprete su resultado y calcule su respectivo coeficiente de determinación parcial.

f) Considere el MRLM incluyendo sólo X_1 y X_2. Sin necesidad de estimar aún el modelo, construya la Tabla ANOVA Incremental y aplique la Prueba F (incremental) para determinar si la inclusión de X_2 resulta significativa. ¿Qué puede concluir?

g) Estime por MCO el MRLM incluyendo sólo X_1 y X_2. Calcule su respectiva Tabla ANOVA. ¿Qué puede concluir? ¿Estos resultados son los esperados?

h) Calcule los coeficientes de correlación parcial de segundo orden entre Y y X_j, j = 3, 4 y 5, aislando el efecto de X_1 y X_2. ¿Qué regresor está más correlacionado con Y aislando el efecto de X_1 y X_2?

i) Considere el MRLM incluyendo sólo X_1, X_2 y X_4. ¿Es significativa la contribución de X_4 al modelo? Aplique la Prueba F (ANOVA incremental) para sustentar su respuesta.

j) Estime por MCO el MRLM incluyendo sólo X_1, X_2 y X_4. ¿Qué puede concluir? ¿Estos resultados son los esperados?

k) ¿Cuál es el modelo final estimado que arroja el método ascendente (*stepwise regression*)?

l) Compare los criterios de información de Akaike, Schwarz y Hannan-Quinn de los modelos de los incisos g) y j). ¿Qué puede concluir al respecto? ¿Esta conclusión es la misma que la del criterio de la R^2 ajustada?

m) Verifique que la Prueba F (ANOVA incremental) del inciso i) se puede obtener a partir de las sumas de cuadrados de residuos del modelo con los regresores X_1 y X_2 (SCR_{12}) y del modelo con los regresores X_1, X_2 y X_4 (SCR_{124}) mediante el estadístico de prueba $F = \frac{[n-(k+1)](SCR_{12} - SCR_{124})}{SCR_{124}} \sim F(1, n-(k+1))$.

E46. Considere nuevamente la información sobre la demanda de carne de pollo en Estados Unidos de 1960 a 1982 del ejercicio E45. Para estos datos se sabe que la forma funcional que permite formular un MRLM es el modelo *Log-Lin*, es decir, considerar el logaritmo natural de cada variable. Construya un MRLM aplicando el método descendente (*backward elimination*) respondiendo a las siguientes preguntas:

a) Estime el MRLM considerando todos los posibles regresores. ¿El modelo resulta significativo?

b) Analice la significancia incremental al 1% de cada uno de los 5 regresores (X_1, X_2, X_3, X_4 y X_5) estimando por MCO los MRLM de 4 regresores. ¿Cuál es el regresor menos significativo? ¿Es necesario eliminarlo?

c) Analice la significancia incremental al 1% de cada uno de los 4 regresores (X_1, X_2, X_4 y X_5) estimando por MCO los MRLM de 3 regresores. ¿Cuál es el regresor menos significativo? ¿Es necesario eliminarlo?

d) Analice la significancia incremental al 1% de cada uno de los 3 regresores (X_1, X_2 y X_4) estimando por MCO los MRLM de 2 regresores. ¿Cuál es el regresor menos significativo? ¿Es necesario eliminarlo?

e) Analice la significancia incremental al 1% de cada uno de los 2 regresores (X_1 y X_2) estimando por MCO los MRLS posibles. ¿Cuál es el regresor menos significativo? ¿Es necesario eliminarlo?

f) ¿Cuál es el modelo final estimado que arroja el método descendente (*backward elimination*)?

E47. A continuación se presenta información sobre salarios y educación para una muestra de 114 empleados en una ciudad industrial del sur de India en 1990.

Género	Edu. Máx. (1)	Tipo Empleo (2)	Salario semanal (Rupias)	Género	Edu. Máx. (1)	Tipo empleo (2)	Salario semanal (Rupias)	Género	Edu. Máx. (1)	Tipo empleo (2)	Salario semanal (Rupias)
Mujer	0	0	120.0	Mujer	0	0	100.0	Mujer	0	0	129.5
Mujer	3	1	224.0	Mujer	0	0	105.0	Mujer	1	0	112.0
Mujer	0	0	132.0	Mujer	2	1	300.0	Mujer	1	0	82.0
Mujer	2	0	75.0	Hombre	2	1	115.3	Mujer	2	1	385.0
Hombre	2	0	111.0	Hombre	3	1	102.5	Hombre	3	1	94.3
Mujer	2	0	127.0	Mujer	1	0	189.5	Mujer	0	1	350.0
Mujer	0	0	30.0	Mujer	2	0	62.5	Mujer	0	0	107.7
Mujer	0	0	24.0	Mujer	1	0	50.0	Mujer	0	0	20.0
Mujer	0	1	119.0	Hombre	3	1	272.5	Hombre	0	0	53.8
Mujer	0	0	75.0	Mujer	2	1	175.0	Mujer	0	1	427.0
Mujer	2	0	324.0	Mujer	1	1	117.0	Mujer	0	0	18.0
Mujer	0	0	42.0	Mujer	3	0	950.0	Mujer	0	0	120.0
Mujer	0	0	100.0	Mujer	0	0	100.0	Mujer	0	0	40.5
Mujer	0	0	135.5	Mujer	0	0	140.0	Mujer	1	1	375.0
Mujer	0	0	107.0	Mujer	2	0	97.0	Mujer	0	0	120.0
Hombre	1	0	50.0	Mujer	0	0	150.0	Mujer	1	1	175.0
Mujer	0	0	90.0	Hombre	0	0	25.0	Hombre	0	0	50.0
Mujer	0	1	377.0	Hombre	0	0	15.0	Mujer	1	1	100.0
Mujer	2	0	150.0	Mujer	0	0	131.0	Hombre	0	1	25.0
Mujer	0	0	162.0	Mujer	0	0	120.0	Mujer	0	1	40.0
Mujer	1	0	18.0	Hombre	0	0	25.0	Mujer	0	1	65.0
Mujer	1	0	127.6	Hombre	0	0	25.0	Hombre	0	1	47.5
Hombre	0	0	47.5	Hombre	0	1	30.0	Mujer	0	1	162.5
Mujer	2	0	135.0	Hombre	0	1	30.0	Hombre	0	1	175.0
Mujer	0	1	400.0	Mujer	0	0	122.0	Hombre	0	1	150.0
Mujer	3	1	91.8	Mujer	2	1	287.5	Mujer	0	1	162.5
Mujer	0	1	139.5	Hombre	0	0	75.0	Mujer	1	1	162.5
Mujer	0	0	49.2	Mujer	0	0	79.0	Hombre	0	1	50.0
Mujer	1	0	30.0	Hombre	1	0	85.3	Mujer	2	1	395.0
Hombre	0	0	40.5	Mujer	2	1	350.0	Hombre	0	1	175.0
Mujer	0	0	81.0	Mujer	0	1	54.0	Mujer	1	0	87.5
Mujer	0	0	105.0	Mujer	0	0	110.0	Mujer	0	0	75.0
Mujer	0	0	200.0	Mujer	0	1	342.0	Mujer	0	1	162.5
Mujer	0	1	139.5	Mujer	0	1	77.5	Mujer	0	1	325.0
Mujer	0	0	80.0	Mujer	0	0	370.0	Mujer	2	0	120.5
Hombre	0	0	47.0	Mujer	0	1	156.0	Mujer	1	0	600.0
Mujer	0	0	125.0	Mujer	0	0	261.2	Mujer	0	0	52.0
Mujer	0	0	500.0	Mujer	2	0	54.0	Mujer	1	0	116.5

Notas: (1) Educación máxima: 0 = Sin primaria, 1 = Hasta primaria, 2 = Hasta secundaria, 3 = Superior a segundaria
(2) 0 = Temporal, 1 = Permanente

Fuente: Gujarati D., Poreter, D. (2009). Econometría, 5a edición. Adaptación de la Tabla 9.7

a) Formule un Modelo ANOVA para explicar el logaritmo natural del sueldo a través de las variables "género" y "tipo de empleo". Estímelo por MCO y analice su significancia estadística.

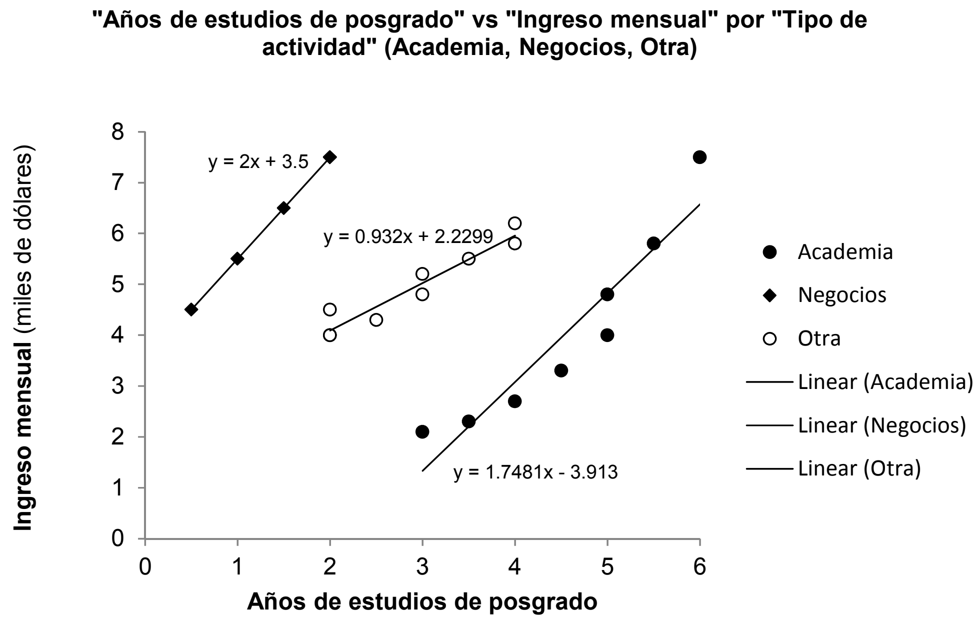
b) Construya una tabla que muestre el ingreso estimado para cada una de los grupos identificados a partir del Modelo ANOVA del inciso anterior.

c) Bajo el supuesto de normalidad, ¿es razonable pensar que el ingreso de las mujeres con empleo permanente es igual al de los hombres con empleo permanente?

d) Formule el Modelo ANOVA para explicar el logaritmo natural del sueldo considerando las 3 variables cualitativas con efectos aditivos. ¿Es significativo?

e) Formule el Modelo ANOVA para explicar el logaritmo natural del sueldo considerando las 3 variables cualitativas con efectos aditivos y multiplicativos. ¿Es significativo? ¿Qué puede concluir?

E48. En el ejercicio E2 se analizó una muestra aleatoria de 22 personas con estudios de posgrado para validar la teoría de algunos Economistas que afirman que un mayor "ingreso mensual" (Y) está asociado a un mayor número de "años de estudios de posgrado" (X). A continuación se presenta uno de los resultados de dicho análisis



a) Formule un Modelo ANCOVA que permita explicar el "ingreso mensual" mediante los "años de estudio de posgrado" y el "tipo de actividad" al que se dedican las personas.

b) Estime el modelo del inciso anterior por MCO. ¿Son coincidentes los resultados frente a los de las regresiones mostradas en la gráfica?

E49. A continuación se muestra el consumo mensual de energía eléctrica de uso doméstico (Y_t, en miles de millones de watts/hora) en México de 2005 a 2012:

Mes	Año							
	2005	2006	2007	2008	2009	2010	2011	2012
Enero	2,599	2,595	2,760	2,837	2,861	2,825	2,998	3,049
Febrero	2,530	2,523	2,732	2,852	2,833	2,769	3,062	2,999
Marzo	2,389	2,430	2,556	2,661	2,686	2,655	2,907	2,990
Abril	2,473	2,507	2,563	2,697	2,798	2,735	3,012	2,983
Mayo	2,594	2,760	2,779	2,952	3,123	2,996	3,349	3,068
Junio	2,743	2,873	2,911	3,121	3,232	3,282	3,537	3,279
Julio	2,960	2,966	3,054	3,205	3,377	3,594	3,688	3,440
Agosto	2,986	3,013	3,201	3,276	3,545	3,544	3,632	3,636
Septiembre	2,996	3,059	3,198	3,332	3,558	3,633	3,792	3,670
Octubre	2,989	3,033	3,119	3,183	3,275	3,422	3,608	3,835
Noviembre	2,788	2,882	2,959	3,012	3,079	3,220	3,274	3,605
Diciembre	2,609	2,774	2,797	2,813	2,792	2,859	3,049	3,436

Fuente: www.inegi.gob.mx

a) Si M_j, j = 1, 2,…, 11, son las variables dicótomas que permiten modelar la estacionalidad mensual que presenta la serie de tiempo Y_t, estime por MCO el modelo $Y_t = \beta_0 + \sum_{j=1}^{11} \beta_j M_{jt} + \beta_{12} t + u_i$, $u_i \sim iid(0,\ \sigma^2)$, es decir, una tendencia lineal con estacionalidad mensual.

b) Calcule el pronóstico puntual del nivel de consumo mensual de energía eléctrica de uso doméstico en México para 2013.

c) Construya una gráfica que muestre las series de tiempo real y ajustada en el período enero 2005 – diciembre 2012, y el pronóstico sobre el periodo enero – diciembre 2013. ¿Qué puede concluir?

E50. Continuando con el ejercicio E49…

a) Calcule la serie de tiempo desestacionalizada del consumo mensual de energía eléctrica de uso doméstico en México en el período enero 2005 – diciembre 2012 y grafíquela junto con la serie de tiempo original.

b) Estime por MCO una tendencia lineal (usando al tiempo como regresor) para explicar la serie de tiempo desestacionalizada del inciso anterior.

c) Calcule el pronóstico puntual del nivel de consumo mensual de energía eléctrica de uso doméstico en México para 2013 partiendo de la tendencia lineal del inciso anterior y agregando los factores estacionales del inciso a).¿Cómo se comparan estos pronósticos contra los obtenido en el ejercicio E49?

d) Se sabe que para enero, febrero y marzo de 2013 el consumo mensual de energía eléctrica de uso doméstico en México fue de 3,126, 3,120 y 2,905 miles de millones de watt/hora, respectivamente. Compare los pronósticos del inciso anterior frente a los del inciso b) del ejercicio E49 y ambos pronósticos frente a los valores reales observados.

E51. El sistema tributario de cierto país establece que las personas físicas que perciben más de 200,000 dólares por año deben pagar un impuesto sobre la renta (ISR) de 35%, mientras que las personas que ganan 200,000 dólares o menos deben pagar un ISR del 25%. El monto de deducciones varía de persona en persona, de modo que para una muestra de 60 personas físicas se identificaron los siguientes ingresos anuales y deducciones fiscales:

Ingreso	Deducciones	Ingreso	Deducciones	Ingreso	Deducciones
159	18.6	135	21.5	79	12.8
287	33.0	252	52.9	167	21.8
129	13.5	179	18.4	192	17.6
222	54.9	303	49.4	134	11.2
243	61.3	179	27.8	251	49.5
186	25.9	215	62.3	247	62.3
254	33.4	138	19.2	181	28.7
288	47.5	142	11.4	143	3.6
214	54.0	265	67.5	159	15.8
205	41.1	288	54.1	233	49.5
152	20.5	165	22.2	95	5.5
277	46.1	342	38.4	149	4.2
195	28.7	100	7.6	126	10.7
243	39.9	172	14.2	196	24.0
226	61.6	139	10.2	127	8.4
187	21.2	139	10.2	122	5.9
224	40.7	180	11.3	202	43.2
190	24.7	231	36.4	172	31.0
165	24.7	255	57.2	186	15.6
130	4.2	135	10.1	123	11.5

Fuente: Ilustrativo

a) Construya la gráfica del impuesto por pagar (ISR menos deducciones) explicado por el ingreso de las personas de esta muestra.

b) Formule el respectivo Modelo de Regresión por Segmentos, estímelo por MCO y grafique las rectas ajustadas. ¿Es relevante mantener el intercepto?

E52. Continuando con el ejercicio E51, determine si es posible validar el supuesto de normalidad en los errores del Modelo de Regresión por Segmentos sin intercepto mediante…

a) La Gráfica Cuantil-Cuantil.

b) El Histograma y la Prueba Jarque-Bera.

c) Las Pruebas Lillieforce, Cramér-Von Mises, Watson y Anderson-Darling.

d) ¿Son coincidentes todos los criterios?

e) Utilice la Transformación Potencia sobre el modelo con intercepto considerando los valores $\lambda = -1, 0, 0.5$ y 1. ¿Es posible validar el supuesto de normalidad en los modelos transformados?

E53. Para comparar la exactitud de las estimaciones MCO de distintos paquetes computacionales, James W. Longley (1967) utilizó datos de Estados Unidos de 1947 a 1962 para explicar el número de personas con trabajo (Y en miles), mediante el índice de deflación implícito (X_1), el PIB (X_2 en millones de dólares), el número de desempleados (X_3 en miles), el número de reclutas en las fuerzas armadas (X_4), la población civil mayor de 14 años de edad (X_5) y el tiempo (X_6). Además, con motivo de la Guerra de Corea en 1951, el número de reclutas se incrementó de manera importante a partir de ese año, por lo que se incluyeron las variables dicótomas X_7 y X_8.

Año	Número de personas con trabajo (Y, miles)	Índice de deflación (X_1)	PIB (X_2, mill. dólares)	Número de desempleados (X_3, miles)	Número de reclutas (X_4)	Población civil mayor de 14 años de edad (X_5)	Tiempo (X_6)	Pre-Guerra de Corea (X_7)	Pos-Guerra de Corea (X_8)
1947	60,323	830	234,289	2,356	1,590	107,608	1	1	0
1948	61,122	885	259,426	2,325	1,456	108,632	2	1	0
1949	60,171	882	258,054	3,682	1,616	109,773	3	1	0
1950	61,187	895	284,599	3,351	1,650	110,929	4	1	0
1951	63,221	962	328,975	2,099	3,099	112,075	5	0	1
1952	63,639	981	346,999	1,932	3,594	113,270	6	0	1
1953	64,989	990	365,385	1,870	3,547	115,094	7	0	1
1954	63,761	1,000	363,112	3,578	3,350	116,219	8	0	1
1955	66,019	1,012	397,469	2,904	3,048	117,388	9	0	1
1956	67,857	1,046	419,180	2,822	2,857	118,734	10	0	1
1957	68,169	1,084	442,769	2,936	2,798	120,445	11	0	1
1958	66,513	1,108	444,546	4,681	2,637	121,950	12	0	1
1959	68,655	1,126	482,704	3,813	2,552	123,366	13	0	1
1960	69,564	1,142	502,601	3,931	2,514	125,368	14	0	1
1961	69,331	1,157	518,173	4,806	2,572	127,852	15	0	1
1962	70,551	1,169	554,894	4,007	2,827	130,081	16	0	1

Fuente: Gujarati, D., Porter, D. (2009). Econometría. 5a Ed. Adaptación de la Tabla 10.8, p. 348

a) Utilice EViews y Excel para estimar por MCO el MRLM con intercepto para explicar el número de personas con trabajo (Y), mediante $X_1, X_2, \ldots, X_8$ de la tabla anterior. ¿Qué puede concluir?

b) Estime nuevamente por MCO el MRLM con intercepto para explicar el número de personas con trabajo (Y), sin considerar las variables dicótomas X_7 y X_8. ¿Hay indicios de multicolinealidad?

c) Calcule la matriz de correlaciones de los regresores $X_1, X_2, \ldots, X_6$. ¿Se verifica la presencia de multicolinealidad?

d) Utilice la Regla de Klein para determinar si la multicolinealidad imperfecta es “grave”.

e) Gujarati y Porter sugieren que: (i) en lugar de incorporar el PIB nominal y el índice de deflación por separado, es posible considerar el PIB real considerando el regresor a $\dfrac{X_2}{X_1}$; (ii) la población civil de más de 14 años de edad (X_5) crece con el tiempo y está muy correlacionada con el tiempo (X_6), por lo que X_6 se puede eliminar; y (iii) no hay razón de peso para explicar el número de personas con trabajo (Y) mediante el número de personas desempleadas (X_3), por lo que X_3 también se puede eliminar. Corrija la multicolinealidad imperfecta del modelo estimado en b) incorporando esta información *a priori*.

f) Tomando en cuenta la "gravedad" de la multicolinealidad detectada por la Regla de Klein, corrija la multicolinealidad imperfecta del modelo estimado en b) considerando alguna transformación de razón que tenga interpretación.

g) Estime por MCO el MRLM considerando los regresores X_1, X_2,…, X_7 y corrija la multicolinealidad imperfecta eliminando variables mediante el método ascendente con significancia del 10% (*stepwise regression*). ¿Cuál de los modelos corregido?

h) ¿Cuál de los modelos libres de multicolinealidad e), f) o g) es el preferido? Justifique su respuesta.

E54. A continuación se muestra el gasto diario (G, en dólares) y el ingreso diario (Y, en dólares) para una muestra de 30 personas de cierto país.

Gasto diario (G, dólares)	**Ingreso diario** (Y, dólares)	**Gasto diario** (G, dólares)	**Ingreso diario** (Y, dólares)
55	80	115	180
65	100	140	225
70	85	120	200
80	110	145	240
79	120	130	185
84	115	152	220
98	130	144	210
95	140	175	245
90	125	180	260
75	90	135	190
74	105	140	205
110	160	178	265
113	150	191	270
125	165	137	230
108	145	189	250

Fuente: Gujarati, D., Porter, D. (2009). Econometría. 5a Ed. Tabla 11.3, p. 384

a) Analice gráficamente la relación entre estas variables, estime por MCO el MRLS con intercepto para explicar el gasto diario mediante el ingreso diario.

b) ¿Se cumple el supuesto de normalidad en los errores del MRLS del inciso a)?

c) Analice gráficamente si se cumple el supuesto de varianza constante en el MRLS estimado en a).

d) Utilice la Prueba Breusch-Pagan-Godfrey para determinar si los residuos del MRLS estimado en a) son heterocedásticos.

e) Utilice la Prueba de Glejser para determinar si los residuos del MRLS estimado en a) son heterocedásticos.

f) Utilice la Prueba de White (con interacciones) para determinar si los residuos del MRLS estimado en a) son heterocedásticos. ¿Cambia su conclusión si se aplica la prueba sin interacciones?

E55. Continuando con el ejercicio E54…

a) Realice las pruebas de igualdad de varianzas Barlett, Levene y Brown-Forsythe considerando 5 grupos. ¿Se confirma la heterocedasticidad de los errores del MRLS $G_i = \beta_0 + \beta_1 Y_i + u_i$?

b) Realice las pruebas de igualdad de varianzas F, Barlett, Levene y Brown-Forsythe considerando 2 grupos. ¿Se obtienen las mismas conclusiones que en el inciso anterior?

E56. A continuación se presentan los salarios anuales (en dólares) de los empleados de 10 sectores de la industria manufacturera de bienes no duraderos, de acuerdo con el tamaño de las empresas del sector (plantilla laboral en número de empleados) en Estados Unidos durante 1958. En las últimas 3 líneas se incluyen media y desviación estándar de los salarios por tamaño de la industria, y la productividad promedio (aumento de los rendimientos en función del trabajo necesario para el producto final, en dólares).

Sector	Tamaño de la plantilla laboral (número de empleados)								
	1 a 4	5 a 9	10 a 19	20 a 49	50 a 99	100 a 249	250 a 499	500 a 999	1,000 a 2,499
Alimentos y similares	2,994	3,295	3,565	3,907	4,189	4,486	4,676	4,968	5,342
Productos del tabaco	1,721	2,057	3,336	3,320	2,980	2,848	3,072	2,969	3,822
Productos textiles	3,600	3,657	3,674	3,437	3,340	3,334	3,225	3,163	3,168
Ropa y productos relacionados	3,494	3,787	3,533	3,215	3,030	2,834	2,750	2,967	3,453
Papel y similares	3,498	3,847	3,913	4,135	4,445	4,885	5,132	5,342	5,326
Impresión y publicación	3,611	4,206	4,695	5,083	5,301	5,269	5,182	5,395	5,552
Productos Químicos y similares	3,875	4,660	4,930	5,005	5,114	5,248	5,630	5,870	5,876
Productos petroleros y carboníferos	4,616	5,181	5,317	5,337	5,421	5,710	6,316	6,455	6,347
Productos de caucho y plásticos	3,538	3,984	4,014	4,287	4,221	4,539	4,721	4,905	5,481
Cuero y productos de cuero	3,016	3,196	3,149	3,317	3,414	3,254	3,177	3,346	4,067
Remuneración promedio	3,396.3	3,787.0	4,012.6	4,104.3	4,145.5	4,240.7	4,388.1	4,538.0	4,843.4
Desviación estándar	742.2	851.4	727.8	805.1	930.0	1,080.6	1,241.5	1,307.7	1,110.7
Productividad promedio	9,355.0	8,584.0	7,962.0	8,275.0	8,389.0	9,418.0	9,795.0	10,281.0	11,750.0

Fuente: Gujarati, D., Porter, D. (2009). Econometría. 5a Ed. Tabla 11.1, p. 369

a) Suponga que se quiere explicar la remuneración promedio (Y) mediante el tamaño de las empresas (X) a través de un MRLS. Por simplicidad considere que X toma los valores 1 para empresas de 1 a 4 empleados, 2 para empresas de 5 a 9 empleados,…., y 9 para empresas de 1,000 a 2,499 empleados. ¿Se puede afirmar que la varianza de los errores es conocida y que es heterocedástica?

b) Estime el modelo $Y_i = \beta_0 + \beta_1 X_i + u_i$, $u_i \sim \left(0, \sigma_i^2\right)$ por MCO y analice gráficamente si los errores son heterocedásticos.

c) Estime el modelo $Y_i = \beta_0 + \beta_1 X_i + u_i$, $u_i \sim \left(0, \sigma_i^2\right)$ por MCP. ¿Se esperan resultados muy distintos a los de MCO?

E57. Continuando con los ejercicios E54 y E55…

a) Corrija la heterocedasticidad del modelo utilizando el procedimiento de White. Compare las estimaciones puntuales de los parámetros, errores estándar y matriz de varianzas y covarianzas frente a los resultados de MCO.

b) ¿Es razonable pensar que $\sigma_i^2 = \sigma^2 Y_i$? En caso afirmativo corrija la heterocedasticidad del modelo utilizando la transformación estabilizadora de varianza respectiva.

E58. A continuación se presentan datos sobre el consumo, ingreso, riqueza y tasa de interés (C, Y, R, I, respectivamente y en dólares reales de 1996) para la economía de en Estados Unidos de 1947 a 2000.

Año	C	Y	R	I
1947	976.4	1,035.2	5,166.8	-10.4
1948	998.1	1,090.0	5,280.8	-4.7
1949	1,025.3	1,095.6	5,607.4	1.0
1950	1,090.9	1,192.7	5,759.5	0.4
1951	1,107.1	1,227.0	6,086.1	-5.3
1952	1,142.4	1,266.8	6,243.9	-0.3
1953	1,197.2	1,327.5	6,355.6	0.6
1954	1,221.9	1,344.0	6,797.0	-0.1
1955	1,310.4	1,433.8	7,172.2	0.3
1956	1,348.8	1,502.3	7,375.2	-0.7
1957	1,381.8	1,539.5	7,315.3	-0.3
1958	1,393.0	1,553.7	7,870.0	-0.6
1959	1,470.7	1,623.8	8,188.1	2.3
1960	1,510.8	1,664.8	8,351.8	1.5
1961	1,541.2	1,720.0	8,971.9	1.3
1962	1,617.3	1,803.5	9,091.5	1.4
1963	1,684.0	1,871.5	9,436.1	2.1
1964	1,784.8	2,006.9	10,003.4	2.0
1965	1,897.6	2,131.0	10,562.8	2.1
1966	2,006.1	2,244.6	10,522.0	2.0
1967	2,066.2	2,340.5	11,312.1	1.2
1968	2,184.2	2,448.2	12,145.4	1.1
1969	2,264.8	2,524.3	11,672.3	1.7
1970	2,317.5	2,630.0	11,650.0	1.2
1971	2,405.2	2,745.3	12,312.9	-0.7
1972	2,550.5	2,874.3	13,499.9	-0.2
1973	2,675.9	3,072.3	13,081.0	1.4

Año	C	Y	R	I
1974	2,653.7	3,051.9	11,868.8	-1.0
1975	2,710.9	3,108.5	12,634.4	-3.5
1976	2,868.9	3,243.5	13,456.8	-0.7
1977	2,992.1	3,360.7	13,786.3	-1.2
1978	3,124.7	3,527.5	14,450.5	0.1
1979	3,203.2	3,628.6	15,340.0	1.7
1980	3,193.0	3,658.0	15,965.0	2.3
1981	3,236.0	3,741.1	15,965.0	4.7
1982	3,275.5	3,791.7	16,312.5	4.4
1983	3,454.3	3,906.9	16,944.8	4.7
1984	3,640.6	4,207.6	17,526.7	5.8
1985	3,820.9	4,347.8	19,068.3	4.3
1986	3,981.2	4,486.6	20,530.0	3.8
1987	4,113.4	4,582.5	21,235.7	2.8
1988	4,279.5	4,784.1	22,332.0	3.3
1989	4,393.7	4,906.5	23,659.8	4.3
1990	4,474.5	5,014.2	23,105.1	3.6
1991	4,466.6	5,033.0	24,050.2	1.8
1992	4,594.5	5,189.3	24,418.2	1.0
1993	4,748.9	5,261.3	25,092.3	0.6
1994	4,928.1	5,397.2	25,218.6	2.2
1995	5,075.6	5,539.1	27,439.7	3.3
1996	5,237.5	5,677.7	29,448.2	3.1
1997	5,423.9	5,854.5	32,664.1	3.1
1998	5,683.7	6,168.6	35,587.0	3.6
1999	5,968.4	6,320.0	39,591.3	3.2
2000	6,257.8	6,539.2	38,167.7	3.6

Considere el MRLM: $\ln(C_t) = \beta_0 + \beta_1 \ln(Y_t) + \beta_2 \ln(R_t) + \beta_3 I_t + u_t$ ….

a) Estime el modelo por MCO y analice gráficamente sus residuos para determinar si existe autocorrelación de primer orden.

b) Aplique la Prueba Durbin-Watson par adeterminar si se confirma la existencia de autocorrelación de primer orden.

c) Corrija la autocorrelación estimando el coeficiente de autocovarianza incorporando el proceso autorregresivo del error.

d) Estime el modelo por MCO pero corrija la matriz de varianzas-covarianzas mediante el Método Newey-West y reporte los errores estándar CHA.