Art of Stat Mobile Apps for Intro Stats

Bernhard Klingenberg Prof. of Statistics Williams College

 \odot

.... 🗢 🗖

ART OF STAT **Explore** Data

Categorical Variables

11:42



One Categorical Variable

Compare Groups on Categorical Variable



Relationship Between Two Categorical Variables

Quantitative Variables



One Quantitative Variable

Compare Groups on

Overview

Conference Theme: "Preparing the Modern Student"

The modern student deserves a modern calculator!

VS.

	* Capture 41 X
NORMAL FLOAT AUTO &+bi DEGREE MP	NORMAL FLOAT AUTO a+bi DEGREE MP
DISTR DRAW 1:normalpdf(2:normalcdf(3:invNorm(4:invT(5:tpdf(6:tcdf(7:X ² pdf(8:X ² cdf(9!Fpdf(CATALOG HELP invNorm(■ (area[,μ,σ]) (area[,μ,σ,tail]) (area[,tail]) LEFT tail:[catalo9] CENTER RIGHT ↓
	PRSTEI ESC
*Capture 42 ×	*Capture 43 ×
NORMAL FLOAT AUTO a+bi DEGREE MP 👖	NORMAL FLOAT AUTO a+bi DEGREE MP

📶 Verizon 奈	11:23	AM	66% 🔲
く тh	ne Normal	Distributio	n (?)
0		1	
Normal Dist P(-1.555 ≤	tribution w X ≤ 1.555)	ith μ = 0, σ = 0.8800	= 1
	88	×=1	.555
-3 -2	-1 C		2 3
Percentage:		pe of Percentile	
88	% C	entral: P(X	≤ x)
Percentile:			
Lower Quar Upper Quar	ntile = -1.5 ntile = 1.55	55 5	
	FIND ROBABILITY	FIND	SIMULATE

8:21	"∥ ≎ ∎
<u>Art o</u> Distrib	F STAT
Continuous Distributio	ns:
Normal Distribution	Student t-Distribution
Normal	t-Distribution
Chi-Squared Distribution	F-Distribution
Chi-Squared	F-Distribution
Exponential Distribution	Uniform Distribution
Discrete Distributions:	
D :	

Overview

Apps span four major themes:



Available for iOS and Android (search for "Art of Stat")





Overview: Explore Data

Art of Stat **Explore** Data

8:22 TestFlight

ART OF STAT **Explore** Data

 \odot

Categorical Variables





Compare Groups on Categorical Variable



Relationship Between Two Categorical Variables

Quantitative Variables



One Quantitative Variable

Compare Groups on

Explore Categorical Variables







Relationship Btw. Two Variables

Explore Quantitative Variables



Distribution of One Variable



Comparing Groups



Relationship Btw. Two Variables

Overview: Distributions



4:53 🖪 🗙 🗟 🛇 52% 🗖 Chi-Squared Distribution **F-Distribution** 10 15 Chi-Squared **F-Distribution Exponential Distribution Uniform Distribution** 4 -0.5 0.0 0.5 1.0 1.5 3 2 Exponential Uniform **Discrete Distributions: Binomial Distribution Geometric Distribution** Number of Successes Number of Trials Binomial Geometric



Continuous Distributions







Discrete Distributions







Overview: Inference

Art of Stat Inference

100%

ART OF STAT Inference

 \odot

Inference About Proportions

8:48



For Population Proportions











Chi-Square Test

50%

• A • B • C

0%

Overview: Inference

Inference in Linear Regression



Illustrating Concepts: Coverage and Errors & Power









8:49

One-Way ANOVA

Inference for Linear Regression Model



Inference in Linear Regression

Illustrating Concepts



''II 🕹 🔲

Overview: Regression

Simple Regression



- Multiple Regression (coming fall '22)
 - Multiple Linear Regression
 - Multiple Logistic Regression



40k

Income

60k

Linear Regression





Let's try it with proportion inference

...

Elon Musk 🕗 @elonmusk · May 13 ... Twitter deal temporarily on hold pending details supporting calculation that spam/fake accounts do indeed represent less than 5% of users

To find out, my team will do a random sample of 100 followers of @twitter.

I invite others to repeat the same process and see what they discover ...

Pranay Pathole @PPathole · May 13 Replying to @elonmusk and @Twitter

Elon, can you elaborate a bit on the "process"? So that we as Twitter can help you in finding out the real percentage of scam/spam/bot accounts

Elon Musk 🕗 @elonmusk · May 13 ... Any sensible random sampling process is fine. If many people independently get similar results for % of fake/spam/duplicate accounts, that will be telling.

I picked 100 as the sample size number, because that is what Twitter uses to calculate <5% fake/spam/duplicate.

Tesla Owners Silicon... @teslaowne... · May 13 ... Replying to @elonmusk and @Twitter How are defining random ? How will you select them

Elon Musk 🕗 @elonmusk · May 13 ... Ignore first 1000 followers, then pick every 10th. I'm open to better ideas.

Coffee Table Tesla (... @coffeet... · May 13 ... Replying to @elonmusk and @Twitter

Why followers of that account, specifically? Why not yours?

Elon Musk 🕗 @elonmusk · May 13 Pick any account with a lot of followers

้าท่e้o 🕉 @realworldNeo · May 14 Replying to @elonmusk and @Twitter I did @nytimes and it had 38 out of 100





Let's try it with proportion inference

Verizon 🗢 12:15	5 PM	63% 🔲
Inference for	a Proportion	?
Enter Data: Summary Statistics		~
Sample Size (n): 100	# of Successes (x 38):
Sample Proportion:		
Number of Successes (x)	38	
Sample Size (n)	100	
Sample Proportion (p̂ = x/r	n) 38/100 = 0	0.3800
Bar Chart of Sample o	of 100 Twitter	

Fake Genuine 62.0% 52.0% 52.0% 62.0% 62.0% 62.0% 62.0% 62.0% 62.0% 62.0% 62.0% 62.0% 62.0% 62.0% 62.0% Failure Fercent (%) Forvide Labels for Success / Failure ENTER DATA CONFIDENCE INTERVAL HYPOTHESIS TEST

Enter Data on First Tab

App Store 📲 🛜 12:20 PM	61% 🔲
Inference for a Pro	portion ⑦
Confidence Interval for the Proportion p:	Population
Statistic	Value
Sample Size (n)	100
Number of Successes	38
Sample Proportion (p̂)	0.380
Standard Error (se) of the Sample Proportion	0.0485
Confidence Level	95%
z-score (a = 5%)	1.960
Margin of Error (me)	± 0.0951
Lower Bound for p	0.285
Upper Bound for p	0.475
p is the population proportion.	

Go to Second Tab "Confidence Interval" and obtain lower and upper bounds (plus intermediate statistics such as ME)



tation of the interval, plus options such as choosing the confidence level



Inference About Means





Let's try it with proportion inference

Verizon 穼 12:19	5 PM	63% 🔲
Inference for	a Proportior	n (?)
Enter Data: Summary Statistics		~
Sample Size (n): 100	# of Successes 38	(x):
Sample Proportion:		
Number of Successes (x)	38	
Sample Size (n)	100	
Sample Proportion (p̂ = x/r	n) 38/100	= 0.3800
Par Chart of Sample (r



?	〈 Inference for a Proportion			(
~	Null Value: 0.05	Alternativ Larger	ve Hypothesis:	~
	Hypothesis Test fo Proportion p:	or the Po	pulation	
	Null Hypothesis		H0: p = 0.05	
	Alternative Hypothesis	5	Ha: p > 0.05	
	Sample Proportion (ĝ)		0.380	
	Null Standard Error (se the Sample Proportion	e0) of ı	0.0218	
	Test Statistic (z)		15.1	
	P-value		< 0.0001	
	n is the nonulation pr	oportion		

10:19 AM

Verizon 穼

Go to Third Tab "Hypothesis Test", provide null value and type of hypothesis, and obtain Pvalue and intermediate steps.

Standard Normal Distribution Test Statistic: z = 15.1, P-value < 0.0001

Get a graphical representation of the distribution of the test statistic under the null, and the P-value.



Enter Data on First Tab

The apps work in offline mode:



	Gree in /	en Ba Airpla	inne ane l	r whe Mode	en e	
*			1:52 PM			51% 🔲
<	Th	ne Norn	nal Dis	stributio	on	?
Mean 100	μ:		s 1	itandard D 0	ev. σ:	
Nori P(X	mal Dis ≤ 80.40	tributio D) = 0.0	n with)250	μ = 100	0, σ = 1	0
2.5	<pre>%</pre>	/				
70	80	90	100	110	120	130
Percer	ntage:		Туре	of Percenti	ile:	
2.5		%	Low	er Tail: F	$P(X \le x)$	~
Perc x = 3	centile: 80.40					
EXPLO	DRE	FIND		FIND	SIM	IULATE

PERCENTILE

NUMBERS

DISTRIBUTION PROBABILITY



- Take screenshots
- Share with students/teachers or entire class, or post to social media.
- Create (narrated) videos through screen capture:

.II Veriz	on 🗢 🞋 3:29 PM Settings	65% 🔳	ull Verizon 중 :	3:29 PM Control Center	65% 🔳		-
	Notifications	>	INCLUDED	CONTROLS		all Verizon 🗢	
()	Sounds & Haptics	>	•	Flashlight			Not Playing
C	Focus	>	0	Timer			
x	Screen Time	>	•	Calculator			
			0	Camera	=	e -	
Ø	General	1 >	• 🖸	Screen Recording	=	Focus	* 4
8	Control Center	>	MORE CON	ITROLS			
AA	Display & Brightness	>	()	Accessibility Shortcuts	5	\sim	
' e	Home Screen	>	0	Alarm		\bigcirc	
Ì	Accessibility	>		Anala TV Damata			

- 1. On iOS, go to Settings, then Control Center
- 2. Activate Screen Recording
- 3. To start screen recording, swipe up from the bottom

On Android, swipe down from the top twice to see if your phone supports Screen Cast

T-Mobile **V** 100

ສີ Screen Cast

🚧 Nearby Share

Thu, May 19

Battery Saver Off

cord

Screen r
 Start

3.47

Include screenshots in documents.

Here is a copy of my Jamboard that I share with students after class.

-	STAT 161 Chapter 10: Comparing	Two Groups <	×	💽 🔹 Share
Ф	♂ Q ✓ Set background	Clear frame		다 Open on a Jam
	Enter Data:	Check Summary Statistics:	Obtain and interpret C.I.:	Obtain and interpret P-value:
	Il Verizon 12:31 PM 71% ■ Inference for Two Means ⑦	Inference for Two Means Inference for Starter and a product Inference for Two Means Inference for Two Means Inference for Two Means 	III Verizon 12:32 PM 70% ■ ✓ Inference for Two Means ⑦	Inference for Two Means ⑦ □ □ □
	Enter Data: Data Separator: Observations V Space	Practice No Practice	Data: n1 = 11, x1 = 0.0700, s1 = 0.0647 n2 = 9, ∞ = 0.1089, s2 = 0.0772	Null HypothesisH0: μ 1 - μ 2 = 0Alternative HypothesisHa: μ 1 - μ 2 \neq 0
k	Label for Group 1: Label for Group 2: Practice No Practice	Sample Size (n) 11 9 Sample Mean (x̄) 0.0700 0.1089	Difference (x1 - x2) -0.0389 Standard Error (se) 0.0323	Difference (x1 - x2) -0.0389 Standard Error (se) 0.0323
Ð	Practice Observations: 0.01 0.05 0.03 0.01 0.06 0.03 0.21 0.05 n = 11	Standard Deviation (s) 0.0647 0.0772 Standard Error (se) 0.0195 0.0257	Confidence Level 95% t-score (df = 15.7, α = 5%) 2.1234	Test Statistic (t) -1.2045 P-value 0.2462
	Enter data, separated by space, or copy & paste	Difference in Sample Means: x1 - x2 = -0.039	Margin of Error (me) ±0.0686 Lower Bound for μ1 - μ2 -0.1074 Users Bound for μ1 - μ2 0.0207	μ1 and μ2 are population means for Practice and No Practice, respectively.
O,	No Practice Observations: 0.25 0.09 0.16 0.04 0.02 0.14 0.01 0.14 n = 9 0.13	Practice	Upper Bound for μ1 - μ2 μ1 and μ2 are population means for Practice and No Practice, respectively.	t-Distribution with df = 15.7 Test Statistic: t = -1.20, P-value = 0.2462
	Enter data, separated by space, or copy & paste Practice No Practice	No Practice -	95% Confidence Interval (-0.1074, 0.0297)	t = -1.20 t = 1.20 t = 3.31%
\$	Sample Size (n) 11 9 Sample Mean (x̄) 0.0700 0.1089 Standard Deviation (s) 0.0647 0.0772	Show Observations on Boxplot	-0.1 0 Difference μ1 - μ2 of Population Means	-4 -3 -2 -1 0 1 2 3 4
	ENTER CONFIDENCE HYPOTHESIS DATA INTERVAL TEST	ENTER CONFIDENCE HYPOTHESIS DATA INTERVAL TEST	ENTER CONFIDENCE HYPOTHESIS DATA INTERVAL TEST	ENTER CONFIDENCE HYPOTHESIS DATA INTERVAL TEST

I show output of other software (here STATA) only after having talked about the app:



Output from commercial software such as Minitab, STATA, SPSS, JMP, but also from R is easier to understand and process once students have seen the app.

- Projecting the app on screen (either in class or remotely)
 - 1. Start a Zoom call from your computer
 - Join the Zoom call on your phone (with no audio and no video)
 - On computer, "admit" phone and make it a "Co-Host"



- Projecting the app on screen (either in class or remotely)
- 4. Press Share Content on your Phone
- 5. Select "Screen"
- 6. Go to the app to start broadcasting it



- Projecting the app on screen (either in class or remotely)
- 7. On your phone, go to the content you want to share (e.g., the Art of Stat App) Select "Screen"
- 8. Present on your phone



This is a video demoing the Linear Regression app and multivariate thinking.

Click on it to start it!

Students might get carried away a bit:

This is a video. Click on it to start it!



Discrete Distributions:

Poisson



2 3 4 5 Number of Events Number of Stars Discrete

James Zinn '22

Data Entry





Type in data



Copy & Paste



Upload CSV File

DATA

One Quantitative Variable ? Enter Data: Individual Observations Variable Name: Data Separator: Textbooks Space \sim Enter numerical observations, separated by a space, or copy & paste from a spreadsheet app (e.g., Google Sheets, Excel or Apple's Numbers app) Observations: Sample Size: 0 ENTER DESCRIPTIVE HISTOGRAM DATA

Recording

♦ 25%

Sample Dataset

These are short videos to show how to upload data into the app. Click each on to play.

Verizon LTE

- Pre-implemented Dataset: Online Lending
 - Data on 200 randomly sampled loans made through a website

	Loan		Interest				
Loan	Amount	Duration	Rate	Grade	Home	Income	FICO Score
1	35000	36	13.6	С	rent	160000	710
2	12000	60	16.1	С	mortgage	65000	675
3	15000	36	11.8	В	rent	34000	690
4	3025	36	15.1	С	mortgage	150000	740
5	9000	36	14.5	С	own	70000	705
6	20000	36	6.5	А	mortgage	100000	785
7	12500	60	7.8	А	rent	41600	700
8	5000	36	18.9	D	mortgage	44000	685
9	21000	60	16	С	mortgage	80000	685
10	35000	36	14.5	С	mortgage	90000	745
]		

Art of Stat Explore Data $\underbrace{\overline{\Box}}_{A} \underbrace{\overline{\Box}}_{B} \underbrace{\underline{\Box}}_{C}$ ıII 🕈 💽

 \odot

ART OF STAT Explore Data

Categorical Variables



One Categorical Variable

Compare Groups on Categorical Variable





Quantitative Variables



One Quantitative Variable

Compare Groupe on

Analyze the variable "Grade of the Loan"



 TestFlight Art of Stat **Explore** Data

8:22

60.

Categorical Variables

Barchart

Side-by-Side Barchart ABC

ART OF STAT **Explore** Data

 \odot

Analyze the variable "Grade of the Loan"



Art of Stat Explore Data 8:22 TestFlight

60.

Categorical Variables

Barchart

Side-by-Side Barchart

ABBC

ART OF STAT Explore Data

 \odot

Other Options:

Verizon LTE	യ 10:34 PM	1	00% 💋
〈 One Ca	ategorical Va	riable	?
Bar Chart			
Loan Grade			
A 31.0%			
B 25.0%			
C 25.0%			
D]12.0%			
E 5.0%			
F 2.0%			
0 8	16	24	32
	Percent (%)		
Horizontal Bar Cha	art 🚺		
Plot Percentages I	nstead of Coun	ts	
Include Labels on	Bars		
x-Axis Label:	y-Axis L	_abel:	
Percent (%)	Loan	Grade	
Plot Title:	Color Pa	alette:	
ENTER	FREQUENCY	BAR 8	
DATA	TABLE	CHA	RIS





Art of Stat Explore Data 8:22 TestFlight

 \odot

ART OF STAT Explore Data

Categorical Variables



Categorical Variable



Relationship Between Two Categorical Variables

Quantitative Variables



One Quantitative Variable

Compare Groupe on

Rate

13.6

16.1

11.8

15.1

14.5

6.5

7.8

18.9

16

14.5

Now, analyze the variable "Interest Rate"



••• Verizon LTE	ල 10:	45 PM	100% 🗲			
〈 One Quantitative Variable ⑦						
Open Sam	ple Datasets		~			
Select Datase Online Len	^{t:} ding		~	Interest		
Loan	Loan Amount	Duration	Interest Rate	Rate data		
1	35000	36	13.6			
2	12000	60	16.1			
3	15000	36	11.8			
4	3025	36	15.1			
5	9000	36	14.5			
6	20000	36	6.5			
7	12500	60	7.8			
Showing fir file, and all	Showing first 100 rows of the 200 rows in the CSV file, and all 8 columns.					
Select Variable Interest Ra	e: ite)	~	Interest Rate		
Sample Siz	e: 200					
ENTER DATA	DESCR STATIS	IPTIVE STICS	HISTOGRAM BOXPLOT			

Art of Stat **Explore** Data

ART OF STAT **Explore** Data

 \odot

Categorical Variables





One Categorical Variable

Compare Groups on Categorical Variable



20 40 60 80 100

Relationship Between Two Categorical Variables

Quantitative Variables



Quantitative Variable Analysis:

Verizon L	TE © 10:45 PI	M 1	00% 🗲			
<	One Quantitative	Variable	?			
Descrip	Descriptive Statistics for Interest Rate:					
	Statistic	Value				
Sample Si	ze (n)	200				
Mean (x̄)		12.813				
Standard	Deviation (s)	5.649				
Minimum		5.3				
First Quar	tile (Q1)	8				
Median		11.7				
Third Qua	rtile (Q3)	16				
Maximum		30.8				
Interquart	ile Range (IQR)	8				
Range (Ma	ax – Min)	25.5				
Find Percentile						
Adjust Number of Digits						
ENTER DATA	DESCRIPTIV	E HISTO	GRAM PLOT			

Descriptive Statistics



8:22 TestFlight

Art of Stat

Explore Data

ART OF STAT **Explore** Data

 \odot

Categorical Variables



One Categorical Variable

Compare Groups on Categorical Variable



Relationship Between Two Categorical Variables

Quantitative Variables



📶 Verizon 奈



Select

the

and

the

Grade as

Grouping

Variable,

Interest

Rate as

Response

Variable

..... 🗢 💽

ART OF STAT Explore Data \odot

Categorical Variables



One Categorical Variable

Compare Groups on Categorical Variable





Finally: Compare Groups on Quantitative Variable

43% 🔲



Row 2 0 20 40 50 50 100 Percent (%)	
Relationship Between Two Categorical Variables	
Quantitative Variable	5:
Histogram & Boxplot	Side-by-Side Boxplots
Scatterplot & Regression Regression Line: $\hat{y} = 2.63 + 0.524(x)$ \hat{y} $\hat{0}$ $\hat{0}$ $\hat{0}$ $\hat{1}$ $\hat{2}$ $\hat{4}$ $\hat{0}$ $\hat{1}$ $\hat{1}$ $\hat{2}$ $\hat{1}$ $\hat{1}$ $\hat{1}$ $\hat{2}$ $\hat{1}$ $\hat{1}$ $\hat{2}$ $\hat{1}$ $\hat{1}$ $\hat{2}$ $\hat{1}$ $\hat{1}$ $\hat{2}$ $\hat{1}$ $\hat{1}$ $\hat{2}$ $\hat{1}$	
Relationship Between Two Quantitative Variables	

2:09 PM

Verizon LTE 12:22 PM			60% 🔲 [,]		
🕻 Compare Groups: Quantitative Respo?					
Enter Data: Open Sample Datasets ~					
Select Dataset: Online Lending ~					
Loan	Loan Amount	Duration	Interest Rate		
1	35000	36	13.6		
2	12000	60	16.1		
3	15000	36	11.8		
4	3025	36	15.1		
5	9000	36	14.5		
6	20000	36	6.5		
7	12500	60	7.8		
Showing first 100 rows of the 200 rows in the CSV file, and all 8 columns.					
Group Variable: Response Variable: Grade \checkmark Interest Rate \checkmark					
ENTER DESCRIPTIVE BOXPLOTS					



ART OF STAT **Explore** Data \odot

Categorical Variables



Compare Groups on Categorical Variable

Group 2



Two Categorical Variables

Quantitative Variables



Finally: Compare Groups on Quantitative Variable

12:24 PM

Compare Groups: Quantitative Respo...?

Side-by-Side Boxplots of Loan Grades

Verizon LTE

30-

25.

Interest Rate

10

Verizon LTE 1	2:22 PM	60% 🔲			
Compare Groups: Quantitative Respo?					
Descriptive Statistics for Interact Pater					
		St Nate.			
Interest Rate	A	В	C		
Sample Size (n)	62	50	5(
Mean (x̄)	7.061	10.918	14.9		
Standard Deviation (s)	0.909	1.060	1.18		
Minimum	5.3	9.4	12.		
First Quartile (Q1)	6.2	10.1	14.		
Median	7.2	10.6	15		
Third Quartile (Q3)	7.8	11.525	16		
Maximum	8.8	13	16.		
Interquartile Range (IQ	R) 1.600	1.425	1.9(
Range (Max – Min)	3.500	3.600	4.3(
Find Percentile					
Adjust Number of Dig	its				
ENTER DE DATA ST		BOXPLO HISTOGR	TS AMS		

Descriptive Statistics



в

D

Grade



With FICO Score

One Categorical Variable

Relationship Between

Explore Data & Inference

Apps handle Contingency Tables

ull Verizon 🛜	3:12	DM	24% 💽	•11 V	
Compare Groups: Categorical Respo ⑦					
Enter Data:				C	
Open Exan	nple Datasets	8	~		
Select Datase				g	
Online Len	ding		~		
Loan	Loan Amount	Duration	Interest Rate	Du	
10	35000	30	14.5		
11	7700	36	6.1		
12	20000	60	18.0	т	
13	16000	36	8.0	Tv	
14	3500	36	17.5	G	
15	5000	36	11.1		
16	12000	60	18.1	Di: Pr	
47	20000		10.1		
Showing fir file, and all	st 100 rows of 8 columns.	the 200 rows	s in the CSV	Ν	

Group Variable: Duration	~	Response V Grade	/ariable:
ENTER	COMF	PUTE	BAR
DATA	PROPOF	RTIONS	CHART

III Verizon 🗢 3:12 P				24%		
〈 Compare Groups: Categorical Respo ⑦						
Conditi	ional Dist	ribution				
Conditional Distribution of Grade, given Duration:						
Grade						
Duration	А	В	С	D		
36	0.393	0.221	0.257	0.093		
60	0.117	0.317	0.233	0.183		
Total	0.310	0.250	0.250	0.120		
Type of Conditional Distribution: Grade, given Duration						
Display: Digits:						
Proportions ~			3	- +		
Marginal Distribution						
Marginal Distribution of Grade:						
A B C D E						
ENTER COMPUTE BAR						

PROPORTIONS



Art of Stat

Explore Data

8:22 TestFlight

ART OF STAT Explore Data

 \odot

Categorical Variables





One Quantitative Variable

Compare Groups on

Your turn!

A couple of prompts:

- The online lending website records home ownership of loan applicants. Explore the distribution of Home Ownership.
- You can either get a 36 months or a 60 months loan. Which one is more popular? Explore the distribution of Loan Duration.
- Investigate the relationship btw. Loan Duration and Grade. Do 60 months loans tend to have lower grades? *
- The online lending website records the FICO score of loan applicants. Explore the distribution of FICO scores.*
- How do FICO scores compare across Home Ownership? *



Row

8:22

TestFlight

Relationship Between Two Categorical Variables

Quantitative Variables



* Requires unlocking



 \odot

Art of Stat **Explore** Data





One Categorical Variable

Stacked Barchart

Col 1 Col 2 Col 3

Compare Groups on Categorical Variable



ART OF STAT



Your turn!



80

120

Mean µ

140

 \odot

ART OF STAT Inference

Inference About Proportions



Difference µ1 - µ2

30

The Online Lending dataset is also available as a Sample Dataset in the Inference app. Verizon 🗢 2:55 PM 30%

31% 🔲

III Verizon 奈 2:46 PM \odot ART OF STAT Inference Inference About Proportions Single Population Difference of Two Proportion **Population Proportions** Art of Stat Website **Offline Mode** Tell a Friend Settings **Restore Purchases** Sample Datasets & Use Cases Cancel

2:55 PM 🖬 Verizon 🗢 Sample Datasets

2. Online Lending

<

The website lendingclub.com is a trading platform for personal loans. Borrowers can ask for unsecured personal loans between \$1,000 and \$40,000, and investors can search and browse the loan listings and select loans they want to invest in based on the information supplied about the borrower. This dataset contains a random sample of 200 funded loans taken from this platform. The following variables have been recorded for each loan:

Loan Amount: in USD, between \$1,000 and \$40,000.

Access its descriptions and prompts for inference by going to Sample Datasets & Use Cases. (Press the three dots in the top right on the landing page.)

Sample Datasets

Inference for a Population Proportion: Find a confidence interval for the proportion of loans that are receiving a grade of A from lendingclub.com. Or, test a hypothesis that more than 50% of loan applicants rent rather than own a home or pay a mortgage.

Compare Two Population Proportions: Find out whether the proportion of loans with a grade of A differs substantially between 36 months and 60 months loans.

Chi-Square Test (Independence/Homogeneity): Is there evidence of an association between a loan grade (A through F) and home ownership (own, mortgage, rent)? If yes, inspect the standardized residuals to describe the type of association.

Inference for a Population Mean: Find a confidence interval for the mean loan amount taken out by users of the lendingclub.com website.

Compare Two Population Means:

Does the average loan amount taken out by users differ by duration of the loan (36 months vs. 60 months). In other words, are the longer loans for higher loan amounts, on average?

Concluding Remarks

Technology is crucial for how students and teachers interact with the material in a course:

- Concepts come to life when ...
 - ... you can "see" and interact with them.
- Concepts and ideas become memorable when ...
 ... you can associate an activity with them.
- Concepts and procedures stay relevant when you can use them on your own data.



Concluding Remarks

The Art of Stat Mobile Apps provide

- A whole ecosystem of apps for (almost) the entire intro stats curriculum.
- Supports students in creating and sharing content, and telling the story.

Thank you and I'm happy to answer any questions!









