Mongiello Associates Strategic Marketing™

mongielloassociates.com

Diamond Weight as a Predictor of Price

Pearson Correlation and Linear Regression

MARKET INSIGHT | DATA ANALYSIS | STRATEGIC POSITIONING



Mongiello Associates Strategic Marketing Firm

Market Insight • Data Analysis • Strategic Positioning

301 Cleveland Avenue
Grover, North Carolina 28073

704-937-2940

mongielloassociates.com

Diamond Weight as a Predictor of Price: Pearson Correlation and Linear Regression

Martin C. J. Mongiello

John F. Kennedy School of Psychology and Social Sciences, National University

PSY-7104: Statistics I

Dr. Mary Bold

October 22, 2026

© 2026 Martin C. J. Mongiello. All Rights Reserved.

Professor Martin C. J. Mongiello, MBA, MA, MCFE, prepared this scholarly work as part of his continuing doctoral studies at National University's John F. Kennedy School of Psychology and Social Sciences. This research received a final academic grade of A on November 1, 2026. During his second year of doctoral study, he has maintained a 4.0 cumulative GPA** and recognition on the Dean's List. This publication represents independent student scholarship and is made publicly available by the author. It does not constitute an official publication, endorsement, certification, or institutional position of National University or its faculty.

ORCID iD: 0009-0001-6619-1861

ORCID URL: https://orcid.org/0009-0001-6619-1861

Zenodo Research Community: Martin Mongiello at National University

Zenodo Community URL: https://zenodo.org/communities/martinmongiello/records

Individual Publication DOI: At CERN, Switzerland, on Zenodo

Diamond Weight as a Predictor of Price: Pearson Correlation and Linear Regression

Descriptive Statistics and Pearson Correlation

Diamond pricing illustrates the distinction between a strong statistical association and a complete explanation of value. The four Cs are carat weight, color, clarity, and cut; cut quality should not be confused with shape (Gemological Institute of America, n.d.). The present analysis examines whether diamond weight predicts listed price in 308 historical Singapore listings. These observations originated in an advertisement in Singapore's *Business Times* on February 18, 2000 (Chu, 2001). Although the original study was published in 2001 and became available through the publisher's online platform in 2017 (about 16 years later), its statistical approach provides useful methodological context for the present analysis. Today, in 2026, it continues to teach and shed light on the diamond and statistics industries. They represent advertised prices, not verified transactions or current market valuations.

The Week 9 assignment calls for a Pearson correlation, simple linear regression, a scatterplot with a fitted line, and an APA-style Results section (National University, n.d.-a). The analysis therefore uses weight as the sole predictor while acknowledging other recorded characteristics. Correlation describes the direction and strength of a linear association, whereas regression supplies an equation for predicting the outcome (Belhekar, 2016). A predictive relationship is not, by itself, evidence of causation (Bennett et al., 2017).

The dataset contained 308 round-cut diamonds. Chu (2001) applied regression analysis to diamond pricing using a dataset of 308 diamonds, identifying carat weight, clarity, color, and certification characteristics as relevant variables. Diamond weight, measured in carats, ranged from 0.18 to 1.10 (M = 0.63, SD = 0.28). Prices, reported in Singapore dollars (S$), ranged from S$638 to S$16,008 ($M$ = S$5,019.48, SD = S$3,403.12). Prices, converted from Singapore

dollars to U.S. dollars using the October 22, 2026, exchange rate of S$1 = US$0.7830 (Wise, 2026), ranged from approximately $499.55 to $12,534.26 ($M$ = $3,930.25, SD = $2,664.64). A two-tailed Pearson product-moment correlation indicated a strong, positive association between weight and price, r (306) = .945, p < .001. Heavier diamonds generally had higher listed prices.

Method

Data Source and Variables

The National University (n.d.-b) spreadsheet contained 308 complete records with a unique diamond identifier, weight, color, clarity, grading organization, and price. Weight was measured in carats and total listed price in Singapore dollars (S$). All stones were round in shape. A separate cut-quality grade was not recorded. Color ranged from D to I, and clarity categories were internally flawless (IF), very very slightly included (VVS1 and VVS2), and very slightly included (VS1 and VS2). The grading organizations were the Gemological Institute of America (GIA), International Gemological Institute (IGI), and Hoge Raad voor Diamant (HRD). All observations were retained. The companion SPSS data file contained the same 308 weight–price pairs, which matched the spreadsheet exactly. Descriptive statistics, correlation, and regression estimates were recalculated from the complete records. Calculations used unrounded values; rounding was applied only for presentation.

Analytical Strategy

A two-tailed Pearson correlation tested the null hypothesis that the population linear correlation was zero. Ordinary least-squares regression predicted price from weight and tested whether the population slope was zero, using $\alpha = .05$. The fitted model was $\hat{Y} = b_0 + b_1X$, where X denotes weight, $\hat{Y}$ denotes predicted price, b_0 is the constant, and b_1 is the unstandardized

slope. Very small probabilities are reported as $p < .001$ rather than $p = .000$, consistent with Abbott's (2016, p. 316) reporting guidance.

The scatterplot and residual behavior were examined rather than assuming that statistical significance establishes model adequacy. Regression normality concerns the errors conditional on the predictors, not a requirement that raw weight and price each be normally distributed (Jones et al., 2025). A supplementary Koenker version of the Breusch–Pagan test examined whether squared residuals varied with weight (whether the prediction errors became more spread out as diamond weight increased). HC3 heteroskedasticity-consistent standard errors and confidence intervals were also calculated as a sensitivity analysis. These checks address unequal error variance, an issue reviewed by Farrar et al. (2025), but do not establish causality or correct an unsuitable mean function.

Results

Descriptive Statistics

Mean diamond weight was 0.63 carats ($SD = 0.28$), with a range of 0.18 to 1.10 carats. Mean listed price was S$5,019.48 ($SD$ = S$3,403.12), with prices ranging from S$638 to S$16,008. Table 1 summarizes these results. All monetary results remain in the dataset's original Singapore-dollar units.

Table 1

Descriptive Statistics for Diamond Weight and Listed Price

Variable	*N*	*M*	*SD*	Range
Weight (carats)	308	0.63	0.28	0.18 to 1.10
Listed price (S$)	308	5,019.48	3,403.12	638 to 16,008

Note. S$ denotes Singapore dollars. Statistics describe all 308 observations in the supplied historical dataset.

Pearson Correlation

There was a strong positive correlation between weight and listed price, $r(306) = .945$, $p < .001$. The conventional two-tailed test rejected the null hypothesis of zero correlation. In this sample, heavier diamonds generally had higher listed prices. The correlation is unitless and does not measure a Singapore-dollar increase or demonstrate that weight alone caused the observed price differences.

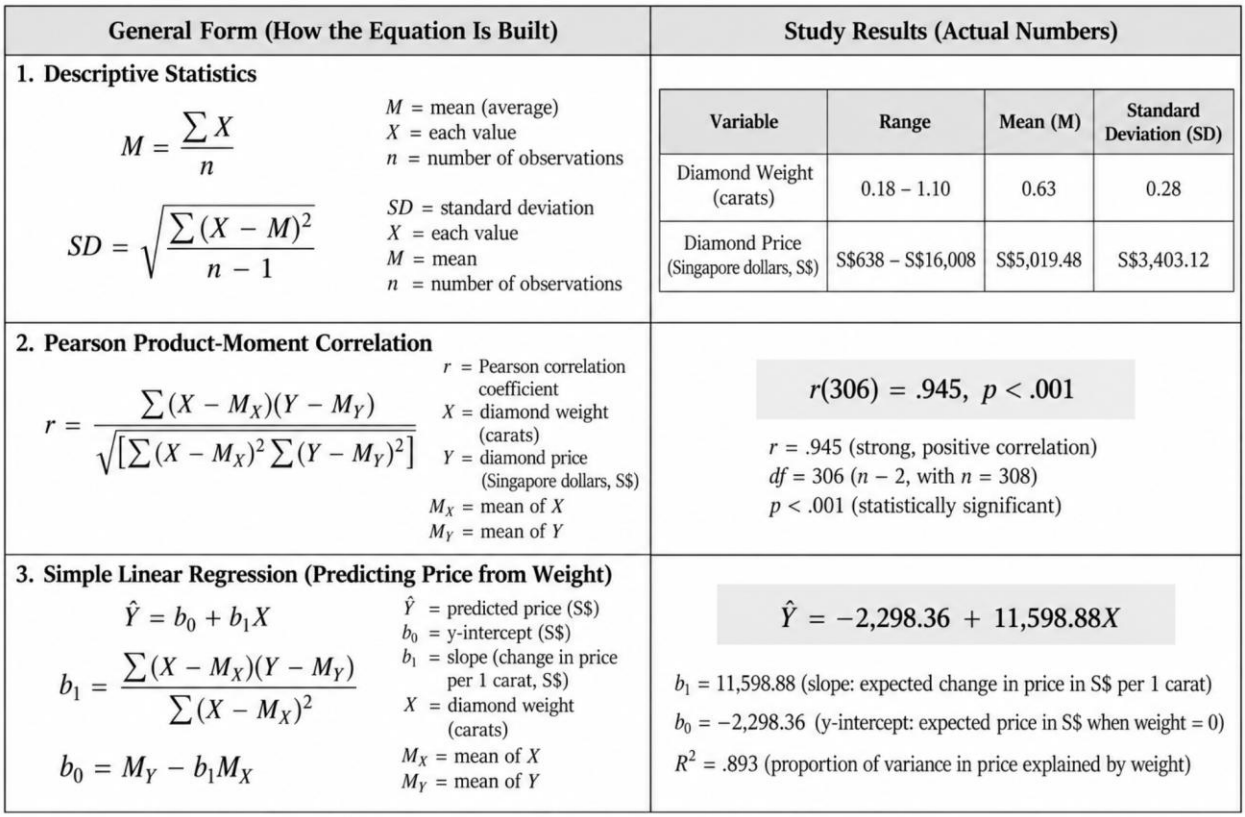
Simple Linear Regression

A simple linear regression was conducted with weight as the predictor and price as the outcome. The regression model was statistically significant, $F(1, 306) = 2{,}540.73$, $p < .001$, and explained 89.3% of the observed variance in price ($R^2 = .893$; adjusted $R^2 = .892$). The unstandardized regression slope was b = S$11,598.88 per additional carat ($SE = 230.11$), $t(306) = 50.41$, $p < .001$, 95% CI [S$11,146.08, S$12,051.68]. The intercept was −S$2,298.36. The fitted regression equation was:

$$\text{Predicted price (S\$)} = -2{,}298.36 + 11{,}598.88 \times \text{weight (carats)}.$$

Figure 1

Two-Panel Pearson Correlation: Formulas and Observed Diamond Data



General Form (How the Equation Is Built)

1. Descriptive Statistics

$$M = \frac{\sum X}{n}$$

M = mean (average)
X = each value
n = number of observations

$$SD = \sqrt{\frac{\sum (X - M)^2}{n - 1}}$$

SD = standard deviation
X = each value
M = mean
n = number of observations

2. Pearson Product-Moment Correlation

$$r = \frac{\sum (X - M_X)(Y - M_Y)}{\sqrt{\left[\sum (X - M_X)^2 \sum (Y - M_Y)^2\right]}}$$

r = Pearson correlation coefficient
X = diamond weight (carats)
Y = diamond price (Singapore dollars, S$)
M_X = mean of X
M_Y = mean of Y

3. Simple Linear Regression (Predicting Price from Weight)

$$\hat{Y} = b_0 + b_1 X$$

$$b_1 = \frac{\sum (X - M_X)(Y - M_Y)}{\sum (X - M_X)^2}$$

$$b_0 = M_Y - b_1 M_X$$

$\hat{Y}$ = predicted price (S$)
b_0 = y-intercept (S$)
b_1 = slope (change in price per 1 carat, S$)
X = diamond weight (carats)
M_X = mean of X
M_Y = mean of Y

Study Results (Actual Numbers)

Variable	Range	Mean (M)	Standard Deviation (SD)
Diamond Weight (carats)	0.18 – 1.10	0.63	0.28
Diamond Price (Singapore dollars, S$)	S$638 – S$16,008	S$5,019.48	S$3,403.12

$r(306) = .945,\ p < .001$

r = .945 (strong, positive correlation)
df = 306 (n − 2, with n = 308)
p < .001 (statistically significant)

$\hat{Y} = -2{,}298.36 + 11{,}598.88X$

b_1 = 11,598.88 (slope: expected change in price in S$ per 1 carat)
b_0 = −2,298.36 (y-intercept: expected price in S$ when weight = 0)
R^2 = .893 (proportion of variance in price explained by weight)

Note. The table illustrates the statistical formulas and corresponding results for 308 round-cut diamonds. Diamond weight was measured in carats, and prices were reported in Singapore dollars (S$). The Pearson correlation demonstrated a strong, statistically significant positive relationship between diamond weight and price, indicating that heavier diamonds generally commanded higher prices. Simple linear regression further showed that diamond weight strongly predicted price, accounting for about 89.3% of the variation in diamond prices.

Table 2

Coefficients for the Simple Linear Regression Predicting Price

Predictor	*b*	*SE*	*t*	*p*	95% CI for *b*
Constant	−2,298.36	158.53	−14.50	< .001	[−2,610.31, −1,986.41]
Weight	11,598.88	230.11	50.41	< .001	[11,146.08, 12,051.68]

Note. N = 308; residual degrees of freedom = 306. The constant is in S$, and the weight coefficient is in S$ per carat. CI = confidence interval. Standard errors and intervals in this table are the conventional equal-variance estimates; the supplementary HC3 results are reported separately.

The fitted regression equation was:

$$" = \ -2{,}298.36 \ + \ 11{,}598.88""X"$$

The slope corresponds to a fitted price difference of S$11,598.88 per carat. A more realistic comparison within the observed weight range is 0.10 carat, associated with a fitted difference of S$1,159.89. For example, predicted prices at 0.50 and 0.60 carat were S$3,501.08 and S$4,660.97, respectively. These predictions do not hold color, clarity, or grading organization constant because those variables were not included in the model.

The constant is the model's predicted price at zero carats, a value outside the observed range with no useful market interpretation. More importantly, the model predicts −S$210.56 at 0.18 carats, which is inside the observed range.

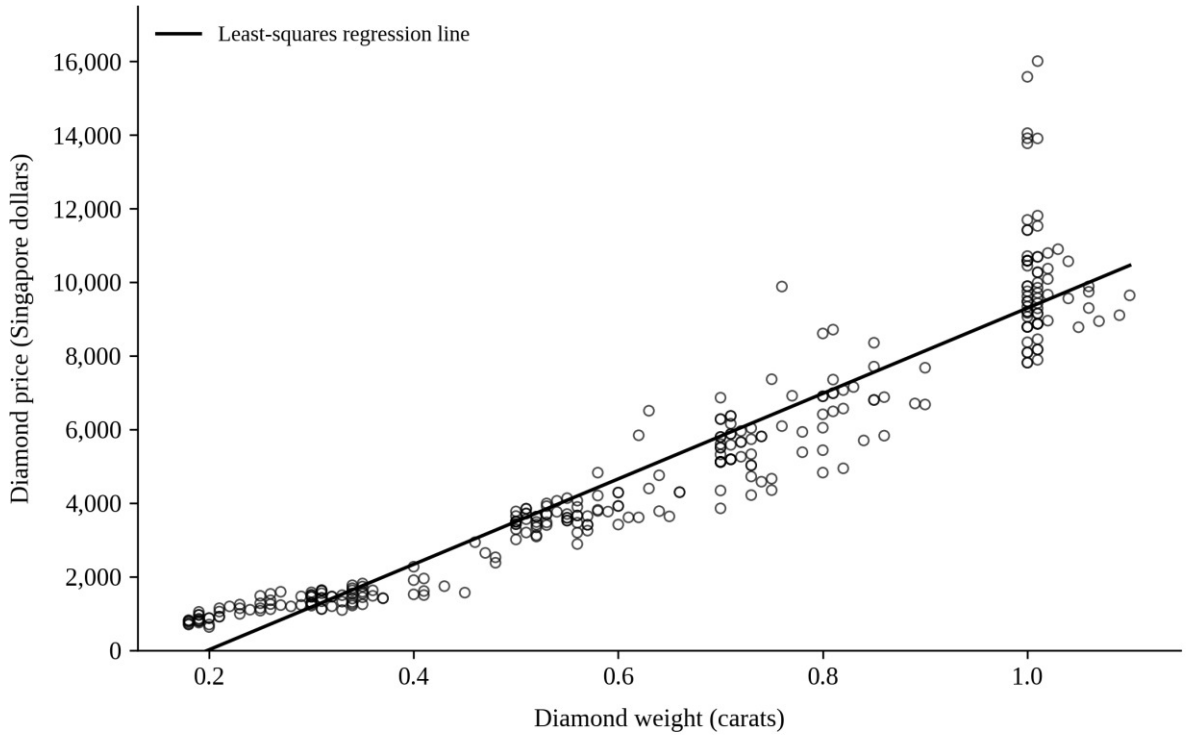
This impossible price shows that a strong overall fit does not make the straight-line equation a suitable appraisal rule for every stone. The Appendix provides the formulas and numerical results side by side.

Scatterplot and Supplementary Diagnostic Results

Figure 2 shows an upward association with substantial vertical price variation among stones of similar weight, especially around one carat. The residual standard error was S$1,117.56. This is a residual variability measure, not a guarantee that an individual prediction will be within S$1,117.56 of its listed price.

Figure 2

Scatterplot of Diamond Weight and Price with the Least-Squares Regression Line



Note. Open circles represent individual diamonds. The solid line represents the fitted weight-only regression. Prices are in Singapore dollars. The displayed vertical axis begins at zero; the fitted line's small negative predictions at the lowest weights are discussed in the text.

The Koenker–Breusch–Pagan diagnostic provided evidence against constant residual variance under the fitted specification, $\chi^2(1) = 18.05$, $p < .001$. With HC3 standard errors, the slope

estimate was unchanged at S$11,598.88 per carat, but its *SE* was 272.04, $t(306) = 42.64$, $p < .001$, 95% CI [11,063.59, 12,134.18]. The positive association therefore remained statistically significant when unequal error variance was allowed. These supplementary calculations do not resolve possible nonlinearity, omitted characteristics, or dependence among listings.

Other Recorded Diamond Characteristics

The sample included 151 GIA-graded diamonds (49.0%), 79 HRD-graded diamonds (25.6%), and 78 IGI-graded diamonds (25.3%). RATER is a three-category grading-organization variable, not a binary measure of whether a diamond was graded. The clarity counts were IF, 44; VVS1, 52; VVS2, 78; VS1, 81; and VS2, 53. Color, clarity, and grading organization were available for a future adjusted analysis, but their separate price contributions were not estimated here.

An illustrative comparison in the spreadsheet shows why weight alone is insufficient. Diamond 116 was a 1.00-carat, D-color, GIA-graded VVS1 stone listed at S$15,582. Diamond 117 had the same recorded weight, color, and grading organization but VS1 clarity and a listed price of S$11,419. The difference was S$4,163, although the weight-only model predicts the same S$9,300.53 for both. This pair illustrates price variation associated with recorded differences; it does not establish a general causal clarity premium.

Discussion

Meaning and Scope of the Findings

The results answer the assignment's central question: weight was positively and significantly related to listed diamond price. However, $R^2 = .893$ describes in-sample variance explained by this fitted model. It does not mean that 89.3% of individual prices were predicted correctly or that weight caused 89.3% of price differences. The remaining variation cannot be

assigned to clarity, color, or grading organization without fitting and evaluating an expanded model.

This distinction is especially important because the slope was not adjusted for other characteristics. Du et al. (2025) explain that omitted-variable bias depends on relationships between omitted variables, included predictors, and the outcome. Their empirical application concerns labor-market data, not diamonds; the relevant contribution here is the methodological warning against interpreting an unadjusted association as an isolated causal effect. The present analysis does not estimate the direction or magnitude of such bias.

Comparison With Diamond-Market Research

Recent studies support examining more than weight, but their outcomes and samples differ from this assignment. Huang et al. (2026) analyzed 183,538 natural and laboratory-grown diamond listings from Blue Nile collected in April 2025, reporting a prominent role for carat weight alongside color and clarity. Their model used logarithmic prices and additional quality characteristics. Those results provide context for the omitted attributes here; they do not validate the numerical slope from the historical Singapore sample.

Cardoso and Chambel (2026) predicted price per carat rather than total price and defined cut as shape, not cut quality. Their three retailer datasets also produced different model rankings. For example, Cubist achieved the highest test-set R^2 for the Miadonna data, whereas random forests (average decision trees) achieved the highest test-set R^2 for the TwelveFifteen (a lab-grown vendor) and natural-diamond data. Consequently, neither their unit-price coefficients nor a blanket claim that one algorithm always performs best should be transferred to the present regression.

Basha and Oveis (2024), Sánchez Sánchez (2024), and Xiao (2024) investigated more complex prediction methods using the widely reused 53,940-record diamond benchmark or samples drawn from it. These are not three wholly independent market samples. OuYang (2024) compared linear regression, decision trees, and random forests in a different dataset with prices in renminbi. Collectively, these studies motivate testing nonlinear and multivariable alternatives, but their reported performance cannot be compared directly with the current in-sample R^2 as though the data, predictors, currencies, and validation procedures were identical.

Sun et al. (2024) examined a different market: 3,741 auction records spanning 2008–2022, with investment-return comparisons beginning in 2009. Their hedonic analysis (feature-based pricing) considered characteristics including weight, color, clarity, shape, brand, and certificate information. Certificate presence was not statistically significant in their reported overall model ($p = .903$), illustrating why grading-related price premiums should not be presumed. That indicator is also different from the GIA/IGI/HRD categories in this spreadsheet. Auction findings and investment returns do not establish current retail values for these historical listings.

Limitations and Future Analysis

The sample represents one historical advertisement rather than a random sample of today's diamond market. Independence of listing errors cannot be established from the file alone, and no held-out test sample was used. Unequal residual variance, substantial dispersion near one carat, and the negative prediction at 0.18 carats limit interpretation as an appraisal model. The HC3 analysis strengthens the sensitivity assessment of the positive slope, but it does not eliminate these limitations.

A logical extension would include color, clarity, and grading organization using appropriate categorical coding, followed by residual diagnostics and out-of-sample evaluation. Chu's (2001) original analysis already explored transformed prices and nonlinear weight terms, demonstrating that refinements need not begin with complex machine learning. Cut quality, seller characteristics, and market timing would require additional information. Such work would be a separate analysis rather than an unreported component of the present weight-only model.

Conclusion

Among these 308 historical listings, mean weight was 0.63 carats and mean price was S$5,019.48. Weight and price were strongly correlated, $r(306) = .945$, $p < .001$. The fitted equation, $\hat{Y} = -2{,}298.36 + 11{,}598.88X$, explained approximately 89.3% of observed price variation. The association remained significant in the supplementary HC3 analysis, but the results support a descriptive and predictive relationship within this dataset, not a causal claim or a universally reliable diamond valuation formula.

References

Abbott, M. L. (2016). *Using statistics in the social and health sciences with SPSS and Excel*. Wiley. https://doi.org/10.1002/9781119121077

Basha, M. S. A., & Oveis, P. M. (2024). Predictive modeling and benchmarking for diamond price estimation: Integrating classification, regression, hyperparameter tuning and execution time analysis. *International Journal of System Assurance Engineering and Management, 15*(11), 5279–5313. https://doi.org/10.1007/s13198-024-02535-0

Belhekar, V. M. (2016). *Statistics for psychology using R*. SAGE Publications. https://doi.org/10.4135/9789353282493

Bennett, J. O., Briggs, W. L., & Triola, M. F. (2017). *Statistical reasoning for everyday life* (5th ed.). Pearson.

Cardoso, M. G. M. S., & Chambel, L. (2026). Predicting lab-grown diamond prices: A comparative view. *International Journal of Data Science and Analytics, 22*(1), Article 70. https://doi.org/10.1007/s41060-026-01048-2

Chu, S. (2001). Pricing the C's of diamond stones. *Journal of Statistics Education, 9*(2). https://doi.org/10.1080/10691898.2001.11910659

Du, S., Wilke, R. A., & Homrighausen, P. (2025). On omitted variables, proxies, and unobserved effects in empirical regression analysis. *Journal of Official Statistics, 41*(2), 725–744. https://doi.org/10.1177/0282423X241312644

Farrar, T., Blignaut, R., Luus, R., & Steel, S. (2025). A review and comparison of methods of testing for heteroskedasticity in the linear regression model. *Journal of Applied Statistics, 52*(16), 3121–3150. https://doi.org/10.1080/02664763.2025.2575038

Gemological Institute of America. (n.d.). *Diamond quality factors*. https://www.gia.edu/diamond-quality-factor

Huang, C., Wang, H., & Zhang, L. (2026). Comparative analysis and price modelling of natural and lab-grown diamonds. *International Journal of Electronic Marketing and Retailing, 18*(5), 1–25. https://doi.org/10.1504/IJEMR.2026.155557

Jones, L., Barnett, A., & Vagenas, D. (2025). Common misconceptions held by health researchers when interpreting linear regression assumptions, a cross-sectional study. *PLOS ONE, 20*(6), Article e0299617. https://doi.org/10.1371/journal.pone.0299617

National University. (n.d.-a). *PSY-7104: Week 9 assignment template* [Course assignment handout]. Navigate. https://resources.nu.edu/ld.php?content_id=66095849

National University. (n.d.-b). *PSY-7104: Week 9 data file* [Data set]. Navigate. https://resources.nu.edu/ld.php?content_id=66095854

OuYang, Z. (2024). Research on the diamond price prediction based on linear regression, decision tree and random forest. *Highlights in Business, Economics and Management, 24*, 248–257. https://doi.org/10.54097/13ccwv59

Sánchez Sánchez, M. (2024). Optimization of diamond price prediction strategies using machine learning techniques. *Journal of Economics and International Finance, 16*(3), 28–38. https://doi.org/10.5897/JEIF2024.1223

Sun, N., Zhou, Q., Zhao, M., Xu, Y., & Zhan, F. (2024). A study on the price and investment return of colourless diamond in auction market. *Journal of Gems & Gemmology, 26*(6), 92–105. https://jogg.cug.edu.cn/en/article/doi/10.15964/j.cnki.027jgg.2024.06.009

Wise. (2026). Singapore dollar to US dollar exchange rate history. https://wise.com/us/currency-converter/sgd-to-usd-rate/history/22-10-2026

Xiao, M. (2024). Diamond price prediction based on regression algorithms: Analyzing key factors and model performance. *Advances in Economics, Management and Political Sciences, 139*, 196–205. https://doi.org/10.54254/2754-1169/2024.19476

Appendix

Formula and Result Comparison

Table A1

Statistical Formulas and Corresponding Results

General form	Numerical result
$M = \frac{\sum X}{N}$	M of weight = 0.63 carats M of price = S$5,019.48
$SD = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}}$	SD of weight = 0.28 carats SD of price = S$3,403.12
$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$	$r(306) = .945$, $p < .001$
$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$	b_1 = S$11,598.88 per carat
$b_0 = \bar{Y} - b_1 \bar{X}$	b_0 = −S$2,298.36
$\hat{Y} = b_0 + b_1 X$	$\hat{Y} = -2{,}298.36 + 11{,}598.88X$
$R^2 = r^2$	$R^2 = .893$

Note. $N = 308$. X = weight in carats; Y = listed price in S$; $\hat{Y}$ = predicted price; M = mean; SD = sample standard deviation; b_0 = constant; b_1 = slope. Subscripts X and Y identify the corresponding variable. Calculations used unrounded data. The equality $R^2 = r^2$ applies to this one-predictor least-squares model with an intercept.

MONGIELLO HOLDINGS, LLC

YOU ARE UNDER OUR PROTECTION

Your time. Your capital. Your goals.

Private consultations and event experiences with Martin Mongiello



★ YOUR JOURNEY WITH US ★

0 · LEVEL 0 · START HERE
COMPLIMENTARY
Explore our published work. Arrive prepared.

1 · LEVEL 1 · A BRIEF HELLO
2 minutes · COMPLIMENTARY
A warm social introduction.

2 · LEVEL 2 · STRATEGIC CONSULTATION
Telephone · 45 minutes
$500 PREPAID

3 · LEVEL 3 · STRATEGIC CONSULTATION
Zoom video · 45 minutes
$900 PREPAID

4 · LEVEL 4 · IN-PERSON CONSULTING
1 hour · By appointment · Grover, NC
$6,000 PREPAID

5 · LEVEL 5 · BESPOKE PROJECTS & STRATEGIC MANDATES
Tailor-made solutions, reports, and deal execution.
PRICED SEPARATELY

COMPLIMENTARY READING

An educated client receives the greatest value.

1. How To Position A Deal So Capital Says Yes
2. How To Turn One Business Into Three Income Streams
3. How To Protect Your Family Office From Costly Mistakes
4. How To Buy Credibility Instead of Building It From Scratch
5. How To Make Your Banker Your Biggest Fan

EXPLORE OUR WEBSITES

https://www.martinmongiello.com
https://www.MongielloAssociates.com
https://www.TheInnofthePatriots.com



KEYNOTE SPEECHES & PRESIDENTIAL EVENT EXPERIENCES ★

$6,000 to $60,000
Based on program scope

Convention and summit programs may include:

- ★ White House Master Chefs State Dinner
- ★ On-site cooking classes
- ★ First Lady's Cocktail Hour

FOR LEVELS 2–4: PAY FIRST, THEN BOOK.

Zelle: marti@mongiello.org
https://calendar.app.google/9z2VBQpBHzqou1zo8

Marti Mongiello
WHITE HOUSE MASTER CHEF

Supported by Erin Borders, CPA and Christopher Klink, Esq.

You are not a transaction.
You are a relationship under our protection.