Learning Module 10: Simple Linear Regression

Q.225 A stock's returns for the past four years are as follows: 12%, 9.5%, 8%, 14.7%. The geometric mean return is *closest to*:

A. 11.02%

B. 11.05%

C. 51.90%

The correct answer is **A.**

$$\text{Geometric return} = (1.12 \times 1.095 \times 1.08 \times 1.147)^{\frac{1}{4}} - 1 = 0.11 \text{ or } 11\%$$

B is incorrect. It denotes the arithmetic mean and not the geometric mean:

$$\text{Arithmetic mean} = [\frac{(0.12 + 0.095 + 0.08 + 0.147)}{4}] = 0.1105 = 11.05\%$$

C is incorrect. It denotes the geometric mean calculation but without the root sign calculation as follows:

$$\text{Geometric mean} = (1.12 \times 1.095 \times 1.08 \times 1.147) - 1 = 0.519 = 51.90\%$$

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

Q.407 Consider the following distribution, 3.5%; 3.8%; 5.9%; 9.6%; 12.4%; 2.3%. The second quintile is *closest to*:

A. 3.64%

B. 3.74%

C. 3.80%

The correct answer is **B.**

We are looking for the second quintile = 40%. We have n=6 (observations). Using the following formula, y=40% so that

$$L_y = (n+1)\frac{y}{100} = (7)\left(\frac{40}{100}\right) = 2.8$$

Arranging the distribution in ascending order we have, 2.3%, 3.5%, 3.8%,5.9%, 9.6%, and 12.4%. Therefore, the second quintile is between the 2nd number 3.5%, and the 3rd number 3.8%. To find the 2.8th number, we interpolate as:

$$0.035 + (2.8 - 2)(0.038 - 0.035) = 0.0374 = 3.74$$

Thus the second quintile is 3.74%.

A is incorrect. It is below the 2nd quantile.

C is incorrect. It is above the 2nd quantile.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

Q.3426 Which of the following is the *most appropriate* description of a parameter?

A. A numerical measure that describes a characteristic of a sample.

B. A numerical measure that describes a characteristic of a population.

C. A statistical inference that describes a characteristic of a population.

The correct answer is **B.**

A parameter is a numerical measure that describes a characteristic of a population, whereas a statistic is a numerical measure that describes a characteristic of a population sample.

A is incorrect. A numerical measure that describes a characteristic of a sample is called a statistic.

C is incorrect. Statistical inference is the process of using a sample to conclude a population from which the sample has been drawn.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

Q.3910 Tracy, senior analyst at CMSSP Capital reviewed Cronin (his junior's) regression analysis. He asked Cronin how the key inputs to the regression could affect the ultimate results. Cronin explained the effects of some of these inputs and assumptions, Cronin made the following comments: **Comment 1**: "The standard error of estimate is an important input for a hypothesis test. Small standard errors result in both tighter confidence intervals and tighter prediction intervals." **Comment 2**: "The estimated value for the variance of the independent variable can also affect hypothesis testing. The higher the assumed variance, the tighter the prediction intervals. However, changes in the assumed variance will have no effect on the confidence interval." Cronin is *most* accurate with respect to:

A. Comment 1 only.

B. Comment 2 only.

C. Both comments 1 and 2.

The correct answer is **C.**

Comment 1 is correct. Smaller standard errors would result in tighter confidence intervals and prediction intervals.

Comment 2 is correct. The higher the variance of the independent variable, the lower the variance of the forecast error, and the tighter the prediction interval. Confidence intervals do not depend on this input.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (d): Describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression.

Q.3911 “For our regression model to be valid, a linear relationship must exist between EPS growth and changes in return spread.” This statement implies that:

A. EPS growth and changes in return spread must be discrete random variables.

B. The correlation coefficient between EPS growth and Changes in return spread must be greater than zero but less than one.

C. The slope coefficient and the intercept are raised to the first power only, and neither of them is divided or multiplied by another regression parameter.

The correct answer is **C.**

For any linear regression model to be valid and reliable, we must assume that there’s a linear relationship between the dependent variable and the independent variable. This, in effect, means that the intercept and the slope coefficient can only be raised to the first power. In addition, none of the two should be multiplied or divided by another regression parameter.

A is incorrect. Linear regression assumes that the independent variable is not random because if the independent variable is random, the relation between the dependent and independent variables will not be random.

B is incorrect. Linear relationship can exist as long as the correlation coefficient lies between -1 and +1.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

Q.3912 Kim Richard has been looking at ways to increase efficiency in the construction process especially with regard to fuel consumption. She ran a regression explaining the variation in fuel consumption as a function of distance. The total variation of the dependent variable was 160.85, the explained variation was 80.15, and the unexplained variation was 100.70. She had 60 monthly observations. The standard error of the estimate in the regression is *closest to*:

A. 1.32.

B. 1.52.

C. 1.74.

The correct answer is **A.**

$$\text{Standard error of the estimate} = \left[\sum_{i=1}^{n} \frac{(Y_i - \hat{b}_0 X_i)^2}{n-2}\right]^{\frac{1}{2}}$$

$$= \left(\frac{\text{Unexplained variation}}{n-2}\right)^{\frac{1}{2}}$$

$$= \left(\frac{100.7}{60-2}\right)^{0.5} = 1.3177$$

B is incorrect. It suggests a standard error of 1.52, which does not align with the calculation based on the provided data.

C is incorrect. It proposes a standard error of 1.74, which is also not supported by the calculation using the given data. The discrepancy indicates a fundamental error in the calculation process or a misinterpretation of the formula's components.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (d): Describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression.

Q.3915 An analyst is developing a regression model to forecast project cost based on the construction costs. He has gathered the following information.

- Multiple R: 0.8821
- R-squared: 0.7651
- Standard Error of Estimate: 0.6346
- Observations: 62

- Variance of mean construction costs = 27.9
- Variance of mean forecasted project price = 18.35
- Mean construction costs = 98.54
- Correlation between mean construction costs and mean forecasted price = 0.75

The standard deviation of the prediction error given independent variable equals 425 is *closest to*:

A. 5.06.

B. 25.64.

C. 41.09.

The correct answer is **A.**

$$
\begin{aligned}
S_f^2 &= s^2\left[1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{(n-1)s_x^2}\right] \\
&= 0.6346^2\left[1 + \frac{1}{62} + \frac{(425 - 98.54)^2}{(62-1)27.9}\right] \\
&= 0.40272(1 + 0.016129 + 62.62185) = 25.63
\end{aligned}
$$

Standard deviation of the prediction error = $\sqrt{25.63} = 5.06$

B is incorrect. 25.64 represents a misunderstanding of the question. It seems to be a misinterpretation of the variance of the prediction error, not the standard deviation.

C is incorrect. 41.09 does not correspond to any calculation related to the standard deviation of the prediction error based on the given data and formula.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (d): Describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression.

Q.3916 Which of the following statement is *most likely* correct?

A. If the sample size is increased, the standard error of the estimated measure will increase. This will reduce the reliability of regression results.

B. If the sample size is increased, the standard error of the estimated measure will decrease. This will increase the reliability of regression results.

C. If the sample size is increased, the standard error of the estimated measure will remain constant. This will not affect the reliability of regression results.

The correct answer is **B.**

An increase in the size of the sample will decrease the sum of squared errors (SSE) and simultaneously increase the denominator of the formula on the right side (see below).

Mean square error (MSE) is the sum of squares error divided by the degrees of freedom, which are $n - k - 1$.

$$\text{MSE} = \sqrt{\frac{\text{SSE}}{n - k - 1}}$$

A is incorrect. This option suggests that increasing the sample size would increase the standard error of the estimated measure, which is not accurate. In statistical terms, the standard error of an estimate is inversely related to the square root of the sample size (n). Mathematically, this relationship can be expressed as:

$$\text{Standard Error} = \frac{\sigma}{\sqrt{n}}$$

where σ is the standard deviation of the population, and n is the sample size. As n increases, the denominator of this fraction becomes larger, resulting in a smaller standard error. Therefore, contrary to the statement, increasing the sample size actually enhances the reliability of regression results by reducing the standard error.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (d): Describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression.

Q.3917 Which of the following statements is *most likely* correct?

A. The standard error of estimate is the standard deviation of the actual values of the independent variable.

B. The standard error of estimate measures the standard deviation of the residual term; its numerator is calculated as the difference between the actual and predicted value of the dependent variable.

C. The standard error of estimate measures the standard deviation of the residual term; its numerator is calculated as the difference between the actual and predicted value of the independent variable.

The correct answer is **B.**

The standard error of estimate is the standard deviation of the residual term in the regression. Its numerator is calculated as the difference between the actual and predicted values of the dependent variable.

A is incorrect. The standard error of estimate is concerned with the residuals of the model, which are derived from the dependent variable's actual and predicted values.

C is incorrect. This option misstates the calculation of the standard error of estimate by suggesting that its numerator is calculated as the difference between the actual and predicted values of the independent variable. The standard error of estimate focuses on the dependent variable in a regression model, not the independent variable. The residuals, which are the basis for calculating the standard error of estimate, are determined by the differences be

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (d): Describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression.

Q.3918 An analyst is forecasting quarterly sales of Smart Inc., a smart TV manufacturer based in Thailand. The regression model is:

$$\text{Sales}_t = b_0 + b_1 \text{Sales}_{t-1} + \epsilon_t$$

The regression results for the smart TV sales model are presented below:

R-squared: 0.7436
Observations: 120

	Coefficient	Standard Error
Intercept	313.24	99.43
Lag 1	0.67	0.16

If TV sales in the first quarter were 1,137, the number of sales forecasted for the second quarter is *closest to*:

A. 762.

B. 1,075.

C. 1,137.

The correct answer is **B.**

The model is correctly specified. Hence, quarterly sales for Smart Inc., can be forecasted using this model:
b_0 = Intercept = 313.24
b_1 = Slope coefficient = 0.67

Thus,

$$\text{Sales}_t = 313.24 + (0.67 \times 1{,}137) = 1{,}075.03$$

A is incorrect. As calculated above.

C is incorrect. This option suggests that the number of sales would remain the same from the first quarter to the second quarter, which disregards the regression model provided. The model indicates that sales are influenced by the sales of the previous quarter through a specific mathematical relationship. Simply repeating the previous quarter's sales ignores the model's dynamics and the effect of the slope coefficient, which is designed to capture the relationship between consecutive quarters' sales.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (e): Calculate and interpret the predicted value for the dependent variable, and a prediction interval for it, given an estimated linear regression model and a value for the independent variable.

Q.3920 An analyst is assessing the contagion effect or spread of market disturbances in financial markets. He picks up four globally recognized indices and prepares a correlation matrix using monthly returns of various stock indices for the last 4 years as shown below:

	DJIA	S&P 500	FTSE 100	CAC 40
DJIA	1.00			
S&P 500	0.78	1.00		
FT SE 100	0.43	0.66	1.00	
CAC 40	0.38	0.33	0.80	1.00

The correlation coefficient is not statistically significant at the 0.01 significance level for which pair of market indices? (See the t-table)

A. CAC 40 with DJIA.

B. FT SE 100 with DJIA.

C. CAC 40 with S&P500.

The correct answer is **C.**

Because Zeng is looking at monthly returns for the last 4 years, n=48. Then, the critical value of the t-test (at the 0.005 significance level, with 48-2 = 46 degrees of freedom) is 2.687. We formulate the null hypothesis that the coefficient = 0 and reject the null when the computed test statistic is outside the range ±2.687.

CAC 40 and DJIA: Correlation = 0.38

The correlation coefficient between CAC 40 and DJIA is 0.38 and test statistic is calculated as follows:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.38\sqrt{48-2}}{\sqrt{1-0.38^2}} = 2.78$$

Hence, the correlation coefficient between FT SE 100 and DJIA is statistically significant.

CAC 40 and S&P 500: Correlation = 0.33

$$t = \frac{0.33\sqrt{48-2}}{\sqrt{1-0.33^2}} = 2.37$$

2.37 **is less than** the upper 99.9% point of the t-distribution with 46 degrees of freedom, i.e., 2.687

Hence, the correlation coefficient between CAC 40 and S&P 500 is not statistically significant.

A is incorrect. The correlation coefficient between CAC 40 and DJIA is 0.38. Using the formula, we find: This t-statistic (2.78) is greater than the critical value of 2.687, indicating that the correlation coefficient between CAC 40 and DJIA is statistically significant at the 0.01 significance level. This suggests that there is a statistically significant linear relationship between these two indices.

B is incorrect. The correlation coefficient between FTSE 100 and DJIA is not directly questioned for its significance in the provided options. However, if we were to calculate it, we would apply the same formula and compare the resulting t-statistic to the critical value to determine its significance.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

Q.3921 Richard Zeng is developing a regression model to predict stock market returns using the GDP growth rate. He considers quarterly returns of the S&P 500 (S&P) as a proxy for stock market returns and quarterly changes in GDP as GDP growth rate (GDP Growth). The linear regression model is as follows:

$$\text{S\&P} = \beta_0 + \beta_1(\text{GDP Growth}) + \epsilon$$

Zeng develops the following partial ANOVA table and regression statistics based on the last 10 years of quarterly data pertaining to the S&P 500 and GDP.

	DF	SS
Regression	1	108
Residual	38	To be calculated
Total	39	155.5

The percentage of variation in the S&P 500 return that can be attributed to the GDP growth rate is *closest to:*

A. 31%.

B. 69%.

C. 100%.

The correct answer is **B.**

The percentage of variation in the S&P 500 return that can be attributed to the GDP growth rate is also called coefficient of determination (R^2).

$$\text{Coefficient of determination} = \frac{\text{RSS}}{\text{TSS}}$$

Where, RSS is the regression sum of squares, or the amount of total variation that is explained in the regression equation, and is the total variation. These numbers are both given in the table. Hence,

$$\text{Coefficient of determination} = \frac{108}{155.5} = 0.6945 \approx 69\%$$

A is incorrect. Suggesting that only 31% of the variation in the S&P 500 returns is explained by the GDP growth rate underestimates the explanatory power of the GDP growth rate in the regression model. This option likely results from misunderstanding the calculation of the coefficient of determination or misinterpreting the values provided in the ANOVA table.

C is incorrect. Claiming that 100% of the variation in the S&P 500 returns is explained by the GDP growth rate is unrealistic in practical economic and financial modeling. It is highly unlikely for a single variable, such as GDP growth, to account for all the variations in a complex market index like the S&P 500. This option disregards the presence of other factors and the inherent randomness in financial markets that can affect stock market returns.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (d): Describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression.

Q.3922 Richard Zeng is developing a regression model to predict stock market returns using the GDP growth rate. He considers quarterly returns of the S&P 500 (S&P) as a proxy for stock market returns and quarterly changes in GDP as GDP growth rate (GDP Growth). The linear regression model is as follows:

$$S\&P = \beta_0 + \beta_1(\text{GDP Growth}) + \epsilon$$

The significance of Zeng's model for predicting the S&P 500 return using the GDP growth rate can be tested by:

A. t-test only.

B. F-test only.

C. either t-test on slope coefficient or F-test model because both will lead to the same conclusion.

The correct answer is **C.**

Either a t-test on the slope coefficient or an F-test on the simple linear regression model can be used. Both tests will lead to the same conclusion (as the F-statistic is simply the square of the t-statistic for the slope coefficient in this case). An F-test is used to determine the effectiveness of independent variables in explaining the variation of the dependent variable.

The F-test can be carried out with more than one independent variable.

However, had this question been about multiple regression analysis, only the F-test can be applied to evaluate the overall statistical significance of the model, and t-tests could be used to evaluate the statistical significance of individual slope coefficients.

A is incorrect. Suggesting that only a t-test can be used to assess the significance of Zeng's model is not entirely accurate. While a t-test is indeed used to determine the statistical significance of individual coefficients in a regression model, implying it is the sole method overlooks the applicability and relevance of the F-test in evaluating the overall model's significance, especially in simple linear regression scenarios.

B is incorrect. Stating that only an F-test can be used to evaluate the significance of Zeng's model is misleading. The F-test is crucial for assessing the joint significance of multiple coefficients in a regression model, particularly in multiple regression scenarios. However, in the context of a simple linear regression model like Zeng's, where there is only one independent variable, the F-test's conclusion about the model's overall significance will mirror the conclusion drawn from a t-test on the slope coefficient. Therefore, stating that only an F-test is applicable neglects the role and relevance of the t-test in this specific context.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

Q.3923 Zeng is developing a regression model to predict stock market returns using the GDP growth rate. He considers quarterly returns of the S&P 500 (S&P) as a proxy for stock market returns and quarterly changes in GDP as GDP growth rate (GDP Growth). The linear regression model is as follows:

$$S\&P = \beta_0 + \beta_1(\text{GDP Growth}) + \epsilon$$

Zeng develops the following partial ANOVA table and regression statistics based on the last 10 years of quarterly data pertaining to the S&P 500 and GDP.

	DF	Sum of Squares
Regression	1	108
Residual	38	To be calculated
Total	39	155.5

The standard error of the estimate for Zeng's model to predict stock market returns using the GDP growth rate is *closest to*:

A. 0.0366.

B. 0.0534.

C. 1.1180.

The correct answer is **C.**

$$\text{Standard Error of Estimate (SSE)} = \sqrt{\frac{\text{sum of square of errors}}{n-2}}$$

Where:

$$\text{Sum of square of Errors} = \text{SST-SSR} = 155.8 - 108 = 47.5$$
$$N = 40 \text{ (Based on 10 years of quarterly data)}$$

Thus,

$$SSE = \sqrt{\frac{47.5}{40-2}} = 1.1180$$

A is incorrect. The value 0.0366 does not correctly represent the standard error of the estimate based on the given data and the correct calculation method. This value significantly underestimates the variability of the residuals around the regression line, suggesting a much higher precision of the model's predictions than is actually the case.

B is incorrect. The value 0.0534 also does not align with the correct calculation of the standard error of the estimate. Similar to option A, this value underestimates the variability of the residuals, implying a level of precision in the model's predictions that is not supported by the data.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (d): Describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression.

Q.3924 The statistic which is used to measure how well a given linear regression model captures the relationship between the dependent and independent variables is *most likely* known as:

A. Standard error of the estimate.

B. Intercept of the regression model.

C. Slope of the independent Variable.

The correct answer is **A.**

An estimate’s standard error measures how well a given linear regression model captures the relationship between the dependent and independent variables.
B is incorrect. The intercept is the estimate of the dependent variable when the independent variable is zero.

C is incorrect. The slope coefficient represents the expected change in the dependent variable for a one-unit change in the independent variable.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

Q.3925 Which of the following statistic is *most likely* used to identify the fraction of the total variation that is explained by the regression?

A. Coefficient of determination.

B. Intercept of the regression model.

C. Slope of the independent variable.

The correct answer is **A.**

The coefficient of determination is the fraction of the total variation in the dependent variable that is explained by the independent variable.

B is incorrect. The intercept is the estimate of the dependent variable when the independent variable is zero.

C is incorrect. The slope coefficient represents the expected change in the dependent variable for a one-unit change in the independent variable.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

Q.3926 Mike Far explains the linear regression model and its underlying assumptions using the following statement: “The estimated parameters in a linear regression model maximize the sum of the squared regression residuals.” The above statement on estimated parameters in a linear regression model is *most likely*:

A. Correct.

B. Incorrect, because the model minimizes the sum of squared regression residuals.

C. Incorrect, because the model minimizes the sum of the regression residuals.

The correct answer is **B.**

A linear regression model computes a line that best fits the observations. It chooses values for the intercept and slope that minimize the sum of the squared vertical distance between the observations and the regression line. Hence, the estimated parameters in a linear regression model minimize the sum of the squared regression residuals.

A is incorrect. It suggests that the estimated parameters in a linear regression model maximize the sum of the squared regression residuals. This statement is the opposite of the actual process involved in linear regression analysis. The goal of linear regression is not to maximize but to minimize the discrepancies between the observed values and the values predicted by the model. By minimizing the sum of squared residuals, the model ensures that the predicted values are as close as possible to the actual observed values, thereby achieving the best fit.

C is incorrect. Residuals can be both positive and negative, and when summed up, positive and negative errors could cancel each other out, giving an impression of a better fit than it actually is. Squaring the residuals before summing them ensures that all errors are treated as positive values, emphasizing larger errors more significantly and leading to a more accurate estimation of the model parameters.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

Q.3927 An analyst has prepared a regression analysis comparing the price of gold to the average cost of purchases of finished gold jewelry of a retailer of fine jewelry and watches. The regression results are shown in Exhibit 1 below.

Exhibit 1: 1983-2013 Annual Data
(31 Observations)

Variable	Coefficient	SE of Coefficient
Intercept	11.06	7.29
Cost of gold	2.897	0.615

*SEE=117.8

The per ounce price of gold that corresponds to the $1,500 cost of finished jewelry is *closest to*:

A. $513.96.

B. $517.77.

C. $521.59.

The correct answer is **A.**

The regression model is of the form:

$$y = \hat{b}_0 + \hat{b}_1 x$$

Where:
y = Cost of the finished jewelry.
$\hat{b}_0$ = Slope coefficient.
x = Cost of gold.

Then, we can solve for to find the cost of gold:

$$1500 = 11.06 + 2.897x$$

Hence,

$$\text{Cost of gold} = \frac{1500 - 11.06}{2.897} = 513.96$$

B is incorrect. The option suggesting a price of $517.77 does not align with the calculation based on the regression model provided. It appears to be a miscalculation or misinterpretation of the regression equation.

C is incorrect. The option suggesting a price of $521.59 also does not follow from the regression model provided in the question. Similar to option B, this appears to be a result of a miscalculation or misunderstanding of how to apply the regression equation.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (e): Calculate and interpret the predicted value for the dependent variable, and a prediction interval for it, given an estimated linear regression model and a value for the independent variable.

Q.3928 Singh, an analyst at Delta Advisory Firm, has prepared a regression analysis comparing the price of gold to the average cost of purchases of finished gold jewelry of a retailer of fine jewelry and watches. The regression results are shown in Exhibit 1 below.

Exhibit 1: 1983-2013 Annual Data
(31 Observations)

Variable	Coefficient	SE of Coefficient
Intercept	11.06	7.29
Cost of gold	2.897	0.615

*SEE=117.8

Singh commented "We may have a problem with parameter instability if the relationship between gold prices and jewelry costs has changed over the past 30 years." Baker computes the test statistic and concluded that "We fail to reject the null hypothesis that the slope coefficient is equal to 4.0 at the 5% significance level." Are Singh (Statement 1) and Baker (Statement 2) correct or incorrect regarding the usefulness of regression results described in Exhibit 1 and the value of the slope coefficient? Use the excerpt of the t-table below.

df	p = 0.10	p = 0.05	p = 0.025	p = 0.01	p = 0.005
⋮					
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
⋮					

A. Both Singh and Baker: Correct.

B. Both Singh and Baker: Incorrect.

C. Singh: Incorrect; Baker: Correct.

The correct answer is **A.**

Both Singh and Baker's statements are correct. The data for regression analysis pertains to a period of more than 30 years, and during this period, the relationship between gold prices and jewelry costs could have changed. This would create parameter instability a regression limitation.

Test statistic is given by:

ˆ

$$\frac{\hat{b}_1 - b_1}{S_{\hat{b}_1}} = \frac{2.987 - 4.0}{0.615} = -1.793$$

The critical value (t-value at 29 dfs and alpha = 0.025) is 2.045.

Our test statistic lies within the non-rejection region (±2.045). We therefore have insufficient evidence to reject the null hypothesis that the slope coefficient is equal to 4.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

Q.3929 Xander Feng, CFA, is a quantitative analyst with Red Star Securities Ltd. Feng is forecasting quarterly sales of Xiomi Inc., a smart phone manufacturer based in China. The regression model is:

$$\text{Sales}_t = b_0 + b_1 \text{Sales}_{t-1} + \epsilon_t$$

The regression results for the smartphone sales model are presented in the exhibits below:

Exhibit 1: Regression statistics for smartphones sales model

	Coefficient	Standard Error
Intercept	313.24	99.43
Lag 1	0.67	0.16

R-squared: 0.7436
Observations: 120

If smartphone sales in first quarter were 1,137, the number of smartphone sales forecasted for the second quarter is *closest to*:

A. 762.

B. 1,075.

C. 1,137.

The correct answer is **B.**

The model is correctly specified. Hence, quarterly sales for Xiomi can be forecasted using this model:

$$\text{Sales}_t = b_0 + b_1\text{Sales}_{t-1} + \epsilon_t = 313.24 + (0.67 \times 1,137) = 1,075.03$$

A is incorrect. It suggests a forecasted sales figure of 762.0, which does not align with the calculation based on the regression model provided.

C is incorrect. It suggests that the number of smartphone sales forecasted for the second quarter is the same as the first quarter, 1,137. This ignores the effect of the regression model, which incorporates both a constant term and a coefficient for the previous quarter's sales to forecast future sales. The regression model indicates that sales are expected to change from quarter to quarter based on the model's parameters, and thus, maintaining the same sales figure does not reflect the model's forecast.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

Q.3931 In which of the following functional forms the dependent variable is linear but the independent variable is logarithmic?

A. The Lin-log model and will be represented as $Y_i = b_0 + b_1 \ln X_i$.

B. The Log-lin model will be represented as $\ln Y_i = b_0 + b_1 X_i$

C. The Log-log model and will be represented as $\ln Y_i = b_0 + b_1 \ln X_i$.

The correct answer is **A.**

Lin-log model: The dependent variable is linear but the independent variable is logarithmic. It is represented as $Y_i = b_0 + b_1 \ln X_i$

B is incorrect. Log-lin model: The dependent variable is logarithmic but the independent variable is linear. It is represented as $\ln Y_i = b_0 + b_1 X_i$

C is incorrect. Log-log model: Both the dependent and independent variables are in logarithmic form. It is represented as $\ln Y_i = b_0 + b_1 \ln X_i$

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (f): Describe different functional forms of simple linear regressions.

Q.3932 In the log-lin model, which of the following statement (s) is *most likely* correct about the slope coefficient?

A. The slope coefficient in the log-lin model provides the absolute change in the dependent variable for a relative change in the independent variable.

B. The slope coefficient in the log-lin model is the relative change in the dependent variable for an absolute change in the independent variable.

C. The slope coefficient in the log-lin model is the relative change in the dependent variable for a relative change in the independent variable.

The correct answer is **B.**

The slope coefficient in the log-lin model is the relative change in the dependent variable for an absolute change in the independent variable. However, because the model involves logarithms, the interpretation of the slope coefficient is based on relative changes rather than absolute changes.

A is incorrect. The slope coefficient in the lin-log model provides the absolute change in the dependent variable for a relative change in the independent variable. In a log-linear model, the slope coefficient provides the relative change in the dependent variable for a one-unit change in the independent variable, not the absolute change.

C is incorrect. The slope coefficient in the log-log model is the relative change in the dependent variable for a relative change in the independent variable. In a log-linear model, the slope coefficient represents the relative change in the dependent variable for a one-unit absolute change in the independent variable.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.

Q.3933 An analyst is comparing two regression models to analyze the relationship between auto sales and bank financing rates of a Country. The model which would better represent the relationship would *least likely* have:

A. higher F-statistic.

B. lower coefficient of determination (R^2).

C. lower standard error of estimate (S_e).

The correct answer is **B.**

A model with a high Coefficient of determination R2, high F-statistic, and low standard error of estimate (SEE) is better.

A is incorrect. A higher F-statistic is indicative of a statistically significant relationship between the dependent and independent variables in the model. The F-statistic tests the null hypothesis that all regression coefficients are equal to zero, meaning no relationship exists between the dependent and independent variables. A higher F-statistic value leads to the rejection of this null hypothesis, suggesting that the model has at least one significant predictor variable. Therefore, a model with a higher F-statistic is generally considered to provide a better representation of the relationship between variables.

C is incorrect. The standard error of estimate (S_e) measures the average distance that the observed values fall from the regression line. A lower S_e indicates that the model has a tighter fit to the observed data, as the observed values are closer to the predicted values. This implies that the model with a lower S_e would more accurately represent the relationship between auto sales and bank financing rates, as it suggests less dispersion of the observed values around the fitted values. Therefore, a model with a lower S_e is preferable for accurately representing the relationship between variables.

CFA Level I, Quantitative Methods, Learning Module 10: Simple Linear Regression. LOS (c): Calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression.
