

Question #1 of 119

A sample of paired points A and B is shown below. What is the covariance between the values of A and B?

Sample	A	B
1	1	2
2	4	5
3	9	7
4	11	10
5	14	12

A) 7.80.

B) 29.76.

C) 20.55.

Question #2 of 119

Consider a sample of 32 observations on variables X and Y in which the correlation is 0.30. If the level of significance is 5%, we:

A) conclude that there is no significant correlation between X and Y.

B) cannot test the significance of the correlation with this information.

C) conclude that there is significant correlation between X and Y.

Question #3 of 119

Regression analysis has a number of assumptions. Violations of these assumptions include which of the following?

A) Residuals that are not normally distributed.

- B)** Independent variables that are not normally distributed.
- C)** A zero mean of the residuals.

Question #4 of 119

Suppose the covariance between Y and X is 12, the variance of Y is 25, and the variance of X is 36. What is the correlation coefficient (r), between Y and X?

- A)** 0.013.
- B)** 0.400.
- C)** 0.160.

Rebecca Anderson, CFA, has recently accepted a position as a financial analyst with Eagle Investments. She will be responsible for providing analytical data to Eagle's portfolio manager for several industries. In addition, she will follow each of the major public corporations within each of those industries. As one of her first assignments, Anderson has been asked to provide a detailed report on one of Eagle's current investments. She was given the following data on sales for Company XYZ, the maker of toilet tissue, as well as toilet tissue industry sales (\$ millions). She has been asked to develop a model to aid in the prediction of future sales levels for Company XYZ. She proceeds by recalling some of the basic concepts of regression analysis she learned while she was preparing for the CFA exam.

Year	Industry Sales (X)	Company Sales (Y)	(X-X) ²
1	\$3,000	\$750	84,100
2	\$3,200	\$800	8,100
3	\$3,400	\$850	12,100
4	\$3,350	\$825	3,600
5	\$3,500	\$900	44,100
Totals	\$16,450	\$4,125	152,000

Coefficient Estimates			
Predictor	Coefficient	Stand. Error of the Coefficient	t-statistic

Intercept	-94.88	32.97	??
Slope (Industry Sales)	0.2796	0.0363	??

Analysis of Variance Table (ANOVA)				
Source	df (Degrees of Freedom)	SS (Sum of Squares)	Mean Square (SS/df)	F-statistic
Regression	1 (# of independent variables)	11,899.50 (SSR)	11,899.50 (MSR)	59.45
Error	3 (n-2)	600.50 (SSE)	200.17 (MSE)	
Total	4 (n-1)	12,500 (SS Total)		

Abbreviated Two-tailed t-table		
df	10%	5%
2	2.920	4.303
3	2.353	3.182
4	2.132	2.776

Standard error of forecast is 15.5028.

Question #5 of 119

Which of the following is the correct value of the correlation coefficient between industry sales and company sales?

- A) 0.9062.
- B) 0.9757.
- C) 0.2192.

Question #6 of 119

Which of the following reports the correct value and interpretation of the R^2 for this regression?

The R^2 is:

- A)** 0.952, indicating that the variability of industry sales explains about 95.2% of the variability of company sales.
 - B)** 0.048, indicating that the variability of industry sales explains about 4.8% of the variability of company sales.
 - C)** 0.952, indicating the variability of company sales explains about 95.2% of the variability of industry sales.
-

Question #7 of 119

What is the predicted value for sales of Company XYZ given industry sales of \$3,500?

- A)** \$900.00.
 - B)** \$883.72.
 - C)** \$994.88.
-

Question #8 of 119

What is the upper limit of a 95% confidence interval for the predicted value of company sales (Y) given industry sales of \$3,300?

- A)** 318.42.
 - B)** 877.13.
 - C)** 827.87.
-

Question #9 of 119

What is the lower limit of a 95% confidence interval for the predicted value of company sales (Y) given industry sales of \$3,300?

- A) 1,337.06.
 - B) 778.47.
 - C) 827.80.
-

Question #10 of 119

What is the t -statistic for the slope of the regression line?

- A) 3.1820.
 - B) 2.9600.
 - C) 7.7025.
-

Question #11 of 119

Limitations of regression analysis include all of the following EXCEPT:

- A) parameter instability.
 - B) regression results do not indicate anything about economic significance.
 - C) outliers may affect the estimated regression line.
-

Question #12 of 119

Consider the following estimated regression equation:

$$\text{AUTO}_t = 0.89 + 1.32 \text{PI}_t$$

The standard error of the coefficient is 0.42 and the number of observations is 22. The 95% confidence interval for the slope coefficient, b_1 , is:

- A) $\{0.444 < b_1 < 2.196\}$.
- B) $\{-0.766 < b_1 < 3.406\}$.

C) $\{0.480 < b_1 < 2.160\}$.

Question #13 of 119

Assume an analyst performs two simple regressions. The first regression analysis has an R-squared of 0.90 and a slope coefficient of 0.10. The second regression analysis has an R-squared of 0.70 and a slope coefficient of 0.25. Which one of the following statements is *most* accurate?

- A) The influence on the dependent variable of a one unit increase in the independent variable is 0.9 in the first analysis and 0.7 in the second analysis.
 - B) The first regression has more explanatory power than the second regression.
 - C) Results of the second analysis are more reliable than the first analysis.
-

Question #14 of 119

Consider the regression results from the regression of Y against X for 50 observations:

$$Y = 5.0 + 1.5 X$$

The standard error of the coefficient is 0.50 and the standard error of the forecast is 0.52. The 95% confidence interval for the predicted value of Y if X is 10 is:

- A) $\{18.980 < Y < 21.019\}$.
 - B) $\{18.954 < Y < 21.046\}$.
 - C) $\{19.480 < Y < 20.052\}$.
-

Question #15 of 119

One hazard of the correlation analysis of two random variables is when two variables are highly correlated, but no economic relationship exists. This condition *most* likely indicates the presence of:

- A) nonlinear relationships.
 - B) multicollinearity.
 - C) spurious correlation.
-

Question #16 of 119

Bea Carroll, CFA, has performed a regression analysis of the relationship between two economic measures. Her analysis indicates a standard error of estimate (SEE) that is high relative to total variability. Which of the following conclusions regarding the relationship between the two economic measures can Carroll most accurately draw from her SEE analysis? The relationship between the two variables is *most likely*:

- A) logarithmic.
 - B) linear and inverse.
 - C) very weak.
-

Question #17 of 119

Which of the following statements about linear regression is *least* accurate?

- A) The correlation coefficient, ρ of two assets x and $y = (\text{covariance}_{x,y}) \times \text{standard deviation}_x \times \text{standard deviation}_y$.
 - B) $R^2 = \text{RSS} / \text{SST}$.
 - C) The independent variable is uncorrelated with the residuals (or disturbance term).
-

Question #18 of 119

A simple linear regression equation had a coefficient of determination (R^2) of 0.8. What is the correlation coefficient between the dependent and independent variables and what is the covariance between the two variables if the variance of the independent variable is 4 and the variance of the dependent variable is 9?

	<u>Correlation coefficient</u>	<u>Covariance</u>
A)	0.89	5.34
B)	0.91	4.80
C)	0.89	4.80

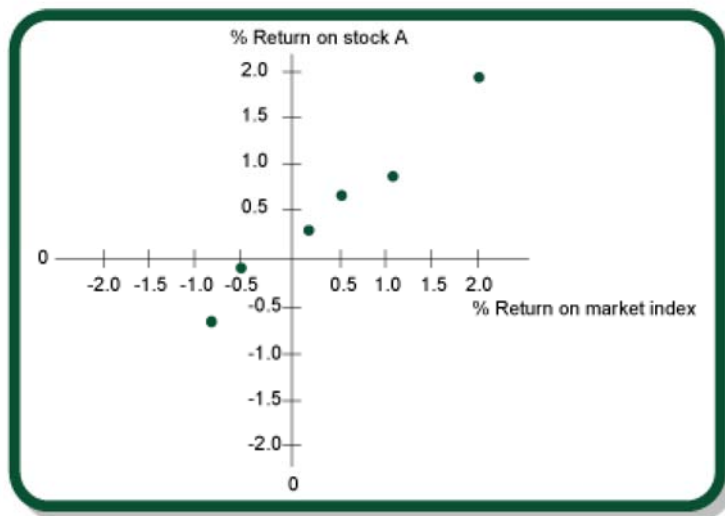
Question #19 of 119

Thomas Manx is attempting to determine the correlation between the number of times a stock quote is requested on his firm's website and the number of trades his firm actually processes. He has examined samples from several days trading and quotes and has determined that the covariance between these two variables is 88.6, the standard deviation of the number of quotes is 18, and the standard deviation of the number of trades processed is 14. Based on Manx's sample, what is the correlation between the number of quotes requested and the number of trades processed?

- A) 0.18.
 - B) 0.78.
 - C) 0.35.
-

Question #20 of 119

In the scatter plot below, the correlation between the return on stock A and the market index is:



- A) positive.
- B) negative.
- C) not discernable using the scatter plot.

Question #21 of 119

Unlike the coefficient of determination, the coefficient of correlation:

- A) indicates whether the slope of the regression line is positive or negative.
- B) indicates the percentage of variation explained by a regression model.
- C) measures the strength of association between the two variables more exactly.

Question #22 of 119

Which of the following statements regarding the coefficient of determination is *least* accurate?

The coefficient of determination:

- A) may range from -1 to $+1$.
- B) cannot decrease as independent variables are added to the model.

C) is the percentage of the total variation in the dependent variable that is explained by the independent variable.

Question #23 of 119

Consider the following analysis of variance (ANOVA) table:

Source	Sum of squares	Degrees of freedom	Mean square
Regression	200	1	200
Error	400	40	10
Total	600	41	

The R^2 and the F-statistic are, respectively:

- A) $R^2 = 33\%$ and $F = 20.0$.
 - B) $R^2 = 33\%$ and $F = 2.0$.
 - C) $R^2 = 50\%$ and $F = 2.0$.
-

Question #24 of 119

The standard error of the estimate measures the variability of the:

- A) values of the sample regression coefficient.
 - B) predicted y-values around the mean of the observed y-values.
 - C) actual dependent variable values about the estimated regression line.
-

Question #25 of 119

Which term is *least likely* to apply to a regression model?

- A) Coefficient of variation.

- B) Coefficient of determination.
 - C) Goodness of fit.
-

Question #26 of 119

Which of the following statements about linear regression analysis is *most* accurate?

- A) An assumption of linear regression is that the residuals are independently distributed.
 - B) When there is a strong relationship between two variables we can conclude that a change in one will cause a change in the other.
 - C) The coefficient of determination is defined as the strength of the linear relationship between two variables.
-

Question #27 of 119

The *most* appropriate measure of the degree of variability of the actual Y -values relative to the estimated Y -values from a regression equation is the:

- A) coefficient of determination (R^2).
 - B) standard error of the estimate (SEE).
 - C) sum of squared errors (SSE).
-

Question #28 of 119

If the correlation between two variables is -1.0 , the scatter plot would appear along a:

- A) straight line running from northwest to southeast.
 - B) straight line running from southwest to northeast.
 - C) a curved line running from southwest to northeast.
-

Question #29 of 119

We are examining the relationship between the number of cold calls a broker makes and the number of accounts the firm as a whole opens. We have determined that the correlation coefficient is equal to 0.70, based on a sample of 16 observations. Is the relationship statistically significant at a 10% level of significance, why or why not? The relationship is:

- A) not significant; the critical value exceeds the t-statistic by 1.91.
 - B) significant; the t-statistic exceeds the critical value by 1.91.
 - C) significant; the t-statistic exceeds the critical value by 3.67.
-

Question #30 of 119

The capital asset pricing model is given by: $R_i = R_f + \text{Beta} (R_m - R_f)$ where R_m = expected return on the market, R_f = risk-free market and R_i = expected return on a specific firm. The dependent variable in this model is:

- A) R_i .
 - B) $R_m - R_f$.
 - C) R_f .
-

Cynthia Jones is Director of Marketing at Vancouver Industries, a large producer of athletic apparel and accessories. Approximately three years ago, Vancouver experienced increased competition in the marketplace, and consequently sales for that year declined nearly 20%. At that time, Jones proposed a new marketing campaign for the company, aimed at positioning Vancouver's product lines toward a younger target audience. Although the new marketing effort was significantly more costly than previous marketing campaigns, Jones assured her superiors that the resulting increase in sales would more than justify the additional expense. Jones was given approval to proceed with the implementation of her plan.

Three years later, in preparation for an upcoming shareholders meeting, the CEO of Vancouver has asked Jones for an evaluation of the marketing campaign. Sales have increased since the inception of the new marketing campaign nearly three years ago, but the CEO is questioning whether the increase is due to the marketing expenditures or can be attributed to other

factors. Jones is examining the following data on the firm's aggregate revenue and marketing expenditure over the last 10 quarters. Jones plans to demonstrate the effectiveness of marketing in boosting sales revenue. She chooses to utilize a simple linear regression model. The output is as follows:

	Aggregate Revenue (Y)	Advertising Expenditure (X)	Y ²	XY	X ²
	300	7.5	90,000	2,250	56.25
	320	9.0	102,400	2,880	81.00
	310	8.5	96,100	2,635	72.25
	335	8.2	112,225	2,747	67.24
	350	9.0	122,500	3,150	81.00
	400	8.5	160,000	3,400	72.25
	430	10.0	184,900	4,300	100.00
	390	10.5	152,100	4,095	110.25
	380	9.0	144,400	3,420	81.00
	430	11.0	184,900	4,730	121.00
TOTAL	3,645	91.2	1,349,525	33,607	842.24

Slope coefficient = 34.74 Standard error of slope coefficient = 9.916629313 Standard error of intercept = 92.2840128

ANOVA			
	Df	SS	MS
Regression	1	12,665.125760	12,665.12576
Residual	8	8,257.374238	1,032.17178
Total	9	20,922.5	

Jones discusses her findings with her market research specialist, Mira Nair. Nair tells Jones that she should check her model for heteroskedasticity. She explains that in the presence of conditional heteroskedasticity, the model coefficients and t-statistics will be biased.

For the questions below, assume a t-value of 2.306.

Question #31 of 119

Which of the following is *closest* to the upper limit of the 95% confidence interval for the slope coefficient?

- A) 62.84.
 - B) 57.61.
 - C) 111.72.
-

Question #32 of 119

Which of the following is *closest* to the lower limit of the 95% confidence interval for the slope coefficient?

- A) 72.84.
 - B) 11.87.
 - C) 12.24.
-

Question #33 of 119

Which of the following is the CORRECT value of the correlation coefficient between aggregate revenue and advertising expenditure?

- A) 0.9500.
 - B) 0.7780.
 - C) 0.6053.
-

Question #34 of 119

Which of the following reports the CORRECT value and interpretation of the R^2 for this regression? The R^2 is:

- A) 0.3947 indicating that the variability of aggregate revenue explains about 39.47% of the variability in advertising expenditure.

- B)** 0.3947 indicating that the variability of advertising expenditure explains about 39.47% of the variability of aggregate revenue.
- C)** 0.6053 indicating that the variability of advertising expenditure explains about 60.53% of the variability in aggregate revenue.
-

Question #35 of 119

Is Nair's statement about conditional heteroskedasticity CORRECT?

- A)** No, because the t-statistics will not be biased.
- B)** Yes, because both the coefficients and t-statistics will be biased.
- C)** No, because coefficients will not be biased.
-

Question #36 of 119

What is the calculated F-statistic?

- A)** 0.1250.
- B)** 12.2700.
- C)** 92.2840.
-

Question #37 of 119

Consider the case when the Y variable is in U.S. dollars and the X variable is in U.S. dollars. The 'units' of the covariance between Y and X are:

- A)** a range of values from -1 to +1.
- B)** U.S. dollars.
- C)** squared U.S. dollars.
-

Question #38 of 119

A study of 40 men finds that their job satisfaction and marital satisfaction scores have a correlation coefficient of 0.52. At 5% level of significance, is the correlation coefficient significantly different from 0?

- A) Yes, $t = 3.76$.
 - B) No, $t = 2.02$.
 - C) No, $t = 1.68$.
-

Question #39 of 119

A dependent variable is regressed against a single independent variable across 100 observations. The mean squared error is 2.807, and the mean regression sum of squares is 117.9. What is the correlation coefficient between the two variables?

- A) 0.30.
 - B) 0.99.
 - C) 0.55.
-

Question #40 of 119

Which of the following statements about the standard error of estimate is *least* accurate? The standard error of estimate:

- A) is the square of the coefficient of determination.
 - B) measures the Y variable's variability that is not explained by the regression equation.
 - C) is the square root of the sum of the squared deviations from the regression line divided by $(n - 2)$.
-

Question #41 of 119

Which of the following statements regarding scatter plots is *most* accurate? Scatter plots:

- A) illustrate the scatterings of a single variable.
 - B) are used to examine the third moment of a distribution (skewness).
 - C) illustrate the relationship between two variables.
-

Question #42 of 119

Which of the following statements regarding a correlation coefficient of 0.60 for two variables Y and X is *most* accurate? This correlation:

- A) is significantly different from zero.
 - B) indicates a positive covariance between the two variables.
 - C) indicates a positive causal relation between the two variables.
-

Question #43 of 119

If X and Y are perfectly correlated, regressing Y onto X will result in which of the following:

- A) the regression line will be sloped upward.
 - B) the standard error of estimate will be zero.
 - C) the alpha coefficient will be zero.
-

Question #44 of 119

The table below shows a sample of returns on two securities:

Period	1	2	3	4	Mean
Security P	0.2%	0.5%	1.1%	-0.6%	0.3%
Security Q	-0.3%	0.9%	1.5%	-0.5%	0.4%

The sample covariance between the two securities' returns is *closest* to:

- A) 0.62.
- B) 0.78.
- C) 0.47.

Milky Way, Inc. is a large manufacturer of children's toys and games based in the United States. Their products have high name brand recognition, and have been sold in retail outlets throughout the United States for nearly fifty years. The founding management team was bought out by a group of investors five years ago. The new management team, led by Russell Stepp, decided that Milky Way should try to expand its sales into the Western European market, which had never been tapped by the former owners. Under Stepp's leadership, additional personnel are hired in the Research and Development department, and a new marketing plan specific to the European market is implemented. Being a new player in the European market, Stepp knows that it will take several years for Milky Way to establish its brand name in the marketplace, and is willing to make the expenditures now in exchange for increased future profitability.

Now, five years after entering the European market, Stepp is reviewing the results of his plan. Sales in Europe have slowly but steadily increased over since Milky Way's entrance into the market, but profitability seems to have leveled out. Stepp decides to hire a consultant, Ann Hays, CFA, to review and evaluate their European strategy. One of Hays' first tasks on the job is to perform a regression analysis on Milky Way's European sales. She is seeking to determine whether the additional expenditures on research and development and marketing for the European market should be continued in the future.

Hays begins by establishing a relationship between the European sales of Milky Way (in millions of dollars) and the two independent variables, the number of dollars (in millions) spent on research and development (R&D) and marketing (MKTG). Based upon five years of monthly data, Hays constructs the following estimated regression equation:

$$\text{Estimated Sales} = 54.82 + 5.97 (\text{MKTG}) + 1.45 (\text{R\&D})$$

Additionally, Hays calculates the following regression estimates:

	Coefficient	Standard Error
Intercept	54.82	3.165
MKTG	5.97	1.825
R&D	1.45	0.987

Question #45 of 119

Hays begins the analysis by determining if both of the independent variables are statistically significant. To test whether a coefficient is statistically significant means to test whether it is statistically significantly different from:

- A) zero.
 - B) the upper tail critical value.
 - C) slope coefficient.
-

Question #46 of 119

The t -statistic for the marketing variable is calculated to be:

- A) 1.886.
 - B) 3.271.
 - C) 17.321.
-

Question #47 of 119

Hays formulates a test structure where the decision rule is to reject the null hypothesis if the calculated test statistic is either larger than the upper tail critical value or lower than the lower tail critical value. At a 5% significance level with 57 degrees of freedom, assume that the two-tailed critical t -values are $t_c = \pm 2.004$. Based on this information, Hays makes the following conclusions:

- Point 1: The intercept term is statistically significant.
- Point 2: Both independent variables are statistically significant in the model explaining sales for Milky Way, Inc.
- Point 3: If an F -test were being used, the null hypothesis would be rejected.

Which of Hays' conclusions are CORRECT?

- A) Points 1 and 3.
- B) Points 1 and 2.

C) Points 2 and 3.

Question #48 of 119

Hays is aware that part, but not all, of the total variation in expected sales can be explained by the regression equation. Which of the following statements correctly reflects this relationship?

- A) $SST = RSS + SSE + MSE$.
 - B) $MSE = RSS + SSE$.
 - C) $SST = RSS + SSE$.
-

Question #49 of 119

Hays decides to test the overall effectiveness of the both independent variables in explaining sales for Milky Way. Assuming that the total sum of squares is 389.14, the sum of squared errors is 146.85 and the mean squared error is 2.576, then:

- A) The R^2 equals 0.242, indicating that the two independent variables account for 24.2% of the variation in monthly sales.
 - B) The R^2 equals 0.623, indicating that the two independent variables together account for 37.7% of the variation in monthly sales.
 - C) The correlation between the actual and predicted values of estimated sales is 0.79.
-

Question #50 of 119

Stepp is concerned about the validity of Hays' regression analysis and asks Hays if he can test for the presence of heteroskedasticity. Hays complies with Stepp's request, and detects the presence of unconditional heteroskedasticity. Which of the following statements regarding heteroskedasticity is *most correct*?

- A) Heteroskedasticity can be detected either by examining scatter plots of the residual or by using the Durbin-Watson test.

- B)** Unconditional heteroskedasticity does create significant problems for statistical inference.
 - C)** Unconditional heteroskedasticity usually causes no major problems with the regression.
-

Question #51 of 119

Wanda Brunner, CFA, is working on a regression analysis based on publicly available macroeconomic time-series data. The most important limitation of regression analysis in this instance is:

- A)** low confidence intervals.
 - B)** limited usefulness in identifying profitable investment strategies.
 - C)** the error term of one observation is not correlated with that of another observation.
-

Question #52 of 119

An analyst is examining the relationship between two random variables, RCRANTZ and GSTERN. He performs a linear regression that produces an estimate of the relationship:

$$\text{RCRANTZ} = 61.4 - 5.9\text{GSTERN}$$

Which interpretation of this regression equation is *least* accurate?

- A)** The intercept term implies that if GSTERN is zero, RCRANTZ is 61.4.
 - B)** The covariance of RCRANTZ and GSTERN is negative.
 - C)** If GSTERN increases by one unit, RCRANTZ should increase by 5.9 units.
-

Question #53 of 119

The *most* appropriate test statistic to test statistical significance of a regression slope coefficient with 45 observations and 2 independent variables is a:

- A)** one-tail *t*-statistic with 42 degrees of freedom.

B) one-tail t -statistic with 43 degrees of freedom.

C) two-tail t -statistic with 42 degrees of freedom.

A study of a sample of incomes (in thousands of dollars) of 35 individuals shows that income is related to age and years of education. The following table shows the regression results:

	Coefficient	Standard Error	t-statistic	P-value
Intercept	5.65	1.27	4.44	0.01
Age	0.53	?	1.33	0.21
Years of Education	2.32	0.41	?	0.01
<i>Anova</i>	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	?	215.10	?	?
Error	?	115.10	?	
Total	?	?		

Question #54 of 119

The standard error for the coefficient of age and t -statistic for years of education are:

A) 0.40; 5.66.

B) 0.53; 2.96.

C) 0.32; 1.65.

Question #55 of 119

The mean square regression (MSR) is:

A) 102.10.

B) 6.72.

C) 107.55.

Question #56 of 119

The mean square error (MSE) is:

- A) 3.60.
 - B) 3.58.
 - C) 7.11.
-

Question #57 of 119

What is the R^2 for the regression?

- A) 76%.
 - B) 65%.
 - C) 62%.
-

Question #58 of 119

What is the predicted income of a 40-year-old married female with 16 years of education and 18 years of work experience?

- A) \$63,970.
 - B) \$82,706.
 - C) \$106,930.
-

Question #59 of 119

What is the F-value?

- A) 29.88.

- B) 1.88.
 - C) 14.36.
-

Question #60 of 119

An analyst performs two simple regressions. The first regression analysis has an R-squared of 0.40 and a beta coefficient of 1.2. The second regression analysis has an R-squared of 0.77 and a beta coefficient of 1.75. Which one of the following statements is *most* accurate?

- A) The R-squared of the first regression indicates that there is a 0.40 correlation between the independent and the dependent variables.
 - B) The second regression equation has more explaining power than the first regression equation.
 - C) The first regression equation has more explaining power than the second regression equation.
-

Question #61 of 119

A sample covariance of two random variables is *most likely* to be used to:

- A) demonstrate that the value of one of the variables determines the value of the other.
 - B) calculate the correlation coefficient, which is a measure of the strength of the variables' linear relationship.
 - C) identify and measure the strength of linear and nonlinear relationships between the two variables.
-

Question #62 of 119

The Y variable is regressed against the X variable resulting in a regression line that is horizontal with the plot of the paired observations widely dispersed about the regression line. Based on this information, which statement is *most likely* accurate?

- A) X is perfectly positively correlated to Y.
 - B) The R^2 of this regression is close to 100%.
 - C) The correlation between X and Y is close to zero.
-

Question #63 of 119

In order to have a negative correlation between two variables, which of the following is *most* accurate?

- A) Either the covariance or one of the standard deviations must be negative.
 - B) The covariance must be negative.
 - C) The covariance can never be negative.
-

Question #64 of 119

Determine and interpret the correlation coefficient for the two variables X and Y. The standard deviation of X is 0.05, the standard deviation of Y is 0.08, and their covariance is -0.003 .

- A) -1.33 and the two variables are negatively associated.
 - B) $+0.75$ and the two variables are positively associated.
 - C) -0.75 and the two variables are negatively associated.
-

Question #65 of 119

A simple linear regression is run to quantify the relationship between the return on the common stocks of medium sized companies (Mid Caps) and the return on the S&P 500 Index, using the monthly return on Mid Cap stocks as the dependent variable and the monthly return on the S&P 500 as the independent variable. The results of the regression are shown below:

	Coefficient	Standard Error of coefficient	t-Value
Intercept	1.71	2.950	0.58
S&P 500	1.52	0.130	11.69
$R^2 = 0.599$			

The strength of the relationship, as measured by the correlation coefficient, between the return on Mid Cap stocks and the return on the S&P 500 for the period under study was:

- A) 0.774.
- B) 0.599.
- C) 0.130.

Question #66 of 119

Consider the following analysis of variance (ANOVA) table:

Source	Sum of squares	Degrees of freedom	Mean square
Regression	500	1	500
Error	750	50	15
Total	1,250	51	

The R^2 and the F-statistic are, respectively:

- A) $R^2 = 0.40$ and $F = 33.333$.
- B) $R^2 = 0.40$ and $F = 0.971$.
- C) $R^2 = 0.67$ and $F = 0.971$.

Question #67 of 119

Paul Frank is an analyst for the retail industry. He is examining the role of television viewing by teenagers on the sales of accessory stores. He gathered data and estimated the following regression of sales (in millions of dollars) on the number of hours watched by teenagers (TV, in hours per week):

$$\text{Sales}_t = 1.05 + 1.6 \text{ TV}_t$$

The predicted sales if television watching is 5 hours per week is:

- A) \$9.05 million.
 - B) \$8.00 million.
 - C) \$2.65 million.
-

Question #68 of 119

The assumptions underlying linear regression include all of the following EXCEPT the:

- A) independent variable is linearly related to the residuals (or disturbance term).
 - B) disturbance term is homoskedastic and is independently distributed.
 - C) disturbance term is normally distributed with an expected value of 0.
-

Question #69 of 119

For the case of simple linear regression with one independent variable, which of the following statements about the correlation coefficient is *least* accurate?

- A) If the regression line is flat and the observations are dispersed uniformly about the line, the correlation coefficient will be +1.
 - B) The correlation coefficient can vary between -1 and +1.
 - C) If the correlation coefficient is negative, it indicates that the regression line has a negative slope coefficient.
-

Craig Standish, CFA, is investigating the validity of claims associated with a fund that his company offers. The company advertises the fund as having low turnover and, hence, low management fees. The fund was created two years ago with only a few uncorrelated assets. Standish randomly draws two stocks from the fund, Grey Corporation and Jars Inc., and measures the variances and covariance of their monthly returns over the past two years. The resulting variance covariance matrix is shown below. Standish will test whether it is reasonable to believe that the returns of Grey and Jars are uncorrelated. In doing the analysis, he plans to address the issue of spurious correlation and outliers.

	Grey	Jars
Grey	42.2	20.8
Jars	20.8	36.5

Standish wants to learn more about the performance of the fund. He performs a linear regression of the fund's monthly returns over the past two years on a large capitalization index. The results are below:

ANOVA				
	df	SS	MS	F
Regression	1	92.53009	92.53009	28.09117
Residual	22	72.46625	3.293921	
Total	23	164.9963		

	Coefficients	Standard Error	t-statistic	P-value
Intercept	0.148923	0.391669	0.380225	0.707424
Large Cap Index	1.205602	0.227467	5.30011	2.56E-05

Standish forecasts the fund's return, based upon the prediction that the return to the large capitalization index used in the regression will be 10%. He also wants to quantify the degree of the prediction error, as well as the minimum and maximum sensitivity that the fund actually has with respect to the index.

He plans to summarize his results in a report. In the report, he will also include caveats concerning the limitations of regression analysis. He lists four limitations of regression analysis that he feels are important: relationships between variables can change over time, multicollinearity leads to inconsistent estimates of regression coefficients, if the error terms are

heteroskedastic the standard errors for the regression coefficient may not be reliable, and if the error terms are correlated with each other over time the test statistics may not be reliable.

Question #70 of 119

Given the variance/covariance matrix for Grey and Jars, in a one-sided hypothesis test that the returns are positively correlated $H_0: \rho \leq 0$ vs. $H_1: \rho > 0$, Standish would:

- A) need to gather more information before being able to reach a conclusion concerning significance.
 - B) reject the null at the 5% but not the 1% level of significance.
 - C) reject the null at the 1% level of significance.
-

Question #71 of 119

In using the correlation coefficient between returns on Grey and Jars, Standish would *most* appropriately question the issue of:

- A) spurious correlation but not the issue of outliers.
 - B) issue of outliers but not the issue of spurious correlation.
 - C) Both spurious correlation and outliers.
-

Question #72 of 119

If the large capitalization index has a 10% return, then the forecast of the fund's return will be:

- A) 16.1.
 - B) 13.5.
 - C) 12.2.
-

Question #73 of 119

The standard deviation of monthly fund returns is *closest* to:

- A) 2.68.
 - B) 12.84.
 - C) 7.17.
-

Question #74 of 119

A 95% confidence interval for the slope coefficient is:

- A) 0.905 to 1.506.
 - B) 0.760 to 1.650.
 - C) 0.734 to 1.677.
-

Question #75 of 119

Of the four caveats of regression analysis listed by Standish, the *least* accurate is:

- A) the relationships of variables change over time.
 - B) if the error terms are heteroskedastic the standard errors for the regression coefficients may not be reliable.
 - C) multicollinearity leads to inconsistent estimates of the regression coefficients.
-

Question #76 of 119

Consider the following estimated regression equation:

$$ROE_t = 0.23 - 1.50 CE_t$$

The standard error of the coefficient is 0.40 and the number of observations is 32. The 95% confidence interval for the slope coefficient, b_1 , is:

- A) $\{-2.317 < b_1 < -0.683\}$.

B) $\{0.683 < b_1 < 2.317\}$.

C) $\{-2.300 < b_1 < -0.700\}$.

Question #77 of 119

Suppose the covariance between Y and X is 0.03 and that the variance of Y is 0.04 and the variance of X is 0.12. The sample size is 30. Using a 5% level of significance, which of the following is *most* accurate? The null hypothesis of:

A) no correlation is rejected.

B) no correlation is not rejected.

C) significant correlation is rejected.

Question #78 of 119

Which of the following is *least likely* an assumption of linear regression? The:

A) residuals are independently distributed.

B) expected value of the residuals is zero.

C) residuals are mean reverting; that is, they tend towards zero over time.

Question #79 of 119

Linear regression is based on a number of assumptions. Which of the following is *least* likely an assumption of linear regression?

A) Values of the independent variable are not correlated with the error term.

B) There is at least some correlation between the error terms from one observation to the next.

C) The variance of the error terms each period remains the same.

Question #80 of 119

A simple linear regression is run to quantify the relationship between the return on the common stocks of medium sized companies (Mid Caps) and the return on the S&P 500 Index, using the monthly return on Mid Cap stocks as the dependent variable and the monthly return on the S&P 500 as the independent variable. The results of the regression are shown below:

	Coefficient	Standard Error of Coefficient	t-Value
Intercept	1.71	2.950	0.58
S&P 500	1.52	0.130	11.69
$R^2 = 0.599$			

Use the regression statistics presented above and assume this historical relationship still holds in the future period. If the expected return on the S&P 500 over the next period were 11%, the expected return on Mid Cap stocks over the next period would be:

- A) 18.4%.
- B) 20.3%.
- C) 33.8%.

Question #81 of 119

What does the R^2 of a simple regression of two variables measure and what calculation is used to equate the correlation coefficient to the coefficient of determination?

R^2 measures: Correlation coefficient

- A) percent of variability of the $R^2 = r \times 2$
- B) percent of variability of the $R^2 = r^2$

- C) percent of variability of the
-

Question #82 of 119

In the estimated regression equation $Y = 0.78 - 1.5 X$, which of the following is *least* accurate when interpreting the slope coefficient?

- A) The dependent variable increases by 1.5 units if X decreases by 1 unit.
B) If the value of X is zero, the value of Y will be -1.5.
C) The dependent variable declines by -1.5 units if X increases by 1 unit.
-

Question #83 of 119

Paul Frank is an analyst for the retail industry. He is examining the role of television viewing by teenagers on the sales of accessory stores. He gathered data and estimated the following regression of sales (in millions of dollars) on the number of hours watched by teenagers (in hours per week):

$$\text{Sales}_t = 1.05 + 1.6 \text{TV}_t$$

Which of the following is the *most* accurate interpretation of the estimated results? If TV watching:

- A) goes up by one hour per week, sales of accessories increase by \$1.6 million.
B) goes up by one hour per week, sales of accessories increase by \$1.60.
C) changes, no change in sales is expected.
-

Question #84 of 119

Which model does not lend itself to correlation coefficient analysis?

- A) $X = Y \times 2$.

B) $Y = X + 2$.

C) $Y = X^3$.

Question #85 of 119

Suppose the covariance between Y and X is 10, the variance of Y is 25, and the variance of X is 64. The sample size is 30. Using a 5% level of significance, which of the following statements is *most* accurate? The null hypothesis of:

A) significant correlation is rejected.

B) no correlation is rejected.

C) no correlation cannot be rejected.

Question #86 of 119

Dan Gates, CFA is forecasting price elasticity of demand for GMX Inc's products. Gates used monthly revenues for the past four years as the dependent variable (\$ millions) and price per unit as the independent variable. The results are shown below.

$$\text{Sales} = 23.45 - 0.6 \text{ Price}$$

$$\text{Standard error (intercept)} = 10.22$$

$$\text{Standard error (slope)} = 0.03$$

$$\text{Standard error of estimate} = 8.32$$

$$\text{Standard error of forecast} = 8.93$$

The 95% confidence interval for predicted value of monthly sales given price was \$2.00 per unit is *closest* to:

A) \$12 million to \$33 million.

B) \$6 million to \$39 million.

C) \$4 million to \$40 million.

Question #87 of 119

Which statement is *most accurate*? Assume a 5% level of significance. The F-statistic is:

Analysis of Variance Table (ANOVA)				
Source	Degrees of freedom (df)	Sum of Squares	Mean Square (SS/df)	F-statistic
Regression	5	18,500	3,700	
Error	94	600.66	6.39	
Total	99	19,100.66		

- A) 579.03 and the regression is said to be statistically significant.
- B) 579.03 and the regression is said to be statistically insignificant.
- C) 0.0017 and the regression is said to be statistically significant.

Question #88 of 119

Sera Smith, a research analyst, had a hunch that there was a relationship between the percentage change in a firm's number of salespeople and the percentage change in the firm's sales during the following period. Smith ran a regression analysis on a sample of 50 firms, which resulted in a slope of 0.72, an intercept of +0.01, and an R^2 value of 0.65. Based on this analysis, if a firm made no changes in the number of sales people, what percentage change in the firm's sales during the following period does the regression model predict?

- A) +0.65%.
- B) +0.72%.
- C) +1.00%.

Question #89 of 119

Joe Harris is interested in why returns on equity differ from one company to another. He has chosen several company-specific variables to explain the return on equity, including financial leverage and capital expenditures. In his model:

- A) financial leverage is an independent variable.
 - B) capital expenditure is a dependent variable.
 - C) return on equity is a control variable.
-

Question #90 of 119

Consider the regression results from the regression of Y against X for 50 observations:

$$Y = 0.78 + 1.2 X$$

The standard error of the estimate is 0.40 and the standard error of the coefficient is 0.45.

Which of the following reports the correct value of the *t*-statistic for the slope and correctly evaluates its statistical significance with 95% confidence?

- A) $t = 2.667$; slope is significantly different from zero.
 - B) $t = 1.789$; slope is not significantly different from zero.
 - C) $t = 3.000$; slope is significantly different from zero.
-

Question #91 of 119

Ron James, CFA, computed the correlation coefficient for historical oil prices and the occurrence of a leap year and has identified a statistically significant relationship. Specifically, the price of oil declined every fourth calendar year, all other factors held constant. James has *most* likely identified which of the following conditions in correlation analysis?

- A) Spurious correlation.
 - B) Time trends.
 - C) Seasonality.
-

Question #92 of 119

Consider a sample of 60 observations on variables X and Y in which the correlation is 0.42. If the level of significance is 5%, we:

- A) conclude that there is no significant correlation between X and Y.
 - B) conclude that there is statistically significant correlation between X and Y.
 - C) cannot test the significance of the correlation with this information.
-

Question #93 of 119

Which of the following statements about covariance and correlation is *least* accurate?

- A) The covariance and correlation are always the same sign, positive or negative.
 - B) There is no relation between the sign of the covariance and the correlation.
 - C) A zero covariance implies a zero correlation.
-

Question #94 of 119

Consider the following analysis of variance (ANOVA) table:

Source	Sum of squares	Degrees of freedom	Mean square
Regression	556	1	556
Error	679	50	13.5
Total	1,235	51	

The R^2 for this regression is:

- A) 0.45.
 - B) 0.55.
 - C) 0.82.
-

Question #95 of 119

Consider the regression results from the regression of Y against X for 50 observations:

$$Y = 5.0 - 1.5 X$$

The standard error of the estimate is 0.40 and the standard error of the coefficient is 0.45. The predicted value of Y if X is 10 is:

- A) -10
 - B) 10
 - C) 20
-

Question #96 of 119

Rafael Garza, CFA, is considering the purchase of ABC stock for a client's portfolio. His analysis includes calculating the covariance between the returns of ABC stock and the equity market index. Which of the following statements regarding Garza's analysis is *most* accurate?

- A) A covariance of +1 indicates a perfect positive covariance between the two variables.
 - B) The covariance measures the strength of the linear relationship between two variables.
 - C) The actual value of the covariance is not very meaningful because the measurement is very sensitive to the scale of the two variables.
-

Question #97 of 119

The independent variable in a regression equation is called all of the following EXCEPT:

- A) predicting variable.
 - B) explanatory variable.
 - C) predicted variable.
-

Question #98 of 119

The purpose of regression is to:

- A) get the largest R^2 possible.
 - B) explain the variation in the dependent variable.
 - C) explain the variation in the independent variable.
-

Question #99 of 119

Consider the regression results from the regression of Y against X for 50 observations:

$$Y = 0.78 - 1.5 X$$

The standard error of the estimate is 0.40 and the standard error of the coefficient is 0.45.

Which of the following reports the correct value of the t -statistic for the slope and correctly evaluates $H_0: b_1 \geq 0$ versus $H_a: b_1 < 0$ with 95% confidence?

- A) $t = 3.750$; slope is significantly different from zero.
 - B) $t = -3.333$; slope is significantly negative.
 - C) $t = -3.750$; slope is significantly different from zero.
-

Question #100 of 119

Which of the following is *least likely* an assumption of linear regression?

- A) The residuals are normally distributed.
 - B) The independent variable is correlated with the residuals.
 - C) The variance of the residuals is constant.
-

Question #101 of 119

Consider the following analysis of variance (ANOVA) table:

Source	Sum of squares	Degrees of freedom	Mean square
Regression	550	1	550.000
Error	750	38	19.834
Total	1,300	39	

The F-statistic for the test of the fit of the model is *closest* to:

- A) 27.730.
 - B) 0.423.
 - C) 0.965.
-

Question #102 of 119

Assume you perform two simple regressions. The first regression analysis has an R-squared of 0.80 and a beta coefficient of 0.10. The second regression analysis has an R-squared of 0.80 and a beta coefficient of 0.25. Which one of the following statements is *most* accurate?

- A) The influence on the dependent variable of a one-unit increase in the independent variable is the same in both analyses.
 - B) Results from the first analysis are more reliable than the second analysis.
 - C) Explained variability from both analyses is equal.
-

Question #103 of 119

A sample covariance for the common stock of the Earth Company and the S&P 500 is -9.50. Which of the following statements regarding the estimated covariance of the two variables is *most* accurate?

- A) The two variables will have a strong tendency to move in opposite directions.
- B) The relationship between the two variables is not easily predicted by the calculated covariance.

C) The two variables will have a slight tendency to move together.

Erica Basenj, CFA, has been given an assignment by her boss. She has been requested to review the following regression output to answer questions about the relationship between the monthly returns of the Toffee Investment Management (TIM) High Yield Bond Fund and the returns of the index (independent variable).

Regression Statistics					
R ²	??				
Standard Error	??				
Observations	20				
ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	23,516	23,516	?	?
Residual	18	?	?		
Total	19	23,644			
Regression Equation					
		<i>Coefficients</i>	<i>Std. Error</i>	<i>t-statistic</i>	<i>P-value</i>
Intercept		5.2900	1.6150	?	?
Slope		0.8700	0.0152	?	?

Question #104 of 119

What is the value of the correlation coefficient?

- A) 0.8700.
- B) 0.9973.
- C) -0.9973.

Question #105 of 119

What is the sum of squared errors (SSE)?

- A) 23,515.
 - B) 23,644.
 - C) 128.
-

Question #106 of 119

What is the value of R^2 ?

- A) 0.9471.
 - B) 0.9946.
 - C) 0.0055.
-

Question #107 of 119

Is the intercept term statistically significant at the 5% level of significance and the 1% level of significance, respectively?

- | | <u>1%</u> | <u>5%</u> |
|--------|-----------|-----------|
| A) Yes | Yes | |
| B) No | No | |
| C) Yes | No | |
-

Question #108 of 119

What is the value of the F-statistic?

- A) 0.9945.

- B) 3,359.
 - C) 0.0003.
-

Question #109 of 119

Heteroskedasticity can be defined as:

- A) independent variables that are correlated with each other.
 - B) error terms that are dependent.
 - C) nonconstant variance of the error terms.
-

Question #110 of 119

The R^2 of a simple regression of two factors, A and B, measures the:

- A) statistical significance of the coefficient in the regression equation.
 - B) impact on B of a one-unit change in A.
 - C) percent of variability of one factor explained by the variability of the second factor.
-

Question #111 of 119

A variable Y is regressed against a single variable X across 24 observations. The value of the slope is 1.14, and the constant is 1.3. The mean value of X is 1.10, and the mean value of Y is 2.67. The standard deviation of the X variable is 1.10, and the standard deviation of the Y variable is 2.46. The sum of squared errors is 89.7. For an X value of 1.0, what is the 95% confidence interval for the Y value?

- A) -1.68 to 6.56.
 - B) 0.59 to 4.30.
 - C) -1.83 to 6.72.
-

Question #112 of 119

One of the limitations of correlation analysis of two random variables is the presence of outliers, which can lead to which of the following erroneous assumptions?

- A) The absence of a relationship between the two variables, when in fact, there is a linear relationship.
 - B) The presence of a nonlinear relationship between the two variables, when in fact, there is no relationship whatsoever between the two variables.
 - C) The presence of a nonlinear relationship between the two variables, when in fact, there is a linear relationship.
-

Question #113 of 119

Jason Brock, CFA, is performing a regression analysis to identify and evaluate any relationship between the common stock of ABT Corp and the S&P 100 index. He utilizes monthly data from the past five years, and assumes that the sum of the squared errors is .0039. The calculated standard error of the estimate (SEE) is *closest* to:

- A) 0.0082.
 - B) 0.0360.
 - C) 0.0080.
-

Question #114 of 119

The standard error of estimate is *closest* to the:

- A) standard deviation of the dependent variable.
 - B) standard deviation of the independent variable.
 - C) standard deviation of the residuals.
-

Question #115 of 119

Which of the following statements about the standard error of the estimate (SEE) is *least* accurate?

- A) The SEE will be high if the relationship between the independent and dependent variables is weak.
- B) The SEE may be calculated from the sum of the squared errors and the number of observations.
- C) The larger the SEE the larger the R^2 .

Question #116 of 119

A regression between the returns on a stock and its industry index returns gives the following results:

	Coefficient	Standard Error	t-value
Intercept	2.1	2.01	1.04
Industry Index	1.9	0.31	6.13

- The t -statistic critical value at the 0.01 level of significance is 2.58
- Standard error of estimate = 15.1
- Correlation coefficient = 0.849

The regression statistics presented indicate that the dispersion of stock returns about the regression line is:

- A) 72.10.
- B) 15.10.
- C) 63.20.

Question #117 of 119

A sample of 200 monthly observations is used to run a simple linear regression: $\text{Returns} = b_0 + b_1 \text{Leverage} + u$. The t-value for the regression coefficient of leverage is calculated as $t = -1.09$. A 5% level of significance is used to test whether leverage has a significant influence on returns. The correct decision is to:

- A) do not reject the null hypothesis and conclude that leverage does not significantly explain returns.
 - B) reject the null hypothesis and conclude that leverage does not significantly explain returns.
 - C) do not reject the null hypothesis and conclude that leverage significantly explains returns.
-

Question #118 of 119

Given: $Y = 2.83 + 1.5X$

What is the predicted value of the dependent variable when the value of an independent variable equals 2?

- A) -0.55
 - B) 2.83
 - C) 5.83
-

Question #119 of 119

The standard error of the estimate in a regression is the standard deviation of the:

- A) dependent variable.
- B) differences between the actual values of the dependent variable and the mean of the dependent variable.
- C) residuals of the regression.