Question #1 of 139

Question ID: 1479918

During the course of a multiple regression analysis, an analyst has observed several items that she believes may render incorrect conclusions. For example, the coefficient standard errors are too small, although the estimated coefficients are accurate. She believes that these small standard error terms will result in the computed *t*-statistics being too big, resulting in too many Type I errors. The analyst has *most likely* observed which of the following assumption violations in her regression analysis?

A) Positive serial correlation.

B) Homoskedasticity.

C) Multicollinearity.

Explanation

Positive serial correlation is the condition where a positive regression error in one time period increases the likelihood of having a positive regression error in the next time period. The residual terms are correlated with one another, leading to coefficient error terms that are too small.

(Module 1.3, LOS 1.i)

Question #2 of 139

Question ID: 1472066

When two or more of the independent variables in a multiple regression are correlated with each other, the condition is called:

A) multicollinearity.

B) conditional heteroskedasticity.

C) serial correlation.

Explanation

Multicollinearity refers to the condition when two or more of the independent variables, or linear combinations of the independent variables, in a multiple regression are highly correlated with each other. This condition distorts the standard error of estimate and the coefficient standard errors, leading to problems when conducting *t*-tests for statistical significance of parameters.

(Module 1.3, LOS 1.j)

Question #3 of 139

Question ID: 1479908

An analyst is trying to determine whether fund return performance is persistent. The analyst divides funds into three groups based on whether their return performance was in the top third (group 1), middle third (group 2), or bottom third (group 3) during the previous year. The manager then creates the following equation: $R = a + b_1D_1 + b_2D_2 + b_3D_3 + \varepsilon$, where R is return premium on the fund (the return minus the return on the S&P 500 benchmark) and D_i is equal to 1 if the fund is in group i. Assuming no other information, this equation will suffer from:

A) multicollinearity. ✓

B) serial correlation. ✗

C) heteroskedasticity. ✗

Explanation

When we use dummy variables, we have to use one less than the states of the world. In this case, there are three states (groups) possible. We should have used only two dummy variables. Multicollinearity is a problem in this case. Specifically, a linear combination of independent variables is perfectly correlated. X1 + X2 + X3 = 1.

There are too many dummy variables specified, so the equation will suffer from multicollinearity.

(Module 1.3, LOS 1.h)

Question #4 of 139

Question ID: 1479904

Which of the following is *least likely* to result in misspecification of a regression model?

A) Inappropriate variable form. ✗

B) Transforming a variable. ✓

C) Omission of an important independent variable. ✗

Explanation

The four types of model specification errors are: omission of an important independent variable, inappropriate variable form, inappropriate variable scaling and data improperly pooled. Transforming an independent variable is usually done to rectify inappropriate variable scaling.

(Module 1.3, LOS 1.g)

Question #5 of 139

Question ID: 1471891

Which of the following statements regarding the results of a regression analysis is *least* accurate? The:

A) slope coefficient in a multiple regression is the change in the dependent variable for a one-unit change in the independent variable, holding all other variables constant. ✗

B) slope coefficients in the multiple regression are referred to as partial betas. ✗

C) slope coefficient in a multiple regression is the value of the dependent variable for a given value of the independent variable. ✓

Explanation

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

(Module 1.1, LOS 1.b)

Question #6 of 139

Question ID: 1471927

One of the underlying assumptions of a multiple regression is that the variance of the residuals is constant for various levels of the independent variables. This quality is referred to as:

A) homoskedasticity. ✓

B) a normal distribution. ✗

C) a linear relationship. ✗

Explanation

Homoskedasticity refers to the basic assumption of a multiple regression model that the variance of the error terms is constant.

(Module 1.1, LOS 1.c)

Question #7 of 139

Question ID: 1471867

Consider the following estimated regression equation, with calculated *t*-statistics of the estimates as indicated:

$$AUTO_t = 10.0 + 1.25\ PI_t + 1.0\ TEEN_t - 2.0\ INS_t$$

with a PI calculated *t*-statistic of 0.45, a TEEN calculated *t*-statistic of 2.2, and an INS calculated *t*-statistic of 0.63.

The equation was estimated over 40 companies. Using a 5% level of significance, which of the independent variables significantly different from zero?

A) TEEN only. ✓

B) PI and INS only. ✗

C) PI only. ✗

Explanation

The critical t-values for 40-3-1 = 36 degrees of freedom and a 5% level of significance are ± 2.028. Therefore, only TEEN is statistically significant.

(Module 1.1, LOS 1.b)

Question #8 of 139

Question ID: 1471873

Henry Hilton, CFA, is undertaking an analysis of the bicycle industry. He hypothesizes that bicycle sales (SALES) are a function of three factors: the population under 20 (POP), the level of disposable income (INCOME), and the number of dollars spent on advertising (ADV). All data are measured in millions of units. Hilton gathers data for the last 20 years and estimates the following equation (standard errors in parentheses):

SALES	= 0.000	+ 0.004 POP	+ 1.031 INCOME	+ 2.002 ADV
	(0.113)	(0.005)	(0.337)	(2.312)

For next year, Hilton estimates the following parameters: (1) the population under 20 will be 120 million, (2) disposable income will be $300,000,000, and (3) advertising expenditures will be $100,000,000. Based on these estimates and the regression equation, what are predicted sales for the industry for next year?

A) $656,991,000. ✗

B) $509,980,000. ✓

C) $557,143,000. ✗

Explanation

Predicted sales for next year are:

$$\text{SALES} = \alpha + 0.004\,(120) + 1.031\,(300) + 2.002\,(100) = 509{,}980{,}000.$$

(Module 1.1, LOS 1.b)

Question #9 of 139

Question ID: 1479883

One possible problem that could jeopardize the validity of the employment growth rate model is multicollinearity. Which of the following would *most likely* suggest the existence of multicollinearity?

A) The variance of the observations has increased over time. ✗

B) The Durbin–Watson statistic is significant. ✗

C) The F-statistic suggests that the overall regression is significant, however the regression coefficients are not individually significant. ✓

Explanation

One symptom of multicollinearity is that the regression coefficients may not be individually statistically significant even when according to the F-statistic the overall regression is significant. The problem of multicollinearity involves the existence of high correlation between two or more independent variables. Clearly, as service employment rises, construction employment must rise to facilitate the growth in these sectors. Alternatively, as manufacturing employment rises, the service sector must grow to serve the broader manufacturing sector.

- The variance of observations suggests the possible existence of heteroskedasticity.
- If the Durbin–Watson statistic may be used to test for serial correlation at a single lag.

(Module 1.2, LOS 1.f)

Question #10 of 139

Question ID: 1586006

Consider the following estimated regression equation:

$$\text{AUTO}_t = 10.0 + 1.25\,\text{PI}_t + 1.0\,\text{TEEN}_t - 2.0\,\text{INS}_t$$

The equation was estimated over 40 companies. The predicted value of AUTO if PI is 4, TEEN is 0.30, and INS = 0.6 is *closest* to:

A) 14.90.

B) 17.50.

C) 14.10.

Explanation

Predicted AUTO

= 10 + 1.25 (4) + 1.0 (0.30) – 2.0 (0.6)

= 10 + 5 + 0.3 – 1.2

= 14.10

(Module 1.2, LOS 1.f)

Ben Sasse is a quantitative analyst at Gurnop Asset Managers. Sasse is interviewing Victor Sophie for a junior analyst position. Sasse mentions that the firm currently uses several proprietary multiple regression models and wants Sophie's opinion about regression models.

Sophie makes the following statements:

Statement 1:	Multiple regression models can be used to forecast independent variables.
Statement 2:	Multiple regression models can be used to test existing theories of relationships among variables.

Sasse then discusses a model that the firm uses to forecast credit spread on investment-grade corporate bonds. Sasse states that while the current model parameters are a secret, the following is an older version of the model:

CSP = 0.22 + 1.04 × DSC – 0.32 × index + 1.33 × D/E

where:

CSP = credit spread (%)

DSC = EBITDA / unsecured debt

index = 1 if the issuer is part of CDX index; 0 otherwise

D/E = long-term debt / equity

Question #11 - 14 of 139

Question ID: 1501587

Regarding Sophie's statement on multiple regression:

A) only Statement 1 is correct. ✗

B) only Statement 2 is correct. ✓

C) both statements are correct. ✗

Explanation

Multiple regression models can be used to *identify* relations between variables, forecast the *dependent* variable, and *test* existing theories. Statement 1 is inaccurate in because it mentions forecast independent (and not dependent) variables.

(Module 1.1, LOS 1.a)

Question #12 - 14 of 139

Question ID: 1501588

Based on the credit spread model, if an issuer gets included in the CDX index and assuming everything else the same, which of the following statements *most accurately* describes the model's forecast?

A) The credit spread on the firm's issue would decrease by 10 bps. ✗

B) The credit spread on the firm's issue will increase by 32 bps. ✗

C) The credit spread on the firm's issue will decrease by 32 bps. ✓

Explanation

The coefficient on the index dummy variable is –0.32, and if the variable takes a value of 1 (inclusion in the index), the credit spread would decrease by 0.32%, or 32 bps.

(Module 1.1, LOS 1.b)

Question #13 - 14 of 139

Question ID: 1501589

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

A) The dependent variable is not serially correlated. ✓

B) There is no linear relationship between the independent variables. ✗

C) The error term is normally distributed. ✗

Explanation

The assumption calls for the residual (or errors) to be not serially correlated. The dependent variable can have serial correlation. Other assumptions are accurate.

(Module 1.1, LOS 1.c)

Question #14 - 14 of 139

Question ID: 1501590

Which assumption of multiple regression is *most likely* evaluated using a QQ plot?

A) Serial correlation of residuals.

B) Conditional heteroskedasticity.

C) Error term is normally distributed.

Explanation

A normal QQ plot of the residuals can visually indicate violation of the assumption that the residuals are normally distributed.

(Module 1.1, LOS 1.c)

Question #15 of 139

Question ID: 1471872

Henry Hilton, CFA, is undertaking an analysis of the bicycle industry. He hypothesizes that bicycle sales (SALES) are a function of three factors: the population under 20 (POP), the level of disposable income (INCOME), and the number of dollars spent on advertising (ADV). All data are measured in millions of units. Hilton gathers data for the last 20 years and estimates the following equation (standard errors in parentheses):

SALES	= α	+ 0.004 POP	+ 1.031 INCOME	+ 2.002 ADV
		(0.005)	(0.337)	(2.312)

The critical t-statistic for a 95% confidence level is 2.120. Which of the independent variables is statistically different from zero at the 95% confidence level?

A) INCOME and ADV.

B) ADV only.

C) INCOME only.

Explanation

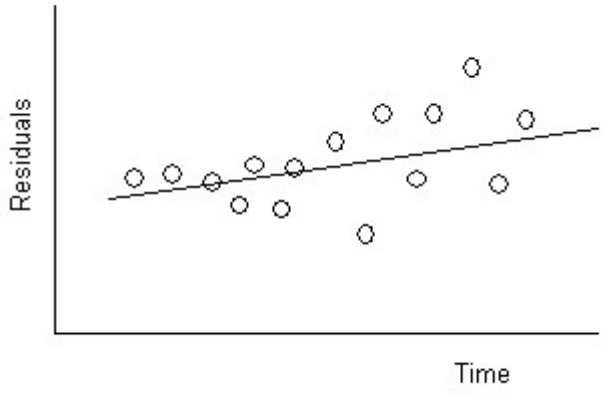
The calculated test statistic is coefficient/standard error. Hence, the t-stats are 0.8 for POP, 3.059 for INCOME, and 0.866 for ADV. Since the t-stat for INCOME is the only one greater than the critical t-value of 2.120, only INCOME is significantly different from zero.

(Module 1.1, LOS 1.b)

Question #16 of 139

Question ID: 1472011

Consider the following graph of residuals and the regression line from a time-series regression:



These residuals exhibit the regression problem of:

A) autocorrelation.

B) homoskedasticity.

C) heteroskedasticity.

Explanation

The residuals appear to be from two different distributions over time. In the earlier periods, the model fits rather well compared to the later periods.

(Module 1.3, LOS 1.h)

Using a recent analysis of salaries (in $1,000) of financial analysts, Timbadia runs a regression of salaries on education, experience, and gender. (Gender equals one for men and zero for women.) The regression results from a sample of 230 financial analysts are presented below, with *t*-statistics in parenthesis.

Salary	= 34.98	+ 1.2 Education	+ 0.5 Experience	+ 6.3 Gender
	(29.11)	(8.93)	(2.98)	(1.58)

Timbadia also runs a multiple regression to gain a better understanding of the relationship between lumber sales, housing starts, and commercial construction. The regression uses a large data set of lumber sales as the dependent variable with housing starts and commercial construction as the independent variables. The results of the regression are:

	Coefficient	Standard Error	t-statistics
Intercept	5.337	1.71	3.14
Housing starts	0.76	0.09	8.44
Commercial construction	1.25	0.33	3.78

Finally, Timbadia runs a regression between the returns on a stock and its industry index with the following results:

	Coefficient	Standard Error
Intercept	2.1	2.01
Industry index	1.9	0.31

- Standard error of estimate = 15.1
- Correlation coefficient = 0.849

Question #17 - 19 of 139

Question ID: 1479889

What is the expected salary (in $1,000) of a woman with 16 years of education and 10 years of experience?

A) 65.48.

B) 59.18.

C) 54.98.

Explanation

34.98 + 1.2(16) + 0.5(10) = 59.18

(Module 1.2, LOS 1.f)

Question #18 - 19 of 139

Question ID: 1479890

If the return on the industry index is 4%, the stock's expected return would be:

A) 9.7%.

B) 7.6%.

C) 11.2%.

Explanation

$$Y = b_0 + bX_1$$

$$Y = 2.1 + 1.9(4) = 9.7\%$$

(Module 1.2, LOS 1.f)

Question #19 - 19 of 139

Question ID: 1479891

The percentage of the variation in the stock return explained by the variation in the industry index return is *closest* to:

A) 72.1%.

B) 63.2%.

C) 84.9%.

Explanation

The coefficient of determination, R^2, is the square the correlation coefficient. 0.849^2, = 0.721.

(Module 1.2, LOS 1.d)

Question #20 of 139

Question ID: 1471980

An analyst runs a regression of monthly value-stock returns on five independent variables over 48 months. The total sum of squares is 430, and the sum of squared errors is 170. Test the null hypothesis at the 2.5% and 5% significance level that all five of the independent variables are equal to zero.

A) Rejected at 5% significance only.

B) Not rejected at 2.5% or 5.0% significance.

C) Rejected at 2.5% significance and 5% significance.

Explanation

The F-statistic is equal to the ratio of the mean squared regression (MSR) to the mean squared error (MSE).

RSS = SST – SSE = 430 – 170 = 260

MSR = 260 / 5 = 52

MSE = 170 / (48 – 5 – 1) = 4.05

F = 52 / 4.05 = 12.84

The critical F-value for 5 and 42 degrees of freedom at a 5% significance level is approximately 2.44. The critical F-value for 5 and 42 degrees of freedom at a 2.5% significance level is approximately 2.89. Therefore, we can reject the null hypothesis at either level of significance and conclude that at least one of the five independent variables explains a significant portion of the variation of the dependent variable.

(Module 1.2, LOS 1.e)

Question #21 of 139

Question ID: 1472075

Jill Wentraub is an analyst with the retail industry. She is modeling a company's sales over time and has noticed a quarterly seasonal pattern. If she includes dummy variables to represent the seasonality component of the sales she must use:

A) three dummy variables.

B) four dummy variables.

C) one dummy variables.

Explanation

Three. Always use one less dummy variable than the number of possibilities. For a seasonality that varies by quarters in the year, three dummy variables are needed.

(Module 1.4, LOS 1.l)

Question #22 of 139

Question ID: 1471947

An analyst regresses the return of a S&P 500 index fund against the S&P 500, and also regresses the return of an active manager against the S&P 500. The analyst uses the last five years of data in both regressions. Without making any other assumptions, which of the following is *most* accurate? The index fund:

A) regression should have higher sum of squares regression as a ratio to the total sum of squares. ✓

B) should have a higher coefficient on the independent variable. ✗

C) should have a lower coefficient of determination. ✗

Explanation

The index fund regression should provide a higher R^2 than the active manager regression. R^2 is the sum of squares regression divided by the total sum of squares.

(Module 1.2, LOS 1.d)

Question #23 of 139

Question ID: 1479923

Assume that in a particular multiple regression model, it is determined that the error terms are uncorrelated with each other. Which of the following statements is *most* accurate?

A) This model is in accordance with the basic assumptions of multiple regression analysis because the errors are not serially correlated. ✓

B) Unconditional heteroskedasticity present in this model should not pose a problem, but can be corrected by using robust standard errors. ✗

C) Serial correlation may be present in this multiple regression model, and can be confirmed only through a Durbin-Watson test. ✗

Explanation

One of the basic assumptions of multiple regression analysis is that the error terms are not correlated with each other. In other words, the error terms are not serially correlated. Multicollinearity and heteroskedasticity are problems in multiple regression that are not related to the correlation of the error terms.

(Module 1.3, LOS 1.i)

Vijay Shapule, CFA, is investigating the application of the Fama-French three-factor model (Model 1) for the Indian stock market for the period 2001–2011 (120 months). Using the dependent variable as annualized return (%), the results of the analysis are shown in **Indian Equities—Fama-French Model**.

Indian Equities—Fama-French Model

Factor	Coefficient	P-Value
Intercept	1.22	<0.001
SMB	0.23	<0.001
HML	0.34	0.003
Rm-Rf	0.88	<0.001
R-squared		0.36
SSE		38.00
AIC		–129.99
BIC		–118.84

Shapule then modifies the model to include a liquidity factor. Results for this four-factor model (Model 2) are shown in **Revised Fama-French Model With Liquidity Factor**

Revised Fama-French Model With Liquidity Factor

Factor	Coefficient	P-Value
Intercept	1.56	<0.001
SMB	0.22	<0.001
HML	0.35	0.012
Rm-Rf	0.87	<0.001
LIQ	–0.12	0.02
R-squared		0.39
SSE		34.00
AIC		–141.34
BIC		–127.40

Question #24 - 27 of 139

Question ID: 1501592

The adjusted R^2 of Model 2 is *closest* to:

A) 0.39.

B) 0.37.

C) 0.36.

Explanation

Given n = 120 months, k = 4 (for Model 2), and R^2 = 0.39:

$$\mathrm{R}_{\mathrm{a}}^2 = 1 - \left[\left(\frac{120-1}{120-4-1}\right) \times (1 - 0.39)\right] = 0.37$$

(Module 1.2, LOS 1.d)

Question #25 - 27 of 139

Question ID: 1501593

The model better suited for prediction is:

A) Model 1 because it has a lower Bayesian information criterion.

B) Model 2 because it has a higher Akaike information criterion.

C) Model 2 because it has a lower Akaike information criterion.

Explanation

The Akaike information criterion (AIC) is used if the goal is to have a better forecast, while the Bayesian information criterion (BIC) is used if the goal is a better goodness of fit. Lower values of both criteria indicate a better model. Both criteria are lower for Model 2.

(Module 1.2, LOS 1.d)

Question #26 - 27 of 139

Question ID: 1501594

The *F*-statistic for testing H_0: coefficient of LIQ = 0 versus H_a: coefficient of LIQ ≠ 0 is *closest* to:

A) 13.33.

B) 5.45.

C) 2.11.

Explanation