Multiple Regression and Issues in Regression Analysis 1

Test ID: 7440339

Questions #1-6 of 100

George Smith, an analyst with Great Lakes Investments, has created a comprehensive report on the pharmaceutical industry at the request of his boss. The Great Lakes portfolio currently has a significant exposure to the pharmaceuticals industry through its large equity position in the top two pharmaceutical manufacturers. His boss requested that Smith determine a way to accurately forecast pharmaceutical sales in order for Great Lakes to identify further investment opportunities in the industry as well as to minimize their exposure to downturns in the market. Smith realized that there are many factors that could possibly have an impact on sales, and he must identify a method that can quantify their effect. Smith used a multiple regression analysis with five independent variables to predict industry sales. His goal is to not only identify relationships that are statistically significant, but economically significant as well. The assumptions of his model are fairly standard: a linear relationship exists between the dependent and independent variables, the independent variables are not random, and the expected value of the error term is zero.

Smith is confident with the results presented in his report. He has already done some hypothesis testing for statistical significance, including calculating a t-statistic and conducting a two-tailed test where the null hypothesis is that the regression coefficient is equal to zero versus the alternative that it is not. He feels that he has done a thorough job on the report and is ready to answer any questions posed by his boss.

However, Smith's boss, John Sutter, is concerned that in his analysis, Smith has ignored several potential problems with the regression model that may affect his conclusions. He knows that when any of the basic assumptions of a regression model are violated, any results drawn for the model are questionable. He asks Smith to go back and carefully examine the effects of heteroskedasticity, multicollinearity, and serial correlation on his model. In specific, he wants Smith to make suggestions regarding how to detect these errors and to correct problems that he encounters.

Question #1 of 100

Question ID: 485683

Suppose that there is evidence that the residual terms in the regression are positively correlated. The *most likely* effect on the statistical inferences drawn from the regressions results is for Smith to commit a:

✓ **A) Type I error by incorrectly rejecting the null hypotheses that the regression parameters are equal to zero.**

✗ **B)** Type I error by incorrectly failing to reject the null hypothesis that the regression parameters are equal to zero.

✗ **C)** Type II error by incorrectly failing to reject the null hypothesis that the regression parameters are equal to zero.

Explanation

One problem with positive autocorrelation (also known as positive serial correlation) is that the standard errors of the parameter estimates will be too small and the t-statistics too large. This may lead Smith to incorrectly reject the null hypothesis that the parameters are equal to zero. In other words, Smith will incorrectly conclude that the parameters are statistically significant when in fact they are not. This is an example of a Type I error: incorrectly rejecting the null hypothesis when it should not be rejected. (Study Session 3, LOS 10.k)

Question #2 of 100

Question ID: 485684

Sutter has detected the presence of conditional heteroskedasticity in Smith's report. This is evidence that:

✗ **A) two or more of the independent variables are highly correlated with each other.**

✗ **B)** the error terms are correlated with each other.

✓ **C)** the variance of the error term is correlated with the values of the independent variables.

Explanation

Conditional heteroskedasticity exists when the variance of the error term is correlated with the values of the independent variables.

Multicollinearity, on the other hand, occurs when two or more of the independent variables are highly correlated with each other. Serial correlation exists when the error terms are correlated with each other. (Study Session 3, LOS 10.k)

Question #3 of 100

Question ID: 485685

Suppose there is evidence that the variance of the error term is correlated with the values of the independent variables. The *most likely* effect on the statistical inferences Smith can make from the regressions results is to commit a:

✓ **A) Type I error by incorrectly rejecting the null hypotheses that the regression parameters are equal to zero.**

✗ **B)** Type II error by incorrectly failing to reject the null hypothesis that the regression parameters are equal to zero.

✗ **C)** Type I error by incorrectly failing to reject the null hypothesis that the regression parameters are equal to zero.

Explanation

One problem with heteroskedasticity is that the standard errors of the parameter estimates will be too small and the t-statistics too large. This will lead Smith to incorrectly reject the null hypothesis that the parameters are equal to zero. In other words, Smith will incorrectly conclude that the parameters are statistically significant when in fact they are not. This is an example of a Type I error: incorrectly rejecting the null hypothesis when it should not be rejected. (Study Session 3, LOS 10.k)

Question #4 of 100

Question ID: 485686

Which of the following is *most likely* to indicate that two or more of the independent variables, or linear combinations of independent variables, may be highly correlated with each other? Unless otherwise noted, significant and insignificant mean significantly different from zero and not significantly different from zero, respectively.

✗ **A) The R^2 is low, the F-statistic is insignificant and the Durbin-Watson statistic is significant.**

✓ **B)** The R^2 is high, the F-statistic is significant and the t-statistics on the individual slope coefficients are insignificant.

✗ **C)** The R^2 is high, the F-statistic is significant and the t-statistics on the individual slope coefficients are significant.

Explanation

Multicollinearity occurs when two or more of the independent variables, or linear combinations of independent variables, may be highly

correlated with each other. In a classic effect of multicollinearity, the R^2 is high and the F-statistic is significant, but the t-statistics on the individual slope coefficients are insignificant. (Study Session 3, LOS 10.I)

Question #5 of 100

Question ID: 485687

Suppose there is evidence that two or more of the independent variables, or linear combinations of independent variables, may be highly correlated with each other. The *most likely* effect on the statistical inferences Smith can make from the regression results is to commit a:

✗ **A) Type I error by incorrectly rejecting the null hypothesis that the regression parameters are equal to zero.**

✓ **B)** Type II error by incorrectly failing to reject the null hypothesis that the regression parameters are equal to zero.

✗ **C)** Type I error by incorrectly failing to reject the null hypothesis that the regression parameters are equal to zero.

Explanation

One problem with multicollinearity is that the standard errors of the parameter estimates will be too large and the t-statistics too small. This will lead Smith to incorrectly fail to reject the null hypothesis that the parameters are statistically insignificant. In other words, Smith will incorrectly conclude that the parameters are not statistically significant when in fact they are. This is an example of a Type II error: incorrectly failing to reject the null hypothesis when it should be rejected. (Study Session 3, LOS 10.I)

Question #6 of 100

Question ID: 485688

Using the Durbin-Watson test statistic, Smith rejects the null hypothesis suggested by the test. This is evidence that:

✓ **A) the error terms are correlated with each other.**

✗ **B)** the error term is normally distributed.

✗ **C)** two or more of the independent variables are highly correlated with each other.

Explanation

Serial correlation (also called autocorrelation) exists when the error terms are correlated with each other.

Multicollinearity, on the other hand, occurs when two or more of the independent variables are highly correlated with each other. One assumption of multiple regression is that the error term is normally distributed. (Study Session 3, LOS 10.k)

Question #7 of 100

Question ID: 461672

An analyst wishes to test whether the stock returns of two portfolio managers provide different average returns. The analyst believes that the portfolio managers' returns are related to other factors as well. Which of the following can provide a suitable test?

✗ **A) Difference of means.**

✓ **B)** Dummy variable regression.

✗ **C)** Paired-comparisons.

Explanation

The difference of means and paired-comparisons tests will not account for the other factors.

Question #8 of 100

Question ID: 461529

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):

$$SALES = \alpha + 0.004\ POP + 1.031\ INCOME + 2.002\ ADV$$

$$(0.005) \qquad (0.337) \qquad (2.312)$$

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) ADV only.**
- ✗ **B)** INCOME and ADV.
- ✓ **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.

Question #9 of 100

Question ID: 461524

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*-statstic of 0.45, a TEEN calculated *t*-statstic of 2.2, and an INS calculated *t*-statstic 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) PI only.**
- ✓ **B)** TEEN only.
- ✗ **C)** PI and INS 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.

Question #10 of 100

Question ID: 461743

Which of the following statements regarding multicollinearity is *least* accurate?

✗ **A) If the *t*-statistics for the individual independent variables are insignificant, yet the F-statistic is significant, this indicates the presence of multicollinearity.**

✗ **B)** Multicollinearity may be a problem even if the multicollinearity is not perfect.

✓ **C)** Multicollinearity may be present in any regression model.

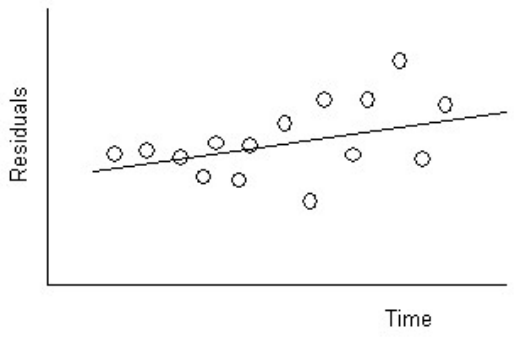
Explanation

Multicollinearity is not an issue in simple linear regression.

Question #11 of 100

Question ID: 461702

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)** heteroskedasticity.

✗ **C)** homoskedasticity.

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.

Question #12 of 100

Question ID: 461673

Consider the following model of earnings (EPS) regressed against dummy variables for the quarters:

$EPS_t = \alpha + \beta_1 Q_{1t} + \beta_2 Q_{2t} + \beta_3 Q_{3t}$

where:

EPS_t is a quarterly observation of earnings per share

Q_{1t} takes on a value of 1 if period t is the second quarter, 0 otherwise

> Q_{2t} takes on a value of 1 if period t is the third quarter, 0 otherwise
>
> Q_{3t} takes on a value of 1 if period t is the fourth quarter, 0 otherwise

Which of the following statements regarding this model is *most* accurate? The:

✗ **A) significance of the coefficients cannot be interpreted in the case of dummy variables.**

✓ **B)** coefficient on each dummy tells us about the difference in earnings per share between the respective quarter and the one left out (first quarter in this case).

✗ **C)** EPS for the first quarter is represented by the residual.

Explanation

The coefficients on the dummy variables indicate the difference in EPS for a given quarter, relative to the first quarter.

Questions #13-18 of 100

Using a recent analysis of salaries (in $1,000) of financial analysts, a regression of salaries on education, experience, and gender is run. (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 #13 of 100

Question ID: 485620

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

✓ **A) 59.18.**

✗ **B)** 65.48.

✗ **C)** 54.98.

Explanation

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

(LOS 10.e)

Question #14 of 100

Question ID: 485621

Holding everything else constant, do men get paid more than women? Use a 5% level of significance.

✗ **A) No, since the t-value does not exceed the critical value of 1.96.**

✗ **B)** Yes, since the t-value exceeds the critical value of 1.56.

✓ **C)** No, since the t-value does not exceed the critical value of 1.65.

Explanation

We cannot reject the null hypothesis.

H_0: $b_{gender} \leq 0$
H_a: $b_{gender} > 0$

For a one-tailed test with a 5% level of significance when degrees of freedom are high (>100), the critical t-value will be approximately 1.65. Because our *t*-value of 1.58 < 1.65 (critical value), we cannot conclude that there is a statistically significant salary benefit for men

(LOS 10.c)

Question #15 of 100

Question ID: 485622

Construct a 95% confidence interval for the slope coefficient for Housing Starts.

✗ **A) 0.76 ± 1.96(8.44).**

✓ **B)** 0.76 ± 1.96(0.09).

✗ **C)** 1.25 ± 1.96(0.33).

Explanation

The confidence interval for the slope coefficient is $b_1 \pm (t_c \times s_{b1})$. With large data set, t_c (α= 5%) = 1.96

(LOS 10.f)

Question #16 of 100

Question ID: 485623

Construct a 95% confidence interval for the slope coefficient for Commercial Construction.

✗ **A) 0.76 ± 1.96(0.09).**

✓ **B)** 1.25 ± 1.96(0.33).

✗ **C)** 1.25 ± 1.96(3.78).

Explanation

The confidence interval for the slope coefficient is $b_1 \pm (t_c \times s_{b1})$. With large data set, t_c (α = 5%) = 1.96

(LOS 10.f)

Question #17 of 100

Question ID: 485624

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

$Y = b_0 + bX_1$
$Y = 2.1 + 1.9(4) = 9.7\%$

(LOS 9.h)

Question #18 of 100

Question ID: 485625

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

- ✗ **A) 84.9%.**
- ✓ **B)** 72.1%.
- ✗ **C)** 63.2%.

Explanation

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

(LOS 9.j)

Question #19 of 100

Question ID: 461608

Wanda Brunner, CFA, is trying to calculate a 95% confidence interval (df = 40) for a regression equation based on the following information:

	Coefficient	*Standard Error*
Intercept	-10.60%	1.357
DR	0.52	0.023
CS	0.32	0.025

What are the lower and upper bounds for variable DR?

- ✗ **A) 0.488 to 0.552.**
- ✗ **B)** 0.481 to 0.559.
- ✓ **C)** 0.474 to 0.566.

Explanation

The critical t-value is 2.02 at the 95% confidence level (two tailed test). The estimated slope coefficient is 0.52 and the standard error is 0.023. The 95% confidence interval is 0.52 ± (2.02)(0.023) = 0.52 ± (0.046) = 0.474 to 0.566.

Question #20 of 100

Question ID: 461596

An analyst is investigating the hypothesis that the beta of a fund is equal to one. The analyst takes 60 monthly returns for the fund and regresses them against the Wilshire 5000. The test statistic is 1.97 and the *p*-value is 0.05. Which of the following is CORRECT?

✗ **A) The proportion of occurrences when the absolute value of the test statistic will be higher when beta is equal to 1 than when beta is not equal to 1 is less than or equal to 5%.**

✗ **B)** If beta is equal to 1, the likelihood that the absolute value of the test statistic is equal to 1.97 is less than or equal to 5%.

✓ **C)** If beta is equal to 1, the likelihood that the absolute value of the test statistic would be greater than or equal to 1.97 is 5%.

Explanation

P-value is the smallest significance level at which one can reject the null hypothesis. In other words, any significance level below the p-value would result in rejection of the null hypothesis. Recognize that we also can reject the null hypothesis when the absolute value of the computed test statistic (i.e., the t-value) is greater than the critical t value. Hence p-value is the likelihood of the test statistic being higher than the computed test statistic value assuming the null hypothesis is true.

Questions #21-26 of 100

Toni Williams, CFA, has determined that commercial electric generator sales in the Midwest U.S. for Self-Start Company is a function of several factors in each area: the cost of heating oil, the temperature, snowfall, and housing starts. Using data for the most currently available year, she runs a cross-sectional regression where she regresses the deviation of sales from the historical average in each area on the deviation of each explanatory variable from the historical average of that variable for that location. She feels this is the most appropriate method since each geographic area will have different average values for the inputs, and the model can explain how current conditions explain how generator sales are higher or lower from the historical average in each area. In summary, she regresses current sales for each area minus its respective historical average on the following variables for each area.

- The difference between the retail price of heating oil and its historical average.
- The mean number of degrees the temperature is below normal in Chicago.
- The amount of snowfall above the average.
- The percentage of housing starts above the average.

Williams used a sample of 26 observations obtained from 26 metropolitan areas in the Midwest U.S. The results are in the tables below. The dependent variable is in sales of generators in millions of dollars.

Coefficient Estimates Table		
Variable	*Estimated Coefficient*	*Standard Error of the*

		Coefficient
Intercept	5.00	1.850
$ Heating Oil	2.00	0.827
Low Temperature	3.00	1.200
Snowfall	10.00	4.833
Housing Starts	5.00	2.333

Analysis of Variance Table (ANOVA)			
Source	*Degrees of Freedom*	*Sum of Squares*	*Mean Square*
Regression	4	335.20	83.80
Error	21	606.40	28.88
Total	25	941.60	

One of her goals is to forecast the sales of the Chicago metropolitan area next year. For that area and for the upcoming year, Williams obtains the following projections: heating oil prices will be $0.10 above average, the temperature in Chicago will be 5 degrees below normal, snowfall will be 3 inches above average, and housing starts will be 3% below average.

In addition to making forecasts and testing the significance of the estimated coefficients, she plans to perform diagnostic tests to verify the validity of the model's results.

Question #21 of 100

Question ID: 485627

According to the model and the data for the Chicago metropolitan area, the forecast of generator sales is:

✗ **A) $55 million above average.**

✓ **B)** $35.2 million above the average.

✗ **C)** $65 million above the average.

Explanation

The model uses a multiple regression equation to predict sales by multiplying the estimated coefficient by the observed value to get:

> [5 + (2 × 0.10) + (3 × 5) + (10 × 3) + (5 × (−3))] × $1,000,000 = $35.2 million.

(Study Session 3, LOS 10.e)

Question #22 of 100

Question ID: 485628

Williams proceeds to test the hypothesis that none of the independent variables has significant explanatory power. He concludes that, at a 5% level of significance:

✓ **A) at least one of the independent variables has explanatory power, because the calculated F-statistic exceeds its critical value.**

✗ **B)** all of the independent variables have explanatory power, because the calculated F-statistic exceeds its critical value.

✗ **C)** none of the independent variables has explanatory power, because the calculated F-statistic does not exceed its critical value.

Explanation

From the ANOVA table, the calculated F-statistic is (mean square regression / mean square error) = (83.80 / 28.88) = 2.9017. From the F distribution table (4 df numerator, 21 df denominator) the critical F value is 2.84. Because 2.9017 is greater than 2.84, Williams rejects the null hypothesis and concludes that at least one of the independent variables has explanatory power. (Study Session 3, LOS 10.g)

Question #23 of 100

Question ID: 485629

With respect to testing the validity of the model's results, Williams may wish to perform:

✗ **A) a Durbin-Watson test, but not a Breusch-Pagan test.**

✗ **B)** a Breusch-Pagan test, but not a Durbin-Watson test.

✓ **C)** both a Durbin-Watson test and a Breusch-Pagan test.

Explanation

Since the model utilized is not an autoregressive time series, a test for serial correlation is appropriate so the Durbin-Watson test would be used. The Breusch-Pagan test for heteroskedasticity would also be a good idea. (Study Session 3, LOS 10.k)

Question #24 of 100

Question ID: 485630

Williams decides to use two-tailed tests on the individual variables, at a 5% level of significance, to determine whether electric generator sales are explained by each of them individually. Williams concludes that:

✓ **A) all of the variables except snowfall are statistically significant in explaining sales.**

✗ **B)** all of the variables are statistically significant in explaining sales.

✗ **C)** all of the variables except snowfall and housing starts are statistically significant in explaining sales.

Explanation

The calculated *t*-statistics are:

- Heating Oil: (2.00 / 0.827) = 2.4184
- Low Temperature: (3.00 / 1.200) = 2.5000
- Snowfall: (10.00 / 4.833) = 2.0691
- Housing Starts: (5.00 / 2.333) = 2.1432

All of these values are outside the *t*-critical value (at (26 − 4 − 1) = 21 degrees of freedom) of 2.080, except the change in snowfall. So Williams should reject the null hypothesis for the other variables and conclude that they explain sales, but fail to reject the null hypothesis with respect to snowfall and conclude that increases or decreases in snowfall do not explain sales. (Study Session 3, LOS 10.c)

Question #25 of 100

Question ID: 485631

When Williams ran the model, the computer said the R^2 is 0.233. She examines the other output and concludes that this is the:

✓ **A) adjusted R^2 value.**

✗ **B)** neither the unadjusted nor adjusted R^2 value, nor the coefficient of correlation.

✗ **C)** unadjusted R^2 value.

Explanation

This can be answered by recognizing that the unadjusted R-square is (335.2 / 941.6) = 0.356. Thus, the reported value must be the adjusted R^2. To verify this we see that the adjusted R-squared is: $1- ((26 - 1) / (26 - 4 - 1)) \times (1 - 0.356) = 0.233$. Note that whenever there is more than one independent variable, the adjusted R^2 will always be less than R^2. (Study Session 3, LOS 10.h)

Question #26 of 100

Question ID: 485632

In preparing and using this model, Williams has *least* likely relied on which of the following assumptions?

✓ **A) There is a linear relationship between the independent variables.**

✗ **B)** The residuals are homoscedastic.

✗ **C)** The disturbance or error term is normally distributed.

Explanation

Multiple regression models assume that there is no linear relationship between two or more of the independent variables. The other answer choices are both assumptions of multiple regression. (Study Session 3, LOS 10.f)

Question #27 of 100

Question ID: 461624

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) a normal distribution.**

✗ **B)** a linear relationship.

✓ **C)** homoskedasticity.

Explanation

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

Question #28 of 100

Question ID: 461605

Test the statistical significance of the independent variable change in oil prices (OIL) on quarterly EPS of SG Inc. (dependent variable). The results of the regression are shown below.

Coefficient	*Coefficient Value*	*Standard error*
Intercept	2.02	1.65

OIL −0.25 0.18
Number of observations = 45

✗ **A) The slope coefficient is statistically significant at 5% level of significance.**

✓ **B)** The slope coefficient is not statistically significant at 10% level of significance.

✗ **C)** The slope coefficient is statistically significant at 10% level of significance but not at 5% level of significance.

Explanation

t = −0.25/0.18 = 1.38
Critical values of t (2-tailed) at 5% level of significance = 1.96
Critical values of t (2-tailed) at 10% level of significance = 1.68
The absolute value of the computed t-statistic is lower than both. The slope coefficient is not statistically significant at 10% level of significance (and therefore cannot be significant at 5% level of significance).

Question #29 of 100

Question ID: 461669

A fund has changed managers twice during the past 10 years. An analyst wishes to measure whether either of the changes in managers has had an impact on performance. The analyst wishes to simultaneously measure the impact of risk on the fund's return. R is the return on the fund, and M is the return on a market index. Which of the following regression equations can appropriately measure the desired impacts?

✓ **A) $R = a + bM + c_1D_1 + c_2D_2 + \varepsilon$, where $D_1 = 1$ if the return is from the first manager, and $D_2 = 1$ if the return is from the third manager.**

✗ **B)** The desired impact cannot be measured.

✗ **C)** $R = a + bM + c_1D_1 + c_2D_2 + c_3D_3 + \varepsilon$, where $D_1 = 1$ if the return is from the first manager, and $D_2 = 1$ if the return is from the second manager, and $D_3 = 1$ is the return is from the third manager.

Explanation

The effect needs to be measured by two distinct dummy variables. The use of three variables will cause collinearity, and the use of one dummy variable will not appropriately specify the manager impact.

Question #30 of 100

Question ID: 461745

An analyst further studies the independent variables of a study she recently completed. The correlation matrix shown below is the result. Which statement *best* reflects possible problems with a multivariate regression?

	Age	*Education*	*Experience*	*Income*
Age	1.00			
Education	0.50	1.00		
Experience	0.95	0.55	1.00	

Income	0.60	0.65	0.89	1.00

✓ **A) Experience may be a redundant variable.**

✗ **B)** Education may be unnecessary.

✗ **C)** Age should be excluded from the regression.

Explanation

The correlation coefficient of experience with age and income, respectively, is close to +1.00. This indicates a problem of multicollinearity and should be addressed by excluding experience as an independent variable.

Question #31 of 100

Question ID: 461710

An analyst is estimating whether a fund's excess return for a month is dependent on interest rates and whether the S&P 500 has increased or decreased during the month. The analyst collects 90 monthly return premia (the return on the fund minus the return on the S&P 500 benchmark), 90 monthly interest rates, and 90 monthly S&P 500 index returns from July 1999 to December 2006. After estimating the regression equation, the analyst finds that the correlation between the regressions residuals from one period and the residuals from the previous period is 0.199. Which of the following is *most* accurate at a 0.05 level of significance, based solely on the information provided? The analyst:

✗ **A) cannot conclude that the regression exhibits either serial correlation or multicollinearity.**

✓ **B)** can conclude that the regression exhibits serial correlation, but cannot conclude that the regression exhibits multicollinearity.

✗ **C)** can conclude that the regression exhibits multicollinearity, but cannot conclude that the regression exhibits serial correlation.

Explanation

The Durbin-Watson statistic tests for serial correlation. For large samples, the Durbin-Watson statistic is approximately equal to two multiplied by the difference between one and the sample correlation between the regressions residuals from one period and the residuals from the previous period, which is 2 × (1 − 0.199) = 1.602, which is less than the lower Durbin-Watson value (with 2 variables and 90 observations) of 1.61. That means the hypothesis of no serial correlation is rejected. There is no information on whether the regression exhibits multicollinearity.

Question #32 of 100

Question ID: 461706

Which of the following is *least* accurate regarding the Durbin-Watson (DW) test statistic?

✓ **A) If the residuals have positive serial correlation, the DW statistic will be greater than 2.**

✗ **B)** If the residuals have positive serial correlation, the DW statistic will be less than 2.

✗ **C)** In tests of serial correlation using the DW statistic, there is a rejection region, a region over which the test can fail to reject the null, and an inconclusive region.

Explanation

A value of 2 indicates no correlation, a value greater than 2 indicates negative correlation, and a value less than 2 indicates a positive correlation. There is a range of values in which the DW test is inconclusive.

Question #33 of 100

Question ID: 461654

Which of the following statements regarding the R^2 is *least* accurate?

✓ **A) The R^2 is the ratio of the unexplained variation to the explained variation of the dependent variable.**

✗ **B)** The R^2 of a regression will be greater than or equal to the adjusted-R2 for the same regression.

✗ **C)** The F-statistic for the test of the fit of the model is the ratio of the mean squared regression to the mean squared error.

Explanation

The R^2 is the ratio of the explained variation to the total variation.

Question #34 of 100

Question ID: 479303

Consider the following regression equation:

> $Sales_i = 10.0 + 1.25\ R\&D_i + 1.0\ ADV_i - 2.0\ COMP_i + 8.0\ CAP_i$
> where Sales is dollar sales in millions, R&D is research and development expenditures in millions, ADV is dollar amount spent on advertising in millions, COMP is the number of competitors in the industry, and CAP is the capital expenditures for the period in millions of dollars.

Which of the following is NOT a correct interpretation of this regression information

✓ **A) If R&D and advertising expenditures are $1 million each, there are 5 competitors, and capital expenditures are $2 million, expected Sales are $8.25 million.**

✗ **B)** One more competitor will mean $2 million less in Sales (holding everything else constant).

✗ **C)** If a company spends $1 million more on capital expenditures (holding everything else constant), Sales are expected to increase by $8.0 million.

Explanation

Predicted sales = $10 + 1.25 + 1 - 10 + 16 = $18.25 million.

Question #35 of 100

Question ID: 461754

A high-yield bond analyst is trying to develop an equation using financial ratios to estimate the probability of a company defaulting on its

bonds. Since the analyst is using data over different economic time periods, there is concern about whether the variance is constant over time. A technique that can be used to develop this equation is:

✓ **A) logit modeling.**

✗ **B)** dummy variable regression.

✗ **C)** multiple linear regression adjusting for heteroskedasticity.

Explanation

The only one of the possible answers that estimates a probability of a discrete outcome is logit modeling.

Question #36 of 100

Question ID: 461719

Which of the following statements regarding serial correlation that might be encountered in regression analysis is *least* accurate?

✓ **A) Serial correlation occurs least often with time series data.**

✗ **B)** Negative serial correlation causes a failure to reject the null hypothesis when it is actually false.

✗ **C)** Positive serial correlation typically has the same effect as heteroskedasticity.

Explanation

Serial correlation, which is sometimes referred to as autocorrelation, occurs when the residual terms are correlated with one another, and is most frequently encountered with time series data.

Question #37 of 100

Question ID: 461704

Which of the following conditions will *least likely* affect the statistical inference about regression parameters by itself?

✓ **A) Unconditional heteroskedasticity.**

✗ **B)** Conditional heteroskedasticity.

✗ **C)** Multicollinearity.

Explanation

Unconditional heteroskedasticity does not impact the statistical inference concerning the parameters.

Question #38 of 100

Question ID: 461530

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:

SALES = α + 0.004 (120) + 1.031 (300) + 2.002 (100) = 509,980,000.

Question #39 of 100

Question ID: 461539

When interpreting the results of a multiple regression analysis, which of the following terms represents the value of the dependent variable when the independent variables are all equal to zero?

✗ **A) *p*-value.**

✗ **B)** Slope coefficient.

✓ **C)** Intercept term.

Explanation

The intercept term is the value of the dependent variable when the independent variables are set to zero.

Question #40 of 100

Question ID: 461643

Which of the following statements about the *F*-statistic is *least* accurate?

✗ **A) $df_{numerator} = k$ and $df_{denominator} = n - k - 1$.**

✗ **B)** F = MSR/MSE.

✓ **C)** Rejecting the null hypothesis means that only one of the independent variables is statistically significant.

Explanation

An *F*-test assesses how well the set of independent variables, as a group, explains the variation in the dependent variable. That is, the *F*-statistic is used to test whether *at least one* of the independent variables explains a significant portion of the variation of the dependent variable.

Questions #41-46 of 100

Miles Mason, CFA, works for ABC Capital, a large money management company based in New York. Mason has several years of experience as a financial analyst, but is currently working in the marketing department developing materials to be used by ABC's sales team for both existing and prospective clients. ABC Capital's client base consists primarily of large net worth individuals and Fortune 500 companies. ABC invests its clients' money in both publicly traded mutual funds as well as its own investment funds that are managed in-house. Five years ago, roughly half of its assets under management were invested in the publicly traded mutual funds, with the remaining half in the funds managed by ABC's investment team. Currently, approximately 75% of ABC's assets under management are invested in publicly traded funds, with the remaining 25% being distributed among ABC's private funds. The managing partners at ABC would like to shift more of its client's assets away from publicly-traded funds into ABC's proprietary funds, ultimately returning to a 50/50 split of assets between publicly traded funds and ABC funds. There are three key reasons for this shift in the firm's asset base. First, ABC's in-house funds have outperformed other funds consistently for the past five years. Second, ABC can offer its clients a reduced fee structure on funds managed in-house relative to other publicly traded funds. Lastly, ABC has recently hired a top fund manager away from a competing investment company and would like to increase his assets under management.

ABC Capital's upper management requested that current clients be surveyed in order to determine the cause of the shift of assets away from ABC funds. Results of the survey indicated that clients feel there is a lack of information regarding ABC's funds. Clients would like to see extensive information about ABC's past performance, as well as a sensitivity analysis showing how the funds will perform in varying market scenarios. Mason is part of a team that has been charged by upper management to create a marketing program to present to both current and potential clients of ABC. He needs to be able to demonstrate a history of strong performance for the ABC funds, and, while not promising any measure of future performance, project possible return scenarios. He decides to conduct a regression analysis on all of ABC's in-house funds. He is going to use 12 independent economic variables in order to predict each particular fund's return. Mason is very aware of the many factors that could minimize the effectiveness of his regression model, and if any are present, he knows he must determine if any corrective actions are necessary. Mason is using a sample size of 121 monthly returns.

Question #41 of 100

Question ID: 485662

In order to conduct an F-test, what would be the degrees of freedom used ($df_{numerator}$; $df_{denominator}$)?

✗ **A) 11; 120.**

✗ **B)** 108; 12.

✓ **C)** 12; 108.

Explanation

Degrees of freedom for the F-statistic is k for the numerator and n − k − 1 for the denominator.

> k = 12
>
> n − k − 1 = 121 − 12 − 1 = 108

(Study Session 3, LOS 10.g)

Question #42 of 100

Question ID: 485663

In regard to multiple regression analysis, which of the following statements is *most* accurate?

✓ **A) Adjusted R^2 is less than R^2.**

✗ **B)** Adjusted R^2 always decreases as independent variables increase.

✗ **C)** R^2 is less than adjusted R^2.

Explanation

Whenever there is more than one independent variable, adjusted R^2 is less than R^2. Adding a new independent variable will increase R^2, but may either increase or decrease adjusted R^2.

> R^2 adjusted = $1 - [((n - 1) / (n - k - 1)) \times (1 - R^2)]$
> Where:
> n = number of observations
> K = number of independent variables
> R^2 = unadjusted R^2

(Study Session 3, LOS 10.h)

Question #43 of 100

Question ID: 485664

Which of the following tests is *most likely* to be used to detect autocorrelation?

✓ **A) Durbin-Watson.**

✗ **B)** Dickey-Fuller.

✗ **C)** Breusch-Pagan.

Explanation

Durbin-Watson is used to detect autocorrelation. The Breusch-Pagan test is used to detect heteroskedasticity. The Dickey Fuller test is a test for unit root. (Study Session 3, LOS 10.k)

Question #44 of 100

Question ID: 485665

One of the most popular ways to correct heteroskedasticity is to:

✓ **A) use robust standard errors.**

✗ **B)** improve the specification of the model.

✗ **C)** adjust the standard errors.

Explanation

Using generalized least squares and calculating robust standard errors are possible remedies for heteroskedasticity. Improving specifications remedies serial correlation. The standard error cannot be adjusted, only the coefficient of the standard errors. (Study Session 3, LOS 10.k)

Question #45 of 100

Question ID: 485666

Which of the following statements regarding the Durbin-Watson statistic is *most* accurate? The Durbin-Watson statistic:

✗ **A) is approximately equal to 1 if the error terms are not serially correlated.**

✓ **B)** only uses error terms in its computations.

✗ **C)** can only be used to detect positive serial correlation.

Explanation

The formula for the Durbin-Watson statistic uses error terms in its calculation. The Durbin-Watson statistic is approximately

equal to 2 if there is no serial correlation. A Durbin-Watson statistic significantly less than 2 may indicate positive serial correlation, while a Durbin-Watson statistic significantly greater then 2 may indicate negative serial correlation. (Study Session 3, LOS 10.k)

Question #46 of 100

Question ID: 485667

If a regression equation shows that no individual t-tests are significant, but the F-statistic is significant, the regression probably exhibits:

✗ **A) heteroskedasticity.**

✓ **B)** multicollinearity.

✗ **C)** serial correlation.

Explanation

Common indicators of multicollinearity include: high correlation (>0.7) between independent variables, no individual t-tests are significant but the F-statistic is, and signs on the coefficients that are opposite of what is expected. (Study Session 3, LOS 10.l)

Question #47 of 100

Question ID: 472469

The F-statistic is the ratio of the mean square regression to the mean square error. The mean squares are provided directly in the analysis of variance (ANOVA) table. Which of the following statements regarding the ANOVA table for a regression is *most* accurate?

✗ **A)** $\mathbf{R^2 = SS_{Error} / SS_{Total}}$.

✗ **B)** $R^2 = SS_{Regression} - SS_{Error} / SS_{Total}$.

✓ **C)** $R^2 = SS_{Regression} / SS_{Total}$.

Explanation

The coefficient of determination is the proportion of the total variation of the dependent variable that is explained by the independent variables.

Questions #48-53 of 100

Manuel Mercado, CFA has performed the following two regressions on sales data for a given industry. He wants to forecast sales for each quarter of the upcoming year.

Model ONE	
Regression Statistics	
Multiple R	0.941828
R^2	0.887039
Adjusted R^2	0.863258
Standard Error	2.543272
Observations	24

Durbin-Watson test statistic = 0.7856

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	965.0619	241.2655	37.30006	9.49E−09
Residual	19	122.8964	6.4682		
Total	23	1087.9583			

	Coefficients	*Standard Error*	*t-Statistic*
Intercept	31.40833	1.4866	21.12763
Q1	−3.77798	1.485952	−2.54246
Q2	−2.46310	1.476204	−1.66853
Q3	−0.14821	1.470324	−0.10080
TREND	0.851786	0.075335	11.20848

Model TWO	
Regression Statistics	
Multiple R	0.941796
R^2	0.886979
Adjusted R^2	0.870026
Standard Error	2.479538
Observations	24

Durbin-Watson test statistic = 0.7860

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	964.9962	321.6654	52.3194	1.19E−09
Residual	20	122.9622	6.14811		
Total	23	1087.9584			

	Coefficients	*Standard Error*	*t-Statistic*
Intercept	31.32888	1.228865	25.49416
Q1	−3.70288	1.253493	−2.95405
Q2	−2.38839	1.244727	−1.91881
TREND	0.85218	0.073991	11.51732

The dependent variable is the level of sales for each quarter, in $ millions, which began with the first quarter of the first year. Q1, Q2, and Q3 are seasonal dummy variables representing each quarter of the year. For the first four observations the dummy variables are as follows: Q1:(1,0,0,0), Q2:(0,1,0,0), Q3:(0,0,1,0). The TREND is a series that begins with one and increases by one each period to end with 24. For all tests, Mercado will use a 5% level of significance. Tests of coefficients will be two-tailed, and all others are one-tailed.

Question #48 of 100

Question ID: 485634

Which model would be a better choice for making a forecast?

✓ **A) Model TWO because it has a higher adjusted R^2.**

✗ **B)** Model ONE because it has a higher R^2.

✗ **C)** Model TWO because serial correlation is not a problem.

Explanation

Model TWO has a higher adjusted R^2 and thus would produce the more reliable estimates. As is always the case when a variable is removed, R^2 for Model TWO is lower. The increase in adjusted R^2 indicates that the removed variable, Q3, has very little explanatory power, and removing it should improve the accuracy of the estimates. With respect to the references to autocorrelation, we can compare the Durbin-Watson statistics to the critical values on a Durbin-Watson table. Since the critical DW statistics for Model ONE and TWO respectively are 1.01 (>0.7856) and 1.10 (>0.7860), serial correlation is a problem for both equations. (Study Session 3, LOS 10.h)

Question #49 of 100

Question ID: 485635

Using Model ONE, what is the sales forecast for the second quarter of the next year?

✗ **A) $56.02 million.**

✓ **B)** $51.09 million.

✗ **C)** $46.31 million.

Explanation

The estimate for the second quarter of the following year would be (in millions):

> 31.4083 + (−2.4631) + (24 + 2) × 0.851786 = 51.091666. (Study Session 3, LOS 10.e)

Question #50 of 100

Question ID: 485636

Which of the coefficients that appear in both models are not significant at the 5% level in a two-tailed test?

✗ **A) The coefficients on Q1 and Q2 only.**

✓ **B)** The coefficient on Q2 only.

✗ **C)** The intercept only.

Explanation

The absolute value of the critical T-statistics for Model ONE and TWO are 2.093 and 2.086, respectively. Since the *t*-statistics for Q2 in Models ONE and TWO are −1.6685 and −1.9188, respectively, these fall below the critical values for both models. (Study Session 3, LOS 10.a)

Question #51 of 100

Question ID: 485637

If it is determined that conditional heteroskedasticity is present in model one, which of the following inferences are *most* accurate?

✗ **A) Both the regression coefficients and the standard errors will be biased.**

✗ **B)** Regression coefficients will be biased but standard errors will be unbiased.

✓ **C)** Regression coefficients will be unbiased but standard errors will be biased.

Explanation

Presence of conditional heteroskedasticity will not affect the consistency of regression coefficients but will bias the standard errors leading to incorrect application of t-tests for statistical significance of regression parameters. (Study Session 3, LOS 10.k)

Question #52 of 100

Question ID: 485638

Mercado probably did not include a fourth dummy variable Q4, which would have had 0, 0, 0, 1 as its first four observations because:

✗ **A) it would have lowered the explanatory power of the equation.**

✓ **B)** the intercept is essentially the dummy for the fourth quarter.

✗ **C)** it would not have been significant.

Explanation

The fourth quarter serves as the base quarter, and for the fourth quarter, Q1 = Q2 = Q3 = 0. Had the model included a Q4 as specified, we could not have had an intercept. In that case, for Model ONE for example, the estimate of Q4 would have been 31.40833. The dummies for the other quarters would be the 31.40833 plus the estimated dummies from the Model ONE. In a model that included Q1, Q2, Q3, and Q4 but no intercept, for example:

> Q1 = 31.40833 + (−3.77798) = 27.63035

Such a model would produce the same estimated values for the dependent variable. (Study Session 3, LOS 10.j)

Question #53 of 100

Question ID: 485639

If Mercado determines that Model TWO is the appropriate specification, then he is essentially saying that for each year, value of sales from quarter three to four is expected to:

✗ **A) remain approximately the same.**

✓ **B)** grow, but by less than $1,000,000.

✗ **C)** grow by more than $1,000,000.

Explanation

The specification of Model TWO essentially assumes there is no difference attributed to the change of the season from the third to fourth quarter. However, the time trend is significant. The trend effect for moving from one season to the next is the coefficient on TREND times $1,000,000 which is $852,182 for Equation TWO. (Study Session 3, LOS 11.a)

Questions #54-59 of 100

In preparing an analysis of HB Inc., Jack Stumper is asked to look at the company's sales in relation to broad based economic indicators. Stumper's analysis indicates that HB's monthly sales are related to changes in housing starts (H) and changes in the mortgage interest rate (M). The analysis covers the past ten years for these variables. The regression equation is:

> S = 1.76 + 0.23H - 0.08M

> Number of

observations:	123
Unadjusted R^2:	0.77
F statistic:	9.80
Durbin Watson statistic	0.50
p-value of Housing Starts	0.017
t-stat of Mortgage Rates	−2.6

Variable Descriptions

S = HB Sales (in thousands)

H = housing starts (in thousands)

M = mortgage interest rate (in percent)

November 20x6 Actual Data

HB's monthly sales: $55,000

Housing starts: 150,000

Mortgage interest rate (%): 7.5

Critical Values for Student's t-Distribution

Degrees of Freedom	Level of significance for one-tailed test					
	10%	5%	2.5%	1%	0.5%	0.05%
	Level of significance for two-tailed test					
	20%	10%	5%	2%	1%	0.1%
10	1.372	1.812	2.228	2.764	3.169	4.587
20	1.325	1.725	2.086	2.528	2.845	3.850
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
120	1.289	1.658	1.980	2.358	2.617	3.373

Question #54 of 100

Question ID: 485585

Using the regression model developed, the closest prediction of sales for December 20x6 is:

✗ **A) $44,000**

✗ **B)** $55,000

✓ **C)** $36,000

Explanation

1.76 + 0.23 * (150) − 0.08 * (7.5) = 35.66.

(Study Session 3, LOS 10.e)

Question #55 of 100

Question ID: 485586

Will Stumper conclude that the housing starts coefficient is statistically different from zero and how will he interpret it at the 5% significance level:

✗ **A) different from zero; sales will rise by $100 for every 23 house starts**

✓ **B)** different from zero; sales will rise by $23 for every 100 house starts

✗ **C)** not different from zero; sales will rise by $0 for every 100 house starts

Explanation

A p-value (0.017) below significance (0.05) indicates a variable which is statistically different from zero. The coefficient of 0.23 indicates that sales will rise by $23 for every 100 house starts.

(Study Session 3, LOS 10.a)

Question #56 of 100

Question ID: 485587

Is the regression coefficient of changes in mortgage interest rates different from zero at the 5 percent level of significance?

✓ **A) yes, because 2.6 > 1.98**

✗ **B)** yes, because 2.6 > 2.23

✗ **C)** no, because 2.6 < 2.62

Explanation

The correct degrees of freedom for critical t-statistic is n-k-1 = 123-2-1 = 120. From the t-table, 5% L.O.S, 2-tailed, critical t-value is 1.98. Note that the t-stat for the coefficient for mortgage rate is directly given in the question (-2.6).

(Study Session 3, LOS 10.c)

Question #57 of 100

Question ID: 485588

In this multiple regression, the F-statistic indicates the:

✓ **A) the joint significance of the independent variables**

✗ **B)** deviation of the estimated values from the actual values of the dependent variable

✗ **C)** degree of correlation between the independent variables

Explanation

The F-statistic indicates the joint significance of the independent variables. The deviation of the estimated values from the actual values of the dependent variable is the standard error of estimate. The degree of correlation between the independent variables is the coefficient of correlation.

(Study Session 3, LOS 10.g)

Question #58 of 100

Question ID: 485589

The regression statistics above indicate that for the period under study, the independent variables (housing starts, mortgage interest rate) together explained approximately what percentage of the variation in the dependent variable (sales)?

✓ **A) 77.00**

✗ **B)** 9.80

✗ **C)** 67.00

Explanation

The question is asking for the coefficient of determination.

(Study Session 3, LOS 10.h)

Question #59 of 100

Question ID: 485590

In this multiple regression, if Stumper discovers that the residuals exhibit positive serial correlation, the *most likely* effect is?

✗ **A) standard errors are too high but coefficient estimate is consistent.**

✗ **B)** standard errors are not affected but coefficient estimate is inconsistent

✓ **C)** standard errors are too low but coefficient estimate is consistent.

Explanation

Positive serial correlation does not affect the consistency of coefficients (i.e., the coefficients are still consistent) but the estimated standard errors are too low leading to artificially high t-statistics.

(Study Session 3, LOS 10.k)

Question #60 of 100

Question ID: 461626

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

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

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

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.

Question #61 of 100

Question ID: 461707

An analyst is estimating whether company sales is related to three economic variables. The regression exhibits conditional heteroskedasticity, serial correlation, and multicollinearity. The analyst uses Hansen's procedure to adjust for the standard errors. Which of the following is *most* accurate? The:

✓ **A) regression will still exhibit multicollinearity, but the heteroskedasticity and serial correlation problems will be solved.**

✗ **B)** regression will still exhibit heteroskedasticity and multicollinearity, but the serial correlation problem will be solved.

✗ **C)** regression will still exhibit serial correlation and multicollinearity, but the heteroskedasticity problem will be solved.

Explanation

The Hansen procedure simultaneously solves for heteroskedasticity and serial correlation.

Question #62 of 100

Question ID: 461753

Which of the following questions is *least likely* answered by using a qualitative dependent variable?

✓ **A) Based on the following executive-specific and company-specific variables, how many shares will be acquired through the exercise of executive stock options?**

✗ **B)** Based on the following company-specific financial ratios, will company ABC enter bankruptcy?

✗ **C)** Based on the following subsidiary and competition variables, will company XYZ divest itself of a subsidiary?

Explanation

The number of shares can be a broad range of values and is, therefore, not considered a qualitative dependent variable.

Question #63 of 100

Question ID: 461718

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) Multicollinearity.**

✗ **B)** Homoskedasticity.

✓ **C)** Positive serial correlation.

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.

Question #64 of 100

Question ID: 461705

Which of the following is *least likely* a method of detecting serial correlations?

✗ **A) The Durbin-Watson test.**

✓ **B)** The Breusch-Pagan test.

✗ **C)** A scatter plot of the residuals over time.

Explanation

The Breusch-Pagan test is a test of the heteroskedasticity and not of serial correlation.

Question #65 of 100

Question ID: 461676

The amount of the State of Florida's total revenue that is allocated to the education budget is believed to be dependent upon the total revenue for the year and the political party that controls the state legislature. Which of the following regression models is *most appropriate* for capturing the effect of the political party on the education budget? Assume Y_t is the amount of the education budget for Florida in year t, X is Florida's total revenue in year t, and D_t = {1 if the legislature has a Democratic majority in year t, 0 otherwise}.

✗ **A)** $\mathbf{Y_t = b_1D_t + b_2X_t + e_t.}$

✗ **B)** $Y_t = b_0 + b_1D_t + e_t.$

✓ **C)** $Y_t = b_0 + b_1D_t + b_2X_t + e_t.$

Explanation

In this application, b_0, b_1, and b_2 are estimated by regressing Y_t against a constant, D_t, and X_t.The estimated relationships for the two parties are:

> Non-Democrats: $\hat{Y} = b_0 + b_2X_t$
> Democrats: $\hat{Y} = (b_0 + b_1) + b_2X_t$

Questions #66-71 of 100

A real estate agent wants to develop a model to predict the selling price of a home. The agent believes that the most important variables in determining the price of a house are its size (in square feet) and the number of bedrooms. Accordingly, he takes a random sample of 32 homes that has recently been sold. The results of the regression are:

	Coefficient	*Standard Error*	*t-statistics*
Intercept	66,500	59,292	1.12
House Size	74.30	21.11	3.52
Number of Bedrooms	10306	3230	3.19

$R^2 = 0.56$; $F = 40.73$

Selected F- table values for significance level of 0.05:

	1	2
28	4.20	3.34
29	4.18	3.33
30	4.17	3.32
32	4.15	3.29

(Degrees of freedom for the numerator in columns; Degrees of freedom for the denominator in rows)

Additional information regarding this multiple regression:

1. Variance of error is not constant across the 32 observations.
2. The two variables (size of the house and the number of bedrooms) are highly correlated.
3. The error variance is not correlated with the size of the house nor with the number of bedrooms.

Question #66 of 100

Question ID: 485669

The predicted price of a house that has 2,000 square feet of space and 4 bedrooms is *closest* to:

✗ **A) $292,000.**

✓ **B)** $256,000.

✗ **C)** $114,000.

Explanation

66,500 + 74.30(2,000) + 10,306(4) = $256,324

(LOS 10.e)

Question #67 of 100

Question ID: 485670

The conclusion from the hypothesis test of H_0: $b_1 = b_2 = 0$, is that the null hypothesis should:

✓ **A) be rejected as the calculated F of 40.73 is greater than the critical value of 3.33.**

✗ **B)** be rejected as the calculated F of 40.73 is greater than the critical value of 3.29.

✗ **C)** not be rejected as the calculated F of 40.73 is greater than the critical value of 3.29.

Explanation

We can reject the null hypothesis that coefficients of both independent variables equal 0. The F value for comparison is $F_{2,29}$ = 3.33. The degrees of freedom in the numerator is 2; equal to the number of independent variables. Degrees of freedom for the denominator is 32 − (2+1) = 29. The critical value of the F-test needed to reject the null hypothesis is thus 3.33. The actual value of the F-test statistic is 40.73, so the null hypothesis should be rejected, as the calculated F of 40.73 is greater than the critical value of 3.33.

(LOS 10.g)

Question #68 of 100

Question ID: 485671

The regression results indicate that at a 5% level of significance:

✗ **A) the slopes and the intercept are both statistically significant.**

✓ **B)** the slopes are significant but the intercept is not.

✗ **C)** the slopes are not significant but the intercept is significant.

Explanation

df = n − k − 1 = 32 − 2 − 1 = 29. The *t*-critical value at 5% significance for a 2-tailed test with 29 df is 2.045. T-values for the slope coefficients are 3.52 and 3.19, which are both greater than the 2.045 critical value. For the intercept, the *t*-value of 1.12 is less than the critical *t*-value of 2.045.

(LOS 10.c)

Question #69 of 100

Question ID: 485672

Which of the following is *most likely* to present a problem in using this regression for forecasting?

✗ **A) autocorrelation.**

✗ **B)** heteroskedasticity.

✓ **C)** multicollinearity.

Explanation

Multicollinearity is present in a regression model when some linear combination of the independent variables are highly correlated. We are told that the two independent variables in this question are highly correlated. We also recognize that unconditional heteroskedasticity is present - but this would not pose any major problems in using this model for forecasting. No information is given about autocorrelation in residuals, but this is generally a concern with time series data (in this case, the model uses cross-sectional data).

(LOS 10.k,l)

Question #70 of 100

Question ID: 485673

Based on the information given in this question, heteroskedasticity is:

✓ **A) present but a statistical inference is still reliable.**

✗ **B)** present and a statistical inference is unreliable.

✗ **C)** not present and a statistical inference is reliable

Explanation

Variance of error is not constant across the 32 observations, however and the error variance is not correlated with the size of the house nor with the number of bedrooms. It appears that unconditional heteroskedasticity exists in the model. This form of heteroskedasticity is not as severe as conditional heteroskedasticity and statistical inference is still possible.

(LOS 10.k)

Question #71 of 100

Question ID: 485674

For this regression model, which condition is *most likely*?:

✗ **A) Coefficient estimates may be inconsistent but standard error will be unbiased.**

✗ **B)** Coefficient estimates will be consistent but standard error may be biased.

✓ **C)** Coefficient estimates may be unreliable and standard error may be biased.

Explanation

There are two issues with this regression: multicollinearity and unconditional heteroskedasticity. Unconditional heteroskedasticity does not pose any serious issues with statistical reliability. Multicollinearity causes coefficient estimates to be unreliable and standard errors to be biased.

(LOS 10.k,l)

Question #72 of 100

Question ID: 461548

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

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

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

✗ **C)** 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.

Explanation

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

Question #73 of 100

Question ID: 461703

Which of the following statements regarding heteroskedasticity is *least* accurate?

✗ **A) Heteroskedasticity results in an estimated variance that is too small and, therefore, affects statistical inference.**

✗ **B)** Heteroskedasticity may occur in cross-sectional or time-series analyses.

✓ **C)** The assumption of linear regression is that the residuals are heteroskedastic.

Explanation

The assumption of regression is that the residuals are homoskedastic (i.e., the residuals are drawn from the same distribution).

Questions #74-79 of 100

William Brent, CFA, is the chief financial officer for Mega Flowers, one of the largest producers of flowers and bedding plants in the Western United States. Mega Flowers grows its plants in three large nursery facilities located in California. Its products are sold in its company-owned retail nurseries as well as in large, home and garden "super centers". For its retail stores, Mega Flowers has designed and implemented marketing plans each season that are aimed at its consumers in order to generate additional sales for certain high-margin products. To fully implement the marketing plan, additional contract salespeople are seasonally employed.

For the past several years, these marketing plans seemed to be successful, providing a significant boost in sales to those specific products highlighted by the marketing efforts. However, for the past year, revenues have been flat, even though

marketing expenditures increased slightly. Brent is concerned that the expensive seasonal marketing campaigns are simply no longer generating the desired returns, and should either be significantly modified or eliminated altogether. He proposes that the company hire additional, permanent salespeople to focus on selling Mega Flowers' high-margin products all year long. The chief operating officer, David Johnson, disagrees with Brent. He believes that although last year's results were disappointing, the marketing campaign has demonstrated impressive results for the past five years, and should be continued. His belief is that the prior years' performance can be used as a gauge for future results, and that a simple increase in the sales force will not bring about the desired results.

Brent gathers information regarding quarterly sales revenue and marketing expenditures for the past five years. Based upon historical data, Brent derives the following regression equation for Mega Flowers (stated in millions of dollars):

> Expected Sales = 12.6 + 1.6 (Marketing Expenditures) + 1.2 (# of Salespeople)

Brent shows the equation to Johnson and tells him, "This equation shows that a $1 million increase in marketing expenditures will increase the independent variable by $1.6 million, all other factors being equal." Johnson replies, "It also appears that sales will equal $12.6 million if all independent variables are equal to zero."

Question #74 of 100

Question ID: 485578

In regard to their conversation about the regression equation:

✗ **A) Brent's statement is correct; Johnson's statement is incorrect.**

✓ **B)** Brent's statement is incorrect; Johnson's statement is correct.

✗ **C)** Brent's statement is correct; Johnson's statement is correct.

Explanation

Expected sales is the dependent variable in the equation, while expenditures for marketing and salespeople are the independent variables. Therefore, a $1 million increase in marketing expenditures will increase the *dependent* variable (expected sales) by $1.6 million. Brent's statement is incorrect.

Johnson's statement is correct. 12.6 is the intercept in the equation, which means that if all independent variables are equal to zero, expected sales will be $12.6 million. (Study Session 3, LOS 10.a)

Question #75 of 100

Question ID: 485579

Using data from the past 20 quarters, Brent calculates the *t*-statistic for marketing expenditures to be 3.68 and the *t*-statistic for salespeople at 2.19. At a 5% significance level, the two-tailed critical values are t_c = +/- 2.127. This *most likely* indicates that:

✗ **A) the null hypothesis should not be rejected.**

✓ **B)** both independent variables are statistically significant.

✗ **C)** the *t*-statistic has 18 degrees of freedom.

Explanation

Using a 5% significance level with degrees of freedom (df) of 17 (20 - 2 - 1), both independent variables are significant and contribute to the level of expected sales. (Study Session 3, LOS 10.a)

Question #76 of 100

Question ID: 485580

Brent calculated that the sum of squared errors (SSE) for the variables is 267. The mean squared error (MSE) would be:

✗ **A) 14.831.**

✓ **B)** 15.706.

✗ **C)** 14.055.

Explanation

The MSE is calculated as SSE / (n − k − 1). Recall that there are twenty observations and two independent variables. Therefore, the MSE in this instance [267 / (20 − 2 − 1)] = 15.706. (Study Session 3, LOS 9.j)

Question #77 of 100

Question ID: 485581

Brent is trying to explain the concept of the standard error of estimate (SEE) to Johnson. In his explanation, Brent makes three points about the SEE:

- Point 1: The SEE is the standard deviation of the differences between the estimated values for the independent variables and the actual observations for the independent variable.
- Point 2: Any violation of the basic assumptions of a multiple regression model is going to affect the SEE.
- Point 3: If there is a strong relationship between the variables and the SSE is small, the individual estimation errors will also be small.

How many of Brent's points are *most* accurate?

✗ **A) All 3 of Brent's points are correct.**

✗ **B)** 1 of Brent's points are correct.

✓ **C)** 2 of Brent's points are correct.

Explanation

The statements that if there is a strong relationship between the variables and the SSE is small, the individual estimation errors will also be small, and also that any violation of the basic assumptions of a multiple regression model is going to affect the SEE are both correct.

The SEE is the standard deviation of the differences between the estimated values for the *dependent* variables (not independent) and the actual observations for the dependent variable. Brent's Point 1 is incorrect.

Therefore, 2 of Brent's points are correct. (Study Session 3, LOS 9.f)

Question #78 of 100

Question ID: 485582

Assuming that next year's marketing expenditures are $3,500,000 and there are five salespeople, predicted sales for Mega Flowers will be:

✓ **A) $24,200,000.**

✗ **B)** $11,600,000.

✗ **C)** $2,400,000.

Explanation

Using the regression equation from above, expected sales equals 12.6 + (1.6 x 3.5) + (1.2 x 5) = $24.2 million. Remember to check the details - i.e. this equation is denominated in millions of dollars. (Study Session 3, LOS 10.e)

Question #79 of 100

Question ID: 485583

Brent would like to further investigate whether at least one of the independent variables can explain a significant portion of the variation of the dependent variable. Which of the following methods would be *best* for Brent to use?

✗ **A) The multiple coefficient of determination.**

✓ **B)** The *F*-statistic.

✗ **C)** An ANOVA table.

Explanation

To determine whether at least one of the coefficients is statistically significant, the calculated *F*-statistic is compared with the critical *F*-value at the appropriate level of significance. (Study Session 3, LOS 10.g)

Question #80 of 100

Question ID: 461709

An analyst is estimating whether a fund's excess return for a month is dependent on interest rates and whether the S&P 500 has increased or decreased during the month. The analyst collects 90 monthly return premia (the return on the fund minus the return on the S&P 500 benchmark), 90 monthly interest rates, and 90 monthly S&P 500 index returns from July 1999 to December 2006. After estimating the regression equation, the analyst finds that the correlation between the regressions residuals from one period and the residuals from the previous period is 0.145. Which of the following is *most* accurate at a 0.05 level of significance, based solely on the information provided? The analyst:

✗ **A) can conclude that the regression exhibits serial correlation, but cannot conclude that the regression exhibits heteroskedasticity.**

✓ **B)** cannot conclude that the regression exhibits either serial correlation or heteroskedasticity.

✗ **C)** can conclude that the regression exhibits heteroskedasticity, but cannot conclude that the regression exhibits serial correlation.

Explanation

The Durbin-Watson statistic tests for serial correlation. For large samples, the Durbin-Watson statistic is equal to two multiplied by the difference between one and the sample correlation between the regressions residuals from one period and the residuals from the previous period, which is 2 × (1 − 0.145) = 1.71, which is higher than the upper Durbin-Watson value (with 2 variables and 90 observations) of 1.70. That means the hypothesis of no serial correlation cannot be rejected. There is no information on whether the regression exhibits heteroskedasticity.

Question #81 of 100

Question ID: 461748

When utilizing a proxy for one or more independent variables in a multiple regression model, which of the following errors is *most likely* to occur?

✓ **A) Model misspecification.**

✗ **B)** Multicollinearity.

✗ **C)** Heteroskedasticity.

Explanation

By using a proxy for an independent variable in a multiple regression analysis, there is some degree of error in the measurement of the variable.

Questions #82-87 of 100

Werner Baltz, CFA, has regressed 30 years of data to forecast future sales for National Motor Company based on the percent change in gross domestic product (GDP) and the change in retail price of a U.S. gallon of fuel. The results are presented below.

Predictor	*Coefficient*	*Standard Error of the Coefficient*
Intercept	78	13.710
Δ GDP	30.22	12.120
Δ $ Fuel	−412.39	183.981

Analysis of Variance Table (ANOVA)

Source	*Degrees of Freedom*	*Sum of Squares*
Regression		291.30
Error	27	132.12
Total	29	423.42

Baltz is concerned that violations of regression assumptions may affect the utility of the model for forecasting purposes. He is especially concerned about a situation where the coefficient estimate for an independent variable could take on opposite sign to that predicted.

Baltz is also concerned about important variables being left out of the model. He makes the following statement:

"If an omitted variable is correlated with one of the independent variables included in the model, the standard errors and coefficient estimates will be inconsistent."

Question #82 of 100

Question ID: 485655

If GDP rises 2.2% and the price of fuels falls $0.15, Baltz's model will predict Company sales to be (in $ millions) closest to:

✗ **A) $128.**

✗ **B)** $82.

✓ **C)** $206.

Explanation

Sales will be closest to $78 + ($30.22 × 2.2) + [(−412.39) × (−$0.15)] = $206.34 million.

(LOS 10.e)

Question #83 of 100

Question ID: 485656

Baltz proceeds to test the hypothesis that none of the independent variables has significant explanatory power. He concludes that, at a

5% level of significance:

✗ **A) none of the independent variables has explanatory power, because the calculated F-statistic does not exceed its critical value.**

✓ **B)** at least one of the independent variables has explanatory power, because the calculated F-statistic exceeds its critical value.

✗ **C)** all of the independent variables have explanatory power, because the calculated F-statistic exceeds its critical value.

Explanation

MSE = SSE / [n − (k + 1)] = 132.12 ÷ 27 = 4.89. From the ANOVA table, the calculated F-statistic is (mean square regression / mean square error) = 145.65 / 4.89 = 29.7853. From the F distribution table (2 df numerator, 27 df denominator) the F-critical value may be interpolated to be 3.36. Because 29.7853 is greater than 3.36, Baltz rejects the null hypothesis and concludes that at least one of the independent variables has explanatory power.

(LOS 10.g)

Question #84 of 100

Question ID: 485657

Baltz then tests the individual variables, at a 5% level of significance, to determine whether sales are explained by changes in GDP and fuel prices. Baltz concludes that:

✗ **A) only GDP changes explain changes in sales.**

✓ **B)** both GDP and fuel price changes explain changes in sales.

✗ **C)** neither GDP nor fuel price changes explain changes in sales.

Explanation

From the ANOVA table, the calculated *t*-statistics are (30.22 / 12.12) = 2.49 for GDP and (−412.39 / 183.981) = −2.24 for fuel prices. These values are both beyond the critical t-value at 27 degrees of freedom of ±2.052. Therefore, Baltz is able to reject the null hypothesis that these coefficients are equal to zero, and concludes that both variables are important in explaining sales.

(LOS 10.c)

Question #85 of 100

Question ID: 485658

With regards to violation of regression assumptions, Baltz should *most appropriately* be concerned about:

✗ **A) Serial correlation.**

✓ **B)** Multicollinearity.

✗ **C)** Conditional Heteroskedasticity.

Explanation

Multicollinearity is a violation of regression assumptions that may affect consistency of estimates of slope coefficients and possibly lead to estimates having the opposite sign to that expected. Heteroskedasticity and serial correlation do not affect consistency of coefficient estimates.

(LOS 10.k,l)

Question #86 of 100

Question ID: 485659

Regarding the statement about omitted variables made by Baltz, which of the following is *most accurate*? The statement:

✗ **A) is incorrect about coefficient estimates but correct about standard errors.**

✗ **B)** is incorrect about standard errors but correct about coefficient estimates.

✓ **C)** is correct.

Explanation

Baltz's statement is correct. If an omitted variable is correlated with one of the independent variables in the model, the coefficient estimates will be biased and inconsistent and standard errors will be inconsistent.

(LOS 10.m)

Question #87 of 100

Question ID: 485660

Presence of conditional heteroskedasticity is *least likely* to affect the:

✗ **A) computed F-statistic.**

✗ **B)** computed t-statistic.

✓ **C)** coefficient estimates.

Explanation

Conditional heteroskedasticity results in consistent coefficient estimates, but it biases standard errors, affecting the computed t-statistic and F-statistic

(LOS 10.k)

Question #88 of 100

Question ID: 461655

Which of the following statements regarding the R^2 is *least* accurate?

✗ **A) It is possible for the adjusted-R^2 to decline as more variables are added to the multiple regression.**

✓ **B)** The adjusted-R^2 is greater than the R^2 in multiple regression.

✗ **C)** The adjusted-R^2 not appropriate to use in simple regression.

Explanation

The adjusted-R^2 will always be less than R^2in multiple regression.

Question #89 of 100

Question ID: 461671

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) four dummy variables.**

✗ **B)** one dummy variables.

✓ **C)** three 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.

Question #90 of 100

Question ID: 461747

An analyst runs a regression of portfolio returns on three independent variables. These independent variables are price-to-sales (P/S), price-to-cash flow (P/CF), and price-to-book (P/B). The analyst discovers that the p-values for each independent variable are relatively high. However, the F-test has a very small p-value. The analyst is puzzled and tries to figure out how the F-test can be statistically significant when the individual independent variables are not significant. What violation of regression analysis has occurred?

✗ **A) conditional heteroskedasticity.**

✗ **B)** serial correlation.

✓ **C)** multicollinearity.

Explanation

An indication of multicollinearity is when the independent variables individually are not statistically significant but the F-test suggests that the variables as a whole do an excellent job of explaining the variation in the dependent variable.

Question #91 of 100

Question ID: 461750

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

✗ **A) Measuring independent variables with errors.**

✗ **B)** Using a lagged dependent variable as an independent variable.

✓ **C)** Transforming a variable.

Explanation

A basic assumption of regression is that the dependent variable is linearly related to each of the independent variables. Frequently, they are not linearly related and the independent variable must be transformed or the model is misspecified. Therefore, transforming an independent variable is a potential solution to a misspecification. Methods used to transform independent variables include squaring the variable or taking the square root.

Question #92 of 100

Question ID: 461595

A dependent variable is regressed against three independent variables across 25 observations. The regression sum of squares is 119.25, and the total sum of squares is 294.45. The following are the estimated coefficient values and standard errors of the coefficients.

Coefficient	*Value*	*Standard error*
1	2.43	1.4200
2	3.21	1.5500
3	0.18	0.0818

For which of the coefficients can the hypothesis that they are equal to zero be rejected at the 0.05 level of significance?

✓ **A) 3 only.**

✗ **B)** 2 and 3 only.

✗ **C)** 1 and 2 only.

Explanation

The values of the *t*-statistics for the three coefficients are equal to the coefficients divided by the standard errors, which are 2.43 / 1.42 = 1.711, 3.21 / 1.55 = 2.070, and 0.18 / 0.0818 = 2.200. The statistic has 25 − 3 − 1 = 21 degrees of freedom. The critical value for a p-value of 0.025 (because this is a two-sided test) is 2.080, which means only coefficient 3 is significant.

Question #93 of 100

Question ID: 461751

What is the main difference between probit models and typical dummy variable models?

✓ **A) A dummy variable represents a qualitative independent variable, while a probit model is used for estimating the probability of a qualitative dependent variable.**

✗ **B)** There is no difference--a probit model is simply a special case of a dummy variable regression.

✗ **C)** Dummy variable regressions attempt to create an equation to classify items into one of two categories, while probit models estimate a probability.

Explanation

Dummy variables are used to represent a qualitative independent variable. Probit models are used to estimate the probability of occurrence for a qualitative dependent variable.

Questions #94-99 of 100

Kathy Williams, CFA, and Nigel Faber, CFA, have been managing a hedge fund over the past 18 months. The fund's objective is to eliminate all systematic risk while earning a portfolio return greater than the return on Treasury Bills. Williams and Faber want to test whether they have achieved this objective. Using monthly data, they find that the average monthly return for the fund was 0.417%, and the average return on Treasury Bills was 0.384%. They perform the following regression (Equation I):

$$(\text{fund return})t = b_0 + b_1\,(\text{T-bill return})_t + b_2\,(\text{S\&P 500 return})_t + b_3\,(\text{global index return})_t + e_t$$

The correlation matrix for the independent variables appears below:

	S&P 500	*Global Index*
T-bill	0.163	0.141
S&P 500		0.484

In performing the regression, they obtain the following results for Equation I:

Variable	*Coefficient*	*Standard Error*
Intercept	0.232	0.098
T-bill return	0.508	0.256
S&P 500 Return	−0.0161	0.032
Global index return	0.0037	0.034

R^2 = 22.44%

adj. R^2 = 5.81%

standard error of forecast = 0.0734 (percent)

Williams argues that the equation may suffer from multicollinearity and reruns the regression omitting the return on the global index. This time, the regression (Equation II) is:

$$(\text{fund return})_t = b_0 + b_1\,(\text{T-bill return})_t + b_2\,(\text{S\&P 500 return})_t + e_t$$

The results for Equation II are:

Variable	*Coefficient*	*Standard Error*
Intercept	0.232	0.095
T-bill return	0.510	0.246
S&P 500 return	−0.015	0.028

R^2 = 22.37%

adj. R^2 = 12.02%

standard error of forecast = 0.0710 (percent)

Based on the results of equation II, Faber concludes that a 1% increase in t-bill return leads to more than one half of 1% increase in the fund return.

Finally, Williams reruns the regression omitting the return on the S&P 500 as well. This time, the regression (Equation III) is:

$$(\text{fund return})_t = b_0 + b_1\,(\text{T-bill return})_t + e_t$$

The results for Equation III are:

Variable	*Coefficient*	*Standard Error*
Intercept	0.229	0.093
T-bill return	0.4887	0.2374

R^2 = 20.94%

adj. R^2 = 16.00%

standard error of forecast = 0.0693 (percent)

Question #94 of 100

Question ID: 485599

In the regression using Equation I, which of the following hypotheses can be rejected at a 5% level of significance in a two-tailed test? (The corresponding independent variable is indicated after each null hypothesis.)

✓ **A) H_0: b_0 = 0 (intercept)**

✗ **B)** H_0: b_2 = 0 (S&P 500)

✗ **C)** H_0: b_1 = 0 (T-bill)

Explanation

The critical *t*-value for 18 − 3 − 1 = 14 degrees of freedom in a two-tailed test at a 5% significance level is 2.145. Although the *t*-statistic for T-bill is close at 0.508 / 0.256 = 1.98, it does not exceed the critical value. Only the intercept's coefficient has a significant *t*-statistic for the indicated test: *t* = 0.232 / 0.098 = 2.37. (Study Session 3, LOS 10.e)

Question #95 of 100

Question ID: 485600

In the regression using Equation II, which of the following hypothesis or hypotheses can be rejected at a 5% level of significance in a two-tailed test? (The corresponding independent variable is indicated after each null hypothesis.)

✓ **A) H_0: b_0 = 0 (intercept) only.**

✗ **B)** H_0: b_0 = 0 (intercept) and b1 = 0 (T-bill) only.

✗ **C)** H_0: b_1 = 0 (T-bill) and H_0: b_2 = 0 (S&P 500) only.

Explanation

The critical *t*-value for 18 − 2 − 1 = 15 degrees of freedom in a two-tailed test at a 5% significance level is 2.131. The *t*-statistics on the intercept, T-bill and S&P 500 coefficients are 2.442, 2.073, −0.536, respectively. Therefore, only the coefficient on the intercept is significant. (Study Session 3, LOS 10.e)

Question #96 of 100

Question ID: 485601

With respect to multicollinearity and Williams' removal of the global index variable when running regression Equation II, Williams had:

✓ **A) reason to be suspicious and took the correct step to cure the problem.**

✗ **B)** reason to be suspicious, but she took the wrong step to cure the problem.

✗ **C)** no reason to be suspicious, but took a correct step to improve the analysis.

Explanation

Investigating multicollinearity is justified for two reasons. First, the S&P 500 and the global index have a significant degree of correlation. Second, neither of the market index variables are significant in the first specification. The correct step is to remove one of the variables, as Williams did, to see if the remaining variable becomes significant. (Study Session 3, LOS 10.n)

Question #97 of 100

Question ID: 485602

Regarding Faber's conjecture about impact of t-bill return in equation II, the most appropriate null hypothesis and most appropriate conclusion (at a 5% level of significance) is:

	Null Hypothesis	Conclusion
✗ **A)**	**H_0: $b_1 \leq 0.5$**	**Reject H_0**
✗ **B)**	H_0: $b_1 \geq 0.5$	Fail to reject H_0
✓ **C)**	H_0: $b_1 \leq 0.5$	Fail to reject H_0

Explanation

Null hypothesis is opposite to Faber's conclusion. The critical *t*-value for 18 − 2 − 1 = 15 degrees of freedom in a one-tailed test at a 5% significance level is 1.753.

t = (0.51 − 0.50)/0.246 = 0.04065 (<1.753). Hence fail to reject the null hypothesis. (Study Session 3, LOS 10.e)

Question #98 of 100

Question ID: 485603

Which of the following problems, multicollinearity and/or serial correlation, can bias the estimates of the slope coefficients?

- ✓ **A) Neither multicollinearity, nor serial correlation.**
- ✗ **B)** Both multicollinearity and serial correlation.
- ✗ **C)** Multicollinearity, but not serial correlation.

Explanation

Neither multicollinearity not serial correlation affects the consistency (i.e. make them biased) of regression coefficients. Multicollinearity can however make the regression coefficients *unreliable*. Both multicollinearity and serial correlation biases the standard errors of the slope coefficients. (Study Session 3, LOS 10.n)

Question #99 of 100

Question ID: 485604

If we expect that next month the T-bill rate will equal its average over the last 18 months, using Equation III, calculate the 95% confidence interval for the expected fund return.

- ✗ **A) 0.259 to 0.598.**
- ✗ **B)** 0.296 to 0.538.
- ✓ **C)** 0.270 to 0.564.

Explanation

The forecast is 0.417 = 0.229 + 0.4887 × (0.384). The 95% confidence interval is Y ± (t_c × s_f) and t_c for 16 degrees of freedom for a 2 tailed test = 2.120. The 95% confidence interval = 0.417 ± (2.120)(.0693) = 0.270 to 0.564. (Study Session 3, LOS 10.g)

Question #100 of 100

Question ID: 461659

Which of the following statements regarding the analysis of variance (ANOVA) table is *least* accurate? The:

✗ **A) standard error of the estimate is the square root of the mean square error.**

✓ **B)** F-statistic cannot be computed with the data offered in the ANOVA table.

✗ **C)** F-statistic is the ratio of the mean square regression to the mean square error.

Explanation

The F-statistic can be calculated using an ANOVA table. The F-statistic is MSR/MSE.