Multiple Regression and Issues in Regression Analysis 2

Test ID: 7440356

Question #1 of 106

Question ID: 461642

Consider the following analysis of variance (ANOVA) table:

Source	*Sum of squares*	*Degrees of freedom*	*Mean square*
Regression	20	1	20
Error	80	40	2
Total	100	41	

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

✗ **A) 0.10.**

✓ **B)** 10.00.

✗ **C)** 0.25.

Explanation

The F-statistic is equal to the ratio of the mean squared regression to the mean squared error.

F = MSR/MSE = 20 / 2 = 10.

Questions #2-7 of 106

An analyst is interested in forecasting the rate of employment growth and instability for 254 metropolitan areas around the United States. The analyst's main purpose for these forecasts is to estimate the demand for commercial real estate in each metro area. The independent variables in the analysis represent the percentage of employment in each industry group.

Regression of Employment Growth Rates and Employment Instability on Industry Mix Variables for 254 U.S. Metro Areas				
	Model 1		*Model 2*	
Dependent Variable	Employment Growth Rate		Relative Employment Instability	
Independent Variables	Coefficient Estimate	*t*-value	Coefficient Estimate	*t*-value
Intercept	-2.3913	-0.713	3.4626	0.623
% Construction Employment	0.2219	4.491	0.1715	2.096
% Manufacturing Employment	0.0136	0.393	0.0037	0.064
% Wholesale Trade Employment	-0.0092	-0.171	0.0244	0.275

% Retail Trade Employment	-0.0012	-0.031	-0.0365	-0.578
% Financial Services Employment	0.0605	1.271	-0.0344	-0.437
% Other Services Employment	0.1037	2.792	0.0208	0.338
R^2	0.289		0.047	
Adjusted R^2	0.272		0.024	
F-Statistic	16.791		2.040	
Standard error of estimate	0.546		0.345	

Question #2 of 106

Question ID: 485606

Based on the data given, which independent variables have both a *statistically* and an *economically* significant impact (at the 5% level) on metropolitan employment growth rates?

✗ A) **"% Manufacturing Employment," "% Financial Services Employment," "% Wholesale Trade Employment," and "% Retail Trade" only.**

✗ B) "% Wholesale Trade Employment" and "% Retail Trade" only.

✓ C) "% Construction Employment" and "% Other Services Employment" only.

Explanation

The percentage of construction employment and the percentage of other services employment have a statistically significant impact on employment growth rates in U.S. metro areas. The *t*-statistics are 4.491 and 2.792, respectively, and the critical *t* is 1.96 (95% confidence and 247 degrees of freedom). In terms of economic significance, construction and other services appear to be significant. In other words, as construction employment rises 1%, the employment growth rate rises 0.2219%. The coefficients of all other variables are too close to zero to ascertain any economic significance, and their *t*-statistics are too low to conclude that they are statistically significant. Therefore, there are only two independent variables that are both statistically and economically significant: "% of construction employment" and "% of other services employment".

Some may argue, however, that financial services employment is also economically significant even though it is not statistically significant because of the magnitude of the coefficient. Economic significance can occur without statistical significance if there are statistical problems. For instance, the multicollinearity makes it harder to say that a variable is statistically significant. (Study Session 3, LOS 10.o)

Question #3 of 106

Question ID: 485607

The coefficient standard error for the independent variable "% Construction Employment" under the relative employment instability model is closest to:

✗ A) **0.3595.**

✓ B) 0.0818.

✗ C) 2.2675.

Explanation

The *t*-statistic is computed by *t*-statistic = slope coefficient / coefficient standard error. Therefore, the coefficient standard error

=

$s_{\hat{b}_1}$ = slope coefficient/the t-statistic = 0.1715/2.096 = 0.0818. (Study Session 3, LOS 10.a)

Question #4 of 106

Question ID: 485608

Which of the following *best* describes how to interpret the R^2 for the employment growth rate model? Changes in the value of the:

✗ A) **employment growth rate explain 28.9% of the variability of the independent variables.**

✗ B) independent variables cause 28.9% of the variability of the employment growth rate.

✓ C) independent variables explain 28.9% of the variability of the employment growth rate.

Explanation

The R^2 indicates the percent variability of the dependent variable that is explained by the variability of the independent variables. In the employment growth rate model, the variability of the independent variables explains 28.9% of the variability of employment growth. Regression analysis does not establish a causal relationship. (Study Session 3, LOS 10.h)

Question #5 of 106

Question ID: 485609

Using the following forecasts for Cedar Rapids, Iowa, the forecasted employment growth rate for that city is *closest* to:

Construction employment	10%
Manufacturing	30%
Wholesale trade	5%
Retail trade	20%
Financial services	15%
Other services	20%

✓ A) **3.15%.**

✗ B) 5.54%.

✗ C) 3.22%.

Explanation

The forecast uses the intercept and coefficient estimates for the model. The forecast is:

$\hat{Y}$ = −2.3913 + (0.2219)(10) + (0.0136)(30) + (−0.0092)(5) + (−0.0012)(20) + (0.0605)(15) + (0.1037)(20) = 3.15%. (Study Session 3, LOS 10.e)

Question #6 of 106

Question ID: 485610

The 95% confidence interval for the coefficient estimate for "% Construction Employment" from the relative employment instability model is *closest* to:

✓ **A) 0.0111 to 0.3319.**

✗ **B)** 0.0897 to 0.2533.

✗ **C)** -0.0740 to 0.4170.

Explanation

With a sample size of 254, and 254 − 6 − 1 = 247 degrees of freedom, the critical value for a two-tail 95% *t*-statistic is very close to the two-tail 95% statistic of 1.96. Using this critical value, the formula for the 95% confidence interval for the *j*th coefficient estimate is:

95% confidence interval = $\hat{b}_j \pm 1.96(s_{\hat{b}_j})$. But first we need to figure out the coefficient standard error:

$$s_{\hat{b}_1} = \frac{\hat{b}_j}{t_j} = \frac{0.1715}{2.096} = 0.08182.$$

Hence, the confidence interval is 0.1715 ± 1.96(0.08182).

With 95% probability, the coefficient will range from 0.0111 to 0.3319, 95% CI = {0.0111 < b1 < 0.3319}. (Study Session 3, LOS 9.f)

Question #7 of 106

Question ID: 485611

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 Durbin-Watson statistic differs sufficiently from 2.**

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

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

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 differs sufficiently from 2, this is a sign that the regression errors have significant serial correlation.

(Study Session 3, LOS 10.l)

Question #8 of 106

Question ID: 461756

Mary Steen estimated that if she purchased shares of companies who announced restructuring plans at the announcement and held them for five days, she would earn returns in excess of those expected from the market model of 0.9%. These returns are statistically significantly different from zero. The model was estimated without transactions costs, and in reality these would approximate 1% if the strategy were effected. This is an example of:

✓ **A) statistical significance, but not economic significance.**

✗ **B)** statistical and economic significance.

✗ **C)** a market inefficiency.

Explanation

The abnormal returns are not sufficient to cover transactions costs, so there is no economic significance to this trading strategy. This is not an example of market inefficiency because excess returns are not available after covering transactions costs.

Question #9 of 106

Question ID: 461597

Seventy-two monthly stock returns for a fund between 1997 and 2002 are regressed against the market return, measured by the Wilshire 5000, and two dummy variables. The fund changed managers on January 2, 2000. Dummy variable one is equal to 1 if the return is from a month between 2000 and 2002. Dummy variable number two is equal to 1 if the return is from the second half of the year. There are 36 observations when dummy variable one equals 0, half of which are when dummy variable two also equals zero. The following are the estimated coefficient values and standard errors of the coefficients.

Coefficient	*Value*	*Standard error*
Market	1.43000	0.319000
Dummy 1	0.00162	0.000675
Dummy 2	−0.00132	0.000733

What is the p-value for a test of the hypothesis that the beta of the fund is greater than 1?

✓ **A) Between 0.05 and 0.10.**

✗ **B)** Lower than 0.01.

✗ **C)** Between 0.01 and 0.05.

Explanation

The beta is measured by the coefficient of the market variable. The test is whether the beta is greater than 1, not zero, so the *t*-statistic is equal to (1.43 − 1) / 0.319 = 1.348, which is in between the *t*-values (with 72 − 3 − 1 = 68 degrees of freedom) of 1.29 for a *p*-value of 0.10 and 1.67 for a p-value of 0.05.

Questions #10-15 of 106

Autumn Voiku is attempting to forecast sales for Brookfield Farms based on a multiple regression model. Voiku has constructed the following model:

> sales = b_0 + (b_1 × CPI) + (b_2 × IP) + (b_3 × GDP) + ε_t
>
> Where:
>
> sales = $ change in sales (in 000's)

> CPI = change in the consumer price index
>
> IP = change in industrial production (millions)
>
> GDP = change in GDP (millions)
>
> All changes in variables are in percentage terms.

Voiku uses monthly data from the previous 180 months of sales data and for the independent variables. The model estimates (with coefficient standard errors in parentheses) are:

sales = 10.2 + (4.6 × CPI) + (5.2 × IP) + (11.7 × GDP)

(5.4) (3.5) (5.9) (6.8)

The sum of squared errors is 140.3 and the total sum of squares is 368.7.

Voiku calculates the unadjusted R^2, the adjusted R^2, and the standard error of estimate to be 0.592, 0.597, and 0.910, respectively.

Voiku is concerned that one or more of the assumptions underlying multiple regression has been violated in her analysis. In a conversation with Dave Grimbles, CFA, a colleague who is considered by many in the firm to be a quant specialist, Voiku says, "It is my understanding that there are five assumptions of a multiple regression model:"

> Assumption 1: There is a linear relationship between the dependent and independent variables.
>
> Assumption 2: The independent variables are not random, and there is zero correlation between any two of the independent variables.
>
> Assumption 3: The residual term is normally distributed with an expected value of zero.
>
> Assumption 4: The residuals are serially correlated.
>
> Assumption 5: The variance of the residuals is constant.

Grimbles agrees with Miller's assessment of the assumptions of multiple regression.

Voiku tests and fails to reject each of the following four null hypotheses at the 99% confidence interval:

> Hypothesis 1: The coefficient on GDP is negative.
>
> Hypothesis 2: The intercept term is equal to -4.
>
> Hypothesis 3: A 2.6% increase in the CPI will result in an increase in sales of more than 12.0%.
>
> Hypothesis 4: A 1% increase in industrial production will result in a 1% decrease in sales.

Figure 1: Partial table of the Student's t-distribution (One-tailed probabilities)

df	*p* = 0.10	*p* = 0.05	*p* = 0.025	*p* = 0.01	*p* = 0.005
170	1.287	1.654	1.974	2.348	2.605
176	1.286	1.654	1.974	2.348	2.604
180	1.286	1.653	1.973	2.347	2.603

Figure 2: Partial F-Table critical values for right-hand tail area equal to 0.05

	df1 = 1	*df1* = 3	*df1* = 5
df2 = 170	3.90	2.66	2.27

df2 = 176	3.89	2.66	2.27
df2 = 180	3.89	2.65	2.26

Figure 3: Partial F-Table critical values for right-hand tail area equal to 0.025

	df1 = 1	*df1 = 3*	*df1 = 5*
df2 = 170	5.11	3.19	2.64
df2 = 176	5.11	3.19	2.64
df2 = 180	5.11	3.19	2.64

Question #10 of 106

Question ID: 461564

Concerning the assumptions of multiple regression, Grimbles is:

✗ **A)** **correct to agree with Voiku's list of assumptions.**

✗ **B)** incorrect to agree with Voiku's list of assumptions because one of the assumptions is stated incorrectly.

✓ **C)** incorrect to agree with Voiku's list of assumptions because two of the assumptions are stated incorrectly.

Explanation

Assumption 2 is stated incorrectly. Some correlation between independent variables is unavoidable; and high correlation results in multicollinearity. However, an exact linear relationship between linear combinations of two or more independent variables should not exist.

Assumption 4 is also stated incorrectly. The assumption is that the residuals are serially uncorrelated (i.e., they are not serially correlated).

Question #11 of 106

Question ID: 461565

For which of the four hypotheses did Voiku incorrectly fail to reject the null, based on the data given in the problem?

✗ **A)** **Hypothesis 3.**

✓ **B)** Hypothesis 2.

✗ **C)** Hypothesis 4.

Explanation

The critical values at the 1% level of significance (99% confidence) are 2.348 for a one-tail test and 2.604 for a two-tail test (df = 176).

The t-values for the hypotheses are:
Hypothesis 1: 11.7 / 6.8 = 1.72
Hypothesis 2: 14.2 / 5.4 = 2.63
Hypothesis 3: 12.0 / 2.6 = 4.6, so the hypothesis is that the coefficient is greater than 4.6, and the t-stat of that hypothesis is (4.6 − 4.6) / 3.5 = 0.
Hypothesis 4: (5.2 + 1) / 5.9 = 1.05

Hypotheses 1 and 3 are one-tail tests; 2 and 4 are two-tail tests. Only Hypothesis 2 exceeds the critical value, so only

Hypothesis 2 should be rejected.

Question #12 of 106

Question ID: 461566

The *most* appropriate decision with regard to the F-statistic for testing the null hypothesis that all of the independent variables are simultaneously equal to zero at the 5 percent significance level is to:

✓ **A) reject the null hypothesis because the F-statistic is larger than the critical F-value of 2.66.**

✗ **B)** reject the null hypothesis because the F-statistic is larger than the critical F-value of 3.19.

✗ **C)** fail to reject the null hypothesis because the F-statistic is smaller than the critical F-value of 2.66.

Explanation

RSS = 368.7 - 140.3 = 228.4, F-statistic = (228.4 / 3) / (140.3 / 176) = 95.51. The critical value for a one-tailed 5% F-test with 3 and 176 degrees of freedom is 2.66. Because the F-statistic is greater than the critical F-value, the null hypothesis that all of the independent variables are simultaneously equal to zero should be rejected.

Question #13 of 106

Question ID: 461567

Regarding Voiku's calculations of R^2 and the standard error of estimate, she is:

✗ **A) incorrect in her calculation of the unadjusted R^2 but correct in her calculation of the standard error of estimate.**

✗ **B)** correct in her calculation of the unadjusted R^2 but incorrect in her calculation of the standard error of estimate.

✓ **C)** incorrect in her calculation of both the unadjusted R^2 and the standard error of estimate.

Explanation

SEE = $\sqrt{[140.3 / (180 - 3 - 1)]}$ = 0.893

unadjusted R^2 = (368.7 − 140.3) / 368.7 = 0.619

Question #14 of 106

Question ID: 461568

The multiple regression, as specified, *most likely* suffers from:

✓ **A) multicollinearity.**

✗ **B)** heteroskedasticity.

✗ **C)** serial correlation of the error terms.

Explanation

The regression is highly significant (based on the F-stat in Part 3), but the individual coefficients are not. This is a result of a regression with significant multicollinearity problems. The t-stats for the significance of the regression coefficients are, respectively, 1.89, 1.31, 0.88, 1.72. None of these are high enough to reject the hypothesis that the coefficient is zero at the 5% level of significance (two-tailed critical value of 1.974 from t-table).

Question #15 of 106

Question ID: 461569

A 90 percent confidence interval for the coefficient on GDP is:

- ✗ **A) -1.5 to 20.0.**
- ✓ **B)** 0.5 to 22.9.
- ✗ **C)** -1.9 to 19.6.

Explanation

A 90% confidence interval with 176 degrees of freedom is coefficient ± $t_c(s_e)$ = 11.7 ± 1.654 (6.8) or 0.5 to 22.9.

Question #16 of 106

Question ID: 461625

Which of the following statements *least* accurately describes one of the fundamental multiple regression assumptions?

- ✗ **A) The independent variables are not random.**
- ✗ **B)** The error term is normally distributed.
- ✓ **C)** The variance of the error terms is not constant (i.e., the errors are heteroskedastic).

Explanation

The variance of the error term **IS** assumed to be constant, resulting in errors that are homoskedastic.

Questions #17-22 of 106

Consider a study of 100 university endowment funds that was conducted to determine if the funds' annual risk-adjusted returns could be explained by the size of the fund and the percentage of fund assets that are managed to an indexing strategy. The equation used to model this relationship is:

> $ARAR_i = b_0 + b_1 Size_i + b_2 Index_i + e_i$
>
> Where:
>
> $ARAR_i$ = the average annual risk-adjusted percent returns for the fund *i* over the 1998-2002 time period.
>
> $Size_i$ = the natural logarithm of the average assets under management for fund *i*.
>
> $Index_i$ = the percentage of assets in fund *i* that were managed to an indexing strategy.

The table below contains a portion of the regression results from the study.

Partial Results from Regression ARAR on Size and Extent of Indexing			
	Coefficients	*Standard Error*	*t-Statistic*
Intercept	???	0.55	−5.2

Size	0.6	0.18	???
Index	1.1	???	2.1

Question #17 of 106

Question ID: 485557

Which of the following is the *most* accurate interpretation of the slope coefficient for size? ARAR:

✗ **A)** **will change by 1.0% when the natural logarithm of assets under management changes by 0.6, holding index constant.**

✓ **B)** will change by 0.6% when the natural logarithm of assets under management changes by 1.0, holding index constant.

✗ **C)** and index will change by 1.1% when the natural logarithm of assets under management changes by 1.0.

Explanation

A slope coefficient in a multiple linear regression model measures how much the dependent variable changes for a one-unit change in the independent variable, *holding all other independent variables constant*. In this case, the independent variable size (= ln average assets under management) has a slope coefficient of 0.6, indicating that the dependent variable ARAR will change by 0.6% return for a one-unit change in size, assuming nothing else changes. Pay attention to the units on the dependent variable. (Study Session 3, LOS 10.a)

Question #18 of 106

Question ID: 485558

Which of the following is the estimated standard error of the regression coefficient for index?

✗ **A)** **1.91.**

✗ **B)** 2.31.

✓ **C)** 0.52.

Explanation

The *t*-statistic for testing the null hypothesis H_0: $\beta_i = 0$ is $t = (b_i -; 0) / \beta_i$, where β_i is the population parameter for independent variable *i*, b_i is the estimated coefficient, and β_i is the coefficient standard error. Using the information provided, the estimated coefficient standard error can be computed as b_{Index} / t = β_{Index} = 1.1 / 2.1 = 0.5238.

(Study Session 3, LOS 10.c)

Question #19 of 106

Question ID: 485559

Which of the following is the *t*-statistic for size?

✗ **A)** **0.70.**

✓ **B)** 3.33.

✗ **C)** 0.30.

Explanation

The *t*-statistic for testing the null hypothesis H_0: $\beta_i = 0$ is $t = (b_i - 0) / \sigma_i$, where β_i is the population parameter for independent variable *i*, b_i is the estimated coefficient, and σ_i is the coefficient standard error. Using the information provided, the *t*-statistic for size can be computed as t = b_{Size} / σ_{Size} = 0.6 / 0.18 = 3.3333.

(Study Session 3, LOS 10.c)

Question #20 of 106

Question ID: 485560

Which of the following is the estimated intercept for the regression?

✗ **A)** **−9.45.**

✗ **B)** −0.11.

✓ **C)** −2.86.

Explanation

The *t*-statistic for testing the null hypothesis H_0: $\beta_i = 0$ is $t = (b_i - 0) / \sigma_i$, where σ_i is the population parameter for independent variable *i*, b_i is the estimated parameter, and σ_i is the parameter's standard error. Using the information provided, the estimated intercept can be computed as $b_0 = t \times \sigma_0 = -5.2 \times 0.55 = -2.86$.

(Study Session 3, LOS 10.c)

Question #21 of 106

Question ID: 485561

Which of the following statements is *most* accurate regarding the significance of the regression parameters at a 5% level of significance?

✓ **A)** **All of the parameter estimates are significantly different than zero at the 5% level of significance.**

✗ **B)** The parameter estimates for the intercept and the independent variable size are significantly different than zero. The coefficient for index is not significant.

✗ **C)** The parameter estimates for the intercept are significantly different than zero. The slope coefficients for index and size are not significant.

Explanation

At 5% significance and 97 degrees of freedom (100 − 3), the critical *t*-value is slightly greater than, but very close to, 1.984. The *t*-statistic for the intercept and *index* are provided as −5.2 and 2.1, respectively, and the *t*-statistic for size is computed as 0.6 / 0.18 = 3.33. The absolute value of the all of the regression intercepts is greater than $t_{critical}$ = 1.984. Thus, it can be concluded that all of the parameter estimates are significantly different than zero at the 5% level of significance.

(Study Session 3, LOS 10.c)

Question #22 of 106

Question ID: 485562

Which of the following is NOT a required assumption for multiple linear regression?

✗ **A)** **The error term is normally distributed.**

✓ **B)** The error term is linearly related to the dependent variable.

✗ **C)** The expected value of the error term is zero.

Explanation

The assumptions of multiple linear regression include: linear relationship between dependent and independent variable, independent variables are not random and no exact linear relationship exists between the two or more independent variables, error term is normally distributed with an expected value of zero and constant variance, and the error term is serially uncorrelated. (Study Session 3, LOS 10.f)

Question #23 of 106

Question ID: 461526

Consider the following regression equation:

> $Sales_i = 20.5 + 1.5\ R\&D_i + 2.5\ ADV_i - 3.0\ COMP_i$
>
> 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, and COMP is the number of competitors in the industry.

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

✓ **A) If a company spends $1 more on R&D (holding everything else constant), sales are expected to increase by $1.5 million.**

✗ **B)** If R&D and advertising expenditures are $1 million each and there are 5 competitors, expected sales are $9.5 million.

✗ **C)** One more competitor will mean $3 million less in sales (holding everything else constant).

Explanation

If a company spends $1 *million* more on R&D (holding everything else constant), sales are expected to increase by $1.5 million. Always be aware of the units of measure for the different variables.

Question #24 of 106

Question ID: 461749

When constructing a regression model to predict portfolio returns, an analyst runs a regression for the past five year period. After examining the results, she determines that an increase in interest rates two years ago had a significant impact on portfolio results for the time of the increase until the present. By performing a regression over two separate time periods, the analyst would be attempting to prevent which type of misspecification?

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

✗ **B)** Forecasting the past.

✓ **C)** Incorrectly pooling data.

Explanation

The relationship between returns and the dependent variables can change over time, so it is critical that the data be pooled correctly. Running the regression for multiple sub-periods (in this case two) rather than one time period can produce more accurate results.

Question #25 of 106

Question ID: 461591

Seventy-two monthly stock returns for a fund between 2007 and 2012 are regressed against the market return, measured by the Wilshire 5000, and two dummy variables. The fund changed managers on January 2, 2010. Dummy variable one is equal to 1 if the return is from a month between 2010 and 2012. Dummy variable number two is equal to 1 if the return is from the second half of the year. There are 36 observations when dummy variable one equals 0, half of which are when dummy variable two also equals 0. The following are the estimated coefficient values and standard errors of the coefficients.

Coefficient	*Value*	*Standard error*
Market	1.43000	0.319000
Dummy 1	0.00162	0.000675
Dummy 2	−0.00132	0.000733

What is the *p*-value for a test of the hypothesis that the new manager outperformed the old manager?

✓ **A) Lower than 0.01.**

✗ **B)** Between 0.01 and 0.05.

✗ **C)** Between 0.05 and 0.10.

Explanation

Dummy variable one measures the effect on performance of the change in managers. H_0: Dummy 1<=0 vs. Dummy 1>0 (this is a one-tailed test). The *t*-statistic is equal to 0.00162 / 0.000675 = 2.400, which is higher than the *t*-value (with 72 - 3 - 1 = 68 degrees of freedom) of approximately 2.39 for a *p*-value of between 0.01 and 0.005 for a 1 tailed test.

Question #26 of 106

Question ID: 461657

May Jones estimated a regression that produced the following analysis of variance (ANOVA) table:

Source	*Sum of squares*	*Degrees of freedom*	*Mean square*
Regression	20	1	20
Error	80	40	2
Total	100	41	

The values of R^2 and the F-statistic for the fit of the model are:

✗ **A) R^2 = 0.25 and F = 0.909.**

✗ **B)** R^2 = 0.25 and F = 10.

✓ **C)** R^2 = 0.20 and F = 10.

Explanation

R^2 = RSS / SST = 20 / 100 = 0.20

The F-statistic is equal to the ratio of the mean squared regression to the mean squared error.

F = 20 / 2 = 10

Questions #27-32 of 106

Som Muttney has been asked to forecast the level of operating profit for a proposed new branch of a tire store. His forecast is one component in forecasting operating profit for the entire company for the next fiscal year. Muttney decide to conduct multiple regression analysis using "branch store operating profit" as the dependent variable and three independent variables. The three independent variables are "population within 5 miles of the branch," "operating hours per week," and "square footage of the facility." Muttney used data on the company's existing 23 branches to develop the model (n=23).

Regression of Operating Profit on Population, Operating Hours, and Square Footage

Dependent Variable	*Operating Profit (Y)*	
Independent Variables	*Coefficient Estimate*	*t-value*
Intercept	103,886	2.740
Population within 5 miles (X_1)	4.372	2.133
Operating hours per week (X_2)	214.856	0.258
Square footage of facility (X_3)	6.767	2.643
Regression sum of squares	6,349	
Sum of squares total	10,898	

	Two-tailed Significance				
Degrees of Freedom	.20	.10	.05	.02	.01
3	1.638	2.353	3.182	4.541	5.841
19	1.328	1.729	2.093	2.539	2.861
23	1.319	1.714	2.069	2.50	2.807

In his research report, Muttney claims that when the square footage of the store is increased by 1%, operating profit will increase by more than 5%

Question #27 of 106

Question ID: 485592

The 95% confidence interval for slope coefficient for independent variable "population" is closest to:

✓ **A) 0.081 − 8.66**

✗ **B)** −0.81 − 9.56

✗ **C)** −0.086 − 8.83

Explanation

The degrees of freedom are [n − k − 1]. Here, n is the number of observations in the regression (23) and k is the number of independent variables (3). df = [23 − 3 − 1] = 19. t_c (for α = 5/2 = 2.5%) = 2.093.
S_e (beta for population) = beta/t-value = 4.372 / 2.133 = 2.05
95% confidence interval = Coefficient ± t_c x S_e = 4.372 ± 2.093 x 2.05 = 0.08135 - 8.66265

(LOS 10.e)

Question #28 of 106

Question ID: 485593

The probability of finding a value of *t* for variable X_1 that is as-large or larger than |2.133| when the null hypothesis is true is:

✗ **A)** **between 1% and 2%.**

✗ **B)** between 5% and 10%.

✓ **C)** between 2% and 5%.

Explanation

The degrees of freedom is = (n − k − 1)

= (23 − 3 − 1)

= 19

In the table above, for 19 degrees of freedom, the value 2.133 would lie between a 2% chance (alpha of 0.02) or 2.539 and a 5% chance (alpha of 0.05) or 2.093.

(LOS 10.b)

Question #29 of 106

Question ID: 485594

The correlation between the actual values of operating profit and the predicted value of operating profit is closest to:

✗ **A)** **0.36**

✓ **B)** 0.76

✗ **C)** 0.53

Explanation

R^2 = RSS/SST = 6,349/10,898 = 0.58. Correlation between predicted and actual values of dependent variable = $(0.58)^{0.5}$ = 0.76

(LOS 10.h)

Question #30 of 106

Question ID: 485595

Regarding Muttney's claim about a 5% increase in operating profit for a 1% increase in square footage, the most appropriate null hypothesis and conclusion (at a 5% level of significance) are:

> Null Hypothesis Conclusion

✗ **A)** **H_0: $b_3 \leq 5$** **Reject H_0**

✓ **B)** H_0: b3 ≤ 5 Fail to reject H_0

✗ **C)** H_0: $b_3 \geq 5$ Fail to reject H_0

Explanation

S_e (beta for sq footage) = beta/t-value = 6.767/2.643 = 2.56
t_c(alpha = 5%, one-tailed, dof = 19) = 1.729

t= beta - $beta_0/S_e$ = 6.767 - 5 /2.56 = 0.69. We fail to reject the null hypothesis

(LOS 10.c)

Question #31 of 106

Question ID: 485596

The standard deviation of regression residuals is closest to:

- ✓ **A) 15.47**
- ✗ **B)** 0.42
- ✗ **C)** 239.42

Explanation

SSE = SST - RSS = 10,898 - 6,349 = 4,549
MSE = SSE/(n-k-1) = 4,549/19 = 239.42
SEE = $(MSE)^{0.5}$ = 15.47

t= beta - $beta_0/S_e$ = 6.767 - 5 /2.56 = 0.69. We fail to reject the null hypothesis.

(LOS 10.i)

Question #32 of 106

Question ID: 485597

The operating profit model as specified is *most likely* a:

- ✗ **A) Time series regression**
- ✓ **B)** Cross-sectional regression
- ✗ **C)** Autoregressive model

Explanation

Cross-sectional data involve many observations for the same time period. Time-series data uses many observations from different time periods for the same entity.

(LOS 10.a)

Questions #33-38 of 106

Dave Turner is a security analyst who is using regression analysis to determine how well two factors explain returns for common stocks. The independent variables are the natural logarithm of the number of analysts following the companies, Ln(no. of analysts), and the natural logarithm of the market value of the companies, Ln(market value). The regression output generated from a statistical program is given in the following tables. Each p-value corresponds to a two-tail test.

Turner plans to use the result in the analysis of two investments. WLK Corp. has twelve analysts following it and a market capitalization of $2.33 billion. NGR Corp. has two analysts following it and a market capitalization of $47 million.

Table 1: Regression Output

Variable	*Coefficient*	*Standard Error of the Coefficient*	*t-statistic*	*p-value*
Intercept	0.043	0.01159	3.71	< 0.001
Ln(No. of Analysts)	−0.027	0.00466	−5.80	< 0.001

Ln(Market Value)	0.006	0.00271	2.21	0.028

Table 2: ANOVA

	Degrees of Freedom	*Sum of Squares*	*Mean Square*
Regression	2	0.103	0.051
Residual	194	0.559	0.003
Total	196	0.662	

Question #33 of 106

Question ID: 485564

In a one-sided test and a 1% level of significance, which of the following coefficients is significantly different from zero?

✗ **A) The coefficient on ln(no. of Analysts) only.**

✓ **B)** The intercept and the coefficient on ln(no. of analysts) only.

✗ **C)** The intercept and the coefficient on ln(market value) only.

Explanation

The p-values correspond to a two-tail test. For a one-tailed test, divide the provided p-value by two to find the minimum level of significance for which a null hypothesis of a coefficient equaling zero can be rejected. Dividing the provided p-value for the intercept and ln(no. of analysts) will give a value less than 0.0005, which is less than 1% and would lead to a rejection of the hypothesis. Dividing the provided p-value for ln(market value) will give a value of 0.014 which is greater than 1%; thus, that coefficient is not significantly different from zero at the 1% level of significance. (Study Session 3, LOS 10.a)

Question #34 of 106

Question ID: 485565

The 95% confidence interval (use a t-stat of 1.96 for this question only) of the estimated coefficient for the independant variable Ln(Market Value) is *closest to*:

✗ **A) 0.014 to -0.009**

✓ **B)** 0.011 to 0.001

✗ **C)** -0.018 to -0.036

Explanation

The confidence interval is 0.006 ± (1.96)(0.00271) = 0.011 to 0.001
(Study Session 3, LOS 10.e)

Question #35 of 106

Question ID: 485566

If the number of analysts on NGR Corp. were to double to 4, the change in the forecast of NGR would be *closest* to?

✗ **A) −0.055.**

✓ **B)** −0.019.

✗ **C)** −0.035.

Explanation

Initially, the estimate is 0.1303 = 0.043 + ln(2)(−0.027) + ln(47000000)(0.006)

Then, the estimate is 0.1116 = 0.043 + ln(4)(−0.027) + ln(47000000)(0.006)

0.1116 − 0.1303 = −0.0187, or −0.019

(Study Session 3, LOS 10.a)

Question #36 of 106

Question ID: 485567

Based on a R^2 calculated from the information in Table 2, the analyst should conclude that the number of analysts and ln(market value) of the firm explain:

✓ **A)** **15.6% of the variation in returns.**

✗ **B)** 84.4% of the variation in returns.

✗ **C)** 18.4% of the variation in returns.

Explanation

R^2 is the percentage of the variation in the dependent variable (in this case, variation of returns) explained by the set of independent variables. R^2 is calculated as follows: R^2 = (SSR / SST) = (0.103 / 0.662) = 15.6%. (Study Session 3, LOS 10.h)

Question #37 of 106

Question ID: 485568

What is the F-statistic from the regression? And, what can be concluded from its value at a 1% level of significance?

✗ **A)** **F = 5.80, reject a hypothesis that both of the slope coefficients are equal to zero.**

✓ **B)** F = 17.00, reject a hypothesis that both of the slope coefficients are equal to zero.

✗ **C)** F = 1.97, fail to reject a hypothesis that both of the slope coefficients are equal to zero.

Explanation

The F-statistic is calculated as follows: F = MSR / MSE = 0.051 / 0.003 = 17.00; and 17.00 > 4.61, which is the critical F-value for the given degrees of freedom and a 1% level of significance. However, when F-values are in excess of 10 for a large sample like this, a table is not needed to know that the value is significant. (Study Session 3, LOS 10.g)

Question #38 of 106

Question ID: 485569

Upon further analysis, Turner concludes that multicollinearity is a problem. What might have prompted this further analysis and what is intuition behind the conclusion?

✗ **A)** **At least one of the *t*-statistics was not significant, the F-statistic was not significant, and a positive relationship between the number of analysts and the size of the firm would be expected.**

✓ **B)** At least one of the *t*-statistics was not significant, the F-statistic was significant, and a positive relationship between the number of analysts and the size of the firm would be expected.

✗ **C)** At least one of the *t*-statistics was not significant, the F-statistic was significant, and an intercept not significantly different from zero would be expected.

Explanation

Multicollinearity occurs when there is a high correlation among independent variables and may exist if there is a significant F-statistic for the fit of the regression model, but at least one insignificant independent variable when we expect all of them to be significant. In this case the coefficient on ln(market value) was not significant at the 1% level, but the F-statistic was significant. It would make sense that the size of the firm, i.e., the market value, and the number of analysts would be positively correlated. (Study Session 3, LOS 10.l)

Question #39 of 106

Question ID: 461755

Which of the following is NOT a model that has a qualitative dependent variable?

✗ **A)** **Discriminant analysis.**

✗ **B)** Logit.

✓ **C)** Event study.

Explanation

An event study is the estimation of the abnormal returns--generally associated with an informational event-that take on quantitative values.

Question #40 of 106

Question ID: 461700

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

✗ **A)** **Multicollinearity is a potential problem only in multiple regressions, not simple regressions.**

✓ **B)** Heteroskedasticity only occurs in cross-sectional regressions.

✗ **C)** The presence of heteroskedastic error terms results in a variance of the residuals that is too large.

Explanation

If there are shifting regimes in a time-series (e.g., change in regulation, economic environment), it is possible to have heteroskedasticity in a time-series.

Questions #41-46 of 106

John Rains, CFA, is a professor of finance at a large university located in the Eastern United States. He is actively involved with his local chapter of the Society of Financial Analysts. Recently, he was asked to teach one session of a Society-sponsored CFA review course, specifically teaching the class addressing the topic of quantitative analysis. Based upon his familiarity with the CFA exam, he decides that the first part of the session should be a review of the basic elements of quantitative analysis, such as hypothesis testing, regression and multiple regression analysis. He would like to devote the second half of the review session to the practical application of the topics he covered in the first half.

Rains decides to construct a sample regression analysis case study for his students in order to demonstrate a "real-life" application of the concepts. He begins by compiling financial information on a fictitious company called Big Rig, Inc. According to the case study, Big Rig is the primary producer of the equipment used in the exploration for and drilling of new oil and gas wells in the United States. Rains has based the information in the problem on an actual equity holding in his personal portfolio, but has simplified the data for the purposes of the review course.

Rains constructs a basic regression model for Big Rig in order to estimate its profitability (in millions), using two independent variables: the number of new wells drilled in the U.S. (WLS) and the number of new competitors (COMP) entering the market:

> Profits = $b_0 + b_1$WLS - b_2COMP + ε

Based on the model, the estimated regression equation is:

> Profits = 22.5 + 0.98(WLS) − 0.35(COMP)

Using the past 5 years of quarterly data, he calculated the following regression estimates for Big Rig, Inc:

	Coefficient	*Standard Error*
Intercept	22.5	2.465
WLS	0.98	0.683
COMP	0.35	0.186

Question #41 of 106

Question ID: 485676

Using the information presented, the t-statistic for the number of new competitors (COMP) coefficient is:

✗ **A) 1.435.**

✓ **B)** 1.882.

✗ **C)** 9.128.

Explanation

To test whether a coefficient is statistically significant, the null hypothesis is that the slope coefficient is zero. The t-statistic for the COMP coefficient is calculated as follows:

(0.35 - 0.0) / 0.186 = 1.882

(Study Session 3, LOS 9.g)

Question #42 of 106

Question ID: 485677

Rains asks his students to test the null hypothesis that states for every new well drilled, profits will be increased by the given multiple of the coefficient, all other factors remaining constant. The appropriate hypotheses for this two-tailed test can best be stated as:

✗ **A) H_0: b_1 = 0.35 versus H_a: b_1 ≠ 0.35.**

✓ **B)** H_0: b_1 = 0.98 versus H_a: b_1 ≠ 0.98.

✗ **C)** H_0: b_1 ≤ 0.98 versus H_a: b_1 > 0.98.

Explanation

The coefficient given in the above table for the number of new wells drilled (WLS) is 0.98. The hypothesis should test to see whether the coefficient is indeed equal to 0.98 or is equal to some other value. Note that hypotheses with the "greater than" or "less than" symbol are used with one-tailed tests. (Study Session 3, LOS 9.g)

Question #43 of 106

Question ID: 485678

Continuing with the analysis of Big Rig, Rains asks his students to calculate the mean squared error(MSE). Assume that the sum of squared errors (SSE) for the regression model is 359.

✗ **A)** **18.896.**

✗ **B)** 17.956.

✓ **C)** 21.118.

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 = 359 / (20 − 2 − 1) = 21.118. (Study Session 3, LOS 9.j)

Question #44 of 106

Question ID: 485679

Rains now wants to test the students' knowledge of the use of the *F*-test and the interpretation of the *F*-statistic. Which of the following statements regarding the *F*-test and the *F*-statistic is the *most* correct?

✗ **A)** **The *F*-test is usually formulated as a two-tailed test.**

✓ **B)** The *F*-statistic is used to test whether at least one independent variable in a set of independent variables explains a significant portion of the variation of the dependent variable.

✗ **C)** The *F*-statistic is almost always formulated to test each independent variable separately, in order to identify which variable is the most statistically significant.

Explanation

An *F*-test assesses how well a set of independent variables, as a group, explains the variation in the dependent variable. It tests all independent variables as a group, and is always a one-tailed test. The decision rule is to reject the null hypothesis if the calculated *F*-value is greater than the critical *F*-value. (Study Session 3, LOS 9.j)

Question #45 of 106

Question ID: 485680

One of the main assumptions of a multiple regression model is that the variance of the residuals is constant across all observations in the sample. A violation of the assumption is known as:

✓ **A)** **heteroskedasticity.**

✗ **B)** positive serial correlation.

✗ **C)** robust standard errors.

Explanation

Heteroskedasticity is present when the variance of the residuals is not the same across all observations in the sample, and there are sub-samples that are more spread out than the rest of the sample. (Study Session 3, LOS 10.k)

Question #46 of 106

Question ID: 485681

Rains reminds his students that a common condition that can distort the results of a regression analysis is referred to as serial correlation. The presence of serial correlation can be detected through the use of:

✗ **A) the Breusch-Pagen test.**

✗ **B)** the Hansen method.

✓ **C)** the Durbin-Watson statistic.

Explanation

The Durbin-Watson test ($DW \approx 2(1 - r)$) can detect serial correlation. Another commonly used method is to visually inspect a scatter plot of residuals over time. The Hansen method does not detect serial correlation, but can be used to remedy the situation. Note that the Breusch-Pagen test is used to detect heteroskedasticity. (Study Session 3, LOS 10.k)

Question #47 of 106

Question ID: 472463

Consider the following estimated regression equation, with standard errors of the coefficients as indicated:

> $Sales_i = 10.0 + 1.25\ R\&D_i + 1.0\ ADV_i - 2.0\ COMP_i + 8.0\ CAP_i$
>
> where the standard error for R&D is 0.45, the standard error for ADV is 2.2, the standard error for COMP 0.63, and the standard error for CAP is 2.5.

Sales are in millions of dollars. An analyst is given the following predictions on the independent variables: R&D = 5, ADV = 4, COMP = 10, and CAP = 40.

The predicted level of sales is *closest* to:

✗ **A) $310.25 million.**

✓ **B)** $320.25 million.

✗ **C)** $300.25 million.

Explanation

Predicted sales = $10 + 1.25 (5) + 1.0 (4) −2.0 (10) + 8 (40)

= 10 + 6.25 + 4 − 20 + 320 = $320.25

Question #48 of 106

Question ID: 461641

Consider the following analysis of variance table:

Source	*Sum of Squares*	*Df*	*Mean Square*
Regression	20	1	20
Error	80	20	4
Total	100	21	

The F-statistic for a test of the overall significance of the model is *closest* to:

✗ **A) 0.05**

✗ **B)** 0.20

✓ **C)** 5.00

Explanation

The F-statistic is equal to the ratio of the mean squared regression to the mean squared error.

F = MSR / MSE = 20 / 4 = 5.

Question #49 of 106

Question ID: 461752

An analyst is building a regression model which returns a qualitative dependant variable based on a probability distribution. This is *least likely* a:

✓ **A)** **discriminant model.**

✗ **B)** probit model.

✗ **C)** logit model.

Explanation

A probit model is a qualitative dependant variable which is based on a normal distribution. A logit model is a qualitative dependant variable which is based on the logistic distribution. A discriminant model returns a qualitative dependant variable based on a linear relationship that can be used for ranking or classification into discrete states.

Question #50 of 106

Question ID: 461609

Wanda Brunner, CFA, is trying to calculate a 98% 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

Which of the following are closest to the lower and upper bounds for variable CS?

✗ **A)** **0.274 to 0.367.**

✗ **B)** 0.267 to 0.374.

✓ **C)** 0.260 to 0.381.

Explanation

The critical t-value is 2.42 at the 98% confidence level (two tailed test). The estimated slope coefficient is 0.32 and the standard error is 0.025. The 98% confidence interval is 0.32 ± (2.42)(0.025) = 0.32 ± (0.061) = 0.260 to 0.381.

Question #51 of 106

Question ID: 461593

Consider the following estimated regression equation, with standard errors of the coefficients as indicated:

> $Sales_i = 10.0 + 1.25\ R\&D_i + 1.0\ ADV_i - 2.0\ COMP_i + 8.0\ CAP_i$
>
> where the standard error for R&D is 0.45, the standard error for ADV is 2.2, the standard error for COMP 0.63, and the standard error for CAP is 2.5.

The equation was estimated over 40 companies. Using a 5% level of significance, what are the hypotheses and the calculated test statistic to test whether the slope on R&D is different from 1.0?

✓ **A) H_0: $b_{R\&D}$ = 1 versus H_a: $b_{R\&D} \neq 1$; t = 0.556.**

✗ **B)** H_0: $b_{R\&D}$ = 1 versus H_a: $b_{R\&D} \neq 1$; t = 2.778.

✗ **C)** H_0: $b_{R\&D} \neq 1$ versus H_a: $b_{R\&D}$ = 1; t = 2.778.

Explanation

The test for "is different from 1.0" requires the use of the "1" in the hypotheses and requires 1 to be specified as the hypothesized value in the test statistic. The calculated *t*-statistic = (1.25-1)/.45 = 0.556

Questions #52-57 of 106

Quin Tan Liu, CFA is looking at the retail property sector for her manager. He is undertaking a top down review as she feels this is the best way to analyze the industry segment. To predict U.S property starts (housing), she has used regression analysis.

Liu included the following variables in his analysis:

- Average nominal interest rates during each year (as a decimal)
- Annual GDP per capita in $'000

Given these variables the following output was generated from 30 years of data:

Exhibit 1 - Results from regressing housing starts (in millions) on interest rates and GDP per capita

		Coefficient	Standard Error	T-statistic
Intercept		*0.42*		*3.1*
Interest rate		*− 1.0*		*− 2.0*
GDP per capita		0.03		0.7
ANOVA	**df**	**SS**	**MSS**	**F**
Regression	*2*	*3.896*	*1.948*	*21.644*
Residual	*27*	*2.431*	*0.090*	
Total	29	6.327		
Observations	*30*			
	1.22			

Durbin Watson

Exhibit 2 - Critical Values for Student's t-Distribution

Degrees of Freedom	**Area in Upper Tail**	
	5%	**2.5%**
26	1.706	2.056
27	1.703	2.052
28	1.701	2.048
29	1.699	2.045
30	1.697	2.040
31	1.696	2.040

Exhibit 3 - Critical Values for F-Distribution at 5% Level of Significance

Degrees of Freedom for the Denominator	Degrees of Freedom (df) for the Numerator		
	1	2	3
26	4.23	3.37	2.98
27	4.21	3.35	2.96
28	4.20	3.34	2.95
29	4.18	3.33	2.93
30	4.17	3.32	2.92
31	4.16	3.31	2.91
32	4.15	3.30	2.90

> The following variable estimates have been made for 20X7
> GDP per capita = $46,700
> Interest rate = 7%

Question #52 of 106

Question ID: 485613

Using the regression model represented in Exhibit 1, what is the predicted number of housing starts for 20X7?

✓ **A) 1,751,000**

✗ **B)** 1,394

✗ **C)** 1,394,420

Explanation

Housing starts = 0.42 − (1x 0.07) + (0.03 x 46.7) = 1.751 million

(Study Session 3, LOS 10.e)

Question #53 of 106

Question ID: 485614

The 90% confidence interval for the interest rate coefficient is:

- ✗ **A)** **−3.000 to +1.000**
- ✓ **B)** −1.852 to −0.149
- ✗ **C)** −1.850 to −0.151

Explanation

The general format for a confidence interval is:

estimated coefficient ± (critical t-stat x coefficient standard error)

The standard error of the interest rate coefficient must be 0.5 since its t-stat is −2.0 and this is derived from the estimated coefficient of −1.0 divided by its standard error.

The critical t-stat is taken from exhibit 2 with 5% in each tail and degrees of freedom = n − k − 1 = 30 − 2 − 1 = 27. This gives a value of 1.703. Hence the 90% confidence interval is:

> −1.0 ± (1.703 x 0.5) = −1.852 to −0.149

(Study Session 3, LOS 10.e)

Question #54 of 106

Question ID: 485615

Is the regression coefficient for the interest rate significantly different from zero at the 5% level of significance?

- ✓ **A)** **No, because │−2.0│< 2.052**
- ✗ **B)** No, because │−2.0│< 2.045
- ✗ **C)** Yes, because │−2.0│>1.703

Explanation

This requires a two-tailed test with 27 degrees of freedom (20−2−1) and 5 percent split between both tails. The critical t-stat is therefore 2.052.

H_0: b= 0
H_a: $b \neq 0$

Since the actual t-stat of −2.0 does not lie in the tail (it is too small) we cannot reject the null that the coefficient in the population is 0.

(Study Session 3, LOS 10.e)

Question #55 of 106

Question ID: 485616

Which of the following statements best describes the explanatory power of the estimated regression?

- ✓ **A)** **The independent variables explain 61.58% of the variation in housing starts.**
- ✗ **B)** The residual standard error of only 0.3 indicates that the regression equation is a good fit for the sample data
- ✗ **C)** The large F statistic indicates that both independent variables help explain changes in housing starts.

Explanation

The coefficient of determination is the statistic used to identify explanatory power. This can be calculated from the ANOVA table as 3.896/6.327 x 100 = 61.58%.

The residual standard error of 0.3 indicates that the standard deviation of the residuals is 0.3 million housing starts. Without knowledge of the data for the dependent variable it is not possible to assess whether this is a small or a large error.

The F statistic does not enable us to conclude on both independent variables. It only allows us the reject the hypothesis that all regression coefficients are zero and accept the hypothesis that at least one isn't.

(Study Session 3, LOS 10.g)

Question #56 of 106

Question ID: 485617

The estimated standard deviation of housing starts (in millions) is *closest* to:

✗ **A)** **0.3**

✗ **B)** 0.22

✓ **C)** 0.47

Explanation

Housing starts is the dependent variable.
Variance of dependent variable = SST/(n-1) = 6.327/29 = 0.22
Standard deviation = $(0.22)^{0.5}$ = 0.467

(Study Session 3, LOS 9.j)

Question #57 of 106

Question ID: 485618

Which of the following is the least appropriate statement in relation to R-square and adjusted R-square:

✓ **A)** **Adjusted R-square is a value between 0 and 1 and can be interpreted as a percentage**

✗ **B)** Adjusted R-square decreases when the added independent variable adds little value to the regression model

✗ **C)** R-square typically increases when new independent variables are added to the regression regardless of their explanatory power

Explanation

Adjusted R-square can be negative for a large number of independent variables that have no explanatory power. The other two statements are correct.

(Study Session 3, LOS 10.h)

Question #58 of 106

Question ID: 461698

An analyst is trying to estimate the beta for a fund. The analyst estimates a regression equation in which the fund returns are the dependent variable and the Wilshire 5000 is the independent variable, using monthly data over the past five years. The analyst finds that the correlation between the square of the residuals of the regression and the Wilshire 5000 is 0.2. Which of the following is *most* accurate, assuming a 0.05 level of significance? There is:

✗ **A)** **evidence of conditional heteroskedasticity but not serial correlation in the regression equation.**

✗ **B)** evidence of serial correlation but not conditional heteroskedasticity in the regression equation.

✓ **C)** no evidence that there is conditional heteroskedasticity or serial correlation in the regression equation.

Explanation

The test for conditional heteroskedasticity involves regressing the square of the residuals on the independent variables of the regression and creating a test statistic that is $n \times R^2$, where n is the number of observations and R^2 is from the squared-residual regression. The test statistic is distributed with a chi-squared distribution with the number of degrees of freedom equal to the number of independent variables. For a single variable, the R^2 will be equal to the square of the correlation; so in this case, the test statistic is $60 \times 0.2^2 = 2.4$, which is less than the chi-squared value (with one degree of freedom) of 3.84 for a p-value of 0.05. There is no indication about serial correlation.

Question #59 of 106

Question ID: 461699

An analyst is estimating whether a fund's excess return for a quarter is related to interest rates and last quarter's excess return. The model residuals exhibit unconditional heteroskedasticity. Residuals from an earlier model with only interest rates as independent variable exhibited serial correlation. Which of the following is *most* accurate? Parameter estimates for the regression model of excess returns on interest rates and prior quarter's excess returns will be:

✗ **A)** **accurate but statistical inference about the parameters will not be valid.**

✓ **B)** inaccurate and statistical inference about the parameters will not be valid.

✗ **C)** inaccurate but statistical inference about the parameters will be valid.

Explanation

Given that prior model without lagged dependent variable had residuals with serial correlation, including the lagged dependent variable as an independent variable indicates functional form of model misspecification, leading to inaccurate parameter estimates and inaccurate statistical inference. Unconditional heteroskedasticity never impacts statistical inference or parameter accuracy.

Question #60 of 106

Question ID: 461661

An analyst is trying to determine whether stock market returns are related to size and the market-to-book ratio, through the use of multiple regression. However, the analyst uses returns of portfolios of stocks instead of individual stocks in the regression. Which of the following is a valid reason why the analyst uses portfolios? The use of portfolios:

✗ **A)** **will remove the existence of multicollinearity from the data, reducing the likelihood of type II error.**

✓ **B)** reduces the standard deviation of the residual, which will increase the power of the test.

✗ **C)** will increase the power of the test by giving the test statistic more degrees of freedom.

Explanation

The use of portfolios reduces the standard deviation of the returns, which reduces the standard deviation of the residuals.

Question #61 of 106

Question ID: 461592

David Black wants to test whether the estimated beta in a market model is equal to one. He collected a sample of 60 monthly returns on a stock and estimated the regression of the stock's returns against those of the market. The estimated beta was 1.1, and the standard error of the coefficient is equal to 0.4. What should Black conclude regarding the beta if he uses a 5% level of significance? The null hypothesis that beta is:

✓ **A) equal to one cannot be rejected.**

✗ **B)** equal to one is rejected.

✗ **C)** not equal to one cannot be rejected.

Explanation

The calculated t-statistic is t = (1.1 − 1.0) / 0.4 = 0.25. The critical *t*-value for (60 − 2) = 58 degrees of freedom is approximately 2.0. Therefore, the null hypothesis that beta is equal to one cannot be rejected.

Question #62 of 106

Question ID: 461674

Suppose the analyst wants to add a dummy variable for whether a person has an undergraduate college degree and a graduate degree. What is the CORRECT representation if a person has both degrees?

	Undergraduate Degree Dummy Variable	Graduate Degree Dummy Variable
✓ **A)**	**1**	**1**
✗ **B)**	0	1
✗ **C)**	0	0

Explanation

Assigning a zero to both categories is appropriate for someone with neither degree. Assigning one to the undergraduate category and zero to the graduate category is appropriate for someone with only an undergraduate degree. Assigning zero to the undergraduate category and one to the graduate category is appropriate for someone with only a graduate degree. Assigning a one to both categories is correct since it reflects the possession of both degrees.

Questions #63-68 of 106

Vikas Rathod, an enrolled candidate for the CFA Level II examination, has decided to perform a calendar test to examine whether there is any abnormal return associated with investments and disinvestments made in blue-chip stocks on particular days of the week. As a proxy for blue-chips, he has decided to use the S&P 500 index. The analysis will involve the use of dummy variables and is based on the past 780 trading days. Here are selected findings of his study:

RSS	0.0039
SSE	0.9534
SST	0.9573
R-squared	0.004
SEE	0.035

Jessica Jones, CFA, a friend of Rathod, overhears that he is interested in regression analysis and warns him that whenever heteroskedasticity is present in multiple regression this could undermine the regression results. She mentions that one easy way to spot conditional heteroskedasticity is through a scatter plot, but she adds that there is a more formal test. Unfortunately, she can't quite remember its name. Jessica believes that heteroskedasticity can be rectified using White-corrected standard errors. Her son Jonathan who has also taken part in the discussion, hears this comment and argues that White correction would typically reduce the number of Type I errors in financial data.

Question #63 of 106

Question ID: 485648

How many dummy variables should Rathod use?

✗ **A)** **Five**

✓ **B)** Four

✗ **C)** Six

Explanation

There are 5 trading days in a week, but we should use (n − 1) or 4 dummies in order to ensure no violations of regression analysis occur.

(Study Session 3, LOS 10.j)

Question #64 of 106

Question ID: 485649

What is most likely represented by the intercept of the regression?

✗ **A)** **The intercept is not a driver of returns, only the independent variables**

✓ **B)** The return on a particular trading day

✗ **C)** The drift of a random walk

Explanation

The omitted variable is represented by the intercept. So, if we have four variables to represent Monday through Thursday, the intercept would represent returns on Friday.

(Study Session 3, LOS 10.j)

Question #65 of 106

Question ID: 485650

What can be said of the overall explanatory power of the model at the 5% significance?

- ✓ **A) There is no value to calendar trading**
- ✗ **B)** The coefficient of determination for the above regression is significantly higher than the standard error of the estimate, and therefore there is value to calendar trading
- ✗ **C)** There is value to calendar trading

Explanation

This question calls for a computation of the F-stat. F = (0.0039/4)/(0.9534/(780−4−1) = 0.79. The critical F is somewhere between 2.37 and 2.45 so we fail to reject the Null that all the coefficients are equal to zero.

(Study Session 3, LOS 10.g)

Question #66 of 106

Question ID: 485651

The test mentioned by Jessica is known as the:

- ✗ **A) Breusch-Pagan, which is a two-tailed test**
- ✗ **B)** Durbin-Watson, which is a two-tailed test
- ✓ **C)** Breusch-Pagan, which is a one-tailed test

Explanation

The Breusch-Pagan is used to detect conditional heteroskedasticity and it is a one-tailed test. This is because we are only concerned about large values in the residuals coefficient of determination.

(Study Session 3, LOS 10.k)

Question #67 of 106

Question ID: 485652

Are Jessica and her son Jonathan, correct in terms of the method used to correct for heteroskedasticity and the likely effects?

- ✓ **A) Both are correct**
- ✗ **B)** One is correct
- ✗ **C)** Neither is correct

Explanation

Jessica is correct. White-corrected standard errors are also known as robust standard errors. Jonathan is correct because White-corrected errors are higher than the biased errors leading to lower computed t-statistics and therefore less frequent rejection of the Null Hypothesis (remember incorrectly rejecting a true Null is Type I error).

(Study Session 3, LOS 10.k)

Question #68 of 106

Question ID: 485653

Assuming the a_1 term of an ARCH(1) model is significant, the following can be forecast:

- ✓ **A) The variance of the error term**
- ✗ **B)** The square of the error term
- ✗ **C)** A significant a1 implies that the ARCH framework cannot be used

Explanation

A Model is ARCH(1) if the coefficient a1 is significant. It will allow for the estimation of the variance of the error term.

(Study Session 3, LOS 11.m)

Question #69 of 106

Question ID: 461538

Which of the following statements *most* accurately interprets the following regression results at the given significance level?

Variable	*p*-value
Intercept	0.0201
X1	0.0284
X2	0.0310
X3	0.0143

✗ A) **The variable X2 is statistically significantly different from zero at the 3% significance level.**

✗ B) The variables X1 and X2 are statistically significantly different from zero at the 2% significance level.

✓ C) The variable X3 is statistically significantly different from zero at the 2% significance level.

Explanation

The *p*-value is the smallest level of significance for which the null hypothesis can be rejected. An independent variable is significant if the *p*-value is less than the stated significance level. In this example, X3 is the variable that has a *p*-value less than the stated significance level.

Question #70 of 106

Question ID: 461658

Wilson estimated a regression that produced the following analysis of variance (ANOVA) table:

Source	*Sum of squares*	*Degrees of freedom*	*Mean square*
Regression	100	1	100.0
Error	300	40	7.5
Total	400	41	

The values of R^2 and the F-statistic for the fit of the model are:

✓ A) **R^2 = 0.25 and F = 13.333.**

✗ B) R^2 = 0.25 and F = 0.930.

✗ C) R^2 = 0.20 and F = 13.333.

Explanation

R^2 = RSS / SST = 100 / 400 = 0.25

The F-statistic is equal to the ratio of the mean squared regression to the mean squared error.

F = 100 / 7.5 = 13.333

Questions #71-76 of 106

Damon Washburn, CFA, is currently enrolled as a part-time graduate student at State University. One of his recent assignments for his course on Quantitative Analysis is to perform a regression analysis utilizing the concepts covered during the semester. He must interpret the results of the regression as well as the test statistics. Washburn is confident in his ability to calculate the statistics because the class is allowed to use statistical software. However, he realizes that the interpretation of the statistics will be the true test of his knowledge of regression analysis. His professor has given to the students a list of questions that must be answered by the results of the analysis.

Washburn has estimated a regression equation in which 160 quarterly returns on the S&P 500 are explained by three macroeconomic variables: employment growth (EMP) as measured by nonfarm payrolls, gross domestic product (GDP) growth, and private investment (INV). The results of the regression analysis are as follows:

Coefficient Estimates		
Parameter	*Coefficient*	*Standard Error of Coefficient*
Intercept	9.50	3.40
EMP	-4.50	1.25
GDP	4.20	0.76
INV	-0.30	0.16

Other Data:

- Regression sum of squares (RSS) = 126.00
- Sum of squared errors (SSE) = 267.00
- Durbin-Watson statistic (DW) = 1.34

Abbreviated Table of the Student's t-distribution (One-Tailed Probabilities)					
df	*p = 0.10*	*p = 0.05*	*p = 0.025*	*p = 0.01*	*p = 0.005*
3	1.638	2.353	3.182	4.541	5.841
10	1.372	1.812	2.228	2.764	3.169
50	1.299	1.676	2.009	2.403	2.678
100	1.290	1.660	1.984	2.364	2.626
120	1.289	1.658	1.980	2.358	2.617
200	1.286	1.653	1.972	2.345	2.601

Critical Values for Durbin-Watson Statistic (α = 0.05)										
	K=1		K=2		K=3		K=4		K=5	
n	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u

20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
>100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Question #71 of 106

Question ID: 485571

How many of the three independent variables (not including the intercept term) are statistically significant in explaining quarterly stock returns at the 5.0% level?

✗ **A) One of the three is statistically significant.**

✓ **B)** Two of the three are statistically significant.

✗ **C)** All three are statistically significant.

Explanation

To determine whether the independent variables are statistically significant, we use the student's t-statistic, where t equals the coefficient estimate divided by the standard error of the coefficient. This is a two-tailed test. The critical value for a 5.0% significance level and 156 degrees of freedom (160-3-1) is about 1.980, according to the table.

The t-statistic for employment growth = -4.50/1.25 = -3.60.

The t-statistic for GDP growth = 4.20/0.76 = 5.53.

The t-statistic for investment growth = -0.30/0.16 = -1.88.

Therefore, employment growth and GDP growth are statistically significant because the absolute values of their t-statistics are larger than the critical value, which means two of the three independent variables are statistically significantly different from zero. (Study Session 3, LOS 10.a)

Question #72 of 106

Question ID: 485572

Can the null hypothesis that the GDP growth coefficient is equal to 3.50 be rejected at the 1.0% confidence level versus the alternative that it is not equal to 3.50? The null hypothesis is:

✗ **A) accepted because the t-statistic is less than 2.617.**

✓ **B)** not rejected because the t-statistic is equal to 0.92.

✗ **C)** rejected because the t-statistic is less than 2.617.

Explanation

The hypothesis is:

> H_0: $b_{GDP} = 3.50$
>
> H_a: $b_{GDP} \neq 3.50$

This is a two-tailed test. The critical value for the 1.0% significance level and 156 degrees of freedom (160 − 3 − 1) is about 2.617. The t-statistic is (4.20 − 3.50)/0.76 = 0.92. Because the t-statistic is less than the critical value, we cannot reject the null hypothesis. Notice we cannot say that the null hypothesis is accepted; only that it is not rejected. (Study Session 3, LOS 10.c)

Question #73 of 106

Question ID: 485573

The percentage of the total variation in quarterly stock returns explained by the independent variables is *closest* to:

- ✓ **A) 32%.**
- ✗ **B)** 42%.
- ✗ **C)** 47%.

Explanation

The R^2 is the percentage of variation in the dependent variable explained by the independent variables. The R^2 is equal to the $SS_{Regession}/SS_{Total}$, where the SS_{Total} is equal to $SS_{Regression} + SS_{Error}$. R^2 = 126.00/(126.00+267.00) = 32%. (Study Session 3, LOS 10.h)

Question #74 of 106

Question ID: 485574

According to the Durbin-Watson statistic, there is:

- ✗ **A) no significant positive serial correlation in the residuals.**
- ✓ **B)** significant positive serial correlation in the residuals.
- ✗ **C)** significant heteroskedasticity in the residuals.

Explanation

The Durbin-Watson statistic tests for serial correlation in the residuals. According to the table, d_l = 1.61 and d_u = 1.74 for three independent variables and 160 degrees of freedom. Because the DW (1.34) is less than the lower value (1.61), the null hypothesis of no significant positive serial correlation can be rejected. This means there is a problem with serial correlation in the regression, which affects the interpretation of the results. (Study Session 3, LOS 10.k)

Question #75 of 106

Question ID: 485575

What is the predicted quarterly stock return, given the following forecasts?

- Employment growth = 2.0%
- GDP growth = 1.0%
- Private investment growth = -1.0%

- ✗ **A) 4.7%.**
- ✓ **B)** 5.0%.
- ✗ **C)** 4.4%.

Explanation

Predicted quarterly stock return is 9.50% + (−4.50)(2.0%) + (4.20)(1.0%) + (−0.30)(−1.0%) = 5.0%. (Study Session 3, LOS 10.e)

Question #76 of 106

Question ID: 485576

What is the standard error of the estimate?

- ✗ **A) 0.81.**
- ✓ **B)** 1.31.
- ✗ **C)** 1.71.

Explanation

The standard error of the estimate is equal to $[SSE/(n - k - 1)]^{1/2} = [267.00/156]^{1/2}$ = approximately 1.31. (Study Session 3, LOS 9.j)

Question #77 of 106

Question ID: 461741

An analyst is testing to see whether a dependent variable is related to three independent variables. He finds that two of the independent variables are correlated with each other, but that the correlation is spurious. Which of the following is *most* accurate? There is:

✗ **A) no evidence of multicollinearity and serial correlation.**

✓ **B)** evidence of multicollinearity but not serial correlation.

✗ **C)** evidence of multicollinearity and serial correlation.

Explanation

Just because the correlation is spurious, does not mean the problem of multicollinearity will go away. However, there is no evidence of serial correlation.

Question #78 of 106

Question ID: 461701

Which of the following is *least likely* a method used to detect heteroskedasticity?

✗ **A) Test of the variances.**

✗ **B)** Breusch-Pagan test.

✓ **C)** Durbin-Watson test.

Explanation

The Durbin-Watson test is used to detect serial correlation. The Breusch-Pagan test is used to detect heteroskedasticity.

Question #79 of 106

Question ID: 461594

63 monthly stock returns for a fund between 1997 and 2002 are regressed against the market return, measured by the Wilshire 5000, and two dummy variables. The fund changed managers on January 2, 2000. Dummy variable one is equal to 1 if the return is from a month between 2000 and 2002. Dummy variable number two is equal to 1 if the return is from the second half of the year. There are 36 observations when dummy variable one equals 0, half of which are when dummy variable two also equals 0. The following are the estimated coefficient values and standard errors of the coefficients.

Coefficient	*Value*	*Standard error*
Market	1.43000	0.319000
Dummy 1	0.00162	0.000675
Dummy 2	0.00132	0.000733

What is the *p*-value for a test of the hypothesis that performance in the second half of the year is different than performance in the first half of the year?

- ✗ **A) Lower than 0.01.**
- ✓ **B)** Between 0.05 and 0.10.
- ✗ **C)** Between 0.01 and 0.05.

Explanation

The difference between performance in the second and first half of the year is measured by dummy variable 2. The *t*-statistic is equal to 0.00132 / 0.000733 = 1.800, which is between the *t*-values (with 63 − 3 − 1 = 59 degrees of freedom) of 1.671 for a *p*-value of 0.10, and 2.00 for a *p*-value of 0.05 (note that the test is a two-sided test).

Question #80 of 106

Question ID: 461746

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

- ✗ **A) conditional heteroskedasticity.**
- ✓ **B)** multicollinearity.
- ✗ **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.

Question #81 of 106

Question ID: 461757

An analyst has run several regressions hoping to predict stock returns, and wants to translate this into an economic interpretation for his clients.

> Return = 0.03 + 0.020Beta - 0.0001MarketCap (in billions) + ε

A correct interpretation of the regression *most likely* includes:

- ✗ **A) prediction errors are always on the positive side.**
- ✓ **B)** a billion dollar increase in market capitalization will drive returns down by 0.01%.
- ✗ **C)** a stock with zero beta and zero market capitalization will return precisely 3.0%.

Explanation

The coefficient of MarketCap is 0.01%, indicating that larger companies have slightly smaller returns. Note that a company with no market capitalization would not be expected to have a return at all. Error terms are typically assumed to be normally distributed with a mean of zero.

Question #82 of 106

Question ID: 461607

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. The predicted value of AUTO if PI is 4, TEEN is 0.30, and INS = 0.6 is *closest* to:

✓ **A) 14.10.**

✗ **B)** 14.90.

✗ **C)** 17.50.

Explanation

Predicted AUTO = 10 + 1.25 (4) + 1.0 (0.30) - 2.0 (0.6)

= 10 + 5 + 0.3 - 1.2

= 14.10

Question #83 of 106

Question ID: 461645

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 2.5% significance and 5% significance.**

✗ **B)** Rejected at 5% significance only.

✗ **C)** Not rejected at 2.5% or 5.0% 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.

Questions #84-89 of 106

Lynn Carter, CFA, is an analyst in the research department for Smith Brothers in New York. She follows several industries, as well as the top companies in each industry. She provides research materials for both the equity traders for Smith Brothers as well as their retail customers. She routinely performs regression analysis on those companies that she follows to identify any emerging trends that could affect investment decisions.

Due to recent layoffs at the company, there has been some consolidation in the research department. Two research analysts have been laid off, and their workload will now be distributed among the remaining four analysts. In addition to her current workload, Carter will now be responsible for providing research on the airline industry. Pinnacle Airlines, a leader in the industry, represents a large holding in Smith Brothers' portfolio. Looking back over past research on Pinnacle, Carter recognizes that the company historically has been a strong performer in what is considered to be a very competitive industry. The stock price over the last 52-week period has outperformed that of other industry leaders, although Pinnacle's net income has remained flat. Carter wonders if the stock price of Pinnacle has become overvalued relative to its peer group in the market, and wants to determine if the timing is right for Smith Brothers to decrease its position in Pinnacle.

Carter decides to run a regression analysis, using the monthly returns of Pinnacle stock as the dependent variable and monthly returns of the airlines industry as the independent variable.

Analysis of Variance Table (ANOVA)			
Source	*df* (*Degrees of Freedom*)	*SS* (*Sum of Squares*)	*Mean Square* (*SS/df*)
Regression	1	3,257 (RSS)	3,257 (MSR)
Error	8	298 (SSE)	37.25 (MSE)
Total	9	3,555 (SS Total)	

Question #84 of 106

Question ID: 485641

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

✗ **A)** **The variance of the residuals is constant.**

✗ **B)** A linear relationship exists between the dependent and independent variables.

✓ **C)** The independent variable is correlated with the residuals.

Explanation

Although the linear regression model is fairly insensitive to minor deviations from any of these assumptions, the independent variable is typically uncorrelated with the residuals. (Study Session 3, LOS 9.d)

Question #85 of 106

Question ID: 485642

Carter wants to test the strength of the relationship between the two variables. She calculates a correlation coefficient of 0.72. This means that the two variables:

✗ **A)** **have a positive but non-linear relationship.**

✓ **B)** have a positive linear association.

✗ **C)** have no relationship.

Explanation

If the correlation coefficient (r) is greater that 0 and less than 1, then the two variables are said to be positively correlated. Positive correlation coefficient indicates a positive linear association between the two variables.(Study Session 3, LOS 9.a)

Question #86 of 106

Question ID: 485643

Based upon the information presented in the ANOVA table, what is the standard error of the estimate?

✗ **A)** **37.25.**

✓ **B)** 6.10.

✗ **C)** 57.07.

Explanation

The standard error of the estimate (SEE) measures the "fit" of the regression line, and the smaller the standard error, the better the fit. The SSE can be calculated as $\sqrt{(MSE} = \sqrt{37.25} = 6.10$ (Study Session 3, LOS 10.i)

Question #87 of 106

Question ID: 485644

Based upon the information presented in the ANOVA table, what is the coefficient of determination?

✓ **A)** **0.916, indicating that the variability of industry returns explains about 91.6% of the variability of company returns.**

✗ **B)** 0.084, indicating that the variability of industry returns explains about 8.4% of the variability of company returns.

✗ **C)** 0.839, indicating that company returns explain about 83.9% of the variability of industry returns.

Explanation

The coefficient of determination (R^2) is the percentage of the total variation in the dependent variable explained by the independent variable.

The R^2 = (RSS / SS) Total = (3,257 / 3,555) = 0.916. This means that the variation of independent variable (the airline industry) explains 91.6% of the variations in the dependent variable (Pinnacle stock). (Study Session 3, LOS 10.i)

Question #88 of 106

Question ID: 485645

Based upon her analysis, Carter has derived the following regression equation: $\hat{Y} = 1.75 + 3.25X_1$. The predicted value of the Y variable equals 50.50, if the:

✗ **A)** **coefficient of the determination equals 15.**

✗ **B)** predicted value of the dependent variable equals 15.

✓ **C)** predicted value of the independent variable equals 15.

Explanation

Note that the easiest way to answer this question is to plug numbers into the equation.

The predicted value for Y = 1.75 + 3.25(15) = 50.50.

The variable X_1 represents the independent variable. (Study Session 3, LOS 11.a)

Question #89 of 106

Question ID: 485646

Carter realizes that although regression analysis is a useful tool when analyzing investments, there are certain limitations. Carter made a list of points describing limitations that Smith Brothers equity traders should be aware of when applying her research to their investment decisions.

- Point 1: Data derived from regression analysis may be homoskedastic.
- Point 2: Data from regression relationships tends to exhibit parameter instability.
- Point 3: Results of regression analysis may exhibit autocorrelation.
- Point 4: The variance of the error term may change over time.

When reviewing Carter's list, one of the Smith Brothers' equity traders points out that not all of the points describe regression analysis limitations. Which of Carter's points *most* accurately describes the limitations to regression analysis?

✗ **A) Points 1, 3, and 4.**

✓ **B)** Points 2, 3, and 4.

✗ **C)** Points 1, 2, and 3.

Explanation

One of the basic assumptions of regression analysis is that the variance of the error terms is constant, or homoskedastic. Any violation of this assumption is called heteroskedasticity. Therefore, Point 1 is incorrect, but Point 4 is correct. Points 2 and 3 also describe limitations of regression analysis. (Study Session 3, LOS 9.k)

Question #90 of 106

Question ID: 461528

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. Which of the follow regression equations *correctly* represents Hilton's hypothesis?

✗ **A) SALES = α x β_1 POP x β_2 INCOME x β_3 ADV x ε.**

✗ **B)** INCOME = α + β_1 POP + β_2 SALES + β_3 ADV + ε.

✓ **C)** SALES = α + β_1 POP + β_2 INCOME + β_3 ADV + ε.

Explanation

SALES is the dependent variable. POP, INCOME, and ADV should be the independent variables (on the right hand side) of the equation (in any order). Regression equations are additive.

Questions #91-96 of 106

Raul Gloucester, CFA, is analyzing the returns of a fund that his company offers. He tests the fund's sensitivity to a small capitalization index and a large capitalization index, as well as to whether the January effect plays a role in the fund's performance. He uses two years of monthly returns data, and runs a regression of the fund's return on the indexes and a January-effect qualitative variable. The "January" variable is 1 for the month of January and zero for all other months. The results of the regression are shown in the tables below.

Regression Statistics	

Multiple R	0.817088
R^2	0.667632
Adjusted R^2	0.617777
Standard Error	1.655891
Observations	24

ANOVA			
	df	*SS*	*MS*
Regression	3	110.1568	36.71895
Residual	20	54.8395	2.741975
Total	23	164.9963	

	Coefficients	*Standard Error*	*t-Statistic*
Intercept	-0.23821	0.388717	-0.61282
January	2.560552	1.232634	2.077301
Small Cap Index	0.231349	0.123007	1.880778
Large Cap Index	0.951515	0.254528	3.738359

Gloucester will perform an *F*-test for the equation. He also plans to test for serial correlation and conditional and unconditional heteroskedasticity.

Jason Brown, CFA, is interested in Gloucester's results. He speculates that they are economically significant in that excess returns could be earned by shorting the large capitalization and the small capitalization indexes in the month of January and using the proceeds to buy the fund.

Question #91 of 106

Question ID: 461678

The percent of the variation in the fund's return that is explained by the regression is:

✗ A) **61.78%.**

✗ B) 81.71%.

✓ C) 66.76%.

Explanation

The R^2 tells us how much of the change in the dependent variable is explained by the changes in the independent variables in the regression: 0.667632.

Question #92 of 106

Question ID: 461679

In a two-tailed test at a five percent level of significance, the coefficients that are significant are:

✓ **A)** **the large cap index only.**

✗ **B)** the January effect and the small capitalization index only.

✗ **C)** the January effect and the large capitalization index only.

Explanation

For a two-tailed test with 20 = 24 - 3 - 1 degrees of freedom and a five percent level of significance, the critical *t*-statistic is 2.086. Only the coefficient for the large capitalization index has a *t*-statistic larger than this.

Question #93 of 106

Question ID: 461680

Which of the following *best* summarizes the results of an F-test (5 percent significance) for the regression? The F-statistic is:

✗ **A)** **9.05 and the critical value is 3.86.**

✓ **B)** 13.39 and the critical value is 3.10.

✗ **C)** 13.39 and the critical value is 3.86.

Explanation

The F-statistic is the ratio of the Mean Square of the Regression divided by the Mean Square Error (Residual): 13.39 = 36.718946 / 2.74197510. The F-statistic has 3 and 20 degrees of freedom, so the critical value, at a 5 percent level of significance = 3.10.

Question #94 of 106

Question ID: 461681

The *best* test for unconditional heteroskedasticity is:

✓ **A)** **neither the Durbin-Watson test nor the Breusch-Pagan test.**

✗ **B)** the Breusch-Pagan test only.

✗ **C)** the Durbin-Watson test only.

Explanation

The Durbin-Watson test is for serial correlation. The Breusch-Pagan test is for *conditional* heteroskedasticity; it tests to see if the size of the independent variables influences the size of the residuals. Although tests for unconditional heteroskedasticity exist, they are not part of the CFA curriculum, and unconditional heteroskedasticity is generally considered less serious than conditional heteroskedasticity.

Question #95 of 106

Question ID: 461682

In the month of January, if both the small and large capitalization index have a zero return, we would expect the fund to have a return equal to:

✗ **A)** **2.799.**

✗ **B)** 2.561.

✓ **C)** 2.322.

Explanation

The forecast of the return of the fund would be the intercept plus the coefficient on the January effect: 2.322 = -0.238214 + 2.560552.

Question #96 of 106

Question ID: 461683

Assuming (for this question only) that the F-test was significant but that the t-tests of the independent variables were insignificant, this would *most likely* suggest:

✓ **A) multicollinearity.**

✗ **B)** serial correlation.

✗ **C)** conditional heteroskedasticity.

Explanation

When the F-test and the t-tests conflict, multicollinearity is indicated.

Question #97 of 106

Question ID: 461656

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) should have a higher coefficient on the independent variable.**

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

✗ **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.

Question #98 of 106

Question ID: 461525

Consider the following estimated regression equation, with the standard errors of the slope coefficients as noted:

> $Sales_i = 10.0 + 1.25\ R\&D_i + 1.0\ ADV_i - 2.0\ COMP_i + 8.0\ CAP_i$
>
> where the standard error for the estimated coefficient on R&D is 0.45, the standard error for the estimated coefficient on ADV is 2.2 , the standard error for the estimated coefficient on COMP is 0.63, and the standard error for the estimated coefficient on CAP is 2.5.

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

✗ **A) R&D, ADV, COMP, and CAP.**

✗ **B)** ADV and CAP only.

✓ **C)** R&D, COMP, and CAP only.

Explanation

The critical t-values for 40-4-1 = 35 degrees of freedom and a 5% level of significance are ± 2.03.

The calculated t-values are:
t for R&D = 1.25 / 0.45 = 2.777
t for ADV = 1.0/ 2.2 = 0.455
t for COMP = -2.0 / 0.63 = -3.175
t for CAP = 8.0 / 2.5 = 3.2
Therefore, R&D, COMP, and CAP are statistically significant.

Question #99 of 106

Question ID: 461644

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

What is the p-value for the test of the hypothesis that all three of the coefficients are equal to zero?

✗ **A) Between 0.05 and 0.10.**
✓ **B)** lower than 0.025.
✗ **C)** Between 0.025 and 0.05.

Explanation

This test requires an F-statistic, which is equal to the ratio of the mean regression sum of squares to the mean squared error.

The mean regression sum of squares is the regression sum of squares divided by the number of independent variables, which is 119.25 / 3 = 39.75.

The residual sum of squares is the difference between the total sum of squares and the regression sum of squares, which is 294.45 − 119.25 = 175.20. The denominator degrees of freedom is the number of observations minus the number of independent variables, minus 1, which is 25 − 3 − 1 = 21. The mean squared error is the residual sum of squares divided by the denominator degrees of freedom, which is 175.20 / 21 = 8.34.

The F-statistic is 39.75 / 8.34 = 4.76, which is higher than the F-value (with 3 numerator degrees of freedom and 21 denominator degrees of freedom) of 3.07 at the 5% level of significance and higher than the F-value of 3.82 at the 2.5% level of significance. The conclusion is that the p-value must be lower than 0.025.

Remember the p-value is the probability that lies above the computed test statistic for upper tail tests or below the computed test statistic for lower tail tests.

Question #100 of 106

Question ID: 461675

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.

Question #101 of 106

Question ID: 461670

The management of a large restaurant chain believes that revenue growth is dependent upon the month of the year. Using a standard 12 month calendar, how many dummy variables must be used in a regression model that will test whether revenue growth differs by month?

- ✗ **A) 12.**
- ✗ **B)** 13.
- ✓ **C)** 11.

Explanation

The appropriate number of dummy variables is one less than the number of categories because the intercept captures the effect of the other effect. With 12 categories (months) the appropriate number of dummy variables is 11 = 12 - 1. If the number of dummy variables equals the number of categories, it is possible to state any one of the independent dummy variables in terms of the others. This is a violation of the assumption of the multiple linear regression model that none of the independent variables are linearly related.

Question #102 of 106

Question ID: 461708

Alex Wade, CFA, is analyzing the result of a regression analysis comparing the performance of gold stocks versus a broad equity market index. Wade believes that serial correlation may be present, and in order to prove his theory, should use which of the following methods to detect its presence?

- ✗ **A) The Hansen method.**
- ✗ **B)** The Breusch-Pagan test.
- ✓ **C)** The Durbin-Watson statistic.

Explanation

The Durbin-Watson statistic is the most commonly used method for the detection of serial correlation, although residual plots can also be utilized. For a large sample size, $DW \approx 2(1-r)$, where r is the correlation coefficient between residuals from one period and those from a previous period. The DW statistic is then compared to a table of DW statistics that gives upper and lower critical values for various sample sizes, levels of significance and numbers of degrees of freedom to detect the presence or absence of serial correlation.

Question #103 of 106

Question ID: 461540

Jacob Warner, CFA, is evaluating a regression analysis recently published in a trade journal that hypothesizes that the annual performance of the S&P 500 stock index can be explained by movements in the Federal Funds rate and the U.S. Producer Price Index (PPI). Which of the following statements regarding his analysis is *most* accurate?

✗ **A)** **If the *t*-value of a variable is less than the significance level, the null hypothesis cannot be rejected.**

✗ **B)** If the *p*-value of a variable is less than the significance level, the null hypothesis cannot be rejected.

✓ **C)** If the *p*-value of a variable is less than the significance level, the null hypothesis can be rejected.

Explanation

The *p*-value is the smallest level of significance for which the null hypothesis can be rejected. Therefore, for any given variable, if the *p*-value of a variable is less than the significance level, the null hypothesis can be rejected and the variable is considered to be statistically significant.

Question #104 of 106

Question ID: 461653

An analyst is estimating a regression equation with three independent variables, and calculates the R^2, the adjusted R^2, and the F-statistic. The analyst then decides to add a fourth variable to the equation. Which of the following is *most* accurate?

✗ **A)** **The R^2 and F-statistic will be higher, but the adjusted R^2 could be higher or lower.**

✗ **B)** The adjusted R^2 will be higher, but the R^2 and F-statistic could be higher or lower.

✓ **C)** The R^2 will be higher, but the adjusted R^2 and F-statistic could be higher or lower.

Explanation

The R^2 will always increase as the number of variables increase. The adjusted R^2 specifically adjusts for the number of variables, and might not increase as the number of variables rise. As the number of variables increases, the regression sum of squares will rise and the residual T sum of squares will fall-this will tend to make the F-statistic larger. However, the number degrees of freedom will also rise, and the denominator degrees of freedom will fall, which will tend to make the F-statistic smaller. Consequently, like the adjusted R^2, the F-statistic could be higher or lower.

Question #105 of 106

Question ID: 461744

Which of the following is a potential remedy for multicollinearity?

✓ **A)** **Omit one or more of the collinear variables.**

✗ **B)** Add dummy variables to the regression.

✗ **C)** Take first differences of the dependent variable.

Explanation

The first differencing is not a remedy for the collinearity, nor is the inclusion of dummy variables. The best potential remedy is to attempt to eliminate highly correlated variables.

Question #106 of 106

Question ID: 461742

A variable is regressed against three other variables, x, y, and z. Which of the following would NOT be an indication of multicollinearity? X is closely related to:

✗ **A)** **3.**

✓ **B)** y^2.

✗ **C)** 3y + 2z.

Explanation

If x is related to y^2, the relationship between x and y is not linear, so multicollinearity does not exist. If x is equal to a constant (3), it will be correlated with the intercept term.