Question #1 of 191

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

A) Multicollinearity may be a problem even if the multicollinearity is not perfect.

B) Multicollinearity may be present in any regression model.

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

Explanation

Multicollinearity is not an issue in simple linear regression.

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

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 191

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

C) "% Wholesale Trade Employment" and "% Retail Trade" 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, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #3 of 191

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

A) 0.3595.

B) 2.2675.

C) 0.0818.

Explanation

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

$$s_{\hat{b}_j}$$

= slope coefficient/the t-statistic = 0.1715/2.096 = 0.0818.

(Study Session 3, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #4 of 191

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

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

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

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, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #5 of 191

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) 5.54%.

B) 3.15%.

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, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #6 of 191

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

A) –0.0740 to 0.4170.

B) 0.0897 to 0.2533.

C) 0.0111 to 0.3319.

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:

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

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, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #7 of 191

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

B) The variance of the observations has increased over time.

C) The Durbin–Watson statistic differs sufficiently from 2.

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, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #8 of 191

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

A) one dummy variables.

B) three dummy variables.

C) four dummy variables.

Explanation

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

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #9 of 191

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

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

B) F = MSR/MSE.

C) $df_{numerator}$ = k and $df_{denominator}$ = n − k − 1.

Explanation

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

(Study Session 3, Module 8.3, LOS 8.g)

Related Material

SchweserNotes - Book 1

Question #10 of 191

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.

(Study Session 3, Module 8.2, LOS 8.c)

Related Material

SchweserNotes - Book 1

Question #11 of 191

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. The model residuals exhibit unconditional heteroskedasticity and serial correlation due to inclusion of lagged dependent variable. 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 but statistical inference about the parameters will be valid.

C) inaccurate and statistical inference about the parameters will not be valid.

Explanation

Given that the residuals exhibit serial correlation due to inclusion of the lagged dependent variable as an independent variable, we have a functional form of model misspecification, which leads to inaccurate parameter estimates and inaccurate statistical inference. Unconditional heteroskedasticity never impacts statistical inference or parameter accuracy.

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #12 of 191

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

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

What are the lower and upper bounds for variable DR?

A) 0.488 to 0.552.

B) 0.481 to 0.559.

C) 0.474 to 0.566.

Explanation

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

(Study Session 3, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

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

$$S = 1.76 + 0.23H - 0.08M$$

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

Variable Descriptions

S = HB Sales (in thousands)

H = housing starts (in thousands)

M = mortgage interest rate (in percent)

November 20x6 Actual Data

HB's monthly sales: $55,000

Housing starts: 150,000

Mortgage interest rate (%): 7.5

Critical Values for Student's t-Distribution

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

Question #13 of 191

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

A) $55,000

B) $44,000

C) $36,000

Explanation

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #14 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #15 of 191

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

A) yes, because 2.6 > 2.23 ✗

B) no, because 2.6 < 2.62 ✗

C) yes, because 2.6 > 1.98 ✓

Explanation

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #16 of 191

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

A) the joint significance of the independent variables ✓

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

C) degree of correlation between the independent variables ✗

Explanation

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #17 of 191

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

A) 9.80 ✗

B) 77.00 ✓

C) 67.00 ✗

Explanation

The question is asking for the coefficient of determination.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #18 of 191

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

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

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

C) standard errors are not affected but coefficient estimate is inconsistent. ✗

Explanation

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #19 of 191

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

A) conditional heteroskedasticity.

B) serial correlation.

C) multicollinearity.

Explanation

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

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

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

how current conditions explain how generator sales are higher or lower from the historical average in each area. In summary, she regresses current sales for each area minus its respective historical average on the following variables for each area.

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

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

Coefficient Estimates Table		
Variable	**Estimated Coefficient**	**Standard Error of the Coefficient**
Intercept	5.00	1.850
$ Heating Oil	2.00	0.827
Low Temperature	3.00	1.200
Snowfall	10.00	4.833
Housing Starts	5.00	2.333

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

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

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

Question #20 of 191

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

A) $65 million above the average.

B) $35.2 million above the average.

C) $55 million above average.

Explanation

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

$$[5 + (2 \times 0.10) + (3 \times 5) + (10 \times 3) + (5 \times (-3))] \times \$1{,}000{,}000 = \$35.2 \text{ million}.$$

(Study Session 3, Module 8.3, LOS 8.g)

Related Material

SchweserNotes - Book 1

Question #21 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.3, LOS 8.g)

Related Material

SchweserNotes - Book 1

Question #22 of 191

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

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

B) both a Durbin-Watson test and a Breusch-Pagan test.

C) a Breusch-Pagan test, but not a Durbin-Watson test.

Explanation

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

(Study Session 3, Module 8.3, LOS 8.g)

Related Material

SchweserNotes - Book 1

Question #23 of 191

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

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

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

C) all of the variables are statistically significant in explaining sales.

Explanation

The calculated *t*–statistics are:

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

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

(Study Session 3, Module 8.3, LOS 8.g)

Related Material

SchweserNotes - Book 1

Question #24 of 191

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

A) adjusted R^2 value.

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

C) unadjusted R^2 value.

Explanation

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

(Study Session 3, Module 8.3, LOS 8.g)

Related Material

SchweserNotes - Book 1

Question #25 of 191

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

A) The disturbance or error term is normally distributed.

B) The residuals are homoscedastic.

C) There is a linear relationship between the independent variables.

Explanation

Multiple regression models assume that there is no linear relationship between two or more of the independent variables. The other answer choices are both assumptions of multiple regression.

(Study Session 3, Module 8.3, LOS 8.g)

Related Material

SchweserNotes - Book 1

Question #26 of 191

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

A) *p*-value.

B) Slope coefficient.

C) Intercept term.

Explanation

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #27 of 191

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.

(Study Session 3, Module 8, LOS 8.i)

Related Material

SchweserNotes - Book 1

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

Smith is confident with the results presented in his report. He has already done some hypothesis testing for statistical significance, including calculating a t-statistic and conducting a two-tailed test where the null hypothesis is that the regression coefficient is equal to zero

versus the alternative that it is not. He feels that he has done a thorough job on the report and is ready to answer any questions posed by his boss.

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

Question #28 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #29 of 191

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

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

B) the variance of the error term is correlated with the values of the independent variables.

C) the error terms are correlated with each other.

Explanation

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

Multicollinearity, on the other hand, occurs when two or more of the independent variables are highly correlated with each other. Serial correlation exists when the error terms are correlated with each other.

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #30 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #31 of 191

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

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

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

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

Explanation

Multicollinearity occurs when two or more of the independent variables, or linear combinations of independent variables, may be highly correlated with each other. In a classic effect of multicollinearity, the R^2 is high and the F-statistic is significant, but the t-statistics on the individual slope coefficients are insignificant.

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #32 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #33 of 191

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

A) the error terms are correlated with each other.

B) two or more of the independent variables are highly correlated with each other.

C) the error term is normally distributed.

Explanation

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

Multicollinearity, on the other hand, occurs when two or more of the independent variables are highly correlated with each other. One assumption of multiple regression is that the error term is normally distributed.

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

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 #34 of 191

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

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

B) 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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #35 of 191

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) -0.018 to -0.036 ✗

B) 0.011 to 0.001 ✓

C) 0.014 to -0.009 ✗

Explanation

The confidence interval is 0.006 ± (1.96)(0.00271) = 0.011 to 0.001

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #36 of 191

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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #37 of 191

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) 18.4% of the variation in returns.

B) 84.4% of the variation in returns.

C) 15.6% 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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #38 of 191

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 = 1.97, fail to reject a hypothesis that both of the slope coefficients are equal to zero.

C) F = 17.00, 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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #39 of 191

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

B) 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.

C) 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

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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #40 of 191

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 0.05 and 0.10. ✗

C) Between 0.01 and 0.05. ✗

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.

(Study Session 3, Module 8.2, LOS 8.c)

Related Material

SchweserNotes - Book 1

Question #41 of 191

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

A) Serial correlation does not affect consistency of regression coefficients. ✗

B) Positive serial correlation and heteroskedasticity can both lead to Type I errors. ✗

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

Explanation

Serial correlation, which is sometimes referred to as autocorrelation, occurs when the residual terms are correlated with one another, and is most frequently encountered with time series data. Positive serial correlation can lead to standard errors that are too small, which will cause computed t-statistics to be larger than they should be, which will lead to too many Type I errors (i.e. the rejection of the null hypothesis when it is actually true). Serial correlation however does not affect the consistency of the regression coefficients.

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #42 of 191

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) Not rejected at 2.5% or 5.0% significance.

C) Rejected at 5% significance only.

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.

(Study Session 3, Module 8.3, LOS 8.g)

Related Material

SchweserNotes - Book 1

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	dl	du	dl	du	dl	du	dl	du	dl	du
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 #43 of 191

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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #44 of 191

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

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

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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #45 of 191

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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #46 of 191

According to the Durbin-Watson statistic, there is:

A) significant heteroskedasticity in the residuals. ✗

B) no significant positive serial correlation in the residuals. ✗

C) significant positive serial correlation 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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #47 of 191

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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #48 of 191

What is the standard error of the estimate?

A) 1.31.

B) 0.81.

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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #49 of 191

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) $R^2 = 0.20$ and F = 13.333.

B) $R^2 = 0.25$ and F = 0.930.

C) $R^2 = 0.25$ 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.

$$F = 100 / 7.5 = 13.333$$

(Study Session 3, Module 8, LOS 8.i)

Related Material

SchweserNotes - Book 1

Question #50 of 191

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

Coefficient	Coefficient Value	Standard error
Intercept	2.02	1.65
OIL	−0.25	0.18
Number of observations = 45		

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

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

C) The slope coefficient is statistically significant at 5% level of significance.

Explanation

$$t = -0.25/0.18 = 1.38$$

Critical values of t (2-tailed) at 5% level of significance = 2.02

Critical values of t (2-tailed) at 10% level of significance = 1.68

The absolute value of the computed t-statistic is lower than both. The slope coefficient is not statistically significant at 10% level of significance (and therefore cannot be significant at 5% level of significance).

(Study Session 3, Module 8.2, LOS 8.d)

Related Material

SchweserNotes - Book 1

Question #51 of 191

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.

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #52 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #53 of 191

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) 13. ✗

B) 11. ✓

C) 12. ✗

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.

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #54 of 191

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

A) Unconditional heteroskedasticity. ✓

B) Conditional heteroskedasticity. ✗

C) Multicollinearity. ✗

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #55 of 191

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

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

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

A) 1 and 2 only. ✗

B) 2 and 3 only. ✗

C) 3 only. ✓

Explanation

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

(Study Session 3, Module 8.2, LOS 8.c)

Related Material

SchweserNotes - Book 1

Question #56 of 191

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 highly 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.

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

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 #57 of 191

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

ARAR:

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

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

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

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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #58 of 191

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

A) 2.31.

B) 1.91.

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 $b_{Index} / t = \beta_{Index} = 1.1 / 2.1 = 0.5238$.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #59 of 191

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

A) 0.30.

B) 3.33.

C) 0.70.

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 $t = b_{Size} / \sigma_{Size} = 0.6 / 0.18 = 3.3333$.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #60 of 191

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

A) −9.45.

B) −2.86.

C) −0.11.

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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #61 of 191

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

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

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

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

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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #62 of 191

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

A) The expected value of the error term is zero.

B) The error term is normally distributed.

C) The error term is linearly related to the dependent variable.

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, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #63 of 191

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

A) Paired-comparisons.

B) Dummy variable regression.

C) Difference of means.

Explanation

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

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #64 of 191

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

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

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

A) INCOME only.

B) INCOME and ADV.

C) ADV only.

Explanation

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #65 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.4, LOS 8.h)

Related Material

SchweserNotes - Book 1

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

Model ONE	
Regression Statistics	
Multiple R	0.941828
R_2	0.887039
Adjusted R_2	0.863258
Standard Error	2.543272
Observations	24

Durbin-Watson test statistic = 0.7856

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

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

Model TWO
Regression Statistics

Multiple R	0.941796
R^2	0.886979
Adjusted R^2	0.870026
Standard Error	2.479538
Observations	24

Durbin-Watson test statistic = 0.7860

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

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

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

Question #66 of 191

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

A) Model ONE because it has a higher R^2.

B) Model TWO because serial correlation is not a problem.

C) Model TWO because it has a higher adjusted R^2.

Explanation

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

(Study Session 3, Module 8.4, LOS 8.h)

Related Material

SchweserNotes - Book 1

Question #67 of 191

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

A) $56.02 million.

B) $46.31 million.

C) $51.09 million.

Explanation

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

$$31.4083 + (-2.4631) + (24 + 2) \times 0.851786 = 51.091666.$$

(Study Session 3, Module 8.4, LOS 8.h)

Related Material

SchweserNotes - Book 1

Question #68 of 191

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

A) The coefficients on Q1 and Q2 only.

B) The intercept only.

C) The coefficient on Q2 only.

Explanation

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

(Study Session 3, Module 8.4, LOS 8.h)

Related Material

SchweserNotes - Book 1

Question #69 of 191

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

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

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

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

Explanation

Presence of conditional heteroskedasticity will not affect the consistency of regression coefficients but will bias the standard errors leading to incorrect application of t-tests for statistical significance of regression parameters.

(Study Session 3, Module 8.4, LOS 8.h)

Related Material

SchweserNotes - Book 1

Question #70 of 191

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

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

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

C) it would not have been significant. ✗

Explanation

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

$$Q1 = 31.40833 + (-3.77798) = 27.63035$$

Such a model would produce the same estimated values for the dependent variable.

(Study Session 3, Module 8.4, LOS 8.h)

Related Material

SchweserNotes - Book 1

Question #71 of 191

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

A) grow, but by less than $1,000,000.

B) grow by more than $1,000,000.

C) remain approximately the same.

Explanation

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

(Study Session 3, Module 8.4, LOS 8.h)

Related Material

SchweserNotes - Book 1

Question #72 of 191

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

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

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

A) $509,980,000.

B) $557,143,000.

C) $656,991,000.

Explanation

Predicted sales for next year are:

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #73 of 191

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) $R^2 = 0.20$ and F = 10.

B) $R^2 = 0.25$ and F = 0.909.

C) $R^2 = 0.25$ 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.

$$F = 20 / 2 = 10$$

(Study Session 3, Module 8, LOS 8.i)

Related Material

SchweserNotes - Book 1

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

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

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

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

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

	Coefficient	Standard Error
Intercept	2.1	2.01
Industry Index	1.9	0.31

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

Question #74 of 191

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

A) 54.98.

B) 59.18.

C) 65.48.

Explanation

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

(Study Session 3, Module 8.6, LOS 8.f)

Related Material

SchweserNotes - Book 1

Question #75 of 191

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

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

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

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

Explanation

We cannot reject the null hypothesis.

H_0: $b_{gender} \leq 0$

H_a: $b_{gender} > 0$

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

(Study Session 3, Module 8.6, LOS 8.f)

Related Material

SchweserNotes - Book 1

Question #76 of 191

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

A) 0.76 ± 1.96(8.44).

B) 1.25 ± 1.96(0.33).

C) 0.76 ± 1.96(0.09).

Explanation

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

(Study Session 3, Module 8.6, LOS 8.f)

Related Material

SchweserNotes - Book 1

Question #77 of 191

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

A) 0.76 ± 1.96(0.09).

B) 1.25 ± 1.96(0.33).

C) 1.25 ± 1.96(3.78).

Explanation

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

(Study Session 3, Module 8.6, LOS 8.f)

Related Material

SchweserNotes - Book 1

Question #78 of 191

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

A) 7.6%.

B) 11.2%.

C) 9.7%.

Explanation

$$Y = b_0 + bX_1$$

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

(Study Session 3, Module 8.6, LOS 8.f)

Related Material

SchweserNotes - Book 1

Question #79 of 191

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

A) 63.2%.

B) 72.1%.

C) 84.9%.

Explanation

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

(Study Session 3, Module 8.6, LOS 8.f)

Related Material

SchweserNotes - Book 1

Question #80 of 191

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

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

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

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

(Study Session 3, Module 8.2, LOS 8.c)

Related Material

SchweserNotes - Book 1

A real estate agent wants to develop a model to predict the selling price of a home. The agent believes that the most important variables in determining the price of a house are its size (in

square feet) and the number of bedrooms. Accordingly, he takes a random sample of 32 homes that has recently been sold. The results of the regression are:

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

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

Selected F- table values for significance level of 0.05:

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

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

Additional information regarding this multiple regression:

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

Question #81 of 191

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

A) $256,000 ✓

B) $114,000 ✗

C) $185,000 ✗

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #82 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #83 of 191

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

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

B) the slopes are not significant but the intercept is significant. ✗

C) the slopes are significant but the intercept is not. ✓

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #84 of 191

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

A) autocorrelation.

B) heteroskedasticity.

C) multicollinearity.

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #85 of 191

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

A) present but a statistical inference is still reliable.

B) not present and a statistical inference is reliable.

C) present and a statistical inference is unreliable.

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #86 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #87 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #88 of 191

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

$$= \$10 + 1.25\ (5) + 1.0\ (4) - 2.0\ (10) + 8\ (40)$$

$$= 10 + 6.25 + 4 - 20 + 320 = \$320.25$$

(Study Session 3, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #89 of 191

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

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

B) Heteroskedasticity only occurs in cross-sectional regressions.

C) Multicollinearity is a potential problem only in multiple regressions, not simple regressions.

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.

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #90 of 191

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.

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

Question #91 of 191

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

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

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

C) The desired impact cannot be measured.

Explanation

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

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

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

For the past several years, these marketing plans seemed to be successful, providing a significant boost in sales to those specific products highlighted by the marketing efforts. However, for the past year, revenues have been flat, even though marketing expenditures increased slightly. Brent is concerned that the expensive seasonal marketing campaigns are simply no longer generating the desired returns, and should either be significantly modified or eliminated altogether. He proposes that the company hire additional, permanent salespeople to focus on selling Mega Flowers' high-margin products all year long. The chief operating officer, David Johnson, disagrees with Brent. He believes that although last year's results were disappointing, the marketing campaign has demonstrated impressive results for the past five

years, and should be continued. His belief is that the prior years' performance can be used as a gauge for future results, and that a simple increase in the sales force will not bring about the desired results.

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

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

Brent shows the equation to Johnson and tells him, "This equation shows that a $1 million increase in marketing expenditures will increase the independent variable by $1 .6 million, all other factors being equal." Johnson replies , "It also appears that sales will equal $12.6 million if all independent variables are equal to zero."

Question #92 of 191

In regard to their conversation about the regression equation:

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

B) Brent's statement is correct; Johnson's statement is correct. ✗

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

Explanation

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

Johnson's statement is correct. 12.6 is the intercept in the equation, which means that if all independent variables are equal to zero, expected sales will be $12.6 million.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #93 of 191

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

A) both independent variables are statistically significant. ✓

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

C) the null hypothesis should not be rejected. ✗

Explanation

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #94 of 191

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

A) 14.055. ✗

B) 15.706. ✓

C) 14.831. ✗

Explanation

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #95 of 191

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

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

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

A) 1 of Brent's points are correct.

B) All 3 of Brent's points are correct.

C) 2 of Brent's points are correct.

Explanation

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

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

Therefore, 2 of Brent's points are correct.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #96 of 191

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

A) $24,200,000.

B) $11,600,000.

C) $24,000,000.

Explanation

Using the information provided, expected sales equals 12.6 + (1.6 x 3.5) + (1.2 x 5) = $24.2 million. Remember to check the details - i.e. this equation is denominated in millions of dollars.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #97 of 191

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

A) An ANOVA table.

B) The *F*-statistic.

C) The multiple coefficient of determination.

Explanation

To determine whether at least one of the coefficients is statistically significant, the calculated *F*-statistic is compared with the critical *F*-value at the appropriate level of significance.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #98 of 191

Bill Samuels, a summer intern at Capulito Securities, is evaluating the results of a regression model. The model was developed by the Capulito's senior economist several years ago, and Samuels decided to evaluate the model using more recent data. The model provides a forecast for the price of oil (per barrel in USD) based on the following independent variables:

LNG: The natural log of the global GDP (in trillions of USD) for the last quarter

USD: The trade-weighted value of the US dollar versus a basket of global currencies

GLD: The average price of an ounce of gold over the last quarter (in USD).

The regression output using the last ten years of quarterly data is shown below:

Variable	Coefficient	Standard Error
Intercept	23.12	1.975
LNG	4.83	0.972
USD	-1.22	0.25
GLD	0.012	0.0008

At a five percent level of significance, the coefficient for LNG is *most likely*:

A) significantly different from 4.

B) significantly greater than 3.

C) significantly less than 6.

Explanation

In this case, n=40 (for 10 years of quarterly data), k = the number of independent variables = 3. The critical t-value for 36 degrees of freedom and a 5% level of significance is 2.02 for a two-tailed test and 1.68 for a one-tailed test.

Hypotheses:

1. H_0: $b_1 \leq 3$ vs. H_a: $b_1 > 3$ (one-tailed test)

   T = (4.83-3.00)/0.972 = 1.88 > critical t value of 1.68. Reject the null (because the coefficient is significantly greater than 3.)

2. H_0: $b_1 = 4$ vs. H_a: $b_1 \neq 4$ (two-tailed test)

   t = (4.83-4.00)/0.972 = 0.85 < critical t value of 2.02. Fail to reject the null (i.e., the coefficient is NOT significantly different than 4).

3. H_0: $b_1 \geq 6$ vs. H_a: $b_1 < 6$ (one-tailed test)

   t = (4.83-6.00)/0.972 = -1.20. Compare the absolute value of 1.20 to the critical t-value of 1.68. We fail to reject the null (i.e., the coefficient is NOT significantly less than 6).

(Study Session 3, Module 8.2, LOS 8.d)

Related Material

SchweserNotes - Book 1

Question #99 of 191

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 = $\alpha \times \beta_1$ POP $\times \beta_2$ INCOME $\times \beta_3$ ADV $\times \varepsilon$. ✗

B) INCOME = $\alpha + \beta_1$ POP + β_2 SALES + β_3 ADV + ε. ✗

C) SALES = $\alpha + \beta_1$ 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.

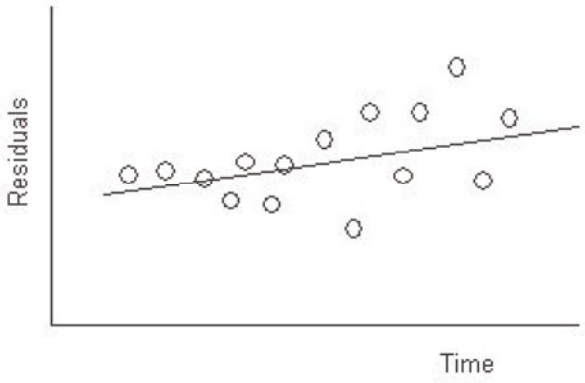
(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #100 of 191

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



These residuals exhibit the regression problem of:

A) homoskedasticity.

B) autocorrelation.

C) heteroskedasticity.

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #101 of 191

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) heteroskedasticity. ✗

B) serial correlation. ✗

C) multicollinearity. ✓

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.

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #102 of 191

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 will be higher, but the adjusted R^2 and F-statistic 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 and F-statistic will be higher, but the adjusted R^2 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.

(Study Session 3, Module 8.4, LOS 8.h)

Related Material

SchweserNotes - Book 1

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 #103 of 191

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, Module 8, LOS 8.i)

Related Material

SchweserNotes - Book 1

Question #104 of 191

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 linear association. ✓

B) have no relationship. ✗

C) have a positive but non-linear 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, Module 8, LOS 8.i)

Related Material

SchweserNotes - Book 1

Question #105 of 191

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

A) 6.10.

B) 57.07.

C) 37.25.

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 $\sqrt{(\text{MSE}} = \sqrt{37.25} = 6.10$

(Study Session 3, Module 8, LOS 8.i)

Related Material

SchweserNotes - Book 1

Question #106 of 191

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, Module 8, LOS 8.i)

Related Material

SchweserNotes - Book 1

Question #107 of 191

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) predicted value of the dependent variable equals 15. ✗

B) coefficient of the determination 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, Module 8, LOS 8.i)

Related Material

SchweserNotes - Book 1

Question #108 of 191

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 2, 3, and 4. ✓

B) Points 1, 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, Module 8, LOS 8.i)

Related Material

SchweserNotes - Book 1

Question #109 of 191

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) ADV and CAP only.

B) R&D, ADV, COMP, and CAP.

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.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #110 of 191

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

A) $Y_t = b_0 + b_1 D_t + e_t$.

B) $Y_t = b_1 D_t + b_2 X_t + e_t$.

C) $Y_t = b_0 + b_1 D_t + b_2 X_t + e_t$.

Explanation

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

Non-Democrats: $\hat{Y} = b_0 + b_2 X_t$

Democrats: $\hat{Y} = (b_0 + b_1) + b_2 X_t$

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #111 of 191

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) Between 0.01 and 0.05.

B) Lower than 0.01.

C) Between 0.05 and 0.10.

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).

(Study Session 3, Module 8.2, LOS 8.c)

Related Material

SchweserNotes - Book 1

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 #112 of 191

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

A) 61.78%.

B) 66.76%.

C) 81.71%.

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.

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #113 of 191

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

A) the January effect and the large capitalization index only.

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

C) the large cap 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.

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #114 of 191

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

A) 13.39 and the critical value is 3.86.

B) 9.05 and the critical value is 3.86.

C) 13.39 and the critical value is 3.10.

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.

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #115 of 191

The *best* test for unconditional heteroskedasticity is:

A) the Durbin-Watson test only. ✗

B) the Breusch-Pagan test only. ✗

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

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.

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #116 of 191

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) 2.322. ✓

B) 2.799. ✗

C) 2.561. ✗

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.

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #117 of 191

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.

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #118 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #119 of 191

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 0.025 and 0.05. ✗

B) lower than 0.025. ✓

C) Between 0.05 and 0.10. ✗

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.

(Study Session 3, Module 8.3, LOS 8.g)

Related Material

SchweserNotes - Book 1

Question #120 of 191

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

A) Positive serial correlation. ✓

B) Multicollinearity. ✗

C) Homoskedasticity. ✗

Explanation

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

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #121 of 191

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) 10.00. ✓

B) 0.25. ✗

C) 0.10. ✗

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.

(Study Session 3, Module 8.3, LOS 8.g)

Related Material

SchweserNotes - Book 1

Question #122 of 191

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) 0.2

B) 5

C) 0.05

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.

(Study Session 3, Module 8.3, LOS 8.g)

Related Material

SchweserNotes - Book 1

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:

$$\text{Profits} = b_0 + b_1\text{WLS} - b_2\text{COMP} + \varepsilon$$

Based on the model, the estimated regression equation is:

$$\text{Profits} = 22.5 + 0.98(\text{WLS}) - 0.35(\text{COMP})$$

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 #123 of 191

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

A) 9.128.

B) 1.435.

C) 1.882.

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, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #124 of 191

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: $b_1 = 0.98$ versus H_a: $b_1 \neq 0.98$.

B) H_0: $b_1 = 0.35$ versus H_a: $b_1 \neq 0.35$.

C) H_0: $b_1 \leq 0.98$ 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, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #125 of 191

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) 17.956.

B) 18.896.

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, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #126 of 191

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

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, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #127 of 191

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 *most likely* to be described as:

A) unstable remnant deviation.

B) heteroskedasticity.

C) positive serial correlation.

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, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #128 of 191

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 Durbin-Watson statistic. ✓

B) the Breusch-Pagen test. ✗

C) the Hansen method. ✗

Explanation

The Durbin-Watson test (DW ≈ 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, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #129 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8, LOS 8.i)

Related Material

SchweserNotes - Book 1

Question #130 of 191

Which of the following is a potential remedy for multicollinearity?

A) Take first differences of the dependent variable.

B) Omit one or more of the collinear variables.

C) Add dummy variables to the regression.

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.

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

Question #131 of 191

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

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

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

C) The proportion of occurrences when the absolute value of the test statistic will be higher when beta is equal to 1 than when beta is not equal to 1 is less than or equal

Explanation

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

(Study Session 3, Module 8.2, LOS 8.c)

Related Material

SchweserNotes - Book 1

Question #132 of 191

Consider the following regression equation:

$$Sales_i = 10.0 + 1.25\ R\&D_i + 1.0\ ADV_i - 2.0\ COMP_i + 8.0\ CAP_i$$

where Sales is dollar sales in millions, R&D is research and development expenditures in millions, ADV is dollar amount spent on advertising in millions, COMP is the number of competitors in the industry, and CAP is the capital expenditures for the period in millions of dollars.

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

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

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

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

Explanation

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #133 of 191

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

A) a normal distribution.

B) homoskedasticity.

C) a linear relationship.

Explanation

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

(Study Session 3, Module 8.6, LOS 8.f)

Related Material

SchweserNotes - Book 1

Question #134 of 191

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 Breusch-Pagan test.

B) The Durbin-Watson statistic.

C) The Hansen method.

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 ≈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.

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #135 of 191

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

A) $R^2 = SS_{Regression} - SS_{Error} / SS_{Total}$.

B) $R^2 = SS_{Regression} / SS_{Total}$.

C) $R^2 = SS_{Error} / SS_{Total}$.

Explanation

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

(Study Session 3, Module 8, LOS 8.i)

Related Material

SchweserNotes - Book 1

Question #136 of 191

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

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

where:

EPS_t is a quarterly observation of earnings per share

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

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

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

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

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

B) EPS for the first quarter is represented by the residual.

C) significance of the coefficients cannot be interpreted in the case of dummy variables.

Explanation

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

(Study Session 3, Module 8.5, LOS 8.j)

Related Material

SchweserNotes - Book 1

Question #137 of 191

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

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

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.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

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

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

Question #138 of 191

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

A) 108; 12.

B) 11; 120.

C) 12; 108.

Explanation

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

k = 12

n − k − 1 = 121 − 12 − 1 = 108

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #139 of 191

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

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

B) R^2 is less than adjusted R^2.

C) Adjusted R^2 always decreases as independent variables increase.

Explanation

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

$$R^2 \text{ adjusted} = 1 - [((n - 1) / (n - k - 1)) \times (1 - R^2)]$$

where:

n = number of observations

K = number of independent variables

R^2 = unadjusted R^2

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #140 of 191

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

A) Dickey-Fuller.

B) Breusch-Pagan.

C) Durbin-Watson.

Explanation

Durbin-Watson is used to detect autocorrelation. The Breusch-Pagan test is used to detect heteroskedasticity. The Dickey Fuller test is a test for unit root.

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #141 of 191

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

A) adjust the standard errors.

B) improve the specification of the model.

C) use robust standard errors.

Explanation

Using generalized least squares and calculating robust standard errors are possible remedies for heteroskedasticity. Improving specifications remedies serial correlation. The standard error cannot be adjusted, only the coefficient of the standard errors.

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #142 of 191

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

A) can only be used to detect positive serial correlation.

B) is approximately equal to 1 if the error terms are not serially correlated.

C) only uses error terms in its computations.

Explanation

The formula for the Durbin-Watson statistic uses error terms in its calculation. The Durbin-Watson statistic is approximately equal to 2 if there is no serial correlation. A Durbin-Watson statistic significantly less than 2 may indicate positive serial correlation, while a Durbin-Watson statistic significantly greater then 2 may indicate negative serial correlation.

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #143 of 191

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

A) heteroskedasticity.

B) serial correlation.

C) multicollinearity.

Explanation

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

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #144 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #145 of 191

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) 17.50.

B) 14.10.

C) 14.90.

Explanation

Predicted AUTO

$$= 10 + 1.25\ (4) + 1.0\ (0.30) - 2.0\ (0.6)$$

$$= 10 + 5 + 0.3 - 1.2$$

$$= 14.10$$

(Study Session 3, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

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

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

Analysis of Variance Table (ANOVA)		
Source	**Degrees of Freedom**	**Sum of Squares**
Regression		291.30
Error	27	132.12
Total	29	423.42

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

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

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

Question #146 of 191

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

A) $82.00 ✗

B) $206.00 ✓

C) $128.00

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #147 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #148 of 191

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

A) only GDP changes explain changes in sales.

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

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

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #149 of 191

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

A) Serial correlation.

B) Conditional Heteroskedasticity.

C) Multicollinearity.

Explanation

Multicollinearity is a violation of regression assumptions that may cause estimates of the regression coefficients to become extremely imprecise and unreliable and possibly lead to estimates having the opposite sign to that expected.

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #150 of 191

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

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

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

C) is correct.

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #151 of 191

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

A) computed F-statistic.

B) coefficient estimates.

C) computed t-statistic.

Explanation

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

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #152 of 191

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) 0.260 to 0.381.

C) 0.267 to 0.374.

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.

(Study Session 3, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #153 of 191

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

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

B) The adjusted-R^2 not appropriate to use in simple regression.

C) The adjusted-R^2 is greater than the R^2 in multiple regression.

Explanation

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

(Study Session 3, Module 8.4, LOS 8.h)

Related Material

SchweserNotes - Book 1

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

$$\text{sales} = b_0 + (b_1 \times \text{CPI}) + (b_2 \times \text{IP}) + (b_3 \times \text{GDP}) + \varepsilon_t$$

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 #154 of 191

Concerning the assumptions of multiple regression, Grimbles is:

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

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

C) correct to agree with Voiku's list of assumptions.

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).

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #155 of 191

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

A) Hypothesis 2.

B) Hypothesis 3.

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.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #156 of 191

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) fail to reject the null hypothesis because the F-statistic is smaller 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) reject the null hypothesis because the F-statistic is larger 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.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #157 of 191

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

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

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

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

Explanation

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

$$\text{unadjusted } R^2 = (368.7 - 140.3) / 368.7 = 0.619$$

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #158 of 191

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

A) multicollinearity.

B) serial correlation of the error terms.

C) heteroskedasticity.

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).

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #159 of 191

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.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

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 #160 of 191

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

(Study Session 3, Module 8.1, LOS 8.b)

Related Material

SchweserNotes - Book 1

Question #161 of 191

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 2% and 5%.

C) between 5% and 10%.

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.

(Study Session 3, Module 8.1, LOS 8.b)

Related Material

SchweserNotes - Book 1

Question #162 of 191

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

A) 0.53

B) 0.76

C) 0.36

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

(Study Session 3, Module 8.1, LOS 8.b)

Related Material

SchweserNotes - Book 1

Question #163 of 191

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)	$H_0: b_3 \geq 5$	Fail to reject H_0	✗
B)	$H_0: b_3 \leq 5$	Reject H_0	✗
C)	$H_0: b_3 \leq 5$	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

(Study Session 3, Module 8.1, LOS 8.b)

Related Material

SchweserNotes - Book 1

Question #164 of 191

The standard deviation of regression residuals is *closest* to:

A) 239.42 ✗

B) 0.42 ✗

C) 15.47 ✓

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.

(Study Session 3, Module 8.1, LOS 8.b)

Related Material

SchweserNotes - Book 1

Question #165 of 191

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.

(Study Session 3, Module 8.1, LOS 8.b)

Related Material

SchweserNotes - Book 1

Question #166 of 191

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

A) Scatter plot. ✗

B) Durbin-Watson test. ✓

C) Breusch-Pagan test. ✗

Explanation

The Durbin-Watson test is used to detect serial correlation. The Breusch-Pagan test is a formal test used to detect heteroskedasticity while a scatter plot can give visual clues about presence of heteroscedasticity.

(Study Session 3, Module 8.6, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #167 of 191

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) 3y + 2z.

B) 9y - 4z + 3

C) y^2.

Explanation

Multicollinearity is when one independent variable in a multiple regression model can be linearly predicted from the other independent variables. If x is related to y^2, the relationship between x and y is not linear, so multicollinearity need not exist.

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

Question #168 of 191

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

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

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

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

Explanation

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

(Study Session 3, Module 8.6, LOS 8.f)

Related Material

SchweserNotes - Book 1

Question #169 of 191

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

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

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

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

A) PI and INS only.

B) PI only.

C) TEEN only.

Explanation

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

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #170 of 191

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

B) If the *t*-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.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #171 of 191

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) should have a lower coefficient of determination.

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

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.

(Study Session 3, Module 8.4, LOS 8.h)

Related Material

SchweserNotes - Book 1

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	StandardError	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			
Durbin Watson	1.22			

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 #172 of 191

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

A) 1,394,420 ✗

B) 1,394 ✗

C) 1,751,000 ✓

Explanation

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

(Study Session 3, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #173 of 191

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

A) −3.000 to +1.000

B) −1.850 to −0.151

C) −1.852 to −0.149

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, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #174 of 191

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

A) No, because $|-2.0| < 2.045$

B) No, because $|-2.0| < 2.052$

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 ≠ 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, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #175 of 191

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

A) The large F statistic indicates that both independent variables help explain changes 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 independent variables explain 61.58% of the variation 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, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #176 of 191

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

A) 0.47 ✓

B) 0.22 ✗

C) 0.3 ✗

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, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #177 of 191

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

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

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

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, Module 8.2, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #178 of 191

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

	Age	Education	Experience	Income
Age	1.00			
Education	0.50	1.00		
Experience	0.95	0.55	1.00	
Income	0.60	0.65	0.89	1.00

A) Age should be excluded from the regression.

B) Experience may be a redundant variable.

C) Education may be unnecessary.

Explanation

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

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

Question #179 of 191

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) Lower than 0.01.

B) Between 0.01 and 0.05.

C) Between 0.05 and 0.10.

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.

(Study Session 3, Module 8.2, LOS 8.c)

Related Material

SchweserNotes - Book 1

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

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

The correlation matrix for the independent variables appears below:

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

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

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

R^2 = 22.44%

adj. R^2 = 5.81%

standard error of forecast = 0.0734 (percent)

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

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

The results for Equation II are:

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

R^2 = 22.37%

adj. R^2 = 12.02%

standard error of forecast = 0.0710 (percent)

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

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

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

The results for Equation III are:

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

$R^2 = 20.94\%$

adj. $R^2 = 16.00\%$

standard error of forecast = 0.0693 (percent)

Question #180 of 191

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

A) $H_0: b_1 = 0$ (T-bill)

B) $H_0: b_2 = 0$ (S&P 500)

C) $H_0: b_0 = 0$ (intercept)

Explanation

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

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

Question #181 of 191

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

A) H_0: $b_0 = 0$ (intercept) only.

B) H_0: $b_0 = 0$ (intercept) and $b_1 = 0$ (T-bill) only.

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

Explanation

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

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

Question #182 of 191

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

A) reason to be suspicious, but she took the wrong step to cure the problem.

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

C) reason to be suspicious and took the correct step to cure the problem.

Explanation

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

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

Question #183 of 191

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

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

Explanation

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

t = (0.51 − 0.50)/0.246 = 0.04065 (< 1.753). Hence fail to reject the null hypothesis.

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

Question #184 of 191

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

A) Neither multicollinearity, nor serial correlation. ✓

B) Both multicollinearity and serial correlation. ✗

C) Multicollinearity, but not serial correlation. ✗

Explanation

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

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

Question #185 of 191

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

A) 0.296 to 0.538.

B) 0.259 to 0.598.

C) 0.270 to 0.564.

Explanation

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

(Study Session 3, Module 8.8, LOS 8.l)

Related Material

SchweserNotes - Book 1

Question #186 of 191

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 X3 is statistically significantly different from zero at the 2% significance level.

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

C) The variable X2 is statistically significantly different from zero at the 3% 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.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #187 of 191

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

A) The variance of the error terms is not constant (i.e., the errors are heteroskedastic).

B) The error term is normally distributed.

C) The independent variables are not random.

Explanation

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

(Study Session 3, Module 8.6, LOS 8.f)

Related Material

SchweserNotes - Book 1

Question #188 of 191

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

C) evidence of serial correlation but not conditional heteroskedasticity 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.

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #189 of 191

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

A) A scatter plot of the residuals over time.

B) The Durbin-Watson test.

C) The Breusch-Pagan test.

Explanation

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

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #190 of 191

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

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

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

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

Explanation

The Hansen procedure simultaneously solves for heteroskedasticity and serial correlation.

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1

Question #191 of 191

An analyst is estimating whether a fund's excess return for a month is dependent on interest rates and whether the S&P 500 has increased or decreased during the month. The analyst collects 90 monthly return premia (the return on the fund minus the return on the S&P 500 benchmark), 90 monthly interest rates, and 90 monthly S&P 500 index returns from July 1999 to December 2006. After estimating the regression equation, the analyst finds that the correlation between the regressions residuals from one period and the residuals from the previous period is 0.145 (DW=1.71). Which of the following is *most* accurate at a 0.05 level of significance, based solely on the information provided? The analyst:

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

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

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

Explanation

The Durbin-Watson statistic tests for serial correlation., which is higher than the lower Durbin-Watson value (with 2 variables and 90 observations) of 1.61. That means the hypothesis of no serial correlation cannot be rejected. There is no information on whether the regression exhibits heteroskedasticity.

(Study Session 3, Module 8.7, LOS 8.k)

Related Material

SchweserNotes - Book 1