CHAPTER 1

MULTIPLE REGRESSION

1. (A) 5.

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.

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

2. (B) Homoskedasticity

Explanation

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

(Module 1.1, LOS 1.c)

Related Material

SchweserNotes - Book 1

3. (C) Incorrectly pooling data.

Explanation

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

(Module 1.3, LOS 1.g)

Related Material

SchweserNotes - Book 1

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

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.

(Module 1.4, LOS 1.l)

Related Material

SchweserNotes - Book 1

5. (B) **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.

(Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

6. (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.

(Module 1.3, LOS 1.j)

Related Material

SchweserNotes - Book 1

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

Salary = 34.98 + 1.2 Education + 0.5 Experience + 6.3 Gender

(29.11) qw (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 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

7. (B) 59.18

Explanation

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

(Module 1.2 LOS 1.f)

Related Material

SchweserNotes - Book 1

8. (B) 9.7%

Explanation

$Y = b_0 + bX_1$

Y = 2.1 + 1.9(4) = 9.7%

(Module 1.2 LOS 1.f)

Related Material

SchweserNotes - Book 1

9. (B) 72.1%

Explanation

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

(Module 1.2 LOS 1.d)

Related Material

SchweserNotes - Book 1

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

Sophie makes the following statements:

Statements 1: Multiple regression models can be used to forecast independent variables.

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

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

$CSP = 0.22 + 1.04 \times DSC - 0.32 \times index + 1.33 \times D/E$

Where:

CSP = credit spread (%)

DSC = EBITDA/unsecured debt.

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

D/E = long-term debt/equity

10. (C) only Statement 2 is correct.

Explanation

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

(Module 1.2 LOS 1.a)

Related Material

SchweserNotes - Book 1

11. (B) The credit spread on the firm's issue will decrease by 32 bps.

Explanation

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

12. (B) The dependent variable is not serially correlated.

Explanation

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

(Module 1.1, LOS 1.c)

Related Material

SchweserNotes - Book 1

13. (B) **Error term is normally distributed.**

Explanation

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

(Module 1.1, LOS 1.c)

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

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

Additional information regarding this multiple regressions.

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.

14. (B) **$256,000.**

Explanation

66,500 + 74,30(2,000) + 10,306(4) = #256,324

(Module 1.2, LOS 1.f)

Related Material

SchweserNotes - Book 1

15. (B) 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.

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

16. (B) 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).

(Module 1.3, LOS 1.j)

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

17. (B) 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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

18. (B) 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 $b_{Index} / t = \beta_{Index} = 1.1 / 2.1 = 0.5238$.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

19. (B) 3.33.

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

20. (A) –2.86.

Explanation

The t-statistic for testing the null hypothesis H_0: β_i; = 0 is t = (b_i – 0) / σ_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 x σ_0 = – 5.2 x 0.55 = – 2.86.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

21. (B) **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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

22. (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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

23. (B) **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.

(Module 1.1. LOS 1.b)

SchweserNotes - Book 1

24. (A) 1 1

Explanation

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

(Module 1.4, LOS 1.l)

Related Material

SchweserNotes - Book 1

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

Explanation

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

(Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

26. (B) **Omit one or more of the collinear variables.**

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.

(Module 1.3, LOS 1.i)

Related Material

SchweserNotes - Book 1

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

Indian Equities-Fama-French Model

Factor	Coefficient	P-value	VIF
Intercept	1.22	< 0.001	
SMB	0.23	< 0.001	3
HML	0.34	0.003	3
Rm-Rf	0.88	< 0.001	2

R-squared		0.36	
SSE		38.00	
BG (lag 1)		2.11	
BG (lag 2)		1.67	

Partial F-Table (5% Level of Significance)

Degrees of Freedom Denominator	Degree of Freedom Numerator		
	1	2	3
112	3.93	3.08	2.69
113	3.93	3.08	2.68
114	3.92	3.08	2.68
115	3.92	3.08	2.68
116	3.92	3.07	2.68
117	3.92	3.07	2.68

Partial Chi-Square Table (5% Level of Significance)

Degrees of Freedom	Critical Value
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
6	12.59

27. (B) **Because the test statistic of 7.20 is lower than the critical value of 7.81, we fail to reject the null hypothesis of no conditional heteroskedasticity in residuals.**

Explanation

The chi-square test statistic = n x R^2 = 120 x 0.06 = 7.20.

The one-tailed critical value for a chi-square distribution with k = 3 degrees of freedom and α of 5% is 7.81. Therefore, we should not reject the null hypothesis and conclude that we don't have a problem with conditional heteroskedasticity.

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

28. (C) **Data improperly pooled.**

Explanation

Out of the four forms of model misspecifications, serial correlation in residuals may be caused by omission of important variables (not an answer choice) and by improper data pooling.

(Module 1.3, LOS 1.g)

Related Material

SchweserNotes - Book 1

29. (B) **No.**

Explanation

The BG test statistic has an F-distribution with *p* and n – *p* – *k* – 1 degrees of freedom, where p = the number of lags tested. Given *n* = 120 and *k* = 3, critical F-values (5% level of significance) are 3.92 (p = 1) and 3.08 (p = 2). BG stats in **Indian Equities—Fama-French Model** are lower than the critical F-values; therefore, serial correlation does not seem to be a problem for both lags.

(Module 1.3, LOS 1.i)

Related Material

SchweserNotes - Book 1

30. (C) **No.**

Explanation

Multicollinearity is detected using the variance inflation factor (VIF). VIF values greater than 5 (i.e., $R^2 > 80\%$) warrant further investigation, while values above 10 (i.e., $R^2 > 90\%$) indicate severe multicollinearity. None of the variables have VIF > 5.

(Module 1.3, LOS 1.j)

Related Material

SchweserNotes - Book 1

31. (C) **Unconditional heteroskedasticity.**

Explanation

Unconditional heteroskedasticity does not impact the statistical inference concerning the parameters. Misspecified models have inconsistent and biased regression parameters. Multicollinearity results in unreliable estimates of regression parameters.

(Module 1.3, LOS 1.h)

(Module 1.3, LOS 1.i)

(Module 1.3, LOS 1.j)

Related Material

SchweserNotes - Book 1

32. (B) **heteroskedasticity.**

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.

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

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

Explanation

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

(Module 1.3, LOS 1.i)

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.

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

Explanation

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

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

(Module 1.3, LOS 1.j)

Related Material

SchweserNotes - Book 1

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

Explanation

One problem with conditional heteroskedasticity while working with financial data, 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.

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

36. (B) 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.

(Module 1.3, LOS 1.j)

Related Material

SchweserNotes - Book 1

37. (C) the error terms are correlated with each other.

Explanation

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

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

(Module 1.3, LOS 1.i)

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 statistics = 0.7856

ANOVA					
	df	SS	MS	F	Significance F
Regression	**4**	**965.0619**	**241.2655**	**37.30006**	**9.49E-09**
Residual	19	122.8964	6.4685		
Total	23	1087.9583			

	Coefficients	Standard Error	t-statistics
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 ONE	
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 = 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-statistics
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.

38. (B) 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 (> 7860), serial correlation is a problem for both equations.

(Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

39. (B) **$51.09 million.**

Explanation

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

31.4083 + (– 2.4631) + (24 + 2) x 0.851786 = 51.091666.

(Module 1.2, LOS 1.f)

Related Material

SchweserNotes - Book 1

40. (C) **Inappropriate variable scaling.**

Explanation

Inappropriate variable scaling may lead to multicollinearity or heteroskedasticity in residuals. Omission of important variable may lead to biased and inconsistent regression parameters and also heteroskedasticity/serial correlation in residuals. Inappropriate variable form can lead to heteroskedasticity in residuals.

(Module 1.3, LOS 1.g)

Related Material

SchweserNotes - Book 1

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

Explanation

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

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

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

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.

(Module 1.4, LOS 1.l)

Related Material

SchweserNotes - Book 1

43. (C) grow, but by less than $1,000,000.

Explanation

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

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

Indian Equities-Farma-French Model

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

Shapule then modifies the model to include a liquidity factor. Results for this four-factor model (Model 2) are shown in

Revised Fama-French Model With Liquidity Factor

Revised Fama-French Model With Liquidity Factor

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

44. (B) 0.37.

Explanation

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

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

(Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

45. (B) Model 2 because it has a lower Akaike information criterion.

Explanation

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

(Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

46. (B) 13.33.

Explanation

$$F = \frac{(SSE_R - SSE_U)/q}{(SSE_U)/(n-k-1)}$$

where n = 120, k = 4, and q = 1.

$$\frac{(38-34)/1}{(34)/(120-4-1)} = 13.33$$

Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

47. (C) 6.80%.

Explanation

Model 1:

Return = 1.22 + 0.23 x SMB + 0.34 x HML + 0.88 Rm-Rf

= 1.22 + 0.23 x 3.30 + 0.34 x 1.25 + 0.88 x 5 = 6.80%.

(Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

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

Explanation

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

(Module 1.1, LOS 1.c)

Related Material

SchweserNotes - Book 1

49. (A) **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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

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

Explanation

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

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

(Module 1.2, LOS 1.f)

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 (Degree 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)**	

51. (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.

(Module 1.1, LOS 1.c)

Related Material

SchweserNotes - Book 1

52. (C) **0.916, indicating that the variability of industry returns explains about 91.6% of the variability of company 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).

(Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

53. (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.

(Module 1.2, LOS 1.f)

Related Material

SchweserNotes - Book 1

54. (C) **Points 2, 3, and 4.**

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 because it describes conditional heteroskedasticity, which results in unreliable estimates of standard errors. Points 2 and 3 also describe limitations of regression analysis.

(Module 1.1, LOS 1.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**	

	Coefficient	Standard Error	t-Statistics
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

Exhibit 1: Partial F-Table (5% Level of Significance)

Degree of Freedom Denominator	Degree of Freedom Numerator		
	1	2	3
18	4.41	3.55	3.16
19	4.38	3.52	3.13
20	4.35	3.49	3.10
21	4.32	3.47	3.07
22	4.30	3.44	3.05
23	4.28	3.42	3.03

Gloucester plans to test for serial correlation and conditional and unconditional heteroskedasticity.

55. (A) **66.76%.**

Explanation

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

(Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

56. (A) **No, because the BG statistic is less than the critical test statistic of 3.55, we don't have evidence of serial correlation.**

Explanation

Number of lags tested = p = 2. The appropriate test statistic for BG test is F-stat with (p = 2) and (n – p – k – 1 = 18) degrees of freedom. From the table, critical value = 3.55.

(Module 1.3, LOS 1.i)

Related Material

SchweserNotes - Book 1

57. (A) **of 1.30 indicates that we cannot reject the hypothesis that the coefficient of small-cap index is not significantly different from 0.**

Explanation

$SSE_R = SST - RSS_R = 164.9963 - 106.3320 = 58.6643$

$F = [(SSE_R - SSE_U) / q] / [SSE_U / (n - k - 1)] = [(58.6643 - 54.8395) / 1] / (54.8395 / 20) = 3.8248 / 2.742 = 1.30$

Critical $F_{(1, 20)} = 4.35$ (from Exhibit 1)

Since the test statistic is not greater than the critical value, we cannot reject the null hypothesis that $b_2 = 0$.

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

58. (C) **neither the Durbin-Watson test nor the Breusch-Pagan test.**

Explanation

Breusch-Godfrey and Durbin-Watson tests are 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.

(Module 1.3, LOS 1.h)

(Module 1.3, LOS 1.i)

Related Material

SchweserNotes - Book 1

59. (A) **2.322.**

Explanation

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

(Module 1.2, LOS 1.f)

Related Material

SchweserNotes - Book 1

60. (B) **multicollinearity**

Explanation

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

(Module 1.3, LOS 1.j)

Related Material

SchweserNotes - Book 1

61. (A) 10.00

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.

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

62. (A) heteroskedasticity

Explanation

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

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

63. (A) Breusch-Godfrey test

Explanation

The Breusch-Godfrey 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.

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

64. (C) Model misspecification.

Explanation

When data are improperly pooled over multiple economic environments in a multiple regression analysis, the model would be misspecified.

(Module 1.3, LOS 1.g)

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:

Sales = $b_0 + (b_1 \times CPI) + (b_2 \times IP) + (b_3 \times GDP) + \varepsilon_t$

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 independents variables are not random, and there is zero correlation between any two of the independent variables.
Assumption 3:	The residual term is normally distributes with an expected value of zero.
Assumption 4:	The residual are serially correlated.
Assumption 5:	The variance of the residuals id 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 increases 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

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

Explanation

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

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

66. (B) Hypothesis 2.

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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

67. (A) **reject the 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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

68. (B) **incorrect in her calculation of both the unadjusted R^2 and the standard error of estimate.**

Explanation

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

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

69. (C) **multicollinearity**

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).

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

70. (A) **0.5 to 22.9**

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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

71. (A) **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. $X_1 + X_2 + X_3 = 1$.

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

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

72. (B) **INCOME only.**

Explanation

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

73. (B) **Rejected at 2.5% significance and 5% significance.**

Explanation

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

RSS = SST – SSE = 430 – 170 = 260

MSR = 260 / 5 = 52

MSE = 170 / (48 – 5 – 1) = 4.05

F = 52 / 4.05 = 12.84

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

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

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

Explanation

The test for conditional heteroskedasticity involves regressing the square of the residuals on the independent variables of the regression and creating a test statistic that is n x 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 x 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.

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

75. (C) **Heteroskedasticity only occurs in cross-sectional 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. Unconditional heteroskedasticity occurs when the heteroskedasticity is not related to the level of the independent variables. Unconditional heteroskedasticity causes no major problems with the regression. Breusch-Pagan statistic has a chi-square distribution and can be used to detect conditional heteroskedasticity.

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

76. (B) **The variable X3 is statistically significantly different from zero at the 2% significance level.**

Explanation

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

Dave Turner is security analyst 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 analysis following the companies. Ln(no. of analysis), and the 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 correspondence to a two-tail test.

Turner plants 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 marker 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**	

77. (B) The intercept and the coefficient on ln(no. of analysts) 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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

78. (A) 0.011 to 0.001.

Explanation

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

79. (C) –0.019.

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

80. (A) **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%.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

81. (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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

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

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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

83. (A) **Intercept term.**

Explanation

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

84. (B) **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 = 1.25/0.145 = 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.)

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

85. (C) **The Durbin-Watson statistic.**

Explanation

The Durbin-Watson statistic is the most commonly used method for the detection of serial correlationat the first lag, although residual plots can also be utilized. For testing of serial correlation beyond the first lag, we can instead use the Breusch-Godfrey test (but is not one of the answer choices).

(Module 1.3, LOS 1.i)

Related Material

SchweserNotes - Book 1

86. (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).**

Explanation

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

(Module 1.4, LOS 1.l)

Related Material

SchweserNotes - Book 1

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

Explanation

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

(Module 1.3 LOS 1.h)

Related Material

SchweserNotes - Book 1

88. (B) **logistic regression model.**

Explanation

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

(Module 1.4, LOS 1.m)

Related Material

SchweserNotes - Book 1

89. (A) 14.10.

Explanation

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

= 10 + 5 + 0.3 – 1.2

= 14.10

(Module 1.2, LOS 1.f)

Related Material

SchweserNotes - Book 1

90. (A) **The assumption of linear regression is that the residuals are heteroskedastic.**

Explanation

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

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

91. (B) **multicollinearity**

Explanation

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

(Module 1.3, LOS 1.j)

Related Material

SchweserNotes - Book 1

92. (B) **If R&D and advertising expenditure are $1 million each, there are 5 competitors, and capital expenditure are $2 million, expected Sales are $8.25 million.**

Explanation

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

93. (A) **$509,980,000.**

Explanation

Predicted sales for next year are:

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

Werner Baltz, CFA, has regressed 30 years of data for forecast future sales for National Motor Company based on the percent change in gross domestic (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	13.170
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**

94. (C) $206.00.

Explanation

Sales will be closest to $78 + ($30.22 x 2.2) + [(–412.39) x (–$0.15)] = $206.34 million

(Module 1.2, LOS 1.f)

Related Material

SchweserNotes - Book 1

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

Explanation

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

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

96. (B) coefficient estimates.

Explanation

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

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

97. (C) The regression will still exhibit multicollinearity, but the heteroskedasticity and serial correlation problems will be solved.

Explanation

The correction mentioned solves for heteroskedasticity and serial correlation. (Module 1.3, LOS 1.h)

(Module 1.3, LOS 1.i)

Related Material

SchweserNotes - Book 1

98. (A) Multicollinearity does not seem to be a problem with the model.

Explanation

Multicollinearity occurs when an independent variable is highly correlated with a linear combination of the remaining independent variables. VIF values exceeding 5 need to be investigated while values exceeding 10 indicate strong evidence of multicollinearity.

(Module 1.3, LOS 1.j)

Related Material

SchweserNotes - Book 1

99. (A) 11.

Explanation

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

(Module 1.4, LOS 1.l)

Related Material

SchweserNotes - Book 1

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

Explanation

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

(Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

101. (A) **Transforming a variable.**

Explanation

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

(Module 1.3, LOS 1.g)

Related Material

SchweserNotes - Book 1

William Brent, CFA, is the chief financial officer for Mega Flowers, one of the largerest producers of flowers and bedding plants in the Western United States. Mega Flowers its plants in three large nursery facilities located in California. Its products are sold in its company-owned retail nurseries as well as large, home and garden "super centers". For it 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 plain, 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 through 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 result 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 (states in million 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 shown that a $1 million increase in marketing expenditures will increase the independent variable by $1.6 million 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".

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

Explanation

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

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

103. (C) **both independent variables are statistically significant.**

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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

104. (B) 15.706.

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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

105. (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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

106. (C) $24,200,000.

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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

107. (B) The F-statistic.

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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

108. (A) R^2 = 0.20 and F= 10.

Explanation

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

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

F= 20 /2 = 10

(Module 1.2, LOS 1.d)

(Module 1.2, LOS 1.e)

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 in 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 sale 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 use a sample of 26 observation obtained from 26 metropolitan areas in the Midwest U.S. The results are in the tables below. The dependent variable is in sales of generation is million 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**	

Table of the F-Distribution

Critical values for right-hand tail area equal to 0.05

Numerator: df1 and Denominator: df2

	df1				
df2	1	2	4	10	20
1	161.45	199.50	224.58	241.88	248.01
2	18.513	19.000	19.247	19.396	19.446
4	7.7086	6.9443	6.3882	5.9644	5.8025
10	4.9646	4.1028	3.4780	2.9782	2.7740
20	4.3512	3.4928	2.8661	2.3479	2.1242

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 plants to perform diagnostic tests to verify the validity of the model's results.

109. (B) $35.2 million above the average.

Explanation

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

[5 + (2 x 0.10) + (3 x 5) + (10 x 3) + (5 x (-3))] x $1,000,000 = $35.2 million.

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

110. (C) at least one of the independent variables has explanatory power.

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.

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

111. (B) both a Breusch-Godfrey test and a Breusch-Pagan test.

Explanation

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

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

112. (C) adjusted 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)) x (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.

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

113. (A) 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.

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

114. (A) Multiple regression model

Explanation

Fye wants to test a theory of January effect on stock returns (dependent variable) using a dummy (January = 1, other months = 0), market cap, and beta (independent variables). A multiple regression model would be most appropriate. Because the dependent variable (stock returns) is not a qualitative variable, a logistic regression would not apply.

(Module 1.1, LOS 1.a)

Related Material

SchweserNotes - Book 1

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

Liu included the following variables in her 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- Result from Regressing Housing Starts (in Millions) on Interest Rates and GDP Per Capita

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

Exhibit 2: 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%

115. (B) **1,751,000**

Explanation

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

(Module 1.2, LOS 1.f)

Related Material

SchweserNotes - Book 1

116. (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.

(Module 1.2, LOS 1.d)

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

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

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.

(Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

118. (A) **slope coefficient in a multiple regression is the value of the dependent variable for a given value of the independent variable.**

Explanation

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

(Module 1.1, LOS 1.b)

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

Variables 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-Distributions

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

119. (A) **$36,000.**

Explanation

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

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

Explanation

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

121. (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).

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

122. (A) **the joint significance of 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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

123. (C) **77.00.**

Explanation

The question is asking for the coefficient of determination.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

124. (A) standard errors are too low but coefficient estimate is consistent.

Explanation

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

Miles Mason, CFA, works for ABC Capital, a large money management company based in New York. Manson 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. 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 way from publicly traded funds into ABC's proprietary funds, ultimately returning to 50/50 spilt 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 can offer its clients 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 to 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 nor promising any measure of future performance, project possible return scenarios. He decides to conduct a regression analysis on all of ABC' in-house funds. He is going to use 12 independent economic variable 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 known he must determine if any corrective actions are necessary. Mason is using a sample size of 121 monthly returns.

125. (C) Breusch-Pagan.

Explanation

Durbin-Watson and Breusch-Godfrey test statistic are used to detect autocorrelation. The Breusch-Pagan test is used to detect heteroskedasticity.

(Module 1.3, LOS 1.i)

Related Material

SchweserNotes - Book 1

126. (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.

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

127. (B) multicollinearity

Explanation

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

(Module 1.3, LOS 1.j)

Related Material

SchweserNotes - Book 1

128. (C) $320.25 million.

Explanation

Predicted sales

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

= 10 + 6.25 + 4 – 20 + 320 = $320.25

(Module 1.2, LOS 1.f)

Related Material

SchweserNotes - Book 1

129. (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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

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

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

Jessica Jones, CFA, a friend of Peter, overhears that he is interested in regression analysis and warms him that whenever heteroskedasticity is present in multiple regression, it could undermine the regressions results. She mentions that one easy way to spot conditional heteroskedasticity it through a scatter plot, but she adds that there is a more formal test.

Unfortunately, she can't quire remember its name. Jessica believes that heteroskedasticity can be rectified using White-corrected standard errors. Her son Jonathan who has also taken part in the discussion, hears this comment and argues that White corrections would typically reduce the number of Type II error in financial data.

130. (C) **The return on a particular trading day.**

Explanation

The omitted variable is represented by the intercept. So, if we have four variables to represent Monday through Thursday, the intercept would represent returns on Friday. Remember when we want to distinguish between "n" classes we always use one less dummy variable the number of classes (n – 1).

(Module 1.4, LOS 1.l)

Related Material

SchweserNotes - Book 1

131. (B) **There is no value to calendar trading.**

Explanation

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

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

132. (C) **Breusch-Pagan, which is a one-tailed test.**

Explanation

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

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

133. (A) **Both are correct.**

Explanation

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

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

134. (A) **Positive serial correlation.**

Explanation

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

(Module 1.3, LOS 1.i)

Related Material

SchweserNotes - Book 1

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

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

In preparing an analysis of Treefell Company, Jack Lumber is asked to look at the company's relation to broad-based economic indicators. Lumber's analysis indicates that Treefell's monthly sales are related to changes in housing starts (H) and changes in the mortgage interest rate (M). The analysis covers the past 10 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 Mortgage Rates:	–2.6

Variable Descriptions

S = Treefell Sales (in thousands)

H = housing starts (in thousands)

M = mortgage interest rate (in percent)

November 20X6 Actual data

Treefell's monthly sales: $55,000

Housing starts: 150,000

Mortgage interest rate(%): 7.5

Partial Chi-Square Table (5% Level of significance)

Degrees of Freedom	Critical Value
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
6	12.59

136. (B) **$36,000.**

Explanation

1.76 + 0.23 x (150) – 0.08 x (7.5) = 35.66.

(Module 1.2, LOS 1.f)

Related Material

SchweserNotes - Book 1

137. (B) **Different from zero; sales will rise by $23 for every 100 house starts.**

Explanation

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

Remember the rule p-value < significance, then reject null

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

138. (A) **the joint significance of the independent variables.**

Explanation

The F-statistic is for the general linear F-test to test the null hypothesis that slope coefficients on all variables are equal to zero.

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

139. (A) **77.00.**

Explanation

The question is asking for the coefficient of determination.

(Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

140. (B) **With a test statistic of 13.53, we can conclude the presence of conditional heteroskedasticity.**

Explanation

Chi-square = n x R^2 = 123 x 0.11 = 13.53. Critical Chi-square (degree of freedom = k = 2) = 5.99. Because the test statistic exceeds the critical value, we reject the null hypothesis (of no conditional heteroskedasticity).

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

141. (B) **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

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

142. (C) **three dummy variables.**

Explanation

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

(Module 1.4, LOS 1.l)

SchweserNotes - Book 1

Damon Washburn, CFA, is currently enrolled as a part-time graduate student at State University. One of his recent assignment 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 realize 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 return on the S&P are explained by three macroeconomic variables: employed 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 of Durbin-Watson Statistics (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

143. (B) Two of the 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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

144. (A) **not rejected because the t-statistic is equal to 0.92.**

Explanation

The hypothesis is:

H_0: $b_{GDP} = 3.50$

H_a: $b_{GDP} \neq 0$ 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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

145. (A) **32%.**

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 $SS_{Regression}/SS_{Total}$, where the SS_{Total} is equal to $SS_{Regression} + SS_{Error}$. R^2 = 126.00/ (126.00 + 267.00) = 32%.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

146. (A) **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.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

147. (B) **5.0%.**

Explanation

Predicted quarterly stock return is 9.50% + (–4.50)(2.0%) + (4.20)(1.0%) + (–0.30)(–1.0%) = 5.0%.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

148. (C) 1.31.

Explanation

The standard error of the estimate is equal to $[SSE/(n – k – 1)]^{1/2}$

$= [267.00/156]^{1/2}$ = approximately 1.31.

(Module 1.1, LOS 1.b)

Related Material

SchweserNotes - Book 1

Jessica Jenkins, CFA, is looking at the retail property sector for her manager. She in undertaking a top down review as she feels this is the best way to analyse the industry segment. To predict U.S. property starts (housing), she has used regression analysis.

Jessica included the following variables in her analysis:

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

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

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

		Coefficient	Standard Error	T-statistic
Intercept Interest rate		0.42 – 1.0		3.1 – 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	1.27			

Exhibit 2 – 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%

149. (B) 1,751,000.

Explanation

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

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

150. (A) **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.

(Module 1.2, LOS 1.d)

(Module 1.2, LOS 1.e)

Related Material

SchweserNotes - Book 1

151. (C) Adjusted R-square can be higher than the coefficient of determination for a model with a good fit.

Explanation

Adjusted R-squared cannot exceed R-squared (or coefficient of determination) for a multiple regression.

(Module 1.2, LOS 1.d)

Related Material

SchweserNotes - Book 1

Philip Lee works for Song Bank as a quantitative analyst. He is currently working on a model to explain the returns (in %) of 20 hedge funds for the past year. He includes three independent variables:

- Market return = return on a broad-based stock index (in %)
- Closed = dummy variable (= 1 if the fund is closed to new investors; 0 otherwise)
- Prior period alpha = fund return for the prior 12 months – return on market (in %)

Estimated model: hedge funds return = 3.2 + 0.22 market return + 1.65 closed – 0.11 prior period alpha

Less is concerned about the impact of outliers on the estimated regression model and collects the following information:

Observation	1	2	3	4	5	6	7	8	9	10
Cook's D	0.332	0.219	0.115	0.212	0.376	0.232	0.001	0.001	0.233	0.389
Observation	11	12	13	14	15	16	17	18	19	20
Cook's D	0.089	0.112	0.001	0.001	0.219	0.001	0.112	0.044	0.517	0.212

Additionally, Lee wants to estimate the probability of a hedge fund closing to new investors, and he uses tow variables:

- Fund size = log of assets under management.
- Prior period alpha (defined earlier)

Results are shown as follows:

Variable	Coefficient
Intercept	– 3.76
Fund size	– 2.98
Prior period alpha	– 2.99

152. (B) **A closed fund is estimated to have an extra return of 1.65% relative to funds that are not closed.**

Explanation

The interpretation of the coefficient is the extra return relative to the alternative outcome.

(Module 1.4, LOS 1.l)

Related Material

SchweserNotes - Book 1

153. (C) **Studentized residuals.**

Explanation

Studentized residuals are used to identify outliers (in the dependent variable). Leverage is used to identify high-leverage observations (in the independent variable), while Cook's D is a composite measure (combines both independent and dependent variables) to identify influential observations.

(Module 1.4, LOS 1.k)

Related Material

SchweserNotes - Book 1

154. (A) **Observations 10 and 19.**

Explanation

Influential observations are those that, when excluded, cause a significant change to the model coefficients.

Observations where Cook's D > $\frac{\sqrt{k}}{n} = \frac{\sqrt{3}}{20} = 0.3873.$

Observations 10 (D = 0.389) and 19 (D = 0.517) satisfy this criteria.

(Module 1.4, LOS 1.k)

Related Material

SchweserNotes - Book 1

155. (C) **4.83%.**

Explanation

Odds = e^{coeff} (fund size) = $e^{-2.98}$ = 0.0508.

Probability = odds / (1 + odds) = 0.0483 = 4.83%.

(Module 1.4, LOS 1.m)

Related Material

SchweserNotes - Book 1

156. (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.

(Module 1.3, LOS 1.h)

Related Material

SchweserNotes - Book 1

157. (C) The Breusch-Pagan test.

Explanation

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

(Module 1.3, LOS 1.i)

Related Material

SchweserNotes - Book 1

158. (B) incorrect variable form.

Explanation

Incorrect variable form misspecification occurs if the relationship between dependent and independent variables is nonlinear.

(Module 1.3, LOS 1.g)

Related Material

SchweserNotes - Book 1