1 MULTIPLE REGRESSION

1. Consider the following analysis of variance table:

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

The F-statistic for a test of joint significance of all the slope coefficients is closest to:

(A) 5.

(B) 0.2.

(C) 0.05.

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

(A) a linear relationship.

(B) homoskedasticity.

(C) a normal distribution.

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

(A) Inappropriate variable form.

(B) Inappropriate variable scaling.

(C) Incorrectly pooling data.

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

(A) $R = a + bM + c_1 D_1 + c_2D_2 + c_3D_3 + \varepsilon$, where $D_1 = 1$ if the return is from the first manager, and $D_2 = 1$ if the return is from the second manager, and $D_3 = 1$ is the return is from the third manager.

(B) The desired impact cannot be measured.

(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.

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

(A) should have a higher coefficient on the independent variable.

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

(C) should have a lower coefficient of determination.

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

(A) serial correlation.

(B) conditional heteroskedasticity.

(C) multicollinearity.

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

(A) 65.48.

(B) 59.18.

(C) 54.98.

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

(A) 11.2%.

(B) 9.7%.

(C) 7.6%.

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

(A) 63.2%.

(B) 72.1%.

(C) 84.9%.

10. Regarding Sophie's statement on multiple regression:
    (A) both statements are correct.
    (B) only Statement 1 is correct.
    (C) only Statement 2 is correct.

11. Based on the credit spread model, if an issuer gets included in the CDX index and assuming everything else the same, which of the following statements most accurately describes the model's forecast?
    (A) The credit spread on the firm's issue would decrease by 10 bps.
    (B) The credit spread on the firm's issue will decrease by 32 bps.
    (C) The credit spread on the firm's issue will increase by 32 bps.

12. Which of the following is least likely an assumption of multiple linear regression?
    (A) There is no linear relationship between the independent variables.
    (B) The dependent variable is not serially correlated.
    (C) The error term is normally distributed.

13. Which assumption of multiple regression is most likely evaluated using a QQ plot?
    (A) Serial correlation of residuals.
    (B) Error term is normally distributed.
    (C) Conditional heteroskedasticity.

14. The predicted price of a house that has 2,000 square feet of space and 4 bedrooms is closest to:
    (A) $114,000.
    (B) $256,000.
    (C) $185,000.

15. The conclusion from the hypothesis test of H_0: $b_1 = b_2 = 0$, is that the null hypothesis should:
    (A) not be rejected as the calculated F of 40.73 is greater than the critical value of 3.29.
    (B) be rejected as the calculated F of 40.73 is greater than the critical value of 3.33.
    (C) be rejected as the calculated F of 40.73 is greater than the critical value of 3.29.

16. Which of the following is most likely to present a problem in using this regression for forecasting?
    (A) Heteroskedasticity.
    (B) Multicollinearity.
    (C) Autocorrelation.

17. Which of the following is the most accurate interpretation of the slope coefficient for size? ARAR:
    (A) and index will change by 1.1% when the natural logarithm of assets under management changes by 1.0.
    (B) will change by 0.6% when the natural logarithm of assets under management changes by 1.0, holding index constant.
    (C) will change by 1.0% when the natural logarithm of assets under management changes by 0.6, holding index constant.

18. Which of the following is the estimated standard error of the regression coefficient for index?
    (A) 2.31.
    (B) 0.52.
    (C) 1.91.

19. Which of the following is the t-statistic for size?
    (A) 0.70.
    (B) 3.33.
    (C) 0.30

20. Which of the following is the estimated intercept for the regression?
    (A) –2.86.
    (B) –9.45.
    (C) –0.11.

21. Which of the following statements is most accurate regarding the significance of the regression parameters at a 5% level of significance?
    (A) The parameter estimates for the intercept are significantly different than zero. The slope coefficients for index and size are not significant.
    (B) All of the parameter estimates are significantly different than zero at the 5% level of significance.
    (C) The parameter estimates for the intercept and the independent variable size are significantly different than zero. The coefficient for index is not significant.

22. Which of the following is NOT a required assumption for multiple linear regression?
    (A) The error term is normally distributed.
    (B) The expected value of the error term is zero.
    (C) The error term is linearly related to the dependent variable.

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

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

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

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

(A) PI and INS only.
(B) TEEN only.
(C) PI only.

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

	Business Degree Dummy Variable	Engineering Degree Dummy Variable
(A)	1	1
(B)	0	1
(C)	0	0

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

(A) The adjusted-R^2 not appropriate to use in simple regression.
(B) The adjusted-R^2 is greater than the R^2 in multiple regression.
(C) It is possible for the adjusted-R^2 to decline as more variables are added to the multiple regression.

26. Which of the following is a potential remedy for multicollinearity?

(A) Add dummy variables to the regression.
(B) Omit one or more of the collinear variables.
(C) Take first differences of the dependent variable.

27. Salve runs a regression using the squared residuals from the model using the original dependent variables. The coefficient of determination of this model is 6%. Which of the following is the most appropriate conclusion at a 5% level of significance?

(A) 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.
(B) Because the test statistic of 7.20 is higher than the critical value of 3.84, we reject the null hypothesis of no conditional heteroskedasticity in residuals.
(C) Because the test statistic of 3.60 is lower than the critical value of 3.84, we reject the null hypothesis of no conditional heteroskedasticity in residuals.

28. Which of the following misspecifications is most likely to cause serial correlation in residuals?
    (A) Improper variable scaling.
    (B) Improper variable form.
    (C) Data improperly pooled.

29. Should Salve be concerned about residual serial correlation?
    (A) Yes, for two lags only.
    (B) No.
    (C) Yes, for one lag only.

30. Should Salve be concerned about residual multicollinearity?
    (A) Yes, and Salve should exclude either variable SMB or HML from the model.
    (B) Yes, and Salve should exclude variable Rm-Rf from the model.
    (C) No.

31. Which of the following conditions will least likely affect the statistical inference about regression parameters by itself?
    (A) Multicollinearity.
    (B) Model misspecification.
    (C) Unconditional heteroskedasticity.

32. One of the main assumptions of a multiple regression model is that the variance of the residuals is constant across all observations in the sample. A violation of the assumption is most likely to be described as:
    (A) unstable remnant deviation.
    (B) heteroskedasticity.
    (C) positive serial correlation.

33. Assume that in a particular multiple regression model, it is determined that the error terms are uncorrelated with each other. Which of the following statements is most accurate?
    (A) Serial correlation may be present in this multiple regression model, and can be confirmed only through a Durbin-Watson test.
    (B) This model is in accordance with the basic assumptions of multiple regression analysis because the errors are not serially correlated.
    (C) Unconditional heteroskedasticity present in this model should not pose a problem, but can be corrected by using robust standard errors.

34. Sutter has detected the presence of conditional heteroskedasticity in Smith's report. This is evidence that:
    (A) the error terms are correlated with each other.
    (B) the variance of the error term is correlated with the values of the independent variables.
    (C) two or more of the independent variables are highly correlated with each other.

35. Suppose there is evidence that the variance of the error term is correlated with the values of the independent variables. The most likely effect on the statistical inferences Smith can make from the regressions results using financial data is to commit a:
    (A) Type I error by incorrectly failing to reject the null hypothesis that the regression parameters are equal to zero.
    (B) Type II error by incorrectly failing to reject the null hypothesis that the regression parameters are equal to zero.
    (C) Type I error by incorrectly rejecting the null hypotheses that the regression parameters are equal to zero.

36. Which of the following is most likely to indicate that two or more of the independent variables or linear combinations of independent variables, may be highly correlated with each other? Unless otherwise noted, significant and insignificantly mean significantly different from zero and not significantly different from zero, respectively.
    (A) The R^2 is low, the F-statistic is insignificant and the Durbin-Watson statistic is significant.
    (B) The R^2 is high, the F-statistic is significant and the t-statistics on the individual slope coefficients are insignificant.
    (C) The R^2 is high, the F-statistic is significant and the t-statistics on the individual slope coefficients are significant.

37. Using the Durbin-Watson test statistic, Smith rejects the null hypothesis suggested by the test. This is evidence that:
    (A) two or more of the independent variables are highly correlated with each other.
    (B) the error term is normally distributed.
    (C) the error terms are correlated with each other.

38. Which model would be a better choice for making a forecast?
    (A) Model ONE because it has a higher R^2.
    (B) Model TWO because it has a higher adjusted R^2.
    (C) Model TWO because serial correlation is not a problem.

39. Using Model ONE, what is the sales forecast for the second quarter of the next year?
    (A) $56.02 million.
    (B) $51.09 million.
    (C) $46.31 million.

40. Which model misspecification is most likely to cause multicollinearity?
    (A) Inappropriate variable form.
    (B) Ommission of important variable(s).
    (C) Inappropriate variable scaling.

41. If it is determined that conditional heteroskedasticity is present in model one, which of the following inferences are most accurate?
    (A) Both the regression coefficients and the standard errors will be biased.
    (B) Regression coefficients will be unbiased but standard errors will be biased.
    (C) Regression coefficients will be biased but standard errors will be unbiased.

42. Mercado probably did not include a fourth dummy variable Q4, which would have had 0, 0, 0, 1 as its first four observations because:
    (A) it would not have been significant.
    (B) the intercept is essentially the dummy for the fourth quarter.
    (C) it would have lowered the explanatory power of the equation.

43. If Mercado determines that Model TWO is the appropriate specification, then he is essentially saying that for each year, value of sales from quarter three to four is expected to:
    (A) grow by more than $1,000,000.
    (B) remain approximately the same.
    (C) grow, but by less than $1,000,000.

44. The adjusted R^2 of Model 2 is closest to:
    (A) 0.36.
    (B) 0.37.
    (C) 0.39.

45. The model better suited for prediction is:
    (A) Model 1 because it has a lower Bayesian information criterion.
    (B) Model 2 because it has a lower Akaike information criterion.
    (C) Model 2 because it has a higher Akaike information criterion.

46. The F-statistic for testing H_0: coefficient of LIQ = 0 versus H_a: coefficient of LIQ # 0 is closest to:
    (A) 5.45.
    (B) 13.33.
    (C) 2.11.

47. What is the predicted return for a stock using Model 1 when SMB = 3.30, HML = 1.25 and Rm-Rf = 5?
    (A) 7.88%.
    (B) 9.58%.
    (C) 6.80%.

48 Which of the following statements least accurately describes one of the fundamental multiple regression assumptions?
    (A) The variance of the error terms is not constant (i.e., the errors are heteroskedastic.
    (B) The independent variables are not random.
    (C) The error term is normally distributed.

49 Henry Hilton, CFA, is undertaking an analysis of the bicycle industry. He hypothesizes that bicycle sales (SALES) are a function of three factors: the population under 20 (POP), the level of disposable income (INCOME), and the number of dollars spent on advertising (ADV). All data are measured in millions of units. Hilton gathers data for the last 20 years. Which of the follow regression equations correctly represents Hilton's hypothesis?
    (A) SALES = $\alpha + \beta_1$ POP + β_2 INCOME + β_3 ADV + ε.
    (B) SALES = $\alpha \times \beta_1$ POP $\times \beta_2$ INCOME $\times \beta_3$ ADV $\times \varepsilon$.
    (C) INCOME = $\alpha + \beta_1$ POP + β_2 SALES + β_3 ADV + ε.

50 One possible problem that could jeopardize the validity of the employment growth rate model is multicollinearity. Which of the following would most likely suggest the existence of multicollinearity?
    (A) The F-statistic suggests that the overall regression is significant, however the regression coefficients are not individually significant.
    (B) The variance of the observations has increased over time.
    (C) The Durbin—Watson statistic is significant.

51. Which of the following is least likely to be an assumption regarding linear regression?
    (A) The variance of the residuals is constant.
    (B) A linear relationship exists between the dependent and independent variables.
    (C) The independent variable is correlated with the residuals.

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

(A) 0.084, indicating that the variability of industry returns explains about 8.4% of the variability of company returns.

(B) 0.839, indicating that company returns explain about 83.9% of the variability of industry returns.

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

53. Based upon her analysis, Carter has derived the following regression equation: $\hat{Y} = 1.75 + 3.24X_1$.

The predicated value of the Y variable equals 50.50 if the:

(A) coefficient of the determination equals 15.

(B) predicated value of the dependent variable equals 15.

(C) predicated value of the independent variable equals 15.

54. Carter realize that although regression is a useful tool when analysing investments, there are certain limitations. Carter made a list of points describing limitations that Smith Brothers equality traders should be aware of when applying her research to their investment decision.

- Point 1: Regression residuals may be homoskedastic.
- Point 2: Data from regressions relationship tends to exhibit parameter instability.
- Point 3: Regression residuals may exhibit autocorrelation.
- Point 4: The variance of the error term may change with one or more independent variables.

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

(A) Points 1, 2, and 3.

(B) Points 1, 3, and 4.

(C) Points 2, 3, and 4.

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

(A) 66.76%

(B) 61.78%

(C) 81.71%

56. Suppose the Breusch-Godfrey statistic is 3.22. At a 5% level of significance, which of the following is the most accurate conclusion regarding the presence of serial correlation (at two lags) in the residuals?

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

(B) No, because the BG statistic is less than the critical test statistic of 3.49, we don't have evidence of serial correlation.

(C) Yes, because the BG statistic exceeds the critical test statistic of 3.16, there is evidence of serial correlation.

57. Gloucester subsequently revises the model to exclude the small cap index and finds that the revised model has a RSS of 106.332. Which of the following statements is most accurate? At a 5% level of significance, the test statistic.

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

(B) of 4.35 indicates that we cannot reject the hypothesis that the coefficient of small-cap index is significantly different from 0.

(C) of 13.39 indicates that we cannot reject the hypothesis that the coefficient of small-cap index is significantly different from 0.

58. The best test for unconditional heteroskedasticity is:

(A) the Breusch-Godfrey test only.

(B) the Breusch-Pagan test only.

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

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

(A) 2.322.

(B) 2.799.

(C) 2.561.

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

(A) serial correlation.

(B) multicollinearity.

(C) Conditional heteroskedasticity.

61. Consider the following analysis of variance (ANOVA) table:

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

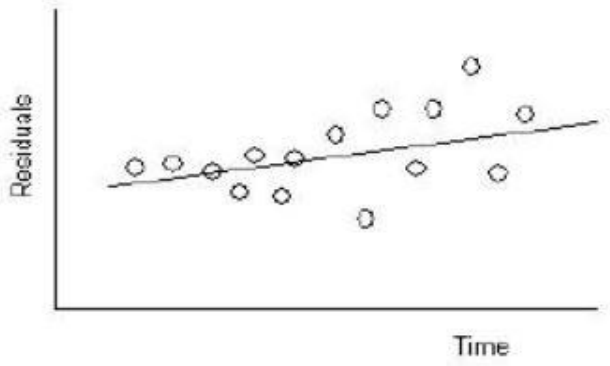
The F-statistic for the test of the fit of the model is closest to:

(A) 10.00.

(B) 0.10.

(C) 0.25.

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



These residuals exhibit the regression problem of:

(A) heteroskedasticity.

(B) autocorrelation.

(C) homoskedasticity.

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

(A) Scatter plot.

(B) Breusch-Pagan test.

(C) Breusch-Godfrey test.

64. When pooling the samples over multiple economic environments in a multiple regression model, which of the following errors is most likely to occur?

(A) Multicollinearity.

(B) Heteroskedasticity.

(C) Model misspecification.

65. Concerning the assumptions of multiple regression, Grimbles is:
   (A) incorrect to agree with Voiku's list of assumptions because one of the assumptions is stated incorrectly.
   (B) correct to agree with Voiku's list of assumptions.
   (C) incorrect to agree with Voiku's list of assumptions because two of the assumptions are stated incorrectly.

66. For which of the four hypotheses did Voiku incorrectly fail to reject the null, based on the data given in the problem?
   (A) Hypothesis 3.
   (B) Hypothesis 2.
   (C) Hypothesis 4.

67. The most appropriate decision with regard to the F-statistic for testing the null hypothesis that all of the independent variables are simultaneously equal to zero at the 5 percent significance level is to:
   (A) reject the null hypothesis because the F-statistic is larger than the critical F-value of 2.66.
   (B) fail to reject the null hypothesis because the F-statistic is smaller than the critical F-value of 2.66.
   (C) reject the null hypothesis because the F-statistic is larger than the critical F-value of 3.19.

68. Regarding Voiku's calculations of R^2 and the standard error of estimate, she is:
   (A) incorrect in her calculation of the unadjusted R^2 but correct in her calculation of the standard error of estimate.
   (B) incorrect in her calculation of both the unadjusted R^2 and the standard error of estimate.
   (C) correct in her calculation of the unadjusted R^2 but incorrect in her calculation of the standard error of estimate.

69. The multiple regressions, as specified, most likely suffers form:
   (A) heteroskedasticity.
   (B) serial correlation of the error terms.
   (C) multicollinearity.

70. A 90 percent confidence interval for the coefficient on GDP is:
   (A) 0.5 to 22.9.
   (B) –1.5 to 20.2.
   (C) –1.9 to 19.6.

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

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

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

The critical t-statistic for a 95% confidence level is 2.120. Which of the independent variables is statistically different from zero at the 95% confidence level?
(A) ADV only.
(B) INCOME only.
(C) INCOME and ADV.

73. An analyst runs a regression of monthly values-stock returns on five independent variables over 48 months. The total sum of squares is 430, and the sum of squared errors is 170. Test the null hypothesis at the 2.5% significance level that all five of the independent variables are equal to zero.
(A) Not rejected at 2.5% or 5.0% significance.
(B) Rejected at 2.5% significance and 5% significance.
(C) Rejected at 5% significance only.

74. An analyst is trying to estimate the beta for a fund. The analyst estimates a regression equation in which the fund returns are the dependent variable and the Wilshire 5000 is the independent variable, using monthly data over the past five years. The analyst finds that the correlation between the square of the residuals of the regression and the Wilshire 5000 is 0.2. Which of the following is most accurate, assuming a 0.05 level of significance? There is:
(A) evidence of level serial correlation but not conditional heteroskedasticity in the regression equation.
(B) evidence of conditional heteroskedasticity but not serial correlation in the regression equation.
(C) no evidence that there is conditional heteroskedasticity or serial correlation in the regression equation.

75. Which of the following statements regarding heteroskedasticity is least accurate?

(A) Conditional heteroskedasticity can be detected using the Breusch-Pagan chi-square statistic.

(B) When not related to independent variables, heteroskedasticity does not pose any major problems with the regression.

(C) Heteroskedasticity only occurs in cross-sectional regressions.

76. which of the following statements most accurately intercepts the following regression results at the given significance level?

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

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

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

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

77. In a one-side test and a 1% level significance, which of the following coefficients is significantly difference from zero?

(A) The intercept and the coefficient on ln (market value) only.

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

(C) The coefficient on ln(no. of Analysts) only.

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

(A) 0.011 to 0.001.

(B) –0.018 to –0.036.

(C) 0.014 to –0.009.

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

(A) –0.035.

(B) –0.055.

(C) –0.019.

80. Base on a R^2 calculated from the information in Table 2, the analyst should conclude that the number of analysts and ln (market value) of the firm explain:
(A) 15.6 of the variation in returns.
(B) 18.4 of the variation in returns.
(C) 84.4% of the variation in returns.

81. What is the F-statistic from the regression? And, what can be concluded from its value at a 1% level of significance?
(A) F = 1.97, fail to reject a hypothesis that both of the slope coefficient are equal to zero.
(B) F = 5.80, reject a hypothesis that both of the slope coefficients are equal to zero.
(C) F = 17.00, reject a hypothesis that both of the slope coefficients are equals to zero.

82. Upon further analysis, Turner concludes that multicollinearity is a problem. What might have prompted this further analysis and what is intuition the conclusion?
(A) At least one of the t-statistics was not significant, the F-statistic was not significant, and a positive relationship between the number of analysts and the size of the firm would be expected.
(B) At least one of the t-statistics was not significant, the F-statistic was not significant, and a positive relationship between the number of analysts and the size of the firm would be expected.
(C) At least one of the t-statistics was not significant, the F-statistic was significant, and an intercept not significantly different from zero would be expected.

83. When interpreting the results of a multiple regression analysis, which of the following terms represents the value of the dependent variable when the independent variables are all equal to zero?
(A) Intercept term.
(B) Slope coefficient.
(C) p-value.

84. Consider the following estimated regression equation, with the standard errors of the slope coefficient as noted:
$Sales_i = 10.0 + 1.25\ R\&D_i + 1.01\ ADV_i - 2.0\ COMP_i + 8.0\ CAP_i$
Where the standard error for the estimated coefficient on R&D is 0.45, the standard error for the estimated coefficient on ADV is 2.2, the standard error for the estimated coefficient on COMP is 0.63, and the standard error for the estimated coefficient on CAP is 2.5.
The equation was estimated over 40 companies. Using a 5% level of significance, which of the estimate coefficient are significantly different from zero?
(A) ADV and CAP only.
(B) R&D, COMP, and CAP only.
(C) R&D, ADV, COMP, and CAP.

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

(A) The Hansen method.

(B) The Breusch-Pagan test.

(C) The Durbin-Watson statistic.

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

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

Where:

EPS_t is a quarterly observation of earnings per shares

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

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

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

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

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

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

(C) Significance of the coefficient cannot be intercepted in the case of dummy variables.

87. Which of the following questions is least likely answered by using a qualitative dependent variables?

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

(B) Based on the following subsidiary and competition variables, will company XYZ divest itself of a subsidiary?

(C) Based on the following company-specific financial ratios, will company ABC enter bankruptcy?

88. A high-yield bond analyst is trying to develop an equation using financial ratios to estimate the probability of a company defaulting on its bonds. A technique that can be used to develop this equation is:

(A) Dummy variable regression.

(B) Logistic regression model.

(C) Multiple linear regression adjusting for heteroskedasticity.

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

$AUTO_t = 10.0 + 1.25\ PI_t + 1.01\ TEEN_t - 2.0\ INS_t$

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

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

(A) 14.10

(B) 17.50.

(C) 14.90.

90. Which of the following statements regarding heteroskedasticity is least accurate?

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

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

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

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

(A) conditional heteroskedasticity.

(B) multicollinearity.

(C) serial correlation.

92. Consider the following regression equation:

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

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

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

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

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

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

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

(Standard errors in parentheses):

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

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

(A) $509,980,000.

(B) $557,143,000.

(C) $656,991,000.

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

(A) $82.00

(B) $128.00.

(C) $206.00.

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

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

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

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

96. Presence of conditional heteroskedasticity is least likely to affect the:

(A) Computed F-statistic

(B) coefficient estimates.

(C) computed t-statistic.

97. An analyst is estimating whether company sales is related to three economic variables. The regression exhibits conditional heteroskedasticity, serial correlation, and multicollinearity. The analyst uses White and Newey-West standard errors. Which of the following is most accurate?

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

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

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

98. A regression with three independent variables have VIF values of 3, 4 and 2 for the first, second, and third independent variables, respectively. Which of the following conclusions is most appropriate?

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

(B) Only variable two has a problem with multicollinearity.

(C) Total VIF of 9 indicates a serious multicollinearity problem.

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

(A) 11.

(B) 13.

(C) 12.

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

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

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

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

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

(A) Transforming a variable.

(B) Inappropriate variable form.

(C) Omission of an important independent variable.

102. In regard to their conversion about the regression equation:
   (A) Brent's statement is correct; Johnson's statement is incorrect.
   (B) Brent's statement is incorrect; Johnson's statement is correct.
   (C) Brent's statement is correct; Johnson's statement is correct.

103. Using data from the past 20 quarters, Brent calculates the t-statistic for marketing expenditures to be 3.68 and the t-statistic for salespeople at 2.19. At a 5% significance level, the two-tailed critical values are t_c = +/– 2.127. This most likely indicated that:
   (A) the t-statistic has 18 degrees of freedom.
   (B) the null hypothesis should not be rejected.
   (C) both independent variables are statistically significant.

104. Brent calculated that the sum of squared errors (SSE) for the variables is 267. The means squared error (MSE) would be:
   (A) 14.055.
   (B) 15.706.
   (C) 17.831.

105. Brent is trying to explain the concept of the standard error of estimate (SEE) to Johnson. In this explanation, Brent makes three points about the SEE.
   - Point 1: The SEE is the standard deviation of the differences between the estimated values for the independent variable and the actual observations for the independent variable.
   - Point 2: Any violation of the basic assumptions of a multiple regressions model is going to affect the SEE.
   - Point 3. If there is a strong relationship between the variables and the SEE is small, the individual estimation errors will also be small.

   How many of Brent's points are most accurate?
   (A) 1 of Brent's points are correct.
   (B) All 3 of Brent's points are correct.
   (C) 2 of Brent's points are correct.

106. Assuming that next year's marketing expenditures are $3,500,000 and there are five salespeople, predicated sales for Mega Flowers should will be:
   (A) $24,000,000.
   (B) $11,600,000.
   (C) $24,200,000.

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

(A) The multiple coefficient of determination.

(B) The F-statistic.

(C) An ANOVA table.

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

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

The values of R^2 and the F-statistic for joint test of significance of all the slope coefficients are:

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

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

(C) $R^2 = 0.25$ and $F = 10$.

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

(A) $55 million above average.

(B) $35.2 million above average.

(C) $65 million above average.

110. Williams proceeds to test the hypothesis that none of the independent variables has significant explanatory power. Using the joint F-test for the significance of all slope coefficients, at a 5% level of significance:

(A) all of the independent variables have explanatory power.

(B) none of the independent variables has explanatory power.

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

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

(A) a Breusch-Godfrey test, but not a Breusch-Pagan test.

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

(C) a Breusch-pagan test, but not Breusch-Godfrey.

112. When Williams ran the model, the computer said the R^2 is 0.233. She examines the other output and conclusion that this is the:
   (A) neither the unadjusted nor adjusted R^2 value, nor the coefficient of correlation.
   (B) unadjusted R^2 value.
   (C) adjusted R^2 value.

113. In preparing and using this model, Williams has least likely relied on which of the following assumptions?
   (A) There is a linear relationship between the independent variables.
   (B) The disturbance or error term is normally distributed.
   (C) The residuals are homoscedastic.

114. Jason Fye, CFA, wants to check for seasonality in monthly stock returns (i.e., the January effect) after controlling for market cap and systematic risk. The type of model that Fye would most appropriately select is:
   (A) Multiple regression model.
   (B) logistic regression model.
   (C) Neither multiple regression nor logistic regression.

115. Using the regression model represented in Exhibit 1, What is the predicated number of housing starts for 20X7.
   (A) 1,394,420
   (B) 1,751,000
   (C) 1,394

116. Which of the following statements best describes the explanatory power of the estimated regression?
   (A) The residual standard error of only 0.3 indicated that the regression equation is a good fit for the sample data.
   (B) The large F-statistic indicates that both independent variables help explain changes in housing starts.
   (C) The independent variables explain 61.58% of the variation in housing starts.

117. Which of the following is the least appropriate statement in relation to R-square and adjusted R-square:
   (A) Adjusted R-square is a value between 0 and 1 can be interpreted as a percentage.
   (B) R-square typically increase when new independent variables are added to the regression regardless of their explanatory power.
   (C) Adjusted R-square decrease when the added independent variable adds little value to the regression model.

118. Which of the following statements regarding the results of a regression analysis is least accurate?

The:

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

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

(C) slope coefficient in the multiple regression are referred to as partial betas.

119. Using the regression model developed, the closest prediction of sales for December 20X6 is:

(A) $36,000

(B) $44,000.

(C) $35,000.

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

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

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

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

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

(A) yes, because 2.6 > 2.23.

(B no, because 2.6 < 2.62.

(C) yes, because 2.6 > 1.98.

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

(A) the joint significance of the independent variables.

(B) deviation of the estimated value from the actual values of the dependent variable.

(C) degree of correlation between the independent variables.

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

(A) 9.80.

(B) 67.00.

(C) 77.00.

124. In this multiple regression, if Stumper discovers that the residuals exhibit positive serial correlation, the most likely effect is:
   (A) standard errors are too low but coefficient estimate is consistent.
   (B) standard errors are too high but coefficient estimate is consistent.
   (C) standard errors are not affected but coefficient estimate is inconsistent.

125. Which of the following tests is least likely to be used to detect autocorrelation?
   (A) Durbin-Watson.
   (B) Breusch-Godfrey.
   (C) Breusch-Pagan.

126. One of the most popular ways to correct heteroskedasticity is to:
   (A) Improve the specification of the model.
   (B) adjust the standard errors.
   (C) use robust standard errors.

127. If a regression equation shown that no individual t-tests are significance, but the F-statistic is significance, the regression probably exhibits:
   (A) serial correlation.
   (B) multicollinearity.
   (C) heteroskedasticity.

128. Consider the following estimated regression equation, with standard errors of the coefficients as indicated:
   $Sales_i = 10.0 + 1.25\ R\&D_i + 1.0\ ADV_i - 2.0\ COMP_i + 8.0\ CAP_i$
   Where the standard error for R&D is 0.45, the standard error for ADV is 2.2, the standard error for COMP 0.63, and the standard error for CAP is 2.5.
   Sales are in millions of dollars. An analyst is given the following predictions on the independent variables: R&D = 5, ADV = 4, COMP = 10, and CAP = 40.
   The predicated level of sales is closest to:
   (A) $310.25 million.
   (B) $300.25 million.
   (C) $320.25 million.

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

(A) If the t-value of a variable is less than the significance level, the null hypothesis should be rejected.

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

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

130. What is most likely represented by the Y intercept of the regression?

(A) The intercept is not a driver of returns, only the independent variables.

(B) The drift of a random walk.

(C) The return on a particular trading day.

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

(A) The coefficient of determination for the above regression is significantly higher than the standard error of the estimate, and therefore there is value to calendar trading.

(B) There is no value to calendar trading.

(C) There is value to calendar trading.

132. The test mentioned by Jessica is known as the:

(A) Breusch-Pagan, which is a two-tailed test.

(B) Durbin-Watson, which is a two-tailed test.

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

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

(A) Both are correct.

(B) One is correct.

(C) Neither is correct.

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

(A) Positive serial correlations.

(B) Multicollinearity.

(C) Homoskedasticity.

135. Consider the following regression equation:

$Sales_i = 20.5 + 1.5\ R\&D_i + 2.5\ ADV_i - 3.0\ COMP_i$

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

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

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

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

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

136. Using the regression model developed, the closest predication sales for December 20X6 is:

(A) $44,000.

(B) $36,000.

(C) $55,000.

137. Will Jack conclude that the housing starts coefficient is statistically different from zero and how will he intercept it at the 5% significance level?

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

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

(C) Not different from zero; sales will rise by $0 for every 100 house starts.

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

(A) the joint significance of the independent variables.

(B) deviation of the estimated values from the actual values of the dependent variable.

(C) degree of correlation between the independent variables.

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

(A) 77.00.

(B) 9.80.

(C) 67.00.

140. For this question only, assume that the regression of squared residuals on the independent variables has R^2 = 11%. At a 5% level significance, which of the following conclusions is most accurate?

(A) Because the critical value is 3.84, we reject the null hypothesis of no conditional heteroskedasticity.

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

(C) With a test statistics of 0.22, we cannot reject the null hypothesis of no conditional heteroskedasticity.

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

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

The values of R^2 and the F-statistic to test the null hypothesis that slope coefficient on all variables are equal to zero are:

(A) R^2 = 0.20 and F = 13.333.

(B) R^2 = 0.25 and F = 13.333.

(C) R^2 = 0.25 and F = 0.930.

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

(A) one dummy variables.

(B) four dummy variables.

(C) three dummy variables.

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

(A) All there are statistically significant.

(B) Two of the three are statistically significant.

(C) One of the three is statistically significant.

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

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

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

(C) accepted because the t-statistic is less than 2.617.

145. The percentage of the total variation in quarterly stock returns explained by the independent variables closest to:
   (A) 32%.
   (B) 47%.
   (C) 42%.

146. According to the Durbin-Watson statistic, there is:
   (A) significant positive serial correlation in the residuals.
   (B) significant heteroskedasticity in the residuals.
   (C) no significant positive serial correlation in the residuals.

147. What is predicated quarterly stock return, given the following forecasts?
   - Employment growth = 2.0%
   - GDP growth = 1.0%
   - Private investment growth = –1.0%

   (A) 4.4%.
   (B) 5.0%.
   (C) 4.7%.

148. What is the standard error of the estimate?
   (A) 0.81.
   (B) 1.71.
   (C) 1.31.

149. Using the regression model represented in Exhibit 1, what is the predicated number of housing starts for 20X7?
   (A) 1,394,420.
   (B) 1,751,000.
   (C) 1,394.

150. Which of the following statements best describes the explanatory power of the estimated regression?
   (A) The independent variables explain 61,58% of the variation in housing starts.
   (B) The large F-statistic indicates that both independent variable help explain changes in housing starts.
   (C) The residual standard error of only 0.3 indicates that the regression is a good fit for the sample data.

151. Which of the following is the least appropriate statement in relation to R-square and adjusted R-square?
   (A) Adjusted R-square decreases when the added independent variable adds little value to the regression model.
   (B) R-square typically increases when new independent variables adds to the regression.
   (C) Adjusted R-square can be higher than the coefficient of determination for a model with a good fit.

152. What is the correct interpretation of the coefficient of closed in the first regression?
   (A) If a model is closed to new investors, the expected excess fund return is 1.65%.
   (B) A closed fund is estimated to have an extra returns of 1.65% relative to funds that are not closed.
   (C) A closed fund is likely to generate a return of 1.65%.

153. To check for only the outliers in the sample, Lee should most appropriately use:
   (A) leverage.
   (B) Cook's D.
   (C) Studentized residuals.

154. Which observations, when excluded, cause a significance change to model coefficients?
   (A) Observation 10 and 19.
   (B) Observation 1, 10, and 11.
   (C) Observation19.

155. What is the change probability of fund closure for a 1% increase in Ln(assets under management)?
   (A) 5.08%
   (B) 2.33%.
   (C) 4.83%.

156. Which of the following statement regarding serial correlation that might be encountered in regression analysis is least accurate?
   (A) Serial correlation does not affect consistency of regression coefficients.
   (B) Positive serial correlation and heteroskedasticity can both lead to Type I errors.
   (C) Serial correlation occurs least often with the time series data.

157. Which of the following is least likely a method of detecting serial correlation?

(A) The Breusch-Godfrey test.

(B) A scatter plot of the residuals over time.

(C) The Breusch-Pagan test.

158. A multiple regression model has included independent variables that are not linearly related to the dependent variable. The model is most likely misspecificed due to:

(A) incorrect data pooling.

(B) incorrect variable form.

(C) incorrect variable scaling.