MULTIPLE REGRESSION

1

1. 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.

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)   (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

2. 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.

3. 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%.

4. 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%.

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 × DSC − 0.32 × index + 1.33 × 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

5. Regarding Sophie's statement on multiple regression:

(A) both statements are correct.

(B) only Statement 1 is correct.

(C) only Statement 2 is correct.

6. 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.

7. 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.

8. 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.

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
tercept	66,500	59,292	1.12
ouse Size	74.30	21.11	3.52
umber 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.

9. 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.

10. 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.

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

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

12. 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.

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

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

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

16. 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.

17. 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.

18. 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.

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

20. 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.

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.

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. 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.

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. 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.

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.

del ONE	
Regression Statistics	
ltiple R	0.941828
	0.887039
justed R_2	0.863258
ndard Error	2.543272
servations	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
ercept	31.40833	1.4866	21.12763
	−3.77798	1.485952	−2.54246
	−2.46310	1.476204	−1.66853
	−0.14821	1.470324	−0.10080
END	0.851786	0.075335	11.20848

Model ONE	
Regression Statistics	
ltiple R	0.941796
	0.886979
justed R^2	0.870026
ndard Error	2.479538
servations	24

Durbin-Watson test statistic = statistic = 0.7860

	df	SS	MS	F	Significance F
gression	3	964.9962	21.6654	52.3194	1.19E–09

sidual	20	122.9622	6.14811		
tal	23	1087.9584			

	Coefficients	Standard Error	t-statistics
ercept	31.32888	1.228865	25.49416
	−3.70288	1.253493	−2.95405
	−2.38839	1.244727	−1.91881
END	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. 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.

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
R-Squared	−0.12	0.02

LIQ		0.39
SSE		34.00
AIC		−141.34
BIC		−127.40

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 = α + β_1 POP + β_2 INCOME + β_3 ADV + ε.

(B)   SALES = α × β_1 POP × β_2 INCOME × β_3 ADV × ε.

(C)   INCOME = α + β_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.

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	SS	Mean Square

	(Degree of Freedom)	(Sum of Squares)	(SS/df)
Regression	1	3,257 (RSS)	3,257 (MSR)
Error	8	298 (SSE)	37.25 (MSE)
Total	**9**	**3,555 (SS Total)**	

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) coeffieient 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.

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. 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:
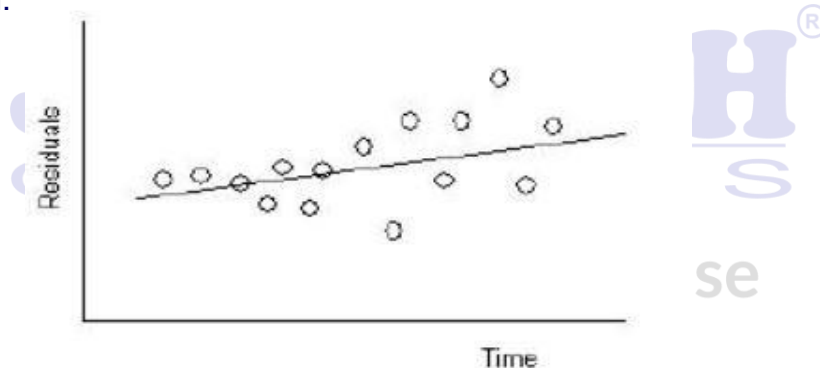
These residuals exhibit the regression problem of:
    (A) heteroskedasticity.
    (B) autocorrelation.
    (C) homoskedasticity.

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.

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) + ε_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 $\times$ CPI)	+ (5.2 $\times$ IP)	+ (11.7 $\times$ 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. 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.

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. 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 In (market value) only.

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

(C) The coefficient on In(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 In (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-Wastson 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 earnnings 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.

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. 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 exceeds 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.

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. 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.

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. 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.

    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. 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.

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. 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 estimates is consistent.
   (B) standard errors are too high but coefficient estimate is consistent.
   (C) standard errors are not affected but coefficient estimate is inconsistent.

   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. 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.

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. 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.   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.

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

132. 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%.

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

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

135. 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%.

136. 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.

137. 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.

138. A multiple regression model has included independent variables that are not linearly related to the dependent variable. The model is most likely misspecified due to:
    (A) incorrect data pooling.
    (B) incorrect variable form.
    (C) incorrect variable scaling.

139. 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.

140. 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.

141. 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.

142. 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.

143. 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.

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

144. Using the regression model developed, the closest predication sales for December 20X6 is:
   (A) $44,000.
   (B) $36,000.
   (C) $55,000.

145. 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.

146. 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.

147. 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.

148. 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.

149. 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.

150. 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.

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

J.K. SHAH®
C L A S S E S
a Veranda Enterprise

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 ($\alpha = 0.05$)									
	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

151. 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.

152. 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.

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

154. 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.

155. 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%.

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

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		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	1.27			

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%

157. 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.

158. 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.