Questions 1 – 6 deal with the following learning outcomes:

LO.a: Formulate a multiple regression equation to describe the relation between a dependent variable and several independent variables and determine the statistical significance of each independent variable.

LO.b: Interpret estimated regression coefficients and their p-values.

LO.c: Formulate a null and an alternative hypothesis about the population value of a regression coefficient, calculate the value of the test statistic, and determine whether to reject the null hypothesis at a given level of significance.

LO.d: Interpret the results of hypothesis tests of regression coefficients.

LO.e: Calculate and interpret 1) a confidence interval for the population value of a regression coefficient and 2) a predicted value for the dependent variable, given an estimated regression model and assumed values for the independent variables.

LO.g: Calculate and interpret the F-statistic, and describe how it is used in regression analysis.

LO.i: Evaluate how well a regression model explains the dependent variable by analyzing the output of the regression equation and an ANOVA table.

LO.o: Evaluate and interpret a multiple regression model and its results.
Use the following information to answer Questions 1 to 7.

An analyst obtained the following regression results:

	Coefficient	**Standard Error**	**t-statistics**
Intercept (b_0)	240.33	48.59	4.95
b_1	1.39	0.19	7.07
b_2	-3.65	9.68	-0.38

ANOVA	**df**	**SS**
Regression	2	3,940.29
Residual	97	4,268.19
Total	99	8,208.47

Regression equation: $Y_1 = b_0 + b_1X1i + b_2X_{2i} + e_i$
(For p = 0.05, the critical t-values with 97 degrees of freedom is 1.660. For p = 0.025, the critical t-values with 97 degrees of freedom is 1.984)

1. The analyst wants to test whether there is a positive relationship between X_1 and Y. Given a 5% level of significance, the hypothesis used and the result of the hypothesis test are *most likely*:

	Hypothesis	Conclusion
A.	H_0: $b_1 \geq 0$, versus H_a: $b_1 < 0$	Reject the null hypothesis and conclude that there is not a positive relationship between X_1 and Y.
B.	H_0: $b_1 \leq 0$, versus H_a: $b_1 > 0$	Reject the null hypothesis and conclude that there is a positive relationship between X_1 and Y.
C.	H_0: $b_1 \geq 0$, versus H_a: $b_1 < 0$	Fail to reject the null hypothesis and conclude that there is a positive relationship between X_1 and Y.

2. The analyst wants to test whether there is a negative relationship between X_2 and Y. Given a 5% significance level, the hypothesis used and the result of the hypothesis test are *most likely*:

	Hypothesis	Conclusion
A.	H_0: $b_2 \geq 0$, versus H_a: $b_2 < 0$	Reject the null hypothesis and conclude that there is a negative relationship between X_2 and Y.
B.	H_0: $b_2 \leq 0$, versus H_a: $b_2 > 0$	Reject the null hypothesis and conclude that the relationship between X_2 and Y is not negative.
C.	H_0: $b_2 \geq 0$, versus H_a: $b_2 < 0$	Fail to reject the null hypothesis and conclude that the relationship between X_2 and Y is not negative.

3. The analyst wants to test whether b_0 is different from zero. Given a 5% significance level, the hypothesis used and the result of the hypothesis test are *most likely*:

	Hypothesis	Conclusion
A.	H_0: $b_0 = 0$, versus H_a: $b_0 \neq 0$	Reject the null hypothesis and conclude that b_0 is not equal to zero.
B.	H_0: $b_0 \neq 0$, versus H_a: $b_0 = 0$	Reject the null hypothesis and conclude that b_0 is equal to zero.
C.	H_0: $b_0 = 0$, versus H_a: $b_0 \neq 0$	Fail to reject the null hypothesis and conclude that b_0 is equal to zero.

4. The F-stat is *closest* to:
   A. 0.22.
   B. 525.02.
   C. 44.77.

5. Given that X_1 equals 2.42 and X_2 equals 1.75, the predicted value of Y according to the regression equation is closest to:
   A. 250.08
   B. 237.31
   C. 236.38

6. The 95% confidence interval for the slope coefficient b_2 is closest to:
   A. -19.73 to 12.43
   B. -15.57 to 22.86
   C. -22.87 to 15.57

LO.f: Explain the assumptions of a multiple regression model.

7. Which of the following is *least likely* an assumption of the classic normal multiple linear regression model?
   A. A linear relationship exists between the dependent variable and the independent variable.
   B. The independent variables are random.
   C. No exact linear relation exists between two or more of the independent variables.

LO.h: Distinguish between and interpret the R^2 and adjusted R^2 in multiple regression.

8. Analyst 1: R^2 is a more reliable as a measure of goodness of fit in a regression with more than one independent variable than in a one independent- variable regression.
   Analyst 2: Adjusted R^2 does not necessarily increase when one adds an independent variable.
   A. Analyst 1 is correct.
   B. Analyst 2 is correct.
   C. Both analysts are correct.

LO.j: Formulate a multiple regression equation by using dummy variables to represent qualitative factors and interpret the coefficients and regression results.

9. An analyst is concerned about the seasonality in stock returns and wants to test whether stock returns differ during the different quarters of the year. How many dummy variables will be need in his regression equation?
   A. 3.
   B. 4.
   C. 5.

10. An analyst constructs a regression model to test whether stock returns differ during the different quarters of the year. He obtains the following regression output.

ANOVA	df	SS	MSS	F	Significance F
Regression	3	0.024	0.008	0.879	0.1526
Residual	96	0.876	0.0091		
Total	99	0.9			

At 5% level of significance, which of the following conclusions is *most accurate*?
   A. The quarters of the year effect is significant for explaining stock returns.
   B. The quarter of the year effect is not significant for explaining stock returns.
   C. Stock returns increase during the first quarter of the year and decrease during the last quarter.

LO.k: Explain the types of heteroskedasticity and how heteroskedasticity and serial correlation affect statistical inference.

11. Which of the following is the preferred approach to correct for heteroskedasticity?

A. Using robust standard errors.
B. Using generalized least squares.
C. Using Hansen method.

12. Which test is used to detect serial correlation?
    A. Breuch-Pagan test.
    B. Hansen test.
    C. Durbin-Watson test.

LO.l: Describe multicollinearity and explain its causes and effects in regression analysis.

13. Which of the following violations of regression assumptions will most likely increase the chances of making Type – II errors?
    A. Positive serial correlation.
    B. Conditional hetroskedasticity.
    C. Multicollinearity.

LO.m: Describe how model misspecification affects the results of a regression analysis and describe how to avoid common forms of misspecification.

14. Analyst 1: If one or more important variables are omitted from regression, it will lead to model misspecification.
    Analyst 2: If a function of a dependent variable is included as an independent variable, it will lead to model misspecification.
    A. Analyst 1 is correct.
    B. Analyst 2 is correct.
    C. Both analysts are correct.

LO.n: Describe models with qualitative dependent variables.

15. An analyst wants to predict whether a company will go bankrupt based on its debt-to-equity ratio and its interest coverage ratio. Which of the following models should *least likely* be used for this analysis?
    A. Discriminant analysis.
    B. Multiple regression with dummy variables.
    C. Probit model.

Solutions:

1. B is correct. The null hypothesis is the position the analyst is looking to reject. The alternate hypothesis is the position he is looking to validate. Therefore:
   H_0: $b_1 \leq 0$, versus H_a: $b_1 > 0$
   t-stat = 7.07
   Given a 5% significance level, the critical t-value with 97 degrees of freedom is 1.660
   This is a one-tailed test. Since the test-stat is greater than the critical t-value, the analyst can reject the null hypothesis and conclude that there is a positive relationship between X_1 and Y.

2. C is correct. The null hypothesis is the position the analyst is looking to reject. The alternate hypothesis is the position he is looking to validate. Therefore:
   H_0: $b_2 \geq 0$, versus H_a: $b_2 < 0$
   t-stat = -0.38
   Given a 5% significance level, the critical t-value with 97 degrees of freedom is -1.660
   This is a one-tailed test. Since the t-stat is not less than the negative critical t-value – 1.660, the analyst cannot reject the null hypothesis. He concludes that the relationship between X_2 and Y is not negative.

3. A is correct. The hypothesis will be structured as:
   H_0: $b_0 = 0$, versus H_a: $b_0 \neq 0$
   t-stat = 4.95
   This is a two-tailed test. Given a 5% significance level, the critical t-values with 97 degrees of freedom are + 1.984 and -1.984. Since the test-stat is greater than the upper critical value, the analyst can reject the null hypothesis and conclude that b_0 is significantly different from zero.

4. C is correct.
   F-stat = MSR/MSE = (RSS/k)/[SSE/(n-k-1)] = (3940.29/2)[4268.19/(100-2-1)] = 44.77

5. B is correct.
   Regression equation: $Y_1 = b_0 + b_1X1i + b_2X_{2i} + e_i$
   $Y_1 = 240.33 + (1.39 \times 2.42) + (-3.65 \times 1.75) = 237.31$

6. C is correct.
   Confidence interval = Estimated value ± (Critical t-value x Standard Error)
   Given 97 degrees of freedom, the absolute value of the critical t-value for a 95% confidence interval equals 1.984. Therefore:
   Confidence interval = -3.65 ± (1.984 x 9.69)
   Confidence interval = -22.87 to 15.57

7. B is correct. The assumptions of classical normal multiple linear regression model are as follows:
   1. A linear relation exists between the dependent variable and the independent variables.
   2. The independent variables are not random. Also, no exact linear relation exists between two or more of the independent variables.

3. The expected value of the error term, conditioned on the independent variables, is 0.
4. The variance of the error term is the same for all observations.
5. The error term is uncorrelated across observations.
6. The error term is normally distributed.

8. B is correct. R2 is non decreasing in the number of independent variables, so it is less reliable as a measure of goodness of fit in a regression with more than one independent variable than in a one independent- variable regression.

9. A is correct. For n states of the world, we need n-1 dummy variables. Therefore for 4 quarters we need 3 dummy variables.

10. B is correct. The p-value of 0.1526 means that the smallest level of significance at which we can conclude that – 'The quarters of the year effect is significant for explaining stock returns' is 15.26%. Hence at 5% level of significance we cannot conclude that the quarter of the year effect is significant for explaining stock returns. Hence, Option B is correct.

11. A is correct. The preferred approach to correct heteroskedasticity is to use robust standard errors.

12. C is correct. The Durbin-Watson test is used to detect serial correlation.

13. C is correct. In multicollinearity the t-stats of coefficients are artificially small thereby increasing the chances of a Type – II error. In serial correlation and heteroskedasticity the t-stats are artificially high thereby increasing the chances of a Type – I error.

14. C is correct. Both statements are accurate.

15. B is correct. The analyst needs to use a qualitative dependent variable. This dummy variable may be given a value of 1 if the company is bankrupt and 0 if the company is not bankrupt. We can use probit/ logit models and discriminant analysis for qualitative dependent variables. Linear regression is not the right technique here.