Final Exam - First Semester
Regression Analysis (Math413)
4th Year Stat. & Comp. Sci.
Deptartment of Mathematics

Helwan University
Faculty of Science

Exam. Date: January 2020
Time allawed: 90 minutes
Dr. Amany E. Aly
Total mark: Seventy Marks

Part I: Choose the correct answer (12 Marks)

1. Which one of the following is **NOT** appropriate for studying the relationship between two quantitative variables?

   (a) Scatterplot   (b) Bar chart
   (c) Correlation   (d) Regression

2. Which of the following can **NOT** be answered from a regression equation?

   (a) Predict the value of y at a particular value of x.
   (b) Estimate the slope between y and x.
   (c) Estimate whether the linear association is positive or negative.
   (d) Estimate whether the association is linear or non-linear

3. A scatterplot is a

   (a) one-dimensional graph of randomly scattered data.
   (b) two-dimensional graph of a straight line.
   (c) two-dimensional graph of a curved line.
   (d) two-dimensional graph of data values.

   For the next 4 questions :
   The simple linear regression equation can be written as $\widehat{Y} = \beta_0 + \beta_1 X$

4. In the simple linear regression equation, the symbol $\widehat{Y}$ represents the

   (a) average or predicted response   (b) estimated intercept
   (c) estimated slope   (d) explanatory variable

5. In the simple linear regression equation, the term β_0 represents the

   (a) estimated or predicted response   (b) estimated intercept
   (c) estimated slope   (d) explanatory variable

6. In the simple linear regression equation, the term β_1 represents the

   (a) estimated or predicted response   (b) estimated intercept
   (c) estimated slope   (d) explanatory variable

7. In the simple linear regression equation, the symbol X represents the

   (a) estimated or predicted response   (b) estimated intercept
   (c) estimated slope   (d) explanatory variable

8. What is the effect of an outlier on the value of a correlation coefficient?

    (a) An outlier will always decrease a correlation coefficient.

    (b) An outlier will always increase a correlation coefficient.

    (c) An outlier might either decrease or increase a correlation coefficient, depending on where it is in relation to the other points.

    (d) An outlier will have no effect on a correlation coefficient.

9. A scatter plot and regression line can be used for all of the following EXCEPT

    (a) to determine if any (x,y) pairs are outliers.

    (b) to predict y at a specific value of x..

    (c) to estimate the average y at a specific value of x.

    (d) to determine if a change in x causes a change in y.

10. The smaller the p-value, the

    (a) stronger the evidence against the alternative hypothesis

    (b) stronger the evidence for the null hypothesis

    (c) stronger the evidence against the null hypothesis

    (d) none of the above

For the next 2 questions :
A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation: $Y = 10.9 + 0.23\,X$

11. One student in the sample was 73 inches tall with a foot length of 29 cm. What is the predicted foot length for this student?

    (a) 17.57 cm   (b) 27.69 cm

    (c) 29 cm   (d) 33 cm

12. One student in the sample was 73 inches tall with a foot length of 29 cm. What is the residual for this student?

    (a) 29 cm   (b) 1.31 cm

    (c) 0.00 cm   (d) −1.31 cm

Part II: Answer the following questions

Question 1/ 21 Marks

A student is investigating the relationship between the price (y pence) of 100 g of chocolate and the percentage ($x\%$) of cocoa solids in chocolate. Thee following data is obtained

X (% cocoa)	10	20	30	35	40	50	60	70
Y (pence)	35.	55	40	100.	60	90.	110.	130

Given that:

$\sum x_i = 315, \sum y_i = 620, \sum x_i^2 = 15225, \sum y_i^2 = 56550, \sum x_i y_i = 28750, S_{xx} = 2821.88, S_{xy} = 4337.5,$
$S_{yy} = 8500.$

Final Exam
Regression Analysis (Math413)
4th Year Stat. & Comp. Sci.
Deptartment of Mathematics

Helwan University
Faculty of Science

Exam. Date: February 2021
Time allawed: 90 minutes
Dr. Amany E. Aly
Total mark: Seventy Marks

M

Part I: Choose the correct answer (30 Marks)

1. What is the effect of an outlier on the value of a correlation coefficient?

   (a) An outlier will always decrease a correlation coefficient.

   (b) An outlier will always increase a correlation coefficient.

   (c) An outlier might either decrease or increase a correlation coefficient, depending on where it is in relation to the other points.

   (d) An outlier will have no effect on a correlation coefficient.

2. A scatter plot and regression line can be used for all of the following **EXCEPT**

   (a) to determine if any (x,y) pairs are outliers.

   (b) to predict y at a specific value of x.

   (c) to estimate the average y at a specific value of x.

   (d) to determine if a change in x causes a change in y.

3. The smaller the p-value, the

   (a) stronger the evidence against the alternative hypothesis

   (b) stronger the evidence for the null hypothesis

   (c) stronger the evidence against the null hypothesis

   (d) none of the above

4. A regression equation was computed to be $Y = 35 + 6X$. The value of 35 indicates that

   (a) An increase in one unit of X will result in an increase of 35 in Y

   (b) The coefficient of correlation is 35

   (c) The coefficient of determination is 35

   (d) The regression line crosses the Y-axis at 35

5. Which of the following can **NOT** be answered from a regression equation?

   (a) Predict the value of y at a particular value of x.

   (b) Estimate the slope between y and x.

   (c) Estimate whether the linear association is positive or negative.

   (d) Estimate whether the association is linear or non-linear.

6. $e_i = Y_i - \widehat{Y}_i$ is called:

    (a) Difference between independent and dependent variables   (b) Residual

    (c) Difference between slope and intercept   (d) Sum of residual

7. Null plot means a scatter plot with what characteristic(s)?

    (a) mean zero and constant variance

    (b) constant variance and no separated point

    (c) mean zero, constant variance and no separated point

    (d) mean zero, constant variance, no separated point and several leverage points

8. When regression line passes through the origin, then:

    (a) Intercept is zero   (b) Regression coefficient is zero

    (c) Correlation is zero   (d) Association is zero

9. The independent variable in a regression line is:

    (a) Non-random   (b) Random   (c) Qualitative   (d) None of the above

10. The problem of determining the height of a person when his weight is given is:

    (a) Correlation   (b) Association   (c) Regression   (d) Qualitative

11. Given $(x_1, y_1), ..., (x_n, y_n)$ best fitting data to $y = f(x)$ by least squares requires minimization of

    (a) $\sum_{i=1}^{n} [y_i - f(x_i)]$ (b) $\sum_{i=1}^{n} |y_i - f(x_i)|$ (c) $\sum_{i=1}^{n} [y_i - f(x_i)]^2$ (d) $\sum_{i=1}^{n} [y_i - \overline{y}]^2$

12. Which of the following is the unbiased estimation of σ^2 as in the simple linear regression model?

    (a) $\dfrac{\sum \widehat{e}_i^2}{n-2}$ (b) $\dfrac{RSS}{n}$ (c) $\dfrac{S_{XX} - \dfrac{S_{XY}^2}{S_{XX}}}{n-2}$ (d) $\dfrac{S_{XX} - \dfrac{S_{XX}^2}{S_{XY}}}{n-2}$

13. Which is correct about the RSS in simple linear regression?

    (a) $RSS = S_{YY} - \dfrac{S_{XY}^2}{S_{XX}}$ (b) $RSS = S_{XX} - \dfrac{S_{XY}^2}{S_{YY}}$ (c) $RSS = n\,\widehat{\sigma}^2$ (d) $RSS = (n-1)\,\widehat{\sigma}^2$

For the next 4 questions :

The simple linear regression equation can be written as $\widehat{Y} = \beta_0 + \beta_1 X$

14. In the simple linear regression equation, the symbol $\widehat{Y}$ represents the

    (a) predicted response   (b) estimated intercept   (c) estimated slope   (d) explanatory variable

15. In the simple linear regression equation, the term β_0 represents the

    (a) predicted response   (b) estimated intercept   (c) estimated slope   (d) explanatory variable

16. In the simple linear regression equation, the term β_1 represents the

    (a) predicted response   (b) estimated intercept   (c) estimated slope   (d) explanatory variable

17. In the simple linear regression equation, the symbol X represents the

    (a) predicted response   (b) estimated intercept   (c) estimated slope   (d) explanatory variable

18. The regression equation for predicting number of speeding tickets (Y) from information about driver age (X) is $Y = -0.065(X) + 5.57$. How many tickets would you predict for a 20-year-old?

    (a) 6   (b) 4.27   (c) 5.57   (d) 1

19. Which one of the following is **NOT** appropriate for studying the relationship between two quantitative variables?

    (a) Scatterplot   (b) Bar chart   (c) Correlation   (d) Regression

20. In simple linear regression, the numbers of unknown parameters are:

    (a) One   (b) Two   (c) Three   (d) Four

21. The graph showing the paired points of (X_i, Y_i) is called:

    (a) Scatter diagram   (b) Histogram   (c) Historigram   (d) Pie diagram

22. The range of regression coefficients is:

    (a) -1 to 1 (b) 0 to 1 (c) $-\infty$ to ∞ (d) 0 to ∞

23. In simple linear regression model $Y = \beta_0 + \beta_1 X + \epsilon$ where β_0 and β_1 are called:

    (a) Estimates   (b) Parameters   (c) Random errors   (d) Variables

24. A regression model may be:

    (a) Linear   (b) Non-linear   (c) Both (a) and (b)   (d) Neither(a) nor (b)

25. A process by which we estimate the value of dependent variable on the basis of one or more independent variables is called:

    (a) Correlation   (b) Regression   (c) Residual   (d) Slope

26. The method of least squares dictates that we choose a regression line where the sum of the square of deviations of the points from the line is:

    (a) Maximum   (b) Minimum   (c) Zero   (d) Positive

27. All data points falling along a straight line is called:

    (a) Linear relationship   (b) Non linear relationship   (c) Residual   (d) Scatter diagram

28. The regression equation always passes through:

    (a) (X, Y) (b) (a, b) (c) $(\overline{X}, \overline{Y})$ (d) $(\overline{X}, Y)$

29. In the regression line $Y = \beta_0 + \beta_1 X$:

    (a) $\sum X = \sum \widehat{X}$ (b) $\sum Y = \sum \widehat{Y}$ (c) $\sum X = \sum Y$ (d) $X = Y$

30. The sum of the difference between the actual values of Y and its values obtained from the fitted regression line is always:

    (a) Zero   (b) Positive   (c) Negative   (d) Minimum

31. $E(Y - \widehat{Y})$ will be:

    (a) < 0 (b) > 0 (c) $= 0$ (d) $\neq 0$

Mid. Term
Module: Regression Analysis (Math413)
Grade: 4ᵗʰ Year Stat. & Comp. Sci
Deptartment of Mathematics

Helwan University
Faculty of Science

Exam. Date: 19. Nov. 2018
Time allawed: One hour
Instructor: Dr.Amany E. Aly
Total mark: fifteen Marks

Answer the following Questions:

Question 1: 7 Marks

Circle the correct answer.

1. The regression equation for predicting number of speeding tickets (Y) from information about driver age (X) is $Y = -0.065(X) + 5.57$. How many tickets would you predict for a twenty-year-old?

   (a) 6

   (b) 4.27

   (c) 5.57

   (d) 1

2. A regression equation was computed to be $Y = 35 + 6X$. The value of 35 indicates that

   (a) An increase in one unit of X will result in an increase of 35 in Y

   (b) The coefficient of correlation is 35

   (c) The coefficient of determination is 35

   (d) The regression line crosses the Y-axis at 35

3. Null plot means a scatter plot with what characteristic(s)?

   (a) mean zero and constant variance

   (b) constant variance and no separated point

   (c) mean zero, constant variance and no separated point

   (d) mean zero, constant variance, no separated point and several leverage points

4. CHAN Tai Man collected a data set of size $n = 5$, namely, $\{x_i, y_i\}, i \in \{1, 2, 3, 4, 5\}$, where y_i denotes the ith student's GPA while x_i denote his/her I.Q. score. CHAN is a student of Professor YAU, and asked him for help. Prof. YAU suggested that a simple linear regression model may be applied i.e. $y_i = \beta_0 + \beta_1 x_i + e_i$. You are given that $\{\hat{e}_1, ..., \hat{e}_4\} = \{0.1, -0.2, 0.3, -0.1\}$ and $S_{YY} = 3.3$ Then the value of R^2 is.

   (a) 0.9545

   (b) 0.9515

   (c) 0.0485

   (d) 0.9538

5. Which is correct about the RSS in simple linear regression?

   (a) $RSS = S_{YY} - \dfrac{S_{XY}^2}{S_{XX}}$

   (b) $RSS = S_{XX} - \dfrac{S_{XY}^2}{S_{YY}}$

   (c) $RSS = n\,\hat{\sigma}^2$

   (d) $RSS = (n-1)\,\hat{\sigma}^2$

6. Which of the following is a stronger correlation than -0.54?

   (a) 0

   (b) -0.1

   (c) -1

   (d) 0.5