King Fahd University of Petroleum and Minerals
Mathematics Department

Dhahran, Saudi Arabia

STAT 310: Regression Analysis

T231

Exam 1

October 2nd , 2023

Name:_____

ID:_____

Score:

Q1	/10	Q3	/60	Total	/100
Q2	/30				

Results you need to solve question 1:

I. If X is random variable, then: $E(X) = \mu_X$ $\quad Var(X) = E(X^2) - \big(E(X)\big)^2 = \sigma_X^2$

II. If Z is a random variable and is given by: $Z = aX + b$ then: $E(Z) = E(aX + b) = aE(X) + b = a\mu_X + b$, and $Var(Z) = Var(aX + b) = a^2 Var(X) = a^2\sigma_X^2$

III. If U is a random variable, then: $Var(X \pm U) = Var(X) + Var(U) \pm Cov(X, U)$, where $Cov(X, U) = E(XU) - E(X)E(U)$

IV. If $X\ and\ U$ are Independent random variables, then: $Var(X \pm U) = Var(X) + Var(U)$, where $Cov(X, U) = 0$

V. As was derived in class: $E(\hat{\beta}_o) = \beta_o$ $\quad$ and $\quad E(\hat{\beta}_1) = \beta_1$

Question 1

Use the above information to prove the statements in (1.1, 1.2):

1.1) Show that when $x = \bar{x}$, then the regression equation $\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x$ passes through the mean of y (i.e. $\hat{y} = \bar{y}$ when $x = \bar{x}$)

1.2) $(y_i - \hat{y}_i) = (y_i - \bar{y}) - \widehat{\beta_1}\,(x_i - \bar{x})$

Question 2

The following data is uploaded on the blackboard "consult", and it is taken from a consulting company that compares hours worked (in a week) with the salary of the individuals in ($ 1000's) and number of clients.

Hours worked	67	80	52	68	48	60	78	55	32	58	60
Salary	14	17.5	14.5	12	11	14	20	15.5	11.5	13.3	15.7
# of clients	3	5	2	2	1	3	6	4	3	4	5

2.1) If we were to construct a **simple linear regression** between the Salary and one of the other variables which one is a better predictor for the Salary and give the justification for your answer

2.2) Construct two estimated regression equations once with each variable, and write the interpretation of the slope and the intercepts in for each model in this context.

2.3) Describe the estimated correlation between the three variables.

2.4) Is the linear relationship between the salary and the hours worked significant? Conduct the appropriate hypothesis test

2.5) Using the model with the salary and the number of clients, predict the salary when the number of clients is 2.

2.6) Using the model with the salary and the hours worked, predict the salary when the weekly hours worked is 63 hours.

Question 3

The data "performance" on the blackboard shows 5 variables:

- Hours Studied: The total number of hours spent studying by each student (daily).
- Previous Scores: The scores obtained by students in previous tests.
- Sleep Hours: The average number of hours of sleep the student had per day.
- Sample Question: The number of sample question papers the student practiced.

Target Variable:

- Performance Index: A measure of the overall performance of each student. The performance index represents the student's academic performance and has been rounded to the nearest integer, with higher values indicating better performance. (out of a 100)

The objective is to examine the factors influencing academic student performance. The dataset consists of 10,000 student records, with each record containing information about various predictors and a performance index.

Using this data answer the following:

3.1) What is the estimated regression equation and interpret the estimated coefficients

3.2) Which variable is most significant when predicting the performance of the students. Justify

3.3) Plot a scatter for the following and comment on the relationship:

4.4.a) Performance Vs Sleep hours

4.4.b) Performance Vs Sample Question

3.4) Estimate the standard error for this model:

3.5) Find the following:
3.5a) The regression sum of squares for each independent variable.
3.6b) The total regression sum of squares. which variable has the highest contribution?

3.6) predict the value of performance when the daily hours studied is 4 hours, the previous score is 80, the sleep hours is 3 hours, and the number of Sample Question Papers Practiced is 20. And interpret this interval.

3.7) Construct a 98% confidence interval on the mean value of the performance when the daily hours studied is 4 hours, the previous score is 80, the sleep hours is 3 hours, and the number of Sample Question Papers Practiced is 20. And interpret this interval.

3.8) Construct a 98% prediction interval of the individual value of a performance when the daily hours studied is 4 hours, the previous score is 80, the sleep hours is 3 hours, and the number of Sample Question Papers Practiced is 20. And interpret this interval.

3.9) Is the linear relationship between the performance index and the sleep hours significant? Conduct the appropriate hypothesis test

3.10) Construct a 92% confidence interval for the population slope between the performance index and the previous scores

For the next 4 parts, construct another regression model without the variable "previous scores"

3.11) What is the new estimated regression equation

3.12) predict the value of performance when the daily hours studied is 4 hours, the sleep hours is 3 hours, and the number of Sample Question Papers Practiced is 20.

3.13) Estimate the standard error for this new model and compare it with standard error of the previous model:

3.14) Compare the coefficient of determination between the first and second model