
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DEPARTMENT OF MATHEMATICS & STATISTICS
DHAHRAN, SAUDI ARABIA

STAT 319: Probability & Statistics for Engineers & Scientists

Semester 143

Final Exam

Thursday Aug 13, 2015

Please circle your instructor name:

Nasir Abbas

Saddam Abbasi

Muhammad Riaz

Farah Saleh

Name:

ID #:

Section #:

Serial #:

Question No	Full Marks	Marks Obtained
1	15	
2	11	
3	9	
4	10	
5	10	
6	15	
7	10	
Total	80	

Q.No.1 (15 x 1 = 15 points).

Identify and correct the error, if any, in each of the following statements? If the statement is true, just write True.

- a. If there are 6 digits in an automobile license tag, and each digit must be one of the 10 integers 0, 1, ..., 9, then there are 10^6 possible license tags.

- b. Two events A and B are independent with $P(A) = P(B) = 0.5$. The $P(A \cap B)$ is 1.00

- c. The hypergeometric distribution is associated with sampling with replacement from a finite population of N objects.

- d. For a fixed value of the standard deviation, a 95% confidence interval on the population mean will expand if the sample size increases.

- e. If the P -value for a certain hypothesis test is 0.0125, we can say that the null hypothesis is rejected at the level of significance 0.01.

- f. You are testing $H_0 : \mu = 10, H_0 : \mu > 10$ with unknown standard deviation and a sample size of $n = 15$. The computed value of the test statistic is $t_0 = 2.45$. Because $t_{0.01,14} = 2.625$ we can say that the null hypothesis can be rejected at the 0.01 level of significance.

- g. You are testing $H_0 : \mu = 10, H_0 : \mu > 10$ with unknown standard deviation and a sample size of $n = 15$. The computed value of the test statistic is $t_0 = 2.45$. Because $t_{0.05,14} = 1.76, t_{0.025,14} = 2.148$, and $t_{0.01,14} = 2.625$ we can say that the P -value for this test is greater than 0.025.

- h. If a 95% CI on the difference in two means is $-9.4 \leq \mu_1 - \mu_2 \leq 3.2$ we cannot reject the null hypothesis $H_0 : \mu_1 = \mu_2$ at the 0.05 level of significance.

- i. a fitted linear regression model is $\hat{y} = 10 + 2x$. If $x = 1$ and the corresponding observed value of $y = 11$, the residual at this observation is 0.
- j. If the error or residual sum of squares from fitting a simple linear regression model to 20 observations is 18, the estimate of the variance of the model errors is 2.
- k. If $SS_T = 100$ and $SS_E = 15$, the value of R^2 is 0.90.
- l. The sample correlation coefficient between X and Y is 0.375. It has been found out that the p – value is 0.256 when testing $H_0: \rho = 0$ vs. $H_0: \rho \neq 0$. To test $H_0: \rho \leq 0$ vs. $H_0: \rho > 0$ at significance level of 10%, the p – value is 0.265.
- m. Suppose that A an event in a sample space S , then $P(A) + P(\bar{A}) = 1$, where \bar{A} is the complementary of the event A
- n. Consider a normal population with a known standard deviation. Then the confidence level of the interval $\bar{x} - \frac{2.81\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{2.81\sigma}{\sqrt{n}}$ is equal to 95%.
- o. The demand for a product is $-1, 0, 1, \text{ and } 2$ per day with probability 0.2, 0.1, 0.4, and 0.3 respectively, the expected demand equal to 1.

Q.No. 3 (4+5 = 9 points).

a). A shipment contains 100 printed circuit boards. A sample of 10 boards will be tested. If 2 or fewer defectives are found, the shipment will be accepted. Assuming that 10% of the boards in the shipment are defective, find the probability it will be accepted.

b). A project manager is creating the design for a new engine. He judges that there will be a 50-50 chance that it will have high-energy (H) consumption instead of low (L). Historically, 30% of all high-energy engines have been approved (A) with the rest disapproved (D), while 60% of all low-energy engines have been approved. If the design is approved, what is the probability that it has a high energy consumption?

Q.No. 5 (5+3+2 = 10 points).

In a random sample of 200 Phoenix residents who drive a domestic car, 165 reported wearing their seat belt regularly, while another sample of 250 Phoenix residents who drive a foreign car revealed 198 who regularly wore their seat belt.

- (a) Is there a statistically significant difference in seat belt usage between domestic and foreign car drivers? Set your probability of a type I error to 0.05. Use p-value approach to make your decision.

- (b) Construct 95% confidence interval for the difference in the proportion of seat belt usage between domestic and foreign car drivers?

- (c) Use the confidence interval obtained in part (b) to verify your decision of part (a).

Q6. (3 +4+1+5+2= 15 points):

As machines are used over long periods of time, the output product can get off target. Data is collected on the average value of how much off target a product is getting manufactured as a function of machine use. The intensity of being off target (Y) is measured in Millimeters and the time of machine use (X) is measured in hours. Based on seven randomly collected observations we have the following summary quantities.

$$n = 7, \quad \sum y_i = 8.91, \quad \sum x_i = 260, \quad \sum y_i^2 = 11.4159, \quad \sum x_i^2 = 9852, \quad \sum x_i y_i = 334.68$$

- a) Estimate the degree of linear correlation between intensity of being off target and the time of machine use. Interpret this quantity

Interpretation:

- b) Find the estimated regression line.

- c) What change in mean intensity of being off target would be expected for a 1 hour change in the time of machine use?

- d) At 1% level of significance, test that the more the time of machine use, the higher the intensity of being off target. How does your conclusion match with that in part (a)?

Comparison with part (a): _____

- e) What are the assumptions required for the different components of regression analysis?

Q7. (4+3+3= 10 points):

The manager of a car plant wishes to investigate how plant's electricity usage depends upon the amount plant's production. He had the following data

	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
Production (\$million)	4.51	3.58	4.31	5.06	5.64	4.99	5.29	5.83	4.7	5.61	4.9	4.2
Electricity Usage (million KWh)	2.48	2.26	2.47	2.77	2.99	3.05	3.18	3.46	3.03	3.26	2.67	2.53

A simple linear regression was fit using MINITAB and the results are:

Term	Coef	SE Coef	T-Value	P-Value
Constant	0.409	0.386	1.06	0.314
Production (\$million)	0.4988	0.0784	6.37	0.000

a) Complete the ANOVA table

Source	DF	SS	MS	F
Regression	___	1.21238	_____	_____
Error	___	_____	_____	
Total	___	_____		

b) Does production affect electricity usage? Justify your answer.

c) Find a 90% confidence interval for the mean electricity usage, when the production is worth 5 \$million.