PRACTICE EXAM – EMSE 4765/6765

Student Name:

Exam by: J. René van Dorp[1]

www.seas.gwu.edu/~dorpjr

Exam products are this workbook with answers,
an excel spreadsheet with your calculations and a Minitab File.

Use the excel template "PRACTICE_EXAM_STUDENTNAME.XLSX"

SAVE FREQUENTLY

Please include your name in the filenames of these files.

[1] Department of Engineering Management and Systems Egineering, School of Engineering and Applied Science, The George Washington University, 1776 G Street, N.W., Suite 110, Washington D.C. 20052. E-mail: dorpjr@gwu.edu.

Question 1: (Total 20 points)

A maintenance firm has gathered the following information regarding failure mechanisms for air conditioning systems:

		Evidence of Gas Leaks	
		Yes	No
Evidence of Electical Failure	Yes	55	17
	No	32	3

The units without evidence of gas leaks or electrical failure showed other types of failure. Answer the questions below using the following event definitions

A = {An AC failure involves a gas leak},

B = {An AC failure involves an electical failure},

C = {An AC failure involves no gas leak nor an electical failure},

a. (6 Points) If this is a representative sample of AC failure, find the probabilities that an AC failure involves a gas leak, an AC failure involves an electrical failure and an AC failure involved no gas leak nor an electrical failure. **Provide explanations below or in EXCEL. Perform your calculations in EXCEL in the worksheet Question 1.**

Total Number of Failures: $55 + 17 + 32 + 3 = 107$

$Pr(A) = \frac{55+32}{107} \approx 0.813; Pr(B) = \frac{55+17}{107} \approx 0.673;$

$Pr(C) = \frac{3}{107} \approx 0.028$

b. (1 Point) True or False? The events A, B and C are mutually exclusive.

FALSE, $Pr(A) + Pr(B) + Pr(C) > 1$

c. (1 Point) True or False? The events A, B and C are collectively exhaustive.

TRUE, there are no other possibilities. Gas Failure, Electrical Failures and Other Failures.

d. (3 Points) Find the probability that there is evidence of electrical failure given that there was a gas leak. **Provide explanations below or in EXCEL. Perform your calculations in EXCEL in the worksheet Question 1.**

$$Pr(B|A) = \frac{Pr(A \cap B)}{Pr(A)} = \frac{55/107}{(55+32)/107} = \frac{55}{55+32} \approx 0.632$$

e. (3 Points) Find the probability that there is evidence of a gas leak given that there was an electrical failure. **Provide explanations below or in EXCEL. Perform your calculations in EXCEL in the worksheet Question 1.**

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{55/107}{(55+17)/107} = \frac{55}{55+17} \approx 0.764$$

f. (3 Points) Are the events A and B independent? Explain why or why not. **Provide explanations below or in EXCEL.**

P(B|A) is not equal to P(B), hence the events A and B are not independent

g. (3 Points) What general conclusion can you draw from the analysis above regarding electical failures and and gas leak failures of AC units.

While formally the events A and B are not independent P(B|A) and P(B) are close and P(A|B) and P(A) are close. Hence, information about an electrical failure does not tell you much about a change in likelihood of a gas failure and vice versa.

Question 2: (Total 20 points)

Let the random variables X_1 and X_2 denote the length and width, respectively, of a manufactured part. Assume that X_1 is normal with $E[X_1] = 2$ centimers and standard deviation 0.1 centimeters and that X_2 is normal with $E[X_2] = 5$ centimeters and standard deviation 0.2 centimeter.

a. (5 Points) Assuming that X_1 and X_2 are independent. Determine the probability that the perimeter exceeds 14.5 centimeters. **Show your derivations here and perform your calculations in Microsoft Excel in the worksheet Question 2. (Hint: Use NORMDIST function in Excel).**

$$Y = Perimeter, Y = 2X_1 + 2X_2$$

$$E[Y] = 2E[X_1] + 2E[X_2] = 4 + 10 = 14$$

$$V[Y] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j COV[X_i, X_j] =$$
$$4Var[X_1] + 4Var[X_2] + 2 \times 4Cov[X_1, X_2]$$

$$X_1, X_2 \text{ are independent, thus } Cov[X_1, X_2] = 0$$

$$V[Y] = 4 \times (0.1)^2 + 4 \times (0.2)^2 = 0.2$$

Using NORMDIST function in Excel: $Pr(Y > 14.5) \approx 0.132$

b. (5 Points) Assuming that the correlation between X_1 and X_2 equals 0.5 determined the probability that the perimeter exceeds 14.5 centimers. **Show your derivations here and perform your calculations in Microsoft Excel in the worksheet Question 2.**

$$Y = Perimeter, Y = 2X_1 + 2X_2$$

$$E[Y] = 2E[X_1] + 2E[X_2] = 4 + 10 = 14$$

$$V[Y] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j COV[X_i, X_j] =$$
$$4Var[X_1] + 4Var[X_2] + 2 \times 4Cov[X_1, X_2]$$

$Cor[X_1, X_2] = 0.5$, thus $Cov[X_1, X_2] = 0.5 \times 0.1 \times 0.2 = 0.01$

$$V[Y] = 4 \times (0.1)^2 + 4 \times (0.2)^2 + 8 \times 0.01 = 0.28$$

Using NORMDIST function in Excel: $Pr(Y > 14.5) \approx 0.172$

c. (5 Points) Assuming that the correlation between X_1 and X_2 equals -0.5 determined the probability that the perimeter exceeds 14.5 centimers. **Show your derivations here and perform your calculations in Microsoft Excel in the worksheet Question 2.**

$$Y = Perimeter, Y = 2X_1 + 2X_2$$

$$E[Y] = 2E[X_1] + 2E[X_2] = 4 + 10 = 14$$

$$V[Y] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j COV[X_i, X_j] =$$
$$4Var[X_1] + 4Var[X_2] + 2 \times 4Cov[X_1, X_2]$$

$$Cor[X_1, X_2] = -0.5, \text{ thus}$$
$$Cov[X_1, X_2] = -0.5 \times 0.1 \times 0.2 = -0.01$$

$$V[Y] = 4 \times (0.1)^2 + 4 \times (0.2)^2 - 8 \times 0.01 = 0.12$$

Using NORMDIST function in Excel: $Pr(Y > 14.5) \approx 0.074$

d. (5 Points) Please provide an explanation in words for your outcomes under a, b and c.

Under B, the length and width are positively dependent. Hence large (small) values of length are associated with large (small) values of width. Hence, since Y=2X1+2X2 they influence the linear combination in the same direction and hence the tail probability of the normal distribution of Y is thicker end hence a larger Pr(Y>14.5) under B than under question A

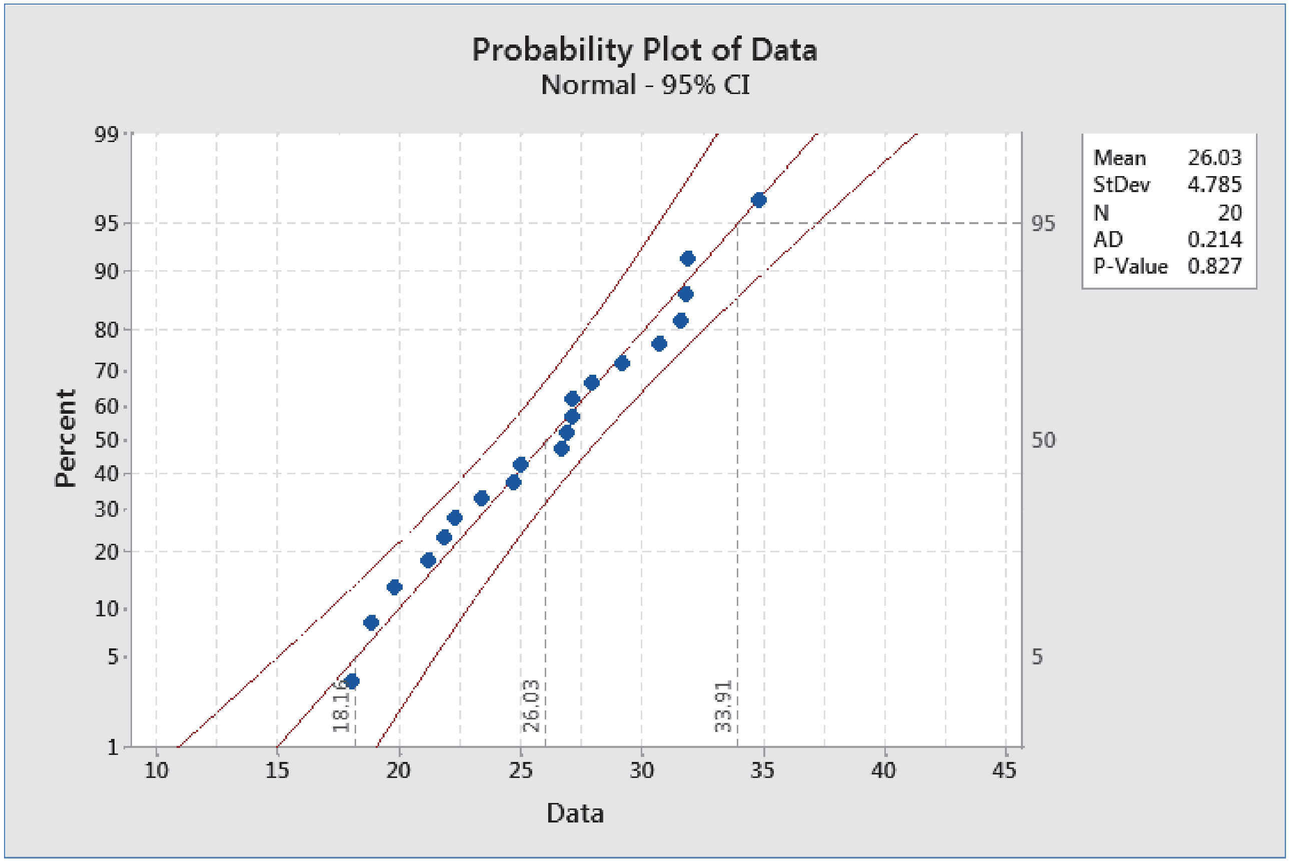
Under C, the length and width are negatively dependent. Hence large values of length are associated with small values of width and vice versau. Hence, since Y=2X1+2X2 they influence the linear combination in opposite directiona and hence the tail probability of the normal distribution of Y is thinner end hence a smaller Pr(Y>14.5) under C than under question A

Question 3: (Total 24 points)
Cloud seeding has been studied for many decades as a weather modification procedure The rainfall in acre-feet from 20 clouds that were selected at random and seeded with silver nitrate are:

18.00	30.70	19.80	27.10	22.30	18.80	31.80	23.40	21.20	27.90
31.90	27.10	25.00	24.70	26.90	21.80	29.20	34.80	26.70	31.60

(The data is also provided in the midterm EXCEL workbook).

a. (6 points) Copy and past the data from EXCEL workbook into MINITAB and create a normal probability plot in MINITAB, with 5%, 50% and 95% percentiles. Save the MINITAB file with your name in the filename. Write down your conclusions below based on this normal probability plot.



All points fall within the confidence boundaries hence the normal distribution seems to be a good fit. In addition, the p-value of the Anderson Darling test equals 0.827, hence we fail to reject the normality hypothesis test at significance levels 1%, 5% and 10%

b. (6 points) Write down the formula for a two-sided 90% confidence interval for the mean rainfall and calculate it. **Perform your calculations in Microsoft Excel in the worksheet Question 3.** Write down an interpretation of this interval below.

$$\mu \in [\overline{X} - \frac{S \times t_{n-1,0.95}}{\sqrt{n}}, \overline{X} + \frac{S \times t_{n-1,0.95}}{\sqrt{n}}]$$

$$n = 20,\ \overline{x} = 26.035,\ s = 4.785,\ \frac{s}{\sqrt{n}} = 1.070,\ t_{19,0.95} = 1.729$$

90% Confidence interval $[24.185, 27.885]$

Intepretation: 90% confidence interval is a realization of a random interval that has a 90% probability of capturing the mean μ. The calculated confidence interval above either contains the value μ or it does not. No probability interpretation can be assigned to the 90% confidence interval.

c. (6 points) Write down the formula for the T-statistic to test the hypothesis

$$H_0 : \mu = 24.5,\ H_1 : \mu > 24.5$$

Calculate the value of the T-statistic for the data at hand and calculate the p-value of this hypothesis test. If your significance level $\alpha = 10\%$ what is your conclusion and why? Write down your conclusion below. **Perform your calculations in Microsoft Excel in the worksheet Question 3.**

$$t_0 = \frac{\overline{x}-\mu_0}{s/\sqrt{n}}, t_0 = 1.43706,\ p - value = Pr(T_{19} > t_0) = 0.083$$

Conclusion: Reject H_0 at 10% significance level

d. (6 points) Calculate and interpret a two-sided 90% credibility interval for the rainfall in the midterm worksheet Question 3. **Perform your calculations in Microsoft Excel in the worksheet Question 3.** Write down an interpretation of this interval below.

Set $\mu = \overline{x}$, $\sigma = s$,

$$LB = NormInv(0.05, \overline{x}, s) = 18.165$$
$$UB = NormInv(0.95, \overline{x}, s) = 33.905$$

A 90% credibility interval for the rainfall is also a random interval with a 90% probability of contain the value of X. The calculated interval above is a realization of that interval and $Pr(X \in [LB, UB]) \approx 0.90$

Question 4:

In the first phase of a study of cost of transporting milk from farms to dairy plants, a survey was taken of firms engaged in milk transportation. Cost data on $X_1 = fuel$, $X_2 = repair$ and $X_3 = capital$, all measured on a per-mile basis, are presented in two worksheets. The data for the 36 gasoline trucks is presented in the worksheet "Question 4 Gasoline" and the data for 23 diesel trucks are presented in the worksheet "Question 4 Diesel".

a. (3 points) Calculate the sample means, sample standard deviations and sample variance of of X_1, X_2, X_3 for the gasoline trucks in the worksheet "Question 4 Gasoline". **See Spreadsheet Midterm Solution.**

b. (4 points) Calculate the sample covariance matrix of of X_1, X_2, X_3 for the gasoline trucks in the worksheet "Question 4 Gasoline" using matrix multiplication **See Spreadsheet Midterm Solution.**

c. (3 points) Calculate the sample means, sample standard deviations and sample variance of of X_1, X_2, X_3 for the gasoline trucks in the worksheet "Question 4 Diesel". **See Spreadsheet Midterm Solution.**

d. (4 points) Calculate the sample covariance matrix of of X_1, X_2, X_3 for the gasoline trucks in the worksheet "Question 4 Diesel" using matrix multiplication. **See Spreadsheet Midterm Solution.**

e. (10 points) In the worksheet "Question 4 Hotelling T2" test for differences of the mean vectors of the gasoline data and the diesel data. Set the significance level $\alpha = 0.10$. Also determine the p-value of the Hotelling T^2 test. **See Spreadsheet Midterm Solution.**

f. (9 points) In the worksheet "Question 4 Variances" test for differences of the variances of X_1, X_2 and X_3 of the gasoline data and the diesel data. Set the significance levels at $\alpha = 0.10$. **See Spreadsheet Midterm Solution.**

g. (5 points) Write down your conclusions or comments with respect to the validity of the analysis conducted. **See Spreadsheet Midterm Solution.**

Normality of the data has to be tested for this example using probability plots. Even without failing to reject this hypothesis, the validaty of the analysis depends on the covariance matrices of both samples being the same. The equality of the variance in fuel cost and gasoline cost is rejected even at the 1% significance level. Formally, however, Box's M hypothesis test would have to be performed comparing both sample variance covariance matrices.