EMSE 4765/EMSE 6765 - SPRING 2021 MIDTERM EXAM

Student Name:

Tuesday, March 9, 2021 at 12:45PM

Allotted time: Two and a half hours

Print out the exam booklet. For full credit, **clearly show all work and indicate how the solution was determined**. Most answers can be typed in the Micro Soft Excel spreadsheet "**Midterm_Exam_SP_2021_Student_name.xlsx". Calculations need to be completed with Excel formulas in MicroSoft Excel (not Google Sheets as Excel formulas may disappear when saving the Google Sheet). You will need to copy some graphs from MINITAB into the MS Excel spreadsheet.**

This is an OPEN NOTES, CLOSED HOMEWORK SOLUTIONS exam.

You may not search the internet or use any form of social media during the exam with any type of device. This will be considered a violation of Academic Integrity. You will be required to turn your camera on as you take the exam.

At the end of the exam you need to upload the Midterm Exam MS Excel Spreadsheet first in BLACKBOARD and hand in the exam booklet.

After uploading your MS Excel Spreadsheet with EXCEL formulas first on BLACKBOARD, you will need to take pictures of page 6 and 7 of Question 3 and upload them on BLACKBOARD. Alternatively, you can scan the entire exam booklet in a pdf file and upload the pdf file on BLACKBOARD.

Question 1: (Total 50 points)
Four psychological tests were given to 32 men and 32 females. The data are recorded in in the worksheets "Question 1 MALE" and "Question 1 FEMALE". The variables were

X_1 = pictorial inconsistencies
X_2 = paper form board
X_3 = tool recognition
X_4 = vocabulary

a. (5 points) Calculate the sample means, sample standard deviations and sample variance of X_1, X_2, X_3 and X_4 for the 32 males in the worksheet "Question 1 MALE". **Perform your calculations using Microsoft Excel Formulas.**

b. (5 points) Calculate the sample covariance matrix of of X_1, X_2, X_3 and X_4 for the 32 males in the worksheet "Question 1 MALE" using matrix multiplication. **Perform your calculations using Microsoft Excel Formulas.**

c. (5 points) Calculate the sample correlation matrix of of X_1, X_2, X_3 and X_4 for the 32 males in the worksheet "Question 1 MALE" using matrix multiplication. **Perform your calculations using Microsoft Excel Formulas.**

d. (5 points) Calculate the sample means, sample standard deviations and sample variance of X_1, X_2, X_3 and X_4 for the 32 females in the worksheet "Question 1 FEMALE". **Perform your calculations using Microsoft Excel Formulas.**

e. (5 points) Calculate the sample covariance matrix of of X_1, X_2, X_3 and X_4 for the 32 females in the worksheet "Question 1 FEMALE" using matrix multiplication. **Perform your calculations using Microsoft Excel Formulas.**

f. (5 points) Calculate the sample correlation matrix of of X_1, X_2, X_3 and X_4 for the 32 males in the worksheet "Question 1 FEMALE" using matrix multiplication. **Perform your calculations using Microsoft Excel Formulas.**

g. (5 points) What can you conclude from the sample correlation matrices calculated under c. and f. above. Provide you answer in the worksheet "Question 1 Hotelling T2".

h. (10 points) In the worksheet "Question 1 Hotelling T2" test the hypothesis test

$$H_0 : \boldsymbol{\mu_1} - \boldsymbol{\mu_2} = \boldsymbol{\Delta},\ H_1 : \boldsymbol{\mu_1} - \boldsymbol{\mu_2} \neq \boldsymbol{\Delta}$$

where $\boldsymbol{\Delta}^{\boldsymbol{T}} = (0,\ 0,\ 0, 0)$

Set the significance level at $\alpha = 1\%$. Also determine the *p*-value of the Hotelling T^2 test. **Perform your calculations using Microsoft Excel Formulas.**

i. (5 points) Write down your conclusions and comments with respect to the analysis conducted under h. Provide you answer in the worksheet "Question 1 Hotelling T2"

Question 2: (Total 50 points)
Data for the lengths (in minutes) of nine-innings games for the first week of the 2001 major league baseball season are provided below and in the MS Worksheet Question 2.

194	160	176	203	187	163	162	183	152	177
177	151	173	188	179	194	149	165	186	187
187	177	187	186	187	173	136	150	173	173
136	153	152	149	152	180	186	166	174	176
198	193	218	173	144	148	174	163	184	155
151	172	216	149	207	212	216	166	190	165
176	158	198							

Assume that this is a realization of a random sample for nine-innings games.

a. **(5 points)** Calculate the 90% confidence interval for the population mean μ. **Perform your calculations using Microsoft Excel Formulas in the cells below in worksheet "Question 2"**.

b. **(5 points)** What is the complete interpretation of the 90% confidence interval under a. . Write the complete interpretation in the worksheet "Question 2" in EXCEL.

c. **(2 points)** Copy and past the data from the EXCEL workbook into MINITAB and create a normal probability plot in MINITAB for the lengths (in minutes) of nine-innings games. Copy and past the graph in your EXCEL spreadsheet in the worksheet "Question 2".

d. **(4 points)** What conclusions can you draw from the probability plot. Write your conclusions in the worksheet "Question 2" in EXCEL.

e. **(2 points)** Evaluate a 90% credibility interval in MINITAB by adding percentile lines to that probability plot. Copy and past the probability plot with the percentile lines in the EXCEL spreadsheet.

f. **(3 points)** What is the complete interpretation of the 90% credibility interval under e. Write the complete interpretation in the worksheet "Question 2" in EXCEL.

g. **(5 points)** Evaluate/Caclulate the 90% credibility interval evaluated by MINITAB under e in EXCEL. **Perform your calculations using Microsoft Excel Formulas in the worksheet "Question 2"**.

h. **(5 points)** Test at a 5% significance level the following hypothesis

$$H_0 : \mu = \mu_0,\ H_1 : \mu < 179$$

by evaluating the t_0-statistic estimate for the null-hypothesis above and the p-value of the hypothesis test. **Perform your calculations using Microsoft Excel Formulas in the worksheet "Question 2"**.

i. **(4 points)** State your conclusion from the analysis under h. in words in EXCEL and provide the interpretation of the p-value under h. Write your answers in the worksheet "Question 2" in EXCEL.

j. **(5 points)** By calculating an estimate for the χ^2_{n-1} estimator

$$\chi^2_{n-1} = \frac{(n-1)S^2}{\sigma_0^2},$$

test at a 5% significance level the hypothesis

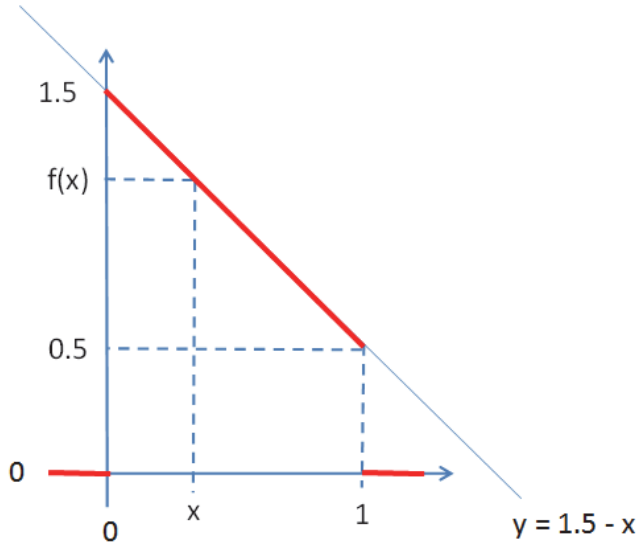
$$H_0 : \sigma^2 = \sigma_0^2,\ H_1 : \sigma^2 > \sigma_0^2$$

where $\sigma_0^2 = 275$ for the null-hypothesis H_0 above by evaluating the p-value of the hypothesis test. **Perform your calculations using Microsoft Excel Formulas in the worksheet "Question 2".**

k. **(4 points)** State your conclusion from the analysis under j. in words in EXCEL and Provide the interpretation of the p-value under j. Write your answers in the worksheet "Question 2" in EXCEL.

l. **(5 points)** What is the relevance of the conclusions under d. for the mean hypothesis test under h. and the variance hypothesis test under l? Write your answers in the worksheet "Question 2" in EXCEL.

Question 3 (50 Points).
Consider the probability density function defined by the graph below (in red) with strict positive density values on the interval $[0, 1]$ and **zero density values elsewhere**.



a. **(5 Points) Explain why the function (in red) in the graph above is a probability density function. Show your derivations below.**

$$f(x) \geq 0 \quad \forall x.$$

$$\int_{-\infty}^{\infty} f(x)\, dx$$

$$= \int_0^1 1.5 - x\, dx$$

$$= 1.5x - \frac{1}{2}x^2 \Big|_0^1$$

$$= 1$$

b. **(5 Points) Derive the cumulative distribution function of this density function (in red) in the graph above. Show your derivations below.**

$$F(x) = \int_{-\infty}^{x} f(y)\,dy$$

$$= \int_0^x 1.5 - y\,dy$$

$$= 1.5y - \frac{1}{2}y^2 \Big|_0^x$$

$$= 1.5x - \frac{1}{2}x^2,$$

So. $F(x) = \begin{cases} 0, & x \le 0 \\ 1.5x - \frac{1}{2}x^2, & 0 < x < 1 \\ 1, & x > 1 \end{cases}$

c. **(5 Points). Derive the inverse of the cumulative distribution function that you derived under b. Show your derivations below.**

Hint: The solutions of $ax^2 + bx + c = 0$ are: $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

Since $y = 1.5x - \frac{1}{2}x^2$,

we know $\frac{1}{2}x^2 - 1.5x + y = 0$

$$x = \frac{1.5 \pm \sqrt{(1.5)^2 - 2y}}{1}$$

$$= \frac{3}{2} \pm \sqrt{\frac{9}{4} - 2y}$$

So, $f^{-1}(x) = \frac{3}{2} \pm \sqrt{\frac{9}{4} - 2y}$, if $0 \le x \le 1$.

d. **(10 Points).** Complete the **CHI-SQUARED goodness of fit table A** in the worksheet named "QUESTION 3" **for the density function under a.** using **the equal bin-width method and the 50 data points in that sheet**. E_i in that table is the expected number of observations in Bin i with lower bound LB_i and upper bound UB_i and O_i is the actual number of observations in Bin i with lower bound LB_i and upper bound UB_i. **Perform your calculations using Microsoft Excel Formulas.**

e. **(5 Points).** Complete the **CHI-SQUARED goodness of fit analysis under d. in Table B** in the worksheet named "QUESTION 3" in the designated cells using a 25% significance level. What is your conclusion? Type your conclusion in the worksheet "QUESTION 3" in Excel. **Perform your calculations using Microsoft Excel Formulas.**

f. **(10 Points).** Complete the **CHI-SQUARED goodness of fit table C** below in the worksheet named "QUESTION 3" **for the density function under a.** using **the equal probability method and the 50 data points in that sheet**. E_i in the table below is the expected number of observations in Bin i with lower bound LB_i and upper bound UB_i. O_i in the table below is the actual number of observations in Bin i with lower bound LB_i and upper bound UB_i. **Perform your calculations using Microsoft Excel Formulas.**

g. **(5 Points).** Complete the **CHI-SQUARED goodness of fit analysis under f. in Table D** in the worksheet named "QUESTION 3" in the designated cells using a 25% significance level. What is your conclusion? Type your conclusion in the worksheet "QUESTION 3" in Excel. **Perform your calculations using Microsoft Excel Formulas.**

h. **(5 Points).** Do you prefer the analysis under d. and e. or the analysis under f. and g. Explain your answer. Type your reasoning in the worksheet "QUESTION 3" in Excel.