Istanbul Technical University-Spring 2018 2 / 5 / 2018

Statistics (MATH 271)

Homework Assignment 3 KEY

Instructor: Halil Bayraktar

Due data: 11 / 5 / 2018 until 1:00 pm. No late submission will be accepted.

HW3 is prepared to help you doing practice about statistical inference that you have learned in lecture 6-11. There are 3 questions.

Please download and print this pdf document. You should only use this document to write your answers/solutions in a provided space in each question. You should also download other files needed to answer questions.

You should write clearly and concisely. Put your final answer to the box given in each question for full credit. You have to show all your work for full credit.

When finished your homework, you can drop your papers to the box outside my office. and submit it before the deadline.

It is not allowed to take another student's solution. You cannot give your solution/results to your classmates.

Good luck.

(Type in capital letters)
First Name:
Last Name:
ID:
Signature:

Question	Score
Q1 (50)	
Q2 (25)	
Q3 (25)	
Total (100)	

1. (50 points) - Biology question.

In this question, you will analyze the effect of different siRNA molecules that are tested to increase the rate of apoptosis (cell death) in cancer cells. The decrease in cell numbers indicates the efficacy of molecules. The results are summarized in the following table. N, x (percentage average) and s (standart deviation) indicates the number of tested samples, average cell death (percentage) and standard deviation.

	n	X	S
siRNA A	20	70.75	14.68
siRNA B	20	62.75	11.98
siRNA C	20	55.62	11.56

a) Determine all the entries of the ANOVA table given the information above. Show all your calculations clearly for full credit.

$$SS_{Groups} = \sum_{groups} n_i(\bar{x}_i - \bar{x})^2$$

$$x_{grups} = x_1 + x_2 + x_3 = 63.04$$

$$SS = 20 \times (70.75 - 63.04)^2 + 20 \times (62.75 - 63.04)^2 + 20 \times (55.62 - 63.04)^2$$

$$MS = \frac{SS}{df} = \frac{2290}{2} = 1145$$

$$SS_{Errors} = \sum_{errors} (n_i - 1) s_i^2$$

$$SS = 19 \times (14.68)^2 + 19 \times (11.98)^2 + 19 \times (11.56)^2$$

$$SS = 9367$$

$$MS = \frac{SS}{df} = \frac{9367}{57} = 164$$

$$f = \frac{MSG}{MSE} = \frac{1145}{164}$$

$$f = 6.97$$

ANOVA	df	SS	MS	F
Groups	2	2290	1145	6.97
Error	57	9367	164	
Total	59	11657		

b) Determine the entry name and value in the ANOVA table that returns the pooled estimate of the gene death variance, s_p^2?

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2}{(n_1-1)+(n_2-1)+(n_3-1)}$$

Name: (MSE) Mean-Squared Error	s_p^2: 164

c) Please state the null (Ho) and alternative (Ha) hypothesis for this study. What is the distribution of the F statistic under Ho? Use F table to obtain an approximate p-value and test at the α=0.05 significance level.

$X_{siRNA A} \sim N(\mu_{siRNA A}, \sigma)$

$X_{siRNA B} \sim N(\mu_{siRNA B}, \sigma)$

$X_{siRNA C} \sim N(\mu_{siRNA C}, \sigma)$

$f = 6.97$, from table, $4.98 < f < 7.77$

$0.001 < P < 0.01$

$P = 0.002$ (from software)

Ha: $\mu_{siRNA A} = \mu_{siRNA B} = \mu_{siRNA C}$

Ho: Not all of the μ are equal.

F: $4.98 < f < 7.77$

d) Now you want compare the cell death from siRNA-A and siRNA-B. Test at the α=0.05 level wether the cell deaths are different in the presence of siRNA A vs. siRNA B. (use two-sided test)

$$t = \frac{x_A - x_B}{\left(\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}\right)^{1/2}} = \frac{(70.75-62.75)}{\left(\frac{14.68^2}{20} + \frac{11.98^2}{20}\right)} = 1.88$$

$df = 19$, $\alpha = 0{,}05$

$1.729 < t < 2.093$ (from table)

$0.025 < P < 0.05$ x2 (one sided)

$0.05 < P < 0.1$ (two-sided)

t: 1.88	P: $0.05 < P < 0.1$

Null Hypothesis: Reject or Accept (circle your answer)

2. (25 points) - Drug discovery question.

A drug discovery study is used to determine the effect of a new small molecule to control the blood sugar levels in diabetes patients. The treatment includes 24 participants. A control group of 12 people received a placebo. Other 12 people received small molecule. Here are the summary of the measurements.

Molecule A	Placebo
156	204
136	200
196	192
208	216
156	240
188	200
148	200
172	160
204	148
180	184
152	188
148	228

a) Compute the mean and the standard deviation of each sample and complete the table shown below.

$$\bar{X}_A = \frac{1}{12} \sum_{i=1}^{12} x_i$$

$$S_A = \left(\frac{\sum_{i=1}^{12} (\bar{x} - x_i)}{n-1} \right)^{1/2}$$

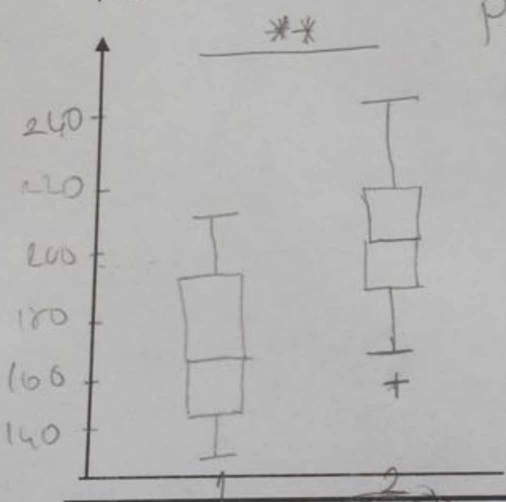
	n	X	S
Molecule A	12	170.3	24.3
Placebo	12	196.6	25.7

b) Can we conclude that the average blood sugar levels in the presence of molecule A and placebo are not the same? Compute the pooled standard deviation by a formula,

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

s_p^2: 628	t= 2.57

c) Draw a boxplot with 5 critical values and show both data on the plot? State Ho and Ha for this experiment. Determine the p value and determine if null hypothesis (Ho, x1=x2) is rejected at the 0.05 significance level? Show the corresponding p value on the plot.



$P<0.01$

$df=22,\ t=2.57$

$2.508<t<2.819$

$0.005<P<0.01$

$P_{exact}=0.0087$

Null Hypothesis: Reject or Accept (circle your answer)	P = $0.005<P<0.01$

d) What is the difference in means at the 90% confidence level?

$$\bar{X}_{Placebo} - \bar{X}_A \pm t^x \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right) = 26.3 \pm 1.717 \times 10.23$$

$t^x = 1.717$ for 90% confidence level

(from table)

$df=22$

$\Delta\dot{x}$= (8.82, 43.83)

3. (25 points) - Social science question.

A following survey data indicates the use of Instagram. The table shown below indicates the response with respect to the gender. X represents the answer of yes

	n	X
Male	1750	850
Female	1950	1250

a) Find the sample proportions of Instagram use for male and females.

$$P_{Male} = \frac{850}{1750} = 0.48$$

$$P_{female} = \frac{1250}{1950} = 0.64$$

Pmale= 0.48

Pfemale= 0.64

b) Report the difference in the proportions and standard error of the difference.

$$D = \hat{p}_1 - \hat{p}_2 = P_{female} - P_{Male} = 0.64 - 0.48 = 0.16$$

$$SE = \left(\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2} \right)^{1/2}$$

$$SE = \left(\frac{(0.64)(0.36)}{1950} + \frac{(0.48)(0.52)}{1750} \right)^{1/2} = 0.01$$

Difference= 0.16

SE= 0.01

c) Find the test statistics (z) and the p value (two-sided). Is there a difference at the 0.05 significance level? Give a 95% confidence interval (CI) for the difference in proportions and briefly summarize what the data show.

$H_0: p_1 = p_2$

$H_a: p_1 \neq p_2$

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{850 + 1250}{1750 + 1950} =$$

$$\hat{p} = \frac{2100}{3700} = 0.56$$

$$SE = \left(\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)^{1/2}$$

$$SE = \left((0.56)(0.44)\left(\frac{1}{1750} + \frac{1}{1950}\right)\right) = 0.0163$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{SE} = \frac{0.16}{0.0163} = 9.52 \Rightarrow p < 1.0 \times 10^{-5}$$

or

$p \cong 0$

$$CI = \hat{p} \pm zSE = 0.16 \pm (1.96)(0.01)$$
$$= 0.16 \pm 0.02$$

Yes there is a difference at $\alpha = 0.05$

z = 9.52	P = $< 1.0 \times 10^{-5}$	CI= (0.14, 0.18)

or

$P \cong 0$