1	2	3	4	Total

Name: Answers
Number:

BLG560E - Statistics and Estimation for Computer Science

Spring 2021-2022
Final Exam

09.06.2022

Rules:

- Duration is 90 min.
- Show your work, do not write the result directly.
- Use the attached distribution lookup tables if required.
- Do not make any approximations between distributions.
- Do not ask any questions during exam. If you think something is wrong or missing, write your assumption(s) and solve the question according to your assumption.
- You can round floating point numbers to two decimal places.
- Solve each question within the corresponding frame. Anything outside the frame **will not be** graded.

Questions:

1. (25 pts) Following data give the number of crimes by days of the week in Istanbul.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
# of crimes	75	97	94	83	107	100	109

Test the hypothesis that a crime is equally likely to occur on any of the 7 days of the week. Use significance level of 0.05.

Let x_i = # of crimes in day i

Expected # of crimes per day = $\frac{1}{7}\sum x_i$

$= 95$

$$X^2 = \sum_i \frac{(x_i - 95)^2}{95} = 9.6$$

At $\alpha = 0.05$ and dof = 6

$$X_c^2 = 12.592$$

Since $X^2 < X_c^2$, H_0 should be retained.

2. (25 pts) An experiment was initiated to study the effect of a newly developed gasoline detergent on mileage. Following data represents km per litre before and after the detergent was added for each of 8 cars.

	Car 1	Car 2	Car 3	Car 4	Car 5	Car 6	Car 7	Car 8
Mileage with detergent	23.5	29.6	32.3	17.6	25.3	25.4	22.2	20.7
Mileage without detergent	24.2	30.4	32.7	19.8	25.0	24.9	20.6	20.7

Find the **p-value** of the test of the hypothesis that mileage is not affected by the addition of detergent using sign test and Wilcoxon signed rank test. Use Normal distribution approximation. Do not use any continuity correction.

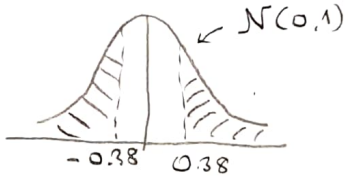
(a) (13 pts) Sign test

Car 1 2 3 4 5 6 7 8

Sign − − − − + + + ∅ ← ignore

neg = 4, pos = 3 } N = 7

Normal apr. $\sim \mathcal{N}\left(\underbrace{Np}_{3.5}, \underbrace{Np(1-p)}_{1.75}\right)$



$$z = \frac{4-3.5}{\sqrt{1.75}} = \frac{0.5}{1.32} = 0.38$$

$$\text{p value} = 2 \times \underbrace{(1 - \Phi(0.38))}_{0.65} = 2 \times .35 = 0.7$$

(b) (12 pts) Wilcoxon signed-rank test

Car	1	2	3	4	5	6	7	8
di	-0.7	-0.8	-0.4	-2.2	0.3	0.5	1.6	0
rank	4	5	2	7	1	3	6	ignore

$W^+ = 1+3+6 = 10$

$W^- = 4+5+2+7 = 18$

$W^+ + W^- = \frac{7 \times 8}{2} = 28$

$W = \min(W^-, W^+) = 10$

$W \sim N\left(\frac{7\times 8}{4}, \frac{7\times 8\times 15}{24}\right)$

$\sim N(14, 35)$

$$z = \frac{W-14}{\sqrt{35}} = \frac{-4}{5.92} = -0.68$$

$$\text{p value} = 2 \times \underbrace{(1 - \Phi(0.68))}_{0.75} = 0.5$$

1	2	3	4	Total

Name:
Number: Answers

3. (25pts) Preliminary studies indiate a possible connection between one's natural hair color and threshold for pain. A sample of 12 women were classified as to having light, medium and dark hair. Each was the given a pain sensitivity test, with the following result.

Light	Medium	Dark
63	60	45
72	48	33
52	44	57
60	53	40

Are the given data sufficient to establish that hair color affects the results of a pain sensitivity test. Use significance level of 0.05.

Using Anova

Light	Medium	Dark
63	60	45
72	48	33
52	44	57
60	53	40
$\bar{x}$ 61.75	51.25	43.75

$\bar{\bar{x}} = 52.25$

$$SSTr = [(61.75-52.25)^2 + (51.25-52.25)^2 + (43.75-52.25)^2] \times 4$$

$$= 163.5 \times 4 = 654$$

$$MSTr = \frac{SSTr}{k-1} = \frac{654}{2} = 327$$

$$SSE = (63-61.75)^2 + \dots + (60-61.75)^2 +$$
$$(60-51.25)^2 + \dots + (53-51.25)^2 +$$
$$(45-43.75)^2 + \dots + (40-43.75)^2 = 654.25$$

$$MSE = \frac{654.25}{9} = 72.69 \qquad F = \frac{MSTr}{MSE} = \frac{327}{72.69} = 4.5$$

$F_{c, 0.05, 2/9} = 4.26$

As $F > F_c$ H_0 should be rejected.

4. (25 pts) It is generally accepted that by increasing the number of produced units, cost per unit can be decreased linearly. A manufacturer records the number of units and cost per unit as follows:

# of units	10	20	50	100	150	200
Cost per unit	9.4	9.2	9.0	8.5	8.1	7.4

Round floating point numbers that are smaller than 0.01 to four decimal places.

(a) (13 pts) Predict the cost per unit when 125 units are produced.

Using simple linear regression

input: # of items
output: cost/item

$$S_{xx} = \sum x_i^2 - N\bar{x}^2 = 28683.3$$

$$S_{xy} = \sum x_i y_i - N\bar{x}\bar{y} = -285$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = -0.0099$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 9.478$$

Cost per item for 125 item

$$= -0.0099 \times 125 + 9.478 = 8.2$$

(b) (12 pts) Estimate the variance of the cost in part (a).

$$\sigma^2 = \frac{SSE}{N-2}$$

$$SSE = \sum_i (y_i - \hat{y}_i)^2 = 0.0282$$

$$\sigma^2 = \frac{SSE}{4} = 0.0071$$