BLG560E - Statistics and Estimation in Computer Science

Spring 19-20 Exam

July 8, 2020

Rules:

- Duration is 100 min.
- Open book, internet and notes
- You must solve these questions alone
- You can use scripts to solve questions
- Write your name and student id on the top right position in the first page of your solutions. Sign under your id. Insert page numbers in right bottom for all pages.
- Show your work, do not write the result directly.
- You can use lookup tables or use software (Excel, Python etc.).
- Upload your solutions to Ninova in a single pdf file. Please do not use any other format such as jpg, png etc.
- You have 20 extra minutes to scan your results into a single pdf file.
- You can use mobile apps such as Adobe Scan, Microsoft Office Lens, Camscanner etc.

Questions:

Each question is 20 pts.

1. Five brands of electric cars are tested for the distance they can go with a fully charged battery. A car from each brand is fully charged and used until their battery runs out. The total amount of distances are recorded as follows in kilometers for 4 trials with each brand.

	Trial 1	Trial 2	Trial 3	Trial 4
Brand A	176.46	156.00	164.68	183.61
Brand B	180.76	187.50	165.06	190.40
Brand C	174.37	140.82	142.08	133.91
Brand D	176.83	156.55	151.45	122.05
Brand E	150.76	157.50	135.06	160.40

   (a) Test the null hypothesis that claims all brands can travel same average distance with a fully charged battery. Use significance level of 0.05.

   (b) Find the p-value.

2. Recent research indicate that there might be a relation between social media usage (in hours/day) and body weight (kg). To test this hypothesis, a random sample of children is observed for their social media usage and their weight. Number of children with certain amounts of social media usage and weight are recorded in the following table.

	Underweight	Normal	Overweight	Obese
less than 1 hour/day	10	32	10	5
1 to 3 hours/day	9	25	15	8
more than 3 hours/day	12	20	18	10

   (a) Test if social media usage and weight are independent or not using significance level of 0.05.

   (b) Find the p-value.

3. It is claimed that average car accidents in İstanbul is 240 (per month) with standard deviation of 30 in 2019. For 2020, average car accidents between January and June are recorded as follows:

	Jan.	Feb.	Mar.	Apr.	May	June
accidents/month	255	244	215	254	241	238

   (a) Test the hypothesis that average car accidents per month increased in 2020 compared to 2019. Use significance level of 0.1.

   (b) Find the p-value.

(c) Find the critical value of average number of accidents per month that would lead to rejection of the null hypothesis.

4. In order to see the effects of age, iq (measure of intellectual ability), education (years) and network (number of friends) on monthly income a random sample of size 10 is collected and following values are obtained. It is suggested to use a script (Matlab, Python etc.) for at least matrix computations.

	age	iq	education	# of friends	income
Person 1	66	74	10	165	10515
Person 2	46	130	14	134	8576
Person 3	41	80	10	116	7762
Person 4	12	109	12	212	9169
Person 5	35	97	19	266	11979
Person 6	34	106	13	199	9905
Person 7	38	65	14	125	7888
Person 8	30	131	20	175	9006
Person 9	39	102	16	50	5679
Person 10	32	77	14	266	11832

(a) Find the most effective factor on monthly income amongst age, iq, education and network. Use R^2 metric to find this factor. Find the coefficient of the factor and intercept values.

(b) Find the most effective **two** factors on monthly income amongst age, iq, education and network. Use R^2 metric to find these two factors. Find the coefficients of the two factors and intercept values.

5. An instructor wanted to see if students would perform equally well on two different examinations. 12 students were selected. Six of the students were given examination A, and the other six examination B. On the next day the students were tested on the examination they had not yet taken. Thus, each of the 12 students took both examinations. The following pairs of scores were obtained by the students on the two examinations:

	Students											
Examination	1	2	3	4	5	6	7	8	9	10	11	12
A	763	419	586	920	881	758	262	332	717	909	940	835
B	797	404	576	855	762	707	195	341	728	817	947	849

(a) Test the following hypothesis using signed-rank test using significance level of 0.05.

H_0: The students perform equally well on exams A & B
H_1: The students' performance are different for exams A & B

(b) Find p-value using normal approximation and continuity correction