

1	2	3	4	Total

Name:
Number: *Answers*

BLG560E - Statistics and Estimation in Computer Science

Final Exam

31.05.2018

Rules:

- Duration is 90 min.
- Show your work, do not write the result directly.
- Use the attached distribution lookup tables if required.
- If you believe, critical information is missing in a question, make an assumption, and solve the question using this assumption.
- Solve each question within the corresponding frame. Anything outside the frame **will not be** graded.

Questions:

1. (25 pts) It is claimed that the amount of walking (steps) per day effects the cholesterol level. Following data is collected.

Group	Steps per day	n_i	\bar{x}_i
1	Less than 7500	25	188
2	More than 7500	16	181

In this table, n_i denotes the number of observations in the group, \bar{x}_i denotes the group mean.

It is known that population variance for both groups is equal to 100.

Use the data to test:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

where μ_i denotes the population mean for group i .

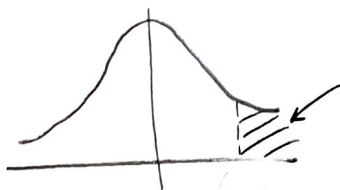
- 15 a) Write the formula of an appropriate test statistic T . Find its value for the given data.

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{100}{25} + \frac{100}{16}}} = \frac{7}{3.2} \approx 2.19$$

- b) Find the p-value for the observed data.

10

From z-table



$$1 - 0.9857 = 0.0143$$

$$p\text{-value} = 2 \times 0.0143 = 0.0286$$

2. (25 pts) A pharmaceutical company develops a new drug that is claimed to reduce cholesterol level. For testing its effectiveness, cholesterol level of four subjects were recorded before and after taking this new drug. Following data is recorded.

Subject	1	2	3	4
Before drug	320	280	324	305
After drug	300	310	334	310

Remember that the binomial pmf of rv X is:

$$P(X = k) = \binom{N}{k} p^k (1-p)^{N-k}$$

with expected value and variance are equal to Np and $Np(1-p)$ respectively.

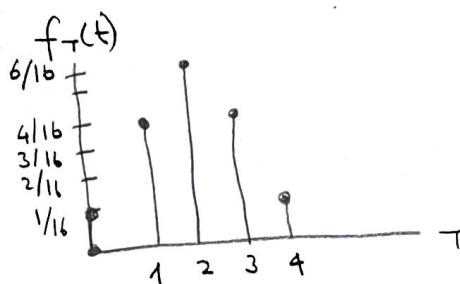
Let T be the number of observations where cholesterol level decreases.

- a) Find and draw the exact pmf of test statistics T for 4 subjects. Assume there is no equal levels of cholesterol level before and after drug. **Note:** This distribution is independent from the observed data.

20

1	2	3	4	T
-	-	-	-	4
-	-	-	+	3
-	-	+	-	3
-	-	+	+	2
-	+	-	-	3
-	+	-	+	2
-	+	+	-	2
-	+	+	+	1
+	-	-	-	3
+	-	-	+	2
+	-	+	-	2
+	-	+	+	1
+	+	-	-	2
+	+	-	+	1
+	+	+	-	1
+	+	+	+	0

T	0	1	2	3	4
occurrence	1	4	6	4	1



- b) Use sign test to find the exact p-value of the data.

5

For the given data

Subject 1 2 3 4 $\Rightarrow T=1$
 Sign - + + +

$$P = 2 \left(P(T=1) + P(T=0) \right) = \frac{10}{16} = \frac{5}{8}$$

1	2	3	4	Total

Name: Answers
Number: _____

3. (25 pts) A pharmaceutical company develops a new drug that lowers cholesterol level. For testing its effectiveness, cholesterol level four subjects were observed before and after taking this new drug.

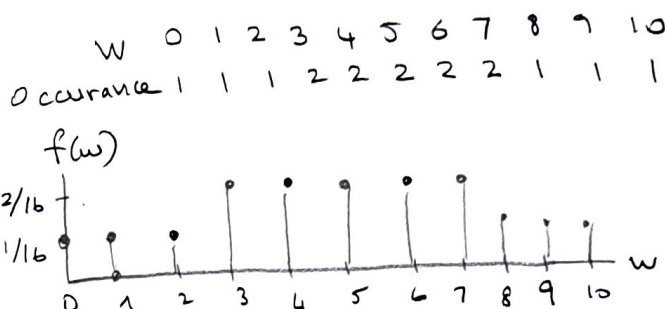
Subject	1	2	3	4
Before drug	320	280	324	305
After drug	300	310	334	310

Let W be the sum-of-ranks of observations where cholesterol level decreases.

- a) Find and draw the exact pmf of the test statistic W for four subjects. Assume there is no equal levels of cholesterol level before and after drug. **Note:** This distribution is independent from the observed data.

20

1	2	3	4	W
-	-	-	-	10
-	-	-	+	6
-	-	+	-	7
-	-	+	+	3
-	+	-	-	8
-	+	-	+	4
-	+	+	-	5
-	+	+	+	1
+	-	-	-	9
+	-	-	+	5
+	-	+	-	6
+	-	+	+	2
+	+	-	-	7
+	+	-	+	3
+	+	+	-	4
+	+	+	+	0



- b) Use sign test to find the exact p-value of the data.

5

For the observed data

Subject	1	2	3	4	sort				
diff.	+20	-30	-10	-5		-5	-10	+20	-30
					rank	1	2	3	4

$W = 7$

$$P = 2 \times P(W \geq 7) = 2 \times \frac{5}{16} = \frac{5}{8}$$

4. (25 pts) It is claimed that the amount of walking (steps) per day effects the cholesterol level. Following data is collected to test:

H_0 : No difference of cholesterol levels between groups.

H_1 : There is difference of cholesterol levels between groups.

Group	Steps per day	n_i	\bar{x}_i	s_i^2
1	Less than 7500	15	188	190
2	Between 7500-10000	10	181	211
3	More than 10000	5	174	202

In this table, n_i denotes the number of observations in a group, \bar{x}_i denotes the group mean, and s_i^2 denotes the group variance.

- a) Analyze the variance of 3 groups and find the value of F-statistics.

20

$$SSE = 14 \times 190 + 9 \times 211 + 4 \times 202 = 5,367 \quad d.o.f = 30 - 3 = 27$$

$$\bar{X} = (15 \times 188 + 10 \times 181 + 5 \times 174) / 30 = 183.33$$

$$\begin{aligned} SS_{Tr} &= 15 \times (188 - 183.33)^2 + 10 \times (181 - 183.33)^2 + 5 \times (174 - 183.33)^2 \\ &= 327.13 + 54.29 + 435.24 \\ &= 816.66 \quad d.o.f = 3 - 1 = 2 \end{aligned}$$

$$MSE = \frac{5367}{27} = 198.78$$

$$MSTr = \frac{816.66}{2} = 408.33$$

$$F = \frac{MSTr}{MSE} = \frac{408.33}{198.78} \approx 2.05$$

- b) For significance level of $\alpha = 0.05$, should you reject null hypothesis?

5

$$\text{For } \alpha = 0.05 \quad F_{2,27} = 3.35$$

As $F < F_{2,27}$ do NOT reject H_0 .