1	2	3	4	Total

Name:
Number: Answers

BLG560E - Statistics and Estimation for Computer Science

Spring 2021-2022
Midterm 2

11.05.2022

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) A researcher wants to determine if there is a significant difference of test scores between left-handed and right-handed students. In this research, null hypothesis claims student groups have equal mean test scores.

   Student scores are as follows:
   Left handed: 12, 10, 12, 14, 12, 10, 8
   Right handed: 8, 10, 10, 12, 11, 6, 7

   Assume both left and right handed students' score variances are known to be 4.

   (a) (15 pts) Should null hypothesis be rejected or retained at 0.05 significance level?

Since population variances are known

$$\bar{X}_1 - \bar{X}_2 \sim N\left(0, \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}\right) \quad \text{if } H_0 \text{ is correct}$$

then $$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \sim N(0,1)$$

$\bar{x}_1 = 11.14$
$\bar{x}_2 = 9.14$

$$Z = \frac{11.14 - 9.14}{\sqrt{\frac{4}{7} + \frac{4}{7}}} = 1.87$$

$Z_{\alpha/2} = 1.96$

Since $Z < Z_{\alpha/2}$

H_0 should be retained

   (b) (10 pts) Find the p-value.

$$\text{p-value} = 2\left(1 - \Phi(1.87)\right) = 0.06$$

2. (25 pts) Assume that you would like to test if people with high income prefers newspaper A to newspaper B. Null hypothesis claims income level and newspaper preference are independent.

(a) (5 pts) For this test, a random sample of 40 subjects is selected. According to the answers of these subjects

- 23 subjects were low income
- 7 subjects from high income indicate that they prefer newspaper B.
- 8 subjects of low income indicate that they prefer newspaper A.

Form the contingency table for this test.

Observed	A	B	Total
Low income	8	15	23
High income	10	7	17
Total	18	22	40

(b) (10 pts) Should this hypothesis be rejected at 0.05 significance level? Do not apply Yates correction.

Expected values

	A	B	Total
Low income	10.35	12.65	23
High income	7.65	9.35	17
Total	18	22	40

$$X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = 2.28$$

$$X^2_{c,0.05} = 3.84$$

Since $X^2 < X^2_{c,0.05}$ H_0 should be retained.

(c) (10 pts) Compute p-value using Yates correction. Use cumulative distribution function (cdf) $F_{\chi^2}()$ for p-value (no need to compute/lookup p-value).

$$X^2 = \sum_i \frac{(|O_i - E_i| - 0.5)^2}{E_i} = 1.41$$

$$\text{p-value} = 1 - F^{-1}_{\chi^2}(1.41)$$

$$= 0.23$$

1	2	3	4	Total

Name:
Number:

3. (25pts) A researcher wants to determine if the variance test scores of left-handed students is smaller than 4. Left handed student scores are as follows:
Left handed: 12, 10, 12, 14, 12, 10, 8

(a) (15 pts) Test $H_0 : \sigma^2 < 4$ with significance level of 0.01.

Use

$$T = \frac{(N-1)s^2}{\sigma^2} \sim \chi^2_{N-1}$$

$$s^2 = 3.81 \Rightarrow T = \frac{(7-1)\,3.81}{4} = 5.71$$

$T_{c,0.01} = 16.81$ one-sided test and upper tail is considered

Since $T < T_{c,0.01}$ H_0 should be retained

(b) (10 pts) Compute the corresponding p-value. Use inverse cumulative distribution function (cdf) $F^{-1}_{\chi^2}()$ for p-value (no need to compute/lookup p-value).

$$p = 1 - F_{\chi^2_6}(5.71) = 0.46$$

4. (25 pts) Circle the correct answer and **explain your answer very briefly** inside the frames under the choices.

(a) (5 pts) With extremely small size samples, the probability of rejecting null hypothesis (when null hypothesis is correct)

Decreases	Does not change	Increases

probability of rejecting H_0, when H_0 is correct $= \alpha$
This is a selected value and does not change with sample size

(b) (5 pts) By increasing significance level, the probability of rejecting null hypothesis (when null hypothesis is correct)

Decreases	Does not change	Increases

significance level $= \alpha =$ prob. of rejecting H_0 when H_0 is correct

(c) (5 pts) When size sample increases, the probability of type-1 error

Decreases	Does not change	Increases

prob of type 1 error $= \alpha$ regardless of sample size

(d) (5 pts) When size sample increases, the test power

Decreases	Does not change	Increases

test power $= 1 - \beta$
as $N \uparrow$ $\beta \downarrow$ $1-\beta \uparrow$

(e) (5 pts) When significance level decreases, the probability of type-2 error

Decreases	Does not change	Increases

$\alpha \downarrow$ $\beta \uparrow$ $1-\beta \downarrow$