**פתרון שאלות 4 ו 5 משיעורי הבית**

- הניסוח של השאלות מעט שונה אך השאלות זהות

<u>תרגיל 1</u>

בקפיטריה של הפקולטה להנדסה יש בד"כ תור ליד הקופה. התפלגות זמן ההמתנה היא מעריכית עם פרמטר $\theta$. בעל הקפיטריה טוען כי זמן ההמתנה הממוצע הוא 5 דקות ואילו הסטודנטים טוענים כי זמן ההמתנה הוא 10 דקות (ומכלה את כל משך ההפסקה).

אחד הסטודנטים החליט לבדוק את הטענות על סמך זמן ההמתנה שלו. הוא קבע כי אם ימתין 10 דקות ומעלה אז זמן ההמתנה הממוצע הוא אכן עשר דקות.

א. נסחו את ההשערות במונחי התוחלת, ציינו באופן מפורש את אזור הדחייה של הסטודנט וחשבו את ההסתברויות לטעות מסוג ראשון ושני.

נסמן זמן ההמתנה $X_{waiting} \sim exp(\theta)$ . תוחלתו שווה ל- $\frac{1}{\theta}$.

- ניסוח ההשערות:

$$H_0:\ \mu = 5 \ \rightarrow \ \theta = \frac{1}{5}$$
$$H_1:\ \mu = 10 \ \rightarrow \ \theta = \frac{1}{10}$$

- זהו מבחן השערות על התוחלת. המבחן הינו חד-צדדי ימני .
- סטטיסטי המבחן: זמן ההמתנה של סטודנט אחד. מתפלג $X_1 \sim exp(\theta)$
- ההסתברות לטעות מסוג ראשון-

$$\alpha = P_{H_0}(C)$$
$$C = \{X_1 / X_1 > 10\}$$
$$H_0: \theta = 0.2$$
$$X_1 \sim exp(0.2)$$
$$P(X_1 > a) = e^{-\theta \cdot a} \Rightarrow P(X_1 > 10) = e^{-0.2 \cdot 10} = e^{-2} = 0.135$$

- ההסתברות לטעות מסוג שני -

$$\beta = P_{H_1}(\underline{C})$$
$$\bar{C} = \{X_1 / X_1 \leq 10\}$$
$$H_1: \theta = 0.1$$
$$X_1 \sim exp(0.1)$$
$$P(X_1 \leq a) = 1 - e^{-\theta \cdot a} \Rightarrow P(X_1 \leq 10) = 1 - e^{-0.1 \cdot 10} = 1 - e^{-1} = 0.632$$

עוצמת המבחן: $1-\beta=0.368$

ב. בעל הקפיטריה טוען כי מדגם של סטודנט אחד אינו מייצג, ורק אם יתקבל זמן המתנה הגדול מ-15 דקות יהיה ניתן לדחות את טענתו. חשבו את ההסתברות לטעות מסוג ראשון ושני וחוו דעתכם.

מערכת ההשערות לא השתנתה.

גם סטטיסטי המבחן נשאר זהה - זמן ההמתנה של סטודנט יחיד.

אזור הדחייה השתנה: נדחה את השערת האפס רק אם זמן ההמתנה המתקבל גדול מ-15 דקות.

- ההסתברות לטעות מסוג ראשון-

$$\alpha = P_{H_0}(C)$$
$$C = \{X_1 / X_1 > 15\}$$

$$H_0: \theta = 0.2$$

$$X_1 \sim exp(0.2)$$

$$P(X_1 > a) = e^{-\theta \cdot a} \Rightarrow P(X_1 > 15) = e^{-0.2 \cdot 15} = e^{-3} = 0.049$$

- הסתברות לטעות מסוג שני:

$$\beta = P_{H_1}(\underline{C})$$
$$\bar{C} = \{X_1 / X_1 \leq 15\}$$

$$H_1: \theta = 0.1$$

$$X_1 \sim exp(0.1)$$

$$P(X_1 \leq a) = 1 - e^{-\theta \cdot a} \Rightarrow P(X_1 \leq 15) = 1 - e^{-0.1 \cdot 15} = 1 - e^{-1.5} = 0.7769$$

עוצמת המבחן: $1\text{-}\beta=0.2231$

<u>פרשנות:</u> לפי הצעתו של בעל הקפיטריה, ההסתברות לטעות מסוג ראשון קטנה אך ההסתברות לטעות מסוג שני גדלה. זה מתאים לעבודה שקשה יותר לדחות את השערת האפס (אזור הדחייה הצטמצם). במקרה כזה, ההחלטה על המבחן הטוב יותר היא בידי החוקר. בכל מקרה, מדגם של זמן המתנה אחד אכן <u>אינו מייצג</u>, ונדרש להגדילו.

1. על מנת ליישב את המחלוקת, החליט בעל הקפיטריה ביחד עם ועד הנדסה כי 100 סטודנטים ימתינו בתור בקפיטריה (לא בו זמנית אלא בניסויים נפרדים כמו בסעיף א' ו-ב'). סטטיסטי המבחן יהיה ממוצע זמני ההמתנה של כל 100 הסטודנטים. בהנחה שזמני ההמתנה של סטודנטים שונים הם ב"ת, מצאו את אזור הדחייה ואת ההסתברות לטעות מסוג ראשון, כאשר הועד דורש מבחן שעוצמתו 95%.

- מערכת ההשערות לא השתנתה.
- סטטיסטי המבחן השתנה - ממוצע של 100 זמני המתנה ב"ת.

עבור התפלגות המעריכית:

$$E(X_{waiting}) = \frac{1}{\theta} \quad V(X_{waiting}) = \frac{1}{\theta^2}$$

n>30 ולכן מקבלים ממשפט הגבול המרכזי:

$$\bar{X} \sim N\left(\frac{1}{\theta}, \frac{\frac{1}{\theta^2}}{n}\right) \rightarrow \bar{X} \sim N\left(\frac{1}{\theta}, \frac{1}{n\theta^2}\right)$$

- אזור הדחייה הוא מהצורה: $C = \{\bar{X} / \bar{X} \geq k\}$

על מנת למצוא את הערך הקריטי $k$ שמגדיר את אזור הדחייה, עלינו להיעזר בדרישה למבחן שעוצמתו 95%.

המשוואה שיש לפתור:

$$1 - \beta = P_{H_1}(C) = P_{H_1}(\underline{X} \geq k)$$

תחת $H_1$, התוחלת והשונות של כל תצפית הן: $E\left(X_{waiting}\right) = 10, \qquad V\left(X_{waiting}\right) = 100$

התוחלת והשונות של הממוצע הן: $E\left(\underline{X}\right) = 10, \qquad V\left(\underline{X}\right) = \frac{\sigma^2}{n} = \frac{100}{100} = 1$

ולכן:

$$P_{H_1}(C) = 1 - \Phi\left(\frac{k - 10}{1}\right) = 0.95 \ \rightarrow \ \Phi\left(\frac{10 - k}{1}\right) = 0.95$$

$$\frac{10 - k}{1} = z_{0.95} = 1.645 \ \rightarrow \ k = 8.355$$

אזור הדחייה: $\boldsymbol{C = \{\bar{X} / \bar{X} \geq 8.355\}}$

- ההסתברות לטעות מסוג ראשון:

$$\alpha = P_{H_0}(C) = P_{H_0}(\underline{X} \geq 8.355)$$

תחת $H_0$, התוחלת והשונות של כל תצפית הן: $E\left(X_{waiting}\right) = 5, \qquad V\left(X_{waiting}\right) = 25$

התוחלת והשונות של הממוצע הן: $E\left(\underline{X}\right) = 5, \qquad V\left(\underline{X}\right) = \frac{\sigma^2}{n} = \frac{25}{100} = 0.25$

$$\alpha = P_{H_0}\left(\underline{X} \geq 8.355\right) = 1 - \Phi\left(\frac{8.355 - 5}{\sqrt{0.25}}\right) = 1 - \Phi(6.71) \approx 1 - 1 = 0$$

**תרגיל 2**

במסגרת מבצע לטיפוח החזות של דירות בעיר נערכה ספירה של מספר העציצים במרפסות של מדגם מקרי של דירות בשתי שכונות בעיר. להלן התוצאות שנמצאו במדגם :

| | גודל מדגם | ממוצע | סטיית תקן |
|---|---|---|---|
| **שכונה א׳** | 15 | 7.2 | 4.1 |
| **שכונה ב׳** | 20 | 12.5 | 5.5 |

האם ניתן לומר כי מספר העציצים הממוצע בשכונה ב׳ גדול ביותר מ-1 ממספר העציצים הממוצע בשכונה א׳?
הנח התפלגות נורמלית של הנתונים ורמת מובהקות של 10%.

**תשובה**

$\mu_1$ - תוחלת מספר העציצים לדירה בשכונה א׳. $\mu_2$ - תוחלת מספר העציצים לדירה בשכונה ב׳.

- מערכת ההשערות : $H_0: \mu_2-\mu_1=1$ , $H_1: \mu_2-\mu_1>1$
- מדובר במדגמים ב״ת ושונויות לא ידועות. נבצע מבחן לשוויון שונויות. $H_0: \sigma_1^2 = \sigma_2^2$ , $H_1: \sigma_1^2 \neq \sigma_2^2$
- כלל החלטה : דחה אם $F > f_{1-\frac{\alpha}{2}}^{n_1-1,n_2-1}$ או $F < f_{\frac{\alpha}{2}}^{n_1-1,n_2-1}$, כאשר $F=\frac{s_1^2}{s_2^2}$.

  הערך של סטטיסטי המבחן הינו : $F=\frac{4.1^2}{5.5^2}=0.556$

##### לא רלבנטי #####

נדרשים הערכים הקריטיים $f_{0.95}^{14,19}$ וכן $f_{0.05}^{14,19}$. הם לא מופיעים בטבלה שבידינו, ולכן נשתמש בטיעון הבא :

לפי כלל הדחייה הנתון, מקבלים את $H_0$ אם $f_{\frac{\alpha}{2}}^{n_1-1,n_2-1} \le F \le f_{1-\frac{\alpha}{2}}^{n_1-1,n_2-1}$.

הגבול העליון מקיים : $2.23 = f_{0.95}^{(15,19)} > f_{0.95}^{(14,19)} = f_{1-\frac{\alpha}{2}}^{n_1-1,n_2-1}$. כלומר, הגבול העליון נמוך מ-2.23.

הגבול התחתון מקיים : $0.418 = \frac{1}{2.39} = \frac{1}{f_{0.95}^{(20,14)}} < \frac{1}{f_{0.95}^{(19,14)}} = f_{0.05}^{(14,19)} = f_{\frac{\alpha}{2}}^{n_1-1,n_2-1}$. כלומר, הגבול התחתון נמוך מ-0.418.

כלומר, אנחנו לא יודעים בדיוק מהו אזור הקבלה האמיתי – אבל ברור שהתחום [0.418 , 2.23] כלול בו.
והפלא ופלא – הערך של סטטיסטי המבחן נופל בתוך התחום הזה, ולכן באזור הקבלה.

####################

מסקנה: לא דוחים את $H_0$, כלומר השונויות שוות ברמת מובהקות 10%, ולכן ודאי גם ברמת מובהקות 5%.

- כלל ההחלטה עבור הפרש התוחלות עם שונויות לא ידועות אך שוות (מקרה 2):

דחה אם $\bar{y}-\bar{x} > d_0 + t_{1-\alpha}^{(n_1+n_2-2)} s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

שימו לב שההשערות מנוסחות כתוחלת ב' פחות תוחלת א', ולכן זהו גם הסדר שבו מופיעים הממוצעים (ממוצע ב' פחות ממוצע א').

$$s_p^2 = \frac{(15-1)4.1^2+(20-1)5.5^2}{15+20-2} = 24.54$$

דחה אם $\bar{y}-\bar{x} > 1 + t_{0.9}^{33}\sqrt{24.54}\sqrt{\frac{1}{15}+\frac{1}{20}} = 3.22$.

מתקיים: $\bar{y}-\bar{x} = 12.5-7.2 = 5.3$.

נדחה את השערת האפס ברמת מובהקות 10%. מספר העציצים הממוצע בשכונה ב' אכן גדול ביותר מ-1 מזה של שכונה א'.