<u>תרגיל 1</u>

נתונות 16 תצפיות מאוכלוסייה שמתפלגת נורמלית עם שונות 100:

23.2	**10.6**	**-1.5**	**-0.7**	**21.8**	**11.5**	**4**	**4.3**	**20.6**	**11.4**	**17.1**	**15.9**	**19.5**	**16**	**-1.4**	**15.7**

השתמש בממוצע המדגם $\bar{X}$ ע"מ לבדוק את ההשערה (H_0) כי $\mu=8$ לעומת ההשערה האלטרנטיבית (H_1) כי $\mu=12$. רמת המובהקות הנדרשת היא 1%.

1. נסחו את מערכת ההשערות. האם זו מערכת השערות חד-צדדית או דו-צדדית?

ניסוח ההשערות:

$$H_0: \mu=8$$
$$H_1: \mu=12$$

שימו לב: האלטרנטיבה $\mu = 12$ משמעה בעצם $\mu > 8$. לכן זוהי מערכת השערת חד-צדדית ימנית! אם הייתה בידינו מערכת השערות דו"צ, השערת האלטרנטיבה הייתה צריכה להיות $H_1: \mu \neq 8$.

2. מה צריכים להיות איזורי הדחייה והקבלה?

בהמשך לסעיף א', אם ממוצע המדגם יהיה מספיק גבוה, נאמר שיש הוכחה מספיק חזקה לדחיית H_0. לכן אזור הדחייה יהיה מהצורה: $kC = \{\underline{X} \geq k\}$. נקרא "הערך הקריטי" ועלינו למצוא אותו, כך שיתאים לדרישה של רמת מובהקות 0.01.

האוכלוסייה מתפלגת נורמלית ולכן:

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \rightarrow \bar{X} \sim N(\mu, \frac{100}{16})$$

תחת H_0 מתקיים:

$$H_0: \bar{X} \sim N(8, \frac{100}{16})$$

$$\alpha = P_{H_0}(C) = P_{H_0}\left(\underline{X} \geq k\right) = 1 - \Phi\left(\frac{k-8}{\sqrt{\frac{100}{16}}}\right) = 0.01$$

$$\Phi\left(\frac{k-8}{\sqrt{\frac{100}{16}}}\right) = 0.99 \ \rightarrow \ \frac{k-8}{\sqrt{\frac{100}{16}}} = Z_{0.99} = 2.326 \ \rightarrow \ k = 13.815$$

$$\alpha = P(C / H_0) = 0.01$$

אזור הדחייה: $\boxed{C = \left\{ \bar{X} \; / \; \bar{X} \geq 13.815 \right\}}$, אזור הקבלה: $\boxed{\bar{C} = \left\{ \bar{X} \; / \; \bar{X} < 13.815 \right\}}$

"הערך הקריטי" הוא 13.815 – מעליו איזור הדחייה ומתחתיו איזור הקבלה.

3. מהי עוצמת המבחן?

$$1 - \beta = P_{H_1}(C) = P_{H_1}\left(\underline{X} \geq 13.815\right) = 1 - \Phi\left(\frac{13.815 - 12}{\sqrt{\frac{100}{16}}}\right) = 1 - \Phi(0.726) = 1 - 0.7642$$
$$= 0.2358$$

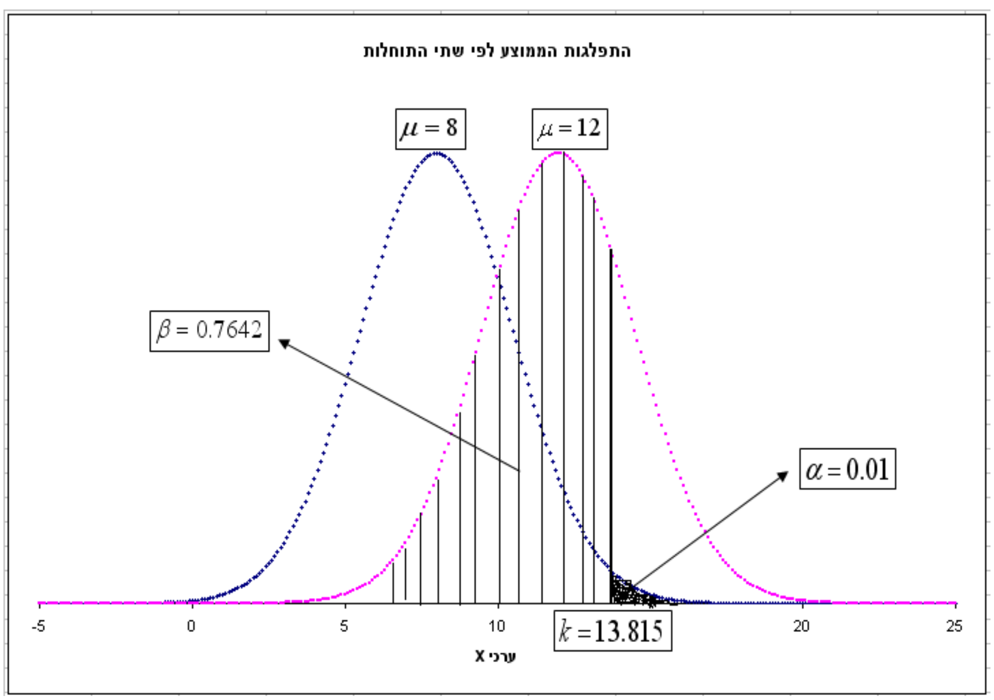
שימו לב שתחת H_1 התוחלת של הממוצע הינה 12!

4. מהי מסקנת המחקר?

נחשב את ממוצע המדגם, עפ"י התצפיות הנתונות בשאלה: $\bar{X} = \frac{\sum_{I=1}^{16} X_i}{16} = 11.75 < 13.815$

הממוצע קטן מהערך הקריטי, ולכן אנו נמצאים באזור הקבלה (של H_0).

מסקנה: לא ניתן לדחות את H_0, כלומר לא ניתן לקבוע כי $\mu = 12$ ברמת מובהקות 1%.



<u>תרגיל 2</u>

בקפיטריה של הפקולטה להנדסה יש בד"כ תור ליד הקופה. התפלגות זמן ההמתנה היא מעריכית עם פרמטר θ. 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 (**α):**

$$\alpha = P_{H_0}(C)$$

$C = \{X_1 / X_1 > 10\}$

$H_0 : \theta = 0.2$

$X_1 \sim \exp(0.2)$

$P(X_1 > a) = e^{-\theta \cdot a} \Rightarrow P(X_1 > 10) = e^{-0.2 \cdot 10} = e^{-2} = 0.135$

- ההסתברות לטעות מסוג שני (**β):**

$$\beta = P_{H_1}(\underline{C})$$

$\bar{C} = \{X_1 / X_1 \le 10\}$

$H_1 : \theta = 0.1$

$X_1 \sim \exp(0.1)$

$P(X_1 \le a) = 1 - e^{-\theta \cdot a} \Rightarrow P(X_1 \le 10) = 1 - e^{-0.1 \cdot 10} = 1 - e^{-1} = 0.632$

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.

- הסתברות 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:

$$\beta = P_{H_1}(\underline{C})$$

$$\bar{C} = \{X_1 / X_1 \le 15\}$$
$$H_1 : \theta = 0.1$$
$$X_1 \sim \exp(0.1)$$
$$P(X_1 \le a) = 1 - e^{-\theta \cdot a} \Rightarrow P(X_1 \le 15) = 1 - e^{-0.1 \cdot 15} = 1 - e^{-1.5} = 0.7769$$

עוצמת המבחן: $1-\beta=0.2231$

<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.

5. 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: $\bar{X} \sim N\left(\frac{1}{\theta}, \frac{1/\theta^2}{n}\right) \to \bar{X} \sim N\left(\frac{1}{\theta}\ ,\ \frac{1}{n\theta^2}\right)$

- אזור הדחייה הוא מהצורה: $C = \left\{ \bar{X} \,/\, \bar{X} \ge k \right\}$

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

המשוואה שיש לפתור:

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

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

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

ולכן:

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

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

אזור הדחייה: $\boxed{C = \left\{ \bar{X} \,/\, \bar{X} \ge 8.355 \right\}}$

- ההסתברות לטעות מסוג ראשון:

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

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

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

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