ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

<u>2.8.12-הרצאה 3</u>

<u>מקרה 1 : בדיקת השערות לתוחלת האוכלוסייה כאשר השונות ידועה :</u>

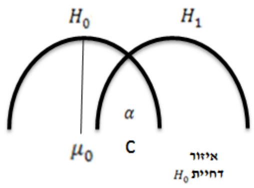
בבדידה השערות קיימות שתי השערות :

- H_0, השערת בסיס, השערת האפס, מה שהיה ידוע ומקבול עד כה
- H_1, השערת אלטרנטיבית, השערת חוקר/השערה חדשה

עלינו להחליט מי מבין שני ההשערות מתאימה יותר.

<u>3 סוגי מבחנים :</u>

1) מבחן חד כיווני ימני :

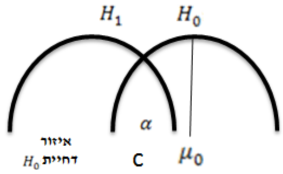


$H_0: \mu = \mu_0$ השערת בסיס.

$H_1: \mu > \mu_0$ השערת חוקר.

$$C = \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$$

2) מבחן חד כיווני שמאלי :

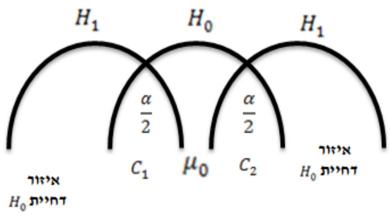


$H_0: \mu = \mu_0$ השערת בסיס.

$H_1: \mu < \mu_0$ השערת חוקר

$$C = \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}.$$

3) מבחן דו צדדי :



$H_0: \mu = \mu_0$ השערת בסיס.

$H_1: \mu \neq \mu_0$ השערת חוקר.

$$C_2 = \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, C_1 = \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

<u>הטעויות הסטטיסטיות בבדיקה השערות :</u>

<u>מצב בפועל</u>

	H_0 נכונה	H_1 נכונה
דוחים H_0	α טעות מסוג ראשון	$1-\beta$ עצמת המבחן
דוחים H_1	$1-\alpha$	β טעות מסוג שני

השאיפה היא שהטעויות β, α יהיו כמה שיותר קטנות, במידת האפשר.

בין β, α קיים קשר הפוך.

- ככל שα גדל, β קטן ו$1-\beta$ גדל
- ככל שקטן , β גדל ו$1-\beta$ קטן

דוגמה :

ידוע שממוצע הכנסה חודשי הוא 6000 שח עם סטיית תקן של 500 שח. חוקר טוען כי ממוצע ההכנסה בשנה האחרונה עלה. לצורך בדיקת הטענה נלקח מדגם מקרי של 100 אנשים, ונמצא כי הממוצע 6125.

א) כתבו השערות

ב) בדקו את ההשערות ברמת מובהקות של 5%.

ג) מה מסקנתכם עבור רמת מובהקות של 10%? האם יש צורך בחישוב נוסף?

פתרון :

א) $H_0: \mu = 6000$. $H_1: \mu > 6000$.

ב) $Z_{0.95} = 1.645$, $\alpha = 5\%$, $C = \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$, מכאן ש $C = 6000 + 1.645 \cdot \frac{500}{\sqrt{100}}$
מתקבל כי $C = 6082.25$ ולכן דחינו את H_0 ומסיקים את H_1.

ג) אין צורך בחישוב נוסף כי אם דחינו ב5% ברור שנדחה גם ב10%.

<u>הגדרה</u>: p. Value (α) : זוהי ה α המינימאלית עבורה דוחים את H_0 . לכל α שגדולה מ p.v. דוחים את H_0.

במבחן ימני : $p.v = p(z > \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}})$

במבחן שמאלי: $p.v = p(z < \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}})$

במבחן דו צדדי: $p.v = 2p.v.$ לצד המתאים.

בהמשך לדוגמה, חשבו את p.v. :

$$p.v = p\left(z > \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right) = p\left(z > \frac{6125-6000}{500/\sqrt{100}}\right) = p(z > 2.5) = 1 - p(z \le 2.5) = 1 - \Phi(2.5) = 0.0062$$

עבור $\alpha > 0.62\%$ דוחים את H_0.

<u>מקרה 2</u> : רווח סמך ובדיקת השערות לתוחלת μ כאשר השונות באוכלוסייה לא ידועה.

במקרה זה בו השונות באוכלוסיה לא ידועה נצטרך להשתמש במקום בשונות של המדגם :
$S^2 = \frac{\sum_{i=1}^{n}(x-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}$. בנוסף, במקום בטבלת ההתפלגות הנורמאלית הסטנדרטית נשתמש בהתפלגות t, כאשר התפלגות t היא התפלגות סימטרית שטוחה יותר בהשוואה להתפלגות נורמאלית.
התפלגות t תלויה בגודל המדגם.

רווח סמך : $\bar{x} - t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}} \le \mu \le \bar{x} + t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}$, את $Z_{\bar{x}}$ מחליף $t_{\bar{x}} = \frac{\bar{x}-\mu_0}{\hat{s}/\sqrt{n}}$.

דוגמה :

משקל של 7 קופסאות מסוימות נתון : 100,98,104,104,98,102,96. בהנחה שמספר הקופסאות מתפלג נורמאלית, בנו רווח סמך לתוחלת ברמת בטחון של 99%.

פתרון :

מאחר והשונות לא ידועה, נחשב את הערכים הבאים : $\bar{x} = \frac{100+\cdots+96}{7} = 100$, $\hat{s}^2 = \frac{104^2+\cdots-7\cdot 100^2}{7-1} = 8$, ומכאן נקבל כי $\hat{s} = \sqrt{8} = 2.83$. רוצים רמת בטחון של 99%, לכן נחשב את $t_{7-1,1-\frac{0.01}{2}} = t_{6,0.995} = 3.707$

נציב הכל ונקבל כי: $100 - 3.707 \cdot \frac{2.83}{\sqrt{7}} \le \mu \le 100 + 3.707 \cdot \frac{2.83}{\sqrt{7}}$, ולכן $96.034 \le \mu \le 103.965$ ברמת בטחון של 99%.

בבדיקת השערות :

במבחן ימני : $p.v = p(t_{n-1} > \frac{\bar{x}-\mu_0}{\hat{s}/\sqrt{n}})$

במבחן שמאלי: $p.v = p(t_{n-1} < \frac{\bar{x}-\mu_0}{\hat{s}/\sqrt{n}})$

במבחן דו צדדי: $p.v = 2p.v.$ לצד המתאים.

דוגמה :

בעיר מסוימת ידוע שהציון הממוצע במתמטיקה מתפלג נורמאלית עם $\mu = 70$. מנהל בית הספר בעיר טוען שהממוצע גבוה יותר. לצורך כך נלקח מדגם מקרי של 8 תלמידים , להלן ציוניהם :
68,71,79,76,81,69,72,66

א) כתבו השערות

ב) מצאו את p.v.

פתרון :

א) $H_0: \mu = 70$. $H_1: \mu > 70$.

ב) $\bar{x} = 72.75$, $\hat{s} = 5.391$, $p.v = p\left(t_7 > \frac{72.75-70}{5.391/\sqrt{8}}\right) = p(t_7 > 1.44)$

$0.9 \le 1 - p.v. \le 0.95$, $0.05 \le p.v. \le 0.1$. מתחת לגבול תחתון, לא דוחים את H_0. בין לגבולות, לא ידוע, ומעל לגבול העליון בטוח דוחים את השערת הבסיס. מסקנה : לככל α שגדולה מ10% דוחים את H_0.

<u>הקשר בין רווח סמך לבדיקת השערות</u> :

לפעמים ניתן לדעת אם דוחים את H_0 או לא לפי רווח סמך כאשר יש התאמה ברמת המובהקות α . אם μ_0 נמצא בגבולות בתוך גבולות רווח הסמך אזי לא דוחים את H_0, ואם הוא נמצא מחוץ לגבולות, ניתן לדחות את השערת הבסיס.

דוגמה :

ברמת בטחון של 96% התקבל רו״ס: $76 \le \mu \le 84$. לפי רו״ס זה בדקו האם נדחה H_0 במקרים הבאים :

א) $H_0: \mu = 82$ $H_1: \mu \neq 82$, $\alpha = 4\%$

ב) $H_0: \mu = 86$ $H_1: \mu \neq 86$, $\alpha = 6\%$

פתרון :

א) השערת הבסיס נמצאת בתוך גבולות הרו״ס ולכן לא דוחים אותה. (רווח הסמך הוא ברמת בטחון 96%)

ב) אם ב4% 86 מחוץ לגבולות רווח הסמך, ברור שב6% זה גם כן יקרה, ולכן H_0 נדחה.

<u>נוסחה חשובה</u> : (כאשר השונות באוכלוסייה ידועה)

$n \ge \left(\frac{\sigma(Z_{1-\alpha}+Z_{1-\beta})}{\mu_1-\mu_0}\right)^2$ מציאת גודל מדגם מינימאלי שחוקר צריך לקחת על מנת להגביל את גודל הטעויות α, β.

דוגמה :

מהו גודל המדגם המינימאלי שחוקר צריך לקחת כאשר $\sigma = 200$ ומעוניין ש$\alpha < 5\%, \beta < 1\%$. עבור ההשערות $H_0: \mu = 3500$ $H_1: \mu = 3600$

פתרון :

נציב בנוסחא $n \ge \left(\frac{\sigma(Z_{1-\alpha}+Z_{1-\beta})}{\mu_1-\mu_0}\right)^2$, $n \ge \left(\frac{200(Z_{0.95}+Z_{0.99})}{3600-3500}\right)^2$, $n \ge 63.07 \approx 64$

<u>מקרה 3 : בדיקת השערות ורו״ס למדגמים תלויים=מזווגים</u> :

במדגמים אלו לוקחים 2 קבוצות התלויות אחת בשנייה. נגדיר 2 סדרות, x,y כאשר סדרת ההפרשים = d.

אני נעסוק בתוחלת ההפרשים, μ_d. $\left(\bar{d} = \frac{\sum_{i=0}^{n} d_i}{n}\right)$

רו"ס ל-μ_d מוגדר כך: $\bar{d} - t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\hat{s}d}{\sqrt{n}} \le \mu_d \le \bar{d} + t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\hat{s}d}{\sqrt{n}}$.

ועבור בדיקת השערות: $H_0: \mu_d = 0$ $H_1: \mu_d > 0, < 0, \neq 0$ בהתאם לסוג המבחן.

מניחים שסדרת ההפרשים $d_i \sim N(\mu_d, \hat{\sigma}_d^2)$, כאשר $\hat{\sigma}_d^2 = \frac{\sum_{i=1}^{n}(d_i - d)^2}{n-1}$, ונסמן גם $t_{\bar{d}} = \frac{\bar{d} - d_0}{\hat{s}_d / \sqrt{n}} \sim t_{n-1}$.

ימני: $p.v. = p(t_{n-1} > \frac{\bar{d} - d_0}{\hat{s}_d / \sqrt{n}})$, שמאלי: $p.v. = p(t_{n-1} < \frac{\bar{d} - d_0}{\hat{s}_d / \sqrt{n}})$. דו"צ: p.v.=2p.v. לצד המתאים.

דוגמה:

רופא סיני טוען כי בידו טיפול המפחית עישון. להלן נתוני מטופליו על מס' הסיגריות היומיות לפני ואחרי הטיפול.

מטופל	1	2	3	4	5	6
X לפני	10	20	15	20	30	40
Y אחרי	5	10	10	10	15	10
d=x-y	5	10	5	10	15	30

א) נסחו את ההשערות ומצאו את $p.v.$:

ב) בנו רו"ס לממוצע השינוי בכמות הסיגריות ברמת בטחון של 98%:

פתרון:

א) $H_0: \mu_d = 0$, $H_1: \mu_d > 0$. $\bar{d} = \frac{30+15+\cdots+5}{6} = 12.5$. $\hat{S}_d = 9.35$. $p.v. = p(t_5 > \frac{12.5-0}{9.35/\sqrt{6}})$, $p.v. = p(t_5 > 3.27)$. מכאן ש $97.5\% \le 1 - p.v. \le 99\%$, ולכן $1\% \le p.v. \le 2.5\%$. לכל α שגדול מ2.5% נקבל שהטיפול אכן יעיל.

ב) $1 - \frac{\alpha}{2} = 0.99$. $t_{5,99\%} = 3.365$. $\bar{x} - t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}} \le \mu \le \bar{x} + t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}$, נציב ונקבל כי: $-0.34 \le \mu \le 25.34$, ברמת בטחון של 98%.

<u>מקרה 3: בדיקת השערות ורו"ס לפרופורציה (p) באוכלוסייה</u>

נגדיר: $\hat{p} = \frac{x}{n}$, ונסמן $1 - p = q$.

נניח כי $\hat{P} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$, וגם כי $Z_{\hat{p}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$.

רו"ס: $\hat{p} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

בדיקת השערות: $H_0: P = p_0$ $H_1: P, >, <, \neq, p_0$ בהתאם לסוג המבחן.

ימני: $p.v. = p(Z > \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}})$, שמאלי: $p.v. = p(Z < \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}})$. דו"צ: p.v.=2p.v. לצד המתאים.

מציאת גודל מדגם מינימאלי:

1) להגבלת אורך רווח הסמך (L): $n \geq \left(\frac{2Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\hat{p}(1-\hat{p})}}{L}\right)^2$. הערה: כשהפרופורציה במדגם לא ידועה לוקחים p=0.5.

2) להגבלת הטעויות α, β, כאשר $H_0: P = P_0$ $H_1: P = P_1$.
$n \geq \left(\frac{\sqrt{p_0(1-p_0)} \cdot Z_{1-\alpha} + \sqrt{p_1(1-p_1)} \cdot Z_{1-\beta}}{P_1 - P_0}\right)^2$. ולהציב ☺

דוגמה:

רוצים לבדוק האם מוטיבציה של מתגייסים לרבי ירדה. מנתוני העבר ידוע כי 40% משרתים בקרבי. נלקח מדגם של 100 מתגייסים חדשים ונמצא כי 30 מתוכם הלכו לקרבי.

א) נסחו השערות ומצאו את רמת המובהקות המינימאלית לדחיית H_0.

ב) בנו רו"ס לפרופורציית המתגייסים לקרבי ברמת בטחון של 95%.

פתרון:

א) *השערות*: $H_0: P = 0.4$, $H_1: P < 0.4$. ורמת מובהקות מינימלית:

$$p.v. = p\left(Z < \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right) = p\left(Z < \frac{0.3-0.4}{\sqrt{\frac{0.4(1-0.4)}{100}}}\right) = \phi(-2.04) = 1 - \phi(2.04) =$$

0.0207. עבור כל α שגדולה מ2.07% דוחים את השערת הבסיס.

ב) $1 - \alpha = 0.95$, $1 - \frac{\alpha}{2} = 0.975$, $Z_{1-\frac{\alpha}{2}=0.975} = 1.96$,

$\hat{p} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$,

$0.3 - 1.96 \cdot \sqrt{\frac{0.3 \cdot 0.7}{100}} \leq p \leq 0.3 + 1.96 \cdot \sqrt{\frac{0.3 \cdot 0.7}{100}}$, $0.21 \leq p \leq 0.39$. ברמת בטחון של 95%