

פתרון תרגיל בית 11 – מבוא להסתברות וסטטיסטיקה 88-165 – קיץ 2013

1. הנתונים: $\sigma = 3, n = 35, \alpha = 95\%$.

$$P(\bar{x} > k) = 0.95$$

$$P\left(z > \frac{k-0}{3/\sqrt{35}}\right) = 0.05$$

$$1 - \Phi\left(\frac{k}{3/\sqrt{35}}\right) = 0.05$$

$$\frac{k}{3/\sqrt{35}} = Z_{0.95}$$

$$k = 1.645 \cdot \frac{3}{\sqrt{35}} = \boxed{0.8342}$$

2. הנתונים: $n = 100; \bar{x} = 98,400; s = 32,200$.

הכנסה מעל 90,000:

$$\begin{aligned} P(k > 90,000) &= P\left(t > \frac{90,000 - 98,400}{32,200/\sqrt{100}}\right) \\ &= P\left(t > \frac{90,000 - 98,400}{32,200/\sqrt{100}}\right) \\ &= P(t > -2.6087) = \boxed{0.9948} \end{aligned}$$

אכן ניתן לקבוע בוודאות מעל 99% שההכנסה הממוצעת גדולה מ-90,000.

רווח סמך להכנסה הממוצעת:

$$t_{99;0.95} = 1.66$$

הערה: בגלל שדרגות החופש 99, אפשר לקחת בקירוב טוב $Z_{0.95} = 1.645$.

$$\mu \in \bar{x} \pm t_{n-1;1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \Rightarrow \mu \in 98,400 \pm 1.66 \cdot \frac{32,200}{\sqrt{100}} \Rightarrow \mu \in \boxed{[93,054.8, 103,745]}$$

3. הנתונים: $n = 26; \bar{x} = 106; \sigma = 14$.

הנוסחה לרווח סמך לתוחלת כאשר השונות ידועה היא: $\mu \in \bar{x} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

מהטבלה הנורמלית הסטנדרטית: $Z_{0.95} = 1.645, Z_{0.975} = 1.96, Z_{0.995} = 2.576$

$$\mu \in 106 \pm 1.645 \cdot \frac{14}{\sqrt{26}} \Rightarrow \mu \in \boxed{[101.483, 110.517]} : \text{רווח סמך } 90\%$$

$$\mu \in 106 \pm 1.96 \cdot \frac{14}{\sqrt{26}} \Rightarrow \mu \in [100.619, 111.381] : 95\% \text{ סמך רווח}$$

$$\mu \in 106 \pm 2.576 \cdot \frac{14}{\sqrt{26}} \Rightarrow \mu \in [98.927, 113.073] : 99\% \text{ סמך רווח}$$

4. הנתונים: $n = 41$; $s^2 = 0.42$; $1 - \alpha = 0.95$.

(א). השערות המבחן:

$$\begin{cases} H_0 : \sigma^2 \leq 25 \\ H_1 : \sigma^2 > 25 \end{cases}$$

חישוב סטטיסטי המבחן:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(41-1) \cdot 0.42}{0.25} = 67.2$$

מטבלת חי-בריבוע:

$$\chi_{(41-1);0.95}^2 = 55.76$$

$$\chi_{(41-1);0.95}^2 = 55.76 < \chi^2 = 67.2 \text{ מכיוון שמתקיים:}$$

אזי נדחה את H_0 ברמת בטחון 95%. כלומר: טענת החברה איננה נכונה, ברמת בטחון 95%.

$$(ב). \text{ מטבלת חי-בריבוע: } \chi_{(41-1);0.95}^2 = 55.76$$

החסם העליון לשונות אחוז השומנים הוא:

$$P\left(\chi^2 \leq \frac{(n-1)s^2}{\sigma_0^2}\right) = 0.95 \Rightarrow \frac{40 \cdot 0.42}{\sigma_0^2} = 55.76 \Rightarrow \sigma_0^2 = 0.30 = \boxed{30\%}$$

5. הנתונים: $n_1 = 150$; $\bar{x}_1 = 6,500$; $\sigma_1 = 234$

$n_2 = 200$; $\bar{x}_2 = 6,800$; $\sigma_2 = 216$

צריך 90% וודאות, מכאן $\alpha = 0.1$. מהטבלה הנורמלית $Z_{0.95} = 1.645$.

נוסחת רווח סמך להפרש התוחלות בהנחה שהשונות ידועות:

$$(\mu_1 - \mu_2) \in (\bar{X}_1 - \bar{X}_2) \pm Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

הצבה בנוסחה נותנת:

$$(\mu_1 - \mu_2) \in 300 \pm 1.645 \cdot \sqrt{\frac{234^2}{150} + \frac{216^2}{200}} = 300 \pm 40.2377$$

שימוש ברווח הסמך שקול למבחן דו-צדדי. מכיוון ש-0 לא נכלל ברווח הסמך, ניתן לומר שיש הבדל מובהק בין ממוצעי ההכנסות בשתי הערים, ברמת המובהקות הנתונה.

6. הנתונים: $n_1 = 14$; $\bar{x}_1 = 5.9$; $\sigma_1 = 4.1$

$n_2 = 18$; $\bar{x}_2 = 4.1$; $\sigma_2 = 3.7$

רמת הבטחון 95%, כלומר $\alpha = 0.05$. מהטבלה הנורמלית $Z_{0.95} = 1.645$.

השערות המבחן:

$$\begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 > 0 \end{cases}$$

הערך הקריטי של המבחן:

$$K = 0 + 1.645 \cdot \sqrt{\frac{4.1^2}{14} + \frac{3.7^2}{18}} = 2.084$$

כלל ההחלטה: דחה H_0 אם $\bar{x}_1 - \bar{x}_2 > K$

$$\bar{x}_1 - \bar{x}_2 = 5.9 - 4.1 = 1.8 < 2.084$$

ההשוואה: לא נדחה H_0 ברמת בטחון 95%. לא ניתן להסיק שדיאטת הפחתת אכילה יותר אפקטיבית.

7. ישנן 800 הטלות של קובייה סה"כ, מתוכן מס' התוצאות הזוגיות:

$$0 \cdot 10 + 1 \cdot 41 + 2 \cdot 70 + 3 \cdot 57 + 4 \cdot 22 = 440$$

מכאן שהאומדן ל- p מתוצאות המדגם:

$$\hat{p} = \frac{440}{800} = 0.55$$

נבדוק את התאמת המדגם להתפלגות $Bin(4, p = 0.55)$:

0	1	2	3	4	מס' קוביות זוגיות
10	41	70	57	22	שכיחות (O_i)
8.2	40.1	73.5	59.9	18.3	ערך צפוי (E_i)

ערכי E_i שבטבלה חושבו ע"י הנוסחה: $E_i = n \cdot P_i$.

$$E_2 = 200 \cdot \binom{4}{2} 0.55^2 \cdot 0.45^2 = 73.5$$

כך לדוגמה עבור 2 קוביות זוגיות:

השערות המבחן:

H_0 : התצפיות מתאימות להתפלגות הנבדקת:

H_1 : התצפיות לא מתאימות להתפלגות הנבדקת:

סטטיסטי המבחן:

$$\chi^2 = \sum_{i=1}^5 \frac{(O_i - E_i)^2}{E_i} = \frac{(10 - 8.2)^2}{8.2} + \dots + \frac{(22 - 18.3)^2}{18.3} = 1.47$$

הערך המתאים מטבלת חי-בריבוע, ניקח רמת בטחון של 95%: $\chi^2_{(5-1);0.95} = 9.488$

מכיוון ש: $\chi^2 = 1.47 < \chi^2_{4;0.95} = 9.488$ לא דוחים את H_0 .

המסקנה: התצפיות מתאימות להתפלגות $Bin(4, p = 0.55)$, ברמת המובהקות הנתונה.

8. בעזרת הנתונים נעריך את הפרמטר $\lambda = \bar{X}$:

$$\bar{X} = \frac{0 \cdot 77 + 1 \cdot 90 + 2 \cdot 55 + 3 \cdot 30 + 4 \cdot 5 + 5 \cdot 3}{260} = 1.25$$

נבנה טבלה של הערכים הצפויים (E_i):

x	P(X = x)	Ei
0	0.2865	260 x 0.2865 = 74.50
1	0.3581	260 x 0.3581 = 93.11
2	0.2238	260 x 0.2238 = 58.19
3	0.0933	260 x 0.0933 = 24.26
4	0.0291	260 x 0.0291 = 7.57
5	0.0073	260 x 0.0073 = 1.90

השערות המבחן:

H_0 : התצפיות מתאימות להתפלגות הנבדקת

H_1 : התצפיות לא מתאימות להתפלגות הנבדקת

נחשב את סטטיסטי המבחן:

$$\begin{aligned} \chi^2 &= \sum_{i=1}^6 \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(77 - 74.50)^2}{74.50} + \frac{(90 - 93.11)^2}{93.11} + \frac{(55 - 58.19)^2}{58.19} + \frac{(30 - 24.26)^2}{24.26} \\ &\quad + \frac{(5 - 7.57)^2}{7.57} + \frac{(3 - 1.90)^2}{1.90} = 2.62 \end{aligned}$$

נבדוק ברמת מובהקות 5%. מטבלת חי-בריבוע:

$$\chi^2_{(6-1-1);0.95} = \chi^2_{4;0.95} = 9.488$$

מכיוון ש: $\chi^2 = 2.62 < \chi^2_{4;0.95} = 9.488$, לא דוחים את H_0 .

המסקנה: התצפיות מתאימות להתפלגות פואסון עם פרמטר $\lambda = 1.25$, ברמת המובהקות הנתונה.

9. מתוצאות המדגם נחשב אומדנים עבור התוחלת והשונות של ההתפלגות הנורמלית.

$$\hat{\mu} = \bar{x} = \frac{1}{175} (7 \cdot 27 + 22 \cdot 29 + 36 \cdot 31 + 45 \cdot 33 + 33 \cdot 35 + 28 \cdot 37 + 4 \cdot 39) = \boxed{33}$$

$$\begin{aligned} \hat{\sigma}^2 = s^2 &= \frac{1}{174} [7 \cdot (27 - 33)^2 + 22 \cdot (29 - 33)^2 + 36 \cdot (31 - 33)^2 + 45 \cdot (33 - 33)^2 \\ &\quad + 33 \cdot (35 - 33)^2 + 28 \cdot (37 - 33)^2 + 4 \cdot (39 - 33)^2] = \boxed{8.46} \end{aligned}$$

נבנה את טבלת הערכים הצפויים (E_i):

$x_a - x_b$	$P(x_a < X < x_b)$	E_i
26-28	0.034804	$175 \times 0.034804 = 6.09$
28-30	0.108408	18.97
30-32	0.21427	37.50
32-34	0.268885	47.05
34-36	0.21427	37.50
36-38	0.108408	18.97
38-40	0.034804	6.09

נחשב את סטטיסטי המבחן:

$$\chi^2 = \sum_{i=1}^7 \frac{(O_i - E_i)^2}{E_i} = \frac{(7 - 6.09)^2}{6.09} + \frac{(22 - 18.97)^2}{18.97} + \dots + \frac{(4 - 6.09)^2}{6.09} = 6.322$$

מטבלת חי-בריבוע, ברמת מובהקות 5%:

$$\chi^2_{(7-2-1);0.95} = \chi^2_{4;0.95} = 9.488$$

מכיוון ש: $\chi^2 = 6.322 < \chi^2_{4;0.95} = 9.488$ לא דוחים את H_0 .

המסקנה: התצפיות מתאימות להתפלגות נורמלית עם הפרמטרים $N(\mu = 33, \sigma^2 = 8.46)$.

10. נארגן את התצפיות בטבלה ונוסיף עמודת סיכום לכל שורה ולכל עמודה:

דעה	יחידה 1	יחידה 2	יחידה 3	סה"כ
טוב מאוד	40	60	55	155
טוב	50	35	40	125
בסדר	30	30	20	80
רע	15	10	10	35
רע מאוד	10	10	10	30
סה"כ	145	145	135	425

נבצע בדיקת טיב-התאמה לאי-תלות בין המשתנים "יחידה" לבין "דעה".

ההשערות המבחן:

H_0 : אין תלות בין שביעות הרצון ליחידה בה מאושפזים:

H_1 : יש תלות בין שביעות הרצון ליחידה בה מאושפזים:

טבלת הערכים הצפויים (E_i):

דעה	יחידה 1	יחידה 2	יחידה 3
טוב מאוד	52.88	52.88	49.24
טוב	42.65	42.65	39.71
בסדר	27.29	27.29	25.41
רע	11.94	11.94	11.12
רע מאוד	10.24	10.24	9.53

הערה: הטבלה לעיל חושבה ע"י הנוסחה-

$$E_{i,j} = \frac{A \cdot B}{n}, \text{ כאשר } A \text{ (סה"כ) השכיחות של קטגוריה } i, \text{ ו- } B \text{ (סה"כ) השכיחות של קטגוריה } j.$$

סטטיסטי המבחן:

$$\chi^2 = \sum_{i=1}^3 \sum_{j=1}^5 \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} = \frac{(40 - 52.88)^2}{52.88} + \dots + \frac{(10 - 9.53)^2}{9.53} = 10.347$$

$$\chi^2_{(3-1)(5-1);0.95} = \chi^2_{8;0.95} = 15.51: 95\% \text{ עבור רמת בטחון}$$

$$\chi^2 > \chi^2_{(3-1)(5-1); 1-\alpha} \text{ הוא: } H_0 \text{ איזור הדחייה של}$$

$$\chi^2 = 10.347 < \chi^2_{8;0.95} = 15.51 \text{ מכיוון שמתקיים:}$$

אזי המסקנה: לא נדחה את H_0 ברמת בטחון 95%.
כלומר: לא ניתן להצביע על תלות מובהקת בין שביעות הרצון לבין יחידת האישפוז, ברמת בטחון 95%.

11. הנתונים:

שנה	2000	2001	2002	2003	2004
מספר בוגרים	88	92	96	95	114

בשאלה שלנו המשתנה הבלתי תלוי (X) הוא השנה, והמשתנה התלוי (Y) הוא מספר הבוגרים.

נוסחת קו הרגרסיה של Y כתלות ב- X : $\hat{y} = bx + a$

$$b = \frac{\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2} \quad \text{כאשר: } a = \bar{x} - b\bar{y}$$

המקדמים המתקבלים מהחישובים: $b = 5.5$, $a = -10914$.

לכן משוואת קו הרגרסיה היא: $\hat{y} = 5.5x - 10914$

$$5.5 \cdot 2008 - 10914 = \boxed{130} \text{ אומדן למספר הבוגרים ב-2008:}$$

12. הנתונים:

שנה	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
סכום מיכרות	132.5	142.6	143.8	171.0	203.8	203.8	228.5	217.8	294.2	303.5

בשאלה שלנו המשתנה הבלתי תלוי (X) הוא השנה, והמשתנה התלוי (Y) הוא סכום המכירות.

נשתמש בנוסחת הרגרסיה, כאשר לפני כן ביצענו את טרנספורמציה לוגריתמית:

$$y = cd^x \Rightarrow \log y = \log c + x \log d$$

הנתונים לאחר הטרנספורמציה:

x	y	ln y
0	132.5	4.8866
1	142.6	4.9600
2	143.8	4.9684
3	171.0	5.1417
4	203.8	5.3171
5	203.8	5.3171
6	228.5	5.4315
7	217.8	5.3836
8	294.2	5.6843
9	303.5	5.7154

- הערה 1: השתמשנו בלוגריתם הטבעי מטעמי נוחות. ניתן להשתמש בבסיס אחר.
הערה 2: העתקנו את המשתנה הבלתי תלוי (X) מהתחום 2000-2009 לתחום 0-9. (שוב מטעמי נוחות).
לצרכי ניבוי, יש לחסר 2000 מערך המשתנה הבלתי תלוי (X).

כעת נציב במשוואת קו הרגרסיה ונקבל: $\hat{y} = 0.0938x + 4.8586$

מכאן, ע"י הטרנספורמציה ההפוכה נקבל:

$$c = e^{4.8586} \approx 128.84; \quad d = e^{0.0938} \approx 1.10$$