

תרגול כיתה 2 – מבוא להסתברות וסטטיסטיקה

מדדים, התפלגות נורמלית, רווח סמך

טרנספורמציה לינארית של מדדים

השפעה כללית טרנספורמציה לינארית $ax + b$ על מדדי המרכז ומדדי הפיזור -

$$\tilde{M} = aM + b \quad : M \text{ מרכז}$$

$$\tilde{D} = g(a)D \quad : D \text{ פיזור}$$

באופן מפורט:

הפעלת טרנספורמציה לינארית על כל המדגם כך שכל תצפית x במדגם הופכת ל: $y = ax + b$

משפיעה על המדדים באופן הבא:

$$Me_y = a \cdot Me_x + b \quad \text{חציון:}$$

$$\bar{y} = a \cdot \bar{x} + b \quad \text{ממוצע:}$$

$$Mo_y = a \cdot Mo_x + b \quad \text{שכיח:}$$

$$MR_y = a \cdot MR_x + b \quad \text{אמצע טווח:}$$

$$Range_y = a \cdot Range_x \quad \text{טווח:}$$

$$S_y^2 = a^2 \cdot S_x^2 \quad \text{שונות:}$$

$$S_y = |a| \cdot S_x \quad \text{סטיית תקן:}$$

שאלה 1

במבחן סוף סמסטר התקבלו המדדים הבאים עבור הציונים:

ממוצע = 61, חציון = 65, שונות = 25, טווח = 80.

שני הבודקים הציעו לשנות את הציונים בדרכים הבאות:

- בודק א' הציע לתת תוספת של 12 נק' לכל סטודנט.

- בודק ב' הציע לחלק כל ציון ב-2 והוספת 50 נק' לתוצאה.

כיצד ישפיעו הצעות הבודקים על כ"א מהמדדים הנ"ל?

פתרון:

(א). השינוי לפי הצעת בודק א' $\tilde{x} = x + 12$. ההשפעה על המדדים -

הממוצע:

$$\tilde{\bar{x}} = \frac{\sum_{i=1}^n (x_i + 12)}{n} = \frac{\sum_{i=1}^n x_i + 12n}{n} = \bar{x} + 12 = 61 + 12 = 73$$

החציון:

אין שינוי בסדר הנתונים תחת הטרנספורמציה לכן אותה תצפית נשארת החציון.

$$\tilde{x}_{0.50} = x_{0.50} + 12 = 65 + 12 = 77$$

השונות:

כאמור הממוצע השתנה $\bar{\tilde{x}} = \bar{x} + 12$ נציב בנוסחת השונות

$$\tilde{s}^2 = \frac{\sum_{i=1}^n (\tilde{x}_i - \bar{\tilde{x}})^2}{n-1} = \frac{\sum_{i=1}^n [(x_i + 12) - (\bar{x}_i + 12)]^2}{n-1} = s^2 = 25$$

הטווח:

$$R(\tilde{x}) = \tilde{x}_{\max} - \tilde{x}_{\min} = (x_{\max} + 12) - (x_{\min} + 12) = x_{\max} - x_{\min} = R(x) = 80$$

(ב). השינוי לפי הצעת בודק ב' $\tilde{x} = x/2 + 50$. ההשפעה על המדדים-

הממוצע:

$$\bar{\tilde{x}} = \frac{\sum_{i=1}^n (\frac{1}{2}x_i + 50)}{n} = \frac{\frac{1}{2}\sum_{i=1}^n x_i + 50n}{n} = \frac{1}{2}\bar{x} + 50 = \frac{1}{2} \cdot 61 + 50 = 80.5$$

החציון:

הטרנספורמציה לא משנה את הסדר, איבר החציון הוא אותו איבר שעבר את הטרנספורמציה:

$$\tilde{x}_{0.50} = \frac{1}{2}x_{0.50} + 50 = \frac{1}{2} \cdot 65 + 50 = 82.5$$

השונות:

כאמור הממוצע $\bar{\tilde{x}} = \frac{1}{2}\bar{x} + 50$ נציב בנוסחת השונות

$$\tilde{s}^2 = \frac{\sum_{i=1}^n (\tilde{x}_i - \bar{\tilde{x}})^2}{n-1} = \frac{\sum_{i=1}^n [(\frac{1}{2}x_i + 50) - (\frac{1}{2}\bar{x}_i + 50)]^2}{n-1} = \frac{1}{4}s^2 = \frac{1}{4} \cdot 25 = 6.25$$

$$R(\tilde{x}) = \tilde{x}_{\max} - \tilde{x}_{\min} = (\frac{1}{2}x_{\max} + 50) - (\frac{1}{2}x_{\min} + 50) = \frac{1}{2}(x_{\max} - x_{\min}) = \frac{1}{2}R(x) = \frac{1}{2} \cdot 80 = 40$$

התפלגות נורמלית

המעבר מהתפלגות נורמלית כלשהי להתפלגות נורמלית סטנדרטית (תקנון)

$$X \sim N(\mu, \sigma^2) \Rightarrow \frac{X - \mu}{\sigma} \sim N(0,1)$$

כללים לשימוש בטבלת ההתפלגות הנורמלית סטנדרטית z

$$(a \geq 0)$$

$$P(Z \leq a) \equiv \phi(a); \quad P(Z \geq a) = 1 - \phi(a); \quad P(Z \leq -a) = 1 - \phi(a)$$

$$P(|Z| \leq a) = 2\phi(a) - 1; \quad P(a \leq Z \leq b) = \phi(b) - \phi(a)$$

תרגול דוגמאות לשימוש בטבלה הנורמלית

- $\Phi(2.42) = 0.9922$
- $\Phi(-1.5) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668$
- $\Phi(0) = 0.5$
- $\Phi(-2.22 < Z \leq 1.55) = \Phi(1.55) - \Phi(-2.22) = \Phi(1.55) - [1 - \Phi(2.22)]$
 $= 0.9394 - 1 + 0.9868 = 0.9262$
- $\Phi(Z \geq -2.33) = \Phi(Z \leq 2.33) = \Phi(2.33) = 0.9901$

סימון מקובל: $\Phi^{-1}(p) = Z_p$. כאשר p – האחוזון ($0 \leq p \leq 1$).

אחוזונים שימושיים:

$$Z_{0.75} = 0.6745$$

$$Z_{0.25} = -Z_{1-0.25} = -Z_{0.75} = -0.6745$$

$$Z_{0.90} = 1.282$$

$$Z_{0.95} = 1.645$$

$$Z_{0.975} = 1.960$$

$$Z_{0.025} = -Z_{1-0.025} = -Z_{0.975} = -1.960$$

$$Z_{0.99} = 2.33$$

$$Z_{0.995} = 2.576$$

$$Z_{0.999} = 3.09$$

שאלה 2

נתון ש- X מתפלג נורמלית עם תוחלת 10 ושונות 100. $X \sim N(\mu = 10, \sigma^2 = 100)$. חשב:

א. $P(X \leq 0)$

ב. מהו האחוזון ה-15.87 של ההתפלגות? כלומר מהו x כך שיתקיים $P(X \leq x) = 0.1587$.

ג. מהו האחוזון ה-10 של ההתפלגות?

פתרון:

א. $P(X \leq 0) = P\left(\frac{X-10}{10} \leq \frac{0-10}{10}\right) = P(Z \leq -1) = \Phi(-1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$

ב. $P(X \leq x) = 0.1587 = P\left(\frac{X-10}{10} \leq \frac{x-10}{10}\right) \Rightarrow \frac{x-10}{10} = Z_{0.1587}$
 $\Rightarrow x = 10Z_{0.1587} + 10 \Rightarrow x = -10Z_{1-0.1587} + 10 = -10 \cdot 1 + 10 = 0$

ג. $P\left(\frac{X-10}{10} \leq \frac{x-10}{10}\right) = 0.1$
 $\Rightarrow \frac{x-10}{10} = Z_{0.10} \Rightarrow \frac{x-10}{10} = -Z_{0.90} \Rightarrow \frac{x-10}{10} = -1.282 \Rightarrow x = -2.82$

שאלה 3

המכירות של שופרות משחת שיניים מסוימת מתפלג נורמלית עם תוחלת של 10,000 וסטיית תקן של 1,500 יחידות בשבוע.

א. מהי ההסתברות שיותר מ-12,000 שופרות תמכרנה בשבוע?

ב. כמה שופרות לפחות צריך לייצר כדי להיות בטוחים ב-95% שלא ייגמר המלאי?

פתרון:

א. נסמן ב- X את מספר השופרות המיוצר בשבוע, אזי $X \sim N(10,000; 1,500^2)$.

אנו נדרשים לחשב את ההסתברות שתמכרנה יותר מ-12,000 שופרות, דהיינו $P(X > 12,000)$.

כדי להשתמש בטבלה נתקן לנורמלית סטנדרטית $Z \sim N(0,1)$, מקבלים

$$P(X > 12,000) = P\left(Z > \frac{12,000 - 10,000}{1,500}\right) = 1 - P\left(Z \leq \frac{4}{3}\right) = 1 - \Phi\left(\frac{4}{3}\right) = 0.091$$

(ב). נדרשים למצוא את הקבוע c כך שמתקיים $P(X \leq c) = 0.95$.

בעזרת הטבלה הנורמלית (Z), נמצא את האחוזון ה-95% של ההתפלגות: $Z_{0.95} = 1.645$. לפיכך, מספר השפופרות שצריך לייצר כדי להבטיח שהמלאי יתאים לביקוש ב-95% מהפעמים הוא-

$$\Phi\left(\frac{c - 10,000}{1,500}\right) = 0.95$$

$$c = Z_{0.95} \cdot 1,500 + 10,000 = 1.645 \cdot 1,500 + 10,000 = 12,467.5$$

דהיינו: 12,468 שפופרות.

שאלה 4

נתון שקבוצת התצפיות מתפלגת נורמלית עם ממוצע m וסטיית תקן s .

א. מהו אחוז התצפיות שמתחת לערך $(m - 0.675s)$?

ב. מצא קבוע a כך שאחוז התצפיות בתוך הטווח $(m \pm as)$ הוא 75% מהקבוצה.

פתרון:

(א). צריך למצוא $P(X \leq m - 0.675s)$

נתקבן:

$$P\left(\frac{x - m}{s} \leq \frac{m - 0.675s - m}{s}\right)$$

$$\Rightarrow P(Z \leq -0.675) = \Phi(-0.675) = 1 - \Phi(0.675) = 1 - 0.75 = 0.25$$

(ב). צריך למצוא a כך שמתקיים $P(m - as \leq X \leq m + as) = 0.75$

נתקבן:

$$\Rightarrow P\left(\frac{m - as - m}{s} \leq \frac{X - m}{s} \leq \frac{m + as - m}{s}\right) = 0.75$$

$$\Rightarrow P(-a \leq Z \leq a) = 0.75$$

$$\Rightarrow \Phi(a) - (1 - \Phi(a)) = 0.75$$

$$\Rightarrow 2\Phi(a) - 1 = 0.75$$

$$\Phi(a) = 0.875$$

$$a = \Phi^{-1}(0.875) = 1.15$$

הסקה סטטיסטית – אמידהמונחים:

- אוכלוסייה – קבוצה/אוסף פריטים שמעוניינים במידע מסוים עליהם.
מדגם – קבוצה חלקית של פריטים מהאוכלוסייה.
פרמטר – מספר קבוע המאפיין את האוכלוסייה (יסומן ע"י: θ).
אומדן/אמד – סטטיסטי המשמש לאמידת פרמטר באוכלוסייה (יסומן ע"י: $\hat{\theta}$).
אומדן – הערך המספרי של האמד.

רווח סמך

זוהי אמידה מרווחית (באינטרוול) של פרמטר באוכלוסייה, ע"י שימוש בתוצאות מדגם. רווח הסמך נבנה כך שההסתברות שהפרמטר יהיה בתוך הקטע נקבעת מראש ושווה $1 - \alpha$ ("רמת סמך").
 $0 \leq \alpha \leq 1$ – רמת מובהקות.

רווח סמך עבור תוחלת האוכלוסייה, כאשר שונותה ידועה

$$\bar{X} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} : \mu \text{ עבור } 1 - \alpha \text{ ברמת סמך}$$

רווח סמך עבור תוחלת האוכלוסייה, כאשר שונותה איננה ידועה

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x_i^2 - n\bar{x}^2}{n-1}} : \text{נסמן את סטיית התקן המדגמית}$$

$$\bar{X} \pm t_{\left(n-1, 1-\frac{\alpha}{2}\right)} \frac{S}{\sqrt{n}} : \mu \text{ עבור } 1 - \alpha \text{ ברמת סמך}$$

רווח סמך לפרופורציה באוכלוסייה

p – פרופורציית התכונה באוכלוסייה. $\hat{p} = x/n$ – פרופורציה מדגמית.

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} : p \text{ עבור } 1 - \alpha \text{ ברמת סמך}$$

שאלה 5

- מכונה מייצרת ברגים שאורכם מתפלג נורמלית עם תוחלת 4 ס"מ וסטיית תקן 0.2 ס"מ.
 עקב תקלה במכונה הועלה חשד כי המכונה אינה מייצרת ברגים באורך הנדרש.
 לשם בדיקה נלקח מדגם מקרי של 25 ברגים אשר יוצרו במכונה לאחר גילוי התקלה ונמצא כי האורך הממוצע שלהם הוא 3.9 ס"מ. נניח כי השונות נותרה ללא שינוי.
- מהו רווח סמך (רו"ס) לתוחלת אורך הברגים לאחר התקלה ברמת מובהקות (ר"מ) 2%?
 - בונים רווח סמך לתוחלת ברמת בטחון של 95%, ע"י לקיחת מדגם של מספר מסויים של ברגים. כמה ברגים לפחות צריך לקחת על מנת שאורכו של רווח הסמך יהיה לכל היותר 0.1 ס"מ?
 - בנו רו"ס בגודל 0.118 ס"מ בעזרת מדגם בגודל 49 ברגים. מה רמת המובהקות α ?

פתרון:

(א) נסמן ב- X את אורך הברגים לאחר התקלה, $X \sim N(\mu, 0.2^2)$. כעת לא ידוע.כמו כן נתון: $n = 25$ ו- $\bar{X} = 3.9$ רמת מובהקות של $2\% \Leftrightarrow \alpha = 0.02 \Leftrightarrow Z_{1-\frac{\alpha}{2}} = 2.326$ רו"ס ברמת מובהקות של α לתוחלת כאשר השונות ידועה:

$$\bar{X} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$3.81 \leq \mu \leq 3.99 \Leftrightarrow 3.9 - 2.236 \cdot \frac{0.2}{5} \leq \mu \leq 3.9 + 2.236 \cdot \frac{0.2}{5}$$

(ב) נשתמש בנוסחה לרו"ס ברמת מובהקות של α לתוחלת כאשר השונות ידועה:

$$\bar{X} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$L = 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

במקרה שלנו, רמת ביטחון של $95\% \Leftrightarrow \alpha = 0.05 \Leftrightarrow Z_{1-\frac{\alpha}{2}} = 1.96$ נרצה שאורך רווח הסמך יהיה לכל היותר 0.1 לכן:

$$2 \cdot 1.96 \cdot \frac{0.2}{\sqrt{n}} \leq 0.1 \Rightarrow \sqrt{n} \geq 7.84 \Rightarrow n \approx 61.47$$

ז"א שיש צורך במדגם בגודל 62 ברגים לפחות.

$$L = 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

במקרה שלנו $\alpha = ?$

$$2Z_{1-\frac{\alpha}{2}} \cdot \frac{0.2}{\sqrt{49}} = 0.118 \Rightarrow Z_{1-\frac{\alpha}{2}} = 2.06 \Rightarrow 1 - \frac{\alpha}{2} = 0.98 \Rightarrow \alpha = 0.04$$

שאלה 6

חוקר מציע סוג של דשן שעשוי להגדיל את משקלם של אפרסקים. נלקח מדגם מקרי של 9 אפרסקים והתקבלו

$$\sum x = 500, \sum x^2 = 28166 \text{ (בגרמים)}$$

א. אמוד ברו"ס את משקל האפרסקים ברמת סמך של 90% על סמך תוצאות המדגם הנ"ל.ב. לאחר האמידה התברר שחלה תקלה שיטתית במדידות ויש להוסיף 8 גרם לכל ערך שנמדד.

הסבר בקצרה, האם וכיצד ישפיע תיקון הטעות על גודל ומיקום רווח הסמך.

פתרון:

(א)

נחשב את ממוצע המדגם וסטיית התקן המדגמית:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{500}{9} = 55.556, \quad S = \sqrt{\frac{\sum x_i^2 - n\bar{x}^2}{n-1}} = \sqrt{\frac{28166 - 9 \cdot 55.556^2}{9-1}} = 6.966$$

הערך המתאים לנתוני השאלה – מטבלת t:

$$t_{8,0.95} = 1.86 \leq 1 - \alpha = 0.9, \quad n = 9$$

נציב בנוסחת רו"ס:

$$55.55 - 1.86 \cdot \frac{6.966}{3} \leq \mu \leq 55.55 + 1.86 \cdot \frac{6.966}{3}$$

$$51.231 \leq \mu \leq 59.868$$

(ב) השפעת תיקון הטעות:

הוספה של קבוע לא משפיעה על השונות (וסטיית התקן), לכן גודל רווח הסמך לא ישתנה. הוספה של קבוע מעלה את הממוצע בקבוע, לכן הממוצע יגדל ב-8. מיקום רו"ס יוזז ב-8 ימינה.

שאלה 7

חוקר מעוניין לאמוד את אחוז האלרגיים באוכלוסייה לפריחת עצי אורן. נלקח מדגם מקרי של 80 אנשים ונמצא כי 20 מהם סובלים מאלרגיה.

- מהו רווח הסמך לאחוז האלרגיים באוכלוסייה כולה ברמת בטחון של 99%?
- מהו גודל המדגם שעל החוקר לדגום ע"מ להבטיח ברמת ביטחון של 90% לפחות שתוצאת המדגם לא תסטה מהתוצאה האמתית באוכלוסייה ביותר מ-5%?
- אם רוצים לשמור על גודל רווח הסמך, כיצד ישתנה גודל המדגם אם נגדיל את רמת הביטחון ל-99.9% (הסבירו ללא חישוב גודל המדגם מחדש).

פתרון:

$$X \sim (80, p) - \text{מספר האלרגיים במדגם.}$$

$$n = 80, \hat{p} = 20/80 = 0.25 \text{ הפרופורציה של האלרגיים במדגם:}$$

א.

$$Z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2.57 \cdot \sqrt{\frac{0.25 \cdot 0.75}{80}} = 0.124$$

$$0.25 - 0.124 \leq p \leq 0.25 + 0.124 \Rightarrow 0.126 \leq p \leq 0.374$$

אחוז האלרגיים באוכלוסייה כולה נע בין 12.6% לבין 37.4% ברמת בטחון של 99%.

(ב) נסמן L – גודל רווח הסמך. מחפשים את הסטייה - $L/2 \leq 0.05$. כלומר -

$$Z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0.05 \Rightarrow (Z_{0.95} = 1.645)$$

$$\Rightarrow 1.645 \cdot \sqrt{\frac{0.25 \cdot 0.75}{n}} \leq 0.05 \Rightarrow \boxed{n \geq 202.952}$$

מכאן שנדרש מדגם של לפחות 203 אנשים.

(ג) באופן כללי: הגדלת רמת הבטחון מגדילה את רווח הסמך. כמו כן, הגדלת המדגם מקטינה את רווח הסמך. לכן נצטרך להגדיל את המדגם כדי לשמור על רווח הסמך. מבחינה מספרית, נרצה לשמור על

$$\text{היחס: } K = \frac{Z_{(1-\alpha/2)}}{\sqrt{n}} - \text{קבוע.}$$