פתרון – תרגיל בית 1 – מבוא להסתברות וסטטיסטיקה

שאלה 1

נמדד מרחק בלימה של משאית על כביש מצופה קרח. נערכו 85 נסיונות בלימה והתוצאות סוכמו בטבלה הבאה:

מרחק בלימה (מטרים)	f (שכיחות)
210 - 220	1
220 - 230	3
230 – 240	12
240 – 250	15
250 – 260	22
260 - 270	7
270 - 280	7
280 - 290	5
290 - 300	2
300 – 310	10
310 - 320	1

א. סטיית התקן המדגמית לנתונים מקובצים נתונה ע"י הנוסחה: $s^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2$

הראה שנוסחה זו שקולה לנוסחה: $s^2=\frac{\sum_{i=1}^{n}x_i^2-n\overline{x}^2}{n-1}$.

ב. הוסף עמודת שכיחות מצטברת F(x) לטבלה.

ג. חשב: (1). ממוצע; (2). חציון; (3). שכיח.

ד. חשב: (1) סטיית התקן המדגמית; (3). תחום בין רבעוני (IQR).

פתרון:

א. נתון: $s^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2$ נפתח את הביטוי-

$$\begin{aligned} s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i^2-2x_i\overline{x}+\overline{x}^2\right) \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2-2\overline{x}\sum_{i=1}^{n}x_i+n\overline{x}^2\right] \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2-2n\overline{x}^2+n\overline{x}^2\right] \\ & \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2-n\overline{x}^2\right] \quad \square \end{aligned}$$

כאשר השתמשנו בזהות- $\overline{x}=\frac{\sum_{i=1}^{n}x_i}{n}\Rightarrow\sum_{i=1}^{n}x_i=n\overline{x}$.

ב. הערה: אין צורך העמודה השמאלית ביותר (#). היא נוספה מטעמי נוחות בלבד.

#	מרחק בלימה	f(x)	F(x)
1	210-220	1	1
2	220-230	3	4
3	230-240	12	16
4	240-250	15	31
5	250-260	22	53
6	260-270	7	60
7	270-280	7	67
8	280-290	5	72
9	290-300	2	74
10	300-310	10	84
11	310-320	1	85

ב.

(1) הממוצע: $\overline{x}=\frac{\sum_i (f_i \cdot x_i)}{n}=\frac{22155}{85}=260.647$

(2) החציון:

נציב בנוסחת החציון: $Q_2=x_{0.5}=250+\left[\frac{0.5\cdot 85-31}{22}\right](260-250)=250+5.6818=255.682$

(3) השכיח: הקבוצה בעלת השכיחות הגדולה ביותר היא (250-260). המחלקות שוות ברוחבן, לכן השכיח הוא מרכז המחלקה: $Mo=255$.

ג.

(1) סטיית התקן המדגמית:

$$s^2=\frac{\sum f\left(x-\overline{x}\right)^2}{n-1}=\frac{47689.4}{84}=567.731 \qquad \Rightarrow s=\sqrt{567.731}=23.8271$$

(2)

הרבעון הראשון Q1: $Q_1=240+\left[\frac{0.25\cdot 85-16}{15}\right](10)=243.5$

הרבעון השלישי Q3: $Q_3=270+\left[\frac{0.75\cdot 85-60}{7}\right](10)=275.357$

התחום הבין רבעוני (IQR): $IQR=Q3-Q1=275.357-243.5=31.857$

<u>שאלה 2</u>

נתונה התפלגות הציונים בקורס בו השתתפו 60 סטודנטים. ידוע שהתפלגות הציונים סימטרית וכן שהשונות (s^2) שווה ל- $116\frac{2}{3}$.

ציון (x)	שכיחות (f)
50 – 60	
60 – 70	
70 – 80	20
80 – 90	
90 – 100	

א. מלא את טבלת השכיחויות הנ"ל. אין לנחש ערכים, יש להצדיק המילוי ע"י חישוב/ים מתאים.
ב. 5 סטודנטים שקיבלו ציון בין 60 – 50 נבחנו במועד ב' ושיפרו ציוניהם. הציונים החדשים הם: 61, 64, 69, 73, 74. כיצד ישתנו המדדים הבאים לאחר השיפור: ממוצע, שכיח, חציון, סטיית תקן.
תן את כיוון השינוי והסבר קצר (לא יותר מ- 3-2 שורות לכל מדד). אין צורך בחישוב מדוייק.
ג. מהו הציון של סטודנט שציון התקן שלו שווה ל- 1.5–.

פתרון:

א. נסמן – השכיחות במחלקה ה- i. $i = 1,2,3,4,5$.

ההתפלגות סימטרית, ולכן: $\bar{x} = \frac{50+100}{2} = 75$ וכמו כן: $f_1 = f_5 = a$ ו־ $f_2 = f_4 = b$

נתון: $S^2 = \frac{\sum (x_i-\bar{x})^2 \cdot f_i}{N} = 116\frac{2}{3}$

$$116\frac{2}{3} = \frac{(55-75)^2\cdot a+(65-75)^2\cdot b+(75-75)^2\cdot 20+(85-75)^2\cdot b+(95-75)^2\cdot a}{60}$$

$$116\frac{2}{3}\cdot 60 = 400a+100b+100b+400a$$

$$4a+b=35$$

סכום השכיחויות: $\sum f_i = 60$

$$f_1+f_2+20+f_4+f_5=60$$

$$a+b+20+b+a=60$$

$$a+b=20$$

$$\begin{array}{r} 4a+b=35 \\ -\quad a+b=20 \\ \hline 3a=15 \end{array} \Rightarrow \boxed{f_1=f_5=5}\ ;\ \boxed{f_2=f_4=15}$$

לפיכך, הטבלה המלאה:

x - ציון	f - שכיחות	x_i - מרכז מחלקה	F - שכיחות מצטברת
50 – 60	$f_1 = 5$	55	5
60 – 70	$f_2 = 15$	65	20
70 – 80	20	75	40
80 – 90	$f_4 = 15$	85	55
90 – 100	$f_5 = 5$	95	60
	60		

ב.

כל 5 התלמידים שקיבלו ציון 50-60 <u>שיפרו</u> את ציוניהם, ולכן:

<u>הממוצע - יגדל</u> - התלמידים בעלי הציונים הנמוכים ביותר שיפרו את ציוניהם. כלומר: קיבלו ציונים גבוהים יותר, ולכן הממוצע גדל.

<u>סטיית התקן - תקטן</u> - לאחר השיפור, הציונים יתקרבו לממוצע. הפיזור יקטן ולכן סטיית התקן תקטן.

<u>השכיח - ישאר ללא שינוי</u>

<u>החציון - יגדל מעט</u> - מחלקת החציון לא תשתנה, אך נוספו שני מקרים בתוך מחלקת החציון, לפני החציון, לפני שינוי הציונים (75), ולכן הוא יגדל מעט.

ג. נתון: $S^2 = 116\frac{2}{3}$ $\Leftarrow$ $S = \sqrt{116\frac{2}{3}} = 10.8$

הממוצע חושב בסעיף א': $\bar{x} = 75$

נתון ציון התקן של הסטודנט: $Z = -1.5$

מכאן- $\boxed{x = 58.8}$ $\Rightarrow$ $-1.5 = \frac{x-75}{10.8}$ $\Rightarrow$ $Z = \frac{x-\bar{x}}{S}$

<u>שאלה 3</u>

בכל אחד מהסעיפים הבאים ענה **נכון/לא נכון** והסבר. <u>תשובה ללא הסבר לא תתקבל</u>.

הנתונים הבאים עבור **סעיפים א' וב'** בלבד:
במפעל 2 דרגות שכר, x_1 נמוכה ו- x_2 גבוהה. מספר הפועלים שמקבלים את הדרגה הנמוכה (x_1), גדול ממספר הפועלים שמקבלים את הדרגה הגבוהה (x_2).

א. מתקיים $\bar{x} < \frac{x_1 + x_2}{2}$ ($\bar{x}$ – השכר הממוצע). הבא הסבר כמותי/פרמטרי לתשובתך.

ב. הממוצע שווה לחציון $(\bar{x} = Me)$. הבא הסבר כמותי/פרמטרי לתשובתך.

ג. בשתי כיתות המונות 50 תלמידים בכל כיתה, ערכו מבחן ידע כללי. הציונים שהתקבלו בשתי הכיתות היו 5, 7, 9 בלבד (סולם הציונים [0,1,2,…,10]). ממוצע הציונים בשתי הכיתות היה 7, אך מספר התלמידים שקיבל 7 בכיתה הראשונה גדול פי 2 ממספר התלמידים שקיבלו 7 בכיתה השניה.
קבע ונמק האם **נכונה/לא נכונה** הטענה הבאה:
"סטיית התקן בכיתה הראשונה שווה לסטיית התקן בכיתה השנייה".
הבא הסבר כמותי/פרמטרי לתשובתך.

<u>פתרון:</u>

א. **נכון**

נתון שמספר הפועלים שמקבלים את הדרגה הנמוכה (x_1), גדול ממספר הפועלים שמקבלים את הדרגה הגבוהה (x_2): $f_1 > f_2$ (f – שכיחות). נתון גם שרמות השכר: $x_1 < x_2$.

כעת, $\bar{x} = \dfrac{x_1 \cdot f_1 + x_2 \cdot f_2}{f_1 + f_2}$. אם $f_1 = f_2$ אזי $\bar{x} = \dfrac{x_1 + x_2}{2}$, אך כאמור $x_1 < x_2$, $f_1 > f_2$ ולכן $\bar{x} < \dfrac{x_1 + x_2}{2}$.

ב. **לא נכון**

מהנתון $f_1 > f_2$, החציון: $Me = x_1$

הממוצע: $\bar{x} = \dfrac{x_1 \cdot f_1 + x_2 \cdot f_2}{f_1 + f_2} \Leftarrow x_1 < \bar{x} < x_2$

ולכן: $Me < \bar{x}$.

ג. **לא נכון**

סטיית התקן ראשונה קטנה יותר.

כיתה 1

ציון x_1	f
5	$\frac{25-b}{2}$
7	$2b$
9	$\frac{25-b}{2}$
	$n_1 = 50$

כיתה 2

ציון x_2	f
5	$\frac{50-b}{2}$
7	b
9	$\frac{50-b}{2}$
	$n_2 = 50$

בכיתה 1 יש יותר ציונים מסביב לממוצע (הממוצע 7 בשתי הכיתות), מאשר בכיתה 2 (2b לעומת b), ולכן בכיתה 1 פיזור הציונים קטן יותר, משמע שסטיית התקן קטנה יותר.

שאלה 4

בקבוצת סטודנטים מסוימת, הציונים בבחינה מתפלגים נורמלית עם ממוצע 60 וסטיית תקן 10.

א. ציון עובר בבחינה הוא 60. נבחר סטודנט באקראי, מה הסיכוי שהוא עבר את הבחינה?

ב. מה הסיכוי לקבל לפחות ציון 80 בבחינה?

ג. מה הסיכוי לקבל לכל היותר ציון 20 בבחינה?

ד. מהו העשירון העליון של הציונים? ומהו העשירון התחתון?

ה. אם נתון כעת שהקבוצה מונה 100 סטודנטים. כמה צפויים לקבל ציון מעל 70?

ו. נניח שנתנו פקטור של תוספת 10% לכל ציון. ענה שוב על סעיף ב', לאחר השינוי. השווה את התוצאה לזו שהתקבלה קודם לכן, האם הסיכוי עלה/ירד/לא השתנה.

פתרון:

נתון: $X \sim N\left(\mu = 60,\ \sigma^2 = 10^2\right)$

א. הסיכוי שסטודנט שנבחר באקראי עבר את הבחינה-

$$P(X \geq 60) = P\left(\frac{X-60}{10} \geq \frac{60-60}{10}\right) = P(Z \geq 0) = P(Z \leq 0) = \Phi(0) = \mathbf{0.5}$$

ב. הסיכוי לקבל לפחות ציון 80-

$$P(X \geq 80) = P\left(\frac{X-60}{10} \geq \frac{80-60}{10}\right) = P(Z \geq 2) = 1 - P(Z \leq 2) = 1 - \Phi(2) = 1 - 0.9772 = \mathbf{0.0228}$$

ג. הסיכוי לקבל לכל היותר ציון 20-

$$P(X \leq 20) = P\left(\frac{x-60}{10} \leq \frac{20-60}{10}\right) = P(Z \leq -4) \cong 0.0000316$$

<u>הערה</u>: אם הטבלה שבידכם נגמרת לפני הערך (4-), אפשר לעגל את התוצאה ל- 0.

ד. העשירון העליון-

$$P(X \leq x) = 0.9 \Leftrightarrow P\left(\frac{X-60}{10} \leq \frac{x-60}{10}\right) = 0.9 \Leftrightarrow \frac{x-60}{10} = Z_{0.9} = 1.282 \Leftrightarrow \boldsymbol{x} = \mathbf{72.82}$$

העשירון התחתון-

$$P(X \leq x) = 0.1 \Leftrightarrow P\left(\frac{X-60}{10} \leq \frac{x-60}{10}\right) = 0.1$$

$$\Leftrightarrow \frac{x-60}{10} = Z_{0.1} = -Z_{0.9} = -1.282 \Leftrightarrow \boldsymbol{x} = \mathbf{47.18}$$

ה. אחוז הסטודנטים שקיבלו ציון מעל 70-

$$P(X \geq 70) = P\left(\frac{X-60}{10} \geq \frac{70-60}{10}\right) = P(Z \geq 1) = 1 - P(Z \leq 1)$$

$$= 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

מכיוון שהקבוצה מונה 100 סטודנטים, אזי מספר הסטודנטים שצפויים לקבל ציון מעל 70-

$$100 \cdot 0.1587 = 15.87 \approx 16$$

ו. לאחר הפקטור, נקבל את השינויים הבאים בהתפלגות האוכלוסייה:

התוחלת וסטיית התקן החדשות- $\sigma_{new} = 1.1 \cdot 10 = 11$ $\quad \mu_{new} = 1.1 \cdot 60 = 66$

כלומר: $X_{new} \sim N\left(\mu = 66,\ \sigma^2 = 11^2\right)$

הסיכוי לקבל לפחות ציון 80 בבחינה הפעם-

$$P(X \geq 80) = P\left(Z \geq \frac{80-66}{11}\right) = P(Z \geq 1.2727) = 1 - P(Z \leq 1.2727)$$
$$= 1 - \Phi(1.2727) = 1 - 0.898 = 0.102$$

התוצאה שהתקבלה גדולה מסעיף ב', כיוון שלאחר הפקטור ברור שהסיכוי גדל לקבל ציון מעל 80.

<u>שאלה 5</u>

חוקר מעלה השערה ששתיית אלכוהול בעת היריון משפיעה על משקל הילודים. לבדיקת השערתו השקה באלכוהול 10 עכברים בהיריון. הוא שקל את הוולדות ומצא את המשקלים הבאים בגרמים:
.17, 17, 20, 19, 18, 19, 16, 21, 15, 22
ידוע שמשקל ממוצע של וולד רגיל הוא 20 גרם.

א. בנה רווח סמך ברמת בטחון 95%, למשקל הוולדות כפי שהתקבל מהמדגם. הסבר (במשפט אחד) מה ניתן להסיק ממנו לגבי השפעת האלכוהול באוכלוסיית העכברים.

ב. בהנחה שנתון כעת סטיית התקן באוכלוסייה היא 2.5 גרם, בנה רווח סמך ברמת בטחון 95% למשקל הוולדות.

ג. הסבר בקצרה (לא יותר מ- 5 שורות) מדוע רווח הסמך בסעיף ב' קטן יותר מרווח הסמך בסעיף א', למרות שסטיית התקן בסעיף ב' גדולה יותר.

<u>פתרון:</u>

יש לנו 10 תצפיות: $n = 10$.
בעזרת הנתונים נחשב את ממוצע המדגם ואת סטיית התקן המדגמית, בהם נעזר בהמשך:

$$\bar{x} = \frac{17+17+\ldots+22}{10} = 18.4 \qquad s = \left(\frac{1}{10-1}\right)\sqrt{(17-18.4)^2 + \ldots + (22-18.4)^2} = 2.221$$

א. לא נתונה סטיית התקן, לכן יש להשתמש בס. התקן המדגמית s ובהתפלגות t,

רווח הסמך נתון ע"י- $\bar{X} \pm t_{\left(n-1,1-\frac{\alpha}{2}\right)} \frac{s}{\sqrt{n}}$.

החישוב: $$18.4 \pm t_{(9,0.975)} \frac{2.221}{\sqrt{10}} = 18.4 \pm 2.62 \frac{2.221}{\sqrt{10}} \Rightarrow (16.81,\ 19.99)$$

המסקנה: האלכוהול גורם אפשרי לירידת המשקל הממוצע של וולדות עכברים ל- 18.4 גרם.
רווח הסמך ברמת בטחון 95% הוא $(17.112,\ 19.687)$.

ב. כעת נתונה סטיית התקן, לכן נשתמש בה ובהתפלגות נורמלית לבניית רווח הסמך שנוסחתו-

$$\bar{X} \pm z_{\left(1-\frac{\alpha}{2}\right)} \frac{\sigma}{\sqrt{n}}$$

החישוב:

$$18.4 \pm z_{(0.975)} \frac{2.5}{\sqrt{10}} = 18.4 \pm 1.96 \frac{2.5}{\sqrt{10}} \Rightarrow (16.85,\ 19.95)$$

ג. בסעיף א' לא נתונה סטיית התקן של האוכלוסייה, לכן אמדנו אותה בעזרת סטיית התקן המדגמית והשתמשנו בהתפלגות t שנותנת ערכים גדולים יותר מזו של ההתפלגות הנורמלית (עבור אותה רמת בטחון) ולכן קיבלנו את ההבדלים בגודל רו"ס בין הסעיפים.