תרגול כיתה 1 – מבוא להסתברות וסטטיסטיקה

סטטיסטיקה תיאורית – מדדי מרכז ופיזור

מתרגלים: ליאור דקל ואדם צ'פמן

שאלה 1 (סיווג משתנים)

א) סווג את המשתנים הבאים לבדידים/רציפים:

(1). גובה מעל פני הים. (2). צבע עיניים. (3). קצב המכוניות שנכנסות לחניון בשעה.

ב). סווג את המשתנים הבאים לפי סולמות מדידה (שמי/סדר/מרווחי/יחס)

(1). דרגה צבאית. (2). ציון במבחן כלשהו. (3). כמות דלק במיכל. (4). התשובות הבאות בשאלון טעימת גלידה חדשה: 1=אוהב; 2=אדיש; 3=לא אוהב.

פתרון: [הערה: הסיווג לא תמיד חד-משמעי ולעתים משתנה לפי הגדרת המשתנים והבעיה]

א). בדידים/רציפים:

(1). גובה מעל פני הים = רציף (לא נתון קנה המידה למדידה. אם היתה השאלה למשל 'במטרים' אז בדיד)

(2). צבע עיניים = בדיד (מקובל)

(3). קצב = רציף (מס' מכוניות חלקי יח' הזמן).

ב). סולמות מדידה:

(1). דרגה צבאית = סדר.

(2). ציון במבחן כלשהו = מרווחי (ניתן לקבוע ציונים: 200-800, 0-100 וכו').

יכול להיות גם יחס (תלוי בקובע סולם הציון ואין טעם להקדיש לסעיף אנרגיה מרובה)

(3). כמות דלק במיכל = יחס (הכמות 0 = אין נוזל במיכל).

(4). התשובות בשאלון טעימת גלידה = שמי.

שאלה 2 (חישוב מדדים)

לפניכם נתונים על שנות וותק בעבודה של 12 עובדים:

6, 7, 5, 8, 5, 6, 7, 7, 8, 6, 4, 6

א). חשב את המדדים המרכזיים של סדרת הנתונים: מינימום, מקסימום, ממוצע, חציון.

ב). חשב את מדדי הפיזור של הסדרה: טווח, שונות, סטיית תקן מדגמית, טווח בין-רבעוני.

ג). התברר כי חלה טעות ברישום הוותק לעובד בעל 4 שנות וותק, והוותק שלו הוא 5 שנים. הסבירו איכותית, כיצד ישפיע התיקון על כל אחד מהמדדים: השכיח, החציון והממוצע.

פתרון:

נסדר את הסדרה מהתצפית הקטנה לגדולה: 8, 8, 7, 7, 7, 6, 6, 6, 6, 5, 5, 4.

מינימום: התצפית הקטנה ביותר $X_{\min}$. בשאלה: 4

מקסימום: התצפית הגדולה ביותר $X_{\max}$. בשאלה: 8

ממוצע (Mean): הנוסחה- $\bar{x} = \frac{1}{n}\sum_i x_i$ בשאלה: $\bar{x} = \frac{1}{12}(4+\ldots+8) = 6.25$

שכיח (Mo-Mode): אותו ערך המופיע בשכיחות הגבוהה ביותר. ייתכנו מספר שכיחים. בשאלה: 6

אמצע הטווח (MR): $\frac{1}{2}(X_{\min}+X_{\max})$ בשאלה: 6=2/(8+4)

חציון (Me/Md-Median): כאשר נתונה רשימת תצפיות מסודרת –

אם מספר התצפיות אי-זוגי - זהו ערך התצפית המרכזית שמיקומה: $X_{(N+1)/2}$.

אם מספרן זוגי – זוהי התצפית שמיקומה: $\frac{1}{2}\left[X_{(N/2)}+X_{(N/2+1)}\right]$.

בשאלה: 6=2/(6+6) <= $N=12$

הערה: חציון הוא האחוזון ה-50. (ראה בהמשך)

ב). <u>מדדים לפיזור הנתונים:</u>

טווח - Range : $X_{max} - X_{min}$ <u>בשאלה</u>: 8 - 4 = 4

שונות המדגם ((sample variance:

אפשר להשתמש באחת משתי הנוסחאות-
$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}$$

<u>בשאלה</u>: $S^2 = 1.477$

סטיית התקן המדגמית (sd) – היא שורש השונות. $sd = \sqrt{s^2}$ <u>בשאלה</u>: sd = 1.215

טווח בין רבעוני – (Inter-Quantile Range): $\text{IQR} = Q_3 - Q_1$ – מחושב ע"י השלבים הבאים

למציאת אחוזונים Q_i(נשתמש בנוסחאות האמפיריות הבאות) -

<u>**אחוזונים (שברונים, ערכי חלוקה) - Quantiles**</u>

<u>**:**</u>

האחוזון ה- p הנו הערך x_p אשר $p\%$ מהתצפיות קטנות או שוות ממנו

ו- $(100-p)\%$ מהתצפיות גדולות או שוות ממנו.

דוגמאות לאחוזונים:

1. חציון = median = האחוזון ה- 50. נסמנו Q_2 או $x_{0.50}$.
2. רבעון תחתון = האחוזון ה- 25. נסמנו Q_1 או $x_{0.25}$.
3. רבעון עליון = האחוזון ה- 75. נסמנו Q_3 או $x_{0.75}$.

שלבי חישוב האחוזון p (נתונים בדידים)**:**

1) מיין את התצפיות מהקטנה לגדולה.

2) יהי $0 < p < 100$ האחוזון המבוקש, אזי חשב $Q = np/100$.

3) אם מתקבל מספר שלם, אזי האחוזון המבוקש: $\left(X_{(Q)} + X_{(Q+1)}\right)/2$.

4) אם מתקבל שבר, אזי האחוזון המבוקש: $X_{(k)}$, $k = \lceil Q \rceil$

($\lceil Q \rceil$ – פונקציית התקרה. לדוגמה: $\lceil 4.1 \rceil = 5$).

<u>חישוב הטווח הבין רבעוני</u>: $\text{IQR} = Q_3 - Q_1$

רבעון תחתון = האחוזון ה- 25:

$$N = 12, p = 25 \Rightarrow 12 \cdot 25/100 = 3 \Rightarrow Q_1 = \tfrac{1}{2}\left(X_{(3)} + X_{(4)}\right) = \tfrac{1}{2}(5+6) = 5.5$$

רבעון עליון = האחוזון ה-75 :

$$N=12, p=75 \Rightarrow 12\cdot 75/100=9 \Rightarrow Q_1=\tfrac{1}{2}\left(X_{(9)}+X_{(10)}\right)=\tfrac{1}{2}(7+7)=7$$

לכן: $\text{IQR}=Q_3-Q_1=7-5.5=1.5$

ג). הסדרה המתוקנת: 8, 8, 7, 7, 7, 6, 6, 6, 6, 5, 5, <u>5</u>.

- רואים שהחציון לא השתנה, התצפית המתוקנת לא שינתה את סדר התצפיות הקיים.
- ספירת המופעים של כל תצפית מראה שהשכיח לא השתנה.
- יש לנו אותו מס' תצפיות (N=12) כשערך תצפית אחת גדל, לכן הממוצע יגדל.

(חישוב מדויק יראה שהוא יגדל ב- $1/12$, אבל לא נשאלנו כמותית).

<u>שאלה 3 (טרנספורמציה לינארית/אפינית של מדדים)</u>

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?

<u>*פתרון:*</u>

השפעת טרנספורמציה לינארית/אפינית (לינארית=השם המקובל בספרות) $x \mapsto ax+b$

השפעת טרנספורמציה לינארית על מדד מרכז M: $\tilde{M}=aM+b$

השפעת טרנספורמציה לינארית על מדד פיזור D: $\tilde{D}=g(a)\,D$

באופן מפורט:

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

<u>משפיעה על המדדים באופן הבא</u>:

חציון: $Me_y=a\cdot Me_x+b$

ממוצע: $\bar{y}=a\cdot\bar{x}+b$

שכיח: $Mo_y=a\cdot Mo_x+b$

טווח: $Range_y=a\cdot Range_x$

אמצע טווח: $MR_y=a\cdot MR_x+b$

שונות: $S_y^2=a^2\cdot S_x^2$

סטיית תקן: $S_y=|a|\cdot S_x$

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

הממוצע:

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

החציון:

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

היא עברה את השינוי $\tilde{x}_{0.50}=x_{0.50}+12=65+12=77$

השונות:

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

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

הטווח:

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

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

הממוצע:

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

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

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

הטווח:

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

קומבינטוריקה ומרחב הסתברות אחיד (סימטרי)

<u>תמורה (פרמוטציה)</u> – בחירה של סדרת איברים כך שהסדר חשוב.

- תמורה חלקית (חליפה)- מאוכלוסיה בגודל n בוחרים k עצמים כך שהסדר חשוב.
  מס' התמורות האפשריות הוא-

$$(n)_k = A_n^k = P_{(n,k)} := \frac{n!}{(n-k)!} = n(n-1)(n-2)(n-3)....(n-(k-1))$$

- בתמורה מלאה, מונים בכמה דרכים ניתן לסדר קבוצה של n עצמים.
  (שקול לבחירת n איברים מתוך קבוצה בגודל n).
  $n! = n(n-1)(n-2)\cdots 2\cdot 1$
  תמורה מלאה היא פונקציה חח"ע ועל מהקבוצה על עצמה.

<u>תמורות במעגל</u> –מספר התמורות הוא (n-1)!, שמכיוון שבמעגל אין נקודת התחלה.

<u>צירוף (קומבינציה)</u>- בחירת קבוצה חלקית של k מתוך n עצמים כך <u>שהסדר אינו חשוב</u>.

$C_n^k = \binom{n}{k} := \frac{n!}{k!(n-k)!}$. כשיש חשיבות לסדר פשוט נחלק ב- k!.

<u>מקדמים מולטינומיים</u> – במקרה בו אנו רוצים לחלק n עצמים שונים זה מזה ל r קבוצות בגדלים $n_1, n_2, ..., n_r$ כאשר $\sum_{i=1}^{r} n_i = n$.

מס' החלוקות האפשרי הוא: $\binom{n}{n_1, n_2, ..., n_r} := \frac{n!}{n_1! n_2! \cdot ... \cdot n_r!}$

טבלת בחירת k עצמים מתוך n עצמים

	ללא החזרה	**עם החזרה**
ללא חשיבות לסדר הבחירה	$\binom{n}{k} = \frac{n!}{(k!)\cdot(n-k)!}$ דוגמה: בחירת ועד בגודל k מכיתה בגודל n.	$\binom{n+k-1}{n-1} = \binom{n+k-1}{k}$ דוגמה: k פריטים מסודרים בשורה, וצריך למקם n-1 מחיצות בינהם.
עם חשיבות לסדר הבחירה	$(n)_k = \frac{n!}{(n-k)!}$ דוגמה: בחירת רה"מ, שר-ביטחון ושר- חוץ מתוך כלל חברי הכנסת (n=120, k=3).	n^k דוגמאות: 1) יצירת מספר בין k ספרות כשמותר להשתמש ב- n ספרות. 2) חלוקת k כדורים ל- n תאים.

<u>מרחב הסתברות אחיד (סימטרי)</u>

מאורע – תוצאה אפשרית של ניסוי.

מרחב הסתברות אחיד – מרחב שבו לכל אחד מהמאורעות האפשריים הסתברות שווה.

ההסתברות שמאורע A יתרחש היא: $P(A)=\dfrac{|A|}{|\Omega|}$

כאשר Ω – מרחב המדגם. $|\Omega|$– גודל מרחב המדגם. $|A|$– גודל קבוצת המאורע A.

<u>שאלה 1</u>

במגירת גרביים: 7 גרביים שחורות, 8 כחולות, 9 ירוקות. הגרביים מפוזרות באקראי במגירה. 2 גרביים נבחרות באקראי:

א. **מה ההסתברות שהן מאותו הצבע ?**
ב. **מה ההסתברות שנבחרה לפחות גרב אחת שחורה ?**
ג. **מה ההסתברות שנבחרה גרב שחורה וגרב כחולה ?**

<u>פתרון:</u>

במגירה יש סה"כ 24 גרביים, מתוכם צריך לבחור 2 ולכן גודל מרחב המדגם $|\Omega|\equiv|N|=\binom{24}{2}=\frac{24\cdot 23}{2}=276$.

א. **לבחור 2 גרביים שחורות יש** $\binom{7}{2}=21$ **אפשרויות, לבחור 2 גרביים כחולות יש** $\binom{8}{2}=28$ **אפשרויות ולבחור 2 גרביים ירוקות יש** $\binom{9}{2}=36$ **אפשרויות.**

אילו מאורעות זרים שאפשר לחבר את הסתברותם, ונקבל שההסתברות היא:

$$\frac{21+28+36}{276}=\frac{85}{276}$$

ב. **נפתור בעזרת המאורע המשלים, לבחור 2 גרביים שאף אחת מהן איננה שחורה. יש 17 גרביים שאינן שחורות, ולכן יש** $\binom{17}{2}=136$ **אפשרויות לבחור 2 גרביים שאינן שחורות. כלומר,**

ההסתברות המבוקשת היא: $P=1-\overline{P}=1-\frac{136}{276}=\frac{140}{276}$.

ג. **לבחור גרב כחולה יש 8 אפשרויות, לבחור גרב שחורה יש 7 אפשרויות ולכן ההסתברות היא:** $\frac{8\cdot 7}{276}=\frac{56}{276}=\frac{14}{69}$.

<u>שאלה 2</u>

4 כדורים מוכנסים אקראית ל-3 תאים. מה ההסתברות ש-

א. תא מס' 1 נותר ריק.

ב. רק תא מס' 1 נותר ריק.

ג. יש בדיוק תא ריק אחד (כלשהו).

ד. יש תא ריק (כלשהו).

ה. יש תא עם בדיוק 2 כדורים.

<u>פתרון</u>:

גודל מרחב המדגם $|\Omega| = 3^4$.

א. **לכל כדור אפשרות להיות ב-2 תאים, ל-4 כדורים** 2^4 **אפשרויות. לכן:** $P = \frac{2^4}{3^4} = \frac{16}{81}$

ב. **מתוך 16 האפשרויות להימצאות הכדורים ב-2 תאים (חושב בסעיף א'), יש 2 אפשרויות "אסורות" בסעיף זה- האפשרות שכל הכדורים יימצאו בתא 2 או שכולם יימצאו בתא 3.**

לכן: $P = \frac{2^4 - 2}{2^4} = \frac{14}{81}$

ג. **המאורע בסעיף זה מורכב מאיחוד 3 מאורעות זרים ושווי הסתברות (מטעמי סימטריה)- (רק תא 1 ריק) U (רק תא 2 ריק) U (רק תא 3 ריק).**

הסתברות כל מאורע כזה שווה להסתברות שחושבה בסעיף ב', לכן:

$$P(C) = 3 \cdot \frac{14}{81} = \frac{42}{81}$$

ד. **נפרק למאורעות זרים-**

$$P = P(1 \ cell \ exactly \ empty) + P(2 \ cells \ exactly \ empty)$$
$$= P(C) + P(All \ in \ one \ cell)$$

ולכן: $P = \frac{42}{81} + 3 \cdot \frac{1}{81} = \frac{45}{81}$

ה. **שוב נפרק למאורעות זרים-**

$$P = P(1 \ cell \ exactly \ with \ 2 \ balls) + P(2 \ cells \ exactly \ with \ 2 \ balls)$$

ולכן: $P = \frac{\binom{3}{1} \cdot \binom{4}{2} \cdot 2! + \binom{3}{2} \cdot \binom{4}{2}}{3^4} = \frac{54}{81}$

<u>שאלה 3</u>

במכונית 4 נוסעים (כולל הנהג). מחלקים באקראי 11 בקבוקי שתייה זהים בין הנוסעים (שים לב שייתכן ונוסע לא מקבל כלל בקבוקי שתייה).

1) מה ההסתברות שכל נוסע יקבל לפחות בקבוק אחד?

2) כל נוסע מטיל קוביית משחק הוגנת. מה ההסתברות שלפחות שני נוסעים יקבלו אותה תוצאה?

<u>פתרון</u>:

א). נסמן A –כל נוסע יקבל לפחות בקבוק אחד. Ω – סה"כ מס' האפשרויות לחלוקת הבקבוקים.

צריך למצוא: $P(A) = \frac{|A|}{|\Omega|}$

<u>תחילה נחשב את סה"כ מס' האפשרויות לחלוקת הבקבוקים</u>

ניתוח: יש 4 תאים (הנוסעים) כלומר בינהם נמצאות 3 מחיצות. לתאים אלה יש לחלק 11 עצמים.אם נסמן מחיצה ע"י 1 ובקבוק ע"י 0, נוכל לייצג חלוקה ע"י המחרוזת: 0010000110000 (נוסע <u>א</u> מקבל

2 בקבוקים, נוסע ב 5 בקבוקים, נוסע ג לא מקבל בקבוקים ונוסע ד מקבל 4 בקבוקים). כל חלוקה כזו מיוצגת ע"י 14 סימנים: שלושה "1" ואחד-עשר "0".

לכן סה"כ מספר אפשרויות החלוקה: $|\Omega| = \binom{11+4-1}{11} = \frac{14!}{11!3!} = 364$

)זו גם הנוסחה של בחירת k עצמים מתוך n עצמים, עם החזרה, ללא חשיבות לסדר הבחירה(.

<u>כעת נחשב את מס' האפשרויות בהן כל נוסע מקבל לפחות בקבוק אחד</u>
כדי להבטיח שכל נוסע יקבל לפחות בקבוק אחד ניתן תחילה לכל אחד מהם בקבוק.
בשלב זה נשארו לחלוקה עוד 7 בקבוקים, לכן: $|A| = \binom{7+4-1}{7} = \binom{10}{7} = \frac{10!}{7!3!} = 120$

לכן ההסתברות המבוקשת: $P(A) = \frac{|A|}{|\Omega|} = \frac{120}{364} \approx 0.33$

ב). גודל מרחב המדגם $|\Omega| = 6^4$
נוח יותר לחשב לפי המאורע המשלים E^C: "כל אחת מהקוביות מראה מספר שונה".
מס' האפשרויות לקבלת 4 תוצאות שונות הוא $|E^C| = 6 \cdot 5 \cdot 4 \cdot 3 = (6)_4 = 360$.

לכן ההסתברות המבוקשת $P(E) = 1 - \frac{|E^C|}{|\Omega|} = 1 - \frac{360}{6^4} = \frac{13}{18} \approx 0.7222$

<u>שאלה 4</u>
קבוצה של 10 מטיילים צועדת בטור. 5 הראשונים חובשים כובעים אדומים, 2 אח"כ חובשים כובעים כחולים, ו- 3 האחרונים חובשים כובעים ירוקים.
1) לאחר שהתפזרו למנוחה קצרה, הקבוצה מסתדרת שוב לצעידה בצורה אקראית. מה ההסתברות שטור המטיילים ישוב לסדר צבע הכובעים כמוקדם?
2) בערב יושבים כל המטיילים סביב מדורה בסדר אקראי . 7 לובשים סוודרים כחולים ו- 3 ירוקים. מה ההסתברות שכל לובשי הסוודרים הירוקים יושבים זה לצד זה?

<u>פתרון</u>: [הדגמת שימוש במקדמים מולטינומיים]
א). סה"כ יש לסדר הצעידה לפי צבע הכובעים:
$$|\Omega| = \binom{10}{5,2,3} = \frac{10!}{5!\cdot 2!\cdot 3!} = 2520$$
לפני המנוחה צעדה הקבוצה בסדר צבעים מסויים (יחיד), לכן $|A| = 1$.

מכאן שההסתברות שטור המטיילים ישוב לסדר צבע הכובעים כמקודם היא: $P(A) = \frac{|A|}{|\Omega|} = \frac{1}{2520}$

ב). סה"כ מס' אפשרויות הישיבה של 10 אנשים במעגל: $|\Omega| = (10-1)! = 9! = 362880$
נסמן ב-A את האירוע שכל לובשי הירוק יושבים ביחד. מס' האפשרויות של אירוע זה:
$$|A| = 7!\cdot 3! = 30240$$
(הסבר: גוש הירוקים ו-7 הכחולים מהווים 8 עצמים שונים. אפשר לסדרם במעגל ב- 7! אופנים שונים. לכך יש להוסיף ולשקלל גם את הסדר הפנימי בגוש הירוקים 3!).
לכן ההסתברות המבוקשת היא:
$$P(A) = \frac{|A|}{|\Omega|} = \frac{30240}{362880} = \frac{1}{12}$$