הטכניון - מכון טכנולוגי לישראל

הפקולטה להנדסת תעשיה וניהול ע"ש ויליאם משה דוידסון

Technion - Israel Institute of Technology

The William Davidson Faculty of Industrial Engineering and Management

---

תאריך הבחינה: 18.09.15

שם המרצה: ד"ר אלה לוין

מבוא להסתברות וסטטיסטיקה – 094481

מבחן מועד ב' - סמסטר אביב תשע"ה

מבחן מועד א' - סמסטר קיץ תשע"ה

טור א' – כולל פתרון

משך הבחינה: שלוש שעות.

חומר מותר: שני דפי עזר A4 כתובים משני הצדדים, מחשבון כיס פשוט.

בנוסף מצורפים דפי התפלגויות מיוחדות, טבלאות התפלגויות דגימה וטבלת בדיקת השערות.

למבחן שני חלקים. **יש לענות על כל השאלות** בכל אחד מחלקי הבחינה:

א. החלק הראשון מכיל שאלות רב-ברירתיות (סגורות). **על חלק זה יש לענות רק בדף המיועד לכך המצורף למבחן**. אין לסמן יותר מתשובה אחת לכל שאלה. הציון יינתן על סמך הסימונים בדף התשובות בלבד!

ב. החלק השני מכיל שתי שאלות פתוחות, עליהן יש לענות **במחברת המצורפת ולא על גבי שאלון זה**. בחלק זה עליכם לפרט ולהסביר את כל השלבים שביצעתם בדרך לפתרון. את הפתרון יש לכתוב בעט שחור או כחול בלבד. ניתן להשתמש בטיוטה שלא תיבדק.

ב ה צ ל ח ה !

חלק ראשון – יש לענות על כל השאלות בחלק זה. משקל כל שאלה: 7 נק'

שאלה 1

נלקח מדגם בגודל 30 תצפיות מהתפלגות נורמלית $X \sim N(-4,2)$.

ההסתברות שהמ"מ X יקבל ערך חיובי היא:

א. $[0.002,0.006)$
ב. $[0.02,0.03)$
ג. $[0,0.0001)$
ד. $[0.5,0.6)$

פתרון:

$$X \sim N(-4, \sqrt{2}^2)$$
$$P(X \geq 0) = 1 - P(X \leq 0) = 1 - \phi(\frac{0+4}{\sqrt{2}}) = 1 - 0.9977 = 0.0023$$

ה.

שאלה 2

יהיו U_1, U_2, U_3 שלושה מ"מ ב"ת בעלי התפלגות אחידה בקטע $[0,1]$. נגדיר: $M = \max\{U_1, U_2, U_3\}$ ו- $W = \min\{U_1, U_2, U_3\}$. ההסתברות $P(W \geq 0.2, M \leq 0.8)$ נמצאת בקטע:

א. $[0, 0.2)$
ב. $[0.2, 0.4)$
ג. $[0.4, 0.6)$
ד. $[0.6, 0.8)$

פתרון

$$P(W \geq 0.2, M \leq 0.8) = \mathrm{P}(0.2 \leq \mathrm{U}_1 \leq 0.8,\ 0.2 \leq \mathrm{U}_2 \leq 0.8,\ 0.2 \leq \mathrm{U}_3 \leq 0.8) =$$
$$\underset{Ui\text{'s are independent}}{=} \mathrm{P}(0.2 \leq \mathrm{U}_1 \leq 0.8) \cdot \mathrm{P}(0.2 \leq \mathrm{U}_2 \leq 0.8) \cdot \mathrm{P}(0.2 \leq \mathrm{U}_3 \leq 0.8) =$$
$$\underset{Ui\text{'s are } equaly\ distributed}{=} \left(\mathrm{P}(0.2 \leq \mathrm{U}_1 \leq 0.8)\right)^3 =$$
$$\underset{Ui \sim U(0,1)}{=} \left(\frac{0.6}{1}\right)^3 = 0.216$$

שאלה 3

התפלגות תכונה מסוימת באוכלוסיה היא נורמלית עם שונות ידועה σ^2. 150 הוא גודל המדגם הדרוש על מנת להבטיח ברמת בטחון של 95% שהסטייה המקסימלית בין התוחלת לממוצע תהיה 0.5. מהו גודל המדגם הדרוש על מנת להבטיח סטייה מקסימלית זו ברמת בטחון של 99% ?

א. 211
ב. 257
ג. 260
ד. 156

פתרון:

d=0.5 הסטייה המקסימלית בין התוחלת לממוצע

עבור רמת בטחון של 95% :

עבור רמת בטחון של 99% :

$$n \geq \left[\frac{z_{0.995}\sigma}{d}\right]^2 = \left[\frac{2.576 \cdot 3.12}{0.5}\right]^2 = 259.102$$

גודל המדגם הדרוש הוא 260.

שאלה 4

אלברט משחק שחמט כל יום. הוא משחֵק משחַק אחד ביום בסיכוי של 0.8 או שני משחקים ביום בסיכוי של 0.2.

אם אלברט משחֵק משחַק אחד הוא מנצח בסיכוי של 0.7. אם הוא משחק שני משחקים הוא מנצח באחד מהם בלבד בסיכוי של 0.4 ובשניהם בסיכוי של 0.2.

מהי תוחלת מספר משחקי השחמט בהם אלברט ינצח ביום שישי הקרוב?

א. 0.736
ב. 0.56
ג. 0.72
ד. 0.80

פתרון

תוחלת מספר משחקי השחמט בהם אלברט ינצח ביום שישי הקרוב (אפשר לפתור גם לפי עץ החלטה)

$$1*(0.8*0.7+0.2*0.4)+2*(0.2*0.2)=0.72$$

טענה 3 נכונה

שאלה 5

על מנת לבחון האם למידה בקורס קריאה מהירה משפרת את כושר הקריאה של ילדים, ערך חוקר ניסוי. בניסוי נבחר מדגם מקרי של 10 ילדים והם חולקו לשתי קבוצות שוות באופן אקראי. הקבוצה הראשונה המשיכה בלימודים רגילים. הקבוצה השנייה המשיכה בלימודים בתוספת הקורס בקריאה מהירה. התוצאות היו כדלקמן:

	ציון ממוצע בכושר קריאה	שונות
קבוצה ראשונה	65	12
קבוצה שנייה	75	10

טענה I. על סמך הנתונים, ברמת מובהקות של 0.05, אפשר להסיק כי הקורס בקריאה מהירה משפר את כושר הקריאה.

טענה II. התפלגות הדגימה שמשתמשים בה בחישובים היא התפלגות נורמלית סטנדרטית, משום שהשונויות של ההתפלגויות ידועות.

א. טענה I נכונה, טענה II לא נכונה.
ב. טענה II נכונה, טענה I לא נכונה.
ג. שתי טענות נכונות
ד. שתי טענות לא נכונות

פתרון:

נסמן:

μ_1 - תוחלת ציון כושר הקריאה של תלמיד בקבוצה ראשונה

μ_2 - תוחלת ציון כושר הקריאה של תלמיד בקבוצה שנייה (לאחר הקורס)

$\bar{X}_i$ - ציון ממוצע בכושר קריאה של תלמידים בקבוצה i. $i=1,2$

א. רווח סמך להפרש התוחלות:

$$(\bar{X}_2-\bar{X}_1)+t^{(n_1+n_2-2)}_{1-\alpha/2}\sqrt{S_p^{\,2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\le \mu_2-\mu_1\le (\bar{X}_2-\bar{X}_1)+t^{(n_1+n_2-2)}_{1-\alpha/2}\sqrt{S_p^{\,2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

לשברון ה-t עם מס' ד"ח מתאים: $t^{(n_1+n_2-2)}_{1-0.025}=t^{(8)}_{0.975}=2.306$

$$S_p^{\,2}=\frac{(n_1-1)S_1^{\,2}+(n_2-1)S_2^{\,2}}{n_1+n_2-2}=\frac{4\cdot 10+4\cdot 12}{8}=11$$

$$\mu_2-\mu_1\in 10\pm 1.86\sqrt{11\left(\frac{2}{5}\right)}$$

$$6.01\le \mu_2-\mu_1\le 13.9$$

רווח הסמך שקיבלנו הוא חיובי ועל כן בר"מ של 0.05 ניתן להסיק כי המבחן משפר יכולות הקריאה. טענה 1 נכונה.

טענה 2 אינה נכונה משום שהשוניות האמתיות אינן ידועות. נתונות רק השוניות המדגמיות.

<u>שאלה 6</u>

להלן 3 טענות:

<u>טענה I:</u> $E\sum_{i=1}^{n}(x_i-\mu)^2=\sigma^2$

טענה II: $E\sum_{i=1}^{n}\left(\mu-\bar{x}\right)^{2}=n\sigma^{2}$

טענה III: $E\left(\sum_{i=1}^{n}\left(x_i-\mu\right)\left(\mu-\bar{x}\right)\right)=-\sigma^{2}$

א. כל הטענות נכונות.
ב. בדיוק טענה אחת נכונה.
ג. בדיוק שתי טענות נכונות.
ד. אף טענה לא נכונה.

פתרון

טענה I:

$$E\sum_{i=1}^{n}\left(x_i-\mu\right)^{2}=\sum_{i=1}^{n}E\left(x_i-\mu\right)^{2}=\sum_{i=1}^{n}Var(x_i)=n\sigma^{2}$$

טענה II:

$$E\sum_{i=1}^{n}\left(\mu-\bar{x}\right)^{2}=\sum_{i=1}^{n}E\left(\mu-\bar{x}\right)^{2}=\sum_{i=1}^{n}V\operatorname{ar}\bar{x}=\sum_{i=1}^{n}\sigma^{2}/n=\sigma^{2}$$

טענה III:

$$E\left(\sum_{i=1}^{n}\left(x_i-\mu\right)\left(\mu-\bar{x}\right)\right)=E\left(\mu-\bar{x}\right)\sum_{i=1}^{n}\left(x_i-\mu\right)=$$
$$=E\left(\mu-\bar{x}\right)\left(\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\mu\right)=E\left(\mu-\bar{x}\right)\left(n\bar{x}-n\mu\right)=-nE\left(\mu-\bar{x}\right)^{2}=$$
$$=-n\frac{\sigma^{2}}{n}=-\sigma^{2}$$

רק הטענה השלישית נכונה

שאלה 7

בעיסה המיועדת להכנת 50 עוגיות הוכנסו 150 פתיתי שוקולד. במאפיה מבטיחים "לפחות שני פתיתי שוקולד בעוגיה, או שכספכם יוחזר".

בהנחה שמספר הפתיתים בעוגיה, X, מפולג פואסונית.

א. ההסתברות שלא תצטרך לשלם עבור העוגיה שווה ל- 1/2
ב. ההסתברות שלא תצטרך לשלם עבור העוגיה קטן מ- 1/2
ג. ההסתברות שלא תצטרך לשלם עבור העוגיה גדול מ- 1/2
ד. לא ניתן לחשב את ההסתברות שלא תצטרך לשלם עבור העוגיה

פתרון

$$X \sim Poiss(3)$$
$$\lambda = 3$$
$$P(X < 2) = P(0) + P(1) = e^{-3} + 3e^{-3} = 4e^{-3} = 0.199 \approx 0.2$$

חלק שני - שאלות פתוחות. יש לענות על שתי השאלות שבחלק זה.

שאלה 8 (33 נקודות)

חברת סלולארי מעוניינת לחקור את ההשערה שמשך שיחה ממוצע של נשים גבוה מזה של גברים. ידוע כי סטיית התקן של משך השיחה אצל נשים שווה ל-12 דקות, וסטיית התקן של משך השיחה אצל גברים שווה ל-9.75 דקות. במדגם מקרי שנערך השתתפו שתי קבוצות ב"ת: 90 נשים ו-70 גברים. מכל פרט נלקחה באופן אקראי תצפית בודדת של משך שיחה. הוחלט שאם ממוצע משך השיחות אצל נשים יהיה גבוה ביותר מ-4 דקות מאשר הממוצע אצל הגברים נדחה את השערת האפס.

הניחו כי משך השיחות בכל מדגם מתפלג נורמלית.

א. נסחו את ההשערות והציגו את סטטיסטי המבחן המתאים. כיצד מתפלג סטטיסטי המבחן תחת השערת האפס? יש לציין את הפרמטרים של ההתפלגות.
ב. מהו אומד נקודתי לפרמטר הנחקר, כיצד הוא מפולג תחת השערת האפס? יש לציין את הפרמטרים של ההתפלגות. האם הוא חסר הטיה או מוטה (הוכח)?

ג. מהי רמת המובהקות של המבחן עליו הוחלט?

ד. אם תוחלת משך השיחה אצל נשים אמנם גבוהה ב-7 דקות מאשר אצל הגברים, מהי עוצמת המבחן? מהי ההסתברות שנסיק על סמך המדגם כי אין הבדל בתוחלות? כיצד נקראת הסתברות זו?

ה. נתון כי ממוצע השיחות אצל גברים (במדגם) היה 8.8 דקות וממוצע השיחות אצל נשים (במדגם) היה 11.9 דקות. בדוק את ההשערה שמשך שיחה ממוצע של נשים גבוה מזה של גברים באמצעות סטטיסטי המבחן ו-P-value עבור רמת מובהקות של 0.01. מהי מסקנתך?

ו. על סמך P-value מהי מסקנתך עבור רמת בטחון של 95%?

פתרון

X_1 - אורך שיחה של אישה.

X_2 - אורך שיחה של גבר.

$$X_1 \sim N\left(\mu_1, 12^2\right) \qquad X_2 \sim N\left(\mu_2, 9.75^2\right)$$

$n_1 = 90$ - מדגם מקרי של נשים (ולכן השיחות ב״ת).

$n_2 = 70$ - מדגם מקרי של גברים (ולכן השיחות ב״ת).

ובנוסף, שני המדגמים ב״ת ביניהם.

א. מערכת ההשערות:

$$H_0: \mu_1 - \mu_2 = 0 \qquad (\mu_1 = \mu_2)$$
$$H_1: \mu_1 - \mu_2 > 0 \qquad (\mu_1 > \mu_2)$$

סטטיסטי המבחן עבור הפרש תוחלות כאשר שני המדגמים ב״ת והשונויות ידועות (מופיע בדף מבחני השערות):

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \overset{H_0}{\sim} N(0,1)$$

ב. אומד נקודתי לפרמטר $\mu_1 - \mu_2$ הוא הפרש ממוצעים $\bar{X}_1 - \bar{X}_2$ כך ש-

$$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

$$E\left(\bar{X}_1 - \bar{X}_2\right) = \mu_1 - \mu_2$$

$$Var\left(\bar{X}_1 - \bar{X}_2\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$E\left(\bar{X}_1 - \bar{X}_2\right) = E\left(\bar{X}_1\right) - E(\bar{X}_2) = \left(\mu_1 - \mu_2\right)$$

לכן, אומד חסר הטיה.

ג. נתון כלל הכרעה: הוחלט שאם משך השיחה הממוצע אצל נשים יהיה גבוה ביותר מ-4 דקות מאשר אצל הגברים, נדחה את השערת האפס. אזי:

$$C_\alpha = \left(\mu_1 - \mu_2\right) + z_{1-\alpha} * \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 0 + z_{1-\alpha} * \sqrt{\frac{12^2}{90} + \frac{9.75^2}{70}} = 0 + z_{1-\alpha} * \sqrt{2.958} = 4$$

$$\Rightarrow z_{1-\alpha} = \frac{4}{1.72} = 2.326$$

$$\Rightarrow 1 - \alpha = \Phi(2.326) = 0.99$$

לכן $\alpha = 0.01$ (רמת המובהקות שווה ל-0.01).

ד. נחשב את עוצמת המבחן במקרה שההפרש האמיתי הוא 7 דקות:

$$\pi = P_{\mu_1 - \mu_2 = 7}\begin{pmatrix} reject \\ H_0 \end{pmatrix} = P_{\mu_1 - \mu_2 = 7}\left(\bar{X}_1 - \bar{X}_2 > 4\right) = P_{\mu_1 - \mu_2 = 7}\left(\frac{(\bar{X}_1 - \bar{X}_2) - 7}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > \frac{4 - 7}{\sqrt{2.958}}\right) =$$

$$= 1 - \Phi(-1.744) = \Phi(1.744) \approx 0.9595$$

אנו יודעים את התפלגותו של סטטיסטי המבחן במדויק ולכן ניתן לחשב את ההסתברות.

ההסתברות שנסיק לפי המדגם כי אין הבדל בתוחלות היא **ההסתברות לטעות מסוג שני** –תוחלת משך השיחה אצל נשים אכן גבוהה יותר ב-7 דקות מאשר התוחלת אצל הגברים , אך בטעות לא דוחים את H_0 .

נחשב את ההסתברות הזאת :

$$\beta = 1 - \pi = 1 - 0.9595 = 0.0405$$

ה. נחשב ראשית את ערך סטטיסטי המבחן שהתקבל בפועל :

וכעת ניתן לחשב pvalue :

p -

ראינו בסעיף ב׳ כי רמת המובהקות הינה 0.01 ועל כן לא נדחה את השערת האפס (התקבל pvalue גדול מרמת המובהקות).

ו. עבור רמת בטחון של 95%

$\alpha > P - value$ לכן נדחה את H_0 ברמת מובהקות 0.05

<u>שאלה 9 (18 נקודות)</u>

משרד הבריאות טוען כי ישנו קשר בין מספר שעות הצפייה השבועיות בתכניות ריאליטי (X) לבין רמת התקשורת הבין אישית במשפחה (Y). לצורך בדיקת הטענה, נבדקו בחמש משפחות הרגלי הצפייה ורמת התקשורת הבין אישית במשפחה. הנתונים:

משפחה	**1**	**2**	**3**	**4**	**5**	**ממוצע**	**סטיית תקן**
שעות צפייה	9	3	3	1	4	4	3
תקשורת	78	104	95	119	104	100	15

נתון גם : $\sum_{i=1}^{n}(x_i \cdot y_i) - n \cdot \bar{x} \cdot \bar{y} = -166$

א. בעזרת איזה מדד יש לבדוק את כיוון ועוצמת הקשר הלינארי? חשב מדד זה והסק לגבי טיב הקשר וכיוונו.

ב. מהו רווח הסמך של שיפוע הרגרסיה ברמת מובהקות של 0.05? האם לפי רווח סמך זה ניתן לומר שהרגרסיה הלינארית מובהקת?

ג. חשב את האומדנים עבור השיפוע והחותך של קו הרגרסיה.

ד. משפחת קדמוני מבלה 4 שעות בשבוע בצפייה בטלוויזיה. מהי עתידה להיות רמת התקשורת הבין אישית במשפחה זו?

<u>פתרון:</u>

$$S_{xx} = S_x{}^2(n-1) = 3^2 \cdot 4 = 36$$
$$S_{yy} = S_y{}^2(n-1) = 15^2 \cdot 4 = 900$$

א. המדד בו נשתמש הוא מקדם המתאם הלינארי:

$$R = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} = \frac{-166}{\sqrt{36 \cdot 900}} = -0.9212$$

ערכו של מקדם המתאם מעיד על קשר לינארי שלילי חזק בין X ל- Y.

ב. רווח סמך לשיפוע הרגרסיה:

$$T = \sqrt{\frac{R^2(n-2)}{1-R^2}} = 4.1$$

$$\hat{\beta}_1 = R \cdot \frac{S_Y}{S_x} = -4.61$$

$$T = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \qquad \rightarrow \qquad \hat{\sigma}_{\hat{\beta}_1} = \frac{\left|\hat{\beta}_1\right|}{T} = \frac{4.61}{4.1} = 1.124$$

$$\hat{\beta}_1 - t_{n-2}^{1-\alpha/2}\hat{\sigma}_{\hat{\beta}_1} \le \hat{\beta}_1 \le \hat{\beta}_1 + t_{n-2}^{1-\alpha/2}\hat{\sigma}_{\hat{\beta}_1}$$
$$t_3^{0.975} = 3.182$$
$$\beta_1 \in \left[-4.61 \pm 3.182 \cdot 1.124\right]$$
$$\beta_1 \in \left[-8.19, -1.03\right]$$

.ג

$$\hat{\beta}_1 = R \cdot \frac{S_Y}{S_x} = -4.6$$
$$\overline{Y} = \hat{\beta}_1 \cdot \overline{X} + \hat{\beta}_0$$
$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \cdot \overline{X} = 118.4$$
$$\Downarrow$$
$$\hat{Y} = -4.6X + 118.4$$

.ד

$$\hat{Y}(X = 4) = -4.6 \cdot 4 + 118.4 = 100$$