הסקה סטטיסטית שיעור 9

הקדמה

חומר קריאה 9א

תכנון מבחן NHST

- נגדיר השערת אפס והשערה אלטרנטיבית.
- נבחר סטטיסטי מבחן שאנחנו יודעים את התפלגות האפס והאלטרנטיבית שלו.
- נגדיר תחום דחייה. בדרך כלל זה נעשה באופן מרומז, על ידי הגדרת רמת מובהקות α ושיטה לחישוב ערכי p לפי הזנבות של התפלגות האפס.
- נחשב את העוצמה לפי ההתפלגות האלטרנטיבית.

ביצוע NHST

- נאסוף דאטא ונחשב את סטטיסטי המבחן.
- נבדוק אם סטטיסטי המבחן נמצא בטווח הדחייה. בדרך כלל על ידי בדיקה האם $p < \alpha$. אם כן, נדחה את השערת האפס H_0. אחרת, **לא נדחה** את השערת האפס.

דוגמה – פרמטרים אוכלוסייתיים וסטטיסטיקות מדגמיות

אם נבחר 10 גברים באקראי מאוכלוסיה ונמדוד את גובהם, נאמר ש**דגמנו את הגובה** מהאוכלוסייה. במקרה זה, ממוצע המדגם $\bar{x}$ הוא ממוצע הגבהים שנדגם. זה סטטיסטי, ואנחנו יודעים את הערך שלו. לעומת זאת, ממוצע הגובה האמיתי באוכלוסייה μ אינו ידוע וניתן להעריך אותו בלבד. μ נקרא **פרמטר של אוכלוסייה.**

המטרה העיקרית של בדיקות מובהקות היא להשתמש בסטטיסטיקות מדגם כדי להסיק דברים על פרמטרים של אוכלוסייה. לדוגמה, נוכל לבדוק אם ממוצע הגובה של גברים באוכלוסייה מסוימת גדול מ-70 אינץ' (בערך 1.78 מטר).

מבחני מובהקות נפוצים הקשורים להתפלגות נורמלית

נראה מספר מבחנים שכולם מניחים שהדאטא נורמלית. שימו לב לסוגי ההשערות שהמבחנים מיועדים להבדיל ביניהן, ולהנחות הנדרשות על הנתונים כדי שהמבחנים יהיו תקפים. התפלגויות האפס בכל המבחנים האלו קשורות להתפלגות נורמלית על ידי נוסחה מוגדרת.

מבחן z

- שימוש: בדיקה אם ממוצע אוכלוסייתי שווה לממוצע משוער.
- דאטא: $x_1, \dots x_n$
- הנחות: הדאטא היא דגימות נורמליות אקראיות בת"ל: $x_i \sim N(\mu, \sigma^2)$ כאשר μ לא ידוע, σ ידוע.
- השערת האפס H_0: עבור μ_0 נתון, $\mu = \mu_0$.
- השערה אלטרנטיבית H_A:
  - דו צדדית: $\mu \neq \mu_0$
  - חד צדדי "גדול מ": $\mu > \mu_0$
  - חד צדדי "קטן מ": $\mu < \mu_0$
- סטטיסטי המבחן: $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$
- התפלגות האפס: $f(z \mid H_0)$ היא הPDF של $Z \sim N(0,1)$
- ערכי p:
  - דו צדדי: $p = P(|Z| > z)$
  - חד צדדי "גדול מ": $p = P(Z > z)$
  - חד צדדי "קטן מ": $p = P(Z < z)$

מבחן t למדגם אחד לבדיקת ממוצע

- שימוש: בדיקה אם ממוצע אוכלוסייתי שווה לממוצע משוער.
- דאטא: $x_1, \dots x_n$
- הנחות: הדאטא היא דגימות נורמליות אקראיות בת"ל: $x_i \sim N(\mu, \sigma^2)$ כאשר μ, σ לא ידועים.
- השערת האפס H_0: עבור μ_0 נתון, $\mu = \mu_0$.
- השערה אלטרנטיבית H_A:
  - דו צדדית: $\mu \neq \mu_0$
  - חד צדדי "גדול מ": $\mu > \mu_0$
  - חד צדדי "קטן מ": $\mu < \mu_0$
- סטטיסטי המבחן: $z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$, כאשר $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ הוא השונות של המדגם.
- התפלגות האפס: $f(t \mid H_0)$ היא הPDF של $T \sim t(n-1)$ (התפלגות סטודנט t עם $n-1$ דרגות חופש).
- ערכי p:
  - דו צדדי: $p = P(|T| > t)$
  - חד צדדי "גדול מ": $p = P(T > t)$
  - חד צדדי "קטן מ": $p = P(T < t)$

מבחן *t* לשני מדגמים להשוואת ממוצעים – המקרה של שונויות שוות

- שימוש: בדיקה אם הממוצעים האוכלוסייתיים משתי אוכלוסיות שונים במידה משוערת.
- דאטא: $x_1, \dots x_n$, $y_1, \dots, y_n$.
- הנחות: שתי הקבוצות הן דגימות נורמליות בת"ל:
  - $x_i \sim N(\mu_x, \sigma^2)$
  - $y_i \sim N(\mu_y, \sigma^2)$
  - כאשר μ_x, μ_y ידועים ואולי שונים. השונות σ^2 לא ידועה, אבל זהה.
- השערת האפס H_0: עבור μ_0 נתון, $\mu_x - \mu_y = \mu_0$.
- השערה אלטרנטיבית H_A:
  - דו צדדית: $\mu_x - \mu_y = \mu_0$
  - חד צדדי "גדול מ": $\mu_x - \mu_y > \mu_0$
  - חד צדדי "קטן מ": $\mu_x - \mu_y < \mu_0$
- סטטיסטי המבחן:

$$t = \frac{\bar{x} - \bar{y} - \mu_0}{s_P}$$

כאשר s_x^2, s_y^2 הם השונויות של המדגמים, ו-s_P הוא השונות הכוללת:

$$s_P^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right), \qquad df = n + m - 2$$

- התפלגות האפס: $f(t \mid H_0)$ היא הPDF של $T \sim t(df)$ (התפלגות סטודנט t עם $n+m-2$ דרגות חופש).
- ערכי p:
  - דו צדדי: $p = P(|T| > t)$
  - חד צדדי "גדול מ": $p = P(T > t)$
  - חד צדדי "קטן מ": $p = P(T < t)$

המקרה של שונויות שונות

נקרא גם **מבחן *t* של וולש.** דומה מאוד למקרה של שונויות שוות, עם כמה הבדלים:

- הנחות: שתי הקבוצות הן דגימות נורמליות בת"ל:
  - $x_i \sim N(\mu_x, \sigma_x^2)$
  - $y_i \sim N(\mu_y, \sigma_y^2)$
  - כאשר μ_x, μ_y ידועים ואולי שונים. השונויות σ_x^2, σ_y^2 לא ידועות, ולא מניחים שהן שוות.
- סטטיסטי המבחן זהה. השונות הכוללת:

$$s_P^2 = \frac{s_x^2}{n} + \frac{s_y^2}{m}, \qquad df = \frac{\left((s_x^2/n) + (s_y^2/m)\right)^2}{\frac{(s_x^2/n)^2}{n-1} + \frac{\left(s_y^2/m\right)^2}{m-1}}$$

- התפלגות האפס: אותו דבר עם ה- df החדש.

מבחן t לזוגות בשני מדגמים (*paired two-sample t-test*)

כאשר הדאטא מגיעה באופן טבעי בזוגות (x_i, y_i), אפשר להשתמש במבחן t לזוגות בשני מדגמים (אם ההנחות תקפות).

דוגמאות: כדי למדוד את ההשפעה של תרופה להורדת כולסטרול, נבדוק כל מטופל לפני ואחרי הטיפול. או, כדי לבדוק את ההשפעה של טיפול לסרטן, נשווה אדם שקיבל את הטיפול מול אדם שלא קיבל. ונרצה להתאים זוגות לפי גיל, משקל וכו.

שימוש: כדי לבדוק אם ההבדל הממוצע בין החלקים של הזוג שווה לערך משוער.

דאטא: זוגות של $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

הנחות: ההבדלים $w_i = x_i - y_i$ הם מ"מ בת"ל, ומתפלגים נורמלית $N(\mu, \sigma^2)$ עם פרמטרים לא ידועים.

נשים לב: זה פשוט מבחן t למדגם יחיד עם w_i.

השערת האפס H_0: עבור μ_0 ספציפי, $\mu = \mu_0$.

ההשערה האלטרנטיבית H_A: דו צדדי: $\mu \neq \mu_0$, חד צדדי גדול: $\mu > \mu_0$, חד צדדי קטן: $\mu < \mu_0$.

סטטיסטי המבחן:

$$t = \frac{\overline{w} - \mu_0}{s/\sqrt{n}}$$

כאשר s היא שונות המדגם:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(w_i - \overline{w})^2$$

התפלגות האפס: $f(t \mid H_0)$ היא ה-PDF של $T \sim t(n-1)$. (התפלגות סטודנט עם $n-1$ דרגות חופש).

ערך *p-value*: דו צדדי: $p = P(|T| > t)$, חד צדדי גדול: $P(T > t)$, חד צדדי קטן: $P(T < t)$.

דוגמה: כדי למדוד את ההשפעה של עישון על הצטברות טסיות דם, לוין (1973) לקח דגימות דם מ-11 נבדקים לפני ואחרי עישון סיגריה, ומדד את רמת ההצטברות. הדאטא:

Before	25	25	27	44	30	67	53	53	52	60	28
After	27	29	37	56	46	82	57	80	61	59	43
Difference	2	4	10	12	16	15	4	27	9	-1	15

השערת האפס היא שאין השפעה, כלומר $\mu_0 = 0$. המבחן נתן ערך *p-value* 0.001633.

מבחן *ANOVA* בכיוון אחד (*F-test for equal means*)

שימוש: כדי לבדוק אם ממוצעי האוכלוסייה מ-n קבוצות כולם שווים.

דאטא: n קבוצות עם m דגימות מכל אחת: (איזה כיף מטריצות)

$$\begin{array}{cccc} x_{1,1}, & x_{1,2}, & \dots, & x_{1,m} \\ x_{2,1}, & x_{2,2}, & \dots, & x_{2,m} \\ & & \dots & \\ x_{n,1}, & x_{n,2}, & \dots, & x_{n,m} \end{array}$$

הנחות: הדאטא של כל קבוצה היא דגימה בת"ל נורמלית מתוך התפלגויות עם ממוצעים שיכולים להיות שונים, אבל עם אותה שונות:

$$x_{1,j} \sim N(\mu_1, \sigma^2), x_{2,j} \sim N(\mu_2, \sigma^2), \dots, x_{n,j} \sim N(\mu_n, \sigma^2)$$

הממוצעים והשונות לא ידועים.

השערת האפס H_0: כל הממוצעים שווים $\mu_1 = \mu_2 = \cdots = \mu_n$.

ההשערה האלטרנטיבית H_A: לא כל הממוצעים שווים.

סטטיסטי המבחן: $w = \frac{MS_B}{MS_W}$, כאשר:

$\bar{x}_i$ הוא הממוצע של קבוצה i:

$$\bar{x}_i = \left(x_{i,1} + x_{i,2} + \cdots + x_{i,m}\right)/m$$

$\bar{x}$ הוא הממוצע של כל הדאטא.

s_i^2 הוא שונות המדגם של קבוצה i:

$$s_i^2 = \frac{1}{m-1}\sum_{j=1}^{m}\left(x_{i,j} - \bar{x}_i\right)^2$$

MS_B השונות בין קבוצות. m כפול שונות המדגם של ממוצעי הקבוצות:

$$MS_B = \frac{m}{n-1}\sum_{i=1}^{n}(\bar{x}_i - \bar{x})^2$$

MS_W הממוצע של השונויות בתוך הקבוצות:

$$MS_W = (s_1^2 + s_2^2 + \cdots + s_n^2)/n$$

הרעיון: אם ה μ_i כולם שווים, היחס הזה יהיה קרוב ל-1.אם הם שונים אז MS_B יהיה גדול יותר ו- MS_W יישאר בערך אותו דבר, אז w יגדל.

התפלגות האפס: $f(w|H_0)$ היא ה-PDF של $W \sim F\left(n-1,\ n(m-1)\right)$. התפלגות F עם $(n-1)$, ו-$(n(m-1))$ דרגות חופש.

***p-value*:** $p = P(W > w)$.

הערות: $ANOVA$ בודק האם כל הממוצעים שווים. הוא לא בודק האם יש תת קבוצה של ממוצעים שהם שווים.

דוגמה: רמות הכאב המורגשות של מטופלים (בסולם 1 עד 6), אחרי 3 טיפולים שונים:

T_1	T_2	T_3
2	3	2
4	4	1
1	6	3
5	1	3
3	4	5

ה p-$value$ המתקבל הוא 0.729. בכל רמת מובהקות סבירה, נדחה את השערת האפס.

מבחן כי-בריבוע לבדיקת טיב התאמה (*Chi-square test for goodness of fit*)

בודק עד כמה התפלגות משוערת מתאימה לדאטא. סטטיסטי המבחן נקרא **סטטיסטי כי-בריבוע** והתפלגות האפס היא **התפלגות כי בריבוע**. היא מסומנת $\chi^2(df)$, כאשר הפרמטר df נקרא דרגות חופש.

נניח שיש לנו PMF לא ידועה הנתונה ע"י הטבלה הבאה:

Outcomes	ω_1	ω_2	$\dots$	ω_n
Probabilities	p_1	p_2	$\dots$	p_n

אנחנו משערים קבוצה של ערכים להסתברויות. בדרך כלל נשער שההסתברויות מגיעות מהתפלגות ידועה עם פרמטרים מסויימים (בינומית, פואסון, מולטינומית). המבחן מנסה לקבוע האם סביר שהאוסף הזה של הסתברויות יניב את הדאטא שקיבלנו.

שימוש: בדיקה האם דאטא בדידה מתאימה ל-PMF סופית מסויימת.

דאטא: ספירה נצפית O_i לכל תוצאה אפשרית ω_i.

הנחות: אין.

השערת האפס H_0: הדאטא הגיעה מהתפלגות בדידה מסויימת.

ההשערה האלטרנטיבית H_A: הדאטא הגיעה מהתפלגות אחרת.

סטטיסטי המבחן: הדאטא היא ספירה O_i לכל ω_i. מהשערת האפס, נקבל קבוצת ספירות צפויות E_i. יש שני סטטיסטים אפשריים:

$$\text{Likelihood ratio statistic: } G = 2 \cdot \sum O_i \ln\left(\frac{O_i}{E_i}\right)$$

$$\text{Pearson's chi-square statistic: } X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

תחת השערת האפס, $G \approx X^2$ שניהם בערך כי-בריבוע.

דרגות חופש df: במבחני כי-בריבוע, מספר דרגות החופש יכול להיות מסובך. במקרה הזה, $df = n - 1$. הוא מחושב לפי הספירות בתאים שיכולים להיות מוגדרים בחופשיות תחת H_A, בהתאם לסטטיסטיקות הנדרשות כדי לחשב את הספירות המצופות בהנחת H_0.

התפלגות האפס: בהנחת H_0, שני הסטטיסטיים מתפלגים (בערך) לפי התפלגות כי-בריבוע עם df דרגות חופש. כלומר ל- $f(G \mid H_0)$ ול- $f(X^2 \mid H_0)$ יש את אותה PDF כמו $Y \sim \chi^2(df)$.

***p-value*:** $p = P(Y > G),\ p = P(Y > X^2)$.

הערה: כאשר משתמשים ב-G, המבחן נקרא גם **מבחן G** או ***likelihood ration test***.

דוגמה

נניח שיש לנו ניסוי שנותן דאטא מספרית. בניסוי הזה התוצאות האפשריות הן: 1, 2, 3, 4, 5 או יותר. ביצענו 51 בדיקות וספרנו את התדירות של כל תוצאה:

Outcomes	0	1	2	3	4	≥ 5
Observed counts	3	10	15	13	7	3

השערת האפס היא שהדאטא מגיעה מהתפלגות $bin(8, 0.5)$.

Outcomes	0	1	2	3	4	≥ 5
Observed counts	3	10	15	13	7	3
H_0 probabilities	0.0039	0.0313	0.1094	0.2188	0.2734	0.3633
Expected counts	0.19	1.53	5.36	10.72	13.40	17.80

מתקיים: $X^2 = 116.41, G = 66.08$

הסטטיסטי היחיד שהשתמשנו בו הוא מספר הניסויים, 51. אז יש 5 דרגות חופש (כי אפשר לקבוע את 5 התאים הראשונים בחופשיות, ורק האחרון נקבע כי צריך שהמספר הכולל יהיה 51).

שני ה p-$value$ הם בפועל כמעט 0. אז נדחה את השערת האפס בכל רמת מובהקות כמעט.

דוגמה 10 – דרגות חופש

נניח שיש את אותה הדאטא מהניסוי הקודם, אבל השערת האפס היא שהדאטא מגיעה מניסויים בת"ל שמתפלגים $bin(8, \theta)$, עם פרמטר θ שיכול להיות כל דבר. במקרה הזה נצטרך להעריך את θ מתוך הדאטא, על ידי MLE. בסך הכל חישבנו שני ערכים מתוך הדאטא: מספר הספירות הכולל וההערכה של θ. אז דרגות חופש: $6 - 2 = 4$.

דוגמה 11 – ניסויים גנטיים של גרגור מנדל

באחד הניסויים, הוא שילב 556 זוגות של אפונה: זכר צהוב חלק, עם נקבה ירוקה מקומטת. בהנחה שהגנים של קמטים וחלק באים באותה תדירות, נצפה שלרבע מהאוכלוסייה יהיו שני גנים חלקים (SS), רבע שני גנים מקומטים (ss), וחצי עם אחד ואחד (Ss). נצפה לאותם האחוזים של צהוב (Y) וירוק (y). אם הצבע והקמטים מורשים באופן בת"ל, וחלק וצהוב הם הדומיננטיים, נצפה לקבל (זוהי השערת האפס):

	Yellow	Green	
Smooth	9/16	3/16	3/4
Wrinkled	3/16	1/16	1/4
	3/4	1/4	1

אז נצפה לראות $556 \cdot \frac{9}{16} = 312.75$ צהובים חלקים.

הטבלה הבאה מתארת את הספירות המצופות ובפועל:

	Observed count	Expected count
Smooth yellow	315	312.75
Smooth green	108	104.25
Wrinkled yellow	102	104.25
Wrinkled green	31	34.75

הסטטיסטים:

$$\text{Likelihood ratio statistic: } G = 2 \cdot \sum O_i \ln\left(\frac{O_i}{E_i}\right) =$$

$$2\left(315 \ln\left(\frac{325}{312.75}\right) + 108 \ln\left(\frac{108}{104.25}\right) + 102 \ln\left(\frac{102}{104.25}\right) + 131 \ln\left(\frac{31}{34.75}\right)\right) = 0.618$$

$$\text{Pearson's chi-square statistic: } X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{2.75}{312.75} + \frac{3.75}{104.25} + \frac{2.25}{104.25} + \frac{3.75}{34.75} = 0.604$$

יש 4 דברים שנספרים ויחס אחד ביניהם (הסכום צריך להיות 556), אז יש 3 דרגות חופש. אז תחת השערת האפס, G יתפלג $\chi^2(3)$. הערך p המתקבל הוא 0.892. כלומר נדחה את השערת האפס בכל רמה סבירה של מובהקות.

מבחן כי-בריבוע להומוגניות (*Chi-square test for homogeneity*)

מבחן ש

שימוש: בדיקה האם m קבוצות בת"ל של דאטא דיסקרטית מגיעות כולן מאותה התפלגות.

תוצאות: התוצאות האפשריות הן $\omega_1, \omega_2, \ldots, \omega_n$. זהות לכל קבוצה של דאטא.

דאטא: נניח שיש m קבוצות בת"ל שנותנות ספירה לכל אחת מהתוצאות האפשריות. כלומר, לקבוצה i יש ספירה נצפית $O_{i,j}$ לכל תוצאה אפשרית ω_j.

הנחות: אין.

השערת האפס H_0: כל קבוצת דאטא מגיעה מאותה התפלגות (לא מציינים מהי).

סטטיסטי המבחן: נראה את הדוגמה למטה. יש $m \times n$ תאים, עם הספירה של כל תוצאה לכל קבוצת דאטא. עם התפלגו האפס אפשר להעריך את הספירה המצופה לכל קבוצת דאטא. הסטטיסטיים G, X^2 מחושבים כמו לעיל.

דרגות חופש: $(m-1)(n-1)$. (בדוגמה למטה).

התפלגות האפס: $\chi^2(df)$. ערכי p מחושבים כמו במבחן כי-בריבוע לבדיקת טיב התאמה.

דוגמה – שייקספיר

נבדוק אם מחזה שנמצא הוא פרי יצירתו של שייקספיר. ניקח 12 עמודים אקראיים מתוך *King Lear*, ונספור מילים נפוצות. ונעשה אותו דבר ל"מחזה האבוד" שנמצא:

Word	a	an	this	that
King Lear	150	30	30	90
Long lost work	90	20	10	80

עם רמת מובהקות 0.1, נבדוק האם זה אכן מחזה של שייקספיר. השערת האפס היא שיש למילים שבדקנו תדירות דומה בשתי היצירות. הספירה הכוללת היא 500, אז ה-MLE בהנחת H_0 היא הספירה של כל מילה חלקי הספירה הכוללת:

Word	a	an	this	that	Total count
King Lear	150	30	30	90	300
Long lost work	90	20	10	80	200
totals	240	50	40	170	500
rel. frequencies under H_0	240/500	50/500	40/500	170/500	500/500

הספירה המצופה של כל מילה היא הספירה הכוללת של המילה כפול התדירות היחסית:

Word	a	an	this	that	Totals
King Lear	(150, 144)	(30, 30)	(30, 24)	(90, 102)	(300, 300)
Long lost work	(90, 96)	(20, 20)	(10, 16)	(80, 68)	(200, 200)
Totals	(249, 240)	(50, 50)	(40, 40)	(170, 170)	(500, 500)

$$\text{Pearson's chi-square statistic: } X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{6^2}{144} + \frac{0^2}{30} + \frac{6^2}{24} + \frac{12^2}{102} + \frac{6^2}{96} + \frac{0^2}{20} + \frac{6^2}{16} + \frac{12^2}{68} \approx 7.9$$

יש 8 תאים וכל הספירות השוליות הן קבועות כי הן נצרכות בשביל לקבוע את הספירה המצופה. כדי להיות עקביים עם הסטטיסטיים האלה, אפשר לקבוע בחופשיות את הערכים של 3 תאים (נגיד הכחולים) ואז השאר נקבעים כדי שהספירות השוליות יצאו נכונות. כלומר $df = 3$. או שנוכל להשתמש בזה ש $df = (m-1)(n-1) = 3 \cdot 1 = 3$.

מתקיים ש $p = 0.048$, אז נדחה את השערת האפס. אם נניח שלכל היצירות של שייקספיר יש את אותה תדירות של מילים, אז נסיק שהיצירה האבודה שנמצאה היא לא של שייקספיר.

עוד מבחנים

יש יותר מדי. נראה חלק מהם בשיעורים. העיקר הוא להבין את הפרדיגמה של *NHST* ואת החשיבות של בחירת סטטיסטי מבחן עם התפלגות האפס ידועה.

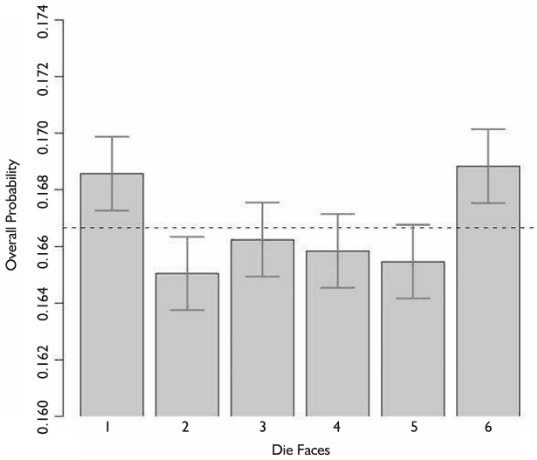
הרצאה

מצגת שיעור 9א

הקוביות של וולדון ולבי (*Weldon's and Labby's dice*)

26,306 זריקות קוביה. התוצאות:

Outcome	Observed	Expected
1	53,222	52,612
2	52,118	52,612
3	52,465	52,612
4	52,338	52,612
5	52,244	52,612
6	53,285	52,612
Total	315,672	315,672



השערת האפס: לכל מספר בקוביה יש את אותה ההסתברות. ההשערה האלטרנטיבית: ההסתברות לא שווה.

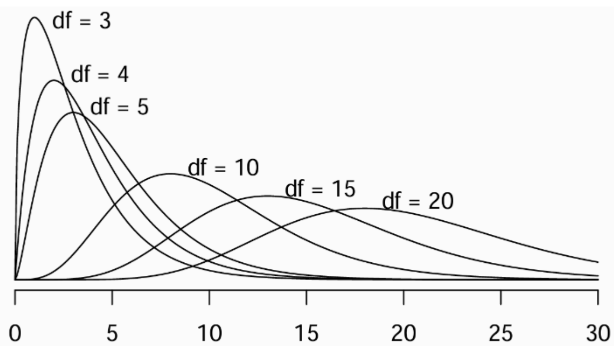
נבדוק עם מבחן כי-בריבוע:

$$\text{Pearson's chi-square statistic: } X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

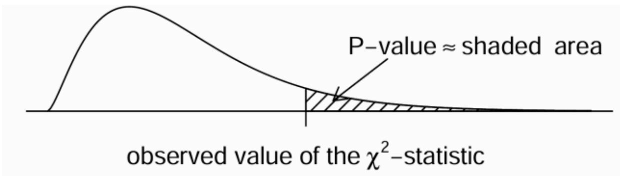
ככל שהתדירויות הנצפות שונות מהתדירויות המצופות, χ^2 יהיה גדול יותר, וזה עדות חזקה יותר נגד H_0.

Outcome	Observed	Expected	$\frac{(O-E)^2}{E}$
1	53,222	52,612	$\frac{(53{,}222-52{,}612)^2}{52{,}612} = 7.07$
2	52,118	52,612	$\frac{(52{,}118-52{,}612)^2}{52{,}612} = 4.64$
3	52,465	52,612	$\frac{(52{,}465-52{,}612)^2}{52{,}612} = 0.41$
4	52,338	52,612	$\frac{(52{,}338-52{,}612)^2}{52{,}612} = 1.43$
5	52,244	52,612	$\frac{(52{,}244-52{,}612)^2}{52{,}612} = 2.57$
6	53,285	52,612	$\frac{(53{,}285-52{,}612)^2}{52{,}612} = 8.61$
Total	315,672	315,672	24.73

התפלגות המדגם של χ^2 עבור דרגות חופש שונות:



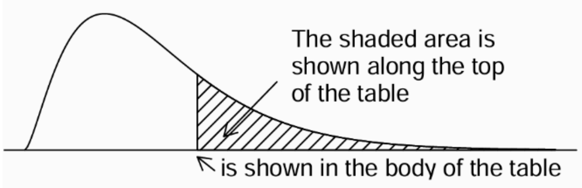
אם המדגם גדול, אז ל-χ^2 יש התפלגות בערך χ^2 עם $k-1$ דרגות חופש, כאשר k הוא מספר האיברים בסכום של χ^2. ה-*p-value* הוא השטח של הזנב מתחת הגרף של χ^2.



עבור הניסוי שלנו, יש 5 דרגות חופש.

כלל אצבע עבור גודל המדגם: כל הספירות צריכות להיות $5 \le$.

טבלת ההסתברויות:



Upper tail	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df 1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52
6	7.23	8.56	10.64	12.59	15.03	16.81	18.55	22.46
7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32
8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88
10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59
11	12.90	14.63	17.28	19.68	22.62	24.72	26.76	31.26
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

דוגמאות

אם $\chi^2 = 10.3$ עם $df = 6$:

$$p = P\left(\chi^2_{df=6} > 10.3\right)$$

Upper tail	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df 1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52
6	7.23	8.56	10.64	12.59	15.03	16.81	18.55	22.46
7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32

אם $\chi^2 = 17.56$, עם $df = 9$:

$$P\left(\chi^2_{df=9} > 17.56\right)$$

Upper tail	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df 7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32
8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88
10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59

בניסוי שלנו, יש $\chi^2 = 24.67$ עם 5 דרגות חופש:

עם $p < 0.001$, נדחה את השערת האפס. נאמר שאין הסתברות שווה לכל מספר.

Upper tail	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001	→
df 1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83	
2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82	
3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27	
4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47	
5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52	→

תזכורת: מבחן כי-בריבוע לבדיקת התאמה

אם יש לנו השערה למודל של התפלגות של דאטא שמתפלגת לפי קטגוריות:

Category	1	2	⋯	k
Probability	p_1	p_2	⋯	p_k

ויש לנו דאטא:

Observed Counts	O_1	O_2	⋯	O_k

ואנחנו רוצים לבדוק האם הדאטא מתאימה להתפלגות.

השערת האפס H_0: הנקודות בדאטא הן מ"מ בת"ל מההתפלגות.

ההשערה האלטרנטיבית H_A: הנקודות בדאטא הן **לא** מ"מ בת"ל מההתפלגות.

אם H_0 נכונה, אז הספירה המצופה (*expected counts*) של הקטגוריה ה-i היא $n \cdot p_i$, כאשר n הוא מספר הנקודות:

category	1	2	⋯	k	Total
observed counts	O_1	O_2	⋯	O_k	n
expected counts	np_1	np_2	⋯	np_k	n

זה לא בהכרח מספר שלם!

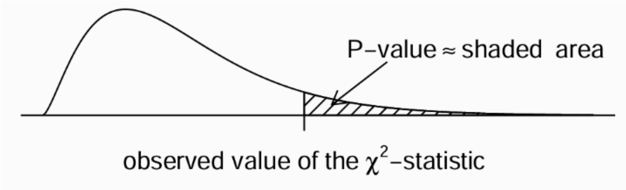
הסטטיסטי χ^2 הוא:

$$\chi^2 = \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}} = \sum_i \frac{(O_i - n \cdot p_i)^2}{n \cdot p_i}$$

דרגות החופש: $k - 1$ (מספר הקטגוריות פחות 1). למה? כי יש k פרמטרים עם הדרישה: $p_1 + p_2 + \cdots + p_k = 1$. כלומר $k - 1$ מהם חופשיים, והאחרון צריך להשלים ל-1.

ה *p-value* הוא בערך השטח בזנב הימני של ה χ^2 עם $k-1$ דרגות חופש:

השיערוך *chi-square approximation* עובד היטב כאשר כל הספירות המצופות הן לפחות 5.



דוגמה: המודל הגנטי של מנדל

כדי לבדוק עמידות של זנים של אורז למזיקים. בדקו 374 שורות, והתוצאות:

	Number of lines
All plants resistant	97
Mixed: some plants resistant, some susceptible	184
All plants susceptible	93

לפי מודל *IRRI*, השורות בת"ל. לכל שורה יש הסתברות 25% להיות עמידה, 50% להיות מעורבת, ו-25% להיות לא עמידה. נבדוק:

Type	Model	observed count	expected count	(Obs-Exp)^2/Obs
Resistent	25%	97	$374 \times 0.25 = 93.5$	$\frac{(97-93.5)^2}{93.5} \approx 0.1310$
Mixed	50%	184	$374 \times 0.50 = 187$	$\frac{(184-187)^2}{187} \approx 0.0481$
Susceptible	25%	93	$374 \times 0.25 = 93.5$	$\frac{(93-93.5)^2}{93.5} \approx 0.0027$
Total	100%	374	374	χ^2-statistics $= 0.1818$

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82

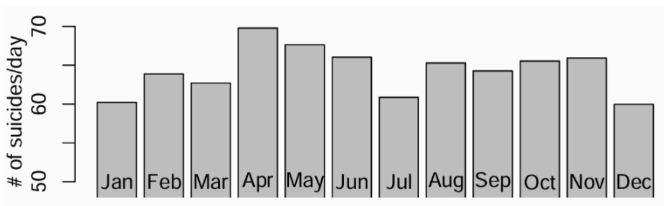
דרגות חופש: $3-1=2$. הסטטיסטי יצא 0.18, שזה הרבה יותר קטן מ-2.41 (שזה הערך שנותן *p-value* 0.3). כלומר ה *p-value* שלנו גדול יותר, אז לא נדחה את השערת האפס – כלומר המודל מתאים לדאטא.

דוגמה: שינויים באחוזי התאבדות

אם האחוזים קבועים מיום ליום, אז הספירה המצופה:

Suicide Counts in US by month in 1970

Month	# of suicides	days/ month	expected counts
Jan	1867	31	2021.889
Feb	1789	28	1826.222
Mar	1944	31	2021.889
Apr	2094	30	1956.667
May	2097	31	2021.889
Jun	1981	30	1956.667
July	1887	31	2021.889
Aug	2024	31	2021.889
Sept	1928	30	1956.667
Oct	2032	31	2021.889
Nov	1978	30	1956.667
Dec	1859	31	2021.889
Total	23480	365	23480



סטטיסטי המבחן χ^2:

$$\frac{(1867-2021.889)^2}{2021.889}+\frac{(1789-1826.222)^2}{1826.222}+\cdots+\frac{(1859-2021.889)^2}{2021.889}=$$
$$=51.18$$

עם $11 = 1 - 12$ דרגות חופש, ה *p-value* פחות מ-0.001. כלומר יש שינוי לפי חודש.

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
	4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
	5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52
	6	7.23	8.56	10.64	12.59	15.03	16.81	18.55	22.46
	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32
	8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
	9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88
	10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59
	11	12.90	14.63	17.28	19.68	22.62	24.72	26.76	31.26
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

מקרה פרטי: 2 קטגוריות

אם יש רק 2 קטגוריות, המבחן χ^2 הוא כמו מבחן דו-צדדי במדגם יחיד של $H_0: p = p_0$.

Category	Success	Failure	Total
probability under H_0	p_0	$1-p_0$	1
observed counts	X	$n-X$	n
expected counts	np_0	$n(1-p_0)$	n

$$\chi^2=\sum_i \frac{(O-E)^2}{E}=\frac{(X-np_0)^2}{np_0}+\frac{\left(n-X-n(1-p_0)\right)^2}{n(1-p_0)}=\frac{(X-np_0)^2}{np_0}+\frac{(X-np_0)^2}{n(1-p_0)}$$
$$=\frac{(X-np_0)^2}{n}\left(\frac{1}{p_0}+\frac{1}{1-p_0}\right)=\frac{(X-np_0)^2}{n}\left(\frac{p_0+(1-p_0)}{p_0(1-p_0)}\right)=\frac{(X-np_0)^2}{np_0(1-p_0)}=\left(\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}\right)^2$$

כאשר $\hat{p}=X/n$. כלומר, זה פשוט הריבוע של סטטיסטי z. בנוסף, ההתפלגות χ^2 עם דרגת חופש 1 היא הריבוע של $N(0,1)$. אז שני המבחנים נותנים את אותו ערך p.

תרגול

מבחן $t-test$.

נניח שיש דאטא $x_1,x_2,\dots,x_n$ עם ממוצע ושונות לא ידועים. נרצה למצוא את התוחלת μ. סטטיסטי המבחן הוא:

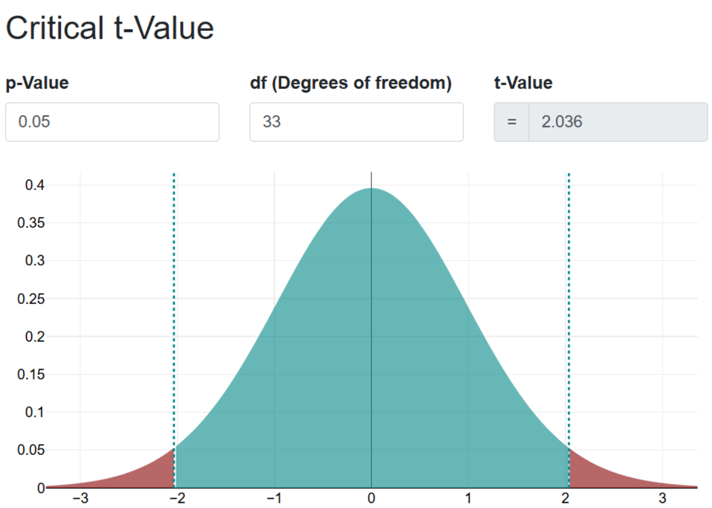
$$t=\frac{\bar{x}-\mu_0}{S_x/\sqrt{n}},\qquad S_x=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

נגיד שיש לנו 34 נשים עם אבא גבוה (מעל 182). נתון $\bar{x}=169.21$, $S_x=6.764$. הממוצע בארץ הוא 162.22.

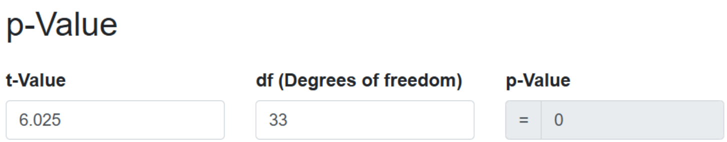
עם $\alpha=0.05$, נרצה לבדוק האם יש השפעה לכך שיש להן אבא גבוה – כלומר לבדוק אם הן לפי ההתפלגות הרגילה. אז:

$$t=\frac{169.21-162.22}{6.764/\sqrt{34}}=6.025$$

הוא מתפלג $t(33)$. אנחנו צריכים לקבוע את תחום הדחייה:



ובגלל שהוא בתחום הדחייה, נדחה את השערת האפס. נבדוק עד כמה העדות חזקה:



(כתוב 0 אבל זה פשוט ממש קטן). כלומר העדות חזקה.

מבחן כי בריבוע:

בהתפלגות דיסקרטית. הדאטא הוא ניסוי שקורה הרבה פעמים, ויש הסתברות כלשהי לכל תוצאה אפשרית. נספור את השכיחות של כל אחד. סטטיסטי המבחן הוא:

$$\chi^2 = \sum_e \frac{(f_0 - f_e)^2}{f_e}$$

לדוגמה, ימי הולדת: נגיד שיש 152 אנשים, אם נניח שההתפלגות שווה בין חודשים, אנחנו מצפים (נניח שיש לכל חודש אותו מספר ימים) שבכל חודש יהיו $152/12 = 12.6667$ ימי הולדת. נבדוק אם הדאטא הבאה מתאימה להתפלגות אחידה:

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
18	*8*	*9*	*11*	*1*	*18*	*15*	*18*	*11*	*16*	*7*	*1*

אז לדוגמה, האיבר הראשון בסכום:

$$\frac{(18 - 12.6667)^2}{12.6667}$$

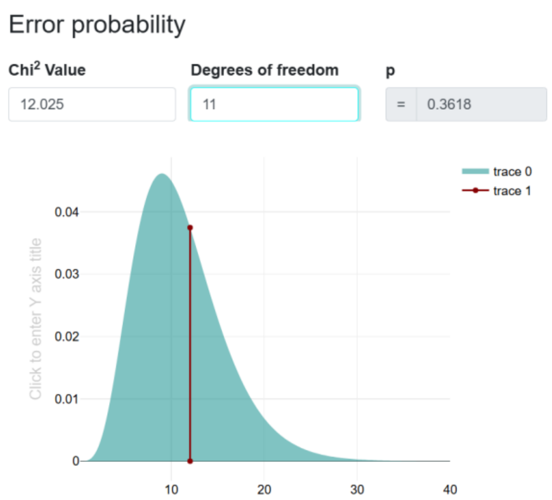
וזה מתפלג:

$$\chi^2 = 12.025 \sim \chi^2(11)$$

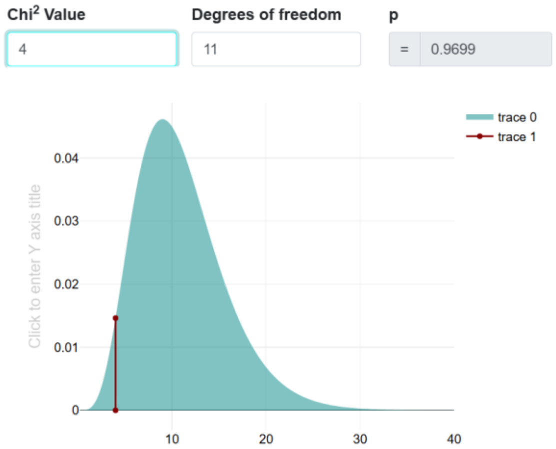
נבדוק בטבלה, עם $\alpha = 0.05$:

Significance level Alpha	**0.995**	**0.975**	**0.2**	**0.1**	**0.05**	**0.025**	**0.02**	**0.01**
Degrees of freedom								
1	0	0.001	1.642	2.706	3.841	5.024	5.412	6.635
2	0.01	0.051	3.219	4.605	5.991	7.378	7.824	9.21
3	0.072	0.216	4.642	6.251	7.815	9.348	9.837	11.345
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277
5	0.412	0.831	7.289	9.236	11.07	12.833	13.388	15.086
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812
7	0.989	1.69	9.803	12.017	14.067	16.013	16.622	18.475
8	1.344	2.18	11.03	13.362	15.507	17.535	18.168	20.09
9	1.735	2.7	12.242	14.684	16.919	19.023	19.679	21.666
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209
11	2.603	3.816	14.631	17.275	19.675	21.92	22.618	24.725

ונבדוק את החוזק של העדות:



כלומר העדות די חלשה אבל לא חלשה מאוד. למה? כי נגיד, כדי להגיע ל-*p-value* קרוב ל-1:



יש "מרחק" גדול מאד (כל ההסתברות באמצע).