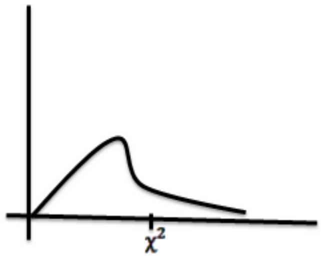
ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

הרצאה 4-5.8.12



התפלגות χ^2, (חי בריבוע) היא התפלגות אי סימטרית ימנית (חיובית) בעלת ערכים חיוביים בלבד. התפלגות זו תלויה בגודל המדגם. השימוש בטבלת χ^2 היא כמו בטבלת t.

<u>χ^2 לטיב התאמה:</u>

במבחן זה נרצה לבדוק האם תופעה מסוימת מתפלגת כצפוי. יש טיב התאמה H_0 , או אין טיב התאמה H_1 .

כלל ההכרעה במבחן זה: $\chi^2=\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}>\chi^2_{k-1,1-\alpha}$ אזי דוחים את H_0. [K-מספר הקטגוריה], [O-מה שנצפה בפועל (observe) E זה המצופה (expected).]

לדוגמה:

על ידי הטלת קובייה 120 פעמים התקבלו התוצאות הבאות:

תוצאה	1	2	3	4	5	6	סה"כ
O_i	18	15	22	25	17	23	120
E_i	20	20	20	20	20	20	120

א) בדקו ברמת מובהקות של 5% האם הקובייה הוגנת, סימטרית, וכן נסחו השערות.

ב) מצאו את α מינימאלית עבורה נוכל לדחות את H_0.

פתרון:

א) השערת הבסיס, H_0 היא שהקובייה סימטרית, וH_1 הוא שאין טיב התאמה (קובייה לא סימטרית). $\chi^2_{5,0.95}=11.07$, זה של הטבלה, ועכשיו למחושב: (ע"פ הנוסחא) $\chi^2=3.8$. קיבלנו כי המחושב קטן מהערך של הטבלה, ולכן לא דוחים את H_0.

ב) נחפש ערך χ^2 של טבלה שיהיה קטן מהערך המחושב: 3.8. ז"א, $\chi^2<3.8$ טבלה. נביט בטבלה ונקבל כי $0.5\leq 1-p.v.\leq 0.25$, ז"א ש $0.5\leq p.v.\leq 0.75$, ז"א שלכל α גדולה מ75% דוחים את H_0.

<u>מבחן χ^2 לאי תלות (אי קשר):</u>

מבחן זה בודק האם קשר (תלות) בין שני משתנים איכותיים.

$$E_{ij}=\frac{\text{סהכ שורה X סהכ עמודה}}{\text{סהכ כללי}}$$

השערת הבסיס היא שאין קשר (H_0) , וH_1 היא שיש קשר.

כלל ההכרעה הוא: אם $\chi^2=\sum_{ixj}^{RxC}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}>\chi^2_{(R-1)(C-1),1-\alpha}$ כאשר R מספר השורות C מספר העמודות, אזי דוחים את השערת הבסיס H_0.

דוגמה: רוצים לבדוק אם קיים קשר/תלות בין מגדר (זכר/נקבה) לבין צבע שיער (שחור/חום/בלונד) לצורך כך נלקח מדגם של 200 והתקבלו התוצאות הבאות:

E מגדר/ O צבע	שחור	חום	בלונד	סה"כ
זכר	50	40	30	120
נקבה	30	30	20	80
סה"כ	80	70	50	200

א) כתבו השערות ובדקו ברמה של 5%.

ב) מצאו את $P.V.$

פתרון:

א) H_0: אין תלות, ו- H_1: יש תלות בין המגדר לצבע השיער. $\chi^2_{(2-1)(3-1),95\%}=5.991$, ועכשיו נמצא את המחושב: $\chi^2=\sum_{ixj}^{RxC}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}=\frac{(50-48)^2}{48}+\dots+\frac{(30-32)^2}{32}=0.466$, $\chi^2_{\text{מחושב}}<\chi^2_{\text{טבלה}}$ ולכן לא דוחים את H_0.

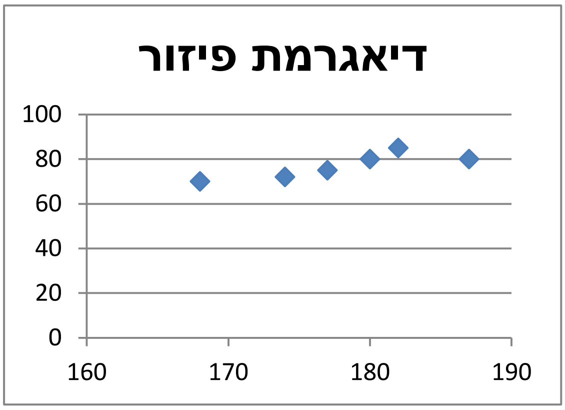
ב) $\chi^2_{\text{מחושב}}=0.466>\chi^2_{\text{טבלה}}$, לכן $0.1\le 1-p.v.\le 0.25$, ומכאן ש $0.75\le p.v.\le 0.9$, לכן עבור כל α שגדולה מ90% נדחה את השערת הבסיס ונסיק כי יש תלות בין המשתנים.

מתאם רגרסיה:

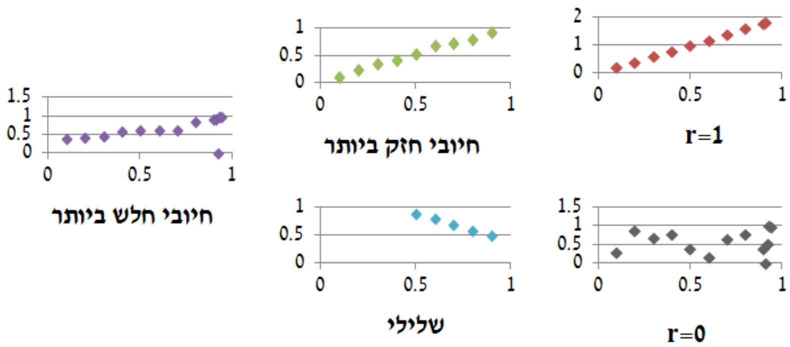
נרצה לבדוק האם קיים קשר/תלות ליניארי בין 2 משתנים כמותיים. ישנם שני סוגי קשרים: חיובי ושלילי.
בקשר *חיובי*: ככל שערך משתנה 1 *עולה* כך *עולה* גם השני.
בקשר *שלילי*: ככל שערך משתנה אחד *עולה* כך *יורד* השני.

דוגמה:

	1	2	3	4	5	6
גובה	168	174	177	180	182	187
משקל	70	72	75	80	85	80



מקדם המתאם של פירסון: $-1\le r_{xy}\le 1$.
כאשר *r=1* קשר חיובי מושלם.
r=-1 קשר שלילי מושלם.
$0.7\le|r|<1$ קשר חזק.
$0.4\le|r|<0.7$ קשר בינוני.
$0<|r|<0.4$ קשר חלש.
r=0 אין קשר.
(כאשר אין קשר יכול להתקבל גם קו אופקי מקביל לציר איקס, הציר התחתון)



חישוב r_{xy}: $r_{xy}=\frac{cov(x,y)}{S_x\cdot S_y}$ כאשר $cov(x,y)=\frac{\sum_{i=1}^{n}X_i\cdot Y_i}{n}-\bar{X}\cdot\bar{Y}$, זה יכול להיות גם שלילי וגם חיובי...
ואת $S_x\cdot S_y$ נגדיר ע״י: $S_y=\frac{\sum_{i=1}^{n}Y_i^2}{n}-\bar{Y}^2$, $S_x=\frac{\sum_{i=1}^{n}X_i^2}{n}-\bar{X}^2$.

דוגמה:

הגובה (x) הממוצע של המדגם מהשאלה הקודמת הוא 178 ס״מ, המשקל (y) הממוצע הוא 77 קילוגרם, ו$S_x=6.027$, $S_y=5.164$ מצא איזה סוג קשר זה. Cov(X,Y)=26.166 וע״פ נוסחא, $r_{xy}=0.84$ ולכן קשר חיובי חזק.

רגרסיה:

אם קיים קשר בין שני משתנים נרצה למצוא את הקשר ע״י משוואת רגרסיה פשוטה ליניארית מהצורה

$\hat{y} = bx + a$ניבוי y ע"י x או $\hat{x} = b'y + a'$, ניבוי x ע"י y.

חישוב a,b: עבור $\hat{y} = bx + a$, $b = \frac{cov(x,y)}{S_x^2} = r\frac{S_y}{S_x}$, ואת a מחשבים ע"י הצבה של הממוצעים $a = \bar{y} - b\bar{x}$. ועבור $\hat{x} = b'y + a'$, נגדיר $b' = \frac{cov(x,y)}{S_y^2} = r\frac{S_y}{S_x}$ ושוב את a' ע"י הצבה של הממוצעים.

דוגמה:

המשך גובה ומשקל (דוגמה קודמת). על סמך הנתונים הקודמים:

א) מצאו את משוואת הרגרסיה לניבוי משקל *y* לפי גובה.

ב) מה תהיה הערכתם למשקל של אדם שגובהו 181 ס"מ?

פתרון:

א) $b = \frac{26.166}{6.027^2} = 0.72$ נמצא את *a* ע"פ הנוסחה ונקבל *a=-51.16*, ולכן משוואת הרגרסיה לניבוי הגובה היא: $\hat{y} = 0.72x - 51.16$.

ב) נציב במקום *x* את הגובה 181 ונקבל משקל *79.21* קילוגרם.

הסתברות:

ניתן להגדיר את המושג הסתברות של מאורע באמצעותו שכיחותו היחסית. נתאר לנו ניסוי, שמרחב המדגם שלו הוא S. נניח שמבצעים ניסוי זה שוב ושוב באותם תנאים בדיוק. לכל מאורע E במרחב המדגם S נסמן את השכיחות של מאורע E בV(E) דהיינו מספר הפעמים שהמאורע E מתרחש בn החזרות הראשונות על הניסוי. ואז נגדיר את ההסתברות של המאורע E ע"י $p(E) = \lim_{n\to\infty}\frac{n(E)}{n}$. ז"א ההסתברות p(E) מוגדרת כגבול של המספר היחסי של הפעמים שבהן E מתרחש.(p(E היא השכיחות היחסית הגבולית של E.

p(E) מקיים את 3 האקסיומות הבאות:

א) $0 \le p(E) \le 1$

ב) p(S)=1

ג) לכל סדרה של מאורעות זרים $E_1, E_2, ..$ כלומר מאורעות המקיימים $E_i \cap E_j = \phi$ לכל i≠j מתקיים $p(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} E_i$. (לכל סדרת מאורעות זרים, ההסתברות שלפחות אחד מהם יתרחש שווה לסכום ההסתברות של המאורעות המהווים את הסדרה)

<u>הסתברות מותנית ואי תלות</u>:

הגדרה: ההסתברות המותנית של מאורע E בתנאי/בהנתן מאורע F מסומנת p(E/F) אם p(F)>0 אזי ניתן לחשב זאת ע"י הנוסחה $p(E/F) = \frac{p(E\cap F)}{p(F)}$

דוגמה:

מטילים שני קוביות, אדומה וכחולה. נניח שלכל אחת מ36 האפשרויות סיכויים שווים להתרחש, כלומר ההסתברות של כל תוצאה היא *1/36*. נניח שראינו שסכום מה שהתקבל בקוביות שווה לשמונה. מה הסיכוי שהקובייה האדומה הראתה שלוש?

פתרון:

E= סכום של 8 ב2 הקוביות

F= תוצאת הקובייה האדומה היא 3. $p(F) = \{(6,2),(5,3),(4,4),(3,5),(2,6),(4,4)\}$.

נציב בהתאם: $\frac{1}{6} = \frac{1/36}{6/36} = \frac{p(E\cap F)}{p(F)} = p(E/F)$.

טענה: $p(E \cap F) = p(E/F) \cdot p(F)$.

דוגמה: קרן מהססת אם לבחור קורס בצרפתית או בכימיה. היא מעריכה שתקבל 10 בצרפתית בהסתברות של 0.5 ותקבל 10 בכימיה בהסתברות של $\frac{2}{3}$. אם קרן מחליטה לבחור בין הקורסים ע"י הטלת מטבע, מה ההסתברות שתקבל 10 בכימיה?

פתרון: ע"פ הטענה נקבל כי $p(E\cap F)=p(E/F)\cdot p(F)$ כאשר E זה לקבל 10 וF לבחור בכימיה. לכן ההסתברות היא מכפלה של $\frac{1}{2}\cdot\frac{2}{3}=\frac{1}{3}$.

נרחיב את הטענה לדבר הבא: לחישוב הסתברות של חיתוך מספר כלשהו של מאורעות:
נוסחת הכפל: $p(E_1\cap\ldots\cap E_n)=p(E_1)\cdot p(E_2/E_1)\cdots p(E_n\ /\ E_1\cap\cdots\cap E_{n-1})$. ונמחיש זאת בדוגמה:

דוגמה:
מחלקים באקראי חפיסה של 52 קלפים ל4 ערמות, 13 בכל אחת. חשב את ההסתברות שבכל ערימה יש אס אחד בדיוק.
פתרון:
נגדיר $E_1=$ אס עלה נמצא באחד הערמות. $E_2=$ אב עלה ואס לב בערמות שונות. $E_3=$ אס עלה, לב ומעוין בערמות שונות. $E_4=$ ארבעת האסים בערמות שונות.
ע"פ הנוסחא: $p(E_1\cap\ldots\cap E_4)=p(E_1)\cdot p(E_2/E_1)\cdot p(E_3/E_1\cap E_2)\cdot p(E_4\ /\ E_1\cap E_2\cap E_3)$.
נציב ונקבל כי $p(E_1\cap\ldots\cap E_4)=1\cdot\frac{39}{51}\cdot\frac{26}{50}\cdot\frac{13}{49}=0.1054$.

<u>נוסחת בייס (Bayes):</u>
יהיו E וF מאורעות. אנו יכולים לבטא את המאורע E ע"י $E=(E\cap F)\cup(E\cap F^C)$. המאורעות זרים ולכן מאקסיומה שלוש ומהטענה הקודמת נקבל כי: $p(E)=p(E\cap F)+p(E\cap F^C)$
$p(E)=p(E/F)\cdot p(F)+p(E/F^C)\cdot p(F^c)$ ולכן $=p(E/F)\cdot p(F)+p(E/F^C)\cdot[1-p(F)]$.

דוגמה:
חברת ביטוח מאמינה שניתן לחלק בני אדם לקבוצות: מועדים לתאונות ושאינם. מהסטטיסטיקה של חברת הביטוח עולה שבתוך תקופה של שנה עבור קבוצה 1 קורות תאונות בהסתברות 0.4 וקבוצה 2 בהסתברות 0.2. אם נניח ש30% מהאוכלוסייה מועדים לתאונות, מה ההסתברות שלמבוטח חדש תקרה תאונה בשנה הראשונה?
פתרון:
נסמן, E למבוטח תקרה תאונה בשנה הראשונה. F מסמל את המאורע "לקוח מועד לתאונות". נציב בנוסחא ונקבל כי $p(E)=0.2\cdot 0.7+0.3\cdot 0.4=0.26$ זוהי ההסתברות המבוקשת.

המשך:
מה ההסתברות שהוא היה מועד לתאונות אם הוא (המבוטח) עבר תאונה בשנה הראשונה?
פתרון:
ע"פ נוסחת ההסתברות המותנית נקבל כי $p(F/E)=\frac{p(E\cap F)}{p(E)}=\frac{0.3\cdot 0.4}{0.26}=0.461$

הגדרה: יחס סיכויים של מאורע A מוגדר ע"י $\frac{p(A)}{1-p(A)}=\frac{p(A)}{p(A^C)}$, בהינתן עובדה חדשה B נקבל את הנוסחא $p(A/B)=\frac{p(B/A)\cdot p(A)}{p(B)}$ ואז $p(A^C/B)=\frac{p(B/A^C)\cdot p(A^C)}{p(B)}$ ואז מספר הסיכויים החדש הוא:
$\frac{p(A/B)}{p(A^C/B)}=\frac{p(A)}{p(A^C)}\cdot\frac{p(B/A)}{p(B/A^C)}$. ז"א מתקבל יחס סיכויים ישן כפול יחס סיכויים מותנית.

הכללה של הטענה הקודמת:
יהיו $F_1,\ldots,F_n$ מאורעות זרים במרחב מדגם S המקיימים $S=\bigcup_{i=1}^n F_i$, נסמן בE מאורע כלשהו במרחב המדגם: $E=\bigcup_{i=1}^n E\cap F_i$, המאורעות $E\cap F$ זרות לכל i, נקבל:
$p(E)=\sum_{i=1}^n p(E\cap F_i)=\sum_{i=1}^n p(E/F_i)\cdot p(F_i)$. נניח כעת שE התרחש ואנו מעוניינים לקבוע איזה מהF-ים התרחש, אזי $p(F_j/E)=\frac{p(E\cap F_j)}{p(E)}=\frac{p(E/F_j)\cdot p(F_j)}{\sum_{i=1}^n p(E/F_i)\cdot p(F_i)}$.