סטטיסטיקה - תרגול 13

## מבחן חי בריבוע של פירסון לטיב התאמה

הוגרלו 250 מספרים התקבלו התוצאות הבאות

| | 0 | ... | | 9 | |
|---|---|---|---|---|---|
| obs | 17 | 31 | 29 | 36 | 250 |
| exp | 25 | 25 | ... | 25 | |

בגלל שזה אחידה אני מצפים

השערת האפס - התפ' אחידה $H_0: X_i \sim Unif\{0,\dots,9\}$

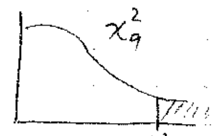
האלטרנטיבה - שלא. $H_1: \neg H_0$

הקורס הלך של חלמה משנה שעברה לא ישתנה

מבחנים לסטטיסטיקה

$$X^2=\sum\frac{o_i-e_i}{e_i}=\underbrace{\frac{(25-17)^2}{25}}_{2.56}+\dots+\dots=23.28$$

$$X^2\overset{H_0}{\sim}\chi^2_{10-1}$$



$\chi^2_{9,0.95}=19.92$

* פקודה ב-R ל-$\chi^2$: pchisq, qchisq

קיבלנו שאיזור הדחייה מתחיל בערך הקריטי 19.92 הערך שקיבלנו ב-$X^2$ היה 23.28.

נחשב את עוצמת המבחן:

$$\pi=P_{H_1}(X^2>3.84)=$$ במידה הנתונים לא ניתן לחשב

הסיבה שלא ניתן לחשב את $\pi$ היא כיוון שהאלטרנטיבה אינה מוגדרת באופן ספציפי ולכן אנו לא יודעים כיצד האלטרנטיבה צפויה להתפלג.

### מבחן חי בריבוע לאי תלות

במחקר לבדיקת קשר בין מגדר והרגלי עישון, התקבלו התוצאות

| | בת | בן | |
|---|---|---|---|
| מעשן | 10 | 20 | 30 |
| לא מעשן | 30 | 40 | 70 |
| | 40 | 60 | |

השערת האפס היא שאין קשר בין המשתנים:

$$H_0: \forall i,j \quad P(\text{מין}=i,\ \text{עישון}=j)=P(\text{מין}=i)\,P(\text{עישון}=j)$$

$$H_1: \neg H_0$$

הסיכוי שבן ומעשן הוא $\frac{60}{100}\cdot\frac{30}{100}$, אני צריך מספר ולא אחוזים ולכן נכפיל בטבלה (כמות זה נכפיל בה-100). נחשב שכיח ה קריטריון ונקבל:

| | בת | בן | |
|---|---|---|---|
| מעשן | 12 | 18 | 30 |
| לא מעשן | 28 | 42 | 70 |
| | 40 | 60 | |

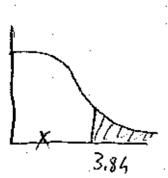
$\frac{(18-20)^2}{18}=\dots$

נחשב אומדן לסטטיסטי $X^2$:

$$X^2=\sum\frac{(e_i-o_i)^2}{e_i}=\dots=0.79$$

/* יחידת המידה כאן היא counts */

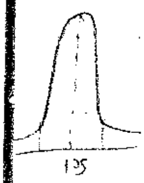
$$X^2\overset{H_0}{\sim}\chi^2_{(2-1)(2-1)}=\chi^2_1$$



נבחן את ההשערה ברמת בטחון $\alpha=0.025$ $\chi^2_{1,0.975}=3.84$, כלומר אנו לא דוחים את השערת ה-0.

תנאי טכני: $X^2 \overset{H_0}{\sim} \chi^2$. כלומר, ההתפלגות היא בקירוב. לכן, חייב להיות בכל תא בטבלה לפחות 5 (חייב להיות בכל תא ב-expected.
יש מקורות שטוענים שצריך לבדוק את שתי הטבלאות.

## הסקה ואמידה בייסיאנית




$f(y_i)$

נניח שהגבהים במדינת ישראל $y_i \sim N(\mu, \sigma_y^2)$. ננסה להעריך בתושבי המדינה שהגובה הטיפוסי $\mu \sim N(\mu_0=175, \sigma_\mu=3.12)$. כלומר, זו הערכה סובייקטיבית וזו הדרך למדל את אי הוודאות שלנו. 3.12 זה הגיוני תחת ההנחה ש-80%, והפרש האמיתי נמצא כ-4 פה לכאן ולכאן. כלומר בין 171-179 צריך להיות 80% מהמסה

$$171 = 175 - \sigma \cdot Z_{0.9} \Rightarrow \sigma \frac{4}{Z_{0.9}} \underset{\text{אמור להיות}}{\rightarrow} 3.12$$

הנחה זו נקראת prior. נסמן את התפלגות הפרמטר $\pi(\mu)$

אנו רוצים כעת לחשב את post. כעת לאחר ע"י דגימה ועדכון ההשערה שלנו.

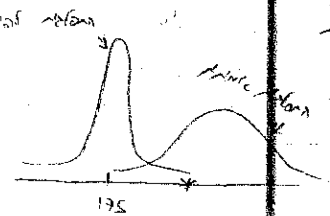
$$f(\mu|y^n) = \frac{f(\mu, y^n)}{f(y^n)} = \frac{f(y^n|\mu) f(\mu)}{f(y)} = const \cdot f(y^n|\mu) \cdot \pi(\mu) =$$

$$= c \cdot \left(\frac{1}{\sqrt{2\pi}\sigma_y}\right)^n \cdot \exp\left(-\frac{1}{2\sigma_y^2} \cdot \sum (y_i-\mu)^2\right) \cdot \left(\frac{1}{\sqrt{2\pi}\cdot\sigma_\mu}\right) \cdot \exp\left(-\frac{1}{2\sigma_\mu^2}(\mu-\mu_0)^2\right)$$

$$= c \cdot e^{-\frac{1}{2}\underbrace{\left(\frac{n}{\sigma_y^2}+\frac{1}{\sigma_\mu^2}\right)}_{\#}\left(\mu - \underbrace{\left(\frac{\sigma_\mu^2}{\sigma_y^2/n+\sigma_\mu^2}\bar{y} + \frac{\sigma_y^2/n}{\sigma_y^2/n+\sigma_\mu^2}\cdot\mu_0\right)}_{\text{כקומבינציה של שני התוחלות מהווה את הממוצע (נסמן כ-*)}}\right)^2}$$

מבנה כללי שהתפלגות גאוסיאנית היא $c \cdot e^{(\text{משהו})}$ ולכן.

מכאן, נקבל כי $\mu|y^n \sim N(*, \#^{-1})$




כתוצאה מתוספת הנתונים אי הודאות שלנו קטנה.

$$\left(\frac{1}{\sigma_\mu^2}\right)^{-1} > \left(\frac{1}{\sigma_\mu^2}+\frac{n}{\sigma_y^2}\right)^{-1}$$