

סטטיסטיקה – מטלה 2

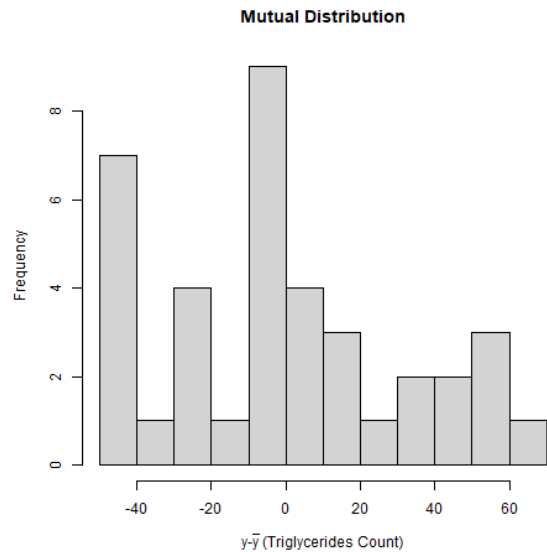
אלעזר פיין

1.

$H_0: \mu_{\text{CHD}} = \mu_{\text{control}}$

$H_1: \mu_{\text{CHD}} > \mu_{\text{control}}$

התפלגות נתונים משותפת:



מבחן לבדיקות שוויון שונות:

נבדוק בעזרת מבחן לווין:

h_levene = Table: 2 x 3 ...	View Table
> Df = 1 36	
> F value = 1.897481	NA
> Pr(>F) = 0.1768648	NA

$p > 0.05$ כלומר מתקיים שוויון שונות (אי אפשר לדחות את השערת האפס כי הן שוות).

מבחן לבדיקת נורמליות של ההתפלגות המשותפת:

נבדוק בעזרת מבחן שפירו-ווילק:

```
h_shapiro = List[1:4]
  statistic = W ... View
  p.value = 0.05276915
  method = "Shapiro-Wilk normality test"
  data.name = "df_flat$shifted"
```

$p.value > 0.05$ כלומר ההתפלגות המשותפת נורמלית (אי אפשר לדחות את השערת האפס כי היא נורמלית).

מבחן להשוואת התוחלות:

לפי עץ ההחלטה:

1. המדגמים unpaired
2. התפלגות משותפת מתפלגת נורמלית (בדקנו עם שפירו)
3. השוניות שוות (לפי לוויין)

לכן (התנאים ל-t-test מתקיימים) נבדוק באמצעות t-test:

```
h_ttest = List[1:10]
  statistic = t ... View
  parameter = df ... View
  p.value = 0.01233947
  > conf.int = 7.110335 Inf ... View
  > estimate = mean in group CHD mean in group control ... View
  null.value = difference in means ... View
  stderr = 10.8332
  alternative = "greater"
  method = "Two Sample t-test"
  data.name = "values by ind"
```

$p.value < 0.05$ כלומר דוחים את H_0 ומסיקים כי רמת השומנים של בעלי המחלת לב גבוהה מזו של אנשי קבוצת הביקורת.

2.

נבנה לוח שכיחות של רמות הדם:

	CHD	control
NORMAL LEVEL COUNT	10	18
ABNORMAL LEVEL COUNT	5	5

טבלה 2×2 עם גודל מדגמים קטן מ-50 ולא יותר מחמישית הטבלה ערכים קטנים (5) – עדיף fishertest מאשר chi-squared (יותר מדויק אך יותר קשה לחישוב, לכן טוב לכמות אינפורמציה קטנה)

נבדוק בעזרת מבחן פישר:

```
h_fisher = List[1:7]
p.value = 0.4725592
conf.int = 0.1009823 3.1118393 ... View
estimate = odds ratio ... View
null.value = odds ratio ... View
alternative = "two.sided"
method = "Fisher's Exact Test for Count Data"
data.name = "cont_table"
```

$p.value > 0.05$ כלומר לא ניתן לדחות את השערת האפס כי אין קשר בין המדגמים לרמת דם גבוהה/נורמלית וכך גם נסיק. (בניסוח שונה [של צ'י – סקוורד] המדגמים מתפלגים בצורה דומה ביחס לרמת דם גבוהה/נורמלית).

3.

א. בשאלה 1 קיבלנו כי רמת השומנים של חולי הלב גבוהה מזו של קבוצת הביקורת ובשאלה 2 קיבלנו כי אין תלות לבין הקבוצה לבין אם אדם ממנה בעלה רמת דם גבוהה/נורמלית. בשתייהן אנו מניחים כי אין הבדל בין המדגמים, בשאלה 1 אנו מקבלים במובהקות 95% כי **יש הבדל בתוחלות** ובשאלה 2 אנו מקבלים במובהקות 95% כי **אין הבדל בהתפלגויות** (ביחס למשתנה הקטגורי).

ב. במבט שטחי נראה כי המסקנות כביכול סותרות אך למעשה הם עונות על שאלות שונות לחלוטין, שאלה 1 שואלת על **התוחלות** של המדגמים ככלל ביחס למשתנה נומרי – רמת השומנים בדם. ושאלה 2 שואלת על **ההתפלגויות** של המדגמים ביחס למשתנה קטגורי – רמת שומנים **גבוהה/נמוכה**. לגבי אינפורמציה הדרושה (R): בשאלה 1 ל-ttest צריך רק את הערכים הנתונים (אחרי פעולות סידור) ולעומת זאת בשאלה 2 ל-fishertest צריך יותר אינפורמציה - לבנות לוח שכיחות בהתבסס על הערכים.