

תרגול 13

Pairwise Comparison of Means

תרגול זה עוסק בהמשך אנאליזת השונויות שביצענו בקובץ הנתונים "MultiWeightliftingScores.csv". בקובץ זה מופיעים ביצועי נבחרות סין, ארה"ב ורוסיה (הדימיוניים) בתחרות הרמת משקולות. עליכם למצוא בין אילו זוגות מדינות ביצועי הנבחרות (תוחלות ביצועי הנבחרות) שונים זה מזה ובאילו דומים זה לזה. לצורך כך:

1. העלו את הקובץ לטבלת נתונים ב-R (dataframe).
2. העלו מחדש את חישוב השונות התוך קבוצתית (MSE).
3. חשבו את האומדנים הסטטיסטיים של ההבדלים בתוחלות בין כל אחד מהזוגות חשבו את הסיכוי לדחיית השערת האפס (P_v) (לצורך כך הניחו - studentized t-distribution) נסחו מסקנתכם עבור כל אחת מהשוואות - האם התוחלות שוות או שונות?
4. ערכו את המבחן התואם לחישוב שביצעתם בשאלה 3 ב-R ובדקו את תשובותיכם.
5. ערכו מבחן השוואות מרובות מבוסס על t-distribution, עם תיקוני בונפרוני, הולמס, ובניימיני. השוו את התוצאות המתקבלות עם מבחן Tukey.

Correlation

תרגול זה עוסק בחישוב מתאם בין שני משתנים.

- לצורך כך נעשה שימוש בבסיס הנתונים "Ozone" מספריית mlbench בה עשינו שימוש במטלה 1.
1. העלו את הנתונים. צרו טבלת נתונים הכוללת רק את המשתנים "V4" = "ozone" ו-"V5" = "Pressure". הציגו סיכום (summary) של הנתונים להתרשמות כללית.
 2. הציגו איור פיזור נתוני האוזון כפונקציה של הלחץ שנמדדו. (plot פשוט יעשה את העבודה, לא לשכוח לכתוב כותרות מתאימות לכל אחד מהצירים)
 3. חשבו את וקטור r למציאת המתאם בין אוזון ללחץ, וממנו חשבו את מקדם המתאם הכללי, r . להזכירכם:

חשבו תחילה את ה-covariance של כל משתנה בנפרד:

$$Z_x = \frac{(x_i - \bar{X})}{S_x}$$

ומהם חשבו את וקטור המתאם:

$$r_i = Z_x \cdot Z_y$$

מהוקטור של r חשבו את מקדם המתאם:

$$r = \frac{1}{n-1} \sum_{i=1}^n r_i$$

4. נסחו השערה סטטיסטית להוכחת המתאם בין המשתנים.

חשבו את האומדן הסטטיסטי להשערה:

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$$

ובעזרתו חשבו את הסיכוי לדחיית השערת האפס בטעות (P_v) (קבלת מקדם מתאם משמעותי בטעות)

נסחו את מסקנתכם מהתוצאה.

5. בצעו מבחן למציאת הסיגניפיקנטיות הסטטיסטית של הקורלציה בין רמות האוזון ללחץ הנמדד בעזרת פונקציית R המתאימה (cor.test). וודאו שהחישובים שביצעתם ב-4 נכונים, ושהמסקנות תואמות לתוצאות המבחן הסגור.