

שיטות מחקר - תרגיל 9

מגישות - אלה דובדבן 305564866 זוהר בוחניק 311142293

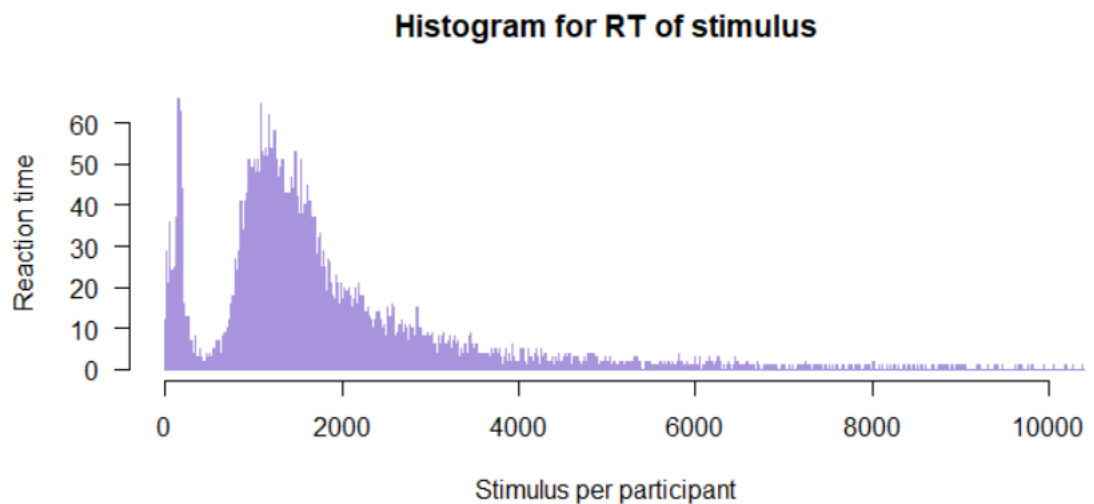
17 ביוני 2019

שאלה 2:

סעיף א'-ג'

ניתן לראות בקובץ *R*. ניקינו את ה-*data* כמו שנדרש.

סעיף ד'



מה שלדעתנו עשוי להיות בעייתי בנתונים הגולמיים הוא שיש מידדות בודדות אשר נמצאות במרחק גדול (של כמה סטיות תקן) מריכוז הדגימות שהתקבלו. דגימות קיצוניות כאלה עלולות להביא למסקנות שונות מהמצב בעולם ולגרום לטעויות מסוג ראשון או שני.

סעיף ה'

הוספנו את *RT_clean* ל-*data*.

סעיף ו'

קיבלנו כי- 9.63% מהשורות היו עם ערך קיצוני ולכן סוננו מהדאטא של זמני התגובה. לדעתנו חוקרים לעיתים מדווחים על האחוז שנתונים כאלו מהווים מכלל הדאטא כיוון שזה מעיד על כמה דגימות נחשבו "קיצוניות" וסוננו החוצה מהדאטא. במידה ואחוז זה הינו גבוה, ייתכן ומדובר בנתח מהאוכלוסייה, שאחרת לא יבוא לידי ביטוי. כמו כן, זה יכול להעלות ספקות לגבי התוקף הפנימי והחיצוני של המבחן אשר בוצע ועל אמיתות התוצאות.

סעיפים ח'-י"ד:

בקובץ *R*

שאלה 3:

סעיפים א'-ג':

בקובץ R

סעיף ד':

ביצענו מבחן קורולציה בין הציונים אשר קיבלו המשתתפים בשאלון ה-TPS המייצג את המבחן האקספליציטי, לבין הציון אשר חושב בשאלה הקודמת - qIAT, המייצג את המבחן האימפליציטי. התוצאות אשר קיבלנו:

```
Pearson's product-moment correlation  
  
data: data_qIAT$qIATscore_corrected and data_qIAT$TPS_score  
t = 2.8008, df = 44, p-value = 0.007546  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 0.1112454 0.6103576  
sample estimates:  
      cor  
0.3889815
```

המסקנות:

עפ"י הנחות המבחן - השערת האפס היא שלא קיימת קורולציה בין המבחנים. היות וקיבלנו כי $p_value = 0.0075$, ניתן לדחות את השערת האפס. כלומר, כן קיימת התאמה בין תוצאות המבחנים. בהתאמה, קיבלנו כי הקורולציה בין המבחנים היא 0.3889. התוקף אשר התוצאות מעידות עליו הוא תוקף הניבוי אשר נמדד ע"י המתאם בין המבחן האימפליציטי למבחן האקספליציטי.

סעיף ה':

סינון הנתונים אחרי קבלת התוצאות (Post hoc) יכול להביא ל"זריקה" של נתונים אשר נראים מסיחים (outliers) כי אולי לא תואמים את ההיפותזה שלנו, אך בעצם הם מייצגים את האוכלוסייה. סינון זה יכול להביא לשינוי המידע כך שיעיד על קיום של אפקט למרות שאינו קיים - כלומר לטעות מסוג ראשון. ברגע שאנחנו רואים את הנתונים אנחנו כבר מוסיפים ולא אובייקטיבים לקבל החלטות כאלה.

סעיף ו':

כפי שטענו בסעיף ו' בשאלה הקודמת, חשוב לבדוק מה אחוז הדגימות שסוננו החוצה מתוך כלל הדגימות. באופן כללי, במצב בו אחוז זה גבוה, אז נכון יותר להשאיר את הדגימות שכן יש חשד שהן מייצגות אוכלוסייה אשר לא תבוא לידי ביטוי אם נסנן את הדגימות. על מנת לתת תוקף חיצוני למבחן כך שנוכל להכליל אותו על האוכלוסייה הכללית, חשוב לתת ביטוי גם לאוכלוסייה זו. כמובן שכל זה תלוי במשתנה שממנו אנו מסננים תוצאות ובהפעלת הגיון סביר לתוצאות אפשריות עבור המשתנה הזה לעומת תוצאות שהן לא תקינות ומהוות רעש שיכול להסיט את המסקנות. במצב בו מדובר בדגימות בודדות שאכן לא נראה סביר כי הן התקבלו מתוך האוכלוסייה יהיה נכון לסנן אותן. נזכר כי ראינו בתרגול כי גם דגימה בודדת שהיא outliers יכולה לשנות את המסקנות.

עבור התוצאות בניסוי הספציפי אשר ביצענו, נטען כי מרקוס טועה, וכי זאת הייתה החלטה נכונה לסנן את הדגימות. זמן תגובה של 300 מ"ש, הנתון לפיו סיננו, הוא זמן מהיר מידי מכדי לעבד את השאלה ולענות תשובה כנה ונראה כי אותו נבדק לא באמת קרא את השאלה וענה ברצינות אלא לחץ בלי מחשבה. לכן תשובות אלו לדעתנו הן כן outliers, הן רעש שיכול להטות את המסקנות ונכון לסנן אותן החוצה.

סעיף ז':

הרצנו רגרסיה לינארית על המשתנה התלוי $diff_time$ עם המנבא TPS_score .

```
Call:
lm(formula = all_data$DIFF_time ~ all_data$TPS_score)

Residuals:
    Min       1Q   Median       3Q      Max
-7.9120 -2.6636  0.1372  3.0456  6.3862

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.58463    2.92454   3.619 0.000759 ***
all_data$TPS_score -0.05235    0.03716  -1.409 0.165993
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.22 on 44 degrees of freedom
Multiple R-squared:  0.04315,    Adjusted R-squared:  0.0214
F-statistic: 1.984 on 1 and 44 DF,  p-value: 0.166
```

ניתן לראות כי מבחן ה- F , אשר בודק האם הציונים המנובאים ע"י הרגרסיה הלינארית והציונים האמיתיים הם מאותה ההתפלגות. השערת האפס במקרה זה הינה שאין קשר לינארי בין המנבא לבין המשתנה התלוי, ולכן היות וקיבלנו $p_value = 0.166$, לא ניתן לדחות את השערת האפס, ולכן לא נוכל לטעון שום דבר לגבי הקשר בין המשתנים. יתכן כי הוא לא לינארי אבל קיים, או שהוא לא קיים בכלל. בשונה מסעיף ד', תוקף הניבוי שקיבלנו הוא נמוך. בפועל, התוצאות מעידות על כך שציון ה- TPS לא חוזה את מדד ה- $diff_time$.

סעיף ח':

משתנה מתערב לדוגמה הינו סוגיית התעדוף בין הקורסים השונים. ייתכן כי בימים אשר בין פרסום השאלון לבין המענה עליו, היו תאריכי יעד של מטלות הגשה נוספות אשר דרשו את השקעת הזמן. במקרה זה, ה"דחיינות" אשר הוצעה כפירוש של ה- $diff_time$ אינה נובעת מתכונת הדחיינות האיטית של הסטודנטים, אלא מניהול הזמנים שלהם. כלומר, יש לקחת בחשבון כמשתנה מתערב את העומס של הסטודנטים במהלך הזמן אשר בין פרסום השאלון לבין זמן המענה עליו. הצעה נוספת, המערערת על תוקף המשתנה, היא כי יתכן שאנשים יעידו על עצמם באופן טבעי כי הם פחות דחיינים ממה שהם בפועל, כמו באימרה ידועה - "אם היו שואלים כל אחד כמה הוא ביחס לממוצע, 95% היו אומרים שהם מעל לממוצע".

שאלה 4:

סעיף א':

נעלה את השאלה - האם יש הבדל בין גברים לנשים בתכונה- נועם הליכות. על מנת לבדוק את השאלה נבצע מבחן t דו זנבי עם $\alpha = 0.05$.

```
Welch Two Sample t-test

data:  women_A and men_A
t = 1.9802, df = 44.84, p-value = 0.05384
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.04146973  4.85373026
sample estimates:
mean of x mean of y
 67.51724  65.11111
```

לא ניתן היה לדחות את השערת האפס. קיבלנו $p_value = 0.05384$. זהו ערך גדול מ-0.05 ולכן לא דחינו את השערת האפס ולא ניתן להסיק כלום. במבט על הממוצעים ניתן לראות הבדלים קטנים מאוד בין הקבוצות. היו לנו 29 דגימות של נשים ו-18 דגימות של גברים. נראה כי גודל מדגם זה לא משקף את האוכלוסייה וכי לא ניתן להסיק ממנו אפקט על המשתנים.

סעיף ב':

multiple comparisons - הבעיה היא שאספנו מספר גדול של משתנים, וכאשר יש יותר מניתוח אחד על אותם משתנים, הסיכוי למציאת אפקט מובהק מקרי כלשהו בין שני תנאים הוא גדול מאלפא (הגדלת הטעות מסוג ראשון). זוהי השמטת תנאים בעייתית. על מנת להיות חוקרים טובים אנו צריכים לדווח באופן כנה על כל המשתנים והתנאים שנכללו בניסוי וגם על המבחנים בהם לא נמצא אפקט. את מתוך ההנחה כי אם נעשה מספיק מבחנים על מספיק תנאים- משהו יצא מובהק. בעיה נוספת היא שבמידה ואין אפקט

ולא נפרסם זאת, אנו יכולים "לבזבז" זמן של חוקר אחר שיחשוב על השאלה הזאת ויחקור אותה כאשר אין לו את הידע המוקדם שנוכל לספק לו על התוצאות של המחקר שלנו.

סעיף ג':

לדעתנו ניתן להתגבר על בעיה כמו זו שעולה מסעיף ב' על ידי:

1. מחשבה מראש- לחשוב מראש על המשתנים הבלתי תלויים שאנו רוצים לבדוק בניסוי. לא לכל תנאים ומשתנים שלא לצורך הממוקד של שאלת המחקר
2. דיווח כנה- לדווח בכנות על המשתנים והתנאים שנכללו בניסוי.
3. להגביל את עצמנו מראש ולהחליט איזה מבחנים סטטיסטיים אנו מעוניינים לבצע ועל איזה נתונים ולהשתדל שתהיה כמה שפחות תלות בין המשתנים שנבחנו למבחנים השונים. אם השתמשנו במשתנה תלוי כלשהו ואנו בודקים את השפעת משתנה בלתי תלוי עליו, נבחר אותו בחכמה ולא ננסה את כל המשתנים הבלתי תלויים האפשריים על אותו משתנה תלוי. אולי אפשר לבדוק בנפרד עבור על משתנה בלתי תלוי נתונים חדשים של המשתנה התלוי ולא להשתמש באותם נבדקים להכל.
4. דיווח מלא של התוצאות- רווח בר סמך, ממוצעים, גודל האפקט, עוצמת המבחן וכן הלאה..

סעיף ד':

על מנת ליצור התפלגות דגימה באמצעים לא פרמטריים בעזרת *bootstrap* ננהג בשלבים הבאים:

1. עבור T איטרציות נבצע:

- נדגום עם החזרות, m דגימות מתוך כלל הדגימות. זאת כאשר m קטן ממספר הדגימות ונבחר באופן אחיד מראש.
- נבצע את המבחן הסטטיסטי על הדגימה שהגרלנו ונשמור את התוצאות

2. נספור כמה פעמים מתוך T האיטרציות התקבלה מובהקות סטטיסטית על הדגימה.

3. נחליט לפי טרשהולד שנקבע אם היה אפקט או לא.

במבחן **פרמוטציה** ננהג באופן דומה רק שהפעם נדגום בלי החזרות m דגימות מתוך כלל הדגימות שקיבלנו. היתרון של תהליך זה הוא שהוא מקטין את השונות ובכך "מחפה" על חלק מהבעייתיות של הליך בדיקת ההשערות. הוא יכול לגרום ליותר דיוק של רווח הבר סמך ובכך תורם להקטנה הטעות מסוג ראשון. אינו דורש הנחת יסוד שהדאטא לקוח מהתפלגות פרמטרית כלשהי וזה טוב כי לעיתים לא ניתן להסתמך על הנחות פרמטריות. בבוטסטראפ מניח שהדגימה בגודל m עם החזרות מדמה באופן טוב את תהליך הדגימה מתוך אוכלוסייה גדולה יותר ועל כן מבצע את ההחזרות. לדעתי לא בהכרח היינו מקבלים את אותה התוצאה כמו המבחן הפרמטרי. יתכן כי המשתנה שבחרנו- נועם הליכות הוא לא פרמטרי ולכן לא מיוצג טוב במספרים האלו.