שיטות מחקר־ תרגיל 7

מגישות זוהר בוחניק 311142293 אלה דובדבן 305564866

27 במאי 2019

שאלה 1:

ככל שהשונות הכללית בין־נבדקים רבה יותר נעדיף להשתמש במערך **תוך נבדקי**. זאת מכיוון שבניגוד למערך תוך־נבדקי, במערך בין־נבדקי שונות כללית רבה בין נבדקים **תגדיל** את הסיכוי לטעות מסוג **שני**.

כאשר אנו מבצעים מבחן תוך נבדקי אנחנו בעצם יודעים להסביר מקור נוסף לשונות – אנחנו יודעים שהוא מגיע ממקור של נבדקים. על ידי כך המודל שלנו נהיה טוב יותר ובמידה והשערת האפס לא נכונה, במערך תוך נבדקי תחת הנתונים של השונות בין הנבדקים שבולטת, יש סיכוי גדול יותר שנצליח לדחות את השערת האפס. כלומר שימוש במערך בין נבדקי במצב של שונות כללית רבה בין נבדקים מגדיל את הסיכוי לכך שלא נדחה את השערת האפס למרות שהיא לא נכונה, שזה בדיוק טעות מסוג שני.

שאלה 2:

סעיף א'

העלנו את ה־$data$ ל־R בעזרת הכלים שלמדנו בתרגול. ניתן לראות זאת בקובץ R המצורף

סעיף ב'

```
Error: Subject
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  7 649583   92798

Error: Subject:facial_expression
                  Df Sum Sq Mean Sq F value  Pr(>F)
facial_expression  2 135625   67812   11.71 0.00102 **
Residuals         14  81042    5789
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

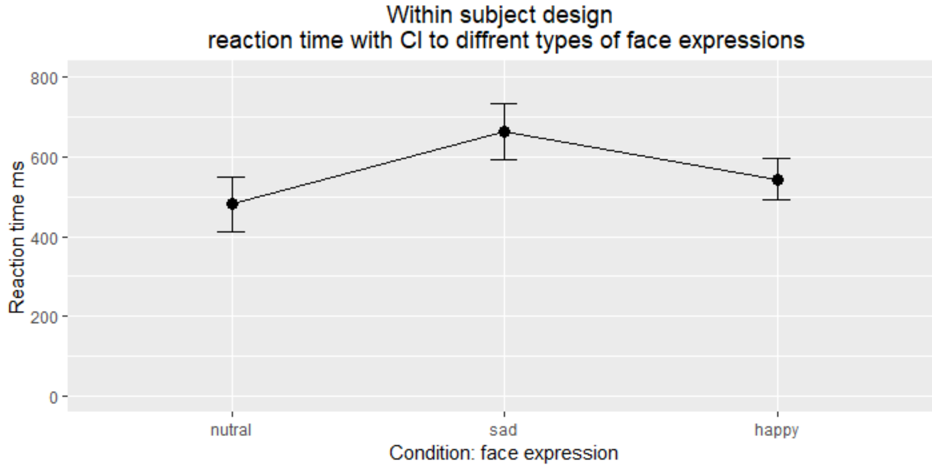
נסכם את התוצאות שהתקבלו:

- ערך F ־ 11.71
- דרגות חופש־ 2 ו־ 14 (כלומר $F_{(2,14)}$)
- ערך p ־ 0.00102
- מובהקות־ כן, כלומר דחינו את השערת האפס (וגם ערך ה־ p קטן בהרבה מ 0.05)

מה שניתן להסיק מהניתוח הוא שערך ה p קטן מאוד, ובפרט קטן מ 0.05 ולכן ברמת מובהקות של 95 אחוז ניתן לדחות את השערת האפס ולהסיק כי סוג ההבעה של פרצוף אכן משפיע על מהירות העיבוד כמו שהחוקרת שיערה.

סעיף ג'

גרף המתאר את הנתונים:



סעיף ד'

ביצענו את המבחן כמבחן $anova\,1\,way$ כך שמתייחסים ל24 נבדקים שונים.
ואלו התוצאות:

```
> summary(anova_test)
                  Df Sum Sq Mean Sq F value Pr(>F)
facial_expression  2 135625   67812   1.949  0.167
Residuals         21 730625   34792
```

ההבדלים מהתוצאות שקיבלנו בסעיף ב' :

- ערכי ה-F קטן משמעותית ל- 1.949 כך שלא ניתן לדחות את השערת האפס. כלומר ניתן לראות ממש בבירור את ההשפעה של השונות בין הנבדקים על התוצאות
- $p-value$ בעת שווה ל 0.167 אשר גדול מ 0.05 וכמובן גדול בהרבה מהערך p במבחן הקודם.
- SSים- במבחן תוך נבדקי היה לנו $SST = SSB + SSs + SSI$ כלומר ידענו להסביר יותר מהשונות בעזרת השונות בין הנבדקים עצמם. כעת כשעברנו למבחן אנובה חד גורמי בין נבדקי עברנו ל- $SST = SSB + SSW$ כלומר החלק של השונות הבלתי מוסברת נהיה גדול יותר. כצפוי ה- SSB נשאר אותו דבר, אמנם ה- SSs ו- SSI התאחדו ל- SSW לכן בעצם הסכום של $SSI + SSs$ בדיוק שווה ל- SSW $649583 + 81042 = 730625$.
- מובהקות- אין מובהקות תחת המבחן הזה, לא ניתן לדחות את השערת האפס.
- דרגות החופש- עלו ל $(2, 21)$ כי "הגדלנו" את מספר הנבדקים.

על מקור ההבדלים פירטנו כאשר הסברנו על השינוי ב SSים.

שאלה 3:

בניתוח שונות תוך-נבדקי ייתכן מצב בו $SST = SSB + SSs$? תשובה: **נכון**.
$SST = SSB + SSs + SSI$ ולכן מצב בו $SST = SSB + SSs$ שקול למצב בו $SSI = 0$. נשים לב כי כאשר יש לנו רק עמודה אחת, כלומר אנו בודקים רק תנאי אחד נקבל:

$$SSI = \sum_{i=1}^{n}\sum_{j=1}^{K}\left(y_{ij} - \overline{y}_i - \overline{y}_j + \overline{y}\right)^2$$

מכיוון ש- $K = 1$ נקבל כי $SSI = \sum_{i=1}^{n}\left(y_{i1} - \overline{y}_i - \overline{y}_1 + \overline{y}\right)^2$ אמנם לכל $i \in [n]$ מתקיים $y_{i1} = \overline{y}_i$, בגלל שיש רק איבר אחד בשורה אז ממוצע השורה $\overline{y}_i$ שווה לערך של האיבר בשורה y_{i1}. בנוסף מתקיים $\overline{y}_1 = \overline{y}$ כיוון שהממוצע הכללי שווה לממוצע העמודה הראשונה, שוב כי יש רק עמודה אחת. לבסוף נקבל כי $SSI = \sum_{i=1}^{n}\left(y_{i1} - \overline{y}_i - \overline{y}_1 + \overline{y}\right)^2 = \sum_{i=1}^{n}\left(0+0\right)^2 = 0$ כנדרש.

שאלה 4:

תשובה־ **ב'**
יש לנו כאן אפקט תקרה־ הוא פוגע במהימנות (לא מראה שונות אמיתית בין נבדקים), ובתוקף (סביר שיש הבדלים במשתנה התיאורטי, אבל אנחנו לא תופסים אותם כי המבחן קל\קשה מדי).

שאלה 5:

תשובה **ג'**.
כמוזכר בשאלה, מערך ניסויי מאפשר לנו שליטה ונטרול של משתנים חיצוניים רבים ־ ובכך מאפשר לוודא שהתוצאות שהתקבלו במשתנה התלוי נובעים מהמשתנה הבלתי תלוי ולא מגורמים חיצוניים ־ ולכן הוא בעל תוקף פנימי גבוה.
לעומת זאת, במבחן מתאמי ־ לא מבצעים מניפולציות על המשתנה הבלתי תלוי, שכן רוצים למצוא קשר \ מתאם בין שני הגורמים. חוסר התערבות זה משאיר את תוצאות המבחן קרובות יותר למציאות הכללית, ולכן הינו בעל תוקף חיצוני גבוה.

שאלה 6:

נבדוק את ההשפעה של הפעולות הבאות על גודל האפקט ועל עוצמת המבחן.
גודל האפקט – מידת ההשפעה של המניפולציה. מיוצגת ע"י $\delta = \frac{\mu_x - \mu_y}{\sigma}$.
עצמת המבחן – מבוטאת ע"י $1 - \beta$ כאשר β הינה הטעות מסוג שני (המקרה בו לא דחינו את השערת האפס, למרות שהשערת המחקר נכונה).

1. הגדלת המניפולציה הניסויית כך שהממוצע בקבוצה הניסויית יתרחק מהממוצע של קבוצת הביקורת.

(א) גודל האפקט **יגדל־** אנו מגדילים את $\mu_x - \mu_y$ כלומר את ההפרש בין הממוצע של הקבוצה הניסויית ושל קבוצת הביקורת ועל ידי כך מגדילים את האפקט.

(ב) עוצמת המבחן **תגדל־** הרחקת הממוצעים זה מזה בעצם מקטינה את השטח החופף של ההתפלגויות עבור ההשערות השונות, ועל ידי כך מביאה להורדת ההסתברות לקבל את הממוצע של הקבוצה הניסויית תחת הנחה H_0, כלומר להעלאת ההסתברות לדחיית H_0.
היות והטעות מסוג שני תלויה בשטח החפיפה בין ההתפלגויות ־ הקטנת שטח החפיפה מקטינה טעות מסוג שני, מה שמביא להגדלת עוצמת המבחן.

2. הגדלת המדגם.

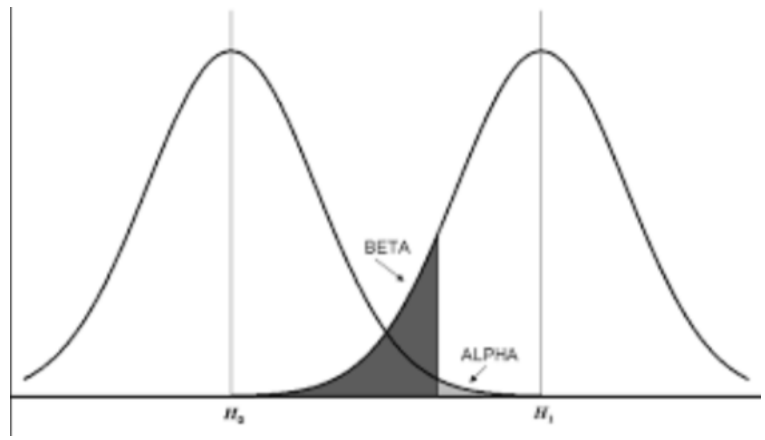
(א) גודל האפקט **יגדל־** כפי שהראינו בכיתה, הגדלת המדגם מביאה להצרה של עקומת הגאוס של ההתפלגות, כלומר להקטנת סטיית התקן. מתוך הנחה שמידת ההשפעה לא משתנה (הפער בין הממוצעים) ־ נטען כי גודל האפקט גדל עם הגדלת המדגם.

(ב) עוצמת המבחן **תגדל־** נשים לב שהטעות מסוג שני מוגדרת כהסתברות לאי דחייה של השערת האפס כאשר השערת המחקר (H_1) היא הנכונה. ככל שנגדיל את המדגם כמו שציינו קודם, שטח החפיפה בין ההתפלגויות יקטן ובעקבות כך כמו שלמדנו נקטין את הטעות מסוג ראשון ושני, לכן בפרט β יקטן ועוצמת המבחן תגדל.

3. הגדלת אלפא.

(א) גודל האפקט **לא ישתנה־** שכן אין הוא תלוי ב־α.

(ב) עוצמת המבחן **תגדל־** ככל ש־α גדולה יותר, יהיה יותר "קל" לדחות את השערת האפס, בעקבות כך בשל ה־$trade\,off$ בין α ו־ β , הטעות מסוג 2 תקטן, מה שיגדיל את עוצמת המבחן. ניתן לראות בגרף את ההשפעה של הגדלת α על הקטנה של β:


BETA
ALPHA
H_0
H_1