

שיטות מחקר תרגיל 5

מגישות:

מיכל דגן 315657064

נועה בן דרור 316163260

שאלה 1

א. על מנת לבדוק את השערתה של מירב, נבצע מבחן t להפרש בין תוחלות, וזאת מכיוון שצוין כי ניתן להניח שכמות החרקים מתפלגת נורמלית, וכי השוניות זהות בין הקבוצות.

ניסוח השערות:

השערת H_0 – הפרש התוחלות בין ספריי A (x) לספריי B (y) קטן או שווה ל-0.

$$H_0: \mu_x - \mu_y \leq 0$$

השערת H_1 – הפרש התוחלות בין ספריי A (x) לספריי B (y) גדול מ-0.

$$H_1: \mu_x - \mu_y > 0$$

קביעת אזור דחייה:

נקבע אלפא = 0.05 (5%). מדובר בהשערה חד זנבית.

ביצוע המבחן:

ערכנו מבחן t , וזו התוצאה שקיבלנו:

```
data: x[x$spray == "A", ]$count and x[x$spray == "B", ]$count
t = -0.45352, df = 22, p-value = 0.6727
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
-3.988519      Inf
sample estimates:
mean of x mean of y
14.50000 15.33333
```

הסקת מסקנה:

נשים לב כי קיבלנו p-value שערכו 0.6727, זהו ערך גדול מהאלפא שבחרנו (0.05), ולכן לא נדחה את השערת H_0 .

ב. על מנת לבדוק את השערתו של גדעון, נבצע מבחן one way anova, וזאת מכיוון שצוין כי גדעון מעוניין לבדוק האם קיים הבדל בין 6 סוגי הספריי – אלו 6 קבוצות בלתי תלויות. כמו כן, צוין כי ניתן להניח שכמות החרקים מתפלגת נורמלית, וכי השוניות זהות בין הקבוצות.

ניסוח השערות:

השערת H_0 – לכל ששת סוגי הספריי יש תוחלת שווה.

השערת H_1 – יש הבדל בין תוחלת ששת סוגי הספריי.

קביעת אזור דחייה:

נקבע אלפא = 0.05 (5%).

ביצוע המבחן:

ערכנו מבחן R^2 anova, וזו התוצאה שקיבלנו:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
spray	5	2669	533.8	34.7	<2e-16 ***
Residuals	66	1015	15.4		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

הסקת מסקנה:

נשים לב כי קיבלנו p-value שערכו קטן מהאלפא שבחרנו (0.05), ולכן **נדחה את השערת H_0** . כלומר, בסבירות של 95%, אכן קיים הבדל בין ששת סוגי הספריי ביחס לכמות החרקים שנותרה אחרי השימוש בו.

ג. ההבדלים בין סעיף א' לב':

- בסעיף א' בדקנו שתי קבוצות בלתי תלויות, ובסעיף ב' בדקנו שש קבוצות בלתי תלויות. לכן, בסעיף א' ערכנו מבחן t, ובסעיף ב' ערכנו מבחן one way anova.
- בסעיף א' בדקנו הפרש תוחלות בין שתי הקבוצות, ואילו בסעיף ב' בדקנו האם יש הבדל בתוחלות בין שש הקבוצות.
- בסעיף א' לא דחינו את השערת האפס, בסעיף ב' כן דחינו את השערת האפס. יתכן כי אין הבדל בין ספריי A לבין ספריי B, ולכן לא דחינו את השערת האפס בסעיף א'. עם זאת, יכול להיות הבדל בין ספריי A לבין ספריי C למשל, ואז נדחה את השערת האפס, כפי שבאמת עשינו בסעיף ב'.

נקודות הדמיון בין סעיף א' לב':

- בשני הסעיפים בדקנו השערה שקשורה להבדלים בין תוחלות של שתי קבוצות, או יותר.
- בשני הסעיפים הנחנו שכמות החרקים מתפלגת נורמלית, וכי השוניות זהות בין הקבוצות.

שאלה 2

א. סט של קונטרסטים מתוכננים התואמים את ההשערות הבאות :

1. הכושר הגופני הדרוש למשחקי כדור נמוך ממקצועות אחרים -

כלומר החוקרת תרצה לבדוק את ההבדל בין הכושר הגופני הדרוש למשחקי כדור לבין הכושר הגופני הדרוש למקצועות אחרים ולכן הקונטרסט הבא יתאים :

ריצה	אופניים	כדורגל	כדורסל
1	1	-1	-1

2. הכושר הגופני הדרוש לכדורגל גבוה יותר מזה הנדרש לכדורסל -

כלומר הפעם החוקרת מתעניינת רק בהבדל בין הכושר הגופני הדרוש לכדורגל לעומת זה הדרוש לכדורסל, ולכן הקונטרסט המתאים יהיה :

ריצה	אופניים	כדורגל	כדורסל
0	0	1	-1

ב. נבנה קונטרסט נוסף שיבחן את השאלה : הכושר הגופני הדרוש לריצה גבוה יותר מזה הדרוש לאופניים

ריצה	אופניים	כדורגל	כדורסל
1	-1	0	0

קונטרסט זה אורתוגונלי לשני הקונטרסטים האחרים שהגדרנו מאחר והוא בודק השוואה בין קבוצה שונה ולא קשורה להשוואות בסעיפים הקודמים. קבוצה זו לא תלויה בתוצאות ההשוואות הקודמות, כלומר לא קיימת חפיפה בין השונויות המוסברות בכל השוואה, כתוצאה מכך תוצאות ניתוח של השוואה אחת לא יצביעו על תוצאות הניתוח של השוואה אחרת.

ג. נראה אי תלות בין שלושת הקונטרסטים שהגדרנו :

ראשית, נשווה בין הקונטרסט הראשון לקונטרסט השני :

$$1*0 + 1*0 + -1*1 + -1*-1 = -1+1=0$$

מהחישוב עולה שהקונטרסט הראשון והשני בלתי תלויים.

שנית, נשווה בין הקונטרסט השני לקונטרסט השלישי :

$$0*1 + 0*-1 + 1*0 + -1*0 = 0$$

מהחישוב עולה שהקונטרסט השני והשלישי בלתי תלויים.

לבסוף, נשווה בין הקונטרסט הראשון לקונטרסט השלישי :

$$1*1 + 1*-1 + -1*0 + -1*0 = 1+-1 = 0$$

לאור ההשוואות עולה כי שלושת הקונטרסטים בלתי תלויים.

ד. תוצאות הבדיקה בR :

```
Confidence level used: 0.95
> contrast(leastsquare, weights, adjust="none")
contrast      estimate    SE df t.ratio p.value
c(-1, -1, 1, 1)      5.17 1.65 20   3.135 0.0052
c(-1, 1, 0, 0)       1.33 1.17 20   1.144 0.2661
c(0, 0, -1, 1)      -0.50 1.17 20  -0.429 0.6725
```

מסקנות התוצאות:

- א. בקונטרסט הראשון קיבלנו $p\text{-val}$ קטן מאלפא ולכן ניתן לדחות את השערת האפס, כלומר הכושר הגופני הדרוש למשחקי כדור אכן נמוך ממקצועות אחרים.
- ב. בקונטרסטים האחרים ניתן לראות כי קיבלנו $p\text{-val}$ גדול מאלפא ולכן לא ניתן לדחות את השערת האפס, כלומר לא ניתן להסיק באופן סטטיסטי שהכושר הגופני הדרוש לכדורגל גבוה יותר מזה הנדרש לכדורסל וכן שהכושר הגופני הדרוש לריצה גבוה יותר מזה הדרוש לאופניים.

שאלה 3

- א. ערך ה-F יקטן, משום שניתן לראות כי ביחס לשאר הקבוצות, ממוצע זמן ההגעה של כלל קבוצת "רכב פרטי" נמוך יותר משאר הקבוצות. לכן, אם נעלה את זמן ההגעה בקבוצה זו ב-4 דקות, למעשה "נקרב" את הממוצע של קבוצה זו לממוצע הכללי של הקבוצות, ואז השונות תקטן. בנוסף, מאחר והשונות בתוך הקבוצות לא תשתנה - ערך ה-F יקטן.
- ב. ערך ה-F יישאר כפי שהיה, מכיוון שהשונות לא תשתנה - החסרנו 10 דקות מכל האנשים מכל הקבוצות, ולכן היחסים נשארו אותו דבר, וערך ה-F נשאר אותו דבר.

שאלה 4

דפנה קיבלה ערך של $F = -0.79$, מוטי צודק, שכן דפנה קיבלה ערך F שלילי, וכפי שלמדנו התפלגות F לא יכולה לקבל ערכים שליליים – תמיד חיובית (כיוון ששונות לא יכולות לקבל ערכים שליליים) ולכן כנראה שיש טעות בחישוביה של דפנה.

שאלה 5

לכל אחת מהתצפיות בקבוצת "ירושלים" סטודנט הוסיף נקודה אחת במשתנה התלוי (רמת החרדה), התשובה הנכונה היא: ב'

"בעקבות טעותו, סכום הסטיות הריבועיות בתוך הקבוצות SSW קטן"

הדבר אינו אפשרי שכן SSW מתאר את סכום ריבועי הסטיות בתוך הקבוצות :

$$\sum_{j=1}^c \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$$

נשים לב שאם הסטודנט הוסיף לכל נבדק נקודה, אזי ההפרש בין ערך הנבדק לבין ממוצע קבוצתו יכול רק לגדול או להישאר זהה, אך לא לקטון :

נסמן m סכום התצפיות לפני ההגדלה ו- m מספר התצפיות בקבוצה אזי מתקיים :

$$m/n = \text{הממוצע לפני ההגדלה}$$

כעת, אם הוספנו לכל תצפית נקודה (m נקודות הוספנו) למעשה נקבל : $m/n + 1 = m/n + n/n = (m+n)/n$

כלומר הממוצע גדל ב-1, אך גם כל תצפית גדלה ב-1 ולכן ההפרש בין כל תצפית לממוצע הקבוצה יישאר זהה.