

## Multiple Comparison Analysis Tests

בתרגיל זה נבצע מבחן השוואות מרובות ונתקן את רמת המובהקות על ידי שליטה ב-FWER וב-FDR.

מטרת התרגיל היא למצוא באילו ערים אחוז הנדבקים בקורונה עובר סף קריטי, הקובע כי העיר מתוייגת כאזור "אדום", ובאילו ערים אחוז הנדבקים אינו עובר את הסף, והן מוגדרות כאזור "ירוק".

לצורך כך, העלו את קובץ הנתונים "multipleTestingData.csv".

בקובץ זה, העמודה הראשונה בשם "criticalDistribution" מתארת את כמות הנדבקים הקריטית להגדרת עיר כעיר בסיכון מוגבר (1 מ-100). אם אחוז הנדבקים בעיר עולה על האחוז הקריטי, העיר תוגדר כ"אדומה".

השורה positive מתארת את כמות הנדבקים שנמצאה בבדיקה, והשורה negative מתארת את כמות הנבדקים שנמצאו בריאים בבדיקה באותה העיר.

שאר העמודות מתארות את ממצאי הבדיקה בכל עיר.

משימה:

עליכם להחליט איזו עיר היא אדומה ואיזה עיר היא ירוקה, תוך שימוש בשלושה סוגי תיקונים לקריטריוני הסף:

1. תיקון בונפרוני לרמת המובהקות (הגבלת ה-FWER)
2. תיקון הולמס לרמת המובהקות (הגבלת ה-FWER)
3. תיקון בנימיני-הוכברג לרמת המובהקות (להגבלת ה-FDR).

אופן הפתרון:

0. הגדירו את סדרת השערות האפס הנבדקת במבחן.
1. על מנת לחשב את אחוז הסטייה הסטטיסטית (ה-Pv) של כל עיר מההתפלגות הקריטית הצפויה (ה-"criticalDistribution") יש לבצע מבחן פישר נפרד לכל עיר אל מול ההתפלגות הקריטית, ולבחון האם ה-OR הוא קטן משמעותית מהצפוי יש לשמור את ערכי ה-Pv המתקבלים בטבלת תוצאות נפרדת רמז: על מנת לחלץ את ערך ה-pv מתוצאות המבחן ניתן לכתוב:  

```
pv = as.numeric(h_fisher$p.value[1])
```
2. יש לסדר את ערכי ה-Pv שהתקבלו בסדר עולה.
3. יש לקבוע קריטריון דחיה מתוקן לפי בונפרוני:  
 $\text{significance level} = \alpha/m, \alpha = 0.05$   
ולקבל / לדחות את השערת האפס לכל עיר בהתאם לקריטריון זה.

4. יש לקבוע קריטריון דחיה מתוקן לפי הולמס:

$$\text{significance level} = \alpha / (m+1-i)$$

ולקבל / לדחות את השערת האפס לכל עיר בהתאם לקריטריון זה.

5. יש לקבוע קריטריון דחיה מתוקן לפי בניימיני-הוכברג:

$$\text{significance level} = \alpha * i / m$$

ולקבל / לדחות את השערת האפס לכל עיר בהתאם לקריטריון זה.

לסיכום:

מה ניתן ללמוד מתהליך קבלה/דחיה זה לגבי שליטה ב-FDR לעומת שליטה ב-FWER?