

## סטטיסטיקה לכריית מידע (תשפ"ו)

### פרויקט - משימה 3

#### בדיקת השערות, ניתוח שונות – ANOVA, מבחני "חי" (טיב התאמה ותלות)

עבור משתנה אחד מרכזי (או יותר = בונס) מבסיס הנתונים שבידכם -

1. עבור 2 קבוצות שונות, כלשהן (מתוך הלפחות 4 שמרכיבות את בסיס הנתונים, מהמשימה הראשונה), דגמו 6% מהנתונים בקבוצה האחת, ו- 10% מהקבוצה השנייה (עגל לשלם הקרוב), בצעו מבחן על הפרש תוחלות, עם השערה כלשהיא אלטרנטיבית ( $H_1$ ), שתשער/תנסח בהתאם למדגם ולבסיס הנתונים. מה תהיה קביעתך ברמות הביטחון: 90%, 95%, 99% (-שלושה מקרים נפרדים).
2. בהמשך לשאלה הקודמת, עבור 2 קבוצות האחרות שלא נדגמו, דגמו לפחות 32 תצפיות ("לב"), מכל אחת מהן, ובצע מבחן על הפרש תוחלות, עם השערה כלשהיא אלטרנטיבית ( $H_1$ ), ורצוי – השערה השונה מהנעשה בשאלה הקודמת, ואשר תהיה בהתאם למדגם ולבסיס הנתונים. מה תהיה קביעתך ברמות הביטחון: 90%, 95%, 99% (-שלושה מקרים נפרדים).
3. מהשאלות הקודמות - עבור 2 קבוצות ומהנתונים שנדגמו ונבחנו במבחן על הפרש תוחלות, תוך כדי שימוש בהתפלגות "סטודנט" ( $T_0$ ), בצע מבחן על יחס שונות, עם השערה כלשהיא אלטרנטיבית ( $H_1$ ), שתשער/תנסח בהתאם למדגם ולבסיס הנתונים, רמות הביטחון: 90%, 95%, 99% (-שלושה מקרים נפרדים). מה תהיה קביעתך בהמשך לנעשה בנתונים אלו ובמבחן  $T_0$  - האם המבחן על הפרש תוחלות היה תקין?  
עד כמה יש שוני/סטטיה מהשונות האמיתיות של כלל האוכלוסייה (-הסיגמות הידועות, שהופקו מכלל האוכלוסייה).
4. עבור הקבוצות השונות (מתוך הלפחות 4 שמרכיבות את בסיס הנתונים, מהמשימה הראשונה), בחרו אקראית/הגרילי' כ – 8% לקבוצה הראשונה, 12% לקבוצה השנייה, 18% לקבוצה השלישית ועל זה הדרך (עגל לשלם הקרוב) לשאר הקבוצות. חשבו את ה-  $SSB$ ,  $SST$  ו-  $SSW$ , עבור מבחן ה-  $ANOVA$ . מה תהיה קביעתך ברמות הביטחון: 90%, 95%, 99% (-שלושה מקרים נפרדים).
5. עבור כלל בסיס הנתונים, הצע בדיקת טיב התאמה, להתפלגותן של הנתונים (-חשוב מה הכי יתאים), ולאחר מכן דגום כ – 30% מהתצפיות וערוך מבחן "חי" לטיב התאמה ולהתפלגות זו. מה תהיה קביעתך ברמות הביטחון: 90%, 95%, 99% (-שלושה מקרים נפרדים).
6. הצע 2 משתנים שלדעתך אולי תלויים/לא תלויים (-וכך נראה בהתפלגות/בחינה ראשונית), וערוך מבחן "חי" לתלות המשתנים. מה תהיה קביעתך ברמות הביטחון: 90%, 95%, 99% (-שלושה מקרים נפרדים).

הפרויקטים ניתנים לעשייה לבד, או בזוג (עדיף).

נא לרכז בקובץ אחד (אקסל), עם גיליונות לא מרובים, וכדאי עם קובץ הסבר/מלווה ב-  $WORD$ .  
ההגשה תעשה בסוף המשימות והסמסטר, בתיבה ייעודית (במודל).