

## אוניברסיטת חיפה – החוג לסטטיסטיקה

### מודלים סטטיסטיים ב'

#### עבודת בית מס' 2

#### מודל ניתוח שונות חד-כיווני

#### הנחיות:

- את עבודת הבית ניתן ומומלץ לבצע בזוגות (אבל לא חובה).
- את הגשת העבודה יש להגיש באמצעות *טופס הגשה* (ראה טופס בהמשך התרגיל).
- יש לרשום את התשובות באופן ברור ונקי בטופס ההגשה.
- בנוסף יש להדפיס את כל הפלטים הרלוונטיים ולצרף לטופס ההגשה באופן מסודר.
- את הנתונים יש להוריד מאתר הקורס (Data Sets -> Data Sets for HW #2). הוראות "טכניות" הקשורות לקובצי נתונים נמצאות באתר.
- **שימו לב: יש לייצר את הנתונים על פי ת"ז כמפורט באתר (כמו שבוצע עבור ת"ב מס' 1).** כמו כן, **שימו לב להנחיות Paste Special / Message.pdf.**
- יש לציין את מספר ת"ז, כולל סיפרת ביקורת, התואם לנתונים במקום המיועד לכך בטופס ההגשה (זהו אחד מתעודות הזהות של המגישים).
- יש לבצע את כל הניתוחים באמצעות SAS.
- בכל מקום שבו נדרשים לבצע מבחן סטטיסטי יש להשתמש ברמת מובהקות  $\alpha = 5\%$ .
- בכל מקום שבו נדרשים לציין בטופס ההגשה P-Value אשר התקבל, יש להקפיד לרשום P-Value אשר מתאים למערכת ההשערות הנבדקת. **שימו לב:** יש להעתיק את הערכים המספריים הרלבנטיים לטופס ההגשה באופן מדויק, כפי שמופיעים בפלט של SAS.
- יש להקפיד על עבודה מדויקת. ציון יינתן ברובו על סמך נכונות הערכים המספריים המצוינים בטופס ההגשה.
- המועד האחרון להגשת התרגיל 18/06/07 (יום ב'), **לתא מזכירות החוג!** סטודנט אשר לא יגיש את התרגיל בזמן או/ו לא ימלא אחר הוראות ההגשה אחרות בצורה קפדנית יקבל ציון "0" באופן אוטומטי עבור תרגיל זה.

## שאלה 1

**חלק 1:** במחקר השוואתי שנערך ע"י הלשכה המרכזית לסטטיסטיקה נבדק, בין השאר, השכר החודשי (במוצע שנתי) של עובדי ההי-טק בשלוש הערים הגדולות בישראל: ירושלים, תל אביב וחיפה. נתוני השכר ב- שני מופיעים בקובץ "ID\_GEN\_wage.xls".

(א) בדקו את הנחות המודל:

(1) נורמליות של השגיאות באמצעות הגרף Q-Q-PLOT. האם מסתמן שהנחת הנורמליות מתקיימת בכל קבוצה (כן או לא)? רשום תשובתך בטופס ההגשה וצרף את הפלטים.

(2) שוויון שוניות ע"י מבחן Bartlett. האם מתקיים שוויון שוניות? יש לציין את ערך ה- P.V. בטופס ההגשה.

בסעיפים הבאים הניחו שהנחות שבדקתם בסעיף א' מתקיימות.

(ב) האם שכר עובדי ההי-טק זהה בשלוש הערים הללו?

(1) כתבו את ההשערה במפורש בטופס ההגשה.

(2) בצעו מבחן סטטיסטי ורישמו את ערך סטטיסטי המבחן.

במידה ומצאתם ששכר העובדים בשלוש הערים אינו זהה, המשיכו לסעיפים הבאים.

(ג) ערכו מבחן המשך מסוג Bonferroni לצורך מציאת הבדלי השכר.

(1) כתבו את ההשערה/ות הרלבנטיות/ות.

(2) מהי רמת המובהקות הרלבנטית לכל השוואה?

(3) מהו המרחק הקריטי המהווה קריטריון ההחלטה?

(4) מהי המסקנה מהמבחן הנ"ל (האם קיים הבדל מובהק בין הקבוצות 1, 2 ו-3 בזוגות)?

יש לכתוב תשובה קצרה ועניינית בטופס ההגשה לסעיף זה.

ד) ערכו מבחן המשך מסוג Tukey לצורך מציאת הבדלי השכר.

1) כתבו את כתבו את ההשערה/ות הרלבנטיות/ות.

2) מהו המרחק הקריטי המהווה קריטריון ההחלטה?

3) מהי המסקנה מהמבחן הנ"ל (האם קיים הבדל מובהק בין הקבוצות 1, 2 ו-3 בזוגות)?

יש לכתוב תשובה קצרה ועניינית בטופס ההגשה לסעיף זה.

ה) האם השכר בירושלים שווה למוצע השכר בתל אביב וחיפה.

1) רשמו את ההשערה המתאימה.

2) בידקו את ההשערה ורשמו את ערך סטטיסטי המבחן.

3) מהי המסקנה (דחיה/אי דחיה).

1) בסעיף זה השתמשו בתיאוריה של קונטרסטים אורתוגונליים.

1) רשמו את ההשערה הטוענת בעת ובעונה אחת שני דברים:

- השכר בירושלים שווה למוצע השכר בתל אביב וחיפה.

- השכר בתל אביב אינו שונה מהשכר בחיפה.

2) מבלי לבצע את המבחן ובהסתמך על הסעיפים הקודמים בלבד, מהו ערך סטטיסטי

המבחן הרלבנטי? הסבירו. (התשובה ללא הסבר או עם הסבר שגוי כלשהו!).

לא תזכה בניקוד

**המשך שאלה 1 - חלק 2:** בסקר השכר העולמי מופיעים הערים הגדולות בלבד. הנציגה היחידה מישראל הינה תל אביב. על מנת שניתן יהיה להשוות את כל שאר הערים מישראל לערים בחו"ל קיים צורך, בשלב ראשון, להשוות ערים שונות בישראל לתל אביב. לפי כך תל אביב הוגדרה כקבוצת ייחוס, כאשר אליה הושוו ירושלים וחיפה.

א) רשמו בטופס ההגשה מהו המבחן המתאים לצורך ביצוע ההשוואה הנ"ל? בצעו את המבחן ורישמו

את המרחק הקריטי המהווה קריטריון ההחלטה.

(ב) בנו מודל רגרסיה (באמצעות 2 משתני דמי) אשר מתאים למודל ניתוח שונות. בחרו את משתני הדמי כך שתל-אביב תהיה "קבוצת הייחוס". כתבו את המודל.

(ג) הריצו את המודל ובצעו מבחני T ל-  $\beta_1$  ול-  $\beta_2$ . מה משמעות התוצאות?

(ד) מה הקשר של תוצאות סעיף ג' לאלה מסעיף א'?

## שאלה 2

במפעל כימי נערך ניסוי שמטרתו להשוות את ביצועיהם של החומר החדש (B) והקיים (A) לקטילת חרקים. לצורך זה נלקחו 80 מיכלים של חרקים (כמות זהה של חרקים בכל מיכל), שרוססו ע"י שני החומרים. יעילות החומר נמדדת במספר החרקים, המתים תוך הדקה הראשונה של הניסוי. הנתונים הרלבנטיים נמצאים בקובץ ID\_GEN\_insects.xls.

הניחו כי מתקיים שוויון שונויות ונורמאליות השגיאות.

(א) רשום בטופס ההגשה את ההשערה הנבדקת.

(ב) מהם שני המבחנים האפשריים שניתן לבצע במקרה זה? נא לרשום את התשובה בטופס ההגשה.

(ג) נא לבצע את שני המבחנים הנ"ל ולרשום את סטטיסטי המבחן של שניהם בטופס ההגשה.

(ד) מהו הקשר בין ערכי הערכים הקריטיים (טבלתים) של שני המבחנים הנ"ל?

(ה) מהי המסקנה מהניסוי (דחייה/קבלה)?

**בהצלחה!**

מודלים סטטיסטיים ב'

עבודת בית מס' 2  
טופס תשובות להגשה

שמות המגישים:

\_\_\_\_\_

מספרי ת"ז:

\_\_\_\_\_

מס' ת"ז התואם לנתונים:

\_\_\_\_\_

שאלה 1

**חלק 1.**

1-א

\_\_\_\_\_

2-א

\_\_\_\_\_

1-ב

\_\_\_\_\_

2-ב

\_\_\_\_\_

1-ג

\_\_\_\_\_

2-ג

\_\_\_\_\_

3-ג

\_\_\_\_\_

4-ג

\_\_\_\_\_

1-ד

\_\_\_\_\_

2-ד

\_\_\_\_\_

3-ד

\_\_\_\_\_

1-ה

\_\_\_\_\_

2-ה

\_\_\_\_\_

3-ה

\_\_\_\_\_

1-ו

\_\_\_\_\_

2-ו

\_\_\_\_\_

**המשך שאלה 1 - חלק 2.**

א

\_\_\_\_\_

\_\_\_\_\_ ב

\_\_\_\_\_ ג

\_\_\_\_\_ ד

**שאלה 2**

\_\_\_\_\_ א

\_\_\_\_\_ ב

\_\_\_\_\_ ג

\_\_\_\_\_ ד

\_\_\_\_\_ ה