

האוניברסיטה העברית בירושלים  
המחלקה לסטטיסטיקה

עקרונות של תכנון ניסויים ושיטות דגימה (52710)

פרופ' דוד צוקר

תשס"ד סמסטר ב', מועד ב'

תאריך: כ"א במנחם אב תשס"ד, 8.8.04

משך הבחינה: שעתיים וחצי

חומר עזר: מחשב כיס ודפי רשימות (שני דפים בגודל A4, שני צדדים)

חל איסור מוחלט להעתיק. תלמיד שייתפס יורחק לשנה מלימודיו.

הערה: מכיוון שכמה מהשאלות במבחן זה דורשות בעיקר תשובות מילוליות, יש להקפיד מאד על כתב יד ברור. תשובות בכתב יד בלתי קריא לא ייבדקו.

בהצלחה!!

שאלה 1 (24 נקודות)

נערך סקר בו נשאלו 745 אנשים את השאלה הבאה: האם יש לאפשר לילדים חולי איידס ללמוד עם ילדים אחרים בבתי ספר (כן/לא)? הסקר התבצע באמצעות דגימת שכבות לפי רמת השכלה (שלוש רמות אפשריות). נלקח מדגם מקרי פשוט ללא החזרה מכל שכבה. נניח אוכלוסיה גדולה מאד כך שניתן להתעלם מכופלי סופיות. להלן התוצאות שהתקבלו.

| מספר אומרי "כן" במדגם | גודל המדגם בשכבה | משקל ( $W_h$ ) | שכבה                |
|-----------------------|------------------|----------------|---------------------|
| 86                    | 181              | 0.3            | פחות מהשכלה תיכונית |
| 259                   | 450              | 0.5            | השכלה תיכונית       |
| 84                    | 114              | 0.2            | סיימו קולג'         |

- [8] א. מה יהיה האומדן לפרופורציית אומרי "כן" באוכלוסיה?  
[8] ב. כיצד תאמדו את שונות האומדן? הציגו נוסחא לאומדן לשונות ואח"כ חשבו שת האומדן המתאים.  
[8] ג. כעת נשקול מה היה קורה לו בצענו את הסקר באמצעות דגימה מקרית פשוטה והשתמשנו באומדן הפשוט לפרופורציית אומרי "כן" באוכלוסיה. השתמשו בתוצאה של סעיף א' על מנת לאומדן את השונות של האומדן הפשוט תחת דגימה מקרית פשוטה. מה ניתן ללמוד מהשוואת אומדן השונות שהתקבל בחישוב הזה עם האומדן שהתקבל בסעיף ב'?

שאלה 2 (20 נקודות)

נערך סקר (באמצעות דגימה מקרית פשוטה ללא החזרה) לאמוד את צריכת הדלק החודשית הממוצעת של אוכלוסיית תושבי ירושלים. מכמה מהמשפחות שנבחרו למדגם החוקרים לא הצליחו להשיג את המידע הנדרש על צריכת דלק. אמנם מספר הנפשות במשפחה היה ידוע עבור כל המשפחות.

[8] א. הציגו שיטה לניצול המידע על מספר הנפשות במשפחה לבנות אומדן לצריכת הדלק הממוצעת באוכלוסיה. (יש להציג שיטה אחת בלבד).

[12] ב. באילו תנאים ניתן לקבל בשיטה שהצעת אומדן עקיב לפרמטר הרלוונטי? נמקו.

### שאלה 3 (30 נקודות)

נערך ניסוי לבדוק את היעילות של משחה חדשה נגד אבעבועות. השתתפו בניסוי 6 בחורים סובלי אבעבועות בגילאים 14-18. כל משתתף קיבל שתי שפופרות משחה דומות, שסומנו א' ו-ב'. שפופרת אחת הכילה את התרופה והשפופרת השנייה הכילה משחת פלצבו. עבור כל אחד מהמשתתפים ההסתברות ששפופרת א' הכילה את התרופה הייתה 0.50, באופן בלתי תלוי במשתתפים אחרים. נאמר למשתתפים לשים את משחה א' בלחי הימנית ואת משחה ב' בלחי השמאלית, בבוקר ובערב לחודש ימים. אחרי תקופת הניסוי, רמת השיפור בשתי הלחיים נמדדה על ידי רופא עור מומחה בסולם סטנדרטי שנע מ-0 עד 100. התוצאות להלן:

| משתתף | ראובן | שמעון | לוי | דן | נפתלי | יששכר |
|-------|-------|-------|-----|----|-------|-------|
| תרופה | 52    | 19    | 35  | 43 | 58    | 81    |
| פלצבו | 33    | 24    | 21  | 34 | 63    | 67    |

החוקרים החליטו להשוות בין התרופה לבין פלצבו באמצעות ההפרש בין ציון הממוצע תחת טיפול בתרופה לבין ציון הממוצע תחת פלצבו.

[15] א. בצעו מבחן רנדומזציה על מנת לבחון את השערת האפס שאין לתרופה שום השפעה. חשבו את ה-p-value הרלוונטי (חד-צדדי).  
[15] ב. ציינו שלוש בעיות אפשריות עם ניסוי במבנה כזה. הסבירו כל בעיה ב-1-2 משפטים (ולא יותר).

### שאלה 4 (14 נקודות)

נערך מחקר לבדוק את ההשפעה של עישון במשך הריון על הסיכון ללדת תינוק עם ציון אפגר גרוע (3 או פחות). כל בתי החולים במדינת Hatzaplatz השתתפו במחקר זה לכל שנת 2001.

התוצאות היו כדלהלן:

לא מעשנות: 203,761 לידות, 22 עם אפגר גרוע  
מעשנות: 107,523 לידות, 87 עם אפגר גרוע

חוו דעת על ערכו של מחקר זה בנוגע להסקת מסקנות על הקשר (או חוסר קשר) בין עישון לבין אפגר גרוע.

איזה מידע נוסף היה מועיל לפענוח התוצאות ?

### שאלה 5 (12 נקודות)

[6] א. תנו דוגמא של שיטת דגימה הסתברותית באשר גודל המדגם הוא משתנה מקרי.  
[6] ב. נתייחס לשיטת דגימה הסתברותית כלשהי. נסמן את גודל האוכלוסייה ב-N ואת גודל המדגם ב-n. נגדיר  $\pi_i = \Pr(i \text{ is in the sample})$ . הוכיחו כי:  $E[n] = \pi_1 + \pi_2 + \dots + \pi_N$ .