

## תרגיל 7- רב"סים ובדיקת השערות

סעיפים בדרגת קושי גבוהה יותר סומנו בכוכבית(\*). כדאי להשתדל לענות עליהם אבל אין הכרח.

1. תהי  $x$  תצפית **בודדת** מהתפלגות  $f_X(x, \theta) = 2\theta x + 1 - \theta; x \in [0, 1]$ .

(a) האם זו פונקציית צפיפות לגיטימית?

(b) מהו סטטיסטי המבחן ומהו כיוון הדחייה אם ההשערות הן  $H_0: \theta = 0; H_1: \theta = 1$  ?

(c) מיצאו מבחן ברמת מובהקות  $\alpha$  לבדיקת ההשערות לעיל.

(d) מה ייקרה לאיזור הדחייה אם ההשערות הן (לא צריך לחשב ממש אלא רק לתאר את הכיוון של השינוי):

i.  $H_0: \theta = 0; H_1: \theta = 2$

ii.  $H_0: \theta = 0.5; H_1: \theta = 1$

2. מנסיון העבר נראה כי זמן ההגעה של טכנאי לטיפול בתקלות MOODLE מתפלג מעריכית.

גם נאמר וטכנאים של MOODLE מגושמים קצת ולצורך פתרון תקלה, צריך שלושים טכנאים שמגיעים אחד אחרי השני. (הראשון מגיע, קורא מייד לשני וכן הלאה כך שזמני ההגעה הם i.i.d) עד הטכנאי השלושים שמצליח לתקן את התקלה.

(a) מה הסיכוי שהתקלה תתוקן תוך פחות מ 200 שעות אם זמן ההגעה הטיפוסי של טכנאי הוא 8 שעות?

(b) האם הנחת iid ריאלית לדעתך?

(c) כיצד ייראה רב"ס לזמן ההגעה הטיפוסי של טכנאי המבוסס על זמני התיקון של m תקלות?

(d) מה יהיה הרב"ס מסעיף קודם ברמת בטחון של 90% אם הזמן הממוצע לתיקון של 20 תקלות היה 220 שעות.

(e) כיצד ייראה איזור דחייה לבדיקת ההשערה שזמן ההגעה הטיפוסי הוא 8 שעות מול אלרטנטיבה שהוא למעשה 7 שעות?

i. מהו סטטיסטי המבחן?

ii. מה יהיה הכיוון של איזור הדחייה?

iii. מה יהיה סף הדחייה (הערך הקריטי) כפונקציה של רמת המובהקות וגודל המדגם?

3. מספר השאלות בשרתי האוניברסיטה מתפלג פואסונית.

מרכז החישובים טוען שבשנייה טיפוסית מבוצעות שתי שאילתות ולכן יש לקנות עוד שרתים.

חשב האוניברסיטה טוען שבשנייה טיפוסית מבוצעת שאילתא בודדת ולכן אין צורך בעוד שרתים ואפשר לחסוך את העלויות.

(a) נסח את ההשערות במונחים סטטיסטיים כאשר טענת מרכז החישובים היא הטענה לדחייה.

(b) סטטיסטי מבחן אפשרי הוא מספר השאילתות הממוצע בשנייה. האם נדחה לערכים גבוהים או נמוכים שלו?

**הערה:** אפשר לחשוב על מקרי קצה- לדוגמה שבממוצע לא בוצעו כלל שאילתות- האם זה סביר יותר תחת  $H_0$  או  $H_1$  ?

(c) כיצד מתפלג **סכום** השאילתות בארבע שניות אם מרכז החישובים צודק? ואם החשב צודק?

(d) אם הוחלט להכריע בהתבסס של מספר השאילתות הכולל שבוצעו בארבע שניות כלשהן שנבחרו באקראי ( $n=4$ ). הציעו מבחן שישמור על סיכוי לטעות קטן מ 5% ( $\alpha=0.05$ ).

(e) מהו איזור הדחייה אם תעדיפו לעבוד עם מספר השאילתות **הממוצע** במקום מספר השאילתות הכולל? (כלומר סטטיסטי המבחן הוא הממוצע ולא הסכום).

(f) מה יהיה הכיוון של איזור הדחייה של מספר השאילתות הממוצע בתרחישים הבאים?

i.  $[H_0: X \sim Pois(\lambda=2); H_1: X \sim Pois(\lambda=1.5)]$

ii.  $[H_0: X \sim Pois(\lambda=2); H_1: X \sim Pois(\lambda>2)]$

iii.  $[H_0: X \sim Pois(\lambda=2); H_1: X \sim Pois(\lambda<2)]$

iv.  $[H_0: X \sim Pois(\lambda=2); H_1: X \sim Pois(\lambda \neq 2)]$

v.  $[H_0: X \sim Pois(\lambda=2.5); H_1: X \sim Pois(\lambda=1)]$

vi.  $[H_0: X \sim Pois(\lambda=1.5); H_1: X \sim Pois(\lambda=1)]$

(g) מה תהייה ההתפלגות המקורבת של מספר השאילתות הכולל אם ישנם מספיק מחוברים כך שמשפט הגבול המרכזי תקף (סמנו את גודל המדגם על ידי  $n$ )?

(h) חיזרו על סעיף d בהתבסס על מדגם בגודל 100.

(i) כיצד ייראה רב"ס (מקורב) למספר השאילתות בשנייה טיפוסית ( $\lambda$ ) ?

(j) מה יהיה הרב"ס מהסעיף הקודם ברמת 95% אם מספר השאילתות הממוצע ב 100 שניות שנבחרו באקראי היה 1.54

4. הלשכה המרכזית לסטטיסטיקה אומדת מידי חודש את שיעור הבלתי מועסקים בארץ.

(a) אם אנשי הלמ"ס מעוניינים להבטיח שרק ב 5% מהמדגמים האומדן שלהם "יחטיא" את פרופורציית המובטלים האמיתית ביותר מאחז, כמה אנשים עליהם לדגום?

**הערה:** אי אפשר לחשב את גודל המדגם ממש בלי לדעת את שיעור המובטלים האמיתי, אבל אפשר לחסום אותו.

## מבוא לסטטיסטיקה למדמ"ח- תשס"ט

- (b) אם יש לאנשי הלמ"ס סיבה להאמין שפרופורציית המובטלים במדינה קטנה מ 20%, כמה אנשים יהיה עליהם לדגום כעת? האם המדגם גדל או קטן? האם זה מפתיע?
- (c) בהיעדר כל מידע על שיעור המובטלים במדינה, מה יהיה גודל המדגם אשר יבטיח שטווח ברוחב שני אחוזים ייתפוס את שיעור המובטלים האמיתי במדינה ב 95% אחוז מהמדגמים?
- (d) שכנעו את עצמכם שסעיפים (a) ו(c) הם אותה הבעייה בניסוחים שונים, ובפרט יביאו לאותה המסקנה על גודל המדגם הדרוש.
- (e) חשבו רב"סים (רגיל ושמרני) לפרופורציית המובטלים במדינה אם במדגם מקרי בגודל 100, התגלה כי 9% מהנדגמים הינם מובטלים ועם/בלי ההנחה מסעיף b (סה"כ 3 רב"סים).

5. קיראו את הקטע "אמת וסקר" (ב"קבצים לשיעורים" שבאתר הקורס) והסבירו בקצרה את ההטיות המודגמות בפרק.