

תרגיל 9- בדיקת השערות

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

1. מאָךר המחשבים השכונתי מנסה למכור לי שרת קבצים...
ידוע לי כי מהירות הגישה של שרת קבצים ניתנת לקירוב על ידי התפלגות נורמלית כאשר זמן הגישה הטיפוסי של שרתים מסוג א' הוא 2 ms, זמן הגישה הטיפוסי של שרתים מסוג ב' הוא 3 ms והשונות זהה ($\sigma=0.5$).
בכוונתי לבדוק האם השרת שהוצע לי הוא מסוג א' ולצורך הבדיקה אניח שהוא אכן כזה.
(a) אם בכוונתי למדוד את הזמנים של שלוש שליפות לפני שאכריע איזה שרת הוצע לי- באיזה סטטיסטי מבחן כדאי לי להשתמש?
(b) כיצד לנסח את הבעייה במונחים סטטיסטיים?
האם אופן ניסוח הבעייה תלוי בסטטיסטי המבחן בו אבחר?
(c) מהו המונח הסטטיסטי להחלטה שהשרת הוא מסוג ב' אם הוא בעצם מסוג א'?
(d) מהו המונח הסטטיסטי להחלטה שהשרת הוא מסוג א' אם הוא מסוג ב'?
(e) מהו המונח הסטטיסטי לסיכוי שאצליח לזהות שהשרת הוא מסוג ב' אם הוא אכן כזה?
(f) האם ערכים גבוהים של סטטיסטי המבחן יחזקו את יחלישו את אמונתי שהשרת הוא מסוג א' (כלומר האם אדחה לערכים גבוהים או נמוכים של סטטיסטי המבחן)?
(g) החלטתי שאם זמן הגישה הממוצע יהיה גדול מ-2.5ms אז כנראה שיש בידי שרת מסוג ב'. מהיא העוצמה של מבחן זה? מהו הסיכוי לטעות מסוג ראשון? מהו הסיכוי לטעות מסוג שני?
(h) איזה שרת המאכר כנראה מציע לי אם הזמן הממוצע של שלוש השליפות היה 2.45?
(i) מהי מובהקות התוצאה עבור זמן השליפה הממוצע מהסעיף קודם?
(j) כיצד אוכל להגדיל את העוצמה של המבחן על ידי שנוי סף ההחלטה שלי (כלומר איזור הדחייה-C)? מה החסרון בכך?
(k) האם ישנן דרכים מלבד שינוי איזור הדחייה בהן אוכל להגדיל את העוצמה של המבחן?
(l) מהי העוצמה של המבחן אם אשתמש בממוצע של **ארבעה** שליפות?
מהו הסיכוי לטעות מסוג ראשון? מהי העוצמה של המבחן?
(m) איזה שרת המאכר כנראה מציע לי אם הזמן הממוצע של **ארבעה** השליפות היה 2.45?
(n) מהי מובהקות התוצאה עבור זמן השליפה הממוצע מהסעיף קודם?
האם המובהקות גדלה או קטנה ביחס לסטטיסטי מבחן המבוסס על שלוש שליפות?
האם זה מפתיע?

מבוא לסטטיסטיקה למדמ"ח- תשס"ט

(o) מה הייתה מובהקות התוצאה אם האלטרנטיבה הייתה שזה פשוט אינו שרת מסוג א'?
(p) חשבו את העוצמה, את הסיכוי לטעות מסוג ראשון ואת הסיכוי לטעות מסוג שני של המבחן המוצע בסעיף (g) אם זמן הגישה הטיפוסי של שרתים מסוג ב' הוא למעשה 2.8ms ולא 3 כפי שהנחנו במקור.

(q) השווה את התוצאה מהסעיף הקודם עם התוצאה מסעיף (g):
כיצד השפיעו ההנחות השונות (כלומר השינוי ב H_1) על העוצמה של המבחן? ועל הסיכוי לטעות מסוג ראשון?
האם יכולנו לצפות שהעוצמה של המבחן תגדל גם ללא חישובים?
האם יכולנו לצפות את ההשפעה של השינוי ב H_1 על הסיכוי לטעות מסוג ראשון גם ללא חישובים?

(r) סעיף זה נועד להמחיש את הרעיון מאחורי בדיקת ההשערות באמצעות סימולציה ומתבסס על השוואת שרתי הקבצים אצל המאכר השכונתי:

i. הגרילו 1000 מדגמים של 3 זמני גישה (כל אחד) מההתפלגות $X_i \sim N(\mu=2, \sigma=0.5)$ (כלומר, כאילו מולכם שרת מסוג א').

ii. עבור כל מדגם, חשבו את זמן הגישה הממוצע.

iii. באיזה אחוז מן המדגמים זמן הגישה הממוצע יביא אותנו למסקנה שזה אינו שרת מסוג א' (כלומר זמן השליפה הממוצע היה גדול מ 2.5 ms) השוו את התוצאה עם סעיף (g).

iv. כעת הגרילו את הנתונים כאילו שלפניכם שרת מסוג ב' כלומר $X_i \sim N(\mu=2.8, \sigma=0.5)$.

v. באיזה אחוז מן המדגמים לא תצליחו לזהות שזה שרת מסוג ב' (טעות מסוג II)?

2. אם בידיכם מטבע לא מאוזן והחלטתם לבדוק האם הסיכוי לעץ הוא 20% או 40% על בסיס שלוש הטלות כשאתם מניחים שהסיכוי לעץ הוא 40%...

(a) נסחו את ההנחות בצורה סטטיסטית.

(b) מהם כל איזורי הדחייה האפשריים (הרציפים) המבוססים על מספר ה"עצים" בשלוש הטלות? (ישנם שמונה כאלו)
כדאי לסמן כל איזור דחייה באות לצורכי נוחות בהמשך השאלה.

(c) בנו טבלה של העוצמה, הסיכוי לטעות מסוג ראשון והסיכוי לטעות מסוג שני של כל איזור דחייה/מבחן שכזה.

(d) ציירו תרשים של העוצמה כנגד הסיכוי לטעות מסוג ראשון.

(e) איזה מאזורי הדחייה/מבחנים אפשר לפסול על הסף? (פשוט כי ישנם מבחנים אחרים טובים יותר בשני הקריטריונים).

3. כרגיל, עבודה על שריר האינטואיציה....

- (a) מובהקות התוצאה (p -value) של מבחן סטטיסטי כלשהו הייתה 0.04. האם השערת האפס תידחה ברמת בטחון של חמישה אחוזים ($\alpha=0.05$)?
- (b) אם במדגם של תצפית בודדת ובמדגם של מאה אלף תצפיות התקבלה בדיוק אותה **מובהקות תוצאה** (p -value), האם ייתכן שנגיע למסקנות **שונות**? כלומר שנדחה את H_0 במדגם אחד ובשני לא עבור אותה רמת מובהקות α ?
- (c) אם במדגם של תצפית בודדת ובמדגם של מאה אלף תצפיות התקבל אותו **סטטיסטי מבחן**, האם ייתכן שנגיע למסקנות שונות עבור אותה רמת מובהקות?
- (d) האם בשביל **לחשב** את סטטיסטי המבחן צריך הנחות על המודל ההסתברותי ממנו מגיעים הנתונים?
- (e) האם בשביל לחשב את **מובהקות התוצאה** צריך הנחות על המודל ההסתברותי ממנו מגיעים הנתונים?
- (f) אם כבר ניסחנו את ההשערות במונחים סטטיסטיים (כלומר יש לנו H_0 ויש H_1), האם סטטיסטי המבחן הוא **יחיד**? כלומר האם יש לנו את החופש להשתמש ביותר מסטטיסטי מבחן אחד?
- (g) אם כבר בחרתם קריטריון החלטה (כלומר איזור הדחייה מוגדר היטב), האם אפשר לחשב את הסיכוי לטעות מסוג ראשון ללא השערה אלטרנטיבית? וללא השערת אפס?
- (h) אם כבר בחרתם קריטריון החלטה (כלומר איזור הדחייה מוגדר היטב), האם אפשר לחשב את העוצמה של מבחן ללא השערת אפס? וללא השערה אלטרנטיבית?
- (i) אפשר למצוא אנלוגיה בין מבחנים סטטיסטיים ובין מערכת המשפט. מהם המונחים הסטטיסטיים המתאימים למצבים הבאים:
- הסיכוי שאדם אשם נמצא זכאי.
 - הסיכוי שאדם זכאי נמצא אשם.
 - הסיכוי שאדם אשם אכן יימצא אשם.
 - רמת הבטחון שדורש השופט לפני שהוא שולח אדם זכאי לכלא.
- (j) מה תהייה העוצמה של מבחן עבורו הסיכוי לטעות מסוג ראשון הוא α כאשר השערת האפס והאלטרנטיבה **זהות לחלוטין**.
- (k) (*) בנו תרחיש שבו קיים מבחן עם בטחון מלא ($\alpha=0$) ובעל עוצמה גדולה מאפס. הכיוון לפתרון: כל שצריך הוא למצוא השערת אפס והשערה אלטרנטיבית שיאפשרו כזה מצב. הפתרון אינו יחיד וישנם מצבים די פשוטים שיאפשרו זאת.
- (l) (*) בנו תרחיש שבו כל מבחן עם בטחון מלא ($\alpha=0$) לעולם לא יביא לדחיית השערת האפס. כלומר $\pi=0$.
- הכיוון לפתרון: גם כאן כל שצריך למצוא הוא שתי השערות שיביאו למצב זה. גם כאן הפתרון אינו יחיד.

4. שני חברים רבים מהו שיעור האתרים שאינם תומכים בפיירפוקס באינטרנט?

מבוא לסטטיסטיקה למדמ"ח- תשס"ט

אבנר 0 טוען ששיעור האתרים הוא 10% ואבנר 1 טוען ששיעור האתרים הוא 20%. הם מחליטים לגלוש עד שייתקלו לראשונה באתר שלא תומך בפיירפוקס. היות שהם עברו שיעור בסטטיסטיקה, הם יודעים שהמבחנים הכי רגישים להבדל בין ההשערות (בעלי עוצמה מקסימלית) הם מבחני יחס הנראות....

(a) מהו סטטיסטי המבחן ומהו הכיוון של איזור הדחייה?

(b) מהו מספר האתרים המקסימלי שבו יחליטו שאבנר 1 צודק אם הם מוכנים לחיות עם סיכוי של 20% שיגיעו למסקנה הלא נכונה?

i. לאיזו החלטה יגיעו אם האתר השני בו ביקרו לא תמך בפיירפוקס?

ii. מהי העוצמה של מבחן זה?

(c) מהו מספר האתרים המקסימלי שבו יחליטו שאבנר 1 צודק אם הם מוכנים לחיות עם סיכוי של 5% שיגיעו למסקנה הלא נכונה?

i. לאיזו החלטה יגיעו אם האתר השני בו ביקרו לא תמך בפיירפוקס?

ii. מהי העוצמה של מבחן זה?

(d) בנה מבחן יחס נראות למקרה כללי של $H_0: X \sim G(p_0)$ מול $H_1: X \sim G(p_1)$.

i. מהו הכיוון של איזור הדחייה?

ii. מהו סף ההחלטה (הערך הקריטי) אם $p_1 < p_0$ והסיכוי המותר לטעות הוא α ?

(e) הפעם אבנר 0 טוען ששיעור האתרים שלא תומכים בפיירפוקס הוא 20%, ואבנר 1 טוען שהשיעור הוא 80%.

(f) מהו מספר האתרים המקסימלי שבו יחליטו שאבנר 1 צודק אם הם מוכנים לחיות עם סיכוי של 5% שיגיעו למסקנה הלא נכונה?

i. לאיזו החלטה יגיעו אם האתר השני בו ביקרו לא תמך בפיירפוקס?

ii. מהי העוצמה של מבחן זה?

(g) מהו מספר האתרים המקסימלי שבו יחליטו שאבנר 1 צודק אם הם מוכנים לחיות עם סיכוי של 20% שיגיעו למסקנה הלא נכונה?

i. לאיזו החלטה יגיעו אם האתר השני בו ביקרו לא תמך בפיירפוקס?

ii. מהי העוצמה של מבחן זה?

iii. האם השינוי באיזור הדחייה ובעוצמה של המבחן מפתיע לאור ההבדלים בהשערות?

5. מספר השאלות בשרתי האוניברסיטה מתפלג פואסונית.

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

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

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

מבוא לסטטיסטיקה למדמ"ח- תשס"ט

(b) בנו את יחס הנראות עבור מקרי כללי של מדגם של מספר השאילות ב n שניות שנבחרו באקראי.

i. האם יחס הנראות עולה או יורד ככל **שסכום** השאילות עולה?

ii. מהו כיוון הדחייה של המבחן? האם התוצאה עקבית עם האינטואיציה שלכם?

(c) אם הוחלט להכריע בהתבסס על מספר השאילות שבוצעו בארבע שניות כלשהן שנבחרו באקראי ($n=4$). האם מבחן המבוסס על סטטיסטי זה יהיה בעל עוצמה מקסימלית (לכל רמת מובהקות)?

(d) ענו על סעיף c בהנחה והתפלגות השאילות בשניה אינה פואסונית אך יחס הנראות עדיין **מונוטוני** במספר השאילות? (אין צורך לחשב ממש)

(e) ענו על סעיף c בהנחה והתפלגות השאילות בשניה אינה פואסונית ויחס הנראות אינו **מונוטוני** במספר השאילות? (אין צורך לחשב ממש)

(f) מהי העוצמה של מבחן עוצמה מקסימלית ברמת מובהקות 0.05? תזכורת- זוהי שאלה שפתרתם בשיעורי הבית ואיזור הדחייה המתקבל הוא $\{0, \dots, 3\}$.