

סטטיסטיקה למדעי המחשב – תרגיל בית שבוע 8

שאלה 1 (25 נקודות, 5 נקודות לכל סעיף)

נערכו שני מבחנים סטטיסטיים חד-צדדיים ברמת מובהקות 0.05 לבדיקת ההשערה $H_0: \mu = \mu_0$ מול $H_A: \mu > \mu_0$, מהתפלגות נורמלית עם שונות ידועה. בשני המבחנים משתמשים בממוצע בתור סטטיסטי המבחן, ובאיזור הדחייה ה"אינטואיטיבי" של דחייה עבור ערכים גבוהים של הסטטיסטי. במבחן הראשון נלקח מדגם של 20 תצפיות, ובשני נלקח מדגם של 40 תצפיות מאותה אוכלוסייה.

א. לאיזה מהמבחנים הסתברות גבוהה יותר לטעות מסוג שני?

נתון שממוצע התצפיות היה זהה בשני המדגמים.

במידה וידוע שבמבחן הראשון (זה עם ה-20 תצפיות) דחו את H_0 :

ב. האם ניתן לדעת מה תהיה המסקנה במבחן השני?

ג. אם במקום מבחנים חד-צדדיים, היינו מבצעים את המבחנים עבור השערה

האלטרנטיבית דו-צדדית, $H_A: \mu \neq \mu_0$, האם המסקנה במבחן הראשון הייתה עשויה להשתנות?

במידה וידוע שבמבחן הראשון (זה עם ה-20 תצפיות) לא דחו את H_0 :

ד. האם ניתן לדעת מה תהיה המסקנה במבחן השני?

ה. אם במקום מבחנים חד-צדדיים, היינו מבצעים את המבחנים עבור השערה

האלטרנטיבית דו-צדדית, $H_A: \mu \neq \mu_0$, האם המסקנה במבחן הראשון הייתה עשויה להשתנות?

שאלה 2 (25 נקודות, 5 נקודות לכל סעיף)

אתנול (Ethanol) ומתנול (Methanol) הם שני סוגים של אלכוהול. בעוד כמות קטנה של אתנול גורמת לבני אדם להיות שמחים לכמה שעות ולקום יום אחרי זה עם כאב ראש, אותה כמות של מתנול גורמת לעיוורון.

א' רצה להשתכר והחליט לקנות מיכל של אלכוהול מסוחר מפוקפק שהבטיח שזה אתנול. א' יודע שמתנול זול יותר לייצור, ולכן סוחרים מפוקפקים נוטים לפעמים למכור מתנול במסווה של אתנול. מאחר ו-א' לא מעוניין להתעוור, הוא החליט לעשות ניסוי כדי לבדוק את אמינותו של הסוחר. הוא לקח 6 דגימות קטנות מהמיכל באופן בלתי תלוי, חימם אותן עד לרתיחה ומדד את הטמפרטורה בעת הרתיחה. להלן התוצאות שקיבל (במעלות צלזיוס):

68.02, 65.71, 69.07, 74.35, 79.74, 72.35

ניתן להניח שהתצפיות מגיעות מהתפלגות נורמלית, עם שונות ידועה של 120.

ידוע שטמפרטורת הרתיחה של אתנול היא 78.5 מעלות צלזיוס ואילו טמפרטורת הרתיחה של מתנול היא 64.5.

- א. הגדירו את השערות המבחן, נמקו את שיקולכם בבחירת השערת האפס.
- ב. הגדירו את סטטיסטי המבחן, ומצאו את התפלגותו תחת השערת האפס ותחת ההשערה האלטרנטיבית.
- ג. מצאו את אזור הדחייה של המבחן ברמה 0.05, ואת העוצמה של המבחן.
- ד. על סמך המדגם שהתקבל והמבחן שביצעתם, האם הייתם ממליצים ל א' לשתות את האלכוהול?
- ה. כיצד המלצתכם הייתה משתנה במידה והייתם מחליפים בין השערת האפס לאלטרנטיבה כפי שהגדרתם בסעיף א'? בצעו את המבחן המתאים והסבירו במילים שלכם האם ומדוע אתם עדיין חושבים שנכון לבצע את המבחן שהוגדר בסעיף א'.

שאלה 3 (25 נקודות, 5 נקודות לכל סעיף)

בעל חנות מזמין אתכם על מנת לעזור לו בבעיה. בחוזה שלו עם חברת התפוזים הוא הסכים שכל ארגז שהוא מקבל יכיל בדיוק 20% תפוזים רקובים. התפוזים מגיעים בארגזים של 1000 תפוזים כל אחד. בעל החנות רוצה לדגום 100 תפוזים עם החזרה מתוך הארגז על מנת שיוכל לקבוע אם הארגז מכיל יותר 20% תפוזים רקובים ואז יוכל להחזיר את הארגז תמורת פיצוי כספי.

- א. נסחו במילים את השערת האפס והשערת המחקר. הגדירו פרמטר מתאים ונסחו את ההשערות באמצעות פרמטר זה.
- ב. השתמשו בכלים שלמדתם עד כה על מנת להגדיר את המבחן - מהו סטטיסטי המבחן ומהו איזור הדחיה ברמת מבחן של 5%? הסבירו באילו קרובים השתמשתם במידת הצורך.
- ג. מדוע חשוב שבעל החנות יבצע דגימה עם החזרה?
- ד. הסבירו לבעל החנות את המשמעות של טעות מסוג ראשון ושני.
- ה. אם כל הארגזים מכילים 25% תפוזים רקובים, מה הסיכוי לטעות סוג שני?
- ו. (**בונוס 5 נקודות**) בעל החנות מעוניין שבמידה ו-25% תפוזים רקובים, הסיכוי לטעות מסוג שני לא יעלה על 0.2, וזאת מבלי לפגוע ברמת המבחן. מה הייתם מציעים לו על מנת לעמוד בתנאי זה?

שאלה 4 (25 נקודות, 5 נקודות לכל סעיף)

נתון מבחן סטטיסטי חד צדדי ברמה α לבדיקת הגובה של תלמידים. השערת האפס היא שהתוחלת שווה ל-175, והאלטרנטיבה היא שהתוחלת שווה ל-170. ידוע שהנתונים מתפלגים נורמלית עם שונות 100.

- א. עבור $\alpha = 0.05$, $n = 25$, הגדירו את סטטיסטי המבחן ומצאו על ידי חישוב מדויק את איזורי הדחיה ואת הסיכוי לטעות מסוג 2.
- ב. הראו שהסיכוי לטעות מסוג 1 והסיכוי לטעות מסוג 2 אכן שווים לערכים הצפויים מהסעיף הקודם באמצעות סימולציה בפייתון. על הסימולציה לכלול הגרלת מדגמים מקריים מתוך ההתפלגות המתאימה, ביצוע המבחן ביחס לאותם המדגמים, ומציאת שיעור הדחייה של השערת האפס.
- ג. התאימו את המבחן למצב שבו השונות אינה ידועה. האם במצב זה ניתן לחשב את איזורי הדחיה באופן מדויק? האם ניתן לחשב את הסיכוי לטעות מסוג 2 באופן מדויק באותו אופן שביצעתם בסעיף א'? נמקו.
- ד. בצעו סימולציה על מנת למצוא באופן אמפירי את הסיכויים לטעות מסוג 1 וטעות סוג 2 עבור המבחן הנוכחי (כלומר, תחת ההנחה שהשונות לא ידועה לחוקר, אבל השונות בפועל היא 100). הסבירו את התוצאות.
- ה. נכון או לא נכון: "בכל מדגם שעבורו דחינו את השערת האפס במבחן עם שונות לא ידועה, בהכרח היינו דוחים את השערת האפס גם במבחן עם שונות ידועה". הוכיחו אם נכון או הציגו דוגמה נגדית.