

# שיטות מחקר בקוגניציה (6132)

## תרגיל 1

4 באפריל 2022

### 1 נהלי עבודה והגשה

יש להגיש את התרגיל בתיבת ההגשה הייעודית במודל, עד לתאריך 18.4 בשעה 23:59.

יש להעלות לתיבת ההגשה קובץ בפורמט PDF ובו תשובות מילוליות לכל השאלות בתרגיל, בצירוף הגרפים הנדרשים. בנוסף, יש להעלות לתיבת ההגשה קובץ R שמכיל את כל הקוד הנדרש בתרגיל. על קובץ הקוד לרוץ ללא שגיאות ולהפיק את כל התוצרים (לרבות הגרפים) שמוצגים במסמך ה-PDF. הקוד ייבדק ידנית, אנא הקפידו על תיעוד ועל קריאות. בגרפים, הקפידו על כותרות רלוונטיות לגרף ולכל אחד מהצירים.

שאלות על התרגיל ניתן לפרסם, גם באנונימיות, בפורום יחידה 1. להוראות נוספות, ראו מסמך נהלי הקורס. בהצלחה!

### 2 חלק א' - תכנותי

הציון על חלק זה מהווה 75% מציון התרגיל.

#### שאלה 1

א. עבור  $N = 3$ , דגמו  $N$  תצפיות מהתפלגות נורמלית עם ממוצע 70 וסטיית תקן 10. הציגו היסטוגרמה של המדגם.

דווחו את ממוצע המדגם ואת סטיית התקן שלו.

ב. חזרו על סעיף א' לכל  $N \in \{3, 10, 50, 100, 250, 500\}$ .

ג. בהתבסס על תוצאות סעיף ב', הציגו גרף של ממוצע המדגם כפונקציה של  $N$ . הוסיפו לגרף קו אופקי שמציין את ממוצע האוכלוסייה.

ד. בהתבסס על תוצאות סעיף ב', הציגו בנפרד גרף של סטיית התקן במדגם כפונקציה של  $N$ . הוסיפו לגרף קו אופקי שמציין את סטיית התקן של האוכלוסייה.

ה. הריצו שוב את הקוד לסעיפים א'-ד' מספר פעמים (אין צורך להציג את התוצאות). כיצד משתנים ממוצע המדגם וסטיית התקן שלו כתלות בגודל המדגם וביחס לפרמטרים המקבילים להם באוכלוסייה? התייחסו לתכונת העקביות (קונסיסטנטיות).

#### שאלה 2

בשאלה זו נכיר פרמטר שטרם הכרנו, נגדיר סטטיסטי, ונשערך את התפלגות הדגימה של הסטטיסטי כדי לבדוק האם הוא עונה על הקריטריונים של אומד טוב לפרמטר.

**ציוד** (skewness) של התפלגות הוא מדד לחוסר הסימטריה שלה (ר' ויקיפדיה). עבור משתנה מקרי  $X$  עם ממוצע  $\mu$  וסטיית תקן  $\sigma$ , ציוד מוגדר באופן הבא:

$$\text{Skew}[X] = \mathbb{E} \left[ \left( \frac{X - \mu}{\sigma} \right)^3 \right]$$

כאשר  $\mathbb{E}$  הוא אופרטור התוחלת.

בהינתן מדגם  $\{x_1, \dots, x_n\}$ , נגדיר את הסטטיסטי הבא:

$$g_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3}$$

כאשר  $\bar{x}$  הוא ממוצע המדגם ו- $s$  סטיית התקן במדגם (ללא תיקון Bessel):

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

א. כתבו פונקציה שמקבלת מדגם ומחזירה את ערך הסטטיסטי  $g_1$  עבורו.

ב. עבור  $N = 50$ , שלפו 10,000 מדגמים מקריים בגודל  $N$  מהתפלגות נורמלית עם ממוצע 70 וסטיית תקן 10. חשבו

את ערך  $g_1$  עבור כל אחד מן המדגמים. 10,000 ערכי  $g_1$  שהתקבלו הם השערוך להתפלגות הדגימה של  $g_1$ . חשבו

את הממוצע ואת השונות של התפלגות הדגימה המשוערכת של  $g_1$ .

ג. חזרו על סעיף ב' עבור כל ערכי  $N$  בין 50 ל-500 בקפיצות של 50.

ד. בהתבסס על תוצאות סעיף ג', הציגו גרף של ממוצע התפלגות הדגימה של  $g_1$  כתלות ב- $N$ . הוסיפו לגרף קו אופקי

שמציין את ציוד האוכלוסייה, 0.

ה. בהתבסס על תוצאות סעיף ג', הציגו בנפרד גרף של שונות התפלגות הדגימה של  $g_1$  כתלות ב- $N$ .

ו. האם  $g_1$  הוא אומד מוטה לציוד האוכלוסייה? האם הוא אומד עקבי (קונסיסטנטי)?

### שאלה 3

בכיתה ראינו כיצד William Gosset ביצע סימולציה של התפלגות הדגימה של סטטיסטי  $t$  בכלים שהיו זמינים ב-1908. כעת נבצע סימולציה דומה עם כלים הזמינים ב-2022.

א. בחרו התפלגות אוכלוסייה כלשהי (בעלת שונות סופית) וגודל מדגם  $n$  כלשהו.

ב. שלפו 10,000 מדגמים מקריים בגודל  $n$  מהאוכלוסייה שבחרתם וחשבו לכל אחד מהם את ערך הסטטיסטי  $t$ :

$$t = \frac{\bar{x} - \mu}{\hat{\sigma} / \sqrt{n}}$$

כאשר  $\bar{x}$  הוא ממוצע המדגם ו- $\hat{\sigma}$  הוא האומד המתוקן לסטיית התקן:

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

ג. הציגו היסטוגרמה של 10,000 ערכי  $t$  שחישבתם.

ד. הציגו גרף של התפלגות  $t$  התיאורטית המתאימה לסימולציה שביצעתם.

ה. מה תפקידה של התפלגות  $t$  בתהליך בדיקת ההשערות (ספציפית, במבחן  $t$  למדגם יחיד)?

## שאלה 4

תרשים quantile-quantile (Q-Q plot) הוא דרך גרפית להשוות בין התפלגויות או מדגמים ע"י השוואת ה-quantiles שלהם (ר' ויקיפדיה). אנחנו נתבונן במקרה הפרטי של השוואה בין מדגם לבין התפלגות נורמלית סטנדרטית, שמשמש לעתים לבחינה איכותנית של השאלה האם מדגם מגיע מהתפלגות נורמלית. הגרף נבנה כך שכל נקודה תואמת ל-quantile מסוים, ערך ה-x שלה הוא ה-quantile הזה בהתפלגות הנורמלית הסטנדרטית וערך ה-y הוא ה-quantile הזה במדגם. ככל שהמדגם דומה יותר להתפלגות הנורמלית הסטנדרטית, כך ה-Q-Q plot המתקבל דומה יותר לישר הזהות  $y = x$ . ב-R, הפונקציה qqnorm מקבלת מדגם ומציגה את ה-Q-Q plot המתאים להשוואתו להתפלגות נורמלית סטנדרטית.

א. השתמשו בפונקציה qqnorm כדי להשוות בין התפלגות  $t$  מהסימולציה שערכתם בשאלה 5 לבין ההתפלגות הנורמלית הסטנדרטית. כלומר, אם 10,000 ערכי ה- $t$  מהסימולציה נמצאים בווקטור בשם `t_vec`, הריצו את השורה הבאה:

```
qqnorm(t_vec)
```

ב. הוסיפו לגרף את ישר הזהות  $y = x$  בעזרת הפונקציה `abline`.

ג. באילו אזורים של הגרף ניכרת החריגה המשמעותית ביותר מישר הזהות? אילו ממאפייניה של התפלגות  $t$  הם הגורמים לחריגה באזורים אלו?

ד. **בונוס:** הגדירו את "הנחת הנורמליות", כפי שנלמדה בהקשר של מבחני  $Z$  ו- $t$ .

ה. **בונוס:** במצבים אמיתיים במחקר, אין לנו ידע על התפלגות האוכלוסייה: יש בידינו מדגם אחד בעקבות ניסוי שביצענו. במקרה שנרצה לבצע על המדגם מבחן  $t$ , האם הייתם ממליצים להשתמש ב-Q-Q plot כדי לוודא את קיומה של הנחת הנורמליות? מדוע?

## שאלה 5

א. הגדירו מהו רווח סמך ברמת בטחון 95% עבור פרמטר כלשהו של האוכלוסייה. כיצד נפרש את טווח הערכים הזה?

ב. עבור  $N = 3$ ,  $\mu = 70$  ו- $\sigma = 2$ , דגמו  $N$  תצפיות מהתפלגות נורמלית עם ממוצע  $\mu$  וסטטיית תקן  $\sigma$ . חשבו רווח סמך ברמת ביטחון של 95% עבור תוחלת האוכלוסייה.

ג. חזרו על סעיף ב' לכל  $N \in \{3, 10, 50, 100, 250, 500\}$ . הציגו גרף של רוחב רווח הסמך (ההפרש בין ערכי העליון לערכי התחתון) כפונקציה של  $N$ .

ד. חזרו על סעיף ב' עם  $N = 50$ , לכל  $\mu \in \{70, 100, 250, 500, 750\}$ . הציגו גרף של רוחב רווח הסמך כפונקציה של  $\mu$ .

ה. חזרו על סעיף ב' עם  $N = 50$ , לכל  $\sigma \in \{2, 4, 6, 8, 10\}$ . הציגו גרף של רוחב רווח הסמך כפונקציה של  $\sigma$ .

ו. כיצד תלוי רוחב רווח הסמך בכל אחד מהפרמטרים  $N, \mu, \sigma$ ?

## שאלה 6

הזנחת צד (Neglect) היא תסמונת הנגרמת בדרך כלל מפגיעה באונה הפריאטלית הימנית. התסמונת מתאפיינת בכך שהמושפעים ממנה מתקשים בזיהוי, בהפניית קשב ובדיווח על גירויים המופיעים בצד הקונטרה-לטרלי להמיספרה הפגועה, ללא פגיעה ביכולת העיבוד החושית או הפעילות המוטורית הגורמת לכך. תסמונת זו מזוהה בין היתר על ידי מבחנים התנהגותיים שונים שמטרתם לכמת ולאבחן את יכולת הפניית הקשב ואת מידת המודעות של החולה לצד המוזנח. אחד המבחנים הוותיקים והמוכרים לבדיקה ולכמות של מידת ההזנחה הוא ה-Bisection Task, בו על הנבדקים לסמן את מרכזו של קו המופיע על דף. נבדקים בעלי הזנחת צד יסמנו את מרכזו של הקו בהטיה, לעתים ניכרת מאוד, לצד המנוגד

לצד המוזנח (בדרך כלל, לצד ימין). מנגד, ידוע שגם לנבדקים ללא פגיעה נוירולוגית יש הטיה, אם כי קטנה יותר. לצורך<sup>4</sup> השאלה נניח שלנבדקים בריאים בגיל העמידה, הגיל בו נמצאים רוב החולים בהזנחת צד, יש הטיה לצד ימין שמתבטאת בסימון מרכז הקו בנקודה שהחלק היחסי של הקו משמאל (אורך הקו עד הנקודה חלקי אורך הקו הכולל) הוא כ-0.55. בקובץ neglect\_data.csv תמצאו נתונים של 25 נבדקות הסובלות מהזנחת צד, שביצעו bisection task, 15 חזרות (trials) כל אחת\*. בשאלה הבאה נבצע מבחן  $t$  למדגם יחיד על התוצאות, על מנת להשוותן לנבדקות בריאות. הנתונים בקובץ בנויים באופן הבא: אורכי הקווים שווים לכל נבדקת בין trials כדי להימנע מהשפעה של הביטוי מוטורית לאורך המטלה. כל שורה בקובץ מייצגת trial אחד של נבדקת אחת. העמודה הראשונה מכילה את מספר השורה בקובץ, העמודה השנייה מכילה את מספר הנבדקת, העמודה השלישית את מספר הטרייל, העמודה הרביעית את אורך הקו המלא בטרייל הזה (במילימטרים), העמודה החמישית את המיקום בו הנבדקת סימנה את מרכז הקו (מרחק במילימטרים מגבולו השמאלי של הקו).

א. טענו את הנתונים מהקובץ ל-data frame ב-R. "נקו" את הדאטא ע"י השמטה של עמודות מיותרות, מתן שמות בעלי משמעות לעמודות וכו'. השמיטו שורות עם ערכים חסרים (NA).

ב. הוסיפו עמודה שבה לכל טרייל יופיע האורך היחסי שמשמאל לנקודה שסומנה מתוך אורכו המלא של הקו. להלן, משתנה זה ייקרא "חלק שמאלי".

ג. בחרו ארבע נבדקות. דווחו רווח סמך ברמת בטחון של 95% לתוחלת החלק השמאלי עבור כל אחת (ארבעה רווחי סמך סה"כ).

שימו לב: בכיתה למדנו חישוב רווח סמך רק במקרים שבהם שונות האוכלוסייה הייתה ידועה. שאלה זו מדמה מצב מחקרי מציאותי יותר, וסטיית התקן של האוכלוסייה אינה ידועה. נדרשת התאמה קלה באופן החישוב של רווח הסמך: יש להשתמש באומד המתקן לסטיית התקן של האוכלוסייה. כיוון שמשתמשים באומד זה, רווח הסמך ייקבע ע"י quantiles של התפלגות  $t$  ולא של התפלגות נורמלית סטנדרטית.

ד. צרו וקטור חדש ושמרו בו את ממוצע החלק השמאלי, על פני כל הטריילים, עבור כל נבדקת. הווקטור צריך להכיל ערך אחד לכל נבדקת.

ה. חשבו רווח סמך ברמת בטחון של 95% לתוחלת החלק שמאלי על פני כל הנבדקות. האם התוצאות דומות לאלו של סעיף ג'؟ הסבירו מדוע.

ו. חשבו סטטיסטי  $t$  המתאים להשערת האפס שתוחלת החלק השמאלי של הנבדקות שנדגמו היא 0.55 (כלומר, שהן אינן לוקות בנגלקט).

ז. השתמשו בפונקציה pt כדי לחשב  $p$ -value חד-זנבי ימני (כלומר, שהמדגם מגיע מאוכלוסייה עם תוחלת גבוהה יותר) עבור השערת האפס הנ"ל.

ח. השתמשו בפונקציה t.test כדי לערוך מבחן  $t$  חד-זנבי ימני (כלומר, שהמדגם מגיע מאוכלוסייה עם תוחלת גבוהה יותר) עבור השערת האפס הנ"ל. האם תוצאות המבחן זהות לאלו שהתקבלו בסעיף ז' אם לא, מדוע?

ט. אם התקבלה תוצאה מובהקת בסעיף ח', ביחס לסף מובהקות  $\alpha = 0.05$ , חשבו ודווחו את גודל האפקט.

\*הערה: הנתונים בשאלה זו אינם לקוחים מתוצאות ניסוי אמיתי, הם נוצרו על-ידינו לטובת התרגיל.

### 3 חלק ב' - מתודולוגי

הציון על חלק זה מהווה 25% מציון התרגיל.

**השאלות מתייחסות למאמר המצורף:**

Chang, D., Song, D., Zhang, J., Shang, Y., Ge, Q., & Wang, Z. (2018). Caffeine caused a widespread increase of resting brain entropy. Scientific reports, 8(1), 1-7.

לפי ארגון הקפה העולמי (ה-ICO), אוכלוסיית העולם שותה מעל למיליארד וחצי כוסות קפה ביום. סגולותיו של הקפה והשפעותיו על יכולות ריכוז ושמירה על קשב ועוררות גבוהים קנו לקפה מקום של כבוד כאחד המשקאות הפופולאריים ביותר ברחבי כדור הארץ. מחקרים שונים שבחנו את ההשפעה של קפה מצאו כי הוא מביא לשיפור במטלות קוגניטיביות שונות ומגדיל את הגמישות הקוגניטיבית, את היכולת לבצע פעילות לאורך זמן ואת יעילותן של פונקציות שונות של קשב. אך האם ניתן לראות ביטוי להשפעה של שתיית קפה על הפעילות המוחית?

המחקר הנוכחי בודק את ההשפעה של צריכת קפאין על המח באמצעות מדד האנטרופיה. אנטרופיה הינו מונח השאול במקורו מתחום התרמודינמיקה ומאוחר יותר מתחום המכניקה הסטטיסטית. מדד זה, בהקשר של פעילות מערכת מוחית, התפתח בשנים האחרונות למדידת תנודות בפעילות המוחית במצב מנוחה אשר יכולות להעיד על מאפיינים שונים של המח הנבדק.

**קראו את המאמר המצורף וענו על השאלות הבאות.** שימו לב: אין צורך להבין את המאפיינים הטכניים של סריקת fMRI על מנת לענות על השאלות, וכן אין צורך להבין בדיוק כיצד מודדים אנטרופיה. התמקדו בעיקר בפרקי השיטה והתוצאות במאמר.

א. האם מערך המחקר הינו מתאמי או ניסויי? הסבירו.

ב. האם המחקר הינו קונפירמטורי או אקספלורטורי? הסבירו.

ג. מנו את המשתנים התלויים והבלתי תלויים במחקר. אין להתייחס לאזורים מוחיים או לוקסלים כמשתנה.

ד. מדוע עשו החוקרים שימוש בשתי קבוצות שונות של נבדקים? איזו תרומה מתודולוגית יכולה להיות לכך לדעתכם?

ה. בחלק שכותרתו "Design" מדווחים כותבי המאמר על בקורות שונות (הן סטטיסטיות והן מתודולוגיות) אשר בהן

נעשה שימוש על מנת לשפר את תוקף המחקר. בחרו שלוש בקורות שלדעתכם משפרות את יכולת הסקת המסקנות

מהתוצאות והסבירו כיצד הן עושות זאת.