

תרגיל 5- התפלגות נורמלית ודגימה

מעבר לעזרה שבתוכנה עצמה, מומלץ להיעזר באתר: <http://wiki.r-project.org/rwiki/doku.php>
הדרכה בעברית ניתן למצוא באתר הקורס.
ניתן ורצוי להתשמש בפורום הקורס להתייעצות.
יש לצרף את הקוד ששימש לפתרון אך אין הוא תחליף לתשובה סופית.
סעיפים בדרגת קושי גבוהה יותר סומנו בכוכבית(*). כדאי להשתדל לענות עליהם אבל אין הכרח.

ההתפלגות הנורמלית

1. חשבו את ההסתברויות הבאות באמצעות הטבלה הנורמלית
 $\Phi(-6), \Phi(-5), \Phi(0), \Phi(-1.5), \Phi(-1.5), \Phi(2.22), \Phi(2.2), \Phi(2)$

- (a) מקובל לסמן משתנה מקרי נורמלי סטנדרטי על ידי האות Z : $Z \sim N(\mu=0, \sigma=1)$.
חשבו את ההסתברויות הבאות:
 $P(-2.22 \leq Z \leq 1.55), P(-1 \leq Z \leq -0.5), P(2 \leq Z \leq 3),$
 $P(Z \geq -2.5), P(Z \geq -2.33), P(3 \leq Z)$
- (b) חשבו את האחוזונים הבאים של ההתפלגות הנורמלית הסטנדרטית:
 $Z_{0.25}, Z_{0.75}, Z_{0.95}, Z_{0.975}, Z_{0.9}, Z_{0.025}$

2. X מתפלג נורמלית עם תוחלת 10 ושונות 100 $X \sim N(\mu=10, \sigma^2=100)$

- (a) מה הסיכוי שההתפלגות הבאה של המשתנה תהיה שלילית? $P(X \leq 0)$
- (b) מהו האחוזון ה-15.86553% של התפלגות זו?
כלומר מהו x כך שיקיים $P(X \leq x) = 0.1586553$?
- (c) מהו האחוזון ה-50 של התפלגות זו? האחוזון האפס? האחוזון ה-100?
- (d) מהו האחוזון ה-10 של התפלגות זו?

3. מקובל לקרב התפלגות גבהים באוכלוסייה על ידי התפלגות נורמלית. נסו להצדיק טיעון זה תוך שימוש במשפט הגבול המרכזי.

4. תנו שלוש דוגמאות לסוגי נתונים אותם לא סביר לתאר באמצעות התפלגות נורמלית (כי אין לצפות שתהיה סימטריה לדוגמה, כי לא כל תוצאה היא אפשרית וכו...)

5. הציונים בבחינה מתפלגים נורמלית עם ממוצע 60 ועם סטיית תקן 10.

- (a) מה הסיכוי לעבור את הבחינה (ציון עובר: 60)?
- (b) האם יכולתם לחשב את הסעיף הקודם רק עם הידיעה שההתפלגות סימטרית (אבל לא בהכרח נורמלית)? האם יכולתם לחסום את ההסתברות?
- (c) מה הסיכוי לקבל לפחות 80? לפחות 90?
- (d) מה הסיכוי לקבל לכל היותר 20 בבחינה?
- (e) מהו העשירון העליון של הציונים? ומהו העשירון התחתון?
- (f) מתוך 100 תלמידים כמה עברו את הבחינה וכמה קבלו מעל 70?
- (g) נניח שנתנו פקטור של 10% תוספת לכל הציונים. ענו שוב על כל הסעיפים והשוו את התוצאות לאלו שהתקבלו קודם לכן.

6. הטמפרטורות במדינת ישראל בשנת 2005 התפלגו בקירוב נורמלית עם תוחלת 15 וסטיית תקן שש $T \sim N(\mu=15, \sigma=6)$.

- (a) מה הייתה הטמפרטורה הכי גבוהה ב 30% הימים הכי קרים בשנה? (האחוזון ה 30)
- (b) באיזה אחוז מהימים היה חם יותר מעשרים מעלות?
- (c) הקשר בין הטמפרטורה בצלזיוס ופרנהייט נתון על ידי הנוסחה: $T_c = 5/9 \cdot (T_f - 32)$. מהי הייתה תוחלת הטמפרטורה בפרנהייט? האם צריך להכיר את תכונות ההתפלגות הנורמלית לשם הפתרון?
- (d) מה הייתה שונות הטמפרטורה (בפרנהייט)? האם צריך להכיר את תכונות ההתפלגות הנורמלית לשם הפתרון?
- (e) מה הייתה הטמפרטורה הכי גבוהה ב 30% הימים הכי קרים בשנה (בפרנהייט)? האם צריך להכיר את תכונות ההתפלגות הנורמלית לשם הפתרון?
- (f) מה הייתה הטמפרטורה הכי גבוהה ב 44% הימים הכי קרים בשנה (בפרנהייט)?
- (g) מדוע לפתרון שאלה זו צריך טבלה ולקודמת לא?

7. X מ"מ שמתפלג $B(n, p)$.

- (a) חשב את ההסתברות **המדוייקת** למאורע $P(X=n \cdot p); n=10, p=1/10$.
- (b) השתמש בצפיפות הנורמלית לקירוב הסיכוי למאורע לעיל.
- (c) חשב את ההסתברות **המדוייקת** למאורע $P(X=n \cdot p); n=100, p=1/4$.
- (d) השתמש בצפיפות הנורמלית לקירוב הסיכוי למאורע לעיל.
- (e) מה תוכל לאמר על טיב הקירוב בשני המקרים השונים?

8. נניח כי הגבהים במדינת ישראל ניתנים לקירוב על ידי התפלגות נורמלית עם תוחלת

μ ועם סטיית תקן σ .

- (a) מה הסיכוי שהאדם הבא שתפגשו ברחוב יהיה **גבוה** בסטיית תקן אחת לפחות מהגובה הטיפוסי (התוחלת)? ונמוך בסטיית תקן אחת לפחות מהגובה הטיפוסי?
- (b) מה הסיכוי שהאדם הבא שתפגשו יהיה **גבוה או נמוך** בסטיית תקן אחת לפחות מהגובה הטיפוסי?
- (c) מה הסיכוי שהאדם הבא שתפגשו יהיה **שונה** בשתי סטיות תקן מהגובה הטיפוסי? ובשלוש סטיות תקן?
- (d) אוכלוסיית שוודיה ידועה כגבוהה יותר מישראל. את התפלגות הגבהים בשוודיה ניתן לתאר בקירוב על ידי התפלגות נורמלית עם תוחלת $\mu^* > \mu$ ועם סטיית תקן $\sigma^* < \sigma$. מה הסיכוי שהשוודי הבא שתפגשו ברחוב (בשוודיה) יהיה גבוה בסטיית תקן אחת לפחות מהתוחלת בשוודיה?
- (e) לאור התובנות שלכם עד כה על משמעות סטיית התקן, האם סביר שסטיית התקן של הגבהים בישראל היא חצי מטר? נמקו כמובן.
- (f) (*) במבוא להסתברות למדתם שניתן לחסום את הסיכוי שמשתנה מקרי רחוק מהתוחלת על ידי שיון צ'בישב. מה ההבדל בין הדרכים ואיזו תעדיפו בתרחיש שבשאלה?

9. (*) סעיף זה נועד לתת תחושה של **משפט הגבול המרכזי** באמצעות סימולציה:

- (a) הגרילו 100 משתנים מהתפלגות אחידה רציפה² בקטע $[0,1]$ באמצעות הפקודה `runif`. וודאו כי כל הערכים סבירים באותה המידה באמצעות היסטוגרמה או תרשים צפיפות.
- (b) חיזרו על הסעיף הקודם 1000 פעמים ובכל חזרה סיכמו את 100 המשתנים שהגרלתם. כעת יש בידיכם 1000 תצפיות על התפלגות משתנה מקרי חדש שנוצר על ידי סכום של 100 משתנים מקריים אחרים. התבוננו בהתפלגות עשרת המשתנים הללו באמצעות היסטוגרמה או תרשים צפיפות. האם כל הערכים עדיין סבירים באותה המידה?
- (c) בסעיף הקודם ראיתם כי סכום משתנים מקריים לא מאותה משפחת ההתפלגויות כמו כל מחובר בודד. מה ייגרום להתפלגות סכום משתנים (מהתפלגות אחידה במקרה זה) להיות יותר נורמלית? יותר משתנים אחידים שנכנסים לסכום או יותר סכומים (באותו אורך כל אחד)?

דגימה

1. הניחו שאלו הם מספר הצימוקים בארבע עוגות במאפיית "פחמימות ובניו":

עוגה	1	2	3	4
צימוקים	54	60	44	56

(a) האם ארבעת עוגות אלו הן **אוכלוסייה** או **מדגם** או **מדגם מקרי**³ בתרחישים הבאים:

- i. אלו העוגות היחידות שיוצרו אי פעם במאפייה.
- ii. אלו עוגות שבחרתם באקראי בביקור האחרון שלכם במאפייה כמייצגות את מספר הצימוקים בעוגות אותו היום.
- iii. אלו עוגות שבחרתם באקראי בביקור האחרון שלכם במאפייה כמייצגות את מספר הצימוקים בכלל העוגות במאפייה.
- iv. אלו עוגות אקראיות שנאספו במועדים אקראיים מאז שהמאפייה נפתחה כמייצגות את מספר הצימוקים בכלל העוגות במאפייה.
- v. אלו העוגות היחידות בטעם כבש שנבחרו לייצג את מספר הצימוקים בכלל העוגות במאפייה.
- vi. אלו העוגות היחידות בטעם כבש שנבחרו לייצג את מספר הצימוקים בכלל העוגות בטעם כבש במאפייה.

בהמשך השאלה, הניחו כי אלו העוגות היחידות שיוצרו אי פעם במאפייה.

(b) מהי תוחלת מספר הצימוקים בעוגה? מהי שונות מספר הצימוקים?

(c) ידיעת מספר הצימוקים בכל עוגה היא מצב לא סביר במציאות. ספירת הצימוקים היא תהליך ארוך שמשחית את העוגות ולכן יותר סביר שבשביל לדעת את תוחלת מספר הצימוקים נצטרך **לאמוד** אותו...

i. בכמה דרכים שונות אפשר לבחור שתי עוגות **שונות**? האם דגימה שכזו מקיימת את הגדרות הדגימה המקרית? (בפרט את השערת אי-התלות בין התצפיות).

ii. בכמה דרכים שונות אפשר לבחור שתי עוגות **שאינן בהכרח שונות** (כלומר מותר לבחור את אותה העוגה פעמיים)? האם דגימה שכזו מקיימת את הגדרות הדגימה המקרית?

iii. ענו על השאלות הבאות אם החלטנו לאמוד את מספר הצימוקים הטיפוסי (התוחלת) על ידי **ממוצע של שתי עוגות** שנדגמו **מקרית**:

A. אילו ערכים אפשריים יש לאומד? מה הסיכוי לכל ערך?

B. על בסיס כל התוצאות האפשריות והסיכויים שלהן מהסעיף הקודם, חשבו את **תוחלת האומד**? האם הוא חסר הטייה?

C. מהי השונות של האומד?

D. מהו ה MSE של האומד?

E. (*) הציעו מדד אלטרנטיבי (שאינו MSE) לאיכות של אומד וחשבו אותו על הממוצע כאומד לתוחלת באותם הנתונים.

³ זאת הפעם האחרונה בה נתעכב על ההבדל בין "מדגם" ל"מדגם מקרי". אזכיר שמדגם **מקרי** אנו דורשים שכל אחת ואחת מהדגימות מתפלגת כמו האוכלוסייה אותה היא נועדה לייצג. דרישה זו מכתובה כמובן אי תלות (אחרת הדגימה השנייה כבר לא תתפלג כמו הראשונה). בעתיד, המילה "מדגם" תתייחס רק **למדגם מקרי** אלא אם נאמר אחרת.

מבוא לסטטיסטיקה למדמ"ח- תשס"ט

i. (*) אולי בכלל עדיף להשתמש בממוצע גיאומטרי כאומד ולא בממוצע חשבוני. חשבו את ה MSE של הממוצע הגיאומטרי המבוסס על שתי תצפיות והשוו אותו לממוצע החשבוני. איזה אומד עדיף?

$$\text{geomean}(\vec{x}) = (x_1 \cdot \dots \cdot x_p)^{1/p} \quad \text{תזכורת- ממוצע גיאומטרי:}$$