מגישים: שיר השקוביץ &
אורי גביר

שיטות מחקר בקוגניציה (6177) - תרגיל 1

שאלה 1 (30 נק׳)

התוחלת (expected value) של משתנה מקרי מוגדרת כסכום משוקלל של כל התוצאות האפשריות לו וההסתברות להן. הממוצע (mean) של מדגם מתוך התפלגותו של המשתנה המקרי הוא סכום התצפיות חלקי מספר התצפיות.

הדוגמא האהובה על קורסים סטטיסטיים : נתונה קוביה לא הוגנת. ההסתברות שהקוביה תיפול על כל אחת מהפאות המציגות מספר זוגי הוא 0.2. ההסתברות שהיא תיפול על כל אחת מהפאות המציגה מספר אי זוגי הוא $\frac{2}{15}$. נגדיר משתנה מקרי x שהוא התוצאה של הטלה יחידה של הקוביה, $x \in \{1,2,3,4,5,6\}$.

1. מהי התוחלת של x?
2. בR, צרו וקטור הכולל את כל התוצאות האפשריות לx.
3. השתמשו באופציית העזרה בR (?) על מנת לקרוא על השימוש בפונקציה sample, שהכרנו בתרגול. כתבו : מה עושה הפונקציה, ומהו השימוש של הפרמטרים האופציונליים prob ו - replace?
4. הפונקציה ()set.seed מאפשרת ״לקבע״ את פעולתן של פונקציות המחזירות ערך אקראי, כך שהרצתן מחדש תגריל את אותם הערכים בכל פעם. לצורך התרגיל הנוכחי, הריצו את הפקודה set.seed(6177).
5. הגדירו וקטור נוסף הכולל את ההסתברות לקבלת הערכים השונים של x, והשתמשו בפונקציה sample על מנת לדגום מדגם של 15 הטלות של הקוביה הנתונה.
6. מהו הממוצע של המדגם? חשבו את הממוצע ידנית, ותארו את הקשר בין ערכי המדגם שדגמתם לבין ההתפלגות של x.
7. מה הקשר בין כפילת כל ערך אפשרי בהסתברותו בחישוב התוחלת, ובין סכימת הערכים של המדגם והחלוקה בn בחישוב הממוצע? חישבו שוב על הוקטור הנתון של המדגם שדגמתם.

שאלה 2 (30 נק׳)

בשנות ה70 של המאה שעברה התנהל ויכוח עז בקרב הקהילה המחקרית על טבע הדמיון האנושי, שעסק בשאלה האם הדמיון הוא ׳תמונתי׳ – האם כאשר נדמיין משהו, האובייקט ייוצג במוח באופן ויזואלי – או שמא מדובר באנלוגיה בלבד, ובפועל תוכן האובייקט המדומיין מיוצג במוח בדרכים שאינן ויזואליות (לדוגמא, כמידע סמנטי). הויכוח הוכרז כ׳הוכרע׳ על ידי פרסומם של שני מחקרים, שהראו שכאשר נבדקים/ות נדרשו לדמיין תמונה ואז להפעיל עליה פעולה, זמן התגובה שלהם/ן היה תלוי במידת

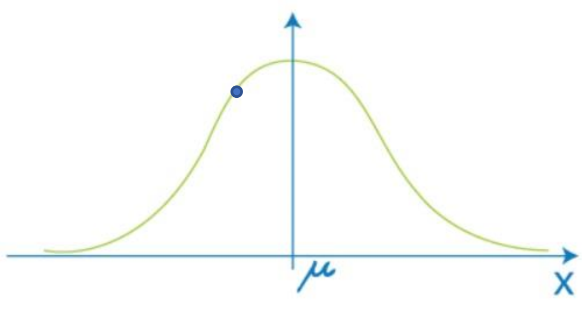
המניפולציה שהיה צורך להפעיל על התמונה, ושלכן, הסיקו, דמיון הוא תמונתי. בניסוי הראשון הנבדקים/ות נדרשו לסובב צורה תלת מימדית, ולבדוק האם גרסתה המסובבת תואמת או לא לגירוי שני. החוקרים מצאו שזמן התגובה של הנבדקים/ות היה ארוך יותר ככל שנדרשו לסובב את הגירוי בזווית גדולה יותר. בניסוי השני נדרשו הנבדקים/ות לדווח על פרט מסוים בתמונה ועל הקשר שלו לפרט אחר בתמונה. הנבדקים/ות למדו את התמונה, והיא לא הייתה זמינה בזמן המענה על השאלה. זמן התגובה של הנבדקים היה גדול יותר ככל שהפרט היה רחוק יותר מהפרט השני. (הקישורים לקריאה נוספת, אין צורך לקרוא את המאמרים כדי לענות על השאלה).

1. נסחו את שאלת המחקר באופן תאורטי (לפני אופרציונליזציה). מהם המשתנים התיאורטיים בשאלת המחקר – הבלתי תלוי והתלוי?
2. בכל אחד מהניסויים – מהו המשתנה התלוי האופרציונלי? מהו המשתנה הבלתי תלוי האופרציונלי?
3. האם מערך הניסוי הראשון הוא ניסויי או מתאמי? אם המערך ניסויי, האם המניפולציה בלבדית? אם המערך מתאמי, מה יכול להיות הסבר אלטרנטיבי לתוצאות?
4. האם מערך הניסוי השני הוא ניסויי או מתאמי?

שאלה 3 (40 נק׳)

משתנה מקרי המתפלג נורמלית הוא משתנה רציף (בשונה מבדיד). ניזכר משיעורי הסתברות, שבשונה מהמקרה של משתנה מקרי בדיד, בו ההסתברות לקבלת ערך מסוים נקבעת על ידי פונקציית הסתברות, ההסתברות שמשתנה רציף יקבל ערך מסוים מוגדרת על ידי פונקציית צפיפות (probability density function, PDF). מטבעו של משתנה רציף, ההסתברות לקבל ערך מסוים (לדוגמא, 3.145) שואפת לאפס בגלל אינסוף הערכים הקיימים בין כל שני ערכים שלמים. לכן, במקרה של משתנה רציף, נדבר על ההסתברות לקבל טווח ערכים. על מנת לחשב את ההסתברות שהמשתנה המקרי יקבל ערך שיפול בטווח נתון, נחשב את האינטגרל המסוים של הPDF בין הקצוות של הטווח הזה.

1. מה עלינו לדעת על מנת לחשב את ההסתברות שמשתנה מקרי נתון, המתפלג נורמלית, יהיה קטן או שווה לאפס?



2. נתון גרף המציג את עקומת הפעמון המוכרת של התפלגות נורמלית. עקומת הפעמון מציגה את ה PDF של ההתפלגות הנורמלית. מסומנת נקודה על הגרף של הפונקציה.
   a. מה משמעות הגובה של הנקודה?
   b. מה משמעות השטח שמתחת לפונקציה, משמאל לנקודה?
   c. מה משמעות השטח מתחת לפונקציה, מימין לנקודה?
   d. קראו על הפונקציה pnorm ותארו את שימושה.
   e. הניחו כי הגרף מתאר התפלגות נורמלית בעלת תוחלת 0 וסטיית תקן 1, וערך הנקודה המסומנת על ציר הx הוא -0.5. חשבו ודווחו את השטח שמשמאל לנקודה בעזרת הפונקציה pnorm.
3. כיצד ישתנה הגרף בעקבות שינוי של תוחלת המשתנה המקרי? כיצד ישתנה הגרף בעקבות שינוי של השונות של המשתנה המקרי?
4. נחזור לדוגמת הגובה בה דנו בתרגול. ידוע שגובה מתפלג נורמלית. קובי מאמץ גורה חדשה, בת שנה, ותוהה מאיזה סוג היא. ידוע שהגובה הממוצע של זן א׳ בגיל שנה הוא 100 ס״מ, ושל זן ב׳ הוא 90 ס״מ. סטיית התקן של ההתפלגות של זן א׳ היא 10 ס״מ ושל זן ב׳, 2 ס״מ. הכלבה החדשה של קובי בגובה 80 ס״מ, ומגיעה מאחד מסוגים אלה. נערוך סימולציה של ההתלבטות של קובי:
   a. בעזרת הפונקציה rnorm אותה הכרנו בתרגול, דגמו דגימה אחת מכל אחת מההתפלגויות, ושמרו אותה במשתנה. על סמך ערך הדגימות בלבד, מאיזו התפלגות נדמה שהגיעה כל אחת מהדגימות?
   b. חזרו על סעיף *a* כמה פעמים (אין צורך לדווח את התוצאות). האם המסקנה שהסקתם/ן בו משתנה?
   c. כעת, דגמו מדגם של 10 דגימות מכל אחת מההתפלגויות, ושמרו כל אחד מהמדגמים בוקטור.
   d. חשבו את הממוצע של כל אחד מהמדגמים (בעזרת הפונקציה mean או ידנית), ושמרו במשתנה נפרד. בדקו, בנוסף, כמה אילו מהערכים בכל אחד מהמדגמים גדול מהתוחלת של כל אחת מהתפלגויות המקור (השתמשו באופרטורים לוגיים).
   e. חזרו על c+d כמה פעמים (אין צורך לדווח את התוצאות). כיצד משתנה קרבת ממוצע המדגם לתוחלת המשתנה בין כל אחת מההתפלגויות? והתנהגות הערכים סביב התוחלת? מדוע?

f. כעת דגמו מכל התפלגות מדגם בגודל 1000. השתמשו בפונקציה hist על מנת להציג היסטוגרמה של כל אחד מהמדגמים, תארו אותן, וצרפו את הפלט לתרגיל.

g. מה ניתן לומר על היכולת להסיק את גזע הכלבה החדשה של קובי מתוך ההיסטוגרמות של כל אחד מהמדגמים?

שאלה 1:

1. מהי התוחלת של x?

$$E[x]=\frac{(1+3+5)\cdot 2}{15}+0.2(2+4+6)=\frac{18}{15}+0.2\cdot 12=1.2+2.4=3.6$$

3. השתמשו באופציית העזרה בR (?) על מנת לקרוא על השימוש בפונקציה sample, שהכרנו בתרגול.

כתבו : מה עושה הפונקציה, ומהו השימוש של הפרמטרים האופציונליים prob ו - replace?

הפונקציה sample מחזירה size דגימות אקראיות מ- x, כאשר x יכול להיות וקטור או מספר $1<x$ (ואז הדגימה תהיה מ- $\{1,2,...,x\}$). פונקציית sample יכולה לקבל גם ארגומנט prob שהוא וקטור משקולות שמנחה מה ההסתברות שצפויה יש לדגום ערכים שונים מ-x (סכום ההסתברויות לא חייב להיסכם ל-1, נלקח באופן פרופורציונלי).
ארגומנט replace קובע אם מותר לדגום עם חזרה ערכים או לא, כאשר הדיפולט הוא לדגום ללא חזרה.
replace=TRUE משמע לדגום עם חזרה, replace=FALSE משמע לדגום ללא חזרה.

ארגומנטים של הפונקציה

6. מהו הממוצע של המדגם? חשבו את הממוצע ידנית, ותארו את הקשר בין ערכי המדגם שדגמתם

לבין ההתפלגות של x.

$$\frac{1}{15}(2+6+3+4+6+6+4+2+2+2+4+6+6+1+3)=3.8$$

נבחין שהתקבלו בעיקר מספרים זוגיים באופן שמתאים להתפלגות, לפיה הסיכוי לקבל מספר זוגי היא כמעט פי 2 מהסתברות לקבל מספר אי זוגי.

7. מה הקשר בין כפילת כל ערך אפשרי בהסתברותו בחישוב התוחלת, ובין סכימת הערכים של המדגם

והחלוקה בn בחישוב הממוצע? חישבו שוב על הוקטור הנתון של המדגם שדגמתם.

המטרה של כפילת כל ערך אפשרי בהסתברותו בחישוב התוחלת היא ליצור ממוצע משוקלל של הערכים השונים באוכלוסייה כאשר משקלים כל ערך בחלק היחסי שלו באוכלוסייה (החלק היחסי = מס׳ הופעות הערך באוכלוסייה / גודל האוכלוסייה).
בחישוב ממוצע המדגם, סכימת הערכים של המדגם והחלוקה ב-n ממלאת את אותו תפקיד - בחישוב ממוצע המדגם שדגמנו מן האוכלוסייה כופלים כל ערך שהופיע במדגם במספר הפעמים שהופיע במדגם (מתבטא בסכימה) ומחלקים ב-n, גודל המדגם. כך יוצא ממוצע משוקלל של הערכים השונים שהתקבלו במדגם כך שכל ערך כזה משוקלל בחלקו היחסי במדגם (החלק היחסי = מס׳ הדגימות ששוות לערך מתוך המדגם / גודל המדגם). הקשר הזה מאפשר לממוצע המדגם להיות אומד טוב לתוחלת האוכלוסייה (ממוצע המדגם הוא אומד unbised לתוחלת האוכלוסיה כפי שראינו בשיעור).
כך למשל בוקטור של המדגם שקיבלנו בסעיף הקודם יש יותר הופעות של מספרים זוגיים מאשר של מספרים אי זוגיים, כך שמשקלו מספר זוגי גדול יותר ממשקלו של מספר אי זוגי, באופן דומה לשקלול בחישוב תוחלת האוכלוסייה (בו המשקל הוא לפי ההסתברות. וההסתברות לקבל מספר זוגי גדולה מההסתברות לקבל מספר אי זוגי).
נשים לב שאכן קיבלנו בשאלה הקודמת שממוצע המדגם שווה ל- 3.8, ערך שקרוב לתוחלת- 3.6.

שאלה 2 (30 נק')

בשנות ה70 של המאה שעברה התנהל ויכוח עז בקרב הקהילה המחקרית על טבע הדמיון האנושי, שעסק בשאלה האם הדמיון הוא 'תמונתי' – האם כאשר נדמיין משהו, האובייקט ייוצג במוח באופן ויזואלי – או שמא מדובר באנלוגיה בלבד, ובפועל תוכן האובייקט המדומיין מיוצג במוח בדרכים שאינן ויזואליות (לדוגמא, כמידע סמנטי). הויכוח הוכרז כ'הוכרע' על ידי פרסומם של שני מחקרים, שהראו שכאשר נבדקים/ות נדרשו לדמיין תמונה ואז להפעיל עליה פעולה, זמן התגובה שלהם/ן היה תלוי במידת

המניפולציה שהיה צורך להפעיל על התמונה, ושלכן, הסיקו, דמיון הוא תמונתי. בניסוי הראשון הנבדקים/ות נדרשו לסובב צורה תלת מימדית, ולבדוק האם גרסתה המסובבת תואמת או לא לגירוי שני. החוקרים מצאו שזמן התגובה של הנבדקים/ות היה ארוך יותר ככל שנדרשו לסובב את הגירוי בזווית גדולה יותר. בניסוי השני נדרשו הנבדקים/ות לדווח על פרט מסוים בתמונה ועל הקשר שלו לפרט אחר בתמונה. הנבדקים/ות למדו את התמונה, והיא לא הייתה זמינה בזמן המענה על השאלה. זמן התגובה של הנבדקים היה גדול יותר ככל שהפרט היה רחוק יותר מהפרט השני. (הקישורים לקריאה נוספת, אין צורך לקרוא את המאמרים כדי לענות על השאלה).

שאלה 2:

1. נסחו את שאלת המחקר באופן תאורטי (לפני אופרציונליזציה). מהם המשתנים התיאורטיים בשאלת המחקר – הבלתי תלוי והתלוי?

שאלת המחקר- האם כאשר אנחנו מדמיינים אובייקט הוא מיוצג במוח באופן ויזואלי?
המשתנה הבלתי תלוי- האובייקט המדומיין
המשתנה התלוי- צורת הייצוג של האובייקט המדומיין במוח

2. בכל אחד מהניסויים – מהו המשתנה התלוי האופרציונלי? מהו המשתנה הבלתי תלוי האופרציונלי?

ניסוי 1-
המשתנה הבלתי תלוי האופרציונלי- שינוי זווית הסיבוב של הצורה התלת מימדית
המשתנה התלוי האופרציונלי- זמן התגובה של הנבדק

ניסוי 2-
המשתנה הבלתי תלוי האופרציונלי- המרחק בין שני פריטים בתמונה
המשתנה התלוי האופרציונלי- זמן התגובה של הנבדק

3. האם מערך הניסוי הראשון הוא ניסויי או מתאמי? אם המערך ניסויי, האם המניפולציה בלבדית?
אם המערך מתאמי, מה יכול להיות הסבר אלטרנטיבי לתוצאות?

מערך הניסוי הוא ניסויי מכיוון שמבצעים מניפולציה על המשתנה הב"ת, על הזווית של הצורה שנדרשים לדמיין, בכדי להכריע התאמה לצורה השניה, במטרה לבדוק כיצד שינוי של הזווית משפיע על המשתנה התלוי, זמן התגובה של הנבדק.
מתוצאות הניסוי גם עולה שככל שהזווית גדלה, זמן התגובה עלה, כלומר נמצא קשר סיבתי ולא רק קורלציה.
ניתן שבוצעה מניפולציה בלבדית, מכיוון שהמניפולציה משפיעה באופן ישיר על זווית הסיבוב (ועליה בלבד) של האובייקט שנדרש לדמיין ובכך נראה שהיא שוללת הסברים אלטרנטיביים לשינוי שנמדד בזמן התגובה. כלומר נראה שהמניפולציה נבחרה באופן שמשפיע אך ורק את המשתנה הב"ת.

4. האם מערך הניסוי השני הוא ניסויי או מתאמי?

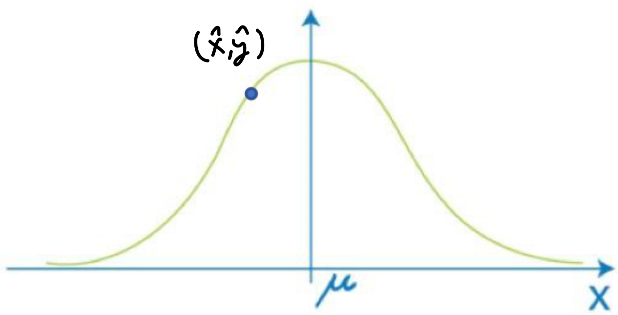
מערך הניסוי השני הוא ניסויי גם כן, מכיוון שבוצעה מניפולציה את המשתנה הבלתי תלוי, הגורם המסביר, המרחק, ובדקו כיצד זה משפיע את המשתנה התלוי, הגורם המוסבר. הזמן שנדרש לדמיין שני פריטים בתמונה ולדווח את הקשר ביניהם (זמן התגובה). כלומר בוצעה מניפולציה את המשתנה הב"ת במטרה לבדוק אם היא מובילה לשינויים במשתנה התלוי. מתוצאות הניסוי עולה שהגדלת המרחק בין הפריטים מגדילה את זמן התגובה, כלומר נמצא קשר סיבתי ולא רק מתאם כמו במערך מתאמי.

שאלה 3:

1. מה עלינו לדעת על מנת לחשב את ההסתברות שמשתנה מקרי נתון, המתפלג נורמלית, יהיה קטן או שווה לאפס?

מ"מ נורמלי הוא מ"מ רציף ולכן התפלגותו נקבעת על פי פונקציית הצפיפות המתאימה לו. פונקציית הצפיפות של מ"מ נורמלי X נקבעת באופן מוחלט ע"פ μ,σ^2, השונות והתוחלת של X בהתאמה. על כן האינטגרל בתחום $[-\infty,0]$ של פונקציית הצפיפות, שבה מחושבת ההסתברות ש-X יהיה קטן שווה ל-0 נקבעת ע"פ הם.

2)



a. מה משמעות הגובה של הנקודה? $\hat{y} = PDF(\hat{x})$

b. מה משמעות השטח שמתחת לפונקציה, משמאל לנקודה?

$P(X \le \hat{x}) = \int_{-\infty}^{\hat{x}} PDF(x)dx =$ ההסתברות לקבל ערך קטן או שווה ל-$\hat{x}$

c. מה משמעות השטח מתחת לפונקציה, מימין לנקודה?

$P(X > \hat{x}) = \int_{\hat{x}}^{\infty} PDF(x)dx =$ ההסתברות לקבל ערך גדול מ-$\hat{x}$

d. קראו על הפונקציה pnorm ותארו את שימושה.

עבור מ"מ $X\sim N(\mu,\sigma^2)$ וערך k, pnorm מחזירה את b, ההסתברות לקבלת ערך קטן או שווה ל-k, כאשר הארגומנט lower.tail=TRUE, ומחזירה את c, ההסתברות לקבל ערך גדול מ-k, כאשר lower.tail=FALSE.
באופן דיפולטי $\sigma^2=1$, $\mu=0$, ואפשר גם להזין μ,σ^2 אחרים כארגומנטים ל-pnorm.

e. הניחו כי הגרף מתאר התפלגות נורמלית בעלת תוחלת 0 וסטיית תקן 1, וערך הנקודה המסומנת על ציר ה-x הוא -0.5. חשבו ודווחו את השטח שמשמאל לנקודה בעזרת הפונקציה pnorm.

0.3085375

3. כיצד ישתנה הגרף בעקבות שינוי של תוחלת המשתנה המקרי? כיצד ישתנה הגרף בעקבות שינוי של השונות של המשתנה המקרי?

שינוי בתוחלת של המ"מ מבצע הזזה של הגרף על ציר ה-X (מרכז שונה). הגדלת התוחלת- הזזה ימינה, הקטנת התוחלת- הזזה שמאלה.
שינוי בשונות של המ"מ משפיע על הכיווץ / ההרחבה של הפונקציה. הקטנה של השונות-כיווץ, הגדלה של השונות- הרחבה.

שאלה 4:

a. בעזרת הפונקציה rnorm אותה הכרנו בתרגול, דגמו דגימה אחת מכל אחת מההתפלגויות, ושמרו אותה במשתנה. על סמך ערך הדגימות בלבד, מאיזו התפלגות נדמה שהגיעה כל אחת מהדגימות?

הדגימה מההתפלגות $N(100,100)$ של זן א' היא: ≈ 77.9
הדגימה מההתפלגות $N(90,4)$ של זן ב' היא: ≈ 89.7.
ניכר שהסיכוי לדגום ערך שקרוב לתוחלת בלכל היותר שתי סטיות תקן מן הממוצע הוא 95%~, ובלכל היותר סטיית תקן אחת הוא 68%~.

לכן נראה שהדגימה הראשונה הגיעה מההתפלגות של זן א' (לא סביר לקבל 78 אם זה זן ב', יש פחות מ-5%~. לעומת זאת עבור זן א' סביר בהרבה כי קרוב למרחק של שתי סטיות תקן 68%~). כמו כן נראה שהדגימה ב' הגיעה מההתפלגות של זן ב', מכיוון שהתוחלת 90 וסטיית התקן קטנה (סביר גם שהגיע מהתפלגות זן א', אך פחות, מכיוון ששם יש יותר ערכים בטווח של סטיית תקן אחת ליפול עליהם?).

b. חזרו על סעיף *a* כמה פעמים (אין צורך לדווח את התוצאות). האם המסקנה שהסקתם/ן בו משתנה?

כאשר חוזרים על סעיף א' מספר פעמים בכל פעם מקבלים תוצאות שונות, בגלל השונות של ההתפלגויות. אולם, ניתן לראות שבהגרלות יחסית גבוהה הדגימה הראשונה רחוקה יחסית מ-100 (וכן מ-90) ולכן בהגרלות גבוהה נראה שהיא שייכת להתפלגות ממנה היא הגיעה, שכן עבור ההתפלגות של זן ב' $\sigma = 2$ ולכן לא סביר לקבל ערכים שרחוקים יחסית מ-90.
כמו כן, בהגרלות גבוהה בדגימה שמתקבלת מההתפלגות של זן ב' נקבל ערכים שקרובים מאוד ל-90 ובמקרה זה נסיק שהיא אכן מגיעה מההתפלגות של זן ב'.
באופן כללי, כאשר השונות גבוהה מאוד סביר לקבל ערכים שרחוקים יחסית מהתוחלת, וכאשר השונות קטנה מאוד, סביר לקבל ערכים קרובים לתוחלת. לפיכך נחליט מאיזה התפלגות נראה שדגימה נלקחה, אך מהחזרות השונות ניתן ללמוד שלפעמים נסיק מסקנות שגויות (למשל, ישנן הרצות בהן הדגימה מזן ב' רחוקה למדי מהתוחלת 90)

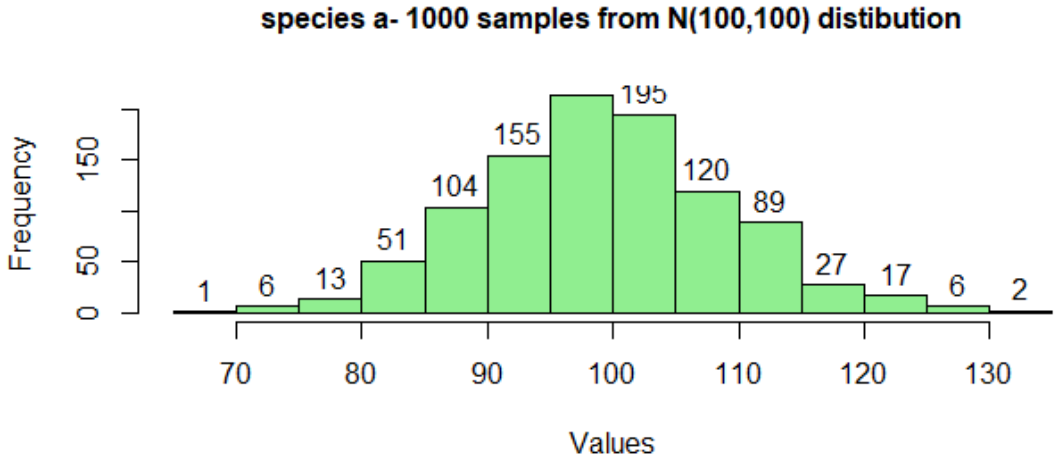
d. חשבו את הממוצע של כל אחד מהמדגמים (בעזרת הפונקציה mean או ידנית), ושמרו במשתנה נפרד. בדקו, בנוסף, כמה אילו מהערכים בכל אחד מהמדגמים גדול מהתוחלת של כל אחת מהתפלגויות המקור (השתמשו באופרטורים לוגיים).

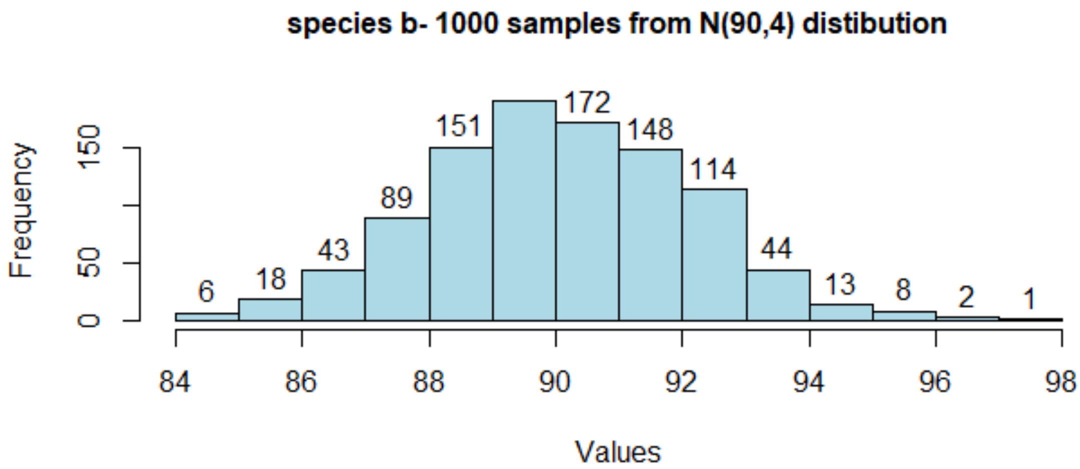
ממוצע המדגם של זן א': ≈ 101.28, 2 מהערכים גדולים מהתוחלת של התפלגות המקור
ממוצע המדגם של זן ב': ≈ 89.08, 4 מהערכים גדולים מהתוחלת של התפלגות המקור

e. חזרו על c+d כמה פעמים (אין צורך לדווח את התוצאות). כיצד משתנה קרבת ממוצע המדגם לתוחלת המשתנה בין כל אחת מההתפלגויות? והתנהגות הערכים סביב התוחלת? מדוע?

המרחק בין הממוצע לבין התוחלת עבור התפלגות של זן א' גדול יותר כמעט תמיד מאשר זן ב'. כמו כן הערכים שנדגמו מזן א' מקובצים פחות סביב התוחלת ונראה פיזור גדול יותר מאשר הדגימות מזן ב'.
זה קורה מכיוון שהשונות 100 גדולה בהרבה מהשונות 4. השונות היא מדד לפיזור של הערכים סביב התוחלת. כאשר השונות קטנה, יש יותר סיכוי לקבל ערך קרוב לתוחלת. לכן גם יותר סביר שהממוצע של מעט דגימות (10) יהיה קרוב יותר לתוחלת - כי רוב הדגימות קרובות לתוחלת בהסתברות גבוהה ולכן ממוצען קרוב לתוחלת. לעומת זאת עבור שונות גבוהה נדרשות הרבה יותר דגימות כדי שממוצע הדגימות יתקרב אל התוחלת, מכיוון שסביר יחסית לדגום ערכים מאוד שונים זה מזה.

f. כעת דגמו מכל התפלגות מדגם בגודל 1000. השתמשו בפונקציה hist על מנת להציג היסטוגרמה של כל אחד מהמדגמים, תארו אותן, וצרפו את הפלט לתרגיל.





g. מה ניתן לומר על היכולת להסיק את גזע הכלבה החדשה של קובי מתוך ההיסטוגרמות של כל אחד מהמדגמים?

כל אחת מן ההיסטוגרמות מתארת גבהים של 1,000 כלבים שנדגמו באופן בלתי תלוי ומקרי מההתפלגות של הזן שבכותרתה. 1,000 דגימות זהו מדגם יחסית גדול ולכן משקף באופן יחסית טוב את האוכלוסיות.
מן ההיסטוגרמות נראה סביר להסיק שהכלבה החדשה של קובי שגובהה 80 הוא זן א', מאחר שנראה שאין כלל כלבים מזן ב' בגובה 80 ואין בקושי בטווח זה, בעוד שישנם לא מעט כלבים מזן א' בטווח גובה זה.