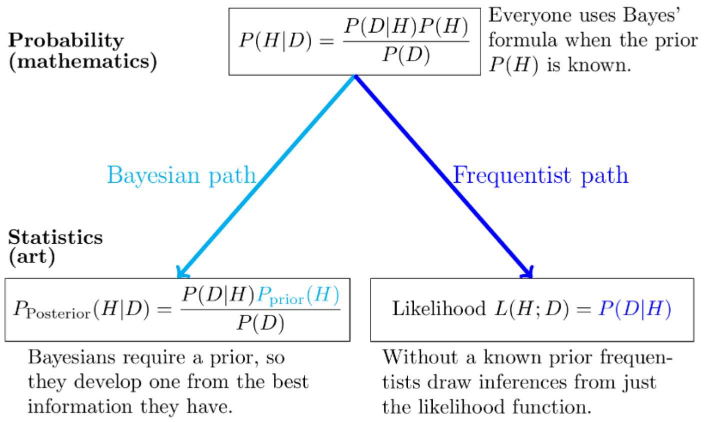
הסקה סטטיסטית שיעור 7

הקדמה

חומר קריאה 7א, חומר קריאה 7ב, חומר קריאה 7ג

הגישה השכיחותנית:



הסקה בייסיאנית תלויה ב prior – בעצם, בהנחה שלנו לגבי המציאות.
הגישה השכיחותית לא משתמשת ב prior. היא מסתמכת רק על תוצאות של ניסויים.

דוגמה 1

נניח שיש לנו מטבע מוטה עם הסתברות לא ידועה θ לעץ. הערך של θ לא ידוע, אבל הוא ערך קבוע. אז בגישה השכיחותית, אין prior pdf $f(\theta)$.
לעומת זאת, הגישה הבייסיאנית מסכימה שיש ל- θ ערך קבוע, אבל מתייחסת ל- $f(\theta)$ בתור ייצוג לאי-הוודאות לגבי הערך. שניהם מסכימים ש $p(heads \mid \theta) = \theta$, כי זה מה שייצא בניסויים לאורך זמן.

לסיכום: בגישה הבייסיאנית יש הסתברות להכל – גם ההשערות וגם הדאטא. בגישה השכיחותית יש הסתברויות לדאטא בהינתן השערה.
כשמדברים על דאטא עם התפלגות לא ידועה, רק ל likelihood יש משמעות. אין משמעות ל prior או ל posterior.

הגדרה: סטטיסטי

סטטיסטי הוא כל דבר שאפשר לחשב מתוך הדאטא. לפעמים נדייק ונאמר שהסטטיסטי הוא **כלל** לחישוב משהו מתוך הדאטא, ו**הערך** של הסטטיסטי הוא מה שמחושב. זה יכול לכלול גם חישוב של likelihoods כשאנחנו משערים מה הערכים של הפרמטרים של המודל. זה לא כולל כל דבר שדורש שנדע את הערך האמיתי של פרמטר לא ידוע.
הממוצע של דאטא זה סטטיסטי. כנ"ל המינימום או המקסימום של דאטא.
נניח ש $x \sim N(\mu, 9)$ עם פרמטר μ לא ידוע. אז ה likelihood:

$$p(x \mid \mu = 7) = \frac{\exp\left(\frac{(x-7)^2}{-18}\right)}{3\sqrt{2\pi}}$$

הוא סטטיסטי. כי הוא לא תלוי ב μ האמיתי.

בדיקת השערת האפס – NHST: Null Hypothesis Significance Testing (חומר קריאה 7ב)

נשקול את הדאטא רק דרך פונקציית ה likelihood, בלי קשר ל prior.

המרכיבים של NHST:

השערת האפס – H_0: ברירת המחדל במודל שמייצר את הדאטא.

ההשערה האלטרנטיבית – H_A: אם נדחה את השערת האפס, נקבל את ההשערה הזו בתור ההסבר הכי טוב לדאטא.

סטטיסטי המבחן – X: נחשב מתוך הדאטא.

תחום הדחייה – rejection region: אם X נמצא בתחום הזה אז נדחה את H_0 ונקבל את H_A.

תחום אי-דחייה – non-rejection region: המשלים לתחום הדחייה. אם X נמצא בתחום הזה, **לא נדחה** את H_0. זה לא אומר שנקבל את H_0, זה פשוט אומר שהדאטא לא קובעת שצריך לדחות את H_0.

השערת האפס היא בדרך כלל השערה פשוטה, או ברירת המחדל, שנדחה רק אם נקבל מספיק ממצאים נגדה.

טרמינולוגיה של NHST:

דוגמה – מטבע: כדי לבדוק אם מטבע נתון הוא הוגן, נטיל אותו 10 פעמים. אם נקבל מספר גבוה או נמוך במיוחד של עץ, נחשוד שהמטבע מוטה. נפרמל את זה במונחים של NHST: יהי θ ההסתברות שהמטבע יוצא עץ בהטלה אחת.

השערת האפס H_0 – "המטבע הוגן", כלומר $\theta = 0.5$.

השערה אלטרנטיבית H_A – "המטבע מוטה", כלומר $\theta \neq 5$.

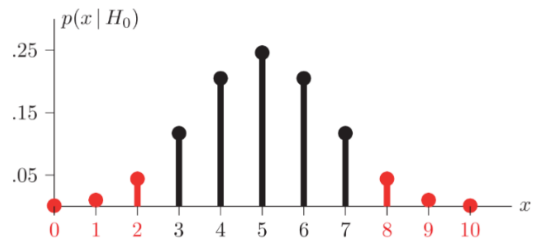
סטטיסטי המבחן X – מספר העץ ב 10 הטלות.

התפלגות האפס – פונקציית ההסתברות המתבססת על השערת האפס: $p(x|\theta = 0.5) \sim Bin(10, 0.5)$

טבלת ההסתברויות עבור התפלגות השערת האפס:

x	0	1	2	3	4	5	6	7	8	9	10
$p(x \mid H_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

תחום דחייה – תחת השערת האפס, נצפה לקבל בערך 5 עץ מתוך 10 הטלות. אז נדחה את H_0 אם X רחוק מ 5. נגדיר את תחום הדחייה $\{0, 1, 2, 8, 9, 10\}$. אפשר לסכם את זה בגרף וטבלת ההסתברויות:



x	0	1	2	3	4	5	6	7	8	9	10
$p(x\|H_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

כמה הערות לגבי הדוגמה:

1. השערת האפס היא ברירת מחדל זהירה: לא נטען שהמטבע מוטה אם אין לנו ראיות משכנעות.
2. תחום הדחייה מורכב מדאטא שנחשבת קיצונית תחת השערת האפס. כלומר, תוצאות שנמצאות בזנבות של התפלגות האפס – רחוק מהמרכז של הסתברות גבוהה. בהמשך נדבר על עד כמה רחוק – זה תלוי ב α, רמת המשמעות של המבחן.
3. אם נקבל 3 עץ, אז סטטיסטי המבחן נמצא בתחום אי-הדחייה. נאמר שהדאטא "לא תומכת בדחיית השערת האפס". אפילו אם נקבל 5 עץ, **לא** נטען שהדאטא מוכיחה את השערת האפס.

שאלה: אם יש לנו מטבע הוגן, מה ההסתברות שנחליט באופן שגוי שהמטבע מוטה?

זה יקרה אם נקבל $\{0, 1, 2, 8, 9, 10\}$. ההסתברות לזה היא סכום ההסתברויות האדומות בטבלה, 0.11.

השערות פשוטות ומורכבות

השערה פשוטה היא כזו שאנחנו יכולים להגדיר את ההתפלגות שלה לגמרי. השערה נפוצה כזו היא ההשערה שפרמטר מקבל ערך מסויים.
השערה מורכבת אם אנחנו לא יכולים להגדיר את ההתפלגות שלה. לדוגמה, ההשערה שפרמטר נמצא בטווח של ערכים.

בדוגמה של המטבע, הערת האפס היא $\theta = 0.5$, והתפלגות האפס היא $Bin(10, 0.5)$. ההתפלגות מוגדרת אז ההשערה פשוטה. ההשערה האלטרנטיבית היא $\theta \neq 0.5$, שזה בעצם הרבה השערות ביחד. אז היא מורכבת.

דוגמה: נניח שיש לנו דאטא $x_1, \ldots x_n$. וההשערות שלנו הן:
H_0 – הדאטא מגיעה מהתפלגות $N(0,1)$, H_A – הדאטא מגיעה מהתפלגות $N(1, 1)$. שתיהן השערות פשוטות.

אם ההשערות הן:
H_0 – הדאטא מגיעה מהתפלגות פואסון עם פרמטר לא ידוע, H_A – הדאטא לא מגיעה מהתפלגות פואסון. אז שתיהן השערות מורכבות.

דוגמה – ESP (תפיסה על-חושית, "חוש שישי"):

במבחן לבדיקת ESP, הנבדק התבקש לזהות את הצורה של 100 קלפים שנבחרו עם החזרה מחפיסה. יהי T מספר ההצלחות. השערת האפס היא שלנבדק אין ESP, והיא השערה פשוטה: $H_0: T \sim Bin(100, 0.25)$

ההשערה האלטרנטיבית היא שלנבדק יש ESP, השערה מורכבת: $H_A: T \sim Bin(100, p),\ \ with\ p > 0.25$

עוד השערה אלטרנטיבית היא שמתרחש משהו חוץ מאקראיות, כלומר לנבדק יש ESP או אנטי-ESP. נתונה ע"י:

$$H_A:\ T \sim Bin(100, p), \qquad with\ p \neq 0.25$$

ערכים של $p < 0.25$ מייצגים מצב שיש לנבדק סוג של אנטי-ESP.

סוגים של טעויות

		המצב האמיתי	
		H_0	H_A
ההחלטה שלנו	לדחות את H_0	טעות סוג 1	החלטה נכונה
	לא לדחות את H_0	החלטה נכונה	טעות סוג 2

יש שני סוגים של טעויות: לדחות את השערת האפס כשהיא אמיתית, או לא לדחות את השערת האפס כשהיא שגויה. הם נקראות טעות סוג 1 וסוג 2 (*type I, type II*). טעות סוג 1 זה לדחות את H_0 כשההשערה אמיתית. טעות סוג 2 זה לא לדחות את H_0 כשהיא שגויה.

מובהקות סטטיסטית, ועוצמה סטטיסטית

משמשים למדידת האיכות של מבחן המובהקות. אידיאלית, במבחן מובהקות לא יהיו טעויות. כלומר, לא נדחה את H_0 אם היא אמיתית, וכן נדחה אותה ונקבל את H_A אם H_A אמיתית. בסה"כ יש 4 הסתברויות חשובות, שמתאימות לטבלה לעיל:

$$P(\text{reject } H_0 \mid H_0), \qquad P(\text{reject } H_0 \mid H_A), \qquad P(\text{don't reject } H_0 \mid H_0), \qquad P(\text{don't reject } H_0 \mid H_A)$$

ההסתברויות שנתמקד בהן:

מובהקות = $P(\text{reject } H_0 \mid H_0)$ = ההסתברות שנדחה את H_0 באופן שגוי = $P(\text{type I error})$.

עוצמה = ההסתברות שנדחה את H_0 באופן מוצדק = $P(\text{reject } H_0 \mid H_A) = 1 - P(\text{type II error})$.

אידיאלית, למבחן יהיו מובהקות נמוכה (קרובה ל 0) ועוצמה גבוהה (קרובה ל 1).

אנלוגיות כדי שיהיה קל לזכור את כל זה

אפשר לחשוב על H_0 בתור ההשערה ש"אין פה שום דבר מעניין". לדוגמה, "המטבע הוגן" או "התרופה לא יותר יעילה מאשר פלצבו". לעומת זאת, H_A זה ההיפך – "משהו מעניין קורה כאן". המטבע מוטה, התרופה עובדת. אזי, העוצמה זה ההסתברות שנגלה משהו מעניין כשהוא קורה והמובהקות היא ההסתברות שבטעות נחשוב שמשהו מעניין קרה.

חזקת חפות מפשע – אדם זכאי כל עוד לא הוכח אחרת מעבר לכל ספק סביר. במונחים של *NHST*:

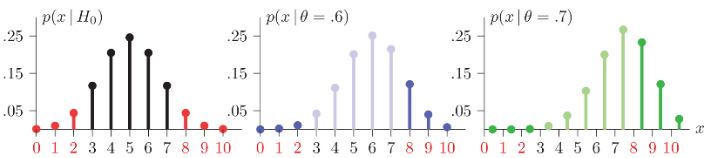
H_0 = האדם זכאי, H_A = האדם אשם.

מובהקות היא ההסתברות שאדם זכאי יואשם. עוצמה היא ההסתברות שאדם אשם אכן יואשם. "מעבר לכל ספק סביר" – נדרוש שהמובהקות תהייה קטנה מאוד.

השערות מורכבות

בדוגמה של המטבע, H_A הוא מורכב, אז המובהקות שונה עבור ערכי שונים של θ. נרחיב את טבלת ההסתברות שתכלול ערכים אחרים לθ. כמו תמיד ב*NHST*, נסתכל על ה *likelihoods*: ההסתברות של דאטא בהינתן השערה.

x	0	1	2	3	4	5	6	7	8	9	10
$H_0 : p(x\|\theta = 0.5)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001
$H_A : p(x\|\theta = 0.6)$	.000	.002	.011	.042	.111	.201	.251	.215	.121	.040	.006
$H_A : p(x\|\theta = 0.7)$	.000	.0001	.001	.009	.037	.103	.200	.267	.233	.121	.028



מובהקות = ההסתברות שנדחה את H_0 אם היא נכונה
= ההסתברות שסטטיסטי המבחן יהיה בתחום הדחייה אם H_0 נכון
= ההסתברות שסטטיסטי המבחן הוא בתחום הדחייה של השורה של H_0 בטבלה
= סכום המשבצות האדומות בשורה של $(\theta = 0.5) = 0.11$

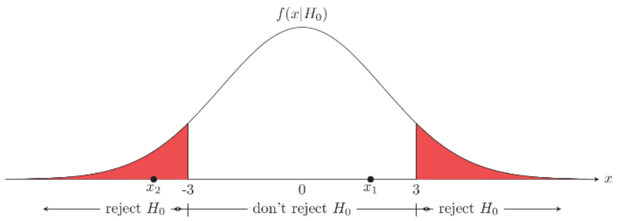
עוצמה כאשר $(\theta = 0.6)$ = ההסתברות שנדחה את H_0 כאשר $(\theta = 0.6)$
= ההסתברות שסטטיסטי המבחן בתחום הדחייה כאשר $(\theta = 0.6)$
= ההסתברות שסטטיסטי המבחן הוא בתחום הדחייה של השורה של $(\theta = 0.6)$ בטבלה
= סכום המשבצות בצבע כחול כהה $= 0.18$

עוצמה כאשר $(\theta = 0.7)$ = ההסתברות שנדחה את H_0 כאשר $(\theta = 0.7)$
= ההסתברות שסטטיסטי המבחן בתחום הדחייה כאשר $(\theta = 0.7)$
= ההסתברות שסטטיסטי המבחן הוא בתחום הדחייה של השורה של $(\theta = 076)$ בטבלה
= סכום המשבצות בצבע ירוק כהה $= 0.384$

ניתן לראות שהעוצמה גבוהה יותר עבור $\theta = 0.7$ מאשר $\theta = 0.6$. זה לא מפתיע, כי נצפה שיהיה קל יותר לראות שמטבע עם הטיה 0.7 הוא מוטה מאשר מטבע של 0.6. באופן כללי, אנחנו מקבלים עוצמה גבוהה יותר כאשר H_A רחוקה מ H_0.

המחשות

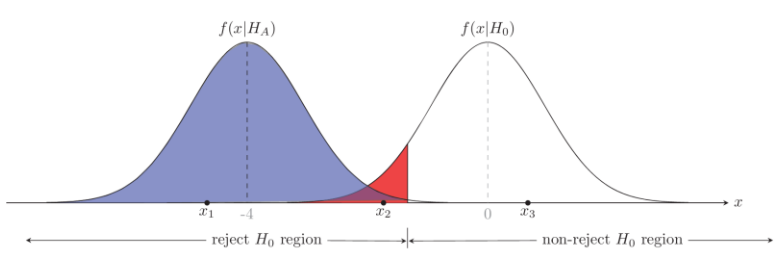
לשם הדגמה, יש שתי אפשרויות לסטטיסטי המבחן: x_1, x_2. אם הדאטא מניבה את x_1, אז נדחה את H_0. אם הדאטא מניבה את x_2, אז לא נדחה את H_0.



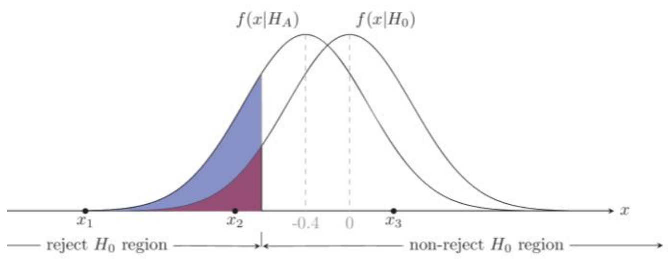
נשים לב לכמה דברים:

1. תחום הדחייה הוא ערכים רחוקים מהמרכז של התפלגות האפס.
2. תחום הדחייה הוא דו צדדי. יהיו גם דוגמאות לתחום דחייה חד צדדי.
3. ההשערה האלטרנטיבית לא מוזכרת. אנחנו דוחים או לא דוחים את H_0 רק לפי ה $likelihood$ $f(x|H_0)$. כלומר, ההסתברות של סטטיסטי המבחן בהינתן H_0. כפי שנראה בהמשך, צריך לקחת בחשבון את H_A כשוחרים תחום דחייה, אבל באופן פורמלי היא לא קשורה להחלטה האם לקבל או לא לקבל את H_0.
4. לפעמים אומרים "נקבל את השערת האפס" במקום "לא נדחה את השערת האפס". זה טכנית לא נכון כי אף פעם לא באמת נקבל את השערת האפס. או שנדחה אותה או שנאמר שהדאטא לא מספיקה כדי לדחות אותה. **אי אפשר להוכיח את השערת האפס.**

מבחן חזק:



מבחן חלש:



ניזכר: האיזור הצבוע מתחת ל $f(x|H_0)$ היא המובהקות – ההסתברות שנדחה את H_0 כשהיא נכונה, כלומר ההסתברות שסטטיסטי המבחן יצא בתחום הדחייה בהינתן H_0. האיזור הצבוע מתחת ל $f(x|H_A)$ הוא העוצמה – ההסתברות שסטטיסטי המבחן יצא בתחום הדחייה אם H_A נכון. בשני המבחנים המובהקות זהה, אבל אם ל $f(x|H_A)$ יש חפיפה גדולה עם $f(x|H_0)$ אז העוצמה יורדת.

בשני המבחנים ההתפלגויות הן נורמלי סטנדרטי. התפלגות האפס, תחום הדחייה והמובהקות זהים. באיור הראשון רואים שהממוצעים של ההתפלגויות רחוקים 4 סטיות תקן אחד מהשני. מכיוון שיש מעט מאוד חפיפה בין התחומים, למבחן יש עוצמה גבוהה יותר. כלומר, אם הדאטא x מגיעה מ H_A היא כמעט בוודאות תהיה בתחום הדחייה. לדוגמה בהתפלגות H_A, x_3 היא תוצאה לא מאוד סבירה. באיור השני רואים שהממוצעים הם במרחק רק 0.4 סטיות תקן. למבחן יש עוצמה נמוכה יותר, כי אם x מגיע מ H_A, סביר להניח שזה יהיה בתחום אי-דחייה. לדוגמה בהתפלגות H_A, x_3 היא תוצאה הרבה יותר סבירה.

בדרך כלל, נוכל להגדיל את העוצמה של מבחן על ידי הגדלת כמות הדאטא ובכך להקטין את השונויות של ההתפלגויות. בתכנון מבחן, חשוב לקבוע מראש את מספר החזרות או הבדיקות הנדרשים כדי לקבל עוצמה מספקת.

דוגמה – תרופה

נניח שמשווים תרופה מול פלצבו: H_0 = התרופה לא עובדת יותר טוב מהפלצבו, H_A = התרופה כן עובדת יותר טוב מהפלצבו.

העוצמה של המבחן היא ההסתברות שהמבחן יקבע שהתרופה יותר טובה, אם היא באמת יותר טובה. המובהקות היא ההסתברות שהמבחן יקבע שהתרופה יותר טובה גם אם היא לא.

תכנון מבחן

1. נבחר את השערת האפס - H_0.

בחירת השערת האפס וההשערה האלטרנטיבית היא לא תהליך מתמטי – זה תהליך אומנותי. יש יותר מדרך אחת נכונה ויש מנהגים שונים. בדרך כלל נבחר את H_0 להיות ההסבר הכי פשוט או זהיר. כלומר – התרופה לא משפיעה, אין ESP, המטבע הוגן. (המצב המשעמם).

2. נחליט אם H_A הוא דו צדדי או חד צדדי.
אם תחום הדחייה הוא קיצון כלשהו לשני הכיוונים (לדוגמה אם מספר הטלות שיוצאות עץ הוא גבוה או נמוך), אז זה יהיה תחום דו צדדי.

3. נבחר סטטיסטי מבחן.
לדוגמה: הממוצע, סכום, או שונות של המדגם. הרבה פעמים הבחירה ברורה. סטטיסטיים נפוצים שנראה הם z, t, χ^2.

4. נבחר מובהקות ונקבע את תחום הדחייה.
בדרך כלל נסמן ב α את המובהקות. פרדיגמת ניימן-פירסון (*Neyman-Pearson*) היא לבחור את α מראש. ערכים נפוצים הם 0.1, 0.05, 0.01. ניזכר שרמת המובהקות היא ההסתברות לטעות מסוג 1, כלומר שנדחה את H_0 כשהיא נכונה. הערך שנבחר תלוי בהשלכות של טעות מסוג 1.

אחרי שבחרנו את המובהקות, אפשר לקבוע את תחום הדחייה בזנב של התפלגות האפס. בדוגמה של המטבע, H_A הוא דו צדדי אז תחום הדחייה מתחלק בין שני הזנבות של התפלגות האפס. ההתפלגות נתונה בטבלה:

x	0	1	2	3	4	5	6	7	8	9	10
$p(x\|H_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

אם נקבע $\alpha = 0.05$ אז תחום הדחייה צריך לכלול לכל היותר הסתברות 0.05. עבור תחום דחיה דו צדדי נקבל: $\{0,1,9,10\}$. אם נקבע את $\alpha = 0.01$ נקבל: $\{0, 10\}$.

אם נשנה את H_A להיות "המטבע מוטה לטובת עץ" (כלומר, עכשיו השערת האפס היא "המטבע לא מוטה לטובת עץ". הוא יכול להיות מוטה לטובת פלי, או הוגן), אז עכשיו יש לנו השערה חד צדדית $\theta > 0.5$. תחום הדחייה שלנו הוא בזנב הימני כי אנחנו לא רוצים לדחות את H_0 אם נקבל מספר קטן של עץ. עכשיו עבור $\alpha = 0.05$ נקבל $\{9,10\}$ ועבור $\alpha = 0.01$ נקבל רק $\{10\}$.

5. נקבע את העוצמות. אחרי שקבענו את תחום הדחייה אפשר לקבוע את עוצמת המבחן בערכים שונים של H_A.

דוגמה – השלכות של מובהקות

אם $\alpha = 0.1$, זה אומר שנצפה ל-10% תדירות של טעות מסוג 1. כלומר, מתוך המקרים שבהם H_0 נכונה, ב-10% מהם נדחה אותה. התשובה לשאלה האם מובהקות 0.1 היא סבירה, תלויה בהחלטות שיתקבלו ממנה.

לדוגמה, אם אנחנו מבצעים ניסוי כדי לבדוק אם שוקולד מסויים הוא יותר מ 72% קקאו, אז תדירות של 10% טעות מסוג 1 היא כנראה בסדר. כלומר, אם בטעות נחשוב ששוקולד 72% הוא יותר מ 72%, זה בסדר. מצד שני, אם אנחנו מבצעים בדיקה לזיהוי טביעות אצבע בחקירת רצח, אז הסתברות של 10% שנפליל בטעות היא לא הסתברות מקובלת.

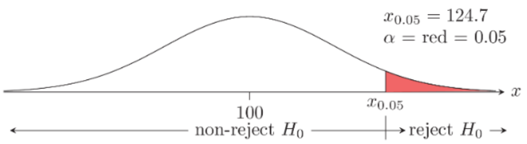
מובהקות בהשערת אפס מורכבת. אם H_0 מורכבת, אז ההסתברות $P(\text{type I error})$ תלויה באיזה מהמרכיבים של H_0 הוא האמיתי. במקרה הזה, המובהקות תהיה המקסימלית מבין ההסתברויות האלה.

ערכים קריטיים

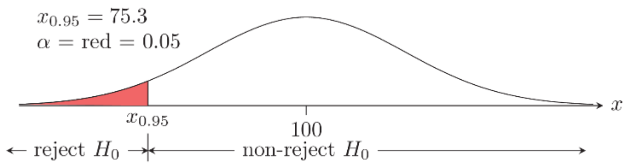
ערכים קריטיים הם כמו שברונים (*quantiles*) אבל הם מתייחסים להסתברות מצד ימין של הערך במקום צד שמאל.

דוגמה – ערכים קריטיים ותחומי דחיה: נניח שלסטטיסטי המבחן שלנו יש התפלגות אפס $N(100, 15^2)$. כלומר, $f(x|H_0) \sim N(100, 15^2)$. נניח גם שתחום הדחייה שלנו הוא רק בצד ימין ויש מובהקות 0.05. נמצא את הערך הקריטי ונשרטט את התפלגות האפס ותחום הדחייה.

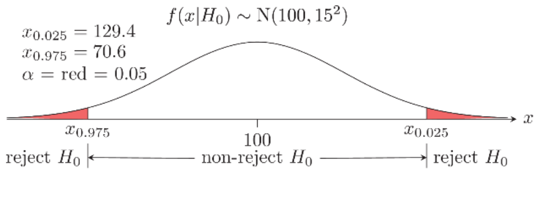
הסימון לערך קריטי עם זנב ימני שמכיל הסתברות 0.05 הוא $x_{0.05}$. הערך הקריטי של $x_{0.05}$ הוא השברון ה-0.95, כלומר יש לו הסתברות של 5% מצד ימין ולכן 95% משמאל. החישוב שבוצע הוא `qnorm(0.95, 100, 15)`:



אם תחום הדחייה הוא שמאלי, אז `qnorm(0.05, 100, 15)`



אם התחום דו צדדי, אז ההסתברות מתחלקת בין שני צדדים:
`qnorm(0.975, 100, 15)` `qnorm(0.025, 100, 15)`:



ערכי p

בפועל, לעיים קרובות נציין את רמת המובהקות ונבצע את בדיקת המובהקות לפי ערכי p. נגדיר מהו, ונראה שמתקיים:

אם ערך ה-p הוא פחות מהמובהקות α אז אנחנו דוחים את H_0. אחרת, לא דוחים.

הגדרה: ערך ה-p הוא ההסתברות, בהנחה שהשערת האפס מתקיימת, שנראה דאטא קיצונית לפחות כמו הדאטא בניסוי. המשמעות של "קיצונית לפחות כמו" תלויה במבנה הניסוי.

דוגמה חד צדדית – מבחן *z* להשערה נורמלית: בדרך כלל, IQ מתפלג נורמלית $N(100, 15^2)$ באוכלוסייה. אנחנו חושדים שלרוב הסטודנטים ב-MIT יש IQ מעל הממוצע, אז נתאר את ההשערות הבאות:

H_0 = אין הבדל בהתפלגות = ה-IQ של תלמידי MIT מתפלג $N(100, 15^2)$.

H_A = לסטודנטים ב-MIT נוטה להיות IQ גבוה יותר = הממוצע של ה-IQ גבוה מ-100.

נשים לב ש H_A הוא חד צדדי.

נניח שבדקנו 9 סטודנטים ומצאנו שיש להם ממוצע $\bar{x} = 112$. האם נוכל לדחות את H_0 ברמת מובהקות $\alpha = 0.05$?

כדי לחשב את p, נעשה סטנדרטיזציה לדאטא: תחת השערת האפס, $\bar{x} \sim N(100, 15^2)$ ולכן:

$$z = \frac{\bar{x} - 100}{15/\sqrt{9}} = \frac{36}{15} = 2.4 \sim N(0,1)$$

כלומר, התפלגות האפס של z היא נורמלית סטנדרטית. נקרא ל-z **סטטיסטי z** (*z-statistic*), ונשתמש בו בתור סטטיסטי המבחן. עבור H_A ימנית, "דאטא קיצונית לפחות כמו הדאטא בניסוי" משמעותה הזנב הימני של z. אז ערך ה-p הוא:

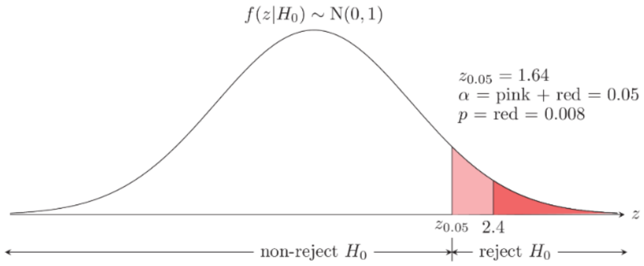
$$p = P(Z \geq 2.4) = 1 - \text{pnorm}(2.4, 0, 1) = 0.0081975$$

מכיוון ש $p \leq \alpha$, נדחה את השערת האפס. נאמר פורמלית:

דחינו את השערת האפס לטובת ההשערה האלטרנטיבית שלסטודנטים ב-MIT יש בממוצע IQ גבוה יותר.
ברמת מובהקות 0.05 עם ערך p 0.008.

נשים לב:

1. הממוצע $\bar{x} = 112$ הוא משתנה מקרי. אם נבצע את הניסוי שוב ייתכן שנקבל תוצאות שונות.
2. אפשר להשתמש בסטטיסטי $\bar{x}$ ישירות. אבל ביצוע סטנדרטיזציה היא דבר סטנדרטי[1], כי נוכל לפתח אינטואיציה למשמעות של ערכי z שונים.
3. ההצדקה לדחיית H_0 כאשר $p \leq \alpha$ מוסברת באיור הבא:



המשך *NHST* (חומר קריאה 7ג)

יש שני סוגי מבחנים: מבחן t למדגם יחיד, ומבחן t לשני מדגמים. כל מבחן מניח משהו לגבי הדאטא – בדרך כלל, נניח שהיא מגיעה מהתפלגות נורמלית. נשים לב גם שכל המבחנים הולכים לפי תבנית דומה. הדבר היחיד שמשתנה הוא החישוב של סטטיסטי המבחן והתפלגות האפס.

ניזכר בשלבים של תכנון מבחן:

1. נתכנן ניסוי שיאסוף דאטא, ונבחר סטטיסטי מבחן x שנחשב מתוך הדאטא. המפתח הוא שאנחנו צריכים לדעת את התפלגות האפס, $f(x|H_0)$. כדי לחשב את העוצמה, נצטרך גם לדעת את ההתפלגות האלטרנטיבית $f(x|H_A)$.
2. נבחר האם המבחן הוא חד צדדי או דו צדדי. לפי המבנה של H_A ו-H_0.
3. נבחר רמת מובהקות α לדחיית H_0. אם זה רלוונטי, נחשב גם את עוצמת המבחן.
4. נבצע את הניסוי ונאסוף את הדאטא $x_1, \dots, x_n$.
5. נחשב את סטטיסטי המבחן x.

---

[1] אה, באמת?

6. נחשב את ערך ה-p שמתאים ל-x, לפי התפלגות האפס.
7. אם $p < \alpha$, נדחה את H_0.

הערות:

1. במקום לבחור רמת מובהקות, אפשר קודם לבחור תחום דחייה. רמת המובהקות שנובעת מזה היא ההסתברות ש x יוצא בתחום הדחייה אם H_0 נכון.
2. השערת האפס היא ההשערה ה"בטוחה" יותר. ככל שהמובהקות נמוכה, נצטרך יותר "ממצאים" כדי לדחות אותה. מקובל לפרסם גם את ערך ה-p עצמו, כדי שאחרים יוכלו להסיק מסקנות בעצמם.
3. **רמת מובהקות 0.05 לא אומרת שהמבחן טועה רק 5% מהזמן!**

זה אומר רק **שאם** השערת האפס נכונה, אז ההסתברות שסטטיסטי המבחן ייצא בתחום הדחייה הוא 5%. העוצמה של מבחן היא רמת הדיוק של המבחן אם H_A נכונה. נסכם:

מובהקות – ההסתברות לדחות את H_0 אם היא נכונה. $P(\text{reject } H_0|H_0)$

עוצמה – ההסתברות לדחות את H_0 אם היא לא נכונה. $P(\text{reject } H_0|H_A)$

ההסתברות לא לדחות את H_0 אם היא לא נכונה: $P(\text{not reject } H_0|H_A) = 1 - P(\text{reject } H_0|H_A)$

הבנת ה-*NHST*

שאלות שנרצה לשאול:

1. איך הדאטא נאספה? מה מבנה הניסוי?
2. מהם השערת האפס וההשעה האלטרנטיבית?
3. באיזה סוג מבחן מובהקות השתמשו? האם הדאטא מתאים לקריטריונים הנדרשים לסוג כזה של מבחן? עד כמה המבחן עמיד (*robust*) בפני סטיות מהקריטריונים האלה?
4. לדוגמה, חלק מהמבחנים שמשווים שתי קבוצות של דאטא מניחים שהקבוצות נדגמו מתוך התפלגויות בעלות אותה שונות. צריך לוודא את זה לפני שמיישמים את המבחן. לעיתים קרובות הבדיקה נעשית על ידי מבחן מובהקות אחר שנועד להשוות את השונויות של שתי קבוצות הדאטא.
5. איך ה *p-value* מחושב? מבחן מובהקות מגיע עם סטטיסטי מבחן והתפלגות אפס. ברוב המקרים, ה *p-value* הוא:
$$p = P(\text{the data is at least as extreme as what we got}|H_0)$$
המשמעות של "קיצוני לפחות כמו" משתנה בהתאם למבחן חד צדדי או דו צדדי.
6. מהי רמת המובהקות α במבחן?

מבחני *t*

הרבה מבחני מובהקות מניחים שהדאטא מגיעה מהתפלגות נורמלית, אז צריך לבדוק את הדאטא כדי לבדוק אם זה הגיוני. היסטוגרמה היא התחלה טובה. המבחנים שנראה: מבחן z, מבחן t למדגם יחיד, מבחן t לשני מדגמים – מתחילים מהנחת הנורמליות הזו.

מבחן *z*

דאטא: נניח ש $x_1, \dots, x_n \sim N(\mu, \sigma^2)$, עם μ לא ידוע ו-σ ידועה. השערת האפס: $\mu = \mu_0$ עבור μ_0 כלשהו.

סטטיסטי המבחן הוא הממוצע אחרי סטנדרטיזציה:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

התפלגות האפס: $f(z|H_0)$ היא ה-*PDF* של $Z \sim N(0,1)$.

ערכי p: ימני - $p = P(Z > z \mid H_0)$, שמאלי - $p = P(Z < z \mid H_0)$, דו צדדי - $p = P(|Z| > |z| \mid H_0)$.

דוגמה

נניח שנתון לנו $\sigma^2 = 4$, ויהיו $H_0 \coloneqq \mu_0 = 0$, $H_A \coloneqq \mu > 0$.

נניח שהדאטא: $1, 2, 3, 6, -1$. עם רמת מובהקות $\alpha = 0.05$, האם נדחה את H_0?

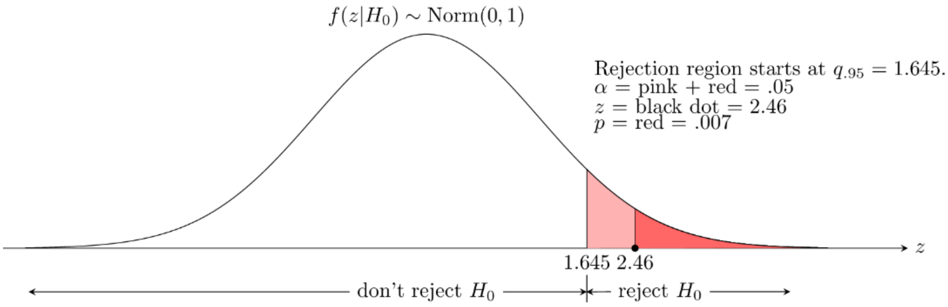
הממוצע הוא $\bar{x} = 2.2$. נבדוק במבחן z:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.2 - 0}{2/\sqrt{5}} = 2.46$$

נבדוק מה הערך p:

$$p = P(Z > z) = P(Z > 2.46) = 0.007$$

מכיוון ש $p < 0.05$, נדחה את השערת האפס.



התפלגות t "סטודנט"

התפלגות t היא סימטרית ובצורת פעמון כמו התפלגות נורמלית. יש לו פרמטר df – דרגות חופש (*degrees of freedom*). עבור df קטן, להתפלגות יש יותר הסתברות בזנבות מאשר הנורמלית. כאשר df גדל, $t(df)$ נהיה יותר דומה ל-$N(0,1)$.

מבחן t למדגם יחיד

במבחן z, הנחנו שאנחנו יודעים את השונות של ההתפלגות של הדאטא. במקרים שהשונות לא ידועה, צריך להעריך אותה מתוך הדאטא. במקרים כאלה נשתמש במבחן t למדגם יחיד במקום מבחן z.

דאטא: נניח ש $x_1, \ldots, x_n \sim N(\mu, \sigma^2)$, עם μ, σ לא ידועים. השערת האפס: $\mu = \mu_0$ עבור μ_0 כלשהו.

סטטיסטי המבחן הוא:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

כאשר:

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

כאן, t נקרא ה ***studentized mean*** ו- s^2 נקרא **שונות המדגם** (*sample variance*). זה הערכה של השונות האמיתית σ^2.

התפלגות האפס: $f(t|H_0)$ היא ה-*PDF* של $T \sim t(n-1)$. כלומר, התפלגות t עם $n-1$ דרגות חופש.

יש משפט מוכח שאומר שאם הדאטא נורמלית עם ממוצע μ_0 אז ה- *studentized mean* מתפלגת t כלשהו.

ערכי p: ימני - $p = P(T > t \mid H_0)$, שמאלי - $p = P(T < t \mid H_0)$, דו צדדי - $p = P(|T| > |t| \mid H_0)$.

דוגמה

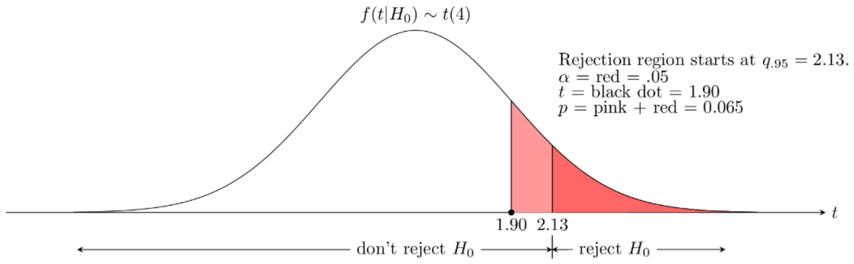
ניקח את הדוגמה הקודמת, אבל נניח שהשונות לא ידועה. כלומר: יהיו $H_0 \coloneqq \mu_0 = 0$, $H_A \coloneqq \mu > 0$. נניח שהדאטא: $1, 2, 3, 6, -1$. עם רמת מובהקות $\alpha = 0.05$, האם נדחה את H_0?

הממוצע הוא $\bar{x} = 2.2$. יש לנו דאטא נורמלית עם תוחלת ושונות לא ידועים, נבצע מבחן t למדגם יחיד. נחשב:

$$s^2 = \frac{1}{4}\left((1-2.2)^2 + (2-2.2)^2 + (3-2.2)^2 + (6-2.2)^2 + (-1-2.2)^2\right) = 6.7$$

סטטיסטי המבחן הוא:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{2.2 - 0}{\sqrt{6.7}/\sqrt{5}} = 1.901$$



ה *p-value*:

$$p = P(T > t) = P(T > 1.901) = 0.065$$

מכיוון ש $p < 0.05$, לא נדחה את השערת האפס.

מבחן t לשני מדגמים עם שונות שווה

מאפשר לנו להשוות בין הממוצעים של שני מדגמים. לדוגמה, אם נרצה להשוות בין ההשפעה של שני טיפולים רפואיים.

דאטא: נניח שיש לנו שתי קבוצות דאטא, שנדגמו מתוך התפלגויות נורמליות:

$$x_1, x_2, \dots, x_n \sim N(\mu_1, \sigma^2)$$

$$y_1, y_2, \dots, y_m \sim N(\mu_2, \sigma^2)$$

כאשר כל הפרמטרים לא ידועים. נשים לב להנחה שלשתי ההתפלגויות יש את אותה שונות.

השערת האפס: $H_0 := \mu_1 = \mu_2$. סטטיסטי המבחן:

$$t = \frac{\bar{x} - \bar{y}}{s_p}, \qquad s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)$$

כאשר s_x^2 הוא שונות המדגם של כל ה-x_i, ו- s_y^2 הוא שונות המדגם של כל ה-y_i.

התפלגות האפס: $f(t|H_0)$ היא הPDF של $T \sim t(n+m-2)$.

דוגמה

מחקר אמיתי שבו 1408 נשים אושפזו בבית חולים ליולדות, או בגלל סיבות רפואיות (1) או במקרה חירום (2). אורך ההיריון נמדד בשבועות שלמים מתחילת הווסת האחרון. אפשר לסכם את הדאטא כך:

רפואי: 775 מקרים, עם $\bar{x}_M = 39.08,\ s_M^2 = 7.77$. חירום: 633 מקרים עם $\bar{x}_E = 39.6,\ s_E^2 = 4.95$

נרצה לבדוק האם הממוצע משתנה בין הקבוצות. השונות הכוללת היא:

$$s_p^2 = \frac{774(7.77) + 632(4.95)}{1406} = \left(\frac{1}{775} + \frac{1}{633}\right) = 0.0187$$

סטטיסטי המבחן t להשערת האפס הוא:

$$t = \frac{\bar{x}_M - \bar{y}_E}{s_p} = -3.8064$$

יש דרגת חופש 1406. נחשב את ה *p-value* הדו צדדי:

$$p = P(|T| > |t|) = 0.00015$$

p קטן יותר מ-$\alpha = 0.01$, אז נדחה את השערת האפס. כלומר יש הבדל בין הממוצעים.

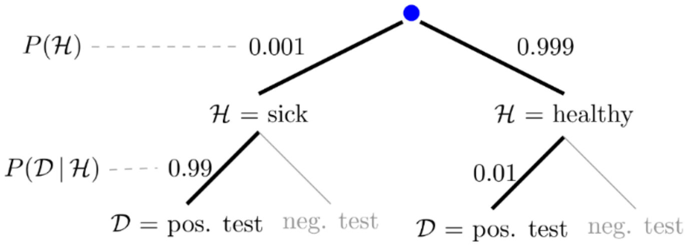
יכולנו גם לשים לב שעם 1406 דרגות חופש, אז ההתפלגות t היא כמעט נורמלית סטנדרטית. אז:

$$P(|t| > 3.8064) \approx P(|z| > 3.8064) < 0.01$$

הנחנו שלקבוצות יש את אותה שונות, אבל בהינתן ההבדל הגדול בין השונויות של המדגמים, יכול להיות שההנחה הזו לא לגיטימית. יש מבחנים אחרים שאפשר לבצע כדי לבדוק אם לקבוצות יש את אותה שונות. בפועל, כנראה נבצע אותם קודם כדי לראות אם מבחן t בכלל רלוונטי.

הרצאה

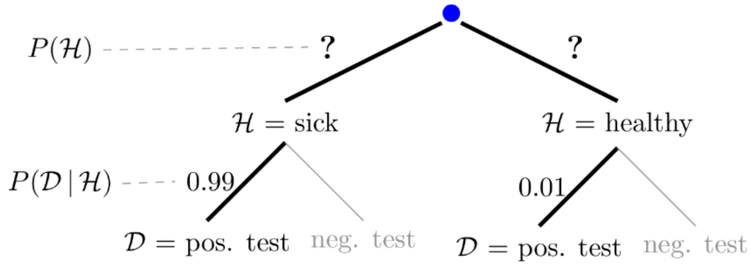
מצגת 7א, מצגת 7ב



ההבדל בין גישה בייסיאנית לשכיחותנית. לדוגמה אם בדיקה רפואית יצאה חיובית:

בגישה הבייסיאנית, אנחנו מניחים שה *prior* ידועה לנו. ואז נוכל להשתמש בנוסחת בייס:

$$P(\text{sick}|\text{positive test}) = \frac{0.001 \cdot 0.99}{0.001 \cdot 0.99 + 0.999 \cdot 0.01} \approx 0.1$$



לעומת זאת, הגישה השכיחותנית אומרת שאנחנו לא יודעים את ה *prior*. יודעים רק את ה *likelihood*.

תרגיל כיתה – ג'יין

כל יום, ג'יין מאחרת לשיעור ב-X שעות, כאשר $X \sim U(0,\theta)$ עם פרמטר θ לא ידוע. ג'ון ממדל את ההשערה שלו לגבי θ על ידי *prior*, $f(\theta)$. אחרי שג'יין מאחרת x שעות ביום למחרת, הוא מחשב את פונקציית ה *likelihood* $f(x|\theta)$ ואת ה *posterior PDF*, $f(\theta|x)$.

מבחינת הגישה השכיחותנית, רק ה *likelihood* היא חישוב תקין להסתברות, כי היא מחשבת את ההסתברות שנראה דאטא x בהינתן פרמטר קבוע θ. ה-*prior* וה-*PDF* הם הסתברויות על ערך של פרמטר לא ידוע, והגישה השכיחותנית לא מקבלת את זה, כי זה דורש שנאמין משהו לגבי הפרמטר, שאנחנו לא יודעים.

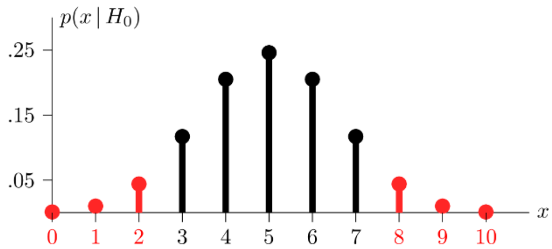
NHST

נזכיר: השערת האפס H_0, השערה אלטרנטיבית H_A. סטטיסטי המבחן x. תחום הדחייה: ערכים אפשריים של x, שאם נקבל אותם, נדחה את H_0. $p(x|H_0)$ או $f(x|H_0)$, התפלגות האפס.

דוגמה – מטבע

מטבע עם הסתברות θ לעץ. H_0: המטבע הוגן, $\theta = 0.5$. H_A: המטבע מוטה, $\theta \neq 0.5$. סטטיסטי המבחן: X, מספר העצים ב-10 הטלות. יש שתי גישות:

1. נבחר תחום דחייה ואז נחשב את המובהקות.
2. נבחר את המובהקות ואז נחשב את תחום הדחייה.



x	0	1	2	3	4	5	6	7	8	9	10
$p(x\|H_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

אם קבענו את תחום הדחייה, נגיד $\{0,1,2,8,9,10\}$:

מהי המובהקות? $P(\text{reject } H_0|H_0)$. ההסתברות שנדחה את השערת האפס היא ההסתברות שייצא x בתחום הדחייה, בהינתן H_0:

$$\alpha = (0.001 + 0.01 + 0.044) \cdot 2 = 0.11$$

אם קבענו את המובהקות, נגיד $\alpha = 0.05$:

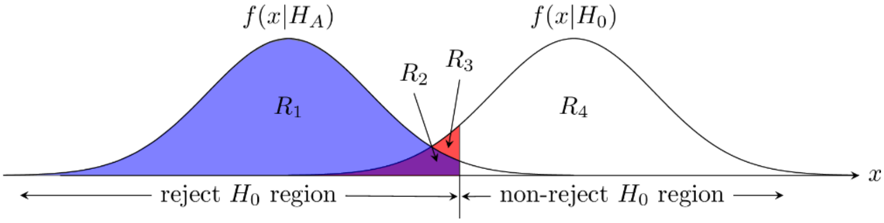
x	0	1	2	3	4	5	6	7	8	9	10
$p(x\|H_0)$	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

תחום הדחייה צריך להיות כך שסכום ההסתברות לדחייה בהינתן H_0, הוא לכל היותר α.

$$2(0.001 + 0.1) = 0.022 < \alpha < 0.11 = 2(0.001 + 0.01 + 0.044)$$

דוגמה

המובהקות היא ההסתברות שנדחה את H_0 גם אם הוא נכון, כלומר $R_2 + R_3$.



מבחן z, ערכי p

נניח שיש לנו דאטא בת"ל, בהתפלגות נורמלית: $x_1, \dots, x_n$, עם μ לא ידוע ו-σ ידועה.

ההשערות: H_0: $x_i \sim N(\mu_0, \sigma^2)$, H_A: דו צדדי $\mu \neq \mu_0$, או חד צדדי $\mu > \mu_0$. ערך z: $\bar{x}$ מנורמל: $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.

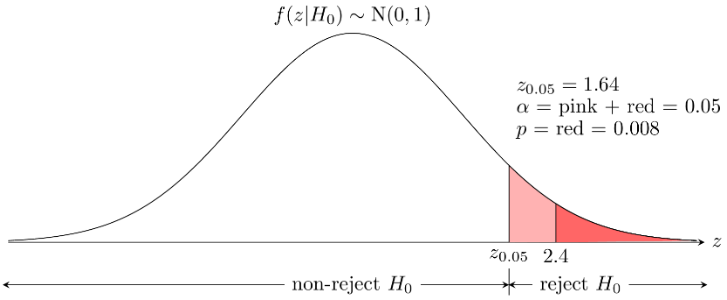
סטטיסטי המבחן: z. התפלגות האפס: בהנחה ש H_0 נכון, $z \sim N(0,1)$. ערכי p: חד צדדי ימני: $p = P(Z > z|H_0)$.

דו-צדדי: $p = P(|Z| > z|H_0)$. מובהקות: עבור $p \leq \alpha$, נדחה את H_0.

יכולנו גם להשתמש ב-$\bar{x}$ בתור סטטיסטי המבחן, ואז $N(\mu_0, \sigma^2)$ היא התפלגות האפס.

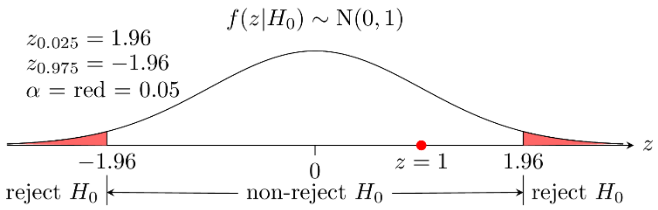
נניח ש $\sigma = 15$. נגדיר: השערת האפס H_0: $\mu = 100$.
השערה אלטרנטיבית חד צדדית H_A: $\mu > 100$.

אספנו 9 נקודות דאטא, $\bar{x} = 112$. כלומר $z = \frac{112-100}{15/3} = 2.4$.
במובהקות 0.05, האם נוכל לדחות את H_0?



תרגיל כיתה

נתון: השערת האפס H_0: הדאטא מתפלגת $N(5, 10^2)$. השערה אלטרנטיבית H_A: הדאטא מתפלגת $N(\mu, 10^2)$ עם $\mu \neq 5$. סטטיסטי המבחן: z, שהוא $\bar{x}$ מנורמל. הדאטא: 64 נקודות, עם ממוצע $\bar{x} = 6.25$. מובהקות נדרשת $\alpha = 0.05$.



תחת H_0, מתקיים: $z = \frac{\bar{x}-5}{10/\sqrt{64}} = \frac{\bar{x}-5}{5/4} \sim N(0,1)$. נמצא את תחום הדחייה: התחום שההסתברות שלו תחת $N(0,1)$ היא לכל היותר α.

לפי הדאטא, מתקיים $z = \frac{6.25-5}{5/4} = \frac{1.25}{1.25} = 1$. כלומר z לא בתחום הדחייה. נחשב *p-value* דו צדדי: $p = P(|Z| > 1) = 0.32$. זה יותר גדול מהמובהקות הנדרשת. שני הדברים האלה שקולים, והמשמעות היא שלא נדחה את H_0.

תרגיל כיתה – שני מטבעות

למטבע C_1 יש הסתברות 0.5 לעץ, למטבע C_2 יש הסתברות 0.6. נבחר אחד באופן מקרי ואחיד, נטיל 8 פעמים ונקבל 6 עץ.

יש שני דרכים ללכת בהן:

1. השערת האפס H_0: המטבע הוא C_1. ההשערה האלטרנטיבית H_A: המטבע הוא C_2. במובהקות 0.05, האם נדחה את H_0?
2. השערת האפס H_0: המטבע הוא C_2. ההשערה האלטרנטיבית H_A: המטבע הוא C_1. במובהקות 0.05, האם נדחה את H_0?

מתקיים:

k	0	1	2	3	4	5	6	7	8
$p(k\|\theta = 0.5)$	.004	.031	.109	.219	.273	.219	.109	.031	.004
$p(k\|\theta = 0.6)$	.001	.008	.041	.124	.232	.279	.209	.090	.017

עבור 1, מכיוון ש $0.6 > 0.5$, נשתמש בתחום דחייה חד צדדי ימני: $\{7,8\}$. לא נדחה את H_0 במקרה הזה.

עבור 2, תחום דחייה שמאלי: $\{0,1,2\}$. גם כאן, לא נדחה את H_0.

כלומר תלוי בהשערת האפס, קיבלנו מסקנה שונה. זה מבטא את אי הסימטריות של *NHST*: השערת האפס היא ההשערה השמרנית והבטוחה יותר. נדחה אותה רק אם הדאטא מאוד לא סבירה בהינתן H_0.

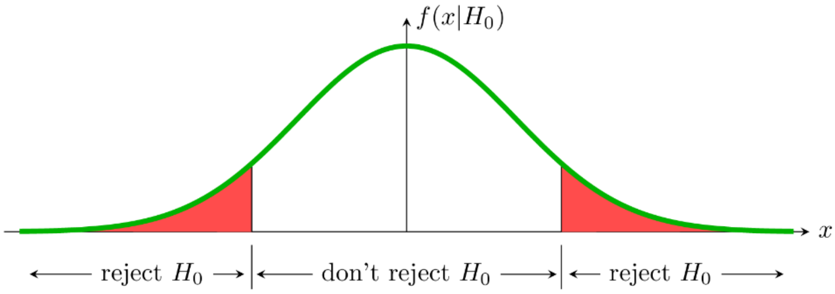
תזכורת

x – סטטיסטי המבחן.

$f(x|H_0)$ – ה PDF של התפלגות האפס (הקו הירוק)

תחום הדחייה הוא תחום של ציר ה-x.

מובהקות – ההסתברות מעל תחום הדחייה (השטח האדום)



השערות פשוטות ומורכבות

השערה פשוטה: התפלגות המדגם מוגדרת במפורש. בדרך כלל לפרמטר המעניין יש ערך ספציפי.

השערה מורכבת: התפלגות המדגם לא מוגדרת במפורש. בדרך כלל לפרמטר המעניין יש תחום ערכים.

דוגמה

למטבע יש הסתברות θ לעץ. נטיל 30 פעמים ויהי x מספר העצים. ההשערה ש $\theta = 0.4$ היא פשוטה. $x \sim bin(30, 0.4)$. ההשערה $\theta > 0.4$ היא מרוכבת. $x \sim bin(30, \theta)$.

ערך p

כלי כדי לבדוק את סטטיסטי המבחן נמצא בתחום הדחייה. מוגדר:

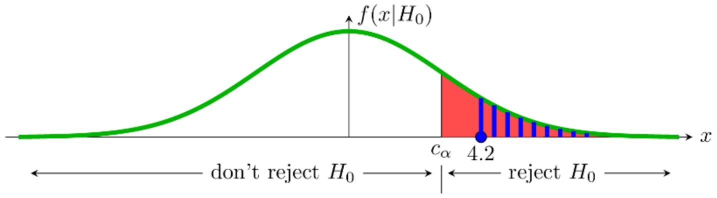
$$p = P(\text{data at least as extreme as } x \mid H_0)$$

המשמעות של "קיצוני לפחות כמו" משתנה לפי תחום הדחייה (חד או דו צדדי).

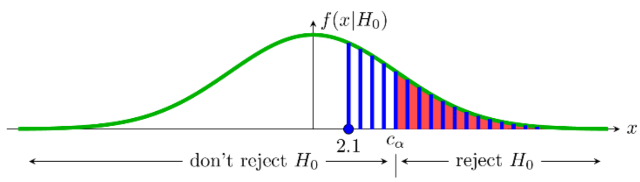
דוגמה

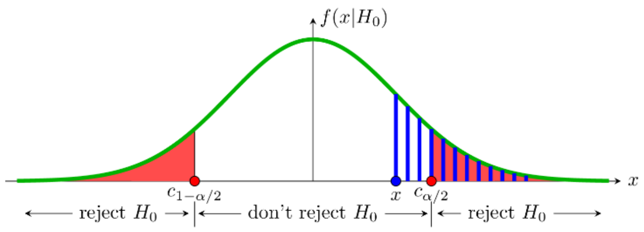
סטטיסטי המבחן נמצא בתחום הדחייה.
השטח הכחול (*p-value*) קטן מהשטח האדום (α).

בעצם, ההסתברות לקבל ערך קיצוני יותר מה-x שקיבלנו, קטנה יותר מההסתברות ש-x יצא בתחום הדחייה.



סטטיסטי המבחן לא בתחום הדחייה:





במקרה השני, אנחנו רוצים למצוא נוסחה ל *p-value*. נחשב את ההסתברות לקבל ערך קיצוני יותר מתחום הדחייה:

$$p = 2 \cdot \min(\text{left tail prob of } x, \text{right tail prob of } x)$$

כלומר, ניקח את ההסתברות הקטנה יותר מבין הזנבות, כפול 2.

ערכים קריטיים

הקצוות של תחום הדחייה. מתוייגים לפי ההסתברות מצד ימין שלהם. המשלים לשיברונים (*quantiles*). $c_{0.1} = q_{0.9}$. לדוגמה: עבור נורמלי סטנדרטי, $c_{0.025} = 1.96, c_{0.975} = -1.96$.

תרגול

תרגיל 1

יהיו $x_1 \dots x_n \sim N(\mu, \sigma^2)$, כאשר σ ידוע. נגיד $H_0 \coloneqq \mu \le 0, H_A \coloneqq \mu > 0$. סטטיסטי המבחן יהיה $\bar{x} = \Sigma x_i / n$. נגדיר תחום דחייה: $\bar{x} > c$. זה מבחן חד צדדי.

נחשב את פונקציית העוצמה: $\beta(\mu) = (\bar{x} > c | \mu) = P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} > \frac{c-\mu}{\sigma/\sqrt{n}} | \mu\right) = 1 - \Phi\left(\frac{c-\mu}{\sigma/\sqrt{n}}\right)$

נחשב את המובהקות: $\alpha = \sup_{\mu \in H_0} \beta(\mu)$. הפונקציה β עולה, אז ה μ המקסימלי הוא 0 (הערך המקסימלי של μ בהנחה ש H_0 נכונה)

$$\alpha = 1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right) \Longrightarrow \Phi\left(\frac{c\sqrt{n}}{\sigma}\right) = 1 - \alpha \Longrightarrow \frac{c\sqrt{n}}{\sigma} = \Phi^{-1}(1-\alpha) \Longrightarrow c = \frac{\Phi^{-1}(1-\alpha)\sigma}{\sqrt{n}}$$

תרגיל 2

ניקח דאטא $x_1 \dots x_n \sim U(0, \theta)$. בגלל שהדאטא לא מתפלגת נורמלי, המבחנים הרגילים לא מתאימים. $H_0 \coloneqq \theta = 0.5, H_A \coloneqq \theta > 0.5$. נגדיר $y = \max\{x_1 \dots x_n\}$. המבחן: $y > c$. נמצא את c. פונקציית העוצמה של H_0:

$$\alpha = \beta(H_0) = P(y > c \mid H_0) = 1 - P(y < c \mid H_0) = 1 - \Pi_{i=1}^{n} \text{clip}\left(\frac{c}{\theta}\right)^n = 1 - \Pi_{i=1}^{n} \text{clip}(2c)^n$$

$$c = 0.5 \cdot (1-\alpha)^{1/n}$$

$$\text{clip}(t) = \begin{cases} 1, & 1 < t \\ t, & 0 \le t \le 1 \\ 0, & t < 0 \end{cases}$$

נגיד ש $n = 20, y = 0.48$. נחשב את ה *p-value*: ה α הכי קטן שעבורו $y > 0.5 \cdot (1-\alpha)^{1/n}$.

$$y^n > 0.5(1-\alpha) \Longrightarrow 2y^n > 1 - \alpha \Longrightarrow \alpha > 1 - 2y^n$$

קיבלנו $\alpha > 1 - 2 \cdot 0.48^{20} = 0.55$.