פרק 4: ההנחות הקלאסיות וקריטריונים לאמידה

4.1. פתיחה

בפרק זה יוצגו "ההנחות הקלאסיות" העוסקות בהתנהגותה של ההפרעה המקרית ובניסוחו הפורמאלי של המודל. חשוב להדגיש כי הנחות אלה אינן דרושות לצורך האמידה עצמה, אלא נעזר בהן רק לאחריה לצורך בחינת תכונות האומדים. לאחר הצגת ההנחות נחזור לדון בבעיה המקורית: בהינתן מדגם של תצפיות כיצד יש לאמוד את הפרמטרים הלא ידועים? לצורך המחשה של בעית האמידה העומדת לפנינו נזכור כי הפרמטרים שברצונינו לאמוד מגדירים קו לינארי במישור X-Y ואילו המדגם מוצג "כענן תצפיות" באותו מישור באמצעות דיאגרמת הפיזור. בעית האמידה היא, אם כן, כיצד להעביר את הקו הטוב ביותר דרך ענן התצפיות כך שיציג בדרך הטובה ביותר את קשר בין המשתנים במדגם.

4.2. ההנחות הקלאסיות

4.2.1. הנחה 1: נכונות המודל

הנחה: המודל $Y_i = \alpha + \beta X_i + \varepsilon_i$ מתאר נכונה את התנהגות המשתנה Y.

הנחה זו גורסת כי ברשותינו המודל הנכון. בפרט, אין גורמים נוספים פרט ל-X המשפיעים באופן סיסטמתי על Y, כמובן שקיימת גם השפעתה של ההפרעה המקרית אך היא נובעת מגורמים אקראיים שלא ניתן לאתרם באופן עקבי. בנוסף, השפעתו של X על Y היא לינארית. הנחה זו גורסת כי **הספציפיקציה** הנכונה של המודל היא לינארית וידועה לנו במלואה.

4.2.2. הנחה 2: תוחלת אפס להפרעה המקרית

הנחה: $E(\varepsilon_i) = 0 \quad \forall i$.[1]

הנחה זו נועדה לצורכי נוחות בלבד. אם לדוגמא נניח כי $E(\varepsilon_i) = \mu \neq 0$, הרי שנוכל להציג את ההפרעה המקרית ע"י המשוואה $\varepsilon_i = \mu + v_i$, כאשר v_i הוא גורם מקרי. במצב זה נוכל להציג את המודל כך: $Y_i = \alpha + \mu + \beta X_i + v_i$, כאשר הפעם v_i הוא הפרעה מקרית שתוחלתה אפס, שכן $v_i = \varepsilon_i - \mu$ ולכן $E(v_i) = E(\varepsilon_i - \mu) = \mu - \mu = 0$. במודל החדש, החותך, כלומר ערכו של Y כאשר $X = 0$, מיוצג על ידי הביטוי $\alpha + \mu$, במודל זה אמנם לא נוכל לאמוד את α ו-μ בנפרד אלא רק את סכומם,[2] אך הנקודה החשובה לענייננו היא שכעת קיבלנו מודל חדש עם הפרעה מקרית, v_i, בעלת תוחלת אפס כפי שהנחנו.

---

[1] הסימן $\forall$ משמעותו "לכל", לכן הכתיבה $\forall i$ משמעותה לכל התצפיות, כלומר $i = 1,2,3,\ldots,n$.
[2] את הסיבה לכך נראה בהמשך, כאשר נדון במולטיקולינאריות.

4.2.3. הנחה 3: שונות קבועה

הנחה: $Var(\varepsilon_i) = \sigma^2 \quad \forall i$, $\sigma^2 < \infty$.

הנחה זו דורשת כי לכל ההפרעות המקריות במדגם אותה שונות. משמעות ההנחה היא שמידת הדיוק של התצפיות השונות בתיאור המודל זהה, ומכאן שבתהליך האמידה אין סיבה לתת משקל יתר לתצפיות מסויימות על פני אחרות. אם לדוגמא, היינו יודעים כי עבור תצפיות מסויימות שונות ההפרעה המקרית גדולה יותר לעומת תצפיות אחרות, הרי שהיינו רוצים שתצפיות אלה יקבלו משקל מופחת בתהליך האמידה שכן הן פחות מדוייקות בתיאור החלק הסיסטמטי של המודל אותו אנו אומדים.

4.2.4. הנחה 4: חוסר מתאם בין ההפרעות המקריות

הנחה: $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$.

ההפרעות המקריות מהתצפיות השונות אינן מתואמות זו עם זו, כלומר שבהפרעה המקרית של תצפית i, ε_i, אין מידע התורם להערכת ההפרעה המקרית בתצפית אחרת, ε_j, עבור $i \neq j$. נשים לב כי עבור $i = j$ נקבל $Cov(\varepsilon_i, \varepsilon_i) = Var(\varepsilon_i)$, וכפי שראינו בהנחה 3 $Var(\varepsilon_i) = \sigma^2$.

4.2.5. הנחה 5: חוסר מתאם בין ההפרעות המקריות והמשתנים המסבירים

הנחה: $Cov(\varepsilon_i, X_j) = 0 \quad \forall i, j$.

כלומר, ההפרעה המקרית אינה מתואמת עם המשתנה המסביר. יש לשים לב לאינדקסים של ההפרעה המקרית והמשתנה המסביר; הנחה זו גורסת כי לא רק שאין מתאם בו זמני בין ε ובין X, אלא אף יותר מכך, ההפרעה בכל תצפית נתונה איננה מתואמת עם אף משתנה מסביר בכל התצפיות האחרות.

4.3. הנחות נוספות

בהתקיים ההנחות אשר הוצגו עד כה האומדים שנקבל בהמשך אכן יקיימו את אותן תכונות רצויות אשר מנינו בפרק הקודם. יחד עם זאת, נהוג להניח עוד שתי הנחות אשר אינן נכללות בהנחות הקלאסיות. הנחות אלה מקלות על הטיפול באומדים ועל הניתוח שלאחר האמידה.

4.3.1. הנחה I: נורמליות

הנחה: $\varepsilon_i \sim N$.

כלומר, ההפרעה המקרית מתפלגת נורמלית. אם נצרף לכך את הנחות 2 ו-3 נקבל $\varepsilon_i \sim N(0, \sigma^2)$.
חשוב לדעת את התפלגות ההפרעה המקרית מכיוון שהתפלגות זו תועבר לאומדים שנקבל (במידה והם אכן לינארים). ללא הנחת ההתפלגות לא נוכל לבצע מבחני השערות על האומדים. הנחה זו חשובה כאשר יש בידינו מדגמים קטנים; במדגמים גדולים ניתן לחשב את ההתפלגות האסימפטוטית של האומדים גם ללא הנחה ספציפית על התפלגות ההפרעה המקרית אולם אנו לא נעסוק בכך ולצורך הפשטות ניעזר בהנחת הנורמליות.

כמו כן, בשילוב עם ההנחות הקלאסיות 2-4 מתקבל כי ההפרעות המקריות הן משתנה מקרי **iid** (Identical Independent Distributed), כלומר שכל ההפרעות באות מאותה התפלגות והן בלתי תלויות זו בזו.

4.3.2. הנחה II: המשתנים המסבירים

הנחה: ה-X-ים אינם משתנים מקריים.

תחילה יש להדגיש כי הנחה זו נועדה לצורכי נוחות בלבד. משמעות ההנחה היא כי לחוקר שליטה מלאה על ערכי המשתנה המסביר (X) בתצפיות השונות. מובן כי במציאות אין הדבר כך, לפחות לא בכלכלה. כך למשל, באמידת השפעת הריבית (משתנה מסביר, X) על האינפלציה (משתנה תלוי, Y) אין החוקר יכול לבחור את ערכי הריביות במדגם, אלא הוא צופה בהן לאחר התממשותן ללא כל יכולת להשפיע על ערכיהן. כלומר, המשתנה המסביר בדוגמא זו הוא משתנה מקרי עבור החוקר.

יחד עם זאת, קיימות דיסיפלינות בהן ערכי המשתנים המסבירים אכן נקבעים ע"י החוקר, הדבר מתרחש בעיקר בניסויי מעבדה. למשל, אם נרצה לבדוק את השפעת כמות המים על התפוקה המתקבלת מגידול כלשהו, נוכל להשקות חלקות שונות בכמויות מים שונות על פי בחירתנו, כך שאנו שולטים בצורה מלאה על כמות המים בכל חלקה. כלומר, אנו שולטים בערכי המשתנה המסביר, ולכן בדוגמא זו ה-X-ים אינם משתנים מקריים עבור החוקר אלא קבועים.

יש לציין כי תכונות האומדים שנקבל לא ישתנו גם אם נניח כי ה-X-ים הם משתנים מקריים, אך הוכחת התכונות תהיה קשה יותר. כלומר, הנחה זו נועדה לצורכי נוחות החישובים בלבד.

4.4. קריטריונים לאמידה

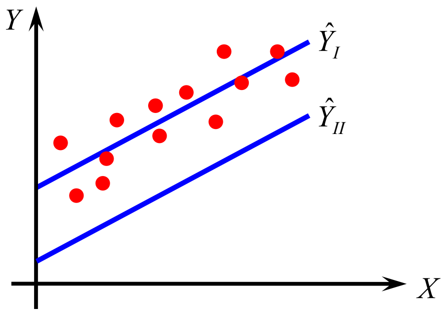
כזכור, בהתאם להנחה הקלאסית הראשונה, המודל מניח כי המשתנה Y נקבע כפונקציה ליניארית של X ולכן גראפית הוא ניתן לתיאור באמצעות קו ישר במישור X-Y. הבעיה העומדת בפנינו היא בחירת האומדים "הטובים ביותר" לחותך הקו, α, ולשיפועו, β. מכיוון שהמודל הנאמד לינארי, הרי שבחירת אומדים אלה כמוה כבחירת הקו כולו. בחלק זה נבחן בצורה אינטואיטיבית קריטריונים לבחירת הקו "הטוב ביותר", ובפרקים הבאים נמצא את האומדים $\hat{\alpha}$ ו- $\hat{\beta}$ הנגזרים מקריטריון האמידה הנבחר ונבחן את תכונותיהם.

כעת ניגש למציאת קריטריון אמידה, תחילה ננקוט בגישה אינטואיטיבית באמצעות בחינה גראפית של הקו הנבחר מול דיאגרמת הפיזור של המדגם, ולאחר מכן ננסח אותו בצורה פורמלית.

נבחן את דיאגרמת הפיזור המוצגת באיור 4.1. באיור מוצעים שני קוים, השאלה העומדת בפנינו בשלב זה היא לקבוע מי מהם מתאר טוב יותר את הקשר בין X ל-Y על בסיס המדגם, $\hat{Y}_I$ או $\hat{Y}_{II}$?

כאשר $\hat{Y} = \hat{\alpha} + \hat{\beta}X$, כלומר $\hat{Y}$ הוא האומד ל-Y בהינתן X, או במילים אחרות $\hat{Y}$ הוא הקו הנאמד.

איור 4.1: בחירת קו לתיאור הקשר בין X ל-Y



מהתבוננות באיור 4.1 מובן כי נעדיף את הקו $\hat{Y}_I$ על פני $\hat{Y}_{II}$, שכן אם נשתמש ב-$\hat{Y}_{II}$ הרי שככל הנראה ברוב המקרים אם ננסה להעריך את Y בהינתן ערך כלשהו של X נקבל כי ערכו הנאמד של Y נמוך מדי. כלומר, אמידת Y באמצעות $\hat{Y}_{II}$ מוטה כלפי מטה. אם נשתמש ב-$\hat{Y}_I$ כנראה **שבממוצע** לא נטעה, לעיתים נעריך את Y יתר על המידה ולעיתים נייחס לו ערך נמוך מידי אך הטעות הממוצעת תהיה קרובה לאפס.

כעת, ננסה לתת ביטוי פורמאלי לאינטואיציה שתוארה לעיל. לשם כך נגדיר את השגיאה הנאמדת (נהוג לסמנה ב-$\hat{\varepsilon}_i$ או e_i). השגיאה הנאמדת היא ההפרש בין ערכו של המשתנה המוסבר, Y_i, וערך אומדנו, $\hat{Y}_i$, כלומר:

$$e_i \equiv Y_i - \hat{Y}_i$$

כאשר $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$. באיור 4.1 קל לראות כי עבור $\hat{Y}_{II}$ מתקבל כי בכל התצפיות $e_i > 0$, לעומת זאת עבור הקו $\hat{Y}_I$ מתקבל כי חלק מהשגיאות הנאמדות חיובי וחלק אחר שלילי. יתרה מזאת, נרצה כי ההפרעות החיוביות "יתקזזו" עם השליליות. מכאן שנחפש אומדים אשר יתנו כי סכום ההפרעות הנאמדות יהיה אפס.

קריטריון ראשון לאמידה: נבחר את $\hat{\alpha}$ ו-$\hat{\beta}$ כך שנקבל $\sum_{i=1}^{N} e_i = 0$

בשלב זה בו יש בידינו קריטריון לאמידה, היינו רוצים למצוא את אותם $\hat{\alpha}$ ו-$\hat{\beta}$ אשר יביאו את סכום השגיאות הנאמדות לאפס. אולם, אם ננסה לעשות זאת נגלה כי קיימים אינסוף צירופים של $\hat{\alpha}$ ו-$\hat{\beta}$ אשר מקיימים קריטריון זה. משפט 4.1 מנסח את הדברים בצורה מדויקת יותר.

משפט 4.1: כל קו העובר דרך נקודת הממוצעים $(\bar{X}, \bar{Y})$, כלומר כל קו המקיים $\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$, מביא את סכום הסטיות הנאמדות לאפס, כלומר: $\sum e_i = 0$.

הוכחה: קו העובר דרך נקודת הממוצעים מקיים: $\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$. נכפיל את שני האגפים ב-$N$ ונעביר אגפים לקבלת:

$$\sum Y_i - N\hat{\alpha} - \hat{\beta}\sum X_i = 0$$

נכנס איברים תחת סימן סכימה אחד:

$$\sum (Y_i - \hat{\alpha} - \hat{\beta}X_i) = 0$$

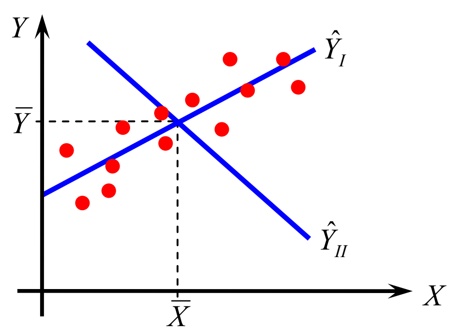
אך מכיוון שבהגדרה $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$, וכן $e_i \equiv Y_i - \hat{Y}_i$, אנו מקבלים:

$$\sum e_i = 0$$

■

המסקנה העולה מתוצאה זו היא שקריטריון זה איננו חד מספיק כיוון שקיימים אינסוף צירופים של $\hat{\alpha}$ ו-$\hat{\beta}$ המקיימים אותו. בכדי להבין טוב יותר את הבעייתיות שבקריטריון די לבחון את איור 4.2 המציג שני קווים שונים אשר שניהם מקיימים $\sum e_i = 0$. נראה כי קל יהיה להגיע להסכמה כי באיור הקו $\hat{Y}_I$ טוב יותר לתיאור הקשר בין X ל-Y, למרות ששני הקווים המוצגים מקיימים את קריטריון האמידה הראשון.

איור 4.2: שני קווים המקיימים $\sum e_i = 0$



מהאמור לעיל עולה כי יש לשכלל את קריטריון האמידה שהוצע כך שיבחר רק קו אחד. נחזור לעיין באיור 4.2 וננסה להבין, ולנסח בצורה פורמלית, מדוע הקו $\hat{Y}_I$ עדיף על $\hat{Y}_{II}$. קל לראות כי התצפיות קרובות יותר ל-$\hat{Y}_I$, או במילים אחרות הפיזור של התצפיות לאורך $\hat{Y}_I$ קטן יותר בהשוואה ל-$\hat{Y}_{II}$. אמנם שני הקווים מקיימים את המשוואה $\sum e_i = 0$, ולכן בשניהם שגיאות חיוביות מתקזזות עם שליליות, אך אם נתייחס גם לערכן המוחלט של השגיאות נמצא כי **סכום הערכים המוחלטים** שלהן לאורך $\hat{Y}_{II}$ גדול יותר. מכאן שאולי נהיה מעוניינים באומדים אשר יניבו סכום ערכים מוחלטים קטן ככל האפשר לשגיאות הנאמדות.

קריטריון שני לאמידה: נבחר את $\hat{\alpha}$ ו- $\hat{\beta}$ כך שיביאו למינימום את הביטוי $\sum_{i=1}^{N}|e_i|$

אם ננסה ליישם קריטריון זה, נרצה לגזור את הביטוי $\sum_{i=1}^{n}|e_i|$ לפי $\hat{\alpha}$ ו- $\hat{\beta}$ ולהשוות את הנגזרות לאפס. אולם אז נתקל בבעיה כיוון שפונקצית הערך המוחלט איננה גזירה בנקודת האפס, דבר אשר יסרבל את החישובים. מכאן, שגם אם קריטריון זה יניב אומדים בעלי תכונות רצויות, קשה יהיה להשתמש בו. לכן, נרצה לשנותו במעט רק בכדי שהטיפול הטכני בבעית האמידה יהיה נח יותר.

נרצה לשמור על העיקרון העומד מאחורי קריטריון האמידה השני, אך יחד עם זאת להתגבר על הקושי הטכני הכרוך בו. בחירה טבעית במקום פונקצית הערך המוחלט היא העלאת כל שגיאה נאמדת בריבוע, ובמקום למזער את סך הערכים המוחלטים נשאף להביא למינימום את סכום ריבועי השגיאות. בדרך זו אנו משתמשים בפונקציה גזירה, דבר שיקל עלינו בבואנו לפתור את בעית המינימיזציה, ובה בעת אנו שומרים על האינטואיציה שהנחתה אותנו בבחירת קריטריון האמידה השני אלא שבמקום מרחק כל תצפית מהקו אנו משתמשים בריבוע המרחק.

קריטריון שלישי לאמידה: נבחר את $\hat{\alpha}$ ו- $\hat{\beta}$ כך שיביאו למינימום את הביטוי $\sum_{i=1}^{N} e_i^2$

קריטריון זה נקרא קריטריון ריבועים פחותים (Least Squares), ואנו נבחר בו לצורך אמידת הפרמטרים הלא ידועים. האומדים המתקבלים כתוצאה משימוש בקריטריון זה נקראים **אומדי ריבועים פחותים (*LSE - Least Squares Estimators*)**.

לסיום יש להעיר כי קריטריון זה נותן משקל גדול יותר לתצפיות הרחוקות יחסית מהקו הנאמד עקב העלאתן של השגיאות בריבוע. תכונה זו איננה רצויה שכן אולי נרצה לתת משקל גדול יותר דווקא לתצפיות הנמצאות קרוב יותר לקו כיוון שהן מייצגות טוב יותר את התנהגות המדגם. למרות זאת נראה בפרקים הבאים כי קריטריון הריבועים הפחותים אכן מניב אומדים בעלי כל אותן תכונות רצויות אשר מנינו בפרק הקודם.

סיכום פרק 4:

- ההנחות הקלאסיות:
  - המודל הנכון הוא: $Y_i = \alpha + \beta X_i + \varepsilon_i$
  - $\forall i \quad E(\varepsilon_i) = 0$
  - $\forall i \quad Var(\varepsilon_i) = \sigma^2$
  - $\forall i \neq j \quad Cov(\varepsilon_i, \varepsilon_j) = 0$
  - $\forall i, j \quad Cov(\varepsilon_i, X_j) = 0$
- הנחות נוספות:
  - $\varepsilon_i \sim N$
  - ה-X-ים אינם משתנים מקריים.
- השגיאה המקרית **הנאמדת** תסומן ב-e_i, כאשר $e_i = Y_i - \hat{\alpha} + \hat{\beta} X_i$.
- האומדים יבחרו על פי קריטריון ריבועים פחותים, כלומר נחפש את $\hat{\alpha}$ ו-$\hat{\beta}$ אשר מביאים למינימום את סכום ריבועי הסטיות הנאמדות:

$$\underset{\hat{\alpha},\hat{\beta}}{Min} \sum_{i=1}^{N} e_i^2$$