פרק 5: אמידה בשיטת הריבועים הפחותים

5.1. פתיחה

פרק זה משתמש בקריטריון הריבועים הפחותים שהוצג בפרק הקודם למציאת אומדים לפרמטרים הלא ידועים, α ו-β, במודל $Y_i = \alpha + \beta X_i + \varepsilon_i$. שיטת האמידה נקראת ריבועים פחותים רגילים (OLS – Ordinary Least Squares), והאומדים המתקבלים ממנה נקראים אומדי ריבועים פחותים (LSE – Least Squares Estimators).

הפרק מתחיל בפיתוח האומדים באמצעות פתרון בעית מינימיזציה לסכום ריבועי השגיאות הנאמדות, לאחר מכן נדון במשמעותם של תנאי סדר ראשון הנגזרים מבעית המינימיזציה, ולבסוף נציג דרכים אלטרנטיביות לכתיבת האומדים. אולם לפני שנפתח יש להציג סימון מקובל לכתיבה מקוצרת אשר יקל על כתיבת נוסחאות בהמשך.

סימון: אותיות קטנות ייצגו מרחק של משתנה מערכו הממוצע במדגם, כך למשל $x_i \equiv X_i - \overline{X}$.

5.2. פיתוח אומדי ריבועים פחותים

בהינתן המודל $Y_i = \alpha + \beta X_i + \varepsilon_i$ אנו מחפשים את $\hat{\alpha}$ ו-$\hat{\beta}$ אשר מביאים למינימום את סכום ריבועי הסטיות הנאמדות, $\sum_{i=1}^{N} e_i^2$, כאשר $e_i \equiv Y_i - \hat{\alpha} - \hat{\beta} X_i$ ו-N הוא מספר התצפיות.

פורמאלית ניתן לכתוב את בעית האמידה כך:

$$\underset{\hat{\alpha},\hat{\beta}}{Min} \sum_{i=1}^{N} (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2$$

תחת הצגה זו, הבעיה הפכה לטכנית במהותה. כל שנשאר לעשות הוא לגזור את הביטוי אותו ברצוננו להביא למינימום (סכום ריבועי הסטיות הנאמדות) לפי $\hat{\alpha}$ ו-$\hat{\beta}$, להשוות את הנגזרות הראשונות לאפס, ולחלץ ממשוואות אלה את הנעלמים המבוקשים ($\hat{\alpha}$ ו-$\hat{\beta}$). כמו כן, נצטרך לוודא כי אכן מדובר בנקודת מינימום, כלומר לבדוק אם תנאי סדר שני מתקיימים.[1]

5.2.1. החותך, $\hat{\alpha}$

נגזור את סכום ריבועי השגיאות לפי $\hat{\alpha}$, נשווה את הנגזרת לאפס, ונקבל:

$$\frac{\partial\{\cdot\}}{\partial\hat{\alpha}} = -2\sum (Y_i - \hat{\alpha} - \hat{\beta} X_i) = 0$$

חלוקת שני האגפים ב-2– ופתיחת סימן הסכום מניבה:

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

[1] בדיקת תנאי סדר שני מוצגת בנספח 4.1 בסוף פרק זה.

כעת נבודד את $\hat{\alpha}$ ונקבל את אומד הריבועים הפחותים לחותך:

$$\boldsymbol{\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}}$$

5.2.2. השיפוע, $\hat{\beta}$

נגזור את סכום ריבועי השגיאות לפי $\hat{\beta}$, נשווה את הנגזרת לאפס, ונקבל:

$$\frac{\partial\{\cdot\}}{\partial\hat{\alpha}} = -2\sum(Y_i - \hat{\alpha} - \hat{\beta}X_i)X_i = 0$$

חלוקה ב-2– והצבת הביטוי שמצאנו ל- $\hat{\alpha}$, מניבה:

$$\sum[Y_i - (\bar{Y} - \hat{\beta}\bar{X}) - \hat{\beta}X_i]X_i = \sum[(Y_i - \bar{Y}) - \hat{\beta}(X_i - \bar{X})]X_i = 0$$

כעת נשתמש באותיות קטנות לייצוג מרחק המשתנים מערכם הממוצע במדגם כפי שהוצג בפתיחת פרק זה, ולכן נכתוב:

$$\sum(y_i - \hat{\beta}x_i)X_i = 0$$

פתיחת הסוגריים וסימן הסכום מניבות:

$$\sum y_iX_i - \hat{\beta}\sum x_iX_i = \sum x_iy_i - \hat{\beta}\sum x_i^2 = 0$$

כאשר במעבר מאותיות גדולות לקטנות השתמשנו בתוצאות שהוצגו בפרק המבוא. לבסוף, מבידוד $\hat{\beta}$ נקבל את אומד הריבועים הפחותים לשיפוע:

$$\boldsymbol{\hat{\beta} = \frac{\sum x_iy_i}{\sum x_i^2}}$$

קיבלנו אם כן אומדים ל-α ו-β, באמצעותם ניתן לתאר את הקשר **הנאמד** בין X ל-Y באמצעות המשוואה $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$. כלומר, בהינתן ערך כלשהו של X נוכל להעריך את ערכו של Y. הקו המתואר באמצעות המשוואה $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$ נקרא **קו הרגרסיה**. יש להדגיש כי קו הרגרסיה אינו מתאר במדויק את הקשר בין המשתנים כפי שהוא מתרחש במציאות, אלא הוא מהווה אומדן בלבד לקשר הנ"ל.

מכאן עולה השאלה, אם תהליך האמידה אינו מניב את הקשר האמיתי בין המשתנים במה הוא תורם להבנת קשר זה? התשובה טמונה בתכונות האומדים, אנו נראה כי אומדי ריבועים פחותים אכן מקיימים את כל אותן תכונות רצויות אותן מנינו בפרק 3 (במידה וההנחות הקלאסיות מתקיימות). כך למשל, אם נווכח כי האומדים חסרי הטיה נוכל להסיק כי אמנם באמידה בודדת סביר כי איננו מדייקים בתיאור הקשר בין X ל-Y, אך אין באמידה טעות סיסטמטית המביאה להטיה כלשהי בתיאור הקשר הנ"ל. הדיון בתכונות האומדים יעשה בפרק הבא, אך לפני כן נחזור ונתעכב על תנאי הסדר הראשון שהתקבלו במהלך האמידה.

5.3. תנאי סדר ראשון: המשוואות הנורמליות

נשים לב שמגזירה לפי $\hat{\alpha}$ והשוואת הנגזרת לאפס קיבלנו $\sum (Y_i - \hat{\alpha} - \hat{\beta}X_i) = 0$. הביטוי בסוגריים הוא e_i, ולכן:

$$\sum e_i = 0$$

משוואה זו נקראת **המשוואה הנורמלית הראשונה**.

כזכור בפרק 4, כאשר חיפשנו קריטריון לאמידה, רצינו בשלב ראשון כי סכום השגיאות הנאמדות יהיה אפס. בחינת המשוואה הנורמלית הראשונה מעלה כי שימוש בקריטריון הריבועים הפחותים אכן מניב תוצאה זו למרות שלא דרשנו אותה, היא מתקבלת כתוצר לוואי של האמידה. יש לשים לב כי משוואה זו נובעת מהנגזרת לפי החותך, $\hat{\alpha}$, ומכאן שרק קיומו של חותך במודל מבטיח כי סכום השגיאות הנאמדות יהיה אפס. כמו כן, בפרק הקודם ראינו כי כל קו המקיים $\sum e_i = 0$ עובר דרך נקודת הממוצעים $(\bar{X}, \bar{Y})$ ומכאן שגם אומדי ריבועים פחותים מניבים קו העובר דרך נקודה זו.[2]

נקודה נוספת שיש להעיר ביחס למשוואה הנורמלית הראשונה היא הדמיון הרב בינה ובין ההנחה הקלאסית כי תוחלת ההפרעה המקרית הוא אפס, $E(\varepsilon_i) = 0 \quad \forall i$. אם נחלק את שני אגפי המשוואה הנורמלית הראשונה ב-N הרי שנקבל $\bar{e} = 0$, כלומר ממוצע ההפרעות הנאמדות **במדגם** הוא אפס. הדמיון בין ההנחה הקלאסית למשוואה הנורמלית נובע מכך כי "האומד הטבעי" לתוחלת הוא הממוצע. תוצאה זו מרמזת על שיטת אמידה נוספת: ניתן לתרגם את ההנחות הקלאסיות למשוואות אשר מהן נחלץ את האומדים. שיטת אמידה זו נקראת שיטת המומנטים, אך אנו לא נתמקד בה.

כעת נעבור לדיון בתנאי סדר ראשון ביחס ל-$\hat{\beta}$. מגזירה לפי $\hat{\beta}$ והשוואת הנגזרת לאפס קיבלנו $\sum (Y_i - \hat{\alpha} - \hat{\beta}X_i)X_i = 0$. הביטוי בסוגריים הוא e_i, ולכן:

$$\sum e_i X_i = 0$$

משוואה זו נקראת **המשוואה הנורמלית השניה**. גם משוואה זו מזכירה את אחת ההנחות הקלאסיות, הפעם זוהי ההנחה כי ההפרעה המקרית אינה מתואמת עם המשתנה המסביר, $Cov(\varepsilon_i, X_j) = 0 \quad \forall i, j$, לפחות עבור X-ים ו-ε-ים הבאים מאותה תצפית (i=j). נשים לב שמכיוון שתוחלת ההפרעה המקרית היא אפס הרי ש: $Cov(\varepsilon_i, X_i) = E(\varepsilon_i \cdot X_i)$; כעת אם נחלק את שני אגפי המשוואה הנורמלית השניה ב-N הרי שנקבל כי אומדן השונות המשותפת בין X ל-ε במדגם הוא אפס. שוב, הדמיון בין ההנחה הקלאסית למשוואה הנורמלית נובע מכך כי "האומד הטבעי" לתוחלת הוא הממוצע. מכאן שגם במקרה זה אומד הריבועים הפחותים זהה לאומד שיטת המומנטים המתרגמת את ההנחה בדבר השונות המשותפת באוכלוסיה בין ההפרעות המקריות למשתנה המסביר לזו המתקבלת במדגם.

---

[2] למעשה בפרק 4 הראינו את הקשר ההפוך: כל קו העובר דרך נקודת הממוצעים מקיים $\sum e_i = 0$, אך בכדי להוכיח את הכיוון ההפוך כל שיש לעשות הוא להתחיל מכך שסכום ההפרעות הנאמדות הוא אפס ולבצע את אותן פעולות בסדר הפוך.

שתי המשוואות הנורמליות מצביעות על תכונות "יפות" של אומדי הריבועים הפחותים. משוואות אלה מבטיחות כי ההפרעות הנאמדות יקיימו **במדגם** שתי תכונות מאלה שהנחנו לגבי התנהגות ההפרעה המקרית באוכלוסיה, וזאת אף על פי שלא דרשנו אותן במפורש במהלך האמידה. עם זאת, יש לשים לב כי דווקא קיומן של תכונות אלה במדגם מקשה עלינו לבדוק האם ההנחות אכן מתקיימות במציאות, שכן קיום התכונות במדגם מובטח בין אם ההנחות נכונות או שגויות.

5.4. ביטויים אלטרנטיבים לאומדי ריבועים פחותים

בחלק זה נציג שני ביטויים נוספים לכל אחד מהאומדים, $\hat{\alpha}$ ו- $\hat{\beta}$. תחילה נציג את האומדים כפונקציה לינארית של ה-Y-ים במדגם, ולאחר מכן יוצגו האומדים כפונקציה של הפרמטרים האמיתיים בתוספת איבר מקרי כלשהו. ביטויים אלה יוכחו כנוחים לעבודה בפרקים הבאים כאשר ניגש לחקירת תכונות האומדים, לכן כדאי להתעכב כעת ולפתחם.

5.4.1. אומדים לינארים

כעת נציג את האומדים כפונקציה לינארית של ה-Y-ים במדגם. נתחיל מהאומד לשיפוע, $\hat{\beta}$.

כזכור $y_i = Y_i - \bar{Y}$, נציב את y_i בביטוי שמצאנו עבור $\hat{\beta}$ ונקבל:

$$\hat{\beta} = \frac{\sum x_i\left(Y_i - \bar{Y}\right)}{\sum x_i^2} = \frac{\sum x_i Y_i - \bar{Y}\sum x_i}{\sum x_i^2}$$

הביטוי $\sum x_i$ שווה לאפס (סכום סטיות מהממוצע), ולכן:

$$\hat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2} = \frac{x_1 Y_1 + \ldots + x_N Y_N}{\sum x_i^2} = \frac{x_1}{\sum x_i^2} Y_1 + \ldots + \frac{x_N}{\sum x_i^2} Y_N$$

או בכתיב מקוצר:

$$\hat{\beta} = \sum\left(\frac{x_i}{\sum x_i^2}\right) Y_i$$

בהצגה זו ניתן לראות כי האומד לשיפוע הוא קומבינציה לינארית של ה-Y-ים במדגם, כאשר כל Y_i מוכפל בקומבינציה של ה-X-ים, $x_i \big/ \sum x_i^2$, אשר על פי הנחותינו אינם משתנים מקריים.

נעבור כעת להצגת החותך כקומבינציה לינארית של ה-Y-ים. מאמידה בריבועים פחותים קיבלנו $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$. נציב את הביטוי ל- $\hat{\beta}$ שמצאנו זה עתה ואת הנוסחה ל- $\bar{Y}$ לקבלת:

$$\hat{\alpha} = \underbrace{\sum Y_i / N}_{\bar{Y}} - \bar{X}\underbrace{\sum\left(\frac{x_i}{\sum x_i^2}\right) Y_i}_{\hat{\beta}}$$

כעת, מכינוס שני הביטויים תחת סימן סכימה אחד והוצאת Y_i כגורם משותף, נקבל את הביטוי המבוקש:

$$\hat{\alpha} = \sum \left(\frac{1}{N} - \frac{\bar{X} \cdot x_i}{\sum x_i^2} \right) Y_i$$

כלומר שגם החותך הוא קומבינציה לינארית של ה-Y-ים במדגם.

סיכום ביניים: הראינו אם כן, כי גם את החותך וגם את השיפוע ניתן להציג **כסכום משוקלל** של ה-Y-ים במדגם, כאשר המשקולות מורכבים מגורמים שאינם מקריים (מה-X-ים במדגם), באופן הבא:

$$\boldsymbol{\hat{\alpha} = \sum \left(\frac{1}{N} - \frac{\bar{X} \cdot x_i}{\sum x_i^2} \right) Y_i \qquad , \qquad \hat{\beta} = \sum \left(\frac{x_i}{\sum x_i^2} \right) Y_i}$$

5.4.2. פירוק האומדים למרכיב קבוע וסטוכסטי

נעבור כעת לפירוק האומדים למרכיב קבוע, שיבוטא באמצעות הפרמטר הנאמד, ולמרכיב סטוכסטי המושפע מההפרעות המקריות. לצורך כך נשתמש במשוואת המודל $Y_i = \alpha + \beta X_i + \varepsilon_i$.

נתחיל בשיפוע. מהפיתוח בסעיף 5.4.1. ראינו כי:

$$\hat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2}$$

במקום Y_i נציב את משוואת המודל $Y_i=\alpha+\beta X_i+\varepsilon_i$, ונקבל:

$$\hat{\beta} = \frac{\sum x_i (\alpha + \beta X_i + \varepsilon_i)}{\sum x_i^2} = \frac{\alpha \sum x_i + \beta \sum x_i X_i + \sum x_i \varepsilon_i}{\sum x_i^2}$$

אנו יודעים (מפרק 1) כי: $\sum x_i = 0$, $\sum x_i X_i = \sum x_i^2$, ולכן:

$$\hat{\beta} = \frac{\beta \sum x_i^2 + \sum x_i \varepsilon_i}{\sum x_i^2} = \frac{\beta \sum x_i^2}{\sum x_i^2} + \frac{\sum x_i \varepsilon_i}{\sum x_i^2} = \beta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}$$

מכאן שהצלחנו אם כן לפרק את $\hat{\beta}$ לשני מרכיבים, הפרמטר הנאמד (β) ואיבר נוסף סטוכסטי המורכב מההפרעות המקריות ומה-X-ים.

פירוק החותך יעשה בדרך דומה. נתחיל מהצגתו של $\hat{\alpha}$ כקומבינציה לינארית של ה-Y-ים כפי שמצאנו בסעיף 5.4.1., במקום Y_i נציב את משוואת המודל $Y_i=\alpha+\beta X_i+\varepsilon_i$ לקבלת:

$$\hat{\alpha} = \sum \left(\frac{1}{N} - \frac{\bar{X} \cdot x_i}{\sum x_i^2} \right) (\alpha + \beta X_i + \varepsilon_i)$$

נפרק את הביטוי לשלושה איברים :

$$\hat{\alpha} = \alpha\sum\left(\frac{1}{N} - \frac{\overline{X}\cdot x_i}{\sum x_i^2}\right) + \beta\sum\left(\frac{1}{N} - \frac{\overline{X}\cdot x_i}{\sum x_i^2}\right)X_i + \sum\left(\frac{1}{N} - \frac{\overline{X}\cdot x_i}{\sum x_i^2}\right)\varepsilon_i$$

נפעיל את סימן הסכום על שני הביטויים הראשונים :

$$\hat{\alpha} = \alpha\left(\frac{N}{N} - \frac{\overline{X}\sum x_i}{\sum x_i^2}\right) + \beta\left(\frac{\sum X_i}{N} - \frac{\overline{X}\sum x_i X_i}{\sum x_i^2}\right) + \sum\left(\frac{1}{N} - \frac{\overline{X}\cdot x_i}{\sum x_i^2}\right)\varepsilon_i$$

שוב נשתמש בתוצאות (מפרק 1) : $\sum x_i = 0$, $\sum x_i X_i = \sum x_i^2$, וכן $\sum X_i / n = \overline{X}$ לקבלת :

$$\hat{\alpha} = \alpha(1-0) + \beta\underbrace{\left(\overline{X} - \frac{\overline{X}\sum x_i^2}{\sum x_i^2}\right)}_{0} + \sum\left(\frac{1}{N} - \frac{\overline{X}\cdot x_i}{\sum x_i^2}\right)\varepsilon_i$$

ולכן :

$$\hat{\alpha} = \alpha + \sum\left(\frac{1}{N} - \frac{\overline{X}\cdot x_i}{\sum x_i^2}\right)\varepsilon_i$$

מכאן שהצלחנו אם כן לפרק גם את $\hat{\alpha}$ לשני מרכיבים, הפרמטר הנאמד (α) ואיבר נוסף סטוכסטי המורכב מההפרעות המקריות ומה-X-ים.

סיכום ביניים : הראינו כי גם את החותך וגם את השיפוע ניתן להציג כסכום של הפרמטרים הנאמדים ואיבר סטוכסטי, באופן הבא :

$$\boldsymbol{\hat{\alpha} = \sum\left(\frac{1}{N} - \frac{\overline{X}\cdot x_i}{\sum x_i^2}\right)\varepsilon_i} \quad , \quad \boldsymbol{\hat{\beta} = \beta + \frac{\sum x_i\varepsilon_i}{\sum x_i^2}}$$

סיכום פרק 5:

- אומד ריבועים פחותים לחותך נתון על ידי:

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

- אומד ריבועים פחותים לשיפוע נתון על ידי:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

- הקו הנאמד נקרא קו הרגרסיה והוא מתואר באמצעות המשוואה:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$$

- מאמידת ריבועים פחותים מתקבלות המשוואות הנורמליות, הראשונה $\sum e_i = 0$, והשניה $\sum e_i X_i = 0$.
- בדומה להנחה הקלאסית בדבר **תוחלת אפס של ההפרעה המקרית באוכלוסיה**, המשוואה הנורמלית הראשונה מראה כי **ממוצע ההפרעות הנאמדות במדגם הוא אפס**. תכונה זו מתקבלת מבלי לכפות אותה במפורש במהלך האמידה ותקפותה מותנת בקיומו של חותך במודל הנאמד.
- בדומה להנחה הקלאסית בדבר **חוסר מתאם בין ההפרעות המקריות ל-X-ים באוכלוסיה**, המשוואה הנורמלית השניה מראה כי **המתאם במדגם בין ההפרעות הנאמדות וה-X-ים הוא אפס**. תכונה זו מתקבלת מבלי לכפות אותה במפורש במהלך האמידה.
- האומד לחותך הוצג גם ע"י המשוואות:

$$\hat{\alpha} = \sum\left(\frac{1}{N} - \frac{\overline{X}\cdot x_i}{\sum x_i^2}\right)Y_i = \alpha + \sum\left(\frac{1}{N} - \frac{\overline{X}\cdot x_i}{\sum x_i^2}\right)\varepsilon_i$$

- האומד לשיפוע הוצג גם ע"י המשוואות:

$$\hat{\beta} = \sum\left(\frac{x_i}{\sum x_i^2}\right)Y_i = \beta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}$$

נספח 4.1. תנאי סדר שני

בכדי למצוא את אומדי הריבועים הפחותים הבאנו למינימום את סכום ריבועי השגיאות הנאמדות. לשם כך גזרנו את הביטוי $\sum e_i^2$ והשוונו את הנגזרות הראשונות לאפס, אך לא בדקנו האם תנאי סדר ראשון אכן מביאים אותנו למינימום. בנספח זה נראה כי תנאי סדר שני מתקיימים, כלומר האומדים שמצאנו אכן מביאים את סכום ריבועי הסטיות הנאמדות לאפס.

תזכורת: עבור פונקציה כלשהי בשני משתנים $f(x,y)$, הנקודה (x_0,y_0) היא נקודת מינימום אם מתקיימים שני התנאים הבאים:

1. ערכי הנגזרות השניות בנקודה (x_0,y_0) הן חיוביות, כלומר:

$$\left.\frac{\partial^2 f}{\partial x^2}\right|_{x_0,y_0} > 0 \quad , \quad \left.\frac{\partial^2 f}{\partial y^2}\right|_{x_0,y_0} > 0$$

2. וכן תנאי נוסף המבטיח כי אכן מדובר בנקודת קיצון (ולא אוכף למשל):

$$\left.\frac{\partial^2 f}{\partial x^2}\right|_{x_0,y_0} \cdot \left.\frac{\partial^2 f}{\partial y^2}\right|_{x_0,y_0} - \left(\left.\frac{\partial^2 f}{\partial x \partial y}\right|_{x_0,y_0}\right)^2 > 0$$

במקרה שלנו פונקצית המטרה היא $\sum e_i^2$, וגזרנו אותה לפי $\hat{\alpha}$ ו- $\hat{\beta}$. לכן, התנאים אותם יש לבדוק הם:

$$\frac{\partial^2 \sum e_i^2}{\partial \hat{\alpha}^2} > 0 \quad , \quad \frac{\partial^2 \sum e_i^2}{\partial \hat{\beta}^2} > 0$$

וכן:

$$\frac{\partial^2 \sum e_i^2}{\partial \hat{\alpha}^2} \cdot \frac{\partial^2 \sum e_i^2}{\partial \hat{\beta}^2} - \left(\frac{\partial^2 \sum e_i^2}{\partial \hat{\alpha} \partial \hat{\beta}}\right)^2 > 0$$

כאשר הנגזרות מוערכות בנקודת האומדים שמצאנו.

כזכור, מגזירה לפי $\hat{\alpha}$ קיבלנו:

$$\frac{\partial\{\cdot\}}{\partial \hat{\alpha}} = -2\sum (Y_i - \hat{\alpha} - \hat{\beta} X_i)$$

נגזור פעם שניה:

$$\frac{\partial^2\{\cdot\}}{\partial \hat{\alpha}^2} = -2\sum(-1) = 2N > 0$$

ביטוי זה חיובי תמיד (ובפרט בנקודת האומדים).

מגזירה לפי $\hat{\beta}$ קיבלנו:

$$\frac{\partial\{\cdot\}}{\partial \hat{\beta}} = -2\sum (Y_i - \hat{\alpha} - \hat{\beta} X_i) X_i$$

נגזור פעם שניה:

$$\frac{\partial^2 \{\cdot\}}{\partial \hat{\beta}^2} = -2\sum(-X_i^2) = 2\sum X_i^2 > 0$$

גם ביטוי זה חיובי תמיד (ובפרט בנקודת האומדים). מכאן שתנאי מספר 1 מתקיים.

לבדיקת תנאי מספר 2 נחשב תחילה את הנגזרת הצולבת של $\sum e_i^2$. לצורך כך ניקח את הנגזרת הראשונה לפי $\hat{\alpha}$ ונגזור אותה לפי $\hat{\beta}$ (ניתן גם לגזור לפי $\hat{\alpha}$ את הנגזרת הראשונה לפי $\hat{\beta}$ ולקבל אותה תוצאה בדיוק):

$$\frac{\partial^2 \sum e_i^2}{\partial \hat{\alpha} \partial \hat{\beta}} = -2\sum(-X_i) = 2\sum X_i = 2N\overline{X}$$

כעת נציב תוצאה זו ואת הנגזרות השניות שחישבנו עבור תנאי מספר 1 בביטוי שקיבלנו עבור תנאי 2, ונקבל:

$$\frac{\partial^2 \sum e_i^2}{\partial \hat{\alpha}^2} \cdot \frac{\partial^2 \sum e_i^2}{\partial \hat{\beta}^2} - \left(\frac{\partial^2 \sum e_i^2}{\partial \hat{\alpha} \partial \hat{\beta}} \right)^2 = 2N \cdot 2\sum X_i^2 - \left(2N\overline{X}\right)^2 = 4N(\sum X_i^2 - N\overline{X}^2)$$

כעת נשים לב כי הביטוי בסוגריים שווה ל- $\sum x_i^2$,[3] ולכן:

$$\frac{\partial^2 \sum e_i^2}{\partial \hat{\alpha}^2} \cdot \frac{\partial^2 \sum e_i^2}{\partial \hat{\beta}^2} - \left(\frac{\partial^2 \sum e_i^2}{\partial \hat{\alpha} \partial \hat{\beta}} \right)^2 = 4N\sum x_i^2 > 0$$

ומכאן שגם תנאי 2 מתקיים, והאומדים שמצאנו, $\hat{\alpha}$ ו- $\hat{\beta}$, אכן מוצאים נקודת מינימום.

[3] תוצאה זו הוצגה בפרק 1.