סיכום
מבוא לאקונומטריקה

מאת ניר דגן <http://he.nirdagan.com/bio>
4 בדצמבר 2008

חשבונות עם סכומים

הגדרת סימון הסכום: $\sum_{i=1}^{n} X_i$

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + \ldots + X_n$$

תכונות בסיסיות של סכומים

$$\sum_{i=1}^{n} aX_i = a\sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} a = na$$

$$\sum_{i=1}^{n} (X_i + Y_i) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i$$

$$\sum_{i=1}^{n} (X_i - Y_i) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i$$

$$\left(\sum_{i=1}^{n} X_i\right)^2 = \sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} X_i X_j$$

ממוצע של סדרת מספרים

הגדרת הממוצע

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

תכונות הממוצע

$$\sum_{i=1}^{n} \left(X_i - \overline{X}\right) = 0$$

$$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)=\sum_{i=1}^{n}\left(X_i-\overline{X}\right)Y_i=\sum_{i=1}^{n}X_iY_i-n\overline{X}\overline{Y}$$

$$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\sum_{i=1}^{n}\left(X_i-\overline{X}\right)X_i=\sum_{i=1}^{n}X_i{}^2-n\overline{X}^2$$

מושגי יסוד בהסתברות

תכונות של תוחלת

יהי a קבוע ו-$X,\ Y$ משתנים מקריים.

$$E(a)=a$$
$$E(aX)=aE(X)$$
$$E(X+Y)=E(X)+E(Y)$$

הגדרת שונות ושונות משותפת

$$Var(X)=E(X-E(X))^2$$
$$Cov(X,\ Y)=E(X-E(X))(Y-E(Y))$$

תכונות של שונות ושונות משותפת

יהיו $a,\ b$ קבועים ו-$X,\ Y$ משתנים מקריים.

$$Var(aX+b)=a^2Var(X)$$
$$Var(X+Y)=Var(X)+Var(Y)+2Cov(X,\ Y)$$

רגרסיה עם משתנה מסביר יחיד

המודל הבסיסי

$$Y_i=\alpha+\beta X_i+u_i$$

ההנחות הקלסיות:

0. X_i אינם משתנים מקריים.
1. לכל i: $E(u_i)=0$
2. לכל i: $E\left(u_i^2\right)=\sigma_u^2$
3. לכל $i\neq j$: $E\left(u_iu_j\right)=0$
4. לכל $i\neq j$: $u_i,\ u_j$ בלתי תלויים, ומתפלגים לפי ההתפלגות הנורמלית.

אומדי ריבועים פחותים

נגדיר עבור האומדים $\hat{\alpha},\ \hat{\beta}$ את e_i

$$e_i = Y_i - \hat{\alpha} - \hat{\beta} X_i$$

אומדי ריבועים פחותים הם פיתרון לבעיית האופטימיזציה הבאה:

$$\min_{\hat{\alpha},\hat{\beta}} \sum_{i=1}^{n} e_i^2$$

או בצורה מפורשת:

$$\min_{\hat{\alpha},\hat{\beta}} \sum_{i=1}^{n} \left(Y_i - \hat{\alpha} - \hat{\beta} X_i \right)^2$$

תנאי סדר ראשון לפתרון הבעיה הם:

$$-2 \sum_{i=1}^{n} \left(Y_i - \hat{\alpha} - \hat{\beta} X_i \right) = 0$$

$$-2 \sum_{i=1}^{n} \left(Y_i - \hat{\alpha} - \hat{\beta} X_i \right) X_i = 0$$

או בכתיבה מקוצרת:

$$\sum_{i=1}^{n} e_i = 0$$

$$\sum_{i=1}^{n} e_i X_i = 0$$

מתוך תנאי הסדר הראשון ניתן לפתור את האומדים:

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

$$\hat{\beta} = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X} \right) \left(Y_i - \overline{Y} \right)}{\sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2}$$

תוחלת ושונות האומדים

כדי לחשב את התוחלת, השונות והשונות המשותפת של האומדים, נציב בנוסחאות האומדים שקיבלנו את צד ימין של המודל האמיתי במקום המשתנה המוסבר. נתחיל עם $\hat{\beta}$.

$$\hat{\beta} = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X} \right) \left(Y_i - \overline{Y} \right)}{\sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2} = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X} \right) Y_i}{\sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2} = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X} \right) (\alpha + \beta X_i + u_i)}{\sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2} = \beta + \frac{\sum_{i=1}^{n} \left(X_i - \overline{X} \right)}{\sum_{i=1}^{n} \left(X_i - \overline{X} \right)}$$

עכשיו נחשב את התוחלת.

$$E\left(\hat{\beta}\right)=E\left(\beta+\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}\right)=E(\beta)+E\left(\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}\right)=\beta+\frac{E\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)u_i\right)}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}$$

$$=\beta+\frac{\sum_{i=1}^{n}E\left(\left(X_i-\overline{X}\right)u_i\right)}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}=\beta+\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)E(u_i)}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}=\beta$$

עכשיו נחשב את השונות של $\hat{\beta}$.

$$Var\left(\hat{\beta}\right)=\sigma_{\hat{\beta}}^2=E\left(\hat{\beta}-E\left(\hat{\beta}\right)\right)^2=E\left(\beta+\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}-\beta\right)^2=E\left(\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}\right)^2$$

$$=\frac{E\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)u_i\right)^2}{\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)^2}=\frac{E\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2u_i^2+\sum_{i=1}^{n}\sum_{j\neq i}\left(X_i-\overline{X}\right)\left(X_j-\overline{X}\right)u_iu_j\right)}{\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)^2}$$

$$=\frac{\sum_{i=1}^{n}E\left(\left(X_i-\overline{X}\right)^2u_i^2\right)+\sum_{i=1}^{n}\sum_{j\neq i}E\left(\left(X_i-\overline{X}\right)\left(X_j-\overline{X}\right)u_iu_j\right)}{\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)^2}$$

$$=\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2E\left(u_i^2\right)+\sum_{i=1}^{n}\sum_{j\neq i}\left(X_i-\overline{X}\right)\left(X_j-\overline{X}\right)E\left(u_iu_j\right)}{\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)^2}$$

$$=\frac{\sigma_u^2\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)^2}=\frac{\sigma_u^2}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}$$

באופן דומה ניתן לחשב עבור $\hat{\alpha}$.

$$E\left(\hat{\alpha}\right)=\alpha$$

$$Var\left(\hat{\alpha}\right)=\sigma_{\hat{\alpha}}^2=\frac{\sigma_u^2\sum_{i=1}^{n}X_i^2}{n\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}=\sigma_u^2\left(\frac{1}{n}+\frac{\overline{X}^2}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}\right)$$

חיזוי בעזרת אומדי ריבועים פחותים

$$\hat{Y}\Big|_{X=X_0}=E\hat{(Y)}\Big|_{X=X_0}=\hat{\alpha}+\hat{\beta}X_0$$

$$Var\left(\hat{Y}\Big|_{X=X_0}\right)=\sigma_u^2\left(1+\frac{1}{n}+\frac{\left(X_0-\overline{X}\right)^2}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}\right)$$

$$Var\left(E\hat{(Y)}\Big|_{X=X_0}\right)=\sigma_u^2\left(\frac{1}{n}+\frac{\left(X_0-\overline{X}\right)^2}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}\right)$$

מקדם ההסבר R^2

הגדרת R^2

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2}$$

משפטים

$$0 \le R^2 \le 1$$

יהי $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$. מתקבל:

$$\overline{\hat{Y}} = \overline{Y}$$
$$\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{n} \left(\hat{Y}_i - \overline{Y}\right)^2 + \sum_{i=1}^{n} e_i^2$$

מקדם המתאם: r_{xy}

הגדרת r_{xy}

$$r_{xy} = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2}}$$

משפטים

$$-1 \le r_{xy} \le 1$$
$$R^2 = r_{xy}^2$$

אומד לשונות של u_i

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$$

המודל ללא משתנה מסביר

$$Y_i = \alpha + u_i$$

אומד ריבועים פחותים

נגדיר עבור האומד $\hat{\alpha}$ את e_i

$$e_i = Y_i - \hat{\alpha}$$

אומד ריבועים פחותים הוא פיתרון לבעיית האופטימיזציה הבאה:

$$\min_{\hat{\alpha}} \sum_{i=1}^{n} e_i^2$$

או בצורה מפורשת:

$$\min_{\hat{\alpha}} \sum_{i=1}^{n} \left(Y_i - \hat{\alpha}\right)^2$$

תנאי סדר ראשון לפתרון הבעיה הוא:

$$-2\sum_{i=1}^{n} \left(Y_i - \hat{\alpha}\right) = 0$$

או בכתיבה מקוצרת:

$$\sum_{i=1}^{n} e_i = 0$$

מתוך תנאי הסדר הראשון ניתן לפתור את האומד:

$$\hat{\alpha} = \overline{Y}$$

תוחלת ושונות האומד

כדי לחשב את התוחלת והשונות של האומד, נציב בנוסחת האומד שקיבלנו את צד ימין של המודל האמיתי במקום המשתנה המוסבר.

$$\hat{\alpha} = \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} (\alpha + u_i) = \alpha + \frac{1}{n}\sum_{i=1}^{n} u_i$$

עכשיו נחשב את התוחלת.

$$E\left(\hat{\alpha}\right) = E\left(\alpha + \frac{1}{n}\sum_{i=1}^{n} u_i\right) = E(\alpha) + E\left(\frac{1}{n}\sum_{i=1}^{n} u_i\right) = \alpha + \frac{1}{n}E\left(\sum_{i=1}^{n} u_i\right) = \alpha + \frac{1}{n}\sum_{i=1}^{n} E(u_i) = \alpha$$

עכשיו נחשב את השונות.

$$Var\left(\hat{\alpha}\right) = E\left(\hat{\alpha} - E\left(\hat{\alpha}\right)\right)^2 = E\left(\alpha + \frac{1}{n}\sum_{i=1}^{n} u_i - \alpha\right)^2 = E\left(\frac{1}{n}\sum_{i=1}^{n} u_i\right)^2 = \frac{1}{n^2}E\left(\sum_{i=1}^{n} u_i\right)^2$$

$$= \frac{1}{n^2}E\left(\sum_{i=1}^{n} u_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} u_i u_j\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n} E\left(u_i^2\right) + \sum_{i=1}^{n}\sum_{j\neq i} E\left(u_i u_j\right)\right) = \frac{\sigma_u^2}{n}$$

המודל ללא חותך

$$Y_i = \beta X_i + u_i$$

אומד ריבועים פחותים

נגדיר עבור האומד $\hat{\beta}$ את e_i

$$e_i = Y_i - \hat{\beta} X_i$$

אומד ריבועים פחותים הוא פיתרון לבעיית האופטימיזציה הבאה:

$$\min_{\hat{\beta}} \sum_{i=1}^{n} e_i^2$$

או בצורה מפורשת:

$$\min_{\hat{\beta}} \sum_{i=1}^{n} \left(Y_i - \hat{\beta} X_i\right)^2$$

תנאי סדר ראשון לפתרון הבעיה הוא:

$$-2 \sum_{i=1}^{n} \left(Y_i - \hat{\beta} X_i\right) X_i = 0$$

או בכתיבה מקוצרת:

$$\sum_{i=1}^{n} e_i X_i = 0$$

מתוך תנאי הסדר הראשון ניתן לפתור את האומד:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

תוחלת ושונות האומד

כדי לחשב את התוחלת, והשונות של האומד, נציב בנוסחת האומד שקיבלנו את צד ימין של המודל האמיתי במקום המשתנה המוסבר.

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} = \frac{\sum_{i=1}^{n} X_i (\beta X_i + u_i)}{\sum_{i=1}^{n} X_i^2} = \beta + \frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}$$

עכשיו נחשב את התוחלת.

$$E\left(\hat{\beta}\right) = E\left(\beta + \frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}\right) = E(\beta) + E\left(\frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}\right) = \beta + \frac{E\left(\sum_{i=1}^{n} X_i u_i\right)}{\sum_{i=1}^{n} X_i^2}$$
$$= \beta + \frac{\sum_{i=1}^{n} E(X_i u_i)}{\sum_{i=1}^{n} X_i^2} = \beta + \frac{\sum_{i=1}^{n} X_i E(u_i)}{\sum_{i=1}^{n} X_i^2} = \beta$$

עכשיו נחשב את השונות של β.

$$Var\left(\hat{\beta}\right) = E\left(\hat{\beta} - E\left(\hat{\beta}\right)\right)^2 = E\left(\beta + \frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2} - \beta\right)^2 = E\left(\frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}\right)^2$$

$$= \frac{E\left(\sum_{i=1}^{n} X_i u_i\right)^2}{\left(\sum_{i=1}^{n} X_i^2\right)^2} = \frac{E\left(\sum_{i=1}^{n} X_i^2 u_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} X_i X_j u_i u_j\right)}{\left(\sum_{i=1}^{n} X_i^2\right)^2}$$

$$= \frac{\sum_{i=1}^{n} E\left(X_i^2 u_i^2\right) + \sum_{i=1}^{n} \sum_{j \neq i} E\left(X_i X_j u_i u_j\right)}{\left(\sum_{i=1}^{n} X_i^2\right)^2}$$

$$= \frac{\sum_{i=1}^{n} X_i^2 E\left(u_i^2\right) + \sum_{i=1}^{n} \sum_{j \neq i} X_i X_j E\left(u_i u_j\right)}{\left(\sum_{i=1}^{n} X_i^2\right)^2}$$

$$= \frac{\sigma_u^2 \sum_{i=1}^{n} X_i^2}{\left(\sum_{i=1}^{n} X_i^2\right)^2} = \frac{\sigma_u^2}{\sum_{i=1}^{n} X_i^2}$$

טעויות בספציפיקציה והשלכותיהן

השמטת החותך כשהמודל האמיתי עם חותך

נתבונן במקרה בו המודל הינו:

$$Y_i = \alpha + \beta X_i + u_i$$

אך בטעות אמדו את β ע"י אמידת המודל השגוי:

$$Y_i = \beta X_i + u_i$$

מה התוחלת של $\hat{\beta}$?

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

כדי לחשב את התוחלת של האומד, נציב בנוסחת האומד שקיבלנו את צד ימין של המודל האמיתי במקום המשתנה המוסבר.

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} = \frac{\sum_{i=1}^{n} X_i(\alpha + \beta X_i + u_i)}{\sum_{i=1}^{n} X_i^2} = \beta + \frac{\alpha \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2} + \frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}$$

עכשיו נחשב את התוחלת.

$$E\left(\hat{\beta}\right) = E\left(\beta + \frac{\alpha \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2} + \frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}\right) = E\left(\beta + \frac{\alpha \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2}\right) + E\left(\frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}\right)$$

$$= \beta + \frac{\alpha \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2} + \frac{E\left(\sum_{i=1}^{n} X_i u_i\right)}{\sum_{i=1}^{n} X_i^2} = \beta + \frac{\alpha \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2} + \frac{\sum_{i=1}^{n} E(X_i u_i)}{\sum_{i=1}^{n} X_i^2} = \beta + \frac{\alpha \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2} + \frac{\sum_{i=1}^{n} X_i E(u_i)}{\sum_{i=1}^{n} X_i^2}$$

$$= \beta + \frac{\alpha \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2} \neq \beta$$

הוספת חותך כאשר במודל האמיתי אין חותך

נתבונן במקרה בו המודל הינו:

$$Y_i = \beta X_i + u_i$$

אך בטעות אמדו את β ע"י אמידת המודל השגוי:

$$Y_i = \alpha + \beta X_i + u_i$$

מה התוחלת של $\hat{\beta}$?

$$\hat{\beta} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)Y_i}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)(\beta X_i + u_i)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} = \beta + \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)u_i}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

הביטוי שהתקבל לעיל הינו זהה לזה שהתקבל כאשר יש חותך במודל האמיתי. אי לכך:

$$E\left(\hat{\beta}\right) = \beta$$

$$Var\left(\hat{\beta}\right) = \frac{\sigma_u^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

יחד עם זאת ראינו שאומד ריבועים פחותים למודל בלי חותך הוא עדיף מפני ששונותו קטנה יותר.

רווחי סמך

ניתן לבנות רווחי סמך לפרמטרים של המודל α, β ולתחזית ספציפית ותחזית לתוחלת של המשתנה המוסבר. מפני שהאומדים שקיבלנו מתפלגים לפי ההתפלגות הנורמלית, והאומד לשונות של u_i מתפלג לפי התפלגות חי בריבוע עם $n-2$ דרגות חופש X_{n-2}^2 ניתן לבנות את רווחי הסמך באופן המפורט להלן:

רווח סמך לפרמטר (או תחזית) q

$$\hat{q} - c\sigma \le q \le \hat{q} + c\sigma$$

כאשר:

- $\hat{q}$ הוא האומד ל-q.
- σ היא סטיית התקן או האומד לסטיית התקן של q.
- אם סטיית התקן ידועה, c הוא הערך הקריטי מטבלת ההתפלגות הנורמלית.
- אם סטיית התקן איננה ידועה, אלא נאמדת, c הוא הערך הקריטי מטבלת התפלגות t עם $n-2$ דרגות חופש.