מודלים ושיטות חיזוי

פרק 1 - מבוא לקורס

תוכן העניינים

מבוא לקורס:

רקע:

הגדרות וסימונים:

משתנה אמפירי – תוצאותיו ידועות מראש (למשל: רמת הכנסה, גיל, מסי שנות לימוד במדגם מסוים).
משתנה מקרי – תוצאותיו לא ידועות מראש (כגון תוצאה בהטלת קובייה או בהטלת מטבע). באקונומטריקה נעסוק בעיקר במשתנים מקריים.

שני סוגי המשתנים יסומנו באות לועזית עם אינדקס (למשל: Y_t או X_t).
קבוע – מקבל ערך אחד בלבד (מסומן באות לועזית ללא אינדקס – למשל a או b).
לכל משתנה מקרי X_t יש <u>תוחלת</u> המייצגת את מרכז ההתפלגות (μ_X או $E(X)$).
השונות – מייצגת את מידת הפיזור של ההתפלגות (σ_X^2 או $V(X)$).
סטית התקן – היא השורש של השונות (σ_X).
שונות משותפת (covariance) – מדד להתפלגות המשותפת של שני משתנים מקריים ומייצגת את הכיוון של הקשר ביניהם ($\mathrm{Cov}(X,Y)$):
$\mathrm{Cov}(X,Y)=0 \Leftrightarrow X,Y$ בלתי מתואמים.
$\mathrm{Cov}(X,Y)>0 \Leftrightarrow$ מתאם חיובי בין המשתנים.
$\mathrm{Cov}(X,Y)<0 \Leftrightarrow$ מתאם שלילי בין המשתנים.
X,Y בלתי תלויים $\Leftarrow X,Y$ בלתי מתואמים.
מקדם המתאם של פירסון – מדד לכיוון ולעוצמת הקשר הליניארי בין שני משתנים: $\eta_{xy}=\dfrac{\mathrm{cov}(x,y)}{\sigma_x\cdot\sigma_y}$.
$-1\le\eta\le1$
$\eta=1$ מתאם ליניארי חיובי מלא בין שני המשתנים.
$\eta=-1$ מתאם ליניארי שלילי מלא בין שני המשתנים.
$\eta=0$ לא קיים מתאם ליניארי בין שני המשתנים.

אמידה:

פרמטר – ערך המשתנה הנחקר המתאר את כל האוכלוסייה.
סטטיסטי/אומד – ערך המשתנה הנחקר המתאר את המדגם.

אוכלוסייה	מדגם
$E(X)=\mu$	$\bar{X}=\frac{\sum_{i=1}^{n} X_i}{n}$
$V(X)=\sigma^2=E(X-E(X))^2$	$S_X^2=\frac{S_{XX}}{n}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$
$\text{cov}(X,Y)=E(X-E(X))(Y-E(Y))$	$\frac{S_{XY}}{n}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n}$
$\eta_{XY}=\frac{\text{cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$	$r_{XY}=\frac{S_{XY}}{\sqrt{S_{XX}}\sqrt{S_{YY}}}$

נוסחאות וחוקים בסטטיסטיקה:

יהיו X ו- Y משתנים מקריים, ו- a, b קבועים:

<u>חוקי הסיגמה</u>:

1. $\sum_{t=1}^{T} X_t = X_1+X_2+X_3+\ldots+X_T$.
2. סכום של קבוע: $\sum_{t=1}^{T} a = Ta$.
3. סכום של קבוע כפול משתנה = לקבוע כפול הסכום: $\sum_{t=1}^{T} aX_i = a\sum_{t=1}^{T} X_t$.
4. סכום של סכום/הפרש = לסכום/הפרש הסכומים: $\sum_{t=1}^{T}(X_t \pm Y)_t = \sum_{t=1}^{T} X_t \pm \sum_{t=1}^{T} Y_t$.
5. יש לשים לב כי: $\sum_{t=1}^{T} X_t^2 \neq \left(\sum_{t=1}^{T} X_t\right)^2$

$\sum_{t=1}^{T} X_t Y_t \neq \sum_{t=1}^{T} X_t \sum_{t=1}^{T} Y_t$.

הגדרות ופיתוחים:

1. סכום הסטיות מהממוצע = 0: $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)=0$.
2. סכום הסטיות הריבועיות מהממוצע (מונה השונות):
$$\sum_{t=1}^{T}\left(X_t-\bar{X}\right)X_t=\sum_{t=1}^{T}X_t^2-T\bar{X}^2=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)^2=S_{XX}.$$
3. מונה של השונות המשותפת:
$$\sum_{t=1}^{T}\left(Y_t-\bar{Y}\right)X_t=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)Y_t=\sum_{t=1}^{T}X_tY_t-T\bar{X}\bar{Y}=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)\left(Y_t-\bar{Y}\right)=S_{XY}.$$

חוקי התוחלת:

1. תוחלת של קבוע = קבוע: $E(a)=a$.
2. תוחלת של סכום/הפרש = לסכום/הפרש התוחלות:
$$E(X\pm Y)=E(X)\pm E(Y)$$
$$E\left(\sum(X_i)\right)=\sum E(X_i).$$
3. תוחלת של כפל/חילוק $\neq$ לכפל/חילוק התוחלות:
$$E(X\cdot Y)\neq E(X)\cdot E(Y)$$
$$E\left(\frac{X}{Y}\right)\neq\frac{E(X)}{E(Y)}.$$
$$E\left(X^2\right)\neq\left[E(X)\right]^2$$
4. השפעת טרנספורמציה ליניארית על התוחלת:
$$E\left(a/\frac{1}{a}X\pm b\right)=a/\frac{1}{a}\cdot E(X)\pm b.$$

חוקי השונות:

1. עבור X ו- Y בלתי תלויים/בלתי מתואמים מתקיים:
שונות של סכום/הפרש = סכום השונויות:
$$V(X\pm Y)=V(X)+V(Y)$$
$$V\sum(X_i)=\sum V(X_i).$$
2. עבור X ו- Y תלויים/מתואמים מתקיים:
שונות של סכום/הפרש $\neq$ סכום השונויות:
$$V(X\pm Y)=V(X)+V(Y)\pm 2\cdot Cov(X,Y).$$

3. שונות של קבוע = 0:
$$V(a)=0$$
$$V(a\pm x)=V(X)$$

4. השפעת טרנספורמציה ליניארית על השונות: $V(aX+b)=a^2V(X)$.

- חוקי התוחלת והשונות מתייחסים למשתנים אמפיריים כאל קבועים (יוצאים מחוץ לתוחלת או לשונות).
חוקי הסכימה מתייחסים למשתנים אמפיריים כמשתנים הנשארים בתוך הסיגמא (רק הקבועים ייצאו מחוץ לסיגמא).

<u>חוקי השונות המשותפת</u>:

1. שונות משותפת בין משתנה לקבוע = 0: $\mathrm{cov}(X,a)=0$.
2. שונות משותפת של משתנים המוכפלים בקבוע: $\mathrm{cov}(aX,bY)=ab\cdot\mathrm{cov}(X,Y)$.
3. שונות משותפת של משתנה עם עצמו = שונות המשתנה:
$$\mathrm{cov}(X,X)=V(X)$$
$$\mathrm{cov}(Y,Y)=V(Y)$$

המודל האקונומטרי:

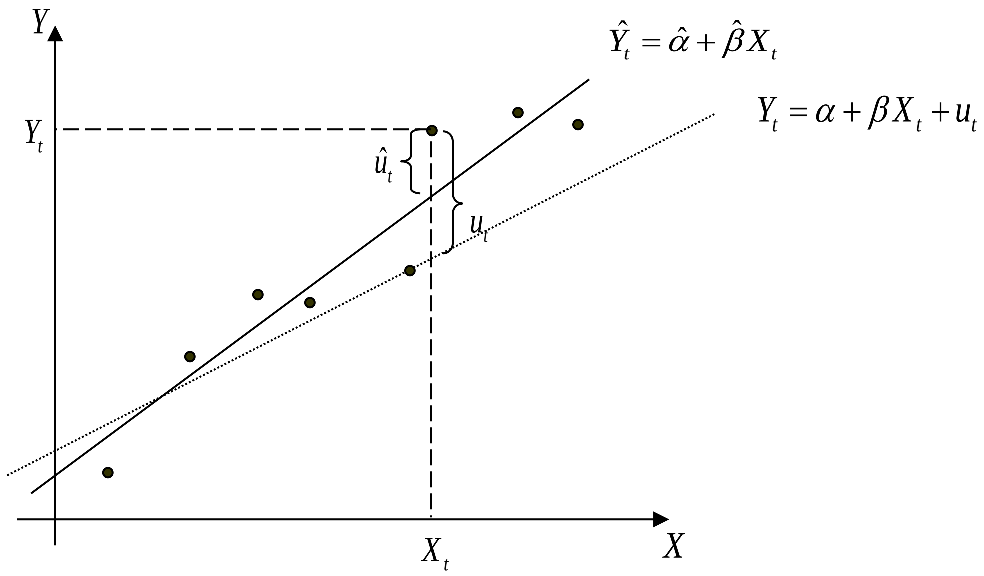
$$Y=\alpha+\beta X+u$$

משתנה מוסבר	פרמטרים: α - חותך, β - שיפוע	משתנה מסביר	הפרעה אקראית

1. במודל: $Y_t=\alpha+\beta X_t+u_t$, α ו- β הם מספרים קבועים אך לא ידועים. אנו יכולים להעריך אותם ולקבל אומדים (תהליך קבלת האומדנים נקרא אמידה).
2. $\hat{\alpha}$ הוא האומד ל- α ו- $\hat{\beta}$ הוא האומד ל- β.
3. אומדי ריבועים פחותים (אר"פ) הם אומדים שחושבו בשיטת הריבועים הפחותים. מסומנים בד"כ ע"י 'כובע' - $\hat{\beta}$.
אומדים אחרים מסומנים בד"כ ע"י 'תלתל' - $\tilde{\beta}$.

4. בעוד α ו-β הם קבועים, $\hat{\alpha}$ ו-$\hat{\beta}$ הם משתנים מקריים כיוון שבכל מדגם מתקבלים $\hat{\alpha}$ ו-$\hat{\beta}$ אחרים.
5. את α ו-β ו-u_t לא ניתן לדעת (אלא רק לאמוד מנתוני המדגם) – הקו האמיתי באוכ' לא ידוע.
6. אפשר לדעת את $\hat{u}_t$, שהיא הסטיה מקו הרגרסיה במדגם:

- עבור X_t, הערך הצפוי של המשתנה המוסבר $(\hat{Y}_t)$ המתקבל לפי הרגרסיה הוא: $\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t$.
- הסטיה של התצפית (Y_t) מהערך הצפוי לפי הרגרסיה $(\hat{Y}_t)$ היא: $\hat{u}_t = Y_t - \hat{Y}_t$.



— קו הרגרסיה הנאמד (במדגם)
..... קו הרגרסיה האמיתי באוכלוסייה)
• תצפית בודדת

שאלות:

1) הבא נוכיח את הזהויות הבאות:

א. $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)^2=\sum_{t=1}^{T}X_t^2-T\bar{X}^2$.

ב. $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)^2=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)X_t$.

ג. $\sum\left(X_t-\bar{X}\right)=0$.

ד. $\sum\frac{\left(X_t-\bar{X}\right)^2}{\sum\left(X_t-\bar{X}\right)X_t}=1$.

ה. $\sum\left(\hat{\alpha}+\hat{\beta}x_i\right)\left(x_i-\bar{x}\right)=\hat{\beta}\left(x_i-\bar{x}\right)^2$.

ו. $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)\left(Y_t-\bar{Y}\right)=\sum_{t=1}^{T}X_tY_t-T\bar{X}\bar{Y}$.

ז. $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)\left(Y_t-\bar{Y}\right)=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)Y_t$.

2) בטא באמצעות: $\operatorname{cov}(x,y),\operatorname{var}(x),\operatorname{var}(y)$ והקבועים a ו- b את הביטויים הבאים:

א. $Var(ax)$.

ב. $Var(x+y)$.

ג. $Var(ax+b)$.

ד. $Cov(x,ay)$.

ה. $Cov(x+a,y+b)$.

ו. מקדם המתאם בין x ל- y.

תשובות סופיות:

1) הוכחה.

2) א. $a^2\operatorname{var}(x)$. ב. $\operatorname{var}(x)+\operatorname{var}(y)+2\operatorname{cov}(x,y)$. ג. $a^2\operatorname{var}(x)$. ד. $a\operatorname{cov}(x,y)$.

ה. $\operatorname{cov}(x,y)$. ו. $r=\frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{var}(x)}\cdot\sqrt{\operatorname{var}(y)}}$.