נוסחאות

הממוצע של X והממוצע של Y (X ו- Y "גדולים"):

$$\bar{X} = \frac{\Sigma X_i}{n} \qquad \bar{Y} = \frac{\Sigma Y_i}{n}$$

הממוצע לאחר טרנספורמציה ליניארית (הוספת קבוע לכל ערכי המשתנה ו/או הכפלתם בקבוע):

$$\overline{a + bX} = a + b\bar{X} \qquad \overline{c + dY} = c + d\bar{Y}$$

סטיות מהממוצע:

x_i "קטן" הוא הסטייה של X_i "גדול" מהממוצע $\bar{X}$ $\quad \bar{x} = 0 \quad \sum x_i = 0 \quad x_i = X_i - \bar{X}$

y_i "קטן" הוא הסטייה של Y_i "גדול" מהממוצע $\bar{Y}$ $\quad \bar{y} = 0 \quad \sum y_i = 0 \quad y_i = Y_i - \bar{Y}$

הנוסחאות הקושרות בין ה"קטנים" ל"גדולים":

$$\sum xy = \sum XY - n \cdot \bar{X} \cdot \bar{Y} \qquad \sum y^2 = \sum Y^2 - n\bar{Y}^2 \qquad \sum x^2 = \sum X^2 - n\bar{X}^2$$

השונות של X במדגם היא S_X^2 וסטיית התקן של X במדגם היא S_X (השורש החיובי של השונות) :

$$S_X^2 = \frac{\sum X^2}{n} - \bar{X}^2 \qquad S_X^2 = \frac{\sum x^2}{n} \qquad S_X^2 = \frac{\sum (X - \bar{X})^2}{n}$$

ולאחר טרנספורמציה ליניארית: $\quad S_{a+bX} = |b| S_X \quad S_{a+bX}^2 = b^2 S_X^2$

השונות המשותפת (covariance) של X ושל Y, במדגם:

$$S_{X,Y} = \frac{\sum XY}{n} - \bar{X} \cdot \bar{Y} \qquad S_{X,Y} = \frac{\sum xy}{n} \qquad S_{X,Y} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n}$$

ולאחר טרנספורמציות ליניאריות: $\quad S_{a+bX,\, c+dY} = b \cdot d \cdot S_{X,Y}$

מקדם המתאם (של פירסון) בין X לבין Y במדגם :

$$r_{X,Y} = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}} \qquad r_{X,Y} = \frac{S_{X,Y}}{S_X \cdot S_Y}$$

לאחר טרנספורמציות ליניאריות ערכו המוחלט לא משתנה, $\quad |r_{a+bX,\, c+dY}| = |r_{X,Y}|$

ורק אם אחד, ורק אחד, מהכופלים (b או d) הוא שלילי– סימנו מתהפך.

מקדם ההסבר (של פירסון) בין X לבין Y הוא $r_{X,Y}^2$ (מקדם המתאם בריבוע).

ולאחר טרנספורמציות ליניאריות הוא נשאר ללא שינוי: $\quad r_{a+bX,\, c+dY}^2 = r_{X,Y}^2$

רגרסיה פשוטה (מסביר אחד וחותך) **.** קו הרגרסיה הוא $\hat{Y} = a + bX$.

עבור כל תצפית (X_i, Y_i):

$$Y_i = a + bX_i + e_i$$

$$Y_i = \hat{Y}_i + e_i$$

הסטייה מקו הרגרסיה:

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

מקדמי הקו נאמדים בשיטת הריבועים הפחותים הרגילים (OLS), השיטה המביאה למינימום את סכום הסטיות הריבועיות, $\min \Sigma e_i^2$.

השיפוע:

$$\boxed{b = r\frac{S_Y}{S_X}} \quad \boxed{b = \frac{\Sigma Y_i X_i - n\bar{Y}\bar{X}}{\Sigma X_i^2 - n\bar{X}^2}} \quad \boxed{b = \frac{\Sigma x_i y_i}{\Sigma x_i^2}} \quad \boxed{b = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2}}$$

$$\boxed{b = \frac{(\Sigma Y_i X_i)/n - \bar{X}\cdot\bar{Y}}{(\Sigma X_i^2)/n - \bar{X}^2}} \quad \boxed{b = \frac{S_{XY}}{S_X^2}} \quad \boxed{b = \frac{n\Sigma Y_i X_i - (\Sigma Y_i)(\Sigma X_i)}{n\Sigma X_i^2 - (\Sigma X_i)^2}}$$

החותך:

$$\boxed{a = \bar{Y} - b\bar{X}} \quad \boxed{a = \frac{\bar{Y}\Sigma X_i^2 - \bar{X}\Sigma X_i Y_i}{\Sigma X_i^2 - \bar{X}\Sigma X_i}}$$

תכונת פירוק השונות:

$$\frac{\sum y_i^2}{n} = \frac{\sum \hat{y}_i^2}{n} + \frac{\sum e_i^2}{n} \quad \text{או} \quad S_Y^2 = S_{\hat{Y}}^2 + S_e^2$$

R^2 מקדם ההסבר (או מקדם ההסבר המרובה או מקדם הקביעה) מוגדר כפרופורציית השונות המוסברת :

$$R^2 = 1 - \frac{\sum e^2}{\sum y^2}$$

סכום הסטיות הריבועיות הוא $\sum e^2 = (1 - R^2)\sum y^2$.

רק ברגרסיה פשוטה: $R^2 = r_{X,Y}^2$ (פרופורציית השונות המוסברת שווה למקדם ההסבר של פירסון)

מקדם ההסבר: נסמן את $r_{X,Y}^2$ ב-r^2.

$$\boxed{r^2 = 1 - \frac{\Sigma e_i^2}{\Sigma y_i^2} = 1 - \frac{S_e^2}{S_Y^2} = \frac{S_{\hat{Y}}^2}{S_Y^2}}$$

נציב $S_{\hat{Y}}^2 = S_{a+bX}^2 = b^2 \cdot S_X^2$ **ונקבל,** $\boxed{r = b\frac{S_X}{S_Y}}$ $\boxed{r^2 = b^2\frac{S_X^2}{S_Y^2}}$

נציב $b = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2}$ $\quad S_Y = \sqrt{\frac{\Sigma(Y_i - \bar{Y})^2}{n}}$ $\quad S_X = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n}}$

ונקבל

$$\boxed{r = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2 \Sigma(Y_i - \bar{Y})^2}} = \frac{\Sigma x_i y_i}{\sqrt{\Sigma x_i^2 \cdot \Sigma y_i^2}}}$$

או

$$\boxed{r = \frac{n\Sigma X_i Y - (\Sigma X_i)(\Sigma Y_i)}{\sqrt{[n\Sigma X_i^2 - (\Sigma X_i)^2][n\Sigma Y_i^2 - (\Sigma Y_i)^2]}}}$$

נציב, $S_X = \sqrt{\frac{\Sigma x_i^2}{n}}$ $\quad S_Y = \sqrt{\frac{\Sigma y_i^2}{n}}$ $\quad S_{XY} = \frac{\Sigma x_i y_i}{n}$ **ונקבל ,** $\boxed{r = \frac{S_{XY}}{S_X \cdot S_Y}}$

האומדים לשונויות:

האומד ל- σ^2, שונות ההפרעה המקרית: $S^2 = \frac{\sum e_i^2}{n-1-1}$

האומד ל- $V(b)$, שונות השיפוע: $S_b^2 = \frac{S^2}{\sum x^2}$

האומד ל- $V(a)$, שונות החותך: $S_a^2 = S^2[\frac{1}{n} + \frac{\bar{X}^2}{\sum x^2}]$

האומד ל- $COV(a,b)$, **השונות המשותפת שבין החותך לשיפוע:** $S_{a,b} = \frac{-\bar{X}\ S^2}{\sum x^2}$

מבחני השערה ורווחי סמך:

מבחן השערה על β: $\frac{b - \beta_{H_0}}{S_b} \sim t_{n-1-1}$ (ועוד נוסחא שימושית $\frac{b-0}{S_b} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$)

מבחן השערה על α: $\frac{a - \alpha_{H_0}}{S_a} \sim t_{n-1-1}$

רווח סמך ל- β: $b \pm t_{n-1-1,\ \alpha/2} \cdot S_b$

רווח סמך ל- α: $a \pm t_{n-1-1,\ \alpha/2} \cdot S_a$

ניבוי התוחלת של Y עבור $X = X_0$**:**

אומד נקודתי ל- $E(Y\big|_{X=X_0})$ הוא $\hat{Y}\big|_{X=X_0} = a + bX_0$.

רווח סמך ל- $E(Y\big|_{X=X_0})$ הוא $\hat{Y}\big|_{X=X_0} \pm t_{n-1-1,\ \alpha/2} \cdot S\sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x^2}}$

ניבוי Y עבור $X = X_0$ (תחזית):

אומד נקודתי ל- $Y\big|_{X=X_0}$ הוא $\hat{Y}\big|_{X=X_0} = a + bX_0$.

רווח סמך ל- $Y\big|_{X=X_0}$ הוא $\hat{Y}\big|_{X=X_0} \pm t_{n-1-1,\ \alpha/2} \cdot S\sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x^2}}$

רגרסיה הפוכה $\hat{X} = a' + b'Y$:

$X = a' + b'Y + e'$

$b' = \frac{\sum xy}{\sum y^2}$ $\quad a' = \bar{X} - b'\bar{Y}$

$b \cdot b' = r^2$

$e_i' = X_i - \hat{X}_i$ $\quad R^2 = 1 - \frac{\sum e'^2}{\sum x^2}$ $\quad S'^2 = \frac{\sum e'^2}{n-2}$

רגרסיה מרובה: מודל עם שני משתנים מסבירים $(k=2)$ וחותך $Y=\alpha+\beta_{01.2}X_1+\beta_{02.1}X_2+u$

$$\hat{Y}=a+b_{01.2}X_1+b_{02.1}X_2$$

$$b_{01.2}=\frac{\sum yx_1\sum x_2^2-\sum yx_2\sum x_1x_2}{\sum x_1^2\sum x_2^2-(\sum x_1x_2)^2} \qquad b_{02.1}=\frac{\sum yx_2\sum x_1^2-\sum yx_1\sum x_1x_2}{\sum x_1^2\sum x_2^2-(\sum x_1x_2)^2}$$

$$a=\bar{Y}-b_{01.2}\bar{X}_1-b_{02.1}\bar{X}_2$$

הקשר בין מקדמי המרובה למקדמי הפשוטה: $b_{01}=b_{01.2}+b_{21}b_{02.1}$ $\quad b_{02}=b_{02.1}+b_{12}b_{01.2}$

האומד לשונות ההפרעה המקרית (σ^2) : $S^2=\dfrac{\sum e^2}{n-k-1}$

האומדנים למקדמים במודל עם שני משתנים מסבירים וחותך:

$$S_{b_{01.2}}^2=\frac{S^2}{(1-r_{12}^2)\sum x_1^2} \qquad S_{b_{02.1}}^2=\frac{S^2}{(1-r_{12}^2)\sum x_2^2} \qquad S_{b_{01.2},b_{02.1}}=\frac{-S^2r_{12}^2}{(1-r_{12}^2)\sum x_1x_2}$$

מבחן השערה על מקדם הרגרסיה β_j **:** $\dfrac{b_j-(\beta_j)_{H_0}}{S_{b_j}}\sim t_{n-k-1}$

מבחן השערה על קשר ליניארי בין מקדמים: $H_0\ :\ \lambda_1\beta_1\pm\lambda_2\beta_2=\lambda_3$

$$\frac{\lambda_1b_1\pm\lambda_2b_2-\lambda_3}{\sqrt{\lambda_1^2S_{b_1}^2+\lambda_2^2S_{b_2}^2\pm2\lambda_1\lambda_2\,\mathrm{c\hat{o}v}(b_1,b_2)}}\sim t_{n-k-1}$$

רווח סמך ל- α : $a\pm t_{n-k-1,\alpha/2}\cdot S_a$

רווח סמך ל- β_j : $b_j\pm t_{n-k-1,\alpha/2}\cdot S_{b_j}$

בחינת מובהקות לרגרסיה, $H_0:\rho^2=0$ $\quad \dfrac{R^2/k}{(1-R^2)/(n-k-1)}\sim F_{k,n-k-1}$

מבחן WALD (לפי רגרסיה מוגבלת): $\dfrac{\left(\sum^{RES}e^2-\sum^{UNRES}e^2\right)/m}{\sum^{UNRES}e^2/(n-k-1)}\sim F_{m,n-k-1}$

ואם ל-RES ול_UNRES יש אותו משתנה מוסבר $\quad \dfrac{(R^2_{UNRES}-R^2_{UNRES})/m}{(1-R^2_{UNRES})/(n-k-1)}\sim F_{m,n-k-1}$

כאשר m הוא מספר ההצהרות ב- H_0 ו- k הוא מספר המשתנים המסבירים ברגרסיה הלא מוגבלת.

מבחן LM:

למובהקות תת קבוצה של מקדמים: בר"מ α דוחים H_0 אם $nR^2>\chi^2_{m,\alpha}$

למובהקות הרגרסיה: בר"מ α דוחים H_0 אם $nR^2>\chi^2_{k,\alpha}$

מקדם ההסבר המתוקנן:

$$\bar{R}^2=1-\frac{\sum e^2/(n-k-1)}{\sum y^2/(n-1)}$$

$$\bar{R}^2=1-(1-R^2)\frac{n-1}{n-k-1}$$