

מתאם ורגרסיה

$$Cov(X,Y) = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{n} = \frac{\sum_i X_i Y_i}{n} - \bar{X} \bar{Y} \quad (\text{covariance}) \quad \text{שונויות משותפת}$$

$$r = \frac{cov(X,Y)}{S_x S_y} \quad \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2} \quad \text{מקדם המתאם}$$

$$Y = a + bX \quad b = \frac{cov(X,Y)}{S_x^2} \quad a = \bar{Y} - b\bar{X} \quad \text{קו הרגרסיה של y על x}$$

$$S_{y,x}^2 = S_y^2(1-r^2) \quad \text{שונויות קו הרגרסיה}$$

$$X = a' + b'Y \quad b' = \frac{cov(X,Y)}{S_y^2} \quad a' = \bar{X} - b'\bar{Y} \quad \text{קו הרגרסיה של x על y}$$

$$S_{x,y}^2 = S_x^2(1-r^2) \quad \text{שונויות קו הרגרסיה}$$

דוגמא: X = השקעה בפרסום באלפי ₪.
Y = אלפי יחידות נמכרות.

X:	400	600	1000	200	600	800
Y:	6	7	10	4	7	8

$$\bar{X} = 600 \quad \bar{Y} = 7 \quad S_x^2 = 66666.667 \quad S_y^2 = 3.33 \quad cov(X,Y) = 466.667 \quad r = \frac{cov(X,Y)}{S_x S_y} = 0.99$$

$$Y = a + bX \quad b = \frac{cov}{S_x^2} = 0.007 \quad a = \bar{Y} - b\bar{X} = 2.8$$

$$Y = 2.8 + 0.007X \quad S_{y,x}^2 = S_y^2(1-r^2) = 0.069$$

$$X = a' + b'Y \quad b' = \frac{cov}{S_y^2} = 140 \quad a' = \bar{X} - b'\bar{Y} = -380$$

$$X = -380 + 140Y \quad S_{x,y}^2 = S_x^2(1-r^2) = 1326.67$$

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r} \quad \text{נגדיר:} \quad \text{כנגד} \quad H_1: \rho > \rho_0 \quad H_0: \rho = \rho_0 > 0 \quad \text{לבדיקת}$$

$$\frac{Z - \mu_z}{\sigma_z} \sim N(0,1) \quad \sigma_z = \frac{1}{\sqrt{n-3}} \quad \text{וסטיית תקן} \quad \mu_z = \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0} \quad \text{עם תוחלת}$$

$$\frac{Z_0 - \mu_z}{\sigma_z} > z_{1-\alpha} \quad \text{נדחה את השערת האפס בר"מ } \alpha \text{ אם}$$

דוגמא: בדוגמא הקודמת נבדוק: $H_0: \rho = 0.7$ כנגד $H_1: \rho > 0.7$

$$Z = \frac{1}{2} \ln \frac{1.99}{0.01} \quad \mu_z = \frac{1}{2} \ln \frac{1.7}{0.3} \quad \sigma_z = \frac{1}{\sqrt{3}} \quad \frac{Z_0 - \mu_z}{\sigma_z} = 3.082 > z_{0.995} = 2.58$$

ברמת מובהקות 0.005 נדחה את השערת האפס, ונסיק כי $\rho > 0.7$.

רגרסיה - המשך

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

באוכלוסיה:

$$Y = a + bX$$

במדגם:

$$S_{y.x}^2 = S_y^2(1-r^2)$$

$$\hat{S}_{y.x}^2 = \frac{n}{n-2} S_{y.x}^2 = \frac{n}{n-2} S_y^2(1-r^2)$$

$$\hat{S}_b^2 = \frac{\hat{S}_{y.x}^2}{n S_x^2} = \frac{S_y^2(1-r^2)}{S_x^2(n-2)}$$

$$\frac{b-\beta}{\hat{S}_b} \sim t_{n-2}$$

$$\hat{S}_a^2 = \frac{\hat{S}_b^2 \left(\sum X_i^2 \right)}{n}$$

$$\frac{a-\alpha}{\hat{S}_a} \sim t_{n-2}$$

$$\hat{S}_{y(x_0)}^2 = \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n S_x^2} \right) \hat{S}_{y.x}^2$$

$$\frac{y_{x_0} - Y_{x_0}}{\hat{S}_{y(x_0)}} \sim t_{n-2}$$

ומכאן: רווח סמך ל- β ברמת סמך של $1-\alpha$:

$$b - \hat{S}_b t_{n-2, 1-\frac{\alpha}{2}} \leq \beta \leq b + \hat{S}_b t_{n-2, 1-\frac{\alpha}{2}}$$

רווח סמך ל- α ברמת סמך של $1-\alpha$:

$$a - \hat{S}_a t_{n-2, 1-\frac{\alpha}{2}} \leq \alpha \leq a + \hat{S}_a t_{n-2, 1-\frac{\alpha}{2}}$$

רווח סמך ל- Y_{x_0} ברמת סמך של $1-\alpha$:

$$y_{x_0} - \hat{S}_{y(x_0)} t_{n-2, 1-\frac{\alpha}{2}} \leq Y_{x_0} \leq y_{x_0} + \hat{S}_{y(x_0)} t_{n-2, 1-\frac{\alpha}{2}}$$