בס"ד

הפרמטר ρ בדגימת אשכולות

בכיתה הגדרנו

$$\rho = \frac{m(S_b^2/S_Y^2) - 1}{m-1}$$

כאשר

$$S_b^2 = \frac{1}{c-1} \sum_{i=1}^{c} (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$S_Y^2 = \frac{1}{N-1} \sum_{i=1}^{c} \sum_{j=1}^{m} (Y_{ij} - \bar{Y}_{..})^2$$

כאן נפרט כמה תכונות של ρ.

1 נוכל לכתוב

$$(c-1) S_b^2 = \sum_{i=1}^{c} (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$= \sum_{i=1}^{c} \left[\frac{1}{m} \sum_{j=1}^{m} (Y_{ij} - \bar{Y}_{..}) \right]^2$$

$$= \left(\frac{1}{m}\right)^2 \sum_{i=1}^{c} \left[\sum_{j=1}^{m} (Y_{ij} - \bar{Y}_{..})\right]\left[\sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})\right]$$

בס"ד

$$= \left(\frac{1}{m}\right)^2 \sum_{i=1}^{c} \sum_{j=1}^{m} \sum_{k=1}^{m} (Y_{ij} - \bar{Y}_{..})(Y_{ik} - \bar{Y}_{..})$$

$$= \left(\frac{1}{m}\right)^2 \sum_{i=1}^{c} \sum_{j=1}^{m} (Y_{ij} - \bar{Y}_{..})^2$$

$$+ \left(\frac{1}{m}\right)^2 \sum_{i=1}^{c} \sum_{\substack{j=1 \\ j \neq k}}^{m} \sum_{k=1}^{m} (Y_{ij} - \bar{Y}_{..})(Y_{ik} - \bar{Y}_{..})$$

$$= \left(\frac{1}{m^2}\right)(N-1)\, S_Y^2 + \frac{1}{m^2} Q$$

כאשר

$$Q = \sum_{i=1}^{c} \sum_{\substack{j=1 \\ j \neq k}}^{m} \sum_{k=1}^{m} (Y_{ij} - \bar{Y}_{..})(Y_{ik} - \bar{Y}_{..})^2$$

לכן

$$S_b^2 = \frac{N-1}{m^2(c-1)} S_Y^2 + \frac{1}{m^2(c-1)} Q$$

$$\rho = \left[m\left(S_b^2 / S_Y^2\right) - 1\right] / (m-1)$$

$$= \frac{1}{m-1}\left[\frac{N-1}{m(c-1)} - 1 + \frac{1}{m(c-1) S_Y^2} Q\right]$$

$$= \frac{1}{m-1}\left[\frac{N-1}{N-m} - 1 + \frac{1}{(N-m) S_Y^2} Q\right]$$

בס"ד

אם C גדול, אזי

$$\frac{N-1}{N-m} = \frac{Cm-1}{Cm-m} \doteq 1$$

ואז נוכל לכתוב

$$\rho \doteq \left[(m-1)(N-1)S_Y^2\right]^{-1} Q$$

לאור זה, נגדיר

$$\tilde{\rho} = \left[(m-1)(N-1)S_Y^2\right]^{-1} Q$$

② <u>טענה</u>: נתייחס לתהליך הבא:

א. נבחר אשכול אחד באופן מקרי. נסמן את האינדקס של האשכול שבחרנו בסימן ϑ.

ב. מתוך האשכול שבחרנו בשלב א' לעיל, נבחר שני פרטים זה אחר זה באופן מקרי ללא החזרה. נסמן:

g_1 = האינדקס של הפרט שנבחר בבחירה הראשונה

g_2 = האינדקס של הפרט שנבחר בבחירה השנייה

נניח כי ϑ ו- (g_1, g_2) בלתי-תלויים.

נגדיר

$$U = Y_{\vartheta g_1} , \quad V = Y_{\vartheta g_2}$$

אזי

$$Corr(U, V) = \tilde{\rho}$$

<u>הערה</u>: התוצאה הנ"ל מחזקת את האינטרפרטציה של הפרמטר ρ (השווה בהרצאה ל- $\tilde{\rho}$) ה...

בס"ד

<u>הוכחה</u>: נזכיר כי

$$Corr(U,V) = Cov(U,V) / [Var(U)\,Var(V)]^{1/2}$$

$$Cov(U,V) = E[(U-E[U])(V-E[V])]$$ כאשר

נחשב את:

$$E[U],\ E[V]$$
$$Var(U),\ Var(V)$$
$$Cov(U,V)$$

לפי התהליך המתואר, יש לנו

$$Pr(\ell = i) = \frac{1}{e}, \quad i = 1, \dots, e$$

$$Pr(g_1 = j,\ g_2 = k) = \frac{1}{m(m-1)}$$

$$j, k = 1, \dots, m \quad (j \neq k)$$

$$Pr(\ell = i,\ g_1 = j,\ g_2 = k) = \frac{1}{e}\,\frac{1}{m(m-1)}$$

$$Pr(g_1 = j) = \frac{1}{m} \qquad j = 1, \dots, m$$

$$Pr(g_2 = k) = \frac{1}{m} \qquad k = 1, \dots, m$$

$$Pr(\ell = i,\ g_1 = j) = \frac{1}{em} = \frac{1}{N}$$

$$Pr(\ell = i,\ g_2 = k) = \frac{1}{em} = \frac{1}{N}$$

⑥ בס״ד

כעת:

$$E[U] = E[Y_{\theta g_1}]$$
$$= \sum_{i=1}^{e}\sum_{j=1}^{m} \Pr(\theta=i, g_1=j)\, Y_{ij}$$
$$= \frac{1}{N}\sum_{i=1}^{e}\sum_{j=1}^{m} Y_{ij} = \bar{Y}_{..}$$

$$Var(U) = E[(U - E[U])^2]$$
$$= E[(Y_{\theta g_1} - \bar{Y}_{..})^2]$$
$$= \sum_{i=1}^{e}\sum_{j=1}^{m} \Pr(\theta=i, g_1=j)(Y_{ij} - \bar{Y}_{..})^2$$
$$= \frac{1}{N}\sum_{i=1}^{e}\sum_{j=1}^{m}(Y_{ij} - \bar{Y}_{..})^2 = \frac{N-1}{N} S_Y^2$$

מטעמי סימטריות:

$$E[V] = E[U] = \bar{Y}_{..}$$
$$Var(V) = Var(U) = \frac{N-1}{N} S_Y^2$$

בס״ד

כעת:

$$\mathrm{Cov}(U,V) = E[(U-E[U])(V-E[V])]$$

$$= E[(Y_{\ell g_1}-\overline{Y}_{..})(Y_{\ell g_2}-\overline{Y}_{..})]$$

$$= \sum_{i=1}^{c}\sum_{\substack{j=1\\ j\neq k}}^{m}\sum_{k=1}^{m} P_r(\ell=i, g_1=j, g_2=k)\,(Y_{ij}-\overline{Y}_{..})(Y_{ik}-\overline{Y}_{..})$$

$$= \frac{1}{cm(m-1)}\sum_{i=1}^{c}\sum_{\substack{j=1\\ j\neq k}}^{m}\sum_{k=1}^{m}(Y_{ij}-\overline{Y}_{..})(Y_{ik}-\overline{Y}_{..})$$

$$= \frac{1}{N(m-1)}\,Q$$

באשר Q כמוגדר בסע׳ (2).

לכן:

$$\mathrm{Corr}(U,V)$$

$$= \left[\frac{1}{N}\,\frac{1}{m-1}\,Q\right] \Big/ \left[\left(\frac{N-1}{N}S_Y^2\right)\left(\frac{N-1}{N}S_Y^2\right)\right]^{1/2}$$

$$= \frac{1}{(N-1)(m-1)S_Y^2}\,Q = \tilde{\rho}\,.$$

מ.ש.ל. ☺