Multiple and Partial Correlations

Date

Result -

$X_1^* = E(X_1 | X_2 = x_2) = a + bx_2$ is the regression line of X_1 on X_2. b is the regression coefficient. It follows that

$$\int x_1 f(x_1|x_2)\, dx_1 = a + bx_2 \quad \text{— (1)}$$

multiplying both sides by $f(x_2)$ and integrating w.r.t. x_2, we get.

$$\iint x_1 f(x_1|x_2) f(x_2)\, dx_1\, dx_2 = a \int f(x_2)\, dx_2 + b \int x_2 f(x_2)\, dx_2$$

$$\Rightarrow \iint x_1 \left[f(x_1, x_2)\, dx_2 \right] dx_1 = a + b\mu_2$$

$$\Rightarrow \int x_1 f(x_1)\, dx_1 = a + b\mu_2$$

$$\Rightarrow \mu_1 = a + b\mu_2 \quad \text{— (2)}$$

Again multiplying (1) by $x_2 f(x_2)$ and integrating w.r.t. x_2 we get.

$$\int x_1 x_2 f(x_1|x_2) f(x_2)\, dx_1\, dx_2 = a \int x_2 f(x_2)\, dx_2 + b \int x_2^2 f(x_2)\, dx_2$$

$$\Rightarrow E(X_1 X_2) = a\mu_2 + b\, E(X_2^2)$$

$$\Rightarrow \sigma_{12} + \mu_1\mu_2 = a\mu_2 + b(\mu_2^2 + \sigma_2^2) \quad \text{— (3)}$$

$Cov(X_1, X_2) = E(X_1 X_2) + E(X_1) E(X_2)$

Solving (2) & (3), we get.

$$-\mu_1\mu_2 = -a\mu_2 + b\mu_2^2$$

$$\sigma_{12} + \mu_1\mu_2 = a\mu_2 + b\mu_2^2 + b\sigma_2^2$$

$$\sigma_{12} = b\sigma_2^2$$

$$\Rightarrow b = \frac{\sigma_{12}}{\sigma_2^2} = \frac{\rho\sigma_1\sigma_2}{\sigma_2^2} = \frac{\rho\sigma_1}{\sigma_2}$$

Date

Substituting in (2), we get.

$$a = \mu_1 - \frac{\rho\sigma_1}{\sigma_2}\mu_2$$

$$\Rightarrow x_1^* = E(x_1 | x_2 = x_2) = \mu_1 - \frac{\rho\sigma_1}{\sigma_2}\mu_2 + \frac{\rho\sigma_1}{\sigma_2}x_2$$

$$= \mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2 - \mu_2)$$

Multiple correlation -

If x_1 is the first component of $\underline{X}$ & $\underline{X}^{(2)}$ the vector of remaining $(p-1)$ components. We express X_1 as a linear combination of $\underline{X}^{(2)}$ defined by the relation

$$X_1^* = \mu_1 + \underset{1\times p}{\underline{\beta}'}\,(\underline{X}^{(2)} - \underline{\mu}^{(2)})_{p\times 1}$$

$\underline{\beta}'$ is determined by minimizing.

$$U = E[X_1 - X_1^*]^2$$
$$= E[X_1 - \mu_1 - \underline{\beta}'(X^{(2)} - \mu^{(2)})]^2$$

Differentiating w.r.t. $\underline{\beta}$ and equating to 0,

$$-2E[(\underline{X}^{(2)} - \underline{\mu}^{(2)})(X_1 - \mu_1 - \underline{\beta}'(X^{(2)} - \underline{\mu}^{(2)})] = 0$$

$$\Rightarrow E(\underline{X}^{(2)} - \underline{\mu}^{(2)})(X_1 - \mu_1) - E(\underline{X}^{(2)} - \underline{\mu}^{(2)})(\underline{X}^{(2)} - \underline{\mu}^{(2)})'\underline{\beta} = 0$$

$$\Rightarrow \underline{\sigma}_{12} = \Sigma_{22}\underline{\beta}$$

$$\Rightarrow \hat{\underline{\beta}} = \Sigma_{22}^{-1}\underline{\sigma}_{12} \quad \text{or} \quad \hat{\underline{\beta}}' = \underline{\sigma}_{12}'\Sigma_{22}^{-1}$$

Date

Therefore the best linear fun. of X_1 in terms of $\underline{X}^{(2)}$ is

$$X_1^* = \mu_1 + \underline{\sigma}_{12}' \Sigma_{22}^{-1} (\underline{X}^{(2)} - \underline{\mu}^{(2)}) = \hat{X}_1 .$$

The corr. coeff. between X_1 & its best linear estimate in terms of $\underline{X}^{(2)}$ is called multiple correlation between X_1 & $X_2, \dots, X_p$ and is denoted by

$$\rho_{1(2,\dots p)} = \frac{Cov(X_1, \hat{X}_1)}{\sqrt{V(X_1)\, V(\hat{X}_1)}} , \quad V(X_1) = E[X_1 - E(X_1)]^2 = \sigma_{11}$$

$$V(\hat{X}_1) = E[(\hat{X}_1 - E(\hat{X}_1))][\hat{X}_1 - E(\hat{X}_1)]'$$
$$= E[\mu_1 + \underline{\sigma}_{12}' \Sigma_{22}^{-1} (\underline{X}^{(2)} - \underline{\mu}^{(2)}) - \mu_1][\mu_1 + \underline{\sigma}_{12}' \Sigma_{22}^{-1} (\underline{X}^{(2)} - \underline{\mu}^{(2)}) - \mu_1]'$$

$$= \underline{\sigma}_{12}' \Sigma_{22}^{-1} E(\underline{X}^{(2)} - \underline{\mu}^{(2)})(\underline{X}^{(2)} - \underline{\mu}^{(2)})' \Sigma_{22}^{-1} \underline{\sigma}_{12}$$

$$= \underline{\sigma}_{12}' \Sigma_{22}^{-1} \Sigma_{22} \Sigma_{22}^{-1} \underline{\sigma}_{12}$$

$$= \underline{\sigma}_{12}' \Sigma_{22}^{-1} \underline{\sigma}_{12}$$

$$Cov(X_1, \hat{X}_1) = E(X_1 - E(X_1))(\hat{X}_1 - E(\hat{X}_1))'$$
$$= E(X_1 - \mu_1)(\underline{\sigma}_{12}' \Sigma_{22}^{-1} (\underline{X}^{(2)} - \underline{\mu}^{(2)}))'$$

$$= E(X_1 - \mu_1)(\underline{X}^{(2)} - \underline{\mu}^{(2)})' \Sigma_{22}^{-1} \underline{\sigma}_{12}$$

$$= \underline{\sigma}_{12}' \Sigma_{22}^{-1} \underline{\sigma}_{12}$$

$$\Rightarrow \rho_{1(2,\dots p)} = \frac{\underline{\sigma}_{12}' \Sigma_{22}^{-1} \underline{\sigma}_{12}}{\sqrt{\sigma_{11}(\underline{\sigma}_{12}' \Sigma_{22}^{-1} \underline{\sigma}_{12})}}$$

$$= \sqrt{\frac{\underline{\sigma}_{12}' \Sigma_{22}^{-1} \underline{\sigma}_{12}}{\sigma_{11}}} = \sqrt{\frac{\underline{\beta}' \Sigma_{22} \underline{\beta}}{\sigma_{11}}}$$

Date

where $\underline{\sigma}_{12}$ & Σ_{22} are defined as $\begin{bmatrix} \sigma_{11} & \underline{\sigma}_{12}' \\ \underline{\sigma}_{12} & \Sigma_{22} \end{bmatrix}$

$\Rightarrow$ Since numerator is $\sqrt{V(\hat{X}_1)} \Rightarrow \rho_{1(2\cdots p)} \geq 0$; i.e.,

$$0 \leq \rho_{1(2\cdots p)} \leq 1$$

Estimation of multiple correlation coeff -

The multiple correlation coeff. in the popn. is

$$\rho_{1(2,\cdots p)} = \sqrt{\frac{\underline{\beta}' \Sigma_{22} \underline{\beta}}{\sigma_{11}}}$$

Given $\underline{x}_\alpha$ $(\alpha = 1 \cdots n)$; $n > p$, we estimate Σ by

$$\hat{\Sigma} = \frac{A}{n} = \frac{n-1}{n} S \text{,} \quad \text{where } A = \sum_\alpha (\underline{x}_\alpha - \bar{\underline{x}})(\underline{x}_\alpha - \bar{\underline{x}})'$$

Now A is partitioned as.

$$\frac{A}{n} = \begin{bmatrix} \frac{a_{11}}{n} & \frac{\underline{a}_{12}'}{n} \\ \frac{\underline{a}_{12}}{n} & \frac{A_{22}}{n} \end{bmatrix}$$ and the estimate of $\underline{\beta}$ is

$$\hat{\underline{\beta}}' = \underline{\sigma}_{12}' \Sigma_{22}^{-1} = \frac{\underline{a}_{12}'}{n} \left(\frac{A_{22}}{n}\right)^{-1} = \underline{a}_{12}' A_{22}^{-1}$$

Using the above estimates the sample multiple correlation coff. of X_1 on $X_2, \cdots, X_p$ is

$$R_{1(2,\cdots p)} = \sqrt{\frac{\hat{\underline{\sigma}}_{12}' \hat{\Sigma}_{22}^{-1} \hat{\underline{\sigma}}_{12}}{\hat{\sigma}_{11}}} = \sqrt{\frac{\underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11}}}$$

and $$1 - R^2_{1(2\cdots p)} = \frac{a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11}} = \frac{(a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{12})\, |A_{22}|}{a_{11} |A_{22}|}$$

Date

$$= \frac{|A|}{a_{11}|A_{22}|}$$

Distribution of sample multiple correlation coefficient in null case:-

The sample multiple correlation coefficient between X_1 and $\underline{X}^{(2)}$ is defined by relation

$$R^2 = \frac{\underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11}} \quad \& \quad 1 - R^2 = \frac{a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11}}$$

where $R^2 = R^2_{1(2,\dots,p)}$ & $A = \begin{bmatrix} a_{11} & \underline{a}_{12}' \\ \underline{a}_{12} & A_{22} \end{bmatrix}$

Therefore,

$$\frac{R^2}{1-R^2} = \frac{\underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11.2}}$$

We know that, if A is partitioned as.

$$A = \begin{array}{c} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{array}{l} q \\ p-q \end{array} \end{array}$$ & $A \sim W_p(n-1, \Sigma)$ then $A_{11} \sim W_q(n-1, \Sigma_{11})$ and $A_{11.2} \sim W_q(n-1-(p-q)), \Sigma_{11.2})$

Thus, in our case.

$$a_{11} \sim W_1(n-1, \sigma_{11}) \Rightarrow \frac{a_{11}}{\sigma_{11}} \sim \chi^2_{n-1}$$

In null case $\rho_{1(2,3,\dots p)} = 0$

$$\Sigma_{11.2} = \sigma_{11} - \underline{\sigma}_{12}' \Sigma_{22}^{-1} \underline{\sigma}_{12} = \sigma_{11}, \quad \text{since } \underline{\sigma}_{12}' = 0$$

So that $\quad a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{12} \sim W_1(n-1-(p-1), \sigma_{11})$

$$\Rightarrow \frac{a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{\sigma_{11}} \sim \chi^2_{n-p}$$

Date

Consider.

$$\frac{a_{11}}{\sigma_{11}} = \frac{a_{11} - a_{12}' A_{22}^{-1} a_{12}}{\sigma_{11}} + \frac{a_{12}' A_{22}^{-1} a_{12}}{\sigma_{11}}$$

or $Q = Q_1 + Q_2$ (say)

where $Q \sim \chi^2_{n-1}$, $Q_1 \sim \chi^2_{n-p}$

from, Fisher Cochran's thm. Q_2 is independently distributed as $\chi^2_{n-1-(n-p)}$ i.e. $Q_2 \sim \chi^2_{p-1}$ & is ind. of Q_1, hence.

$$f = \frac{R^2}{1-R^2} \times \frac{n-p}{p-1} = \frac{a_{12}' A_{22}^{-1} a_{12} / \sigma_{11}}{a_{11\cdot 2} / \sigma_{11}} \times \frac{n-p}{p-1}$$

$$= \frac{\chi^2_{p-1} / p-1}{\chi^2_{n-p} / n-p} \sim F_{p-1,\, n-p}$$

The distn. of the statistic f is.

$$df(F) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} f^{\frac{\nu_1}{2}-1}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)\left(1 + \frac{\nu_1}{\nu_2} f\right)^{\frac{\nu_1+\nu_2}{2}}} \cdot df$$

where $\nu_1 = p-1$, $\nu_2 = n-p$.

In this put $f = \frac{R^2}{1-R^2} \frac{\nu_2}{\nu_1}$ then $df = \frac{dR^2}{(1-R^2)^2} \frac{\nu_2}{\nu_1}$

$$df(R^2) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \left(\frac{R^2}{1-R^2} \cdot \frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{2}-1}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)\left(1 + \frac{R^2}{1-R^2}\right)^{\frac{\nu_1+\nu_2}{2}}} \frac{\nu_2}{\nu_1} \frac{dR^2}{(1-R^2)^2}$$

$$= \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2} - \frac{\nu_1}{2} + 1 - 1} \left(\frac{R^2}{1-R^2}\right)^{\frac{\nu_1}{2}-1}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)\left(\frac{1}{1-R^2}\right)^{\frac{\nu_1+\nu_2}{2}}} \frac{dR^2}{(1-R^2)^2}$$

Date

$$= \frac{1}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} (R^2)^{\frac{\nu_1}{2}-1} (1-R^2)^{\frac{\nu_1+\nu_2}{2} - \frac{\nu_1}{2} + 1 - 2} dR^2$$

$$= \frac{1}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \cdot (R^2)^{\frac{\nu_1}{2}-1} (1-R^2)^{\frac{\nu_2}{2}-1} dR^2$$

Put $dR^2 = 2R\,dR$ in the dist^n of R,

$$df(R) = \frac{1}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} (R)^{\nu_1 - 2} (1-R^2)^{\frac{\nu_2}{2}-1} \cdot 2R\,dR.$$

$$= \frac{2R^{(\nu_1 - 1)} (1-R^2)^{\frac{\nu_2}{2}-1}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)}$$

$$= \frac{2R^{p-2} (1-R^2)^{\frac{n-p}{2}-1}}{B\left(\frac{p-1}{2}, \frac{n-p}{2}\right)} dR \quad , 0 < R < 1.$$