Multivariate Statistics

Lecture 09

Fudan University

Outline

Outline

The Distribution of the Sample Correlation Coefficient

If one has a sample (of p-component vectors) $\mathbf{x}_1, \dots, \mathbf{x}_N$ from a normal distribution, the maximum likelihood estimator of the correlation between the i-th component and the j-th component is

$$r_{ij} = \frac{\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)}{\sqrt{\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)^2}\sqrt{\sum_{\alpha=1}^{N}(x_{j\alpha} - \bar{x}_j)^2}},$$

where $x_{i\alpha}$ is the i-th component of $\mathbf{x}_\alpha$ and

$$\bar{x}_i = \frac{1}{N}\sum_{\alpha=1}^{N} x_{i\alpha}.$$

We shall treat that r_{ij} and need only consider the joint distribution of $(x_{i1}, x_{j1}), (x_{i2}, x_{j2}), \dots, (x_{iN}, x_{jN})$.

The Distribution of the Sample Correlation Coefficient

We reformulate the problems to be considered a bivariate normal distribution. Let $\mathbf{x}_1^*, \ldots, \mathbf{x}_N^*$ be observation from

$$\mathcal{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2\end{bmatrix}\right), \quad \text{where } -1 < \rho < 1.$$

We shall consider the sample correlation coefficient

$$r = \frac{a_{12}}{\sqrt{a_{11}}\sqrt{a_{22}}}$$

where

$$a_{ij} = \sum_{\alpha=1}^{N} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j), \qquad \bar{x}_i = \frac{1}{N}\sum_{\alpha=1}^{N} x_{i\alpha}$$

and $x_{i\alpha}$ is the i-th component of $\mathbf{x}_\alpha^*$.

Let $n = N - 1$. We see that a_{ij} are distributed like

$$a_{ij} = \sum_{\alpha=1}^{n} z_{i\alpha} z_{j\alpha}$$

where

$$\begin{bmatrix} z_{1\alpha} \\ z_{2\alpha} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix} \right).$$

and the pair $(z_{12}, z_{22}), \ldots, (z_{1N}, z_{2N})$ are independent.

The Distribution of the Sample Correlation Coefficient

Define the n-component vectors $\mathbf{v}_i = [z_{i1}, \ldots, z_{in}]^\top$ for $i = 1, 2$.

1. The correlation coefficient between $\mathbf{v}_1$ and $\mathbf{v}_2$ is the cosine of the angle, say θ, between $\mathbf{v}_1$ and $\mathbf{v}_2$, that is

$$\cos\theta = \frac{\mathbf{v}_1^\top \mathbf{v}_2}{\|\mathbf{v}_1\|_2 \|\mathbf{v}_2\|_2}.$$

2. If we let $b = \mathbf{v}_2^\top \mathbf{v}_1 / (\mathbf{v}_1^\top \mathbf{v}_1)$ then $\mathbf{v}_2 - b\mathbf{v}_1$ is orthogonal to $\mathbf{v}_1$ and

$$\cot\theta = \frac{b\|\mathbf{v}_1\|_2}{\|\mathbf{v}_2 - b\mathbf{v}_1\|_2}.$$

3. We shall show that $\cot\theta$ is proportional to a t-variable when $\rho = 0$.

The Distribution of the Sample Correlation Coefficient

Theorem 1

If the pairs $(z_{11}, z_{21}), \ldots, (z_{1n}, z_{2n})$ are independent and each pair are distributed according to

$$\begin{bmatrix} z_{1\alpha} \\ z_{2\alpha} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}\right), \quad \text{where } \alpha = 1, \ldots, n,$$

then given $z_{11}, z_{12}, \ldots, z_{1n}$, the conditional distributions of

$$b = \frac{\sum_{\alpha=1}^n z_{2\alpha} z_{1\alpha}}{\sum_{i=1}^n z_{1\alpha}^2} \quad \text{and} \quad \frac{u}{\sigma^2} = \sum_{\alpha=1}^n \frac{(z_{2\alpha} - bz_{1\alpha})^2}{\sigma^2}$$

are $\mathcal{N}\left(\beta, \sigma^2/c^2\right)$ and χ^2-distribution with $n-1$ degrees of freedom, respectively; and b and U are independent, where

$$\beta = \frac{\rho\sigma_2}{\sigma_1}, \quad \sigma^2 = \sigma_2^2(1-\rho^2) \quad \text{and} \quad c^2 = \sum_{i=1}^n z_{1\alpha}^2.$$

The Distribution of the Sample Correlation Coefficient

We can write

$$\cot\theta = \frac{b\left\|\mathbf{v}_1\right\|_2}{\left\|\mathbf{v}_2 - b\mathbf{v}_1\right\|_2} = \frac{cb/\sigma}{\sqrt{u/\sigma^2}}$$

If $\rho = 0$, then $\beta = 0$, and $b \sim \mathcal{N}(0, \sigma^2/c^2)$, and

$$\frac{cb/\sigma}{\sqrt{\frac{u/\sigma^2}{n-1}}} \sim \frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}}$$

has a conditional t-distribution with $n-1$ degrees of freedom.

We require the following lemma.

Lemma 1

If $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are independently distributed, if

$$\mathbf{y}_\alpha = \begin{bmatrix} \mathbf{y}_\alpha^{(1)} \\ \mathbf{y}_\alpha^{(2)} \end{bmatrix}$$

has the density $f(\mathbf{y}_\alpha)$ and if the conditional density of $\mathbf{y}_\alpha^{(2)}$ given $\mathbf{y}_\alpha^{(1)}$ is $f\left(\mathbf{y}_\alpha^{(2)} \mid \mathbf{y}_\alpha^{(1)}\right)$ for $\alpha = 1, \ldots, n$. Then in the conditional distribution of $\mathbf{y}_1^{(2)}, \ldots, \mathbf{y}_N^{(2)}$ given $\mathbf{y}_1^{(1)}, \ldots, \mathbf{y}_N^{(1)}$, the random vectors $\mathbf{y}_1^{(2)}, \ldots, \mathbf{y}_N^{(2)}$ are independent and the density of $\mathbf{y}_\alpha^{(2)}$ is $f\left(\mathbf{y}_\alpha^{(2)} \mid \mathbf{y}_\alpha^{(1)}\right)$.

The Distribution of the Sample Correlation Coefficient

We also use the following lemma with $x_\alpha = z_{2\alpha}$ and matrix $\mathbf{C}$ whose the first row is $\mathbf{v}_1^\top / c$, where $c = \|\mathbf{v}_1\|_2$.

Lemma 2

Suppose $\mathbf{x}_1, \dots, \mathbf{x}_N$ are independent, where $\mathbf{x}_\alpha \sim \mathcal{N}_p(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma})$. Let $\mathbf{C} \in \mathbb{R}^{N \times N}$ be an orthogonal matrix, then

$$\mathbf{y}_\alpha = \sum_{\gamma=1}^{N} c_{\alpha\gamma} \mathbf{x}_\gamma \sim \mathcal{N}_p(\boldsymbol{\nu}_\alpha, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\nu}_\alpha = \sum_{\gamma=1}^{N} c_{\alpha\gamma} \boldsymbol{\mu}_\gamma$ for $\alpha = 1, \dots, N$ and $\mathbf{y}_1, \dots, \mathbf{y}_N$ are independent.

The Distribution of the Sample Correlation Coefficient

Theorem 2

if x and y are independently distributed, x having the distribution $\mathcal{N}(0,1)$ and y having the χ^2-distribution with m degrees of freedom, then

$$t = \frac{x}{\sqrt{y/m}}$$

has the density of t-distribution such that

$$f(t;m) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\,\Gamma\left(\frac{m}{2}\right)}\left(1+\frac{t^2}{m}\right)^{-\frac{m+1}{2}}.$$

The Distribution of the Sample Correlation Coefficient

Recall that $a_{ij} = \sum_{\alpha=1}^{n} z_{i\alpha} z_{j\alpha}$ and $\mathbf{v}_i = [z_{i1}, \ldots, z_{in}]^\top$ for $i = 1, 2$, then

$$b = \frac{\sum_{\alpha=1}^{n} z_{2\alpha} z_{1\alpha}}{\sum_{i=1}^{n} z_{1\alpha}^2} = \frac{a_{12}}{a_{11}}, \quad c^2 = \sum_{i=1}^{n} z_{1\alpha}^2 = a_{11}$$

$$u = \sum_{\alpha=1}^{n} (z_{2\alpha} - b z_{1\alpha})^2 = \sum_{\alpha=1}^{n} (z_{2\alpha}^2 - b^2 z_{1\alpha}^2) = a_{22} - \frac{a_{12}^2}{a_{11}}.$$

Hence, we can write the above conditional t-distributed random variable with $n-1$ degrees of freedom as

$$\begin{aligned}
\frac{cb/\sigma}{\sqrt{\frac{u/\sigma^2}{n-1}}} &= \sqrt{n-1} \cdot \frac{cb}{\sqrt{u}} \\
&= \sqrt{n-1} \cdot \frac{a_{12}/\sqrt{a_{11} a_{22}}}{\sqrt{1 - a_{12}^2/(a_{11} a_{22})}} \\
&= \sqrt{n-1} \cdot \frac{r}{\sqrt{1-r^2}}.
\end{aligned}$$

The Distribution of the Sample Correlation Coefficient

The conditional density of

$$t = \frac{cb/\sigma}{\sqrt{\frac{u/\sigma^2}{n-1}}} = \sqrt{n-1} \cdot \frac{r}{\sqrt{1-r^2}}$$

given $\mathbf{v}_1$ is

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{(n-1)\pi}\,\Gamma\left(\frac{n-1}{2}\right)} \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}.$$

Then the conditional density of r given $\mathbf{v}_1$ is

$$k_N(r) = \frac{\Gamma\left(\frac{N-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{N-2}{2}\right)} (1-r^2)^{\frac{N-4}{2}}, \quad \text{where} \quad N = n+1.$$

We can verify that

$$\mathbb{E}\left[r^{2m}\right] = \frac{\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(m+\frac{1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{N-1}{2}+m\right)}.$$

Outline

Tests for the Hypothesis of Lack of Correlation

Consider the hypothesis $H : \rho_{ij} = 0$ for some particular pair (i, j).

1. For testing H against alternatives $\rho_{ij} > 0$, we reject H if $r_{ij} > r_0$ for some positive r_0. The probability of rejecting H when H is true is

$$\int_{r_0}^{1} k_N(r)\,\mathrm{d}r.$$

2. For testing H against alternatives $r_{ij} < 0$, we reject H if $r_{ij} < -r_0$.
3. For testing H against alternatives $r_{ij} \neq 0$, we reject H if $r_{ij} > r_1$ or $r_{ij} < -r_1$ for some positive r_1. The probability of rejection when H is true is

$$\int_{-1}^{-r_1} k_N(r)\,\mathrm{d}r + \int_{r_1}^{1} k_N(r)\,\mathrm{d}r.$$

Tests for the Hypothesis of Lack of Correlation

We have shown that

$$\sqrt{N-2} \cdot \frac{r_{ij}}{\sqrt{1 - rij^2}}$$

has the t-distribution with $N - 2$ degrees of freedom.

We can also use t-tables. For $\rho_{ij} \neq 0$, reject H if

$$\sqrt{N-2} \cdot \frac{|r_{ij}|}{\sqrt{1 - r_{ij}^2}} > t_{N-2}(\alpha),$$

where $t_{N-2}(\alpha)$ is the two-tailed significance point of the t-statistic with $N - 2$ degrees of freedom for significance level α.

The Distribution in the Case of $\rho \neq 0$

Conditional on $\mathbf{v}_1$ held fixed, the random variables

$$b = \frac{a_{12}}{a_{11}} \quad \text{and} \quad \frac{u}{\sigma^2} = \frac{a_{22} - a_{12}^2/a_{11}}{\sigma^2},$$

which are distributed independently according to $\mathcal{N}(\beta, \sigma^2/c^2)$ and χ^2-distribution with $n-1$ degrees of freedom, respectively.

Theorem 3

The correlation coefficient in a sample of N from a bivariate normal distribution with correlation ρ is distributed with density

$$\frac{2^{n-2}(1-\rho^2)^{\frac{n}{2}}(1-r^2)^{\frac{n-3}{2}}}{(n-2)!\pi} \sum_{\alpha=0}^{\infty} \frac{(2\rho r)^\alpha}{\alpha!} \Gamma^2\left(\frac{n+\alpha}{2}\right),$$

where $-1 \leq r \leq 1$ and $n = N-1$.

The Distribution in the Case of $\rho \neq 0$

It should be pointed out that any test based on r is invariant under transformations of location and scale, that is,

$$x_{i\alpha}^* = b_i x_{i\alpha} + c_i,$$

for $b_i \neq 0$ and $i = 1, 2$.

Test $\rho = \rho_0$ by the Likelihood Ratio Criterion

The likelihood ratio criterion:

1. Let $L(\mathbf{x}, \boldsymbol{\theta})$ be the likelihood function of the observation $\mathbf{x}$ and the parameter vector $\boldsymbol{\theta} \in \Omega$.
2. Let a null hypothesis be defined by a proper subset ω of Ω. The likelihood ratio criterion is
$$\lambda(\mathbf{x}) = \frac{\sup_{\boldsymbol{\theta} \in \omega} L(\mathbf{x}, \boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Omega} L(\mathbf{x}, \boldsymbol{\theta})}.$$
3. The likelihood ratio test is the procedure of rejecting the null hypothesis when $\lambda(\mathbf{x})$ is less than a predetermined constant.

Test $\rho = \rho_0$ by the Likelihood Ratio Criterion

Let us consider the likelihood ratio test of the hypothesis that $\rho = \rho_0$ based on a sample $\mathbf{x}_1, \ldots, \mathbf{x}_N$ from the bivariate normal distribution

$$\mathcal{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2\end{bmatrix}\right).$$

The set Ω consists of $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ such that

$$\sigma_1 > 0, \quad \sigma_2 > 0 \quad \text{and} -1 < \rho < 1$$

and the set ω is the subset for which $\rho = \rho_0$.

The likelihood ratio criterion is

$$\frac{\sup_\omega L(\mathbf{x}, \boldsymbol{\theta})}{\sup_\Omega L(\mathbf{x}, \boldsymbol{\theta})} = \left(\frac{(1-\rho_0^2)(1-r^2)}{(1-\rho_0 r)^2}\right)^{\frac{N}{2}}.$$

Test $\rho = \rho_0$ by the Likelihood Ratio Criterion

The likelihood ratio criterion is

$$\frac{\sup_{\omega} L(\mathbf{x}, \boldsymbol{\theta})}{\sup_{\Omega} L(\mathbf{x}, \boldsymbol{\theta})} = \left(\frac{(1-\rho_0^2)(1-r^2)}{(1-\rho_0 r)^2} \right)^{\frac{N}{2}}.$$

The likelihood ratio test is

$$\frac{(1-\rho_0^2)(1-r^2)}{(1-\rho_0 r)^2} \leq c$$

where c is chosen by the prescribed significance level.

Test $\rho = \rho_0$ by the Likelihood Ratio Criterion

The critical region can be written equivalently as

$$(\rho_0^2 c - \rho_0^2 + 1) r^2 - 2\rho_0 c r + c - 1 + \rho_0^2 \geq 0,$$

that is,

$$r > \frac{\rho_0 c + (1 - \rho_0^2)\sqrt{1 - c}}{\rho_0^2 c - \rho_0^2 + 1} \quad \text{and} \quad r < \frac{\rho_0 c - (1 - \rho_0^2)\sqrt{1 - c}}{\rho_0^2 c - \rho_0^2 + 1}.$$

Thus the likelihood ratio test of $H: \rho = \rho_0$ against alternatives $\rho \neq \rho_0$ has a rejection region of the form $r > r_1$ and $r < r_2$ (not chosen so that the probability of each inequality is $\alpha/2$ when H is true).

Outline

The Asymptotic Distribution of Sample Correlation

For a sample $\mathbf{x}_1, \ldots, \mathbf{x}_N$ from a normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we are interested in the sample correlation coefficient

$$r(n) = \frac{a_{ij}(n)}{\sqrt{a_{ii}(n)}\sqrt{a_{jj}(n)}}$$

where $n = N - 1$,

$$a_{ij}(n) = \sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) \sim \sum_{\alpha=1}^{n} \begin{bmatrix} z_{i\alpha} \\ z_{j\alpha} \end{bmatrix} \begin{bmatrix} z_{i\alpha} & z_{j\alpha} \end{bmatrix}$$

with

$$\begin{bmatrix} z_{i\alpha} \\ z_{j\alpha} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ji} & \sigma_{jj} \end{bmatrix}\right) \quad \text{and} \quad \bar{x}_i = \frac{1}{N}\sum_{\alpha=1}^{N} x_{i\alpha}.$$

The Asymptotic Distribution of Sample Correlation

We can also write

$$r(n) = \frac{c_{ij}(n)}{\sqrt{c_{ii}(n)}\sqrt{c_{jj}(n)}},$$

with

$$c_{ii}(n) = \frac{a_{ii}(n)}{\sigma_{ii}}, c_{ij}(n) = \frac{a_{ij}(n)}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} \quad \text{and} \quad c_{jj}(n) = \frac{a_{ii}(n)}{\sigma_{jj}}.$$

Then we have

$$c_{ij}(n) = \sum_{\alpha=1}^{n} \begin{bmatrix} z_{i\alpha}^* \\ z_{j\alpha}^* \end{bmatrix} \begin{bmatrix} z_{i\alpha}^* & z_{j\alpha}^* \end{bmatrix}$$

with

$$\begin{bmatrix} z_{i\alpha}^* \\ z_{j\alpha}^* \end{bmatrix} = \begin{bmatrix} \frac{z_{i\alpha}}{\sqrt{\sigma_{ii}}} \\ \frac{z_{j\alpha}}{\sqrt{\sigma_{jj}}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \quad \text{and} \quad \rho = \frac{\sigma_{ij}}{\sqrt{\sigma_{jj}}\sqrt{\sigma_{jj}}}.$$

The Asymptotic Distribution of Sample Correlation

Let

$$\mathbf{u}(n) = \frac{1}{n}\begin{bmatrix} c_{ii}(n) \\ c_{jj}(n) \\ c_{ij}(n) \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ \rho \end{bmatrix}$$

The vector

$$\sqrt{n}(\mathbf{u}(n) - \mathbf{b}) = \frac{1}{\sqrt{n}}\left(\begin{bmatrix} c_{ii}(n) \\ c_{jj}(n) \\ c_{ij}(n) \end{bmatrix} - n\mathbf{b}\right)$$

has a limiting normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\begin{bmatrix} 2 & 2\rho^2 & 2\rho \\ 2\rho^2 & 2 & 2\rho \\ 2\rho & 2\rho & 1+\rho^2 \end{bmatrix}.$$

The Asymptotic Distribution of Sample Correlation

Apply the following theorem with $\mathbf{A}(n) = \mathbf{C}(n)$ and $\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

Theorem 4

Let

$$\mathbf{A}(n) = \sum_{\alpha=1}^{N} \left(\mathbf{x}_\alpha - \bar{\mathbf{x}}_N\right)\left(\mathbf{x}_\alpha - \bar{\mathbf{x}}_N\right)^\top,$$

where $\mathbf{x}_1, \dots, \mathbf{x}_N$ are independently distributed according to $\mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ and $n = N - 1$. Then the limiting distribution of

$$\mathbf{B}(n) = \frac{1}{\sqrt{n}}\left(\mathbf{A}(n) - n\mathbf{\Sigma}\right)$$

is normal with mean $\mathbf{0}$ and covariance $\mathbb{E}\left[b_{ij}(n) b_{kl}(n)\right] = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}$.

The Asymptotic Distribution of Sample Correlation

The sample correlation coefficient can be written as $r = \frac{u_3}{\sqrt{u_1}\sqrt{u_2}}$.

Theorem 5 [Serfling (1980), Section 3.3]

Let $\{\mathbf{u}(n)\}$ be a sequence of m-component random vectors and $\mathbf{b}$ a fixed vector such that

$$\lim_{n\to\infty} \sqrt{n}(\mathbf{u}(n) - \mathbf{b}) \sim \mathcal{N}(\mathbf{0}, \mathbf{T}).$$

Let $\mathbf{f}(\mathbf{u})$ be a vector-valued function of $\mathbf{u}$ such that each component $f_j(\mathbf{u})$ has a nonzero differential at $\mathbf{u} = \mathbf{b}$, and let

$$\frac{\partial f_j(\mathbf{u})}{\partial u_i}\bigg|_{\mathbf{u}=\mathbf{b}}$$

be the (i, j)-th component of $\mathbf{\Phi_b}$. Then $\sqrt{n}(\mathbf{f}(\mathbf{u}(n)) - f(\mathbf{b}))$ has the limiting distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Phi_b}^\top \mathbf{T} \mathbf{\Phi_b})$.

The Asymptotic Distribution of Sample Correlation

Applying Theorem 5 with $r = f(\mathbf{u}) = u_3 u_1^{-\frac{1}{2}} u_2^{-\frac{1}{2}}$, we have $f(\mathbf{b}) = \rho$ and

$$\mathbf{\Phi}_{\mathbf{b}} = \begin{bmatrix} \frac{\partial r}{\partial u_1}\Big|_{\mathbf{u}=\mathbf{b}} \\ \frac{\partial r}{\partial u_2}\Big|_{\mathbf{u}=\mathbf{b}} \\ \frac{\partial r}{\partial u_3}\Big|_{\mathbf{u}=\mathbf{b}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} u_3 u_1^{-\frac{3}{2}} u_2^{-\frac{1}{2}}\Big|_{\mathbf{u}=\mathbf{b}} \\ -\frac{1}{2} u_3 u_1^{-\frac{1}{2}} u_2^{-\frac{3}{2}}\Big|_{\mathbf{u}=\mathbf{b}} \\ u_1^{-\frac{1}{2}} u_2^{-\frac{1}{2}}\Big|_{\mathbf{u}=\mathbf{b}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\rho \\ -\frac{1}{2}\rho \\ 1 \end{bmatrix}.$$

Thus, the covariance of the limiting distribution of $\sqrt{n}(r(n) - \rho)$ is

$$\begin{bmatrix} -\frac{1}{2}\rho & -\frac{1}{2}\rho & 1 \end{bmatrix} \begin{bmatrix} 2 & 2\rho^2 & 2\rho \\ 2\rho^2 & 2 & 2\rho \\ 2\rho & 2\rho & 1+\rho^2 \end{bmatrix} \begin{bmatrix} -\frac{1}{2}\rho \\ -\frac{1}{2}\rho \\ 1 \end{bmatrix} = (1-\rho^2)^2$$

and we have $\lim_{n\to\infty} \frac{\sqrt{n}(r(n) - \rho)}{1-\rho^2} \sim \mathcal{N}(0,1)$.

The Asymptotic Distribution of Sample Correlation

If $f(x)$ is differentiable at $x = \rho$ with non-zero differential, then

$$\sqrt{n}(f(r) - f(\rho))$$

is asymptotically normally distributed with mean zero and variance

$$\left(\frac{\partial f}{\partial x}\Big|_{x=\rho} \right)^2 \left(1 - \rho^2\right)^2.$$

Theorem 6 [Fisher's z]

Let

$$z = \frac{1}{2} \log \frac{1+r}{1-r} \quad \text{and} \quad \zeta = \frac{1}{2} \log \frac{1+\rho}{1-\rho}$$

where r is the correlation coefficient of a sample of $N = n + 1$ from a bivariate normal distribution with correlation ρ. Then $\sqrt{n}\,(z - \zeta)$ has a limiting normal distribution with mean 0 and variance 1.

The Asymptotic Distribution of Sample Correlation

Fisher's z approaches to normality much more rapid than for r. We have

$$\mathbb{E}[z] \simeq \zeta + \frac{\rho}{2n} \qquad \text{and} \qquad \mathbb{E}\left[z - \zeta - \frac{\rho}{2n}\right]^2 \simeq \frac{1}{n-2}.$$

See "Hotelling, H. (1953). New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society. Series B (Methodological)*, 15(2), 193-232."

We wish to test the hypothesis $\rho = \rho_0$ on the basis of a sample of N against the alternatives $\rho \neq \rho_0$.

1. We compute r and $z = \frac{1}{2}\log\frac{1+r}{1-r}$.
2. Let $\zeta_0 = \frac{1}{2}\log\frac{1+\rho_0}{1-\rho_0}$.
3. Then a region of rejection at the 5% significance interval is

$$\sqrt{N-3}\left|z - \zeta_0 - \frac{\rho_0}{2(N-1)}\right| > 1.96.$$

Outline

Partial Correlation Coefficients

Consider the normal distribution $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

then the conditional distribution of $\mathbf{x}^{(1)}$ given $\mathbf{x}^{(2)}$ is

$$\mathbf{x}^{(1)} \mid \mathbf{x}^{(2)} \sim \mathcal{N}\left(\boldsymbol{\mu}^{(1)} + \mathbf{B}\big(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}\big), \boldsymbol{\Sigma}_{11.2}\right),$$

where

$$\mathbf{B} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \quad \text{and} \quad \boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

Partial Correlation Coefficient

The partial correlations of $\mathbf{x}^{(1)}$ given $\mathbf{x}^{(2)}$ are the correlations calculated in the usual way from $\boldsymbol{\Sigma}_{11.2}$.

Suppose $\mathbf{x}^{(1)}$ has q components and let

$$\boldsymbol{\Sigma}_{11.2} = \begin{bmatrix} \sigma_{11\cdot q+1,\dots,p} & \sigma_{12\cdot q+1,\dots,p} & \cdots & \sigma_{1q\cdot q+1,\dots,p} \\ \sigma_{21\cdot q+1,\dots,p} & \sigma_{22\cdot q+1,\dots,p} & \cdots & \sigma_{2q\cdot q+1,\dots,p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{q1\cdot q+1,\dots,p} & \sigma_{q2\cdot q+1,\dots,p} & \cdots & \sigma_{qq\cdot q+1,\dots,p} \end{bmatrix} \in \mathbb{R}^{q\times q}.$$

We define

$$\rho_{ij\cdot q+1,\dots,p} = \frac{\sigma_{ij\cdot q+1,\dots,p}}{\sqrt{\sigma_{ii\cdot q+1,\dots,p}}\sqrt{\sigma_{jj\cdot q+1,\dots,p}}}$$

as the partial correlation between x_i and x_j holding $x_{q+1},\dots,x_p$ fixed.

Partial Correlation Coefficient

Corollary 1

If on the basis of a given sample $\hat{\theta}_1, \ldots, \hat{\theta}_m$ are maximum likelihood estimators of the parameters $\theta_1, \ldots, \theta_m$ of a distribution, then $\phi_1(\hat{\theta}_1, \ldots, \hat{\theta}_m), \ldots, \phi_m(\hat{\theta}_1, \ldots, \hat{\theta}_m)$ are maximum likelihood estimator of $\phi_1(\theta_1, \ldots, \theta_m), \ldots, \phi_m(\theta_1, \ldots, \theta_m)$ if the transformation from $\theta_1, \ldots, \theta_m$ to $\phi_1, \ldots, \phi_m$ is one-to-one. If the estimators of $\theta_1, \ldots, \theta_m$ are unique, then the estimators of $\theta_1, \ldots, \theta_m$ are unique.

The Estimation of Partial Correlation Coefficient

Theorem 6

Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be a sample from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and partition the variables as

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Define $\mathbf{B} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$,

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{\mathbf{x}}^{(1)} \\ \bar{\mathbf{x}}^{(2)} \end{bmatrix} = \frac{1}{N}\sum_{\alpha=1}^{N} \begin{bmatrix} \mathbf{x}_\alpha^{(1)} \\ \mathbf{x}_\alpha^{(2)} \end{bmatrix} \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \sum_{\alpha=1}^{N} (\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})^\top.$$

Then the maximum likelihood estimators of $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$, $\mathbf{B}$, $\boldsymbol{\Sigma}_{11.2}$ and $\boldsymbol{\Sigma}_{22}$ are

$$\hat{\boldsymbol{\mu}}^{(1)} = \bar{\mathbf{x}}^{(1)}, \quad \hat{\boldsymbol{\mu}}^{(2)} = \bar{\mathbf{x}}^{(2)}, \quad \hat{\mathbf{B}} = \mathbf{A}_{12}\mathbf{A}_{22}^{-1},$$

$$\hat{\boldsymbol{\Sigma}}_{11.2} = \frac{1}{N}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}) \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_{22} = \frac{1}{N}\mathbf{A}_{22}.$$

The Estimation of Partial Correlation Coefficient

Then the maximum likelihood estimators of the partial correlation coefficients are

$$\hat{\rho}_{ij\cdot q+1,\ldots,p} = \frac{\hat{\sigma}_{ij\cdot q+1,\ldots,p}}{\sqrt{\hat{\sigma}_{ii\cdot q+1,\ldots,p}}\sqrt{\hat{\sigma}_{jj\cdot q+1,\ldots,p}}},$$

where $\hat{\sigma}_{ij\cdot q+1,\ldots,p}$ is the (i,j)-th element of $\hat{\mathbf{\Sigma}}_{11.2}$.

We can also write

$$\hat{\rho}_{ij\cdot q+1,\ldots,p} = \frac{a_{ij\cdot q+1,\ldots,p}}{\sqrt{a_{ii\cdot q+1,\ldots,p}}\sqrt{a_{jj\cdot q+1,\ldots,p}}},$$

where $a_{ij\cdot q+1,\ldots,p}$ is the (i,j)-th element of $\mathbf{A}_{11.2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$.

The Distribution of Partial Correlation Coefficient

To obtain the distribution of ρ_{ij} we showed that $\mathbf{A}$ was distributed as

$$\mathbf{A} = \sum_{\alpha=1}^{N-1} \mathbf{z}_\alpha \mathbf{z}_\alpha^\top$$

where $\mathbf{z}_\alpha$ are distributed independently according to $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$.

Here we want to show that $\mathbf{A}_{11.2}$ is distributed as

$$\mathbf{A}_{11.2} = \sum_{\alpha=1}^{N-1-(p-q)} \mathbf{u}_\alpha \mathbf{u}_\alpha^\top$$

where $\mathbf{u}_\alpha$ are distributed independently according to $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{11.2})$.

The Distribution of Partial Correlation Coefficient

Theorem 7

Suppose $\mathbf{y}_1, \ldots, \mathbf{y}_m$ are independent with $\mathbf{y}_\alpha$ distributed according to $\mathcal{N}(\boldsymbol{\Gamma}\mathbf{w}_\alpha, \boldsymbol{\Phi})$, where $\mathbf{w}_\alpha$ is an r-component vector. Let $\mathbf{H} = \sum_{\alpha=1}^m \mathbf{w}_\alpha \mathbf{w}_\alpha^\top$ assumed non-singular, $\mathbf{G} = \sum_{\alpha=1}^m \mathbf{y}_\alpha \mathbf{w}_\alpha^\top \mathbf{H}^{-1}$ and

$$\mathbf{C} = \sum_{\alpha=1}^{m} (\mathbf{y}_\alpha - \mathbf{G}\mathbf{w}_\alpha)(\mathbf{y}_\alpha - \mathbf{G}\mathbf{w}_\alpha)^\top = \sum_{\alpha=1}^{m} \mathbf{y}_\alpha \mathbf{y}_\alpha^\top - \mathbf{G}\mathbf{H}\mathbf{G}^\top.$$

Then $\mathbf{C}$ is distributed as $\sum_{\alpha=1}^{m-r} \mathbf{u}_\alpha \mathbf{u}_\alpha^\top$ and where $\mathbf{u}_1, \ldots, \mathbf{u}_{m-r}$ are independently distributed according to $\mathcal{N}(\mathbf{0}, \boldsymbol{\Phi})$ independently of $\mathbf{G}$.

Corollary 2

If $\Gamma = \mathbf{0}$, the matrix $\mathbf{G}\mathbf{H}\mathbf{G}^\top$ defined in Theorem 7 is distributed as $\sum_{\alpha=m-r+1}^{m} \mathbf{u}_\alpha \mathbf{u}_\alpha^\top$, where $\mathbf{u}_{m-r+1}, \ldots, \mathbf{u}_m$ are independently distributed, each according to $\mathcal{N}(\mathbf{0}, \boldsymbol{\Phi})$.

The Distribution of Partial Correlation Coefficient

We can write $\mathbf{A} = \sum_{\alpha=1}^{N-1} \mathbf{z}_\alpha \mathbf{z}_\alpha^\top$, where $\mathbf{z}_1, \ldots, \mathbf{z}_{N-1}$ are independent, each with distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$.

Let $\mathbf{z}_\alpha$ be partitioned into two subvectors of q and $p-q$ components, that is $\mathbf{z}_\alpha^\top = \left[\left(\mathbf{z}_\alpha^{(1)}\right)^\top, \left(\mathbf{z}_\alpha^{(2)}\right)^\top\right]$. Then $\mathbf{A}_{ij} = \sum_{\alpha=1}^{N-1} \mathbf{z}_\alpha^{(i)} \left(\mathbf{z}_\alpha^{(j)}\right)^\top$.

By Lemma 2, conditionally on $\mathbf{z}_1^{(2)}, \ldots, \mathbf{z}_{N-1}^{(2)}$, the random vectors $\mathbf{z}_1^{(1)}, \ldots, \mathbf{z}_{N-1}^{(1)}$ are independently distributed, with $\mathbf{z}_\alpha^{(1)} \sim \mathcal{N}\left(\mathbf{B}\mathbf{z}_\alpha^{(2)}, \mathbf{\Sigma}_{11.2}\right)$, where $\mathbf{B} = \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}$ and $\mathbf{\Sigma}_{11.2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$.

Now we apply Theorem 7 with $\mathbf{y}_\alpha = \mathbf{z}_\alpha^{(1)}$, $\mathbf{w}_\alpha = \mathbf{z}_\alpha^{(2)}$, $m = N-1$, $r = p-q$, $\mathbf{\Gamma} = \mathbf{B}$, $\mathbf{\Phi} = \mathbf{\Sigma}_{11.2}$, $\sum_{\alpha=1}^{m} \mathbf{y}_\alpha \mathbf{y}_\alpha^\top = \mathbf{A}_{11}$, $\mathbf{G} = \mathbf{A}_{12}\mathbf{A}_{22}^{-1}$, $\mathbf{H} = \mathbf{A}_{22}$, then the conditional distribution of

$$\mathbf{A}_{11.2} = \mathbf{A}_{11} - (\mathbf{A}_{12}\mathbf{A}_{22}^{-1})\mathbf{A}_{22}(\mathbf{A}_{12}\mathbf{A}_{22}^{-1})^\top$$

given $\mathbf{z}_1^{(2)}, \ldots, \mathbf{z}_{N-1}^{(2)}$ is distributed as $\sum_{\alpha=1}^{N-1-(p-q)} \mathbf{u}_\alpha \mathbf{u}_\alpha^\top$ and where $\mathbf{u}_1, \ldots, \mathbf{u}_{N-1-(p-q)}$ are independent, each with distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{11.2})$.

The Distribution of Partial Correlation Coefficient

Since the distribution of $\mathbf{A}_{11.2} = \sum_{\alpha=1}^{N-1-(p-q)} \mathbf{u}_\alpha \mathbf{u}_\alpha^\top$ does not depend on $\mathbf{z}_\alpha^{(2)}$, we obtain the following theorem:

Theorem 8

The matrix $\mathbf{A}_{11.2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}\mathbf{A}_{12}^\top$ is distributed as $\sum_{\alpha=1}^{N-1-(p-q)} \mathbf{u}_\alpha \mathbf{u}_\alpha^\top$, where $\mathbf{u}_1, \ldots, \mathbf{u}_{N-1}$ are independently distributed, each according to $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{11.2})$, and independently of $\mathbf{A}_{12}$ and $\mathbf{A}_{22}$.

Corollary 3

If $\mathbf{\Sigma}_{12} = \mathbf{0}$ (or $\mathbf{B} = \mathbf{0}$), the matrix $\mathbf{A}_{11.2}$ is distributed as $\sum_{\alpha=1}^{N-1-(p-q)} \mathbf{u}_\alpha \mathbf{u}_\alpha^\top$ and the matrix $\mathbf{A}_{12}\mathbf{A}_{22}\mathbf{A}_{12}^\top$ is distributed as $\sum_{\alpha=N-(p-q)}^{N-1} \mathbf{u}_\alpha \mathbf{u}_\alpha^\top$, where $\mathbf{u}_1, \ldots, \mathbf{u}_{N-1}$ are independently distributed, each according to $\mathcal{N}(0, \mathbf{\Phi})$.

The Distribution of Partial Correlation Coefficient

The distribution of $r_{ij.q+l,\dots,p}$ and the related tests of hypotheses based on N observations is the same as that of a simple correlation coefficient based on $N-(p-q)$ observations with a corresponding population correlation value of $r_{ij.q+l,\dots,p}$.