Chapter 9: Multiple Regression: Random $\boldsymbol{x}$'s

In the random-x case, $k+1$ variables $y, x_1, x_2, \ldots, x_k$ are measured on each of the n subjects, and we have

$$\mathrm{Cov}((y, x_1, \ldots, x_k)^T) = \Sigma,$$

where Σ is not a diagonal matrix.

1 Multivariate Normal Regression Model

Under the normality assumption that $(y, \boldsymbol{x}^T)^T$ is distributed as $N_{k+1}(\boldsymbol{\mu}, \Sigma)$ with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} \sigma_{yy} & \boldsymbol{\sigma}_{yx}^T \\ \boldsymbol{\sigma}_{yx} & \Sigma_{xx} \end{pmatrix}. \tag{1}$$

By the property of the multivariate Gaussian distribution, we have

$$E(y|\boldsymbol{x}) = \mu_y + \boldsymbol{\sigma}_{yx}^T \Sigma_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x) = \beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{x},$$

where $\beta_0 = \mu_y - \boldsymbol{\sigma}_{yx}^T \Sigma_{xx}^{-1} \boldsymbol{\mu}_x$ and

$$\boldsymbol{\beta}_1 = \Sigma_{xx}^{-1} \boldsymbol{\sigma}_{yx}. \tag{2}$$

We also have

$$\mathrm{Var}(y|\boldsymbol{x}) = \sigma_{yy} - \boldsymbol{\sigma}_{yx}^T \Sigma_{xx}^{-1} \boldsymbol{\sigma}_{yx} := \sigma^2.$$

2 Estimation and Testing in Multivariate Normal Regression

Theorem 2.1. *If $(y_1, \boldsymbol{x}_1^T)$, $(y_2, \boldsymbol{x}_2^T)$, $\ldots$, $(y_n, \boldsymbol{x}_n^T)$ is a random sample from $N_{k+1}(\boldsymbol{\mu}, \Sigma)$, the maximum likelihood estimators are*

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} \bar{y} \\ \bar{\boldsymbol{x}} \end{pmatrix},$$

$$\hat{\Sigma} = \frac{n-1}{n} \boldsymbol{S} = \frac{n-1}{n} \begin{pmatrix} s_{yy} & \boldsymbol{s}_{yx}^T \\ \boldsymbol{s}_{yx} & \boldsymbol{S}_{xx} \end{pmatrix}. \tag{3}$$

Theorem 2.2. *The MLE of a function of one or more parameters is the same function of the corresponding estimators; that is, if* $\hat{\boldsymbol{\theta}}$ *is the MLE of the vector or matrix of parameters* $\boldsymbol{\theta}$*, then* $g(\hat{\boldsymbol{\theta}})$ *is the MLE of* $g(\boldsymbol{\theta})$*.*

Example 2.1. *We illustrate the use of the invariance property in Theorem 2.2 by showing that the sample correlation matrix* $\boldsymbol{R}$ *is the MLE of the population correlation matrix* $\boldsymbol{P}_\rho$ *when sampling from the multivariate Gaussian distribution.*

Theorem 2.3. *If* $(y_1, \boldsymbol{x}_1^T)$*, ...,* $(y_n, \boldsymbol{x}_n^T)$ *is a random sample from* $N_{k+1}(\boldsymbol{\mu}, \Sigma)$*, the MLEs of* β_0*,* $\boldsymbol{\beta}_1$ *and* σ^2 *are given by*

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \boldsymbol{s}_{yx}^T \boldsymbol{S}_{xx}^{-1} \bar{\boldsymbol{x}}, \\ \hat{\boldsymbol{\beta}} - 1 &= \boldsymbol{S}_{xx}^{-1} \boldsymbol{s}_{yx}, \\ \hat{\sigma}^2 &= \frac{n-1}{n} s^2, \end{aligned}$$

where $s^2 = s_{yy} - \boldsymbol{s}_{yx}^T \boldsymbol{S}_{xx}^{-1} \boldsymbol{s}_{yx}$*.*

3 Standardized Regression Coefficients

We now show that the regression coefficient vector $\hat{\boldsymbol{\beta}}_1$ can be expressed in terms of sample correlations. The sample correlation matrix can be written in partitioned form as

$$\boldsymbol{R} = \begin{pmatrix} 1 & \boldsymbol{r}_{yx}^T \\ \boldsymbol{r}_{yx} & \boldsymbol{R}_{xx} \end{pmatrix}, \tag{4}$$

where r_{yx} is the vector of correlation between y and x's, and $\boldsymbol{R}_{xx}$ is the correlation matrix for the x's. In particular, we have

$$r_{y,x_j} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_{ij} - \bar{x}_j)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^b (x_{ij} - \bar{x}_j)^2}}.$$

$\boldsymbol{R}$ can be converted to $\boldsymbol{S}$ by

$$\boldsymbol{S} = \boldsymbol{D}\boldsymbol{R}\boldsymbol{D},$$

where $\boldsymbol{D} = [diag(S)]^{1/2} = \mathrm{diag}(s_y, \sqrt{s_{11}}, \ldots, \sqrt{s_{kk}})$. Therefore

$$\boldsymbol{S} = \begin{pmatrix} s_{yy} & \boldsymbol{s}_{yx}^T \\ \boldsymbol{s}_{yx} & \boldsymbol{S}_{xx} \end{pmatrix} = \begin{pmatrix} s_y^2 & s_y \boldsymbol{r}_{yx}^T \boldsymbol{D}_x \\ \boldsymbol{s}_y \boldsymbol{D}_x & \boldsymbol{D}_x \boldsymbol{R}_{xx} \boldsymbol{D}_x \end{pmatrix},$$

where $\boldsymbol{D}_x = diag(s_1, s_2, \ldots, s_k)$ and $s_i = \sqrt{s_{ii}}$ for $i = 1, 2, \ldots, k$. By the partition of $\boldsymbol{S}$, we have

$$\boldsymbol{S}_{xx} = \boldsymbol{D}_x \boldsymbol{R}_{xx} \boldsymbol{D}_x, \quad \boldsymbol{s}_{yx} = s_y \boldsymbol{D}_x \boldsymbol{r}_{yx}.$$

Therefore,

$$\hat{\boldsymbol{\beta}}_1 = \boldsymbol{S}_{xx}^{-1}\boldsymbol{s}_{yx} = s_y \boldsymbol{D}_x^{-1}\boldsymbol{R}_{xx}^{-1}\boldsymbol{r}_{yx}.$$

We illustrate the formula for $k = 2$. Consider the centered the model:

$$\hat{y}_i = \bar{y} + \hat{\beta}_1(x_{i1} - \bar{x}_1) + \hat{\beta}_2(x_{i2} - \bar{x}_2),$$

which can be expressed in terms of standardized variables as

$$\frac{\hat{y}_i - \bar{y}}{s_y} = \frac{s_1}{s_y}\hat{\beta}_1\left(\frac{x_{i1} - \bar{x}_1}{s_1}\right) + \frac{s_2}{s_y}\hat{\beta}_2\left(\frac{x_{i2} - \bar{x}_2}{s_2}\right).$$

We this define the standardized coefficient as

$$\hat{\beta}_j^* = \frac{s_1}{s_y}\hat{\beta}_j.$$

In the matrix-vector form, we have

$$\hat{\boldsymbol{\beta}}_1^* = \frac{1}{s_y}\boldsymbol{D}_x\hat{\boldsymbol{\beta}}_1,$$

which can be further written as

$$\hat{\boldsymbol{\beta}}_1^* = \boldsymbol{R}_{xx}^{-1}\boldsymbol{r}_{yx}.$$

4 R^2 in Multivariate Normal Regression

The population multiple correlation coefficient $\rho_{y|\boldsymbol{x}}$ is defined as the correlation between y and the linear function $w = \mu_y + \boldsymbol{\sigma}_{yx}^T \Sigma_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x)$:

$$\rho_{y|\boldsymbol{x}} = corr(y, w) = \frac{\sigma_{yw}}{\sigma_y \sigma_w},$$

where w is equal to $E(y|\boldsymbol{x})$. As $\boldsymbol{x}$ varies randomly, w becomes a random variable.

It is easy to establish that

$$\text{Cov}(y, w) = \text{Var}(w) = \boldsymbol{\sigma}_{yx}^T \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx},$$

by the fact that $y = w + e$, and e denotes a random error vector.

Then the population multiple correlation $\rho_{y|\boldsymbol{x}}$ becomes

$$\rho_{y|\boldsymbol{x}} = \frac{\text{Cov}(y, w)}{\text{Var}(y)\text{Var}(w)} = \sqrt{\frac{\boldsymbol{\sigma}'_{yx}\Sigma_{xx}^{-1}\boldsymbol{\sigma}_{yx}}{\boldsymbol{\sigma}_{yx}}}.$$

The *population coefficient of determination* or population squared multiple correlation $\rho_{y|\boldsymbol{x}}^2$ is given by

$$\rho_{y|\boldsymbol{x}}^2 = \frac{\boldsymbol{\sigma}'_{yx}\Sigma_{xx}^{-1}\boldsymbol{\sigma}_{yx}}{\boldsymbol{\sigma}_{yx}}.$$

In what follows, we list some properties of $\rho_{y|\boldsymbol{x}}$ and $\rho_{y|\boldsymbol{x}}^2$:

1. $\rho_{y|\boldsymbol{x}}$ is the maximum correlation between y and any linear function of $\boldsymbol{x}$; that is, $\rho_{y|\boldsymbol{x}} = \max_{\boldsymbol{a}} \rho_{y,\boldsymbol{a}'\boldsymbol{x}}$.

2. $\rho^2_{y|\boldsymbol{x}}$ can be expressed in terms of determinants:

$$\rho^2_{y|\boldsymbol{x}} = 1 - \frac{|\Sigma|}{\sigma_{yy}|\Sigma_{xx}|},$$

   where Σ and Σ_{xx} are as defined in (**??**).

3. $\rho^2_{y|\boldsymbol{x}}$ is invariant to linear transformations on y or on the x's; that is, if $u = ay$ and $\boldsymbol{v} = \boldsymbol{Bx}$, where $\boldsymbol{B}$ is nonsingular, then $\rho^2_{u|\boldsymbol{v}} = \rho^2_{y|\boldsymbol{x}}$.

4. Using $\text{Var}(w) = \boldsymbol{\sigma}'_{yx}\boldsymbol{\sigma}^{-1}_{xx}\boldsymbol{\sigma}_{yx}$, $\rho^2_{y|\boldsymbol{x}}$ can be written in the form

$$\rho^2_{y|\boldsymbol{x}} = \frac{\text{Var}(w)}{\text{Var}(y)}.$$

5. $\text{Var}(y|\boldsymbol{x})$ can be expressed in terms of $\rho^2_{y|\boldsymbol{x}}$:

$$\text{Var}(y|\boldsymbol{x}) = \sigma_{yy} - \boldsymbol{\sigma}^T_{yx}\Sigma^{-1}_{xx}\boldsymbol{\sigma}_{yx} = \sigma_{yy} - \sigma_{yy}\rho^2_{y|\boldsymbol{x}} = \sigma_{yy}(1 - \rho^2_{y|\boldsymbol{x}}).$$

6. If we consider $y - w$ as a residual or error term, then $y - w$ is uncorrelated with the x's,

$$\text{Cov}(y - w, x) = 0.$$

We can obtain a MLE for $\rho^2_{y|\boldsymbol{x}}$, which is

$$R^2 = \frac{\boldsymbol{s}'_{yx}S^{-1}_{xx}\boldsymbol{s}_{yx}}{\boldsymbol{s}_{yy}}.$$

R is called the sample multiple correlation coefficient.

We now list several properties of R and R^2, some of which are analogous to properties of $\rho^2_{y|\boldsymbol{x}}$ above.

1. R is equal to the correlation between y and $\hat{y}$.

2. R is equal to the maximum correlation between $\boldsymbol{y}$ and ant linear combination $\boldsymbol{a}'\boldsymbol{x}$:

$$R = \max_{\boldsymbol{a}} r_{y,\boldsymbol{a}'\boldsymbol{x}}.$$

3. R^2 can be expressed in terms of correlations:

$$R^2 = \boldsymbol{r}^T_{yx}\boldsymbol{R}^{-1}_{xx}\boldsymbol{r}_{yx},$$

   where $\boldsymbol{r}_{yx}$ and $\boldsymbol{R}_{xx}$ are from the samle correlation matrix $\boldsymbol{R}$ partitioned as in (4).

4. R^2 can be obtained from $\boldsymbol{R}^{-1}$:

$$R^2 = 1 - \frac{1}{r^{yy}},$$

where r^{yy} is the first diagonal element of $\boldsymbol{R}^{-1}$.

5. R^2 can be expressed in terms of determinants:

$$R^2 = 1 - \frac{|\boldsymbol{S}|}{s_{yy}|\boldsymbol{S}_{xx}|} = 1 - \frac{|\boldsymbol{R}|}{|\boldsymbol{R}_{xx}|},$$

where $\boldsymbol{S}_{xx}$ are $\boldsymbol{R}_{xx}$ are defined in (3) and (4).

6. If $\rho^2_{y|\boldsymbol{x}} = 0$, the expected value of R^2 is given by

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

R^2 is biased when $\rho^2_{y|\boldsymbol{x}} = 0$.

7. $R^2 \geq \max_j r^2_{yj}$, where r_{yj} is an element of $\boldsymbol{r}_{yx} = (r_{y1}, r_{y2}, \ldots, r_{yk})^T$.

8. R^2 is invariant to full linear transformations on y or on the x's.

5 Tests and Confidence Intervals for R^2

Note that $\rho^2_{y|\boldsymbol{x}} = 0$ becomes

$$\rho^2_{y|\boldsymbol{x}} = \frac{\boldsymbol{\sigma}^T_{yx}\Sigma^{-1}_{xx}\boldsymbol{\sigma}_{yx}}{\sigma_{yy}} = 0,$$

which leads to $\boldsymbol{\sigma}_{yx} = 0$ since Σ_{xx} is positive. Further, by (2), $\boldsymbol{\beta}_1 = \Sigma^{-1}_{xx}\boldsymbol{\sigma}_{yx}$, $H_0 : \rho^2_{y|\boldsymbol{x}} = 0$ is equivalent to $H_0 : \boldsymbol{\beta}_1 = 0$.

The F statistic for fixed x's is given by

$$\begin{aligned} F &= \frac{(\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - n\bar{y}^2)/k}{(\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y})/(n-k-1)} \\ &= \frac{R^2/k}{(1-R^2)/(n-k-1)}. \end{aligned} \tag{5}$$

The test statistic in (5) can be obtained by the likelihood ratio approach in the case of random x's.

Theorem 5.1. *If $(y_1, \boldsymbol{x}'_1)$, $(y_2, \boldsymbol{x}'_2)$, ..., $(y_n, \boldsymbol{x}'_n)$ is a random sample from $N_{k+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the likelihood ratio test for $H_0 : \boldsymbol{\beta}_1 = 0$ or equivalently $H_0 : \rho^2_{y|\boldsymbol{x}} = 0$ can be based on F in (5). We reject H_0 if $F > F_{\alpha,k,n-k-1}$.*

When $k = 1$, F in (5) reduces to $F = (n-2)r^2/(1-r^2)$. Hence,

$$t = \frac{\sqrt{n-2}r}{\sqrt{1-r^2}}$$

has a t-distribution with $n-2$ degrees of freedom when (y, x) has a bivariate normal distribution with $\rho = 0$.

If (y, x) is bivariate normal and $\rho \neq 0$, then $\text{Var}(r) = (1-\rho^2)^2/n$ and the function

$$u = \frac{\sqrt{n}(r-\rho)}{1-\rho^2},$$

is approximately standard normal for large n. However, the distribution of u approaches normality very slowly as n increases. Fisher (1921) found a function of r

$$z = \frac{1}{2}\log\frac{1+r}{1-r} = \tanh^{-1}(r),$$

approaches normality much faster than does u. The approximated mean and variance of z are

$$E(z) \approx \frac{1}{2}\log\frac{1+\rho}{1-\rho},$$

$$\text{Var}(z) \approx \frac{1}{n-3}.$$

We can use Fisher's transformation to test the hypotheses such as $H_0 : \rho = \rho_0$ vs. $H_1 : \rho \neq \rho_0$, we calculate

$$v = \frac{z - \tanh^{-1}(\rho_0)}{\sqrt{1/(n-3)}},$$

which is approximately distributed as the standard normal $N(0, 1)$.

6 Sample Partial Correlations

The population partial correlation $\rho_{ij\cdot rs\cdots q}$ is the correlation between y_i and y_j in the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$, where y_i and y_j are in $\boldsymbol{y}$ and the subcripts $r, s, \ldots, q$ represent all the variables in $\boldsymbol{x}$:

$$\rho_{ij\cdot rs\cdots q} = \frac{\sigma_{ij\cdot rs\cdots q}}{\sqrt{\sigma_{ii\cdot rs\cdots q}\sigma_{jj\cdot rs\cdots q}}},$$

where $\sigma_{ij\cdot rs\cdots q}$ is the (ij)th element of $\boldsymbol{\Sigma}_{y\cdot x} = \text{Cov}(y|x)$.

To simplify exposition, we illustrate with $r_{12\cdot 3}$. The sample partial correlation of y_1 and y_2 with y_3 held fixed is usually given as

$$r_{12\cdot 3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}},$$

where r_{12}, r_{13} and r_{23} are the ordinary correlations between y_1 and y_2, y_1 and y_3, y_2 and y_3, respectively.

Theorem 6.1. *The expression for $r_{12\cdot 3}$ is equal to $r_{y_1-\hat{y}_1, y_2-\hat{y}_2}$, where $y_1 - \hat{y}_1$ and $y_2 - \hat{y}_2$ are residuals from regression of y_1 on y_3 and y_2 on y_3.*