Means and covariances

If A is an $p \times n$ matrix of constants and $\mathbf{Y}$ is a $n \times 1$ random vector, then

$$\mathrm{E}(\mathrm{A}\mathbf{Y}) = A\,\mathrm{E}(\mathbf{Y}).$$

If B is a $q \times n$ matrix of constants, then

$$\mathrm{Cov}(\mathrm{A}\mathbf{Y}, \mathrm{B}\mathbf{Y}) = \mathrm{A}\,\mathrm{Cov}(\mathbf{Y})\mathrm{B}^\top.$$

Multivariate normal

The density of a multivariate normal vector $\mathbf{Y} \sim \mathrm{MVN}_n(\boldsymbol{\mu}, \Sigma)$ is

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right),$$

The moment generating function is

$$M_{\mathbf{Y}}(\mathbf{t}) = \mathrm{E}\left[e^{\mathbf{t}^\top \mathbf{Y}}\right] = \exp\left(\mathbf{t}^\top \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}\right).$$

Any $r \times 1$ sub-vector of $\mathbf{Y}$ has an r-variate normal distribution with mean vector and covariance matrix equal to the corresponding subvector of $\boldsymbol{\mu}$ and submatrix of Σ.

If $\mathbf{Y}$ and $\mathbf{X}$ are jointly multivariate normal, then the conditional distribution of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ is multivariate normal with

$$\mathrm{E}(\mathbf{Y} \mid \mathbf{x}) = \boldsymbol{\mu}_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x),$$

$$\mathrm{Cov}(\mathbf{Y} \mid \mathbf{x}) = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}.$$

Mahalanobis transform: $\Sigma^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}) \sim \mathrm{MVN}(\mathbf{0}, \mathrm{I})$.

χ^2, F and t distributions

The density of

$$V = \sum_{i=1}^{n} Y_i^2 = \mathbf{Y}^\top \mathbf{Y},$$

where the Y_i are independently distributed as $\mathrm{N}(\mu_i, 1)$, $\mathbf{Y} \sim \mathrm{MVN}_n(\boldsymbol{\mu}, \mathrm{I})$, is called the noncentral chi-square distribution and is denoted by $\chi^2(n, \lambda)$. The noncentrality parameter λ is defined as

$$\lambda = \frac{1}{2}\sum_{i=1}^{n} \mu_i^2 = \frac{1}{2}\,\boldsymbol{\mu}^\top \boldsymbol{\mu}.$$

If $U \sim \chi^2(p, \lambda)$, $V \sim \chi^2(q)$, and U and V are independent, then $Z = \frac{U/p}{V/q}$ is distributed as $\mathrm{F}(p, q, \lambda)$.

If $Y \sim \mathrm{N}(\mu, 1)$, $U \sim \chi^2(p)$, and Y and U are independent, then $T = \frac{Y}{\sqrt{U/p}}$ is distributed as $\mathrm{t}(p, \mu)$.

Quadratic forms

Let A be an $n \times n$ symmetric matrix of constants of rank r, let B be a $k \times n$ matrix of constants and let $\mathbf{Y} \sim \mathrm{MVN}_n(\boldsymbol{\mu}, \Sigma)$ unless otherwise specified.

Then

$$\mathrm{E}(\mathbf{Y}^\top \mathrm{A} \mathbf{Y}) = \mathrm{tr}(\mathrm{A}\Sigma) + \boldsymbol{\mu}^\top \mathrm{A} \boldsymbol{\mu}$$

and

$$\mathrm{Var}(\mathbf{Y}^\top \mathrm{A} \mathbf{Y}) = 2 \ \mathrm{tr}\big((\mathrm{A}\Sigma)^2\big) + 4 \, \boldsymbol{\mu}^\top \mathrm{A} \Sigma \mathrm{A} \boldsymbol{\mu}.$$

Further,

$$\mathbf{Y}^\top \mathrm{A} \mathbf{Y} \sim \chi^2(r, \lambda),$$

with $\lambda = \frac{1}{2} \, \boldsymbol{\mu}^\top \mathrm{A} \boldsymbol{\mu}$, if and only if $\mathrm{A}\Sigma$ is idempotent.

If $\mathbf{Y} \sim \mathrm{MVN}(\mathbf{0}, \mathrm{I})$, then $\mathbf{Y}^\top \mathrm{A} \mathbf{Y} \sim \chi^2(r)$ if and only if A is idempotent.

If $\mathbf{Y} \sim \mathrm{MVN}(\boldsymbol{\mu}, \sigma^2 \mathrm{I})$, then $\frac{\mathbf{Y}^\top \mathrm{A} \mathbf{Y}}{\sigma^2} \sim \chi^2\Big(r, \frac{\boldsymbol{\mu}^\top \mathrm{A} \boldsymbol{\mu}}{2\sigma^2}\Big)$ if and only if A is idempotent.

Independence:

$\mathrm{B}\mathbf{Y}$ and $\mathbf{Y}^\top \mathrm{A} \mathbf{Y}$ are independent if and only if $\mathrm{B}\Sigma\mathrm{A} = \mathbf{0}$.

Let A and B be symmetric matrices of constants. Then $\mathbf{Y}^\top \mathrm{A} \mathbf{Y}$ and $\mathbf{Y}^\top \mathrm{B} \mathbf{Y}$ are independent if and only if $\mathrm{A}\Sigma\mathrm{B} = \mathbf{0}$.

Multiple linear regression

Model: $\mathbf{Y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with the design matrix X $n \times (k+1)$ and full rank.

Assumptions: $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathrm{I}$, and $\boldsymbol{\varepsilon} \sim \mathrm{MVN}(\mathbf{0}, \sigma^2 \mathrm{I})$.

Estimators: $\hat{\boldsymbol{\beta}} = (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \mathbf{Y}$ and $s^2 = \frac{1}{n-k-1} (\mathbf{Y} - \mathrm{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathrm{X}\hat{\boldsymbol{\beta}})$, with $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}) = s^2 (\mathrm{X}^\top \mathrm{X})^{-1}$.

Sums of squares: $\text{SST} = \text{SSR} + \text{SSE}$, where

$$\text{SSE} = \mathbf{Y}^\top \left(\text{I} - \text{X}(\text{X}^\top\text{X})^{-1}\text{X}^\top\right)\mathbf{Y} = \mathbf{Y}^\top(\text{I} - \text{H})\mathbf{Y},$$

$$\text{SST} = \mathbf{Y}^\top \left(\text{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top\right)\mathbf{Y} = \mathbf{Y}^\top \left(\text{I} - \frac{1}{n}\text{J}\right)\mathbf{Y},$$

and

$$\text{SSR} = \mathbf{Y}^\top \left(\text{H} - \frac{1}{n}\text{J}\right)\mathbf{Y}.$$

Coefficient of determination: $R^2 = \frac{\text{SSR}}{\text{SST}}$, and $R^2_{\text{a}} = \frac{n-1}{n-k-1}\left(R^2 - \frac{k}{n-1}\right)$

Generalized least squares: $\mathbf{Cov}(\mathbf{Y}) = \sigma^2\text{V}$

Estimators: $\hat{\boldsymbol{\beta}} = (\text{X}^\top\text{V}^{-1}\text{X})^{-1}\text{X}^\top\text{V}^{-1}\mathbf{Y}$, and $s^2 = \frac{(\mathbf{Y}-\text{X}\hat{\boldsymbol{\beta}})^\top\text{V}^{-1}(\mathbf{Y}-\text{X}\hat{\boldsymbol{\beta}})}{n-k-1}$.

Tests of hypotheses

F test overall regression:

$$F = \frac{\text{SSR}/k}{\text{SSE}/(n-k-1)}.$$

Partial F test: (* denotes reduced model)

$$F = \frac{(\text{SSR} - \text{SSR}^*)/q}{\text{SSE}/(n-k-1)}.$$

General linear hypothesis F test:

$$H_0: \ \text{C}\boldsymbol{\beta} = \mathbf{0} \quad \text{vs} \quad H_1: \ \text{C}\boldsymbol{\beta} \neq \mathbf{0},$$

$$F = \frac{(\text{C}\hat{\boldsymbol{\beta}})^\top\left[\text{C}(\text{X}^\top\text{X})^{-1}\text{C}^\top\right]^{-1}(\text{C}\hat{\boldsymbol{\beta}})/q}{\text{SSE}/(n-k-1)}.$$

T test:

$$T = \frac{\hat{\beta}_j}{\sqrt{s^2\left[(\text{X}^\top\text{X})^{-1}\right]_{jj}}}$$

Prediction interval for a new observation

$$\hat{y}_0 \pm t(\alpha/2, n-k-1)\, s\sqrt{1 + \mathbf{x}_0^\top(\text{X}^\top\text{X})^{-1}\mathbf{x}_0}.$$

Model check

Residuals: $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathrm{X}\hat{\boldsymbol{\beta}} = \mathbf{Y} - \hat{\mathbf{Y}}$, $\hat{\boldsymbol{\varepsilon}} \sim \mathrm{MVN}_n(\mathbf{0}, \sigma^2(\mathrm{I} - \mathrm{H}))$

Studentized residuals: $r_i = \frac{\hat{\varepsilon}_i}{s\sqrt{1-h_{ii}}}$.

Cook's distance: $D_i = \frac{r_i^2}{k+1}\frac{h_{ii}}{1-h_{ii}}$.

Variance inflation factor: $\mathrm{VIF}_j = \frac{1}{1-R_j^2}$, where R_j^2 is the coefficient of determination from the regression of the j-th predictor on all other predictors.

Linear mixed models

Random intercept:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{0i} + \varepsilon_{ij}, \quad \gamma_{0i} \sim \mathrm{N}(0, \tau_0^2), \quad \varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2),$$

Random intercept and slope:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{0i} + \gamma_{1i} x_{ij} + \varepsilon_{ij},$$

$$\begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix} \sim \mathrm{MVN}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{pmatrix} \right), \quad \varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2).$$

ANOVA

One-way:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \ldots, m,\ j = 1, \ldots, n_i, \quad \sum_i \alpha_i = 0.$$

Two-way:

$$Y_{ijk} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \varepsilon_{ijk}, \quad i = 1, \ldots, m_\alpha,\ j = 1, \ldots, m_\gamma,\ k = 1, \ldots, n_{ij},$$

$$\sum_i \alpha_i = 0, \quad \sum_j \gamma_j = 0, \quad \sum_i (\alpha\gamma)_{ij} = 0, \quad \sum_j (\alpha\gamma)_{ij} = 0.$$