

1 Multivariate normal

a) Define $\mathbf{Y} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \mathbf{X}$ with $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$

Then, $E\mathbf{Y} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ k\mu \end{bmatrix} = \begin{bmatrix} (1+k)\mu \\ (1-k)\mu \end{bmatrix}$ and using, $\text{Cov}(\mathbf{A}\mathbf{X}) = \mathbf{A}\text{Cov}(\mathbf{X})\mathbf{A}^T$,

$$\text{Cov}(\mathbf{Y}) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{Cov}(\mathbf{X}) \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^T = \begin{bmatrix} 2(\sigma^2 + \rho) & 0 \\ 0 & 2(\sigma^2 - \rho) \end{bmatrix}.$$

Yes they are independent since a covariance of zero between normally distributed random variables implies independence.

b) $\mu = 1, k = 2, \sigma^2 = 2^2, \rho = 2$

Using $X_1 \sim N(1, 2^2)$, $P(X_1 \leq 2) = P(Z \leq 0.5) = 0.6915$.

We have $X_1|X_2 = 2 \sim N(1, 3)$, so $P(X_1 \leq 2|X_2 = 2) = P(Z \leq 1/\sqrt{3}) \approx \Phi(0.58) = 0.7190$.

The distribution of $2X_1 - 3X_2$ is normal with mean $2 - 6 = -4$ and variance $16 + 36 - 24 = 28$. Then, $P(2X_1 - 3X_2 \geq 0) = 1 - P(Z \leq \frac{4}{\sqrt{28}}) \approx 1 - \Phi(0.76) = 0.2236$.

2 Sums of squares

a) The sum is univariate normal with mean $n\mu$ and variance $n\sigma^2$.

First $(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y} = \mathbf{Y} - \frac{1}{n}\mathbf{J}\mathbf{Y}$. Here $\frac{1}{n}\mathbf{J}\mathbf{Y}$ is an n -vector with $\bar{Y}$ in all entries, so that for some row i of the n -vector $(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y}$ we have $Y_i - \bar{Y}$. Furthermore, since $(\mathbf{I} - \frac{1}{n}\mathbf{J})$ is symmetric and idempotent,

$$\mathbf{Y}^T(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y} = \mathbf{Y}^T(\mathbf{I} - \frac{1}{n}\mathbf{J})^T(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y} = \left[(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y} \right]^T (\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y}$$

and we can now observe that the expression is the sum of the squares $(Y_i - \bar{Y})^2$ for $i = 1$ to n .

From the course we know that if $\mathbf{X}$ is an n -vector of identically distributed and independent normal variables with mean $\boldsymbol{\mu}$ variance $\sigma^2\mathbf{I}$ and $\mathbf{A}$ is symmetric and idempotent, then

$$(\mathbf{X} - \boldsymbol{\mu})^T \mathbf{A}(\mathbf{X} - \boldsymbol{\mu}) / \sigma^2$$

is χ^2 -distributed with degrees of freedom equal to the rank of $\mathbf{A}$.

Consider $(\mathbf{I} - \frac{1}{n}\mathbf{J})(\mathbf{Y} - \boldsymbol{\mu})$ with $\boldsymbol{\mu}$ an n -vector with all entries equal to μ . The average of all entries in $\boldsymbol{\mu}$ is also equal to μ . We then have $(\mathbf{I} - \frac{1}{n}\mathbf{J})\boldsymbol{\mu} = \mathbf{0}$. Thus,

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 / \sigma^2 = \mathbf{Y}^T(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y} / \sigma^2 = (\mathbf{Y} - \boldsymbol{\mu})^T (\mathbf{I} - \frac{1}{n}\mathbf{J})(\mathbf{Y} - \boldsymbol{\mu}) / \sigma^2$$

with $(\mathbf{I} - \frac{1}{n}\mathbf{J})$ symmetric and idempotent so the χ^2 -distribution follows. Furthermore, the rank of $(\mathbf{I} - \frac{1}{n}\mathbf{J})$ is equal to the trace of $\mathbf{I}$ minus the trace of $\frac{1}{n}\mathbf{J}$ which is $n - n/n = n - 1$ so the degrees of freedom of the χ^2 -distribution is $n - 1$.

b) From the course we know that if $\mathbf{X}$ is an n -vector of identically distributed and independent normal variables with mean $\boldsymbol{\mu}$ variance $\sigma^2\mathbf{I}$ and $\mathbf{A}$ and $\mathbf{B}$ are symmetric and idempotent, then $\mathbf{AB} = \mathbf{0}$ implies independence of $(\mathbf{X} - \boldsymbol{\mu})^T \mathbf{A}(\mathbf{X} - \boldsymbol{\mu})$ and $(\mathbf{X} - \boldsymbol{\mu})^T \mathbf{B}(\mathbf{X} - \boldsymbol{\mu})$.

From a) we know that the first expression can be written in matrix format as

$$(\mathbf{Y} - \boldsymbol{\mu})^T (\mathbf{I} - \frac{1}{n}\mathbf{J})(\mathbf{Y} - \boldsymbol{\mu})$$

The second expression can when $\mu = 0$ be expressed as

$$(\mathbf{Y} - \boldsymbol{\mu})^T \frac{1}{n}\mathbf{J}(\mathbf{Y} - \boldsymbol{\mu}).$$

We compute

$$(\mathbf{I} - \frac{1}{n}\mathbf{J})\frac{1}{n}\mathbf{J} = \frac{1}{n}\mathbf{J} - \frac{1}{n^2}n\mathbf{J} = \mathbf{0}$$

which implies independence.

3 One-factor ANOVA

a) The parameter μ is the ‘global mean’; the expected test score if we randomly draw a student from the entire group (all regions), the parameter α_l represents the ‘effect’ (deviation from global mean) of a student being from region l , and σ^2 is the variability in test scores among students from the same region (assumed equal for all regions).

$$E(Y_{kl}) = \mu + \alpha_l$$

$$\text{Covar}(Y_{kl}, Y_{k'l}) = \text{Covar}(\varepsilon_{kl}, \varepsilon_{k'l}) = 0$$

b)

$$E(Y_{klm}) = \mu + \alpha_l$$

$$\text{Covar}(Y_{klm}, Y_{klm'}) = \text{Covar}(u_{kl} + \varepsilon_{klm}, u_{kl} + \varepsilon_{klm'}) = \text{Var}(u_{kl}) = \tau^2$$

$$\text{Covar}(Y_{klm}, Y_{k'lm}) = \text{Covar}(u_{kl} + \varepsilon_{klm}, u_{k'l} + \varepsilon_{k'lm}) = 0$$

4 Linear regression

a) We know that $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$. Since the columns of $\mathbf{X}$ are orthogonal, we have

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \mathbf{x}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{x}_2^T \mathbf{x}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{x}_p^T \mathbf{x}_p \end{bmatrix},$$

i.e. a $p \times p$ diagonal matrix with $\mathbf{x}_j^T \mathbf{x}_j$ along the diagonal for $j = 1, \dots, p$. The inverse is also diagonal with $1/\mathbf{x}_j^T \mathbf{x}_j$ along the diagonal and the j -th row of $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ is $(1/\mathbf{x}_j^T \mathbf{x}_j) \mathbf{x}_j^T$. Then $\hat{\beta}_j = (1/\mathbf{x}_j^T \mathbf{x}_j) \mathbf{x}_j^T \mathbf{Y}$.

b) The design matrix has 4 rows and 4 columns; $\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

and we let the parameter vector be $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_{12})^T$.

The first component β_0 is the model intercept and will here be the overall mean of Y , i.e. the expected measurement value across all experiments. The component β_1 is 1/2 the main effect of factor 1 (i.e. $2\beta_1$ represents the main effect of changing factor 1 from low to high level), and similarly for β_2 for factor 2. The final component β_{12} is 1/2 the interaction effect of factors 1 and 2. Calculating the expected difference in measurements of changing factor 1 from low to high level we have $2\beta_1 + 2x_2\beta_{12}$. If for example factor 2 is at low level ($x_2 = -1$) then we subtract $2\beta_{12}$ to the main effect of the first factor to get the expected difference in measurements.

Orthogonal columns:

$$\mathbf{x}_1^T \mathbf{x}_2 = -1 + 1 - 1 + 1 = 0,$$

and similarly for $\mathbf{x}_1^T \mathbf{x}_3, \mathbf{x}_1^T \mathbf{x}_4$. Also,

$$\mathbf{x}_2^T \mathbf{x}_3 = (-1) \cdot (-1) + 1 \cdot (-1) + (-1) \cdot 1 + 1 \cdot 1 = 0$$

$$\mathbf{x}_2^T \mathbf{x}_4 = (-1) \cdot 1 + 1 \cdot (-1) + (-1) \cdot (-1) + 1 \cdot 1 = 0$$

$$\mathbf{x}_3^T \mathbf{x}_4 = (-1) \cdot 1 + (-1) \cdot (-1) + 1 \cdot (-1) + 1 \cdot 1 = 0$$

Estimates:

$$\hat{\beta}_0 = \bar{y} = 30.3$$

$$\hat{\beta}_1 = \frac{1}{2} \left(\frac{1}{2}(32.3 + 32.7) - \frac{1}{2}(29.5 + 26.7) \right) = 2.2$$

$$\hat{\beta}_2 = \frac{1}{2} \left(\frac{1}{2}(26.7 + 32.7) - \frac{1}{2}(29.5 + 32.3) \right) = -0.8$$

$$\hat{\beta}_{12} = \frac{1}{2} \left(\frac{1}{2}(29.5 + 32.7) - \frac{1}{2}(32.3 + 26.7) \right) = 0.6$$

c) Hypothesis: $H_0 : \beta_1 = 0, H_1 : \beta_1 \neq 0$. Test statistic $T = \frac{\hat{\beta}_1}{\text{SE}(\hat{\beta}_1)}$. The realization of T is therefore $t = 2.0250/0.6726 = 3.010705 \approx 3.011$. The distribution of T if H_0 is true is a t -distribution with $8 - 2 = 6$ degrees of freedom and the p -value is found by

$$P(T > 3.011) + P(T < -3.011) = 2 \cdot P(T < -3.011) \approx 2 \cdot P(T < -3) = 2 \cdot 0.012 = 0.024$$

In general, the F-test is constructed for the null hypothesis that all coefficients (except the intercept) of a linear regression model are zero, against the alternative hypothesis that at least one coefficient differs from zero. With only one covariate (x_1) in the model, the F-test will be a test of $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$. Here $k = 1$. The test statistic can be rewritten as

$$F = \frac{SSR/k}{SSE/(n-k-1)} = \frac{SSR}{\hat{\sigma}^2}$$

where

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} - \bar{y})^2 = \sum_{i=1}^n (\bar{y} - \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_1 x_{1i} - \bar{y})^2 = \hat{\beta}_1^2 \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2.$$

Therefore,

$$F = \frac{SSR}{\hat{\sigma}^2} = \frac{\hat{\beta}_1^2}{\hat{\sigma}^2 / \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2}.$$

We also have

$$T = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 c_{11}}}$$

where c_{11} is the 2nd diagonal element of $(\mathbf{X}^T \mathbf{X})^{-1}$. With only one covariate,

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} n & n\bar{x}_1 \\ n\bar{x}_1 & \sum_i x_{1i}^2 \end{bmatrix}$$

and the 2nd diagonal element of $(\mathbf{X}^T \mathbf{X})^{-1}$ is

$$\frac{n}{n \sum_i x_{1i}^2 - n^2 \bar{x}_1^2} = \frac{1}{\sum_i (x_{1i} - \bar{x}_1)^2}$$

It follows that $T = F^2$ for this problem.

d) The local significance level is $0.1/3 \approx 0.033$ so the two main effects are considered significant but not the interaction.

The estimated standard error of a coefficient j is in general

$$\hat{\sigma} \sqrt{c_{jj}}$$

where c_{jj} is the $j+1$ th diagonal element of $(\mathbf{X}^T \mathbf{X})^{-1}$. For the type of design matrix considered here, with $\mathbf{x}_j^T \mathbf{x}_j = n$ and $\mathbf{x}_j^T \mathbf{x}_{j'} = 0$, we have

$$(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{n} \mathbf{I}$$

Therefore, the estimated standard error of all coefficient estimates are

$$\hat{\sigma} / \sqrt{n} = 1.0235 / \sqrt{8} \approx 0.3619.$$