Statistics 203: Introduction to Regression and Analysis of Variance

Multiple Linear Regression + Multivariate Normal

Jonathan Taylor

Today

- Multiple linear regression
- Some proofs: multivariate normal distribution.

Multiple linear regression

- Specifying the model.
- Fitting the model: least squares.
- Interpretation of the coefficients.

Model

- Basically, rather than one predictor, we more than one predictor, say $p-1$.
- $$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$
- Errors $(\varepsilon_i)_{1 \leq i \leq n}$ are assumed independent $N(0, \sigma^2)$, as in simple linear regression.
- Coefficients are called (partial) regression coefficients because they “allow” for the (partial) effect of other variables.

Design matrix

- Define the $n \times p$ matrix

$$X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1,p-1} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & X_{n1} & X_{n2} & \dots & X_{n,p-1} \end{pmatrix}$$

  and the column vectors $X_j = (X_{1j}, \dots, X_{nj})$.
- Model can be expressed as

$$Y = X\beta + \varepsilon.$$

Fitting the model: SSE

- Just as in simple linear regression, model is fit by minimizing

$$SSE(\beta_0, \dots, \beta_p) = \sum_{i=1}^{n} (Y_i - (\beta_0 + \sum_{j=1}^{p} \beta_j X_{ij}))^2.$$

- Minimizers: $\widehat{\beta} = (\widehat{\beta}_0, \dots, \widehat{\beta}_p)$ are the "least squares estimates" and are also normally distributed as in simple linear regression.
- Explicit expression when X is full rank (next slide)

$$\widehat{\beta} = (X^t X)^{-1} X^t Y.$$

Solving for $\widehat{\beta}$

- Normal equations

$$\frac{\partial}{\partial \beta_j} SSE \Big|_{\widehat{\beta}} = -2 \left(Y - X\widehat{\beta}\right)^t X_j = 0, \qquad 0 \leq j \leq p-1.$$

- Equivalent to

$$\begin{aligned} (Y - X\widehat{\beta})^t X &= 0 \\ Y^t X &= \widehat{\beta}^t (X^t X) \\ X^t Y &= (X^t X)\widehat{\beta} \\ \widehat{\beta} &= (X^t X)^{-1} X^t Y \end{aligned}$$

- Properties: after some facts about multivariate normal random vectors.

Multivariate normal

- $Z = (Z_1, \dots, Z_n) \in \mathbb{R}^n$ is *multivariate Gaussian* if, for every $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$, $\langle \alpha, Z \rangle = \sum_{i=1}^n \alpha_i Z_i$ is Gaussian.
- Mean vector: $\mu \in \mathbb{R}^n$ has components

$$\mu_i = \mathbb{E}(Z_i).$$

- Covariance matrix: Σ a non-negative definite $n \times n$ matrix

$$\Sigma_{ij} = \mathrm{Cov}(Z_i, Z_j).$$

- Non-negative (positive) definite: for any $\alpha \in \mathbb{R}^n$

$$\alpha^t \Sigma \alpha \geq (>) 0.$$

- We write $Z \sim N(\mu, \Sigma)$.

Multivariate normal

- For any $m \times n$ matrix A

$$AZ \sim N(A\mu, A\Sigma A^t).$$

- If Σ is positive definite then the density of Z is

$$f_Z(z) = (2\pi)^{-n/2}|\Sigma|^{-1/2}e^{-(z-\mu)^t\Sigma^{-1}(z-\mu)/2}.$$

- If Σ is only non-negative definite (i.e. rank of $\Sigma < n$) then Z lives on a lower dimensional space and has no density on $\mathbb{R}^n$.

Projections

- If an $n \times n$ matrix P satisfies
  - $P^2 = P$ (idempotent)
  - $P = P^t$ (symmetric)

  then P is a projection matrix.
- That is, there exists a subspace $L \subset \mathbb{R}^n$ of dimension $r \leq n$ such that for any $z \in \mathbb{R}^n$ Pz is the projection of z onto L. We write P_L to denote the subspace L projects onto.
- Given any orthonormal basis $\{w_1, \dots, w_r\}$ of L

$$P_L z = \sum_{j=1}^{r} \langle z, w_j \rangle w_j.$$

- If P_L is a projection matrix then

$$I - P_L = P_{L^\perp}$$

  is also a projection matrix which projects onto $L^\perp$, the orthogonal complement of L in $\mathbb{R}^n$.

Projections

- Let $\{X_1, \dots, X_r\}$ be a set of linearly independent vectors in $\mathbb{R}^n$ and

$$X = \begin{pmatrix} X_1 & X_2 & \dots & X_r \end{pmatrix}$$

is the $n \times r$ matrix made by "concatenating" the X_i's.

- If

$$L = \mathrm{span}(X_1, \dots, X_r)$$

is the subspace of $\mathbb{R}^n$ spanned by $\{X_1, \dots, X_r\}$ then

$$P_L = X(X^tX)^{-1}X^t.$$

Identity covariance, projections & χ^2

- If $\Sigma = \sigma^2 I$ and L is a subspace of $\mathbb{R}^n$ then

$$P_L Z \sim N(P_L \mu, \sigma^2 P_L)$$

where P_L is the projection matrix onto L.

- If $P_L \mu = 0$ then

$$\|P_L Z\|^2 \sim \chi^2_{\dim(L)}$$

and

$$\dim(L) = \mathrm{Tr}(P_L).$$

- If $P_L \mu \neq 0$ then

$$\|P_L Z\|^2 \sim \chi^2_{\dim(L), \|P_L \mu\|^2}$$

has a *non-central* χ^2 distribution.

Properties of multiple regression estimates

- $\widehat{\beta} \sim N(\beta, \sigma^2 (X^t X)^{-1})$.
- As in simple regression

$$\widehat{\sigma}^2 = MSE = \frac{SSE}{n-p} \sim \sigma^2 \cdot \frac{\chi^2_{n-p}}{n-p}$$

  independent of $\widehat{\beta}$.
- The least squares estimates are "minimum variance" linear unbiased estimators. (Gauss-Markov theorem)