STAT 331

Simple Linear Regression

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \;,\; \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \;,\; \text{MSS} = \hat{\beta}_1^2 S_{xx}$$

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-p-1}\sum_{i=1}^{n} r_i^2$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}} \;,\; \text{se}(y_p - \hat{y_p}) = \sqrt{\hat{\sigma}^2\left[1 + \frac{1}{n} + \frac{(x_p-\bar{x})^2}{S_{xx}}\right]}$$

Random Vectors

$$\vec{Y} = (Y_1, Y_2, \ldots, Y_n)^T$$

$$\mathbb{E}(\vec{Y}) = (\mathbb{E}(Y_1), \mathbb{E}(Y_2), \ldots, \mathbb{E}(Y_n))^T$$

$$\text{Var}\,\vec{Y} = \mathbb{E}[(\vec{Y} - \mathbb{E}(\vec{Y}))(\vec{Y} - \mathbb{E}(\vec{Y}))^T] = \mathbf{\Sigma}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \text{Var}(Y_1) & \text{Cov}(Y_1,Y_2) & \cdots & \text{Cov}(Y_1,Y_n) \\ \text{Cov}(Y_2,Y_1) & \text{Var}(Y_2) & \cdots & \text{Cov}(Y_2,Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \text{Var}(Y_n) \end{bmatrix}$$

Properties of $\mathbf{\Sigma}$

(1) $\mathbf{\Sigma}$ is **Symmetric**. $\mathbf{\Sigma} = \mathbf{\Sigma}^T$

(2) $\mathbf{\Sigma}$ is **positive semi-definite** ($\vec{a}^t\mathbf{\Sigma}\vec{a} \geq 0 \;\; \forall \vec{a} \in \mathbb{R}^n$)

(3) If $Y_1, \ldots, Y_n$ are independent then $\text{Cov}(Y_i, Y_j) = 0 \forall i \neq j$. $\mathbf{\Sigma}$ is a diagonal matrix.

Basic Properties of Random Vectors

Let A be a $n \times n$ matrix of constants and $\vec{b}$ be a $n \times 1$ vector of constants.

(1) $\mathbb{E}(b^T\vec{Y}) = b^T\mathbb{E}(\vec{Y})$

(2) $\text{Var}(b^T\vec{Y}) = b^T\,\text{Var}(\vec{Y})b$

(3) $\mathbb{E}[A\vec{Y} + b] = A\mathbb{E}[Y] + b$

(4) $\text{Var}(A\vec{Y} + b) = A\,\text{Var}(\vec{Y})A^T$

Multivariate Normal Distribution

$\vec{Y} \sim MVN$ has the p.d.f

$$f(\vec{y}) = (2\pi)^{\frac{-n}{2}}|\mathbf{\Sigma}|^{\frac{-1}{2}}\exp\left(\frac{-1}{2}(\vec{y}-\vec{\mu})^T\mathbf{\Sigma}^{-1}(\vec{y}-\vec{\mu})\right)$$

$$\mathbb{E}[\vec{Y}] = \vec{\mu}$$

$$\text{Var}(\vec{Y}) = \mathbf{\Sigma}$$

Properties of MVN

(1) $A\vec{Y} + b \sim \text{MVN}(A\vec{\mu} + b, A\mathbf{\Sigma}A^T)$, $b^T\vec{Y} \sim N(b^T\vec{\mu}, b^T\mathbf{\Sigma}b)$

(2) $Y_i \sim N(\mu_i, \sigma_i^2)$, $\sigma_i^2 = \mathbf{\Sigma}_{ii}$

(3) If $\vec{Y} \sim \text{MVN}$ then $\mathbf{\Sigma}$ is a diagonal matrix $\iff$ $Y_1, \ldots, Y_n$ are independent.

(4) $\vec{V} = A\vec{Y}, \vec{W} = B\vec{Y}$. V, W are independent $\iff A\mathbf{\Sigma}B^T = 0$

Vector-Matrix Differentiation

$$\frac{\partial}{\partial\vec{x}}\left(a^T\vec{x}\right) = a = \frac{\partial}{\partial\vec{x}}\left(\vec{x}^Ta\right)$$

$$\frac{\partial}{\partial\vec{x}}\left(\vec{x}^TA\vec{x}\right) = 2A\vec{x}$$

Multiple Linear Regression Models

$$Y_i = \beta_0 + \sum_{j=1}^{p}\beta_j x_{ij} + R_i \;,\; R_i \sim N(0, \sigma^2)$$

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \cdots \beta_p x_{ip} = \begin{pmatrix}1 & x_{i1} & \cdots & x_{ip}\end{pmatrix}^T\vec{\beta}$$

$$\mathbf{Y} = \begin{pmatrix} \beta_0+\beta_1x_{11}+\cdots\beta_px_{1p} \\ \beta_0+\beta_1x_{21}+\cdots\beta_px_{2p} \\ \vdots \\ \beta_0+\beta_1x_{p1}+\cdots\beta_px_{pp} \end{pmatrix} + \vec{R} = \mathbf{X}\vec{\beta} + \vec{R}$$

Least Square Estimation

$$\arg\min_{\vec{\beta}}\sum_{i=1}^{n}(y_i - \mathbf{X}_i^T\vec{\beta})^2 = (\vec{y} - \mathbf{X}\vec{\beta})(\vec{y} - \mathbf{X}\vec{\beta})^T$$

$$\frac{\partial}{\partial\vec{\beta}}\left(\vec{y} - \mathbf{X}\vec{\beta}\right)(\vec{y} - \mathbf{X}\vec{\beta})^T = -2X^T\vec{y} + 2(X^TX)\vec{\beta}$$

$$\hat{\vec{\beta}} = (X^TX)^{-1}X^T\vec{y} \;,\; \tilde{\vec{\beta}} = (X^TX)^{-1}X^T\vec{Y}$$

$$\tilde{\sigma^2} = \frac{1}{n-p-1}\sum_{i=1}^{n}\tilde{r_i^2}$$

Sampling Distribution of OLS

$$\tilde{\beta} = (X^TX)^{-1}X^T\vec{Y}$$

$$\mathbb{E}(\tilde{\beta}) = \mathbb{E}((X^TX)^{-1}X^T\vec{Y})$$

$$= (X^TX)^{-1}X^T\mathbb{E}(\vec{Y})$$

$$= (X^TX)^{-1}X^TX\vec{\beta} = \beta$$

$$\text{Var}(\tilde{\beta}) = \sigma^2(X^TX)^{-1}$$

$$\tilde{\beta} \sim \text{MVN}(\beta, \sigma^2(X^TX)^{-1}) \;,\; \tilde{\beta}_j \sim N(\beta_j, \text{Var}(\tilde{\beta}_j))$$

$$\frac{\sum_{i=1}^{n}\tilde{r_i^2}}{\sigma^2} \sim \chi^2_{n-p-1}$$

$$\mathbb{E}\left[\frac{\sum_{i=1}^{n}\tilde{r_i^2}}{\sigma^2}\right] = n-p-1 \Rightarrow \mathbb{E}\left[\frac{1}{n-p-1}\sum_{i=1}^{n}\tilde{r_i^2}\right] = \sigma^2$$

Fitted Values

$$\hat{\mu} = X\hat{\beta} = \underbrace{X(X^TX)^{-1}X^T}_{H}\vec{Y}$$

Let H be the **Hat Matrix**.

(1) H is symmetric $H^T = H$

(2) H is idempotent $H = H^2$

(3) $I - H$ is idempotent $(I-H) = (I-H)(I-H)$

$$\vec{r} = \vec{y} - \hat{\vec{\mu}} = \vec{y} - H\vec{y} = (I-H)\vec{y}$$

$$\sum_{i=1}^{n}\hat{r}_i = 0$$

$$\sum_{i=1}^{n}\hat{r}_i x_{ij} = 0 \;,\; \mathbf{X}^T\hat{\vec{r}} = \vec{0}$$

$$\sum_{i=1}^{n}\hat{r}_i\hat{\mu}_i = 0$$

Estimation of σ^2

$$\tilde{\sigma^2} = \frac{1}{n-p-1}\sum_{i=1}^{n}\hat{r}_i$$

Inference in MLR

Inference about β_j

$$\tilde{\beta}_j \sim N(\beta_j, \sigma^2(X^TX)^{-1}_{jj})$$

$$\frac{\tilde{\beta}_j - \beta_j}{\sqrt{\tilde{\sigma^2}(X^TX)^{-1}_{jj}}} \sim t_{n-p-1}$$

$$\text{se}(\tilde{\beta}_j) = \sqrt{\tilde{\sigma^2}(X^TX)^{-1}_{jj}}$$

$(1-\alpha)100\%$ CI for $\tilde{\beta}_j$ is

$$\hat{\beta}_j \pm t_{n-p-1}\left(1 - \frac{\alpha}{2}\right)\text{se}(\tilde{\beta}_j)$$

$H_0 : \beta_j = 0$ then the t value is

$$|t| = \frac{\hat{\beta}_j - \beta_j}{\text{se}(\tilde{\beta}_j)}$$

Inference about Mean Response

$$\tilde{\mu}(c) = c^T\tilde{\vec{\beta}}$$

$$\tilde{\mu}(c) \sim N\left(c^T\vec{\beta}, c^T\left[\sigma^2(X^TX)^{-1}\right]c\right)$$

$$\frac{\tilde{\mu}(c) - \mu(c)}{\sqrt{c^T\left[\sigma^2(X^TX)^{-1}\right]c}} \sim t_{n-p-1}$$

$$\text{se}(\tilde{\mu}(c)) = \sqrt{c^T\left[\sigma^2(X^TX)^{-1}\right]c}$$

$(1-\alpha)100\%$ CI for $\mu(c)$ is

$$\hat{\mu}(c) \pm t_{n-p-1}\left(1 - \frac{\alpha}{2}\right)\text{se}(\tilde{\mu}(c))$$

Prediction Interval

$Y_p = c^T\vec{\beta} + R_p$

$\mathbb{E}[Y_p - \tilde{Y}_p] = 0$

$$\mathrm{Var}(Y_p - \tilde{Y}_p) = \sigma^2 + c^T \underbrace{\left[\sigma^2(X^TX)^{-1}\right]}_{\mathrm{Var}(\tilde{\beta})} c$$

$\mathrm{se}(Y_p - \tilde{Y}_p) = \sqrt{\sigma^2 + c^T\,\mathrm{Var}(\tilde{\beta})c}$

$\dfrac{Y_p - \tilde{Y}_p}{\mathrm{se}(Y_p - \tilde{Y}_p)} \sim t_{n-p-1}$

$(1-\alpha)100\%$ Prediction Interval for Y_p is

$$\hat{Y}_p \pm t_{n-p-1}\left(1 - \frac{\alpha}{2}\right)\mathrm{se}(Y_p - \tilde{Y}_p)$$

ANOVA in MLR

Source	Sum of Squares	$d.f$	MS
Model	$\sum_{i=1}^{n}(\hat{\mu}_i - \bar{y})^2$	p	$\frac{1}{p}\sum_{i=1}^{n}(\hat{\mu}_i - \bar{y})^2$
Residual	$\sum_{i=1}^{n}\hat{r}_i^2$	$(n-p-1)$	$\frac{1}{n-p-1}\sum_{i=1}^{n}\hat{r}_i^2$
Total	$\sum_{i=1}^{n}(y_i - \bar{y})^2$	$(n-1)$	$\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$

F-test for Significance of Model. $H_0 : \beta_0 = \beta_1 = \cdots = 0$, $H_a : \exists\beta_j \neq 0$.

$$F = \frac{\mathrm{MMS}}{\mathrm{RMS}} = \frac{MSS/p}{RSS/(n-p-1)}$$

If H_0 is true then $F \sim F_{p,n-p-1}$. $p-$value $= \mathbb{P}(F_{p,n-p-1} > F)$

Coefficient of Determination

$$R^2 = 1 - \frac{RSS}{TSS} = \frac{MSS}{TSS}$$

$$R^2_{\mathrm{adj}} = 1 - \frac{RMS}{TMS} = 1 - \frac{n-1}{n-p-1}(1-R^2)$$

Geometric Interpretation

$\mathbf{X} = [\,\mathbb{1}_n\ \mathbf{x_1}\ \cdots\ \mathbf{x_p}\,]\ \ \vec{\mu} = \mathbf{X}\vec{\beta}$

Geometrically $\vec{\mu} \in \mathrm{Span}(\mathbb{1}_n, \mathbf{X_1}, \ldots, \mathbf{X_p}) = \mathrm{Col}(\mathbf{X})$.
We need to choose $\hat{\mu} \in \mathrm{Col}(\mathbf{X})$ such that $\hat{\mu}$ is the **closest** to observed $\mathbf{y}$.
The point which makes the residue vector $\hat{\vec{r}} = y - \hat{y}$ the smallest is when it is perpendicular to $\mathrm{Col}(\mathbf{X})$.

$$\hat{\vec{r}} \perp \mathrm{Col}(\mathbf{X})$$

$$\iff X_j^T\hat{\vec{r}} = 0 \quad \iff \quad \mathbf{X}^T\hat{\vec{r}} = \mathbb{0}_{p+1}$$

$$\mathbf{X}^T\hat{\vec{r}} = \mathbb{0}_{p+1}\ ,\ \ \mathbf{X}^T(y - \mathbf{X}\hat{\beta}) = \mathbb{0}_{p+1}$$

$$\hat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^Ty$$

$$\hat{\vec{\mu}} = \mathbf{X}\hat{\vec{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^Ty = \mathbf{H}y$$

Test General Linear Hypothesis

$C_{q\times(p+1)}, \vec{b}_{q\times 1}$

$H_0 : C\vec{\beta} = \vec{b}$

C has a row rank of q. Each row corresponds to a linear restriction on $\vec{\beta}$

$C\tilde{\beta} \sim \mathrm{MVN}(C\beta, \sigma^2C(X^TX)^{-1}C^T)$

$(C\tilde{\beta} - C\beta)^T[\sigma^2C(X^TX)^{-1}C^T]^{-1}(C\tilde{\beta} - C\beta) \sim \chi^2_q$

$F-$test for $H_0 : C\vec{\beta} = \vec{b}$. If H_0 is true then

$$F = \frac{(C\tilde{\beta} - C\beta)^T[\sigma^2C(X^TX)^{-1}C^T]^{-1}(C\tilde{\beta} - C\beta)}{q} \sim F_{q,n-p-1}$$

$p-$value $\mathbb{P}(F_{q,n-p-1} > F)$. q is the number of constraints/restrictions on $\vec{\beta}$.
Restricted Sum of Squares: RSS_C residual sum of squares for the restricted model under the linear constraints. We can write F statistic as

$$F = \frac{(RSS_C - RSS)/q}{RSS/(n-p-1)} \sim F_{q,n-p-1}$$

Model Diagnostic

Assumptions in Linear Regression

- Linearity $\mu_i = \mathbb{E}(Y_i) = \beta_0 + \beta_1x_{i1} + \cdots + \beta_px_{ip}$
- Independence R_i are independent
- Homoscedasticity: $\forall i\ \ \mathrm{Var}(R_i) = \sigma^2$ (Equal varriance)
- Normality: $R_i \sim N(0, \sigma^2)$

Residual Plots

$\hat{r}_i = y_i - \hat{\mu}_i$

$\tilde{\hat{\vec{r}}}_i = \vec{Y} - \tilde{\hat{\vec{\mu}}} = \vec{Y} - X\tilde{\hat{\vec{\beta}}} = \vec{Y} - H\vec{Y} = (I - H)\vec{Y}$

$\vec{R} \sim \mathrm{MVN}(\vec{\mathbb{0}}, \sigma^2I) \Rightarrow \tilde{\vec{r}} \sim \mathrm{MVN}(\vec{\mathbb{0}}, \sigma^2(I-H))$

$\sigma^2(I-H)$ might not have non-zero non-diagonal entries. If H is small relative to I then $\tilde{\vec{r}} \approx \vec{R}$ and $\tilde{\vec{r}} \sim\approx \mathrm{MVN}(\vec{\mathbb{0}}, \sigma^2I)$.
$\hat{\mathrm{Var}}(\tilde{r}_i) = \sqrt{\hat{\sigma^2}(1-h_{ii})}$. Where h_ii is the $i-$th diagonal of H.

Standardized Residuals: $\dfrac{\hat{r_i}}{\hat{\sigma}}$

Studentized Residuals: $\dfrac{\hat{r_i}}{\sqrt{\hat{\sigma^2}(1-h_{ii})}}$

Data Transformation

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(y_i) & \lambda = 0 \end{cases} \qquad \text{(Box-Cox)}$$

Weighted Least Squares

$Y_i = \vec{x_i}^T\vec{\beta} + R_i$

$\mathbb{E}[R_i] = 0\ ,\ \mathrm{Var}(R_i) = \dfrac{\sigma^2}{w_i}$

$w_iY_i = \sqrt{w_i}\vec{x_i}^T\vec{\beta} + \sqrt{w_i}R_i$

$R_i^* = \sqrt{w_i}R_i\ ,\ \mathrm{Var}(R_i^*) = \sigma^2$

WLS is equivalent to regressing transformed $\sqrt{w_i}y_i$ on the transformed co-variates $\sqrt{w_i}\vec{x}_i$. Where W needs to be known (or estimated).

$$W = \begin{pmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{pmatrix} \qquad \text{(Weights)}$$

$$s(\vec{\beta}) = \sum_{i=1}^{n}\left(\sqrt{w_i}y_i - \sqrt{w_i}\vec{x}_i^T\vec{\beta}\right)^2 = (\vec{y} - X\vec{\beta})^TW(\vec{y} - X\vec{\beta})$$

$\dfrac{\partial s}{\partial\vec{\beta}} = -2X^TW(\vec{y} - X\vec{\beta})$

$\hat{\beta}_{WLS} = (X^TWX)^{-1}X^TW\vec{y}$

$\mathrm{Var}\,\tilde{\beta}_{WLS} = (X^TWX)^{-1}X^TW\,\mathrm{Var}(\vec{Y})WX(X^TWX)^{-1}$

$\mathrm{Var}\,\tilde{\beta}_{WLS} = (X^TWX)^{-1}$ if $\mathrm{Var}(\vec{Y}) = W^{-1}$

$\mathrm{Var}\,\tilde{\beta}_{WLS} = \sigma^2(X^TWX)^{-1}$ if $\mathrm{Var}(\vec{Y}) = \sigma^2W^{-1}$

$$\hat{\sigma}^2 = \frac{1}{n-p-1}\sum_{i=1}^{n}w_i(y_i - \vec{x}_i^T\hat{\beta}_{WLS})^2$$

Computing weights:

```
residfit = lm(abs(rstand.ols) ~ muhat.ols)
wts = 1/(fitted(residfit)^2)
```

Influential Points

$\dfrac{\mathrm{d}\hat{\mu}_i}{\mathrm{d}y_i} = h_{ii}$ **leverage** of point ith observation

$\dfrac{\mathrm{d}\hat{\mu}_i}{\mathrm{d}\hat{r}_i} = \dfrac{h_{ii}}{1-h_{ii}}$ **elasticity** of point ith observation

$\sum_{i=1}^{n}h_{ii} = \mathrm{trace}(H) = p\ ,\ h_{\mathrm{avg}} = \dfrac{p}{n}$

$$d_i = \frac{\hat{r}_i}{1-h_{ii}}\ ,\ \frac{d_i}{\mathrm{se}(\tilde{d}_i)} = \frac{\hat{r}_i}{\sqrt{\hat{\sigma}^2_{(i)}1 - h_{ii}}} \qquad \text{(deleted residual)}$$

$$D_i = \frac{\left\|\hat{\mu}^{(i)} - \hat{\mu}\right\|}{\hat{\sigma}^2(p+1)} = d_i^2\frac{h_i}{1-h_{ii}}\cdot\frac{1}{p+1} \qquad \text{(Cook's Distance)}$$

Variance Inflation Factors (VIF) $\mathrm{VIF}_i = \frac{1}{1-R_j^2}$.

Model Selection

$$C_p = \frac{\mathrm{SSR}_p}{\sigma^2_{\mathrm{full}}} - n + 2(p+1) \qquad \text{(Mallow's } C_p\text{)}$$

$$\mathbf{AIC} = 2(p+1) - 2\ln\left(L(\hat{\theta})\right) \qquad \text{(Akaike's IC)}$$

$$\mathbf{AIC}_{\mathrm{Gaussian}} = 2(p+1) - \frac{n}{2}\ln(\mathrm{SSR})$$

$$\mathbf{BIC} = (p+1)\ln(n) - 2\ln\left(L(\hat{\theta})\right) \qquad \text{(Bayesian IC)}$$