$$\frac{y-\bar{y}}{SD_y} = r\frac{x-\bar{x}}{SD_x}$$

$$y = \alpha + \beta x$$

$$\mathsf{RSS} = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

$$\sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i(y_i - \bar{y}) = \sum_i (x_i - \bar{x})y_i$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x},$$

$$\hat{\beta}_1 = \frac{\mathsf{Sxy}}{\mathsf{Sxx}} = r_{\mathsf{XY}}\left(\frac{\mathsf{Syy}}{\mathsf{Sxx}}\right)^{1/2}$$

$$\mathsf{Sxy} = \sum(x_i - \bar{x})(y_i - \bar{y}),$$

$$\mathsf{Sxx} = \sum(x_i - \bar{x})^2, \quad \mathsf{Syy} = \sum(y_i - \bar{y})^2,$$

$$r_{\mathsf{XY}} = \frac{\mathsf{Sxy}}{\sqrt{(\mathsf{Sxx})(\mathsf{Syy})}} \quad \text{(the sample correlation)}$$

Fitted value

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

Residual

$$r_i = y_i - \hat{y}_i \quad \sum_i r_i = 0, \quad \sum_i r_i x_i = 0.$$

$$\mathsf{RSS} = \sum_{i=1}^{n} r_i^2$$

$$\hat{\sigma}^2 = \frac{RSS}{n-p}$$

TSS $\qquad R^2 = r_{\mathsf{XY}}^2$

$$\begin{aligned}\sum_i (y_i - \bar{y})^2 &= \sum_i (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \sum_i (r_i + \hat{y}_i - \bar{y})^2 \\ &= \sum_i r_i^2 + \sum_i (\hat{y}_i - \bar{y})^2 \\ &= \mathsf{RSS} + \mathsf{FSS},\end{aligned}$$

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{\mathsf{FSS}}{\mathsf{TSS}} = \frac{\mathsf{TSS} - \mathsf{RSS}}{\mathsf{TSS}} = 1 - \frac{\mathsf{RSS}}{\mathsf{TSS}}.$$

Regression Through the Origin

$$\sum_i y_i^2 = \sum_i (y_i - \hat{y}_i + \hat{y}_i)^2 = \sum_i (y_i - \hat{y}_i)^2 + \sum_i \hat{y}_i^2$$

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \qquad \tilde{R}^2 = \frac{\sum_i \hat{y}_i^2}{\sum_i y_i^2} = 1 - \frac{\mathsf{RSS}}{\sum_i y_i^2}$$

Properties of LS Estimates

$$\mathbb{E}[y_i \mid x_i] = \beta_0 + \beta_1 x_i, \quad \mathsf{Cov}[y_i, y_j \mid x_i, x_j] = \sigma^2\delta_{ij}$$

$\delta_{ij} = 1$ if $i = j$ and 0 if $i \neq j$.

unbiased

$$\hat{\beta}_1 = \sum_i \frac{(x_i - \bar{x})}{\mathsf{Sxx}} y_i = \sum_i c_i y_i, \quad \sum_i c_i = 0$$

$$\mathbb{E}\hat{\beta}_1 = \sum_i c_i \mathbb{E}y_i = \sum_i c_i(\beta_0 + \beta_1 x_i) = \beta_1\left(\sum_i c_i x_i\right) = \beta_1$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$\mathbb{E}\hat{\beta}_0 = \left(\frac{1}{n}\sum_i \mathbb{E}y_i\right) - \bar{x}\cdot\mathbb{E}\hat{\beta}_1 = \beta_0 + \beta_1\bar{x} - \beta_1\bar{x} = \beta_0$$

$$\mathsf{Var}(\hat{\beta}_1) = \mathsf{Var}\left(\sum_i c_i y_i\right) = \sigma^2\sum c_i^2 = \sigma^2\frac{1}{\mathsf{Sxx}}$$

$$\mathsf{Var}(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\mathsf{Sxx}}\right)$$

y_i indep. $\sim \mathsf{N}(\beta_0 + \beta_1 x_i, \sigma^2)$ $\qquad e_i$ iid $\sim \mathsf{N}(0, \sigma^2)$

Multiple Linear Regression (MLR)

Least Squares Estimation

$$\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \mathbf{e}_{n\times 1}, \quad \mathbf{e} \sim \mathsf{N}_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$$

$$\mathsf{RSS} = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^t(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}$$

$$= \frac{1}{n\sum x_i^2 - (n\bar{x})^2}\begin{pmatrix}\sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n\end{pmatrix}\begin{pmatrix} n\bar{y} \\ \sum x_i y_i\end{pmatrix}$$

$$\hat{\beta}_1 = \frac{-n^2\bar{x}\bar{y} + n\sum x_i y_i}{n\sum x_i^2 - (n\bar{x})^2} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\mathsf{Sxy}}{\mathsf{Sxx}}$$

Fitted value

$$\hat{\mathbf{y}}_{n\times 1} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y} = \mathbf{H}_{n\times n}\mathbf{y}_{n\times 1}$$

Residuals $\quad \mathbf{r}_{n\times 1} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y}.$

normal equation $\quad \mathbf{X}^t(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}.$

$\mathbf{r} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ satisifies: $\mathbf{X}^t\mathbf{r} = \mathbf{0}$, $\hat{\mathbf{y}}^t\mathbf{r} = \hat{\boldsymbol{\beta}}^t\mathbf{X}^t\mathbf{r} = 0$

$\mathbf{r}$ is orthogonal to each column of $\mathbf{X}$ and $\hat{\mathbf{y}}$

The Hat Matrix

$$\mathbf{H}_{n\times n} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t \quad \mathbf{H}\mathbf{X} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X} = \mathbf{X}$$

Symmetric: $\mathbf{H}^t = \left[\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\right]^t = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t = \mathbf{H}$

Idempotent [a]: $\mathbf{H}\mathbf{H} = \mathbf{H}\mathbf{H}^t = \mathbf{H} \qquad \mathbf{H}(\mathbf{I} - \mathbf{H}) = \mathbf{0}_{n\times n}.$

$$\begin{aligned}\mathbb{E}(\hat{\boldsymbol{\beta}}) &= \mathbb{E}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y} \\ &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbb{E}\mathbf{y} \\ &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta} \\ \mathsf{Cov}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathsf{Cov}(\mathbf{y})\left[(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\right]^t \\ &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\sigma^2\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^t\mathbf{X})^{-1} \\ \mathbb{E}(\hat{\mathbf{y}}) &= \mathbf{X}\boldsymbol{\beta}, \quad \mathsf{Cov}(\hat{\mathbf{y}}) = \sigma^2\mathbf{H}; \\ \mathbb{E}(\mathbf{r}) &= \mathbf{0}, \quad \mathsf{Cov}(\mathbf{r}) = \sigma^2(\mathbf{I}_n - \mathbf{H}) \\ \mathbb{E}(\hat{\sigma}^2) &= \frac{1}{n-p}\mathbb{E}\mathbf{r}^t\mathbf{r} = \frac{1}{n-p}\mathsf{tr}\left[\mathbb{E}\mathbf{r}^t\mathbf{r}\right] = \frac{1}{n-p}\mathsf{tr}\left[\mathbb{E}\mathbf{r}\mathbf{r}^t\right] = \sigma^2\end{aligned}$$

Distributions of LS Estimates $\quad \mathbf{y} \sim \mathsf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}\mathbf{y} \sim \mathsf{N}_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^t\mathbf{X})^{-1})$$

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} \sim \mathsf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{H})$$

$$\mathbf{r} = (\mathbf{I}_n - \mathbf{H})\mathbf{y} \sim \mathsf{N}_n(\mathbf{0}, \sigma^2(\mathbf{I}_n - \mathbf{H}))$$

$$\mathbb{E}\hat{\mathbf{y}} = \mathbf{H}\mathbb{E}\mathbf{y} = \mathbf{H}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} \qquad \mathsf{Cov}(\hat{\mathbf{y}}) = \mathbf{H}\sigma^2\mathbf{H}^t = \sigma^2\mathbf{H}$$

$$\mathbb{E}\mathbf{r} = (\mathbf{I}_n - \mathbf{H})\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \qquad \mathsf{Cov}(\mathbf{r}) = (\mathbf{I}_n - \mathbf{H})\sigma^2(\mathbf{I}_n - \mathbf{H})^t = \sigma^2(\mathbf{I}_n - \mathbf{H})$$

$$\hat{\sigma}^2 = \frac{\|\mathbf{r}\|^2}{n-p} \sim \sigma^2\frac{\chi^2_{n-p}}{n-p}$$

Source	df	SS	MS	F
Regression	$p-1$	FSS	FSS$/(p-1)$	MS(reg)/MS(err)
Error	$n-p$	RSS	RSS$/(n-p)$	
Total	$n-1$	TSS		

$$\frac{\mathsf{MS(reg)}}{\mathsf{MS(err)}} \sim F_{(p-1),n-p}$$

Nested Models

$$F = \frac{(\mathsf{RSS}_0 - \mathsf{RSS}_a)/q}{\mathsf{RSS}_a/(n-p)} \sim F_{q,n-p}$$

ANCOVA

$$y = \beta_0+\beta_1 x+\beta_2 d+\beta_3(x\cdot d)+e = \begin{cases} \beta_0+\beta_1 x+e, \\ (\beta_0+\beta_2)+(\beta_1+\beta_3)x+e, \end{cases}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & x_1 & 0 \\ 1 & 0 & x_2 & 0 \\ 1 & 1 & x_3 & x_3 \\ 1 & 1 & x_4 & x_4 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_2 \\ \beta_1 \\ \beta_3 \end{pmatrix}$$

categorical variable D produces an additive change in Y and also changes the effect of X on Y

$$\frac{(\text{RSS}_0-\text{RSS}_a)/(k-1)}{\text{RSS}_A/(n-p_A)} = \frac{(\text{RSS}_0-\text{RSS}_a)/(k-1)}{\hat\sigma^2_A}$$

which follows $F_{k-1,n-p_A}$ under the null, where RSS_A denotes the RSS from the biggest model $Y \sim X+D+X:D$ and $p_A = 2k$.

```
anova(lm(Y ~ X + D + X:D))
```

H_0	H_a
$Y \sim 1$	$Y \sim X$
$Y \sim X$	$Y \sim X+D$
$Y \sim X+D$	$Y \sim X+D+X:D$

Parallel regression lines

$$y = \beta_0+\beta_2 d+\beta_1 x+e$$

β_2: measures the change of the additive effect

Regression lines with equal intercepts but different slopes:

$$y = \beta_0+\beta_1 x+\beta_3(x\cdot d)+e$$

β_3: measures the change of the slope.

The QR Decomposition (*)

How is the LS estimate $\hat\beta$ solved in R? Denote the QR decomposition of $\mathbf{X}$ as

$$\mathbf{X}_{n\times p} = \mathbf{Q}_{n\times p}\mathbf{R}_{p\times p}$$

where $\mathbf{Q}$ is an orthogonal matrix (i.e., $\mathbf{Q}^t\mathbf{Q} = \mathbf{I}_p$) and $\mathbf{R}$ is an upper triangular matrix, i.e., all the entries in $\mathbf{R}$ below the diagonal are equal to 0.

$$\begin{aligned} \hat\beta &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y} \\ (\mathbf{X}^t\mathbf{X})^{-1} &= (\mathbf{R}^t\mathbf{R})^{-1} = \mathbf{R}^{-1}(\mathbf{R}^t)^{-1} \\ \hat\beta &= \mathbf{R}^{-1}\mathbf{Q}^t\mathbf{y} \\ \mathbf{R}\hat\beta &= \mathbf{Q}^t\mathbf{y} \end{aligned}$$

Gram-Schmidt (*)

One method for computing the QR decomposition is the *Gram-Schmidt* algorithm. Let's work with a matrix

$$\mathbf{A}_{n\times p} = \big[\mathbf{a}_1 \mid \mathbf{a}_2 \mid \cdots \mid \mathbf{a}_p\big],$$

where $\mathbf{a}_j$ denotes the jth column of $\mathbf{A}$. Then

- $\mathbf{e}_1 = \mathbf{a}_1, \quad \mathbf{u}_1 = \frac{\mathbf{e}_1}{\|\mathbf{e}_1\|}$
- $\mathbf{e}_2 = \mathbf{a}_2 - (\mathbf{a}_2^t\mathbf{u}_1)\mathbf{u}_1, \quad \mathbf{u}_2 = \frac{\mathbf{e}_2}{\|\mathbf{e}_2\|}$
- $\cdots$
- $\mathbf{e}_{k+1} = \mathbf{a}_{k+1} - \sum_{j=1}^k(\mathbf{a}_j^t\mathbf{u}_j)\mathbf{u}_j, \quad \mathbf{u}_{k+1} = \frac{\mathbf{e}_{k+1}}{\|\mathbf{e}_{k+1}\|}$

The resulting QR decomposition is

$$\mathbf{A} = \big[\mathbf{a}_1 \mid \mathbf{a}_2 \mid \cdots |\mathbf{a}_p\big] = \big[\mathbf{u}_1 \mid \cdots \mid \mathbf{u}_p\big]\mathbf{R} = \mathbf{QR}.$$

- Affine transformations: $\mathbf{W} = \mathbf{a}_{n\times 1}+\mathbf{B}_{n\times m}\mathbf{Z}$,

$$\mathbb{E}[\mathbf{W}] = \mathbf{a}+\mathbf{B}\boldsymbol{\mu}, \quad \text{Cov}(\mathbf{W}) = \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^t.$$

Especially, for $W = v_1Z_1+\cdots v_mZ_m = \mathbf{v}^t\mathbf{Z}$,

$$\begin{aligned} \mathbb{E}[W] &= \mathbf{v}^t\boldsymbol{\mu} = \sum_{i=1}^m v_i\mu_i, \\ \text{Var}(W) &= \mathbf{v}^t\Sigma\mathbf{v} = \sum_{i=1}^m v_i^2\text{Var}(Z_i)+2\sum_{i<j}v_iv_j\text{Cov}(Z_i,Z_j) \end{aligned}$$

$$\mathbf{x}_*^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{x}_* = \frac{1}{n}+\frac{1}{n-1}(\mathbf{z}^*-\bar{\mathbf{z}})^t\hat\Sigma^{-1}(\mathbf{z}^*-\bar{\mathbf{z}}),$$

which is the so-called Mahalanobis distance from $\mathbf{x}_i$ to the center of the

Hypothesis Testing

$$t = \frac{\hat\beta_j-c}{\text{se}(\hat\beta_j)} = \frac{\hat\beta_j-c}{\hat\sigma\sqrt{[(\mathbf{X}^t\mathbf{X})^{-1}]_{jj}}} \sim T_{n-p}$$

SLR

$$\frac{\hat\beta_1-c}{\text{se}(\hat\beta_1)} = \frac{\hat\beta_1-c}{\hat\sigma/\sqrt{\text{Sxx}}} \sim T_{n-2}$$

$$\left(\hat\beta_j \ \pm \ t_{n-p}^{(\alpha/2)}\text{se}(\hat\beta_j)\right)$$

CI $\quad \hat\beta_0+\hat\beta_1x_* \ \pm \ t_{n-2}^{(\alpha/2)}\hat\sigma\sqrt{\frac{1}{n}+\frac{(x_*-\bar x)^2}{\text{Sxx}}}$

PI $\quad \hat\beta_0+\hat\beta_1x_* \ \pm \ t_{n-2}^{(\alpha/2)}\hat\sigma\sqrt{1+\frac{1}{n}+\frac{(x_*-\bar x)^2}{\text{Sxx}}}$

$\text{se}(\hat\mu^*) = \hat\sigma\sqrt{(\mathbf{x}^*)^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{x}^*}$.

$\text{se}(\hat y^*) = \hat\sigma\sqrt{1+(\mathbf{x}^*)^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{x}^*}$.

The Gauss-Markov theorem tells us that the BLUE of μ^* is

$$\hat\theta_{LS} = \mathbf{c}^t\hat{\boldsymbol\beta} \qquad \hat\mu^* = (\mathbf{x}^*)^t\hat{\boldsymbol\beta}^t = (\mathbf{x}^*)^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}.$$

This is just a linear transformation of $\mathbf{y}$, so we can easily derive its variance, and find its standard error.

$$\text{se}(\hat\mu^*) = \hat\sigma\sqrt{(\mathbf{x}^*)^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{x}^*}.$$

A CI for μ^* is given by

$$\left(\hat\mu^*-t_{n-p}^{(\alpha/2)}\text{se}(\hat\mu^*),\ \hat\mu^*+t_{n-p}^{(\alpha/2)}\text{se}(\hat\mu^*)\right).$$

Joint Confidence Region

Just as we can use estimated standard errors and t-stats to form confidence intervals for a single parameter, we can also obtain a $(1-\alpha)\times 100\%$ confidence region for the entire vector $\boldsymbol\beta$. In particular

$$\boldsymbol\beta-\hat{\boldsymbol\beta} \sim \text{N}(\mathbf{0},\sigma^2(\mathbf{X}^t\mathbf{X})^{-1}).$$

Thus, the quadratic form

$$\frac{(\boldsymbol\beta-\hat{\boldsymbol\beta})^t\mathbf{X}^t\mathbf{X}(\boldsymbol\beta-\hat{\boldsymbol\beta})}{p\hat\sigma^2} \sim F_{p,n-p}.$$

Simultaneous CIs/PIs

- Consider a simple linear regression $y_i = \beta_0+\beta_1x_i+e_i$.
- Given a new value x^*, the $(1-\alpha)$ CI for $\mu^* = \beta_0+\beta_1x^*$ is

$$I(x^*) = \left(\hat\mu^* \ \pm \ t_{n-2}^{(\alpha/2)}\text{se}(\hat\mu^*)\right),$$

where

$$\hat\mu^* = \hat\beta_0+\hat\beta_1x^*, \quad \text{se}(\hat\mu^*) = \hat\sigma\sqrt{\frac{1}{n}+\frac{(x^*-\bar x)^2}{\sum_{i=1}^n(x_i-\bar x)^2}}.$$