第7章 方差分析和回归分析

《统计推断》 第11章

感谢清华大学自动化系江瑞教授提供PPT

Analysis of Variance (ANOVA)

统计学方法及其应用 统计模型 方差分析

ANOVA data

	Treatment					
Index	**1**	**2**	**3**	**…**	$\boldsymbol{k-1}$	$\boldsymbol{k}$
1	Y_{11}	Y_{21}	Y_{31}		$Y_{(k-1)1}$	Y_{k1}
2	Y_{12}	Y_{22}	Y_{32}		$Y_{(k-1)2}$	Y_{k2}
$\cdots$	Y_{13}	Y_{23}	Y_{33}		$Y_{(k-1)3}$	Y_{k3}
	$\cdots$	$\cdots$	$\cdots$		$\cdots$	$\cdots$
	Y_{1n_1}				$Y_{(k-1)n_{(k-1)}}$	
			Y_{3n_3}			
		Y_{2n_2}				
						Y_{kn_k}
$\boldsymbol{\theta}$	θ_1	θ_2	θ_3	$\cdots$	θ_{k-1}	θ_k
N	n_1	n_2	n_3	$\cdots$	n_{k-1}	n_k
$\overline{\overline{Y}}$	$\overline{Y}_{1\bullet}$	$\overline{Y}_{2\bullet}$	$\overline{Y}_{3\bullet}$		$\overline{Y}_{(k-1)\bullet}$	$\overline{Y}_{k\bullet}$

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_k \quad \text{versus} \quad H_1 : \theta_i \neq \theta_j \text{ for some } i, j$$

ANOVA model

Random variables Y_{ij} are observed according to the model

$$Y_{ij} = \theta_i + \varepsilon_{ij}, i = 1, \ldots, k, j = 1, \ldots, n_i,$$

where

(i) $\mathrm{E}\varepsilon_{ij} = 0$, $\mathrm{Var}\varepsilon_{ij} = \sigma_i^2 < \infty$, for all i and j.

$\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{st}) = 0$ for all $i, j, s,$ and t unless $i = s$ and $j = t$.

(ii) The ε_{ij} are independent and normally distributed (normal errors).

(iii) $\sigma_i^2 = \sigma^2$ for all i (equality of variance, homoscedasticity).

Partitioning sums of squares

$$\mathbf{SST = SSB + SSW}$$

$$\sum_{j=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\overline{\overline{Y}})^2 = \sum_{i=1}^{k} n_i(\overline{Y}_{i\bullet}-\overline{\overline{Y}})^2 + \sum_{j=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\bullet})^2$$

Dividing them by σ^2

$$\chi^2_{N-1} = \chi^2_{k-1} + \chi^2_{N-k}$$

$$\overline{Y}_{i\bullet} = \frac{1}{n_i}\sum_{i=1}^{k} Y_{ij}, \overline{\overline{Y}} = \frac{1}{N}\sum_{i=1}^{k} n_i \overline{Y}_{i\bullet} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}$$

ANOVA table

Source of variation	Degrees of freedom	Sum of squares	Mean square	*F* statistic	*p* value
Between treatment groups	$k-1$	**SSB** $\sum_{i=1}^{k} n_i(\bar{y}_{i\bullet}-\bar{\bar{y}})^2$	**MSB** $\frac{\mathrm{SSB}}{k-1}$	$F=\frac{\mathbf{MSB}}{\mathbf{MSW}}$	$1-F_{k-1,N-k}(F)$
Within treatment groups	$N-k$	**SSW** $\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\bullet})^2$	**MSW** $\frac{\mathrm{SSW}}{N-k}$		
Total	$N-1$	**SST** $\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{\bar{y}})^2$			

Practice

```
aov(formula, data, ...);

Formula:
     response ~ group



anova(lm(formula, data, ...));

Formula:
     response ~ predictor
```

对比(Contrast)

- Motivation：上述考虑的零假设

$$H_0 : \theta_1 = \cdots = \theta_k$$

  往往是一个不真的假设。实验的目的在于找出哪个处理更好。

- 有必要将ANOVA的假设拆分成更小的、更容易描述的部分。

对比(Contrast)

- 定义：设 $t=(t_1,\cdots,t_k)$ 是变量的集合，这些变量是参数或统计量，$a=(a_1,\cdots,a_k)$ 是k个已知常数，那么函数

$$\sum_{i=1}^{k} a_i t_i$$

称作t的线性组合。如果 $\sum a_i = 0$, 此时线性组合称作对比。

对比的用处

- 对比可以用来比较处理的均值。

- 例子：如果 $\theta_1, \cdots, \theta_k$ 为均值，取 $a = (1, -1, 0, \cdots, 0)$ 则

$$\sum_i a_i \theta_i = \theta_1 - \theta_2$$

为比较 θ_1 和 θ_2 的对比。

ANOVA零假设的对比表示

- 定理：设$\theta = (\theta_1, \cdots, \theta_k)$为任意参数，则

$$\theta_1 = \cdots = \theta_k \leftrightarrow \sum_{i=1}^{k} a_i \theta_i = 0, \forall a \in \mathcal{A}$$

其中 $\mathcal{A}$ 为常数集合

$$\mathcal{A} = \left\{ a = (a_1, \cdots, a_k) : \sum a_i = 0 \right\}$$

即ANVOA的零假设等价于所有对比为0.

ANOVA并-交检验

- 对每个 $a \in \mathcal{A}$, 定义以a为Index的零假设

$$\Theta_a = \{\theta = (\theta_1, \theta_2, \cdots, \theta_k) : \sum_{i=1}^{k} a_i \theta_i = 0\}$$

- 于是ANOVA的零假设可以写成交的形式

$$\theta \in \{\theta : \theta_1 = \cdots = \theta_k\} \leftrightarrow \theta \in \cap_{a \in \mathcal{A}} \Theta_a$$

分布性质

- 对任何常数向量 $a=(a_1,\cdots,a_k)$, 容易验证

$$\sum_{i=1}^{k} a_i \overline{Y}_{i\cdot} \sim N(*,*)$$

$$E(\sum_{i=1}^{k} a_i \overline{Y}_{i\cdot}) = \sum_{i=1}^{k} a_i\theta_i, Var(\sum_{i=1}^{k} a_i \overline{Y}_{i\cdot}) = \sigma^2 \sum_{i=1}^{k} \frac{a_i^2}{n_i}$$

$$\frac{\sum_{i=1}^{k} a_i \overline{Y}_{i\cdot} - \sum_{i=1}^{k} a_i\theta_i}{\sqrt{\sigma^2 \sum_{i=1}^{k} \frac{a_i^2}{n_i}}} \sim N(0,1)$$

分布性质

- 如果定义

$$S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\cdot})^2,\ i = 1, \cdots, k$$

$$S_p^2 = \frac{1}{N-k}\sum_{i=1}^{k}(n_i - 1)S_i^2 = \frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\cdot})^2$$

- 容易验证

$$(n_i - 1)S_i^2/\sigma^2 \sim \chi^2_{n_i-1}, (N-k)S_p^2/\sigma^2 \sim \chi^2_{N-k}$$

$$\frac{\sum_{i=1}^k a_i\overline{Y}_{i\cdot} - \sum_{i=1}^k a_i\theta_i}{\sqrt{S_p^2\sum_{i=1}^k \frac{a_i^2}{n_i}}} \sim t_{N-k}$$

ANOVA并-交检验

- 对于任何a, 考虑假设检验

$$H_{0a}: \theta \in \Theta_a \leftrightarrow H_{1a}: \theta \notin \Theta_a$$

- 采用T统计量

$$T_a = \left| \frac{\sum_{i=1}^k a_i \overline{Y}_{i\cdot} - \sum_{i=1}^k a_i \theta_i}{\sqrt{S_p^2 \sum_{i=1}^k \frac{a_i^2}{n_i}}} \right|$$

给出检验a的拒绝域 $\{x : T_a > c\}$

ANOVA并-交检验

- 由并-交检验，如果每个a的检验都能够拒绝，那么对于极大化T_a的a也能够拒绝。
- 因此作为并-交检验，ANOVA的零假设当

$$\sup_a T_a > c$$

时被拒绝，其中临界值c由

$$Pr_{H_0}(\sup_a T_a > c) = \alpha$$

所确定

定理

- 定理：极大值T_a可以由如下公式给出

$$\sup_{a:\sum a_j=0} T_a^2 = \frac{\sum_{i=1}^k n_i \left((\overline{Y}_{i\cdot} - \overline{\overline{Y}}) - (\theta_i - \overline{\theta}) \right)^2}{S_p^2}$$

$$\overline{\overline{Y}} = \sum n_i \overline{Y}_{i\cdot} / \sum n_i,\ \overline{\theta} = \sum n_i \theta_i / \sum n_i.$$

且在ANOVA零假设下，

$$\sup_{a:\sum a_i=0} T_a^2 \sim (k-1) F_{k-1,N-k}$$

Linear Regression

统计学方法及其应用 > 统计学模型 > 线性回归

ANOVA model

Random variables Y_{ij} are observed according to the model

$$Y_{ij} = \mu + \theta_i + \varepsilon_{ij}, i = 1, \ldots, k, j = 1, \ldots, n_i,$$

where

(i) $\mathrm{E}\varepsilon_{ij} = 0$, $\mathrm{Var}\varepsilon_{ij} = \sigma_i^2 < \infty$, for all i and j.

$\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{st}) = 0$ for all i, j, s, and t unless $i = s$ and $j = t$.

(ii) The ε_{ij} are independent and normally distributed (normal errors).

(iii) $\sigma_i^2 = \sigma^2$ for all i (equality of variance, homoscedasticity).

(iv) $\sum_{i=1}^{k} \theta_i = 0$.

(θ_i: unique effect due to an individual treatment).

(θ_i: unique error due to an individual treatment).

ANOVA data

	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\ldots$	$\boldsymbol{k-1}$	$\boldsymbol{k}$
	Y_{11}	Y_{21}	Y_{31}		$Y_{(k-1)1}$	Y_{k1}
	Y_{12}	Y_{22}	Y_{32}		$Y_{(k-1)2}$	Y_{k2}
	Y_{13}	Y_{23}	Y_{33}		$Y_{(k-1)3}$	Y_{k3}
	$\cdots$	$\cdots$	$\cdots$		$\cdots$	$\cdots$
	Y_{1n_1}				$Y_{(k-1)n_{(k-1)}}$	
			Y_{3n_3}			
		Y_{2n_2}				
						Y_{kn_k}
$\boldsymbol{\mu}$	μ	μ	μ	μ	μ	μ
$\boldsymbol{\theta}$	θ_1	θ_2	θ_3	$\cdots$	θ_{k-1}	θ_k

From another point of view

	0	**1**	**2**	**3**	**…**	$k-1$	k
Y_{11}	**1**	**1**	0	0	0	0	0
$\dots$	**1**	**1**	0	0	0	0	0
Y_{1n1}	**1**	**1**	0	0	0	0	0
Y_{21}	**1**	0	**1**	0	0	0	0
$\dots$	**1**	0	**1**	0	0	0	0
Y_{2n2}	**1**	0	**1**	0	0	0	0
$\dots$	**1**						
Y_{k1}	**1**	0	0	0	0	0	**1**
$\dots$	**1**	0	0	0	0	0	**1**
$Y_{k\ nk}$	**1**	0	0	0	0	0	**1**
$\boldsymbol{\theta}$	$\boldsymbol{\mu}$	θ_1	θ_2	θ_3	$\dots$	θ_{k-1}	θ_k

From another point of view

Random variables Y_{ij} are observed according to the model

$$Y_{1j} = \mu + \theta_1 + 0\times\theta_2 + 0\times\theta_3 + \cdots + 0\times\theta_k + \varepsilon_{1j}, j = 1,\ldots,n_1,$$

$$Y_{2j} = \mu + \theta_2 + 0\times\theta_1 + 0\times\theta_3 + \cdots + 0\times\theta_k + \varepsilon_{1j}, j = 1,\ldots,n_2,$$

$$\cdots$$

$$Y_{kj} = \mu + \theta_k + 0\times\theta_1 + 0\times\theta_2 + \cdots + 0\times\theta_{k-1} + \varepsilon_{1j}, j = 1,\ldots,n_k.$$

$$\Rightarrow \qquad Y_{ij} = \mu + \theta_1 \underbrace{I(i=1)}_{x_1} + \theta_2 \underbrace{I(i=2)}_{x_2} + \cdots + \theta_k \underbrace{I(i=k)}_{x_k} + \varepsilon_{ij}$$

$$\Rightarrow \qquad Y_{ij} = \mu + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k + \varepsilon_{ij}, i = 1,\ldots,k, j = 1,\ldots,n_i$$

$$\Rightarrow \qquad Y_{ij} = \mu + \sum_{i=1}^{k} \theta_i x_i + \varepsilon_{ij}, i = 1,\ldots,k, j = 1,\ldots,n_i, \sum_{i=1}^{k} \theta_i = 1$$

ANOVA data

	0	1	2	$\ldots$	$k-1$	k
Y_1						
Y_2						
Y_3						
$\ldots$	1		**Binary indicators**			
Y_n						
Means	μ	θ_1	θ_2		θ_{k-1}	θ_k

Beyond ANOVA data

	$\mathbf{0}$	$\mathbf{1}$	$\mathbf{2}$	$\ldots$	$\boldsymbol{k-1}$	$\boldsymbol{k}$
Y_1						
Y_2						
Y_3						
$\ldots$	1		**Real numbers**			
Y_n						
$\boldsymbol{\beta}$	β_0	β_1	β_2	$\ldots$	β_{k-1}	β_k

Linear regression model

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i, i = 1, \ldots, n$$

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & & \\ \ldots & & \ddots & \\ 1 & x_{n1} & & x_{nk} \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_k \end{pmatrix} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}$$

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$$

Y_i : ***response*** **variable**; x_i : ***predictor*** **variable**

Regression: representing a relationship between variables

Linear in parameters

Linear:

$$Y_i = \alpha + \beta x_i + \varepsilon_i$$

$$Y_i = \alpha + \beta x_i^2 + \varepsilon_i$$

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i$$

$$\log Y_i = \beta_0 + \beta_1 x_i + \beta_2 (1 / x_i) + \varepsilon_i$$

$$\Phi^{-1}\left[P(Y_i = 1)\right] = \alpha + \beta x_i + \varepsilon_i$$

$$\log \frac{P(Y_i = 1)}{1 - P(Y_i = 1)} = \alpha + \beta x_i + \varepsilon_i$$

Non-linear:

$$Y_i = \alpha + \beta x_i + \beta^2 x_i + \varepsilon_i$$

Simple linear regression

Simple linear regression

$$Y_i = \alpha + \beta x_i + \varepsilon_i, i = 1, \ldots, n$$

Population regression function

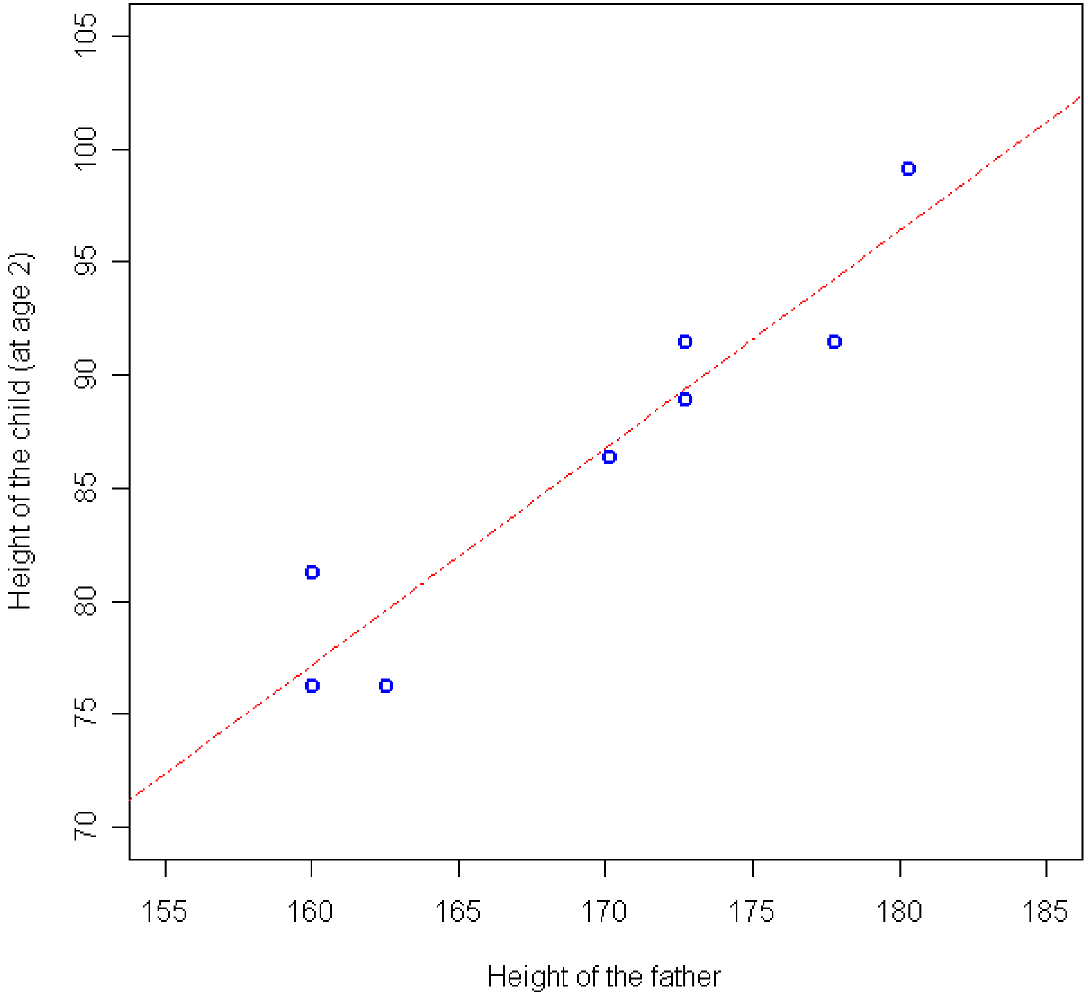
$$\mathrm{E}\, Y_i = \alpha + \beta x_i, i = 1, \ldots, n$$

Y_i : ***response*** **variable**; x_i : ***predictor*** **variable**

Regression: representing a relationship between variables

Linear ⇒ a straight line

Child	Father
99	180
76	160
81	160
86	170
89	173
91	173
91	178
76	163



Regression $\Rightarrow$ go back to the mean

- In the 1800s, **Sir Francis Galton** investigated the relationship between heights of fathers and heights of sons
- He found that tall fathers tend to have tall sons, and short fathers tend to have short sons
- However, he also found that very tall fathers tend to have shorter sons, and very short fathers tend to have taller sons
- Galton called this phenomenon **regression toward the mean (to go back to the mean)**, and from this usage we get the present use of the word **regression**

Two steps to the analysis

- Data fitting step
  - We attempt to summarize the observed data via data fitting
  - We are interested only in the data at hand
  - We do not have to make any assumption about parameters
- Statistical analysis step
  - We attempt to infer conclusions about the relationship in the population
  - We need to make assumptions about the parameters

Data Fitting

统计学方法及其应用 → 方差分析与回归分析 → 线性回归

Regression data

	$\mathbf{1}$	$\boldsymbol{x}$
Y_1		x_1
Y_2		x_2
Y_3		x_3
	$\mathbf{1}$	
$\dots$		**Real numbers**
Y_n		x_n
	$\boldsymbol{a}$	$\boldsymbol{b}$

Sum of squares

Sample means

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Sum of squares

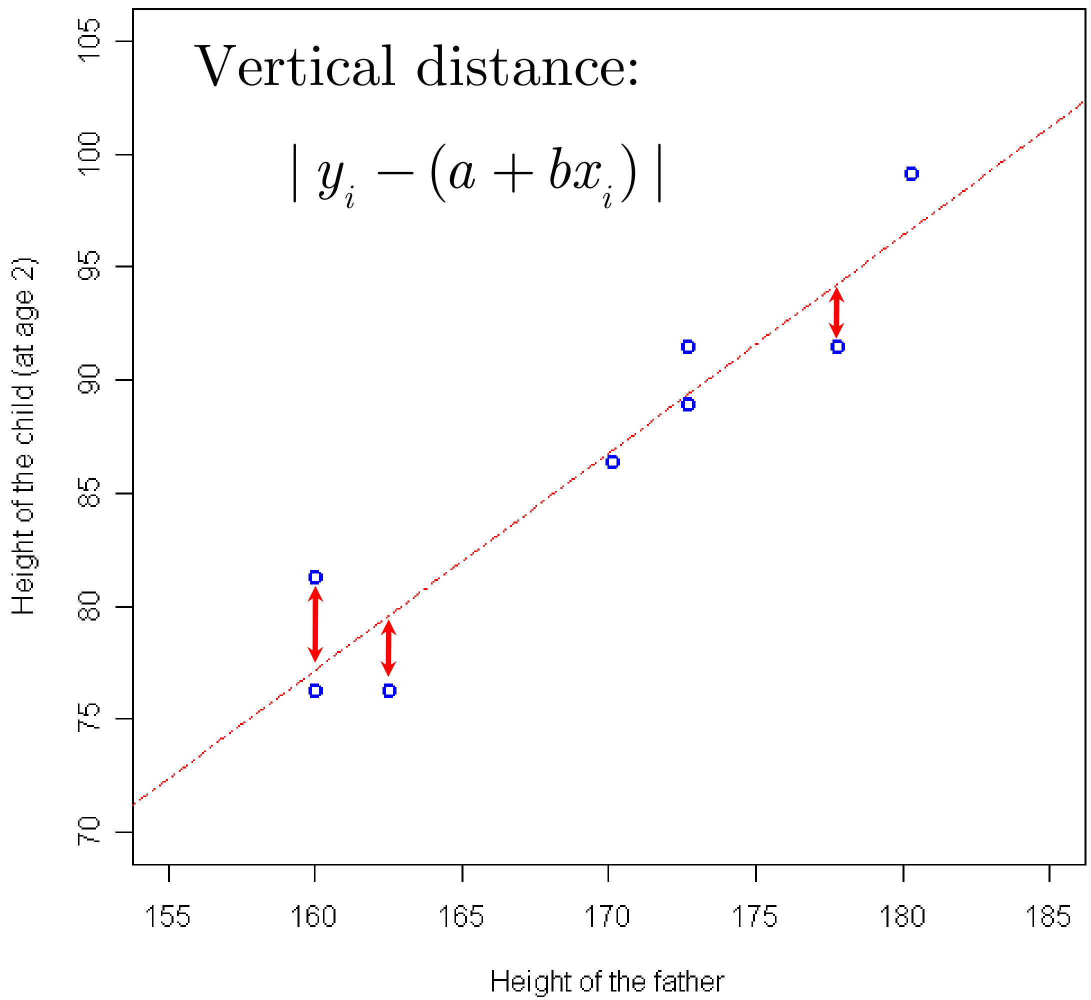
$$S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} (x_i - \overline{x})x_i = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} (y_i - \overline{y})y_i = \sum_{i=1}^{n} y_i^2 - n\overline{y}_i^2$$

Sum of cross-products

$$S_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} (x_i - \overline{x})y_i = \sum_{i=1}^{n} x_i(y_i - \overline{y})$$

Define a distance



Least squares solutions

$$\mathbf{RSS} = \sum_{i=1}^{n} [\boldsymbol{y}_i - (\boldsymbol{a} + \boldsymbol{b}\boldsymbol{x}_i)]^2$$

$$\text{Solve} \quad \min_{a,b} \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$$

$$\sum_{i=1}^{n} [y_i - (a + bx_i)]^2 = \sum_{i=1}^{n} [(y_i - bx_i) - a]^2 \Rightarrow a = \frac{1}{n} \sum_{i=1}^{n} (y_i - bx_i) = \overline{y} - b\overline{x}$$

$$\sum_{i=1}^{n} [(y_i - bx_i) - (\overline{y} - b\overline{x})]^2 = \sum_{i=1}^{n} [(y_i - \overline{y}) - b(x_i - \overline{x})]^2 = b^2 S_{xx} - 2bS_{xy} + S_{yy}$$

$$= S_{xx} \left(b - \frac{S_{xy}}{S_{xx}} \right)^2 + \frac{S_{xx} S_{yy} - S_{xy} S_{xy}}{S_{xx}}$$

$$\Rightarrow b = \frac{S_{xy}}{S_{xx}}, a = \overline{y} - b\overline{x}$$

RSS: Residual Sum of Squares

Property of the sample mean

Let $x_1, \ldots, x_n$ be any numbers and $\overline{x} = (x_1 + \cdots + x_n)/n$.

Then $$\min_a \sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

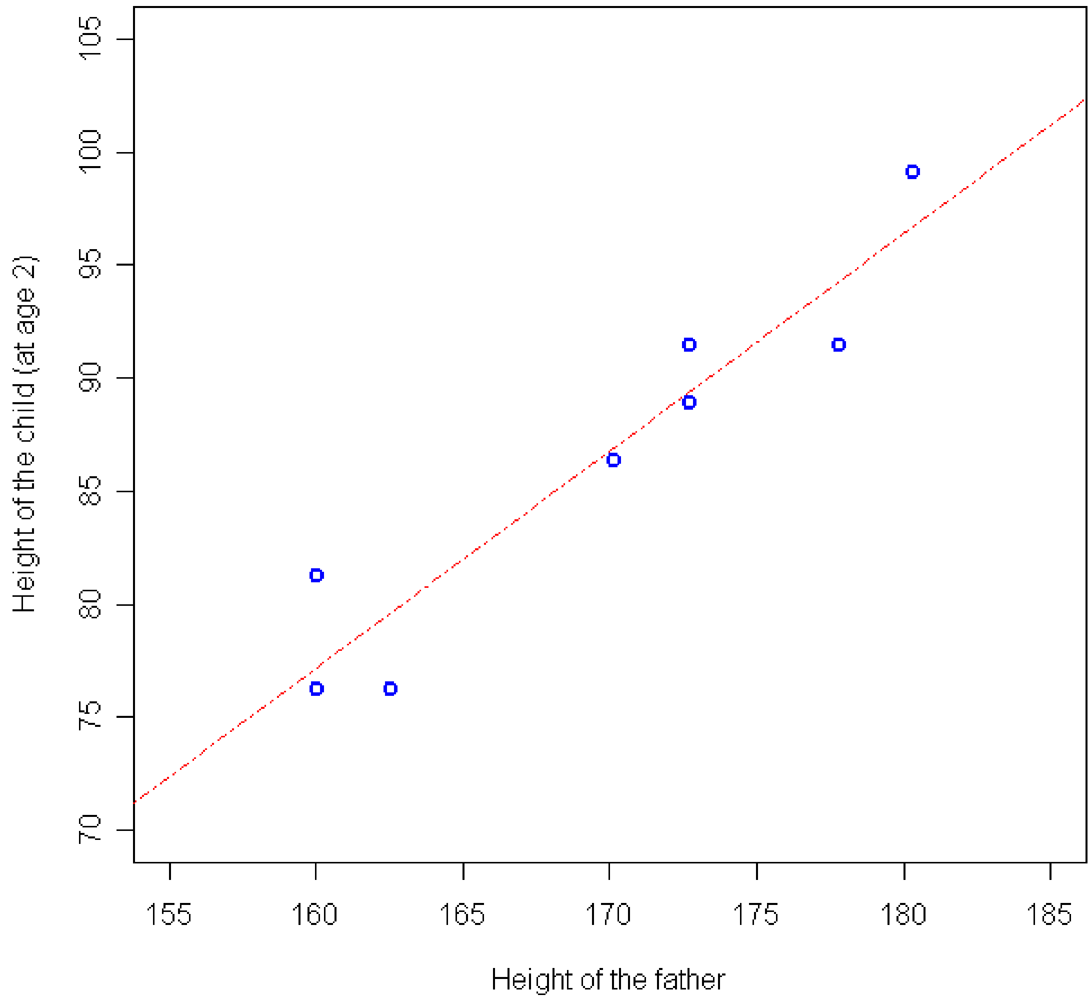
$$\begin{aligned}
\sum_{i=1}^{n} (x_i - a)^2 &= \sum_{i=1}^{n} (x_i - \overline{x} + \overline{x} - a)^2 \\
&= \sum_{i=1}^{n} (x_i - \overline{x})^2 + \sum_{i=1}^{n} (\overline{x} - a)^2 + 2\sum_{i=1}^{n} (x_i - \overline{x})(\overline{x} - a) \\
&= \sum_{i=1}^{n} (x_i - \overline{x})^2 + n(\overline{x} - a)^2 + 2(\overline{x} - a)\underbrace{\sum_{i=1}^{n} (x_i - \overline{x})}_{=0} \\
&= \sum_{i=1}^{n} (x_i - \overline{x})^2 + n(\overline{x} - a)^2
\end{aligned}$$

The minimum of $\sum_{i=1}^{n} (x_i - a)^2$ is $\sum_{i=1}^{n} (x_i - \overline{x})^2$ obtained at $a = \overline{x}$.

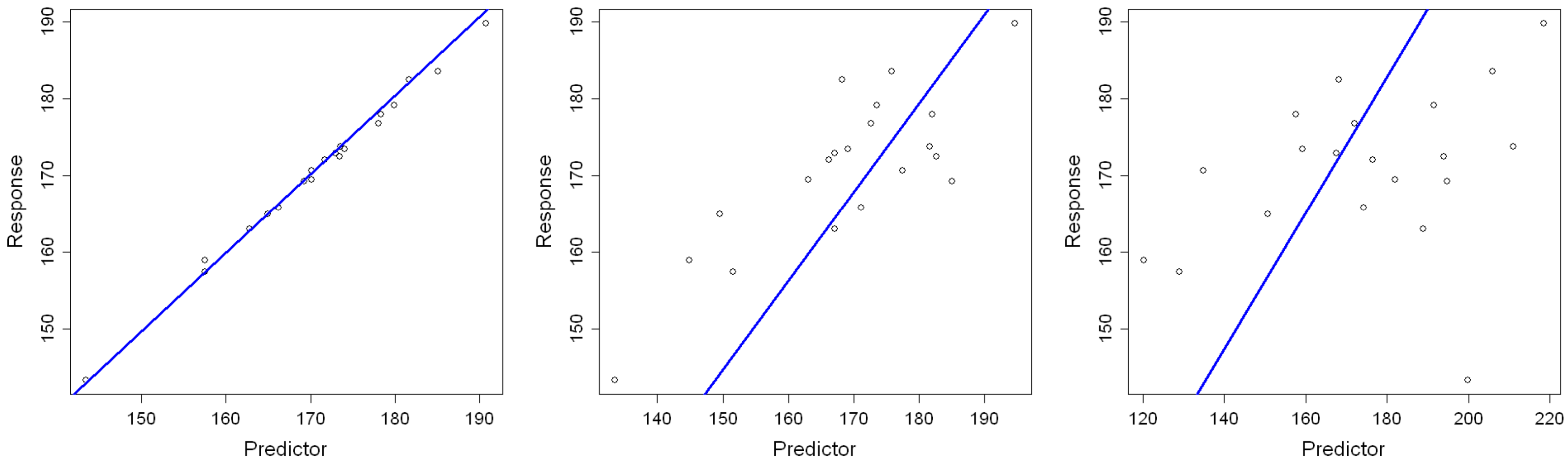
Fitted straight line

Child	Father
99	180
76	160
81	160
86	170
89	173
91	173
91	178
76	163

$$\begin{aligned}
\bar{x} &= 169.6 \\
\bar{y} &= 86.4 \\
S_{xx} &= 435.5 \\
S_{xy} &= 419.4 \\
S_{yy} &= 451.6 \\
a &= -76.9 \\
b &= 0.963 \\
a &: \text{Intercept} \\
b &: \text{Slope}
\end{aligned}$$



Fitted lines




```
w <- rnorm(20, 170, 10);
x <- rnorm(20, w,    1); #10, 20
m <- lm(x~w);
plot(x, w, xlab="Predictor", ylab="Response");
abline(m$coefficient[1], m$coefficient[2], lwd=3, col="blue");
```

Statistical Analysis

统计学方法及其应用 → 方差分析与回归分析 → 线性回归

Linear estimators

$$\begin{aligned} Y_i &= \alpha + \beta x_i + \varepsilon_i, && i = 1, \ldots, n \\ \mathrm{E}\, Y_i &= \alpha + \beta x_i && (\mathrm{E}\varepsilon_i = 0) \\ \mathrm{Var}\, Y_i &= \sigma^2 && (\mathrm{Var}\varepsilon_i = \sigma^2) \end{aligned}$$

We attempt to find estimators (functions of $\mathbf{Y}$) of α and β

Now, restrict our attention to the class of **linear estimators**, say,

$$\hat{\alpha} = \sum_{i=1}^{n} \xi_i Y_i \quad \text{and} \quad \hat{\beta} = \sum_{i=1}^{n} \delta_i Y_i,$$

where ξ_i and δ_i are known, fixed constants.

We are interested in **unbiased and minumum variance** estimators.

Unbiased estimator of the slope

$$\hat{\beta} = \frac{S_{xY}}{S_{xx}} \qquad \hat{\beta} = \sum_{i=1}^{n} \delta_i Y_i$$

$$\mathrm{E}\hat{\beta} = \mathrm{E}\left[\sum_{i=1}^{n} \delta_i Y_i\right] = \sum_{i=1}^{n} \delta_i \mathrm{E}\, Y_i = \sum_{i=1}^{n} \delta_i (\alpha + \beta x_i) = \alpha \sum_{i=1}^{n} \delta_i + \beta \sum_{i=1}^{n} \delta_i x_i$$

$\hat{\beta}$ is unbiased if and only if $\sum_{i=1}^{n} \delta_i = 0$ and $\sum_{i=1}^{n} \delta_i x_i = 1$.

Now, consider $\delta_i = \dfrac{x_i - \bar{x}}{S_{xx}}$

$$\sum_{i=1}^{n} \delta_i = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{S_{xx}} = \frac{1}{S_{xx}} \sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

$$\sum_{i=1}^{n} \delta_i x_i = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{S_{xx}} x_i = \frac{1}{S_{xx}} \left[\sum_{i=1}^{n} (x_i - \bar{x}) x_i + \sum_{i=1}^{n} (x_i - \bar{x}) \bar{x}\right] = \frac{1}{S_{xx}} \sum_{i=1}^{n} (x_i - \bar{x})^2 = 1$$

Therefore, $\hat{\beta} = \sum_{i=1}^{n} \delta_i Y_i = \sum_{i=1}^{n} \dfrac{x_i - \bar{x}}{S_{xx}} Y_i = \dfrac{1}{S_{xx}} \sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y}) = \dfrac{S_{xY}}{S_{xx}}$ is unbiased.

Unbiased estimator of the intercept

$$\hat{\boldsymbol{\alpha}} = \bar{\boldsymbol{Y}} - \hat{\boldsymbol{\beta}}\bar{\boldsymbol{x}} \qquad \hat{\alpha} = \sum_{i=1}^{n} \xi_i Y_i$$

$$\mathrm{E}\hat{\alpha} = \mathrm{E}\left(\sum_{i=1}^{n} \xi_i Y_i\right) = \sum_{i=1}^{n} \xi_i \mathrm{E}\, Y_i = \sum_{i=1}^{n} \xi_i (\alpha + \beta x_i) = \alpha \sum_{i=1}^{n} \xi_i + \beta \sum_{i=1}^{n} \xi_i x_i$$

$\hat{\alpha}$ is unbiased if and only if $\sum_{i=1}^{n} \xi_i = 1$ and $\sum_{i=1}^{n} \xi_i x_i = 0$

Now, consider $\xi_i = \frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}}$

$$\sum_{i=1}^{n} \xi_i = \sum_{i=1}^{n} \left[\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}}\right] = 1 - \frac{1}{S_{xx}} \sum_{i=1}^{n} (x_i - \bar{x})\bar{x} = 1$$

$$\sum_{i=1}^{n} \xi_i x_i = \sum_{i=1}^{n} \left[\frac{1}{n} x_i - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}} x_i\right] = \bar{x} - \frac{\bar{x}}{S_{xx}} \sum_{i=1}^{n} (x_i - \bar{x}) x_i = 0$$

Therefore, $\hat{\alpha} = \sum_{i=1}^{n} \xi_i Y_i = \bar{Y} - \hat{\beta}\bar{x}$ is unbiased.

Best linear unbiased estimator (BLUE)

$$\hat{\boldsymbol{\beta}} = \frac{\boldsymbol{S}_{xY}}{\boldsymbol{S}_{xx}}$$

$$\hat{\beta} = \sum_{i=1}^{n} \delta_i Y_i, \ \ \delta_i = \frac{x_i - \overline{x}}{S_{xx}}$$

$$\mathrm{Var}\hat{\beta} = \mathrm{Var}\left(\sum_{i=1}^{n} \delta_i Y_i\right) = \sum_{i=1}^{n} \delta_i^2 \mathrm{Var}\, Y_i = \sum_{i=1}^{n} \delta_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^{n} \delta_i^2 = \frac{\sigma^2}{S_{xx}}$$

$$\text{because } \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{S_{xx}}\right)^2 = \frac{S_{xx}}{S_{xx} S_{xx}} = \frac{1}{S_{xx}}$$

It can be proved that $\mathrm{Var}\hat{\beta}$ is the minimum.

Therefore, $\hat{\beta}$ is the best linear unbiased estimator (BLUE) of β.

Best linear unbiased estimator (BLUE)

$$\hat{\boldsymbol{\alpha}} = \bar{\boldsymbol{Y}} - \hat{\boldsymbol{\beta}}\bar{\boldsymbol{x}}$$

$$\hat{\alpha} = \sum_{i=1}^{n} \xi_i Y_i, \ \ \xi_i = \frac{1}{n} - \frac{(x_i - \overline{x})\overline{x}}{S_{xx}}$$

$$\mathrm{Var}\hat{\alpha} = \mathrm{Var}\left(\sum_{i=1}^{n} \xi_i Y_i\right) = \sum_{i=1}^{n} \xi_i^2 \mathrm{Var}\, Y_i = \sum_{i=1}^{n} \xi_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^{n} \xi_i^2 = \sigma^2 \left(\frac{1}{nS_{xx}} \sum_{i=1}^{n} x_i^2\right)$$

$$\sum_{i=1}^{n} \xi_i^2 = \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{(x_i - \overline{x})\overline{x}}{S_{xx}}\right)^2 = \sum_{i=1}^{n} \frac{1}{n^2} + \sum_{i=1}^{n} \left(\frac{(x_i - \overline{x})\overline{x}}{S_{xx}}\right)^2 = \frac{1}{n} + \frac{\overline{x}^2}{S_{xx}} = \frac{1}{nS_{xx}} \sum_{i=1}^{n} x_i^2$$

It can be proved that $\mathrm{Var}\hat{\alpha}$ is the minimum.

Therefore, $\hat{\alpha}$ is the best linear unbiased estimator (BLUE) of α.

Conditional normal model

$$Y_i = \alpha + \beta x_i + \varepsilon_i, i = 1, \ldots, n$$

Assume that

(i) $\mathrm{E}\varepsilon_i = 0,\ \mathrm{Var}\varepsilon_i = \sigma_i^2 < \infty,$ for all.

$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for all i and j unless $i = j$.

(ii) The ε_i are independent and normally distributed (normal errors).

(iii) $\sigma_i^2 = \sigma^2$ for all i (equality of variance, homoscedasticity).

(iv) Y_i independent (but not identically distributed), $i = 1, \ldots, n$.

(iv) x_i known and fixed (not random variables), $i = 1, \ldots, n$.

It then follows that

$$\begin{aligned} Y_i \mid x_i &\sim \mathrm{N}(\alpha + \beta x_i, \sigma^2) \\ \mathrm{E}\,Y_i &= \alpha + \beta x_i \\ \mathrm{Var}\,Y_i &= \sigma^2 \end{aligned}$$

Bivariate normal model

$$Y_i = \alpha + \beta X_i, i = 1, \ldots, n$$

Assume that (X_i, Y_i) is bivariate normal $(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, $i = 1, \ldots, n$. That is

$$f(x, y \mid \mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times$$
$$\exp\left\{-\frac{1}{2\left(1-\rho^2\right)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}.$$

It follows that,

$$\begin{aligned} Y_i \mid x_i &\sim \mathrm{N}(\mu_Y + \rho\sigma_Y / \sigma_X(x_i - \mu_X), \sigma_Y^2(1-\rho^2)) \\ \mathrm{E}(Y_i \mid x_i) &= \mu_Y + \rho\sigma_Y / \sigma_X(x_i - \mu_X) = (\mu_Y - \rho(\sigma_Y / \sigma_X)\mu_X) + \rho(\sigma_Y / \sigma_X)x_i \\ \mathrm{Var}\, Y_i &= \sigma_Y^2(1-\rho^2) \end{aligned}$$

Point estimation

$$
\begin{aligned}
L(\alpha,\beta,\sigma^2 \mid \mathbf{x},\mathbf{y}) &= f(\mathbf{y} \mid \alpha,\beta,\sigma^2,\mathbf{x}) \\
&= \prod_{i=1}^{n} f_i(y_i \mid \alpha,\beta,\sigma^2,x_i) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{\frac{[y_i-(\alpha+\beta x_i)]^2}{2\sigma^2}\right\} \\
&= (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp\left\{\frac{\sum_{i=1}^{n}[y_i-(\alpha+\beta x_i)]^2}{2\sigma^2}\right\}
\end{aligned}
$$

$$
\begin{aligned}
l(\alpha,\beta,\sigma^2 \mid \mathbf{x},\mathbf{y}) &= \log L(\alpha,\beta,\sigma^2 \mid \mathbf{x},\mathbf{y}) \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}(\sigma^2)^{-1}\sum_{i=1}^{n}[y_i-(\alpha+\beta x_i)]^2
\end{aligned}
$$

Maximum likelihood estimates (MLE)

$$\max \quad -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}(\sigma^2)^{-1}\sum_{i=1}^{n}\left[y_i - (\alpha + \beta x_i)\right]^2$$

$$\begin{aligned}
\frac{\partial}{\partial \beta} l(\alpha, \beta, \sigma^2 \mid \mathbf{x}, \mathbf{y}) &= \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(y_i - \overline{y}) - \beta(x_i - \overline{x})\right]x_i = 0 \\
&\Rightarrow \hat{\beta} = S_{xy} \,/\, S_{xx} \\
\frac{\partial}{\partial \alpha} l(\alpha, \beta, \sigma^2 \mid \mathbf{x}, \mathbf{y}) &= \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(y_i - \beta x_i) - \alpha\right] = 0 \\
&\Rightarrow \hat{\alpha} = \overline{y} - \hat{\beta}\overline{x} \\
\frac{\partial}{\partial \sigma^2} l(\alpha, \beta, \sigma^2 \mid \mathbf{x}, \mathbf{y}) &= -\frac{n}{2}(\sigma^2)^{-1} + \frac{1}{2}(\sigma^2)^{-2}\sum_{i=1}^{n}\left[y_i - (\alpha + \beta x_i)\right]^2 = 0 \\
&\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - (\hat{\alpha} + \hat{\beta} x_i)\right]^2
\end{aligned}$$

$$\sum_{i=1}^{n}(y_i - \overline{y})\overline{x} = 0, \;\; \sum_{i=1}^{n}(x_i - \overline{x})\overline{x} = 0$$

Biased estimator of the variance

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}[Y_i^2 - (\alpha + \hat{\beta}x_i)]^2$$

$$\widehat{\varepsilon}_i = Y_i - (\widehat{\alpha} + \widehat{\beta}x_i), \ \text{(residuals from the regression)}$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_i^2 \ \Rightarrow \ \mathrm{E}\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(\widehat{\varepsilon}_i^2) = \frac{n-2}{n}\sigma^2$$

$$\mathrm{E}\widehat{\varepsilon}_i = \mathrm{E}\left[Y_i - (\widehat{\alpha} + \widehat{\beta}x_i)\right] = \mathrm{E}\,Y_i - (\mathrm{E}\widehat{\alpha} + x_i\mathrm{E}\widehat{\beta}) = (\alpha + \beta x_i) - (\alpha + \beta x_i) = 0$$

$$\begin{aligned}
\mathrm{Var}\widehat{\varepsilon}_i &= \mathrm{E}(\widehat{\varepsilon}_i^2) = \mathrm{E}\left[Y_i - \widehat{\alpha} - \widehat{\beta}x_i\right]^2 \\
&= \mathrm{E}\left[(Y_i - \alpha - \beta x_i) - (\widehat{\alpha} - \alpha) - (\widehat{\beta} - \beta)x_i\right]^2 \\
&= \mathrm{E}\left[(Y_i - \alpha - \beta x_i)^2\right] + \mathrm{E}\left[(\widehat{\alpha} - \alpha)^2\right] + \mathrm{E}\left[(\widehat{\beta} - \beta)x_i\right]^2 + 2\mathrm{E}\left[(\widehat{\alpha} - \alpha)(\widehat{\beta} - \beta)x_i\right] \\
&\quad - 2\mathrm{E}\left[(Y_i - \alpha - \beta x_i)(\widehat{\alpha} - \alpha)\right] - 2\mathrm{E}\left[(Y_i - \alpha - \beta x_i)(\widehat{\beta} - \beta)x_i\right] \\
&= \mathrm{Var}\,Y_i + \mathrm{Var}\widehat{\alpha} + x_i^2\mathrm{Var}\widehat{\beta} - 2\mathrm{Cov}(Y_i, \widehat{\alpha}) - 2x_i\mathrm{Cov}(Y_i, \widehat{\beta}) + 2x_i\mathrm{Cov}(\widehat{\alpha}, \widehat{\beta})
\end{aligned}$$

Unbiased estimator of the variance

$$S^2 = \frac{n}{n-2}\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}[y_i - (\alpha + \hat{\beta}x_i)]^2 = \frac{1}{n-2}\sum_{i=1}^{n}\epsilon_i^2$$

Because

$$\mathrm{E}[\hat{\sigma}^2] = \frac{n}{n-2}\sigma^2$$

We have

$$\mathrm{E}[S^2] = \sigma^2$$

Recall biased and unbiased estimators for the normal variance

Sampling distribution of the slope

$$\hat{\boldsymbol{\beta}} = \frac{\boldsymbol{S}_{xY}}{\boldsymbol{S}_{xx}} \sim \mathbf{N}\left(\boldsymbol{\beta}, \frac{\boldsymbol{\sigma}^2}{\boldsymbol{S}_{xx}}\right)$$

$$\hat{\beta} = \sum_{i=1}^{n} \delta_i Y_i, \delta_i = \frac{x_i - \overline{x}}{S_{xx}}, E(\hat{\beta}) = \beta, Var(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$$

Y_i is normally distributed, therefore the linear combination $\hat{\beta} = \sum_{i=1}^{n} \delta_i Y_i$ is also normally distributed. In other words, $\hat{\beta}$ has a normal distribution

$$\Rightarrow \qquad \hat{\beta} \sim \mathrm{N}(\beta, \sigma^2 / S_{xx}) \text{ or } \frac{\hat{\beta} - \beta}{\sqrt{\sigma^2 / S_{xx}}} \sim \mathrm{N}(0,1)$$

Sampling distribution of the intercept

$$\boldsymbol{\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x} \sim \mathrm{N}\left(\alpha, \frac{\sigma^2}{nS_{xx}}\sum_{i=1}^{n} x_i^2\right)}$$

$$\hat{\alpha} = \sum_{i=1}^{n} \xi_i Y_i, \xi_i = \frac{1}{n} - \frac{(x_i - \overline{x})\overline{x}}{S_{xx}}, E(\hat{\alpha}) = \alpha, Var(\hat{\alpha}) = \frac{\sigma^2}{nS_{xx}}\sum_{i=1}^{n} x_i^2$$

Y_i is normally distributed, therefore the linear combination $\hat{\alpha} = \sum_{i=1}^{n} \xi_i Y_i$ is also normally distributed. In other words, $\hat{\alpha}$ has a normal distribution

$$\Rightarrow \qquad \hat{\alpha} \sim \mathrm{N}\left(\alpha, \frac{\sigma^2}{nS_{xx}}\sum_{i=1}^{n} x_i^2\right) \text{ or } \frac{\hat{\alpha} - \alpha}{\sqrt{\sigma^2 \left(\sum_{i=1}^{n} x_i^2\right) / \left(nS_{xx}\right)}} \sim \mathrm{N}(0,1)$$

Covariance of the intercept and slope

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{x}, \quad \hat{\beta} = \frac{S_{xY}}{S_{xx}}$$

$$\operatorname{cov}(\hat{\alpha}, \hat{\beta}) = \operatorname{cov}\left(\sum_{i=1}^{n} \xi_i Y_i, \sum_{i=1}^{n} \delta_i Y_i\right) = \sum_{i=1}^{n} \xi_i \delta_i \operatorname{Var} Y_i = \sigma^2 \sum_{i=1}^{n} \xi_i \delta_i$$

$$\begin{aligned}
\sum_{i=1}^{n} \xi_i \delta_i &= \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{(x_i - \overline{x})\overline{x}}{S_{xx}}\right)\left(\frac{x_i - \overline{x}}{S_{xx}}\right) \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{S_{xx}}\right) - \frac{\overline{x}}{S_{xx}} \sum_{i=1}^{n} \left(\frac{(x_i - \overline{x})(x_i - \overline{x})}{S_{xx}}\right) \\
&= -\frac{\overline{x}}{S_{xx}}
\end{aligned}$$

$$\Rightarrow \qquad \operatorname{cov}(\hat{\alpha}, \hat{\beta}) = -\frac{\sigma^2 \overline{x}}{S_{xx}}$$

Sampling distribution of the variance

$$S^2 = \frac{n}{n-2}\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left[Y_i - (\hat{\alpha} + \hat{\beta}x_i)\right]^2 = \frac{1}{n-2}\sum_{i=1}^{n}\varepsilon_i^2$$

(1) $(\hat{\alpha}, \hat{\beta})$ **and** S^2 **are independent**

(2) $$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$$

$(n-2)S^2 = \sum_{i=1}^{n}\varepsilon_i^2$, **Residual sum of squares**

Sampling distribution of the intercept

$$\widehat{\alpha} = \overline{Y} - \widehat{\beta}\overline{x}$$

$$\frac{\widehat{\alpha} - \alpha}{\sqrt{\sigma^2 (\sum_{i=1}^{n} x_i^2)/(nS_{xx})}} \sim \mathrm{N}(0,1)$$

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2} \Rightarrow$$

$$\frac{\widehat{\alpha} - \alpha}{\sqrt{S^2 (\sum_{i=1}^{n} x_i^2)/(nS_{xx})}} = \frac{\dfrac{\widehat{\alpha} - \alpha}{\sqrt{\sigma^2 (\sum_{i=1}^{n} x_i^2)/(nS_{xx})}}}{\sqrt{\dfrac{(n-2)S^2/\sigma^2}{n-2}}} = \frac{\mathrm{N}(0,1)}{\sqrt{\chi^2_{n-2}/(n-2)}} \sim T_{n-2}$$

$$\frac{(\widehat{\alpha} - \alpha)^2}{S^2 (\sum_{i=1}^{n} x_i^2)/(nS_{xx})} = \left(\frac{\widehat{\alpha} - \alpha}{\sqrt{S^2 (\sum_{i=1}^{n} x_i^2)/(nS_{xx})}} \right)^2 = \frac{\chi^2_1/1}{\chi^2_{n-2}/(n-2)} \sim F_{1,n-2}$$

$$\frac{\mathrm{N}(0,1)}{\sqrt{\chi^2_q/q}} \sim t_q, \left(\frac{\mathrm{N}(0,1)}{\sqrt{\chi^2_q/q}} \right)^2 = \frac{\mathrm{N}(0,1)^2}{\chi^2_q/q} = \frac{\chi^2_1/1}{\chi^2_q/q} \sim F_{1,q}$$

Sampling distribution of the slope

$$\widehat{\beta} = \frac{S_{xY}}{S_{xx}}$$

$$\frac{\widehat{\beta} - \beta}{\sqrt{\sigma^2 / S_{xx}}} \sim \mathrm{N}(0,1)$$

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2} \Rightarrow$$

$$\frac{\widehat{\beta} - \beta}{\sqrt{S^2 / S_{xx}}} = \frac{\dfrac{\widehat{\beta} - \beta}{\sqrt{\sigma^2 / S_{xx}}}}{\sqrt{\dfrac{(n-2)S^2 / \sigma^2}{n-2}}} = \frac{\mathrm{N}(0,1)}{\sqrt{\chi^2_{n-2} / (n-2)}} \sim T_{n-2}$$

$$\frac{\left(\widehat{\beta} - \beta\right)^2}{S^2 / S_{xx}} = \left(\frac{\widehat{\beta} - \beta}{\sqrt{S^2 / S_{xx}}}\right)^2 = \frac{\chi^2_1 / 1}{\chi^2_{n-2} / (n-2)} \sim F_{1,n-2}$$

Hypothesis testing of the intercept

$$H_0 : \alpha = \alpha_0 \quad \text{versus} \quad H_1 : \alpha \neq \alpha_0$$

Since

$$\frac{\hat{\alpha} - \alpha}{\sqrt{S^2 \left(\sum_{i=1}^{n} x_i^2\right) / (nS_{xx})}} \sim T_{n-2}$$

We could reject H_0 at level ρ if and only if

$$\frac{|\hat{\alpha} - \alpha_0|}{\sqrt{S^2 \left(\sum_{i=1}^{n} x_i^2\right) / (nS_{xx})}} > t_{n-2,\rho/2}$$

p-value is

$$p = 2P\left(T_{n-2} \geq \frac{|\hat{\alpha} - \alpha_0|}{\sqrt{S^2 \left(\sum_{i=1}^{n} x_i^2\right) / (nS_{xx})}}\right)$$

Hypothesis testing of the intercept

$$H_0 : \alpha = \alpha_0 \quad \text{versus} \quad H_1 : \alpha \neq \alpha_0$$

Since

$$\frac{(\hat{\alpha} - \alpha)^2}{S^2 \left(\sum_{i=1}^{n} x_i^2\right) / (nS_{xx})} \sim F_{1,n-2}$$

We could reject H_0 at level ρ if and only if

$$\frac{(\hat{\alpha} - \alpha_0)^2}{S^2 \left(\sum_{i=1}^{n} x_i^2\right) / (nS_{xx})} > F_{1,n-2,\rho}$$

p-value is

$$p = P\left(F_{1,n-2} \geq \frac{(\hat{\alpha} - \alpha_0)^2}{S^2 \left(\sum_{i=1}^{n} x_i^2\right) / (nS_{xx})}\right)$$

Hypothesis testing of the slope

$$H_0 : \beta = \beta_0 \quad \text{versus} \quad H_1 : \beta \neq \beta_0$$

Since

$$\frac{\hat{\beta} - \beta}{\sqrt{S^2 / S_{xx}}} \sim T_{n-2}$$

We could reject H_0 at level ρ if and only if

$$\frac{|\hat{\beta} - \beta_0|}{\sqrt{S^2 / S_{xx}}} > t_{n-2,\rho/2}$$

p-value is

$$p = 2P\left(T_{n-2} \geq \frac{|\hat{\beta} - \beta_0|}{\sqrt{S^2 / S_{xx}}}\right)$$

Hypothesis testing of the slope

$$H_0 : \beta = \beta_0 \quad \text{versus} \quad H_1 : \beta \neq \beta_0$$

Since

$$\frac{(\hat{\beta} - \beta)^2}{S^2 / S_{xx}} \sim F_{1,n-2}$$

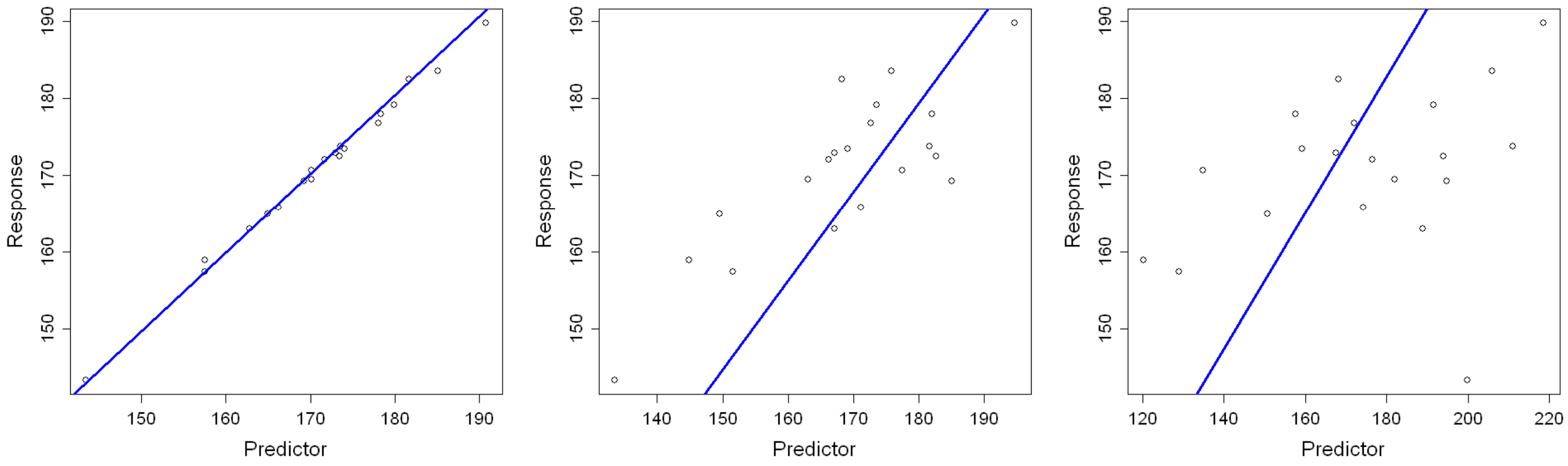
We could reject H_0 at level ρ if and only if

$$\frac{(\hat{\beta} - \beta_0)^2}{S^2 / S_{xx}} > F_{1,n-2,\rho}$$

p-value is

$$p = P\left(F_{1,n-2} \geq \frac{(\hat{\beta} - \beta_0)^2}{S^2 / S_{xx}} \right)$$

Examples of hypothesis testing




```
summary(lm(x~w))

Residuals:
    Min      1Q  Median      3Q     Max
-43.853 -18.201   2.282  20.020  49.410

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  23.0179    99.3394   0.232    0.819
w             0.8885     0.5804   1.531    0.143

Residual standard error: 26.24 on 18 degrees of freedom
Multiple R-Squared: 0.1152,     Adjusted R-squared: 0.06603
F-statistic: 2.343 on 1 and 18 DF,  p-value: 0.1432
```

Hypothesis testing of the slope

$$H_0 : \beta = 0 \quad \text{versus} \quad H_1 : \beta \neq 0$$

We could reject H_0 at level ρ if and only if

$$\frac{\hat{\beta}^2}{S^2 / S_{xx}} > F_{1,n-2,\rho}$$

The test statistic

$$\frac{\hat{\beta}^2}{S^2 / S_{xx}} = \frac{(S_{xy} / S_{xx})^2}{S^2 / S_{xx}} = \frac{S_{xy}^2 / S_{xx}}{(n-2)S^2 / (n-2)}$$

S_{xy}^2 / S_{xx} : Regression sum of squares

$(n-2)S^2 = \sum_{i=1}^{n} \hat{\varepsilon}_i^2$: Residual sum of squares

$n-2$: Degree of freedom

Partitioning of the sum of squares

$$\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

SST = Reg.SS + RSS

$$\begin{aligned}
\sum_{i=1}^{n}(y_i-\overline{y})^2 &= \sum_{i=1}^{n}(y_i-\hat{y}_i+\hat{y}_i-\overline{y})^2 \\
&= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2+2\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\overline{y}) \\
\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\overline{y}) &= \sum_{i=1}^{n}[y_i-(\hat{\alpha}+\hat{\beta}x_i)][(\hat{\alpha}+\hat{\beta}x_i)-\overline{y}] \\
&= \sum_{i=1}^{n}[y_i-(\overline{y}-\hat{\beta}\overline{x}+\hat{\beta}x_i)][(\overline{y}-\hat{\beta}\overline{x}+\hat{\beta}x_i)-\overline{y}] \\
&= \sum_{i=1}^{n}[(y_i-\overline{y})-\hat{\beta}(x_i-\overline{x})][\hat{\beta}(x_i-\overline{x})] \\
&= \hat{\beta}\sum_{i=1}^{n}(y_i-\overline{y})(x_i-\overline{x})-\hat{\beta}^2\sum_{i=1}^{n}(x_i-\overline{x})^2 \\
&= (S_{xy}/S_{xx})S_{xy}-(S_{xy}/S_{xx})^2S_{xx}=0
\end{aligned}$$

Simple linear regression ANOVA table

Source of variation	Degrees of freedom	Sum of squares	Mean square	*F* statistic	*p* value
Regression (slope)	$\mathbf{1}$	$\mathbf{SS(Reg.)}$ $\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$	$\mathbf{MS(Reg.)}$ $\frac{S_{xy}^2}{S_{xx}}$	$\boldsymbol{F} = \frac{\mathbf{MS(Reg.)}}{\mathbf{MS(Res.)}}$	$1 - F_{1,n-2}(F)$
Residual	$\boldsymbol{n} - \mathbf{2}$	$\mathbf{SS(Res.)}\ \mathbf{RSS}$ $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$\mathbf{MS(Res.)}$ $\frac{\mathrm{RSS}}{n-2}$		
Total	$\boldsymbol{n} - \mathbf{1}$	$\mathbf{SST}$ $\sum_{i=1}^{n}(y_i - \overline{y})^2$			

$$H_0 : \beta = 0 \quad \text{versus} \quad H_1 : \beta \neq 0$$

Coefficient of determination

$$r^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{\textbf{Regression sum of squares}}{\textbf{Total sum of squares}}$$

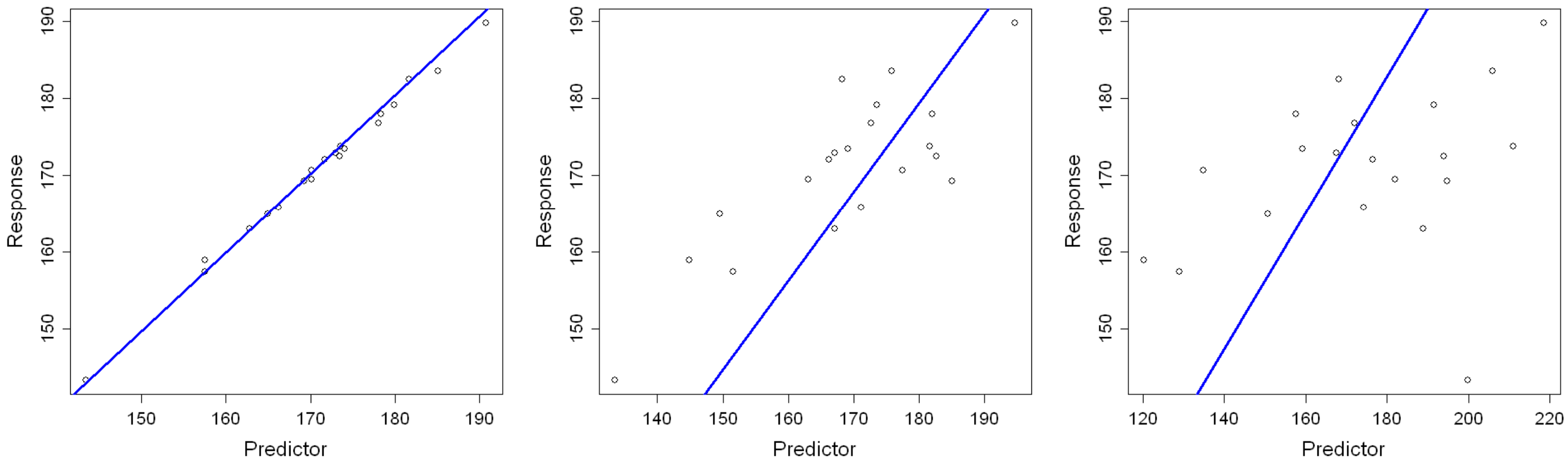
$$0 \le r^2 \le 1$$

$r^2 = 1$: all y_is are on the fitted line.

$r^2 \approx 0$: all y_is are far away from the fitted line.

Measure the proportion of the total variation in $\mathbf{y}$ that is explained by the fitted line.

Examples of hypothesis testing




```
summary(lm(x~w))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -28.8807    33.6304  -0.859    0.402
w             1.1573     0.1965   5.890 1.41e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.882 on 18 degrees of freedom
Multiple R-Squared: 0.6584,     Adjusted R-squared: 0.6394
F-statistic: 34.69 on 1 and 18 DF,  p-value: 1.414e-05
```

Interval estimation of the slope

Inverting the test

$$H_0 : \beta = \beta_0 \quad \text{versus} \quad H_1 : \beta \neq \beta_0$$

We have a $1-\rho$ confidence interval for the slope

$$\hat{\beta} - t_{n-2,\rho/2}\sqrt{S^2 / S_{xx}} \leq \beta \leq \hat{\beta} + t_{n-2,\rho/2}\sqrt{S^2 / S_{xx}}$$

How about using pivotal quantities?

Interval estimation of the intercept

Inverting the test

$$H_0 : \alpha = \alpha_0 \quad \text{versus} \quad H_1 : \alpha \neq \alpha_0$$

We have a $1-\rho$ confidence interval for the intercept

$$\widehat{\alpha} - t_{n-2,\rho/2}\sqrt{S^2\left(\sum_{i=1}^{n} x_i^2\right)/(nS_{xx})} \leq \alpha \leq \widehat{\alpha} + t_{n-2,\rho/2}\sqrt{S^2\left(\sum_{i=1}^{n} x_i^2\right)/(nS_{xx})}$$

How about pivoting a CDF?

Summary

Observation:

$$(y_1, x_1), \ldots, (y_n, x_n)$$

Model:

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \varepsilon_i \sim \mathrm{N}(0, \sigma^2), i = 1, \ldots, n$$

Parameter:

$$\alpha, \beta, \text{ and } \sigma^2$$

Point estimation:

$$\hat{\alpha}, \hat{\beta}, \text{ and } S^2$$

Hypothesis testing:

$$H_0 : \alpha = \alpha_0 \text{ versus } H_1 : \alpha \neq \alpha_0$$

$$H_0 : \beta = \beta_0 \text{ versus } H_1 : \beta \neq \beta_0$$

Interval estimation:

$$\hat{\beta} - t_{n-2,\rho/2}\sqrt{S^2/S_{xx}} \le \beta \le \hat{\beta} + t_{n-2,\rho/2}\sqrt{S^2/S_{xx}}$$

$$\hat{\alpha} - t_{n-2,\rho/2}\sqrt{S^2(\sum_{i=1}^{n} x_i^2)/(nS_{xx})} \le \alpha \le \hat{\alpha} + t_{n-2,\rho/2}\sqrt{S^2(\sum_{i=1}^{n} x_i^2)/(nS_{xx})}$$

Prediction at a single point

Given a new observation $x = x_0$, we have a random variable

$$Y_0 = \alpha + \beta x_0 + \varepsilon_0$$

with the mean being

$$\mathrm{E}(Y_0 \mid x_0) = \alpha + \beta x_0$$

What's the distribution of the estimator

$$\hat{\mu}_{Y_0} = \hat{\alpha} + \hat{\beta} x_0$$

What's the distribution of the random variable

$$Y_0 = \hat{\alpha} + \hat{\beta} x_0 + \varepsilon_0$$

Estimation at a single point

Obviously, $\hat{\mu}_{Y_0} = \hat{\alpha} + \hat{\beta} x_0$ has a normal distribution

$$E\hat{\mu}_{Y_0} = E(\hat{\alpha} + \hat{\beta}x_0) = E(\hat{\alpha}) + x_0 E(\hat{\beta}) = a + \beta x_0$$

$$Var\hat{\mu}_{Y_0} = Var(\hat{\alpha} + \hat{\beta}x_0) = Var(\hat{\alpha}) + x_0^2 Var(\hat{\beta}) + 2x_0 Cov(\hat{\alpha}, \hat{\beta})$$

$$= \frac{\sigma^2}{nS_{xx}} \sum_{i=1}^{n} x_i^2 + \frac{\sigma^2 x_0^2}{S_{xx}} - \frac{2\sigma^2 x_0 \overline{x}}{S_{xx}} = \frac{\sigma^2}{S_{xx}} \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2 + x_0^2 - 2x_0\overline{x} \right)$$

$$= \frac{\sigma^2}{S_{xx}} \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2 + x_0^2 - 2x_0\overline{x} + \overline{x}^2 \right)$$

$$= \frac{\sigma^2}{S_{xx}} \left\{ \frac{1}{n} \left[\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 \right] + (x_0 - \overline{x})^2 \right\}$$

$$= \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}} \right]$$

Sampling distribution

$$\hat{\boldsymbol{\mu}}_{\mathbf{Y}_0} = \hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\beta}}\boldsymbol{x}_0 \sim \mathbf{N}\left(\boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{x}_0, \boldsymbol{\sigma}^2\left[\frac{1}{\boldsymbol{n}} + \frac{(\boldsymbol{x}_0 - \overline{\boldsymbol{x}})^2}{\boldsymbol{S}_{\boldsymbol{xx}}}\right]\right)$$

$$\frac{(\widehat{\alpha} + \widehat{\beta}x_0) - (\alpha + \beta x_0)}{\sqrt{\sigma^2\left[\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right]}} \sim \mathrm{N}(0,1)$$

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2} \Rightarrow$$

$$\frac{(\widehat{\alpha} + \widehat{\beta}x_0) - (\alpha + \beta x_0)}{\sqrt{S^2\left[\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right]}} = \frac{(\widehat{\alpha} + \widehat{\beta}x_0) - (\alpha + \beta x_0)}{\sqrt{\sigma^2\left[\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right]}} \frac{1}{\sqrt{\frac{(n-2)S^2}{(n-2)\sigma^2}}} \sim T_{n-2}$$

Interval estimation

$$\frac{(\widehat{\alpha}+\widehat{\beta}x_0)-(\alpha+\beta x_0)}{\sqrt{S^2\left[\frac{1}{n}+\frac{(x_0-\overline{x})^2}{S_{xx}}\right]}} \sim T_{n-2}$$

It is a pivotal quantity.
Therefore, a $1-\rho$ confidence interval for $\alpha+\beta x_0$ is

$$\hat{\alpha}+\hat{\beta}x_0 - t_{n-2,\rho/2}S\sqrt{\frac{1}{n}+\frac{(x_0-\overline{x})^2}{S_{xx}}} \leq \alpha+\beta x_0 \leq \hat{\alpha}+\hat{\beta}x_0 + t_{n-2,\rho/2}S\sqrt{\frac{1}{n}+\frac{(x_0-\overline{x})^2}{S_{xx}}}$$

Prediction at a single point

Assume that

$$Y_0 = \alpha + \beta x_0 + \varepsilon_0 \sim \mathrm{N}(\alpha + \beta x_0, \sigma^2)$$

Then,

$$Y_0 - \hat{\mu}_{Y_0} \text{ has a normal distribution}$$

$$E(Y_0 - \hat{\mu}_{Y_0}) = EY_0 - E\hat{\mu}_{Y_0} = (\alpha + \beta x_0) - (\alpha + \beta x_0) = 0$$

$$Var(Y_0 - \hat{\mu}_{Y_0}) = VarY_0 + Var\hat{\mu}_{Y_0} + 2Cov(Y_0, \hat{\mu}_{Y_0})$$

$$= \sigma^2 + \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}} \right]$$

$$= \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}} \right]$$

Sampling distribution

$$\boldsymbol{Y}_0 - (\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\beta}} \boldsymbol{x}_0) \sim \mathbf{N}\left(0, \boldsymbol{\sigma}^2 \left[1 + \frac{1}{\boldsymbol{n}} + \frac{(\boldsymbol{x}_0 - \overline{\boldsymbol{x}})^2}{\boldsymbol{S}_{xx}}\right]\right)$$

$$\frac{Y_0 - (\widehat{\alpha} + \widehat{\beta} x_0)}{\sqrt{\sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right]}} \sim \mathrm{N}(0,1)$$

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2} \Rightarrow$$

$$\frac{Y_0 - (\widehat{\alpha} + \widehat{\beta} x_0)}{\sqrt{S^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right]}} = \frac{Y_0 - (\widehat{\alpha} + \widehat{\beta} x_0)}{\sqrt{\sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right]}} \frac{1}{\sqrt{\frac{(n-2)S^2}{(n-2)\sigma^2}}} \sim T_{n-2}$$

Prediction interval

A $1-\alpha$ **prediction interval** for an unobserved random variable Y based on the observed data $\mathbf{X}$ is a random interval $[L(\mathbf{X}), U(\mathbf{X})]$ with the property that

$$P_\theta(Y \in [L(\mathbf{X}), U(\mathbf{X})]) \geq 1-\alpha$$

for all values of the parameter θ.

Prediction interval estimation

$$\frac{Y_0 - (\widehat{\alpha} + \widehat{\beta} x_0)}{\sqrt{S^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right]}} \sim T_{n-2}$$

It is a pivotal quantity.

Therefore, a $1-\rho$ prediction interval for Y_0 is

$$\widehat{\alpha} + \widehat{\beta} x_0 - t_{n-2,\rho/2} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}} \le Y_0 \le \widehat{\alpha} + \widehat{\beta} x_0 + t_{n-2,\rho/2} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}$$

$$\hat{\alpha} + \hat{\beta} x_0 - t_{n-2,\rho/2} S \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}} \le \alpha + \beta x_0 \le \hat{\alpha} + \hat{\beta} x_0 + t_{n-2,\rho/2} S \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}$$

Estimation at multiple points

Instead of making interval estimation at a single point x_0, we like to make interval estimation for a series of points

$$x_{01}, \ldots, x_{0m}.$$

Confidence band without correction

$$\frac{(\widehat{\alpha}+\widehat{\beta}x_{0i})-(\alpha+\beta x_{0i})}{\sqrt{S^2\left[\frac{1}{n}+\frac{(x_{0i}-\overline{x})^2}{S_{xx}}\right]}} \sim T_{n-2}$$

It is a pivotal quantity for any $i = 1, \ldots, m$.

Therefore, a $1-\rho$ confidence interval for $\alpha+\beta x_{0i}$ is

$$\widehat{\alpha}+\widehat{\beta}x_{0i}-t_{n-2,\rho/2}S\sqrt{\frac{1}{n}+\frac{(x_{0i}-\overline{x})^2}{S_{xx}}} \leq \alpha+\beta x_{0i} \leq \widehat{\alpha}+\widehat{\beta}x_{0i}+t_{n-2,\rho/2}S\sqrt{\frac{1}{n}+\frac{(x_{0i}-\overline{x})^2}{S_{xx}}}$$

Any problem?

Confidence band with Bonferroni C Correction

$$\frac{(\hat{\alpha}+\hat{\beta}x_{0i})-(\alpha+\beta x_{0i})}{\sqrt{S^2\left[\frac{1}{n}+\frac{(x_{0i}-\overline{x})^2}{S_{xx}}\right]}} \sim T_{n-2}$$

It is a pivotal quantity for any $i = 1, \ldots, m$.

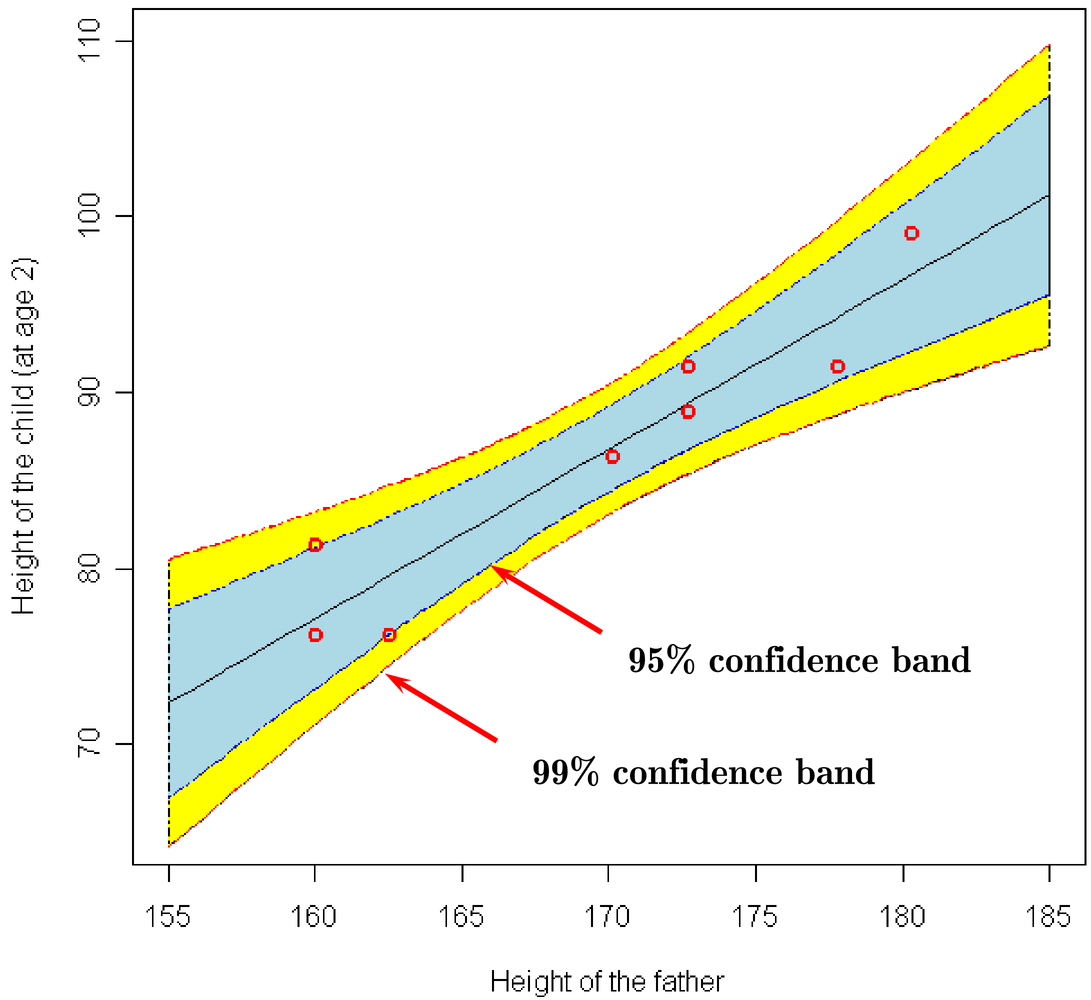
Therefore, a $1-\rho$ confidence interval for $\alpha + \beta x_{0i}$ is

$$\hat{\alpha}+\hat{\beta}x_{0i} - t_{n-2,\rho/(2m)} S\sqrt{\frac{1}{n}+\frac{(x_{0i}-\overline{x})^2}{S_{xx}}}$$

$$\leq \alpha + \beta x_{0i} \leq$$

$$\hat{\alpha}+\hat{\beta}x_{0i} + t_{n-2,\rho/(2m)} S\sqrt{\frac{1}{n}+\frac{(x_{0i}-\overline{x})^2}{S_{xx}}}$$

Illustration of confidence bands



Prediction at multiple points

Instead of making prediction at a single point x_0, we like to make prediction for a series of points

$$x_{01}, \ldots, x_{0m}.$$

Prediction band estimation correction

$$\frac{Y_{0i} - (\widehat{\alpha} + \widehat{\beta} x_{0i})}{\sqrt{S^2 \left[1 + \frac{1}{n} + \frac{(x_{0i} - \overline{x})^2}{S_{xx}}\right]}} \sim T_{n-2}$$

It is a pivotal quantity for any $i = 1, \ldots, m$.
Therefore, a $1-\rho$ prediction interval for Y_{0i} is

$$\hat{\alpha} + \hat{\beta} x_{0i} - t_{n-2,\rho/2} S \sqrt{1 + \frac{1}{n} + \frac{(x_{0i} - \overline{x})^2}{S_{xx}}}$$

$$\leq Y_{0i} \leq$$

$$\hat{\alpha} + \hat{\beta} x_{0i} + t_{n-2,\rho/2} S \sqrt{1 + \frac{1}{n} + \frac{(x_{0i} - \overline{x})^2}{S_{xx}}}$$

Any problem?

Prediction band with Bonferroni correction

$$\frac{Y_{0i} - (\widehat{\alpha} + \widehat{\beta} x_{0i})}{\sqrt{S^2 \left[1 + \frac{1}{n} + \frac{(x_{0i} - \overline{x})^2}{S_{xx}} \right]}} \sim T_{n-2}$$

It is a pivotal quantity for any $i = 1, \ldots, m$.

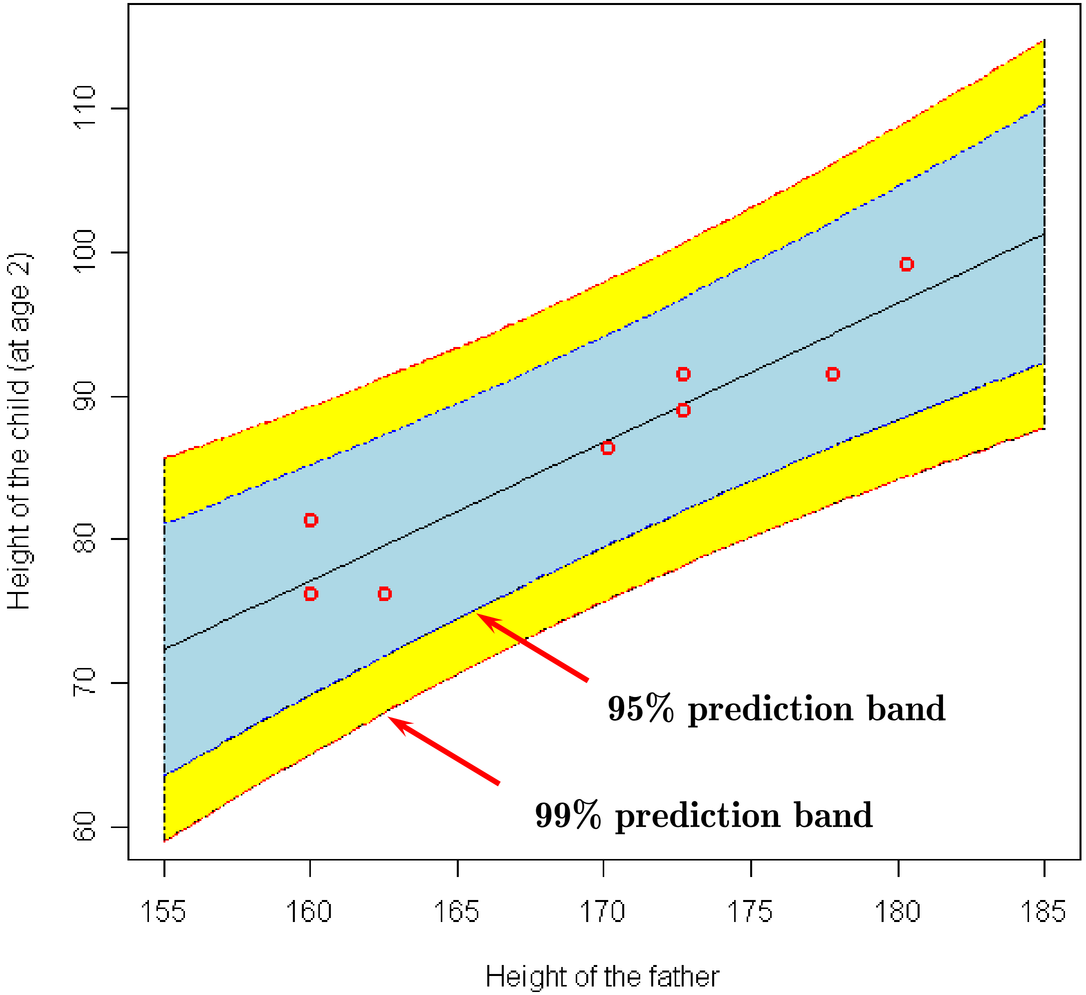
Therefore, a $1-\rho$ prediction interval for Y_{0i} is

$$\hat{\alpha} + \hat{\beta} x_{0i} - t_{n-2,\rho/(2\boldsymbol{m})} S \sqrt{1 + \frac{1}{n} + \frac{(x_{0i} - \overline{x})^2}{S_{xx}}}$$

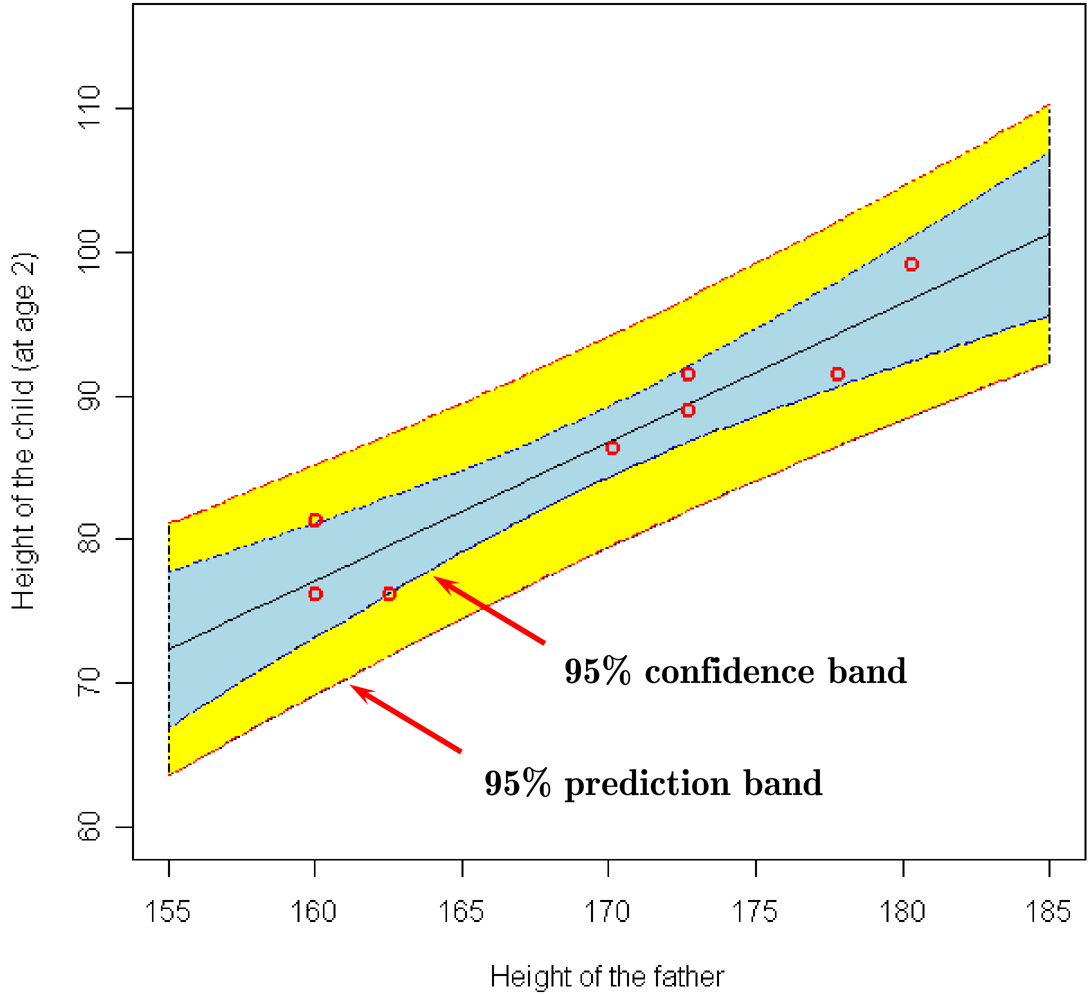
$$\leq Y_{0i} \leq$$

$$\hat{\alpha} + \hat{\beta} x_{0i} + t_{n-2,\rho/(2\boldsymbol{m})} S \sqrt{1 + \frac{1}{n} + \frac{(x_{0i} - \overline{x})^2}{S_{xx}}}$$

Illustration of prediction bands



Confidence and prediction bands



Regression functions

lm

    linear regression
    analysis of variance

predict

    confidence band
    prediction band

summary

    summarization of model fitting functions

Regression with errors in variables

Simple linear regression model

$$Y_i = \alpha + \beta x_i + \epsilon_i$$
$$\mathrm{E}\, Y_i = \alpha + \beta x_i$$

where we assume that the x's are already **KNOWN**.

Errors in variables (EIV) model, or ***measurement error model***

$$Y_i = \alpha + \beta x_i + \epsilon_i$$

where we measure a random variable whose mean is x

General model

In the general EIV model, we assume that we observe paired samples $(x_i,\ y_i)$ of random variables $(X_i,\ Y_i)$ whose means satisfy the linear relationship

$$\mathrm{E}\,Y_i = \alpha + \beta \mathrm{E} X_i$$

Define $\mathrm{E}\,Y_i = \eta_i$ and $\mathrm{E}X_i = \xi_i$, we have

$$\eta_i = \alpha + \beta \xi_i$$

A linear relationship between the means of the two random variables. η_i and ξ_i are often called ***latent variables***, a term that refers to quantities that cannot be directly measured (may be not only impossible to measure directly but impossible to measure at all)

General model

$$Y_i = \alpha + \beta\xi_i + \epsilon_i; \qquad \epsilon_i \sim N(0, \sigma_\epsilon^2)$$

$$X_i = \xi_i + \delta_i; \qquad \delta_i \sim N(0, \sigma_\delta^2)$$

1. The assumption of normality, though common, is not necessary
2. In the case that $\delta_i = 0$, the model becomes simple linear regression, since there is now no measurement error, we can directly observe the ξ_i's
3. In the case that $\alpha = 0$ (and possibly unequal variances), we have the *Behrens-Fisher problem*, the problem of interval estimation and hypothesis testing concerning the difference between the means of two normal populations when the variances of the two populations are not assumed to be equal, based on two independent samples

Linear functional relationship

$$Y_i = \alpha + \beta\xi_i + \epsilon_i; \qquad \epsilon_i \sim N(0, \sigma_\epsilon^2)$$

$$X_i = \xi_i + \delta_i; \qquad \delta_i \sim N(0, \sigma_\delta^2)$$

1. We assume the functional relationship between the means of the paired random variables $(X_i,\ Y_i)$
2. We assume that ξ_i are fixed, and errors are unknown and independent
3. We are interested in estimate parameters α and β
4. The inference is made using the joint distribution of $(X_i,\ Y_i)$, ***conditional on*** ξ_i, for $i = 1,\ 2,\ \ldots,\ n$

Linear structural relationship

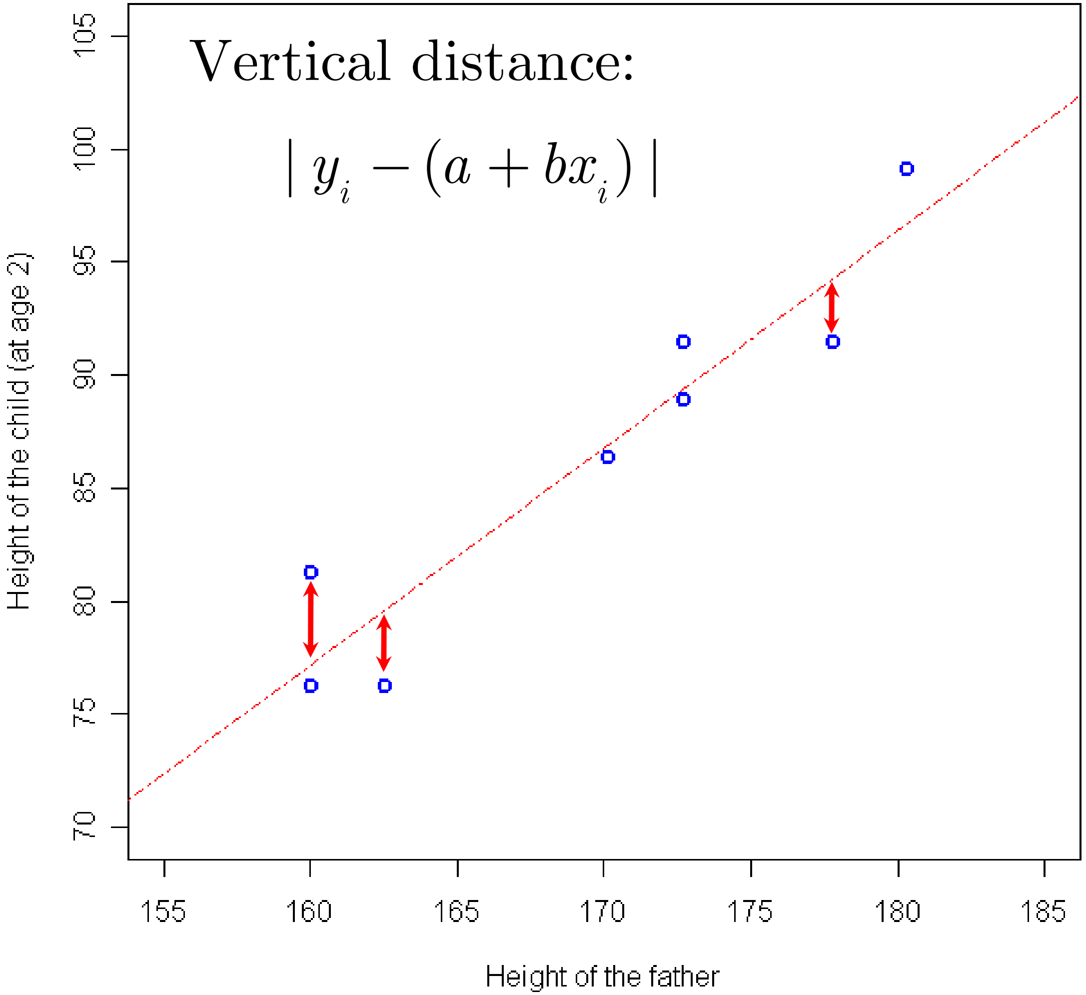
$$Y_i = \alpha + \beta\xi_i + \epsilon_i; \qquad \epsilon_i \sim N(0, \sigma_\epsilon^2)$$

$$X_i = \xi_i + \delta_i; \qquad \delta_i \sim N(0, \sigma_\delta^2)$$

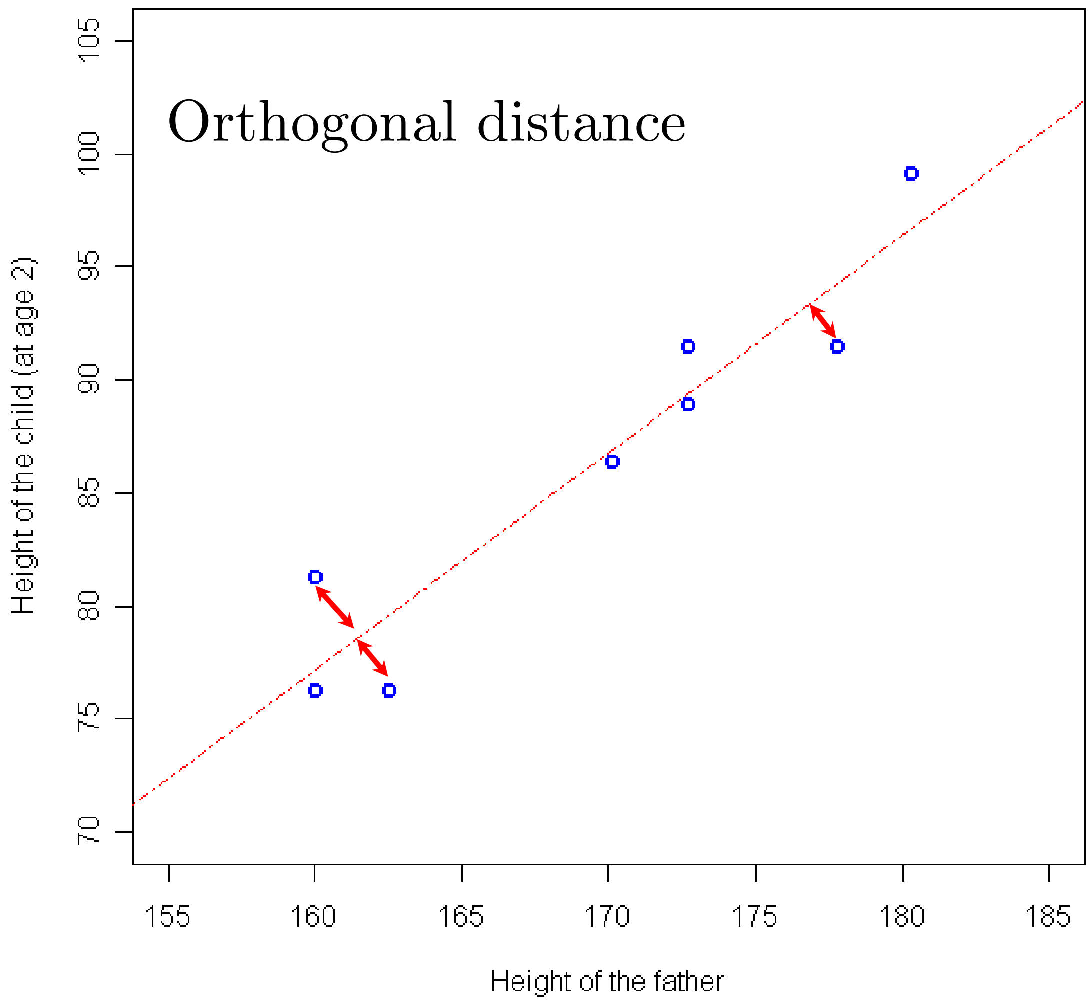
$$\xi_i = \xi + \rho_i; \qquad \rho_i \sim N(0, \sigma_\xi^2)$$

1. We assume the functional relationship between the means of the paired random variables $(X_i,\ Y_i)$
2. We assume that ξ_i are random variables, and errors are unknown and independent
3. We are interested in estimate parameters α and β
4. The inference is made using the joint distribution of $(X_i,\ Y_i)$, ***unconditional on*** ξ_i, for $i = 1, 2, \ldots, n$

Ordinary least squares



Orthogonal least squares



Maximum likelihood

For the linear functional relationship model, we commonly assume that ***the ratio between the variances of the two variables is fixed and known***, and then apply the maximum likelihood method to estimate all parameters

When the ratio is fixed as 1, the MLEs agree with the orthogonal lease square solutions

When the ratio is fixed as 0, the MLEs agree with the ordinary lease square solutions

Bivariate normal model

For the structural relationship model, we commonly assume that ***the two parameters follow a bivariate normal model***

$$(X_i, Y_i) \sim \text{Bivariate normal } (\xi, \alpha + \beta\xi, \sigma_\delta^2 + \sigma_\xi^2, \sigma_\epsilon^2 + \beta^2\sigma_\xi^2, \beta^2\sigma_\xi^2)$$

and then we apply the maximum likelihood method to estimate all parameters

Extension of the predictor variable

$$Y_i = \alpha + \beta x_i + \epsilon_i$$
$$\mathrm{E}\, Y_i = \alpha + \beta x_i$$

How about

$$\mathrm{E}\, Y_i = \alpha + \beta h(x_i)$$

For example

$$h(x) = \frac{1}{x}$$
$$h(x) = e^x$$
$$h(x) = \log x$$
$$\dots$$

Make transforms

$$\mathrm{E}\,Y_i = \alpha + \beta\frac{1}{x_i}$$

Introduce

$$x_i^* = \frac{1}{x_i}$$

The model becomes

$$\mathrm{E}\,Y_i = \alpha + \beta x_i^*$$

Everything then becomes familiar

Make transforms

$$\mathrm{E}\,Y_i = \alpha + \beta e^{x_i}$$

Introduce

$$x_i^* = e^{x_i}$$

The model becomes

$$\mathrm{E}\,Y_i = \alpha + \beta x_i^*$$

Everything then becomes familiar

Make transforms

$$\mathrm{E}\,Y_i = \alpha + \beta \log x_i$$

Introduce

$$x_i^* = \log x_i$$

The model becomes

$$\mathrm{E}\,Y_i = \alpha + \beta x_i^*$$

Everything then becomes familiar

Multiple linear regression

$$Y_i = \alpha + \beta x_i + \epsilon_i$$
$$\mathrm{E}\, Y_i = \alpha + \beta x_i$$

How about

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i$$
$$\mathrm{E}\, Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$$

Matrix representation

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i$$

Let

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & & \\ \dots & & \ddots & \\ 1 & x_{n1} & & x_{nk} \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{pmatrix} \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}$$

We have

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$$

X: Design matrix

Polynomial regression

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_i^j + \varepsilon_i$$

Let

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & & x_2^k \\ \cdots & & & \ddots & \\ 1 & x_n & x_n^2 & & x_n^k \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_k \end{pmatrix} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix}$$

We have

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$$

Least square solution

$$\mathrm{RSS} = \sum_{i=1}^{n}\left[y_i - \left(\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}\right)\right]^2$$

Equivalently

$$\begin{aligned}
\mathrm{RSS} &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
\frac{\partial}{\partial \boldsymbol{\beta}}\mathrm{RSS} &= -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
\frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}\mathrm{RSS} &= -2\mathbf{X}^T\mathbf{X}
\end{aligned}$$

$$\text{Set } \frac{\partial}{\partial \boldsymbol{\beta}}\mathrm{RSS} = 0 \Rightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Sampling distribution

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2$$

Parallel to the univariate case

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2)$$

Unbiased estimate of the variance

A typical unbiased estimate of the variance σ^2 is

$$S^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

And

$$\frac{(n-k-1)S^2}{\sigma^2} \sim \chi^2_{n-k-1}$$

Significance of a single coefficient

From

$$\hat{\boldsymbol{\beta}} \sim \mathrm{N}\left(\boldsymbol{\beta}, (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2\right)$$

$$\frac{(n-k-1)S^2}{\sigma^2} \sim \chi^2_{n-k-1}$$

We have

$$\frac{\hat{\beta}_j - \beta_j}{S\sqrt{v_j}} \sim T_{n-k-1}$$

Where v_j is the j-th diagonal element of $(X^TX)^{-1}$.

Consider how to test the hypothesis

$$H_0 : \hat{\beta}_j = 0 \text{ versus } H_1 : \hat{\beta}_j \neq 0$$

Significance of a group of coefficients

We have

$$\frac{(\mathrm{RSS}_0 - \mathrm{RSS}_1)\,/\,(k_1 - k_0)}{\mathrm{RSS}_1\,/\,(n - k_1 - 1)} \sim F_{k_1 - k_0,\, n - k_1 - 1}$$

Where RSS_0 is the residual sum of squares for the simple model that has k_0+1 parameters; RSS_1 is the residual sum of squares for the complex model that has k_1+1 parameters.

H_0 : the extra parameters in the complex model can be removed

versus

H_1 : H_0 is not true

Ridge regression

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i$$

Minimizing the residual sum of squares

$$\mathrm{RSS} = \sum_{i=1}^{n} \left[y_i - \left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \right) \right]^2$$

yields the classical multiple linear regression model.

Minimizing the penalized residual sum of squares

$$\mathrm{RSS}_{\mathrm{ridge}} = \sum_{i=1}^{n} \left[y_i - \left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \right) \right]^2 + \lambda \sum_{j=1}^{k} \beta_j^2$$

yields the *ridge* regression model.

Lasso regression

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i$$

Minimizing the residual sum of squares

$$\text{RSS} = \sum_{i=1}^{n} \left[y_i - \left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \right) \right]^2$$

yields the classical multiple linear regression model.

Minimizing the penalized residual sum of squares

$$\text{RSS}_{\text{lasso}} = \sum_{i=1}^{n} \left[y_i - \left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \right) \right]^2 + \lambda \sum_{j=1}^{k} | \beta_j |$$

yields the *lasso* regression model.