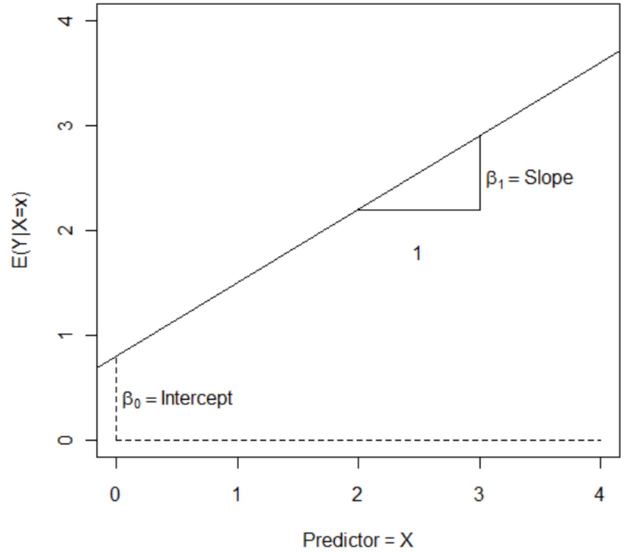
Chapter 2

Simple Linear Regression

Motivation

- The Simplest model relating Response Y and Predictor X.
  - Simple Linear Regression
- Mean function:
  - $E(Y|X=x)= \beta_0 + \beta_1 x$
- Variance function
  - $Var(Y|X=x)=\sigma^2$
- Parameters to estimates:
  - β_0 = intercept (Expectation of Y when X=0)
  - β_1 = slope (changes in E(Y) for 1 unit increase of X)
  - σ^2 = error variance

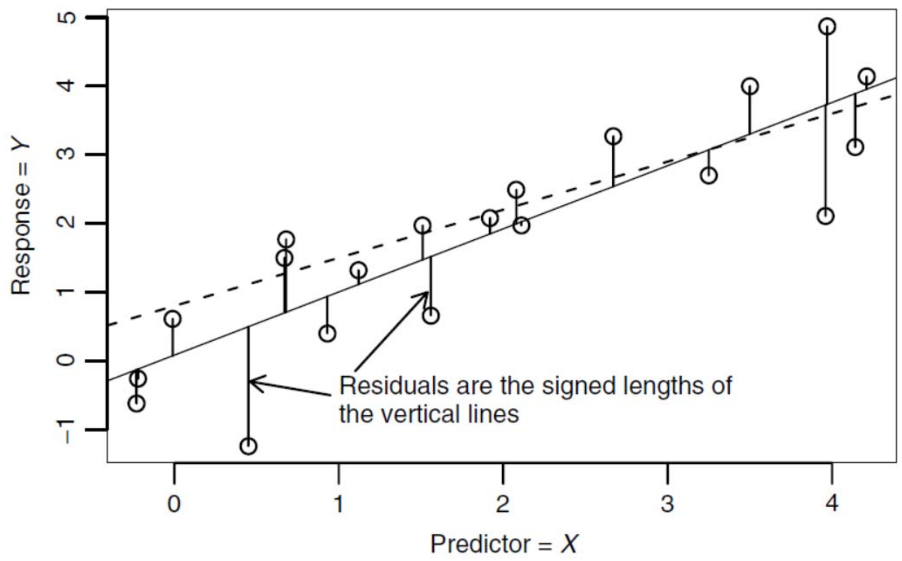


Simple Linear Regression

- Instead of expressing $E(Y|X)$ and $Var(Y|X)$ separately, we may say
  - $y_i = \beta_0 + \beta_1 x_i + e_i$
  - $E(e_i) = 0$, $Var(e_i) = \sigma^2$, e_i's are i.i.d.
- Compare with
  - Mean function: $E(Y|X=x) = \beta_0 + \beta_1 x$
  - Variance function: $Var(Y|X=x) = \sigma^2$
- e_i : statistical error
  - Vertical distance between y and the "truth" $E(Y|X=x_i)$
  - $e_i = y_i - \beta_0 - \beta_1 x_i$ = Random error component
- We can assume x is known and e, y are random

How to find the estimators?

- $y_i = \beta_0 + \beta_1 x_i + e_i$
- Vary β_0 , β_1 to find the best fit line

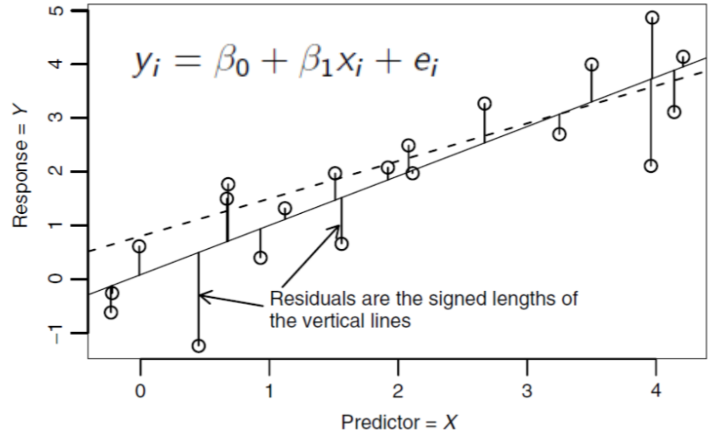


- What's meant by the "best fit"?

How to find the estimators?

- Method of Least Square.
- Residual sum of squares

$$\boxed{RSS(\beta_0, \beta_1) = \sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_i)]^2}$$



- Residual = vertical dist b/w y_i and its fitted value on the fitted regression line
- We find (β_0 , β_1) that minimize RSS

$$(\hat{\beta}_0, \hat{\beta}_1) = \mathrm{argmin}_{\beta_0,\beta_1} RSS(\beta_0, \beta_1)$$

Least Squares estimators

- Minimize Residual sum of squares

$$\boxed{RSS(\beta_0, \beta_1) = \sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_i)]^2}$$

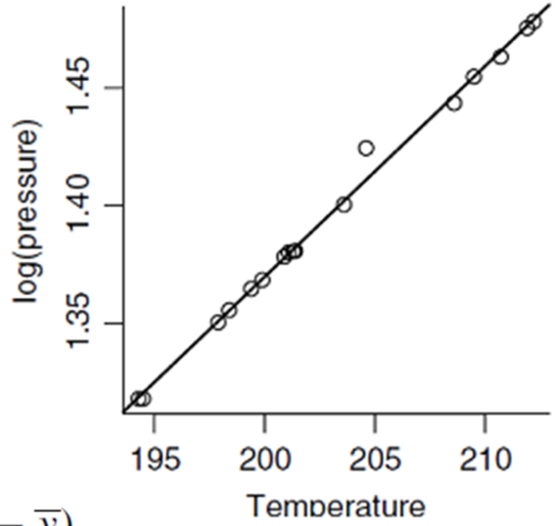
- Step 1. Differentiate RSS w.r.t. β_0 & β_1
- Step 2. Solving 2 equations 2 unknowns
- Answer:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}, \quad \hat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}$$

(see textbook P.273)

Example – Forbes Data
Measure pressure from boiling point of water

Case Number	*Temp* (°F)	*Pressure* (Inches Hg)	$Lpres = 100 \times \log(Pressure)$
1	194.5	20.79	131.79
2	194.3	20.79	131.79
3	197.9	22.40	135.02
⋮	⋮	⋮	⋮
17	212.2	30.06	147.80



$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}, \quad \hat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}$$

Results:

$$\boxed{\hat{\mathrm{E}}(Lpres \mid Temp) = -42.138 + 0.895\, Temp}$$

Least Squares estimator

- Notations used in the book

Sample mean

$$\bar{x} = \frac{1}{n}\sum_i x_i,$$

$$\bar{y} = \frac{1}{n}\sum_i y_i,$$

Sum of squares

$$SXX = \sum_i (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2,$$

$$SYY = \sum_i (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2,$$

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

Estimators

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}, \quad \hat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} = \frac{SXY}{SXX}$$

Fitted value for case i

$$\hat{y}_i = \hat{\mathrm{E}}(Y|X = x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

residual for case i

$$\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

Least Squares estimator

- Hat is important!!!

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

$$\hat{y}_i = \hat{\mathrm{E}}(Y|X = x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

- Be familiar with the terms: observed data, parameter, parameter estimates, error, residual, fitted value…
- With hat = Known estimated value
  = Involve data only
- No hat = Observed/Unknown true value

How to estimate σ^2 ?

- There are 3 parameters
  - $y_i = \beta_0 + \beta_1 x_i + e_i$
  - $E(e_i) = 0$, $Var(e_i) = \sigma^2$, e_i's are i.i.d.
- Note that $\sigma^2 = Var(e_i) = E(e^2_i)$,
  - Looking at $\sum \hat{e}_i^2$ helps estimate σ^2
  - Fact: $E(\sum \hat{e}_i^2) = (n-2)\,\sigma^2$
  - We estimate σ^2 by

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-2}$$

How to estimate σ^2 ?

- Computations
  - $$\sum \hat{e}_i^2 \overset{\text{def}}{=} \sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 \overset{\text{def}}{=} RSS(\hat{\beta}_0, \hat{\beta}_1) \overset{\text{def}}{=} RSS$$
  - $$RSS = \sum_{i=1}^{n}[y_i - \bar{y} + \bar{y} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

$$= \sum_{i=1}^{n}[y_i - \bar{y}\,]^2 + 2\sum_{i=1}^{n}[\bar{y} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)][y_i - \bar{y}\,] + \sum_{i=1}^{n}[\bar{y} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

$$= \sum_{i=1}^{n}[y_i - \bar{y}\,]^2 - \hat{\beta}_1^{\,2} \sum_{i=1}^{n}[\,x_i - \bar{x}\,]^2$$

$$\boxed{\hat{\beta}_1[\;\bar{x} - x_i\,]}$$

$$= SYY - \hat{\beta}_1^{\,2} SXX = SYY - \frac{SXY^2}{SXX}$$

- Therefore

$$\boxed{\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-2} = \frac{SYY - \frac{SXY^2}{SXX}}{n-2}}$$

How to estimate σ^2 ?

$$\boxed{\widehat{\sigma}^2 = \frac{\sum \widehat{e}_i^2}{n-2} = \frac{SYY - \frac{SXY^2}{SXX}}{n-2}}$$

- Forbe's data

$$\begin{aligned} RSS &= 427.79402 - \frac{475.31224^2}{530.78235} \\ &= 2.15493 \end{aligned}$$

square root of $\widehat{\sigma}^2$, $\widehat{\sigma} = \sqrt{0.14366} = 0.37903$ is also called the standard error of regression

Properties of least squares estimators

1. $\widehat{\beta}_0$ and $\widehat{\beta}_1$ can be written as a linear combination of y_i
   - $\widehat{\beta}_1 = \sum (\frac{x_i - \bar{x}}{SXX}) y_i \quad \hat{\beta}_0 = \sum [\frac{1}{n} - \bar{x}\left(\frac{x_i - \bar{x}}{SXX}\right)]\ y_i$
2. The fitted line passes through $(\bar{x},\bar{y})$
   - Check the derivative of RSS w.r.t. β_0
3. The sum of residuals=0 (not errors!)
   - Check the derivative of RSS w.r.t. β_0
4. The estimators are unbiased (Expectation=true value)
   - $\mathrm{E}(\widehat{\beta}_0|X) = \beta_0, \quad \mathrm{E}(\widehat{\beta}_1|X) = \beta_1, \quad \mathrm{E}(\widehat{\sigma}^2|X) = \sigma^2$

Variance of least squares estimators

Unknown True Variance	Estimators of Variance
$\mathrm{Var}(\widehat{\beta}_1 \vert X) = \sigma^2 \frac{1}{SXX}$	$\widehat{\mathrm{Var}}(\widehat{\beta}_1 \vert X) = \widehat{\sigma}^2 \frac{1}{SXX}$
$\mathrm{Var}(\widehat{\beta}_0 \vert X) = \sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{SXX})$	$\widehat{\mathrm{Var}}(\widehat{\beta}_0 \vert X) = \widehat{\sigma}^2 (\frac{1}{n} + \frac{\bar{x}^2}{SXX})$
$\mathrm{Cov}(\widehat{\beta}_0, \widehat{\beta}_1 \vert X) = -\sigma^2 \frac{\bar{x}}{SXX}$	$\widehat{\mathrm{Cov}}(\widehat{\beta}_0, \widehat{\beta}_1 \vert X) = -\widehat{\sigma}^2 \frac{\bar{x}}{SXX}$
$\rho(\widehat{\beta}_0, \widehat{\beta}_1) = \frac{-\bar{x}}{\sqrt{SXX/n + \bar{x}^2}}$	$\widehat{\rho}(\widehat{\beta}_0, \widehat{\beta}_1) = \frac{-\bar{x}}{\sqrt{SXX/n + \bar{x}^2}}$

See textbook p.273-275

2.6 Comparing models: Analysis of variance (ANOVA)

- Regression is the study of dependence of variables
  - $y_i = \beta_0 + \beta_1 x_i + e_i$
    - $\beta_1 = 0 \rightarrow$ x and y are independent
    - $\beta_1 \neq 0 \rightarrow$ x and y are dependent
- Question:
  - Are x and y dependent?

2.6 Comparing models: Analysis of variance (ANOVA)

- Regression is the study of dependence of variables
  - $y_i = \beta_0 + \beta_1 x_i + e_i$
    - $\beta_1 = 0 \rightarrow$ x and y are dependent
    - $\beta_1 \neq 0 \rightarrow$ x and y are not dependent
- Question:
  - Are x and y dependent?
- Answer:
  - Method 1) test whether $\beta_1 = 0$
  - Method 2) Compare the two models
    - $E(y|x) = \beta_0$ i.e. $y_i = \beta_0 + e_i$
    - $E(y|x) = \beta_0 + \beta_1 x$ i.e. $y_i = \beta_0 + \beta_1 x_i + e_i$

2.6 Comparing models: Analysis of variance (ANOVA)

- Analysis of variance (ANOVA) is a method that compares two models of mean functions
  - $E(y|x)=\beta_0$
  - $E(y|x)=\beta_0+\beta_1 x$
- For the first model: $E(y|x)=\beta_0$
  - β_0 can be estimated by minimizing $\sum(y_i-\beta_0)^2$
    - differentiate w.r.t. β_0 gives $\hat{\beta}_0=\overline{y}$
  - Residual sum of square RSS is

$$\sum(y_i-\hat{\beta}_0)^2=\sum(y_i-\overline{y})^2=SYY$$

2.6 Comparing models: Analysis of variance (ANOVA)

- Compare
  - $E(y|x)=\beta_0$
  - $E(y|x)=\beta_0+\beta_1 x$
- For the first model: $E(y|x)=\beta_0$
  - Residual sum of square, RSS_1, is

$$\sum (y_i - \hat{\beta}_0)^2 = \sum (y_i - \overline{y})^2 = SYY$$

- For the second model: $E(y|x)=\beta_0+\beta_1 x$
  - Residual sum of square, RSS_2, is

$$\sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 = SYY - \frac{(SXY)^2}{SXX}$$

- $RSS_1 > RSS_2$… Is the 2nd model always better?

2.6 Comparing models: Analysis of variance (ANOVA)

- RSS_1 for $E(y|x)=\beta_0$ → SYY
- RSS_2 for $E(y|x)=\beta_0+\beta_1 x$ → $SYY - \frac{(SXY)^2}{SXX}$
- The second model is useful only if $RSS_1 >>> RSS_2$
- Difference sum of square due to regression (SSreg)
  - SSreg = $RSS_1 - RSS_2 = \frac{(SXY)^2}{SXX}$
  - Large SSreg → 2nd model explains much more variation
  - How large is large?

2.6 Comparing models: Analysis of variance (ANOVA)

- Difference sum of square due to regression (SSreg)
  - SSreg = $RSS_1 - RSS_2 = \frac{(SXY)^2}{SXX}$
  - Large SSreg ➔ 2^{nd} model explains much more variation
  - How large is large?
- Study the distribution of SSreg under model 1 (idea)
  - After some algebra, $\text{SSreg} = [\sum (\frac{x_i - \bar{x}}{\sqrt{SXX}}) y_i]^2$
  - By CLT, $\sum (\frac{x_i - \bar{x}}{\sqrt{SXX}}) y_i$ is approximately $N(0,\sigma^2)$
  - SSreg ~ $\sigma^2 (N(0,1))^2 = \sigma^2 X^2_1$ (Chi-square with d.f. 1)
  - $\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-2}$ ~ $\sigma^2 X^2_{n-2}/(n-2)$ (Chi-square with d.f. n-2)
  - SSreg/$\hat{\sigma}^2$ ~ $X^2_1 / (X^2_{n-2}/(n-2)) = F(1,n-2)$

2.6 Comparing models: Analysis of variance (ANOVA)

- Statistical facts we have used:
  - Degree of freedom = number of values in the final calculation of a statistic that are free to vary
  - $\sum_{i=1}^{n} y_i$: d.f.=n, each of the y_i is free to vary
  - $\sum (y_i - \bar{y})^2$: d.f.=n-1, there is a constraint: $\sum_{i=1}^{n} y_i = n\bar{y}$
  - $RSS = \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$: d.f.=n-2, there are 2 constraints, $\hat{\beta}_0$, $\hat{\beta}_1$

  - About distributions
    - If $z \sim N(0,1)$, then $z^2 \sim X^2_1$ (Chi-square with d.f. 1)
    - If $z_i \sim N(0,1)$, then $z_1^2 + z_2^2 + \ldots + z_k^2 \sim X^2_k$ (Chi-square with d.f. k)
    - If $Y_1 \sim X^2_m$ and $Y_2 \sim X^2_n$, then $[Y_1/m]/[Y_2/n] \sim F(m,n)$

2.6 Comparing models: Analysis of variance (ANOVA)

- ANOVA table: a break-down of squares (variation)

Source	df	SS	MS	F	p-value
Regression	1	$SSreg$	$SSreg/1$	$MSreg/\hat{\sigma}^2$	
Residual	$n-2$	RSS	$\hat{\sigma}^2 = RSS/(n-2)$		
Total	$n-1$	SYY			

$$\boxed{\begin{array}{ccccc} \sum_{i=1}^{n}[y_i - \bar{y}]^2 & = & \sum_{i=1}^{n}[y_i - \hat{y}_i]^2 & + & \sum_{i=1}^{n}[\hat{y}_i - \bar{y}]^2 \\ TSS = SYY & = & RSS & + & SSreg \end{array}}$$

Variation of the data	Variation not explained by regression	Variation explained by regression

2.6 Comparing models: Analysis of variance (ANOVA)

- ANOVA table: a break-down of squares (variation)

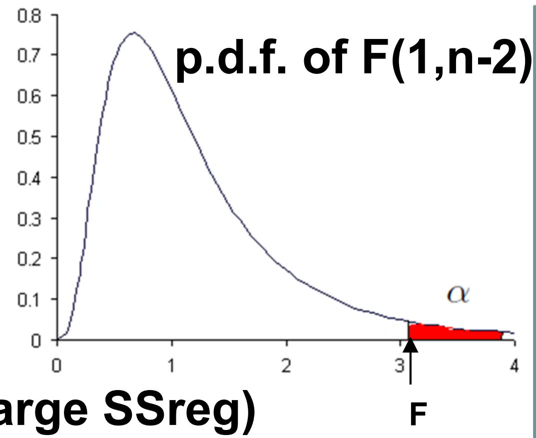
Source	df	SS	MS	F	p-value
Regression	1	$SSreg$	$SSreg/1$	$MSreg/\hat{\sigma}^2$	
Residual	$n-2$	RSS	$\hat{\sigma}^2 = RSS/(n-2)$		
Total	$n-1$	SYY			

F test for Regression

$$\text{NH} : \mathrm{E}(Y|X=x) = \beta_0$$

$$\text{AH} : \mathrm{E}(Y|X=x) = \beta_0 + \beta_1 x$$

statistics: $F = \dfrac{SSreg/1}{RSS/(n-2)}$ **~ F(1,n-2) under NH**



Idea:
- **larger F means regression is effective (large SSreg)**
- **Under NH, F~F(1,n-2), it is unlikely to be very big**
- **If the red area (α) is small, F is large → NH is suspicious**

α is the p-value = P(observing a test stat more extreme than F)
If p-value is small, e.g. <0.05, we reject the NH

2.6 Comparing models: Analysis of variance (ANOVA)

- Forbe's data

Source	df	SS	MS	F	p-value
Regression on *Temp*	1	425.639			
Residual	15	2.155			

  - conclusion?

2.6 Comparing models: Analysis of variance (ANOVA)

- Forbe's data

▶

Source	df	SS	MS	F	p-value
Regression on *Temp*	1	425.639	425.639	2962.79	≈ 0
Residual	15	2.155	0.144		

2.7 Coefficient of Determination, R^2

- Definition

$$R^2 = \frac{SSreg}{SYY}$$

  - Proportion of variability explained by regression
- Scale-free one number summary of strength of relationship between x and y.
- Connections to sample correlation r^2_{xy}

$$R^2 = \frac{SSreg}{SYY} = \frac{(SXY)^2}{SXX\ SYY} = r^2_{xy}$$

- R^2 is always between 0 and 1.
  - Close to 1→ good fit
  - Close to 0→ bad fit

2.8 Confidence intervals and tests

- Regression model:
  - $E(Y|X=x)= \beta_0 + \beta_1 x$
- Quantities of interests
  - Intercept: β_0
  - Slope: β_1
  - Prediction: If we observe x_*, what is the y?
  - Fitted value: $E(Y|X=x)$ for different values of x
- Confidence intervals give estimates for the above quantities of interests

2.8 Confidence intervals and tests

- Statistics facts:
  - If z_i are independent, then $z_1 + z_2 + \ldots + z_k \sim$ Normal (Central limit theorem)
  - If $z_i \sim$ iid $N(0,1)$, then $z_1^2 + z_2^2 + \ldots + z_k^2 \sim X^2_k$
  - If $Z \sim N(0,1)$ and $Y \sim X^2_m$, then $Z/\sqrt{(Y/m)} \sim t(m)$
- Recall what we have shown
  - $\hat{\beta}_0 = \sum [\frac{1}{n} - \bar{x}\left(\frac{x_i - \bar{x}}{SXX}\right)]\ y_i \quad \widehat{\mathrm{Var}}(\hat{\beta}_0 | X) = \hat{\sigma}^2 (\frac{1}{n} + \frac{\bar{x}^2}{SXX})$
  - $\widehat{\beta_1} = \sum (\frac{x_i - \bar{x}}{SXX}) y_i \quad \widehat{\mathrm{Var}}(\widehat{\beta}_1 | X) = \widehat{\sigma}^2 \frac{1}{SXX}$
  - $\widehat{\sigma}^2 = \frac{\sum \widehat{e}_i^2}{n-2}$ $\sim \sigma^2 X^2_{n-2}/(n-2)$ (Chi-square with d.f. n-2)

2.8 Confidence intervals and tests

- Intercept: β_0
- Testing:

$$\begin{aligned} &\text{NH:} \quad \beta_0 = \beta_0^*, \quad \beta_1 \text{ arbitrary} \\ &\text{AH:} \quad \beta_0 \neq \beta_0^*, \quad \beta_1 \text{ arbitrary} \end{aligned}$$

$$t = \frac{\hat{\beta}_0 - \beta_0^*}{\text{se}(\hat{\beta}_0)} \quad t\text{-distribution with } n-2 \text{ df}$$

where $se(\hat{\beta}_0) = \hat{\sigma}(\frac{1}{n} + \frac{\bar{x}^2}{SXX})^{\frac{1}{2}}$

- (1-α)x100% confidence interval:

$$\hat{\beta}_0 - t(\alpha/2, n-2)\text{se}(\hat{\beta}_0) \leq \beta_0 \leq \hat{\beta}_0 + t(\alpha/2, n-2)\text{se}(\hat{\beta}_0)$$

2.8 Confidence intervals and tests

- slope: β_1
- Testing:

$$\begin{array}{ll} \text{NH:} & \beta_1 = \beta_1^*, \quad \beta_0 \text{ arbitrary} \\ \text{AH:} & \beta_1 \neq \beta_1^*, \quad \beta_0 \text{ arbitrary} \end{array}$$

$$t = \frac{\hat{\beta}_1 - \beta_1^*}{\text{se}(\hat{\beta}_1)} \qquad t\text{-distribution with } n-2 \text{ df}$$

where $se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{SXX}}$

- (1-α)x100% confidence interval:

$$\hat{\beta}_1 - t(\alpha/2, n-2)\text{se}(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t(\alpha/2, n-2)\text{se}(\hat{\beta}_1)$$

2.8 T-test = F-test in testing β_1=0!

- A common test of slope
  - NH: $\beta_1=0$
  - AH: $\beta_1\neq 0$
- It is equivalent to comparing the models
  - $y_i= \beta_0 +e_i$
  - $y_i= \beta_0 + \beta_1 x_i+e_i$
- F-stat=(t-stat)^2

$$t\text{-statistic:}\ \ t = \frac{\hat{\beta}_1 - 0}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{SXX}}$$

$$t^2 = \frac{\hat{\beta}_1^2}{\hat{\sigma}^2/SXX} = \frac{\hat{\beta}_1^2 SXX}{\hat{\sigma}^2} = \frac{(SXY)^2}{\hat{\sigma}^2 SXX} = \frac{SSreg}{\hat{\sigma}^2} = F\text{-statistic}$$

- $F(1,m)= X^2_1/[X^2_m/m]=N(0,1)^2/[X^2_m/m] =t(m)^2$

2.8 Confidence intervals and tests

- Prediction: If we observe x_*, what is the y_*?
  - New observation (x_*,?):
  - Prediction ▸ $\tilde{y}_* = \hat{\beta}_0 + \hat{\beta}_1 x_*$
  - Prediction uncertainty
    - $Var(y_* - \tilde{y}_* \mid x_*) = Var\left((\beta_o - \hat{\beta}_o) + (\beta_1 - \hat{\beta}_1)x_* + e\right)$
    - (predicting a particular observation incorporate the error, giving σ^2)

$$\mathrm{Var}(\tilde{y}_*|x_*) = \sigma^2 + \sigma^2\left(\frac{1}{n} + \frac{(x_* - \bar{x})^2}{SXX}\right)$$

$$\mathrm{sepred}(\tilde{y}_*|x_*) = \hat{\sigma}\left(1 + \frac{1}{n} + \frac{(x_* - \bar{x})^2}{SXX}\right)^{\frac{1}{2}}$$

This is the notation used in the textbook. It should be understood as $Var(y_* - \tilde{y}_* \mid x_*)$

- Prediction interval for y_* (pointwise)

$$\boxed{\tilde{y}_* \pm t_{(\frac{\alpha}{2}, n-2)}\, \mathrm{sepred}(\tilde{y}_*|x_*)}$$

2.8 Confidence intervals and tests

- Fitted value: E(Y|X=x) for different values of x
  - What is the mean daughter height if mothers height=x?
  - Fitted value: $E(Y|X=x)=\beta_0+\beta_1 x$
  - Estimation: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
  - Estimation uncertainty: $Var(E(Y\mid x) - \hat{y}\mid x) = Var\left((\beta_o - \hat{\beta}_o) + (\beta_1 - \hat{\beta}_1)x\right)$
    - It is not a prediction, no error term, so no σ^2

$$\text{sefit}(\hat{y}|x) = \hat{\sigma}\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{SXX}\right)^{\frac{1}{2}}$$

- Confidence interval for E(Y|X=x): (pointwise)

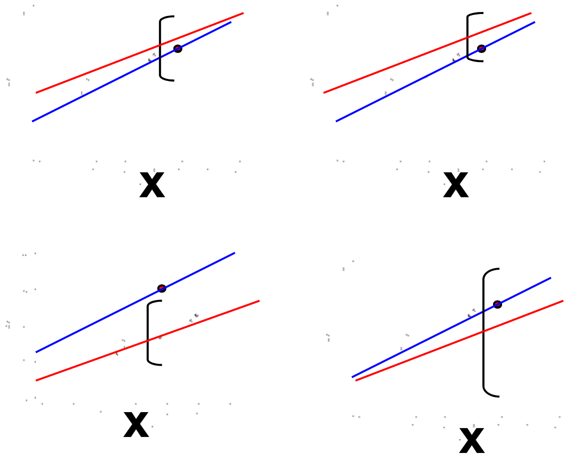
$$\boxed{\hat{y} \pm t_{(\frac{\alpha}{2},n-2)}\,\text{sefit}(\hat{y}\,|x)}$$

- Confidence band for E(Y|X=x): (for entire line)

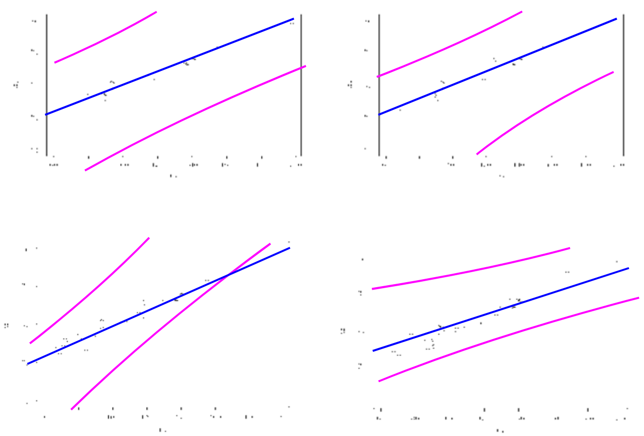
$$\boxed{(\hat{\beta}_0 + \hat{\beta}_1 x) \pm [2F(\alpha;2,n-2)]^{1/2}\text{sefit}(\hat{y}|x)}$$

2.8 Confidence intervals and bands

- Confidence interval (at each point x)
  - For <u>each</u> of x, P(E(Y|X=x) in C.I.)=1-α
- Confidence band (for the entire line)
  - P(For <u>all</u> x, E(Y|X=x) in C.B.)= 1-α



For n C.I.s, n(1-α) of them covers the true value at x

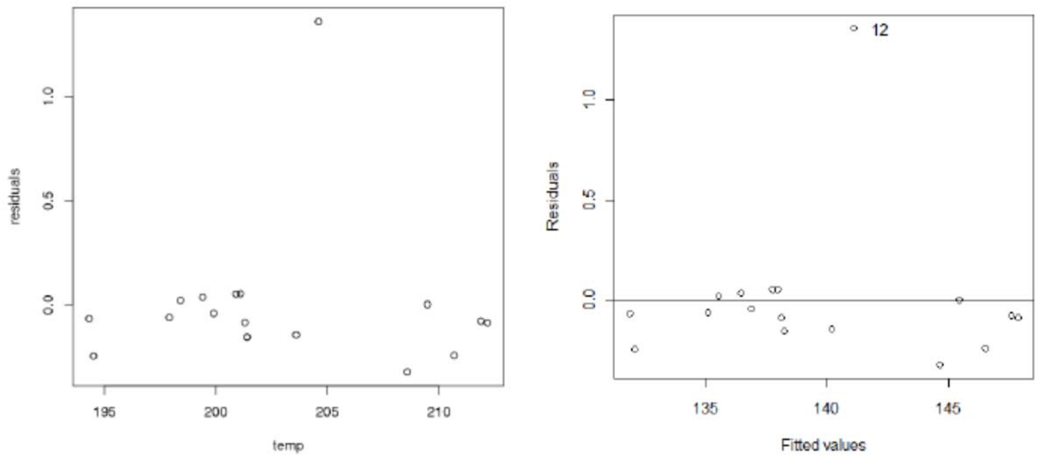


For n C.B.s, n(1-α) of them covers the whole true regression line

2.9 Residuals

- $$\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$
- Check the goodness of regression fit
- Common plots:

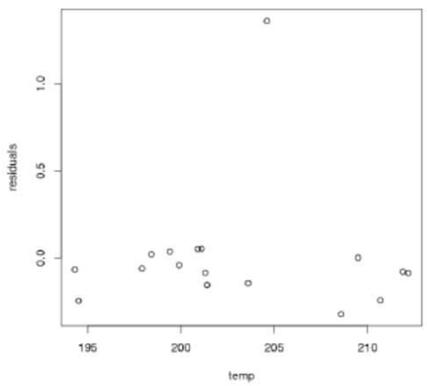
Forbes' data

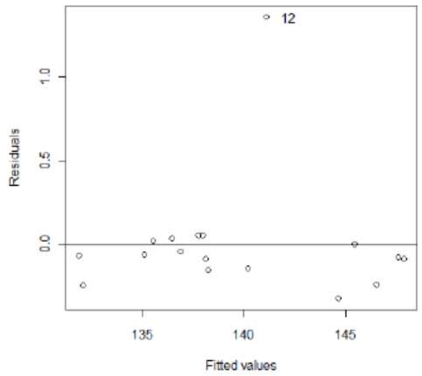


Residuals v.s. predictor Residuals v.s. fitted value

2.9 Residuals

Forbes' data



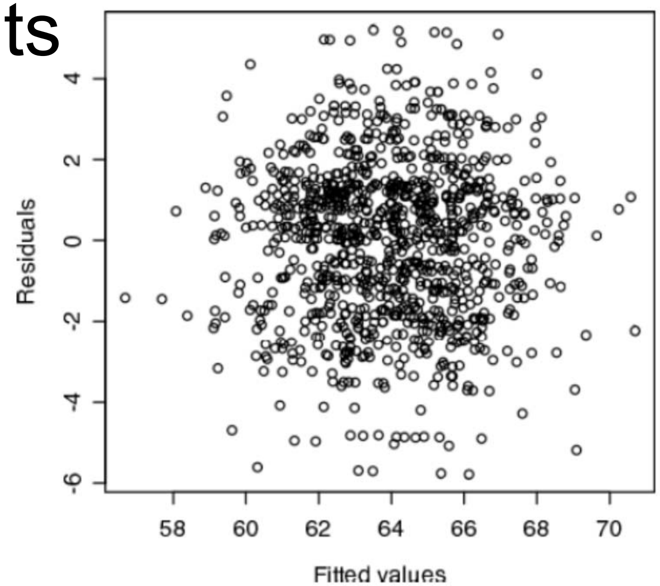


- Case 12: possible outlier
- remove Case 12 and re-do the regression
- Summary Statistics for Forbes' Data with All Data and with Case 12 deleted

Quantity	All Data	Delete Case 12
$\hat{\beta}_0$	-42.138	-41.308
$\hat{\beta}_1$	0.895	0.891
se($\hat{\beta}_0$)	3.340	1.001
se($\hat{\beta}_1$)	0.016	0.005
$\hat{\sigma}$	0.379	0.113
R^2	0.995	1.000

2.9 Residuals

- Null plot:
  - Mean function = 0
  - Variance function = constant
  - No separated points
- A null plot is a good residual plot



Chapter 2 summary

- Estimators

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}, \quad \hat{\beta}_1 = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2} = \frac{SXY}{SXX} \qquad \hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-2} = \frac{SYY - \frac{SXY^2}{SXX}}{n-2}$$

- Variances

$$\widehat{\mathrm{Var}}(\hat{\beta}_1|X) = \hat{\sigma}^2\frac{1}{SXX} \quad \widehat{\mathrm{Var}}(\hat{\beta}_0|X) = \hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SXX}\right) \quad \widehat{\mathrm{Cov}}(\hat{\beta}_0, \hat{\beta}_1|X) = -\hat{\sigma}^2\frac{\bar{x}}{SXX} \quad \hat{\rho}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x}}{\sqrt{SXX/n + \bar{x}^2}}$$

- ANOVA table

Source	df	SS	MS	F	p-value
Regression	1	$SSreg$	$SSreg/1$	$MSreg/\hat{\sigma}^2$	
Residual	$n-2$	RSS	$\hat{\sigma}^2 = RSS/(n-2)$		
Total	$n-1$	SYY			

- Tests for dependence of x and y
  - F-test: $F = \frac{SSreg/1}{RSS/(n-2)}$ **~ F(1,n-2)**
  - T-test $t = \frac{\hat{\beta}_1 - \beta_1^*}{\mathrm{se}(\hat{\beta}_1)}$ ~ $t(n-2)$

- $R^2 = \frac{SSreg}{SYY}$
- Confidence band for E(Y|X=x): $(\hat{\beta}_0 + \hat{\beta}_1 x) \pm [2F(\alpha; 2, n-2)]^{1/2}\hat{\sigma}\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{SXX}\right)^{\frac{1}{2}}$
- Confidence interval for E(Y|X=x): $(\hat{\beta}_0 + \hat{\beta}_1 x) \pm t_{(\frac{\alpha}{2}, n-2)}\,\hat{\sigma}\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{SXX}\right)^{\frac{1}{2}}$
- Prediction interval for new obs: $(\hat{\beta}_0 + \hat{\beta}_1 x) \pm t_{(\frac{\alpha}{2}, n-2)}\,\hat{\sigma}\left(1 + \frac{1}{n} + \frac{(x_* - \bar{x})^2}{SXX}\right)^{\frac{1}{2}}$