- It is equivalent to comparing the models
  - $y_i = \beta_0 + e_i$
  - $y_i = \beta_0 + \beta_1 x_i + e_i$

$$\hat{\beta}_1 = \frac{SXY}{SXX}$$

- F-stat=(t-stat)^2

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

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

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

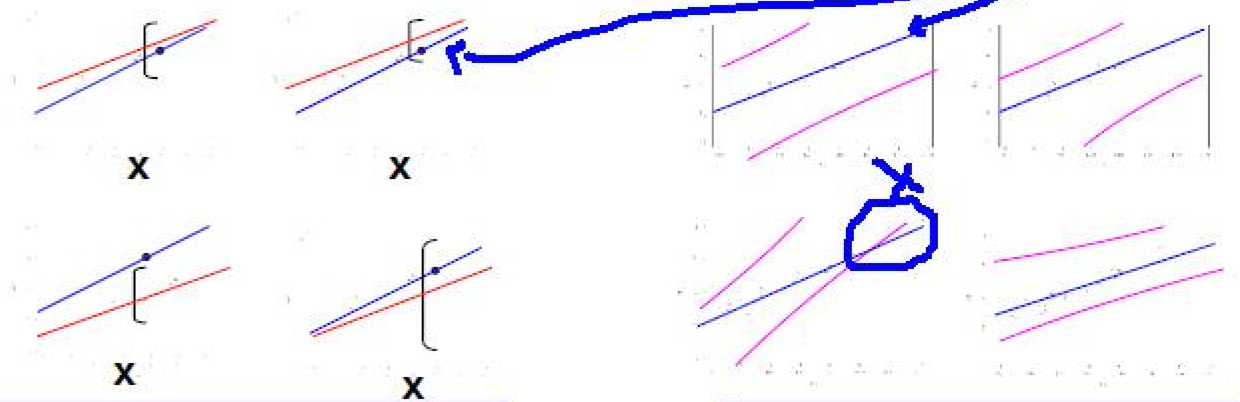
$$t(m) = \frac{N(0,1)}{\sqrt{X^2_m/m}}$$

2.8 Confidence intervals and bands

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

$Y = \beta_0 + \beta_1 x$

true regression line



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

2.8 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: $Var(E(Y|x) - \hat{y}|x) = Var\left((\beta_o - \hat{\beta}_o) + (\beta_1 - \hat{\beta}_1)x\right)$
    - 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)

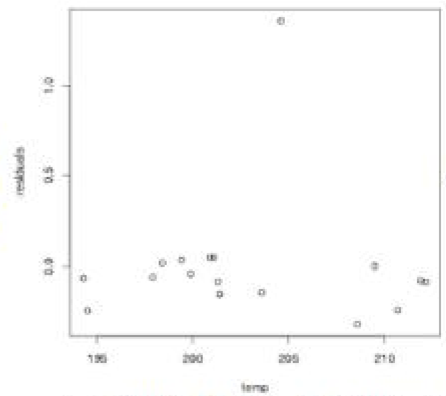
$$\boxed{\underbrace{(\hat{\beta}_0 + \hat{\beta}_1 x)}_{\hat{y}} \pm [2F(\alpha; 2, n-2)]^{1/2} \text{sefit}(\hat{y}|x)}$$

2.9 Residuals

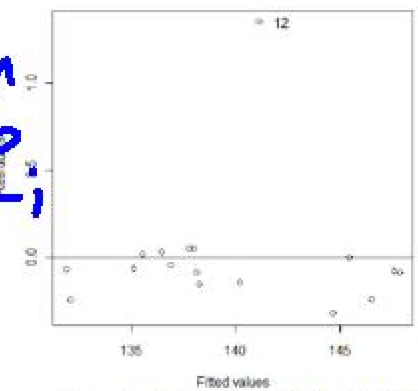
- $\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$
- Check the goodness of regression fit
- Common plots:

Forbes' data

$\hat{e}_i$

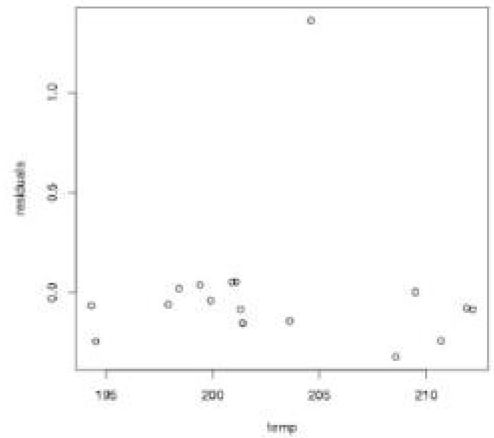


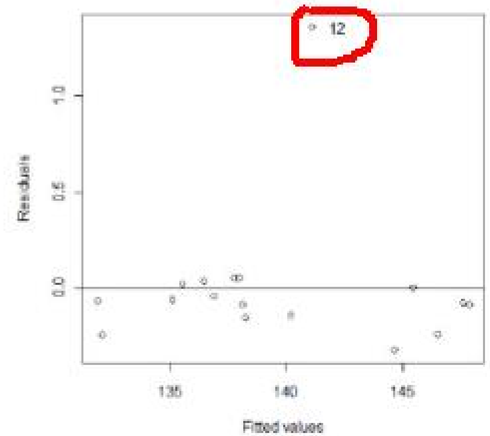
$\hat{e}_i$



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

x_i $\hat{y}_i$





- 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

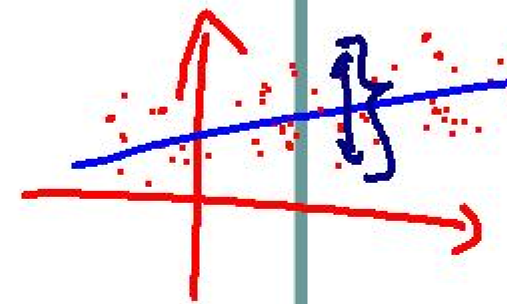
$$\frac{SSreg}{TSS(=SYY)} \rightarrow$$

- **Regression Model**

$(x_1, x_2 \cdots x_p)$

$$E(Y|X) = \beta_0 (1) + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

$$\mathrm{Var}(Y|X) = \sigma^2$$



- **Observed value in Matrix form:**

case	y	predictor 1		predictor p
1	y_1	x_{11}	$\cdots$	x_{1p}
2	y_2	x_{21}	$\cdots$	x_{2p}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_n	x_{n1}	$\cdots$	x_{np}

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

$n \times (p+1)$

intercept

Matrix Notation for Multiple Regression

$$
\mathbf{Y} = \begin{array}{c} \text{nx1} \\ \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \end{array}
\quad
\mathbf{X} = \begin{array}{c} \text{nx(p+1)} \\ \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \end{array}
\quad
\beta = \begin{array}{c} \text{(p+1)x1} \\ \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \end{array}
\quad
\mathbf{e} = \begin{array}{c} \text{nx1} \\ \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \end{array}
$$

multiple linear regression in matrix notation

$$\boxed{\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}}$$

$\Rightarrow$ the ith row is $\boxed{y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i}$

$\mathrm{Cov}(e_i, e_j) = 0$

about the vector of errors $\mathbf{e}$:

$$\mathrm{E}(\mathbf{e}) = \mathbf{0},\ \mathrm{Var}(\mathbf{e}) = \begin{pmatrix} Var(e_1) & Cov(e_1,e_2) & \ldots & Cov(e_1,e_m) \\ Cov(e_2,e_1) & Var(e_2) & \ldots & Cov(e_2,e_m) \\ \vdots & \vdots & \ldots & \vdots \\ Cov(e_m,e_1) & \ldots & \ldots & Var(e_m) \end{pmatrix} = \sigma^2 \mathbf{I_n}$$

$\Uparrow$

$e_i \rightarrow E(e_i) = 0 \quad Var(e_i) = \sigma^2 \quad e_i \overset{iid}{\sim}$