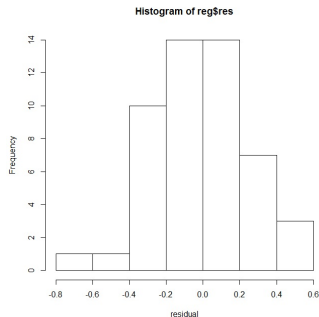
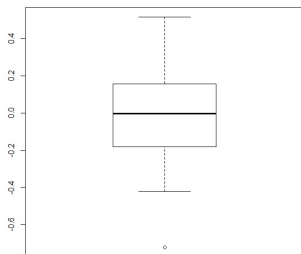


Stat GR5205 Lecture 6

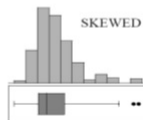
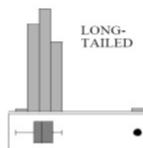
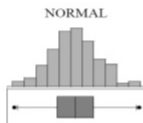
Jingchen Liu

Department of Statistics
Columbia University

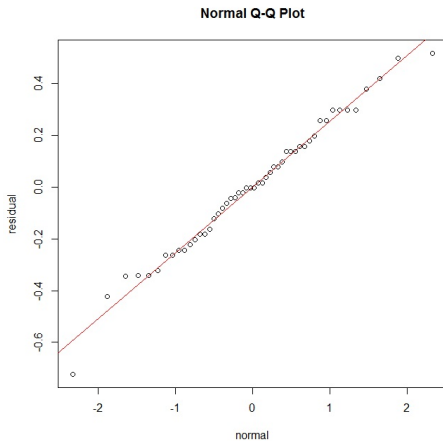
Diagnosis – graphical analysis



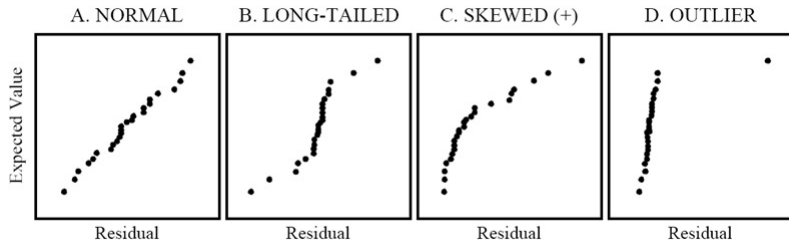
Diagnosis – box plot



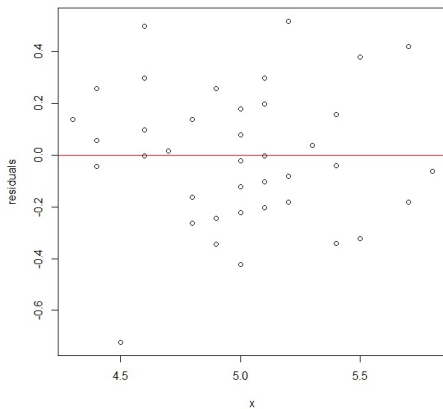
Diagnosis – quantile-quantile plot



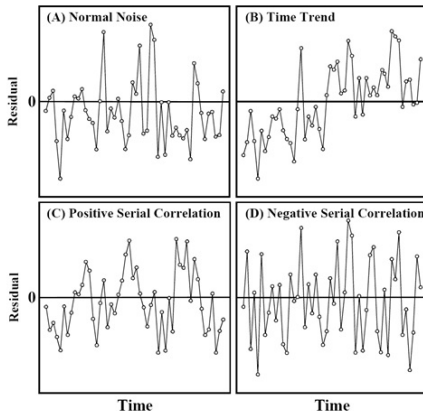
Diagnosis – quantile-quantile plot



Diagnosis – residuals



Some possible deviation away from the assumption



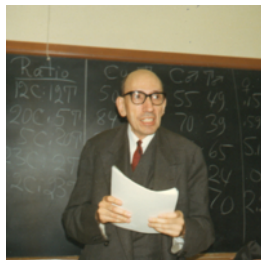
Diagnostic test

- ▶ Pure significant test
- ▶ Equal variance test: Brown-Forsythe test and Levene's test

Diagnostic test

- ▶ Pure significant test
- ▶ Equal variance test: Brown-Forsythe test and Levene's test

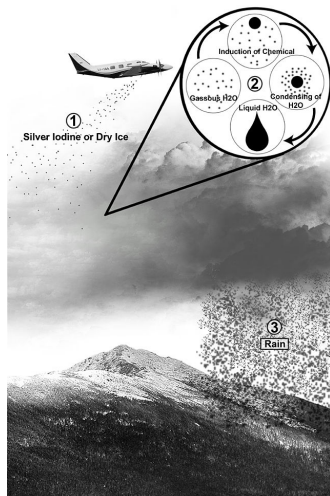
Howard Levene



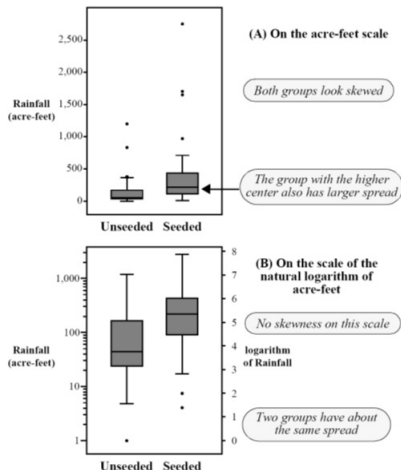
Transformation 101

- ▶ Logarithm!

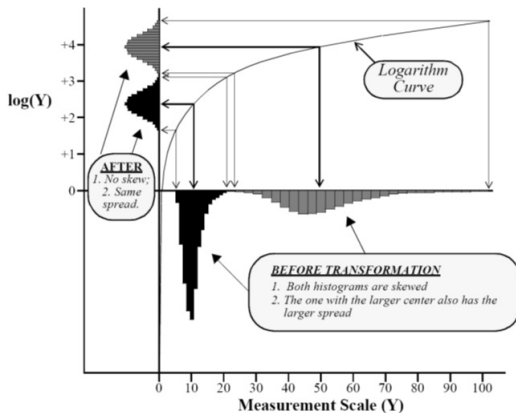
Logarithm



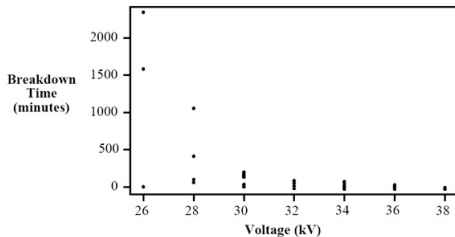
Logarithm



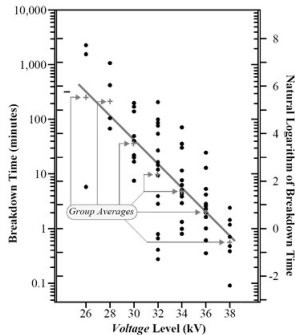
Logarithm



Logarithm



Logarithm



Box-Cox tranformation

$$f_{\lambda}(y)$$

- ▶ $f_{\lambda}(y) = (y^{\lambda} - 1)/\lambda$ if $\lambda \neq 0$; $f_{\lambda}(y) = \log y$, if $\lambda = 0$.
- ▶ Choice of λ : maximum likelihood estimate
- ▶ `library{car}`: `box.cox.powers`

Box-Cox tranformation

$$f_{\lambda}(y)$$

- ▶ $f_{\lambda}(y) = (y^{\lambda} - 1)/\lambda$ if $\lambda \neq 0$; $f_{\lambda}(y) = \log y$, if $\lambda = 0$.
- ▶ Choice of λ : maximum likelihood estimate
- ▶ `library{car}: box.cox.powers`

Box-Cox tranformation

$$f_{\lambda}(y)$$

- ▶ $f_{\lambda}(y) = (y^{\lambda} - 1)/\lambda$ if $\lambda \neq 0$; $f_{\lambda}(y) = \log y$, if $\lambda = 0$.
- ▶ Choice of λ : maximum likelihood estimate
- ▶ `library{car}`: `box.cox.powers`

Some other transformations

$$y \in [0, 1]$$

- ▶ Logit transform: $\log \frac{y}{1-y}$
- ▶ Probit transform: $F^{-1}(y)$ where $F(x) = P(Z \leq x)$
- ▶ etc.

Some other transformations

$$y \in [0, 1]$$

- ▶ Logit transform: $\log \frac{y}{1-y}$
- ▶ Probit transform: $F^{-1}(y)$ where $F(x) = P(Z \leq x)$
- ▶ etc.

Some other transformations

$$y \in [0, 1]$$

- ▶ Logit transform: $\log \frac{y}{1-y}$
- ▶ Probit transform: $F^{-1}(y)$ where $F(x) = P(Z \leq x)$
- ▶ etc.

Variance estimation

- ▶ An unbiased estimator

$$\hat{\sigma}^2 = \frac{(Y - \hat{Y})^\top (Y - \hat{Y})}{n - p - 1}$$

- ▶ The distribution of $\hat{\sigma}^2$

- ▶ Prediction

$$\hat{Y} = X(X^\top X)^{-1}X^\top Y$$

- ▶ Hat matrix

$$H = X(X^\top X)^{-1}X^\top$$

Variance estimation

- ▶ An unbiased estimator

$$\hat{\sigma}^2 = \frac{(Y - \hat{Y})^\top (Y - \hat{Y})}{n - p - 1}$$

- ▶ The distribution of $\hat{\sigma}^2$

- ▶ Prediction

$$\hat{Y} = X(X^\top X)^{-1}X^\top Y$$

- ▶ Hat matrix

$$H = X(X^\top X)^{-1}X^\top$$

Variance estimation

- ▶ An unbiased estimator

$$\hat{\sigma}^2 = \frac{(Y - \hat{Y})^\top (Y - \hat{Y})}{n - p - 1}$$

- ▶ The distribution of $\hat{\sigma}^2$
- ▶ Prediction

$$\hat{Y} = X(X^\top X)^{-1}X^\top Y$$

- ▶ Hat matrix

$$H = X(X^\top X)^{-1}X^\top$$

Variance estimation

- ▶ An unbiased estimator

$$\hat{\sigma}^2 = \frac{(Y - \hat{Y})^\top (Y - \hat{Y})}{n - p - 1}$$

- ▶ The distribution of $\hat{\sigma}^2$

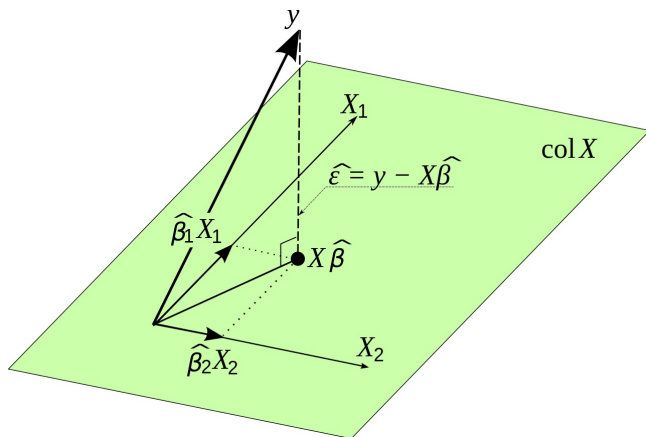
- ▶ Prediction

$$\hat{Y} = X(X^\top X)^{-1}X^\top Y$$

- ▶ Hat matrix

$$H = X(X^\top X)^{-1}X^\top$$

Projection



Joint distribution

- ▶ $\hat{\beta}$ and $\hat{\sigma}^2$ are independent.

Hypothesis testing

- ▶ Hypothesis testing

$$H_0 : \beta_i = \beta_i^0 \quad H_1 : \beta_i \neq \beta_i^0$$

- ▶ Z-statistic

$$Z - stat = \frac{\hat{\beta}_i - \beta_i^0}{SD(\hat{\beta}_i)}$$

- ▶ t-statistic

$$t - stat = \frac{\hat{\beta}_i - \beta_i^0}{SE(\hat{\beta}_i)}$$

Hypothesis testing

- ▶ Hypothesis testing

$$H_0 : \beta_i = \beta_i^0 \quad H_1 : \beta_i \neq \beta_i^0$$

- ▶ Z-statistic

$$Z - stat = \frac{\hat{\beta}_i - \beta_i^0}{SD(\hat{\beta}_i)}$$

- ▶ t-statistic

$$t - stat = \frac{\hat{\beta}_i - \beta_i^0}{SE(\hat{\beta}_i)}$$

Hypothesis testing

- ▶ Hypothesis testing

$$H_0 : \beta_i = \beta_i^0 \quad H_1 : \beta_i \neq \beta_i^0$$

- ▶ Z-statistic

$$Z - stat = \frac{\hat{\beta}_i - \beta_i^0}{SD(\hat{\beta}_i)}$$

- ▶ t-statistic

$$t - stat = \frac{\hat{\beta}_i - \beta_i^0}{SE(\hat{\beta}_i)}$$

Confidence interval

- ▶ σ^2 known

$$\hat{\beta}_i \pm q(1 - \alpha/2)SD(\hat{\beta}_i)$$

- ▶ σ^2 unknown

$$\hat{\beta}_i \pm t_{n-p-1}(1 - \alpha/2)SE(\hat{\beta}_i)$$

Confidence interval

- ▶ σ^2 known

$$\hat{\beta}_i \pm q(1 - \alpha/2)SD(\hat{\beta}_i)$$

- ▶ σ^2 unknown

$$\hat{\beta}_i \pm t_{n-p-1}(1 - \alpha/2)SE(\hat{\beta}_i)$$

Prediction

- ▶ Prediction $x^\top \hat{\beta}$
- ▶ Prediction error

$$SD(x^\top \hat{\beta}) = x^\top SD(\hat{\beta})x = \sigma^2 x^\top (X^\top X)^{-1}x$$

- ▶ Prediction of future observations
- ▶ Simultaneous interval

$$\hat{Y} \pm \lambda SE(\hat{Y})$$

where

$$\lambda^2 = (p+1)F(1-\alpha; p+1, n-p-1).$$

Prediction

- ▶ Prediction $x^\top \hat{\beta}$
- ▶ Prediction error

$$SD(x^\top \hat{\beta}) = x^\top SD(\hat{\beta})x = \sigma^2 x^\top (X^\top X)^{-1}x$$

- ▶ Prediction of future observations
- ▶ Simultaneous interval

$$\hat{Y} \pm \lambda SE(\hat{Y})$$

where

$$\lambda^2 = (p+1)F(1-\alpha; p+1, n-p-1).$$

Prediction

- ▶ Prediction $x^\top \hat{\beta}$
- ▶ Prediction error

$$SD(x^\top \hat{\beta}) = x^\top SD(\hat{\beta})x = \sigma^2 x^\top (X^\top X)^{-1}x$$

- ▶ Prediction of future observations
- ▶ Simultaneous interval

$$\hat{Y} \pm \lambda SE(\hat{Y})$$

where

$$\lambda^2 = (p+1)F(1-\alpha; p+1, n-p-1).$$

Prediction

- ▶ Prediction $x^\top \hat{\beta}$
- ▶ Prediction error

$$SD(x^\top \hat{\beta}) = x^\top SD(\hat{\beta})x = \sigma^2 x^\top (X^\top X)^{-1}x$$

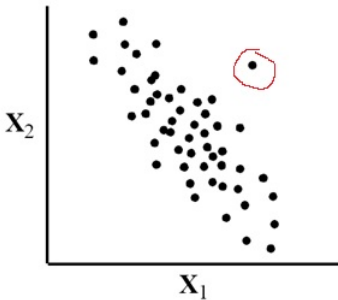
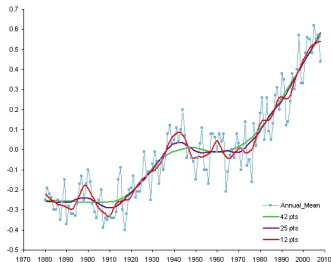
- ▶ Prediction of future observations
- ▶ Simultaneous interval

$$\hat{Y} \pm \lambda SE(\hat{Y})$$

where

$$\lambda^2 = (p+1)F(1-\alpha; p+1, n-p-1).$$

Extrapolation: one predictor and multiple predictors



Linear model is a good local approximation.

Analysis of variance

- ▶ ANOVA

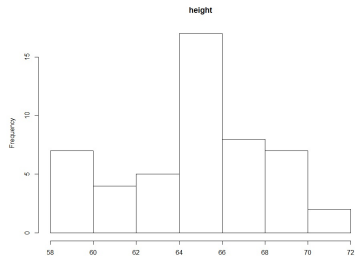
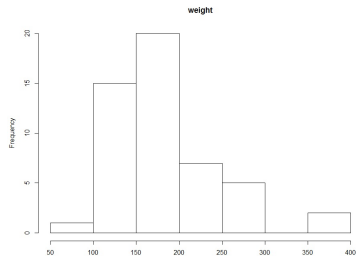
$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

- ▶ $R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

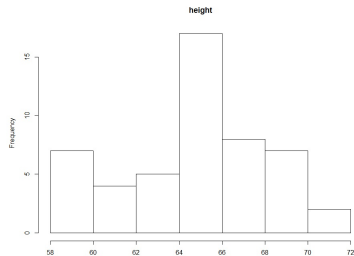
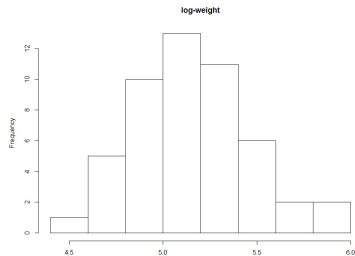
Example

- ▶ 50 samples
- ▶ 5 Asian, 15 African American, 30 Whites
- ▶ Weight and height
- ▶ Coding of the design matrix

Example



Example



Example

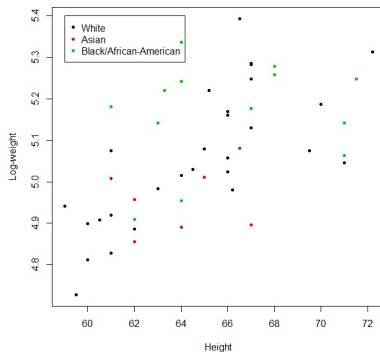


Figure: Height (inch) versus log-weight (log-lb)

Example



$$\log(\text{weight}) = \beta_0 + \beta_1 \text{height} + \varepsilon$$



$$\log(\text{weight}) = \beta_0 + \beta_{Asian} I_{Asian} + \beta_{Black} I_{Black} + \beta_1 \text{height} + \varepsilon$$



$$\begin{aligned} \log(\text{weight}) = & \beta_0 + \beta_{Asian} I_{Asian} + \beta_{Black} I_{Black} + \beta_1 \text{height} \\ & + \beta_{Asian,H} I_{Asian} \text{height} + \beta_{Black,H} I_{Black} \text{height} + \varepsilon \end{aligned}$$

Example



$$\log(\text{weight}) = \beta_0 + \beta_1 \text{height} + \varepsilon$$



$$\log(\text{weight}) = \beta_0 + \beta_{Asian} I_{Asian} + \beta_{Black} I_{Black} + \beta_1 \text{height} + \varepsilon$$



$$\begin{aligned} \log(\text{weight}) = & \beta_0 + \beta_{Asian} I_{Asian} + \beta_{Black} I_{Black} + \beta_1 \text{height} \\ & + \beta_{Asian,H} I_{Asian} \text{height} + \beta_{Black,H} I_{Black} \text{height} + \varepsilon \end{aligned}$$

Example



$$\log(\text{weight}) = \beta_0 + \beta_1 \text{height} + \varepsilon$$

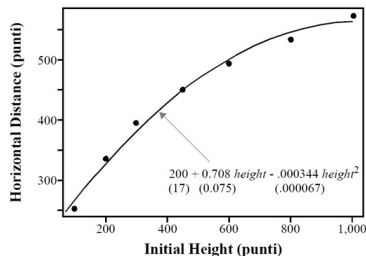
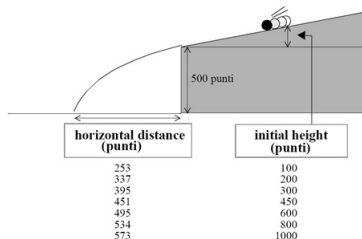


$$\log(\text{weight}) = \beta_0 + \beta_{Asian} I_{Asian} + \beta_{Black} I_{Black} + \beta_1 \text{height} + \varepsilon$$



$$\begin{aligned} \log(\text{weight}) = & \beta_0 + \beta_{Asian} I_{Asian} + \beta_{Black} I_{Black} + \beta_1 \text{height} \\ & + \beta_{Asian,H} I_{Asian} \text{height} + \beta_{Black,H} I_{Black} \text{height} + \varepsilon \end{aligned}$$

Galileo's experiment



$$\text{distance} = \beta_0 + \beta_1 \text{height} + \beta_2 \text{height}^2 + \varepsilon$$

Galileo's experiment

variable	coefficient	standard error	<i>t</i> -statistic	<i>p</i> -value
intercept	199.91	16.8	11.93	0.0003
height	0.71	0.075	9.5	0.0007
height ²	- 0.00034	0.000067	5.15	0.007

$$R^2 = 0.99 \quad \hat{\sigma} = 13.6$$

Galileo's experiment

$$distance = \beta_0 + \beta_1 height + \beta_2 height^2 + \beta_3 height^3 + \varepsilon$$

The extra sum-of-squares F test

- ▶ Question: is there any difference among the three groups aside from that due to height difference
- ▶ The formulation

$$\log(\text{weight}) = \beta_0 + \beta_{Asian} + \beta_{Black} + \beta_1 \text{height} + \varepsilon$$

- ▶ The hypotheses

$$H_0 : \beta_{Asian} = \beta_{Black} = 0 \quad H_1 : \text{otherwise}$$

The extra sum-of-squares F test

- ▶ Question: is there any difference among the three groups aside from that due to height difference
- ▶ The formulation

$$\log(\text{weight}) = \beta_0 + \beta_{Asian} + \beta_{Black} + \beta_1 \text{height} + \varepsilon$$

- ▶ The hypotheses

$$H_0 : \beta_{Asian} = \beta_{Black} = 0 \quad H_1 : \text{otherwise}$$

The extra sum-of-squares F test

- ▶ Question: is there any difference among the three groups aside from that due to height difference
- ▶ The formulation

$$\log(\text{weight}) = \beta_0 + \beta_{Asian} + \beta_{Black} + \beta_1 \text{height} + \varepsilon$$

- ▶ The hypotheses

$$H_0 : \beta_{Asian} = \beta_{Black} = 0 \quad H_1 : \text{otherwise}$$

The full model versus the reduced model

- ▶ Full model (H_1)

$$\log(\text{weight}) = \beta_0 + \beta_{Asian} + \beta_{Black} + \beta_1 \text{height} + \varepsilon$$

- ▶ Reduced model (H_0)

$$\log(\text{weight}) = \beta_0 + \beta_1 \text{height} + \varepsilon$$

- ▶ Comparing the full model against the reduce model

The full model versus the reduced model

- ▶ Full model (H_1)

$$\log(\text{weight}) = \beta_0 + \beta_{Asian} + \beta_{Black} + \beta_1 \text{height} + \varepsilon$$

- ▶ Reduced model (H_0)

$$\log(\text{weight}) = \beta_0 + \beta_1 \text{height} + \varepsilon$$

- ▶ Comparing the full model against the reduce model

The full model versus the reduced model

- ▶ Full model (H_1)

$$\log(\text{weight}) = \beta_0 + \beta_{Asian} + \beta_{Black} + \beta_1 \text{height} + \varepsilon$$

- ▶ Reduced model (H_0)

$$\log(\text{weight}) = \beta_0 + \beta_1 \text{height} + \varepsilon$$

- ▶ Comparing the full model against the reduce model

Analysis of variance

- ▶ ANOVA of the full model

$$SST = SSR_{full} + SSE_{full}$$

- ▶ ANOVA of the reduced model

$$SST = SSR_{reduced} + SSE_{reduced}$$

- ▶ Inequality

$$SSE_{extra} = SSE_{reduced} - SSE_{full} > 0$$

- ▶ Reject H_0 if SSE_{extra} is large
- ▶ Distribution of SSR_{extra}

Analysis of variance

- ▶ ANOVA of the full model

$$SST = SSR_{full} + SSE_{full}$$

- ▶ ANOVA of the reduced model

$$SST = SSR_{reduced} + SSE_{reduced}$$

- ▶ Inequality

$$SSE_{extra} = SSE_{reduced} - SSE_{full} > 0$$

- ▶ Reject H_0 if SSE_{extra} is large
- ▶ Distribution of SSR_{extra}

Analysis of variance

- ▶ ANOVA of the full model

$$SST = SSR_{full} + SSE_{full}$$

- ▶ ANOVA of the reduced model

$$SST = SSR_{reduced} + SSE_{reduced}$$

- ▶ Inequality

$$SSE_{extra} = SSE_{reduced} - SSE_{full} > 0$$

- ▶ Reject H_0 if SSE_{extra} is large
- ▶ Distribution of SSR_{extra}

Analysis of variance

- ▶ ANOVA of the full model

$$SST = SSR_{full} + SSE_{full}$$

- ▶ ANOVA of the reduced model

$$SST = SSR_{reduced} + SSE_{reduced}$$

- ▶ Inequality

$$SSE_{extra} = SSE_{reduced} - SSE_{full} > 0$$

- ▶ Reject H_0 if SSE_{extra} is large
- ▶ Distribution of SSR_{extra}

Analysis of variance

- ▶ ANOVA of the full model

$$SST = SSR_{full} + SSE_{full}$$

- ▶ ANOVA of the reduced model

$$SST = SSR_{reduced} + SSE_{reduced}$$

- ▶ Inequality

$$SSE_{extra} = SSE_{reduced} - SSE_{full} > 0$$

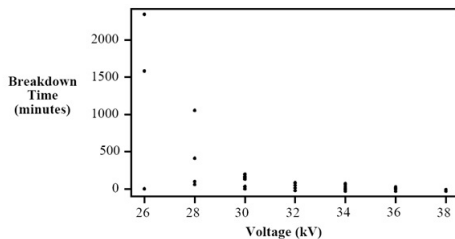
- ▶ Reject H_0 if SSE_{extra} is large
- ▶ Distribution of SSR_{extra}

Extra sums of squares test

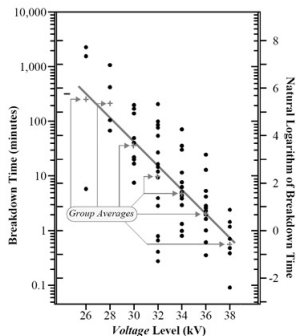
- ▶ Test statistic

$$F - statistic = \frac{SSE_{extra} / (p_{full} - p_{reduced})}{SSE_{full} / (n - p_{full})}$$

Lack-of-fit test



Lack-of-fit test



Lack-of-fit test

source	sum of sq	d.f.	mean sq	<i>F</i> -stat	<i>p</i> -value
regression	190	1	190	78	< 0.0001
residual	180	74	2.4		
total	370	75			

source	sum of sq	d.f.	mean sq	<i>F</i> -stat	<i>p</i> -value
between group	196	6	33	13	< 0.0001
residual	174	69	2.5		
total	370	75			

Lack-of-fit test

$$F - \text{statistic} = \frac{(196 - 190)/5}{174/69} = 0.48$$