AMS 256
Chapter 1: Examples of the General Linear Model

Spring 2016

$$a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^n$$

$$1_n = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \Bigg\} n$$

$$0_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$I_n = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} \quad J_{m\times n} = \overbrace{\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \\ 1 & & 1 \end{bmatrix}}^{n}$$

† Notation

1. Vectors: boldface lowercase.
   e.g $\mathbf{a} = [a_1, \ldots, a_n]^T \in \mathbb{R}^n$, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, $\mathbf{1}$, $\mathbf{0}$

2. Matrices: boldface uppercase.
   e.g. $\underbrace{\mathbf{A}}_{n \times m}$

   $\mathbf{A}_{.j.} \in \mathbb{R}^n$: the j^{th} column

3. Transpose: $\mathbf{a}^T$, $\mathbf{A}^T$

The General Linear Model

In this course we will focus on models that can be written in the following form:

$\star$ Linear Model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e},$$

where

- $\mathbf{y}$ is a $n \times 1$ vector of *observed* responses,
- $\mathbf{X}$ is a $n \times p$ matrix of *fixed constants*,
- β is a $p \times 1$ vector of *fixed but unknown* parameters, and
- $\mathbf{e}$ is a $n \times 1$ vector of *unobserved* errors

$\star\star$ Note: Linear in the *parameters*

LSE

① define a distance measure

② minimize distance between y and $f(X)\beta$.

General linear model form

$\hat{\beta} = (X^T X)^{-} X^T \underline{y}$

A more general linear model form is

$$\mathbf{y} = f(\mathbf{X})\boldsymbol{\beta} + \mathbf{e}.$$

Some topics that will be discussed in this class include:

- Estimation:
  - Least squares estimation and generalized least squares estimation.
  - BLUE (Best Linear Unbiased Estimators). Gauss-Markov theorem. $\Rightarrow$ $E(e_n) = 0_n$ & $cov(e_n) = \sigma^2 I_n$
  - MLE. $\Rightarrow$ $e_n \sim MVN(0_n, \sigma^2 I_n)$
- Estimability
- Hypothesis testing and confidence regions

To discuss these, we need to cover

- Linear algebra — vector spaces, subspaces, orthogonality, and projections. (for **LSE**)
- Distribution theory– multivariate normal distribution and properties of quadratic forms. (for **MLE**)

⋆ The theory of linear statistical models underlies several important and widely used procedures such as

- Univariate and multivariate regression analysis
- Analysis of variance (ANOVA)
- Analysis of covariance (ANACOVA)
- Random-effect modeling
- Time series analysis
- Spatial analysis ...

$(X^T X)^{-1}$

$\star$ Two Main Applications to consider in this course

- Regression Analysis: X_j, $j = 1, \ldots, k$ are continuous or categorical variables ($\Rightarrow$ full rank linear models) $\times$
- Analysis of Variance (ANOVA): the explanatory variables generally correspond to levels of different factors of interest. ($\Rightarrow$ less than full rank linear models) $\times$

$$Y = \underbrace{X\beta}_{\text{tendency}} + e$$
scatter due to random error

Regression models

Y

- Dependent variable vs Independent variable the variable of interest (dependent variable, response variable, output) is predicted from other variables (independent variables, predictor variables, explanatory variables) using an equation.

X

⋆ Regression analysis: A regression model is a formal means of expressing the two essential ingredients of a statistical relation.

- A tendency of the response variable y to vary with the predictor variable **x** in a systematic fashion.
- A scatter of points around the curve of statistical relationship.

◇ simple linear regression vs multiple linear regression

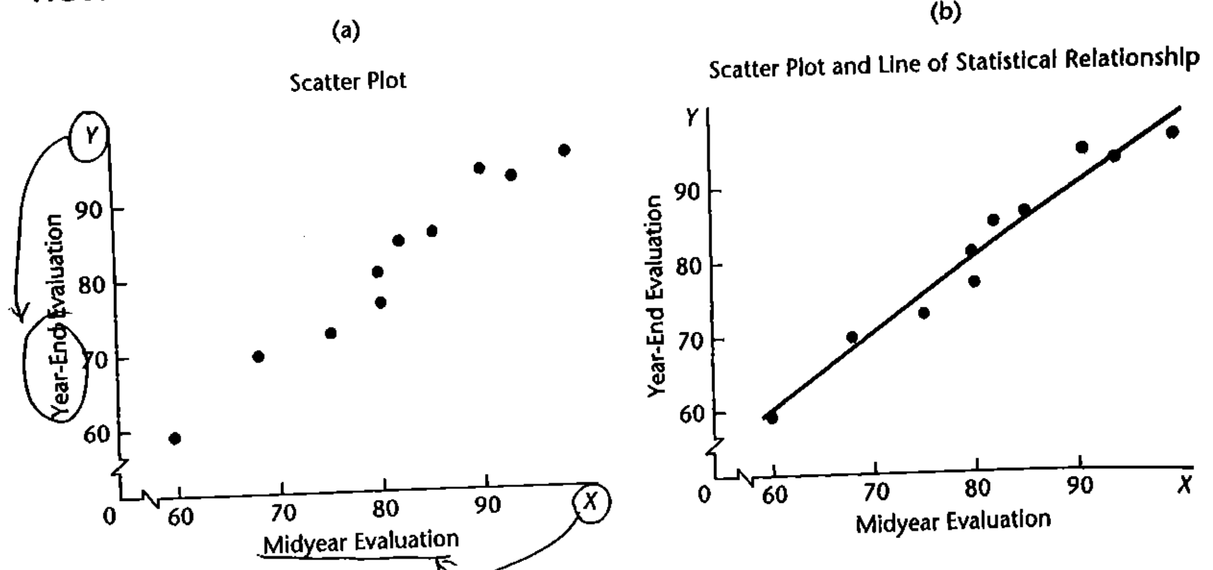
$k=1$ $\quad$ $k>1$

◇ multivariate regression vs multiple linear regression

$Y_j: \ j=1,\dots,p\ (>1)$ $\quad$ $k>1$, one single Y

◇eg.: Performance evaluations for 10 employees were obtained at midyear and at year-end.

FIGURE 1.2 **Statistical Relation between Midyear Performance Evaluation and Year-End Evaluation.**



$$\begin{bmatrix} y_1 \\ \vdots \\ y_{10} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_{10} \end{bmatrix}}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_{10} \end{bmatrix}$$

$$\Rightarrow \quad Y = X\beta + e$$

$\Diamond$eg.: The figure below presents data on age and level of a steroid in plasma for 27 healthy females between 8 and 15 years old

FIGURE 1.3 Curvilinear Statistical Relation between Age and Steroid Level in Healthy Females Aged 8 to 25.

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{27} \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{27} & x_{27}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_{27} \end{bmatrix}$$

$$\Rightarrow \quad Y = X\beta + e$$

Examples

$\star$ Common mean model

Assume that the y_is are independently sampled from a common distribution with $\mathrm{E}(y_i) = \mu$ and $\mathrm{Var}(y_i) = \sigma^2$ for all i. In this case

$$\mathbf{X}\beta = \mathbf{1}\mu, \qquad \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \mu + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

and so $\mathbf{X} = \mathbf{1} = (1, \ldots, 1)^T$ and $\beta = \mu$.

$\Leftrightarrow$ we assume that the e_is are independent with $E(e_i) = 0$ and $\mathrm{Var}(e_i) = \sigma^2$ for all i.

Examples

⋆ Simple linear regression

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1 : n.$$

What are **y**, **X**, $\boldsymbol{\beta}$ and **e**?

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

$$\mathbf{y} = \mathbf{X} \times \beta + e$$

• Note that if the x_is were measured with error then we may need a joint model for **X** and **y**. $\begin{bmatrix} x \\ y \end{bmatrix}$ $y \mid x$

Examples

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & & x_{kn} \end{bmatrix}$$

⋆ Multiple regression

Any model that can be written in the form

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + e_i, \quad i = 1 : n,$$

where the e_is are typically assumed to be uncorrelated with zero mean and constant variance σ^2 is a multiple regression model (only if $x_{i1}, \ldots, x_{ik}$ are fixed).

In this case $\mathbf{y}$ is a $n \times 1$ vector, $\mathbf{X}$ is a $n \times (k+1)$ matrix, $\boldsymbol{\beta}$ is a $(k+1) \times 1$ vector, and $\mathbf{e}$ is $n \times 1$.

Examples

The models

$$y_i = \beta_0 + \beta_1 i + \beta_2 i^2 + \beta_3 i^3 + e_i,$$

and

$$y_i = \beta_0 + \beta_1 \cos(2\pi i/7) + \beta_2 \log(i) + e_i,$$

are multiple linear regression models, however, the models

$$y_i = \beta_0 + \beta_1 \exp(-\beta_2 x_i) + e_i,$$

and

$$y_i = \frac{\beta_0}{\beta_1 + \beta_2 x_i} + \epsilon_i$$

are not.

Examples

$\star$ Bernoulli models

Let $u_1, \ldots, u_r$ be iid Bernoulli r.v. with probability p and $v_1, \ldots, v_s$ be iid Bernoulli r.v. with probability q. Let
$\mathbf{y} = (\underbrace{u_1, \ldots, u_r}_{p}, \underbrace{v_1, \ldots, v_s}_{q})^T$ and

$$\mathbf{X}\beta = \begin{pmatrix} \mathbf{1}_r & \mathbf{0}_r \\ \mathbf{0}_s & \mathbf{1}_s \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}.$$

We have that $\mathrm{E}(\mathbf{y}) = \mathbf{X}\beta$ and in this case $\mathbf{e} = \mathbf{y} - \mathrm{E}(\mathbf{y})$.

Examples

$Y = X\beta + e$

$\mathrm{cov}(e) \neq \sigma^2 I_n$

⋆ Regression with AR errors

$$\begin{aligned} y_t &= \beta_0 + \beta_1 t + e_t, \quad t = 1:n \\ e_t &= \nu + \rho e_{t-1} + a_t, \quad a_t \sim N(0, \sigma^2). \end{aligned}$$

This model can be written as

$$y_t = \beta_0(1-\rho) + \eta + \rho\beta_1 + \beta_1(1-\rho)t + \rho \boxed{y_{t-1}} + a_t,$$

however, this is not a linear model because y_{t-1} is a random variable and so the estimators and distributional results are not applicable to the estimators of ρ.

- Alternative: use a linear model with $\mathrm{Cov}(e_t, e_s) = \sigma^2 \rho^{|t-s|}$.

Design of Experiments

observational study

⋆ **Terminology**

◇ Experiments An experiment deliberately imposes a treatment on a group of objects or subjects in the interest of observing the response.

◇ Experimental units A unit is a person, animal, plant or thing which is actually studied by a researcher; the basic objects upon which the study or experiment is carried out. For example, a person; a sample of soil;

◇ Factor A factor of an experiment is a controlled independent variable; a variable whose levels are set by the experimenter.

◇ Treatment In experiments, a treatment is something that researchers administer to experimental units.

Factor A, a levels

Factor B, b levels

$a \times b$

Design of Experiments

The three basic principles of experimental design are *replication*, *randomization*, and *blocking*.

⋆ **Terminology** (contd)

◇ Replication A repetition of the basic experiment.

◇ Randomization Both the allocation of the experimental material and the order in which the individual runs or trials of the experiment are to be performed are randomly determined.

◇ Blocking Grouping experimental unites into the units that will act similarly into blocks(homogeneous clusters) and then (randomly) applying the treatments to the units in each block.

Design of Experiments

⋆ **Terminology** (contd)

◇ Complete vs incomplete "Complete" means that each block contains all the treatments. "Incomplete" means that every treatment is not present in every block.

◇ Fixed vs random effects Fixed effects are constant across individuals, and random effects vary.

◇ Balanced vs unbalanced A balanced design has an equal number of observations for each treatment. An unbalanced design has an unequal number of observations.

$\star$ The design of experiments determines the class of linear models: e.g.

- Completely randomized design

$\diamondsuit$ one factor vs multi-factor

- Randomized block design

$\diamondsuit$ complete vs incomplete

- The Latin square design: complete block designs that incorporate two separate forms of blocking

- ...

Examples

⋆ **One-way ANOVA** (statistical method for data from a completely randomized design with a single factor)

"The one-way analysis of variance (ANOVA) is used to determine whether there are any significant differences between the means of two or more independent (unrelated) groups (although you tend to only see it used when there are a minimum of three, rather than two groups). "

Assume we want to compare I treatments and we assign n_i experimental units to each treatment and measure y_{ij}.

To estimate μ and α_i, $i=1,2,3$, we place some restrictions

① Set $\sum \alpha_i = 0$ $\Rightarrow$ μ : the ~~mean~~ response over all levels

α_i : difference from the overall ~~mean~~ mean for trt i

$E(y_{1j})$

② Set $\alpha_1 = 0$ $\Rightarrow$ μ : the expected mean response ~~for~~ for level 1

$= E(\widehat{\mu + \alpha_1} + e_{1j})$

$= \mu + \alpha_1 = \mu$

α_i, $i \neq 1$: difference between level i and level 1

Examples

⋆ **One-way ANOVA** (Contd)

E.g., y_{ij} is the jth measurement of nitrogen concentration in the soil that received treatment i.

$$y_{ij} = \overset{+c}{\mu} + \overset{-c}{\alpha_i} + e_{ij}, \quad = \underbrace{\overset{10+5}{(\mu + c)}}_{=\mu'} + \underbrace{\overset{15-5}{(\alpha_i - c)}}_{=\alpha_i'} + e_{ij}$$

for $i = 1, 2, 3$, $j = 1 : n_i$. and $n_i = (3, \cancel{1}, 2)$.

- What are **y**, **X**, β and **e**?

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{31} \\ e_{32} \end{bmatrix}$$

- Are there alternative parameterizations of this model?

$y_{ij} = \theta_i + e_{ij}$ $\qquad \mu = 0$

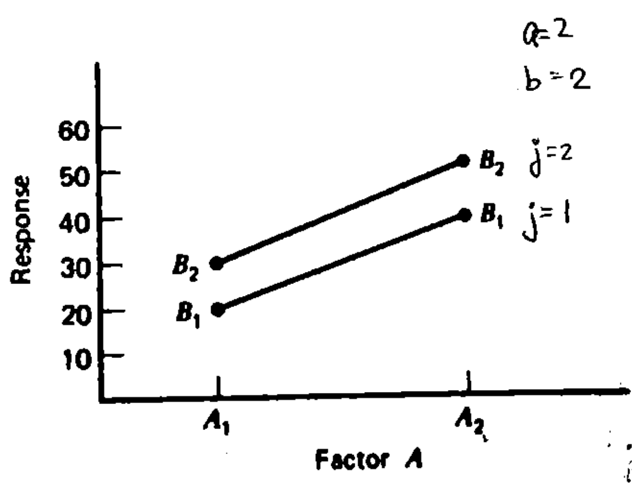
θ_i : the expected mean response for level i

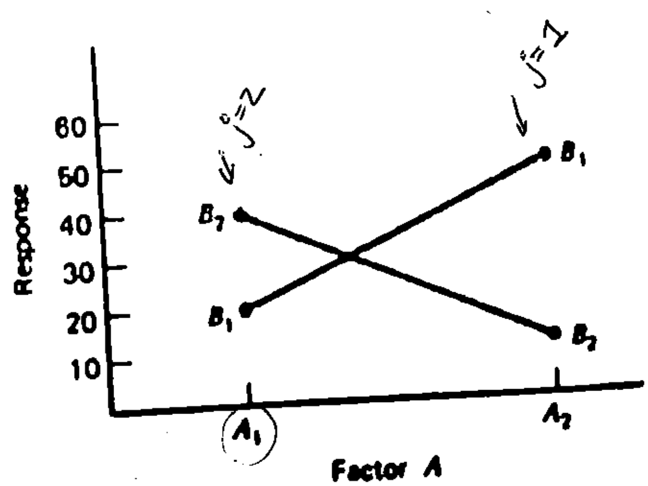
$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{31} \\ e_{32} \end{bmatrix}$$

Examples

⋆ **Interaction:** In some experiments, we may find that the difference in response between levels of one factor is not the same at all levels of the other factors. When this occurs, there is an interaction.

Consider two factors, A and B with $a = 2$ and $b = 2$, respectively.





Examples

⋆ **Two-way ANOVA models**

Assume we have two fixed factors and that we have a levels of one factor (A) and b levels of another factor (B). Also assume that the two-way layout is crossed when every level of Factor A occurs with every level of Factor B.

So the two-way crossed model (two-way ANOVA) with interactions is:

$$y_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_{i,j} + e_{i,j,k}.$$

for $i = 1 : a$, $j = 1 : b$, and $k = 1 : n$.

- μ: the common value (grand mean)
- α_i: the effect of level i of Factor A
- β_j: the effect of level j of Factor B
- $\gamma_{i,j}$: the interaction effect between α_i and β_j

Examples

⋆ Two-way ANOVA models

Consider the two-way crossed model with $n = 1$: $\quad n > 1$

$n = 1$

$$y_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_{i,j} + e_{i,j,k}.$$

for $i = 1 : a$, $j = 1 : b$, and $k = 1 : n$.

Suppose $n = 1$ for all (i, j) (that is, no replicate)

$\Rightarrow$ can't estimate γ_{ij}

$\Rightarrow$ drop γ_{ij} from the model

$\Rightarrow$ $y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$, $\quad i = 1, \dots, a$

$\quad j = 1, \dots, b$

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1b} \\ y_{21} \\ \vdots \\ y_{ab} \end{bmatrix} = \begin{array}{c} \begin{array}{ccccc} \mu & \alpha_1 & \alpha_2 & \cdots \ \alpha_a & \\ \swarrow & \swarrow & \downarrow & \swarrow & \end{array} \\ \begin{bmatrix} 1_b & 1_b & 0_b & \cdots & 0_b & I_b \\ 1_b & 0_b & 1_b & \cdots & 0_b & I_b \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1_b & 0_b & 0_b & \cdots & 1_b & I_b \end{bmatrix} \end{array} \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \\ \beta_1 \\ \vdots \\ \beta_b \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{ab} \end{bmatrix}$$

Examples

Consider a randomized complete block design:

$$y_{i,j} = \mu + \underset{\text{factor}}{\alpha_i} + \underset{\text{block}}{\beta_j} + e_{i,j}, \quad i = 1 : a, j = 1 : b.$$

where

- α_is represent the **treatment effects –fixed**
- β_js represent the **block effects** (e.g., cars and drivers).

$\Rightarrow$ Just a two-way ANOVA without interaction

♣ Note:

- Usually no test for block effects.
- If the block effects are considered random effects, this is a linear mixed model with errors given by $\beta_j + e_{i,j}$ that are no longer uncorrelated.

Examples

Little more about random effects Assume

- β_j are independent random variables with $\mathrm{E}(\beta_j) = 0$ and $\mathrm{Var}(\beta_j) = \sigma_\beta^2$
- the β_j's and $e_{i,j}$'s are independently distributed.

$\Rightarrow$ Easy to verify

$$\mathrm{Cov}(y_{i,j}, y_{i',j'}) = \begin{cases} \sigma_\beta^2 + \sigma^2 & i = i', j = j' \\ \sigma_\beta^2 & i \neq i', j \neq j' \\ 0 & i \neq i', j \neq j' \end{cases}$$

Examples

⋆ **Two-way nested models:** Consider two factors, A and B. Sometimes factor levels are **nested**, not **crossed.** If B is nested within A, this means that each level of factor A will contain different levels of factor B.

♣ eg: Suppose we have two factor, say A states, and B_i counties within each state i. In this case we have:

State 1 county A 1, B 2, C 3

$$y_{i,j,k} = \mu + \alpha_i + \beta_{i,j} + e_{i,j,k}.$$

State 2 α 1, β 2, γ 3

For example, if we have two states with two counties in the first state and three counties in the second state, $i = 1:2$, and for $i = 1$ $j = 1:2$, while for $i = 2$, $j = 1:3$. Let $k = 1:n_{i,j}$.

Examples

$\star$ **ANCOVA models**: Suppose one or more measurements are made on a group of experimental units. The concomitant observations come into play as regression variables that are added to the basic experimental design model (reducing the variability of treatment comparisons like blocking)

♣ eg: Recall the nitrogen example. Suppose that the amount of nitrogen is measured in each plot of ground prior to the application of any treatments. Suppose we have a treatments of a given factor (factor A) and a covariate x (e.g., amount of nitrogen before the treatment), then we can consider a model of the form

$$y_{i,j} = \mu + \underline{\alpha_i} + \beta x_{i,j} + e_{i,j}, \quad i = 1 : a, \; j = 1 : n_i.$$

We can also consider a model of the form

$$y_{i,j} = \mu + \alpha_i + \beta_i x_{i,j} + e_{i,j}, \quad i = 1 : a, \; j = 1 : n_i.$$

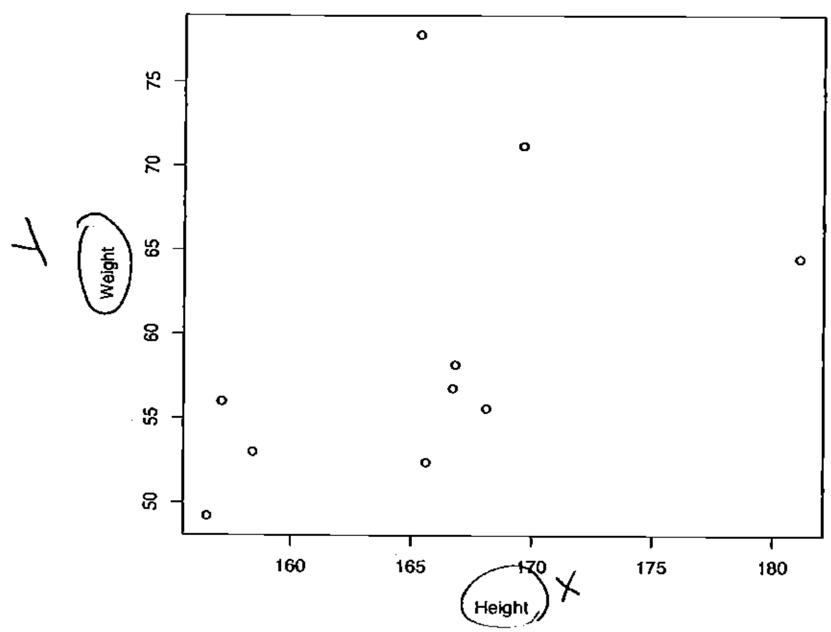
What comes next?

Estimation and hypothesis testing:

1. Consider $\mathbf{y} = \mathbf{X}\beta + e$ as a numerical approximation problem and find the best approx. to $\mathbf{y}$ as a linear function of the columns of $\mathbf{X}$.
2. Assume $E(\mathbf{e}) = \mathbf{0}$ and find unbiased estimators of $\lambda'\beta$.
3. Assume $E(e_i) = 0$, $Var(e_i) = \sigma^2$ and $Cov(e_i, e_j) = 0$ and find the linear estimator of $\lambda'\beta$ with smallest variance. BLUE
4. MLE
5. Hypothesis testing: we need distribution theory.

Simple Linear Regression

```
#Height and weight data from S Weisberg (Applied Linear Regression)
height<-c(169.6,166.8,157.1,181.1,158.4,165.6,166.7,156.5,168.1,165.3)
weight<-c(71.2,58.2,56.0,64.5,53.0,52.4,56.8,49.2,55.6,77.8)
plot(height,weight,ylab="Weight",xlab="Height")
#plot(weight~height)
```



Simple Linear Regression

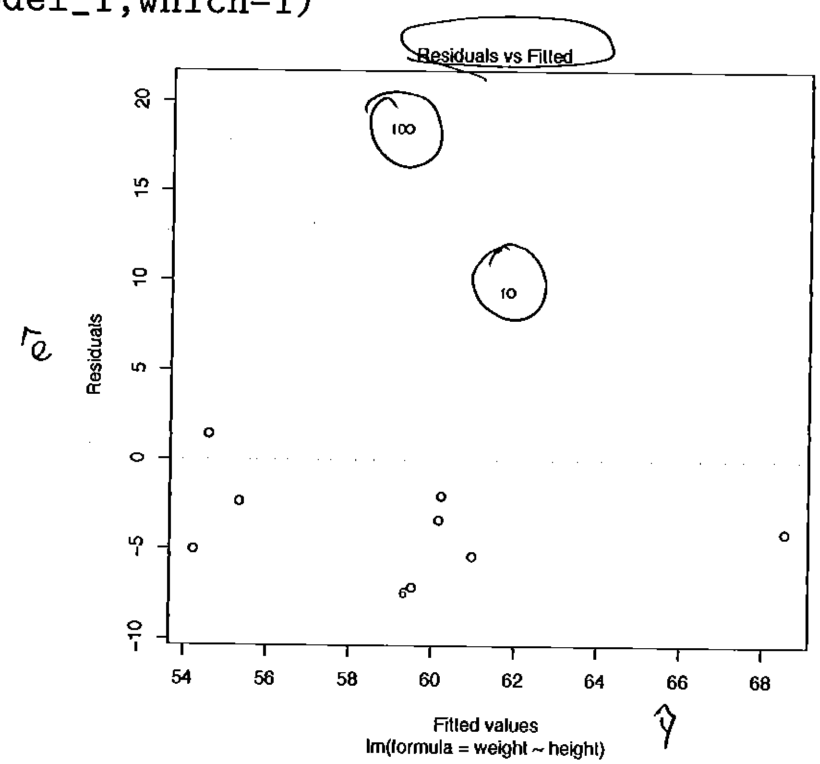
```
lm(weight~height)
#Call:
#lm(formula = weight ~ height)
#
#Coefficients:
#(Intercept)        height
#   -36.8759 β̂0     0.5821 β̂1

model_1=lm(weight~height)
anova(model_1)
#Analysis of Variance Table
#Response: weight
#           Df Sum Sq Mean Sq F value Pr(>F)
#height      1 159.95  159.95   2.237 0.1731
#Residuals   8 572.01   71.50
```

Simple Linear Regression

```
summary(model_1)
#Call:
#lm(formula = weight ~ height)
#Residuals:
#    Min      1Q  Median      3Q     Max
#-7.1166 -4.7744 -2.8412  0.5696 18.4581
#Coefficients:
#            Estimate Std. Error t value Pr(>|t|)
#(Intercept) -36.8759    64.4728  -0.572    0.583
#height        0.5821     0.3892   1.496    0.173
#
#Residual standard error: 8.456 on 8 degrees of freedom
#Multiple R-Squared: 0.2185 Adjusted R-squared: 0.1208
#F-statistic: 2.237 on 1 and 8 DF,  p-value: 0.1731
```

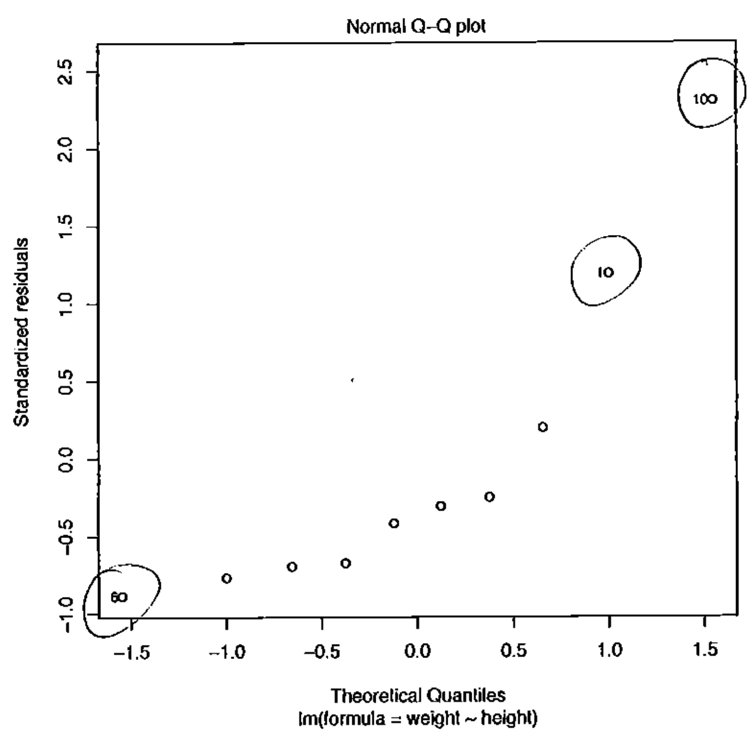
Simple Linear Regression

```
plot(model_1)
# will show various residual plots...
plot.lm(model_1,which=1)
```



Simple Linear Regression

```
plot.lm(model_1,which=2)
```



Simple Linear Regression

$\hat{y}$

```
new=data.frame(height=c(170,180))
predict.lm(model_1,new,se.fit=TRUE,
interval="confidence",level=0.95)
$fit
       fit      lwr      upr
1 62.07772 54.71651 69.43892
2 67.89852 53.51465 82.28239

$se.fit
       1        2
3.192190 6.237573

$df
[1] 8

$residual.scale
[1] 8.455868
```

Simple Linear Regression

```
predict.lm(model_1,new,se.fit=TRUE,
interval="prediction",level=0.95)
$fit
       fit      lwr      upr
1 62.07772 41.23525 82.92019
2 67.89852 43.66802 92.12902

$se.fit
       1        2
3.192190 6.237573

$df
[1] 8

$residual.scale
[1] 8.455868
```