This course was developed as a part of the VLIR-UOS Cross-Cutting project s:

- Statistics: 2011-2016, 2017.
- Statistics: 2017.
- Statistics for development : 2018-2020.

The >eR-Biostat initative

Making R based education materials in statistics accessible for all

Basic concepts in statistical modeling using R: simple linear regression

Developed by

Legesse Kassa Debusho (UNISA, South Africa) and Ziv Shkedy (Hasselt University)

contents

- Simple linear regression:
  - Introduction and model formulation.
  - Fitting a simple linear regression model using the lm() function in R.
  - Model diagnostic.
  - Model diagnostic in R.

YouTube tutorials

- YouTube tutorials are available for:
  - Simple Linear regression in R (host by Mike Marin): https://www.youtube.com/watch?v=66z_MRwtFJM&list=PLqzoL9-eJTNBJrvFcN-ohc5G13E7Big0e
  - Checking Linear Regression Assumptions in R (host by Mike Marin): https://www.youtube.com/watch?v=eTZ4VUZHzxw

YouTube tutorials

- YouTube tutorials are available for:
  - Simple Linear regression in R (host by Mike Marin): https://www.youtube.com/watch?v=66z_MRwtFJM&list=PLqzoL9-eJTNBJrvFcN-ohc5G13E7Big0e
  - Checking Linear Regression Assumptions in R (host by Mike Marin): https://www.youtube.com/watch?v=eTZ4VUZHzxw

R program and Datasets

- Simple linear regression:
  - Introduction and model formulation.
  - Fitting a simple linear regression model using the lm() function in R.
  - Model diagnostic.
  - Model diagnostic in R.

Part 1
Simple linear regression

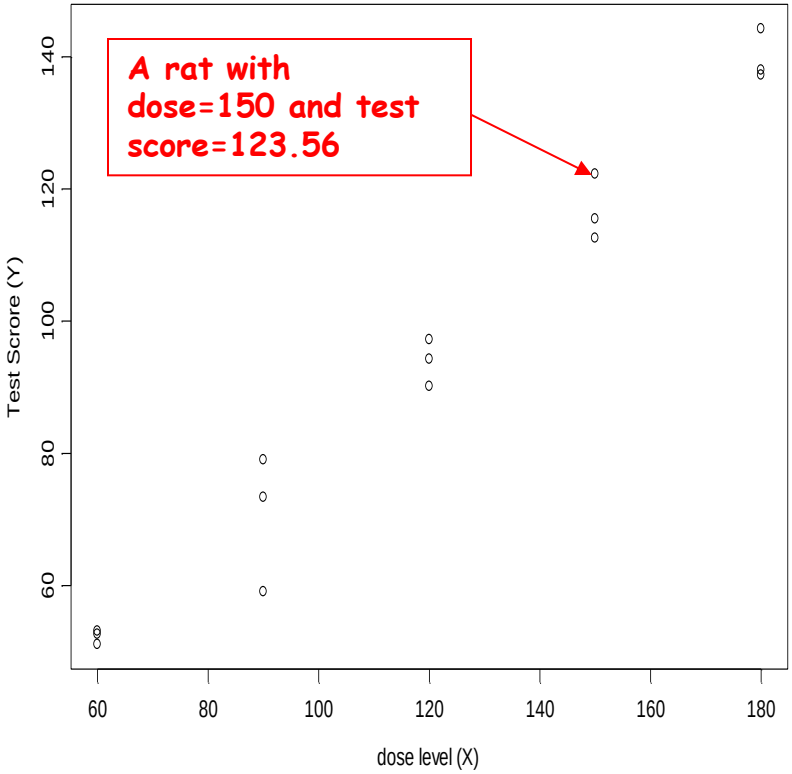
Introduction

A Biopharmaceutical Problem

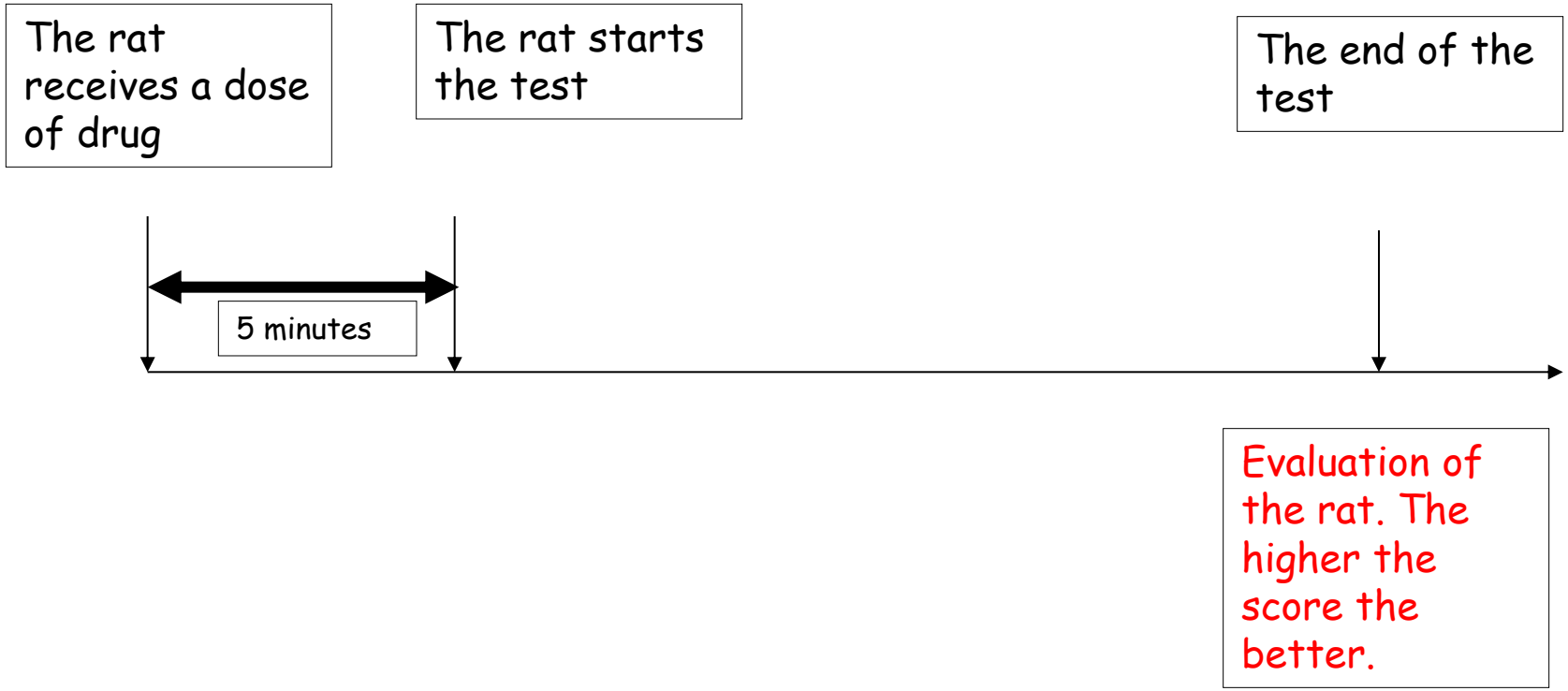
- A group of 15 rats received a dose of a drug and then had to complete a test.

- It is assumed that the performance of the rat depends on the dose level.

The data

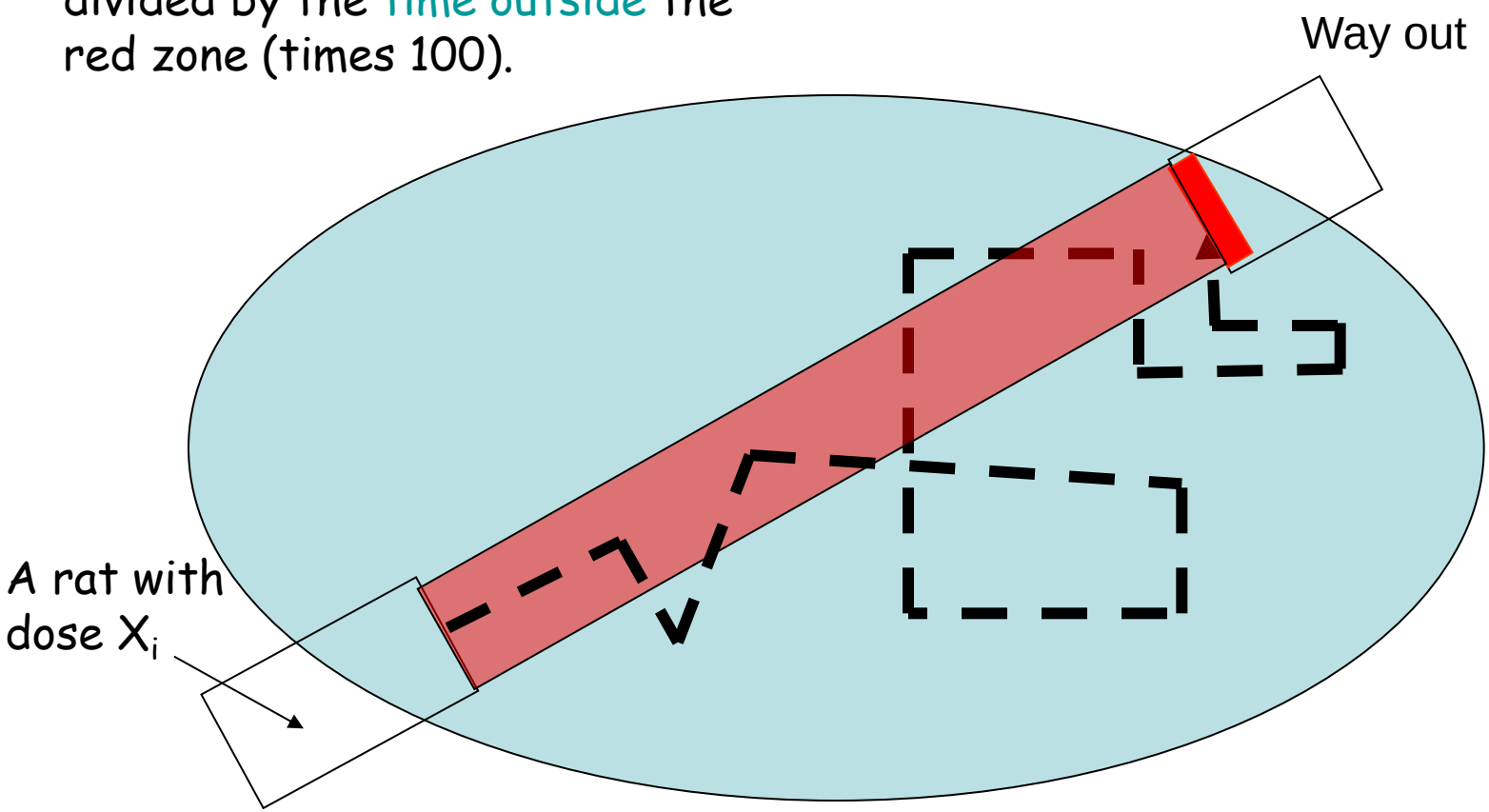


Description of the Experiment



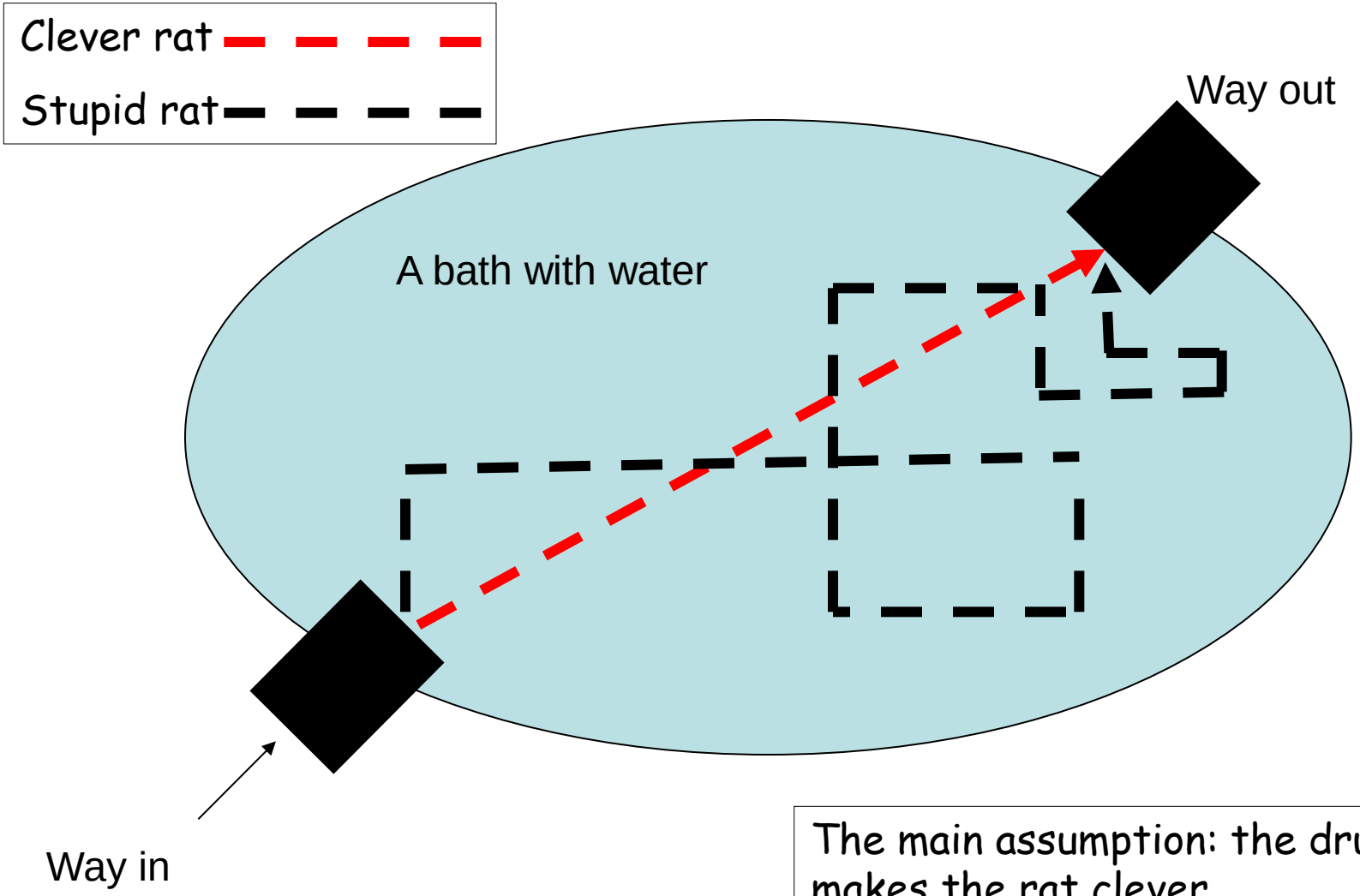
The Evaluation of the Rat

Y_i is the time inside the red zone divided by the time outside the red zone (times 100).



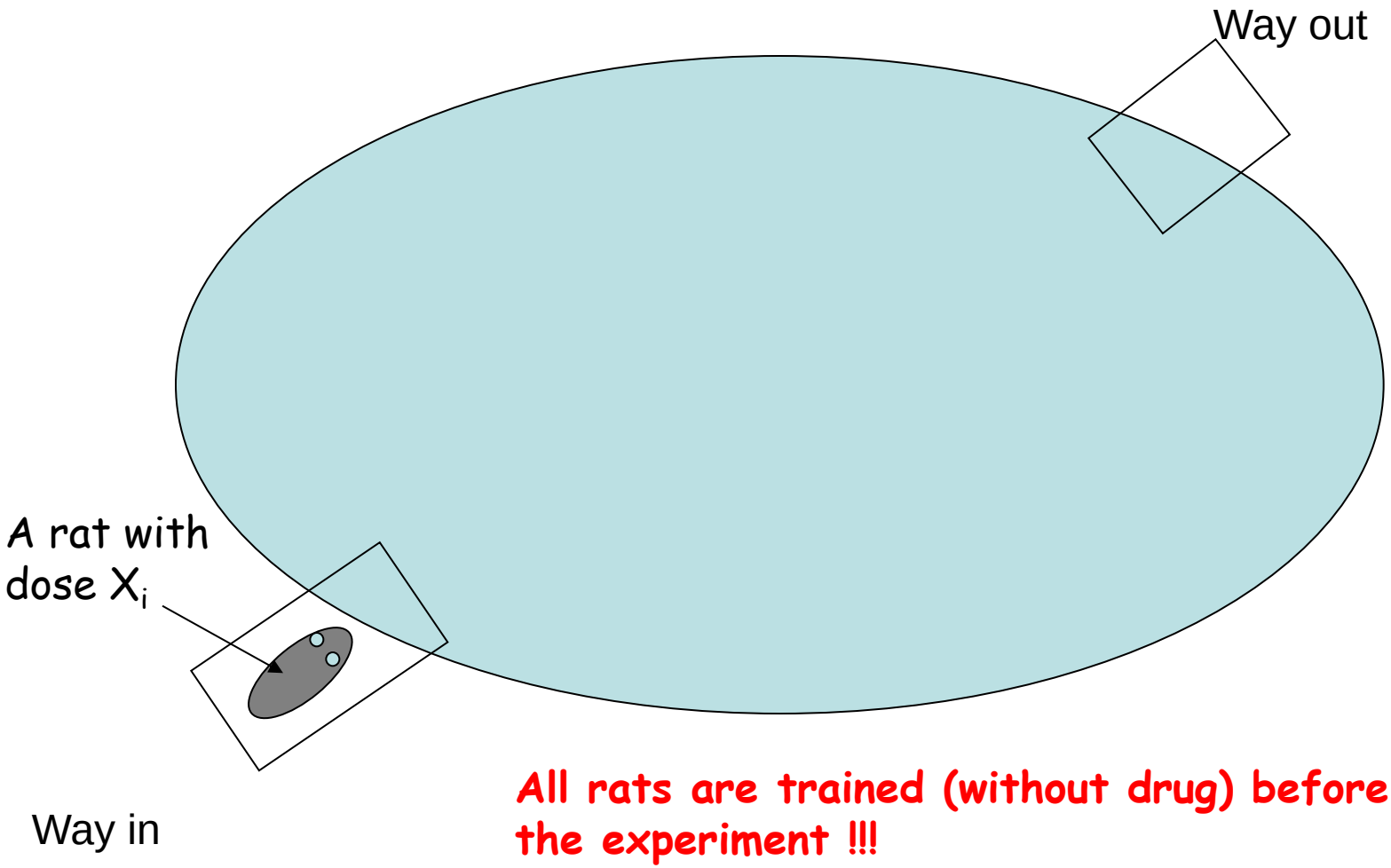
The mission of the rat: swim directly to the other side

Description of the Experiment



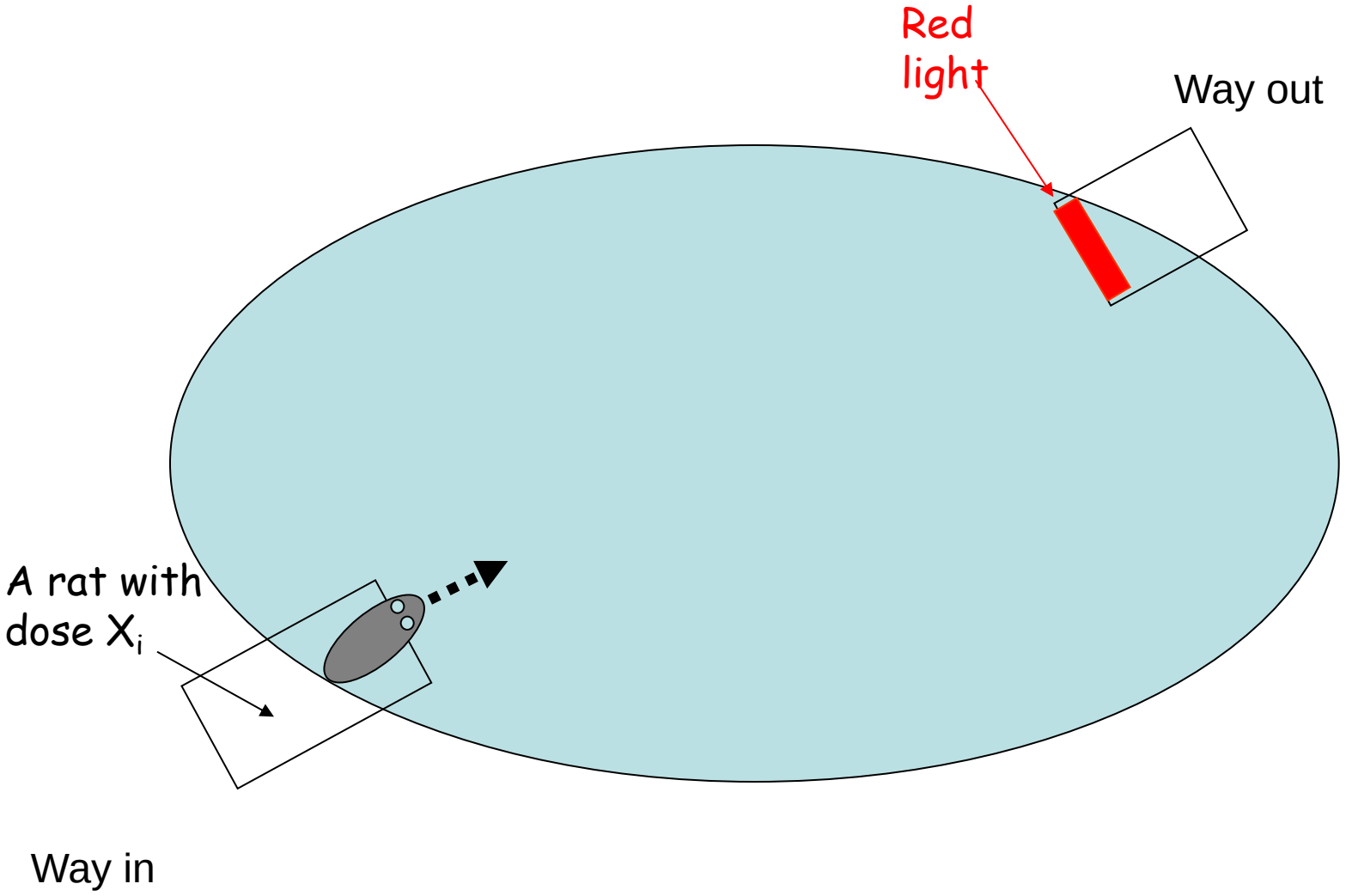
The main assumption: the drug makes the rat clever.

The Evaluation of the Rat



All rats are trained (without drug) before the experiment !!!

The Evaluation of the Rat



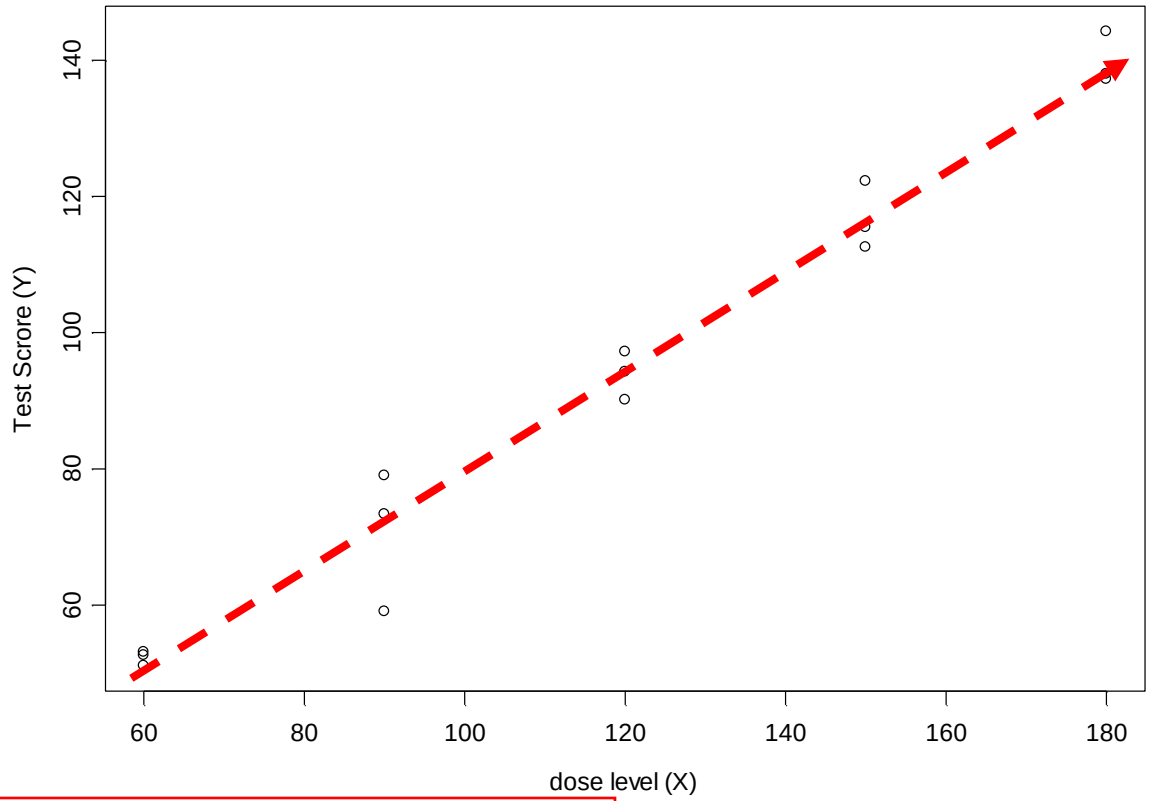
The Scientific Question

- Does the performance of the rat depend on the dose level ?

A good drug is expected to improve the rats' performance

The scientists expect that: the higher the dose the better the performance

The Data

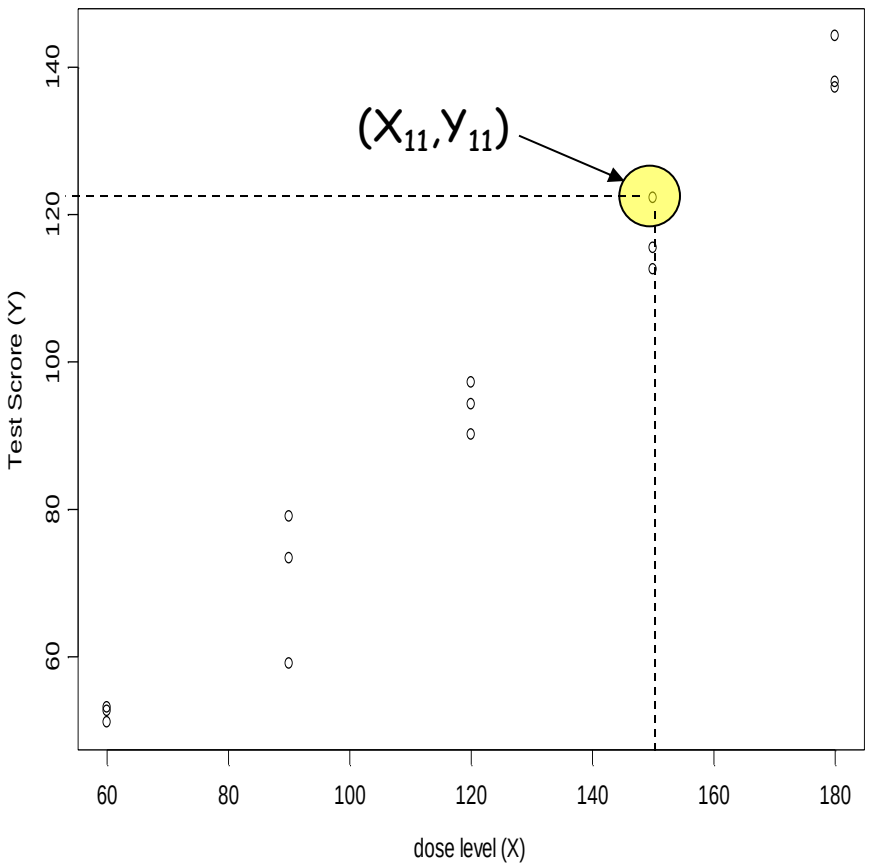


Upward trend: in general test score increases with dose level

Regression Terminology

- The test score (Y_i) is the dependent variable. It depends on the dose level (X_i).
- The dose level is called the independent variable or the predictor.
- The observation unit:

$$(X_i, Y_i),\ i=1,2,\ldots,n.$$



Data Structure

(X_1,Y_1)
(X_2,Y_2)
. .
. .
(X_i,Y_i)
. .
. .
(X_n,Y_n)

Dependent variable y_i

Dose level x_i

The dose response data

60	56.07362
60	49.45516
60	56.07840
90	74.18539
90	73.13873
90	77.35170
120	95.37789
120	93.03198
120	92.46663
150	117.61100
150	123.56117
150	119.12260
180	130.81847
180	137.31600
180	139.09742

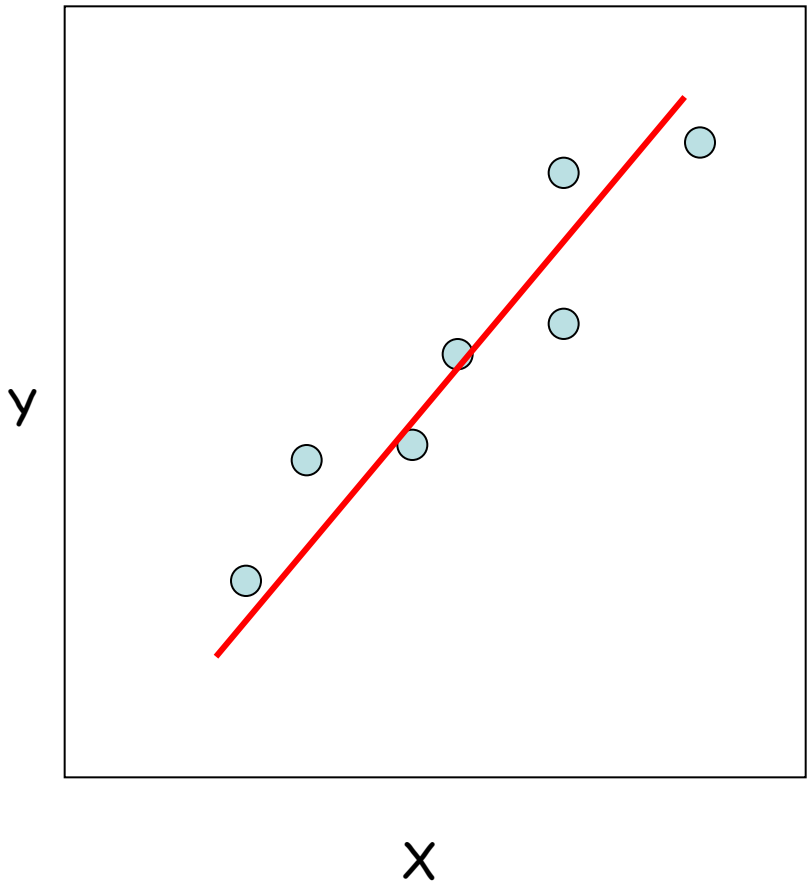
(X_{11},Y_{11})

Dose level x_i

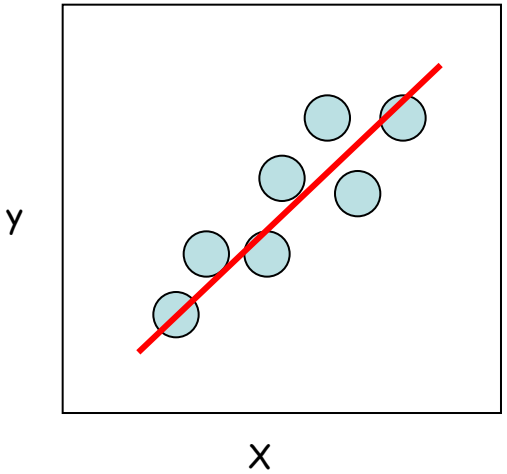
Test score y_i

What is a Simple Linear Regression Model ?

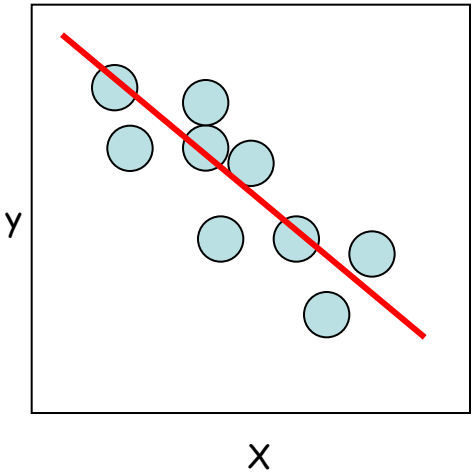
- A regression model is a statistical model which aims to describe the relationship between a predictor (the dose level) and the dependent variable (test score) with a **straight line**.



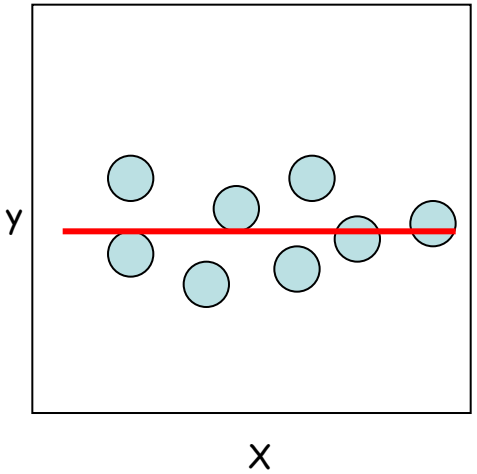
Properties of Simple Linear Regression Models : Trends



Upward trend

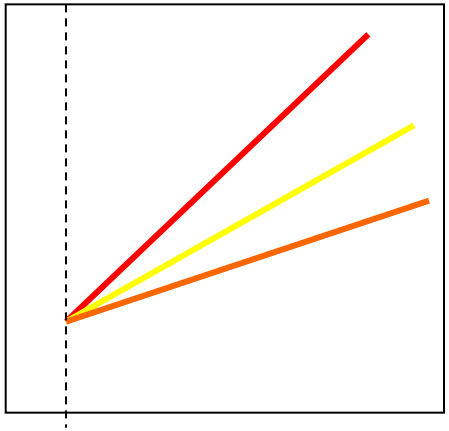


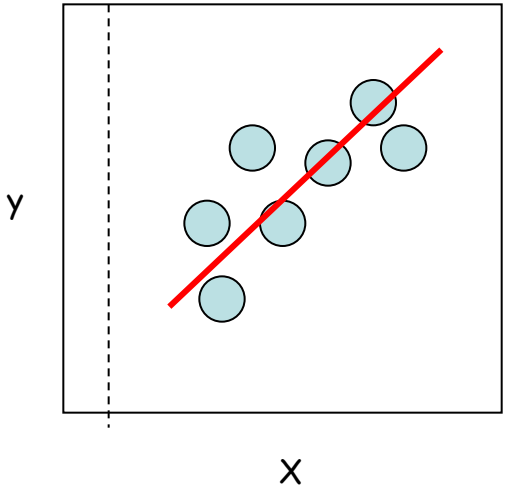
Downward trend



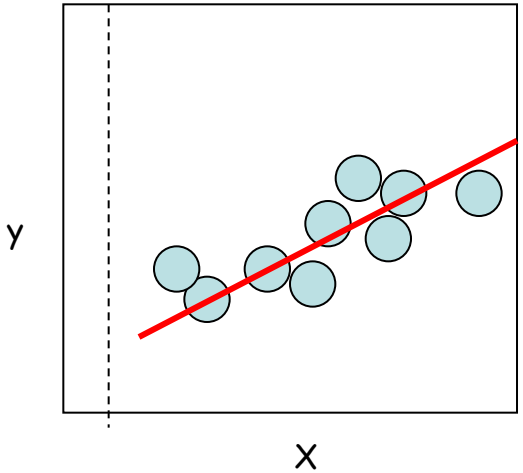
Y does not depend on X

Properties of Simple Linear Regression Models : Slope

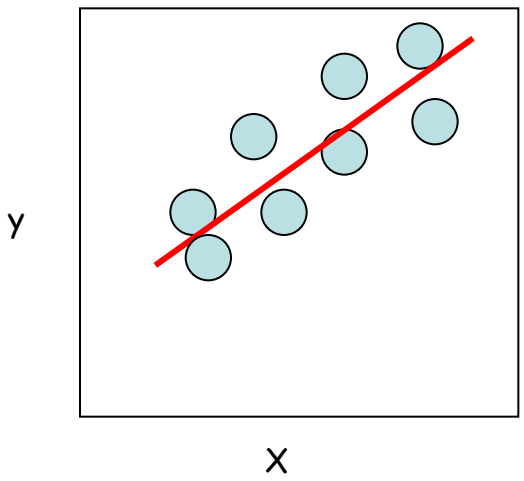


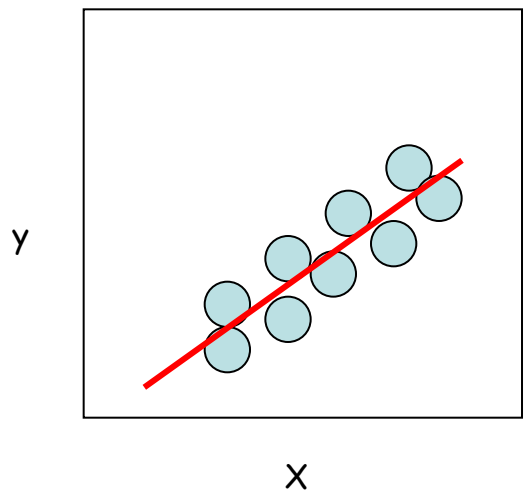


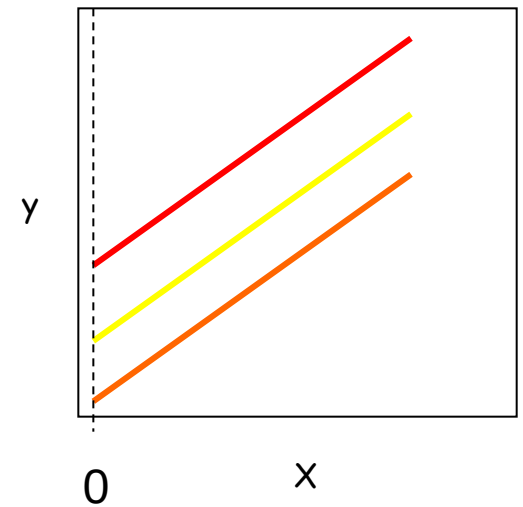
The sploe is the change in the mean of Y for a unit change in X



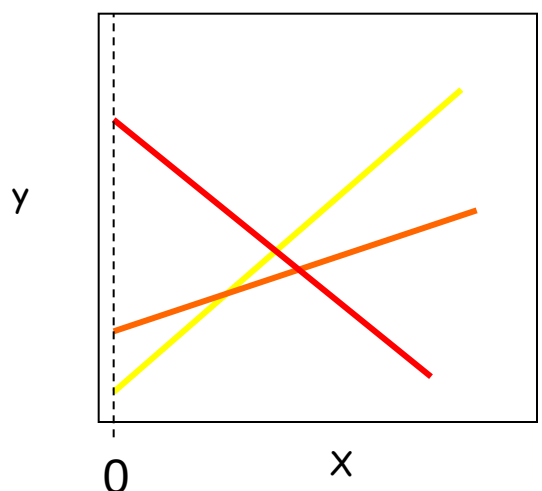
Properties of Simple Linear Regression Models : Intercept







The intercept is the highet of the regression line when x=0.



A Simple Linear Regression Model

- We assume that the relationship between the predictor and the response can be describe with the model:

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

Dependent variable Y_i

Predictor X_i

observed

Two parameters: intercept and slope

Random error

unobserved

Estimation (I)

- We need to estimate the unobserved parameters of the model:

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

$$\hat{Y}_i = a + bX_i$$

- The estimator for the random error:

$$e_i = \hat{Y}_i - Y_i$$

predicted value for the test score

(the estimator for the test score)

- **a and b are the estimators for alpha and beta**
- **e_i (the residual) is the estimator for the random error**

Regression Model for the Data

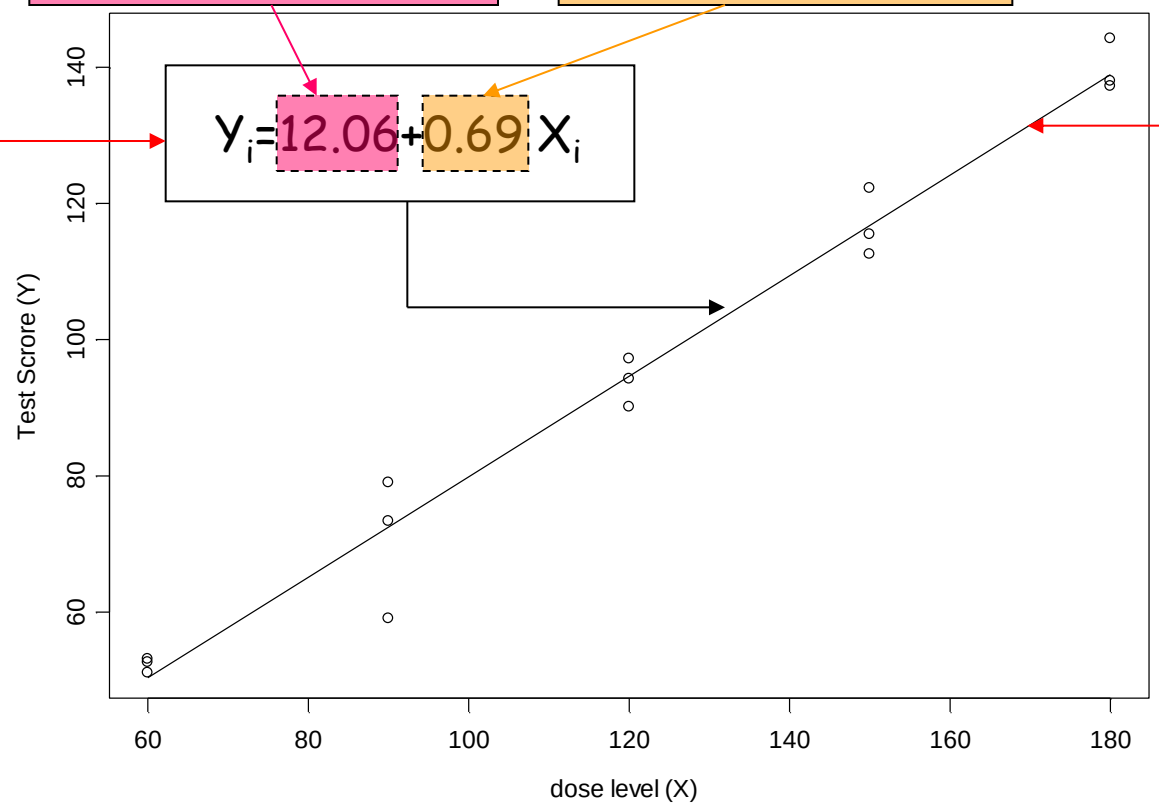
a: the estimate for alpha

b: the estimate for beta

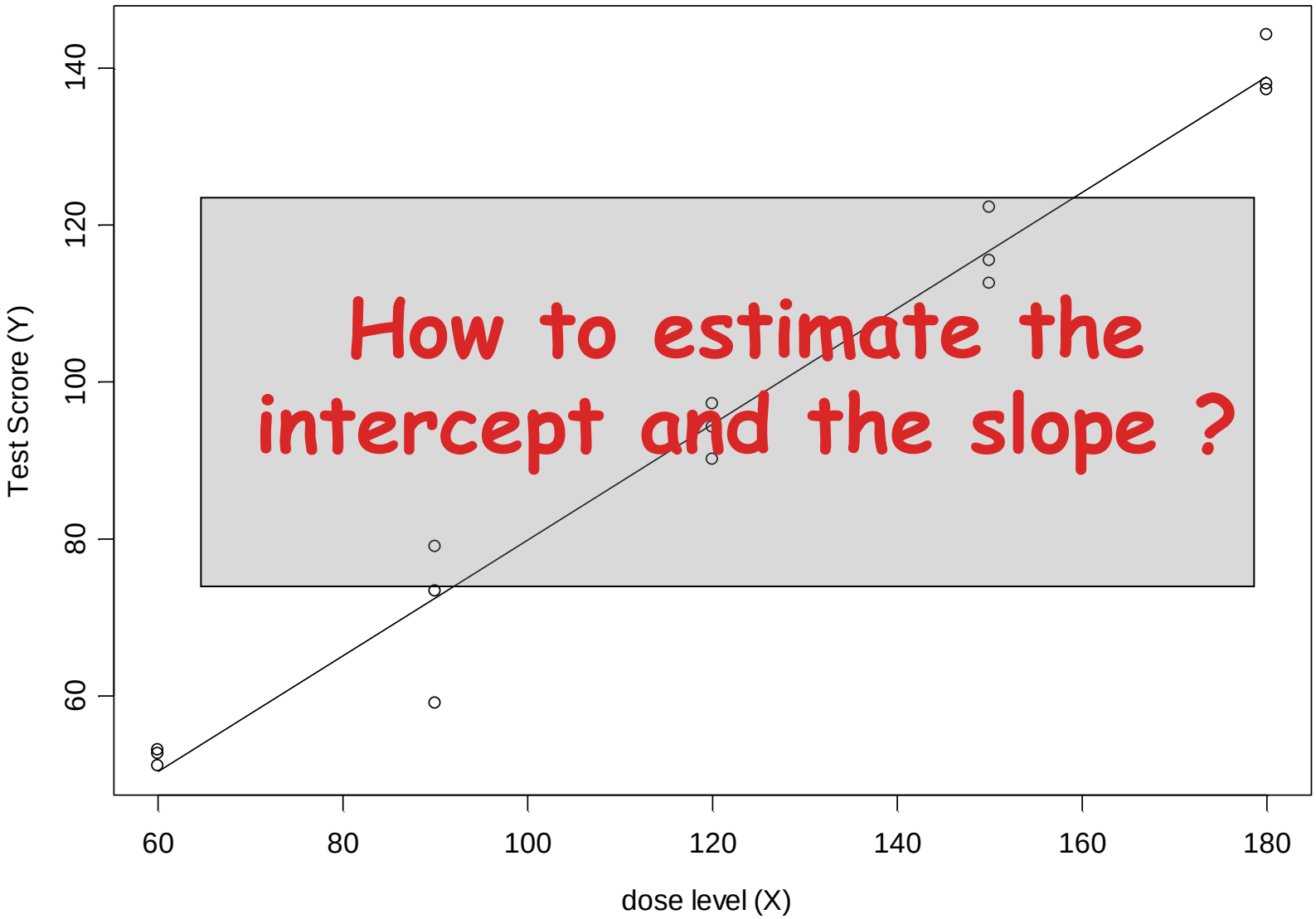
The estimated (fitted) model

$Y_i = 12.06 + 0.69\, X_i$

The regression line

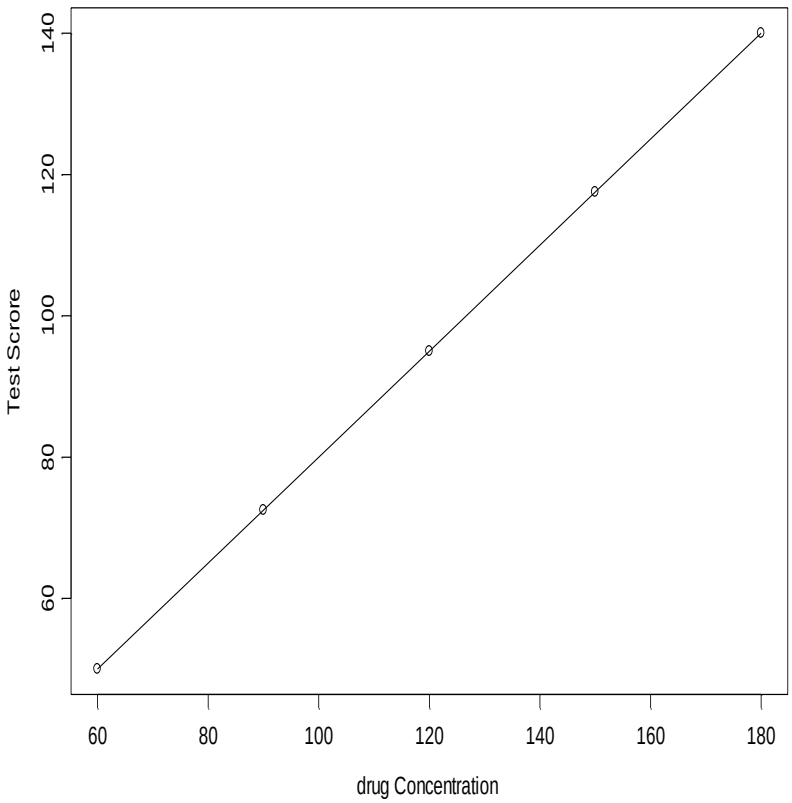


$$Y_i = \alpha + \beta X_i + \varepsilon_i$$ The model

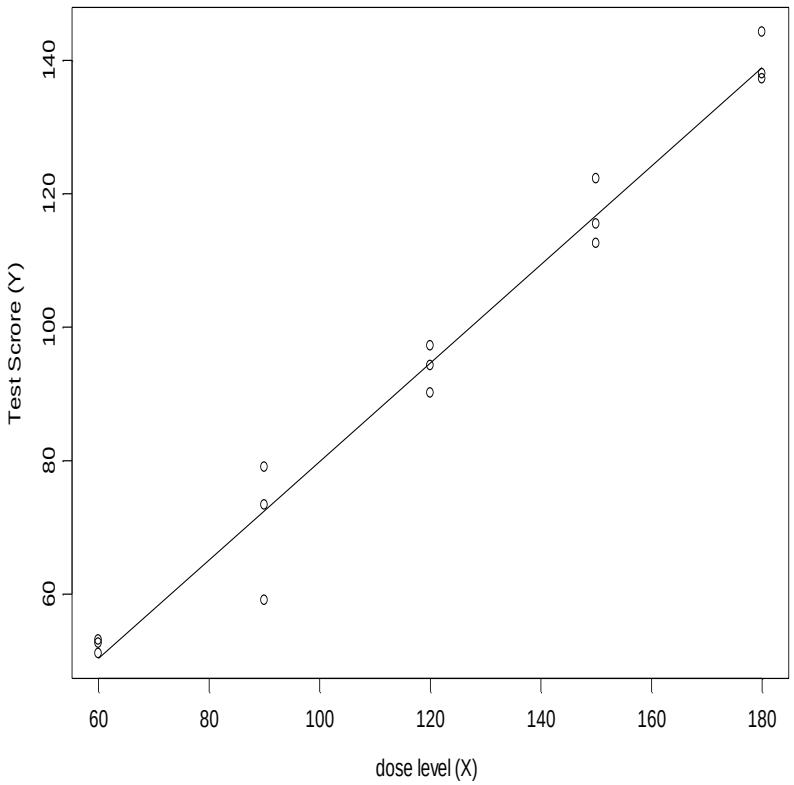

How to estimate the intercept and the slope ?
Test Scrore (Y)
60
80
100
120
140
dose level (X)
60
80
100
120
140
160
180

Regression Model for the Data

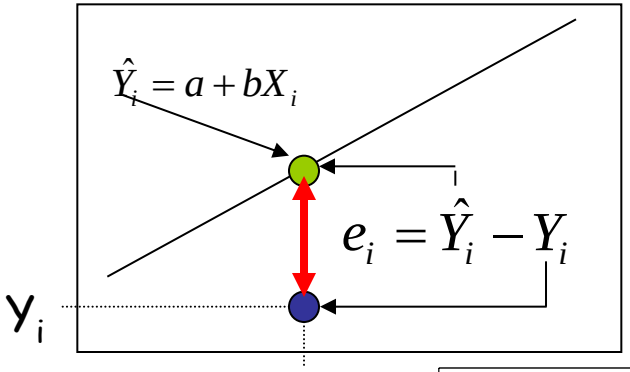
A EXAMPLE OF PERFECT FIT…

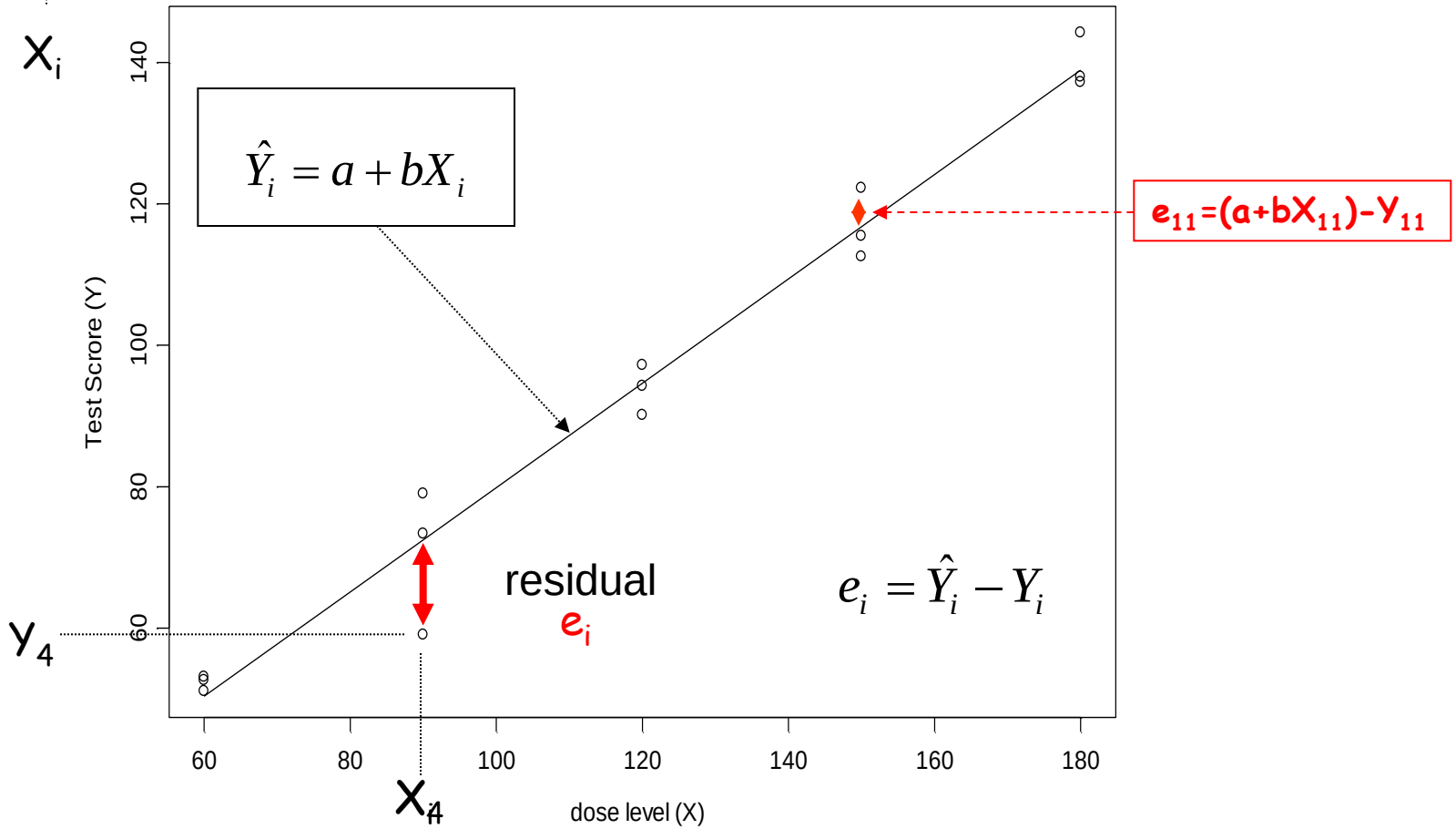


….BUT NOBODY IS PERFECT !



The Residuals

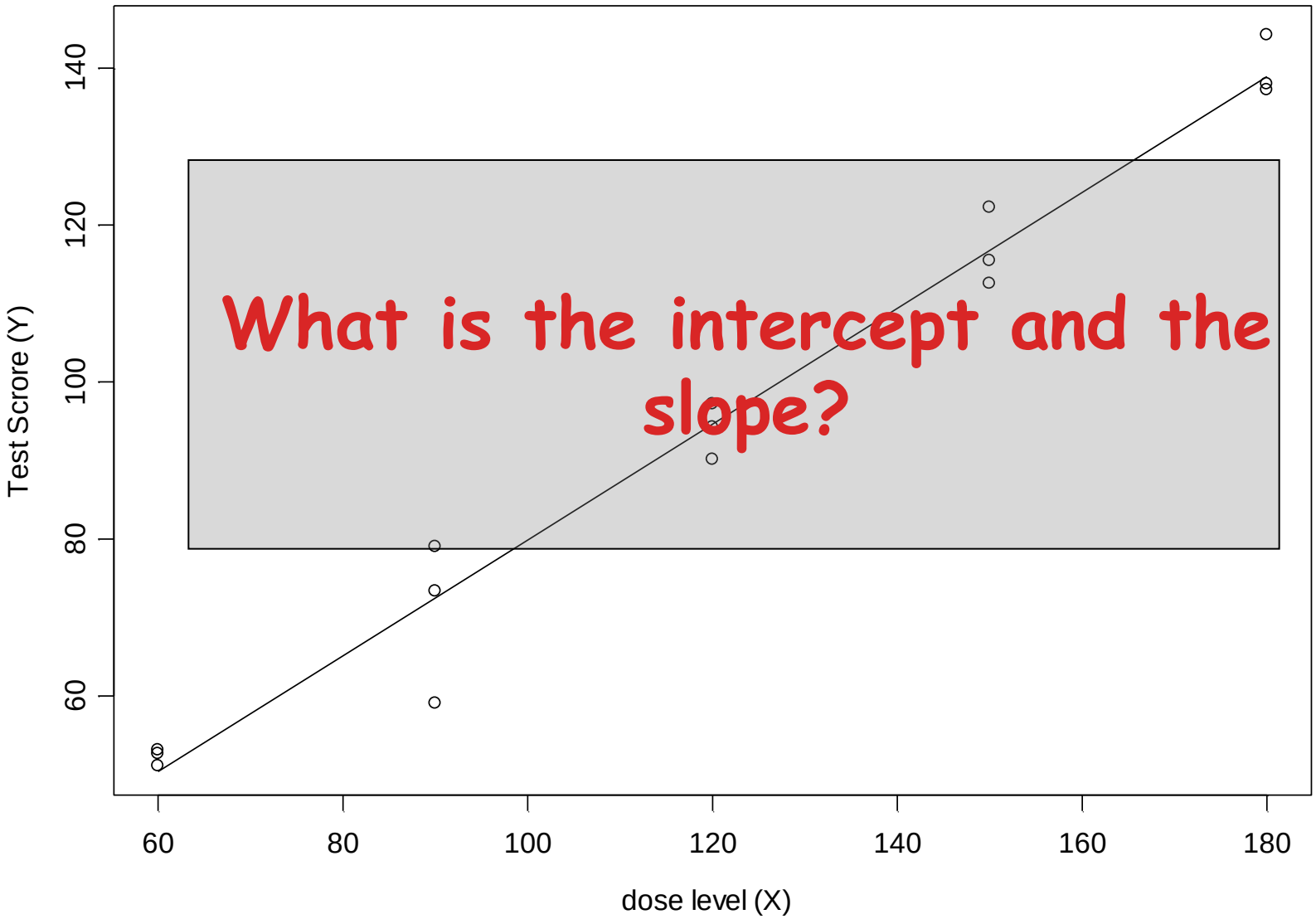




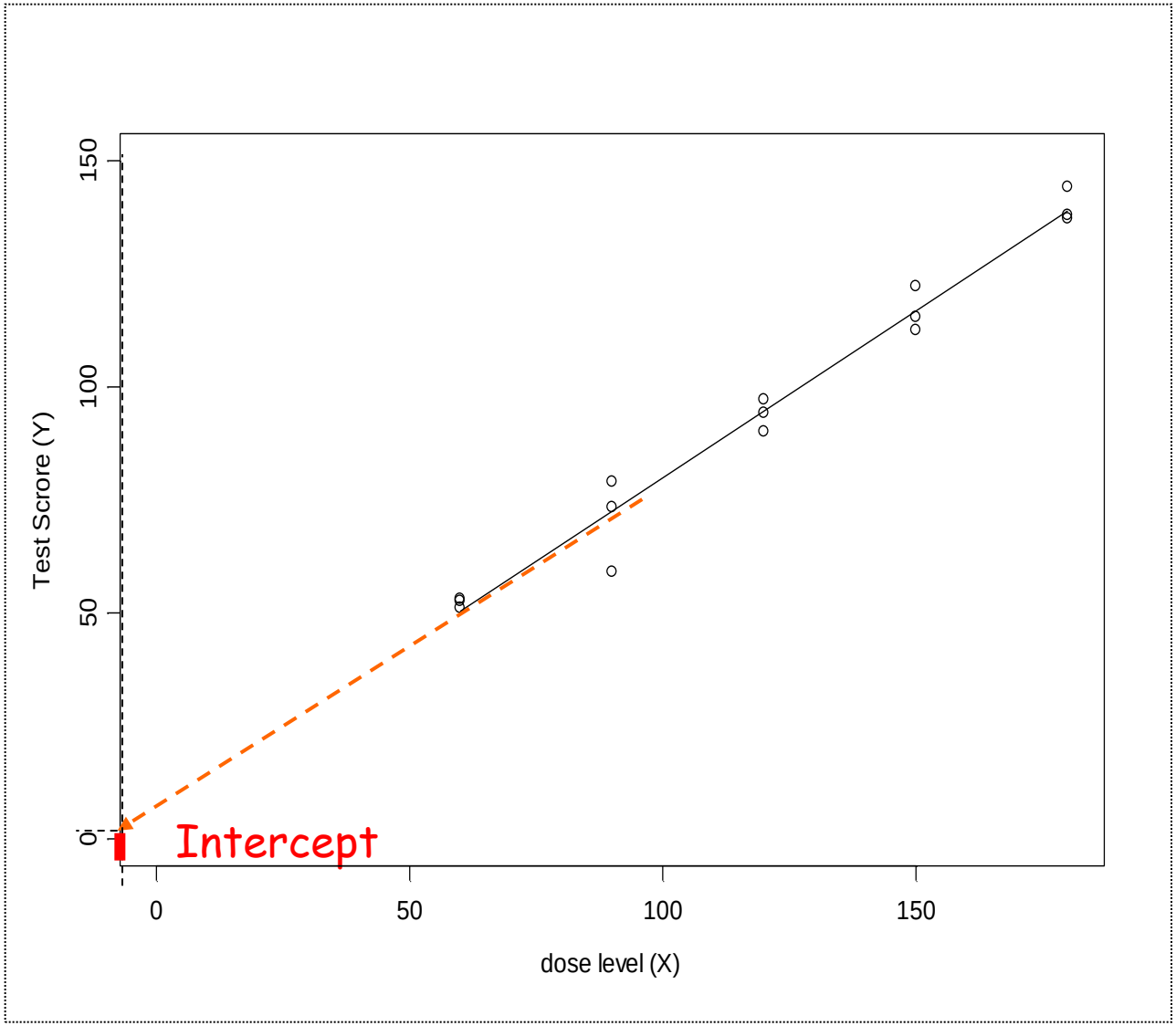
Estimation (II): The Least Squares Criterion

- How to estimate the intercept and slope?
- We want that the fitted model (the line which describes the relationship between Y and X) will be "close" to the data.
- The residual sum of squares = sum (residual)^2.
- The least squares criterion: choose intercept and slope which minimize the residual sum of squares

$$RSS = \sum_{i=1}^{n} \left(\hat{Y}_i - Y_i\right)^2$$


What is the intercept and the slope?
Test Scrore (Y)
60
80
100
120
140
dose level (X)
60
80
100
120
140
160
180

Interpretation: The Intercept



The intercept is the predicted test score for dose level zero:

For $X_i=0$ we have:

Predicted test score=12.09+0.

Interpretation: The Slope

Suppose that we have two rats: the first received a dose of 100 and the second dose of 101.

WHAT IS THE DIFFERENCE BETWEEN THE PREDICTED VALUES OF THE TWO RATS ?

- Dose level 100:

Predicted value=12.09+0.69*100

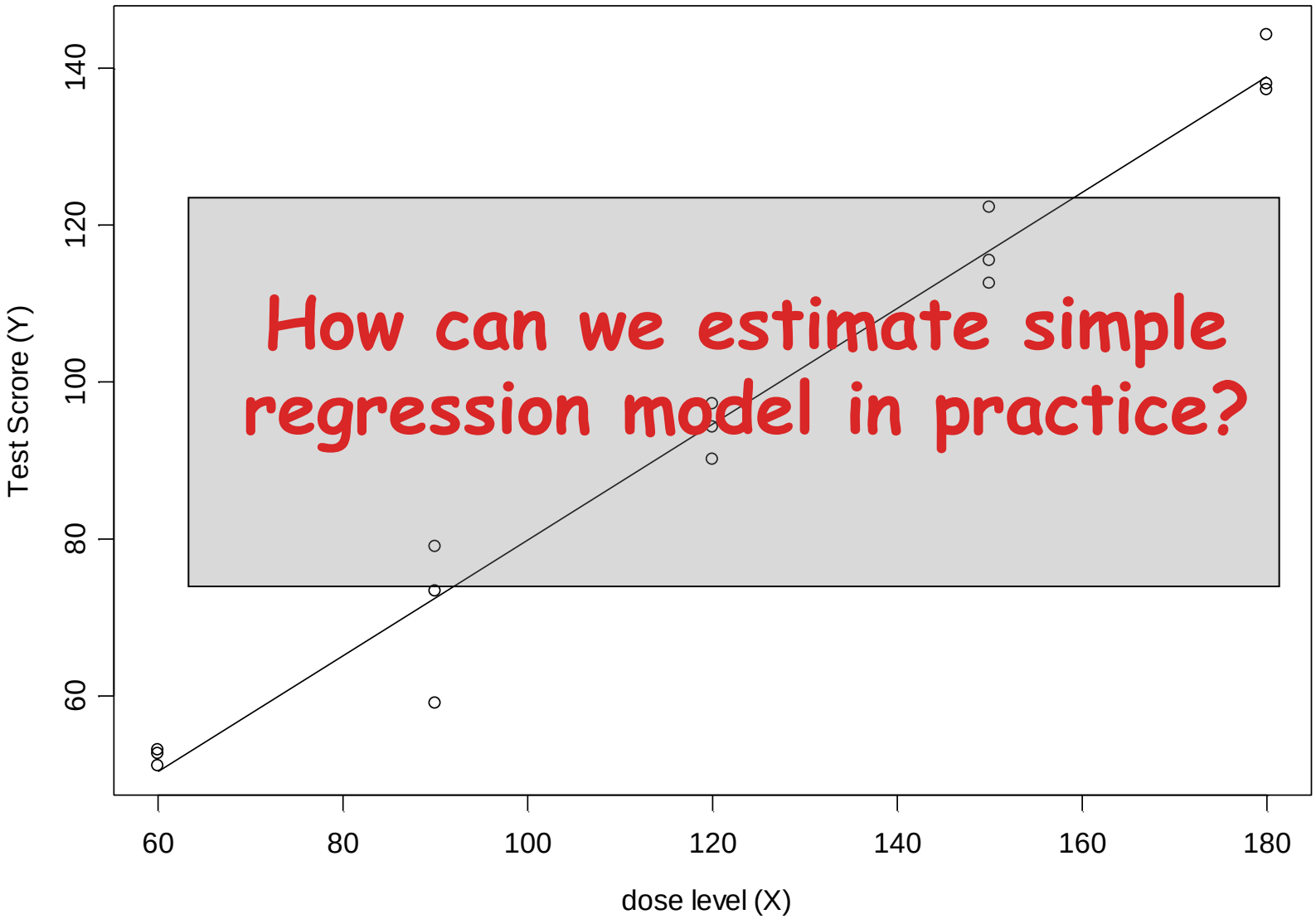
- Dose level 101:

Predicted value=12.09+0.69*101

(12.09+0.69*101)-(12.09+0.69*100)=0.69

The difference is equal to 0.69 which is the value of b

The slope is the change in the (expected) response level for a unit change in the predictor


How can we estimate simple regression model in practice?
Test Scrore (Y)
60
80
100
120
140
dose level (X)
60
80
100
120
140
160
180

Part 2

Fitting simple linear regression in R using the `lm()` function in R

The Data in R

```
Dose <- c(60,60,60,90,90,90, 120,120,120,150,150,150,180,180,180)
Score <- (56.07362,49.45516,56.07840,74.18539,73.13873,77.35170,95.37789,93.03198,
          92.46663,117.61100,123.56117,119.12260,130.81847,137.31600,139.09742)
dose.data <- cbind(Dose, Score)
print(dose.data)
       Dose     Score
 [1,]   60  56.07362
 [2,]   60  49.45516
 [3,]   60  56.07840
 [4,]   90  74.18539
 [5,]   90  73.13873
 [6,]   90  77.35170
 [7,]  120  95.37789
 [8,]  120  93.03198
 [9,]  120  92.46663
[10,]  150 117.61100
[11,]  150 123.56117
[12,]  150 119.12260
[13,]  180 130.81847
[14,]  180 137.31600
[15,]  180 139.09742
```

Name of the data

Dependent variable

Predictor

The function lm( ) in R

- Simple linear regression model can be fitted in R using the function lm( ).

- The model statement:

Score ~ Dose

Example of R script for function lm( )

```
fit.dose <- lm(Score ~ Dose,data = dose.data)
```

Dependent variable

Predictor

Fitting the model in R

```
Dose <- c(60,60,60,90,90,90, 120,120,120,150,150,150,180,180,180)
Score <- (56.07362,49.45516,56.07840,74.18539,73.13873,77.35170,95.37789,93.03198,
          92.46663,117.61100,123.56117,119.12260,130.81847,137.31600,139.09742)
dose.data <- cbind(Dose, Score)
print(dose.data)
       Dose     Score
 [1,]   60  56.07362
 [2,]   60  49.45516
 [3,]   60  56.07840
 [4,]   90  74.18539
 [5,]   90  73.13873
 [6,]   90  77.35170
 [7,]  120  95.37789
 [8,]  120  93.03198
 [9,]  120  92.46663
[10,]  150 117.61100
[11,]  150 123.56117
[12,]  150 119.12260
[13,]  180 130.81847
[14,]  180 137.31600
[15,]  180 139.09742
```

Name of the data

Dependent variable

Predictor

```
> fit.dose <- lm(Score ~ Dose)

> summary(fit.dose)
```

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

Output

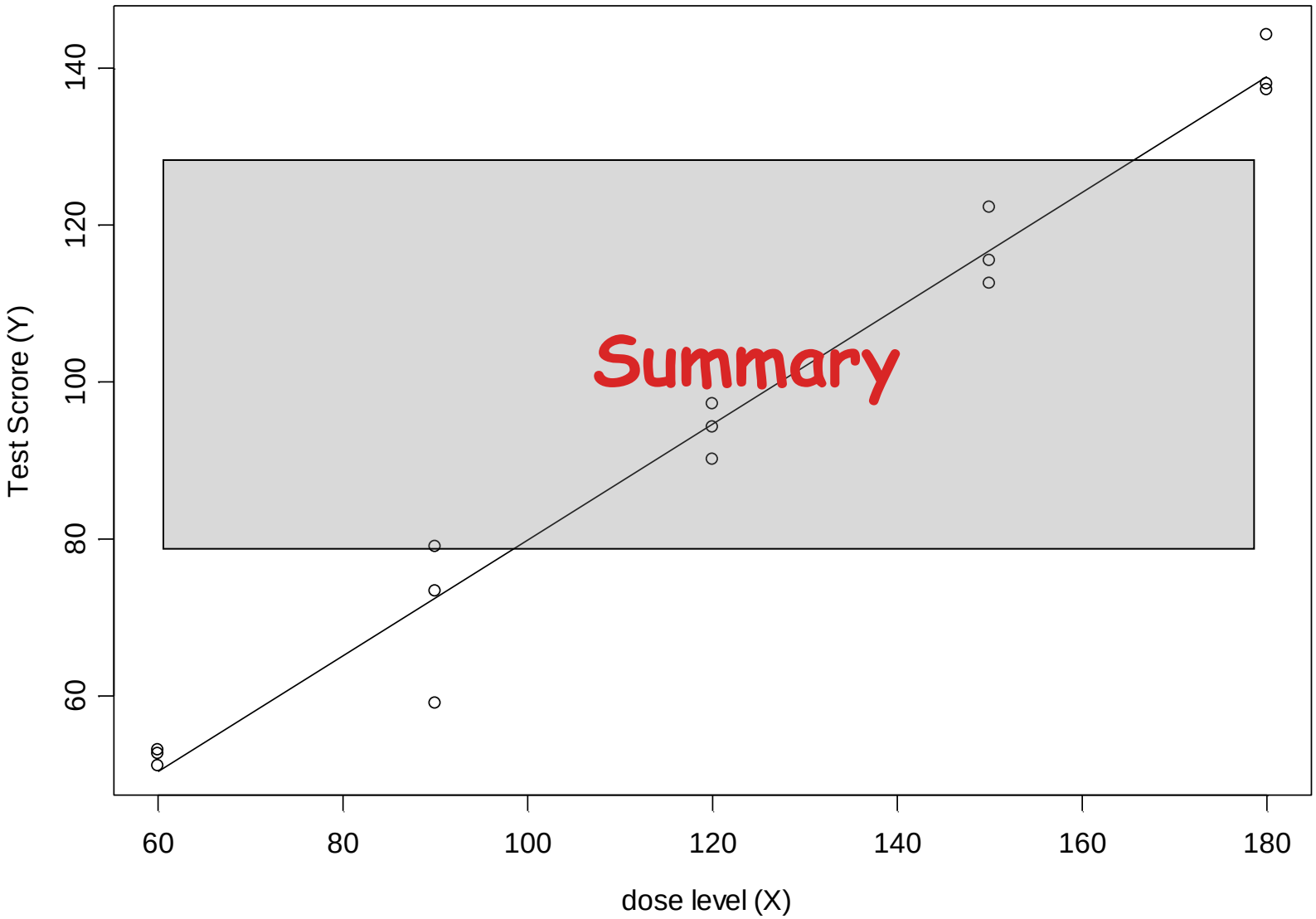
ESTIMATION

INFERENCE

```
Coefficients:
            Estimate      Std. Error t value Pr(>|t|)
(Intercept) 12.06329      2.71389    4.445 0.000661 ***
Dose         0.69652      0.02132   32.666 7.28e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.504 on 13 degrees of freedom
Multiple R-squared: 0.988,      Adjusted R-squared: 0.987
F-statistic:  1067 on 1 and 13 DF,  p-value: 7.279e-14
```

The intercept: what is the test score for dose=0

The slope: how much the response change for a unit change in the predictor


Summary
Test Scrore (Y)
60
80
100
120
140
dose level (X)
60
80
100
120
140
160
180

Technical Details (Estimation)

- A simple linear regression model has the form:

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

- α and β are the parameters in the model and ε is the random error.

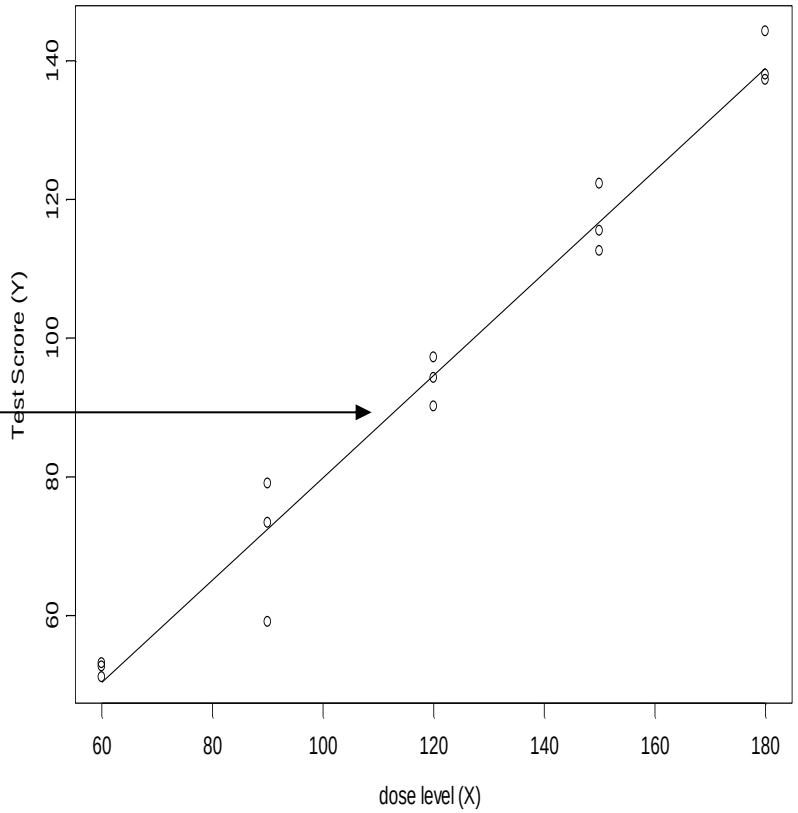
- We can estimate α and β by minimizing the residual sum of squares

$$RSS = \sum_{i=1}^{n} \left(\alpha - \beta X_i - Y_i\right)^2$$

Technical Details (Estimation)

- The estimated model

$$\hat{Y}_i = a + bX_i$$



- The residual

$$e_i = \hat{Y}_i - Y_i$$

Technical Details (Estimation)

We assume that the relationship between Y_i and X_i can be described with a statistical model $Y_i = \alpha + \beta X_i + \varepsilon_i$

We assume that the random error ε is normally distributed.	$\varepsilon \sim N(0, \sigma^2)$
The mean of ε is equal to zero	$E(\varepsilon_i) = 0$
The conditional mean of Y_i (given the value of X_i)	$E(Y_i \mid X_i) = \alpha + \beta X_i$
The estimator for the conditional mean of Y_i (the fitted model=the regression line)	$\hat{E}(Y_i \mid X_i) = a + bX_i = \hat{Y}_i$
The residual: the estimator for ε	$e_i = \hat{Y}_i - Y_i$
Least square criterion: choose a and b that minimize the residuals sum of squares	$RSS = \sum_{i=1}^{n} (\alpha - \beta X_i - Y_i)^2$

Part 3
Model diagnostic

Simple regression model and it's assumptions

In this SLW we focus on model diagnostic. We consider the following linear regression model

$$Y_i = \alpha + \beta \times X_i + \varepsilon_i$$

The random error is **assumed** to be normal distributed:

$$\varepsilon_i \sim N(0, \sigma^2)$$

We also assume that the variance is constant, i.e., $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are came from the normal distribution with mean zero and equal variances.

How to check the model assumptions? (1)

- The random error, ε_i, is unknown but we can estimate ε_i with the residuals

$$e_i = Y_i - \hat{Y}_i$$

Observed Predicted

- The residuals can be used in order to check the model asaumptions.

- We focus on two things:
  1) the distribution of e_i
  2) the variability of e_i

How to check the model assumptions? (2)

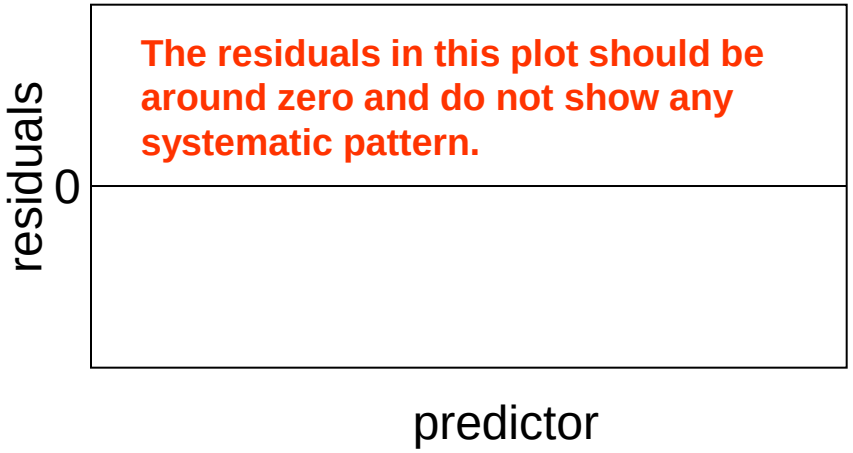
- We assume that the mean of Y_i is linear with respect to X:

$$E(Y)_i = \alpha + \beta \times X_i$$

- This is true only if

$$E(\varepsilon_i) = 0$$

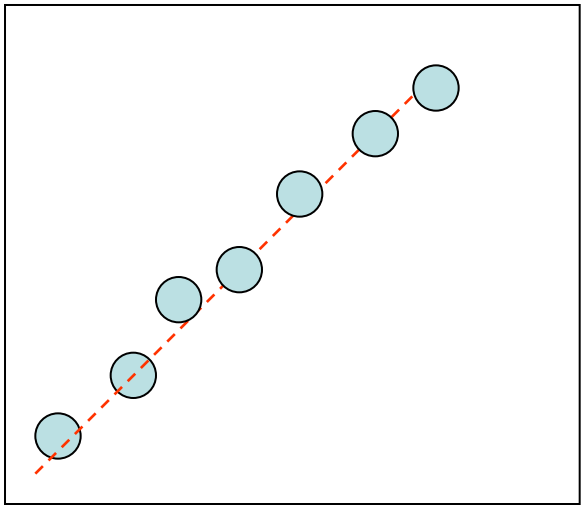
- Once again, the residuals can be used in order to check the linearity asaumption .



Assumption 1: The distribution of e_i

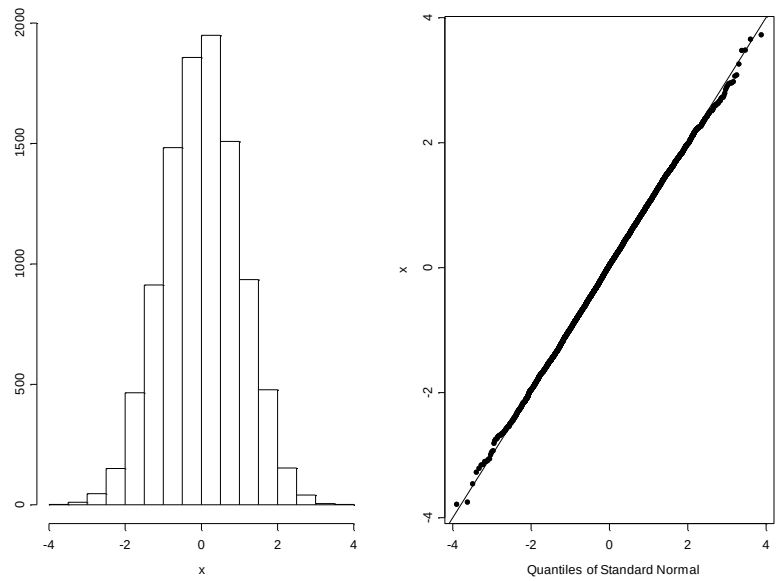
- The distribution of e_i is expected to be normal with mean zero and variance σ^2
- qq-normal plot (or normal probability plot) is a graphical tool that can be used in order to asses the normality assumption.

If the normalty assumption holds we expect qq-normal plot will be a srtight line.



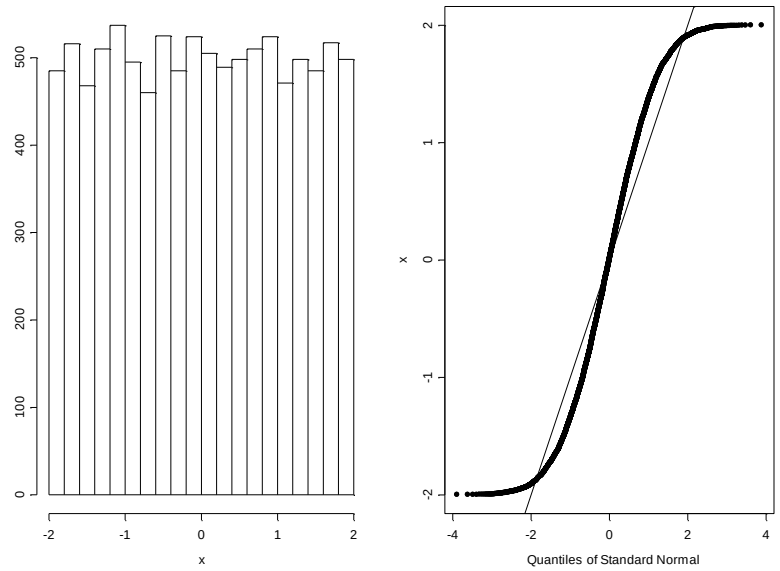
Example of qqnormal plot form N(0,1)

- Sample of 10000 observations from N(0,1)
- The qqnormal plot is a stright line.
- If the random error ε_i is normal distributed, the qqnormal plot of the residuals should be a stright line.



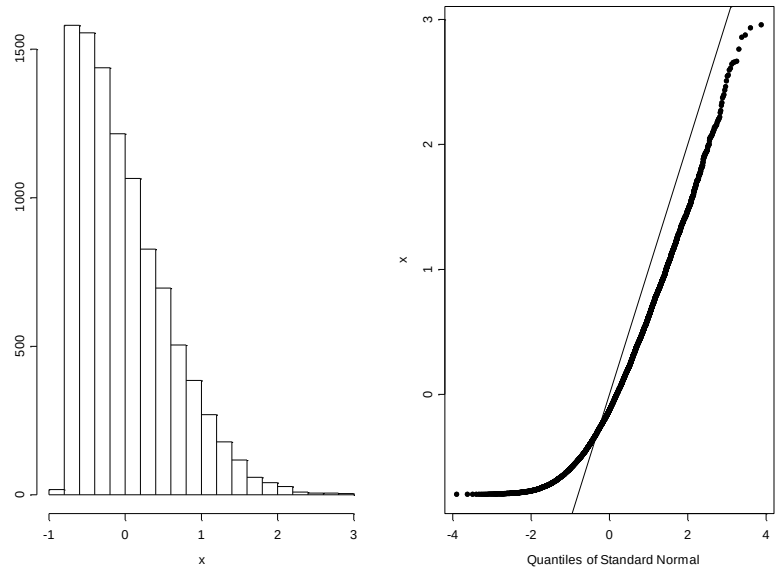
Example of qqnormal polt from havy tailed distribution

- Sample of 10000 observations from U(-2,2).
- S shape of the qqnormal plot.
- This is an example of a symatric distribution with more observations (relativly to the normal distribution) at the tails.



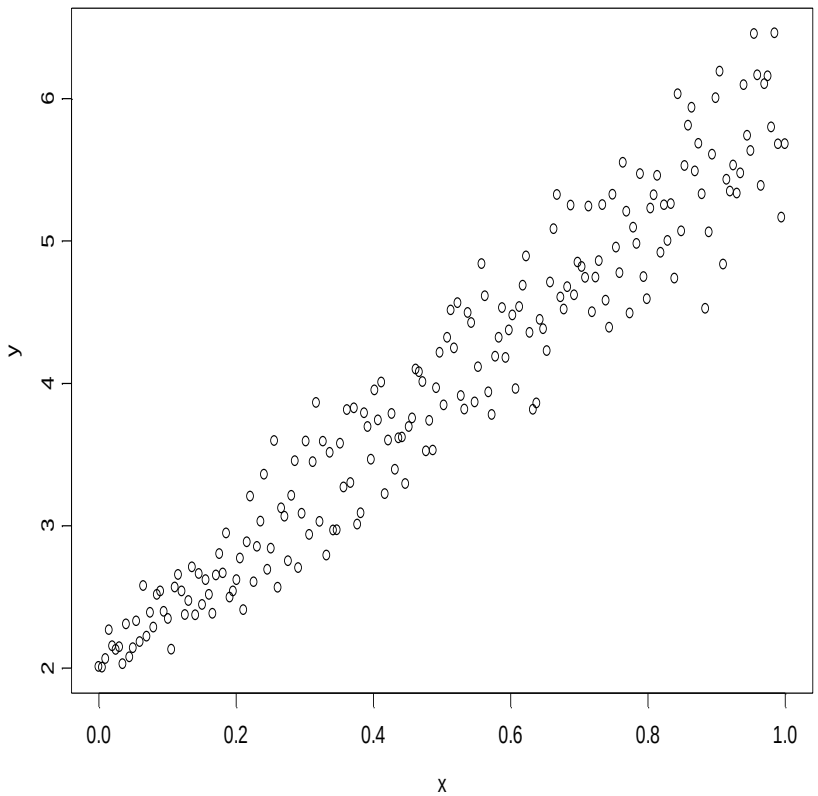
Example of qqnormal polt from skwed distribution

- Sample of 10000 observations from a skewed distribution.
- The distribution is sewed to the right and the points in the qqplot are not follow the stright line.



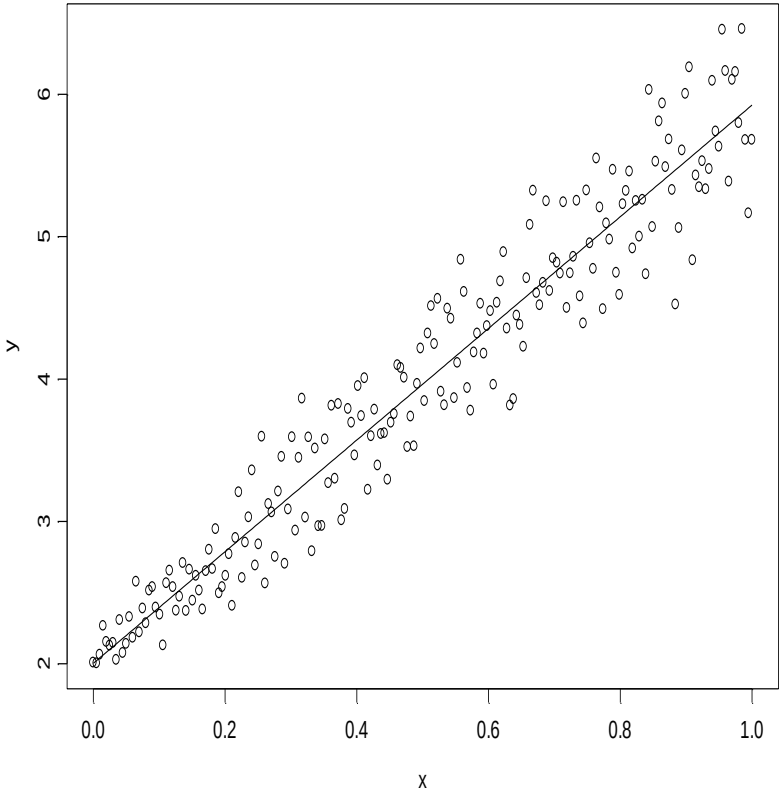
Assumption 2: Constant variance

- This is an example of a dataset in which the variance is not constant.
- The variance increases when the value of X increases.
- However, there is a linear relatioship between the predictor and the response.



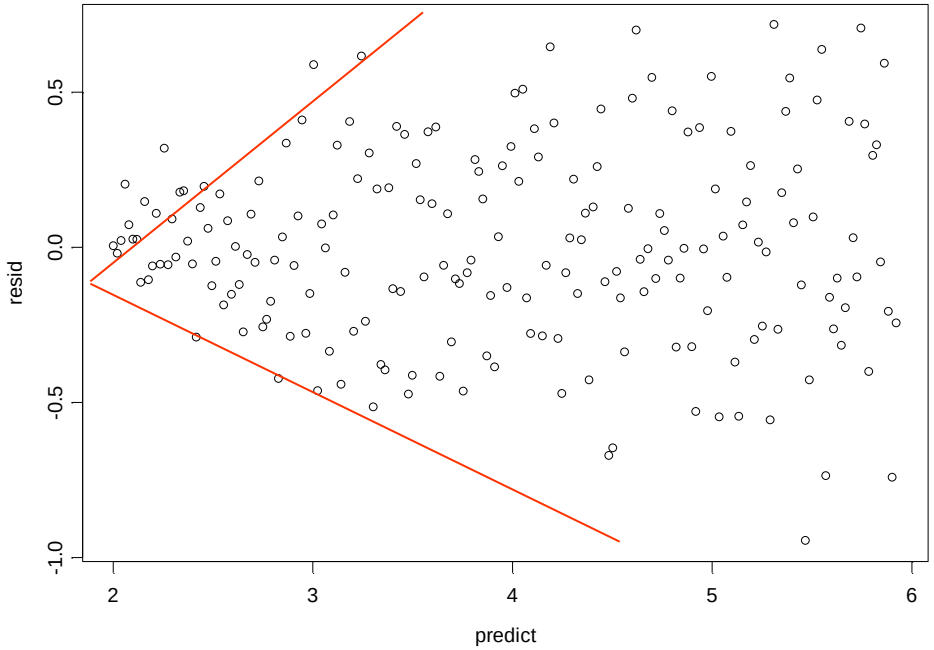
The data and the resrssion line

- The model seems to fit the data well in the sense that is captures that structure of the mean.



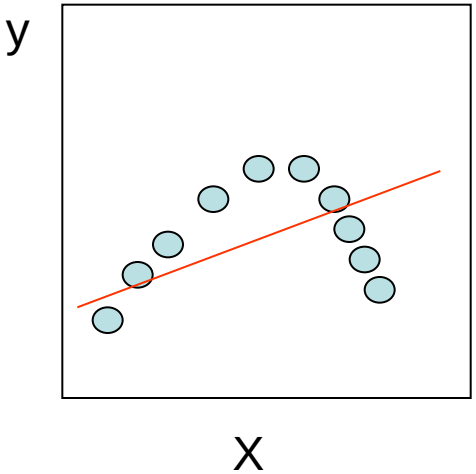
Residuals plot: Residuals versus the predicted values

- In this plot we can see clearly a pattern. As the predicted values increase the variability among the residuals increase (a "megaphone" shape).

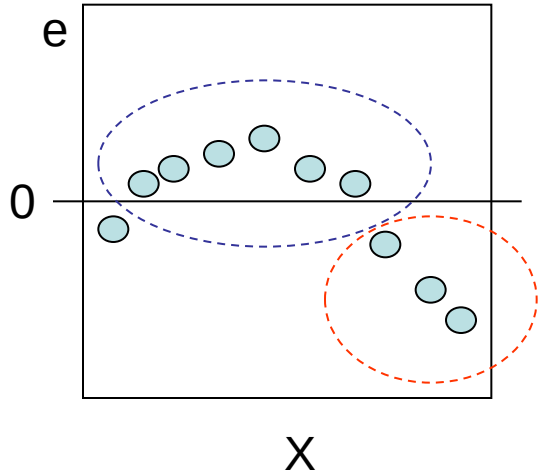


Assumptio 3: Linearity

The data



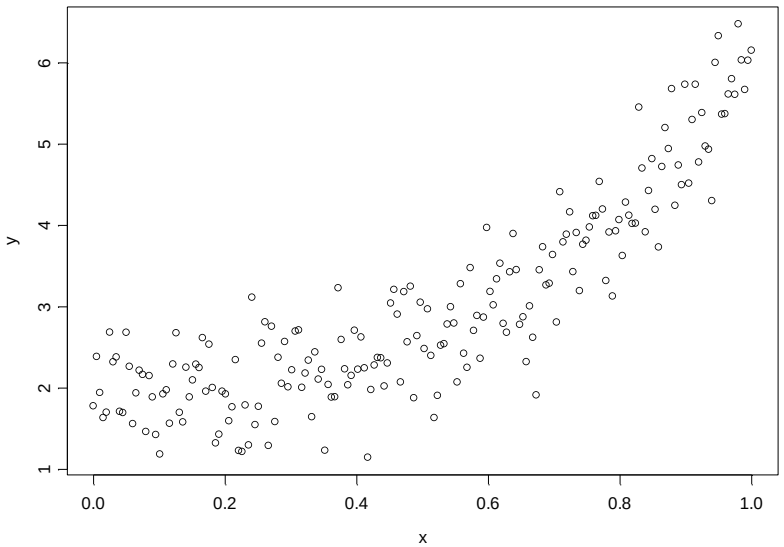
e versus x



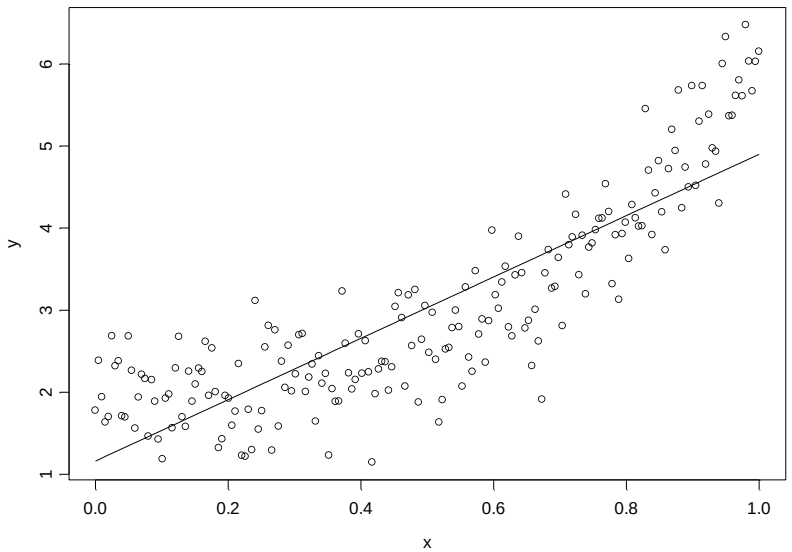
Example of systematic pattern in the residual plot.

The scatterplot of the data reveals that the association between the response and the predictor is not linear. The residuals plot (in the right) reveals a clear pattern among the residuals which depends on the value of X.

Systematic patterns



data

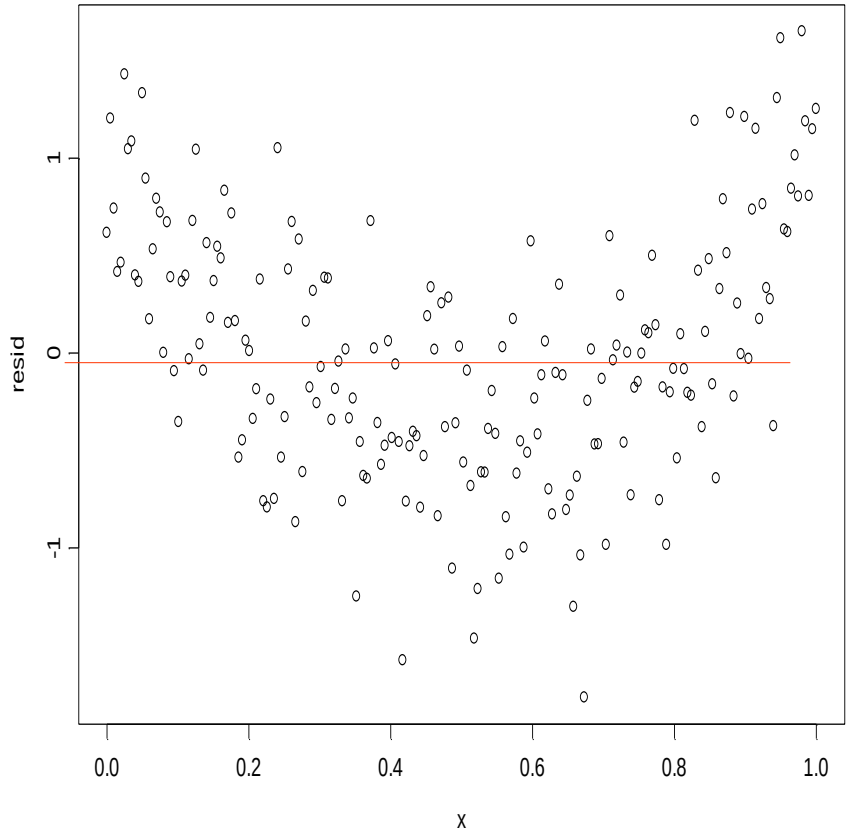


data and fitted model

The model underestimates the value of Y when the value of X is relatively small or large.

Linearity: Residuals plot

This plot reveals clear systematic pattern among the residuals: the residuals are positive for small and large value of X and negative in the middle. This means that there is structure in the data that the linear regression model did not capture.



Bottom line about model diagnostic

- We let the residuals to tell us the story.
- Departure from model assumptions (constant variance, normality and linearity) can be investigated using qq-plot and residuals plots.

Part 4
Model diagnostic using R

Fitting the model in R

```
Dose <- c(60,60,60,90,90,90, 120,120,120,150,150,150,180,180,180)
Score <- (56.07362,49.45516,56.07840,74.18539,73.13873,77.35170,95.37789,93.03198,
           92.46663,117.61100,123.56117,119.12260,130.81847,137.31600,139.09742)
dose.data <- cbind(Dose, Score)
print(dose.data)
       Dose     Score
 [1,]   60  56.07362
 [2,]   60  49.45516
 [3,]   60  56.07840
 [4,]   90  74.18539
 [5,]   90  73.13873
 [6,]   90  77.35170
 [7,]  120  95.37789
 [8,]  120  93.03198
 [9,]  120  92.46663
[10,]  150 117.61100
[11,]  150 123.56117
[12,]  150 119.12260
[13,]  180 130.81847
[14,]  180 137.31600
[15,]  180 139.09742
```

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

```
> fit.dose <- lm(Score ~ Dose)
```

Diagnostic plots

```
> par(mfrow = c(2,2))
> plot(fit.dose)
```

This statemnt produces the plot of the residuals versus the predicted values (to check if the variance is constant), qqnormal plot (to check normality), scale-location plot (to check if the variance is constant), the plot of residuals versus leverage (to check if there is an influencial observation)

The output

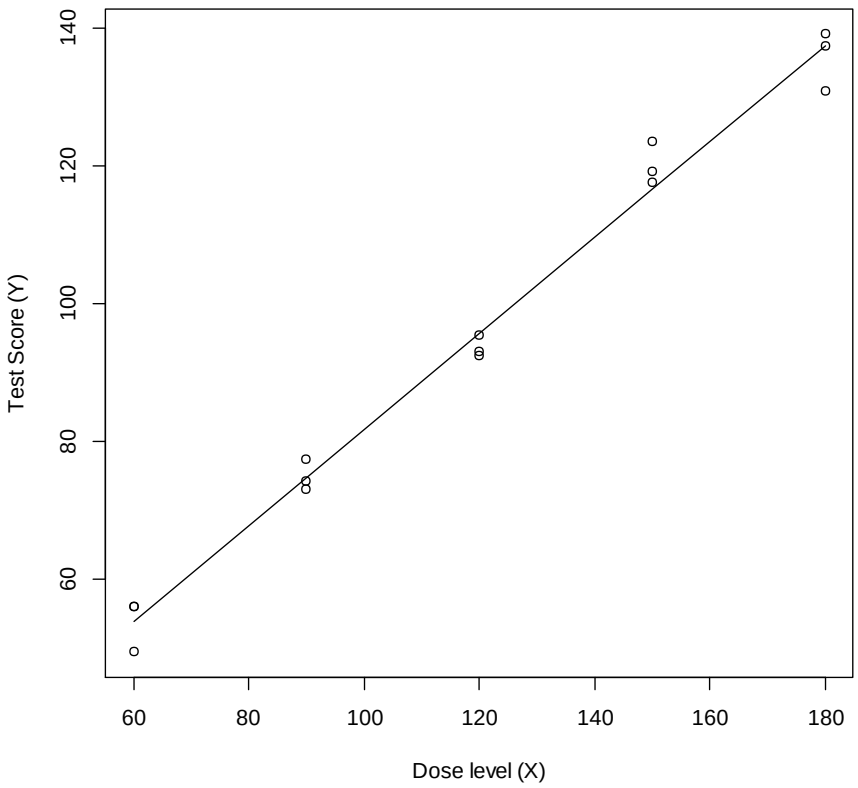
```
>summary(fit.dose)
Call:
lm(formula = Test_score ~ Dose_level, data = dose)

Residuals:
   Min      1Q Median      3Q     Max
-6.619 -2.113 -0.121  2.221  7.020

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.06329    2.71389   4.445 0.000661 ***
Dose_level   0.69652    0.02132  32.666 7.28e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.504 on 13 degrees of freedom
Multiple R-squared: 0.988,      Adjusted R-squared: 0.987
F-statistic:  1067 on 1 and 13 DF,  p-value: 7.279e-14
```

INFERENCE

ESTIMATION

Data and predicted model

```
>plot(Dose,Score,
      ylab = "TestScore (Y)",
      xlab = "Doselevel (X)")
>x <- Dose
>y <- fit.dose$fit
>lines(x,y)
```



The output

ANOVA Table:

```
> aov(fit.dose)

Call:
   aov(formula = fit.dose)

Terms:
                    Dose_level Residuals
Sum of Squares   13098.798   159.579
Deg. of Freedom          1        13

Residual standard error: 3.503618
Estimated effects may be unbalanced
```

Regression Sum of Squares

RSS=Residual Sum Squares

Graphical output

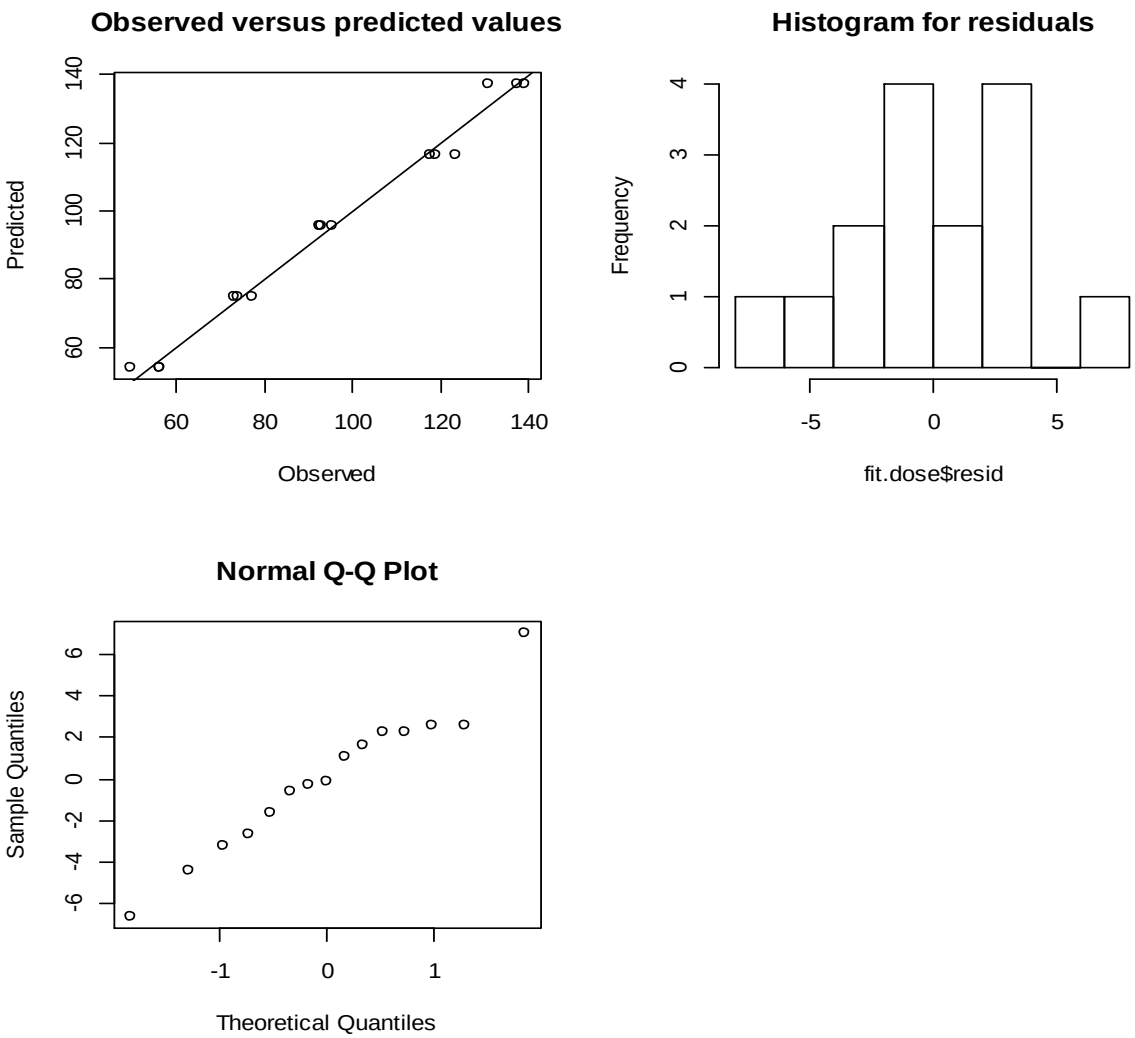
This statemnt produces the Histogram of residuals (to check normality)

```
> par(mfrow=c(2,2))
> plot(fit.dose$fit,xlab="Observed",
   ylab="Predicted", main = "Observed versus,
   predicted values")
> abline(0,1)
> hist(fit.dose$resid,col=0,main="Histogram for
+ residuals")
> qqnorm(fit.dose$resid)
```

This statemnt produces the qqnormal plot (to check normality)

This statemnt produces the plot of the observed versus the predicted values (to check if the variance is constant)

Graphical output

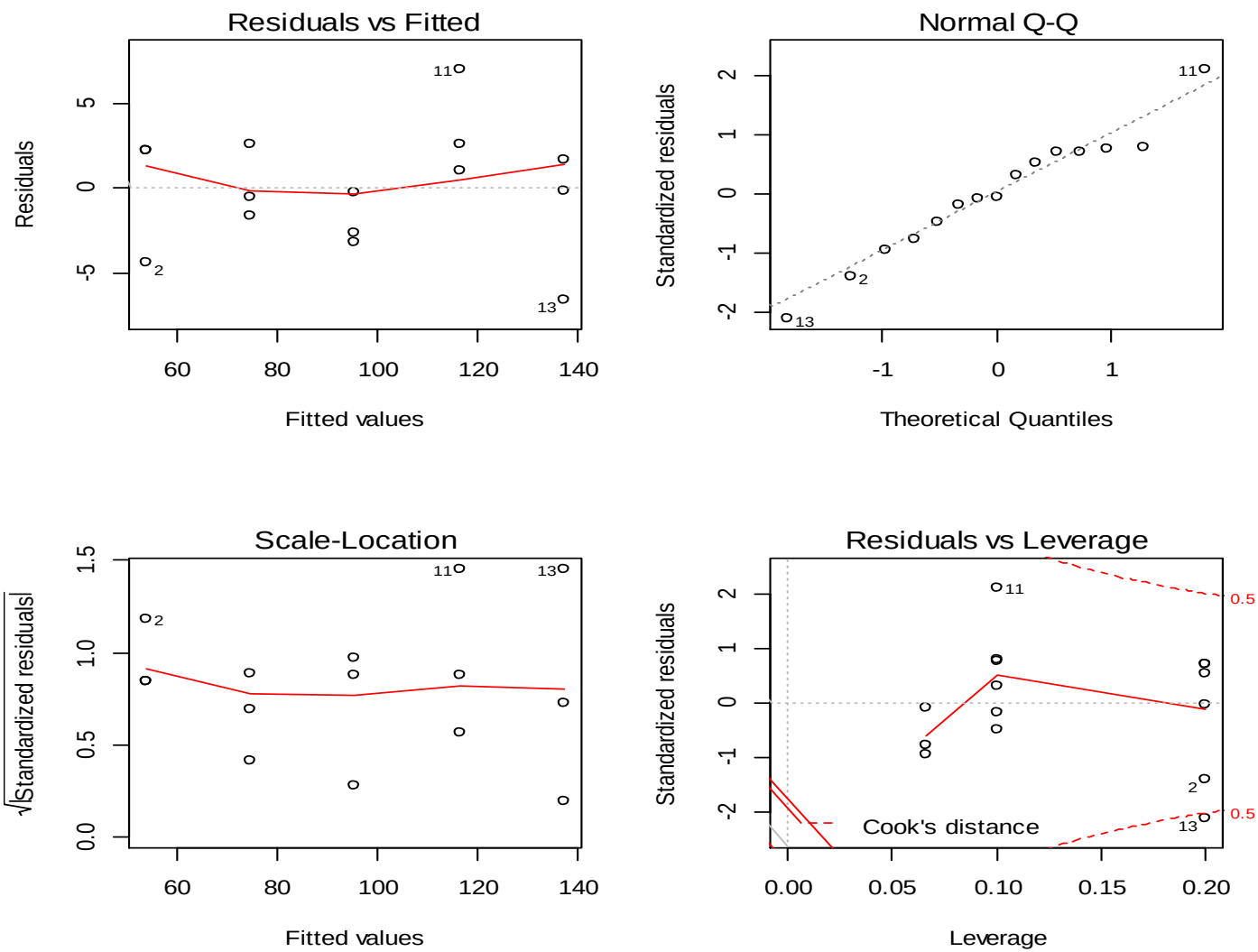


Diagnostic plots

```
> par(mfrow = c(2,2))
> plot(fit.dose)
```

This statemnt produces the plot of the residuals versus the predicted values (to check if the variance is constant), qqnormal plot (to check normality), scale-location plot (to check if the variance is constant), the plot of residuals versus leverage (to check if there is an influencial observation)

Diagnostic plots



Residual analyses

```
## Normality test ##
> shapiro.test(residuals(fit.dose))
```

```
 Shapiro-Wilk normality test

data:  residuals(fit.dose)
W = 0.9723, p-value = 0.8907
```

```
##Constant variance test ##
> library(lmtest)
> bptest(fit.dose)
```

```
 studentized Breusch-Pagan test

data:  fit.dose
BP = 1.0129, df = 1, p-value = 0.3142
```

```
#Testing the Independence Assumption #
library(lmtest)
dwtest(fit.dose, alternative =
+ "two.sided")
```

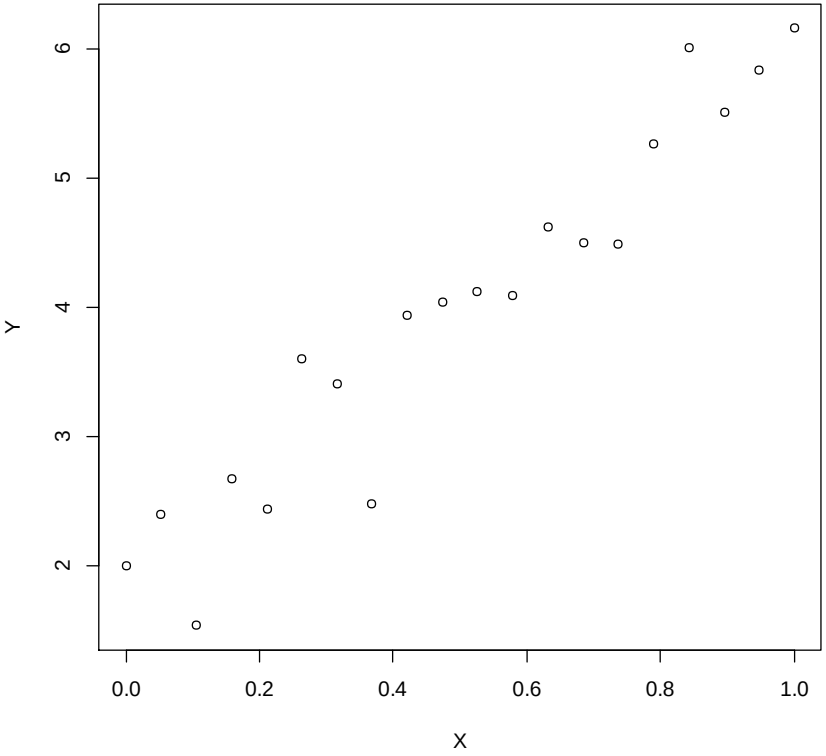
```
Durbin-Watson test

data:  fit.dose
DW = 2.0775, p-value = 0.8863
alternative hypothesis: true autocorrelation is
not 0
```

Four examples

- Example 1: all model asgsumptions hold.
- Example 2: the variance is not constant.
- Example 3: structure in the residuals.
- Example 4: the distribution of the residuals is not normal.

Example 1: The data

- The sample size is equal to 20.
- The observation unit (x_i, y_i), $i=1,\ldots,20$.
- The relatioship between X and Y seems to be linear.



Example 1 External data: example1.txt

Formulation of the model

We consider a linear regression model of the form

$$Y_i = \alpha + \beta \times X_i + \varepsilon_i$$

It is further assumed that the random error is normal distributed with mean 0 and constant variance σ^2.

$$\varepsilon_i \sim N(0, \sigma^2)$$

ANOVA table and parameter estimates

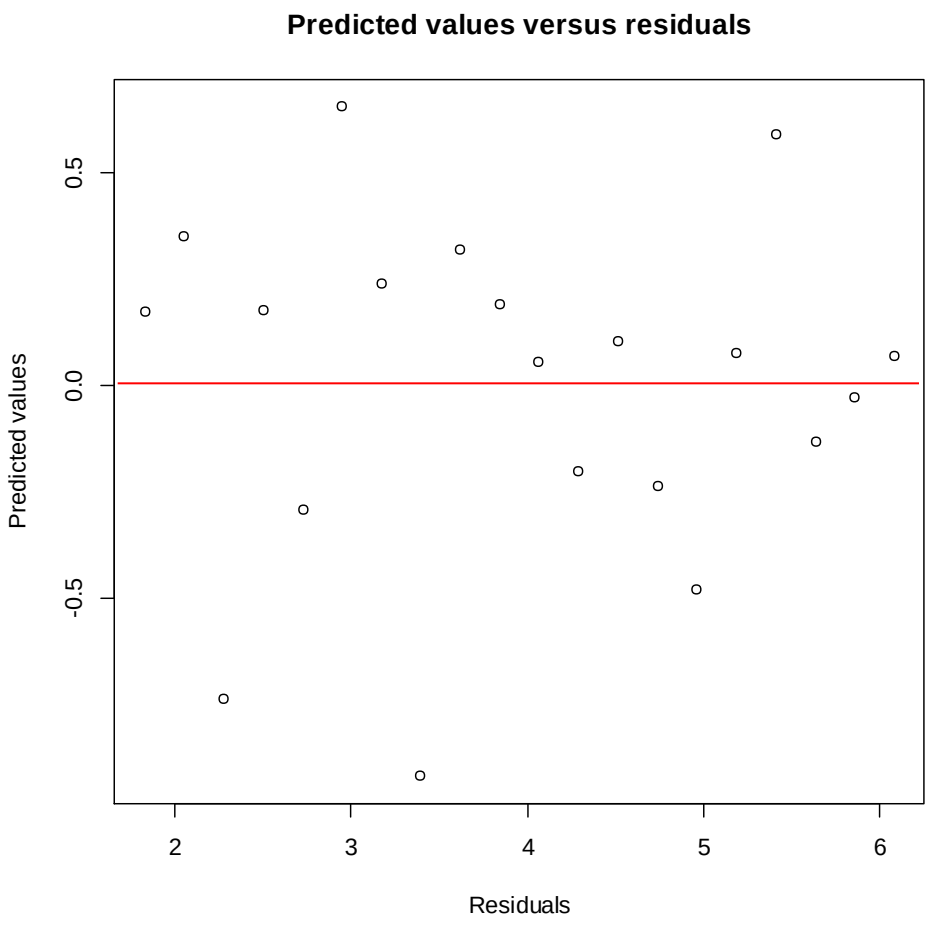
```
Call:
   aov(formula = fit.example1)

Terms:
                        x Residuals
Sum of Squares  33.38747   2.99696
Deg. of Freedom         1        18

Residual standard error: 0.4080414
Estimated effects may be unbalanced
```

ANOVA table and parameter estimates

```
Call:
lm(formula = y ~ x, data = example1)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.91702 -0.21027  0.07406  0.20531  0.65608 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.8260     0.1759   10.38 4.99e-09 ***
x             4.2582     0.3007   14.16 3.36e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.408 on 18 degrees of freedom
Multiple R-squared: 0.9176,     Adjusted R-squared: 0.9131 
F-statistic: 200.5 on 1 and 18 DF,  p-value: 3.364e-11
```

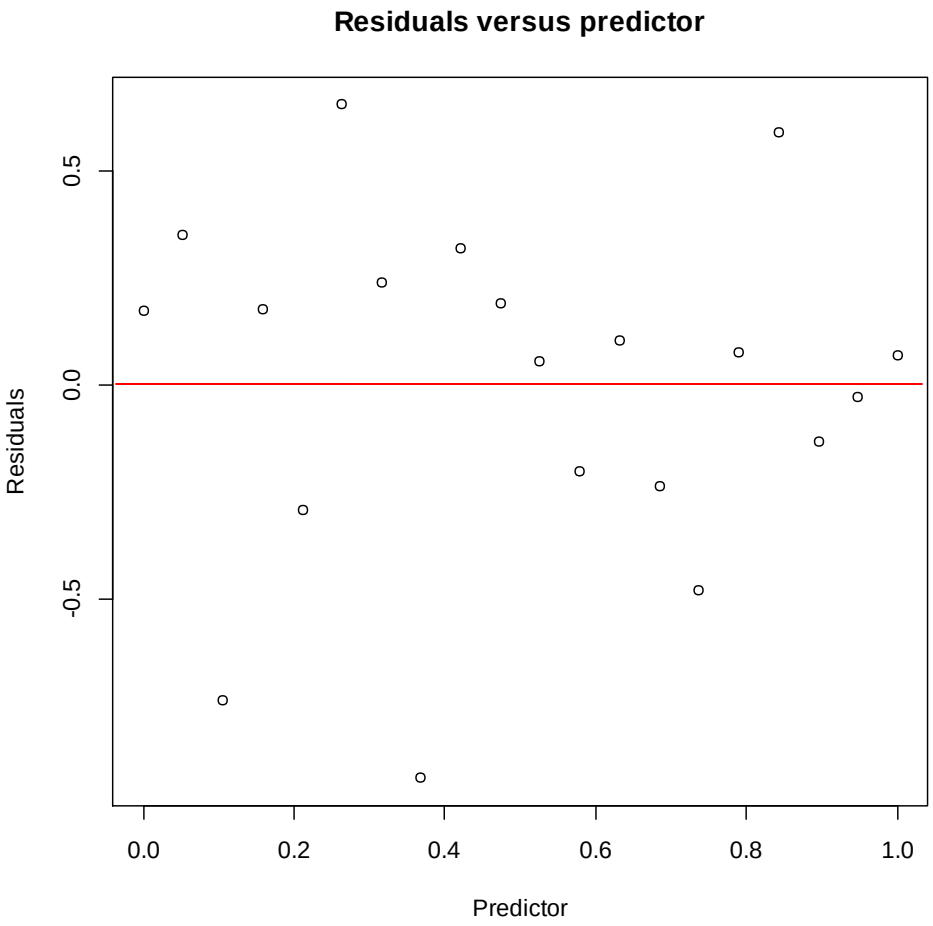
Constant variance: residuals versus predicted values

- We focus in this plot on the variabilty, if it constant we do not expect to patterns in this plot.



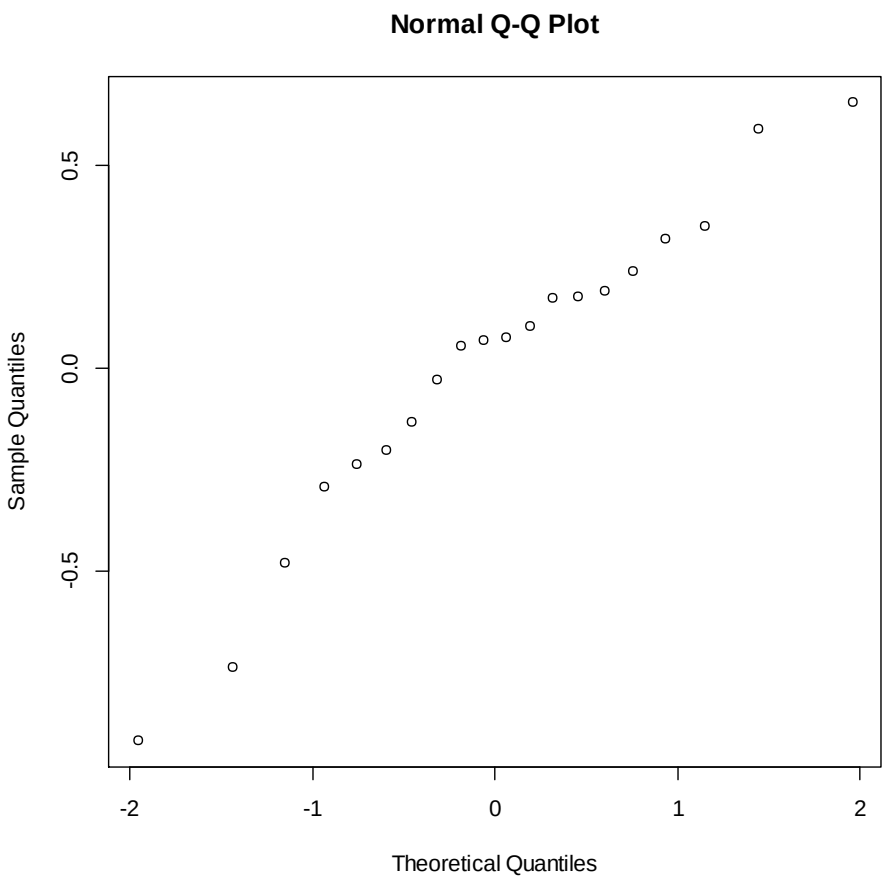
Linearity: residuals versus the predictor

- If the linear model is a "good model" (this means that the assumption that the mean of Y is linear with respect to X) we do not expect to patterns in this plot.

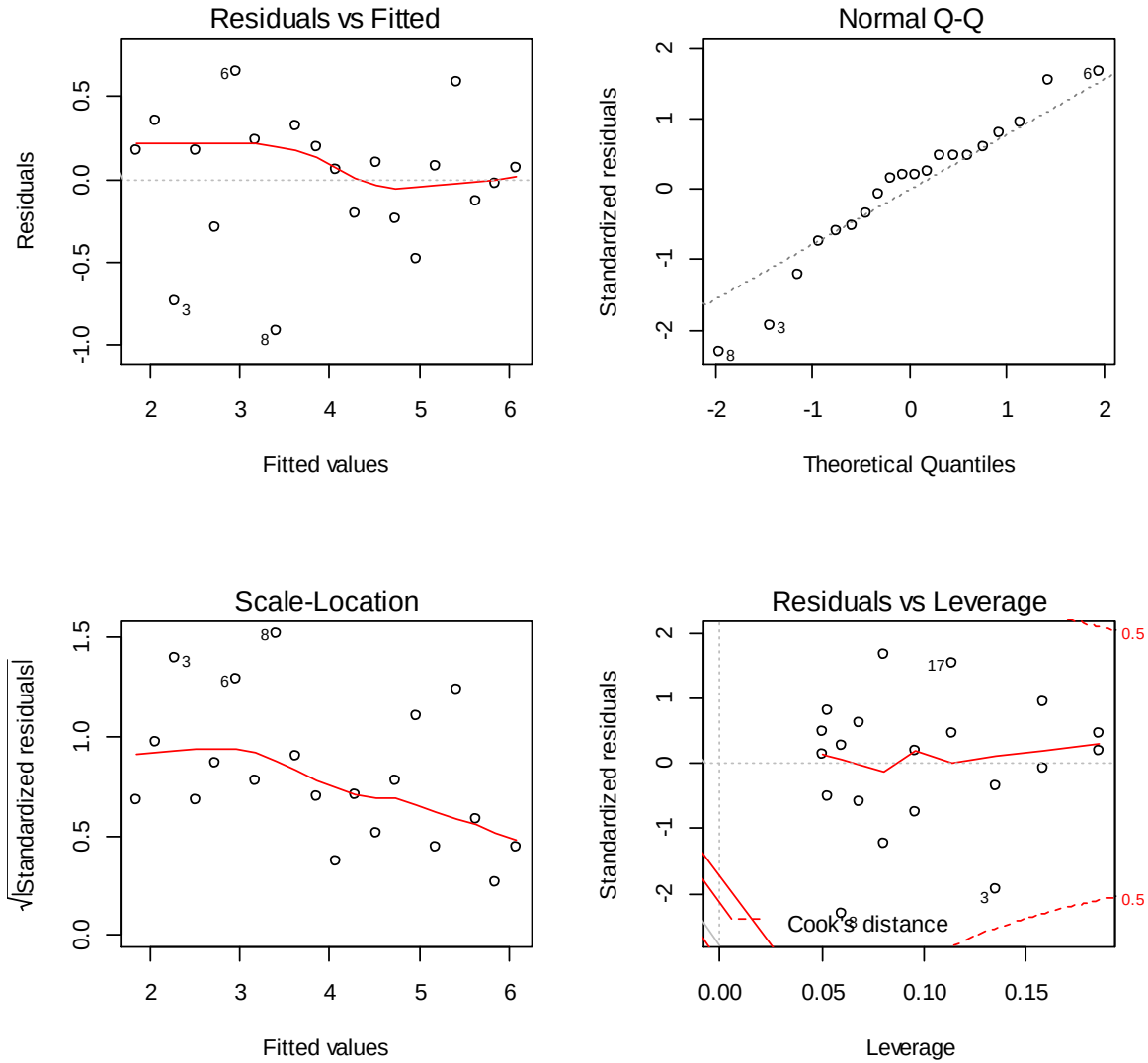


Normality: qqpolt for the residuals

- If the random error is normal distributed the points in the qqnormal plot should follow a stright line pattern.

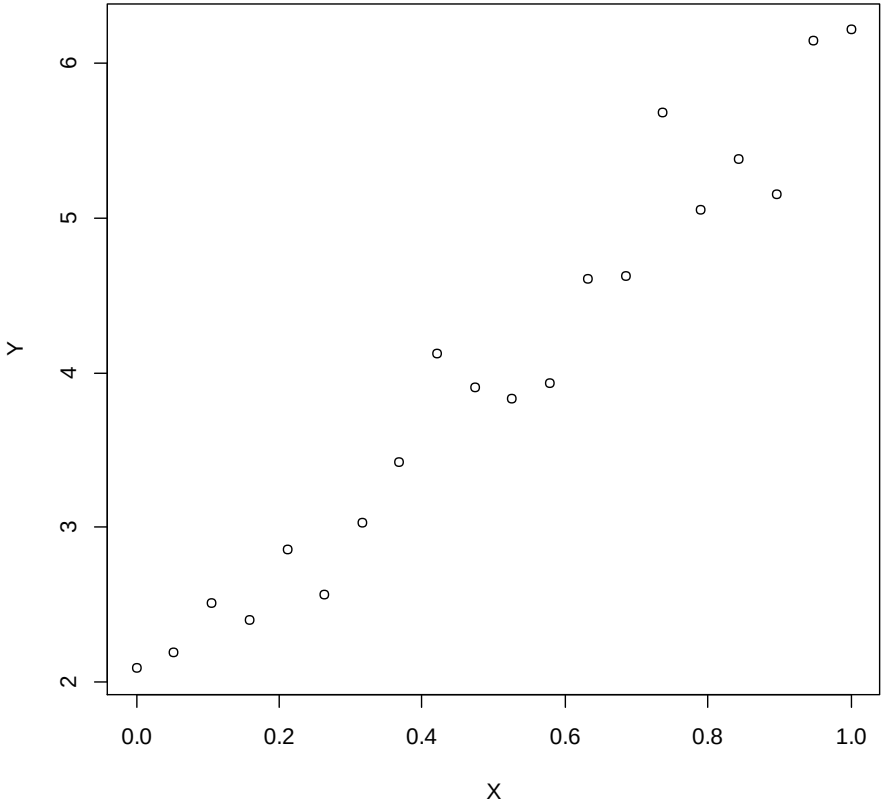


Diagnostic plots



Example 2: The data

- Sample size is 20.
- The relatioship seems to be linear. So the regression model should fit the data.



Example 2 External data: example2.txt

ANOVA table and Parameter estimates

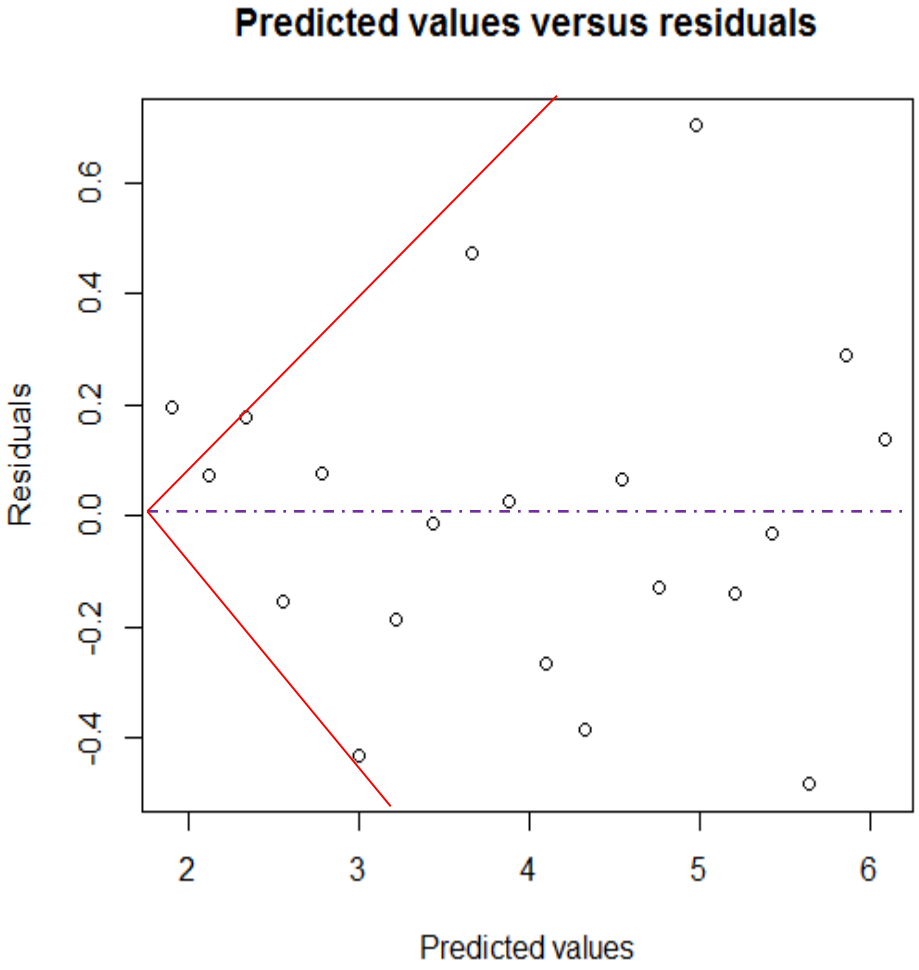
```
> aov(fit.example2)

Call:
   aov(formula = fit.example2)

Terms:
                        x Residuals
Sum of Squares  32.28117   1.64056
Deg. of Freedom         1        18

Residual standard error: 0.3018981
Estimated effects may be unbalanced
```

ANOVA table and Parameter estimates

```
> summary(fit.example2)

Call:
lm(formula = y ~ x, data = example2)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.48422 -0.16228  0.00692  0.14724  0.70333 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.8938     0.1301   14.55 2.13e-11 ***
x             4.1870     0.2225   18.82 2.75e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3019 on 18 degrees of freedom
Multiple R-squared: 0.9516,     Adjusted R-squared: 0.9489 
F-statistic: 354.2 on 1 and 18 DF,  p-value: 2.745e-13
```

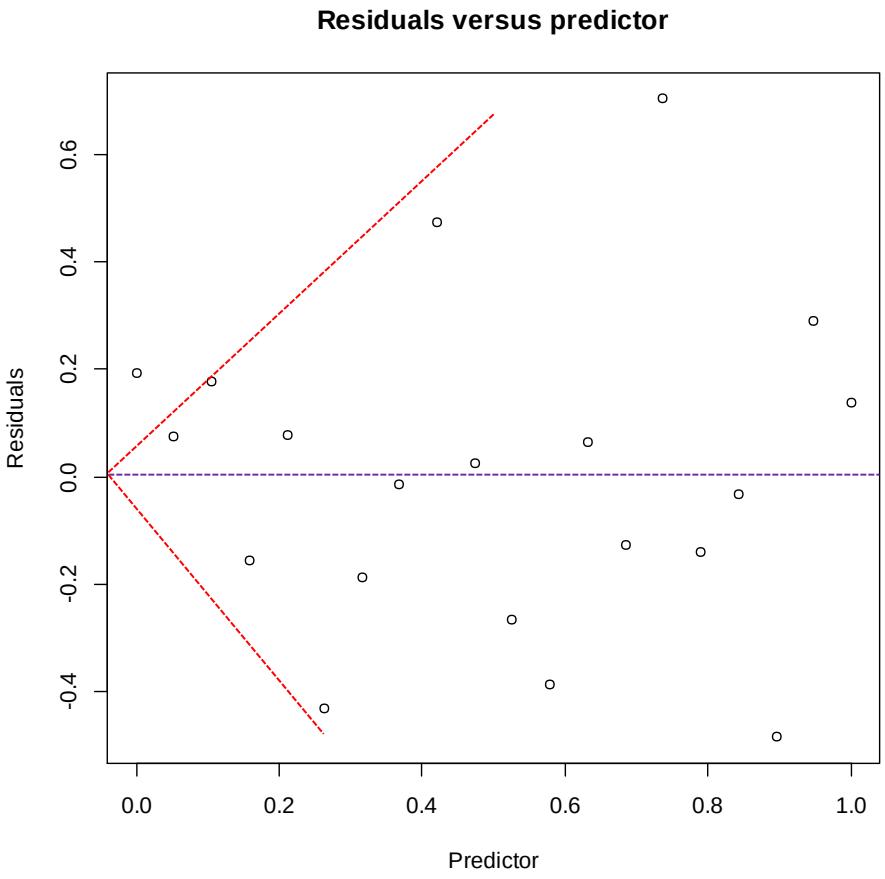
Constant variability

- A “megaphon” shape.
- The variability is not constant.
- The variability increase as the predicted values increase.



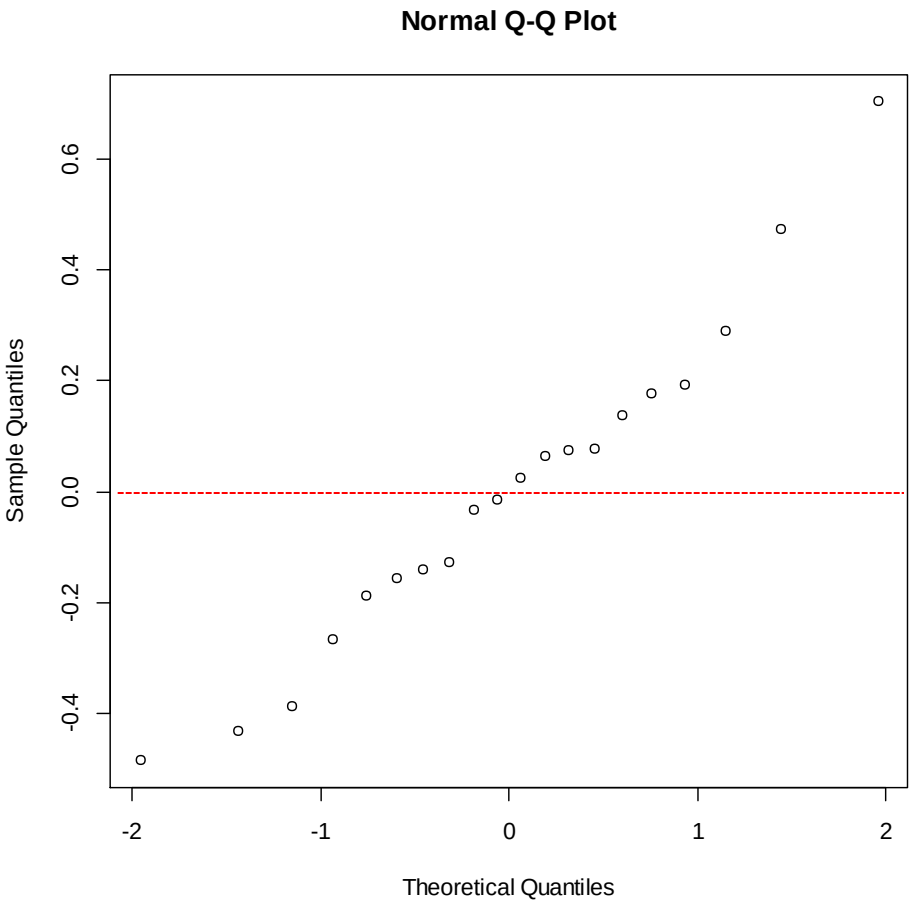
Linearity and constant variability: residuals versus the predictor

- Residuals distributed around zero. This means that the linear regression model captures the main pattern in the data BUT it is clear that the variability is not constant.

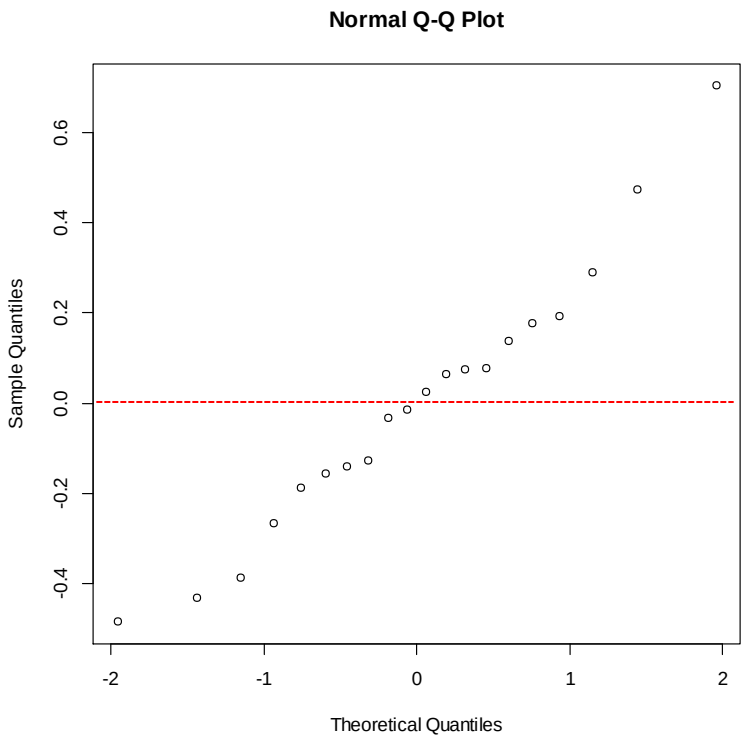


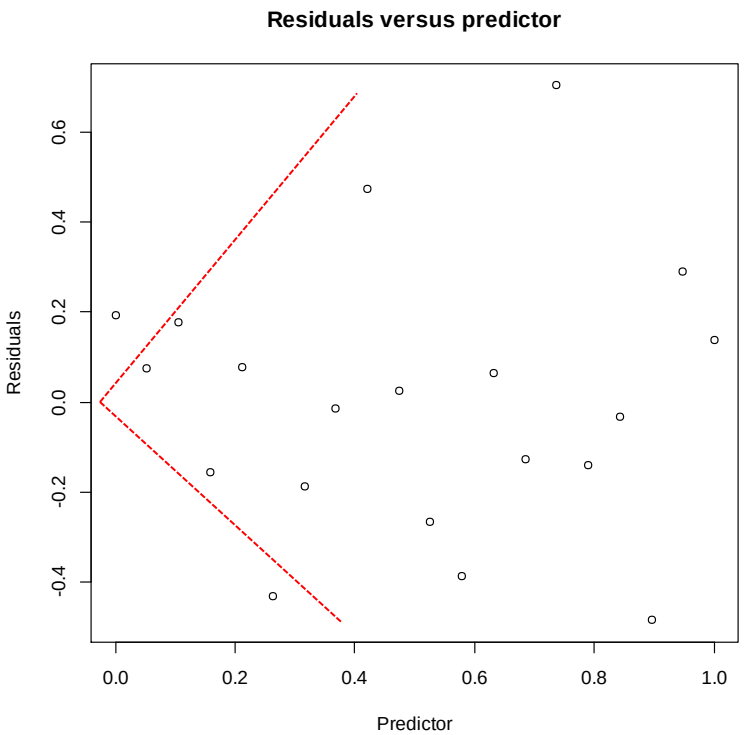
qqnormal plot

- No pattern is detected so we conclude that the random error is normal distributed.



Do not use only one residuals plot for model diagnostic

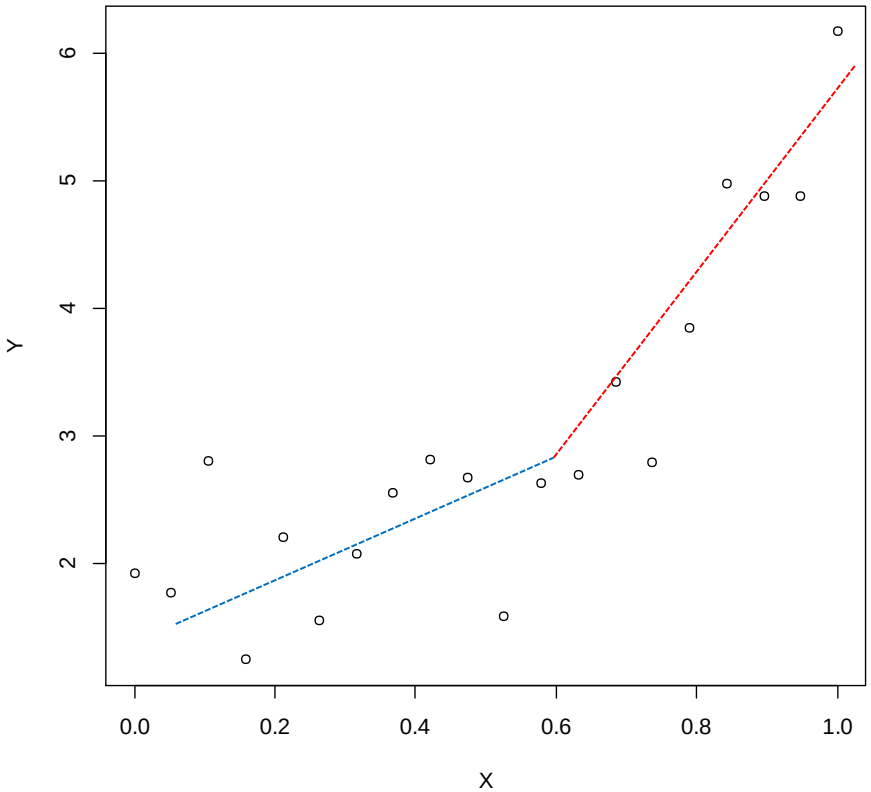




Another example in which the qqnormal plot indicate that the random error is normal distributed and the the plot with the residuals versus the predictor indicates on non constant variance.

Example 3: The data

- Sample size is 20.
- The relatioship seems to be linear BUT NOT A STRIGHT LINE.
- This means that a simple linear regression model will not be able to capture all structure of the data.



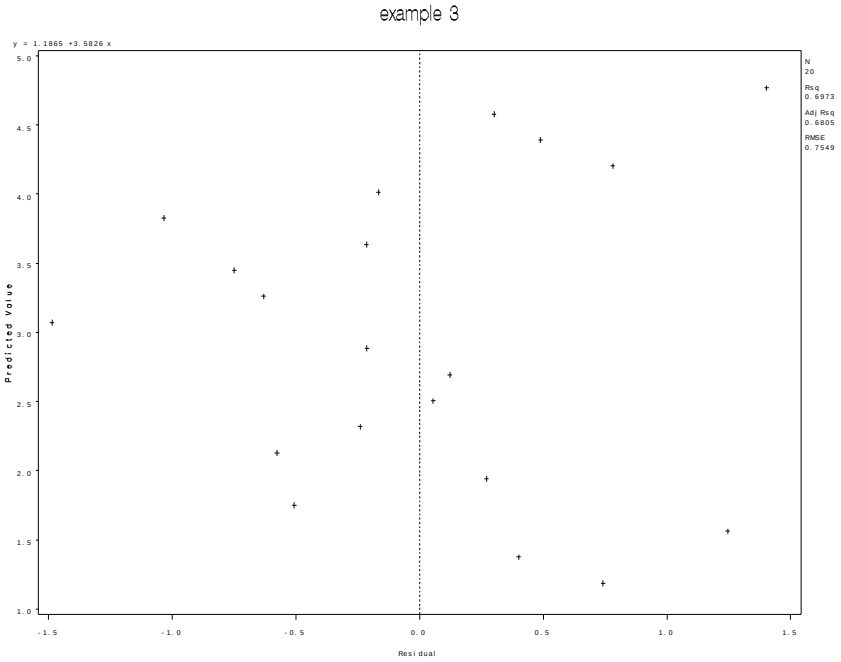
Example 3 External data: example3.txt

ANOVA table and Parameter estimates

```
> aov(fit.example3)

Call:
   aov(formula = fit.example3)

Terms:
                        x Residuals
Sum of Squares  23.63339  10.25865
Deg. of Freedom         1        18

Residual standard error: 0.7549339
Estimated effects may be unbalanced
```

Constant variability

- The residuals plot do not reavel any pattern which indicates that the variance is not constant.

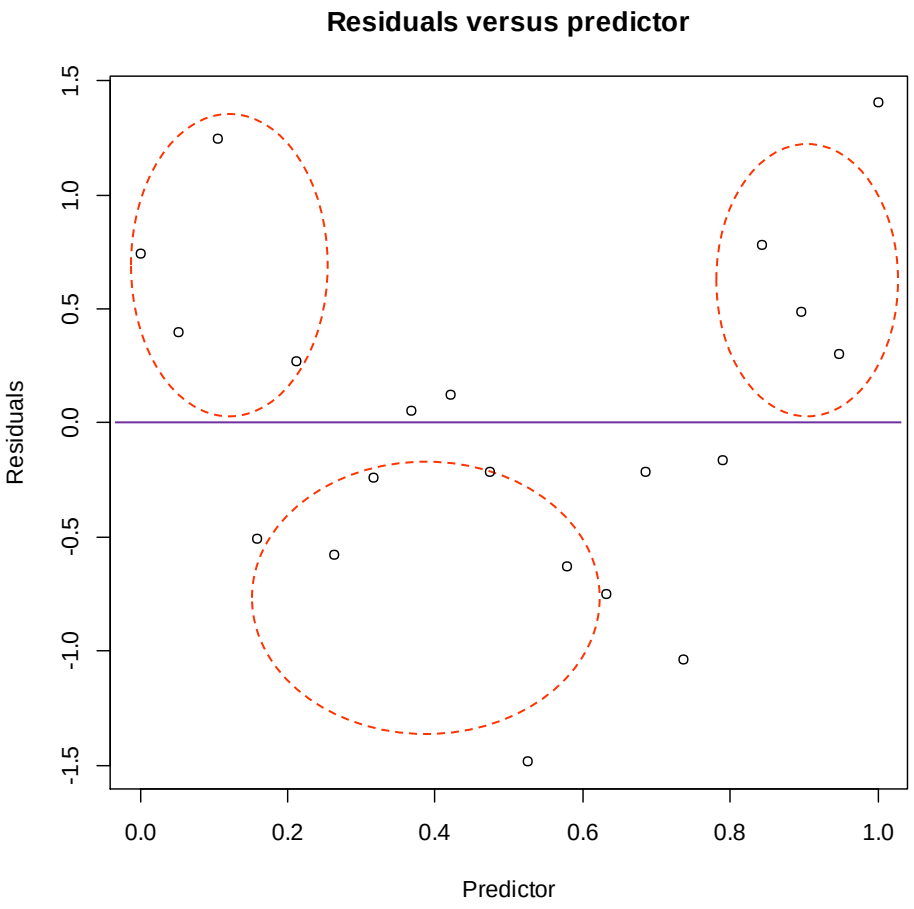


ANOVA table and Parameter estimates

```
> summary(fit.example3)

Call:
lm(formula = y ~ x, data = example3)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.48495 -0.52386 -0.05503  0.42272  1.40292 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.1865     0.3254   3.647  0.00185 ** 
x             3.5826     0.5563   6.440 4.64e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7549 on 18 degrees of freedom
Multiple R-squared: 0.6973,     Adjusted R-squared: 0.6805 
F-statistic: 41.47 on 1 and 18 DF,  p-value: 4.64e-06
```

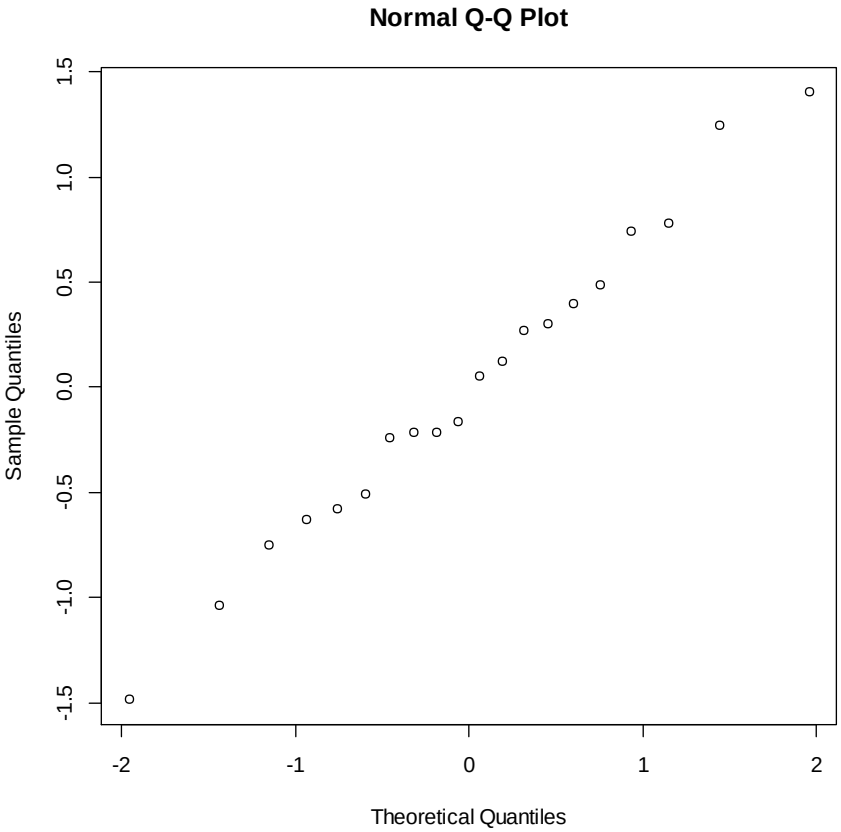
Linearity

- Pattern in the residual plot.
- We observed groups with positive and negative residuals.
- This means that the model does not capture all structure in the data.



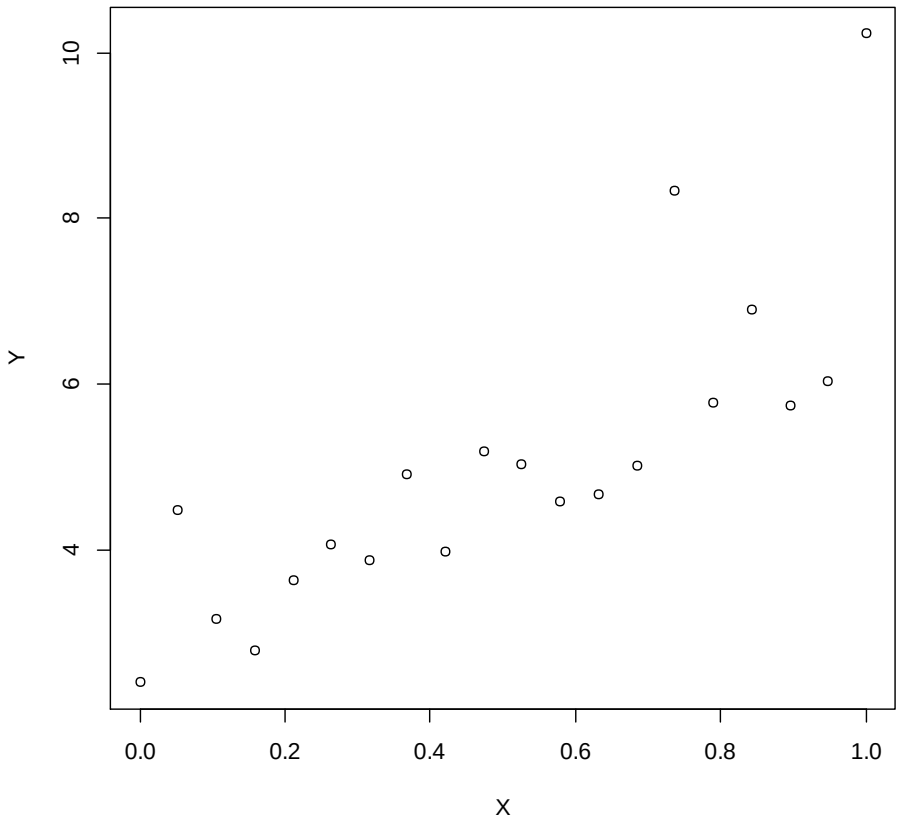
Normality

- A pattern of a sgtight line in the qqnormal plot.
- This indicate that the residuals follow a normal distribution.



Example 4: The data

- This is an example in which the three residuals plots reveal the same problem of the model which is not related to linearity and constant variability.



Example 4 External data: example4.txt

ANOVA table and Parameter estimates

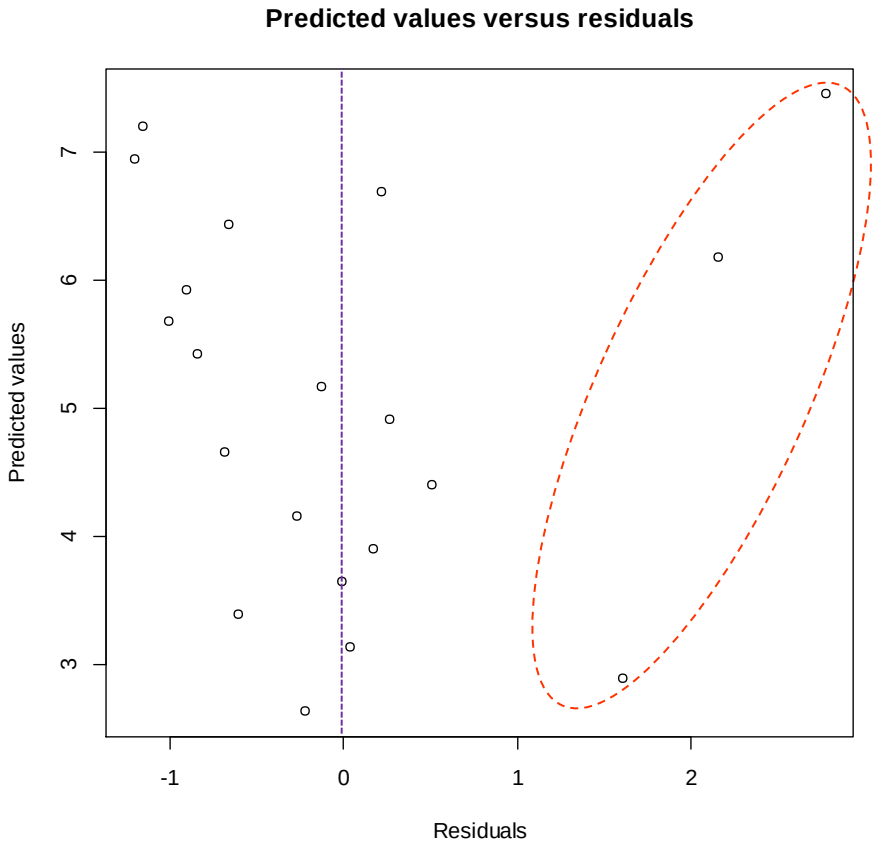
```
> aov(fit.example4)

Call:
   aov(formula = fit.example4)

Terms:
                        x Residuals
Sum of Squares  42.91607  22.07751
Deg. of Freedom         1        18

Residual standard error: 1.107487
Estimated effects may be unbalanced
```

ANOVA table and Parameter estimates

```
> summary(fit.example4)

Call:
lm(formula = y ~ x, data = example4)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.2074 -0.7238 -0.1791  0.2265  2.7738 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.6287     0.4773   5.507 3.14e-05 ***
x             4.8277     0.8161   5.915 1.34e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.107 on 18 degrees of freedom
Multiple R-squared: 0.6603,     Adjusted R-squared: 0.6414 
F-statistic: 34.99 on 1 and 18 DF,  p-value: 1.341e-05
```

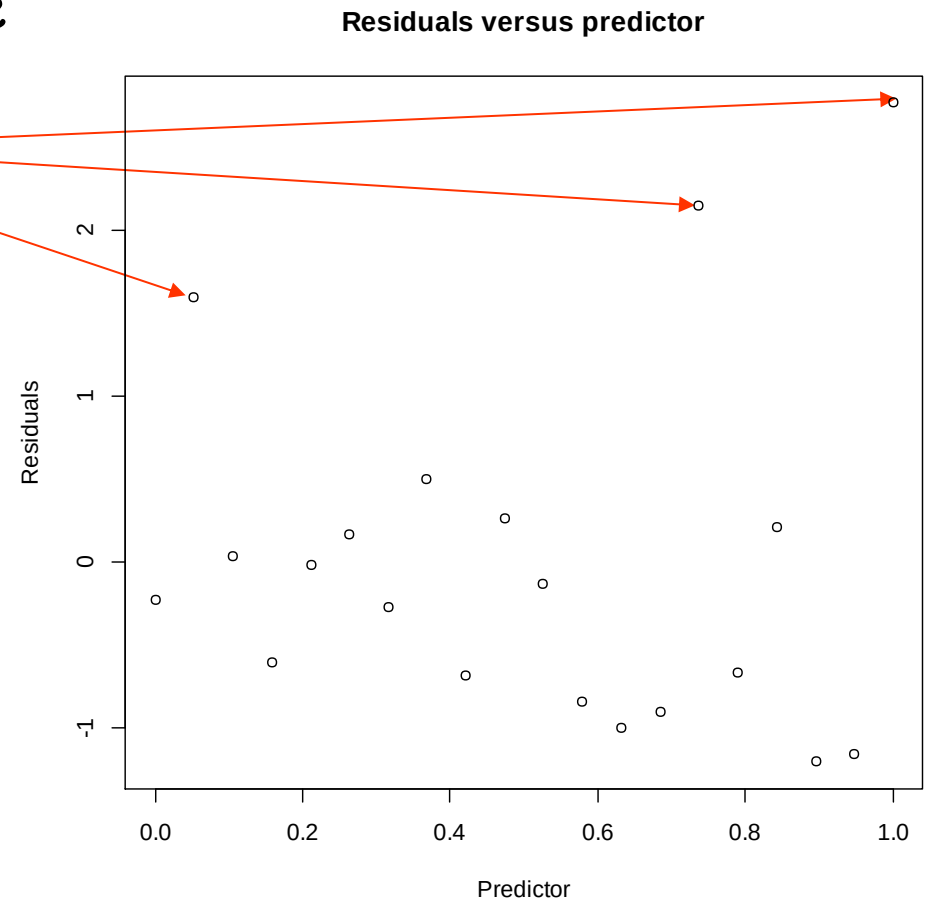
Residuals versus predicted values

- Three positive outliers.



Residuals versus the predictor

- There are more negative residuals than positive residuals and three positive outliers.



Normality

- The pattern in the qqnormal plot indicates on departure from normality.
- Mind the three outliers.