MA 542

REGESSION ANALYSIS

FALL -2018

INSTRUCTOR: Buddika Peiris, PhD (e-mail : tbpeiris@wpi.edu)

LECTURE: M 5.30 -8.20 pm, SH 106

OFFICE: SH 100 (phone: 508 831 5940)

OFFICE HOURS: F 3.00-5.00pm, M 1.00-3.00pm (or by appointment)

TA: Shiao Liu (e-mail : sliu5@wpi.edu)

TEXT BOOK:

This course uses Mathematics extensively, and the students are required to give Mathematical proofs, clear algebraic arguments, and to combine Statistical concepts. The text book,

Applied Linear Regression Models (Fourth edition), by Kutner, Nachtsheim, and Neter

will be used and the chapters 1-11 and 14 will be covered. Some of the material will be omitted and, some new materials will be presented. To benefit from this course, you are required to have a reasonable understanding of Probability and Statistics at the level of MA 511 and matrix algebra.

Other texts that would be useful for the course are:

- Linear Algebra and Its Applications, by David Lay. This has been used as the textbook for MA 2071 (one of the requirements for the course).
- Applied Statistics for Engineers and Scientists, by Joseph Petruccelli, Balgobin Nandram, and Minghui Chen. This has been the textbook for MA2611 and MA2612 (the other requirement for the course).
- Learning R: A Step-by-step Function Guide to Data Analysis By Richard Cotton O'Reilly Media, September 2013.

COURSE WEBSITE: https://canvas.wpi.edu

The website is the main platform through which this course will be managed. It contains the syllabus (this document), and lecture notes, announcements, and other course materials. You are responsible for knowing the information in the materials that appear there.

COMPUTING BACKGROUND FOR THE COURSE:

You will need to be able to get your hands dirty playing with, processing, and plotting data using your favorite computer language. The textbook does not assume any particular computer language, and you are free to do the homework assignments using any computer language you like. However, I will only be able to provide assistance for R. This is not intended to be a programming course (i.e., your code will not be graded, or even collected), but actually working with data will be extremely important (i.e., the results of the code will be graded). A handout with R codes will be provided for each chapter.

R (Statistical software)

• R itself can be found at: http://cran.r-project.org

• I also highly recommend the RStudio front end. It makes developing R code much easier. It can be found at: http://www.rstudio.com

• Note, RStudio requires that you have R itself already installed (so you have to access both of the web pages above).

• Good place to start: A Step-by-Step Function Guide to Data Analysis By Richard Cotton O'Reilly Media, September 2013 available for free from the library.

COURSE OUTLINE:

About a week will be devoted to each of the sections of this course.

- Linear Regression with One Predictor Variable
- Inference in Regression and Correlation
- Diagnostics and Remedial Measures
- Simultaneous Inferences and Other Topics in Regression Analysis
- Matrix Approach to Simple Linear Regression Analysis
- Multiple Regression
- Regression Models for Quantitative and Qualitative Predictors
- Model Selection and Validation
- Diagnostics Measures
- Remedial Measures
- Logistic Regression, Poisson Regression, and Generalized Linear Models

HOMEWORK:

There will be a homework assignment every week for your benefit and practice - they can also serve as a test of your level of materials being covered in class. Homework will help you to

- Gain a solid understanding of the course material.
- Be creative and think beyond the course material.
- Do better in the exams.

You can informally discuss some problems with your classmates but the final work should be based on your own effort. Please feel free to see me if you have any question.

Advice on homework:

- Make sure your home-works are clearly written and stand alone.
- Make sure that everything appears in your homework write-up!
- Make sure it is clear where each part of each question is answered.
- Make sure to **submit each homework before the class starts** on it's due data.

QUIZZES:

Eight ten minutes' open book quizzes will be held. **No electronics are allowed** during the quizzes except for a simple calculator. Calculator apps on a smartphone, tablet, kindle, etc are not allowed. You should bring a calculator to each quiz.

EXAMS:

There will be 2 exams based on the material covered until the latest lecture before each. One double-sided **hand written** sheet is allowed for each exam. **No electronics are allowed** during the exams except for a simple calculator. Calculator apps on a smartphone, tablet, kindle, etc are not allowed. No makeup exam will be given unless a student notify me with a legitimate excuse by writing prior to the exam. **Makeup exam may be harder than the original exam**.

Make sure you do not select classes with conflicting exam dates.

GRADIN CRITERIA:

- 10 HWs (20%)
- 8 Quizzes (20%)
- Test-1 (25%)
- Test-2 (35%)

GRADIN SCALE:

- A: Overall ≥ 90 **and** grade for each component (HWs, Quizzes, Test-1 and Test-2) ≥ 80.
- B: Overall ≥ 80 **and** grade for each component (HWs, Quizzes, Test-1 and Test-2) ≥ 70.
- C: Overall ≥ 70 **and** grade for each component (HWs, Quizzes, Test-1 and Test-2) ≥ 60.
- NR: Overall < 70 **or** grade for at least one component (HWs, Quizzes, Test-1 and Test-2) < 60.

STUDENTS WITH DISABILITIES:

You should contact the Disabilities Services Office so an appropriate accommodation can be implemented. Please contact dso@wpi.edu or phone x-5235. See me as early as possible in the term so I can address your specific needs.

ACADEMIC HONESTY:

The academic honesty policy can be accessed at:
http://www.wpi.edu/Pubs/Policies/Honesty/Students/

TENTATIVE DATES:

Class	Date	Assignments
1	Jan-10	
2	Jan-22	Homework-1
3	Jan-29	Homework-2, Quiz-1
4	Feb-05	Homework-3, Quiz-2
5	Feb-12	Homework-4, Quiz-3
6	Feb-19	Homework-5, Quiz-4
7	Feb-26	Exam-1
8	Mar-12	
9	Mar-19	Homework-6
10	Mar-26	Homework-7, Quiz-5
11	Apr-02	Homework-8, Quiz-6
12	Apr-09	Homework-9, Quiz-7
13	Apr-23	Homework-10, Quiz-8
14	Apr-30	Final Exam

HW - 1 : M (1-22)

MA 542: Regression Analysis

Class-1

Defn Regression

Regression analysis is a statistical methodology that utilizes the relationship between two or more quantitative variables so that the outcome variable (response variable) can be predicted from the other (or others).

Eg:- The age of unborn baby is estimated using SFH (fundal height).

can not observe

can observe

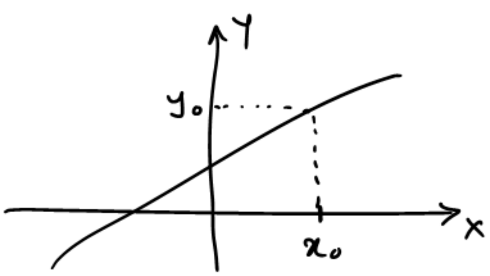
Relationship between variables

There are two types.

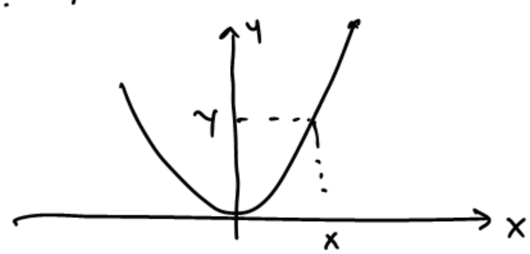
1) Functional Relations (perfect)

A functional relation (Mathematical relation) is of the form $y = f(x)$ and it is a perfect relation.

Eg:- $Y = 3X + 2$



2. $Y = 5X^2$



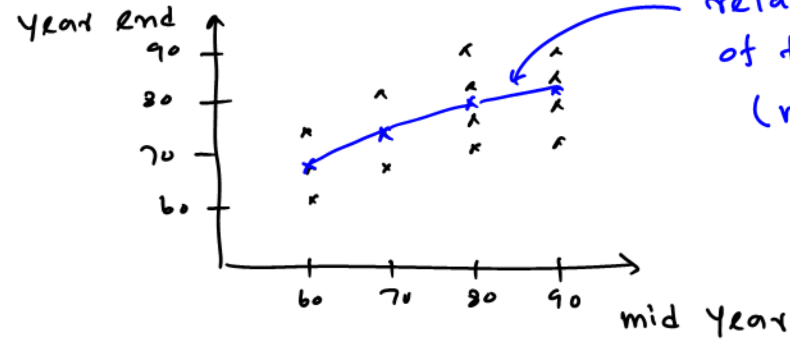
2) Statistica Relations

Statistical relations are not perfect.

Eg:- Performance evaluations for 10 employs were obtained at mid year and the year end.

$(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$

Scatter Plot:



relation between X and the mean of the response (Y) at each X-level (mathematical relation)

Here relation between X and Y is not perfect.

It is a Statistical relation.

A regression model decribes the the Statistical relations.

Problem:
How to construct a regression model?

There 3 steps

1) Identify the response and Selection of predictors. (How many, what ?).

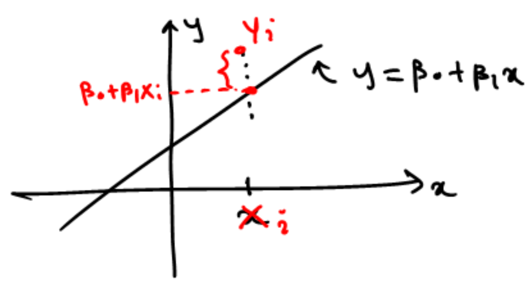
2) Function form of the regression relation. (linear or non-linear)

3) Scope of the model (ie Range of each predictor variable).

Chapter-1: Linear Regression with One predictor Variable

* Simple Linear Regression Model.

$$\underbrace{Y_i}_{\text{random variable}} = \underbrace{\beta_0 + \beta_1 X_i}_{\text{fixed}} + \underbrace{\varepsilon_i}_{\text{random error}},$$



where

β_0, β_1 — Parameters

X_i — Known constant, the value of the predictor variable for the i^{th} trial.

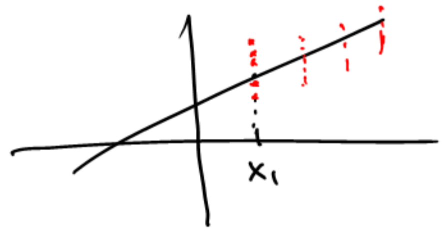
ε_i — random error for the i^{th} trial.

Assumptions:

1) $E(\varepsilon_i) = 0$ for all $i = 1, 2, \ldots n$.

2) $var(\varepsilon_i) = \sigma^2$ for all $i = 1, 2, \ldots n$.

3) error terms are independent.



Note:

* The model is called simple because the model has only one predictor variable.
* It is called linear because it is linear in parameters and in X.

$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{constant}} + \underbrace{\varepsilon_i}_{\text{random}}$$

$$E[aX + b] = aE(X) + b$$

rv

$$* E[Y_i] = E[\beta_0 + \beta_1 X_i + \varepsilon_i]$$
$$= \beta_0 + \beta_1 X_i + \underbrace{E(\varepsilon_i)}_{=0}$$
$$= \beta_0 + \beta_1 X_i$$

$$var(aX + b) = a^2 var(X)$$

$$var(Y_i) = var(\beta_0 + \beta_1 X_i + \varepsilon_i)$$
$$= var(\varepsilon_i)$$
$$= \sigma^2$$

Further since ε_i and ε_j $(i \neq j)$ are independent, Y_i and Y_j are also independent.

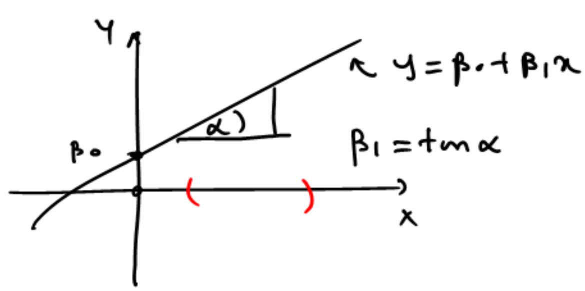
Interpretation of Parameters

The mean response,

$$E[Y_i] = \beta_0 + \beta_1 X_i$$

Intercept parameter

Slope parameter.



β_0 and β_1 are also called as regression coefficients.

* β_1 - Slope parameter : change in mean response with a unit increase in X.

* β_0 - Intercept parameter : value of the mean response when $X=0$, if 0 in the scope of X. Otherwise there is no particular meaning for β_0.

An alternative model

Data: $X_1, X_2, \dots X_n$

centralized data:

$X_1-\bar{X}, X_2-\bar{X}, \dots, X_n-\bar{X}$.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$= \beta_0 + \beta_1(X_i - \bar{X}) + \beta_1\bar{X} + \varepsilon_i$$

$$= \underbrace{\beta_0 + \beta_1\bar{X}} + \beta_1(\underbrace{X_i - \bar{X}}_{\text{centralized data}}) + \varepsilon_i$$

$$\boxed{Y_i = \beta_0^{*} + \beta_1(X_i - \bar{X}) + \varepsilon_i}$$, where $\beta_0^{*} = \beta_0 + \beta_1\bar{X}$.

Random Sample

$\binom{30}{10} = {}^{30}C_{10} = \frac{30!}{10!\,20!}$

Defn

A sequence of independent and identically distributed random variables $X_1, X_2, \dots, X_n$ is called a random sample.

random sample

Sample #	X_1	X_2	X_3	- - - - -	X_{10}	$\bar{X}$	Var(X)	$\tilde{X}$
1	×	×	×	× . . .	×	$\bar{x}_1$		
2	□	□	□	□ . . .	□	$\bar{x}_2$		
3	o	o	o	o	o	$\bar{x}_3$		
⋮						⋮		
$\binom{30}{10}$	×	×	×	×	×	$\bar{x}_n$		

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$ — Estimator (Random variable)

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$ — Estimate (a value).

Estimation of Regression Function.

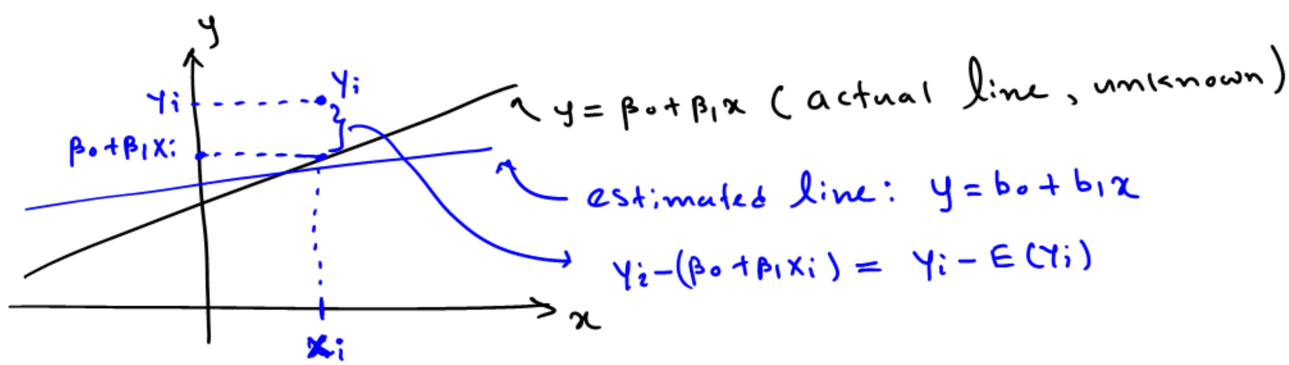
Let $Y_1, Y_2, \ldots Y_n$ is a random sample and we have $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad , \quad Var(\varepsilon_i) = \sigma^2$$

There 3 parameters to estimate.

To estimate β_0 and β_1, we use the method of least squares.

Idea:



$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

Finding $\hat{\beta}_0$ (ie estimator of $\beta_0 = b_0$) and $\hat{\beta}_1$ (ie b_1) which minimze Q is called least square estimation.

Take the partial derivatives

$$\frac{\partial Q}{\partial \beta_0} = 2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)(-1) \overset{\text{set}}{=} 0 \longrightarrow ①$$

$$\frac{\partial Q}{\partial \beta_1} = 2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)(-X_i) \overset{\text{set}}{=} 0 \longrightarrow ②$$

Also check

$$\frac{\partial^2 Q}{\partial \beta_0^2} = 2n > 0$$

$$\frac{\partial^2 Q}{\partial \beta_1^2} = 2 \sum X_i^2 > 0$$

$$① \Rightarrow \quad \sum Y_i = \sum b_0 + \sum b_1 X_i = n b_0 + b_1 \sum X_i$$

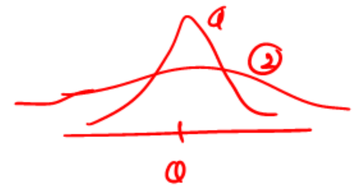
$$② \Rightarrow \quad \sum X_i Y_i = b_0 \sum X_i + b_1 \sum X_i^2$$

These are called normal equations.

By solving,

$$\hat{\beta}_1 = b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = b_0 = \frac{1}{n}\left(\sum Y_i - b_1 \sum X_i\right) = \bar{Y} - b_1 \bar{X}$$



Properties of least square estimators

Gauss Markov theorem

under the condition of regression model, the least square estimators b_0 and b_1 are unbiased and have minimum variance among all unbiased linear estimators.

ie $E(b_0) = \beta_0$ and $E(b_1) = \beta_1$

random variables

constants.

Note:

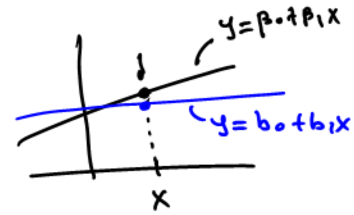
$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \underbrace{\sum k_i Y_i}_{\text{linear combination of } Y_i \text{ (linear estimator)}},$$

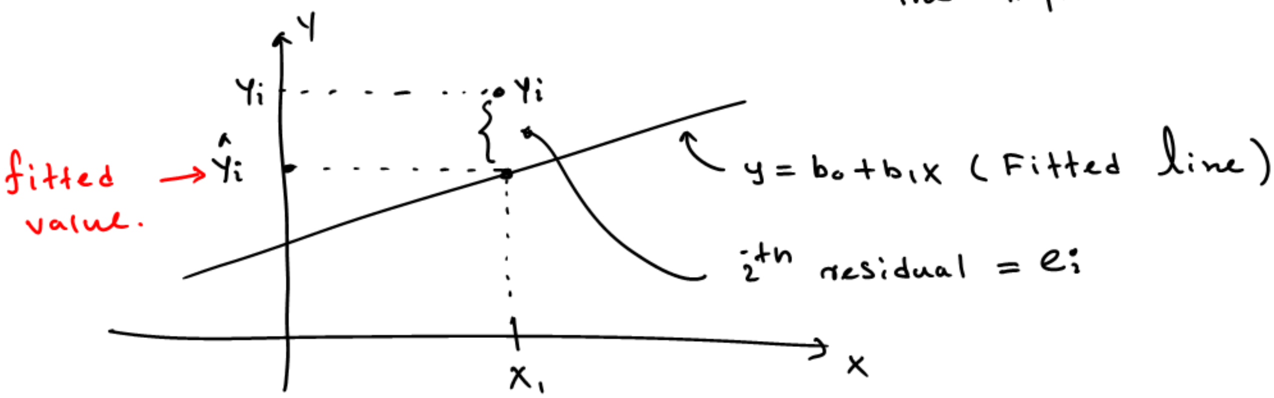
where $k_i = \dfrac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2}$.

Point estimation of mean response

Mean response : $E(Y) = \beta_0 + \beta_1 X$ — parameter.

Point Estimator : $\hat{Y} = b_0 + b_1 X$ – the value of the regression function at X.





* $\hat{Y}_i = b_0 + b_1 X_i$, $i = 1, 2, \dots n$ is the estimated value of the regression function at $X = X_i$.

* The difference between Y_i and $\hat{Y}_i$
  ie $e_i = Y_i - \hat{Y}_i$ is calle the i^{th} residual.

$$e_i = Y_i - \hat{Y}_i, \quad i = 1, 2, \dots n.$$

$$\text{ie } e_i = Y_i - (b_0 + b_1 X_i), \quad i = 1, 2 \dots n.$$

Properties of the fitted regression line

1) $\sum e_i = 0$

Proof:

$$\sum e_i = \sum (Y_i - b_0 - b_1 X_i) = \sum Y_i - nb_0 - b_1 \sum X_i$$
$$= 0 \text{ (1st normal equation)}$$

2) $\sum e_i^2$ is minimum.

$$Q = \sum (Y_i - \beta_0 - \beta_1 X_i)^2$$

Proof:

$\sum e_i^2 = \sum (Y_i - b_0 - b_1 X_i)^2$ is minimum because b_0 and b_1 are least square estimators.

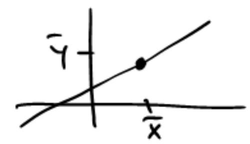
3) $\sum Y_i = \sum \hat{Y}_i$

(ie sum of the observed values = sum of the fitted values)

Proof – HW.

4) $\sum X_i e_i = 0$ (ie sum of weighted residuals is zero)

Proof – HW.



5) $\sum \hat{Y}_i e_i = 0$

Proof - HW.

6) The fitted regression line always goes through the point $(\bar{X}, \bar{Y})$

<u>Proof:</u>

$$\underset{\sim}{b_0} + b_1 \bar{X} = \underbrace{\bar{Y} - b_1 \bar{X}} + b_1 \bar{X} = \bar{Y}$$

So $(\bar{X}, \bar{Y})$ is on the line $Y = b_0 + b_1 X$.

<u>Estimation of error variance</u>

<u>Recall:</u>

Estimation of variance for a single population.

Let $\underline{Y_1, Y_2, \ldots, Y_n}$ be a random sample from a population with mean μ and variance σ^2.

$\frac{Y_1 + Y_2 \cdots + Y_n}{n}$

then $\hat{\mu} = \bar{Y}$ and

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{df} = \frac{\sum (Y_i - \bar{Y})^2}{n-1}.$$

By following the same argument,

$\| \quad \|$

$b_0 + b_1 X_i$

$$\text{point estimator of the error variance} = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{df}$$

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

Here $\sum (y_i - \hat{y}_i)^2$ is called the Sum of squares of errors (SSE) and the degrees of freedom of SSE is $n-2$

Further $\dfrac{\sum (y_i - \hat{y}_i)}{n-2}$ = mean square error.

$$\text{ie} \quad \hat{\sigma}^2 = MSE = \frac{SSE}{n-2} = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} = \frac{\sum e_i^2}{n-2}$$

* point estimator of $\sigma = \sqrt{MSE}$.

Normal Error Regression Model

The normal error regression model

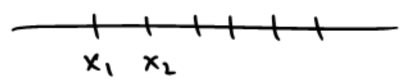
$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad , \; i = 1, 2, \dots n,$$

where ε_i's are independent, identically normaly distributed with mean 0 and variance σ^2.

Recall:

$$E(y_i) = \beta_0 + \beta_1 X_i \quad \text{and} \quad var(y_i) = \sigma^2$$

Further $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, $i = 1, 2, \dots n$.



* Here note that $y_i \overset{\text{independent}}{\sim} N(\beta_0 + \beta_1 X_i, \sigma^2)$

not identical

Recall: Pdf of normal distribution: If $X \sim N(\mu, \sigma^2)$

$$\text{Pdf}: \quad f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \quad : \; -\infty < x < \infty.$$

* Maximum Likelihood Estimators of normal Error regression model

Recall : Likelihood function:

Likelihood function is the joint pdf, when we consider it as a function of parameters.

Eg:- Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then the likelihood function is

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{\Sigma(x_i-\mu)^2}{2\sigma^2}}.$$

R Codes for Chapter-1

Importing data from internet

When downloading data from internet, use "**read.table()".** In the arguments of the function:

- header: if TRUE, tells R to include variables names when importing,
- sep: tells R how the entries in the data set are separated.
  - sep=",": when entries are separated by COMMAS
  - sep="\t": when entries are separated by TAB
  - sep=" ": when entries are separated by SPACE

E.g:- The following command is used to import the data for Plastic Hardness example (exercise 1.22)

```
> data<-read.table("http://www.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/

data/textdatasets/KutnerData/Chapter%20%201%20Data%20Sets/CH01PR22.txt ",

header= FALSE , sep="")

> data
    V1 V2
1  199 16
2  205 16
3  196 16
4  200 16
5  218 24
6  220 24
7  215 24
8  223 24
9  237 32
10 234 32
11 235 32
12 230 32
13 250 40
14 248 40
15 253 40
16 246 40
```

Importing data from the computer:

First, you need to save data in a folder in your computer. Then use **read.table()** as follows.

```
> datart=read.table("R:\\Teaching\\2017\\MA 542 F\\R Codes\\

Plastic Hardnes.csv",header = FALSE)

> datart
    V1 V2
1  199 16
2  205 16
.

.
```

This is only a part of the output.

Fitting the Simple Linear Regression (SLR) Model

The command **"lm"** can be used to fit the SLR model in R. To perform use the command:

lm (response ~ Predictor)

Here the terms response and Predictor in thecommand should be replaced by the names of the response and predictor variables, respectively, used in the analysis.

Ex. Plastic Hardness (Problem 1.22), Y=Hardness in Brinell units, X=Elapsed time in hours.

```
> Hardness=data[,1]
> Time=data[,2]
```

The following command crates a data frame, which is needed for most of the commands.

```
> dataf=data.frame(Hardness, Time)
> dataf
   Hardness Time
1       199   16
2       205   16
3       196   16
```

To fit a simple linear regression model, use the command:

name Y X

```
> SLR=lm(Hardness~Time,data=dataf)
> SLR
```

data frame.

```
Call:
lm(formula = Hardness ~ Time, data = dataf)

Coefficients:
(Intercept)         Time
    168.600        2.034
```

This output indicates that the fitted model is given by $\widehat{Y} = 168.600 + 2.034\,X$.

We can access more details about the fitted model by typing:

```
> summary(SLR)
```

name

```
Call:
lm(formula = Hardness ~ Time, data = dataf)

Residuals:
    Min      1Q  Median      3Q     Max
-5.1500 -2.2188  0.1625  2.6875  5.5750
```

C(1,1)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 168.60000    2.65702   63.45  < 2e-16 ***
Time          2.03438    0.09039   22.51 2.16e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$\sqrt{MSE} = \hat{\sigma}$

```
Residual standard error: 3.234 on 14 degrees of freedom
Multiple R-squared:  0.9731,   Adjusted R-squared:  0.9712
F-statistic: 506.5 on 1 and 14 DF,  p-value: 2.159e-12
```

Extracting Estimators:

```
> b0=summary(SLR)$coefficients[1,1]
> b0
[1] 168.6

> b1=summary(SLR)$coefficients[2,1]
> b1
[1] 2.034375
```

The following command extracts the **least square estimator of the error standard deviation** $\hat{\sigma}$.

```
> sigmahat=summary(SLR)$sigma #Least square estimator.
> sigmahat
[1] 3.234027
```

$= \sqrt{MSE}$

We need to calculate the MLE of the error standard deviation manually.

```
> DoFR=df.residual(SLR) #Extracting error degrees of freedom:
> DoFR
[1] 14
>
> mle_sigmahat=sqrt(summary(SLR)$sigma^2*DoFR/(length(Hardness)))
> mle_sigmahat
[1] 3.025155
```

Fitted Values:

To calculate the fitted values, use the following command.

```
> Fitvals=fitted.values(SLR)
> Fitvals
      1       2       3       4       5       6       7       8       9 
201.150 201.150 201.150 201.150 217.425 217.425 217.425 217.425 233.700 
     10      11      12      13      14      15      16 
233.700 233.700 233.700 249.975 249.975 249.975 249.975 
```

Residuals:

Residuals for the fitted regression model are calculated as follows.

```
> Res=residuals(SLR)
> Res
     1      2      3      4      5      6      7      8      9     10 
-2.150  3.850 -5.150 -1.150  0.575  2.575 -2.425  5.575  3.300  0.300 
    11     12     13     14     15     16 
 1.300 -3.700  0.025 -1.975  3.025 -3.975 
```

MLE of σ in a different way:

```
> mles=sqrt(sum(Res*Res)/(length(Hardness)))
> mles
[1] 3.025155
```

Checking the Properties of residuals:

1. $\sum_{i=1}^{n} e_i = 0$

```
> sumei=sum(Res)
> sumei
[1] -1.998401e-15
```

2. $\sum_{i=1}^{n} X_i e_i = 0$

```
> sumXiei=sum(Time*Res)
> sumXiei
[1] -6.306067e-14
```

3. $\sum_{i=1}^{n} \widehat{Y}_i e_i = 0$

```
> sumyihatei=sum(Fitvals*Res)
> sumyihatei
[1] -5.782042e-13
```

4. $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \widehat{Y}_i$

```
> sumyi_sumyihat=sum(Hardness)-sum(Fitvals)
> sumyi_sumyihat
[1] 0
```

5. Fitted Regression line passes through the point $(\bar{X}, \bar{Y})$.

```
> XbarYbar=mean(Hardness)-(b0+b1*mean(Time))
> XbarYbar
[1] 0
```