Biological Statistics II

BTRY3020/STSCI3200 - Spring 2024

This page is intentionally blank

Use this page to make a table of contents.

This page is intentionally blank

Chapter 1: Vectors & Matrices

Section 1: Basics

Definition:
An $r \times c$ **matrix** X is an $r \times c$ array of values where r is the number of rows and c is the number of columns. The element in the i^{th} row and j^{th} column is denoted by $x_{i,j}$.[1]

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1c} \\ x_{21} & x_{22} & \cdots & x_{2c} \\ \vdots & \vdots & x_{ij} & \vdots \\ x_{r1} & x_{r2} & \cdots & x_{rc} \end{bmatrix}$$

```
# Matrix X
X <- matrix(6:1, nrow = 2, ncol = 3, byrow = TRUE)
X
     [,1] [,2] [,3]
[1,]    6    5    4
[2,]    3    2    1
X[2, 3]
[1] 1
```

Definition:
The **dimension** of a matrix is the ordered pair (r, c).

```
dim(X)
[1] 2 3
```

Section 2: Special Matrices

Definition:
A **column vector** is an $r \times 1$ matrix.[2]

$$X = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{i1} \\ \vdots \\ x_{r1} \end{bmatrix} = \begin{bmatrix} x_{1} \\ x_{2} \\ \vdots \\ x_{i} \\ \vdots \\ x_{r} \end{bmatrix}$$

Definition:
A **row vector** is a $1 \times c$ matrix.[3]

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1c} \end{bmatrix} = \begin{bmatrix} x_{1} & x_{2} & \cdots & x_{j} & \cdots & x_{c} \end{bmatrix}$$

[1] When it will not cause confusion, the comma is left out.
[2] When it will not cause confusion, only the row index will be listed.
[3] When it will not cause confusion, only the column index will be listed.

Definition:
A **square** matrix is an $r \times c$ matrix where $r = c$.

```
# Matrix Z
Z <- matrix(c(-1, 0, 5, 7, 10, -1, 9, 5, 2), nrow = 3, ncol = 3, byrow = TRUE)
Z
     [,1] [,2] [,3]
[1,]   -1    0    5
[2,]    7   10   -1
[3,]    9    5    2
dim(Z)
[1] 3 3
```

Definition:
The **transpose** of an $r \times c$ matrix $X = [x_{ij}]$ is an $c \times r$ matrix $X^T = [(X^T)_{ij} = x_{ji}]$.

$$X^T = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{c1} \\ x_{12} & x_{22} & \cdots & x_{c2} \\ \vdots & \vdots & x_{ji} & \vdots \\ x_{1r} & x_{2r} & \cdots & x_{cr} \end{bmatrix}$$

```
#### Original Matrix
X
     [,1] [,2] [,3]
[1,]    6    5    4
[2,]    3    2    1

#### Result
t(X)
     [,1] [,2]
[1,]    6    3
[2,]    5    2
[3,]    4    1
```

Definition:
Suppose X is a square matrix. If $X = X^T$ then X is called **symmetric**.

Definition:
A **diagonal** matrix X is a matrix where the elements $x_{ij} = 0$ whenever $i \neq j$. The elements on the **main diagonal** of X are those elements x_{ij} where $i = j$.

```
D <- diag(c(5, 7, 9), nrow = 3, ncol = 5)
D
     [,1] [,2] [,3] [,4] [,5]
[1,]    5    0    0    0    0
[2,]    0    7    0    0    0
[3,]    0    0    9    0    0
```

Definition:
An $n \times n$ **identity** matrix I_n is a square diagonal matrix with ones on the main diagonal.

```
I <- diag(3)
I
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

Section 3: Operations

Scalar Multiplication:

The **scalar multiple** of a $r \times c$ matrix $X = [x_{ij}]$ by constant k is given by $kX = [(kX)_{ij} = kx_{ij}]$

```
#### Original Matrix
X
     [,1] [,2] [,3]
[1,]    6    5    4
[2,]    3    2    1
#### Scalar
k <- 3

#### Result
Y <- k * X
Y
     [,1] [,2] [,3]
[1,]   18   15   12
[2,]    9    6    3
```

Matrix Addition:

The sum of two $r \times c$ matrices $X = [x_{ij}]$ and $Y = [y_{ij}]$ is given by $X + Y = [(X+Y)_{ij} = x_{ij} + y_{ij}]$[4]

```
#### Original Matrices
X
     [,1] [,2] [,3]
[1,]    6    5    4
[2,]    3    2    1

Y
     [,1] [,2] [,3]
[1,]   18   15   12
[2,]    9    6    3

#### Result
X + Y
     [,1] [,2] [,3]
[1,]   24   20   16
[2,]   12    8    4
```

Inner Product:

The **inner product** of two $n \times 1$ column vectors X and Y is given by

$$\langle X, Y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

$$\langle X, Y \rangle \quad = \quad \left\langle \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \end{bmatrix} \right\rangle \quad = \quad (1 \times 5) + (2 \times 6) + (3 \times 7) + (4 \times 8).$$

[4] Matrices can not be added if they do not have the same dimensions. A row and column vector can not be added together. Remember, a vector is R is not the same thing as a row/column vector.

Matrix Multiplication

Matrix Multiplication of a row and column vector:
The product XY of a $1 \times n$ row matrix X (on the left) and an $n \times 1$ column matrix Y (on the right) is given by

$$XY = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

$$(= \langle X^T, Y \rangle).$$

$$XY = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \end{bmatrix} = (1 \times 5) + (2 \times 6) + (3 \times 7) + (4 \times 8)$$

General Matrix Multiplication:
The product of an $r \times d$ matrix $X = [x_{ij}]$ and a $d \times c$ matrix $Y = [y_{ij}]$ is given by an $r \times c$ matrix $XY = [(XY)_{ij} = X_{i.}Y_{.j}]$.

$$\begin{bmatrix} \leftarrow & X_{1.} & \rightarrow \\ \leftarrow & X_{2.} & \rightarrow \\ \vdots & \vdots & \vdots \\ \leftarrow & X_{r.} & \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ Y_{.1} & Y_{.2} & \cdots & Y_{.c} \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} = \begin{bmatrix} X_{1.}Y_{.1} & X_{1.}Y_{.2} & \cdots & X_{1.}Y_{.c} \\ X_{2.}Y_{.1} & X_{2.}Y_{.2} & \cdots & X_{2.}Y_{.c} \\ \vdots & \vdots & X_{i.}Y_{.j} & \vdots \\ X_{r.}Y_{.1} & X_{r.}Y_{.2} & \cdots & X_{r.}Y_{.c} \end{bmatrix}$$

Example:
Determine the dimension of each of the following products:

$$XY = \begin{bmatrix} 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ -3 & 4 \\ -5 & 6 \end{bmatrix} \qquad YX = \begin{bmatrix} -1 & 2 \\ -3 & 4 \\ -5 & 6 \end{bmatrix} \begin{bmatrix} 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix}$$

Example:
Determine the dimension of each of the following products and compute the entries:

a. $XZ = \begin{bmatrix} 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 5 \\ 7 & 10 & -1 \\ 9 & 5 & 2 \end{bmatrix}$

```
X %*% Z
     [,1] [,2] [,3]
[1,]   65   70   33
[2,]   20   25   15
```

b. $ZX = \begin{bmatrix} -1 & 0 & 5 \\ 7 & 10 & -1 \\ 9 & 5 & 2 \end{bmatrix} \begin{bmatrix} 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix}$

```
Z %*% X
Error in Z %*% X: non-conformable arguments
```

Facts about Matrix Multiplication

1. Suppose X is an $r \times t$ matrix and Y is a $v \times c$ matrix.
    a. The product XY is only defined when $t = v$.
    b. If $t \neq r$, the matrices are said to be **non-conformable**.

2. Matrix Multiplication is **not** commutative.
    a. XY may not equal YX

3. Matrix Multiplication is distributive
    a. $X(Y + Z) = XY + XZ$
    b. $(Y + Z)X = YX + ZX$

4. Matrix Multiplication is associative
    a. $(XY)Z = X(YZ)$

5. $(XY)^T = Y^T X^T$

Section 4: Inverses

Recall:
An $n \times n$ **identity** matrix I_n is a square diagonal matrix with ones on the main diagonal.

Definition:
The **inverse** of a matrix X, if such a matrix exists, denoted by X^{-1} is any matrix such that $X(X^{-1}) = (X^{-1})X = I$. A matrix with an inverse is called **invertible**. Otherwise, it is called **singular**.

```
mat <- matrix(c(1, 2, 4, 1), nrow = 2)
mat
     [,1] [,2]
[1,]    1    4
[2,]    2    1

matInv <- solve(mat)
matInv
           [,1]       [,2]
[1,] -0.1428571  0.5714286
[2,]  0.2857143 -0.1428571

mat %*% matInv
     [,1] [,2]
[1,]    1    0
[2,]    0    1
```

1. Suppose X is an $r \times c$, then
    a. $I_r X = X$
    b. $X I_c = X$

2. If a matrix is invertible then it is a square matrix.
    a. A square matrix can be singular.

3. If a matrix is singular then one row(column) is a sum of multiples of the other rows(columns).

4. If X is symmetric and invertible, then $(X^T)^{-1} = X^{-1}$

Section 5: Random Matrices

Definition:
A **random matrix** X is an $r \times c$ matrix whose elements are random variables X_{ij}.

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1c} \\ X_{21} & X_{22} & \cdots & X_{2c} \\ \vdots & \vdots & X_{ij} & \vdots \\ X_{r1} & X_{r2} & \cdots & X_{rc} \end{bmatrix}.$$

Definition:
A **random column vector** is an $r \times 1$ random matrix. A **random row vector** is an $1 \times c$ random matrix[5].

<u>Expected Value</u>

The **mean**, or **expected value**$\left(\mu_X, E(X)\right)$, of a random variable X is

$$E(X) = \mu = \begin{cases} \displaystyle\sum_{x \in S} x\, p(x) & X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} x\, f(x)\, dx & X \text{ is continuous.} \end{cases}$$

Definition:
The **expected value** $E(X)$ of a $r \times c$ random matrix $X$is given by

$$E(X) = \begin{bmatrix} E(X_{11}) & E(X_{12}) & \cdots & E(X_{1c}) \\ E(X_{21}) & E(X_{22}) & \cdots & E(X_{2c}) \\ \vdots & \vdots & E(X_{ij}) & \vdots \\ E(X_{r1}) & E(X_{r2}) & \cdots & E(X_{rc}) \end{bmatrix}.$$

Facts:
1. A random sample $X_1, X_2, \ldots, X_n$ can be represented with a random column vector $X = [X_1 \quad X_2 \quad \cdots \quad X_n]^T$.
2. Matrix operations can be carried out between random and non-random matrices.
3. If A a matrix of constants, X is a random vector, b is a constant vector, and $Y = AX + b$, then $E(Y) = AE(X) + b$.

Example:
Suppose X is a random vector with $n \geq 1$ independent and identically distributed elements, and 1_n is an n element row vector whose elements are all ones. Determine $E\left(\frac{1}{n} 1_n X\right)$.

[5] First, unless it will cause confusion, when a random column/row vector is being used, only one indexing value will be used. Second, assume a given vector is a column vector if there is no indication given as to whether it is a row or column vector.

<u>Covariance</u>

Definition:
The **covariance** of a two random variables X, Y is $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$.

$$Cov(X,Y) = \begin{cases} \displaystyle\sum_{(x,y)\in S} (x-\mu_X)\,(y-\mu_Y)p(x,y) & (X,Y) \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)\,(y-\mu_Y)f(x,y)\,dx\,dy & f(x,y) \text{ is continuous.} \end{cases}$$

The **variance** of a random variable X is $Cov(X,X)$.

$$V(X) = \begin{cases} \displaystyle\sum_{x\in S} (x-\mu)^2\,p(x) & X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} (x-\mu)^2\,f(x)\,dx & X \text{ is continuous.} \end{cases}$$

Definition:
The **variance/covariance matrix** $V(X)$ of a n element random vector X with mean vector $\mu = E(X) = [\mu_1 \quad \mu_2 \quad \cdots \quad \mu_n]^T$ is given by

$$\begin{aligned} V(X) &= E[(X-\mu)(X-\mu)^T] \\ &= E\left(\begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ \vdots \\ X_n-\mu_n \end{bmatrix} [X_1-\mu_1 \quad X_2-\mu_2 \quad \cdots \quad X_n-\mu_n]\right) \\ &= E\begin{bmatrix} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & \cdots & (X_1-\mu_1)(X_n-\mu_n) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & \cdots & (X_2-\mu_2)(X_n-\mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & & \\ (X_n-\mu_n)(X_1-\mu_1) & (X_n-\mu_n)(X_2-\mu_2) & \cdots & (X_n-\mu_n)^2 \end{bmatrix} \end{aligned}$$

Facts:
Suppose X is a random vector with n elements.

1. $V(X)$ is an $n \times n$(square) symmetric matrix.
    a. The values on the main diagonal represent the variances of each element of X.
    b. The values off the main diagonal represent covariances.
2. If the elements of X are independent, then $V(X)$ is a diagonal matrix.
3. If A a matrix of constants, b is a constant vector, $Y = AX + b$, and the matrix operations make sense, then $V(Y) = AV(X)A^T$.

Example:
Suppose X is a random vector with $n \geq 1$ independent and identically distributed elements, and 1_n is an n element row vector whose elements are all ones. Determine $V\left(\frac{1}{n}1_nX\right)$.

Chapter 2: Regression Analysis

Definition:
Regression Analysis is a statistical methodology that examines the relationship between a (collection of) predictor variable(s) X and a response variable Y. The goal is to predict a value for the response based on a specified values of the predictor(s).

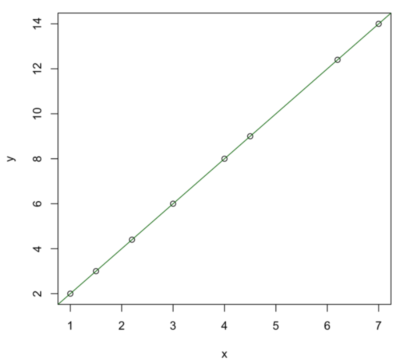
In some situations,

a. predictor variables could be referred to as independent or explanatory variables;
b. response variables could be referred to as dependent or outcome variables.

Section 1: Relations between Variables

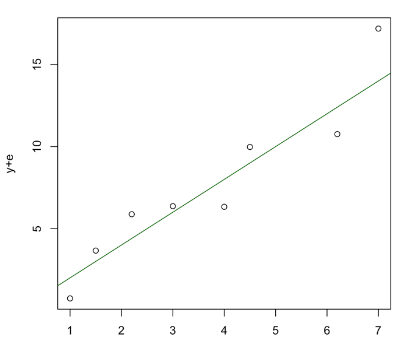
Functional Relationship:
The response variable Y is a function $f(X)$ of the predictor variable(s) X.

```
set.seed(1)
par(mfrow = c(1, 1))
x <- c(1, 3, 4, 7, 1.5, 6.2, 4.5, 2.2)
ln <- length(x)
y <- 2 * x
plot(x, y)
abline(0, 2, col = "darkgreen")
```



Statistical Relationship:
The response variable Y is a not a function of the predictor variable(s) X. However, given a value of X, something is known about Y.[6]

```
y.plus.e <- y + rnorm(ln, sd = 2)
plot(x, y.plus.e, ylab = "y+e")
abline(0, 2, col = "darkgreen")
```



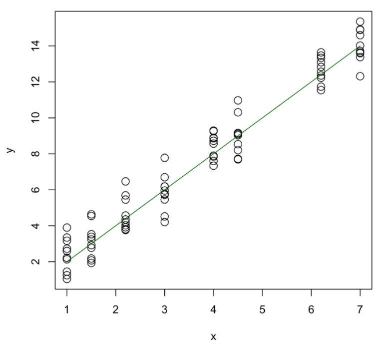
[6] In some cases, there is no interesting relationship between the predictor(s) and the response.

Section 2: Regression Models

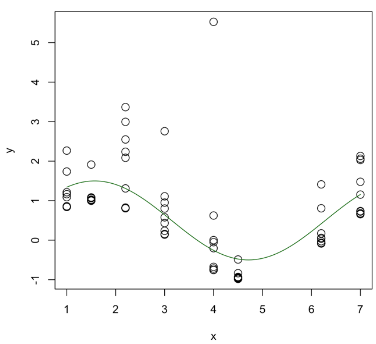
Definition:
A **regression model** is a formal description of the statistical relationship. The model specifies two characteristics:

1. The probability distribution of the response variable at each level of the predictor variables.
2. The functional relationship between the mean of the response variable and the predictor variables.

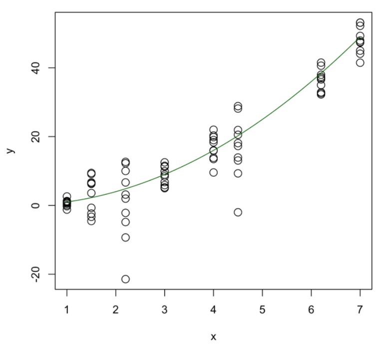
```
set.seed(4)
x <- rep(c(1, 3, 4, 7, 1.5, 6.2, 4.5, 2.2), times = 10)
ln <- length(x)
y <- 2 * x + rnorm(ln)
plot(x, y, cex = 1.5)
curve(2 * x, from = 1, to = 7, col = "darkgreen", add = T)
```



```
y <- sin(x) + rchisq(ln, df = 0.5)
plot(x, y, cex = 1.5)
curve(sin(x) + 0.5, from = 1, to = 7, col = "darkgreen", add = T)
```



```
y <- x^2 + rnorm(ln, sd = 1:(ln/10))
plot(x, y, cex = 1.5)
curve(x^2, from = 1, to = 7, col = "darkgreen", add = T)
```



Section 3: Simple Linear Regression - Unspecified Error

Simple Linear Regression Model:
A **simple linear regression** model relates the response variable Y_i and a single predictor variable X_i as follows:

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

where

1. Y_i is the value of the response in the i^{th} trial.

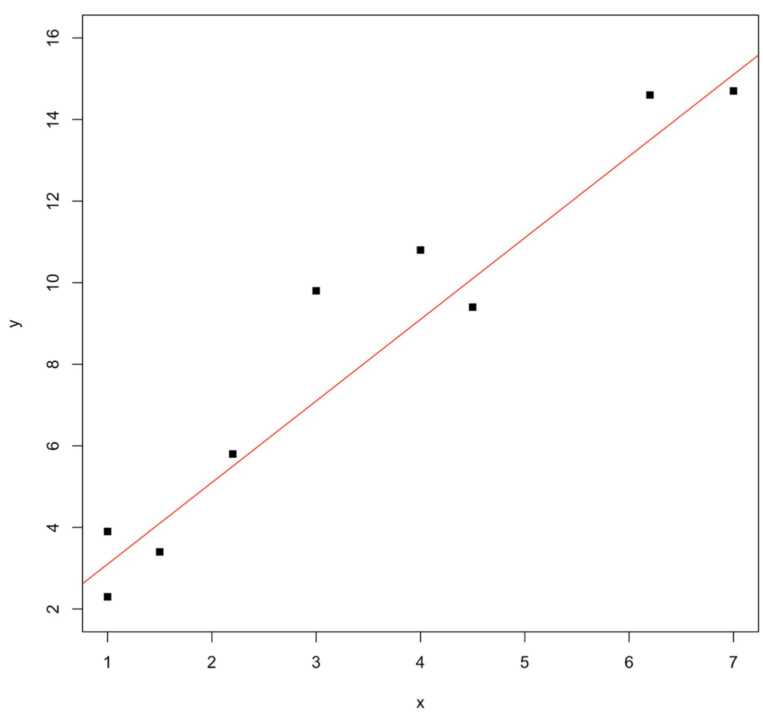
2. β_0, β_1 are unknown intercept and slope parameters.

3. X_i is the value of the predictor in the i^{th} trial, and is a known constant.

4. ϵ_i is the random error (variable) on the i^{th} trial where $E(\epsilon_i) = 0$, $V(\epsilon_i) = \sigma^2 > 0$, and $Cov(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

```
(1,2.3)  (3,9.8)  (4,10.8)  (7,14.7)  (1.5,3.4)  (6.2,14.6)  (4.5,9.4)  (2.2,5.8)  (1,3.9)
```




```
set.seed(14)
x <- c(1, 3, 4, 7, 1.5, 6.2, 4.5, 2.2, 1)
error <- rchisq(9, df = 1) - 1
y <- 2 * x + 1.1 + error
y <- round(y, digits = 1)
plot(x, y, pch = 15, ylim = c(2, 16))
abline(a = 1.1, b = 2, col = "red")
```

Section 4: Estimation of Regression Function

Method of Least Squares

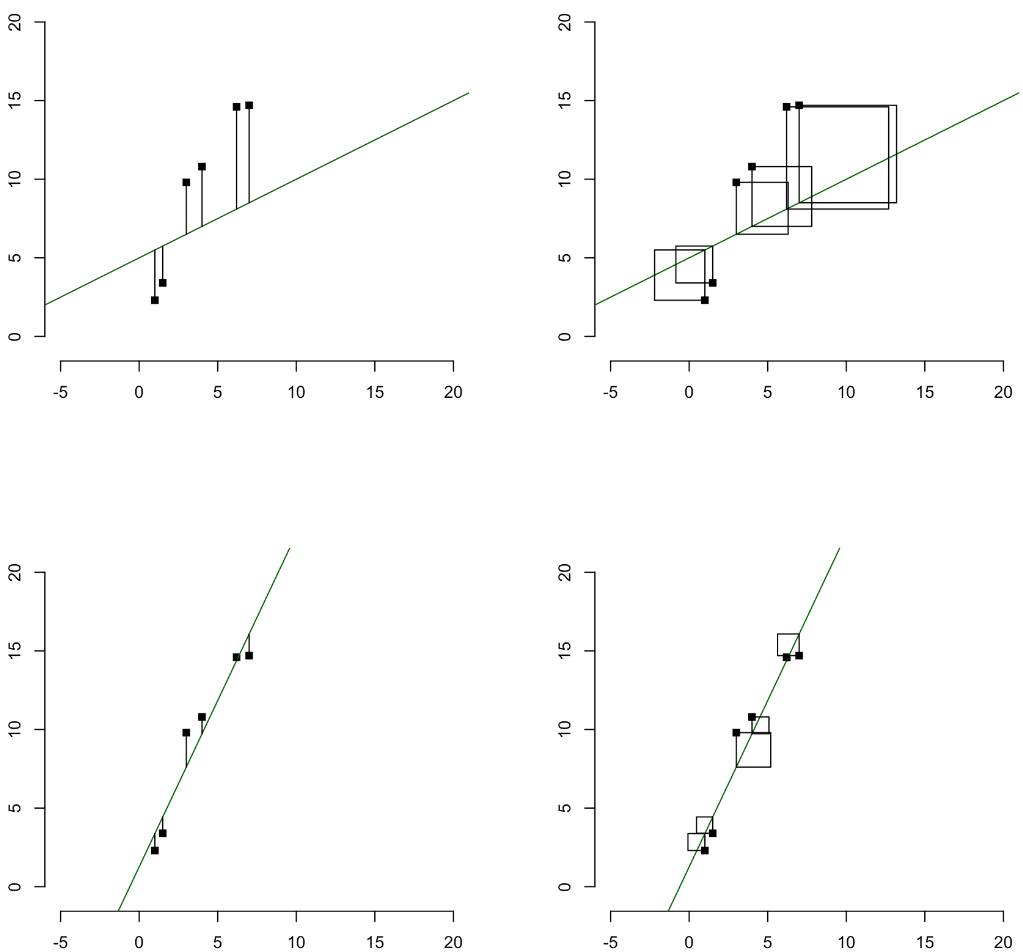
Definition:
Given a data set $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$,, the **method of least squares** estimates the intercept β_0 and slope β_1 parameters. It works by minimizing the function

$$\begin{aligned} SS(b_0^*, b_1^*) &= \sum_{i=1}^{n} [y_i - (b_0^* + b_1^* X_i)]^2 \\ &= (Y - Xb^*)^T (Y - Xb^*) \end{aligned}$$

The values of b_0^* and b_1^* that minimize $SS(b_0^*, b_1^*)$ will be denoted by b_0 and b_1.

a. b_0 will be called the **least squares estimate** of the intercept β_0.
b. b_1 will be called the **least squares estimate** of the slope β_1.

$$SS(b_0, b_1) \leq \min_{\substack{\text{all intercepts } b_0^* \\ \text{all slopes } b_1^*}} SS(b_0^*, b_1^*)$$



Least Squares Estimators

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

1. For the simple linear regression model, the least squares estimates of β_0 and β_1 are given by

$$b_0 = \bar{y} - \bar{x} b_1 \qquad b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

2. Different methods exist for minimizing $SS(b_0^*, b_1^*)$.
   a. Guess and Check

   b. Calculus

   c. Matrix Algebra

3. For the simple linear regression model, b_0 and b_1 are the solutions to the **normal equations**.

$$\begin{aligned} \sum Y_i &= nb_0 + b_1 \sum X_i \\ \sum X_i Y_i &= b_0 \sum X_i + b_1 \sum X_i^2 \end{aligned} \qquad X^T Y = X^T X b$$

4. For the simple linear regression model, the least squares estimates b_0 and b_1 are unbiased estimates of β_0 and β_1.

5. For the simple linear regression model, the least squares estimates b_0 and b_1 are have the smallest variance among all possible unbiased linear estimates of β_0 and β_1.

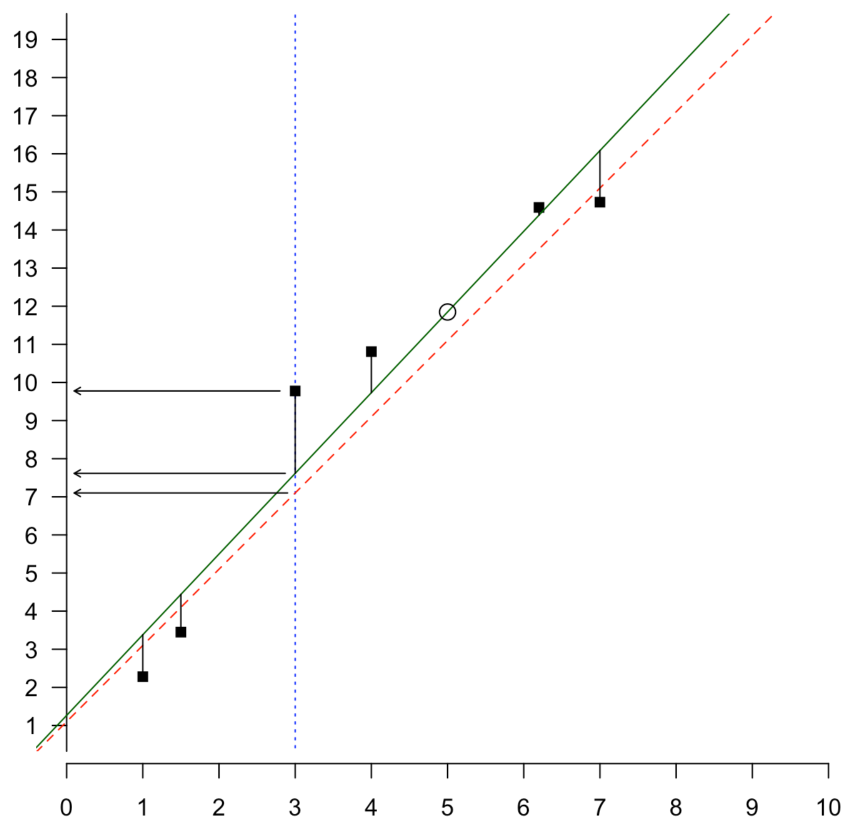
Point Estimate of the Mean Response

Definition:
The **equation of the least squares regression line** is given by

$$\hat{Y} = b_0 + b_1 X,$$

where b_0 and b_1 are estimates of β_0 and β_1.

Definition:
The **mean response** at X is given by $E(Y) = \beta_0 + \beta_1 X$. The estimate of the mean response at X is given by evaluating the equation of the least squares regression line at X. When X is one of the observed values of the independent variable, the **mean response** at X is called a **fitted value**.



Residuals

Definition:
The i^{th} **residual** e_i is the difference between the i^{th} observed value and the i^{th} fitted value. That is,

$$e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} - \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}$$

Definition:
The i^{th} **residual** e_i is the difference between the i^{th} observed value and the i^{th} fitted value. That is,

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

Properties of the Fitted Regression Line

1. The sum of the residuals is zero. $\sum e_i = 0$

2. The sum of the squared residuals is a minimum.

3. The sum of the observed values is equal to the sum of the fitted values: $\sum Y_i = \sum \hat{Y}_i$.

4. The sum of the weighted residuals is zero, when the weight on the i^{th} residual is X_i: $\sum X_i e_i = 0$.

5. The sum of the weighted residuals is zero, when the weight on the i^{th} residual is $\hat{Y}_i$: $\sum \hat{Y}_i e_i = 0$.

6. The regression line always goes through the point $(\bar{X}, \bar{Y})$: $\bar{Y} = b_0 + b_1\bar{X}$

Section 5: Estimating the Variance of the Error

$\sigma^2 = E[(Y - \mu_Y)^2]$

<u>Single Population</u>

<u>Known Mean</u>

$s^2 = \frac{\sum_{i=1}^{n}(Y_i - \mu_Y)^2}{n}$

<u>Unknown Mean</u>

$$SSE = \sum(Y_i - \bar{Y})^2$$

$$DFE = n - 1$$

$$s^2 = \frac{SSE}{DFE}$$

<u>Regression</u> <u>$\sigma^2 = V(\epsilon) = V(Y \mid X)$</u>

The **sum of square error (SSE)** or **sum of square residuals (SSR)** is given by $SSE = \sum\left(Y_i - \hat{Y}_i\right)^2$.

The **degrees of freedom error (DFE)**[7] or **degrees of freedom residuals (DFR)** is given by $DFE = n - 2$.

$$SSE = \sum\left(Y_i - \hat{Y}_i\right)^2$$

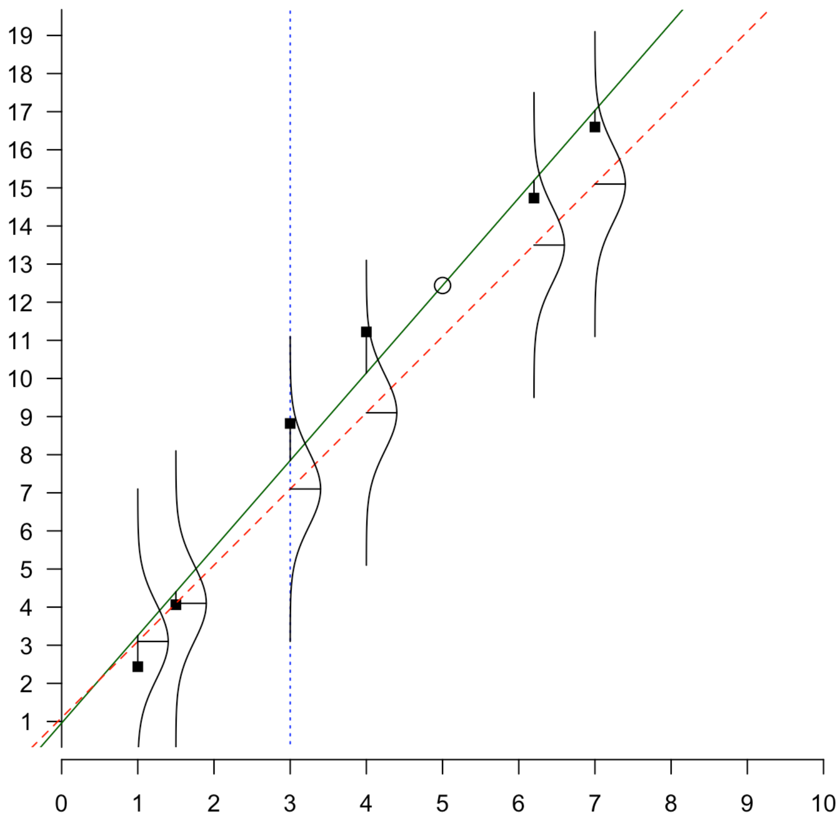
$$DFE = n - 2$$

$$s^2 = \frac{SSE}{DFE}$$

[7] In many cases, the degrees of freedom can be computed using the following guide: DF = number of independent pieces of information (n) minus number of estimated parameters.

Section 6: Normal Errors

Normal Error Regression Model:
A **normal error regression** model is a simple linear regression model where the ϵ_i are independent normally distributed random variables with mean zero and common variance σ^2.



Maximum Likelihood Estimators

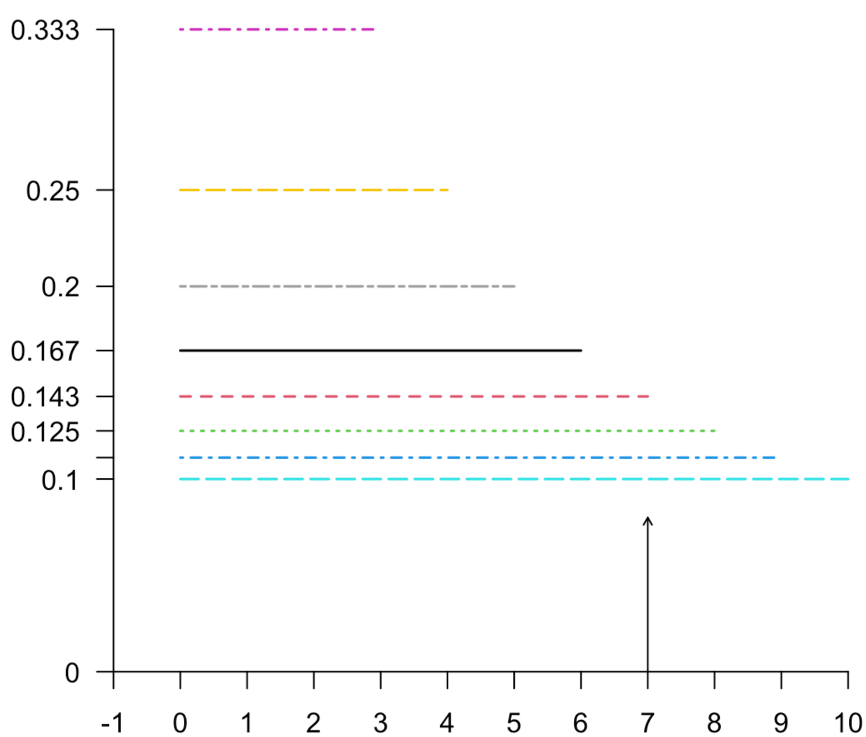
When the distribution of the error terms are specified, another method can be used to estimate β_0, β_1, and σ^2.

Definition:
Given a random variable X with density function $f(x|\theta)$, that is dependent on a parameter θ, and an observed value of $X = x$, the **likelihood function** is $L(\theta|x) = f(x|\theta)$.

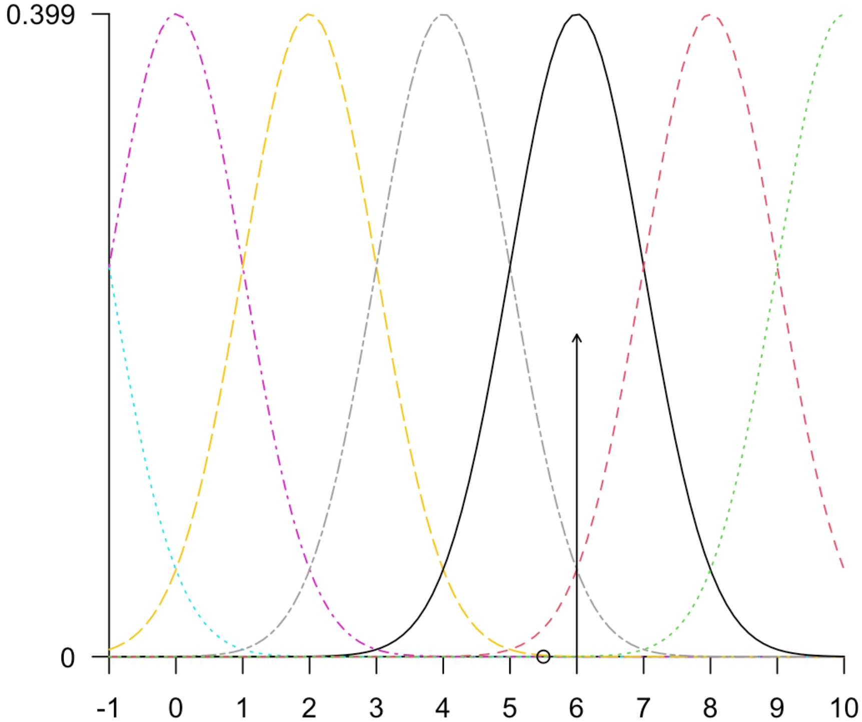
Definition:
The **maximum likelihood estimate** of a parameter θ is the value (mle θ) that makes the observed data most likely.

$$L(\text{ mle } \theta|x) = \max_{\text{all possible } \theta} L(\theta|x)$$

Example:
Suppose a population follows a uniform distribution on the interval $[0, b]$. You have taken a sample of size 1. The observed value is 7. Determine the maximum likelihood estimate of b.



Example:
Suppose a population follows a normal distribution with a standard deviation of one. You have taken a sample of size 1. The observed value is 5.5. Determine the maximum likelihood estimate of the mean of the distribution.



<u>Properties of Maximum Likelihood Estimators</u>

1. If $\hat{\theta}$ is a maximum likelihood estimator of a parameter θ, and $g(\theta)$ is a function of θ, then $g(\hat{\theta})$ is a maximum likelihood estimator of a parameter $g(\theta)$.

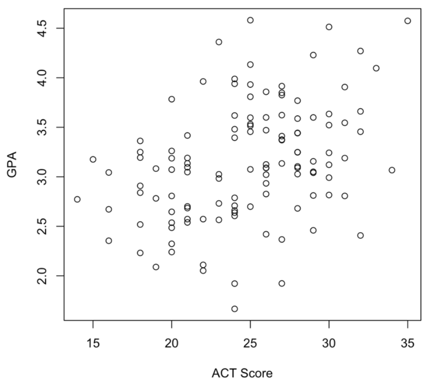
2. Using the normal regression model, the maximum likelihood estimators of β_0 and β_1 are b_0 and b_1, respectively.

3. Using the normal regression model, the maximum likelihood estimators of σ^2 is $\frac{\sum(Y_i - \hat{Y}_i)^2}{n}$.

Section 7: Example & Code

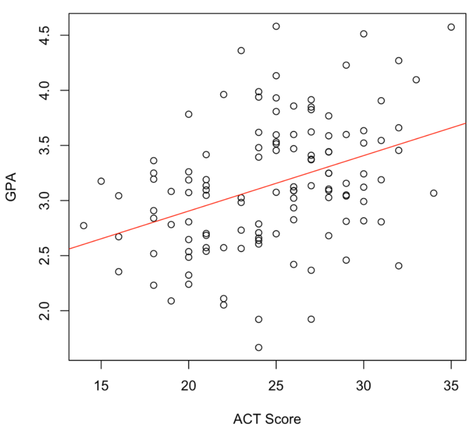
Example:
The director of admissions at a college wants to describe the GPA of a freshmen at the end of their freshmen year based upon their ACT score. They will fit a simple linear model.

```
#### A portion of the dataset
head(SLRSet1)
       GPA ACT
1 2.539076  21
2 2.772057  14
3 2.680529  28
4 3.962266  22
5 3.134735  21
6 2.806456  31

#### Sample Size
nrow(SLRSet1)
[1] 120

plot(SLRSet1$ACT, SLRSet1$GPA, xlab = "ACT Score", ylab = "GPA")
```




```
#### Fitting Linear Model
lm.fit <- lm(GPA ~ ACT, data = SLRSet1)

#### Output of Fitted Model
lm.fit

Call:
lm(formula = GPA ~ ACT, data = SLRSet1)

Coefficients:
(Intercept)          ACT
    1.89783      0.05034

#### Summarized Output of Fitted Model
summary(lm.fit)

Call:
lm(formula = GPA ~ ACT, data = SLRSet1)

Residuals:
     Min       1Q   Median       3Q      Max
-1.44000 -0.37148 -0.04281  0.36769  1.42492

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.89783    0.28280   6.711 7.11e-10 ***
ACT          0.05034    0.01126   4.472 1.80e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5492 on 118 degrees of freedom
Multiple R-squared:  0.1449,    Adjusted R-squared:  0.1377
F-statistic:    20 on 1 and 118 DF,  p-value: 1.796e-05
```

```
#### Contents of Fitted Model
names(lm.fit)
 [1] "coefficients"  "residuals"     "effects"       "rank"
 [5] "fitted.values" "assign"        "qr"            "df.residual"
 [9] "xlevels"       "call"          "terms"         "model"

coefficients(lm.fit)
(Intercept)         ACT
 1.89783001  0.05033694

#### Computing Fitted Values
predict(lm.fit)[1:5]
       1        2        3        4        5
2.954906 2.602547 3.307264 3.005243 2.954906

fitted.values(lm.fit)[1:5]
       1        2        3        4        5
2.954906 2.602547 3.307264 3.005243 2.954906


#### Computing Estimates of Mean Response
coefficients(lm.fit)[1] + coefficients(lm.fit)[2] * c(4, 0, 2.1)
[1] 2.099178 1.897830 2.003538


nd <- data.frame(ACT = c(4, 0, 2.1))
predict(lm.fit, newdata = nd)
       1        2        3
2.099178 1.897830 2.003538

#### Computing residual
SLRSet1$GPA[1:5] - fitted.values(lm.fit)[1:5]
         1          2          3          4          5
-0.4158302  0.1695100 -0.6267357  0.9570229  0.1798295


residuals(lm.fit)[1:5]
         1          2          3          4          5
-0.4158302  0.1695100 -0.6267357  0.9570229  0.1798295
sd(residuals(lm.fit))
[1] 0.5468474

#### Alternative Graphing Method
plot(GPA ~ ACT, data = SLRSet1, xlab = "ACT Score", ylab = "GPA")
abline(coefficients(lm.fit), col = "red")
```



Chapter 3: Inference with Simple Linear Regression

Data

$(X_1, Y_1), (X_2, Y_2), \cdots, (X_i, Y_i), \cdots, (X_n, Y_n)$

Linear Model

$$
\begin{array}{ccccccc}
Y_1 & = & \beta_0 \cdot 1 & + & \beta_1 X_1 & + & \epsilon_1 \\
Y_2 & = & \beta_0 \cdot 1 & + & \beta_1 X_2 & + & \epsilon_2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
Y_i & = & \beta_0 \cdot 1 & + & \beta_1 X_i & + & \epsilon_i \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
Y_n & = & \beta_0 \cdot 1 & + & \beta_1 X_n & + & \epsilon_n
\end{array}
$$

$$
\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_n \end{bmatrix}
$$

$$
\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_i \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_n \end{bmatrix}
$$

Parameter Estimates

Slope & Intercept:

$$
b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \bar{Y} - b_1 \bar{X} \\ \dfrac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \end{bmatrix} = (X^T X)^{-1} X^T Y
\qquad
E(b) = \begin{bmatrix} E(b_0) \\ E(b_1) \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}
$$

Standard Errors:

$$
V(b) = \sigma^2 (X^T X)^{-1} = \begin{bmatrix} \dfrac{\sigma^2}{n} + \dfrac{\bar{X}^2 \sigma^2}{\sum (X - \bar{X})^2} & \dfrac{-\bar{X}\sigma^2}{\sum (X - \bar{X})^2} \\ \dfrac{-\bar{X}\sigma^2}{\sum (X - \bar{X})^2} & \dfrac{\sigma^2}{\sum (X - \bar{X})^2} \end{bmatrix}
$$

Section 1: Inference about β_1

1. This inference assumes a normal error regression model.
2. For a 1 unit change in the predictor, there will be a β_1 change in mean response.

Sampling Distribution of b_1

The random variable b_1 is a normally distributed random variable with mean and variance given by

$$E(b_1) = \beta_1$$

$$V(b_1) = \frac{\sigma^2}{\sum(X_i - \bar{X})^2} \approx \frac{s^2}{\sum(X_i - \bar{X})^2} = \frac{MSE}{\sum(X_i - \bar{X})^2}.$$

$$b_1 \sim \text{Normal}\big(\beta_1, V(\beta_1)\big)$$

$$\sigma^2 = V(\epsilon) = V(Y \mid X)$$

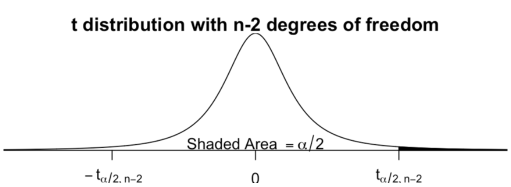
Sampling Distribution of $t = \frac{b_1 - \beta_1}{s_{b_1}}$

The statistic $t = \frac{b_1 - \beta_1}{s_{b_1}}$ has a t distribution with $n-2$ degrees of freedom where s_{b_1} is the estimated standard error of b_1.

$100(1-\alpha)\%$ Confidence Interval for β_1

Given a random sample of size n from a population, the $100(1-\alpha)\%$ confidence interval for β_1 is given by

$$b_1 - t_{\alpha/2,n-2}s_{b_1} < \beta_1 < b_1 + t_{\alpha/2,n-2}s_{b_1}.$$



Testing for β_1

Hypotheses:
H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$

Test Statistic:
$t = \frac{b_1}{s_{b_1}}$

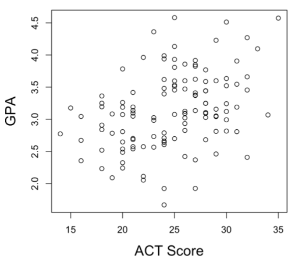
Level α Rejection Rule:
Reject H_0 if

a. $t < -t_{\alpha/2,n-2}$ or $t > t_{\alpha/2,n-2}$
b. p - value $\leq \alpha$
   - p - value = 2 × pt($|t|$, df = n-2, lower.tail = F)

Example:

The director of admissions at a college wants to describe the GPA of a freshmen at the end of their freshmen year based upon their ACT score. They will fit a simple linear model[8].(Kutner)

```
#### A portion of the data set
head(SLRSet1, n = 2)
       GPA ACT
1 2.539076  21
2 2.772057  14

#### Sample Size
n <- nrow(SLRSet1)
n
[1] 120

#### Graph
plot(GPA ~ ACT, data = SLRSet1, xlab = "ACT Score", ylab = "GPA", cex.lab = 1.5)
```




```
#### Fitting Linear Model
lm.fit <- lm(GPA ~ ACT, data = SLRSet1)

#### Summarized Output of Fitted Model
summary(lm.fit)

Call:
lm(formula = GPA ~ ACT, data = SLRSet1)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.44000 -0.37148 -0.04281  0.36769  1.42492 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.89783    0.28280   6.711 7.11e-10 ***
ACT          0.05034    0.01126   4.472 1.80e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5492 on 118 degrees of freedom
Multiple R-squared:  0.1449,	Adjusted R-squared:  0.1377 
F-statistic:    20 on 1 and 118 DF,  p-value: 1.796e-05

#### Confidence Interval
confint(lm.fit, parm = 2, level = 0.99)
         0.5 %     99.5 %
ACT 0.02086496 0.07980892

#### Hand Calculations
cv <- qt(0.995, df = n - 2, lower.tail = TRUE)
b1 <- coefficients(lm.fit)[2]
b1
       ACT 
0.05033694 
sdres <- sd(lm.fit$residuals) * sqrt(((n - 2) + 1)/(n - 2))
sdres
[1] 0.5491597
ssx <- var(SLRSet1$ACT) * ((n - 2) + 1)
se.b1 <- sdres/sqrt(ssx)
b1 + c(-cv, cv) * se.b1
[1] 0.02086496 0.07980892
```

[8] Without checking, it is assumed that a normal error regression model applies.

Section 2: Inference about β_0

1. This inference assumes a normal error regression model.
2. The mean response when the predictor is equal to zero.
3. Important when the predictors are near, or include, zero.

Sampling Distribution of b_0

The random variable b_0 is a normally distributed random variable with mean and variance given by

$$E(b_0) = \beta_0 \qquad\qquad b_0 \sim \text{Normal}\big(\beta_0, V(\beta_0)\big)$$

$$V(b_0) = \sigma^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}\right] \qquad\qquad \sigma^2 = V(\epsilon) = V(Y \mid X)$$

$$\approx s^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}\right] = MSE\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}\right].$$

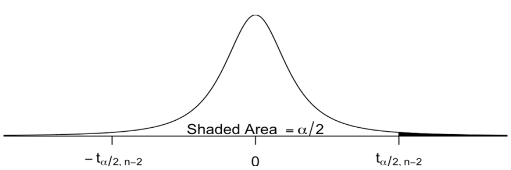
Sampling Distribution of $t = \frac{b_0 - \beta_0}{s_{b_0}}$

The statistic $t = \frac{b_0 - \beta_0}{s_{b_0}}$ has a t distribution with $n-2$ degrees of freedom where s_{b_0} is the estimated standard error of b_0.

$100(1-\alpha)\%$ Confidence Interval for β_0

Given a random sample of size n from a population, the $100(1-\alpha)\%$ confidence interval for β_0 is given by

$$b_0 - t_{\alpha/2,n-2}s_{b_0} < \beta_0 < b_0 + t_{\alpha/2,n-2}s_{b_0}.$$



Testing for β_0

Hypotheses:
H_0: $\beta_0 = 0$
H_1: $\beta_0 \neq 0$

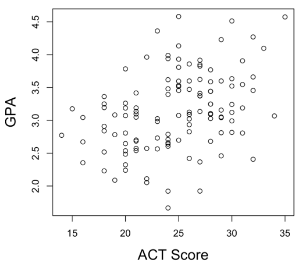
Test Statistic:
$t = \frac{b_0}{s_{b_0}}$

Level α Rejection Rule:
Reject H_0 if

a. $t < -t_{\alpha/2,n-2}$ or $t > t_{\alpha/2,n-2}$
b. p - value $\leq \alpha$
   – p - value = 2 × pt(| t |, df = n-2, lower.tail = F)

Example:
The director of admissions at a college wants to describe the GPA of a freshmen at the end of their freshmen year based upon their ACT score. They will fit a simple linear model[9].(Kutner)

```
#### Graph
plot(GPA ~ ACT, data = SLRSet1, xlab = "ACT Score", ylab = "GPA", cex.lab = 1.5)
```




```
#### Fitting Linear Model
lm.fit <- lm(GPA ~ ACT, data = SLRSet1)

#### Summarized Output of Fitted Model
summary.fit <- summary(lm.fit)
summary.fit

Call:
lm(formula = GPA ~ ACT, data = SLRSet1)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.44000 -0.37148 -0.04281  0.36769  1.42492 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.89783    0.28280   6.711 7.11e-10 ***
ACT          0.05034    0.01126   4.472 1.80e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5492 on 118 degrees of freedom
Multiple R-squared:  0.1449,    Adjusted R-squared:  0.1377 
F-statistic:    20 on 1 and 118 DF,  p-value: 1.796e-05

names(summary.fit)
 [1] "call"          "terms"         "residuals"     "coefficients"
 [5] "aliased"       "sigma"         "df"            "r.squared"   
 [9] "adj.r.squared" "fstatistic"    "cov.unscaled" 

#### Confidence Interval
confint(lm.fit, parm = 1, level = 0.9)
                 5 %     95 %
(Intercept) 1.428977 2.366683

#### Hand Calculations
n <- nrow(SLRSet1)
dof <- n - 2
alpha <- 0.1
cv <- qt(alpha/2, df = dof, lower.tail = FALSE)
b0 <- coefficients(summary.fit)[1]
b0
[1] 1.89783
sdres <- summary.fit$sigma
sdres
[1] 0.5491597
Xs <- 1/n + (mean(SLRSet1$ACT)^2)/(var(SLRSet1$ACT) * (n - 1))
se.b0 <- sdres * sqrt(Xs)
b0 + c(-cv, cv) * se.b0
[1] 1.428977 2.366683
```

[9] Without checking, it is assumed that a normal error regression model applies.

Section 3: Interval Estimation of $E[Y_h]$

Assumptions:

1. This inference assumes a normal error regression model.
2. Let $X_i, i = 1,2,\dots,n$ be the observed values of the predictor variable.
3. Let $X_h, h = 1,2,\dots,H$ be values for the predictor variable where one wants to estimate the mean response $E(Y \mid X_h)$.

Definition:

The point estimate of mean response at X_h, $E[Y_h \mid X_h] = \beta_0 + \beta_1 X_h$, is $\hat{Y}_h = b_0 + b_1 X_h$.

Sampling Distribution of $\hat{Y}_h$

The random variable $\hat{Y}_h$ is a normally distributed random variable with mean and variance given by

$$E[\hat{Y}_h \mid X_h] = E[Y_h]$$

$$V[\hat{Y}_h \mid X_h] = \sigma^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right]$$

$$\approx s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right] = MSE\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right].$$

$$\hat{Y}_h \sim \text{Normal}\left(E[\hat{Y}_h \mid X_h], V[\hat{Y}_h \mid X_h]\right)$$

$$\sigma^2 = V(\epsilon) = V(Y \mid X)$$

Controlling Size of Variance:

1. For any n, $V[\hat{Y}_h \mid X_h] \geq \sigma^2\left[\frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right]$.
2. If $X_h = \bar{X}$, $V[\hat{Y}_h \mid X_h = \bar{X}] = \frac{\sigma^2}{n}$.
3. The size of $V[\hat{Y}_h \mid X_h]$ depends on values that can be selected:
   - Select X_h near $\bar{X}$
   - Select X_i so that $\sum(X_i - \bar{X})^2$ is large.

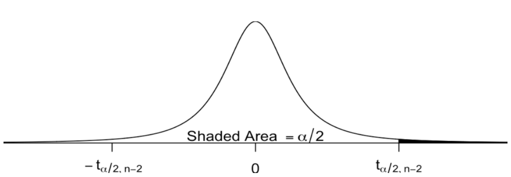
Sampling Distribution of t

The statistic $t = \frac{\hat{Y}_h - E[Y_h]}{s_{\hat{Y}_h}}$ has a t distribution with $n - 2$ degrees of freedom where $s_{\hat{Y}_h}$ is the estimated standard error of $s_{\hat{Y}_h}$.

$100(1-\alpha)\%$ Confidence Interval for $E[Y_h|X_h]$

Given a random sample of size n from a population, the $100(1-\alpha)\%$ confidence interval for $E[Y_h|X_h]$ is given by

$$\hat{Y}_h - t_{\alpha/2,n-2}s_{\hat{Y}_h} < E[Y_h|X_h] < \hat{Y}_h + t_{\alpha/2,n-2}s_{\hat{Y}_h}.$$



Section 4: Prediction from a New Observation

If a new observation is to be collected, predict the individual response.

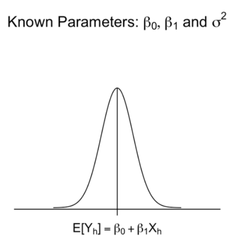
Assumptions:
1. This inference assumes a normal error regression model.
2. Let $X_i, i = 1,2,\dots,n$ be the observed values of the predictor variable.
3. Let $X_h, h = 1,2,\dots,H$ be values for the predictor variable where one wants to predict the response $Y_{h(new)}$ if additional observations were to be taken.

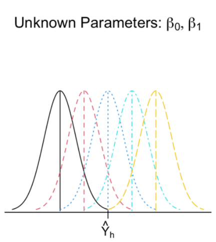
Definition:
If an additional observation is taken at X_h, then predicted response, $Y_{h(new)} = \beta_0 + \beta_1 X_h + \epsilon_h$, is $Y_{h(new)_h} = b_0 + b_1 X_h$.
- $Y_{h(new)}$ is not a parameter
- $Y_{h(new)}$ is random variable.

<u>Sampling Distribution of $\frac{Y_h - \hat{Y}_h}{s_{pred}}$</u>

The random variable $\frac{Y_{h(new)} - \hat{Y}_h}{s_{\widehat{pred}}}$ has a t distribution with $n-2$ degrees of freedom.

$$V(pred) = \sigma^2 + V(\hat{Y}_h)$$





<u>$100(1-\alpha)\%$ **Prediction Interval for $Y_{h(new)}$**</u>

Given a sample of size n from a population that satisfies the normal error regression model, a $100(1-\alpha)\%$ prediction interval for Y_h is given by

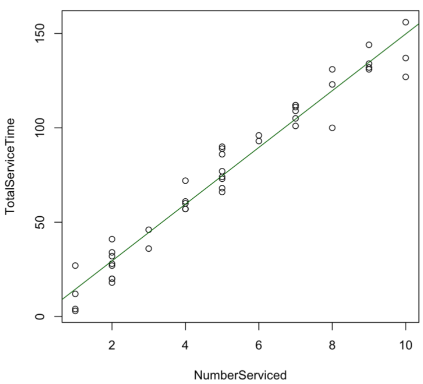
$$\hat{Y}_h - t_{\alpha/2,n-2} s_{pred} < Y_{h(new)} < \hat{Y}_h + t_{\alpha/2,n-2} s_{pred}.$$

Section 5: Example

CopyCo rents office copiers. As part of a copier rental, CopyCo provides preventative maintenance and repair services. The collected data is for forty-five recent preventative maintenance visits. For each visit, the number of copiers serviced was recorded, as well as the total service time per visit[10].(Kutner)

1. Does there appear to be a linear relationship between the number of copiers serviced and the total service time per visit?

```
#### Partial Data Frame
head(CopyCo)
  TotalServiceTime NumberServiced
1               20              2
2               60              4
3               46              3
4               41              2
5               12              1
6              137             10
n <- nrow(CopyCo)

#### Fitting Linear Model
lm.fit <- lm(TotalServiceTime ~ NumberServiced, data = CopyCo)

#### Graph
plot(CopyCo[2:1])
abline(coefficients(lm.fit), col = "darkgreen")
```



2. Estimate the change in the average service time when the number copiers being serviced increases by 1.

```
#### Summarized Output of Fitted Model
summary.fit <- summary(lm.fit)
summary.fit

Call:
lm(formula = TotalServiceTime ~ NumberServiced, data = CopyCo)

Residuals:
     Min       1Q   Median       3Q      Max 
-22.7723  -3.7371   0.3334   6.3334  15.4039 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -0.5802     2.8039  -0.207    0.837    
NumberServiced  15.0352     0.4831  31.123   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.914 on 43 degrees of freedom
Multiple R-squared:  0.9575,    Adjusted R-squared:  0.9565 
F-statistic: 968.7 on 1 and 43 DF,  p-value: < 2.2e-16

#### Confidence Interval
confint(lm.fit, parm = 2, level = 0.9)
                    5 %     95 %
NumberServiced 14.22314 15.84735
```

[10] Without checking, it is assumed that a normal error regression model applies.

3. The manufacturer of the copy machines rented by CopyCo advertises that the average total service time should not increase by more than 14 minutes for each additional copier serviced during a service visit.

```
#### Summarized Output of Fitted Model
names(summary.fit)
 [1] "call"          "terms"         "residuals"     "coefficients"
 [5] "aliased"       "sigma"         "df"            "r.squared"
 [9] "adj.r.squared" "fstatistic"    "cov.unscaled"

summary.fit$coefficients
                 Estimate Std. Error    t value     Pr(>|t|)
(Intercept)    -0.5801567  2.8039411 -0.2069076 8.370587e-01
NumberServiced 15.0352480  0.4830872 31.1232581 4.009032e-31

#### Test Statistic
b1 <- summary.fit$coefficients[2, 1]
b1
[1] 15.03525
se.b1 <- summary.fit$coefficients[2, 2]
se.b1
[1] 0.4830872
hypothesized.value <- 14

t <- (b1 - hypothesized.value)/se.b1
t
[1] 2.142984

#### Critical Values
dof <- n - 2
qt(0.05, df = dof, lower.tail = FALSE)
[1] 1.681071

#### Test Statistic
pt(t, df = dof, lower.tail = FALSE)
[1] 0.01890766
```

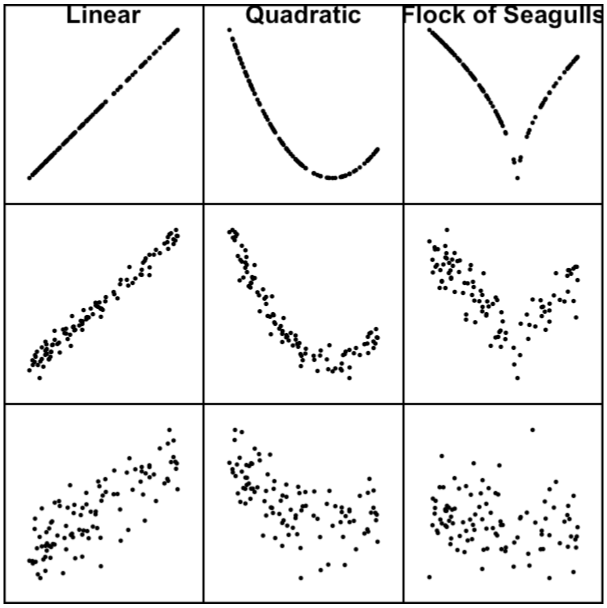
4. Confidence and Prediction Intervals when $X_h = 5{,}6$

```
#### Predictor Values
X_h <- data.frame(NumberServiced = c(5, 6))

#### Confidence Intervals
predict(lm.fit, newdata = X_h, interval = "confidence", level = 0.9)
       fit      lwr      upr
1 74.59608 72.36054 76.83162
2 89.63133 87.28387 91.97880

#### Prediction Intervals
predict(lm.fit, newdata = X_h, interval = "prediction", se.fit = TRUE)
$fit
       fit      lwr       upr
1 74.59608 56.42133  92.77084
2 89.63133 71.43628 107.82639

$se.fit
       1        2
1.329831 1.396411

$df
[1] 43

$residual.scale
[1] 8.913508
```

```
#### Variance/Covariance Matrix
var(CopyCo)
                 TotalServiceTime NumberServiced
TotalServiceTime        1826.7455     116.333333
NumberServiced           116.3333       7.737374
mean(CopyCo$NumberServiced)
[1] 5.111111
```

Section 6: Covariance & Pearson Correlation Coefficient

Definition:
Two numerical variables X and Y are called **correlated** if they are associated.



Definition:
The **covariance** of random variables X and Y is $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$. Covariance measures the tendency of two random variables to change together along a straight line.

Definition:
The **linear correlation coefficient** of random variables X and Y is $\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$. Correlation measures the strength and direction of the linear association between two variables.

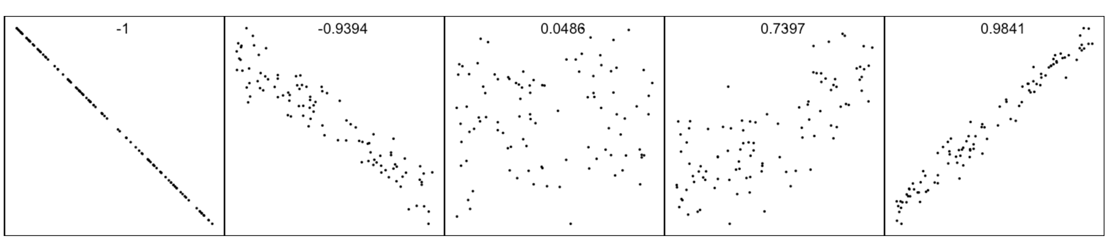
Point Estimates

Given a random sample (x_i, y_i) of size n, the estimates are as follows:

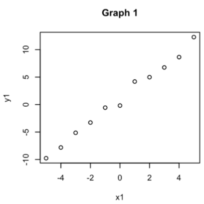
$$cov(x,y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1} \qquad\qquad r = \frac{cov(x,y)}{s_x s_y}$$

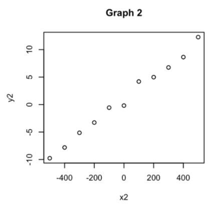
Facts about Covariance and Correlation

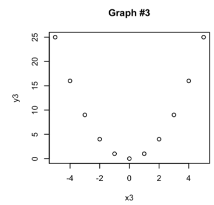
1. Linear correlation coefficient is bounded above by 1 and below by -1.



2. The scale of observations affects size of covariance, but it does not affect the size of correlation







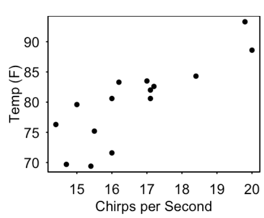

```
#### Graph 1 Data
x1 <- -5:5
ln <- length(x1)
error <- rnorm(ln)
y1 <- 2 * x1 + 1 + error
cov(x1, y1)
[1] 22.49821
cor(x1, y1)
[1] 0.9836314

#### Graph #2 Data
x2 <- x1 * 100
y2 <- 100 * y1
cov(x2, y2)
[1] 224982.1
cor(x2, y2)
[1] 0.9836314
```

3. Linear correlation and covariance only measure straight line relationships.

```
#### Graph #3 Data
x3 <- -5:5
y3 <- x3^2

cor(x3, y3)
[1] 0
```

4. If two random variables X and Y are independent, then Cov(X,Y) and $\rho_{X,Y}$ equal zero. (But not necessarily the other way around.)

Example:
The temperature and the number of times a cricket chirped per second from "The Song of Insects" by Dr.G.W. Pierce.




```
#### Partial Dataframe
head(chirps, n = 2)
  Chirps Temp
1     20 88.6
2     16 71.6

#### Variance/Covariance Matrix
cov(chirps)
         Chirps      Temp
Chirps 2.896952  9.534143
Temp   9.534143 44.988286

#### Correlation Matrix
cor(chirps)
          Chirps      Temp
Chirps 1.0000000 0.8351438
Temp   0.8351438 1.0000000
```

Chapter 4: Multiple Linear Regression

Section 1: General Linear Model with normal error

Definition:
A **general linear regression model with normal error** relates the response variable Y_i and $p-1$ predictor variables $X_1, X_2, \ldots, X_{p-1}$ as follows:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i \\ &= \beta_0 + \sum_{j=1}^{p-1} \beta_j X_{ij} + \epsilon_i \end{aligned}$$

where
1. Y_i is the value of the response in the i^{th} trial
2. $\beta_0, \beta_1, \ldots, \beta_{p-1}$ are parameters
3. $X_{i,j}$ is the value of the j^{th} predictor in the i^{th} trial, and is a known constant.
4. ϵ_i are independent Normally distributed random variables with mean 0 and variance σ^2.

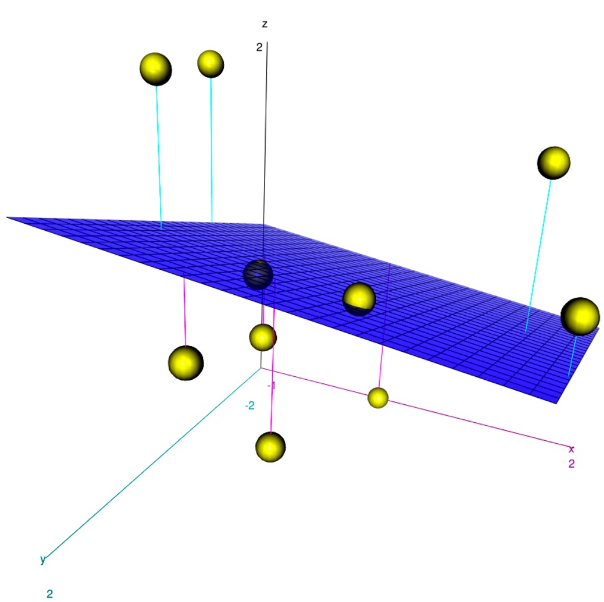
$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,p-1} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,p-1} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \qquad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Definition:
A **case** is a collection of values for the predictor variables and possibly the response variable.[11]

a. Predictors only: $(X_1, X_2, \ldots, X_{p-1})$.
   - The vector $X_i = [X_{i,1}, X_{i,2}, \ldots, X_{i,j} \ldots, X_{i,p-1}]^T$ may be referred to as the i^{th} case or i^{th} observation on the predictors.

b. Predictors and Response: $(X_1, X_2, \ldots, X_{p-1}, Y)$

```
Data:  (1,0,2.5,-11.7)  (3,5,6.4,-21)  (4,2.4,9.6,-51.3)  (7,1.7,5.5,-19)  (1.5,10,9.8,-33.4)  (6.2,9,5.1,4.1)
```



[11] Context should make the difference clear.

Regression Coefficients

β_j represents the change in $E(Y)$ with a one unit change in X_j when all other variables are held constant.[12]

$$\begin{aligned} E[Y \mid X_j + 1] &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_j (X_j + 1) + \cdots + \beta_{p-1} X_{p-1} \\ E[Y \mid X_j] &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_j X_j + \cdots + \beta_{p-1} X_{p-1} \\ E[Y \mid X_j + 1] - E[Y \mid X_j] &= \beta_j \end{aligned}$$

$p - 1$ predictors

$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i$

There are p parameters: $\beta_0, \beta_1, \beta_2, \ldots, \beta_j, \ldots, \beta_{p-1}$. Since β_0 is the intercept, so it doesn't get a predictor. Thus, there are $p - 1$ predictors is in this standard model.

Qualitative Predictors

Definition:
A **qualitative predictor** is a predictor variable that represents categories and not numerical values.

Definition:
An **indicator variable** is a variable that indicates that an outcome A has occurred, or not occurred, by taking a value of one, or zero, respectively.

$$X_A = \begin{cases} 1 & \text{if } A \text{ occurred} \\ 0 & \text{otherwise} \end{cases}$$

A qualitative predictor variable with c categories is represented by $c - 1$ indicator variables.

Example:
A qualitative predictor measures satisfaction. Its possible categories: Low, Medium, High.

$$X_{Medium} = \begin{cases} 1 & \text{Medium} \\ 0 & \text{Otherwise} \end{cases} \qquad X_{High} = \begin{cases} 1 & \text{High} \\ 0 & \text{Otherwise} \end{cases}$$

Definition:
The **reference** is the category that all others are generally compared to. It will be the category that is not represented by any indicator.

[12] Situations arise where this interpretation does not hold. This occurs when predictors are related in some way. In that case, a change in one predictor can't be made without another predictor also changing.

Multiple Linear Regression Techniques can be applied to may situations. Usually, this is done with a suitable substitution of variables.

Polynomial Regression

Suppose the relationship between the response and a single predictor looks to follow a polynomial with degree $k > 1$.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \cdots + \beta_j X_1^j + \cdots + \beta_k X_1^k + \epsilon$$

A substitution of variables turns this into a multiple linear regression problem.

Substituting $Z_1 = X_1, Z_2 = X_1^2, \ldots, Z_j = X_1^j, \ldots Z_k = X_1^k$ will result in a relationship[13] such as

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + \epsilon.$$

Transformed Variables

A: $\log(Y_i) = \beta_0 + \sum_{j=1}^{p-1} \beta_j X_{ij} + \epsilon_i$ B: $Y_i = \dfrac{1}{\beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i}$

A: Substituting $Y^* = \log(Y_i)$ yields $Y^* = \beta_0 + \sum_{j=1}^{p-1} \beta_j X_{ij} + \epsilon_i$
B: Substituting $Y^* = \frac{1}{Y_i}$ yields $Y^* = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i$

Interactions Effects

Definition:
An **interaction** between predictor is present when the change in the response based on a change in one predictor is dependent on the value of another predictor.

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon_i \\ &= \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \epsilon_i \end{aligned}$$

$$\begin{aligned} Z_1 &= X_1 \\ Z_2 &= X_2 \\ Z_3 &= X_1 X_2 \end{aligned}$$

Example:
Suppose $E[Y \mid X_1, X_2] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$ where X_1, X_2 are two indicator variables and β_1, β_2 and β_3 are all nonzero.

$$X_1 = \begin{cases} 1 & \text{Success Type 1} \\ 0 & \text{Otherwise} \end{cases} \qquad X_2 = \begin{cases} 1 & \text{Success Type 2} \\ 0 & \text{Otherwise} \end{cases}$$

$$E[Y \mid X_1, X_2] = \begin{cases} \beta_0 & X_1 = 0, X_2 = 0 \\ \beta_0 + \beta_1 & X_1 = 1, X_2 = 0 \\ \beta_0 + \beta_2 & X_1 = 0, X_2 = 1 \\ \beta_0 + \beta_1 + \beta_2 + \beta_3 & X_1 = 1, X_2 = 1 \end{cases}$$

[13] The drawback here is that the interpretation of a β_j will not be the same. It is not possible to hold all except one Z_j variable constant. They all change together.

Section 2: Estimation of Parameters

Definition:
Given a data set $(X_{1,1}, X_{2,1}, \dots, X_{p-1,1}, Y_1), (X_{1,2}, X_{2,2}, \dots, X_{p-1,2}, Y_2), \dots, (X_{1,n}, X_{2,n}, \dots, X_{p-1,n}, Y_n)$, the **method of least squares** estimates the intercept β_0 and slope $\beta_i, i = 1,2,\dots,p-1$ parameters. It works by minimizing the function

$$\begin{aligned} SS(b_0^*, b_1^*, \dots, b_{p-1}^*) &= \sum_{i=1}^{n}\left[Y_i - \left(b_0^* + b_1^* X_{i,1} + b_2^* X_{i,2} + \cdots + b_{p-1}^* X_{i,p-1}\right)\right]^2 \\ &= (Y - Xb^*)^T(Y - Xb^*) \end{aligned}$$

The values of $b_0^*, b_1^*, \dots, b_{p-1}^*$ that minimize $SS(b_0^*, b_1^*, \dots, b_{p-1}^*)$ will be denoted by $b_0, b_1, \dots, b_{p-1}$.

a. b_0 will be called the **least squares estimate** of the intercept β_0.
b. $b_i, i = 1,2,\dots,p-1$ will be called the **least squares estimate** of the slope β_i.

$$SS(b_0, b_1, \dots, b_{p-1}) \le \min_{\substack{\text{all intercepts } b_0^* \\ \text{all slopes } b_1^* \dots, b_{p-1}^*}} SS(b_0^*, b_1^*, \dots, b_{p-1}^*)$$

$$X = \begin{bmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,p-1} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,p-1} \end{bmatrix} \qquad X^T X b = X^T Y \qquad b = (X^T X)^{-1} X^T Y = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix} \qquad E(b) = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

Point Estimate of the Mean Response

Definition:
The **mean response function** at $X_1, X_2, \dots, X_{p-1}$ is given by $E[Y \mid X_1, X_2, \dots, X_{p-1}] = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1}$.

The **equation of the least squares regression function** is given by $\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_{p-1} X_{p-1}$, where $b_0, b_1, \dots, b_{p-1}$ are unbiased estimates of $\beta_0, \beta_1, \dots, \beta_{p-1}$.

The estimates of the mean response at the observed cases of $(X_{i,1}, X_{i,2}, \dots, X_{i,p-1})$ is given by evaluating the equation of the least squares regression function at $(X_{i,1}, X_{i,2}, \dots, X_{i,p-1})$.

$$\begin{aligned} \hat{Y} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} &= \begin{bmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,p-1} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,p-1} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix} \\ &= Xb \\ &= X(X^T X)^{-1} X^T Y = HY \end{aligned}$$

$$H = X(X^T X)^{-1} X^T \qquad \text{(hat matrix)}$$

When $(X_{i,1}, X_{i,2}, \dots, X_{i,p-1})$ is one of the observed cases of predictors, the **mean response** at this case is also called a **fitted value**.

Residuals

Definition:
The i^{th} **residual** e_i is the difference between the i^{th} observed value and the i^{th} fitted value. That is, $e_i = Y_i - \hat{Y}_i$.

$$e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} Y_1 - \hat{Y}_1 \\ Y_2 - \hat{Y}_2 \\ \vdots \\ Y_n - \hat{Y}_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} - \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = Y - \hat{Y}$$

Notes about residuals:

1. e is a random vector
2. e_i estimate ϵ_i
3. e_i are dependent.
4. $e = (I - H)Y$

$V(e) = \sigma^2(I - H)$

5. $V(e)$ is a variance/covariance matrix.
   a. $V(e)_{i,j} = V(e_i)$ for $i = j$
   b. $V(e)_{i,j} = Cov(e_i, e_j)$ for $i \neq j$

Estimating the Variance of the Error

$SSE \quad = \sum(Y_i - \hat{Y}_i)^2$

$DFE \quad = n - p$

$s^2 \quad = \frac{SSE}{DFE}$

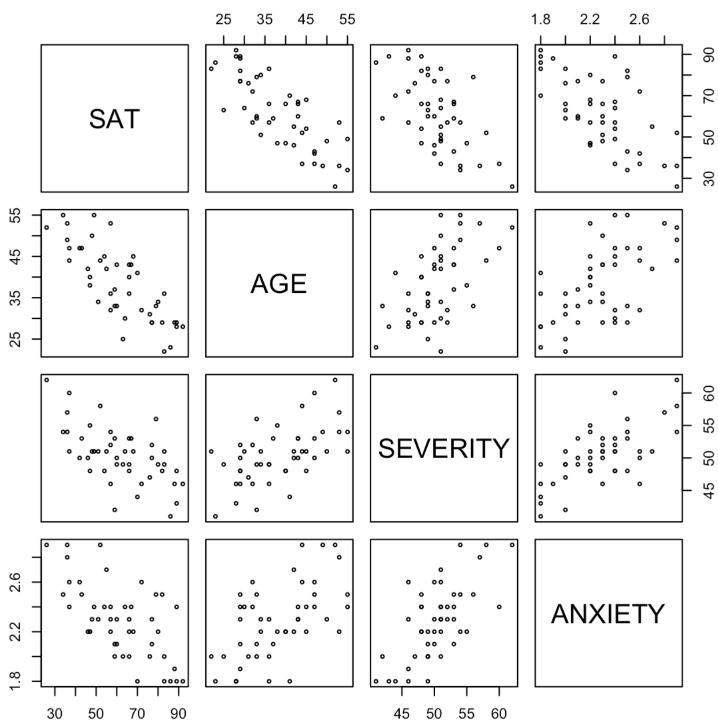
Properties of the Fitted Regression

1. The sum of the residuals is zero.
2. The sum of the squared residuals is a minimum.
3. The sum of the observed values is equal to the sum of the fitted values.
4. The sum of the weighted residuals is zero, when the weight on the i^{th} residual is X_i.
5. The sum of the weighted residuals is zero, when the weight on the i^{th} residual is $\hat{Y}_i$.

Example

A hospital administrator wished to study the relationship between patient satisfaction Y and patient's age X_1, severity of illness X_2, and anxiety level X_3. The administrator randomly selected 46 patients. The data is held in the PatSat dataset. Larger values for Y, X_2 and X_3 indicate higher satisfaction, severity and anxiety. (Kutner)

```
#### Partial Dataframe
head(PatSat)
  SAT AGE SEVERITY ANXIETY
1  48  50       51     2.3
2  57  36       46     2.3
3  66  40       48     2.2
4  70  41       44     1.8
5  89  28       43     1.8
6  36  49       54     2.9

#### Scatterplot Matrix
pairs(PatSat, cex = 0.5)
```




```
#### Variance/Covariance Matrix
var(PatSat)
                SAT         AGE     SEVERITY     ANXIETY
SAT       297.095652 -120.937198 -44.8289855 -3.32579710
AGE      -120.937198   79.532367  21.8483092  1.52077295
SEVERITY  -44.828986   21.848309  18.6067633  0.86579710
ANXIETY    -3.325797    1.520773   0.8657971  0.08960386



#### Correlation Matrix
cor(PatSat)
                SAT        AGE   SEVERITY    ANXIETY
SAT       1.0000000 -0.7867555 -0.6029417 -0.6445910
AGE      -0.7867555  1.0000000  0.5679505  0.5696775
SEVERITY -0.6029417  0.5679505  1.0000000  0.6705287
ANXIETY  -0.6445910  0.5696775  0.6705287  1.0000000
```

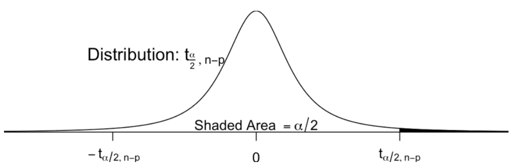
```
#### Fitting Linear Model
satisfaction.fit <- lm(SAT ~ AGE + SEVERITY + ANXIETY, data = PatSat)


#### Components of Fitted Model
names(satisfaction.fit)
 [1] "coefficients"  "residuals"     "effects"       "rank"
 [5] "fitted.values" "assign"        "qr"            "df.residual"
 [9] "xlevels"       "call"          "terms"         "model"

satisfaction.fit$df.residual
[1] 42



#### Summarized Output of Fitted Model
satisfaction.summary <- summary(satisfaction.fit)
satisfaction.summary

Call:
lm(formula = SAT ~ AGE + SEVERITY + ANXIETY, data = PatSat)

Residuals:
     Min       1Q   Median       3Q      Max
-18.3524  -6.4230   0.5196   8.3715  17.1601

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 158.4913    18.1259   8.744 5.26e-11 ***
AGE          -1.1416     0.2148  -5.315 3.81e-06 ***
SEVERITY     -0.4420     0.4920  -0.898   0.3741
ANXIETY     -13.4702     7.0997  -1.897   0.0647 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.06 on 42 degrees of freedom
Multiple R-squared:  0.6822,    Adjusted R-squared:  0.6595
F-statistic: 30.05 on 3 and 42 DF,  p-value: 1.542e-10



#### Components of Summary
names(satisfaction.summary)
 [1] "call"          "terms"         "residuals"     "coefficients"
 [5] "aliased"       "sigma"         "df"            "r.squared"
 [9] "adj.r.squared" "fstatistic"    "cov.unscaled"

#### Residual Standard Error (s)
satisfaction.summary$sigma
[1] 10.05798


#### Fitted Values
satisfaction.fit$fitted.values[1:2]
       1        2
47.88707 66.07965



#### Hand Computations of Fitted Values
X <- as.matrix(cbind(rep(1, nrow(PatSat)), PatSat[2:4]))

b <- solve(t(X) %*% X) %*% t(X) %*% PatSat[[1]]
b
                            [,1]
rep(1, nrow(PatSat)) 158.4912517
AGE                   -1.1416118
SEVERITY              -0.4420043
ANXIETY              -13.4701632


(X %*% b)[1:2]
[1] 47.88707 66.07965
```

Section 3: Inference with Multiple Slope Parameters

1. This inference assumes a normal error regression model.
2. Holding all other variables constant, for a 1 unit change in X_i, there will be a $\beta_k, k > 0$ change in mean response.
3. $V(b) = \sigma^2(X^T X)^{-1}$ is a variance/covariance matrix.
   a. $V(b)_{i,j} = V(b_i)$ for $i = j$
   b. $V(b)_{i,j} = Cov(b_i, b_j)$ for $i \neq j$

Sampling Distribution of $t = \frac{b_k - \beta_k}{s_{b_k}}$

The statistic $t = \frac{b_k - \beta_k}{s_{b_k}}$ has a t distribution with $n - p$ degrees of freedom where s_{b_k} is the estimated standard error of b_k.



Tests for β_k

If all other variables are included, is $\beta_k = 0$ or not?

Hypotheses:
H_0: $\beta_k = 0$
H_1: $\beta_k \neq 0$

Test Statistic:
$t = \frac{b_k}{s_{b_k}}$

Level α Rejection Rule:
Reject H_0 if
a. $t < -t_{\alpha/2,n-p}$ or $t > t_{\alpha/2,n-p}$
b. p - value $\leq \alpha$
   – p - value = 2 × pt(| t |, df = n-p, lower.tail = F)

Interval Estimation of of β_k

Given a random sample of size n from a population, the $100(1-\alpha)\%$ confidence interval for β_k is given by

$$b_k - t_{\alpha/2,n-p}s_{b_k} < \beta_k < b_k + t_{\alpha/2,n-p}s_{b_k}.$$

Joint Inferences (Bonferroni Correction)

To estimate $2 \leq g \leq p$ coefficients simultaneously, use g confidence intervals given by $b_k \pm t_{\alpha/2g,n-p}s_{b_k}$.

- To make $P(\text{all intervals are correct}) \geq 1 - \alpha$, each individual interval should have a confidence level of $100 \times \left(1 - \frac{\alpha}{2g}\right)\%$.
- To make $P(\text{no type I errors}) \leq \alpha$, each individual two-sided test should have a significance level of $\frac{\alpha}{2g}$.

Example:
Suppose you want to compute a confidence interval of β_0, β_2 and β_6. Additionally, you want a **simultaneous** confidence level of 98%. What confidence level would you use for each individual confidence interval.

Interval Estimation of $E(Y_h)$

Assumptions:

1. This inference assumes a normal error regression model.
2. The mean response at predictor vector X_h is given by evaluating the regression equation at X_h, namely

$$E[Y_h|X_h] = \beta_0 + \beta_1 X_{h,1} + \cdots + \beta_{h,p-1} X_{h,p-1}.$$

3. The estimate of the mean response at predictor vector X_h is given by evaluating the equation of the least squares regression function at X_h, namely

$$\hat{Y}_h = b_0 + b_1 X_{h,1} + \cdots + b_{h,p-1} X_{h,p-1}.$$

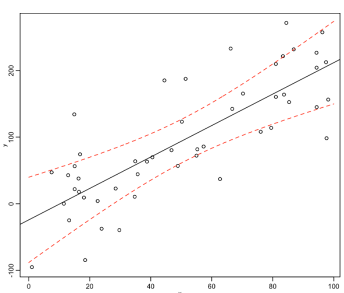
$$X_h = \begin{bmatrix} 1 \\ X_{h,1} \\ \vdots \\ X_{h,p-1} \end{bmatrix} \qquad b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix} \qquad \hat{Y}_h = X_h^T b \qquad \hat{Y}_h \pm t_{\alpha/2,n-p} s_{\hat{Y}_h} \qquad \begin{aligned} V(\hat{Y}_h) &= X_h^T V(b) X_h \\ &= \sigma^2 X_h^T (X^T X)^{-1} X_h \end{aligned}$$

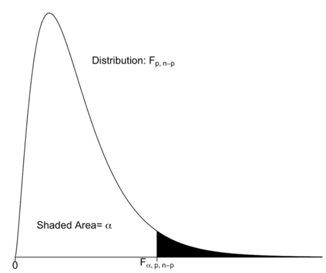
Confidence Band/Region for Regression Function

Compute confidence intervals for **all** X_h simultaneously using

$$\hat{Y}_h \pm W s_{\hat{Y}_h}.$$

a. $W^2 \sim pF_{\alpha,p,n-p}$
b. $F_{\alpha,p,n-p}$ is a critical value take from an F distribution with degrees of freedom $df1 = p, df2 = n - p$.



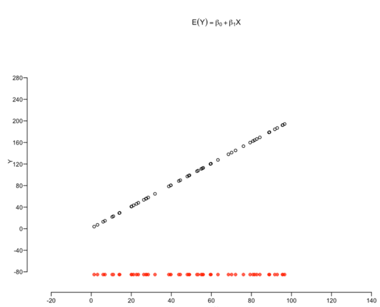


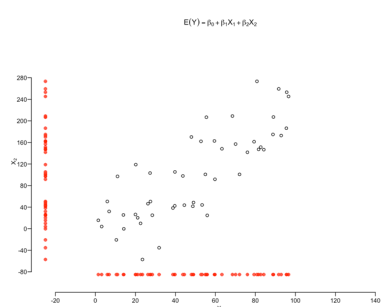
Prediction of $Y_{h(new)}$

$$V(pred) = \sigma^2 + V(\hat{Y}_h) \qquad \hat{Y}_h - t_{\alpha/2,n-p} s_{pred} < Y_{h(new)} < \hat{Y}_h + t_{\alpha/2,n-p} s_{pred}$$

Scope & Extrapolation

Definitions:
The **scope** of a dataset is the range of observed predictors. **Extrapolation** occurs when the the mean response is estimated/predicted outside the scope of the observed data.





Example

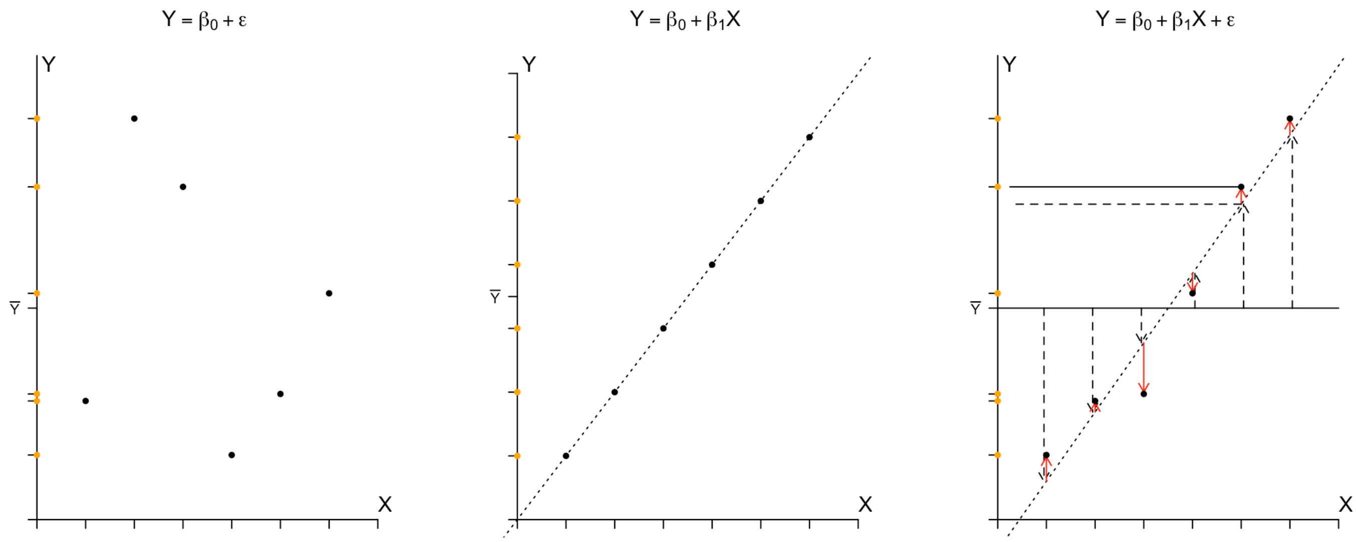
A hospital administrator wished to study the relationship between patient satisfaction Y and patient's age X_1, severity of illness X_2, and anxiety level X_3. The administrator randomly selected 46 patients. The data is held in the PatSat dataset. Larger values for Y, X_2 and X_3 indicate higher satisfaction, severity and anxiety. (K)

```
#### Partial Dataframe
head(PatSat, n = 3)
  SAT AGE SEVERITY ANXIETY
1  48  50       51     2.3
2  57  36       46     2.3
3  66  40       48     2.2

#### Fitting Linear Model
satisfaction.fit <- lm(SAT ~ AGE + SEVERITY + ANXIETY, data = PatSat)

#### Components of Fitted Model
names(satisfaction.fit)
 [1] "coefficients"  "residuals"     "effects"       "rank"
 [5] "fitted.values" "assign"        "qr"            "df.residual"
 [9] "xlevels"       "call"          "terms"         "model"

#### Summarized Output of Fitted Model
satisfaction.summary <- summary(satisfaction.fit)
satisfaction.summary

Call:
lm(formula = SAT ~ AGE + SEVERITY + ANXIETY, data = PatSat)

Residuals:
     Min       1Q   Median       3Q      Max
-18.3524  -6.4230   0.5196   8.3715  17.1601

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 158.4913    18.1259   8.744 5.26e-11 ***
AGE          -1.1416     0.2148  -5.315 3.81e-06 ***
SEVERITY     -0.4420     0.4920  -0.898   0.3741
ANXIETY     -13.4702     7.0997  -1.897   0.0647 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.06 on 42 degrees of freedom
Multiple R-squared:  0.6822,    Adjusted R-squared:  0.6595
F-statistic: 30.05 on 3 and 42 DF,  p-value: 1.542e-10


names(satisfaction.summary)
 [1] "call"          "terms"         "residuals"     "coefficients"
 [5] "aliased"       "sigma"         "df"            "r.squared"
 [9] "adj.r.squared" "fstatistic"    "cov.unscaled"

#### Confidence Intervals
confint(satisfaction.fit, parm = c(3, 4), level = 0.95)
              2.5 %    97.5 %
SEVERITY  -1.434831 0.5508228
ANXIETY  -27.797859 0.8575324



#### Hand Computations
X <- as.matrix(cbind(rep(1, nrow(PatSat)), PatSat[2:4]))
b <- solve(t(X) %*% X) %*% t(X) %*% PatSat[[1]]
b
                              [,1]
rep(1, nrow(PatSat)) 158.4912517
AGE                   -1.1416118
SEVERITY              -0.4420043
ANXIETY              -13.4701632
dof <- satisfaction.fit$df.residual
dof
[1] 42
qt(0.025, df = dof, lower.tail = F)
[1] 2.018082
```

```
#### Computing Fitted Values
satisfaction.fit$fitted.values[1:2]
       1        2 
47.88707 66.07965 

yHat <- X %*% b
yHat[1:2]
[1] 47.88707 66.07965

predict(satisfaction.fit)[1:2]
       1        2 
47.88707 66.07965 


#### Confidence Intervals for Mean Response
nd <- data.frame(AGE = c(37, 38), SEVERITY = c(50, 55), ANXIETY = c(2.8, 2.5))
nd
  AGE SEVERITY ANXIETY
1  37       50     2.8
2  38       55     2.5

predict(satisfaction.fit, newdata = nd, interval = "confidence", level = 0.9)
       fit      lwr      upr
1 56.43494 49.49290 63.37699
2 57.12436 52.92048 61.32824


#### Prediction Intervals
predict(satisfaction.fit, newdata = nd, interval = "prediction", level = 0.99)
       fit      lwr      upr
1 56.43494 27.10185 85.76803
2 57.12436 29.16194 85.08678

nd <- as.matrix(nd[1, ])
x.h <- c(1, nd)
y.h <- t(x.h) %*% b
y.h
         [,1]
[1,] 56.43494

#### Mean Square Error
MSE <- satisfaction.summary$sigma^2

#### Variance/Covariance Matrix
vcov(satisfaction.fit)
            (Intercept)         AGE    SEVERITY    ANXIETY
(Intercept) 328.5478428  0.93283693 -6.87207388 -6.8081417
AGE           0.9328369  0.04613853 -0.03223004 -0.4716488
SEVERITY     -6.8720739 -0.03223004  0.24203030 -1.7916031
ANXIETY      -6.8081417 -0.47164876 -1.79160306 50.4051837


MSE * solve(t(X) %*% X)
                     rep(1, nrow(PatSat))         AGE    SEVERITY    ANXIETY
rep(1, nrow(PatSat))          328.5478428  0.93283693 -6.87207388 -6.8081417
AGE                             0.9328369  0.04613853 -0.03223004 -0.4716488
SEVERITY                       -6.8720739 -0.03223004  0.24203030 -1.7916031
ANXIETY                        -6.8081417 -0.47164876 -1.79160306 50.4051837
sqrt(MSE * t(x.h) %*% solve(t(X) %*% X) %*% x.h)
         [,1]
[1,] 4.127372

qt(0.95, df = dof, lower.tail = T)
[1] 1.681952

#### Confidence Bands
cv.f <- qf(0.05, df1 = 4, df2 = dof, lower.tail = FALSE)
W <- sqrt(4 * cv.f)
W
[1] 3.221343
```

Chapter 5: Explaining Variation in Response from its Mean

Section 1: Partitioning Variation



Definition:
The i^{th} **deviation** of a response value from the overall mean is $Y_i - \bar{Y}$.

$$Y_i - \bar{Y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})$$

Note:

1. If $\beta_1 \neq 0$, then the i^{th} deviation of the response can be partially explained by the regression relationship between the response and the predictor.
   - This difference is given by $\hat{Y}_i - \bar{Y}$.
   - There is a difference between the overall mean of the response **AND** the mean of the response at a particular value of the predictor.
2. If $V(\epsilon > 0)$, then the deviation of the response can't be entirely explained by the regression relationship.
   - The response varies away from the regression line.
   - This variation is from the random error.
   - It is called **unexplained** variation. Unexplained because the regression relationship can't explain it.

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

$$SSTotal = SSError + SSRegression$$

Total Variation = Unexplained Variation + Explained Variation

Section 2: ANOVA TABLE - Simple Linear Regression - Overall F Test

The partitioning of the total variation in the response can be recorded in an **AN**alysis **O**f **V**ariance table. Otherwise known as an ANOVA table.

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F	Expected(MS)
Regr.	1	$SSReg = \sum\left(\hat{Y}_i - \bar{Y}\right)^2$	$MSReg = \dfrac{SSReg}{1}$	$F = \dfrac{MSReg}{MSE}$	$\sigma^2 + \beta_1^2 \sum (X_i - \bar{X})^2$
Error	$n-2$	$SSE = \sum\left(Y_i - \hat{Y}_i\right)^2$	$MSE = \dfrac{SSE}{n-2}$		σ^2
Total	$n-1$	$SST = \sum (Y_i - \bar{Y})^2$			

Degrees of Freedom - A Partition of Information

- $n-1$ independent pieces of information to be partitioned.
- $n-2$ of these are used in the Error row.
- 1 is left over for the regression row.

Expected Mean Squares

$E(\text{MSReg}) = \sigma^2 + \beta_1^2 \sum (X_i - \bar{X})^2$ $\qquad$ $E(\text{MSE}) = \sigma^2$

Motivation for a test:

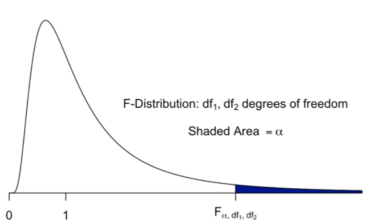
Consider the ration of $E(\text{MSReg})$ to $E(\text{MSE})$.[14] Namely,

$$\frac{E(\text{MSReg})}{E(\text{MSE})} = \begin{cases} \dfrac{\sigma^2 + \beta_1^2 \sum (X_i - \bar{X})^2}{\sigma^2} & \text{if } \beta_1 \neq 0 \\ \dfrac{\sigma^2}{\sigma^2} & \text{if } \beta_1 = 0 \end{cases}.$$

Section 3: F-Test $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$

Sampling Distribution of F^*

For the normal error regression model, if $\beta_1 = 0$ then $F^* = \frac{MSReg}{MSE}$ has an F distribution with 1 numerator degree of freedom and $n-2$ denominator degrees of freedom.

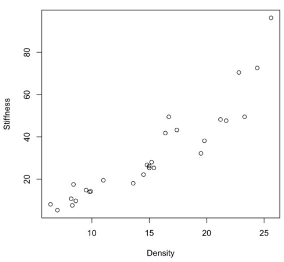


[14] Do not misinterpret this to thing that $E\left(\frac{\text{MSReg}}{\text{MSE}}\right) = \frac{E(\text{MSReg})}{E(\text{MSE})}$

Example:

A materials engineer at a furniture manufacturing site wants to assess the stiffness of the particle board that the manufacturer uses. The engineer collects stiffness data from particle board pieces that have various densities at different temperatures.(Minitab,Mont)

```
#### Partial Dataframe
head(PartBoard, n = 3)[1:2]
  Density Stiffness
1     9.5    14.814
2     8.4    17.502
3     9.8    14.007

#### Plot Data
plot(Stiffness ~ Density, data = PartBoard, xlab = "Density", ylab = "Stiffness")
```




```
#### Fitting the Model
particle.fit <- lm(Stiffness ~ Density, data = PartBoard)

#### Summarizing the Fitted Model
summary(particle.fit)

Call:
lm(formula = Stiffness ~ Density, data = PartBoard)

Residuals:
     Min       1Q   Median       3Q      Max 
-15.3565  -6.6346   0.7408   3.7357  27.1077 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -21.5935     4.8647  -4.439 0.000148 ***
Density       3.5465     0.3041  11.662 7.87e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.037 on 26 degrees of freedom
Multiple R-squared:  0.8395,    Adjusted R-squared:  0.8333 
F-statistic:   136 on 1 and 26 DF,  p-value: 7.873e-12

#### ANOVA TABLE
particle.anova <- anova(particle.fit)
particle.anova
Analysis of Variance Table

Response: Stiffness
          Df  Sum Sq Mean Sq F value    Pr(>F)    
Density    1 11106.5 11106.5  136.01 7.873e-12 ***
Residuals 26  2123.2    81.7                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#### Components of ANOVA
names(particle.anova)
[1] "Df"      "Sum Sq"  "Mean Sq" "F value" "Pr(>F)" 
dof <- particle.anova$Df
dof
[1]  1 26

#### Critical Value
qf(0.05, df1 = dof[1], df2 = dof[2], lower.tail = F)
[1] 4.225201
```

Section 4: ANOVA TABLE - Multiple Linear Regression - Overall F Test

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F	Expected(MS)
Regr.	$p-1$	$SSReg = \sum(\hat{Y}_i - \bar{Y})^2$	$MSReg = \frac{SSReg}{p-1}$	$F = \frac{MSReg}{MSE}$	$\sigma^2 + g(\beta_0, \beta_1, \dots, \beta_{p-1})$
Error	$n-p$	$SSE = \sum(Y_i - \hat{Y}_i)^2$	$MSE = \frac{SSE}{n-p}$		σ^2
Total	$n-1$	$SST = \sum(Y_i - \bar{Y})^2$			$g(\beta_0, \beta_1, \dots, \beta_{p-1}) \geq 0$

<u>Degrees of Freedom - A Partition of Information</u>

- $n-1$ independent pieces of information to be partitioned.
- $n-p$ of these are used in the Error row.
- $p-1$ is left over for the regression row.

<u>Expected Mean Squares</u>

$$E(\text{MSReg}) = \sigma^2 + \sum[\beta_1(X_{i1} - \bar{X}_1) + \beta_2(X_{i2} - \bar{X}_2) + \cdots]^2 = \sigma^2 \qquad\qquad E(\text{MSE}) = \sigma^2$$

Motivation for a test:

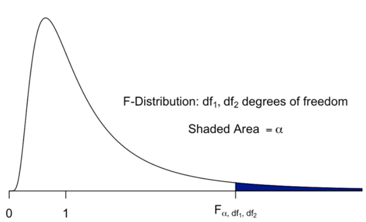
Consider the ration of $E(\text{MSReg})$ to $E(\text{MSE})$.[15] Namely,

$$\frac{E(\text{MSReg})}{E(\text{MSE})} = \begin{cases} \dfrac{\sigma^2 + g(\beta_0, \beta_1, \dots, \beta_{p-1})}{\sigma^2} & \text{if } \beta_j \neq 0 \text{ for some } j \text{ in } 1,2,\dots,p-1 \\[2ex] \dfrac{\sigma^2}{\sigma^2} & \text{otherwise} \end{cases}.$$

Section 5: F-Test $H_0: \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$ vs. H_1: At least one $\beta_i \neq 0$

<u>Sampling Distribution of F^*</u>

For the normal error regression model, if $\beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$ then $F^* = \frac{MSReg}{MSE}$ has an F distribution with $p-1$ numerator degree of freedom and $n-p$ denominator degrees of freedom.



[15] Do not misinterpret this to thing that $E\left(\frac{\text{MSReg}}{\text{MSE}}\right) = \frac{E(\text{MSReg})}{E(\text{MSE})}$

Example:

A hospital administrator wished to study the relationship between patient satisfaction Y and patient's age X_1, severity of illness X_2, and anxiety level X_3. The administrator randomly selected 46 patients. The data is held in the PatSat dataset. Larger values for Y, X_2 and X_3 indicate higher satisfaction, severity and anxiety.(K)

```
#### Fitting the Model
satisfaction.fit <- lm(SAT ~ AGE + SEVERITY + ANXIETY, data = PatSat)

#### Components of the Fitted Model
names(satisfaction.fit)
 [1] "coefficients"  "residuals"     "effects"       "rank"
 [5] "fitted.values" "assign"        "qr"            "df.residual"
 [9] "xlevels"       "call"          "terms"         "model"

#### Summarizing the Fitted Model
satisfaction.summary <- summary(satisfaction.fit)
satisfaction.summary

Call:
lm(formula = SAT ~ AGE + SEVERITY + ANXIETY, data = PatSat)

Residuals:
     Min       1Q   Median       3Q      Max
-18.3524  -6.4230   0.5196   8.3715  17.1601

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 158.4913    18.1259   8.744 5.26e-11 ***
AGE          -1.1416     0.2148  -5.315 3.81e-06 ***
SEVERITY     -0.4420     0.4920  -0.898   0.3741
ANXIETY     -13.4702     7.0997  -1.897   0.0647 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.06 on 42 degrees of freedom
Multiple R-squared:  0.6822,    Adjusted R-squared:  0.6595
F-statistic: 30.05 on 3 and 42 DF,  p-value: 1.542e-10

names(satisfaction.summary)
 [1] "call"          "terms"         "residuals"     "coefficients"
 [5] "aliased"       "sigma"         "df"            "r.squared"
 [9] "adj.r.squared" "fstatistic"    "cov.unscaled"

#### SEQUENTIAL ANOVA TABLE
satisfaction.anova <- anova(satisfaction.fit)
satisfaction.anova
Analysis of Variance Table

Response: SAT
          Df Sum Sq Mean Sq F value    Pr(>F)
AGE        1 8275.4  8275.4 81.8026 2.059e-11 ***
SEVERITY   1  480.9   480.9  4.7539   0.03489 *
ANXIETY    1  364.2   364.2  3.5997   0.06468 .
Residuals 42 4248.8   101.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#### NOT THE ANOVA TABLE WE WANT DOES NOT INCLUDE OVERALL F TEST
```

Section 6: General Linear Test

This approach to testing allows you to include some predictors and test whether others should be included in the model.

Definition:
A **full** or **unrestricted** model is a model that includes all predictors of possible interest.

Definition:
The **reduced** or **restricted** model is a version of the full model with pre-specified parameters set to zero.[16]

$$\overbrace{\underbrace{Y \sim X_1 + X_2 + \cdots + X_r}_{\text{}}\!\!\!\!\!\!\!\overbrace{}^{} }^{}$$

$$\overbrace{\overbrace{Y \sim X_1 + X_2 + \cdots + X_r}^{\text{Reduced Model}} + X_{r+1} + X_{r+2} + \cdots + X_{r+t}}^{\text{Full Model}}$$

$$\begin{aligned} \text{Full Model:} \quad & Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_r X_{i,r} + \beta_{r+1} X_{i,r+1} + \cdots + \beta_{r+t} X_{i,r+t} + \epsilon_i \\ \text{Reduced Model:} \quad & Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_r X_{i,r} + \epsilon_i \end{aligned}$$

Hypotheses:

H_0: $\beta_{r+1} = \beta_{r+1} = \cdots = \beta_{r+t} = 0$ (Reduced Model)
H_1: at least one of $\beta_{r+1}, \ldots, \beta_{r+t} \neq 0$ (Full Model)

$$\overbrace{\overbrace{Y \sim X_1 + X_2 + \cdots + X_r}^{\substack{\text{Null Hypothesis} \\ \beta_{r+1} = \cdots = \beta_{r+t} = 0}} + X_{r+1} + X_{r+2} + \cdots + X_{r+t}}^{\substack{\text{Alternative Hypothesis} \\ \text{At least one of } \beta_{r+1}, \ldots, \beta_{r+t} \neq 0}}$$

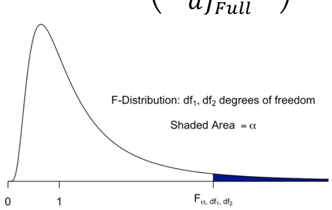
<u>**Least Squares**</u>

$$\min \mathrm{SS}(b_0^*, b_1^*, \cdots, b_r^*, b_{r+1}, \cdots, b_{r+t}) = \quad SSE(Full) \leq SSE(Reduced) \quad = \min \mathrm{SS}(b_0^*, b_1^*, \cdots, b_r^*, 0, \cdots, 0)$$

Test Statistic:

Assuming the assumptions for the normal error multiple linear regression model are satisfied, the test statistic F for the General Linear Test will have an F distribution with $df_1 = DF_{Reduce} - DF_{FULL}$ and $df_2 = DF_{FULL}$. The test statistic is given by

$$F = \frac{\left(\dfrac{SSE(Reduced) - SSE(Full)}{df_{Reduced} - df_{Full}}\right)}{\left(\dfrac{SSE(Full)}{df_{Full}}\right)}$$



- $SSE(Full) =$ Amount of Variation NOT explained by FULL Model
- $SSE(Reduced) - SSE(Full) = SSREG(Full) - SSREG(Reduced) =$
  Amount of variation explained by full model but NOT by the reduced model

[16] The most reduced model you can have does not use any predictor variables. The model is built only on the intercept. It is commonly written as $Y \sim 1$.

Simple Linear Regression Example:

A materials engineer at a furniture manufacturing site wants to assess the stiffness of the particle board that the manufacturer uses. The engineer collects stiffness data from particle board pieces that have various densities at different temperatures.(Minitab,Mont)

```
#### Partial Dataframe
head(PartBoard)
  Density Stiffness     Temp
1     9.5    14.814 70.61056
2     8.4    17.502 73.34893
3     9.8    14.007 66.15377
4    11.0    19.443 70.05781
5     8.3     7.573 69.33919
6     9.9    14.191 69.12882

#### Fitting Full Model
particle.fit <- lm(Stiffness ~ Density, data = PartBoard)

#### Fitting Reduced Model
particle.intercept.only <- lm(Stiffness ~ 1, data = PartBoard)
```

```
#### Summarizing Full Model
summary(particle.fit)

Call:
lm(formula = Stiffness ~ Density, data = PartBoard)

Residuals:
     Min       1Q   Median       3Q      Max
-15.3565  -6.6346   0.7408   3.7357  27.1077

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -21.5935     4.8647  -4.439 0.000148 ***
Density       3.5465     0.3041  11.662 7.87e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.037 on 26 degrees of freedom
Multiple R-squared:  0.8395,    Adjusted R-squared:  0.8333
F-statistic:   136 on 1 and 26 DF,  p-value: 7.873e-12
```

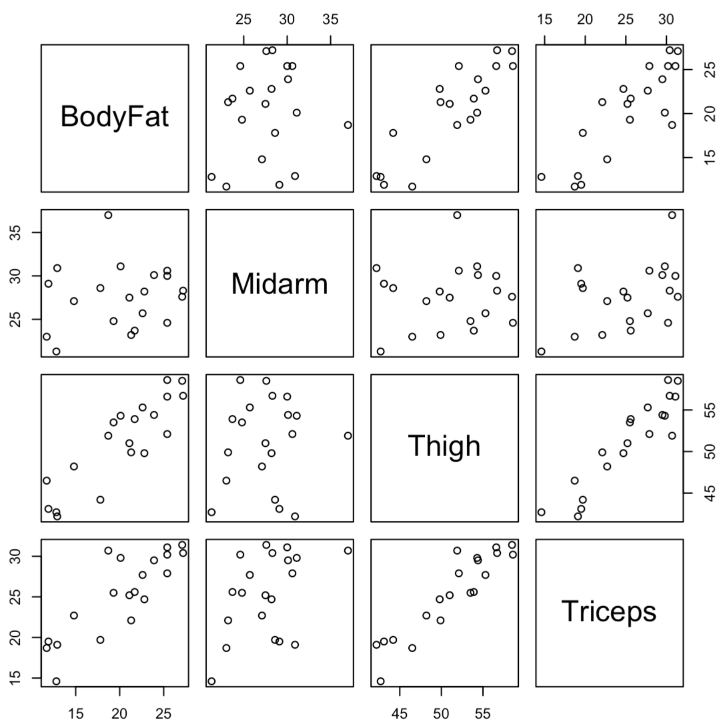
```
#### ANOVA TABLE - OVERALL F-TEST:
particle.anova <- anova(particle.fit)
particle.anova
Analysis of Variance Table

Response: Stiffness
          Df  Sum Sq Mean Sq F value    Pr(>F)
Density    1 11106.5 11106.5  136.01 7.873e-12 ***
Residuals 26  2123.2    81.7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#### GENERAL LINEAR TEST
anova(particle.intercept.only, particle.fit)
Analysis of Variance Table

Model 1: Stiffness ~ 1
Model 2: Stiffness ~ Density
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     27 13229.7
2     26  2123.2  1     11106 136.01 7.873e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example:

A researcher is exploring the relationship between body fat (Y) and three predictors: Tricep Skinfold Thickness, Thigh Circumference, and Midarm Circumference.(Kutner)

```
#### Partial Dataframe
head(BodyFat)
  BodyFat Midarm Thigh Triceps
1    11.9   29.1  43.1    19.5
2    22.8   28.2  49.8    24.7
3    18.7   37.0  51.9    30.7
4    20.1   31.1  54.3    29.8
5    12.9   30.9  42.2    19.1
6    21.7   23.7  53.9    25.6

#### Scatterplot Matrix
pairs(BodyFat)
```




```
#### Correlation Matrix
cor(BodyFat)
          BodyFat    Midarm     Thigh   Triceps
BodyFat 1.0000000 0.1424440 0.8780896 0.8432654
Midarm  0.1424440 1.0000000 0.0846675 0.4577772
Thigh   0.8780896 0.0846675 1.0000000 0.9238425
Triceps 0.8432654 0.4577772 0.9238425 1.0000000
```

Overall F Test - Multiple Linear Regression

Recall:
The overall F test is used for the following hypotheses:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$ $\quad Y \sim 1$ (Reduced Model)
H_1: at least one of $\beta_1, \ldots, \beta_{p-1} \neq 0$ $\quad Y \sim X_1 + X_2 + \cdots + X_{p-1}$ (Full Model)

```
#### Fitting Full Model
body.fit.all <- lm(BodyFat ~ Midarm + Thigh + Triceps, data = BodyFat)

#### Summarizing Full Model
summary(body.fit.all)

Call:
lm(formula = BodyFat ~ Midarm + Thigh + Triceps, data = BodyFat)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.7263 -1.6111  0.3923  1.4656  4.1277 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
Midarm        -2.186      1.595  -1.370    0.190
Thigh         -2.857      2.582  -1.106    0.285
Triceps        4.334      3.016   1.437    0.170

Residual standard error: 2.48 on 16 degrees of freedom
Multiple R-squared:  0.8014,    Adjusted R-squared:  0.7641 
F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

```
#### SEQUENTIAL ANOVA TABLE FULL MODEL
body.all.anova <- anova(body.fit.all)
body.all.anova
Analysis of Variance Table

Response: BodyFat
          Df Sum Sq Mean Sq F value    Pr(>F)    
Midarm     1  10.05   10.05  1.6343    0.2193    
Thigh      1 374.23  374.23 60.8471 7.684e-07 ***
Triceps    1  12.70   12.70  2.0657    0.1699    
Residuals 16  98.40    6.15                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Source	DF	SS	MS	F	P-value
Model	3	397	132	22	$7.343e - 06$
Error	16	98	6.2		
Total	19	495			

```
#### Fitting Full Model & Intercept Only Model
body.fit.all <- lm(BodyFat ~ Midarm + Thigh + Triceps, data = BodyFat)
body.intercept.only <- lm(BodyFat ~ 1, data = BodyFat)

#### Summarizing Full Model
anova(body.intercept.only, body.fit.all)
Analysis of Variance Table

Model 1: BodyFat ~ 1
Model 2: BodyFat ~ Midarm + Thigh + Triceps
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)    
1     19 495.39                                  
2     16  98.40  3    396.98 21.516 7.343e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

General Linear Tests - Multiple Linear Regression

Example:
A researcher is exploring the relationship between body fat (Y) and three predictors: Tricep Skinfold Thickness, Thigh Circumference, and Midarm Circumference.(Kutner)

```
#### Fitting Full Model & Midarm Only Model
body.fit.all <- lm(BodyFat ~ Midarm + Thigh + Triceps, data = BodyFat)
body.midarm.only <- lm(BodyFat ~ Midarm, data = BodyFat)

#### Summarizing Full Model
summary(body.fit.all)

Call:
lm(formula = BodyFat ~ Midarm + Thigh + Triceps, data = BodyFat)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.7263 -1.6111  0.3923  1.4656  4.1277 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
Midarm        -2.186      1.595  -1.370    0.190
Thigh         -2.857      2.582  -1.106    0.285
Triceps        4.334      3.016   1.437    0.170

Residual standard error: 2.48 on 16 degrees of freedom
Multiple R-squared:  0.8014,    Adjusted R-squared:  0.7641 
F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

Midarm:	H_0:	BodyFat $\sim$ Thigh + Tricep	H_a:BodyFat $\sim$ Thigh + Tricep + Midarm
Thigh:	H_0:	BodyFat $\sim$ Tricep + Midarm	H_a:BodyFat $\sim$ Tricep + Midarm + Thigh
Triceps:	H_0:	BodyFat $\sim$ Thigh + Midarm	H_a:BodyFat $\sim$ Thigh + Midarm + Tricep
F:	H_0:	BodyFat $\sim$ 1	H_a:BodyFat $\sim$ Thigh + Tricep + Midarm

```
#### SEQUENTIAL ANOVA TABLE FULL MODEL
anova(body.fit.all)
Analysis of Variance Table

Response: BodyFat
          Df Sum Sq Mean Sq F value    Pr(>F)    
Midarm     1  10.05   10.05  1.6343    0.2193    
Thigh      1 374.23  374.23 60.8471 7.684e-07 ***
Triceps    1  12.70   12.70  2.0657    0.1699    
Residuals 16  98.40    6.15                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Midarm:	H_0:	BodyFat $\sim$ 1	H_a:BodyFat $\sim$ Midarm
Thigh:	H_0:	BodyFat $\sim$ Midarm	H_a:BodyFat $\sim$ Midarm + Thigh
Triceps:	H_0:	BodyFat $\sim$ Midarm + Thigh + Midarm	H_a:BodyFat $\sim$ Midarm + Thigh + Tricep

```
#### Summarizing Full Model - Midarm only vs All
anova(body.midarm.only, body.fit.all)
Analysis of Variance Table

Model 1: BodyFat ~ Midarm
Model 2: BodyFat ~ Midarm + Thigh + Triceps
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)    
1     18 485.34                                  
2     16  98.40  2    386.93 31.456 2.856e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

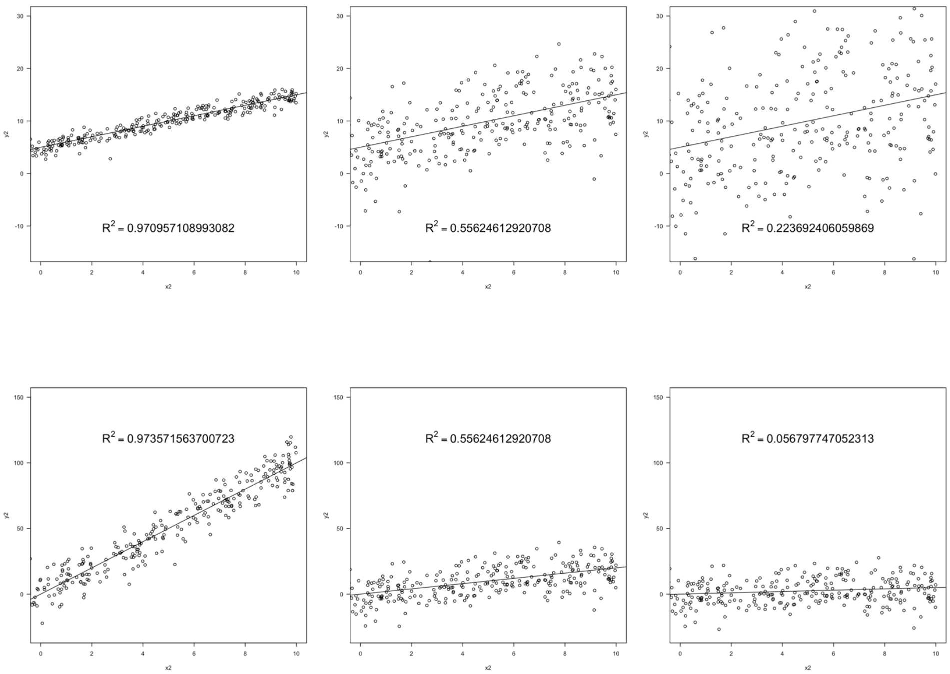
```
#### Fitting Full Model - Order of Predictors Reversed
body.fit.rev <- lm(BodyFat ~ Triceps + Thigh + Midarm, data = BodyFat)

#### ANOVA Table - Order of Predictors Reversed
anova(body.fit.rev)
Analysis of Variance Table

Response: BodyFat
          Df Sum Sq Mean Sq F value    Pr(>F)    
Triceps    1 352.27  352.27 57.2768 1.131e-06 ***
Thigh      1  33.17   33.17  5.3931   0.03373 *  
Midarm     1  11.55   11.55  1.8773   0.18956    
Residuals 16  98.40    6.15                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#### ANOVA Table - Original Order of Predictors
anova(body.fit.all)
Analysis of Variance Table

Response: BodyFat
          Df Sum Sq Mean Sq F value    Pr(>F)    
Midarm     1  10.05   10.05  1.6343    0.2193    
Thigh      1 374.23  374.23 60.8471 7.684e-07 ***
Triceps    1  12.70   12.70  2.0657    0.1699    
Residuals 16  98.40    6.15                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#### Summarizing Full Model - Order of Predictors Reversed
summary(body.fit.rev)

Call:
lm(formula = BodyFat ~ Triceps + Thigh + Midarm, data = BodyFat)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.7263 -1.6111  0.3923  1.4656  4.1277 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
Triceps        4.334      3.016   1.437    0.170
Thigh         -2.857      2.582  -1.106    0.285
Midarm        -2.186      1.595  -1.370    0.190

Residual standard error: 2.48 on 16 degrees of freedom
Multiple R-squared:  0.8014,	Adjusted R-squared:  0.7641 
F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06

#### Summarizing Full Model - Original Order of Predictors
summary(body.fit.all)

Call:
lm(formula = BodyFat ~ Midarm + Thigh + Triceps, data = BodyFat)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.7263 -1.6111  0.3923  1.4656  4.1277 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
Midarm        -2.186      1.595  -1.370    0.190
Thigh         -2.857      2.582  -1.106    0.285
Triceps        4.334      3.016   1.437    0.170

Residual standard error: 2.48 on 16 degrees of freedom
Multiple R-squared:  0.8014,	Adjusted R-squared:  0.7641 
F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

Section 7: Coefficient Of Determination

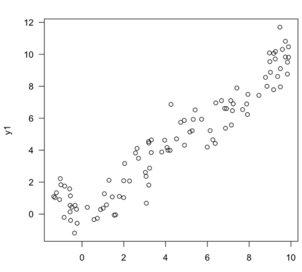
Definition:
The **coefficient of determination** R^2 is the proportion of deviation/variance explained by regression.

$$R^2 = \frac{SSReg}{SST} = 1 - \frac{SSE}{SST}$$



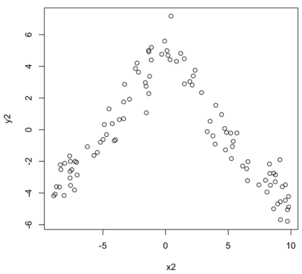
<u>A high R^2 value doesn't necessarily indicate that regression line is a good fit.</u>

```
par(las = 1)
set.seed(6)
n <- 100
x1 <- runif(n, min = -2, max = 10)
y1 <- abs(x1) + rnorm(n, sd = 1)
summary(lm(y1 ~ x1))$r.squared
[1] 0.8930766
plot(x1, y1)
```



<u>An R^2 value near zero doesn't necessarily indicate that the variables are not related.</u>

```
set.seed(6)
x2 <- runif(n, min = -10, max = 10)
y2 <- -abs(x2) + rnorm(n, sd = 1) + 5
summary(lm(y2 ~ x2))$r.squared
[1] 0.03905343
plot(x2, y2)
```



<u>As the number of explanatory variables increases, R^2 increases, even if they are meaningless variables.</u>

Example:

A hospital administrator wished to study the relationship between patient satisfaction Y and patient's age X_1, severity of illness X_2, and anxiety level X_3. The administrator randomly selected 46 patients. The data is held in the PatSat dataset. Larger values for Y, X_2 and X_3 indicate higher satisfaction, severity and anxiety. (K)

```
#### Fitting SAT ~ AGE
satisfaction.fit1 <- lm(SAT ~ AGE, data = PatSat)

summary(satisfaction.fit1)$r.squared  #### R-Squared
[1] 0.6189843

#### Fitting SAT ~ AGE + SEVERITY
satisfaction.fit2 <- lm(SAT ~ AGE + SEVERITY, data = PatSat)
summary(satisfaction.fit2)$r.squared  #### R-Squared
[1] 0.6549559

#### Fitting SAT ~ AGE + SEVERITY + ANXIETY
satisfaction.fit.all <- lm(SAT ~ AGE + SEVERITY + ANXIETY, data = PatSat)
summary(satisfaction.fit.all)$r.squared  #### R-Squared
[1] 0.6821943

#### Generating Junk Variable
junk.Variable <- rnorm(nrow(PatSat), sd = 10)

#### Fitting SAT ~ AGE + SEVERITY + ANXIETY +junk.Variable
satisfaction.fit.junk <- lm(SAT ~ AGE + SEVERITY + ANXIETY + junk.Variable, data = PatSat)
summary(satisfaction.fit.junk)

Call:
lm(formula = SAT ~ AGE + SEVERITY + ANXIETY + junk.Variable,
    data = PatSat)

Residuals:
     Min       1Q   Median       3Q      Max
-16.4690  -5.9766   0.3852   9.1106  15.3285

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   158.0685    18.0311   8.766 6.03e-11 ***
AGE            -1.1949     0.2181  -5.478 2.39e-06 ***
SEVERITY       -0.3912     0.4911  -0.797   0.4303
ANXIETY       -13.4459     7.0612  -1.904   0.0639 .
junk.Variable   0.1917     0.1587   1.208   0.2341
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10 on 41 degrees of freedom
Multiple R-squared:  0.6931,    Adjusted R-squared:  0.6632
F-statistic: 23.15 on 4 and 41 DF,  p-value: 4.626e-10
```

Section 8: Multiple Inference

Which Test? Which Variables?

1. Overal F Test $H_0: \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$

$$X_1, X_2, \cdots, X_{p-1}$$

2. Tests on Subsets: $H_0: \beta_{r+1} = \beta_{r+2} = \cdots = \beta_{r+t}$

$$X_1, X_2, \cdots, X_r \qquad\qquad X_{r+1}, X_{r+2}, \cdots, X_{r+t}$$

Bonferonni Inequality

Suppose that $A_1, A_2, \ldots, A_k$ are events. Using the Addition Rule, we have the following inequalities:

$$P(A_1 \text{ or } A_2) = P(A_1) + P(A_2) - P(A_1 \text{ and } A_2) \leq P(A_1) + P(A_2)$$

$$P(A_1 \text{ or } A_2 \text{ or } A_3) = P([A_1 \text{ or } A_2] \text{ or } A_3)$$

$$P(A_1 \text{ or } A_2 \text{ or } \cdots \text{ or } A_k) \leq P(A_1) + P(A_2) + \cdots + P(A_k)$$

Multiple Tests

Definition:
Suppose you are performing k hypothesis tests. Define the event $A_i =$ Type I error on the i^{th} test, $i = 1,2, \ldots, k$. Further, let $\alpha_i =$ significance level of the i^{th} test. Then **overall/experiment-wise error rate** on k simultaneous hypothesis tests is $\alpha_e = P(\text{at least one Type I error in } k \text{ tests})$.

Upper Bound:
An upper bound can be found on the experiment-wise error rate, α_e, using the inequalities given above.

$$\begin{aligned} P(A_1 \text{ or } A_2 \text{ or } \cdots \text{ or } A_k) &\leq P(A_1) + P(A_2) + \cdots + P(A_k) \\ &= \alpha_1 + \alpha_2 + \cdots + \alpha_k \end{aligned}$$

To control the experiment-wise error rate:

a. Select a value α_{max} that represents the maximum experiment-wise error rate that you will allow.
b. Select $\alpha_1, \alpha_2, \ldots, \alpha_k$ for each test so that $\alpha_1 + \alpha_2 + \cdots + \alpha_k \leq \alpha_{max}$.

Then the actual experiment-wise error rate $\alpha_e \leq \alpha_1 + \alpha_2 + \cdots + \alpha_k \leq \alpha_{max}$.

Example:
Suppose $\alpha_{max} = 0.05$ and $k = 10$. Assuming you will use equal Type I error rates on each test. What should be the Type I error rates on each test?

Multiple Confidence Intervals

Definition:
Suppose you are to construct k confidence intervals. Define the event $A_i = i^{th}$ confidence interval misses its target, $i = 1,2, \dots, k$. Further, let $\alpha_i =$ significance level of corresponding to the i^{th} confidence interval. Then the **simultaneous confidence level** $CL_{all}\%$ on the k confidence intervals is $100 \times (1 - \alpha_e)\%$ where $(1 - \alpha_e) = P(\text{all } k \text{ confidence intervals are correct})$.

Lower Bound:
An lower bound can be found on the simultaneous confidence level, $CL_{all}\%$, using the inequalities given on the previous page.

$$\begin{aligned} CL_{all}/100 &= P(\text{all } k \text{ confidence intervals are correct}) \\ &= 1 - P(\text{at least one CI misses its target}) \\ &= 1 - P(A_1 \text{ or } A_2 \text{ or } \cdots \text{ or } A_k) \\ &\geq 1 - \big(P(A_1) + P(A_2) + \cdots + P(A_k)\big) \\ &= 1 - (\alpha_1 + \alpha_2 + \cdots + \alpha_k) \end{aligned}$$

To control the simultaneous confidence level:

a. select a value $CL_{min}\%$ that represents the minimum simultaneous confidence level that you will allow.
b. Determine the corresponding value α_{max} where $1 - \alpha_{max} = CL_{max}/100$.
c. Set the confidence level for each individual confidence interval to $100(1 - \alpha_1)\%, 100(1 - \alpha_2)\%, \dots, 100(1 - \alpha_k)\%$ so that $\alpha_1 + \alpha_2 + \cdots + \alpha_k \leq \alpha_{max}$.

Then the actual simultaneous confidence level is

$$CL_{all} \geq 100 \times \big(1 - (\alpha_1 + \alpha_2 + \cdots + \alpha_k)\big) \geq CL_{min}$$

Example:
Suppose $CL_{min} = 95\%$ and $k = 10$. Assuming you will use equal individual confidence levels for each confidence intervals. What should be the confidence level be for each confidence interval?