STA302H5
Regression Analysis

Al Nosedal

Fall 2018

"Simple can be harder than complex: You have to work hard to get your thinking clean to make it simple. But it's worth it in the end because once you get there, you can move mountains."

Steve Jobs.

PART 1: LINEAR MODELS AND ESTIMATION BY LEAST SQUARES

The kind of problem we want to solve.

Companies like Best Buy depend on large computer systems to manage and store millions of customer transactions, inventory records, payroll information, and other types of company data. So the company must have enough computing power to be able to process and retrieve that data quickly and efficiently. For a growing company, assuring enough computing capacity and speed is critical.

Each year Best Buy purchases mainframe computing, measured in MIPS (Millions of Instructions Per Second). For planning and budgeting purposes they also want to forecast the number of MIPS needed the following year. Figure 1 shows monthly mainframe computing use and the number of stores Best Buy had between August 1996 and July 2000.

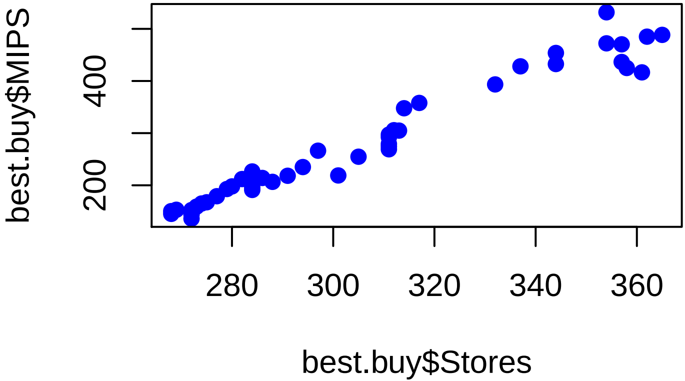
Scatter plot (reading txt file)

```r
#Step 1. Entering data;
# url of Best Buy;
best_url = "http://www.math.unm.edu/~alvaro/Best-Buy.txt"
# importing data into R;
best.buy= read.table(best_url, header = TRUE);
names(best.buy)

## [1] "Month"  "MIPS"   "Stores"
```

Scatterplot

```
# Step 2. Making scatterplot;

plot(best.buy$Stores,best.buy$MIPS,col="blue",pch=19);
```

Figure 1



From the scatterplot, you can see that the relationship between computer capacity and number of stores is positive, linear, and strong. But the strength of the relationship is only part of the picture. In 2000, management might have wanted to predict how many MIPS they'd need to support the 419 stores they projected they'd have by the end of fiscal 2001. That's is a reasonable business question, but we can't read the answer directly from the scatterplot. We need a model for the trend.

Homework

This course uses R. R is an open-source computing package which has seen a huge growth in popularity in the last few years. R can be downloaded from https://cran.r-project.org

Please, download R and bring your laptop next time.

Functional Relation between Two Variables

A functional relation between two variables is expressed by a mathematical formula. If X denotes the *independent variable* and Y the *dependent variable*, a functional relation is of the form:

$$Y = f(X)$$

Given a particular value of X, the function f indicates the corresponding value of Y.

Statistical Relation between Two Variables

A statistical relation, unlike a functional relation, is not a perfect one. In general, the observations for a statistical relation do not fall directly on the curve of relationship.

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

1. A tendency of the response variable Y to vary with the predictor variable X in a systematic fashion.
2. A scattering of points around the curve of statistical relationship.

These two characteristics are embodied in a regression model by postulating that:

1. There is a probability distribution of Y for each level of X.
2. The means of these probability distributions vary in some systematic fashion with X.

LINEAR ALGEBRA (REVIEW)

1. A vector **u** in N-dimensional space is an array of the form

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{pmatrix}$$

2. Vector addition is defined componentwise by

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_N \end{pmatrix} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \\ \vdots \\ u_N + v_N \end{pmatrix}$$

Example

If $\mathbf{v} = [1, -3, 2]^T$ and $\mathbf{w} = [4, 2, 1]^T$, then

$$\mathbf{v} + \mathbf{w} = \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix} + \begin{pmatrix} 4 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ -1 \\ 3 \end{pmatrix}$$

3. Multiplication by a scalar c, is defined componentwise by

$$c \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{pmatrix} = \begin{pmatrix} cu_1 \\ cu_2 \\ cu_3 \\ \vdots \\ cu_N \end{pmatrix}$$

Example

If $\mathbf{v} = [1, -3, 2]^T$ and $\mathbf{w} = [4, 2, 1]^T$, then

$$2\mathbf{v} = \begin{pmatrix} 2 \\ -6 \\ 4 \end{pmatrix}$$

$$-\mathbf{w} = \begin{pmatrix} -4 \\ -2 \\ -1 \end{pmatrix}$$

Note that $\mathbf{v} - \mathbf{w} = \mathbf{v} + (-\mathbf{w})$.

4. The span of a set of vectors

$$\mathbf{u}_1 = \begin{pmatrix} u_{11} \\ u_{21} \\ u_{31} \\ \vdots \\ u_{N1} \end{pmatrix} \cdots \mathbf{u}_k = \begin{pmatrix} u_{1k} \\ u_{2k} \\ u_{31} \\ \vdots \\ u_{Nk} \end{pmatrix}$$

is the set of all vectors of the form

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_k\mathbf{u}_k$$

(that is, **all** linear combinations of $\mathbf{u}_1, \cdots, \mathbf{u}_k$). Any such span is termed a subspace of N-dimensional space.

5. The squared length (norm) of a vector $\mathbf{y}$ is

$$||\mathbf{y}||^2 = y_1^2 + y_2^2 + \cdots + y_N^2.$$

A unit vector $\mathbf{u}$, is a vector of length (norm) 1; that is $||\mathbf{u}|| = 1$.

Examples

The norm of the vector $\mathbf{u} = [-3, 2, 1]^T$ is

$$||\mathbf{u}|| = \sqrt{||\mathbf{u}||^2} = \sqrt{(-3)^2 + (2)^2 + (1)^2} = \sqrt{6}$$

The norm of the vector $\mathbf{v} = [1, 1, 1]^T$ is

$$||\mathbf{v}|| = \sqrt{||\mathbf{v}||^2} = \sqrt{(1)^2 + (1)^2 + (1)^2} = \sqrt{3}$$

$\mathbf{w} = \frac{\mathbf{v}}{\sqrt{3}}$ is a unit vector.

6. The angle, θ, between two vectors $\mathbf{u}$ and $\mathbf{v}$ is given by

$$\cos\theta = \frac{u_1 v_1 + u_2 v_2 + \cdots + u_n v_n}{||u||||v||} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||u||||v||}$$

($u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$ is termed the dot product of $\mathbf{u}$ and $\mathbf{v}$). Vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if $\theta = 90^0$. This occurs if and only if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Example

Consider the vectors $\mathbf{u} = [2, -1, 1]^T$ and $\mathbf{v} = [1, 1, 2]^T$. Find $\langle \mathbf{u}, \mathbf{v} \rangle$ and determine the angle θ between $\mathbf{u}$ and $\mathbf{v}$.

Solution

$$\begin{aligned}\langle \mathbf{u}, \mathbf{v} \rangle &= u_1 v_1 + u_2 v_2 + u_3 v_3 \\ &= (2)(1) + (-1)(1) + (1)(2) \\ &= 2 - 1 + 2 = 3\end{aligned}$$

$$\begin{aligned}||\mathbf{u}||^2 = \langle \mathbf{u}, \mathbf{u} \rangle &= u_1 u_1 + u_2 u_2 + u_3 u_3 \\ &= (2)^2 + (-1)^2 + (1)^2 \\ &= 6\end{aligned}$$

$$\begin{aligned}||\mathbf{v}||^2 = \langle \mathbf{v}, \mathbf{v} \rangle &= v_1 v_1 + v_2 v_2 + v_3 v_3 \\ &= (1)^2 + (1)^2 + (2)^2 \\ &= 6\end{aligned}$$

Solution

$$\begin{aligned}\cos\theta &= \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{||u||||v||} \\ &= \frac{3}{\sqrt{6}\sqrt{6}} \\ &= \frac{1}{2}\end{aligned}$$

Thus, $\theta = 60^0$.

7. An orthonormal coordinate system for N-space is a set of N orthogonal unit vectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_N$.

Example

Let $\mathbf{v}_1 = [1, 1, 1]^T$, $\mathbf{v}_2 = [-2, 1, 1]^T$, and $\mathbf{v}_3 = [0, -1, 1]^T$ and assume that $\Re^3$ has the Euclidean inner product (dot product). It is easy to show that the set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is orthogonal since $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{v}_3 \rangle = \langle \mathbf{v}_2, \mathbf{v}_3 \rangle = 0$.
The (Euclidean) norms of the vectors in our last example are: $||\mathbf{v}_1|| = \sqrt{3}$, $||\mathbf{v}_2|| = \sqrt{6}$, and $||\mathbf{v}_3|| = \sqrt{2}$.

Example

Let $\mathbf{u}_1 = \frac{1}{\sqrt{3}}[1, 1, 1]^T$, $\mathbf{u}_2 = \frac{1}{\sqrt{6}}[-2, 1, 1]^T$, and $\mathbf{u}_3 = \frac{1}{\sqrt{2}}[0, -1, 1]^T$.
I leave for you to verify that $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is orthonormal.

8. The orthonormal decomposition of an arbitrary vector $\mathbf{y}$ in terms of such a coordinate system is

$$\mathbf{y} = \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \cdots + \langle \mathbf{y}, \mathbf{u}_N \rangle \mathbf{u}_N$$

Example+

Let $\mathbf{u}_1 = \frac{1}{\sqrt{3}}[1,1,1]^T$, $\mathbf{u}_2 = \frac{1}{\sqrt{6}}[-2,1,1]^T$, and $\mathbf{u}_3 = \frac{1}{\sqrt{2}}[0,-1,1]^T$. Express the vector $\mathbf{y} = [1.1, 1.4, 1.4]^T$ as a linear combination of the vector in S, and find the coordinate vector $(\mathbf{y})_S$.

Solution

$$\begin{aligned}
\langle \mathbf{y}, \mathbf{u}_1 \rangle &= \frac{1}{\sqrt{3}}(1.1 + 1.4 + 1.4) \\
&= \frac{3.9}{\sqrt{3}} \\
&= \sqrt{3}(1.3)
\end{aligned}$$

Solution

$$
\begin{aligned}
\langle \mathbf{y}, \mathbf{u}_2 \rangle &= \frac{1}{\sqrt{6}}(-2.2 + 1.4 + 1.4) \\
&= \frac{0.6}{\sqrt{6}} \\
&= \sqrt{6}(0.1)
\end{aligned}
$$

Solution

$$\begin{aligned}\langle \mathbf{y}, \mathbf{u}_3 \rangle &= \frac{1}{\sqrt{2}}(0) \\ &= 0\end{aligned}$$

Solution

Finally,

$$\mathbf{y} = 1.3(\sqrt{3})\mathbf{u}_1 + 0.1(\sqrt{6})\mathbf{u}_2 + (0)\mathbf{u}_3$$

The coordinate vector of $\mathbf{y}$ relative to S is

$$\begin{aligned}(\mathbf{y})_S &= [\langle \mathbf{y}, \mathbf{u}_1 \rangle, \langle \mathbf{y}, \mathbf{u}_2 \rangle, \langle \mathbf{y}, \mathbf{u}_3 \rangle]^T \\ &= [1.3\sqrt{3}, 0.1\sqrt{6}, 0]^T\end{aligned}$$

9. If $\mathbf{u}$ and $\mathbf{a}$ are vectors in N-space and if $\mathbf{a} \neq \mathbf{0}$, then

$$proj_{\mathbf{a}}\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{a} \rangle}{||\mathbf{a}||^2}\mathbf{a}$$

(vector component of $\mathbf{u}$ along $\mathbf{a}$)

$$\mathbf{u} - proj_{\mathbf{a}}\mathbf{u} = \mathbf{u} - \frac{\langle \mathbf{u}, \mathbf{a} \rangle}{||\mathbf{a}||^2}\mathbf{a}$$

(vector component of $\mathbf{u}$ orthogonal to $\mathbf{a}$)

Example^{++}

Let $\mathbf{y} = [1.1, 1.4, 1.4]^T$ and $\mathbf{u}_1 = \frac{1}{\sqrt{3}}[1, 1, 1]^T$. Find the vector component of $\mathbf{y}$ along $\mathbf{u}_1$ and the vector component of $\mathbf{y}$ orthogonal to $\mathbf{u}_1$.

Solution

$$
\begin{aligned}
\langle \mathbf{y}, \mathbf{u}_1 \rangle &= \frac{(1.1)(1) + (1.4)(1) + (1.4)(1)}{\sqrt{3}} \\
&= \frac{3.9}{\sqrt{3}} \\
&= 1.3\sqrt{3} \\
&= \sqrt{3}\bar{\mathbf{y}}
\end{aligned}
$$

$$||\mathbf{u}_1||^2 = 1$$

Solution

Thus, the vector component of $\mathbf{y}$ along $\mathbf{u}_1$ is

$$
\begin{aligned}
proj_{\mathbf{u}_1}\mathbf{y} &= 1.3\sqrt{3}\left[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right]^T \\
&= [1.3, 1.3, 1.3]^T \\
&= [\bar{\mathbf{y}}, \bar{\mathbf{y}}, \bar{\mathbf{y}}]^T \\
&= \begin{pmatrix} \bar{\mathbf{y}} \\ \bar{\mathbf{y}} \\ \bar{\mathbf{y}} \end{pmatrix}
\end{aligned}
$$

Solution

The vector component of $\mathbf{y}$ orthogonal to $\mathbf{u}_1$ is

$$\begin{aligned}
\mathbf{y} - proj_{\mathbf{u}_1}\mathbf{y} &= [1.1, 1.4, 1.4]^T - 1.3\sqrt{3}\left[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right]^T \\
&= [1.1, 1.4, 1.4]^T - [1.3, 1.3, 1.3]^T \\
&= [-0.2, 0.1, 0.1]^T \\
&= [\mathbf{y}_1 - \bar{\mathbf{y}}, \mathbf{y}_2 - \bar{\mathbf{y}}, \mathbf{y}_3 - \bar{\mathbf{y}}]^T \\
&= \begin{pmatrix} \mathbf{y}_1 - \bar{\mathbf{y}} \\ \mathbf{y}_2 - \bar{\mathbf{y}} \\ \mathbf{y}_3 - \bar{\mathbf{y}} \end{pmatrix}
\end{aligned}$$

Note that $\langle proj_{\mathbf{u}_1}\mathbf{y}, \mathbf{y} - proj_{\mathbf{u}_1}\mathbf{y}\rangle = 0$.

10. The nearest point of $\mathbf{y}$ in M, the span of $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$, $k \leq N$, is

$$\langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \cdots + \langle \mathbf{y}, \mathbf{u}_k \rangle \mathbf{u}_k$$

This vector is called the projection of $\mathbf{y}$ onto the subspace M.

11. Pythagora's theorem in N-space says that

$$||\mathbf{y}||^2 = \langle \mathbf{y}, \mathbf{u}_1 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_2 \rangle^2 + \cdots + \langle \mathbf{y}, \mathbf{u}_N \rangle^2$$

That is,
$(\text{length})^2 =$ sum of squared lengths of projections.

Example 1

$y_1, y_2, y_3 \sim \mathrm{N}(\mu, \sigma^2)$ y_is independent. Our aim is to estimate μ and σ^2 and to test the null hypothesis $H_0 : \mu = 0$ vs $H_a : \mu \neq 0$.
Our observation vector is:

$$\mathbf{y} = [1.1, 1.4, 1.4]^T = \begin{pmatrix} 1.1 \\ 1.4 \\ 1.4 \end{pmatrix}$$

Model

$$y_i = \mu + e_i$$

$$\mathbf{y} = \begin{pmatrix} \mu \\ \mu \\ \mu \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

where e_i are independent $\mathrm{N}(0, \sigma^2)$.

Model

$$y_i = \mu + e_i$$

$$\begin{pmatrix} 1.1 \\ 1.4 \\ 1.4 \end{pmatrix} = \mu \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

Model space M

$$M = span\left\{\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}$$

Fitting Model

We use the method of least squares and choose $\hat{\mu}$ (the estimate of μ) so that $\hat{\mu}$ is closest to $\mathbf{y}$. We must project $\mathbf{y}$ onto M. From example^{++}, if $\mathbf{u}_1 = \frac{1}{\sqrt{3}}[1, 1, 1]^T$

$$
\begin{aligned}
proj_{\mathbf{u}_1}\mathbf{y} &= 1.3\sqrt{3}\left[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right]^T \\
&= 1.3\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\
&= \bar{y}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}
\end{aligned}
$$

or $\bar{y}$ is the least squares approximation to μ.

Decomposition

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \bar{y} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ y_3 - \bar{y} \end{pmatrix}$$

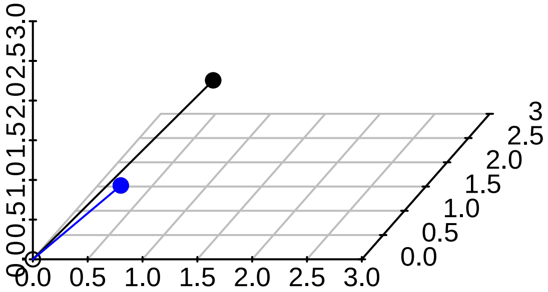
or observation vector = mean vector + residual vector.

Decomposition

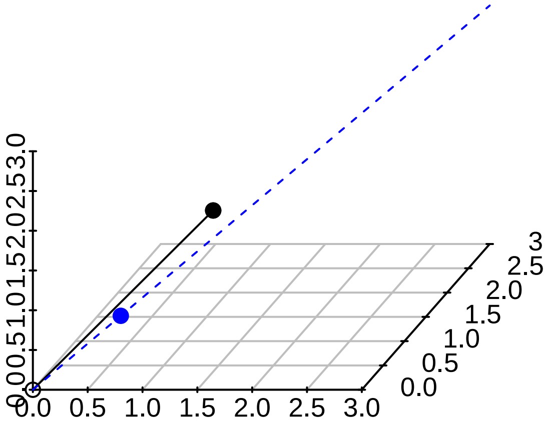
$$\begin{pmatrix} 1.1 \\ 1.4 \\ 1.4 \end{pmatrix} = 1.3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -0.2 \\ 0.1 \\ 0.1 \end{pmatrix}$$

Note that we have broken $\mathbf{y}$ into two **orthogonal** components, a model vector and a residual vector.

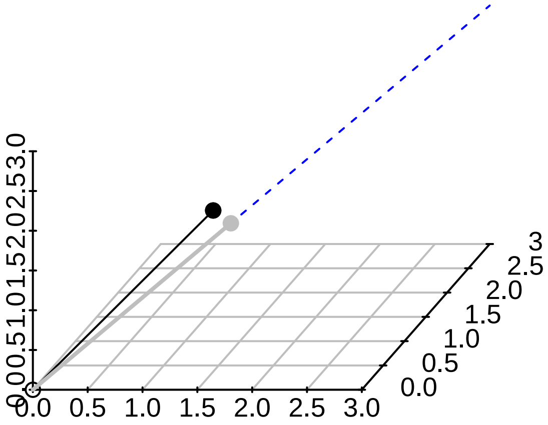
y vs unit vector



y vs model space (M)



y vs estimate (projection of y onto M)



Testing hypothesis

Suppose now that we wish to test $H_0 : \mu = 0$ vs $H_a : \mu \neq 0$. We take any orthonormal coordinate system for 3-space that includes $\mathbf{u}_1$. From example$^+$, $\mathbf{u}_1 = \frac{1}{\sqrt{3}}[1, 1, 1]^T$, $\mathbf{u}_2 = \frac{1}{\sqrt{6}}[-2, 1, 1]^T$, and $\mathbf{u}_3 = \frac{1}{\sqrt{2}}[0, -1, 1]^T$ should work. The space spanned by $\mathbf{u}_2$ and $\mathbf{u}_3$ we term the residual space, since the residual vector always lies in this space.

Testing hypothesis

Pythagora's theorem now tells us that

$$||\mathbf{y}||^2 = \langle \mathbf{y}, \mathbf{u}_1 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_2 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_3 \rangle^2$$

or

$$5.13 = 5.07 + 0.06$$

We can easily show that

$$\langle \mathbf{y}, \mathbf{u}_1 \rangle = \frac{y_1 + y_2 + y_3}{\sqrt{3}} \sim N(\sqrt{3}\mu, \sigma^2)$$

$$\langle \mathbf{y}, \mathbf{u}_2 \rangle = \frac{-2y_1 + y_2 + y_3}{\sqrt{6}} \sim N(0, \sigma^2)$$

$$\langle \mathbf{y}, \mathbf{u}_3 \rangle = \frac{-y_2 + y_3}{\sqrt{2}} \sim N(0, \sigma^2)$$

Testing hypothesis

If H_0 holds

$$\frac{\langle \mathbf{y}, \mathbf{u}_1 \rangle^2}{\sigma^2} \sim \chi^2(1)$$

$$\frac{\langle \mathbf{y}, \mathbf{u}_2 \rangle^2}{\sigma^2} \sim \chi^2(1)$$

$$\frac{\langle \mathbf{y}, \mathbf{u}_3 \rangle^2}{\sigma^2} \sim \chi^2(1)$$

Test statistic

Assuming that $\langle \mathbf{y}, \mathbf{u}_1 \rangle$, $\langle \mathbf{y}, \mathbf{u}_2 \rangle$ and $\langle \mathbf{y}, \mathbf{u}_3 \rangle$ are independent and if H_0 holds

$$\frac{\frac{1}{\sigma^2}(\langle \mathbf{y}, \mathbf{u}_1 \rangle)^2}{\frac{1}{2\sigma^2}\left(\langle \mathbf{y}, \mathbf{u}_2 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_3 \rangle^2\right)} \sim F_{1,2}$$

which is equivalent to

$$\frac{\langle \mathbf{y}, \mathbf{u}_1 \rangle^2}{\frac{1}{2}\left(\langle \mathbf{y}, \mathbf{u}_2 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_3 \rangle^2\right)} \sim F_{1,2}$$

For our data,

$$\frac{5.07}{\frac{0.06}{2}} = \frac{5.07}{0.03} = 169$$

Using an F table, we reject $H_0 : \mu = 0$ at the 1% significance level.

Showing independence of test statistic components

Using the Multivariate Transformation Method we can show that $\langle \mathbf{y}, \mathbf{u}_1 \rangle, \langle \mathbf{y}, \mathbf{u}_2 \rangle$ and $\langle \mathbf{y}, \mathbf{u}_3 \rangle$ are independent.
Let $X_1 = \frac{Y_1+Y_2+Y_3}{\sqrt{3}}$, $X_2 = \frac{-2Y_1+Y_2+Y_3}{\sqrt{6}}$, and $X_3 = \frac{Y_3-Y_2}{\sqrt{2}}$.
The random variables X_1 X_2 and X_3 are jointly continuous with joint density function given by
$f_{X_1,\ X_2,\ X_3}(x_1,\ x_2,\ x_3) = f_{Y_1,\ Y_2,\ Y_3}(y_1,\ y_2,\ y_3)|J(x_1, x_2, x_3)|^{-1}$,

Showing independence of test statistic components

Letting $x_1 = g_1(y_1,\ y_2,\ y_3) = \frac{y_1+y_2+y_3}{\sqrt{3}}$ and $x_2 = g_2(y_1,\ y_2,\ y_3) = \frac{-2Y_1+Y_2+Y_3}{\sqrt{6}}$, and $x_3 = \frac{y_3-y_2}{\sqrt{2}}$ we see that

$\frac{\partial x_1}{\partial y_1} = \frac{1}{\sqrt{3}}$

$\frac{\partial x_1}{\partial y_2} = \frac{1}{\sqrt{3}}$

$\frac{\partial x_1}{\partial y_3} = \frac{1}{\sqrt{3}}$

Showing independence of test statistic components

Letting $x_1 = g_1(y_1,\ y_2,\ y_3) = \frac{y_1+y_2+y_3}{\sqrt{3}}$ and
$x_2 = g_2(y_1,\ y_2,\ y_3) = \frac{-2Y_1+Y_2+Y_3}{\sqrt{6}}$, and $x_3 = \frac{y_3-y_2}{\sqrt{2}}$ we see that
$\frac{\partial x_2}{\partial y_1} = \frac{-2}{\sqrt{6}}$
$\frac{\partial x_2}{\partial y_2} = \frac{1}{\sqrt{6}}$
$\frac{\partial x_2}{\partial y_3} = \frac{1}{\sqrt{6}}$

Showing independence of test statistic components

Letting $x_1 = g_1(y_1,\ y_2,\ y_3) = \frac{y_1+y_2+y_3}{\sqrt{3}}$ and $x_2 = g_2(y_1,\ y_2,\ y_3) = \frac{-2Y_1+Y_2+Y_3}{\sqrt{6}}$, and $x_3 = \frac{y_3-y_2}{\sqrt{2}}$ we see that

$\frac{\partial x_3}{\partial y_1} = 0$

$\frac{\partial x_3}{\partial y_2} = \frac{-1}{\sqrt{2}}$

$\frac{\partial x_3}{\partial y_3} = \frac{1}{\sqrt{2}}$

Showing independence of test statistic components

$$J(x_1, x_2, x_3) = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \frac{\partial x_1}{\partial y_3} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \frac{\partial x_2}{\partial y_3} \\ \frac{\partial x_3}{\partial y_1} & \frac{\partial x_3}{\partial y_2} & \frac{\partial x_3}{\partial y_3} \end{pmatrix}$$

Showing independence of test statistic components

$$J(x_1, x_2, x_3) = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

Showing independence of test statistic components

Hence

$$|J(x_1, x_2, x_3)| = \frac{1}{6} + \frac{2}{6} + \frac{1}{6} + \frac{2}{6} = 1$$

and

$$|J(x_1, x_2, x_3)|^{-1} = 1$$

Showing independence of test statistic components

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

$$\mathbf{X} = \mathbf{QY}$$

Showing independence of test statistic components

$$\mathbf{Q}^T = \begin{pmatrix} 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \end{pmatrix}$$

It is easy to show that

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_{3\times 3}$$

Showing independence of test statistic components

Which means that the inverse transformations are given by

$$y_1 = g_1^{-1}(x_1,\ x_2,\ x_3) = \frac{x_1}{\sqrt{3}} - \frac{2x_2}{\sqrt{6}}$$

$$y_2 = g_2^{-1}(x_1,\ x_2,\ x_3) = \frac{x_1}{\sqrt{3}} + \frac{x_2}{\sqrt{6}} - \frac{x_3}{\sqrt{2}}$$

$$y_3 = g_3^{-1}(x_1,\ x_2,\ x_3) = \frac{x_1}{\sqrt{3}} + \frac{x_2}{\sqrt{6}} + \frac{x_3}{\sqrt{2}}$$

Showing independence of test statistic components

$$y_1^2 = \frac{x_1^2}{3} + \frac{4x_2^2}{6} - \frac{4x_1x_2}{\sqrt{3}\sqrt{6}}$$

$$y_2^2 = \frac{x_1^2}{3} + \frac{x_2^2}{6} + 2\frac{x_1x_2}{\sqrt{3}\sqrt{6}} + \frac{x_3^2}{2} - 2\frac{x_1x_3}{\sqrt{3}\sqrt{2}} - 2\frac{x_2x_3}{\sqrt{6}\sqrt{2}}$$

$$y_3^2 = \frac{x_1^2}{3} + \frac{x_2^2}{6} + \frac{x_3^2}{2} + 2\frac{x_1x_2}{\sqrt{3}\sqrt{6}} + 2\frac{x_1x_3}{\sqrt{3}\sqrt{2}} + 2\frac{x_2x_3}{\sqrt{6}\sqrt{2}}$$

Showing independence of test statistic components

As the joint density function of Y_1, Y_2 and Y_3 is

$$f_{Y_1,\ Y_2,\ Y_3}(y_1, y_2, y_3) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^3 e^{-(y_1^2+y_2^2+y_3^2)/2\sigma^2}$$

we see that the joint density function of X_1, X_2, and X_3, is given by

$$f_{X_1,\ X_2,\ X_3}(x_1,\ x_2,\ x_3) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^3 e^{-[(g_1^{-1})^2+(g_2^{-1})^2+(g_3^{-1})^2]/2\sigma^2}$$

(after using the expressions from our previous slide and a "little" algebra...)

$$f_{X_1,\ X_2,\ X_3}(x_1,\ x_2,\ x_3) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^3 e^{-(x_1^2+x_2^2+x_3^2)/2\sigma^2}$$

Showing independence of test statistic components

Hence, we have shown that if Y_1, Y_2, and Y_3 are independent Normally distributed random variables with mean zero and variance σ^2, then X_1, X_2, and X_3 are also independent Normally distributed random variables with mean zero and variance σ^2.

Example 2

$y_1, y_2, y_3, y_4, y_5 \sim \mathrm{N}(\beta_0 + \beta_1 x_i, \sigma^2)$, where y_is independent and $\mathbf{x} = [1, 2, 3, 4, 5]^T$. Our aim is to estimate β_0, β_1 and σ^2 and to test the null hypothesis $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$.
Our observation vector is:

$$\begin{aligned} \mathbf{y} &= [2.1, 3.1, 3.0, 3.8, 4.3]^T \\ &= \begin{pmatrix} 2.1 \\ 3.1 \\ 3.0 \\ 3.8 \\ 4.3 \end{pmatrix} \end{aligned}$$

Model

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

$$\mathbf{y} = \begin{pmatrix} \beta_0 + \beta_1(1) \\ \beta_0 + \beta_1(2) \\ \beta_0 + \beta_1(3) \\ \beta_0 + \beta_1(4) \\ \beta_0 + \beta_1(5) \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{pmatrix}$$

where e_i are independent $N(0, \sigma^2)$.

Model

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

$$\begin{pmatrix} 2.1 \\ 3.1 \\ 3.0 \\ 3.8 \\ 4.3 \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{pmatrix}$$

where e_i are independent $\mathrm{N}(0, \sigma^2)$.

Re-writing Model

We are going to apply Gram-Schmidt method to find an orthogonal basis for the Model space.
We have $\{\mathbf{v}_1, \mathbf{v}_2\} = \{[1, 1, 1, 1, 1]^T, [1, 2, 3, 4, 5]^T\}$. Now, let's define $\{\mathbf{w}_1, \mathbf{w}_2\}$ as follows:

$$\mathbf{w}_1 = \mathbf{v}_1$$

and

$$\mathbf{w}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{||w_1||^2}\mathbf{w}_1$$

Re-writing Model

In this case,

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

and

$$\mathbf{w}_2 = \begin{pmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \\ x_4 - \bar{x} \\ x_5 - \bar{x} \end{pmatrix}$$

"New" Model

$$y_i = \beta_0^* + \beta_1^*(x_i - \bar{x}) + e_i^*$$

$$\begin{pmatrix} 2.1 \\ 3.1 \\ 3.0 \\ 3.8 \\ 4.3 \end{pmatrix} = \beta_0^* \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta_1^* \begin{pmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{pmatrix} + \begin{pmatrix} e_1^* \\ e_2^* \\ e_3^* \\ e_4^* \\ e_5^* \end{pmatrix}$$

where e_i^* are independent $\mathrm{N}(0, \sigma^2)$.

Model space M

$$M = span\left\{\begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}-2\\-1\\0\\1\\2\end{pmatrix}\right\}$$

We use the same method as before and choose $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$ (estimates of β_0^* and β_1^*, respectively) so that $\hat{\mathbf{y}}$ is closest to $\mathbf{y}$. We must project $\mathbf{y}$ onto M. If $\mathbf{u}_1 = \frac{1}{\sqrt{5}}[1,1,1,1,1]^T$ and $\mathbf{u}_2 = \frac{1}{\sqrt{10}}[-2,-1,0,1,2]^T$

$$\langle \mathbf{y}, \mathbf{u}_1 \rangle = \frac{16.3}{\sqrt{5}}$$

$$\begin{aligned} \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 &= \frac{16.3}{5} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \\ &= 3.26 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \end{aligned}$$

Fitting Model

$$\langle \mathbf{y}, \mathbf{u}_2 \rangle = \frac{5.1}{\sqrt{10}}$$

$$\begin{aligned}
\langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 &= \frac{5.1}{10} \begin{pmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{pmatrix} \\
&= 0.51 \begin{pmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{pmatrix}
\end{aligned}$$

Decomposition

$$\begin{pmatrix} 2.1 \\ 3.1 \\ 3.0 \\ 3.8 \\ 4.3 \end{pmatrix} = 3.26 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + 0.51 \begin{pmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{pmatrix} + \begin{pmatrix} -0.14 \\ 0.35 \\ -0.26 \\ 0.03 \\ 0.02 \end{pmatrix}$$

Note that we have broken $\mathbf{y}$ into two **orthogonal** components, a model vector and a residual vector.

Testing hypothesis

Suppose now that we wish to test $H_0 : \beta_1^* = 0$ vs $H_a : \beta_1^* \neq 0$.
We can easily show that under $H_0 : \beta_1^* = 0$,

$$\langle \mathbf{y}, \mathbf{u}_2 \rangle = \frac{-2y_1 - y_2 + y_4 + 2y_5}{\sqrt{10}} \sim N(0, \sigma^2).$$

Test statistic

So, if $\mathbf{u}_3$, $\mathbf{u}_4$, and $\mathbf{u}_5$ complete our coordinate system,

$$\frac{\frac{1}{\sigma^2}(\langle \mathbf{y}, \mathbf{u}_2 \rangle)^2}{\frac{1}{3\sigma^2}\left(\langle \mathbf{y}, \mathbf{u}_3 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_4 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_5 \rangle^2\right)} \sim F_{1,3}$$

which is equivalent to

$$\frac{\langle \mathbf{y}, \mathbf{u}_2 \rangle^2}{\frac{1}{3}\left(\langle \mathbf{y}, \mathbf{u}_3 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_4 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_5 \rangle^2\right)} \sim F_{1,3}$$

For our data,

$$\frac{26.01/10}{\frac{0.211}{3}} = \frac{2.601}{0.0703} \approx 37$$

Using an F table, we reject $H_0 : \beta_1 = 0$ at the 1% significance level.

A few comments about decomposition and test statistic

Alternative Decomposition

We claim that we can express $\mathbf{y}$ as a linear combination of $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, $\mathbf{u}_4$, and $\mathbf{u}_5$, that is

$$\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5$$

Moreover, we claim that $c_i = \langle \mathbf{y}, \mathbf{u}_i \rangle$.

Model space M

$$M = span \left\{ \begin{pmatrix} 1/\sqrt{5} \\ 1/\sqrt{5} \\ 1/\sqrt{5} \\ 1/\sqrt{5} \\ 1/\sqrt{5} \end{pmatrix}, \begin{pmatrix} -2/\sqrt{10} \\ -1/\sqrt{10} \\ 0 \\ 1/\sqrt{10} \\ 2/\sqrt{10} \end{pmatrix} \right\}$$

Residual space R

$$R = span\left\{ \begin{pmatrix} 2/\sqrt{14} \\ -1/\sqrt{14} \\ -2/\sqrt{14} \\ -1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix}, \begin{pmatrix} -1/\sqrt{10} \\ 2/\sqrt{10} \\ 0 \\ -2/\sqrt{10} \\ 1/\sqrt{10} \end{pmatrix}, \begin{pmatrix} 1/\sqrt{70} \\ -4/\sqrt{70} \\ 6/\sqrt{70} \\ -4/\sqrt{70} \\ 1/\sqrt{70} \end{pmatrix} \right\}$$

Inner products (c_is)

$$\langle \mathbf{y}, \mathbf{u}_1 \rangle = \frac{16.3}{\sqrt{5}}$$
$$\langle \mathbf{y}, \mathbf{u}_2 \rangle = \frac{5.1}{\sqrt{10}}$$
$$\langle \mathbf{y}, \mathbf{u}_3 \rangle = \frac{-0.1}{\sqrt{14}}$$
$$\langle \mathbf{y}, \mathbf{u}_4 \rangle = \frac{0.8}{\sqrt{10}}$$
$$\langle \mathbf{y}, \mathbf{u}_5 \rangle = \frac{-3.2}{\sqrt{70}}$$

Checking that it works

```
# Model space
m1=c(1,1,1,1,1);
m2=c(-2,-1,0,1,2);

# Residual space
r1=c(2,-1,-2,-1,2);
r2=c(-1,2,0,-2,1);
r3=c(1,-4,6,-4,1);

# Verifying they form a basis

M=cbind(m1,m2);
R=cbind(r1,r2,r3);
Q=cbind(M,R);

t(Q)%*%Q;
```

Checking that it works

```
##     m1 m2 r1 r2 r3
## m1   5  0  0  0  0
## m2   0 10  0  0  0
## r1   0  0 14  0  0
## r2   0  0  0 10  0
## r3   0  0  0  0 70
```

Note. They are orthogonal to each other!

Finding coefficients

```
# Finding norms;
sq.norms=t(Q)%*%Q;
sq.norms=diag(sq.norms);

# Observed vector;
y=c(2.1,3.1,3,3.8,4.3);
y=matrix(y,ncol=1);

# Coefficients;
coeff1=t(y)%*%Q[ ,1];
coeff2=t(y)%*%Q[ ,2];
coeff3=t(y)%*%Q[ ,3];
coeff4=t(y)%*%Q[ ,4];
coeff5=t(y)%*%Q[ ,5];

coeff1;
```

Finding coefficients

```
##       [,1]
## [1,] 16.3
##       [,1]
## [1,]  5.1
##       [,1]
## [1,] -0.1
##       [,1]
## [1,]  0.8
##       [,1]
## [1,] -3.2
```

Finding projections and decomposition

```
proj1=matrix(coeff1/sq.norms[1]*Q[ ,1],ncol=1);

proj2=matrix(coeff2/sq.norms[2]*Q[ ,2],ncol=1);

proj3=matrix(coeff3/sq.norms[3]*Q[ ,3],ncol=1);

proj4=matrix(coeff4/sq.norms[4]*Q[ ,4],ncol=1);

proj5=matrix(coeff5/sq.norms[5]*Q[ ,5],ncol=1);

decomp=proj1+proj2+proj3+proj4+proj5;

decomp;
```

Finding projections and decomposition

```
##        [,1]
## [1,]   2.1
## [2,]   3.1
## [3,]   3.0
## [4,]   3.8
## [5,]   4.3
```

Note. Our decomposition yields **y**!

Test statistic

We claim that we **don't need** $\mathbf{u}_3$, $\mathbf{u}_4$, and $\mathbf{u}_5$ to compute our test statistic. To do so, we only need $\mathbf{y}$, $\mathbf{u}_1$, and $\mathbf{u}_2$.
Let $\hat{\mathbf{y}} = \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2$ (predicted vector) and
$\hat{\epsilon} = \langle \mathbf{y}, \mathbf{u}_3 \rangle \mathbf{u}_3 + \langle \mathbf{y}, \mathbf{u}_4 \rangle \mathbf{u}_4 + \langle \mathbf{y}, \mathbf{u}_5 \rangle \mathbf{u}_5$ (residual vector).
We have already established that

$$\mathbf{y} = \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \langle \mathbf{y}, \mathbf{u}_3 \rangle \mathbf{u}_3 + \langle \mathbf{y}, \mathbf{u}_4 \rangle \mathbf{u}_4 + \langle \mathbf{y}, \mathbf{u}_5 \rangle \mathbf{u}_5$$

Therefore,

$$||\mathbf{y}||^2 = \langle \mathbf{y}, \mathbf{u}_1 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_2 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_3 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_4 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_5 \rangle^2$$

Which implies that

$$||\mathbf{y} - \hat{\mathbf{y}}||^2 = ||\hat{\epsilon}||^2$$

Checking that it works

```
y.hat=proj1+proj2;

residual=proj3+proj4+proj5;

y.norm2=t(y)%*%y;
y.hat.norm2=t(y.hat)%*%y.hat;
res.norm2=t(residual)%*%residual;

res.norm2;

t(y-y.hat)%*%(y-y.hat);
```

Checking that it works

```
##         [,1]
## [1,] 0.211
##         [,1]
## [1,] 0.211
```

Homework?

Show that $E(\langle \mathbf{y}, \mathbf{u} \rangle) = 0$ for any vector $\mathbf{u}$ in the error space.

SIMPLE LINEAR REGRESSION

Assumptions

1. The mean of an observation is assumed to depend on the x value with which it is associated, via the straight line relationship $E(Y) = \beta_0 + \beta_1(x - \bar{x})$. Here β_0 and β_1 are the unknown parameters of the line.
2. For each x value, Y is assumed to be Normally distributed about this mean.
3. For each x value, Y is assumed to have a common variance of σ^2.
4. We assume that in sampling, our errors, the deviation of our observations from the line, are independent values from a $N(0, \sigma^2)$ distribution.

Initial Model

$$y_i = \beta_0^* + \beta_1^* x_i + e_i^*$$

$$\mathbf{y} = \begin{pmatrix} \beta_0^* + \beta_1^*(x_1) \\ \beta_0 * + \beta_1^*(x_2) \\ \beta_0 * + \beta_1^*(x_3) \\ \beta_0 * + \beta_1^*(x_4) \\ \vdots \\ \beta_0 * + \beta_1^*(x_n) \end{pmatrix} + \begin{pmatrix} e_1^* \\ e_2^* \\ e_3^* \\ e_4^* \\ \vdots \\ e_n^* \end{pmatrix}$$

where e_i^* are independent $\mathrm{N}(0, \sigma^2)$.

Model

$$y_i = \beta_0^* + \beta_1^* x_i + e_i^*$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \beta_0^* \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1^* \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} e_1^* \\ e_2^* \\ e_3^* \\ \vdots \\ e_n^* \end{pmatrix}$$

where e_i^* are independent $\mathrm{N}(0, \sigma^2)$.

Re-writing Model

We are going to apply Gram-Schmidt method to find an orthogonal basis for the Model space.
We have $\{\mathbf{v}_1, \mathbf{v}_2\} = \{[1, 1, 1, 1, 1]^T, [x_1, x_2, x_3, \cdots, x_n]^T\}$. Now, let's define $\{\mathbf{w}_1, \mathbf{w}_2\}$ as follows:

$$\mathbf{w}_1 = \mathbf{v}_1$$

and

$$\mathbf{w}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{||w_1||^2} \mathbf{w}_1$$

In this case,

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

and

$$\mathbf{w}_2 = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} - \frac{\sum x_i}{n} \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}$$

"New" Model

$$y_i = \beta_0 + \beta_1(x_i - \bar{x}) + e_i$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{pmatrix}$$

where e_i are independent $\mathrm{N}(0, \sigma^2)$.

Model space M

$$M = span\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} \right\}$$

Fitting Model

We must project $\mathbf{y}$ onto M. Note that $\mathbf{u}_1$ and $\mathbf{u}_2$ form an orthonormal coordinate system for our model space, where

$$\mathbf{u}_1 = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sqrt{\sum (x_i - \bar{x})^2}} \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}$$

Fitting Model

$$\langle \mathbf{y}, \mathbf{u}_1 \rangle = \frac{\sum y_i}{\sqrt{n}}$$

$$\begin{aligned}
\langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 &= \frac{\sum y_i}{\sqrt{n}} \begin{pmatrix} \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{n}} \\ \vdots \\ \frac{1}{\sqrt{n}} \end{pmatrix} \\
&= \bar{y} \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}
\end{aligned}$$

Fitting Model

$$\langle \mathbf{y}, \mathbf{u}_2 \rangle = \frac{\sum (x_i - \bar{x}) y_i}{\sqrt{\sum (x_i - \bar{x})^2}}$$

$$\langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 \quad = \quad \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}$$

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} = \frac{\langle \mathbf{y}, \mathbf{u}_2 \rangle}{||\mathbf{x} - \bar{\mathbf{x}}||}$$

Decomposition

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \hat{\beta}_0 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \hat{\beta}_1 \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} + \begin{pmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \hat{e}_3 \\ \vdots \\ \hat{e}_n \end{pmatrix}$$

Note that we have broken $\mathbf{y}$ into two **orthogonal** components, a model vector and a residual vector.

Testing hypothesis

The hypothesis $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$ can be tested using the test statistic

$$\frac{\langle \mathbf{y}, \mathbf{u}_2 \rangle^2}{\frac{\langle \mathbf{y}, \mathbf{u}_3 \rangle^2 + \langle \mathbf{y}, \mathbf{u}_4 \rangle^2 + \cdots + \langle \mathbf{y}, \mathbf{u}_n \rangle^2}{n-2}}$$

which comes from $F_{1,n-2}$ if H_0 is true.

Testing hypothesis (equivalent test statistic)

The hypothesis $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$ can be tested using the test statistic

$$\frac{\hat{\beta}_1^2 ||\mathbf{x} - \bar{\mathbf{x}}||^2}{\frac{||\text{residual vector}||^2}{n-2}}$$

which comes from $F_{1,n-2}$ if H_0 is true.

Distribution of test statistic under H_0

First, recall that

$$\langle \mathbf{y}, \mathbf{u}_2 \rangle = \frac{\sum (x_i - \bar{x}) y_i}{\sqrt{\sum (x_i - \bar{x})^2}} = \sum c_i y_i$$

If H_0 is true ($\beta_1 = 0$), $y_i \sim N(\beta_0, \sigma^2)$,

$$\begin{aligned}
E[\langle \mathbf{y}, \mathbf{u}_2 \rangle] &= E[\sum c_i y_i] \\
&= \sum E[c_i y_i] \\
&= \sum c_i \beta_0 \\
&= \beta_0 \sum c_i \\
&= \frac{\beta_0}{||\mathbf{x} - \bar{\mathbf{x}}||} \sum (x_i - \bar{x}) \\
&= 0
\end{aligned}$$

Distribution of test statistic under H_0

Again, recall that

$$\langle \mathbf{y}, \mathbf{u}_2 \rangle = \frac{\sum (x_i - \bar{x}) y_i}{\sqrt{\sum (x_i - \bar{x})^2}} = \sum c_i y_i$$

If H_0 is true ($\beta_1 = 0$), $y_i \sim N(\beta_0, \sigma^2)$,

$$\begin{aligned} V[\langle \mathbf{y}, \mathbf{u}_2 \rangle] &= V[\sum c_i y_i] \text{ (using independence)} \\ &= \sum V[c_i y_i] \\ &= \sum c_i^2 V[y_i] \\ &= \sum c_i^2 \sigma^2 \\ &= \frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \sigma^2 \\ &= \sigma^2 \end{aligned}$$

Distribution of test statistic under H_0

Therefore, if H_0 holds, we have that

$$\langle \mathbf{y}, \mathbf{u}_2 \rangle \sim N(0, \sigma^2)$$

and

$$\frac{\langle \mathbf{y}, \mathbf{u}_2 \rangle}{\sigma} \sim N(0, 1)$$

and

$$\frac{\langle \mathbf{y}, \mathbf{u}_2 \rangle^2}{\sigma^2} \sim \chi^2(1)$$

Distribution of test statistic under H_0

Using result from your homework problem, it is easy to show that

$$\frac{\langle \mathbf{y}, \mathbf{u}_3 \rangle^2}{\sigma^2} \sim \chi^2(1)$$

$$\frac{\langle \mathbf{y}, \mathbf{u}_4 \rangle^2}{\sigma^2} \sim \chi^2(1)$$

$$\vdots$$

$$\frac{\langle \mathbf{y}, \mathbf{u}_n \rangle^2}{\sigma^2} \sim \chi^2(1)$$

Showing independence of test statistic components

Using the Multivariate Transformation Method and generalizing ideas presented in Example 1, we can show that

$$\frac{\langle \mathbf{y}, \mathbf{u}_2 \rangle^2}{\sigma^2}$$

$$\frac{\langle \mathbf{y}, \mathbf{u}_3 \rangle^2}{\sigma^2}$$

$$\frac{\langle \mathbf{y}, \mathbf{u}_4 \rangle^2}{\sigma^2}$$

$$\vdots$$

$$\frac{\langle \mathbf{y}, \mathbf{u}_n \rangle^2}{\sigma^2}$$

are independent.

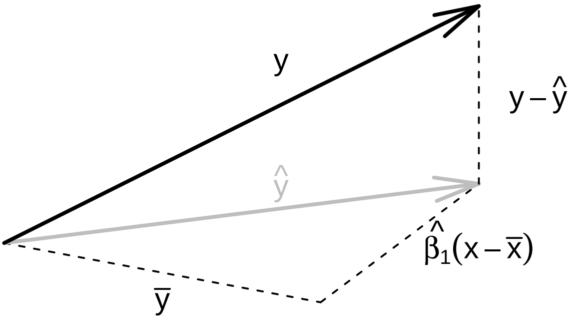
ANOVA table

The fitted model is

$$\mathbf{y} = \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \text{residual vector}$$

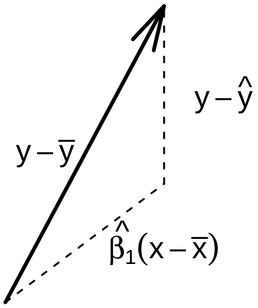
$$\mathbf{y} = \bar{\mathbf{y}} + \hat{\beta}_1 (\mathbf{x} - \bar{\mathbf{x}}) + \text{residual vector}$$

$$\mathbf{y} = \hat{\mathbf{y}} + \text{residual vector}$$


y
y – ŷ
ŷ
β̂₁(x – x̄)
ȳ

ANOVA table (simplified decomposition)

$$\mathbf{y} - \bar{\mathbf{y}} = \hat{\beta}_1(\mathbf{x} - \bar{\mathbf{x}}) + \text{residual vector}$$



$y - \hat{y}$
$y - \bar{y}$
$\hat{\beta}_1(x - \bar{x})$

ANOVA table

The simplified orthogonal decomposition leads to

$$||\mathbf{y} - \bar{\mathbf{y}}||^2 = \hat{\beta}_1^2 ||\mathbf{x} - \bar{\mathbf{x}}||^2 + ||\text{residual vector}||^2.$$

$$||\mathbf{y} - \bar{\mathbf{y}}||^2 = \hat{\beta}_1^2 ||\mathbf{x} - \bar{\mathbf{x}}||^2 + ||\mathbf{y} - \hat{\mathbf{y}}||^2.$$

This in turn leads to the ANOVA table shown below.

Calculations (example 2, again)

Step 1. Find $\bar{x}$, $\mathbf{x} - \bar{\mathbf{x}}$, and $||\mathbf{x} - \bar{\mathbf{x}}||^2$.
$\bar{x} = \frac{1+2+3+4+5}{5} = 3$

$$\mathbf{x} - \bar{\mathbf{x}} = \begin{pmatrix} 1-3 \\ 2-3 \\ 3-3 \\ 4-3 \\ 5-3 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$||\mathbf{x} - \bar{\mathbf{x}}||^2 = 10$

Calculations (example 2, again)

Step 2. Find $\bar{y}$, $\mathbf{y} - \bar{\mathbf{y}}$, and $||\mathbf{y} - \bar{\mathbf{y}}||^2$.
$\bar{y} = \frac{2.1+3.1+3.0+3.8+4.3}{5} = 3.26$

$$\mathbf{y} - \bar{\mathbf{y}} = \begin{pmatrix} 2.1 - 3.26 \\ 3.1 - 3.26 \\ 3.0 - 3.26 \\ 3.8 - 3.26 \\ 4.3 - 3.26 \end{pmatrix} = \begin{pmatrix} -1.16 \\ -0.16 \\ -0.26 \\ 0.54 \\ 1.04 \end{pmatrix}$$

$||\mathbf{y} - \bar{\mathbf{y}}||^2 = 2.812$

Calculations (example 2, again)

Step 3. Find $\hat{\beta}_1$ and $\hat{\beta}_1^2$.
Recall that
$\hat{\beta}_1 = \frac{\langle \mathbf{y}, \mathbf{x} - \bar{\mathbf{x}} \rangle}{||\mathbf{x} - \bar{\mathbf{x}}||^2} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}$
$\langle \mathbf{y}, \mathbf{x} - \bar{\mathbf{x}} \rangle = 5.1$
$\hat{\beta}_1 = \frac{5.1}{10} = 0.51$
$\hat{\beta}_1^2 = 0.51 = 0.2601$

Step 4. Use ANOVA table.

ANOVA table

Source of Variation	df	SS	MS	F
Regression	1	$\hat{\beta}_1^2 \|\|\mathbf{x} - \bar{\mathbf{x}}\|\|^2$	$\frac{\hat{\beta}_1^2 \|\|\mathbf{x} - \bar{\mathbf{x}}\|\|^2}{1}$	$\frac{\hat{\beta}_1^2 \|\|\mathbf{x} - \bar{\mathbf{x}}\|\|^2}{\|\|\mathbf{y} - \hat{\mathbf{y}}\|\|^2 / n-2}$
Error	n-2	$\|\|\mathbf{y} - \hat{\mathbf{y}}\|\|^2$	$\frac{\|\|\mathbf{y} - \hat{\mathbf{y}}\|\|^2}{n-2}$	
Total	n-1	$\|\|\mathbf{y} - \bar{\mathbf{y}}\|\|^2$		

ANOVA table

Source of Variation	df	SS	MS	F
Regression	1	**2.601**	2.601	$\frac{2.601}{0.0703} = 37$
Error	5-2 =3	2.812 - 2.601 = 0.211	$\frac{0.211}{3} = 0.0703$	
Total	5-1=4	**2.812**		

Estimation of σ^2

We have transformed the original independent set of random variables $y_i \sim N(\beta_0 + \beta_1(x_i - \bar{x}), \sigma^2)$, into a new independent set of Normal random variables, $\langle \mathbf{y}, \mathbf{u}_i \rangle$. With this new set, $\langle \mathbf{y}, \mathbf{u}_1 \rangle$ and $\langle \mathbf{y}, \mathbf{u}_2 \rangle$ are used to estimate parameters β_0 and β_1, the remaining $n - 2$ random variables are used to estimate σ^2.

Estimating σ^2

Let's define $\frac{W}{\sigma^2}$ as follows

$$\frac{W}{\sigma^2} = \frac{\langle \mathbf{y}, \mathbf{u}_3 \rangle^2}{\sigma^2} + \frac{\langle \mathbf{y}, \mathbf{u}_4 \rangle^2}{\sigma^2} + \cdots + \frac{\langle \mathbf{y}, \mathbf{u}_n \rangle^2}{\sigma^2} \sim \chi^2(n-2).$$

Then,

$$E\left(\frac{W}{\sigma^2}\right) = \frac{1}{\sigma^2} E(W) = n - 2$$

Therefore,

$$\frac{W}{n-2} = \frac{\sum_{i=3}^{n} \langle \mathbf{y}, \mathbf{u}_i \rangle^2}{n-2} = \frac{||\text{residual vector }||^2}{n-2}$$

is an unbiased estimator of σ^2.

$\hat{\sigma}^2$

We have that :

$$\hat{\sigma}^2 = \frac{||\text{residual vector }||^2}{n-2} = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{Y})^2 = \frac{SSE}{n-2} = S^2$$

The quantity SSE is also called the **sum of squares for error**.

PROPERTIES OF LEAST-SQUARES ESTIMATORS: SIMPLE LINEAR REGRESSION

We know that :

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})Y_i}{\sum(x_i - \bar{x})^2} = \frac{\sum(x_i - \bar{x})Y_i}{S_{xx}} = \sum c_i Y_i$$

where $c_i = \frac{(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = \frac{(x_i - \bar{x})}{S_{xx}}$

$$
\begin{aligned}
E(\hat{\beta}_1) &= E[\sum c_i Y_i] \\
&= \sum c_i E[Y_i] \\
&= \sum c_i[\beta_0 + \beta_1(x_i - \bar{x})] \\
&= \sum \beta_0 c_i + \beta_1 \sum c_i(x_i - \bar{x}) \\
&= \sum \beta_0 c_i + \beta_1 \frac{\sum(x_i - \bar{x})^2}{S_{xx}} \\
&= \beta_0 \sum c_i + \beta_1 \\
&= \beta_1
\end{aligned}
$$

Note that

$$\sum c_i = \frac{\sum(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = 0.$$

$V(\hat{\beta}_1)$

$$
\begin{aligned}
V(\hat{\beta}_1) &= V[\sum c_i Y_i] \text{ (Y_is are independent)} \\
&= \sum c_i^2 V[Y_i] \\
&= \sum c_i^2 \sigma^2 \\
&= \sigma^2 \sum c_i^2 \\
&= \sigma^2 \sum \left[\frac{(x_i - \bar{x})}{S_{xx}} \right]^2 \\
&= \sigma^2 \frac{\sum (x_i - \bar{x})^2}{S_{xx}^2} \\
&= \sigma^2 \frac{S_{xx}}{S_{xx}^2} \\
&= \frac{\sigma^2}{S_{xx}} = \frac{\sigma^2}{||\mathbf{x} - \bar{\mathbf{x}}||^2}
\end{aligned}
$$

$cov(\hat{\beta}_0, \hat{\beta}_1)$

We know that :

$$\hat{\beta}_0 = \bar{Y} = \frac{Y_i}{n}.$$

Then,

$$cov(\hat{\beta}_0, \hat{\beta}_1) = cov\left(\sum \frac{1}{n} Y_i, \sum c_i Y_i\right)$$

where $c_i = \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{(x_i - \bar{x})}{S_{xx}}$.
We can easily show that $cov(\hat{\beta}_0, \hat{\beta}_1) = 0$... Homework?

INFERENCES CONCERNING THE PARAMETERS β_i.

Testing the hypothesis of no linear relationship

We can also test hypotheses about the slope β_1. The most common hypothesis is

$$H_0 : \beta_1 = 0.$$

A regression line with slope 0 is horizontal. That is, the mean of y does not change at all when x changes. So this H_0 says that there is no true linear relationship between x and y.

Significance test for regression slope

To test the hypothesis $H_0 : \beta_1 = 0$, compute the t statistic

$$t_* = \frac{\hat{\beta}_1}{\frac{S}{||\mathbf{x} - \bar{\mathbf{x}}||}} = \frac{\hat{\beta}_1}{\frac{S}{\sqrt{S_{xx}}}}.$$

In terms of a random variable T having the $t(n-2)$ distribution, the P-value for a test of H_0 against $H_a : \beta \neq 0$ is $2P(T \geq |t_*|)$.

Development (numerator)

If $Y_1, Y_2, \cdots Y_n$ constitute a random sample from a Normal population with mean $\beta_0 + \beta_1(x_i - \bar{x})$ and variance σ^2 then,

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{||\mathbf{x} - \bar{\mathbf{x}}||^2}\right)$$

equivalent to

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right).$$

Which implies that

$$\frac{\hat{\beta}_1 - \beta_1}{\frac{\sigma}{||\mathbf{x} - \bar{\mathbf{x}}||}} \sim N(0, 1).$$

Now, recall that

$$\frac{W}{\sigma^2} = \frac{\langle \mathbf{y}, \mathbf{u}_3 \rangle^2}{\sigma^2} + \frac{\langle \mathbf{y}, \mathbf{u}_4 \rangle^2}{\sigma^2} + \cdots + \frac{\langle \mathbf{y}, \mathbf{u}_n \rangle^2}{\sigma^2} \sim \chi^2(n-2).$$

$$
\begin{aligned}
t_* &= \frac{\hat{\beta}_1 - \beta_1}{\frac{S}{||\mathbf{x}-\bar{\mathbf{x}}||}} \\
&= \frac{(\hat{\beta}_1 - \beta_1)||\mathbf{x} - \bar{\mathbf{x}}||}{S} \\
&= \frac{\frac{(\hat{\beta}_1-\beta_1)||\mathbf{x}-\bar{\mathbf{x}}||}{\sigma}}{\frac{S}{\sigma}} \\
&= \frac{\frac{\frac{(\hat{\beta}_1-\beta_1)}{\sigma}}{||\mathbf{x}-\bar{\mathbf{x}}||}}{\frac{1}{\sigma}\sqrt{\frac{W}{n-2}}} \\
&= \frac{\frac{\frac{(\hat{\beta}_1-\beta_1)}{\sigma}}{||\mathbf{x}-\bar{\mathbf{x}}||}}{\sqrt{\frac{1}{n-2}\frac{W}{\sigma^2}}} \sim t(n-2)
\end{aligned}
$$

Note that

$$
\begin{aligned}
t_*^2 &= \frac{\frac{(\hat{\beta}_1-\beta_1)^2||\mathbf{x}-\bar{\mathbf{x}}||^2}{\sigma^2}}{\frac{1}{n-2}\frac{W}{\sigma^2}} \\
&= \frac{\hat{\beta}_1^2||\mathbf{x}-\bar{\mathbf{x}}||^2}{\frac{||\mathbf{y}-\hat{\mathbf{y}}||^2}{n-2}} \sim F(\text{num } = 1, \text{denom} = n-2)
\end{aligned}
$$

INFERENCES CONCERNING LINEAR FUNCTIONS OF MODEL PARAMETERS: SIMPLE LINEAR REGRESSION

Function of model parameters (θ)

Suppose that we wish to make an inference about the linear function

$$\theta = a_0\beta_0 + a_1\beta_1,$$

where a_0 and a_1 are constants.

Estimator ($\hat{\theta}$)

$$\hat{\theta} = a_0\hat{\beta}_0 + a_1\hat{\beta}_1.$$

$E(\hat{\theta})$

$$
\begin{aligned}
E(\hat{\theta}) &= E(a_0\hat{\beta}_0 + a_1\hat{\beta}_1) \\
&= a_0 E(\hat{\beta}_0) + a_1 E(\hat{\beta}_1) \\
&= a_0\beta_0 + a_1\beta_1 \\
&= \theta
\end{aligned}
$$

$$
\begin{aligned}
V(\hat{\theta}) &= V(a_0\hat{\beta}_0 + a_1\hat{\beta}_1) \\
&= a_0^2 V(\hat{\beta}_0) + a_1^2 V(\hat{\beta}_1) + 2a_0 a_1 cov(\hat{\beta}_0, \hat{\beta}_1) \\
&\quad \text{(by hw problem } cov(\hat{\beta}_0, \hat{\beta}_1) = 0\text{)} \\
&= a_0^2 V(\hat{\beta}_0) + a_1^2 V(\hat{\beta}_1) \\
&= a_0^2 \frac{\sigma^2}{n} + a_1^2 \frac{\sigma^2}{||\mathbf{x} - \bar{\mathbf{x}}||^2} \\
&= \left[\frac{a_0^2}{n} + \frac{a_1^2}{||\mathbf{x} - \bar{\mathbf{x}}||^2} \right] \sigma^2
\end{aligned}
$$

Distribution of $\hat{\theta}$

Recalling that $\hat{\beta}_0$ and $\hat{\beta}_1$ are Normally distributed, it is clear that $\hat{\theta}$ is Normally distributed with mean θ and standard deviation

$$\sigma_{\hat{\theta}} = \sigma \sqrt{\frac{a_0^2}{n} + \frac{a_1^2}{||\mathbf{x} - \bar{\mathbf{x}}||^2}}$$

$$\sigma_{\hat{\theta}} = \sigma \sqrt{\frac{a_0^2}{n} + \frac{a_1^2}{S_{xx}}}$$

A test for θ

To *test the hypothesis* $H_0 : \theta = \theta_0$, compute the *test statistic*

$$t^* = \frac{\hat{\theta} - \theta_0}{S\sqrt{\frac{a_0^2}{n} + \frac{a_1^2}{||\mathbf{x} - \bar{\mathbf{x}}||^2}}} = \frac{\hat{\theta} - \theta_0}{S\sqrt{\frac{a_0^2}{n} + \frac{a_1^2}{S_{xx}}}}$$

In terms of a variable T having the $t(n-2)$ distribution, the P-value for a test of H_0 against
$H_a : \mu > \mu_0$ is $P(T \geq t^*)$.
$H_a : \mu < \mu_0$ is $P(T \leq t^*)$.
$H_a : \mu \neq \mu_0$ is $2P(T \geq |t^*|)$.

Development (Numerator)

We know that *th$\hat{e}$ta* is Normally distributed with mean θ and standard deviation $\sigma_{\hat{\theta}}$, then

$$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$$

has a Normal distribution with mean 0 and standard deviation 1.

Development (Denominator)

We also know that $\frac{W}{\sigma^2} = \frac{\sum_{i=3}^{n} \langle \mathbf{y}, \mathbf{u}_i \rangle^2}{\sigma^2}$ has a $\chi^2(n-2)$ and that $S^2 = \frac{W}{n-2}$ is an unbiased estimator of σ^2.

Development (t^*)

$$
\begin{aligned}
t^* &= \frac{Z}{\sqrt{\frac{W}{(n-2)\sigma^2}}} \quad \text{(using definition of T distribution)} \\
&= \frac{\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}}}{\frac{1}{\sigma}\sqrt{\frac{W}{n-2}}} \\
&= \frac{\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}}}{\frac{S}{\sigma}} \\
&= \frac{(\hat{\theta}-\theta)\sigma}{S\sigma_{\hat{\theta}}} \\
&= \frac{(\hat{\theta}-\theta)\sigma}{S\sigma\sqrt{\frac{a_0^2}{n}+\frac{a_1^2}{S_{xx}}}} = \frac{\hat{\theta}-\theta}{S\sqrt{\frac{a_0^2}{n}+\frac{a_1^2}{S_{xx}}}}
\end{aligned}
$$

Confidence Interval for θ

A $100(1-\alpha)\%$ Confidence Interval for $\theta = a_0\beta_0 + a_1\beta_1$

$$\hat{\theta} \pm t_{\alpha/2} S \sqrt{\frac{a_0^2}{n} + \frac{a_1^2}{||\mathbf{x} - \bar{\mathbf{x}}||^2}}$$

$$\hat{\theta} \pm t_{\alpha/2} S \sqrt{\frac{a_0^2}{n} + \frac{a_1^2}{S_{xx}}}$$

where the tabulated $t_{\alpha/2}$ is based on $n-2$ df.

A $100(1-\alpha)\%$ Confidence Interval for $E(Y) = \beta_0 + \beta_1(x_* - \bar{x})$

$$\hat{\beta}_0 + \hat{\beta}_1(x_* - \bar{x}) \pm t_{\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x_* - \bar{x})^2}{||\mathbf{x} - \bar{\mathbf{x}}||^2}}$$

$$\hat{\beta}_0 + \hat{\beta}_1(x_* - \bar{x}) \pm t_{\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x_* - \bar{x})^2}{S_{xx}}}$$

where the tabulated $t_{\alpha/2}$ is based on $n-2$ df.

PREDICTING A PARTICULAR VALUE OF Y BY USING SIMPLE LINEAR REGRESSION

Confidence Interval for Y when $x = x_*$

A $100(1-\alpha)\%$ **Prediction** Interval for Y when $x = x_*$

$$\hat{\beta}_0 + \hat{\beta}_1(x_* - \bar{x}) \pm t_{\alpha/2} S \sqrt{1 + \frac{1}{n} + \frac{(x_* - \bar{x})^2}{||\mathbf{x} - \bar{\mathbf{x}}||^2}}$$

$$\hat{\beta}_0 + \hat{\beta}_1(x_* - \bar{x}) \pm t_{\alpha/2} S \sqrt{1 + \frac{1}{n} + \frac{(x_* - \bar{x})^2}{S_{xx}}}$$

where the tabulated $t_{\alpha/2}$ is based on $n-2$ df.

Development

Note that $Y_* = \hat{Y}_* + (Y_* - \hat{Y}_*) = \text{prediction} + \text{error}$. Then,

$$\begin{aligned} Var(Y_*) &= Var(\text{prediction} + \text{error}) \\ &= Var(\text{prediction}) + Var(\text{error}) + 2Cov(\text{prediction}, \text{error}) \end{aligned}$$

It can be shown that $Cov(\text{prediction}, \text{error}) = -Var(\text{prediction})$.
Therefore,

$$Var(Y_*) = Var(\text{error}) - Var(\text{prediction})$$

$$Var(Y_*) + Var(\text{prediction}) = Var(\text{error})$$

$Cov(\text{prediction}, \text{error})$

We will show that

$$Cov(\text{prediction}, \text{error}) = -Var(\text{prediction})$$

Cov(prediction, error)

$$
\begin{aligned}
cov[U, V-U] &= E\{[U-\mu_U][(V-U)-(\mu_V-\mu_U)]\} \\
&= E\{[U-\mu_U][(V-\mu_V)-(U-\mu_U)]\} \\
&= E\{[U-\mu_U][(V-\mu_V)]-[U-\mu_U]^2\} \\
&= cov(U, V) - Var(U)
\end{aligned}
$$

Let $U = \hat{Y}_*$ and $V = Y_*$, then

$$cov(\hat{Y}_*, Y_* - \hat{Y}_*) = cov(\hat{Y}_*, Y_*) - Var(\hat{Y}_*) = -Var(\hat{Y}_*)$$

(recall that $\hat{Y}_*$ and Y_* are independent).

$Var(\text{error})$

We will show that

$$Var(\text{error}) = \left[1 + \frac{1}{n} + \frac{(x_* - \bar{x})^2}{||\mathbf{x} - \bar{\mathbf{x}}||^2}\right] \sigma^2$$

$Var(\text{error})$

$$
\begin{aligned}
Var(\text{error}) &= V\left[Y_* - \hat{Y}_*\right] \\
&= V[Y_*] + V[\hat{Y}_*] - 2cov[Y_*, \hat{Y}_*] \\
&= V[Y_*] + V[\hat{Y}_*] \;\; (Y_* \text{ and } \hat{Y}_* \text{ are independent}) \\
&= \sigma^2 + \left[\frac{1}{n} + \frac{(x_* - \bar{x})^2}{||\mathbf{x} - \bar{\mathbf{x}}||^2}\right]\sigma^2 \\
&= \left[1 + \frac{1}{n} + \frac{(x_* - \bar{x})^2}{||\mathbf{x} - \bar{\mathbf{x}}||^2}\right]\sigma^2
\end{aligned}
$$

CORRELATION

Definition

The correlation coefficient, r, is defined as the cosine of the angle between the vectors $\mathbf{y} - \bar{\mathbf{y}}$ and $\mathbf{x} - \bar{\mathbf{x}}$.

$$r = cos\theta = \frac{\langle \mathbf{x} - \bar{\mathbf{x}}, \mathbf{y} - \bar{\mathbf{y}} \rangle}{||\mathbf{x} - \bar{\mathbf{x}}|| \, ||\mathbf{y} - \bar{\mathbf{y}}||}$$

Note that

$$\langle \mathbf{x} - \bar{\mathbf{x}}, \mathbf{y} - \bar{\mathbf{y}} \rangle = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = S_{XY}$$

$$||\mathbf{x} - \bar{\mathbf{x}}|| = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \sqrt{S_{XX}}$$

$$||\mathbf{y} - \bar{\mathbf{y}}|| = \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \sqrt{S_{YY}}$$

Therefore,

$$cos\theta = \frac{S_{XY}}{\sqrt{S_{XX}}\sqrt{S_{YY}}}$$

We think of r as an estimator of the true correlation coefficient ρ, where

$$\rho = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sqrt{E[(X - \mu_X)^2]}\sqrt{E[(Y - \mu_Y)^2]}} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

The square of the correlation coefficient, r^2, is known as the coefficient of determination; r^2 is the proportion of the total corrected sum of squares explained by the regression.

By rewriting the regression sum of squares as

$$||\mathbf{y} - \bar{\mathbf{y}}||^2 = r^2||\mathbf{y} - \bar{\mathbf{y}}||^2 + (1 - r^2)||\mathbf{y} - \bar{\mathbf{y}}||^2.$$

Therefore, the test statistic used for testing $H_0 : \beta_1 = 0$ can be written in terms of r^2 as follows:

$$F_* = \frac{r^2||\mathbf{y} - \bar{\mathbf{y}}||^2}{(1 - r^2)||\mathbf{y} - \bar{\mathbf{y}}||^2/(n-2)} = \frac{(n-2)r^2}{1 - r^2}$$

Exercise

Archaeopterix is an extinct beast having feathers like a bird but with teeth and a long bony tail like a reptile. Here are the lengths in centimeters of the femur (a leg bone) and the humerus (a bone in the upper arm) for the five fossil specimens that preserve both bones:

Femur	38	56	59	64	74
Humerus	41	63	70	72	84

Test the hypothesis $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$. Use femur length as the explanatory variable and $\alpha = 0.05$.

Summary statistics

$\bar{x} = 58.2$
$\bar{y} = 66$
$\sum(x_i - \bar{x})^2 = 696.8$
$\sum(y_i - \bar{y})^2 = 1010$
$\sum(x_i - \bar{x})(y_i - \bar{y}) = 834$
$\sum(x_i - \bar{x})(y_i) = 834$

Calculations (exercise)

Step 1. Find $\bar{x}$, $\mathbf{x} - \bar{\mathbf{x}}$, and $||\mathbf{x} - \bar{\mathbf{x}}||^2$.
$\bar{x} = 58.2$

$$\mathbf{x} - \bar{\mathbf{x}} = \begin{pmatrix} -20.2 \\ -2.2 \\ 0.8 \\ 5.8 \\ 15.8 \end{pmatrix}$$

$||\mathbf{x} - \bar{\mathbf{x}}||^2 = 696.8$

Calculations (exercise)

Step 2. Find $\bar{y}$, $\mathbf{y} - \bar{\mathbf{y}}$, and $||\mathbf{y} - \bar{\mathbf{y}}||^2$.
$\bar{y} = 66$

$$\mathbf{y} - \bar{\mathbf{y}} = \begin{pmatrix} -25 \\ -3 \\ 4 \\ 6 \\ 18 \end{pmatrix}$$

$||\mathbf{y} - \bar{\mathbf{y}}||^2 = 1010$

Step 3. Find $\hat{\beta}_1$ and $\hat{\beta}_1^2$.
Recall that
$\hat{\beta}_1 = \frac{\langle \mathbf{y}, \mathbf{x} - \bar{\mathbf{x}} \rangle}{||\mathbf{x} - \bar{\mathbf{x}}||^2} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}$
$\hat{\beta}_1 = \frac{834}{696.8} = 1.1969001$
$\hat{\beta}_1^2 = 0.51 = 1.4325699$

Step 4. Use ANOVA table.

Method 1. ANOVA Table

Using our four-step procedure, you should be able to construct the following table:

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x          1 998.21  998.21   254.1 0.0005368 ***
## Residuals  3  11.79    3.93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

Method 2. T test

Using our t test, you should get something like this:

```
##              Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) -3.659587 4.45896232 -0.8207261 0.4719439905
## x            1.196900 0.07508543 15.9405098 0.0005368404
```

Method 3. Using correlation coefficient, r

Using our t test, you should get something like this:

```
##    value    numdf    dendf
## 254.0999   1.0000   3.0000
```

R CODE

Example

One effect of global warming is to increase the flow of water into the Arctic Ocean from rivers. Such an increase might have major effects on the world's climate. Six rivers (Yenisey, Lena, Ob, Pechora, Kolyma, and Severnaya Dvina) drain two-thirds of the Arctic in Europe and Asia. Several of these are among the largest rivers on earth. File arctic-rivers.txt contains the total discharge from these rivers each year from 1936 to 1999[2]. Discharge is measured in cubic kilometers of water.

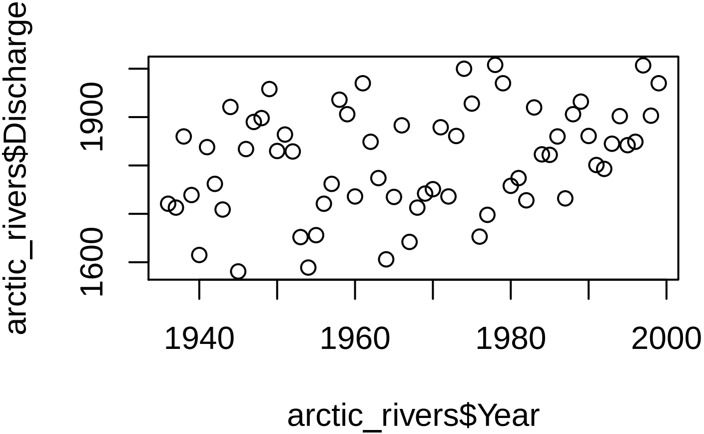
Reading our data

```
# url of arctic rivers data;

riv_url = "http://www.math.unm.edu/~alvaro/arctic-rivers.txt"

# import data in R;

arctic_rivers = read.table(riv_url, header = TRUE);
```

Scatterplot (R code)

```
plot(arctic_rivers$Year,arctic_rivers$Discharge);
```

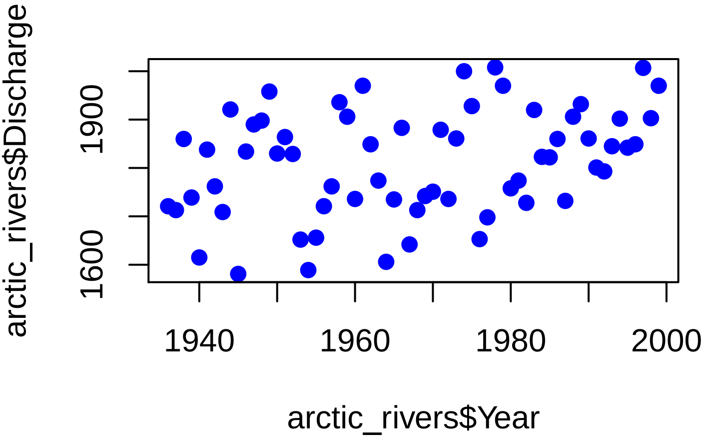
Scatterplot (R code)



Scatterplot (R code)

```
plot(arctic_rivers$Year,arctic_rivers$Discharge,
pch=19,col="blue");
```

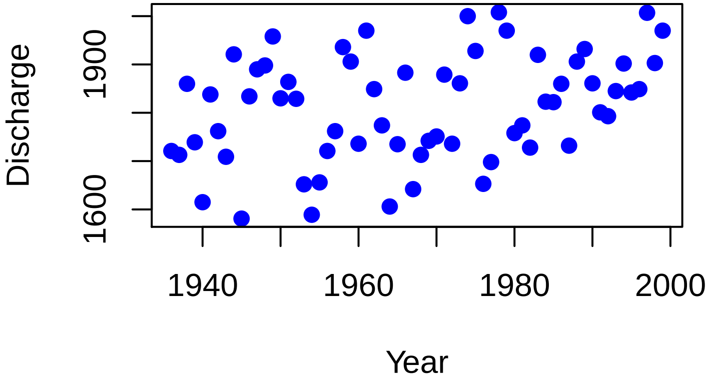
Scatterplot (R code)



Scatterplot (R code)

```
plot(arctic_rivers$Year,arctic_rivers$Discharge,
pch=19,col="blue", xlab="Year",
ylab="Discharge");
```

Scatterplot (R code)



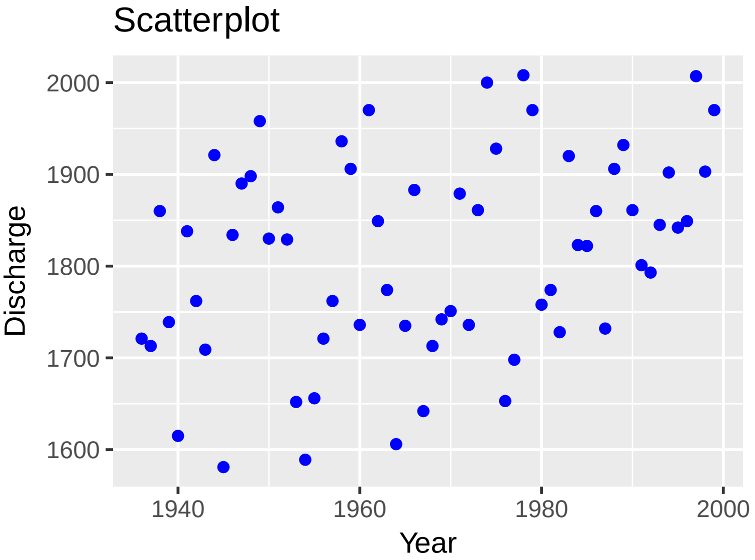
Scatterplot (using ggplot2)

```
install.packages("ggplot2", dependencies=TRUE);
# R will download ggplot2;

library(ggplot2);
# R will load ggplot2;
```

Scatterplot (R code)

```
plot=qplot(arctic_rivers$Year,arctic_rivers$Discharge,
colour=I("blue"),xlab="Year",ylab="Discharge",
main="Scatterplot");

plot;
```

Scatterplot (R code)



The scatterplot shows a weak positive, linear relationship.

R code

```
explanatory=arctic_rivers$Year;

response=arctic_rivers$Discharge

rivers.reg=lm(response~explanatory);
```

R code

```
names(rivers.reg);

##  [1] "coefficients"  "residuals"      "effects"        "rank'
##  [5] "fitted.values" "assign"         "qr"             "df.re
##  [9] "xlevels"       "call"           "terms"          "model
```

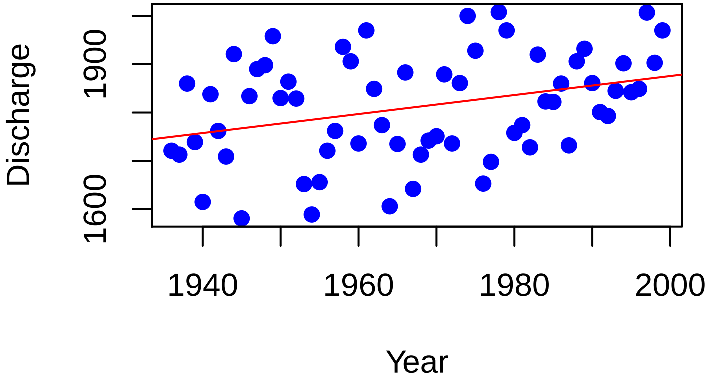
$\hat{\beta}_0$ and $\hat{\beta}_1$

```
rivers.reg$coef;

##  (Intercept)  explanatory
## -2056.769460     1.966163
```

Scatterplot with least-squares line

```
plot(explanatory,response,
pch=19,col="blue", xlab="Year",
ylab="Discharge");

abline(rivers.reg$coef, col="red");
```

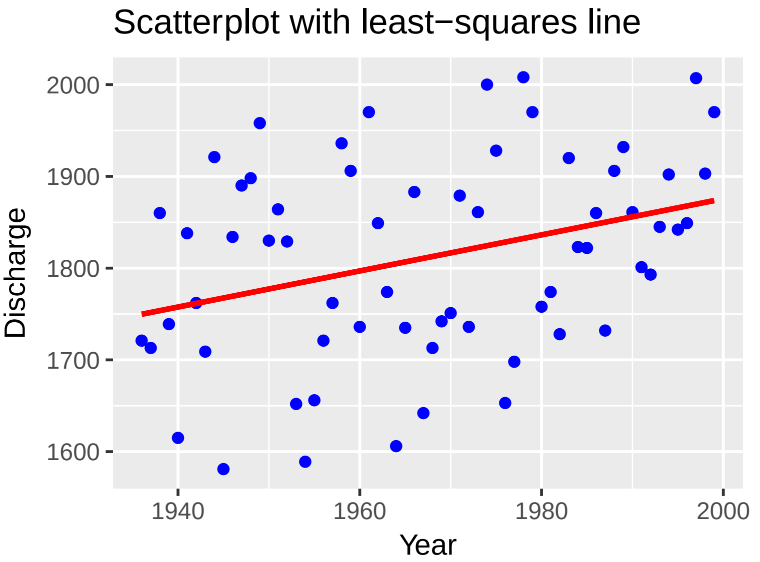
Scatterplot with least-squares line



Scatterplot with least-squares line (with ggplot2)

```
plot=qplot(arctic_rivers$Year,arctic_rivers$Discharge,
colour=I("blue"),xlab="Year",ylab="Discharge",
main="Scatterplot with least-squares line");

new.plot=plot+stat_smooth(method="lm",
se=FALSE,colour=I("red"));
new.plot;
```

Scatterplot with least-squares line (with ggplot2)



Residuals

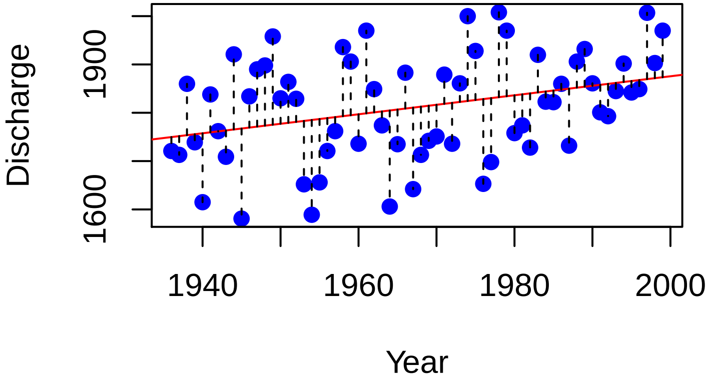
A **residual** is the difference between an observed value of the response variable and the value predicted by the regression line. That is,

$$residual = \text{observed } y - \text{predicted } y = y - \hat{y}.$$

Scatterplot with residual line segments

```
plot(explanatory,response,
pch=19,col="blue", xlab="Year",
ylab="Discharge");

abline(rivers.reg$coef, col="red");

segments(explanatory, fitted(rivers.reg),
explanatory,response, lty=2, col="black");
```

Scatterplot with residual line segments



Residual Plots

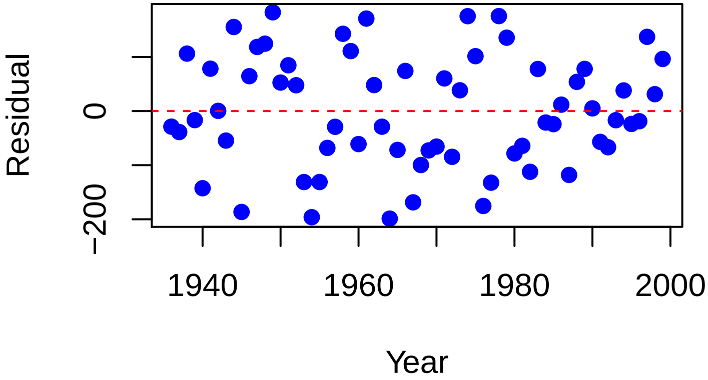
A **residual plot** is a scatterplot of the regression residuals against the explanatory variable. Residual plots help us assess the fit of a regression line.

A residual plot magnifies the deviations of the points from the line and makes it easier to see unusual observations and patterns.

Residual plot

```
plot(explanatory,resid(rivers.reg),
pch=19,col="blue", xlab="Year",
ylab="Residual");

abline(h=0, col="red",lty=2);
```

Residual plot



The Method of Least Squares

The Method of Least Squares can be illustrated simply by fitting a straight line to a set of data points. Suppose that we wish to fit the model

$$E(Y) = \beta_0 + \beta_1 x.$$

The least-squares procedure for fitting a line through a set of n data points is similar to the method that we might use if we fit a line by eye; that is, we want the differences between the observed values and corresponding points on the fitted line to be "small" according to some criterion. A convenient way to accomplish this is to minimize the sum of squares of the vertical deviations from the fitted line.

Thus, if

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

is the predicted value of the ith y value, then the deviation of the observed value y_i from $\hat{y}_i$ is the difference $y_i - \hat{y}_i$ and the sum of squares of deviations to be minimized is

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2.$$

The quantity SSE is also called the **sum of squares for error**.

If SSE possesses a minimum, it will occur for values of β_0 and β_1 that satisfy the equations, $\frac{\partial SSE}{\partial \hat{\beta}_0} = 0$ and $\frac{\partial SSE}{\partial \hat{\beta}_1} = 0$. These equations are called the **least-squares equations** for estimating the parameters of a line.

You can verify that the solutions are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

(Further, it can be shown that the simultaneous solution for the two least-squares equations yields values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize SSE. I leave this for you to prove).

Least-Squares Estimators for Simple Linear Regression Model

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}},$$

where $S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ and $S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})^2$.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}.$$

Estimating the Model Parameters in R

```
lin.reg=lm(y~x);
```

The general form of the command, in pseudo code, is
"the name you choose" = lm(response variable ~ explanatory variables).

Basic Inference for the Model

The following command will give you typical regression output:

```
summary(lin.reg);
```

Regression Standard Error

The **regression standard error** is

$$s = \sqrt{\frac{1}{n-2}\sum \text{residual}^2} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y})^2}.$$

Use s to estimate the **unknown** σ in the regression model.

Regression Standard Error

```
residuals<-resid(rivers.reg);

n<-length(residuals);

s<-sqrt(sum(residuals^2)/(n-2));

s;

## [1] 104.0026
```

Regression Standard Error (another way)

```
summary(rivers.reg)$sigma

## [1] 104.0026
```

Confidence intervals for the regression slope

A level C confidence interval for the slope β_1 of the true regression line is

$$\hat{\beta}_1 \pm t^* SE_{\hat{\beta}_1}.$$

In this formula, the standard error of the least-squares slope $\hat{\beta}_1$ is

$$SE_{\hat{\beta}_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}}$$

and t^* is the critical value for the $t(n-2)$ density curve with area C between $-t^*$ and t^*.

Example

We will use the data in arctic-rivers.txt to give a 90% confidence interval for the slope of the true regression of Arctic river discharge on year.

Confidence interval for slope

```
summary(rivers.reg)$coef

##                  Estimate    Std. Error   t value   Pr(>|t|)
## (Intercept) -2056.769460 1384.6873683 -1.485367 0.14251371
## explanatory     1.966163    0.7037491  2.793841 0.00692068
```

Confidence interval for slope

```
b<-summary(rivers.reg)$coef[2,1];

SEb<-summary(rivers.reg)$coef[2,2];

# lower bound;
b-qt(0.95,df=n-2)*SEb;

## [1] 0.7910398

# upper bound;
b+qt(0.95,df=n-2)*SEb;

## [1] 3.141286
```

R output gives $b = 1.966163$ and $SE_b = 0.7037491$. There were $n = 64$ observations, so $df = 62$. Our 90% Confidence Interval for β is given by (0.7910398, 3.1412862). Because this interval does not contain 0, we have evidence that β (the rate at which discharge is increasing) is positive.

Testing the hypothesis of no linear relationship

We can also test hypotheses about the slope β_1. The most common hypothesis is

$$H_0 : \beta_1 = 0.$$

A regression line with slope 0 is horizontal. That is, the mean of y does not change at all when x changes. So this H_0 says that there is no true linear relationship between x and y.

Significance test for regression slope

To test the hypothesis $H_0 : \beta_1 = 0$, compute the t statistic

$$t = \frac{\hat{\beta}_1}{SE_{\hat{\beta}_1}}.$$

In terms of a random variable T having the $t(n-2)$ distribution, the P-value for a test of H_0 against $H_a : \beta \neq 0$ is $2P(T \geq |t|)$.

Our example

The most important question we ask of the data in arctic-rivers.txt is this: Is the increasing trend visible in your plot statistically significant? If so, changes in the Arctic may already be affecting earth's climate. Use R to answer this question.

```
t.statistic<-b/SEb;

t.statistic;

## [1] 2.793841

p.value<-2*(1-pt(t.statistic,n-2));

p.value;

## [1] 0.00692068
```

```
summary(rivers.reg)$coef

##                  Estimate    Std. Error   t value    Pr(>|t|)
## (Intercept) -2056.769460 1384.6873683 -1.485367 0.14251371
## explanatory     1.966163    0.7037491  2.793841 0.00692068
```

The t statistic for testing $H_0 : \beta = 0$ is therefore $t = 2.7938409$. This has $df = 62$; R gives a P-value of 0.0069207. There is significant evidence (at $\alpha = 0.01$ significance level) that β is nonzero.

Another way (ANOVA table)

```
anova(rivers.reg)
```

Another way (ANOVA table)

```
## Analysis of Variance Table
##
## Response: response
##             Df Sum Sq Mean Sq F value   Pr(>F)
## explanatory  1  84429   84429  7.8055 0.006921 **
## Residuals   62 670625   10817
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```