9 – Linear Models

A major part of inferential statistics is to create *models* that represent relationships between variables. Models can be used to predict outcomes where we lack data.

Example A *linear* statistical model might be used to predict the **Pulse** rate of a student based on their **Age**, **Height**, and **Sex**.

	Sex	Wr.Hnd	NW.Hnd	W.Hnd	Fold	Pulse	Clap	Exer	Smoke	Height	M.I	Age
1	Female	18.5	18.0	Right	R on L	92	Left	Some	Never	173.00	Metric	18.250
2	Male	19.5	20.5	Left	R on L	104	Left	None	Regul	177.80	Imperial	17.583
3	Male	18.0	13.3	Right	L on R	87	Neither	None	Occas	NA	NA	16.917
4	Male	18.8	18.9	Right	R on L	NA	Neither	None	Never	160.00	Metric	20.333
5	Male	20.0	20.0	Right	Neither	35	Right	Some	Never	165.00	Metric	23.667
6	Female	18.0	17.7	Right	L on R	64	Right	Some	Never	172.72	Imperial	21.000
7	Male	17.7	17.7	Right	L on R	83	Right	Freq	Never	182.88	Imperial	18.833
…												
235	Female	17.5	16.5	Right	R on L	NA	Right	Some	Never	170.00	Metric	18.583
236	Male	21.0	21.5	Right	R on L	90	Right	Some	Never	183.00	Metric	17.167
237	Female	17.6	17.3	Right	R on L	85	Right	Freq	Never	168.50	Metric	17.750

Age = 21
Height = 170 → predicted **Pulse** =
Sex = Female

9 – 线性模型

推断统计的一个重要部分是建立模型，以表示变量之间的关系。这些模型可用于在缺乏数据的情况下预测结果。

示例 线性统计模型可用于根据学生的**脉搏**率、**年龄**、**身高**和**性别**进行预测。

	Sex	Wr.Hnd	NW.Hnd	W.Hnd	Fold	Pulse	Clap	Exer	Smoke	Height	M.I	Age
1	Female	18.5	18.0	Right	R on L	92	Left	Some	Never	173.00	Metric	18.250
2	Male	19.5	20.5	Left	R on L	104	Left	None	Regul	177.80	Imperial	17.583
3	Male	18.0	13.3	Right	L on R	87	Neither	None	Occas	NA	NA	16.917
4	Male	18.8	18.9	Right	R on L	NA	Neither	None	Never	160.00	Metric	20.333
5	Male	20.0	20.0	Right	Neither	35	Right	Some	Never	165.00	Metric	23.667
6	Female	18.0	17.7	Right	L on R	64	Right	Some	Never	172.72	Imperial	21.000
7	Male	17.7	17.7	Right	L on R	83	Right	Freq	Never	182.88	Imperial	18.833
…												
235	Female	17.5	16.5	Right	R on L	NA	Right	Some	Never	170.00	Metric	18.583
236	Male	21.0	21.5	Right	R on L	90	Right	Some	Never	183.00	Metric	17.167
237	Female	17.6	17.3	Right	R on L	85	Right	Freq	Never	168.50	Metric	17.750

年龄 = 21**身**
高 = 170 → 预测的**脉搏** =
性别 = 女性

Linear Correlation Coefficient

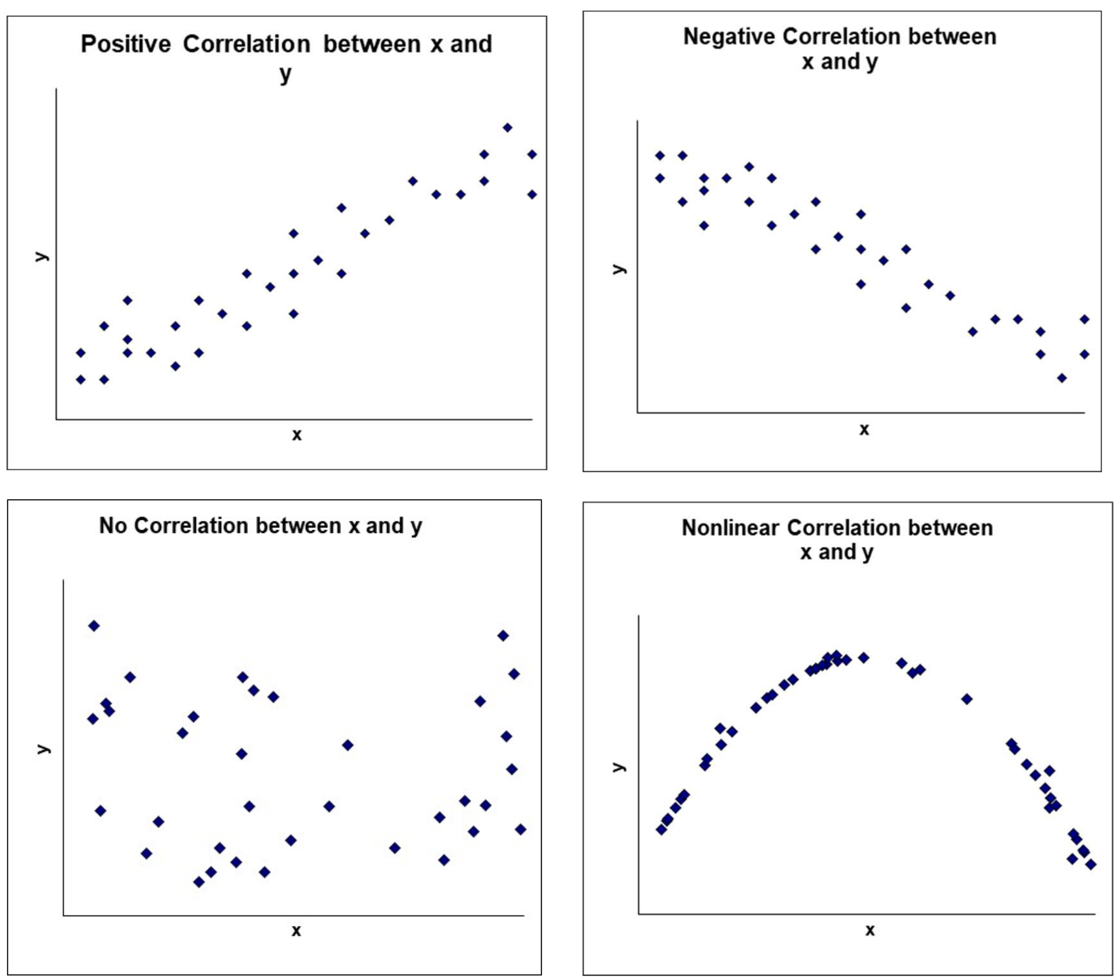
If X and Y are paired numerical variables with data $X = x_1, x_2, \ldots, x_n$ and $Y = y_1, y_2, \ldots, y_n$, then we define the *linear correlation coefficient*:

Pearson's Correlation Coefficient

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \cdot \sqrt{n(\sum y^2) - (\sum y)^2}}$$

or using R:

```
r <- cor(X, Y)
```



线性相关系数

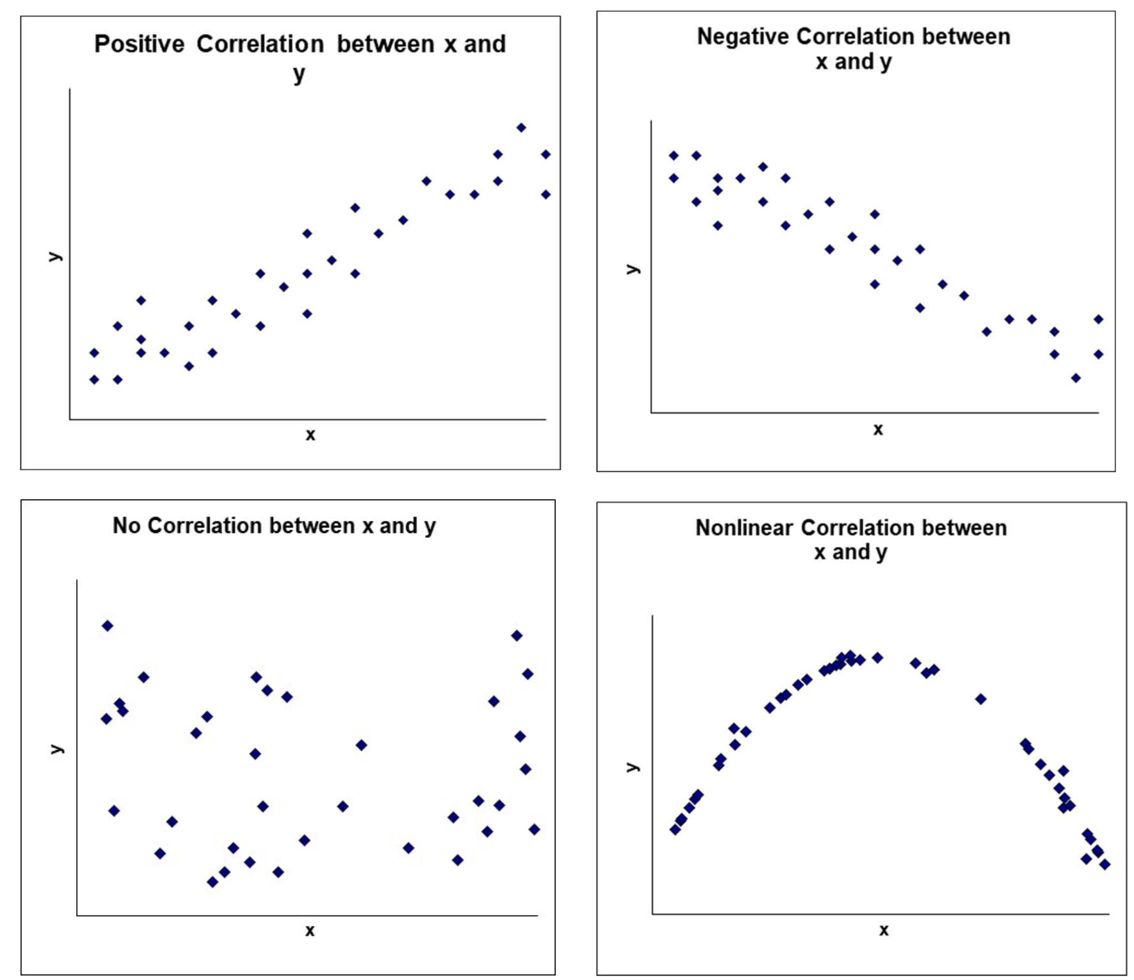
如果X 和Y 是一对数值变量，其数据为$X = x_1, x_2, \ldots, x_n$ 和$Y = y_1, y_2, \ldots, y_n$，那么我们定义线性相关系数*ff*:

皮尔逊相关系数

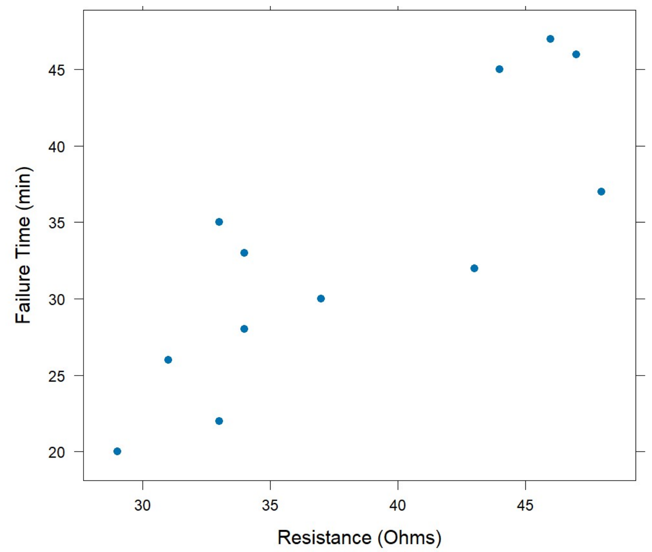
$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \cdot \sqrt{n(\sum y^2) - (\sum y)^2}}$$

或使用R：

```
r <- cor(X, Y)
```



Example (Electric Circuit) Suppose we make $n = 12$ measurements of an electric circuit's Resistance (in Ohms) and Fail.Time (in minutes).

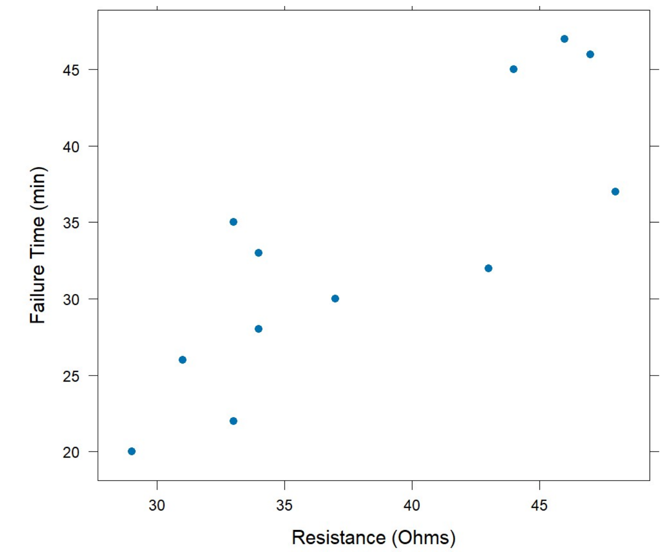


	Resistance	Fail.time
1	43	32
2	29	20
3	44	45
4	33	35
5	33	22
6	47	46
7	34	28
8	31	26
9	48	37
10	34	33
11	46	47
12	37	30

Let's calculate r using the data for **Resistance** and **Fail.Time**.

	Resistance	**Fail.Time**			
	x	y	xy	x^2	y^2
	43	32	1376	1849	1024
	29	20	580	841	400
	44	45	1980	1936	2025
	33	35	1155	1089	1225
	33	22	726	1089	484
	47	46	2162	2209	2116
	34	28	952	1156	784
	31	26	806	961	676
	48	37	1776	2304	1369
	34	33	1122	1156	1089
	46	47	2162	2116	2209
	37	30	1110	1369	900
Total	459	401	15907	18075	14301

示例（电路）假设我们对一个电路的电阻（单位：欧姆）和失效时间（单位：分钟）进行了 $n = 12$ 次测量。



	Resistance	Fail.time
1	43	32
2	29	20
3	44	45
4	33	35
5	33	22
6	47	46
7	34	28
8	31	26
9	48	37
10	34	33
11	46	47
12	37	30

让我们使用**电阻**和**失效时间**的数据来计算r。

	电阻 失效时间				
	x	y	xy	x^2	y^2
	43	32	1376	1849	1024
	29	20	580	841	400
	44	45	1980	1936	2025
	33	35	1155	1089	1225
	33	22	726	1089	484
	47	46	2162	2209	2116
	34	28	952	1156	784
	31	26	806	961	676
	48	37	1776	2304	1369
	34	33	1122	1156	1089
	46	47	2162	2116	2209
	37	30	1110	1369	900
总计	459	401	15907	18075	14301

Properties of the Linear Correlation Coefficient, r

Suppose X and Y are two numerical variables and we calculate r for a random sample.

1. The possible values of r are:
$$-1 \leq r \leq 1$$
2. The value of r does not change if X or Y are expressed in different *units*.
3. Swapping X and Y does not affect r.
4. Linear correlation coefficient r only measures the strength of *linear* relationships. It will not indicate the existence of non-linear relationships.
5. We use r to denote linear correlation for a *sample*.
We use ρ to denote linear correlation of a *population*.

$r = 1$	$\Rightarrow$ perfect positive linear correlation
r "close" to 1	$\Rightarrow$ strong positive linear correlation
$r = 0$	$\Rightarrow$ no linear correlation
r "close" to -1	$\Rightarrow$ strong negative linear correlation
$r = -1$	$\Rightarrow$ perfect negative linear correlation

IMPORTANT NOTE:

If X and Y are correlated ($r \neq 0$), it does not mean that higher values of X *cause* higher values of Y.

"Correlation does not imply causation"

Example The amount of ice cream consumed (X) and number of boating accidents (Y) each day are positively correlated. But eating ice cream does not cause boating accidents!

线性相关系数 r 的性质

假设有两个数值变量 X 和 Y，我们计算一个随机样本的 r。

1. r 的可能取值为：
$$-1 \leq r \leq 1$$
2. 如果 X 或 Y 以不同的 单位 表示，则 r 的值不会改变。
3. 交换 X 和 Y 不会影响 r。
4. 线性相关系数 r 仅用于衡量 线性 关系的强度，它不会反映非线性关系的存在。
5. 我们用 r 表示 样本 的线性相关性。
我们用 ρ 表示 总体 的线性相关性。

$r = 1$	$\Rightarrow$ 完全正线性相关
r “接近” 1	$\Rightarrow$ 强正线性相关
$r = 0$	$\Rightarrow$ 无线性相关
r “接近” -1	$\Rightarrow$ 强负线性相关
$r = -1$	$\Rightarrow$ 完全负线性相关

重要提示：

如果 X 和 Y 相关（$r \neq 0$），并不意味着 X 的较高值会 导致 Y 的较高值。

“相关性并不意味着因果关系”

示例 每天消耗的冰淇淋数量 (X) 与划船事故数量 (Y) 呈正相关。但吃冰淇淋并不会导致划船事故！

Hypothesis Testing for ρ

When we calculated $r = 0.8324$ for the **Resistance** and **Fail.Time** measurements, we were *certain* that the *sample* measurements were positively correlated.

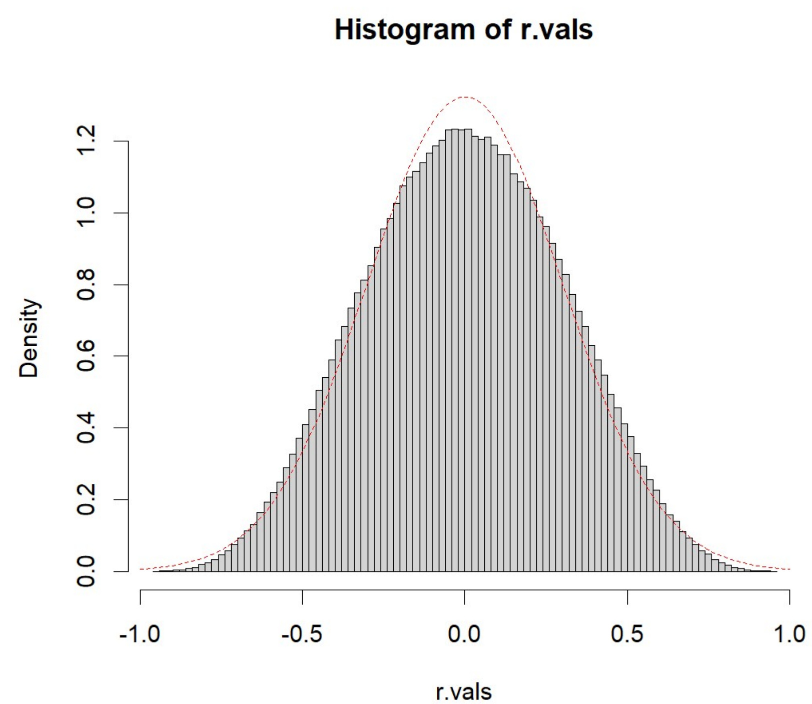
Can we conclude that **Resistance** and **Fail.Time** are correlated at the *population* level?

If they are not, we would have to say that $r = 0.8324$ occurred *by random chance*.

Sampling Distribution of r

Suppose X and Y are two normal variables that are *un*correlated ($\rho = 0$) and follow a *bivariate normal* distribution (more on this below).

If we take random samples of size n and calculate the correlation coefficient r, then r follows a distribution that is *roughly* normal.



In fact, it turns out that $t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$ follows a Student t-distribution with df $= n-2$.

[This is because $\sqrt{\frac{1-r^2}{n-2}}$ is an unbiased estimator of σ_r, the standard deviation of r.]

This allows us to perform a *hypothesis test* for the claim that $\rho \neq 0$.

ρ的假设检验

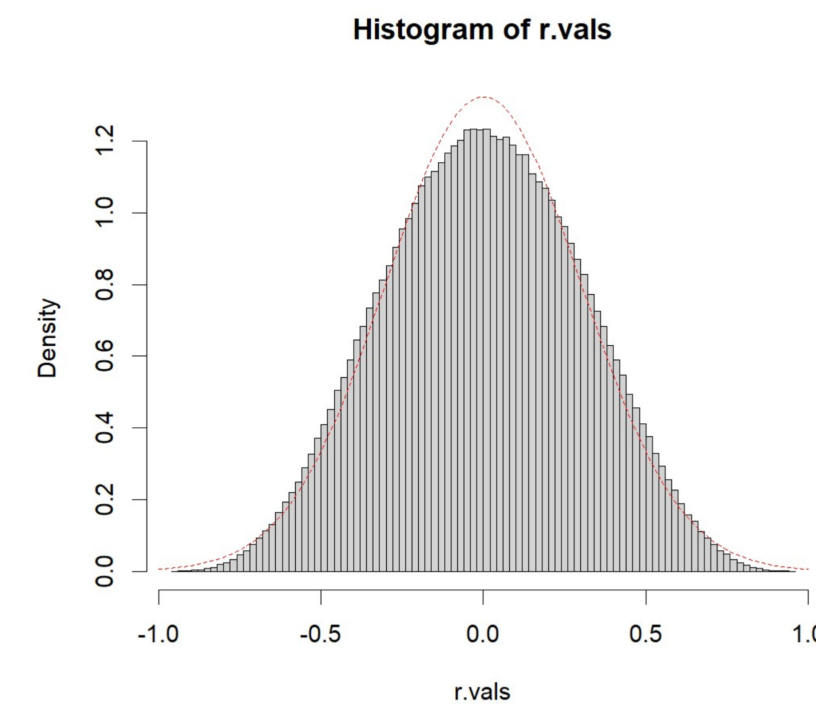
当我们计算 **电阻** 和 **失效时间** 测量值的 $r = 0.8324$ 时，我们确信 样本 测量值呈正相关。

我们能否得出结论：**电阻**和**失效时间**在总体层面上存在相关性？如果它们不存在相关性，我们就必须认为$r = 0.8324$ 是由随机因素导致的。

r的抽样分布

假设X 和Y 是两个不相关的正常变量，并且服从二元正态分布（下文将进一步说明）。

如果我们取大小为n 的随机样本并计算相关系数ff$r$，则r服从一个大致正态的分布。



事实上，$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$ 服从自由度为 df $= n-2$ 的学生 t 分布。

[这是因为是σ_r的$\sqrt{\frac{1-r^2}{n-2}}$无偏估计量，即r的标准差。]

这使我们能够对以下主张进行假设检验：tha t $\rho \neq 0$.

Example Using the sample statistics for **Resistance** and **Fail.Time** ($n = 12, r = 0.8324$), test the claim that $\rho \neq 0$.

1. (Claim): The claim is that $\rho \neq 0$.

2. (Hypotheses):

3. (Test statistic): $t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} =$

4. (P-value):

5. (Decision):

6. (Conclusion):

示例 使用**电阻** 和**失效时间**的样本统计量 ($n = 12, r = 0.8324$)，检验关于 $\rho \neq 0$ 的主张。

1. （声明）：该声明为 $\rho \neq 0$。

2. （假设）：

3. （检验统计量）：$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} =$

4. （P 值）：

5. （决策）：

6. （结论）：

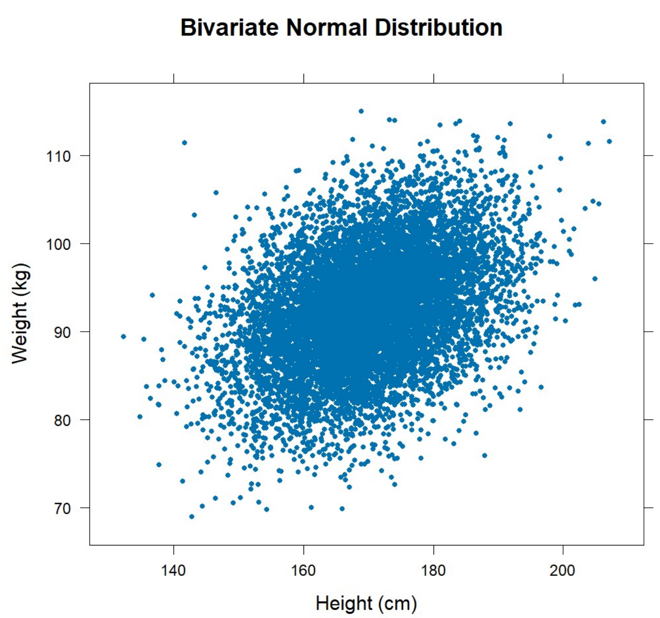
Assumptions for Hypothesis Test about ρ

Note that using the test statistic

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

is only correct under the assumption that X and Y are normally distributed for *each specific value* of the other variable. This is called a *bivariate normal* distribution.

Example Shown here is a scatter plot of variables X (Height) and Y (Weight) that follow a bivariate normal distribution.



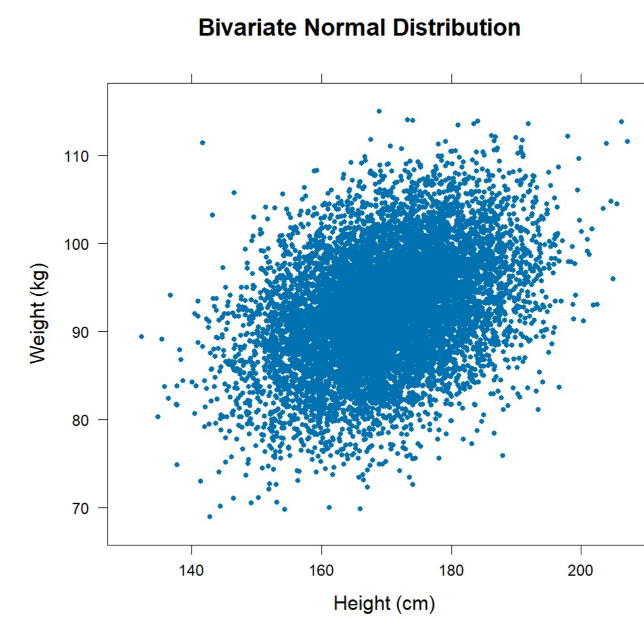
关于ρ的假设检验的前提假设

请注意使用该检验统计量

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

只有在假设X 和Y 对于另一变量的每一个特定取值都呈正态分布时才成立。这称为双变量正态分布。

示例 此处显示的是变量 X （身高）和 Y （体重）的散点图，它们服从二元正态分布。



124

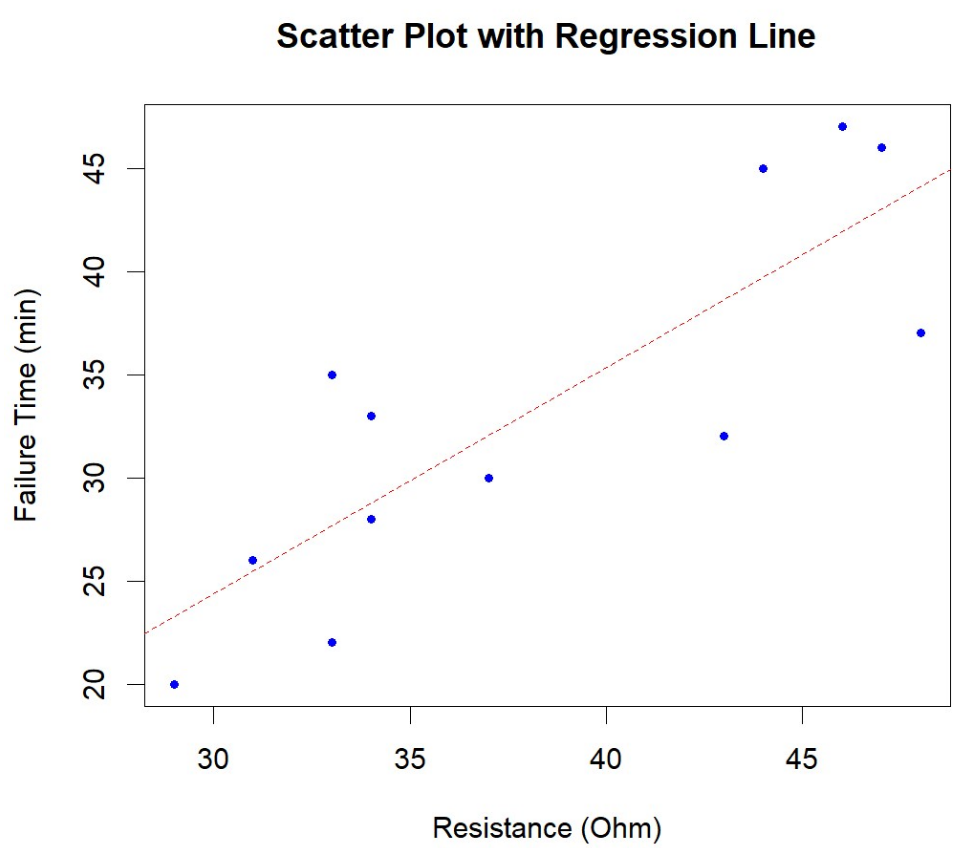
Linear Regression

Once we have determined that two variables X and Y are correlated at the population level, (i.e., they have a significant linear relationship), we typically describe that relationship using a *regression line* (i.e., "best fitting line").

In this context, we say:

X is the *independent* or *predictor* variable

Y is the *dependent* or *response* variable



We will write this regression line in the form

$$\hat{y} = a + bx$$

where $\hat{y}$ is the *predicted* Y value for a given value $X = x$. The coefficients are:

$a = y$-intercept of the line (α at population level)

$b =$ slope of the line (β at population level)

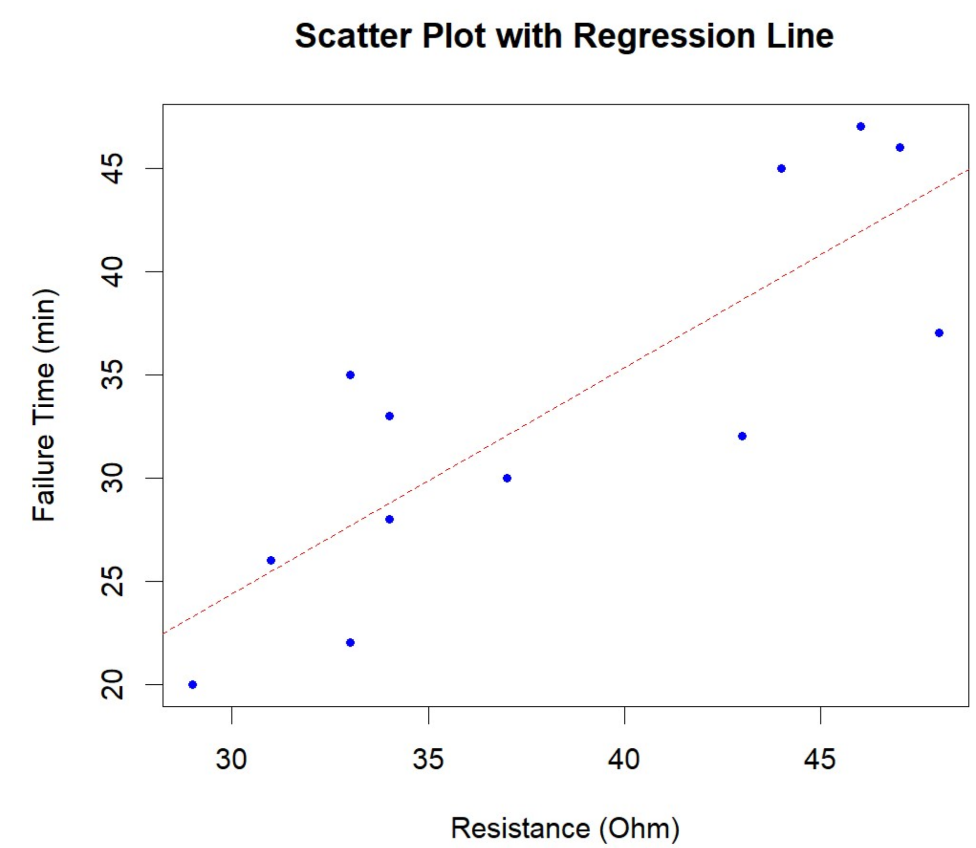
线性回归

一旦我们确定两个变量 X 和 Y 在总体层面上存在相关性（即它们具有显著的线性关系），通常会使用 回归直线 （即“最佳拟合直线”）来描述这种关系。

在此背景下，我们称：

X 是自变量或预测变量

Y 是因变量或响应变量



我们将以如下形式写出这条回归线

$$\hat{y} = a + bx$$

其中 $\hat{y}$ 是对于给定值 $X = x$ 的 预测 Y 值。系数为：

直线的$a = y$截距 （在总体水平上的α）

直线的$b =$ 斜率 （在总体水平上的β）

Linear Regression Coefficients: $\hat{y} = a + bx$

Using methods from calculus (or from linear algebra), it is possible to derive the formulas

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \qquad = r \cdot \frac{s_Y}{s_X}$$

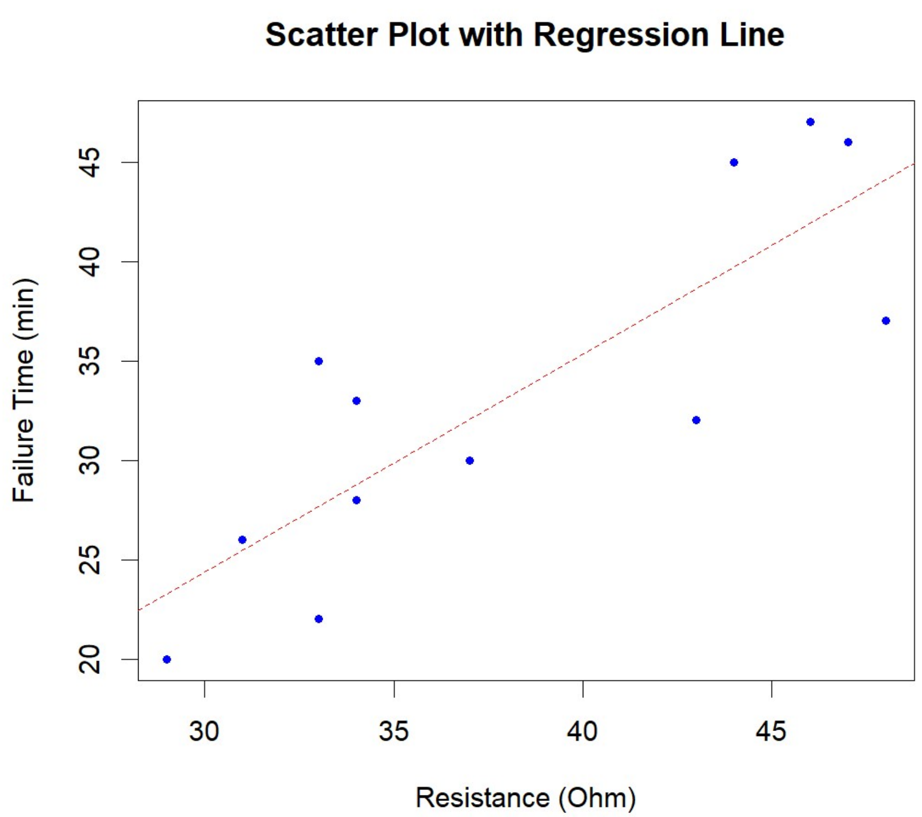
$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \qquad = \bar{Y} - b \cdot \bar{X}$$

where

s_X = sample std. dev. of X

s_Y = sample std. dev. of Y

The resulting line, $\hat{y} = a + bx$ is sometimes called the "least-squares" regression line, because it creates the least possible sum of the vertical errors (squared).



线性回归系数：ff$\hat{y} = a + bx$

利用微积分（或线性代数）的方法，可以推导出以下公式

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \qquad = r \cdot \frac{s_Y}{s_X}$$

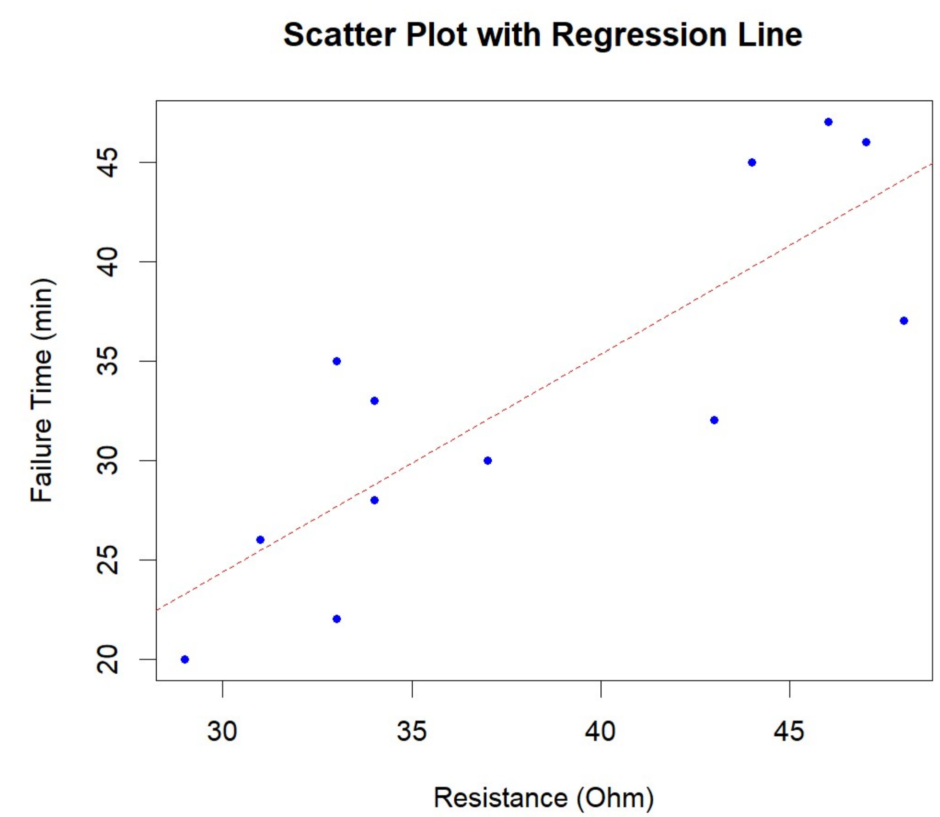
$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \qquad = \bar{Y} - b \cdot \bar{X}$$

其中

s_X = sample std. dev. of X

s_Y = sample std. dev. of Y

所得直线 $\hat{y} = a + bx$ 有时被称为"最小二乘"回归直线，因为它使得垂直误差的平方和达到最小。



Example Use the sample data for **Resistance** and **Fail.time** to find the regression line.

	Resistance	Fail.Time			
	x	y	xy	x^2	y^2
	43	32	1376	1849	1024
	29	20	580	841	400
	44	45	1980	1936	2025
	33	35	1155	1089	1225
	33	22	726	1089	484
	47	46	2162	2209	2116
	34	28	952	1156	784
	31	26	806	961	676
	48	37	1776	2304	1369
	34	33	1122	1156	1089
	46	47	2162	2116	2209
	37	30	1110	1369	900
Total	459	401	15907	18075	14301

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} =$$

$$r \cdot \frac{s_Y}{s_X} =$$

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$\bar{Y} - b \cdot \bar{X} =$$

示例 使用 **电阻** 和 **失效时间** 的样本数据来求回归直线。

	电阻 失效时间				
	x	y	xy	x^2	y^2
	43	32	1376	1849	1024
	29	20	580	841	400
	44	45	1980	1936	2025
	33	35	1155	1089	1225
	33	22	726	1089	484
	47	46	2162	2209	2116
	34	28	952	1156	784
	31	26	806	961	676
	48	37	1776	2304	1369
	34	33	1122	1156	1089
	46	47	2162	2116	2209
	37	30	1110	1369	900
总计	459	401	15907	18075	14301

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} =$$

$$r \cdot \frac{s_Y}{s_X} =$$

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

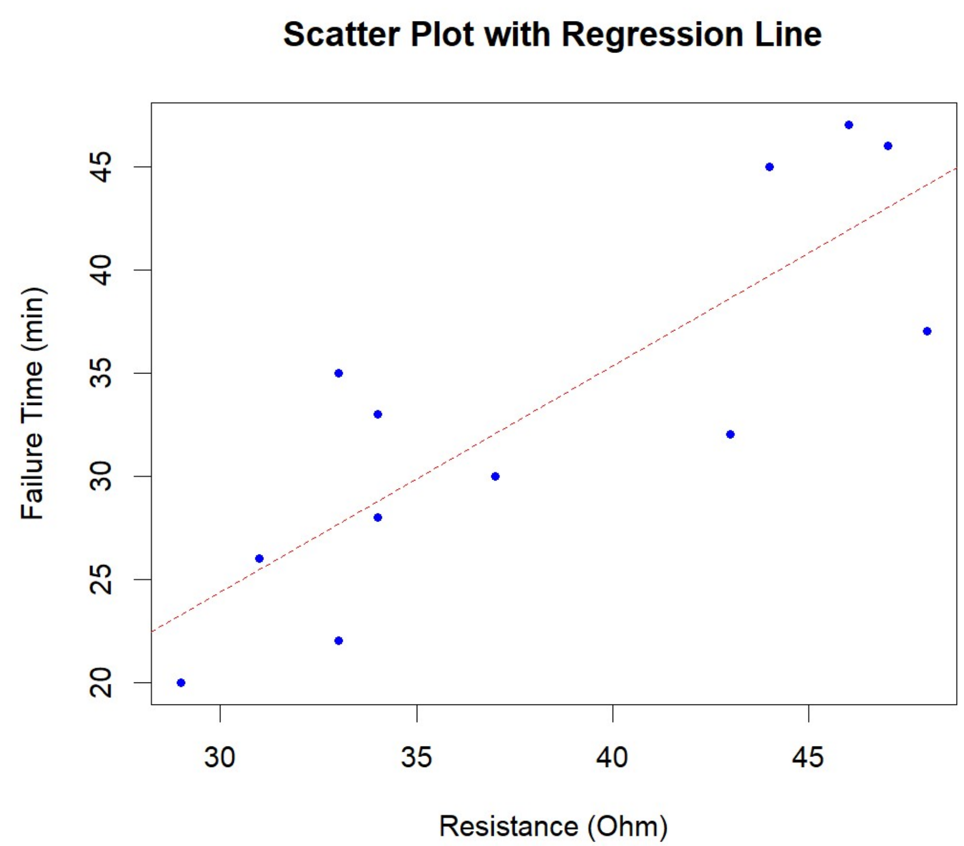
$$\bar{Y} - b \cdot \bar{X} =$$

Making Predictions

Knowing the model (equation) that best fits the data allows us to make predictions for values that were not measured.

Example Use the regression line to predict **Fail.Time** if **Resistance** is 40 Ohms.

$$\hat{y} = -8.56 + 1.10x$$



Note: It is NOT reliable to make predictions for X values that are outside the range of your data set. (This is called *extrapolation*.)

In this scenario, our X values go from 29 to 47. Using $X = 70$ to predict

$$\hat{y} = -8.56 + 1.10 \times 70 = 68.4 \text{ min}$$

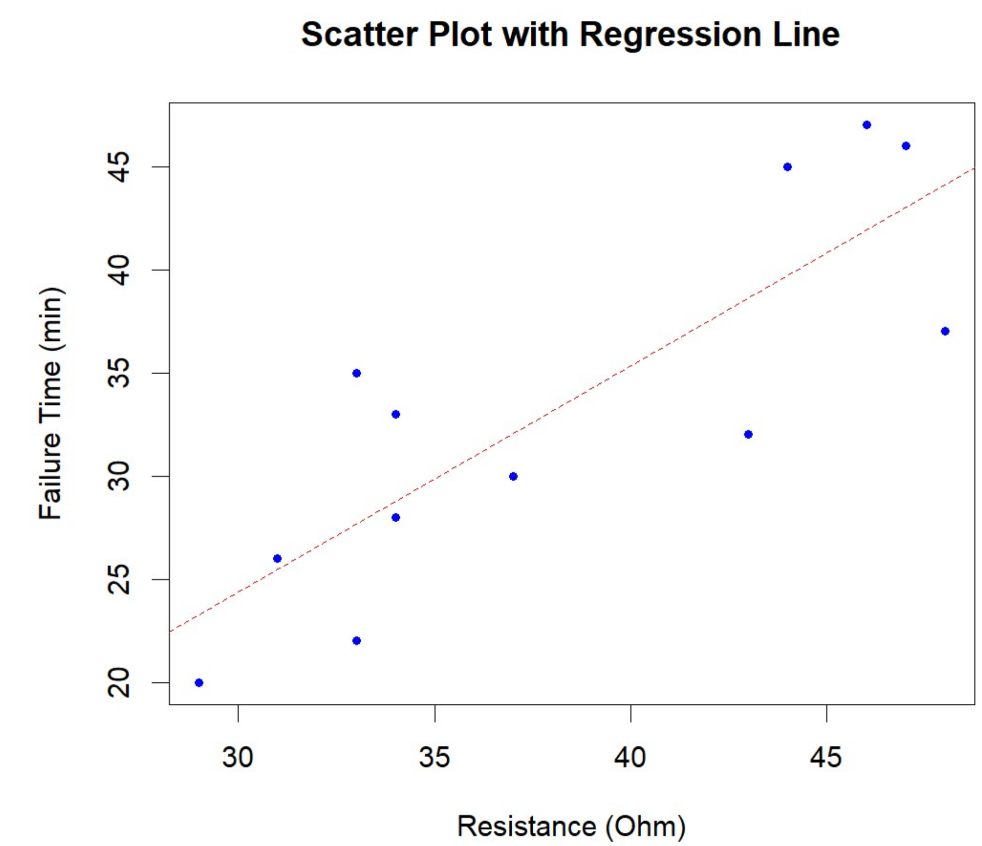
would be unreliable, since we do not know if the linear relationship continues for $X > 47$.

进行预测

了解最能拟合数据的模型（方程）可使我们对未测量的值进行预测。

示例 使用回归直线预测当**电阻**为40欧姆时的**失效时间**。

$$\hat{y} = -8.56 + 1.10x$$



注意： 对于超出数据集范围的 X 值进行预测是不可靠的。（这被称为外推法。）

在此场景中，我们的 X 值范围是从 29 到 47。使用 $X = 70$ 进行预测

$$\hat{y} = -8.56 + 1.10 \times 70 = 68.4 \text{ min}$$

将是不可靠的，因为我们不知道线性关系在$X > 47$处是否仍然成立。

Prediction Intervals

In the previous example we obtained a point estimate $\hat{y} = 35.4$ for $X = 40$ Ohm.

It is even more useful to generate an *interval estimate* for Y. In this context, we call such an interval a *prediction interval*.

To do this, we need to make certain assumptions about X and Y.

Assumptions

- X and Y follow a bivariate distribution
- The variance of Y is the same for all specific values of X (*homoscedasticity)*
- The mean of Y at each X level lies along a line for different values of X (*linearity*)

Under these assumptions, it is possible to prove that the prediction error $(Y - \hat{y})$ follows a normal distribution with a standard deviation that we can estimate using:

$$S_e = \sqrt{\frac{\sum(Y-\hat{y})^2}{n-2}} = \sqrt{\frac{(\sum y^2) - a(\sum y) - b(\sum xy)}{n-2}}$$

The quantify S_e is called the *Standard Error of the Estimate*.

Think of S_e as the typical vertical distance between the regression line and a point (X, Y).

Example Calculate S_e using the sample data for **Resistance** and **Fail.Time**.

$$S_e = \sqrt{\frac{(\sum y^2) - a(\sum y) - b(\sum xy)}{n-2}}$$

$$= \sqrt{\frac{14301 - (-8.56041)(401) - (1.09744)(15907)}{12-2}} =$$

预测区间

在前面的例子中，我们得到了$X = 40$ 欧姆的一个点估计值$\hat{y} = 35.4$。

为Y生成一个区间估计甚至更有用。在此背景下，我们将这样的区间称为预测区间。

为此，我们需要对X 和Y做出某些假设。

假设

- X 和Y 遵循二元分布
- 对于X 的所有特定取值，Y 的方差相同（同方差性）
- 在每个X 水平上，Y 的均值位于一条直线上，适用于X 的不同取值（线性）

在这些假设下，可以证明预测误差（$Y - \hat{y}$）服从正态分布，其标准差可使用以下公式进行估计：

$$S_e = \sqrt{\frac{\sum(Y-\hat{y})^2}{n-2}} = \sqrt{\frac{(\sum y^2) - a(\sum y) - b(\sum xy)}{n-2}}$$

该量 S_e 称为 估计标准误差。

将S_e 视为回归线与点(X, Y)之间的典型垂直距离。

示例 使用**电阻** 和**失效时间**的样本数据计算S_e 。

$$S_e = \sqrt{\frac{(\sum y^2) - a(\sum y) - b(\sum xy)}{n-2}}$$

$$= \sqrt{\frac{14301 - (-8.56041)(401) - (1.09744)(15907)}{12-2}} =$$

To construct a prediction interval for Y at $X = x_0$ with a confidence level $1 - \alpha$, we use

$$\hat{y} - E < Y < \hat{y} + E$$

where

$$\hat{y} = a + bx_0$$

$$E = t_{\alpha/2} \cdot S_e \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{X})^2}{(n-1)s_X^2}}$$

The critical t value $t_{\alpha/2}$ has $n - 2$ degrees of freedom.

Example Calculate a 95% prediction interval for Y with $X = 40$ Ohm.

为了在置信水平$1 - \alpha$下为Y 在$X = x_0$ 处构建预测区间，我们使用

$$\hat{y} - E < Y < \hat{y} + E$$

其中

$$\hat{y} = a + bx_0$$

$$E = t_{\alpha/2} \cdot S_e \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{X})^2}{(n-1)s_X^2}}$$

临界值 t $t_{\alpha/2}$ 具有 $n - 2$ 个自由度。

示例 使用 $X = 40$ Ohm 计算 Y 的 95% 预测区间。

Confidence Intervals for α and β

We can construct confidence intervals for the parameters α and β in our regression model.

$$\hat{y} = \alpha + \beta X$$

As usual, the interval depends mainly on a formula for the *margin of error, E*.

The confidence interval (at level $1-\alpha$) for the slope parameter β is

$$b - E < \beta < b + E$$

where

$$E = t_{\alpha/2}\frac{S_e}{\sqrt{n-1}\cdot s_X} \quad (\text{where df} = n-2)$$

Likewise, it is possible to perform a hypothesis test on the regression slope β using the test statistic

$$t = \frac{b-\beta}{\dfrac{S_e}{\sqrt{n-1}\cdot s_X}} = \frac{b-\beta}{S_e}\cdot\sqrt{n-1}\cdot s_X \quad (\text{df} = n-2)$$

Example Continue our example involving time to failure of resistors. Find the 95% confidence interval for the regression slope β. Also test the hypothesis that the population regression slope is non-zero, at the 5% significance level.

α 和 β 的置信区间

我们可以为回归模型中的参数 α 和 β 构建置信区间。

$$\hat{y} = \alpha + \beta X$$

与往常一样，该区间主要取决于 误差范围 E。

斜率参数 β 在水平 $1-\alpha$ 上的置信区间为

$$b - E < \beta < b + E$$

其中

$$E = t_{\alpha/2}\frac{S_e}{\sqrt{n-1}\cdot s_X} \quad (\text{where df} = n-2)$$

同样，可以使用检验统计量对回归斜率 β 进行假设检验

$$t = \frac{b-\beta}{\dfrac{S_e}{\sqrt{n-1}\cdot s_X}} = \frac{b-\beta}{S_e}\cdot\sqrt{n-1}\cdot s_X \quad (\text{df} = n-2)$$

示例 继续我们关于电阻器失效时间的例子。求回归斜率 β 的 95% 置信区间。同时在 5 % 的显著性水平下检验总体回归斜率不为零的假设。

```
> # Much of this can be read off of the output of the follow:
> model

Call:
lm(formula = Fail.time ~ Resistance, data = circuit.df)

Coefficients:
(Intercept)   Resistance  
     -8.561        1.097  

> summary(model)

Call:
lm(formula = Fail.time ~ Resistance, data = circuit.df)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.1167 -3.8628 -0.1064  4.4551  7.3449 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -8.5605     8.9685  -0.955 0.362328    
Resistance    1.0974     0.2311   4.749 0.000781 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.261 on 10 degrees of freedom
Multiple R-squared:  0.6928,	Adjusted R-squared:  0.6621 
F-statistic: 22.55 on 1 and 10 DF,  p-value: 0.0007813
```



```
> # Much of this can be read off of the output of the follow:
> model

Call:
lm(formula = Fail.time ~ Resistance, data = circuit.df)

Coefficients:
(Intercept)   Resistance  
     -8.561        1.097  

> summary(model)

Call:
lm(formula = Fail.time ~ Resistance, data = circuit.df)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.1167 -3.8628 -0.1064  4.4551  7.3449 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -8.5605     8.9685  -0.955 0.362328    
Resistance    1.0974     0.2311   4.749 0.000781 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.261 on 10 degrees of freedom
Multiple R-squared:  0.6928,	Adjusted R-squared:  0.6621 
F-statistic: 22.55 on 1 and 10 DF,  p-value: 0.0007813
```


132