Simple Comparative Experiments

Lin Wang

Some Terminology

- Factor: Process input variable that may cause a change in the output
- Response: the output variable
- Experimental run: experimental trial
- Treatment: the combination of the settings of one or several factors
- Design: A set of experimental runs, arranged as a table or a matrix, where a row accommodates a run, and a column accommodates the settings of a factor

The goal of experimental design is typically to study the treatment effects accurately.

Toy Example (Drug Experiment)

Factor

- Drug dose: Low (10 mg), High (30 mg)
- Dosing frequency: Once daily, Twice daily

Response

- Change in symptom score (baseline → Week 4)

Treatment

- (10 mg, once daily)
- (10 mg, twice daily)
- (30 mg, once daily)
- (30 mg, twice daily)

Experimental run

- Treat one patient under one treatment and record response

Design

Dose (mg)	Frequency
10	Once
10	Twice
30	Once
30	Twice

Planning, Conducting & Analyzing an Experiment

1. Recognition & statement of problem
2. Choice of factors, levels, and ranges
3. Selection of the response variable(s)
4. Choice of design
5. Conducting the experiment
6. Statistical analysis
7. Drawing conclusions, recommendations, decisions

A Simple Comparative Experiment
----Portland Cement Formulation (Section 2.1)

- An engineer is studying the formulation of a Portland cement. He wants to see if adding a polymer latex emulsion impacts and tension bond strength of the cement.
- Goal: Compare two formulations: two treatments or two levels of the factor formulations.
- Design of the experiment: 10 samples of the original formulation and 10 samples of the modified formulation, performed in a random order.

- Factor: cement formulation

  (Original formulation, Modified formulation)
- Response: tension bond strength
- Treatment: One specific formulation
- Design: 20 total runs (10 original,10 modified) Randomized run order

Run	Formulation	Run	Formulation
1	M	11	U
2	U	12	M
3	M	13	U
4	M	14	M
5	U	15	M
6	U	16	U
7	M	17	M
8	U	18	U
9	M	19	U
10	U	20	M

The Basic Principles of DOE

- **Randomization**
  - Running the trials in an experiment in random order
  - Notion of balancing out effects of “lurking” variables
- **Replication**
  - Improve precision of effect estimation, estimation of error or background noise
  - Difference between replication and repeat measurements
- **Blocking**
  - Dealing with controllable nuisance factors

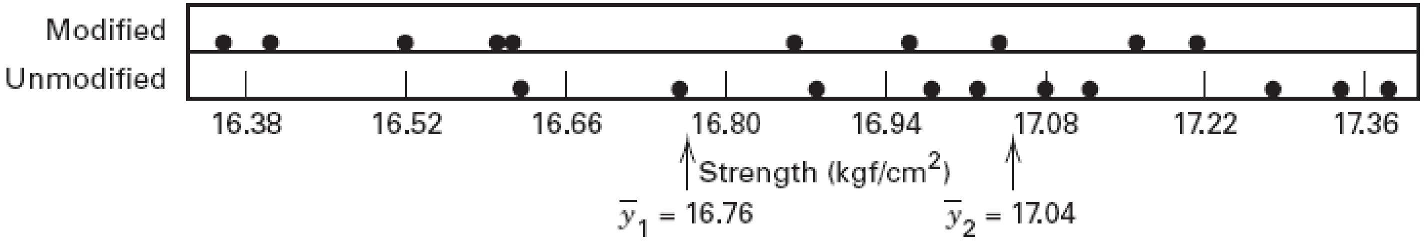
Conducting the experiment

■ **TABLE 2.1**

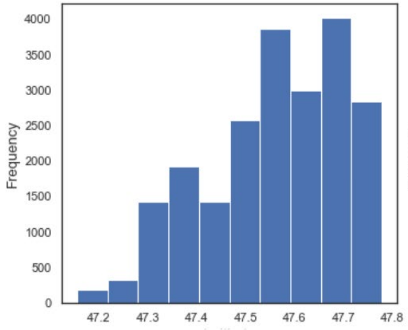
Tension Bond Strength Data for the Portland Cement Formulation Experiment

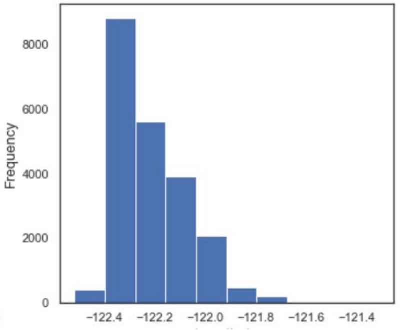
j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

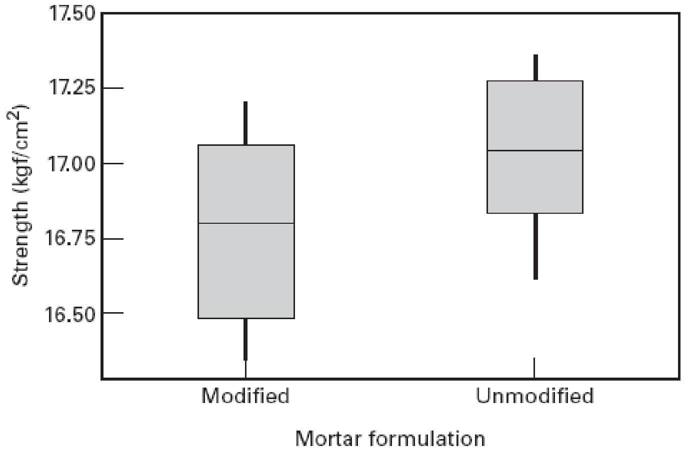
Graphical View of the Data



If you have a large sample, a histogram may be useful







■ **FIGURE 2.3** Box plots for the Portland cement tension bond strength experiment

Note: The histograms are for illustration purpose. The Portland cement data may not be big enough to draw a good histogram.

Any ideas on the analysis of the data?

Linear Regression Anylasis

Define an indicator variable:

$$X_i = \begin{cases} 0, & \text{U (Unmodified)} \\ 1, & \text{M (Modified)} \end{cases}$$

Regression model

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

where

- β_0: mean response for **Unmodified** cement
- β_1: **treatment effect** of polymer latex
- ε_i: random error, mean 0, variance σ^2

Inference goal

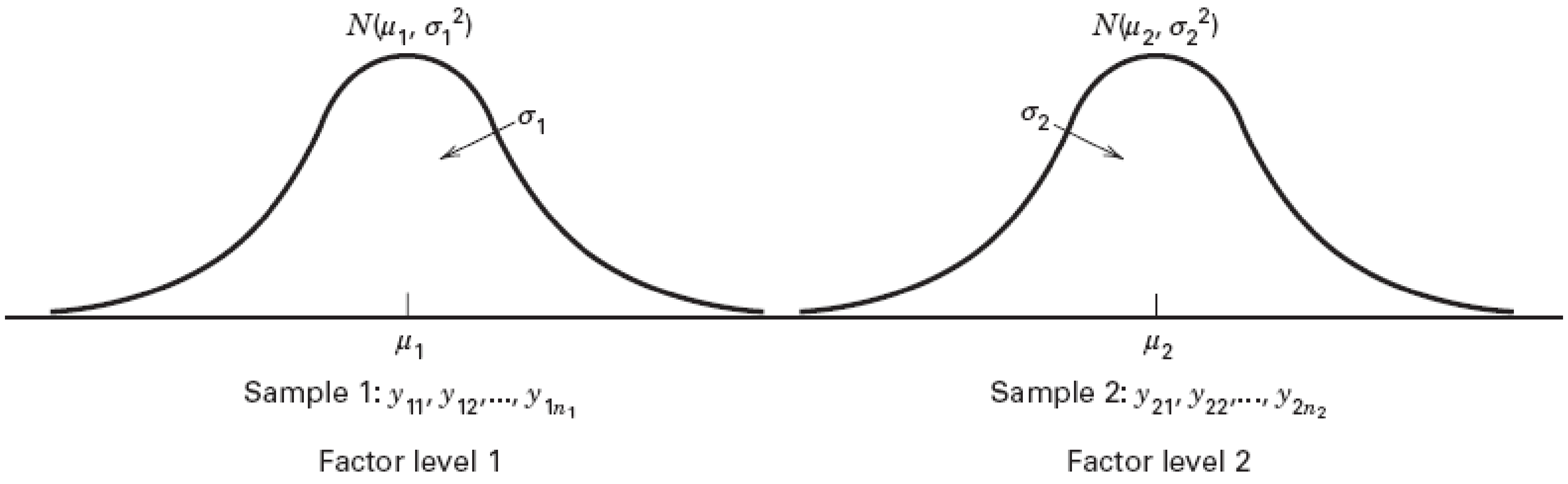
- Test

$$H_0 : \beta_1 = 0$$

Equivalent to a **two-sample test**

(no effect of polymer latex)

Two-sample Hypothesis Testing



■ **FIGURE 2.9** The sampling situation for the two-sample t-test

- Assume data are sampled from two **normal** distributions
- Statistical hypotheses:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Two-Sample *Z*-Test: If we knew the two variances

Use the sample means to draw inferences about the population means

$\bar{y}_1 - \bar{y}_2 = 16.76 - 17.04 = -0.28$

$$\frac{\text{Difference in sample means}}{\text{Standard deviation of the difference in sample means}}$$

$$\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n}, \quad \text{and } \sigma_{\bar{y}_1 - \bar{y}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}, \bar{y}_1 \text{ and } \bar{y}_2 \text{ independent}$$

This suggests a statistic:

$$Z_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

If the variances were known we could use the normal distribution as the basis of a test

Z_0 has a N(0,1) distribution if the two population means are equal

Estimation of Parameters

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ estimates the population mean } \mu$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \text{ estimates the variance } \sigma^2$$

Modified Mortar

“New recipe”

$$\bar{y}_1 = 16.76$$
$$S_1^2 = 0.100$$
$$S_1 = 0.316$$
$$n_1 = 10$$

Unmodified Mortar

“Original recipe”

$$\bar{y}_2 = 17.04$$
$$S_2^2 = 0.061$$
$$S_2 = 0.248$$
$$n_2 = 10$$

Two-Sample *Z*-Test: If we knew the two variances

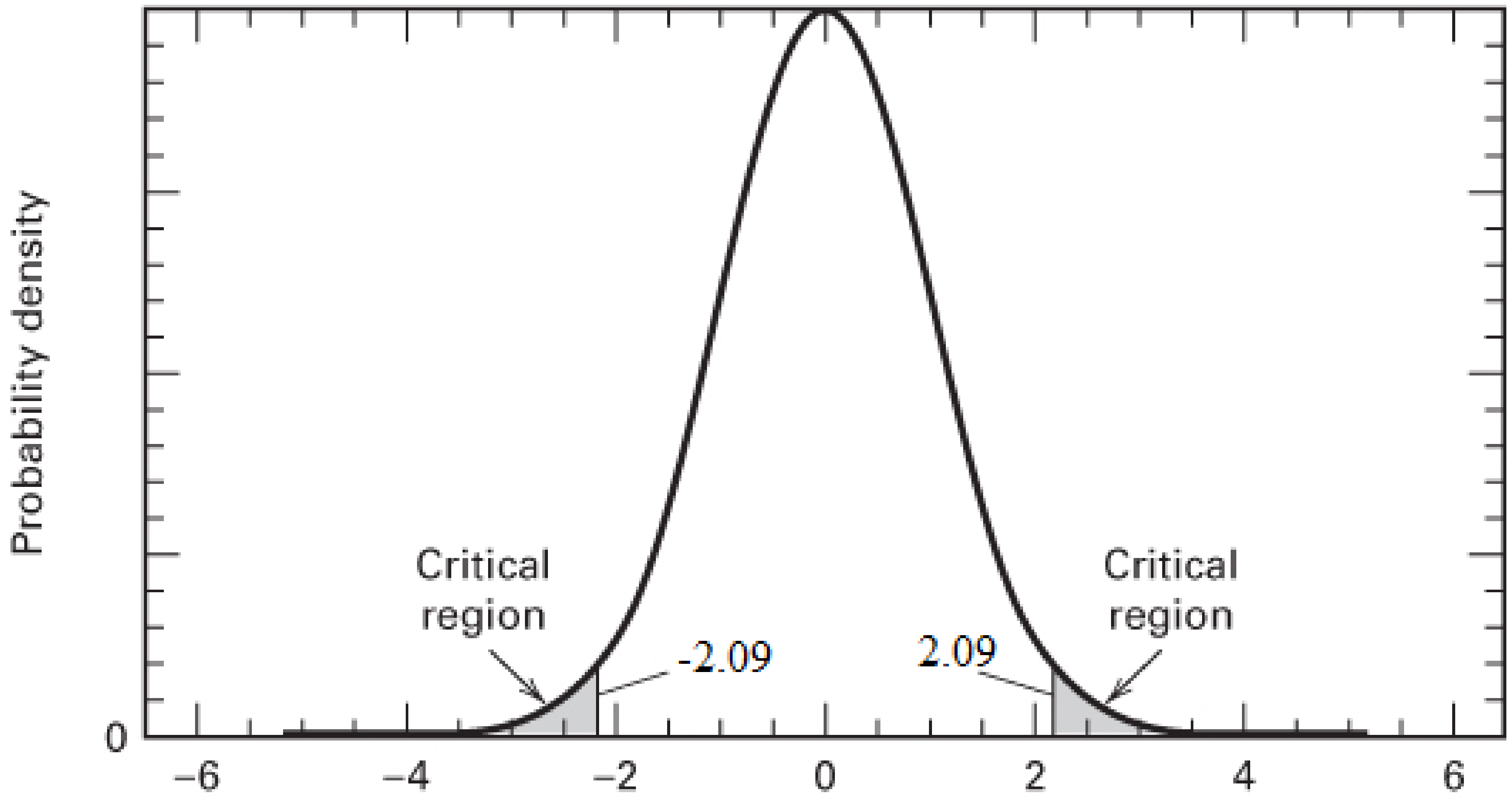
Suppose that $\sigma_1 = \sigma_2 = 0.30$. Then we can calculate

$$Z_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{-0.28}{\sqrt{\frac{0.3^2}{10} + \frac{0.3^2}{10}}} = \frac{-0.28}{0.1342} = -2.09$$

How “unusual” is the value $Z_0 = -2.09$ if the two population means are equal?

It turns out that 95% of the area under the standard normal curve (probability) falls between the values $Z_{0.025} = 1.96$ and $-Z_{0.025} = -1.96$.

So the value $Z_0 = -2.09$ is pretty unusual in that it would happen less that 5% of the time if the population means were equal


Probability density
Critical region
Critical region
-2.09
2.09
0
−6
−4
−2
0
2
4
6

P-value

In R, use pnorm() to find the p-value

```
> pnorm(-2.09)
[1] 0.0183089
```

The *P*-value is twice this probability, or 0.03662.

So we would reject the null hypothesis at any level of significance that is less than or equal to 0.03662.

Typically 0.05 is used as the cutoff.

The two-sample *t*–Test (when we don’t know the variances)

- The Z-test just described would work perfectly if we knew the two population variances.
- Since we usually don’t know the true population variances, what would happen if we just **plugged in the sample variances**?
- The answer is that if the sample sizes were large enough (say both $n > 30$ or 40) the Z-test would work just fine. It is a good approximation of Z-statistic and is a good **large-sample test**.
- But what if the sample size is small?
- It turns out that if the sample size is small we can no longer use the Z-test because Z_0 does not follow the N(0,1) distribution.

How the Two-Sample *t*-Test Works:

Use S_1^2 and S_2^2 to estimate σ_1^2 and σ_2^2

The previous ratio becomes $\dfrac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$

Assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (different case to be covered later)

Pool the individual sample variances:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

How the Two-Sample *t*-Test Works:

The test statistic is

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If H_0 holds, t_0 follows a t-distribution with $n_1 + n_2 - 2$ degrees of freedom

- t_0 is a "distance" measuring how far apart the averages are expressed in standard deviation units
- Values of t_0 that are near zero are consistent with the null hypothesis
- Values of t_0 that are very different from zero are consistent with the alternative hypothesis

The Two-Sample *t*-Test for the cement data (Table 2.1)

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{9(0.100) + 9(0.061)}{10 + 10 - 2} = 0.081$$

$$S_p = 0.284$$

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{16.76 - 17.04}{0.284\sqrt{\frac{1}{10} + \frac{1}{10}}} = -2.20$$

The two sample means are a little over two standard deviations apart

Is this a "large" difference?

P-value

In R, use pt(, df) to find the p-value

```
> pt(-2.20,18)
[1] 0.02055429
```

The *P*-value is twice this probability 0.042.

The *P*-value is a measure of how unusual the value of the test statistic is, given that the null hypothesis is true

The *P*-value is the risk of **wrongly rejecting** the null hypothesis of equal means

Confidence Intervals (Section 2.4.1)

- Hypothesis testing gives an objective statement concerning the difference in means, but it doesn't specify "how different" they are
- **General form** of a confidence interval

$$L \le \theta \le U \text{ where } P(L \le \theta \le U) = 1 - \alpha$$

- In a comparative experiment, we are interested in the 100(1- α)% **confidence interval** on the difference in two means:

$$\mu_1 - \mu_2$$

$$\frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \quad \text{is distributed as } t_{n_1+n_2-2}.$$

$$P\left(-t_{\alpha/2,n_1+n_2-2} \le \frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \le t_{\alpha/2,n_1+n_2-2}\right) = 1 - \alpha$$

$$P\left(\bar{y}_1 - \bar{y}_2 - t_{\alpha/2,n_1+n_2-2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le \bar{y}_1 - \bar{y}_2 + t_{\alpha/2,n_1+n_2-2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 1 - \alpha$$

$100(1-\alpha)$ percent confidence interval for $\mu_1 - \mu_2$:

$$\bar{y}_1 - \bar{y}_2 - t_{\alpha/2,n_1+n_2-2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le \bar{y}_1 - \bar{y}_2 + t_{\alpha/2,n_1+n_2-2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

The actual 95 percent confidence interval estimate for the difference in mean tension bond strength for the formulations of Portland cement mortar is found by substituting in Equation 2.30 as follows:

$$16.76 - 17.04 - (2.101)0.284\sqrt{\tfrac{1}{10} + \tfrac{1}{10}} \leq \mu_1 - \mu_2 \leq 16.76 - 17.04 + (2.101)0.284\sqrt{\tfrac{1}{10} + \tfrac{1}{10}}$$

$$-0.28 - 0.27 \leq \mu_1 - \mu_2 \leq -0.28 + 0.27$$

$$-0.55 \leq \mu_1 - \mu_2 \leq -0.01$$

Thus, the 95 percent confidence interval estimate on the difference in means extends from -0.55 to -0.01 kgf/cm^2. Put another way, the confidence interval is $\mu_1 - \mu_2 = -0.28 \pm 0.27$ kgf/cm^2, or the difference in mean strengths is -0.28 kgf/cm^2, and the accuracy of this estimate is ± 0.27 kgf/cm^2. Note that because $\mu_1 - \mu_2 = 0$ is *not* included in this interval, the data do not support the hypothesis that $\mu_1 = \mu_2$ at the 5 percent level of significance (recall that the *P*-value for the two-sample *t*-test was 0.042, just slightly less than 0.05). It is likely that the mean strength of the unmodified formulation exceeds the mean strength of the modified formulation.

Previous Example: Portland Cement Formulation

- Now let's consider linear regression for the Portland cement data

■ **TABLE 2.1**
Tension Bond Strength Data for the Portland Cement Formulation Experiment

j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

Linear regression with categorical predictors:
X is the cement formulation with two levels 0 (unmodified) and 1 (modified)
Y is the tension bond strength

```
> x=c(rep(1,10),rep(0,10))
> x
 [1] 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
> y=c(16.85,16.40,17.21,16.35,16.52,17.04,16.96,
17.15,16.59,16.57,16.62,16.75,17.37,17.12,16.98,
16.87,17.34,17.02,17.08,17.27)
```

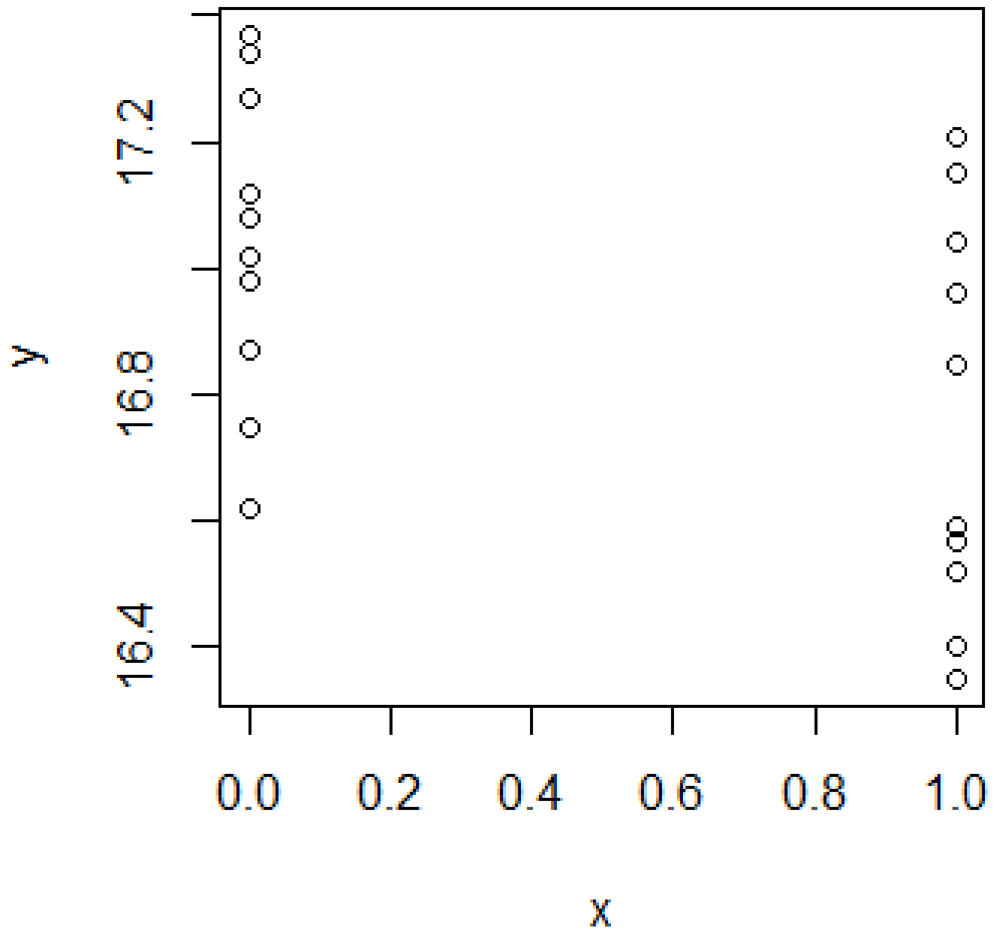
```
> summary(lm(y~x))

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.4220 -0.2065  0.0080  0.2400  0.4460 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 17.04200    0.08989 189.590   <2e-16 ***
x           -0.27800    0.12712  -2.187   0.0422 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2843 on 18 degrees of freedom
Multiple R-squared:  0.2099,	Adjusted R-squared:  0.166 
F-statistic: 4.782 on 1 and 18 DF,  p-value: 0.0422
```

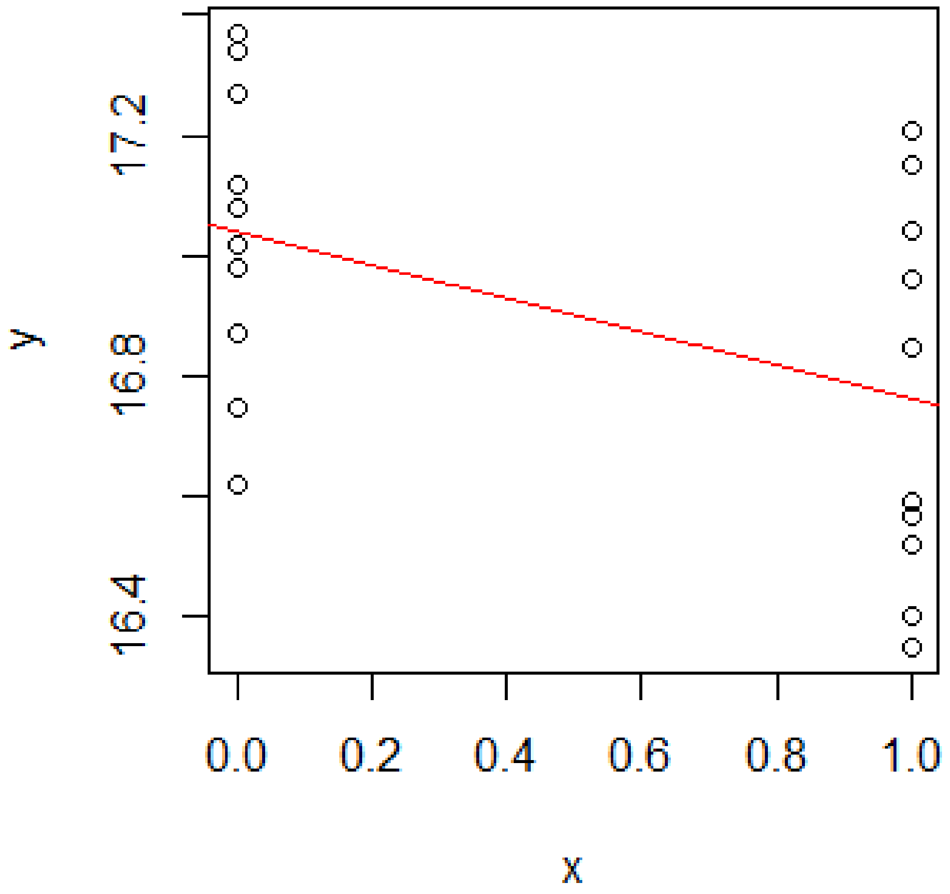
Requires normal assumption for p-value

```
> t.test(y[1:10],y[11:20],var.equal=TRUE)

        Two Sample t-test

data:  y[1:10] and y[11:20]
t = -2.1869, df = 18, p-value = 0.0422
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.54507339 -0.01092661
sample estimates:
mean of x mean of y
   16.764    17.042
```

The t-statistic is the same as the t-statistic for testing β_1=0
The p-value is the same as the p-value for testing β_1=0
The CI is the same as the CI for β_1

```
> plot(x,y)
```





17.2
16.8
16.4
y
0.0
0.2
0.4
0.6
0.8
1.0
x

What if the Two Variances (in the two-sample t-test) are Different?

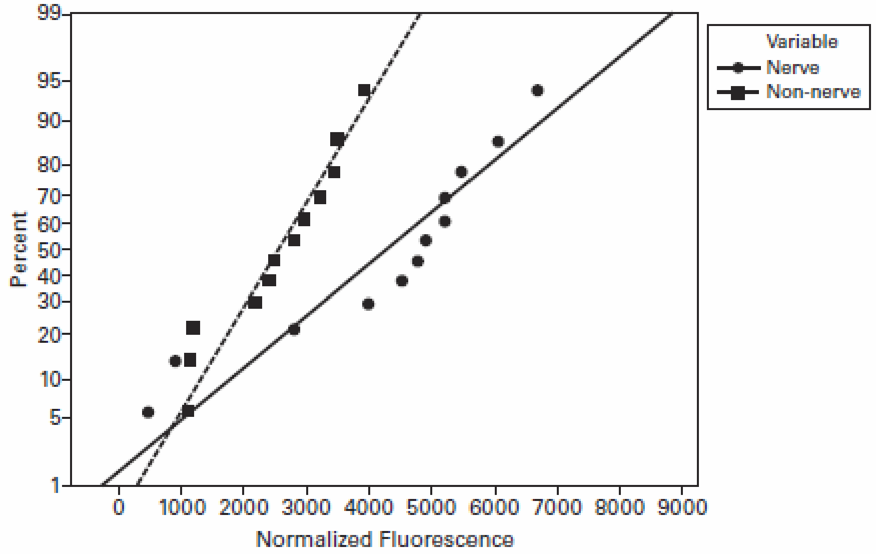
EXAMPLE 2.1

Nerve preservation is important in surgery because accidental injury to the nerve can lead to post-surgical problems such as numbness, pain, or paralysis. Nerves are usually identified by their appearance and relationship to nearby structures or detected by local electrical stimulation (electromyography), but it is relatively easy to overlook them. An article in *Nature Biotechnology* ("Fluorescent Peptides Highlight Peripheral Nerves During Surgery in Mice," Vol. 29, 2011) describes the use of a fluorescently labeled peptide that binds to nerves to assist in identification. Table 2.3 shows the normalized fluorescence after two hours for nerve and muscle tissue for 12 mice

TABLE 2.3

Normalized Fluorescence After Two Hours

Observation	Nerve	Muscle
1	6625	3900
2	6000	3500
3	5450	3450
4	5200	3200
5	5175	2980
6	4900	2800
7	4750	2500
8	4500	2400
9	3985	2200
10	900	1200
11	450	1150
12	2800	1130



■ FIGURE 2.14 Normalized Fluorescence Data from Table 2.3

The test statistic becomes

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

This statistic is not distributed exactly as t. However, the distribution of t_0 is well approximated by t if we use

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}$$

as the number of degrees of freedom

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{4228 - 2534}{\sqrt{\frac{(1918)^2}{12} + \frac{(961)^2}{12}}} = 2.7354$$

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}} = \frac{\left(\frac{(1918)^2}{12} + \frac{(961)^2}{12}\right)^2}{\frac{\left[(1918)^2/12\right]^2}{11} + \frac{\left[(961)^2/12\right]^2}{11}} = 16.1955$$

(roughly 16)

Use t-distribution with 16 degrees of freedom to compute p-value

In this example, we want to pay attention to the test:

Two-sided

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

or

One-sided

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

Use one-sided test for this example:

$t_0 = 2.74$, DF = 16, p-value = 0.007