APS606 - Lecture Note (Week 4)

Norman Lo, M.S. in Quantitative Economics

9/30/2020

Introduction to Hypothesis Testing

I. Hypothesis Testing with Normal Distribution II. Hypothesis Testing with Student's t Distribution III. Hypothesis Testing with Two Independent Samples IV. Hypothesis Testing with Two Related Samples

I. Hypothesis Testing with Normal Distribution

Introduction to Statistical Inference

In this section, we are shifting our focus to statisical inference. One of the popular methods in statistical inference is **hypothesis testing**. Genearlly speaking, hypothesis testing is the use of statistics to determine the probability that a given hypothesis is true.

Statistical Hypothesis:

The best way to determine whether a statistical hypothesis is true would be to examine the entire population. Since that is often impractical, researchers typically examine a random sample from the population. If sample data are not consistent with the statistical hypothesis, the hypothesis is rejected.

There are two types of statistical hypotheses.

Null hypothesis: The null hypothesis, denoted by H_o, is usually the hypothesis that sample observations result purely from chance.

Alternative hypothesis: The alternative hypothesis, denoted by H_o or H_a, is the hypothesis that sample observations are influenced by some non-random cause.

For example, suppose we wanted to determine whether a coin was fair and balanced. A null hypothesis might be that half the flips would result in Heads and half, in Tails. The alternative hypothesis might be that the number of Heads and Tails would be very different. Symbolically, these hypotheses would be expressed as

$$H_o : P = 0.5$$
$$H_a : P \neq 0.5$$

Suppose we flipped the coin 50 times, resulting in 40 Heads and 10 Tails. Given this result, we would be inclined to reject the null hypothesis. We would conclude, based on the evidence, that the coin was probably not fair and balanced.

Hypothesis Tests:

Statistical hypothesis testing is used to assess the strength of the evidence in a random sample against a stated null hypothesis concerning a population parameter. A null hypothesis is a conjecture about a population parameter that is stated as a mathematical equation. The usual process of hypothesis testing consists of four steps.

1. Formulate the **null hypothesis** H_0 (commonly, that the observations are the result of pure chance) and the **alternative hypothesis** H_a (commonly, that the observations show a real effect combined with a component of chance variation).
2. Identify a **test statistic** that can be used to assess the truth of the null hypothesis.
3. Compute the **p-value**, which is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true. The smaller the p-value, the stronger the evidence against the null hypothesis.
4. Compare the p-value to an acceptable significance value alpha (sometimes called an alpha value). If $p \leq \alpha$, that the observed effect is statistically significant, the null hypothesis is ruled out, and the alternative hypothesis is valid.

In fixed level testing, a significance level is chosen prior to collecting the sample, and the following decision rule is used:

If the p-value of the test statistic is less than or equal to the significance level (α), reject the null hypothesis
If the p-value of the test statistic is greater than the significance level (α), fail to reject the null hypothesis

Note: In this section, we mainly focus on using the t distribution for the hypothesis testing because population standard deviation σ is hardly be given in most of the real world problems.

Decision Errors:

Two types of errors can result from a hypothesis test.

Type I error: A Type I error occurs when the researcher rejects a null hypothesis when it is true. The probability of committing a Type I error is called the significance level. This probability is also called alpha, and is often denoted by α.

Type II error: A Type II error occurs when the researcher fails to reject a null hypothesis that is false. The probability of committing a Type II error is called Beta, and is often denoted by β. The probability of not committing a Type II error is called the Power of the test.

Example:

9. The psychology department is gradually changing its curriculum by increasing the number of online course offerings. To evaluate the effectiveness of this change, a random sample of n = 36 students who registered for introductory Psychology is placed in the online version of the course. At the end of the semester, all students take the same final exam. The average score for the sample is $\bar{x} = 76$. For the general population of students taking the traditional lecture class, the final exam scores form a normal distribution with a mean of $\mu = 71$.

If the final exam scores for the population have a standard deviation of $\sigma = 12$, does the sample provide enough evidence to conclude that the new online course is significantly different from the traditional class? Use a two-tail test with $\alpha = 0.05$.

Solution:

Step 1: State the hypothesis statements

$$H_o : \mu = \bar{x}$$
$$H_a : \mu \neq \bar{x}$$

Reject Null Hypothesis Criteria: $\alpha = 0.05$

Step 2: Convert $\bar{x}$ to the standardized z score (σ is known)

$$z = \frac{76 - 71}{\frac{12}{\sqrt{36}}} = \frac{76 - 71}{2} = 2.5$$

Step 3: Calculate the p-value or confidence interval to determine the rejection of the null hypothesis.

P-value Approach:

Reject the null hypothesis if p-value is smaller than 5%.

```
# Calculate the probability with the z score
p <- round(2*(1 - pnorm(2.5)), 3)
p
```

```
## [1] 0.012
```

```
# Alternative, calculate the probability with the average time
round(2*(1 - pnorm(76, mean = 71, sd = 2)), 3)
```

```
## [1] 0.012
```

```
print(paste("The p-value is ", p))
```

```
## [1] "The p-value is  0.012"
```

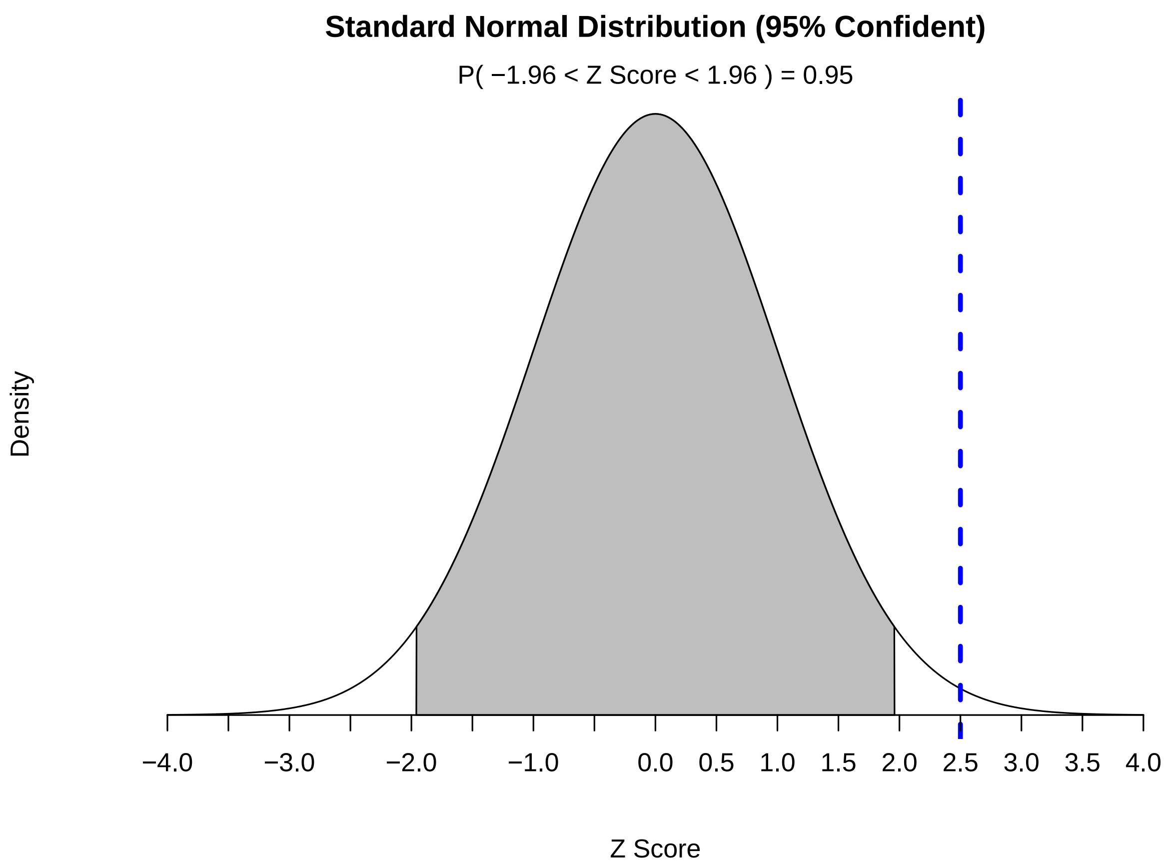
Confidence Interval Approach:

If the testing value ($/barx = 76$) is not within 95% confidence interval, we reject the null hypothesis.

```
# Calculate the 95% confidence interval
z <- qnorm(0.025, lower.tail=FALSE)
margin_of_error <- z *  12 / sqrt(36)
Upper <- 71 + margin_of_error
Lower <- 71 - margin_of_error
print(paste("95% confidence interval ", round(Lower, digits = 3),"and ", round(Upper, digits = 3),"."))
```

```
## [1] "95% confidence interval  67.08 and  74.92 ."
```

Graphical Solution:



Step 4: Make a final decision for the test.

Based on the testing result, we are 95% confident to reject the null hypothesis, so that the exam score of the new online course is statistically significant different from the traditional class.

II. Hypothesis Testing with t Distribution

When the population standard deviation is unknown in the study, we need to apply the t distribution for our hypothesis testing. However, the process are basically the same as we did with the sample mean distribution (z distribution).

Example:

The average US household spends \$90 per day. Assume a sample of 30 households in Corning, NY, showed a sample mean daily expenditure of \$84.50 with a sample standard deviation of \$14.50.

Test the hypothesis $Ho : \mu = 90$ and $Ha : \mu \neq 90$ to see whether the population mean in Corning, NY, differs from the U.S. mean. Use α=0.05 significance level. What is your conclusion?

Solution:

Step 1: State the hypothesis statements

$$H_0 : \mu = \bar{x}$$

$$H_a : \mu \neq \bar{x}$$

Reject Null Hypothesis Criteria: $\alpha = 0.05$

Step 2: Convert $\bar{x}$ to the standardized t statistics (σ is unknown)

$$t = \frac{84.5 - 90}{\frac{14.5}{\sqrt{30}}} = \frac{84.5 - 90}{2.647} = 2.078$$

Step 3: Calculate the p-value or confidence interval to determine the rejection of the null hypothesis.

Critical Value Approach:

Reject the null hypothesis if the t statistic is less than the critical value.

```
# Calculate the t statistics with the given information
t = (84.50 - 90) / (14.50 / sqrt(30))

# Identify the critical t statistics with df = 29
tcrit = qt(0.025, 29)

# Print the results
print(paste("t value is", round(t, digits = 3)))
```

```
## [1] "t value is -2.078"
```

```
print(paste("t critical value is", round(tcrit, digits = 3)))
```

```
## [1] "t critical value is -2.045"
```

```
print("Is t value less than t critical value?")
```

```
## [1] "Is t value less than t critical value?"
```

```
t < tcrit
```

```
## [1] TRUE
```

p-value Approach:

Reject the null hypothesis if the p-value is smaller than 5%.

```
# Calculate the t statistics with the given information
t = (84.50 - 90) / (14.50 / sqrt(30))

# Calculate the probability on two tails given the t statistics with df = 29
p = 2 * pt(t, 29)

# Print the results
print(paste("p value= ", round(p, digits = 4)))
```

```
## [1] "p value=  0.0467"
```

```r
print("Is p value less than the significance level 0.05?")
```

```
## [1] "Is p value less than the significance level 0.05?"
```

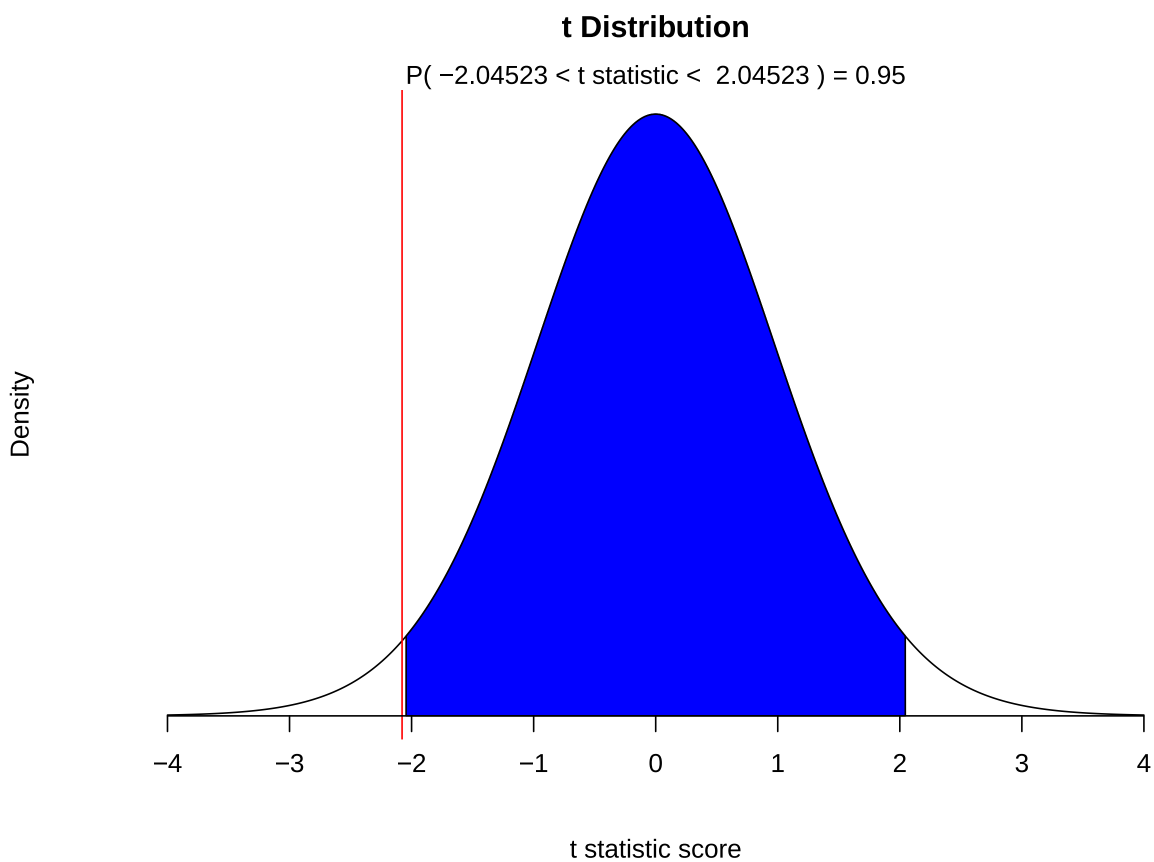
```r
p < 0.05
```

```
## [1] TRUE
```

Step 4: Make the final decision for the test.

Reject H_0 at 5% level of significance. The population mean in Corning, NY, differs from the U.S. mean at a 5% level of significance.

Graphical Solution:



One-Tail Test Example

Example:

The Coca-Cola Company reported that the mean per capita annual sales of its beverages in the United States was 423 eight-ounce servings (Coca-Coca Company Website, February 3, 2009).

Suppose you are curious whether the consumption of Coca-Cola beverages is higher in Atlanta, Georgia, the location of Coca-Cola's corporate headquarters. A sample of 36 individuals from the Atlanta area showed a sample mean annual consumption of 460.4 eight-ounce servings with a standard deviation of s=101.9 ounces. Using α =0.05, do the sample results support the conclusion that mean annual consumption of Coca-Cola beverage products is higher in Atlanta?

Solution:

Step 1: State the hypothesis statemeents

Let μ denote the mean annual consumption of Coca-Cola beverage products in Atlanta.

$$H_0 : \mu \leq \bar{x}$$

$$H_a : \mu > 423$$

Reject Null Hypothesis Criteria: $\alpha = 0.05$

Step 2: Convert $\bar{x}$ to the standardized t statistics (σ is unknown)

$$t = \frac{460.4 - 423}{\frac{101.9}{\sqrt{36}}} = \frac{460.4 - 423}{16.983} = 2.202$$

Step 3: Calculate the p-value or confidence interval to determine the rejection of the null hypothesis.

Critical Value Approach:

Reject the null hypothesis if the t statistic is less than the critical value.

```
# Calculate the t statistics from the given information
t = (460.4 - 423) / (101.9 / sqrt(36))

# Calculate the critical region locations
tcrit = qt(0.05, 35, lower.tail=F)

# Print the results
print(paste("t value is", round(t, digits = 3)))
```

```
## [1] "t value is 2.202"
```

```
print(paste("t critical value is", round(tcrit, digits = 3)))
```

```
## [1] "t critical value is 1.69"
```

```
print("Is t value greater than t critical value?")
```

```
## [1] "Is t value greater than t critical value?"
```

```
t > tcrit
```

```
## [1] TRUE
```

p-value Approach:

Reject the null hypothesis if the p-value is smaller than 5%.

```r
# Calculate the t statistic
t = (460.4 - 423) / (101.9 / sqrt(36))

# Calculate the probability for rejecting H0 based on the given t statistics
p = pt(t, 35, lower.tail=F)

# Print the results
print(paste("p value= ", round(p,digits=4)))
```

```
## [1] "p value=  0.0172"
```

```r
print("Is p value less than the significance level 0.05?")
```

```
## [1] "Is p value less than the significance level 0.05?"
```

```r
p < 0.05
```

```
## [1] TRUE
```

Conclusion:

Reject H_0 at a 5% level of significance. The sample results support the conclusion that mean annual consumption of Coca-Cola beverage products is higher in Atlanta at a 5% level of significance.

III. Hypothesis Testing with Two Independent Samples

A common task in research is to compare two populations or groups. Researchers may wish to compare the income level of two regions, the nitrogen content of two lakes, or the effectiveness of two drugs. An initial question that arises is what aspects (parameters) of the populations should be compared. We might consider comparing the averages, the medians, the standard deviations, the distributional shapes, or maximum values.

The comparison parameter is based on the particular problem. The typical comparison of two distributions is the comparison of means. If we can safely make the assumption of the data in each group following a normal distribution, we can use a two-sample t-test to compare the means of random samples drawn from these two populations.

Assumptions:

1. Random samples
2. Independent observations
3. The population of each group is normally distributed.

t Statistic Calculation

The test statistic for an Independent Samples t Test is denoted t. There are actually two forms of the test statistic for this test, depending on whether or not equal variances are assumed.

EQUAL VARIANCES ASSUMED:

When the two independent samples are assumed to be drawn from populations with identical population variances, $\sigma_1^2 = \sigma_2^2$, the test statistic t is computed as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

with

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

where

$\bar{x}_1$ = Mean of first sample $\bar{x}_2$ = Mean of second sample n_1 = Sample size of the first sample n_2 = Sample size of the second sample s_1 = Standard deviation of first sample s_2 = Standard deviation of second sample s_p = Pooled standard deviation of first and second sample

The calculated t value is then compared to the critical t value from the t distribution table with degrees of freedom df = n1 + n2 - 2 and chosen confidence level. If the calculated t value is greater than the critical t value, then we reject the null hypothesis.

Because we assume equal population variances, it is OK to "pool" the sample variances (s_p). However, if this assumption is violated, the pooled variance estimate may not be accurate, which would affect the accuracy of our test statistic (and hence, the p-value).

EQUAL VARIANCES NOT ASSUMED:

When the two independent samples are assumed to be drawn from populations with unequal variances, $\sigma_1^2 \neq \sigma_2^2$, the test statistic t is computed as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where

$\bar{x}_1$ = Mean of first sample $\bar{x}_2$ = Mean of second sample n_1 = Sample size of the first sample n_2 = Sample size of the second sample s_1 = Standard deviation of first sample s_2 = Standard deviation of second sample

The calculated t value is then compared to the critical t value from the t distribution table with degrees of freedom

$$df = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{1}{n_1 - 1}(\frac{s_1^2}{n_1})^2 + \frac{1}{n_2 - 1}(\frac{s_2^2}{n_2})^2}$$

and chosen confidence level. If the calculated t value > critical t value, then we reject the null hypothesis.

Fortunately, we do not need to do all these math by hand anymore because the statistical software can compute the test statistic in one line of code.

Example:

Does posting calorie content for menu items affect people's choices in fast food restaurants? According to results obtained by Elbel, Gyamfi, and Kersh (2011), the answer is no. The researchers monitored the calorie content of the food purchases for children and adolescents in four large fast food chain before and after mandatory labeling began in New York City. Although most of the adolescents reported noticing the calorie labels, apparently the labels had no effect on their choices. Data similar to the results obtained show an average of $\bar{x}_1 = 786$ calories per meal with $s_1 = 85$ for $n_1 = 100$ children and adolescents before the labeling, compared to an average of $\bar{x}_2 = 772$ calories with $s_2 = 91$ for a similar sample of $n_2 = 100$ after the mandatory posting. Assume the population variances are the same.

Use a two-tailed test with $\alpha = 0.05$ to determine whether the mean number of calories after the posting is significantly different than before calorie content was posted.

Step 1: State the Hypothesis Statements

$$H_o : \bar{x}_1 = \bar{x}_2$$
$$H_a : \bar{x}_1 \neq \bar{x}_2$$

$$\alpha = 0.05$$

Step 2: Calculate the t statistic

Since the population variances are assumed to be the same, we can then apply the pooled variance calculation to find the t statistic in this problem.

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$s_p = \sqrt{\frac{(100 - 1)85^2 + (100 - 1)91^2}{100 + 100 - 2}}$$

$$s_p = 88.051$$

Then, we can calculate the t statistic using the following equation,

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{786 - 772}{88.051\sqrt{\frac{1}{100} + \frac{1}{100}}}$$

$$t = \frac{14}{12.415} = 1.128$$

Step 3: Calculate the p-value or critical value for rejecting the null hypothesis

```r
# Calculate the t statistics with the given information
t = 1.128

# Identify the critical t statistics with df = 100 + 100 - 2 = 198
tcrit = qt(0.025, 198)

# Print the results
print(paste("t value is", round(t, digits = 3)))
```

```
## [1] "t value is 1.128"
```

```r
print(paste("t critical value is", round(tcrit, digits = 3)))
```

```
## [1] "t critical value is -1.972"
```

```r
print("Is t value less than t critical value?")
```

```
## [1] "Is t value less than t critical value?"
```

```r
t < tcrit
```

```
## [1] FALSE
```

p-value Approach:

Reject the null hypothesis if the p-value is smaller than 5%.

```r
# Calculate the t statistics with the given information
t = 1.128

# Calculate the probability on two tails given the t statistics with df = 198
p = 2 * pt(t, df = 198)

# Print the results
print(paste("p value= ", round(p, digits = 4)))
```

```
## [1] "p value=  1.7393"
```

```r
print("Is p value less than the significance level 0.05?")
```

```
## [1] "Is p value less than the significance level 0.05?"
```

```r
p < 0.05
```

```
## [1] FALSE
```

Step 4: Making the Decision for Rejecting the Null Hypothesis

We failed to reject the null hypothesis at 5% signficiant level and conclude that there was no significant change in calorie consumption after the mandatory posting.

Example:

Suppose we are given two simulated sample data sets (x and y). We would like to test the hypothesis that if the two sample data sets have the same means at 5% significant level.

```r
# Create two simulated sample data sets
x <- rnorm(65, mean = 55.8, sd = 5.2)
y <- rnorm(72, mean = 57.6, sd = 4.8)
```

Step 1: State the hypothesis statemeent

$$H_o : \bar{x} = \bar{y}$$
$$H_a : \bar{x} \neq \bar{y}$$

$$\alpha = 0.05$$

Step 2: Use the t.test() function to calculate the test statistic and p-value

Option 1: When the population variances are assumed to be the same.

```
# t test, assume population variances are the same
t.test(x, y, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  x and y
## t = -2.1465, df = 135, p-value = 0.03362
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.6689856 -0.1501989
## sample estimates:
## mean of x mean of y
##  55.95937  57.86896
```

Option 2: When the population variances are assumed to be different.

```
# t test, assume population variances are not the same
t.test(x, y, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  x and y
## t = -2.1632, df = 134.69, p-value = 0.03229
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.6554840 -0.1637005
## sample estimates:
## mean of x mean of y
##  55.95937  57.86896
```

Step 3: Making the Decision for Rejecting the Null Hypothesis

We reject the null hypothesis at 5% signficiant level and conclude that the mean scores from sample x and y are statistically different.

IV. Hypothesis Testing with Two Related Samples

The **paired sample t-test**, sometimes called the **dependent sample t-test**, is a statistical procedure used to determine whether the mean difference between two sets of observations is zero. In a paired sample t-test, each subject or entity is measured twice, resulting in pairs of observations. Common applications of the paired sample t-test include case-control studies or repeated-measures designs. Suppose you are interested in evaluating the effectiveness of a company training program. One approach you might consider would be to measure the performance of a sample of employees before and after completing the program, and analyze the differences using a paired sample t-test.

Like many statistical procedures, the paired sample t-test has two competing hypotheses, the null hypothesis and the alternative hypothesis. The null hypothesis assumes that the true mean difference between the paired samples is zero. Under this model, all observable differences are explained by random variation. Conversely, the alternative hypothesis assumes that the true mean difference between the paired samples is not equal to zero. The alternative hypothesis can take one of several forms depending on the expected outcome. If the direction of the difference does not matter, a two-tailed hypothesis is used. Otherwise, an upper-tailed or lower-tailed hypothesis can be used to increase the power of the test. The null hypothesis remains the same for each type of alternative hypothesis. The paired sample t-test hypotheses are formally defined below:

- The null hypothesis (H_o) assumes that the true mean difference (μ_d) is equal to zero.
- The two-tailed alternative hypothesis (H_a) assumes that μ_d is not equal to zero.
- The upper-tailed alternative hypothesis (H_a) assumes that μ_d is greater than zero.
- The lower-tailed alternative hypothesis (H_a) assumes that μ_d is less than zero.

The mathematical representations of the null and alternative hypotheses are defined below:

$$H_o : \mu_d = 0$$
$$H_a : \mu_d \neq 0$$
$$H_a : \mu_d > 0$$
$$H_a : \mu_d < 0$$

It is important to remember that hypotheses are never about data, they are about the processes which produce the data. In the formulas above, the value of μ_d is unknown. The goal of hypothesis testing is to determine the hypothesis (null or alternative) with which the data are more consistent.

Procedure for Carrying Out a Paired t-test:

Suppose a sample of n students were given a diagnostic test before studying a particular module and then again after completing the module. We want to find out if, in general, our teaching leads to improvements in students' knowledge/skills (i.e. test scores). We can use the results from our sample of students to draw conclusions about the impact of this module in general. Let x = test score before the module, y = test score after the module To test the null hypothesis that the true mean difference is zero, the procedure is as follows:

1. Calculate the difference, $d_i = y_i - x_i$, between the two observations on each pair, make sure you distinguish between positive and negative differences.
2. Calcluate the mean difference, $\bar{d}$.
3. Calculate the standard deviation of the difference, s_d, and use this to calculate the standard error of the mean difference, $se_{\bar{d}} = \frac{s_d}{\sqrt{n}}$.
4. Calculate the t statistic, which is given by $t = \frac{\bar{d}}{se_{\bar{d}}}$. Under the null hypothesis, this statistic follows a t distribution with n - 1 degrees of freedom.
5. Use R to calculate the p-value for the paired t-test.

NOTE: For this test to be valid the differences only need to be approximately normally distributed. Therefore, it would not be advisable to use a paired t-test where there were any extreme outliers.

Example:

Using the above example with n = 20 students, the following results were obtained:

Student	Pre-module Score	Post-module Score	Difference
1	18	22	4
2	21	25	4
3	16	17	1
4	22	24	2
5	19	16	-3
6	24	29	5
7	17	20	3
8	21	23	2
9	23	19	-4
10	18	20	2
11	14	15	1
12	16	15	-1
13	16	18	2
14	19	26	7
15	18	18	0
16	20	24	4
17	12	18	6
18	22	25	3
19	15	19	4
20	17	16	-1

We would like to test if the students result a statistically different test score after completing the module at 1% significant level.

$$H_o : \bar{d} = 0$$
$$H_a : \bar{d} \neq 0$$
$$\alpha = 0.01$$

Calculating the mean and standard deviation of the differences gives:

$$\bar{d} = 2.05$$
$$s_d = 2.837$$
$$se_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{2.837}{\sqrt{20}} = 0.634$$

Then, we calculate the t statistic,

$$t = \frac{\bar{d}}{se_{\bar{d}}} = \frac{2.05}{0.634} = 3.231$$

$$df = n - 1 = 20 - 1 = 19$$

Using the p-value approach,

```r
# Calculate the t statistics with the given information
t = 3.2313

# Calculate the probability on two tails given the t statistics with df = 19
p = 2 * pt(t, df = 19, lower.tail = FALSE)

# Print the results
print(paste("p value= ", round(p, digits = 4)))
```

```
## [1] "p value=  0.0044"
```

```r
print("Is p value less than the significance level 0.01?")
```

```
## [1] "Is p value less than the significance level 0.01?"
```

```r
p < 0.01
```

```
## [1] TRUE
```

Alternative approach, using t.test() function in R.

```r
t.test(pre, post, paired = TRUE, conf.level = 0.99)
```

```
##
##  Paired t-test
##
## data:  pre and post
## t = -3.2313, df = 19, p-value = 0.004395
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
##  -3.8650595 -0.2349405
## sample estimates:
## mean of the differences
##                   -2.05
```

We reject the null hypothesis at 1% signficiant level and conclude that the test score result from students who completed the module are statistically different to the ones who did not complete the module.