

R Intermediate Short Course

Session 3 - Hypothesis Testing (Two sample T-tests)

Facilitator: Rajesh Lakan

The University of the West Indies, St Augustine

Wednesday 20th July 2022
(5:00pm - 7:00pm)
(online)

For **two-sample** hypothesis tests we compare the means of two samples that are assumed to come from a normal population with either known or unknown variance.

Two sample t-tests can either be:

1. Dependent - one sample depends on the other, or
2. Independent - both samples come from different populations and are thus independent of each other.

The hypotheses differ for dependent and independent tests but they are tested in the same way as follows:

We reject the null hypothesis H_0 if,

1. The confidence interval does not contain the value at H_0
2. The $|$ test statistic $| \geq$ critical value
3. P-value $< \alpha$ (level of significance)

Recall, the critical value for a two-tailed test is $t_{\alpha/2, df}$ and for a one-tailed test is $t_{\alpha, df}$ where,

```
qt(alpha/2, df, lower.tail=F, log.p=F)      # t\alpha/2, df
qt(alpha, df, lower.tail=F, log.p=F)          # t\alpha, df
```

The degree of freedom here, depends on the type of t-test being conducted.

1. Dependent or Paired T-test

The paired t-test examines the mean difference in each pair of data points for a single sample under two different conditions for example the weight of individuals before and after dieting.

Weight before (kg)	Weight after (kg)	Difference, d_i
65	64	1
55	57	-2
88	80	8
73	68	5
62	62	0

$$\text{mean difference, } \bar{d} = \frac{\sum d_i}{n} = \frac{1}{5}[1 + (-2) + 8 + 5 + 0] = 2.4$$

Usually we estimate the population mean μ to the sample mean \bar{x} . Here we estimate the population mean difference, μ_D to the sample mean difference, \bar{d} .

The types of hypotheses for a paired t-test are:

1. Two tailed test $H_0 : \mu_D = 0$ vs $H_1 : \mu_D \neq 0$
2. Upper tailed test $H_0 : \mu_D \leq 0$ vs $H_1 : \mu_D > 0$
3. Lower tailed test $H_0 : \mu_D \geq 0$ vs $H_1 : \mu_D < 0$

Performing Paired T-test in R

Two tailed test:

```
t.test(data1, data2, alternative="two.sided",
       conf.level=CL, paired=T)
```

Upper tailed test:

```
t.test(data1, data2, alternative="greater",
       conf.level=CL, paired=T)
```

Lower tailed test:

```
t.test(data1, data2, alternative="less",
       conf.level=CL, paired=T)
```

Example: A recent study showed that meditation influences chemical substances in blood. To test this theory, the chemical levels before and after meditation were measured in 11 subjects. These measurements are given in the table below. Test at a 10% level of significance whether there is a difference in chemical levels.

before	73.1	43.8	91.3	70.0	88.3	111.4
after	80.3	50.2	96.3	69.3	84.3	101.4
before	93.0	78.4	61.2	85.9	63.2	
after	96.4	83.2	60.1	89.7	70.4	

Required to test $H_0: \mu_D = 0$ vs $H_1: \mu_D \neq 0$ at $CL = 0.90$.

R Console

```
> before=c(73.1,43.8,91.3,70.0,88.3,111.4,93.0,78.4,61.2,85.9,63.2)
> after=c(80.3,50.2,96.3,69.3,84.3,101.4,96.4,83.2,60.1,89.7,70.4)
> t.test(before,after,alternative="two.sided",conf.level=0.90,paired=T)

    Paired t-test

data: before and after
t = -1.2267, df = 10, p-value = 0.248
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
-4.9549231  0.9549231
sample estimates:
mean of the differences
                  -2

> qt(0.05,df=10,lower.tail=F,log.p=F)
[1] 1.812461
```

The 90% confidence interval (-4.955, 0.955) contains 0,
the test statistic, $|t|$ ($= 1.227 < 1.812$) and
the p-value ($= 0.248 > 0.10$).

We do not reject H_0 at a 10% level of significance concluding that there is no difference in the chemical levels after meditation.

2. Independent T-test

There are two cases of the independent t-test,

- (i) The population variance for both groups are assumed to be equal
- (ii) The population variances are unequal for both samples.

The three main types of hypotheses are:

1. **Two tailed test** $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$
or written as $H_0 : \mu_1 - \mu_2 = 0$ vs $H_1 : \mu_1 - \mu_2 \neq 0$
2. **Upper tailed test** $H_0 : \mu_1 \leq \mu_2$ vs $H_1 : \mu_1 > \mu_2$
3. **Lower tailed test** $H_0 : \mu_1 \geq \mu_2$ vs $H_1 : \mu_1 < \mu_2$

Where the null hypothesis can be written as $H_0 : \mu_1 = \mu_2$ regardless of the type of test.

Performing equal variance T-test in R

Two tailed test:

```
t.test(data1, data2, alternative="two.sided",
       conf.level=CL, var.equal=T)
```

Upper tailed test:

```
t.test(data1, data2, alternative="greater",
       conf.level=CL, var.equal=T)
```

Lower tailed test:

```
t.test(data1, data2, alternative="less",
       conf.level=CL, var.equal=T)
```

Performing unequal variance T-test in R

Two tailed test:

```
t.test(data1, data2, alternative="two.sided",
       conf.level=CL, var.equal=F)
```

Upper tailed test:

```
t.test(data1, data2, alternative="greater",
       conf.level=CL, var.equal=F)
```

Lower tailed test:

```
t.test(data1, data2, alternative="less",
       conf.level=CL, var.equal=F)
```

Example 1

In a study to compare the purity of water, 16 samples of water were each collected from two different locations. The levels of chloride were then found and tabulated as follows,

location 1	25.4	22.0	24.3	25.2	28.0	23.9	23.6	24.8
location 2	24.8	24.3	22.0	20.9	21.5	22.6	25.0	20.8
location 1	23.1	27.2	27.0	26.5	23.2	26.7	22.2	25.5
location 2	20.4	21.6	22.3	24.4	24.0	20.9	21.1	24.0

Assuming that the population variances for both groups are equal, test the hypotheses $H_0: \mu_1 \leq \mu_2$ vs $H_1: \mu_1 > \mu_2$. Use a 5% level of significance.

STEP 1: Manually enter the data in R

```
> loc1 = c(25.4,22.0,24.3,25.2,28.0,23.9,23.6,24.8,  
        23.1,27.2,27.0,26.5,23.2,26.7,22.2,25.5)  
> loc2 = c(24.8,24.3,22.0,20.9,21.5,22.6,25.0,20.8,  
        20.4,21.6,22.3,24.4,24.0,20.9,21.1,24.0)
```

STEP 2: Perform equal variance t-test (upper tailed)

```
> t.test(loc1, loc2, alternative="greater",
          conf.level=0.95, var.equal=T)
```

STEP 3: Simulate a suitable critical value (this is a one tailed test)

```
> qt(0.05,df=30,lower.tail=F,log.p=F)
```

The screenshot shows an R console window titled "R Console". The console displays the following R code and its output:

```
> loc1 = c(25.4,22.0,24.3,25.2,28.0,23.9,23.6,24.8,
+ 23.1,27.2,27.0,26.5,23.2,26.7,22.2,25.5)
> loc2 = c(24.8,24.3,22.0,20.9,21.5,22.6,25.0,20.8,
+ 20.4,21.6,22.3,24.4,24.0,20.9,21.1,24.0)
> t.test(loc1, loc2, alternative="greater", conf.level=0.95, var.equal=T)

Two Sample t-test

data: loc1 and loc2
t = 3.8772, df = 30, p-value = 0.0002673
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.335327      Inf
sample estimates:
mean of x mean of y
 24.9125    22.5375

> qt(0.05,df=30,lower.tail=F,log.p=F)
[1] 1.697261
```

STEP 4: Make a decision and interpret the results.

1. The 95% confidence interval $(1.335, \infty)$ does not contain 0.
Recall the value at H_0 is 0 (i.e. $\mu_1 \leq \mu_2 \Rightarrow \mu_1 - \mu_2 \leq 0$).
2. The test statistic ($t = 3.877$) is greater than the critical value ($t_{0.05,30} = 1.697$).
3. The p-value ($= 0.0003$) is less than the level of significance ($\alpha = 0.05$).

The conditions are satisfied therefore we can reject H_0 at a 5% level of significance. The chloride level in location 1 is significantly greater than the chloride level in location 2.

Example 2

The R built-in data set *sleep* shows the effects of two drugs on sleep. The variable "extra" indicates the increase in sleep hours while the variable "group" indicates the drug used. Test at a 10% level of significance whether the effects of both drugs are the same.

Required to test $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$ at CL = 0.90

Notice that no equal variance assumption was made so we must assume that the population variances are unequal.

STEP 1: Sort the data set to create 2 vectors that represent the effect of sleep for each drug.

```
> drug1 = sleep[1:10, 1]  
> drug2 = sleep[11:20, 1]
```

STEP 2: Perform unequal variance t-test (two-tailed)

```
> t.test(drug1, drug2, alternative="two.sided",  
         conf.level=0.90, var.equal=F)
```

STEP 3: Simulate a suitable critical value (two tailed test $\therefore t_{\alpha/2, df}$)

```
> qt(0.05, df=18, lower.tail=F, log.p=F)
```

R Output:

```
R Console
> drug1 = sleep[1:10,1]
> drug2 = sleep[11:20,1]
> t.test(drug1, drug2, alternative="two.sided", conf.level=0.90, var.equal=F)

Welch Two Sample t-test

data: drug1 and drug2
t = -1.8608, df = 17.776, p-value = 0.07939
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
-3.0533815 -0.1066185
sample estimates:
mean of x mean of y
0.75      2.33

> qt(0.05,df=18,lower.tail=F,log.p=F)
[1] 1.734064
```

STEP 4: Make a decision and interpret the results.

1. The 90% confidence interval (-3.053, -0.107) does not contain 0.
2. The test statistic $|t| (= 1.861) >$ the critical value $t_{0.05,18} (= 1.734)$.
3. The p-value ($= 0.079$) $< \alpha (= 0.10)$.

Since the conditions are satisfied, we can reject H_0 at a 10% level of significance. Therefore, there is a significant difference in the effect of both drugs on sleep.

Example 3

The *Butterfat.txt* file contains the level of butterfat found in two brands of milk. Assuming that the population variance for both brands are the same, test whether the average butterfat level in brand 1 is significantly less than that of brand 2. Conduct this test at a 1% level of significance.

Required to test $H_0 : \mu_1 \geq \mu_2$ vs $H_1 : \mu_1 < \mu_2$ at CL = 0.99
(equal variance assumed)

STEP 1: Read and attach the data set in R.

```
> butterfat = read.table("Butterfat.txt", header=T)
> attach(butterfat)
```

STEP 2: Perform equal variance t-test (lower tailed)

```
> t.test(Brand1, Brand2, alternative="less",
          conf.level=0.99, var.equal=T)
```

STEP 3: Simulate a suitable critical value (one tailed test)

```
> qt(0.01, df=46, lower.tail=F, log.p=F)
```

The screenshot shows the R Console window with the following content:

```
R Console
```

```
> butterfat=read.table("Butterfat.txt",header=T)
> attach(butterfat)
> t.test(Brand1, Brand2, alternative="less", conf.level=0.99, var.equal=T)

Two Sample t-test

data: Brand1 and Brand2
t = -0.46246, df = 46, p-value = 0.323
alternative hypothesis: true difference in means is less than 0
99 percent confidence interval:
-Inf 0.2649835
sample estimates:
mean of x mean of y
4.341667 4.404583

> qt(0.01,df=46,lower.tail=F,log.p=F)
[1] 2.410188
```

STEP 4: Decision and interpretation

1. The 99% confidence interval $(-\infty, 0.265)$ contains 0.
2. The test statistic $|t| (= 0.462) <$ the critical value $t_{0.01,46} (= 2.410)$.
3. The p-value ($= 0.323$) $> \alpha (= 0.01)$.

The conditions are not satisfied so we fail to reject H_0 at a 1% level of significance. There is insufficient evidence to conclude that the average level of butterfat in brand 1 is less than brand 2.

Hypothesis Testing Proportions

The **proportion** of an event is a fraction, probability or percentage of the entire sample that contains that event.

The population proportion is denoted by p and the estimated or sample proportion is denoted by \hat{p} where $\hat{p} = \frac{x}{n}$ for a sample of size n .

For example if 9 students in a class of 40 students obtain a grade A, the proportion of getting a grade A can be estimated by,

$$\hat{p} = \frac{x}{n} = \frac{9}{40} = 0.225 \text{ (or } 22.5\%)$$

where x = number of students that obtained a grade A and n = number of students in the class.

In hypothesis testing, proportions follow a normal probability distribution.

Note that the normal distribution is also used when sample size, $n > 30$.

Note that if the sample size is small $n < 30^*$, we will have to use another distribution.

A continuous random variable is said to have a **normal** distribution with parameters μ and σ^2 if the probability density function of X is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ for } -\infty < x < \infty$$

where $E(X) = \mu$ and $Var(X) = \sigma^2$

The normal distribution can be represented as a smooth curve which is bell-shaped or symmetric about the mean.

The **standard normal** distribution denoted by Z , is the normal distribution with parameters $\mu = 0$ and $\sigma^2 = 1$ i.e. $Z \sim Normal(0, 1)$

To generate a sample of random values which all follow a $Normal(\mu, \sigma^2)$:

```
x = rnorm(sample size, mu, sigma)
```

Probabilities associated with this distribution are computed by:

<code>dnorm(x, mu, sigma, log=F)</code>	# $P(X=x)$
<code>pnorm(x, mu, sigma, lower.tail=T, log.p=F)</code>	# $P(X \leq x)$

Example: Let X be a random variable that follows a normal distribution with mean $\mu = 80$ and variance $\sigma^2 = 100$.

- By first generating a sample of 5000 values which follow this distribution, estimate the sample mean and sample variance of X .
- Find the value of the probability density function for which $X = 100$ and $P(65 \leq X \leq 100)$.

$$X \sim Normal(80, 10^2)$$

a) > `x = rnorm(5000, 80, 10)`

> `mean(x)`

[1] 79.8711

> `var(x)`

[1] 98.41425

b) > `dnorm(100, 80, 10, log=F)`

[1] 0.005399097

> `pnorm(100, 80, 10, lower.tail=T, log.p=F) - pnorm(65, 80, 10, lower.tail=T, log.p=F)`

[1] 0.9104427

One sample proportion test

The three main types of hypothesis tests are:

1. **Two tailed test** $H_0 : p = p_0$ vs $H_1 : p \neq p_0$
2. **Upper tailed test** (one tail) $H_0 : p \leq p_0$ vs $H_1 : p > p_0$
3. **Lower tailed test** (one tail) $H_0 : p \geq p_0$ vs $H_1 : p < p_0$

Where the null hypothesis can be written as $H_0 : p = p_0$ regardless of the type of test.

Recall, we reject the null hypothesis H_0 if:

1. The confidence interval does not contain the value at H_0
2. The $|$ test statistic $| \geq$ critical value
3. P-value $< \alpha$ (level of significance)

The test statistic derived from the data given is denoted by z .

The critical value for a two-tailed test is $z_{\alpha/2}$ and a one-tailed test is z_α .

These can be determined by:

```
qnorm(1 - alpha/2)      # $z_{\alpha/2}$   
qnorm(1 - alpha)        # $z_\alpha$ 
```

Performing one-sample proportion tests in R:

The command `prop.test` in R conducts a proportion test that generates the Pearson's Chi-squared test statistic, X . As such we cannot compare the test statistic with the critical value above when testing hypotheses.

We can however use the remaining two methods i.e. reject H_0 if:

1. The confidence interval does not contain the value at H_0
2. P-value $< \alpha$ (level of significance)

Two tailed test:

```
prop.test(x, n, p, alternative = "two.sided",  
          conf.level = CL, correct = T/F)
```

Upper tailed test:

```
prop.test(x, n, p, alternative = "greater",  
          conf.level = CL, correct = T/F)
```

Lower tailed test:

```
prop.test(x, n, p, alternative = "less",  
          conf.level = CL, correct = T/F)
```

where,

x = number or successes

n = number of trials

p = the value of p_0 tested in the null hypothesis

and correct = TRUE/FALSE represents Yates correction for continuity.

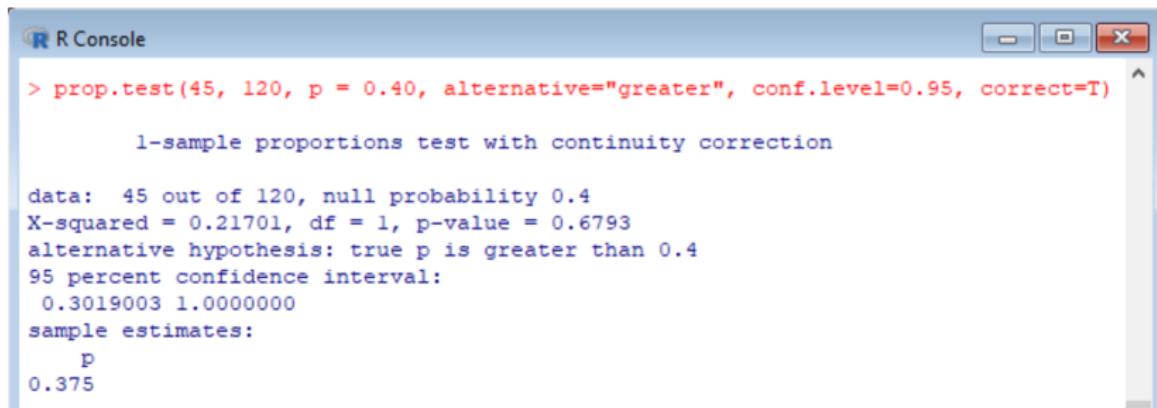
Yates continuity correction is used to account for bias.

Example 1:

In a village of 120 residents, it was found that 45 persons suffered losses due to a recent flooding. Test whether the proportion of residents affected exceeds 0.40. Use $\alpha = 0.05$ for your test.

Required to test $H_0: p \leq 0.40$ vs $H_1: p > 0.40$ at CL = 0.95.

parameters: $x = 45$ $n = 120$ $p = 0.40$



R Console

```
> prop.test(45, 120, p = 0.40, alternative="greater", conf.level=0.95, correct=T)

 1-sample proportions test with continuity correction

data: 45 out of 120, null probability 0.4
X-squared = 0.21701, df = 1, p-value = 0.6793
alternative hypothesis: true p is greater than 0.4
95 percent confidence interval:
 0.3019003 1.0000000
sample estimates:
 p
0.375
```

1. The confidence interval $(0.302, 1.000)$ contains 0.4
2. The p-value $(= 0.6793) > 0.05$

Therefore we fail to reject H_0 at $\alpha = 0.05$. There is insufficient evidence to conclude that the proportion of residents affected by flooding exceeds 0.40.

Example 2:

A local newspaper conducted a survey to determine whether readers were satisfied with their services. Among the 200 respondents, 50 readers claimed they were satisfied. Test at a 10% level of significance whether less than 32% of their readers are satisfied.

Required to test $H_0: p \geq 0.32$ vs $H_1: p < 0.32$ at CL = 0.90.

parameters: $x = 50$ $n = 200$ $p = 0.32$

```
R R Console
> prop.test(50, 200, p = 0.32, alternative ="less", conf.level = 0.90, correct=T)

 1-sample proportions test with continuity correction

data: 50 out of 200, null probability 0.32
X-squared = 4.1877, df = 1, p-value = 0.02036
alternative hypothesis: true p is less than 0.32
90 percent confidence interval:
 0.0000000 0.2937763
sample estimates:
 p
0.25
```

1. The confidence interval (0.000, 0.294) does not contain 0.32
2. The p-value (= 0.020) < 0.10

Therefore we reject H_0 at $\alpha = 0.10$. The proportion of satisfied readers is significantly less than 32%.