Hypothesis Testing

Ozancan Özdemir

April 13, 2018

Statistical Inference

Hypothesis Testing

A hypothesis is a statement on the population paramater. The aim of the hypoythesis test is to decide which one of the complementary hypothesis is true based on sample values from the population.

One sample Z-test

It is a test that used to test the hypothesis about the sample mean. It is only used to test sample mean. Here is the type of hypotheses

Ho: µ1=µ vs H1= µ1 != µ (Two Tail)

Ho: µ1<=µ vs H1= µ1> µ (Right Tail)

Ho: µ1>=µ vs H1= µ1=µ vs H1= µ1

One Sample T-test

The one-sample t-test is used to determine whether a sample comes from a population with a specific mean. This population mean is not always known, but is sometimes hypothesized. It is used when sample mean and variance is unknown.

When we use student t- test

-Data points should be independent from each other.

-Your data should be normally distributed.

-Your data should be randomly selected from a population, where each item has an equal chance of being selected.

t= $\frac{\bar{x}-\mu}{s/n}$

In this equation, $\bar{x}$ is sample mean, μ is the population mean given in the question, s is the standard deviation of the sample.

The rejection rules for t-test are as follows;

	Lower Tail	Upper Tail	Two Tail
Hypothesis	Ho:$\mu_{1\geq}\mu$ $H_1=\mu_{1<}\mu$	Ho:$\mu_{1\leq}\mu$ $H_1=\mu_{1>}\mu$	Ho:$\mu_{1=}\mu$ $H_1=\mu_{1\neq}\mu$
Test	t	t	t
P value	$p\leq\alpha$	$p\leq\alpha$	$p\leq\alpha$
C.V	$T<-t_{1-\alpha,n-1}$	$T>t_{1-\alpha,n-1}$	$\|T\|>t_{1-\alpha/2,n-1}$

Consider mtcars data set. We want to test whether mean value of mpg is different than 15 or not.

```r
#Checking normality at first
shapiro.test(mtcars$mpg)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mtcars$mpg
## W = 0.94756, p-value = 0.1229
```

```r
t.test(mtcars$mpg,alternative ="two.sided",mu=15)
```

```
##
##  One Sample t-test
##
## data:  mtcars$mpg
## t = 4.778, df = 31, p-value = 4.054e-05
## alternative hypothesis: true mean is not equal to 15
## 95 percent confidence interval:
##  17.91768 22.26357
## sample estimates:
## mean of x
##  20.09062
```

We want to test whether mean value of mpg is less than 15 or not.

```r
t.test(mtcars$mpg,alternative ="less",mu=15)
```

```
##
##  One Sample t-test
##
## data:  mtcars$mpg
## t = 4.778, df = 31, p-value = 1
## alternative hypothesis: true mean is less than 15
## 95 percent confidence interval:
##      -Inf 21.89707
## sample estimates:
## mean of x
##  20.09062
```

We want to test whether mean value of mpg is greater than 15 or not.

```r
t.test(mtcars$mpg,alternative ="greater",mu=15)
```

```
##
##  One Sample t-test
##
## data:  mtcars$mpg
## t = 4.778, df = 31, p-value = 2.027e-05
## alternative hypothesis: true mean is greater than 15
## 95 percent confidence interval:
##  18.28418      Inf
## sample estimates:
## mean of x
##  20.09062
```

Two-Sample-T-test

A two-sample t-test is used when you want to compare two independent groups to see if their means are different. There are two options for estimating the variances for the 2-sample t-test with independent samples; using pooled variances or using separate variances.

```r
knitr::include_graphics("3.png")
```

Pooled Variances	***Separate Variances***
$s_p = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$	$t^* = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
$t^* = \dfrac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$df = \dfrac{(n_1-1)\cdot(n_2-1)}{(n_2-1)C^2 + (1-C)^2(n_1-1)}$
	$C = \dfrac{s_1^2/n_1}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Assumptions

-Samples should be taken from Normal Distribution randomly with unknown variances.

-Two samples must be independent of each other.

4. 6 subjects were given a drug (treatment group) and an additional 6 subjects a placebo (control group). Their reaction time to a stimulus was measured (in ms). We want to perform a two-sample t-test for comparing the means of the treatment and control groups.

Control = 91, 87, 99, 77, 88, 91

Treat = 101, 110, 103, 93, 99, 104

Ho: µ1???µ2 vs H1= µ1<µ2 (Left Tail)

```r
Control = c(91, 87, 99, 77, 88, 91)
Treat = c(101, 110, 103, 93, 99, 104)
```

```r
#Check Normality
shapiro.test(Control) #Follows normal distribution
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Control
## W = 0.93942, p-value = 0.6545
```

```r
shapiro.test(Treat) #Follows normal distribution
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Treat
## W = 0.98231, p-value = 0.9624
```

Checking Homogeneity of Variances

The variances of two samples can be checked in a two way. The first one is drawing a box plot. However, it is obtained more precise result using statistical test which is F-test.

Ho: Variances are equal.

H1: They are not equal.

```
var.test(Control,Treat) #Variances are equal
```

```
##
##  F test to compare two variances
##
## data:  Control and Treat
## F = 1.6119, num df = 5, denom df = 5, p-value = 0.6131
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##   0.2255582 11.5194293
## sample estimates:
## ratio of variances
##           1.611925
```

There are two ways to hold the hypotesis.

First way

```
sp<-sqrt( ( (length(Control)-1)*var(Control)+(length(Treat)-1)*var(Treat))/(length(Control)+length(Trea

tobs<-(mean(Control)-mean(Treat))/(sp*sqrt( (1/length(Control))+(1/length(Treat)))) #observed t value
tobs
```

```
## [1] -3.445613
```

```
tcv<-qt(0.05,df=length(Control)+length(Treat)-2,lower.tail = TRUE) #tabulated t-value
tcv
```

```
## [1] -1.812461
```

```
#If you hold a right tailed test, you have to write lower.tail=FALSE


pval<-pt(tobs, df=length(Control)+length(Treat)-2,lower.tail = TRUE) #calculate p value by hand
pval
```

```
## [1] 0.003136062
```

```
#If you hold a right tailed test, you have to write lower.tail=FALSE
```

Second Way

```
t.test(Control,Treat,alternative="less", var.equal=T,conf.level=0.95)
```

```
##
##  Two Sample t-test
##
## data:  Control and Treat
## t = -3.4456, df = 10, p-value = 0.003136
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.082744
## sample estimates:
## mean of x mean of y
##  88.83333 101.66667
```

Therefore,The mean of control group is significantly less than the mean value of treat group. ##Paired T-Test A paired t-test is used to compare two population means where you have two samples in which observations in one sample can be paired with observations in the other sample.

Examples

-Before-and-after observations on the same subjects (e.g. students' diagnostic test results before and after a particular module or course).

-A comparison of two different methods of measurement or two different treatments where the measurements/treatments are applied to the same subjects (e.g. blood)

It is essentially same as one-sample t-test.

Steps

1. Calculate the difference (d=x-y) between the two observations on each pair, making sure you distinguish between positive and negative differences.
2. Calculate the mean difference $\bar{d}$
3. Calcuate the t statistics, which is given by $t = \frac{\bar{d}}{s(d)/n}$. Under the null hypothesis, this statistic follows a t-distribution with n-1 degrees of freedom.

4.Then, compare your results with the tabulated value.

Assumptions

-The difference (d) must be normally distributed and independent of each other.

7. A study was performed to test whether cars get better mileage on premium gas than on regular gas. Each of 10 cars was first filled with either regular or premium gas, decided by a coin toss, and the mileage for that tank was recorded. The mileage was recorded again for the same cars using the other kind of gasoline. Use a paired t-test to determine whether cars get significantly better mileage with premium gas.

reg = 16, 20, 21, 22, 23, 22, 27, 25, 27, 28

prem = 19, 22, 24, 24, 25, 25, 26, 26, 28, 32

```
reg = c(16, 20, 21, 22, 23, 22, 27, 25, 27, 28)
prem = c(19, 22, 24, 24, 25, 25, 26, 26, 28, 32)
t.test(prem,reg,alternative="greater", paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  prem and reg
## t = 4.4721, df = 9, p-value = 0.0007749
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  1.180207      Inf
## sample estimates:
## mean of the differences
##                       2
```