

# One-sample model and sample size

André Meichtry

## Contents

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Continuous data: One-sample <math>t</math>-test</b> | <b>1</b> |
| 1.1      | Statistical model . . . . .                            | 1        |
| 1.2      | Intro in sample size calculation . . . . .             | 1        |
| 1.2.1    | Power approach . . . . .                               | 2        |
| 1.2.2    | Precision approach . . . . .                           | 2        |
| 1.3      | Simulation of example data . . . . .                   | 3        |
| 1.4      | Analysis . . . . .                                     | 3        |
| 1.4.1    | As one-sample $t$ -Test . . . . .                      | 3        |
| 1.4.2    | As linear model . . . . .                              | 4        |
| 1.4.3    | Meaning of Type I error rate . . . . .                 | 4        |
| 1.4.4    | Residual analysis . . . . .                            | 5        |
| 1.4.5    | Small samples without normality . . . . .              | 5        |
| <b>2</b> | <b>Equivalence testing</b>                             | <b>6</b> |
| 2.1      | Philosophical background . . . . .                     | 6        |
| 2.2      | Implementation . . . . .                               | 6        |
| 2.3      | Package TOSTER functions . . . . .                     | 7        |

## 1 Continuous data: One-sample $t$ -test

One-sample  $t$ -tests look at just one parameter of interest  $\mu$  (in addition to  $\sigma$ ) as data-generation mechanism. (see shinyApp).

### 1.1 Statistical model

- $Y_i = \mu + \epsilon_i$ ;  $\epsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$  (or with  $\epsilon_i$  i.i.d. and “large” sample size),  $i = 1, \dots, n$
- $Y_i \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$  (equivalent)

This is the simplest example of  $Y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$ , the **only intercept model** with  $\beta_1 = \mu$  (see IntroLinearModels).

### 1.2 Intro in sample size calculation

Let us first ask the **how many do I need**-question. There are two kinds of *sample size calculations*, the power approach and the precision approach.

Consider the problem  $H_0 : \mu = \mu_0$  versus  $H_1 : \mu \neq \mu_0$ .

### 1.2.1 Power approach

A priori sample size calculation, see pages 77-78 AMT. Let  $\beta$  be the Type II error (The probability of **not** rejectig  $H_0$  **if** a specified alternative  $H_1$  is true.  $1 - \beta$  is the probability of the complement, of rejectig  $H_0$  **if** a specified alternative  $H_1$  is true).

**Question:** What  $n$  we need to assure

- to reject  $H_0 : \mu = \mu_0$
- with probability  $1 - \beta$
- with false-positive rate  $\alpha$
- **if** the specified alternative  $H_1 : \mu = \mu_1$  is true?

Assume  $\mu_0 = 0$ , thus  $\delta = \mu_1 - \mu_0 = \mu_1$ . One can show hat

$$n \geq \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{(\delta/\sigma)^2}$$

The approximation  $n \approx 8/(\delta/\sigma)^2$  holds for  $\alpha = 0.05$  and  $\beta = 0.2$ .

Assume we estimate or “guess”  $\sigma$  from another study with  $\hat{\sigma} = 22$  and we want that  $H_0$  is rejected with probability 0.8 **if**  $\mu_1 = 13$  is true.

```
(qnorm(.975)+qnorm(.8))^2/((mu1-mu0)/sigmahat)^2
```

```
[1] 22.478
```

In R, you can use the *exact* version, `power.t.test()`. Read the help file `help(power.t.test)`. You have to specify  $\delta$ ,  $\sigma$ , power and the type of test:

```
power.t.test(delta=mu1-mu0,sd=sigmahat,power=0.8,type="one.sample")
```

```
One-sample t test power calculation
```

```
      n = 24.469
delta = 13
sd = 22
sig.level = 0.05
power = 0.8
alternative = two.sided
```

### 1.2.2 Precision approach

In this approach, we do not need a Type II error. We have **not** to specify the alternative, which **most often makes more sense**, since there are **many (even infinite) options for the alternative**  $\mu \neq \mu_0$ . We estimate the sample size by specifying the precision we want for the estimation, that is, we specify a priori the maximal width of the  $100 \times (1 - \alpha)\%$  CI.

An approximative  $100 \times (1 - \alpha)\%$  CI is given by  $\bar{x} \pm \delta$ , with  $\delta = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ . Solving for  $n$  gives

$$n \geq \frac{z_{1-\alpha/2}^2}{(\delta/\sigma)^2}$$

Using again  $\hat{\sigma} = 22$ :

```
delta<-c(2,3,4,5,8)
alpha<-0.05
data.frame(delta=delta,n=qnorm(1-alpha/2)^2*(sigmahat/delta)^2)
```

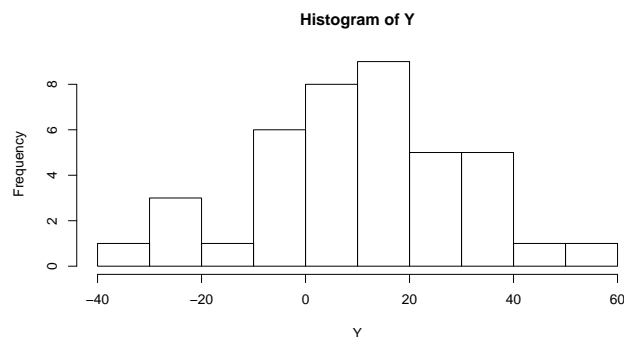
```
  delta      n
1      2 464.817
2      3 206.585
3      4 116.204
4      5  74.371
5      8  29.051
```

### 1.3 Simulation of example data

Let us simulate some data from population with **assumed known parameters**  $\mu$  and  $\sigma^2$ . We simulate to understand the emergence of data. Of course, in reality, we do not know the data-generating process.

```
set.seed(55)
mu<-10
sigma<-20
n<-40
Y<-rnorm(n,mu,sigma)
```

```
hist(Y)
```



```
summary(Y)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-31.505  -0.214   11.220   11.309   25.104   57.107
```

### 1.4 Analysis

#### 1.4.1 As one-sample *t*-Test

```
t.test(Y)
```

```
One Sample t-test
```

```
data: Y
t = 3.66, df = 39, p-value = 0.00074
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 5.0604 17.5574
```

```
sample estimates:
mean of x
    11.309
```

### 1.4.2 As linear model

This is equivalent with a linear model with only an intercept as parameter.

```
mod1<-lm(Y~1)
summary(mod1)
```

```
Call:
lm(formula = Y ~ 1)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-42.81 -11.52  -0.09   13.80   45.80
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.31      3.09    3.66 0.00074
```

Residual standard error: 19.5 on 39 degrees of freedom

Of course, the results are the same. The  $t$ -test is a the simplest linear model.

Estimation of  $\sigma$  “by hand” (this quantity is given in the output above):

```
sqrt(sum(mod1$residuals^2)/(n-1))
```

```
[1] 19.538
```

Confidence interval for  $\mu$ :

```
confint(mod1)
```

```
              2.5 % 97.5 %
(Intercept) 5.0604 17.557
```

### 1.4.3 Meaning of Type I error rate

```
nrep<-1000 ##number of simulations
mu<-0 ##true mean
alpha<-0.05 ##Type 1 error rate
n<-40 ##Sample Size
sigma<-20 ## True sigma
decision<-rep(0,nrep) ## Preparation of Decision vector
for (i in 1:nrep)
{
  Y2 <- rnorm(n=n,mean=mu,sd=sigma)
  decision[i] <- (t.test(Y2,)$p.value)< alpha ### nrep simulations and analysis
}
decision<-factor(decision,labels=c("do not reject H0","reject H0"))
decision <- prop.table(table(decision))
decision
```

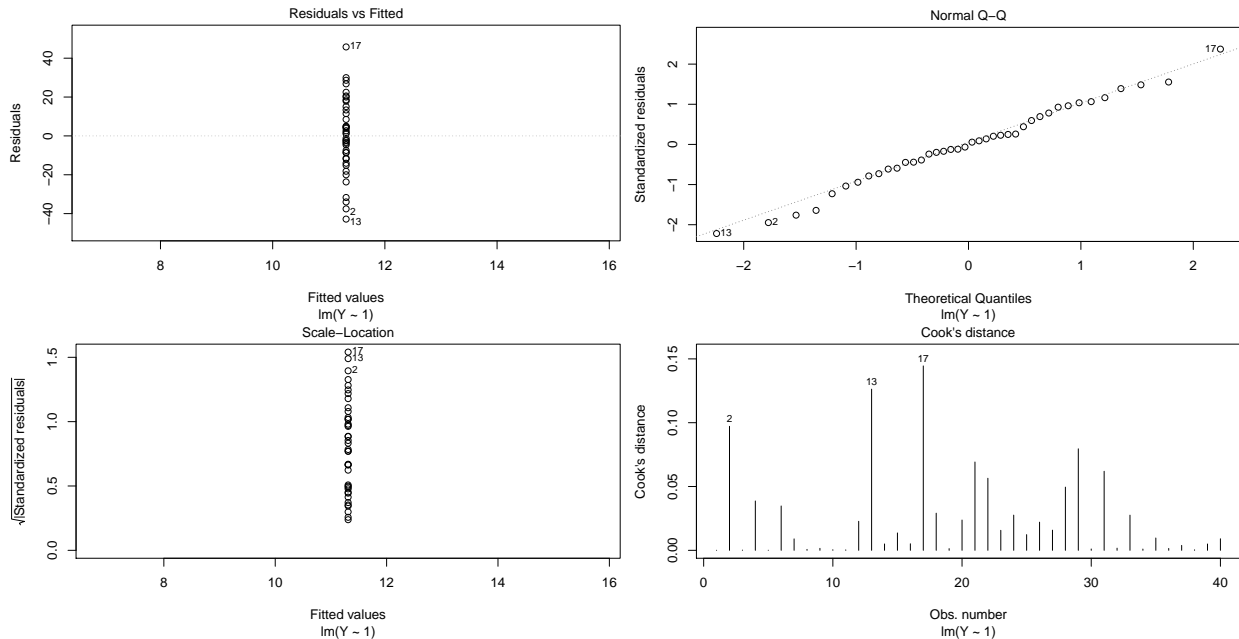
```

decision
do not reject H0      reject H0
0.945                 0.055

```

#### 1.4.4 Residual analysis

```
plot(mod1,which=c(1,2,3,4))
```



#### 1.4.5 Small samples without normality

In the example above, the assumptions are met since we simulated data from a normal distribution. In the absence of normality, this assumption of the  $t$ -test is not met. We can then perform a non-parametric test.

- The `wilcox.test` is the non-parametric alternative to the  $t$ -test.
- With the actual data, results are similar since we have simulated from a normal distribution.

```
wilcox.test(Y,conf.int=TRUE)
```

Wilcoxon signed rank test

```

data: Y
V = 655, p-value = 0.00068
alternative hypothesis: true location is not equal to 0
95 percent confidence interval:
 5.141 17.842
sample estimates:
(pseudo)median
11.487

```

## 2 Equivalence testing

### 2.1 Philosophical background

Very often, it makes not much sense to test nulls such as  $H_0 : \mu = 0$  versus  $H_1 : \mu \neq 0$ . Assume a theory predicts a range for  $\mu$ , i.e. that  $\mu$  lies in a region  $[-\epsilon, +\epsilon]$ .

You then have the following test situation:

$H_0 : \mu \leq -\epsilon$  OR  $\mu \geq +\epsilon$  versus  $H_1 : -\epsilon < \mu < \epsilon$ , meaning that

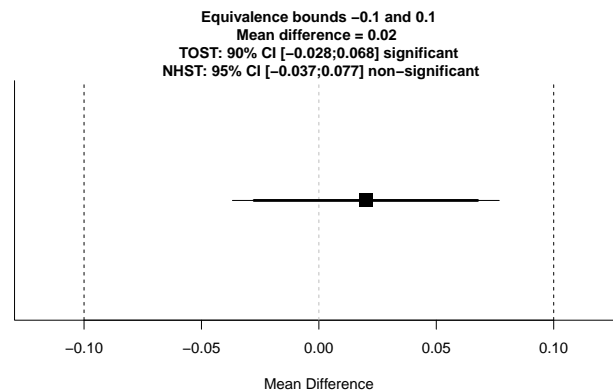
- Null: The true parameter is outside a tolerance region, “irrelevance”.
- Alternative: The true parameter is inside a tolerance region, “relevance”.

**Rejecting  $H_0$  now really means rejecting irrelevance (defined by the margins  $\epsilon$ ).**

### 2.2 Implementation

This analysis can be performed with TOST (Two One-sided  $t$ -Tests), this is implemented in the package TOSTER. For the one-sample situation, we have the function `TOSTone.raw()`.

```
library(TOSTER)
## Test observed mean of 0.52 and standard deviation of 0.5 in sample of 300 participants
## against 0.5 given equivalence bounds in raw units of -0.1 and 0.1, with an alpha = 0.05.
TOSTone.raw(m=0.52,mu=0.5,sd=0.5,n=300,low_eqbound=-0.1, high_eqbound=0.1, alpha=0.05)
```



TOST results:

t-value lower bound: 4.16      p-value lower bound: 0.00002

t-value upper bound: -2.77      p-value upper bound: 0.003

degrees of freedom : 299

Equivalence bounds (raw scores):

low eqbound: -0.1

high eqbound: 0.1

TOST confidence interval:

lower bound 90% CI: -0.028

upper bound 90% CI: 0.068

NHST confidence interval:

lower bound 95% CI: -0.037

upper bound 95% CI: 0.077

Equivalence Test Result:

The equivalence test was significant,  $t(299) = -2.771$ ,  $p = 0.00297$ , given equivalence bounds of  $-0.100$  and  $0.100$ .

Null Hypothesis Test Result:

The null hypothesis test was non-significant,  $t(299) = 0.693$ ,  $p = 0.489$ , given an alpha of 0.05.

Based on the equivalence test and the null-hypothesis test combined, we can conclude that the observed difference is not statistically significant.

## 2.3 Package TOSTER functions

If you need such analysis in the future, look at `help(package="TOSTER")` for other functions in the package, such as functions for sample size estimation.