**Point Estimate of Population Mean**

For any particular random sample, we can always compute its sample [mean](http://www.r-tutor.com/node/35). Although most often it is not the actual population mean, it does serve as a good **point estimate**. For example, in the data set [survey](http://www.r-tutor.com/node/61), the survey is performed on a sample of the student population. We can compute the sample mean and use it as an estimate of the corresponding population parameter.

**Problem**

Find a point estimate of mean university student height with the sample data from survey.

**Solution**

For convenience, we begin with saving the survey data of student heights in a variable height.survey.

> library(MASS)                  # load the MASS package   
> height.survey = survey$Height

It turns out not all students have answered the question, and we must filter out the missing values. Hence we apply the mean function with the "na.rm" argument as TRUE.

> mean(height.survey, na.rm=TRUE)  # skip missing values   
[1] 172.38

**Answer**

A point estimate of the mean student height is 172.38 centimeters.

**Interval Estimate of Population Mean with Known Variance**

After we found a [point estimate of the population mean](http://www.r-tutor.com/node/62), we would need a way to quantify its accuracy. Here, we discuss the case where the [population variance](http://www.r-tutor.com/node/42) *σ*2 is assumed known.

Let us denote the 100(1 −*α∕*2) [percentile](http://www.r-tutor.com/node/38) of the [standard normal distribution](http://www.r-tutor.com/node/58) as *zα∕*2. For random sample of sufficiently large size, the end points of the **interval** **estimate**at (1 − *α*) confidence level is given as follows:

        σ
¯x± zα∕2√--
        n


**Problem**

Assume the population standard deviation *σ*of the student height in [survey](http://www.r-tutor.com/node/61) is 9.48. Find the margin of error and interval estimate at 95% confidence level.

**Solution**

We first filter out missing values in survey$Height with the na.omit function, and save it in height.response.

> library(MASS)                  # load the MASS package   
> height.response = na.omit(survey$Height)

Then we compute the standard error of the mean.

> n = length(height.response)   
> sigma = 9.48                   # population standard deviation   
> sem = sigma/sqrt(n); sem       # standard error of the mean   
[1] 0.65575

Since there are two tails of the normal distribution, the 95% confidence level would imply the 97*.*5*th* percentile of the normal distribution at the upper tail. Therefore, *zα∕*2 is given by qnorm(.975). We multiply it with the standard error of the mean sem and get the margin of error.

> E = qnorm(.975)∗sem; E         # margin of error   
[1] 1.2852

We then add it up with the sample mean, and find the confidence interval as told.

> xbar = mean(height.response)   # sample mean   
> xbar + c(−E, E)   
[1] 171.10 173.67

**Answer**

Assuming the population standard deviation *σ*being 9.48, the margin of error for the student height survey at 95% confidence level is 1.2852 centimeters. The confidence interval is between 171.10 and 173.67 centimeters.

**Alternative Solution**

Instead of using the textbook formula, we can apply the z.test function in the TeachingDemospackage. It is not a core R package, and must be installed and loaded into the workspace beforehand.

> library(TeachingDemos)         # load TeachingDemos package   
> z.test(height.response, sd=sigma)   
   
       One Sample z−test   
   
data:  height.response   
z = 262.88, n = 209.000, Std. Dev. = 9.480,   
Std. Dev. of the sample mean = 0.656, p−value < 2.2e−16   
alternative hypothesis: true mean is not equal to 0   
95 percent confidence interval:   
 171.10 173.67   
sample estimates:   
mean of height.response   
                 172.38

**Interval Estimate of Population Mean with Unknown Variance**

After we found a [point estimate of the population mean](http://www.r-tutor.com/node/62), we would need a way to quantify its accuracy. Here, we discuss the case where the [population variance](http://www.r-tutor.com/node/42) is not assumed.

Let us denote the 100(1 −*α∕*2) [percentile](http://www.r-tutor.com/node/38) of the [Student t distribution](http://www.r-tutor.com/node/59) with *n*− 1 degrees of freedom as *tα∕*2. For random samples of sufficiently large size, and with [standard deviation](http://www.r-tutor.com/node/43) *s*, the end points of the **interval estimate**at (1 −*α*) confidence level is given as follows:

        s
¯x± tα∕2√--
        n


**Problem**

Without assuming the population standard deviation of the student height in [survey](http://www.r-tutor.com/node/61), find the margin of error and interval estimate at 95% confidence level.

**Solution**

We first filter out missing values in survey$Height with the na.omit function, and save it in height.response.

> library(MASS)                  # load the MASS package   
> height.response = na.omit(survey$Height)

Then we compute the sample standard deviation.

> n = length(height.response)   
> s = sd(height.response)        # sample standard deviation   
> SE = s/sqrt(n); SE             # standard error estimate   
[1] 0.68117

Since there are two tails of the Student t distribution, the 95% confidence level would imply the 97*.*5*th* percentile of the Student t distribution at the upper tail. Therefore, *tα∕*2 is given by qt(.975, df=n-1). We multiply it with the standard error estimate SE and get the margin of error.

> E = qt(.975, df=n−1)∗SE; E     # margin of error   
[1] 1.3429

We then add it up with the sample mean, and find the confidence interval.

> xbar = mean(height.response)   # sample mean   
> xbar + c(−E, E)   
[1] 171.04 173.72

**Answer**

Without assumption on the population standard deviation, the margin of error for the student height survey at 95% confidence level is 1.3429 centimeters. The confidence interval is between 171.04 and 173.72 centimeters.

**Alternative Solution**

Instead of using the textbook formula, we can apply the t.test function in the built-in statspackage.

> t.test(height.response)   
   
       One Sample t−test   
   
data:  height.response   
t = 253.07, df = 208, p−value < 2.2e−16   
alternative hypothesis: true mean is not equal to 0   
95 percent confidence interval:   
 171.04 173.72   
sample estimates:   
mean of x   
   172.38

**Sampling Size of Population Mean**

The quality of a sample survey can be improved by increasing the sample size. The formula below provide the sample size needed under the requirement of population mean interval estimate at (1 −*α*) confidence level, margin of error *E*, and population [variance](http://www.r-tutor.com/node/42) *σ*2. Here, *zα∕*2 is the 100(1 − *α∕*2) [percentile](http://www.r-tutor.com/node/38) of the [standard normal distribution](http://www.r-tutor.com/node/58).

         2 2
n = (zα∕2)σ--
      E2


**Problem**

Assume the population standard deviation *σ*of the student height in [survey](http://www.r-tutor.com/node/61) is 9.48. Find the sample size needed to achieve a 1.2 centimeters margin of error at 95% confidence level.

**Solution**

Since there are two tails of the normal distribution, the 95% confidence level would imply the 97*.*5*th* percentile of the normal distribution at the upper tail. Therefore, *zα∕*2 is given by qnorm(.975).

> zstar = qnorm(.975)   
> sigma = 9.48   
> E = 1.2   
> zstar^2 ∗ sigma^2/ E^2   
[1] 239.75

**Answer**

Based on the assumption of population standard deviation being 9.48, it needs a sample size of 240 to achieve a 1.2 centimeters margin of error at 95% confidence level.

**Point Estimate of Population Proportion**

Multiple choice questionnaires in a survey are often used to determine the the proportion of a population with certain characteristic. For example, we can estimate the proportion of female students in the university based on the result in the sample data set [survey](http://www.r-tutor.com/node/61).

**Problem**

Find a point estimate of the female student proportion from survey.

**Solution**

We first filter out missing values in survey$Sex with the na.omit function, and save it in gender.response.

> library(MASS)                  # load the MASS package   
> gender.response = na.omit(survey$Sex)   
> n = length(gender.response)    # valid responses count

To find out the number of female students, we compare gender.response with the factor ’Female’, and compute the sum. Dividing it by n gives the female student proportion in the sample survey.

> k = sum(gender.response == "Female")   
> pbar = k/n; pbar   
[1] 0.5

**Answer**

The point estimate of the female student proportion in survey is 50%.

**nterval Estimate of Population Proportion**

After we found a [point sample estimate of the population proportion](http://www.r-tutor.com/node/66), we would need to estimate its confidence interval.

Let us denote the 100(1 −*α∕*2) [percentile](http://www.r-tutor.com/node/38) of the [standard normal distribution](http://www.r-tutor.com/node/58) as *zα∕*2. If the samples size *n*and population proportion *p*satisfy the condition that *np*≥ 5 and *n*(1 − *p*) ≥ 5, than the end points of the interval estimate at (1 − *α*) confidence level is defined in terms of the sample proportion as follows.

       ∘--------
¯p± z     ¯p(1-−-¯p)
    α∕2    n


**Problem**

Compute the margin of error and estimate interval for the female students proportion in [survey](http://www.r-tutor.com/node/61)at 95% confidence level.

**Solution**

We first determine the proportion point estimate. Further details can be found in the [previous tutorial](http://www.r-tutor.com/node/66).

> library(MASS)                  # load the MASS package   
> gender.response = na.omit(survey$Sex)   
> n = length(gender.response)    # valid responses count   
> k = sum(gender.response == "Female")   
> pbar = k/n; pbar   
[1] 0.5

Then we estimate the standard error.

> SE = sqrt(pbar∗(1−pbar)/n); SE     # standard error   
[1] 0.032547

Since there are two tails of the normal distribution, the 95% confidence level would imply the 97*.*5*th* percentile of the normal distribution at the upper tail. Therefore, *zα∕*2 is given by qnorm(.975). Hence we multiply it with the standard error estimate SE and compute the margin of error.

> E = qnorm(.975)∗SE; E              # margin of error   
[1] 0.063791

Combining it with the sample proportion, we obtain the confidence interval.

> pbar + c(−E, E)   
[1] 0.43621 0.56379

**Answer**

At 95% confidence level, between 43.6% and 56.3% of the university students are female, and the margin of error is 6.4%.

**Alternative Solution**

Instead of using the textbook formula, we can apply the prop.test function in the built-in statspackage.

> prop.test(k, n)   
   
       1−sample proportions test without continuity   
           correction   
   
data:  k out of n, null probability 0.5   
X−squared = 0, df = 1, p−value = 1   
alternative hypothesis: true p is not equal to 0.5   
95 percent confidence interval:   
 0.43672 0.56328   
sample estimates:   
  p   
0.5

**Sampling Size of Population Proportion**

The quality of a sample survey can be improved by increasing the sample size. The formula below provide the sample size needed under the requirement of population proportion interval estimate at (1 − *α*) confidence level, margin of error *E,*and *planned*proportion estimate *p*. Here, *zα∕*2 is the 100(1 − *α∕*2) [percentile](http://www.r-tutor.com/node/38) of the [standard normal distribution](http://www.r-tutor.com/node/58).

         2
n = (zα∕2)-p(1−-p)
         E2


**Problem**

Using a 50% planned proportion estimate, find the sample size needed to achieve 5% margin of error for the female student [survey](http://www.r-tutor.com/node/61) at 95% confidence level.

**Solution**

Since there are two tails of the normal distribution, the 95% confidence level would imply the 97*.*5*th* percentile of the normal distribution at the upper tail. Therefore, *zα∕*2 is given by qnorm(.975).

> zstar = qnorm(.975)   
> p = 0.5   
> E = 0.05   
> zstar^2 ∗ p ∗ (1−p) / E^2   
[1] 384.15

**Answer**

With a planned proportion estimate of 50% at 95% confidence level, it needs a sample size of 385 to achieve a 5% margin of error for the survey of female student proportion.