Confidence Intervals

Vladislav Kargin

September 2022

Confidence intervals

Small sample confidence interval, one sample.

An example about lobsters

The carapace lengths of ten lobsters examined in a study of the infestation of the *Thenus orientalis* lobster by two types of barnacles, *Octolasmis tridens* and *O. lowei*, are given in the following table. (Table is attached.) Find a 95% confidence interval for the mean carapace length (in millimeters, mm) of *T. orientalis* lobsters caught in the seas in the vicinity of Singapore.

This is a small sample estimation problem for μ. Let us find a standard error and CI:

```
x = c(78,66,65,63,60,60,58,56,52,50)
mean(x) # sample mean
```

```
## [1] 60.8
```

```
sum((x-60.8)^2)/(10-1)  # sample variance
```

```
## [1] 63.51111
```

```
#alternatively
var(x)
```

```
## [1] 63.51111
```

```
sqrt(sum((x-60.8)^2)/(10-1))  # sample standard deviation
```

```
## [1] 7.969386
```

```
#alternatively
sd(x)
```

```
## [1] 7.969386
```

```
qt(0.975,9) # 0.025 percentage point
```

```
## [1] 2.262157
```

```
              # for t distribution with df=9
             #Note that it is different from 1.96, which is the critical
            #value for the normal distribution
qt(0.975,9)*sd(x)/sqrt(10)
```

```
## [1] 5.700955
```

```
#Confidence interval:
cat('(',mean(x)-qt(0.975,9)*sd(x)/sqrt(10),',',mean(x)+qt(0.975,9)*sd(x)/sqrt(10),')',sep = "")
```

```
## (55.09904,66.50096)
```

Answer: 60.8 ± 5.700955. Note that we have divided by $\sqrt{9}$ when we estimated S and then again by $\sqrt{10}$ when we estimated $\sigma_{\overline{X}}$. Do not forget the second division.

Alternatively, we can get the confidence interval by using the function t.test, which is used for testing hypothesis about the mean of a sample (or about the equality of means of two samples).

```
t.test(x,conf.level = 0.9)
```

```
##
##  One Sample t-test
##
## data:  x
## t = 24.126, df = 9, p-value = 1.727e-09
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  56.1803 65.4197
## sample estimates:
## mean of x
##      60.8
```

Two-sample example

To reach maximum efficiency in performing an assembly operation in a manufacturing plant, new employees require approximately a 1-month training period. A new method of training was suggested, and a test was conducted to compare the new method with the standard procedure. Two groups of nine new employees each were trained for a period of 3 weeks, one group using the new method and the other following the standard training procedure. The length of time (in minutes) required for each employee to assemble the device was recorded at the end of the 3-week period. The resulting measurements are as shown in Table 8.3 (see the book). Estimate the true mean difference ($\mu_1 - \mu_2$) with confidence coefficient .95. Assume that the assembly times are approximately normally distributed, that the variances of the assembly times are approximately equal for the two methods, and that the samples are independent.

```
x = c(32,37,35,28,41,44,35,31,34)
y = c(35,31,29,25,34,40,27,32,31)

t.test(x,y, conf.level = 0.95, var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  x and y
## t = 1.6495, df = 16, p-value = 0.1185
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.045706  8.379039
## sample estimates:
## mean of x mean of y
##  35.22222  31.55556
```

```
t.test(x,y, conf.level = 0.95, var.equal = F) #now we do not assume that variance are equal. Note that
```

```
##
##  Welch Two Sample t-test
##
## data:  x and y
```

```
## t = 1.6495, df = 15.844, p-value = 0.1187
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.049486  8.382820
## sample estimates:
## mean of x mean of y
##  35.22222  31.55556
```

Confidence interval for variance

Suppose that you wished to describe the variability of the carapace lengths of this population of lobsters. Find a 90% confidence interval for the population variance σ^2.

```
x = c(78,66,65,63,60,60,58,56,52,50)
n = length(x)
n
```

```
## [1] 10
```

```
sum(( x-mean(x) )^2) # the numerator of sample variance and the CI LB and UB
```

```
## [1] 571.6
```

```
#alternatively
(n - 1)*var(x)
```

```
## [1] 571.6
```

and then we calculate the denominators of the CI LB and UB

```
qchisq(0.05,9)
```

```
## [1] 3.325113
```

```
qchisq(0.95,9)
```

```
## [1] 16.91898
```

and the inteval is

```
(n - 1)*var(x)/qchisq(0.95,9)
```

```
## [1] 33.78455
```

```
(n - 1)*var(x)/qchisq(0.05,9)
```

```
## [1] 171.9039
```

This can be done also with the variance one-sample test. However, in order to use it, we should import the library of statistical functions name “EnvStats”. If it is not on your computer, you should install it either by using the command “install.packages” or by using the Tools menu.

```
#install.packages("EnvStats") #This needs to be done only once on your computer
```

```
library(EnvStats)
```

```
##
## Attaching package: 'EnvStats'
```

```
## The following objects are masked from 'package:stats':
##
##     predict, predict.lm
```

```
## The following object is masked from 'package:base':
##
##     print.default
```

```r
varTest(x, conf.level = 0.90)
```

```
## $statistic
## Chi-Squared
##       571.6
##
## $parameters
## df
##  9
##
## $p.value
## [1] 0
##
## $estimate
## variance
## 63.51111
##
## $null.value
## variance
##        1
##
## $alternative
## [1] "two.sided"
##
## $method
## [1] "Chi-Squared Test on Variance"
##
## $data.name
## [1] "x"
##
## $conf.int
##       LCL       UCL
##  33.78455 171.90394
## attr(,"conf.level")
## [1] 0.9
##
## attr(,"class")
## [1] "htestEnvStats"
```