07 July 2017

Tutorial Exercises

Exercise T.16

We want to analyse the daily expenditures of (adult) summer tourists in Lutenblag (Molwanîen). It is assumed that the expenditures are normally distributed, where neither the expectation μ nor the variance σ^2 are known. In a survey, 100 tourists were asked about their expenditures. From this data, it was found that on average 105 Euro per day were spent with a sample variance of 144. Determine a one-sided lower confidence interval at the level 0.95 for μ.

Hint:

```
round(qt(c(0.95,0.975), df = 99), 2)

## [1] 1.66 1.98
```

Homework Exercises

Exercise H.17

The amount of milk from 9 cows of a certain farm in one lactation period (about 300 milk days) was saved in the variable `Milk`. The following R commands were performed:

```
Milk

## [1] 5535 5659 4853 5020 5691 5564 5090 5217 6371

c(mean(Milk), var(Milk))

## [1]   5444.444 211893.028
```

It has been shown that the normal distribution $N(\vartheta_1, \vartheta_2)$ is a suitable distribution model for the amount of milk.

a) Using `R`, calculate the two-sided confidence interval for the expectation ϑ_1 of the assumed normal distribution (or of the normal distribution model) at the level $1 - \alpha = 0.90$ and 0.95. Explain why the 95% confidence interval is longer than the 90% confidence interval.

b) Assume now that you got the same empirical mean (or average) and the same empirical variance for sample sizes $n = 25$ and $n = 60$. Now calculate again by hand (i.e. use `R` only for computation of quantiles of normal distributions) a two-sided confidence interval at the level $1 - \alpha = 0.95$. Compare the lengths of the three confidence intervals calculated at the level 0.95 up to now.

c) Calculate a one-sided confidence interval for the expectation ϑ_1 in the case case $n = 60$ at the level $1 - \alpha = 0.96$ by hand (i.e. use `R` only for computation of quantiles of normal distributions). Decide yourself which one-sided interval is more interesting to the owner of the cows. Give reasons for your decision *Hint: Assume that the empirical average and the empirical variance remain the same.*

d) Finally calculate a two-sided confidence interval for the variance ϑ_2 for the given data set, i.e. for $n = 9$, at the level $1 - \alpha = 0.97$ by hand (i.e. use `R` only for computation of quantiles of normal distributions).

R-Exercises

Exercise R.9

With respect to confidence intervals, one often says: "The true unknown parameter ϑ of the distribution belongs to the confidence interval with probability $(1 - \alpha)100\%$". This statement is generally not correct since the unknown true parameter ϑ is a constant. The correct statement is that the confidence interval covers the true unknown parameter ϑ with probability $(1 - \alpha)100\%$. This means if we sample several samples and determine the corresponding confidence intervals then these confidence intervals cover the true unknown parameter ϑ in approximately $(1 - \alpha)100\%$ of the cases. Check this statement using the R-code below:

```r
# Specify parameters
m     <- 50
n     <- 100
theta <- 0.5
alpha <- 0.1
# Draw pseudo-random numbers
theta.hat <- rbinom( m, n, theta)

confidence_intervals <- matrix(rep(0, 2 * m), nrow = m)
# Loop to compute 50 confidence intervals with binom.test()
for( i in 1:m){
  confidence_intervals[i,] <- binom.test(theta.hat[i], n = n,
                                         conf.level = 1 - alpha)$conf.int[1:2]
}
# Count how many times the true parameter value is covered
sum(confidence_intervals[, 1] < theta & theta < confidence_intervals[, 2])/m
# Plot all confidence intervals
matplot(t(confidence_intervals), rbind(1:m, 1:m), type = "l", lty = 1,
        xlab="Confidence intervals", ylab="Simulation number")
# The true parameter value is  denoted by a yellow vertical line
abline(v = theta, lwd = 3, col = "yellow" )
```

How many out of the 50 confidence intervals for the success probability ϑ cover the true probability $\vartheta = 0.5$?