Chapter 3: Estimators and confidence intervals

1 Session 1

Objectives: understand the notion of confidence interval; compute confidence intervals for the parameters of a normal distribution.

Steps of the session:

1. Individual work: Practice with the instructions given on the web site.
2. Share and work: discuss what you understand.
3. Download and source function 'confint.normal.R'.
4. Individual work: Practice with the exercises of Session 1.
5. Share and work: discuss your scripts together.
6. Individual work at home: Finish exercises of this session.

Exercise 1.1. The function `confint.normal` takes a number of samples `N`, a sample size `n`, values of μ, σ, and a confidence `level`. It draws N samples of size n of the normal distribution $\mathcal{N}(\mu, \sigma)$. It computes the N confidence intervals for μ assuming σ known, for μ, assuming σ unknown, for σ. It splits the graphic window into 3. On the 3 subwindows, it represents the intervals by blue horizontal segments, and the true value of the parameter by a red vertical line. It returns the 3 proportions of intervals that contain the true value of the parameter. Run `confint.normal` at least for `N=100`, `n=10,100`, `mu=0,100`, `sigma=1,10`, `level=0.95,0.99,0.999`.

1. Download and source function 'confint.normal.R'.
2. How many intervals contain the true value of the parameter in each case? How does the width of the confidence intervals vary as `n` increases? as `level` increases? as `sigma` increases?
3. What difference can you see between the two cases σ known and σ unknown? Is that difference perceptible for `n` large?
4. What difference can you see between the confidence intervals for μ and for σ?

Exercise 1.2. Upload `dimer.csv`, read data description.

1. Assign column `log.dimer` to variable `D`, column `age` to variable `A`. Display a boxplot of `D` against `A`. How does log dimer vary with age?

2. Display a `qqnorm` plot of `D`. Compute estimates for the mean and standard deviation of `D`. Add the straight line with intercept equal to the estimated mean, slope equal to the estimated standard deviation.

3. Compute a 95% confidence interval for the mean of `D`, assuming the standard deviation unknown with the function 't.test(D)'.

4. Compute again the confidence intervals at level 98%, then 99%. How do the results change?

5. Assign to variable `D1` the subsample of `D` corresponding to age class 1. Compute a 98% confidence interval for the expectation of `D1`.

6. Assign to variable `D3` the subsample of `D` corresponding to age class 3. Compute a 98% confidence interval for the expectation of `D3`. What can you conclude?

7. Compute 98% confidence intervals for the proportion of negative values in `D1` and `D3`. What can you conclude?

8. Compute 95% confidence intervals for the proportion of negative values in `D1` and `D3`. What can you conclude?

Exercise 1.3. The fluorescein concentration of a given solution has been measured 90 times. An empirical mean of 4.38 mg/l and a standard deviation of 0.08 mg/l have been observed.

1. Give a confidence interval for the true concentration of the solution, at confidence levels 0.95 and 0.99.

2. What would be the confidence level of the interval 4.38 ± 0.01?

3. What should the sample size be for the interval 4.38 ± 0.01 to have confidence level 0.98?

Exercise 1.4. In order to study the influence of X-rays on the spermatogenesis of Bombyx mori, males have been exposed to radiation on the second day and on the fourth day of the larval stage. These males have been mated with non exposed females, and the number of fertile eggs laid by the females have been counted: out of a total of 5646 eggs laid, 4998 were fertile. In a control group of non exposed males and females, 5834 fertile eggs out of 6221 were obtained.

1. Find a 95% confidence interval for the proportion of fertile eggs after radiation exposure of males.

2. Find a 95% confidence interval for the proportion of fertile eggs of non exposed couples.

3. Do the two intervals overlap? What does this tell about the influence of radiation exposure on fertility?

2 Session 2

Objectives: compute confidence intervals on your dataset.

Steps of the session:

1. Team work: Compute confidence intervals on your project dataset. Compare sub-groups of individuals.

2. Team work: Write the report corresponding to your results

3. Individual work at home: Read the chapter *Statistical tests*.