MATH 4753 Laboratory 11: Classical confidence intervals

In this lab we will learn how to use R to create confidence intervals. This will cover the remainder of chapter 7. Bootstrap intervals are easy to make once an R script has been created, because it relies on simulation each run of the algorithm will result in slightly different intervals. Classical analytical methods give constant results and everyone will calculate the same values for the endpoints of the confidence intervals (provided there are no arithmetic mistakes).

There are two interpretations that you will need to grapple with:

* Theoretical – a long run probabilistic understanding. This relates to many future samples.
* Practical – a statement of “confidence”. This relates to the sample or samples in hand.

You will need practice at using these interpretations.

For this course it is NOT required that you memorize formulae for confidence interval calculation. These will always be supplied. You will need to select the correct formula, make the calculation and interpret the interval made. The most important part is the interpretation. The selection of the correct formula proceeds from a good understanding of the sampling situation and the intent of the experiment.

There are some easy ways to calculate intervals in R, however they rely on the use of a function that is made to primarily carry out a test (Ch 8). The confidence interval is a by-product and it is this that we want to retrieve.

### Tasks

There will be two files you need to upload:

* An html
* An RMD

Do not make unnecessarily big documents.

**Note: All plots you are asked to make should be recorded through RMD**

**You are expected to adjust the functions as needed to answer the questions within the tasks below.**

* Task 1
  + Make a folder LAB11
  + Download the file “lab11.r”
  + Place this file with the others in LAB11.
  + Start Rstudio
  + Open “lab11.r” from within Rstudio.
  + Go to the “session” menu within Rstudio and “set working directory” to where the source files are located.
  + Issue the function getwd() and copy the output here.
  + Create your own R file and record the R code you used to complete the lab.
* Task 2
  + This relates to one sample from a population where we want to estimate .
  + Suppose we wish to build a confidence interval for the mean.
  + A ram press makes washers for agricultural machinery. The mean diameter (cm) of the washers is of interest. Suppose the following are the diameters of washers from a random sample taken from the press:

d=c(5.0581, 4.9707, 5.0893, 4.9334, 4.9777, 5.0285, 4.8555, 4.9565,

4.9769, 4.9722, 4.999, 4.9925, 4.9686, 5.0662, 4.9239, 4.9781,

5.0485, 5.0014, 4.9957, 5.0195, 5.0118, 4.9928, 5.0361, 5.0185,

4.9879)

* + Create the following intervals for the mean using R as a calculator
    - 95% ci
    - 90% ci
    - 80% ci
    - 50% ci
  + Create an 80% ci for the mean using t.test()using the option conf.level=0.80, you will need to make an object and then extract the confidence interval using $
  + Now we will concentrate on the population variance .
  + Create the following intervals for the population variance using R as a calculator
    - 95% ci
    - 90% ci
    - 80% ci
    - 50% ci
* Task 3
  + In this task we will examine confidence intervals for . Suppose two large (independent random samples are taken from two populations (not necessarily normal).
  + For small sample sizes and the following is the appropriate formula:
  + where
  + Suppose that two random samples of fish were caught, 20 blue cod fish and 15 snapper. The lengths of the fish were measured in inches. Assume pop. Variances are equal.

blue=c(21.65, 17.48, 20.1, 21.57, 14.82, 19.17, 21.08, 18.23, 22.93,

15.66, 20.89, 21.66, 18.5, 20.59, 18.63, 18.91, 19.53, 17.7, 16.5, 19.03)

* + snapper=c(31.65, 27.48, 30.1, 31.57, 24.82, 29.17, 31.08, 28.23, 32.93,
  + 25.66, 30.89, 31.66, 28.5, 30.59, 28.63)
  + Choosing the correct ci formula make a 95% confidence interval for by using R as a calculator.
  + What is the 95% ci for
  + Give a practical interpretation of the above interval.
    - **This interval is obviously an interval that we can state the mean lies between with 95% confidence, but also, this interval is the same formulae as a confidence interval with one set of data theoretically.**
  + N.B small sample ci for the difference in means with unequal variances are given on page 291 MS. We will use t.test() and let the software manage the correct formula.
  + Use t.test(,var.equal=TRUE)$conf.int to calculate a
    - 95% ci for
    - 85% ci for
    - 75% ci for
    - 25% ci for
  + What happens to the interval as the confidence level decreases?
    - **The size of the interval slims along with the confidence level decreasing. This is due to the allowance of a higher level of inaccuracy with each lower level.**
* Task 4
  + Paired samples (dependent) occur when two measurements are made on the same experimental units.
  + If the population variance is unknown and use
  + In 2012 MATH 4753 had two exams, the following were the results:

Exam1=c(40.98, 59.36, 46.69, 41.8, 61.63, 65.31, 62.96, 60.21, 56.89,

78.41, 53.44, 75.2, 60.54, 52.43, 41.41, 70.79, 73.55, 55.65,

61.43, 63.84, 58.07, 53.79, 54.45, 67.18, 44.46)

Exam2=c(50.22, 66.19, 58.75, 51.88, 66.61, 70.86, 74.25, 70.23, 69.55,

87.18, 63.62, 81.7, 70.5, 66.02, 51.35, 80.92, 85.65, 65.44,

74.37, 75.28, 67.86, 59.92, 64.42, 73.57, 57.15)

* + Make a 95% ci for
  + Interpret the interval practically
    - **We can say with 95% confidence that lies between the interval (53.38, 73.18)**
  + Make the following intervals using the appropriate options in t.test():
    - 90% ci for
    - 80% ci for
    - 70% ci for
    - 60% ci for
    - 10% ci for
* Task 5

,

* + ci for a population
  + The following table is derived from the data set NZBIRDS.
  + Verify the table by reading in the data set and using a suitable function. HINT: with(), table().

|  |  |  |
| --- | --- | --- |
| Bird population | Number of species sampled | Number of flightless species |
| Extinct | 38 | 21 |
| Non extinct | 78 | 7 |

* + Use R as a calculator and find a 95% ci for the difference in proportion of flightless birds for extinct and non-extinct species.
* Task 6
  + A ci for the ratio of two population variances, .
  + Two samples taken independently from Normal populations.
  + You will need qf(…,df1,df2) in R
  + Use the following simulated samples to obtain a 95% ci for the ratio (1 to 2) of the population variances.
  + set.seed(35);sam1=rnorm(25,mean=10,sd=5); set.seed(45);sam2=rnorm(34,mean=40,sd=8)
  + Use var.test() to create ci’s for the same population variances with the following confidences
    - 80%
    - 70%
    - 60%
    - 50%

################### LAB FINISHES HERE ###############################

* Task 7: Extra for experts!
  + Make a bootstrap() function for 2 sample intervals