MATH 4753 Laboratory 12:

Hypothesis testing.

### In this lab we will investigate hypothesis testing which fuses closely with two-sided confidence intervals. There are two hypotheses that you need to learn and understand:

1. The NULL hypothesis, which is the skeptic’s choice – “there is nothing going on, there is no difference”
2. The Alternate hypothesis, which is the researchers hunch, “Something is going on, there is a difference”

The function of hypothesis testing is to establish if there is evidence against the NULL. This is carried out by assuming the NULL is true and then determining if the data collected is consistent with this assumption.

When performing Hypothesis test there are two errors that are a part of the process. We call them type 1 and type 2 errors.

1. Type 1: Rejecting the NULL when it is true
2. Type 2: Accepting the NULL when it is false

We can represent these events with a table. The top row is the TRUTH

|  |  |  |
| --- | --- | --- |
| Action\ State of truth | NULL is ***TRUE*** | NULL is ***FALSE*** |
| Reject NULL | **Type 1 error** | Correct (POWER) |
| Accept NULL | Correct | **Type 2 error** |

The power of a test is defined as

### Tasks

All output made using RMD.

Knit html (and pdf if possible) – upload html and rmd documents.

**Note: All plots you are to be made using RMD and included in the knitted documents.**

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

* Task 1
  + Make a folder LAB12
  + Download the file “lab12.r”
  + Place this file with the others in LAB12.
  + Start Rstudio
  + Open “lab12.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 test a specific NULL hypothesis. There are assumptions made for this test and all others in chapter 8.
  + Lab12.R contains code and data needed to carry out this task and the others.
  + Using the sample x1, and assume that you know neither the population mean nor the population variance, perform a one sample t test using the function t.test() with the following NULL hypotheses: (Say if the NULL is rejected in favour of the alternate or not and copy in the P-value and ci)
    - ,
    - ,
    - ,
    - ,
    - ,
  + Make a boxplot of the data and plot the sample mean, and the 95% confidence interval (see the code in Lab 12.R)
  + Follow the logic:
  + We will investigate if the data is consistent with the NULL hypothesis by assuming that the data was generated from a population with mean =. We will use a **pivotal statistic** to **transform the data** since we will be able to obtain distributional information from it.
  + is the random variable with a t distribution and degrees of freedom.
  + If the P- value is small then the assumption that is true is unlikely or appears to contradict the sample data. If suitably small we will reject the NULL in favour of the alternate hypothesis .
  + Find , with
  + Use mypvalue() to create the plot of the p-value (alpha=0.05) using x1
  + What is the rejection region?
  + What is the p-value that will determine if we reject or not?
  + Is in the rejection region?
  + Construct bootstrap p-values using the same data set x1 for the following Hypotheses:
    - ,
    - ,
    - ,
    - ,
    - ,
  + Compare these results with the ones above. What do you conclude?
* Task 3
  + Two sample t tests are done using the same function t.test(), you must be careful with the NULL hypothesis. Remember to express all NULL’s as , is called the NULL value. You will use this in the t.test(…,mu=).
  + Let x and y be defined as in the R code. i.e

set.seed(30);x=rnorm(15,mean=10,sd=7)

set.seed(40);y=rnorm(20,mean=12,sd=4)

* + Perform an equality of variance test using var.test()
  + What do you conclude?
  + What will you assign for var.equal= ### inside t.test()?
  + Perform a t.test for the following hypotheses
  + Summarize what you have learnt.
* Task 4
  + Let x and y be defined as in the R code, namely

set.seed(30);x=rnorm(15,mean=10,sd=4)

set.seed(40);y=rnorm(20,mean=12,sd=4)

* + Perform a test of the equality of variances.
  + What will you assign for var.equal= ### inside t.test()?
  + Perform a t.test for the following hypotheses
  + Summarize what you have learnt.
* Task 5
  + Perform bootstrap testing using the function boot2pval() on Task 3 (No var.test() needed)
  + Record p values and plots
* Task 6
  + Perform bootstrap testing using the function boot2pval() on Task 4 (No var.test() needed)
  + Record p values and plots
* Task 7
  + Explain the lines indicated with a #
  + > t.test(x1,mu=23) # A
  + One Sample t-test # B
  + data: x1
  + t = 2.3563, df = 29, p-value = 0.02543 # C
  + alternative hypothesis: true mean is not equal to 23 #D
  + 95 percent confidence interval: #E
  + 23.30198 27.27320 #F
  + sample estimates:
  + mean of x
  + 25.28759 #G

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

* Task 8: Extra for experts!
  + MAKE YOUR OWN BOOTSTRAP FUNCTION THAT WILL CALCULATE P-VALUES TESTING see MS page 426