Week 8 Live Session

w203 Instructional Team

Feb 27, 2017

Announcements

You may take Quiz 1 at any time in the next week, before the week 9 live session.

- Time limit is 2 hours There are 15 questions.
- The quiz is located on the ISVC. Look under the **Assignments and Assessments** section of the Coursework tab.
- You instructors will supply the password you need to access the quiz.
- Note that you can go forwards and backwards through the questions. There is a navigation frame on the right that lets you skip around to any question.

T-Distribution Functions in R

Suppose $X_1, X_2, ..., X_n$ is a random sample from a normal distribution with mean μ and unknown standard deviation.

Let $T = (\bar{X} - \mu)/(s/\sqrt{n})$ where $\bar{X}$ is the mean and s is the sample standard deviation (sd) of $X_1, \ldots, X_n$.

Then T is distributed as a t-distribution with df= n-1 degrees of freedom.

In R, we can use the following functions to find the values needed for t-tests. Specifically, we can find the density, distribution function, quantile function and random generation for the t distribution with df degrees of freedom (and optional non-centrality parameter ncp):

dt gives the density, pt gives the distribution function, qt gives the quantile function, and rt generates random deviates.

Usage

dt(x, df, ncp, log = FALSE)

pt(q, df, ncp, lower.tail = TRUE, log.p = FALSE)

qt(p, df, ncp, lower.tail = TRUE, log.p = FALSE)

rt(n, df, ncp)

Arguments

x, q vector of quantiles.

p vector of probabilities.

n number of observations. If length(n) > 1, the length is taken to be the number required.

df degrees of freedom (> 0, maybe non-integer). df = Inf is allowed.

ncp non-centrality parameter delta; currently except for rt(), only for abs(ncp) <= 37.62. If omitted, use the central t distribution.

log, log.p logical; if TRUE, probabilities p are given as log(p).

lower.tail logical; if TRUE (default), probabilities are $P(X \leq x)$, otherwise, $P(X > x)$.

Practice with T-Distributions

What is the 2.5th and 97.5th percentiles of the t distribution with 5 degrees of freedom?

What is the 2.5th and 97.5th percentiles of the t distribution with 1 to 10 degrees of freedom?

P-Values and Confidence Intervals

The actual voltages of power packs labeled as 12 volts are as follows: 11.77, 11.90, 11.64, 11.84, 12.13, 11.99, and 11.77.

1) Find the mean and the standard deviation.
2) What is the critical value for a 95% confidence interval for this sample?
3) Find the 95% confidence interval for the sample.
4) Define a test statistic, i.e. a variable T, as a function of the values from (1) that tests whether the mean is 12.
5) Calculate the p-value using the t statistic. Should you reject the null hypothesis? (Also, could you predict what would happen from your answer to (3)?)
6) Suppose you were to use a normal distribution instead of a t-distribution to test your hypothesis. What would your p-value be for the z-test?

Executing T-tests in R

The file athlet2.Rdata contains data on college football games. The data is provided by Wooldridge and was collected by Paul Anderson, an MSU economics major, for a term project. Football records and scores are from 1993 football season.

```
load("athlet2.RData")
```

We are especially interested in the variable, dscore, which represents the score differential, home team score - visiting team score. We would like to test whether a home team really has an advantage over the visiting team.

a. The instructor will assign you to one of two teams. Team 1 will argue that the t-test is appropriate to this scenario. Team 2 will argue that the t-test is invalid. Take a few minutes to examine the data, then formulate your best argument.
b. Should you perform a one-tailed test or a two-tailed test? What is the strongest argument for your answer?
c. Execute the t-test and interpret every component of the output.
d. Based on your output, suggest a different hypothesis that would have led to a different test result. Try executing the test to confirm that you are correct.

Assumptions Behind the T-test

For each of the following scenarios, give your best argument for why a t-test may be *invalid*.

a. You have a sample of 50 CEO salaries, and you want to know whether the mean salary is greater than $1 million.

b. You have data on 1000 students that are sampled from 10 randomly chosen public universities, and you want to know whether public university students sleep less than 7 hours a night.

c. A nonprofit organization measures the percentage of students that pass an 8th grade reading test in 40 neighboring California counties. You are interested in whether the percentage of students that pass in California is over 80%

Your own T-Test Function

1. Using your understanding of the procedure for a one-sample t-test, write your own function to execute the test. You may use the following function header.

```
my.t.test = function(values, alpha = 0.05, mu = 0)
{
  # Your code here
}
```

2. Autogenerate a sequence of 20 values and use it as the input of your function. Did you reject or accept the null hypothesis?
3. How can you test that the function we created works properly?