- hadoop
- animation
- time series
- ggplot2
- ggplot
- LaTeX
- trading
- excel
- finance
- PCA
- googlevis
- how to import image file to R
- quantmod
- RStudio
- market research
- eclipse
- knitr
- rattle
- SQL
- coplot
- rcmdr
- tutorial
- map

Recent Posts

- R-based data maps in PowerBI
- Toying with models: The Game of Life with selection
- Batch Forecasting in R
- New CRAN package gunsales
- My Baby Boomer Name Might Have Been "Debbie"
- Mapping Birds with Choroplethr
- Network Visualization with Plotly and Shiny
- A Tale of Two Charting Paradigms: Vega-Lite vs R+ggplot2
- Mortality by Weekday and Age
- Upload plots as PNG file to your wordpress
- The first official Johannesburg R Users' Meetup!
- Create Vega-Lite specs & widgets with the vegalite package
- Control Structures Loops in R
- Principal Component Analysis using R
- Japans ageing population, animated with R

Other sites

- Jobs for R-users
- Statistics of Israel
- SAS blogs

Two sample Student's t-test #1

July 24, 2009
By Todos Logos

(This article was first published on **Statistic on aiR**, and kindly contributed to R-bloggers)

t-Test to compare the means of two groups under the assumption that both samples are random, independent, and come from normally distributed population with unknow but equal variances

Here I will use the same data just seen in a previous post. The data are given below:

A: 175, 168, 168, 190, 156, 181, 182, 175, 174, 179
B: 185, 169, 173, 173, 188, 186, 175, 174, 179, 180

To solve this problem we must use to a Student's t-test with two samples, assuming that the two samples are taken from populations that follow a Gaussian distribution (if we cannot assume that, we must solve this problem using the non-parametric test called **Wilcoxon-Mann-Whitney test**; we will see this test in a future post). Before proceeding with the *t-test*, it is necessary to evaluate the sample variances of the two groups, using a **Fisher's F-test** to verify the *homoskedasticity* (*homogeneity of variances*). In R you can do this in this way:

```
a = c(175, 168, 168, 190, 156, 181, 182, 175, 174, 179)
b = c(185, 169, 173, 173, 188, 186, 175, 174, 179, 180)

var.test(a,b)

     F test to compare two variances

data: a and b
F = 2.1028, num df = 9, denom df = 9, p-value = 0.2834
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
   0.5223017 8.4657950
sample estimates:
 ratio of variances
2.102784
```

We obtained p-value greater than 0.05, then we can assume that the two variances are homogeneous. Indeed we can compare the value of F obtained with the tabulated value of F for alpha = 0.05, degrees of freedom of numerator = 9, and degrees of freedom of denominator = 9, using the function `qf(p, df.num, df.den)`:

```
qf(0.95, 9, 9)
[1] 3.178893
```

Note that the value of F computed is less than the tabulated value of F, which leads us to accept the null hypothesis of homogeneity of variances.
NOTE: The F distribution has only one tail, so with a confidence level of 95%, `p = 0.95`. Conversely, the *t-distribution* has two tails, and in the R's function `qt(p, df)` we insert a value `p = 0975` when you're testing a two-tailed alternative hypothesis.

Then call the function t.test for homogeneous variances (`var.equal = TRUE`) and independent samples (`paired = FALSE`: you can omit this because the function works on independent samples by default) in this way:

```
t.test(a,b, var.equal=TRUE, paired=FALSE)

Two Sample t-test

data: a and b
t = -0.9474, df = 18, p-value = 0.356
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -10.93994 4.13994
sample estimates:
mean of x mean of y
    174.8     178.2
```

We obtained p-value greater than 0.05, then we can conclude that the averages of two groups are significantly similar. Indeed the value of t-computed is less than the tabulated t-value for 18 degrees of freedom, which in R we can calculate:

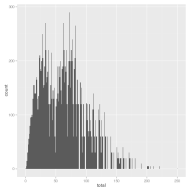
```
qt(0.975, 18)
[1] 2.100922
```

This confirms that we can accept the null hypothesis H0 of equality of the means.

Related

Two sample Z-test
Similar post

Two sample Student's t-test #2
In "R bloggers"



What's the smallest amount you can't make with 5 coins ?
In "R bloggers"

To **leave a comment** for the author, please follow the link and comment on their blog: **Statistic on aiR**.

R-bloggers.com offers **daily e-mail updates** about R news and tutorials on topics such as: Data science, Big Data, R jobs, visualization (ggplot2, Boxplots, maps, animation), programming (RStudio, Sweave, LaTeX, SQL, Eclipse, git, hadoop, Web Scraping) statistics (regression, PCA, time series, trading) and more...

If you got this far, why not **subscribe for updates** from the site? Choose your flavor: e-mail, twitter, RSS, or facebook...

Tags: F-test, Fisher, Homoskedasticity test, Independent data-set, t-test, Two sample

Comments are closed.

Recent popular posts

- Principal Component Analysis using R
- Network Visualization with Plotly and Shiny
- A Tale of Two Charting Paradigms: Vega-Lite vs R+ggplot2
- Control Structures Loops in R

Most visited articles of the week

1. Is deep learning a Markov chain in disguise?
2. Installing R packages
3. In-depth introduction to machine learning in 15 hours of expert videos
4. Using apply, sapply, lapply in R
5. How to Learn R
6. First steps with Non-Linear Regression in R
7. Principal Component Analysis using R
8. How to perform a Logistic Regression in R
9. Computing and visualizing PCA in R

Sponsors

Jobs for R users

- Sr. Manager Data Science @ LA
- Data Scientist
- Root word and Noun Phrase Extraction
- Data analyst
- Statistician R developer for agriculture statistics
- Freelance Intern R and shiny developer
- R Shiny Developer

Search & Hit Enter
Full list of contributing R-bloggers