

R- stats package

Online tutorial: <http://www.statmethods.net/stats/index.html>

Example (you can run the code – in blue - and observe the results):

We have results for four stochastic algorithms for GRN model inference, representing the difference (MSE) between real and inferred parameters: file “algs.txt”, download from:

<http://www.computing.dcu.ie/~asirbu/algs.txt>

We want to compare these four algorithms, and order them based on performance (the lower the MSE, the better). The input file contains results for each algorithm (columns) in multiple runs (last algorithm was run less times, so has missing values).

Read in data into a data.frame:

```
> data=read.table('algs.txt')
```

Show descriptive statistics:

```
>summary(data)
```

Descriptive statistics indicate alg2 as best, followed by alg3, alg1 and alg4 (based on median and mean values). However, this conclusion is not necessarily *statistically significant*, as we are dealing with a stochastic algorithm.

Visualise data using boxplots (gives an indication of data distribution):

```
>boxplot(data)
```

This shows again alg2 as best. We can perform t-tests to check if indeed there is significant difference between alg2 and alg1:

```
>t.test(data$alg1, data$alg2)
Welch Two Sample t-test

data: data$alg1 and data$alg2
t = 4.0643, df = 24.651, p-value = 0.0004293
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.057375 3.233082
sample estimates:
mean of x mean of y
2.2099976 0.0647695
```

This shows significant difference with 95% confidence interval. Now, let's compare alg1 and alg3:

```
>t.test(data$alg1, data$alg3)
      Welch Two Sample t-test

data:  data$alg1 and data$alg3
t = 1.1528, df = 41.458, p-value = 0.2556
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.5432524  1.9894236
sample estimates:
mean of x mean of y
 2.209998  1.486912
```

In this case, the p-value is large, so the difference between the two algorithms is not significant.

The 'stats' package also has methods for ANOVA, MANOVA and other statistical tools that you can use for your project (see tutorial at the top of this document for an introduction and examples).