

Tilastollisten ja numeeristen menetelmien jatkokurssi, Harkka 1

1. Perform the following steps:

- Create artificial data samples ($n_1 = 20$, $n_2 = 18$) from two normal distributions with mean values ($\mu_1 = 7.2$, $\mu_2 = 8.5$) and standard deviations ($\sigma_1 = 4.2$, $\sigma_2 = 6.1$) (use Matlab's function `randn` to generate the samples).
- From these samples calculate the sample means μ_{s1} , μ_{s2} and their difference $D = \mu_{s1} - \mu_{s2}$. Calculate also the sample variances var_{s1} , var_{s2} and their quotient $Q = var_{s1}/var_{s2}$.
- Repeat the procedure 50, 1000, 10000 times. What would you expect in the behaviour of D and Q ? Verify the result by plotting the histograms of both (use Matlab's function `histogram`).

2. Load data `examgrades` (given in Matlab). The data contains 5 columns. The following piece of code tests whether the first column in the matrix comes from a Student's t-distribution with a location parameter of 75, a scale parameter of 10, and one degree of freedom.

```
load examgrades;
x = grades(:,1);
% prepare theoretical CDF matrix
test_cdf = [x,cdf('tlocation',x,75,10,1)];
h = kstest(x,'CDF',test_cdf)
```

Perform the test on all the columns of the data. What are the test results and the respective p-values?

3. Load the given data file `RandD.mat` (provided in Moodle). The data represents companies' expenditure for R&D per revenue level.

- Use functions `expfit` and `wblfit` (or `fitdist` for both) to estimate parameters of exponential and Weibull distributions based on the data.
- Plot normalized histogram of the data (use function `histogram` and change its `Normalization` attribute to `'pdf'`). Use `pdf` function to plot in the same figure the theoretical distributions – exponential and Weibull with parameters estimated from the data.
- Use `kstest` and `adtest` functions to test whether the data comes from any of these two distributions.
- For each of the two, plot the empirical CDF of the data, and the CDF of the hypothesized theoretical distribution in the same plot. (you can use Matlab's functions, for instance `ecdf`, `cdfplot`, `cdf`).