Experiment Design – Case Studies

Alex Tsukuba (99999999)

Summary

The four case studies below are examples of experiments that can be analysed by the techniques that we have discussed in class. In each case, a description of the experiment goals and methods, as well as a sample data set, is given.

Your task is to perform the analysis of the data set following the descriptions and goals for the experiment. After performing the necessary calculations, present the results and draw appropriate conclusions.

This file (completed for all four case studies) must be submitted as report 3 until June 26th on MANABA. Please submit *both* the Rmd file, and the resulting PDF or HTML file.

Activities

Your analysis should follow the following simple pattern:

1. Describe the experimental design required to answer the technical question of interest. Detail the hypotheses being tested and the relevant design for testing those hypotheses.
2. Perform the statistical analysis using the observations contained in the data file provided. This includes:

    a. Perform the actual test of statistical significance;
    b. Estimate the effect size (including the confidence interval);
    c. Check the assumptions of your test;
    d. Discuss the power of your test, **if relevant**
    e. Describe your conclusions and recommendations. (what conclusions can we make? How can we improve the experiment?)

Case Study 1 – Coin Counting

The experiment

In a classroom we collected estimates from the students regarding the following two quantities:

1. The total number of coins contained in a glass recipient A;
2. The total monetary value contained in a glass recipient B;

The collection of the first estimates (*regarding the number of coins in recipient A*) was made in a blind fashion. In other words, each student made their estimate in writing, without interacting with the other students.

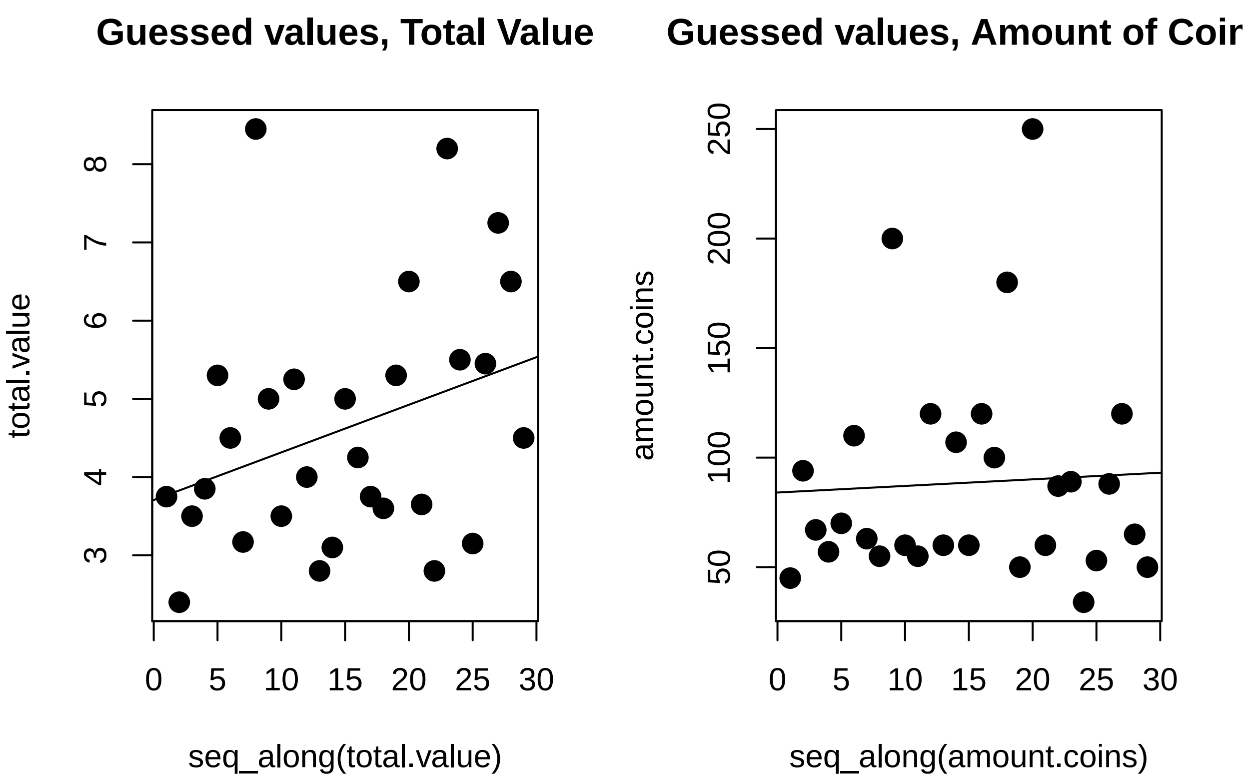
After a while, the collection of the second estimates (*regarding the total value contained in recipient B*) was made in an open fashion. In other words, each student made an estimate out loud, sequentially.

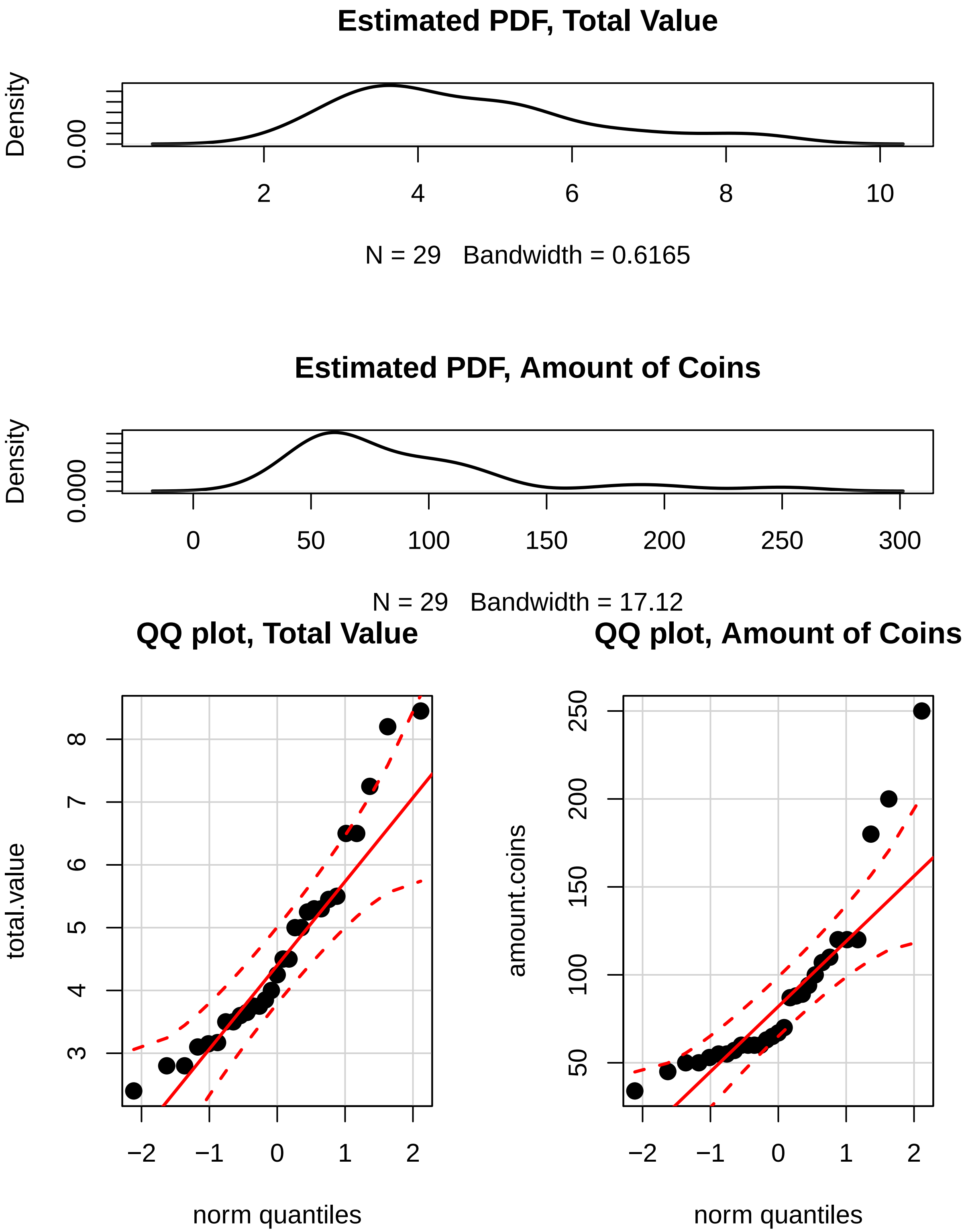
In this case study, we want to investigate whether this particular classroom is a good estimator of small quantities of coins. Based on the data collected, we must try to answer the following two questions:

1. How many coins are contained in recipient A?
2. How much money is contained in recipient B?

The data related to this experiment is located in two files: *01_coins_value.csv* and *01_coins_amount.csv*.

Data Analysis





Testing for Normality and independence of the data

```
##
##  Shapiro-Wilk normality test
##
## data:  total.value
```

```
## W = 0.91928, p-value = 0.02923
```

```
##
##  Shapiro-Wilk normality test
##
## data:  amount.coins
## W = 0.79755, p-value = 7.551e-05
```

```
##
## Call:
## lm(formula = total.value ~ seq_along(total.value))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2457 -0.9914 -0.1002  0.7455  4.2563
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             3.70680    0.57906   6.401  7.4e-07 ***
## seq_along(total.value)  0.06086    0.03371   1.805   0.0822 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.519 on 27 degrees of freedom
## Multiple R-squared:  0.1077, Adjusted R-squared:  0.07463
## F-statistic: 3.258 on 1 and 27 DF,  p-value: 0.08223
```

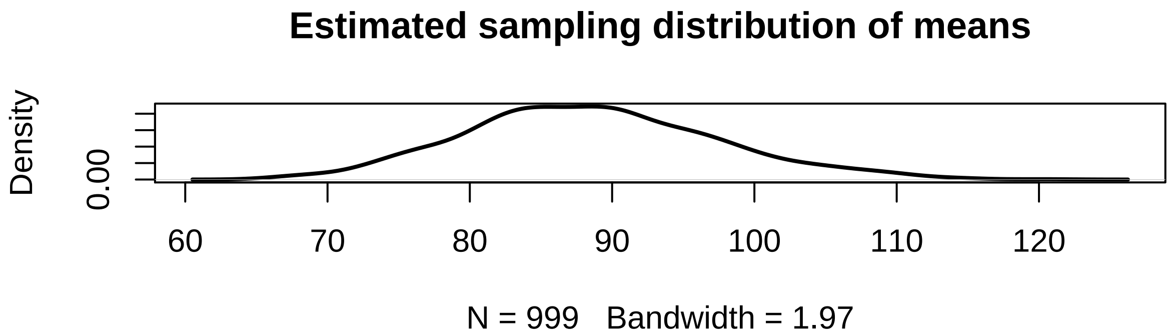
```
##  lag Autocorrelation D-W Statistic p-value
##    1      0.07397356      1.841015   0.628
##  Alternative hypothesis: rho != 0
```

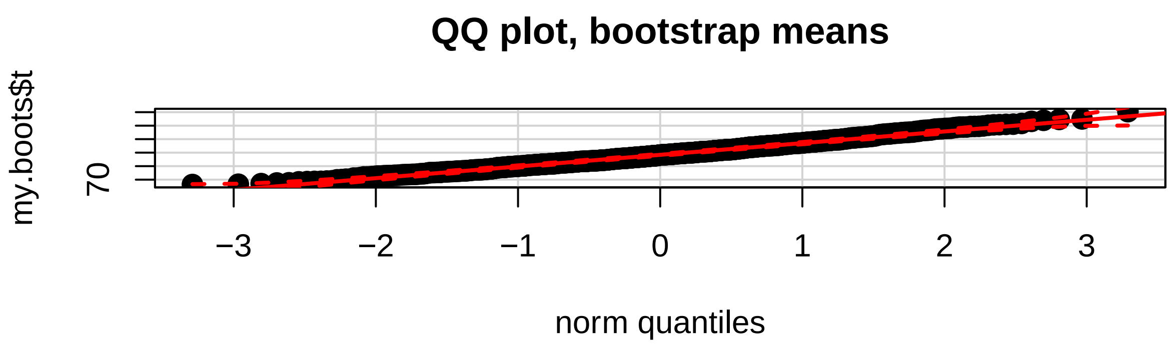
Calculating confidence intervals

```
## [1] 4.019002 5.220309
## attr(,"conf.level")
## [1] 0.95
```

```
##
## Attaching package: 'boot'
```

```
## The following object is masked from 'package:car':
##
##     logit
```






```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 999 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = my.boots, conf = 0.95, type = "norm")
##
## Intervals :
## Level      Normal
## 95%   ( 71.42, 105.90 )
## Calculations and Intervals on Original Scale
```

Discussion about the results

Discuss here what you learned from your analysis.

Case Study 2 – Algorithm Analysis

The experiment

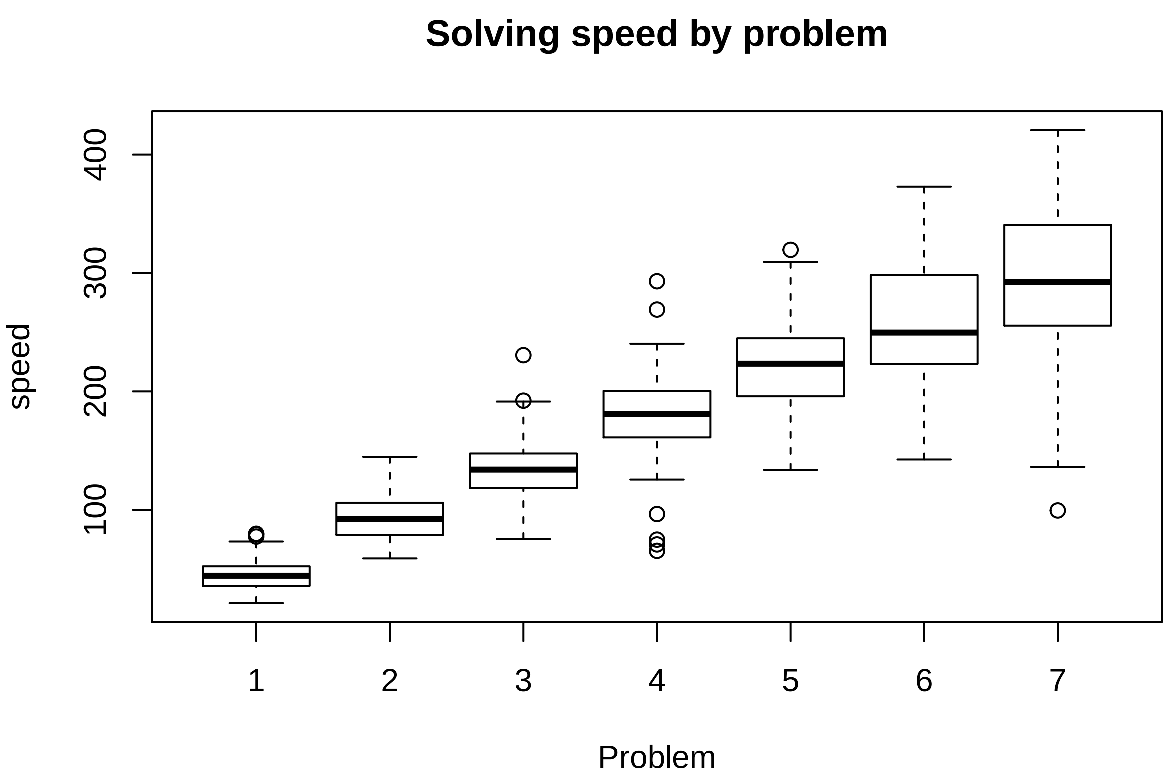
A student was performed to compare two algorithms, *A* and *B*, in regards to their speed when solving optimization problems.

The student ran both algorithms 20 times on 7 representative problems in a benchmark, and collected the running time of each execution. The student made sure that the order of the runs were randomized, and that as much as possible the execution environment was similar.

The results of this experiment are locates in the file *02_algorithm.csv*, and the student now wants to know if there is any significant difference between the running times of both algorithms, of an order of at least 10 seconds.

Data Analysis

```
##  Problem Algorithm      Time
##  1:60    A:210     Min.   : 21.23
##  2:60    B:210     1st Qu.:100.30
##  3:60              Median :174.47
##  4:60              Mean   :175.48
##  5:60              3rd Qu.:240.96
##  6:60              Max.   :420.61
##  7:60
```



The performance of both algorithms seem to depend strongly on the problem, so we summarize the measure times for each problem:algorithm combination.

```
##    Problem Algorithm      Time
```

```
## 1          1          A  37.62860
## 2          2          A  83.93047
## 3          3          A 126.01392
## 4          4          A 169.77983
## 5          5          A 211.75310
## 6          6          A 249.12982
## 7          7          A 289.73297
## 8          1          B  53.38058
## 9          2          B 103.93440
## 10         3          B 146.95828
## 11         4          B 187.67396
## 12         5          B 233.59827
## 13         6          B 266.38702
## 14         7          B 296.78549
```

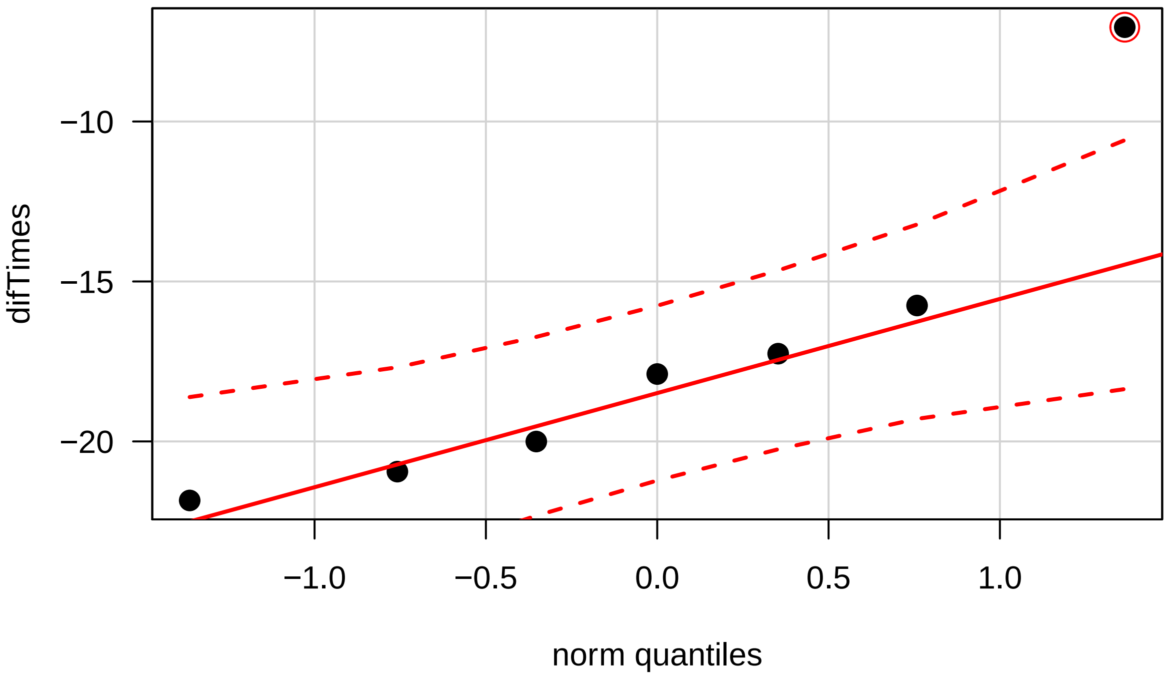
We are interested in testing whether there is a difference in performance between the two algorithms. In this case, we formulate our hypothesis over the size of the difference between the algorithms:

$$\begin{cases} H_0 : \mu_A - \mu_B = 0 \\ H_1 : \mu_A - \mu_B \neq 0 \end{cases}$$

And perform a paired T-test:

```
##
##  Paired t-test
##
## data:  Time by Algorithm
## t = -9.1585, df = 6, p-value = 9.54e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.85862 -12.64118
## sample estimates:
## mean of the differences
##                -17.2499
```

The result of the paired T-test allows us to reject the null hypothesis. Now we need to test the assumptions made by the T-test: Normality of the residuals and Independence




```
##
##   Shapiro-Wilk normality test
##
## data:  difTimes
## W = 0.83866, p-value = 0.09655

## [1] "P-value of the t-test without the outlier 6.17974290361148e-06"

## [1] "CI of the mean difference without the outlier -21.4185570144635"
## [2] "CI of the mean difference without the outlier -16.4803652385927"
```

Discussion about the results

What can you conclude based on the observed results?

Case Study 3 – Biodiversity

The Experiment[1]

A field ecologist is interested in examining the effect of sewage disposal in the diversity of invertebrates normally found in rivers. She dispatches a team of graduate students to count how many different species can be found in 100-Liter samples of water from seven rivers that have untreated sewage disposal at some point in their course, from both 100m before and after the point where the sewage is disposed. For each river, 10 samples are collected (one each month, from february to november). The simulated data is available in the file *03_biodiversity.csv*.

Based on the available sample, your task is to answer the following question:

Does the disposal of raw sewage in rivers affect the mean number of species?

Data Analysis

Discussion about the results

[1]Based on M.J. Crawley (2007), “The R Book”, Ch. 8.

Case Study 4 – Laboratory Compliance

The Experiment[2]

A ballistics laboratory is in the process of being certified for the evaluation of shielding technology. As part of this process, the laboratory must provide evidence that a certain callibration procedure produces results that are consistent with a reference equipment from the Department of Defence.

The procedure in question consists of shooting a standardized steel cube against a 320mm-thick alluminum target and measuring the resulting hole area. From previous measurements under similar conditions, the standard deviations of the observations of this laboratory and of the Department of Defence can be roughly estimated as:

- $\hat{\sigma}_{Lab}$: 5 mm^2
- $\hat{\sigma}_{DD}$: 10 mm^2

The certification authority demands that the mean hole area generated by this procedure in the lab be the same as the one from the reference equipment, and tolerates deviations no greater than 4 mm^2. Since this certification is quite important for the laboratory, the engineer in charge of the process decides that he wants a significance level $\alpha = 0.01$ and a power of $(1 - \beta) = 0.9$ for the smallest effect size of practical significance.

Your task is to answer the following question:

Is the mean hole size generated by the laboratory in conformity with the one generated by the reference equipment?

The data related to this experiment is located in two files: *04_compliance.csv*.

Data Analysis

Discussion about the results

[2] Adapted from Mason *et al.*, *Statistical Design and Analysis of Experiments with Applications to Engineering and Science*, Wiley-Interscience 2003.