EEE933 - Design and Analysis of Experiments

Case Study 01

This version was compiled on September 13, 2018

Experiment: performance of a new software version.

The experiment

The **current** version of a given system is known, based on extensive past experience, to have a distribution of execution costs with populational mean $\mu = 50$ and populational variance $\sigma^2 = 100$ (the specific units are not important in this particular experiment). Notice that these are considered populational parameters, not sample estimates.

A **new** version of this software is developed, and we wish to investigate whether it results in *performance improvements* (e.g., smaller mean execution cost and/or smaller variance) in relation to the current standard. To investigate this particular question, an experimental analysis will be performed.

To **simulate** the data collection procedures for the **new** software version, the following routines will be used. First, to set up the data collection procedure, use:

```
# Install required package and set up simulation
install.packages("ExpDE") # <-- you only need to install it once
```

```
# Set up the data-generating procedure
library(ExpDE)
mre <- list(name = "recombination_bin", cr = 0.9)
mmu <- list(name = "mutation_rand", f = 2)
mpo <- 100
mse <- list(name  = "selection_standard")
mst <- list(names = "stop_maxeval", maxevals = 10000)
mpr <- list(name  = "sphere", xmin  = -seq(1, 20), xmax  = 20 + 5 * seq(5, 24))
my.seed <- 1998  # <--- ATTENTION: USE THE BIRTH YEAR OF YOUNGEST TEAM MEMBER
```

A **single** observation of running cost for the **new** version can be obtained by running:

```
# Generate a single observation
ExpDE(mpo, mmu, mre, mse, mst, mpr, seed = my.seed,
      showpars = list(show.iters = "none"))$Fbest
```

```
#  [1] 61.18181
```

Activities

For the test on the mean cost. For this test, assume a desired significance level $\alpha = 0.01$, a minimally relevant effect size of $\delta^* = 4$, and a desired power of $\pi = 1 - \beta = 0.8$. Each team must perform the following activities:

- Define the statistical hypotheses to be tested (null/alternative).
- Define the sample size to be used in this experiment, and collect the sample.
- Test the hypotheses and decide for rejecting (or not) the null hypothesis.
- Calculate the confidence interval on the mean.
- Validate and discuss the assumptions of the test.
- Discuss the power of the test (if needed), and the adequacy of the sample size used for this particular test.

For the test on the variance of the cost. For this test, assume a desired significance level $\alpha = 0.05$. Assume also that reductions in standard deviation is a secondary benefit, to be investigated using the same observations collected for the first test. Each team must perform the following activities:

- Define the statistical hypotheses to be tested (null/alternative).
- Test the hypotheses and decide for rejecting (or not) the null hypothesis.
- Calculate the confidence interval on the variance.
- Validate and discuss the assumptions of the test.

After performing the activities related to each test individually, the team must:

- Draw conclusions and provide recommendations regarding the adoption (or not) of the new software version.
- Discuss possible ways to improve this experiment.

[Bonus question]. Can you also provide a 90% tolerance interval (using a confidence level $\alpha = 0.05$) for the population of running costs? Notice that the data may not be Normal.

Report

Each team must prepare a short report detailing the experiment and the analysis performed. The report will be evaluated according to the following criteria:

- Use of the predefined format (see below);
- Reproducibility of the analyses;
- Technical quality;
- Logical structure;
- Correct use of language (grammar, orthography, *etc.*);

The report must **necessarily** be prepared using R Markdown, and must contain the full code needed to reproduce the analysis performed by the team, embedded in the form of *code blocks*. Each team must deliver the following files:

- The report file, compiled in **.pdf** .
- The original (**.Rmd** source file) of the report.
- The data file (**.csv**) generated in this experiment.

The **.Rmd** file must be able to be recompiled, if needed (tip: save your **.Rmd** file using UTF-8 encoding, to prevent compilation problems in other operational systems.

Report templates are available on https://git.io/vHk0F, and an example of report structure can be consulted on https://git.io/vHk0j. This document can also be used as a template.

Important: Please include in the report the roles of each team member (Coordinator, Recorder, Checker and, for 4-member teams, Monitor)

Important: Reports can be prepared in either Portuguese or English.

Deadline

The report files (pdf + rmd + csv) must be compressed into a single file (use **.ZIP**) and uploaded to the activity **Case Study 01** on Moodle, until **Monday, September 24 2018, 11:59p.m.**. After that deadline the system will be closed.

Important: Only ONE submission is required for each team.

Important: Reports will NOT be received by e-mail or in printed form.