

# APTS Design of Experiments - Assessment

## September 2020

Dave Woods (D.Woods@southampton.ac.uk  
(mailto:D.Woods@southampton.ac.uk) ; <http://www.southampton.ac.uk/~davew>  
(<http://www.southampton.ac.uk/~davew>))  
Statistical Sciences Research Institute, University of Southampton

These questions provide an opportunity to review concepts from the course and to practise the application of Design of Experiments methods in R.

The functions referred to in Questions 1 and 3 below are available in the `apts.doe` package, which can be installed from `GitHub`. Make sure you have the latest version of this package before attempting the assessment.

```
library(devtools)
install_github("statsdavew/apts.doe", quiet = T)
library(apts.doe)
```

**1. Fractional factorial designs** This question is based on a real experiment performed by Tribologists at Southampton.

Engineers in the Materials Science group at the University of Southampton wish to run an experiment to investigate the impact of contamination on material corrosion in a “pin-on-disc” tribometer (<https://en.wikipedia.org/wiki/Tribometer>) (<https://en.wikipedia.org/wiki/Tribometer>).

There are six factors that can be varied, each of which can take two values:

- A: disc material - steel (coded -1) or silicon (+1)
- B: pin material - steel (-1) or silicon (+1)
- C: addition of soot - 0% by weight (-1) or 10% by weight (+1)
- D: oxidation - 0 hours (-1) or 10 hours (+1)
- E: addition of acid - 0mM (-1) or 25mM (+1)
- F: moisture content - 0% (-1) or 2.5% (+1)

The experimenters can afford  $n = 16$  runs. Prior knowledge from previous experiments suggests the following interactions may be important:

- C:D
- C:E
- D:E
- E:F

Design an appropriate fractional factorial experiment.

The function `material.response` simulates data from this experiment. As input, it takes a vector of six values, each either `-1` or `+1`, representing a single run of the experiment. Use this function to generate data for your chosen design, and perform an appropriate analysis. What conclusions can you draw?

**Note:** the `Effects` function will not work with `lm` objects which are over-parameterised (ie which contain missing (NA) coefficients). You will therefore need to make sure your `lm` object is not over-parameterised to draw main and interaction effect plots. As the design is orthogonal (in the non-aliased effects), the easiest way to do this is to refit a linear model in just the effects of interest, eg to plot the interaction between A and B:

```
materials.lmAB <- lm(y ~ A*B, data = materials.frame)
```

Of course, this doesn't break the aliasing in the design! So in this case, `AB` is probably aliased with another interaction (depending on the design you have chosen).

## 2. Bayesian optimal design

In Practical Lab 2, we found designs for a two-compartment model, of the type used in pharmacokinetics studies (the practical sheet and some example solutions can be found here) (<https://statsdavew.github.io/apts.doe/>).

Find Bayesian  $D$ -optimal designs (using the asymptotic approximation) for this model for  $n = 11, \dots, 20$  design points under the prior distributions

1.  $\theta_1 \sim \text{Uniform}(0.1, 0.2)$ ,  $\theta_2 \sim \text{Uniform}(0.3, 8.3)$ ,  $\theta_3 = 20$
2.  $\theta_1 \sim \text{Uniform}(0.1, 0.5)$ ,  $\theta_2 \sim \text{Uniform}(0.3, 18.3)$ ,  $\theta_3 = 20$
3.  $\theta_1 \sim \text{Uniform}(0.1, 1)$ ,  $\theta_2 \sim \text{Uniform}(0.3, 28.3)$ ,  $\theta_3 = 20$

Assess how the expected utility changes with  $n$ , and graphically compare the designs obtained under different prior distributions.

## 3. Computer experiments

A common mathematical model used to demonstrate computer experiment methodology in the literature describes the flow of ground water through a borehole from an upper aquifer to a lower aquifer separated by a impermeable rock layer. The model has eight input variables:

- $r_w \in [0.05, 0.15]$ : radius of the borehole ( $m$ )
- $r \in [100, 50000]$ : radius of influence ( $m$ )
- $T_u \in [63070, 115600]$ : transmissivity of the upper aquifer ( $m^2/yr$ )
- $H_u \in [990, 1110]$ : potentiometric head of the upper aquifer ( $m$ )
- $T_l \in [63.1, 116]$ : transmissivity of the lower aquifer ( $m^2/yr$ )
- $H_l \in [700, 820]$ : potentiometric head of the lower aquifer ( $m^2/yr$ )
- $L \in [1120, 1680]$ : length of the borehole ( $m$ )
- $K_w \in [9855, 12045]$ : hydraulic conductivity of borehole ( $m/yr$ )

The output from the model is  $y$ , water flow rate ( $m^3/yr$ ). The function `borehole` implements this model. The input to the model should be a numeric vector of length 8, giving values for the different input variables on the ranges defined above.

Design a suitable computer experiment, use the `borehole` function to perform the experiment, and fit a Gaussian process (GP) emulator. From the GP output, which variables look to have most impact on the response?

(I would suggest you fit the Gaussian process with all input variables scaled to  $[0, 1]$  to improve interpretability).

Perform a Monte Carlo sensitivity analysis to obtain Sobol' indices and total effect indices for each variable, assuming the following distributions

- $r_w \sim N(0.10, 0.0161812^2)$
- $r \sim \text{Lognormal}(7.71, 1.0056^2)$  (where the parameters are the mean and variance of  $\log r$ )
- $T_u \sim \text{Uniform}(63070, 115600)$
- $H_u \sim \text{Uniform}(990, 1110)$
- $T_l \sim \text{Uniform}(63.1, 116)$
- $H_l \sim \text{Uniform}(700, 820)$
- $L \sim \text{Uniform}(1120, 1680)$
- $K_w \sim \text{Uniform}(9855, 12045)$

Do these indices back up your conclusions from the GP output? Produce a plot of the predicted response against the most important variable. If you identified any potential interactions, find a way to visualise these on your plot(s).