Performance Engineering Tutorial
Design of Experiments

Exercise 1. We wish to investigate the effect of a GCC compiler optimization flag on the execution time (E) of a code that calculates digits of π. We use a 2^k factorial design without replication and with $k = 2$ factors. The first factor is the GCC optimization fl ag[1] (F), with ON/OFF levels, and the second factor is the number of digits D that we require the generator to compute, with levels 10000 or 20000. We have collected following measurements:

$E\ [s]$	$F = OFF$	$F = ON$
$D = 10000$	1.1	1.5
$D = 20000$	5.9	4.3

Question 1.1 Give the sign table for the design and quantify the effects q_0, q_F, q_D, q_{FD}.

Solution:
Setting OFF=-1, ON=+1, 10000=-1, 20000=+1, we get

I	F	D	FD	$Response$
+1	-1	-1	+1	1.1
+1	+1	-1	-1	1.5
+1	-1	+1	-1	5.9
+1	+1	+1	+1	4.3

Therefore,

$$\begin{aligned} q_0 &= (1.1 + 1.5 + 5.9 + 4.3)/4 = 3.2, \\ q_F &= (-1.1 + 1.5 - 5.9 + 4.3)/4 = -0.3, \\ q_D &= (-1.1 - 1.5 + 5.9 + 4.3)/4 = 1.9, \\ q_{FD} &= (1.1 - 1.5 - 5.9 + 4.3)/4 = -0.5, \end{aligned}$$

Question 1.2 Quantify the percentages of variation explained by the factors and their interaction.
Solution:

$$SST = 4(q_F^2 + q_D^2 + q_{FD}^2) = 4(0.3^2 + 1.9^2 + 0.5^2) = 15.8$$

$$SSF = 4(q_F^2) = 4(0.3^2) = 0.36$$

$$SSD = 4(q_D^2) = 4(1.9^2) = 14.44$$

$$SSFD = 4(q_{FD}^2) = 4(0.5^2) = 1$$

Therefore, the flag explains $SSF/SST = 0.36/15.8 = 2.28\%$ of the total variation, the number of digits explains $SSD/SST = 91.39\%$, and the interaction $SSFD/SST = 6.33\%$.

[1]The values shown in this exercise are actual numbers. The optimization flag is *-O2* with gcc 4.6.3.

Exercise 2. Consider a 2^3 full factorial design where the factors can take the following levels:

Factor	Low level	High level
Number of users (A)	300	500
Number of cores (B)	20	30
Available memory (C)	6GB	8GB

Question 2.1 List all possible experiments for this design.
Solution:
We first need to replace levels with -1 and 1. For example, we might set low levels to -1 and high levels to 1. The sign table is obtained by enumerating all possible combinations, which are 2^k for a design with k factors, thus:

Experiment	A	B	C
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1

Question 2.2 Generate a sign table for this design.
Solution:
We now need to extend the list of experiments to represent all possible terms in the regression model.

I	A	B	C	AB	AC	BC	ABC
1	-1	-1	-1	1	1	1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	-1	-1	1	-1	1
1	1	1	-1	1	-1	-1	-1
1	-1	-1	1	1	-1	-1	1
1	1	-1	1	-1	1	-1	-1
1	-1	1	1	-1	-1	1	-1
1	1	1	1	1	1	1	1

As a double check, it is useful to verify that any pair of columns is orthogonal. For example, for columns A and B we must have

$$\sum_{i=1}^{8} x_{Ai} x_{Bi} = 0$$

Using this definition, it is simple to verify that the columns in the above table are pairwise orthogonal.

Question 2.3 Suppose now that we wish to add to the experiment other four factors: L1 Cache size (D), L2 Cache size (E), Processor speed (F), Utilization level (G). If we want to use a 2^{7-4} fractional factorial design, how should we revise the sign table?

Solution:
This is a 2^{k-p} design with $p = 4$. Since the 2^3 full factorial design has exactly 4 terms for interactions, we can replace these terms with the p new factors. This brings the new sign table:

I	A	B	C	D	E	F	G
1	-1	-1	-1	1	1	1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	-1	-1	1	-1	1
1	1	1	-1	1	-1	-1	-1
1	-1	-1	1	1	-1	-1	1
1	1	-1	1	-1	1	-1	-1
1	-1	1	1	-1	-1	1	-1
1	1	1	1	1	1	1	1

Question 2.4 What are the confoundings for A, B and C in the 2^{7-4} design?
Solution:
To determine the confoundings, we consider the four generator relationships:

$$D = AB, E = AC, F = BC, G = ABC$$

And apply the following rules:

- I is treated as the identity, e.g., $IA = A$
- A factor that appears twice is removed, e.g., $ABCA = BC$

We need now to manipulate the generators to obtain A, B and C:

$$D = AB \Rightarrow AD = AAB \Rightarrow \boxed{AD = B}$$

$$D = AB \Rightarrow BD = BAB \Rightarrow \boxed{BD = A}$$

$$F = BC \Rightarrow BF = BBC \Rightarrow \boxed{BF = C}$$

$$F = BC \Rightarrow CF = CBC \Rightarrow \boxed{CF = B}$$

$$E = AC \Rightarrow AE = AAC \Rightarrow \boxed{AE = C}$$

$$E = AC \Rightarrow CE = CAC \Rightarrow \boxed{CE = A}$$

$$G = ABC \Rightarrow AG = AABC \Rightarrow AG = BC \Rightarrow GAG = GBC \Rightarrow A = GBC \Rightarrow \boxed{A = GF}$$

$$G = ABC \Rightarrow BG = BABC \Rightarrow BG = AC \Rightarrow GBG = GAC \Rightarrow B = GAC \Rightarrow \boxed{B = GE}$$

$$G = ABC \Rightarrow CG = CABC \Rightarrow CG = AB \Rightarrow GCG = GAB \Rightarrow C = GAB \Rightarrow \boxed{C = GD}$$

Where in the last three derivations we used again at the end one of the generator relationships. Therefore:

$$A = GF = CE = BD$$

$$B = AD = CF = GE$$

$$C = GD = AE = BF$$

and this will be a Resolution III design, which is not a good design since first-order effects and second-order interactions are confounded.

Exercise 3 (Requires calculator). Consider three factors A, B, C with 2 levels each. Analyze the 2^3 design shown in the following table:

	A_1		A_2	
	C_1	C_2	C_1	C_2
B_1	100	15	120	10
B_2	40	30	20	50

For example, the measured value 15 corresponds to the experiment where $A = A_1$, $B = B_1$, and $C = C_2$.

Question 3.1 Quantify the effects and all interactions.

Solution:
Using the sign table developed in Question 2.2, assigning -1 to the low levels (e.g., A_1) and $+1$ to the high levels (e.g., A_2), we have

I	A	B	C	AB	AC	BC	ABC	*Response*
1	-1	-1	-1	1	1	1	-1	100
1	1	-1	-1	-1	-1	1	1	120
1	-1	1	-1	-1	1	-1	1	40
1	1	1	-1	1	-1	-1	-1	20
1	-1	-1	1	1	-1	-1	1	15
1	1	-1	1	-1	1	-1	-1	10
1	-1	1	1	-1	-1	1	-1	30
1	1	1	1	1	1	1	1	50

The calculations now are as follows

$$q_0 = (100 + 120 + 40 + 20 + 15 + 10 + 30 + 50)/2^3 = 48.125$$
$$q_A = (-100 + 120 - 40 + 20 - 15 + 10 - 30 + 50)/2^3 = 1.875$$
$$q_B = (-100 - 120 + 40 + 20 - 15 - 10 + 30 + 50)/2^3 = -13.125$$
$$q_C = (-100 - 120 - 40 - 20 + 15 + 10 + 30 + 50)/2^3 = -21.875$$
$$q_{AB} = (100 - 120 - 40 + 20 + 15 - 10 - 30 + 50)/2^3 = -1.875$$
$$q_{AC} = (100 - 120 + 40 - 20 - 15 + 10 - 30 + 50)/2^3 = 1.875$$
$$q_{BC} = (100 + 120 - 40 - 20 - 15 - 10 + 30 + 50)/2^3 = 26.875$$
$$q_{ABC} = (-100 + 120 + 40 - 20 + 15 - 10 - 30 + 50)/2^3 = 8.125$$

Question 3.2 Quantify percentages of variation explained. Then sort the factors in order of decreasing importance.

Solution:

$$SST = 2^3(q_A^2 + q_B^2 + q_C^2 + q_{AB}^2 + q_{AC}^2 + q_{BC}^2 + q_{ABC}^2) = 11596.875$$

Therefore

$$A : 2^3 q_A^2/SST = 0.24\%$$
$$B : 2^3 q_B^2/SST = 11.88\%$$
$$C : 2^3 q_C^2/SST = 33.01\%$$
$$AB : 2^3 q_{AB}^2/SST = 0.24\%$$
$$AC : 2^3 q_{AC}^2/SST = 0.24\%$$
$$BC : 2^3 q_{BC}^2/SST = 49.82\%$$
$$ABC : 2^3 q_{ABC}^2/SST = 4.55\%$$

Sorting from largest to smallest: BC, C, B, ABC, A, AB, AC Thus, BC is the most important factor, while A, AB, AC. are jointly the least important, suggesting that factor A may be removed from future experiments, provided that the 4.55% variation explained by ABC is deemed small by the experimenter.