

1

A simple comparative experiment

1.1 Key concepts

1. Good experimental designs allow for precise estimation of one or more unknown quantities of interest. An example of such a quantity, or parameter, is the difference in the means of two treatments. One parameter estimate is more precise than another if it has a smaller variance.
2. Balanced designs are sometimes optimal, but this is not always the case.
3. If two design problems have different characteristics, they generally require the use of different designs.
4. The best way to allocate a new experimental test is at the treatment combination with the highest prediction variance. This may seem counterintuitive but it is an important principle.
5. The best allocation of experimental resources can depend on the relative cost of runs at one treatment combination versus the cost of runs at a different combination.

Is A different from B? Is A better than B? This chapter shows that doing the same number of tests on A and on B in a simple comparative experiment, while seemingly sensible, is not always the best thing to do. This chapter also defines what we mean by the best or optimal test plan.

2 OPTIMAL DESIGN OF EXPERIMENTS

1.2 The setup of a comparative experiment

Peter and Brad are drinking Belgian beer in the business lounge of Brussels Airport. They have plenty of time as their flight to the United States is severely delayed due to sudden heavy snowfall. Brad has just launched the idea of writing a textbook on tailor-made design of experiments.

[Brad] I have been playing with the idea for quite a while. My feeling is that design of experiments courses and textbooks overemphasize standard experimental plans such as full factorial designs, regular fractional factorial designs, other orthogonal designs, and central composite designs. More often than not, these designs are not feasible due to all kinds of practical considerations. Also, there are many situations where the standard designs are not the best choice.

[Peter] You don't need to convince me. What would you do instead of the classical approach?

[Brad] I would like to use a case-study approach. Every chapter could be built around one realistic experimental design problem. A key feature of most of the cases would be that none of the textbook designs yields satisfactory answers and that a flexible approach to design the experiment is required. I would then show that modern, computer-based experimental design techniques can handle real-world problems better than standard designs.

[Peter] So, you would attempt to promote optimal experimental design as a flexible approach that can solve any design of experiments problem.

[Brad] More or less.

[Peter] Do you think there is a market for that?

[Brad] I am convinced there is. It seems strange to me that, even in 2011, there aren't any books that show how to use optimal or computer-based experimental design to solve realistic problems without too much mathematics. I'd try to focus on how easy it is to generate those designs and on why they are often a better choice than standard designs.

[Peter] Do you have case studies in mind already?

[Brad] The robustness experiment done at Lone Star Snack Foods would be a good candidate. In that experiment, we had three quantitative experimental variables and one categorical. That is a typical example where the textbooks do not give very satisfying answers.

[Peter] Yes, that is an interesting case. Perhaps the pastry dough experiment is a good candidate as well. That was a case where a response surface design was run in blocks, and where it was not obvious how to use a central composite design.

[Brad] Right. I am sure we can find several other interesting case studies when we scan our list of recent consulting jobs.

[Peter] Certainly.

[Brad] Yesterday evening, I tried to come up with a good example for the introductory chapter of the book I have in mind.

[Peter] Did you find something interesting?

A SIMPLE COMPARATIVE EXPERIMENT 3

[Brad] I think so. My idea is to start with a simple example. An experiment to compare two population means. For example, to compare the average thickness of cables produced on two different machines.

[Peter] So, you'd go back to the simplest possible comparative experiment?

[Brad] Yep. I'd do so because it is a case where virtually everybody has a clear idea of what to do.

[Peter] Sure. The number of observations from the two machines should be equal.

[Brad] Right. But only if you assume that the variance of the thicknesses produced by the two machines is the same. If the variances of the two machines are different, then a 50–50 split of the total number of observations is no longer the best choice.

[Peter] That could do the job. Can you go into more detail about how you would work that example?

[Brad] Sure.

Brad grabs a pen and starts scribbling key words and formulas on his napkin while he lays out his intended approach.

[Brad] Here we go. We want to compare two means, say μ_1 and μ_2 , and we have an experimental budget that allows for, say, $n = 12$ observations, n_1 observations from machine 1 and $n - n_1$ or n_2 observations from machine 2. The sample of n_1 observations from the first machine allows us to calculate a sample mean \bar{X}_1 for the first machine, with variance σ^2/n_1 . In a similar fashion, we can calculate a sample mean \bar{X}_2 from the n_2 observations from the second machine. That second sample mean has variance σ^2/n_2 .

[Peter] You're assuming that the variance in thickness is σ^2 for both machines, and that all the observations are statistically independent.

[Brad] Right. We are interested in comparing the two means, and we do so by calculating the difference between the two sample means, $\bar{X}_1 - \bar{X}_2$. Obviously, we want this estimate of the difference in means to be precise. So, we want its variance

$$\text{var}(\bar{X}_1 - \bar{X}_2) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

or its standard deviation

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

to be small.

[Peter] Didn't you say you would avoid mathematics as much as possible?

[Brad] Yes, I did. But we will have to show a formula here and there anyway. We can talk about this later. Stay with me for the time being.

Brad empties his Leffe, draws the waiter's attention to order another, and grabs his laptop.

[Brad] Now, we can enumerate all possible experiments and compute the variance and standard deviation of $\bar{X}_1 - \bar{X}_2$ for each of them.

4 OPTIMAL DESIGN OF EXPERIMENTS

Table 1.1 Variance of sample mean difference for different sample sizes n_1 and n_2 for $\sigma^2 = 1$.

n_1	n_2	$\text{var}(\bar{X}_1 - \bar{X}_2)$	$\sigma_{\bar{X}_1 - \bar{X}_2}$	Efficiency (%)
1	11	1.091	1.044	30.6
2	10	0.600	0.775	55.6
3	9	0.444	0.667	75.0
4	8	0.375	0.612	88.9
5	7	0.343	0.586	97.2
6	6	0.333	0.577	100.0
7	5	0.343	0.586	97.2
8	4	0.375	0.612	88.9
9	3	0.444	0.667	75.0
10	2	0.600	0.775	55.6
11	1	1.091	1.044	30.6

Before the waiter replaces Brad’s empty glass with a full one, Brad has produced Table 1.1. The table shows the 11 possible ways in which the $n = 12$ observations can be divided over the two machines, and the resulting variances and standard deviations.

[Brad] Here we go. Note that I used a σ^2 value of one in my calculations. This exercise shows that taking n_1 and n_2 equal to six is the best choice, because it results in the smallest variance.

[Peter] That confirms traditional wisdom. It would be useful to point out that the σ^2 value you use does not change the choice of the design or the relative performance of the different design options.

[Brad] Right. If we change the value of σ^2 , then the 11 variances will all be multiplied by the value of σ^2 and, so, their relative magnitudes will not be affected. Note that you don’t lose much if you use a slightly unbalanced design. If one sample size is 5 and the other is 7, then the variance of our sample mean difference, $\bar{X}_1 - \bar{X}_2$, is only a little bit larger than for the balanced design. In the last column of the table, I computed the efficiency for the 11 designs. The design with sample sizes 5 and 7 has an efficiency of $0.333/0.343 = 97.2\%$. So, to calculate that efficiency, I divided the variance for the optimal design by the variance of the alternative.

[Peter] OK. I guess the next step is to convince the reader that the balanced design is not always the best choice.

Brad takes a swig of his new Leffe, and starts scribbling on his napkin again.

[Brad] Indeed. What I would do is drop the assumption that both machines have the same variance. If we denote the variances of machines 1 and 2 by σ_1^2 and σ_2^2 , respectively, then the variances of \bar{X}_1 and \bar{X}_2 become σ_1^2/n_1 and σ_2^2/n_2 . The variance of our sample mean difference $\bar{X}_1 - \bar{X}_2$ then is

$$\text{var}(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2},$$

A SIMPLE COMPARATIVE EXPERIMENT 5

Table 1.2 Variance of sample mean difference for different sample sizes n_1 and n_2 for $\sigma_1^2 = 1$ and $\sigma_2^2 = 9$.

n_1	n_2	$\text{var}(\bar{X}_1 - \bar{X}_2)$	$\sigma_{\bar{X}_1 - \bar{X}_2}$	Efficiency (%)
1	11	1.818	1.348	73.3
2	10	1.400	1.183	95.2
3	9	1.333	1.155	100.0
4	8	1.375	1.173	97.0
5	7	1.486	1.219	89.7
6	6	1.667	1.291	80.0
7	5	1.943	1.394	68.6
8	4	2.375	1.541	56.1
9	3	3.111	1.764	42.9
10	2	4.600	2.145	29.0
11	1	9.091	3.015	14.7

so that its standard deviation is

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

[Peter] And now you will again enumerate the 11 design options?

[Brad] Yes, but first I need an a priori guess for the values of σ_1^2 and σ_2^2 . Let's see what happens if σ_2^2 is nine times σ_1^2 .

[Peter] Hm. A variance ratio of nine seems quite large.

[Brad] I know. I know. I just want to make sure that there is a noticeable effect on the design.

Brad pulls his laptop a bit closer and modifies his original table so that the thickness variances are $\sigma_1^2 = 1$ and $\sigma_2^2 = 9$. Soon, he produces Table 1.2.

[Brad] Here we are. This time, a design that requires three observations from machine 1 and nine observations from machine 2 is the optimal choice. The balanced design results in a variance of 1.667, which is 25% higher than the variance of 1.333 produced by the optimal design. The balanced design now is only $1.333/1.667 = 80\%$ efficient.

[Peter] That would be perfect if the variance ratio was really as large as nine. What happens if you choose a less extreme value for σ_2^2 ? Can you set σ_2^2 to 2?

[Brad] Sure.

A few seconds later, Brad has produced Table 1.3.

[Peter] This is much less spectacular, but it is still true that the optimal design is unbalanced. Note that the optimal design requires more observations from the machine with the higher variance than from the machine with the lower variance.

[Brad] Right. The larger value for n_2 compensates the large variance for machine 2 and ensures that the variance of \bar{X}_2 is not excessively large.

6 OPTIMAL DESIGN OF EXPERIMENTS

Table 1.3 Variance of sample mean difference for different sample sizes n_1 and n_2 for $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$.

n_1	n_2	$\text{var}(\bar{X}_1 - \bar{X}_2)$	$\sigma_{\bar{X}_1 - \bar{X}_2}$	Efficiency (%)
1	11	1.182	1.087	41.1
2	10	0.700	0.837	69.4
3	9	0.556	0.745	87.4
4	8	0.500	0.707	97.1
5	7	0.486	0.697	100.0
6	6	0.500	0.707	97.1
7	5	0.543	0.737	89.5
8	4	0.625	0.791	77.7
9	3	0.778	0.882	62.4
10	2	1.100	1.049	44.2
11	1	2.091	1.446	23.2

[Peter, pointing to Table 1.3] Well, I agree that this is a nice illustration in that it shows that balanced designs are not always optimal, but the balanced design is more than 97% efficient in this case. So, you don't lose much by using the balanced design when the variance ratio is closer to 1.

Brad looks a bit crestfallen and takes a gulp of his beer while he thinks of a comeback line.

[Peter] It would be great to have an example where the balanced design didn't do so well. Have you considered different costs for observations from the two populations? In the case of thickness measurements, this makes no sense. But imagine that the two means you are comparing correspond to two medical treatments. Or treatments with two kinds of fertilizers. Suppose that an observation using the first treatment is more expensive than an observation with the second treatment.

[Brad] Yes. That reminds me of Eric Schoen's coffee cream experiment. He was able to do twice as many runs per week with one setup than with another. And he only had a fixed number of weeks to run his study. So, in terms of time, one run was twice as expensive as another.

[Peter, pulling Brad's laptop toward him] I remember that one. Let us see what happens. Suppose that an observation from population 1, or an observation with treatment 1, costs twice as much as an observation from population 2. To keep things simple, let the costs be 2 and 1, and let the total budget be 24. Then, we have 11 ways to spend the experimental budget I think. One extreme option takes one observation for treatment 1 and 22 observations for treatment 2. The other extreme is to take 11 observations for treatment 1 and 2 observations for treatment 2. Each of these extreme options uses up the entire budget of 24. And, obviously, there are a lot of intermediate design options.

Peter starts modifying Brad's table on the laptop, and a little while later, he produces Table 1.4.

A SIMPLE COMPARATIVE EXPERIMENT 7

Table 1.4 Variance of sample mean difference for different designs when treatment 1 is twice as expensive as treatment 2 and the total cost is fixed.

n_1	n_2	$\text{var}(\bar{X}_1 - \bar{X}_2)$	$\sigma_{\bar{X}_1 - \bar{X}_2}$	Efficiency (%)
1	22	1.045	1.022	23.2
2	20	0.550	0.742	44.2
3	18	0.389	0.624	62.4
4	16	0.313	0.559	77.7
5	14	0.271	0.521	89.5
6	12	0.250	0.500	97.1
7	10	0.243	0.493	100.0
8	8	0.250	0.500	97.1
9	6	0.278	0.527	87.4
10	4	0.350	0.592	69.4
11	2	0.591	0.769	41.1

[Peter] Take a look at this.

[Brad] Interesting. Again, the optimal design is not balanced. Its total number of observations is not even an even number.

[Peter, nodding] These results are not quite as dramatic as I would like. The balanced design with eight observations for each treatment is still highly efficient. Yet, this is another example where the balanced design is not the best choice.

[Brad] The question now is whether these examples would be a good start for the book.

[Peter] The good thing about the examples is that they show two key issues. First, the standard design is optimal for at least one scenario, namely, in the scenario where the number of observations one can afford is even, the variances in the two populations are identical and the cost of an observation is the same for both populations. Second, the standard design is often no longer optimal as soon as one of the usual assumptions is no longer valid.

[Brad] Surely, our readers will realize that it is unrealistic to assume that the variances in two different populations are exactly the same.

[Peter] Most likely. But finding the optimal design when the variances are different requires knowledge concerning the magnitude of σ_1^2 and σ_2^2 . I don't see where that knowledge might come from. It is clear that choosing the balanced design is a reasonable choice in the absence of prior knowledge about σ_1^2 and σ_2^2 , as that balanced design was at least 80% efficient in all of the cases we looked at.

[Brad] I can think of a case where you might reasonably expect different variances. Suppose your study used two machines, and one was old and one was new. There, you would certainly hope the new machine would produce less variable output. Still, an experimenter usually knows more about the cost of every observation than about its variance. Therefore, the example with the different costs for the two populations is

8 OPTIMAL DESIGN OF EXPERIMENTS

possibly more convincing. If it is clear that observations for treatment 1 are twice as expensive as observations for treatment 2, you have just shown that the experimenter should drop the standard design, and use the unbalanced one instead. So, that sounds like a good example for the opening chapter of our book.

[Peter, laughing] I see you have already lured me into this project.

[Brad] Here is a toast to our new project!

They clink their glasses, and turn their attention toward the menu.

1.3 Summary

Balanced designs for one experimental factor at two levels are optimal if all the runs have the same cost, the observations are independent and the error variance is constant. If the error variances are different for the two treatments, then the balanced design is no longer best. If the two treatments have different costs, then, again, the balanced design is no longer best.

A general principle is that the experimenter should allocate more runs to the treatment combinations where the uncertainty is larger.