POINTS OF SIGNIFICANCE

Two-factor designs

When multiple factors can affect a system, allowing for interaction can increase sensitivity.

When probing complex biological systems, multiple experimental factors may interact in producing effects on the response. For example, in studying the effects of two drugs that can be administered simultaneously, observing all the pairwise level combinations in a single experiment is more revealing than varying the levels of one drug at a fixed level of the other. If we study the drugs independently we may miss biologically relevant insight about synergies or antisynergies and sacrifice sensitivity in detecting the drugs' effects.

The simplest design that can illustrate these concepts is the 2 × 2 design, which has two factors (A and B), each with two levels (*a*/*A* and *b*/*B*). Specific combinations of factors (*a*/*b*, *A*/*b*, *a*/*B*, *A*/*B*) are called treatments. When every combination of levels is observed, the design is said to be a complete factorial or completely crossed design. So this is a complete 2 × 2 factorial design with four treatments.

Our previous discussion about experimental designs was limited to the study of a single factor for which the treatments are the factor levels. We used ANOVA[1] to determine whether a factor had an effect on the observed variable and followed up with pairwise *t*-tests[2] to isolate the significant effects of individual levels. We now extend the ANOVA idea to factorial designs. Following the ANOVA analysis, pairwise *t*-tests can still be done, but often analysis focuses on a different set of comparisons: main effects and interactions.

Figure 1 illustrates some possible outcomes in a 2 × 2 factorial experiment (values in **Table 1**). Suppose that both factors correspond to drugs and the observed variable is liver glucose level. In **Figure 1a**, drugs A and B increase glucose levels by 1 unit. Because neither drug influences the effect of the other we say there is no interaction and that the effects are additive. In **Figure 1b**, the effect of A in the presence of B is larger than the sum of their effects when they are administered separately (3 vs. 0.5 + 1). When the effect of the levels of a factor depends on the levels of other factors, we say that there is an interaction between the factors. In this case, we need to be careful about defining the effects of each factor.

The main effect of factor A is defined as the difference in the means of the two levels of A averaged over all the levels of B. For **Figure 1b**, the average for level *a* is $\tau = (0 + 1)/2 = 0.5$ and for level *A* is $\tau = (0.5 + 3)/2 = 1.75$, giving a main effect of 1.75 – 0.5 = 1.25 (**Table 1**). Similarly, the main effect of B is 2 – 0.25 = 1.75. The interaction compares the differences in the mean of A at the two levels of B (2 – 0.5 = 1.5; in the Δ row) or, equivalently, the differences in the mean of B at the two levels of A (2.5 – 1 = 1.5). Interaction plots are useful to evaluate effects when the number of factors is small (line plots, **Fig 1b**). The *x* axis represents levels of one factor and lines correspond to levels of other factors. Parallel lines indicate no interaction. The more the lines diverge, or cross, the greater the interaction.

Figure 1c shows an interaction effect with no main effect. This can happen if one factor increases the response at one level of the other factor but decreases it at the other. Both factors have the same average value for each of their levels, $\tau = 0.5$. However, the two factors do interact because the effect of one drug is different depending on the presence of the other.

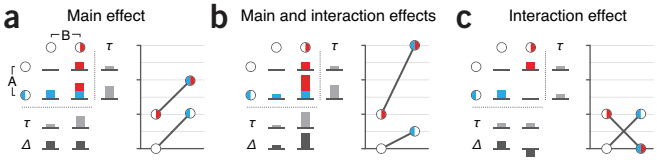

Figure 1 | When studying multiple factors, main and interaction effects can be observed, shown here for two factors (A, blue; B, red) with two levels each. (**a**) The main effect is the difference between τ values (light gray), which is the response for a given level of a factor averaged over the levels of other factors. (**b**) The interaction effect is the difference between effects of A at the different levels of B or vice versa (dark gray, Δ). (**c**) Interaction effects may mask main effects.

There are various ways in which effects can combine; their clear and concise reporting is important. For a 2 × 2 design with two levels per factor, effects can be estimated directly from treatment means. In this case, effects should be summarized with their estimated value and a confidence interval (CI) and graphically reported as a plot of means with error bars[2]. Optionally, a two-sample *t*-test can be used to provide a *P* value for the null hypothesis that the two treatments have the same effect—a zero difference in their means. For example, with levels *a*/*A* and *b*/*B* we have four treatment means μ_{ab}, μ_{Ab}, μ_{aB} and μ_{AB}. The effect of A at level *b* is $\mu_{Ab} - \mu_{ab}$, which is estimated by substituting the observed sample means. The standard error of this estimate is s.e. $= s\sqrt{(1/n_{Ab} + 1/n_{ab})}$, where *s* is the estimate of the population standard deviation, estimated by $\sqrt{MS_E}$, where MS_E is the residual mean square from the ANOVA, and n_{ij} is the observed sample size for treatment A = *i* and B = *j*. If the design is balanced, $n_{Ab} = n_{ab} = n$ and s.e. $= \sqrt{(2MS_E/n)}$. The *t*-statistic is $t = (\bar{x}_{Ab} - \bar{x}_{ab})/\text{s.e.}$ The CI can be constructed using $\bar{x}_{Ab} - \bar{x}_{ab} \pm t^* \times \text{s.e.}$, where t^* is the critical value for the *t*-statistic at the desired α. Note, however, that the degrees of freedom (d.f.) are the error d.f. from the ANOVA, not $2(n - 1)$ as in the usual two-sample *t*-test[2], because the MS_E rather than the sample variances is used in the s.e. computation.

When there are more factors or more levels, the main effects and interactions are summarized over many comparisons as sums of squares (SS) and usually only the test statistic (*F*-test), its d.f. and the *P* value are reported. If there are statistically significant interactions, pairwise comparisons of different levels of one factor for fixed levels of the other factors (sometimes called simple main effects) are often computed in the manner described above. If the interactions are not significant, we typically compute differences between levels of one factor averaged over the levels of the other factor. Again, these are pairwise comparisons between means that are handled as just described, except that the sample sizes are also summed over the levels.

To illustrate the two-factor design analysis, we'll use a simulated data set in which the effect of levels of the drug and diet were tested in two different designs, with 8 mice and 8 observations (**Fig. 2a**). We'll assume an experimental protocol in which a mouse liver tissue sample is tested for glucose levels using two-way ANOVA. Our simulated simple effects are shown in **Figure 1b**—the increase in the response variable is 0.5 (*A*/*b*), 1 (*a*/*B*) and 3 (*A*/*B*). The two drugs are synergistic—A is 4× as potent in the presence of B, as can be seen by $(\mu_{AB} - \mu_{aB})/(\mu_{Ab} - \mu_{ab}) = \Delta_B/\Delta_b = 2/0.5 = 4$ (**Table 1**). We'll assume the same variation due to mice and measurement error, $\sigma^2 = 0.25$.

Table 1 | Quantities used to determine main and interaction effects from data in **Figure 1**

	Main effect			Main and interaction effects			Interaction effect		
	b	***B***	τ	***b***	***B***	τ	***b***	***B***	τ
a	0	1	0.5	0	1	0.5	0	1	0.5
A	1	2	1.5	0.5	3	1.75	1	0	0.5
τ	0.5	1.5		0.25	2		0.5	0.5	
Δ	1	1		0.5	2		1	−1	

Treatment values shown are means for *a/b*, *a/B*, *A/b* and *A/B* level combinations. A main effect is observed if the difference between τ values (e.g., 1.5 − 0.5 = 1) is nonzero. An interaction effect is observed if Δ, the difference between the mean levels of A, varies across levels of B or vice versa.

We'll use a completely randomized design with each of the 8 mice randomly assigned to one of the four treatments in a balanced fashion each providing a single liver sample (**Fig. 2a**). First, let's test the effect of the two factors separately using one-way ANOVA, averaging over the values of the other factor. If we consider only A, the effects of B are considered part of the residual error and we do not detect any effect ($P = 0.48$, **Fig. 2b**). If we consider only B, we can detect an effect ($P = 0.04$) because B has a larger main effect (2.0 − 0.25 = 1.75) than A (1.75 − 0.5 = 1.25).

When we test for multiple factors, the ANOVA calculation partitions the total sum of squares, SS_T, into components that correspond to A (SS_A), B (SS_B) and the residual (SS_E) (**Fig. 2b**). The additive two-factor model assumes that there is no interaction between A and B—the effect of a given level of A does not depend on a level of B. In this case, the interaction component is assumed to be part of the error. If this assumption is relaxed, we can partition the total variance into four components, now accounting for how the response of A varies with B. In our example, the SS_A and SS_B terms remain the same, but SS_E is reduced by the amount of SS_{AB} (4.6), to 2.0 from 6.6. The resulting reduction in MS_E (0.5 vs. 1.3) corresponds to the variance explained by the interaction between the two factors. When interaction is accounted for, the sensitivity of detecting an effect of A and B is increased because the *F*-ratio, which is inversely proportional to MS_E, is larger.

To calculate the effect and interaction CIs, as described above, we start with the treatment means $\bar{x}_{ab} = 0.27$, $\bar{x}_{Ab} = -0.39$, $\bar{x}_{aB} = 0.86$ and $\bar{x}_{AB} = 3.23$, each calculated from two values. To calculate the main effects of A and B, we average over four measurements to find $\bar{x}_a = 0.57$, $\bar{x}_A = 1.42$, $\bar{x}_b = -0.06$ and $\bar{x}_B = 2.05$. The residual error $MS_E = 0.5$ is used to calculate the s.e. of main effects: $\sqrt{(2MS_E/n)} = \sqrt{(2 \times 0.5/4)} = 0.5$. The critical *t*-value at $\alpha = 0.05$ and d.f. = 4 is 2.78, giving a 95% CI for the main effect of A to be 0.9 ± 1.4 ($F_{1,4} = 2.9$), where d.f. = (1,4) and of B to be 2.1 ± 1.4 ($F_{1,4} = 17.6$). The CIs reflect that we detected the main effect of B but not of A. For the interaction, we find $(\bar{x}_{AB} - \bar{x}_{aB}) - (\bar{x}_{Ab} - \bar{x}_{ab}) = 3.0$ with s.e. = 1 and a CI of 3.0 ± 2.8 ($F_{1,4} = 9.1$).

To improve the sensitivity of detecting an effect of A, we can mitigate biological variability in mice by using a randomized complete block approach[1] (**Fig. 2a**). If the mice share some characteristic, such as litter or weight which contributes to response variability, we could control for some of the variation by assigning one complete replicate to each batch of similar mice. The total number of observations will still be 8, and we will track the mouse batch across measurements and use the batch as a random blocking factor[2]. Now, in addition to the effect of interaction, we can further reduce the MS_E by the amount of variance explained by the block (**Fig. 2b**).

The sum-of-squares partitioning and *P* values for the blocking scenario are shown in **Figure 2b**. In each case, the SS_E value is proportionately lower than in the completely randomized design, which makes the tests more sensitive. Once we incorporate blocking and interaction, we are able to detect both main and interaction effects and account for nearly all of the variance due to sources other than measurement error ($SS_E = 0.8$, $MS_E = 0.25$). The interpretation of $P = 0.01$ for the blocking factor M is that the biological variation due to the blocking factor has a nonzero variance. Effects and CIs are calculated just as for the completely randomized design—although the means have two sources of variance (block effect and MS_E), their difference has only one (MS_E) because the block effect cancels.

With two factors, more complicated designs are also possible. For example, we might expose the whole mouse to a drug (factor A) *in vivo* and then expose two liver samples to different *in vitro* treatments (factor B). In this case, the two liver samples from the same mouse form a block that is nested in mouse.

We might also consider factorial designs with more levels per factor or more factors. If the response to our two drugs depends on genotype, we might consider using three genotypes in a 2 × 2 × 3 factorial design with 12 treatments. This design allows for the possibility of interactions among pairs of factors and also among all three factors. The smallest factorial design with *k* factors has two levels for each factor, leading to 2^k treatments. Another set of designs, called fractional factorial designs, used frequently in manufacturing, allows for a large number of factors with a smaller number of samples by using a carefully selected subset of treatments.

Complete factorial designs are the simplest designs that allow us to determine synergies among factors. The added complexity in visualization, summary and analysis is rewarded by an enhanced ability to understand the effects of multiple factors acting in unison.

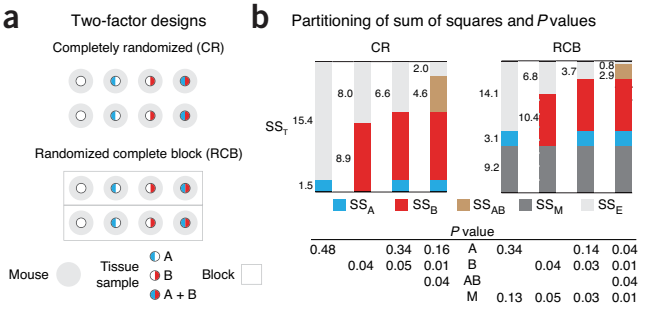

Figure 2 | In two-factor experiments, variance is partitioned between each factor and all combinations of interactions of the factors. (**a**) Two common two-factor designs with 8 measurements each. In the CR scenario, each mouse is randomly assigned a single treatment. Variability among mice can be mitigated by grouping mice by similar characteristics (e.g., litter or weight). The group becomes a block. Each block is subject to all treatments. (**b**) Partitioning of the total sum of squares (SS_T; CR, 16.9; RCB, 26.4) and *P* values for the CR and RCB designs in **a**. M represents the blocking factor. Vertical axis is relative to the SS_T. The total d.f. in both cases = 7; all other d.f. = 1.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Martin Krzywinski & Naomi Altman

1. Krzywinski, M. & Altman, N. *Nat. Methods* **11**, 699–700 (2014).
2. Krzywinski, M. & Altman, N. *Nat. Methods* **11**, 215–216 (2014).
3. Montgomery, D.C. *Design and Analysis of Experiments* 8th edn. (Wiley, 2012).

Martin Krzywinski is a staff scientist at Canada's Michael Smith Genome Sciences Centre. Naomi Altman is a Professor of Statistics at The Pennsylvania State University.