

2/28/2023

Factorial designs for ANOVA analysis

Suppose you are designing an experiment to test p different factors on a measurement. If each of those factors has n_i levels, then the total number of potential combinations those levels is $\prod_{i=1}^p n_i$. This can be a significant amount of data to collect, so the experimenter might want to run a test experiment of those p factors each with just two levels (presumably a “high” and a “low” one) to see if the factor matters at all before conducting the full experiment with all possible levels. Such experiments are called 2^p factorial experiments.

Unlike a Latin squares design, having at least one observation in each of the 2^p combinations of levels means that interaction effects can be identified if they exist.

We have previously looked at Three-Way ANOVA, and 2^3 factorial experiments follow this pattern of main effects, combine with possible two-way interactions, a three-way interaction and the rest is error.

We can review some of our previous discussion of three-way ANOVA from last semester. 2^3 factorial experiments will be a bit easier to compute since there are only two levels, some of these values will reduce to 1 in our computations.

As with our previous ANOVA models, we aren’t going to compute these by hand, but it’s worth seeing how the analysis develops and expands to account to the various effects and interactions.

$$\mu_{ijk} = \mu + \alpha_i + \beta_j + \delta_k + \gamma_{ij}^{AB} + \gamma_{ik}^{AC} + \gamma_{jk}^{BC} + \gamma_{ijk}$$

The notation in our text (Devore) uses gamma for the interaction terms as it did in the two-way model case, leaving delta for the third primary effect (if you know Greek, you know this is out-of-order now). Some sources will use gamma as the third primary effect, and may use multiple letters for the interaction effects going down the line. The subscripts here really tell us what is going on. The primary effects have only one subscript referring to each of the main treatment variables. The combination effects have pairs or triples of subscripts that correspond to the combination of main effects at play. Devore further adds subscripts to clarify which factor variables are involved (when i and j and k are replaced with specific values, this helps to disambiguate). And then the error term ϵ_{ijkl} is the random error on each measurement with mean of zero and standard deviation of σ .

Relevant sums of squares are

$$\begin{aligned}
 SST &= \sum_i \sum_j \sum_k \sum_l (X_{ijkl} - \bar{X}_{\dots})^2 & df &= IJKL - 1 \\
 SSA &= \sum_i \sum_j \sum_k \sum_l \hat{\alpha}_i^2 = JKL \sum_i (\bar{X}_{i\dots} - \bar{X}_{\dots})^2 & df &= I - 1 \\
 SSAB &= \sum_i \sum_j \sum_k \sum_l (\hat{\gamma}_{ij}^{AB})^2 & df &= (I - 1)(J - 1) \\
 &= KL \sum_i \sum_j (\bar{X}_{ij\dots} - \bar{X}_{i\dots} - \bar{X}_{\dots j} + \bar{X}_{\dots})^2 \\
 SSABC &= \sum_i \sum_j \sum_k \sum_l \hat{\gamma}_{ijk}^2 = L \sum_i \sum_j \sum_k \hat{\gamma}_{ijk}^2 & df &= (I - 1)(J - 1)(K - 1) \\
 SSE &= \sum_i \sum_j \sum_k \sum_l (X_{ijkl} - \bar{X}_{ijk})^2 & df &= IJK(L - 1)
 \end{aligned}$$

This list doesn't include the SSAC and SSBC sum of squares, but they can be generalized from the SSAB equation.

Null Hypothesis	Test Statistic Value
H_{0A} : all α_i 's = 0	$f_A = \frac{MSA}{MSE}$
H_{0AB} : all γ_{ij}^{AB} 's = 0	$f_{AB} = \frac{MSAB}{MSE}$
H_{0ABC} : all γ_{ijk} 's = 0	$f_{ABC} = \frac{MSABC}{MSE}$

Reminder that these are examples. For the full three-way model, we have three main effect hypotheses, three two-way interactions and one three-way interaction: 7 hypotheses to test.

A full ANOVA table for this type of model will have to fill in as follows. These formulas assume the sample size is the same for all components, but recall that we made some adjustments to account for different sample sizes in the various treatment combinations and similar adjustments can be made here. As our models become more complex, it becomes increasingly common to have only one observation in each treatment combination (or less).

SOURCE OF VARIATION	DF	SUM OF SQUARES	MEAN SQUARE	F
A	$I - 1$	SSA	MSA	F_A
B	$J - 1$	SSB	MSB	F_B
C	$K - 1$	SSC	MSC	F_C
AB	$(I - 1)(J - 1)$	$SSAB$	$MSAB$	F_{AB}
AC	$(I - 1)(K - 1)$	$SSAC$	$MSAC$	F_{AC}
BC	$(J - 1)(K - 1)$	$SSBC$	$MSBC$	F_{BC}
ABC	$(I - 1)(J - 1)(K - 1)$	$SSABC$	$MSABC$	F_{ABC}
ERROR	$IJK(L - 1)$	SSE	MSE	
TOTAL	$IJKL - 1$	SST		

An example of a three-way model in R using the mtcars data set using the factors gear, am and carb to predict mpg.

```

      Df Sum Sq Mean Sq F value    Pr(>F)
gear    2  483.2   241.62   26.382 1.87e-06 ***
carb    5  425.1    85.02    9.283 8.80e-05 ***
am      1   14.1    14.08    1.538  0.229
carb:am  2   11.3     5.64    0.616  0.550
Residuals 21  192.3     9.16
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

However, if we remove the interaction term, all the variable become significant.

```

      Df Sum Sq Mean Sq F value    Pr(>F)
gear    2  483.2   241.62   27.294 8.46e-07 ***
am      1   72.8    72.80    8.224 0.008699 **
carb    5  366.4    73.28    8.278 0.000136 ***
Residuals 23  203.6     8.85
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

As we noted, the full model of all possible interactions is quite complex, and we would need a complete layout (at least one observation for all conditions) in order to detect those effects which often do not rise to the level of significance. We can limit our model to just main effects and then we don't have to collect quite as much data. We can carefully chose the combinations of factors to measure. These are called incomplete layouts. They don't have the power to test interactions, but they are more efficient, especially as we continue to add more variables with more treatment levels.

As we increase the value of p to be greater than three, the number of possible interactions increases. We will have p main effects, $\binom{p}{2}$ two-way interactions, $\binom{p}{3}$ three-way interactions and so on until we have $\binom{p}{p} = 1$ p -way interactions. For $p = 4$, this means 4 main effects, 6 two-interactions, 4 three-way interactions and one 4-way interaction that are possible and should be included in the full model. If $p = 5$, then there are 5 main effects, 10 two-way interactions, 10 three-way interactions, 5 four-way interactions, and 1 five-way interaction. While these higher-level interactions tend to get rejected right away, as we saw with even two-way and three-way ANOVA, sometimes the interaction will not be rejected, but the main effect might be, creating potential issues of interpretation for the model.

As we increase the number of factors, we increase the possibility of other potential complications. One possible complication is confounding variables. It may not be possible to conduct all the experiments in the same, identical environment for a variety of real-world reasons. In some cases it may be possible to break the experiment down into smaller blocks to conduct 2^r (with $r < p$) experiments in an identical environment. While block designs are valid, this may reduce the sensitivity of the experiment and make some main effects or interactions impossible to estimate. However, this does generally allow lower-order interactions or main effects to be estimable. The design of the experiment can assist on determining which effects could be confounded.

Consider the 2^3 design with the following blocks.



This type of blocking permits the two-way interactions to be estimated, and the main effects, but the three-way effects will be confounded.

We can also use more than two blocks. In a five-factor experiment (A, B, C, D, and E), the two three-factor interactions BCD and CDE might be chosen for confounding. The third effect confounded is then the generalized interaction of the two, obtained by writing the two chosen effects side by side and then cancelling any letters common to both: (BCE)(CDE)=BE. Notice that if ABC and CDE are chosen for confounding, their generalized interaction is (ABC)(CDE)=ABDE, so that no main effects or two-factor interactions are confounded.

We can then select the blocks with even/odd and odd/odd contrasts with the confounded effects. An example of such a design is shown below.

<i>Block 1</i>	<i>Block 2</i>	<i>Block 3</i>	<i>Block 4</i>
(1) (0, 0)	<i>d</i> (0, 1)	<i>a</i> (1, 0)	<i>c</i> (1, 1)
<i>ab</i> (2, 0)	<i>e</i> (0, 1)	<i>b</i> (1, 0)	<i>ad</i> (1, 1)
<i>de</i> (0, 2)	<i>ac</i> (2, 1)	<i>cd</i> (1, 2)	<i>ae</i> (1, 1)
<i>acd</i> (2, 2)	<i>bc</i> (2, 1)	<i>ce</i> (1, 2)	<i>bd</i> (1, 1)
<i>ace</i> (2, 2)	<i>abd</i> (2, 1)	<i>ade</i> (1, 2)	<i>be</i> (1, 1)
<i>bcd</i> (2, 2)	<i>abe</i> (2, 1)	<i>bde</i> (1, 2)	<i>abc</i> (3, 1)
<i>bce</i> (2, 2)	<i>acde</i> (2, 3)	<i>abcd</i> (3, 2)	<i>cde</i> (1, 3)
<i>abde</i> (2, 2)	<i>bcde</i> (2, 3)	<i>abce</i> (3, 2)	<i>abcde</i> (3, 3)

Block 1, containing (1), is called the principal block. A second block can be obtained by selecting any experimental condition not in the principal block and obtaining its generalized interaction with every condition in the principal block. The other blocks are then constructed in the same way by first selecting a condition not in a block already constructed and finding generalized interactions with the principal block.

As we saw with earlier designs with many levels, as the number of factor variables grows, the number of conditions we need to observe can become quite large. The Latin Square design was one method of reducing the number of observations needed to study main effects in those cases. In 2^p factorial designs, one sometimes prefers to do a half-replicate experiment, where only 2^{p-1} observations are collected. If the five-fold interaction is chosen to be the confounded effect, then we can use effects to pair by using aliases, or its generalized interaction with the confounded effect.

$\{A, BCDE\}$	$\{B, ACDE\}$	$\{C, ABDE\}$	$\{D, ABCE\}$	$\{E, ABCD\}$
$\{AB, CDE\}$	$\{AC, BDE\}$	$\{AD, BCE\}$	$\{AE, BCD\}$	$\{BC, ADE\}$
$\{BD, ACE\}$	$\{BE, ACD\}$	$\{CD, ABE\}$	$\{CE, ABD\}$	$\{DE, ABC\}$

We can then choose the first element of these pairs, or the second to conduct our half-replicate observations on. Since the 4-factor effects are generally negligible, this allows the experimenter to test for the presence of main effects.

A quarter-replicate experiment can also be constructed similarly resulting in quadruples shown below.

$\{A, CE, ABDE, BCD\}$ $\{B, ABCE, DE, ACD\}$ $\{C, AE, BCDE, ABD\}$
 $\{D, ACDE, BE, ABC\}$ $\{E, AC, BD, ABCDE\}$ $\{AB, BCE, ADE, CD\}$
 $\{AD, CDE, ABE, BC\}$

Depending on the size of p , we could extend our analysis to $\frac{1}{2^r}$ -replicate experiments.

If you are interested in learning more about the details of factorial experimental design in general or in R specifically, I've linked several sources below. We will primarily concern ourselves with analyzing the results of our ANOVA tests on data collected in this fashion, so it's important to know upfront what the limitations are, but we will let R do the ANOVA calculations for us. I won't ask you to design such an experiment, but it may be something you want to consider as part of your final project if it interests you.

References:

1. https://faculty.ksu.edu.sa/sites/default/files/probability_and_statistics_for_engineering_and_the_sciences.pdf
2. <http://www.r-tutor.com/elementary-statistics/analysis-variance/factorial-design>
3. <https://towardsdatascience.com/design-of-experiments-with-r-e54167fac490>
4. https://www.researchgate.net/publication/344597074_FACTORIAL_ANOVA_EXPERIMENTAL_DESIGN_IN_R
5. <https://search.r-project.org/CRAN/refmans/ez/html/00Index.html>
6. <https://maxrabe.com/2020/designr-tutorial/>