FACTORIAL DESIGNS

INTRODUCTION

Overview

In a factorial design, all possible combinations of the levels of two or more factors are investigated.

R.A. Fisher argued that factorial designs are more efficient and more informative than one-factor-at-a-time experiments. In particular, they allow us to detect interactions.

In a factorial design, all possible combinations of the levels of two or more factors are investigated.

R.A. Fisher argued that factorial designs are more efficient and more informative than one-factor-at-a-time experiments. In particular, they allow us to detect interactions.

Example

Example. Two factors, A and B, each with two levels. Consider this table of the true means:

	B	
A	1	2
1	μ_{11}	μ_{12}
2	μ_{21}	μ_{22}

Let's start with a simple example where we study two factors, A and B, each with two levels.

Consider this table of true means. Each cell of the table corresponds to a combination of a level of factor A and a level of factor B.

Simple effects, main effect, interaction

In a factorial experiment, the treatment effect of one factor can vary as the level of the other factor changes. To help fix ideas, we define

1. The ***simple effect*** of a factor is the effect of that factor when holding the level of the other factor fixed:
   - The simple effect of factor A in level 1 of factor B is $\mu_{21} - \mu_{11}$.
   - The simple effect of factor A in level 2 of factor B is $\mu_{22} - \mu_{12}$.
   - The simple effect of factor B in level 1 of factor A is $\mu_{12} - \mu_{11}$.
2. The ***main effect*** of factor A is the average of its (two) simple effects:

$$\frac{(\mu_{21} - \mu_{11}) + (\mu_{22} - \mu_{12})}{2} = \frac{(\mu_{21} + \mu_{22}) - (\mu_{11} + \mu_{12})}{2} = \bar{\mu}_{2\cdot} - \bar{\mu}_{1\cdot}$$

3. If the simple effect of A differs between the levels of B, we have ***interaction***, or ***effect modification*** (see figures on the next page).

In a factorial experiment, the treatment effect of one factor can vary as the factor level of the other changes. To help fix ideas, we define simple effect, main effect, and interaction.

The simple effect of a factor is the effect of that factor when holding the level of the other factor fixed.

For example, the simple effect of factor A in level 1 of factor B is (mu_{21} - mu_{11});

The simple effect of factor A in level 2 of factor B is (mu_{22} - mu_{12});

Note that there is one simple effect of factor A for each level of factor B. Similarly, there is one simple effect of factor B for each level of factor A.

The simple effect of factor B in level 1 of factor A is (mu_{12} - mu_{11}).

The main effect of factor A is the average of its simple effects.

If the simple effect of A differs between the levels of B, we have interaction, or effect modification (see figures on the next page).

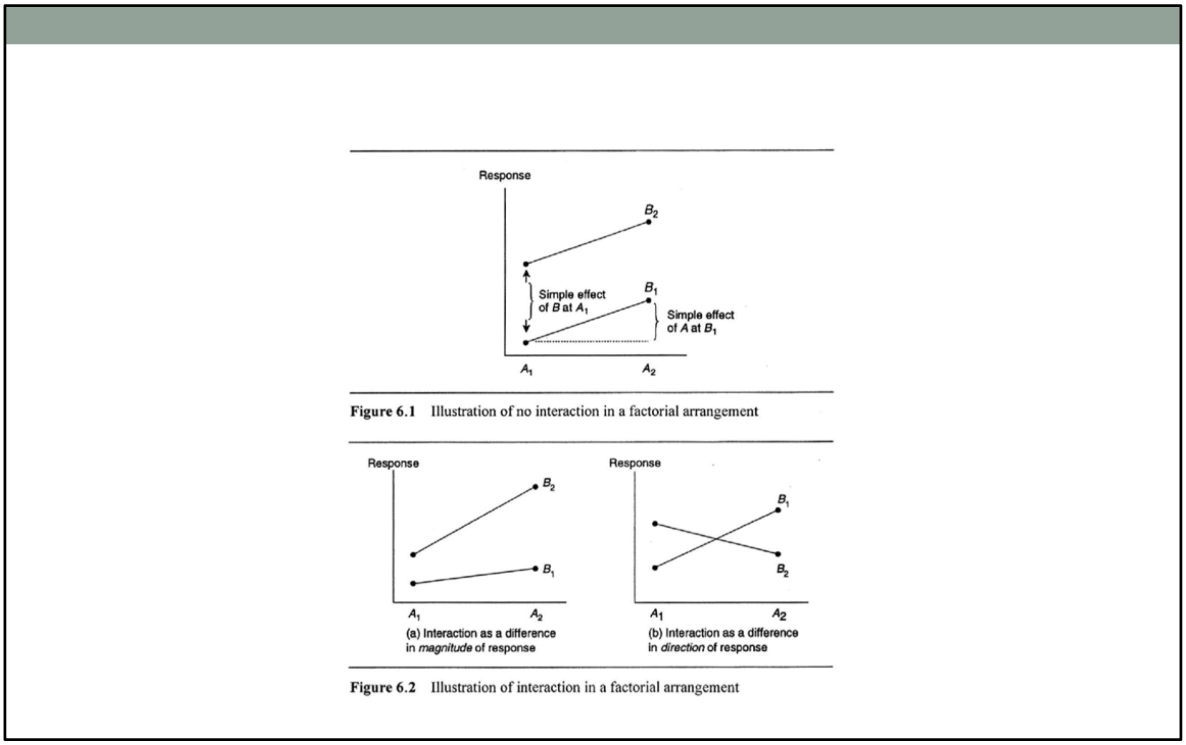


Figure 6.1 Illustration of no interaction in a factorial arrangement

Figure 6.2 Illustration of interaction in a factorial arrangement

In these Figures, we illustrate the simple effect, the main effect and interaction.

In each plot, each line segment connecting the two points indicates one level of factor B, and thus the simple effect of A is the height difference between the two end points of each line segment.

The simple effects of factor B correspond vertical distance between the two points that have the same X coordinate.

In the top plot, we see that the two simple effects of A are the same, so there is no interaction in this case.

In the bottom two plots, we see the simple effects of A are different in the two levels of factor B. So there are interactions.

In particular, in the bottom right plot, the simple effects of A have different signs under the two levels of factor B.

Note that if all simple effects of A are the same, one can prove that all simple effects of B must be the same too. If the simple effects of A are not all the same, one can prove that the simple effects of B are not all the same either.

Example

In a study of the effects of aggregate type ($a = 2$ levels) and compaction method ($b = 4$ levels) on the strength of concrete specimens, each of the 8 treatment combinations was replicated three times ($r = 3$) using a completely randomized design: 24 concrete specimens were prepared and tested for strength in random order.

Exercise: Can you list the elements of the experimental design for this example?

In a study of the effects of two aggregate types and four compaction methods on the strength of concrete specimens, each of the 8 treatment combinations was replicated three times using a completely randomized design: 24 concrete specimens were prepared and tested for strength in random order.

Can you list the elements of the experimental design for this example?

Elements of experimental design

1. Research question: the effects of two aggregate types and four compaction methods on the strength of concrete specimens.
2. Treatments: the 8 treatment combinations of aggregate type and compaction methods
3. Assignment of treatments to experimental units: the 8 treatments were randomly assigned to 24 specimens in a completely randomized experiment.
4. Experimental units: the 24 specimens
5. Measurement: strength of the concrete specimens

Here we list the elements of experimental design for this experiment:

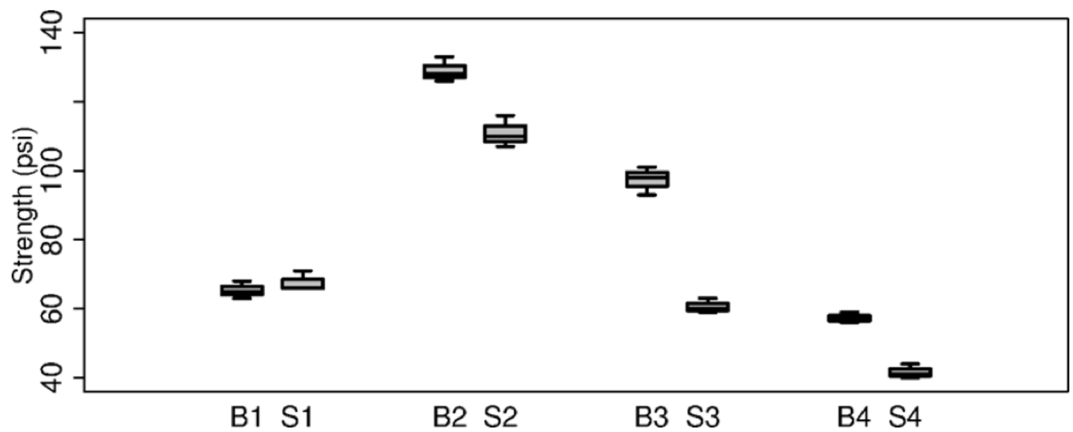
The research question is the study of the effects of two aggregate types and four compaction methods on the strength of concrete specimens.

The treatments are the 8 treatment combinations of aggregate type and compaction methods.

The 8 treatments were randomly assigned to 24 specimens in a completely randomized experiment.

The experimental units are the 24 specimens.

And the measurement is strength of the concrete specimens.



Strength of concrete specimens for different combinations of aggregate type (B = basalt, S = siliceous) and compaction method (1 through 4).

Now let's look at the data from this experiment. Remember that we always want to look at the data.

The side-by-side boxplots of strength by treatment combination show an interaction between aggregate type and compaction method: for example, we see that the difference in response means between the two aggregate types changes as the compaction method change.

As we see here, the side-by-side boxplots are a useful tool for visualizing data from a factorial experiment. It can help us spot interactions easily.

STATISTICAL MODEL FOR THE TWO-FACTOR FACTORIAL DESIGN

The cell-means model (parameterization)

$$y_{ijk} = \mu_{ij} + e_{ijk}, i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, r,$$

where μ_{ij} is the mean of the treatment combination $A_i B_j$, and e_{ijk} is random error, assumed i.i.d. $N(0, \sigma^2)$.

Note that in the model equation, the cell means are the fixed, but unknown parameters. The error term e_{ijk} is random.

Recall that we can assume that e_{ijk}'s are iid, if we completely randomize the experimental units to all treatment levels (here A and B have $a \times b$ level combinations).

Using a cell-means parameterization, the model equation for the two-factor experiment looks almost the same as the cell-means model equation for a one-way ANOVA model with completely randomized design. There is one parameter (mu i j) for each combination of (A, B) levels.

In fact, a two-factor experiment can be viewed as a one-factor experiment if we treat all the combinations of (A, B) levels as factor levels one single factor, as long as we completely randomize the experimental units to the combined treatment levels.

In the model equation, the cell means are the fixed, but unknown parameters. The error term (e i j k) is random.

Recall that we can assume that (e i j k)'s are iid, if we completely randomize the experimental units to all treatment levels. Here the treatment levels are all (a times b) level combinations of factors A and B.

The effects model (parameterization)

This is the same model as, but a different parameterization of, the cell-means model ($\mu_{ij} = \mu + \alpha_i + \beta_j$):

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk},$$

where

- μ = overall mean
- α_i = the effect of ith level of A
- β_j = effect of jth level of B
- $(\alpha\beta)_{ij}$ = effect of the interaction between α_i and β_j; and
- e_{ijk} = random error.

The α's, β's and $(\alpha\beta)$'s sum to zero (see Equation 6.10 on p. 183 of Kuehl).

When fitting regression model in R, the model equation is often specified as the effects model.

The effects model is a different parameterization of the same model as the cell-means model.

Under this parameterization, the mean parameter (mu i j) is partitioned into contributions from an A-level effect (alpha i), a B-level effect (beta j), and the interaction effect of (alpha beta i j).

In particular, this parameterization will allow us to test for the interaction effects. If the interaction effects are not significant, the model will be easier to interpret than the cell-means model.

The alpha's, beta's and (alpha beta)'s sum to zero (see Equation 6.10 on p. 183 of Kuehl).

When fitting regression model in R, the model equation is often specified as the effects model.

ANOVA, HYPOTHESIS TEST, CONFIDENCE INTERVALS

Decomposition of sum of squares in a balanced two-factor experiment

$$\text{SS Total} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(y_{ijk} - \bar{y}_{\bullet\bullet\bullet})^2 \qquad \text{df} = abr - 1$$

$$= \text{SSA} + \text{SSB} + \text{SSAB} + \text{SSE},$$

where

$$\text{SSA} = br\sum_{i=1}^{a}(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2 \qquad \text{df} = a - 1$$

$$\text{SSB} = ar\sum_{j=1}^{b}(\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet})^2 \qquad \text{df} = b - 1$$

$$\text{SSAB} = r\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet})^2$$

$$\text{df}_{AB} = (\text{df for cells}) - (\text{df for main effects})$$
$$= (ab - 1) - (a - 1) - (b - 1) = (a - 1)(b - 1)$$

$$\text{SSE} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(y_{ijk} - \bar{y}_{ij\bullet})^2$$

$$\text{df}_E = (r - 1 \text{ df per cell}) \times (ab \text{ cells}) = ab(r - 1)$$

This page summarizes the partitioning of the total sum of squares into SSA, SSB, SSAB, and SSE in a balanced two-factor experiment.

In a balanced experiment, each level combination is replicated the same number of times.

The numbers of d.f. corresponding to SSA, SSB, SSAB and SSE are a-1, b-1, (a-1) times (b-1), and (a b times r -1) respectively.

Decomposition of the treatment effect

The treatment effect, $\overline{y}_{ij.} - \overline{y}_{...}$, can be decomposed into the sum of

- factor A main effect $\overline{y}_{i..} - \overline{y}_{...}$,
- factor B main effect $\overline{y}_{.j.} - \overline{y}_{...}$, and
- interaction $(\overline{y}_{ij.} - \overline{y}_{...}) - [(\overline{y}_{i..} - \overline{y}_{...}) + (\overline{y}_{.j.} - \overline{y}_{...})] = (\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...})$.

That is

$$\overline{y}_{ij.} - \overline{y}_{...} = (\overline{y}_{i..} - \overline{y}_{...}) + (\overline{y}_{.j.} - \overline{y}_{...}) + (\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...}).$$

Square both sides and sum over all observations, we get

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\overline{y}_{ij.} - y_{...})^2 = SSA + SSB + SSAB$$

The left-hand side sum of squares is called **SS Treatment**. It has $(ab - 1)$ d.f..

Note: the "cross terms" arising from squaring the right hand side of the equation sum up to 0 for a balanced two-factor design.

The treatment effect, the difference between each cell-mean and the overall mean, can be decomposed into a sum of three effects:

a factor A main effect, the difference between the an A-factor level mean and the overall mean,

a factor B main effect, the difference between a B-factor level mean and the overall mean,

and an interaction term, the difference between the treatment effect the sum of the corresponding A-factor and B-factor main effects.

Square both sides of this decomposition equation and sum over all observations, we see that

the sum of squares for treatment effects equals SSA + SSB + SSAB.

The sum of squares on the left-hand side of this equation is called SS Treatment. It has (ab - 1) d.f.. (ab-1) is the sum of the numbers of d.f. for SSA, SSB, and SSAB.

Note that the "cross terms" arising from squaring the right-hand side of the above equation sum up to 0 for a balanced two-factor design.

SS Treatment versus SSA, SSB, SSB

Conceptually, one can think that the SS Treatment corresponds to the cell-means parameterization and SSA, SSB, SSAB correspond to the effects parameterization.

Since the two parameterizations correspond to the same model:

$$\text{SS Treatment} = SSA + SSB + SSAB,$$

and the corresponding numbers of d.f. also add up:

$$(ab-1) = (a-1) + (b-1) + (a-1)(b-1).$$

One can think that the SS Treatment correspond to the cell-means parameterization and SSA, SSB, SSAB correspond to the effects model.

Since the two parameterizations correspond to the same model, SS Treatment equals the sum of SSA, SSB and SSAB, and the corresponding numbers of d.f. also add up.

ANOVA table and expected mean squares (for fixed effects)

Table 6.5 Analysis of variance for a two-factor treatment design

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square
Total	$rab-1$	SS Total		
Factor A	$a-1$	SSA	MSA	$\sigma_e^2 + rb\theta_a^2$
Factor B	$b-1$	SSB	MSB	$\sigma_e^2 + ra\theta_b^2$
AB Interaction	$(a-1)(b-1)$	$SS(AB)$	$MS(AB)$	$\sigma_e^2 + r\theta_{ab}^2$
Error	$ab(r-1)$	SSE	MSE	σ_e^2

$$\theta_a^2 = \sum_{i=1}^{a} (\overline{\mu}_{i.} - \overline{\mu}_{..})^2/(a-1) \qquad \theta_b^2 = \sum_{j=1}^{b} (\overline{\mu}_{.j} - \overline{\mu}_{..})^2/(b-1)$$

$$\theta_{ab}^2 = \sum_{i=1}^{a}\sum_{j=1}^{b} (\mu_{ij} - \overline{\mu}_{i.} - \overline{\mu}_{.j} + \overline{\mu}_{..})^2/(a-1)(b-1)$$

Kuehl, p. 186.

The ANOVA table summarizes the sum of squares, mean squares, the numbers of d.f. for the mean squares, and expected mean squares.

The column of sum of squares gives the partitioning of the total sum of squares.

The column of the expected mean squares tells us the expected value of the MSE under the null distribution. This helps us determine which quantity to use as the denominator of F-test.

For example, test for no factor A effect, the null hypothesis will correspond to all A factor level means equal, and which implies that (theta_a squared) equals 0. Under this null, MSA = MSE. This indicates that we should use the MSE as the denominator in an F-test for no factor A effect.

Similarly, MSE should also be used as the denominator in F-tests for no factor B effect or no AB interaction.

Hypothesis tests (for fixed effects)

To test H_0's, calculate F_0 and compare to the appropriate F-distribution:

H_0	F_0	F-distribution
$\alpha_1 = \cdots = \alpha_a = 0$	MSA / MSE	$F_{a-1,\, ab(r-1)}$
$\beta_1 = \cdots = \beta_b = 0$	MSB / MSE	$F_{b-1,\, ab(r-1)}$
$(\alpha\beta)_{ij} = 0$ for all i, j	MSAB / MSE	$F_{(a-1)(b-1),\, ab(r-1)}$

This table summarize the null hypotheses and F-test for the main factor effects and the interaction.

The numbers of d.f. for each of the F-test are the numbers of d.f. for the numerator and denominator mean squares.

Concrete strength data: ANOVA table

Source	SS	d.f.	MS	F	P-value
type	1734.0	1	1734.0	182.53	< 0.0001
method	16243.5	3	5414.5	569.95	< 0.0001
interaction (type $\times$ method)	1145.0	3	381.7	40.18	< 0.0001
residual	152.0	16	9.50		

This is the ANOVA table for the concrete strength data.

We see that the F-tests for the two main effects and the interaction are all significant.

Concrete strength data: hypothesis test for interaction effects

To test $H_0: (\alpha\beta)_{11} = \cdots = (\alpha\beta)_{24} = 0$ (no interaction), calculate

$$F_0 = 381.7/9.50 = 40.18.$$

The p-value is $Pr(F_{3,16} > 40.18) < 0.0001.$

We have strong evidence of an interaction between compaction method and material type: the strength difference between the two materials changes significantly as the compaction method changes.

Here are more details on computing the F-test by hand based on the mean squares from he ANOVA table.

To test the null hypothesis of no interaction, which corresponds to all (alpha beta)'s equal 0, we compute the F statistic as the ratio MS Interaction over MSE and compare it to a F-distribution with (3, 16) degrees of freedom. 3 is the number of d.f. for MS interaction and 16 is the number of d.f. for MSE.

The p-value is less than 0.0001. We have strong evidence of an interaction between compaction method and material type. That is the strength difference between the two materials changes significantly as the compaction method changes.

Mean table for the concrete strength data

Estimated cell and marginal means of tensile strength of the concrete specimens:

	Compaction Method				
Type	1	2	3	4	Means ($\bar{y}_{i\bullet\bullet}$)
Basalt	65.3	129.0	97.3	57.3	87.3
Siliceous	67.7	111.0	60.7	41.7	70.3
Means ($\bar{y}_{\bullet j \bullet}$)	66.5	120.0	79.0	49.5	$\bar{y}_{\bullet\bullet\bullet} = 78.8$

This table summarizes the cell, marginal, and overall sample means for the concrete strength data set.

For estimates of the treatment effects, see Kuehl, p. 226.

Parameter estimation

Point estimation of the simple effects and main effects are straightforward: we replace the population means with corresponding sample means.

Consider the **simple effects** of method 3 vs. method 1 for the two material types:

- Basalt: $\overline{y}_{13\cdot} - \overline{y}_{11\cdot} = 97.3 - 65.3 = 32.0$
- Siliceous: $\overline{y}_{23\cdot} - \overline{y}_{21\cdot} = 60.7 - 67.7 = -7.0$

The estimated **main effect** of method 3 vs. method 1 is $\overline{y}_{\cdot 3\cdot} - \overline{y}_{\cdot 1\cdot} = 79.0 - 66.5 = 12.5$, which is not a useful summary for either of the two material types.

Point estimation of the simple effects and main effects are straightforward: we replace the population means with corresponding sample means.

Consider the simple effects of method 3 vs. method 1 for the two material types:

For material type Basalt, the effect is 32. For Siliceous, the effect is -7.

The estimated main effect of method 3 versus method 1 is 12.5, which is the average of the two simple effects.

Note that the two simple effects have opposite signs, the main effect in this case is not a useful summary for either of the two material types.

Confidence intervals for means and contrasts

If $\overline{y}$ is the mean of n independent observations, then $Var(\overline{y}) = Var(y)/n = \sigma^2/n$.

We usually estimate σ^2 with the MSE from the fitted model. Therefore, the estimated standard error for different types of sample means are

$$\widehat{Var}(\overline{y}_{ij\cdot}) = MSE/r$$
$$\widehat{Var}(\overline{y}_{i\cdot\cdot}) = MSE/br$$
$$\widehat{Var}(\overline{y}_{\cdot j\cdot}) = MSE/ar$$
$$\widehat{Var}(\overline{y}_{\cdot\cdot\cdot}) = MSE/abr$$

In each case, we simply need to count how many observations are used to compute that mean.

To construct confidence intervals for means and contrasts, we need to know the standard errors.

Recall that when y bar is the mean of n independent observations, the variance of (y bar) is the variance of a single observation divided by n.

We usually estimate population variance of a single observation, (sigma squared), by the MSE from the fitted model.

Therefore, the estimated standard error for different types of sample means are as listed here.

In each case, we simply need to count how many observations are used to compute that mean.

Exercise

Construct a 95% CI for the mean strength of basalt specimens.

As an exercise, construct a 95% CI for the mean strength of basalt specimens.

Once you're done, you can check the solution on next page.

Exercise

Construct a 95% CI for the mean strength of basalt specimens.

Solution:

$$\bar{y}_{1..} \pm t_{0.025,16} \cdot \sqrt{\frac{MSE}{br}} = 87.3 \pm 2.12 \cdot \sqrt{\frac{9.50}{4\times 3}} = 87.3 \pm 2.12 \cdot 0.890$$
$$= (85.4, 89.2).$$

We can find the mean value from the mean table shown earlier, and find the MSE from the ANOVA table.

Since this particular mean is computed from 12 observations, we can estimate its variance by MSE divided by 12.

The number of d.f. for the t-critical value is the number of d.f. of the MSE.

Exercise

Construct a 95% CI for the difference in mean strength of basalt specimens compacted by method 1 vs. method 3.

In this exercise, you are asked to construct a 95% CI for the difference in mean strength of basalt specimens compacted by method 1 vs. method 3.

Once you're done, you can check the solution on next page.

Exercise

Solution:

The difference in mean strength of basalt specimens compacted by method 1 vs. method 3 is a contrast. Let's denote it by C.

Out estimate of the contrast C is

$$c = \overline{y}_{13\cdot} - \overline{y}_{11\cdot} = 97.3 - 65.3 = 32$$

The estimated variance of c is

$$s_c^2 = MSE \cdot (1/3 + 1/3) = 9.50 \cdot (2/3) = 6.33.$$

Note that $\overline{y}_{13\cdot}$ and $\overline{y}_{11\cdot}$ are independent and their variances add up.

A 95% CI for the difference in mean strength of basalt specimens compacted by method 1 vs. method 3 is

$$C \pm t_{0.025,16} \cdot s_c = 32 \pm 2.12 \cdot \sqrt{6.33} = 32 \pm 5.33 = (26.7,37.3).$$

The difference in mean strength of basalt specimens compacted by method 1 vs. method 3 is a contrast.

To estimate this contrast, we replace the population mean parameters in the contrast by corresponding sample means.

The estimate is 32.

The estimated variance of the estimate is 6.33. Note that, in this case, each mean value is computed from 3 observations. Note also that the two sample means here are independent and their variances add up.

A 95% CI for the difference in mean strength of basalt specimens compacted by method 1 vs. method 3 is (26.7, 37.3).

Again, the number of d.f. of the t-critical value is the number of d.f. of the MSE used in estimating the standard error.

R code

See `script2.html` for the R code for analyzing the concrete strength data.

Residual analysis

Residual analysis can be used to check model assumption (Kuehl, p. 190), and to look for usual patterns or observations.

Residual analysis can be used to check model assumption, and to look for usual patterns or observations.

Power calculations for two-way ANOVA (completely randomized design)

These are based on comparing the within-group variation of the response to the variation among the means for the different treatment combinations, ignoring the factorial structure (see Kuehl, p. 208).

In other words, we treat all factor level combinations as factor levels of a single factor, then proceed as if it were a one-way ANOVA experiment.

This is assuming that we will use a completely randomized design.

See `script2.html` for details.

Power calculation for two-way ANOVA model are based on comparing the within-group variation of the response to the variation among the means for the different treatment combinations, ignoring the factorial structure.

In other words, we treat all factor level combinations as factor levels of a single factor, then proceed as if it were a one-way ANOVA experiment.

This is assuming that we will use a completely randomized design.

Summary

1. Design and analysis of a two-factor factorial experiment using a completely randomized design.
2. The cell-means model and the effects model.
3. The ANOVA table
   - Partitioning of the sum of squares
   - F-tests for the effect of A, B or AB
4. Inference on mean parameters and contrasts
   - How to estimate the variance of different types of sample means.
5. Power and sample size calculation for two-way ANOVA with completely randomized design.
6. Make sure to study the R code in script2.html

In this lecture, we discuss the design and analysis of a two-factor factorial experiment using a completely randomized design.

We discussed the cell-means model and the effects model.

The ANOVA table summarizes the partitioning of the sum of squares and the F-tests for the effect of A, B or AB.

We discussed inference for mean parameters and contrasts. In particular, how to estimate the variance of different types of sample means.

Make sure to study the R code in script2.html.