Factorial Design

UBCO MDS — DATA 543

When the ANOVA is significant

- When the null hypothesis is rejected, it tells us that not all means are the same.
- Generally, we will want to know more, such as which means are statistically significantly different from each other.
- Hence, further investigation should be conducted to find out exactly which means differ.
- In some cases the investigator will have preplanned comparisons he would like to make; in other situations he may have no idea what differences to look for.

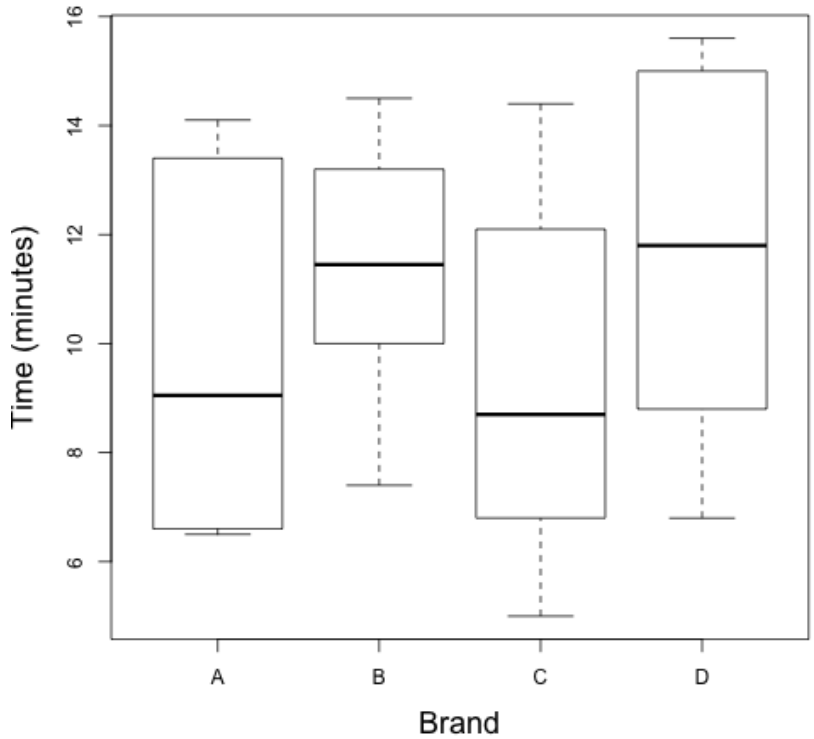
Multiple Comparisons

- Producing all pair-wise comparisons of means is often referred to as multiple comparisons
- There are a number of accepted methods for doing so:
  - Tukey-HSD (honest significant difference) — requires equal sample sizes, results in the relatively precise (smaller) intervals.
  - Tukey-Kramer — does not require same sample sizes, but will result in a smaller 'true' α than set by user.
  - Scheffé — can also be used for multiple contrasts, generally results in less precise (wider) intervals.
  - Bonferroni — can be used in most general situations, but overestimates α even more than Tukey-Kramer.

See (Montgomery, 2017, Section 3.5)

Contrasts

Last day we were comparing completion times using software A, B, C, and D



Looking at the boxplots we might want to test whether the average time from A and C is equal to the average time from B and D. You will test this in the lab.

Contrast Example One

There are many ways we can set up the contrasts by adjusting the weights a_i.

- In comparing 3 drugs, for example, a more complicated contrast is the difference between the effect of treatment 1 (drug A) and the average of the effects of treatments 2 and 3 (drugs B and C).
- In our notation, we have

$$\theta = \tau_1 - \frac{1}{2}(\tau_2 + \tau_3)$$

  $a_1 = 1$, and $a_2 = a_3 = -\frac{1}{2}$, confirming this is also a contrast.
- Notice that $\sum_i a_i = 0$ is one of our constraints.

Contrast Example Two

- Why might we be interested in different contrasts?
- In our three drug example, suppose the three drugs are made by increasing the amount chemical X. i.e. Drug A has one unit, Drug B has two units and Drug C has three units.
- We would like to test if the effect of increasing chemical X by one unit is linear. If the effect is linear then

$$\tau_2 - \tau_1 = \tau_3 - \tau_2$$

- To rephrase this as a contrast, we are interested in knowing if the following contrast is zero.

$$\begin{aligned} \theta &= [\tau_2 - \tau_1] - [\tau_3 - \tau_2] \\ &= -\tau_1 + 2\tau_2 - \tau_3 \end{aligned}$$

Factorial Treatment Structure and Interaction

In the randomized block design we framed our research question in terms of a treatment of interest, and used blocking to help identify potential treatment effects.

Blocks therefore acted as a **nuisance** variable: we (usually) don't care about whether there's a difference between blocks, although if there is it can help us investigate the treatment we *do* care about.

Factorial Treatment Structure and Interaction

What if instead of a treatment and a block, we have two **factors** that we were interested in comparing?

- e.g. with our eye treatment study, we had prescription (6 or 2 hours of patching) and additional therapy (glasses, atropine drops or none).

A **treatment** is now any combination of levels of these two factors (e.g. 6 hours and drops).

- This is a **factorial treatment structure**, as part of a **factorial** design.

Factorial Treatment Structure and Interaction

Important note: the main difference between a factorial and blocked design is perspective/intention.

- e.g. Our eye study example could be a blocked design if (e.g.) the research question was only interested in the hours of patching, but wanted to use the additional therapy as a blocking variable.

Another important aspect of factorial treatment structures is **interaction**.

- Interaction occurs if the effect of one factor level depends on the level of a second factor.

Sidenote: the RCBD we discussed only had one subject per "cell" (ie treatment/block combination), hence the interaction term cannot be extracted.

Introduction

- When multiple factors are under study, one classical approach is to study each separately while holding all others constant.
- We'll see an example of the inherent flaw with that approach momentarily
- This material is covered in (Lawson, 2014, Chapter 3) and (Montgomery, 2017, Chapter 5)

Introduction

- A much better strategy for experimenting with multiple factors is to use a *factorial design*.
- These designs consider all possible combinations of the levels of the factors under study.
- Additional pros to factorial designs include:
  - they allow for estimation of interdependency of effects (or interactions)
  - the same power or precision can be obtained with fewer replicates.
  - allow the effects of a factor to be estimated at several levels of the other factors, yielding conclusions that are valid over a range of experimental conditions.

Lawson (2014)'s Helicopter Example

Paper helicopters can be cut from one half of an $8\times$ 11 sheet of paper. An experiment can be performed by constructing a helicopter, dropping it from a fixed height, and clocking the time it takes to rotate to the floor. The wing length could be varied by trimming some paper off the top prior to folding the wings. Trimming some paper off would reduce the weight of the helicopter, but would also result in less surface area on the blades. You could experiment to determine if changing the wing length affects the flight time.

Example: helicopters

- The original example only considers one factor: the wing length
- However, to maximize the flight time of paper helicopters, it would be advisable to consider more than one factor.
- For example, consider varying wing length over 4 levels as before, *and* the body width over four levels, such as: 4.25", 4.0", 3.75", and 3.5".

Example: helicopters

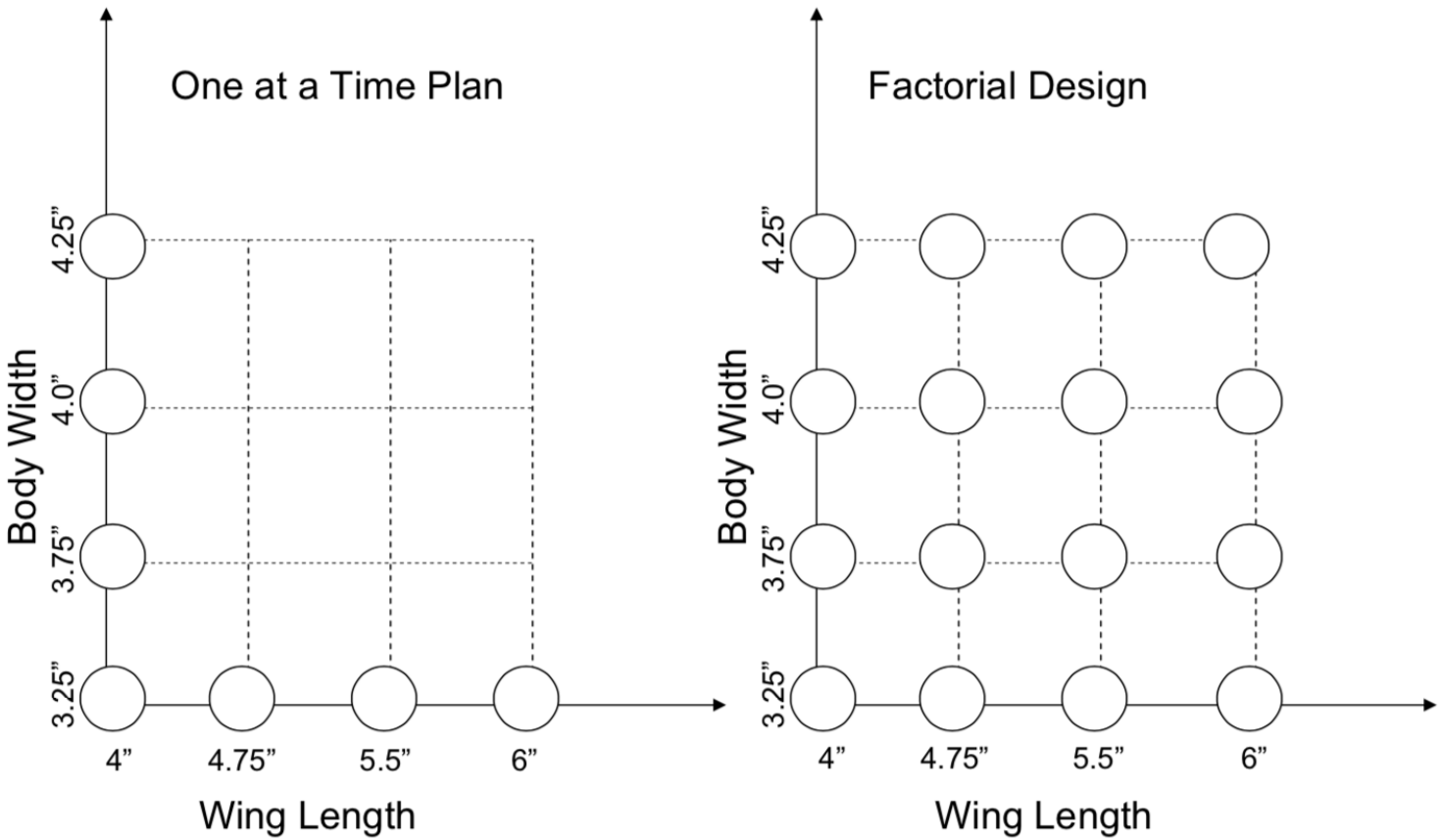
(Lawson, 2014, Fig 3.1)

- The next slide represents the classical plan in which one factor is varied at a time.
- The circles in the diagram represent experiments or runs.
- Using this approach, the experiments across the bottom of the figure would be completed by varying wing length while holding body width constant at 3.25".
- Next, the three additional experiments up the left side of the figure would be completed by varying body width while holding the wing length constant at its low level of 4.0".

Example: helicopters

Source: (Lawson, 2014, Fig 3.1)

Figure 3.1 *Comparison of One-at-a-Time and Factorial Designs*



Example: helicopter

- If the objective were to find the combination with the longest flight time, the classical approach would be to complete the experiments with one factor first.
- Next one would calculate the cell means and then select the level with the highest mean.
- Finally, the second factor would be varied while holding the first constant at its optimal level
- **Flaws:** If any unknown forces changed after the first set of experiments, the results could be biased. Additionally, the optimal level of one factor may depend upon the level of the other factor. Therefore, varying one factor at a time, the overall optimum may be missed.

Interactions

- If there is an *interaction* or joint effect between two factors, the effect of one factor upon the response will differ depending on the level of the other factor.
- Since interactions are common in factorial experiments, it is important to learn how to explain or interpret an interaction in order to clearly present the results of research studies.
- This is best done by describing the effect of one factor upon the response, and then contrasting or comparing how that effect changes depending on the level of the other factor.

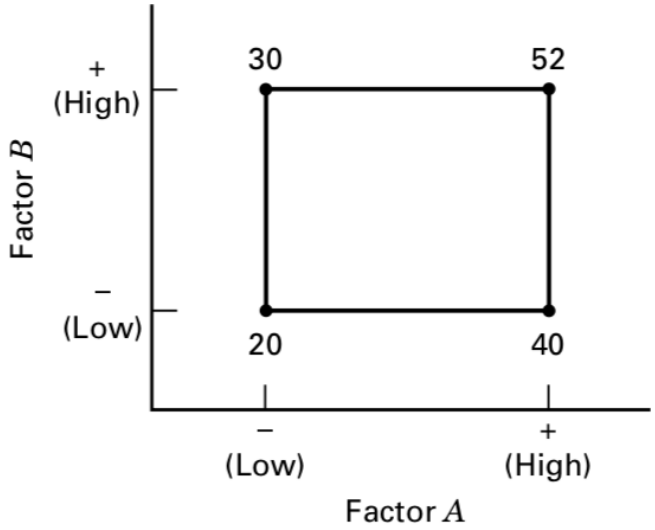
The *main effect* is the change in response produced by a change in the level of the factor.

- Let's consider a two-factor factorial experiment with both design factors at two levels.
- We have called these levels "low" $(-)$ and "high" $(+)$.
- The main effect of factor A in this two-level design can be thought of as the difference between the average response at the low level of A and the average response at the high level of A.

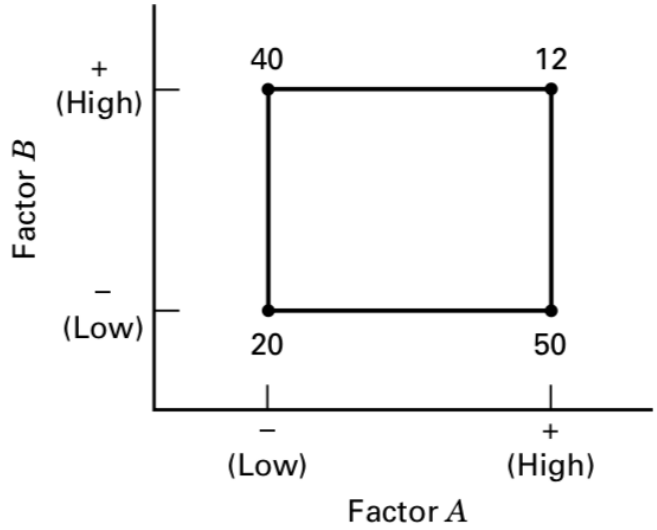
  - Main effect of A $= \dfrac{40+52}{2} - \dfrac{20+30}{2} = 21$

  - Increasing factor A from the low level to the high level causes an average response increase of 21 units

Source: Montgomery (2017)



■ **FIGURE 5.1** **A two-factor factorial experiment, with the response (y) shown at the corners**



■ **FIGURE 5.2** **A two-factor factorial experiment with interaction**

Similarly, the main effect of B $= \frac{30+52}{2} - \frac{20+40}{2} = 11$

Interaction

- We may find that the difference in response between the levels of one factor is not the same at all levels of the other factors.
- When this occurs, there is an *interaction* between the factors, eg.
  - At the low level of factor B (or B^-), the A effect is

$$A = 50 - 20 = 30$$

  - at the high level of factor B (or B^+), the A effect is

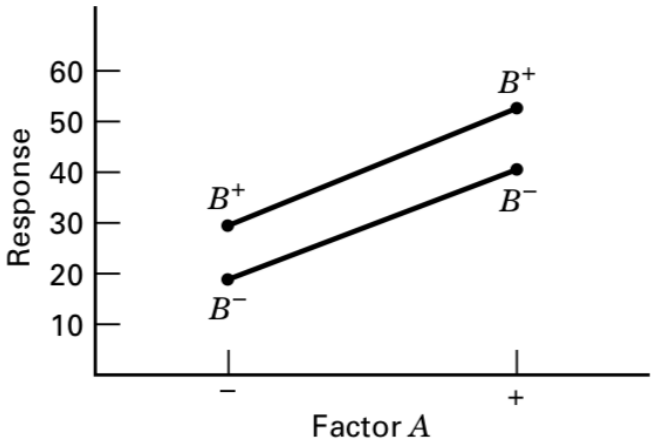
$$A = 12 - 40 = -28$$

- Because the effect of A depends on the level chosen for factor B, we see that there is *interaction* between A and B.
- The *magnitude* of the interaction effect is the average *difference* in these two A effects, or AB = (-28-30)/2 = -29

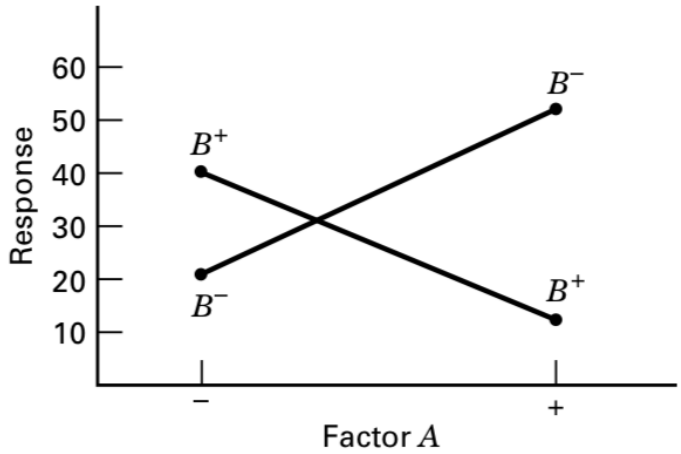
Interaction

- These ideas may be illustrated graphically in the next slide.
- Figure 5.3 plots the response data in Figure 5.1 against factor A for both levels of factor B
  - Since the lines for B^+ and B^- are approximately parallel, this indicates a lack of interaction between factors A and B.
- Similarly, Figure 5.4 plots the response data in Figure 5.2.
  - Since the lines for B^+ and B^- are *not* parallel, this indicates an interaction between factors A and B.

Source: Montgomery (2017)



■ **FIGURE 5.3** **A factorial experiment without interaction**



■ **FIGURE 5.4** **A factorial experiment with interaction**

- These *interaction plots* are frequently very useful in interpreting significant interactions and in reporting results to nonstatistically trained personnel.
- However, they should not be utilized as the sole technique of data analysis because their interpretation is subjective and their appearance is often misleading.

Two-factor Factorial

- The simplest types of factorial designs involve only two factors or sets of treatments.

- For example, if there are a levels of factor A and b levels of factor B these are arranged in a factorial design.
  - When each factor has two levels (i.e., 2×2 design) its called a two-way ANOVA
  - When each factor has three levels, the design is called a 3^2 factorial design.

- Each replicate contains all ab treatment combinations. In general, there are r replicates.

Factorial Design Model

The model is given by:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + R_{ijk} \quad \begin{matrix} i = 1, \ldots, a \\ j = 1, \ldots, b \\ k = 1, \ldots, r \end{matrix} \tag{1}$$

- μ: overall mean effect
- τ_i: treatment effect of the ith level of the row factor A,
- β_j: is the effect of the jth level of column factor B
- $(\tau\beta)_{ij}$: is the effect of the interaction between i and j
- R_{ijk}: is a random error component.

Constraints: $\sum_{i=1}^{a} \tau_i = 0$, $\sum_{j=1}^{b} \beta_j = 0$ and $\sum_{i=1}^{a} \sum_{j=1}^{b} (\tau\beta)_{ij} = 0$.

- In the two-factor factorial, both row and column factors (or treatments), A and B, are of equal interest. Specifically, we are interested in testing hypotheses about the equality of row treatment effects, say

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0 \qquad H_1 : \text{at least one } \tau_i \neq 0$$

and the equality of column treatment effects, say

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0 \qquad H_1 : \text{at least one } \beta_j \neq 0$$

- We are also interested in determining whether row and column treatments *interact*. Thus, we also wish to test:

$$H_0 : (\tau\beta)_{ij} = 0 \text{ for all } (i, j) \qquad H_1 : \text{at least one } (\tau\beta)_{ij} \neq 0$$

Decomposition of the SS

As you may have guessed, testing boils down to a decomposition of the sum of squares. Now we have:

$$SSTo = SS_A + SS_B + SS_{AB} + SS_E$$

The degrees of freedom associated with each sum of squares is:

Effect	**Degrees of Freedom**
A	$a - 1$
B	$b - 1$
AB interaction	$(a - 1)(b - 1)$
Error	$ab(r - 1)$
Total	$abr - 1$

As before we get the mean squares (MS*) by dividing the corresponding sum of squares (SS*) by its associated degrees of freedom (df*). See (Montgomery, 2017, Ch 5) for explicit formulas.

Interpretation

- If the null hypotheses of no row treatment effects, no column treatment effects, and no interaction are true, then MSA, MSB, MSAB, and MSE all estimate σ^2.
  - if there are differences between row treatment effects, then $MSA > MSE$
  - if there are column treatment effects, then $MSB > MSE$
  - or interaction present, , then $MSAB > MSE$
- Therefore, to test the significance of both main effects and their interaction, simply divide the corresponding mean square by the error mean square.
- Large values of this ratio imply that the data do not support the null hypothesis.

- If we assume model (1) is adequate and that the error terms R_{ijk} are normally and independently distributed with constant variance σ^2, then each of the ratios of mean squares

$$\frac{MSA}{MSE} \qquad \frac{MSB}{MSE} \qquad \frac{MSAB}{MSE}$$

is distributed as F with $a-1$, $b-1$, and $(a-1)(b-1)$ numerator degrees of freedom, respectively, and $ab(r-1)$ denominator degrees of freedom [1].

- The assumptions in a factorial ANOVA are the same as in a one-way ANOVA: equality of cell variances, normality of cell distributions, and independence.

[1] the critical region would be the upper tail of the F distribution

Factorial ANOVA table

Source: (Montgomery, 2017, Ch 5)

■ **TABLE 5.3**

The Analysis of Variance Table for the Two-Factor Factorial, Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
A treatments	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
B treatments	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
Interaction	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
Error	SS_E	$ab(n-1)$	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	SS_T	$abn-1$		

To account for interaction in our ANOVA code in R simply use:

`aov(y ~ A*B)`

where A and B are factors comprised of a, and b levels respectively.

Weight-loss Experiment

Example: Suppose we are interested in the effects of diet and exercise on weight-loss.

We create four weight-loss programs (treatments) by varying the

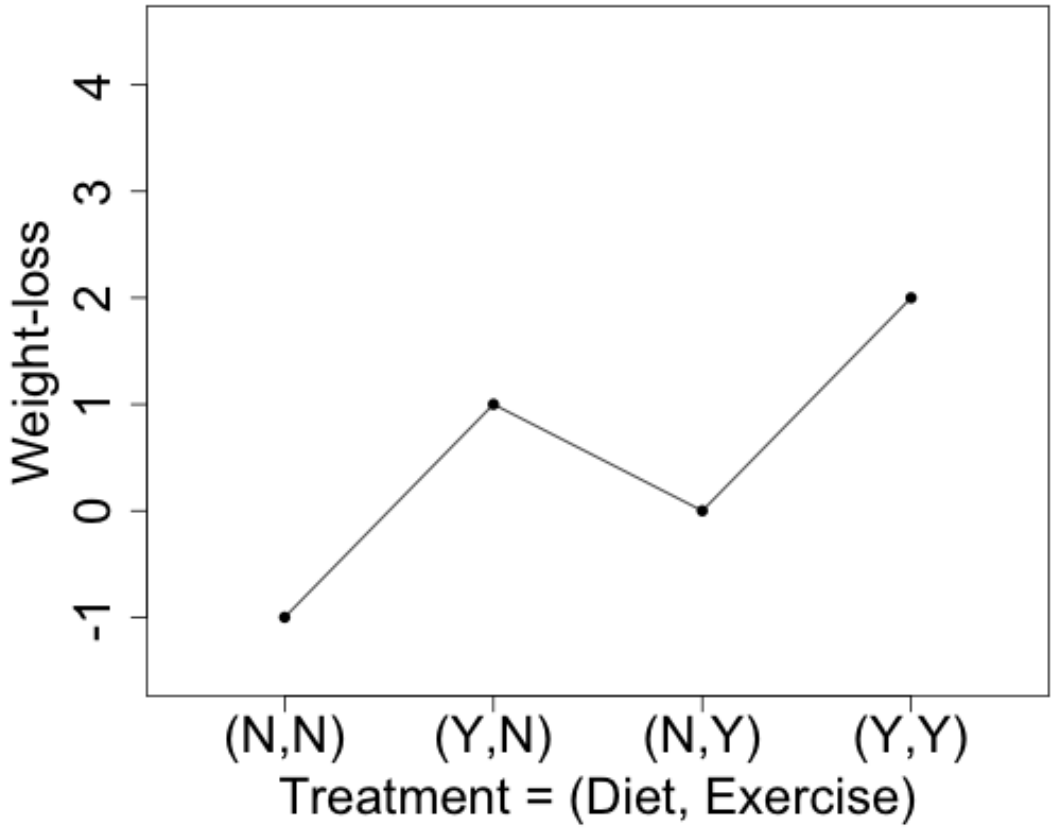
- Diet: 'Yes' or 'No' and
- Exercise: 'Yes' or 'No'.

The the response is the amount of weight-loss after two weeks.

We can summarize the experiment by plotting the average score for each combination of the two factors.

Adding the individual points would be a good way to quantify the variation but here we are focusing on the averages.

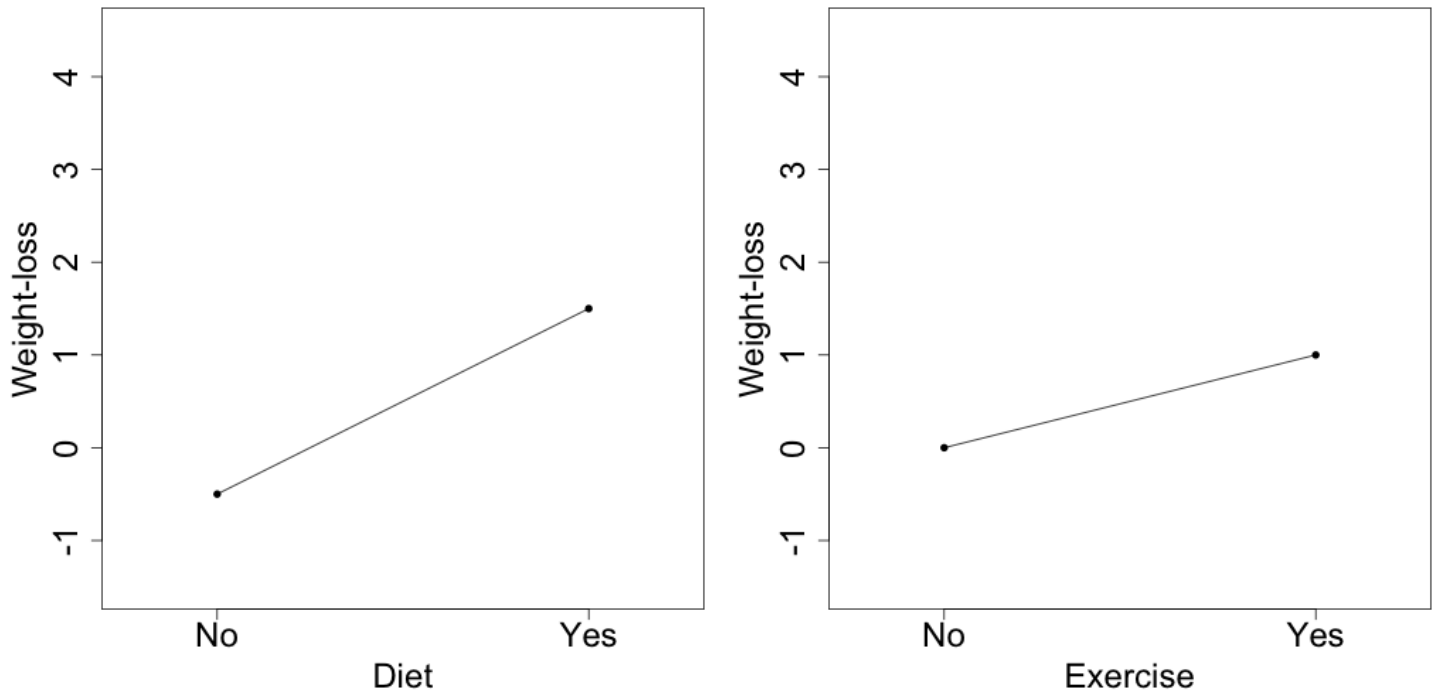
Weight-loss: Treatment Plot



Instead of plotting the four separate treatments. We could

- plot the average response when each factor is Yes/No or plot
  - avg response when Diet is Yes/No and Exercise is No and
  - avg response when Diet is Yes/No and Exercise is Yes.

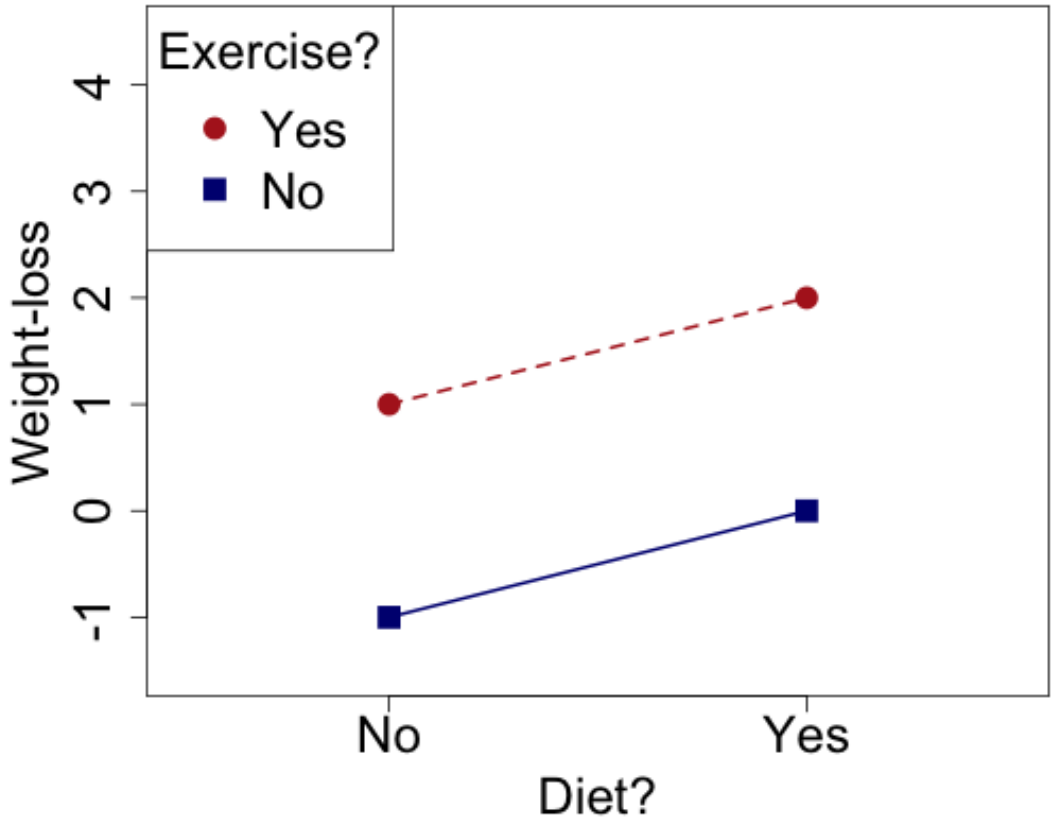
Weight-loss: Main Effect Plots



These are known as main effect plots. They

- plot the average response when Diet is Yes/No and
- plot the average response when Exercise is Yes/No.

Weight-loss: Interaction Plot



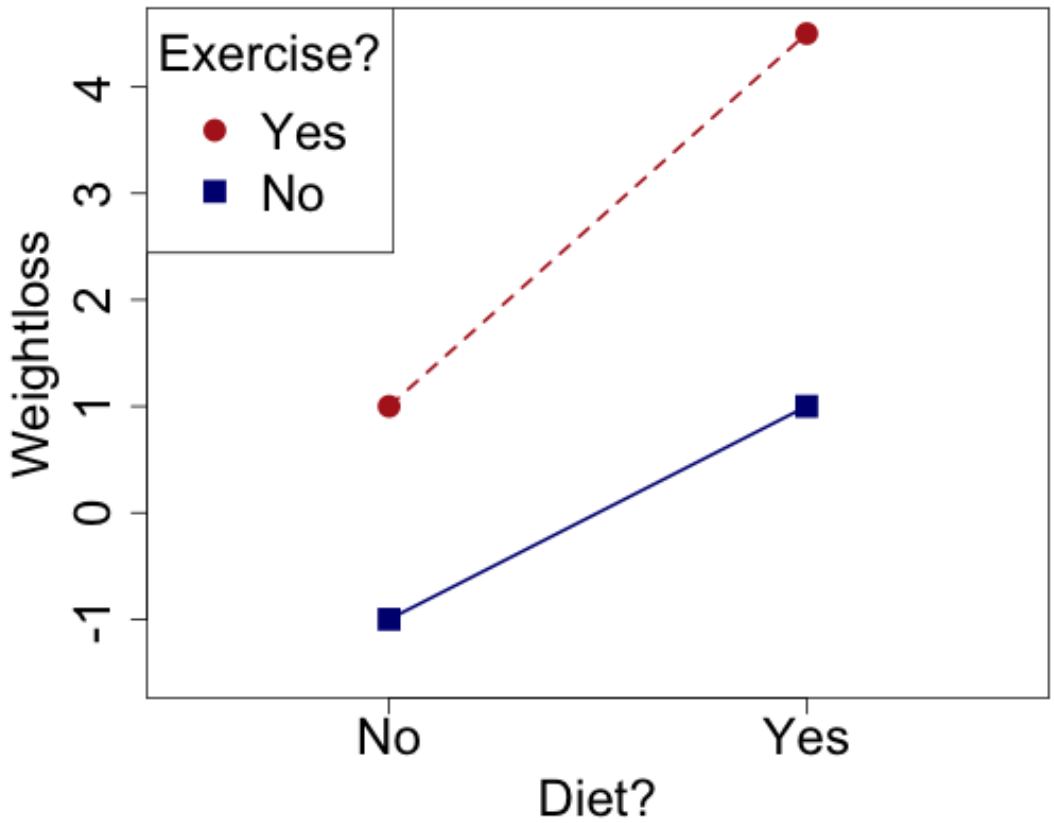
We have rearranged the treatment plot to instead plot

- average response when Diet is Yes/No and Exercise is No and
- average response when Diet is Yes/No and Exercise is Yes.

Key observation: the lines are parallel.

Weight-loss: Interaction Plot - Synergy

If exercise is more beneficial when dieting, then this might look like



Key observation: the lines are not parallel

Weight-loss: Interaction Plot - Synergy

This Interaction plot suggests that

- the benefit of exercise will yield a 2 to your weight-loss when not dieting and
- the benefit of exercise will yield a 3.5 to your weight-loss when dieting.

In other words, the effect of moving from 'no Exercise' to 'Exercise' **depends on the other factor**; i.e., whether you are on a diet.

In this example, we see **interaction** between the two factors.

This is an example of **positive** interaction:

- the change from no exercise to exercise is more beneficial if you also change from no diet to diet.

Fries Example: Description

Consider the following options to add to fries:

- Ketchup or no ketchup.
- Gravy/cheese curds, or no gravy/cheese curds.

This generates 4 possible treatments:

1. Plain fries (no ketchup or cheese curds/gravy).
2. Fries with ketchup.
3. Fries with gravy/cheese curds (i.e., poutine).
4. Fries with gravy/cheese curds and ketchup (i.e., ketchup poutine).

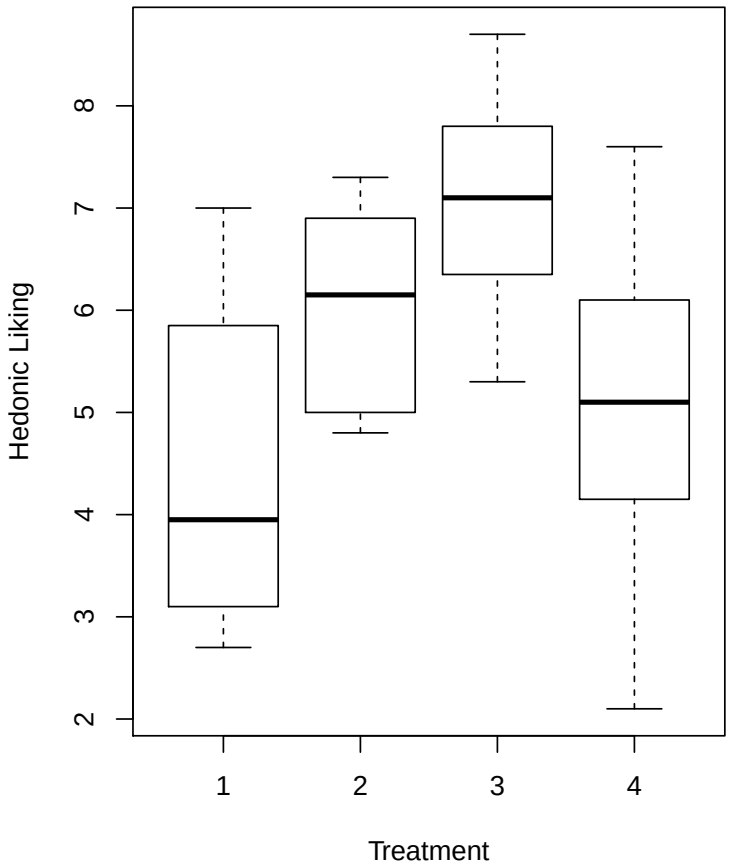
Fries Example: Summary Data

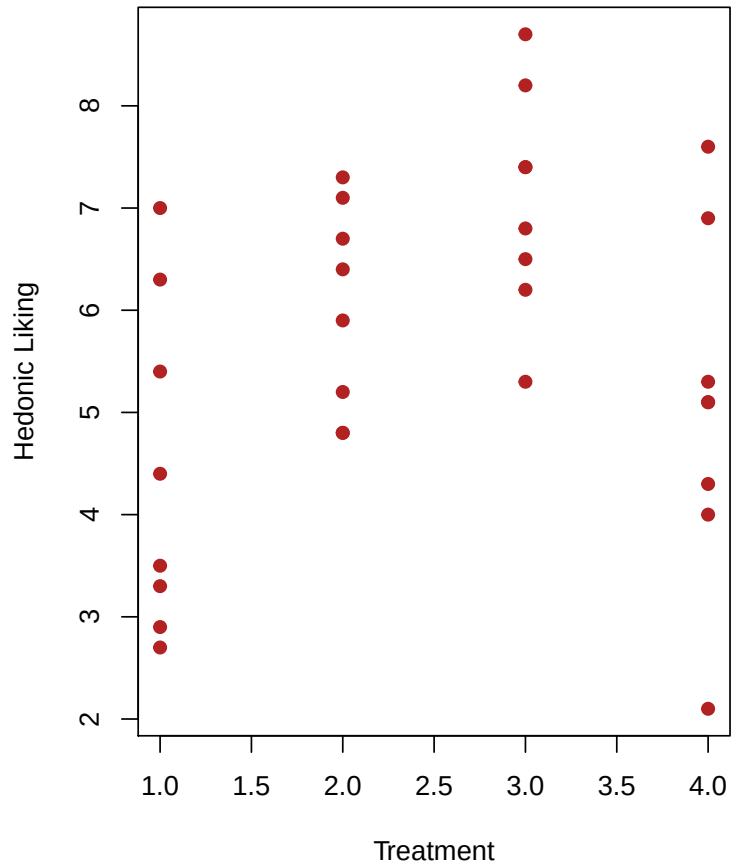
Suppose we make 32 poutines, with 8 of each type, and ask volunteers to score them in terms of deliciousness (with higher numbers preferred) on the Hedonic (Liking) Scale. We can generate a table of average responses:

Treatment	K	G+C	Average
1	No	No	4.44
2	No	Yes	6.03
3	Yes	No	7.06
4	Yes	Yes	5.05

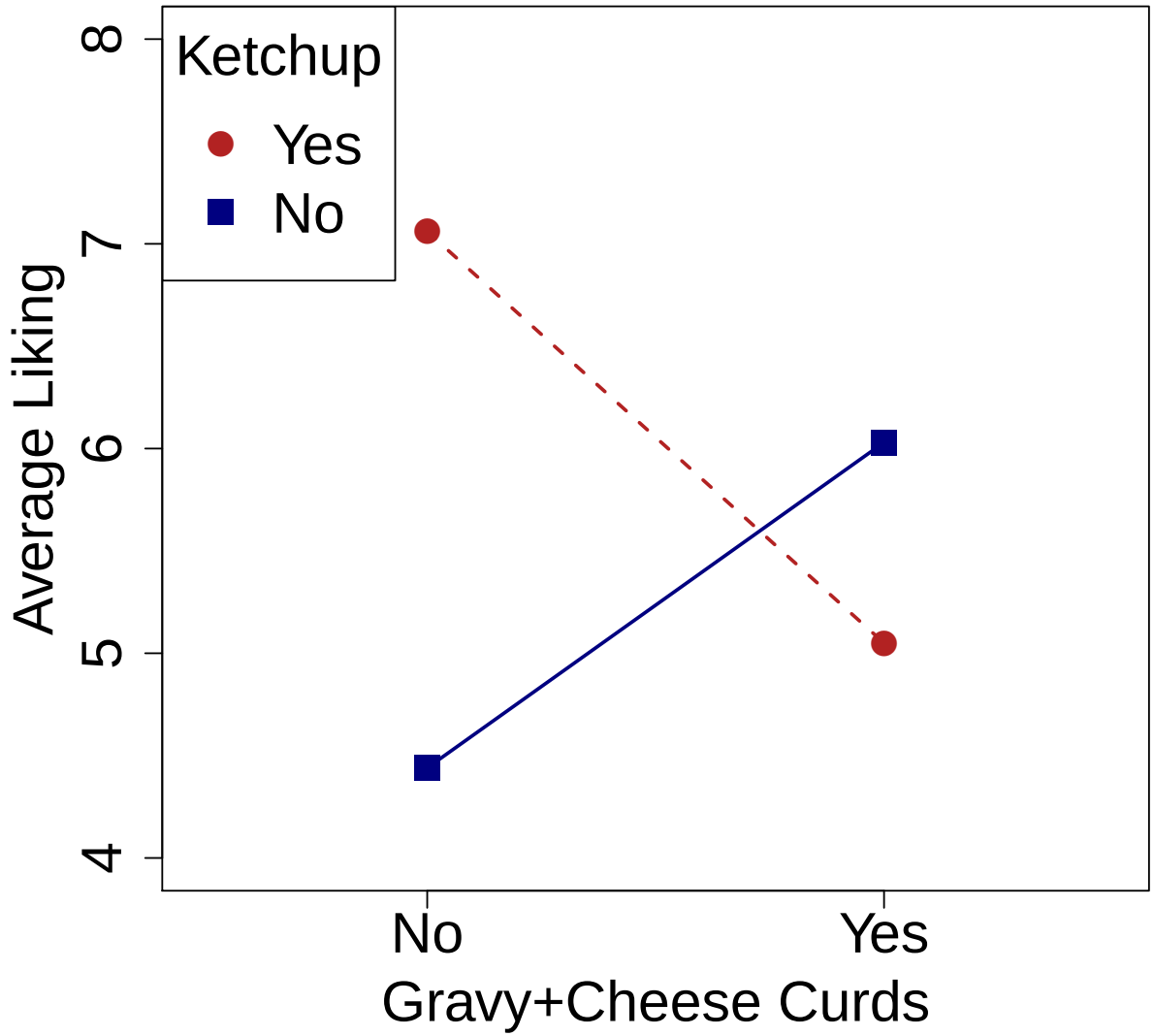
Note: when there's no ketchup, changing from no gravy+cheese to gravy+cheese *increases* the response, whereas when there is ketchup this *decreases* the average response.

This suggests there is an interaction between the two factors.


Hedonic Liking
2
3
4
5
6
7
8
1
2
3
4
Treatment


Hedonic Liking
2
3
4
5
6
7
8
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Treatment

Fries Example: Interaction Plot



Interaction: Fries Example

Treatment	K	G+C	Average
1	No	No	4.44
2	No	Yes	6.03
3	Yes	No	7.06
4	Yes	Yes	5.05

	Gravy+Cheese Curds		
	No	Yes	Average
Ketchup No	4.44	6.03	5.23
Yes	7.06	5.05	6.06
Average	5.75	5.54	5.64

Fries Example: Questions of Interest

The questions of interest are:

1. Is there any evidence of a difference among the treatments?
   - We can test this with a one-way ANOVA

2. Is the effect of adding ketchup the same regardless of whether the fries have gravy/cheese curds?
   - We can test this with an two-way ANOVA or
   - We test this formally using the following contrast under the CRD model

$$\theta = (\tau_4 - \tau_3) - (\tau_2 - \tau_1)$$

One-way ANOVA: Fries example

1. Is there any evidence of a difference among the treatments?

Treating this as a CRD and testing whether:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \qquad H_1 : \mu_i \neq \mu_j$ for at least one pair (i, j)

we see that there is a significant treatment effect (p-value$=0.0045 < \alpha$):

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(t)	3	31.73	10.58	5.44	0.0045
Residuals	28	54.39	1.94		

Two-way ANOVA: Fries example

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
k	1	5.45	5.45	2.80	0.1052
gc	1	0.36	0.36	0.19	0.6696
k:gc	1	25.92	25.92	13.34	0.0011
Residuals	28	54.39	1.94		

2. Is the effect of adding ketchup the same regardless of whether the fries have gravy/cheese curds?

This is equivalent to testing for interaction:

$$H_0 : (\tau\beta)_{ij} = 0 \text{ for all } (i, j) \qquad H_1 : \text{at least one } (\tau\beta)_{ij} \neq 0$$

Based on the p-value from above (0.0011) we conclude that there is a significant interaction between ketchup and and gravy/cheese curds.

Fries Example: Interaction as a Contrast

Alternatively, we could test this formally using the following contrast

$$\theta = (\tau_4 - \tau_3) - (\tau_2 - \tau_1)$$

(Double-check this is a contrast!)

Our estimate is $\hat{\theta} = (5.05 - 7.06) - (6.03 - 4.43) = -3.6$ and the variance of the corresponding estimator is

$$Var(\tilde{\theta}) = \sigma^2 \frac{\sum_i a_i^2}{r} = \sigma^2 \frac{4}{8} = \frac{\sigma^2}{2}$$

Contrasts Estimator (results from CRD)

- The corresponding estimator is $\tilde{\theta} = \sum_i a_i \tilde{\tau}_i = \sum_i a_i \overline{Y}_{i\cdot}$,
- with expectation

$$E[\tilde{\theta}] = E[\sum_i a_i \overline{Y}_{i\cdot}] = \sum_i a_i E[\overline{Y}_{i\cdot}] = \sum_i a_i(\mu + \tau_i) = \sum_i a_i \tau_i$$

- and variance

$$Var(\tilde{\theta}) = Var(\sum_i a_i \overline{Y}_{i\cdot}) = \sum_i a_i^2 Var(\overline{Y}_{i\cdot}) = \frac{\sigma^2}{r} \sum_i a_i^2$$

Contrasts Estimator (results from CRD)

The preceding results give us

$$\tilde{\theta} \sim N(\sum_i a_i \tau_i, \frac{\sigma^2}{r} \sum_i a_i^2)$$

and allows us to derive the test statistic

$$\frac{\sum_i a_i \tilde{\tau}_i - \sum_i a_i \tau_i}{\frac{\tilde{\sigma}}{\sqrt{r}} \sqrt{\sum_i a_i^2}} \sim t_{ab(r-1)}$$

from which confidence intervals and hypothesis tests follow. In general, this t-distribution will have the same degrees of freedom associated with the MSE term.

Fries Example: Interaction Test

The test statistic, with $H_0 : \theta = 0$ is:

$$t_{obs} = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}/\sqrt{2}} = \frac{-3.6}{\sqrt{1.94/2}} = -3.65$$

where $\theta_0 = 0$; and $\hat{\sigma}$ can be obtained from the usual formulas or the ANOVA table. We compare t_{obs} with the t-distribution with 28 degrees of freedom, and find

$$P(|t_{28}| \geq 3.65) = 0.001$$

so there is strong evidence of an interaction! Notice how this is the same p-value that we got from the ANOVA table.

```
2*pt(3.652809, 28, lower.tail = FALSE)
[1] 0.00105709
pf(13.343, 1, 28, lower.tail = F)
## [1] 0.001057095
```

Testing Main Effects: Fries Example

- To test difference of the main effects for Gravy+Cheese Curds. We use the following contrast.

$$\theta = (\tau_1 + \tau_3)/2 - (\tau_2 + \tau_4)/2$$

- Our estimate is

$$\hat{\theta} = (4.44 + 7.06)/2 - (6.03 + 5.05)/2 = 5.75 - 5.54 = 0.21$$

- and recall the variance of the corresponding estimator is

$$Var(\tilde{\theta}) = \sigma^2 \frac{\sum_i a_i^2}{r} = \frac{\sigma^2}{8}$$

Testing Main Effects: Fries Example

The test statistic, with $H_0 : \theta = 0$ is:

$$t_{obs} = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}/\sqrt{8}} = \frac{-0.21}{\sqrt{1.94/8}} = -0.426$$

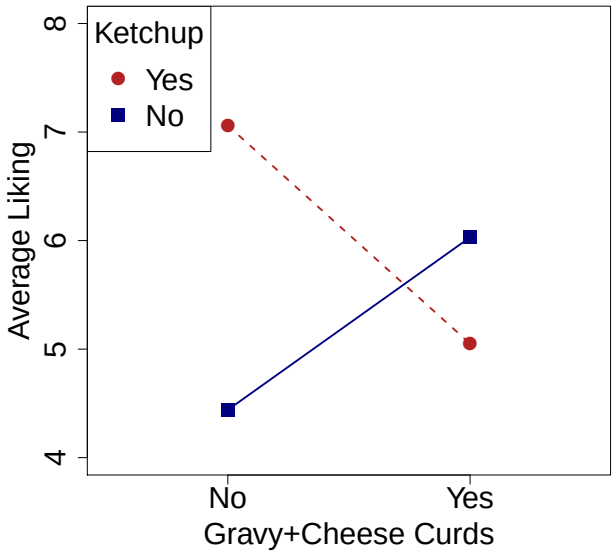
where $\theta_0 = 0$; and $\hat{\sigma}$ can be obtained from the usual formulas or the ANOVA table.

We compare t_{obs} with the t-distribution with 28 degrees of freedom, and find a p-value equal to

$$P(|t_{28}| \geq |-0.426|) = 0.673$$

so there is no evidence against the null hypothesis.

Interaction: Fries Example



Does the factor Gravy+Cheese Curds matter? Yes because the interaction was significant.

Conclusion: If an interaction is detected, it does not make sense to consider the factors separately (i.e. the main effects) because the effect of one factor on the response variate depends on the level of the second.

References I

Lawson, J. (2014), *Design and Analysis of Experiments with R*, Vol. 115.

Montgomery, D. (2017), *Design and analysis of experiments*, John Wiley and sons.