

Factorial Designs

Dr Austin R Brown

Kennesaw State University

Introduction

- ▶ So far, we have discussed experimental situations in which we are interested in the levels of a single treatment factor.
- ▶ All of our designs, including CRD, RBD, and LSD only have one treatment factor.
 - ▶ And of course, 0, 1, and 2 blocking factors, respectively.
- ▶ But could there be instances where we are interested in the effects of two or more treatment factors?
 - ▶ Let's consider an example.

Experiment Example: American Football Helmets

- ▶ In American football, players are required to wear helmets as a means of protecting their head and face from a variety of potential injuries. One of the biggest head injury risks a player faces during game play is a concussion. It is in the interest of players at all levels to reduce this risk both in terms of the rules of the game as well as the quality of the equipment.
- ▶ Suppose a football equipment manufacturer is developing new helmet technology aimed at reducing concussion risk. They want to investigate the effectiveness of two design features:
 1. **Outer Helmet Material:** (1) Traditional Polycarbonate and (2) New Composite
 2. **Internal Padding System:** (1) Standard Foam Padding and (2) Advanced Gel-Based Padding

Experiment Example: American Football Helmets

- ▶ The manufacturer creates four different helmet prototypes, one for each combination of the two factors. To collect data, they set up a standardized rig where a stationary helmet is struck by a mechanically controlled mallet at a fixed speed, simulating the forces experienced during a head-on collision (check out this **video** for a visual example).
- ▶ Inside the helmet, a force sensor records the amount of force transmitted by the blow to inside the helmet (measured in G-Force).
- ▶ For this experiment, the manufacturer produced 10 of the four prototypes for minimizing replication error. The data are contained in the Football Helmet.xlsx file.

Designing the Experiment

- ▶ As with all of our examples, we can list out our
 1. Outcome: G-Forces measured inside the helmet after being struck
 2. Treatments: Outer Helmet Material and Internal Padding System
 3. Potential Lurking Variables: Could happen if we have different operators conducting the runs of the experiment

- ▶ Next, we need to decide on the design. Here, we have two treatment variables (i.e., independent variables) and no blocking/confounding/lurking variables.
 - ▶ So what's our design?

Completely Randomized Two-Factor Design

- ▶ In this case, because we have two treatment factors and no blocking factors, our experimental design would be a ***Completely Randomized Two-Factor Design***.
- ▶ Mathematically:

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

Completely Randomized Two-Factor Design

- ▶ You may be asking yourself, “how does this conceptually differ from RBD?” Great question!
- ▶ The answer is subtle. In RBD and LSD, we assumed the blocking factors didn't interact with the treatment factor (or each other in the case of LSD).
 - ▶ In this two-factor design, we are assuming that there may be an **interaction** between Treatment 1 (denoted by α) and Treatment 2 (denoted by β).
- ▶ This interaction effect is denoted by $(\alpha\beta)_{ij}$.

Completely Randomized Two-Factor Design

- ▶ Additionally, in RBD and LSD, we completely ignored the F-tests associated with the blocking factors.
- ▶ In this two-factor design, we are interested in the test for both treatments as well as their interactions.
- ▶ Let's take a look at the data and start working through this example in a similar manner to what we've done before.

Descriptive Analysis: Summary Statistics

- Let's start by examining the means and standard deviations of the Shell type:

```
library(tidyverse)
library(readxl)
library(rstatix)
## Read in the Data ##
helmet <- read_excel("Football Helmet.xlsx")
## Calculate Means & SDs of Shell Type ##
helmet |>
  group_by(Shell) |>
  get_summary_stats(Response, type='mean_sd') |>
  select(Shell, mean, sd)
```

```
# A tibble: 2 x 3
  Shell      mean    sd
  <chr>    <dbl> <dbl>
1 New Composite  77.9  21.0
2 Polycarbonate  94.5  24.4
```

Descriptive Analysis: Summary Statistics

- ▶ So we can see that in general, regardless of Padding type, the New Composite helmets experience a substantially lesser force compared to the Polycarbonate helmets.
 - ▶ This may indicate this shell-type is better at limiting head injuries.
- ▶ Let's check out the Padding type now:

Descriptive Analysis: Summary Statistics

```
## Calculate Means & SDs of Padding Type ##  
helmet |>  
  group_by(Padding) |>  
  get_summary_stats(Response, type='mean_sd') |>  
  select(Padding, mean, sd)
```

```
# A tibble: 2 x 3  
  Padding mean    sd  
  <chr>   <dbl> <dbl>  
1 Foam    98.1  23.4  
2 Gel     74.4  18.4
```

Descriptive Analysis: Summary Statistics

- ▶ Here, we can see that the Gel padding type has a substantially lower force experienced compared to the Foam padding type.
- ▶ But here, unlike the cases we saw with one-factor RBD and LSD, we assume that our two treatment factors may have an interaction effect with one another.
- ▶ We can observe this effect both in the summary statistics as well as data visualizations.

Descriptive Analysis: Summary Statistics

```
helmet |>
  group_by(Shell,Padding) |>
  get_summary_stats(Response,type='mean_sd') |>
  select(Shell,Padding,mean,sd)
```

```
# A tibble: 4 x 4
```

	Shell	Padding	mean	sd
	<chr>	<chr>	<dbl>	<dbl>
1	New Composite	Foam	84.4	23.7
2	New Composite	Gel	71.4	16.5
3	Polycarbonate	Foam	112.	13.6
4	Polycarbonate	Gel	77.3	20.5

Descriptive Analysis: Summary Statistics

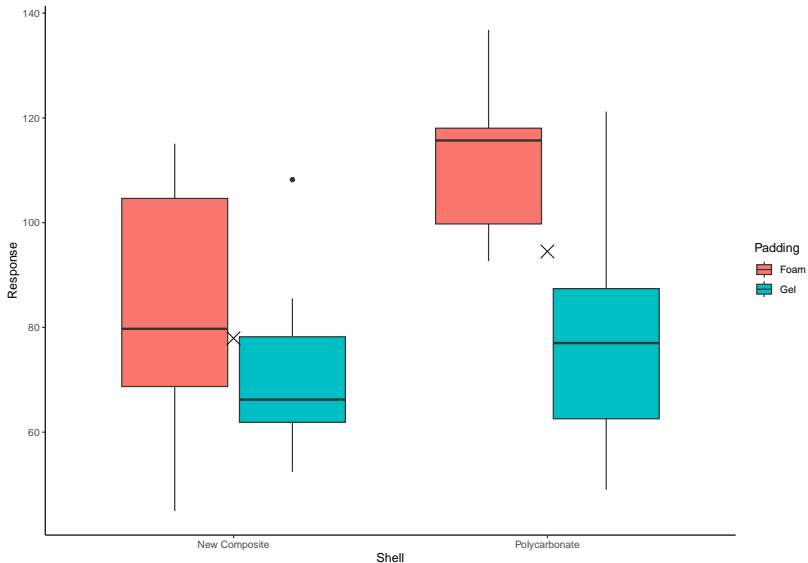
- ▶ Observing the interactions, we can see that the polycarbonate helmets generally experience the greatest force but the gel padding does reduce the force experienced for both shell types.
- ▶ If there **is not** a significant interaction, then the change in the mean response between the Shell types would be roughly equivalent regardless of Padding type.
 - ▶ In other words, the difference between New Composite/Gel and Polycarbonate/Gel would be the same as what's observed between New Composite/Foam and Polycarbonate/Foam.
- ▶ If there is evidence for a significant interaction, then the change in the mean response between the Shell types would be different depending on the type of Padding.
 - ▶ In other words, the difference between New Composite/Gel and Polycarbonate/Gel would be meaningfully different compared to what's observed between New Composite/Foam and Polycarbonate/Foam.

Descriptive Analysis: Boxplots

- ▶ As before, we can use boxplots to visually assess the same information we generated tabularly:

```
## Let's look at a boxplot to visualize the data. ##  
helmet |>  
  ggplot(aes(x=Shell)) +  
  geom_boxplot(aes(y=Response,fill=Padding)) +  
  geom_point(data = helmet |>  
    group_by(Shell) |>  
    get_summary_stats(Response,type='mean_sd'),  
    aes(x=Shell,y=mean),shape=4,size=5) +  
  theme_classic()
```

Descriptive Analysis: Boxplots



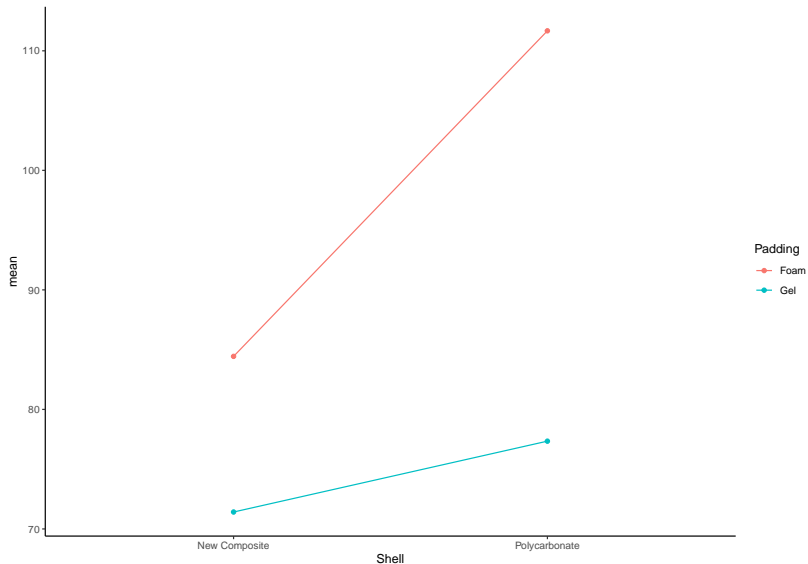
Descriptive Analysis: Boxplots

- ▶ Here, the “X” between the boxes denotes the mean of the Shell.
 - ▶ Clearly, the Polycarbonate shell has a greater mean, as we saw in the summary statistics.
- ▶ We can also see that the Foam padding in the Polycarbonate shell had a greater force experienced compared to the Foam padding in the New Composite shell.
 - ▶ This may indicate a possible interaction effect between shell and padding type.
- ▶ To better examine the interaction effect, we can use an interaction plot!

Descriptive Analysis: Interaction Plot

```
helmet |>  
  group_by(Shell,Padding) |>  
  get_summary_stats(Response,type='mean_sd') |>  
  ggplot(aes(x=Shell,y=mean,group=Padding,color=Padding)) +  
  geom_point() +  
  geom_line() +  
  theme_classic()
```

Descriptive Analysis: Interaction Plot

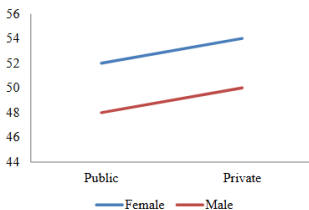


Descriptive Analysis: Interaction Plot

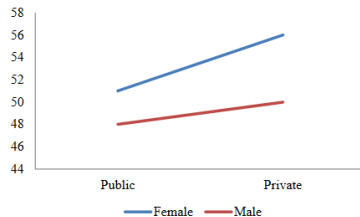
- ▶ Let's take a step back and think more generally about what an *interaction* is really referring to in context.
- ▶ Here, we're ultimately asking: does the Padding type change the force experienced differently depending on the Shell type?
 - ▶ In other words, does foam/gel work better with polycarbonate or new composite?
- ▶ If the answer to this question is “no,” then the prior plot will show us two parallel lines. If the lines are not parallel, then this may indicate the presence of a significant interaction.

Descriptive Analysis: Interaction Plot

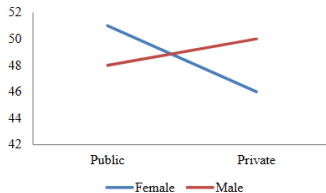
Parallel Lines → No interaction



Nonparallel Lines → interaction



Nonparallel Lines → interaction



Inferential Analysis

- ▶ Now that we have performed our appropriate descriptive analyses, we can go ahead and perform our inferential analyses (assuming all relevant assumptions have been tested...see the associated R code for details).
- ▶ Here, since we have two treatment effects and an interaction effect that are part of our experiment, we technically have three sets of statistical hypotheses we test.

Inferential Analysis

- ▶ **Shell Type:** $H_0 : \alpha_{\text{Polycarbonate}} = \alpha_{\text{New Composite}} = 0$
vs. $H_1 : \alpha_{\text{Polycarbonate}}$ or $\alpha_{\text{New Composite}}$ do not equal 0.
- ▶ **Padding Type:** $H_0 : \beta_{\text{Foam}} = \beta_{\text{Gel}} = 0$ vs. $H_1 : \beta_{\text{Foam}}$ or β_{Gel}
do not equal 0.
- ▶ **Interaction:**
 $H_0 : (\alpha\beta)_{PF} = (\alpha\beta)_{PG} = (\alpha\beta)_{NF} = (\alpha\beta)_{NG} = 0$ vs. $H_1 :$
at least one of these do not equal 0.

Inferential Analysis

- ▶ You may be asking yourself “why are the hypotheses laid out this way, with respect to the effects rather than the means?”
- ▶ The answer: this is what is actually being tested in the F-tests. Since we can write the treatment and interaction group means as functions of these effects, it is sufficient (and traditional) to focus on the effects.
- ▶ But to reiterate, because the effects are functions of the group means, we can still think of these tests as being related to the group means, because they are!

Inferential Analysis

► Our ANOVA table is:

Source	df	SS	MS	F-Statistic
A	$I - 1$	SSA	$SSA/(I - 1)$	MSA/MSE
B	$J - 1$	SSB	$SSB/(J - 1)$	MSB/MSE
AB	$(I - 1)(J - 1)$	$SSAB$	$\frac{SSAB}{(I-1)(J-1)}$	$MSAB/MSE$
Error	$(r - 1)IJ$	SSE	$\frac{SSE}{(r-1)IJ}$	
Total	$rIJ - 1$	SST		

Inferential Analysis

► Let's see what our ANOVA table is in this example:

```
## Fit ANOVA model ##  
mod <- aov(Response ~ Shell*Padding, data=helmet)  
## Obtain ANOVA table ##  
library(broom)  
mod |>  
  tidy()
```

A tibble: 4 x 6

	term <chr>	df <dbl>	sumsq <dbl>	meansq <dbl>	statistic <dbl>	p.value <dbl>
1	Shell	1	2749.	2749.	7.64	0.00895
2	Padding	1	5605.	5605.	15.6	0.000353
3	Shell:Padding	1	1135.	1135.	3.15	0.0842
4	Residuals	36	12958.	360.	NA	NA

Inferential Analysis

- ▶ It is recommended that we evaluate our three hypothesis tests beginning with the interaction.
- ▶ Here, our p-value is 0.08, indicating that our data support the null hypothesis that the interaction effects are not meaningfully different from 0.
- ▶ In other words, while we did see a difference in slope in the interaction plot, the difference was not found to be statistically meaningful.

Inferential Analysis

- ▶ Note, if we were to have found a significant interaction effect, and especially if we were to have seen an intersection of the lines in the interaction plot, then we would have to proceed with caution in our interpretation of the treatment effects (sometimes called the “main effects”)
- ▶ Why is this? The differences we might see when only examining the treatment effects aren't telling the whole story.
- ▶ We need to make sure we are telling the complete story by appropriately accounting for significant interactions.

Inferential Analysis

- ▶ Now, moving into the tests for the two treatment effects, we can see that both Shell and Padding are statistically significant at the 0.05 level.
- ▶ This implies that we have observed a statistically meaningful difference in mean force by Shell and by Padding type.
- ▶ In this case, since we only have two levels of each treatment effect, we don't necessarily have to perform a post-hoc test in order to know which group is greater and which group is lesser in terms of the mean response.
 - ▶ But we could!

Inferential Analysis

```
## Tukey's PostHoc Test ##
```

```
TukeyHSD(mod) |>
```

```
  tidy() |>
```

```
  filter(term %in% c("Shell","Padding")) |>
```

```
  select(term,contrast,estimate,adj.p.value)
```

```
# A tibble: 2 x 4
```

	term	contrast	estimate	adj.p.value
	<chr>	<chr>	<dbl>	<dbl>
1	Shell	Polycarbonate-New Composite	16.6	0.00895
2	Padding	Gel-Foam	-23.7	0.000353

Inferential Analysis

- ▶ So for our contextual conclusion, we would say that if our goal was to manufacture helmets that are safest for players, we should probably use the New Composite material for the shell and the Gel for the interior padding.

Final Thoughts

- ▶ Factorial ANOVA can be extended to more than two treatment effects and include multiway interactions.
- ▶ In theory, while we could have a four or five way interaction, this is almost never practically interpretable. Thus, in most higher-order designs like this, we completely ignore those and mostly focus on the two-way interactions and the main effects.
- ▶ As is often the case in statistical methodology, increased model complexity often comes at the expense of practical interpretability.