

Comparative Experiments - Post Hoc Tests

Dr Austin R Brown

Kennesaw State University

Introduction

- ▶ In the last class, we learned about techniques we can use in a completely randomized design.
 - ▶ The independent means t -test and one-way ANOVA
- ▶ In both the Dog Toys and the Can Cooler example, we only had two groups. So at the conclusion of our t -test, we had explicit evidence that these specific two group means were or were not substantially different from each other.
 - ▶ But what if we have more than two groups?

Example: Breakfast Foods

- ▶ Suppose we are the head chef for a breakfast restaurant chain. We are considering preparing our fried eggs using one of three different techniques:
 - ▶ Technique 1: Fried in butter
 - ▶ Technique 2: Fried in olive oil
 - ▶ Technique 3: Fried in bacon grease
- ▶ Obviously, we want to choose the technique that produces the best tasting eggs. We decide to conduct an experiment to determine which technique produces the best tasting eggs.

Example: Breakfast Foods

- ▶ To conduct the experiment, we decide to host a focus group. Among the 90 focus group participants, we randomly assign 30 to each of the three techniques.
- ▶ After the focus group, we ask each participant to rate the taste of the eggs on a scale of 1 to 10 with higher scores indicating a more positive rating and vice versa. The data is contained in the “Eggs Rating.xlsx” file.

Descriptive Analysis

- In this CRD, where we have a quantitative outcome (egg rating) and the cooking techniques serving as the treatment groups (and the IV), we will first begin by performing some descriptive analysis:

```
library(tidyverse)
library(readxl)
## Read in the Data ##
egg_data <- read_excel("Egg Rating.xlsx")
## Get a Glimpse of the Data ##
egg_data |>
  glimpse()
```

Rows: 90

Columns: 2

\$ Technique <chr> "Butter", "Butter", "Butter", "Butter", "Butter", "Bu

\$ Rating <dbl> 8, 5, 5, 4, 4, 4, 6, 6, 5, 9, 7, 6, 7, 3, 7, 9, 1, 7,

Descriptive Analysis

```
## Get Means and SDs ##  
library(rstatix)  
egg_data |>  
  group_by(Technique) |>  
  get_summary_stats(Rating,type="mean_sd") |>  
  select(Technique,mean,sd)
```

```
# A tibble: 3 x 3  
  Technique      mean      sd  
  <chr>      <dbl> <dbl>  
1 Bacon Grease  8.17  1.39  
2 Butter       5.9   2.23  
3 Olive Oil    6.87  2.21
```

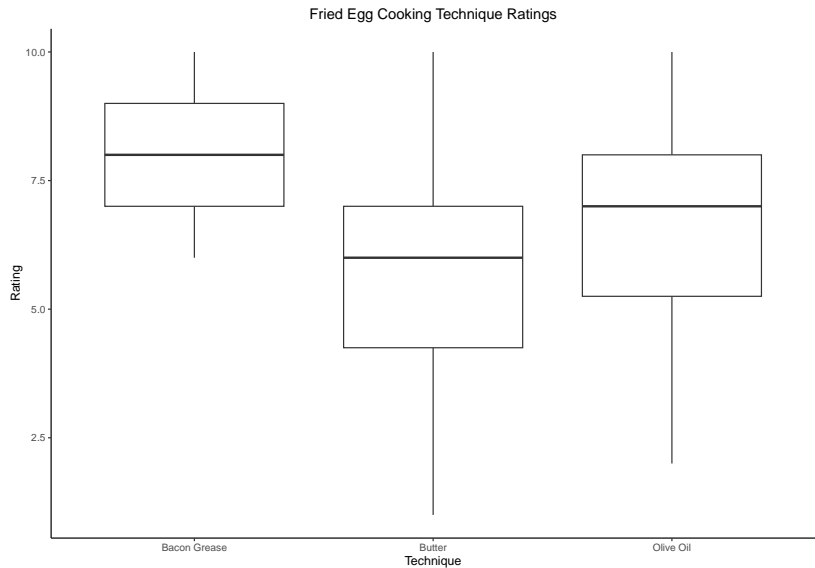
Descriptive Analysis

- ▶ As we can see, the Bacon Grease technique has the highest rating followed by olive oil and butter.
- ▶ Based on the standard deviations, it seems reasonable to assume that our ultimate conclusion will be that Bacon Grease should be the method we go with.
- ▶ But let's continue our systematic evaluation of our experimental data! Let's build a boxplot!

Descriptive Analysis

```
## Build a Boxplot ##  
egg_data |>  
  ggplot(aes(x=Technique,y=Rating)) +  
  geom_boxplot() +  
  labs(title = "Fried Egg Cooking Technique Ratings") +  
  theme_classic() +  
  theme(plot.title=element_text(hjust=0.50))
```


Descriptive Analysis



Descriptive Analysis

- ▶ Here, we can see that the Bacon Grease box doesn't have much overlap with the Butter and Olive Oil boxes, which would indicate that Bacon Grease is likely substantially different than the other two groups.
- ▶ The Butter and Olive Oil boxes do have a good amount of overlap which would potentially indicate that the response patterns between those two groups may not be substantially different.
- ▶ Let's keep this information in mind as we begin our inferential analysis!

Inferential Analysis

- ▶ Here, we will be performing a one-way ANOVA model. The hypotheses we will specifically be testing are:

$$H_0 : \mu_{\text{Bacon Grease}} = \mu_{\text{Butter}} = \mu_{\text{Olive Oil}}$$

$$H_1 : \text{At least one pair of means differ substantially}$$

Inferential Analysis

- ▶ The model we will be building is:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

- ▶ where y_{ij} denotes the egg rating provided by the j th participant in the i th treatment group, μ represents the overall mean egg rating, τ_i represents the treatment effect of the i th group, and ε_{ij} represents the j th random error term in the i th treatment group.

Inferential Analysis

- ▶ Now that we have clearly specified our hypotheses and model, we can go ahead and fit the one-way ANOVA model using the `aov` function.

```
library(broom)
egg_mod <- aov(Rating ~ Technique, data=egg_data)
egg_mod |>
  tidy()
```

```
# A tibble: 2 x 6
```

	term	df	sumsq	meansq	statistic	p.value
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Technique	2	77.6	38.8	9.86	0.000138
2	Residuals	87	342.	3.93	NA	NA

Inferential Analysis

- ▶ Our $F_{\text{Stat}} = 9.86$ and has an associated p-value of 0.0001.
- ▶ This indicates that our data more strongly support the alternative hypothesis, if we use our traditional α -level of 0.05.

Next Steps: Post-Hoc Analysis

- ▶ Okay, great! We've worked our way through the analysis and have concluded that at least two of our techniques may have differing mean ratings.
- ▶ However, when we have more than two treatment groups, what this test cannot tell us is *which* groups differ.
- ▶ Obviously we have a guess based on what we saw in the descriptive analysis. But to evaluate this using an inferential method, we must employ a post-hoc test.

Next Steps: Post-Hoc Analysis

- ▶ There exist many techniques for post-hoc analysis.
- ▶ The one we will use in this course is called *Tukey's Honestly Significant Difference (HSD)* method.
- ▶ Let's see how it works!!

Tukey's HSD Post-Hoc Method

- ▶ We perform this test for each pair of means. For a given pair of means, say i and i^* , our statistical hypotheses are:

$$H_0 : \mu_i = \mu_{i^*}$$

$$H_1 : \mu_i \neq \mu_{i^*}$$

Tukey's HSD Post-Hoc Method

- ▶ Our test statistic is:

$$q_{\text{Stat}} = \frac{|\bar{x}_i - \bar{x}_{i^*}|}{\sqrt{\frac{MSE}{2} \left(\frac{1}{r_i} + \frac{1}{r_{i^*}} \right)}}$$

- ▶ This test statistic follows a distribution known as the *Studentized Range* distribution, typically denoted as q .

Tukey's HSD Post-Hoc Method

- ▶ If $q_{\text{Stat}} > q_{cv}$, then this indicates to us that these two groups, i and i^* , have means which differ substantially.
- ▶ While we could do this by hand, we can have R perform this analysis for us using the TukeyHSD function

Tukey's HSD Post-Hoc Method

```
## Perform Tukey's HSD ##  
TukeyHSD(egg_mod) |>  
  tidy() |>  
  select(contrast, estimate, adj.p.value)
```

A tibble: 3 x 3

contrast	estimate	adj.p.value
<chr>	<dbl>	<dbl>
1 Butter-Bacon Grease	-2.27	0.0000818
2 Olive Oil-Bacon Grease	-1.30	0.0342
3 Olive Oil-Butter	0.967	0.148

Tukey's HSD Post-Hoc Method

- ▶ These results, as shown by the adjusted $p < 0.05$, indicate that the mean rating for Bacon Grease is greater than both the Butter and Olive Oil mean rating.
- ▶ However, the difference in mean rating between Olive Oil and Butter was not found to be statistically different.
- ▶ So as we noted during the descriptive analysis, the head chef should cook fried eggs in bacon grease based on the result of this experiment.

Why Post-Hoc?

- ▶ You may be asking yourself, why couldn't we just use several independent means t -tests to draw the same conclusion?
- ▶ While this may seem like a good idea, we actually run into a big problem when performing multiple comparisons.
- ▶ Namely, our Type I error rate (the probability of a test signaling in favor of H_1 when H_0 is actually true) increases!

Why Post-Hoc?

- ▶ Post-Hoc methods, like Tukey's HSD, adjust the Type I error rate to maintain the desired overall significance level, α .
- ▶ Moreover, Post-Hoc tests simultaneously compare all group means, which provides a more efficient approach in identifying potentially significant differences.
- ▶ Finally, good Post-Hoc tests, like Tukey's HSD, are designed to work with, not independently of, the ANOVA F-test.