

# Comparative Experiments - An Introduction to the Effects Model and Completely Randomized Designs

Dr Austin R Brown

Kennesaw State University

# Introduction

- ▶ In the last section, we learned some of the basics of experimental design.
- ▶ In this section, we will begin to learn about specific *designs* we can employ to analyze our data and answer our research questions.
- ▶ Let's begin by discussing an example.

# Introduction

- ▶ Suppose we work for a company that manufactures high-end canned beverage coolers (like koozies!). We are considering launching a new cooler and have two designs in mind. One design uses hard plastic and the other uses stainless steel. Since our goal is to maximize the amount of time the beverage stays cold, we want to know which design is better at keeping the beverage cold.

# Planning the Experiment

- ▶ Now that we have a general sense of what our goal is with the experiment, we need to plan the experiment.
  
- ▶ Recall from the last section that the steps in planning a successful experiment are:
  1. Define the objective
  2. Decide what the outcome is (and how it will be measured)
  3. Determine the independent variables and possible lurking variables
  4. Choose the design
  5. Be clear on data collection processes/procedures
  6. Be clear on which analyses will be performed
  7. Draw conclusions

# Planning the Experiment: Define the Objective

- ▶ First, we must define the objective.
- ▶ Here, our objective is to determine which of the two cooler materials is better at keeping canned beverages cold.

## Planning the Experiment: Decide on the Outcome

- ▶ Next, we must decide what our outcome variable is and how we will measure it.
- ▶ You can probably think of several ways to do this, but for the sake of this example, let's define the outcome as the temperature of the beverage (in degrees Fahrenheit) after it has been in the cooler for two hours.

## Planning the Experiment: Determine the Independent Variable & Possible Lurking Variables

- ▶ Now, we need to decide what our independent variable will be and what potential lurking variables may exist.
- ▶ Our IV will be the material type.
- ▶ While we may have some lurking variables depending on the way we conduct the experiment, we can eliminate many if not all of them by having careful control over:
  - ▶ Starting temperature of the canned beverages
  - ▶ Consistency of room temperature in room where the cans will be set out for two hours

## Planning the Experiment: Determine the Independent Variable & Possible Lurking Variables

- ▶ To control the starting temperature of the canned beverages, we can keep the cans in the same refrigerator for 24 hours prior to the experiment.
- ▶ To control the consistency of the room temperature, we will place the cans on a table in a room kept at a consistent temperature and with the window shades closed.



## Planning the Experiment: Choose the Design

- ▶ In Chapter 1 of our text (and also in our syllabus), you can see that we have lots of different designs to choose from (e.g., completely randomized design, split-plot,  $2^k$ , etc.).
- ▶ The choice of design is largely a function of our independent variable(s) and how we are conducting the experiment.
- ▶ Perhaps the most important element of any experimental design are the concepts of *replication* and *randomization*.

## Planning the Experiment: Choose the Design

- ▶ **Randomization** means that our experimental units are randomly partitioned into treatment groups.
  - ▶ Recall from the email marketing example in the last section, our customer base was randomly assigned to be in either the generic email group or the personalized email group. This is an example of randomization.
- ▶ Randomization is an important way to eliminate potential sources of variability to ensure that the conclusions we draw are accurate.
- ▶ In the email example, imagine we didn't do randomization and we accidentally assigned all female customers to the generic group and all male customers to the personalized group.
  - ▶ Whatever differences we observe between the email groups will be *confounded* with customer sex.

## Planning the Experiment: Choose the Design

- ▶ **Replication** refers to having more than one experimental unit in a specific treatment group.
- ▶ In our cooler example, we would not perform the temperature test by using only one can in the hard plastic cooler and one in the stainless steel cooler. We would want more *replicates* (typically denoted by  $r$ ).
  - ▶ Why? Two reasons!
- ▶ First, without replication, we wouldn't be able to tell if treatment differences are real or just a random manifestation of the particular experimental units used in the study.

## Planning the Experiment: Choose the Design

- ▶ Second and relatedly, when we only have one replicate, we don't have variance (the variance of a constant is 0, as you will learn in mathematical statistics).
- ▶ And if we don't have variance, none of our hypothesis tests will work!
- ▶ Deciding on the value of  $r$  can be a statistical conversation (i.e., power analysis) but often it's a practical one (how much time or money will it cost me to produce a certain number of replicates?).

## Planning the Experiment: Choose the Design

- ▶ Now, circling back to choosing a specific design, in our case, our experimental units (the canned beverages) are essentially the same. Through our procedure of performing the experiment, we've eliminated any other major lurking variables.
- ▶ Thus, the best design for comparing the effects of our two treatment groups would be a **completely randomized design**.
- ▶ In a CRD, all experimental units are randomly assigned to a treatment and each unit has an equal chance of being in one of the treatment groups.
  - ▶ For us, we could number each can and use the R function `sample` in order to determine which can goes into which cooler.

## Planning the Experiment: Choose the Design

- ▶ Suppose we decide to have  $r = 30$  replicates for each of our  $t = 2$  treatment groups to make for a total sample size of  $n = r \times t = 30 \times 2 = 60$ .
- ▶ We would number our cans 1 - 60. Then using sample:

```
set.seed(123)
cans <- 1:60
s1 <- sample(cans,size=30)
plastic <- cans[s1]
steel <- cans[-s1]
print(cbind(plastic,steel)[1:3,])
```

	plastic	steel
[1,]	31	1
[2,]	15	2
[3,]	51	4

## Planning the Experiment: Choose the Design

- ▶ So the first three cans in the plastic group are cans 31, 15, and 51.
- ▶ The first three cans in the stainless steel group are 1, 2, and 4.

# Planning the Experiment: Be Clear on Data Collection Processes and Procedures

- ▶ We've already discussed some of our processes and procedures but we should go into more detail on some others!
- ▶ For instance, the canned beverage we choose should be exactly the same one. So let's choose regular 12 ounce Coca-Cola cans.
- ▶ The 60 Coke cans will go into the fridge at exactly the same time and all removed at the exact same time.



## Planning the Experiment: Be Clear on Data Collection Processes and Procedures

- ▶ It would be best if we could have a team of people working together to place the cans into the coolers to ensure that they spend almost exactly the same amount of time in their respective coolers.
- ▶ To measure beverage temperature, we will open all of the cans at nearly the same time and use a digital thermometer to measure the temperature of each canned beverage.
- ▶ Measurements for each can will be recorded in an Excel Spreadsheet (contained in the "Can Temperature.xlsx" file)

## Planning the Experiment: Be Clear on the Analyses to be Performed

- ▶ Generally when performing any analysis, I like to perform both a descriptive analysis as well as an inferential analysis.
- ▶ Doing both tends to yield more information rather than just one or the other.
- ▶ In our case, since we're comparing a quantitative variable across two groups, we have specific descriptive and inferential techniques to analyze these data:

# Descriptive Analyses: Means and Standard Deviations

- ▶ Perhaps the most straightforward descriptive analyses we could do in the case of the cooler data is calculate the sample mean temperature and sample standard deviation temperature for each group.
- ▶ Recall, the sample mean for group  $i$  is defined as:

$$\bar{y}_{i\cdot} = \frac{1}{r_i} \sum_{j=1}^{r_i} y_{ij}$$

# Descriptive Analyses: Means and Standard Deviations

- ▶ Here,  $r_i$  is the sample size (number of replicates) for group  $i$ .
- ▶ We interpret the mean as the expected value of a group of data points.
- ▶ In other words, it's our best guess as to the value of an experimental unit in a given group.

# Descriptive Analyses: Means and Standard Deviations

- ▶ The sample standard deviation is calculated by:

$$s_i = \sqrt{\frac{1}{r_i - 1} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i.})^2}$$

- ▶ Very simply, we can think of sample standard deviation as our best guess as to how far away a given observation is from the sample mean, on average (in both the positive and negative direction).

## Descriptive Analyses: Means and Standard Deviations

- ▶ To perform these calculations in R, I like to use a combination of the `group_by` function from the `dplyr` package and the `get_summary_stats` function from the `rstatix` package

# Descriptive Analyses: Means and Standard Deviations

```
## Load the required libraries ##  
library(tidyverse)  
library(readxl)  
library(rstatix)  
## Read in the Can Data ##  
can_data <- read_excel("Can Temperature.xlsx")  
can_data |>  
  group_by(Treatment) |>  
  get_summary_stats(Temperature, type = "mean_sd")
```

```
# A tibble: 2 x 5
```

	Treatment	variable	n	mean	sd
	<chr>	<fct>	<dbl>	<dbl>	<dbl>
1	Plastic	Temperature	30	34.4	1.84
2	Steel	Temperature	30	33.1	1.98

## Descriptive Analyses: Means and Standard Deviations

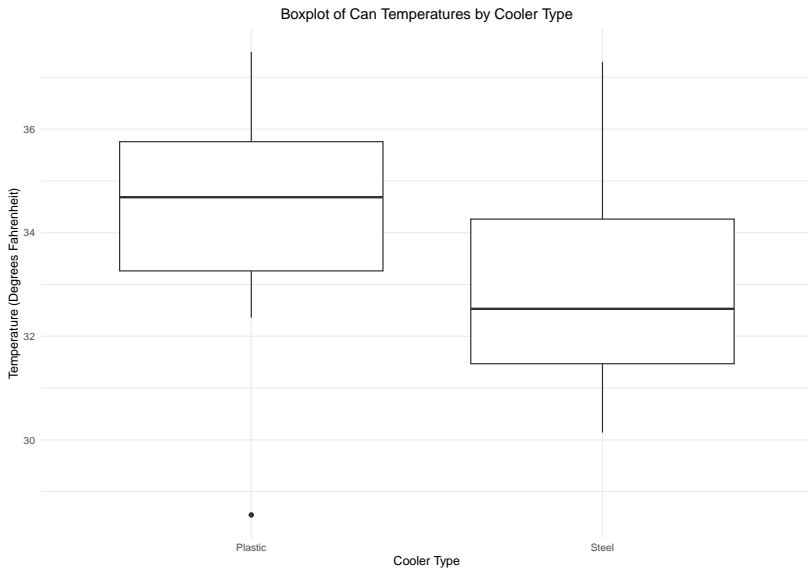
- ▶ So we can see from these data that the mean temperature of the cans of Coca-Cola were 34.43 degrees for the plastic coolers and 33.07 degrees for the steel coolers.
- ▶ In the context of beverage coolers, a roughly 1.5 degree difference may be meaningfully large!
- ▶ Let's see how we can visualize these differences using the boxplot:



# Descriptive Analyses: Boxplots

```
can_data |>
  ggplot(aes(x = Treatment, y = Temperature)) +
  geom_boxplot() +
  labs(title = "Boxplot of Can Temperatures by Cooler Type",
       x = "Cooler Type",
       y = "Temperature (Degrees Fahrenheit)") +
  theme_minimal() +
  theme(plot.title=element_text(hjust=0.50))
```

# Descriptive Analyses: Boxplots



## Descriptive Analyses: Boxplots

- ▶ Remember, a boxplot consists of 5 numbers. When we look at the box itself, the bottom, middle-line, and top of the box represent the 25th, 50th, and 75th percentiles, respectively.
  - ▶ We can think of the width of the box (interquartile range, the difference between the 75th and 25th percentiles) as a measure of variability.
- ▶ The lines going out of the top and bottom of the box have a length of 1.5 times the interquartile range.
- ▶ Any points plotting above or below the upper or lower line can generally be thought of as outliers.

## Descriptive Analyses: Boxplots

- ▶ The way I generally interpret boxplots is by looking at the degree of overlap.
- ▶ If the boxes have substantial overlap, that means they have similar distributions and likely don't meaningfully differ.
- ▶ If the boxes do not substantially overlap, then that means the distributions have some degree (perhaps a meaningful degree!) of difference.
- ▶ In our case, there is some overlap but not one I would consider substantial. It appears, empirically, that we have evidence of a meaningful difference in the can temperatures between the two treatment groups.

# Inferential Analysis: The Effects Model

- ▶ After we perform our descriptive analysis, the next step is to perform the inferential analysis.
- ▶ In general, we can describe pretty much every **supervised learning technique** (that is, a model with an outcome variable and at least one explanatory/independent variable) can be described as a function.
- ▶ Remember from regression, the simple linear regression model can be written as:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

# Inferential Analysis: The Effects Model

- ▶ When working with categorical independent variables, like we are here, we can still write our model as a regression model, but we generally use slightly different notation:

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

- ▶ This specific written model is called the **effects model**.  $y_{ij}$  represents the  $j$ th replicate in the  $i$ th treatment group,  $\mu$  represents the grand or overall mean of all the observations, the  $\tau_i$ 's are called the *effects* (the difference between the individual group means and the grand mean), and  $\varepsilon_{ij}$  denotes the residual (or random error) term associated with the  $j$ th replicate in the  $i$ th treatment group.

# Inferential Analysis: The Effects Model

- ▶ Akin to regression, we have distributional assumptions in the effects model. Specifically:

$$Y_{ij} \sim N(\mu + \tau_i, \sigma^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

# Inferential Analysis: The Effects Model

- ▶ So once we have collected real data like we have in the `can_data` dataframe, how do we go about estimating  $\mu$  and the  $\tau_i$ 's?
- ▶ Just like we do in regression, we use least-squares estimation! The LSEs for the group means are the same calculations given for the sample means previously in these slides.



# Inferential Analysis: The Effects Model

- ▶ The LSE for  $\mu$ , say  $\hat{\mu}$ , is given by:

$$\hat{\mu} = \bar{y}_{..} = \frac{1}{t} \sum_{i=1}^t \bar{y}_i.$$

- ▶ Thus, the  $\tau_i$  LSEs, say  $\hat{\tau}_i$ , is given by:

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$$

# Inferential Analysis: The Effects Model

- ▶ Let's see how we can estimate these values with the `car_data` using R:

```
## Calculate Estimates for Grand Mean ##  
mu_hat <- mean(car_data$Temperature)  
## Now, let's save those group means we calculated earlier ##  
group_means <- car_data |>  
  group_by(Treatment) |>  
  summarise(mean_temp = mean(Temperature))  
## Finally, calculate the Tau hats ##  
tau_hat <- group_means$mean_temp - mu_hat
```

# Inferential Analysis: The Effects Model

```
print(list(`Grand Mean` = mu_hat,  
          `Group Means` = group_means,  
          `Tau Hats` = tau_hat))
```

```
$`Grand Mean`
```

```
[1] 33.75
```

```
$`Group Means`
```

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

```
  Treatment mean_temp
```

```
  <chr>          <dbl>
```

```
1 Plastic          34.4
```

```
2 Steel            33.1
```

```
$`Tau Hats`
```

```
[1] 0.6813333 -0.6813333
```

# Inferential Analysis: The Effects Model

- ▶ Note, in this case where we only have two treatment groups, the  $\hat{\tau}_i$ 's will just be the opposite sign of each other.
- ▶ Okay, now that we've put everything together, what statistical test do we run to determine if the difference we see in the group means and subsequently the  $\hat{\tau}_i$ 's is statistically meaningful or not?

# Inferential Analysis: The Effects Model

- ▶ First, what are our statistical hypotheses?
- ▶ In words, the null hypothesis,  $H_0$ , is that all the group means are equal. In mathematical notation:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_i$$

- ▶ In words, the alternative hypothesis,  $H_1$ , is that at least two of the group means differ. The mathematical notation is essentially the same as the words.

# Inferential Analysis: The Effects Model

► In our case:

$$H_0 : \mu_{\text{Plastic}} = \mu_{\text{Steel}}$$

$$H_1 : \mu_{\text{Plastic}} \neq \mu_{\text{Steel}}$$

# Inferential Analysis: The Effects Model

- Our next step would be to calculate our test statistic. For the effects model, what is our test statistic?

$$F_{\text{Stat}} = \frac{SSTreat/(t-1)}{SSE/(n-t)} \sim F(t-1, n-t)$$

$$SSTreat = \sum_{i=1}^t r_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$SSE = \sum_{i=1}^t \sum_{j=1}^{r_t} (y_{ij} - \bar{y}_{i.})^2$$

# Inferential Analysis: The Effects Model

- ▶ If our  $F_{\text{Stat}} > F_{cv}$ , where we typically choose  $F_{cv}$  such that  $\alpha = 0.05$ , then this indicates that the data more strongly support the alternative hypothesis.
- ▶ If the converse is true,  $F_{\text{Stat}} < F_{cv}$ , then this indicates that the data more strongly support the null hypothesis.



# Inferential Analysis: The Effects Model

- How do we calculate all of this in R? Fortunately for us, we can do this with the `aov` function (`aov` being short for “analysis of variance”):

```
## Build the AOV Model ##  
library(broom)  
can_aov <- aov(Temperature~Treatment,data=can_data)  
can_aov |>  
  tidy()
```

```
# A tibble: 2 x 6  
  term          df sumsq meansq statistic  p.value  
  <chr>      <dbl> <dbl>   <dbl>      <dbl>   <dbl>  
1 Treatment         1  27.9   27.9        7.63  0.00769  
2 Residuals       58 212.    3.65         NA      NA
```

## Inferential Analysis: The Effects Model

- ▶ As we can see, the F-statistic is 7.63 and its associated p-value is 0.008.
- ▶ Since p (which is the probability of observing a test statistic of greater magnitude in a world where we assume  $H_0$  is true) is less than our typically used  $\alpha$  threshold of 0.05, this indicates to us that our data more strongly support the alternative hypothesis.
- ▶ In more contextual language, our data support the notion that the stainless steel cooler keeps canned beverages meaningfully colder than the hard plastic cooler.

## Inferential Analysis: The Two Group Case

- ▶ Now, in the case when we have only two treatment groups, like we do here, this F-test can be somewhat simplified into an independent means  $t$ -test.
- ▶ The statistical hypotheses stay the same but the test statistic can be rewritten as:

$$t_{\text{Stat}} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_{\text{Pooled}}^2 \left( \frac{1}{r_1} + \frac{1}{r_2} \right)}} \sim t(r_1 + r_2 - 2)$$

# Inferential Analysis: The Two Group Case

► where:

$$s_{\text{Pooled}}^2 = \frac{(r_1 - 1)s_1^2 + (r_2 - 1)s_2^2}{r_1 + r_2 - 2}$$

► If  $|t_{\text{Stat}}| > t_{cv}$ , then we conclude that our data more strongly support the alternative hypothesis.

## Inferential Analysis: The Two Group Case

- Note,  $t_{\text{Stat}}^2 = F_{\text{Stat}}$ . So technically an independent means t-test can be performed as a one-way ANOVA with two groups.

## Your Turn!

- ▶ Let's say we work for a dog toy manufacturing company. We want to manufacture rubber tennis balls for dogs to be able to play with and chew on that are designed to last. As you may know, some dogs who are of certain breeds and/or are aggressive chewers can break toys very easily. We want to avoid that as much as possible with our rubber tennis balls. Our engineers have come up with two different formulations of hard rubber that can be used to manufacture these balls. Let's denote these as Formula 1 and Formula 2.
- ▶ To decide which formulation to use, we want to perform a pressure test to simulate the pressure these rubber tennis balls might experience being clamped down on by a dog's jaw. We test 60 balls of each formulation by putting them each under a hydraulic press and recording the amount of pressure exerted (in PSI) when the ball rips.
- ▶ The data for this experiment are contained in the "Dog

# Your Turn!

- ▶ With these data, I want you to:
  1. Specify the outcome and independent variables. What lurking variables might be present?
  2. Briefly explain why a completely randomized design might be appropriate in this context
  3. Perform an appropriate exploratory analysis
  4. Perform an appropriate inferential analysis (including the specification of  $H_0$  and  $H_1$ )
  5. Provide contextual conclusions – which formulation seems more durable?
  6. What questions remain?