Lesson 8: Chip Melting Times

LTC James K Starling

Background

Industrial-scale baking requires a precise understanding of how recipe ingredients behave when baked. For example, bakers want to know and understand how different types of baking chips respond to different oven temperatures. Researchers selected a sample of 31 participants to compare the melting times of chocolate chips and butterscotch chips. The response variable will be the amount of time, from placing a chip on a participant's tongue and holding it steadily to the roof of their mouth, until the chip is completely melted.

```
library(tidyverse)
library(pwr) # see https://www.statmethods.net/stats/power.html for good tutorial.

## Load data
chipmelt <- read.table('http://www.isi-stats.com/isi2/data/MeltingTimesUnstacked.txt', header = TRUE, s

# chipmelt2 <- read.table('MeltingTimes.txt', header = TRUE, stringsAsFactors = TRUE)
```

Step 1: Ask a research question: "Is there an association between chip type and melting time?"

Step 2: Design a study and collect data:

(1) Create a sources of variation table.

Observed variation in:	Sources of explained variation:	Sources of unexplained variation:

Inclusion criteria: ?? (class?
people?)
Design:

(2) Develop a study protocol to investigate this question. What is your inclusion criteria? Will you use randomization?

(3) Think about the sources of variation diagram. Do any of your answers have anything in common? Can you suggest a way to improve the model?

(4) Is a repeated measures type of matched pairs design feasible for this study? Describe how it would work and the benefits. Would randomization still need to apply here? How?

(5) Update the sources of variation table when using a repeated measures, matched pairs design.

Observed variation in:	Sources of explained variation:	Sources of unexplained variation:
Inclusion criteria: ?? (class? people?)		
Design:		

(6) Would a two-sample t-test (or separate means model) as in Chapter 1 be an appropriate analysis to compare the two melting time means for our matched pairs study design? Explain.

Step 3: Explore the Data.

(7) The data set you are currently working with is in what we consider "wide format" where each participant appears in only one row. Often times in our analysis it helps to have our data in "long format" where participants appear multiple times based on the levels of another variable. Use the `pivot_longer()` to pivot the data set and rename it `chips_long`. Keep in mind that you must tell R the columns you want to pivot (`cols =`), the new variable that you want to contain the current variable names of the columns you are pivoting (`names_to =`) and the new variable that you want to contain the current values of the columns you are pivoting (`values_to =`). Hint: use the help file in R for `pivot_longer`. Update the "chip" column to be a factor variable.

```
# chips_long <- chipmelt %>% pivot_longer(...
```

```
# chips_long <- chips_long %>% mutate(...
```

(8) Create a dotplot (using `ggplot` and `geom_dotplot`), stratified by chip type. Does it appear that one chip has a longer melting time than the other?

```
# chips_long %>% ggplot(...
```

(9) Calculate the effect of each type of chip.

Step 4: Draw inferences beyond the data.

(10) Conduct a separate means model using the type of chip as explanatory variable, but use effect encoding (use `contrasts()` and `contr.sum`). This will allow us to use R to find the correct R-squared value (and other things)). Also use `anova()` and `aov()` to calculate SST, SSE, SSM, F statistic, p-value.

```
# contrasts(xxx) <- xxx
# contrasts(xxx)
#
# eff.model <- xxx
# summary(eff.model)
```

(11) How do we take into account the 'person effect'? Two approaches: 1) take the differences for each individual. 2) Use a separate means model. Which method is preferred and why?

(12) Conduct a separate means model using the type of chip as explanatory variable and account for the person to person variation. Use effect encoding (use `contrasts()`) as we did in the previous model. Provide the Rsquared, SST, SSE, SSM, F statistic(s), p-value(s).

```
# contrasts(xxx) <- contr.sum
# # contrasts(chips_long$student)
#
# eff.model2 <- xxx
# summary(eff.model2)
# anova(eff.model2)
```

Step 5: Formulate conclusions.

(13) After accounting for student to student variation is there a significant difference between the melting time of the two chip types?