Lesson 8: Chip Melting Times

LTC James K Starling

Last compiled on 27 January, 2022

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, stringsAsFactors = TRU
E)

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:
melting time	type of chip (chocolate/butterscotch)	body temperature
		saliva amount (by person)
		chip size
Inclusion criteria: ?? (class? people?)		chip temperature
Design: same technique for "melting technique" for each person		Unknown

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

Have a lot of participants and randomly assign them to a chocolate or butterscotch chip. Use a stopwatch to measure how long the chip is completely dissolved. Inclusion criteria can be any group within reason.

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?

Many of the unexplained sources of variation come from the (same) participant. We could improve the model by having each participant melt both types of chips.

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?

Yes a repeated measures, matched pairs design is feasible. Each participant can receive both treatments (both chips). This would keep person-specific variables more constant (and reduce variability). Yes, randomization will still need to apply. If we don't randomize order of taking the chip we will have the order confound the treatment.

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:
melting time	type of chip (chocolate/butterscotch)	chip size
	body temperature	chip temperature
	saliva amount (by person)	Unknown
Inclusion criteria: ?? (class? people?)		
Design: same technique for "melting technique" for each person		

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.

No. The two-sample t-test requires independent observations. Since we have people take both treatments we cannot assume independence.

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`.

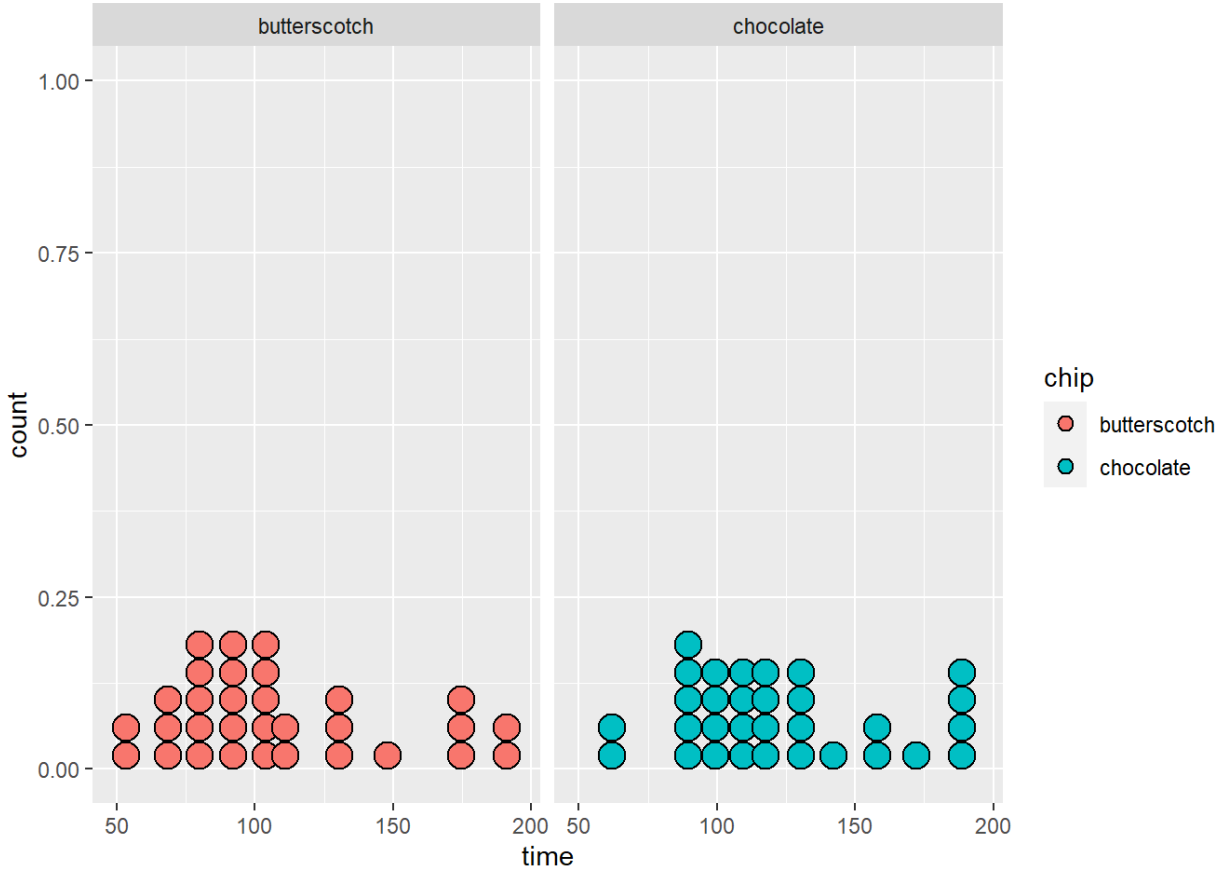
```
chips_long <- chipmelt %>% pivot_longer(cols=-student, names_to = "chip", values_to = "time")
```

Update the “chip” column to be a factor variable.

```
chips_long <- chips_long %>% mutate(chip = as.factor(chip))
```

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(aes(x=time, fill = chip)) + geom_dotplot(binwidth = 10) + facet_wrap(~chip)
```



Yes, it appears that the chocolate chips take slightly longer to melt.

9. Calculate the effect of each type of chip.

```
mn <- mean(chips_long$time)
chips_long %>% group_by(chip) %>% summarise(mean(time) - mn)
```

chip <fct>	mean(time) - mn <dbl>
butterscotch	-7.048387
chocolate	7.048387

2 rows

10. Conduct a separate means model using the type of chip as our explanatory variable. Remember: we can't use the R-squared value here.

```
sep.model <- lm(time~0+chip, data=chips_long)
summary(sep.model) # do not use R-squared!
```

```
##
## Call:
## lm(formula = time ~ 0 + chip, data = chips_long)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -62.484 -27.460  -7.935  20.339  86.613
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## chipbutterscotch  108.387      6.795   15.95   <2e-16 ***
## chipchocolate     122.484      6.795   18.02   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.83 on 60 degrees of freedom
## Multiple R-squared:  0.9061, Adjusted R-squared:  0.903
## F-statistic: 289.7 on 2 and 60 DF,  p-value: < 2.2e-16
```

Step 4: Draw inferences beyond the data.

11. 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(chips_long$chip) <- contr.sum
contrasts(chips_long$chip)
```

```
##             [,1]
## butterscotch    1
## chocolate      -1
```

```
eff.model <- lm(time ~ chip, data = chips_long)
summary(eff.model)
```

```
##
## Call:
## lm(formula = time ~ chip, data = chips_long)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -62.484 -27.460  -7.935  20.339  86.613
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  115.435      4.805  24.024   <2e-16 ***
## chip1         -7.048      4.805  -1.467    0.148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.83 on 60 degrees of freedom
## Multiple R-squared:  0.03462,    Adjusted R-squared:  0.01853
## F-statistic: 2.152 on 1 and 60 DF,  p-value: 0.1476
```

```
aov(eff.model)
```

```
## Call:
##    aov(formula = eff.model)
##
## Terms:
##                      chip Residuals
## Sum of Squares   3080.15  85887.10
## Deg. of Freedom        1        60
##
## Residual standard error: 37.83453
## Estimated effects are balanced
```

```
anova(eff.model)
```

	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
chip	1	3080.145	3080.145	2.151763	0.1476284
Residuals	60	85887.097	1431.452	*NA*	*NA*

2 rows

Rsq = 0.03462, SST = 88,967.24, SSE = 85,894.42, SSM (SSChip) = 3,072.82, F-stat = 2.1518, pvalue = 0.1476.

12. 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?

For the differences model we can conduct a regular t-test. However, this will only tell us that there is a significant association between chip type and melting time. The separate means modle will tell us how much of the variation in mean melting time is explained by the chip type used.

13. 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(chips_long$student) <- contr.sum
contrasts(chips_long$student)
```

```
##     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14]
## BB     1    0    0    0    0    0    0    0    0     0     0     0     0     0
## CL     0    1    0    0    0    0    0    0    0     0     0     0     0     0
## DC     0    0    1    0    0    0    0    0    0     0     0     0     0     0
## EH     0    0    0    1    0    0    0    0    0     0     0     0     0     0
## EP     0    0    0    0    1    0    0    0    0     0     0     0     0     0
## GA     0    0    0    0    0    1    0    0    0     0     0     0     0     0
## HH     0    0    0    0    0    0    1    0    0     0     0     0     0     0
## IQ     0    0    0    0    0    0    0    1    0     0     0     0     0     0
## JO     0    0    0    0    0    0    0    0    1     0     0     0     0     0
## JR     0    0    0    0    0    0    0    0    0     1     0     0     0     0
## JS     0    0    0    0    0    0    0    0    0     0     1     0     0     0
## JvK    0    0    0    0    0    0    0    0    0     0     0     1     0     0
## KH     0    0    0    0    0    0    0    0    0     0     0     0     1     0
## KH2    0    0    0    0    0    0    0    0    0     0     0     0     0     1
## KH3    0    0    0    0    0    0    0    0    0     0     0     0     0     0
## KJ     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## LC     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## LO     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## LR     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## MB     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## MG     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## ML     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## NH     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## NL     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## NN     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## PS     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## RUR    0    0    0    0    0    0    0    0    0     0     0     0     0     0
## SG     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## SS     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## TA     0    0    0    0    0    0    0    0    0     0     0     0     0     0
## TS    -1   -1   -1   -1   -1   -1   -1   -1   -1    -1    -1    -1    -1    -1
##     [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25] [,26]
## BB      0     0     0     0     0     0     0     0     0     0     0     0
## CL      0     0     0     0     0     0     0     0     0     0     0     0
## DC      0     0     0     0     0     0     0     0     0     0     0     0
## EH      0     0     0     0     0     0     0     0     0     0     0     0
## EP      0     0     0     0     0     0     0     0     0     0     0     0
## GA      0     0     0     0     0     0     0     0     0     0     0     0
## HH      0     0     0     0     0     0     0     0     0     0     0     0
## IQ      0     0     0     0     0     0     0     0     0     0     0     0
## JO      0     0     0     0     0     0     0     0     0     0     0     0
```

```
## JR      0     0     0     0     0     0     0     0     0     0     0     0
## JS      0     0     0     0     0     0     0     0     0     0     0     0
## JvK     0     0     0     0     0     0     0     0     0     0     0     0
## KH      0     0     0     0     0     0     0     0     0     0     0     0
## KH2     0     0     0     0     0     0     0     0     0     0     0     0
## KH3     1     0     0     0     0     0     0     0     0     0     0     0
## KJ      0     1     0     0     0     0     0     0     0     0     0     0
## LC      0     0     1     0     0     0     0     0     0     0     0     0
## LO      0     0     0     1     0     0     0     0     0     0     0     0
## LR      0     0     0     0     1     0     0     0     0     0     0     0
## MB      0     0     0     0     0     1     0     0     0     0     0     0
## MG      0     0     0     0     0     0     1     0     0     0     0     0
## ML      0     0     0     0     0     0     0     1     0     0     0     0
## NH      0     0     0     0     0     0     0     0     1     0     0     0
## NL      0     0     0     0     0     0     0     0     0     1     0     0
## NN      0     0     0     0     0     0     0     0     0     0     1     0
## PS      0     0     0     0     0     0     0     0     0     0     0     1
## RUR     0     0     0     0     0     0     0     0     0     0     0     0
## SG      0     0     0     0     0     0     0     0     0     0     0     0
## SS      0     0     0     0     0     0     0     0     0     0     0     0
## TA      0     0     0     0     0     0     0     0     0     0     0     0
## TS     -1    -1    -1    -1    -1    -1    -1    -1    -1    -1    -1    -1
##     [,27] [,28] [,29] [,30]
## BB      0     0     0     0
## CL      0     0     0     0
## DC      0     0     0     0
## EH      0     0     0     0
## EP      0     0     0     0
## GA      0     0     0     0
## HH      0     0     0     0
## IQ      0     0     0     0
## JO      0     0     0     0
## JR      0     0     0     0
## JS      0     0     0     0
## JvK     0     0     0     0
## KH      0     0     0     0
## KH2     0     0     0     0
## KH3     0     0     0     0
## KJ      0     0     0     0
## LC      0     0     0     0
## LO      0     0     0     0
## LR      0     0     0     0
```

```
## MB      0     0     0     0
## MG      0     0     0     0
## ML      0     0     0     0
## NH      0     0     0     0
## NL      0     0     0     0
## NN      0     0     0     0
## PS      0     0     0     0
## RUR     1     0     0     0
## SG      0     1     0     0
## SS      0     0     1     0
## TA      0     0     0     1
## TS     -1    -1    -1    -1
```

```
eff.model2 <- lm(time ~ chip + student, data = chips_long)
summary(eff.model2)
```

```
##
## Call:
## lm(formula = time ~ chip + student, data = chips_long)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -23.452  -4.202   0.000   4.202  23.452
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 115.43548    1.86584  61.868  < 2e-16 ***
## chip1        -7.04839    1.86584  -3.778 0.000701 ***
## student1      0.06452   10.21965   0.006 0.995005
## student2     68.06452   10.21965   6.660 2.24e-07 ***
## student3    -35.93548   10.21965  -3.516 0.001414 **
## student4    -50.93548   10.21965  -4.984 2.44e-05 ***
## student5     -4.93548   10.21965  -0.483 0.632644
## student6    -16.93548   10.21965  -1.657 0.107912
## student7      0.56452   10.21965   0.055 0.956315
## student8     40.56452   10.21965   3.969 0.000415 ***
## student9    -33.43548   10.21965  -3.272 0.002690 **
## student10    18.06452   10.21965   1.768 0.087295 .
## student11    16.06452   10.21965   1.572 0.126457
## student12    -2.43548   10.21965  -0.238 0.813258
## student13    -1.93548   10.21965  -0.189 0.851064
## student14   -16.93548   10.21965  -1.657 0.107912
## student15   -41.93548   10.21965  -4.103 0.000287 ***
## student16   -17.93548   10.21965  -1.755 0.089468 .
## student17   -17.43548   10.21965  -1.706 0.098328 .
## student18    26.06452   10.21965   2.550 0.016103 *
## student19    -7.93548   10.21965  -0.776 0.443537
## student20   -27.43548   10.21965  -2.685 0.011711 *
## student21   -11.93548   10.21965  -1.168 0.252047
## student22    68.06452   10.21965   6.660 2.24e-07 ***
## student23   -14.43548   10.21965  -1.413 0.168087
## student24     5.06452   10.21965   0.496 0.623811
## student25    69.56452   10.21965   6.807 1.50e-07 ***
## student26   -26.43548   10.21965  -2.587 0.014784 *
## student27    70.56452   10.21965   6.905 1.15e-07 ***
## student28   -18.93548   10.21965  -1.853 0.073764 .
## student29    54.56452   10.21965   5.339 8.94e-06 ***
## student30   -28.93548   10.21965  -2.831 0.008198 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.69 on 30 degrees of freedom
## Multiple R-squared:  0.9272, Adjusted R-squared:  0.852
## F-statistic: 12.33 on 31 and 30 DF,  p-value: 3.375e-10
```

```
anova(eff.model2)
```

	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
chip	1	3080.145	3080.1452	14.27016	7.009264e-04
student	30	79411.742	2647.0581	12.26369	4.026186e-10
Residuals	30	6475.355	215.8452	*NA*	*NA*

3 rows

Step 5: Formulate conclusions.

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

Yes, we can see that we account for variation by student, we can see a significant difference between the melting times of the two chips.