Lesson 13 - Hurricanes!

CDT I.M.Smart

A study by Jung et al. (Proc. Natl. Acad. Sci. USA, 2014) looked at hurricane names and their perceived threat. They thought that if hurricanes had male names they would be perceived as more dangerous than if they had female names. In one part of the study, the participants were given a weather map showing a hurricane moving toward the area where they lived. Participants were told the name of the hurricane and then rated the riskiness of the hurricane on a seven-point scale, with 1 being "not at all" (risky) and 7 being "very risky."

```
### Loading packages
library(tidyverse)
library(ggResidpanel)
## Load data

hurrA <- read.table('https://raw.githubusercontent.com/jkstarling/MA376/main/hurricanesA.txt',
                    header = TRUE, stringsAsFactors = TRUE)
hurrB <- read.table('https://raw.githubusercontent.com/jkstarling/MA376/main/hurricanesB.txt',
                    header = TRUE, stringsAsFactors = TRUE)

# setwd('C:/Users/james.starling/OneDrive - West Point/Teaching/MA376/JimsLessons_AY23-1/Block II/LSN13
# hurrA <- read.table('hurricanesA.txt', header = TRUE, stringsAsFactors = TRUE)
# hurrB <- read.table('hurricanesB.txt', header = TRUE, stringsAsFactors = TRUE)

hurrA <- hurrA %>% mutate(Participant=as.factor(Participant))
hurrB <- hurrB %>% mutate(ParticipantID=as.factor(ParticipantID))
```

(1): Start by brainstorming how you might design this study. What variables would you measure? What variables would you control? What would you randomize? Why?

(2): The researchers decided to block on the participant. Explain what it means to "block on the participant" in this study. Why do you think they chose that blocking variable? How does that change your randomization plan?

Suppose the researchers had two participants in their study. Blocking on participant and randomizing to two treatments is a matched pairs design as discussed in Section 2.1, and a special case of the randomized complete block designs discussed in Section 2.2. Suppose they carried out the matched pairs design for these two participants and obtained the following ratings.

Participant	Hurricane Name	Rating from 1-7
1	Arthur (Male)	3.5
1	Bertha (Female)	4
2	Arthur (Male)	5
2	Bertha (Female)	4

(3): Based on this table, do male names tend to receive higher ratings than female names? Or, can you not judge such a main effect because there appears to be an interaction between participant and rating? How are you deciding?

We can't say too much from this study due to the very small sample size. In particular, one participant rated Hurricane Arthur as a larger threat, whereas the other participant picked Bertha as the larger threat. Why is this? Is this a trend? Or just random chance? Is there an interaction between Hurricane Name and Participant?

(4): A way to think of the limitation of this study is in terms of degrees of freedom. Can you fill in the table below? How many treatments are considered?

Source	DF
Hurricane Gender	
Participant	
Hurricane gender x Participant	
Error	
Total	

(5): How could you modify the study in order to test for a statistically significant interaction? Would adding more participants to the study help with this issue?

(6): If we didn't have replication, what type of models could we fit?

(7): Well-designed experiments use random assignment. Where should random assignment be used when conducting this study?

In this study, the researchers weren't that interested in specific hurricane names, but rather the hurricane gender (male or female). Therefore, although it probably wouldn't make sense to have the same participant rate the same name multiple times, they can rate the same hurricane name gender multiple times. These researchers had participants rate 10 different hurricane names selected from the official 2014 Atlantic Hurricane names, 5 of which were female names

(Bertha, Dolly, Fay, Laura, and Hanna) and 5 of which were male names (Arthur, Cristobal, Omar, Kyle, and Marco). Next, we have five repeats for each participant–gender combination, allowing us to estimate the "within participant" variation separately from the between participant and between treatment variation. When you have more than one observation for each block-treatment combination, this is often called a generalized block design.

The data in the `hurricanes`data set contain the perceived threat level rating given by two participants for each of the 10 hurricanes.

(8): Create an interaction plot using the `hurricanes` data set? What are the conclusions you can draw from this plot? Does there appear to be an interaction between name and participant?

```
# plot.data <- hurrA %>%
#               group_by(Participant, HurrGender) %>%
#               summarise(Mean = mean(Rating))
#
# plot.data %>% ggplot() + aes(x = xxx, y = xxx, group = xxx,
#                              color = xxx) +
#               geom_line() + geom_point() + theme_classic()
```

(9): Create an interaction model and anova table for `rating`, using `HurrGender` and `Participant` as explanatory variables? (Hint: pay attention to the type of variables that you have in the data set. Do any of the variables need to be changed to a factor?)

```
# contrasts(hurrA$Participant) = contr.sum
# contrasts(hurrA$HurrGender) = contr.sum
#
# interact.lm <- lm(xxx ~ xxx, data=hurrA)
# summary(interact.lm)
# anova(interact.lm)
```

(10): Is the interaction statistically significant?

In the actual study, the researchers used 346 students from the University of Illinois at Urbana-Champaign—participating for course credit (age 19-25 years, 208 females)—each of whom rated each of the 10 hurricane names (5 of each gender). The researchers examined several research questions but the hurricanesB file contains the results for the hurricane gender and hurricane age variables.

(Use the `hurrB` dataframe that was read in above.)

(11): Using the `hurricanesB` dataset create an interaction plot for `HurricaneGender` and `HurricaneAge`. Is there much evidence of a gender effect? Is there an age effect? An interaction? Describe the nature of the interaction that is present.

```
# plot.data = hurrB %>%
#               group_by(HurricaneGender, HurricaneAge) %>%
#               summarise(Mean = mean(Rating))
#
# plot.data %>% ggplot() + aes(x = xxx, y = xxx, group = xxx,
#                              color = xxx) +
#               geom_line() + geom_point() + theme_classic()
```

(12): Create a two-variable model and anova table, including the interaction between `HurricaneGender` and `HurricaneAge`, and adjusting for the blocking variable (`ParticipantID`). How much variation in ratings of hurricanes is explained in this model? Explain where each degree of freedom comes from.

```
# library(car)
#
# contrasts(hurrB$HurricaneGender) = contr.sum
# contrasts(hurrB$HurricaneAge) = contr.sum
# contrasts(hurrB$ParticipantID) = contr.sum

# factorial.lm <- hurrB %>% lm(xxx ~ xxx, data = .)
### summary(factorial.lm)  ## don't un-comment.
# factorial.anv <- Anova(factorial.lm, type = 3)
# factorial.anv
# rsq <- xxx
# rsq
```

(13): Examine the residual plot for this model. Do you consider the validity conditions for the theory-based F-tests to be met? Include supporting evidence for your conclusions.

```
# resid_panel(factorial.lm, plots = c("hist"))
# hurrB %>% group_by(HurricaneGender, HurricaneAge) %>%
#   summarise(sd(Rating))
```

(14): State the null and alternative hypotheses for the interaction. What do you conclude? Is this consistent with your interaction plot?

(15): Because the interaction is not statistically significant, we will summarize the main effects. Rerun the analysis without including the interaction.

State the null and alternative hypotheses for the interaction. What do you conclude? Is this consistent with your interaction plot?

```
# gender.lm <- hurrB %>% lm(xxx ~ xxx, data = .)

### summary(gender.lm)  ## don't un-comment.
```

```
# gender.anv <- Anova(gender.lm, type = 3)
# gender.anv
```

(16): Does this analysis support the researchers' conjecture that male names would be perceived as more dangerous? (Hint: After adjusting?) Write a short summary of your conclusions, being sure to address significance, estimation, causation, and generalizability.

(17): Is there evidence that blocking on participant was worthwhile? Explain how you are deciding and what this means in context.