Lesson 11: Corporate Credibility, Endorser, and Purchase Intent (§3.1)

LTC James K. Starling

Review

(-3) What types of design did we discuss in Chapter 2?

(-2) We saw from Ch. 1 that $SST = SSM + SSE$. How did this change in what we saw in Ch. 2? *What is the adjusted sum of squares*?

(-1) When would we want to block in a design? In other words, what assumption are we making when we use a block study design?

In this chapter we're going to dive more into some design structures starting with full factorial designs.

Background: Our book talks about a study on corporate credibility where researchers recruited 160 female college students and gave them a pamphlet containing background information (positive or negative) on a fictional shoe company. Each student was then shown a celebrity endorsement for one of the products and asked to rate her attitude towards the brand as well as her purchase intent. The researchers are interested in how corporate credibility affects purchase intent, and whether corporate credibility matters more than endorser credibility (or vice versa).

(1) Create a Sources of Variation diagram. What were the response variables and the explanatory variable(s)? How are these variables classified?

(2) Are any of these explanatory variables blocking variables? Why or why not.

(3) **Design 1:** The first design structure the researchers considered was to randomly assign 80 participants to the company with positive literature and Flo Jo and 80 participants to the company with negative literature and Roseanne. Is this a good design structure? Why or why not?

(4) **Design 2:** Then they thought 'what if we split our population into two groups. One group will focus on the corporate credibility (positive or negative credibility) and the other group will focus on the endorser (Roseanne or Flo Jo)'. Is this a good design structure?

(5) **Design 3:** Instead of manipulating the variables separately the researchers exposed a subset of the participants to every treatment. This is known as a *full factorial design.*

By the way, what's a treatment? What does the 'factor' in factorial refer to? How many factors do we have (what are they)? How many treatments do we have (what are they)?

(6) If we want a balanced full factorial design, how many of our 160 participants should we assign to each treatment? What are the main advantages to this type of study design?

(7) Can you foresee any potential problems with a full factorial design in general?

(8) Consider the graphs below. Explain what you see in the graphs.

```
cred %>% ggplot(aes(x=Treatments, y=PI, color=Treatments)) +
  geom_violin(trim=FALSE) +
  geom_boxplot(width=0.2) +
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.25, binwidth = 1) +
  stat_summary(fun=mean, geom="point", size=5, pch=4) +
  labs(y="purchase intent", x="endorser", title="Boxplot and Dotplot by endorser")
```

Boxplot and Dotplot by endorser
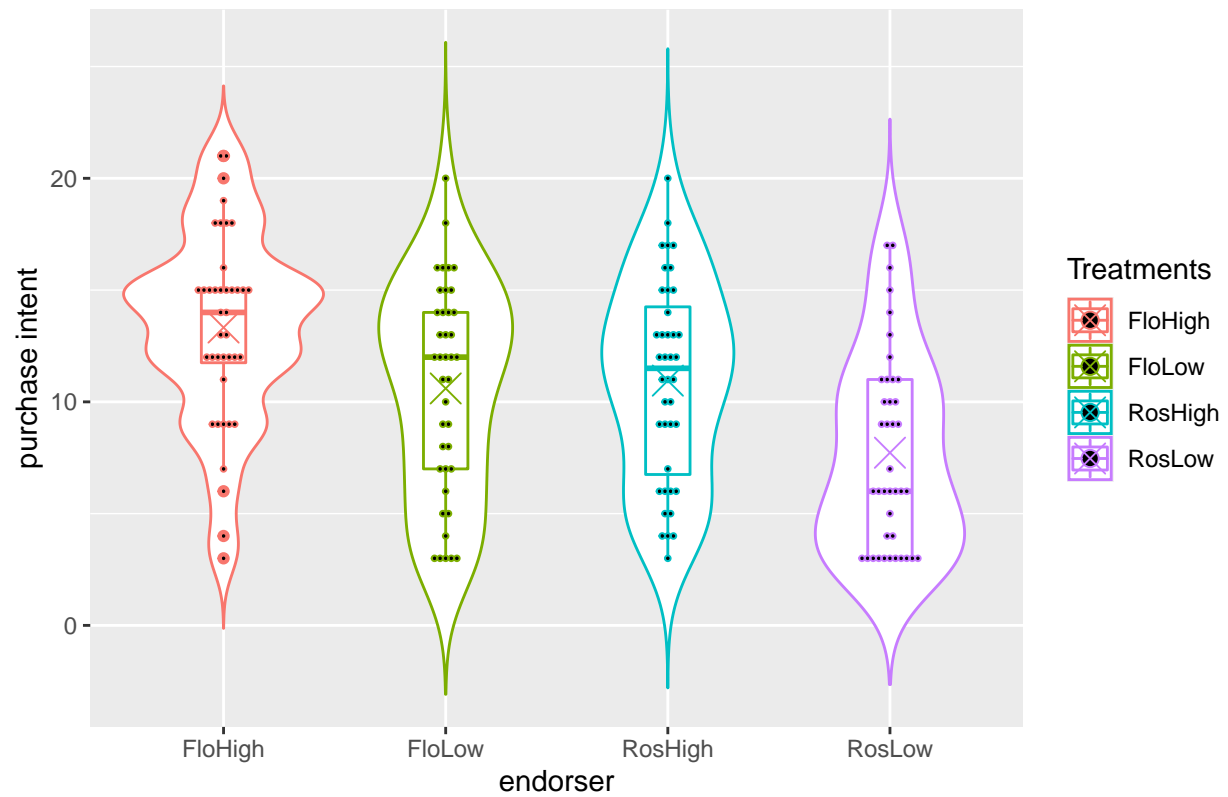

```
cred %>% ggplot(aes(x=Endorser, y=PI, color=Endorser)) +
  geom_violin(trim=FALSE) +
  geom_boxplot(width=0.2) +
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.25, binwidth = 1) +
  stat_summary(fun=mean, geom="point", size=5, pch=4) +
  labs(y="purchase intent", x="endorser", title="Boxplot and Dotplot by endorser")
```

Boxplot and Dotplot by endorser
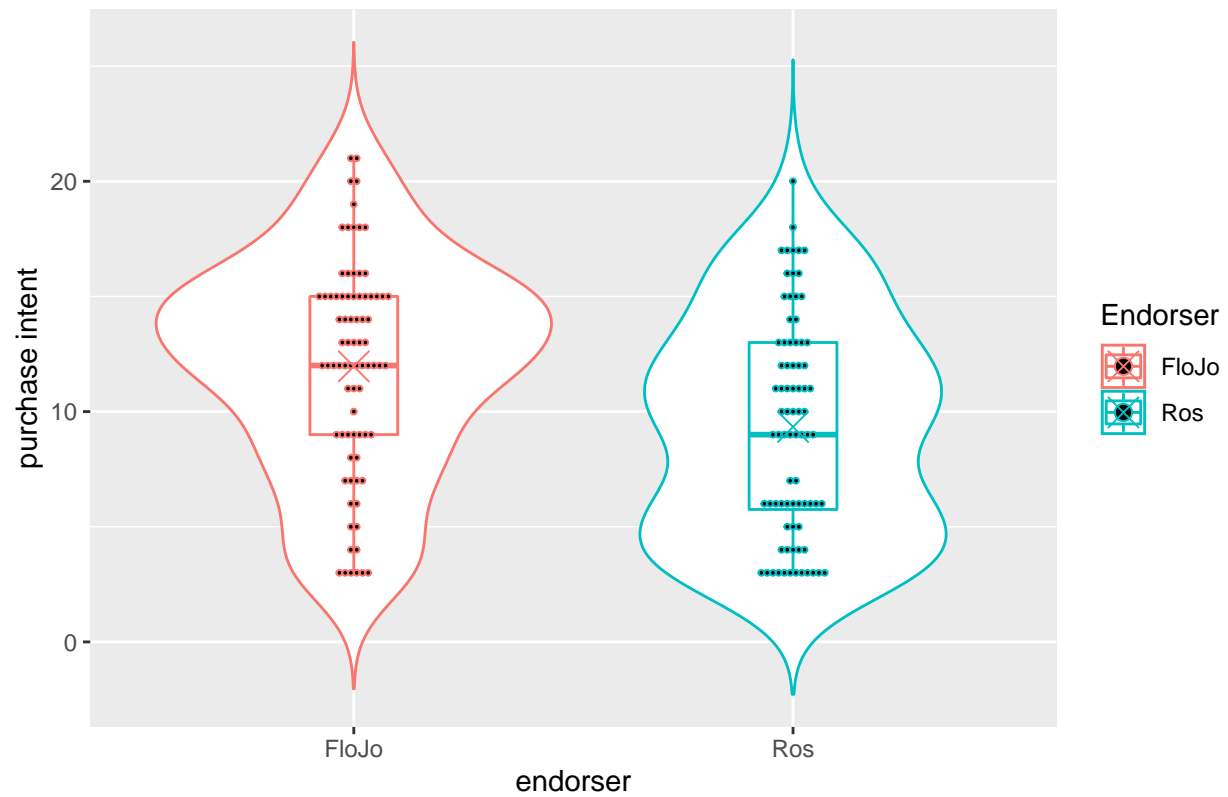

```
cred %>% ggplot(aes(x=CorpCred, y=PI, color=CorpCred)) +
  geom_violin() +
  geom_boxplot(width=0.2) +
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.25, binwidth = 1) +
  stat_summary(fun=mean, geom="point", size=5, pch=4) +
  labs(y="purchase intent", x="corporate credibility", title="Boxplot and Dotplot by corporate credibil
```

Boxplot and Dotplot by corporate credibility
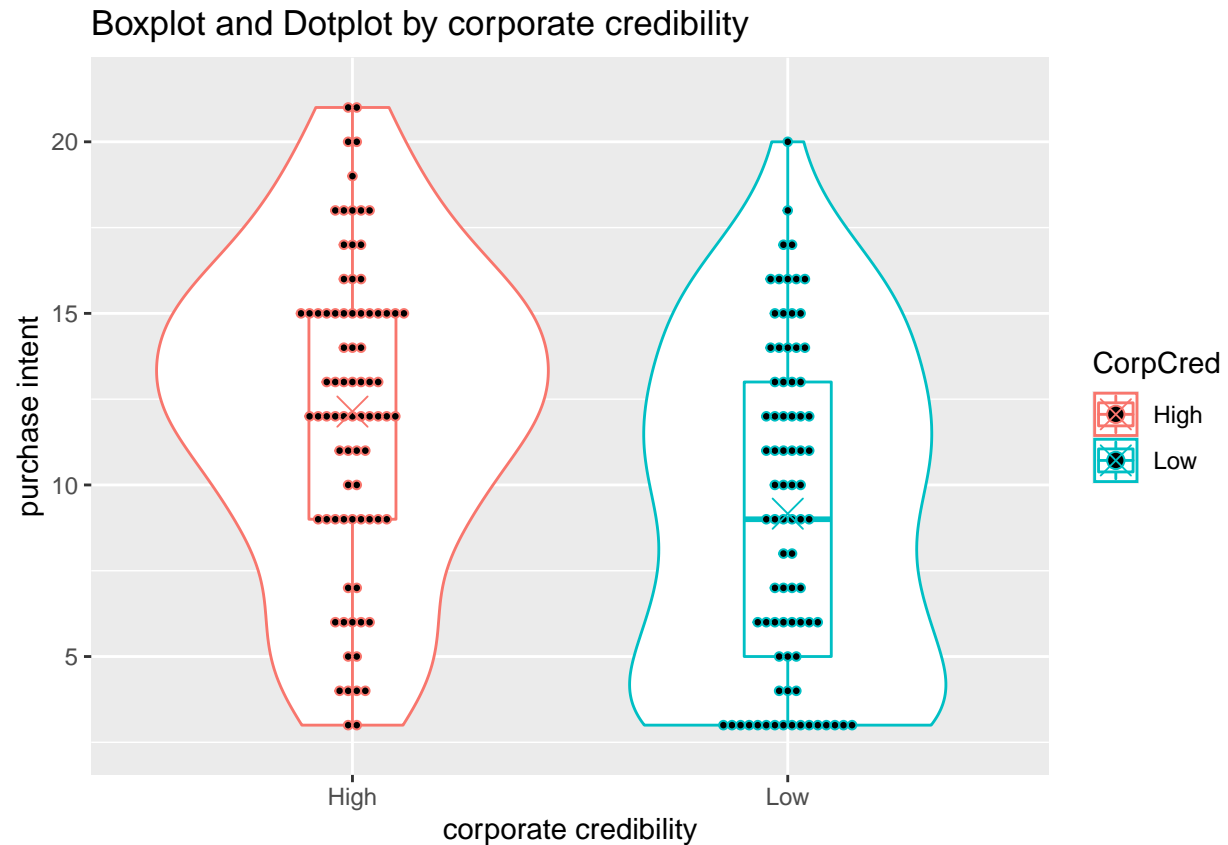
(10) What are the hypotheses for the one-variable analysis of the four treatment groups?

(11) What are the purchase intention means and standard deviations for each treatment? Are the differences in the means statistically significant? What is the size of the difference? To answer these questions let us first fit the the statistical model (multiple means model) and look at the ANOVA table.

```
# #calculate the purchase intention means by treatment
# cred %>% group_by(Treatments) %>% summarise(xxxxx)

# #One-way ANOVA (without constructing model)
# one_var.aov <- cred %>% aov(xxxx)
# summary(one_var.aov)
#
# rsq.1 <- xxx / xxx
# rsq.1
# mse1 <- xxx
# se1 <- sqrt(mse1)
# se1
```

(12) What is the conclusion?

(13) So? We still don't know where (all) the difference are. What can we use to figure out where the differences are? (Hint: think latin. Note: the differences will be a little different that the book).

(14) Is this what we want? Can we really say anything like "FloJo is better than Roseanne," or "High credibility is better than low credibility" from this analysis?

(15) Since the balanced study design has insured there is no confounding between credibility and endorser, we can form a combined prediction model (aka an additive model) that would allow us to make these inferences. Why is there no confounding between the two explanatory variables? (hint: see p. 218) How do we calculate the degrees of freedom for the ANOVA table?

```
# #Two-variable analysis of purchase intent by corporate credibility and endorser using effect coding
# cred2 <- cred
# contrasts(cred2$Endorser)=contr.sum
# contrasts(cred2$CorpCred)=contr.sum
#
# additive.lm <-lm(PI ~ Endorser + CorpCred, data=cred2)
# anova(additive.lm)
# rsq.2 <- xxx /xxx
# rsq.2
# mse2 <- xxx
# se2 <- sqrt(mse2)
# se2
```

(16) Using the summary output from the linear model, what is the predicted purchase intent for high corporate credibility with Roseanne as the endorser? What about high corporate credibility with Flo Jo?

(17) If we compare the two models how did we do in terms of explaining variation in the purchase intent?

(18) Whenever we do theory-based tests for statistical significance, we need to assess our validity conditions to ensure that our results are valid. What are the validity conditions?

```
# library(ggResidpanel)
# resid_panel(additive.lm, plots = 'hist', bins = 10)
```

(19) What are the main effects? Calculate an approximate 95% CI for the impact of each main effect (product endorser/corporate credibility) The 95% CI are calculated using the following: (difference in means) $\pm$ (multiplier) (SE residuals) $\times \sqrt{1/n_i + 1/n_j}$.

```
# credibility 95% CI:
```

```
# endorser 95% CI:
```