Bayesian Modeling Regression Pooling Homework

Business Scenario

Its 2021, and we Nordys outet opened store in the Bay Area. We're trying to forecast Q4 volume of Beauty Products *(which is a little tricky because 2020 was a very weird year and that's all the data we have).*

Data Prep and EDA

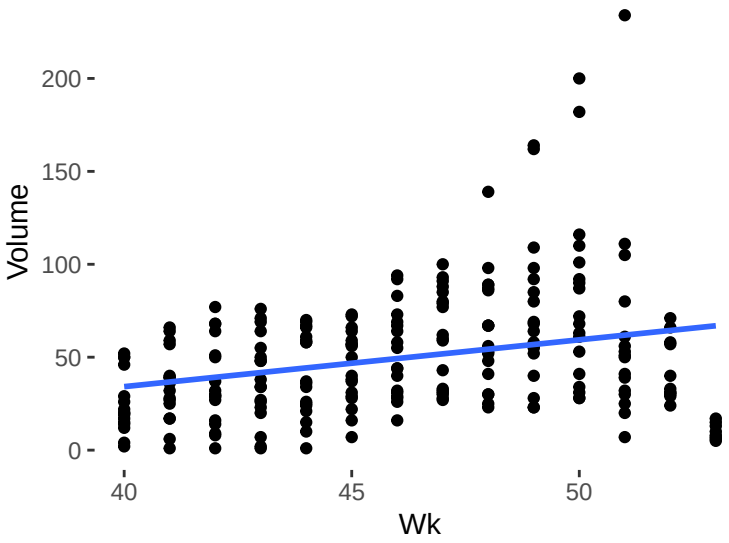
Load the following libraries, and tidy up the Sales data as shown below. We just want Q4 for San Jose and San Francisco, as the new store is in the Bay Area.

```
library(tidyverse)
library(rstan)
library(sn)
library(lubridate)
library(stringr)
library(kableExtra)
library(lme4)
```

```
set.seed(9)

SalesTrans <- read_csv("C:/Users/ellen/Documents/UH/Fall 2020/Class Materials/Section II/Class 1/Data/Sa
Location <- read_csv("C:/Users/ellen/Documents/UH/Fall 2020/Class Materials/Section II/Class 1/Data/Loca
MerGroup <- read_csv("C:/Users/ellen/Documents/UH/Fall 2020/Class Materials/Section II/Class 1/Data/MerG
SalesTrans = SalesTrans %>% inner_join(Location, by = "LocationID")
SalesTrans = SalesTrans %>% inner_join(MerGroup, by = "MerGroup")
LocationID <- as.factor(SalesTrans$LocationID)
SalesTrans$ProductID <- as.factor(SalesTrans$ProductID)
SalesTrans$Description <- as.factor(SalesTrans$Description)
SalesTrans$MerGroup <- as.factor(SalesTrans$MerGroup)
SalesTrans$Qtr <- quarter(SalesTrans$Tdate)
SalesTrans = filter(SalesTrans, Qtr == 4)

SalesTransSummary = SalesTrans %>%
  group_by(Description, Population, Income, MerGroup, MerGroup, Wk ) %>%
  summarise(Volume = n(), TotSales = sum(Amount) )
SalesTransSummary = filter(SalesTransSummary, Description %in% c("San Jose", "San Francisco"))

SalesTransSummary$Population = round(SalesTransSummary$Population/1000, 0)
SalesTransSummary$TotSales = round(SalesTransSummary$TotSales/1000, 0)
SalesTransSummary$MerGroupID <- as.integer(factor(SalesTransSummary$MerGroup))
```

Modeling

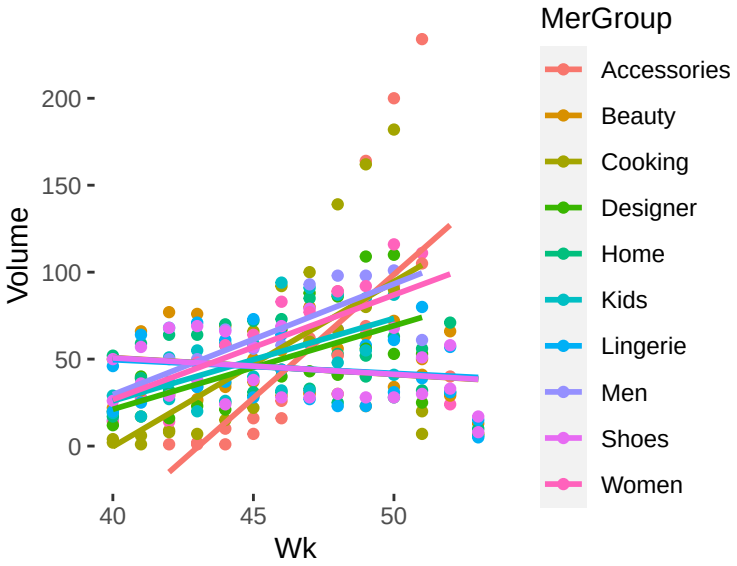
Now let's take a look at the overall trend for sales in Q4 for San Jose and San Francisco, noting the relationship between sales volume and weeks:



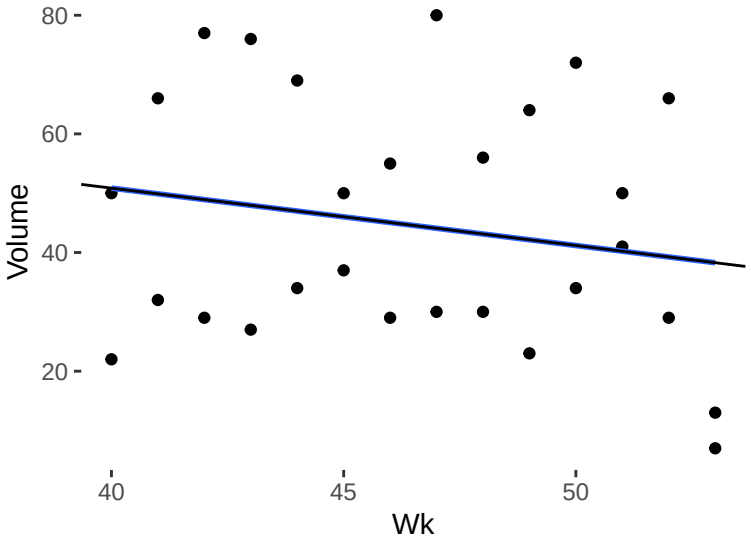
This is our fully pooled model, which gives us an overall trend. Show your coefficients:

	Intercept	Slope	MerGroup	Model
(Intercept)	-66.62	2.52	Beauty	lm-FP

Since we need to forecast Beauty Products for an outlet on the West Coast, lets first take a look at all the MerGroups using a no pooling model, to get a sense for the divergence of MerGroups, and trends, show no-pooled models *(just use ggplot with color = MerGroup to visualize)* :



Now, focusing on Beauty products only, create a separate beauty model *(this is what no-pooling does)*, and show the regression line: Notice that sales trend down at the end of the year:



Note: Management doesn't believe that will be the case this year. They think last year was unusual because women didn't buy beauty products during the pandemic in 2020 *(they didn't go into offices and social events)*. They think Beauty sales will follow the overall trend more.

Let's get the coefficents for Beauty Products - 2020:

Intercept	Slope	MerGroup	Model
-66.62	2.52	Beauty	lm-FP
89.54	-0.97	Beauty	lm-NP

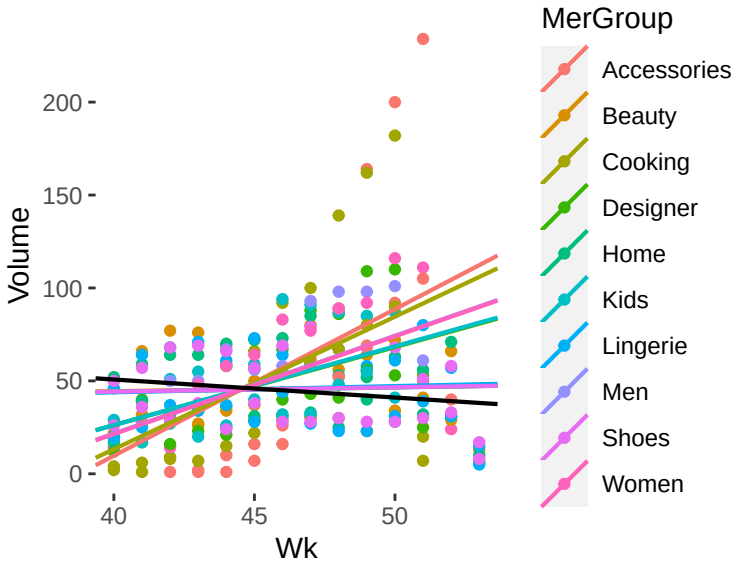
Now, use lmer to get a **Partially Pooled** model *(remember, lmer, using frequentist methods, weighs correlations accoring to a set formula)*:

The lmer coefficients should look something like *(you'll want to capture all the coefficients as you'll use these for priors in the first Bayesian model)*:

```
LmerCoefMap =  lmerCoef %>%
  rownames_to_column("MerGroup") %>%
  mutate(Model = "lmer") %>%
  rename(Intercept = `(Intercept)`, Slope = Wk)

LmerCoefMap$Intercept = round(LmerCoefMap$Intercept, 2)
LmerCoefMap$Slope = round(LmerCoefMap$Slope, 2)

knitr::kable(LmerCoefMap) %>%
  kable_styling(full_width = F, bootstrap_options = "striped", font_size = 9)
```

MerGroup	Intercept	Slope	Model
Accessories	-305.89	7.89	lmer
Beauty	35.63	0.22	lmer
Cooking	-271.97	7.13	lmer
Designer	-141.83	4.20	lmer
Home	34.50	0.24	lmer
Kids	-143.40	4.24	lmer
Lingerie	30.69	0.33	lmer
Men	-188.74	5.26	lmer
Shoes	35.03	0.23	lmer
Women	-188.99	5.26	lmer

Now, let's use CoefMap to visualize the MerGroup trends:



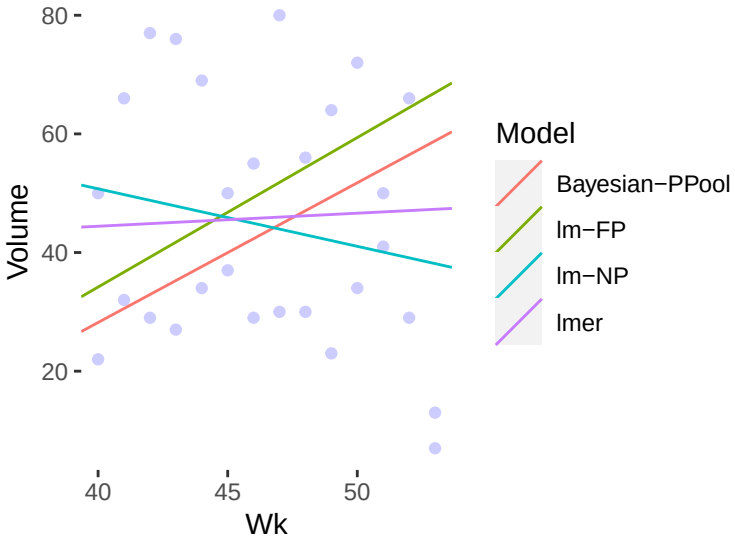
and updating our CoefMap data for Beauty Products:

Intercept	Slope	MerGroup	Model
-66.62	2.52	Beauty	lm-FP
89.54	-0.97	Beauty	lm-NP
35.63	0.22	Beauty	lmer

Bayesian Models (Pooling with Priors)

OK, now let's use Bayesian modeling to adjust priors to management judgement. First, run a model using the lmer parameters with tight priors.

> NOTE: after I stored the lmer priors in p_a and p_b vectors, I used asigma = rep(5,length(p_a)); bsigma = rep(2,length(p_b)). You can experiment with values and see how that affects your model.



Showing the CoefMap data:

Intercept	Slope	MerGroup	Model
-66.00	2.36	Beauty	Bayesian-PPool
-66.62	2.52	Beauty	lm-FP
89.54	-0.97	Beauty	lm-NP
35.63	0.22	Beauty	lmer