Seasonal Affective Disorder and Weather Patterns

S Smolenski

Introduction

Seasonal Affective Disorder is a type of episodic depression, wherein the person's mood is contingent upon local weather. Most commonly, this manifests in wetter climates, which receive high levels of precipitation in the cooler months: a personal with this traditional Seasonal Affective Disorder will become depressed during the colder months. This type of Seasonal Affective Disorder can be successfully treated with more antidepressants and ultraviolet lamps to simulate the sunshine that is so frequently absent during the relevant time of year. A second, less common, form of Seasonal Affective disorder exists during the opposite months; those who have this type become depressed during the hotter summer months.

I was diagnosed with Summer Seasonal Affective Disorder in the summer of 2015. Thereafter, I began recording my mood with the iMoodJournal android app to observe the patterns of depression and anxiety that I experience during the summer months.

This report summarizes my findings as I compare my self-recorded mood data, including information from my medication dosage, to local weather data.

Data

Acquisition

Climate data are downloaded from the US National Climate Data Center. I isolate only the weather from the station closest to my residence from September 1, 2015 onward. The data file, as downloaded from the NCDC website, can be found in the github repository for this project.

Mood data has been uploaded via iMoodJournal android app. The raw data, as exported by the app, are available in github repository.

For more detailed information on the datasets, please consult the respective codebook (available in the github repository).

Cleaning

I do not keep the entire NCDC file, but instead import only the data that are of interest to me: date, average temperature, recorded high, recorded low, inches of precipitation, and code, and a code indicating fog/rain/sleet/hail/thunder/tornado. I also drop the first four records, as these correspond to September 1-4, for which I have no mood data recorded.

```
widths<-c(-14, 8, -4, 4, -73, 5, -4, 4, -3, 4, -9, 6)
cnames<-c("Date", "AvgTemp", "High", "Low", "Prcp", "FRSHTT")
weather<-read.fwf("tempdata.txt",
                  widths=widths,
                  na.strings=c(999.9,9999.9),
                  skip=1,
                  col.names=cnames)
weather$Date<-ymd(weather$Date)
weather<-weather[5:1242, ]
```

From my mood data, I isolate the date, mood level (an integer from 1-10 indicating bad-good mood, respectively), the medication tag, and text comments:

```
read.csv("iMoodJournal.csv",
            stringsAsFactors=FALSE,
            fileEncoding="UTF-8-BOM") %>%
        select(c(1,5,214,7)) %>%
        rename(Wellbutrin=wellbutrin) -> moods
moods$Date<-as.Date(moods$Date,"%B %d, %Y")
```

Medication

During the summer of 2017 I began taking Bupropion (Wellbutrin), beginning with a dose of 100mg/day. From then until now, the dose has been varied throughout the cycle of the year, peaking at 300mg/day during the summer and dropping to 0 or 50 mg/day during the winter. I tagged start/stop/dose change days with #Wellbutrin in iMoodJournal, along with the dose. These tags are represented in the data by a 1 in the Wellbutrin column if #Wellbutrin was present in the comment for that day. Furthermore, I can access the dose by viewing the comment for a particular day.

More useful would be a column indicating what my dosage of Wellbutrin was on a given day. I accomplish this as follows:

```
meds<- which(moods$Wellbutrin==1,arr.ind=TRUE)
meddates<- as.Date(moods$Date[meds])
dose<-integer(0)
for (i in 1:length(meds)){
    index<-meds[i]
    str<- moods$Comment[index]
    dose<-as.numeric(c(dose,str_extract(str,"[[:digit:]]+")))
}
dose[2]<-0
for (i in 0:(length(meds))){
    if(i==0){
        WellbutrinDose<-rep(0,(meds[1]-1))
    }else if (i==length(meds)) {
        n<-(length(moods$Wellbutrin) - meds[i]+1)
        WellbutrinDose<- c(WellbutrinDose, rep(dose[i], n ))
    }else{
        n<- meds[i+1] - meds[i]
        WellbutrinDose<- c(WellbutrinDose, rep(dose[i], n ))
    }
}
```

I have manually included one stop date, where the comment on that date did not follow the pattern for the others (ie, comment reads “stopped #Wellbutrin” rather than "#Wellbutrin 0mg). Later dates have more consistent tagging schemes, so this is the only date I input manually.

Missing Data

During fall of 2017, I neglected to record the exact dates and values of my first attempt to taper off medication. As such, I fill in estimates of dates and doses. These filled in values have the correct doses, but the dates may be off anywhere on the order of days to weeks. For my purposes, this uncertainty is acceptable.

```
WellbutrinDose[1135:1189] <- 100
WellbutrinDose[1190:1243] <- 150
WellbutrinDose[1244:1430] <- 200
WellbutrinDose[1430:1484] <- 100
WellbutrinDose[1485:1541] <- 200
WellbutrinDose[1542:1562] <- 100
```

```
WellbutrinDose[1563:1574] <- 50
moods$Wellbutrin<-WellbutrinDose
```

Cleaned Data

The cleaned and combined data can be summarized as follows:

```
head(data)
```

```
##         Date Level High Wellbutrin Prcp
## 1 2015-09-05     7   64          0 0.19
## 2 2015-09-05     7   64          0 0.19
## 3 2015-09-05     7   64          0 0.19
## 4 2015-09-06     6   77          0 0.00
## 5 2015-09-06     8   77          0 0.00
## 6 2015-09-06     6   77          0 0.00
##
## 1
## 2
## 3                                                                                #calm #content #full #recumbe
## 4                                                                My #wristhurts :(\n#backinriverside\n#alone #h
## 5 #home with my #family, #coolweather !\n#content #awake #chipper #cheerful #cold #good #loved #peac
## 6
##   AvgTemp Low FRSHTT
## 1    45.7  28  10000
## 2    45.7  28  10000
## 3    45.7  28  10000
## 4    49.3  28      0
## 5    49.3  28      0
## 6    49.3  28      0
```

```
str(data)
```

```
## 'data.frame':    2492 obs. of  9 variables:
##  $ Date      : Date, format: "2015-09-05" "2015-09-05" ...
##  $ Level     : int  7 7 7 6 8 6 7 6 7 6 ...
##  $ High      : num  64 64 64 77 77 77 82.9 82.9 82.9 88 ...
##  $ Wellbutrin: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Prcp      : num  0.19 0.19 0.19 0 0 0 0 0 0 0 ...
##  $ Comment   : chr  "" "" "#calm #content #full #recumbent #relaxed #sleepy #relaxed " "My #wristhur
##  $ AvgTemp   : num  45.7 45.7 45.7 49.3 49.3 49.3 56.7 56.7 56.7 62 ...
##  $ Low       : num  28 28 28 28 28 28 28 28 28 36 ...
##  $ FRSHTT    : int  10000 10000 10000 0 0 0 0 0 0 0 ...
```

```
summary(data)
```

```
##       Date                Level           High          Wellbutrin
##  Min.   :2015-09-05   Min.   :2.0   Min.   : 10.00   Min.   :  0.00
##  1st Qu.:2016-06-08   1st Qu.:6.0   1st Qu.: 52.00   1st Qu.:  0.00
##  Median :2017-06-02   Median :6.0   Median : 66.00   Median :  0.00
##  Mean   :2017-05-15   Mean   :6.3   Mean   : 67.27   Mean   : 78.18
##  3rd Qu.:2018-04-21   3rd Qu.:7.0   3rd Qu.: 84.90   3rd Qu.:200.00
##  Max.   :2019-01-29   Max.   :9.0   Max.   :102.00   Max.   :300.00
##                       NA's   :58    NA's   :19       NA's   :58
##       Prcp            Comment             AvgTemp           Low
##  Min.   :0.00000   Length:2492        Min.   :-3.00   Min.   :-8.00
```

```
##  1st Qu.:0.00000   Class :character   1st Qu.:37.30   1st Qu.:23.00
##  Median :0.00000   Mode  :character   Median :48.00   Median :32.00
##  Mean   :0.03932                      Mean   :49.27   Mean   :31.58
##  3rd Qu.:0.01000                      3rd Qu.:62.70   3rd Qu.:39.90
##  Max.   :9.99000                      Max.   :81.60   Max.   :73.90
##  NA's   :10                           NA's   :10      NA's   :10
##      FRSHTT
##  Min.   :     0
##  1st Qu.:     0
##  Median :     0
##  Mean   : 11067
##  3rd Qu.: 10000
##  Max.   :111000
##  NA's   :10
```

Model Building

I hypothesize that mood level depends on temperature, precipitation, and medication dosage. I explore a few multivariable linear regression models in detail.

```
model1 <- lm(Level~High,data)
model2 <- lm(Level~High+Prcp,data)
model3 <- lm(Level~High+Wellbutrin+Prcp,data)
model4 <- lm(Level~High+Wellbutrin,data)
model5 <- lm(Level~Prcp+Wellbutrin,data)
model6 <- lm(Level~AvgTemp,data)
model7 <- lm(Level~AvgTemp+Prcp,data)
model8 <- lm(Level~AvgTemp+Wellbutrin+Prcp,data)
model9 <- lm(Level~AvgTemp+Wellbutrin,data)
model10 <- lm(Level~Low,data)
model11 <- lm(Level~Low+Prcp,data)
model12 <- lm(Level~Low+Wellbutrin+Prcp,data)
model13 <- lm(Level~Low+Wellbutrin,data)
```

I avoid using any of the temperature variables (High, AvgTemp, Low) in combination in any one model, since they are highly correlated variables and I want to avoid confounding and reduce error.

Taking a look at the models individually:

```
summary(model1)
```

```
##
## Call:
## lm(formula = Level ~ High, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3979 -0.5182  0.0895  0.6326  2.9662
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.326949   0.072443  101.14   <2e-16 ***
## High        -0.015231   0.001034  -14.73   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.9529 on 2413 degrees of freedom
##   (77 observations deleted due to missingness)
## Multiple R-squared:  0.08248,    Adjusted R-squared:  0.08209
## F-statistic: 216.9 on 1 and 2413 DF,  p-value: < 2.2e-16
```

```
summary(model2)
```

```
##
## Call:
## lm(formula = Level ~ High + Prcp, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3982 -0.5194  0.0895  0.6317  2.9659
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.329693   0.073407  99.851   <2e-16 ***
## High        -0.015260   0.001042 -14.646   <2e-16 ***
## Prcp        -0.019860   0.085164  -0.233    0.816
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9531 on 2412 degrees of freedom
##   (77 observations deleted due to missingness)
## Multiple R-squared:  0.0825, Adjusted R-squared:  0.08174
## F-statistic: 108.4 on 2 and 2412 DF,  p-value: < 2.2e-16
```

```
summary(model3)
```

```
##
## Call:
## lm(formula = Level ~ High + Wellbutrin + Prcp, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3018 -0.5508  0.1193  0.6305  3.1579
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.4784651  0.0755808  98.947  < 2e-16 ***
## High        -0.0192738  0.0011743 -16.413  < 2e-16 ***
## Wellbutrin   0.0015736  0.0002202   7.145 1.18e-12 ***
## Prcp        -0.0314215  0.0843090  -0.373    0.709
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9434 on 2411 degrees of freedom
##   (77 observations deleted due to missingness)
## Multiple R-squared:  0.1015, Adjusted R-squared:  0.1004
## F-statistic: 90.81 on 3 and 2411 DF,  p-value: < 2.2e-16
```

```
summary(model4)
```

```
##
```

```
## Call:
## lm(formula = Level ~ High + Wellbutrin, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3013 -0.5512  0.1197  0.6315  3.1581
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.4739770  0.0746020 100.185  < 2e-16 ***
## High        -0.0192236  0.0011664 -16.481  < 2e-16 ***
## Wellbutrin   0.0015720  0.0002201   7.141 1.22e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9432 on 2412 degrees of freedom
##   (77 observations deleted due to missingness)
## Multiple R-squared:  0.1015, Adjusted R-squared:  0.1007
## F-statistic: 136.2 on 2 and 2412 DF,  p-value: < 2.2e-16
```

```
summary(model5)
```

```
##
## Call:
## lm(formula = Level ~ Prcp + Wellbutrin, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3107 -0.3070 -0.2785  0.6930  2.7215
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.3069597  0.0260082 242.499   <2e-16 ***
## Prcp         0.1261260  0.0883006   1.428    0.153
## Wellbutrin  -0.0001425  0.0002036  -0.700    0.484
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9947 on 2421 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.001079,   Adjusted R-squared:  0.0002543
## F-statistic: 1.308 on 2 and 2421 DF,  p-value: 0.2705
```

```
summary(model6)
```

```
##
## Call:
## lm(formula = Level ~ AvgTemp, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3738 -0.5254  0.0922  0.6226  2.8781
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  7.190602   0.064881  110.83   <2e-16 ***
## AvgTemp     -0.017991   0.001252  -14.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9551 on 2422 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.07859,    Adjusted R-squared:  0.07821
## F-statistic: 206.6 on 1 and 2422 DF,  p-value: < 2.2e-16
```

```
summary(model7)
```

```
##
## Call:
## lm(formula = Level ~ AvgTemp + Prcp, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3735 -0.5312  0.0923  0.6217  2.8785
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.188792   0.065516 109.727   <2e-16 ***
## AvgTemp     -0.017968   0.001257 -14.291   <2e-16 ***
## Prcp         0.017100   0.085093   0.201    0.841
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9553 on 2421 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.0786, Adjusted R-squared:  0.07784
## F-statistic: 103.3 on 2 and 2421 DF,  p-value: < 2.2e-16
```

```
summary(model8)
```

```
##
## Call:
## lm(formula = Level ~ AvgTemp + Wellbutrin + Prcp, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2793 -0.5465  0.1012  0.6392  3.0313
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.2842628  0.0665704 109.422  < 2e-16 ***
## AvgTemp     -0.0221475  0.0014003 -15.816  < 2e-16 ***
## Wellbutrin   0.0014269  0.0002178   6.551 6.96e-11 ***
## Prcp         0.0164911  0.0843655   0.195    0.845
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9471 on 2420 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.09466,    Adjusted R-squared:  0.09354
```

```
## F-statistic: 84.34 on 3 and 2420 DF,  p-value: < 2.2e-16
```

```
summary(model9)
```

```
##
## Call:
## lm(formula = Level ~ AvgTemp + Wellbutrin, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2795 -0.5464  0.1008  0.6385  3.0309
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.2860114  0.0659535 110.472  < 2e-16 ***
## AvgTemp     -0.0221700  0.0013953 -15.889  < 2e-16 ***
## Wellbutrin   0.0014270  0.0002178   6.553 6.89e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9469 on 2421 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.09465,    Adjusted R-squared:  0.0939
## F-statistic: 126.5 on 2 and 2421 DF,  p-value: < 2.2e-16
```

```
summary(model10)
```

```
##
## Call:
## lm(formula = Level ~ Low, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2583 -0.4632 -0.0004  0.6839  2.7417
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.891438   0.057227   120.4   <2e-16 ***
## Low         -0.018621   0.001693   -11.0   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.971 on 2422 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.04755,    Adjusted R-squared:  0.04715
## F-statistic: 120.9 on 1 and 2422 DF,  p-value: < 2.2e-16
```

```
summary(model11)
```

```
##
## Call:
## lm(formula = Level ~ Low + Prcp, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2574 -0.4584 -0.0058  0.6717  2.7461
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.885949   0.057364 120.040   <2e-16 ***
## Low         -0.018589   0.001693 -10.978   <2e-16 ***
## Prcp         0.115357   0.086127   1.339    0.181
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9709 on 2421 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.04825,    Adjusted R-squared:  0.04747
## F-statistic: 61.37 on 2 and 2421 DF,  p-value: < 2.2e-16
```

```
summary(model12)
```

```
##
## Call:
## lm(formula = Level ~ Low + Wellbutrin + Prcp, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1921 -0.4806 -0.0182  0.6869  2.8117
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.9020693  0.0573994 120.246  < 2e-16 ***
## Low         -0.0209940  0.0018172 -11.553  < 2e-16 ***
## Wellbutrin   0.0007659  0.0002133   3.591 0.000336 ***
## Prcp         0.1272359  0.0859796   1.480 0.139047
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9685 on 2420 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.05329,    Adjusted R-squared:  0.05212
## F-statistic: 45.41 on 3 and 2420 DF,  p-value: < 2.2e-16
```

```
summary(model13)
```

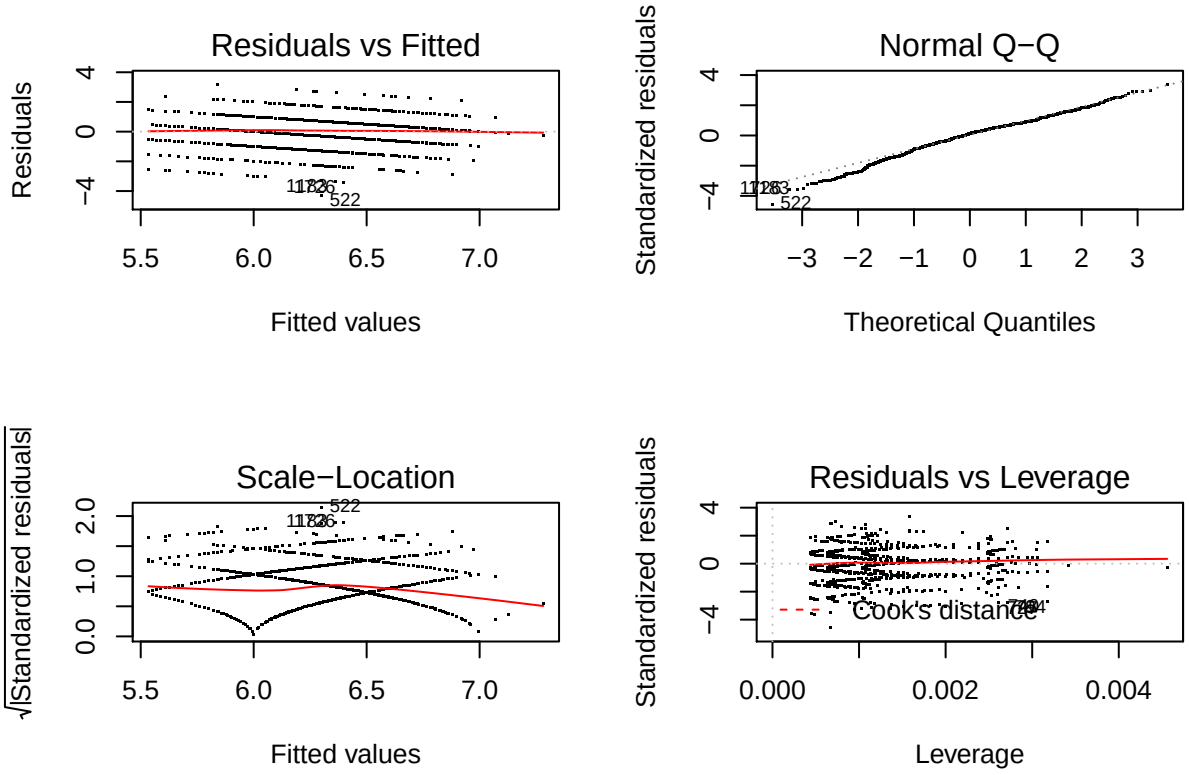
```
##
## Call:
## lm(formula = Level ~ Low + Wellbutrin, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1942 -0.4773 -0.0241  0.6847  2.8058
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.9078582  0.0572800 120.598  < 2e-16 ***
## Low         -0.0209910  0.0018176 -11.549  < 2e-16 ***
## Wellbutrin   0.0007537  0.0002132   3.535 0.000415 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.9687 on 2421 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.05244,    Adjusted R-squared:  0.05166
## F-statistic: 66.99 on 2 and 2421 DF,  p-value: < 2.2e-16
```

We note a strange trend that precipitation seems to be negatively correlated with mood, which is the opposite of my (subjective) personal experience. However, the p values are very high for Prcp coefficients in every model. This implies to me that the precipitation data is so scarce that a solid relationship cannot be drawn between mood and precipitation. For this reason, I will discard the models which include precipitation as a predictor, at least until more data can be gathered.

It seems clear, based on the low p and high t value for models including Wellbutrin, that this *is* an important predictor, and so we are left to determine which of the temperature variables is the strongest predictor of mood. Comparing models 4, 9, and 13, while High, Low, and AvgTemp all seem to predict mood fairly well, High has a marginally higher t value than either of the others. Thus, model 4 seems to be the best fit for the data.

I take a look at the residual plots to get a secondary measure of how good a fit model4 actually is for the data.



Here, the Q-Q plot looks satisfactory, but the residual plots clearly display some sort of pattern. This implies to me that, while the high temperature and medication dosage may be the best predictors for mood, the relationship is either non-linear in one or both variables, or there may be some other variable(s), which affect mood. I do expect a large variance in the data, but the patterns should not be present in the residual plots.

I consider two possible solutions for this below.

Non-Linear Regression

Missing predictor

This problem is less technically complicated, but more difficult to deal with. Of course, it will not be possible to account for the myriad variables that could possibly affect my mood. Some potential predictors I have discarded before this analysis: for instance, the iMoodJournal app displays *Mood vs Time of Day* and *Mood vs Day of Week* plots in-app. My data show no apparent correlation between mood and either of these variables.

One variable, which I have not accounted for here, but could be viable to include in the future is the hormone levels (namely, estrogen and progesterone) associated with the natural menstrual cycle. Premenstrual Dysphoric Disorder (PMDD) is a disorder which can cause depression and anxiety in the weeks preceding menstruation. While I have noticed an increase in depressive symptoms during the relevant weeks, this is merely a qualitative hypothesis as this point. I do, however, have enough data collected from manstrual cycles that I may be able to include hormone level as a predictor in a future version of this model.