Statistical Computations

Joseph S. Tabadero, Jr.

November 1, 2017

Import Datasets

```
library(tidyverse)
theme_set(theme_bw())
```

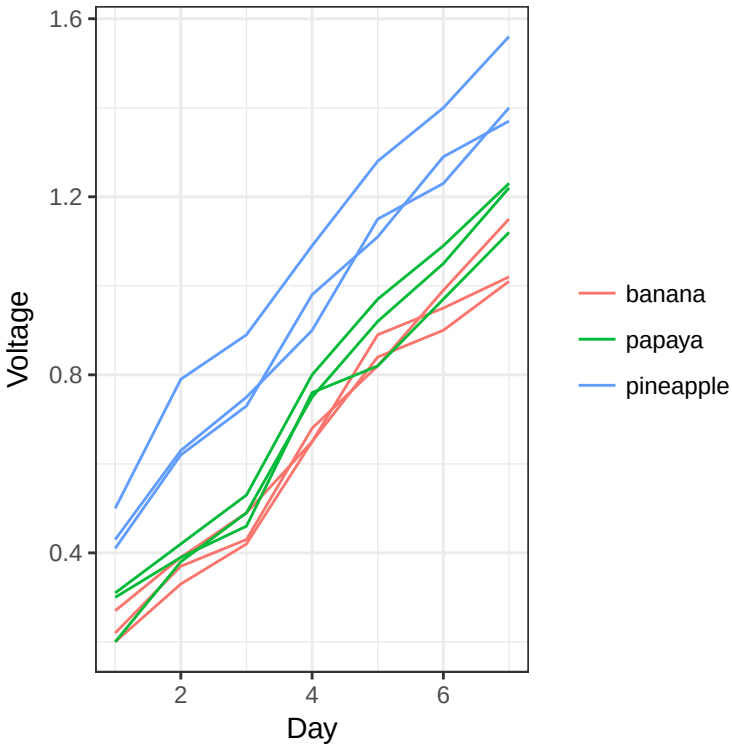
```
banana <- read.table("./data/banana.txt", header=TRUE, sep="")
papaya <- read.table("./data/papaya.txt", header=TRUE, sep="")
pineapple <- read.table("./data/pineapple.txt", header=TRUE, sep="")
to_long <- function(data){
    data %>% gather(., Cell, Voltage, -Day)
}
banana <- to_long(banana)
papaya <- to_long(papaya)
pineapple <- to_long(pineapple)
banana$fruit <- "banana"
papaya$fruit <- "papaya"
pineapple$fruit <- "pineapple"
data <- rbind(banana, papaya, pineapple)
data <- data %>%
  mutate(
    Day = as.integer(Day),
    Cell = as.integer(str_replace_all(Cell, "cell", "")),
    Voltage = as.numeric(str_replace_all(Voltage, "V", ""))
  )
```

Get mean by fruit and day

```
means <- data %>% group_by(Day, fruit) %>%
  summarise(Voltage = mean(Voltage))
```

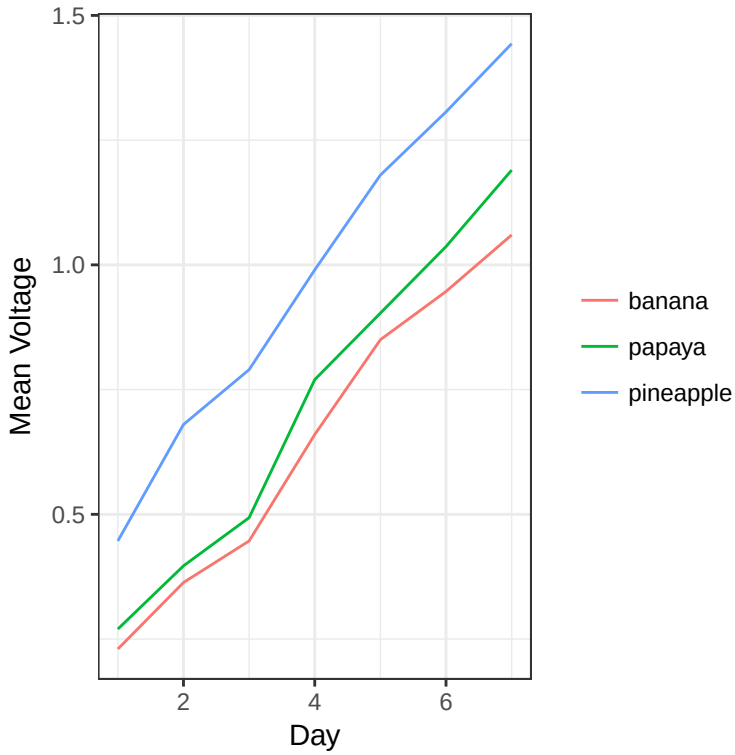
Get the plot of the voltages over 7 days for each type of fruit

```
ggplot(data, aes(Day, Voltage, group = as.factor(Cell), color = fruit)) + geom_line() + theme(legend.ti
```



Get the plot of the mean of the voltages over 7 days for each type of fruit

```
ggplot(means, aes(Day, Voltage, color = fruit)) + geom_line() + ylab("Mean Voltage") + theme(legend.tit
```



Analysis of variance for repeated measures design

Your data set contains the following:

- Cell = blocking factor (random factor, since the effect each cell may contribute to the voltage is random)
- Day = within subject or repeated measures factor (fixed factor)
- fruit = the treatment, which is a between subject factor (fixed factor)
- Voltage = measured (dependent) variable

A one-way analysis of variance will not work here, nor will a randomized complete block design ANOVA will work, because the assumption of independence of observations is violated. This assumption is violated because observations are taken from the same cells over time, and are therefore dependent on the cells.

To model the variation in the voltage outputs, we can use both fixed effects and random effects factors in the data set. The model is called a **linear mixed effects model**. You can learn more about linear mixed effects model from the following links:

- http://www.bodowinter.com/tutorial/bw_LME_tutorial2.pdf
- https://wiki.bcs.rochester.edu/HlpLab/StatsCourses?action=AttachFile&do=get&target=Groningen11.pdf
- https://arxiv.org/ftp/arxiv/papers/1308/1308.5499.pdf
- http://www.theanalysisfactor.com/repeated-measures-approaches/

Based on your study, what we want to model are the following: * There might be interactions between time (Day) and treatment (fruit), that is, we may allow the effect of Day to vary between type of fruit peeling (fruit) * We will allow change over time to differ across participants (i.e., to explicitly model differences in voltage changes among cells over time). Because of this, we will also allow the effect of Day to be *random*.

With these considerations, our model can be written as:

```
Voltage ~ Day + fruit + Day:fruit + (Day|Cell)
```

The analysis will be carried in R (R Core Team, 2017), using the `lmerTest` (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2016) package for modeling linear mixed effects. Post-hoc analysis will be conducted with the help of the R package `multcomp` (Hothorn, Bretz, & Westfall, 2008).

```
library(lmerTest)
mod <- lmer(Voltage ~ fruit*Day + (Day|Cell), data)
anova(mod)
```

```
## Analysis of Variance Table of type III  with  Satterthwaite
## approximation for degrees of freedom
##           Sum Sq Mean Sq NumDF DenDF F.value  Pr(>F)
## fruit      0.064   0.032     2     6      16  0.0037
## Day        2.608   2.608     1     6    1334 2.8e-08
## fruit:Day  0.008   0.004     2     6       2  0.2178
```

From the linear mixed effects model ANOVA table, we can see that the interaction of the type of fruit peeling and time does not affect the voltage output ($F(2, 6) = 2, p = 0.2178$). We can therefore interpret the effects of the type of fruit peelings and Day individually without worrying about their possible interactions. Here, we are primarily interested with the effect of type of fruit peelings. We can see that the voltage outputs differ between types of fruit peelings ($F(2, 6) = 16, p = 0.0037$) at the 0.05 significance level.

References

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). *LmerTest: Tests in linear*

mixed effects models. Retrieved from https://CRAN.R-project.org/package=lmerTest

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/