ST 518 - Homework 4

Nora Quick

```
library(Sleuth3)
library(tidyverse)
library(ggplot2)
library(arm)
library(vcdExtra)
library(magrittr)
```

R Question:

1. Analyze the data by treating the number of women out of 30 people on a venire as a binomial response (that is, you'll change the percent women in the datasheet to a count by multiplying by 30 and rounding) and judge as an explanatory variable .

(a) Is there evidence of over dispersion in these data? Please explain (i.e., don't just answer "yes" or "no").

```
dat <- data.frame(case0502)
dat$success <- round(dat$Percent*30/100)
dat$fail <- 30 - dat$success

glm(formula = cbind(success, fail) ~ Judge, family = binomial, data = dat)
```

```
##
## Call:  glm(formula = cbind(success, fail) ~ Judge, family = binomial,
##     data = dat)
##
## Coefficients:
##  (Intercept)        JudgeB        JudgeC        JudgeD        JudgeE
##     -0.66329       0.01974      -0.23749      -0.34831      -0.37691
##       JudgeF  JudgeSpock's
##     -0.32945      -1.08591
##
## Degrees of Freedom: 45 Total (i.e. Null);  39 Residual
## Null Deviance:       62.92
## Residual Deviance: 31.12     AIC: 208.5
```

No, becuase the DF isn't too large to indicate over dispersion. There is also not a large amount of variance within the values for each judge.

(b) Do the odds of a female on a venire differ for the different judges? Please explain.

```
dat$total <- dat$success + dat$fail

dat %<>% mutate(success = success, fail = fail, total = total, Judge = Judge, Percent = Percent, logPer

mod1 <- glm(data = dat, (success/total) ~ logPer, weights = total, family = "binomial")

mod2 <- glm(data = dat, (success/total) ~ logPer + logPer2, weights = total, family = "binomial")

LRstats(mod1, mod2)
```

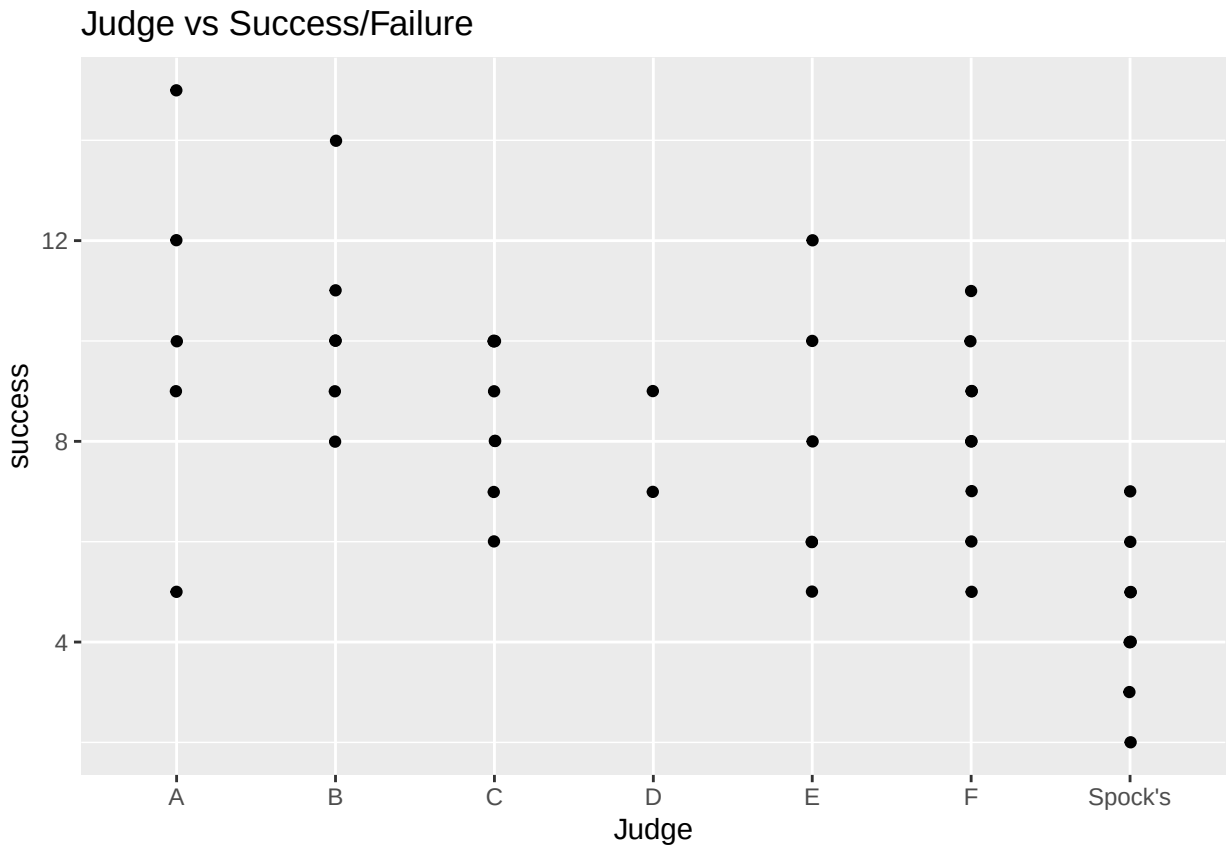
```
## Likelihood summary table:
##         AIC    BIC LR Chisq Df Pr(>Chisq)
## mod1 168.73 172.39  1.32236 44          1
## mod2 169.98 175.46  0.56982 43          1
```

```
anova(mod1, mod2,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: (success/total) ~ logPer
## Model 2: (success/total) ~ logPer + logPer2
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        44    1.32236
## 2        43    0.56982  1  0.75254   0.3857
```

```
ggplot(data = dat, aes(x = Judge, y = success))+
  geom_jitter(width = 0.005, height = 0.01) +
  geom_smooth(se = FALSE) +
  ggtitle("Judge vs Success/Failure")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

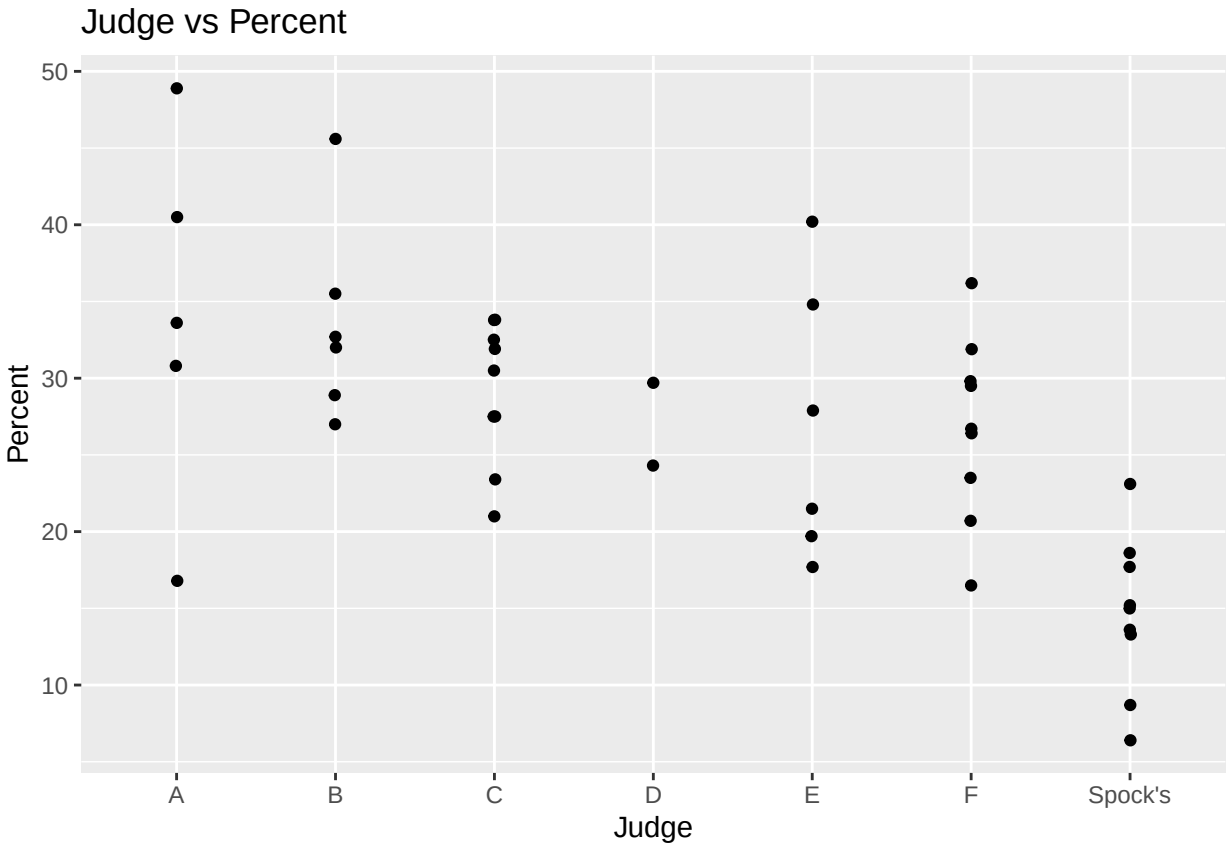


Yes, the odds of a female on a venire differ from judge to judge as we can see above. When looking specifically at success (not failure) there is a clear difference in the most accepting judges and the least accepting judges.

(c) Do judges A-F differ in their probabilities of selecting females to the venire? Please explain.

```
ggplot(data = dat, aes(x = Judge, y = Percent))+
  geom_jitter(width = 0.005, height = 0.01) +
  geom_smooth(se = FALSE) +
  ggtitle("Judge vs Percent")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



No, they do not differ in their probabilities. Some have large spreads while others have their probabilities clustered together but either way the percentages that judges A-F cover are the same span.

(d) How different are the odds of having a woman on the Spock judge's venires from the odds of having a woman on the venires of other judges? Please explain.

Looking at both graphs based on Percent and Success we can see that there is a difference in judges A-F and Spock's. We see A-F have a higher percentage and success rate than Spock. Overall, the odds are much lower for a woman to be picked with Spock as judge.

Conceptual Questions:

2. Give three explanations for why you may see evidence of extra-binomial variation in a logistic regression.

Sample size / variation between observations, an outlier, incorrect models.

3. In lecture, we saw that when we fit a logistic model when extra binomial variation is present, we get standard errors that are 'too small'. Explain why this gives misleading results.

We are overcompensating. We are fighting against something and it skews us into an area where the outcome has been fixed too much.

4. Consider an experiment that studies the number of insects that survive a certain dose of an insecticide, using several batches of insects of size n each. The insects may be sensitive to factors that vary among batches during the experiment but these factors (such as temperature) were unmeasured. Explain why the distribution of the number of insects per batch surviving the experiment might show over dispersion relative to a binomial(n, p)distribution.

There is not enough independence. The life of an animal (even an insect) is determined by many things including temperature. The temperature and effects of the insecticide are not idependent of each other and their effect.