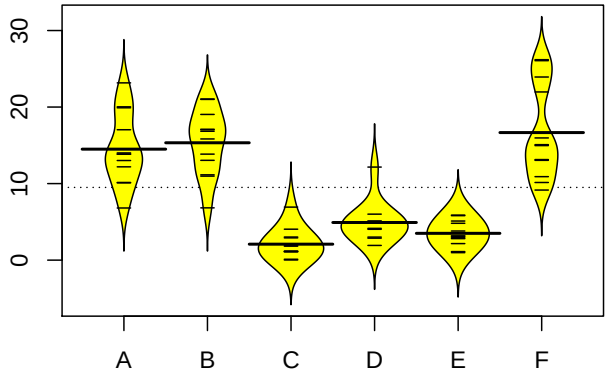
One-Way ANOVA Activity

Stat 217

In an agricultural experiment, 72 plots of land were randomly assigned to be treated with one of six different insecticides.The next day, the number of insects in each plot were counted. We will test to see if there is a difference in the number of insects for at least one of the sprays. Since there is a quantitative response variable and a single categorical explanatory variable, a One-Way ANOVA is an appropriate procedure for these data.

```
require(datasets)
require(ggplot2)
data(InsectSprays)
require(beanplot)
beanplot(count ~ spray, data = InsectSprays, method = "jitter", log = "",
    col = 7)
```




```
require(mosaic)
summary(InsectSprays)
```

```
##      count       spray
##  Min.   : 0.0   A:12
##  1st Qu.: 3.0   B:12
##  Median : 7.0   C:12
##  Mean   : 9.5   D:12
##  3rd Qu.:14.2   E:12
##  Max.   :26.0   F:12
```

1. Describe what you see in the beanplot

2. What is n_1? n_5? n_3?

3. What is $y_{2,1}$ approximately?

4. The Insect Sprays Hypotheses:

   Using cell means coding, what are the hypotheses?

   Using reference coding, what are the hypotheses?:

5. Is the model fit below a reference coded model or a cell means model?

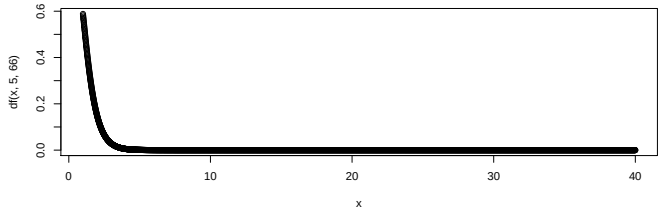
   Write out the reference coded model. Use correct notation.

```
lm1 <- lm(count ~ spray, data = InsectSprays)
anova(lm1)

## Analysis of Variance Table
##
## Response: count
##           Df Sum Sq Mean Sq F value Pr(>F)
## spray      5   2669     534    34.7 <2e-16
## Residuals 66   1015      15
```

6. Using the ANOVA table above, report the test statistic and the p-value.

7. Below is a picture of an F distribution with 5 and 66 degrees of freedom. Draw a vertical line at the test statistic on the plot below and shade in the area that is the p-value.



8. What distribution does the test statistic follow under the null hypothesis?

9. What is your decision (reject or fail to reject)?

10. Using the ANOVA table provided above, write a conclusion to your hypotheses(evidence sentence only).

11. Based on your conclusion, would you choose a single mean model or a separate means model? Why?

12. The model summary output is shown below.

```
summary(lm1)
```

```
##
## Call:
## lm(formula = count ~ spray, data = InsectSprays)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8.33  -1.96  -0.50   1.67   9.33
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   14.500      1.132   12.81  < 2e-16
## sprayB         0.833      1.601    0.52     0.60
## sprayC       -12.417      1.601   -7.76  7.3e-11
## sprayD        -9.583      1.601   -5.99  9.8e-08
## sprayE       -11.000      1.601   -6.87  2.8e-09
## sprayF         2.167      1.601    1.35     0.18
##
## Residual standard error: 3.92 on 66 degrees of freedom
## Multiple R-squared:  0.724,Adjusted R-squared:  0.704
## F-statistic: 34.7 on 5 and 66 DF,  p-value: <2e-16
```

(a) What is the reference group?

(b) What Greek letter does the sprayC row estimate?

(c) What is the average number of insects in the sprayF group?

(d) What is the average number of insects in the sprayA group? The sprayE group?

13. For this example, did we conduct a parametric test or a non-parametric test?