LAB 4: Anova

$H_o : \mu_A = \mu_B = \mu_C = \mu_D = \mu_E = \mu_F$

H_a : at least one of the mean is different from the others

11/2/20

ANOVA

- Analysis of variance (ANOVA)
  - SS, MS, df, F

ANOVA

Goal: **how large our treatment effect is.**

Variability in treatment effects = True Effect Differences + Error

Variability in residuals = Error

Hypothesis testing

H_0: True Effect Differences =0

H_A: The Effect differences are not all equal.

$$SS_{Treatments} = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$

$$SS_{Total} = SS_{Treatments} + SS_E$$

where n is the group size, and a is the number of treatments.

Degrees of freedom

$df\_Treatments = a - 1$

$df\_E = N - a$

N = total number of observations

a = number of treatments

Mean Squares (MS)

MST = SST/df_treatments

MSE = SSE/df_E

F-Ratio

The ratio of the Mean squares is called the f ratio and it is the test statistic for the null hypothesis

F-stat= MST/MSE

ANOVA Assumptions

1. The responses for each factor level have a normal population distribution.
2. These distributions have the same variance.
3. The data are independent.

R-code

p-value=pf(f-stat, df_T, df_E)

```r
summary(aov(chickwts$weight ~ chickwts$feed))
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## chickwts$feed   3 112125   37375   12.28  5e-06 ***
## Residuals      46 140008    3044
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Checking conditions: Shapiro-Wilk normality test

```r
shapiro.test(InsectSprays$count[InsectSprays$spray == "F"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  InsectSprays$count[InsectSprays$spray == "F"]
## W = 0.88475, p-value = 0.1009
```

Independence

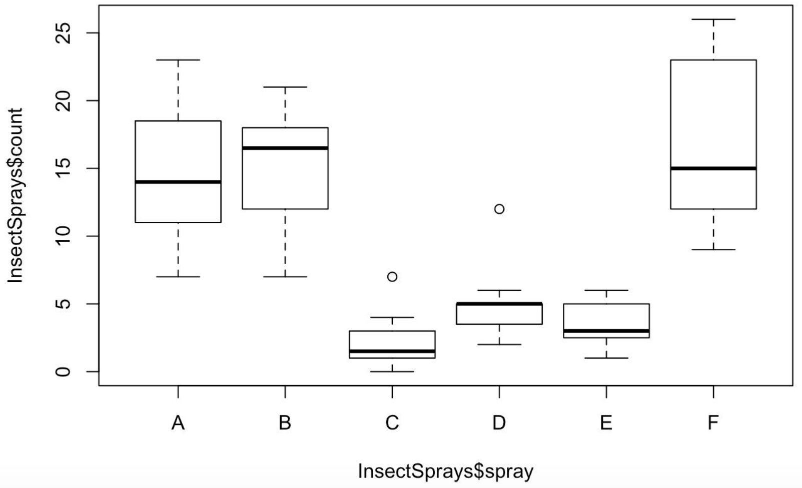
- From context of the problem
- “Randomly selected/assigned”

Use the following R code to read in a subset of the chickwts dataset in R. This data is for an experiment designed to measure the relative effectiveness of various feed supplements on chicken weights. Assume the **different feeds are randomly assigned** to the chickens. We are removing two of the feeds, horsebean and meatmeal, to simplify the problem slightly. To learn more about the full data, use ?chickwts.

Equal Variance

We can check this visually or by examining the standard deviations for each group.

Constant variance assumption is important when the sample sizes are not the same between groups.



Checking equal Variance

Another rule of thumb for equal variances is to compare the smallest and largest sample standard deviations.

$$1 < SD\_max/SD\_min < 2$$