

Analysis of Variance

Math 222 - Workshop

1. The following table contains summary statistics for the On Base Percentage of four groups of baseball players. Use the data to fill in the values in the ANOVA table below.

Position	n	\bar{x}	s
Catcher	39	0.323	0.045
Designated Hitter	14	0.348	0.036
In Field	154	0.332	0.037
Out Field	120	0.334	0.029
All Positions	327	0.332	0.036

Source	DF	Sum of Squares	Mean Square	F
Groups				
Error				
Total				

2. What is the R command to find the p-value of the F-statistic above? What is the result?

ANOVA is often used in medical research to simultaneously assess the effectiveness of several treatments for a given condition. One such study looked at different treatments for dandruff. The treatments were

- 1% pyrithione zinc shampoo (PyrI),
- 1% pyrithione zinc shampoo with instructions to shampoo two times (PyrII),
- 2% ketoconazole shampoo (Keto),
- A placebo shampoo (Placebo).

Subjects were randomized to each group, and initially each group had 112 volunteers, except the placebo group which was only assigned 28. Each volunteer was examined for dandruff flakes before and after six weeks of treatment. Dandruff flaking was measured on a scale from 0 to 80. Initially, there were no significant differences between the groups. During the clinical trial, 3 dropped out from the PyrII group and 6 from the Keto group. No patients dropped out of the other two groups.

3. Download the data and graph it using the following R commands:

```
dandruff <- read.csv("http://people.hsc.edu/faculty-staff/blins/classes/spring18/math222/data/dandruff.txt")
boxplot(flaking ~ treat, data = dandruff)
```

Which treatment seems the best/worst?

4. Use the following commands to get the ANOVA table and F statistic.

```
results <- aov(flaking ~ treat, data = dandruff)
summary(results)
```

What do the results mean?

5. Were the conditions for using ANOVA satisfied in this situation? To check constant variance, use:

```
aggregate(flaking ~ treat, data = dandruff, FUN = sd)
```