Agenda

1. More ANOVA
2. ANOVA lab

ANOVA Statisticians overload the term ANOVA. We use it when we are doing nested F-tests, when we are testing the significance of a categorical variable within a larger linear regression model, and we use it when we are just predicting a quantitative variable with a categorical one.

Just to get some intuition, lets compare the results from a few different approaches to using ANOVA.

```
require(mosaic)
bloodp = read.csv("http://www.math.smith.edu/~bbaumer/mth247/labs/bloodpress.csv")
m1 = lm(BP ~ Weight, data=bloodp)
anova(m1)

## Analysis of Variance Table
##
## Response: BP
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Weight     1 505.47  505.47  166.86 1.528e-10 ***
## Residuals 18  54.53    3.03
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

m2 = lm(BP ~ Age, data=bloodp)
anova(m2)

## Analysis of Variance Table
##
## Response: BP
##           Df Sum Sq Mean Sq F value   Pr(>F)
## Age        1 243.27 243.266  13.825 0.001574 **
## Residuals 18 316.73  17.596
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

m3 = aov(BP~ Weight + Age, data=bloodp)
summary(m3)

##             Df Sum Sq Mean Sq F value   Pr(>F)
## Weight       1  505.5   505.5  1781.3  < 2e-16 ***
## Age          1   49.7    49.7   175.2 2.22e-10 ***
## Residuals   17    4.8     0.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

m4 = aov(BP~ Age +  Weight, data=bloodp)
summary(m4)

##             Df Sum Sq Mean Sq F value   Pr(>F)
## Age          1 243.27  243.27   857.3 5.48e-16 ***
## Weight       1 311.91  311.91  1099.2  < 2e-16 ***
## Residuals   17   4.82    0.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
m5 = aov(BP~., data=bloodp)
summary(m5)

##             Df Sum Sq Mean Sq  F value   Pr(>F)
## Age          1 243.27  243.27 1466.914 9.38e-15 ***
## Weight       1 311.91  311.91 1880.844 1.89e-15 ***
## BSA          1   1.77    1.77   10.660  0.00615 **
## Dur          1   0.34    0.34    2.021  0.17871
## Pulse        1   0.12    0.12    0.742  0.40458
## Stress       1   0.44    0.44    2.666  0.12649
## Residuals   13   2.16    0.17
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

m6 = aov(BP~Stress+Pulse+Dur+BSA+Weight+Age, data=bloodp)
summary(m6)

##             Df Sum Sq Mean Sq  F value   Pr(>F)
## Stress       1  15.04   15.04   90.714 3.16e-07 ***
## Pulse        1 306.93  306.93 1850.831 2.09e-15 ***
## Dur          1   0.72    0.72    4.337   0.0576 .
## BSA          1 173.87  173.87 1048.444 8.15e-14 ***
## Weight       1  27.95   27.95  168.533 8.09e-09 ***
## Age          1  33.33   33.33  200.986 2.76e-09 ***
## Residuals   13   2.16    0.17
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA lab For your convenience, here is the code from the ANOVA lab

```
require(mosaic)
require(Stat2Data)

data(FruitFlies)
head(FruitFlies)

tally(~Treatment, data=FruitFlies)

# Set the reference level
FruitFlies = FruitFlies %>%
  mutate(Treatment = relevel(Treatment, ref="none"))

d1plot = dotPlot(~Longevity, data=FruitFlies)
d2plot = dotPlot(~Longevity | Treatment, data=FruitFlies, layout=c(1,5))
# arrange the two plots horizontally
print(d1plot, position=c(0, 0, 0.5, 1), more=TRUE)
print(d2plot, position=c(0.5, 0, 1, 1))

b1plot = bwplot(~Longevity, data=FruitFlies)
b2plot = bwplot(~Longevity | Treatment, data=FruitFlies, layout=c(1,5))
# arrange the two plots horizontally
print(b1plot, position=c(0, 0, 0.5, 1), more=TRUE)
print(b2plot, position=c(0.5, 0, 1, 1))
```

```
gstats = favstats(Longevity ~ Treatment, data=FruitFlies)
gstats

fm.null = lm(Longevity ~ 1, data=FruitFlies)
summary(fm.null)
mean(~Longevity, data=FruitFlies)
fitted = fitted.values(fm.null)
ladd(panel.abline(v=coef(fm.null), col="red", lwd=3), plot=d2plot)

fm.aov = aov(Longevity ~ Treatment, data=FruitFlies)
summary(fm.aov)
model.tables(fm.aov)

fm.ref = lm(Longevity ~ Treatment, data=FruitFlies)
summary(fm.ref)
anova(fm.ref)
mean(Longevity~Treatment, data = FruitFlies) - mean(~Longevity, data = FruitFlies)

dotPlot(~Longevity | Treatment, data=FruitFlies, layout=c(1,5), panel=function(x,...){
  panel.dotPlot(x, ...)
  fitted = mean(x)
  panel.abline(v=fitted, col="red")
})

par(mfrow=c(2,2))
plot(fm.aov)
par(mfrow=c(1,1))

sd(Longevity~Treatment, data=FruitFlies)
max(sd(Longevity~Treatment, data=FruitFlies)) / min(sd(Longevity~Treatment, data=FruitFlies))

data(Hawks)
```