

Lab 12

Analysis of Variance (ANOVA)

We have used the `anova()` function in R to run an F test to compare **models**. But we can also use F tests to simultaneously compare **group means**. This can be done with the `aov()` function in R (or `anova`, as before). If you look up `anova()` and `aov()` in the help section, the documentation shows that: * `anova()` takes **models** you have already run (like from running `lm`) as arguments, and outputs an ANOVA table regarding model fit.

- The `aov()` function takes a **formula** and **data** as arguments, and returns an ANOVA table regarding a comparison of **groups** (categorical variable from data). This is equivalent to running `lm()` with a categorical independent variable, but the summary is displayed in the traditional language of an ANOVA rather than a linear regression.

Let's look at some data on readings times. This data came from <http://users.stat.ufl.edu/~winner/data/ereader1.dat>. You can read a description of it here: <http://users.stat.ufl.edu/~winner/data/ereader1.txt>.

```
# Read in data directly from site
read <- read.table('http://users.stat.ufl.edu/~winner/data/ereader1.dat')

# Add in column names
colnames(read) <- c('device', 'light', 'read_time')
```

There are three device types, 4 levels of light, and a continuous variable measuring reading time in seconds. The data was being collected to assess reading fatigue (tiredness) with different devices at differing brightness levels. The longer someone could read, the lower the reading fatigue.

Let's say we want to know if there is an effect of device type on reading time. We will run a one-way ANOVA. What will our hypotheses be?

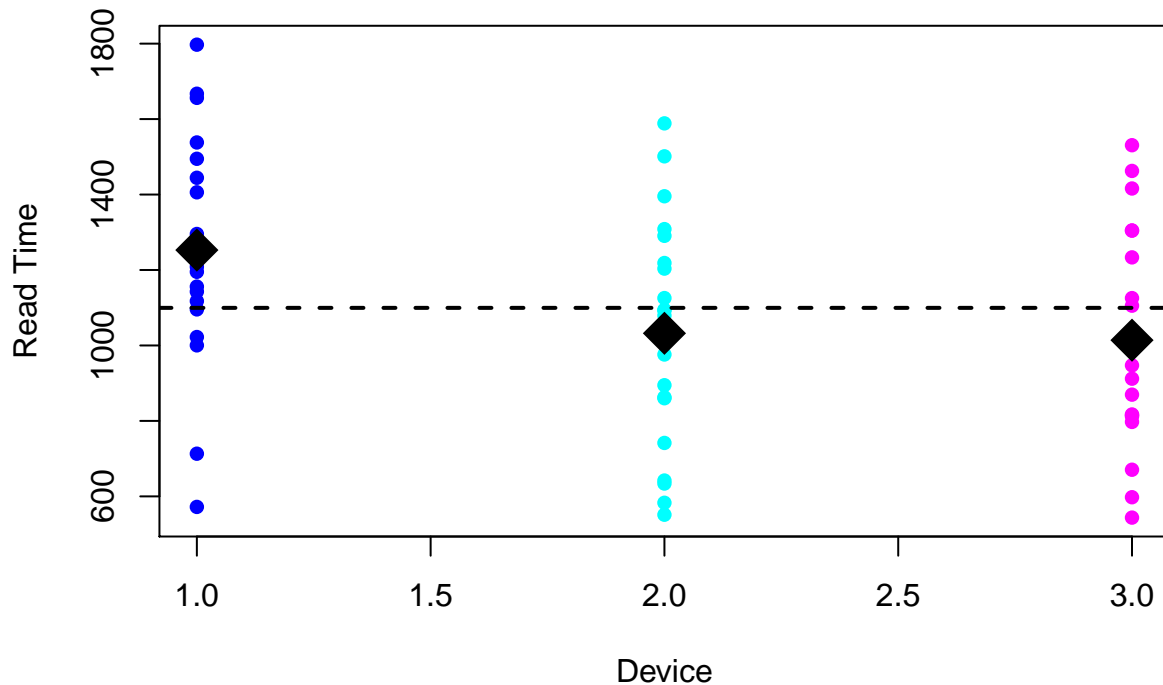
H0: Reading times are the same across different devices.

H1: Reading times are different across different devices (though we do not know where that difference (or differences) occurs).

Run the ANOVA and report the F statistic, degrees of freedom, and p-value. Determine if the null should be rejected at a significance of 0.05 and state your conclusion.

Start with a plot:

```
gpmeans = unlist(lapply(split(read$read_time, read$device), mean))
plot(read$device, read$read_time, col=read$device+3, pch=16, xlab="Device", ylab="Read Time")
points(1:3, gpmeans, pch=18, cex=3)
abline(h=mean(read$read_time), lwd=2, lty=2)
```



```
# Using aov
```

```
dev_model <- aov(read_time ~ factor(device), data = read)
summary(dev_model)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## factor(device) 2  706968   353484    3.91 0.0256 *
## Residuals      57 5152810    90400
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Using lm
```

```
dev_lm <- lm(read_time ~ factor(device), data = read)
summary(dev_lm)
```

```
##
## Call:
## lm(formula = read_time ~ factor(device), data = read)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -680.75 -197.34   -8.65   225.08   556.44
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1252.83     67.23  18.635  <2e-16 ***
## factor(device)2  -220.63     95.08  -2.320   0.0239 *
## factor(device)3  -238.83     95.08  -2.512   0.0149 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 300.7 on 57 degrees of freedom
## Multiple R-squared:  0.1206, Adjusted R-squared:  0.08979
## F-statistic:  3.91 on 2 and 57 DF,  p-value: 0.02562
```

The F statistic is 3.91 with degrees of freedom 2 and 57. The p-value is 0.0256. Since $p < 0.05$, we reject the null that reading times are the same across devices. We conclude that there is evidence that reading times are not the same across devices, though we do not know where those differences occur from this test.

Now let's check for an effect of brightness on reading time (regardless of device type). We will again use a one-way ANOVA, reporting F, the two values for degrees of freedom, and the p-value. And again we will write our hypotheses, whether we reject the null at $\alpha = 0.05$, and our conclusion.

H0: Reading times are the same across different levels of brightness. H1: Reading times are different across different levels of brightness.

```
light_model <- aov(read_time ~ factor(light), data = read)
summary(light_model)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## factor(light)  3 1481064  493688    6.314 0.000914 ***
## Residuals     56 4378714    78191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our F statistic is 6.314 and our degrees of freedom are 3 and 56. The p-value is 0.0009, which is less than our significance level of 0.05, so we reject the null that reading times are the same across different levels of brightness. We conclude that there is evidence that difference in reading times exist across levels of brightness, though we do not know where those differences are from this test alone.

If we want to find out where the differences are between levels of brightness, we can compare each of the four levels to another level in pairs. How many different pair-wise comparisons can we make for four groups?

This is the same as asking, “how many combinations of 2 can I make from the 4 levels?” We can use the `choose()` function to calculate this. To see the actual pairings, we can use the `combn()` function from the `utils` package (already available to you automatically).

```
# Number of combos
choose(n = 4, k = 2)
```

```
## [1] 6
```

```
# Actual pairings
combn(x = 4, m = 2)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    1    1    1    2    2    3
## [2,]    2    3    4    3    4    4
```

Now let's use t tests on our pairs, but we'll use a Bonferroni adjustment to our p-value. Remember, we make this adjustment because the more hypothesis tests we conduct, the more likely we are to obtain a significant result just by chance. In other words, the more tests we conduct, the higher our chances of a Type I error. So we “correct” for this by adjusting the alpha level to lower value (or equivalently, the p-values to a higher value), such that we are more conservative about when we will reject the null. Fortunately, there is a nice function in R that does all the work for us: `pairwise.t.test()`.

```
pairwise.t.test(x = read$read_time, g = read$light, p.adjust.method = 'bonferroni')
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data:  read$read_time and read$light
##
##      1      2      3
## 2 1.0000 -      -
```

```
## 3 0.0103 0.1936 -
## 4 0.0027 0.0658 1.0000
##
## P value adjustment method: bonferroni
```

This function output shows us the adjusted p-values for each test. Which pairs have significantly different reading times? Compare the adjusted p-values to the alpha level of 0.05.

Next, we might want to know if: -Controlling for device, are reading times different by light level?
-Controlling for light level, are reading times different by device?

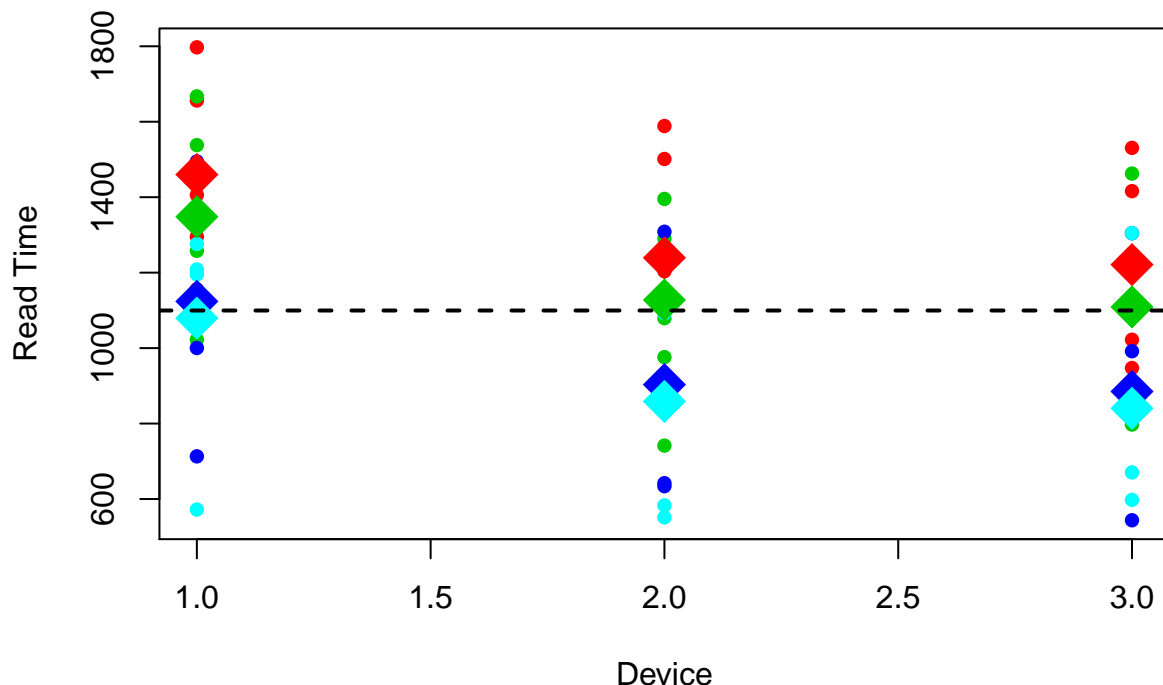
We can do this by adding both factors to our model.

Look at a graph first:

```
addmod = lm(read_time ~ factor(light) + factor(device), data=read)
gpmeans_d1 = c(addmod$coef[1], addmod$coef[1]+addmod$coef[2], addmod$coef[1]+addmod$coef[3],
               addmod$coef[1]+addmod$coef[4])
gpmeans_d2 = gpmeans_d1 + addmod$coef[5]
gpmeans_d3 = gpmeans_d1 + addmod$coef[6]
gpmeans = c(gpmeans_d1, gpmeans_d2, gpmeans_d3)

plot(read$device, read$read_time, col=read$light+1, pch=16, xlab="Device", ylab="Read Time")
points(rep(1,4), gpmeans[1:4], pch=18, cex=3, col=c(2:5))
points(rep(2,4), gpmeans[5:8], pch=18, cex=3, col=c(2:5))
points(rep(3,4), gpmeans[9:12], pch=18, cex=3, col=c(2:5))

abline(h=mean(read$read_time), lwd=2, lty=2)
```



Now model:

```
model_2way <- aov(read_time ~ factor(device) + factor(light), data = read)
summary(model_2way)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## factor(device) 2  706968   353484    5.199 0.008614 **
```

```
## factor(light)      3 1481064 493688 7.261 0.000353 ***
## Residuals         54 3671746 67995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#same thing

```
model_2way2 <- lm(read_time ~ factor(device) + factor(light), data = read)
anova(model_2way2)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: read_time
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## factor(device)  2  706968   353484   5.1987 0.0086140 **
## factor(light)   3 1481064   493688   7.2606 0.0003531 ***
## Residuals      54 3671746    67995
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#same thing

```
model <- lm(read_time ~ 1, data = read)
model1 <- lm(read_time ~ factor(device), data = read)
model2 <- lm(read_time ~ factor(device) + factor(light), data = read)
anova(model, model1, model2)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: read_time ~ 1
```

```
## Model 2: read_time ~ factor(device)
```

```
## Model 3: read_time ~ factor(device) + factor(light)
```

```
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      59 5859778
## 2      57 5152810  2    706968 5.1987 0.0086140 **
## 3      54 3671746  3   1481064 7.2606 0.0003531 ***
```

```
## ---
```

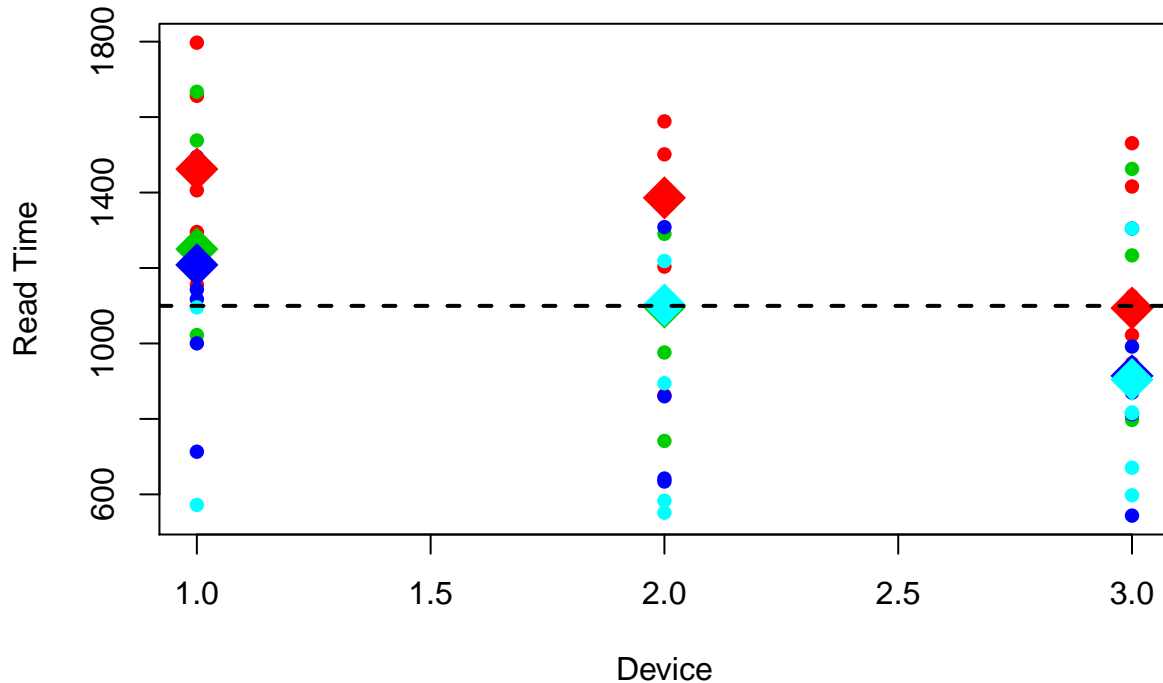
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

How could we use the aov() function to test if device type has different effects for subjects reading with differing levels of brightness? In other words, how do we test if the effect of device type varies by light level?

H0: The effect of device type on reading time does not vary by light level. H1: The effect of device type on reading time does vary by light level.

Let's look at how the picture changed first:

```
gpmeans = read %>%
  dplyr::group_by(device,light) %>%
  dplyr::summarise(means= mean(read_time))
means = gpmeans$means
plot(read$device, read$read_time, col=read$light+1, pch=16, xlab="Device", ylab="Read Time")
points(rep(c(1:4), 3), means, pch=18, cex=3, col=rep(c(1:4), each=3)+1)
abline(h=mean(read$read_time), lwd=2, lty=2)
```



```
int_model <- aov(read_time ~ factor(device) * factor(light), data = read)
summary(int_model)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## factor(device)    2  706968   353484   4.648 0.014279 *
## factor(light)     3 1481064   493688   6.492 0.000891 ***
## factor(device):factor(light) 6   21543    3591   0.047 0.999525
## Residuals        48 3650203   76046
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#same thing

```
int_model1 <- lm(read_time ~ factor(device) * factor(light), data = read)
anova(int_model1)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: read_time
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## factor(device)    2  706968   353484   4.6483 0.0142790 *
## factor(light)     3 1481064   493688   6.4920 0.0008906 ***
## factor(device):factor(light) 6   21543    3591   0.0472 0.9995253
## Residuals        48 3650203   76046
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We check if the interaction term is significant to determine whether we can reject the null. The F statistic is 0.047 and the p-value is 0.9995. At the 0.05 significance level, we fail to reject the null that device effect does not vary by light level. We conclude that we do not have evidence of device type effect varying by light level (the interaction was insignificant).