BMEG 802 – Advanced Biomedical Experimental Design and Analysis

One Way (Between) Analysis of Variance (ANOVA)

Joshua G. A. Cashaback, PhD

Recap

- Regression
  - Bivariate
    - Linear (Derivation)
    - Nonlinear
  - Multiple Regression
- Correlation
  - Pearson's r
  - Spearman's ρ

1 Way (Between) ANOVA

- develop logic & rationale for ANOVA (and formulas) based on a General Linear Model Approach
- any phenomenon is affected by multiple factors
- observed value on dependent variable (DV) = sum of effects of known factors + sum of effects of unknown factors
- similar idea to traditional approach of "accounting for variance" due to various factors
  - equivalent mathematically

General Linear Model (GLM)

Let's develop a model that expresses dependent variable (DV) as a sum of known and unknown factors

$$DV = C + F + R$$

- C = constant factors (known)
- F = factors systematically varied (known)
- R = randomly varying factors (unknown)

Notation looks like this:

$$Y_i = \beta_0 + \beta_1 \cdot X1_i + \beta_2 \cdot X2_i + \ldots + \beta_n \cdot Xn_i + \epsilon_i$$

Single Group Example

- a little artificial (who ever does experiments using just one group?)
- but it will help us develop the ideas
- imagine we collect scores on some DV for a group of subjects
- we want to compare the group mean to some known population mean
- e.g. IQ scores where by definition, $\mu = 100$ and $\sigma = 15$

Single Group Example

We know that:

$H_0 : \bar{Y} = \mu$

$H_1 : \bar{Y} \neq \mu$

Let's reformulate in terms of a GLM of the effects on DV:

$H_0 : \bar{Y} = \mu + \epsilon$; where $\mu = 100$

$H_1 : \bar{Y} = \hat{\mu} + \epsilon$; where $\hat{\mu} = Y_i$

Terminology:

H_0 is the "Restricted Model" — no parameters need to be estimated

H_0 is the "Full Model" — we need to estimate one parameter (can you see what it is?)

Computing Model Error

- how well do these two models fit our data?
- let's use the **sum of squared deviations** of our model from the data, as a measure of goodness of fit

$$H_0 : \sum_{i=1}^{N}(e_i^2) = \sum_{i=1}^{N}(Y_i - 100)^2$$

$$H_1 : \sum_{i=1}^{N}(e_i^2) = \sum_{i=1}^{N}(Y_i - \hat{\mu})^2 = \sum_{i=1}^{N}(Y_i - \hat{Y})^2$$

- SSE about the sample mean is lower than SSE about any other number
- so the error for H_0 will be greater than for H_1
- so the relevant question then is, how much greater must H_0 error be,for us to reject H_0?

Computing Model Error

- Consider the Proportional Increase in Error (PIE)
  - $(E_{restricted} - E_{full})/E_{full}$
- PIE gives error increase for H_0 (restricted) compared to H_1 as (full) a % of H_1 error
- We want a model that is both
  - adequate (low error)
  - simple (few parameters to estimate)
- question: why do we want a simpler model?
  - philosophical reasons (Occam's razer)
  - statistical reasons (over-fitting)

Computing Model Error

- how big is increase in error with H_0 (restricted model), per unit of simplicity?
- let's design a test statistic that takes into account simplicity
- simplicity will be related to the number of parameters we have to estimate
- degrees of freedom (df): # of independent observations in the dataset minus # of independent parameters that need to be estimated
- higher df = a simpler model

Computing Model Error

Let's normalize model errors (PIE) by model df - This is called the F-statistic!

$$F = \frac{(E_{restricted} - E_{full})/(df_{restricted} - df_{full})}{(E_{full}/df_{full})}$$

We can compute F_{obs} and calculate the probability of obtaining that F-statistic by using the F-distribution!

Two Group Example

Let's look at a more realistic situation

- 2 groups, 10 subjects in each group
  - test mean of group 1 vs mean of group 2
  - do we accept H_0 or H_1?
- we will formulate this question as before in terms of 2 linear models
  - full vs restricted model
  - is the error for the restricted model significantly higher than for the full model?
  - is the decrease in error for the full model large enough to justify the need to estimate a greater # parameters?

Two Group Example: Hypotheses and Models

$H_0 : \mu = \mu_1 = \mu_2$

restricted model: $Y_{ij} = \mu + \epsilon_{ij}$

$H_1 : \mu_1 \neq \mu_2$

full model: $Y_{ij} = \mu_j + \epsilon_{ij}$

i = individual (1, 2, . . . , 10)

j = group (1 or 2)

Restricted model:

- each score Y_{ij} is the result of a single population mean plus random error ϵ_{ij}

Full model:

- each score Y_{ij} is the result of a different group mean plus random error ϵ_{ij}

Deciding between full and restricted model

- how do we decide between these two competing accounts of the data?

key question

- will a restricted model with fewer parameters be a significantly less adequate representation of the data than a full model with a parameter for each group?
- we have a trade-off between simplicity (fewer parameters) and adequacy (ability to accurately represent the data)

Error and DF for the Restricted Model

- Sum of squared deviations of each observation from the estimate of the population mean (given by the grand mean of all of the data)
- Here we need to estimate 1 parameter, $\hat{\mu}$

$$E_{restricted} = \sum_{j=1}^{n_j} \sum_{i=1}^{n_i} (Y_{ij} - \hat{\mu})^2$$

$$\hat{\mu} = \frac{1}{n_i \cdot n_j} \sum_{j=1}^{n_j} \sum_{i=1}^{n_i} (Y_{ij})$$

$$df_{restricted} = N - 1$$

(N = number of participants, 1 is the number of estimated means in the restricted model)

Error and DF for the Full Model

- Here we need to estimate 2 parameter (one for each group), $\hat{\mu_1}$ and $\hat{\mu_2}$

$$E_{full} = \sum_{j=1}^{n_j} \sum_{i=1}^{n_i} (Y_{ij} - \hat{\mu_j})^2 = \sum_{i=1}^{n_i} (Y_{i1} - \hat{\mu_1})^2 + \sum_{i=1}^{n_i} (Y_{i2} - \hat{\mu_2})^2$$

$$\hat{\mu_1} = \frac{1}{n_1} \sum_{i=1}^{n_i} (Y_{i1})$$

$$\hat{\mu_2} = \frac{1}{n_2} \sum_{i=1}^{n_i} (Y_{i2})$$

$$df_{full} = N - a = N - 2$$

(N = number of participants, a = number of estimated means in the full model)

Deciding between Full and Restricted Model

Now we formulate our measure of proportional increase in error (PIE) as before using the F-statistic:

$$F = \frac{(E_{restricted} - E_{full})/(df_{restricted} - df_{full})}{(E_{full}/df_{full})}$$

Defining Error and DF For the General Case

$$E_{full} = \sum_{j=1}^{a} \sum_{i=1}^{n} (Y_{ij} - \bar{Y}_j)^2$$

$$df_{full=N-a}$$

$$E_{restricted} - E_{full} = n \sum_{j=1}^{a} (\bar{Y}_j - \bar{Y})^2$$

$$df_{restricted} - df_{full} = a - 1$$

Again, we formulate our measure of proportional increase in error (PIE) as before using the F-statistic:

$$F = \frac{(E_{restricted} - E_{full})/(df_{restricted} - df_{full})}{(E_{full}/df_{full})}$$

Model Comparison Approach vs Traditional Approach to ANOVA

Traditional formulation of ANOVA asks the same question in a different way:

- is the variability between groups (variance due to differences between groups) greater than expected on the basis of the within-group variability (the variability within a group) observed, and random sampling of group members?

Maxwell, Delaney, Kelley, Chapter 3: Proof that these two approaches are mathematically equivelant

- both use sum of squares
- both use F-statistic
- $F = \frac{MS_{between}}{MS_{within}}$ (Same Mean Squared Error on ANOVA table outputs)

Assumptions of the F test

1. the scores on the dependent variable Y are normally distributed in the population (and normally distributed within each group)
2. the population variances of scores on Y are equal for all groups
3. scores are independent of one another

Violation of Assumption

- ANOVA is somewhat robust to violations of normality
- ANOVA is somewhat robust to homogeneity of variance
- ANOVA is NOT robust to violations of independence

Three Group Example in R

- We can use ANOVA / GLM for 1 sample and 2 sample tests.
  - But, equivalent to using a t-test
- ANOVA are most widely used when there is greater than 2 groups. - Acts as the "omnibus" test to decide whether you have "permission" to perform follow-up mean comparisons.
- Lets perform an ANOVA using R for the following simple example. We want to know if the following three groups are different from one another.

Three Group Example in R

Let's determine if there is a main effect of group.

Group 1	Group 2	Group 3
4	7	6
5	4	9
2	6	8
1	3	5
3	5	7
mean $= 3$	mean $= 5$	mean $= 7$

$H_0 : \mu_1 = \mu_2 = \mu_3$

restricted model: $Y_{ij} = \mu + \epsilon_{ij}$

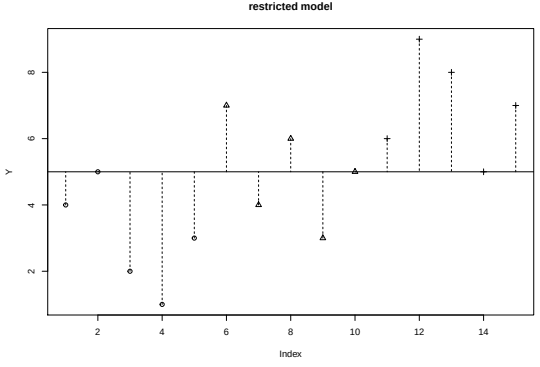
$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$

full model: $Y_{ij} = \mu_j + \epsilon_{ij}$

Three Group Example in R

Plot Restricted Model: Single parameter, μ

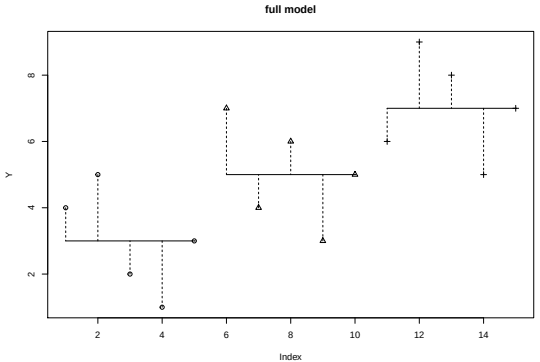
```
Y <- c(4,5,2,1,3,7,4,6,3,5,6,9,8,5,7)
myFac <- c(1,1,1,1,1,2,2,2,2,3,3,3,3,3)
plot(Y, pch=myFac,main="restricted model")
abline(h=mean(Y))
for (i in 1:length(Y)) {
  lines(c(i,i), c(Y[i], mean(Y)), lty=2)}
```



Three Group Example in R

Plot Full Model: three parameters, μ_1, μ_2, μ_3

```
Y <- c(4,5,2,1,3,7,4,6,3,5,6,9,8,5,7)
IVO <- c(1,1,1,1,1,2,2,2,2,3,3,3,3,3)
plot(Y, pch=IVO, main="full model")
for (j in 1:3) {
  w <- which(IVO==j)
  lines(c(min(w),max(w)),c(mean(Y[w]),mean(Y[w])))
  for (i in 1:length(w)) {
    lines(c(w[i],w[i]), c(Y[w[i]], mean(Y[w])), lty=2) }}
```



Three Group Example in R

Using the aov() in R

```
m1 <- aov(Y ~ factor(IV0))
summary(m1)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## factor(IV0)  2     40    20.0       8 0.0062 **
## Residuals   12     30     2.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here we find that there is statistically significant main effect of group ($p = 0.006$)!

Three Group Example in R

Alternatively we can build a linear model, using lm(), and then pass it through aov()

```
m2 <- lm(Y ~ factor(IV0))
summary(m2)
```

```
##
## Call:
## lm(formula = Y ~ factor(IV0))
##
## Residuals:
##    Min     1Q Median     3Q    Max
##     -2     -1      0      1      2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.0000     0.7071   4.243  0.00114 **
## factor(IV0)2   2.0000     1.0000   2.000  0.06866 .
## factor(IV0)3   4.0000     1.0000   4.000  0.00176 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.581 on 12 degrees of freedom
## Multiple R-squared:  0.5714, Adjusted R-squared:    0.5
## F-statistic:     8 on 2 and 12 DF,  p-value: 0.006196
```

In this case the estimate for the first group (called Intercept in the anova output) is 3.0000. The estimate for the mean of group two is equal to the Intercept plus 2.0000, which equals 5.0000. Likewise the estimate for group three is 3.0000 + 4.0000 which equals 7.0000.

Three Group Example in R

. . . now running an F-test on our linear model:

```
anova(m2)
```

```
## Analysis of Variance Table
##
## Response: Y
##             Df Sum Sq Mean Sq F value   Pr(>F)
## factor(IV0)  2     40    20.0       8 0.006196 **
## Residuals   12     30     2.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F-test of the main effect of the factor is called an omnibus test. A significant test indicates only that the population means are not equal — we would need to perform follow-up tests to find out specifically which groups differ.

Testing Normality

Use Shapiro-Wilk test on EACH group to test for normality (i.e., normal: $p > 0.05$).

```
shapiro.test(Y[1:5]) # group 1
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Y[1:5]
## W = 0.98676, p-value = 0.9672
```

```
shapiro.test(Y[6:10]) # group 2
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Y[6:10]
## W = 0.98676, p-value = 0.9672
```

```
shapiro.test(Y[11:15]) # group 3
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Y[11:15]
## W = 0.98676, p-value = 0.9672
```

There were no violations of Normality

Testing Homogeneity of Variances

Bartlett test to test whether variances are similar (i.e., equal homogeneity of variance: $p > 0.05$)

```
bartlett.test(Y ~ factor(myFac))
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Y by factor(myFac)
## Bartlett's K-squared = 0, df = 2, p-value = 1
```

There is no violation of Homogeneity of Variances: $p = 1.0$

- Data not normal? Can use transforms (e.g., log, square root, etc) OR a nonparametric version of the 1-way ANOVA, known as the Kruskal Wallis.
- Can perform Welch's corrections to F-test if variances are not equal using the oneway.test() OR use Kruskal Wallis.

Follow up Mean Comparison's

If ANOVA significant, as well as no violations of either normality or homogeneity of variance, you can perform follow-up mean compomparisons to test for differences between groups. Remember to correct for multiple comparisons!

```
pval_1v2 = t.test(Y[1:5], Y[6:10], alternative = "two.sided")$p.value # G1 vs G2
pval_1v3 = t.test(Y[1:5], Y[11:15], alternative = "two.sided")$p.value # G1 vs G3
pval_2v3 = t.test(Y[6:10], Y[11:15], alternative = "two.sided")$p.value # G2 vs G
pvals = c(pval_1v2, pval_1v3, pval_2v3)
p.adjust(pvals, method = "holm", n = length(pvals))
```

```
## [1] 0.16103248 0.01184932 0.16103248
```

Interpretation: There is a significant main of effect of group [F(2,12) = 8.0,p = 0.006], where Group 3 is significantly greater than Group 1 (p = 0.012).

- note: F(a-1, n - a)

Effect Size for ANOVA

One measure is the standardized measure of effect size, f

$$f = \frac{\sigma_m}{\sigma_\epsilon}$$

$$\sigma_m = \sqrt{\frac{\Sigma(\mu_j - \mu)^2}{k}}$$

$$\mu = \frac{(\Sigma \mu_j)}{k}$$

σ_m = between-group standard deviation

σ_ϵ = within-group standard deviation

- "small" effect: f = 0.10
- "medium" effect: f = 0.25
- "large" effect: f = 0.40

Other Effect Sizes options for ANOVA

- Eta-squared
- Omega squared (gold standard): Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychological methods, 8(4), 434.
- Provides estimates of effect size that are comparable across a variety of research designs
  - "small" effect: $\omega^2 = 0.01$
  - "medium" effect: $\omega^2 = 0.06$
  - "large" effect: $\omega^2 = 0.14$

http://daniellakens.blogspot.com/2015/06/why-you-should-use-omega-squared.html

ANOVA Effect Size in R

install.packages("effectsize")

```
library(effectsize)
cohens_f(m2)
```

```
## For one-way between subjects designs, partial eta squared is equvilant to eta squared.
## Returning eta squared.

## Parameter   | Cohen's f |       90% CI
## ----------------------------------------
## factor(IV0) |      1.15 | [0.46, 1.72]
```

```
eta_squared(m2)
```

```
## For one-way between subjects designs, partial eta squared is equvilant to eta squared.
## Returning eta squared.

## Parameter   | Eta2 |       90% CI
## ---------------------------------
## factor(IV0) | 0.57 | [0.17, 0.75]
```

```
omega_squared(m2)
```

```
## For one-way between subjects designs, partial omega squared is equvilant to omega squared.
## Returning omega squared.

## Parameter   | Omega2 |       90% CI
## -----------------------------------
## factor(IV0) |   0.48 | [0.07, 0.69]
```

Power Analysis on ANOVA

We can also run power analyses for omnibus tests (e.g., number of participants needed to find a sufficiently powered main effect or interaction).

- sometimes suggested by grant reviewers
- may not sufficiently power one for the smallest desired effect size of interest
- Recommendation: Perform on smallest desired effect size (e.g., mean comparison while controlling for multiple corrections).

Often complex tests cannot be performed analytically and you must use numerical methods.

- Same approach we have already done!

Power Analysis on ANOVA

- you are planning a reaction-time study involving three groups ($k = 3$)
- pilot research & data from literature suggest population means might be 400, 450 and 500 ms with a sample within-group standard deviation of 100 ms
- suppose you want a power of 0.80
- how many subjects do you need in each sample group?

Power Analysis on ANOVA

```
power.anova.test(groups=3, n=NULL, between.var=var(c(400,450,500)),
                 within.var=100**2, sig.level=0.05, power=0.80)
```

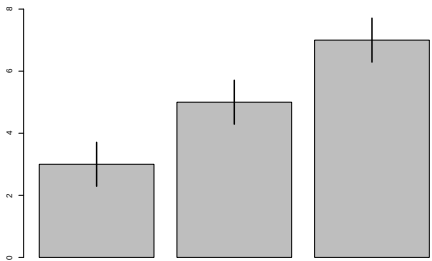
```
##
##      Balanced one-way analysis of variance power calculation
##
##           groups = 3
##                n = 20.30205
##     between.var = 2500
##      within.var = 10000
##       sig.level = 0.05
##           power = 0.8
##
## NOTE: n is number in each group
```

We need 21 participants per group for a sufficiently powered main effect.

Quick Plot of the Data

```
group_x <- c(mean(Y[1:5]), mean(Y[6:10]), mean(Y[11:15]))
group_ste <- c(sd(Y[1:5]) / sqrt(length(Y[1:5])),
sd(Y[6:10]) / sqrt(length(Y[6:10])), sd(Y[11:15])/ sqrt(length(Y[11:15])))

barCenters <- barplot(group_x, ylim=c(0,8))
segments(barCenters, group_x-group_ste,
         barCenters, group_x+group_ste, lwd=3)
```



Graphing with GG Plot

- There is a more powerful graphics package you can add to R called ggplot2
- There is lots of documentation about the ggplot2 package online, I suggest doing a google search.
- The ggplot2 package is also included in the meta-package called tidyverse, which includes other useful packages like dplyr and others.
- The following new (as of 2017) book is a great introduction to ggplot2, dplyr and other basic R functionality: R for Data Science by Hadley Wickham & Garrett Grolemund. O'Reilly (2017) ISBN: 978- 1491910399
- The book has a website where you can read everything online: http://r4ds.had.co.nz

Graphing with GG Plot - data frame

install.packages("ggplot2")

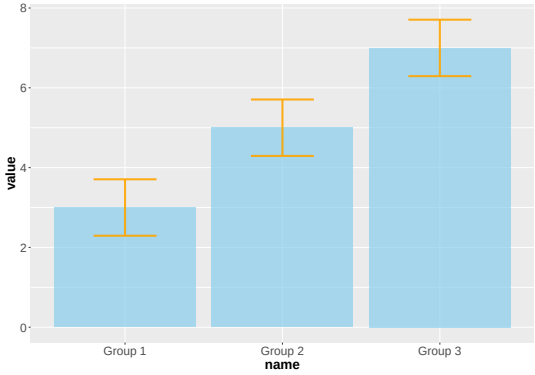
```
library(ggplot2)

group_x <- c(mean(Y[1:5]), mean(Y[6:10]), mean(Y[11:15]))
group_ste <- c(sd(Y[1:5]) / sqrt(length(Y[1:5])),
sd(Y[6:10])/ sqrt(length(Y[6:10])), sd(Y[11:15])/ sqrt(length(Y[11:15])))

# create data frame
data <- data.frame(
  name=c("Group 1", "Group 2", "Group 3"),
  value=group_x,
  sd=group_ste
)
```

Graphing with GG Plot - Bar Plots

```
# Basic error bar
ggplot(data) +
  geom_bar( aes(x=name, y=value), stat="identity", fill="skyblue", alpha=0.7) +
  geom_errorbar( aes(x=name, ymin=value-sd,ymax=value+sd), width=0.4, colour="orange", alpha=0.9, size=1.3) +
  theme(axis.text=element_text(size=16), axis.title=element_text(size=18,face="bold"))
```



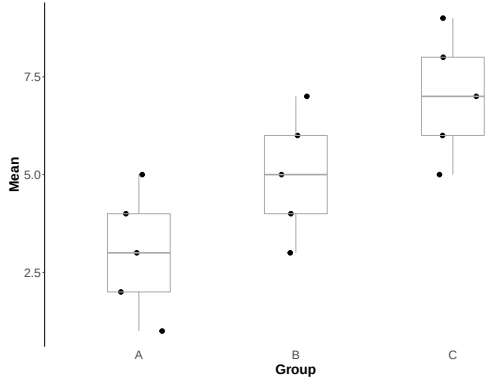
Graphing with GG Plot - Box and Whisker

```
Y <- c(4,5,2,1,3,7,4,6,3,5,6,9,8,5,7)
IV0 <- c("A","A","A","A","A","B","B","B","B","B","C","C","C","C","C")

# create data frame
data1 <- data.frame(
  name=IV0,
  value=Y,
  sd=group_ste
)
```

Graphing with GG Plot - Box and Whisker

```
ggplot(data1, mapping=aes(x=name, y=value)) +
  geom_point(position=position_jitter(width=0.15, height=0), color="black", size=2.5) +
  geom_boxplot(fill=NA, width=0.4, size=0.5, color="DarkGray") +
  theme_classic() +
  theme(axis.title=element_text(size=18,face="bold"), axis.text=element_text(size=16),
        axis.line.x=element_blank(), axis.ticks.x=element_blank()) +
  labs(x="Group", y="Mean")
```



Kruskal Wallis

Nonparametric (use when ANOVA assumptions violated).

Extension of the Mann-Whitney U test to handle 2 or more groups.

Let's perform on the previous example

```
kruskal.test(Y ~ factor(IV0))
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Y by factor(IV0)
## Kruskal-Wallis chi-squared = 8.1159, df = 2, p-value = 0.01728
```

There is a significant main effect of group ($p = 0.01728$). Can follow up with Mann-Whitney U tests for mean comparisons and common language effect size.

Next Week

- Factorial (2-way, 3-way, etc.) ANOVA