One-Way Analysis of Variance: ANOVA

Dr. J. Kyle Roberts

Southern Methodist University
Simmons School of Education and Human Development
Department of Teaching and Learning

Background to ANOVA

- Recall from the Independent Samples t test that we are testing to see if two means drawn from independent samples are statistically significantly different. With such, we are testing:

$$H_0 : \overline{X}_1 = \overline{X}_2$$

- where $\overline{X}_1$ and $\overline{X}_2$ are two sample means drawn from independent populations.
- While this is helpful for when $k = 2$, we must use alternative techniques when $k > 2$.
- In this case, we must use an F test instead of the previously used t test since we now have two sources of dfs.

Hypothesis Test in One-Way ANOVA

- In the ANOVA, we refer to the number of independent variables as either "ways" or "factors."
- The number of divisions of each "way" is referred to as "levels."
- Therefore, an analysis in which we are testing for the mean difference of 3 recognized ethnicites on a single dependent variable would be referred to as a one-way ANOVA with 3 levels.
- The null hypothesis for this test could then be written as:

$$H_0 : \overline{Y}_1 = \overline{Y}_2 = \overline{Y}_3$$

- assuming $\overline{Y}_1$, $\overline{Y}_2$ and $\overline{Y}_3$ are random samples drawn from $Y_{i1} \sim \mathcal{N}(\mu_1, \sigma^2)$, $Y_{i2} \sim \mathcal{N}(\mu_2, \sigma^2)$ and $Y_{i3} \sim \mathcal{N}(\mu_3, \sigma^2)$ distributions, respectively. Or alternatively as $Y_{ij} \sim \mathcal{N}(\mu_j, \sigma^2)$.

The One-Way ANOVA Summary Table

Source	SS	df	MS	F	p	η^2
Between	$SS_t - SS_w$	$k - 1$	$\frac{SS_b}{df_b}$	$\frac{MS_b}{MS_w}$		$\frac{SS_b}{SS_t}$
Within	$SS_t - SS_b$	$df_t - df_b$	$\frac{SS_w}{df_w}$			
Total	$SS_b + SS_w$	$n - 1$				

- The statistical significance of F can be obtained by computing the F-critical value. Determining statistical significance follows the same pattern for the t test only we have two sources of df: between and within.
- For a one-way ANOVA with 5 levels and 50 people, the critical value of F at $\alpha = 0.05$ would be:

```
> qf(0.95, 4, 45)

[1] 2.578739
```

One-Way ANOVA Practice

- Fill in the Missing Values Below
- F-crit=2.690

Source	SS	df	MS	F	p	η^2
Between	50	4				
Within		30				
Total	100					

- Fill in the Missing Values Below
- F-crit=2.922

Source	SS	df	MS	F	p	η^2
Between			10			0.15
Within		30				
Total	200					

Computing the Probability of F

- Based on our F from the ANOVA Summary Table previously, we can compute the probability of matching or exceeding the probability that there are no differences between the groups/levels.
- We compute this in R by:

```
> pf(7.5, 4, 30, lower = FALSE)
[1] 0.0002593994
> pf(1.765, 3, 30, lower = FALSE)
[1] 0.1750938
```

The "effects" form of the model

- An alternative representation of the model for testing mean differences among the way/factor is:

  $$Y_{ij} = \mu + \alpha_j + \epsilon_{ij},\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2),\ j = 1, \ldots, k,\ i = 1, \ldots, n_j$$

  This is a "signal + noise" form like the simple linear model.
- α_j is called the *effect* of level j of the factor.
- This means that an individual's score (Y_{ij}) can be thought of as the sum of the grand mean (μ) plus that individual group's deviation around the grand mean (α_j) and their own deviation around their group mean (ϵ_{ij}).

The Null Hypothesis for One-Way ANOVA

- In the effects form we write the null and alternative hypotheses as

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$$
$$H_a : \text{at least one } \alpha_i \neq 0$$

- Note that in order to reject H_0 we only need to have *one* mean different from the other means.
- This null may also be written as

$$H_0 : \overline{Y}_1 = \overline{Y}_2 = \cdots = \overline{Y}_k$$
$$H_a : \text{at least one } \overline{Y}_j \neq \text{any other } \overline{Y}_j$$

Computation of Sums of Squares

- Computation for SS_{total}.

$$SS_t = \sum_{i=1}^{N} (Y_{ij} - \overline{Y})^2$$

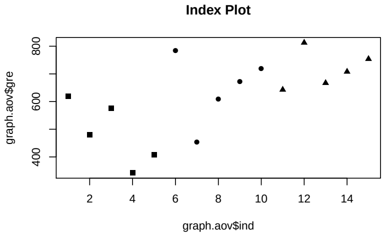
- Computation for $SS_{between}$.

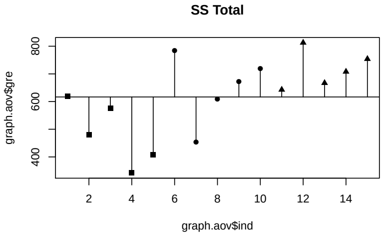
$$SS_b = \sum_{j=1}^{k} n_j (\overline{Y}_j - \overline{Y})^2$$

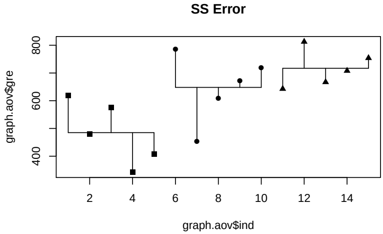
- Computation for SS_{within}.

$$SS_w = \sum_{i=1}^{N} (Y_{ij} - \overline{Y}_{.j})^2$$

Graphical Representation of Computation of SS





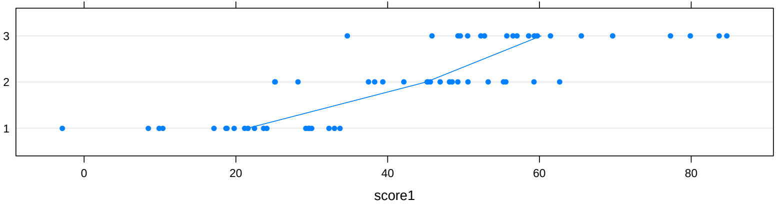


Understanding η^2 as a Measure of Effect Size

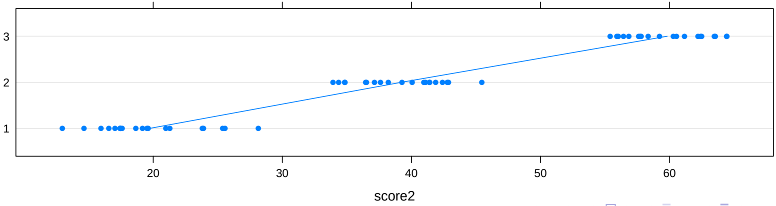
- Recall that for the t test, we measure the size of the effect of the mean difference as a function of the standarzied difference between the two means. This distance we called d or Δ.
- In the case of the ANOVA, we cannot compute a standardized difference of movement since we have no basis by which we would compute the difference in a case where $k > 2$.
- Instead we are going to compute the amount of variance in the dependent variable that is "explained" by the grouping variable.
- We compute this explained variance as $\eta^2 = SS_b/SS_t$, or as a ratio of the SS_t that is in SS_b (thus we can think of this as a percent since SS_b will never exceed SS_t).

Understanding η^2 cont.

- Consider the following two cases where $k = 3$.
- Assessment 1



- Assessment 2



Running ANOVA in R

- Consider the following dataset:
- Ethnicity 1-8,7,6,7,9,11,13
- Ethnicity 2-11,13,14,18,17,14,12,15
- Ethnicity 3-14,13,15,15,20,21,22

```
> ethdata <- data.frame(ethn = factor(rep(1:3, c(7,
+     8, 7))), score = c(8, 7, 6, 7, 9, 11, 13, 11,
+     13, 14, 18, 17, 14, 12, 15, 14, 13, 15, 15, 20,
+     21, 22))
> tapply(ethdata$score, ethdata$ethn, mean)

        1         2         3
 8.714286 14.250000 17.142857

> tapply(ethdata$score, ethdata$ethn, sd)

       1        2        3
2.497618 2.375470 3.716117
```

Assumptions for ANOVA

- Relatively the same number of people in each level
- Normality in the population for each of the levels
- Homogeneity of variance

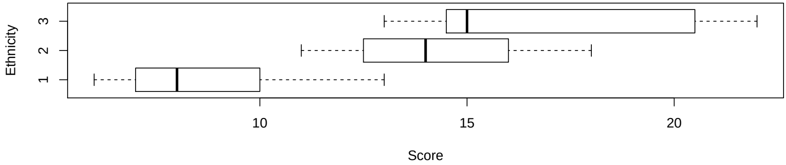
$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

```
> fligner.test(score ~ ethn, ethdata)
```

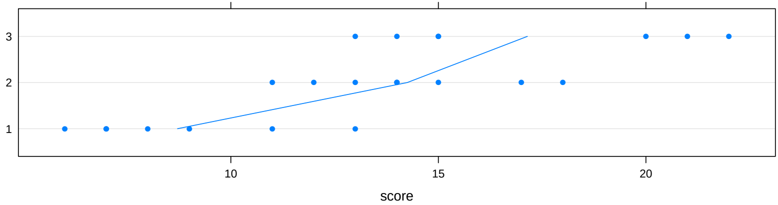
```
Fligner-Killeen test of homogeneity of variances
data:  score by ethn
Fligner-Killeen:med chi-squared = 1.3404, df = 2, p-value
= 0.5116
```

Graphing Ethnicity Data

```
> boxplot(score ~ ethn, ethdata, ylab = "Ethnicity",
+     xlab = "Score", horizontal = TRUE)
```

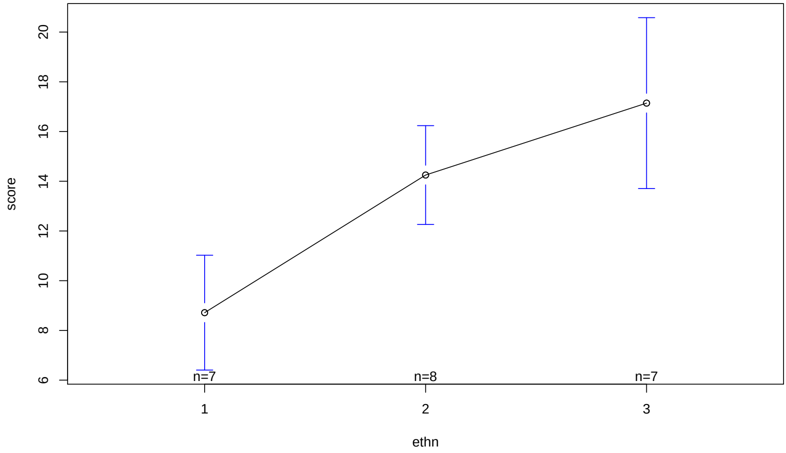



```
> print(dotplot(ethn ~ score, ethdata, type = c("p",
+     "a")))
```



Graphing Ethnicity Data, cont.

```
> library(gregmisc)
> plotmeans(score ~ ethn, ethdata)
```



ANOVA in R cont.

```
> m1 <- aov(score ~ ethn, ethdata)
> anova(m1)

Analysis of Variance Table
Response: score
          Df Sum Sq Mean Sq F value    Pr(>F)
ethn       2 257.53  128.77  15.312 0.0001094
Residuals 19 159.79    8.41
```

- Note that R puts the ANOVA summary table in a slightly different format than we will report. There is no "Total" row and there is no η^2.
- For this case η^2 is computed as $257.53/(257.53 + 159.79) = 0.617$.

Testing Pairwise Comparisons

- If we reject H_0 for the F test, we conclude that there are significant differences between the groups. Usually we follow up and determine which groups are significantly different.
- If we test all possible pairs of groups for significant differences we will perform $\binom{k}{2} = k(k-1)/2$ separate tests. We say we are doing *multiple comparisons*.
- Using t tests without any adjustment for the multiple comparisons inflates the probability of declaring a significant difference when there isn't one.
- There are several techniques for adjusting the tests. One of the most common is Tukey's "honest significant difference" (function *TukeyHSD*), which is based on the Studentized range.

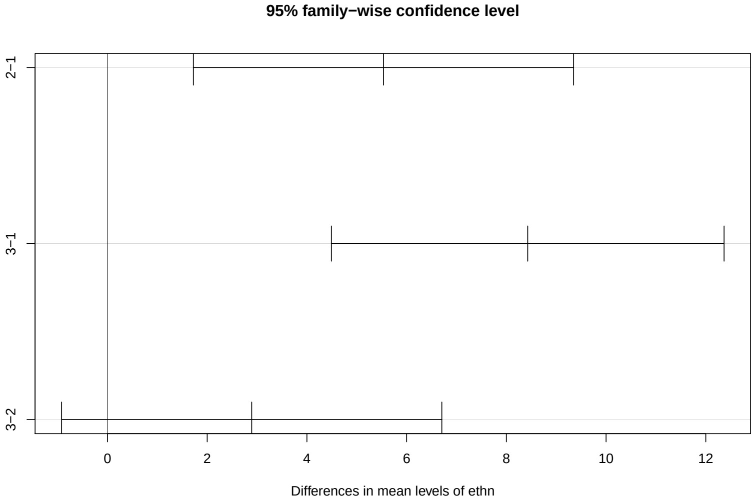
Tukey's HSD for ANOVA

- The Tukey's HSD provides a correction factor to the pairwise comparisons such that the p-value is slightly inflated.
- These adjustments are based on the number of comparisons.

```
> TukeyHSD(m1)

  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = score ~ ethn, data = ethdata)

$ethn
        diff        lwr       upr     p adj
2-1 5.535714  1.7228228  9.348606 0.0042454
3-1 8.428571  4.4906341 12.366509 0.0000860
3-2 2.892857 -0.9200343  6.705749 0.1582960
```

Plotting Tukey HSD

```
> plot(TukeyHSD(m1))
```



What to do if you do not meet the homogeneity of variance assumption

Suppose that you have a dataset in which you run the Fligner Test and discover that $p < 0.05$. In this case, we can apply a correction factor to the within df to compute the relative F when this assumption is not met.

```
> hovdata <- data.frame(score = c(1:10, 200), grp = factor(rep(1
+      c(5, 6))))
> fligner.test(score ~ grp, hovdata)

Fligner-Killeen test of homogeneity of variances
data:  score by grp
Fligner-Killeen:med chi-squared = 0.4948, df = 1, p-value
= 0.4818

> oneway.test(score ~ grp, hovdata)

One-way analysis of means (not assuming equal variances)
data:  score and grp
F = 1.3358, num df = 1.000, denom df = 5.005, p-value =
0.2999
```

Proving the Importance of the Way

- Consider the following dataset in which there are two separate ways `g1` and `g2` for the `dv`.

```
> newtrial <- data.frame(dv = c(1:9, 11), g1 = rep(1:2,
+     5), g2 = rep(1:2, each = 5))
```

- Run two separate ANOVAS for both `g1` and `g2` on `dv`. Why is it that the first ANOVA was not statistically significant and the second one was when we used the *same* dependent variable?

In-Class Practice Example with 5 Levels

- Create the following dataset in R.
- Test all assumptions and run all appropriate post-hoc tests.

Program 1	Program 2	Program 3	Program 4	Program 5
30	32	31	43	44
33	35	34	47	50
31	28	33	53	50
25	29	32	54	49
26	19	29	52	47
29	20	30	55	49
29	20	31	45	49
31		31		

Out of Class Homework Assignment

- Create heuristic data for a one-way ANOVA with 4 levels. You must have at least 10 people in each level.

1. Create the first ANOVA such that results are statistically significant and in which the post-hoc tests reveal:

$$\overline{Y}_1 \neq \overline{Y}_2, \overline{Y}_1 \neq \overline{Y}_3, \overline{Y}_1 \neq \overline{Y}_4$$

   but

$$\overline{Y}_2 = \overline{Y}_3 = \overline{Y}_4$$

2. Create the second ANOVA such that results are statistically significant and in which the post-hoc tests reveal:

$$\overline{Y}_1 = \overline{Y}_2 \text{ and } \overline{Y}_3 = \overline{Y}_4$$

   but

$$\overline{Y}_1 \text{ and } \overline{Y}_2 \neq \overline{Y}_3 \text{ and } \overline{Y}_4$$