PSY 3307

Analysis of Variance (ANOVA)

Jonathan A. Pedroza PhD

Cal Poly Pomona

2021-10-21

Agenda

- Post-hoc tests
  - Tukey HSD
  - Bonferroni Correction
- Effect sizes
- Proportion of Variance Accounted For
- What is an ANCOVA

Post-hoc Tests

- **Tukey HSD (Honestly Significant Difference)** is a post hoc test for ANOVA to compare mean for each level in a factor
    - used only when the `ns` in each level of the factor are equal
- it computes the the minimum difference between two means that is required for them to be significantly different
- Bonferroni Correction
    - based on your alpha, you get that value and divide by the number of tests you need to conduct

From last class

$$\begin{bmatrix} Source \\ Between \\ Within \\ Total \end{bmatrix} \begin{bmatrix} Sum\ of\ Squares \\ 63.33 \\ 84 \\ 147.33 \end{bmatrix} \begin{bmatrix} df \\ 2 \\ 12 \\ 14 \end{bmatrix} \begin{bmatrix} MeanSquare \\ 31.67 \\ 7 \\ \end{bmatrix} \begin{bmatrix} F-ratio \\ 4.52 \\ \\ \end{bmatrix}$$

Tukey Steps

1. Find `q_{k}`, which are values of studentized range statistic
    - We'll come back to studentized values
    - locate the `k`
    - find the `df_{wm}`
2. Compute the HSD

$$HSD = (q_k)(\sqrt{\frac{MS_{wn}}{n}})$$

1. Determine the differences between each pair of means

2. Compare each difference between means to the Tukey HSD value

    - if the absolute difference between two means is greater than the HSD then these differences are significantly different
    - similar to a independent samples t-test with a significant t-obtained value

Example from last slides

MS within = 7 n = 5

For `q_{k}`, we need k and `df_{wn}` df_wn = 12 k = 3 q_k = 3.77

```
3.77*(sqrt(7/5))
```

```
## [1] 4.460724
```

Our Tukey HSD is 4.46

So now we compare the means of our three groups/levels

```
8 - 6 # easy - medium
```

```
## [1] 2
```

```
8 - 3 # easy - hard
```

```
## [1] 5
```

```
6 - 3 # medium - hard
```

```
## [1] 3
```

2 < 4.46 5 > 4.46 3 < 4.46

So only the difference between easy and the hard groups was a statistically significant finding at an alpha of .05

Effect Size

- cohen's d only really used for two-sample designs (t-tests)
- rather, for ANOVA and regression, we will rely on **Proportion of variance accounted for**
    - there are better effect sizes (most would argue that proportion of variance accounted for in regression models is not an effect size)
- greater the eta squared, the more that our IV accounts for the DV
- eta squared is the same as R squared
- **eta squared** is the proportion of variance in the DV that is accounted for with the change in levels in the factor/IV

Formula

$$\eta^2 = \frac{SS_{bn}}{SS_{total}}$$

- SS between is the differences between all the scores in the experiment
  - it is also what we know our model is actually looking at
- because it is proportion, we can easily interpret it as the percentage that the change in groups/levels accounts for the DV score difference

$$ANOVA: \ eta^{squared} = \eta^2$$

$$Regression: \ R^2$$

ANCOVA

- conceptually the same thing as ANOVA but it includes covariates/control variables
- ANCOVA = Analysis of Covariance
- The same thing as multiple regression
- You are still interested in differences between levels on your DV values, but now you include additional variables to help account for some of the variance in your DV

$$\begin{bmatrix} Source \\ Between \\ Within \\ Total \end{bmatrix} \begin{bmatrix} Sum\ of\ Squares \\ SS_{bn} \\ SS_{wn} \\ SS_{total} \end{bmatrix} \begin{bmatrix} df \\ df_{bn} \\ df_{wn} \\ df_{total} \end{bmatrix} \begin{bmatrix} MeanSquare \\ MS_{bn} \\ MS_{wn} \\ \end{bmatrix} \begin{bmatrix} F-ratio \\ F_{obt} \\ \\ \end{bmatrix}$$

Context

You are trying out four different methods for helping participants with their depression. You decide to give out: medication only, individual therapy, group, and medication and individual therapy. Below are each level's depression levels at the end of the experiment. You are interested in a difference between the four groups.

```
data <- data.frame(meds = c(4, 5, 4, 6, 5),
                   ind_therapy = c(5, 4, 3, 5, 5),
                   group = c(8, 8, 6, 9, 10),
                   med_ind = c(3, 1, 3, 2, 1))

data
```

```
##   meds ind_therapy group med_ind
## 1    4           5     8       3
## 2    5           4     8       1
## 3    4           3     6       3
## 4    6           5     9       2
## 5    5           5    10       1
```

Hypotheses

H0: There will be no differences between the therapeutic approaches

H1: There will be differences between the therapeutic approaches

Even better,

H1: The group that receives medication and individual therapy will have significantly better or worse depression rates than every other group.

Step-by-step

What is the k?

What is the number of participants per level?

What is the total number of participants?

What is the sum of each group?

What is the mean of each group?

What is the sum of squared Xs for reach group?

What is the total sum?

What is the total sum of all squared values?

What is the sum of squares total?

$$SS_{total} = \Sigma(X_{ij} - \overline{\overline{X}})^2$$

$$SS_{total} = \Sigma X_{total}^2 - \frac{(\Sigma X_{total})^2}{N}$$

What is the sum of squares between/treatment?

$$SS_{treat} = n\Sigma(\overline{\overline{X_j}} - \overline{\overline{X}})^2$$

$$SS_{bn} = \Sigma(\frac{(\Sigma X \ in \ each \ column)^2}{n \ in \ each \ column}) - \frac{(\Sigma X_{total})^2}{N}$$

What is the sum of squares within/error?

$$SS_{wn} = SS_{total} - SS_{bn}$$

$$SS_{error} = SS_{total} - SS_{treat}$$

What are the degrees of freedom between?

$$df_{bn} = k - 1$$

What are the degrees of freedom within? (N - k)

What are the degrees of freedom total? (N - 1)

What is the mean squares between value?

$$MS_{bn} = \frac{SS_{bn}}{df_{bn}}$$

What is the mean squares within value?

$$MS_{wn} = \frac{SS_{wn}}{df_{wn}}$$

What is the F-obtained value?

$$F_{obt} = \frac{MS_{bn}}{MS_{wn}}$$

Calculate the Tukey HSD

$$HSD = (q_k)(\sqrt{\frac{MS_{wn}}{n}})$$

Effect Size

$$\eta^2 = \frac{SS_{bn}}{SS_{total}}$$

What can we say?

- There is no significant difference between medications and individual therapy in depression scores
- There are lower depression scores in the group with individual therapy compared to group therapy
- There are lower depression scores in the group with medications compared to group therapy
- There are lower depression scores in the group with individual therapy and medications compared to group therapy
- There are lower depression scores in the group with individual therapy and medications compared to medications alone

- There are lower depression scores in the group with individual therapy and medications compared to individual therapy alone

- Effect size: The differences in therapeutic styles accounted for 83.87% of the variance in depression scores in our sample.

```r
data2 <- data.frame(meds = c(4, 5, 4, 6, 5),
                    ind_therapy = c(5, 4, 3, 5, 5),
                    group = c(8, 8, 6, 9, 10),
                    med_ind = c(3, 1, 3, 2, 1),
                    control = c(10, 9, 8, 7, 9))

data2
```

```
##   meds ind_therapy group med_ind control
## 1    4           5     8       3      10
## 2    5           4     8       1       9
## 3    4           3     6       3       8
## 4    6           5     9       2       7
## 5    5           5    10       1       9
```

Step-by-step

What is the k?

What is the number of participants per level?

What is the total number of participants?

What is the sum of each group?

What is the mean of each group?

What is the sum of squared Xs for reach group?

What is the total sum?

What is the total sum of all squared values?

What is the sum of squares total?

$$SS_{total} = \Sigma(X_{ij} - \overline{\overline{X}})^2$$

$$SS_{total} = \Sigma X_{total}^2 - \frac{(\Sigma X_{total})^2}{N}$$

What is the sum of squares between/treatment?

$$SS_{treat} = n\Sigma(\overline{\overline{X_j}} - \overline{\overline{X}})^2$$

$$SS_{bn} = \Sigma(\frac{(\Sigma X \ in \ each \ column)^2}{n \ in \ each \ column}) - \frac{(\Sigma X_{total})^2}{N}$$

What is the sum of squares within/error?

$$SS_{wn} = SS_{total} - SS_{bn}$$

$$SS_{error} = SS_{total} - SS_{treat}$$

What are the degrees of freedom between?

$$df_{bn} = k - 1$$

What are the degrees of freedom within? (N - k)

What are the degrees of freedom total? (N - 1)

What is the mean squares between value?

$$MS_{bn} = \frac{SS_{bn}}{df_{bn}}$$

What is the mean squares within value?

$$MS_{wn} = \frac{SS_{wn}}{df_{wn}}$$

What is the F-obtained value?

$$F_{obt} = \frac{MS_{bn}}{MS_{wn}}$$

Calculate the Tukey HSD

$$HSD = (q_k)(\sqrt{\frac{MS_{wn}}{n}})$$

$$\eta^2 = \frac{SS_{bn}}{SS_{total}}$$

What can we say?

- There is no significant difference between medications and individual therapy in depression scores
- There are lower depression scores in the group with medications compared to group therapy
- There are lower depression scores in the group with individual therapy and medications compared to medications alone
- There are lower depression scores in the group with medications compared to the control group
- There are lower depression scores in the group with individual therapy compared to group therapy

- There are lower depression scores in the group with individual therapy and medications compared to individual therapy alone
- There are lower depression scores in the group with individual therapy compared to the control group
- There are lower depression scores in the group with individual therapy and medications compared to group therapy
- There are no significant differences between the group therapy group and the control group

- There are lower depression scores in the group with individual therarpy and medications compared to the control group

- Effect size: The differences in therapeutic styles accounted for 91.61% of the variance in depression scores in our sample.