

Nested & Incomplete Designs in ANOVAs

Gemma Wallace & Neil Yetz
PSY 653 Module 3 Lab
Feb 12, 2020

A quick review of ANOVAs

An example from last semester:

```
# Conduct an ANOVA
```

```
```{r}
```

```
lm1 <- lm(StressLevel ~ ProgramCode.f, data = stress)
```

```
anova(lm1)
```

```
```
```

Analysis of Variance Table

Response: StressLevel

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---------------|-----|--------|---------|---------|-----------|
| ProgramCode.f | 3 | 54.83 | 18.2750 | 3.5623 | 0.01643 * |
| Residuals | 116 | 595.10 | 5.1302 | | |

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

ANOVAs evaluate how a continuous outcome varies across levels of a categorical predictor(s)

Sum of squares (SS): The sum of squared deviations from each individual value and the mean

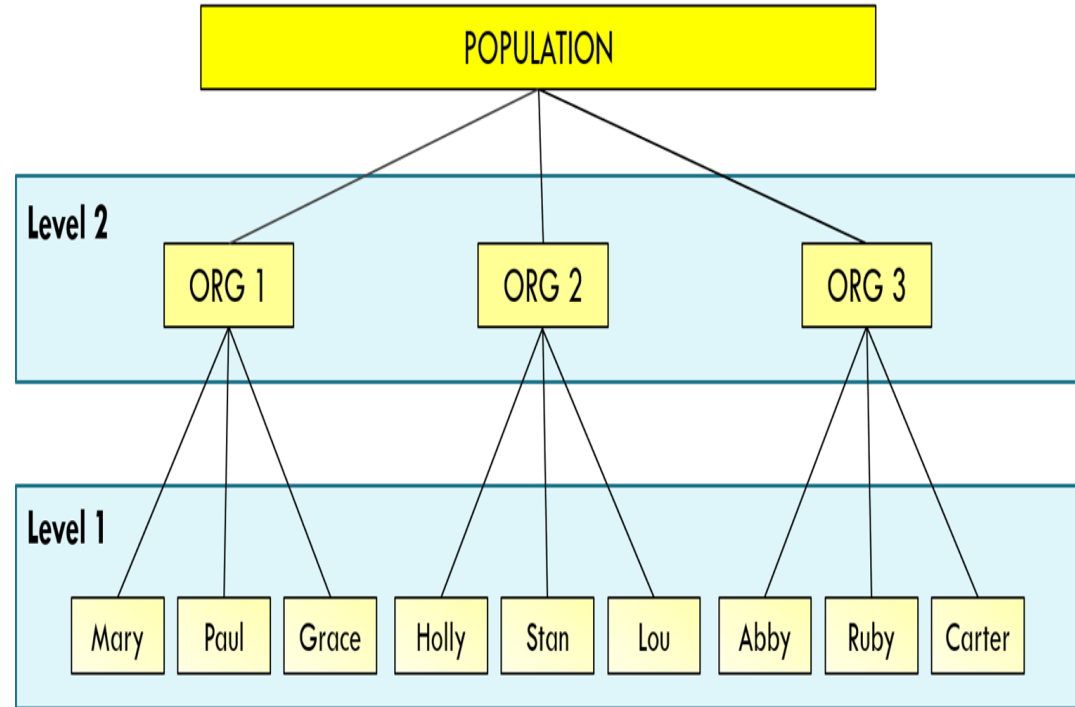
Mean square (MS): The average distance values deviate from the mean (SS/df)

F-statistic: The ratio of model mean square to the residual mean square ($MS_{\text{explained}}/MS_{\text{residual}}$)

The p-value for the model F-test is significant at $p < 0.05$, indicating that there was a significant effect of ProgramCode on stress level.

What is a nested design?

- A **nested design** (aka hierarchical design) is when a factor(s) is *nested* within another factor.
- A **nested ANOVA** (aka hierarchical ANOVA) is used to evaluate the differences within variables while taking into account the nested structure of the data

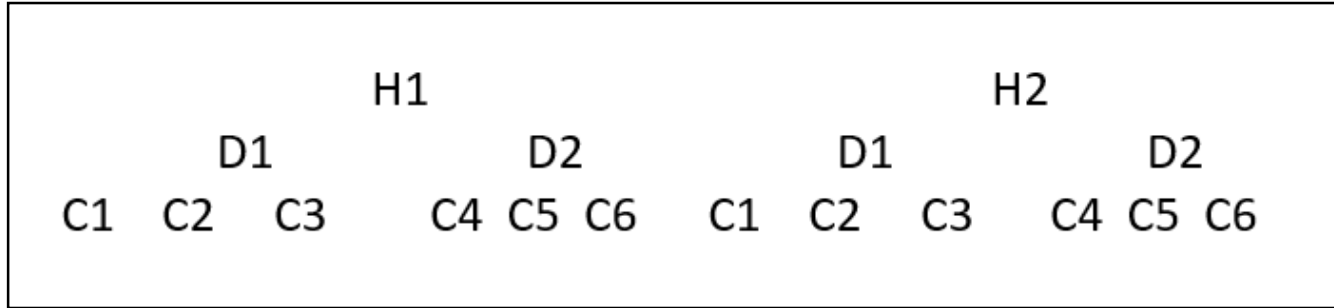


©Kimberly Henry

Practice with nested ANOVAs

- This demo uses data from 114 patients who participated in study to evaluate the effects of different drugs and treatment conditions on health.
- Our overall research question:
 - *Does patient health significantly vary across hospitals, drugs, and treatment conditions?*
 - Outcome = Y (a proxy for health)
 - Predictors = Hospital, Drug, Condition

Hierarchical structure of the data



Drug is not nested within hospital in this data, because the drugs are the same between the two hospitals (i.e., the effects of a particular drug could be different from one hospital to the next).

Condition is not nested within hospital in this data, because the conditions are the same between the two hospitals (i.e., the effects of a particular condition could be different from one hospital to the next).

However, **condition is nested within drug**, because the conditions are different between the two drugs. (i.e., since each condition was not tested for each drug, we cannot assess the possibility that the different conditions are different between the drugs).

Set up a new R notebook and load the following libraries

```
---  
title: "PSY 653 Module 3: Nested Designs"  
subtitle: "February 12, 2020"  
output: html_notebook  
---
```

```
# Part 1: Demo of Nested ANOVAs
```

```
12 ## Load libraries  
13 ```{r, message=FALSE}  
14 library(psych) # using for describe() function  
15 library(tidyverse) # using for read_csv() function, so character variables still kept as character  
16 ```  
17
```

Read in the data

```
15 ## Read data
16 {r, message=FALSE}
17 nested_demo<-read_csv(file="Nested_demo.csv")
18
19
```

| | Y | Hospital | Drug | Condition |
|----|---|----------|------|-----------|
| 1 | 5 | H1 | D1 | C1 |
| 2 | 4 | H1 | D2 | C2 |
| 3 | 3 | H1 | D1 | C3 |
| 4 | 5 | H1 | D2 | C4 |
| 5 | 4 | H1 | D1 | C5 |
| 6 | 3 | H1 | D2 | C6 |
| 7 | 6 | H1 | D1 | C1 |
| 8 | 5 | H1 | D2 | C2 |
| 9 | 4 | H1 | D1 | C3 |
| 10 | 6 | H1 | D2 | C4 |
| 11 | 5 | H1 | D1 | C5 |

- Y = the outcome variable for health
- H = hospital
- D = drug
- C = condition

Check how variables are read and get descriptives

```
## Check how variables are read
```

```
` `{r}  
str(nested_demo)  
` `
```



```
Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame':  
 $ Y          : num  5 4 3 5 4 3 6 5 4 6 ...  
 $ Hospital   : chr   "H1" "H1" "H1" "H1" ...  
 $ Drug       : chr   "D1" "D2" "D1" "D2" ...  
 $ Condition: chr   "C1" "C2" "C3" "C4" ...  
 - attr(*, "spec")=  
 .. cols(  
 ..   Y = col_double(),  
 ..   Hospital = col_character(),  
 ..   Drug = col_character(),  
 ..   Condition = col_character()  
 .. )
```

```
## Get variable descriptives
```

```
` `{r}  
describe(nested_demo)  
` `
```



| | vars
<dbl> | n
<dbl> | mean
<dbl> | sd
<dbl> | median
<dbl> |
|------------|---------------|------------|---------------|-------------|-----------------|
| Y | 1 | 114 | 8.75 | 2.91 | 9 |
| Hospital* | 2 | 114 | NaN | NA | NA |
| Drug* | 3 | 114 | NaN | NA | NA |
| Condition* | 4 | 114 | NaN | NA | NA |

*Descriptives show NaN because
predictors are character variables*

Get descriptives by Hospital, Drug & Condition

```
25 # Calculate the mean values of Y by hospital, drug, and condition
26 ```{r}
27 # aggregate function here just shows the values that we're comparing, but not saving these values
28 aggregate(nested_demo$Y, by=list(Hospital=nested_demo$Hospital), mean)
29 aggregate(nested_demo$Y, by=list(Drug=nested_demo$Drug), mean)
30 aggregate(nested_demo$Y, by=list(Condition=nested_demo$Condition), mean)
31
```

Before running analyses, it's helpful to evaluate how much the mean of Y differs across the levels of each categorical predictor variable

The aggregate function, broken down

```
2 aggregate(nested_demo$Y, by=list(Hospital=nested_demo$Hospital), mean)
```

- `aggregate()` = function
- `nested_demo$Y` = Continuous outcome variable
- `Hospital` = New name given for output (must be 1 word)
- `nested_demo$Hospital` = Categorical predictor variable
- `mean` = Desired descriptive function (i.e. mean, sd, median, etc.)

Get descriptives by Hospital, Drug & Condition

```
25 # Calculate the mean values of Y by hospital, drug, and condition
26 ```{r}
27 # aggregate function here just shows the values that we're comparing, but not saving these values
28 aggregate(nested_demo$Y, by=list(Hospital=nested_demo$Hospital), mean)
29 aggregate(nested_demo$Y, by=list(Drug=nested_demo$Drug), mean)
30 aggregate(nested_demo$Y, by=list(Condition=nested_demo$Condition), mean)
31
```

Aggregate Results

Hospital

<fctr>

x

<dbl>

H1

8.759259

H2

8.733333

Drug

<fctr>

x

<dbl>

D1

8.754386

D2

8.736842

2 rows

Condition

<fctr>

x

<dbl>

C1

8.894737

C2

8.631579

C3

8.578947

C4

9.789474

C5

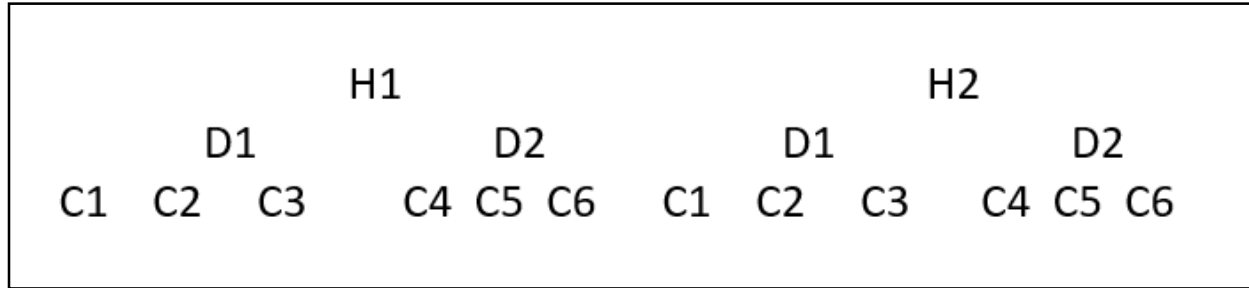
8.789474

C6

7.789474

6 rows

Getting ready to run an ANOVA: get degrees of freedom for each effect



Degrees of freedom in ANOVAS:

$$\text{df-error} = N - k$$

$$\text{df-hyp} = k - 1$$

Remember: k = the # of groups!

Degrees of freedom in our nested design

Hospital df: $(2-1) = 1$

Drug df: $(2-1) = 1$

Condition within Drug df: $(3-1) + (3-1) = 4$

Hospital * Drug df: $(2-1) * (2-1) = 1$

Hospital * (Condition within Drug) df: $(2-1) * [(3-1) + (3-1)] = 4$

ANOVA Table: List factors and df for each

| Source | Degrees of freedom (df) |
|---|-------------------------|
| Hospital | 1 |
| Drug | 1 |
| Condition <i>within</i> Drug | 4 |
| Hospital * Drug | 1 |
| Hospital * (Condition <i>within</i> Drug) | 4 |
| Error | 102 |

These are all of the effects
we can test in this data

We want to know if Y varies across these
categorical predictor variables, and if these
effects are significant

Conduct a Hospital x Drug ANOVA

```
## (5b) Conduct a Hospital x Drug ANOVA, ignoring Condition
```

```
`{r}
```

```
mod1 <- anova(lm(Y~Hospital + Drug + Hospital*Drug, data=nested_demo))
```

```
mod1
```

```
`{r}
```

Analysis of Variance Table

Response: Y

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---------------|-----|--------|---------|---------|--------|
| Hospital | 1 | 0.02 | 0.0191 | 0.0022 | 0.9626 |
| Drug | 1 | 0.01 | 0.0088 | 0.0010 | 0.9747 |
| Hospital:Drug | 1 | 0.08 | 0.0764 | 0.0088 | 0.9254 |
| Residuals | 110 | 953.52 | 8.6684 | | |

* specifies “interaction”

For this first step, we are ignoring condition and just looking at Hospital, Drug, and the interaction between the two as predictors of Y.

Plug results from the Hospital x Drug model to the ANOVA table

| Source | df | Sum Sq | Mean Sq | F value | <i>p</i> |
|---------|-----|--------|---------|---------|----------|
| H | 1 | 0.02 | 0.0191 | 0.0022 | 0.9626 |
| D | 1 | 0.01 | 0.0088 | 0.0010 | 0.9747 |
| D/C | - | - | - | - | - |
| H*D | 1 | 0.08 | 0.764 | 0.0088 | 0.9254 |
| H*(D/C) | - | - | - | - | - |
| Error | 110 | 953.52 | 8.6684 | - | - |

Conduct separate ANOVA for Condition under each possible combination of Drug and Hospital

```
## 5c Conduct separate ANOVA for effect of condition
```

```
```{r}
mod2 <- anova(lm(Y~Drug/Condition + Hospital*(Drug/Condition), data=nested_demo))
mod2 # get same model output as for the full model
```
```

Analysis of Variance Table

Response: Y

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-------------------------|-----|--------|---------|---------|--------|
| Drug | 1 | 0.01 | 0.0088 | 0.0010 | 0.9751 |
| Hospital | 1 | 0.02 | 0.0191 | 0.0021 | 0.9633 |
| Drug:Condition | 4 | 39.30 | 9.8246 | 1.0970 | 0.3622 |
| Drug:Hospital | 1 | 0.08 | 0.0764 | 0.0085 | 0.9266 |
| Drug:Condition:Hospital | 4 | 0.75 | 0.1884 | 0.0210 | 0.9991 |
| Residuals | 102 | 913.47 | 8.9556 | | |

/ specifies “within”
* specifies “interaction”

*Note: order matters when specifying nesting!
List the higher level first*

Plug results from Condition model into the ANOVA table

| Source | df | Sum Sq | Mean Sq | F value | <i>p</i> |
|---------|-----|--------|---------|---------|----------|
| H | 1 | 0.02 | 0.0191 | 0.0021 | 0.9633 |
| D | 1 | 0.01 | 0.0088 | 0.0010 | 0.9751 |
| D/C | 4 | 39.30 | 9.8246 | 1.0970 | 0.3622 |
| H*D | 1 | 0.08 | 0.764 | 0.0085 | 0.9266 |
| H*(D/C) | 4 | 0.75 | 0.1884 | 0.0210 | 0.9991 |
| Error | 102 | 913.47 | 8.9556 | - | - |



= added information

Interpret the ANOVA

Analysis of Variance Table

Response: Y

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-------------------------|-----|--------|---------|---------|--------|
| Hospital | 1 | 0.02 | 0.0191 | 0.0021 | 0.9633 |
| Drug | 1 | 0.01 | 0.0088 | 0.0010 | 0.9751 |
| Drug:Condition | 4 | 39.30 | 9.8246 | 1.0970 | 0.3622 |
| Hospital:Drug | 1 | 0.08 | 0.0764 | 0.0085 | 0.9266 |
| Hospital:Drug:Condition | 4 | 0.75 | 0.1884 | 0.0210 | 0.9991 |
| Residuals | 102 | 913.47 | 8.9556 | | |

None of the F statistics for the predictors are statistically significant at $p < 0.05$, indicating that patient health (Y) did not significantly differ between the two hospitals, the two drugs, and the six treatment conditions.

An alternative method to get the same ANOVA table

- We ran two smaller models to fill in our ANOVA table for all of the factors in the data
- Alternatively, you could run the full model with one line of code to get the same results in fewer steps:

```
## Full model code
```

```
```{r}
```

```
mod_full <- anova(lm(Y~Hospital + Drug + Drug/Condition + Hospital*Drug + Hospital*(Drug/Condition), data=nested_demo))
```

```
mod_full
```

```
```
```

Analysis of Variance Table

Response: Y

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-------------------------|-----|--------|---------|---------|--------|
| Hospital | 1 | 0.02 | 0.0191 | 0.0021 | 0.9633 |
| Drug | 1 | 0.01 | 0.0088 | 0.0010 | 0.9751 |
| Drug:Condition | 4 | 39.30 | 9.8246 | 1.0970 | 0.3622 |
| Hospital:Drug | 1 | 0.08 | 0.0764 | 0.0085 | 0.9266 |
| Hospital:Drug:Condition | 4 | 0.75 | 0.1884 | 0.0210 | 0.9991 |
| Residuals | 102 | 913.47 | 8.9556 | | |