PS303: Week 12

pp. 536-549

Recap

- Factorial ANOVAs describe whether $k \geq 2$ independent variables can singularly or interactively predict variances across a dependent outcome
- The computed F-ratio tells us how likely the present data is if the null hypothesis (H_0) is true. If the present data is 'extremely' unlikely ($p < .05$), than H_0 can be rejected
- The *practical* (beyond statistical) importance of a significant model can be described through effect sizes (e.g.,$\eta_p^2 = \frac{SS_M}{SS_M + SS_R}$)
- For significant models that contain independent variables with ≥ 3 levels, we run post-hoc tests to estimate whether differences between pairs of groups are statistically significant

---

- Assumptions for running conventional ANOVAs include:
  - *Homogeneity of variance*: Are the samples being compared statistically equivalent ($p \geq .05$) along shared variance? Answered using Levene's test.
  - *Normality of data*: Are the residuals of the model normally distributed? Answered using histograms, QQ-plots and Shapiro tests.
  - *Balanced design*: Are observations equally distributed across all combinations of independent levels?

Not meeting these assumptions can generate *biased* outcomes.

Parameter inputs for ANOVAs and OLS regressions are similar in R
(e.g. `lm(DV~IV1+IV2, data)= aov(DV~IV1+IV2, data)`) because both are *linear* models.

Running a balanced ANOVA with post-hoc tests

Let's re-run our earlier ANOVA but this time with an additional cohort from Kiribati. We now have 3 levels for the Location predictor (*Fiji, Singapore, Kiribati*) and 2 levels for the Depression predictor (*Low, High*). This would be a 2×3 independent ANOVA.

ID	Location	Depression	Weekly alcohol consumption (ml)
1	Fiji$_1$	Low$_1$	311
2	Fiji$_2$	Low$_2$	320
3	Fiji$_3$	Low$_3$	313
4	Singapore$_1$	Low$_4$	443
5	Singapore$_2$	Low$_5$	441

ID	Location	Depression	Weekly alcohol consumption (ml)
6	Singapore_3	Low_6	480
7	Kiribati_1	Low_7	320
8	Kiribati_2	Low_8	353
9	Kiribati_3	Low_9	313
10	Fiji_4	High_1	385
11	Fiji_5	High_2	420
12	Fiji_6	High_3	412
13	Singapore_4	High_4	557
14	Singapore_5	High_5	519
15	Singapore_6	High_6	608
16	Kiribati_4	High_7	512
17	Kiribati_5	High_8	487
18	Kiribati_6	High_9	526

We can average across each row (R) and column (C) to respectively extract marginal means for **Location** and **Depression** factors respectively.

	Fiji ($Col1$)	Singapore ($Col2$)	Kiribati ($Col3$)	Marginal row (R) means
Low depression ($Row1$)	314.67	454.67	328.67	$Row1_\mu = 366$
High depression ($Row2$)	405.67	561.33	508.33	$Row2_\mu = 491.78$
Marginal column means (C_μ)	$Col1_\mu = 360.17$	$Col2_\mu = 508$	$Col3_\mu = 418.5$	$Grand_\mu = 366$

We can declare the same null hypotheses as before:

- $H_0 1$: There is no difference in alcohol consumed between participants categorized as low and high depressed ($Row1_\mu = Row1_\mu$)
- $H_0 2$: There is no difference in alcohol consumed between participants from Fiji, Singapore and Kiribati ($Col1_\mu = Col2_\mu = Col3_\mu$)

Let’s set up our data…

```r
ID         <- seq(1:18)                                                                    # 1
8 participants
Location   <- rep(c(rep("Fiji",3),rep("Singapore",3),rep("Kiribati",3)),2)                 # 3
Location levels
Depression <- c(rep("Low",9),rep("High",9))                                                # 2
Depression levels
Alcohol    <- c(311,320,313,443,441,480,320,353,313,385,420,412,557,519,608,512,487,526)   # A
lcohol drunk
df <- cbind.data.frame(ID,Location,Depression,Alcohol)               # Combine into data frame

# Convert non-Alcohol variables into factors
df$ID          <- as.factor(df$ID)
df$Location    <- as.factor(df$Location)
df$Depression  <- as.factor(df$Depression)

# Print the data frame (named 'df')
df
```

```
##    ID  Location Depression Alcohol
## 1   1      Fiji        Low     311
## 2   2      Fiji        Low     320
## 3   3      Fiji        Low     313
## 4   4 Singapore        Low     443
## 5   5 Singapore        Low     441
## 6   6 Singapore        Low     480
## 7   7  Kiribati        Low     320
## 8   8  Kiribati        Low     353
## 9   9  Kiribati        Low     313
## 10 10      Fiji       High     385
## 11 11      Fiji       High     420
## 12 12      Fiji       High     412
## 13 13 Singapore       High     557
## 14 14 Singapore       High     519
## 15 15 Singapore       High     608
## 16 16  Kiribati       High     512
## 17 17  Kiribati       High     487
## 18 18  Kiribati       High     526
```

Now we can run the model and explore the summary

```r
mod3 <- aov(data=df,formula=Alcohol~Depression*Location)
summary(mod3)
```

```
##                     Df Sum Sq Mean Sq F value   Pr(>F)    
## Depression           1  71190   71190 116.008 1.60e-07 ***
## Location             2  66535   33268  54.211 9.79e-07 ***
## Depression:Location  2   6718    3359   5.474   0.0204 *  
## Residuals           12   7364     614                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A 2×3 Type-1 ANOVA revealed a significant interaction between depression scores and participant location, $F_{2,12} = 5.47, p = .02, \eta_p^2 = ?$. We also confirmed significant main effects for depression, $F_{1,12} = 116.01, p < .001, \eta_p^2 = ?$, and location,$F_{2,12} = 54.21, p < .001, \eta_p^2 = ?$ (though the reporting of main effects is typically unnecessary when we find a significant interaction effect). We ran series of post-hoc tests to estimate which groups were significantly different from others.

Post-hoc tests

Tukey's "Honestly Significant Difference" (HSD) test is a go-to strategy for running *pairwise* contrasts across all combinations of the predictor factor levels (imagine multiple t-tests across all combinations, but controlled for **familywise error rates**)

```
TukeyHSD(mod3)   # Apply the Tukey HSD function to the compiled ANOVA model
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Alcohol ~ Depression * Location, data = df)
##
## $Depression
##               diff       lwr       upr p adj
## Low-High -125.7778 -151.2215 -100.3341 2e-07
##
## $Location
##                         diff       lwr      upr     p adj
## Kiribati-Fiji       58.33333  20.17677  96.4899 0.0040346
## Singapore-Fiji     147.83333 109.67677 185.9899 0.0000007
## Singapore-Kiribati  89.50000  51.34344 127.6566 0.0001153
##
## $`Depression:Location`
##                                    diff        lwr         upr     p adj
## Low:Fiji-High:Fiji            -91.00000 -158.93920  -23.060803 0.0073484
## High:Kiribati-High:Fiji       102.66667   34.72747  170.605864 0.0028633
## Low:Kiribati-High:Fiji        -77.00000 -144.93920   -9.060803 0.0234965
## High:Singapore-High:Fiji      155.66667   87.72747  223.605864 0.0000638
## Low:Singapore-High:Fiji        49.00000  -18.93920  116.939197 0.2225518
## High:Kiribati-Low:Fiji        193.66667  125.72747  261.605864 0.0000067
## Low:Kiribati-Low:Fiji          14.00000  -53.93920   81.939197 0.9794125
## High:Singapore-Low:Fiji       246.66667  178.72747  314.605864 0.0000005
## Low:Singapore-Low:Fiji        140.00000   72.06080  207.939197 0.0001804
## Low:Kiribati-High:Kiribati   -179.66667 -247.60586 -111.727470 0.0000148
## High:Singapore-High:Kiribati   53.00000  -14.93920  120.939197 0.1653983
## Low:Singapore-High:Kiribati   -53.66667 -121.60586   14.272530 0.1572091
## High:Singapore-Low:Kiribati   232.66667  164.72747  300.605864 0.0000009
## Low:Singapore-Low:Kiribati    126.00000   58.06080  193.939197 0.0004848
## Low:Singapore-High:Singapore -106.66667 -174.60586  -38.727470 0.0020887
```

> Tukey's HSDs confirmed individuals with low depression drink 125.8 ml *less* alcohol on average relative to individuals with high depression ($p < .001$). Individuals from Kiribati drink 58.3 ml more alcohol than Fijians ($p = .004$). Singaporeans drink on average 147.8 ml more relative to Fijians, and 89.5 ml more relative to Kiribati residents ($p's < .001$).

Remember that the goal of post-hoc tests is to identify *which* group-pairs are significantly different.

Unbalanced ANOVA

Imagine that after we collected our alcohol data, we later find out the researcher mistakenly reported the amount of alcohol *he* was drinking for one of the highly depressed Singaporean's data [$ID : 16$]. This means that the latter has to be excluded from our dataset, which is now *unbalanced* (has unequal observations across conditions)

	Fiji	Kiribati	Singapore
Low	$n = 3$	$n = 3$	$n = 3$

	Fiji	Kiribati	Singapore
High	$n = 3$	$n = 3$	$n = 2$

Let’s remove the erroneous observation from the original data and store it in a new dataframe called `df2`

```
df1 <- df[-13,]   # Remove the 13th row corresponding to the incorrect observation
df1               # Print the data
```

```
##    ID  Location Depression Alcohol
## 1   1      Fiji        Low     311
## 2   2      Fiji        Low     320
## 3   3      Fiji        Low     313
## 4   4 Singapore        Low     443
## 5   5 Singapore        Low     441
## 6   6 Singapore        Low     480
## 7   7  Kiribati        Low     320
## 8   8  Kiribati        Low     353
## 9   9  Kiribati        Low     313
## 10 10      Fiji       High     385
## 11 11      Fiji       High     420
## 12 12      Fiji       High     412
## 14 14 Singapore       High     519
## 15 15 Singapore       High     608
## 16 16  Kiribati       High     512
## 17 17  Kiribati       High     487
## 18 18  Kiribati       High     526
```

There are at least three varieties of ANOVAs that can be run. The default method in `R` , which is the one we have been using so far, is known as a **Type-1** ANOVA. This involves entering predictors in the sequence they were entered into the formula, which is generally not an issue when we have balanced designs. However, this can be problematic when designs are unbalanced.

Consider the initial model where depression was entered *before* location (the erroneous data has **not** been removed):

```
summary(aov(formula = Alcohol ~ Depression * Location, data = df))
```

```
##                     Df Sum Sq Mean Sq F value   Pr(>F)    
## Depression           1  71190   71190 116.008 1.60e-07 ***
## Location             2  66535   33268  54.211 9.79e-07 ***
## Depression:Location  2   6718    3359   5.474   0.0204 *  
## Residuals           12   7364     614                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The outcomes do not change if location was entered before depression:

```
summary(aov(formula = Alcohol ~ Location*Depression, data = df))
```

```
##                     Df Sum Sq Mean Sq F value   Pr(>F)    
## Location             2  66535   33268  54.211 9.79e-07 ***
## Depression           1  71190   71190 116.008 1.60e-07 ***
## Location:Depression  2   6718    3359   5.474   0.0204 *  
## Residuals           12   7364     614                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now let's run the models on the corrected data frame (with the erroneous observation removed)

```
summary(aov(formula = Alcohol ~ Depression * Location, data = df1))
```

```
##                     Df Sum Sq Mean Sq F value   Pr(>F)    
## Depression           1  58598   58598  87.867 1.41e-06 ***
## Location             2  61998   30999  46.482 4.31e-06 ***
## Depression:Location  2   6498    3249   4.872   0.0305 *  
## Residuals           11   7336     667                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov(formula = Alcohol ~ Location*Depression, data = df1))
```

```
##                     Df Sum Sq Mean Sq F value   Pr(>F)    
## Location             2  52039   26019  39.016 1.01e-05 ***
## Depression           1  68557   68557 102.801 6.44e-07 ***
## Location:Depression  2   6498    3249   4.872   0.0305 *  
## Residuals           11   7336     667                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

> Now when the predictors are entered in a different sequence across our unbalanced design, the F-ratios (and their associated p-values) are different. This is not overly problematic here since both sequences provide significant interactions. However, without a sufficiently justified theory, how can we know which sequence is the "correct" one?

Ordering effects during Type-1 ANOVAs mean the first predictor you enter into the model is given theoretical primacy during null hypothesis tests. Consider the following sequences of hypothesis tests.

> *Full Model*: `alcohol~Depression*Location`

$H1_0:$ `alcohol~1`
$H1_A:$ `alcohol~Depression`

The main effect for `Depression` is estimated without taking `Location` into account

$H2_0:$ `alcohol~Depression`
$H2_A:$ `alcohol~Depression*Location`

Full Model: `alcohol~Location*Depression`

$H1_0:$ `alcohol~1`
$H1_A:$ `alcohol~Location`

This time the main effect for `Location` is estimated without taking `Depression` into account

H_02: `alcohol~Location`
H_A2: `alcohol~Location*Depression`

The asymmetry becomes troublesome when sample sizes are unequal, as a significant effect might correspond with *one* sequence over the other.

We might decide to run Type-2 and Type-3 tests, which do not vary along the order of inputs to the model. Both approaches commence with the full model, and then incrementally delete predictors while noting any shifts in model performance.

However, Type-3 tests are reliant on the specific contrast patterns coded at the onset and are difficult to interpret meaningfully otherwise. This is why why we typically run Type-2 tests, which are robust to ordering effects (unlike Type-1) or contrast patterns (unlike Type-3). This allows for easier interpretation of *what* is being reported.

There are no native functions for running Type-2 ANOVAs in R, so we will require functions from external packages.

Type-2 ANOVAs operate along the **marginality principle**, which states that all lower-order terms (main effects) should be entered before higher order terms (interactions). For the full model, main effects and interactions are estimated in consideration of all variables present in the data.

Note that the full model `alcohol~Depression*Location` is short-hand for describing the main effects and interactions, so `alcoholo~Depression+Location+Depression:Location` . In a Type-2 test, the tests for main effects and interactions include the following contrasts:

For estimating the main effect of **Location**

H_0: `alcohol~Depression`
H_A: `alcohol~Depression+Location`

For estimating the main effect of **Depression**

H_0: `alcohol~Location`
H_A: `alcohol~Location+Depression`

For estimating interactions between predictors

H_0: `alcohol~Location+Depression`
H_A: `alcohol~Location+Depression+Location:Depression`

To run a Type-2 ANOVA, we will use the `Anova()` function in the `car` package

```
require(car)
mod4 <- aov(formula = Alcohol ~ Depression * Location, data = df)  # Assign the linear model to
a variable
Anova(mod4,type=2)   # Specify the type of ANOVA in the model
```

```
## Anova Table (Type II tests)
##
## Response: Alcohol
##                      Sum Sq Df  F value    Pr(>F)
## Depression            71190  1 116.0080 1.597e-07 ***
## Location              66535  2  54.2114 9.791e-07 ***
## Depression:Location    6718  2   5.4737   0.02045 *
## Residuals              7364 12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can report our outcomes ‘as is’ without worrying about the order of items entered or the specific contrast patterns across factor levels!

We can run post-hoc tests using Tukey’s test..

```
TukeyHSD(mod4)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Alcohol ~ Depression * Location, data = df)
##
## $Depression
##               diff       lwr       upr p adj
## Low-High -125.7778 -151.2215 -100.3341 2e-07
##
## $Location
##                          diff       lwr      upr     p adj
## Kiribati-Fiji       58.33333  20.17677  96.4899 0.0040346
## Singapore-Fiji     147.83333 109.67677 185.9899 0.0000007
## Singapore-Kiribati  89.50000  51.34344 127.6566 0.0001153
##
## $`Depression:Location`
##                                    diff        lwr         upr     p adj
## Low:Fiji-High:Fiji            -91.00000 -158.93920  -23.060803 0.0073484
## High:Kiribati-High:Fiji       102.66667   34.72747  170.605864 0.0028633
## Low:Kiribati-High:Fiji        -77.00000 -144.93920   -9.060803 0.0234965
## High:Singapore-High:Fiji      155.66667   87.72747  223.605864 0.0000638
## Low:Singapore-High:Fiji        49.00000  -18.93920  116.939197 0.2225518
## High:Kiribati-Low:Fiji        193.66667  125.72747  261.605864 0.0000067
## Low:Kiribati-Low:Fiji          14.00000  -53.93920   81.939197 0.9794125
## High:Singapore-Low:Fiji       246.66667  178.72747  314.605864 0.0000005
## Low:Singapore-Low:Fiji        140.00000   72.06080  207.939197 0.0001804
## Low:Kiribati-High:Kiribati   -179.66667 -247.60586 -111.727470 0.0000148
## High:Singapore-High:Kiribati   53.00000  -14.93920  120.939197 0.1653983
## Low:Singapore-High:Kiribati   -53.66667 -121.60586   14.272530 0.1572091
## High:Singapore-Low:Kiribati   232.66667  164.72747  300.605864 0.0000009
## Low:Singapore-Low:Kiribati    126.00000   58.06080  193.939197 0.0004848
## Low:Singapore-High:Singapore -106.66667 -174.60586  -38.727470 0.0020887
```

The results remain resemble earlier post-hoc tests even after omitting the researcher's drinking record.

Lab Activity

1. Return to the ANOVA outputs in the previous week's slides (p. 8), where we explored whether the factors *Location* (Fiji, Singapore) and *Depression* (Low, High) significantly explained alcohol consumption. Calculate **three** effect sizes (η_p^2) for the interactions and main effects (remember that $\eta_p^2 = \frac{SS_{Factor}}{SS_{Factor}+SS_{Residuals}}$).
2. Across the post-hoc analyses reported in the previous page, report all location~depression level contrasts that were significant across **interactions between Fiji and Kiribati only**. Do not report on any contrasts involving Singaporeans. For example,*highly depressed Fijians drank 91 ml more alcohol on average relative to low depressed Fijians ($p = .007$).*
3. A Type-2 ANOVA on the dataset `df` was run earlier after the 13th observation had been manually removed (see p. of the current document). Return to that data frame, omit the 3rd and 10th observations and assign the remaining values to a new data frame. Then, using the `car` package, run a Type-2 ANOVA and report

whether any significant interactions and/or main effects were found. Run post-hoc tests if any factor levels (e.g., low vs high depression) significantly varied.

This is the final statistics lab for the semester.