Factorial ANOVA

Jonathan A. Pedroza, PhD

Factorial Design

- factorial designs are often also referred to as two-way ANOVA
- JP: I like to draw out my factorial ANOVAs because it also helps with understanding means
  - a two-way ANOVA has 2 IVs as well as an interaction (we'll talk about this shortly)
    - they can often be written out as the number of variables and the number of levels/conditions per variable
    - Ex: penguins (Adelie, Gentoo, and Chinstrap) and sex (Male and Female) would be a 3 x 2 (Three by Two) Factorial Design
    - Ex: Intervention (Control, Physical Activity, Physical Activity + Eating Habits) and method of delivery (Online and In-person) would be __________
    - Ex: Intervention (Control and Treatment) and Age (Young Adult, Adult, Older Adult) and Sex (Male and Female) would be _________
  - an ANCOVA can have two variables, with one being a covariate, but will not include an interaction

Factorial Design

- **independent factorial design** has several IVs/predictors comparing different groups
  - between-subjects design
- **repeated-measures factorial design** has several IVs/predictors that have been measured where all participants receive all conditions
  - within-subjects design
- **mixed design** has several IVs/predictors with some measuring all participants over all conditions while also examining different groups
  - Ex: pretest-posttest while comparing whether or not participants meet a MDD diagnosis
  - one IV is between-subjects and the other IV is a within-subjects variable

Independent Factorial Design

- the best way of deciphering ANOVAs are by what precedes the word ANOVA
  - **one**-way ANOVA
  - **two**-way independent ANOVA

Independent Factorial Design

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + ... + b_n X_{ni} + \epsilon_i$$

- let's talk cars
  - variables = number of cylinders and transmission type
    - levels = 4 vs 6 cylinder (we'll leave out 8 cylinder for the example)
    - levels = manual vs automatic

Independent Factorial Design

- SPSS does not require dummy coding for ANOVA but you can run the same test in a linear regression with dummy coded variables and get the same estimates
  - we'll have reference groups for each variable

$$MPG_i = b_0 + b_1 Cylinder_{1i} + b_2 Transmission_{2i} + \epsilon_i$$

- however, the model above is not a two-way ANOVA
  - it lacks one core component (right now it's only an ANCOVA)

Independent Factorial Design

$$MPG_i = b_0 + b_1 Cylinder_{1i} + b_2 Transmission_{2i} + b_3 Interaction_i + \epsilon_i$$

$$MPG_i = b_0 + b_1 Cylinder_{1i} + b_2 Transmission_{2i} +$$

$$b_3 Cylinder * Transmission_i + \epsilon_i$$

- the **interaction** term is the combined (multiplied) effect of cylinder and transmission type on MPG

Independent Factorial Design

- let's now look at the *b* values

```
               mpg cyl am
Mazda RX4      21.0   1  1
Mazda RX4 Wag  21.0   1  1
Datsun 710     22.8   0  1
Hornet 4 Drive 21.4   1  0
Valiant        18.1   1  0
Merc 240D      24.4   0  0
Merc 230       22.8   0  0
Merc 280       19.2   1  0
Merc 280C      17.8   1  0
Fiat 128       32.4   0  1
Honda Civic    30.4   0  1
Toyota Corolla 33.9   0  1
Toyota Corona  21.5   0  0
Fiat X1-9      27.3   0  1
Porsche 914-2  26.0   0  1
Lotus Europa   30.4   0  1
```

Independent Factorial Design

- combinations we'll have are:
  - 4-cylinder (coded = 0) and automatic (coded = 0)
    - Average MPG = 22.9, SD = 1.45
  - 4-cylinder (coded = 0) and manual (coded = 1)
    - Average MPG = 28.1, SD = 4.48
  - 6-cylinder (coded = 1) and automatic (coded = 0)
    - Average MPG = 19.1, SD = 1.63
  - 6-cylinder (coded = 1) and manual (coded = 1)
    - Average MPG = 20.6, SD = 0.75

Independent Factorial Design

$$MPG_i = b_0 + b_1 Cylinder_{1i} + b_2 Transmission_{2i} +$$

$$b_3 Cylinder * Transmission_i + \epsilon_i$$

$$X_{4,automatic} = b_0 + (b_1 * 0) + (b_2 * 0) + (b_3 * 0) + \epsilon_i$$

$$b_0 = X_{4,automatic}$$

$$b_0 = 22.9$$

- b_0 will be the average MPG for cars that are a **4-cylinder** car and have an **automatic** transmission

Independent Factorial Design

- now we can look at when we have a 6-cylinder car

$$X_{6,automatic} = b_0 + (b_1 * 1) + (b_2 * 0) + (b_3 * 0) + \epsilon_i$$

$$= b_0 + b_1$$

Independent Factorial Design

$$= X_{4,automatic} + b_1$$

$$b_1 = X_{6,automatic} - X_{4,automatic}$$

$$b_1 = 19.1 - 22.9$$

$$b_1 = -3.8$$

Independent Factorial Design

- b_1 is the average difference between a for **6-cylinder** cars that have an **automatic** transmission and **4-cylinder** cars that are **automatic**

Independent Factorial Design

- Let's now move on to $b2$

$$X_{4,manual} = b_0 + (b_1 * 0) + (b_2 * 1) + (b_3 * 0) + \epsilon_i$$

$$= b_0 + b_2$$

Independent Factorial Design

$$= X_{4,automatic} + b_2$$

$$b_2 = X_{4,manual} - X_{4,automatic}$$

$$b_2 = 28.1 - 22.9$$

$$b_2 = 5.2$$

Independent Factorial Design

- b_2 is the average difference between **4-cylinder** cars that are **manual** and **4-cylinder** cars that are **automatic**

Independent Factorial Design

- Finally, let's focus on the **interaction,** or the multiplied terms of cylinder and transmission

$$X_{6,manual} = b_0 + (b_1 * 1) + (b_2 * 1) + (b_3 * 1) + \epsilon_i$$

$$= b_0 + b_1 + b_2 + b_3$$

Independent Factorial Design

$$= X_{4,automatic} + (X_{6,automatic} - X_{4,automatic}) + (X_{4,manual} - X_{4,automat}$$

$$= X_{6,automatic} + X_{4,manual} - X_{4,automatic}$$

$$b_3 = X_{4,automatic} - X_{6,automatic} + X_{6,manual} - X_{4,manual}$$

Independent Factorial Design

$$b_3 = 22.9 - 19.1 + 20.6 - 28.1$$

$$b_3 = 3.8 + (-7.5)$$

$$b_3 = -3.7$$

Independent Factorial Design

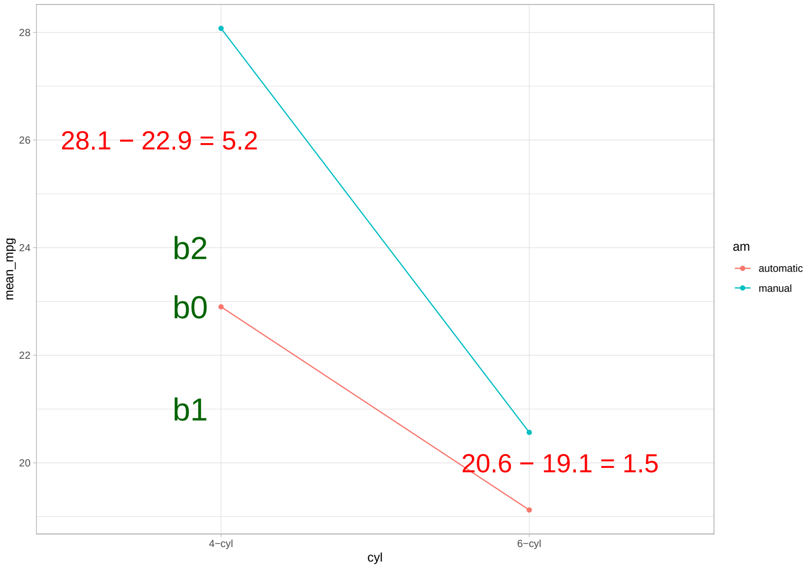
```
Call:
lm(formula = mpg ~ as.factor(cyl) * as.factor(am), data = s

Residuals:
    Min      1Q  Median      3Q     Max 
-6.6750 -1.2500 -0.0125  2.0812  5.8250 

Coefficients:
                              Estimate Std. Error t value
(Intercept)                     22.900      1.915  11.956
as.factor(cyl)1                 -3.775      2.534  -1.490
as.factor(am)1                   5.175      2.246   2.304
as.factor(cyl)1:as.factor(am)1  -3.733      3.386  -1.103
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

Residual standard error: 3.317 on 14 degrees of freedom
Multiple R-squared:  0.634, Adjusted R-squared:  0.5556
```

Independent Factorial Design

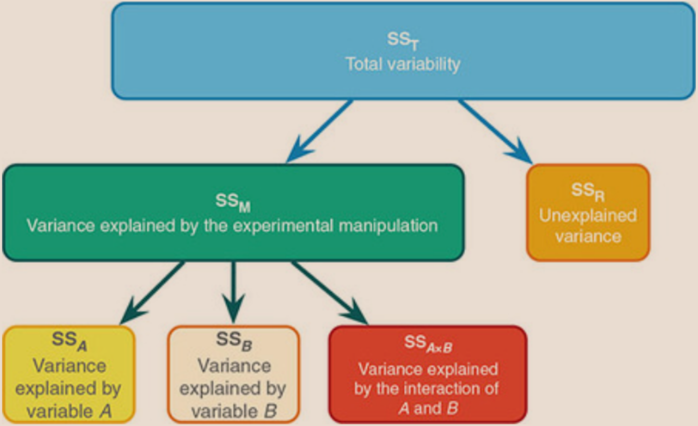


Independent Factorial Design

- the interaction shows us the difference between automatic and manual transmissions when comparing 4-cylinder and 6-cylinder cars in their average MPG
- from the findings, we can conclude
  - a difference between 4-cylinder cars and 6-cylinder cars in average MPG?
  - a difference between automatic and manual transmission cars in average MPG?
  - a difference in average MPG between 4-cylinder and 6-cylinder cars is dependent on whether or not a car has an automatic or manual transmission?

Behind the Scenes of Factorial Designs

Figure 14.3 Breaking down the variance in a two-way factorial design



Behind the Scenes of Factorial Designs

- calculations are very similar to a one-way ANOVA
  - model sum of squares is now broken up into
    - what is explained from our first IV
    - what is explained from our second IV
    - what is explained by the interaction between IV_1 and IV_2

Total Sum of Squares

$$SS_T = \sum_{i=1}^{N}(x_i - \overline{X}_{grand})^2$$

$$SS_T = S^2_{grand}(N-1)$$

- N is still the number of participants in our sample

Model Sum of Squares

$$SS_M = \sum_{g=1}^{k} n_g(\overline{X}_g - \overline{X}_{grand})^2$$

- we are still looking at what the variance our model explains in our outcome
- using the car data from before, we could find the model sum of squares by using the 2 levels of cylinder (4-cyl and 6-cyl) and the 2 levels of transmission (automatic and manual)
  - the mean of each group and subtract the grand mean from it
- degrees of freedom for the model would still follow the same formula $df_M = k - 1$, so since we have 2 levels for both variables we would have 4 different groups

Main Effect of IV_1

$$SS_A = \sum_{g=1}^{k} n_g(\overline{X}_g - \overline{X}_{grand})^2$$

- common letters for the main effects are A and B with the interaction being $A * B$
- if our first variable is cylinder, then we would use the means for all cars that are a 4-cylinder and all cars that are a 6-cylinder **independent** of whether these cars have a manual or automatic transmission
  - only focusing on cylinder types for the main effect of cylinder on MPG
- df for this main effect would still be $k - 1$, but now we are only focusing on our two cylinder groups so it would now be $2 - 1 = 1$

Main Effect of IV_2

$$SS_B = \sum_{g=1}^{k} n_g(\overline{X}_g - \overline{X}_{grand})^2$$

- we'll now do the same thing for transmission types (automatic and manual) **independent** of how many cylinders are car has

Interaction of $IV_1 * IV_2$

- the easiest way of calculating the sum of squares for the interaction is by using the information we already have

$$SS_{A*B} = SS_M - SS_A - SS_B$$

- the degrees of freedom could be calculated the same way

$$df_{A*B} = df_M - df_A - df_B$$

Residual Sum of Squares

- we can calculate the residual sum of squares the same way we did when conducting a one-way ANOVA

$$SS_R = SS_T - SS_M$$

- remember that at this point, we already calculated the SS_T and the SS_M so we can use this formula to get the variance not explained by our model

$$SS_R = \sum_{g=1}^{k} S_g^2(n_g - 1)$$

- our you can use this formula to get the variance of each group and add them all together

The F-statistic

- the only difference between a factorial ANOVA and a one-way ANOVA is that, we will now calculate three different mean squares values for the model and the residual mean squares

$$MS_A = \frac{SS_A}{df_A}$$

$$MS_B = \frac{SS_B}{df_B}$$

$$MS_{A*B} = \frac{SS_A * B}{df_A * B}$$

$$MS_R = \frac{SS_R}{df_R}$$

The F-statistic

- we will also calculate F statistics for both main effects and the interaction
  - since these F tests are all signal-to-noise ratios, we will use the residual mean squares for the noise of each test

$$F_A = \frac{MS_A}{MS_R}$$

$$F_B = \frac{MS_B}{MS_R}$$

$$F_{A*B} = \frac{MS_{A*B}}{MS_R}$$

The F-statistic

```
                             Df Sum Sq Mean Sq F value   Pr
as.factor(cyl)                1 204.89  204.89  18.618 0.00
as.factor(am)                 1  48.61   48.61   4.417 0.05
as.factor(cyl):as.factor(am)  1  13.38   13.38   1.216 0.28
Residuals                    14 154.07   11.00
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

# A tibble: 4 x 4
  cyl      am mean_mpg sd_mpg
  <chr> <dbl>    <dbl>  <dbl>
1 0         0     22.9  1.45
2 0         1     28.1  4.48
3 1         0     19.1  1.63
4 1         1     20.6  0.751
```

Model Assumptions

- if you violate the assumption of homogeneity of variance (homoscedasticity), SPSS will struggle to correct for the assumption with anything more than a $2x2$ design
  - one way to get around this is to bootstrap your post-hoc tests
- we'll look at the residuals and predicted values
- we'll test for normality

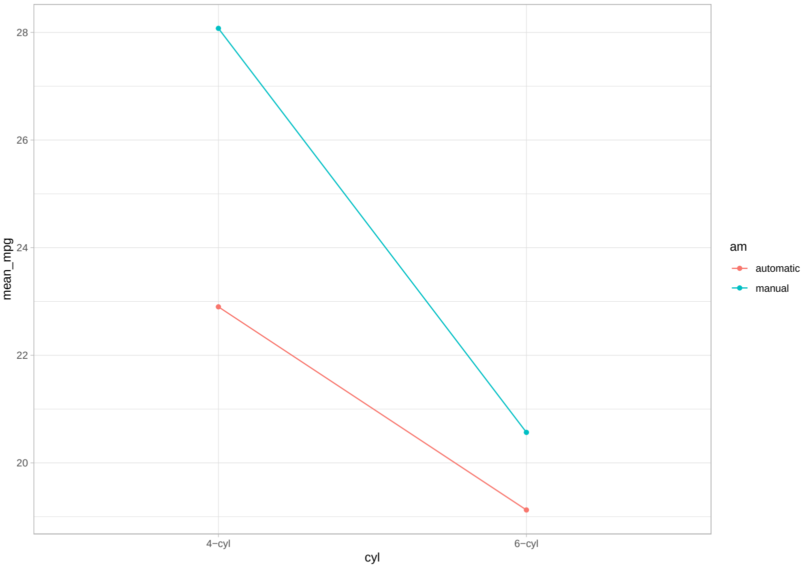
Simple Effects Analysis

- one way to break down interactions is through the use of **simple effects analysis**
  - examines the effect/relationship of one IV at individual levels of the other IV
  - Ex: cylinder –> MPG for automatic cars; cylinder –> MPG for manual cars
- we'll talk about how to conduct simple effect analyses in SPSS syntax
  - which we'll talk about today

Interpreting Interaction Graphs

- "a picture is worth a thousand words"
- if a significant interaction is found, then we would state that the relationship between cylinder and MPG is dependent on what type of transmission a car has

Interpreting Interaction Graphs



Interpreting Interaction Graphs

- non-parallel lines indicate some degree of an interaction
  - Note: this does not mean non-parallel lines always show a statistically significant interaction
- interaction plots that cross are non-parallel and could indicate a possible statistically significant interaction
  - Note: the visual/plot can tell us that there may be a significant interaction; however, only our F-test for the interaction will tell us if the interaction is significant or not

Interpreting Interaction Graphs

- SPSS allows for bar plots or line plots
  - both are not the prettiest but line plots are easier to read (for me personally)
  - you can also include error bars (which is best practice) but for our assignments i will not because it clutters SPSS output

```
                                Df Sum Sq Mean Sq F value   Pr
as.factor(cyl)                   1 204.89  204.89  18.618 0.00
as.factor(am)                    1  48.61   48.61   4.417 0.05
as.factor(cyl):as.factor(am)     1  13.38   13.38   1.216 0.28
Residuals                       14 154.07   11.00
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```