PSYC*6060: 1-way ANOVA

1 ANOVA: Ensure R Results Match SPSS Results

The omnibus hypothesis in single-factor analysis of variance is the hypothesis that means of the groups are all equal. The Field et al book provides one method for conducing a single-factor analysis but I will present another approach here. The reason for this is that the Field et al book uses different procedures for the single-factor and multi-factor analysis of variance. I will use an approach that is similar regardless of the number of factors (note the contrasts will differ though). The advantage of using this process is the consistency across different situations makes it easier to remember.

The reason for this slightly different procedure relates to the fact that they way we typically calculate multi-factor ANOVA values in Psychology (and specifically in SPSS) is very controversial. In fact, the process is so controversial that the creators of R refuse to integrate into base R the same method for calculating ANOVA values that SPSS uses since they believe it is incorrect (i.e., they believe Type III Sums of Squares are incorrect). The capacity to create ANOVA values the same as SPSS in R is not problematic because this functionality has been added in a variety of different R packages. Consequently, it is relatively easy to get Type III Sums of Squares in R. In particular, the car package makes it easy to get the same values as SPSS based on Type III Sums of Squares (car stands for Companion to Applied Regression [a text book]).

To ensure the values you get in R match those from SPSS you need to do three things:

I. Each independent variable is a factor in R (recall the as.factor function)

II. Set the contrasts correctly (critical in multi-factor ANOVA designs)

III. Use Type III Sums of Squares (typically using the car package Anova function)

When all three of these steps are done correctly, the ANOVA values in R will match those from SPSS.

2 Analysis Example

Here is an example of how you would proceed with an one-way ANOVA.

2.1 Load the data

```
library(tidyverse)
oneway.data <- read_csv("Viagra.csv")
```

2.2 I Set the factors correctly

```
oneway.data$dose <- as.factor(oneway.data$dose)
levels(oneway.data$dose) <- list("Placebo"=1, "Low Dose"=2, "High Dose"=3)
```

2.3 Display Descriptive Statistics

```
psych::describeBy(oneway.data$libido, group=oneway.data$dose)
```

```
## group: Placebo
##   vars n mean  sd median trimmed  mad min max range skew kurtosis   se
## 1    1 5  2.2 1.3      2     2.2 1.48   1   4     3 0.26    -1.96 0.58
## ------------------------------------------------------------
## group: Low Dose
##   vars n mean  sd median trimmed  mad min max range skew kurtosis   se
## 1    1 5  3.2 1.3      3     3.2 1.48   2   5     3 0.26    -1.96 0.58
## ------------------------------------------------------------
## group: High Dose
##   vars n mean   sd median trimmed  mad min max range skew kurtosis   se
## 1    1 5    5 1.58      5       5 1.48   3   7     4    0    -1.91 0.71
```

2.4 Test Homogeneity of Variance Assumption

We need to test the assumption that the variances of the groups are all the same (at the population level). If this assumption is true (i.e., Levene's test is non-significant), we can proceed in the normal fashion for 1-way ANOVA. If the assumption is false, (i.e., Levene's test is significant) we need to use a somewhat different approach. Note that for results to match SPSS you must use *median* centering.

```
car::leveneTest(oneway.data$libido, group=oneway.data$dose,center="median")
```

```
## Levene's Test for Homogeneity of Variance (center = "median")
##       Df F value Pr(>F)
## group  2  0.1176   0.89
##       12
```

This test reveals that the homogeneity of variance assumption was not violated, $F(2,12) = 0.12$, $p = .89$.

2.5 Main Analysis: Equal Variances Assumed (i.e., Levene's non-significant)

2.5.1 II Set the contrasts correctly

```
options(contrasts = c("contr.sum", "contr.poly"))
```

2.5.2 Run the ANOVA

```
oneway.results <- lm(libido~dose,data=oneway.data) # Conduct ANOVA as GLM
```

This output reveals $F(2,12) = 5.12$, $p = .03$. But it's easier if we use apaTables - we get effect sizes with confidence intervals.

2.5.3 III Use Type 3 Sums of Squares

```
car::Anova(oneway.results,type=3)
```

```
## Anova Table (Type III tests)
##
## Response: libido
##              Sum Sq Df F value    Pr(>F)
## (Intercept) 180.267  1 91.6610 5.721e-07 ***
## dose         20.133  2  5.1186   0.02469 *
## Residuals    23.600 12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.5.4 APA ANOVA Table

The *apaTables* command apa.aov.table automatically uses Type 3 sums of squares (so you don't need to do the above step). As well, it also provides effect sizes with confidence intervals. Notice that the confidence intervals are 90% not 95%; this is convention when reporting effect sizes for ANOVA. This practices ensure the *significance* of an effect is roughly consistent with the confidence interval reported.

```
library(apaTables)
apa.aov.table(oneway.results,
              table.number = 1,
              filename = "Table1.doc")
```

```
##
##
## Table 1
##
## ANOVA results using libido as the dependent variable
##
##
##    Predictor     SS df     MS     F    p partial_eta2 CI_90_partial_eta2
##  (Intercept) 180.27  1 180.27 91.66 .000
##         dose  20.13  2  10.06  5.12 .025          .46         [.04, .62]
##        Error  23.60 12   1.97
##
## Note: Values in square brackets indicate the bounds of the 90% confidence interval for partial eta-s
```

2.5.5 APA Mean/SD Table

You can make an APA table of means with the commands below:

```
library(apaTables)
apa.1way.table(iv=dose, dv=libido,
               data=oneway.data,
               table.number = 2,
               filename = "Table2.doc",
               show.conf.interval = TRUE)
```

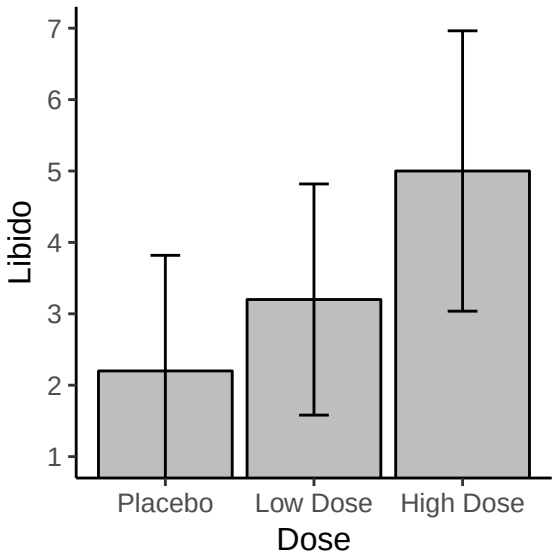
```
##
##
## Table 2
##
## Descriptive statistics for libido as a function of dose.
##
##       dose    M   LL   UL   SD
##    Placebo 2.20 0.58 3.82 1.30
##   Low Dose 3.20 1.58 4.82 1.30
##  High Dose 5.00 3.04 6.96 1.58
##
## Note. M and SD represent mean and standard deviation, respectively.
## LL and UL indicate the lower and upper limits of the 95% confidence interval
## for the mean, respectively.
## The confidence interval is a plausible range of population means that could
## have caused a sample mean (Cumming, 2014).
```

You will also want to read about the apa.d.table command in the posthoc section below. This commands will create a table of d-values for all paired-comparisons (along with CI's).

2.5.6 Graphing

Note that in this graph we use "whiskers" to indicate the confidence interval. This is the recommended approach because confidence intervals provide substantial meaningful information (i.e., plausible range of population values). Some people prefer to use whiskers to indicate standard error of the mean; however, this approach is discouraged because it is not as informative as the confidence interval.

```
myBar<-ggplot(oneway.data,aes(dose,libido)) #order matters the IV must be first
myBar<-myBar + stat_summary(fun.y=mean, geom="bar", fill="Grey", colour="Black")
myBar<-myBar + stat_summary(fun.data=mean_cl_normal, geom="errorbar",width=0.2)
myBar<-myBar + coord_cartesian(ylim=c(1, 7))
myBar<-myBar + scale_y_continuous(breaks=seq(1,7))
myBar<-myBar + labs(x="Dose",y="Libido")
myBar<-myBar + theme_classic(12)
print(myBar)
```



2.5.7 Summary

I found that drug (placebo, low dose, high dose) had an effect on libido, $F(2,12) = 5.12$, $p = .03$, $\eta^2 = .46$, 90% CI [.04, .62] (see Tables 1 and 2). The confidence interval has a low lower-bound and is quite wide indicating that the data are sufficient to suggest a non-zero effect but the sample size is insufficient to meaningfully estimate the size of the effect (it could be anywhere from small/medium to extraordinarily large). (Cohen indicated that partial eta-squared values of .01, 06, and.14 correspond to small, medium, and large, respectively).

You would want to supplement this paragraph with your *a priori* contrast results or *post hoc* parried comparison results.

2.6 Main Analysis: Equal Variances Not Assumed (i.e., Levene's significant)

If the Levene's test was significant, than you can use this procedure instead of the GLM approach. But you need to create your APA ANOVA table by hand.

```
options(contrasts = c("contr.sum", "contr.poly"))
oneway.test(libido~dose,data=oneway.data)
```

3 a priori Contrasts

If you have a specific hypothesis about the direction of the difference between two group means before running the experiment, an *a priori* test is appropriate. See pages 419 to 426 for a detailed review of contrasts in Field et al. In particular, see the description of the contrasts used here in Table 10.4 on p. 421. Don't forget to divide the p-values obtained in these analyses by 2 to reflect the fact they are one-sided tests. (If you don't know which mean will be larger, you should be doing a *post hoc* test).

```
#define the contrasts
placebo.vs.low.and.high.dose=c(-2,1,1)
low.dose.vs.high.dose=c(0,-1,1)
my.contrasts <- cbind(placebo.vs.low.and.high.dose,low.dose.vs.high.dose)

#set the contrasts
contrasts(oneway.data$dose)<-my.contrasts

#conduct the regression with specified contrasts
planned.contrasts.results<-lm(libido~dose,data=oneway.data)

#inspect the results to see the contrast
summary(planned.contrasts.results, type=3)
```

```
##
## Call:
## lm(formula = libido ~ dose, data = oneway.data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -2.0   -1.2   -0.2    0.9    2.0
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       3.4667     0.3621   9.574 5.72e-07 ***
## doseplacebo.vs.low.and.high.dose  0.6333     0.2560   2.474   0.0293 *
## doselow.dose.vs.high.dose         0.9000     0.4435   2.029   0.0652 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.402 on 12 degrees of freedom
## Multiple R-squared:  0.4604, Adjusted R-squared:  0.3704
## F-statistic: 5.119 on 2 and 12 DF,  p-value: 0.02469
```

3.1 Confidence intervals for contrasts

If you want to calculate the confidence interval for a contrast-based comparison use the ci.c command in the MBESS package. There are excellent examples of how to use the command in the help file. Just type the command below to view the help file (after installing MBESS).

```
library(MBESS)
?ci.c
```

3.2 Power analysis for contrasts

There are a number commands for sample size analysis for contrasts. Use the command below to bring up the help file for MBESS. Scroll to the bottom of this page. Click the *index* link in blue. Then scroll down until you see all of the commands that begin: ss.aipe

```
library(MBESS)
?MBESS
```

You may begin by inspecting the general contrast sample size analysis command below:

```
?ss.aipe.c
```

4 post hoc Paired Comparisons

When you don't have specific hypotheses based on past research about the direction of particular paired comparisons then you call those comparisons post hoc comparisons. That is, in light of a significant omnibus test, which indicates the group means are not all the same, you conduct after the fact (i.e., post hoc) comparisons to determine which means are different. To do this, you examine every possible paired comparison among the groups. Note that these comparisons are all two-sided tests (since they are post hoc).

A consequence of examining every possible paired comparison is that you will have a higher overall chance of making a Type I Error (i.e., concluding a difference exists when it does not).

The procedure below conducts every possible paired comparison but does not worry about the inflated Type I Error rate.

```
pairwise.t.test(oneway.data$libido,oneway.data$dose,p.adjust.method="none")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  oneway.data$libido and oneway.data$dose
##
##           Placebo Low Dose
## Low Dose  0.2816  -
## High Dose 0.0083  0.0652
##
## P value adjustment method: none
```

4.1 Controlling Type I Error: Bonferroni Correction

One option for controlling Type I Errors is to use the Bonferroni correction. Typically, with this correction you divide the criteria for determining significance (i.e., alpha = .05) by the number of tests (3 in this case) to get a new criteria for determining significance. That is .05/3=.017. With this approach, you only consider a difference significant if the p-value is less than .017 instead of .05. R uses a different approach.

R uses a mathematically equivalent version of this test. Instead of changing the criterion for significance (i.e., is $p < .05$), R leaves the criteria the same and multiplies the p-value by the number of possible comparisons. In a single-factor ANOVA where the independent variable has 3 levels there are 3 possible comparisons, so the p-value is multiplied by 3. With this output, you simply inspect the p-values and indicate see if any are less than .05 (as usual).

```
pairwise.t.test(oneway.data$libido,oneway.data$dose,p.adjust.method="bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  oneway.data$libido and oneway.data$dose
##
##           Placebo Low Dose
## Low Dose  0.845   -
## High Dose 0.025   0.196
##
## P value adjustment method: bonferroni
```

Notice with this Bonferroni correction the p-values are multiplied by 3 when compared to pairwise.t.test command used above. That is the evaluation criteria of significance stays the same (i.e., alpha = .05) and the p-values are increased to reflect the fact multiple comparisons have been conducted.

Instead of "bonferroni" you could also have used "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", or "fdr". Field et al. provide a discussion of these different options.

4.2 Controlling Type I Error: Tukey

There are many approaches to controlling Type I Error with paired comparisons. Bonferroni is one approach another is the Tukey HSD approach. Note, however, that there are many approaches to controlling Type I Error beyond these two approaches – I encourage you to learn more about all the different approaches. The Tukey approach is illustrated below.

```
library(multcomp)
posthocs<-glht(oneway.results,linfct=mcp(dose="Tukey"))
summary(posthocs)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = libido ~ dose, data = oneway.data)
##
## Linear Hypotheses:
##                            Estimate Std. Error t value Pr(>|t|)
## Low Dose - Placebo == 0     1.0000     0.8869   1.127    0.516
## High Dose - Placebo == 0    2.8000     0.8869   3.157    0.021 *
## High Dose - Low Dose == 0   1.8000     0.8869   2.029    0.148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Oddly some of the statistics reported in this output are slightly wrong (but consistent with how SPSS reports them) even though the p-values are calculated correctly. I suggest using this output to determine if a particular comparison is significant by looking at the p-value but not reporting the t-value or q-value. Just report the p-value and provide an estimate of the effect size (see below).

4.3 Paired comparison effect sizes

4.3.1 Easy way apaTables (Type I Error not Controlled)

You can use the apaTable package to quickly calculate all paired comparison d-values. It also displays the mean and SD for each cell.

```
library(apaTables)
apa.d.table(iv = dose, dv = libido,
            data = oneway.data,
            filename="dValueTable.doc",
            show.conf.interval = TRUE)
```

```
##
##
## Means, standard deviations, and d-values with confidence intervals
##
##
##   Variable     M    SD   1             2
##   1. Placebo   2.20 1.30
##
##   2. Low Dose  3.20 1.30 0.77
##                           [-0.55, 2.04]
##
##   3. High Dose 5.00 1.58 1.93*         1.24
##                           [0.34, 3.44]  [-0.17, 2.59]
##
##
## Note. * indicates p < .05; ** indicates p < .01.
## M and SD are used to represent mean and standard deviation, respectively.
## Values in square brackets indicate the 95% confidence interval for each d-value.
## The confidence interval is a plausible range of population d-values
## that could have caused the sample d-value (Cumming, 2014).
## d-values are unbiased estimates calculated using formulas 4.18 and 4.19
## from Borenstein, Hedges, Higgins, & Rothstein (2009).
## d-values not calculated if unequal variances prevented pooling.
##
```

4.3.2 Old-school way

We begin by getting a subset of the data corresponding to each group.

```
placebo.rows   <- oneway.data %>% filter(dose=="Placebo")
low.dose.rows  <- oneway.data %>% filter(dose=="Low Dose")
high.dose.rows <- oneway.data %>% filter(dose=="High Dose")
```

Comparison 1 Placebo vs Low Dose (smd is the d-value)

```
library(MBESS)

#Notice how we still have to select the column libido (i.e., DV column)
#using the dollar sign in the commands below.
d.value <- smd(Group.1=placebo.rows$libido,
               Group.2=low.dose.rows$libido)
```

```
ci.smd(smd=d.value,n.1=length(placebo.rows$libido),
       n.2=length(low.dose.rows$libido))
```

```
## $Lower.Conf.Limit.smd
## [1] -2.038896
##
## $smd
## [1] -0.766965
##
## $Upper.Conf.Limit.smd
## [1] 0.5481115
```

Comparison 2 Placebo vs High Dose (smd is the *d*-value)

```
d.value <- smd(Group.1=placebo.rows$libido,
               Group.2=high.dose.rows$libido)
ci.smd(smd=d.value,n.1=length(placebo.rows$libido),
       n.2=length(high.dose.rows$libido))
```

```
## $Lower.Conf.Limit.smd
## [1] -3.444831
##
## $smd
## [1] -1.932184
##
## $Upper.Conf.Limit.smd
## [1] -0.3449147
```

Comparison 3 Low Dose vs High Dose (smd is the *d*-value)

```
d.value <- smd(Group.1=low.dose.rows$libido,
               Group.2=high.dose.rows$libido)
ci.smd(smd=d.value,n.1=length(low.dose.rows$libido),
       n.2=length(high.dose.rows$libido))
```

```
## $Lower.Conf.Limit.smd
## [1] -2.587867
##
## $smd
## [1] -1.242118
##
## $Upper.Conf.Limit.smd
## [1] 0.1651003
```