HUDM 5123 - Linear Models and Experimental Design
Lab 03 - One-Way ANOVA

1 The Data

I found the data set we will work with today while browsing through the *Journal of Open Psychology Data*. The data correspond with an attempt to replicate the results of a landmark randomized controlled trial by Seligman, Steen, Park, and Peterson (2005) that found that positive psychology interventions, delivered via the internet, could increase happiness and decrease depression relative to a control group.

The replication study by Woodworth, O'Brien-Malone, Diamond, and Schüz (2017), published in the *Journal of Clinical Psychology*, enrolled 295 participants online and directed them to a website to complete the study. Participants were given a baseline assessment in depression using the Center for Epidemiological Studies, Depression Scale (CES-D) and a baseline assessment in happiness using the Authentic Happiness Index (AHI), prior to any interventions (see the Woodworth paper for details and citations). Participants were then subsequently randomly assigned to one of four intervention groups (quoted descriptions are from Woodworth et al, 2017, pp. 221-222):

1. SS; "Using signature strengths. Participants were e-mailed a special link to a page on the study website (hidden from participants who were not in the using signature strengths condition) where they completed a questionnaire that gave them feedback about their top five character strengths. Participants were instructed to use one of their five signature strengths in a new and different way for a week."

2. TGT; "Three good things in life. Participants were instructed that for a week, they were to write down three good things that happened each day, together with a causal explanation for each thing."

3. GV; "Gratitude visit. Participants were instructed to write and deliver in person a letter of appreciation to someone who had been kind to them but whom they had never properly thanked."

4. EM; "Early memories: placebo control. Participants were instructed to write about their early memories each night for a week."

The intervention period lasted for a week. After the intervention period ended, participants were assessed five more times: an immediate posttest directly after the intervention, 1 week after posttest, 1 month after posttest, 3 months after posttest, and 6 months after posttest.

Research Question 1: Is there an overall treatment effect?

$$\begin{cases} H_0: & \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_1: & \text{At least one of the means differs from the rest.} \end{cases}$$

Research Question 2: Do any of the positive psychology interventions increase happiness or decrease depression more than the control?

$$\text{Set 1: } \begin{cases} H_{0_1}: & \mu_1 <= \mu_4 \\ H_{1_1}: & \mu_1 > \mu_4 \end{cases} \quad \text{Set 2: } \begin{cases} H_{0_2}: & \mu_2 <= \mu_4 \\ H_{1_2}: & \mu_2 > \mu_4 \end{cases} \quad \text{Set 3: } \begin{cases} H_{0_3}: & \mu_3 <= \mu_4 \\ H_{1_3}: & \mu_3 > \mu_4 \end{cases}$$

Research Question 3: Does the average of the three positive psychology interventions increase happiness or decrease depression more than the control group?

$$\begin{cases} H_0: & \frac{\mu_1 + \mu_2 + \mu_3}{3} <= \mu_4 \\ H_1: & \frac{\mu_1 + \mu_2 + \mu_3}{3} > \mu_4 \end{cases}$$

For lab today we will test RQs 1 & 2. We will test RQ 3 after discussing linear contrasts next week. One further complicating aspect of the data is that participants were assessed longitudinally a total of six times over as many months. We will be able to handle the longitudinal data structure more appropriately after we cover such methods. For now, however, we will need to examine the data at a single time point. **Let's look for treatment effects at the one-month follow-up.** One "trick" that can be used to make ANOVA work with two timepoints is to take differences (posttest - pretest). These differences are called *gain scores*.

2 Research Question 1

Read in the data from the csv file directly with

```
dat <- read.csv(file = "FILL IN YOUR PATH HERE")
```

or by browsing with

```
dat <- read.csv(file = file.choose())
```

Note that the data set contains all of the individual item scores for the two inventories. We will only use the total scores, which are at the end of the column list. Also note that the id variable identifies each participant and that most participants are associated with multiple rows that correspond with repeated follow-up tests. These data are in what is referred to as *long format*. That is, each row corresponds with a measurement occasion. To take differences to make gain scores, we would like the data to be in *wide format*, where each participant is associated with only one row of data. First, create a subset, called `rct`, that only contains the variables we will be working with directly:

```
rct <- subset(x = rct,
              select = c("id", "occasion", "intervention",
                         "ahiTotal", "cesdTotal"))
```

Add a factor to the data set based on the intervention codes 1 through 4.

```
rct$int_fact <- factor(x = rct$intervention,
                       levels = 1:4,
                       labels = c("SS", "TGT", "GV", "REM"))
```

Convert the data from long to wide format using `reshape()`. `idvar` is the partcipant identifier. `timevar` is the variable in long format that differentiates multiple records from the same participant. `v.names` gives the names of variables that should be spread out in wide format.

```
rct_wide <- reshape(data = rct,
                    idvar = "id",
                    timevar = "occasion",
                    v.names = c("ahiTotal", "cesdTotal"),
                    direction = "wide")
```

Calculate gain scores for each participant by taking their one-month follow-up score minus their pretest score. Do this for both the AHI and the CES-D as follows.

```
rct_wide$ahiGS <- rct_wide$ahiTotal.3 - rct_wide$ahiTotal.0
rct_wide$cesdGS <- rct_wide$cesdTotal.3 - rct_wide$cesdTotal.0
```

Finally, we will eliminate rows with missing data from the one-month follow-up by eliminating rows missing the AHI gain score variable.

```
rct_wide <- rct_wide[-which(is.na(rct_wide$ahiGS)),]
dim(rct_wide)
[1] 139 17
```

2.1 The Classical ANOVA Approach

Now we are ready to run ANOVA on the gain scores for AHI and CES-D. For notation, let Y_{ij} denote the outcome Recall from last class that the ANOVA table looks like this:

Source	Sum of Squares	*df*	Mean Square	*F* Ratio
Treatment	SSTR $= \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2$	$r-1$	MSTR $= \frac{\text{SSTR}}{r-1}$	$\frac{\text{MSTR}}{\text{MSE}}$
Error	SSE $= \sum_{i=1}^{r} n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$n_T - r$	MSE $= \frac{\text{SSE}}{n_T - r}$	
Total	SST $= \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$	$n_T - 1$		

Task 1 *Create an ANOVA table for the omnibus test of intervention effect on CES-D gain. Be sure to include an informative title with your table (see the APA 6 guide for examples of tables and titles).*

The first step to run one-way ANOVA on the gain scores for the AHI is to run a regression model and save the output to a name. The regression model should have the AHI gain scores as the dependent variable and the categorical intervention factor as the independent variable.

```
lm1 <- lm(ahiGS ~ int_fact, data = rct_wide)
```

Load package **car** and use the `Anova()` function with the output from the regression model. Note the letter "A" in the name of the function is upper case. Also note that the `type` argument is set to 3; we will discuss the differences between types after covering interactions because they don't make a difference for one-way ANOVA. Before running the test, let's recap what we are testing. Since we took post-pre differences, each group mean, $\mu_1, \mu_2, \mu_3, \mu_4$ represents change in averages (or, likewise, the average change) for each intervention group. The ANOVA F test for the main effect of the categorical factor is also known as the *omnibus* test, because it tests for any effect whatsoever. The null and alternative hypotheses for the omnibus test here are

$$\begin{cases} H_0: & \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_1: & \text{At least one of the means (average change from pre to post) differs.} \end{cases}$$

```
library(car)
Anova(lm1, type = 3)
```

```
Response: ahiGS
             Sum Sq  Df F value   Pr(>F)
(Intercept)  1154.8   1  9.8880 0.002046 **
int_fact      256.7   3  0.7327 0.534219
Residuals   15766.2 135
```

The p-value associated with the intervention factor is .53 so for $\alpha = .05$, we fail to reject the null hypothesis. There is not sufficient evidence to conclude that the average change in mean happiness, as measured by the AHI, from pretest to one-month follow-up differs across any of the four intervention groups. For depression the conclusion is much the same.

```
lm2 <- lm(cesdGS ~ int_fact, data = rct_wide)
summary(lm2)
Anova(lm2)
```

```
Response: cesdGS
             Sum Sq  Df F value  Pr(>F)
(Intercept)   226.2   1  2.8371 0.09442 .
int_fact       96.0   3  0.4012 0.75236
Residuals   10765.6 135
```

Task 2 *Write-up the results of the omnibus test for the CES-D variable in context.*

~~**Task 3** *Create an ANOVA table for the omnibus test of intervention effect on CES-D gain. Be sure to include an informative title with your table (see the APA 6 guide for examples of tables and titles).*~~

2.2 The Regression Approach: Model Comparisons

The regression approach to ANOVA is based on comparing nested linear models. A linear model A is said to be *nested* within a linear model B if it is possible to make model B identical to model A by imposing constraints on the coefficients of model B and we refer to the larger model as the *full model* and the nested model as the *reduced model*. For example, the second model below is nested within the first because the first can be made identical to the second by setting $\beta_2 = 0$. The incremental F test, defined below, may be used to test the null hypothesis that the constrained parameters are identical to their constraints, versus the alternative that they are not. For this example, $H_0 : \beta_2 = 0$ would be tested against $H_0 : \beta_2 \neq 0$.

$$
\begin{aligned}
Y_i =& \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \\
Y_i =& \beta_0 + \beta_1 X_{1i} + \epsilon_i
\end{aligned}
$$

As another example, let D_1, D_2, D_3, and D_4 be dummy-coded variables for the four categories of the positive psychology intervention, where group 4 is the control group. The the second model below is nested within the first because the first can be made identical to the second by constraining $\beta_1 = \beta_2 = \beta_3 = 0$. Thus, an incremental F test applied to these two nested models would test $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ against $H_0 : \beta_1 = \beta_2 = \beta_3 \neq 0$.

$$
\begin{aligned}
Y_i =& \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_2 D_{3i} + \epsilon_i \\
Y_i =& \beta_0 + \beta_1 X_{1i} + \epsilon_i
\end{aligned}
$$

ANOVA tests can be framed in terms of nested model comparisons by using the *incremental F test*. The incremental F test describes how to construct a statistic from nested models based on residual sums of squares.

$$
\text{RSS} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2
$$

A full model will always fit the data at least as well as a model nested within it. That is, $\text{RSS}_R \geq \text{RSS}_F$, where R is for "reduced" and F is for "full." The F-statistic for testing the omnibus null hypothesis is

$$
\begin{aligned}
F_0 &= \frac{(RSS_R - RSS_F)/(df_R - df_F)}{RSS_F/df_F}, \text{ or} \\
F_0 &= \frac{(RegSS_F - RegSS_R)/(df_R - df_F)}{RSS_F/df_F}.
\end{aligned}
$$

These formulations are identical because both models have the same TSS, and TSS = RegSS + RSS. The `anova()` function in R is used to run incremental F tests for nested models. You use the function by running two linear regression models, one full and one reduced (i.e., nested in the full model). Then you pass the models to the anova function as shown below, with the reduced model as the first argument and the full model as the second. For happiness:

```
lmF1 <- lm(ahiGS ~ int_fact, data =rct_wide)
lmR1 <- lm(ahiGS ~ 1, data = rct_wide)
anova(lmR1, lmF1)

Analysis of Variance Table

Model 1: ahiGS ~ 1
Model 2: ahiGS ~ int_fact
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1    138 16023
2    135 15766  3    256.72 0.7327 0.5342

lmF2 <- lm(cesdGS ~ int_fact, data = rct_wide)
> lmR2 <- lm(cesdGS ~ 1, data = rct_wide)
> anova(lmR2, lmF2)
Analysis of Variance Table
```

Both of the results above correspond to tests of the null hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. So how does this connect to ANOVA? Recall that in the dummy-coded regression model,

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_2 D_{3i} + \epsilon_i$$

we know that the slope coefficients represent the means of each group minus the reference group (group 4). That is, $\beta_1 = \mu_1 - \mu_4$, $\beta_2 = \mu_2 - \mu_4$, and $\beta_3 = \mu_3 - \mu_4$. Thus, the null hypothesis may be reformulated through substitution as follows:

$\beta_1 = \beta_2 = \beta_3 = 0$
$\beta_1 = 0$ and $\beta_2 = 0$ and $\beta_3 = 0$
$\mu_1 - \mu_4 = 0$ and $\mu_2 - \mu_4 = 0$ and $\mu_3 - \mu_4 = 0$
$\mu_1 = \mu_4$ and $\mu_2 = \mu_4$ and $\mu_3 = \mu_4$
$\mu_1 = \mu_2 = \mu_3 = \mu_4$

3 Research Question 2: Pairwise Comparisons

Research question 2 asks if there are pairwise differences between the positive psychology intervention groups and the control group. The particular hypotheses are framed as one-sided hypotheses because we are presumably only interested in testing if the positive psychology interventions *increase* happiness or *decrease* depression, as theorized. For the happiness outcome, the hypotheses are as follows.

Set 1: $\begin{cases} H_{0_1} : & \mu_1 <= \mu_4 \\ H_{1_1} : & \mu_1 > \mu_4 \end{cases}$ Set 2: $\begin{cases} H_{0_2} : & \mu_2 <= \mu_4 \\ H_{1_2} : & \mu_2 > \mu_4 \end{cases}$ Set 3: $\begin{cases} H_{0_3} : & \mu_3 <= \mu_4 \\ H_{1_3} : & \mu_3 > \mu_4 \end{cases}$

These hypotheses can be set up and tested in the model comparison framework using the incremental F test, but will require some custom control over the dummy variables. Up to

this point, we have been letting R automatically create dummies from the `int_fac` factor variable. Now, we will create our own dummy variables. The function `model.matrix()` creates a design matrix for regression based on a formula, so it is just what we need to create dummies based on categorical factor. Note that the first column below is all ones for the intercept. The remaining three columns represent dummy variables for TGT, GV, and REM.

```
head(model.matrix(object = lmF1))
   (Intercept) int_factTGT int_factGV int_factREM
3            1           0          0           0
11           1           0          1           0
15           1           1          0           0
20           1           0          0           0
24           1           1          0           0
32           1           1          0           0
```

To set up tests of the positive psychology interventions against the control group, we will need set the REM group as the reference category. We may use the `relevel()` function to do this as shown below. Note that after releveling, the REM group appears as the first level; the first level is always used as the reference category when R generates dummy (or other categorical) codes.

```
rct_wide$int_fact2 <- relevel(x = rct_wide$int_fact,
                              ref = "REM")
summary(rct_wide$int_fact2)
REM  SS TGT  GV
 73  72  76  74
```

After releveling, the dummies are now associated with the three positive psychology interventions, as desired.

```
mm1 <- model.matrix(object = lmF1_R)
head(mm1)
   (Intercept) int_fact2SS int_fact2TGT int_fact2GV
3            1           1            0           0
11           1           0            0           1
15           1           0            1           0
20           1           1            0           0
24           1           0            1           0
32           1           0            1           0
```

Next, run the regression of AHI gain scores on the intervention factor that has been releveled.

```
lmF1_R <- lm(ahiGS ~ int_fact2, data = rct_wide)
summary(lmF1_R)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)    3.0000     1.7766   1.689   0.0936 .
int_fact2SS    3.3103     2.6802   1.235   0.2189
int_fact2TGT   0.1628     2.4233   0.067   0.9465
int_fact2GV    2.1667     2.6551   0.816   0.4159
---

Residual standard error: 10.81 on 135 degrees of freedom
Multiple R-squared:  0.01602,Adjusted R-squared:  -0.005844
F-statistic: 0.7327 on 3 and 135 DF,  p-value: 0.5342
```

Now the coefficients and their associated p-values represent the two-sided comparisons between each positive psychology intervention gain score and the control group gain score. To convert to one-sided tests, for AHI, we hypothesized that the positive psychology interventions would increase happiness relative to the control group, so we would be looking for positive regression coefficient estimates. Note that all three of the estimates are positive (i.e., in the expected direction). So, to convert from two-sided to one-sided p-values, we will divide the p-values in half; nevertheless, the smallest is still non-significant at $\alpha = .05$. If, on the other hand, one of the estimates had been negative (i.e., in the opposite of the expected direction), we would have had to have added .5 to the p-value instead.

Task 4 *Run, report, and interpret the results of pairwise comparisons for the CES-D gain score outcome.*

4 Estimated Marginal Means

In this case we find no evidence of a statistically significant treatment effect. Nevertheless, when writing up the results it is important to summarize the data accurately. We are typically interested in group means, standard deviations, confidence intervals, effect size estimates, and the results of any tests of significance including statistic values and p-values. When other continuous covariates are involved, we will present "adjusted means" that control for the values of continuous covariates in the model. Another word for describing these means is "marginal". Thus, the general term that is used to describe model-predicted values, whether they be group means or adjusted group means is "estimated marginal means". To calculate EMMs in R we will use package **emmeans**. The package has a number of useful functions; begin by installing and loading it.

```
install.packages("emmeans")
load(emmeans)
emm1 <- emmeans(object = lmF1,
                specs = ~ int_fact,
                level = .95)
emm1
 int_fact emmean   SE  df lower.CL upper.CL
 SS         6.31 2.01 135   2.3416    10.28
 TGT        3.16 1.65 135  -0.0965     6.42
 GV         5.17 1.97 135   1.2646     9.07
```

```
 REM        3.00 1.78 135  -0.5136     6.51

Confidence level used: 0.95
```

The confidence level can be adjusted via the `level` argument to function `confint()`.

```
ci1 <- confint(object = emm1,
               level = .90)
ci1
 int_fact emmean   SE  df lower.CL upper.CL
 SS         6.31 2.01 135   2.9867     9.63
 TGT        3.16 1.65 135   0.4333     5.89
 GV         5.17 1.97 135   1.8989     8.43
 REM        3.00 1.78 135   0.0575     5.94

Confidence level used: 0.9
```

To get results for pairwise comparisons, use the `pairs()` function. Note that p-values still need to be adjusted for one-sided tests.

```
pairs(emm1, adjust = "none")
 contrast  estimate   SE  df t.ratio p.value
 SS - TGT     3.148 2.60 135  1.212  0.2276
 SS - GV      1.144 2.81 135  0.406  0.6851
 SS - REM     3.310 2.68 135  1.235  0.2189
 TGT - GV    -2.004 2.57 135 -0.779  0.4371
 TGT - REM    0.163 2.42 135  0.067  0.9465
 GV - REM     2.167 2.66 135  0.816  0.4159
```

Task 5 *Calculate estimated marginal means and pairwise differences for the CES-D outcome. Pick one of each (i.e., one emm and one p-w difference) and interpret in context.*

Finally, for effect size, interest for pairwise comparisons is typically in Cohen's d, which is a standardized measure of mean difference, defined as follows,

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_{\text{pooled}}},$$

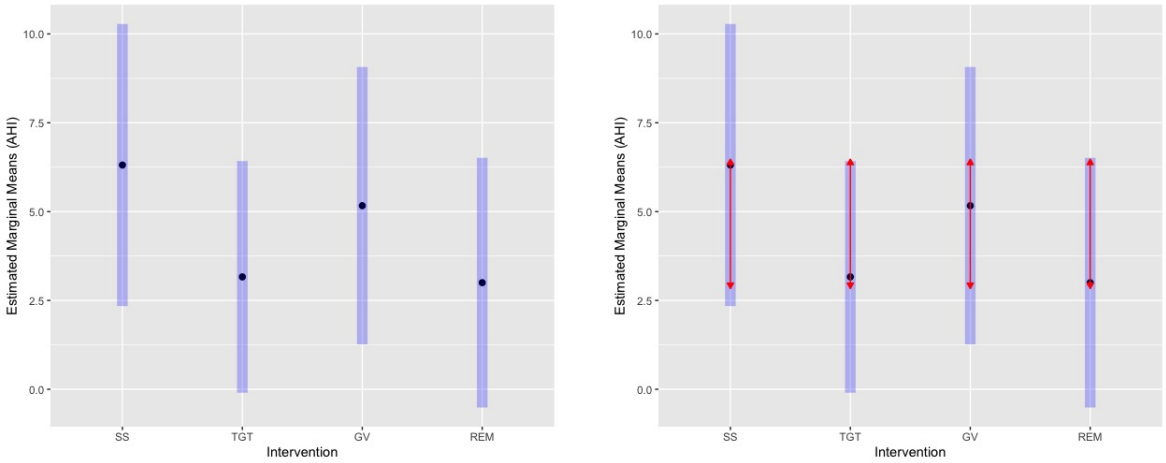
where X_1 and X_2 are the sample means and s_1^2 and s_2^2 are the sample variances of the treated and comparison groups, respectively, and $s_{\text{pooled}} = \sqrt{\frac{(N_1-1)s_1^2+(N_2-1)s_2^2}{N_1+N_2-2}}$ is the pooled standard deviation across groups, where N_1 and N_2 are the treated and comparison group sample sizes, respectively. Note that $N_1 + N_2 - 2$ is an expression for the degrees of freedom for d. Estimating effect sizes with R will be discussed later in the course.
Finally, results may be plotted.

```
?plot.emmGrid
plot(x = emm1, alpha = .05,
     level = .95,
```

```
adjust = "Tukey",
comparisons = TRUE,
horizontal = FALSE,
xlab = "Intervention",
ylab = "Estimated Marginal Means (AHI)")
```

Task 6 *Create and report a plot of estimated marginal means with 95% confidence intervals and comparison arrows. Interpret the plot.*

Figure 1: Plots of marginal means of gain scores for the AHI by intervention group along with 95% confidence intervals. The left panel does not show comparison arrows (i.e., `comparisons = FALSE`); the right panel does (i.e., `comparisons = TRUE`).



4.1 Reporting Descriptives in a Table

See p. 143 of the APA 6 Manual for a table that motivated this style of presentation.

Table 1: Estimated pairwise mean differences in average gains in happiness as measured by the AHI)

Contrast	Estimate	SE	df	t ratio	p-value
SS - REM	3.3103	2.6802	135	1.235	0.2189
TGT - REM	0.1628	2.4233	135	0.067	0.9465
GV - REM	2.1667	2.6551	135	0.816	0.4159

Note. SS = signature strengths, TGT = three good things; GV = gratitude visit; REM = recall early memories (control group)

Task 7 *Create a table for CES-D output and add the 95% confidence interval.*