HUDM 5123 - Linear Models and Experimental Design
Lab 07 - Effect Size and Power

1 Data

Data for lab come from a randomized experiment to study the efficacy of acupuncture for treating headaches. Results of the trial were published in the British Medical Journal in 2004. You may view the paper at the following link: `http://www.bmj.com/content/328/7442/744.full`. The data set includes 301 cases: 140 control (no acupuncture) and 161 treated (acupuncture). Participants were randomly assigned to groups. Variable names and descriptions are as follows:

- `age`; age in years
- `sex`; male = 0, female = 1
- `migraine`; diagnosis of migraines = 1, diagnosis of tension-type headaches = 0
- `chronicity`; number of years of headache disorder at baseline
- `acupuncturist`; ID for acupuncture provider
- `group`; acupuncture treatment group = 1, control group = 0
- `pk1`; headache severity rating at baseline
- `pk5`; headache severity rating 1 year later

The data can be found in a .Rdata file on canvas called acupuncture.Rdata.

2 Was the Treatment Effective?

Begin by testing for a treatment effect via one-way ANOVA. The group variable is not a factor so our first step is to convert it to a factor. Then, we will work with the difference in headache pain score (1 year follow up minus baseline score), but that variable does not yet exist in the data set, so we will need to create it.

```
acupuncture$group <- factor(acupuncture$group,
                            levels = c(0,1),
                            labels = c("C", "T"))

acupuncture$diff <- acupuncture$pk5 - acupuncture$pk1
```

To test the efficacy of the treatment, run a linear model regressing the pain difference on the group and pass the model output to the `Anova()` function from package **car**.

```
options(contrasts = c("contr.sum", "contr.poly"))
lm1 <- lm(diff ~ group, data = acupuncture)
library(car)
Anova(lm1, type = 3)

Anova Table (Type III tests)

Response: diff
            Sum Sq  Df F value    Pr(>F)
(Intercept)  12071   1 88.2814 < 2.2e-16 ***
group         1176   1  8.5967  0.003628 **
Residuals    40885 299
```

Run `emmeans` to estimate the group means.

```
(emm1 <- emmeans(object = lm1, specs = ~ group))
 group emmean    SE  df lower.CL upper.CL
 C      -4.37 0.988 299    -6.31    -2.42
 T      -8.33 0.922 299   -10.14    -6.52

pairs(emm1)
 contrast estimate   SE  df t.ratio p.value
 C - T        3.96 1.35 299 2.932   0.0036
```

Note that the difference was calculated as C - T. T - C is, therefore, -3.96. Thus, the acupuncture treatment was effective at reducing the headache pain by 3.96 points, on average.

2.1 Cohen's d

Cohen's d is defined as follows.

$$d = \frac{\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}}{s_p} = \frac{\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}}$$

We will calculate Cohen's d with a function from the package **effsize** and also using the formula. The function `cohen.d()` is used as follows.

```
cohen.d(diff ~ group,
         data = acupuncture,
         pooled = TRUE)

d estimate: 0.3388233 (small)
```

To use the formula, we need to know the group sample sizes and the group sample variances.

```
table(acupuncture$group)

  C   T
140 161

by(data = acupuncture$diff, INDICES = acupuncture$group, FUN = var)
acupuncture$group: C
[1] 108.208
------------------------------------------------------------------------
acupuncture$group: T
[1] 161.5239
```

Plug in to the formula for s_p.

```
s_pooled <- sqrt((139*108.2080 + 160*161.5239)/(140 + 161 - 2))
s_pooled # 11.6935
 11.69351
(d <- (-8.33 - -4.37)/11.6935)
 -0.3386497
```

2.2 Eta Squared

The *complete* η^2 value for a categorical predictor, A, is calculated as the sum of squares for the factor A, divided by the total sum of squares.

$$\eta^2 = \frac{\mathrm{SS}_A}{\mathrm{TSS}}.$$

Partial η^2:

$$\eta^2_{\text{partial}} = \frac{\mathrm{SS}_A}{\mathrm{SS}_A + \mathrm{SSE}}.$$

It would be unusual to use η^2 instead of d for a two-group comparison, but we will calculated here for completeness. Since this is a single factor design, complete and partial η^2 are the same. The value here is 0.028, which is a small effect.

$$\eta^2 = \frac{\mathrm{SS}_A}{\mathrm{TSS}} = 1176/(1176 + 40885) = 0.028$$

3 Was there an age effect?

The ages of participants range from 18 to 65.
Create a factor based on age cut points.

```
acupuncture$age_3 <- cut(x = acupuncture$age,
                         breaks = c(18, 39, 50, 65))
table(acupuncture$age_3)
```

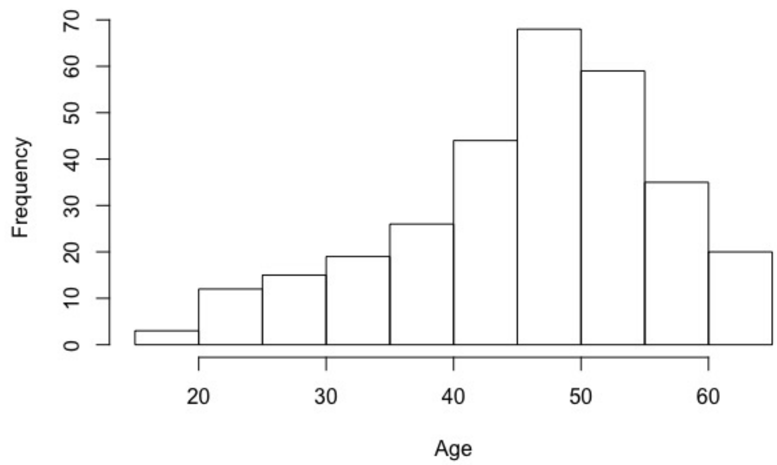

Figure 1: Histogram of ages for the acupuncture data participant sample

```
(18,39] (39,50] (50,65]
     67     119     114
```

Before estimating main effects, we should check to see if age category interacts with treatment group. Perhaps, for example, the acupuncture intervention was more effective for some age groups than others.

```
lm2 <- lm(diff ~ group*age_3, data = acupuncture)
Anova(lm2, type = 3)
Anova Table (Type III tests)

Response: diff
            Sum Sq  Df F value    Pr(>F)
(Intercept)  12910   1 96.2947 < 2.2e-16 ***
group         1457   1 10.8679  0.001098 **
age_3         1103   2  4.1117  0.017331 *
group:age_3    361   2  1.3477  0.261426
Residuals    39417 294
```

Also, examine the interaction plot.

```
emm2 <- emmeans(lm2, ~ group*age_3)
emmip(emm2, group ~ age_3)
```

Since the interaction is not significant, we will interpret the main effect of the age factor and, if significant, we will follow up with main effects pairwise comparisons of the three age groups. Using Shaffer's planned post-omnibus modification we may test all three pairwise comparisons at .05.

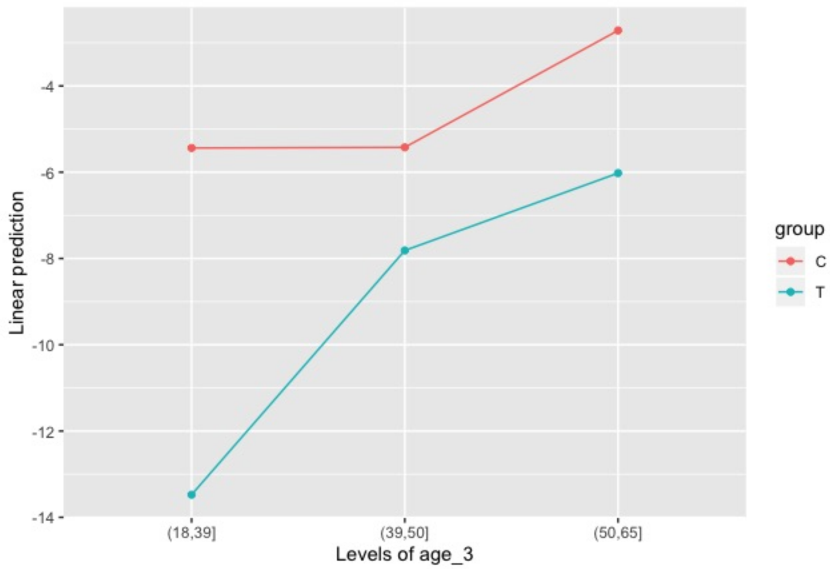

Figure 2: Interaction plot for the effect of age and group factors on pain difference scores

```
emm3 <- emmeans(lm2, ~ age_3)
NOTE: Results may be misleading due to involvement in interactions

pairs(emm3, adjust = "none")
 contrast          estimate   SE  df t.ratio p.value
 (18,39] - (39,50]    -2.84 1.78 294 -1.596  0.1116
 (18,39] - (50,65]    -5.09 1.78 294 -2.855  0.0046
 (39,50] - (50,65]    -2.25 1.53 294 -1.472  0.1421
```

Thus, over the course of the study, pain decreased among 18 to 39 year olds by about 5 points more, on average, than it decreased among 50 to 65 year olds (t(294) = -2.86; p-value = .005).

3.1 Cohen's d

In a design with multiple groups (i.e., here there are three groups), if the constant variance assumption holds then the square root of MSE may be used as a replacement for the pooled SD, s_p. This is because MSE is an unbiased estimator for σ^2, the common error variance. That is, the square root of MSE is an extension of s_p for more than two groups. Note that before using this method, we must check that the constant variance assumption holds across age groups.

```
by(data = acupuncture$diff,
+    INDICES = acupuncture$age_3,
+    FUN = var)
acupuncture$age_3: (18,39]
[1] 154.3251
------------------------------------------------------------------------
acupuncture$age_3: (39,50]
```

```
[1] 128.5534
--------------------------------------------------------------------------
acupuncture$age_3: (50,65]
[1] 138.2354
```

The ratio of largest to smallest variance is only about 1.2. Furthermore, Levene's test null hypothesis is not rejected.

```
leveneTest(y = acupuncture$diff,
           group = acupuncture$age_3)
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   2  0.0153 0.9849
      297
```

We can get MSE from the ANOVA table by calculating SSE/df = 39417/294 = 134.07.

```
Anova(lm4, type = 3)
Anova Table (Type III tests)

Response: diff
            Sum Sq  Df F value    Pr(>F)
(Intercept)  12910   1 96.2947 < 2.2e-16 ***
group         1457   1 10.8679  0.001098 **
age_3         1103   2  4.1117  0.017331 *
group:age_3    361   2  1.3477  0.261426
Residuals    39417 294
```

Therefore, d = -.44.

```
(d_3 <- (-9.46 - -4.37)/sqrt(134.0714))
[1] -0.4395919
```

3.2 Eta Squared

Use the ANOVA table to calculate eta squared for the age effect.


```
Anova(lm4, type = 3)
Anova Table (Type III tests)

Response: diff
            Sum Sq  Df F value    Pr(>F)
(Intercept)  12910   1 96.2947 < 2.2e-16 ***
group         1457   1 10.8679  0.001098 **
age_3         1103   2  4.1117  0.017331 *
group:age_3    361   2  1.3477  0.261426
Residuals    39417 294
```

Complete η^2:

$$\eta^2 = 1103/(1457 + 1103 + 361 + 39417) = .026$$

Partial η^2:

$$1103/(1103 + 39417) = .027$$

Both complete and partial values are indicative of small effects.

4 Do the Youngest Participants Differ from the Average of the Two Older Groups?

To determine if the youngest group differs from the average of the two older groups in terms of pain difference score we will define and test a contrast.

$$L = 1\mu_1 - 1/2\mu_2 - 1/2\mu_3$$

```
contrast(emm3, method = list(c(1, -1/2, -1/2)))
 contrast          estimate   SE  df t.ratio p.value
 c(1, -0.5, -0.5)     -3.96 1.61 294 -2.465  0.0143
```

4.1 Cohen's d

The contrast is significant (t(294) = -2.47; p = .014). The estimated value of the contrast is -3.96. Thus, the average pain rating in the youngest group decreased by about 4 points more than the average pain rating decreased in the older two groups. Using the MSE method for Cohen's d yields the following.

```
(d_4 <- -3.96 / sqrt(134.0714))
[1] -0.3420008
```

4.2 Eta Squared

To get η^2 for the contrast, we first need the sum of squares for the contrast. Recall that the SS for the contrast may be calculated as follows.

$$\begin{aligned}\text{SS}_L =& \frac{\hat{L}^2}{\displaystyle\sum_{i=1}^{r}\left(\frac{c_i^2}{n_i}\right)} \\ =& \frac{-3.96^2}{\frac{1^2}{67} + \frac{(-1/2)^2}{119} + \frac{(-1/2)^2}{114}} = 815.93\end{aligned}$$

With the contrast sum of squares in hand, we can now calculate the value of η^2 for the contrast.

$$\eta^2 = 815.93/(1457 + 1103 + 361 + 39417) = .019.$$

5 Power Analysis

We will use package **pwr** for power calculations; install and load the package. Suppose the above analyses were for a pilot study to plan a larger study. We will use the estimates of effect from above to determine the sample size requirements for sufficient power to detect (a) the group effect, (b) the age effect, and (c) the contrast.

5.1 The Group Effect

Above we estimated Cohen's d to be about -0.34 for the effect of acupuncture on one-year change in pain rating. The functions for power analysis in package **pwr** start with `pwr.` and end with `.test`. The function for the two-group case is called `pwr.t.test` and it takes arguments

- `n`; number of observations (*per group*; assuming equal group sizes)
- `d`; Cohen's d for the effect in the population
- `sig.level`; significance level (Type I error probability α)
- `power`; power of the test (1 - β)

One of these four arguments, `n`, `d`, `sig.level`, and `power`, must be left out or set to `NULL` and it will be calculated. For this situation, we will leave `n` out as we wish to know the sample size required power of .8 with $\alpha = .05$ for $d = -.34$. According to the output (directly below), the required sample size for .8 power is 137 participants per group (i.e., 274 total).

```
pwr.t.test(d = -.34, sig.level = .05, power = .8)

     Two-sample t test power calculation

              n = 136.7605
              d = 0.34
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

5.2 Power Curves

If there is uncertainty about the effect size, it can be instructive to make a power curve (i.e., a plot of power across a number of effect sizes) for a specific sample size. Here, we will create power curves for effect sizes ranging from -0.5 to -0.2 for sample sizes 50, 100, 150, and 200.

```
out50 <- pwr.t.test(d = seq(-.5, -.2, .01), n = 50)
out100 <- pwr.t.test(d = seq(-.5, -.2, .01), n = 100)
out150 <- pwr.t.test(d = seq(-.5, -.2, .01), n = 150)
out200 <- pwr.t.test(d = seq(-.5, -.2, .01), n = 200)
```

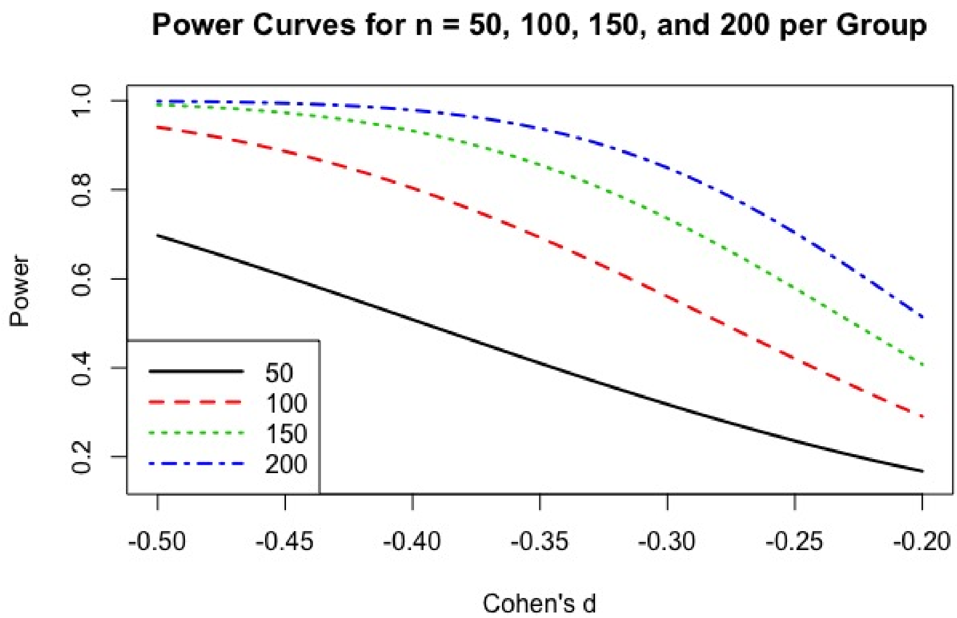

Figure 3: Power curves for the two group case

5.3 The Age Effect

The complete eta squared value for age was .026. The **pwr** package, like most power software, uses Cohen's f instead of ω^2. Recall the relationship between f and ω^2:

$$f = \sqrt{\frac{\omega^2}{(1-\omega^2)}} \quad \text{and} \quad \omega^2 = \frac{f^2}{(f^2+1)}.$$

Also, recall that ω^2 and η^2 are both measures of association that are very similar. So we will use .026 for ω^2 and plug it into the formula to convert ω^2 to f.

$$f = \sqrt{\frac{\omega^2}{(1-\omega^2)}} = \sqrt{.026/(1-.026)} = .16$$

```
pwr.anova.test(k = 3, f = .16,
               sig.level = .05,
               power = .8)

     Balanced one-way analysis of variance power calculation

              k = 3
              n = 126.4556
              f = 0.16
      sig.level = 0.05
```

```
         power = 0.8

NOTE: n is number in each group
```

6 Tasks for Lab 07

Your TASKS for lab 07 have to do with the `sex` variable in the acupuncture data set.

1. TASK 1: Convert the `sex` variable into a factor that has levels `male` and `female`.
2. TASK 2: Assess if there is a sex effect on change in pain score from baseline to one year follow up. Only look for a main effect for `sex`, not an interaction with `group`.
3. TASK 3: Calculate and interpret Cohen's d for the relationship of `sex` on change in pain rating. (Do this whether or not the relationship is significant.)
4. TASK 4: Calculate and interpret complete eta squared for the relationship of `sex` on change in pain. (Do this whether or not the relationship is significant.)
5. TASK 5: Suppose this was a pilot study to generate effect size estimates for the relationship between participant sex and change in pain. Use the estimate of Cohen's d to calculate the sample size needed to detect the effect with $\alpha = .05$ and for power $= .8$ in a two-sample design based on the two-sample t test.
6. TASK 6: Do the same sample size calculation for eta squared using an ANOVA design.