HW3: Power & Individual Comparisons

5/5 Great!

Seung Kim

You seem to have an excellent handle on the material!

October 29, 2020

In a study of a behavioral self-control intervention for problem drinkers, one of the less sensitive dependent variables was number of drinking days per week (Hester & Delaney, 1997). Forty participants were assigned randomly to either receive the intervention immediately or be in a waiting list control group (i.e., n=20 per group). At the initial follow-up assessment, the means and standard deviations on Drinking Days per Week were as follows:

Condition	Mean	SD
Immediate	3.65	1.57
Delayed	4.80	2.55

Assume this set of data is being viewed as a pilot study for a proposed replication.

1. Conduct an ANOVA on these data, and compute as descriptive measures of the observed effect size both d and $\hat{f}_{obv}$.

Solution.

I'm writing a lot to check my understanding (trying to really make sure I *understand* stats)—if you have the patience for it, I would really appreciate some feedback.

Let us start by articulating our models. Since the groups have equal sample sizes, we can quickly get grand mean $\hat{\mu} = 4.225$ by averaging the group means. Then, we have the following models:

- Full: $Y_{ij} = \mu_j + \epsilon_{ijF}$
- Restricted: $Y_{ij} = 4.225 + \epsilon_{ijR}$

Let's work on the ANOVA. ANOVA essentially compares the effect produced by the differences detected between groups (effect from Treatment; accounted for by Model) to the effect produced by the difference detected within each group (effect from Error; accounted for by Residual). The F-statistic measures the ratio between these differences.

$$F = \frac{MS_{between}}{MS_{within}}$$

yes

$MS_{between}$ is approximated by $\frac{E_R - E_F}{df_R - df_F}$, since we can isolate the effect produced by differences between groups by comparing the effect of the Full model (E_F), which assumes that different groups exist, to the effect of the Restricted model (E_R), which assumes that no differences between groups exist.

MS_{within} is approximated by $\frac{E_F}{df_F}$, since we can only get the effect produced by differences within each group through the Full model, which assumes that different groups exist.

yes!

Onto the calculations. Let's first get $E_R - E_F$:

$$(E_R - E_F) = \sum_{j=1}^{a} n_j(\bar{Y}_j - \bar{Y})^2 = 20(3.65 - 4.225)^2 + 20(4.80 - 4.225)^2 = 6.6125 + 6.6125$$
$$= 13.225$$

And we can get the $\frac{E_F}{df_F}$ as well:

$$\frac{E_F}{df_F} = \frac{\sum s_j^2}{a} = \frac{1.57^2 + 2.55^2}{2}$$
$$= 4.484$$

Cool. Now all we need is $df_R - df_F$. For the restricted model, we only estimate the grand mean, so $df_R = 40 - 1 = 39$; for the full model, we estimate the grand mean and one group mean so $df_F = 40 - 2 = 38$. So:

$$(df_R - df_F) = 39 - 38$$
$$= 1$$

So now let's bang out the F-statistic:

$$F = \frac{(E_R - E_F)/(df_R - df_F)}{E_F/df_F} = \frac{13.225/1}{4.484}$$
$$= 2.950$$

And compare to $F_{0.05;1,38}$:

or could compute pf(2.950, 1, 38, lower.tail = F) = .09 (not sig.)

```
[32]: qf(0.05, 1, 38, lower.tail = FALSE)
```

4.09817173088084

Our $F = 2.950$ is smaller than the $F_{0.05;1,38}$. So the full model does not do significantly better than the restricted model, and there is not enough evidence to believe that immediate treatment significantly affects the number of drinking days per week compared to the control.

Let's now talk about the measures of effect size, d and $\hat{f}_{obv}$.

yes

Cohen's d is an estimate of the standardized difference between the population means. We happen to know that $\sqrt{MS_{within}}$ is our best estimate for the population standard deviation (since we assume the homogeneity of variance, both groups are assumed to have the same σ). We cal this figure s_p or pooled standard deviation.

good

To get d, we plug into the following equation:

$$d = \frac{\bar{Y}_1 - \bar{Y}_2}{s_p} = \frac{3.65 - 4.80}{\sqrt{MS_{within}}} = \frac{-1.15}{\sqrt{4.484}} = \frac{-1.15}{2.118}$$
$$= -0.543$$

yes, exactly

So, immediate intervention decreased drinking days per week by 0.543 standard deviations. That is just above Cohen's cutoff for a 'medium' effect.

Cool. The f is the ratio of the standard deviation of populations means to the pooled within-group standard deviation.

$$f = \frac{\sigma_m}{\sigma_\epsilon}$$

The *standard deviation of population means* (σ_m) is estimated by the standard deviation of the group means from the grand mean; the *within-group standard deviation* (σ_ϵ) is just $\sqrt{MS_{within}}$, which is the same thing as standard deviation because of variation inherent to within each group (=error, residual). This constitutes the 'obvious' estimate of f.

Let's now calculate it. Since our grand mean is 4.225,

$$\hat{f}_{obv} = \frac{\sqrt{\sum_1^j(\mu_j - \mu)^2/a}}{\sqrt{MS_{within}}} = \frac{\sqrt{\frac{(3.65-4.225)^2+(4.80-4.225)^2}{2}}}{\sqrt{4.484}} = \frac{\sqrt{0.331}}{\sqrt{4.484}}$$
$$= 0.272$$

exactly!

There's another way to calculate this.

$$\hat{f}_{obv} = \sqrt{\frac{(a-1)F}{N}} = \sqrt{\frac{1 \times 2.950}{40}} = 0.272$$

An $f \geq 0.25$ is considered a 'medium' effect size. But since f_{obv} is usually an overestimate of f_{unb}, f_{unb} is probably smaller and possibly considered 'small'.

2. Determine the sample size that would be required to achieve a power of .80 using an α of .05 if one used the value of $\hat{f}_{obv}$ arrived at in #1 as the effect size measure in a power analysis.

Solution. We would need 55 participants in each group (rounding up), for a total sample size of 110.

```
[9]: library('pwr')
```

```
[10]: pwr.anova.test(k=2, f=0.272, sig.level=0.05, power=0.80)
```

```
     Balanced one-way analysis of variance power calculation

              k = 2
              n = 54.02168
              f = 0.272
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

3. Now compute $f_{.50}$ and $f_{.33}$, the lower bounds of one-sided 50% and 66.7% confidence intervals for this effect size measure. Carry out a revised power analysis based on these two alternate effect size measures. How do the required sample sizes compare to what you found in #2? What does this tell you about how effect size relates to required sample size?

Solution. $\hat{f}_{.50}$ is the median effect size, which is less biased than both $\hat{f}_{obv}$ and $\hat{f}_{unb}$.

```
[12]: library('MBESS')
```

```
[67]: ci.srsnr(F.value=2.950, df.1=1, df.2=38, N=40, alpha.lower=0.5, alpha.upper=0)
```

$Lower.Limit.of.the.Square.Root.of.the.Signal.to.Noise.Ratio
0.269602691831927
$Upper.Limit.of.the.Square.Root.of.the.Signal.to.Noise.Ratio

So we get an $\hat{f}_{med} = 0.270$, which is marginally smaller than $\hat{f}_{obv}$.

```
[71]: pwr.anova.test(k=2, f=0.269602691831927, sig.level=0.05, power=0.80)
```

```
     Balanced one-way analysis of variance power calculation

              k = 2
              n = 54.96887
              f = 0.2696027
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

And we need 55 people per sample for a total of 110 people, but $n = 54.969$ is closer to 55 than the $n = 54.022$ of $\hat{f}_{obv}$.

How about $\hat{f}_{0.33}$, the lower bound for the 66.7% CI for f?

```
[72]: ci.srsnr(F.value=2.950, df.1=1, df.2=38, N=40, alpha.lower=0.33, alpha.upper=0)
```

$Lower.Limit.of.the.Square.Root.of.the.Signal.to.Noise.Ratio
0.198067264880928
$Upper.Limit.of.the.Square.Root.of.the.Signal.to.Noise.Ratio

```
[73]: pwr.anova.test(k=2, f=0.198067264880928, sig.level=0.05, power=0.80)
```

```
     Balanced one-way analysis of variance power calculation

              k = 2
              n = 101.0042
              f = 0.1980673
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

So we get $\hat{f}_{0.33} = 0.198$ which is much smaller than $\hat{f}_{obv}$, and needs a much larger sample size of 102 for each group for a total of 204.

right!

So, as the effect size decreases, we need a larger sample size if we want the same level of power. Which makes sense—if the effect is really big, we wouldn't need to test as many people to know that the effect is really there, but if the effect is small, we would need to test a lot of people to make sure the effect is not just from sampling error.

great

A graduate student has conducted a treatment study involving three treatments to alleviate depression. The first two groups are active treatment groups and the third group is a placebo control group. The following data are obtained:

Active Treatment 1	Active Treatment 2	Control
10	6	13
13	12	16
14	8	13
8	13	19
9	10	11
12	16	13

You may assume homogeneity of variance throughout all parts of this exercise. You may use `aov()` or `lm()` to obtain MS_{within}; otherwise, carry out all analyses by hand. It is fine to use R as a calculator for any necessary quantities – e.g., group means, $\hat{\psi}$, etc. – but please clearly show your calculations for $SS_{contrast}$ and $F_{contrast}$.

4. Test whether the mean of Active Treatment 1 is different from the mean of the Control group.

Solution. OK, let's first take care of the dataset.

```
[28]: AT1 = c(10, 13, 14, 8, 9, 12)
      AT2 = c(6, 12, 8, 13, 10, 16)
      Control = c(13, 16, 13, 19, 11, 13)
```

We're doing an individual mean comparison between the Active Treatment 1 and Control groups. Let's start by articulating our models:

- Full: $Y_{ij} = \mu_j + \epsilon_{ij}$ (all means are different)
- Restricted: $Y_{ij} = \mu_j + \epsilon_{ij}$ where $\mu_1 = \mu_3$ (all means are different, except $\mu_1 = \mu_3$)

Now, we can do our individual comparison to test $H_0 : \mu_1 = \mu_3$. We can go about this in two ways. I'm going to try the long way first.

```
[62]: E_F = sum((AT1-mean(AT1))^2)+sum((AT2-mean(AT2))^2)+sum((Control-mean(Control))^2)
      Y_star = c(AT1, Control)
      E_R = sum((AT1-mean(Y_star))^2)+sum((AT2-mean(AT2))^2)+sum((Control-mean(Y_star))^2)

      # df_F = 15 since we estimate 3 means;
      # df_R = 16 since we only estimate 2.
      df_F = 15
      df_R = 16
      num = (E_R-E_F)/(df_R-df_F)
      denom = E_F/df_F
      F = num/denom
      F
```

good!

3.37593516209476

Cool, so we get $F = 3.376$. Let's try the shortcut way:

```
[27]: mean(AT1)
      mean(AT2)
      mean(Control)
```

11

10.8333333333333

14.1666666666667

```
[30]: sd(AT1)
      sd(AT2)
      sd(Control)
```

2.36643191323985

3.60092580688171

2.85773803324704

$$\hat{\psi} = \sum_{j=1}^{3} c_j\mu_j = (1)11 + (0)10.833 + (-1)14.167$$
$$= -3.167$$

So the Control has a mean that is 3.167 depression units greater than AT1. Let's see if this difference is statistically significant.

We have some tricks up our sleeves in calculating the F-statistic. Since

$$E_R - E_F = SS_{contrast} = SS(\psi) = \frac{\hat{\psi}^2}{\sum_{j=1}^{a}(\frac{c_j^2}{n_j})}$$

Our $E_R - E_F$ is:

$$= \frac{(-3.167)^2}{\frac{1^2}{6} + \frac{0^2}{6} + \frac{(-1)^2}{6}} = \frac{10.030}{\frac{1}{3}}$$
$$= 30.090$$

And, $df_F = 18 - 3 = 15$ since we're estimating three unique μ_js, and $df_R = 18 - 2 = 16$ since we're only estimating two unique μ_js ($\mu_1 = \mu_3, \mu_2$).

All we need now is our E_F/df_F, which is just equal to MS_{within}:

$$\frac{E_F}{df_F} = MS_{within} = \frac{\sum s_j^2}{a} = \frac{2.366^2 + 3.601^2 + 2.858^2}{3}$$
$$= 8.911$$

Awesome. Now we can get our F-statistic.

$$F_{contrast} = \frac{30.090/1}{8.911} = 3.377$$

```
[64]: qf(0.05, 1, 15, lower.tail=FALSE)
```

4.54307716526697

Since our $F_{contrast} = 3.377$ is smaller than $F_{0.05;1,15} = 4.543$, there is no reason to prefer the full model to the restricted one, and not enough evidence to say that the mean of Active Treatment 1 is statistically different from the mean of the Control.

5. Test whether the mean of Active Treatment 2 is different from the mean of the Control group.

We're doing an individual mean comparison between the Active Treatment 2 and Control groups. Let's start by articulating our models:

- Full: $Y_{ij} = \mu_j + \epsilon_{ij}$
- Restricted: $Y_{ij} = \mu_j + \epsilon_{ij}$ where $\mu_2 = \mu_3$

$$\hat{\psi} = \sum_{j=1}^{3} c_j\mu_j = (0)11 + (1)10.833 + (-1)14.167$$
$$= -3.334$$

So the Control's depression units are 3.334 higher than AT2's.

We have some tricks up our sleeves in calculating the F-statistic. Since

haha

$$E_R - E_F = SS_{contrast} = SS(\psi) = \frac{\hat{\psi}^2}{\sum_{j=1}^{a}(\frac{c_j^2}{n_j})}$$

Our $E_R - E_F$ is:

$$= \frac{(-3.334)^2}{\frac{0^2}{6} + \frac{1^2}{6} + \frac{(-1)^2}{6}} = \frac{11.116}{\frac{1}{3}}$$
$$= 33.347$$

And, $df_F = 18 - 3 = 15$ since we're estimating three unique μ_js, and $df_R = 18 - 2 = 16$ since we're only estimating two unique μ_js ($\mu_1, \mu_2 = \mu_3$).

All we need now is our E_F/df_F, which is just equal to MS_{within}:

$$\frac{E_F}{df_F} = MS_{within} = \frac{\sum s_j^2}{a} = \frac{2.366^2 + 3.601^2 + 2.858^2}{3}$$
$$= 8.911$$

Awesome. Now we can get our F-statistic.

$$F_{contrast} = \frac{33.347/1}{8.911} = 3.742$$

Since our $F_{contrast} = 3.742$ is smaller than $F_{0.05;1,15} = 4.543$, there is no reason to prefer the full model to the restricted one, and not enough evidence to say that the mean of Active Treatment 2 is statistically different from the mean of the Control.

6. Test whether the mean of the two Active Treatment groups combined is different from the mean of the Control group (i.e., form a complex comparison).

Solution. A complex comparison is a little bit different. The models are:

- Full: $Y_{ij} = \mu_j + \epsilon_{ij}$
- Restricted: $Y_{ij} = \mu_j + \epsilon_{ij}$ where $\frac{1}{2}(\mu_1 + \mu_2) = \mu_3$

We can articulate the null hypothesis as:

$$H_0 : \frac{1}{2}(\mu_1 + \mu_2) = \mu_3$$

Which is the same as:

$$H_0 : \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3 = 0$$

This constitutes a contrast, which is a *linear combination of means in which the coefficients of the means add up to zero*. We call this ψ.

$$\hat{\psi} = \sum_{j=1}^{3} c_j\mu_j = \left(\frac{1}{2}\right)11 + \left(\frac{1}{2}\right)10.833 + (-1)14.167 = 5.5 + 5.417 - 14.167$$
$$= -3.251$$

So the Control's depression units are 3.251 higher than the Active Treatment groups combined.

We have some tricks up our sleeves in calculating the F-statistic. Since

$$E_R - E_F = SS_{contrast} = SS(\psi) = \frac{\hat{\psi}^2}{\sum_{j=1}^{a}(\frac{c_j^2}{n_j})}$$

Our $E_R - E_F$ is:

$$= \frac{(-3.251)^2}{\frac{0.5^2}{6} + \frac{0.5^2}{6} + \frac{(-1)^2}{6}} = \frac{10.569}{0.25}$$
$$= 42.276$$

As for the degrees of freedom, $df_F = 18 - 3 = 15$ since we estimate three means; $df_R = 18 - 2 = 16$ since we assume that $\frac{1}{2}(\mu_1 + \mu_2) = \mu_3$, so if we estimate two means we can find the third.

All we need now is our E_F/df_F, which is just equal to MS_{within}:

$$\frac{E_F}{df_F} = MS_{within} = \frac{\sum s_j^2}{a} = \frac{2.366^2 + 3.601^2 + 2.858^2}{3}$$
$$= 8.911$$

So let's bang out our F:

$$F = \frac{42.276/1}{8.911} = 4.744$$

Since our $F = 4.744$ is greater than $F_{0.05;1,15} = 4.543$, there is reason to prefer the Restricted model to the Full. There is enough evidence to say that the mean depression score of both the active treatments combined is lower than that of the Control.

```
[65]: pf(4.744, 1, 15, lower.tail=FALSE)
```

0.0457686397756328

7. Form a 95% confidence interval for the mean difference between Active Treatment 1 and the Control group.

Solution. Here are the pieces we have:

- $\hat{\psi} = -3.167$
- $F_{0.05;1,15} = 4.543$
- $MS_{within} = 8.911$
- $\sum_{j=1}^{a}(c_j^2/n_j) = \frac{1}{3}$

So the confidence interval is:

$$\hat{\psi} \pm \sqrt{F_{\alpha;1,N-a}} \times \sqrt{MS_{within}\sum_{j=1}^{a}(c_j^2/n_j)} = -3.167 \pm \sqrt{4.543} \times \sqrt{8.911 \times \frac{1}{3}}$$
$$= -3.167 \pm (2.131 \times 1.723) = -3.167 \pm 3.673$$
$$= [-6.840, 0.\cancel{510}]$$

(within rounding error) 0.503

8. Form a 95% confidence interval for the mean difference between Active Treatment 2 and the Control group.

Solution. Here are the pieces we have:

- $\hat{\psi} = -3.334$
- $F_{0.05;1,15} = 4.543$
- $MS_{within} = 8.911$
- $\sum_{j=1}^{a}(c_j^2/n_j) = \frac{1}{3}$

So the confidence interval is:

$$\hat{\psi} \pm \sqrt{F_{\alpha;1,N-a}} \times \sqrt{MS_{within}\sum_{j=1}^{a}(c_j^2/n_j)} = -3.334 \pm \sqrt{4.543} \times \sqrt{8.911 \times \frac{1}{3}}$$
$$= -3.334 \pm (2.131 \times 1.723) = -3.334 \pm 3.673$$
$$= [-7.007, 0.339]$$

9. Form a 95% confidence interval for the difference between the mean of the two Active Treatment groups and the mean of the Control group.

Solution. Here are the pieces we have:

- $\hat{\psi} = -3.251$
- $F_{0.05;1,15} = 4.543$
- $MS_{within} = 8.911$
- $\sum_{j=1}^{a}(c_j^2/n_j) = \frac{1}{4}$

So the confidence interval is:

$$\hat{\psi} \pm \sqrt{F_{\alpha;1,N-a}} \times \sqrt{MS_{within}\sum_{j=1}^{a}(c_j^2/n_j)} = -3.251 \pm \sqrt{4.543} \times \sqrt{8.911 \times \frac{1}{4}}$$
$$= -3.251 \pm (2.131 \times 1.493) = -3.251 \pm 3.181$$
$$= [-6.432, -0.070]$$

10. Which of your intervals in #7-#9 contain zero and which exclude zero? How does this relate to the tests you performed in #4-#6?

Solution. Only #7 and #8 contain zero; #9 does not. This is significant since we could not reject the null hypothesis for #7 and #8, but we could for #9—that is, we had not enough evidence to say $\psi \neq 0$ for #7 and #8 within the 95% CI, but did have evidence to say $\psi \neq 0$ for #9 within the 95% CI.