Homework 5
Psychology 311
Spring, 2015

Instructions. Answer the following questions. Show your R code, your input, and your output.

1. (60 points).

B_1				B_2			
A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
24	22	31	28	52	26	20	51
22	14	33	31	25	18	25	40
36	24	43	29	37	17	15	54

(a) (15) The table shown above has data for a 2-way ANOVA with $n = 3$ per cell. The data file, *EX9_1.csv*, is available on the course website, should you need it. Cohen's f for the A main effect is defined as

$$f = \frac{\sigma_A}{\sigma_e}$$

where

$$\sigma_A^2 = \frac{\sum_{i=1}^a \alpha_i^2}{a} = \frac{a-1}{a}\theta_A^2$$

Using this fact, and the results given in the expected mean squares table, produce a "reasonable" estimator for Cohen's f based on expected mean squares, and your knowledge of expected value theory. In case you've forgotten, let me give you an extended "tutorial hint."

Hint. Suppose you are trying to estimate a quantity $\gamma = \sqrt{a/b}$, and for a situation with known constants j and n, you have available to you statistics K, M and P with the following known expected values

$$
\begin{aligned}
E(K) &= \sigma^2 + na + jnb \\
E(M) &= nb \\
E(P) &= \sigma^2
\end{aligned}
$$

Notice that you are looking for a quantity with an expected value of a for the numerator, and a different quantity with an expected value of b for the denominator. If $E(M) = nb$, then $E(M/n) = b$. Notice how you "factor out" a constant to construct an unbiased estimator of b.

Now, how about an unbiased estimator of a? Do you see why $E(K - P - jM)/n = a$? So a reasonable (but not unbiased) estimator for γ would be

$$\hat{\gamma} = \sqrt{\frac{\frac{K-P-jM}{n}}{\frac{M}{n}}} = \sqrt{\frac{K - P - jM}{M}}$$

The technique shown above is used over and over again to derive very complicated and impressive-looking formulas that you see in your textbook. It is based on the fact that in general, for constants a, b and random variables X and Y, if $Y = aX + b$,

$$E(Y) = E(aX + b) = aE(X) + b$$

We essentially derived this rule the first day of Psychology 310.

Usually we think of the rule as "going from left to right," i.e., the formula gives what happens to the mean (i.e., expected value) if we linearly transform a random variable. But it can be thought of in reverse, as we now will see.

An unbiased estimator of $E(X)$ is obtained as follows:

$$
\begin{aligned}
E(Y) &= E(aX + b) \\
&= aE(X) + b \\
E(Y) - b &= aE(X) \\
\frac{E(Y) - b}{a} &= E(X)
\end{aligned}
$$

Now that you have manipulated $E(X)$ alone on the right side, you can apply the "trick" that is used over and over in deriving estimates of fit indices in ANOVA.

I have what I want, $E(X)$, on the right. What I have on the left is a linear transformatiof a random variable's expected value.

Since
$$E\left(\frac{Y-b}{a}\right) = \frac{E(Y)-b}{a} = E(X)$$

I have my desired estimator, i.e.,
$$\frac{Y-b}{a}$$

After deriving your formula for $\hat{f}_A$, produce a similar formula for estimating Cohen's f_B for the B main effect. *Second Hint.* If you are careful, you can convert your answer for $\hat{f}_A$ into one for $\hat{f}_B$ in a couple of seconds, because which factor was called A and which was called B was of course arbitrary, and both are fixed effects factors. In answering the question, you may also wish to look at material in RDASA3, p. 222–224.

(b) (5) Produce the interaction plot for the sample means obtained in the data.

(c) (5) Perform the ANOVA and generate a summary table in R.

(d) (10) You can download an article on effect size estimation (Steiger, 2004) from the publications section of my website. Equation 19 in that article gives the relationship between the noncentrality parameter λ for the F statistic for testing any effect θ, and f^2, the square of Cohen's f.

The formula is
$$f_\theta^2 = \frac{\lambda_\theta}{N_{tot}}$$

where N_{tot} is the total n for the entire design.

You can easily use that formula to get a 90% confidence interval for Cohen's f^2 for the B main effect from a confidence interval for λ_B, and of course, you can also generate a confidence interval for f from your confidence interval for f^2 by taking square roots, because f is never negative. Calculated a point estimate for f_B based on the formula you derived earlier. Compare your confidence interval with your point estimate.

(e) (5) Using your point estimate of f_B, use GPower 3 to estimate the sample size required to obtain a power of 0.80 to detect the B main effect in this situation (i.e., with $a = 4$, $b = 2$)

3

(f) (6) Suppose the A effect was a *random effect* and B a fixed effect. Produce the revised source table, including revisions of the F statistics and p values where necessary.

(g) (6) Suppose the B effect was a *random effect* and A a fixed effect. Produce the revised source table, including revisions of the F statistics and p values where necessary.

(h) (8) Now suppose both A and B represent random effects. Produce the revised source table, including revisions of the F statistics and p values where necessary.

2. (20 points). *Statistical Detective.* Suppose you read an article that provides the following table of cell means in a 2×3 ANOVA design. The cell means are based on a sample size of $n = 6$ per cell. The article tells you that the only significant effect in the design is the AB interaction effect, and the F statistic is 8.00. It turns out with just this information, you can reconstruct the entire source table of Sums of Squares, degrees of freedom, Mean Squares, and F statistics for the entire table. All the information you need is contained in the data provided above and the formulas in RDASA3 Table 9.5.

I'm going to give you a couple of hints to get you started. Realize first that you can easily construct row and column means from the table. From those, you can easily construct SS_A and SS_B from the sum of squared deviations of the row means around the grand mean and sum of squared deviation of the column means around the grand mean, respectively.

SS_{cells}, also referred to as "SS between cells" in Table 9.5, can be easily computed from the sum of squared deviations of the cell means around the grand mean.

SS_{AB} can be computed directly from the information referred to above. From there, you should (looking carefully at the F statistic provided) be able to construct the entire source table.

	B_1	B_2	B_3
A_1	12	16	14
A_2	18	14	12

(a) (10) Using the above hints, construct the entire F table, including the p values for all tests of main effects and interactions.

(b) (10) Perform a test of the simple main effect of B at A_2.

3. (20 points). The data in the file *WileyUnequalN.csv* (available on the course website) are from an example in section 9.8.3 of RDASA3. Using functions like `anova` and `Anova` (from the `car` library), as well as `interaction.plot` and `aggregate`, I want you to do the following:

(a) (5) Generate the 2×4 table of cell means corresponding to Table 9.14 in RDASA3. (You do not have to display the cell frequencies.)

(b) (5) Generate an interaction plot with two different-colored lines representing the two levels of the `Format` factor.

(c) (10) Generate R anova output corresponding to each of the 4 situations below, and verify that results in your tables agree with the results in RDASA3 Table 9.15:

   i. ANOVA using Type III SS
   ii. ANOVA using Type II SS
   iii. ANOVA using Type I SS with factor A entered first in the formula
   iv. ANOVA using Type I SS with factor B entered first in the formula