WINTER 2019/20 TERM 2 EPSE 592: ASSIGNMENT 3
Due: Week 10 in class: Mar. 19th

- Please make sure you write your answers to these questions in your own words. Even if you work with a group to formulate your responses, do not just copy someone else's sentences/words.
- There is no need to record more than 3 decimal places for any of these problems.
- All problem data are available online in .csv format.
- All R code is available online in plain text format. Copy and paste **exactly** as recorded into the Rj editor of Jamovi. If Jamovi does not work on your computer: install base R from

  `https://www.r-project.org/`

  Then open R, copy and paste code **exactly** as recorded, and hit 'enter'. **In either platform, the code may take 10-30 seconds to compile.**

Question 1: (The importance of interaction effects and robustness of ANOVA to departures from normality) In this study, we aim to study the effects of potassium chloride on the size of coastal strawberry (*Fragaria chiloensis*) and kinnikinnick (*Arctostaphylos uva-ursi*), two species of native shrubs. We establish 4 test plots on a one acre piece of land near Bamfield, BC. Two plots are treated with potassium chloride enriched soil, while the other two plots are left unaltered from their natural soil composition. We then choose one treated and one untreated plot and plant 10 strawberry bushes in each plot; we do the same for the kinnikinnick bushes in the remaining plots. We let the plants grow without any further intervention, and at the end of one year, we measure the height of each plant in centimetres. Partial data are presented below (access the full dataset on the webpage).

	Species	Soil	Height (cm)
Plant 1	Kinnikinnick	treated	12.3
Plant 2	Kinnikinnick	untreated	1.8
Plant 3	Strawberry	treated	12.9
Plant 4	Strawberry	treated	2.1
Plant 5	Strawberry	untreated	12.5

Table 1: Partial data for soil experiment.

(a) Examine a boxplot of the response vs. species. Looking at the graphic, does there appear to be a species effect on the average response? Do the data appear normally distributed within each group?

(b) Now examine a boxplot of the response vs. soil treatment. Does there appear to be a treatment effect on the average response? Do the data appear normally distributed within each group?

(c) Finally, examine boxplots of the response vs. species and soil treatment simultaneously. Do you see any signs of a possible interaction between the variables? Do the data appear normally distributed within each group?

(d) Perform a (full) two-way ANOVA on the study data and interpret the output. Does your formal analysis agree with the informal analysis you conducted in parts (a)-(c)?

(e) Perform some model diagnostics. As in class, use the following code in the Rj editor of Jamovi to create a plot of the residuals (observed errors) vs. the fitted values (estimated group effects) and a qq-plot:

```
mod1 <- aov(formula=data$Height~data$Species*data$Soil)
plot(mod1)
```

Assess the homoskedasticity and normality assumptions. How does this assessment mesh with your informal assessments in parts (a)-(c)?

(f) Does a modest departure from normality seem to impede your ability to objectively quantify the effects (or lack thereof) you informally saw in parts (a)-(c)?

Question 2: (Exploring the murky world of statistical power for a simple one-way ANOVA) Suppose you are planning a small pilot study to try to assess if two new kinds of physical therapy techniques have a different effect on patient outcomes compared to the current standard technique. Patients of interest are women aged 30-45 who have given birth to at least one child in the past and suffer from chronic lower back pain. The new physical therapy programs under consideration require the therapist to perform the same physical treatments as currently, but the treatments are now to be performed under one of two targeted temperature gradients over the course of a single therapy session: (1) room temperature (RT) to warm (RT+10°C) to room temperature, or (2) room temperature to cool (RT−10°C) to room temperature.

Your outcome of interest is the patient-reported pain index (a subjective ranking on a 10-point scale) after 4 weeks of therapy, administered twice weekly for one-hour at each session. Under the current treatment regime (patient subjected to no temperature gradient), it is known that target patients report a decrease in their pain index of 2 points on average. You would like to make sure your new study is *powerful enough* to detect if either of the new therapies are: (1) *more effective* than the current treatment by at least 1 pain index point on average, or (2) *less effective* than the current treatment by at least 0.5 pain index points on average. You suspect that the warm gradient therapy will have a larger positive effect than the current treatment, while the cold gradient therapy will have a smaller positive effect than the current treatment.

(a) Explain why designing this study with more power to detect negative patient outcomes (vs. current treatment) makes sense with respect to the general medical principle of *first do no harm*.

(b) After reviewing the known data on patient outcomes under the current treatment regimen, you can reasonably justify that the change in pain index data from start to end of therapy is roughly normally distributed with a variance of 1 among target patients. You think you will be able to recruit 30 target patients in total for this study; you plan to randomly assign 10 patients to each of the three treatment groups (current treatment, warm-gradient treatment, or cold-gradient treatment). The data this procedure generates will then be analyzed via a one-way ANOVA. Reality check: what is the response variable and what is the explanatory factor variable for this one-way ANOVA? What is the target population of this study? Will you be randomly sampling from this target population? This is only a pilot study, but what kind of confounding variables could you imagine might influence your results?

(c) **(Effect of target effect sizes on power)** Suppose we set a type I error rate of $\alpha = 0.05$. Use the R code provided to estimate the power of your study if all other parameters (variances of different treatment responses all equal 1, normality of data) remain unchanged, but instead, the positive and negative effect sizes of clinical interest are smaller: treatment is more effective by at least 0.5 pain index points on average or less effective by at least 0.25 pain index points on average. Which combination of clinical effect sizes are most attractive? Why? [Note: as before, use Jamovi's Rj-Editor to run the R code. Simply copy and paste the text from the

posted R code file into the Rj-Editor and then press the green triangle to run the code. Note that some pieces of code may take 10-30 seconds to compile.]

(d) **(Effect of the distribution of the data on power)** Regardless of your answer in part (c), suppose you decide to stick with the original clinical effect sizes of interest. It is good statistical and scientific practice to check how robust your study will be to departures from your assumptions about the distribution of your data (not yet collected, remember). This process is called *sensitivity analysis*.

i. Use the R code provided to estimate the power of your study if the variance you assumed for the response variable is actually bigger than you expect in the different treatment groups, but all other parameters (e.g. normality of data, effect sizes of interest) remain unchanged. What happens to your study's power as the variability of the response increases across the different treatment groups? Why does this change in power make sense?

ii. Use the R code provided to estimate the power of your study if the distribution of your response variable is moderately skewed in the different treatment groups (this skewness will force the response variable to have a slightly higher variance as well). What happens to your study's power if the response variable is skewed in the different treatment groups? In which treatment group does skewness seem to have the greatest effect on power? The boxplot produced along with this R code shows a case when all treatment groups produce skewed data so that you get a sense of how much skewness we are considering here. Do these responses distributions seem that skewed? Why or why not?

iii. The departures from homoskedasticity and normality we just simulated were moderate. How concerned would you be about your study's power to detect the effects of interest in the presence of moderate heteroskedasticity and/or non-normality?

Question 3: (Blocking, confounding, and repeated measures) You and a neighbouring lab have each recruited 30 arthritis patients to participate in a joint study. A third of patients suffer from mild arthritis (L), a third suffer from moderate (M), and a third suffer from severe arthritis (H). Patients are subjected to a non-invasive massage treatment for 3 months. Both your lab and the neighbouring lab have 2 physical therapists qualified to administer the treatment: one has 10 years experience, the other has only 2 years experience. Patient dexterity is measured before and after the 3 month period.

(a) The neighbouring lab (LAB1) decide that they are uncomfortable with their less experienced therapist (II) administering the treatment to patients with more severe arthritis; thus, they decide to give their Therapist II only 2 random patients from the high severity group, 5 random patients from the medium severity group, and 8 patients from the low severity group. The more experienced Therapist I is then assigned 8 patients from the high severity group, 5 patients from the medium severity group, and 2 patients from the low severity group. Give at least one reason why this study design may lead to suboptimal experimental results. Give at least one reason why this study design may nevertheless be desirable given the context of the study.

(b) Analyze the data from the LAB1 study conducted under the above design (see webpage). Use the "Repeated Measures ANOVA" option with SEVERITY and THERAPIST as explanatory (between-subjects) factors.

i. Do the data suggest marginal and/or interaction effects among these between-subject factors? Within the context of the research problem, interpret what the presence of any suggested non-zero effects means, and why they should or should not be expected to be there.

ii. Do the data suggest an effect of treatment on patient dexterity over the 3 month study period? Does treatment appear to interact with any between-subject factors? Interpret these effects (or lack thereof) within the context of the study.

(c) Your lab (LAB2) is less concerned with the experience level of your therapists and decides to perform a randomized block design for your study, with each therapist (the "blocks") administering the treatment to exactly 5 patients from each severity group. Perform a separate RM-ANOVA to analyze these data (see webpage) using the same model fit above.

i. Do the data suggest marginal and/or interaction effects among the between-subject factors? Within the context of the research problem, interpret what the presence of any suggested non-zero effects means, and why they should or should not be expected to be there.

ii. Do the data suggest an effect of treatment on patient dexterity over the 3 month study period? Does treatment appear to interact with any between-subject factors? Interpret these effects (or lack thereof) within the context of the study.

(d) How do the analyses from the two labs compare? What would explain any substantive difference in observed treatment effect between the two labs? Is there any evidence that a treatment effect is affected by (i.e. interacts with) the different experience levels of the study therapists or the severity of the patient's arthritis in either lab? Produce an interaction plot to illustrate. [Note: you can create these easily in Jamovi by selecting the interacting variables you want to plot in the "Estimated Marginal Means" option dialogue box of the RM-ANOVA menu.]

Question 4: (BONUS (optional) - classical t-tests are equivalent to F-tests on two groups) Recall (see Lectures 4, 5) how a Fisher's F-statistic on $(K-1)$ numerator degrees of freedom and $(N-K)$ denominator degrees of freedom is defined within the context of ANOVA:

$$\frac{MS(treatment)}{MS(error)} \sim F(K-1, N-K),$$

where

$$MS(treatment) = \frac{1}{K-1}\sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y_{\cdot j}} - \overline{Y_{\cdot\cdot}})^2,$$

and

$$MS(error) = \frac{1}{N-K}\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y_{\cdot j}})^2.$$

Here, N denotes the total sample size, K denotes the number of groups, n_j denotes the sample size within group j, Y_{ij} denotes the ith observed response in group j, $\overline{Y_{\cdot j}}$ denotes the sample mean of the response in group j, and $\overline{Y_{\cdot\cdot}}$ denotes the overall sample mean.

When there are only *two groups* being compared via this ANOVA F-test (i.e. when $K = 2$), show that the F-statistic reduces to the square of the classical Student's t-statistic traditionally used to test for a mean difference between two groups. That is, show that

$$\frac{MS(treatment)}{MS(error)} = \left[\frac{\overline{Y_{\cdot 1}} - \overline{Y_{\cdot 2}}}{S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right]^2,$$

where

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{N-2}},$$

and S_1^2, S_2^2 are the sample variances of the observed responses from groups 1 and 2 respectively.

[Hints]: First simplify the expression for the MSE. Then, to simplify the expression for the MST, first expand the sum and then use the fact that we can write the grand mean as follows:

$$\overline{Y_{\cdot\cdot}} = \frac{n_1\overline{Y_{\cdot 1}} + n_2\overline{Y_{\cdot 2}}}{N}$$