To receive credit for this lab, turn this sheet in before leaving the lab.

Name: _____________________________________________

Lab Partners: _____________________________________________

1. What is blocking useful for in research design?

2. What are ANOVA assumptions? Is normality satisfied in this case?

3. Why are the MSwithin values for the separate ANOVAs smaller than for the overall ANOVA?

4. List one or two things that are still unclear to you.
General Lab Procedures: The weekly labs are required, key component of the course. They are your opportunity to get assisted, hands-on practice at applying the concepts introduced in class and/or the textbook. They also serve as a bridge towards doing the homeworks.

The lab handouts contain specific questions and spaces to answer them. Writing down answers to these questions is an excellent way of checking your understanding of the material. I encourage you to write your answers (even though they are not turned in) because this is good practice for homework and exams. Note that the ♠ symbol indicates that the current material relates to a question on the turn-in sheet. Be sure to check the sketch of the lab solutions, which will be available by the end of the week. It is your responsibility to verify that your answers are correct.

You are to work in groups of three. This is an excellent way to learn from each other. Work at a steady pace without rushing. If you do not finish during the lab time, finish on your own time, but turn in the lab sheet at the end of the lab. If you don’t know how to perform a task, refer to old handouts, use the online help system, talk to each other, or ask the instructor.

Goals: In this lab you will learn about 1) some principles of research design, 2) one-way ANOVA and 3) blocking.

Part I: Read the following on Some Principles of Research Design (approximately 10min)

A. “The goal of any research design is to arrive at clear answers to questions of interest [about the populations] while expending a minimum of resources.” –Ramsey and Shafer
   1. Keep it simple / Don’t try to answer everything in one experiment
   2. Be objective / Learn

B. Identify sources of experimental variation
   1. Subject to subject variability
   2. Environmental variability
   3. Treatment application variability
   4. Measurement variability

   1. Have groups differ in only one aspect, if possible
   2. Use a control group, if possible
   3. Use blinding, including a placebo, if possible (to avoid the possibility that differences are due to expectations)
   4. Randomize treatment application to reduce the chance of confounding and to allow for causal inference. (A confounder is anything that changes systematically between groups other than the desired treatment.)
   5. Randomize subject selection to reduce the chance of confounding.

D. Principle 2: Promote broad inference. Prevent your experiment from having limited impact
   1. Can we generalize from one age group to others? race? gender? nationality? education?
   2. Can we generalize from carefully controlled environmental conditions to the real world?
   3. Can we generalize from carefully controlled treatment application to the real world?
4. Randomize subject selection to promote broader inference.
5. Stratif\text{y} the population into separate meaningful groups (age, race, etc.), then sample experimental units from each stratum. Assures that low frequency groups are better represented.

E. **Principle 3: Promote power.** Improve your ability to detect real differences, often by reducing within group error.
   1. **Blocking (♠ 1):** Group similar subjects into “blocks” and randomize treatment application within those blocks. Analyze in a way that “pools” results across blocks. Examples of blocks include grouping by age, gender, intelligence, experience, apparatus, location, etc.
   2. Control the four sources of variation.
   3. In **within-subjects designs** each subject is his or her own control, and the subject-to-subject variation is mathematically isolated, reducing the effective experimental error.
   4. Measure what you can’t control, and appropriately include those measurements in your model as covariates.
   5. Assure that your treatments are strong enough.
   6. Assure that you have enough subjects.

F. **Example.** This study explores the effect of race on perception of trustworthiness. Paid volunteers are recruited with signs on buses. They are told that they will be paid $5 for watching a 5 minute video and answering questions about it. The first 20 black, white and Asian subjects are accepted into the study. Each is randomly shown one of two versions of a scripted video for selling real estate, performed by one of the two researchers. One researcher is black and one is white, and both are 40 year old males who received PhDs in psychology from Harvard. Questionnaires are given that ask objective questions about the video, and ask for subjective rankings of the presenters. The question to “rate the trustworthiness of the presenter on a scale from 0 to 100” is used as the outcome. The explanatory variable is the presenter’s race.

1. Interpretability

2. Broad inference

3. Power
Part II: Problem Description
A psychologist wishes to study the Stroup effect. The experimental apparatus is a set of cards each with the name of a color written with colored ink. Some cards have matching ink and name (concordant cards), while others differ (discordant cards). Our goal is to see whether there is a difference in how much time people take to identify and speak the color when the word is a mismatch.

We will consider the Stroup test for all current members of the U.S. Navy (fake). Our subject population is all current US sailors. Our outcome populations are response times for the concordant condition and for the discordant condition for all current Navy members. Our model is that the outcomes are normally distributed with variance \( \sigma^2 \) in each population, and we name the population means \( \mu_1 \) and \( \mu_2 \). Our null hypothesis is that \( \mu_1 = \mu_2 \) and our alternative hypothesis is that \( \mu_1 \neq \mu_2 \).

Task 1: Import non-SAS data into SAS (approximately 5 minutes)

a. The file navyblock.dbf contains a record of the type of Stroup test performed (concordant=1 or discordant=2) and the average time (in seconds) to correctly name 15 colors, for each member of the US Navy. Additional descriptive columns are present. Randomization to the two test conditions was done separately within blocks defined by age and IQ, under the assumptions that these clearly affect speed of verbal responses. Download the file from the course webpage: http://www.stat.psu.edu/~sesa/stat460 navyblock.dbf.

b. In SAS choose Solutions/Analysis/Analyst/. This will open a “New Project”.

c. Choose Open and choose your “navyblock.dbf”. You should see your data in a spreadsheet.

   a. How many variables are there? Which are measurements and which categorical variables? Do you see variable ‘block’

   b. How many observations

Task 2: EDA (approximately 5min)
Hint: “Statistics/Descriptive/Summary Statistics”
Perform the following EDA: Frequencies for block and scatter plots of the outcome vs. age and outcome vs. IQ, both with test type, and graphs and statistics for outcome by test type (using “Graphs/Boxplot” and “Graphs/ScatterPlot”). Use ‘block’ in the “CLASS”. Interpret what you see. Say something about the center, spread, shape, outliers.
Task 3: **ANOVA and blocking (approximately 20min)**

a. Before performing the ANOVA, state the ANOVA assumptions (2).

b. Is normality satisfied? Plot Q-Q plot or Normal plot (“Graphs/Probability Plot” and make sure Normal is for the distribution). Do you think your data satisfy normality assumption?

c. Now perform the one-way ANOVA with “test type” as the factor (ignoring block), “Statistics/ANOVA/OneWayANOVA”. What does $MS_{within}$ always estimate? When does $MS_{between}$ estimate the same quantity? Here is the approximate sampling distribution of the F statistic under $H_0$.

![Sampling Distribution](image)

If your sample happens to have 98 denominator df, then the “critical F value” is 3.94. This indicates that 95% of the time (under the null) F values will be less than 3.94 and 5% of the time they will be above. In other words, before looking at the results, we can state that “if the F statistic is greater than 3.94 then the p-value will be less than 0.05”. Mark your F statistic on the graph, and estimate the area to the right which is your p-value (the entire area equals 1.0). Interpret the meaning of your result.
d. Now let’s split the data in SAS. Choose “Data/Split Column”. In “Split Column” put ‘stroup’ and in ‘Split By’ put ‘block’. Save this new data as a SAS file and open it in the project to do a new analysis. Now rerun ANOVA. You will get separate ANOVAs for each block. IF the blocking is effective the MSwithin values for the six separate ANOVAs will be smaller than for the combined ANOVA. Why (♠3)? The appropriate analysis on blocked data will give more power to detect a true Stroup effect than the unblocked data because of the reduced ‘error’ reflected in MSwithin.