**Statistics 101A-Phase 2**

**11-21-2012**

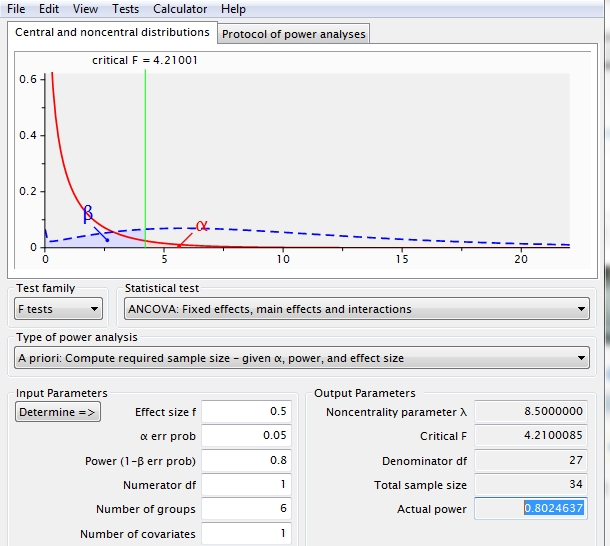
**Group members:**

Name: JiaLin Huang Email: vincenthjl@yahoo.com

Name: Wenting Li Email: mosquitowendy56@yahoo.com

Name: Junjie Qiu Email: ismyucla@g.ucla.edu

**(a)**



Firstly, we need to decide the sample size of our experiment. We will do it in G power.

**Sample size: 36 subjects**

> sd(testform$vocabulary)

[1] 2.039296

According to our data, our standard deviation is roughly 2. When we observe the interaction plot graph, the mean difference is expected to be 1 words as we treat 1 that is the significant increase in memory test after drinking coffee. This means that the effect size that we are looking for is 0.5.

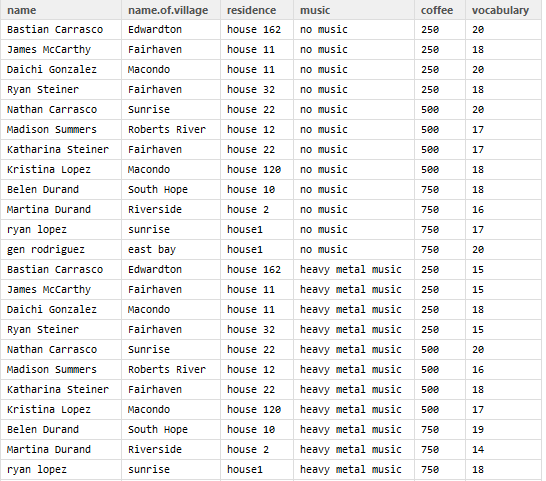
However, we want the effect size to be small, because we want to be able to determine if our experiment work even a little bit. And also, we want the sample size to be small so that we do not need to spend too much time on too many subjects. Eventually, we set our effect size on d=0.5 and power=0.8. Then, we will have approximately 34 subjects in total according to the G power calculation.

After computing the power calculation, we decide to have 36 samples in total, and each block has 6 samples randomly assigned. According to our preliminary data, the required sample size is strongly correlated to our pre-analysis.

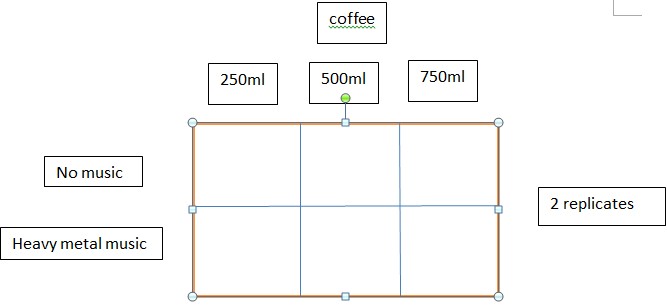
What’s more, since we have the sample size=36, we believe the project can be carried out in the time allowed. With the effect size 0.5, standard deviation in 2, we will have an 80% chance of seeing an effect of 1 word or more.

**(b)**

Our sample data were collected from the website home of the Islanders, which is a virtual population to support learning and teaching in experimental design, epidemiology and statistical reasoning. On this virtual island, there are thirty-nine villages with a population of over eight thousand Islanders. Islanders are the subjects for our experiment. Different islander has different information such as location, age, gender and health condition, which make our data look more real. We looked for a subject by randomly going to a village and a house to obtain the consent of a randomly islander. If the islander agrees to be tested, we will set task for him such as giving him [Heavy Metal Music](http://island.maths.uq.edu.au/task.php?id=2drry&code=musicmetal), some amount of coffee and [Memory Test Vocabulary](http://island.maths.uq.edu.au/task.php?id=2drry&code=memoryvocab) before and after drinking coffee. After the islander finishing each task, we will collect the relative data such as islander’s name, residence, name of village, amount for coffee gave him, whether give him Heavy Metal Music and the result of Memory Test before and after drinking. In addition, we put them in excel and combine these data in a table below.



In the procedure of collecting data, we use randomly assignment. Firstly, for our project proposal, we decide to collect data from different 12 islanders and we make sure that these 12 islanders are from different locations and have different ages. In our experiment, we have treatments coffee and [Heavy Metal Music](http://island.maths.uq.edu.au/task.php?id=2drry&code=musicmetal). So, we randomly select 4 subjects in our sample, gave them difference amount of coffee. And also, we replicate all the subjects in heavy metal music group which also require taking different amount of coffee. That is, after finishing the music test, we also randomly select 4 islanders and given them 250ml of coffee from the whole sample 12 islanders. In the same time, we gave 500ml of coffee to the following 4 islanders and 750ml coffee to the other 4. Totally, we will end up with 24 data in our preliminary experiment. In this way, we make sure that each level of coffee has been randomly assigned. What’s more, we think our blocking used appropriately. And we did it like in the following graph.



**(c)**

In addition, in our experiment, we will exclude some types of people and condition. For example, we collect the data only from the healthy islanders. In other words, we exclude the subjects who are pregnant or have severe disease like Glycuresis. Also, we don’t want to data from the subject who is too young and too old since the test may be harmful to the subjects. Therefore, we try our best to collect the subject whose age between 10 and 60.

What’s more, the reason why we want to exclude them is that it will affect our result. In the procedure of our experiment, we found that if our data include those subjects we want to exclude above, we would obtain large standard error. Also, due to this large standard error, it may affect the effect size and sample size we choose. It is not a good new both to us and the investors since it will require us to have a large sample size, which means more work to do and more cost! That’s why we want to exclude them.

**(d)**

i) In order to random assign the treatment factors to all 12 observations, we give each observation a number from 1 to 12. Since there are three levels of treatment factor-caffeinated drink (coffee) like 250 ml, 500 ml, and 750 ml, we will shuttle all subjects randomly into three groups, and then randomly assign the three levels to the three groups. For the other factor heavy metal music, we offer the 12 observations to listen to the heavy metal music and not listen to the music before they drink coffee, which can reduce the residuals because we test on the same sample subject.

ii) The important part of the experiment is how we apply the treatment factors to all observations. After the shuffle and random assignment, the 12 observations are equally divided into three groups for three different levels of coffee. First, we play heavy metal music to all of them, and then we offer them the assigned amount of coffee they need to drink (250ml, 500 ml, and 750 ml). We will wait till 30 minutes and then offer them memory test of vocabulary, which let they try to memorize 20 words in one minutes and later test those words they can recall in 30 seconds. We create a table to record the number of words they still can memorize. Similarly to the process, we offer 12 observations their assigned amount of drink and do not offer them to listen to heavy metal music. After this, we will wait 30 minutes to offer them memory test of vocabulary again to see how many words they still can remember. After all data collection, we create a full table with observations’ names, locations, residence and treatments we offer them in detail for future comparison.

iii) According to the cause and effect of the experiment, there exist some nuisance factors that will make a difference one our memory test response. But we can try to reduce their effects in the design of the experiment in order to draw a more representative conclusion and reduce errors. The nuisance factors are like age, gender, pre-existing health conditions, personal habits, family history and education levels. That is why when we design the experiment, we try to use each subject as a block and compare their memory test with and without heavy metal music at a certain level of coffee to reduce the errors caused by the differences between each individual observation. When we shuffle and random assign the three levels of coffee, we try to randomize the assignment in order to reduce the effect of the nuisance factor as well.

iv) The design of our experiment is try to show the effects of heavy metal music and coffee on the memory test. Since nowadays, many people think music and coffee can encourage them to work more effectively, so we design our experiment to see whether heavy metal music or amounts of coffee will help them to memorize more stuff. We mainly focus on the two treatment factors heavy metal music and coffee, and there are three levels (250 ml, 500 ml, and 750 ml) within the treatment factor coffee. All different people can follow the process of our experiment mentioned above to carry out the same experiment easily. After revision of the experiment, our design is appropriate and simply tests on the effect of heavy metal music and coffee levels, which match our research question. And we offer clear assignment conditions to the observations, which make the experiment easy to follow.