Biostatistics - Dr. Patrick
BMEN 350; Section 201
Saurabh Dhole 626 002 135
October 29th, 2021

Project 3

A. Read and Reflect

Upon reading the knowledgeable article called "Power and sample size" by Dr. Krzywinski, I was intrigued upon learning about the impact of sample size on different types of experiments. I learned about how experiments with many conditions and experiments with multiple outcomes are negatively affected by low power. Additionally, I learned about some of the causes of underpowered studies. I would like to reflect on my observations here.

In the article, Dr. Krzywinski makes it clear that the ability to discern experimental effects is reduced when a study lacks power. Further, Dr. Krzywinski emphasizes that statistical power is very important but it is sometimes overlooked when designing experiments. I observed that Dr. Krzywinski supported these statements by giving the example that in experiments with many conditions and outcomes (omics studies), large portions of the significant results may indeed be incorrect. This example that Dr. Krzywinski shows is in direct support of the claim that when power is low, important effects may not be detected. I later learned that omics studies are studies done in the fields of genomics, epigenomics, transcriptomics, proteomics, and metabolomics. As a Biomedical Engineer, the information discussed here about how low power can result in significant results from omics studies not being detected, is crucial. I say this because I foresee myself having an increased involvement in omics studies in my undergraduate research position. In this role, I will have to be careful so as to design experiments with sufficient power so that significant results from the omics studies that I am apart of will not be undermined (in the case that the principal investigator asks me to help design a study). In a general sense, this information about how low power can result in significant results from omics studies not being detected, is indeed important for me. I say this because I foresee myself having to participate in omics studies in industrial undergraduate internships. During these industrial internships my supervisors may ask me to analyze the effectiveness of genomics studies that were done. In such situations, I will be sure to apply my new found knowledge of power as it relates to omics studies. I plan to hold on to this article and annotate this section of where it discusses the example of omics studies so that I can revisit it during my research and industrial roles.

Further in Dr. Krzywinski's article, it is discussed that in experiments with multiple outcomes such as in gene expression experiments, there are occasions when only less than 10% of the outcomes may have a chance of having an effect. Dr. Krzywinski further explains that even at the conventional power of 0.8, more than one third of positive results may be wrong! This information was certainly very riveting for me as it implies that if studies are not designed with power in mind, a great chunk of the results can be utterly useless and incorrect. As a Biomedical Engineer, this information that in gene expression experiments large portions of the results can be incorrect, is indeed very important for me. I say this because the work that I do in a research lab largely revolves around genomics. The principal investigator has considered providing me with a more intense role in the upcoming years. This role involves me helping to design experiments. In these scenarios, I must be able to identify if a potential study design has enough power to yield significant results. In a more general sense, this information is essential for my understanding on how sample size and effect size affect the power of a study. I plan to apply this information to module projects and module quizzes in the BMEN 350 course.

It is evident from the above paragraphs that Dr. Krzywinski has discussed in his article, how power can affect different types of experiments, from omics studies to gene expression experiments. I was intrigued upon observing that Dr. Krzywinski has also described some potential causes of underpowered studies. These causes for underpowered studies include fiscal constraints on experimental design and commonplace lack of statistical rigor. As a Biomedical Engineer, this information is very important for me. I say this because in my experiences during undergraduate research and during internships, I noticed that fiscal constraints and lack of statistical rigor were indeed present. I do not know for certain that these limitations caused our studies to be underpowered, but I will remain vigilant from here onwards so that such fiscal constraints and lack of statistical rigor do not adversely affect the power of our studies. If I observe such a situation, I will be sure to notify the principal investigator. In a general sense, this information is important for me as I have always have been daunted by the thought that the experiments that I participate in may not be yielding significant results. I find relief in the fact that I can watch out for things like fiscal constraints, and lack of statistical rigor in the environments that a study is conducted in, so as to be able to raise an alarm of an underpowered study. I plan to hang on to and apply this information in my own research and internship experiences so that I can discuss with investigators and supervisors about the power and effectiveness of a study that is being conducted.

After analyzing the thought-provoking article called “Sample Size: More Than Calculations” by Dr. Parker, I was fascinated upon learning about the approach of calculating sample size, the examples of experiments that were used to calculate sample size, and how sample size should be interpreted. I would like to reflect on my observations on Dr. Parker’s publication here.

Firstly, during the introduction of the article, Dr. Parker provides a clear, chronological, 6-step list (a-f), that outlines the technique of arriving at a comfortable sample size for the experimental design at hand. This sample size determining technique provided by Dr. Parker includes elements such as the necessity to specify study design characteristics such as parallel group, cross-over design, with or without baseline data. I observed that Dr. Parker takes great care in providing more of such guidelines in this 6-step list. As a Biomedical Engineer, this information is indeed very important for me. I say this because I have held quality engineering internship positions in the past and I plan on participating in more catheter R&D internships in the future. In these internships I foresee my supervisors prompting me to calculate and provide sufficient sample sizes for the catheter studies that are conducted in the labs. This is why as a Biomedical Engineer; this information is very important for me. In a more general sense, this information is essential for my understanding of sample sizes and their effects on statistical studies. I say this because the list that Dr. Parker has provided has allowed me to really ponder on how to go about finding an effective sample size. I plan on hoarding and utilizing this information about calculating sample sizes in my research experiences, internship experiences, and in projects in the BMEN 350 course.

Dr. Parker then goes onto explain how exactly the above list of how to calculate sample size is applied in real experiments. Dr. Parker does this by exposing the reader to a pilot study on autoimmune disease. The autoimmune disease pilot study clearly shows the application of the above techniques on calculating sample size. Another example of how the technique of calculating sample size is applied is shown in the phase I and phase II studies in oncology example. I observed that I was able to easily understand the technique of calculating sample size by taking a look at these two examples. As a Biomedical Engineer, this content of applying the sample size calculation technique in unique examples is indeed very crucial for me. I say this because I foresee myself performing clinical research studies in the near future and it is very beneficial for me to actually see the techniques of calculating sample size being applied, so that I may follow these techniques in practice. These examples serve as a blue print for

how to apply these amazing sample size calculating techniques. In a more general sense, this content is essential to my understanding of how sample sizes are selected. I say this because in my daily life, for example when I am watching the news, I am presented with data on Covid-19 positivity rates, vaccination rates and so on. I worry sometimes that I am gullible and I take this data at face value. However, If I remember Dr. Parker's examples on how the techniques of how to calculate sample size, I can analyze the data that I have been presented with instead of being gullible. I plan on annotating and further analyzing these examples so that I can better calculate sample sizes in BMEN 350 module projects.

In the article, Dr. Parker made a statement that I thought was very intriguing. Dr. Parker stated that sample size should not be viewed as a number that is right or wrong, but instead sample size should be used as a factor to evaluate the utility of a study. I observed that this statement by Dr. Parker clearly gave me a better conceptual grasp of how to interpret sample size. As a Biomedical Engineer, this statement by Dr. parker is paramount for my studies and future research experiences. I say this because I now see sample size as something that is similar to statistical power in that it can allow us to assess the usefulness or effectiveness of the results of a study. I did not think of sample size in this way before. This way of thinking can be useful for me during quality engineering internships and future research projects. In a more general sense, this statement on sample size is indeed valuable for me. I say this because my colleagues in the research lab like to chastise me on the small sample size, or "n number" as they call it, of my personal project experiments. I can now explain to them and the principal investigator that I now understand the importance of sample size on the ability to assess the utility of my studies. I plan on saving this quote by Dr. Parker so that I can remember the value that sample size has on assessing the utility of statistical studies. I also plan on remembering this quote when working on BMEN 350 module projects so that I can predict or make a guess about the utility of a study based on the sample size.

B. Problem 1

The G Power Software extension (please see Appendix) was used to find the minimum number of animals needed to determine if there is a difference between the drug candidates and the placebo and between each of the drug candidates. If this study was carried out fully, the null hypothesis would be that there is no difference between the drug candidates and the placebo and between each of the drug candidates. Accordingly, if this study was carried out fully, the alternative hypothesis would be that there is indeed a difference between the drug candidates and the placebo and between each of the drug candidates. The **power level** was set to 0.8. This implies that 80% of the time we should receive a statistically significant result. Further, this would imply that the probability of finding a difference between each of the drug candidates and the placebo and between each of the drug candidates is 80%. However, 20% of the time we will not receive a statistically significant result. Or in other words, 20% is the probability that there would not be any difference between the drug candidates and the placebo and between each of the drug candidates. The **alpha** was set to 0.05 and the **effect size** was set to 0.25. The total **sample size** that resulted was 275. When we take into account the 5% **attrition** rate for the animals, we must order **289 animals**.

If we order 289 animals, the price comes out to $115,600. This is obviously greater than our strict budget of $100,000. Since each animal costs $400, we do the following dimensional analysis to know how many animals we can order: $100,000 / $400 = 250 animals. We can order a maximum of 250 animals.

Let us keep constant the **alpha** of 0.05 and let us keep constant the **attrition** rate of 5%. Let us also keep the same **effect size** from above of 0.25. Let us decrease the **power level** from 0.8 to **0.7**. The

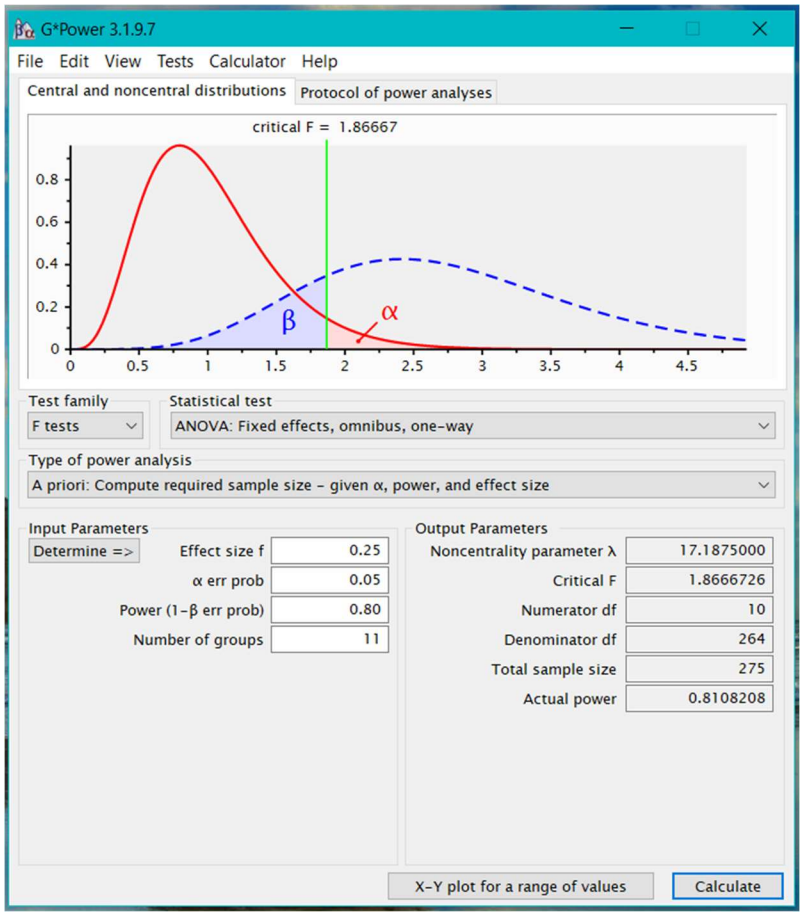
new resulting sample size came out to 231. When we take into account the 5% **attrition** rate for the animals, we must order **243 animals**. The price of ordering 243 animals is $97,200. This new price is well within the strict budget of $100,000.

There are indeed ramifications for staying within this strict budget. The ramification is as follows: **the decreased number of animals comes at a cost of reduced power to the study**. In order to reduce the number of animals ordered from the initial 289 to the new 243, the **power** had to be reduced from 0.8 to 0.7. This means that we should expect to receive a statistically significant result only 70% of the time instead of 80% of the time. Or in other words, 70% is the probability that we find a difference between each of the drug candidates and the placebo and between each of the drug candidates.

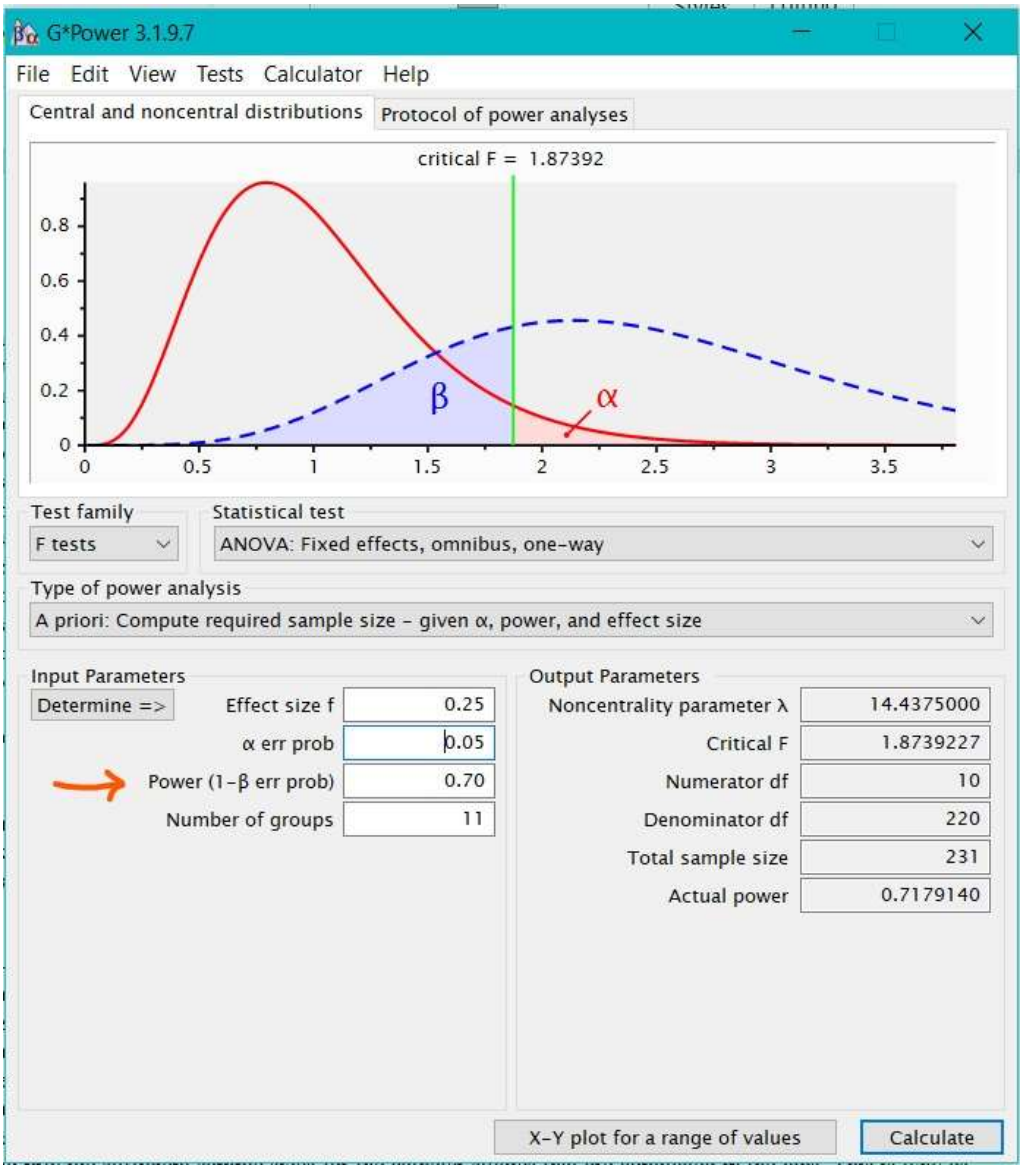
Some of the other parameters can be weakened as well, for example the **alpha** value can be increased to 0.10 while keeping the **effect size** at 0.25, and while keeping the **power level** at 0.80. Increasing the **alpha** from 0.05 to 0.10 has the same effect on the sample size as reducing the **power level** from 0.8 to 0.7 above (please see Appendix). The sample size that is received when the **alpha** is increased from 0.05 to 0.10 is 231 animals. Taking the 5% **attrition** rate into account, we must order **243 animals** in this case. This is another ramification of trying to reduce the number of animals in order to stay within the budget. **The decreased number of animals comes at a cost of increasing the probability of incorrectly rejecting the null hypothesis and decreasing our confidence level.** This is because we had to increase our **alpha** from 0.05 to 0.10 in order to conform with the budget.

C. Appendix

Problem 1:

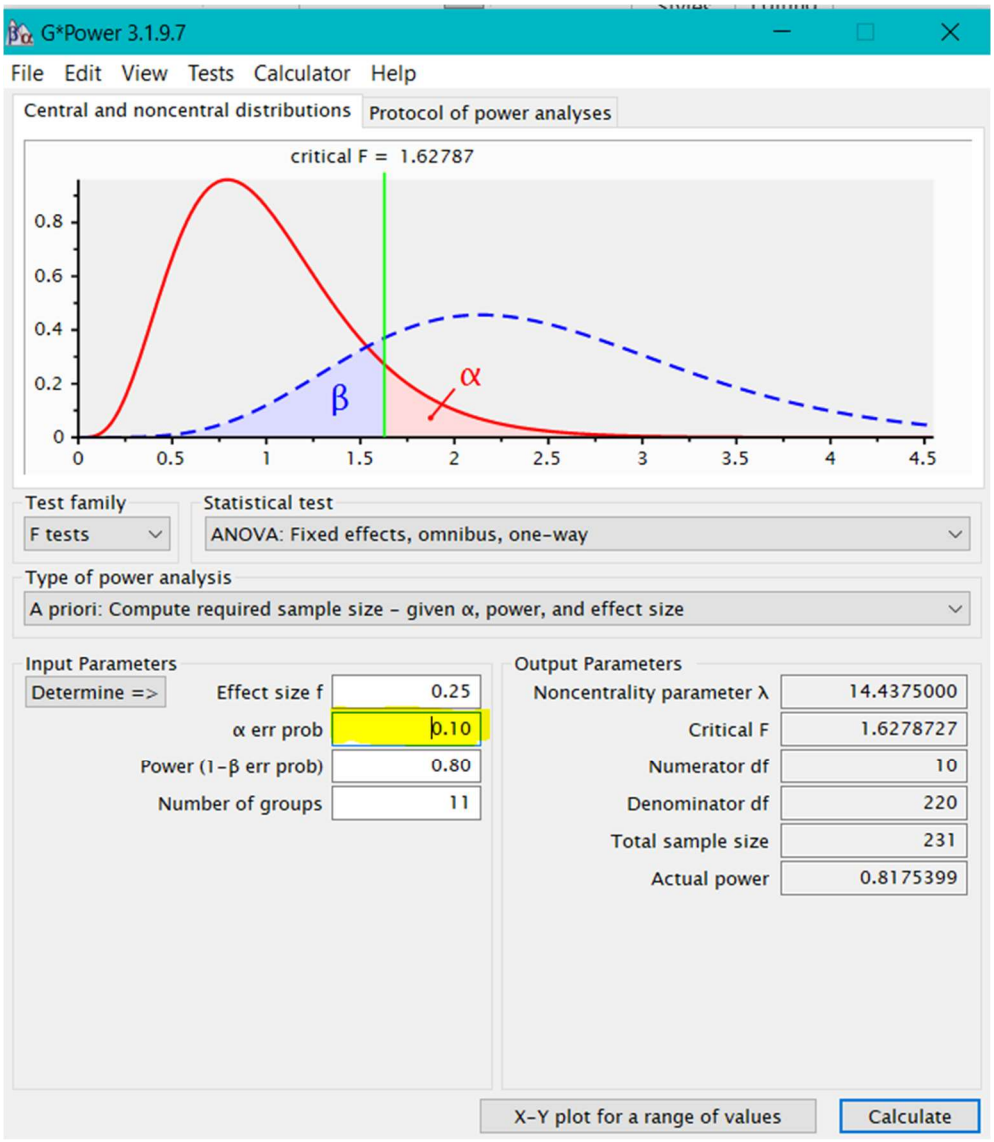


Result when **power** was set to 0.8, **alpha** set to 0.05, and **effect size** set to 0.25. Resulting sample size = 275. Multiply 275 by 1.05 to get 289.



Result when **power** was set to 0.7, **alpha** set to 0.05, and **effect size** set to 0.25. Resulting sample size = 231. Multiply 231 by 1.05 to get 243

Please see next page ….



Result when **power** was set to 0.8, **alpha** set to 0.10, and **effect size** set to 0.25. Resulting sample size = 231. Multiply 231 by 1.05 to get 243

Please note that changing the effect size to 0.40 also reduces the sample size !!!