Cholesterol levels and living altitude

Learning Goals:

- Understand statistical power and how it is impacted by sample size, variability within groups, number of groups, and significance level
- Use statistical power analysis to plan the sample size of a study

Background: Elevated levels of low-density lipoprotein (LDL) (often referred to as “bad” cholesterol) are believed to put a person at risk for several health issues, such as heart diseases. While researchers have found evidence that diet, exercise, and other behaviors (e.g. Pedersen and Saltin, 2006) may affect LDL levels, what about factors such as where you live? Are there differences in LDL levels between people who live at different altitudes?

Suppose that you are planning a study where you will investigate whether and how the altitude at which you live (sea level versus mountains) is associated with LDL levels. Sea level-living will be defined as 1000ft or less, while mountain-living will be defined as living at 2500 to 3500 ft. above sea level. Only people who have lived at the altitude of interest for 10 or more years will be included in the study. LDL levels will be measured via a fasting blood test, and recorded in mg/dL. You will use a significance level of 5%.

How many people should you plan to have in your study? Let’s investigate this question by conducting a statistical power analysis.

1. Identify the explanatory variable and the response variable, and for each identify whether it is categorical or quantitative, as well as units of measurement if it is a quantitative variable.

2. Explain why this would be an observational study rather than a randomized experiment.

3. State the null and the alternative hypotheses for this study.

4. Will the study use random sampling? If yes, then describe how. If not, describe how would choose the study participants.

5. Will the study use random assignment? If yes, then describe how.

6. What will be the Type I error rate in your proposed study? Explain, in context, what a Type I error would represent in your planned study.

7. What size p-value will you need in order to declare a statistically significant result?

But, how large of a sample size do you need in each group (that is, how many people living at low altitude and how many living at high altitude) in order to get a convincing p-value? To investigate this question, let's go to the **Comparing Two Populations** applet. Note, this applet simulates random samples from different theoretical populations (rather than random shuffling).

In this applet, we can explore how our study might play out if we were to run the study as planned, assuming certain characteristics of the populations

Let's assume that (a) those live at sea level the mean LDL cholesterol is 70 mg/dL **and that it is the same for those who live in the mountains**, (b) that the standard deviation of LDL in both populations is 5mgl/dL and (c) that we will have 40 people in each sample.

8. Which hypothesis (null or alternative) as we assuming to be true?

Enter this information into the applet as population model 1 (sea level) and population model 2 (mountains) and leave the population shapes set on Normal. Check the **Show sampling options** box in order to set the sample size to 40 (40 in each group), and press **Draw samples**. The applet has selected 40 random people from the population at sea level and 40 random people from the population living in the mountains.

9. What mean did you get for the sea level group? Standard deviation? What mean did you get for the mountain group? Standard deviation? How different are the means?

10. Now, take 999 more samples. The graph all the way to the right shows the difference in the two sample means for each of the 1000 samples you took. At what value is this distribution centered? Why does that make sense? What is the standard deviation?

11. How large of a difference in sample means do you need in order to be in the top 5% of the distribution? (*Hint*: You will have to try different numbers in the "Count Samples greater than" box until you find a percentage that is close to 5% but below 5%. Use two decimal places.)

12. Based on your answer to the previous question, if you actually did this study (surveying 40 people who lived at sea level and 40 people who lived in the mountains), and got a difference of 3 mg/dL on average between the two sample means, what would your conclusion be? How would your answer change if the difference in sample means was only 1.5 mg/dL? Why?

You should find that you need the difference in sample means to be roughly 1.85 mg/dl or larger to reject the null hypothesis at the 5% level of significance. The region *difference in sample means* $\geq$ *1.85* is called the ***rejection region***. Using this rejection region, the probability we will make a Type I Error should be at most 0.05.

Definition: The ***rejection region*** tells us how different the sample statistic (e.g., difference in group means) need to be in order for us to reject the null hypothesis at a specified level of significance.

13. Suppose we instead planned to sample 80 people in each altitude. How would this change in sample sizes impact the probability of making a Type I error?

So, what does the rejection region tell us about how large our sample sizes need to be? In order to answer this question, we have to make an assumption about what the true difference in average LDL between the two living altitudes is, and how willing we are to make a Type II error.

14. Explain, in context, what a Type II error would represent in your planned study.

What if, in reality, the average LDL for people living at sea level is, say, 3 mg/dL more than the average LDL for people living at the mountain level. How often would we take random samples from these populations but make a Type II error? How often would we make the correct conclusion?

15. Change the population mean for population 2 to be 67 (this is if people living at the mountain level had an average LDL of 67 mg/dL compared to people living at sea level who had an average LDL of 70 mg/dL – for a difference of 3 mg/dL in the two means). Take a single sample of size 40 from each group. How different are the group means?

16. Now, take 999 more samples of each group. What is the mean of the differences in group means across the 1000 (pairs of) samples? What is the standard deviation? How do these values compare to the ones we obtained earlier when we assumed people had the same average LDL regardless of living altitude? Why do these new values make sense?

17. On this new distribution, how often did you get a difference in sample means that falls in the rejection region you found earlier? Interpret this percentage.

Your answer to the previous question, the percentages of samples that would convince us to reject the null hypothesis that the population mean LDL levels are the same (which in reality they differ by 3 mg/dl) is called the ***statistical power*** of the study.

Definition: The ***statistical power*** of a study is the probability that the researchers will find evidence against the null hypothesis and in favor of the alternative hypothesis when the alternative hypothesis is true.

To see that this is what we just found, remember that we (a) first found that we will reject the null hypothesis and find evidence for the alternative hypothesis if the difference in group means is more than about 1.85 and then (b) found that differences larger than 1.85 will happen approximately 84% of the time if the average LDL for the sea level population is 70 mg/dL and 68 mg/dL for the mountain population, a genuine difference of 3 mg/dl

18. What is your estimate of the Type II error rate in this study? How does this value relate to the statistical power?

Factors that Impact Power

There are numerous factors which affect the power of a study. Once decisions are made about these factors, we can determine the sample size necessary to achieve the desired level of power.

Sample Size

What would happen to the power if we plan to sample 30 people from each population?

19. Start by determining the rejection region for a 5% level of significance. How large will the difference in sample means have to be to reject the null hypothesis (p-value $\leq$ 0.05) with samples of 30 people from each altitude's population? (Remember to set both population means at 70 to find the rejection region first.)

20. Now, use the rejection region to estimate the power of the study to detect a difference of 3 mg/dl.

21. What is the relationship between power and sample size? Why does this relationship make intuitive sense?

Size of the actual difference/How wrong the null hypothesis is

What if we thought the actual difference in the population means was 1 mg/dL? How would this impact the power of the study? Start by sampling 40 people from each population again.

22. What is the power if the actual difference in the population means is 1 mg/dL?

23. Did a smaller assumed difference in population means change the power? How so? Explain why this relationship makes intuitive sense.

Choice of Statistic

[Type here]

The *F*-statistic is a standardized statistic with a known theoretical distribution. Instead of comparing the difference in means, let's now use the *F*-statistic to estimate the power.

24. Use the applet to find the rejection region for the *F*-statistic assuming that the mean LDL is the same for both populations (70 mg/dL), the standard deviation is 5 in each population, and you plan to sample 40 people from each population. (*Hint:* Use the Statistic pull-down menu to select *F*-statistic above the distributions of sample means.)

25. Use the rejection region to estimate the power of detecting a difference of 3 mg/dL in the population means with the *F*-statistic.

26. How does this power estimate compare to the one you found in #17 when using the difference in means? Why is it a bit different?

Significance Level

What if we were more concerned about Type I errors and so wanted to set the Type I error rate to 1%? Suppose we plan to sample 40 people from each altitude, the standard deviation is 5 mg/dL in each population, and we are back to using the difference in means as our statistic and a one-sided test. (Don't forget to find the rejection region with a Type I error rate of 1% first.)

27. What is the power to detect a difference of 3 mg/dL (using the difference in means) with a 1% level of significance?

28. How does the power change when the Type I error rate decreases? Explain why this relationship makes intuitive sense.

Population Standard Deviations

To this point, we've been assuming that the standard deviation of LDL measurements is 5 mg/dL for both populations. But what if it isn't?

29. Find the power of detecting a difference of 3 mg/dL if we sample 40 from each altitude, but the standard deviation is 2.5 mg/dL in each population. (Don't forget to find the rejection region when SD is 2.5 mg/dL in each population and mean LDL is 70 mg/dL for both altitudes first.)

30. How is power affected with smaller conjectures for the population SDs?