ADS2 Week 2.4. Problem Sheet
Power and Sample Size

Rob Young (based on work by Xianghua Li and Wanlu Liu)

robert.young@ed.ac.uk

Question 1. Two-sample sample size estimation

An investigator is planning a study the association between coffee consumption and average grade point among college seniors. The plan is to categorize students as heavy drinkers of coffee and rest using 5 or more cups of coffee on a typical day as the criterion for heavy consumption. Mean grade point averages will be compared between students classified as heavy drinker versus non-heavy drinkers, using a two-sample unpaired test of means. The standard deviation in the grade point averages is assumed to be 0.42 and a meaningful difference in grade point averages (relative to coffee consumption status) is 0.25 units. How many college seniors should be enrolled in the study to ensure that the power of the test is 80% to detect a 0.25 unit difference in mean grade point? Use a two-sided test with a 5% level of significance.

Question 2. The relationship between statistical power, sample size, significance level and effect size.

- Take a two-sample, two-sided test, with significance level of 0.05. The standard deviation of our sample is 0.5. If our sample size is 20, what's the statistical power under an effect size (mean difference of two population) of 0.4?
- What happens to our statistical power (does it increase or decrease) if we increase our significance level to 0.1?
- What happens to our statistical power (does it increase or decrease) if we decrease our sample size to 10?
- What happens to our statistical power (does it increase or decrease) if we increase our effect size to 0.8?
- Advanced challenges (optional). Based on the example from above, can you use simulation and plotting to figure out the relationship between statistical power vs significance level, sample size, and effect size? You should draw three curves with y-axis being the statistical power and x-axis being the different significance level, or sample size, or effect size (simulate 100 data points for each plot).

Question 3. The relationship between sample size and p-value

- First, for sample A, generate 5 random numbers from a normal distribution with mean of 10 and sd of 5. Then, for sample B, generate 5 random number from normal distribution with mean of 11 and sd of 5. Now we want to compare whether there is

any significance difference between the mean of sample A and B, what should we do? write out the R code. Is there any significant difference for the mean of sample A and B? Write out the R code. Is there any significant difference for the mean of sample A and B?

- What if we now increase the sample size to 500 (instead of 5) for sample A and B? Is there any significant difference for the mean of sample A and B? Write out R code.
- Advanced challenges (optional). If you can figure it out yourself, you are an absolute 'R master'! Just like we did previously, can you use simulation and plotting to visualize the relationship between sample size and p-value? Plot out the curve with p-value on the y-axis and different sample number on the x-axis (with at least 100 datapoints).