Week 2 Problem Set: Non-parametric methods

(30 pts)

Question 1 (8 pts)

Assume survival data (days from treatment to death) from two groups, a treatment group (=20) and a control/placebo group (=15) as follows:

Treatment = (90, 91, 68, 90, 167, 26, 38, 89, 88, 99, 41, 123, 76, 88, 79, 96, 79, 122, 11, 23)

Control = (52, 104, 146, 27, 46, 120, 5, 15, 11, 48, 30, 40, 8, 42, 74)

1. Use a permutation method to test the null hypothesis that the treatment does not change survival time. Report the p-value associated with the two-tailed test of the null hypothesis. If the critical value , would you reject the null hypothesis? (Would the observed difference be considered significant [i.e. reject the null hypothesis] if we set ? What about ?) (6 pts)
2. Given the context for the analysis, is 0.05 the best cut-off for significance? Why or why not? What might be an argument for using a different cut-off for significance? (2 pts)

Question 2 (16 pts)

In Question 2, we are going to answer this question in a different way, using nonparametric bootstrap rather than permutation. I am asking you to calculate confidence intervals for the difference in means between the two groups using bootstrap sampling. We can do this as follows:

Pseudocode:

1. Sample with replacement for Treatment and calculate the mean of that bootstrapped sample. Store that in a vector.
2. Sample with replacement for Control and calculate the mean of that bootstrapped sample. Store that in a vector.
3. Take the difference between these two means. Store that in a vector.
4. Start back at Step #1, and loop through some number of times.

The distribution for the difference follows

(NB: We have assumed independence between the two groups of data.) That means that the standard deviation of our distribution for should be approximately .Make sure this is at least approximately true for your bootstrapped samples.

a) (2 pts) Using your bootstrap samples, what is the standard error of the mean for the Treatment group (). How does this compare to ? Approximately how many bootstrap samples are required to get a decent estimate (you define what constitutes “decent”)?

b) (2 pts) Using your bootstrap samples, what is the standard error of the mean for the Control group (). How does this compare to ? Approximately how many bootstrap samples are required to get a decent estimate (you define what constitutes “decent”)?

c) (4 pts) Using from 2a and from 2b, calculate the standard error of the difference in means () assuming the two groups are independent. NB: In other words, and . As explained in some detail in the “Algebra of Expectations”, *variances* add. (You cannot simply add the standard errors.) Therefore , and .

d) (4 pts) Use from part c to estimate the 95th percentile confidence interval on the difference in means, assuming that the interval is given by (mean-,mean+). (We will discuss where the 1.96 comes from a bit later.)

e) (4 pts) We're now going to answer for 2d in a slightly different way using the percentiles of the bootstrapped samples. Calculate a 95th percentile confidence interval for the difference in mean survival days () based on the quantiles of the bootstrapped samples.

[Side Note: Both method d and e use the bootstrapped estimates of standard error but they use them in different ways. In part d, we assumed normality of both Treatment and Control means to work out what the CI on the difference might be, whereas in part e, we empirically worked out the distribution of the differences. In most applications, the latter procedure is preferred because it does not assume any particular distribution for the bootstrapped-generated means. While means do become Normally distributed for large sample sizes (because of the Central Limit Theorem), other statistics of interest do not, and so we should not assume that the statistic of interest will always be Normally distributed under bootstrap sampling.]

Question 3 (6 pts)

Briefly ( a paragraph, with drawings as needed) explain the relationship between testing a hypothesis and estimating an effect size with a confidence interval. Can you use an effect size and its confidence interval to test a hypothesis? If so, how? Why might providing an effect size and its confidence interval be preferable to providing just a p-value?

Bonus (5 pts)

Using the ‘fitdistr’ function available from the MASS package, find the distribution that reasonably well fits the treatment and control datasets. (In other words, find the best fitting parameters for that distribution.) I’m not expecting you to do an in-depth model comparison, as we haven’t learned how yet, but I want you to find something that fits reasonably well. You don’t need to use the same distribution for both Treatment and Control, though you can. (Note that we haven’t covered the ‘fitdistr’ function yet, but this is a bonus problem and so you are expected to sort it out from the documentation. Remember, once you have loaded the package you can use ?fitdistr for more information.

Using that distribution, answer Question 2e again using a parametric bootstrap approach.