

Assignment 3

Instructor: Xiner Zhou

January 25, 2023

Exercise 1

August, Hung, and Houck (1974) studied collagen metabolism in children deficient in growth hormone before and after growth hormone therapy. The data are the values of heat-insoluble hydroxyproline in the skin of children before and 3 months after growth hormone therapy. Can we conclude on the basis of these data that growth hormone therapy increases heat-insoluble hydroxyproline in the skin?

Child i	Before	After
1	349	425
2	400	533
3	520	362
4	490	628
5	574	463
6	427	427
7	435	449

Pretend that you are the Statistician/Data Scientist working on this study, and answer the following questions.

- (1) Tell your collaborators what test(s) are appropriate for this problem and why.
- (2) Perform all the test(s) proposed above, since the sample size is relatively small, conduct the exact test(s) (not the large-sample approximation), and make conclusions.
- (3) For each test proposed above, obtain a point estimate of the effect of growth hormone therapy on heat-insoluble hydroxyproline in the skin, and find a 95% confidence interval.
- (4) Summarize your findings in (2) and (3), and make sure to explain if you have tried multiple methods and see different results.

Exercise 2

In this exercise, we are going to do a **Simulation study** to compare several statistical tests. Simulation is a crucial part of almost every statistical research. The goal here is to let you see the reasons why we learn nonparametric methods, especially for nonparametric tests.

Notice that, one-sample t-test, Wilcoxon signed rank test and Fisher's signed test, are tests designed for one-sample location problem. Your goal is to explore the "distribution-free" property of nonparametric tests. That is, why do we say nonparametric tests can work better under a wide range of population distributions.

In addition to normal distribution, pick several (at least 2) continuous distributions. For each population distribution chosen, create a graph similar to what we have in lecture notes, and compare Type I Error and Power of One Sample Nonparametric Tests (Wilcoxon signed rank test, Fisher's signed test) with parametric t-test.

Explain what you observe in your simulation, in terms of relative performance, in type 1 error and power. Does your simulation confirm and help you understand the statement "nonparametric tests tend to work better for a wider range of population distributions, i.e. distribution-free"?