

Homework 1

- This dataset is drawn from a study discussed by Siegel (1956). It involves eight pairs of identical twins who are of nursery school age. In the study, for each pair, one is randomly selected to attend nursery school while the other remains at home. At the end of the study period, all 16 children are given the same social awareness test. The scores are given below. Conduct an appropriate test of the hypothesis that the median of the distribution of difference of scores between school and home trained twins is positive.

school	82	69	73	43	58	56	76	65
home	63	42	74	37	51	43	80	62

- Generate a sample of size 30 from a standard t_2 distribution using the R function `rt`. Use the t-test at $\alpha = 5\%$ to test the hypothesis that the median θ (same as mean) is zero. Use the Wilcoxon signed rank test for the same purpose.

Repeat the above procedure 1000 times and record the empirical size (proportion of times null hypothesis is rejected) of each test.

Now generate a sample of size 30 from a standard t_2 distribution using the R function `rt` and add 0.5 to each observation. Carry out the above procedure to test for $\theta = 0$. Here the actual theta is non-zero. So this will give the empirical power of each test.

Summarize your conclusions regarding comparative size and power of the t-test test and Wilcoxon test for the specific simulation settings of this problem.

- Data was recorded on 150 workplace accidents. The number of days since the previous accident is summarized in the following table

Days	0	1	2	3	4	5	6	7	8	9	10	11	12	13	15	20
Count	37	35	23	23	8	4	7	4	2	2	0	1	1	1	1	1

- First we assume a geometric model for this data.

$$P(X = x) = p(1 - p)^x, \quad x = 0, 1, 2, \dots$$

Find the MLE for of p and calculate its value from the given set of data.

- Calculate the MLE of the probability of the event $A = (X > 3)$ under this geometric model.
- Use a nonparametric estimation method to estimate $\theta = P(X > 3)$.
- Obtain a 95% confidence interval for θ based on the nonparametric estimator.

4. Suppose a health administrator wants to compare the unoccupied bed space for three hospitals located in the same city. She randomly selects 10 different days from the records of each hospital and lists the number of unoccupied beds for each day. The data is given below.
- Hospital 1: 6, 38, 3, 17, 11, 30, 15, 16, 25, 5
Hospital 2: 34, 28, 42, 13, 40, 31, 9, 32, 39, 27
Hospital 3: 13, 35, 19, 4, 29, 0, 7, 33, 18, 24.
- Carry out an appropriate nonparametric test of the hypothesis that the probability distributions of the number of unoccupied beds are the same for all three hospitals.