Sample Questions

- The questions in this file can be considered as samples for Assignment 1.
- The solution will be posted soon.
- Attempt these questions carefully for practice before checking the solution. Then compare your results with the solution to self-assess your performance.
- For Questions 1 – 4, choose T (True) or F (False) to answer each of parts (a) – (c) and explain why your choice is correct.
- Questions of this type will carry marks for each of parts (a) – (c) as follows:
  - 1 mark for a correct choice of T or F, and 0 mark for incorrect choice;
  - 3 marks for convincing reasons, 1 or 2 marks for partially correct reasons, and 0 mark for incorrect or irrelevant reasons;
  - 4 marks maximum for each part.
- For Questions 5 – 8, work out the details and show the steps to solve each problem, including the right theory and methods used, appropriate formulae to calculate the answers, and the steps of calculations.

True/False questions

Question 1

In a hypothesis test of the null hypothesis H_0 against the alternative H_1,

(a) if H_0 is rejected at the 5% level of significance, then H_0 is surely wrong;

(b) if H_0 is rejected at the 5% level of significance, then there is sufficient evidence to support the claim that H_1 is correct;

(c) if H_0 is accepted at the 10% level of significance, then there is sufficient evidence that H_0 is correct;

Question 2

If a statistical procedure is nonparametric, then

(a) it is valid whether the underlying probability distributions of the sample data are parametric or nonparametric;

(b) it does not use any distribution with known mathematical form;

(c) there is no parameter involved in the procedure.

Question 3

Let T^+ be the Wilcoxon signed rank statistic from a random sample $X_1,\ldots,X_8$ with median θ, and R_i the rank of X_i for T^+. The following statements are true:

(a) $\Pr(T^+ = 7) = 5/256$ under $H_0 : \theta = 0$.

(b) The distribution of T^+ is symmetric about 36.

(c) If $X_i < 0 < X_j$, then $R_i < R_j \iff X_i + X_j > 0$.

Question 4

Based on two independent samples, if the Wilcoxon rank sum test rejects $\Delta = 0$ and the Ansari-Bradley test finds little evidence against $\gamma^2 = 1$, then:

(a) we can reasonably conclude that the two samples have a significant difference in location, but not in dispersion.

(b) the difference in location is justified, but not equal dispersion.

(c) neither the Wilcoxon rank sum test nor the Ansari-Bradley test is reliable.

Problem-solving questions

Question 5

In an annual survey to determine whether federal pay scales were commensurate with private sector salaries, government and private workers were matched as closely as possible (with respect to type of job, educational background, years of experience, etc.) and the salaries of the matched pairs were obtained. The data in the table below are the annual salaries (in dollars) for 12 such matched pairs,

Pair *i*	Private	Government
1	12,500	11,750
2	22,300	20,900
3	14,500	14,800
4	32,300	29,900
5	20,800	21,500
6	19,200	18,400
7	15,800	14,500
8	17,500	17,900
9	23,300	21,400
10	42,100	43,200
11	16,800	15,200
12	14,500	14,200

Let X correspond to the government worker's salary, Y to the matched private sector salary, and θ the median of the difference $Z = Y - X$.

Based on the data provided in the above table, perform the following analyses:

(a) Find the exact *p*-value of testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$ by the sign test.

(b) Calculate the large-sample approximation of the above *p*-value. To get a better approximation, use the *continuity correction*: $\Pr(B \geq b) = \Pr(B > b - 0.5)$.

(c) Estimate the median θ and obtain its exact and approximate confidence intervals with at least 95% confidence level based on the sign statistic.

(d) Test the null hypothesis of no difference between government and private sector salaries against the alternative hypothesis that private sector salaries are higher than government by the Wilcoxon signed rank test.

(e) Estimate the median θ of $Y - X$ and obtain its approximate 95% confidence interval based on the Wilcoxon signed ranks.

(f) Compare and comment on the differences in the *p*-values of tests and confidence intervals of θ between the sign test statistic and the Wilcoxon signed ranks.

Question 6

The following two samples were obtained from a study of spinal cord injuries:

$(X_1,\ldots,X_8)=(0.89,0.76,0.63,0.69,0.58,0.79,0.02,0.79)$.

$(Y_1,\ldots,Y_5)=(0.19,0.14,0.02,0.44,0.37)$

where X represents certain health measure on patients with spinal cord injuries and Y on non-injured patients. The two samples are considered as independent.

Assume the location-shift model for the above data with location-shift Δ.

(a) Calculate the approximate p-value of testing $H_0:\Delta=0$ against $H_1:\Delta<0$ by the Wilcoxon rank sum test conditional on the ties in the data, and explain what the test result means.

(b) Determine the exact p-value of the problem in part (a) based on the conditional distribution with the ties given in the data.

(c) Estimate the location-shift parameter Δ and find an approximate 95% confidence interval of Δ based on the Wilcoxon rank sum.

Question 7

Two samples of data from a study are listed below:

$(X_1,\ldots,X_8)=(0.37,0.23,-0.06,0.18,0.44,0.31,-0.22,0.35)$

$(Y_1,\ldots,Y_4)=(0.34,0.14,-0.33,0.41)$

The question of interest is whether there exists a significant difference in variability (dispersion) between the two samples.

(a) Let C denote the Ansari-Bradley two-sample scale statistic. Use enumerations to find the probabilities $\Pr(C=c)$ for $c=6,7,8,9$, and determine the value of c_α such that $\Pr(C\ge c_\alpha)=\alpha$, where α is the largest achievable level at or below 5%. Reminder: C has a symmetric distribution if $N=m+n$ is an even number.

(b) Assume the location-scale parameter model with equal location parameter for the data. Test $H_0:\gamma^2=1$ $(\mathrm{Var}(X)=\mathrm{Var}(Y))$ against $H_1:\gamma^2\ne 1$ $(\mathrm{Var}(X)\ne\mathrm{Var}(Y))$ at the 10% level of significance based on the results in part (a).

(c) Test the hypotheses in part (b) by the large-sample approximate p-value and interpret the result in words.

Question 8

Use programme R to carry out the Wilcoxon signed rank test in Question 5(d), the Wilcoxon rank sum test in Question 6(a), and the Ansari-Bradley in Question 7(b). Show the R-commands and output.