STA3007 Assignment 1

118010350

October 17, 2020

Question 1

(a) T: If direction is true: X is symmetric about $a \in \mathbb{R}$, $F(x)$ is a continuous function on $\mathbb{R} \Rightarrow F(a-x') + F(a+x') = 1$ holds for all $x' \in \mathbb{R} \Rightarrow$ let $x = a+x'$, $F(2a-x) = 1-F(x)$ holds for all $x \in \mathbb{R}$. Only if direction is true, we can justify it by its contrapositive. Suppose $F(x)$ is a discrete function on $\mathbb{R}$. There exists a discrete point x in $F(x)$, such that $F(2a-x)+F(x) > 1$. Thus, $F(2a-x) = 1-F(x)$ does not hold for all $x \in \mathbb{R}$. By contrapositive, the only if direction is true.

(b) T: WLOG, assume $b_1 < b_2 < b_3$ in each combination, then random selection generate $No.\{(b_1, b_2, b_3)\} = \binom{10}{3} = 120$, and thus $Pr(X = x) = No.\{(b_1, b_2, b_3) : b_1b_2 + b_1b_3 + b_2b_3 - 3b_1b_2b_3 = x; b_1 < b_2 < b_3\}/120$.

(c) F: The chance to accept a correct hypothesis H_1 is $Pr(acceptH_1|H_1)$, but the p-value is the probability under H_0. The p-value of the test is 0.05 $\Rightarrow$ H_0 is rejected at 5% significance level $\Rightarrow$ Type I error is 0.05, i.e. the chance to reject a correct H_0 is 5% $\Rightarrow$ there is 95% chance that H_0 is false. Thus, the true statement should be there is 95% chance to reject a wrong hypothesis H_0.

Question 2

(a) F: The Y_i is defined as $Y_i = I_{\{x_i>0\}}$, $i = 1, \cdots, n$, then when $\theta = 0$, each Y_i has a known distribution $Bin(1, 0.5)$, when $\theta \neq 0$, each Y_i has a nonparametric distribution. Thus, the statement that each Y_i has a parametric distribution is false, and the true statement should be each Y_i dose not have a parametric distribution.

(b) F: Under the hypothesis $H_0 : \theta = 0$, $\psi_i \sim 1 - \psi_i$, add with symmetric distribution of $X_1, \ldots, X_n$, $T^+ = \sum_{i=1}^n \psi_i R_i \sim M - T^+ = \sum_{i=1}^n (1-\psi_i)R_i$, then $Pr(T^+ \geqslant t) = Pr(T^+ \leqslant M - t)$. Hence both H_0 and assumption of symmetric distribution ensure the symmetry of T^+. The symmetry of T^+ gives the property $Pr(T^+ \geqslant t) = Pr(T^+ \leqslant M - t)$, where $M = n(n+1)/2$. Thus, the rejection rule, $T^+ \geqslant t_\alpha$ against $H_1 : \theta > 0$, $T^+ \leqslant M - t_\alpha$ against $H_1 : \theta < 0$, either $T^+ \geqslant t_{\alpha/2}$ or $T^+ \leqslant M - t_{\alpha/2}$) against $H_1 : \theta \neq 0$ are effected by the assumption of symmetric distribution for $X_1, \ldots, X_n$.

(c) F: The $100(1-\alpha)\%$ confidence interval of θ is given by $(\theta_L, \theta_U) = (X_{(C_\alpha)}, X_{(n+1-C_\alpha)}) = (X_{(n+1-b_{\alpha/2})}, X_{(b_{\alpha/2})})$, which is not related to the point estimate of θ. Thus, the statement is false, and the true statement should be it is not necessary to first find a point estimate of θ to construct a non-parametric confidence interval of θ.

Question 3

(a) T: The distribution of T^+ under $H_0 : \theta = 0$ is given by $Pr(T^+ = t) = No.\{(r_1, \cdots, r_B) : r_1 + \cdots + r_B = t\}/2^n$, and when $t = 9$, the combination can be $(9), (1,8), (2,7), (3,6), (4,5), (1,2,6), (1,3,5), (2,3,4)$. Thus, $Pr(T^+ = 9) = 8/2^n = 2^{3-n}$.

(b) F: The range of T^+ is $\{0, 1, \cdots, M\}$ with $M = n(n+1)/2$, for $n > 10$, i.e. $n \geqslant 11$, we have $M \geqslant 66$. Since under $H_0 : \theta = 0$, the distribution of T^+ is symmetric, then $Pr(T^+ \geqslant (M+1)/2) \leqslant 0.5$. Since $(M+1)/2 > 30$, then we have $Pr(T^+ \geqslant 30) > 0.5$. Thus, the statement is false, and the true statement should be under $H_0 : \theta = 0$, $Pr(T^+ \geqslant 30) > 0.5$.

(c) T: If $X_{(5)} < 0$, we have at least 5 X_i such that $X_i < 0$. Since the Walsh averages $W_{ij} = (X_i + X_j)/2,$, $i \leqslant j = 1, \cdots, n$, then the Walsh averages $W_{ij} < 0$ for at least 15 pairs $\{(i,j) : 1 \leqslant j \leqslant n\}$ because $\binom{5}{2} + 5 = 15$.

Question 4

(a) T: The sample $(Y_1, \cdots, Y_n)$ has mostly smaller values than $(X_1, \cdots, X_m)$ $\Rightarrow (Y_1, \cdots, Y_n)$ is likely to center at a smaller value than $(X_1, \cdots, X_m) \Rightarrow$ $(Y_1, \cdots, Y_n)$ is likely to have a smaller median than $(X_1, \cdots, X_m)$. The sample $(Y_1, \cdots, Y_n)$ has a substantially wider range than $(X_1, \cdots, X_m) \Rightarrow$ $(Y_1, \cdots, Y_n)$ is likely distributed more widely than $(X_1, \cdots, X_m) \Rightarrow (Y_1, \cdots,$ $Y_n)$ is likely to have a greater variance than $(X_1, \cdots, X_m)$.

(b) F: The Wilcoxon rank sum test is based on the assumption that two samples have the same variance. Since $(Y_1, \cdots, Y_n)$ is likely to have a greater variance than $(X_1, \cdots, X_m)$, though $(Y_1, \cdots, Y_n)$ may have a smaller median than $(X_1, \cdots, X_m)$, the test may not likely to reject $H_0 : \theta_X = \theta_Y$ in favor of $H_1 : \theta_X > \theta_Y$.

(c) F: The Ansari-Bradley rank test is based on the assumption that two samples have the same median. Since $(Y_1, \cdots, Y_n)$ is likely to have a smaller median than $(X_1, \cdots, X_m)$, though $(Y_1, \cdots, Y_n)$ may have a greater variance than $(X_1, \cdots, X_m)$, the test may not likely to reject $H_0 : Var(x) = Var(Y)$ in favor of $H_1 : Var(X) < Var(Y)$.

Question 5

(a) The values of $Z_i = Y_i - X_i$, $i = 1, \cdots, 11$ are calculated as

$$89, 87, -60, 68, 56, 114, -44, 100, 91, 27, -45$$

The data have 8 positive values in the sample of size $n = 11$ and $B \sim Bin(11, 0.5)$ under $H_0 : \theta = 0$. Thus the exact p-value of testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$ by the sign test is

$$Pr(B \geqslant 8) = Pr(B \leqslant 3) = \left[\binom{11}{0} + \binom{11}{1} + \binom{11}{2} + \binom{11}{3}\right](0.5)^{11} = 0.1133$$

(b) The ordered values $Z_{(1)} \leqslant Z_{(2)} \leqslant \cdots \leqslant Z_{(11)}$ of $Z_1, \cdots, Z_{11}$ are

$$-60, -45, -44, 27, 56, 68, 87, 89, 91, 100, 114$$

For $B \sim Bin(11, 0.5)$, $Pr(B \geqslant 9) = 0.0327 < 0.05$ and $Pr(B \geqslant 8) = 0.1133 > 0.05$. Thus the minimum achievable confidence level above 90% is

$$1 - \alpha = 1 - 2(0.0327) = 1 - 0.0654 = 0.9346, \alpha = 0.0654$$

It follows that

$$b_{\alpha/2} = b_{0.0327} = 9, C_\alpha = n + 1 - b_{\alpha/2} = 11 + 1 - 9 = 2$$

The 93.46% confidence interval of θ is given by $(Z_2, Z_9) = (-45, 91)$.

(c) We test $H_0 : \theta = 0$ against $H_1 : \theta > 0$ by the Wilcoxon signed rank test. Calculate the values of $Z_i = Y_i - X_i$, $|Z_i|$, R_i of $|Z_i|$, $\psi_i = I_{\{Z_i > 0\}}$ and $\psi_i R_i$, $i = 1, \cdots, 11$ in the following table

i	Z_i	$\|Z_i\|$	R_i	ψ_i	$\psi_i R_i$
1	89	89	8	1	8
2	87	87	7	1	7
3	-60	60	5	0	0
4	68	68	6	1	6
5	56	56	4	1	4
6	114	114	11	1	11
7	-44	44	2	0	0
8	100	100	10	1	10
9	91	91	9	1	9
10	27	27	1	1	1
11	-45	45	3	0	0

The Wilcoxon signed rank statistic is calculated as

$$T^+ = \sum_{i=1}^{11} \psi_i R_i = 8 + 7 + 6 + 4 + 11 + 10 + 9 + 1 = 56$$

Under $H_0 : \theta = 0$, t^+ follows a symmetric distribution, then $Pr(T^+ \geqslant t) = Pr(T^+ \leqslant M - t)$, where $M = n(n+1)/2 = 11(11+1)/2 = 66$. Hence $Pr(T^+ \geqslant 56) = Pr(T^+ \leqslant 10)$, and we can calculate $Pr(T^+ \leqslant 10)$ by enumeration. All the possible combinations $(r_1, \cdots, r_B)$ for $T^+ = 0, 1, \cdots, 10$ are in the following table

T^+	$(r_1, \cdots, r_B)$	$No.$
0	$\varnothing$	1
1	(1)	1
2	(2)	1
3	(3) (1,2)	2
4	(4) (1,3)	2
5	(5) (1,4) (2,3)	3
6	(6) (1,5) (2,4) (1,2,3)	4
7	(7) (1,6) (2,5) (3,4) (1,2,4)	5
8	(8) (1,7) (2,6) (3,5) (1,2,5) (1,3,4)	6
9	(9) (1,8) (2,7) (3,6) (4,5) (1,2,6) (1,3,5) (2,3,4)	8
10	(10) (1,9) (2,8) (3,7) (4,6) (1,2,7) (1,3,6) (1,4,5) (2,3,5) (1,2,3,4)	10

The exact p-value of testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$ by Wilcoxon signed rank test is

$$Pr(T^+ \leqslant 10) = \frac{3(1) + 2(2) + 3 + 4 + 5 + 6 + 8 + 10}{2^{11}} = 0.0210 < 0.05$$

Since $Pr(T^+ \geqslant 56) = Pr(T^+ \leqslant 10) < 0.05$, then we reject the null hypothesis $H_0 : \theta = 0$ at 5% significance level, that is, there is sufficient evidence at 5% level that the new technology is effective to increase the production of the company.

(d) Let $W_1 \leqslant W_2 \leqslant \cdots \leqslant W_M$ be the ordered values of Walsh averages $\{(Z_i + Z_j)/2, i \leqslant j \leqslant n\}$, where $M = n(n+1)/2 = 11(11+1)/2 = 66$. The values ordered Walsh averages are listed in the following table

k	$W_{(k)}$	k	$W_{(k)}$	k	$W_{(k)}$	k	$W_{(k)}$	k	$W_{(k)}$	k	$W_{(k)}$
1	-60	12	5.5	23	22.5	34	56	45	77.5	56	91
2	-52.5	13	6	24	23	35	57	46	78	57	91
3	-52	14	11.5	25	23.5	36	58	47	78.5	58	93.5
4	-45	15	12	26	27	37	59	48	79.5	59	94.5
5	-44.5	16	13.5	27	27	38	62	49	84	60	95.5
6	-44	17	14.5	28	27.5	39	63.5	50	85	61	100
7	-16.5	18	15.5	29	28	40	68	51	87	62	100.5
8	-9	19	20	30	34.5	41	70.5	52	88	63	101.5
9	-8.5	20	21	31	35	42	71.5	53	89	64	102.5
10	-2	21	21.5	32	41.5	43	72.5	54	89	65	107
11	4	22	22	33	47.5	44	73.5	55	90	66	114

The estimate of median θ based on Wilcoxon signed ranks is

$$\tilde{\theta} = \frac{W_{(M/2)} + W_{(M/2+1)}}{2} = \frac{W_{(33)} + W_{(34)}}{2} = 51.75$$

Based on the formula $Pr(\theta < W_{(C_\alpha)}) = Pr_0(T^+ \geqslant t_{\alpha/2}) = \alpha/2$, where $C_\alpha = M - t_{\alpha/2} + 1$, and $Pr_0(T^+ \geqslant 56) = 0.0210 < 0.025$, $Pr(T^+ \geqslant 55) = Pr(T^+ \leqslant 11) = 0.0269 > 0.025$. Thus the minimum achievable confidence interval above 95% is

$$1 - \alpha = 1 - 2(0.021) = 1 - 0.042 = 0.958, \alpha = 0.042$$

It follows that

$$t_{\alpha/2} = 56, C_\alpha = M - t_{\alpha/2} + 1 = 66 - 56 + 1 = 11$$

The 95.8% confidence interval of θ is given by $(W_{11}, W_{56}) = (4, 91)$.

(e) The p-value of sign test is over 10%, which shows insufficient evidence for $\theta > 0$ at 10% level. The p-value of Wilcoxon signed rank test is below 5%, which shows sufficient evidence for $\theta > 0$ at 5% level. This shows that the Wilcoxon signed rank test is more powerful and efficient than the sign test to difference between paired samples based on the same set of data.

The 93.46% confidence interval of θ given by sign statistic is $(-45, 91)$, and the 95.8% confidence interval given by Wilcoxon signed ranks is $(4, 91)$. Though the confidence level is close, the confidence interval based on the Wilcoxon signed ranks is much shorter, indicating more accurate estimation, than the confidence interval based on the sign statistic.

Question 6

(a) Since X_1 and X_2 are independent, and

$$f_1(x) = 0.5I_{\{|x|\leqslant 1\}} = \begin{cases} 0.5 & -1 \leqslant x \leqslant 1 \\ 0 & otherwise \end{cases}$$

and

$$f_2(x) = e^{-2|x|} = \begin{cases} e^{-2x} & x \geqslant 0 \\ e^{2x} & x < 0 \end{cases}$$

Then, the joint distribution of X_1, X_2 is

$$f(x_1, x_2) = \begin{cases} 0.5e^{-2|x_2|} & |x_1| \leqslant 1 \\ 0 & |x_1| > 1 \end{cases}$$

Hence we can calculate

$$\begin{aligned} Pr(S=2) &= Pr\left(I\{X_1 > 0\} = 0, I\{X_2 > 0\} = 1\right) \\ &= Pr(X_1 \leqslant 0, X_2 > 0) \\ &= Pr(X_1 \leqslant 0)Pr(X_2 > 0) \\ &= 0.5(0.5) = 0.25 \end{aligned}$$

$$\begin{aligned} Pr(X_1 > 0, R_1 = 2, X_2 < 0) &= Pr(X_1 + X_2 > 0, X_1 > 0, X_2 < 0) \\ &= \iint_{x_1+x_2>0, x_1>0, x_2<0} f(x_1, x_2) dx_1 dx_2 \\ &= \int_{-1}^{0} \int_{-x_2}^{1} 0.5e^{-2|x_2|} dx_1 dx_2 \\ &= \int_{-1}^{0} 0.5e^{2x_2} + 0.5x_2 e^{2x_2} dx_2 \\ &= 0.125 + 0.125e^{-2} = 0.1419 \end{aligned}$$

$$\begin{aligned} Pr(X_1 < 0, R_2 = 2, X_2 > 0) &= Pr(X_1 + X_2 > 0, X_1 < 0, X_2 > 0) \\ &= \iint_{x_1+x_2>0, x_1<0, x_2>0} f(x_1, x_2) dx_1 dx_2 \\ &= \int_{0}^{1} \int_{-x_2}^{0} 0.5e^{-2|x_2|} dx_1 dx_2 + \int_{1}^{\infty} \int_{-1}^{0} 0.5e^{-2|x_2|} dx_1 dx_2 \\ &= 0.125 - 0.125e^{-2} = 0.1081 \end{aligned}$$

$$
\begin{aligned}
Pr(T^{+} = 2) = Pr(X_1 > 0, R_1 = 2, X_2 < 0)& \\
+ Pr(X_1 < 0, R_2 = 2, X_2 > 0) &= 2(0.125) = 0.25
\end{aligned}
$$

(b) Similarly, we can have

$$
f_1(x) = f_2(x) = \begin{cases} 1 & -0.5 \leqslant x < 0 \\ 2(1-x)^3 & 0 \leqslant x \leqslant 1 \\ 0 & otherwise \end{cases}
$$

Then, the joint distribution of X_1, X_2 is

$$
f(x_1, x_2) = \begin{cases} 2(1-x_2)^3 & -0.5 \leqslant x_1 < 0, 0 \leqslant x_2 \leqslant 1 \\ 2(1-x_1)^3 & -0.5 \leqslant x_2 < 0, 0 \leqslant x_1 \leqslant 1 \\ \cdots & othercases \end{cases}
$$

Hence we can calculate

$$
\begin{aligned}
Pr(S = 2) &= Pr\left(I\{X_1 > 0\} = 0, I\{X_2 > 0\} = 1\right) \\
&= Pr(X_1 \leqslant 0, X_2 > 0) \\
&= Pr(X_1 \leqslant 0) Pr(X_2 > 0) \\
&= 0.5(0.5) = 0.25
\end{aligned}
$$

$$
\begin{aligned}
Pr(X_1 > 0, R_1 = 2, X_2 < 0) &= Pr(X_1 + X_2 > 0, X_1 > 0, X_2 < 0) \\
&= \iint_{x_1+x_2>0, x_1>0, x_2<0} f(x_1, x_2) dx_1 dx_2 \\
&= \int_{-0.5}^{0} \int_{-x_2}^{1} 2(1-x_1)^3 dx_1 dx_2 \\
&= \int_{-0.5}^{0} 0.5(1+x_2)^4 dx_2 \\
&= 0.1 - 0.1(0.5)^5 = 0.0969
\end{aligned}
$$

$$
\begin{aligned}
Pr(X_1 < 0, R_2 = 2, X_2 > 0) &= Pr(X_1 + X_2 > 0, X_1 < 0, X_2 > 0) \\
&= \iint_{x_1+x_2>0, x_1<0, x_2>0} f(x_1, x_2) dx_1 dx_2 \\
&= \int_{0}^{1} \int_{-x_2}^{0} 2(1-x_2)^3 dx_1 dx_2 \\
&= \int_{0}^{1} 2x_2(1-x_2)^3 dx_1 dx_2 \\
&= 0.1
\end{aligned}
$$

$$Pr(T^+ = 2) = Pr(X_1 > 0, R_1 = 2, X_2 < 0)$$
$$+ Pr(X_1 < 0, R_2 = 2, X_2 > 0) = 0.2 - 0.1(0.5)^5 = 0.1969$$

(c) In part (a), $Pr(S = 2) = Pr(T^+ = 2) = 0.25$. In part (b), $Pr(S = 2) = 0.25$ but $Pr(T^+ = 2) < 0.25$. Hence, under the condition of symmetric distribution, we have $Pr(S = \psi_1 + 2\psi_2) = Pr(T^+ = R_1\psi_1 + R_2\psi_2)$, but without the condition of symmetric distribution, we cannot have $Pr(S = \psi_1+2\psi_2) = Pr(T^+ = R_1\psi_1+R_2\psi_2)$. Therefore, in the Wilcoxon signed rank test, the assumption that the distributions of $X_1, \cdots, X_n$ are symmetric ensures that

$$Pr(T^+ = \sum_{i=1}^{n} R_i\psi i) = Pr(S = \sum_{i=1}^{n} i\psi_i)$$

Question 7

(a) The ordered values of $(X_1, \cdots, X_6, Y_1, \cdots, Y_4)$ are

$$(Z_1, \cdots, Z_{10}) = (-3, -1, -1, 1, 1, 3, 6, 8, 12, 12)$$

And the ranks of $(Z_1, \cdots, Z_{10})$ are

$$(r_1, \cdots, r_{10}) = (1, 2.5, 2.5, 4.5, 4.5, 6, 7, 8, 9.5, 9.5)$$

Then the two-sample Wilcoxon rank sun statistic is

$$W = 2.5 + 4.5 + 7 + 9.5 = 23.5$$

All 4-tuples (r_i, r_j, r_k, r_l) such that $r_i + r_j + r_k + r_l = 26$ and $i < j < k < l$ are listed in the following table

(r_i, r_j, r_k, r_l)	(i, j, k, l)	$No.$
(2.5,6,7,8)	(2,6,7,8) (3,6,7,8)	2
(1,6,7,9.5)	(1,6,7,9) (1,6,7,10)	2
(2.5,4.5,7,9.5)	(2,4,7,9) (2,4,7,10) (2,5,7,9) (2,5,7,10) (3,4,7,9) (3,4,7,10) (3,5,7,9) (3,5,7,10)	8

The total number of (r_i, r_j, r_k, r_l) with $i < j < k < l$ is

$$\binom{10}{4} = \frac{10(9)(8)(7)}{1(2)(3)(4)} = 210$$

Under the null hypothesis of no treatment effect, we can determine

$$Pr(W = 23.5) = \frac{2 + 2 + 8}{210} = \frac{12}{210} = 0.0571$$

(b) The ordered values of $(X_1, \cdots, X_6, Y_1, \cdots, Y_4)$ are

$$(Z_1, \cdots, Z_{10}) = (-3, -1, -1, 1, 1, 3, 6, 8, 12, 12)$$

And the scores of $(Z_1, \cdots, Z_{10})$ are

$$(a_1, \cdots, a_{10}) = (1, 2.5, 2.5, 4.5, 4.5, 5, 4, 3, 1.5, 1.5)$$

Then the Ansari-Bradley test statistic is

$$C = 2.5 + 4 + 4.5 + 1.5 = 12.5$$

All 4-tuples (a_i, a_j, a_k, a_l) such that $a_i + a_j + a_k + a_l = 12.5$ and $i < j < k < l$ are listed in the following table

(a_i, a_j, a_k, a_l)	(i, j, k, l)	$No.$
(1,2.5,5,4)	(1,2,6,7) (1,3,6,7)	2
(1,4.5,4,3)	(1,4,7,8) (1,5,7,8)	2
(2.5,4.5,4,1.5)	(2,4,7,9) (2,4,7,10) (2,5,7,9) (2,5,7,10) (3,4,7,9) (3,4,7,10) (3,5,7,9) (3,5,7,10)	8
(2.5,2.5,4.5,3)	(2,3,4,8) (2,3,5,8)	2
(1,2.5,4.5,4.5)	(1,2,4,5) (1,3,4,5)	2
(4.5,5,1.5,1.5)	(4,6,9,10) (5,6,9,10)	2

The total number of (a_i, a_j, a_k, a_l) with $i < j < k < l$ is

$$\binom{10}{4} = \frac{10(9)(8)(7)}{1(2)(3)(4)} = 210$$

Under the null hypothesis of equal dispersion between the two samples, we can determine

$$Pr(C = 12.5) = \frac{2 + 2 + 2 + 2 + 2 + 8}{210} = \frac{18}{210} = 0.0857$$

Question 8

(a) > x<-c(6.17, 4.78, 3.99, 5.65, 3.87, 4.43, 4.82, 6.68, 4.46, 6.95, 3.02, 4.22, 4.21, 3.97)
> y<-c(9.94, 7.08, 7.14, 5.82, 9.60, 10.09, 8.66, 4.74, 4.14, 10.92, 5.61, 6.47, 5.20, 8.21, 3.55, 9.81)
> wilcox.test(y, x, alternative ="greater")

```
wilcoxon rank sum test
data: y and x
W = 182, p-value = 0.001423
alternative hypothesis: true location shift is greater than 0
```

Since p-value $= 0.001423 < 0.01$, then we reject $H_0 : \Delta = 0$ in favor of $H_1 : \Delta > 0$ at 1% level of significance. Hence there is sufficient evidence (at 1% level of significance) for sample Y have a greater location parameter than sample X.

(b)
```
> x<-c(6.17, 4.78, 3.99, 5.65, 3.87, 4.43, 4.82, 6.68, 4.46, 6.95, 3.02, 4.22,
    4.21, 3.97)
> y<-c(9.94, 7.08, 7.14, 5.82, 9.60, 10.09, 8.66, 4.74, 4.14, 10.92, 5.61, 6.47,
    5.20, 8.21, 3.55, 9.81)
> ansari.test(y, x)
```

```
Ansari-Bradley test
data: y and x
AB = 119, p-value = 0.4846
alternative hypothesis: alternative hypothesis: true ratio of scales is not
equal to 1
```

Since p-value $= 0.4846 > 0.1$, then we accept $H_0 : \gamma^2 = 1$ against $H_1 : \gamma^2 \neq 1$ at 10% level of significance. Hence there is insufficient evidence (at 10% level of significance) for different dispersions between the two samples X and Y.

(c)
```
> x<-c(6.17, 4.78, 3.99, 5.65, 3.87, 4.43, 4.82, 6.68, 4.46, 6.95, 3.02, 4.22,
    4.21, 3.97)
> y<-c(9.94, 7.08, 7.14, 5.82, 9.60, 10.09, 8.66, 4.74, 4.14, 10.92, 5.61, 6.47,
    5.20, 8.21, 3.55, 9.81)
> x<x+2
> ansari.test(y, x)
```

```
Ansari-Bradley test
data: y and x
AB = 96, p-value = 0.007276
alternative hypothesis: alternative hypothesis: true ratio of scales is not
equal to 1
```

Since p-value $= 0.007276 < 0.01$, then we reject $H_0 : \gamma^2 = 1$ in favor of

$H_1 : \gamma^2 \neq 1$ at 1% level of significance. Hence there is sufficient evidence (at 1% level of significance) for different dispersions between the two samples X and Y.

Compared (b) and (c), we can find that when the two samples are not at the same location, though there are different dispersions between the two samples, the Ansari-Bradley rank test still accepts the false hypothesis. Therefore, the Ansari-Bradley rank test is reliable based on the fact that two samples are at the same location, i.e.

$$\frac{X - \theta}{\eta_1} \sim \frac{Y - \theta}{\eta_2}$$

where θ is the location parameter and η_1,η_2 are the scale parameters.