STA 3007 . Exercise 2. 118010350.

Question 1.

(a) True. X has a symmetric distribution about θ.

$Pr(a \le X \le b) = 1$, $Pr(X=a) > 0$, $Pr(X=b) > 0$.

$\Rightarrow$ a, b are symmetric about θ.

$\Rightarrow \theta = \frac{a+b}{2}$.

For $x \ge \theta$, and $Pr(X=x) > 0$. define $y = x - \theta \ge 0$.

$$\Rightarrow E[X] = \sum x Pr(X=x) = \sum(\theta+y) \cdot Pr(X=\theta+y) + \sum(\theta-y) \cdot Pr(X=\theta-y).$$

$$= \sum 2\theta \cdot Pr(X=\theta+y) = \sum 2\theta Pr(X=\theta-y).$$

Since $\sum Pr(X=\theta+y) = \sum Pr(X=\theta-y) = \frac{1}{2}$

then $E[X] = 2\theta \cdot \frac{1}{2} = \theta$.

$\Rightarrow E[X] = \theta = \frac{a+b}{2}$.

(b). True. X has a symmetric distribution about θ.

$\Rightarrow Pr(X=\theta-y) = Pr(X=\theta+y)$. $\forall y \in \mathbb{R}$.

$\Rightarrow Pr(X \le \theta-y) = Pr(X \ge \theta+y)$. $\forall y \in \mathbb{R}$

let $x = \theta - y$. $\Rightarrow Pr(X \le x) = Pr(X \ge 2\theta - x)$, $\forall x \in \mathbb{R} = (-\infty, \infty)$.

(c). False. $Pr(a \le X \le b) = 1$. $Pr(X=b) > 0$.

$\Rightarrow F(b) = Pr(X \le b) = 1$.

$Pr(X=a) > 0 \Rightarrow F(a) = Pr(X \le a) > 0$.

we have $a+b = 2\theta$, but $F(a) + F(b) > 1$.

$\Rightarrow F(\theta+x) + F(\theta-x) = 1$ does not hold for $\forall x \in \mathbb{R}$

Actually. $F(\theta+x) + F(\theta-x) = 1$ does not hold

for every discrete point x except $x = \theta$.

Question 2.

(a). True. All possible combination of $(r_1, \cdots r_8)$..

such that $T^+ = 7$. are (7), (1,6), (2,5), (3,4), (1,2,4).

Under $H_0: \theta = 0$. $Pr(T^+ = 7) = \frac{5}{2^8} = \frac{5}{256}$.

(b) False. The range of T^+ is $\{0, 1, \cdots, M\}$.

$M = \frac{8 \times 9}{2} = 36$, $\Rightarrow T^+$ is not symmetric about 36.

Actually, under $H_0: \theta = 0$, T^+ is symmetric

about $\frac{M}{2} = \frac{36}{2} = 18$.

(c). True. ① Show $R_i > R_j \Rightarrow X_i + X_j < 0$

If $X_i < 0 < X_j$, and $R_i > R_j$.

$\Rightarrow |X_i| > |X_j| \Rightarrow -X_i > X_j \Rightarrow X_i + X_j < 0$.

② Show $X_i + X_j < 0 \Rightarrow R_i > R_j$.

If $X_i < 0 < X_j$, and $X_i + X_j < 0$.

$\Rightarrow -X_i > X_j \Rightarrow |X_i| > |X_j| \Rightarrow R_i > R_j$.

Question 3.

(a). False. Since wilcoxon rank sum test rejects $\Delta = 0$. then $\Delta \neq 0$, But Ansari-Bradley test is based on $\Delta = 0$, thus it is not justified.

(b) False. If equal dispersion is not justified. then $r^2 \neq 1$. But wilcoxon rank sum test is based on $r^2 = 1$, thus it is not justified and its result of different location is also not justified.

The results of both tests are questionable and not well justified is true.

The wilcoxon rank sum test is not reliable because $r^2 = 1$ is not justified; the Ansari-Bradley test is not reliable because $\Delta = 0$ is not justified.

Question 4.

(a). Proof. Since $X_1 \sim U(-1,1)$, $X_2 \sim U(-2,2)$.

Then $Pr(X_1 = x) = \frac{1}{2}$, $-1 \leq x \leq 1$, $Pr(X_2 = x) = \frac{1}{4}$, $-2 \leq x \leq 2$, $\Rightarrow Pr(X_1 = x_1, X_2 = x_2) = \frac{1}{8}$. $-1 \leq x_1 \leq 1$, $-2 \leq x_2 \leq 2$.

① $Pr(T^+ = 0) = Pr(\psi_1 = 0, \psi_2 = 0) = Pr(X_1 \leq 0, X_2 \leq 0)$ X_1, X_2 independent.

$= Pr(X_1 \leq 0) \cdot Pr(X_2 \leq 0)$.

$= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

$Pr(S = 0) = Pr(\psi_1 = 0, \psi_2 = 0) = \frac{1}{4}$.

$\Rightarrow Pr(T^+ = 0) = Pr(S = 0)$.

② $Pr(T^+ = 1) = Pr(\psi_1 = 1, \psi_2 = 0, R_1 = 1) + Pr(\psi_1 = 0, \psi_2 = 1, R_2 = 1)$.

$= Pr(X_1 > 0, X_2 \leq 0, X_1 + X_2 < 0) + Pr(X_1 \leq 0, X_2 > 0, X_1 + X_2 < 0)$.

$Pr(X_1>0, X_2\le 0, X_1+X_2<0) = \int_{-1}^{0}\int_{0}^{-X_2} \frac{1}{8}\, dx_1 dx_2 + \int_{-2}^{-1}\int_{0}^{1} \frac{1}{8}\, dx_1 dx_2.$

$= \frac{1}{16}+\frac{1}{8} = \frac{3}{16}.$

$Pr(X_1\le 0, X_2>0, X_1+X_2<0) = \int_{-1}^{0}\int_{0}^{-X_1} \frac{1}{8}\, dx_2 dx_1$

$= \frac{1}{16}$

$Pr(T^+=1) = \frac{3}{16}+\frac{1}{16} = \frac{1}{4}$

$Pr(S=1) = Pr(\psi_1=1, \psi_2=0) = Pr(X_1>0, X_2\le 0) = \frac{1}{4}$

$\Rightarrow Pr(T^+=1) = Pr(S=1)$

③ $Pr(T^+=2) = Pr(\psi_1=1, \psi_2=0, R_1=2) + Pr(\psi_1=0, \psi_2=1, R_2=2).$

$= Pr(X_1>0, X_2\le 0, X_1+X_2>0) + Pr(X_1\le 0, X_2>0, X_1+X_2>0).$

$Pr(X_1>0, X_2\le 0, X_1+X_2>0) = \int_{0}^{1}\int_{-X_1}^{0} \frac{1}{8}\, dx_2 dx_1$

$= \frac{1}{16}$

$Pr(X_1\le 0, X_2>0, X_1+X_2>0) = \int_{0}^{1}\int_{-X_2}^{0} \frac{1}{8}\, dx_1 dx_2 + \int_{1}^{2}\int_{-1}^{0} \frac{1}{8}\, dx_1 dx_2$

$= \frac{1}{16}+\frac{1}{8} = \frac{3}{16}.$

$Pr(T^+=2) = \frac{1}{16}+\frac{3}{16} = \frac{1}{4}$

$Pr(S=2) = Pr(\psi_1=0, \psi_2=1) = Pr(X_1\le 0, X_2>0) = \frac{1}{4}.$

$\Rightarrow Pr(T^+=2) = Pr(S=2).$

④ $Pr(T^+=3) = Pr(\psi_1=1, \psi_2=1) = Pr(X_1>0, X_2>0)$

$= Pr(X_1>0)\, Pr(X_2>0)$

$= \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$

$Pr(S=3) = Pr(\psi_1=1, \psi_2=1) = \frac{1}{4}.$

$\Rightarrow Pr(T^+=3) = Pr(S=3).$

By ①, ②, ③, $\Rightarrow Pr(T^+=i) = Pr(S=i)$, $i\in\{0,1,2,3\}$.

(b). Suppose the median of X_1, X_2 is θ.

Then $Pr(X_1\le\theta) = Pr(X_2\le\theta) = \int_0^\theta f(x)\,dx = 0.5.$

$\Rightarrow \int_{-\ln 2}^{\theta} 0.5\cdot e^{-x} = 0.5 \Rightarrow \int_{-\ln 2}^{\theta} e^{-x} = 1.$

$\Rightarrow \theta=0$. Check that $Pr(X_1>\theta) = Pr(X_2>\theta) = \int_\theta^\infty f(x)\,dx = 0.5.$

Thus $\theta=0$ is the median of X_1, X_2.

$Pr(X_1=x_1, X_2=x_2) = 0.25\cdot e^{-(x_1+x_2)}$ $x_1\ge -\ln 2$, $x_2\ge -\ln 2$.

By (a), $Pr(T^+=1) = Pr(X_1>0, X_2\le 0, X_1+X_2<0)$

$+ Pr(X_1\le 0, X_2>0, X_1+X_2<0).$

$Pr(X_1>0, X_2\le 0, X_1+X_2<0) = \int_{-\ln 2}^{0}\int_{0}^{-X_2} 0.25\cdot e^{-(X_1+X_2)}\, dx_1 dx_2.$

$= \frac{1}{4} - \frac{1}{4}\ln 2.$

$Pr(X_1 \le 0, X_2 > 0, X_1 + X_2 < 0) = \int_{-\ln 2}^{0} \int_{0}^{-X_1} 0.25 \cdot e^{-(X_1+X_2)} \, dx_2 dx_1$

$= \frac{1}{4} - \frac{1}{4}\ln 2.$

$Pr(T^+ = 1) = \left(\frac{1}{4} - \frac{1}{4}\ln 2\right) \times 2 = \frac{1}{2} - \frac{1}{2}\ln 2.$

$Pr(S=1) = Pr(\varphi_1 = 1, \varphi_2 = 0) = Pr(X_1 > 0, X_2 \le 0)$

$= Pr(X_1 > 0) \, Pr(X_2 \le 0)$

$= \frac{1}{4}$

$Pr(T^+ = 1) \ne Pr(S=1).$

Question 5.

(a). As $m = n = 21$, $N = m + n = 42$. N is even.

Let $z_{(1)} \le z_{(2)} \le \cdots \le z_{(42)}$ denote the ordered values of the combined X and Y values, and r_i denote the rank of $z_{(i)}$, $i = 1, 2, \cdots 42$. Then

$z_{(i)}$	0	1	5	7	8	9	11	12	13	13	13	14	14	15.
r_i	1	2	3	4	5	6	7	8	10	10	10	12.5	12.5	14
$z_{(i)}$	16	16	19	20	23	23	25	26	30	33	34	36	37	39
r_i	15.5	15.5	17	18	19.5	19.5	21	21	20	19	18	17	16	15
$z_{(i)}$	44	45	47	62	62	94	139	145	146	146	154	156	169	290.
r_i	14	13	12	10.5	10.5	9	8	7	5.5	5.5	4	3	2	1

C = sum of Y scores $= 2 + 4 + 10 + 10 + 15.5 + 15.5 + 18 + 19.5$

$+ 21 + 20 + 17 + 16 + 15 + 13 + 10.5 + 9 + 8 + 7 + 5.5 + 2 + 1$

$= 239.5.$

Note: r_i in circle are tied scores of sample Y.

other r_i are tied scores of sample X.

(b). $E_0(C) = \frac{n(N+2)}{4} = \frac{21 \times 44}{4} = 231$

$Var_0(C) = \frac{mn}{N(N-1)} \cdot \left[\sum_{j=1}^{N} r_j^2 - \frac{N(N+2)^2}{16} \right]$

$= \frac{21 \times 21}{42 \times (42-1)} \left[(2 \times 1^2 + 2 \times 2^2 + \cdots + 2 \times 21^2) - \frac{42 \times (42+2)^2}{16} \right]$

$= \frac{64491}{164} = 393.2378.$

$C^* = \frac{C - E_0(C)}{\sqrt{Var_0(C)}} = \frac{239.5 - 231}{\sqrt{393.2378}} = 0.4286.$

Under $H_0: r^2=1$, $C^* \sim N(0,1)$ approximately.

The approximate p-value against $H_1: r^2 \neq 1$ is

$Pr(|C^*| \geq 0.4286) = 0.6682$ is large.

Thus we accept $H_0: r^2=1$, $var(X)=var(Y)$.

reject $H_1: r^2 \neq 1$, $var(X) \neq var(Y)$.