Semester 2	Statistical Tests (Advanced)	2011

Exercise 4 Solutions

1. The hypotheses to be tested are

$$H_0 : \mu_X = \mu_Y \quad \text{vs.} \quad H_1 : \mu_X < \mu_Y,$$

where μ_X and μ_Y denote the average breaking point of the string corresponding to the standard technique and the new technique, respectively.

The ranks corresponding to the standard technique and the new technique are given as follows:

```
Combined: 126 131 132 134 139 143 144 147
   Ranks:  1   2   3   4   5   6   7   8

Standard: 144 131 155 126 134   New: 139 154 132 143 147
   Ranks:  7   2   10  1   4          5   9   3   6   8
```

By using the Wilcoxon rank-sum test, the p-value for testing the hypothesis is given by

$$p = \mathbf{P}(W \le w),$$

where by using the information contained in the sample,

$$w = 7 + 2 + 10 + 1 + 4 = 24.$$

With $n_1 = n_2 = 5$, it follows from the tables that

$$p = \mathbf{P}(W \le 24) = 0.2738.$$

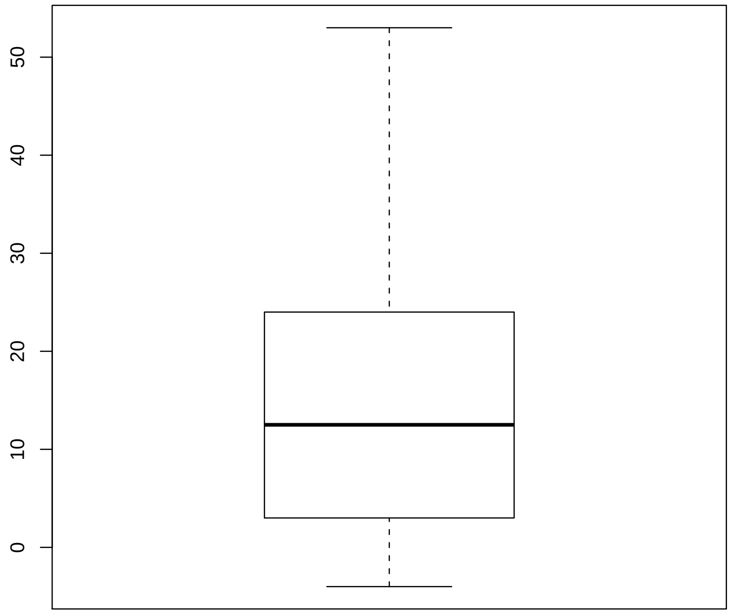
Conclusion: The data are consistent with the null hypothesis H_0.

2. (a)

Before	109	57	53	57	68	72	51	65	52	61
After	56	44	55	40	62	46	48	41	56	49
Difference($x_i - y_i$)	53	13	-2	17	6	26	3	24	-4	12

Comment: The boxplot (next page) does not look perfectly symmetric. Hence it is possible that the sample is not from a normal population. Also, the sample size is small so it might be better to perform a nonparametric test (just in case).

Boxplot:



(b) The hypotheses to be tested are

$$H_0 : \mu = 0 \quad \text{vs.} \quad H_1 : \mu \neq 0,$$

where μ is the median of the difference between before and after treatment. Let X be the number of positive $(x_i - y_i)$'s. Then the observed value of X is 8, and hence the p-value of the sign test is

$$p = 2\mathbf{P}(B(10, 1/2) \geq 8) = 2\mathbf{P}(B(10, 1/2) \leq 2) = 2 \times 0.0547 = 0.1094.$$

The p-value is greater than 0.10. Hence the data are consistent with the null hypothesis H_0.

3. Note

$$\begin{aligned}
\text{Var}(W) &= \frac{mn}{N(N-1)}\left[\sum_{i=1}^{m+n} i^2 - \frac{N(N+1)^2}{4}\right] \\
&= \frac{mn}{N(N-1)}\left[\frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)^2}{4}\right] \\
&= \frac{mn(N+1)}{12(N-1)}[2(2N+1) - 3(N+1)] \\
&= \frac{mn(N+1)}{12}.
\end{aligned}$$

4. The mid-ranks are

$$2, 2, 2, 4.5, 4.5, 6.5, 6.5.$$

To obtain the distribution of W consider all possible samples of size 4 drawn from this list without replacement. There are $\binom{7}{4} = 35$ terms. List of distinct outcomes:

Mid-ranks	W	Number of terms
2, 2, 2, 4.5	10.5	2
2, 2, 2, 6.5	12.5	2
2, 2, 4.5, 4.5	13	3
2, 2, 4.5, 6.5	15	12
2, 2, 6.5, 6.5	17	3
2, 4.5, 4.5, 6.5	17.5	6
2, 4.5, 6.5, 6.5	19.5	6
4.5, 4.5, 6.5, 6.5	22	1

Distribution:

w:	10.5	12.5	13	15	17	17.5	19.5	22
$\mathbf{P}(W = w)$:	$\frac{2}{35}$	$\frac{2}{35}$	$\frac{3}{35}$	$\frac{12}{35}$	$\frac{3}{35}$	$\frac{6}{35}$	$\frac{6}{35}$	$\frac{1}{35}$

Note the distribution is not necessarily symmetric when there are ties.

Check $\mathrm{E}(W) = 16$ and

$$\mathrm{Var}(W) = E(W^2) - [E(W)]^2 = 7.14286.$$

5. The ordered, combined sample and corresponding ranks are:

$$65, 65, 73, 75, 77, 77, 78, 82, 83, 85, 85, 86, 86, 86, 89, 90, 91, 92, 93, 97, 100$$

$$1.5, 1.5, 3, 4, 5.5, 5.5, 7, 8, 9, 10.5, 10.5, 13, 13, 13, 15, 16, 17, 18, 19, 20, 21$$

$m = 10$ and $n = 11$ so $N = 21$.

$$W = 1.5 + 1.5 + 3 + 4 + 5.5 + 7 + 9 + 10.5 + 16 + 20 = 78$$

If the two groups can be modelled by the same distribution then

$$\mathrm{E}(W) = 10 \times 22/2 = 110$$

and, using the mid-ranks and the formula in question 3,

$$\mathrm{Var}(W) = 200.750.$$

Thus the p-value for the two-sided test is

$$\begin{aligned} 2 \times \mathbf{P}(W \leq 78) &\simeq 2 \times \mathbf{P}\left(Z \leq \frac{78 - 110}{\sqrt{200.75}}\right) \\ &= 2 \times \mathbf{P}(Z \leq -2.2585) \\ &= 0.024. \end{aligned}$$

Thus, we conclude that the mean responses are not the same for the two groups. The Huntington's group has a higher mean score.