CHAPTER 13

Nonparametric and Distribution-Free Statistics

Nonparametric tests – these test hypotheses that are not statements about population parameters (e.g., χ^2 tests for goodness of fit and independence).

Distribution-Free tests – tests that make no assumptions about the sampled populations.

Note. These terms, in practice, tend to be used interchangably, most often with both under the umbrella of *nonparametric statistics*.

Advantages of nonparametric statistics

(1) Allow for the testing of hypotheses that are not statements about population parameter values.

(2) May be used when the form of the sampled population is unknown.

(3) Tend to be computationally easier and more quickly applied than parametric procedures.

(4) May be applied when the data being analyzed consist merely of rankings or classifications.

Disadvantages of nonparametric statistics

(1) The use of nonparametric procedures with data that can be handled with parametric procedures is a waste of time.

(2) The application of some nonparametric tests may be laborious for large samples.

Wilcoxon Signed-Rank Test for Location – for testing a null hypothesis about a population mean where neither z (small sample $[n < 30]$ from a population that is grossly nonnormally distributed, so Central Limit Theorem does not apply) nor t (the sampled population does not sufficiently approximate a normal population) is an appropriate test statistic.

Test assumptions on the data:

1) Random variable
2) Continuous variable
3) Population symmetric about μ
4) At least an interval scale

Calculations:

1) For $H_0 : \mu = \mu_0$, let $d_i = x_i - \mu_0$

   For $H_0 : \mu_1 = \mu_2$, let $d_i = x_{1,i} - x_{2,i}$

   Eliminate cases where $d_i = 0$, reducing n accordingly.
2) Rank the usable d_i from the smallest absolute value to the largest absolute value. If two or more of the $|d_i|$ are equal, assign each tied value the mean of the rank positions the tied values occupy.
3) Assign each rank the sign of the d_i that yields that rank.
4) Find T_+, the sum of the ranks with positive signs, and T_-, the sum of the ranks with negative signs.

Test statistic T:

For $H_A : \mu \neq \mu_0$, T is the smaller of T_+ and T_-

For $H_A : \mu < \mu_0$, $T = T_+$

For $H_A : \mu > \mu_0$, $T = T_-$

EXAMPLE (13.4.1). Using the data on page 696 of the text, we test the hypotheses

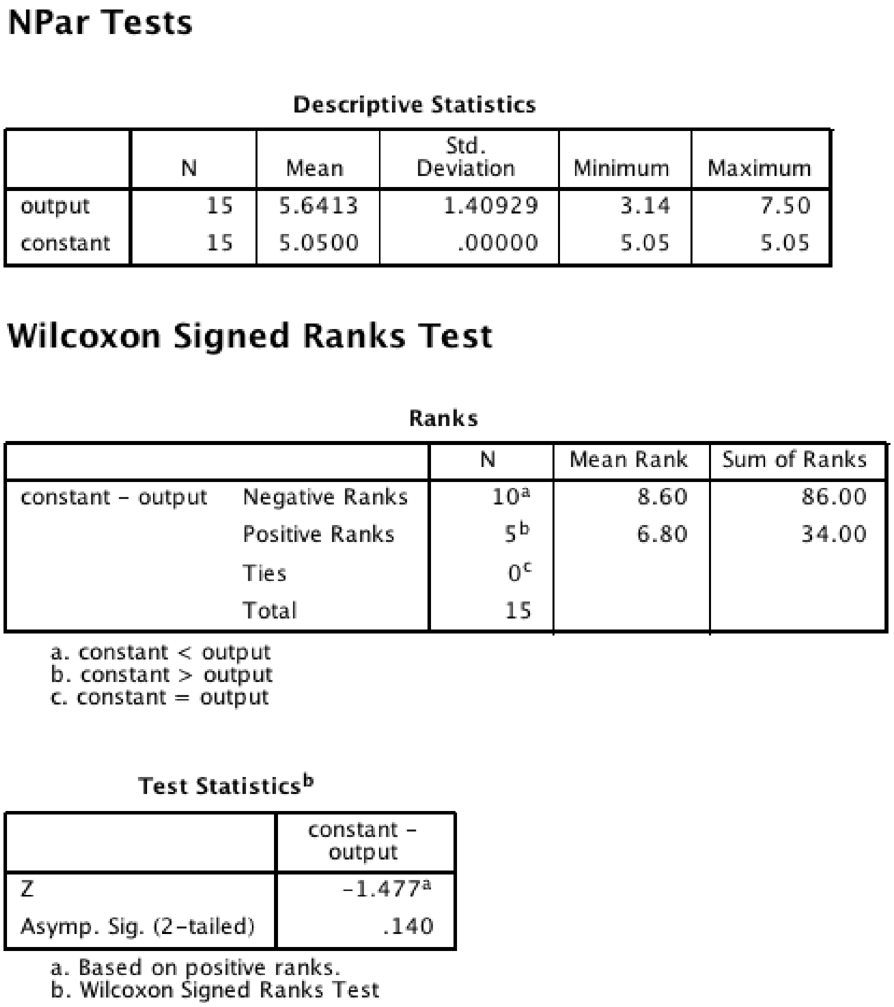
$$H_0 : \mu = 5.05$$

$$H_A : \mu \neq 5.05$$

at the $\alpha = 0.05$ level of significance. The critical value in Table K for $\frac{\alpha}{2} = .025$ and $n = 15$ is, by using .024, $T = 25$.

We compute $T_+ = 86$ and $T_- = 34$, yielding $T = 34$. Since $34 > 25$, we are unable to reject H_0. From Table K, we also get that $p = 2(.0757) = .1514$.

This same data is used on pages 39-40 of my SPSS manual, yielding the following output:



NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
output	15	5.6413	1.40929	3.14	7.50
constant	15	5.0500	.00000	5.05	5.05

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
constant – output	Negative Ranks	10[a]	8.60	86.00
	Positive Ranks	5[b]	6.80	34.00
	Ties	0[c]		
	Total	15		

a. constant < output
b. constant > output
c. constant = output

Test Statistics[b]

	constant – output
Z	-1.477[a]
Asymp. Sig. (2-tailed)	.140

a. Based on positive ranks.
b. Wilcoxon Signed Ranks Test

The Z is the standardized normal approximation to the test statistic, with $p = .140$ here.

Mann-Whitney test for equal medians for two independent samples

Assumptions:

(1) Independent random samples of size m and n

(2) At least ordinal scales

(3) Continuous variables

(4) Populations differ only with respect to medians

We test the hypothesis

$$H_0 : M_1 = M_2$$

against

$$H_A : M_1 \neq M_2, \qquad H_A : M_1 > M_2, \quad \mathit{or} \quad H_A : M1 < M_2$$

with level of significance $\alpha = .05$.

If the two populations are symmetric, so that within each population the mean and the median are the same, the conclusions we reach regarding the two population medians will also apply to the two populations means.

Example (3.6.1). We will consider three cases, all with level of significance $\alpha = .05$:

$$\begin{array}{lll} (a) & H_0 : M_X \geq M_Y & H_A : M_X < M_Y \\ (b) & H_0 : M_X \leq M_Y & H_A : M_X > M_Y \\ (c) & H_0 : M_X = M_Y & H_A : M_X \neq M_Y \end{array}$$

Calculations:

(1) Rank all variables from smallest to largest, yet keeping them separate. Handle ties as in Wilcoxon. See Table 13.6.2 on page 705 of the text.

(2) The test statistic is

$$T = S - \frac{n(n+1)}{2}$$

where n is the number of sample X observations and S is the sum of the ranks assigned to the sample observations from the population X. The choice of which sample's values we label X is arbitrary. In this example,

$$T = 145 - \frac{15(16)}{2} = 25.$$

(3) For (a), we reject H_0 since $25 < 45$ with 45 the critical value obtained from Table L. For (b), we would reject H_0 if $T > nm -$ critical value $= 150 - 45 = 105$. That is not the case here. For (c), we use $\alpha = .025$ in Table L to get critical values of 40 and $150 - 40 = 110$. We reject H_0 if $T < 40$ or $T > 110$. In our case, $T < 40$, so we reject H_0.

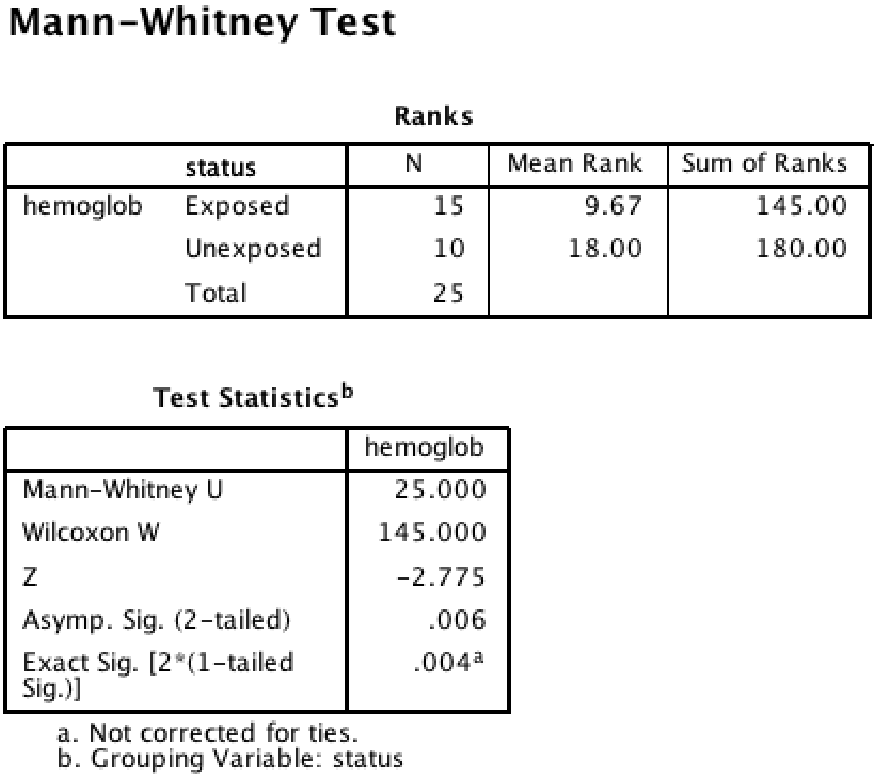
(4) Since $22 < 25 < 30$, we have $.001 < p < .005$ for the one-sided tests and $.002 < p < .01$ for the two-sided test.

(5) Table L does not work for n or m greater than 20. In general, if $nm \leq 400$ and $\frac{mn}{2} + \min(n, m) \leq 220$, the exact significance level is based on an algorithm of Dineen and Blakesley. This is what is in Table L. Otherwise, we compute

$$z = \frac{T - mn/2}{\sqrt{nm(n+m+1)/12}},$$

which is distributed approximately as a standard normal distribution.

The SPSS output for this example is given below.



Mann-Whitney Test

Ranks

	status	N	Mean Rank	Sum of Ranks
hemoglob	Exposed	15	9.67	145.00
	Unexposed	10	18.00	180.00
	Total	25		

Test Statistics[b]

	hemoglob
Mann-Whitney U	25.000
Wilcoxon W	145.000
Z	-2.775
Asymp. Sig. (2-tailed)	.006
Exact Sig. [2*(1-tailed Sig.)]	.004[a]

a. Not corrected for ties.
b. Grouping Variable: status

The Asymp. Sig. is based on Z.