Statistical Inference Theory
Lesson 46
Non-parametric Statistics

46.1-The Sign Test

46.1 - Problem 1:
➤(a).
Let p equal the proportion of supermarkets that charge less than $2.15 a pound.

H_o: $p \leq 0.50$
H_a: $p > 0.50$

H_a is the claim of the Agency.

➤(b).
Step 1:

Super Market	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Price $ per Pound	2.05	2.19	1.98	2.0	1.79	2.77	2.05	2.01	2.18	2.79
+/-	-	+	-	-	-	+	-	-	+	+

Step 2: Since we have no zeros, we assume N = 10.

Step 3: The number of + signs is $N_+ = 4$ and the number of minus signs is $N_- = 6$.

Therefore, there are 6 supermarkets that charge less than $2.15 a pound.

Step 4: Rule 2 assumes that the distribution of signs is binomial.

Step5 : Assume H_o is true. The following is the cumulative binomial distribution for N = 10 and p = 0.50.

k	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P\{N_+ \geq k\}$	1	0.99	0.99	0.94	0.83	0.62	0.38	0.17	0.05	0.01	0

Step 7: Since we have a one-sided test, we use $\alpha = 0.01$.

Using the above cumulation table, P{K ≥ 6} = 0.38 > 0.01, and therefore, we reject H_a . We have no statistical basis for accepting the Agency's claim.

46.2 - The Mann-Whitney U Test

46.2 - Problem 1:

➤(a).

H_o: There is no difference in the team's total scores between playing at home or away.

H_a: The team has greater scores when playing away from home.

➤(b).

We will apply the four rules above.

Step 1: Using rule 1 we have

56	57	60	61	65	66	67	68	72	74	75	77	88	89	91	96	98	99	100

101

Step 2: Using rule 2 we have

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
56	57	60	61	65	66	67	68	72	74	75	77	88	89	91	96	98	99	100

20
101

Step 3:Using rule 3, we have

Final scores of games played at home and rank		**Final scores of games played away from home and rank**	
Grade	**Rank**	**Grade**	**Rank**
67	7	65	5
89	14	91	15
72	9	74	10
66	6	100	19
101	20	56	1

98	17	68	8
77	12	57	2
99	18	60	3
96	16	61	4
88	13	75	11
Sum = R_1 =	132	**Sum = R_2** =	78

Step 4: Since $N_1 = 10$, $N_2 = 10$, $R_1 = 132$,

$$U = N_1N_2 + \frac{N_1(N_1 + 1)}{2} - R_1 = (10)(10) + \frac{10(10 + 1)}{2} - 132 = 23.$$

➤(c).

Step 1: $\mu = \frac{N_1N_2}{2} = \frac{10(10)}{2} = 50$

Step 2: $\sigma^2 = \frac{N_1N_2(N_1 + N_2 + 1)}{12} = \frac{10(10)(21)}{12} \approx 175$

Step 3: Since U is approximately normal, we use

$z = \frac{U - \mu}{\sigma}$ which is normal with mean 0 and variance 1.

Step 4: $z = \frac{U - \mu}{\sigma} = \frac{23 - 50}{13.23} \approx -2.04$

Step 5: Since we have a one-sided test, $\alpha = 0.05$. Therefore, the corresponding z value is $z = -1.64$.

Step 6. Since $-2.04 < -1.64$, we reject H_o and conclude there is a significant difference between the team playing at home or away.

46.3 - The Kruskal-Wallis H Test

46.3 - Problem 1:

➤(a).

H_o: There is no significant weight loss due to different drugs.

H_a: There is a significant weight loss due to different drugs.

➤(b).
Step 1: From the above table, combine and rank the data:

17.6	19.1	19.7	20.8	21.2	21.3	21.5	21.7	22.1	22.1	29	30.5	30.9	32.06	33.43
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Step 2: Apply this ranking for each brand:

Drug A	**Rank**	**Drug B**	**Rank**	**Drug C**	**Rank**
21.30	6	22.11	9	33.43	15
32.06	14	19.18	2	29.00	11
21.78	8	17.66	1	21.26	5
22.12	10	19.79	3	30.58	12
30.98	13	20.89	4	21.55	7
Sum $\mathbf{R_1}$ =	51	**Sum** $\mathbf{R_2}$ =	19	**Sum** $\mathbf{R_3}$ =	50

Step 3: Compute H from the formula:

$N = N_1 + N_2 + N_3 = 5 + 5 + 5 = 15.$

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{N_1} + \frac{R_2^2}{N_2} + \frac{R_3^2}{N_3} + \ldots + \frac{R_k^2}{N_k}\right) - 3(N+1) =$$

$$\frac{12}{15(15+1)}\left(\frac{51^2}{5} + \frac{19^2}{5} + \frac{50^2}{5}\right) - 3(15+1) = 6.62$$

➤(c).
Step 1: The degrees of freedom is k - 1 = 3 - 1 = 2.

Step 2: For $\alpha = 0.01$, the chi-square table give us $\chi^2 = 9.21$.

Step 3: Since H = 6.62 < 5.99, we conclude that at $\alpha = 0.01$ significance level we cannot conclude there is a significant difference in the weight reduction resulting from these drugs.

46.4 - The Spearman's Rank Correlation

46.4 - Problem 1:
Step 1: List the Patients' weight in ascending order and rank:

Patients' weight	175	197	210	211	225	235	237	255	278	298
Rank	1	2	3	4	5	6	7	8	9	10

Step 2: List the cholesterol level in ascending order and rank:

Cholesterol level	172	190	205	210	220	225	250	320	325	256
Rank	1	2	3	4	5	6	7	8	9	10

Step 3: We now place the ranking in the above table and compute D and D^2:

Patients' weight	**Rank**	**Cholesterol level**	**Rank**	**D**	**D^2**
175	1	225	6	1 - 6 = -5	25
197	2	205	3	2 - 3 = -1	1
210	3	220	5	3 - 5 = -2	4
211	4	320	8	4 - 8 = -4	16
225	5	190	2	5 - 2 = 3	9
235	6	172	1	6 - 1 = 5	25
237	7	250	7	7 - 7 = 0	0
255	8	210	4	8 - 4 = 4	16
278	9	325	9	9 - 9 = 0	0
298	10	256	10	10 - 10 = 0	0
					Sum = 96

Step 4: Since N = 7, $r \approx 1 - 58 = 0.42$.

Supplementary Problems

1.

rank	10	9	8	7	6	5	4	3	2	1
y	10	9	8	7	6	5	4	3	2	1

rank	1	2	3	4	5	6	7	8	9	10
x	1	2	3	4	5	6	7	8	9	10

x	**rank**	**y**	**rank**	**D**	**D^2**
1	1	10	10	1 - 10 = -9	81
2	2	9	9	2 - 9 = -7	49
3	3	8	8	3 - 8 = -5	25
4	4	7	7	4 - 7 = -3	9
5	5	6	6	5 - 6 = -1	1
6	6	5	5	6 - 5 = 1	1
7	7	4	4	7 - 4 = 3	9
8	8	3	3	8 - 3 = 5	25
9	9	2	2	9 - 2 = 7	49
10	10	1	1	10 - 1 = 9	81
					Sum = 330

$$\mathbf{r} = 1 - \frac{6(D_1^2 + D_2^2 + \dots + D_N^2)}{N(N^2 - 1)} = 1 - (6)\frac{330}{10(99)} = 1 - 2 = -1$$

2.
Let a = h and b = t.

From the above groups, $N_1 = 11$, $N_2 = 9$.

$$\mu = \frac{2(11)(9)}{11 + 9} + 1 = 10.9$$

$$\sigma^2 = \frac{2(11)(9)[2(11)(9) - 11 - 9]}{(11 + 9)^2(11 + 9 - 1)} \approx 4.64$$

3.

➤a.

Since we don't know the outcome of the sample of 200 chips, we estimate N_1 and N_2 , as follows: $N_1 = 0.05(200) = 10$(number of defective chips) and $N_2 = (0.95)(200) = 190$(number of non-defective chips).

$$\mu_R = \frac{2N_1N_2}{N_1 + N_2} + 1 = \frac{2(10)(190)}{10 + 190} + 1 = 20$$

$$\sigma_R^2 = \frac{2(10)(190)[(2(10)(190) - 10 - 190]}{(10 + 190)^2(10 + 190 - 1)} = 1.72$$

$\sigma_R \approx 1.31$

➤b.

If the system is stable then 5% or less are defective. Therefore,

H_o: $\mu_R \leq 20$ (The system is stable)

H_a: $\mu_R > 20$ (The system is not stable).

➤c.

Decision rule: If $c^* \leq R$ then assume the system is not stable; otherwise assume the system is stable.

$c^* = \mu + z\,\sigma_R = 20 + z1.31$

For $\alpha = 0.05$, from the normal distribution table, $z = 1.64$.

Therefore, $c^* = 20 + 1.64(1.31) \approx 22$.

Decision rule: If $22 \leq R$ then assume the system is not stable; otherwise assume the system is stable.

➤d.
We estimate N_1 = (0.08)(200) =16 (number of defective chips), N_2 = (0.92)(200) = 184 (number of non-defective chips).

$$\mu_R = \frac{2N_1N_2}{N_1 + N_2} + 1 = \frac{2(16)(184)}{16 + 184} + 1 = 30.44$$

$$\sigma_R^2 = \frac{2(16)(184)[(2(16)(184) - 16 - 184]}{(16 + 184)^2(16 + 184 - 1)} \approx 4.21$$

σ_R = 2.05

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{22 - 30.4}{2.05} \approx -4.10$$

Therefore from the normal distribution table, the probability is zero that the system is assumed stable. $\beta = 0$.

➤e.
Since the system is stable, from problem (a), μ = 20 and σ = 1.31.

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{25 - 20}{1.31} \approx 3.82$$

Therefore, from the normal distribution table, the probability is zero that the system will result in at least 25 runs.