Stat 403 - Solution to Assignment 2

Turned in Thursday, September 14, 2000

1. Use the sign test to analyze the data in problem 3.7 of the text. Also construct a confidence interval for the median difference. Use a confidence coefficient as close to 98% as possible. Report the actual level of confidence for your interval.

Specimen	1	2	3	4	5	6	7	8	9	10
Formula A	7.2	4.3	5.8	6.5	4.9	6.8	6.3	7.0	6.5	6.2
Formula B	5.1	4.1	5.5	4.1	5.0	5.1	5.3	7.3	4.8	5.8
Difference	2.1	0.2	0.3	2.4	-0.1	1.7	1.0	-0.3	1.7	0.4

H: $\eta_d = 0$
A: $\eta_d \neq 0$

$S_+ = 8 \quad S_- = 2$

Table F

n	Left S	P	Right S
10	2	0.0547	8

The P-value is 2(0.0547) = 0.1094. Since this P-value is not small, we cannot reject H. Given these data, Formula A and Formula B are about equal in terms of fading after exposure to sunlight.

98% Confidence Interval for the median difference.

–0.3, –0.1, 0.2, 0.3, 0.4, 1.0, 1.7, 1.7, 2.1, 2.4

Table F

n	Left S	P	Right S
10	1	0.0107	9

A 97.86% Confidence interval for the median difference will extend from the 2^{nd} ordered value to the 9^{th} ordered value. That is from –0.1 to 2.1. This covers zero and so a median difference of zero is acceptable.

2. Use the sign test to analyze the data in problem 3.11 of the text. Also construct a confidence interval for the median difference. Use a confidence coefficient as close to 90% as possible. Report the actual level of confidence for your interval.

Cage Number	1	2	3	4	5	6	7	8	9	10
First Rat	2.49	2.42	2.31	2.15	2.83	2.10	2.43	1.88	2.48	2.28
Second Rat	2.24	2.38	2.29	1.80	1.98	2.10	2.39	2.03	2.30	2.06
Difference	0.25	0.04	0.02	0.35	0.85	0.00	0.04	-0.15	0.18	0.22

H: $\eta_d = 0$
A: $\eta_d \neq 0$

$S_+ = 8 \quad S_- = 1$

Table F	Left		Right
n	S	P	S
9	1	0.0195	8

The P-value is 2(0.0195) = 0.0390. Since this P-value is smaller than 0.05, we should reject H. There is a significant difference in median total weight between the first and second rat.

90% Confidence Interval for the median difference.

–0.15, 0.00, 0.02, 0.04, 0.04, 0.18, 0.22, 0.25, 0.35, 0.85

Table F	Left		Right
n	S	P	S
10	2	0.0547	8

A 89.06% Confidence interval for the median difference will extend from the 3^{rd} ordered value to the 8^{th} ordered value. That is from 0.02 to 0.25. This does not cover zero, indicating a significant difference.

3. Salaries of professional athletes receive a good deal of attention in the press. The 1990 salaries of a random sample of 20 non-pitchers in baseball are given below, units are thousands of dollars.

100	100	111	114	165	210	225	225	230	250
410	575	750	900	1200	1900	2100	2100	2650	3300

We wish to investigate the central value of the distribution of salaries of non-pitchers in baseball.

(a) n=20 $\overline{X} = 880.75$ s=986.02

H: $\mu = 2000$
A: $\mu < 2000$

$$t = \frac{\overline{X}-\mu_o}{\frac{s}{\sqrt{n}}} = \frac{880.75-2000}{\frac{986.02}{\sqrt{20}}} = \frac{-1119.25}{220.48} = -5.076$$

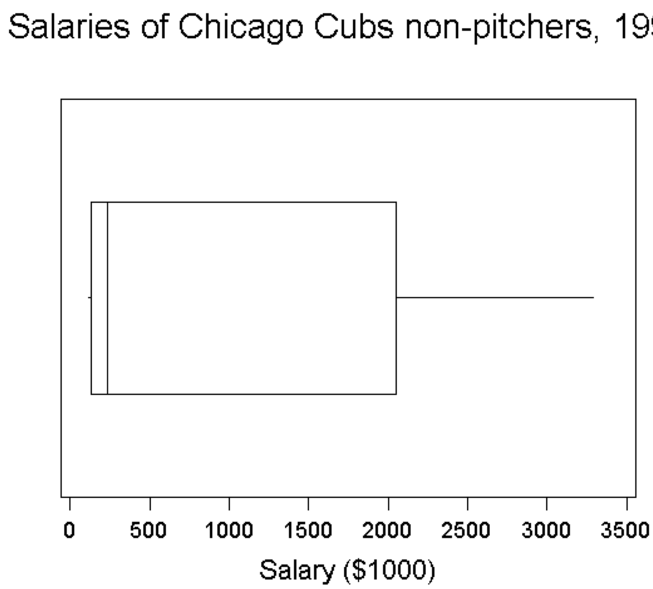
P-value is less than 0.0001. Reject H because the P-value is so small. There is evidence to indicate that the true mean salary is less than $2,000,000.

(b) 95% confidence and 19 df $\Rightarrow$ t* = 2.093

$$\overline{X} \pm t * \frac{s}{\sqrt{n}} = 880.75 \pm 2.093\frac{986.02}{\sqrt{20}} = 880.75 \pm 461.47$$

The 95% confidence interval for the mean salary goes from \$419,300 to \$1,342,200. This interval does not cover \$2,000,000 and so a median salary of \$2,000,000 is not acceptable.

(c) From the box plot we can see that the data is skewed to the right. This is inconsistent with the assumption of normality, since normal distributions are symmetric.



(d) Sign test for the median salary.

H: $\eta = 2000$
A: $\eta < 2000$

$S_+ = 4 \quad S_- = 16$

Table F	Left		Right
n	S	P	S
20	4	0.0059	16

The P-value is 0.0059. Since this P-value is very small, so reject H. There is evidence that the true median salary is less than \$2,000,000.

(e) 95% Confidence interval for the median salary.

Table F	Left		Right
n	S	P	S
20	5	0.0207	15

A 95.86% Confidence interval for the median salary will extend from the 6^{th} ordered value to the 15^{th} ordered value. That is from \$210,000 to \$1,200,000. This interval does not cover \$2,000,000 and so a median salary of \$2,000,000 is not acceptable.

4. The sign test for the one-sample location problem can be stated in terms of the binomial test in the following way. For $H : \eta = \eta_o$ versus $A_+ : \eta > \eta_o$ we define the probability of a success as $\theta = P(X > \eta_o)$ and test $H : \theta = \frac{1}{2}$ versus $A_+ : \theta > \frac{1}{2}$. Instead suppose we are interested in the upper quartile, q_U such that $P(X < q_U) = 0.75$ and $P(X > q_U) = 0.25$. A random sample of 15 students take an achievement test upon entering college. Their scores are given below. Test the hypothesis $H : q_U = 193$ versus $A_+ : q_U > 193$. Be sure to include the value of your test statistic, exact P-value, decision and a conclusion within the context of the problem.

189	233	195	160	212
179	231	185	199	213
202	193	174	166	248

In order to test the hypotheses

$H : q_U = 193 \quad H : \theta = 0.25$
$A : q_U > 193 \quad A : \theta > 0.25$

one needs to count the number of observations less than the hypothesized q_U (this is S_-) and the number of observations greater than the hypothesized q_U (this is S_+). Any observation equal to the hypothesized q_U is deleted and the effective sample size is reduced.

160, 166, 174, 179, 185, 189, 193
195, 199, 202, 212, 213, 231, 233, 248

$S_- = 6 \quad S_+ = 8 \quad n = 14$

$$\begin{aligned} P - value &= Pr(Bin(n = 14, \theta = 0.25) \geq 8) \\ &= 1 - Pr(Bin(n = 14, \theta = 0.25) \leq 7) \\ &= 1 - 0.9897 = 0.0103 \end{aligned}$$

Reject H, since the P-value is small (smaller than 0.05). There is evidence that the upper quartile is larger than 193. Note that the sample upper quartile is 213.