Math 553. 3rd Homework. Due Tuesday, February 27, 2001.

Homework on Chapter 3, The one sample location problem.

Name: ....

Show all your work. Please, send me the Splus programs which you do my e-mail.

1. Two types of fertilizers were used in different portions of 12 fields. Suppose that the yield of these fields were:

Type A	10.7	12.65	10.89	11.7	10.7	10.7	13.5	11.3	12.4	10.5	10.6	12.3
Type B	10.5	12.3	10.5	12.1	9.6	10.3	12.6	9.3	12.2	10.8	9.2	11.2

Find the p–value of this data to check the alternative hypothesis the type A fertilizer is better than the type B fertilizer using the following tests: (a) the sign test is used, (b) the t test is used, (c) the Wilcoxon signed–rank test is used. Find the t-statistic t (for paired observations), the Wilcoxon signed–rank test T^+ and the sign statistic B. Comment on your results.

2. The price index values for 30 homes in thousand of dollars in a suburban area of the Northeast are:

$$0, 50, 56, 65, 72, 80, 82, 83, 99, 101, 110, 112, 117, 120, 140$$

$$144, 145, 150, 180, 201, 210, 220, 240, 290, 309, 320, 325, 400, 500, 507$$

Find $\bar{\theta} = \bar{x}$, $\hat{\theta} = \text{median}\{\frac{Z_i+Z_j}{2} : 1 \le i \le j \le n\}$ and $\tilde{\theta} = \text{median}\{Z_i : 1 \le i \le n\}$ for the previous data. Comment: the considered estimators are competing estimators for the location of a symmetric distribution. If the the distribution is not symmetric every estimator is estimating something else. In this case, the distribution is skewed to the left. So, the mean is expected to be higher than the median.

3. A random sample consisting of 20 people who drove automobiles was selected to see if the alcohol affected reaction time. Each driver's reaction time was measured in a laboratory before and after drinking a specified amount of a beverage containing alcohol. The reaction times in seconds were as follows:

Subject	1	2	3	4	5	6	7	8	9	10
Before	.68	.64	.68	.82	.58	.80	.72	.65	.84	.73
After	.73	.62	.66	.92	.68	.87	.77	.70	.88	.79

Subject	11	12	13	14	15	16	17	18	19	20
Before	.65	.59	.78	.67	.65	.76	.61	.86	.74	.88
After	.72	.60	.78	.66	.68	.77	.72	.86	.72	.97

Does alcohol affect the reaction time? Find the p–value of this data to check whether alcohol affect the reaction time using the following tests: (a) the sign test, (b) the t test, (c) the Wilcoxon signed–rank test. Find the t-statistic t (for paired observations), the Wilcoxon signed–rank test T^+ and the sign statistic B. Comment on your results. Recall that when doing the sign test and the Wilcoxon signed–rank test you have to remove the observations with $Z_i = 0$. Since, there are ties, Splus does not supply the null distribution of T^+. Use simulations to estimate the p-value for the Wilcoxon signed–rank test. Recall that under the null hypothesis T^+ has the distribution of $\sum_{i=1}^n R_i\xi_i$, where $R_1, \dots, R_n$ are the midranks and $\xi_1, \dots, \xi_n$ are i.i.d.r.v.'s with distribution $\Pr\{\xi_i = 1\} = \Pr\{\xi_i = 0\} = 1/2$.

4. The sign statistic is $B = \sum_{i=1}^n I(Z_i > 0)$, where $Z_1, \dots, Z_n$ are independent r.v.'s. Under the null hypothesis $Z_1, \dots, Z_n$ are symmetric and continuous and $P(B = k|n) = \frac{b(k,n)}{2^n}$, where $b(k,n)$ is the number of ways to choose k places with $Z_i > 0$ from a total of n places.

(a) Show that $b(k,n) = b(k, n-1) + b(k-1, n-1)$. Hint: consider separately the number of ways of selecting the k places from $\{1, \dots, n\}$. (Do you remember the Tartaglia triangle?)

(b) Show that, under the null hypothesis, $P(B = k|n) = \frac{P(B=k-1|n-1)+P(B=k|n-1)}{2}$.

5. The Wilcoxon signed–rank test is defined as $T^+ = \sum_{i=1}^n R_i I(Z_i > 0)$, where $R_1, \dots, R_n$ are the ranks of $|Z_1|, \dots, |Z_n|$. Here, $Z_1, \dots, Z_n$ are i.i.d.r.v.'s. Under the null hypothesis, $|Z_1|, \dots, |Z_n|$ have a continuous symmetric distribution. So, we may suppose that $|Z_1|, \dots, |Z_n|$ are all different. Under the null hypothesis $P(T^+ = k|n) = \frac{t^+(k,n)}{2^n}$, where $t^+(k,n)$ is the number of ways to select a collection of ranks from $\{1, \dots, n\}$ which add k.

(a) Show that $t(k,n) = t(k,n) + t(k-n,n)$.

(b) Show that, under the null hypothesis, $P(T^+ = k|n) = \frac{P(T^+=k-n|n-1)+P(T^+=k|n-1)}{2}$.

(c) Use (b) to find the distribution of T^+, under the null hypothesis, when $n = 1, 2, 3, 4$.

6. Do and run Splus programs to find 10,000 estimators $\hat\theta_1, \dots, \hat\theta_{10{,}000}$ for θ using a random sample of size 20 from each of the following distributions: normal, uniform with minimum=-1 and maximum =1, double exponential and Cauchy (stable with index=1

and skewness =0). Find the simulated mse $\hat{MSE} := \frac{1}{10000}\sum_{i=1}^{10000}(\hat{\theta}_i - \theta)^2$. This number estimate the mean square error of the estimator $E[(\hat{\theta} - \theta)^2]$. Repeat for $\bar{\theta}$ and $\tilde{\theta}$. Do a program for each of the distributions! Do a table of $\frac{\hat{MSE}(\bar{\theta})}{\hat{MSE}(\hat{\theta})}$! Is this table close to the table in (3.116) in texbook? Do a table of $\frac{\hat{MSE}(\bar{\theta})}{\hat{MSE}(\tilde{\theta})}$! Is this table close to the table in (3.118) in texbook?

7. Show that $\hat{\theta} = \text{median}\{\frac{Z_i+Z_j}{2} : 1 \le i \le j \le n\}$ is asymptotic normal, i.e. show that

$$n^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} N\left(0, \frac{1}{12(\int_{-\infty}^{\infty} f^2(x)\,dx)^2}\right),$$

assuming that $Z_1, \ldots, Z_n$ are i.i.d.r.v.'s from a distribution which is symmetric about θ and it has a density $f(x)$ with $\int_{-\infty}^{\infty} f^2(x)\,dx < \infty$.