Stat {4601, 6872} Midterm (90 min.) Name_____

- Open book and open note; Use a simple calculator if necessary.
- Show your work. (E.g. Next to each numerical answer, you need to put things like $P_{0.01}(B \geq 3)$ or something equivalent) An asterisks (*) means a hard question.

The following are head width measurements (in micrometer) of 10 common mayfly species, *Stenacron interpunctatum*, in a certain habitat.

#	Head Width	#	Head Width
"1"	36	"6"	33
"2"	31	"7"	27
"3"	30	"8"	18
"4"	27	"9"	19
"5"	20	"10"	28

#1. (60 points) Use procedures based on **signed ranks** to answer the following questions.

(a) What assumptions do you make about the distribution of the measurements when you employ these procedures?

(b) Test the hypothesis that the median head width θ for mayflies from the habitat is 22 μm against alternative that it is greater than 22. Compute the P-value.

(c) [*Use only the five numbers on the first column for the next 3 problems*] Obtain a point estimate of θ.

(d) [*Use only the 5 numbers*] Find a lower confidence bound for θ with confidence coefficient .906.

(e) [*Use only the 5 numbers*] Obtain a confidence interval for θ with confidence coefficient .938.

(f) [*Use only the 5 numbers*] Use the large sample approximation to the confidence bound for θ with the same confidence coefficient as in (e). Be sure to be conservative.

#2. (60+10 points) Use procedures based on **signs** to answer the following questions. (Use all 10 numbers)

(a) What assumptions do you make about the distribution of the measurements if you employ these procedures?

(b) Test the hypothesis that θ is 22 μm against alternative that it is greater than 22.

(c) * What is the power of the test when the true θ is 25 μm?

(d) Obtain a point estimate of the θ.

(e) Find a lower confidence bound for θ with confidence coefficient .9453.

(f) Obtain a confidence interval for θ with confidence coefficient .9786.

(g) Use the large sample approximation to compute the confidence bound for θ with the same confidence coefficient as in (f). Be sure to be conservative.

#3. (40+30 points) Use procedures based on **Z-test** (not t-test) to answer the following questions, assuming population standard deviation is know to be σ=6.12. (Use table A.1) Use the pre-computed sample mean $\overline{Z}$ =26.95.

(a) What assumptions do you make about the distribution of the measurements if you employ these procedures?

(b) * Which of the three assumptions made so far (#1 (a), #2 (a), #3 (a)) is strongest? Which is weakest? What're the advantages of using procedures based on weak assumptions? What're the advantages of using ones based on strong assumptions?

(c) * How would you decide which of the three procedures (#1, #2, #3) to use?

(d) Test the hypothesis that θ is 22 μm against alternative that it is greater than 22.

(e) * What's the power of the test when the true θ is 25 μm?

(f) Obtain a point estimate of θ.

(g) Obtain two confidence intervals for θ with confidence coefficient .938 and. 9786, respectively.

Midterm Solution

#1. (7 points) Use procedures based on **signed ranks** to answer the following questions.

Table for Wilcoxon signed rank test

```
       z       R psi
 [1,] 14 14 10.0   1 10.0
 [2,]  9  9  8.0   1  8.0
 [3,]  8  8  7.0   1  7.0
 [4,]  5  5  4.5   1  4.5
 [5,] -2  2  1.0   0  0.0
 [6,] 11 11  9.0   1  9.0
 [7,]  5  5  4.5   1  4.5
 [8,] -4  4  3.0   0  0.0
 [9,] -3  3  2.0   0  0.0
[10,]  6  6  6.0   1  6.0
```

Here are the Walsh averages:

```
     [,1] [,2] [,3] [,4] [,5]
[1,]   14 11.5 11.0  9.5  6.0
[2,]   NA  9.0  8.5  7.0  3.5
[3,]   NA   NA  8.0  6.5  3.0
[4,]   NA   NA   NA  5.0  1.5
[5,]   NA   NA   NA   NA -2.0
Ordered Walsh averages:
-2.0  1.5  3.0  3.5  5.0  6.0  6.5  7.0  8.0  8.5  9.0  9.5 11.0 11.5 14.0
Ordered Zs.
18.0 19.0 20.0 27.0 27.5 28.0 30.0 31.0 33.0 36.0
```

(a) The values are independent and the distribution is continuous and symmetric around a common median.

(b) T^+=49. For n=10, P(T^+>=49) = .014.

(c) $\hat{\theta}$=22+7.0=29.0

(d) (1-α)=.906.
α=.094.
t_α = 13.
C_α=n(n+1)/2 +1 - t_α = 15+1-13 = 3.
$W^{(3)}$=3.0 (3.0, ∞)
In terms of the original Z, it's (25.0, ∞).

(e) (1-α)=.938.
α/2=.031.
t_α = 15.
Cα=n(n+1)/2 +1 - tα = 15+1- 15= 1.
($W^{(1)}$,$W^{(15)}$)=(-2.0, 14.0)
In terms of the original Z, it's (20, 36).

(f) C_α = n(n+1)/4 - z_α/2 sqrt(n*(n+1)*(2*n+1)/24) =
= 7.5 – 1.866 * 3.7 = 0.58
Closest integer is 1. Answer is same as (e): (20, 36).

#2. (7 points) Use procedures based on **signs** to answer the following questions.

(a) The values are independent and the distribution continuous and has a common median

(b) B=7. For n=10, $P_{1/2}$(B>=7)= 0.172

(c) B doesn't change. (no # between 22 and 25). So the answer is 0.172, just as (b).

(d) $\tilde{\theta}$ =27.5

(e) (1-α)=.9453,
α=.0547,

$C\alpha = n+1-b_{\alpha,1/2} = 10+1-8=3.$

$Z^{(3)}=20.$

(f) $(1-\alpha)=.9786,$

$\alpha/2 = (1-.9786)/2= .0107.$

$C\alpha = n+1-b_{\alpha/2,1/2} = 10+1-9=2.$

$(Z^{(2)}, Z^{(9)})=$ (19, 33).

(g) $C\alpha = n/2 - z\alpha/2 \text{ sqrt}(n/4)$

$= 5 - 2.30*1.58 = 1.36$

Pick $C\alpha=1$. Confidence interval = $(Z^{(1)}, Z^{(10)})$ = (18, 36)

#3. Use procedures based on **Z-test**

(a) The values are independent and the distribution is normal with common median(=mean) and SD.

(b) Normal procedures make strongest assumptiongs (i); sign procedures weakest (ii)

Advantage of (i): efficient if the assumed model is correct

Advantage of (ii): broader applicability; robust

(c) z =(Z-bar – 22)/sqrt(6.12/10) = 2.53.

P-value is P(Z > 2.53) = 0.0028 (if t-test is used, 0.0161)

(d) Field knowledge; visual inspection; formal test.

(e)

(f) $\overline{Z}$ =26.9

(g) (23.3, 30.5) and (22.4, 31.4)