BST 140.652
Problem Set 8

Problem 1. Obstetric records of a group of children who died suddenly and unexpectedly, S.U.D., were compared with those of a group of live control children. Observations on the duration of the 2nd state of labor were as follows:

Time in minutes:

S.U.D.	60,	25,	6,	8,	<5,	10,	25,	15,	10,		
Controls	13,	20,	15,	7,	75,	120*,	10,	100,	9,	25,	30

* Terminated by surgical intervention

Compare the median duration of labor in the two groups and test the significance of the difference using the Wilcoxon test.

Problem 2. Independent random samples of size 4 and 2 are taken from 2 continuous populations. Enumerate all possible rank collections for the smaller sample. Obtain the null distribution of the Mann-Whitney (Wilcoxon-Rank Sum) test statistic.

Problem 3. A large survey of over 100,000 births in South Wales during the period 1956-1962 gave an incidence rate for spina bifida of 4.12 per 1,000 births. In a random sample of 1000 births, compute the probability of observing (i) no cases, (ii) one case, (iii) two cases. Using the following two approaches

a. The exact distribution based on the binomial distribution

b. Approximate probabilities based on a Poisson approximation to the binomial

Problem 4. A cohort study was performed to evaluate the relation between oral-contraceptive use and breast cancer in the age group 40-44. At enrollment, women were classified as current user, past user or never user of oral contraceptives.

OC Use	Cancer Cases	Person-Years
Current user	13	4,761
Past user	164	121,091
Never user	113	98,091

a. Calculate the breast cancer incidence rates in the 3 groups along with 95% confidence intervals. State your assumptions.

b. Is there a significant difference in incidence rates between current users and never users? How about between past users and never users? Perform hypothesis test, report a p-value and interpret your findings.

c. Calculate the ratio of incidence rates comparing current users versus never users along with a 95% confidence interval for this ratio.

Problem 5. Researchers comparing fMRI signals between a resting state and a active state in 10 different regions of the brain, found the following P-values

Region	1	2	3	4	5	6	7	8	9	10
P-value	.081	.011	.053	.0140	.016	.045	.046	.050	.003	.053

a. Controlling the FWER of .05, which regions would be rejected? (Interpret your results.)

b. Controlling the FDR at .05, which regions would be rejected? (Interpret your results.)

Problem 6. The following data show blood glucose levels in mg/kg in rabbits immediately before and two hours after the administration of an analgesic compound.

Investigate the effect of analgesia on blood glucose level by applying (1) the sign test, (2) the Wilcoxon signed rank test, (3) an appropriate parametric test.

Rabbit number	Before analgesia	After analgesia
1	158	206
2	119	134
3	122	204
4	89	105
5	111	96
6	135	171
7	138	212
8	122	134
9	127	177
10	127	136
11	137	137
12	120	117
13	118	127
14	126	140
15	134	153
16	134	147
17	125	131
18	124	131

Interpret your results. Comment on any assumptions the methods make. Please do these calculations first without a computer, then if you like, try with a computer. Describe the computer procedure you used and whether your results agree.

Problem 7. Medical records were reviewed to assess the frequency of medical malpractice. Two reviewers classified the same records. We wish to study the agreement between the reviewers. Perform an analysis to assess agreement between the reviewers ($+$ means classified as malpractice).

		Review A	
		$+$	$-$
Reviewer B	$+$	35	13
	$-$	21	249

Problem 8. A paired comparison experiment will be performed with $n = 15$ pairs of subjects.

(a) To test $H_0 : P+ = .5$ vs. $H_1 : P+ < .5$ with the sign test, determine the rejection region so that α is closest $.10$.

(b) Using the binomial table, find the power of the test corresponding to each of the alternatives $P+ = .05, .1, .2, .3, .4$. Sketch the power curve for all values of $P + (0 < P+ < 1)$.

Problem 9. Consider the Wilcoxon-Signed Rank Test with $n = 4$.

(a) List all possible associations of signs with the ranks $1, 2, 3, 4$ and list corresponding values of T^+ and T^-.

(b) Obtain the probability distribution of T^+ and T^- for $n = 4$ using (a).

(c) From (b) calculate $\mathrm{E}(T^+), \mathrm{Var}(T^+)$. Also compute these from the formulas given for a general n.

Problem 10. A new drug is developed to relieve the ocular symptoms (redness, itching) of hay fever. An experiment is performed on a group of 24 patients with hay fever. In the experiment drug A is administered to a randomly selected eye and a placebo is administered to the other eye and the change from baseline is noted for each eye. The data are given; $+$ represents more improvement in the drug-treated eye; $-$ represents more improvement in the placebo-treated eye; 0 represents equal improvement in both eyes.

Comparison of drug A vs. placebo for the relief of redness and itching in hay fever patients

Subject	Redness	Itching	Subject	Redness	Itching	Subject	Redness	Itching
1	$+$	$+$	9	$-$	$+$	17	$-$	$+$
2	0	$+$	10	$-$	$-$	18	$-$	0
3	$-$	$-$	11	$+$	$+$	19	0	0
4	$+$	$-$	12	0	$+$	20	$-$	$+$
5	$+$	$+$	13	$+$	$+$	21	$-$	0
6	$+$	$+$	14	$-$	0	22	$-$	$+$
7	$+$	$+$	15	$+$	$+$	23	$+$	$+$
8	$-$	0	16	$-$	$-$	24	$-$	$+$

a. Why might a nonparametric statistical test be useful in comparing drug A with placebo for this experiment.
Why was it important to administer the placebo to the second eye of the same person rathe than to a different group of people with hay fever?

b. Compare the drug A and placebo eyes on redness and report an exact p-value. What procedure did you use? Why? Interpret. Also, perform a large sample approximation and compare with the exact result.
Compare the drug A and placebo eyes on itching and report an exact p-value. Also, perform a large sample approximation and compare with the exact result.

Problem 11. Two entities (I and R) in rural India collect data on births report the data below for 1945.

	I-list	
R-list	Present	Not
Present	794	710
Not	741	-

a. Assuming independence between the two groups, estimate the number of births in 1945 (give a confidence interval).

b. Plot a likelihood function for the number of births in 1945.