Math 159, Spring 2005
Professor Hardin
Exam #2 – sample questions

1. We observe: $X_1 = 2.1, X_2 = 1.9, X_3 = 2.6, X_4 = 3.3; Y_1 = 1.9, Y_2 = 2.6, Y_3 = 3.7$.

   (a) Apply the Mann-Whitney test, and find the **exact** p-value without enumerating all $\binom{7}{3}$ options or looking at a table.

   (b) Give the tails of the exact distribution for this situation (that is, values that are within the lower 20% and upper 20%.)

2. Consider the setup to Kendall's τ. Let there be N pairs of points, N_c are concordant, N_d are discordant, and $N - N_c - N_d$ are tied. Argue that:

$$c = \frac{N_c + 0.5(N - N_c - N_d) - N_d}{N}$$

   is a reasonable measure of correlation. Also indicate why we might want to use this measure (or not) over the other correlations we've considered.

3. Consider the following data of mothers' heights and daughters' heights (in inches) of 4 mother daughter pairs.

Mother	70	69	65	64
Daughter	67	63	62	64

   (a) Test for significant association using Spearman's ρ and Kendall's τ.

   (b) Test for significance of the slope for a linear model and find a 90% CI for the slope.

   (c) How would a larger sample size affect your results for (a) and (b)?

4. Students in an introductory statistics class were asked how many brothers and sisters they have and whether their hometown is rural, urban, or suburban.

Number of siblings in rural, urban, and suburban areas

Rural	3	2	1	1	2	3	2	2		3	2	1	1	4	1	1	1	1	6	2	2	2	1	1
Urban	1	0	1	1	0	0	1	1	1	8	1	1	1	0	1	1	2							
Suburban	1	2	1	1	0	2	3	2	2	2	1	1												

   Test whether there is a significant difference between rural and urban areas. (Hint: you can use the parameterization given on page 292, why?)

5. Thirty recreational basketball players were asked to shoot two free throws. Data on whether they made or missed their shots are show in the table below. The question of interest is whether the probability of making a shot on the first attempt is different from the probability of making a shot on the second attempt.

	Made Second Attempt	Missed Second Attempt
Made First Attempt	7	5
Missed First Attempt	14	4

   Let p_1 = the unknown probability that someone who makes the first shot will make the second shot; let p_2 = unknown probability that someone who misses the first shot will make the second shot.

   Use Fisher's Exact Test to asses the significance of:
   $H_o : p_1 = p_2$
   $H_1 : p_1 > p_2$

6. Explain why a randomization test (or permutation test) for the following 4 parameters of interest will give you equivalent results when comparing population 1 to population 2:

   (a) difference of variances

   (b) ratio of variances

   (c) difference of standard deviations

   (d) ratio of standard deviations

7. While waiting for a customer, a caddy saw eight golfers finish their round of golf, pay their caddies, and leave. He estimated the age of each golfer and noted how much they paid their caddies.

	Golfer							
	1	2	3	4	5	6	7	8
Age (estimated)	32	30	33	41	43	47	28	30
Amount Paid	10.00	11.40	9.00	12.00	16.00	17.00	8.75	10.50

   (a) Does there seem to be a tenancy for older golfers to pay their caddies more?

   (b) What other statistical methods could you have used in part (a)? What are the main advantages and disadvantages of each, including the method you used?

8. At the beginning of the year, a first-grade class was randomly divided into two groups. One group was taught to read using a uniform method, where all students progressed from one stage to the next at the same time, following the teacher's direction. The second group was taught to read using an individual method, where each student progressed at his or her own rate according to a programmed workbook, under supervision of the teacher. At the end of the year, each student was given a reading ability test with the following results:

1st Group	227	55	184	174	176	147	194	252	194	88
2nd Group	271	63	19	14	151	184	127	165	235	53

   (a) Test the null hypothesis that there is no difference in the two teaching methods against the alternative that the two population means are different.

      i. Use the exact distribution

      ii. Use the normal approximation

      iii. What are the advantages and disadvantages of exact vs. normal in this situation?

   (b) Find a 98% CI for the difference in means.