University of Regina
Department of Mathematics & Statistics
Final Examination
200730

Statistics 296
Non-Parametric Statistics

Instructor: Michael Kozdron

This exam is due at 5:00 pm on Thursday, December 13, 2007, in my office (College West 307.31).

Read all of the following information before starting this exam.

Please read all instructions carefully, and check your answers. Show all work neatly and in order, and clearly indicate your final answers. Answers must be justified whenever possible in order to earn full credit.

You may use standard notation; however, any new notations or abbreviations that you introduce must be clearly defined. You may cite results from our textbook Introduction to Modern Nonparametric Statistics, *by J.J. Higgins, and the associated laboratory manual without recopying them.*

You may consult any notes or books or websites that you wish, provided that proper acknowledgements are included. However, you may not discuss this exam with anyone other than the instructor before 5:00 pm on December 13, 2007. This includes students, professors, and colleagues.

*This exam has **8** problems, and it is worth a total of **130** points.*

1. (*16 points*) Suppose that a quality-control inspector wants to assess a machine that manufactures compact disks (CDs). The CDs are labelled as either acceptable (A) or defective (D). A sample of size 30 produced the following data:

D A A A A A A D D A D D A A A D D A A A A A A A D D D A A A

(a) Use the runs test to determine whether or not the sequence is random. Perform your test at the $\alpha = 0.05$ level.

(b) The machine is considered reliable if less than 25% of the CDs are defective. Does this data provide evidence at the $\alpha = 0.10$ significance level that the machine is unreliable?

Hint: Let p denote the true proportion of defective CDs and consider testing $H_0 : p = 0.25$ against $H_A : p > 0.25$. Conduct a binomial test (or an appropriate approximation).

2. (*16 points*) Polyphenols are a group of chemical substances that may have antioxidant characteristics with potential health benefits. In particular, they may reduce the risk of cardiovascular disease and cancer. Note that higher levels of polyphenols are desirable. A study comparing the health benefits of red and white wine was conducted and 18 individuals (all of the same gender, and all roughly of the same age and health) were randomly assigned to receive a glass of red wine each evening for two weeks or a glass of white wine. The following data give the percentage change in the level of polyphenols in the blood after the two week period.

Red Wine	3.5	8.1	7.4	4.0	0.7	4.9	8.4	7.0	5.5
White Wine	3.1	0.5	-3.8	4.1	-0.6	2.7	1.9	-5.9	0.1

Does this data support the theory that red wine raises polyphenols more than white wine?

(a) Which two-sample test is most appropriate in this situation? Why?

(b) Conduct the test you selected in **(a)** and make your decisions at the significance level $\alpha = 0.05$.

3. (*16 points*) A very small survey was conducted of undergraduate and graduate students at the University of Regina to see their preferences among four types of (non-alcoholic) beverages. The choices were: Pepsi, Coke, other non-diet sodas, and other diet sodas. Among undergraduates, the responses from 6 people were: 1, 1, 3, 1, favouring Pepsi, Coke, other non-diet sodas, and other diet sodas, respectively. Among graduates, the corresponding responses were 3, 2, 0, and 1. Does there exist a difference in beverage preference between undergraduate and graduate students? *Be sure that you carefully define null and alternative hypotheses. Perform an appropriate test and justify your choice. Make your conclusions at the* $\alpha = 0.10$ *significance level.*

4. (*16 points*) A study was conducted to examine the difference in house prices between four cities. Selling pricing were sampled from four houses in each city. The prices for City A were (in thousands): 289, 312, 308, 294. The prices for City B were: 288, 296, 305, 302. The prices for City C were: 318, 306, 325, 322. The prices for City D were: 324, 304, 475, 382. Based on these data, perform a Kruskal-Wallis test to decide if there is a difference in house prices between these cities. *Be sure that you carefully define null and alternative hypotheses. Make your conclusions at the* $\alpha = 0.10$ *significance level.*

5. (*24 points*) Suppose that a biologist wishes to see if there is a relationship between the heights of trees in a certain forest and their diameters. He selects 10 trees at random from the forest and measures their heights and diameters. To ensure consistent measurements, he records the diameters around the tree from a height of 1.4 metres above the ground.

Diameter (m)	Height (m)
26.0	79.6
24.1	97.8
11.5	66.8
12.8	85.6
19.3	48.5
16.4	25.3
18.0	58.2
14.9	43.0
11.2	70.7
13.9	32.9

Note: In case you are curious, there are some pretty simple ways to measure heights of trees. See `http://warrensburg.k12.mo.us/conundrums/treeht/help.html` for one such way.

(a) Draw a scatterplot of the data.

(b) Determine the equation of the linear regression line.

(c) Test for association using the Spearman rank correlation. *Be sure to carefully state your null and alternative hypothesis, and conduct the test at the* $\alpha = 0.05$ *significance level.*

(d) Test for linear association using Pearson's correlation coefficient. *Be sure to carefully state your null and alternative hypothesis, and conduct the test at the* $\alpha = 0.05$ *significance level.*

(e) Test for significance using Kendall's tau. Perform the test by hand using a normal approximation to determine the p-value. *Carefully state your null and alternative hypothesis, and conduct the test at the* $\alpha = 0.05$ *significance level.*

6. (*10 points*) In a breeding experiment, white chickens with small combs were mated and produced 190 offspring of the type shown below.

TYPE	NUMBER OF OFFSPRING
White feathers, small comb	111
White feathers, large comb	37
Dark feathers, small comb	34
Dark feathers, large comb	8

Use the chi-square goodness-of-fit test at significance level $\alpha = 0.10$ to determine if these data are consistent with the Mendelian expected ratios of 9:3:3:1 for the four types.

7. (*16 points*) One explanation for the widespread incidence of the hereditary condition known as sickle-cell trait is that it confers some protection against malarial infection. In one investigation, 543 African children were checked for the trait and for malaria. The results are shown in the following table.

	Heavily infected (with malaria)	Noninfected or lightly infected (with malaria)	
Yes (has sickle-cell trait)	36	100	136
No (does not have sickle-cell trait)	152	255	407
	188	355	543

(a) Based on the chi-squared test for 2×2 contingency tables, do the data provide evidence in favour of the explanation? *Conduct your test at the* $\alpha = 0.05$ *significance level with a directional alternative.*

(b) Based on Fisher's exact test for 2×2 contingency tables, do the data provide evidence in favour of the explanation? *Conduct your test at the* $\alpha = 0.05$ *significance level with a directional alternative.*

8. (*16 points*) A group of mountain climbers participated in a trial to investigate the usefulness of the drug acetazolamide in preventing altitude sickness. The climbers were randomly assigned to receive either drug or placebo during an ascent of Mt. Ranier. The experiment was to supposed to be double-blind, but the question arose whether some of the climbers might have received clues (perhaps from the presence or absence of side effects or from a perceived therapeutic effect or lack of it) as to which treatment they were receiving. To investigate this possibility, the climbers were asked (after the trial was over) which treatment they had received. The results can be cast in the following contingency table.

	Drug received	Placebo received
Correct guess	20	12
Incorrect guess	11	21

Alternatively, the same results can be rearranged in the following contingency table.

	Drug received	Placebo received
Drug guess	20	21
Placebo guess	11	12

Consider the null hypothesis

H_0: The blinding was perfect. (The climbers received no clues.)

Carry out a chi-squared test of H_0 against the alternative that the climbers did receive clues.

Hint: To clarify the issue for yourself, try inventing a fictitous data set in which most of the climbers have received *strong clues*; then arrange your fictitous data in each of the two contingency table formats and note which table would yield a larger value of χ_s^2.