ECE 315 - INTRODUCTION TO STATISTICS - INVEST 25
NONPARAMETRIC STATISTICS

WINTER 2004 **A.P. FELZER**

To do "well" on this investigation you must not only get the right answers but must also do neat, complete and concise writeups that make obvious what each problem is, how you're solving the problem and what your answer is. You also need to include drawings of all circuits as well as appropriate graphs and tables.

In the last four Investigations on statistics we've making use of the fact that the sample means $\overline{X}$ are normally distributed to estimate the population mean μ. In this Investigation we introduce nonparametric statistic tests that do not depend on the population being Gaussian. Note that despite their name, some nonparametric tests involve parameters like means.

1. We begin with a review problem. Suppose 23 out of 50 patients who took a placebo get better. And 29 out of 50 who took a new drug get better. Is the new drug effective

2. The objective of this problem is to introduce the nonparametric sign test for medians with an example. Suppose a certain university claims that the median starting salaries of its new graduates is $50,000 per year. To test this claim you take a random sample of 100 new graduates and find that 40 have salaries above and 60 below $50,000 per year. Is this reasonable to expect from a population with a median of $50,000 or was the university doing a little wishful thinking. Hint - note that the probability distribution for the number of students k in a sample of size n to be above the median is binomial with p = 0.5 since half the students have salaries at or above the median. So what you need to do is find the probability of 60 or more successes - of 60 or more students above the median - in a sample of 100

3. The objective of this problem is to introduce the **rank-sum test** by Wilconox for determining whether two continuous populations that aren't necessarily Gaussian have the same means. We do this by testing the null hypothesis $H_0\!: \ \mu_1 = \mu_2$ as follows

   (1) Obtain random samples of size n_1 and n_2 of the two populations

   (2) Arrange the $n_1 + n_2$ samples in order from smallest to largest

   (3) Assign the smallest sample value a rank of 1, the next smallest a rank of 2 and so on. If two samples - say the 5'th and 6'th - have the same value then they're both assigned the corresponding average rank 5.5

   (4) The decision to accept or reject the null hypothesis is then based on the values of

   $$u_1 = w_1 - \frac{n_1(n_1+1)}{2} \qquad \text{and} \qquad u_2 = w_2 - \frac{n_2(n_2+1)}{2}$$

   where w_1 = sum of the ranks of the samples from population 1

   w_2 = sum of the ranks of the samples from population 2

   a. Explain why $u_1 > 0$ and $u_2 > 0$
   b. Explain why $u_1 << u_2$ implies $\mu_1 < \mu_2$
   c. Explain why $u_1 >> u_2$ implies $\mu_1 > \mu_2$
   d. As we take different samples the values of u_1 and u_2 as well as u as follows

   $$u = \min\{u_1, u_2\}$$

will change. But we can put together a table for how small u has to be for us to reject the null hypothesis for different values of α. Suppose in particular we have the following samples from two populations

A: 2.3, 4, 1.7, 3, 2.5

B: 3, 2, 2.5, 1, 2, 1.5

Make use of the rank-sum table to decide whether to accept or reject the null hypothesis as follows with a significance of $\alpha = 0.05$

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$