STA 580 — Spring 2009 — Dr. Charnigo

Lecture 9

Nonparametric Techniques and Variable Types

Introduction. Suppose that we wish to compare two populations with respect to central tendency but that the populations are not normal and that the sample sizes are not large enough to invoke the Central Limit Theorem. In Lecture 8 we identified some strategies to handle such a situation:
• nonlinear transformation of the response variable followed by application of Lecture 7 methodology for comparing means;
• discretization of the response variable followed by application of Lecture 8 methodology for comparing proportions; and,
• nonparametric techniques, on which details were promised for Lecture 9.

Nonparametric techniques are also useful when the response variable is categorical and the categories can be placed into a meaningful order. While a concept of central tendency exists for such a variable, there is no mean *per se*. Thus, methodology for comparing means is not applicable.

Before discussing nonparametric techniques in detail, let us distinguish among several types of variables.

Nominal variables. We have a nominal variable, giving rise to nominal data (Definition 9.4), if:

• the possible values are either categories or artificial numerical designations for categories; and,

• the categories do not have a natural ordering.

Gender is a nominal variable. The possible values are "male" and "female". We can artificially designate males as 0's and females as 1's, but we can just as easily designate males as 1's and females as 0's. Other nominal variables include country of origin and first name.

Obviously a nominal variable cannot have a mean. For instance, there is no such thing as a mean gender.[1]

Ordinal variables. We have an ordinal variable, giving rise to ordinal data (Definition 9.3), if:

• the possible values are either categories or artificial numerical designations for categories; and,

• the categories have a natural ordering but differences between possible values are not meaningful.

The rating a faculty member receives for overall teaching effectiveness is an ordinal variable. The possible values are "far above average", "somewhat above average", "average", "somewhat below average", and "far below average". Clearly "far above average" is better than "somewhat above average", but the arithmetic difference

"far above average" minus "somewhat above average"

is meaningless.

The concept of central tendency exists for an ordinal variable. For instance, suppose that a faculty member receives a "far above average" rating from 20% of students, a "somewhat above average" rating from 40% of students, an "average" rating from 30% of students, a "somewhat below average" rating from 5% of students, and a "far below average" rating from 5% of students. We can define the median rating to be that rating which satisfies both of the following conditions: (i) no more than 50% of the students assign a higher rating; and, (ii) no more than 50% of the students

[1]If we designate males as 0's and females as 1's, we can compute the mean of the 0's and 1's. But that is a mean of artificial numerical designations, not a mean of gender. In fact, the mean of the 0's and 1's turns out to be precisely the proportion of females.

assign a lower rating. In our case, the median rating is "somewhat above average" since: (i) 20% ($\leq$ 50%) of students assign a higher rating; and, (ii) 40% ($\leq$ 50%) of students assign a lower rating. You can also verify that conditions (i) and (ii) are not satisfied for any rating except "somewhat above average".

However, an ordinal variable does not have a mean *per se*. For instance, an expression like

$$0.20 \text{ far above average } + 0.40 \text{ somewhat above average } +$$

$$0.30 \text{ average } + 0.05 \text{ somewhat below average } + 0.05 \text{ far below average}$$

is nonsensical.[2]

Cardinal variables. We have a cardinal variable, giving rise to cardinal data (Definition 9.1), if the differences between possible values are meaningful. If the quotients of possible values (assuming a nonzero denominator) are also meaningful, we have a ratio variable; otherwise we have an interval variable (Definition 9.2).

Body weight in kilograms is a ratio variable: someone who weighs 120 kilograms is twice as heavy as someone who weighs 60 kilograms. Outdoor temperature in Celsius degrees is an interval variable: there is not twice as much heat on a 20-degree day as on a 10-degree day.

Any continuous variable is cardinal, but a cardinal variable need not be continuous: consider, for example, the number of influenza cases in a city.

[2] If we designate "far above average" as 4, "somewhat above average" as 3, "average" as 2, "somewhat below average" as 1, and "far below average" as 0, we can compute the mean of the 0's, 1's, 2's, 3's, and 4's. But, again, that is a mean of artificial numerical designations. On the other hand, many people would contend that the mean of artificial numerical designations is informative in this situation. Such a view can be rationalized by regarding the teaching rating as an intrinsically continuous variable that is rounded off to the nearest integer and then assigned a label to expedite the process of faculty evaluation.

The sign test and the signed-rank test

Scenarios. The sign test and the signed-rank test, to be developed presently, apply in either of the following scenarios.

SCENARIO 1. For each of m subjects we obtain measurements on an ordinal or a cardinal variable. Let these be denoted $d_1, \ldots, d_m$. If the variable is ordinal, then we assume that one of the categories has been designated as "0", that all "better" categories have been designated as positive numbers, and that all "worse" categories have been designated as negative numbers. Letting Δ denote the population median, we wish to test

$$H_0 : \Delta = 0 \quad \text{against} \quad H_1 : \Delta \neq 0.$$

Note that, if the variable is cardinal, we do not necessarily have $\Delta = \mu$.[3]

SCENARIO 2. For each of m subjects we obtain measurements of a cardinal variable under two different treatments. Let the within-subject difference scores be denoted $d_1, \ldots, d_m$. Letting Δ denote the population median for the difference scores, we wish to test

$$H_0 : \Delta = 0 \quad \text{against} \quad H_1 : \Delta \neq 0.$$

This scenario is much like the one in which we employ the paired t-test, except that we have not assumed normality (a requirement for the paired t-test) and we are concerned with a population median rather than a population mean.

Sign test. If $H_0 : \Delta = 0$ is true, then we anticipate that about half of the nonzero d_i will be positive. So, we will reject H_0 if a lot more or a lot less than half of the nonzero d_i are positive.

[3] One case in which we do have $\Delta = \mu$ is if the population is normal.

Let n ($\leq m$) be the number of nonzero d_i. When $n \geq 20$ we reject H_0 if the number of positive d_i is (Equation 9.1) either greater than

$$n/2 + 1/2 + z_{1-\alpha/2}\sqrt{n/4}$$

(i.e., a lot more than $n/2$) or less than

$$n/2 - 1/2 - z_{1-\alpha/2}\sqrt{n/4}$$

(i.e., a lot less than $n/2$). Note that I have included the author's continuity correction ($\pm 1/2$) in the criterion for rejecting H_0. The author also provides a formula for the p-value (Equation 9.2) and a version of the sign test that can be used when $n < 20$ (Equation 9.3); you are not responsible for the latter in STA 580.

Example (sign test). Refer to Example 9.7 on page 361. We have $m = 45$ people who are highly susceptible to sunburn place Ointment A on one arm and Ointment B on the other arm. Then each person is exposed to sunlight for one hour. At the end of the hour, each person is assigned two "burn scores", one for the arm with Ointment A and one for the arm with Ointment B. Burn score is a cardinal variable whose possible values are $0, 1, \ldots, 10$ (higher scores represent more burn).[4] We let d_i denote the difference between person i's burn score for the arm with Ointment A and person i's burn score for the arm with Ointment B. Hence, a positive d_i indicates that Ointment B performed better for person i, while a negative d_i indicates that Ointment A performed better for person i.

The data are summarized in Table 9.1 on page 367. In particular, there are 22 negative differences (i.e., 22 people for whom Ointment A performed

[4] Your textbook author says that burn score is an ordinal variable, but if that were so then calculating differences $d_1, \ldots, d_m$ would be nonsensical.

better), 18 positive differences (i.e., 18 people for whom Ointment B performed better), and 5 zero differences (i.e., 5 people for whom Ointments A and B performed similarly). Thus, $n = 22 + 18 = 40$. Since 18 is obviously not "a lot more" than $n/2 = 20$, we only need to check whether 18 is "a lot less" than $n/2 = 20$. The cutoff is

$$20 - 1/2 - 1.96\sqrt{40/4} = 13.3,$$

so we accept H_0 at level $\alpha = 0.05$.

Refer to page 1 of {NONPARExamples.pdf}. In the "Tests for Location" box we see a p-value of 0.6358 in the "Sign" row, which agrees with our decision not to reject H_0. The -2 in the "Sign" row represents the difference between the number of nonzero d_i that were positive (18) and the number that we might have expected if H_0 were true (20).

Signed-rank test. With the sign test we use only the signs of the d_i (i.e., whether they are positive or negative). The signed-rank test uses both the signs and the magnitudes of the d_i.[5]

We begin by assigning ranks to the nonzero d_i using the procedure described in Equation 9.4 and illustrated in Table 9.1. For instance, there are 14 people for whom $|d_i| = 1$, so these people could be ranked 1 through 14. However, since there is no justification to rank one of these people ahead of the others, we assign them all the "average" rank of

$$7.5 = \frac{1 + 2 + \cdots + 14}{14}.$$

[5]The textbook author states that $d_1, \ldots, d_m$ should arise from an underlying continuous distribution that is symmetric about its median. I do not take this statement too literally, as it obviously rules out using the signed-rank test with ordinal data. The analysis of ordinal data was one of the author's principal motivations for introducing the signed-rank test!

Similarly, the 10 people for whom $|d_i| = 2$ get assigned the average rank of

$$19.5 = \frac{15 + 16 + \cdots + 24}{10}.$$

We continue in this manner until, finally, the single person for whom $|d_i| = 8$ is assigned the rank of 40. Then we add up the ranks for all of the people with positive d_i. Let this number be denoted r_1.

If H_0 is true, then we anticipate that r_1 will be close to $n(n+1)/4$. This is because we get

$$1 + 2 + \cdots + n = n(n+1)/2$$

if we add up the ranks for all of the people with nonzero d_i. So, if H_0 is true, adding up the ranks for all of the people with positive d_i should yield about half of $n(n+1)/2$, which is $n(n+1)/4$. More specifically, when $n \geq 16$ we reject H_0 if (Equation 9.5) the test statistic

$$\frac{|r_1 - n(n+1)/4| - 1/2}{\sqrt{n(n+1)(2n+1)/24 - \sum_j (t_j^3 - t_j)/48}}$$

is greater than $z_{1-\alpha/2}$, where t_j is the number of people for whom $|d_i| = j$ (i.e., t_j corresponds to an entry in the "Number of people with same absolute value" column of Table 9.1).

Example (signed-rank test). We continue with the previous example. Since there are 10 people with $d_i = 1$ (assigned the average rank of 7.5), 6 people with $d_i = 2$ (assigned the average rank of 19.5), and 2 people with $d_i = 3$ (assigned the average rank of 28.0), we obtain

$$r_1 = 10 \times 7.5 + 6 \times 19.5 + 2 \times 28.0 = 248.$$

Also, with $n = 40$ we have $n(n+1)/4 = 410$. Noting that $t_1 = 14, t_2 = 10, \ldots, t_8 = 1$, we obtain

$$\sqrt{n(n+1)(2n+1)/24 - \sum_j (t_j^3 - t_j)/48} = \sqrt{5449.75} = 73.82.$$

Hence, the test statistic is

$$\frac{|248 - 410| - 1/2}{73.82} = 2.19.$$

Since 2.19 is greater than 1.96, we reject H_0 at level $\alpha = 0.05$. This differs from our decision to accept H_0 when we carried out the sign test. Thus, only by taking into account both the signs and the magnitudes of the d_i were we able to establish a preference for Ointment A: indeed, some people fared very much better with Ointment A than with Ointment B, while nobody fared much worse with Ointment A.

Refer to page 1 of {NONPARExamples.pdf}. In the "Tests for Location" box we see a p-value of 0.0262 in the "Signed Rank" row, which agrees with our decision to reject H_0. The -162 in the "Signed Rank" row represents the difference between the sum of the ranks corresponding to the positive d_i (248) and the number that we might have expected if H_0 were true (410).

The rank-sum test

Scenario. We obtain measurements of an ordinal or a cardinal variable for n_1 subjects under one treatment and for n_2 different subjects under another treatment. Let $x_1, \ldots, x_{n_1}$ denote the measurements for the first n_1 subjects and $y_1, \ldots, y_{n_2}$ the measurements for the other n_2 subjects. Let Δ_x denote the population median corresponding to the first treatment and Δ_y the population median corresponding to the second treatment. We wish to test

$$H_0 : \Delta_x = \Delta_y \quad \text{against} \quad H_1 : \Delta_x \neq \Delta_y.$$

This scenario is much like the one in which we employ an independent samples t-test, except that we have not assumed normality (a requirement for an independent samples t-test) and we are concerned with population medians rather than population means.

Rank-sum test. We begin by assigning ranks to the x_i and y_i using the procedure described in Equation 9.6 and illustrated in Table 9.3. For instance, all 6 people in the "cluster" of 20-20 visual acuity are assigned the average rank of

$$\frac{1+2+\cdots+6}{6} = 3.5.$$

Then we add up the ranks corresponding to the x_i (i.e., for the subjects receiving the first treatment). Let this sum be denoted by r_1.

If H_0 is true, then we anticipate that r_1 will be close to $n_1(n_1+n_2+1)/2$. When $n_1, n_2 \geq 10$ we reject H_0 if (Equation 9.7) the test statistic

$$\frac{|r_1 - n_1(n_1+n_2+1)/2| - 1/2}{\sqrt{\left(\frac{n_1 n_2}{12}\right)\left[n_1+n_2+1-\frac{\sum_j t_j(t_j^2-1)}{(n_1+n_2)(n_1+n_2-1)}\right]}}$$

is greater than $z_{1-\alpha/2}$, where t_j is the number of people in the j^{th} cluster (i.e., t_j corresponds to an entry in the "Combined sample" column of Table 9.3).[6]

Example (rank-sum test). Refer to Example 9.15 on page 372. We are interested in investigating whether, among retinitis pigmentosa (RP) patients aged 10 through 19, there is an association between visual acuity and the genetic form of RP (dominant or sex-linked). Thus, we assess visual acuity for $n_1 = 25$ dominant-RP patients and $n_2 = 30$ sex-linked-RP patients. The data are summarized in Table 9.3. For instance, 5 dominant-RP patients have 20-20 visual acuity, while only 1 sex-linked-RP patient has

[6] Your textbook author also states that $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$ should arise from underlying continuous distributions. My previous objection applies.

20-20 visual acuity. Each of these 6 patients is assigned the average rank of 3.5.

Since there are 5 dominant-RP patients with average rank 3.5, 9 with average rank 13.5, 6 with average rank 25.5, 3 with average rank 34.0, and 2 with average rank 42.5, we have

$$r_1 = 5 \times 3.5 + 9 \times 13.5 + 6 \times 25.5 + 3 \times 34.0 + 2 \times 42.5 = 479.$$

We also have $n_1(n_1 + n_2 + 1)/2 = 700$. Since $t_1 = 6, t_2 = 14, \ldots, t_8 = 1$, the denominator of the test statistic is

$$\sqrt{\left(\frac{n_1 n_2}{12}\right)\left[n_1 + n_2 + 1 - \frac{\sum_j t_j(t_j^2 - 1)}{(n_1 + n_2)(n_1 + n_2 - 1)}\right]} = \sqrt{3386.74} = 58.2.$$

We compare the test statistic

$$\frac{|479 - 700| - 1/2}{58.2} = 3.79$$

to 1.96, prompting rejection of H_0 at level $\alpha = 0.05$.

Refer to page 3 of {NONPARExamples.pdf}. In the "1" row of the "Wilcoxon Scores" box we recognize 479 under "Sum of Scores", 700 under "Expected Under H_0", and 58.2 under "Std Dev Under H_0". In row "Z" of the "Wilcoxon Two-Sample Test" box we see -3.79 instead of 3.79 because SAS did not use a continuity correction (-1/2) and hence did not bother taking the absolute value of $479 - 700$ when computing the test statistic.