EXPLORATIONS

Activity 13

Some Nonparametric Tests

There are many nonparametric methods. Many of these use the ranks of the sample data obtained by sorting the data. We will cover four topics to give an idea of what is possible with the TI-83. Topics 56 and 57 on the Sign Test and the Wilcoxon Signed-Ranks Test are the nonparametric counterparts to the two-sample *t* test for dependent data in Topic 44. Topics 58 and 59 on the Mann-Whitney-Wilcoxon Test and the Randomization Test (re-sampling) are the nonparametric counterparts to the two-sample *t* test for independent samples in Topic 44.

Topic 56—Sign Test for Two Dependent Samples

To test the claim that a blood pressure medication reduces the diastolic blood pressure, a random sample of ten people with high blood pressure had their pressures recorded. After a few weeks on the medication, their pressures were recorded again. See the data in the table below.

Subject	1	2	3	4	5	6	7	8	9	10
Before (L_1):	94	87	105	92	102	85	110	95	92	93
After (L_2):	87	88	93	87	92	88	96	87	92	86
$L_1 - L_2 = L_3$	+	−	+	+	+	−	+	+	0	+

Letting p represent the proportion in the population whose blood pressure would be reduced by the medication (+ in the above samples), we want to test the following hypotheses:

$H_0 : p = 0.5$ $H_a : p > 0.5$

If the medication has no effect on the blood pressure reading, we would expect about half to do better after taking the medication and half to do worse. The zero difference does not add any information; therefore, we see that of the nine subjects, there are two negative signs and seven positive signs. We would expect 4.5 of each (the average in the long run). This is a binomial distribution with $N = 9$ and $P = 0.5$.

P-value and Conclusion

We will use [2nd] [DISTR] **A:binomcdf(9** [,] **0.5** [,] **2** [)] [ENTER] for screen 1, or a p-value of **0.0898**, which is the probability of getting two or less negative signs. We could also have found the probability of getting seven or more positives, also shown in screen 1.

(1)

```
binomcdf(9,.5,2
        .08984375
1-binomcdf(9,.5,
6)
        .08984375
```

We have very little evidence to reject the null hypothesis that the medication has no effect. (Not significant at the $\alpha = 0.05$ level.)

If the sample size were doubled to 18 and the number of negatives doubled to 4, we would reach the opposite conclusion. (See screen 2.) With the larger sample size, however, we could also see a larger proportion of negatives (for example, six or more), and again, we would not be able to reject the null hypothesis.

(2)

```
binomcdf(18,.5,4
      .0154418945
binomcdf(18,.5,6
      .1189422607
```

Note: *If the sample sizes become greater than 999, you could use the normal distribution to approximate the binomial (as in Topic 24) because it would only take samples greater than size 20 to meet the criteria needed for this approximation with* $p = 0.5$ *(*$n * p$ *and* $n * (1 - p) \geq 10$*).*

If there is sufficient data, the Sign Test is easy to use and explain. When we used more than the sign of the difference of the blood pressure readings for the more restrictive t test in Topic 44, "Dependent Samples," we had strong evidence (p-value = 0.0047) that the medication reduced the pressure. Not rejecting the null hypothesis does not mean that the null hypothesis is true; we just have insufficient evidence to say it is false.

The next topic is less restrictive than the t test but uses more of the information available in the sample than does the sign test.

Topic 57—Wilcoxon Signed-Ranks Test for Two Dependent Samples

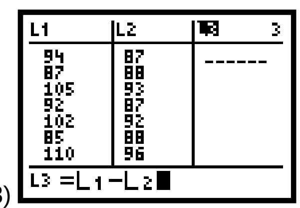
To test the claim that a blood pressure medication reduces the diastolic blood pressure, a random sample of ten people with high blood pressure had their pressures recorded. After a few weeks on the medication their pressure was recorded again. See the data in the table below.

Subject	1	2	3	4	5	6	7	8	9	10
Before (L1):	94	87	105	92	102	85	110	95	92	93
After (L2):	87	88	93	87	92	88	96	87	92	86
L1 - L2 = L3	7	−1	12	5	10	−3	14	8	0	7

Use the following procedure to test the following hypotheses:

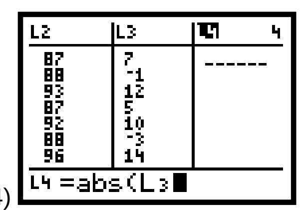
H_0 : The populations of blood pressures are identical.
H_a: The population on medication is shifted to the left.

1. Put the Before readings in **L1** and the After readings in **L2**.



(3)

2. Highlight **L3**, and then enter **L1** [−] **L2** for the bottom line, as shown in screen 3. Press [ENTER] to complete the calculation, as shown in screen 4.

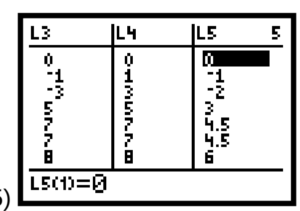


(4)

3. Highlight **L4**, and then press [MATH] **<NUM> 1:abs(L3** for the bottom line, as shown in screen 4. Press [ENTER], and all the values in **L4** will be positive.

4. Press [STAT] **2:SortA(L4** [,] **L3** [ENTER].

   The values in **L4** will be sorted with the smallest value zero coming first. List **L3** will take along the corresponding values (0 comes first even though -3 is the smallest number in **L3**). See the results in screen 5.



(5)

5. Put the ranks of the data in **L4** into **L5** as follows.

   a. Rank **0** as 0.

   b. Rank **1** and **3** as ⁻1 and ⁻2 (negative because they are from negative differences in **L3**).

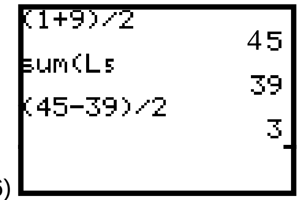
   c. Rank **5** as 3.

   d. Rank the two **7**s by averaging the fourth and fifth ranks (or (4 + 5)/2 = **4.5**).

   e. Rank the other values, which are all different and positive, as 6, 7, 8, and 9.

The sum of the integers from 1 to $n = (1 + n) * n/2$; thus, the sum of the integers from 1 to 9 = (1 + 9) ∗ 9/2 = **45**. T = **sum(L5** = **39** is our test statistic. (See screen 6)

As a check on your work, note that half the difference between the above two values is the absolute value of the sum of the negative ranks, or *abs* (⁻1 + ⁻2) = 3.



(6)

What is the probability of getting a sum of 39 by chance if there are no differences in the two populations?

All possible ranks from 1 to 9, positive or negative, are possible. This is 2^9 = 512 possibilities.

P-Value and Conclusion

The five ways of getting a sum of 39 or more are listed in the table below. Thus, the p-value = 5/512 = 0.009765 = **0.01**.

There is strong evidence that the medication has decreased the blood pressure.

±1	±2	±3	±4	±5	±6	±7	±8	±9	sum
⁻1	⁻2	3	4	5	6	7	8	9	39
1	2	⁻3	4	5	6	7	8	9	39
1	⁻2	3	4	5	6	7	8	9	41
⁻1	2	3	4	5	6	7	8	9	43
1	2	3	4	5	6	7	8	9	45

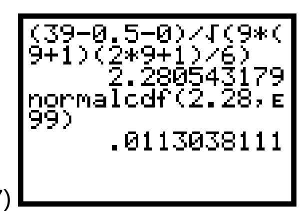
Charts in Text and Normal Approximation

As a sample size increases, it can be a bit tedious to calculate all of the possibilities. Texts will have a chart for small sample sizes (some texts base their chart on the smallest sum of the positives or negatives, ⁻3 in our example). For sizes larger than the charts can handle, the normal approximation below gives good results.

Mean: $\mu_T = 0$ Standard Deviation: $\sigma_T = \sqrt{(n(n+1)(2n+1)/6)}$.

If the two populations are the same, we would expect an equal number of positive and negative ranks of their differences. Thus, on average, the sum of the ranks should be zero.

$z = (T \pm 0.5 - \mu_T)/\sigma_T = (39 - 0.5 - 0)/\sqrt{(9 * 10 * 19/6)}$ = **2.28**, where the -0.5 is a continuity correction. For the right tail, we start at 38.5 to include the class of 39. The area in the right tail of the normal distribution is a p-value of 0.0113 = **0.01** (see Topic 24), very close to the exact value above. (See screen 7.)



(7)

Topic 58– Wilcoxon (Mann-Whitney) Test for Two Independent Samples

Test the claim that teaching Method A results in higher test scores than Method B based on the following scores from random samples of students taught with the two methods.

							mean	Sx	n
Method A (L1)	40	38	39	32	35	38	37	2.966	6
Method B (L2)	34	36	31	37	29		33.4	3.362	5

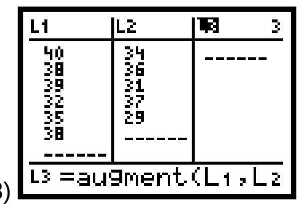
We will test the following hypotheses:

H_0 : The two population distributions are identical.

H_a: Population A is to the right of population B.

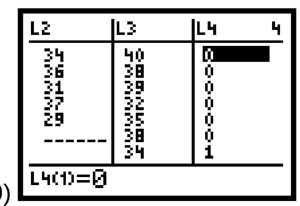
1. Put the Method A scores in **L1** and the Method B scores in **L2**.

(8)



2. Highlight **L3**, and then press [ENTER] **[LIST] <OPS> 9:augment(L1** [,] **L2** for the bottom line, as shown in screen 8. Press [ENTER] and **L1** is pasted above **L2** in **L3**, as shown in screen 9.

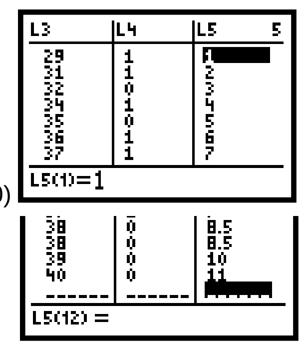
(9)



3. Put zeros next to the six values of the largest sample, Method A, in **L4** and ones next to the five values of Method B. (See screen 9.)

4. Use [STAT] **2:SortA(L3** [,] **L4** [ENTER] to sort the data in **L3** and carry the method identification (0 or 1) in **L4**, as shown in screen 10.

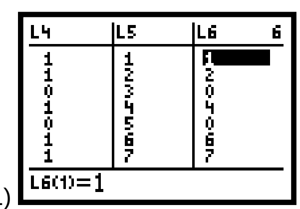
(10)



5. Put the ranks of the values in **L3** in **L5** from 1 for the lowest score of 29, to 11 for the highest score of 40 (screen 10). Because there are two 38s, they both get the mean of the eighth and ninth rank, or 8.5.

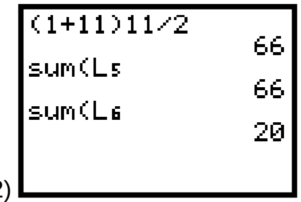
6. Highlight **L6**, and then press **L4** [×] **L5** [ENTER] for the ranks of Method B in **L6** (screen 11). Zeros are in the rows where the ranks of Method A are in L5.

   To check your work, the sum of n ranks is $(1 + n)n/2$ = $(1 + 11)11/2$ = **66**, which should be the sum of **L5**.



(11)

Press [2nd] [LIST] **<MATH> 5:sum(L5** [ENTER] for **66**, as shown in screen 12. The test statistic T = **sum(** L6 = 20 is the sum of the ranks for the smaller sample.



(12)

The average rank value is (the sum of ranks)/n = 66/11 = 6. The smaller sample of five values has an average rank of 20/5 = 4 and a total of 20 instead of the expected 5 ∗ 6 = 30. Is this difference significantly less than what we would expect if the data came from identical populations?

P-value and Conclusion

There are 462 (11 nCr 5 = 462) ways of picking five ranks for Method B and the other six ranks for Method A (see Topic 22). There are 19 possibilities where the sum of the five ranks is 20 or less (see the table below).

P-value = 19/462 = 0.0411.

We reject the null hypothesis and conclude that Method A does significantly better.

sample	rank											
B	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**sum**
1	1	2	3	4	5							15
2	1	2	3	4		6						16
3	1	2	3	4			7					17
4	1	2	3	4				8				18
5	1	2	3	4					9			19
6	1	2	3	4						10		20
7	1	2	3		5	6						17
8	1	2	3		5		7					18
9	1	2	3					8				19
10	1	2	3						9			20
11	1	2	3			6	7					19
12	1	2	3			6		8				20
13	1	2		4	5	6						18
14	1	2		4	5		7					19
15	1	2		4	5			8				20
16	1	2		4		6	7					20
17	1		3	4	5	6						19
18	1		3	4	5		7					20
19		2	3	4	5	6						20

Charts in Texts and Normal Approximation

It is tedious to calculate all the values in the previous table, and it is not necessary because texts covering this topic have charts for your use. If your sample sizes are too large for the table, you can use the following normal approximation.

Let n be the smaller sample size and m the larger.

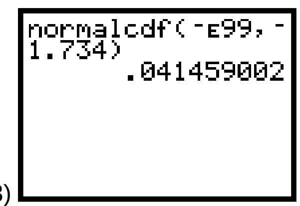
Mean for smaller sample:
$\mu_T = n(n + m + 1)/2 = 5(5 + 6 + 1)/2 = 30$

Standard deviation:
$\sigma_T = \sqrt{(n * m * (n + m + 1)/12)} = \sqrt{(5 * 6 * 12/12)} = 5.477$

$z = (T \pm 0.5 - \mu_T)/\sigma_T = (20 + 0.5 - 30)/5.477 = -1.734$

The 0.5 is the continuity correction as we want the area to the left, including the class with 20 that ends at 20.5.

Find the area in the left tail of the normal distribution, or an approximate p-value = **0.0415** (very close to the exact value of 0.0411 in the previous section), by pressing [2nd] **[DISTR] 2:normalcdf(** [(-)] **E99** [,] [(-)] **1.734** [ENTER]. (See screen 13.)



(13)

Notice that these p-values are less than 0.0493, the value obtained with the more restrictive t test in “Testing Independent Samples” in Topic 44.

Topic 59—Randomization Test for Two Independent Samples

Test the claim that teaching Method A results in higher test scores than Method B based on the following scores from random samples of students taught with the two methods.

							mean	Sx	n
Method A (L1)	40	38	39	32	35	38	37	2.966	6
Method B (L2)	34	36	31	37	29		33.4	3.362	5

Use the following procedure to test the following hypotheses:

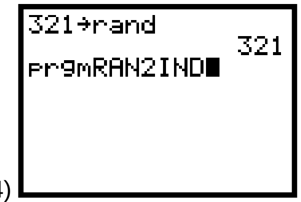
$H_0 : \mu_A - \mu_B = 0$ from identical populations

$H_a: \mu_A - \mu_B > 0$

1. Put the Method A scores in $\mathbf{L_1}$ and the B scores in $\mathbf{L_2}$.

   You will need program **RAN2IND**. This listing is given at the end of this topic.

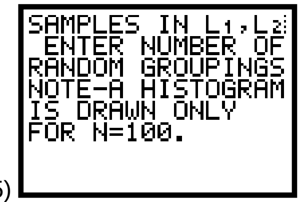
2. Set the random seed as explained in Topic 21 and in the first two lines in screen 14.



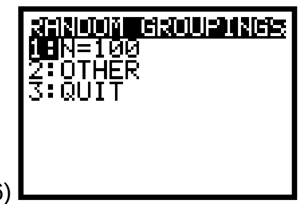
(14)

3. Press [PRGM], and then highlight program **RAND2IND** so **prgmRAN2IND** is pasted to the home screen.

4. Press [ENTER] for screen 15 that reminds you to put the two samples in $\mathbf{L_1}$ and $\mathbf{L_2}$. Press [ENTER] again, and you get the menu screen. (See screen 16.)

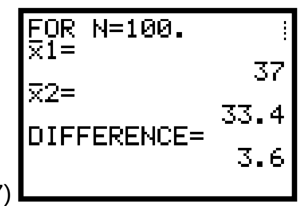


(15)



(16)

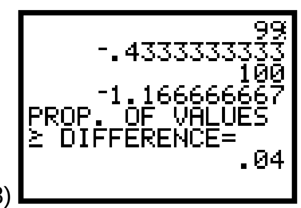
5. Select **1:N=100** for the next screen 17, which tells you the difference between the means of the two samples is **3.6**. Press [ENTER] and the program takes the mean of five randomly selected values from the combined pool of 11 values.



(17)

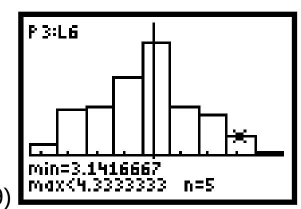
   The 11 values are a representative sample taken from the null hypothesis identical populations. The program will subtract the mean of the five selected values from the mean of the remaining six values and compare it to 3.6. If the difference of the two means is 3.6 or greater, this fact is recorded.The difference of means is stored in $\mathbf{L_6}$.

   This process is repeated 100 times. You see from the results of the 100 repetitions in screen 18 that getting a difference of 3.6 happened four times out of 100 for an estimated p-value of **0.04** (0.08 for a two-tail test). (If you used another seed or a different number of repetitions, your answer could differ.)



(18)

6. Press [ENTER], and the **Histogram** of all the differences is shown, as shown in screen 19. Use [TRACE] to see that values of 3.6 or greater are in the right tail. If values of 3.6 or greater were a more common occurrence, the values would be more toward the center of the distribution, and you would fail to reject the null hypothesis.



(19)

Note that this test is very much like Topic 58, but in Topic 58, you were concerned with all possible totals (or equivalently means) of size 5. Here, you do the same thing, but rather than theoretically figure out the number of possibilities, you just randomly generate them with the actual sample data rather than their ranks. No need to worry about ties.

If you look at the stat list editor or the spreadsheet, you see that the values of the differences have been taken out of their original order in **L6** and placed in descending order in **L4**. (See screens 20 and 21 for the first 12 values.)

The four largest values were **4.3333**, **3.9667**, **3.9667** and **3.6**.

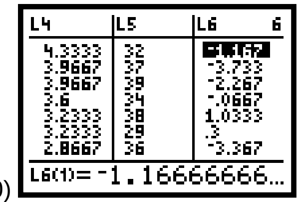
L5 has the last randomly assigned samples. The first five values represent the smaller sample, or Method B {**32**, **37**, **39**, **34**, **38**}. The remaining six values represent Method A {**29**, **36**, **35**, **40**, **31**,**38**}. (See screens 20 and 21.)

If you go down to the bottom of **L6**, or look at screen 18 again, or use **L6** (**100**), as shown in screen 22, you will see the difference of the last means is **-1.167**.

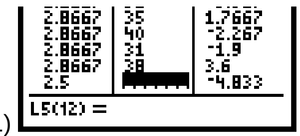
The program can be rerun with different values of *N*, but a **Histogram** will not be automatically plotted.

The program listing is on the next page.

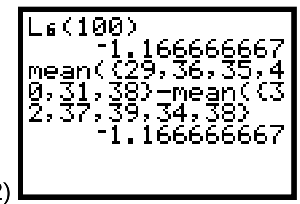
(20)



(21)



(22)



***Note:** It is a coincidence that -1.1667 was also the mean of the first sample.*

Activity 13, Some Nonparametric Tests (cont.)

Program RAN2IND.83P

```
ClrDraw:ClrHome:ClrList L5,L6
Ø→P
Disp "SAMPLES IN L1,L2"
Disp " ENTER NUMBER OF"
Disp "RANDOM GROUPINGS"
Disp "NOTE-A HISTOGRAM"
Disp "IS DRAWN ONLY"
Disp "FOR N=1ØØ.":Pause
Menu("RANDOM
GROUPINGS","N=1ØØ",A,"OTHER",B,"QUIT",
C)
Lbl C:Stop
Lbl A
1ØØ→K:Goto D
Lbl B:Disp "N=":Input K
Lbl D
1-Var Stats L1
x̄→A:Σx→E:n→M
1-Var Stats L2
x̄→B:Σx→F:n→N
abs((A-B)→D
E+F→T:M+N→L
Disp "x̄1=",A
Disp "x̄2=",B
Disp "DIFFERENCE=",D
Pause
L1→L5:M+1→J:For(I,1,N):L2(I)→L5(J):1+J
→J:End
For(I,1,K)
L→C:1→J:Ø→S
If A>B:Then:M→Z:Else:N→Z:End
For(J,1,Z)
int((C*rand+1)→V:S+L5(V)→S
L5(C)→H
L5(V)→L5(C)
H→L5(V)
C-1→C:End
If A>B:Then:S/M-(T-S)/N→U:Else:S/N-(T-
S)/M→U:End
If K<1Ø1:U→L6(I)
If U≥D:1+P→P
Disp I,U
End
Disp "PROP. OF VALUES"
Disp "≥ DIFFERENCE="
Disp P/K:Pause
If K≠1ØØ:Stop
L6→L4
SortD(L4)
FnOff :PlotsOff :PlotsOn
3:Plot3(Histogram,L6,1):ZoomStat
Return
```