Due a week from today!
Different from the HW!

Erick Castillo.
Ryan Agustin
Hovsep Akopyan
Austin Venezia.
2/9/2021.

MATH 440A
SCHILLING

LAB: NONPARAMETRIC ESTIMATION

Using the shift model, our setup and null hypotheses for the two cases we have studied can be expressed as follows:

Paired data: $D_i = Y_i - X_i, i = 1, \ldots, n$ $\quad D_i$'s $\sim F(x-\Delta)$; F symmetric around 0 $\quad H_0$: $\Delta = 0$

Independent samples: $X_1, \ldots, X_{n_1} \sim F(x)$ $\quad Y_1, \ldots, Y_{n_2} \sim G(x) = F(x-\Delta)$ $\quad H_0$: $\Delta = 0$

A logical way to estimate Δ is to see how much the Y's need to be shifted in order to "align them" with the X's. We can consider them aligned if the test statistic then takes its "central value"--its expected value under H_0.

I. Estimating the Shift for Paired Data:

- The *t-test* gives the estimator $\hat{\Delta} = \overline{D}$ (because $t = \dfrac{\overline{D}}{s_D/\sqrt{n}}$; subtracting $\overline{D}$ from all the Y's would make $t = 0$, the central value of the t distribution).
- The *sign test* gives the estimator $\hat{\Delta} = \widetilde{D}$ (because subtracting $\widetilde{D}$ from all the Y's would make $M = n/2$; that is, there would be an equal number of positive and negative D_i's).

Example: Job satisfaction before and after exercise program

Before (X)	After (Y)	Difference (Y−X)	Diff's after shifting Y's down by $\widetilde{D}$:
37	30	-7	-11.5
43	37	-6	-10.5
24	23	-1	-5.5
15	18	3	-1.5
31	35	4	-0.5
16	21	5	0.5
27	44	17	12.5
24	44	20	15.5
29	50	21	16.5
19	59	40	35.5

positive differences: 5

negative differences: 5

$\widetilde{D} = \frac{4+5}{2} = 4.5$

※ • What estimator does the *signed-rank test* produce? Hodges – Lehmann Estimator

We need to shift the Y's by an amount that makes $W^+ = W^-$. To see how to do this, fill in the empty column in the following table and total the entries:

Subject	Before Program	After Program	Differences (D_i's)	#Pairs $i \le j$ with $(D_i+D_j)/2 < 0$	Sign	Signed Rank
1	37	30	-7	6	–	-6
2	43	37	-6	5	–	-5
3	24	23	-1	1	–	-1
4	15	18	3	0	+	2
5	31	35	4	0	+	3
6	16	21	5	0	+	4
7	27	44	17	⋮	+	7
8	24	44	20		+	8
9	29	50	21		+	9
10	19	59	40	0	+	10
				Total: 12		$W^- = 12$

How does the total compare to the value of W^-? EQUAL!

Similarly, you could see that W^+ = #Pairs with $(D_i+D_j)/2 > 0$. Thus to equalize W^+ and W^-, we need to shift the Y's down by the *median* of all the pairs $i \le j$ with $(D_i+D_j)/2$:

$$\hat{\Delta} = \underset{i \le j}{median}\{(D_i + D_j)/2\}.$$

This is called the *Hodges-Lehmann estimator*. To determine the value of the estimator for this example, the following table is quite useful:

Values of $(D_i+D_j)/2$:

	-7	-6	-1	3	4	5	17	20	21	40
-7	-7	-6.5	-4	-2	-1.5	-1	5	6.5	7	16.5
-6		-6	-3.5	-1.5	-1	-0.5	5.5	7	7.5	17
-1			-1	1	1.5	2	8	9.5	10	19.5
3				3	3.5	4	10	11.5	12	21.5
4		27			4	4.5	10.5	12	12.5	22
5						5	11	12.5	13	22.5
17		#'s < $\hat{\Delta}$					17	18.5	19	28.5
20								20	20.5	30
21									21	30.5
40										40

27 #'s > $\hat{\Delta}$

$\hat{\Delta} = 8$

Once you think you have found $\hat{\Delta}$, circle the values in the table that are below $\hat{\Delta}$ and also circle the values that are above $\hat{\Delta}$, making sure that there are an equal number of values in both sets, and indicate that number.

I. Estimating the Shift for Independent Samples:

6.4 5.63

- The *t-test* gives the estimator $\hat{\Delta} = \bar{y} - \bar{x}$ (because subtracting $\bar{y} - \bar{x}$ from all the Y's would make $t = 0$, the central value of the t distribution).

- The Mann-Whitney test shifts the Y's by an amount that makes $U_{X<Y} = U_{Y<X}$. The resulting estimator is called the *Hodges-Lehmann* estimator. (Equivalently, we could shift to make the rank sum of the Y's equal to the rank sum of the X's.)

***Example*: Memory training**

An experiment was performed to assess the effect of a special training method on memory retention. Half of the subjects were given the training method, while the other half served as controls. Each subject was then instructed to listen to a spoken string of 15 digits and then repeat as many as possible starting at the beginning of the string. The results of ten trials per subject are shown below:

Average Number of Digits Correctly Recalled

Control responses: 4.3, 4.9, 5.1, 5.7, 6.2, 7.6 Treatment responses: 4.4, 5.3, 6.0, 6.8, 7.6, 8.3

Note that the Y's tend to be larger than the X's, indicating that the training method did improve memory retention.

If the Y's are shifted down by an amount $\hat{\Delta}$ that makes $U_{X<Y} = U_{Y<X}$, what will the common value of those two U statistics be? 18 (Note: $U_{X<Y}$ is denoted by U_Y in the textbook—see p. 441.)

Plot the data on the two number lines below, then use the "method of sliding" to determine $\hat{\Delta}$. Clip the bottom strip in a position that aligns the X and Y data in the sense that $U_{X<Y} = U_{Y<X}$, and show clearly how you computed $\hat{\Delta}$ from the offset of the two scales:

$\hat{\Delta} = 0.71$

$U_{x<y} = 1+2+2+3+4+6$
$= 18$

$U_{y<x} = 1+3+4+5+5$
$= 18$