

Homework 3 (Two-Sample Inference)

1. (Mann-Whitney Test) In a genetic inheritance study discussed by Margolin [1988], samples of individuals from several ethnic groups were taken. Blood samples were collected from each individual and several variables measured. We shall compare the groups labeled “Native American” and “Caucasian” with respect to the variable MSCE (mean sister chromatid exchange).

The data is as follows:

Native American:	8.50 9.48 8.65 8.16 8.83 7.76 8.63
Caucasian:	8.27 8.20 8.25 8.14 9.00 8.10 7.20 8.32 7.70

Apply the Mann-Whitney or Wilcoxon rank-sum test to see if there is significant difference in MSCE between Native Americans and Caucasians. Obtain the p-value using the large-sample approximation.

Solutions: Let $F_1(x)$ and $F_2(x)$ denote the CDFs of MSCE for the Native American and Caucasian groups, respectively. The hypotheses can be stated as follows:

$$H_0 : F_1(x) = F_2(x) \text{ vs. } H_a : F_1(x) \neq F_2(x)$$

Using the following worksheet, it can be found that $n_1 = 7$, $n_2 = 9$, $T_1 = 75$ and $T_2 = 61$.

Native.American	8.5	9.48	8.65	8.16	8.83	7.76	8.63		
Rank	11	16	13	6	14	3	12		
Caucasian	8.27	8.2	8.25	8.14	9	8.1	7.2	8.32	7.7
Rank	9	7	8	5	15	4	1	10	2

Since there are no tied values in the data, the p-value can be obtained as

$$p\text{-value} = 2 \min \left[\Pr \{ Z \geq z_{11} \}, \Pr \{ Z \leq z_{12} \} \right] = 2 \min [0.0562, 0.9548] = 0.1124,$$

$$\text{where } z_{11} = \frac{T_1 - n_1(n+1)/2 - 1/2}{\sqrt{\frac{n_1 n_2 (n+1)}{12}}} = \frac{75 - 7 \times (16+1)/2 - 1/2}{\sqrt{\frac{7 \times 9 \times (16+1)}{12}}} = 1.587768;$$

$$z_{12} = \frac{T_1 - n_1(n+1)/2 + 1/2}{\sqrt{\frac{n_1 n_2 (n+1)}{12}}} = \frac{75 - 7 \times (16+1)/2 + 1/2}{\sqrt{\frac{7 \times 9 \times (16+1)}{12}}} = 1.693620; \text{ And } Z \text{ is a random}$$

variable following standard normal distribution.

Since the p-value is greater than $\alpha = 0.05$, we cannot reject the null hypothesis.

Some R Codes:

```

Native.American <- scan()
8.50 9.48 8.65 8.16 8.83 7.76 8.63

Caucasian <- scan()
8.27 8.20 8.25 8.14 9.00 8.10 7.20 8.32 7.70

n1 <- length(Native.American);
n2 <- length(Caucasian)
n <- n1 + n2
R <- rank(c(Native.American, Caucasian))
rbind(Native.American, R[1:n1])
T1 <- sum(R[1:n1]); T1
rbind(Caucasian, R[(n1+1):n])
T2 <- sum(R[(n1+1):n]); T2

z11 <- (T1 - n1*(n+1)/2 - 1/2)/sqrt(n1*n2*(n+1)/12); z11
z12 <- (T1 - n1*(n+1)/2 + 1/2)/sqrt(n1*n2*(n+1)/12); z12
cbind(1-pnorm(z11), pnorm(z12))

wilcox.test(Native.American, Caucasian, alternative = "t")
# "t" for "two-sided"

```

2. (*Mann-Whitney Test*) Consider an artificial example:

Group A: $X_1, \dots, X_{n_1} \sim F_A$ Group B: $Y_1, \dots, Y_{n_2} \sim F_B$

Null Hypothesis $H_0 : F_A = F_B$

Artificial Example: gpA: 1.3 3.4 ($n_1 = 2$)
gpB: 4.9 10.3 3.3 ($n_2 = 3$)

Order all observations in the combined sample & assign ranks: (gp A data underlined)

Order	<u>1.3</u>	3.3	<u>3.4</u>	4.9	10.3
Assign ranks	<u>1</u>	2	<u>3</u>	4	5

We consider the one-sided hypothesis:

$$H_0 : F_A(x) = F_B(x) \text{ vs. } H_a : F_A(x) > F_B(x)$$

In other words, under the alternative hypothesis H_a , Group A tends to yield smaller values than Group B.

Solutions:

Since $n_1 = 2 < n_2 = 3$, the test statistic $T_1 = 1 + 3 = 4$ is the sum of ranks for data in Group A.

Under the null hypothesis H_0 , each $n_1 = 2$ -subset of the ranks $\{1, 2, 3, 4, 5\}$ is equally likely

to occur as the ranks in T_1 . There are a total of $\binom{n}{n_1} = \binom{5}{2} = 10$ possible different ways for the

two ranks in Group A, as listed below:

	sum =	R_1
	{1, 2}	3
	{1, 3}	4
	{1, 4}	5
	{1, 5}	6
Possible ranks for X_1, X_2 :	{2, 3}	5
	{2, 4}	6
	{2, 5}	7
	{3, 4}	7
	{3, 5}	8
	{4, 5}	9

Hence the distribution of R_1 under H_0 is given by

$r =$	3	4	5	6	7	8	9
$P_{H_0}(R_1 = r)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{1}{10}$

In our toy example, $R_1 = 4$, the one-sided P-value

$$P = P_{H_0}(R_1 \leq 4) = P(\text{seeing a value as small or smaller than observed}) = \frac{1}{5}$$

3. (*Mann-Whitney Test: With Many Ties*) The carapace lengths (in mm) of crayfish were recorded for samples from two sections of a stream in Kansas.

Stream I :	10	11	13	13				$(n_1 = 4)$
Rank: 1	2.5	5	5					
Stream II:	11	13	15	16	19	19		$(n_2 = 6)$
Rank: 2.5	5	7	8	9.5	9.5			

Perform the Mann-Whitney test to see if crayfish from these two different sections tend to have different carapace lengths.

Solution: Let $F_1(x)$ and $F_2(x)$ denote the CDFs of carapace lengths for these two streams, respectively. The hypotheses can be stated as follows:

$$H_0 : F_1(x) = F_2(x) \text{ vs. } H_a : F_1(x) \neq F_2(x)$$

It can be found that $T_1 = 13.5$ and $T_2 = 41.5$.

Since there are many tied values, the test statistic (using the large sample approximation) is

$$\begin{aligned} p\text{-value} &= 2 \min \left[\Pr \{Z \geq z_{11}\}, \Pr \{Z \leq z_{12}\} \right] \\ &= 2 \min \left[\Pr \{Z \geq -2.0634\}, \Pr \{Z \leq -1.7276\} \right] \\ &= 2 \times \Pr \{Z \leq -1.7276\} = 2 \times 0.0411 = 0.0822. \end{aligned}$$

$$\text{where } z_{11} = \frac{T_1 - n_1(n+1)/2 - 1/2}{\text{s.d.}} = \frac{13.5 - 4 \times (10+1)/2 - 1/2}{4.604} = -2.0634;$$

$$z_{12} = \frac{T_1 - n_1(n+1)/2 + 1/2}{\text{s.d.}} = \frac{13.5 - 4 \times (10+1)/2 + 1/2}{4.604} = -1.7376; \text{ the standard error is}$$

$$\text{s.d.} = \sqrt{\frac{n_1 n_2}{n(n-1)} \sum_{i,j} R_{ij}^2 - \frac{n_1 n_2 (n+1)^2}{4(n-1)}} = \sqrt{\frac{4 \times 6}{10 \times (10-1)} \times 382 - \frac{4 \times 6 \times (10+1)^2}{4 \times (10-1)}} = 4.604; \text{ and}$$

$$\sum_{i,j} R_{ij}^2 = 1^2 + 2.5^2 + 5^2 + 5^2 + 2.5^2 + 5^2 + 7^2 + 8^2 + 9.5^2 + 9.5^2 = 382.$$

4. (*Permutation Test*) A new treatment for post surgical recovery is compared to a standard treatment by observing the recovery times (in days) of patients on each treatment. Of the $N = 7$ subjects available, $n_1 = 4$ are randomly assigned to receive the new treatment and $N - n_1 = 3$ receive the standard treatment. We wish to test the null hypothesis

New Treatment: 19, 22, 25, 26

Standard Treatment: 23, 33, 40

H_0 : There is no difference between the treatments

H_1 : The new treatment decreases recovery times. ← This is a ***one-sided*** hypothesis.

Note that if the null hypothesis is true and there is no difference between the treatments, then the recovery time for each subject will be the same *regardless of which treatment is received*.

- (i) Let (μ_1, μ_2) denote the (population) mean recovery time. Construct a 95% parametric CI for the mean difference.

Solutions: It can be found $\bar{x}_1 = 23$, $\bar{x}_2 = 32$, $s_1 = 1.162$, and $s_2 = 8.544$. Then you can find $s_p = 5.933$. The 95% t CI can be found to be $(-20.648263, 2.648263)$. More details are explained in class notes.

- (ii) The permutation test based on difference in sample means.

Solutions: For observed data, it can be found the difference in sample means is -9.00 . Since there is no tied value, the total number of different permuted samples is

$$\binom{n}{n_1} = \binom{7}{4} = 35.$$

See Table 1 for detailed enumeration of all permuted samples. And the exact p-value = $\frac{\text{number of differences} \leq -9.00}{35} = \frac{3}{35} = 0.0857$.

Thus we cannot reject the null at the significance level of 0.05.

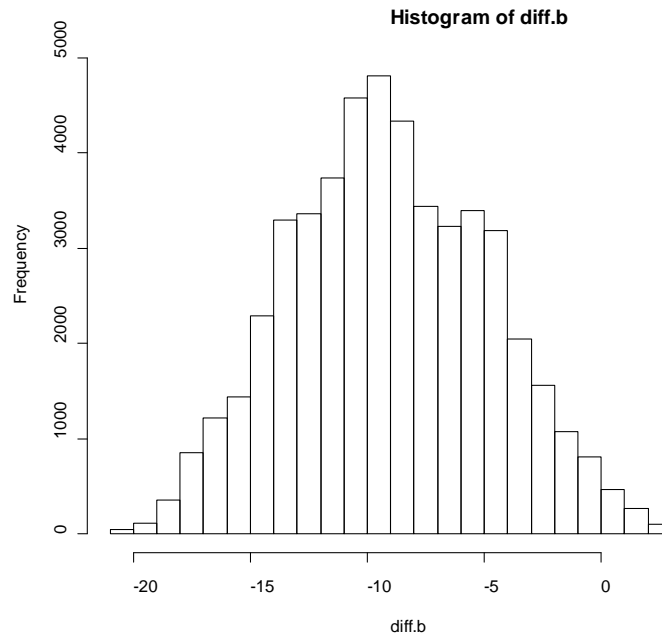
TABLE 1
*All possible randomizations of seven recovery times (days) to two treatment groups of sizes
 $n = 4$ and $m = 3$*

No.	Randomization						Difference in means	Sum of new	Sum of standard	Difference in medians	
	New treatment				Standard treatment						
▲ 1	19	22	25	26	23	33	40	-9.00	92	96	-9.5
2	22	23	25	26	19	33	40	-6.67	96	92	-9.0
3	22	33	25	26	19	23	40	-0.83	106	82	2.5
4	22	25	26	40	19	23	33	3.25	113	75	2.5
5	19	23	25	26	22	33	40	-8.42	93	95	-9.0
6	19	25	26	33	22	23	40	-2.58	103	85	2.5
7	19	25	26	40	22	23	33	1.50	110	78	2.5
8	19	22	23	26	25	33	40	-10.17	90	98	-10.5
9	19	22	26	33	23	25	40	-4.33	100	88	-1.0
10	19	22	26	40	23	25	33	-0.25	107	81	-1.0
11	19	22	23	25	26	33	40	-10.75	89	99	-10.5
12	19	22	25	33	23	26	40	-4.92	99	89	-2.5
13	19	22	25	40	23	26	33	-0.83	106	82	-2.5
14	23	25	26	33	19	22	40	-0.25	107	81	3.5
15	22	23	26	33	19	25	40	-2.00	104	84	-0.5
16	22	23	25	33	19	26	40	-2.58	103	85	-2.0
17	19	23	26	33	22	25	40	-3.75	101	87	-0.5
18	19	23	25	33	22	26	40	-4.33	100	88	-2.0
19	19	22	23	33	25	26	40	-6.08	97	91	-3.5
20	23	25	26	40	19	22	33	3.83	114	74	3.5
21	22	23	26	40	19	25	33	2.08	111	77	-0.5
22	22	23	25	40	19	26	33	1.50	110	78	-2.0
23	19	23	26	40	22	25	33	0.33	108	80	-0.5
24	19	23	25	40	22	26	33	-0.25	107	81	-2.0
25	19	22	23	40	25	26	33	-2.00	104	84	-3.5
26	25	26	33	40	19	22	23	9.67	124	64	7.5
27	22	26	33	40	19	23	25	7.92	121	67	6.5
28	22	25	33	40	19	23	26	7.33	120	68	6.0
29	19	26	33	40	22	23	25	6.17	118	70	6.5
30	19	25	33	40	22	23	26	5.58	117	71	6.0
31	19	22	33	40	23	25	26	3.83	114	74	2.5
32	23	26	33	40	19	22	25	8.50	122	66	7.5
33	23	25	33	40	19	22	26	7.92	121	67	7.0
34	22	23	33	40	19	25	26	6.17	118	70	3.0
35	19	23	33	40	22	25	26	4.42	115	73	3.0

(iii) The permutation test based on difference in sample medians.

Solutions: When based on difference in median, the exact p-value is the same as above, 0.0857. Again, see Table 1 for details.

- (iv) Next, we want to obtain the nonparametric bootstrap 95% confidence interval for the difference in means. Show one example of a bootstrap sample and the relevant calculations involved. Also, shown below are the histogram and the percentiles for the mean differences obtained from $B=5000$ bootstrap samples. What is the 95% bootstrap CI then?



1%	2.5%	5%	10%	20%	33%
-17.917	-17.000	-15.667	-14.667	-12.833	-10.833
50%	66%	75%	95%	97.5%	99%
-9.000	-7.250	-5.917	-1.750	-0.250	1.000

Solution: Here is an example of a bootstrap sample taken for this purpose:

```

trt      boot.smpl.1
-----
"new"    "19"
"new"    "26"
"new"    "22"
"new"    "19"
"standard" "33"
"standard" "23"
"standard" "40"

```

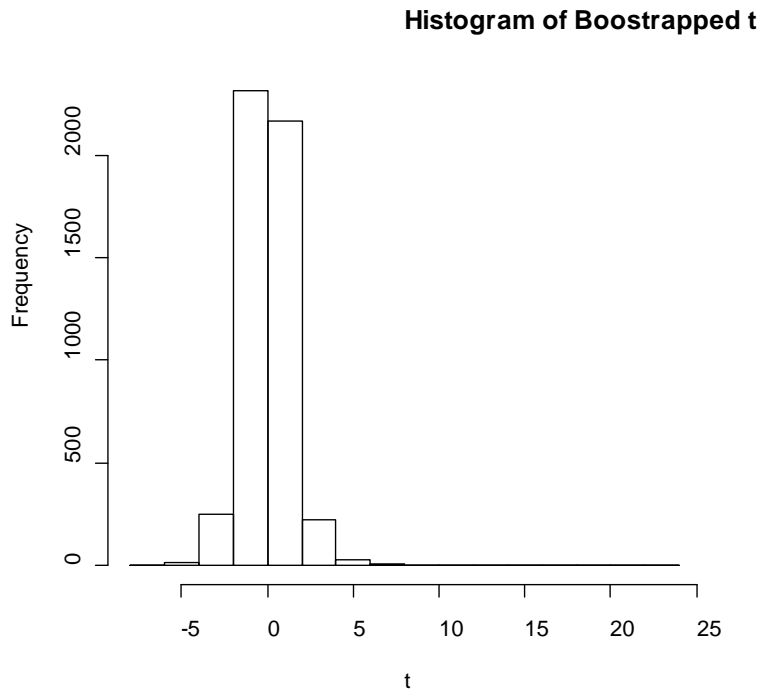
```

  b mean1.b mean2.b diff.b
  1  21.5      32  -10.5

```

The 95% bootstrap CI is (-17.00, -0.250).

- (v) Yet another bootstrap approach, we want to obtain the parametric bootstrap 95% confidence interval for the difference in means. Show one example of a bootstrap sample and the relevant calculations involved. Also, shown below are the histogram and the percentiles for t statistic obtained from B=5000 bootstrap samples. What is the 95% bootstrap CI then?



2.5%	5%	50%	95%	97.5%
-2.49916356	-2.04075554	-0.02312493	2.00614507	2.53552023

Solution: Again, $\bar{x}_1 = 23$, $\bar{x}_2 = 32$, $s_1 = 1.162$, and $s_2 = 8.544$. Then you can find $s_p = 5.933$.

First obtain the observed errors by $\varepsilon_{ij} = x_{ij} - \bar{x}_i$, for $i = 1, 2$ and $j = 1, \dots, n_i$.

New Treatment: -4, -1, 2, 3
 Standard Treatment: -9, 1, 8

Here is one example of bootstrap sampling of ε_{ij} and calculating t value.

Bootstrap Sample:
 New -1 -9 -9 1 Standard -9 8 -4

Mean.new	mean.std	s1.new	s2.std	sp	t
-4.5	-1.666667	5.259911	8.736895	6.8654	-0.54035

Using the 2.5th and 97.5th percentiles from the bootstrap distribution of t , it can be found that the 95% bootstrap CI is (2.3246, -20.32464).

5. (*Comparing Variation between Two Samples*) A blood bank kept a record of the rate of heartbeats for several blood donors, as listed below. Is the variation among the men significantly different from (or greater than) the variation among women?

Men: 68.2, 76.2, 82.4, 79.2, 65.7, 86.0
 Women: 74.1, 76.5, 72.0, 73.6, 75.9

It can be found that $s_1^2 = 63.528$ and $s_2^2 = 3.287$

- (i) Perform a parametric F test to see if there is significant difference between two samples.

Solution: It can be found that the F test statistic is $19.33 > F_{.95}^{(5,4)} = 6.256$

```
s1.sq  s2.sq  F.stat  F.95th  pvalue.F
63.53767 3.287 19.32999 6.256057 0.006637335
```

Also the p-value can be found as 0.00664 in R.

- (ii) Apply the Siegel-Tukey test to see if man has large variation. Provide the asymptotic p-value.

Solution: It can be found that the sorted sample in combination is

```
65.7 68.2 72.0 73.6 74.1 75.9 76.2 76.5 79.2 82.4 86.0
1    4    5    8    9   11   10    7    6    3    2
```

Men: 68.2, 76.2, 82.4, 79.2, 65.7, 86.0
 Women: 74.1, 76.5, 72.0, 73.6, 75.9

The rank sum for women (framed) is $T_2 = 5 + 8 + 9 + 11 + 7 = 40$ and hence

$$T_1 = (11 \times 12) / 2 - 40 = 26$$

The asymptotic p-value can be found to be $\Pr\{T_0 \leq 26\} = \Pr\{Z \leq -1.734\} = 0.04142$.

R Codes for Calculation the P-value:

```
# Siegel-Tukey Test
T1 <- 26; T2 <- 40
n1 <- 6; n2 <- 5;
n <- n1+n2
Z1 <- (T1 - n1*(n+1)/2 - .5)/sqrt(n1*n2*(n+1)/12)
Z2 <- (T1 - n1*(n+1)/2 + .5)/sqrt(n1*n2*(n+1)/12)
twosided.pvalue <- 2*min(pnorm(Z1, lower.tail =F), pnorm(Z2));
pvalue <- pnorm(Z2); # THE DESIRED P-VALUE
p.value
```

(iii) The following table provides a worksheet for performing the permutation test. In particular, the third and fourth columns give $e_{ij} = (x_{ij} - \bar{x}_i)^2$ and $e'_{ij} = e_{ij} \cdot \frac{n_i}{n_i - 1}$. Column 5 and 6 show two permuted samples. Would the results from these two samples be counted towards the calculation of the exact p-value?

trt	x	e	e.prime	sample1	sample2
0	68.2	6.534e+01	7.841e+01	5.408e+00	7.321e+00
0	76.2	6.944e-03	8.333e-03	1.344e+02	4.490e+01
0	82.4	3.741e+01	4.490e+01	8.405e-01	7.841e+01
0	79.2	8.507e+00	1.021e+01	1.133e+02	1.021e+01
0	65.7	1.120e+02	1.344e+02	7.321e+00	2.738e+00
0	86.0	9.441e+01	1.133e+02	1.280e-01	8.405e-01
1	74.1	1.024e-01	1.280e-01	2.738e+00	8.333e-03
1	76.5	4.326e+00	5.408e+00	8.333e-03	5.408e+00
1	72.0	5.856e+00	7.321e+00	7.841e+01	1.280e-01
1	73.6	6.724e-01	8.405e-01	4.490e+01	1.344e+02
1	75.9	2.190e+00	2.738e+00	1.021e+01	1.133e+02

It can be found that

	\bar{e}'_1	\bar{e}'_2	Difference
For observed e'_{ij} :	63.54	3.287	60.25

Find the exact p-value, given that the number of mean differences in permuted e'_{ij} , in absolute value, greater than or equal to 60.25 is 11.

Solution: It can be found that

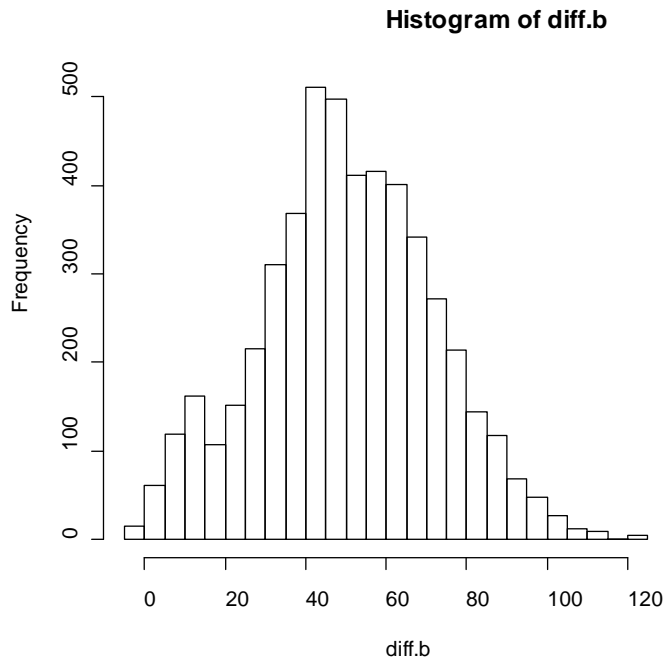
	\bar{e}'_1	\bar{e}'_2	Difference
Sample 1:	43.57	27.25	16.32 < 60.25, will NOT be counted
Sample 2:	24.07	50.65	-26.58 < 60.25, NOT

Note that the exact p-value for the two-sided test is calculated as (number of differences greater than or equal to 60.25 in absolute value) / (the total number of possible different permuted samples).

The total number of possible different permuted samples = $\binom{11}{5} = 462$ and. Thus the exact p-value is $11/462 = 0.02381$.

(iv) 5000 bootstrap samples are taken to compute the difference in variances. The histogram and percentiles of $s_1^2 - s_2^2$ are given below. What is the 95% bootstrap CI for $(\sigma_1^2 - \sigma_2^2)$?

2.5% 5% 50% 95% 97.5%
 7.325 12.020 49.821 86.151 92.180



Solution: The 95% bootstrap CI for difference in variance is (7.325, 92.18).

6. (*Smirnov Two-Sample Test*) We consider the following two-sample data:

\underline{X}	\underline{Y}
17.62	15.2
16.48	21.7
18.96	25.9
18.77	17.5
	26.8
	18.2
	19.1

Solution:

(a). Applying the Smirnov Two Sample test and following the steps outlined in class, it can be found that for the two sided test:

$D = 0.5714$
 $P\text{-value} = .3091$ (asymptotic)

Thus we can not reject the null hypothesis that the two populations from which our samples were drawn have the same distribution function.

(b) Applying the *Cramer-von Mises* test, it can be found that the test statistic is 0.185, which is less than the 95th percentile 0.461. Thus we cannot reject the null either.

7. (*Wilcoxon Signed Rank Sum Test*) Suppose that 16 students in an introductory statistics course are presented with a number of questions (of the sort you encountered in Chapters 5 and 6) concerning basic probabilities. In each instance, the question takes the form "What is the probability of such-and-such?" However, the students are not allowed to perform calculations. Their answers must be immediate, based only on their raw intuitions. They are instructed to frame each answer in terms of a zero to 100 percent rating scale, with 0% corresponding to $P=0.0$, 27% corresponding to $P=.27$, and so forth. They are also told that they can give non-integer answers if they wish to make really fine-grained distinctions; for example, 49.0635...%. (As it turns out, none do.)

The instructor of the course is particularly interested in student's responses to two of the questions, which we will designate as question A and question B. He reasons that if students have developed a good, solid understanding of the basic concepts, they will tend to give higher probability ratings for question A than for question B; whereas, if they were sleeping through that portion of the course, their answers will be mere shots in the dark and there will be no overall tendency one way or the other. The instructor's hypothesis is of course directional: he expects his students have mastered the concepts well enough to sense, if only intuitively, that the event described in question A has the higher probability. The following table shows the probability ratings of the 16 subjects for each of the two questions.

Subj.	X_A	X_B	$X_A - X_B$
1	78	78	0
2	24	24	0
3	64	62	+2
4	45	48	-3
5	64	68	-4
6	52	56	-4
7	30	25	+5
8	50	44	+6
9	64	56	+8
10	50	40	+10
11	78	68	+10
12	22	36	-14
13	84	68	+16
14	40	20	+20
15	90	58	+32
16	72	32	+40
mean difference =			+7.75

The observed results are consistent with the hypothesis. The probability ratings do on average end up higher for question A than for question B. Now we want to determine whether the degree

of the observed difference reflects anything more than some lucky guessing. (Source: <http://faculty.vassar.edu/lowry/ch12a.html>)

Solution: Applying the Wilcoxon Signed Rank-Sum test, we have the following work sheet. Since we have *many ties*, it can be found the test statistic is

$$Z = \frac{\sum_{i=1}^{n'} R_i}{\sqrt{\sum_{i=1}^{n'} R_i^2}} = \frac{T_+ - T_-}{\sqrt{\sum_{i=1}^{n'} R_i^2}} = \frac{67}{\sqrt{1014}}.$$

And, with upper-sided hypothesis, the p-value can be given as

$$\begin{aligned} &= \Pr \left\{ Z > \frac{\sum_{i=1}^{n'} R_i - 0.5}{\sqrt{\sum_{i=1}^{n'} R_i^2}} \right\} = \Pr \left\{ z > \frac{67 - 0.5}{\sqrt{1014}} \right\} \\ &= \Pr \left\{ z > \frac{67 - 0.5}{\sqrt{1014}} = 2.088 \right\} = 0.0184 \end{aligned}$$

Thus we reject the null at sig level 0.05.

Note: An alternative continuity correction for Wilcoxon signed rank sum test is to apply 1

instead of 0.5, i.e., $\Pr \left\{ Z > \left(\sum_{i=1}^{n'} R_i - 1 \right) / \sqrt{\sum_{i=1}^{n'} R_i^2} \right\}$.

Worksheet Used for Wilcoxon Signed Rank Sum Test

1	2	3	4	5	6	7
Subj.	X_A	X_B	original $X_A - X_B$	absolute $X_A - X_B$	rank of absolute $X_A - X_B$	signed rank
1	78	78	0	0	---	---
2	24	24	0	0	---	---
3	64	62	+2	2	1	+1
4	45	48	-3	3	2	-2
5	64	68	-4	4	3.5	-3.5
6	52	56	-4	4	3.5	-3.5
7	30	25	+5	5	5	+5
8	50	44	+6	6	6	+6
9	64	56	+8	8	7	+7
10	50	40	+10	10	8.5	+8.5
11	78	68	+10	10	8.5	+8.5
12	22	36	-14	14	10	-10
13	84	68	+16	16	11	+11
14	40	20	+20	20	12	+12
15	90	58	+32	32	13	+13

16	72	32	+40	40	14	+14
$T_+ = 86.0$ and $T_- = 19.0$ Sum: $86 - 19 = 67$ $n = 16$ and $n' = 14$						