Nonparametric Dispersion and Equality Tests

Nathaniel E. Helwig

Assistant Professor of Psychology and Statistics
University of Minnesota (Twin Cities)

Updated 04-Jan-2017

Copyright

Outline of Notes

1) Dispersion (Ansari-Bradley):

- Overview
- Procedure
- Example 1

2) Dispersion/Location (Lepage):

- Overview
- Procedure
- Example 2

3) Equality (Kolmogorov-Smirnov):

- Overview
- Procedure
- Example 3

Dispersion Test (Ansari-Bradley)

Problem(s) of Interest

Like two-sample location problem, we have $N = m + n$ observations

- $X_1, \ldots, X_m$ are iid random sample from population 1
- $Y_1, \ldots, Y_n$ are iid random sample from population 2

We want to make inferences about difference in distributions

- Let F_1 and F_2 denote distributions of populations 1 and 2
- Null hypothesis is same distribution ($F_1(z) = F_2(z)$ for all z)

Using the location-scale parameter model, we have

- $F_1(z) = G([z - \theta_1]/\eta_1)$ and $F_2(z) = G([z - \theta_2]/\eta_2)$
- θ_j and η_j are median and scale parameters for population j

Assumptions

Within sample independence assumption

- $X_1, \ldots, X_m$ are iid random sample from population 1
- $Y_1, \ldots, Y_n$ are iid random sample from population 2

Between sample independence assumption

- Samples $\{X_i\}_{i=1}^m$ and $\{Y_i\}_{i=1}^n$ are mutually independent

Continuity assumption: both F_1 and F_2 are continuous distributions

Location assumption: $\theta_1 = \theta_2$ or θ_1 and θ_2 are known

Parameter of Interest and Hypothesis

Parameter of interest is the ratio of the variances:

$$\gamma^2 = \frac{V(X)}{V(Y)}$$

so that $\gamma^2 = 1$ whenever $V(X) = V(Y)$.

The null hypothesis about γ^2 is

$$H_0 : \gamma^2 = 1$$

and we could have one of three alternative hypotheses:

- One-Sided Upper-Tail: $H_1 : \gamma^2 > 1$
- One-Sided Lower-Tail: $H_1 : \gamma^2 < 1$
- Two-Sided: $H_1 : \gamma^2 \neq 1$

Test Statistic

Let $\{Z_{(k)}\}_{k=1}^{N}$ denote the order statistics for the combined sample, and assign rank scores

$$R_k^* = \begin{cases} 1, 2, 3, \ldots, \frac{N}{2}, \frac{N}{2}, \ldots, 3, 2, 1 & \text{if } N \text{ is even} \\ 1, 2, 3, \ldots, \frac{N-1}{2}, \frac{N+1}{2}, \frac{N-1}{2}, \ldots, 3, 2, 1 & \text{if } N \text{ is odd} \end{cases}$$

to the combined sample $\{Z_{(k)}\}_{k=1}^{N}$.

The Ansari-Bradley test statistic C is defined as

$$C = \sum_{j=1}^{n} R_j$$

where R_j is the assigned rank score of Y_j for $j = 1, \ldots, n$

Distribution of Test Statistic under H_0

Under H_0 all $\binom{N}{n}$ arrangements of Y-ranks occur with equal probability

- Given (N, n), calculate C for all $\binom{N}{n}$ possible outcomes
- Each outcome has probability $1/\binom{N}{n}$ under H_0

Example null distribution with $m = 3$ and $n = 2$:

Y-ranks	C	Probability under H_0
1,2	3	1/10
1,3	4	1/10
1,4	3	1/10
1,5	2	1/10
2,3	5	1/10
2,4	4	1/10
2,5	3	1/10
3,4	5	1/10
3,5	4	1/10
4,5	3	1/10

Note: there are $\binom{5}{2} = 10$ possibilities

Hypothesis Testing

One-Sided Upper Tail Test:

- $H_0 : \gamma^2 = 1$ versus $H_1 : \gamma^2 > 1$
- Reject H_0 if $C \geq c_\alpha$ where $P(C > c_\alpha) = \alpha$

One-Sided Lower Tail Test:

- $H_0 : \gamma^2 = 1$ versus $H_1 : \gamma^2 < 1$
- Reject H_0 if $C \leq [c_{1-\alpha} - 1]$

Two-Sided Test:

- $H_0 : \gamma^2 = 1$ versus $H_1 : \gamma^2 \neq 1$
- Reject H_0 if $C \geq c_{\alpha/2}$ or $C \leq [c_{1-\alpha/2} - 1]$

Large Sample Approximation

Under H_0, the expected value and variance of C are

- if N is even: $E(C) = \frac{n(N+2)}{4}$ and $V(C) = \frac{mn(N+2)(N-2)}{48(N-1)}$
- if N is odd: $E(C) = \frac{n(N+1)^2}{4N}$ and $V(C) = \frac{mn(N+1)(3+N^2)}{48N^2}$

We can create a standardized test statistic C^* of the form

$$C^* = \frac{C - E(C)}{\sqrt{V(C)}}$$

which asymptotically follows a $\mathrm{N}(0, 1)$ distribution.

Derivation of Large Sample Approximation

Note that we have $C = \sum_{j=1}^{n} R_j$, which implies that

- C/n is the average of the (combined) Y rank scores
- C/n has same distribution as sample mean of size n drawn without replacement from finite population
  $S = \{1, 2, 3, \ldots, \frac{N}{2}, \frac{N}{2}, \ldots, 3, 2, 1\}$ if N is even
  $S = \{1, 2, 3, \ldots, \frac{N-1}{2}, \frac{N+1}{2}, \frac{N-1}{2}, \ldots, 3, 2, 1\}$ if N is odd

Using some basic results of finite population theory, we have

- $E(C/n) = \mu$, where $\mu = \frac{1}{N}\sum_{k=1}^{N} S_k = \begin{cases} \frac{N+2}{4} & \text{if } N \text{ is even} \\ \frac{(N+1)^2}{4N} & \text{if } N \text{ is odd} \end{cases}$
- $V(C/n) = \sigma^2 \frac{N-n}{n(N-1)}$, where
  $\sigma^2 = (\frac{1}{N}\sum_{i=1}^{N} S_k^2) - \mu^2 = \begin{cases} \frac{(N+2)(N-2)}{48} & \text{if } N \text{ is even} \\ \frac{(N+1)(N-1)(3+N^2)}{48N^2} & \text{if } N \text{ is odd} \end{cases}$

Handling Ties

If $Z_i = Z_j$ for any two observations from combined sample $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$, then use the average ranking procedure.

- C is calculated in same fashion (using average ranks)
- Average ranks with null distribution is approximate level α test
- Can still obtain an exact level α test via *conditional distribution*

Large sample approximation variance formulas:

$$V_*(C) = \begin{cases} \frac{mn\left[16\sum_{j=1}^{g} t_j r_j^2 - N(N+2)^2\right]}{16N(N-1)} & \text{if } N \text{ is even} \\ \frac{mn\left[16N\sum_{j=1}^{g} t_j r_j^2 - (N+1)^4\right]}{16N^2(N-1)} & \text{if } N \text{ is odd} \end{cases}$$

where

- g is the number of tied groups
- t_j is the size of the tied group
- r_j is the average rank score for group

Example 1: Data

Some simulated data:

X	R_k	Y	R_k
-0.63	(5)	0.78	(8)
0.18	(9)	-1.24	(2)
-0.84	(3)	-4.43	(1)
1.60	(5)	2.25	(1)
0.33	(10)	-0.09	(7)
-0.82	(4)	-0.03	(8)
0.49	(11)	1.89	(2)
0.74	(9)	1.64	(4)
0.58	(10)	1.19	(7)
-0.31	(6)	1.84	(3)
1.51	(6)		

Example 1: By Hand

X	R_k	Y	R_k
-0.63	(5)	0.78	(8)
0.18	(9)	-1.24	(2)
-0.84	(3)	-4.43	(1)
1.60	(5)	2.25	(1)
0.33	(10)	-0.09	(7)
-0.82	(4)	-0.03	(8)
0.49	(11)	1.89	(2)
0.74	(9)	1.64	(4)
0.58	(10)	1.19	(7)
-0.31	(6)	1.84	(3)
1.51	(6)		
$\sum$	78	$\sum$	43

$$C = \sum_{j=1}^{10} R_j = 43$$

Example 1: Using R (Hard Way)

```
> set.seed(1)
> x = round(rnorm(11),2)
> y = round(rnorm(10,0,2),2)
> m = length(x)
> n = length(y)
> N = m + n
> z = sort(c(x,y),index=TRUE)
> rz = seq(1,(N-1)/2)
> rz = c(rz,(N+1)/2,rev(rz))
> r = rz[sort(z$ix,index=TRUE)$ix]
> sum(r[1:11])
[1] 78
> sum(r[12:21])
[1] 43
```

Example 1: Using R (Easy Way)

```
> set.seed(1)
> x = round(rnorm(11),2)
> y = round(rnorm(10,0,2),2)
> ansari.test(x,y)

 Ansari-Bradley test

data:  x and y
AB = 78, p-value = 0.04563
alternative hypothesis: true ratio of scales is not equal to 1

> ansari.test(x,y,alternative="less")

 Ansari-Bradley test

data:  x and y
AB = 78, p-value = 0.02282
alternative hypothesis: true ratio of scales is less than 1
```

Dispersion/Location (Lepage)

Problem(s) of Interest

Like other two-sample problems, we have $N = m + n$ observations

- $X_1, \ldots, X_m$ are iid random sample from population 1
- $Y_1, \ldots, Y_n$ are iid random sample from population 2

We want to make inferences about difference in distributions

- Let F_1 and F_2 denote distributions of populations 1 and 2
- Null hypothesis is same distribution ($F_1(z) = F_2(z)$ for all z)

Using the location-scale parameter model, we have

- $F_1(z) = G([z - \theta_1]/\eta_1)$ and $F_2(z) = G([z - \theta_2]/\eta_2)$
- θ_j and η_j are median and scale parameters for population j

Assumptions

Within sample independence assumption

- $X_1, \ldots, X_m$ are iid random sample from population 1
- $Y_1, \ldots, Y_n$ are iid random sample from population 2

Between sample independence assumption

- Samples $\{X_i\}_{i=1}^m$ and $\{Y_i\}_{i=1}^n$ are mutually independent

Continuity assumption: both F_1 and F_2 are continuous distributions

Parameters of Interest and Hypothesis

Parameters of interest are the median difference and variance ratio:

$$\delta = \theta_1 - \theta_2 \quad \text{and} \quad \gamma^2 = \frac{V(X)}{V(Y)}$$

so that $\delta = 0$ whenever $\theta_1 = \theta_2$ and $\gamma^2 = 1$ whenever $V(X) = V(Y)$.

The null hypothesis about δ and γ^2 is

$$H_0 : \delta = 0 \;\; \text{and} \;\; \gamma^2 = 1$$

and there is only one alternative hypothesis

$$H_1 : \delta \neq 0 \;\; \text{and/or} \;\; \gamma^2 \neq 1$$

Test Statistic

The Lepage test statistic D is given by

$$D = \frac{[W - E(W)]^2}{V(W)} + \frac{[C - E(C)]^2}{V(C)}$$

- W is the Wilcoxon rank sum test statistic
- C is the Ansari-Bradley test statistic

Hypothesis Testing & Large Sample Approximation

One-Sided Upper Tail Test:

- $H_0 : \delta = 0$ and $\gamma^2 = 1$ versus $H_1 : \delta \neq 0$ and/or $\gamma^2 \neq 1$
- Reject H_0 if $D \geq d_\alpha$ where $P(D > d_\alpha) = \alpha$

This is the only appropriate test here...

- Large $\frac{(W-E(W))^2}{V(W)}$ and $\frac{(C-E(C))^2}{V(C)}$ provide more evidence against H_0
- We only reject H_0 if test statistic D is too large

Under H_0 and as $n \to \infty$, we have that $D \sim \chi^2_{(2)}$

- $\chi^2_{(2)}$ denotes a chi-squared distribution with 2 df
- Reject H_0 if $D \geq \chi^2_{(2);\alpha}$ where $P(\chi^2_{(2)} > \chi^2_{(2);\alpha}) = \alpha$

Example 2: Data

Same simulated data:

X	$[W_{R_k}]$	(C_{R_k})	Y	$[W_{R_k}]$	(C_{R_k})
-0.63	[5]	(5)	0.78	[14]	(8)
0.18	[9]	(9)	-1.24	[2]	(2)
-0.84	[3]	(3)	-4.43	[1]	(1)
1.60	[17]	(5)	2.25	[21]	(1)
0.33	[10]	(10)	-0.09	[7]	(7)
-0.82	[4]	(4)	-0.03	[8]	(8)
0.49	[11]	(11)	1.89	[20]	(2)
0.74	[13]	(9)	1.64	[18]	(4)
0.58	[12]	(10)	1.19	[15]	(7)
-0.31	[6]	(6)	1.84	[19]	(3)
1.51	[16]	(6)			

Example 2: By Hand

X	$[W_{R_k}]$	(C_{R_k})	Y	$[W_{R_k}]$	(C_{R_k})
-0.63	[5]	(5)	0.78	[14]	(8)
0.18	[9]	(9)	-1.24	[2]	(2)
-0.84	[3]	(3)	-4.43	[1]	(1)
1.60	[17]	(5)	2.25	[21]	(1)
0.33	[10]	(10)	-0.09	[7]	(7)
-0.82	[4]	(4)	-0.03	[8]	(8)
0.49	[11]	(11)	1.89	[20]	(2)
0.74	[13]	(9)	1.64	[18]	(4)
0.58	[12]	(10)	1.19	[15]	(7)
-0.31	[6]	(6)	1.84	[19]	(3)
1.51	[16]	(6)			
$\sum$	106	78	$\sum$	125	43

$$W_* = \frac{W-E(W)}{\sqrt{V(W)}} = \frac{W-n(N+1)/2}{\sqrt{mn(N+1)/12}} = \frac{125-110}{\sqrt{201.6667}} = 1.056268$$

$$C_* = \frac{C-E(C)}{\sqrt{V(C)}} = \frac{C-n(N+1)^2/(4N)}{\sqrt{mn(N+1)(3+N^2)/(48N^2)}} = \frac{43-57.61905}{\sqrt{50.75964}} = -2.051917$$

$$D = W_*^2 + C_*^2 = (1.056268)^2 + (-2.051917)^2 = 5.326067$$

Example 2: Using R (Hard Way)

```
> set.seed(1)
> x = round(rnorm(11),2)
> y = round(rnorm(10,0,2),2)
> m = length(x)
> n = length(y)
> N = m + n
> z = sort(c(x,y),index=TRUE)
> rz = seq(1,(N-1)/2)
> rz = c(rz,(N+1)/2,rev(rz))
> r = rz[sort(z$ix,index=TRUE)$ix]
> C = sum(r[12:21])
> rk = rank(c(x,y))
> W = sum(rk[12:21])
> Wstar = (W-n*(N+1)/2)/sqrt(m*n*(N+1)/12)
> Cstar = (C-n*((N+1)^2)/(4*N))/sqrt(m*n*(N+1)*(3+N^2)/(48*(N^2)))
> D = Wstar^2 + Cstar^2
> D
[1] 5.326067
> 1 - pchisq(D,2)
[1] 0.06973637
```

Example 2: Using R (Easy Way)

```
> require(NSM3)
> set.seed(1)
> x = round(rnorm(11),2)
> y = round(rnorm(10,0,2),2)
> pLepage(x,y)
Number of X values:  11 Number of Y values:  10
Lepage D  Statistic:  5.3261
Monte Carlo  (Using  10000 Iterations) upper-tail probability:  0.0643
```

Equality Test (Kolmogorov-Smirnov)

Problem(s) of Interest

Like other two-sample problems, we have $N = m + n$ observations

- $X_1, \ldots, X_m$ are iid random sample from population 1
- $Y_1, \ldots, Y_n$ are iid random sample from population 2

We want to make inferences about difference in distributions

- Let F_1 and F_2 denote distributions of populations 1 and 2
- Null hypothesis is same distribution ($F_1(z) = F_2(z)$ for all z)

Do NOT assume the location-scale parameter model

- More general test than the others
- Interested in *any* differences between F_1 and F_2

Assumptions

Within sample independence assumption

- $X_1, \ldots, X_m$ are iid random sample from population 1
- $Y_1, \ldots, Y_n$ are iid random sample from population 2

Between sample independence assumption

- Samples $\{X_i\}_{i=1}^{m}$ and $\{Y_i\}_{i=1}^{n}$ are mutually independent

Continuity assumption: both F_1 and F_2 are continuous distributions

Parameter of Interest and Hypothesis

Parameter of interest is the maximum absolute difference between the CDFs of X and Y:

$$\omega = \max_{-\infty \leq z \leq \infty} |F_1(z) - F_2(z)|$$

The null hypothesis about ω is

$$H_0 : \omega = 0$$

and there is only one alternative hypothesis

$$H_1 : \omega > 0$$

Test Statistic

Define the maximum absolute difference between the empirical CDFs of X and Y as

$$\hat{\omega} = \max_{k=1,\ldots,N} |\hat{F}_{1,m}(Z_{(k)}) - \hat{F}_{2,n}(Z_{(k)})|$$

where

- $\hat{F}_{1,m}(z) = \frac{\sum_{i=1}^{m} 1_{\{X_i \leq z\}}}{m}$ and $\hat{F}_{2,n}(z) = \frac{\sum_{j=1}^{n} 1_{\{Y_j \leq z\}}}{n}$
- $Z_{(k)}$ denotes the k-th order statistic of the combined sample

The Kolmogorov-Smirnov test statistic K is given by

$$K = \frac{mn}{d}\hat{\omega}$$

where d is greatest common divisor of m and n

Distribution of Test Statistic under H_0

Under H_0 all $\binom{N}{n}$ arrangements of ranks occur with equal probability

- Given (N, n), calculate K for all $\binom{N}{n}$ possible outcomes
- Each outcome has probability $1/\binom{N}{n}$ under H_0

Example null distribution with $m = 3$ and $n = 2$:

Y-ranks	$F_{1,m}(Z_{(k)})$	$F_{2,n}(Z_{(k)})$	$\hat{\omega}$	K	Probability under H_0
1,2	$(0, 0, \frac{1}{3}, \frac{2}{3}, 1)$	$(\frac{1}{2}, 1, 1, 1, 1)$	1	6	1/10
1,3	$(0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1)$	$(\frac{1}{2}, \frac{1}{2}, 1, 1, 1)$	2/3	4	1/10
1,4	$(0, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 1, 1)$	1/2	3	1/10
1,5	$(0, \frac{1}{3}, \frac{2}{3}, 1, 1)$	$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 1)$	1/2	3	1/10
2,3	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1)$	$(0, \frac{1}{2}, 1, 1, 1)$	2/3	4	1/10
2,4	$(\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	$(0, \frac{1}{2}, \frac{1}{2}, 1, 1)$	1/3	2	1/10
2,5	$(\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1, 1)$	$(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 1)$	1/2	3	1/10
3,4	$(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	$(0, 0, \frac{1}{2}, 1, 1)$	2/3	4	1/10
3,5	$(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1, 1)$	$(0, 0, \frac{1}{2}, \frac{1}{2}, 1)$	2/3	4	1/10
4,5	$(\frac{1}{3}, \frac{2}{3}, 1, 1, 1)$	$(0, 0, 0, \frac{1}{2}, 1)$	1	6	1/10

Note: there are $\binom{5}{2} = 10$ possibilities and $d = 1$ is the gcd

Hypothesis Testing & Large Sample Approximation

One-Sided Upper Tail Test:

- $H_0 : \omega = 0$ versus $H_1 : \omega > 0$
- Reject H_0 if $K \geq k_\alpha$ where $P(K > k_\alpha) = \alpha$

This is the only appropriate test here. . .

- Large $\hat{\omega} = |\hat{F}_{1,m}(z) - \hat{F}_{2,n}(z)|$ provide more evidence against H_0
- We only reject H_0 if test statistic K is too large

Under H_0 and as $\min(m, n) \to \infty$, Smirnov (1939) showed that

- $K^* = (mn/N)^{1/2}\hat{\omega} = \frac{d}{(mnN)^{1/2}} K$
- $P(K^* < z) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 z^2}$

Handling Ties

The empirical CDFs $\hat{F}_{1,m}$ and $\hat{F}_{2,n}$ are well defined when ties occur within and/or between the two samples.

Consequently, we do NOT need any adjustment, and we still have a conservative test.

- Significance level will not exceed nominal level α

Example 3: Data

Same simulated data:

k	$Z_{(k)}$	Population	$\hat{F}_{1,m}(Z_{(k)})$	$\hat{F}_{2,n}(Z_{(k)})$
1	-4.43	2	0/11	1/10
2	-1.24	2	0/11	2/10
3	-0.84	1	1/11	2/10
4	-0.82	1	2/11	2/10
5	-0.63	1	3/11	2/10
6	-0.31	1	4/11	2/10
7	-0.09	2	4/11	3/10
8	-0.03	2	4/11	4/10
9	0.18	1	5/11	4/10
10	0.33	1	6/11	4/10
11	0.49	1	7/11	4/10
12	0.58	1	8/11	4/10
13	0.74	1	9/11	4/10
14	0.78	2	9/11	5/10
15	1.19	2	9/11	6/10
16	1.51	1	10/11	6/10
17	1.60	1	11/11	6/10
18	1.64	2	11/11	7/10
19	1.84	2	11/11	8/10
20	1.89	2	11/11	9/10
21	2.25	2	11/11	10/10

Note: $m = 11$ and $n = 10$ so that $d = 1$.

Example 3: By Hand

k	$Z_{(k)}$	Population	$\hat{F}_{1,m}(Z_{(k)})$	$\hat{F}_{2,n}(Z_{(k)})$	$\hat{\omega}$
1	-4.43	2	0/11	1/10	0.1000
2	-1.24	2	0/11	2/10	0.2000
3	-0.84	1	1/11	2/10	0.1091
4	-0.82	1	2/11	2/10	0.0182
5	-0.63	1	3/11	2/10	0.0727
6	-0.31	1	4/11	2/10	0.1636
7	-0.09	2	4/11	3/10	0.0636
8	-0.03	2	4/11	4/10	0.0364
9	0.18	1	5/11	4/10	0.0545
10	0.33	1	6/11	4/10	0.1455
11	0.49	1	7/11	4/10	0.2364
12	0.58	1	8/11	4/10	0.3273
13	0.74	1	9/11	4/10	0.4182
14	0.78	2	9/11	5/10	0.3182
15	1.19	2	9/11	6/10	0.2182
16	1.51	1	10/11	6/10	0.3091
17	1.60	1	11/11	6/10	0.4000
18	1.64	2	11/11	7/10	0.3000
19	1.84	2	11/11	8/10	0.2000
20	1.89	2	11/11	9/10	0.1000
21	2.25	2	11/11	10/10	0.0000

Note: $m = 11$ and $n = 10$ so that $d = 1$.

$$K = (11)(10)(0.4182) = 46$$

Example 3: Using R (Hard Way)

```
> set.seed(1)
> x = round(rnorm(11),2)
> y = round(rnorm(10,0,2),2)
> z = sort(c(x,y),index=TRUE)
> zlab = c(rep("x",11),rep("y",10))
> j = ifelse(zlab[z$ix]=="x",1L,2L)
> F1vec = F2vec = 0
> for(k in 2:22){
+     if(j[k-1]==1L){
+         F1vec = c(F1vec,F1vec[k-1]+1)
+         F2vec = c(F2vec,F2vec[k-1]+0)
+     } else{
+         F1vec = c(F1vec,F1vec[k-1]+0)
+         F2vec = c(F2vec,F2vec[k-1]+1)
+     }
+ }
> F1vec = F1vec[2:22]/11
> F2vec = F2vec[2:22]/10
> omega = abs(F1vec-F2vec)
> max(omega)
[1] 0.4181818
```

Example 3: Using R (Easy Way)

```
> set.seed(1)
> x=round(rnorm(11),2)
> y=round(rnorm(10,0,2),2)
> ks.test(x,y)

 Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.4182, p-value = 0.2586
alternative hypothesis: two-sided
```

Example 3: Using R (Easy Way, More Data)

```
> set.seed(1)
> x=round(rnorm(100),2)
> y=round(rnorm(100,0,2),2)
> ks.test(x,y)

 Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.24, p-value = 0.006302
alternative hypothesis: two-sided

Warning message:
In ks.test(x, y) : p-value will be approximate in the presence of ties
```