Two Sample Tests: II
Bios 662

Michael G. Hudgens, Ph.D.

mhudgens@bios.unc.edu

http://www.bios.unc.edu/∼mhudgens

2008-09-10 15:13

Outline

- Last time: parametric/large sample
- Wilcoxon rank sum
  - Hodges-Lehmann estimator, CIs
- Permutation test
- Kolmogorov-Smirnov

Wilcoxon (Mann-Whitney) Rank Sum Test

1. Assume $Y_{1j}, \ldots, Y_{n_j j}$ iid $F_j(y)$; $j = 1, 2$

$$H_0 : F_1(y) = F_2(y)$$

$$H_A : F_1(y + \Delta) = F_2(y)$$

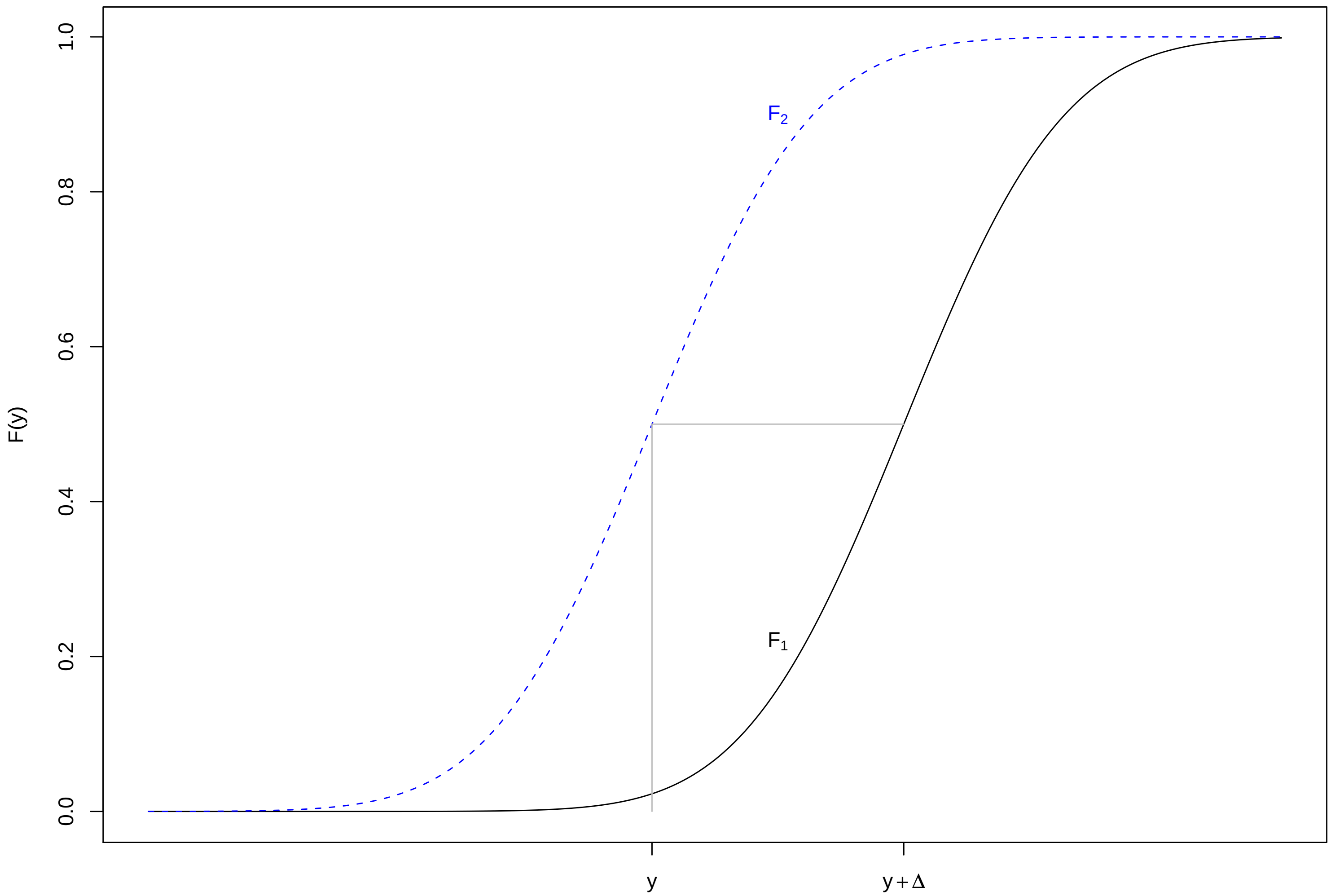
   where Δ is a non-zero constant

2. Pool the two samples

3. Rank them from smallest to largest

4. Compute the sum of the ranks, W_1, in group 1

Wilcoxon (Mann-Whitney) Rank Sum Test



Wilcoxon Rank Sum Test

- There are $N = n_1 + n_2$ subjects in our study
- Thus there are $\binom{N}{n_1}$ possible outcomes
- Under H_0, each is equally likely
- We compute the distribution of W_1 by enumeration

Wilcoxon Rank Sum Test: Example

- A new drug is being test in humans for the first time to assess effect on CD4+ T cells in patients with HIV
- 7 individuals are randomized to 2 groups: control ($n_1 = 3$) or drug ($n_2 = 4$)
- Endpoint is percent change in CD4+ count from baseline
- Null hypothesis is the drug has no effect

$$H_0 : \Delta = 0; H_A : \Delta \neq 0$$

Wilcoxon Rank Sum Test: Example

- Data: control (65, 73, 69); drug (89, 70, 92, 88)
- There are $\binom{7}{3} = 35$ possible outcomes of the study
  i.e. there are 35 possible sets of rankings for group 1

Wilcoxon Rank Sum Test: $n_1 = 3, n_2 = 4$

Ranks	W_1	Ranks	W_1	Ranks	W_1
1,2,3	6	1,5,6	12	2,6,7	15
1,2,4	7	1,5,7	13	3,4,5	12
1,2,5	8	1,6,7	14	3,4,6	13
1,2,6	9	2,3,4	9	3,4,7	14
1,2,7	10	2,3,5	10	3,5,6	14
1,3,4	8	2,3,6	11	3,5,7	15
1,3,5	9	2,3,7	12	3,6,7	16
1,3,6	10	2,4,5	11	4,5,6	15
1,3,7	11	2,4,6	12	4,5,7	16
1,4,5	10	2,4,7	13	4,6,7	17
1,4,6	11	2,5,6	13	5,6,7	18
1,4,7	12	2,5,7	14		

Wilcoxon Rank Sum Test: $n_1 = 3, n_2 = 4$

w	$F(w)$	w	$F(w)$
6	0.0286	13	0.6857
7	0.0571	14	0.8000
8	0.1143	15	0.8857
9	0.2000	16	0.9429
10	0.3142	17	0.9714
11	0.4286	18	1
12	0.5714		

Wilcoxon Rank Sum Test: $n_1 = 3, n_2 = 4$

- Note it is impossible to reject H_0 for a two-sided alternative when $\alpha = 0.05$.
- For a two-sided $\alpha = 0.1$ test

$$C_\alpha = \{6, 18\}$$

- Observed $W_1 = 1 + 2 + 4 = 7$; do not reject H_0

Wilcoxon Rank Sum Test

- Note
$$W_1 + W_2 = \sum_{i=1}^{N} i = \frac{N(N+1)}{2}$$
- Thus
$$E(W_1) = \frac{n_1}{N}\frac{N(N+1)}{2} = \frac{n_1(N+1)}{2}$$
- Similarly
$$V(W_1) = \frac{n_1 n_2(N+1)}{12}$$
(cf Lehmann 1998, Example 3, p 332)

Wilcoxon Rank Sum Test: Large Sample Approx

- If n_1 and n_2 are large

$$Z = \frac{W_1 - E(W_1)}{\sqrt{V(W_1)}}$$

will be approx $N(0, 1)$

- Approximation is good for $n_1, n_2 \geq 12$
- If there are ties

$$V(W_1) = \frac{n_1 n_2 (N+1)}{12} - \frac{n_1 n_2}{12N(N-1)} \sum_{i=1}^{q} t_i(t_i - 1)(t_i + 1)$$

Wilcoxon Rank Sum Test: BW Example

Drug	Rank	Placebo	Rank
6.9	**18**	6.4	11
7.6	**25.5**	6.7	13
7.3	**23.5**	5.4	3
7.6	**25.5**	8.2	**28.5**
6.8	**15**	5.3	2
7.2	22	6.6	12
8.0	27	5.8	**8.5**
5.5	4	5.7	**6.5**
5.8	**8.5**	6.2	10
7.3	**23.5**	7.1	21
8.2	**28.5**	7.0	20
6.9	**18**	6.9	**18**
6.8	**15**	5.6	5
5.7	**6.5**	4.2	1
8.6	30	6.8	**15**

Wilcoxon Rank Sum Test: BW Example

- $H_0 : \Delta = 0$; $H_A : \Delta > 0$
- $C_{.05} = \{z : z > 1.645\}$
- $E(W_1) = \frac{15(31)}{2} = 232.5$
- $V(W_1|\text{no ties}) = \frac{15^2(31)}{12} = 581.25$

Wilcoxon Rank Sum Test: BW Example

- Tie correction:

  $q = 7;\ t_1 = t_2 = 2;\ t_3 = t_4 = 3;\ t_5 = t_6 = t_7 = 2$

$$\sum_{i=1}^{q} t_i(t_i - 1)(t_i + 1) = 78$$

$$V(W_1) = 581.25 - \frac{78(15)^2}{12(30)(29)} = 579.57$$

Wilcoxon Rank Sum Test: BW Example

- $w_1 = 290.5$

$$z = \frac{290.5 - 232.5}{\sqrt{579.57}} = 2.409;$$

- Reject H_0
- $p = 1 - \Phi(2.409) = 0.008$
- Note: without tie correction $z = 2.406$; $p = 0.008$

Wilcoxon Rank Sum Test: BW Example

- SAS

```
proc npar1way wilcoxon correct=no; class trt; var bw;

          Wilcoxon Scores (Rank Sums) for Variable bw
                  Classified by Variable trt

                  Sum of     Expected      Std Dev          Mean
trt       N       Scores     Under H0     Under H0         Score
----------------------------------------------------------------
drug     15       290.50       232.50    24.074239     19.366667
plac     15       174.50       232.50    24.074239     11.633333

              Average scores were used for ties.


                    Wilcoxon Two-Sample Test

              Statistic (S)               290.5000

              Normal Approximation
              Z                             2.4092
              One-Sided Pr >  Z             0.0080
              Two-Sided Pr > |Z|            0.0160
```

Wilcoxon Rank Sum Test: BW Example

- R

```
> wilcox.test(bw$drug,bw$placebo,alternative="greater",exact=F,correct=F)

        Wilcoxon rank sum test

data:  bw$drug and bw$placebo
W = 170.5, p-value = 0.007993
alternative hypothesis: true mu is greater than 0
```

Wilcoxon Rank Sum Exact P-values

- For two-sided alternative, exact p-values are computed (under the null) by

$$\Pr[|W_1 - E(W_1)| \geq |w_1 - E(W_1)|]$$

where

$$E(W_1) = \frac{n_1(N+1)}{2}$$

- Without ties, distribution of W_1 is symmetric about $E(W_1)$

Wilcoxon Rank Sum Exact P-values: Example

- Suppose $\boldsymbol{Y}_1 = (65, 70, 73)$ and $\boldsymbol{Y}_2 = (70, 89)$
- There are $\binom{5}{2} = 10$ possible rankings for group 1

Ranks	W_1	Ranks	W_1
1,2.5,2.5	6	1,4,5	10
1,2.5,4	7.5	2.5,2.5,4	9
1,2.5,4	7.5	2.5,2.5,5	10
1,2.5,5	8.5	2.5,4,5	11.5
1,2.5,5	8.5	2.5,4,5	11.5

- Thus $|w_1 - E(W_1)| = |7.5 - 9| = 1.5 \rightarrow p = 0.5$

Wilcoxon Rank Sum Exact P-values: R

```
> wilcox.exact(c(65,70,73),c(70,89))

        Exact Wilcoxon rank sum test

data:  c(65, 70, 73) and c(70, 89)
W = 1.5, p-value = 0.5
alternative hypothesis: true mu is not equal to 0
```

Wilcoxon Rank Sum Exact P-values: SAS

```
proc npar1way wilcoxon;
 class trt;
 var bw;
 exact wilcoxon;
run;

The NPAR1WAY Procedure

            Wilcoxon Scores (Rank Sums) for Variable bw
                    Classified by Variable trt

                         Sum of      Expected       Std Dev           Mean
trt         N            Scores      Under H0      Under H0          Score
-------------------------------------------------------------------------
1           3              7.50           9.0      1.688194          2.500
2           2              7.50           6.0      1.688194          3.750

                Average scores were used for ties.

     Wilcoxon Two-Sample Test

Statistic (S)                  7.5000

Exact Test
One-Sided Pr >=  S             0.3000
Two-Sided Pr >= |S - Mean|     0.5000
```

Mann-Whitney Test

- Consider all $n_1 n_2$ possible pairs

$$(Y_{1i}, Y_{2j}); i = 1, 2, \ldots, n_1; j = 1, 2, \ldots, n_2$$

- Let U_1 equal the number of pairs with $Y_{1i} < Y_{2j}$
- It can be shown that

$$U_1 = \frac{n_1(N + n_2 + 1)}{2} - W_1$$

- Reject

$$H_0 : \Delta = 0 \text{ vs } H_A : \Delta > 0$$

if W_1 large ie U_1 small

Mann-Whitney Test

- Ie Mann-Whitney and Wilcoxon rank sum test are equivalent
- This explains the R output

$$U_2 = \frac{n_2(N + n_1 + 1)}{2} - W_2 = \frac{15 * 46}{2} - 174.5 = 170.5$$

- Reject

$$H_0 : \Delta = 0 \text{ vs } H_A : \Delta > 0$$

if W_2 small ie U_2 large

Mann-Whitney Test

- Table A.10 text: Critical values

  For example, $n_1 = n_2 = 15$ one-sided test

  $\alpha = 0.01$: $CV = 169$

  $\alpha = 0.005$: $CV = 174$

- Efficiency

  Compared to t-test ARE = 0.955 under normality; never worse than 0.864

Hodges-Lehmann Estimator

- Assume $F_1(y + \Delta) = F_2(y)$ for some constant Δ.
- WRS

$$H_0 : \Delta = 0$$

$$H_A : \Delta \neq 0$$

- Estimate Δ by the amount $\hat{\Delta}$ by which the Y_{2j}'s must be shifted to give the best possible agreement with the Y_{1i}'s

Hodges-Lehmann Estimator

- From Mann-Whitney perspective, want $Y_{1i} > Y_{2j} + \hat{\Delta}$ half of the time
- Thus

  $$\hat{\Delta} = \text{median}\{Y_{1i} - Y_{2j} : i = 1, \ldots, n_1, ; j = 1, \ldots, n_2\}$$

- CI for Δ?

CIs by Inverting a Test

- For each possible value of $\theta_0 \in \Omega$, let $C_\alpha(\theta_0)$ denote the critical region for testing $H_0 : \theta = \theta_0$ at the α level of significance

- Let X denote the corresponding test statistic and

$$S(X) = \{\theta : X \notin C_\alpha(\theta)\}$$

- Claim: $S(X)$ is a $(1 - \alpha) \times 100\%$ CI for θ

- Proof:

$$\Pr_\theta[\theta \in S(X)] = \Pr_\theta[X \notin C_\alpha(\theta)] \geq 1 - \alpha$$

CIs by Inverting a Test

- For given value x of a test statistic X, find all values of θ where we would not reject at the α level of significance

- For example, consider the Wilcoxon Rank Sum test:

  For a given value of W_1, find all values of Δ where we fail to reject H_0

Wilcoxon Rank Sum Test: BW Example in R

```
> wilcox.test(bw$drug,bw$placebo,exact=F,correct=F,conf.int=T)

        Wilcoxon rank sum test

data:  bw$drug and bw$placebo
W = 170.5, p-value = 0.01599
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 0.1000055 1.5000265
sample estimates:
difference in location
             0.8000382

> median(outer(bw$drug,bw$placebo,"-"))
[1] 0.8
```

Wilcoxon Rank Sum Test: BW Example in R

```
> wilcox.test(bw$drug,bw$placebo+.1,exact=F,correct=F)

W = 165, p-value = 0.02927

> wilcox.test(bw$drug,bw$placebo+.10001,exact=F,correct=F)

W = 159, p-value = 0.05366

> wilcox.test(bw$drug,bw$placebo+1.5,exact=F,correct=F)

W = 68, p-value = 0.06442

> wilcox.test(bw$drug,bw$placebo+1.50001,exact=F,correct=F)

W = 63, p-value = 0.03997
```

Permutation Test

- Cf Section 8.9 of text; $H_0 : F_1 = F_2$
- Test statistic $D \equiv \bar{Y}_1 - \bar{Y}_2$
- N subjects randomly assigned to two groups; n_1, n_2
- There are $\binom{N}{n_1}$ possible group assignments and each is equally likely
- Each of these assignments results in a value of $\bar{Y}_1 - \bar{Y}_2$
- Compute $\bar{Y}_1 - \bar{Y}_2$ for each possible assignment

Permutation Test

- Compute the CDF of $\bar{Y}_1 - \bar{Y}_2$ under $H_0 : F_1 = F_2$
- From the CDF, determine the critical region
- Example: HIV study $\binom{7}{3} = 35$ possible group assignments

Example: All Possible Group Assignments

Group 1	Group2	$\bar{Y}_1 - \bar{Y}_2$	Group 1	Group 2	$\bar{Y}_1 - \bar{Y}_2$
65 69 70	73 88 89 92	-17.50	65 69 73	70 88 89 92	-15.75
65 69 88	70 73 89 92	-7.00	65 69 89	70 73 88 92	-6.42
65 69 92	70 73 88 89	-4.67	65 70 73	69 88 89 92	-15.17
65 70 88	69 73 89 92	-6.42	65 70 89	69 73 88 92	-5.83
65 70 92	69 73 88 89	-4.08	65 73 88	69 70 89 92	-4.67
65 73 89	69 70 88 92	-4.08	65 73 92	69 70 88 89	-2.33
65 88 89	69 70 73 92	4.67	65 88 92	69 70 73 89	6.42
65 89 92	69 70 73 92	6.00	69 70 73	65 88 89 92	-12.83
69 70 88	65 73 89 92	-4.08	69 70 89	65 73 88 92	-3.50
69 70 92	65 73 88 89	-1.75	69 73 88	65 70 89 92	-2.33
69 73 89	65 70 88 92	-1.75	69 73 92	65 70 88 89	0.00
69 88 89	65 70 73 92	7.00	69 88 92	65 70 73 89	8.75
69 89 92	65 70 73 88	9.33	70 73 88	65 69 89 92	-1.75
70 73 89	65 69 88 92	-1.17	70 73 92	65 69 88 89	0.58
70 88 89	65 69 73 92	7.58	70 88 92	65 69 73 89	9.33
70 89 92	65 69 73 88	9.92	73 88 89	65 69 70 92	9.33
73 88 92	65 69 70 89	11.08	73 89 92	65 69 70 92	10.67
88 89 92	65 69 70 73	20.42			

EDF of $\bar{Y}_1 - \bar{Y}_2$

d	$\Pr[\bar{Y}_1 - \bar{Y}_2 \leq d]$	d	$\Pr[\bar{Y}_1 - \bar{Y}_2 \leq d]$
-17.50	0.029	6.41	0.714
-15.75	0.057	7.00	0.743
-15.17	0.086	7.58	0.771
-12.83	0.114	8.75	0.800
-7.00	0.143	9.33	0.886
-6.42	0.200	9.92	0.914
-5.83	0.229	10.67	0.943
-4.67	0.286	11.08	0.971
-4.08	0.371	20.42	1.000
-3.50	0.400		
-2.33	0.457		
-1.75	0.543		
-1.17	0.571		
0.00	0.600		
0.58	0.629		
4.67	0.657		
6.00	0.686		

Permutation Test Example

- Symmetric critical region for $\alpha = 0.1$

$$C_{.10} = \{D : D = -17.5 \text{ or } D = 20.42\}$$

where $D = \bar{Y}_1 - \bar{Y}_2$

- Observed $d = -15.75$; do not reject H_0

Permutation Test

- No assumptions except random assignment
- Computations extensive if N is moderately large

  e.g., $N = 20$, no. of permutations $> 2 * 10^{18}$

  However, $\binom{20}{10} = 184,756$
- *Conditional test*: Y's fixed
- Exact: probability of rejecting null when it holds never exceeds the nominal significance level

Kolmogorov-Smirnov Test

- Want to test

$$H_0 : F_1(y) = F_2(y) \text{ for all } y$$

versus general alternative

$$H_A : F_1(y) \neq F_2(y) \text{ for at least one } y$$

- KS test

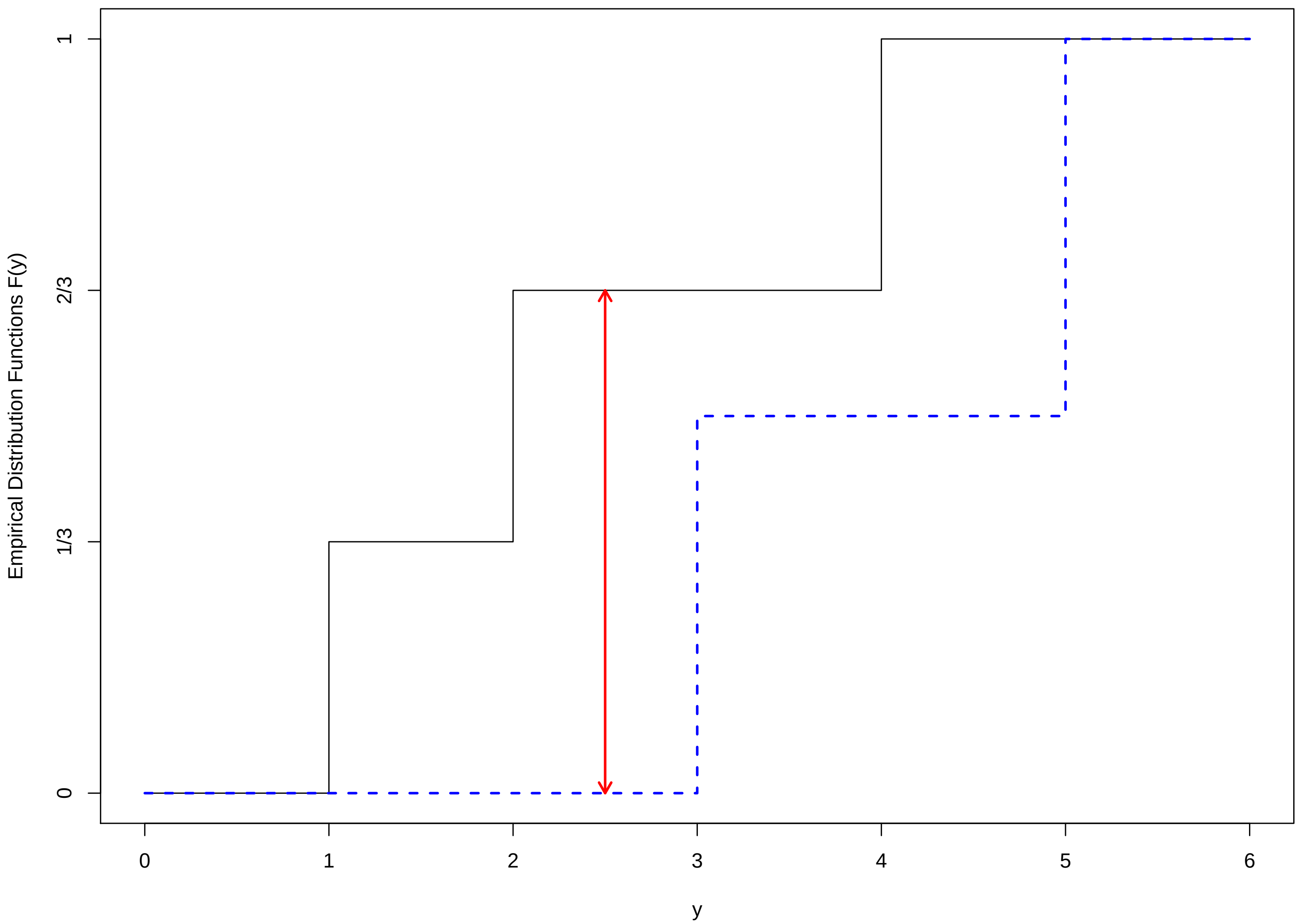
$$D = \max_y |F_{1n}(y) - F_{2m}(y)|$$

where $F_{1n}(y)$ and $F_{2m}(y)$ are the EDFs for samples 1 and 2

Kolmogorov-Smirnov Test

- Can be viewed as a rank test (text p 279)
- Large sample based critical values p 268 text
- For small samples, exact distribution based on enumeration of all possible group assignments as illustrated in the following example
- Example: $\boldsymbol{Y}_1 = (1, 2, 4)$, $\boldsymbol{Y}_2 = (3, 5)$

Kolmogorov-Smirnov Test: Example



Kolmogorov-Smirnov Test: Example

- Observe $d = 2/3$
- There are 10 possible group assignments

$\boldsymbol{Y}_1$	D	$\boldsymbol{Y}_1$	D
1,2,3	1	1,4,5	2/3
1,2,4	2/3	2,3,4	1/2
1,2,5	2/3	2,3,5	1/2
1,3,4	1/2	2,4,5	2/3
1,3,5	1/3	3,4,5	1

- Thus $p = \Pr[D \geq d] = .6$

KS in R and SAS

- R

```
> ks.test(c(1,2,4),c(3,5))

        Two-sample Kolmogorov-Smirnov test

data:  c(1, 2, 4) and c(3, 5)
D = 0.6667, p-value = 0.6
alternative hypothesis: two-sided
```

- SAS

```
proc npar1way;
 class trt;
 var bw;
 exact ks;
run;
```

KS in SAS

```
  The NPAR1WAY Procedure

     Kolmogorov-Smirnov Test for Variable bw
            Classified by Variable trt

                     EDF at    Deviation from Mean
trt         N       Maximum         at Maximum

1           3      0.666667           0.461880
2           2      0.000000          -0.565685
Total       5      0.400000


Kolmogorov-Smirnov Two-Sample Test

D = max |F1 - F2|      0.6667
Asymptotic Pr >  D     0.6604
Exact      Pr >= D     0.6000
```

Discussion

- WRS: Default non-parametric test
- Permutation
  - Asy equivalent to t-test; thus most powerful asymptotically under normality
  - Computationally intensive since unique to each data set
  - Sensitive to outliers
- KS
  - Employ if trying to detect difference in distributions other than location shift