BIOS 662 Fall 2018

Two Sample Tests

David Couper, Ph.D.

david_couper@unc.edu

or

couper@bios.unc.edu

https://sakai.unc.edu/portal

Two Sample Test Settings

- Single cross-sectional sample, comparing two sub-samples
- Comparing samples from two different populations (two cross-sectional samples, unmatched case control study)
- Single sample; subjects randomly allocated to different interventions (experiment, clinical trials)

Fundamentals

- Definition 5.2. Two random variables Y_1 and Y_2 are *independent* if for all y_1 and y_2

$$\Pr[Y_1 \leq y_1, Y_2 \leq y_2] = \Pr[Y_1 \leq y_1] \Pr[Y_2 \leq y_2]$$

- Result 5.1. If Y_1 and Y_2 are independent random variables, then for any two constants a_1 and a_2 the random variable

$$W = a_1 Y_1 + a_2 Y_2$$

has mean and variance

$$E(W) = a_1 E(Y_1) + a_2 E(Y_2)$$

$$\mathrm{Var}(W) = a_1^2 \mathrm{Var}(Y_1) + a_2^2 \mathrm{Var}(Y_2)$$

Fundamentals

- Result 5.2. If Y_1 and Y_2 are independent random variables that are **normally** distributed, then

$$W = a_1 Y_1 + a_2 Y_2$$

is normally distributed with mean and variance given by Result 5.1

- Corollary: If $\bar{Y}_1$ and $\bar{Y}_2$ are based on two independent random samples of size n_1 and n_2 from two normal distributions with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, then

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Fundamentals

- Result 5.3. If $\bar{Y}_1$ and $\bar{Y}_2$ are based on two independent random samples of size n_1 and n_2 from two normal distributions with means μ_1 and μ_2 and the same variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{1/n_1 + 1/n_2}} \sim t_{n_1+n_2-2}$$

where

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Note: If $n_1 = n_2$ then

$$s_p^2 = (s_1^2 + s_2^2)/2$$

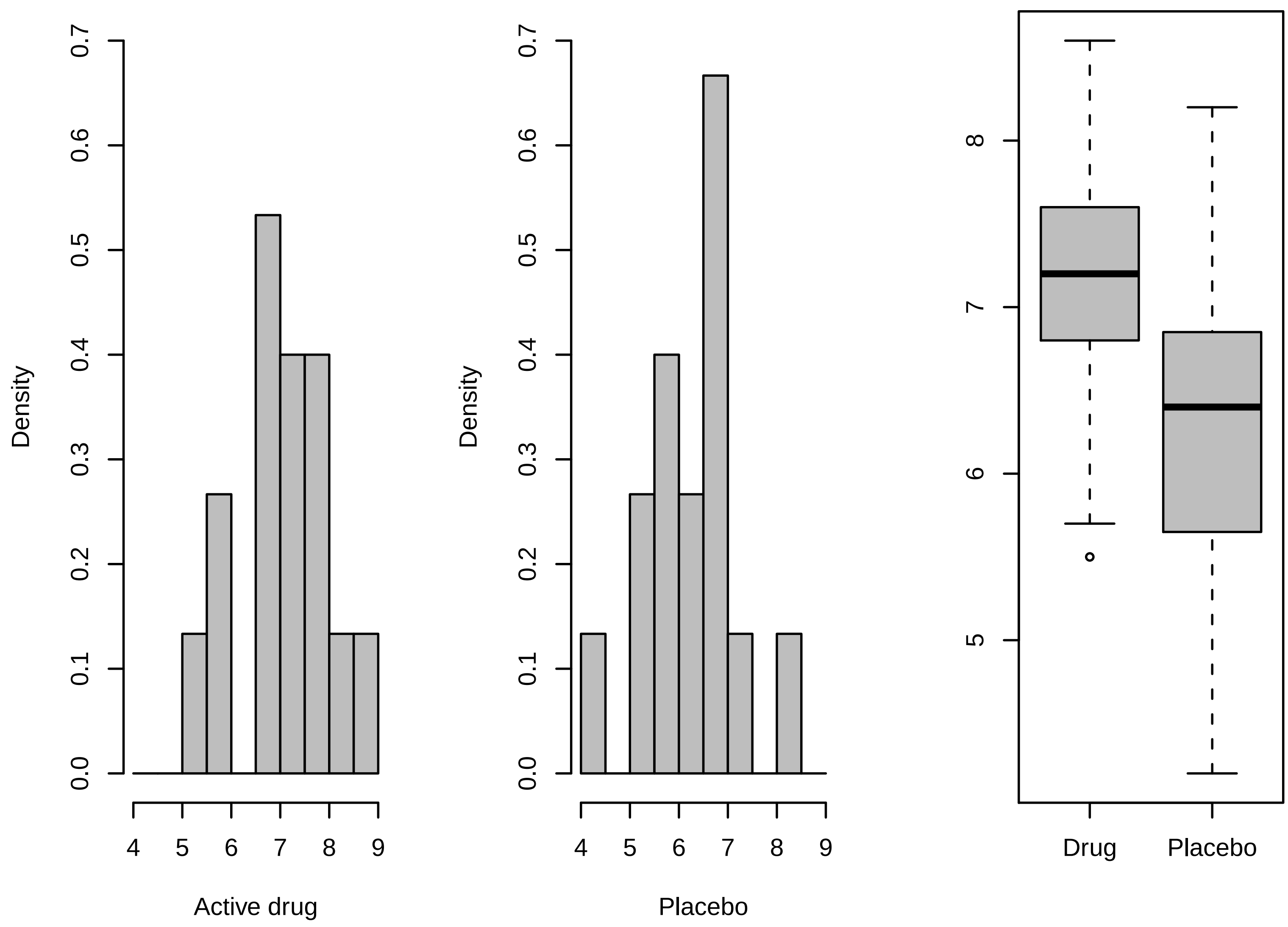
Two Sample t-test Example

- An experiment was conducted to see if a drug could prevent premature birth or low birthweight
- 30 women at risk of premature birth were randomly assigned to take the drug or a matching placebo (15 in each group)
- Endpoint: birthweight (in pounds)
- Let $1 =$ drug, $2 =$ placebo
- Consider $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 > \mu_2$
- $C_\alpha = \{t : t > t_{1-\alpha;28}\}$
- $C_{0.05} = \{t : t > 1.7\}$

Two Sample t-test Example cont.

Drug	Placebo
6.9	6.4
7.6	6.7
7.3	5.4
7.6	8.2
6.8	5.3
7.2	6.6
8.0	5.8
5.5	5.7
5.8	6.2
7.3	7.1
8.2	7.0
6.9	6.9
6.8	5.6
5.7	4.2
8.6	6.8

Two Sample t-test Example cont.



Two Sample t-test Example cont.

- $\bar{y}_1 = 7.08, \;\; s_1 = 0.899$
- $\bar{y}_2 = 6.26, \;\; s_2 = 0.961$
- Thus

$$s_p^2 = \frac{14(0.899)^2 + 14(0.961)^2}{28} = 0.8657$$

$$t = \frac{7.08 - 6.26}{0.930\sqrt{2/15}} = 2.41$$

- Because $t \in C_{0.05}$, reject H_0

$$p = 1 - F_{t_{28}}(2.41) = 0.011$$

Two Sample t-test Example cont.

- R

```
> t.test(bw$drug,bw$placebo,var.equal=TRUE,alternative="greater")

        Two Sample t-test

data:  bw$drug and bw$placebo
t = 2.4136, df = 28, p-value = 0.01129
alternative hypothesis: true difference in means is greater than 0
```

Two Sample t-test Example cont.

- SAS

```
proc ttest; class trt; var bw;

The TTEST Procedure

Variable:  bw

trt              N        Mean     Std Dev     Std Err

drug            15      7.0800      0.8994      0.2322
placebo         15      6.2600      0.9605      0.2480
Diff (1-2)              0.8200      0.9304      0.3397

Method           Variances        DF    t Value    Pr > |t|

Pooled           Equal            28       2.41      0.0226
Satterthwaite    Unequal       27.88       2.41      0.0226
```

Homogeneity of Variance

- Want to test

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs.} \quad H_A : \sigma_1^2 \neq \sigma_2^2$$

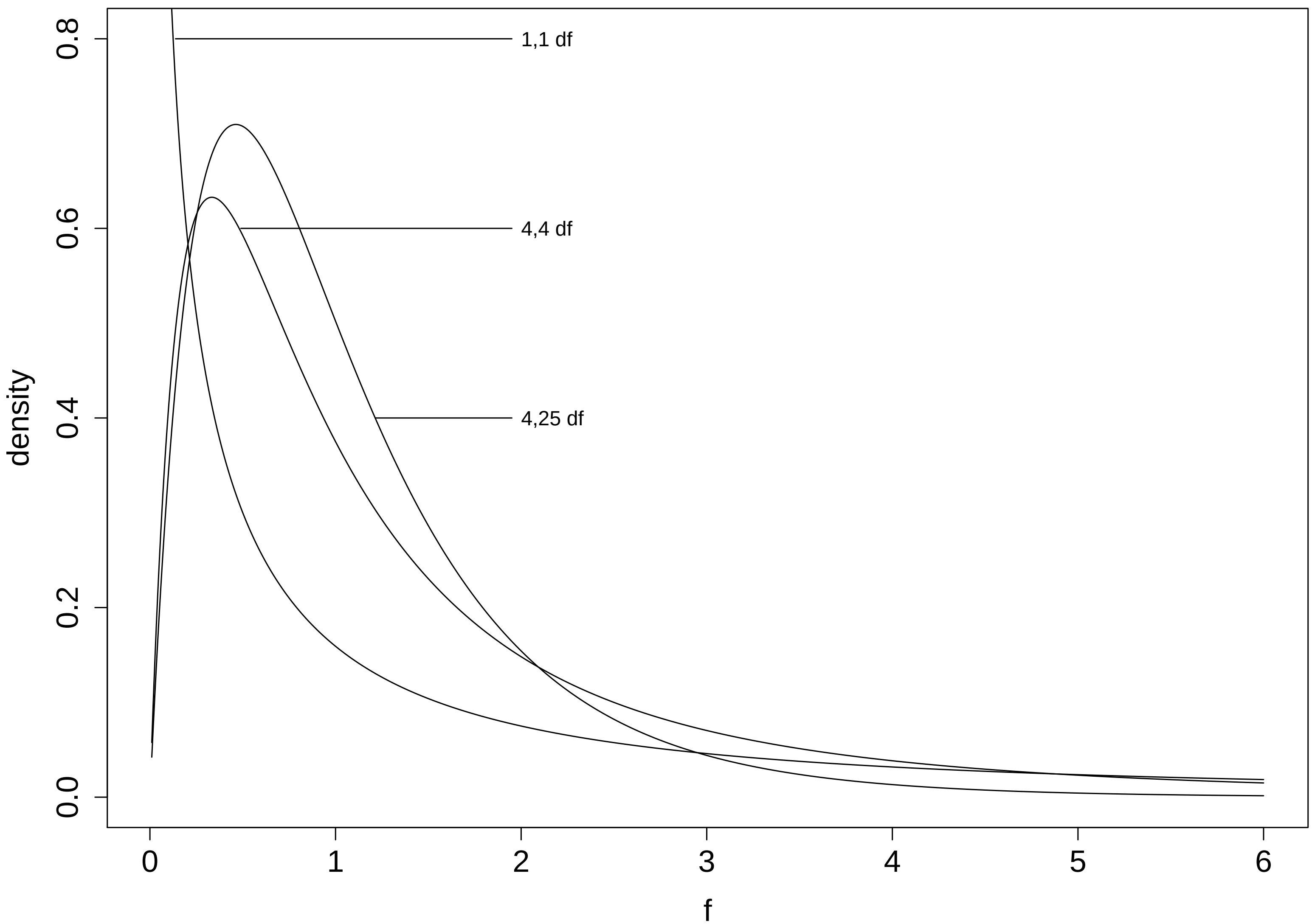
- We know that (assuming normality)

$$\frac{(n_k - 1)s_k^2}{\sigma_k^2} \sim \chi^2_{n_k - 1} \quad \text{for} \quad k = 1, 2$$

- If X_1 and X_2 are independent random variables with $X_1 \sim \chi^2_{v_1}$ and $X_2 \sim \chi^2_{v_2}$, then

$$\frac{X_1/v_1}{X_2/v_2} \sim F_{v_1, v_2}$$

F Distribution



Homogeneity of Variance

- Let

$$X_k = \frac{(n_k - 1)s_k^2}{\sigma_k^2} \quad \text{for} \quad k = 1, 2$$

- It follows that

$$Y = \frac{X_1/(n_1 - 1)}{X_2/(n_2 - 1)} \sim F_{n_1-1,n_2-1}$$

- Thus

$$Y = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$$

Homogeneity of Variance

- Under $H_0 : \sigma_1^2 = \sigma_2^2$,

$$\frac{s_1^2}{s_2^2} \sim F_{n_1-1,n_2-1}$$

- For $H_A : \sigma_1^2 \neq \sigma_2^2$, reject null if s_1^2/s_2^2 is very large or very small (i.e., near zero)

- Formally,

$$C_\alpha = \{f : f < F_{n_1-1,n_2-1,\alpha/2} \text{ or } f > F_{n_1-1,n_2-1,1-\alpha/2}\}$$

where $f = s_1^2/s_2^2$

Homogeneity of Variance

- Note: $F_{v_1,v_2,\alpha} = 1/F_{v_2,v_1,1-\alpha}$
- Table A.5 and A.6 of the text for two-sided $\alpha = 0.10$ and $\alpha = 0.02$; see errata
- R

```
> qf(0.975,14,14)
[1] 2.978588
> qf(0.025,14,14)
[1] 0.3357296
> 1/qf(0.025,14,14)
[1] 2.978588
```

- SAS

```
data; y = finv(0.975,14,14);
```

Homogeneity of Variance: BW Example

- $H_0 : \sigma_1^2 = \sigma_2^2;\;\; H_A : \sigma_1^2 \neq \sigma_2^2$
- For $\alpha = 0.05$,

$$C_{0.05} = \{f : f < F_{14,14,0.025} \text{ or } f > F_{14,14,0.975}\}$$

$$= \{f : f < 0.34 \text{ or } f > 2.98\}$$

- Observed test statistic

$$f = \frac{0.8994^2}{0.9605^2} = 0.8768$$

- Therefore, do not reject H_0

$$p = 2 \times F_{14,14}(0.8768) = 0.809$$

Homogeneity of Variance: BW Example cont.

- SAS

```
proc ttest;  class trt;  var bw;

The TTEST Procedure

Method           Variances        DF    t Value    Pr > |t|

Pooled           Equal            28       2.41      0.0226
Satterthwaite    Unequal       27.88       2.41      0.0226

              Equality of Variances

Method      Num DF    Den DF    F Value    Pr > F

Folded F        14        14       1.14    0.8090
```

Homogeneity of Variance: BW Example cont.

- R

```
> var.test(bw$drug,bw$placebo)

        F test to compare two variances

data:  bw$drug and bw$placebo
F = 0.8767, num df = 14, denom df = 14, p-value = 0.809
alternative hypothesis: true ratio of variances is not equal to 1
```

Testing Homogeneity of Variance

- Cf. page 133 of text
- Genuine interest in whether variances equal
- With respect to testing $H_0 : \mu_1 = \mu_2$
  - For small samples, potential for type II error
  - For large samples, CLT/Slutsky
  - Adjustment for sequential testing
- For additional reading, see Moser and Stevens (*The American Statistician* 1992)

Effect on Testing $\mu_1 = \mu_2$

- What if $\sigma_1^2 \neq \sigma_2^2$ and unknown?
- Solutions

1. Large sample approximation
2. Normality: Welch-Satterthwaite approximation (Behrens-Fisher problem)
3. Transformation
4. Nonparametric methods: Wilcoxon rank sum

Large Sample Approximation

- If n_1 and n_2 are large, homogeneity of variance assumption is not important
- Recall CLT plus Slutsky implies

$$\bar{Y} \dot{\sim} N\left(\mu, \frac{s^2}{n}\right)$$

- Thus

$$\bar{Y}_1 - \bar{Y}_2 \dot{\sim} N\left(\mu_1 - \mu_2, \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)$$

Large Sample Approximation

- Therefore, to test $H_0 : \mu_1 - \mu_2 = \delta$, we can use

$$Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Under H_0, $Z \dot{\sim} N(0, 1)$
- Approximation gets better as $n_1, n_2 \to \infty$
- Generally, require $n_j \geq 25$ for $j = 1, 2$
- Note that the assumption that the Ys are normally distributed is no longer needed either (CLT)

Large Sample Approximation: Example

- A study was done to compare the percent body fat of third graders at schools on two Native American reservations: Tohona and Apache
- $H_0 : \mu_T = \mu_A$ vs. $H_A : \mu_T \neq \mu_A$
- $n_T = 63, \;\; n_A = 35$
- $C_{0.05} = \{z : |z| > 1.96\}$
- $\bar{y}_T = 37.9\%; \;\; s_T = 8.66; \;\; \bar{y}_A = 32.8\%; \;\; s_A = 6.88;$

$$z = \frac{37.9 - 32.8}{\sqrt{\frac{8.66^2}{63} + \frac{6.88^2}{35}}} = 3.2;$$

p=0.0014

Welch-Satterthwaite Approximation

- Assume normality; n_1, n_2 small; $\sigma_1^2 \neq \sigma_2^2$
- Statistic

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{\mathrm{df}}$$

- Note 5.2 of text:

$$\mathrm{df}_{\mathrm{text}} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1+1} + \frac{(s_2^2/n_2)^2}{n_2+1}} - 2$$

Welch-Satterthwaite Approximation

- Welch (Biometrika 1938), SAS and R

$$\mathrm{df}_{\mathrm{welch}} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

- Use $\lfloor \mathrm{df} \rfloor$ if using tables

Welch-Satterthwaite Approximation: Example

- Premature birth example

$$n_1 = n_2 = 15$$

$$s_1 = 0.8994,\ s_2 = 0.9605$$

$$\mathrm{df}_{\mathrm{text}} = 29.86$$

$$\mathrm{df}_{\mathrm{welch}} = 27.88$$

Welch-Satterthwaite Approximation: R

- R

```
> t.test(bw$drug,bw$placebo,var.equal=FALSE,alternative="greater")

        Welch Two Sample t-test

data:  bw$drug and bw$placebo
t = 2.4136, df = 27.88, p-value = 0.01131

alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.2419592       Inf

sample estimates:
mean of x mean of y
     7.08      6.26
```

Summary

Normal	Var known	Var equal	N large	Test statistic
$\surd$	$\surd$			$\frac{(\bar{Y}_1-\bar{Y}_2)-\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$
	$\surd$		$\surd$	
$\surd$		$\surd$		$\frac{(\bar{Y}_1-\bar{Y}_2)-\delta}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$
			$\surd$	$\frac{(\bar{Y}_1-\bar{Y}_2)-\delta}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \sim N(0,1)$
$\surd$				$\frac{(\bar{Y}_1-\bar{Y}_2)-\delta}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \sim t_{df}$
				Transform, nonparametrics

Outline

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

Wilcoxon (Mann-Whitney) Rank Sum Test

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

$$H_0 : F_1(y) = F_2(y)$$
$$H_A : F_1(y + \Delta) = F_2(y)$$

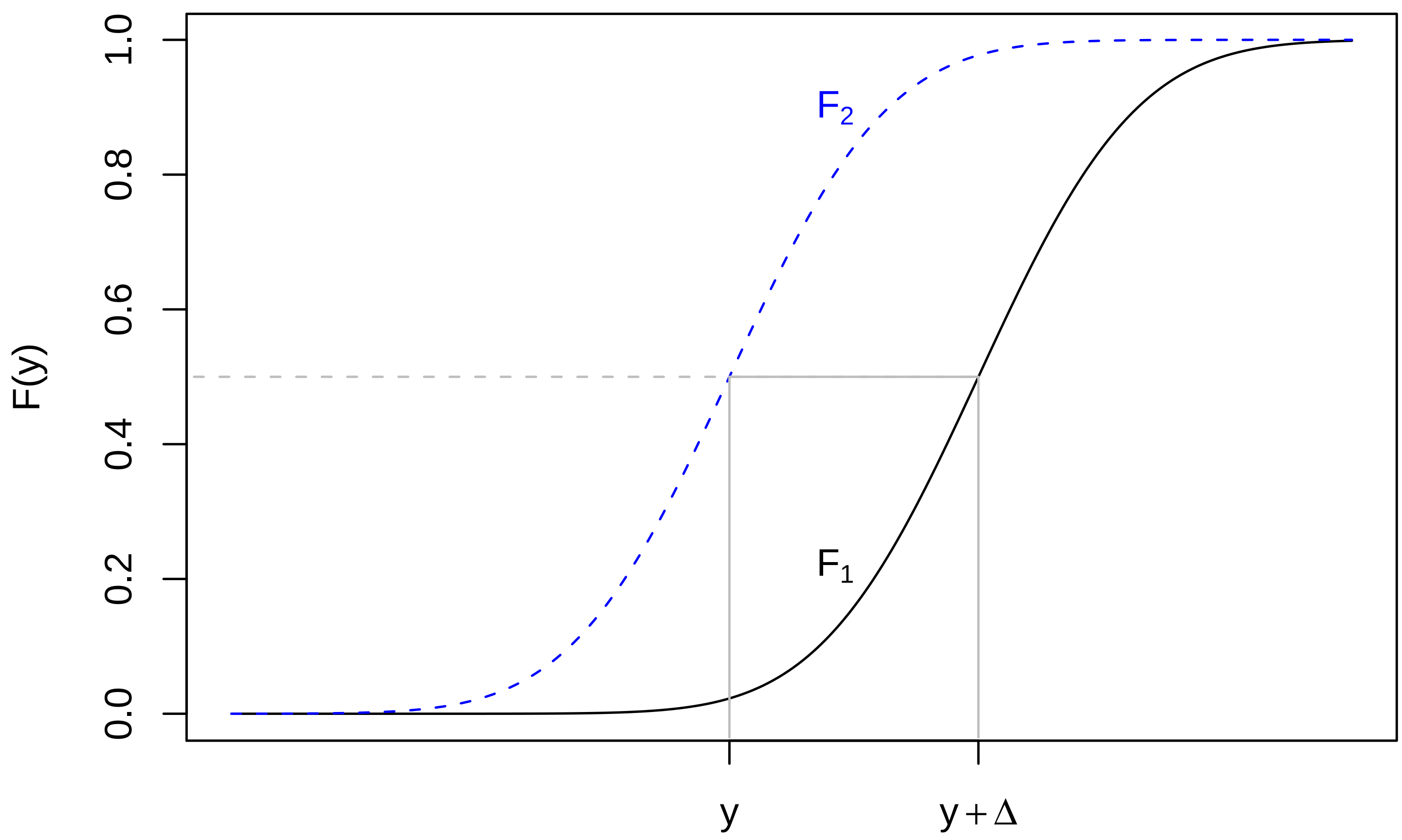
   where Δ is a non-zero constant

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 (in terms of which ranks are in group 1)
- 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 tested in humans for the first time to assess effect on $CD4^+$ T cells in patients with HIV
- 7 individuals are randomized to 2 groups:

$$\text{control} \quad (n_1 = 3) \quad \text{or drug} \quad (n_2 = 4)$$

- Endpoint is percent change in $CD4^+$ count from baseline
- Null hypothesis is that the drug has no effect

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

Wilcoxon Rank Sum Test: Example cont.

- Data: control (65, 73, 69); drug (89, 70, 92, 88)
- There are $\binom{7}{3}$ 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	$\Pr[W = w]$	$F(w)$	$\Pr[W \geq w]$
6	0.0286	0.0286	1
7	0.0286	0.0571	0.9714
8	0.0571	0.1143	0.9429
9	0.0857	0.2000	0.8857
10	0.1143	0.3143	0.8000
11	0.1143	0.4286	0.6857
12	0.1429	0.5714	0.5714
13	0.1143	0.6857	0.4286
14	0.1143	0.8000	0.3143
15	0.0857	0.8857	0.2000
16	0.0571	0.9429	0.1143
17	0.0286	0.9714	0.0571
18	0.0286	1	0.0286

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

- Note that 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 $W_1 = 1 + 2 + 4 = 7$
- So 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, under H_0

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

- Similarly

$$\mathrm{Var}(W_1) = \frac{n_1 n_2 (N+1)}{12}$$

(cf. Lehmann 1998, Example 3, p 332)

Large Sample Approximation

- If n_1 and n_2 are large

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

  will be approximately $N(0,1)$

- Approximation is good for $n_1, n_2 \geq 12$

- If there are ties

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

Wilcoxon Rank Sum Test: BW Example

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

Wilcoxon Rank Sum Test: BW Example cont.

- $H_0 : \Delta = 0; \quad H_A : \Delta > 0$
- $C_{0.05} = \{z : z > 1.645\}$
- $E(W_1) = \frac{15(31)}{2} = 232.5$
- $\text{Var}(W_1 | \text{no ties}) = \frac{15^2(31)}{12} = 581.25$

Wilcoxon Rank Sum Test: BW Example cont.

- Tie correction:

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

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

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

Wilcoxon Rank Sum Test: BW Example cont.

- $w_1 = 290.5$

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

- 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 cont.

- 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 cont.

- 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
```

- Note that `W` here is not the sum of the ranks; the slides on the Mann-Whitney test will explain what it is

Wilcoxon Rank Sum Exact P-values

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

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

where

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

- Without ties, the 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	$\|W_1 - E(W_1)\|$	Ranks	W_1	$\|W_1 - E(W_1)\|$
1,2.5,2.5	6	3	1,4,5	10	1
1,2.5,4	7.5	1.5	2.5,2.5,4	9	0
1,2.5,4	7.5	1.5	2.5,2.5,5	10	1
1,2.5,5	8.5	0.5	2.5,4,5	11.5	2.5
1,2.5,5	8.5	0.5	2.5,4,5	11.5	2.5

- Thus $|w_1 - E(W_1)| = |7.5 - 9| = 1.5$,

  giving $p = 0.5$

Wilcoxon Rank Sum Exact P-values: R

```
> # First need to install package exactRankTests
> 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 group; var y;
   exact wilcoxon;

             Wilcoxon Scores (Rank Sums) for Variable y
                    Classified by Variable group

                        Sum of      Expected       Std Dev           Mean
group        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 $H_0 : \Delta = 0$ in favor of $H_A : \Delta > 0$ if W_1 large, i.e., U_1 small

Mann-Whitney Test

- That is, the Mann-Whitney and Wilcoxon rank sum test are equivalent
- This explains the R output

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

- Reject $H_0 : \Delta = 0$ in favor of $H_A : \Delta > 0$ if W_2 small, i.e., U_2 large

Mann-Whitney Test

- Table A.10 in the text gives critical values

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

  $\alpha = 0.01$: critical value $= 169$

  $\alpha = 0.005$: critical value $= 174$

- Efficiency

  Compared to the t-test, the 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.
- For the Wilcoxon rank sum test:

$$H_0 : \Delta = 0$$

$$H_A : \Delta \neq 0$$

- If we reject H_0, we may want an estimate of Δ
- 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 the Mann-Whitney perspective, we 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 set

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

- 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 for which we would not reject H_0 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 for which we would fail to reject H_0

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
```

Rank Sum Test: BW Example in R cont.

```
> # Obtain CI by inverting test
> # Note that the original observations are to 1 decimal place

> 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 the text
- $H_0 : F_1 = F_2$
- Test statistic $D \equiv \bar{Y}_1 - \bar{Y}_2$
- Suppose N subjects randomly assigned to two groups of sizes n_1, n_2
- There are $\binom{N}{n_1}$ possible group assignments of sizes n_1, n_2 and each is equally likely under H_0
- 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 into groups of sizes 3 and 4

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 88	7.00	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	$\Pr[\bar{Y}_1 - \bar{Y}_2 \leq d]$	$\Pr[\bar{Y}_1 - \bar{Y}_2 \geq d]$	d	$\Pr[\bar{Y}_1 - \bar{Y}_2 \leq d]$	$\Pr[\bar{Y}_1 - \bar{Y}_2 \geq d]$
-17.50	0.029	1.000	6.41	0.686	0.343
-15.75	0.057	0.971	7.00	0.743	0.314
-15.17	0.086	0.943	7.58	0.771	0.257
-12.83	0.114	0.914	8.75	0.800	0.229
-7.00	0.143	0.886	9.33	0.886	0.200
-6.42	0.200	0.857	9.92	0.914	0.114
-5.83	0.229	0.800	10.67	0.943	0.086
-4.67	0.286	0.771	11.08	0.971	0.057
-4.08	0.371	0.714	20.42	1.000	0.029
-3.50	0.400	0.629			
-2.33	0.457	0.600			
-1.75	0.543	0.543			
-1.17	0.571	0.457			
0.00	0.600	0.429			
0.58	0.629	0.400			
4.67	0.657	0.371			

Permutation Test Example

- Symmetric critical region for $\alpha = 0.1$

$$C_{0.10} = \{D : D = -17.5 \text{ or } D = 20.42\}$$

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

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

Permutation Test

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

  e.g., for $N = 20$, the number of permutations is 20!, which is $> 2 \times 10^{18}$

  However, $\binom{20}{10} = 184,756$
- *Conditional test* (conditional on the observed Ys), that is, Ys fixed
- Exact: probability of rejecting the 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

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

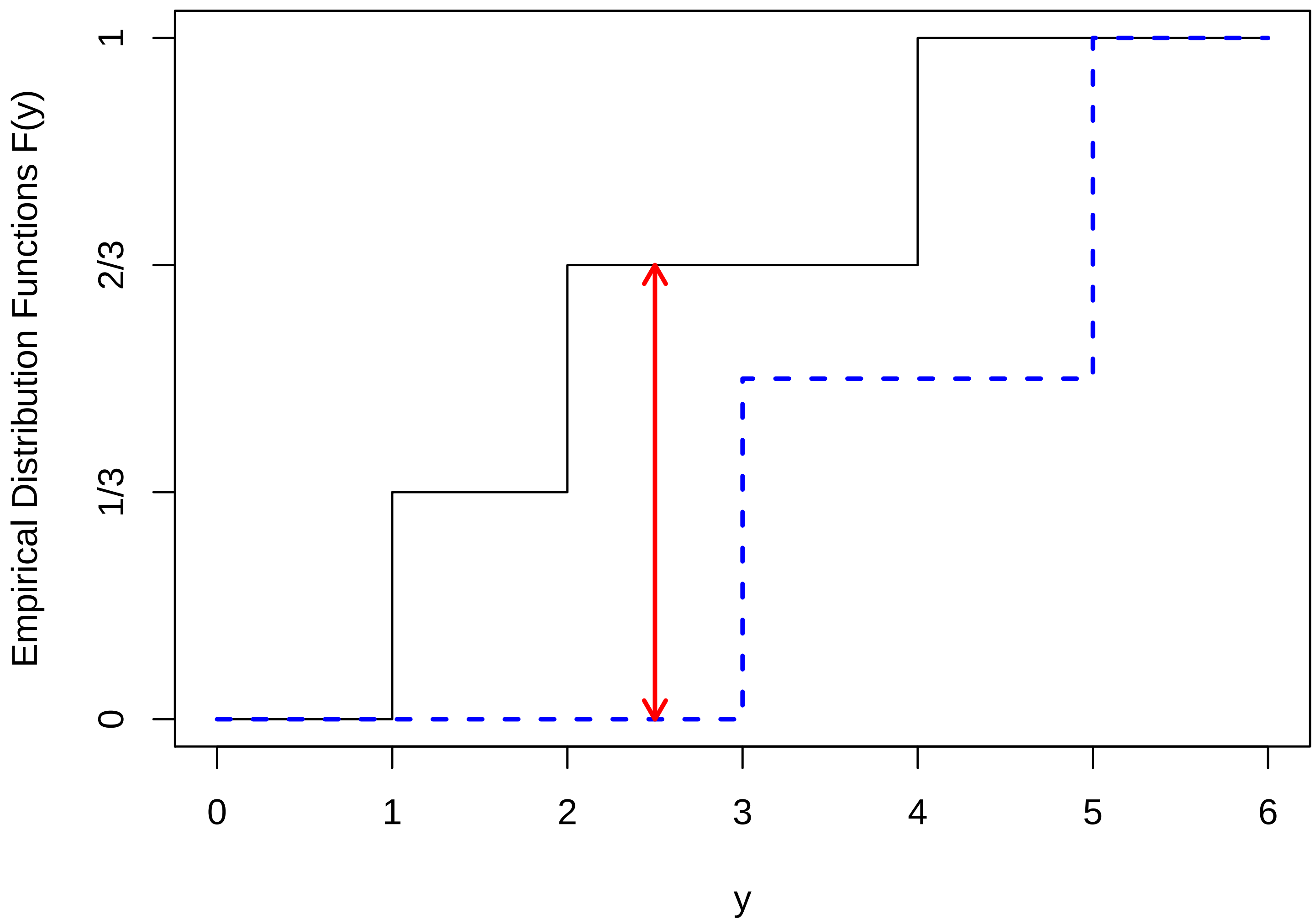
where $F_{1n}(y)$ and $F_{2m}(y)$ are the EDFs for the samples from F_1 and F_2, respectively

- Note that even though the alternative is two-sided, we reject H_0 only for large values of D

Kolmogorov-Smirnov Test

- Can be viewed as a rank test (text p. 279)
- Large sample based critical values on p. 268 of the 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 cont.

- 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 $p = \Pr[D \geq d] = 0.6$

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 group;
 var x;
 exact ks;
run;
```

KS in SAS

```
      Kolmogorov-Smirnov Test for Variable x
           Classified by Variable group

                       EDF at    Deviation from Mean
group       N       Maximum        at Maximum
--------------------------------------------------
1           3      0.666667           0.461880
2           2      0.000000          -0.565685
Total       5      0.400000

   Maximum Deviation Occurred at Observation 2
           Value of x at Maximum = 2.0

Kolmogorov-Smirnov Two-Sample Test

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

D+ = max (F1 - F2)      0.6667
Asymptotic Pr >  D+     0.3442
Exact      Pr >= D+     0.3000

D- = max (F2 - F1)      0.0000
Asymptotic Pr >  D-     1.0000
Exact      Pr >= D-     1.0000
```

Discussion

- Wilcoxon rank sum test: Default non-parametric test
- Permutation test
  - Asymptotically equivalent to t-test; thus most powerful asymptotically under normality
  - Computationally intensive because unique to each data set
  - Sensitive to outliers
- Kolmogorov-Smirnov test
  - Employ if trying to detect difference in distributions other than location shift