4 The two-sample location problem

Go to the example

Data: We obtain N=m+n observations X1,…,Xm and Y1,…,Yn

Assumptions:

A1. The observations X1,…, Xm are random sample (independent and identically distributed) from population 1. The observations Y1,…, Yn are random sample (independent and identically distributed) from population 2.

A2. The Xs and Ys are mutually independent.

A3. Populations 1 and 2 are continuous populations.

4.1 A Distribution-Free Rank Sum Test (Wilcoxon, Mann and Whitney)

Let

- F be the distribution function corresponding to population 1
- G be the distribution function corresponding to population 2

Hypothesis:

$H_0 : F(t) = G(t)$ for every t

- Note the common distribution is not specified.

The location-shift model or translation model

$G(t) = F(t - \Delta)$ for every t

Or, equivalently,

$Y =^d X + \Delta$

Where $=^d$ means "has the same distribution as"

Δ is called the location shift or treatment effect. The expected increase (decrease) due to the treatment.

H_0: $\Delta=0$ (the population means are equal; the treatment has no effect)

4.1.1 Procedure

1. Order the combined samples of N=m+n X and Y-values from the least to the greatest.
2. Let S1,…,Sn denote the ranks of Y1,…,Yn in this join ordering
3. W is the sum of the ranks assigned to the Y-values.

$$W = \sum_{j=1}^{n} S_j$$

One-sided Upper-Tail Test: To test H0 versus H1: Δ>0 at the α level of significance,
Reject H0 if $W \geq w_\alpha$; otherwise do not reject
w_α is chosen to P(type I error)=α. Use Table A.6.

One-sided Lower-Tail Test: To test H0 versus H1: Δ<0 at the α level of significance,
Reject H0 if $W \leq n(m+n+1) - w_\alpha$; otherwise do not reject
w_α is chosen to P(type I error)=α. Use Table A.6.

Two-sided Lower-Tail Test: To test H0 versus H1: $\Delta \neq 0$ at the α level of significance,
Reject H0 if $W \geq w_{\alpha/2}$ or $W \leq n(m+n+1) - w_{\alpha/2}$; otherwise do not reject
w_α is chosen to P(type I error)=α. Use Table A.6. (two-sided symmetric test)

Using Table A.6: when naming the samples, call the Y-sample the one with the smaller sample size. $n \leq m$.

4.1.2 Large sample Approximation

When H0 is true, W has the mean and variance of

$$E_0(W) = \frac{n(n+m+1)}{2}$$

$$\mathrm{var}_0(W) = \frac{mn(m+n+1)}{12}$$

The standardized version

$$W^* = \frac{W - E_0 W}{\left[\mathrm{var}_0 W\right]^{1/2}}$$

has asymptotic distribution N(0,1) as both m and n tend to infinity.

The normal approximation to the above procedures have the usual form of rejecting H0 when
$W^* \geq z_\alpha; W^* \leq -z_\alpha; \left|W^*\right| \geq z_{\alpha/2}$

4.1.3 Ties

Give tied observations the average of the ranks for which those observations are competing. Now the test is approximate rather than exact.

When there are ties, the null mean of W is unaffected but the null variance is reduced. (why?)

4.1.4 Miscellaneous

- Motivarion of the test?
- Testing Δ is equal to some specified nonzero Δ_0.
  - Use $Y_j' = Y_j - \Delta_0$.
- Number of possible outcomes? What're the smallest and the largest values of W? n(n+1)/2, to n(2m+n+1)/2. Why?
- Under H0, the distribution of W is symmetric about its mean, so $P(W \le x) = P(W \ge n(m+n+1) - x)$ for x=n(n+1)/2, …, n(2m+n+1)/2.
- Derivation of the null distribution?

4.1.5 Power results and sample size determination for the Wilcoxon test

Consider the upper tail α-level test of H0: Δ=0 vs H1: Δ>0. Suppose the true shift is Δ. If F is normal with standard deviation σ,

Power $\approx \Phi(A)$ with $A = \left(\sqrt{\frac{3mn}{(N+1)\pi}} \frac{\Delta}{\sigma} \right) - z_\alpha$.

Interpretation?

Let $\delta = P(X < Y)$. An approximate total sample size N so that the α level one-sided test will have approximate power 1-β against an alternative value δ (>1/2). With m=cN, the approximate value of N is

$$N \approx \frac{(z_\alpha + z_\beta)^2}{12c(1-c)(\delta - 1/2)^2}$$

Q. How about α=.05 test with power = 1-β = of at least .90 against an alternative where δ = .7 for m=n so that c=.5.
A. z_α=1.65 and z_β=1.28 We fine m=n=N/2 = 35.8

4.1.5.1 Robustness

The significance level of the test is not preserved if

- two populations differ in shape or dispersion

- dependencies exist among the X's or Y's.

4.2 An estimator associated with Wilcoxon's Rank Sum statistic (Hodges-Lehmann)

To estimate Δ, form the mn differences Y_j-X_i. Then estimate
$\hat{\Delta} = median\{(Y_j - X_i), i = 1,...,m; j = 1,...,n\}$.
Let $U^{(1)}$ <=…,$U^{(mn)}$ denote the ordered values. Then compute their median.

Motivation: Estimate Δ by the amount $\hat{\Delta}$ that the Y sample should be shifted in order that $X_1,..., X_m$ and $Y_1 - \hat{\Delta},...,Y_n - \hat{\Delta}$ appear (when 'viewed' by the rank sum statistic W) as two samples from the same population.
Sensitivity to Gross Errors : $\hat{\Delta}$ is less sensitive to gross errors than its normal theory analog $\overline{Y} - \overline{X}$.

P(X_1<Y_1): (Alternative parameter) $\delta = P(X_1 < Y_1)$, the probability that a single Y observation will be larger than a single X observation. Such parameter can make more sense than Δ.
E.g. P(X<Y)=.76 can make more sense than "$\frac{\mu_2 - \mu_1}{\sigma} = 1$" when X and Y are response to two different treatments.

It is estimated by
$\hat{\delta} = \frac{U}{mn}$ where $U = \sum_{i=1}^{m}\sum_{j=1}^{n}\phi(X_i, Y_j) = \sum_{i=1}^{m}\sum_{j=1}^{n}1(X_i < Y_j)$ is the Mann-Whitney U statistic. The relationship

$$W = U + \frac{n(n+1)}{2}$$

holds. (tests based on U are equivalent to tests based on W! why?) Distribution-free confidence bounds for them are avialabel (p.129 of the text) but are rather complicated.

The estimator $\hat{\Delta}$ is asymptotically normal and efficient.

4.3 Distribution free confidence interval based on Wilcoxon's Rank Sum Test

For (1-α) confidence interval, set

$$C_\alpha = \frac{n(2m+n+1)}{2} + 1 - w_{\alpha/2}$$

and compute

$$\left(\Delta_L, \Delta_U\right) = \left(U^{(C_\alpha)}, U^{(mn+1-C_\alpha)},\right).$$

Picture?

Do NOT confuse between ranks and values!

4.3.1 Large sample approximation

For large m and n, we approximate $C\alpha$ by

$$C_\alpha \approx \frac{mn}{2} - z_{\alpha/2}\left\{\frac{mn(m+n+1)}{12}\right\}^{1/2}$$

Again, be conservative.

Example?

Confidence interval consists of those Δ_0 values for which the two-sided α-level test of $\Delta=\Delta_0$ accepts the null hypothesis.

4.3.2 Confidence Bounds:

Let $C_\alpha^* = \frac{n(2m+n+1)}{2} + 1 - w_\alpha$.

Lower confidence bound for Δ with confidence coefficient 1-α is given by

$$\left(\Delta_L^*, \infty\right) = \left(U^{(C_\alpha^*)}, \infty\right).$$

Upper confidence bound for Δ with confidence coefficient 1-α is given by

$$\left(-\infty, \Delta_U^*\right) = \left(-\infty, U^{(mn+1-C_\alpha^*)}\right).$$

Large sample approximation to them is straightforward.

$$C_\alpha^* \approx \frac{mn}{2} - z_\alpha\left\{\frac{mn(m+n+1)}{12}\right\}^{1/2}$$

4.4 Example (Permeability Constant)

Average pearmeability constant (in 10^{-4} cm/s) for six measurements on each of 15 chorioamnion (a placental membrane) tissues, obtained from 10 term pregrancies (X) and 5 terminated pregrancies (Y)

```
X Y
0.80 1.15
0.83 0.88
1.89 0.90
1.04 0.74
```

```
1.45 1.21
1.38
1.91
1.64
0.73
1.46
```

How would you analyze this data?
What would you do first?

* Draw boxplot. (boxplot(x,y); boxplot X Y; Overlay.)
* Draw normal Q-Q plot. (qqnorm(x), qqnorm(y) vs pplot)

4.4.1.1 Testing

Want to test for H1: $\Delta<0$.
Test at $\alpha=.082$.
What's W? What's the P-value?
Large sample approximation? (P-value)

```
rank(c(x,y))
 [1]  3  4 14  7 11 10 15 13  1 12  8  5  6  2  9
```

Y ranks are 2, 5,6,8,9
W=30
Reject H0 if W<= 28
P(W<=30) = P(W>=50) = .127.
N=5, m=10, x=50.
Large sample approximation W* = -1.225
P-value = .11

4.4.1.2 Estimation

Ordered Yj-Xi values:

```
> outer(y,x,"-")
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8] [,9] [,10]
[1,]  0.35  0.32 -0.74  0.11 -0.30 -0.23 -0.76 -0.49 0.42 -0.31
[2,]  0.08  0.05 -1.01 -0.16 -0.57 -0.50 -1.03 -0.76 0.15 -0.58
[3,]  0.10  0.07 -0.99 -0.14 -0.55 -0.48 -1.01 -0.74 0.17 -0.56
[4,] -0.06 -0.09 -1.15 -0.30 -0.71 -0.64 -1.17 -0.90 0.01 -0.72
[5,]  0.41  0.38 -0.68  0.17 -0.24 -0.17 -0.70 -0.43 0.48 -0.25
>
matrix(sort(c(tmp)), byrow=TRUE, ncol=10)
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9] [,10]
[1,] -1.17 -1.15 -1.03 -1.01 -1.01 -0.99 -0.90 -0.76 -0.76 -0.74
[2,] -0.74 -0.72 -0.71 -0.70 -0.68 -0.64 -0.58 -0.57 -0.56 -0.55
[3,] -0.50 -0.49 -0.48 -0.43 -0.31 -0.30 -0.30 -0.25 -0.24 -0.23
[4,] -0.17 -0.16 -0.14 -0.09 -0.06  0.01  0.05  0.07  0.08  0.10
```

```
[5,]  0.11  0.15  0.17  0.17  0.32  0.35  0.38  0.41  0.42  0.48
>
```

$\hat{\Delta}$=Median(Y_j-X_i) = (-.31-.30)/2 = -.305

4.4.1.3 Confidence Interval

Obtain 96% CI for Δ.
m=10, n=5, $w_{\alpha/2}$ = $w_{.02}$ = 57.
$C_{.04}$ = 9.
$\left(\Delta_L, \Delta_U\right) = \left(U^{(C_\alpha)}, U^{(mn+1-C_\alpha)},\right)$=($U^{(9)}$, $U^{(42)}$) = (-.76, .15)

Large sample approximation:
$C_{.04}$ ~ 8.3. Set it to 8.

4.4.1.4 Using R to answer the above

```
x <- c(0.80, 0.83,1.89,1.04,1.45,1.38,1.91,1.64,0.73,1.46)
y <- c(1.15,0.88,0.90,0.74,1.21)
or read.table

>wilcox.test(x,y, alternative='greater')

        Wilcoxon rank sum test

data:  x and y
W = 35, p-value = 0.1272
alternative hypothesis: true mu is greater than 0

> wilcox.test(x,y, alternative='greater', exact=FALSE, correct=FALSE)

        Wilcoxon rank sum test

data:  x and y
W = 35, p-value = 0.1103
alternative hypothesis: true mu is greater than 0

> wilcox.test(y,x, conf.int=TRUE)

       Wilcoxon rank sum test

data:  y and x
W = 15, p-value = 0.2544
alternative hypothesis: true mu is not equal to 0
95 percent confidence interval:
 -0.76  0.15
sample estimates:
difference in location
                -0.305

> wilcox.test(y,x, conf.int=TRUE, conf.level=.96,
```

```
+               exact=FALSE, correct=FALSE)

        Wilcoxon rank sum test

data:  y and x
W = 15, p-value = 0.2207
alternative hypothesis: true mu is not equal to 0
96 percent confidence interval:
 -0.7599419  0.1500816
sample estimates:
difference in location
            -0.3038673
> >
```

4.4.1.5 Using Minitab to answer the above

```
MTB > mann c1 c2;
SUBC> alt 1.

Mann-Whitney Test and CI: X, Y

    N  Median
X  10  1.4150
Y   5  0.9000


Point estimate for ETA1-ETA2 is 0.3050
95.7 Percent CI for ETA1-ETA2 is (-0.1499,0.7602)
W = 90.0
Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.1223
```

4.5 A robust rank test for the Behrens-Fisher Problem (Graduate)

Let X1,…,Xm and Y1,…Yn be

- independent random samples
- from continuous distribuitons
- that are symmetric about the population medians θx, θy, respectively.

We don't assume

- the two populations have the same form
- variances of the two populations are equal

We are interested in H0': $\theta_x = \theta_y$ versus one sided or two sided alternatives.

5 The two-sample dispersion problem

Data: we obtain N=m+n observations X1,…,Xm and Y1,…,Yn.

Assumptions:

A1. Within group independence and iid from continuous populations

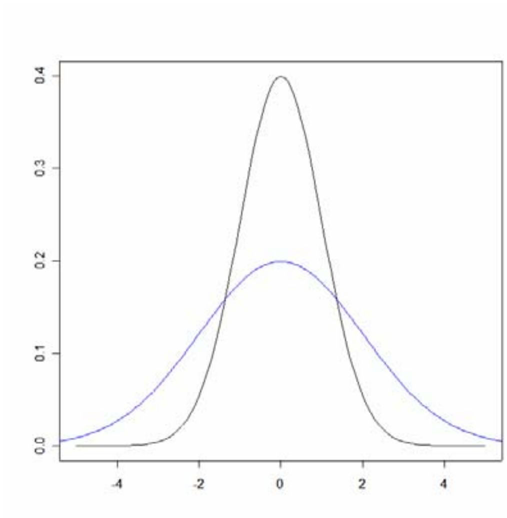
A2. Between group independence

The null hypothesis is $H_0 : F(t) = G(t)$ for every t

Location-scale parameter model:
$F(t) = H\left(\dfrac{t-\theta_1}{\eta_1}\right)$ and $G(t) = H\left(\dfrac{t-\theta_2}{\eta_2}\right)$ for every t, where H(u) is the continuous distribution function with median 0. In other words:
$\dfrac{X-\theta_1}{\eta_1} =^d \dfrac{Y-\theta_2}{\eta_2}$.

Re-write things in terms of $\gamma^2 = \left(\dfrac{\eta_1}{\eta_2}\right)^2 = \left(\dfrac{\mathrm{var}(X)}{\mathrm{var}(Y)}\right)^2$

5.1 Distribution-free rank test for dispersion- Median equal (Ansari-Bradley)



Further assume their medians are equal, i.e.,
A3. $\theta_1 = \theta_2$.

Ansari-Bradley two-sample scale statistic C is computed as follows:
Order the combined sample of N=(m+n) X an dY values from least to greatest.
Assign score i to both i'th smallest and largest observations in the combined sample.
Let R_j denote the score assigned to Y_j.

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

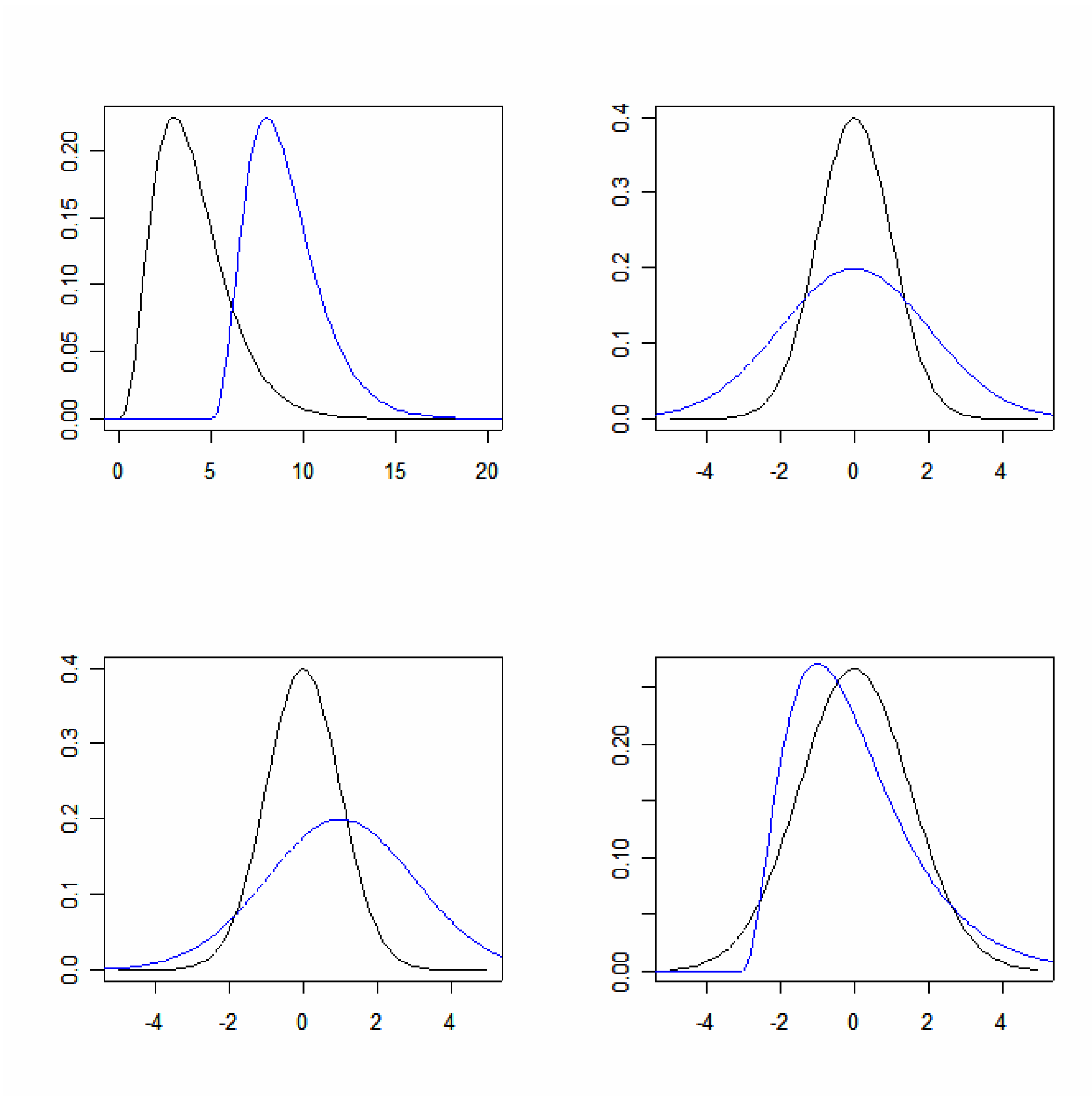
To test H_0:γ^2=1 vs H_1:γ^2>1, H_2:γ^2<1 or H_0:$\gamma^2 \neq 1$ at the α level of significance,
Reject H0 if $C \geq c_\alpha$, if $\leq c_\alpha - 1$, or if $C \geq c_{\alpha 1}$, or $\leq c_{\alpha 2}-1$, respectively, where, $\alpha_1+\alpha_2=\alpha$.
Use Table 8 to obtain the values.

What does this generalize?

```
## Hollander & Wolfe (1973, p. 86f):
## Serum iron determination using Hyland control sera
ramsay <- c(111, 107, 100, 99, 102, 106, 109, 108, 104, 99,
            101, 96, 97, 102, 107, 113, 116, 113, 110, 98)
jung.parekh <- c(107, 108, 106, 98, 105, 103, 110, 105, 104,
            100, 96, 108, 103, 104, 114, 114, 113, 108, 106, 99)
boxplot(ramsay, jung.parekh)
> ansari.test(ramsay, jung.parekh)

       Ansari-Bradley test

data:  ramsay and jung.parekh
AB = 185.5, p-value = 0.1815
alternative hypothesis: true ratio of scales is not equal to 1

> ansari.test(ramsay, jung.parekh, alternative='greater')

       Ansari-Bradley test

data:  ramsay and jung.parekh
AB = 185.5, p-value = 0.09073
alternative hypothesis: true ratio of scales is greater than 1
```

Usual large-sample approximation etc.

5.2 Distribution-free rank test for dispersion based on Jackknife- Medians not necessarily equal



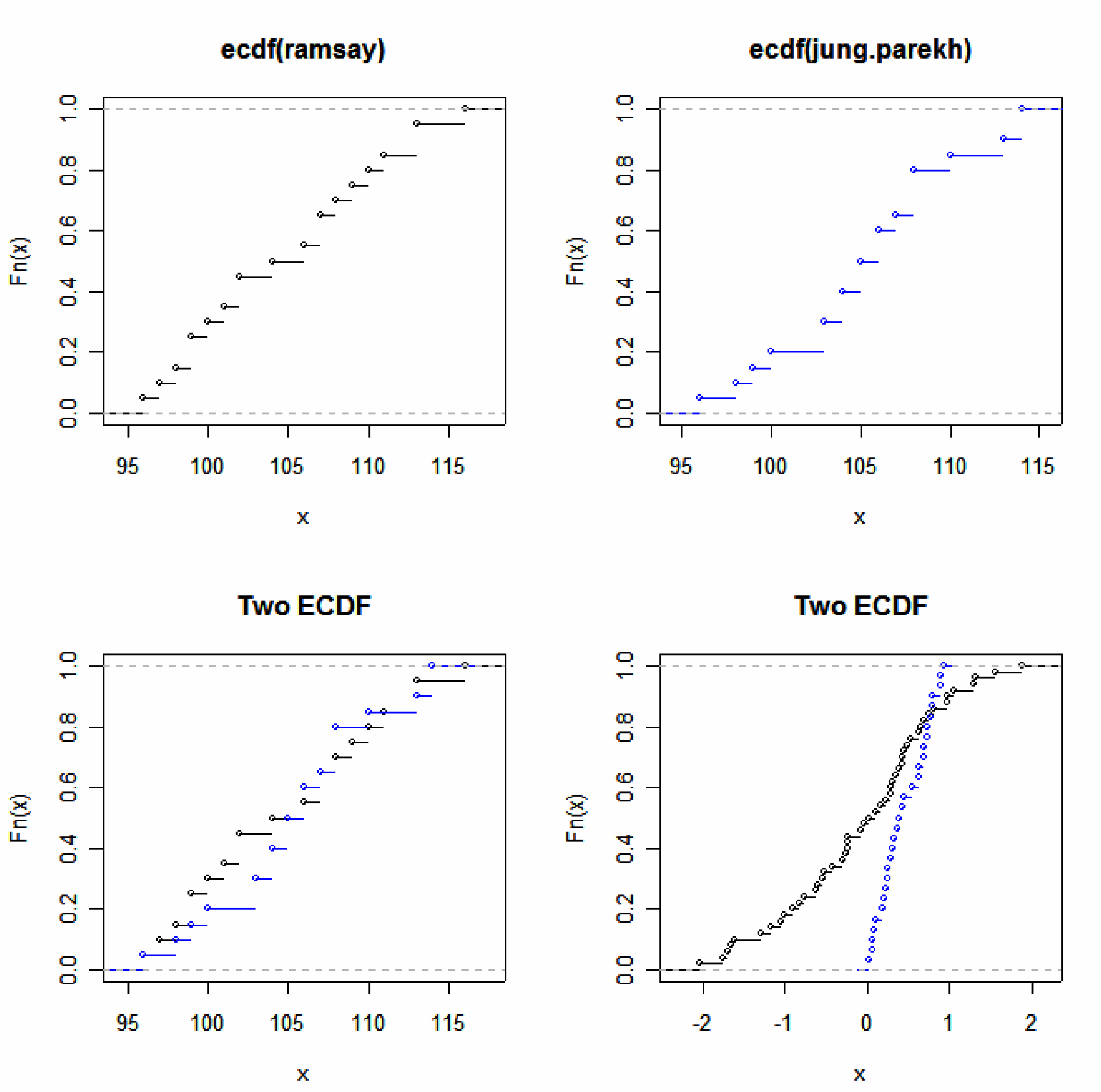
5.3 Distribution-free rank test for either location or dispersion (Lepage)

5.4 Distribution-free test for general differences in two populations (Kolmogorov-Smirnov)

$X_1,\ldots,X_m \sim iid, F; Y_1,\ldots,Y_n \sim iid, G$, continuous.

H_1: There are any differences whatsoever between the X and Y probability distributions.
$H_1 : F(t) \neq G(t)$ for at least one t .

$J = \dfrac{mn}{d} \max_t |F_m(t) - G_n(t)|$ where F_m and G_n are empirical distribution functions for the two samples.



Procedure: Reject H0 if $J \geq j_\alpha$.

```
ks.test(ramsay, jung.parekh)
```

```
       Two-sample Kolmogorov-Smirnov test

data:  ramsay and jung.parekh
D = 0.25, p-value = 0.5596
alternative hypothesis: two.sided


ks.test(x,y)

       Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.5, p-value = 9.065e-05
alternative hypothesis: two.sided
```

* The test is usually too general to have much power. (Trying to detect all alternatives) Not sensitive enough for many purposes.