STA 104 Applied Nonparametric Statistics

Chapter 3: Two-Sample Methods

Xiner Zhou

Department of Statistics, University of California, Davis

In this chapter the data consist of two random samples, a sample from the control population and an independent sample from the treatment population.

On the basis of these samples, we wish to investigate the presence of a treatment effect that results in a shift of location.

Table of contents

Two-Sample Permutation Test

We begin with a simple example. Suppose UCD is trying to decide whether to augment its traditional classroom instruction with hybrid mode. Seven students are selected for a trial. Four are randomly assigned to the new method of hybrid instruction, and the other three are given the traditional instruction. A test is given afterward to compare the two methods.

New Method	Smith (37), Lin (49), Neal (55), Zedillo (57)
Traditional Method	Johnson (23), Green (31), Zook (46)

A two-sample t-test:

A two-sample t-test of the null hypothesis of no difference between the two methods versus the one-sided alternative hypothesis that the mean of the new method is greater than the mean of the traditional method gives $t = 2.08$ and $p = .046$.

Thus, we conclude that the new method produces a significantly higher mean test score than the traditional method at the 5% level of significance.

A two-sample t-test:

The application of the t-test comes at a price by requiring assumptions:

- the observations are independent;
- the populations have normal distributions $\Rightarrow$ null distribution of test statsitics
- the variances of the two populations are the same.

In this case, there is no guarantee that the assumptions of the t-test are met.

If we apply the t-test anyway, we run the risk of misstating the p-value of the statistical test and therefore of declaring a result to be statistically significant when it is not.

Intuition of permutation test

Key question: What is null distribution without any parametric assumptions like t-test?

Intuition of permutation test

If there is no difference between the two methods, then all data sets obtained by randomly assigning four of these scores to the new method and the other three to the traditional method would have an equal chance of being observed in the study. There are

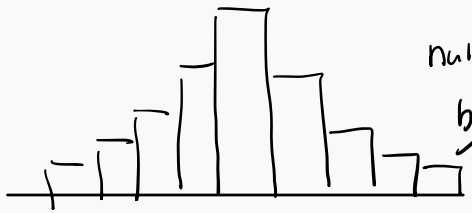
$$\binom{7}{4} = \frac{7!}{4!3!} = 35$$ permutations

such two-sample data sets. Most data sets among the 35 have both large and small scores assigned to each treatment. These are the types of data sets we would expect to observe if the two treatments were not different (equally effective).

Any appropriate statistic that quantifies the difference between groups can work:

t-test statistic

	New Method	Traditional Method	Difference between means
1	46 49 55 57	23 31 37	21.4
2*	37 49 55 57	23 31 46	16.2
3	37 46 55 57	23 31 49	14.4
4	37 46 49 57	23 31 55	10.9
5	37 46 49 55	23 31 57	9.8
6	31 49 55 57	23 37 46	12.7
7	31 46 55 57	23 37 49	10.9
8	31 46 49 57	23 37 55	7.4
9	31 46 49 55	23 37 57	6.3
10	31 37 55 57	23 46 49	5.7
11	31 37 49 57	23 46 55	2.2
12	31 37 49 55	23 46 57	1.0
13	31 37 46 57	23 49 55	0.4
14	31 37 46 55	23 49 57	-0.8
15	31 37 46 49	23 55 57	-4.3
16	23 49 55 57	31 37 46	8.0
17	23 46 55 57	31 37 49	6.3
18	23 46 49 57	31 37 55	2.8
19	23 46 49 55	31 37 57	1.6
20	23 37 55 57	31 46 49	1.0
21	23 37 49 57	31 46 55	-2.5
22	23 37 49 55	31 46 57	-3.7
23	23 37 46 57	31 37 55	-4.3
24	23 37 46 55	31 49 57	-5.4
25	23 37 46 49	31 55 57	-8.9
26	23 31 55 57	37 46 49	-2.5
27	23 31 49 57	37 46 55	-6.0
28	23 31 49 55	37 46 57	-7.2
29	23 31 46 57	37 49 55	-7.8
30	23 31 46 55	37 49 57	-8.9
31	23 31 46 49	37 55 57	-12.4
32	23 31 37 57	46 49 55	-13.0
33	23 31 37 55	46 49 57	-14.2
34	23 31 37 49	46 55 57	-17.7
35	23 31 37 46	49 55 57	-19.4



null distribution by permutation

The procedure we have just described is called a two-sample permutation test, since it is based on permuting the observations among two groups in the original sample.

The distribution of the 35 differences of means is called the permutation distribution for the difference between two means.

- The permutation principle states that the permutation distribution is an appropriate reference distribution for determining the p-value of a test and deciding whether or not a statistical test is statistically significant.
- Under nonparametric permutation approach, the researcher is free to choose a statistic that he or she feels best describes the difference between the two groups and then use the permutation approach to determine whether or not the statistic is significant.

Prior to the computer age, the practical use of permutations tests was limited by prohibitive computations.

> For instance, if two treatments each have 8 observations, then the number of possible two-sample data sets that can be obtained by permuting the 16 observations, 8 to a treatment, is
>
> $$\left(\begin{array}{c} 16 \\ 8 \end{array} \right) = 12,870$$
>
> If just two more observations are added to each treatment, then this number increases more than tenfold to
>
> $$\left(\begin{array}{c} 20 \\ 10 \end{array} \right) = 184,756$$

Simple way to obtain an approximate permutation distribution:

> Rather than using all the permutations, we take a random sample of the permutations and perform the steps involved in a permutation test on the randomly sampled permutations, say 1000.

Procedure

Summary of Steps Used in a Two-Sample Permutation Test

- Permute the $m + n$ observations between the two treatments so that there are m observations for treatment 1 and n observations for treatment 2 . Obtain all possible permutations. The number of possibilities are

$$\begin{pmatrix} m+n \\ m \end{pmatrix} = \frac{(m+n)!}{m!n!}$$

- For each permutation of the data, compute the test statistics. The permutation distribution of the test statistics characterizes the null distribution if the null hypothesis is true.
- Use the permutation distribution to calculate p-value or critical value. For example, upper-tail test

$$P_{\text{upper tail}} = \frac{\text{number of } D \text{ 's } \geq D_{\text{obs}}}{\begin{pmatrix} m+n \\ m \end{pmatrix}}$$

  For a two-sided test, perform a similar procedure on the absolute values

$$P_{\text{two wail}} = \frac{\text{number of } \mid D' \text{ 's} \mid \geq \mid D_{\text{obs}} \mid}{\begin{pmatrix} m+n \\ m \end{pmatrix}}$$

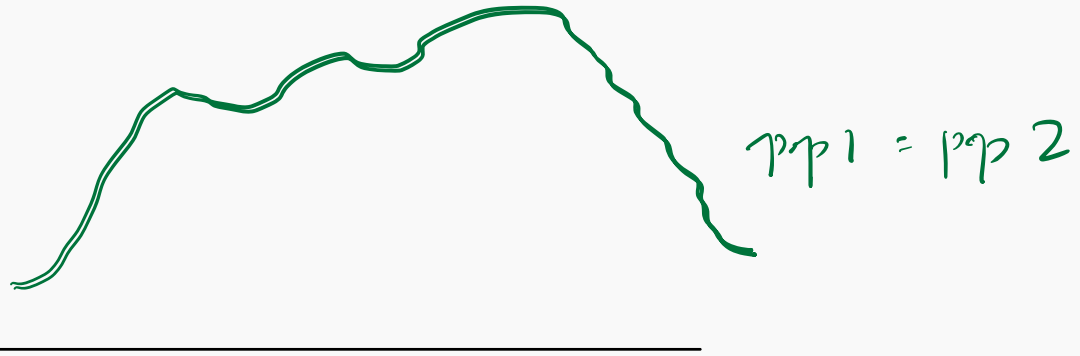
- If a predetermined level of significance has been set, declare the test to be statistically significant if the p-value is less than or equal to this level.

Wilcoxon Rank-Sum Test

Setting

We obtain $N = m + n$ observations $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$.

- The observations $X_1, \ldots, X_m$ are a random sample from population 1; that is, the X 's are independent and identically distributed. The observations $Y_1, \ldots, Y_n$ are a random sample from population 2; that is, the Y 's are independent and identically distributed.
- The X 's and Y 's are mutually independent. Thus, in addition to assumptions of independence within each sample, we also assume independence between the two samples.
- Populations 1 and 2 have continuous distribution.

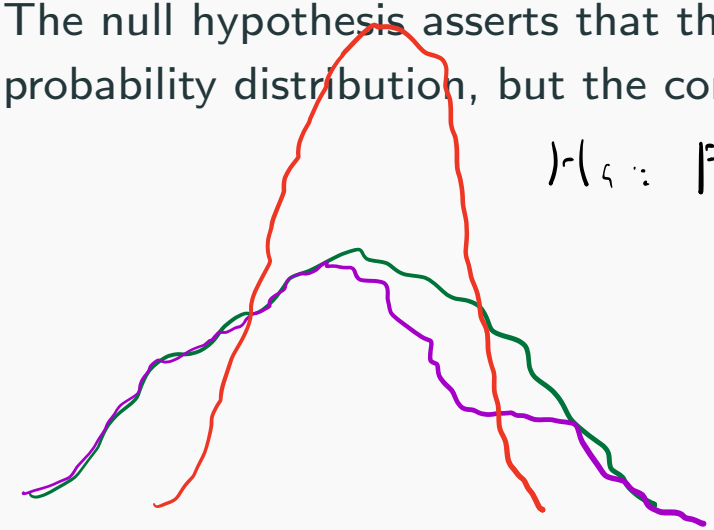


Let F be the distribution function corresponding to population 1 and let G be the distribution function corresponding to population 2.

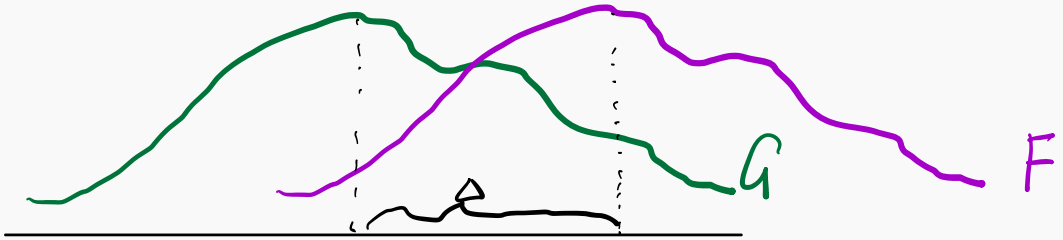
The null hypothesis is

$$H_0 : F(t) = G(t), \qquad \text{for every } t.$$

The null hypothesis asserts that the X variable and the Y variable have the same probability distribution, but the common distribution is not specified.



Two-sample location problem



The alternative hypothesis in a two-sample location problem typically specifies that Y tends to be larger (or smaller) than X. One model that is useful to describe such alternatives is the translation model, also called the location-shift model. The location-shift model is

$$G(t) = F(t - \Delta), \quad \text{for every } t.$$

says that population 2 is the same as population 1 except it is shifted by the amount Δ. Another way of writing this is

$$Y \stackrel{d}{=} X + \Delta$$

where the symbol $\stackrel{d}{=}$ means "has the same distribution as."

The parameter Δ is called the location shift. It is also known as the treatment effect. If X is a randomly selected value from population 1, the control population, and Y is a randomly selected value from population 2 , the treatment population, then Δ is the expected effect due to the treatment. [1]

If Δ is positive, it is the expected increase due to the treatment, and if Δ is negative, it is the expected decrease due to the treatment:

$$\underbrace{\Delta = E(Y) - E(X)}_{\text{Difference in population means}}$$

[1] Athough we find it convenient to use the "treatment" and "control" terminology, many situations will arise in which we want to compare two random samples, neither one of which can be described as a sample from a control population.

Hypothesis

In terms of the location-shift model, the null hypothesis H_0 reduces to

$$H_0 : \Delta = 0, \quad \Leftarrow \quad H_0 : G(t) = F(t)$$

the hypothesis that asserts the population means are equal or, equivalently, that the treatment has no effect.

Two-Sided Test:

$$H_0 : \Delta = 0 \text{ versus } H_a : \Delta \neq 0$$

One-Sided Upper-Tail Test:

$$H_0 : \Delta = 0 \text{ versus } H_a : \Delta > 0$$

One-Sided Lower-Tail Test:

$$H_0 : \Delta = 0 \text{ versus } H_a : \Delta < 0$$

Motivation

$$W = \sum_{i=1}^{n} R(Y_i)$$

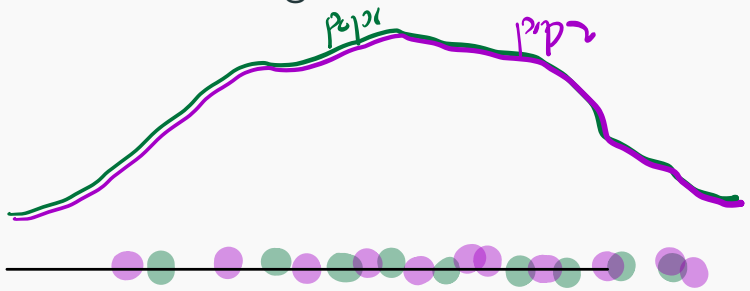
- if $\Delta = 0$:

  Ys will tend to be $\approx$ than Xs

  $R(Y_i)$s will tend to be take $\approx$ values among the ranks $1 \ldots n + m$

  W will tend to be neither too large nor too small

  This suggests neither too large nor too small W in favor of null $\Delta = 0$



$\Rightarrow$ Xs and Ys similar "spread" $\Rightarrow$ similar ranks

$\Rightarrow$ pool two-samples $\{X_1 \ldots X_m, Y_1 \ldots Y_n\}$

rank the pooled sample from 1 ... ... $n+m$

$\Rightarrow$ look at ranks of Ys, sum up ranks of Ys

Motivation

$$W = \sum_{i=1}^{n} R(Y_i)$$

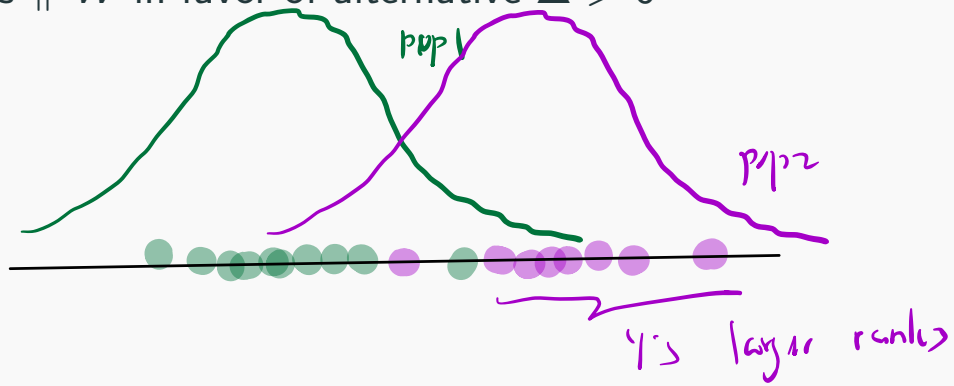
- if $\Delta > 0$:

  Ys will tend to be $\Uparrow$ than Xs

  $R(Y_i)$s will tend to be take $\Uparrow$ values among the ranks $1 \ldots n + m$

  W will tend to be $\Uparrow$

  This suggests $\Uparrow$ W in favor of alternative $\Delta > 0$



Motivation

$$W = \sum_{i=1}^{n} R(Y_i)$$

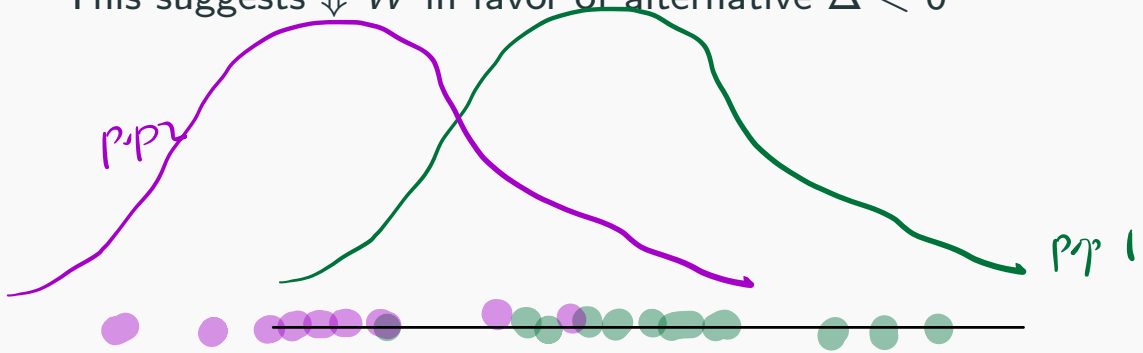
- if $\Delta < 0$:

  Ys will tend to be $\Downarrow$ than Xs

  $R(Y_i)$s will tend to be take $\Downarrow$ values among the ranks $1 \ldots n + m$

  W will tend to be $\Downarrow$

  This suggests $\Downarrow$ W in favor of alternative $\Delta < 0$



Derivation of null distribution using permutation

When H_0 is true: There are

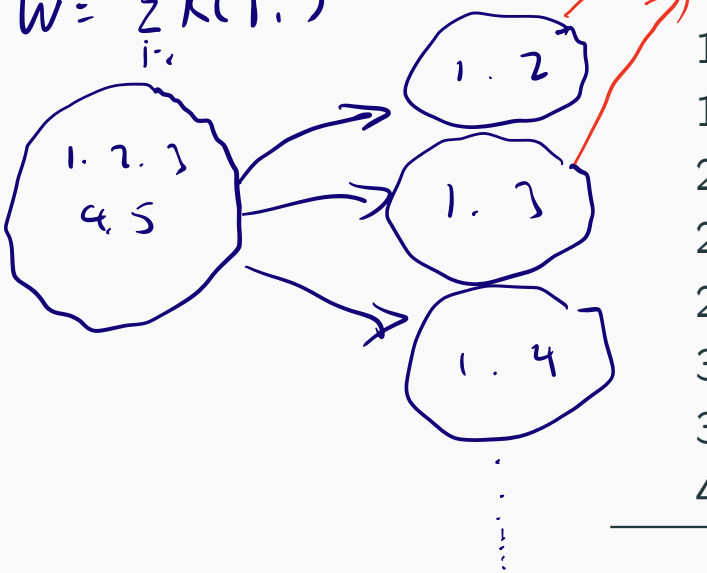
$$\binom{n+m}{n} = \frac{(n+m)!}{n!\,m!}$$

possible assignments for $Y-$ranks (equal likely).

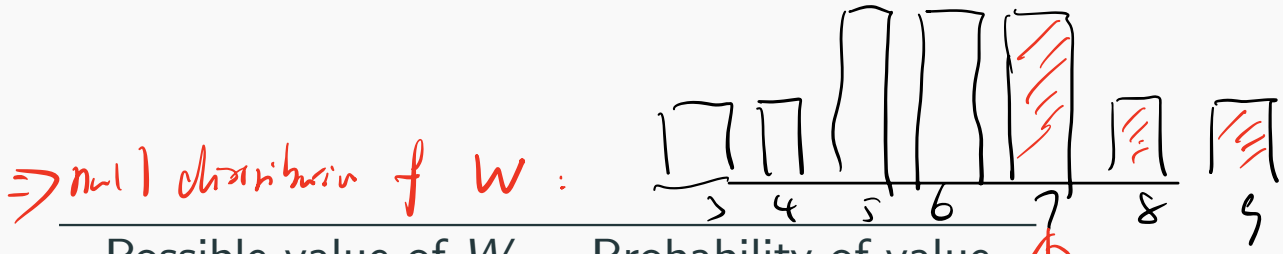
$m = 3, n = 2$

$\binom{3+2}{2} = \binom{5}{2} = \frac{5!}{2!\,3!} = 10$

$W = \sum_{i=1}^{n} R(1;i)$



Y-ranks	Probability	W
1, 2	$\frac{1}{10}$	3
1, 3	$\frac{1}{10}$	4
1, 4	$\frac{1}{10}$	5
1, 5	$\frac{1}{10}$	6
2, 3	$\frac{1}{10}$	5
2, 4	$\frac{1}{10}$	6
2, 5	$\frac{1}{10}$	7
3, 4	$\frac{1}{10}$	7
3, 5	$\frac{1}{10}$	8
4, 5	$\frac{1}{10}$	9



Possible value of W	Probability of value
3	.1
4	.1
5	.2
6	.2
7	.2
8	.1
9	.1

Thus, for example, under H_0, the p-value for an upper-tail test with observed $W = 7$ is the probability that W is greater than or equal to 7

$$\begin{aligned} P_0(W \geq 7) &= P_0(W = 7) + P_0(W = 8) + P_0(W = 9) \\ &= .2 + .1 + .1 = .4 \end{aligned}$$

Large sample approximation of null distribution

$$\frac{W}{n} = \frac{1}{n}\sum_{i=1}^{n} R(Y_i)$$

We want to know the behavior of: Under H_0, sample mean of a random sample of size n drawn without replacement from finite population $\{1 \ldots N = n + m\}$ [2]

[2] Facts from finite population theory:

- The mean is equal to the mean μ_{pop} of the finite population.
- The variance is equal to

$$\frac{\sigma^2_{\text{pop}}}{n} \times \frac{N-n}{N-1},$$

where σ^2_{pop} denotes the variance of the finite population and the factor $(N-n)/(N-1)$ is the finite population correction factor.

Large sample approximation of null distribution

Optional:

For the finite population $\{1, 2, \ldots, N\}$, direct calculations establish

$$\mu_{\text{pop}} = \frac{1 + 2 + \cdots + N}{N} = \frac{N+1}{2}$$

$$\sigma^2_{\text{pop}} = \frac{1}{N}\left\{1^2 + 2^2 + \cdots + N^2\right\} - \left(\frac{N+1}{2}\right)^2 = \frac{(N-1)(N+1)}{12}$$

$$\Rightarrow E\left(\frac{W}{n}\right) = \frac{N+1}{2}$$

$$\Rightarrow \text{var}\left(\frac{W}{n}\right) = \frac{(N-1)(N+1)}{12n} \times \frac{N-n}{N-1} = \frac{m(N+1)}{12n}$$

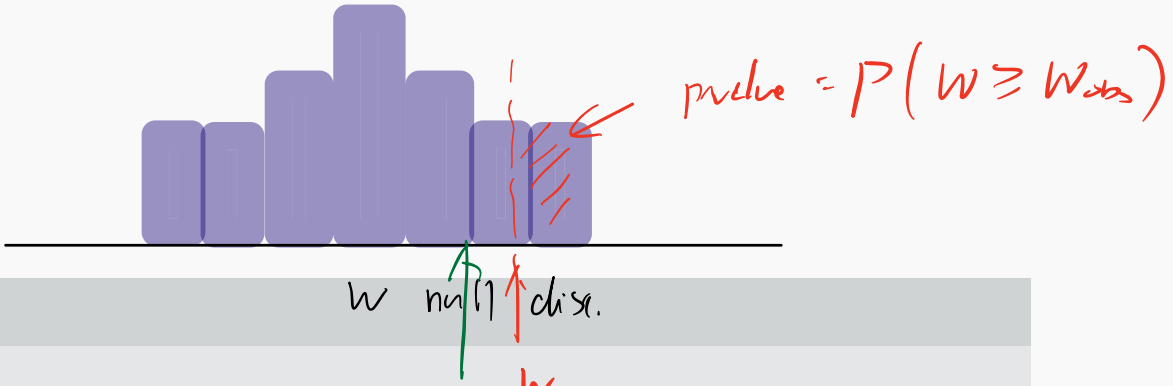
$$\Rightarrow EW = \frac{n(N+1)}{2}$$

$$\Rightarrow \text{var}(W) = \frac{\text{mn}(N+1)}{12}$$

$$\Rightarrow W^* = \frac{W - \frac{n(N+1)}{2}}{\sqrt{\frac{mn(N+1)}{12}}} = \frac{W - E(W)}{\sigma(W)} \sim N(0,1)$$

Asymptotic normality follows from standard theory for the mean of a sample.

Procedure



a. One-Sided Upper-Tail Test.

To test

$$H_0 : \Delta = 0$$

versus

$$H_a : \Delta > 0$$

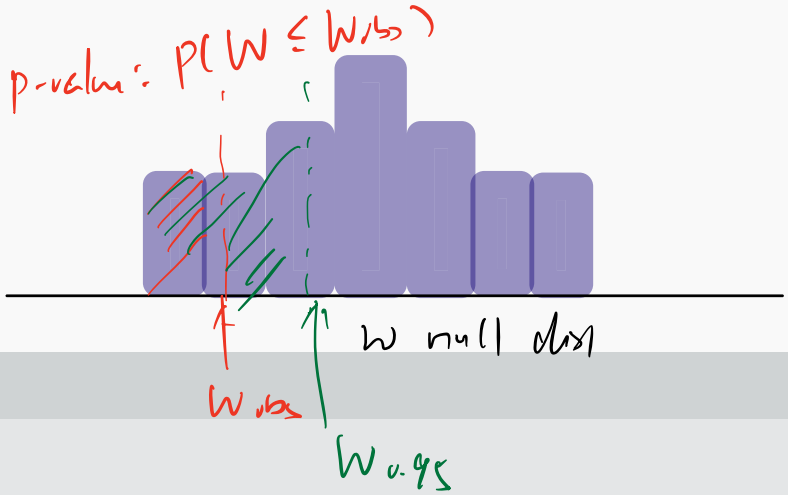
at the α level of significance,

Reject H_0 if $W \geq w_\alpha$; otherwise do not reject, where the constant w_α is chosen to make the type I error probability equal to α. (Or use p-value)

Large-samaple approximation

Reject H_0 if $W^* \geq z_\alpha$; otherwise do not reject.

Procedure



b. One-Sided Lower-Tail Test.

To test

$$H_0 : \Delta = 0$$

versus

$$H_a : \Delta < 0$$

at the α level of significance, Reject H_0 if $W \leq w_{1-\alpha}$; otherwise do not reject.

Large-samaple approximation

Reject H_0 if $W^* \leq -z_\alpha$; otherwise do not reject.

Procedure

c. Two-Sided Test.

To test

$$H_0 : \Delta = 0$$

versus

$$H_a : \Delta \neq 0$$

at the α level of significance, Reject H_0 if $W \geq w_{\alpha/2}$ or $W \leq w_{1-\alpha/2}$; otherwise do not reject,

Large-samaple approximation

Reject H_0 if $|W^*| \geq z_{\alpha/2}$; otherwise do not reject.

An estimator for a shift parameter associated with the Wilcoxon's rank sum statistics (Hodges-Lehmann)

The Mann-Whitney Statistic

$$U = \text{ number of pairs } (X_i, Y_j) \text{ for which } X_i < Y_j$$
$$= \sum_{j=1}^{n} \sum_{i=1}^{m} 1\{X_i \leq Y_j\}$$

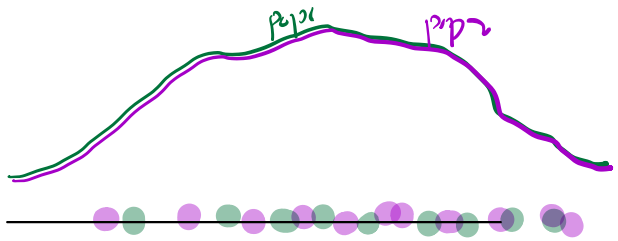
The null hypothesis is that the distributions of the X 's and Y 's are the same. A large value of U indicates that the larger observations tend to occur with treatment 2 (the Y' s), and vice versa if U is small.

$U =$ number of pairs (X_i, Y_j) for which $X_i < Y_j$

$$= \sum_{j=1}^{n} \sum_{i=1}^{m} 1\{X_i \le Y_j\}$$

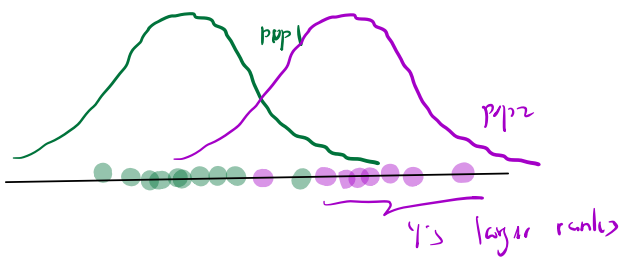
m Xs n Ys m·n

$\Delta : 0$



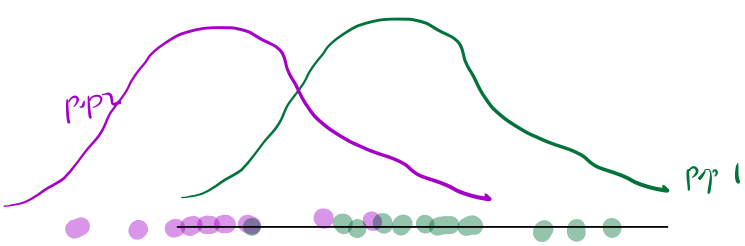
$$U \approx \frac{m \cdot n}{2}$$

$\Delta > 0$



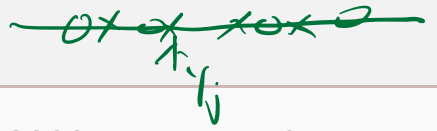
$$U >> \frac{m \cdot n}{2}$$

$\Delta < 0$:



$$U << \frac{n \cdot m}{2}$$

Mann-Whitney statistic = Wilcoxon rank sum statistic



The Mann-Whitney statistic can be shown to be equivalent to the Wilcoxon rank sum statistic.

$$R(Y_j) = (\text{number of } Y's \le Y_j) + (\text{number of } X's \le Y_j)$$

$$= \sum_{i=1}^{m} 1\{X_i \le Y_j\} + \sum_{j'=1}^{n} 1\{Y_{j'} \le Y_j\}$$

For simplicity assume that the Y 's have been arranged from smallest to largest; that is, $Y_1 < Y_2 < \cdots < Y_n$. Since the number of $Y' \le Y_j = j$,

$$\sum_{j'=1}^{n} 1\{Y_{j'} \le Y_j\} = j$$

$$\Rightarrow W = \sum_{j=1}^{n} R(Y_j) = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} 1(X_i \le Y_j) + j \right)$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{m} 1\{X_i \le Y_j\} + 1 + 2 + \cdots + n$$

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

Mann-Whitney statistic plays a useful role in the confidence interval. For statistical testing, there is no reason to prefer one over the other.

Motivation

A reasonable estimator of Δ is the amount $\hat{\Delta}$ (say) that should be subtracted from each Y_j so that the value of U, when applied to the aligned samples $X_1, \ldots, X_m, Y_1 - \widehat{\Delta}, \ldots, Y_n - \widehat{\Delta}$, is appear (when "viewed" by the Mann-Whitney statistic U) as two samples from the same population.

Under H_0, $U = \sum_{j=1}^{n} \sum_{i=1}^{m} 1\{X_i \leq Y_j - \Delta\}$ should be centered at $E[W] - \frac{n(n+1)}{2} = \frac{nm}{2}$.

$$\Rightarrow U = \sum_{j=1}^{n} \sum_{i=1}^{m} 1\{\Delta \leq Y_j - X_i\} \approx \frac{nm}{2}$$

$$\Rightarrow \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} 1\{\Delta \leq Y_j - X_i\} \approx \frac{1}{2}$$

$\Rightarrow \Delta$ is the value s.t. half of $Y_j - X_i$ lies below and above

$\Rightarrow \Delta$ shoud be the sample median of pairwise differences!

i.e. total observations above the true population median is the sample median.

Intuitively, we estimate θ by the amount that the X sample should be shifted in order that $X_1 - \tilde{\theta}, \ldots, X_n - \tilde{\theta}$ appears (when "viewed" by the sign statistic B) as a sample from a population with median 0 .

pop 1
shift by Δ
pop 2
x_5
y_5
Δ

Procedure

Estimate of Δ

To estimate Δ form the mn differences $Y_j - X_i$, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

$$\widehat{\Delta} = \text{median}\left\{\left(Y_j - X_i\right), i = 1, \ldots, m; j = 1, \ldots, n\right\}.$$

Let $U^{(1)} \leq \cdots \leq U^{(mn)}$ denote the ordered values of $Y_j - X_i$.

- if mn is odd, say $mn = 2k + 1$, we have $k = (mn - 1)/2$ and

  $$\widehat{\Delta} = U^{(k+1)},$$

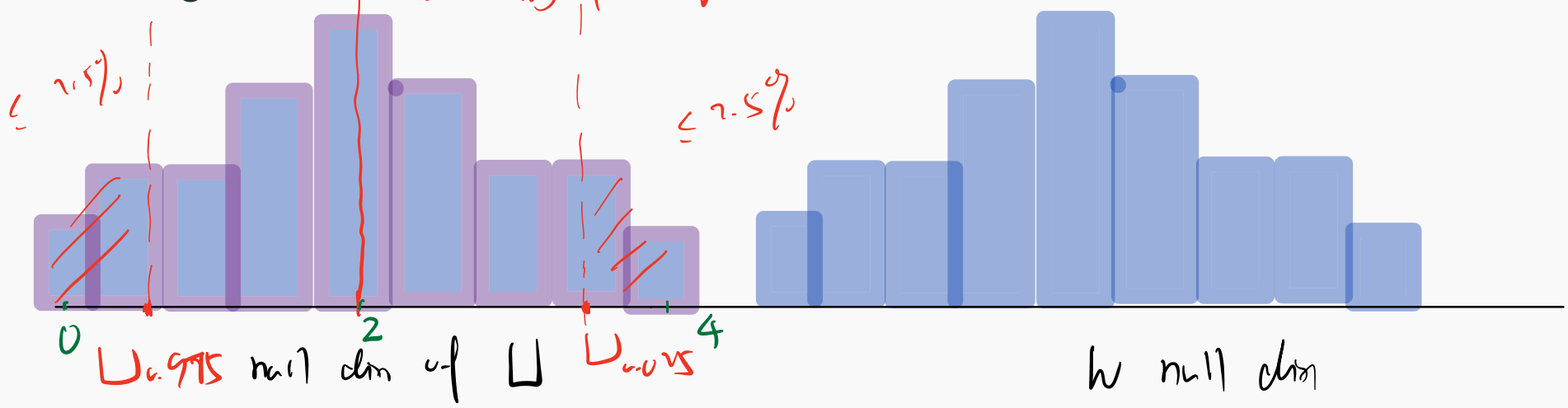
  the value that occupies the position $k + 1$ in the list of the ordered $Y - X$ differences.
- if mn is even, say $mn = 2k$, then $k = mn/2$ and

  $$\widehat{\Delta} = \frac{U^{(k)} + U^{(k+1)}}{2}$$

  That is, $\widehat{\Delta}$ is the average of the two $Y - X$ differences that occupy the positions k and $k + 1$ in the ordered list of the mn differences.

Confidence interval for a shift parameter associated with the Wilcoxon's rank sum statistics

The true shift parameter Δ is the value such that the number of pairwise differences above it is the Mann-Whitney U statisitcs which should be centered at $\frac{nm}{2}$ with some natural variation. So we use the natural variation of U reverse engineer the most probable region of Δ.



$$P(U_{\text{low}} \le U \le U_{\text{up}}) \approx 95\%$$

$$\sum_i \sum_j 1(X_i \le Y_j - \Delta)$$

$$= \sum_i \sum_j 1(\Delta \le Y_j - X_i)$$

$$\Rightarrow (U^{(1)}, \ldots, U^{(3)})$$

Procedure

$100(1-\alpha)\%$ confidence interval

For a symmetric two-sided confidence interval for θ, with confidence coefficient $1-\alpha$, first obtain the upper $(\alpha/2)$ nd percentile point $U_{\alpha/2}$ [3] of the null distribution of U

$$U_{1-\alpha/2} = nm + 1 - U_{\alpha/2}$$

The $100(1-\alpha)\%$ confidence interval $(\Delta_{\mathrm{L}}, \Delta_{\mathrm{U}})$ for Δ that is associated with sign test

$$\Delta_{\mathrm{L}} = U_{1-\alpha/2}, \Delta_{\mathrm{U}} = U_{\alpha/2}$$

where $U^{(1)} \leq \cdots \leq U^{(nm)}$ are the ordered pairwise differences.

Then we have

$$P_{\Delta}\left(\Delta_{\mathrm{L}} < \Delta < \Delta_{\mathrm{U}}\right) = 1-\alpha \text{ for all } \Delta.$$

[3] Since the distribution of U is discrete, it may not be possible to find percentiles such that the level of confidence is precisely. In such cases, we would choose these points so that the level of confidence is at least the stated level but as close as possible.

Example: Alcohol Intakes

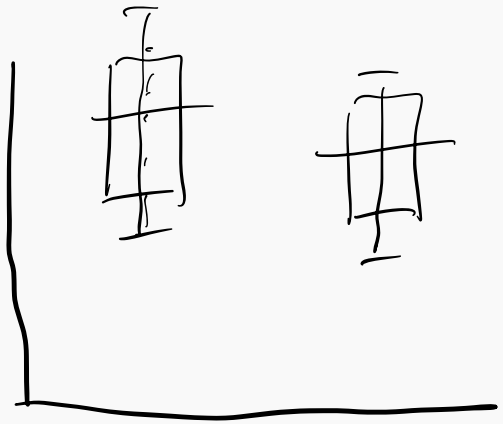
Eriksen, Björnstad, and Götestam (1986) studied a social skills training program for alcoholics. ~~Twenty-four~~ 23 "alcohol-dependent" male inpatients at an alcohol treatment center were randomly assigned to two groups. The control group patients were given a traditional treatment program. The treatment group patients were given the traditional treatment program plus a class in social skills training (SST). After being discharged from the program, each patient reported-in 2-week intervals - the quantity of alcohol consumed, the number of days prior to his first drink, the number of sober days, the days worked, the times admitted to an institution, and the nights slept at home. Reports were verified by other sources (wives or family members).

We are interested in whether SST group tends to have lower alcohol intakes

Control		SST	
1042	(13)	874	(9)
1617	(23)	389	(2)
1180	(18)	612	(4)
973	(12)	798	(7)
1552	(22)	1152	(17)
1251	(19)	893	(10)
1151	(16)	541	(3)
1511	(21)	741	(6)
728	(5)	1064	(14)
1079	(15)	862	(8)
951	(11)	213	(1)
1319	(20)		

$m = 12$ $n = 11$

Xs Ys



Control	SST		
1042	(4)	874	(3)
1617	(5)	389	(2)
1180	(4)	213	(1)

W = 6

Permutation distribution: $\binom{6}{3}$ = 15 permutations

Y-ranks	Probability	W
1, 2, 3		6
1, 2, 4		7
1, 2, 5	$\frac{1}{W}$: no	8
1, 2, 6		9
1, 3, 4		8
1, 3, 5		9
1, 3, 6		10
1, 4, 5		10
1, 4, 6		11
1, 5, 6		12
2, 3, 4		9
2, 3, 5		10
2, 3, 6		11
2, 4, 5		11
2, 4, 6		12
2, 5, 6		13
3, 4, 5		12

3 4 6 — 13

3 5 6 — 14

4 5 6 — 15

W	Prob
6	0.05
7	0.10
8	0.20
9	0.25
10	0.15
11	0.10
12	0.05
13	0.05
14	0.05
15	0.05

P-value = $P(W \in \{6\}) = 0.05$

(y-x) differences:

y	x	y-x	
874	1042	-168	(9)
	1017	-743	(6)
	1180	-306	(8)
389	1042	-653	(7)
	1017	-1228	(2)
	1180	-791	(5)
213	1042	-829	(4)
	1017	-1404	(1)
	1180	-967	(3)

$\Rightarrow \hat{\Delta} = -791$

95% Confidence interval: $U_{n,\alpha}$

$W = U + \frac{n(n+1)}{2} = U + \frac{3\times4}{2}$

$W_{obs} = 15$

$\Rightarrow U_{0.05} = 15 - 6 = 9$

$\Rightarrow U_{1-\alpha} = 3\times3+1 - 9 = 1$

$\Rightarrow$ 95% C.I. for $\hat{\Delta}$:

$U(1), U(9)$

Y-ranks : 11 out 23 ranks {1 ... 23}

$\binom{23}{11} > 1,000,000$ { 1, 1, ... 11

$$W = \sum_{i=1}^{n} R(Y_i) = 81 \qquad W = U + \frac{n(n+1)}{2}$$

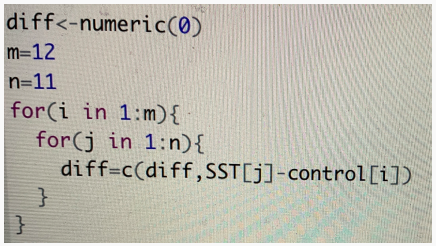
$$W^* = \frac{W - \frac{n(N+1)}{2}}{\sqrt{\frac{mn(N+1)}{12}}} = -3.138833 \qquad \frac{11 \times 12}{2} = 66$$

```
> control<-c(1042,1617,1180,973,1552,1251,1151,1511,728,1079,951,1319)
> SST<-c(874,389,612,798,1152,893,541,741,1064,862,213)
> wilcox.test(SST, control, alternative = c("less"),  exact = T) # Mann-Whitney statistics

        Wilcoxon rank sum exact test

data:  SST and control
W = 15, p-value = 0.0004904
alternative hypothesis: true location shift is less than 0

> # large sample
> pnorm(-3.138833)
[1] 0.0008481104
```

Both the exact and large-sample approximation indicate that there is strong evidence that the SST class in combination with the traditional treatment program tends to lower alcohol intake in alcoholics.

Treatment effect estimate and confidence interval:

```
> diff<-sort(diff)
> sum(diff[132/2]+diff[132/2+1])/2
[1] -435.5
>
> # verify with built-in function
> wilcox.test(SST, control, alternative = c("t"), conf.int=T)

        Wilcoxon rank sum exact test

data:  SST and control
W = 15, p-value = 0.0009807
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 -713 -186
sample estimates:
difference in location
                -435.5
```



```
diff<-numeric(0)
m=12
n=11
for(i in 1:m){
  for(j in 1:n){
    diff=c(diff,SST[j]-control[i])
  }
}
```

Why Ranks? Scoring Systems

Ranks can be thought of as scores that are used in place of the original observations in nonparametric methods. This leads us to consider coming up with other scores to be used in the same way. The key is to figure out a reasonable way to generate scores.

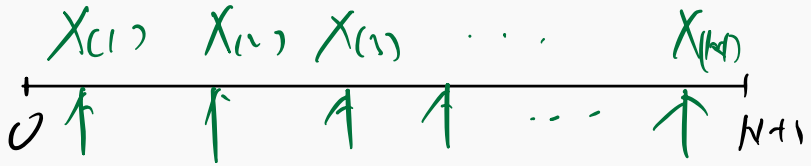
First, consider a particular way in which we might think of ranks. Suppose we take a random sample from a uniform probability distribution on the interval $[0, N+1]$; that is, the population distribution is

$$f(w) = \frac{1}{N+1}, 0 \leq w \leq N+1$$

Let $W_{(1)} < W_{(2)} < \cdots < W_{(N)}$ denote the order statistics of this random sample, where $W_{(1)}$ is the smallest observation, $W_{(2)}$ is the next smallest, and so on. It can be shown that the ranks are just the expected values of the $W_{(i)}$'s; that is, $E\left(W_{(i)}\right) = i$.

N samples:

$E[X_{(i)}] = i$



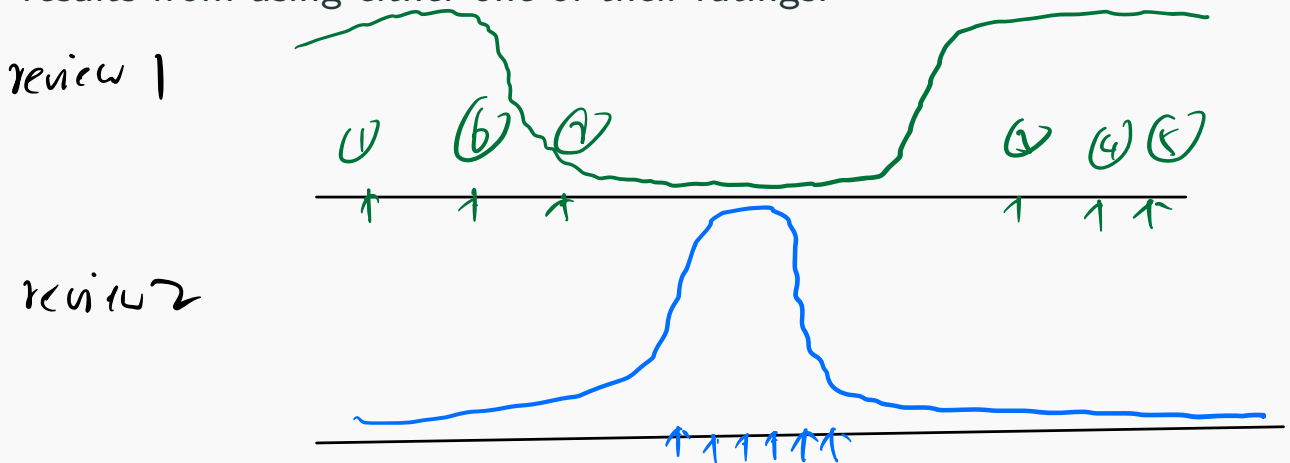
In this vein, we consider scoring systems as looking at the expected order of data, we can generalize based on distributions other than the uniform, which leads to more general scoring systems.

So why ignore the exact measurements we get from the original data?

In some sense we lose partial information coded in the exact numbers of the data, but by replacing in with its relative order, i.e. ranks, so in face value we lost something,

but the bonus is that, we can get procedures that are more generally applicable for many types of data, no distributional assumptions are necessary.

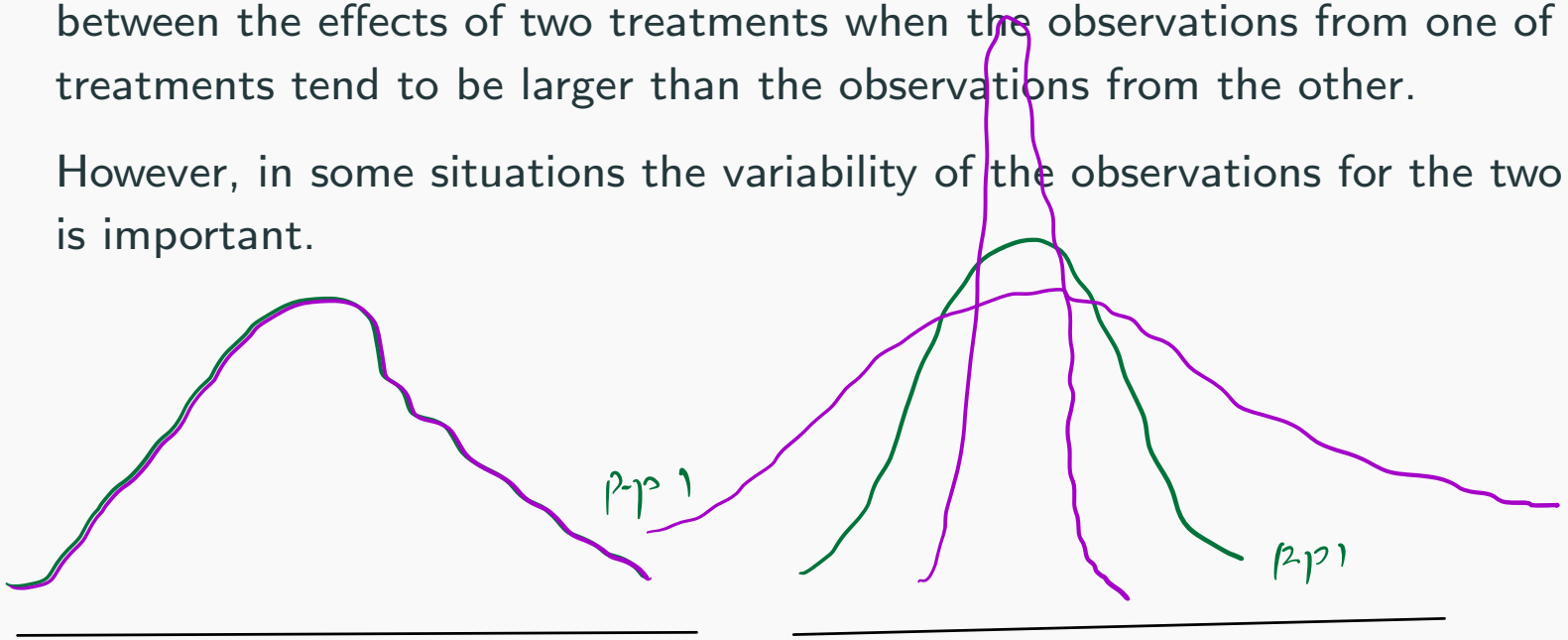
- For instance, two reviewers with similar tastes may rate products based on different numerical systems, as long as their general tastes are the same, we get exact same results from using either one of their ratings.



Tests for Equality of Scale Parameters

The tests we have discussed to this point are particularly designed to distinguish between the effects of two treatments when the observations from one of the treatments tend to be larger than the observations from the other.

However, in some situations the variability of the observations for the two treatments is important.



Suppose a machine for bottling a soft drink is designed to fill containers with 16 ounces of the beverage. Observations on the process may show that the data are centered around 16 as they should be, but there is excessive variability. This finding could lead an engineer to identify a problem that, if fixed, would reduce the variability of product quality.

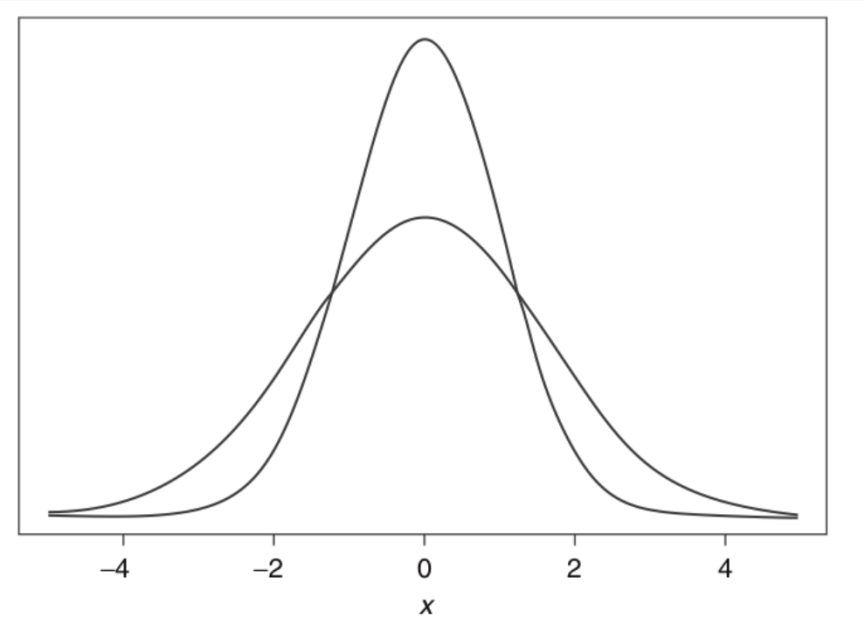


Figure 1: distributions of observations of the type that we would expect to see if the treatments affect the scale parameters of the population distributions but not the location parameters.

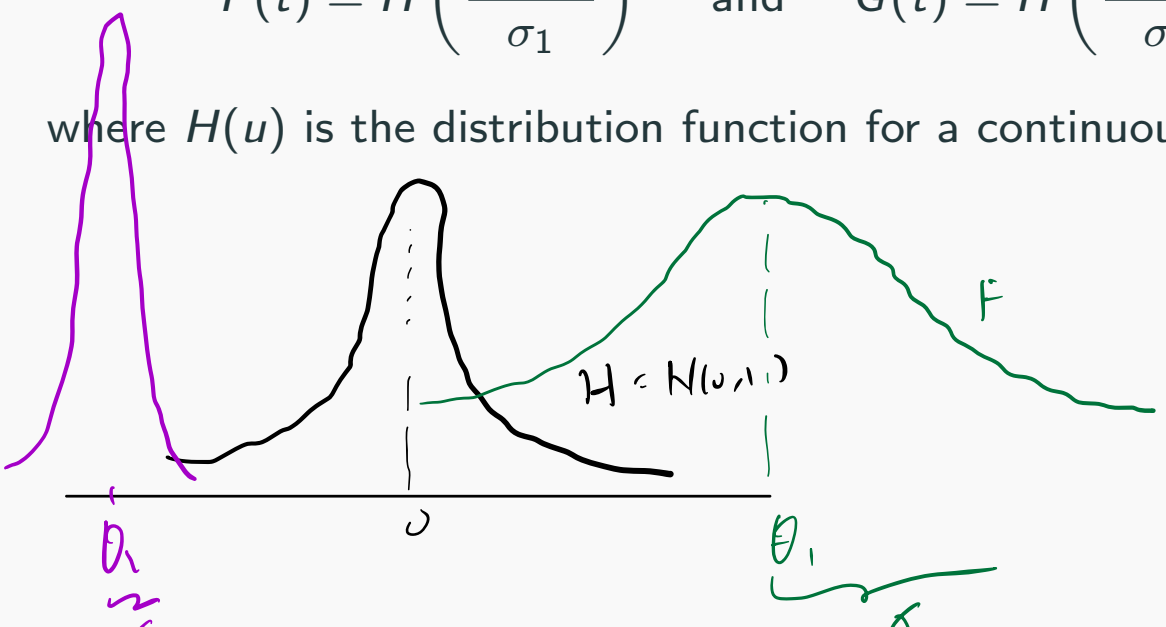
Setting

Let F and G be the distribution functions corresponding to populations 1 and 2, respectively.

The two-sample dispersion problem specifies that the Y population has greater (or less) variability associated with it than does the X population. We work under the location scale parameter model:

$$F(t) = H\left(\frac{t-\theta_1}{\sigma_1}\right) \quad \text{and} \quad G(t) = H\left(\frac{t-\theta_2}{\sigma_2}\right), \quad -\infty < t < \infty,$$

where $H(u)$ is the distribution function for a continuous distribution with median 0.



Setting

General assumptions of location scale model:

- θ_1 and θ_2 are the location parameter (~~medians~~ means) for the X and Y distributions, respectively.
- σ_1 and σ_2 are the scale parameters (variance) associated with the X and Y distributions, respectively.
- Y population has the same general form as the X population, but they could have different ~~medians~~ means and scale parameters.
- Another way to express this is to write

$$\frac{X-\theta_1}{\sigma_1} \stackrel{d}{=} \frac{Y-\theta_2}{\sigma_2},$$

We assume $\theta_1 = \theta_2 = \theta$, i.e. equal median.

$$\frac{X-\theta}{\sigma_1} \stackrel{d}{=} \frac{Y-\theta}{\sigma_2},$$

Hypothesis (Ansari-Bradley test)

The parameter of interest is the ratio of the scale parameters, $\gamma^2 = \frac{\sigma_1^2}{\sigma_2^2}$

Two-Sided Test:

$$H_0 : \gamma^2 = 1 \text{ versus } H_a : \gamma^2 \neq 1$$

One-Sided Upper-Tail Test:

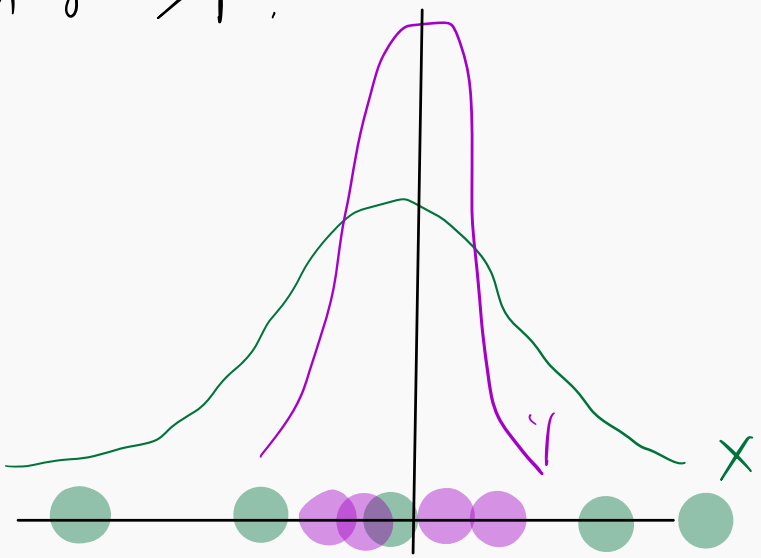
$$H_0 : \gamma^2 = 1 \text{ versus } H_a : \gamma^2 > 1$$

One-Sided Lower-Tail Test:

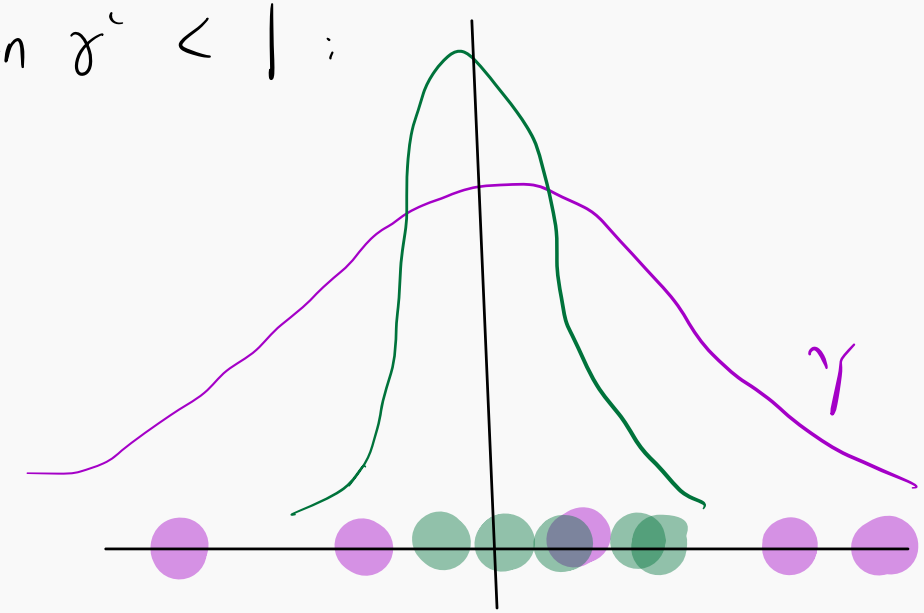
$$H_0 : \gamma^2 = 1 \text{ versus } H_a : \gamma^2 < 1$$

A rank based test for dispersion when median equal (Ansari-Bradley test)

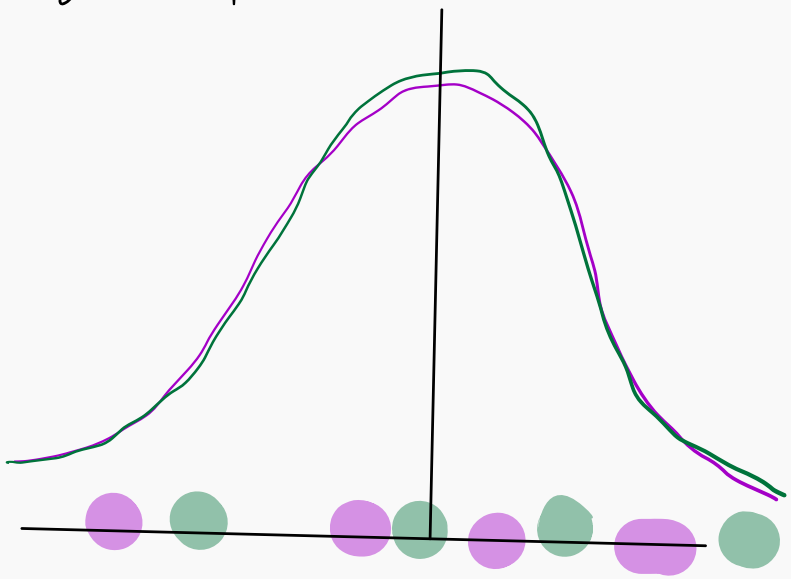
when $\gamma > 1$:



when $\gamma < 1$:



when $\gamma^2 \simeq 1$:



How to capture relative deviation from the center without using the exact values of X and Y?

- Order the combined sample of $N = (m + n)$ X-values and Y-values from least to greatest.
- Assign the score 1 to both the smallest and largest observations in this combined sample, assign the score 2 to the second smallest and second largest, and continue in the manner.
  - If N is an even integer, the array of assigned scores is $1, 2, 3, \ldots, N/2, N/2, \ldots, 3, 2, 1$.
  - If N is an odd integer, the array of assigned scores is $1, 2, 3, \ldots, (N-1)/2, (N+1)/2, (N-1)/2, \ldots, 3, 2, 1$.

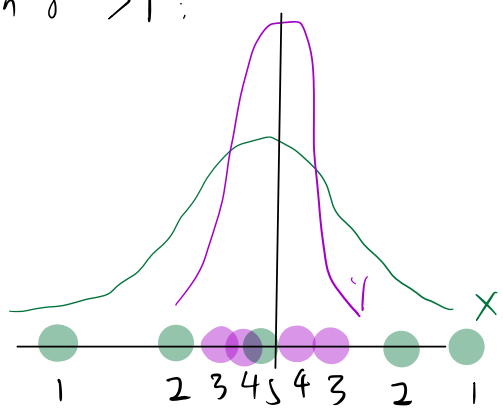
Let $R(Y_j)$ denote the score assigned in this manner to Y_j, for $j = 1, \ldots, n$, and

$$C = \sum_{j=1}^{n} R(Y_j)$$

Closer to the common center : rank ↑

Further away from the common center : rank ↓

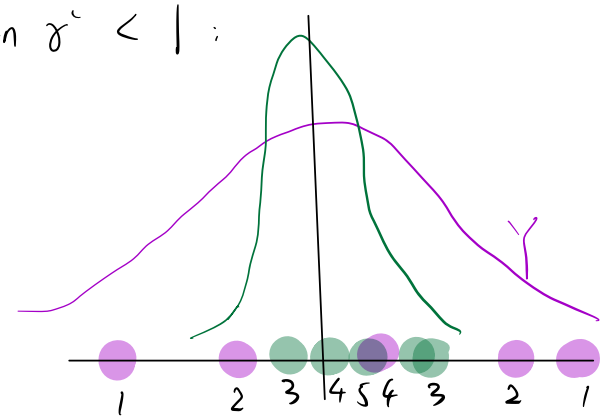
When $\gamma^2 > 1$:



C large

C : 3 + 4 + 4 + 3 = 14

When $\gamma^2 < 1$:

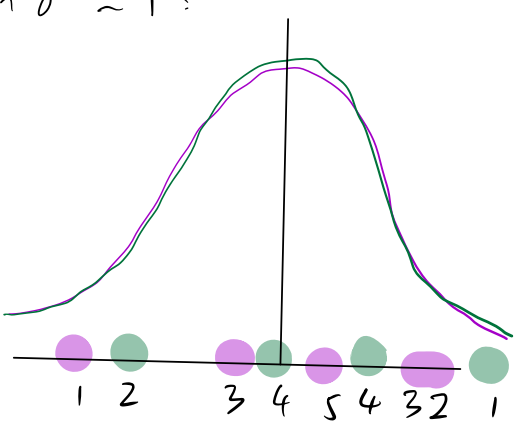


C small

C = 1 + 2 + 4 + 2 + 1

= 10

When $\gamma^2 \simeq 1$:



C not too large nor too small

C : 1 + 3 + 5 + 3 + 2 = 14

Derivation of null distribution using permutation

When H_0 is true: There are

$$\binom{n+m}{n} = \frac{(n+m)!}{n!\,m!}$$

possible assignments for $Y-$ranks (equal likely).

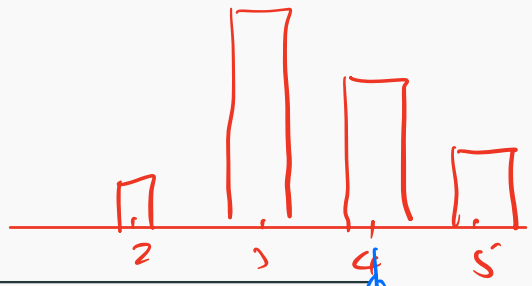
$m = 3, n = 2$ $\binom{5}{2} = \frac{5!}{2!\,3!} = \frac{5\times 4}{2} = 10$ permutations

~~Meshing~~	Probability	~~$(R^{(1)}, R^{(2)})$~~	$C = ~~R^{(1)} + R^{(2)}~~$
YYXXX	$\frac{1}{10}$	$(1,2)$	3
YXYXX	$\frac{1}{10}$	$(1,3)$	4
YXXYX	$\frac{1}{10}$	$(1,2)$	3
YXXXY	$\frac{1}{10}$	$(1,1)$	2
XYYXX	$\frac{1}{10}$	$(2,3)$	5
XYXYX	$\frac{1}{10}$	$(2,2)$	4
XYXXY	$\frac{1}{10}$	$(1,2)$	3
XXYYX	$\frac{1}{10}$	$(2,3)$	5
XXYXY	$\frac{1}{10}$	$(1,3)$	4
XXXYY	$\frac{1}{10}$	$(1,2)$	3

$R(1,) + R(1,)$

C: 2 3 4 5

=> null dist of C :



Possible value of C	Probability under H_0
2	$\frac{1}{10}$
3	$\frac{4}{10}$
4	$\frac{3}{10}$
5	$\frac{2}{10}$

$C_{obs} = 4$

$H_A: \gamma > 1$

Thus, for example, the probability, under H_0, that C is greater than or equal to 4 , for example, is therefore

$$P_0(C \geq 4) = P_0(C = 4) + P_0(C = 5) = .3 + .2 = .5,$$

Large sample approximation of null distribution

$$\frac{C}{n} = \frac{1}{n}\sum_{i=1}^{n} R(Y_i)$$

is the average of the scores assigned to the Y observations.

We want to know the behavior of: Under H_0, sample mean of a random sample of size n drawn without replacement from finite population of scores S_N, where $S_N = \{1, 2, 3, \ldots, N/2, N/2, \ldots, 3, 2, 1\}$ if N is an even number and $S_N = \{1, 2, 3, \ldots, (N-1)/2, (N+1)/2, (N-1)/2, \ldots, 3, 2, 1\}$ if N is odd. [4]

[4] Facts from finite population theory:

- The mean is equal to the mean μ_{pop} of the finite population.
- The variance is equal to

$$\frac{\sigma^2_{\text{pop}}}{n} \times \frac{N-n}{N-1},$$

where σ^2_{pop} denotes the variance of the finite population and the factor $(N-n)/(N-1)$ is the finite population correction factor.

- if N even:

$$S_N = \{1, 2, 3, \ldots, N/2, N/2, \ldots, 3, 2, 1\}$$
$$u_{\text{pop}} = \frac{2}{N} \sum_{i=1}^{N/2} i = \frac{(N/2)[(N/2)+1]}{2(N/2)} = \frac{N+2}{4}$$
$$E_0\left(\frac{C}{n}\right) = \frac{N+2}{4}$$
$$\text{var}_0\left(\frac{C}{n}\right) = \left[\frac{(N+2)(N-2)}{48n}\right]\left[\frac{N-n}{N-1}\right] = \frac{m(N+2)(N-2)}{48n(N-1)}$$
$$\Rightarrow E_0(C) = nE_0\left(\frac{C}{n}\right) = \frac{n(N+2)}{4}$$
$$\Rightarrow \text{var}_0(C) = n^2 \,\text{var}_0\left(\frac{C}{n}\right) = \frac{mn(N+2)(N-2)}{48(N-1)}$$

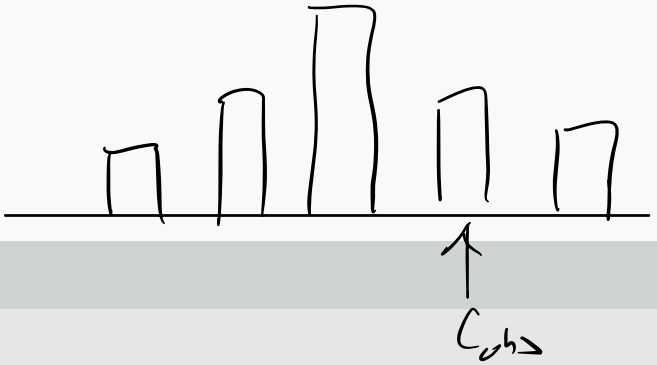
- if N odd:

$$S_N = \{1, 2, 3, \ldots, (N-1)/2, (N+1)/2, (N-1)/2, \ldots, 3, 2, 1\}$$
$$\Rightarrow E_0(C) = \frac{n(N+1)^2}{4N}$$
$$\Rightarrow \text{var}_0(C) = \frac{mn(N+1)\left(3+N^2\right)}{48N^2}$$

Asymptotic normality follows from standard theory for the mean of a sample.

$$C^* = \frac{C - E_0(C)}{\sqrt{\text{var}_0(C)}} \sim N(0, 1)$$

Procedure



a. One-Sided Upper-Tail Test

To test

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

versus

$$H_1 : \gamma^2 > 1$$

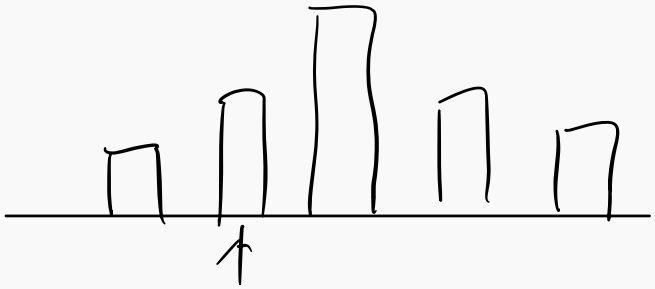
at the α level of significance,

Reject H_0 if $C \geq c_\alpha$; otherwise do not reject, where the constant w_α is chosen to make the type I error probability equal to α. (Or use p-value)

Large-sample approximation

Reject H_0 if $C^* \geq z_\alpha$; otherwise do not reject.

Procedure



b. One-Sided Lower-Tail Test

To test

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

versus

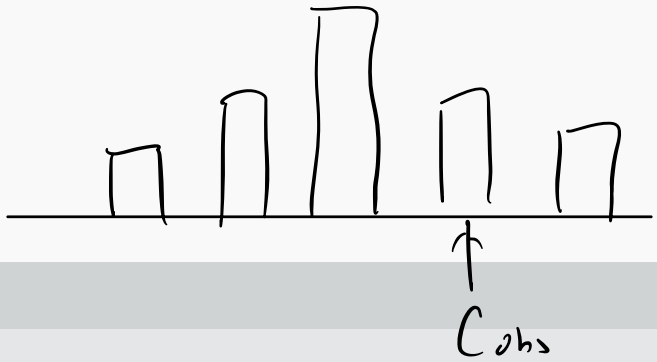
$$H_a : \gamma^2 < 1$$

at the α level of significance, Reject H_0 if $C \leq c_{1-\alpha}$; otherwise do not reject.

Large-sample approximation

Reject H_0 if $C^* \leq -z_\alpha$; otherwise do not reject.

Procedure



c. Two-Sided Test

To test

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

versus

$$H_3 : \gamma^2 \neq 1$$

at the α level of significance, Reject H_0 if $C \geq c_{\alpha/2}$ or $C \leq c_{1-\alpha/2}$; otherwise do not reject

Large-sample approximation

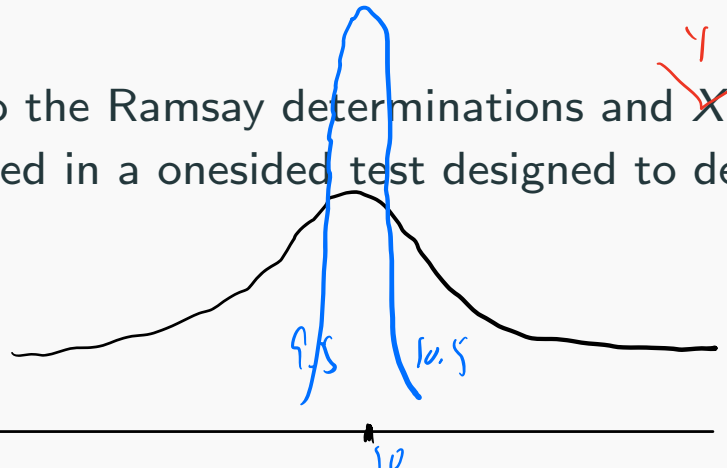
Reject H_0 if $|C^*| \geq z_{\alpha/2}$; otherwise do not reject.

Example:Serum Iron Determination

Jung and Parekh (1970) in a study concerned with techniques for direct determination of serum iron. In particular, they attempted to eliminate some of the problems associated with other commonly used methods, which often result in turbidity of the analyzed serum, as well as requiring large samples and slow, tedious analyses. To accomplish this, the authors proposed an improved method for serum iron determination based on a different detergent.

One of the purposes of their investigation was to study the accuracy of their method for serum iron determination in comparison to a method due to Ramsay (1957). From the point of view of procedural technique, the Jung-Parekh method competes favorably with the Ramsay method for serum iron determination. An additional concern, however, is whether there is a loss of accuracy when the Jung-Parekh procedure is used instead of the Ramsay procedure. As a result, the interest is greater dispersion or variation for the Jung-Parekh method of serum iron determination than for the method of Ramsay.

Hence, letting X correspond to the Ramsay determinations and Y to the Jung-Parekh determinations, we are interested in a onesided test designed to detect the alternative $H_1 : \gamma^2 > 1$.



true iron

old new

Ramsay method	Jung-Parekh method
111	107.5
107	108
99.5	105.5
98.5	98
102	105
106	103
109	110
108.5	106.5
103.5	104
99	100

Ramsay method	Jung-Parekh method
111	107.5
107	108
99.5	~~105.5~~

$m=3 \quad n=2$

order the combined sample : 99.5 107 107.5 108 111

rank : 1 2 3 2 1

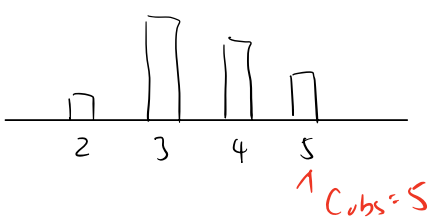
$\Rightarrow C = 5$

Permutation null distribution:

$m = 3, n = 2$

~~Meshing~~	Probability	~~$(R^{(1)}, R^{(2)})$~~	$C = ~~R^{(1)} + R^{(2)}~~$ $R(1,)+R(1,)$
YYXXX	$\frac{1}{10}$	$(1,2)$	3
YXYXX	$\frac{1}{10}$	$(1,3)$	4
YXXYX	$\frac{1}{10}$	$(1,2)$	3
YXXXY	$\frac{1}{10}$	$(1,1)$	2
XYYXX	$\frac{1}{10}$	$(2,3)$	5
XYXYX	$\frac{1}{10}$	$(2,2)$	4
XYXXY	$\frac{1}{10}$	$(1,2)$	3
XXYYX	$\frac{1}{10}$	$(2,3)$	5
XXYXY	$\frac{1}{10}$	$(1,3)$	4
XXXYY	$\frac{1}{10}$	$(1,2)$	3

Possible value of C	Probability under H_0
2	$\frac{1}{10}$
3	$\frac{4}{10}$
4	$\frac{3}{10}$
5	$\frac{2}{10}$



$H_1: \gamma^2 > 1$

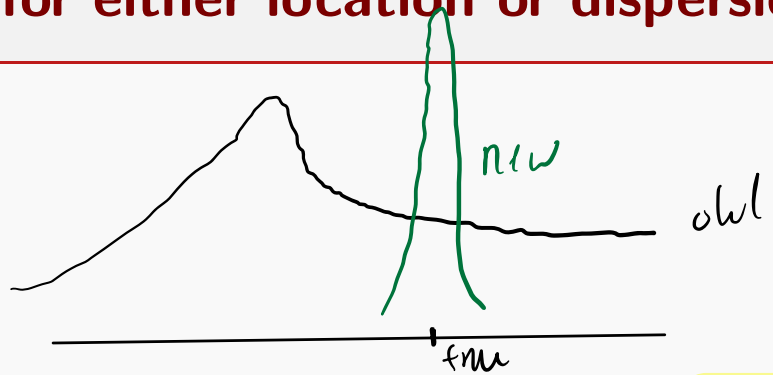
C larger

p-value $= P(C \geq 5) = P(C=5) = 0.2 > 0.05$

```
> ramsay <- c(111, 107, 99.5, 98.5, 102, 106, 109, 108.5, 103.5, 99)
> jung.parekh <- c(107.5, 108, 105.5, 98, 105, 103, 110, 106.5, 104,100)
>
> rank(c(ramsay,jung.parekh))
 [1] 20 14  4  2  6 12 16 17  8  3 15 18 11  1 10  7 19 13  9  5
>
> ansari.test(ramsay, jung.parekh, alternative="greater",exact=T)

        Ansari-Bradley test

data:  ramsay and jung.parekh
AB = 47, p-value = 0.1306
alternative hypothesis: true ratio of scales is greater than 1

> # large-sample normal approximation
> ansari.test(ramsay, jung.parekh, alternative="greater", exact=F)

        Ansari-Bradley test

data:  ramsay and jung.parekh
AB = 47, p-value = 0.1124
alternative hypothesis: true ratio of scales is greater than 1
```

Hence, there is not sufficient evidence to indicate loss of accuracy when the Jung-Parekh method is used instead of the Ramsay method.

A rank based test for either location or dispersion (Lepage test)



In many two-sample situations, we are interested in simultaneously detecting either location or scale differences between the X and Y populations.

We are interested in assessing whether there are differences in either the location parameters (i.e., medians) θ_1 and θ_2 or the scale parameters σ_1 and σ_2 for the X and Y populations.

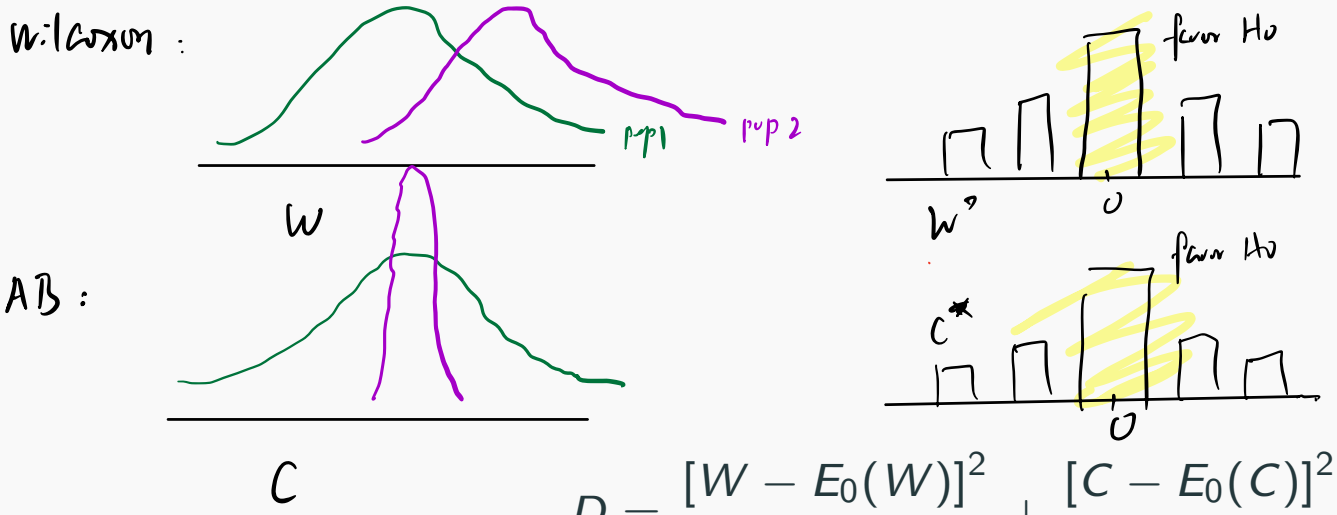
Thus, we are interested in testing

$$H_0 : \theta_1 = \theta_2, \sigma_1 = \sigma_2$$

vs

$$H_a : \theta_1 \neq \theta_2, \text{ or } \sigma_1 \neq \sigma_2$$

Motivation



$$D = \frac{[W - E_0(W)]^2}{\text{var}_0(W)} + \frac{[C - E_0(C)]^2}{\text{var}_0(C)}$$

$$= (W^*)^2 + (C^*)^2$$

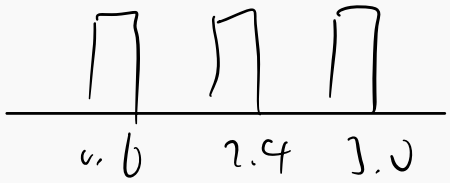
- A large value of $(W^*)^2$ is indicative of a possible difference in locations
- A large value of $(C^*)^2$ is indicative of a possible difference in dispersions
- D will be large if and only if $(W^*)^2$ is large or $(C^*)^2$ is large or both, then such a large value of D is indicative of alternative is true

Derivation of null distribution using permutation

$m = 2, n = 2$

	Probability	$D = (W^*)^2 + (C^*)^2$
$XXYY$	$\frac{1}{6}$	2.4
$XYXY$	$\frac{1}{6}$	.6
$YXXY$	$\frac{1}{6}$	3.0
$XYYX$	$\frac{1}{6}$	3.0
$YXYX$	$\frac{1}{6}$	.6
$YYXX$	$\frac{1}{6}$	2.4

Value of D	Probability under H_0
0.6	$\frac{1}{3}$
2.4	$\frac{1}{3}$
3.0	$\frac{1}{3}$



$n = m = 2 \quad N = n + m = 4$

$$W^* = \frac{W - \frac{n(N+1)}{2}}{\sqrt{\frac{mn(N+1)}{12}}} = \frac{W - 5}{\sqrt{\frac{5}{3}}}$$

$$C^* = \frac{C - \frac{n(N+2)}{4}}{\sqrt{\frac{mn(N+2)(N-2)}{48(N-1)}}} = \frac{C - 3}{\sqrt{\frac{1}{3}}}$$

Permutations	Prob	W	W^*	C	C^*	$D^2 = (W^*)^2 + (C^*)^2$
Y Y X X	$\frac{1}{6}$	3		3		
Y X Y X	$\frac{1}{6}$	4		3		
Y X X Y	$\frac{1}{6}$	5		2		
X Y Y X	$\frac{1}{6}$	5		4		
X Y X Y	$\frac{1}{6}$	6		3		
X X Y Y	$\frac{1}{6}$	7		3		

$$\binom{4}{2} = 6$$

Large sample approximation of null distribution

$$D \sim \chi^2_{df=2}$$

$$D = (W')^2 + (C')^2$$

Example:Effect of Maternal Steroid Therapy on Platelet Counts of Newborn Infants.

Autoimmune thrombocytopenic purpura (ATP) is a disease in which the patient produces antibodies to his/her own platelets. Due to transplacental passage of antiplatelet antibodies during pregnancy, children of women with ATP are often born with low platelet counts. For this reason, there is medical concern that a vaginal delivery for a mother with ATP could result in intracranial hemorrhage for the infant. However, the proper obstetrical management of pregnant women with ATP is controversial. Most doctors have advocated cesarean section as the preferable method of delivery for mothers with ATP. Others suggest that cesarean section, with its obvious complications for both mother and infant, be avoided unless there is some additional obstetrical reason for it. Karpatkin, Porges, and Karpatkin (1981) studied the effect of administering the corticosteroid prednisone to pregnant women with ATP with the intent of raising the infants' platelet counts to safe levels during their deliveries.

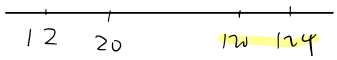
The data are a subset of the data obtained by Karpatkin et al. in their study of the effect that administration of prednisone to pregnant women with ATP had on their infants' platelet counts.

The primary interest in the study is in whether or not the predelivery administration of prednisone typically leads to an increased newborn platelet count. Thus, the principal statistical issue in the study is that of a possible difference in locations for the prednisone and nonprednisone populations. However, there is some concern that the administration of predelivery prednisone could also lead to a rather large increase in variability in the newborn platelet counts. (Such a finding would certainly affect our interpretation of any possible increase in typical platelet count resulting from the prednisone.)

We take the infant platelet count data for mothers given prednisone to be the Y sample ($n = 10$) and the corresponding control (nonprednisone) data to be the X sample ($m = 6$).

Mothers given prednisone	Mothers not given prednisone
120,000	12,000
124,000	20,000
215,000	112,000
90,000	32,000
67,000	60,000
95,000	40,000
190,000	
180,000	
135,000	
399,000	

Mothers given prednisone	Mothers not given prednisone
120,000	12,000
124,000	20,000



$n = m = 2 \qquad N = n + m = 4$

$W = 7$

$$W^* = \frac{W - \frac{n(N+1)}{2}}{\sqrt{\frac{mn(N+1)}{12}}} = \frac{W - 5}{\sqrt{\frac{5}{3}}} = \frac{2}{\sqrt{\frac{5}{3}}}$$

$C = 3$

$$C^* = \frac{C - \frac{n(N+2)}{4}}{\sqrt{\frac{mn(N+2)(N-2)}{48(N-1)}}} = \frac{C - 3}{\sqrt{\frac{1}{3}}} = 0$$

$$\Rightarrow D = \left(\frac{2}{\sqrt{\frac{5}{3}}}\right)^2 + 0^2 = 2.4$$

	Probability	$D = (W^*)^2 + (C^*)^2$
XXYY	$\frac{1}{6}$	2.4
XYXY	$\frac{1}{6}$	.6
YXXY	$\frac{1}{6}$	3.0
XYYX	$\frac{1}{6}$	3.0
YXYX	$\frac{1}{6}$	.6
YYXX	$\frac{1}{6}$	2.4

Value of D	Probability under H_0
0.6	$\frac{1}{3}$
2.4	$\frac{1}{3}$
3.0	$\frac{1}{3}$

$$\Rightarrow \text{p-value} = P(D \geq 2.4) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

$$> 0.05$$

$$W = 10 + 11 + 16 + 7 + 6 + 8 + 14 + 13 + 12 + 15 = 112$$

$$W^* = \frac{112 - \{10(6 + 10 + 1)/2\}}{\{6(10)(6 + 10 + 1)/12\}^{1/2}} = 2.929$$

$$C = 7 + 6 + 2 + 7 + 6 + 8 + 3 + 4 + 5 + 1 = 49$$

$$C^* = \frac{49 - \{10(16 + 2)/4\}}{\left\{\frac{10(6)(16+2)(16-2)}{48(16-1)}\right\}^{1/2}} = .873$$

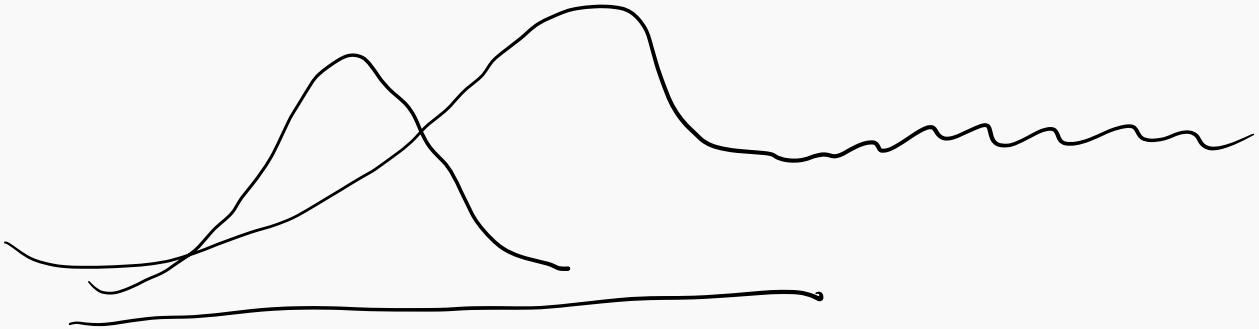
$$D = (2.929)^2 + (.873)^2 = 9.34$$

[redacted]

```
> library(NSM3)
> # large-sample
> pLepage(x,y,method="Asymptotic")
Number of X values: 10 Number of Y values: 6
Lepage D Statistic: 9.3384
Asymptotic upper-tail probability: 0.0094
> # permutation
> pLepage(x,y,method="Exact")
Number of X values: 10 Number of Y values: 6
Lepage D Statistic: 9.3384
Exact upper-tail probability: 0.0025
```

We reject H_0 and there is significant difference in either locations or variabilities between the infant platelet counts for the prednisone and control populations.

An Omnibus Test for general differences in two populations (Kolmogorov-Smirnov test)



Suppose it is not known how a difference between two ~~treatments~~ populations might manifest itself in the data.

It might cause observations in one treatment to be larger than observations in the other, or it might affect the variability of the observations, or it might affect the shapes of the distributions in some other way.

What we would like is an omnibus test -that is, a test designed to pick up differences among ~~treatments~~ populations regardless of the nature of the differences.

The Kolmogorov-Smirnov test is appropriate for this situation.

Setting

We obtain $N = m + n$ observations $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$.

- The observations $X_1, \ldots, X_m$ are a random sample from population 1; that is, the X 's are independent and identically distributed. The observations $Y_1, \ldots, Y_n$ are a random sample from population 2; that is, the Y 's are independent and identically distributed.
- The X 's and Y 's are mutually independent. Thus, in addition to assumptions of independence within each sample, we also assume independence between the two samples.
- Populations 1 and 2 have continuous distributions with distribution functions F and G, respectively.

Hypothesis

We are interested in assessing whether there are any differences whatsoever between the X and Y probability distributions.

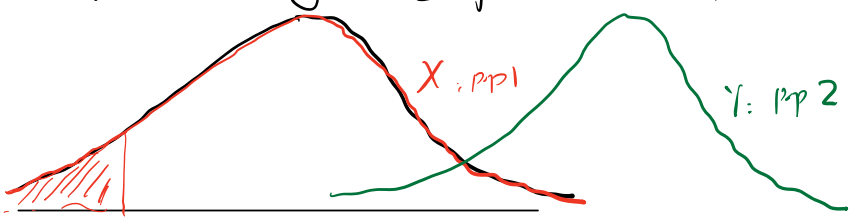
$$H_0 : F(t) = G(t) \text{ for all } t$$

versus

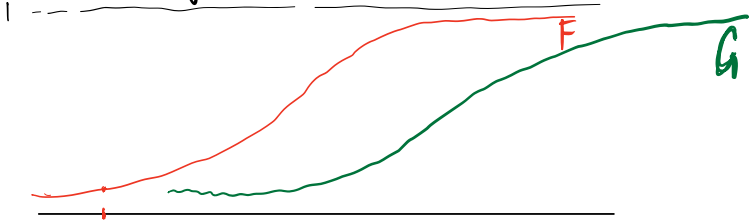
$$H_a : F(t) \neq G(t) \text{ for at least some } t$$

Continuous data (random variables):

p.d.f (probability density function) $f(x)$



distribution function $F(x) \triangleq P(X \le x)$



H_0 : $F(t) = G(t)$ for $\forall t$

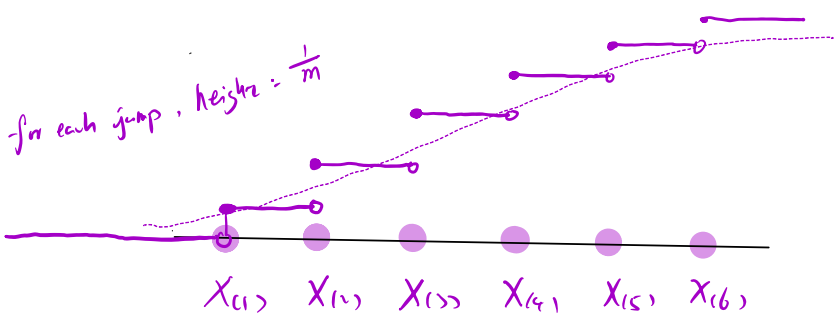
H_a : $F(t) \neq G(t)$ for at least some t

idea: use sample to estimate F, G, and then test

Defn: Empirical distribution function

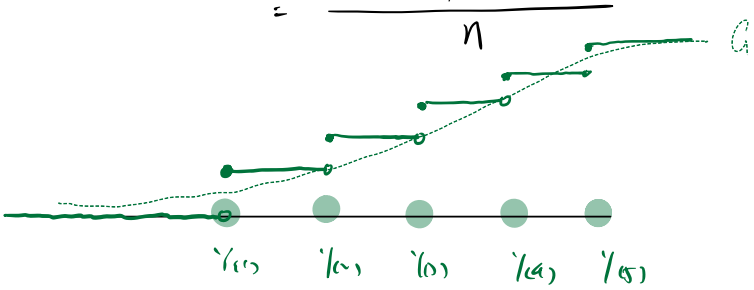
for X pop: $F_m(t)$ = proportion of X samples $\le t$

$$= \frac{\#\ X\text{s} \le t}{m}$$

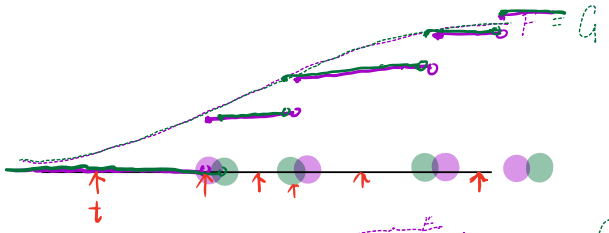


for 1 pop: $G_n(t)$ = proportion of 1 samples $\le t$

$$= \frac{\#\ 1 \text{ samples} \le t}{n}$$

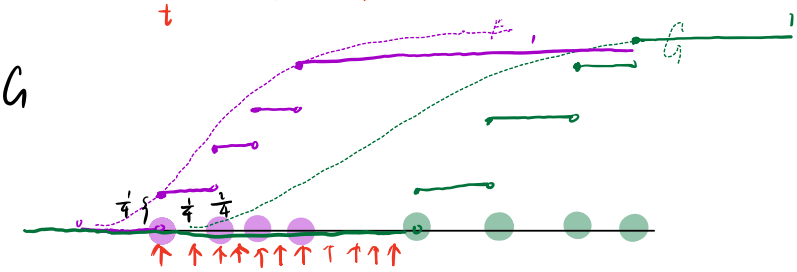


$H_0: F \equiv G$



$D \approx 0$
small in favor H_0

$H_a: F \neq G$



$D \nearrow$
large in favor H_a

$$\underbrace{D}_{\text{difference}} = \max_{-\infty < t < \infty} |F_m(t) - G_n(t)|$$

Motivation

Obtain the empirical distribution functions for the X and Y samples. For every t, let

$$F_m(t) = \frac{\text{number of sample } X \text{ 's } \leq t}{m}$$

and

$$G_n(t) = \frac{\text{number of sample } Y \text{ 's } \leq t}{n}.$$

(The functions $F_m(t)$ and $G_n(t)$ are called the empirical distribution functions for the X and Y samples, respectively.)

- if $F(t) = G(t)$

  $\max_{(-\infty<t<\infty)} \{|F_m(t) - G_n(t)|\}$ tend to $\downarrow$
- if $F(t) \neq G(t)$

  $\max_{(-\infty<t<\infty)} \{|F_m(t) - G_n(t)|\}$ tend to $\uparrow$

Kolmogorov-Smirnov statistic

$$D = \max_{(-\infty < t < \infty)} \{|F_m(t) - G_n(t)|\}$$

$F_m(t)$ and $G_n(t)$ are step functions changing functional values only at the observed X and Y sample observations, respectively. Thus, if we let $Z_{(1)} \leq \cdots \leq Z_{(N)}$ denote the $N = (m + n)$ ordered values for the combined sample of $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, then

$$D = \max_{i=1,\ldots,N} \left\{ \left| F_m\left(Z_{(i)}\right) - G_n\left(Z_{(i)}\right) \right| \right\}$$

$$J = \frac{mn}{d} \max_{i=1 \ldots N} \left\{ \left| F_m(Z_{(i)}) - G_n(Z_{(i)}) \right| \right\}$$

where d : greatest common divisor of m and n

Derivation of null distribution using permutation

$\binom{m+n}{n}$ permutations of (X, Y) samples

calculate D / J for each permutation

$\sim$

KS

Large sample approximation of null distribution

As $\min(m, n)$ tends to infinity,

$$J^* = \left(\frac{mn}{N}\right)^{1/2} \max_{i=1,\ldots,N} \left\{\left|F_m\left(Z_{(i)}\right) - G_n\left(Z_{(i)}\right)\right|\right\} \sim \text{ Kolmogorov distribution}$$

rescaling factor $\quad D$

Example:Effect of Feedback on Salivation Rate

The effect of enabling a subject to hear himself salivate while trying to increase or decrease his salivary rate has been studied by Delse and Feather (1968). Two groups of subjects were told to attempt to increase their salivary rates upon observing a light to the left and decrease their salivary rates upon observing a light to the right. Members of the feedback group received a 0.2-s, 1000-cps tone for each drop collected, whereas members of the no-feedback group did not receive any indication of their salivary rates.

Feedback group	No-Feedback group
−.15	2.55
8.60	12.07
5.00	.46
3.71	.35
4.29	2.69
7.74	−.94
2.48	1.73
3.25	.73
−1.15	−.35
8.38	−.37

i	$z_{(i)}$	$F_{10}(z_{(i)})$	$G_{10}(z_{(i)})$	$\lvert F-G\rvert$
1	−1.15	0.1	0	0.1
2	−0.54	0.1	0.1	0
3	−0.37	0.1	0.2	0.1
4	−0.35	0.1	0.3	0.2
5	−0.15	0.2	0.3	0.1
6	0.35	0.2	0.4	0.2
7	0.46	0.2	0.5	0.3
8	0.73	0.2	0.6	0.4
9	1.73	0.2	0.7	0.5
10	2.48	0.3	0.7	0.4
11	2.55	0.3	0.8	0.5
12	2.69	0.3	0.9	0.6
13	3.25	0.4	0.9	0.5
14	3.71	0.5	0.9	0.4
15	4.29	0.6	0.9	0.3
16	5	0.7	0.9	0.2
17	7.74	0.8	0.9	0.1
18	8.38	0.9	0.9	0
19	8.60	1	0.9	0.1
20	12.07	1	1	0

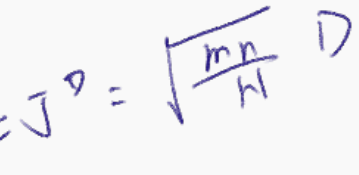
$$\Rightarrow D = 0.6$$

$$\Rightarrow J = \frac{mn}{d} D = \frac{10 \times 10}{\cancel{100}} \times 0.6 = 6$$

i	$Z_{(i)}$	$F_{10}\left(Z_{(i)}\right)$	$G_{10}\left(Z_{(i)}\right)$	$\left\lvert F_{10}\left(Z_{(i)}\right) - G_{10}\left(Z_{(i)}\right)\right\rvert$
1	-1.15	$\frac{1}{10}$	$\frac{0}{10}$	$\frac{1}{10}$
2	$-.94$	$\frac{1}{10}$	$\frac{1}{10}$	0
3	$-.37$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$
4	$-.35$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{2}{10}$
5	$-.15$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{1}{10}$
6	$.35$	$\frac{2}{10}$	$\frac{4}{10}$	$\frac{2}{10}$
7	$.46$	$\frac{2}{10}$	$\frac{5}{10}$	$\frac{3}{10}$
8	$.73$	$\frac{2}{10}$	$\frac{6}{10}$	$\frac{4}{10}$
9	1.73	$\frac{2}{10}$	$\frac{7}{10}$	$\frac{5}{10}$
10	2.48	$\frac{3}{10}$	$\frac{7}{10}$	$\frac{4}{10}$
11	2.55	$\frac{3}{10}$	$\frac{8}{10}$	$\frac{5}{10}$
12	2.69	$\frac{3}{10}$	$\frac{9}{10}$	$\frac{6}{10}$
13	3.25	$\frac{4}{10}$	$\frac{9}{10}$	$\frac{5}{10}$
14	$\frac{5}{10}$	$\frac{9}{10}$	$\frac{9}{10}$	$\frac{4}{10}$
19	$\frac{7}{10}$	$\frac{9}{10}$	$\frac{3}{10}$	
15	8.60	$\frac{9}{10}$	$\frac{9}{10}$	
16	5.00	$\frac{9}{10}$	$\frac{1}{10}$	
17	7.74	$\frac{9}{10}$	0	
18	8.38	$\frac{9}{10}$	0	

$$\max_{i=1,\ldots,20} \left\{ \left| F_{10}\left(Z_{(i)}\right) - G_{10}\left(Z_{(i)}\right) \right| \right\} = \frac{6}{10}$$

```
> x=c(-0.15,8.6,5,3.71,4.29,7.74,2.48,3.25,-1.15,8.38)
> y=c(2.55,12.07,0.46,0.35,2.69,-0.94,1.73,0.73,-0.35,-0.37)
>
> library(NSM3)
> pKolSmirn(x,y,method="Exact")
Number of X values: 10 Number of Y values: 10
Kolmogorov-Smirnov J Statistic: 6
Exact upper-tail probability: 0.0524
> pKolSmirn(x,y,method="Asymptotic")
Number of X values: 10 Number of Y values: 10
Kolmogorov-Smirnov J* Statistic: 1.3416
Asymptotic upper-tail probability: 0.0546
```



Indicate some marginal evidence in the samples that feedback might have an effect on salivation rate.

Feedback group	No-Feedback group
−.15	2.55
8.60	12.07

$\binom{4}{2} = \frac{4!}{2!\,2!} = 6$

i	$Z_{(i)2}$	$F_2(z_{(i)2})$	$G_2(z_{(i)2})$	$\lvert F - G \rvert$
1	−.15	0.5	0	0.5
2	2.55	1	0	0.5
3	8.6	1	0.5	0.5
4	12.07	1	1	0

$\Rightarrow D = 0.5$

$\Rightarrow J = \frac{2 \times 2}{2} D = 1$

i	$Z_{(i)2}$	$F_2(z_{(i)2})$	$G_2(z_{(i)2})$	$\lvert F - G \rvert$
1	−.15	0.5	0	0.5
2	2.55	0.5	0.5	0
3	8.6	1	0.5	0.5
4	12.07	1	1	0

$\Rightarrow D = 0.5$

$J = 1$

i	$Z_{(i)2}$	$F_2(z_{(i)2})$	$G_2(z_{(i)2})$	$\lvert F - G \rvert$
1	−.15	0.5	0	0.5
2	2.55	0.5	0.5	0
3	8.6	0.5	1	0.5
4	12.07	1	1	0

$\Rightarrow D = 0.5$

$J = 1$

i	$Z_{(i)2}$	$F_2(z_{(i)2})$	$G_2(z_{(i)2})$	$\lvert F - G \rvert$
1	−.15	0	0.5	0.5
2	2.55	0.5	0.5	0
3	8.6	1	0.5	0.5
4	12.07	1	1	0

$\Rightarrow D = 0.5$

$J = 1$

i	$z_{(i)}$	$F_1(z_{(i)})$	$G_2(z_{(i)})$	$\lvert F-G \rvert$	
1	-.15	0	.5	.5	
2	2.55	.5	.5	0	$\Rightarrow D = .5$
3	8.6	.5	1	.5	$\Rightarrow D = 1$
4	12.07	1	1	0	

i	$z_{(i)}$	$F_1(z_{(i)})$	$G_2(z_{(i)})$	$\lvert F-G \rvert$	
1	-.15	0	.5	.5	
2	2.55	0	1	1	$\Rightarrow D = 1$
3	8.6	.5	1	.5	$\Rightarrow D = 2$
4	12.07	1	1	0	