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.

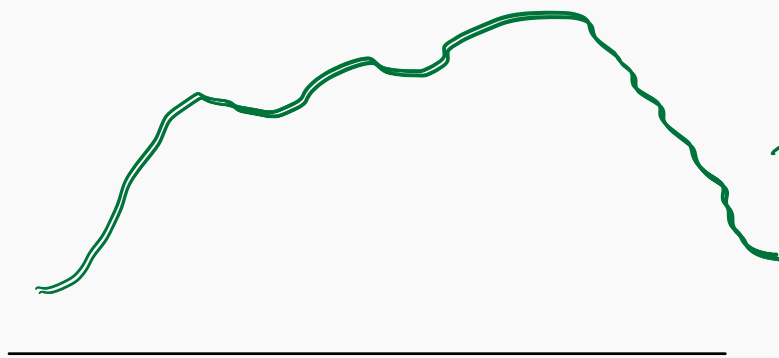
Table of contents

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.



pop 1 = pop 2

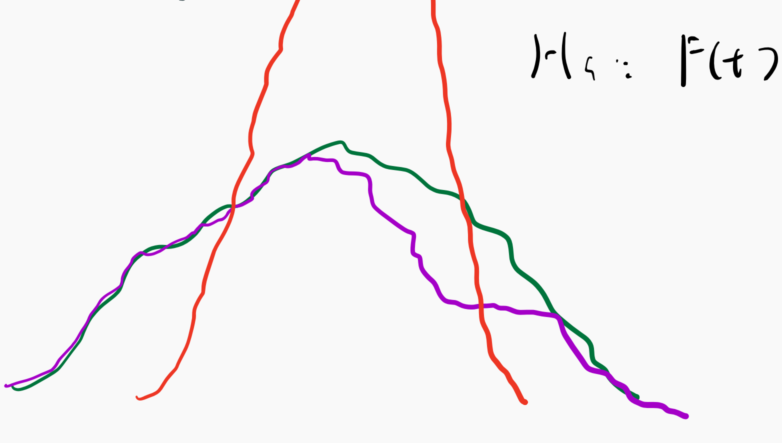
Let F be the distribution function corresponding to population 1 and let G be the distribution function corresponding to population 2.

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.

$H_1 : F(t) \neq G(t)$ for some t



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

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

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

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

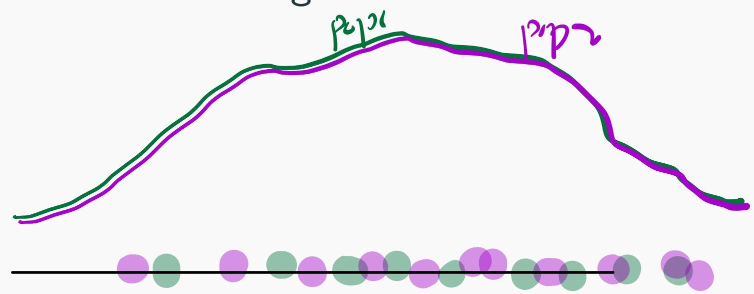
- $\boxed{\text{if } \Delta = 0}$:

  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" = similar ranks

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

rank the pooled sample from $1 \ldots \ldots n+m$

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

Motivation

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

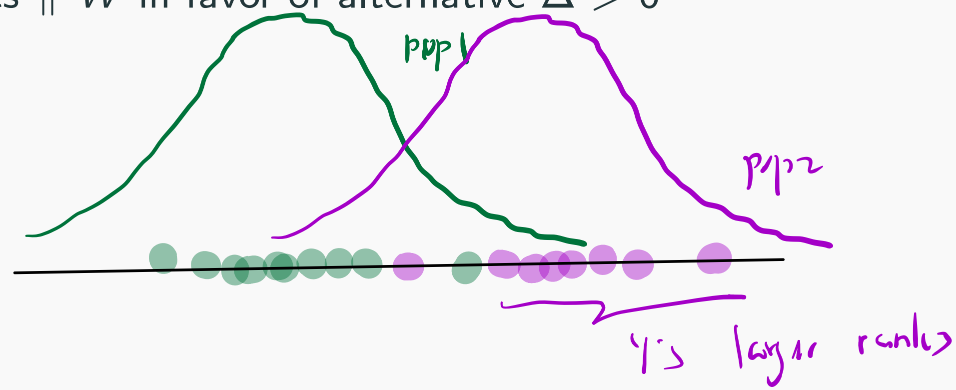
- $\boxed{\text{if } \Delta > 0}$:

  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

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

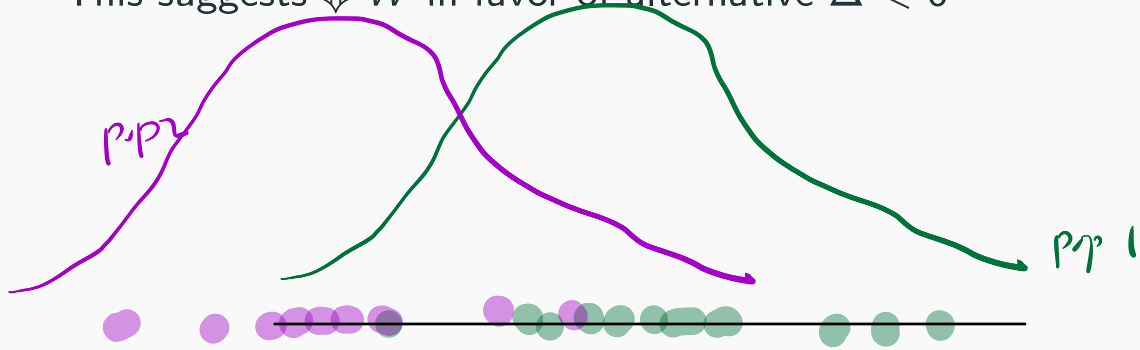
- $\boxed{\text{if } \Delta < 0 :}$

  Ys will tend to be $\Downarrow$ than Xs

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

  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

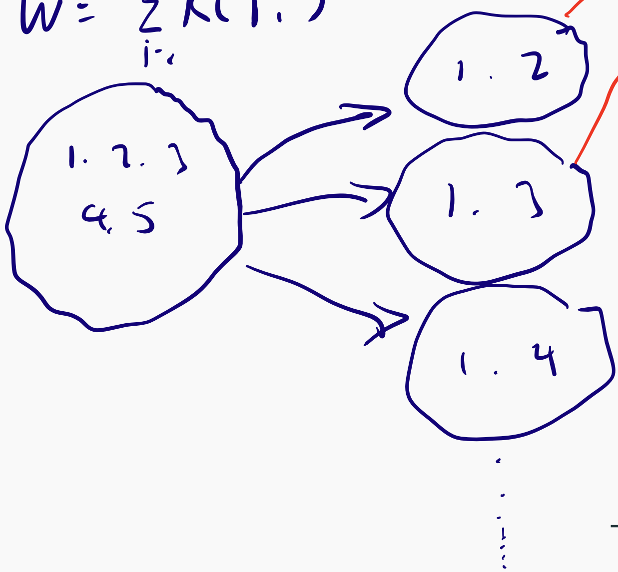
$$\boxed{\binom{n+m}{n}} = \frac{(n+m)!}{n!\,m!}$$

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

$m = 3, n = 2$

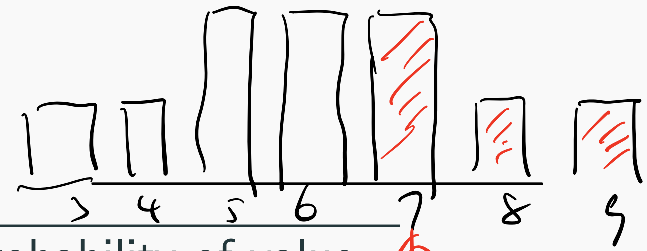
$$\binom{3+2}{2} = \binom{5}{2} = \frac{5!}{2!\,3!} = 10$$

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



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

$\Rightarrow$ null distribution of W :



W_{obs}

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

$$P_0(W \geq 7) = P_0(W = 7) + P_0(W = 8) + P_0(W = 9)$$
$$= .2 + .1 + .1 = .4$$

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

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

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

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

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

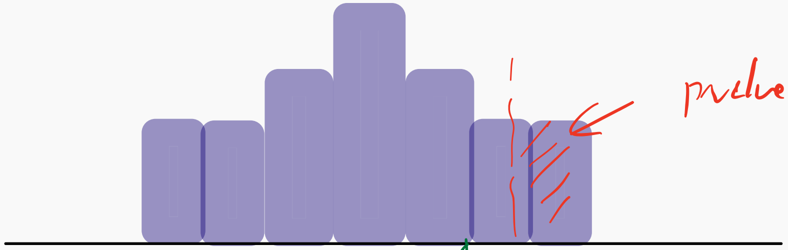
$$\Rightarrow EW = \frac{n(N+1)}{2}$$

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

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

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

Procedure



pvalue $= P(W \geq W_{obs})$

W null distr.

W_{obs}

$W_{0.05}$

a. One-Sided Upper-Tail Test.

To test

$$H_0 : \Delta = 0$$

versus

$$H_a : \Delta > 0$$

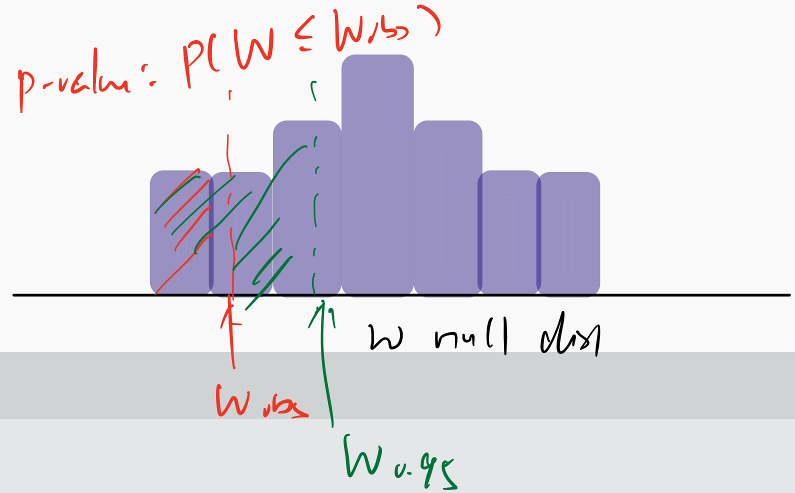
at the α level of significance,

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

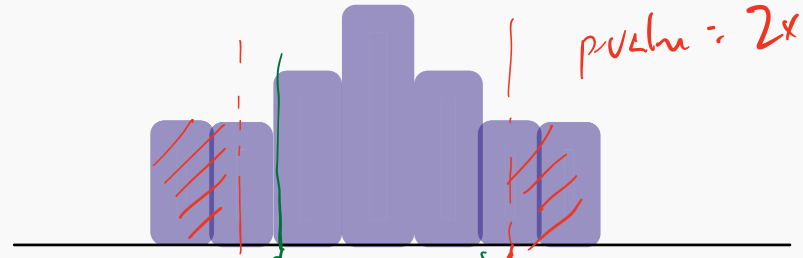
$$H_a : \Delta < 0$$

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

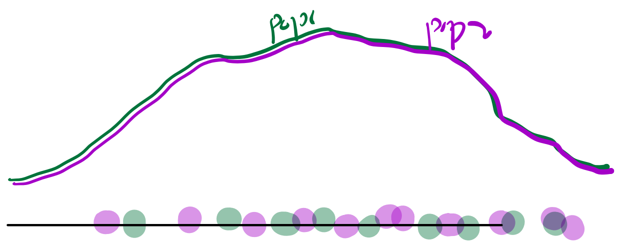
$$\begin{aligned} U = & \text{ number of pairs } \left(X_i, Y_j\right) \text{ for which } X_i < Y_j \\ = & \sum_{j=1}^{n} \sum_{i=1}^{m} 1\left\{X_i \leq Y_j\right\} \end{aligned}$$

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.

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

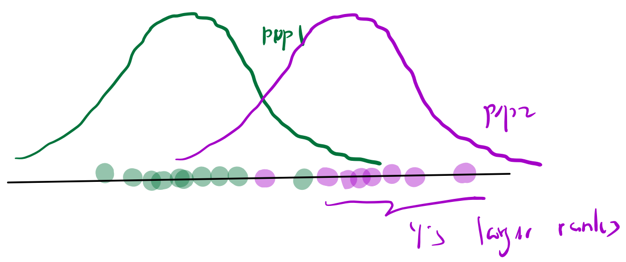
m X's n Y's $m \cdot n$

$\Delta : 0$



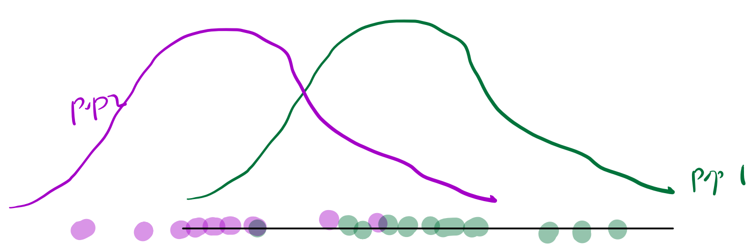
$U \approx \dfrac{m \cdot n}{2}$

$\Delta > 0$



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

$\Delta < 0$:



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

Mann-Whitney statistic = Wilcoxon rank sum statistic

The Mann-Whitney statistic can be shown to be equivalent to the Wilcoxon rank sum statistic. $W = \sum_{j=1}^{n} R(Y_j)$

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

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

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 $Y' \le Y_j = j$,

$$\sum_{j\prime=1}^{n} 1\left\{Y_{j\prime} \le Y_j\right\} = j$$

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

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

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

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

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

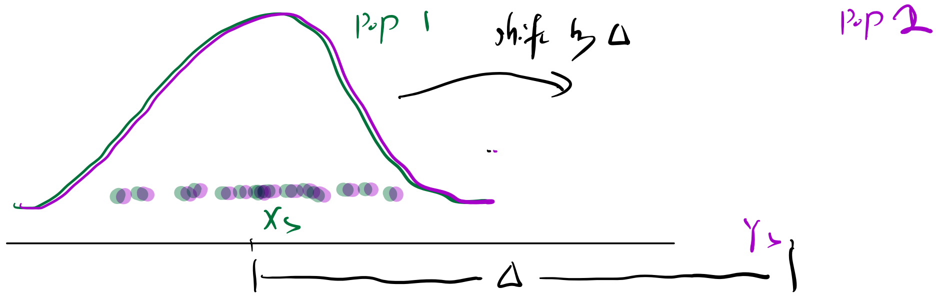
$$\Rightarrow \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} 1\left\{\Delta \leq Y_j - X_i\right\} \approx \frac{1}{2}$$

$\Rightarrow \Delta$ 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
Xs
Ys
Δ

Procedure

Estimate of Δ

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

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

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

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

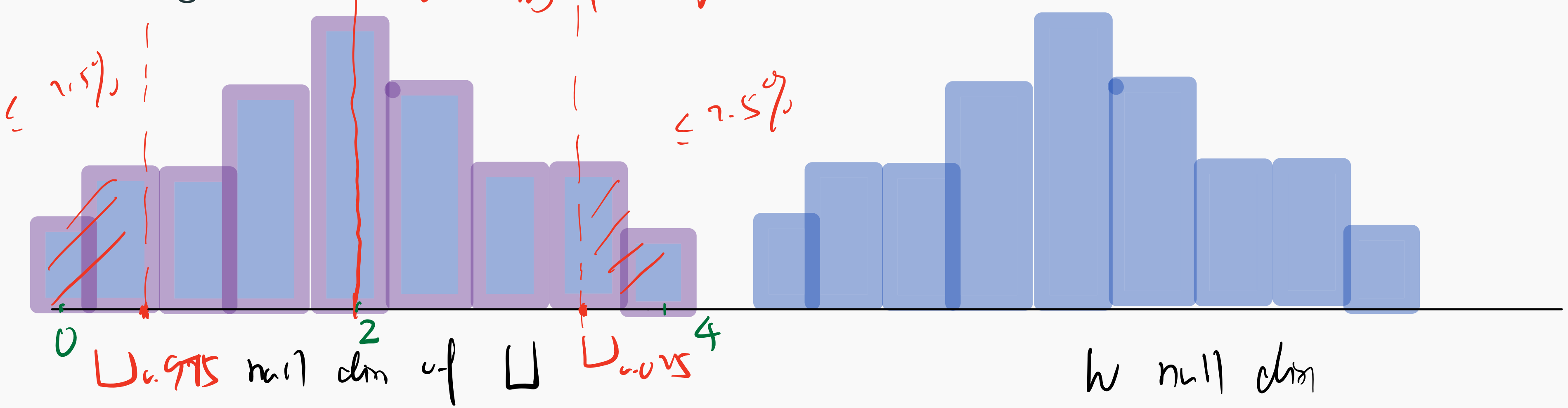
  the value that occupies the position $k + 1$ in the list of the ordered $Y - X$ differences.
- if mn is even, say $mn = 2k$, then $k = mn/2$ and

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

  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.



$$P(\underbrace{U_{0.975}}_{1} \le U \le \underbrace{U_{0.025}}_{3}) \approx 95\% \qquad W = U + \frac{n(n+1)}{2}$$

$m: 2 \quad n: 2$

$$U = \sum_i \sum_j \mathbb{1}(X_i \le Y_j - \Delta)$$

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

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

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

$$U_{1-\alpha/2} = nm + 1 - \boxed{U_{\alpha/2}}. \approx U_{0.025}$$

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

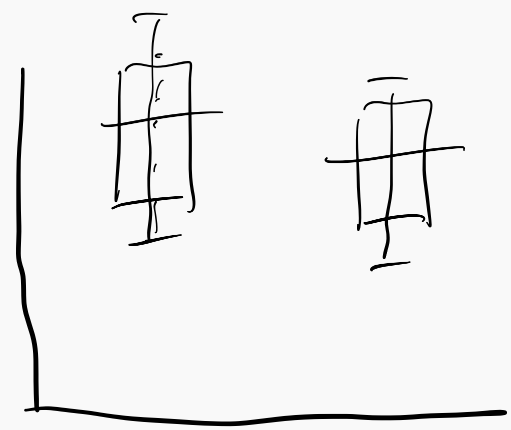
Eriksen, Björnstad, and Götestam (1986) studied a social skills training program for alcoholics. ~~Twenty-four~~ 23 "alcohol-dependent" male inpatients at an alcohol treatment center were randomly assigned to two groups. The control group patients were given a traditional treatment program. The treatment 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	($\cancel{3}$ 4)	874	(3)
1617	($\cancel{5}$ 6)	389	(2)
1180	($\cancel{4}$ 5)	213	(1)

$W = \cancel{3}\ 6$

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

Y-ranks	Probability	W
1, 2, 3		6
1, 2, 4		7
1, 2, 5		8
1, 2, 6	$\frac{1}{w}$ = n of	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	0.05
8	0.1
9	0.15
10	0.15
11	0.15
12	0.15
13	0.1
14	0.05
15	0.05

$$\text{p-value} = P(W \le 6) = 0.05$$

(Y − X) differences:

Y	X	Y − X	
874	1042	−168	(9)
	1617	−743	(6)
	1180	−306	(8)
389	1042	−653	(7)
	1617	−1228	(2)
	1180	−791	(5)
212	1042	−829	(4)
	1617	−1404	(1)
	1180	−967	(3)

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

90% Confidence interval: $U_{0.05}$

$$W = U + \frac{n(n+1)}{2} = U + \underbrace{\frac{3\times 4}{2}}_{6}$$

$$W_{0.05} = 15$$

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

$$\Rightarrow U_{1-\frac{\alpha}{2}} = 3\times 3 + 1 - 9 = 1$$

$\Rightarrow$ 90% C.I. for $\hat{\Delta}$:

$$U_{(1)}, \; U_{(9)}$$

=

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

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

1. 1... 11

$$W = \sum_{i=1}^{n} R(Y_i) = 81$$

$W = U + \underbrace{\frac{n(n+1)}{2}}_{\frac{11 \times 12}{2} = 66}$

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

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

```
> 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
```