Section 4 : Permutation Inference

Yotam Shem-Tov

Fall 2014

Wilcoxons signed rank test (WSRT)

- Wilcoxons signed rank test is used for matched pair experiments, or data designs
- Let $S = N/2$ be the number of pairs, in paired randomized experiment
- Let r_{is} be the outcome of unit i in strata s. The outcome of each observation r_{si} can have many values, $r_{si} \in \mathbb{R}$.
- In each strata s there are two observations $n_s = 2$ and one is assigned to treatment and the other to control, $m_s = 1$
- When should we use WSRT instead of WRST? We want to test the question:

  Is the barley yields of a field in 1931 and 1932 are the same?

```
> library(MASS) # load the MASS package
> head(immer) # the data set
```

Wilcoxons signed rank test

- The Wilcoxons signed rank test statistic is developed as follows,
- compute $|r_{s1} - r_{s2}|$
- Let $d_{si} = rank\left(|r_{s1} - r_{s2}|\right)$
- Let $c_{si} = 1$ if $r_{si} > r_{sj}$ and 0 otherwise, for $i, j \in \{1, 2\}$ and $i \neq j$
- The expression, $\sum_{i=1}^{2} Z_{si} c_{si}$ equals 1 if the treated unit in pair s had a higher response than the control unit, and 0 otherwise
- The test statistic is the sum of the ranks for pairs in which the treated unit had a higher response than the control unit:

$$t(\mathbf{Z}, \mathbf{r}) = \sum_{s=1}^{S} d_{si} \cdot \sum_{i=1}^{2} Z_{si} c_{si}$$

- Are d_s and c_{si} fixed or random under H_0? *No*

Wilcoxons signed rank test: Example

- Let the control group be the chicks in *Confinement*, and treatment is *OpenRange*

```
library(PairedData)
data(ChickWeight)

> head(ChickWeight)
Chicks Confinement OpenRange
1     C01            9         8
2     C02           17        15
3     C03           14        11
4     C04           13        11
5     C05           15         9
6     C06           10        12
```

Wilcoxons signed rank test: Example

- how many different permutations of treatment are possible? $2^{10} = 1024$
- If it was not a paired data set, what is the number of different possible allocations of treatment? $\binom{N}{m} = \binom{20}{10} = 184756$
- Calculate the Wilcoxon sign rank test statistic in R

```
sum_Z_is_c_si <- (OpenRange>Confinement)*1 # c_si:
d_si <- rank(abs(OpenRange-Confinement)) # d_si:
statistic <- sum(d_si*sum_Z_is_c_si)
```

Wilcoxons signed rank test: Example

What is the R function wilcox.test does?

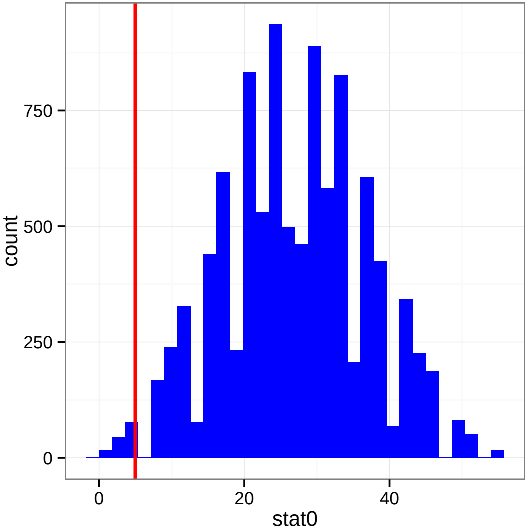
```
> wilcox.test(OpenRange,Confinement,paired=TRUE)

Wilcoxon signed rank test with continuity correction

data:  OpenRange and Confinement
V = 4, p-value = 0.03205
alternative hypothesis: true location shift is not equal to

Warning messages:
1: In wilcox.test.default(OpenRange, Confinement, paired =
  cannot compute exact p-value with ties
2: In wilcox.test.default(OpenRange, Confinement, paired =
  cannot compute exact p-value with zeroes
```

Wilcoxons signed rank test: Example

```
### Permutation diftribution under the null
L = 10000
Y = ChickWeight[,c(2,3)]
stat0 <- rep(999,L)
for (i in c(1:L)){
  OpenRange0 <- rep(999,10)
  Confinement0 <- rep(999,10)
  for(j in c(1:10)){
    id0 <- sample(c(2,3),1)
    OpenRange0[j] <- Y[j,c(2,3) %in% id0]
    Confinement0[j] <- Y[j,!c(2,3) %in% id0]
  }
  sum_Z_is_c_si0 <- (OpenRange0>Confinement0)*1 # c_si:
  d_si0 <- rank(abs(OpenRange0-Confinement0)) # d_si:
  stat0[i] <- sum(d_si0*sum_Z_is_c_si0)
}
```

Wilcoxons signed rank test: Example

permutation distribution



Wilcoxons signed rank test: Example

The P-value according to the permutation distribution we calculated

```
> ### P-value
> min(sum(statistic<=stat0)/L,sum(statistic>=stat0)/L)*2
[1] 0.0278
```

Is there a difference between our results and the wilcox.test function results? Yes, what can explain the difference?

Asymptotic approximation: The CLT

The central limit theorem (CLT) is most useful when considering asymptotic approximations

The CLT:

Suppose $X_1, \dots, X_N$ is a sequence of i.i.d random variables with $\mathbb{E}(X) = \mu$ and $\mathbb{V}(X) = \sigma^2 < \infty$. Then,

$\frac{\sum_{i=1}^{N} X_i - N\mu}{\sigma\sqrt{N}}$ is approximately distributed standard normal, $N(0,1)$

and

$\frac{\bar{X}-\mu}{\sigma/N}$ is approximately distributed standard normal, $N(0,1)$

Asymptotic approximation: technical note

- When should we use a continuity correction?
  Answer: When a discrete distributions supported on the integers are approximated by a continuous distribution, such as the Normal distribution

- WRST and WSRT test statistics are both supported on the integers (assuming no ties)
  Is the KS test statistic supported on the integers? *No*

Asymptotic approximation

- The CLT will be the basis for most of the asymptotic approximations
- In many of the tests (WRST, WSRT, difference in means) the test statistic is a sum
- In order to use an asymptotic approximation we need first to calculate, $\mathbb{E}(W)$ and $\mathbb{V}(W)$
- If we know μ and σ^2 and the CLT can be applied, the asymptotic approximation is simple

Asymptotic approximation: technical note

Recall the following equalities:

$$\sum_{i=1}^{N} i = \frac{N(N+1)}{2}, \quad \sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} i \cdot j = \frac{N^2(N+1)^2}{4}$$

Hence,

$$\sum_{i=1}^{N}\sum_{j\neq i}^{N} i \cdot j = \frac{N^2(N+1)^2}{4} - \frac{N(N+1)(2N+1)}{6}$$

This equalities can be proved using induction.

Asymptotic approximation: Wilcoxons signed rank test

- What is the expectations of $t(\mathbf{Z}, \mathbf{r})$ under the null?

$$
\begin{aligned}
\mathbb{E}t(\mathbf{Z}, \mathbf{r}) &= \mathbb{E}(\sum_{s=1}^{S} d_{si} \cdot \sum_{i=1}^{2} Z_{si} c_{si}) = \sum_{s=1}^{S} d_{si} \cdot \mathbb{E}(\sum_{i=1}^{2} Z_{si} c_{si}) \\
&= \sum_{s=1}^{S} d_{si} \cdot P(\sum_{i=1}^{2} Z_{si} c_{si} = 1) = \sum_{s=1}^{S} d_{si} \cdot \frac{1}{2} \\
&= \sum_{s=1}^{S} i \cdot \frac{1}{2} = \frac{S(S+1)}{4}
\end{aligned}
$$

Asymptotic approximation: Wilcoxons signed rank test

- What is the variance of the test statistic under the null?

$$
\begin{aligned}
\mathbb{V}(t(\mathbf{Z},\mathbf{r})) &= \mathbb{V}(\sum_{s=1}^{S} d_{si} \cdot \sum_{i=1}^{2} Z_{si} c_{si}) = \sum_{s=1}^{S} d_{si}^2 \cdot \mathbb{V}(\sum_{i=1}^{2} Z_{si} c_{si}) \\
&= \sum_{s=1}^{S} d_{si}^2 \cdot P(\sum_{i=1}^{2} Z_{si} c_{si} = 1) \cdot \left(1 - P(\sum_{i=1}^{2} Z_{si} c_{si} = 1)\right) \\
&= \sum_{s=1}^{S} d_{si}^2 \cdot \frac{1}{2} \cdot \frac{1}{2} = \sum_{s=1}^{S} i^2 \cdot \frac{1}{4} = \frac{S(S+1)(2S+1)}{6} \cdot \frac{1}{4} \\
&= \frac{S(S+1)(2S+1)}{24}
\end{aligned}
$$

Hence, when $N \to \infty$, $\frac{t(\mathbf{Z},\mathbf{r}) - \mathbb{E}(\cdot)}{\mathbb{V}(\cdot)} \xrightarrow{D} N(0,1)$

Asymptotic approximation: Wilcoxons rank sum test

- What is the expectation of the test statistic under the null?

$$
\begin{aligned}
\mathbb{E}(t(\mathbf{Z},\mathbf{r})) &= \mathbb{E}\left(\sum_{i=1}^{N} Z_i q_i\right) = \sum_{i=1}^{N} \mathbb{E}\left(Z_i q_i\right) \\
&= \sum_{i=1}^{N} q_i \mathbb{E}\left(Z_i\right) = \sum_{i=1}^{N} q_i P\left(Z_i\right) = \sum_{i=1}^{N} q_i \frac{m}{N} \\
&= \frac{m}{N} \sum_{i=1}^{N} i = \frac{m}{N} \cdot \frac{N(N+1)}{2} = \frac{m(N+1)}{2}
\end{aligned}
$$

- Recall that,

$$
\mathbb{V}\left(\sum_{i=1}^{N} a_i\right) = \sum_{i=1}^{N} \sum_{j=1}^{N} Cov(a_i, a_j) = \sum_{i=1}^{N} \mathbb{V}(a_i) + \sum_{i=1}^{N} \sum_{j \neq i} Cov(a_i, a_j)
$$

Asymptotic approximation: WRST

- What is the variance of the test statistic under the null?
- We will use this problem in order to demonstrate a new and more general approach
- A derivation of the variance in a similar way as was used in the case of WSRT can be found in,

  `http://www.real-statistics.com/non-parametric-tests/wilcoxon-rank-sum-test/wilcoxon-rank-sum-test-advanced/`

Asymptotic approximation: WRST

- Let a population consist of N distinct numbers, $v_1, \ldots, v_N$
- The distinct numbers assumptions is equivalent to assuming no ties
- Let one of the numbers be selected at random and denote it by V.

$$\mathbb{E}(V) = \frac{v_1 + \cdots + v_N}{N} = \bar{v}$$

$$\mathbb{V}(V) \equiv \tau^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} (v_i - \bar{v})^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} v_i^2 - \bar{v}^2$$

$$= \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4} = \ldots = \frac{(N-1)^2}{12}$$

Asymptotic approximation: WRST

- Assume the m units of the v's are selected at random, such that all $\binom{N}{m}$ possible are equally likely $\Leftrightarrow$ m units are selected at random with equal probabilities
- Let $V_1, \dots, V_m$ denote the m selected v values, and let,

$$T = V_1 +, V_2, \dots, + V_m$$

- The variance of T is exactly the variance of the Wilcoxon rank sum test statistic
- Note, $Cov(V_i, V_j) = \lambda$ for all i and j. What is λ?

Asymptotic approximation: WRST

- The variance of T is,

$$\mathbb{V}(\sum_{i=1}^{m} V_i) = \sum_{i=1}^{m} \mathbb{V}(V_i) + \sum_{i=1}^{m} \sum_{j \neq i} Cov(V_i, V_j)$$

$$= m\tau^2 + m(m-1)\lambda \quad (1)$$

- If we select the hull population, i.e $m = N$ the variance of T is <u>zero</u>. Hence,

$$\mathbb{V}(V_1 + \cdots + V_N) = N\tau^2 + N(N-1)\lambda = 0 \Rightarrow \lambda = -\frac{\tau^2}{N-1} \quad (2)$$

- This "trick" allows us a simple way to derive the covariance between each two sampled values

Asymptotic approximation: WRST

- Substituting equation 2 in equation 1 yields (after some simplification),

$$\mathbb{V}(T) = \frac{m(N-m)}{N-1} \cdot \tau^2 = \frac{m(N-m)}{N-1} \cdot \underbrace{\frac{(N-1)^2}{12}}_{\tau}$$

$$= \frac{m(N-m)}{N-1} \cdot \frac{(N-1)(N+1)}{12}$$

$$\Rightarrow \mathbb{V}(T) = \frac{m(N-m)(N+1)}{12} \tag{3}$$

- Hence, when $N \to \infty$, $\frac{T - \mathbb{E}(\cdot)}{\mathbb{V}(\cdot)} \xrightarrow{D} N(0,1)$

Estimation using permutation inference

- How can we make estimation of treatment effect using permutation inference?
- Do we need to assume a model, i.e a functional form of the treatment effect?
  Yes! In order to make estimation, and construct confidence intervals we must have a model of the treatment effect
- Examples of different models are:

$$Y_{i1} = Y_{i0} + \tau$$

$$Y_{i1} = Y_{i0} \cdot \tau$$

$$Y_{i1} = \begin{cases} Y_{i0} + \tau, & Y_{i0} \geq 0 \\ Y_{i0} & Y_{i0} < 0 \end{cases}$$

Confidence intervals

- What do we gain by assuming a model of the treatment effect? *Given one of the potential outcomes either Y_{i1} or Y_{i0} and τ_0, than under the null hypothesis that $H_0 : \tau = \tau_0$ we can calculate the non-observed potential outcome*
- Define Y_i^d as the adjusted response.
- Our objective is to define Y_i^d using the knowledge we have on the treatment effect model such that:

$$Y^d \perp T$$

Confidence intervals

- Assume the true model of the treatment effect, is an additive treatment:

$$Y_{i1} = Y_{i0} + \tau$$

- Let $Y_i^d \equiv Y_i - \tau_0 \cdot T_i$
- Claim: *Under the null hypothesis that: $H_0 : \tau = \tau_0$, the adjusted responses are independent of the treatment assignment*
- Proof:

$$Y_i^d = \begin{cases} Y_{i0}, & \text{if} \quad T_i = 1 \\ Y_{i1} - \tau_0 \cdot 1 = Y_{i0} + \tau - \tau = Y_{i0}, & \text{if} \quad T_i = 0 \end{cases}$$

- Therefore under the sharp null, Y_i^d is independent of the treatment assignment. Hence, the distribution of Y_i^d in the treatment group and control group are the same under the null

Confidence intervals

- How can we construct a confidence interval for τ?
- We will use the method of inverting hypothesis tests
- What is the definition of a confidence interval (confidence set)?
- It is all the values, τ_0, for which we **cannot** reject the null hypothesis that $\tau = \tau_0$

Confidence intervals

- Consider a two sided hypothesis test,

$$H_0 : \tau = \tau_0$$

$$H_1 : \tau \neq \tau_0$$

- All the values of τ_0 for which we cannot reject the null hypothesis, that $\tau = \tau_0$, are in a two sided confidence interval
- Consider a one sided hypothesis test,

$$H_0 : \tau \leq \tau_0$$

$$H_1 : \tau \geq \tau_0$$

- All the values of τ_0 for which we cannot reject the null that $H_0 : \ \tau \leq \tau_0$, should be included in a one-sided confidence set

Confidence intervals: Summary

- A $1 - \alpha$ confidence set is the set of hypothesized values of a parameter not rejected by a level α test
- Let A be the set of all values of τ not rejected with a significance level α test:

$$Pr(\textit{in the acceptance region when testing } \tau) \geq 1 - \alpha$$

This implies that,

$$\Rightarrow Pr(\textit{in the rejection region when testing } \tau) \leq \alpha$$

Example

- The data is,

  ```
  t=c (12,12,12.9,13.6,16.6,17.2,17.5,18.2,19.1,
  19.3,19.8,20.3,20.5,20.6,21.3,21.6,22.1)
  c=c(5,5.4,6.1,10.9,11.8,12,12.3,14.8,15,16.8,
  17.2,17.2,17.4,17.5,18.5,18.7,18.7,19.2)
  ```

- The treatment group is t and the control group is c
- We want to estimate a confidence interval (set) for τ assuming an additive treatment effect model, i.e $Y_{i1} = Y_{i0} + \tau$
- What are the steps we need to do?

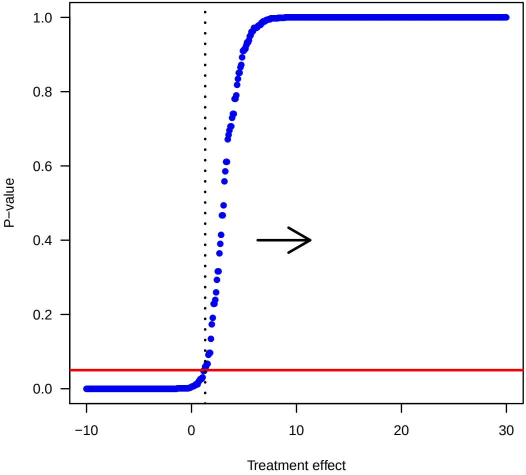
Example: code

- The code for calculating a one-sided confidence set

```
### calculate a one-sided confidence interval:
L = 500
tau.gride = seq(-10,30,length=L)
pv.gride = rep(999,length(tau.gride))

for (j in c(1:length(tau.gride))){
  pv.gride[j] = wilcox.test(t-tau.gride[j],c,
  exact=FALSE,alternative="greater")$p.value
}
```

- What does it mean that we choose the option "exact" in the R function *wilcox.test*? Calculating WRST using a Normal approximation

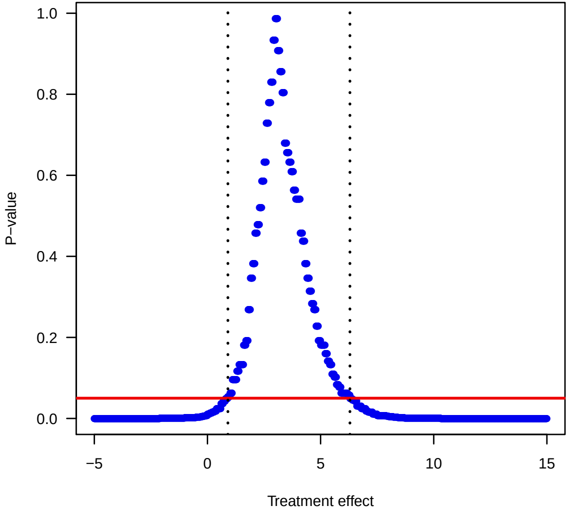
Example: Confidence set illustration



Example: summary

- The one-sided confidence set is: $[1.3, \infty]$
- The t-test one-sided confidence interval is, $[1.4, \infty)$
- Does the two tests coincide?

Example: Two-sided confidence set illustration



Example: summary

- How can we get a point estimate?
- One option (my preferred) is the value of τ_0, which has the heights P-value
- Another option (very similar in practice) is the HodgesLehmann estimator

```
wilcox.test(t,c,exact=FALSE,conf.int=TRUE)
# or
wilcox.test(t,c,exact=FALSE,conf.int=TRUE)$estimate
```

Links for farther reading

- One of the classic text books on non-parametric statistical inference is,
  Nonparametrics: Statistical Methods Based on Ranks
  Erich L. Lehmann
- A good and formal description of permutation inference and permutation tests is in:
  Permutation Tests for Complex Data: Theory, Applications and Software
  Fortunato Pesarin, Luigi Salmaso
- This is the classic test book for bootstrap and chapter 20 describes permutation tests and discusses the difference between permutation tests and bootstrap
  An Introduction to the Bootstrap
  Bradley Efron, R.J. Tibshirani