Two-sample Nonparametric Tests

Jyotishka Datta

Updated: 2023-10-09

Example

The following are 12 independent observations of pouring times of a metal grouped into before and after lunch. The metallurgical engineer suspects that the pouring time **before lunch was shorter than the pouring time after lunch**. We want to find the P-value for the alternative that the median pouring time before lunch is indeed shorter.

```
x = c(12.6,11.2,11.4,9.4,13.2,12)
y = c(16.4,15.4,14.1,14,13.4,11.3)
```

- As you have seen before, specifying the correct alternative decides how the p-value is calculated. Here, we have:

$$H_0 : \mu_{before} = \mu_{after}$$
$$H_1 : \mu_{before} < \mu_{after}$$

Mind the alternative

- The `R` function is `wilcox.test` but pay attention to the `paired = F` argument. It tells us that the two series are not paired, i.e. x_1 and y_1 might not be observed for the same individual.

- The second thing to note here is that since we specify `x` before `y`, the alternative should be `less`, not `greater`.

- Note that the `W` value here is 4. This is the Mann-Whitney U.

```r
wilcox.test(x,y,paired=F,alternative="less")
```

```
##
##      Wilcoxon rank sum exact test
##
## data:  x and y
## W = 4, p-value = 0.01299
## alternative hypothesis: true location shift is less than 0
```

Mind the alternative

- If you switch the order, i.e. specify `y` then `x` then you should specify the alternative as `greater` since $y > x$ is what we want to test.

```
wilcox.test(y,x,paired=F,alternative="g")
```

```
##
##     Wilcoxon rank sum exact test
##
## data:  y and x
## W = 32, p-value = 0.01299
## alternative hypothesis: true location shift is greater than 0
```

- Note that the `W` value here is 32. This is the Mann-Whitney U (but in the other direction!)

R code returns all the values

- You can get hold of all the internal calculations from the R function as well. Here these are stored in the `twotest` object that where we assign the outcome of the `wilcox.test` function.

```
twotest <- wilcox.test(y,x,paired=F,alternative="greater")
names(twotest)
```

```
## [1] "statistic"   "parameter"   "p.value"     "null.value"  "alternative"
## [6] "method"      "data.name"
```

- For example, you can just extract the `statistic` or the `p.value`.

```
twotest$statistic
```

```
##  W
## 32
```

```
twotest$p.value
```

```
## [1] 0.01298701
```

Apply CLT manually

- We can also apply CLT maually. We need the large sample formula for mean and standard deviation for the CLT result.

- Note that the p-value is calculated using `1 - pnorm(z)` which is $P(Z \geq z)$ since our alternative is "greater".

```
U_Y = wilcox.test(y,x,paired=F,alternative="greater")$statistic
m = length(y)
n = length(x)
clt.mu = m*n/2
clt.s2 = m*n*(m+n+1)/12

(Z_score = (U_Y-clt.mu-0.5)/(sqrt(clt.s2))) # Continuity correction,
```

```
##        W
## 2.16173
```

```
(P.value = 1 - pnorm(Z_score)) # P(Z > z) since alt = "greater"
```

```
##          W
## 0.01531949
```

Apply CLT manually

- We can apply CLT with `y` first and `x` second and a "less" alternative too.
- But then the p-value should be calculated using `pnorm(z)` which is $P(Z \leq z)$ since our alternative is now "less".

```
U_X = wilcox.test(x,y,paired=F,alternative="less")$statistic
m = length(x)
n = length(y)

(Z_score = (U_X-clt.mu)/(sqrt(clt.s2))) # continuity correction, add
```

```
##          W
## -2.241794
```

```
(P.value = pnorm(Z_score)) # P(Z <= z) since alt = "less"
```

```
##          W
## 0.01248734
```

Using formula

- Instead of specifying the two "vectors", we can also use formula in a data.frame.
- It should be of the form: `value ~ grp` where `value` column has all the values for both groups and `grp` is a binary categorical variable.

```
xydata = rbind(data.frame(values = x, grp = "x"),
              data.frame(values = y, grp = "y"))
wilcox.test(values ~ grp, data = xydata, alternative = "less")
```

```
##
##      Wilcoxon rank sum exact test
##
## data:  values by grp
## W = 4, p-value = 0.01299
## alternative hypothesis: true location shift is less than 0
```

Class data

- Read the simulated class data from the OpenIntro website. We can directly read from the URL without downloading it.

  > "An instructor decided to run two slight variations of the same exam. Prior to passing out the exams, they shuffled the exams together to ensure each student received a random version. Anticipating complaints from students who took Version B, they would like to evaluate whether the difference observed in the groups is so large that it provides convincing evidence that Version B was more difficult (on average) than Version A."

- The dataset has three exams: a, b, c. We will only consider the first two exams "a" and "b".

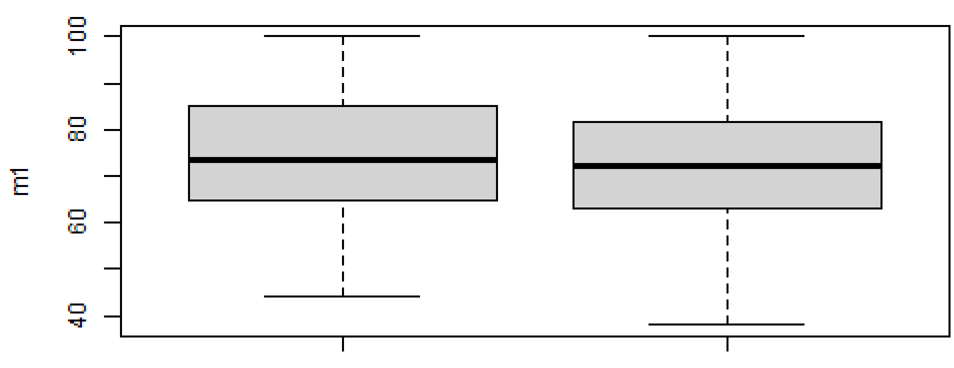
```
classdata <- read.csv("https://www.openintro.org/data/csv/classdata.c
classdata_ab <- classdata[classdata$lecture %in% c("a", "b"),]
```

Visualize & Summarize

```r
library(dplyr)
classdata_ab %>% group_by(lecture) %>% summarize(mean = mean(m1), me
```

```
## # A tibble: 2 × 5
##   lecture  mean median    sd     n
##   <chr>   <dbl>  <dbl> <dbl> <int>
## 1 a        75.1   73.5  13.9    58
## 2 b        72.0   72    13.8    55
```

```r
boxplot(m1 ~ lecture, data = classdata_ab)
```



Testing difference

- Test if the score for the two exams "a" and "b" had the same median?
- Note that we want to test if "b" was more difficult, i.e. your hypotheses are:

$$
\begin{aligned}
H_0 &: \mu_a = \mu_b \\
H_1 &: \mu_a > \mu_b
\end{aligned}
$$

```
## Your code here!
```

- Take 5 minutes!