HW2 - Lucas Fellmeth, Sven Bergmann

2023-10-05

Chapter 7

Problem 3

Diet A was given to a group of 10 overweight boys between the ages of 8 and 10. Diet B was given to another independent group of 8 similar overweight boys. The weight loss is given in the table below. Using WMW test, test the hypothesis that the diets are of comparable effectiveness against the two-sided alternative. Use $\alpha = 5\%$ and normal approximation.

Diet A	7	2	3	-1	4	6	0	1	4	6
Diet B	5	6	4	7	8	9	7	2		

Solution

We test for $H_0 : F_A(x) = F_B(x)$.

First we find

$$W_n = \sum_{n=1}^{n_1} R_i$$

We reject when $W_n \geq k_{0.05}$.

```r
# rm(list=ls())
a <- c(7, 2, 3, -1, 4, 6, 0, 1, 4, 6)
b <- c(5, 6, 4, 7, 8, 9, 7, 2)

n_1 <- length(a)
n_2 <- length(b)
c <- c(a, b)
n <- length(c)
rnk <- rank(c)
W_n <- sum(rnk[1:n_1])  # sum of the ranks of a
U <- W_n - (n_1 * (n_1 + 1)/2)
```

Since there are ties in the ranks

$$W_n \sim \mathcal{N}\left(\frac{n_1(n+1)}{2}, \frac{n_1 n_2 (n+1)}{12} - \frac{n_1 n_2 \sum_{i=1}^{k}(t_i^3 - t_i)}{12n(n+1)}\right)$$

where $t_1, \ldots, t_k$ are the number of different observations among all the observations in the combined sample.

We reject if

$$\frac{W_n - \frac{n_1(n+1)}{2}}{\sqrt{\frac{n_1 n_2 (n+1)}{12} - \frac{n_1 n_2 \sum\limits_{i=1}^{k}(t_i^3 - t_i)}{12n(n+1)}}} > z_{0.05}$$

$$\Leftrightarrow W_n > \underbrace{z_{0.05} * \left(\sqrt{\frac{n_1 n_2 (n+1)}{12} - \frac{n_1 n_2 \sum\limits_{i=1}^{k}(t_i^3 - t_i)}{12n(n+1)}} \right) + \frac{n_1(n+1)}{2}}_{k_0.05}$$

```
t <- tabulate(rnk)
t <- t[t != 0]
t_sum <- sum(t^3 - t)

lower <- qnorm(1 - 0.05) * sqrt((n_1 * n_2 * (n + 1)/12) - (n_1 * n_2 * t_sum)/(12 *
    n * (n + 1))) + (n_1 * (n + 1)/2)

if (W_n > lower) {
    print("We reject the null hypothesis for the alternative.")
} else {
    print("We accept the null hypothesis.")
}
```

```
## [1] "We accept the null hypothesis."
```

If we choose the same test implemented in R we somehow get a different result. See below.

```
wilcox.test(x = a, y = b, alternative = "two.sided", conf.level = 0.95, exact = F)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  a and b
## W = 16.5, p-value = 0.03964
## alternative hypothesis: true location shift is not equal to 0
```

Problem 6

Use the link below to see the results of an experiment on the effect of prior information on the time to fuse random dot stereograms. One group (NV) was given either no information or just verbal information about the shape of the embedded object. A second group (VV) received both verbal information and visual information (e.g. a drawing of the object). Does the median time prove to be greater for the NV group? Compare your results to those from a two-sample t-test.

Solution

This time, our null hypothesis is:

$$H_0 : \mu_{NV}^2 = \mu_{VV}^2 \text{ vs } H_1 : \mu_{NV}^2 > \mu_{VV}^2$$

```r
# rm(list = ls())
stereogram_data <- read.delim2("stereograms.txt")
stereogram_data$fusion.time <- as.numeric(stereogram_data$fusion.time)
stereogram_data$group <- factor(stereogram_data$group)
table(stereogram_data$group)
```

```
##
## NV VV
## 43 35
```

```r
means <- aggregate(fusion.time ~ group, stereogram_data, mean)
print(means)
```

```
##   group fusion.time
## 1    NV    8.560465
## 2    VV    5.551429
```

Yes, the median time does prove to be greater for the NV group.

```r
t.test(x = subset(stereogram_data, subset = (group == "NV"))$fusion.time, y = subset(stereogram_data,
    subset = (group == "VV"))$fusion.time, alternative = "greater", var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  subset(stereogram_data, subset = (group == "NV"))$fusion.time and subset(stereogram_data, subs
## t = 2.0384, df = 70.039, p-value = 0.02264
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.5484007       Inf
## sample estimates:
## mean of x mean of y
##  8.560465  5.551429
```

The two-sample t-test confirms our observation. We reject H_0 for the alternative.

Problem 7

Derive the exact distribution of the Mann–Whitney U statistic in the case that $n_1 = 4$ and $n_2 = 2$.

Solution

The exact distribution can be found as

$$P(W_n = m) = \frac{k_{n_1,n_2}(m)}{\begin{pmatrix} n \\ n_1 \end{pmatrix}}, \frac{n_1(n_1+1)}{2} \leq m \leq \frac{n_1(2n - n_1 + 1)}{2}$$

with $k_{n_1,n_2}(m)$ the number of all arrangements of zeroes and ones in $(S_1(X,Y),\ldots,S_n(X,Y))$ such that

$$W_n = \sum_{i=1}^{n} i \cdot S_i(X,Y) = m.$$

We have

$$n_1 = 4, n_2 = 2, n = n_1 + n_2 = 6$$

so there are $\begin{pmatrix} 6 \\ 4 \end{pmatrix} = \begin{pmatrix} 6 \\ 2 \end{pmatrix} = 15$ distinguishable configurations of the vector $(S_1, S_2, \ldots, S_6)$.

The minimum of W_6 is 10 and the maximum is 18.

Then we get the following distribution:

Table 2: Distribution of W_6 with $n_1 = 4$ and $n_2 = 2$

W_6	Configuration	Probability
10	$(1,2,3,4)$	1/15
11	$(1,2,3,5)$	1/15
12	$(1,2,3,6),(1,2,4,5)$	2/15
13	$(1,2,4,6),(1,3,4,5)$	2/15
14	$(1,3,4,6),(2,3,4,5),(1,2,5,6)$	3/15
15	$(2,3,4,6),(1,3,5,6)$	2/15
16	$(2,3,5,6),(1,4,5,6)$	2/15
17	$(2,4,5,6)$	1/15
18	$(3,4,5,6)$	1/15