SHENDUO ZHANG

Small sample Wilcoxon Signed-rank test(Exact test)

The actual change scores on the electroretinogram (ERG), a measure of electrical activity in the retina are studied. It's record before and after a certain surgery for retinitis pigmentosa(Berson et al.,). Evaluate the significance of the results without assuming the change scores are normally distributed? What do the results indicate? *(Review Question 9A.5 on the book)*

1.1 Data

Patient	Score	Patient	Score
1	-0.238	6	+0.090
2	-0.085	7	-0.736
3	-0.215	8	-0.365
4	-0.227	9	-0.179
5	-0.037	10	-0.048

1.2 Mannual Computation

We want to test the hypothesis that $H_0 := \Delta = 0$ against the two-sided alternative, where Δ is the median change scores, *i.e. ln(ERG amplitude) at follow-up - ln(EGR amplitude) at baseline.* Denote d_i as the score for the i patient.

First we rank all the Scores with its absolute value. We find that there're no ties in the data and the only patient has a positive score, the patient 6, ranked 4 in all ten patient. So rank sum R_1 of positive difference is 7. Assuming the null hypothesis is true, the p-value here can be computed as follow

R_1	Numbers of outcome
0	1
1	1
2	1
3	2
4	2

And there are total of 2^{10} outcome possible. Since $(1+1+1+2+2)/2^{10} = 0.0068 \leq 0.5$, the p-value when H_0 is true is $2 \times 0.0068 = 0.01367$

Hence we shall reject the null hypothesis at significance level 0.05.

1.3 Biological Conclusion

We can assert that the following surgery has a significant effect on the change of ERG amplitude, with the surgery decline the level of ERG amplitude.

Large sample Wilcoxon Signed-rank test(Normal theory test)

Laureysens et al. (2004) measured metal content in the wood of $n = 36$ poplar clones growing in a polluted area, once in August and once in November. Concentrations of aluminum (in micrograms of Al per gram of wood) are shown.

There are two nominal variables: time of year (August or November) and poplar clone (Columbia River, Fritzi Pauley, etc.), and one measurement variable (micrograms of aluminum per gram of wood). We want to see if there's any significance difference of concentration of aluminum in different time.

2.1 Data

Only 5 samples are shown here. The rest is in the code.

Clone	August	November	August-November
Columbia River	18.3	12.7	-5.6
Fritzi Pauley	13.3	11.1	-2.2
Hazendans	16.5	15.3	-1.2
Primo	12.6	12.7	0.1
Raspalje	9.5	10.5	1.0

2.2 Mannual computation

We want to test the hypothesis $H_0 := \Delta = 0$ against the two-sided alternative, where Δ is the median of the change of concentration of aluminum.

First we rank the data with respect to its absolute value and compute the rank sum of positive term $R_1 = 343$. The positive change of concentration of aluminum has $n_1 = 17$ members, the negative one has $nh_2 = 19$. Normal theory test can be adopted.

$$\frac{n(n+1)}{4} = 333 \neq R_1 \tag{1}$$

And there're no ties in the data. Then

$$T = \left[\left|R_1 - \frac{n(n+1)}{4}\right| - \frac{1}{2}\right] \Big/ \sqrt{n(n+1)(2n+1)/24} = 0.1493 \tag{2}$$

Compute the p-value then

$$p = 2 \times [1 - \Phi(T)] = 0.881 \tag{3}$$

where Φ is the c.d.f of standard normal distribution. We see a very large p-value, so we can not reject the null hypothesis at significance level 0.05. Meaning that the mean of the difference is 0 and symmetry.

2.3 Biological conclusion

There's no significance in the level of the concentration of aluminum in the wood between August and November. The level of concentration of metal remains the same in two measurements in August and November.

Programming results for Wilcoxon signed rank test

Our program can automatically determine to use whether exact test or not.

First we run the test in the small sample case, the ERG problem. We see that the following results.

```
> d1<-c(-0.238,-0.085,-0.215,-0.227,-0.037,0.09,-0.736,-0.365,-0.179,-0.048)
> Wilcox_signed_rank_test(d1)
$`value of possitive rank sum`
[1] 4

$`p-value`
[1] 0.01367188

$`reject at significance level 0.05?`
[1] TRUE
```

It gives identical result to our mannual computation.

Then we apply the test to the large sample region, the concentration of metal problem, we see the following results.

```
> d2 <- c(-5.6,-2.2,-1.2,0.1,1.0,2.0,3.1,5.3,6.3,7.2,12.0,12.3,23.4,5.2,-3.3,-9.4,8.9,6.7,4.8,1.1,-2.3,-9.5,-6.6,-6.8,-3.1,-
4.9,6.5,4.2,-10.1,-2.2,-2.9,-6.3,-5.1,-0.3,-0.4,-3.5)
> Wilcox_signed_rank_test(d2)
$`value of positive rank sum`
[1] 343

$`value of T statistics`
[1] 0.1492505

$`p-value`
[1] 0.881356

$`reject at significance level 0.05?`
[1] FALSE
```

It also gives identical resul to our mannual computation.

It's worth noticing that we enumerate all the possible outcomes in the program explicitly. So if you try to run the program, please make sure your memory is **at least 16G**, otherwise it will be very likely that you will run out of memory before the program ends.

Small sample Wilcoxon rank-sum test

We use one problem but excluding 1 sample from each group for distinguishing large and small sample region.

A pilot study is planned to test the efficacy of vitamin E supplementation as a possible preventive agent for Alzheimer's disease. Twenty subjects age 65+ are randomized to either a supplement of vitamin E of 400 IU/day (group 1,$n_1 = 1$), or placebo (group 2, $n_2 = 10$). It is important to compare the total vitamin E intake (from food and supplements) of the two groups at baseline. The baseline intake of each group in IU/day is given.

Compare the baseline vitamin E intake between the two groups. *(Review Questions 9B, Page 359)*

3.1 Data

Group 1($n_1 = 9$)	7.5	12.6	3.8	20.2	6.8	403.3	2.9	7.2	10.5
Group 2($n_2 = 9$)	8.2	13.3	102.0	12.7	6.3	4.8	19.5	8.3	407.1

3.2 Mannual Computation

We want to test the hypothesis $H_0 : F_1(x) = F_2(x)$ against the alternative $H_1 : F_1(x) = F_2(x+\Delta)$, where $\Delta \neq 0$, F_1 is the cdf of group 1 and F_2 is the one for group 2.

There's no enough sample for normal theory test, we adopt exact test here.

First we rank the combined sample and compute the rank sum of the first group $R_1 = 74$. And under the null hypothesis, the distribution is uniform on the sample space. There are total of

$$\binom{n_1+n_2}{n_1} = 48620 \tag{4}$$

of possible out come, in which there're 8268 outcomes has the rank sum smaller than our obervation, which is less than 1/2 in propotion. So the p-value i computed as $p = 8268/48620 = 0.340107$.

This is a large p-value, hence we can not reject the null hypothesis at significance level 0.05.

3.3 Biological conclusion

There's no significance in difference of baseline vitamin E intake between two groups. The underlying distribution of two groups are not significantly different.

Large sample Wilcoxon rank-sum test

This problem use the original full data from the textbook.

A pilot study is planned to test the efficacy of vitamin E supplementation as a possible preventive agent for Alzheimer's disease. Twenty subjects age 65+ are randomized to either a supplement of vitamin E of 400 IU/day (group 1,$n_1 = 1$), or placebo (group 2, $n_2 = 10$). It is important to compare the total vitamin E intake (from food and supplements) of the two groups at baseline. The baseline intake of each group in IU/day is given.

Compare the baseline vitamin E intake between the two groups. *(Review Questions 9B, Page 359)*

4.1 Data

Group 1($n_1 = 10$)	7.5	12.6	3.8	20.2	6.8	403.3	2.9	7.2	10.5	205.4
Group 2($n_2 = 10$)	8.2	13.3	102.0	12.7	6.3	4.8	19.5	8.3	407.1	10.2

4.2 Mannual Computation

We want to test the hypothesis $H_0 : F_1(x) = F_2(x)$ against the alternative $H_1 : F_1(x) = F_2(x+\Delta)$, where $\Delta \neq 0$, F_1 is the cdf of group 1 and F_2 is the one for group 2.

First we rank the combined sample and compute the rank sum of the first group $R_1 = 97$.

And here we observe that $n_1, n_2 \geq 10$ so we adopt normal theory test.

Since

$$\frac{n_1(n_1+n_2+1)}{2} = 105 \neq R_1. \tag{5}$$

There's no ties in the data, so the T statistics can be computed as follow

$$T = \left[\left|R_1 - \frac{n_1(n_1+n_2+1)}{2}\right| - \frac{1}{2}\right] \Big/ \sqrt{\left(\frac{n_1 n_2}{12}\right)(n_1+n_2+1)} = 0.5670 \tag{6}$$

The p-vaue can given as

$$p = 2 \times [1 - \Phi(T)] = 0.5707 \tag{7}$$

We can not reject the null hypothesis at significance level 0.05.

4.3 Biological conclusion

There's no significance in difference of baseline vitamin E intake between two groups. The underlying distribution of two groups are not significantly different.

Programming result for Wilcoxon rank sum test

Our program can automatically determine to use whether exact test or not.

First we run the small sample test, which is the large sample test but with some samples excluded. We see the following results

```
> Fisher_exact_test(TranplantData,0.05,"<")
$`left-handed p-value`
[1] 0.7047553

$`Reject?`
[1] FALSE
```

It gives identical result to our mannual computation

Next we run the large sample test, which is the original problem in the text book. We see the following results

```
> # test for large sample Wilcoxon rank sum tests
> g1 <- c(7.5,12.6,3.8,20.2,6.8,403.3,2.9,7.2,10.5,205.4)
> g2 <- c(8.2,13.3,102.0,12.7,6.3,4.8,19.5,8.3,407.1,10.2)
> Wilcox_rank_sum_test(g1,g2)
$`value of rank sum`
[1] 97

$`value of T statistics`
[1] 0.5669467

$`p-value`
[1] 0.5707504

$`reject null hypothesis at significance level 0.05`
[1] FALSE
```

It gives identical result to our mannual computation.