statstutor

community project

encouraging academics to share statistics support resources

stcp-marquier-WilcoxonR

The following resources are associated:
Paired t-test, Excel dataset 'Video.csv', R script file Wicoxon.R

Wilcoxon Signed-rank test in R
(Non-parametric equivalent to paired t-test)

Used when two measurements of the same dependent variable are taken at different time points or under different conditions for each subject and the assumptions of the paired t-test have not been met.

Dependent: Continuous (scale) or ordinal

Independent: Time/condition

Data: The dataset 'Video' contains results from a study by a charity comparing the effectiveness of videos made to aid understanding of a particular medical condition. Participants watched three videos (A, B, C) and one product demonstration (D) and were asked several Likert style questions about each. These were summed to give an overall score for each e.g. TotalCOld is the total score of the ordinal questions for the old video C. Ordinal questions for understanding the condition (Strongly disagree – strongly agree) for video C and the product demonstration are in the 2nd and 3rd columns of the data in the image below and total preference scores are in the 4th and 5th columns of this example.

Person	VideoCOldUnderstandingCONDITION	DEMOUnderstandingCONDITION	TotalCOld	TotalDDEMO
1	3	5	13	22
2	5	5	20	22
3	2	5	14	23
4	5	5	23	25

The Wilcoxon test can be used to compare individual ordinal questions such as those asking for opinions on understanding the condition, or overall scores such as TotaolCOld if the paired differences are not normally distributed.

Open the video dataset from the place you saved it, call it videoR and attach the data so that variable names can be used. This example refers to the memory stick where the data is stored as stcp-Rdataset-video.

```
videoR<-read.csv("E:\\stcp-Rdataset-video.csv",header=T,sep=",")

attach(videoR)
```

Research question: Do members of the public prefer the demonstration technique to the old video C?

The total scores for video C and the demonstration are numeric so the paired t-test is used if the paired differences are normally distributed. Paired differences can be calculated for each individual using:

```
DiffCD=TotalDDEMO-TotalCOld
```

Produce a histogram of these differences using the `hist` command:

```
hist(DiffCD, main="Distribution of D - C ",xlab="Value")
```


Distribution of the differences between the two related groups
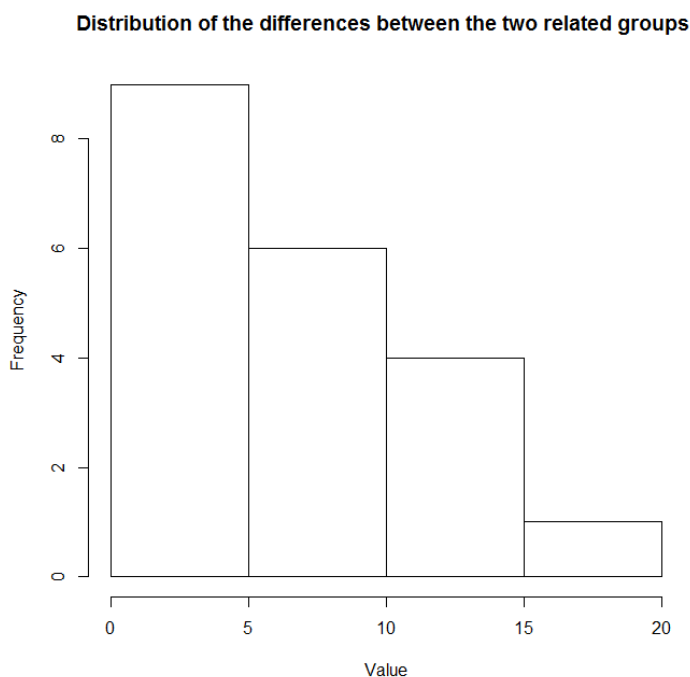
The paired differences are not normally distributed in this dataset so the paired t-test is not reliable and the Wilcoxon signed rank test is more appropriate.

When using a non-parametric test, it is better to summarise using medians rather than means. Examination of the medians using the `median ()` command suggests that people generally preferred the new method C as the median is higher.

```
median(TotalDDEMO)
[1] 22
median(TotalCOld)
[1] 14
```

Conducting the Wilcoxon signed rank test in R

In order to run the Wilcoxon signed rank test, use the command
```
wilcoxon.test (variable1, variable2,Paired=TRUE, exact=FALSE)
```

So for this example use
```
wilcoxon.test(TotalCOld , TotalDDEMO, paired = TRUE, exact=FALSE)
```

The test will give the output below:

```
        Wilcoxon signed rank test with continuity correction
data:  TotalDDEMO and TotalCOld
V = 210, p-value = 9.33e-05
alternative hypothesis: true location shift is not equal to 0
```

The Wilcoxon W **Test Statistic** is simply the lowest sum of ranks but in order to calculate the p-value (Asymp. Sig), R uses an approximation to the standard normal distribution to give the resulting p-value (p = 9.33e-05, which can be written as p < 0.001). This approximation is not as reliable for small samples when an exact test is more reliable. Change the `exact=FALSE` to `exact=TRUE` if you have a small sample.

The output shows that there is significant evidence (p < 0.001) of a difference between the scores for the old video and the new demonstration method.

Reporting a Wilcoxon test

A Wilcoxon signed rank test showed that there was a significant difference (p< 0.001) between scores given for the old video compared to the new demonstration method. The median score for the new demonstration method was 22 compared to 14 for the old video. Therefore, the charity should start using the new method for explaining the condition to the general public.

Effect size

Some people may wish to calculate an effect size which can be done dividing the absolute (positive) Standardised test statistic z by the square root of the number of pairs (20 in our example). R does not give the standardised test statistic in its output but it can be calculated using the p-value from the output and the following formula.

First, give the Wilcoxon out a name e.g. test

```
test<-wilcoxon.test(TotalCOld , TotalDDEMO, paired = TRUE, exact=FALSE)
```

Calculate the standardised z statistic Z and call it Zstat
```
Zstat<-qnorm(test$p.value/2)
```

Calculate the effect size using `abs(Zstat)/sqrt(20)`

```
abs(Zstat)/sqrt(20)
[1] 0.8737202
```
Here the effect size is approximately 0.873 which is very large according to Cohen's classification of effect sizes which is 0.1 (small effect), 0.3 (moderate effect) and 0.5 and above (large effect).