Nonparametric Inference

In this course we will mainly focus on parametric inference, which depends on distributional assumptions. However, if a parametric distribution does not appear to be appropriate you might want to consider a nonparametric test.

Nonparametric, or distribution free tests require few, if any, assumptions about the shapes of the underlying population distributions. For this reason, they are often used in place of parametric tests if/when one feels that the assumptions of the parametric test have been too extremely violated (e.g. if the sample sizes are small and the data do not follow a defined probability distribution).

The parameter of interest for such tests is generally the **median**, since it is a robust measure of location that is not heavily affected by outliers.

Wilcoxon's Signed Ranks Test (One population or paired data)

When we have only one sample of data from a population, the main question of interest is often about estimating the *average* value of a specific variable within the population.

The Wilcoxon Signed Ranks test investigates the null hypothesis that the median of the distribution of the data has a specified value. The idea behind the test is based on differences between each of the data points in the sample and a specified value for the median. If it is true that the population median value is equal to the specified value then we would expect half of the differences to be positive and half of the differences to be negative. If this is not the case we have evidence to reject the null hypothesis.

Similarly, the test is also commonly used for paired data to test the null hypothesis that the median difference between paired data points is zero. The difference between each pair of points is computed and under the null hypothesis that the median difference is zero then we would expect half of the differences to be positive and half to be negative. If this is not the case we have evidence to reject the null hypothesis.

However, we do not wish to deal with the actual numerical values of the differences since these can be affected by outliers in the dataset. The test therefore gives ranks to the absolute value of the differences and computes the sum of the positive and negative ranks.

The test has a small number of **assumptions**:

- The observations come from a symmetric distribution.
- The observations are random and independent.

Example 1 - House price

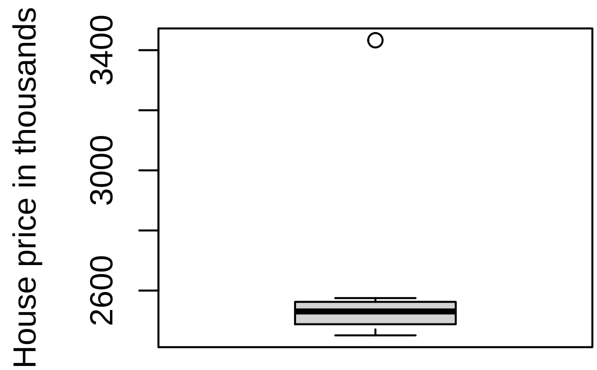
The following values are the house price (in £1,000's) of 8 houses in an extremely affluent area of England:

```
2491    2485    3433    2575    2521    2451    2550    2540
```

Investigate the hypothesis that the median house price of the area is £2,500,000.

Initial Impression

```r
house <- c(2491, 2485, 3433, 2575, 2521, 2451, 2550, 2540)

boxplot(house, ylab="House price in thousands")
```



The plot highlights that there is one extreme data point at £3,433,000.

The observed values (with the exception of the outlier) are consistent with an average house price of £2,500,000. There is some concern about the validity of the Wilcoxon Signed Ranks Test, since the outlier gives the sample data such a strong impression of asymmetry. Each house price can be considered to be independent.

Hypotheses

Let η denote the median house price of the houses in the area. Then,

$H_0 : \eta_A = 2500$

$H_1 : \eta_A \neq 2500$

Test Statistic (TS)

Let $d_i = x_i - 2500$, where x_i is the house price of the ith house.

Observed Value of TS

Ordered Data:	2451	2485	2491	2521	2540	2550	2575	3433
d_i	-49	-15	-9	21	40	50	75	933

d_i	9	15	21	40	49	50	75	933
sign	-	-	+	+	-	+	+	+
rank	1	2	3	4	5	6	7	8

Compute the sum of the positive and negative ranks, denoted $W+$ and $W-$ respectively. The test statistic is whichever of these has the smallest value. In this case, the test statistic is $W-$, the sum of the ranks of the $|d_i|$ corresponding to negative d_i values in the order statistic. (This is chosen because it is smaller than W+).

$W- = 1 + 2 + 5 = 8$

Rejection Region (RR)

An exact p-value can be computed by hand. However, it is cumbersome to do this and so an alternative method is to use critical values from statistical tables to determine whether or not to reject H_0.

The statistical tables provide a critical value for the test to assess the evidence against H_0. Only the sample size and the significance level are required to obtain the critical value.

Since this is a two-sided test we will choose our significance level to be $\alpha = 0.05$. Using Statistical Tables for the Wilcoxon Signed Ranks test at $n = 8$ we get,

$\text{RR} = \{W- : W- \leq 3\}$.

Conclusion

The observed value of W- = 8 does not lie in RR, so we do not have significant evidence to reject the Null Hypothesis.

In other words, there is insufficient evidence that the median house price of houses in the area is different from £2,500,000

We may wish to repeat the test after removing the extreme observation to see if the results of the analyses change.

Analysis in `R`

The following `R` command can be used to perform this test, x here are the original data, the house prices..

```
house <- c(2491, 2485, 3433, 2575, 2521, 2451, 2550, 2540)
wilcox.test(house, mu=2500)
```

```
##
##  Wilcoxon signed rank exact test
##
## data:  house
## V = 28, p-value = 0.1953
## alternative hypothesis: true location is not equal to 2500
```

Statistical packages will return a p-value for the test as illustrated above. Note, the test statistic here is the larger of W+ and W-, `R` is using a slightly different formulation of the test.

Conclusion

Since the p-value > 0.05 at 0.1953, we do not reject H_0 and conclude that there is insufficient evidence of a difference from £2,500,000.

Mann-Whitney U Test

A similar approach can be used to derive a test that compares the medians of two independent populations.

Although this test is often attributed to Mann and Whitney, and called the Mann-Whitney U Test, Wilcoxon himself independently realised how to extend his earlier work on one-sample problems to the two-sample context and obtained the test in an equivalent form.

The idea of the test is that if there is no difference between the two populations, in terms of the values of a specific variable, then we should be able to collect all of the data together, ignoring which population the data come from, and rank the magnitude of the values.

By using ranks we avoid working with the numerical values where outliers may influence the result.

If we then aggregate (add up) the ranks for each population then we should get roughly the same answer for each population. If the sum of the ranks is very different for each population then we have evidence that the values of the two populations are not the same.

Assumptions

- All the recorded values are independent observations;
- The variable of interest has the same distribution in the two populations, except possibly for a difference in the medians. (In other words, the distribution of the variable of interest has the same shape and spread in the two populations.)

Here we will just illustrate the test using the `R` command `wilcox.test()`.

Example 2 - Preferred Room Temperatures

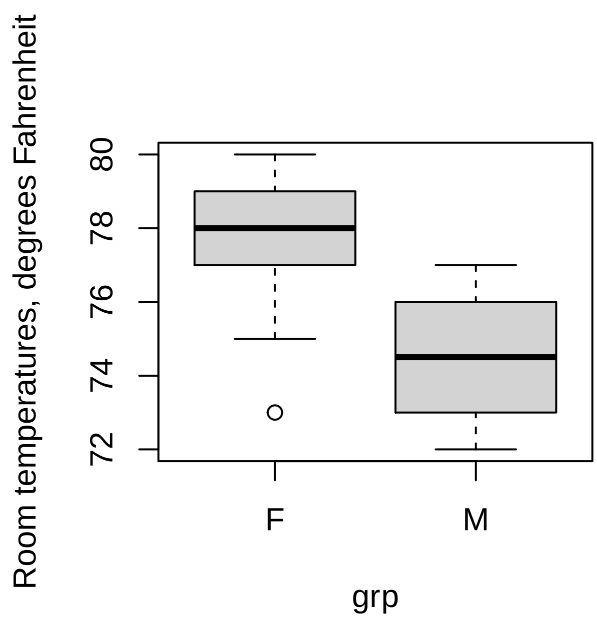
In a controlled environment laboratory, 10 men and 10 women were tested individually to determine the room temperature (^{0}F) they found to be most comfortable. The following results were obtained:

```
Men            74  72  77  76  76  73  75  73  74  75

Women          75  77  78  79  77  73  78  79  78  80
```

Assuming that these values represent a random sample from the respective populations, is the average comfortable temperature the same for men and women?

Initial Impression

```r
Temps <- c(74, 72, 77, 76, 76, 73, 75, 73, 74, 75, 75, 77, 78, 79, 77, 73, 78, 79, 78, 80)
grp <- c(rep("M",10), rep("F",10))

boxplot(Temps~grp, ylab="Room temperatures, degrees Fahrenheit")
```



From the figure above it can be seen that the preferred room temperatures for females are quite a bit higher than those for males.

Assumptions

If X is the preferred room temperature of a randomly-selected individual, then we assume:

1. all the recorded values of X are independent observations;

2. X has the same distribution in the populations of men and women, except possibly for a difference in the medians. (In other words, the distribution of X has the same shape and spread in the two populations.)

The assumptions appear plausible in this example.

Hypotheses

If η_A, η_B are the population median preferred temperatures for men and women, then

$H_0 : \eta_A = \eta_B$

$H_1 : \eta_A \neq \eta_B$

Analysis in `R`

The following `R` command can be used to perform this test:

```
Temps <- c(74, 72, 77, 76, 76, 73, 75, 73, 74, 75, 75, 77, 78, 79, 77, 73, 78, 79, 78, 80)
grp <-c(0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1)
 wilcox.test(Temps~grp)
```

```
## Warning in wilcox.test.default(x = c(74, 72, 77, 76, 76, 73, 75, 73, 74, :
## cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  Temps by grp
## W = 13, p-value = 0.005452
## alternative hypothesis: true location shift is not equal to 0
```

Conclusion

The results in `R` provided a p-value of 0.005 and since this is < 0.05 we reject H_0. Therefore, there is evidence that the preferred median temperatures are different for each population.