

LABORATION 2:

Two sample t tests and confidence intervals, non parametric tests.

1 Introduction

In this lab, I will go on with two sample methods and non parametric methods, using R.

This lab can be seen as a help for solving part of hand-in assignment 2.

2 Two sample methods

2.1 Two samples, common variance

If you have two samples with means μ_1 and μ_2 and unknown but equal variances, taken from the normal distribution, you may get a 95% confidence interval and a p value of the test of $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$ as follows. The data are from "övning 7.6.7" in Alm-Britton.

```
> x=c(4.25,1.48,1.14,6.28,3.32)
> y=c(4.48,7.56,4.42,1.06,6.33,6.72)
> t.test(x,y,var.equal=TRUE)
```

Two Sample t-test

```
data:  x and y
t = -1.3284, df = 9, p-value = 0.2167
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.867998  1.265998
sample estimates:
mean of x mean of y
  3.294    5.095
```

We find that we may not reject H_0 at any reasonable level, and the 95% confidence interval for $\mu_1 - \mu_2$ is about $(-4.87, 1.27)$.

As in the one sample case, you can change the significance level from the default of 95% to something else, for example 90%, by writing

```
> t.test(x,y,var.equal=TRUE,conf.level=0.90)
```

Check what happens if you do so!

2.2 Two samples, different variances

If you do not want to impose that the variances are equal, you get the default choice in R. For the same example data as above, we get

```
> t.test(x,y)
```

Welch Two Sample t-test

```
data: x and y
t = -1.3424, df = 8.9136, p-value = 0.2127
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.840461  1.238461
sample estimates:
mean of x mean of y
   3.294    5.095
```

Observe that in this case, the result is very similar to the one where we specified equal variances.

Just as in the one sample case, you can calculate one-sided confidence intervals and perform one-sided tests. Try that here and see what you get!

2.3 Paired samples

The default of the command `t.test` is that the samples are not paired. If they are, you can add this specification, see the example below with data from example 7.30 in Alm-Britton. As above, we get a confidence interval for $\mu_1 - \mu_2$ and a test of $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$.

```
> x=c(5.2,5.8,4.3,5.2,4.6,4.7,5.8,5.5)
> y=c(5.6,6.3,4.9,5.8,5.5,5.7,6.1,5.4)
> t.test(x,y,paired=TRUE)
```

Paired t-test

```
data: x and y
t = -4.2994, df = 7, p-value = 0.003569
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.8137429 -0.2362571
sample estimates:
mean of the differences
      -0.525
```

As you can see, the P value is about 0.4%, and we can reject H_0 at the 1% level. Check what happens if you do not impose the assumption of paired samples! It turns out that the results become quite different. Why do you think that is?

3 Chi square test

3.1 Tests of goodness of fit

The version of the χ^2 test where you test that the data follows proportions given beforehand in a table, is as follows. We use data from example 8.5 in Alm-Britton. The observations are in the vector `x`, and the given proportions to be tested are in `prob`.

```
> x=c(315,101,108,32)
> prob=c(9/16,3/16,3/16,1/16)
> chisq.test(x,p=prob)
```

Chi-squared test for given probabilities

```
data: x
X-squared = 0.47002, df = 3, p-value = 0.9254
```

We find that the P value is about 93%. Hence, we can not reject the null of given proportions at any reasonable level.

We can also use the χ^2 test to test that the data comes from a specific distribution that contains some unknown parameter that needs to be estimated. Consider the data in problem 803 of AB, also given in the following table:

Number of deaths	Number of cavallery years
0	109
1	65
2	22
3	3
4	1
≥ 5	0

We want to test the null hypothesis that the data comes from the Poisson distribution. At first, we need to estimate the parameter, say m , of this distribution. The MLE is the sample mean, which we obtain through

```
> n=109+65+22+3+1
> m=(1*65+2*22+3*3+4*1)/n
> m
[1] 0.61
```

Observe that, by the rule of thumb for χ^2 approximation, we need the expected number of observations in each cell to be at least 5. The expected number of observations with $X \geq k$ is $nP(X \geq k)$. We have $n = 200$ and $X \sim \text{Po}(m)$ with $m = 0.61$ from above. Hence, by using `ppois` to calculate the Poisson probabilities that we need (cf laboration 1), we calculate the expected numbers of observations strictly greater than 2 or 1, respectively, through

```
> n*ppois(2,m,lower.tail=FALSE)
[1] 4.822934
> n*ppois(1,m,lower.tail=FALSE)
[1] 25.04102
```

So we need to merge the cells for $X = 2, 3, 4$ and $X \geq 5$ of the table of problem 803, into one cell for $X \geq 2$. This means that we should have $22 + 3 + 1 = 26$ observations in this cell.

Now, we can use `dpois` to compute the cell probabilities. We get

```
> x=c(109,65,26)
> prob=c(dpois(0,m),dpois(1,m),ppois(1,m,lower.tail=FALSE))
> chisq.test(x,p=prob)
```

Chi-squared test for given probabilities

```
data:  x
X-squared = 0.062784, df = 2, p-value = 0.9691
```

Observe that the number of degrees of freedom, 2, is not correct here, since R doesn't know that the `prob` vector was obtained after estimating one parameter. The correct number should be $3 - 1 - 1 = 1$. However, the value $Q \approx 0.062784$ is correct. This means that we may get the correct P value from the $\chi^2(1)$ distribution by writing

```
> pchisq(0.062784,1,lower.tail=FALSE)
[1] 0.8021486
```

So the P value is about 80%, and we can not reject that the data comes from the Poisson distribution at any reasonable level.

3.2 Testing independence/homogeneity

Test of independence/homogeneity is done as below, where the data comes from example 8.9 in Alm-Britton. Note the command `rbind` that makes a table `A` out of the rows `a1` and `a2`.

```
> a1=c(52,948)
> a2=c(70,930)
> A=rbind(a1,a2)
> chisq.test(A)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data:  A
X-squared = 2.5227, df = 1, p-value = 0.1122
```

We get a P value of about 11%, so we may not reject the null hypothesis of independence/homogeneity at the 5% level. Note that the observed statistic (2.5227) does not agree with the book (2.828). This is so, because by default, some kind of continuity correction is performed by R. If you want to perform the test without this correction, you write

```
> chisq.test(A,correct=FALSE)
```

Pearson's Chi-squared test

```
data: A
X-squared = 2.8283, df = 1, p-value = 0.09262
```

We then get the same observed statistic as in the book. Now, the P value is 9%, so even in this way, we do not reject the null hypothesis at the 5% level.

4 The sign test

In 'plain' R, there is no procedure for the sign test, but it is fairly straightforward to calculate its P value anyway, by using the Binomial distribution through `pbinom`. For example, take the data from example 8.12 in Alm-Britton. The purpose is to test the null hypothesis that 'kalkning' has no effect vs the alternative of a positive effect (increased 'pH' value). The sign test rejects if the number of positive increases is large. A code for doing this is as follows

```
> x=c(5.2,5.8,4.3,5.2,4.6,4.7,5.8,5.5)
> y=c(5.6,6.3,4.9,5.8,5.5,5.7,6.1,5.4)
> k=sum(y-x>0)
> n=length(x)
> pbinom(k-1,n,1/2,lower.tail=FALSE)
[1] 0.03515625
```

We get a P value of about $0.035 < 0.05$, so we can reject the null hypothesis at the 5% level. But why does the code work? Well, at first we calculate k as the sum of the vector given by $y-x>0$, which contains ones at entries where $y-x>0$ is true, and zeros at the other entries. (What is k here? Check!). Then, with n as the number of observations (what is n here?), we calculate the P value as $P(X \geq k) = P(X > k - 1)$ where $X \sim \text{Bin}(n, 1/2)$. Check that you get the same result by using table 2 in Alm-Britton!

5 The signed rank test

The signed rank test, where we have a paired sample of (X, Y) and the null hypothesis is that X and Y have the same distribution, may be performed as in the example below. The data are as above, from example 8.12 in Alm-Britton, see övning 8.4.1. Observe that the test is one-sided! Also observe that here, when writing y, x , the alternative hypothesis should be 'greater'.

```
> wilcox.test(y,x,paired=TRUE,alternative='greater')
```

Wilcoxon signed rank test

data: y and x

$V = 35$, p-value = 0.007813

alternative hypothesis: true location shift is greater than 0

Hence, the P value is 0.8%, which is a lot smaller than the corresponding P value of the sign test. This is quite natural, since the signed rank test uses more information from the data.

Check what happens with the P value if you perform a two sided test according to the default command

```
> wilcox.test(y,x,paired=TRUE)
```

Also, check what happens if you change the ordering y, x to x, y for both tests.

6 The Wilcoxon two sample test

In case you have two independent samples, and if you want to test the null hypothesis that they come from the same distribution, then you can use the Wilcoxon two sample test.

We take data from example 8.17 in Alm-Britton and get:

```
> A=c(9.5,12.5,13.5,14,14.5,17,18,18.5,21,23,24)
```

```
> B=c(19,20,20,20.5,22,24.5,26.5,27,28,29.5,34,34,35.5)
```

```
> wilcox.test(A,B)
```

Wilcoxon rank sum test with continuity correction

data: A and B

$W = 14$, p-value = 0.0009538

alternative hypothesis: true location shift is not equal to 0

Warning message:

```
In wilcox.test.default(A, B) : cannot compute exact p-value with ties
```

Observe the warning text. It seems that R can not be fully trusted when you have moderately large samples containing ties.

The P value is 0.095%, so for example, we may reject the null hypothesis at the 1% level. This is when (by default) the alternative is two-sided.

To get a one-sided test, where the alternative is that the A sample comes from a distribution that tends to produce smaller values than what the distribution of the B sample does, you write

```
> wilcox.test(A,B,alternative='less')
```

Check the P value! Also, check what happens when you have the opposite alternative, using

```
> wilcox.test(x,y,paired=TRUE,alternative='greater')
```

Now, you obtain a P value which is close to one. Why?

7 Testing for trend

In this course, we have presented two non parametric tests for trend: the runs test and a test based on the Spearman correlation.

7.1 The runs test

The runs test is not available in plain R, but it can be downloaded from some package, for example the package **randtests**. To download a package, you can e.g. go to **Tools** on the upper bar of commands, then choose **Install Packages** and write the package name. You will then get a fairly lengthy output in red text, hopefully including the word **DONE** close to the end.

Then, to complete the installation, for example of the package **randtests**, write

```
> library(randtests)
```

Now, let's take the data from example 8.20 in Alm-Britton. We write the data for 'Kalkmängd' in the vector **x** e.g. as (observe the 'recursive' way of writing to avoid to long lines)

```
> x=c(113.3,149.1,177.9,145.6,158.1,176.5)
> x=c(x,202.4,203.1,222.5,219.8,191.8,190.6)
> x=c(x,213.2,204.0,190.7,204.1,192.0,180.8)
```

To get the version of the runs test that we want, we need to specify that we reject only for few runs (**alternative='left.sided'**). To get the same result as in the book, it also turns out that we need to calculate the P value in an exact manner by adding **pvalue='exact'**. Otherwise, the default choice of normal approximation is used, but for our data, the rule of thumb for that is not fulfilled (why?).

With these aspects taken into account, we get

```
> runs.test(x,alternative='left.sided',pvalue='exact')
```

Runs Test

```
data: x
statistic = -1.4577, runs = 7, n1 = 9, n2 = 9, n = 18, p-value =
0.109
alternative hypothesis: trend
```

i.e. the same as the book (7 runs and P value 0.109).

7.2 Using the Spearman rank correlation

To test for trend, we can also use Spearman's rank correlation. With the year as y variable, we get the resulting statistic from

```
> y=seq(1986,2003)
> n=length(x)
> r=cor(y,x,method='spearman')
> sqrt(n-1)*r
[1] 2.353021
```

Now, we can compare to the corresponding two-sided normal percentile (1.96 for a 5% level test), or we can calculate the P value by writing

```
> 2*pnorm(sqrt(n-1)*r,lower.tail=FALSE)
[1] 0.01862158
```

Does this result fit with the test using the normal percentile?