Lecture 4 - statistical tests

In R it is almost too easy to employ a statistical test!

One has to be careful to understand what the test tests.

What is a test?

We have a hypothesis, and want to test if the data is consistent with this hypothesis. We also have an alternative hypothesis.

Usually we will

1. Decide on a number to calculate from our data.
2. Calculate how likely is one to get this number "or worse" given our hypothesis.

   (Is "or worse" one- or two-sided?)
3. This gives us a p-value, which we compare to pre-defined cutoffs.

The problem: how do we know how likely things are?

We can either use distributions that the best mathematicians in the world figured out, or we can make our own using bootstrapping.

Loading Libraries (packages)

In R many functions are stored in libraries. To use them, we need to load the library:

```
> library(ctest)
>
```

Frequencies

If we observe a family with 1 male offspring, and 9 female offspring, is the sex ratio 50:50?

Our hypothesis is that the sex ratio is 50%.

Since we would have been as surprised to see 1:9 as 9:1, we want to know how likely is it to get a ratio of 1:9 or 9:1 or worse?

How likely are we to get 0:10, 1:9, 9:1, or 10:0 if the ratio is really 50%?

The possibilities are:

F	F	F	F	F	F	F	F	F	F	M	M	M	M	M	M	M	M	M	M
M	F	F	F	F	F	F	F	F	F	F	M	M	M	M	M	M	M	M	M
F	M	F	F	F	F	F	F	F	F	M	F	M	M	M	M	M	M	M	M
F	F	M	F	F	F	F	F	F	F	M	M	F	M	M	M	M	M	M	M
F	F	F	M	F	F	F	F	F	F	M	M	M	F	M	M	M	M	M	M
F	F	F	F	M	F	F	F	F	F	M	M	M	M	F	M	M	M	M	M
F	F	F	F	F	M	F	F	F	F	M	M	M	M	M	F	M	M	M	M
F	F	F	F	F	F	M	F	F	F	M	M	M	M	M	M	F	M	M	M
F	F	F	F	F	F	F	M	F	F	M	M	M	M	M	M	M	F	M	M
F	F	F	F	F	F	F	F	M	F	M	M	M	M	M	M	M	M	F	M
F	F	F	F	F	F	F	F	F	M	M	M	M	M	M	M	M	M	M	F

Each is equally likely. There are 1024 possibilities, 22 of which are as bad as our observation, so:

```
> 22/1024
[1] 0.02148438
>
```

is our p-value.

R can do this calculation thus:

```
> binom.test(1,10)

        Exact binomial test

data:  1 and 10
number of successes = 1, number of trials = 10, p-value = 0.02148
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.002528579 0.445016117
sample estimates:
probability of success
                   0.1

>
```

And, we get exactly the same p-value!

In this case, we were "surprised" to see so few males. But we had no preference for males or females. Therefore our alternative hypothesis was that the sex ratio is not 0.5.

But, maybe in a previous experiment, we already saw few males. Now we just want to confirm that it isn't an equal sex ratio, but that we have few males.

In this case we ask: what is the chance to see 1:9 or worse, i.e. 1 male, 9 females, or 0 males 10 females?

This is a one-sided test.

From above we can see that the p-value in this case is 11/1024.

```
> binom.test(1,10,alternative="less")

        Exact binomial test

data:  1 and 10
number of successes = 1, number of trials = 10, p-value = 0.01074
alternative hypothesis: true probability of success is less than 0.5
95 percent confidence interval:
 0.0000000 0.3941633
sample estimates:
probability of success
                   0.1

>
```

If ahead of time we want to show that there are too few males, but we get 9 males and 1 female, then we still have to test for less!

```
> binom.test(9,10,alternative="less")
```

```
        Exact binomial test

data:  9 and 10
number of successes = 9, number of trials = 10, p-value = 0.999
alternative hypothesis: true probability of success is less than 0.5
95 percent confidence interval:
 0.0000000 0.9948838
sample estimates:
probability of success
                   0.9

>
```

Frequency tables

Let us assume that our hypothesis is that the frequency of green, blue, and brown eyes in the population is 20%, 30%, 50%.

We see the following data:

```
> observed=c(blue=17,blue=25,brown=60)
> observed

 blue  blue brown
   17    25    60

> sum(observed)

[1] 102

>
```

We would expect:

```
> expected=c(0.2,0.3,0.5)*102
> expected

[1] 20.4 30.6 51.0
```

We could now, for example, sum the absolute difference of observed-expected. Or (observed-expected)^2.

But for all those it will be hard to calculate to get that value.

For the following function, we do have an approximate distribution:

(observed value - expected value)^2 / (expected value)

```
> chisq.test(c(17,25,60),p=c(0.2,0.3,0.5))


        Chi-squared test for given probabilities

data:  c(17, 25, 60)
X-squared = 3.1797, df = 2, p-value = 0.2040
```

Another example:

```
> hair=sample(c("blond","black"),90,rep=T,p=c(0.3,0.7))
> eyes=sample(c("green","brown"),90,rep=T,p=c(0.8,0.2))
```

```
> T=table(hair,eyes)
> T
         eyes
hair     brown green
  black  9     58
  blond  4     19
>
```

We would like to know if hair color and eye color are independent. Our hypothesis is that they are independent, and therefore one can calculate the chance for each category from

```
> chisq.test(T)

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

data:  T
X-squared = 0.0149, df = 1, p-value = 0.9027

Warning message:
Chi-squared approximation may be incorrect in: chisq.test(T)
>
```

The reason for the warning is that the chisq.test is only approximate. For small values in the table, the test is not exact. Another test is exact in this case:

```
> fisher.test(T)

        Fisher's Exact Test for Count Data

data:  T
p-value = 0.7327
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.1796318 3.6673071
sample estimates:
odds ratio
 0.7397047
>
```

Comparing means

Often we have data from two sources, and would like to know if we have the same means, i.e. one is not bigger than the other.

As before, we can think of many functions one could compute.

Assuming that the data is normaly distributed, a test called t-test exists:

The t-test calculates the (difference of the means in the two samples)/(std. error of mean)

To use the t-test we have to know if the variances are the same. To do this, there is a test called var.test:

```
> x=rnorm(20,mean=10)
```

```
> y=rnorm(20,mean=10.5)
> var.test(x,y)

        F test to compare two variances

data:  x and y
F = 0.9084, num df = 19, denom df = 19, p-value = 0.8364
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3595679 2.2951056
sample estimates:
ratio of variances
         0.9084307

>
```

Now we can use the t-test:

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

        Welch Two Sample t-test

data:  x and y
t = -1.1162, df = 37.913, p-value = 0.2714
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.1086930  0.3206622
sample estimates:
mean of x mean of y
 10.03693  10.43094

>
```

Notice that the test did not detect the difference in the variables.

If we know or suspect that the variances are not the same, we can use a parameter of t.test:

```
> t.test(x,y,var.eq=F)

        Welch Two Sample t-test

data:  x and y
t = -1.1162, df = 37.913, p-value = 0.2714
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.1086930  0.3206622
sample estimates:
mean of x mean of y
 10.03693  10.43094

>
```

A non-parametric test of means

Non-parametric tests don't assume anything about the distribution of the variable - they use only the rank of the data.

The wilcoxon test calculates the rank of all samples, and then sums the ranks of the smaller sample, and does various things with them

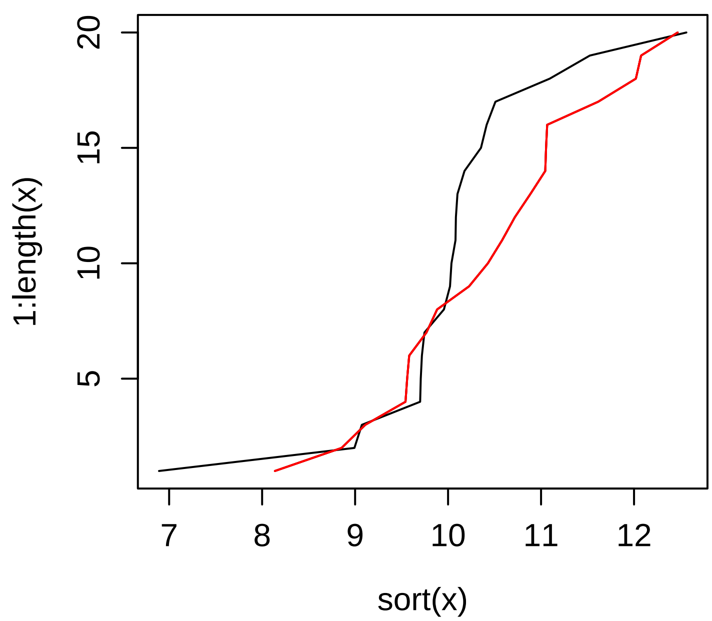
```
> wilcox.test(x,y)

        Wilcoxon rank sum test

data:  x and y
W = 164, p-value = 0.3408
alternative hypothesis: true mu is not equal to 0

>
```

The wilcox test is slightly weaker than the t-test, because it doesn't take into account that the data is normal.

The Kolmogorov-Smirnov test

This is a very nice test that compares two distributions. It will differ even if the mean is the same but the variance is different.

```
> plot(sort(x),1:length(x),type="l")
> lines(sort(y),1:length(y),col=2);v()
```




```
>
```

The kolmogov-smirnov test finds the point at which the proportion of points is most different, and looks at that maximal difference.

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

```
        Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.35, p-value = 0.1745
alternative hypothesis: two.sided

>
```