

# Null-Hypothesis testing

## Ecological Data Management and Analysis

Achaz von Hardenberg

1 Dec 2014

# What is a null-hypothesis test?

As biologists, all the time we pose ourselves questions.

Questions such as:

Is the mean length of the right leg of the Dahu really longer than the left leg?

Or is the the difference we find in our sample of measures due to chance?

This and other kind of questions (comparison of a single sample against an expected distribution, comparison between multiple samples, correlation between variables etc.), are the focus of what in Statistical inference is called **statistical hypothesis testing** or also **confirmatory data analysis** (distinguished from **Exploratory data analysis** in which no particular a-priori hypothesis is tested).

The tool used to reply to this kind of question is the **Null-hypothesis test**

# First step: Null and alternative hypotheses

**Are the right legs of Dahus longer than the left?**

$$H_0 = \mu_R = \mu_L$$

$$H_1 = \mu_R > \mu_L$$

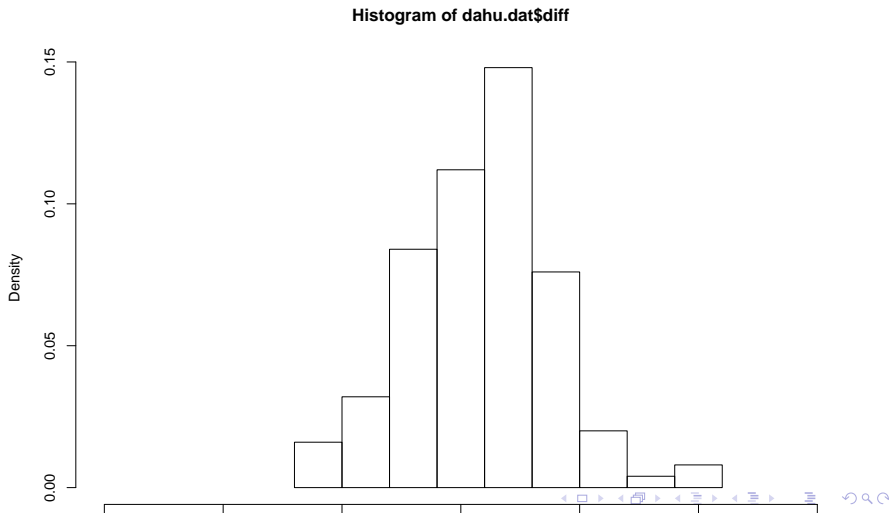
$$H_0 = \mu_R - \mu_L = 0$$

$$H_1 = \mu_R - \mu_L > 0$$

The Null hypothesis is considered **True**, unless it's **rejected** by the alternative hypothesis

## First step: Null and alternative hypotheses

```
dahu.dat<-read.csv("dahu.csv")  
dahu.dat$diff<-dahu.dat$R-dahu.dat$L  
hist(dahu.dat$diff,prob=T,xlim=c(0,30))
```



# Student t-test

$$t = \frac{\overline{X}_{H_1} - \overline{X}_{H_0}}{\frac{sd}{\sqrt{n}}}$$

```
t<-mean(dahu.dat$diff-0)/(sd(dahu.dat$diff)/sqrt(length(dahu.dat$diff)))
```

```
## [1] 58.11
```

# Student t-test

t statistic follows the Student's t distribution

```
t.test(dahu.dat$diff)
```

```
##  
##  One Sample t-test  
##  
## data:  dahu.dat$diff  
## t = 58.11, df = 124, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
##  15.26 16.34  
## sample estimates:  
## mean of x  
##      15.8
```

## Student t-test

```
t.test(dahu.dat$R,dahu.dat$L,paired=T)
```

```
##  
## Paired t-test  
##  
## data: dahu.dat$R and dahu.dat$L  
## t = 58.11, df = 124, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not  
## 95 percent confidence interval:  
## 15.26 16.34  
## sample estimates:  
## mean of the differences  
## 15.8
```

## Student t-test

```
t.test(dahu.dat$diff~dahu.dat$Sex)
```

```
##  
##  Welch Two Sample t-test  
##  
## data:  dahu.dat$diff by dahu.dat$Sex  
## t = -3.414, df = 120.2, p-value = 0.0008741  
## alternative hypothesis: true difference in means is not  
## 95 percent confidence interval:  
##  -2.7682 -0.7359  
## sample estimates:  
## mean in group 0 mean in group 1  
##           14.86           16.62
```

The **p-value** is the probability of obtaining the sample results if the Null hypothesis is true (i.e. that the difference between right and left leg is the same in males and females). If the p-value is lower than a prespecified **significance level** ( $\alpha$  level = 0.05), then we



## 5 common misunderstandings around P-values

1. A null hypothesis can only be **rejected**: a high p-value does **not** mean that the null hypothesis is true!

## 5 common misunderstandings around P-values

1. A null hypothesis can only be **rejected**: a high p-value does **not** mean that the null hypothesis is true!
2. Conversely, a small p-value is **not** the probability that the alternative hypothesis is false.

## 5 common misunderstandings around P-values

1. A null hypothesis can only be **rejected**: a high p-value does **not** mean that the null hypothesis is true!
2. Conversely, a small p-value is **not** the probability that the alternative hypothesis is false.
3. The significance level (0.05) is **not** determined by the p-value, but is arbitrary chosen by the researcher a-priori.

## 5 common misunderstandings around P-values

1. A null hypothesis can only be **rejected**: a high p-value does **not** mean that the null hypothesis is true!
2. Conversely, a small p-value is **not** the probability that the alternative hypothesis is false.
3. The significance level (0.05) is **not** determined by the p-value, but is arbitrary chosen by the researcher a-priori.
4. The p-value is **not** the probability of getting the same result when replicating the study.

## 5 common misunderstandings around P-values

1. A null hypothesis can only be **rejected**: a high p-value does **not** mean that the null hypothesis is true!
2. Conversely, a small p-value is **not** the probability that the alternative hypothesis is false.
3. The significance level (0.05) is **not** determined by the p-value, but is arbitrary chosen by the researcher a-priori.
4. The p-value is **not** the probability of getting the same result when replicating the study.
5. P-value is not an indication of the **effect size**

# Effect size

The **effect size** is the quantitative measurement of the strength of a phenomenon.

the mean difference, the correlation between two variables, the regression coefficient (the  $\beta$  of the regression line) are all measures of effect size.

The larger the absolute value of the effect size, the stronger the effect.

In a statistical test, the p-value may be lower than significance level, and thus making the researcher reject the null hypothesis, but still the effect size may be so low that it has actually no biological relevance!

**Don't look only to the p-value! Always consider also the relevance of the effect size (and the confidence intervals around it)!**

# Type I and Type II errors

**Type I error** is the probability to reject the Null-hypothesis when it is actually true:

*Considering an effect significant when it is actually not (false positive)*

Type I error corresponds to the significance level!

**Type II error** is the probability to fail to reject the Null-hypothesis when it is wrong:

*Failing to detect a significant effect (false negative)*

# Power of a statistical test

**Power** = 1 - Type II error

It tells us the probability of finding a significant effect when it is true given a particular sample size

```
power.t.test(delta=mean(dahu.dat$diff), sd=sd(dahu.dat$diff)
```

```
##  
##      Two-sample t test power calculation  
##  
##              n = 2.077  
##            delta = 15.8  
##             sd = 3.041  
##      sig.level = 0.05  
##             power = 0.8  
##      alternative = two.sided  
##  
## NOTE: n is number in *each* group
```



## Power of a statistical test

```
chick.dat<-chickwts[11:36,]  
t.test(chick.dat$weight~chick.dat$feed)
```

```
##  
##  Welch Two Sample t-test  
##  
## data:  chick.dat$weight by chick.dat$feed  
## t = -1.325, df = 23.63, p-value = 0.198  
## alternative hypothesis: true difference in means is not  
## 95 percent confidence interval:  
##  -70.84  15.49  
## sample estimates:  
## mean in group linseed mean in group soybean  
##                218.8                246.4
```

# Power of a statistical test

```
power.t.test(n=13,delta=mean(chick.dat$weight[1:12])-mean(  
  sd=sd(chick.dat$weight),sig.level=0.05)
```

```
##  
##      Two-sample t test power calculation  
##  
##              n = 13  
##            delta = 27.68  
##            sd = 54.06  
##      sig.level = 0.05  
##            power = 0.24  
##      alternative = two.sided  
##  
## NOTE: n is number in *each* group
```

## Power of a statistical test

```
power.t.test(delta=mean(chick.dat$weight[1:12])-mean(chick  
sd=sd(chick.dat$weight),sig.level=0.05, power=0.80)
```

```
##  
##      Two-sample t test power calculation  
##  
##              n = 60.85  
##            delta = 27.68  
##             sd = 54.06  
##    sig.level = 0.05  
##           power = 0.8  
##    alternative = two.sided  
##  
## NOTE: n is number in *each* group
```

**The lesson is:** we need to feed many more chicks to be reasonably sure that actually there is no difference between feeding lindseed or soybeen!

# Checking for the Normality assumption

The assumption of the t-test is that the sample of data comes from a normal population. We can use a test, like the Shapiro-wilk test, to test if our sample deviates significantly from a normal distribution

```
shapiro.test(chick.dat$weight)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  chick.dat$weight  
## W = 0.963, p-value = 0.4535
```

Our Null hypothesis is that the data conforms to what expected by a normal distribution, therefore a non-significant result is an indication that our data is approximately normal (even though not necessarily!)

## Checking for the Normality assumption

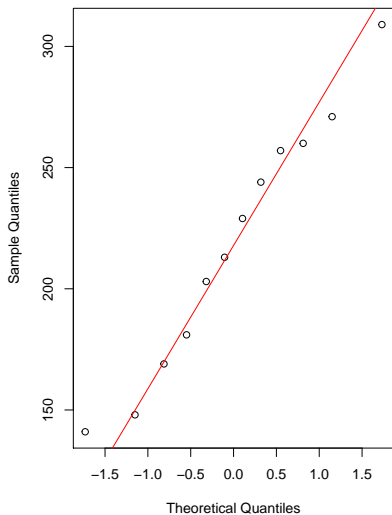
We can also check for normality graphically. The best way to do is using a quantile-quantile normality plot (qqnorm)

```
par(mfrow=c(1,2))  
qqnorm(chick.dat$weight[chick.dat$feed=="linseed"])  
qqline(chick.dat$weight[chick.dat$feed=="linseed"], col="red")  
qqnorm(chick.dat$weight[chick.dat$feed=="soybean"])  
qqline(chick.dat$weight[chick.dat$feed=="soybean"], col="red")
```

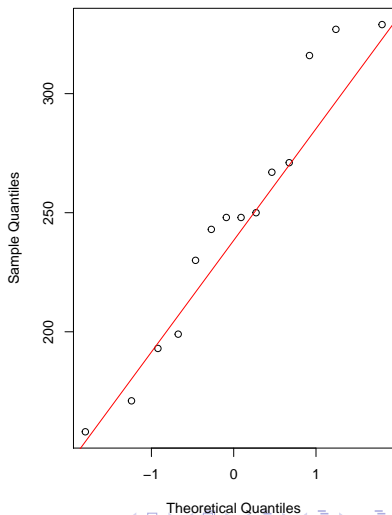
# Checking for the Normality assumption

We can also check for normality graphically. The best way to do is using a quantile-quantile normality plot (qqnorm)

Normal Q-Q Plot



Normal Q-Q Plot

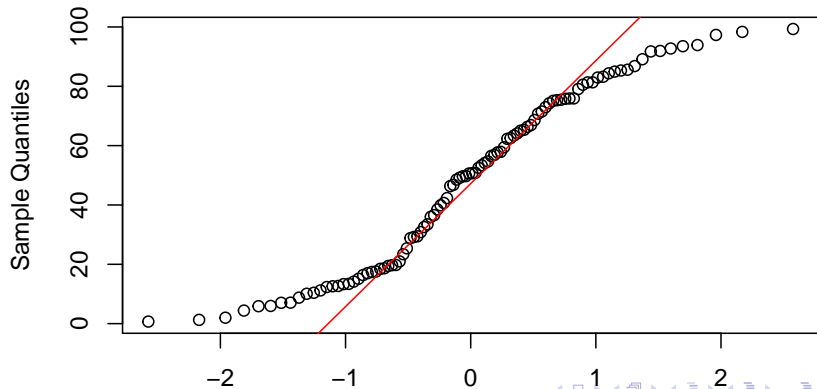


## Checking for the Normality assumption

What does the qqnorm plot look like with not normally distributed data?

```
test.dat<-runif(100,0,100)
qqnorm(test.dat); qqline(test.dat, col="red")
```

**Normal Q-Q Plot**



## Tests for count data: The Chi-square test

Ladybirds (*Adalia bipunctata*), can be red with black points, or black with red points. Does the frequency of the two type differ according of the area in which they live (**Industrial** or **Rural**)?

Example taken from: *Beckerman and Petchey (2012) Getting started with R. An Introduction for Biologists. Oxford University Press*

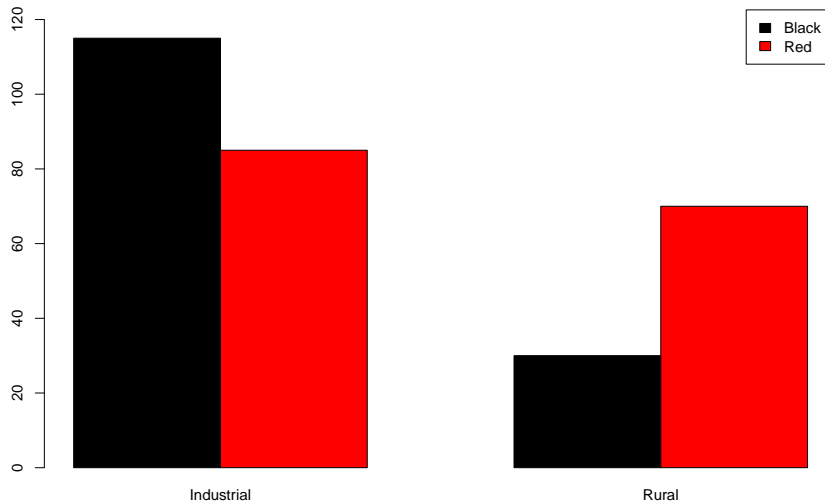
```
ladybird.dat <- matrix(c(115, 30, 85, 70), 2, 2, byrow = TRUE,  
  dimnames = list(c("Black", "Red"), c("Industrial", "Rural")))
```

Color	Industrial	Rural
Black	115	30
Red	85	70



# Tests for count data: The Chi-square test

```
barplot(ladybird.dat, beside = TRUE, col=c("Black", "Red"), y
```



## Tests for count data: The Chi-square test

```
chisq.test(ladybird.dat)
```

```
##
```

```
##  Pearson's Chi-squared test with Yates' continuity corre
```

```
##
```

```
## data:  ladybird.dat
```

```
## X-squared = 19.1, df = 1, p-value = 1.239e-05
```

## Tests for count data: Fisher's exact test

If in any cell you have less than 10 cases, use Fisher's exact test

```
ladybird2.dat <- matrix(c(12, 3, 9, 7), 2, 2, byrow = TRUE,  
  dimnames = list(c("Black", "Red"), c("Industrial", "Rural")))
```

Color	Industrial	Rural
Black	12	3
Red	9	7

# Tests for count data: Fisher's exact test

```
fisher.test(ladybird.dat)
```

```
##  
## Fisher's Exact Test for Count Data  
##  
## data: ladybird.dat  
## p-value = 8.804e-06  
## alternative hypothesis: true odds ratio is not equal to  
## 95 percent confidence interval:  
## 1.841 5.464  
## sample estimates:  
## odds ratio  
## 3.144
```