

PhD Course
Advanced Biostatistics

Lecture 3
Statistical Tests

dr. P. Nazarov
petr.nazarov@lih.lu

26-05-2016

- ◆ **Hypotheses testing for means and proportions (L3.1)**
 - ◆ Hypotheses about means
 - ◆ Hypotheses about proportions
 - ◆ 1-tail vs. 2-tail
- ◆ **Hypotheses testing for means of 2 populations (L3.2)**
 - ◆ Independent and matched samples
 - ◆ Unpaired t-test
 - ◆ Paired t-test
 - ◆ Hypotheses about proportions of 2 populations
- ◆ **Testing hypothesis about variances of 2 populations (L3.3)**
 - ◆ uniform, normal, exponential
- ◆ **Testing hypothesis about correlations (L3.4)**
 - ◆ uniform, normal, exponential
- ◆ **Power of a test (L3.5)**
 - ◆ uniform, normal, exponential

Hypotheses

Here we continue the discussion of statistical inference by showing how hypothesis testing can be used to determine whether a statement about the value of a population parameter should or should not be rejected.

In hypothesis testing we begin by making a tentative assumption about a population parameter, i.e. by formulation of a null hypothesis.

Null Hypothesis

The hypothesis tentatively assumed true in the hypothesis testing procedure, H_0

Alternative hypothesis

The hypothesis concluded to be true if the null hypothesis is rejected, H_a

$$H_0: \mu \leq \text{const}$$

$$H_0: \mu \geq \text{const}$$

$$H_a: \mu > \text{const}$$

$$H_a: \mu < \text{const}$$

$$H_0: \mu = \text{const}$$

$$H_a: \mu \neq \text{const}$$

$$H_0: \mu_1 \leq \mu_2$$

$$H_0: \mu_1 \geq \mu_2$$

$$H_a: \mu_1 > \mu_2$$

$$H_a: \mu_1 < \mu_2$$

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

Errors

Type I error

The error of rejecting H_0 when it is true.

Type II error

The error of accepting H_0 when it is false.

Level of significance

The probability of making a Type I error when the null hypothesis is true as an equality, α

		Population Condition	
		H_0 True	H_a True
Conclusion	Accept H_0	Correct Conclusion	Type II Error
	Reject H_0	Type I Error	Correct Conclusion

poor sensitivity

False Negative, β error

False Positive, α error

poor specificity

One-tailed Test for Mean

One-tailed test

A hypothesis test in which rejection of the null hypothesis occurs for values of the test statistic in one tail of its sampling distribution

$$H_0: \mu \leq \mu_0$$

$$H_0: \mu \geq \mu_0$$

$$H_a: \mu > \mu_0$$

$$H_a: \mu < \mu_0$$

A Trade Commission (TC) periodically conducts statistical studies designed to test the claims that manufacturers make about their products. For example, the label on a large can of Hilltop Coffee states that the can contains 3 pounds of coffee. The TC knows that Hilltop's production process cannot place exactly 3 pounds of coffee in each can, even if the mean filling weight for the population of all cans filled is 3 pounds per can. However, as long as the population mean filling weight is at least 3 pounds per can, the rights of consumers will be protected. Thus, the TC interprets the label information on a large can of coffee as a claim by Hilltop that the population mean filling weight is at least 3 pounds per can. We will show how the TC can check Hilltop's claim by conducting a lower tail hypothesis test.

$$\mu_0 = 3 \text{ lbm}$$

Suppose sample of $n=36$ coffee cans is selected. From the previous studies it's known that $\sigma = 0.18 \text{ lbm}$

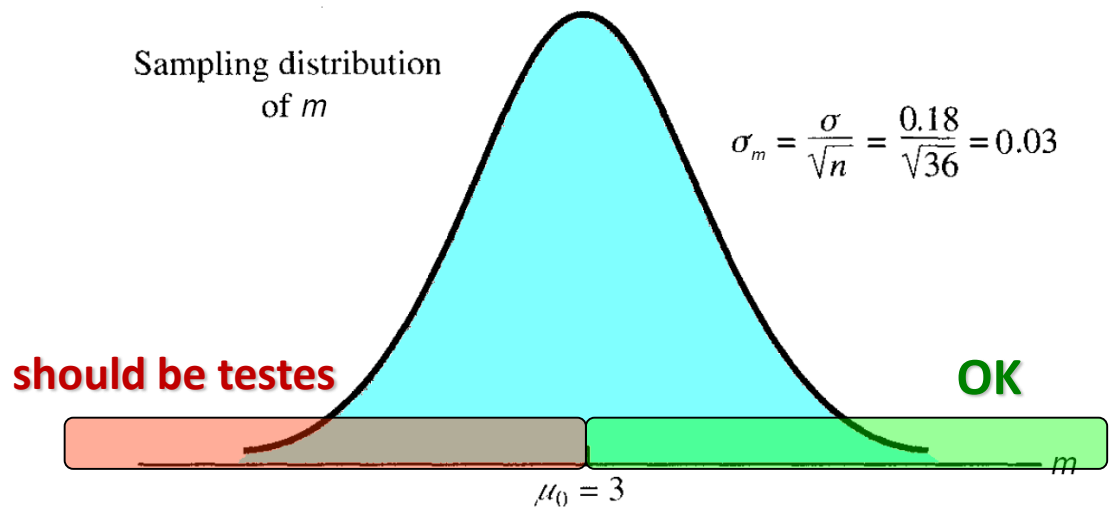
One-tailed Test: Example

$$\mu_0 = 3 \text{ lbm}$$

$$H_0: \mu \geq 3 \quad \text{no action}$$

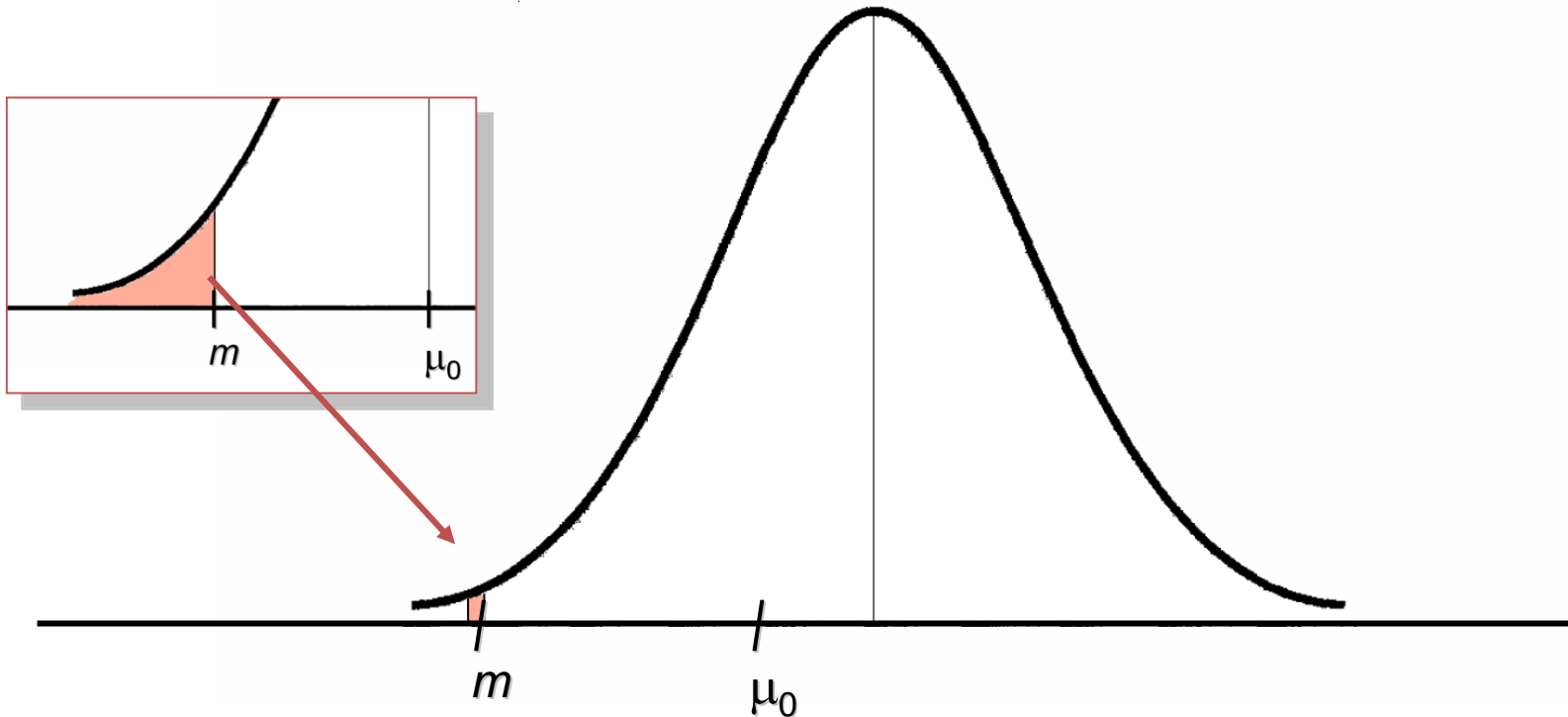
$$H_a: \mu < 3 \quad \text{legal action}$$

Let's say: **in the extreme case**, when $\mu=3$, we would like to be 99% **sure that we make no mistake**, when starting legal actions against Hilltop Coffee. It means that selected significance level is **$\alpha = 0.01$**



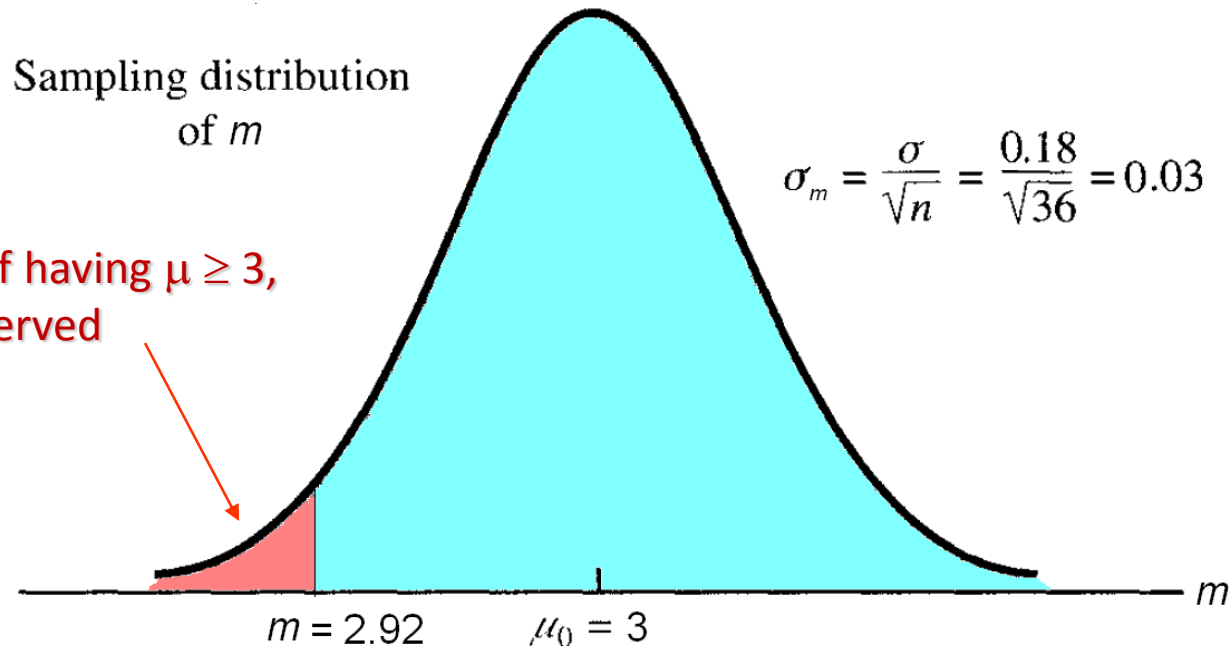
One-tailed Test: Example

Let's find the probability of observation m for all possible $\mu \geq 3$. We start from an extreme case ($\mu=3$) and then probe all possible $\mu > 3$. See the behavior of the small probability area around measured m . What you will get if you summarize its area for all possible $\mu \geq 3$?



$P(m)$ for all possible $\mu \geq \mu_0$ is equal to $P(x < m)$ for an extreme case of $\mu = \mu_0$

p-value



In other words, **red area** characterizes the probability of the null hypothesis.

To be completely correct, the **red area** gives us a **probability of making an error** when rejecting the null hypothesis, or the **p-value**.

Pipeline to Test Hypothesis about Population Mean (manual)

if σ is unknown:

$$\sigma \rightarrow s$$

$$z \rightarrow t$$

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0 : \mu \geq \mu_0$ $H_a : \mu < \mu_0$	$H_0 : \mu \leq \mu_0$ $H_a : \mu > \mu_0$	$H_0 : \mu = \mu_0$ $H_a : \mu \neq \mu_0$
Test Statistic	$t = \frac{m - \mu_0}{s/\sqrt{n}}$	$t = \frac{m - \mu_0}{s/\sqrt{n}}$	$t = \frac{m - \mu_0}{s/\sqrt{n}}$
Rejection Rule: p-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $t \leq -t_\alpha$	Reject H_0 if $t \geq t_\alpha$	Reject H_0 if $t \leq -t_{\alpha/2}$ or if $t \geq t_{\alpha/2}$

Pipeline to Test Hypothesis about Population Mean (R)



In fact in R it is much simpler:

In R use (parametric):

◆ `t.test(x, mu = μ_0 , alternative = ...)`

In R use (non parametric):

◆ `wilcox.test(x, mu = μ_0 , alternative = ...)`

alternative = c("two.sided", "less", "greater")

Pipeline to Test Hypothesis about Population Proportion (manual)

Proportions

π – population proportion

p – sample proportion

π_0 – testing value

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0 : \pi \geq \pi_0$ $H_a : \pi < \pi_0$	$H_0 : \pi \leq \pi_0$ $H_a : \pi > \pi_0$	$H_0 : \pi = \pi_0$ $H_a : \pi \neq \pi_0$
Test Statistic <i>If $np \geq 5, n(1-p) \geq 5$</i>	$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$	$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$	$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$
Rejection Rule: p-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $z \leq -z_{\alpha}$	Reject H_0 if $z \geq z_{\alpha}$	Reject H_0 if $z \leq -z_{\alpha/2}$ or if $z \geq z_{\alpha/2}$

Is used with big n



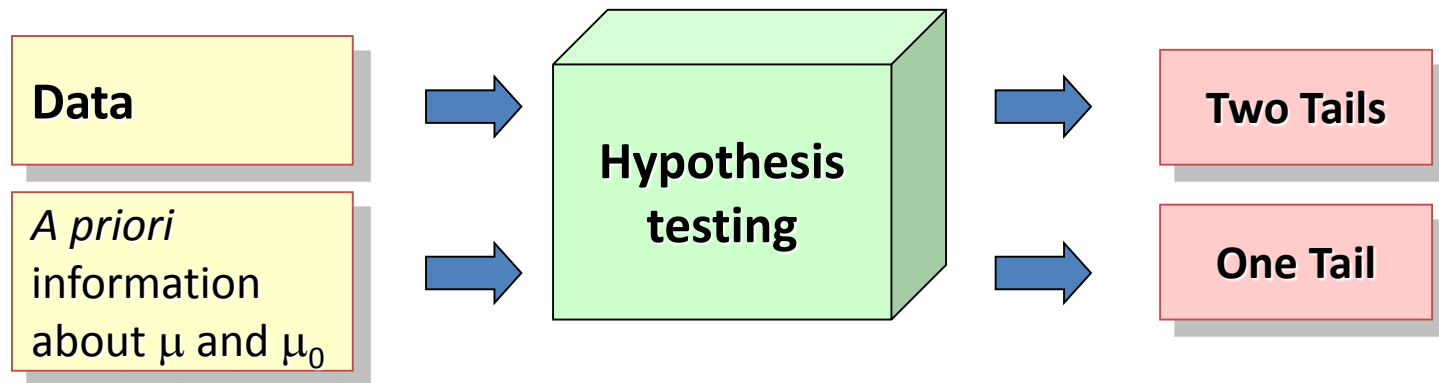
In R use:

◆ `prop.test(x, n, p = π_0)`
◆ `binom.test(x, n, p = π_0)`

Exact test, works always (but is 10 times slower than prop.test)

One-tail Test vs. Two-tail Test

There is a raging controversy (for about the last hundred years) on whether or not it is ever appropriate to use a one-tailed test. The rationale is that if you already know the direction of the difference, why bother doing any statistical tests. While it is **generally safest to use a two-tailed tests**, there are situations where a one-tailed test seems more appropriate. The bottom line is that **it is the choice of the researcher** whether to use one-tailed or two-tailed research questions.



$$2 \times p\text{-value}_{(1 \text{ tail})} = p\text{-value}_{(2 \text{ tails})}$$

Reminder: discussion around NDAs submitted to FDA, USA

Example: Hypothesis about Mean

Number of living cells in **5 wells** under some conditions are given in the table, with average value of **4705**. In a reference literature source authors claimed a mean quantity of **5000** living cells under the same conditions. Is our result significantly different?

Well	Cells
1	5128
2	4806
3	5037
4	4231
5	4322

Two Tails

$$H_0: \mu = 5000$$

$$H_a: \mu \neq 5000$$

Let's use $\alpha=0.05$

$$t = \frac{m - \mu_0}{s / \sqrt{n}}$$

n	5
mean	4704.8
stdev	409.49
mu	5000
t	-1.612
p-value 2 t	0.1823
p-value 1 t	0.0911

```
x =c(5128,4806,5037,4231,4222)
n=length(x)
m=mean(x)
s=sd(x)
mu=5000
t=(m-mu)/s*sqrt(n)
p.val.1 = pt(t,df=n-1)
p.val.2 = 2*pt(t,df=n-1)
```

In R use:

◆ **t.test(x,mu=5000)**

Example: Hypothesis about Proportion

During a study of a new drug against viral infection, you have found that **70 out of 100** mice survived, whereas the survival after the standard therapy is **60%** of the infected population. Is this enhancement statistically significant? Use error level $\alpha=0.05$

One Tail

$$H_0: \pi \leq 0.6$$

$$H_a: \pi > 0.6$$

Let's use $\alpha=0.05$

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$$

In R use:

◆ `prop.test(...)`

◆ `binom.test(...)`

```
p = 0.7
p0 = 0.6
sp = sqrt(p0*(1-p0)/n)
z = (p-p0)/sp
p.val.1 = 1-pnorm(z)
```

```
> p.val.1
[1] 0.02061342
```

```
prop.test(x=70,n=100,p=0.6,
          alternative="greater")
```

```
data: 70 out of 100, null probability 0.6
X-squared = 3.7604, df = 1, p-value = 0.02624
alternative hypothesis: true p is greater
than 0.6
```

Discrepancy for `prop.test()` comes from continuity correction.

```
#####
# L3.1. HYPOTHESES ABOUT MEAN OF A POPULATION
#####
## clear memory
rm(list = ls())
##-----
## L3.1.1. Hypotheses about mean
##-----
##Number of living cells in 5 wells under some conditions are given below.
##In a reference literature source authors claimed a mean quantity of 5000
##living cells under the same conditions. Is our result significantly different?

## put number of cells
x =c(5128,4806,5037,4231,4222)

## make analysis by hands
n=length(x)
m=mean(x)
s=sd(x)
mu=5000
t=(m-mu)/s*sqrt(n)
p.val.1 = pt(t,df=n-1)
p.val.2 = 2*pt(t,df=n-1)

## make analysis t.test()
t.test(x,mu=5000)

##-----
## L3.1.2. Hypotheses about proportion
##-----
##During a study of a new drug against viral infection, you have found that 70
##out of 100 mice survived, whereas the survival after the standard therapy is
##60% of the infected population. Is this enhancement statistically significant?
p = 0.7
p0 = 0.6
n=100
## make analysis by hands
sp = sqrt(p0*(1-p0)/n)
z = (p-p0)/sp
p.val.1 = 1-pnorm(z)

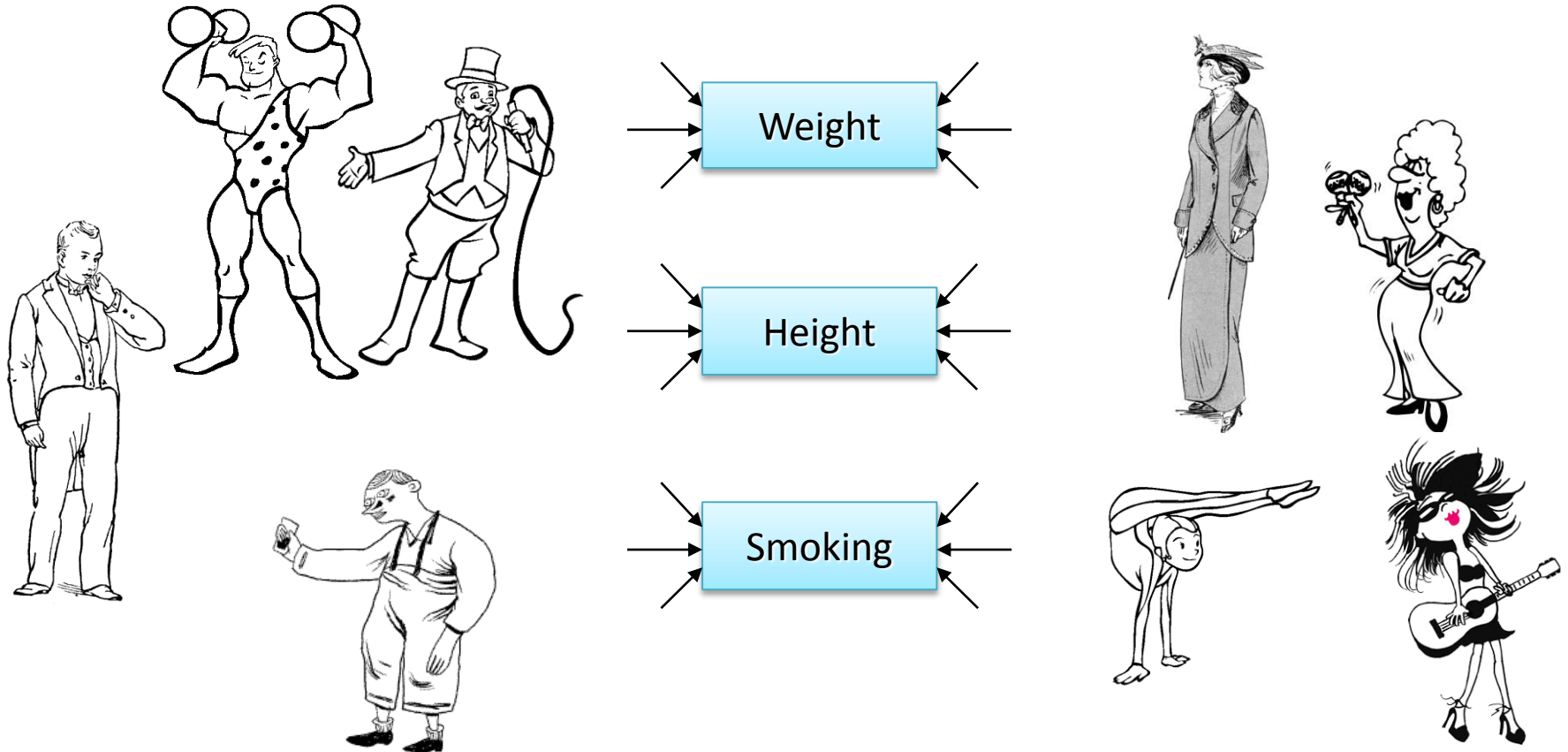
## make analysis prop.test()
prop.test(x=70,n=100,p=0.6,alternative="greater")
binom.test(x=70,n=100,p=0.6,alternative="greater")
```

Task L3.1

Independent Samples

Independent samples

Samples selected from two populations in such a way that the elements making up one sample are chosen independently of the elements making up the other sample.



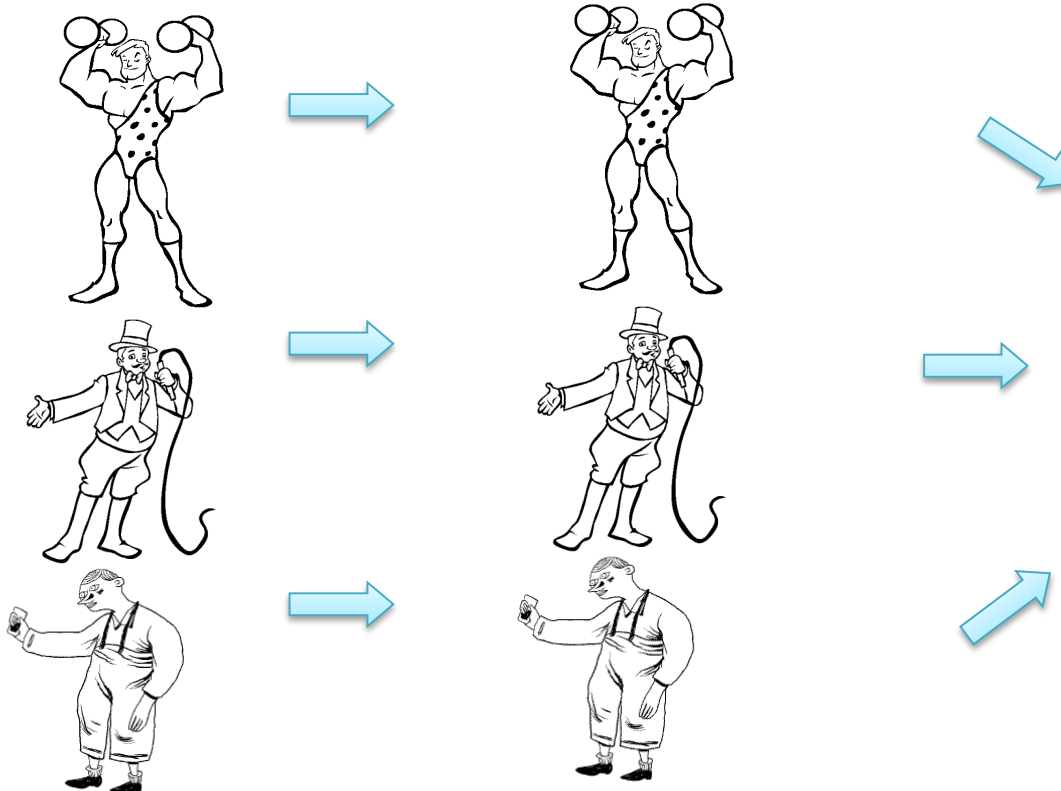
Matched Samples

Matched samples

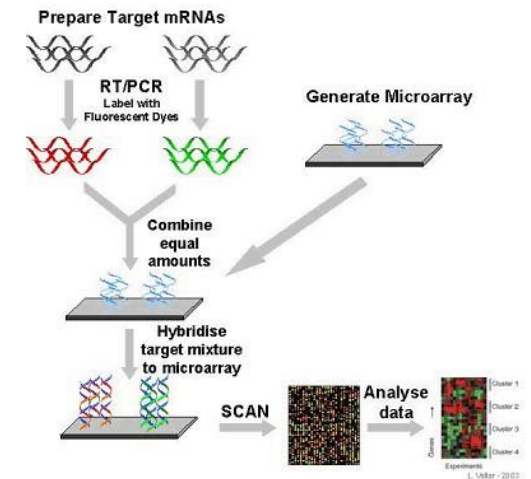
Samples in which each data value of one sample is matched with a corresponding data value of the other sample.

Before treatment

After treatment

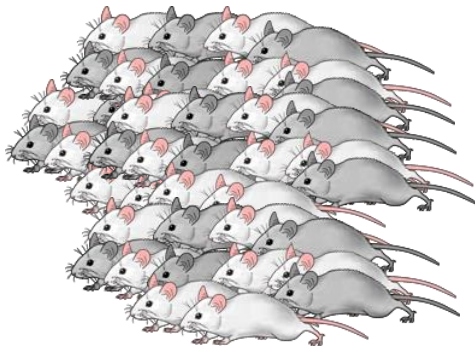


Analysis



Example: Independent Samples

mice.txt

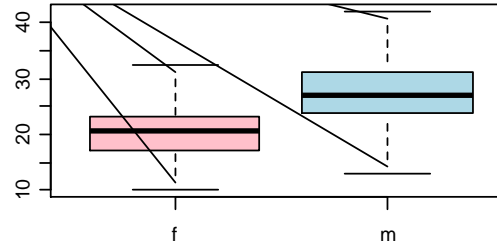


Q1: Is **body weight** for male and female significantly different?

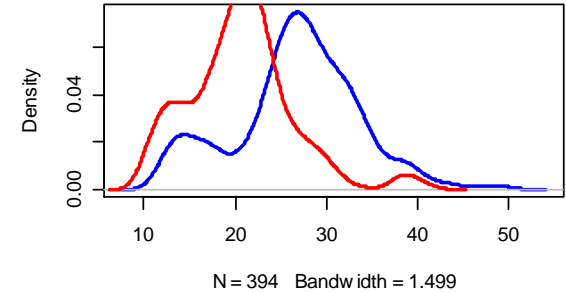
Q2: Is **weight change** for male and female significantly different?

Q3: Is **bleeding time** for male and female significantly different?

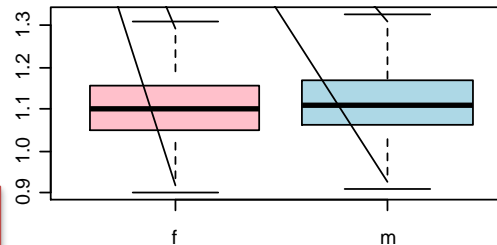
Final body weights (g)



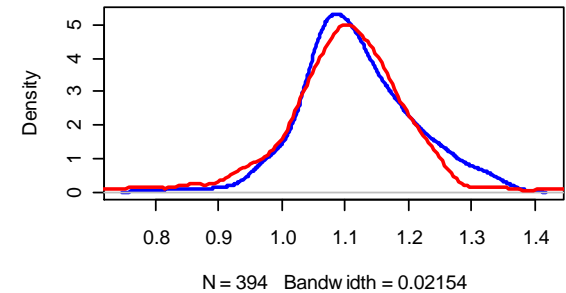
Body weight distributions



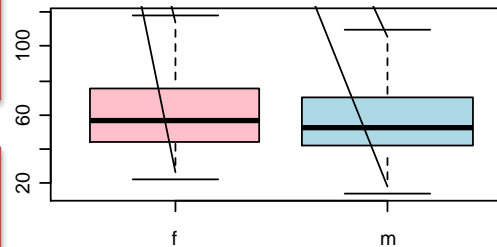
Weights change (g)



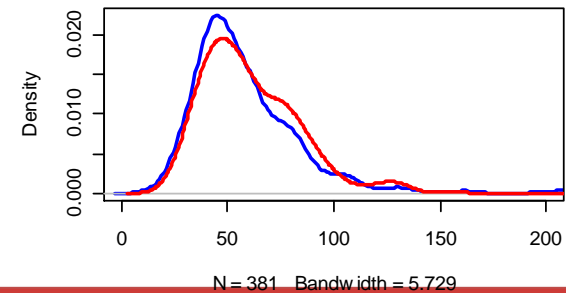
Distributions of weight change



Bleeding time (g)

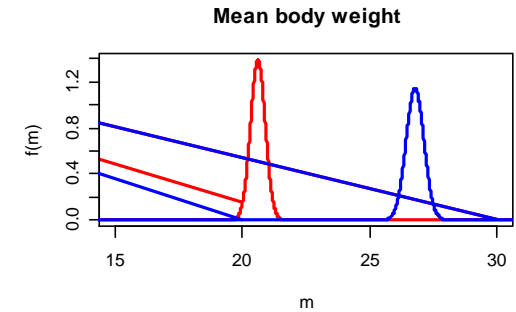
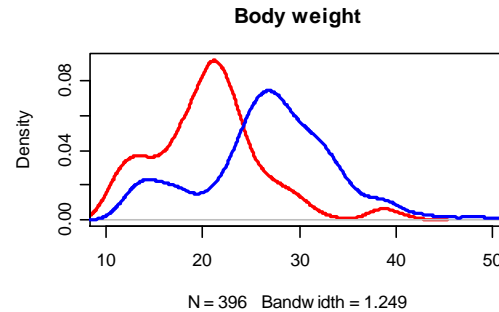


Distributions of bleeding times

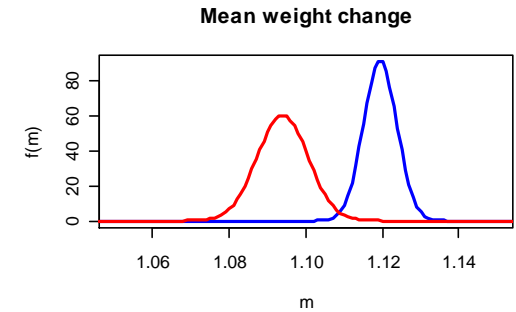
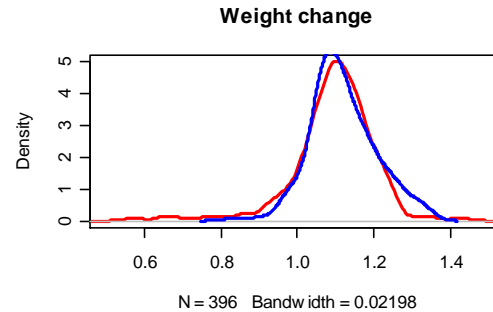


Example: Independent Samples

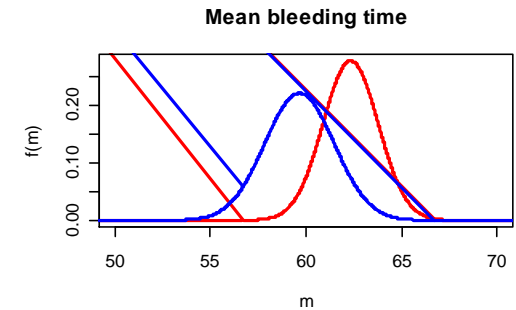
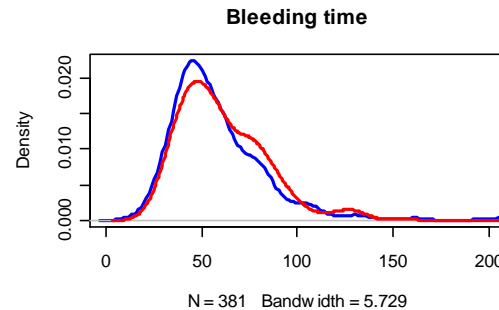
Q1: Is **body weight** for male and female significantly different?



Q2: Is **weight change** for male and female significantly different?



Q3: Is **bleeding time** for male and female significantly different?



Algorithm (manual... only for stat-geeks 😊)

$$H_0: \mu_2 - \mu_1 = D_0$$

$$H_a: \mu_2 - \mu_1 \neq D_0$$

$$D_0 = m_2 - m_1$$

Usually $D_0 = 0$

$$s_{m_2 - m_1} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

1. Build the statistics to be used for hypothesis testing:

$$t = \frac{m_2 - m_1 - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

t-distribution has following degrees of freedom:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2}$$

$$df = (n - 1) \frac{(s_1^2 + s_2^2)^2}{(s_1^4 + s_2^4)}$$

$$(n_1 + n_2) / 2 < df < n_1 + n_2$$

2. Calculate p-value

$$\blacklozenge = 2 * pt(-abs(t), df)$$

In Excel use:

$$\blacklozenge = T.TEST(x, y, 2, 3)$$

In R



In fact we do not need these calculations:

In R use (parametric):

```
◆ t.test( x, y, alternative=...)
```

In R use (non-parametric):

```
◆ wilcox.test ( x, y, alternative=...)
```

`bloodpressure.txt`

Systolic blood pressure (mmHg)

Subject	BP before	BP after
1	122	127
2	126	128
3	132	140
4	120	119
5	142	145
6	130	130
7	142	148
8	137	135
9	128	129
10	132	137
11	128	128
12	129	133

Test	p-value
unpaired	0.414662
paired	0.014506

Paired Samples

The systolic blood pressures of $n=12$ women between the ages of 20 and 35 were measured before and after usage of a newly developed oral contraceptive.

Q: Does the treatment affect the systolic blood pressure?

Unpaired test

`= t.test (x, y)`

Paired test

`= t.test (x, y, paired=T)`

In R use (parametric):

◆ `t.test(x, y, paired=T)`

In R use (non-parametric):

◆ `wilcox.test(x, y, paired=T)`

Hypothesis about Proportions of 2 Populations

$$H_0: \pi_1 = \pi_2$$

$$H_a: \pi_1 \neq \pi_2$$

$$H_0: \pi_1 - \pi_2 = 0$$

$$H_a: \pi_1 - \pi_2 \neq 0$$

$$\sigma_{p_1 - p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Pooled estimator of π

An estimator of a population proportion obtained by computing a weighted average of the point estimators obtained from two independent samples.

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$\sigma_{p_1 - p_2} = \sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$\blacklozenge = 2 * \text{pnorm}(-\text{abs}(z))$$

$$\blacklozenge = 2 * \text{pnorm}(-\text{abs}(z))$$

Example: Hypothesis about Proportions of 2 Populations

SWR/J	MA/MyJ
f	f
f	f
f	f
f	f
f	f
f	f
f	f
f	f
f	f
f	f
f	f
f	m
f	m
m	m
m	m
m	m
m	m
m	m
m	m
m	m
m	m
m	m
m	m
m	m
m	m
m	m
m	m
m	m
m	m

`mice.txt`

Q: Is the male proportion significantly different in these mouse strains (0.47 and 0.65)?

	SWR/J	MA/MyJ	pooled
count male	9	15	24
n	19	23	42
p	0.474	0.652	0.571
z	-1.16		
p-val	0.244658997		

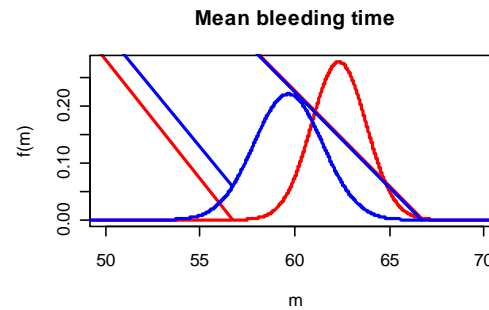
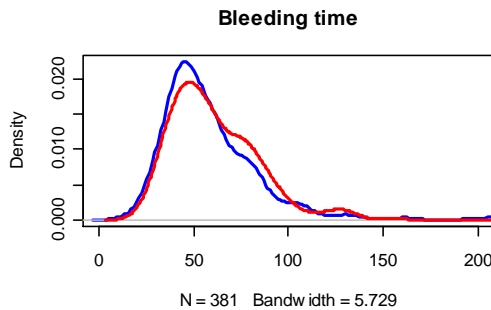
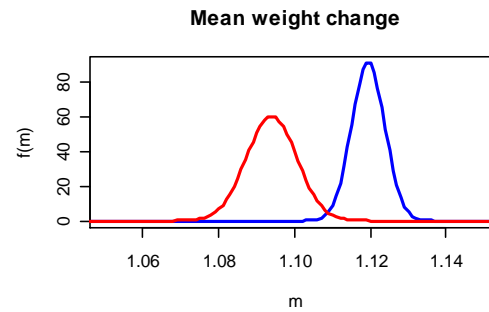
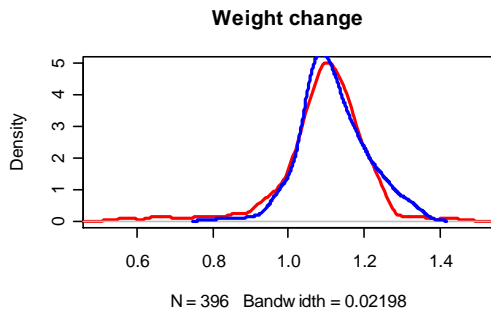
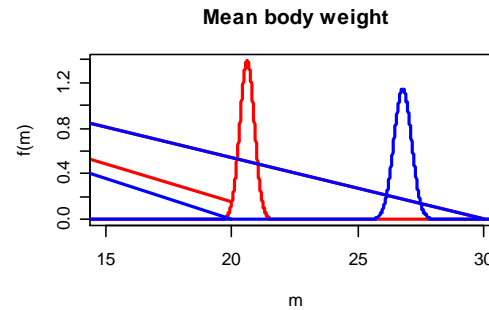
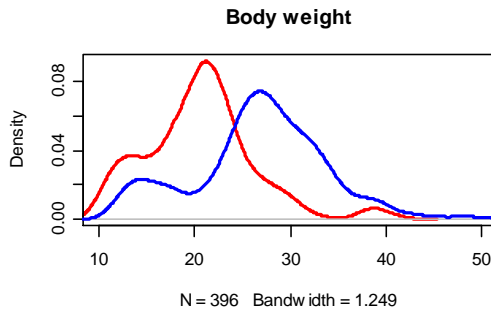
In R use:
◆ `prop.test(...)`

```
prop.test(c(9, 15), n=c(19, 23), correct=F)
prop.test(c(9, 15), n=c(19, 23))
```

p-value = 0.3952

Discrepancy comes from continuity correction.

Non-parametric Tests



◆ **T-test**, p-val < 2.2e-16
 ◆ **Wilcox**, p-val < 2.2e-16

◆ **T-test**, p-val = 0.0014
 ◆ **Wilcox**, p-val = 0.0299

Explanations?

◆ **T-test**, p-val = 0.2487
 ◆ **Wilcox**, p-val = 0.0178

In R use:
 ◆ `wilcox.test(x, y)`

```
#####
# L3.2. HYPOTHESES ABOUT MEANS OF TWO POPULATION
#####

##-----
## L3.2.1. Unmatched samples
##-----

Mice=read.table("http://edu.sablab.net/data/txt/mice.txt",header=T,sep="\t")
str(Mice)

## calculate p-value by hands
y=Mice$Weight.change[Mice$Sex=="m"]
x=Mice$Weight.change[Mice$Sex=="f"]
s1=sd(x)^2/length(x)
s2=sd(y)^2/length(y)
t=(mean(x)-mean(y))/sqrt(s1 + s2)
df=(s1 + s2)^2 / (s1^2/(length(x)-1) + s2^2/(length(y)-1))
2*pt(-abs(t),df)

## calculate p-value by t.test
t.test(x,y)

##-----
## Let us plot the distributions in a nice way
x11()
par(mfrow=c(3,2))
boxplot(Ending.weight~Sex,data=Mice,col=c("pink","lightblue"),outline=F)
title("Ending weight (g)")
plot(density(Mice$Ending.weight[Mice$Sex=="m"]),col="blue",,lwd=2, main="")
lines(density(Mice$Ending.weight[Mice$Sex=="f"]),col="red",lwd=2)
title("Ending weight distributions")

## The distributions of weight change
boxplot(Weight.change~Sex,data=Mice,col=c("pink","lightblue"),outline=F)
title("Weights change")
plot(density(Mice$Weight.change[Mice$Sex=="m"]),col="blue",,lwd=2, main="")
lines(density(Mice$Weight.change[Mice$Sex=="f"]),col="red",lwd=2)
title("Distributions of weight change")

## The distributions of bleeding time
boxplot(Bleeding.time~Sex,data=Mice,col=c("pink","lightblue"),outline=F)
title("Bleeding time")
plot(density(Mice$Bleeding.time[Mice$Sex=="m"],na.rm=T),col="blue",,lwd=2, main="")
lines(density(Mice$Bleeding.time[Mice$Sex=="f"],na.rm=T),col="red",lwd=2)
title("Distributions of bleeding time")

## Perform t.tests (by default - unpaired, two.sided)
t.test(Mice$Ending.weight[Mice$Sex=="m"],Mice$Ending.weight[Mice$Sex=="f"])
t.test(Mice$Weight.change[Mice$Sex=="m"],Mice$Weight.change[Mice$Sex=="f"])
t.test(Mice$Bleeding.time[Mice$Sex=="m"],Mice$Bleeding.time[Mice$Sex=="f"])

## Perform Wilcoxon Rank Sum (equiv. Mann-Whitney test)
wilcox.test(Mice$Ending.weight[Mice$Sex=="m"],Mice$Ending.weight[Mice$Sex=="f"])
wilcox.test(Mice$Weight.change[Mice$Sex=="m"],Mice$Weight.change[Mice$Sex=="f"])
wilcox.test(Mice$Bleeding.time[Mice$Sex=="m"],Mice$Bleeding.time[Mice$Sex=="f"])

## Compare the results!

##-----
## L3.2.2. Matched samples
##-----

BP=read.table("http://edu.sablab.net/data/txt/bloodpressure.txt",header=T,sep="\t")
str(BP)

## Unpaired
t.test(BP$BP.before,BP$BP.after)
wilcox.test(BP$BP.before,BP$BP.after)

## Paired
t.test(BP$BP.before,BP$BP.after,paired=T)
wilcox.test(BP$BP.before,BP$BP.after,paired=T)

##-----
## L3.2.3. Proportions
##-----

Mice=read.table("http://edu.sablab.net/data/txt/mice.txt",header=T,sep="\t")
str(Mice)

x = Mice$Sex[Mice$Strain == "SWR/J"]
y = Mice$Sex[Mice$Strain == "MA/MyJ"]

prop.test(c(sum(x=="m"),sum(y=="m")),n=c(length(x),length(y)))
prop.test(c(9,15),n=c(19,23),correct=F)

## Perform t.tests (by default - unpaired, two.sided)
t.test(Mice$Ending.weight[Mice$Sex=="m"],Mice$Ending.weight[Mice$Sex=="f"])
t.test(Mice$Weight.change[Mice$Sex=="m"],Mice$Weight.change[Mice$Sex=="f"])
t.test(Mice$Bleeding.time[Mice$Sex=="m"],Mice$Bleeding.time[Mice$Sex=="f"])

## Perform Wilcoxon Rank Sum (equiv. Mann-Whitney test)
wilcox.test(Mice$Ending.weight[Mice$Sex=="m"],Mice$Ending.weight[Mice$Sex=="f"])
wilcox.test(Mice$Weight.change[Mice$Sex=="m"],Mice$Weight.change[Mice$Sex=="f"])
wilcox.test(Mice$Bleeding.time[Mice$Sex=="m"],Mice$Bleeding.time[Mice$Sex=="f"])

```

Task L3.2

Hypotheses about Population Variance

$$H_0: \sigma^2 \leq \text{const}$$

$$H_a: \sigma^2 > \text{const}$$

$$H_0: \sigma^2 \geq \text{const}$$

$$H_a: \sigma^2 < \text{const}$$

$$H_0: \sigma^2 = \text{const}$$

$$H_a: \sigma^2 \neq \text{const}$$

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0: \sigma^2 \geq \sigma_0^2$ $H_a: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 \leq \sigma_0^2$ $H_a: \sigma^2 > \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 \neq \sigma_0^2$
Test Statistic	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$
Rejection Rule: p-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $\chi^2 \leq \chi_{(1-\alpha)}^2$	Reject H_0 if $\chi^2 \geq \chi_{\alpha}^2$	Reject H_0 if $\chi^2 \leq \chi_{(1-\alpha/2)}^2$ or if $\chi^2 \geq \chi_{\alpha/2}^2$

Sampling Distribution

In many statistical applications we need a comparison between variances of two populations. In fact well-known ANOVA-method is base on this comparison.

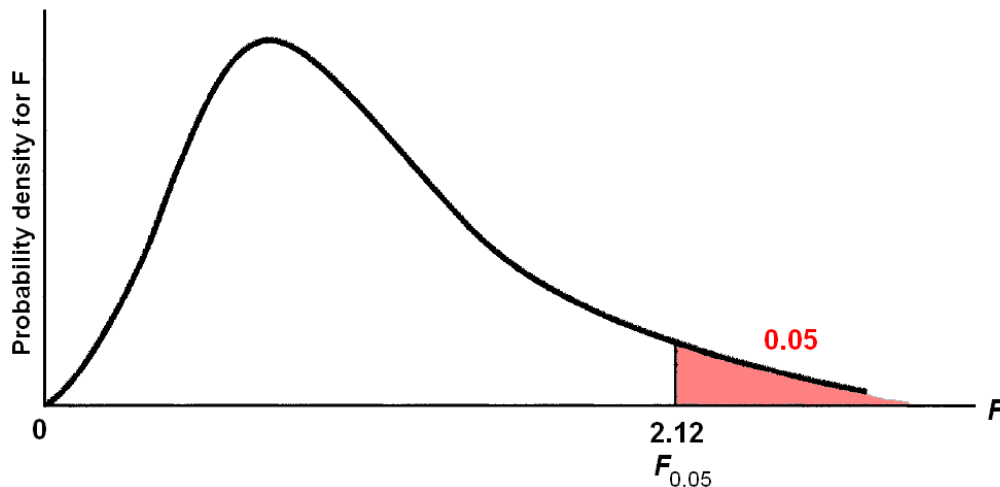
The statistics is build for the following measure:

$$F = \frac{s_1^2}{s_2^2}$$

Sampling distribution of s_1^2/s_2^2 when $\sigma_1^2 = \sigma_2^2$

Whenever a independent simple random samples of size n_1 and n_2 are selected from two normal populations with equal variances, the sampling of s_1^2/s_2^2 has **F-distribution** with n_1-1 degree of freedom for numerator and n_2-1 for denominator.

F-distribution for 20 d.f. in numerator and 20 d.f. in denominator



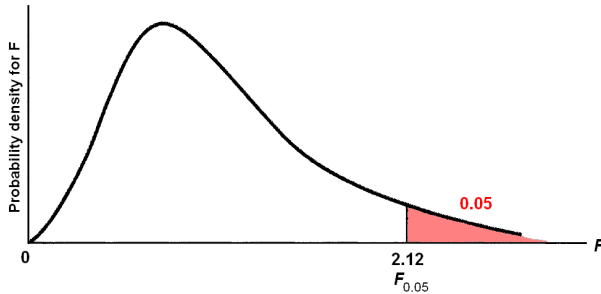
In Excel use functions:

$$\diamond = \text{FTEST}(x, y)$$

In R use:

$$\diamond \text{ var.test}(x, y)$$

Hypotheses about Variances of Two Populations



$$H_0: \sigma_1^2 \leq \sigma_2^2$$

$$H_a: \sigma_1^2 > \sigma_2^2$$

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_a: \sigma_1^2 \neq \sigma_2^2$$

	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0: \sigma_1^2 \leq \sigma_2^2$ $H_a: \sigma_1^2 > \sigma_2^2$	$H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 \neq \sigma_2^2$ <p><i>Note: Population 1 has the larger sample variance</i></p>
Test Statistic	$F = \frac{s_1^2}{s_2^2}$	$F = \frac{s_1^2}{s_2^2}$
Rejection Rule: p-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $F \geq F_\alpha$	Reject H_0 if $F \geq F_\alpha$

Example

schoolbus.txt

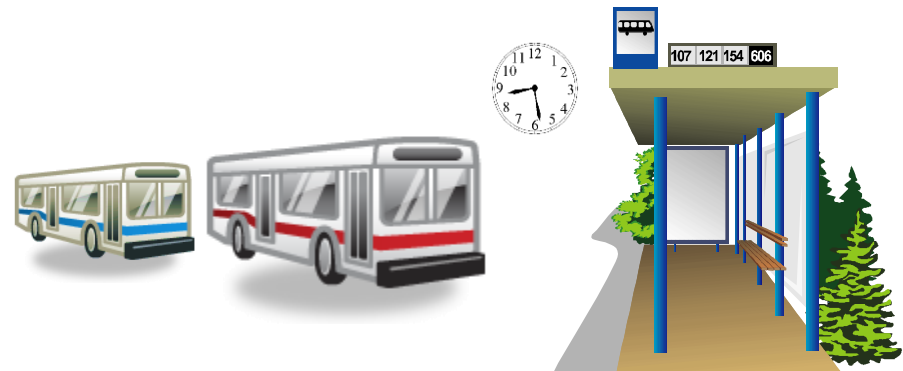
#	Milbank	Gulf Park
1	35.9	21.6
2	29.9	20.5
3	31.2	23.3
4	16.2	18.8
5	19.0	17.2
6	15.9	7.7
7	18.8	18.6
8	22.2	18.7
9	19.9	20.4
10	16.4	22.4
11	5.0	23.1
12	25.4	19.8
13	14.7	26.0
14	22.7	17.1
15	18.0	27.9
16	28.1	20.8
17	12.1	
18	21.4	
19	13.4	
20	22.9	
21	21.0	
22	10.1	
23	23.0	
24	19.4	
25	15.2	
26	28.2	

Dullus County Schools is renewing its school bus service contract for the coming year and must select one of two bus companies, the Milbank Company or the Gulf Park Company. We will use the variance of the arrival or pickup/delivery times as a primary measure of the quality of the bus service. Low variance values indicate the more consistent and higher-quality service. If the variances of arrival times associated with the two services are equal, Dullus School administrators will select the company offering the better financial terms. However, if the sample data on bus arrival times for the two companies indicate a significant difference between the variances, the administrators may want to give special consideration to the company with the better or lower variance service. The appropriate hypotheses follow.

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_a: \sigma_1^2 \neq \sigma_2^2$$

If H_0 can be rejected, the conclusion of unequal service quality is appropriate. We will use a level of significance of $\alpha = .10$ to conduct the hypothesis test.



Example

schoolbus.txt

#	Milbank	Gulf Park
1	35.9	21.6
2	29.9	20.5
3	31.2	23.3
4	16.2	18.8
5	19.0	17.2
6	15.9	7.7
7	18.8	18.6
8	22.2	18.7
9	19.9	20.4
10	16.4	22.4
11	5.0	23.1
12	25.4	19.8
13	14.7	26.0
14	22.7	17.1
15	18.0	27.9
16	28.1	20.8
17	12.1	
18	21.4	
19	13.4	
20	22.9	
21	21.0	
22	10.1	
23	23.0	
24	19.4	
25	15.2	
26	28.2	

1. Let us start from estimation of the **variances** for 2 data sets

Milbank: $s_1^2 = 48$

Milbank: $\sigma_1^2 \approx 48$ (29.5 ÷ 91.5)

Gulf Park: $s_2^2 = 20$

Gulf Park: $\sigma_2^2 \approx 20$ (10.9 ÷ 47.9)

2. Let us calculate the **F-statistics**

$$F = \frac{s_1^2}{s_2^2} = \frac{48}{20} = 2.40$$

3. ... and **p-value** = 0.08

```
var.test(x,y)
```

```
F test to compare two variances
```

```
data: Bus[, 1] and Bus[, 2]
```

```
F = 2.401, num df = 25, denom df = 15, p-value = 0.08105
```

```
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
```

```
0.8927789 5.7887880
```

```
sample estimates:
```

```
ratio of variances
```

```
2.401036
```

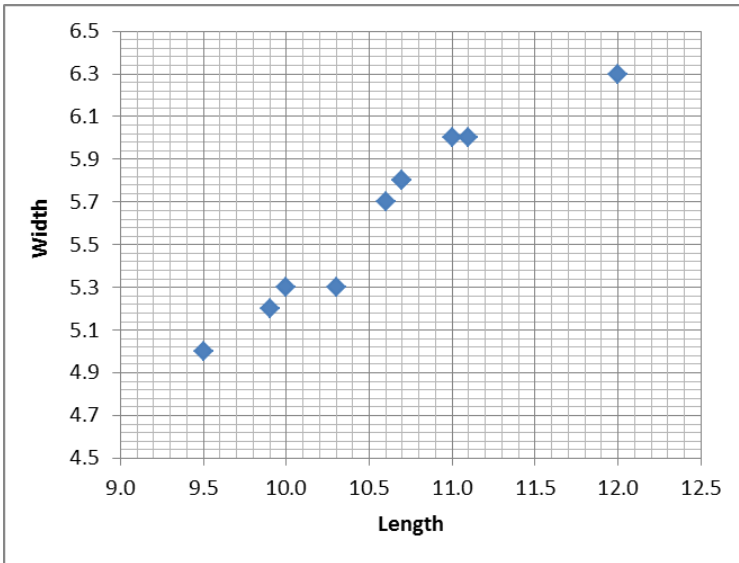
Significance of Correlation

A malacologist interested in the morphology of West Indian chitons, *Chiton olivaceus*, measured the length and width of the eight overlapping plates composing the shell of 10 of these animals.

chiton.txt



Length	Width
10.7	5.8
11.0	6.0
9.5	5.0
11.1	6.0
10.3	5.3
10.7	5.8
9.9	5.2
10.6	5.7
10.0	5.3
12.0	6.3



$r = 0.9692$, is it significant?

Test hypotheses:

$$H_0: \rho = 0$$

$$H_a: \rho \neq 0$$

Assume x, y has normal distributions, $\rho = 0$, then perform a one sample t-test with following parameters:

$$s_r = \sqrt{\frac{1-r^2}{n-2}}$$

Degree of freedom $df = n - 2$

Significance of Correlation

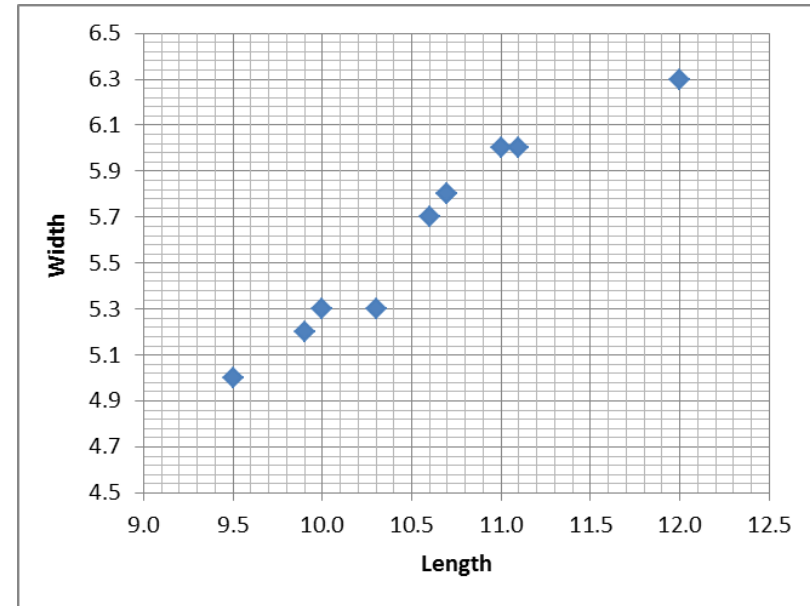
$$r = 0.9692$$

$$s_r = \sqrt{\frac{1-r^2}{n-2}}$$

Degree of freedom $df = n - 2$

$$t = \frac{r - 0}{s_r} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

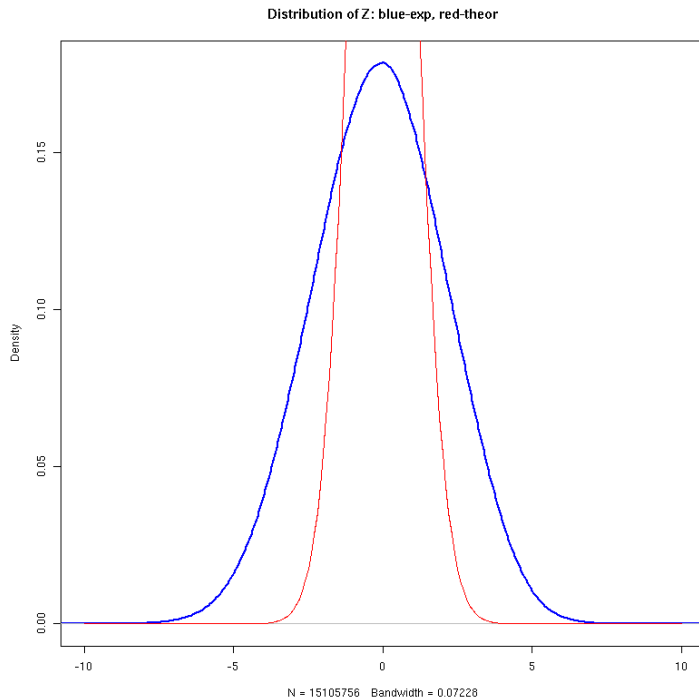
$t = 11.14,$
 $p\text{-value} = 4e-6$



In R use:

◆ `cor.test(x, y)`

Comparison of 2 Correlations



Fisher's transformation

$$z = 0.5 \ln \left(\frac{1+r}{1-r} \right)$$

z-statistics for the difference in correlation

$$\leftarrow Z = \frac{z_1 - z_2}{\sqrt{(n_1 - 3)^{-1} + (n_2 - 3)^{-1}}}$$

Use standard normal distribution to assign p-value of identified Z

Type II Error

Type I error

The error of rejecting H_0 when it is true.

Type II error

The error of accepting H_0 when it is false.

Level of significance

The probability of making a Type I error when the null hypothesis is true as an equality

poor sensitivity

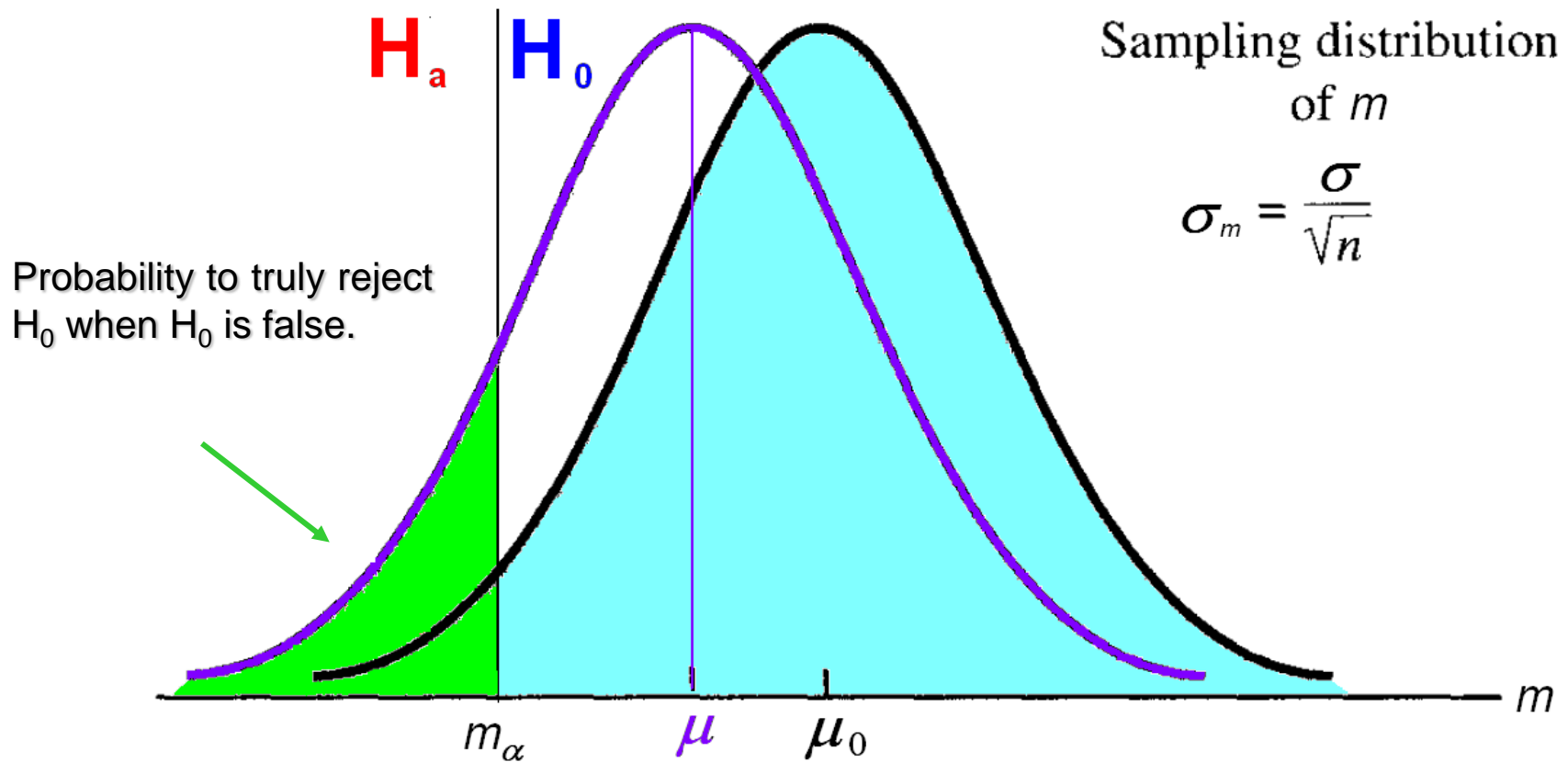
**False Negative,
 β error**

		Population Condition	
		H_0 True	H_a True
Conclusion	Accept H_0	Correct Conclusion	Type II Error
	Reject H_0	Type I Error	Correct Conclusion

**False Positive,
 α error**

poor specificity

Power Curve



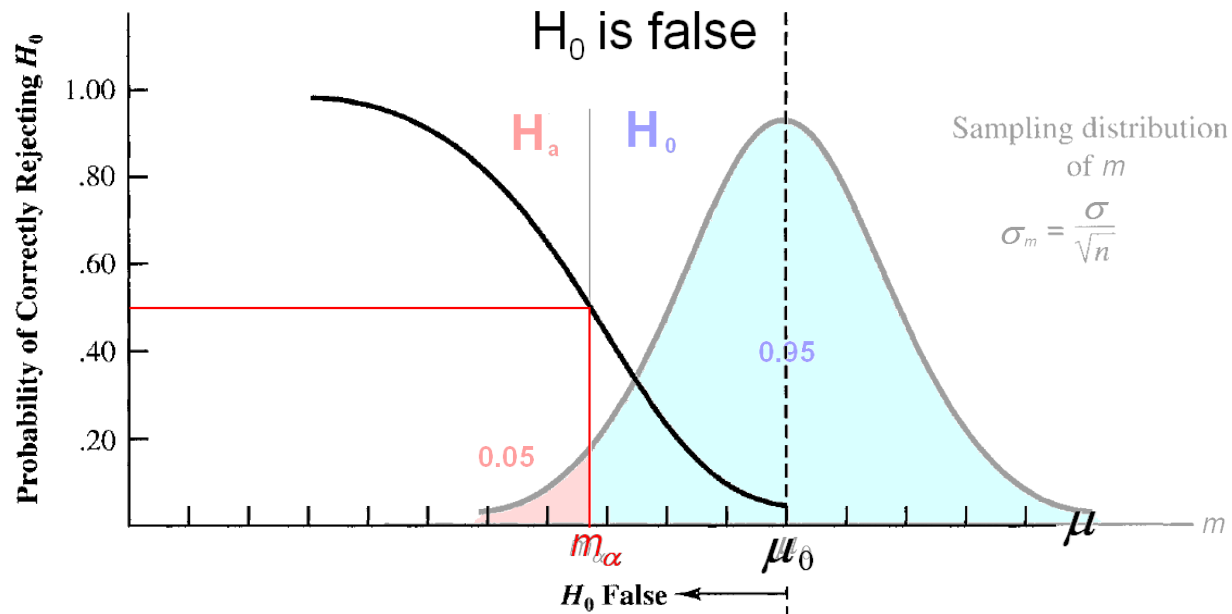
Power Curve

Power

The probability of correctly rejecting H_0 when it is false

Power curve

A graph of the probability of rejecting H_0 for all possible values of the population parameter not satisfying the null hypothesis. The power curve provides the probability of correctly rejecting the null hypothesis



Power Analysis in R

In R use:

- ◆ `power.t.test (...)`
- ◆ `power.prop.test (...)`

Or **pwr** package

- ◆ `pwr.2p.test (...)`
- ◆ `pwr.t.test (...)`
- ◆ ...

<http://www.statmethods.net/stats/power.html>

Thank you for your attention



to be continued...