CENTRE DE RECHERCHE PUBLIC

Microarray Center

APPLIED STATISTICS

Lecture 6

Hypothesis Tests for Means

Petr Nazarov
petr.nazarov@crp-sante.lu

24-11-2009

- **Hypothesis tests for means and proportions**
  - zzz
  - xxx
- **Hypothesis test for means of two populations**
  - interval estimation
  - population mean: σ known
  - population mean: σ unknown
  - Student's distribution
  - estimation the size of a sample
  - population proportion

Null and Alternative 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: $\mu \leq$ const	H_0: $\mu \geq$ const	H_0: $\mu =$ const
H_a: $\mu >$ const	H_a: $\mu <$ const	H_a: $\mu \neq$ const

Developing Null and Alternative Hypotheses: Example 1

Consider a particular automobile model that currently attains an average fuel efficiency of 24 miles per galon. A product research group developed a new fuel injection system specifically designed to increase the miles-per-gallon rating. To evaluate the new system, several will be manufactured, installed in automobiles, and subjected to research-controlled driving tests. Here the product research group is looking for evidence to conclude that the new system increases the mean miles-per-gallon rating. In this case, the research hypothesis is that the new fuel injection system will provide a mean miles-per-gallon rating exceeding 24, that is, $\mu > 24$. As a general guideline, a research hypothesis should be stated as the alternative hypothesis. Hence, the appropriate null and alternative hypotheses for the study are

$$H_0: \mu \leq 24$$

$$H_a: \mu > 24$$

If the sample results indicate that Ho cannot be rejected, researchers cannot conclude the new fuel injection system is better. Perhaps more research and subsequent testing should be conducted. However, if the sample results indicate that Ho can be rejected, researchers can make the inference that H_a: $\mu > 24$ is true. With this conclusion, the researchers gain the statistical support necessary to state that the new system increases the mean number of miles per gallon. Production with the new system should be considered.

Developing Null and Alternative Hypotheses: Example 2

Consider the situation of a manufacturer of soft drinks who states that it fills two-liter containers of its products with an average of at least 67.6 fluid ounces. A sample of two-liter containers will be selected, and the contents will be measured to test the manufacturer's claim. In this type of hypothesis testing situation, we generally assume that the manufacturer's claim is true unless the sample evidence is contradictory. Using this approach for the soft-drink example, we would state the null and alternative hypotheses as follows.

$$H_0: \mu \geq 67.6$$

$$H_a: \mu < 67.6$$

If the sample results indicate H_0 cannot be rejected, the manufacturer's claim will not be challenged. However, if the sample results indicate H_0 can be rejected, the inference will be made that H_a: $\mu < 67.6$ is true. With this conclusion, statistical evidence indicates that the manufacturer's claim is incorrect and that the soft-drink containers are being filled with a mean less than the claimed 67.6 ounces. Appropriate action against the manufacturer may be considered.

Developing Null and Alternative Hypotheses: Example 3

For example, on the basis of a sample of parts from a shipment just received, a quality control inspector must decide whether to accept the shipment or to return the shipment to the supplier because it does not meet specifications. Assume that specifications for a particular part require a mean length of two inches per part. If the mean length is greater or less than the two-inch standard, the parts will cause quality problems in the assembly operation. In this case, the null and alternative hypotheses would be formulated as follows.

$$H_0: \mu = 2$$

$$H_a: \mu \neq 2$$

If the sample results indicate H_0 cannot be rejected, the quality control inspector will have no reason to doubt that the shipment meets specifications, and the shipment will be accepted. However, if the sample results indicate H_0 should be rejected, the conclusion will be that the parts do not meet specifications. In this case, the quality control inspector will have sufficient evidence to return the shipment to the supplier.

HYPOTHESES

Type I 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

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

$$H_0\!: \mu \le \mu_0 \qquad\qquad H_0\!: \mu \ge \mu_0$$

$$H_a\!: \mu > \mu_0 \qquad\qquad H_a\!: \mu < \mu_0$$

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.

μ_0 = 3 lbm

Suppose sample of n=36 coffee cans is selected. From the previous studies it's known that σ = 0.18 lbm

One-tailed Test: Example

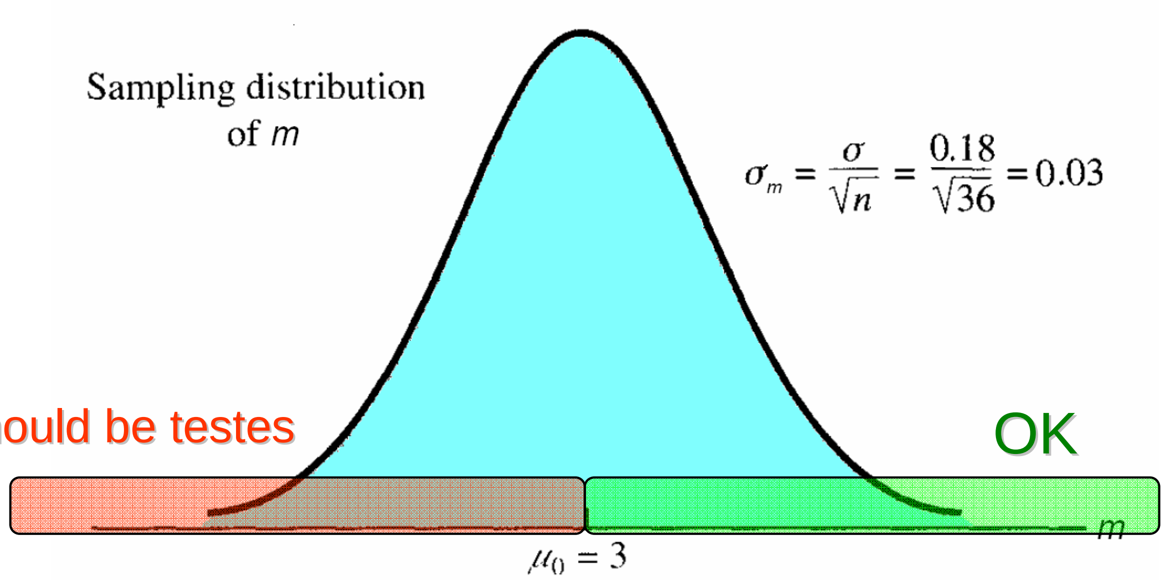
μ_0 = 3 lbm

Suppose sample of n = 36 coffee cans is selected and m = 2.92 is observed. From the previous studies it's known that σ = 0.18 lbm

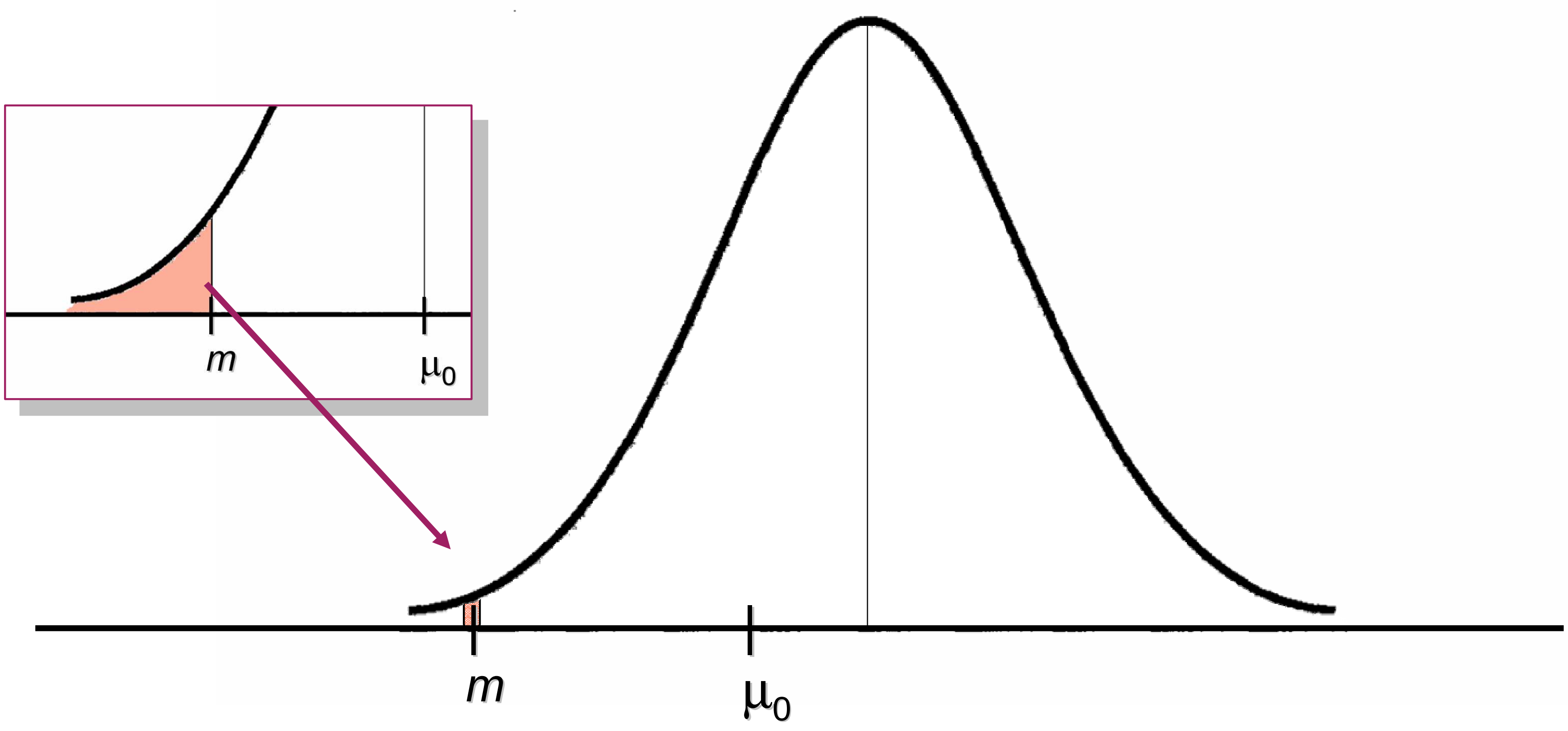
H_0: $\mu \geq 3$ no action

H_a: $\mu < 3$ legal action

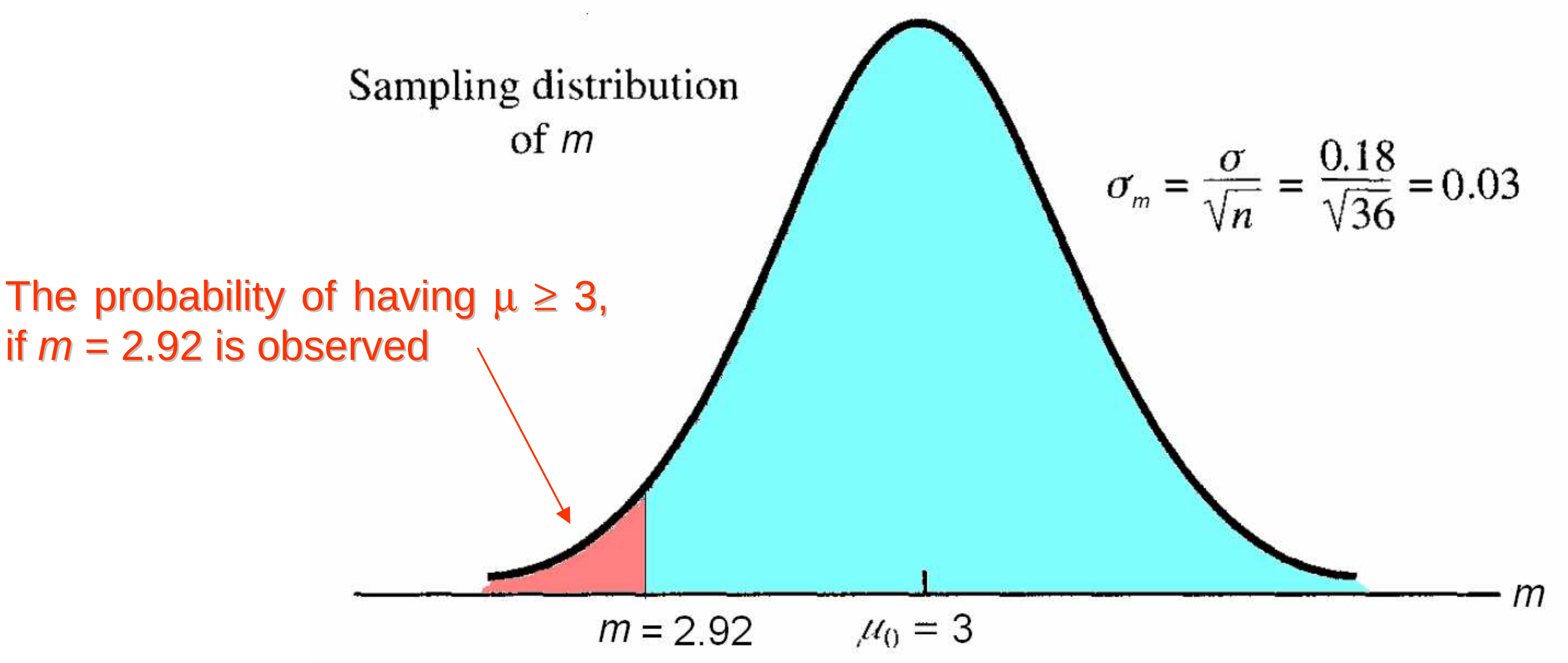
Let's say: in the extreme case, when μ=3, we would like to be 99% sure that we make no mistake, when starting legal actions against Hilltop Coffee. It means that selected significance level is $\alpha = 0.01$



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 $P(\overline{x}<m)$ for an extreme case of $\mu=\mu_0$



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 error** when rejecting the null hypothesis, or the **p-value.**

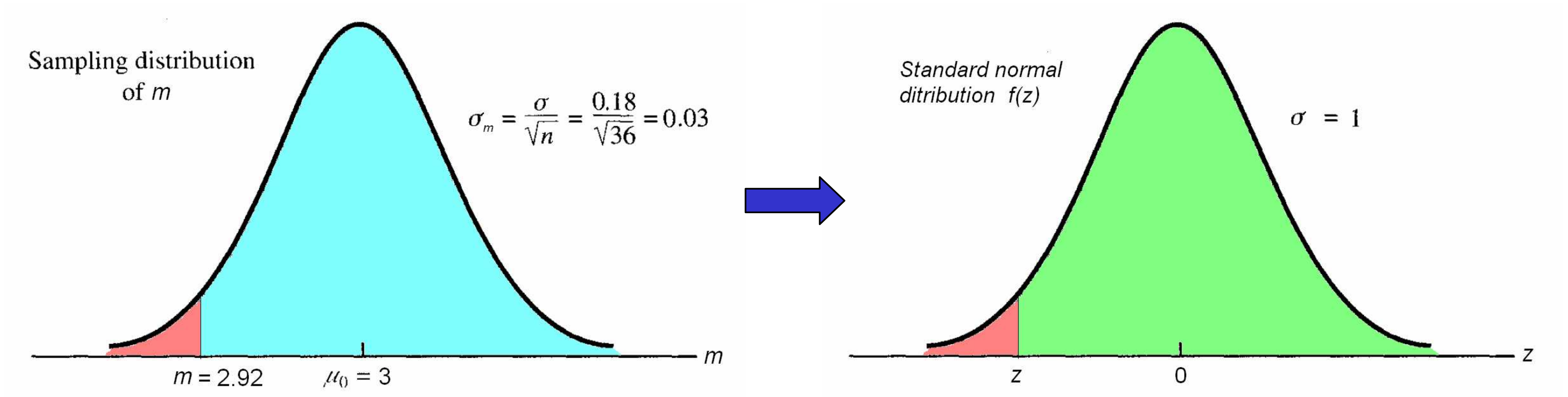
H_0: $\mu \geq 3$

H_a: $\mu < 3$

Step 1. Introduce the test statistics

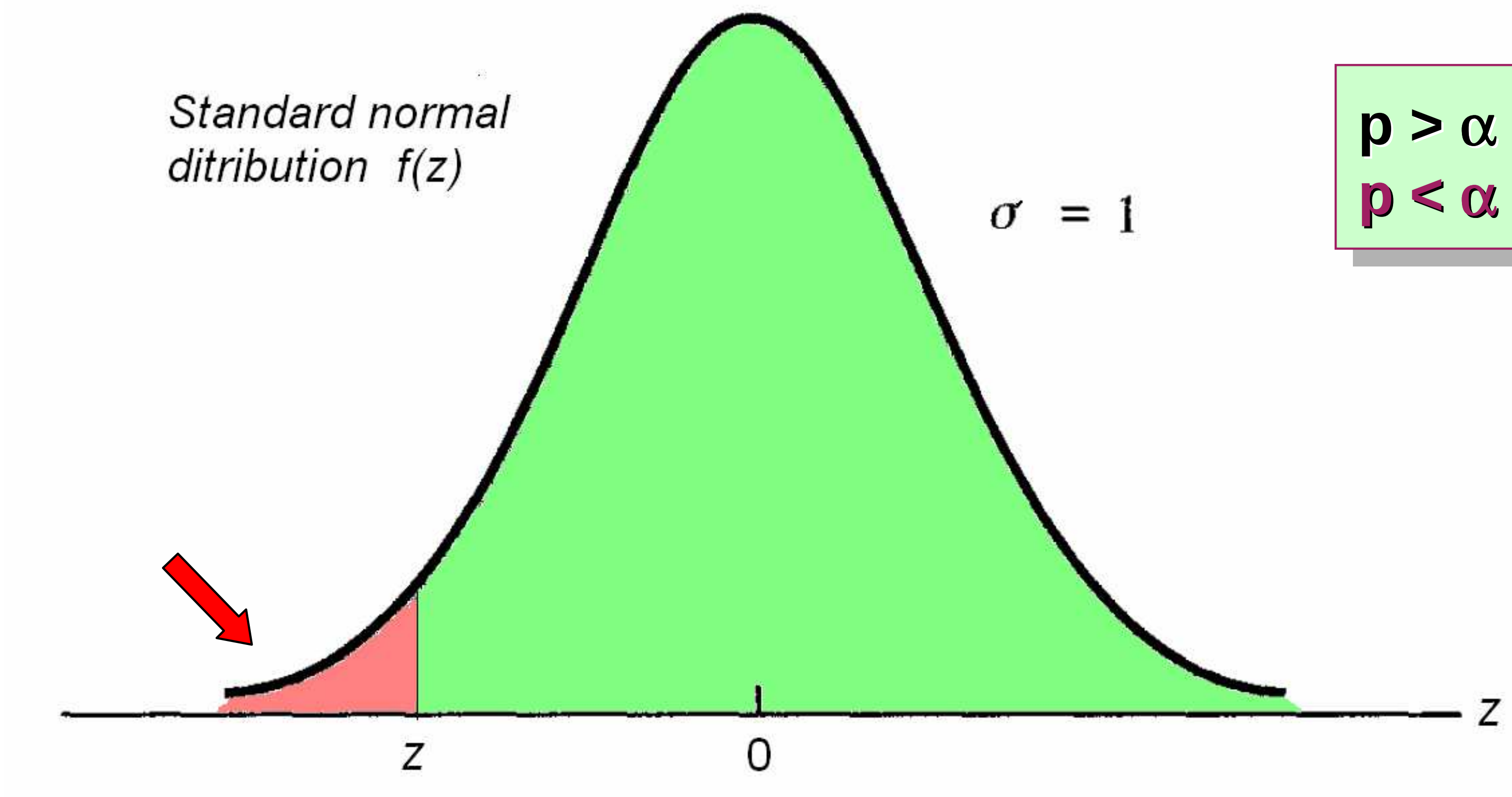
Test statistic
A statistic whose value helps determine whether a null hypothesis can be rejected

$$z = \frac{m - \mu_0}{\sigma / \sqrt{n}}$$



Step 2. Calculate p-value and compare it with α

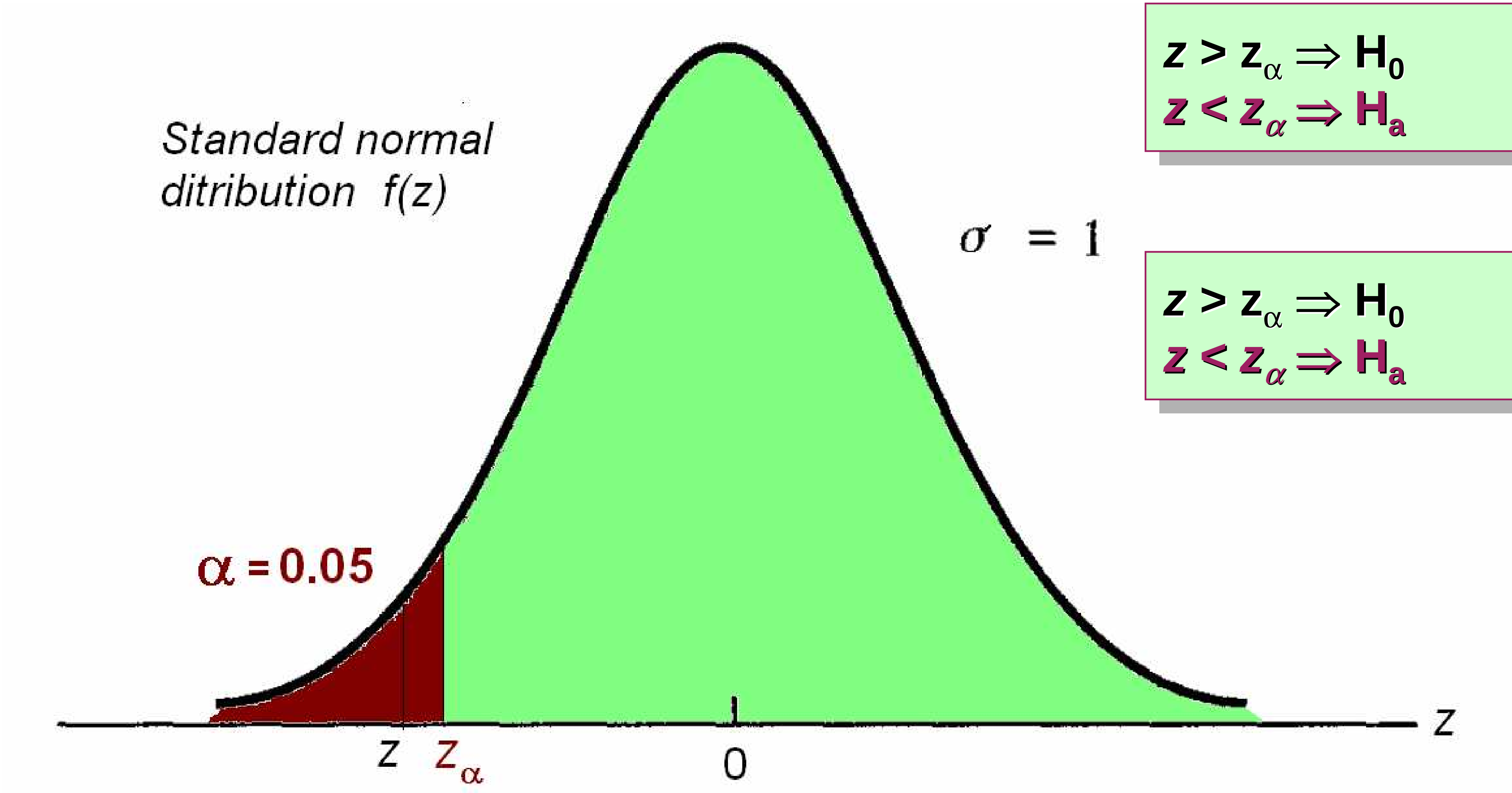
p-value
A probability, computed using the test statistic, that measures the support (or lack of support) provided by the sample for the null hypothesis. It is a probability of making error of type I



$p > \alpha \Rightarrow H_0$
$p < \alpha \Rightarrow H_a$

Critical value

A value that is compared with the test statistic to determine whether H_0 should be rejected



$z > z_\alpha \Rightarrow H_0$

$z < z_\alpha \Rightarrow H_a$

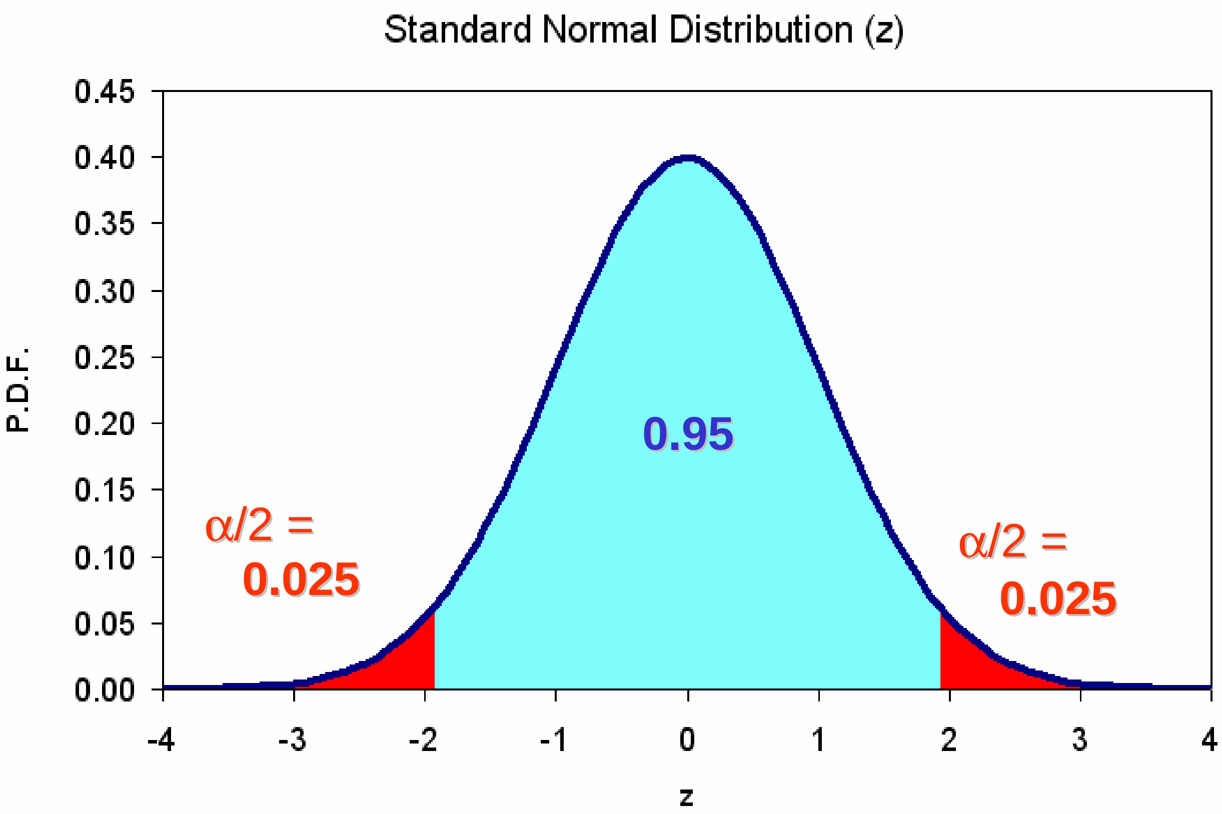
$z > z_\alpha \Rightarrow H_0$

$z < z_\alpha \Rightarrow H_a$

Two-tailed test
A hypothesis test in which rejection of the null hypothesis occurs for values of the test statistic in either tail of its sampling distribution.

H_0: $\mu = \mu_0$

H_a: $\mu \neq \mu_0$



σ is Unknown

if σ in unknown:

$\sigma \longrightarrow s$

$z \longrightarrow t$

	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	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
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 $t \geq t_{\alpha/2}$

HYPOTHESIS TESTING FOR THE PROPORTION

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.

Data

A priori information about μ and μ_0

Hypothesis testing

Two Tails

One Tail

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

HYPOTHESIS TESTING FOR THE MEAN

Example

Number of living cells in **5 wells** under some conditions are given in the table, with average value of **4651**. In a reference literature source authors clamed 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 α=0.05

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

m = AVERAGE(A2:A6)

s = STDEV(A2:A6)

μ_0 = 5000

t = (m- μ_0)/s*SQRT(5)

p-value = **TDIST**(ABS(t);5-1;2)

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

For the proportions:
1) use z-statistics
2) use proper equation for σ_p

$np \geq 5, \; n(1\text{-}p) \geq 5$

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0 : p \geq p_0$ $H_a : p < p_0$	$H_0 : p \leq p_0$ $H_a : p > p_0$	$H_0 : p = p_0$ $H_a : p \neq p_0$
Test Statistic	$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$
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 $z \geq z_{\alpha/2}$

An Easy Way to Two Tail Hypothesis

1) Formulate the hypothesis ($m = \mu_0$)
2) Select α
3) Calculate the interval estimation for $1-\alpha$ confidence
4) Check whether your μ_0 is inside the interval

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

E = **CONFIDENCE** (α, s, n)

m=	4705
E=	359
m-E=	4346
m+E=	5064

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

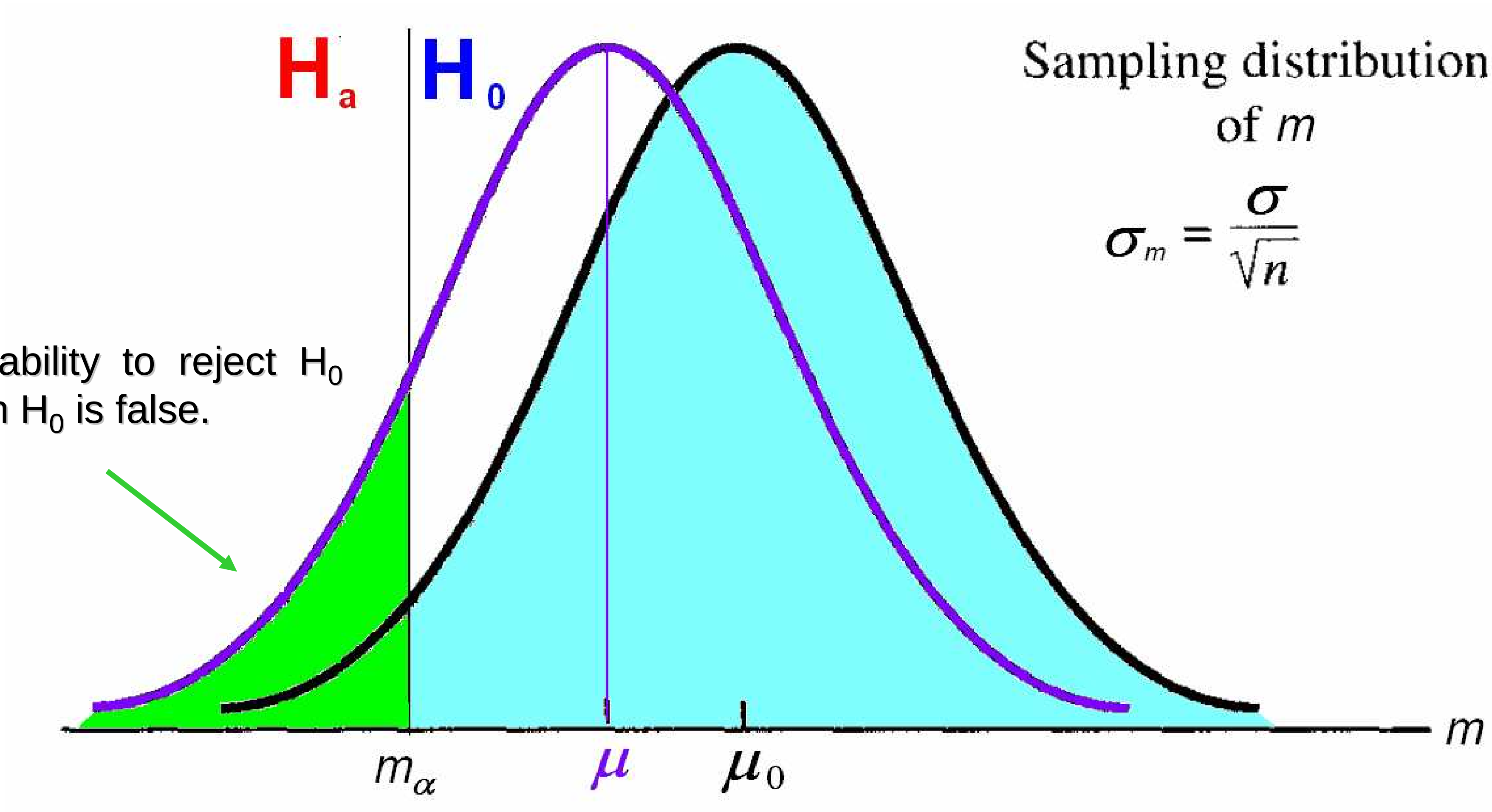
poor sensitivity

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

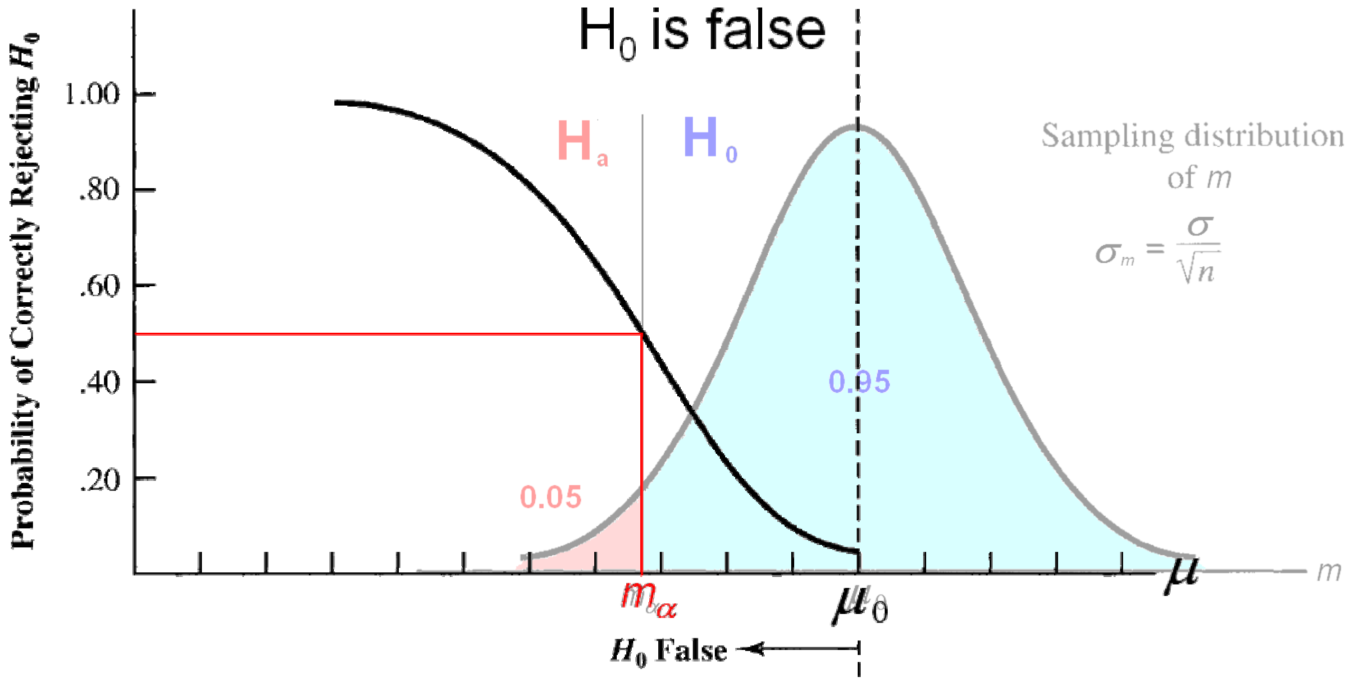
poor specificity



Probability to reject H_0 when H_0 is false.

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



Part II

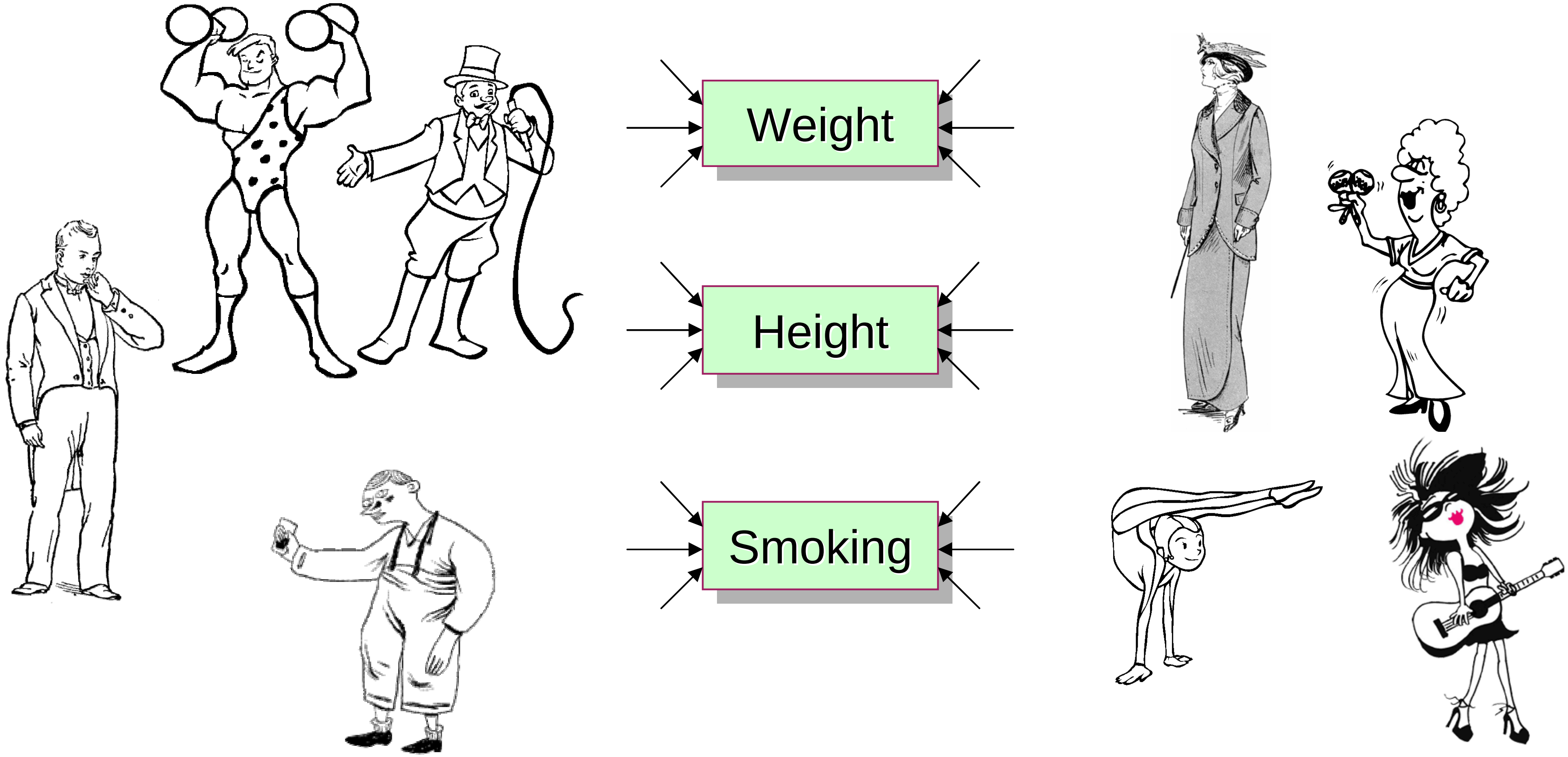
Hypothesis about Means and Proportions of Two Populations

HYPOTHESIS ABOUT MEANS OF 2 POPULATIONS

Independent Samples

Independent samples

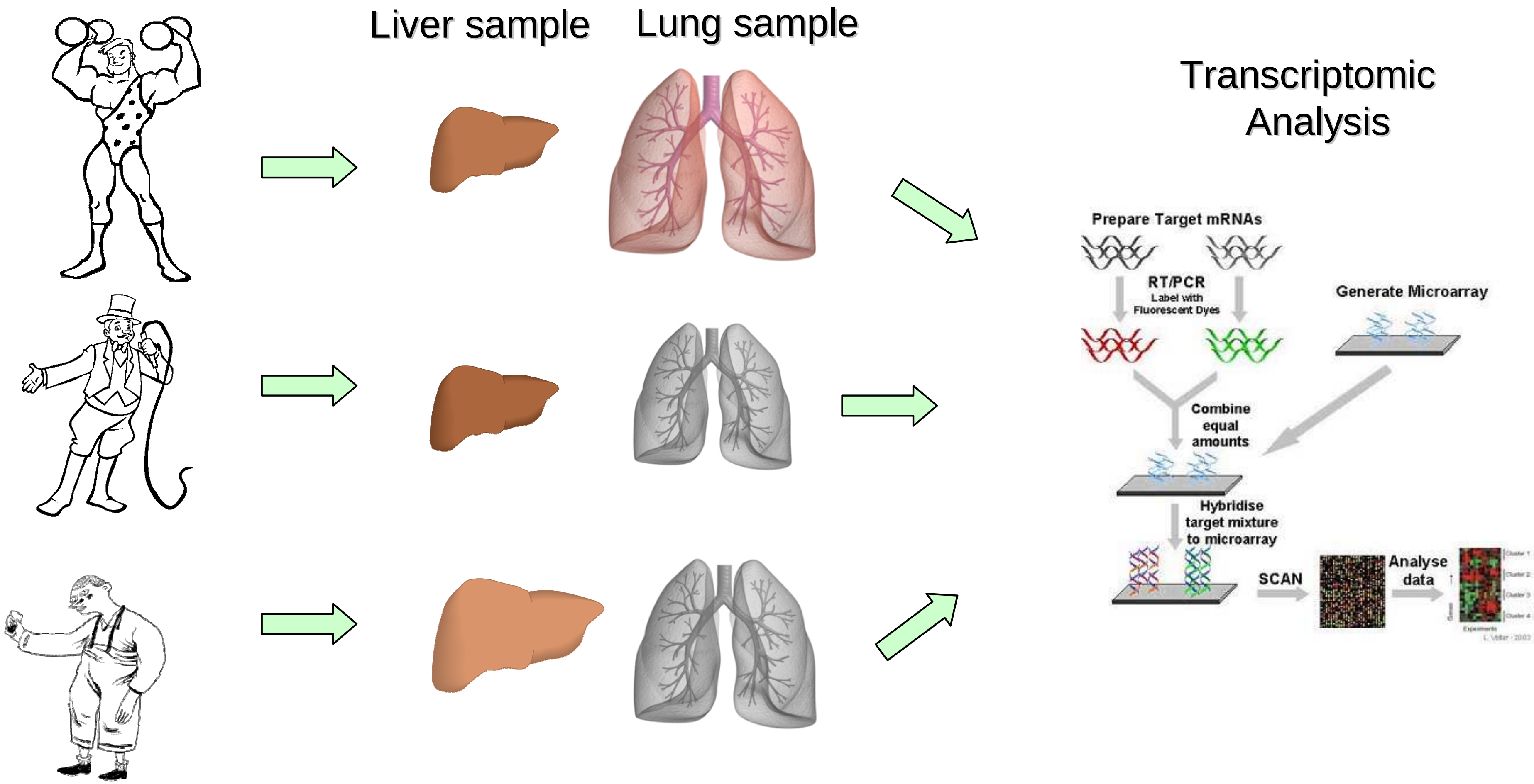
Samples selected from two populations in such a way that the elements making up one sample are chosen independently of the elements making up the other sample.



HYPOTHESIS ABOUT MEANS OF 2 POPULATIONS

Dependent Samples

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

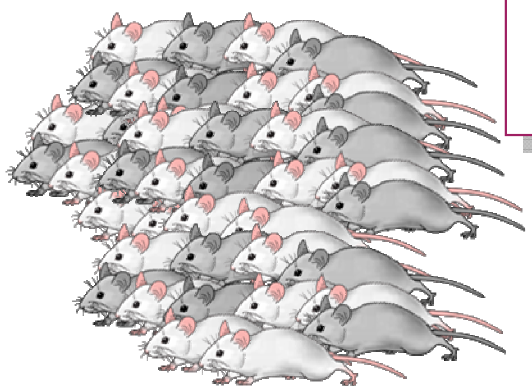


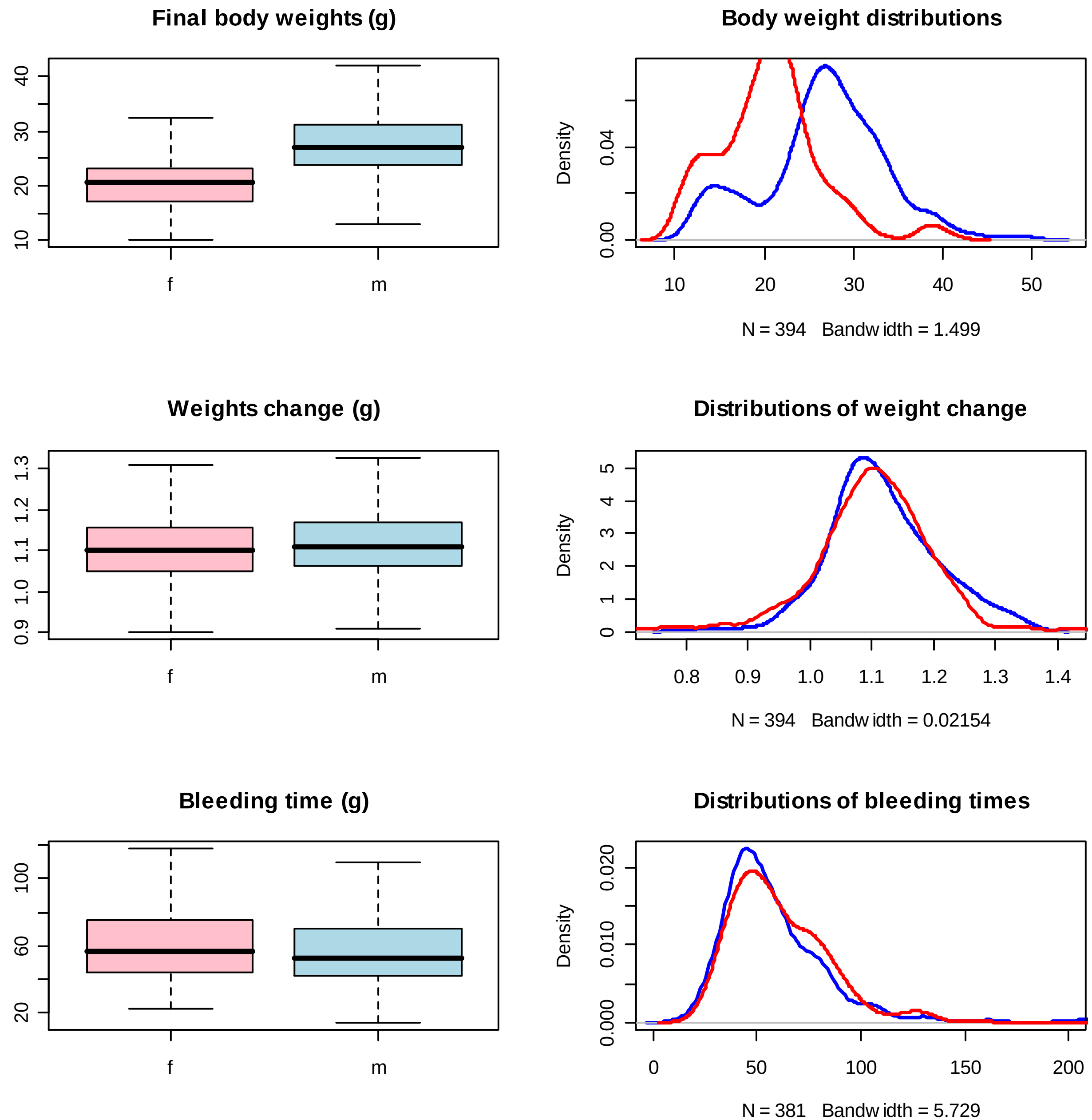
HYPOTHESIS ABOUT MEANS OF 2 POPULATIONS

Example

mice.xls

outliers are removed from boxplots





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?

CENTRE DE RECHERCHE PUBLIC SANTÉ

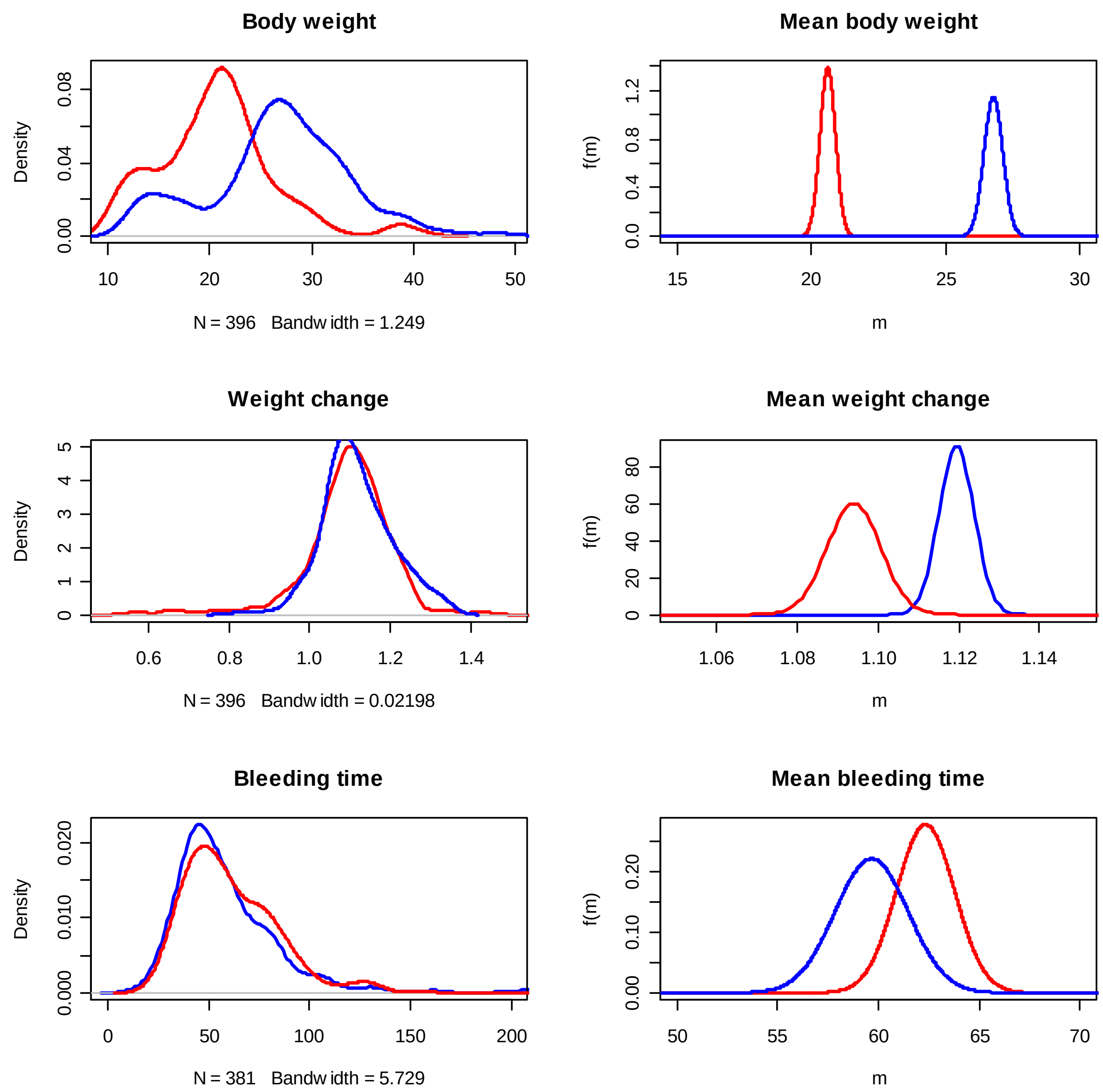
HYPOTHESIS ABOUT MEANS OF 2 POPULATIONS

Example

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?



Two tail hypothesis

H_0: $\mu_1 = \mu_2$

H_a: $\mu_1 \neq \mu_2$

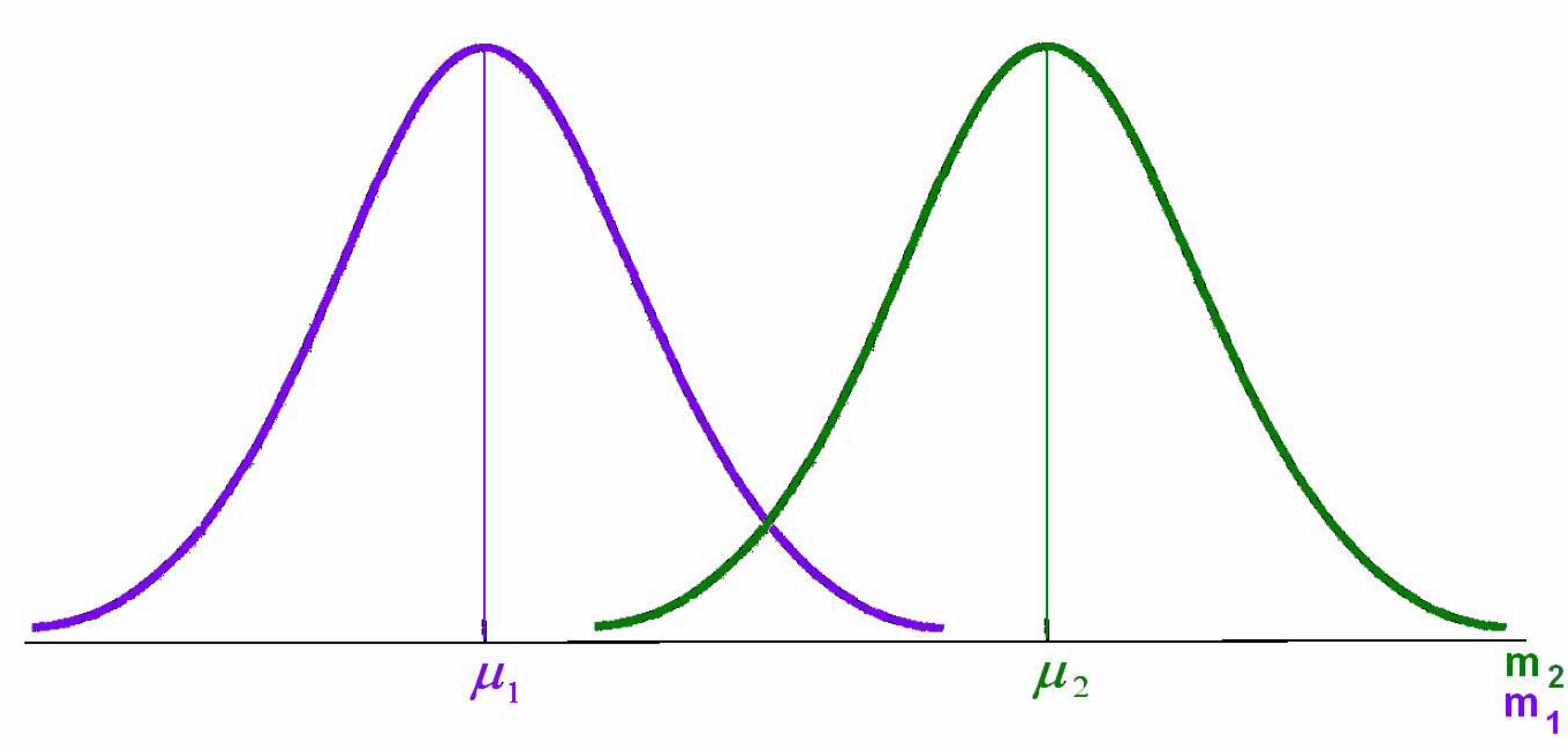
One tail hypothesis

H_0: $\mu_1 \geq \mu_2$

H_a: $\mu_1 < \mu_2$

H_0: $\mu_1 \leq \mu_2$

H_a: $\mu_1 > \mu_2$



HYPOTHESIS ABOUT MEANS OF 2 POPULATIONS

Theory

$H_0: \mu = \mu_0$

$H_a: \mu \neq \mu_0$

As we know how to work with standard hypotheses (comparison with constant μ_0), let us transform our hypothesis:

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

$\Rightarrow$

$H_0: \mu_2 - \mu_1 = 0$

$H_a: \mu_2 - \mu_1 \neq 0$

To use it, we need to know what is the distribution of **D = $m_2 - m_1$**

Distribution of sum or difference of 2 normal random variables

The sum/difference of 2 (or more) normal random variables is a normal random variable with mean equal to sum/difference of the means and variance equal to **SUM** of the variances of the compounds.

Variables	m_1	m_2	$m_2 - m_1$
Means	μ_1	μ_2	$\mu_2 - \mu_1$
Variances	σ_1^2	σ_2^2	$\sigma_1^2 + \sigma_2^2$

HYPOTHESIS ABOUT MEANS OF 2 POPULATIONS

Theory

H_0: $\mu_2 - \mu_1 = D_0$

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

$$D_0 = \mu_2 - \mu_1$$

$$\sigma_{m_2 - m_1} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$D_0 = m_2 - m_1$$

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

Statistics to be used for hypothesis testing:

if σ is known: z-statistics

$$z = \frac{m_2 - m_1 - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

if σ is unknown: t-statistics

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

This is what we call t-test !!!

Unpaired t-test: Algorithm

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:

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

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

$(n_1+n_2)/2 < \boldsymbol{df} < n_1+n_2$

2. Calculate the p-value

- `= TDIST(ABS(t),df,2)`

☺. Or simply do:

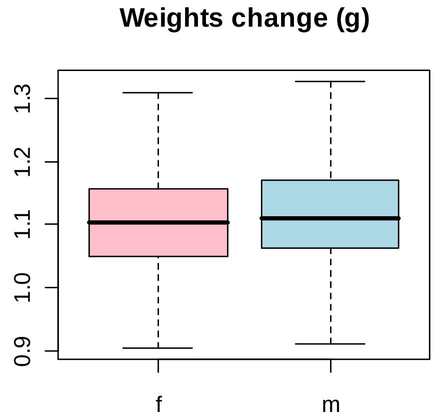
- `= TTEST (array1, array2, 2, 3)`

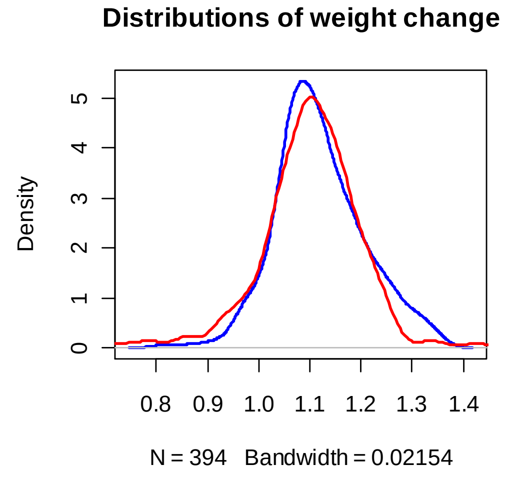
HYPOTHESIS ABOUT MEANS OF 2 POPULATIONS

Example

mice.xls

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

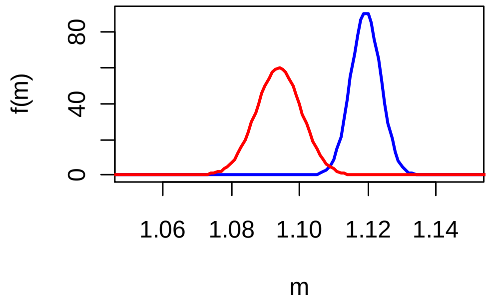




Parameter	Value	Command
Female		
m1	1.0939	=AVERAGE($B$2:$B$397)
s1	0.1320	=STDEV($B$2:$B$397)
n1	396	=COUNTIF($B$2:$B$397,">=0")
Male		
m2	1.1194	=AVERAGE($B$398:$B$791)
s2	0.0867	=STDEV($B$398:$B$791)
n2	394	=COUNTIF($B$398:$B$791,">=0")
m2-m1	0.0255	=F8-F3
s(m2-m1)	0.0079	=SQRT(F4^2/F5+F9^2/F10)
t	3.2067	=F12/F13
df ~	700	
p-value (1)	0.001403856	=TDIST(F14,700,2)
p-value (2)	0.00140541	=TTEST($B$2:$B$397,$B$398:$B$791,2,3)

- **= TTEST(array1, array2, 2, 3)**

p-value = 0.0014



Paired t-test: Example

bloodpressure.xls

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.

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

Q: Does the treatment affect the systolic blood pressure?

Unpaired test

- `= TTEST (array1, array2, 2, 3)`

Paired test

- `= TTEST (array1, array2, 2, 1)`

Test	p-value
unpaired	0.414662
paired	0.014506

HYPOTHESIS ABOUT PROPORTIONS OF 2 POPULATIONS

Theory

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.

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

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

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

```
= 2*(1-NORMDIST(ABS(z),0,1,TRUE))
```

HYPOTHESIS ABOUT PROPORTIONS OF 2 POPULATIONS

Example

SWR/J	MA/MyJ
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

mice.xls

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

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

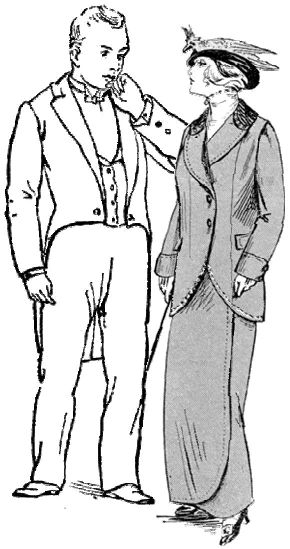
$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

```
= 2*(1-NORMDIST(ABS(z),0,1,TRUE))
```

	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**		

p-value = 0.24

Thank you for your attention



to be continued…