Introduction to statistical testing

Illustrated with XLSTAT

Jean Paul Maalouf
webinar@xlstat.com

Nov. 9, 2016

www.xlstat.com

Test family	Question	Data	Null Hypothesis	Example	Parametric tests	Conditions of validity (parametric tests)	Non-parametr equivalents
Compare locations*	Compare an observed mean to a theoretical one	Measurements on one sample and 1 theoretical mean (1 number)	Observed mean = theoretical mean	Compare an observed pollution rate to a standard	One-sample t-test	2	
	Compare two observed locations* (independent samples)	Measurements on two samples	Locations* are identical	Compare hemoglobin concentration between two groups of patients	t-test on two independent samples	1 ; 2 ; 3 ; 5	Mann-Whitney' test
	Test the equivalence between two samples	Measurements on two samples	Locations* are different	Check if the effect of medication A is the same as the effect of medication B on the concentration of a molecule in mice	Equivalence test (TOST)	1 ; 2 ; 3 ; 5	
	Compare several observed locations* (independent samples)	Measurements on several samples	Locations* are identical	Compare corn yields according to 4 different fertilizers	ANOVA	1 ; 3 ; 4 ; 6	Kruskal-Wallis test; Mood's tes

Link

Goal of this webinar

Let you become independent in using our web stat test selection tool

(whether you're an XLSTAT user or not)

PLAN

- **XLSTAT**: who are we ?
- Statistics: **categories**
- **Reminder** on Descriptive / exploratory statistics
- Statistical tests: **principles**, **steps** & **practice** on XLSTAT
- **Parametric** vs **non parametric** tests – practice on XLSTAT
- Tests on **independent** vs **paired** samples
- Statistical tests: **Comparison** vs **Association**
- Practice on XLSTAT: Fisher's exact test on a contingency table
- Appendix: How to interpret p-value > alfa?

All the data in this webinar were made up unless otherwise specified

XLSTAT

Your data analysis solution

XLSTAT: Who are we?

XLSTAT is a user-friendly statistical add-on software for Microsoft Excel®

XLSTAT

A growing software and team

1993
Thierry Fahmy develops a user-friendly solution for data analysis: XLSTAT is born

1996
XLSTAT realizes its first sale on the Internet

2000
The company Addinsoft is created

2006
New version, VBA interface, C++ computations, 7 languages

2009
New offers adapted to business needs

2015
New products, new website, growing and dynamic team

2016
XLSTAT 365
Cloud version of XLSTAT for Excel 365

XLSTAT Free
Free limited Edition

XLSTAT in a few numbers

XLSTAT

200+ statistical features

General or field-oriented solutions

16 employees

Always receptive to the needs of users

7 languages

50k users

Across the world. Companies, education, research

130k visits/month on the website

Easy tutorials available in 5 languages

400 downloads/day

Statistics: 4 categories

Statistics: 4 categories

XLSTAT

Recording

Recording

Nov. 30

Description

I want to **summarize small data sets** (1-3 variables) using **simple statistics** or **charts** (mean, standard deviation, boxplots...)

Exploration

I want to easily extract **information** from a **large data set without** necessarily **having a precise question** to answer. (PCA, AHC...)

Tests

I want to **accept** / **reject** a very precise **hypothesis** assuming error risks. (t tests, ANOVA, correlation tests, chi-square...)

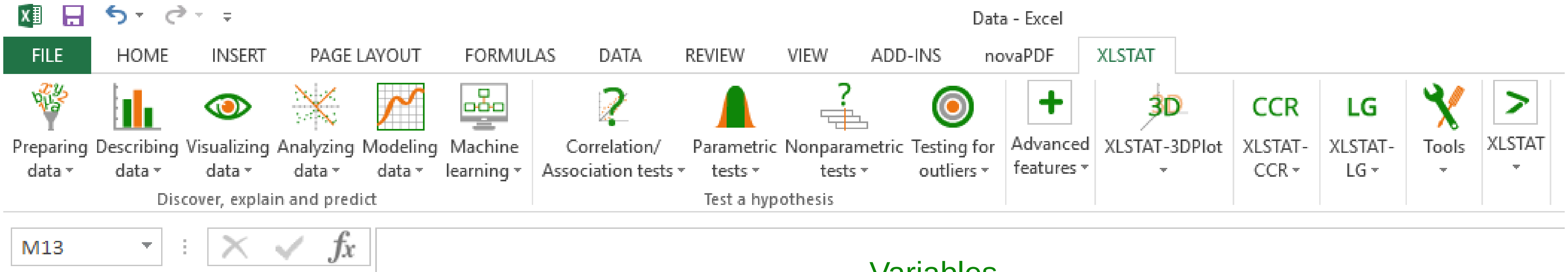
Modeling

I want to understand the way a phenomenon evolves according to a set of parameters. (regression, ANOVA, ANCOVA...)

Reminder on Descriptive / exploratory statistics

Data set: online shoe selling platform

XLSTAT

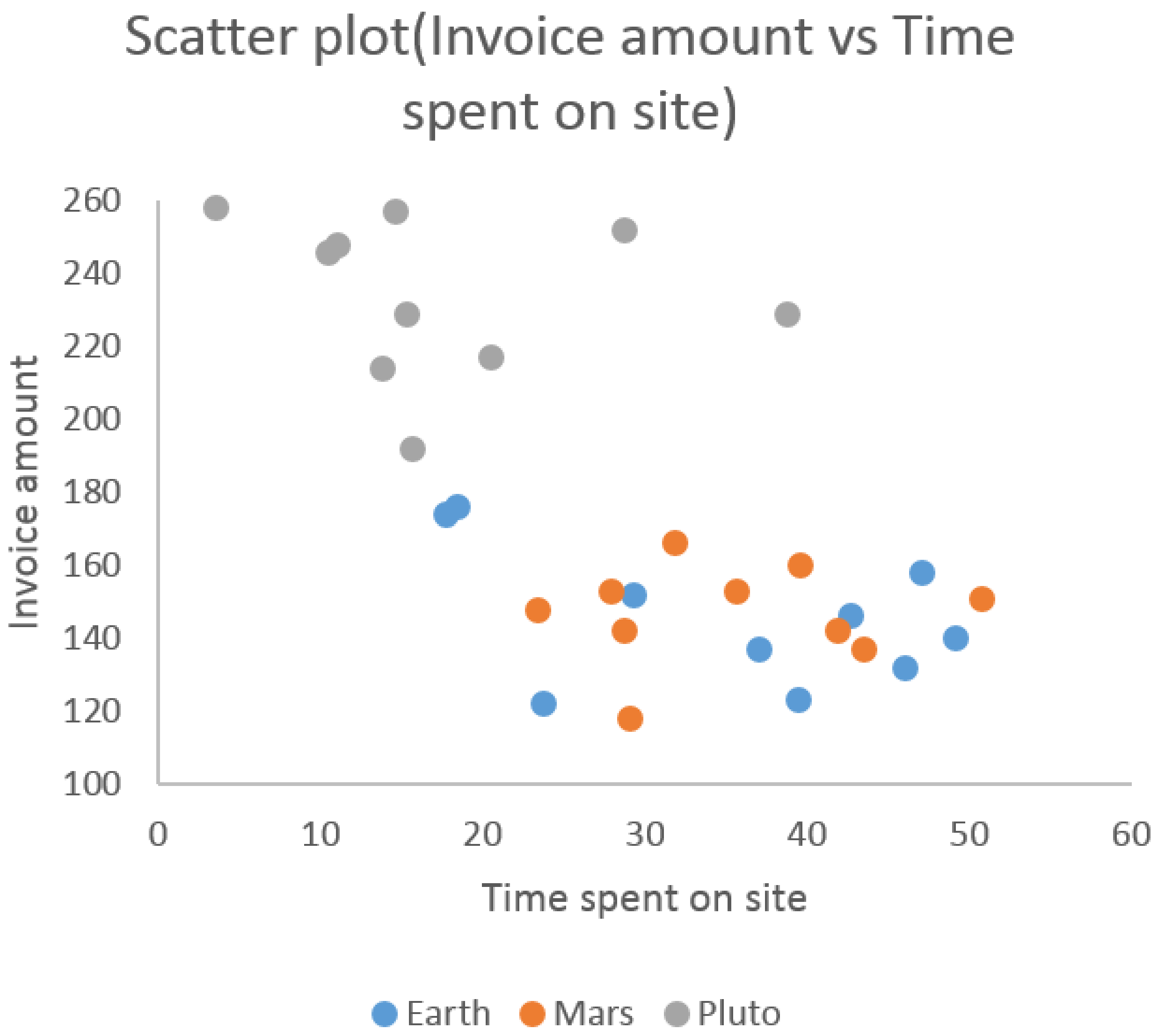


Variables

Individuals

	A	B	C	D	E	F	G	H	I
1	**Client**	**Origin**	**Gender**	**Height**	**Shoe size**	**Weight**	**Time spent on site**	**Preferred brand**	**Invoice amount**
2	Willard	Mars	M	9	26	50	28	Brand A	153
3	Derrick	Pluto	M	102	40	335	39	Brand D	229
4	Olivia	Mars	F	13	31	68	29	Brand A	118
5	Marc	Mars	M	20	29	81	40	Brand C	160
6	Alice	Pluto	F	116	33	378	11	Brand B	248
7	Elijah	Mars	M	22	30	88	44	Brand C	137
8	Clyde	Mars	M	27	29	105	42	Brand C	142
9	Angel	Mars	F	31	28	116	51	Brand A	151
10	Randy	Earth	M	4	29	36	18	Brand A	174

Toward exploratory data analysis: scatter plot colored by group



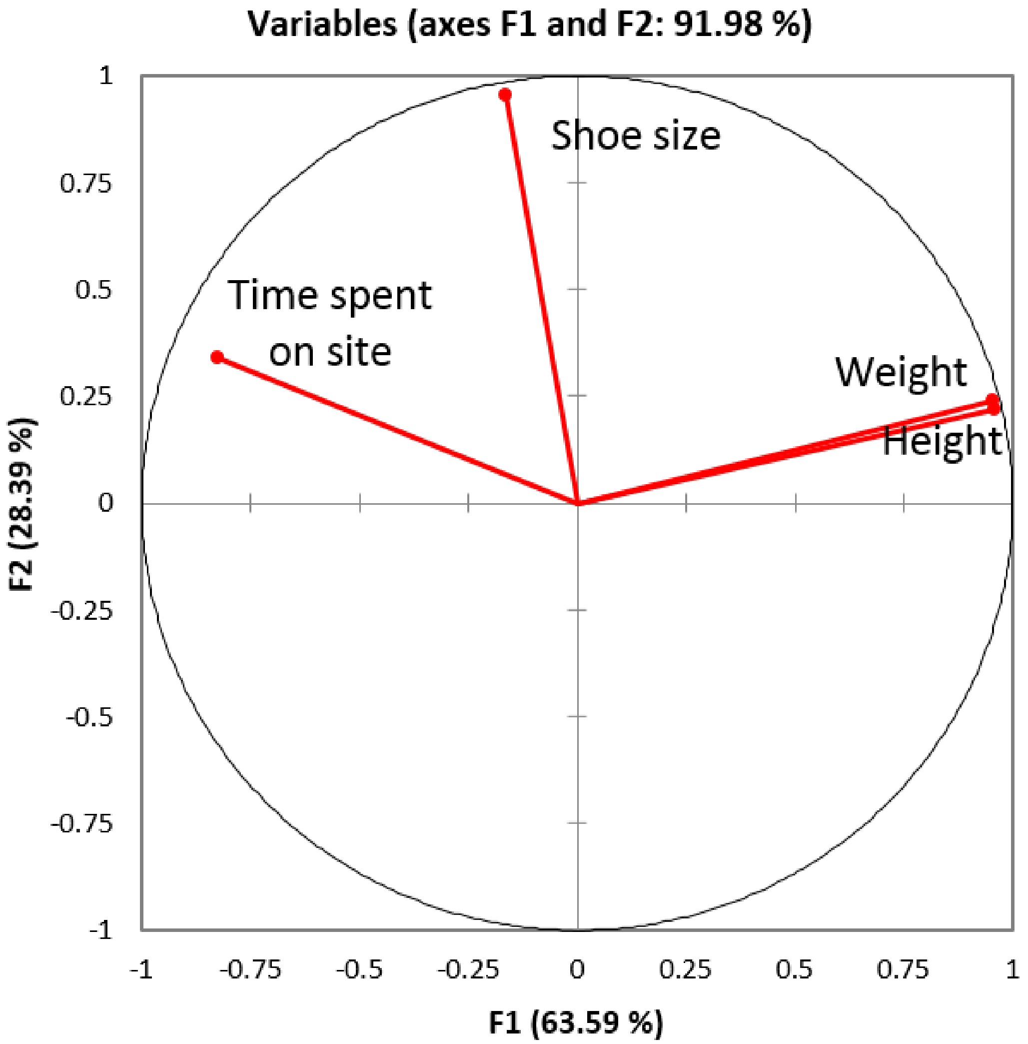
- Invoice amount decreases with time spent on the website.
- Plutonians spend more money on the website compared to others.
- Martians and humans form a relatively homogeneous group
- ...

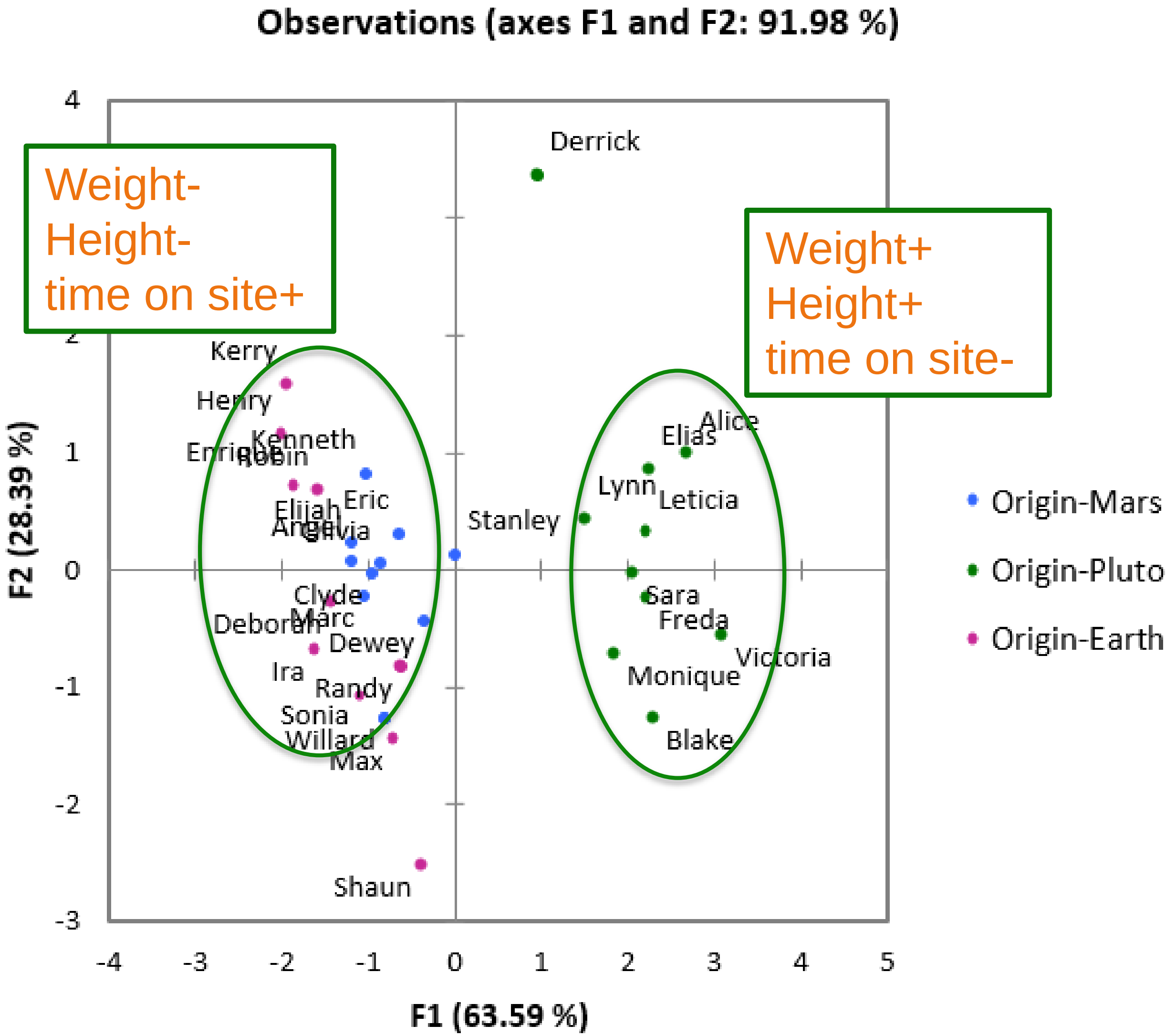
Webinar Recording

The same kind of reasoning on a higher number of variables... Exploratory statistics (or Exploratory Data Analysis)

Principal Component Analysis

Chart 1: correlation circle ; chart 2: observations





Webinar Recording

PCA: explorations ...

- Weight increases with height
- Time spent on site decreases with weight & height
- Looks like there are two clusters in the data
- Shoe size is unrelated to weight / height
- Derrick has big feet. Shaun has small feet.
- And so on...

Data exploration inspired us many hypotheses. Are they valid?

→ Statistical tests

Statistical tests

I want to **accept** / **reject** a very precise **hypothesis** assuming error risks.

Statistical tests usually answer yes/no questions

Statistical testing: steps

Writing up the question (answer: yes/no)

Writing up the null & the alternative hypotheses

Choosing the appropriate statistical test & the alfa risk threshold (check out the guide online)

Gathering the data

Things will be added here later

Running the test

Answering the question: if p-value < alfa, we reject H0 with a risk proportional to p-value of being wrong

Step 1: writing up the question

Question: do fertilizers A & B induce a difference in sugar rate in bananas?

H0

vs

Ha

Step 2: Writing up the null & the alternative hypotheses

Writing up hypotheses

?

Question

Do fertilizers A & B induce a difference in sugar rate in bananas?

H0

Null Hypothesis

Generally implies an idea of equality

H0: mean sugar rate in A-fertilized bananas = mean sugar rate in B-fertilized bananas

Ha

Alternative Hypothesis

Generally implies an idea of difference

Ha: mean sugar rate in A-fertilized bananas ≠ mean sugar rate in B-fertilized bananas

Statistical testing: steps – where are we?

XLSTAT

Writing up the question (answer: yes/no)

Writing up the null & the alternative hypotheses

Choosing the appropriate statistical test & the alfa risk threshold (check out the guide online)

Gathering the data

Things will be added here later

Running the test

Answering the question: if p-value < alfa, we reject H0 with a risk proportional to p-value of being wrong

Step 3a: choosing the appropriate statistical test

Are we comparing **means**?
If yes, how many?

Are we comparing **proportions**?
If yes, how many?

Are we comparing **variances**?
If yes, how many?

Are we testing **associations**?
...

In our case, we want to compare 2 means
→ Student's t-test for two independent samples

Compare two observed means	Measurements on two samples	means* are identical	Compare hemoglobin concentration	t-test on two independent samples

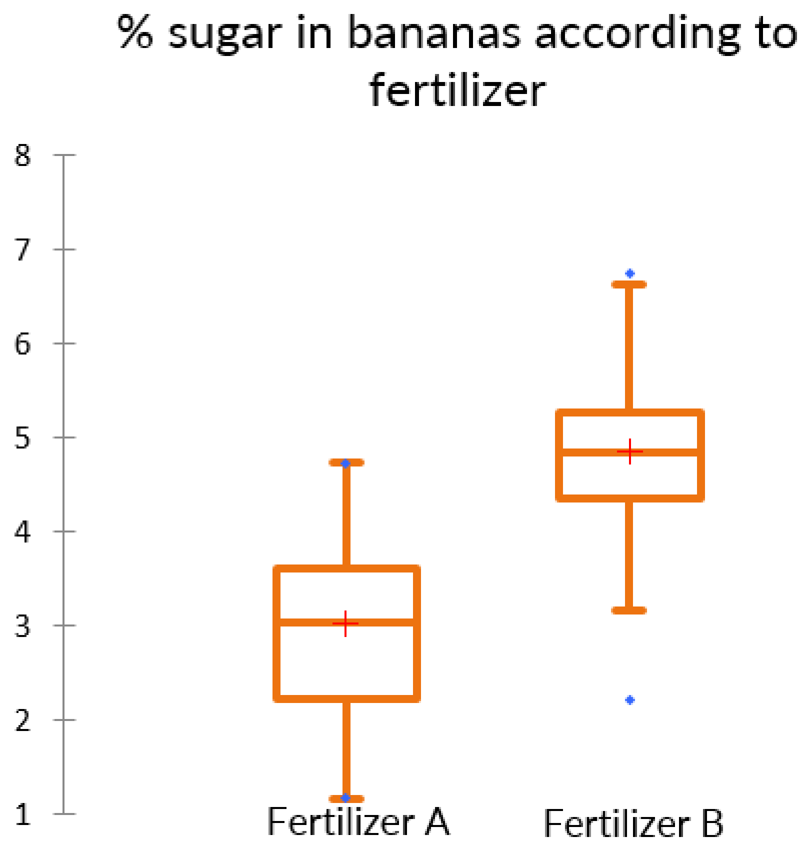
Link: choosing the appropriate statistical test according to your situation

Step 3b: choosing the alfa risk threshold

- The alfa risk threshold (0<alfa<1) is the threshold below which we decide to reject H0
- The more we want to limit the risk of taking a wrong decision, the more we should decrease alfa
- People often set alfa at 0.05. *This is not a reason to do it systematically*

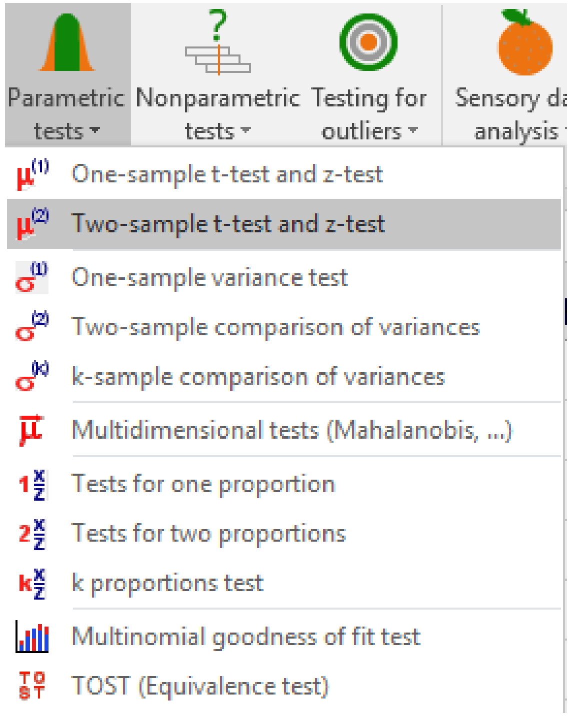
(but this is what we’ll do in our example ☺)

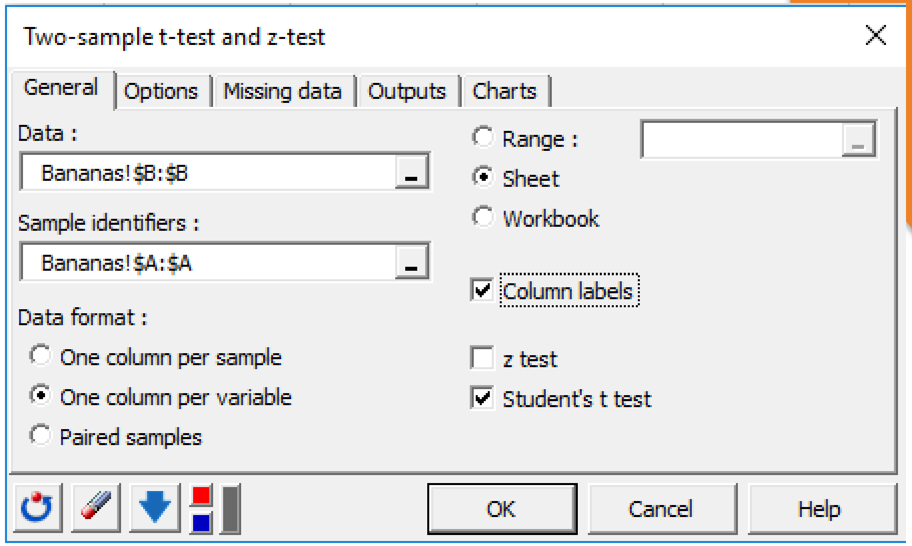
Experiment: 60 banana trees are planted; 30 of them receive fertilizer A, 30 of them receive fertilizer B



Step 4: gathering the data

Step 5: running the test in XLSTAT





p-value

vs

alfa

Step 6: interpreting the result and answering the question

Interpreting the result

XLSTAT

?

Question

Do fertilizers A & B induce a difference in sugar rate in bananas?

H0

Null Hypothesis

Generally implies an idea of equality

H0: mean sugar rate in A-fertilized bananas = mean sugar rate in B-fertilized bananas

Ha

Alternative Hypothesis

Generally implies an idea of difference

Ha: mean sugar rate in A-fertilized bananas ≠ mean sugar rate in B-fertilized bananas

The test computes a number called p-value.
$0 < \text{p-value} < 1$

The p-value is the risk you take of being wrong when rejecting H0 and accepting Ha

Decision *: If p-value < alfa, we reject H0 and accept Ha assuming a risk proportional to p-value of being wrong.*

Interpreting the result

Difference	-1.829
t (Observed value)	-7.385
|t| (Critical value)	2.002
DF	58
p-value (Two-tailed)	< 0.0001
alpha	0.05

Test interpretation:

H0: The difference between the means is equal to 0.

Ha: The difference between the means is different from 0.

As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.

The risk to reject the null hypothesis H0 while it is true is lower than 0.01%.

Decision*: p-value < alfa. We reject H0 & accept Ha with a very low risk of being wrong.*

Answer*: The two means (fertilizer A vs fertilizer B) are significantly different*

Parametric vs non parametric tests

Power

vs

Robustness

Parametric vs non parametric tests

Differences on the way they work

A statistical test can be either parametric or non parametric

- Parametric tests are **reliable** only under certain **conditions** that are linked to the distribution of populations. These conditions can be found on our online statistical testing guide.
- Non parametric tests **do not assume** any underlying **distribution**. Most of them are computed from the ranks of the data.

So why do we still use parametric tests?

Differences on their usefulness

- Non parametric tests: reliable in a larger number of situations than parametric tests → they are more **robust**.
- Parametric tests: more able to reject H0 if it is false, and if applicability conditions are respected → they are more **powerful***.

*Statistical power of a test is its ability to lead to a rejection of H0 if H0 is wrong

So, which type should you choose? Here's a proposition:

Choose an appropriate parametric test

Gather the data

Are assumptions for the parametric test met?

Yes

No

Replace with a non parametric test, less powerful but more robust

Run the test

Tests on independent vs paired samples

Tests on independent vs paired samples

Independent samples

Two or **more** distinct **populations**

Examples : compare a treated group and a control group; compare females and males; compare treated and untreated banana trees.

Paired samples

One single **population**

Examples : measuring the weight of patients before/after a treatment ; follow up companies or surveyed individuals at different dates ; follow photosynthetic capacities of the same banana trees at different dates/

Statistical tests: comparison vs association

Statistical tests: comparison & association

Comparison tests

- Comparing **means** (Student / ANOVA)
- Comparing **variances** (Fisher / Levene)
- Comparing **proportions** (tests on proportions)

Variables association tests

- Test the association between two **qualitative variables** (chi-square & exact Fisher's test)
- Test the association between two **quantitative variables** (Pearson & Spearman correlation coefficients)

Commonly used statistical tests

XLSTAT

Parametric tests and their non parametric equivalents

Question	Independent / paired samples	Parametric tests	Non parametric equivalents
Compare 2 means	Independent	Student's t-test on independent samples	Mann-Whitney's test
	Paired	Student's t-test on paired samples	Wilcoxon's test
Compare k means	Independent	ANOVA	Krukal-Wallis test
	Paired	Repeated measures ANOVA	Friedman's test
Compare 2 variances	Independent	Fisher's test	
Compare k variances		Levene's test	
Association (qualitative var.)	Independent	Chi2 test	Fisher's exact test
	Paired		Cochran's Q test
Association (quantitative var.)	Independent	Pearson correlation	Spearman correlation

Link: choose an appropriate test according to your situation

Association tests: Fisher’s exact test on two qualitative variables

Investigating the significance of a contingency table (= crosstab)

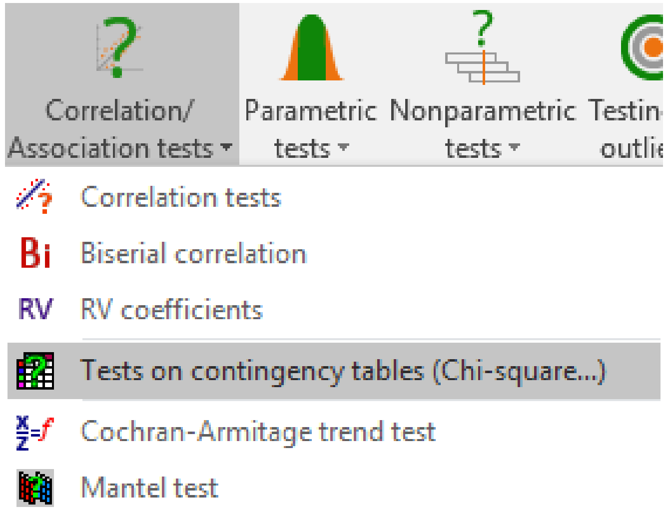
Application: association test (qualitative variables)

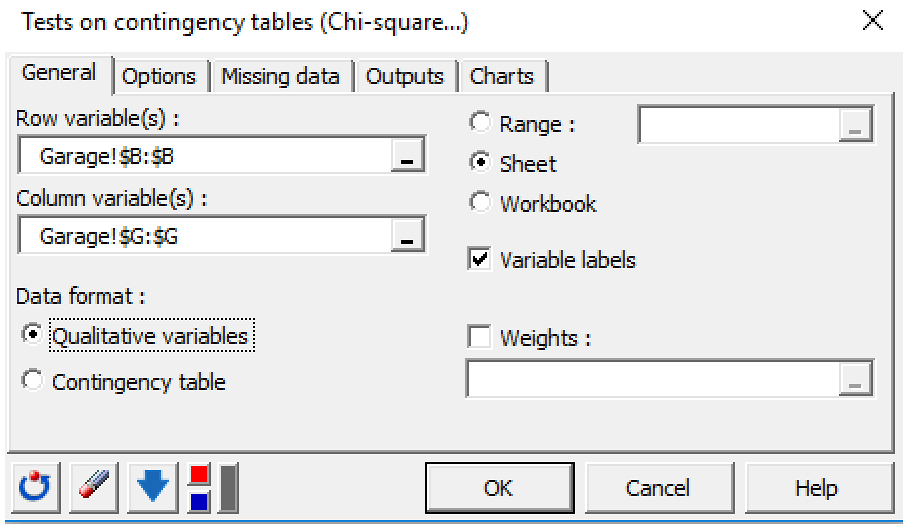
EXAMPLE: car garage, customer satisfaction survey

	A	B	C	D	E	F	G
1	**Client**	**Sociopro category**	**Satisfied**	**Repaired**	**Welcome**	**Q/Price**	**Come back**
2	C1	a	Yes	Yes	5	Yes	Yes
3	C2	b	Yes	Yes	4	Yes	dk
4	C3	a	Yes	Yes	4	Yes	dk
5	C4	b	Yes	dk	4	Yes	dk
6	C5	a	Yes	dk	4	Yes	Yes
7	C6	b	Yes	dk	4	Yes	Yes
8	C7	b	Yes	dk	5	Yes	No
9	C8	b	Yes	dk	3	Yes	No
10	C9	b	Yes	Yes	2	Yes	No
11	C10	a	Yes	Yes	5	No	dk
12	C11	a	Yes	Yes	4	No	dk
13	C12	a	Yes	Yes	3	No	Yes

Launching the test in XLSTAT

EXAMPLE: car garage, customer satisfaction survey





Association test example

EXAMPLE: car garage, customer satisfaction survey

Contingency table (Sociopro category / Come back):

	No	Yes	dk
a	5	3	6
b	6	1	7

Percentages / Column (Sociopro category / Come back):

	No	Yes	dk	Total
a	45	75	46	50
b	55	25	54	50
Total	100	100	100	100

Fisher's exact test:

p-value (Two-tailed)	0.721
alpha	0.05

p-value > alfa. We cannot reject H0.

H0: proportions of categories *a* & *b* **do not change** according to categories no-yes-dk
Ha: proportions of categories *a* & *b* **change** according to categories no-yes-dk

Statistical tests: revisiting the steps, a conclusion

Writing up the question (answer: yes/no)

Writing up the null & the alternative hypotheses

Choosing the appropriate statistical test (comparison / association) & the alfa risk threshold

Gathering the data

Are assumptions for the parametric test met?

Yes

No

Replacing with a non parametric test, less powerful but more robust

Running the test

Answering the question: if p-value < alfa, we reject H0 with a risk proportional to p-value of being wrong

Statistics: 4 categories

XLSTAT

Recording

Recording

Nov. 30

Description

I want to **summarize small data sets** (1-3 variables) using **simple statistics** or **charts** (mean, standard deviation, boxplots...)

Exploration

I want to easily extract **information** from a **large data set without** necessarily **having a precise question** to answer. (PCA, AHC...)

Tests

I want to **accept** / **reject** a very precise **hypothesis** assuming error risks. (t tests, ANOVA, correlation tests, chi-square...)

Modeling

I want to understand the way a phenomenon evolves according to a set of parameters. (regression, ANOVA, ANCOVA...)

Future webinars

Nov. 30, 2016: statistical modeling (click here)

Thanks for attending!

All the tools we saw are available in all XLSTAT solutions (except XLSTAT-Free)

Survey time…

Appendix: How to interpret p > alfa?

Appendix: How to interpret p > alfa?

If p-value < threshold (often 0.05), we **reject** H0 and accept Ha with a risk proportional to p-value of being wrong.

If p-value > threshold, there are two possibilities:

- If **Statistical power* is high** (>0.95)

We **accept H0** and **reject Ha** by taking another risk (Bêta = 1 - Power) of being wrong.

- If **Statistical power is low** (<0.95)

→ The **risk** of being wrong when **accepting H0** is too **high** (power is low)

→ The **risk** of being wrong when **rejecting H0** is too **high** (p-value is high)
We are **unable to take any decision**. Game over.

**(statistical power being the ability of an experiment/a test to make you reject H0 when it is false)*

Statistical power: how to increase it

Statistical power increases with:

- The number of measurements
- Measurement precision
- Size effect
- The alfa threshold
- The statistical test used