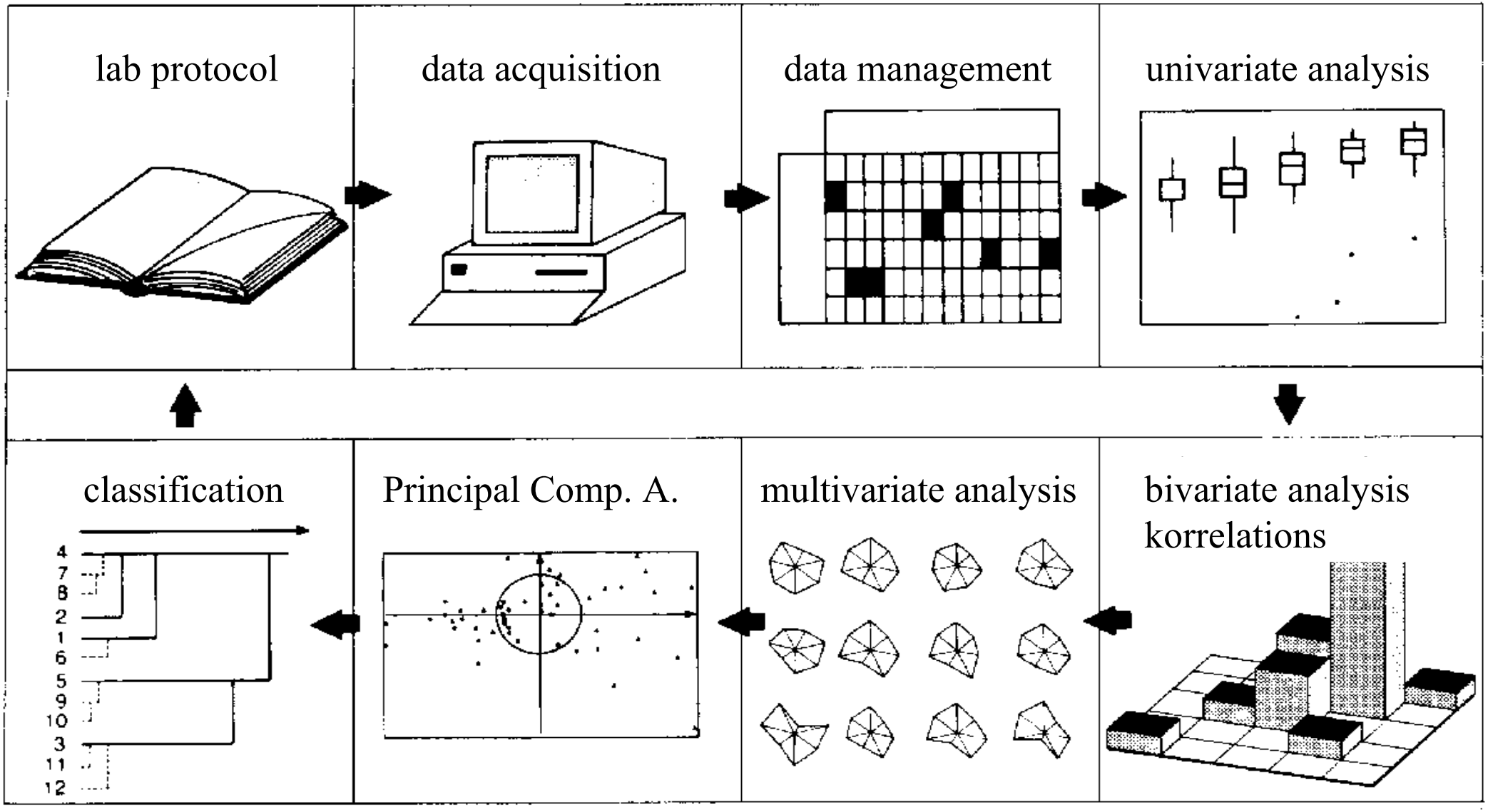
Evaluation of analytical Data - Aims and Purpose

- from measurements to conceptual models -



What is the purpose of STATISTICS?

…….to clarify and quantify:

1. **Make sure:** was the experimental design successful or does it need upscaling?

What kind of experiment – what kind of statistics?

Treatment, n Factors? → Differences? → **ANOVA**

T- test, post hoc tests,...

Observation, n Variables? → Relationship? / Clustering? → **ORDINATON**

Corr.A, PCA, Cluster A, NMDS, Discr.A, ...

2. **Explore**: what are your main results, what sound conclusions may be drawn? how relevant are they?

Results substancial, how big? → % Δ, $\%R^{2,}$,% Variance explained

Results reliable, how reproducible? → p value / % sig.Level

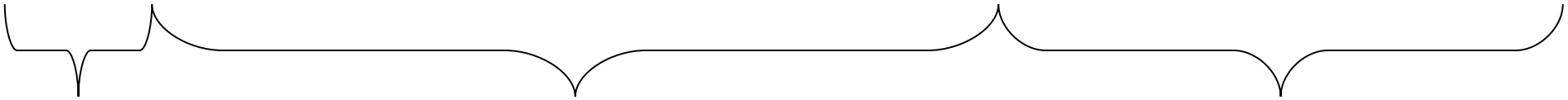
3.**Verify**: answer *and* quantify your **original research questions**

3.a Which presentation graph is best?

4. What are the key parameters for a mathematical **model**?

Data Acquisition: Harvest Procedure → Data Matrix

N	Sample	Treatment	Species	Plant/Branch	Replicate	Treatment Species Var1	VarN
1	Cros 1 1	C	ros	1	1	C ros	
2	Cros 1 2	C	ros	1	2	C ros	
3	Cros 1 3	C	ros	1	3	C ros	
4	Cros 2 1	C	ros	2	1	C ros	
5	Cros 2 2	C	ros	2	2	C ros	
6	Cros 2 3	C	ros	2	3	C ros	
7	Cflo 1 1	C	flo	1	1	C flo	
8	Cflo 1 2	C	flo	1	2	C flo	
9	Cflo 1 3	C	flo	1	3	C flo	
10	Cflo 2 1	C	flo	2	1	C flo	
11	Cflo 2 2	C	flo	2	2	C flo	
12	Cflo 2 3	C	flo	2	3	C flo	
13	Cmin 1 1	C	min	1	1	C min	
14	Cmin 1 2	C	min	1	2	C min	
15	Cmin 1 3	C	min	1	3	C min	
16	Cmin 2 1	C	min	2	1	C min	
17	Cmin 2 2	C	min	2	2	C min	
18	Cmin 2 3	C	min	2	3	C min	
19	Cmax 1 1	C	max	1	1	C max	
20	Cmax 1 2	C	max	1	2	C max	
21	Cmax 1 3	C	max	1	3	C max	



Single samples
(cases/observations)

Sampling data:
group variables
(factors/level code)

Measuring parameter:
measuring (dependent) variable
(variables)

Graphical Plotting, Statistics

first:

simple line or barcharts (never waste time and efforts for hasty formatting!)

Excel: MEDIAN and ERROR INDICATOR:

Standard deviation STDEV: $\sqrt{\frac{nSx^2-\left(Sx\right)^2}{n\left(n-1\right)}}$

Standard error (error of the mean) SE: $s_x=\frac{s}{\sqrt{n}}$

or Confidence Interval CI: $x\pm t\cdot s_x$ (**t** n=1000...1.96)

for table only: Coefficient of Variance CV: $\frac{s}{x}100$

next:

define the aim of your data evaluation

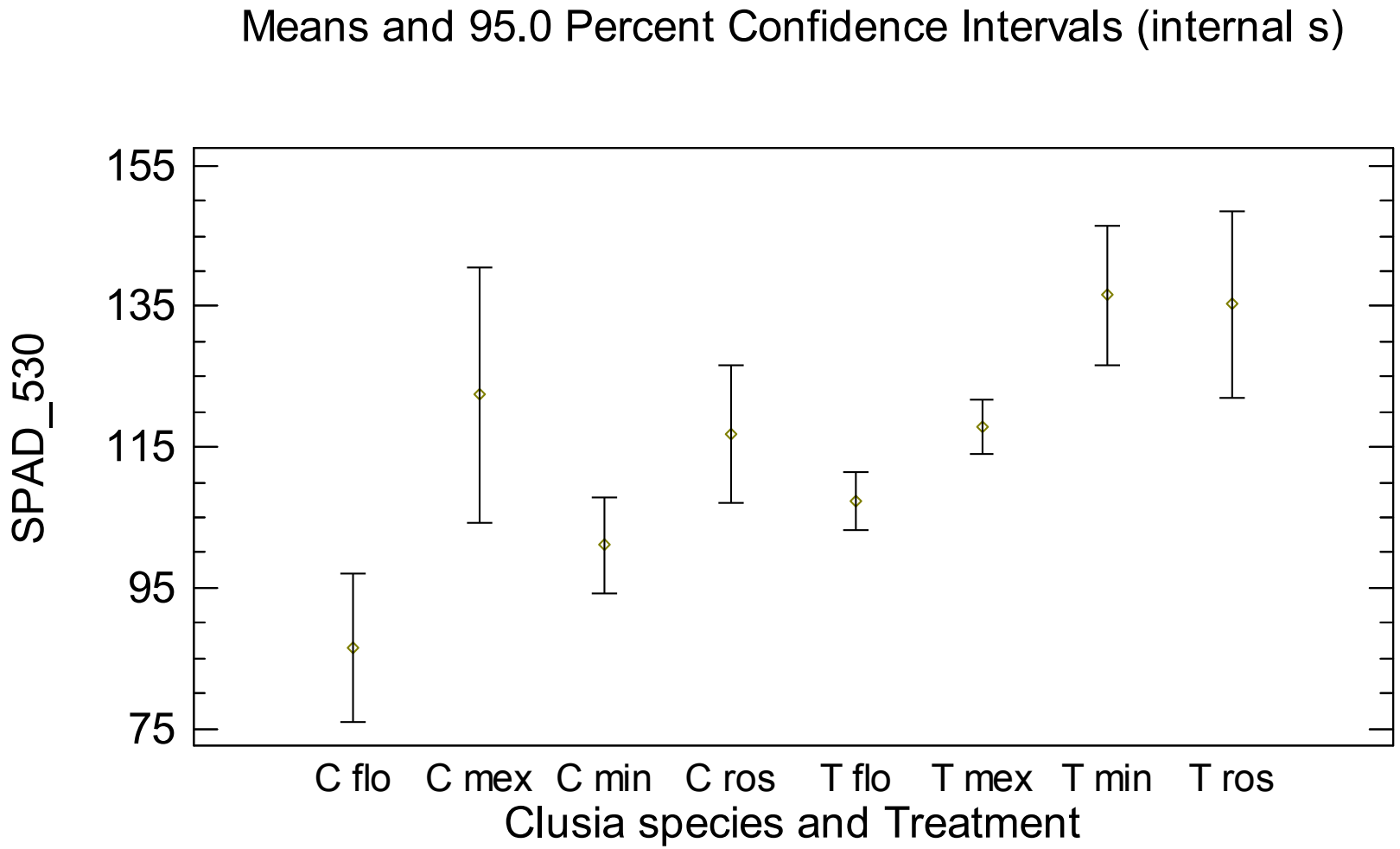
Differences (sig) Similarities temporal Sequence

define the need for statistics

sign.Differences, Korrelation (Ordination) or Classification?

define the aim of your presentation (Poster/Publikation)

Choice of suitable Software -->Formating of the plots



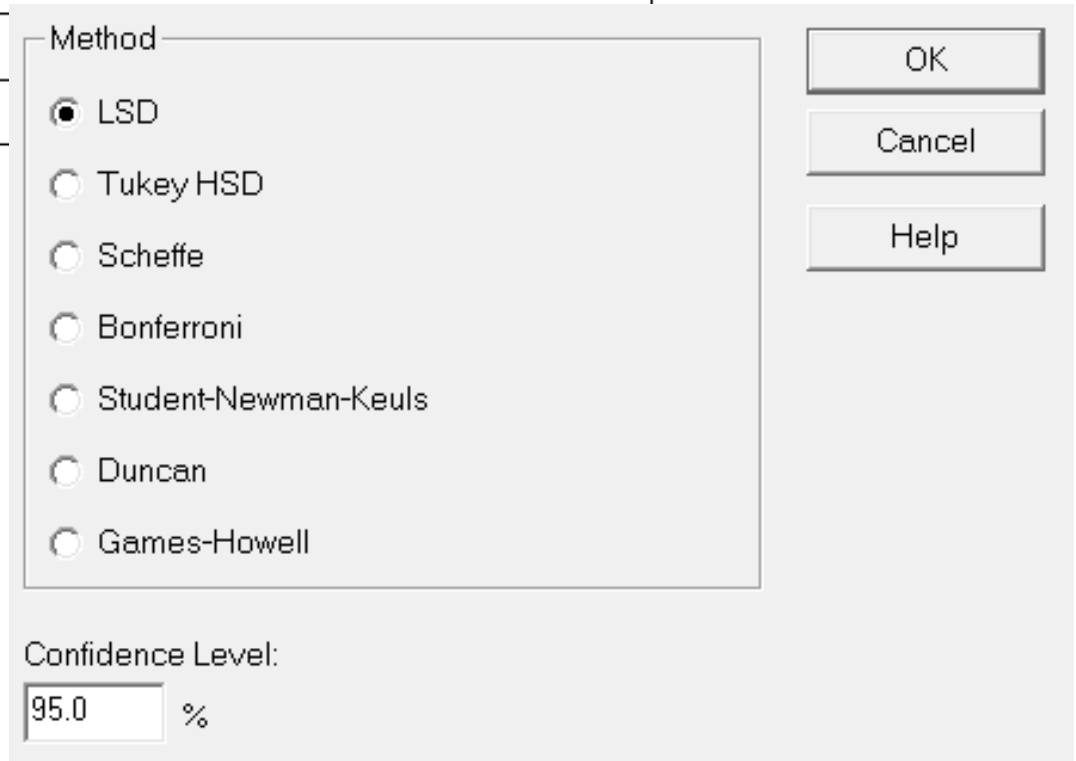
Students T-test: Δ average (1,2) < CI1+ CI2

http://en.wikipedia.org/wiki/Post-hoc_analysis

Multiple Range Tests for SPAD_530 by Clusia species and Treatment

Method: 95.0 percent LSD

Level	*Count*	*Mean*	*Homogeneous Groups*
C flo	6	86.4533	X
C min	6	101.035	X
T flo	6	107.3	XX
C ros	6	116.84	XX
T mex	11	117.873	X
C mex	6	122.383	X
T ros	7	135.27	X
T min	6	136.532	X



The Significance Level,
- the reproducibility of an experiment -

significant differences (univariate Tests, ANOVA -Analysis Of Variance):

Significance level: P + CI = 100 %
P.... Error probability
CI ...Confidence Interval

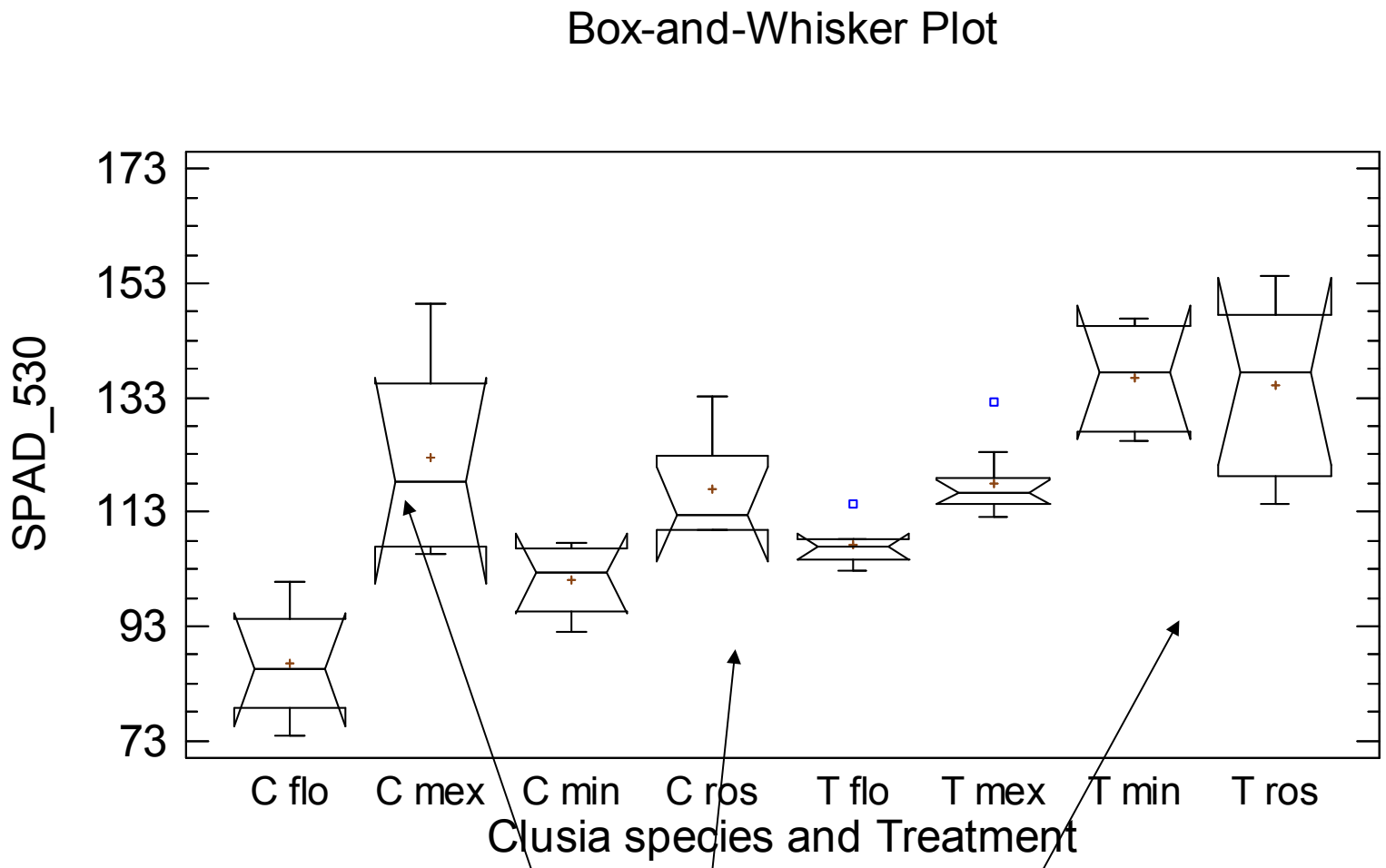
highly significant	P < 0,1% (0,001)	***
significant	P < 1% (0,01)	**
weakly significant	P < 5% (0,05)	*
not(?) significant	P > 5%	n.s.

In any case: do provide the actual significance level !
It is not only of interst ***if***, but first of all ***how*** signifikant your results really are.

4 Stages of Data Evaluation

1) evaluation of the experimental setup:
   - *did it work out (sufficient replicates)?*
2) explorative statistics:
   - *what (else) can be found in the data?*
3) quantification of findings:
   - *how relevant (robust, big, significant)?*
4) conceptual models:
   - *intercorrelation patterns?*
   - *system properties, system comparison?*

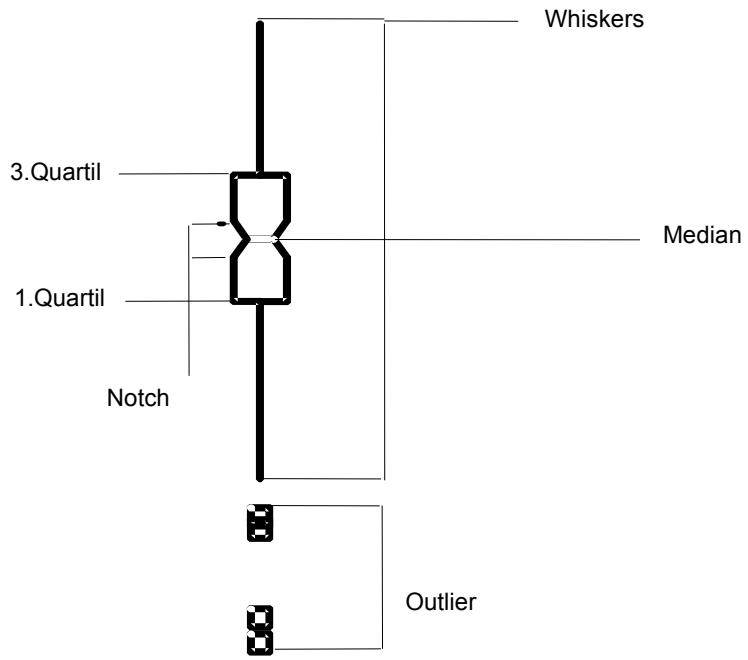
Explorative plots and Methods



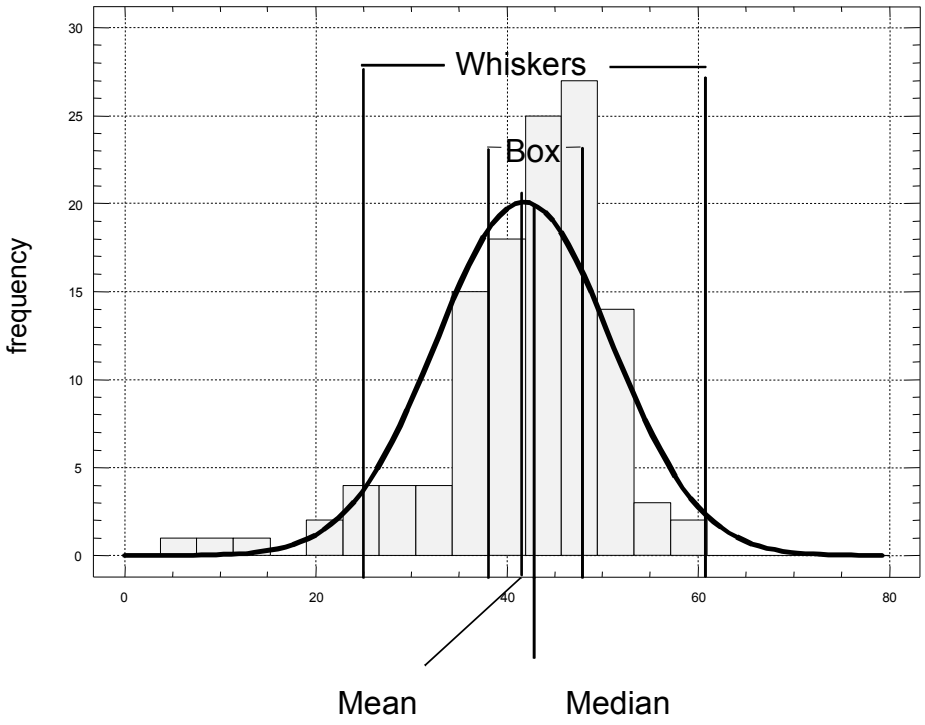
„notched box and whisker plot“

Explanation of notched Box and Whisker Plots

Notched Box and Whisker Plot

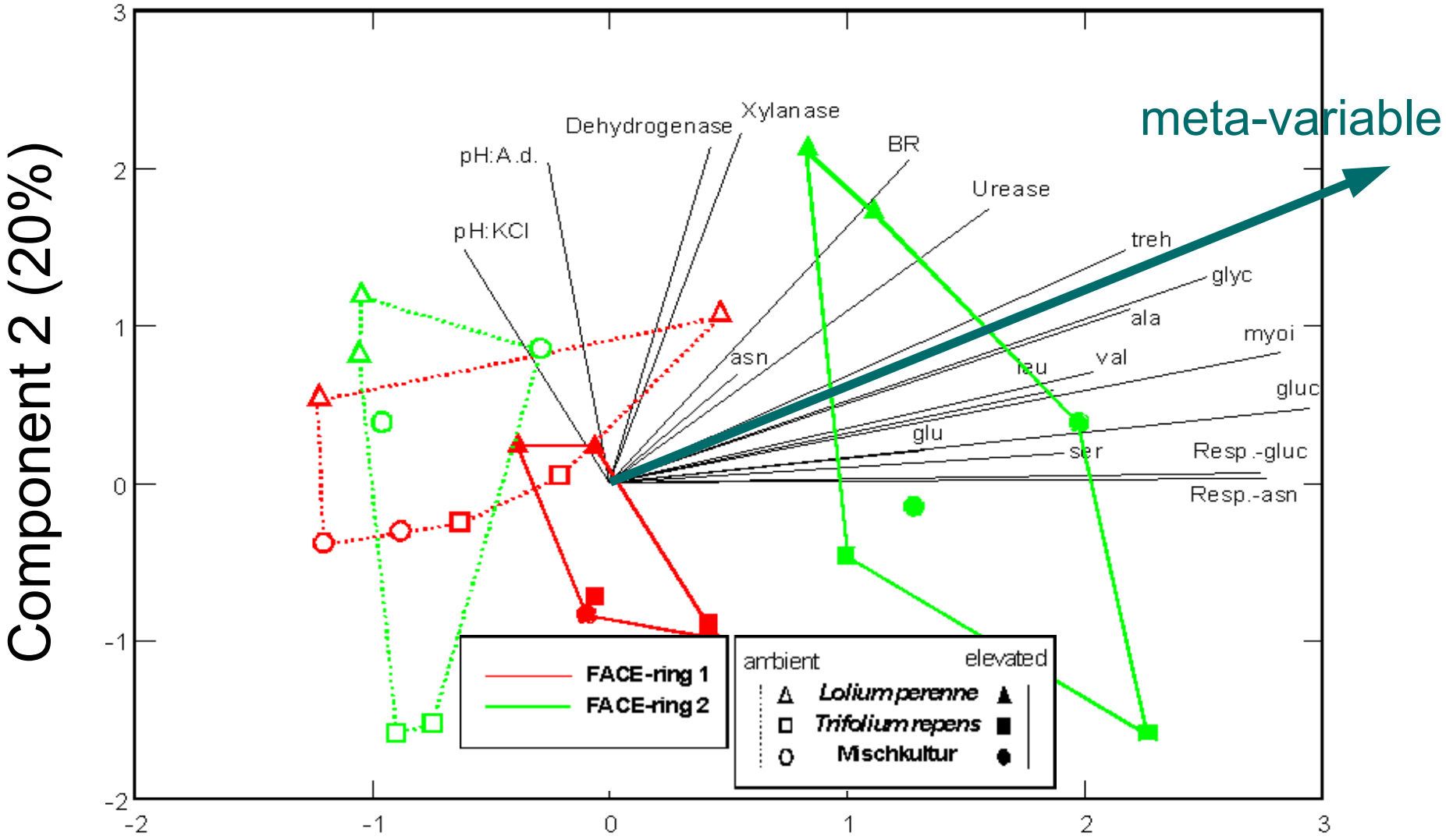


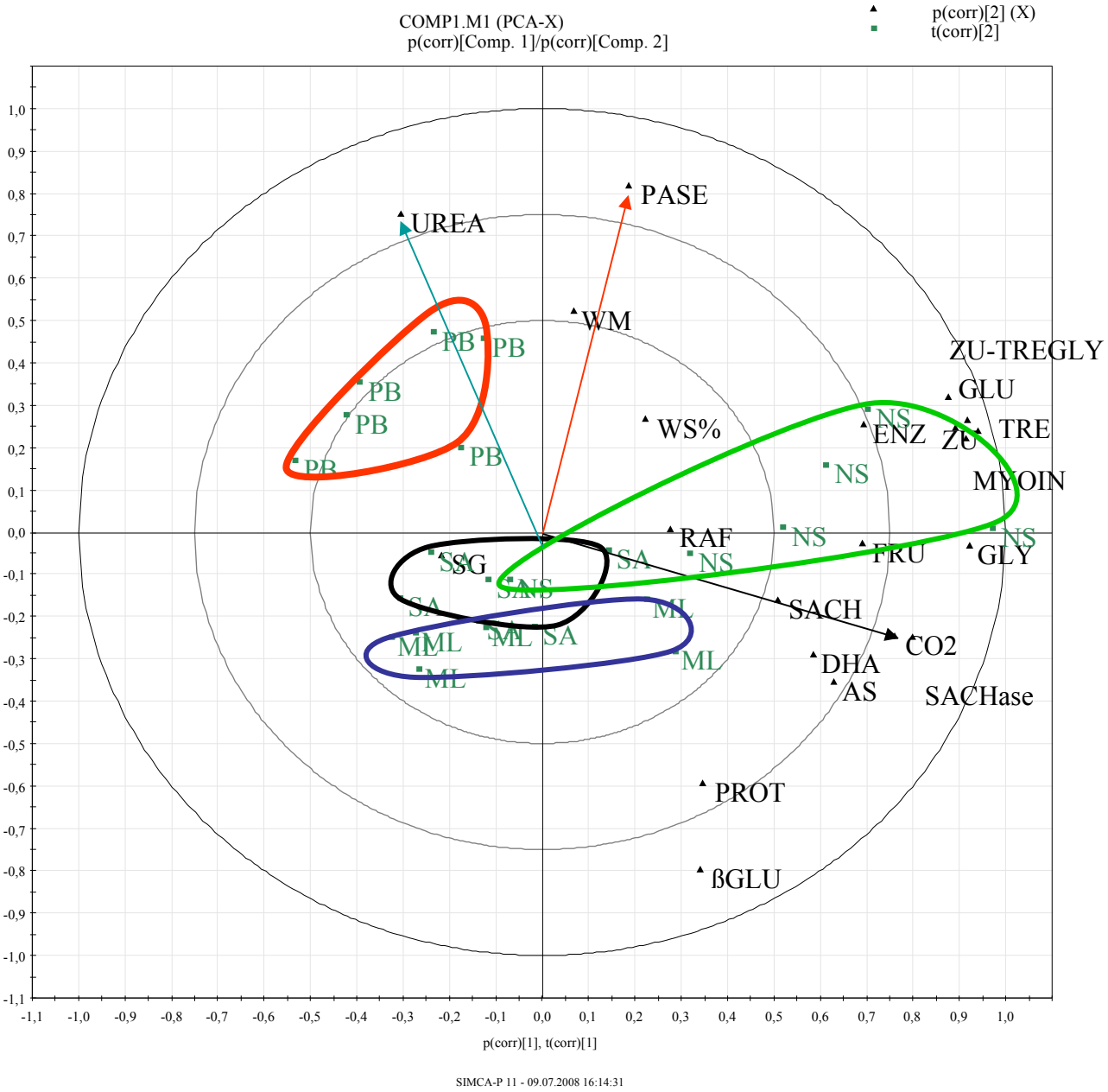
Frequency Histogram



Notch: indicates 95% confidence interval for the Median.
If Notches do not overlap, a significant difference between medians is present

Main Component or Factor analysis



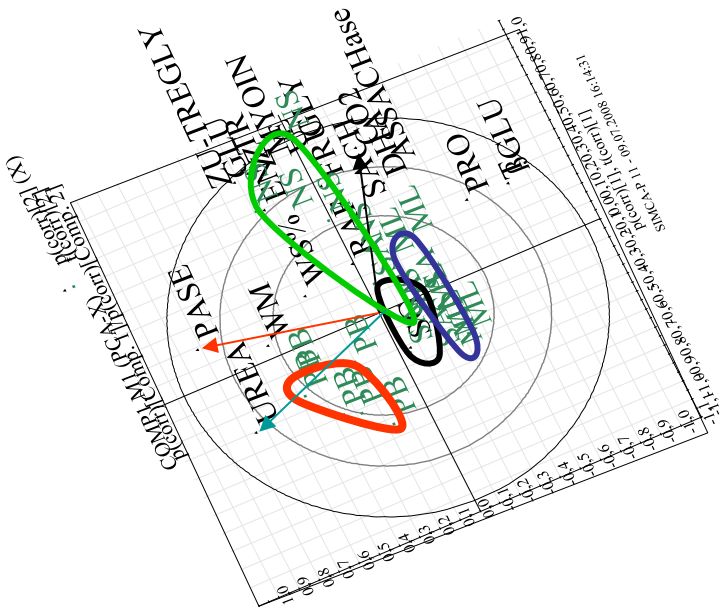


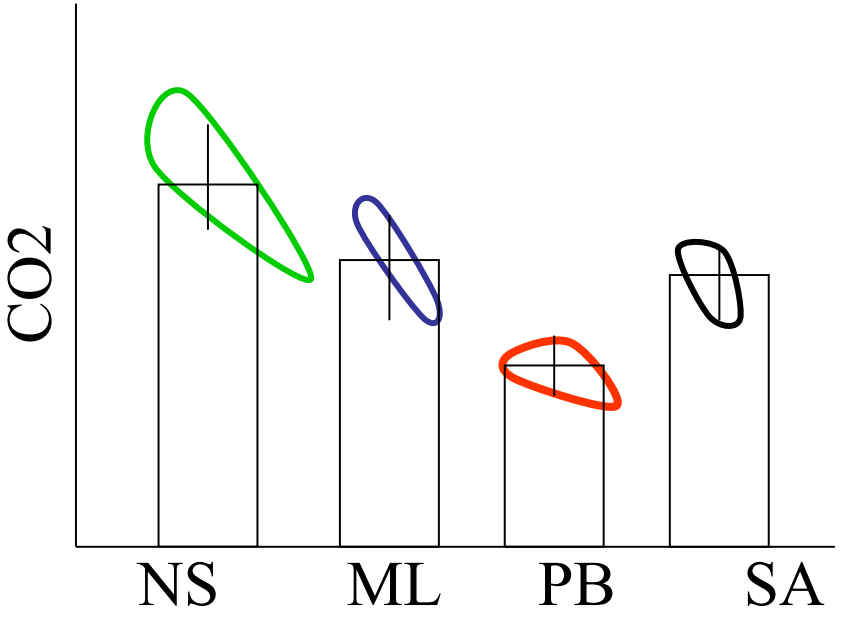
Simca-P
Analysis Loadings Bi-plot
right click, copy
„enh. Metafile“

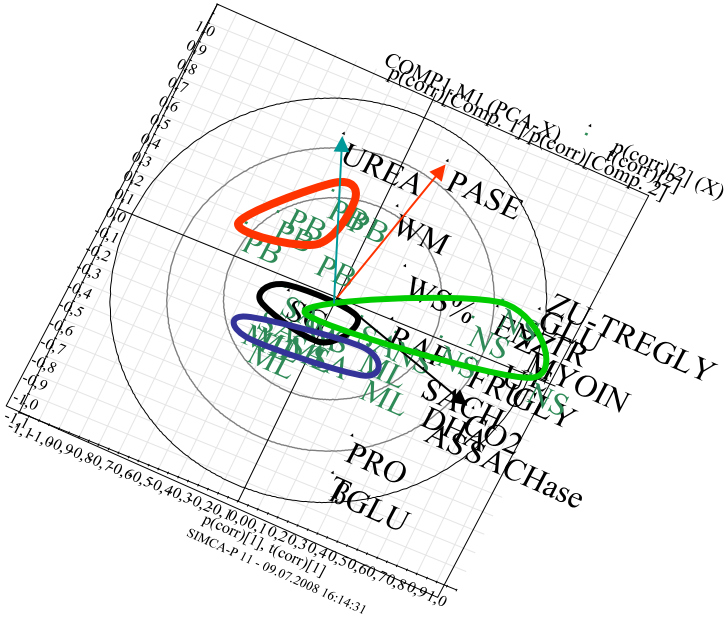
Powerpoint / Impress
reduce size
ungroup
set fonts to 12p
group

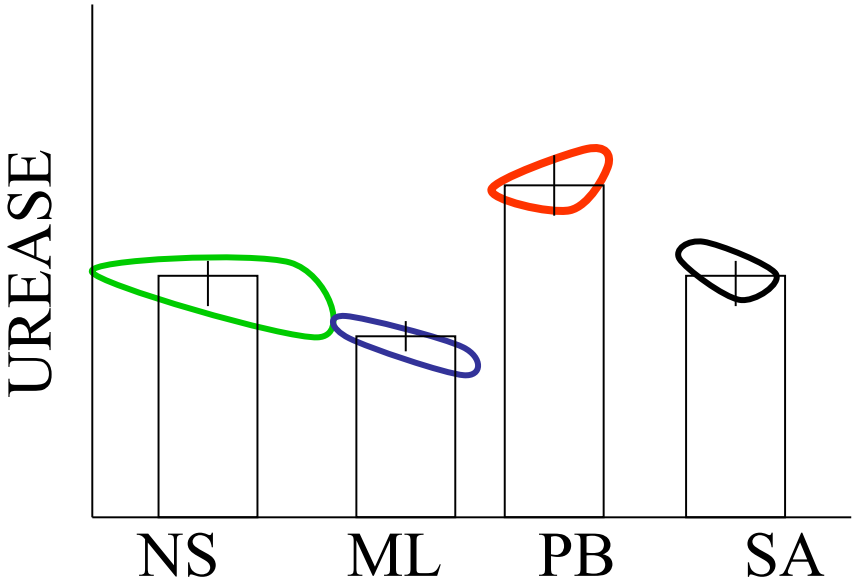
add main vectors
surround cluster by bezier lines

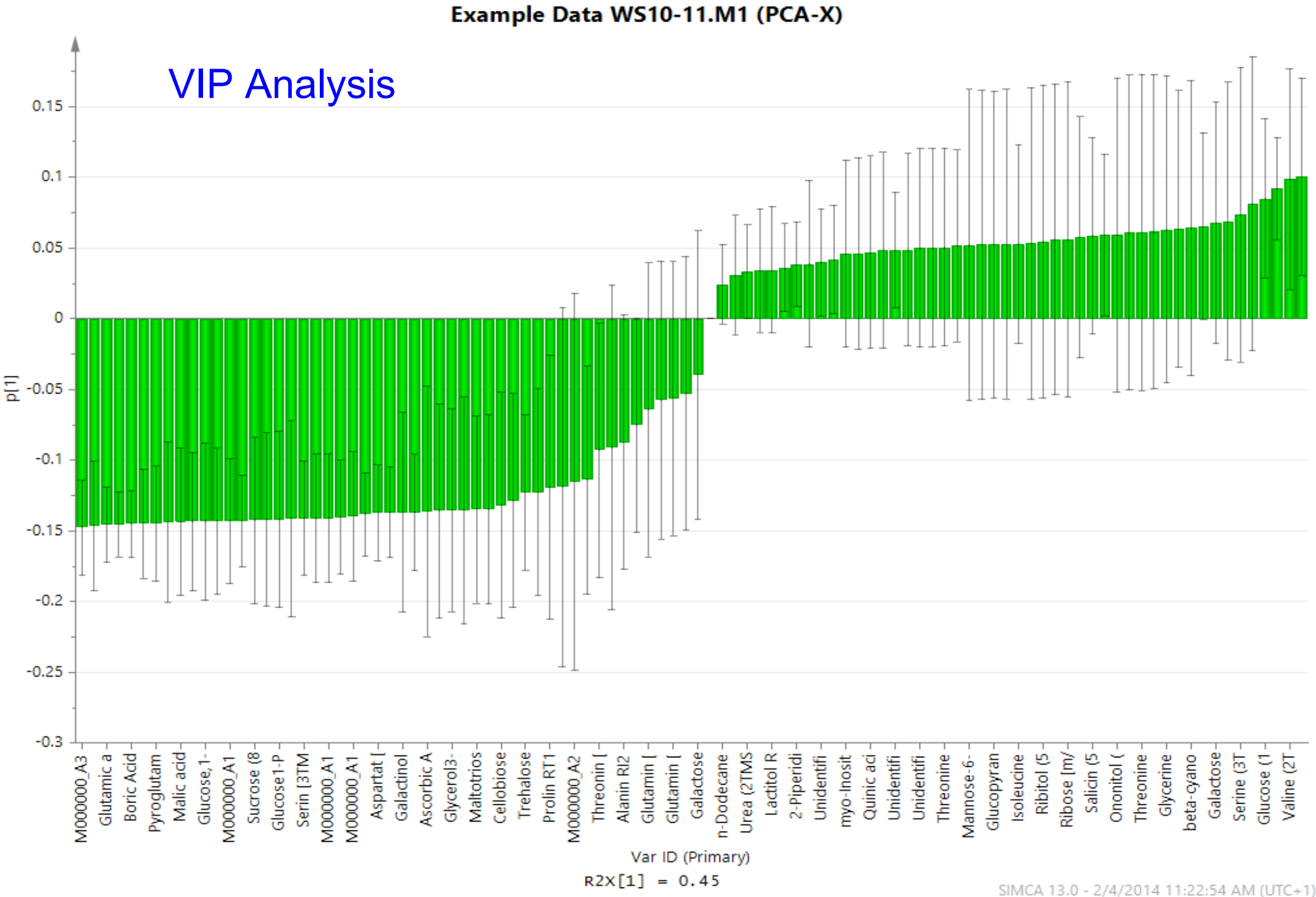
Find the bargraphs (Anova plots) in PCA…

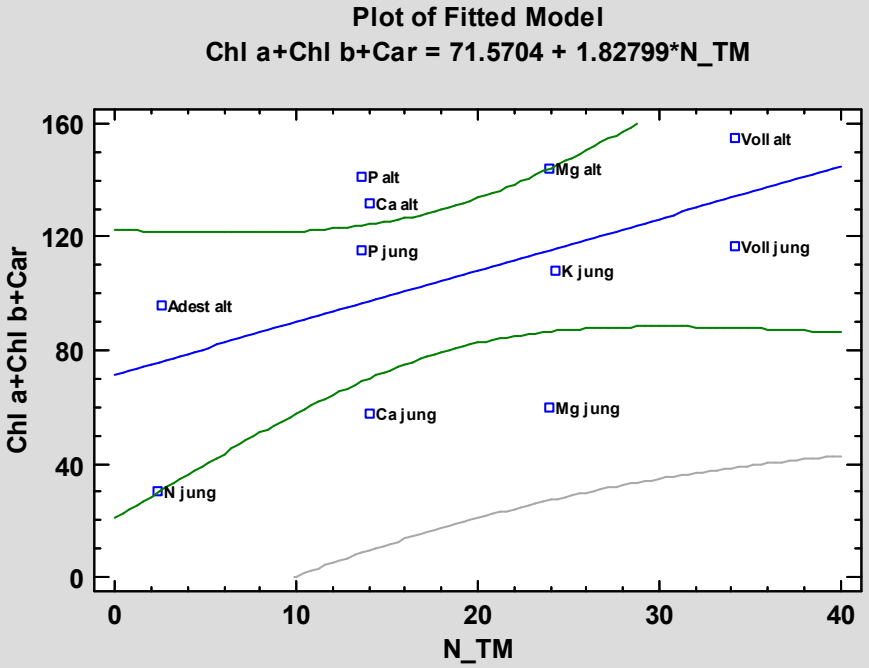








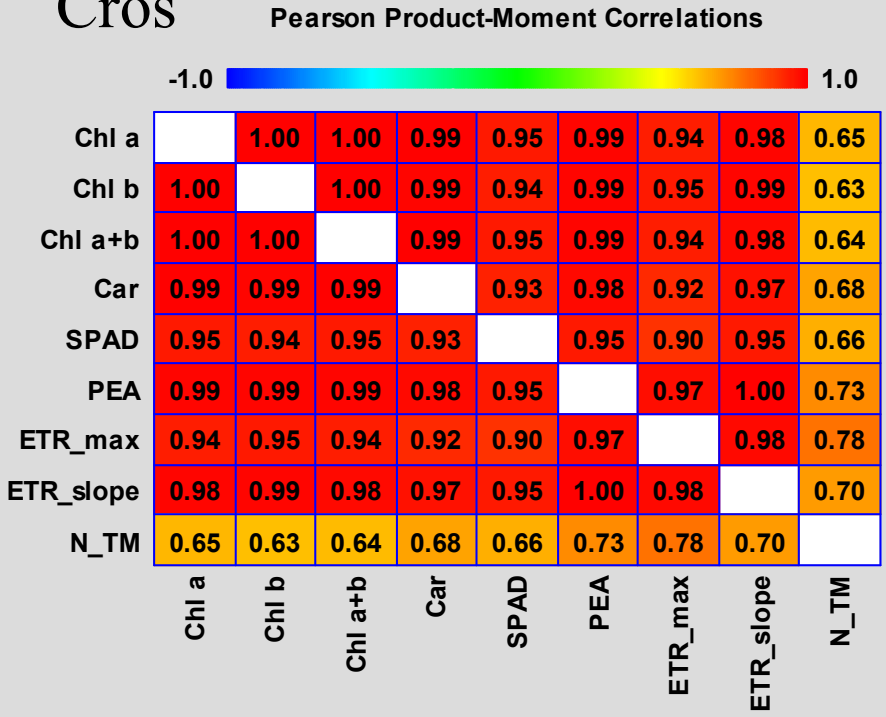

Example Data WS10-11.M1 (PCA-X)
VIP Analysis
p[1]
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
-0.3
M000000_A3
Glutamic a
Boric Acid
Pyroglutam
Malic acid
Glucose,1-
M000000_A1
Sucrose (8
Glucose1-P
Serin [3TM
M000000_A1
M000000_A1
Aspartat [
Galactinol
Ascorbic A
Glycerol3-
Maltotrios
Cellobiose
Trehalose
Prolin RT1
M000000_A2
Threonin [
Alanin RI2
Glutamin [
Glutamin [
Galactose
n-Dodecane
Urea (2TMS
Lactitol R
2-Piperidi
Unidentifi
myo-Inosit
Quinic aci
Unidentifi
Unidentifi
Threonine
Mannose-6-
Glucopyran
Isoleucine
Ribitol (5
Ribose [m/
Salicin (5
Ononitol (
Threonine
Glycerine
beta-cyano
Galactose
Serine (3T
Glucose (1
Valine (2T
Var ID (Primary)
R2X[1] = 0.45
SIMCA 13.0 - 2/4/2014 11:22:54 AM (UTC+1)



Intercorrelation Patterns Heat-Map

Correlation Coefficient = 0.494171
R-squared = 24.4205 percent
R-squared (adjusted for d.f.) = 16.0227 percent
Standard Error of Est. = 36.9063
Mean absolute error = 30.264

Cros



Tros