

Simulated Power Curves in ASCA

Topics in Chemometrics

2023



JOSÉ CAMACHO
JOSECAMACHO@UGR.ES
COMPUTATIONAL DATA SCIENCE LAB

Project *PID2020-113462RB-I00*
funded by:

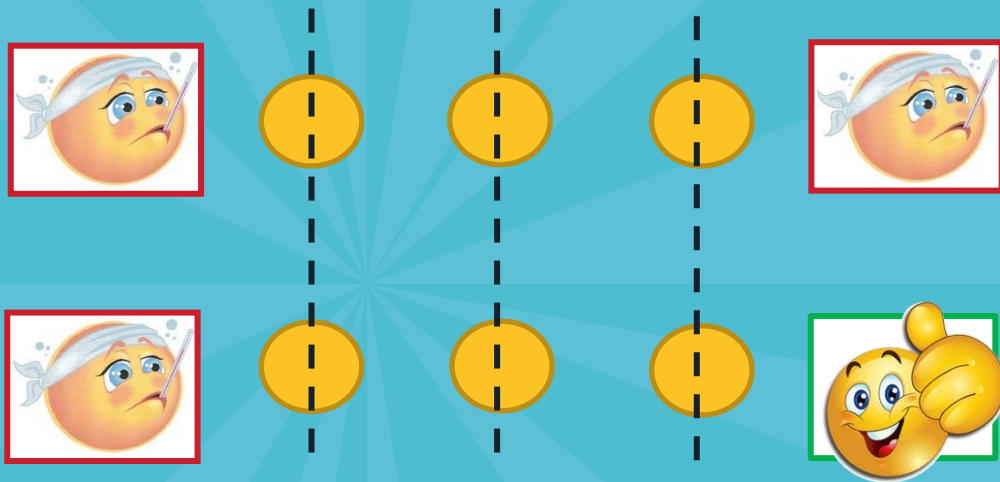


MINISTERIO
DE CIENCIA
E INNOVACIÓN

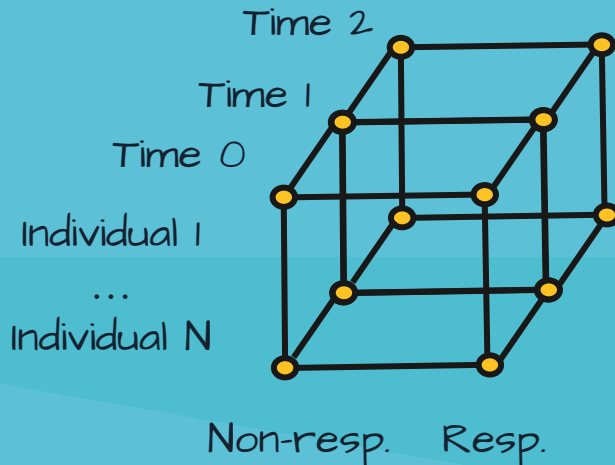


AGENCIA
ESTATAL DE
INVESTIGACIÓN

Context : Multi-way (multi-factor) clinical trial



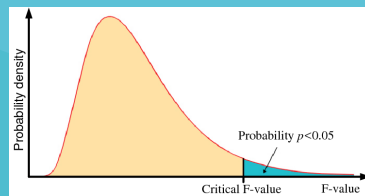
Baseline Treatment Post-T



Biomarker discovery: Statistical inference

$$X = 1m + Xp + Xt + Xi + Xpt + E$$

$$F = \frac{SS_P}{SS_E}$$



So many possibilities...

- Clinical trial ...

... longitudinal, cross-over, multi-center, sequential, observational, experimental, randomized, prospective, retrospective ...

- Factors ...

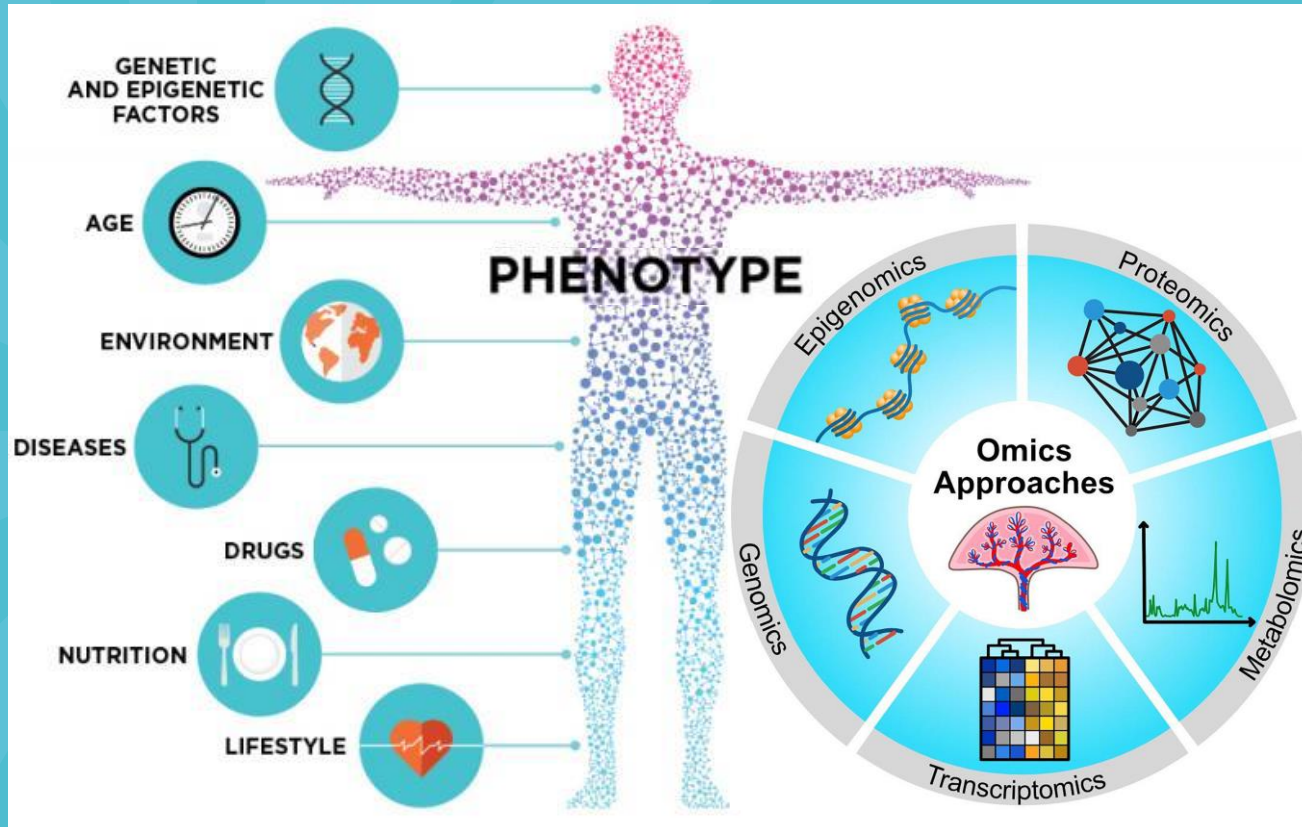
... fixed, random

- Relationships among factors ...

... crossed, nested, interactions

- Complexities ...

... unbalanced, distributions, outliers, missing data ...



Jaremek, Adam, et al. *Frontiers in Cell and Developmental Biology* 9 (2021): 674162.

ANOVA with MULTIVARIATE RESPONSE

$$X - \mathbf{1} \cdot m = b_A F_A + b_B F_B + b_{AB} F_{AB} + E$$

<table style="width: 100%; border-collapse: collapse;"> <tr><td>1.3</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>2.4</td></tr> <tr><td>2.0</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr><td>2.1</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr><td>3.3</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr><td>4.7</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr><td>2.8</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr><td>1.3</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr><td>1.0</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr><td>0.9</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr><td>3.6</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr><td>4.1</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr><td>0.9</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>3.3</td></tr> </table>	1.3	2.4	2.0	2.1	3.3	4.7	2.8	1.3	1.0	0.9	3.6	4.1	0.9	3.3	$- \mathbf{1}m =$	b_A	$+$	b_B	$+$	b_{AB}	$+$	E
1.3	2.4																																																																																						
2.0																																																																																						
2.1																																																																																						
3.3																																																																																						
4.7																																																																																						
2.8																																																																																						
1.3																																																																																						
1.0																																																																																						
0.9																																																																																						
3.6																																																																																						
4.1																																																																																						
0.9	3.3																																																																																						

1	$+$	1	$+$	1	$+$	E
1	$+$	1	$+$	1	$+$	E
1	$+$	1	$+$	1	$+$	E
1	$+$	-1	$+$	-1	$+$	E
1	$+$	-1	$+$	-1	$+$	E
1	$+$	-1	$+$	-1	$+$	E
-1	$+$	1	$+$	-1	$+$	E
-1	$+$	1	$+$	-1	$+$	E
-1	$+$	-1	$+$	1	$+$	E
-1	$+$	-1	$+$	1	$+$	E
-1	$+$	-1	$+$	1	$+$	E

ANOVA with MULTIVARIATE RESPONSE

$$X - \mathbf{1} \cdot m = b_A F_A + b_B F_B + b_{AB} F_{AB} + E$$

General Linear Model

$$b = \operatorname{argmin} \|X - m - bF\|_2$$

$$b = (F' * F)^{-1} F' \cdot (X - m)$$



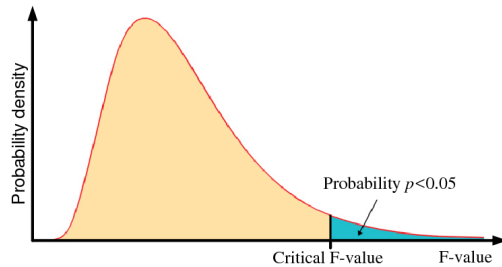
$$X - \mathbf{1} \cdot m = X_A + X_B + X_{AB} + E$$

MULTIVARIATE EXTENSIONS TO ANOVA

$$\|X - \mathbf{1} \cdot m\|_F^2 = \|X_p\|_F^2 + \|X_t\|_F^2 + \|X_i\|_F^2 + \|X_{pt}\|_F^2 + \|E\|_F^2$$

\downarrow \downarrow \downarrow \downarrow \downarrow

SS_X SS_p SS_t SS_i SS_{pt} SS_E

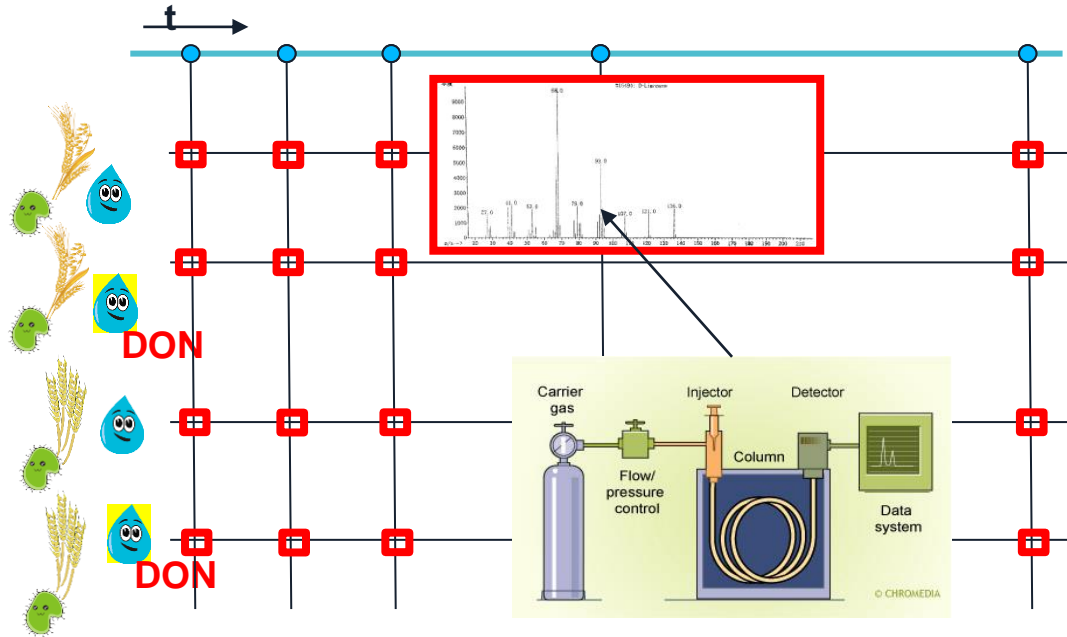
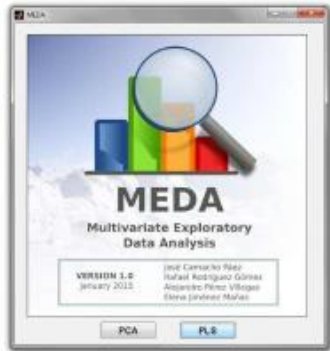


Is SS_p bigger enough than SS_E ?

$$F = \frac{SS_p}{SS_E}$$

ASCA example: Wheat dataset

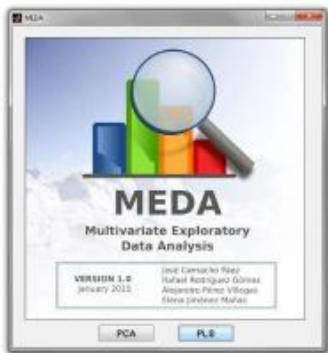
- MEDA Toolbox



$$X - \mathbf{1} \cdot m = \mathbf{b}_{time} \mathbf{F}_{time} + \mathbf{b}_{trait} \mathbf{F}_{trait} + \mathbf{b}_{tre} \mathbf{F}_{tre} + \mathbf{E}$$

ASCA example

- MEDA Toolbox

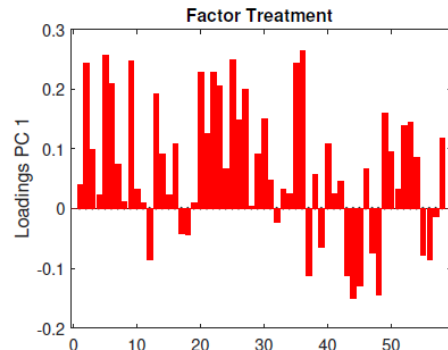
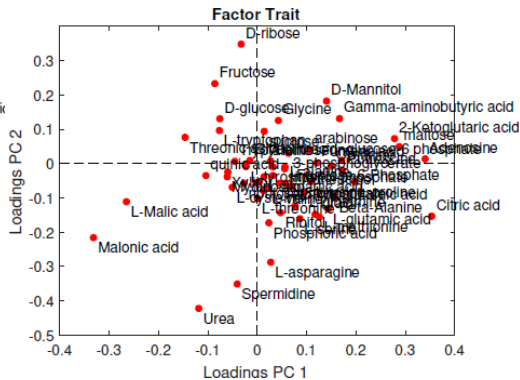
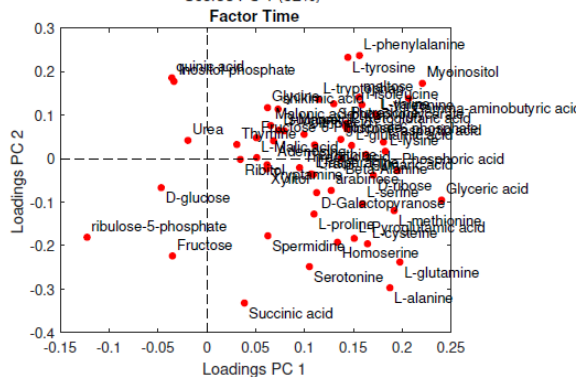
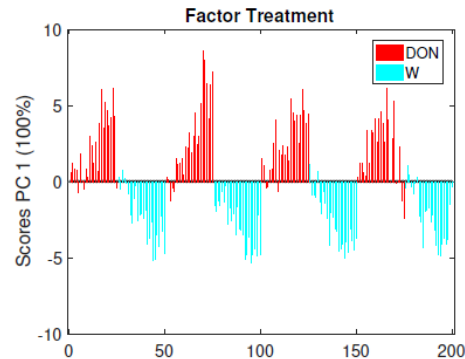
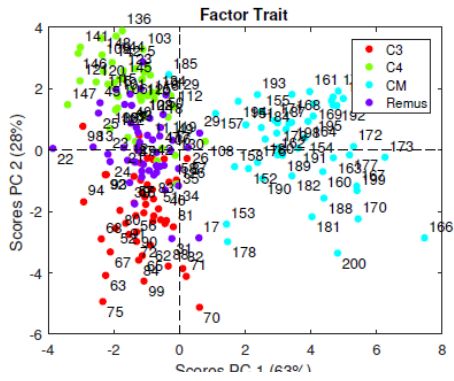
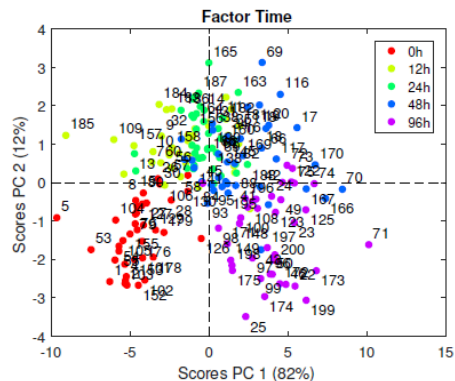


Source	SumSq	PercSumSq	df	MeanSq	F	Pvalue
'Mean'	61649	84.23	1	61649	NaN	NaN
'Factor 1'	2013.4	2.7508	4	503.34	13.87	0.000999
'Factor 2'	1366.8	1.8674	3	455.59	12.554	0.000999
'Factor 3'	1230.4	1.6811	1	1230.4	33.904	0.000999
'Residuals'	6931.5	9.4704	191	36.291	NaN	NaN
'Total'	73191	100	200	365.96	NaN	NaN

$$X - \mathbf{1} \cdot m = X_{time} + X_{trait} + X_{tre} + E$$

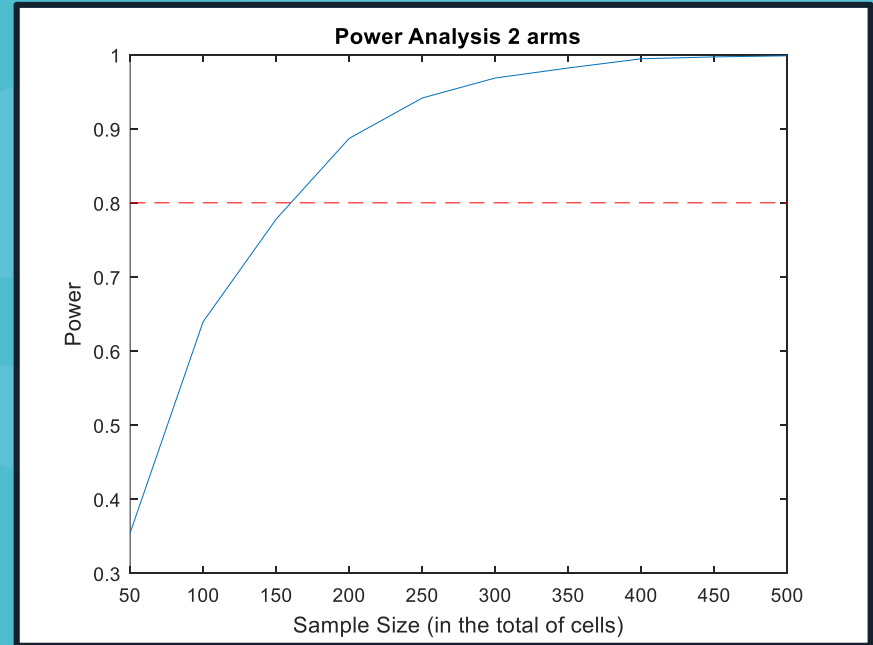
ASCA example

$$X - 1 \cdot m = X_{time} + X_{trait} + X_{tre} + E$$



Power Analysis: Sample Size

- Define effect size
- Estimate response variance
- Define error probability
 - Alpha: false positive (0.05 or 0.01)
 - Beta: 1 - false negative (0.8)



Simulated Power Curves for ASCA

- Estimate RELATIVE statistical POWER of variants of ...

A specific clinical trial / DoE / nested vs crossed / fixed vs random:

$$X - \mathbf{1} \cdot \mathbf{m} = X_{time} + X_{trait} + X_{tre} + E$$

Vs

$$X - \mathbf{1} \cdot \mathbf{m} = X_{time} + X_{tre} + E$$

Vs

$$X - \mathbf{1} \cdot \mathbf{m} = X_{time} + X_{tre} + X_{time \times tre} + E$$

Simulated Power Curves for ASCA

- Estimate RELATIVE statistical POWER of variants of ...

A specific test and/or test statistic (related to the previous):

Journal of Statistical Computation and Simulation,
2003, Vol. 73(2), pp. 85–113

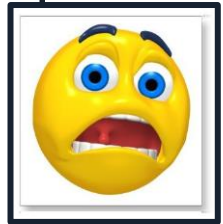


PERMUTATION TESTS FOR MULTI-FACTORIAL ANALYSIS OF VARIANCE

MARTI J. ANDERSON^{a,*} and CAJO J. F. TER BRAAK^{b,†}

^aCentre for Research on Ecological Impacts of Coastal Cities, Marine Ecology Laboratories, A11,
University of Sydney, NSW, 2006, Australia; ^bBiometris, Wageningen University and
Research Centre, Box 100, 6700 AC, Wageningen, The Netherlands

(Received 1 February 2001; In final form 13 August 2002)



Simulated Power Curves for ASCA

- Estimate RELATIVE statistical POWER of variants of ...

A specific data preprocessing choice:

Missing data, transformation, outlier detection, ...

- Materials: <https://github.com/josecamachop/PowerCurvesASCA>

The screenshot shows the GitHub repository page for 'PowerCurvesASCA'. At the top, there is a header with the repository name, a 'Public' badge, and interaction buttons for Pin, Unwatch (3), Fork (0), and Star (2). Below the header, there are navigation options for 'main' (4 branches, 0 tags), 'Go to file', 'Add file', and 'Code'. A notification banner states 'Your main branch isn't protected' with a 'Protect this branch' button. The commit history shows a recent update to 'README.md' by 'josecamachop' on June 9, with 10 commits. A file named 'A' is listed with the description 'Delete unnecessary code' and a commit date of 'last year'. On the right side, there is an 'About' section with a gear icon, containing text about the software corresponding to a paper by Camacho, José, Caridad Diaz, and Pedro Sánchez-Rovira, published in the Journal of Chemometrics.

PowerCurvesASCA Public

Pin Unwatch 3 Fork 0 Star 2

main 4 branches 0 tags

Go to file Add file Code

Your main branch isn't protected
Protect this branch from force pushing or deletion, or require status checks before merging. [Learn more](#) [Protect this branch](#) ×

josecamachop Update README.md 31b24ae on Jun 9 10 commits

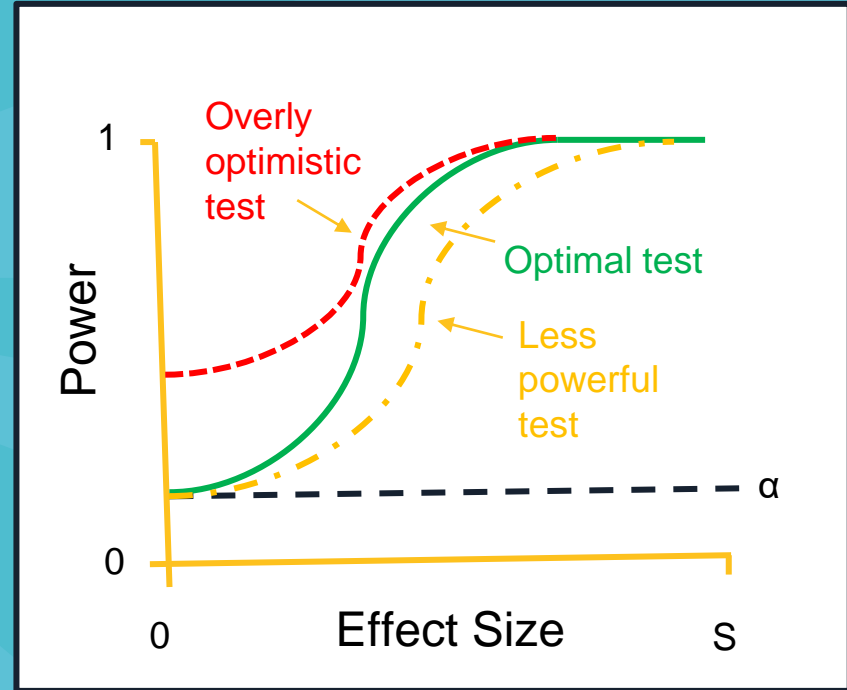
A Delete unnecessary code last year

About

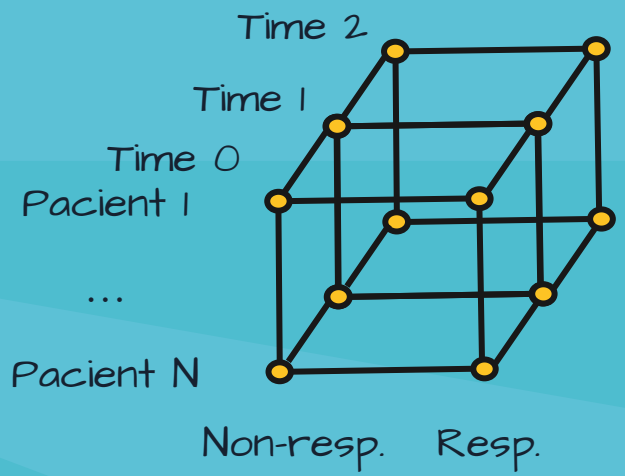
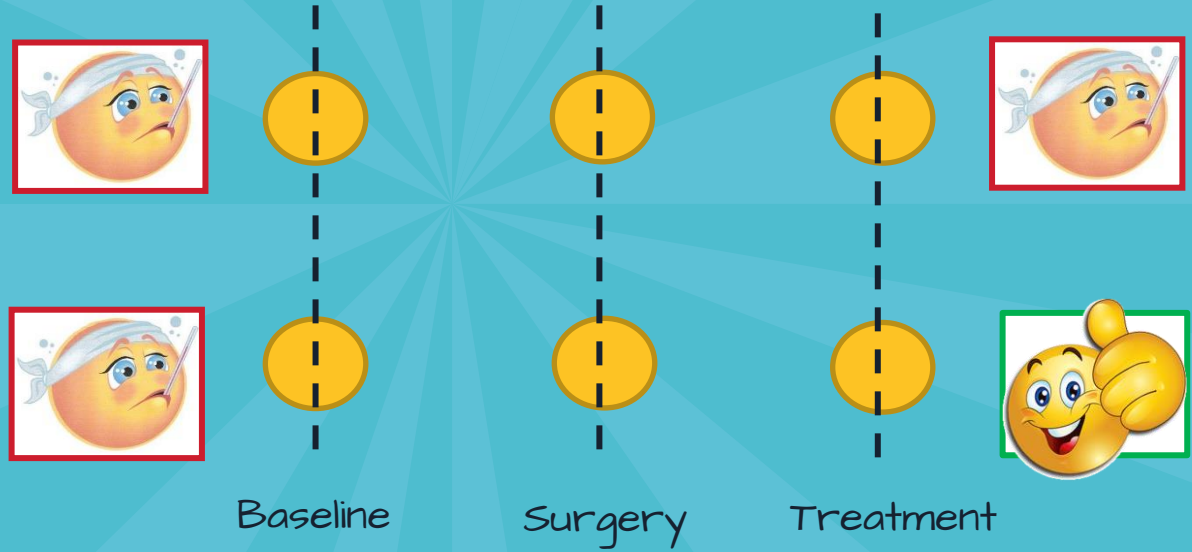
Software corresponding to the paper "Camacho, José, Caridad Diaz, and Pedro Sánchez-Rovira. "Permutation Tests for ASCA in Multivariate Longitudinal Intervention Studies." Journal of Chemometrics: e3398.", <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/full/10.1002/cem.3398>

Simulated Power Curves for ASCA

- Define variants
- Define
 - Alpha: false positive (0.05 or 0.01)
 - Increasing effect size (null to S)
- Design the simulation *



Example I: Predicting dynamic response to neoadjuvant chemotherapy in breast cancer: a novel metabolomics approach



$$X = 1m + \mathbf{A} + B + C(A) + \mathbf{AB} + E$$

Example I: Predicting dynamic response to neoadjuvant chemotherapy in breast cancer: a novel metabolomics approach

- Motivation: Our first results with real data were unexpectedly significant
- Steps:
 - Variants: 15+ variants in DoE and statistical tests
 - alpha = 0.05
 - effect size from 0 to 0.5
- Curves computed: $X = 1m + \mathbf{A} + B + C(A) + \mathbf{AB} + E$

Example 1. Predicting dynamic response to neoadjuvant chemotherapy: a novel

- Motivation: Our first
- Steps:
 - Variants: 15
 - alpha = 0.05
 - effect size
 - Simulation:

```
effectS = 0:0.05:0.5;

Xpac = randn(length(up),length(var_1));
Xpac = Xpac/norm(Xpac);
Xtime = randn(length(ut),length(var_1));
Xtime = Xtime/norm(Xtime);
for i = 1:length(obs_1)
    if strcmp(class{i},'R')
        Xstruct(i,:) = Xpac(vp(i,:),:) + Xtime(vt(i,:),:);
    else
        Xstruct(i,:) = Xpac(vp(i,:),:);
    end
end

Xnoise = randn(length(obs_1),length(var_1));
Xnoise = Xnoise/norm(Xnoise);

for a = 1:length(effectS)

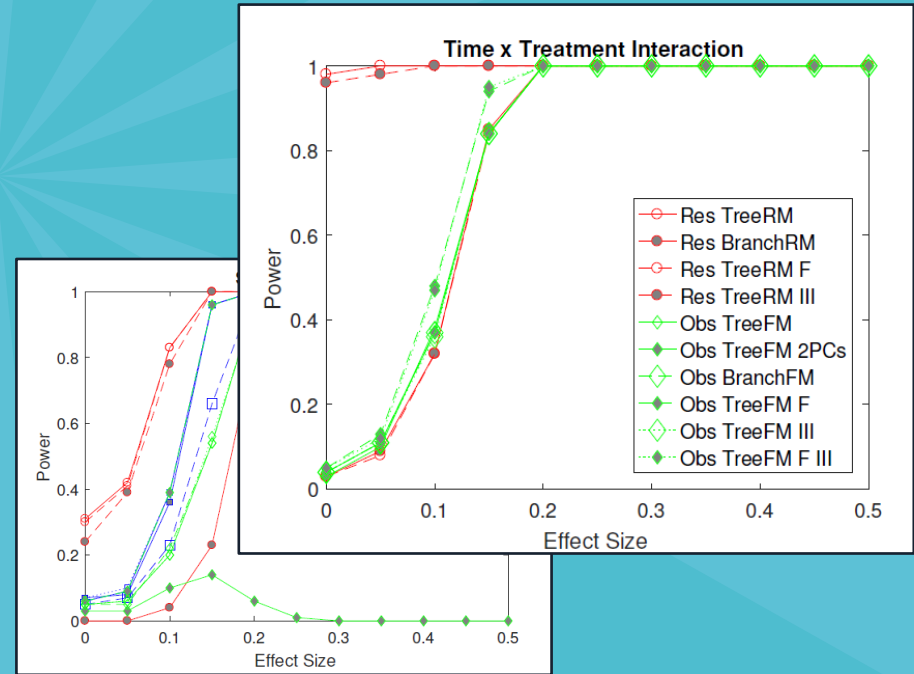
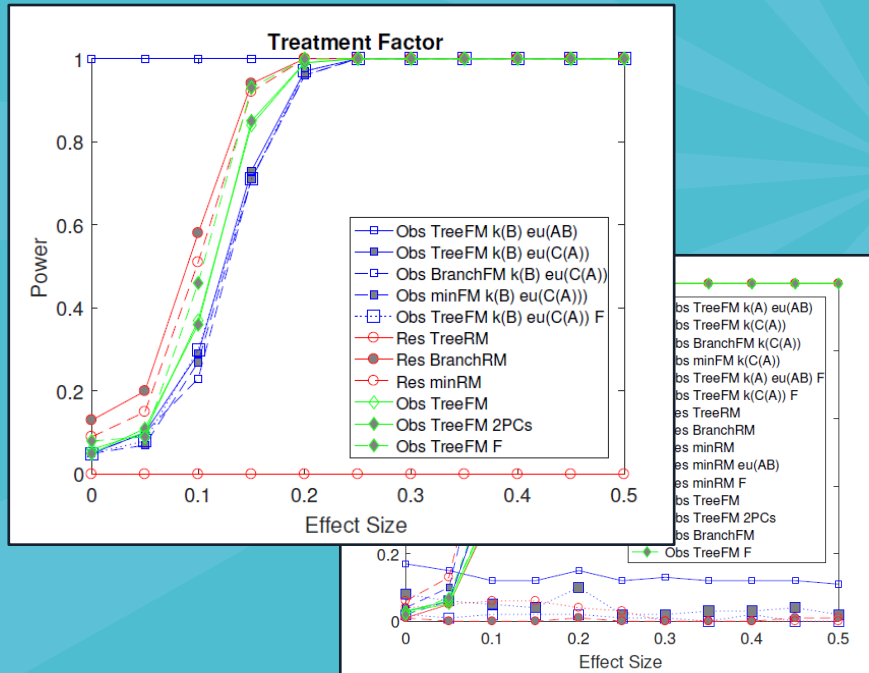
    % Make a blend

    Xm = effectS(a)*Xstruct + (1-effectS(a))*Xnoise;
```

$$X = 1m + A + B + C(A) + AB + E$$

Permutation Tests for ASCA in Multivariate Longitudinal Intervention Studies

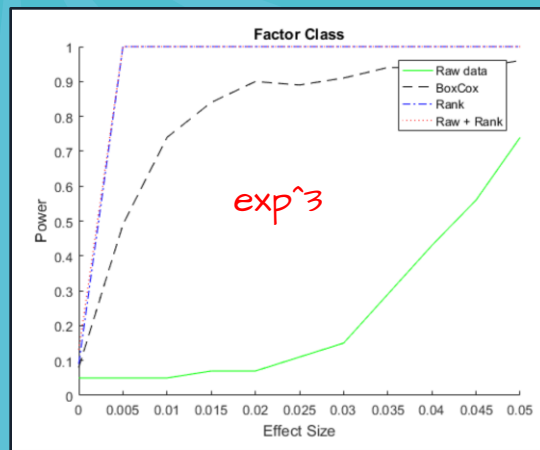
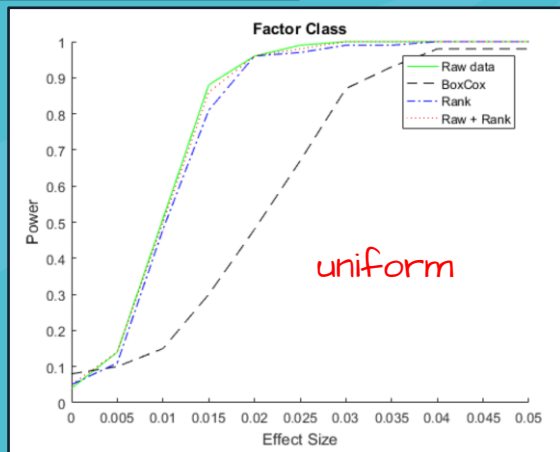
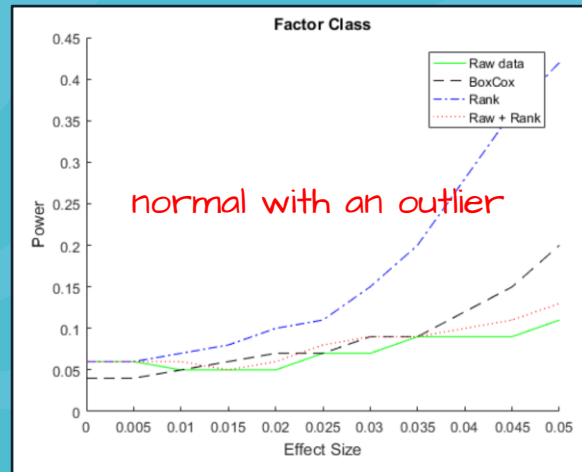
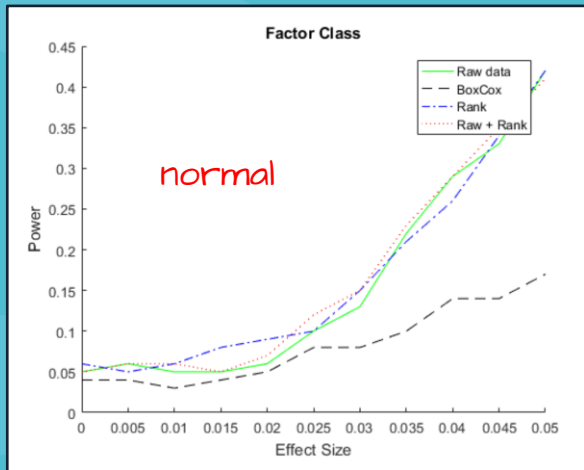
$$X = 1m + \mathbf{A} + B + C(A) + \mathbf{AB} + E$$



Example II: Transformations and outliers

- Motivation: We obtained inconsistent results after outliers isolation and several transformations
- Steps:
 - Variants: Raw data, Box-Cox, Rank, Raw+Rank
 - $\alpha = 0.05$
 - effect size from 0 to 0.05
- Curves computed: residuals distributed as normal, uniform, \exp^3 and normal with an outlier

Example II: Transformations and outliers



CONCLUSION

- ASCA is a really interesting and competitive method for multivariate analysis in designed experiments
- BUT it is still under development
- There are plenty of problems of interest that deserve research
- **Simulated Power Curves** are a simulation framework to explore these problems

THANKS

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik** and illustrations by **Stories**
Please keep this slide for attribution