Simulated Power Curves in ASCA

Topics in Chemometrics

2023

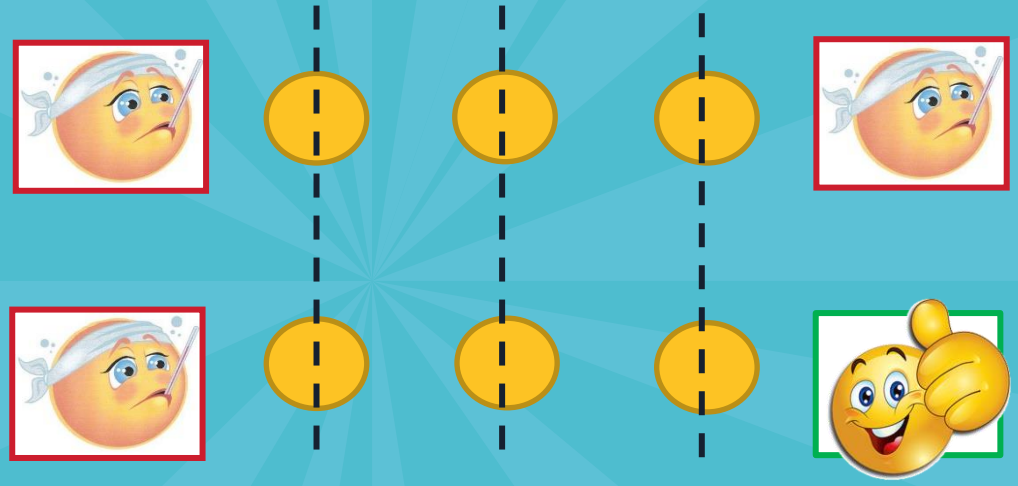
Project *PID2020-113462RB-I00* funded by:

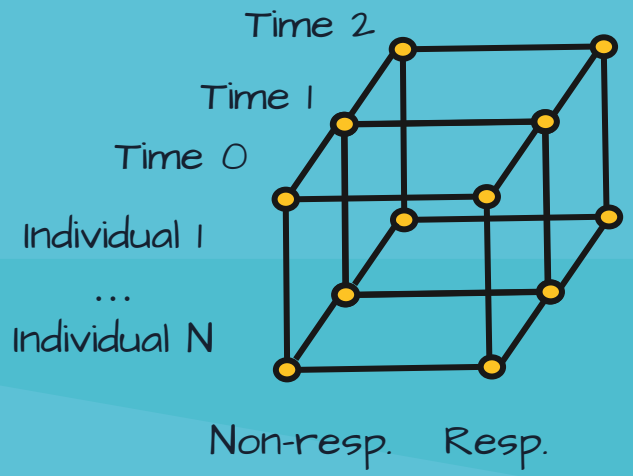
MINISTERIO DE CIENCIA E INNOVACIÓN

AGENCIA ESTATAL DE INVESTIGACIÓN

José Camacho
josecamacho@ugr.es
Computational Data Science Lab

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

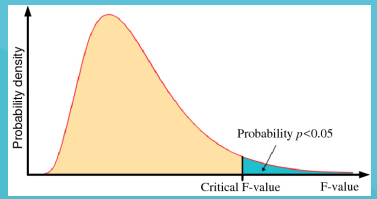




Biomarker discovery: Statistical inference

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

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



So many possibilities…

- Clinical trial …

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

- Factors …

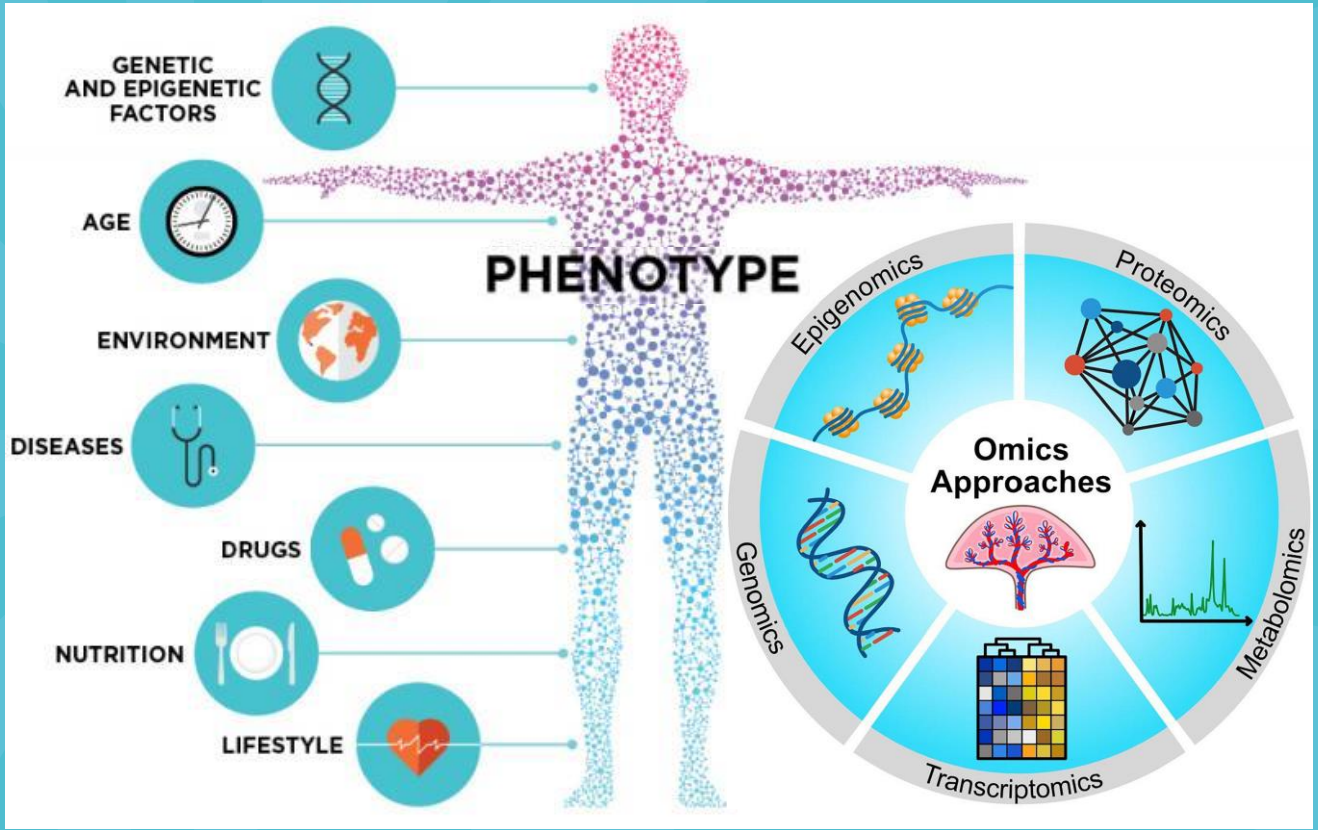
  … fixed, random

- Relationships among factors …

  … crossed, nested, interactions

- Complexities …

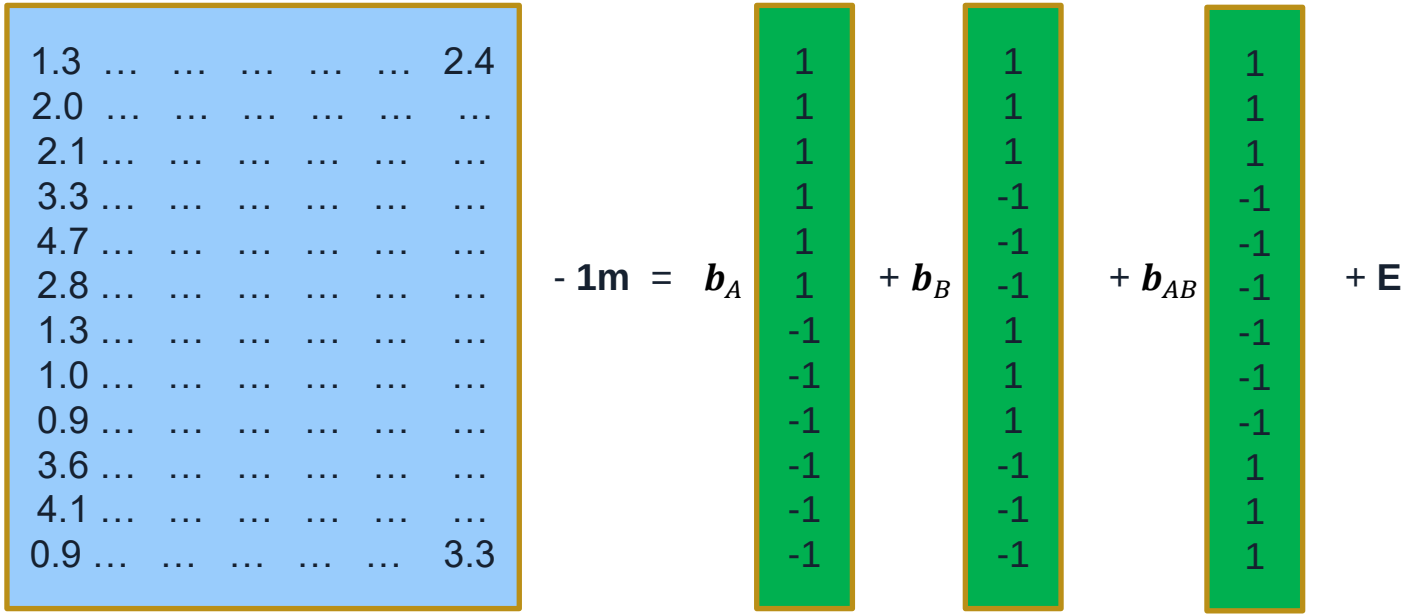
  … unbalanced, distributions, outliers, missing data …



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

ANOVA with MULTIVARIATE RESPONSE

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



ANOVA with MULTIVARIATE RESPONSE

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

General Lineal Model

$$b = argmin \, ||X - m - bF||_2$$

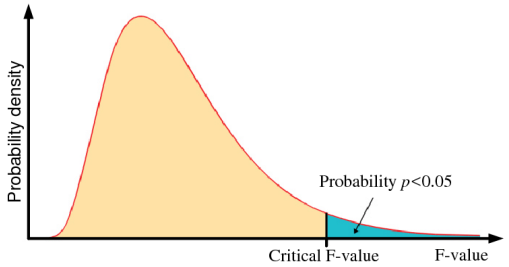
$$b = (F' * F)^{-1} F' \cdot (X - m)$$

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

MULTIVARIATE EXTENSIONS TO ANOVA

$$\|\boldsymbol{X} - \mathbf{1} \cdot \boldsymbol{m}\|_F^2 = \|\boldsymbol{X}_p\|_F^2 + \|\boldsymbol{X}_t\|_F^2 + \|\boldsymbol{X}_i\|_F^2 + \|\boldsymbol{X}_{pt}\|_F^2 + \|E\|_F^2$$

$$SS_X \qquad SS_p \qquad SS_t \qquad SS_i \qquad SS_{pt} \qquad SS_E$$

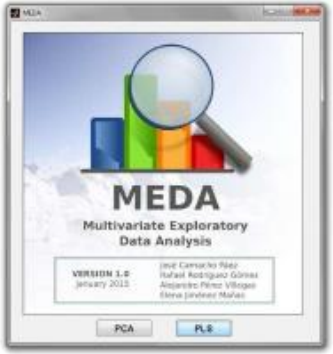


Is SS_p bigger enough than SS_E?

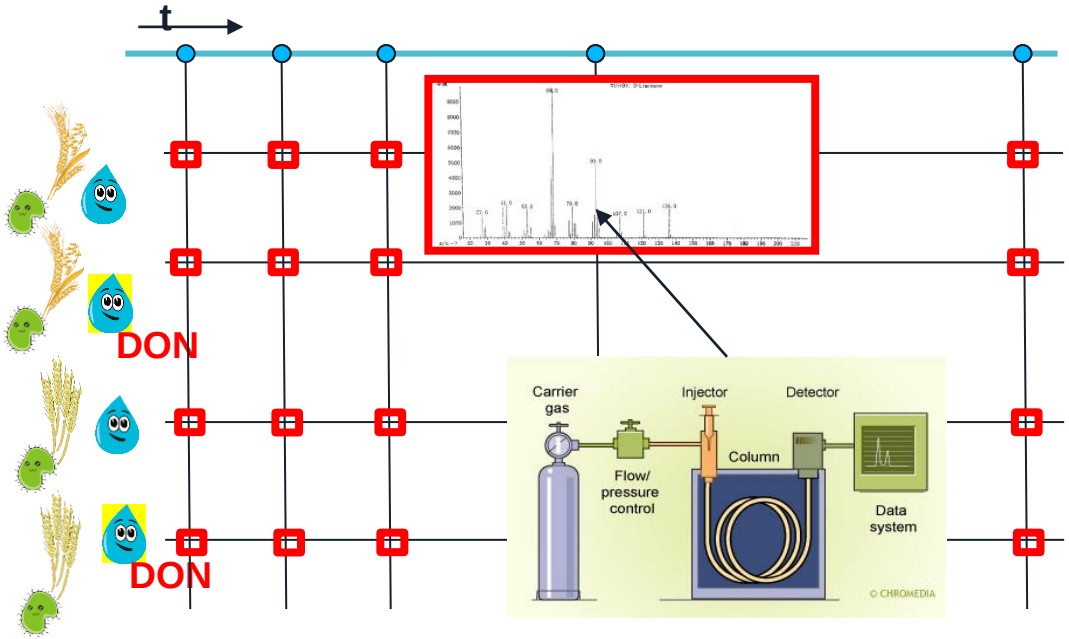
$$F = \frac{SS_p}{SS_E}$$

ASCA example: Wheat dataset

- MEDA Toolbox



MATLAB
MathWorks

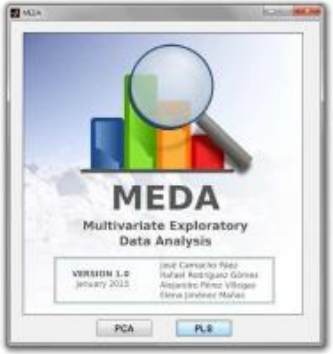


$$X - 1 \cdot m = b_{time}F_{time} + b_{trait}F_{trait} + b_{tre}F_{tre} + E$$

https://github.com/CoDaSLab/MEDA-Toolbox

ASCA example

- MEDA Toolbox



MATLAB
MathWorks

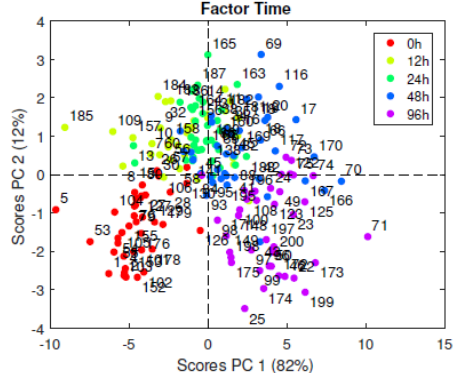
Source	SumSq	PercSumSq	df	MeanSq	F	Pvalue
'Mean'	61649	84.23	1	61649	NaN	NaN
'Factor 1'	2013.4	2.7508	4	503.34	13.87	0.000999
'Factor 2'	1366.8	1.8674	3	455.59	12.554	0.000999
'Factor 3'	1230.4	1.6811	1	1230.4	33.904	0.000999
'Residuals'	6931.5	9.4704	191	36.291	NaN	NaN
'Total'	73191	100	200	365.96	NaN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

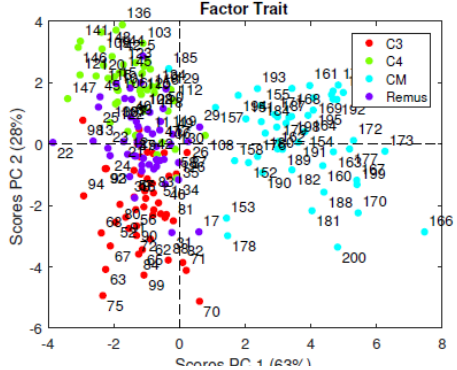
$$\boldsymbol{X} - \boldsymbol{1} \cdot \boldsymbol{m} = \boldsymbol{X}_{time} + \boldsymbol{X}_{trait} + \boldsymbol{X}_{tre} + \boldsymbol{E}$$

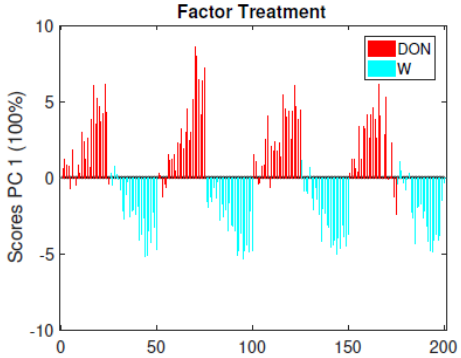
https://github.com/CoDaSLab/MEDA-Toolbox

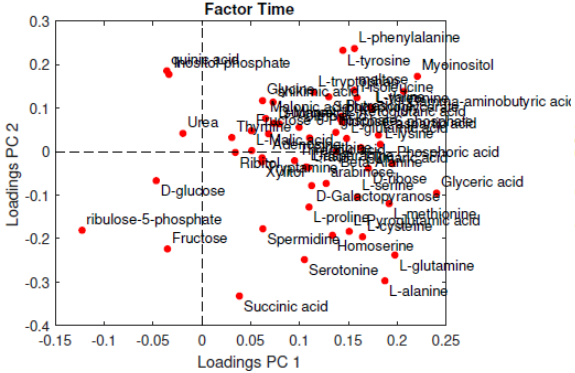
ASCA example

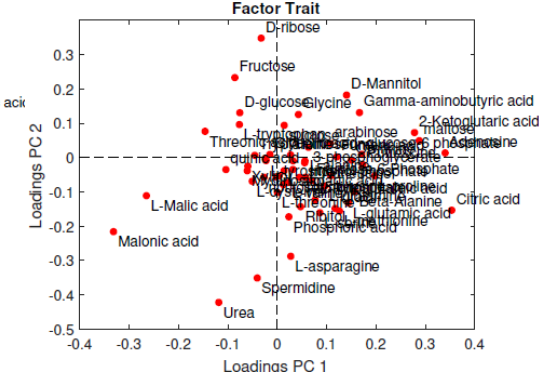
$$X - 1 \cdot m = X_{time} + X_{trait} + X_{tre} + E$$

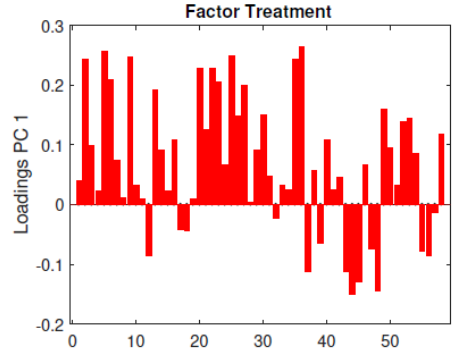








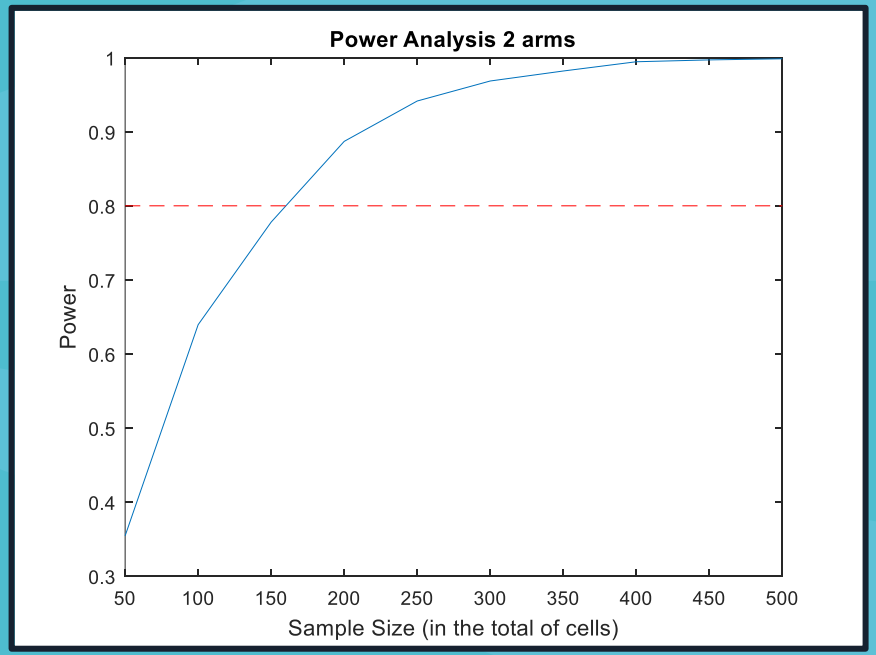




https://github.com/CoDaSLab/MEDA-Toolbox

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:

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

$$Vs$$

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

$$Vs$$

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

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

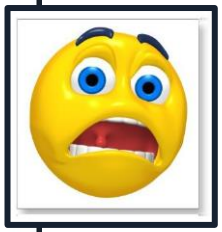
Taylor & Francis
Taylor & Francis Group

PERMUTATION TESTS FOR MULTI-FACTORIAL ANALYSIS OF VARIANCE

MARTI J. ANDERSON[a,*] and CAJO J. F. TER BRAAK[b,†]

[a]*Centre for Research on Ecological Impacts of Coastal Cities, Marine Ecology Laboratories, A11, University of Sydney, NSW, 2006, Australia;* [b]*Biometris, Wageningen University and Research Centre, Box 100, 6700 AC, Wageningen, The Netherlands*

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



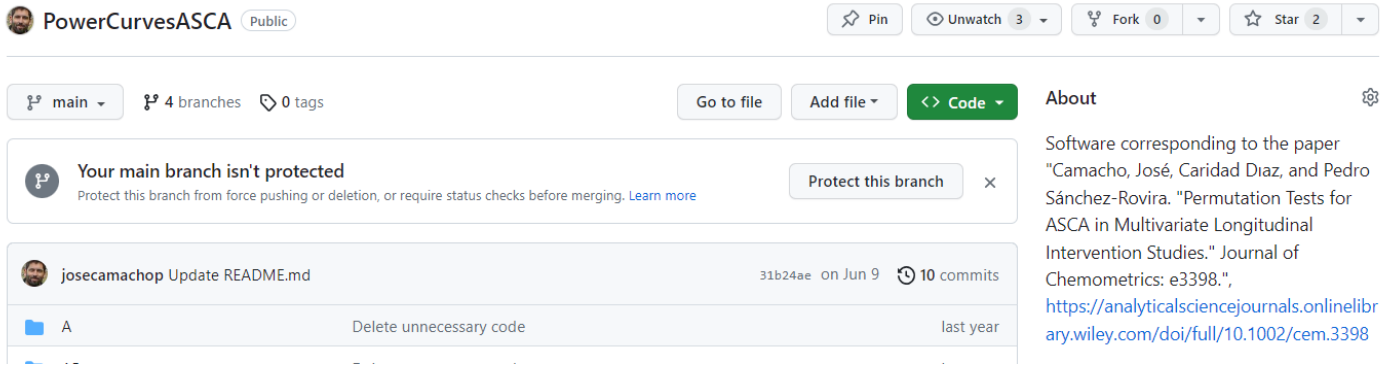
Simulated Power Curves for ASCA

- Estimate RELATIVE statistical POWER of variants of ...

  A specific data preprocessing choice:

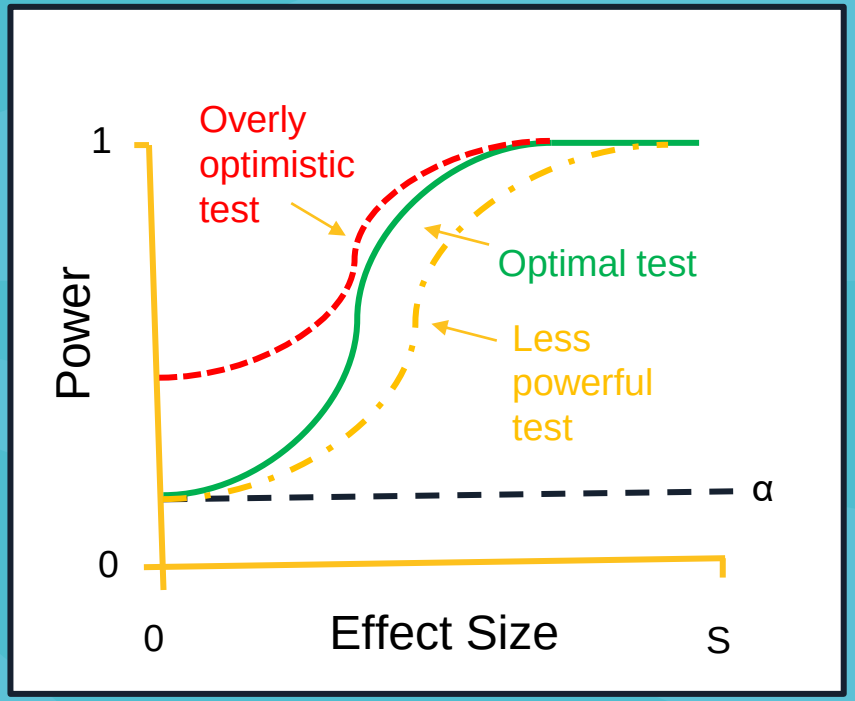
  Missing data, transformation, outlier detection, ...

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

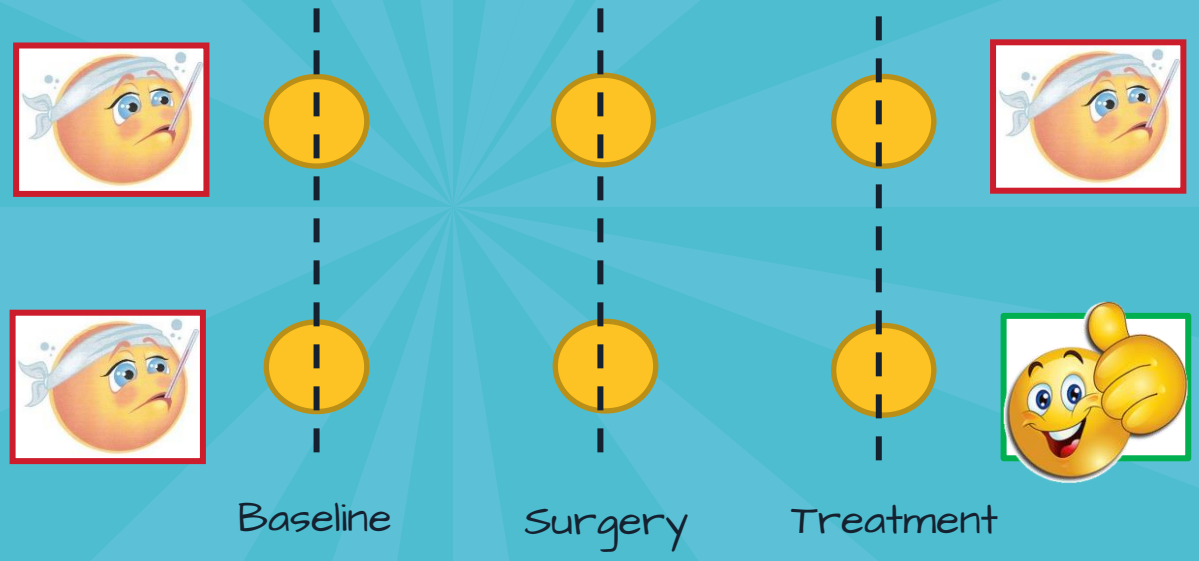


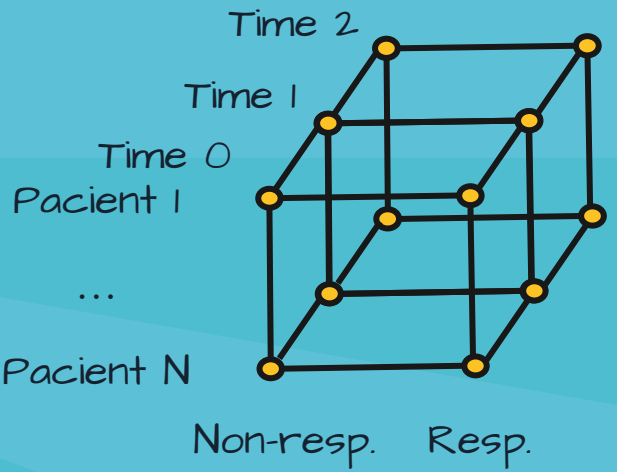
Simulated Power Curves for ASCA

- 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 I: Predicting dynamic response to neoadjuvant c…cer: a novel

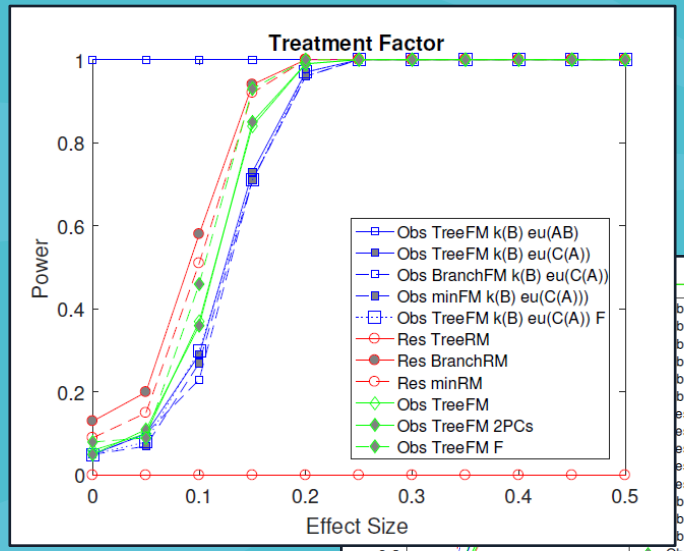
- Motivation: Our fi…
- Steps:
  - Variants: 15…
  - alpha = 0.05…
  - effect size…
  - Simulation:

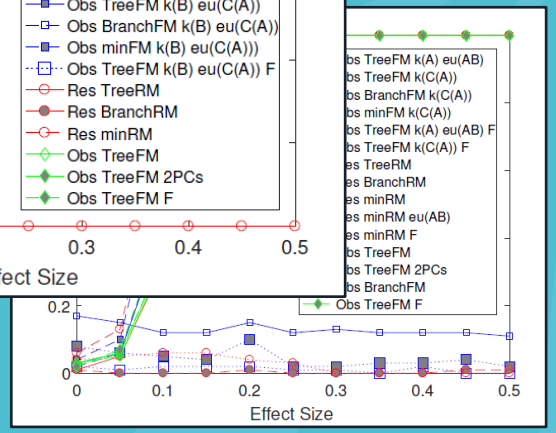
```
effectS = 0:0.05:0.5;

Xpac = randn(length(up),length(var_l));
Xpac = Xpac/norm(Xpac);
Xtime = randn(length(ut),length(var_l));
Xtime = Xtime/norm(Xtime);
for i = 1:length(obs_l)
    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_l),length(var_l));
Xnoise = Xnoise/norm(Xnoise);

for a = 1:length(effectS)

    % Make a blend

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

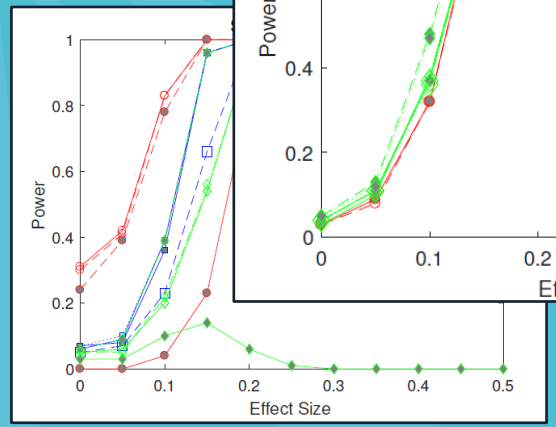
$$X = 1m + A + B + C(A) + AB + E$$

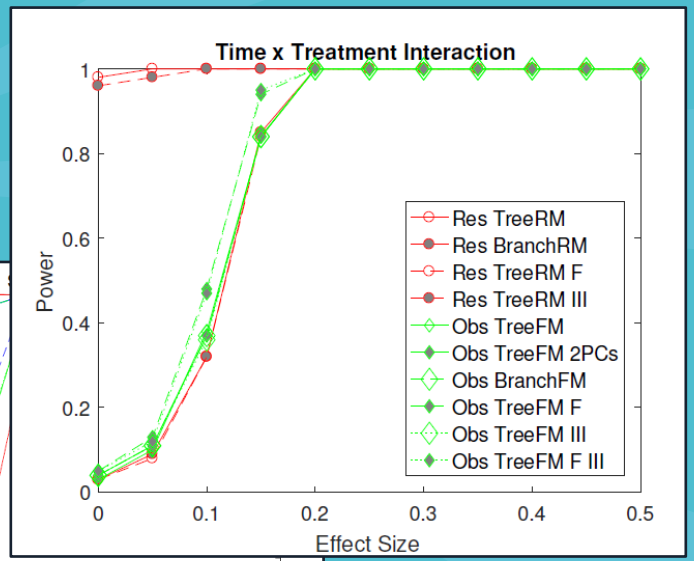
Permutation Tests for ASCA in Multivariate Longitudinal Intervention Studies

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





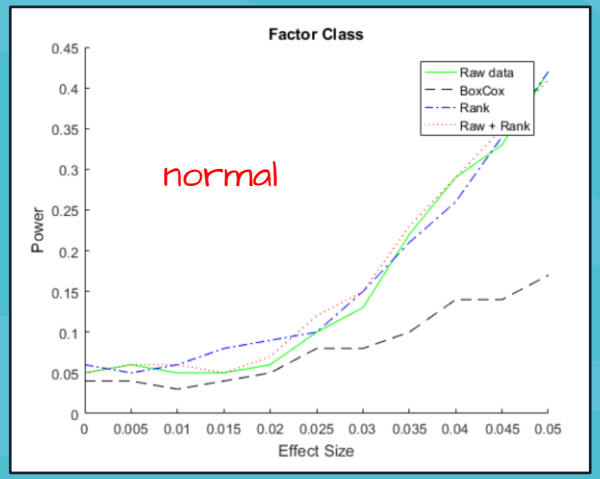


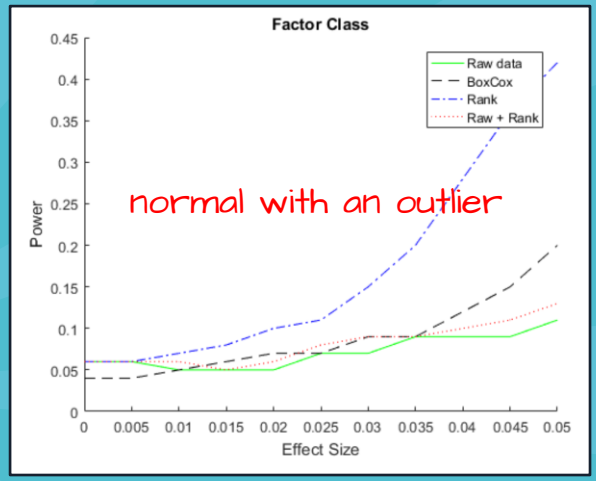


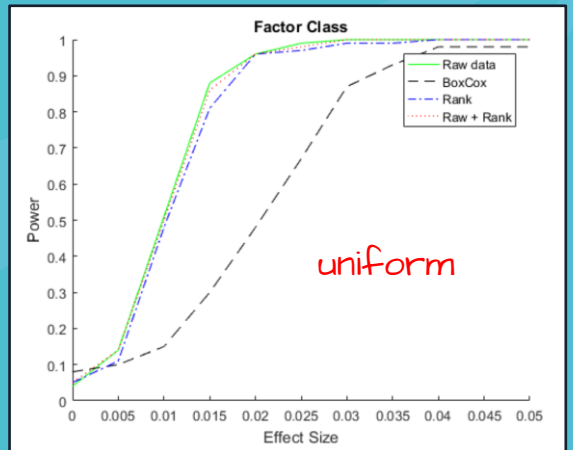
Example II: Transformations and ouliers

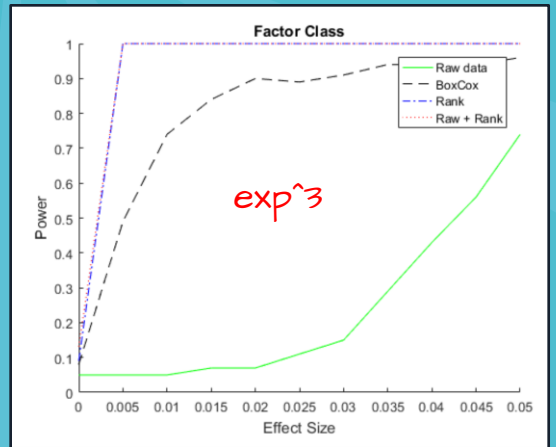
- Motivation: We obtained inconsistent results after outliers isolation and several transformations
- Steps:
    - Variants: Raw data, Box-Cox, Rank, Raw+Rank
    - alpha = 0.05
    - effect size from 0 to 0.05
- Curves computed: residulals distributed as normal, uniform, exp^3 and normal with an outlier

Example II: Transformations and ouliers









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