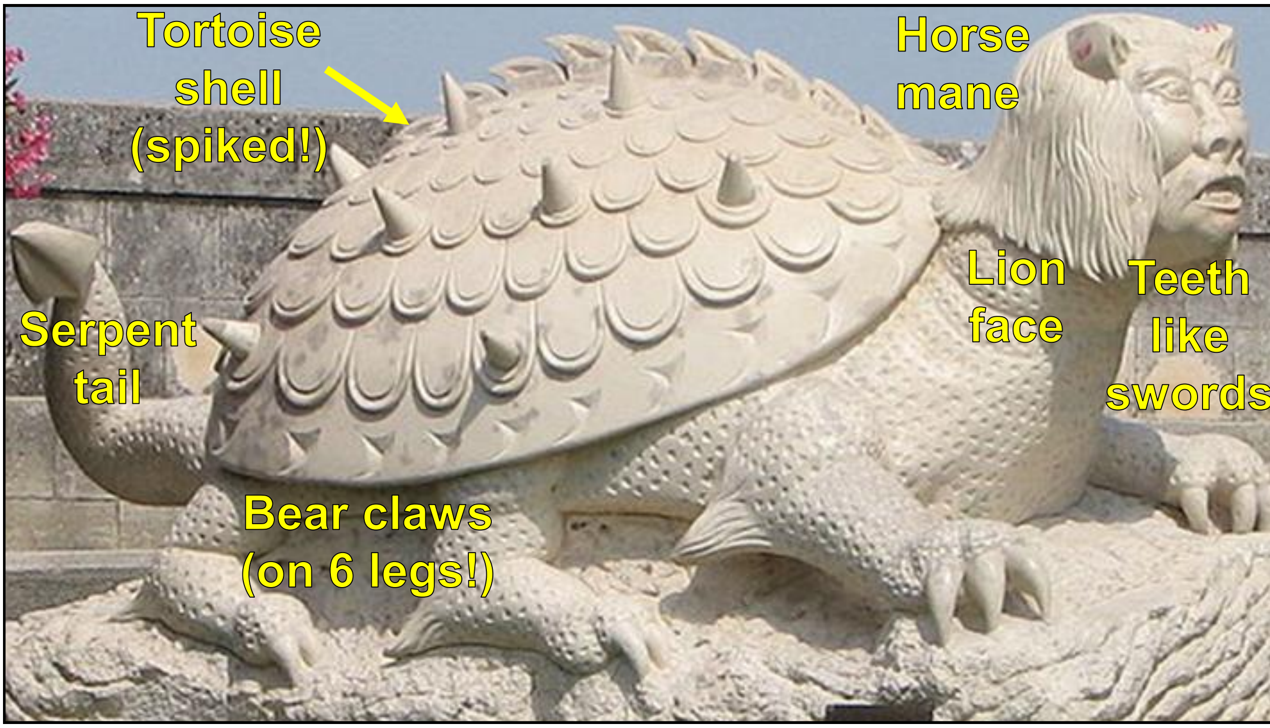
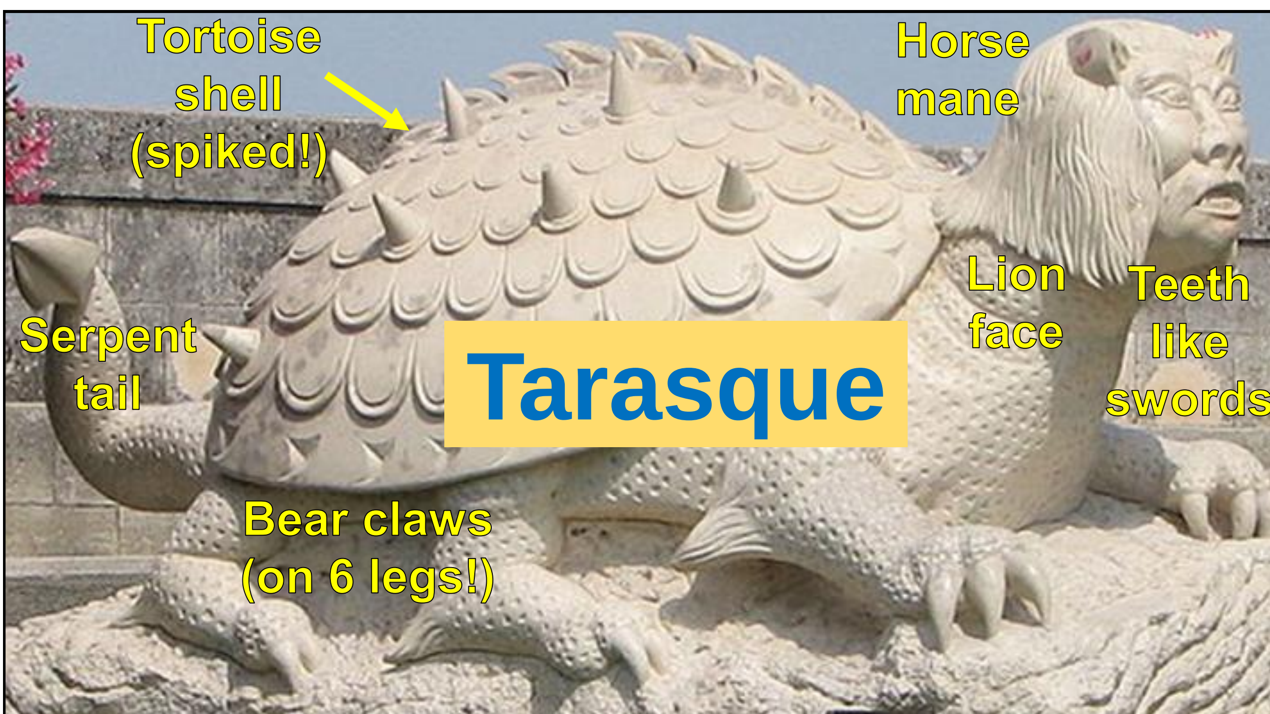
Investigating relationships in sensory and instrumental data using component-based methods

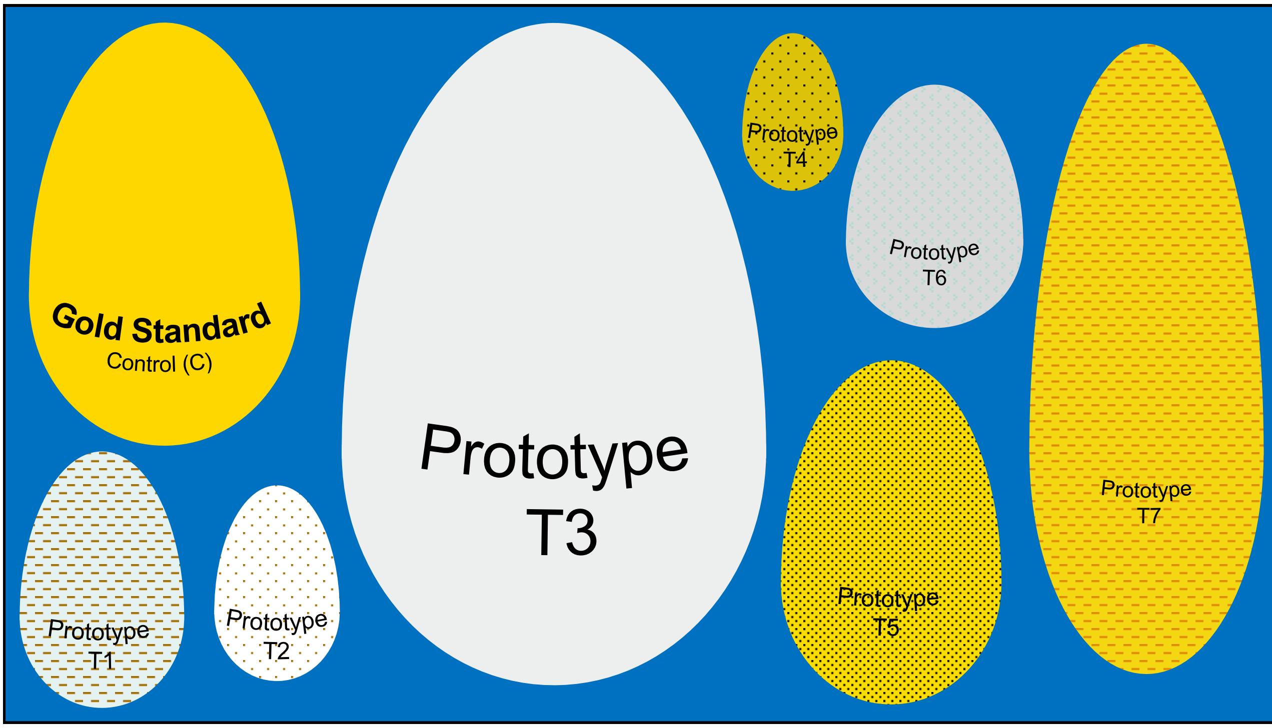
John C. Castura

Compusense

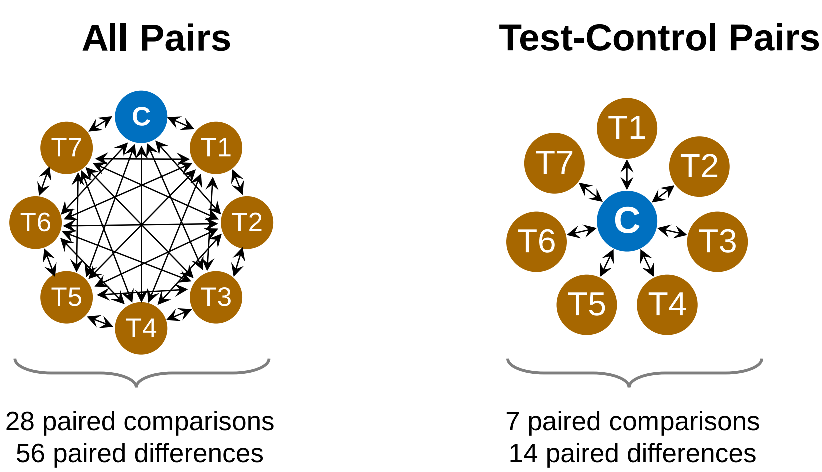
Castura, J.C. (2024). Investigating relationships in sensory and instrumental data using component-based methods. *Society of Sensory Professionals 2024 Conference.* 2-4 October. Pittsburgh, PA, USA. (Oral).


Tortoise
shell
(spiked!)
Horse
mane
Serpent
tail
Lion
face
Teeth
like
swords
Bear claws
(on 6 legs!)

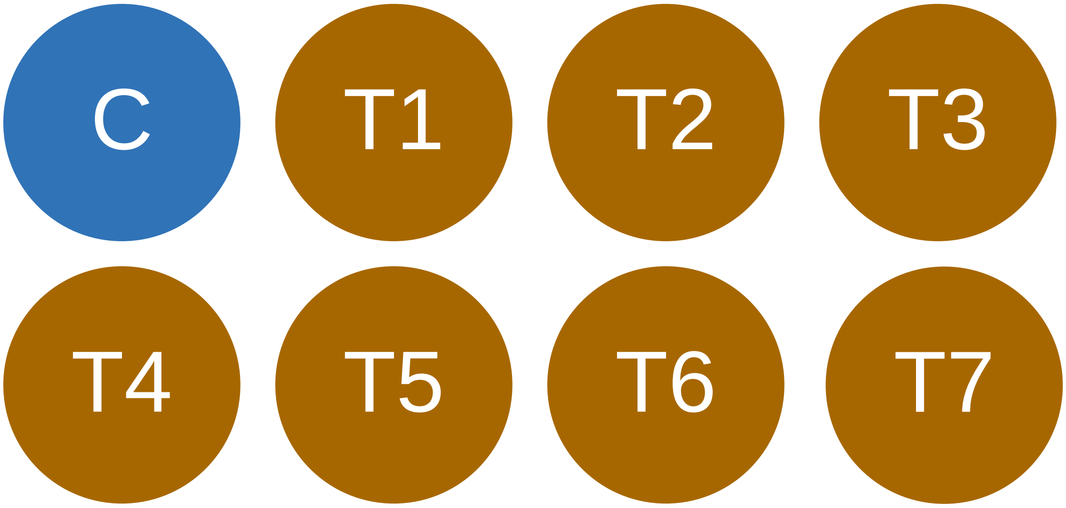

Tortoise
shell
(spiked!)
Horse
mane
Serpent
tail
Lion
face
Teeth
like
swords
Tarasque
Bear claws
(on 6 legs!)



All pairs vs. a subset of paired comparisons

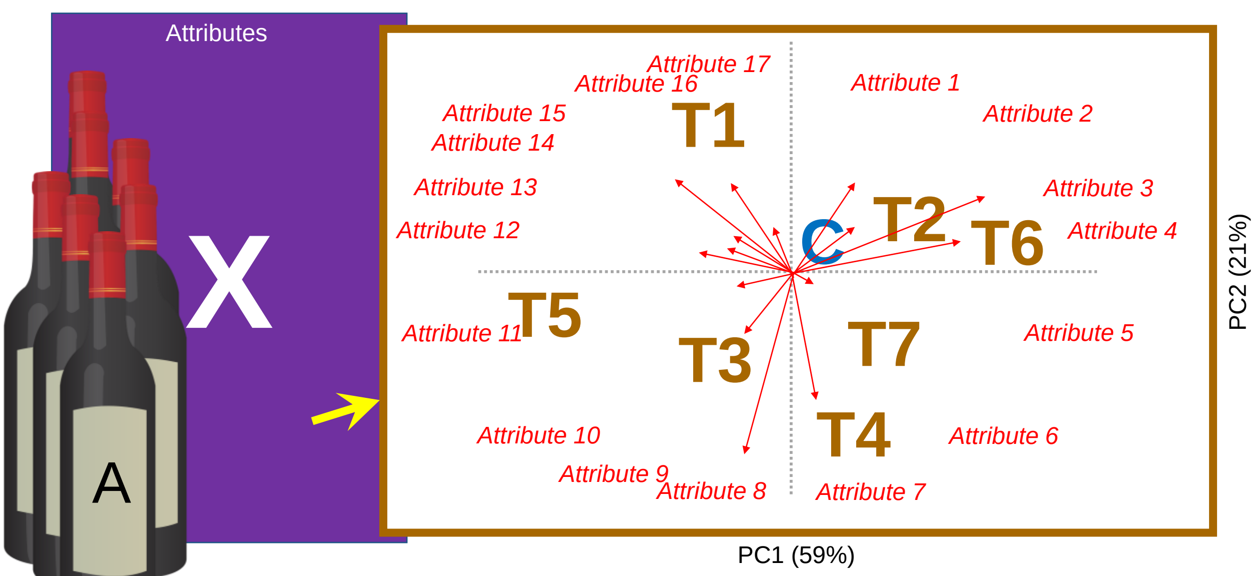


Castura, J.C., Varela, P., & Næs, T. (2023). Investigating only a subset of paired comparisons after principal component analysis. *Food Quality and Preference*, 110, 104941. https://doi.org/10.1016/j.foodqual.2023.104941

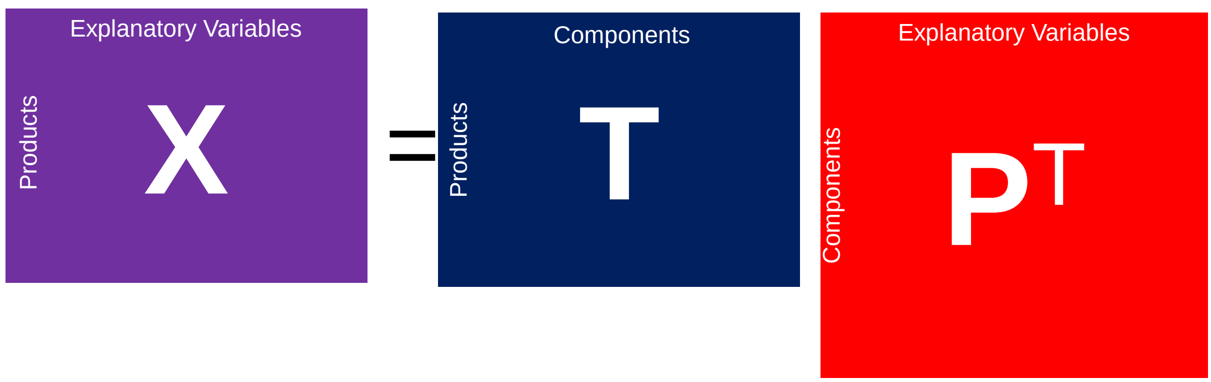


Whatever shall we do?

Sensory evaluation

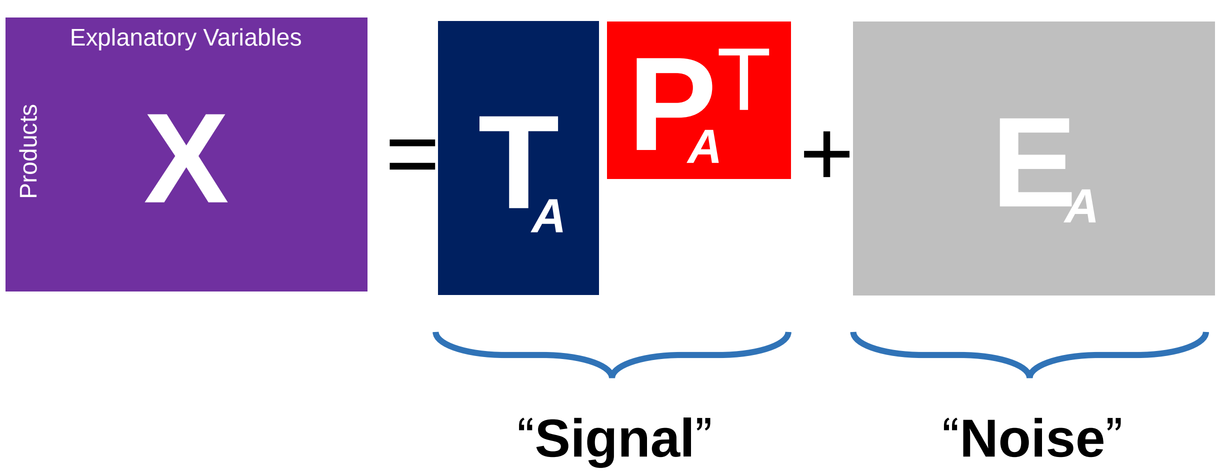


Principal component analysis of X

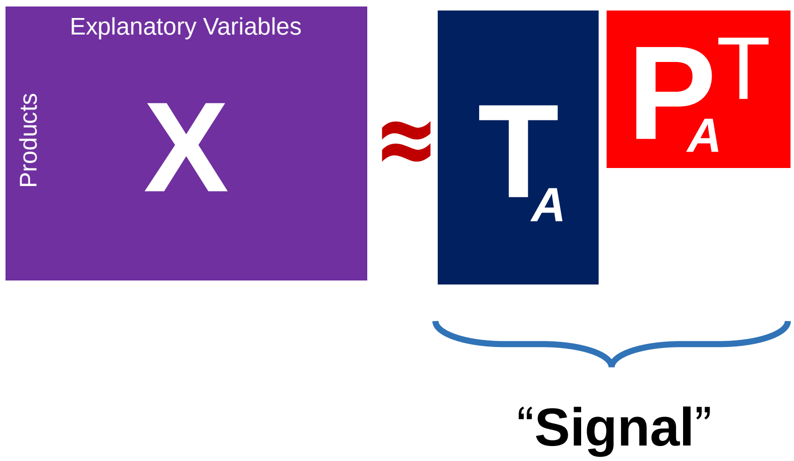


Pearson (1901) and Hotelling (1933)

Dimension reduction to *A* principal components

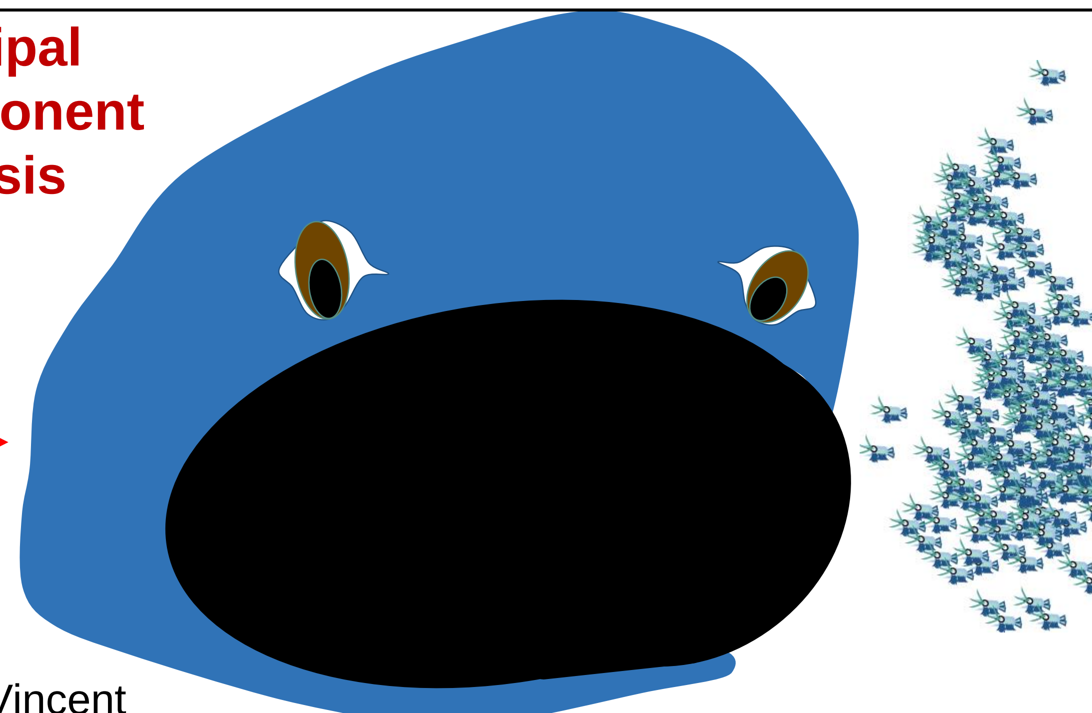


Dimension reduction to *A* principal components



Principal component analysis

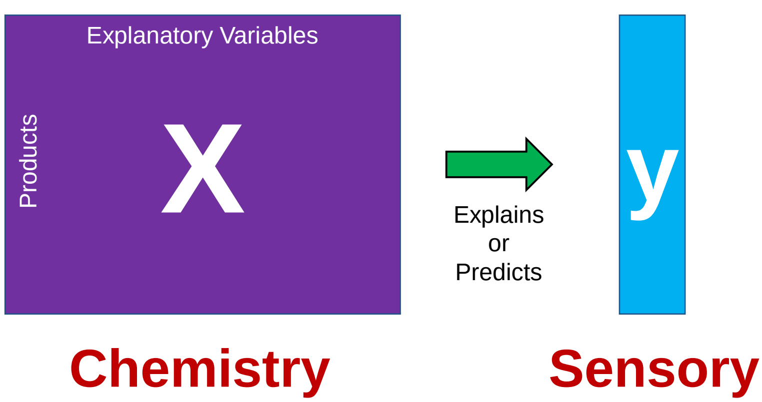
Baleen whale



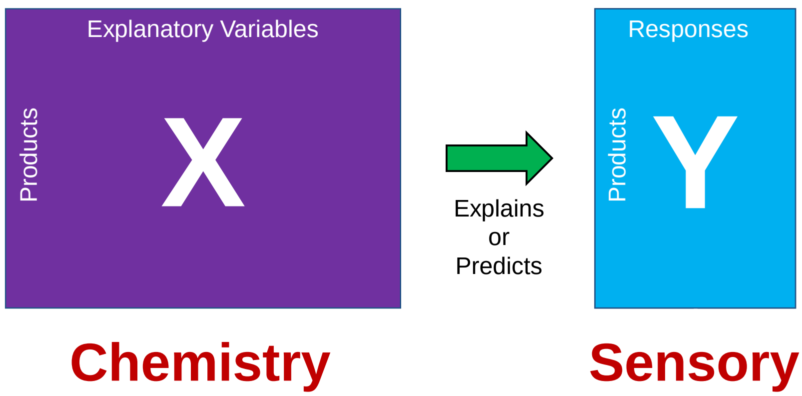
*merci Vincent

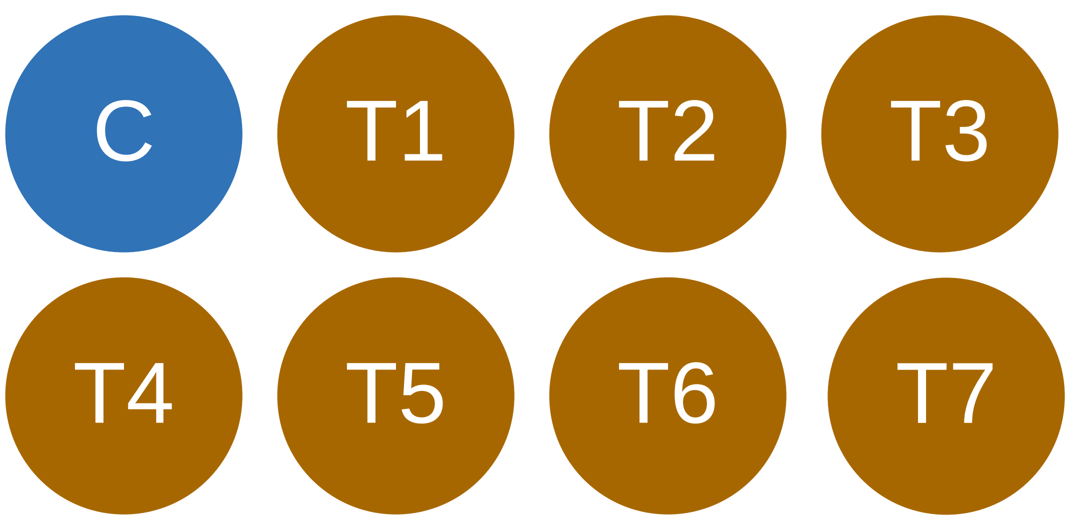
Principal component

Explaining and predicting data relationships



Explaining and predicting data relationships

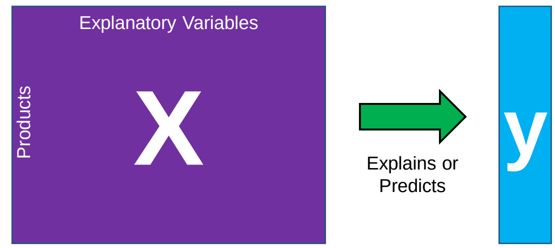




Whatever shall we do?


Whatever shall we do?


Multiple linear regression?
Whatever shall we do?



Chemistry **Sensory**

Multiple linear regression

will not do the job

Multiple Linear Regression

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{f}$$

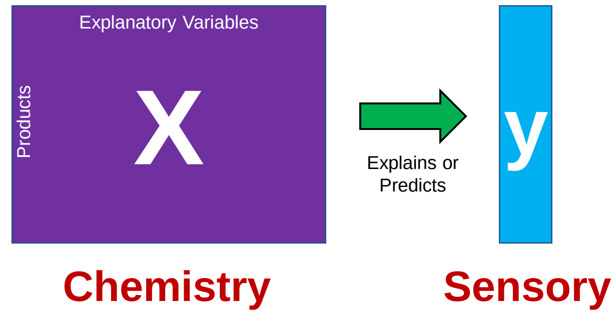
Estimation problems due to…
- **multicollinearity**
- **more variables than objects**


Multiple linear regression?
Whatever shall we do?


Multiple linear regression?
Nee!
Nee!
Nee!
Whatever shall we do?


Whatever shall we do?


Principal component regression?
Whatever shall we do?



Principal component regression

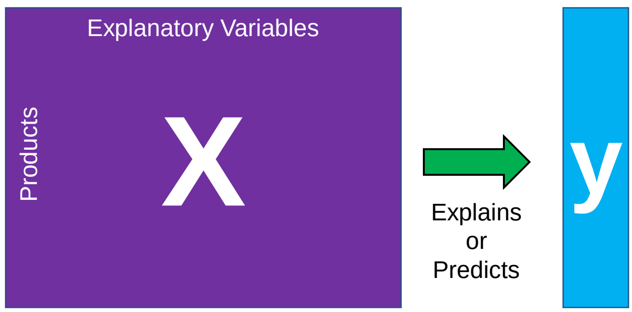
Unsupervised & Supervised

Principal component regression (PCR)

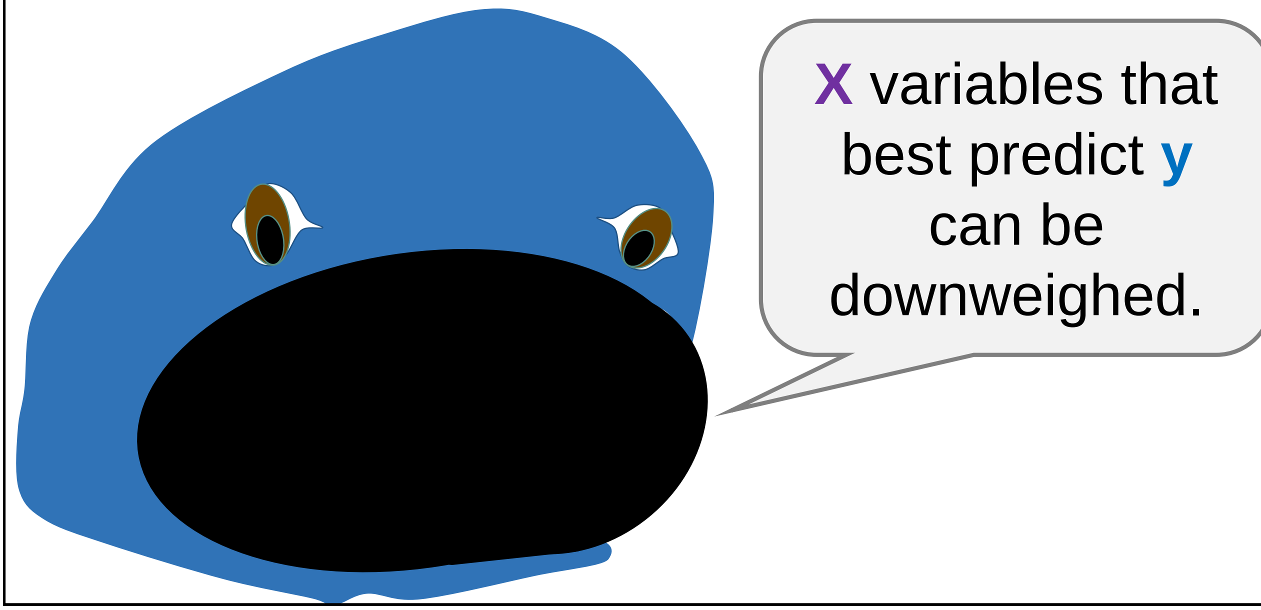
We want to ***explain*** and ***predict*** the response **y** from multivariate **X**.

Response **y** is regressed on principal components of **X**.

Explaining and predicting data relationships



Explaining and predicting data relationships

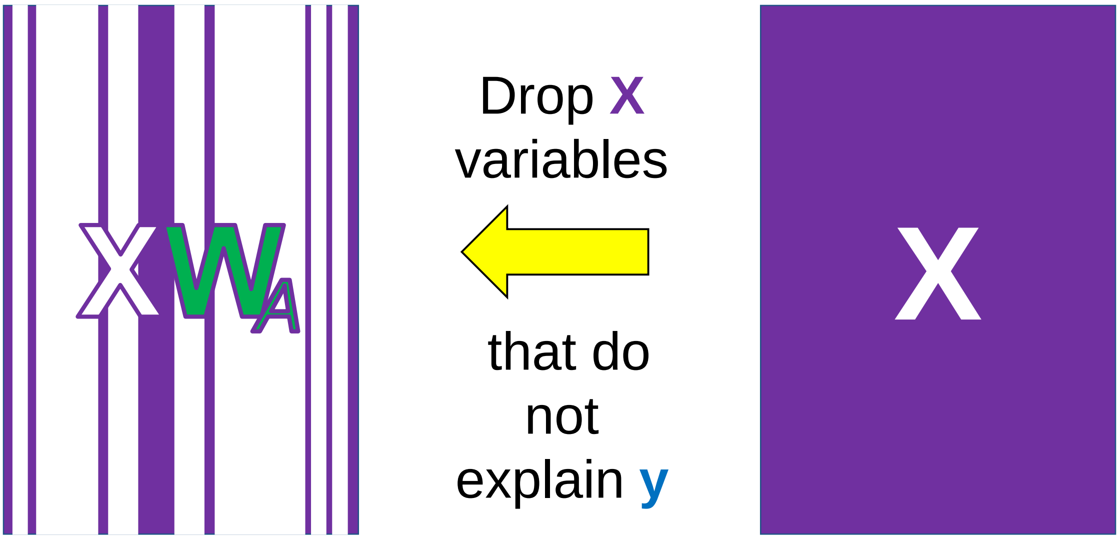


Supervised principal component regression (SPCR)

We want to ***explain*** and ***predict*** the response y from the multivariate X.

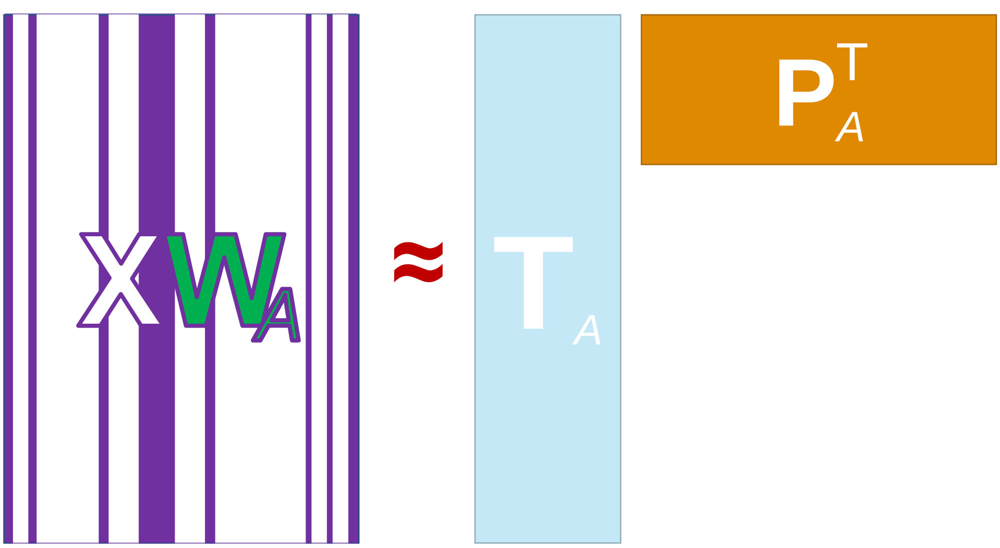
Response y is regressed on principal components of X.

Supervised principal component regression (SPCR)



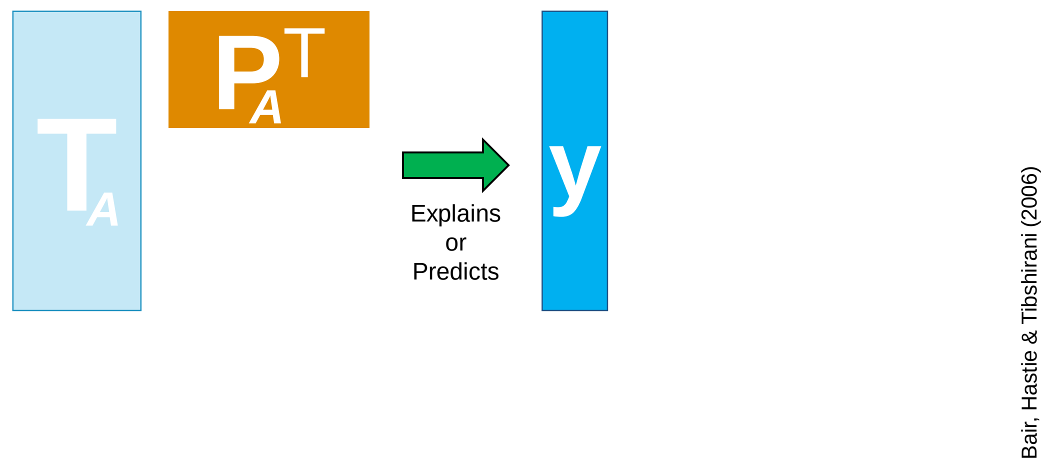
Bair, Hastie & Tibshirani (2006). Prediction by supervised principal components. JASA. oi:10.1198/016214505000000628

Dimension reduction in SPCR

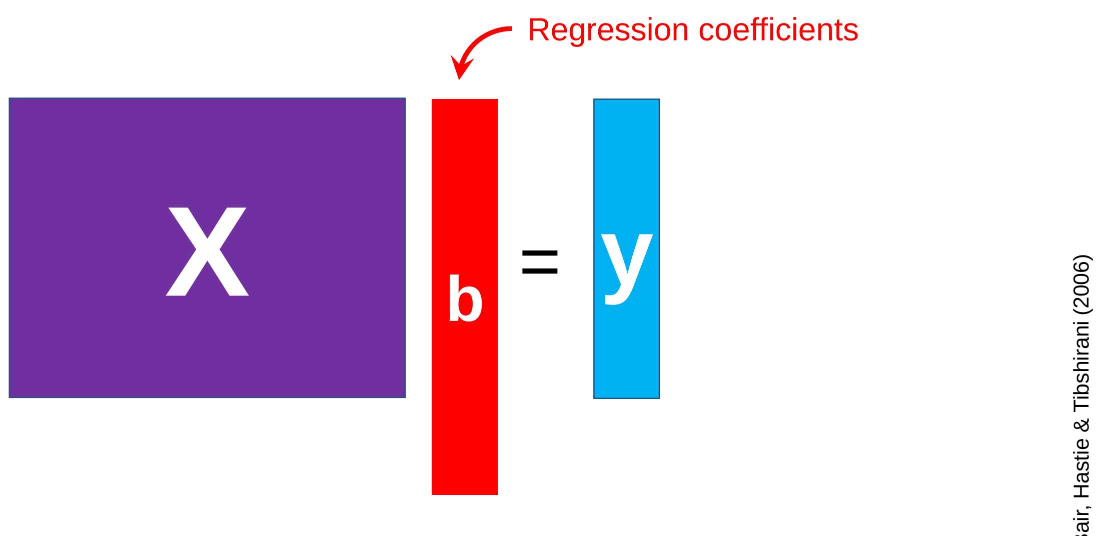


Bair, Hastie & Tibshirani (2006)

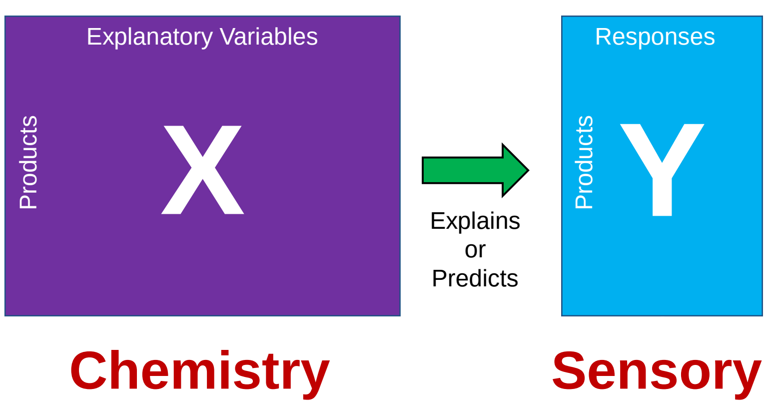
Supervised principal component regression (SPCR)

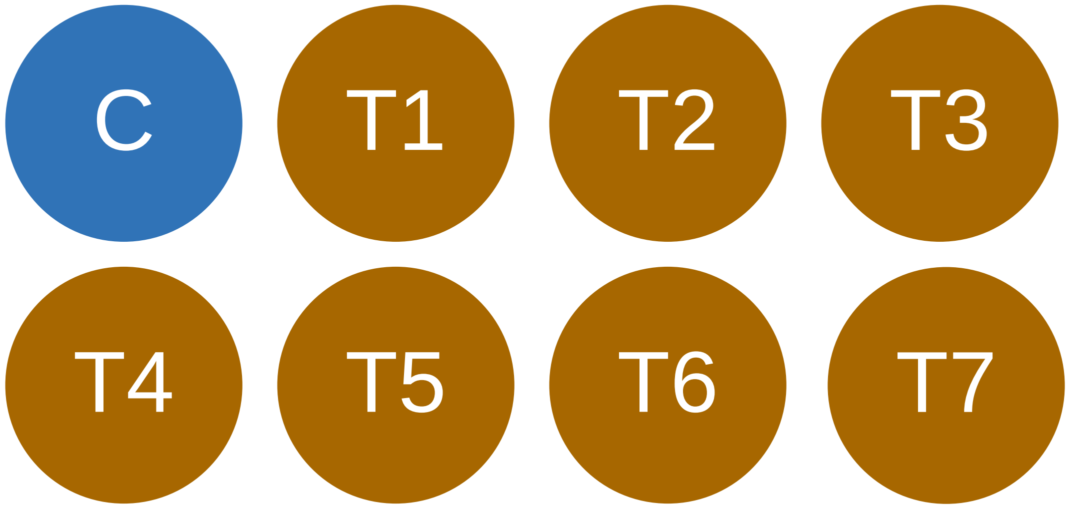


Explaining and predicting data relationships

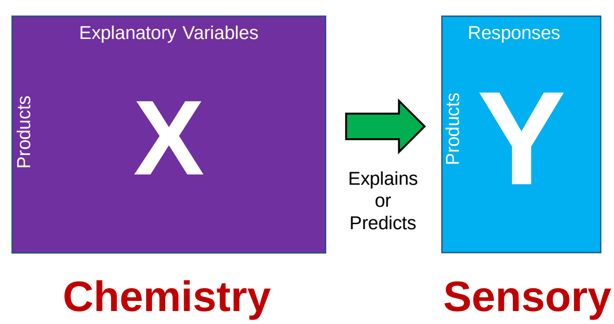


Suppose there are many response variables



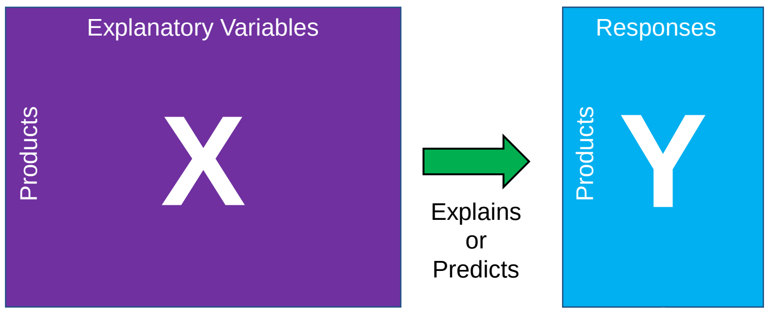


Whatever shall we do?



Partial least squares regression

Explaining and predicting data relationships



Chemistry **Sensory**

Partial least squares regression (PLSR)

We want to ***explain*** and ***predict*** multivariate Y from the multivariate X.

Successive PLS components extract ***covariation*** between X and Y maximally.

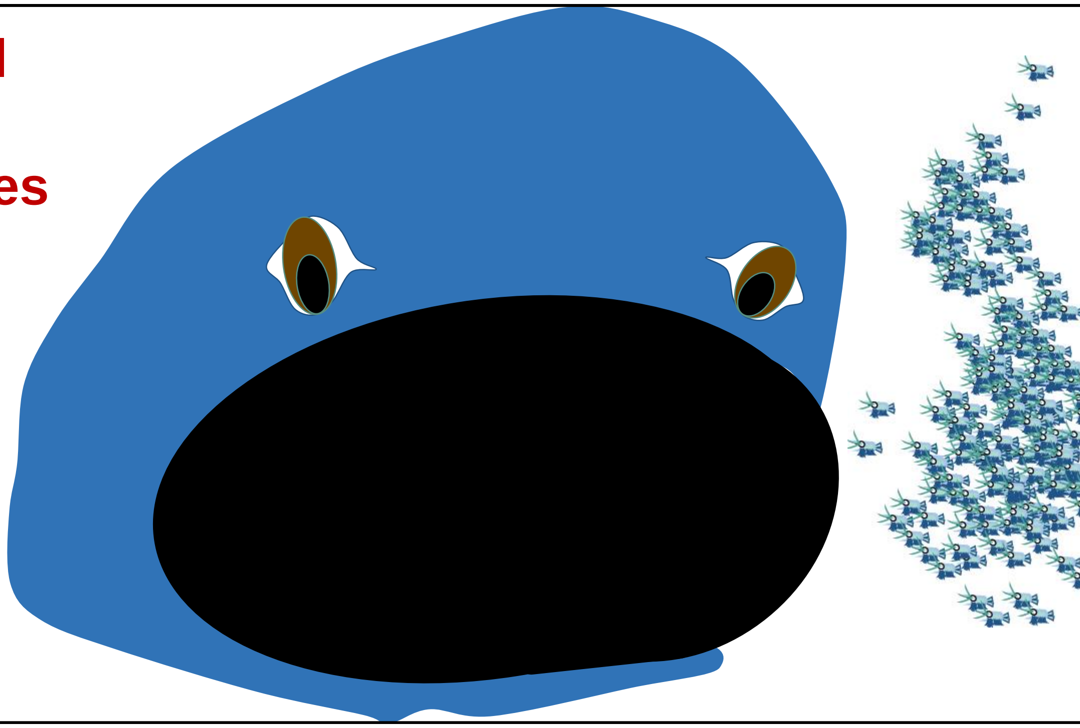
Partial least squares regression (PLSR)

We want to ***explain*** and ***predict*** multivariate Y from the multivariate X.

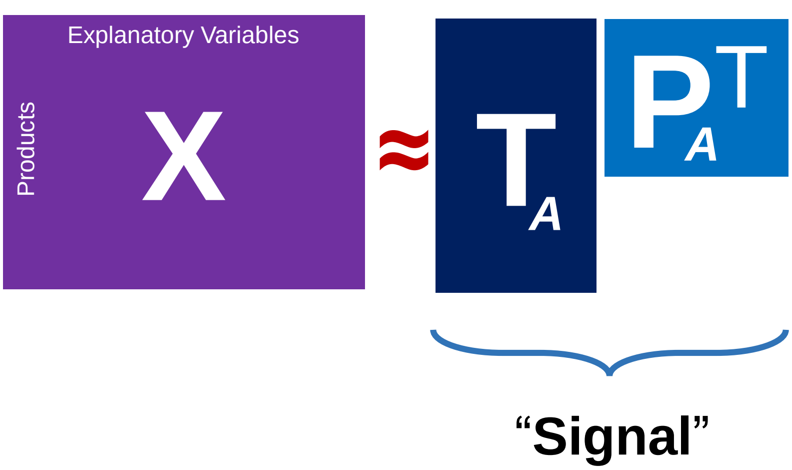
Successive PLS components extract ***covariation*** between X and Y maximally.

PLS component

Partial least squares

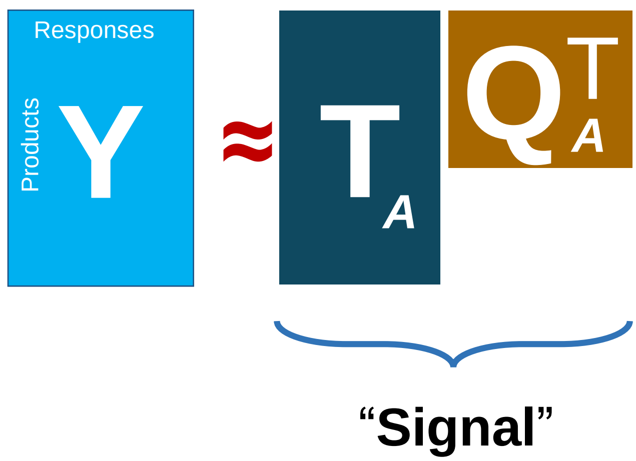


Explaining and predicting data relationships



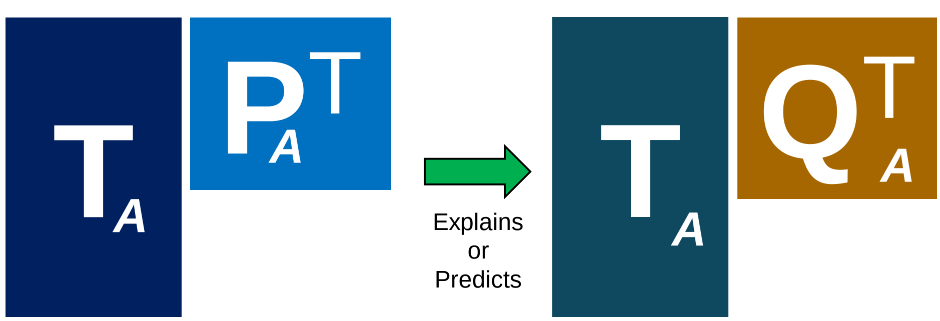
Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In Krishnaiaah, P.R. (ed.). Multivariate Analysis.

Explaining and predicting data relationships



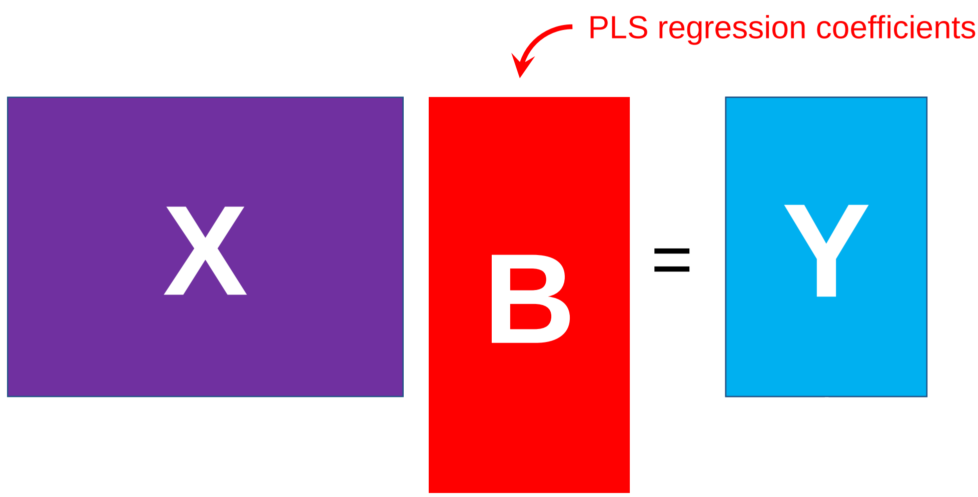
Wold (1966)

Partial least squares regression (PLSR)



Wold (1966)

Explaining and predicting data relationships



Wold (1966)


Whatever shall we do?


What about paired comparisons?
Whatever shall we do?

Investigating paired comparisons

Crossdiff-unfolding*

Subtract every row in **X** from every row in **X**

X

J rows

crossdiff-unfolding →

X⊖**X**

J^2 rows

* See
Castura, J.C., Varela, P., & Næs, T. (2023). Investigating paired comparisons after principal component analysis. *Food Quality and Preference*, 106, 104814. https://doi.org/10.1016/j.foodqual.2023.104814

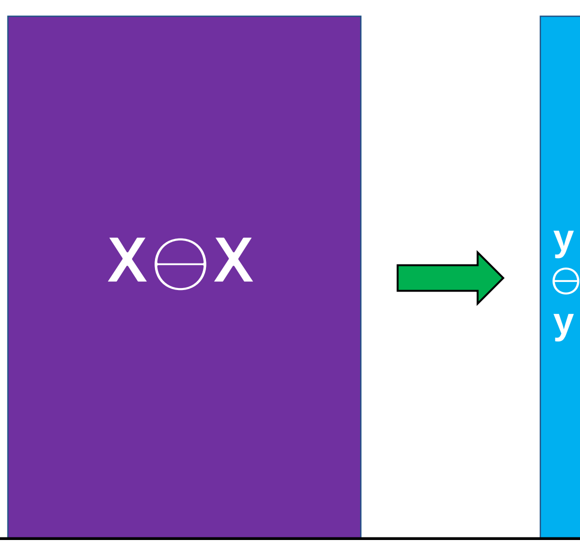
(S)PCR of all paired comparisons

Goal is to ***explain*** and ***predict*** all response paired comparisons from all explanatory paired comparisons.

Regress $\mathbf{y} \ominus \mathbf{y}$ on $\mathbf{X} \ominus \mathbf{X}$.

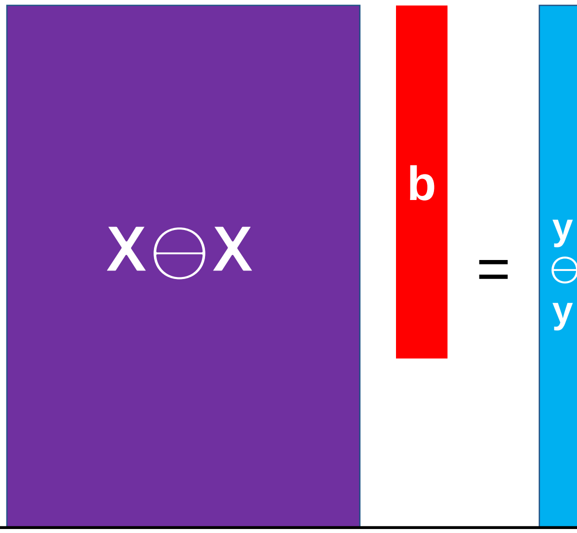
Castura & Tomic (2024)

(S)PCR of all paired comparisons



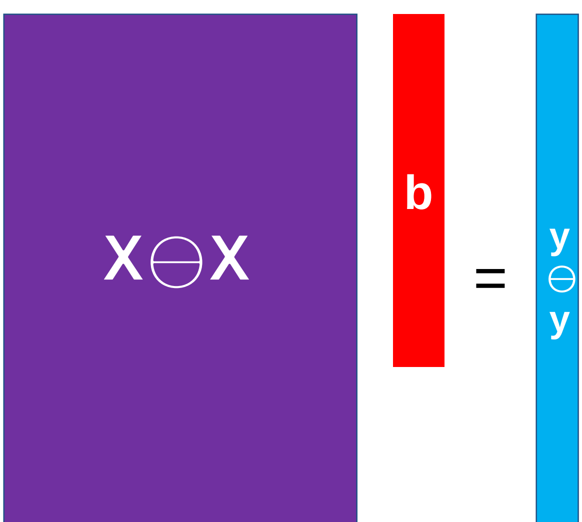
Castura & Tomic (2024)

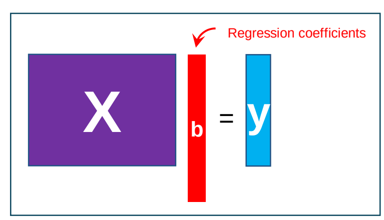
(S)PCR of all paired comparisons



Castura & Tomic (2024)

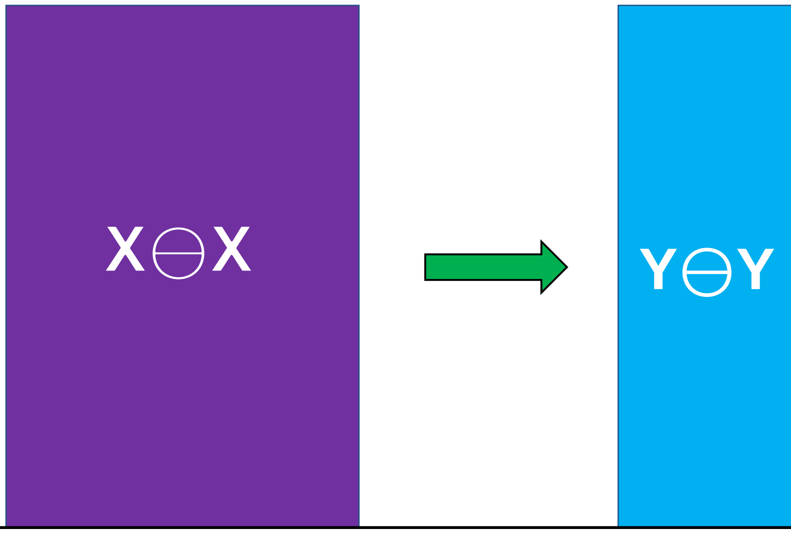
(S)PCR of all paired comparisons





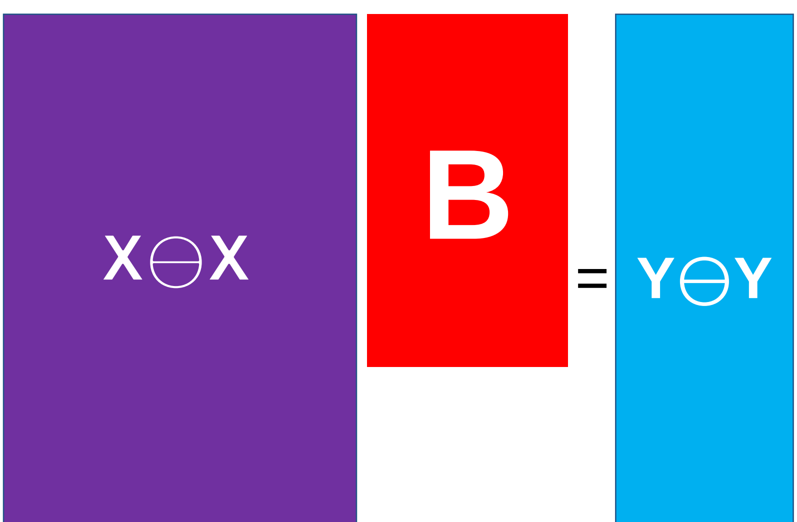
Castura & Tomic (2024)

PLSR of all paired comparisons



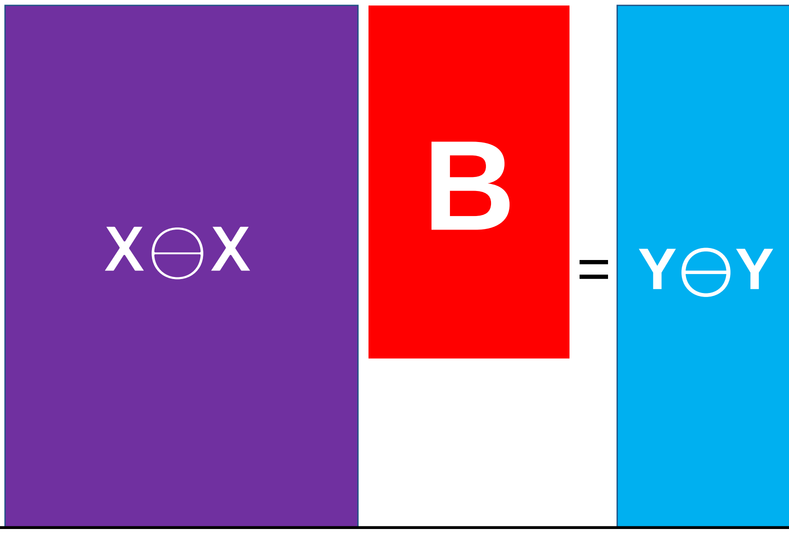
Castura, Tomic & Næs (2024)

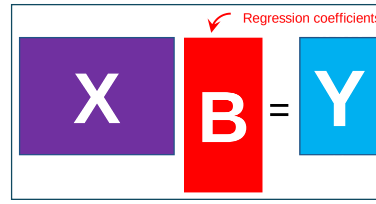
PLSR of all paired comparisons



Castura, Tomic & Næs (2024)

PLSR of all paired comparisons





Castura, Tomic & Næs (2024)

A subset of paired comparisons

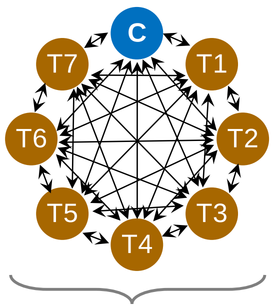
If a subset of paired comparisons is of primary interest, then…

we want to focus on this subset which contains the *relevant* variation

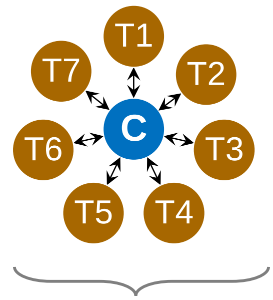
Castura & Tomic (2024)

All pairs vs. a subset of paired comparisons

All Pairs | **Test-Control Pairs**



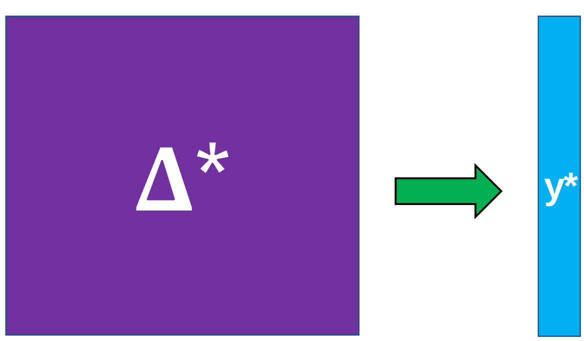
28 paired comparisons
56 paired differences



7 paired comparisons
14 paired differences

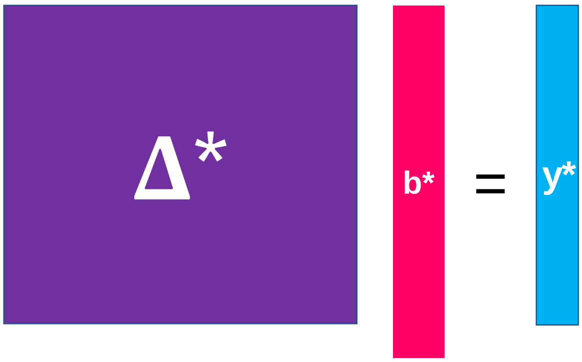
Castura, J.C., Varela, P., & Næs, T. (2023). Investigating only a subset of paired comparisons after principal component analysis. *Food Quality and Preference*, 110, 104941. https://doi.org/10.1016/j.foodqual.2023.104941

(S)PCR of a subset of paired comparisons



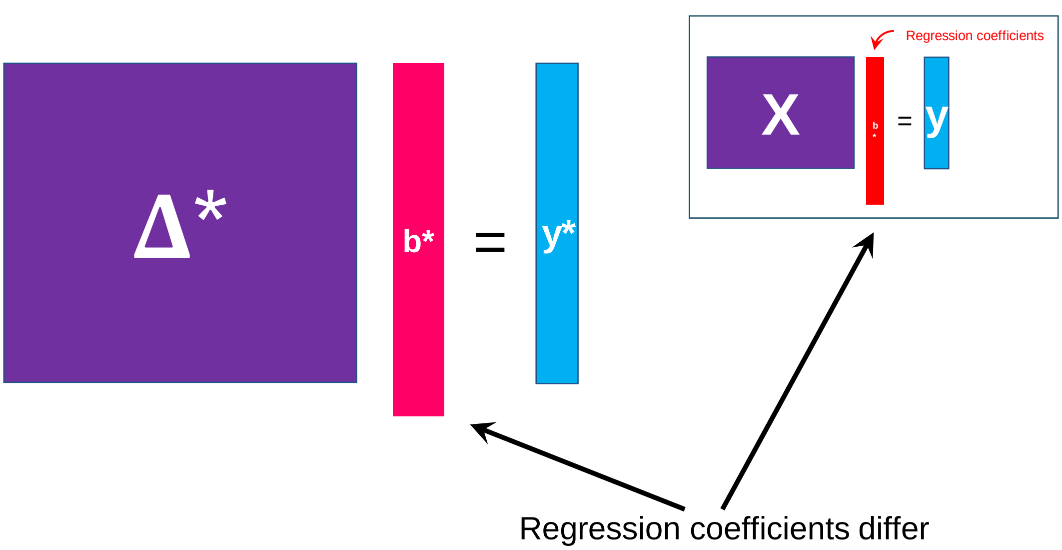
Castura & Tomic (2024)

(S)PCR of a subset of paired comparisons



Castura & Tomic (2024)

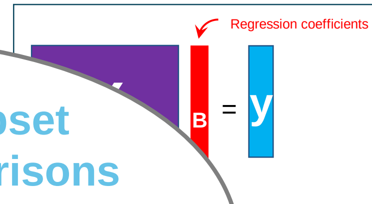
(S)PCR of a subset of paired comparisons



Castura & Tomic (2024)

Castura, J.C. (2024). Investigating relationships in sensory and instrumental data using component-based methods. *Society of Sensory Professionals 2024 Conference.* 2-4 October. Pittsburgh, PA, USA. (Oral).

(S)PCR of a subset of paired comparisons



SPCR of a subset of paired comparisons always explains and often predicts these paired comparisons better than conventional PCR

Regression coefficients differ

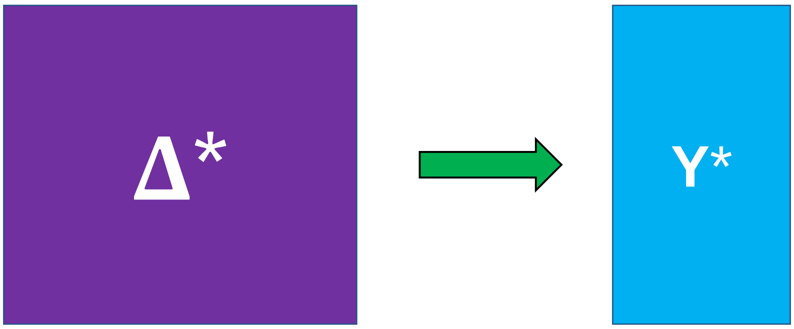
Castura & Tomic (2024)

SPCR of a subset of paired comparisons

Focus on relevant variables / columns

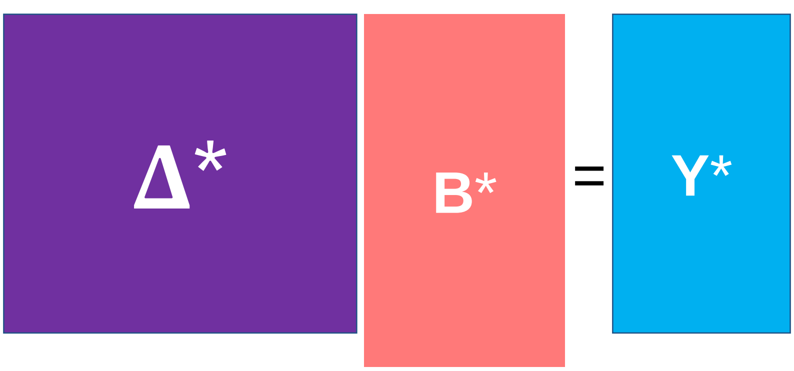
Focus on relevant paired comparisons / rows

PLSR of a subset of paired comparisons

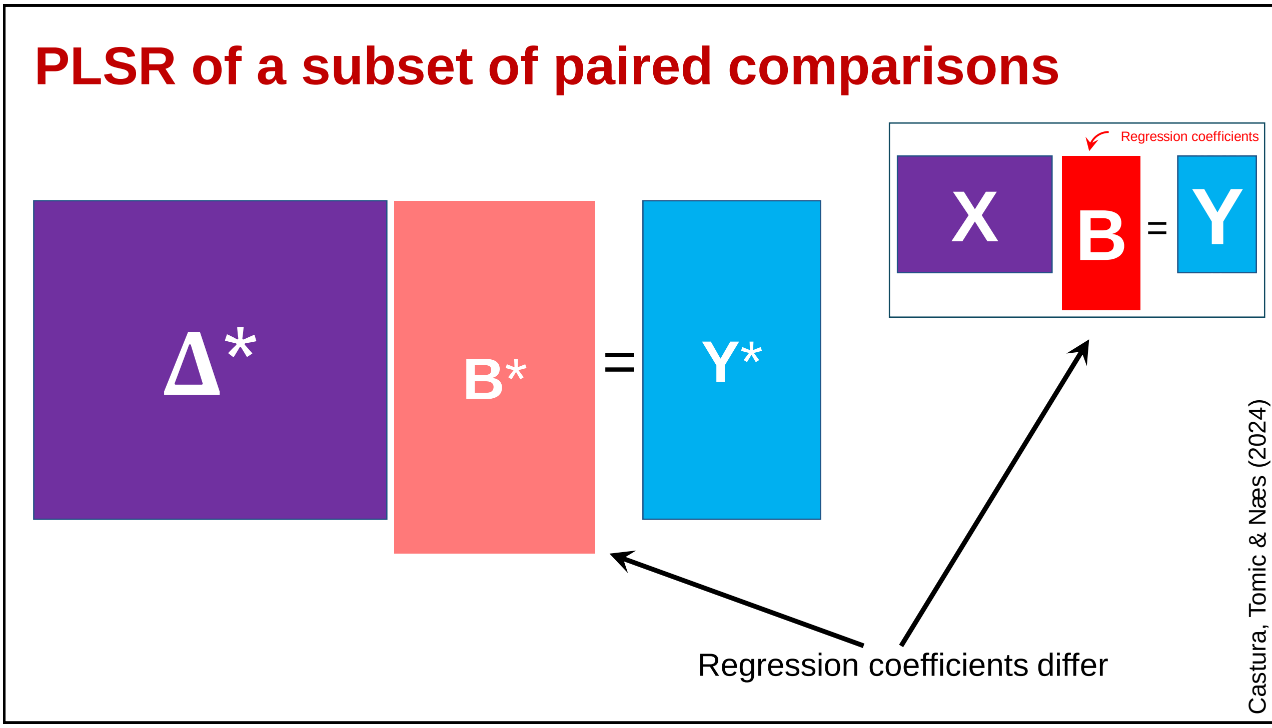


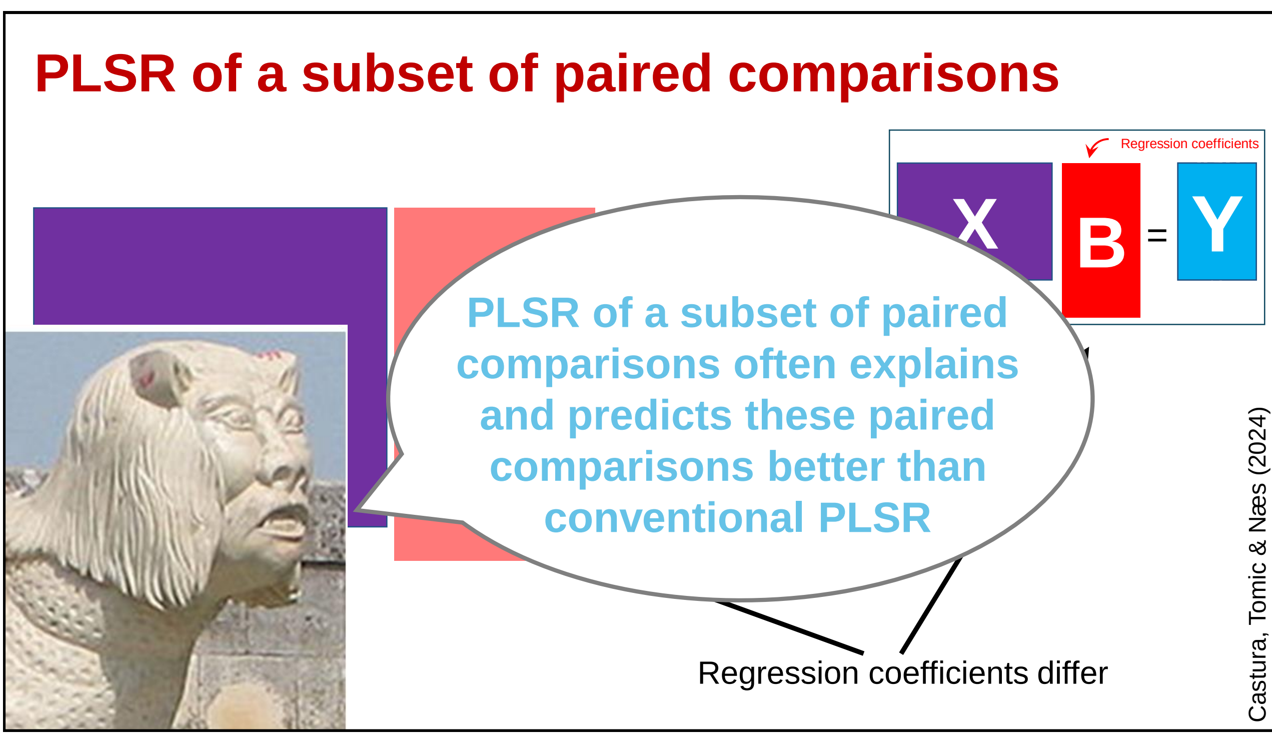
Castura, Tomic & Næs (2024)

PLSR of a subset of paired comparisons



Castura, Tomic & Næs (2024)


PLSR of a subset of paired comparisons
Regression coefficients
X
B
=
Y
Δ*
B*
=
Y*
Regression coefficients differ
Castura, Tomic & Næs (2024)


PLSR of a subset of paired comparisons
Regression coefficients
X
B
=
Y
PLSR of a subset of paired comparisons often explains and predicts these paired comparisons better than conventional PLSR
Regression coefficients differ
Castura, Tomic & Næs (2024)

References from 2024

- Castura, J.C., & Tomic, O. (2024). Supervised principal component regression of select paired comparisons. *Manuscript under review. To be presented at 17th International Weurman Flavour Research Symposium. 24-27 September 2024. Wageningen University & Research, Wageningen, The Netherlands.*
- Castura, J.C., Tomic, O., & Næs, T. (2024). Partial least squares regression of select paired comparisons. *16th AgroStat Conference. 3-6 September 2024. Bragança, Portugal.*
- Castura, J.C., Cariou, V., & Næs, T. (2024). Investigating control-centred results after uncentred principal component analysis. *Zenodo* (preprint). https://doi.org/10.5281/zenodo.11496201

References from 2023

- Castura, J.C., Varela, P., & Næs, T. (2023). Investigating paired comparisons after principal component analysis. *Food Quality and Preference*, 106, 104814. https://doi.org/10.1016/j.foodqual.2023.104814
- Castura, J.C., Varela, P., & Næs, T. (2023) Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis. *Food Quality and Preference*, 107, 104843. https://doi.org/10.1016/j.foodqual.2023.104843
- Castura, J.C., Varela, P., & Næs, T. (2023). Investigating only a subset of paired comparisons after principal component analysis. *Food Quality and Preference*, 110, 104941. https://doi.org/10.1016/j.foodqual.2023.104941
- Næs, T., Varela, P., Castura, J.C., Bro, R., & Tomic, O. (2023). Why use component-based methods in sensory science? *Food Quality and Preference*, 112, 105028. https://doi.org/10.1016/j.foodqual.2023.105028

John Castura

Compusense

Acknowledgements…

Oliver Tomic	Paula Varela	Tormod Næs

Norwegian University of Life Sciences

Nofima

For further information, please contact **jcastura@compusense.com**