

## Statistical Engineering: de la pràctica a la teoria

Xavier Tort-Martorell

7 de març de 2019



### Statistical Engineering

#### La meva (i la de molts altres estadístics) activitat professional favorita:

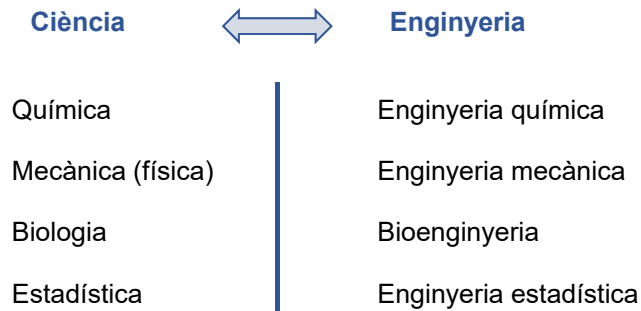
Resoldre problemes complexos fent servir els conceptes mètodes i eines estadístiques juntament amb els d'altres disciplines i per suposat coneixements sobre el tema en qüestió

- Hi ha un mètode o cada “maestrillo”...?
- Hi ha “librillos” millors que d'altres?
- El “librillo” depèn del tema?
- S'aprèn per experiència?
- Es pot ensenyar com fer-ho?
- Això és estadística aplicada?
- ...



## Ciència Vs. Enginyeria

**Enginyeria:** Conjunt de les tècniques que permeten l'aplicació dels coneixements científics (*Institut d'estudis Catalans*)



## Exemples

hp

**Problema:** 4000 impressores de gran format enviant informació. Detectar impressores que requereixen atenció de l'equip de suport

**Background:** moltes variables amb freqüències diferents, dades afectades pel comportament del client (manteniment), dades molt "brutes". Molts tipus d'error, diferents criteris sobre quan requereixen atenció, diferents dashboards, diferents grups dins de hp involucrats.

**Team:** UPC-Datancia, R&D, Engineering, Sales, IT

**Eines:** EDA, Clustering, PCA, Multivariate SPC, Bayesian modeling

**Resultats:** Dos algoritmes en R, implementació en marxa (requereix suport, coordinació, política...)



## Examples

### P&G

**Problem:** Setting fill weight targets on products that have a label net content declaration

**Background:** Various ways for calculating fill weight targets, No way currently to quantify the risk of failing a government inspection, overfill is very costly and calculated in many different ways.

**Team:** Engineering, Manufacturing, Quality Assurance, Regulatory, Statistics. Several sites around the world involved

**Tools used:** Distribution theory with complex assumptions, simulations, Restricted Maximum Likelihood techniques, variance components, DOE

**Results:** A target setting standard process and a soft (AccuTarget™). Solution embedded in work processes



## Examples

### Merk

**Problem:** Potential disruption of several vaccine products due to unpredictable trends in potency of an intermediate product

**Background:** High impact in many senses, very complex process and testing, many people involved: operation personal, process engineers, R&D and scientific experts involved

**Tools used:** control charts (SPC), CUSUM analysis, time series modelling, variance components, simulation

**Results:** Variability reduced to 2/3 of original value, same approach can be used in other processes



## Examples

### Lubrizol

**Problem:** Taylor products to technical needs and demands of customers

**Background:** Since 1990 Lubrizol statisticians have developed models to assist chemists and engineers in determining chemical recipes that satisfy customers demands and pass legal tests

**Tools used:** IT, databases integration, Bayesian modelling algorithms, SPC applied to control models, business protocols

**Results:** A complex predictive trademark registered software (Q.LIFE) used internally and (some parts) externally



## Problemes complexos, rellevants i poc estructurats

- Impacte elevat
- Diversos departaments, grups i funcions involucrats
- Reptes tècnics i d'altres tipus (merder, equip, egos, historia...)
- Requereix dades de diverses fonts (disponibilitat, qualitat...)
- S'avança a "empentes"
- Tècniques estadístiques i d'altres tipus, IT
- Implementar i estandarditzar

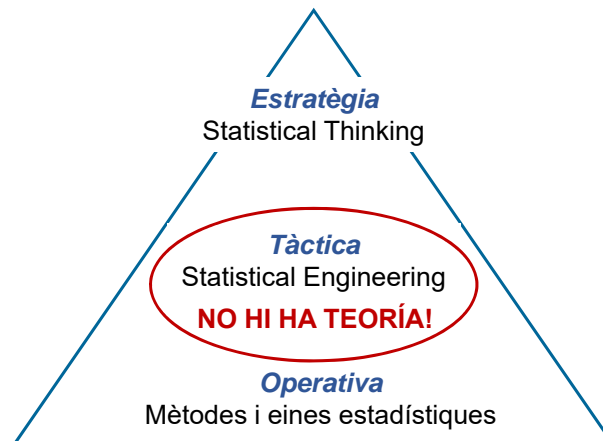
### Exemples:

- Reduir stocks i augmentar la satisfacció del client (t i fiabilitat dels lliuraments)
- Seguretat ciutadana
- Temps efectiu d'utilització dels quiròfans
- Definició del cost d'una assegurança



## Visió sistèmica de l'estadística

Com s'han d'atacar els problemes complexos, rellevants i poc estructurats?



© TSG-EIO, UPC

9



## Definició

La enginyeria estadística es la **disciplina** que estudia com **sistematitzar** la **integració** dels conceptes, mètodes i eines estadístiques, sovint amb les d'altres disciplines per resoldre **problemes importants** de forma sostenible

La idea no es nova:

- Box, Hogg, Cox, Tukey, Hunter, Joiner and many other

Però ha arribat el moment en el que és urgent:

- Meng, X. (2009) "Desired and feared- what to do now and over the next 50 years?"
- ASA (2014) Guidelines for the design of undergraduate statistics programs

© TSG-EIO, UPC

10



## Antecedents

### Six Sigma (Entre d'altres. *Clinical trials?*)

- Organització: rols ben definits
- Metodologia: DMAIC (Define, Measure, Analyze, Improve, Control)
- Eines estadístiques i no estadístiques en seqüència

### Factors d'èxit (llicions)

- Lideratge i compromís
- Project Management
- Equips interdepartamentals i interdisciplines (Team Dynamics)
- Metodologia basada en el Mètode científic
- Dades (mètodes i eines estadístiques seqüenciats)



## Desenvolupament (in progress)

### Principis fonamentals de l'enginyeria estadística

- Entendre el context del problema
- Definir l'estratègia a seguir per resoldre'l
- Vigilar la qualitat de les dades
- Integració amb els coneixements existents (domain knowledge)
- Utilitzar un enfocament seqüencial



## Context

- Com es farà servir la solució?
- Com s'ha arribat a la situació actual?
- Quines solucions s'han intentat abans? Perquè no van funcionar?
- Quins aspectes "polítics estan involucrats" ( i sovint amagats)?
- L'objectiu, en general, és una "solució útil", no "optima" (Millor solució practica, no millor solució estadística!)

Hi ha raons perquè el problema no s'hagi resolt



## Estratègia

- Fent servir les eines de manera lligada i seqüencial (i potser innovadora) millora la efectivitat, l'aprenentatge i l'impacte  
Rara vegada els problemes complexos es poden resoldre amb un sol mètode
- Abordar els problemes pensant en les eines rara vegada funciona  
Analogia del martell i els claus
- Els problemes senzills requereixen l'aplicació de l'eina adient; els problemes complexos requereixen una estratègia  
La estratègia és diferent per cada problema

Els problemes complexos requereixen una estratègia



## Qualitat de les dades

- Observacionals Vs. Experimentals
- No totes les dades son “creades” iguals!
- Versions, modificacions, filtres, neteges, estimacions, alteracions...
- Cal entendre el procés que ha creat les dades
- Els models mai han de ser més complexes que les dades que els suporten
- S’han de documentar les limitacions dels anàlisis i models degudes a la qualitat de les dades

Les dades son “culpables mentre no es demostrí el contrari”



## Integració amb els coneixements existents

- Imprescindible per tal de que els resultats siguin aplicables
- Aquests coneixements són útils tan al recollir les dades com per interpretar els resultats dels anàlisis i aprendre seqüencialment
- L'estadística es inútil si està divorciada dels coneixements sobre el tema en qüestió

L'única raó per l'existència de la estadística com a disciplina és la millora d'altres disciplines (química, enginyeria, psicologia, economia, etc.)

La formació de base de quasi tots els pioners de l'estadística era en enginyeria o ciències





## Enfocament seqüencial

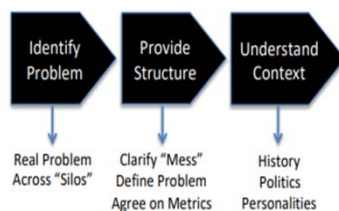
- Les aplicacions estadístiques s'han de concebre com a part de l'aprenentatge resultat d'aplicar el mètode científic i no com estudis aïllats
- La seqüencialitat permet desenvolupar noves teories i adquirir nous coneixements i no merament testejar hipòtesis existent
- Guiar estudis futurs és, sovint, l'aspecte més beneficiós de l'anàlisi de dades observacionals
- L'enfocament seqüencial encaixa perfectament amb tenir una estratègia general per abordar els problemes

El coneixement no es genera a base d'estudis aïllats



## Desenvolupament (in progress)

### Full de ruta per atacar problemes complexos, rellevants i poc estructurats



- Oportunitats Vs. Problemes
- Rellevant (alt impacte)
- Objectiu / abast
- Mètriques/ restriccions
- Stakeholders / historia
- Fonts de dades

**Desenvolupament (in progress)**

**Full de ruta per atacar problemes complexos, rellevants i poc estructurats**

The diagram illustrates a process flow. On the left, two black arrow-shaped boxes represent 'Develop Strategy' and 'Develop & Execute Tactics'. Below 'Develop Strategy' are the sub-points: 'How to Attack', 'Sequential Approach', and 'Use Core Processes'. Below 'Develop & Execute Tactics' are: 'Identify Alternatives', 'Select Methods & Apply'. On the right, a cycle of four rounded rectangular boxes is shown: 'Aprentatge (inferència)' at the top left, 'Recollida de dades' at the top right, 'Anàlisi exploratori' at the bottom right, and 'Modelat' at the bottom left. Arrows connect them in a clockwise cycle. Above this cycle is a box labeled 'Pla d'acció'. Below the cycle, the text 'Core processes' is written in red.

© TIG-IBO, UPC

19

**Desenvolupament (in progress)**

**Full de ruta per atacar problemes complexos, rellevants i poc estructurats**

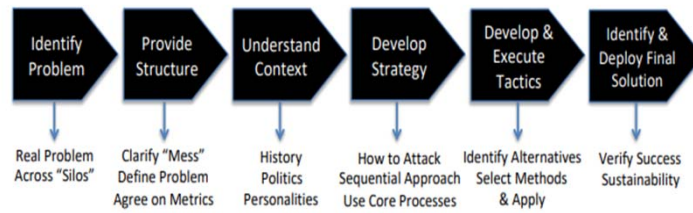
The diagram shows a single black arrow-shaped box labeled 'Identify & Deploy Final Solution'. Below it, an arrow points to the text 'Verify Success Sustainability'.

- Identificar solucions
- Comprovar
- Implementar / resistència al canvi
- Garantir continuïtat

© TIG-IBO, UPC

20

### Full de ruta per atacar problemes complexos, rellevants i poc estructurats



	FASES					
	Identificar oportunitat	Estructurar	Entendre el context	Desenvolupar estratègia	Desenvolupar i executar tàctica	Identificar i implantar solució
<b>Organització</b>	xxx	xx	xx	x	x	xx
<b>Human factors</b>	x		xxx	x	x	xxx
<b>Estadística</b>	x	x	x	xxx	xxx	xx
<b>IT</b>		x		xx	xxx	xx
<b>Altres tecnologies</b>	x	x		xx	xxx	x

ACTORS



## Summary

### Problemes complexos, rellevants i poc estructurats

- Un enfocament “enginyeril” sembla adequat per atacar aquests problemes
- L’actual sistema d’ensenyament de l’estadística no prepara per afrontar-los
- L’enginyeria estadística proporciona un marc per accelerar la corba d’aprenentatge per fer-ho
- Actualment s’està desenvolupant la disciplina (teoria i pràctica)
- Dona resposta a molts reptes actuals de l’estadística

Esperem que tingui un impacte significatiu en l’ensenyament, la recerca i la praxis de l’estadística



## ENBIS Spring Meeting about SE

### ENBIS Spring Meeting 2019

13 – 14 June 2019

Abstract submission: 15 January – 15 April 2019

[Welcome](#) [About the meeting](#) [Location and venue](#) [Travel and hotels](#) [Abstracts](#) [Registration](#)

The 2019 ENBIS Spring Meeting will be organized together by ISEA, the International Statistical Engineering Association ([www.isea-change.org](http://www.isea-change.org)) and ENBIS. It will take place on June 13th and 14th, 2019, in Barcelona (Spain), hosted by UPC|BarcelonaTech.



[WWW.ENBIS.ORG](http://WWW.ENBIS.ORG)



## ENBIS Spring Meeting about SE

	Thursday, June 13	Friday, June 14
9:00 to 10:45	What is SE, why it is needed, why it is the right time for it. What's in it for statisticians. What are the benefits for everybody: society, companies, etc. Implications for statistical training	How SE is different from: Applied statistics and data science. Simple and well-known examples of SE like DMAIC (and why it is not enough), or clinical trials. How the idea originated and evolved. Lessons from examples and how they are the basis for a new science.
11:15 to 12:45	SE cases / examples	What has ISEA done so far and what is planning to do. Development of the SE Book of Knowledge.
12:45 to 13:15	What's the Role of Software in Statistical Engineering?	
	Lunch	
14:30 to 18:00	Contributed Sessions	



## ENBIS Spring Meeting about SE

### Coordinators:

- Lluís Marco i Xavier Tort-Martorell (UPC).

### Presenters:

- Ronald Does (University of Amsterdam)
- Roger Hoerl (Union College)
- Bart de Ketelaere (KU Leuven)
- Volker Kraft and Ian Cox (JMP)
- Murat Kulahci (DTU, Technical university of Denmark)
- Geoff Vining (Virginia Tech)

### Discussants:

- Antje Christensen (Novo Nordisk)
- Alberto Ferrer (Universitat Politècnica de Valencia)
- Laura Freeman (Institute for Defense Analysis)
- Luisa Puerto (Maxam)



## International Statistical Engineering Association (ISEA)

- New Professional Society
- Focus: The Emerging Discipline of Statistical Engineering
- Business Model Based on ENBIS (European Network for Business and Industrial Statistics) (<https://enbis.org/>)
  - Web Based
  - Free Individual Memberships
  - Paid Organizational Memberships
- <https://isea-change.org>



## Statistical Engineering: de la pràctica a la teoria

Xavier Tort-Martorell



UNIVERSITAT POLITÈCNICA DE CATALUNYA  
BARCELONATECH  
Departament d'Estadística  
i Investigació Operativa

7 d'abril de 2019