

A Biomed Data Analyst Training Program

Causality and experimental design

Professor Ron S. Kenett

Chapter 8

Modern Statistics: A Computer Based Approach with Python
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (September 15, 2022)
ISBN-13: 978-3031075650

(c) 2022 Ron Kenett, Shelemyahu Zacks, Peter Gedeck

The code needs to be executed in sequence.

```
In [1]: import os
os.environ['OUTDATED_IGNORE'] = '1'
import warnings
from outdated import OutdatedPackageWarning
warnings.filterwarnings('ignore', category=FutureWarning)
warnings.filterwarnings('ignore', category=OutdatedPackageWarning)
```

Modern analytic methods: Part II

```
In [2]: import networkx as nx

import statsmodels.api as sm
from statsmodels.tsa.stattools import grangercausalitytests
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import mistat
```

[Modern Analytic Methods: Part II](#)

Ron Kenett, Shelemyahu Zacks, Peter Gedeck
Pages 395-419

Chapter 8

Modern Analytic Methods: Part II



8.4 Causality Models

Causality analysis has been studied from two main different points of view, the “probabilistic” view and the “mechanistic” view. Under the probabilistic view, the causal effect of an intervention is judged by comparing the evolution of the system when the intervention is and when it is not present. The mechanistic point of view focuses on understanding the mechanisms determining how specific effects come about. The interventionist and mechanistic viewpoints are not mutually exclusive. For example, when studying biological systems, scientists carry out experiments where they intervene on the system by adding a substance or by knocking out genes. However, the effect of a drug product on the human body cannot be decided only

Conservation of Momentum

The *Law of Conservation of Momentum* states that the total linear momentum (**P**) of a closed system is constant. That means that the momentum of the balls on impact equals the momentum of the second group of balls after impact:

$$\mathbf{P} = m\mathbf{v} = M\mathbf{V}$$

where

- **m** is the total mass of the balls being released
- **v** is the velocity on impact
- **M** is the total mass of the balls moved on the other end
- **V** is the velocity of the second group of balls after impact

Conservation of Energy

The *Law of the Conservation of Energy* states that the total kinetic energy (**KE**) of a system with no external forces acting on it remains constant. That means that the kinetic energy of the moving ball or balls upon impact equals the kinetic energy of the balls leaving the other side of the row of balls.

Note: Although the force of gravity is an external force, it is only a factor in accelerating the first set of balls and in the swing of the second set. The force of gravity is not a factor at the point of impact and with the resulting motion.

$$\mathbf{KE} = m\mathbf{v}^2/2 = M\mathbf{V}^2/2$$

Mass and velocity of balls moved

In the momentum equality, solve for v and square both sides of the equation:

$$mv = MV$$

$$v = MV/m$$

$$v^2 = M^2V^2/m^2$$

Substitute this value of v^2 into the energy equation $mv^2/2$:

$$mv^2/2 = mM^2V^2/2m^2 = M^2V^2/2m$$

Since $mv^2/2 = MV^2/2$:

$$MV^2/2 = M^2V^2/2m$$

$$1 = M/m$$

Thus:

$$M = m$$

This means the mass of the balls leaving equals the incoming mass. Since the balls are of equal mass, that means the same number of balls leave the series as those which impacted the group of balls.

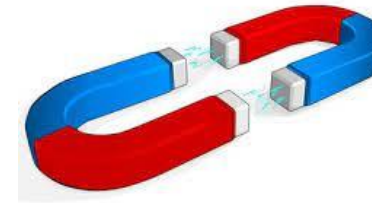
Also, since $M = m$, then:

$$mv^2/2 = mV^2/2$$

and

$$V = v$$

The final velocity of the end balls is the same as the initial velocity.



shutterstock.com · 526441027

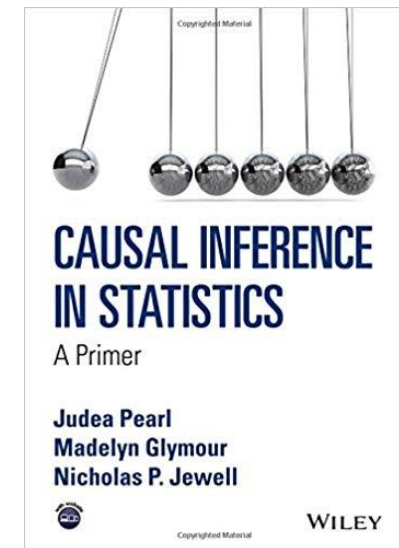


Feynman: Magnets FUN TO IMAGINE 4/ NEW updated higher quality version!

<https://www.youtube.com/watch?v=Q1IL-hXO27Q>

https://www.school-for-champions.com/science/newtons_cradle_derivation.htm#.ZArRjXZBxnI

<https://www.youtube.com/watch?v=lzS3emevxyo>



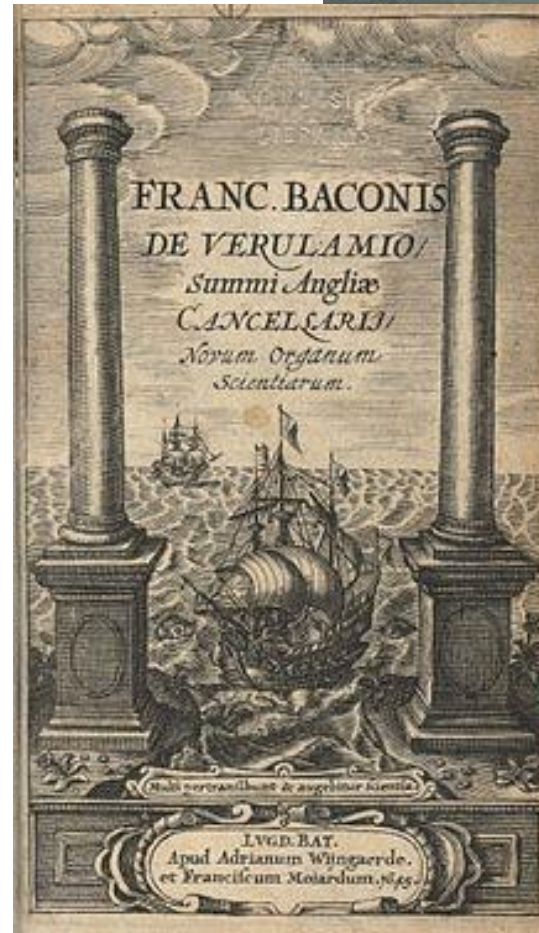
Bacon's scientific methodology can be summarized as follows: 1. The scientist must start with a set of unprejudiced observations; 2. these observations lead infallibly to correct generalizations or axioms; and 3. the test of a correct axiom is that it leads to new discoveries.

https://en.wikipedia.org/wiki/Novum_Organum

*...the true method of experience.
.. first lights the candle, then by means of the candle shows the way; commencing as it does with experience duly ordered and digested, not bungling or erratic, and from it educing axioms, and from established axioms again new experiments. .*

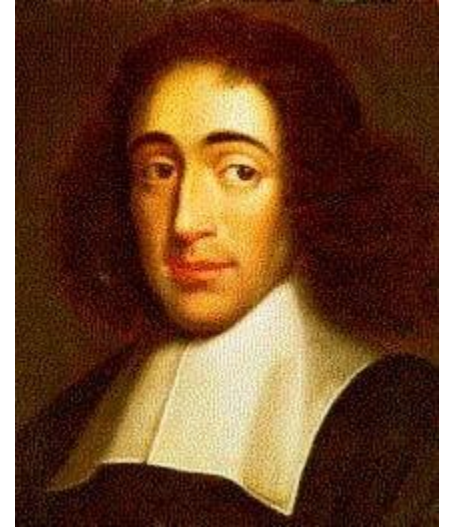


Sir Francis Bacon
1561 - 1626



Baruch Spinoza (1632—1677)

<https://www.youtube.com/watch?v=y9uUBxdWO7E>



Causal Determinism

Spinoza ultimately conceives of the relation between infinite and finite modes, he is clear about one thing – the system of modes is an entirely deterministic system in which everything is fully determined to be and to act:

IP29: *In nature there is nothing contingent, but all things have been determined from the necessity of the divine nature to exist and produce an effect in a certain way.*

Causal Parallelism

Causal relations exist only among modes falling under the same attribute. His explanation for this may be traced back to an axiom set forth at the beginning of Book One: IA4: *The knowledge of an effect depends on, and involves, the knowledge of its cause.*

Given this axiom, if a finite mode falling under one attribute were to have God as its cause insofar as he is considered under a different attribute, i.e., if it were to be caused by a finite mode falling under a different attribute, then the knowledge of that mode would involve the knowledge of that other attribute. Since it does not, that mode cannot have God as its cause insofar as he is considered under some other attribute. In other words, it cannot be caused by a finite mode falling under some other attribute. There can be no causal interaction between ideas and bodies. This does not mean that ideas and bodies are unrelated to one another. Indeed, it is one of the best-known theses in the *Ethics* that the lines of causation that run among them are strictly parallel: IIP7: *The order and connection of ideas is the same as the order and connection of things.*

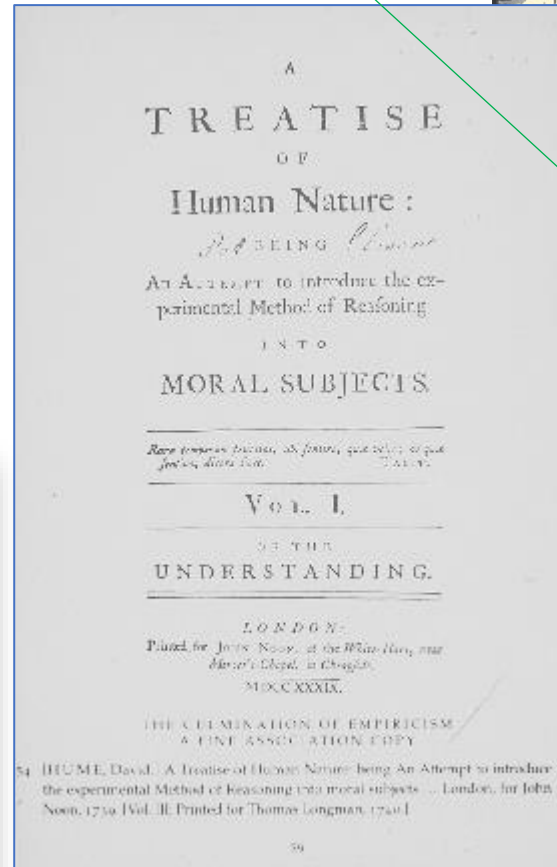
Three Kinds of Knowledge

With the distinction between adequate and inadequate perception, Spinoza introduces a set of further distinctions. He begins with inadequate perception, which he now calls **knowledge of the first kind**, and divides it into two parts. The first consists of **knowledge from random experience** (*experientia vaga*). This is knowledge “from singular things which have been represented to us through the senses in a way which is mutilated, confused, and without order for the intellect”(P40S2). The second consists of **knowledge from signs** (*ex signis*), “for example, from the fact that, having heard or read certain words, we recollect things, and form certain ideas of them, like those through which we imagine the things”(P40S2). What links both of these forms of knowledge is that they lack a rational order. It is obvious that knowledge from random experience follows the order of the affections of the human body, but so does knowledge from signs. A Roman who hears the word ‘*pomum*’, for instance, will think of an apple, not because there is any rational connection between the word and the object, but only because they have been associated in his or her experience. When we reach what Spinoza calls the **second kind of knowledge, reason** (*ratio*), we have ascended from an inadequate to an adequate perception of things. This type of knowledge is gained “from the fact that we have common notions and adequate ideas of the properties of things” (P40S2). What Spinoza has in mind here is the formation of adequate ideas of the common properties of things and the movement by way of deductive inference to the formation of adequate ideas of other common properties. Unlike in the case of knowledge of the first kind, this order of ideas is rational. Spinoza adds a third type, which he regards as superior. He calls this **intuitive knowledge** (*scientia intuitiva*) and tells us that it “proceeds from an adequate idea of the formal essence of certain attributes of God to the adequate knowledge of the [formal] essence of things”(P40S2). He seems to be envisioning a type of **knowledge that gives insight into the essence of some singular thing** together with an understanding of how that essence follows of necessity from the essence of God. Furthermore, the characterization of this kind of knowledge as intuitive indicates that the connection between the individual essence and the essence of God is grasped in a single act of apprehension and is not arrived at by any kind of deductive process. How this is possible is never explained.

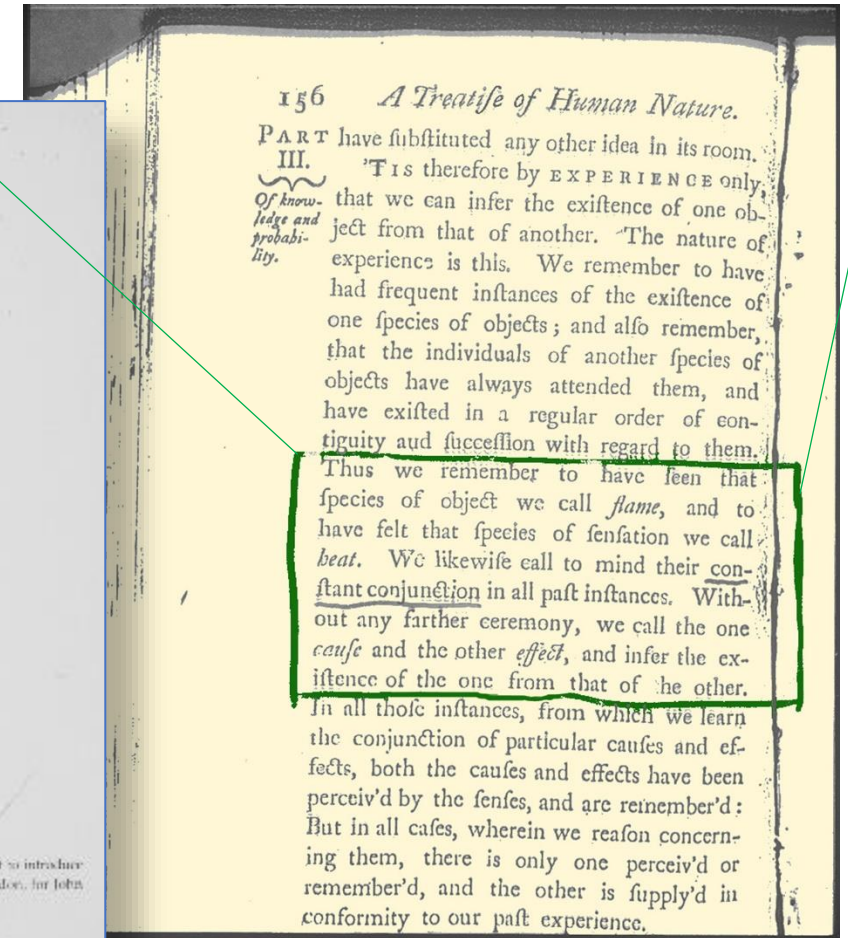
David Hume (1711-1776)

https://en.wikipedia.org/wiki/David_Hume

1. Analytical vs. empirical claims, *the former are product of thoughts, the latter matter of fact*
2. Causal claims are empirical
3. All empirical claims originate from experience.

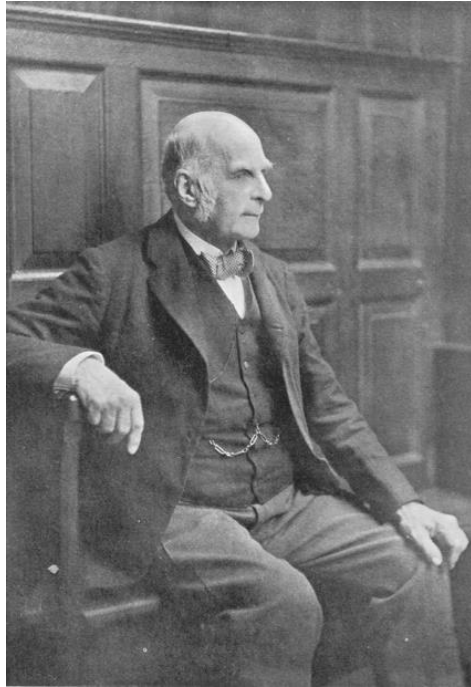


"Thus we remember to have seen that species of object we call *flame*, and to have felt that species of sensation we call *heat*. We likewise call to mind their constant conjunction in all past instances. Without any farther ceremony, we call the one *cause* and the other *effect*, and infer the existence of the one from that of the other."

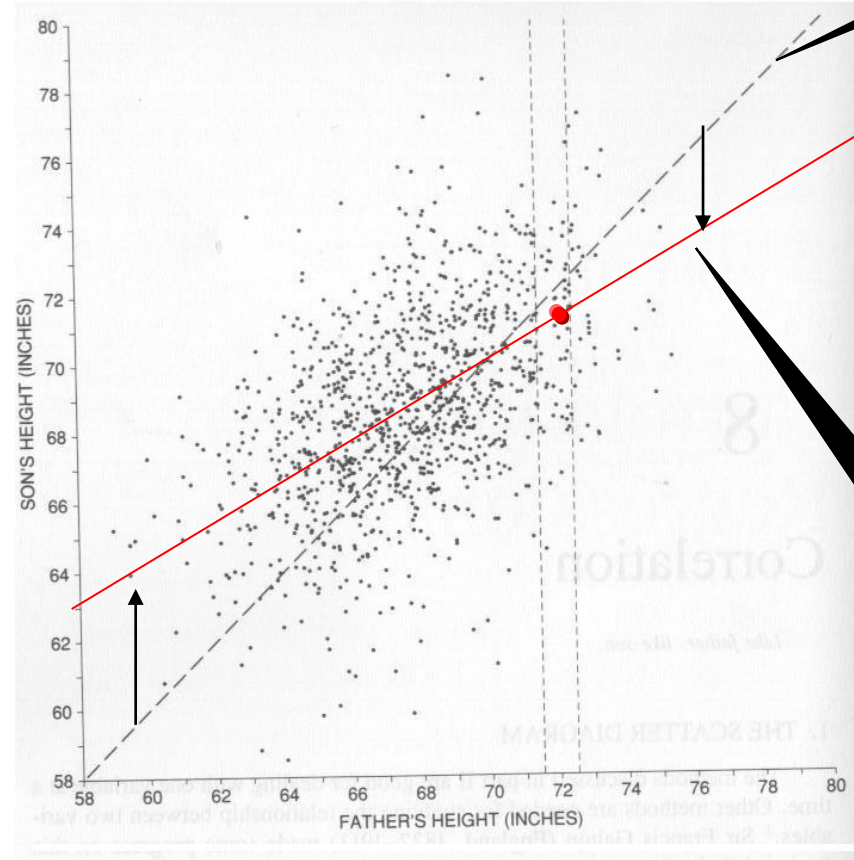


Regression towards the mean....

Equivalence Line



Sir Francis Galton (1822-1911)



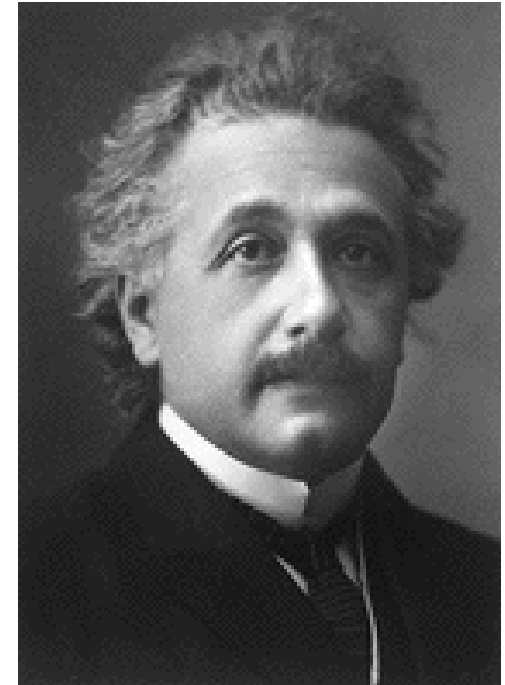
Regression Line

“It is easy to see that consequence of the **co-relation** must be the variation of the two organs being partly due to common causes”

Galton, F. (1886). "Regression towards mediocrity in hereditary stature".
The Journal of the Anthropological Institute of Great Britain and Ireland 15: 246–263

Albert Einstein (1879-1955)

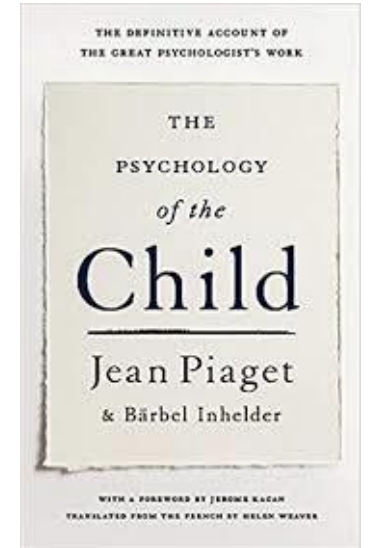
“Development of Western science is based on two great achievements: the invention of the formal logical system (in Euclidean geometry) by the Greek philosophers, and **the discovery of the possibility to find out causal relationships by systematic experiment** (during the Renaissance).”



A. Einstein, April 23, 1953

Jean Piaget (1896 – 1980)

Piaget's (1936) theory of cognitive development explains how a child constructs a mental model of the world. His contributions include a stage theory of child cognitive development, detailed observational studies of cognition in children, and a series of tests to reveal different cognitive abilities.



“The infant’s hand hits a hanging toy. The infant sees it bob about, then repeats the gesture several times, later applying it to other objects as well, developing a striking schema for striking.”

The notion of causality in the infant’s model entails a primitive cause-effect relationship between actions and results. For example if $Z =$ ‘pull string hanging from bassinet hood’ $Y =$ ‘toy shakes’, the infant’s model contains the causal relationship $Z \rightarrow Y$.

W. Edwards Deming (1900-1993)

“Tests of variables that affect a process are useful only if they **predict what will happen if this or that variable is increased or decreased.**

Statistical theory, as taught in the books, is valid and leads to operationally verifiable tests and criteria for an **enumerative study**. Not so with an **analytic problem**, as the conditions of the experiment will not be duplicated in the next trial.

Unfortunately, most problems in industry are analytic.”*



*From preface to *The Economic Control of Quality of manufactured product*
by W. Shewhart, 1931

Cited by

	All
Citations	89932
h-index	98
i10-index	307



Judea Pearl

2011 Turing Award for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning



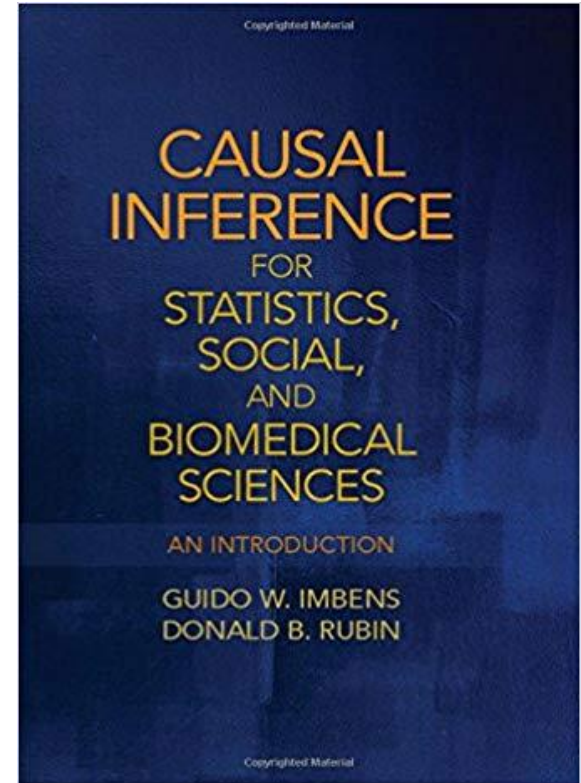
Maximum likelihood from incomplete data via the EM algorithm	55477	1977
AP Dempster, NM Laird, DB Rubin Journal of the royal statistical society. Series B (methodological), 1-38		
Statistical analysis with missing data	24728	2014
RJA Little, DB Rubin John Wiley & Sons		
Bayesian data analysis	22694 *	2004
A Gelman, J Carlin, H Stern, DB Rubin CRC press		
The central role of the propensity score in observational studies for causal effects	21341	1983
PR Rosenbaum, DB Rubin Biometrika 70 (1), 41-55		



Don Rubin

Cited by

	All
Citations	279566
h-index	128
i10-index	364



Guido Imbens



2021

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference*. Cambridge University Press

5,807 views | Apr 18, 2019, 11:53am

Deep Learning And The Limits Of Learning By Correlation Rather Than Causation



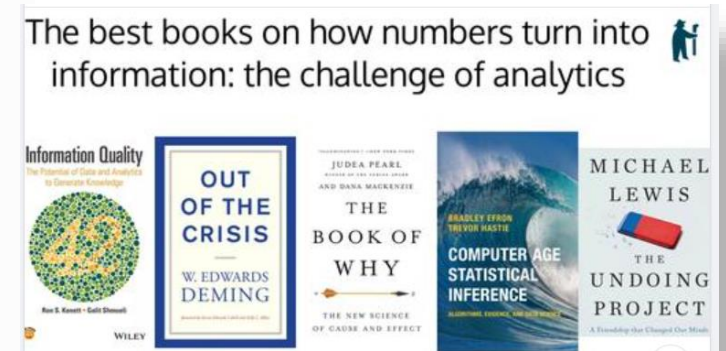
Kalev Leetaru Contributor ⓘ

AI & Big Data

I write about the broad intersection of data and society.

https://www.forbes.com/sites/kalevleetaru/2019/04/18/deep-learning-and-the-limits-of-learning-by-correlation-rather-than-causation/?fbclid=IwAR3MUe_a0cUidgotBtedUXsVwG7cG5HHjboR5ainWcP9xNnRxNQP9yw0Qal

<https://twitter.com/i/status/1186491285919731713>



AI development today has become fixated on singular monolithic models trained end-to-end without any human assistance and encapsulating an almost general intelligence-like variety of tasks together. The resulting models have struggled in areas like [content moderation](#) to sufficiently [abstract](#) beyond their limited training data. Yet, as Waymo [reminds](#) us, the most successful complex AI systems combine multiple deep learning models with traditional hand-coded algorithms to address one of the greatest challenges confronting today's deep learning systems: their [inability](#) to abstract from correlation to causation.



Cause-and-Effect Diagrams

By Ron S. Kenett^{1,2}

Keywords: *scatter plots, Ishikawa diagrams, structural equation models, Bayesian networks, integrated management models*

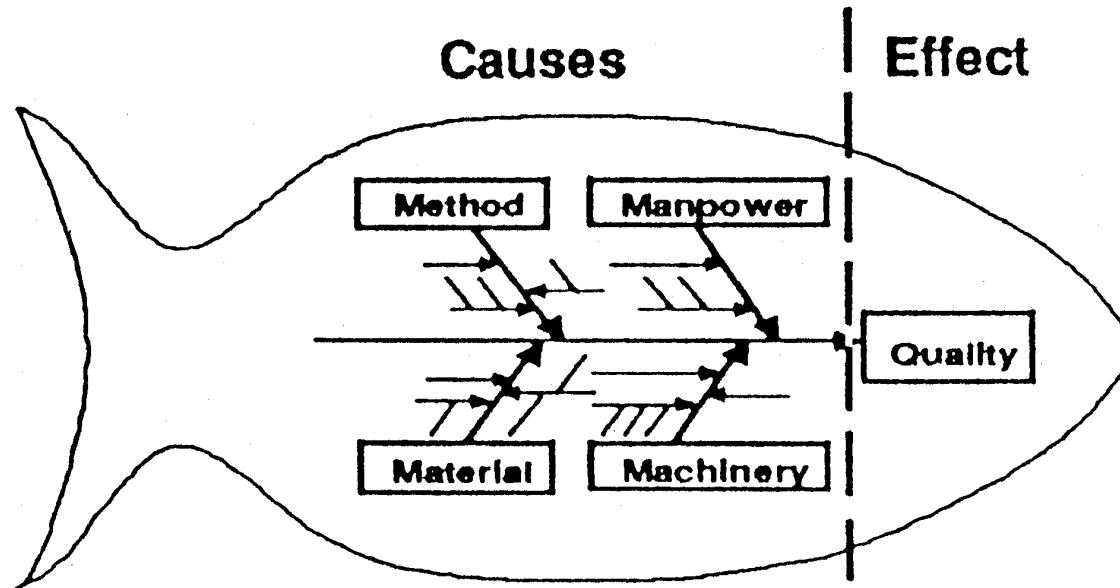
Abstract: Cause and effect is a basic knowledge driven by theoretical and empirical considerations. Several tools have been proposed to map cause and effect relationships, with some more heuristics some highly quantitative. In this section we cover the Ishikawa fishbone diagram, structural equation models, and Bayesian networks.

Cause-Effect Diagram

- Objectives: Visual presentation of relationships between **Effect** and possible **Causes**
- How?: **List** of possible Causes and their **Structure (Fishbone)**
- **Individual** and **Teamwork** tool for improvement program initiation
- Possibility to select critical Causes based on **Expert Knowledge**

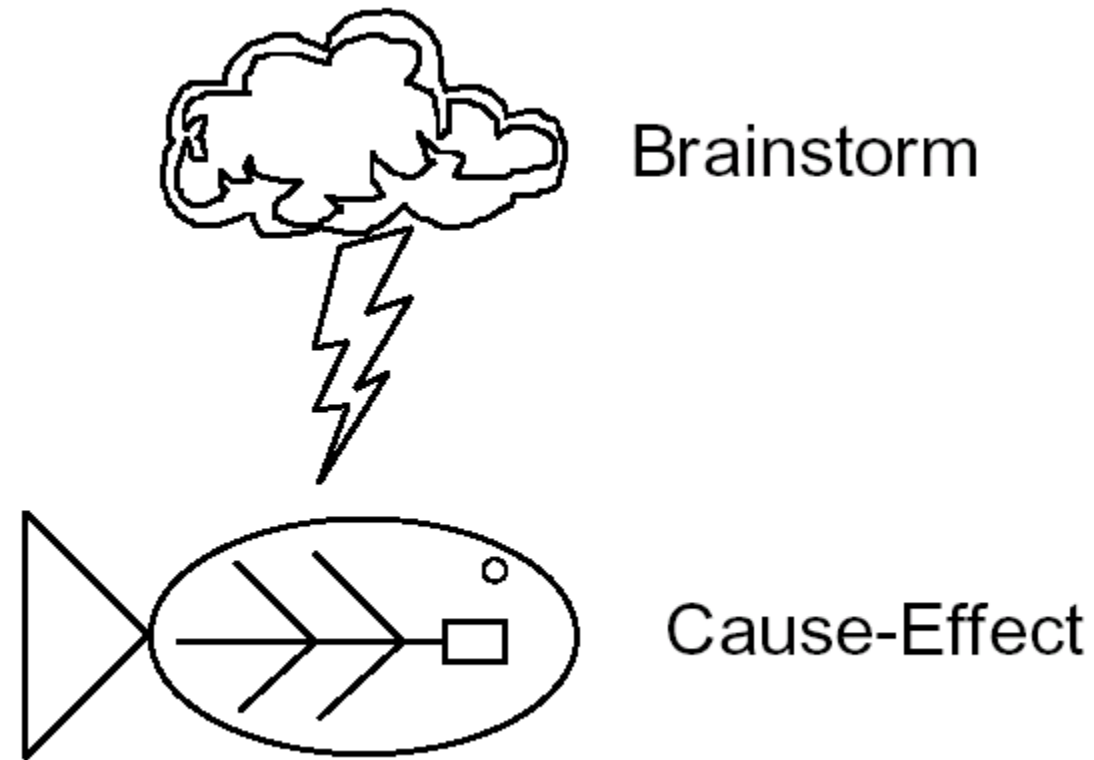
Cause-Effect (Ishikawa) Diagram

(Fishbone Diagram)



Kaoru Ishikawa
1915 - 1989

Cause-Effect Diagram Methodology



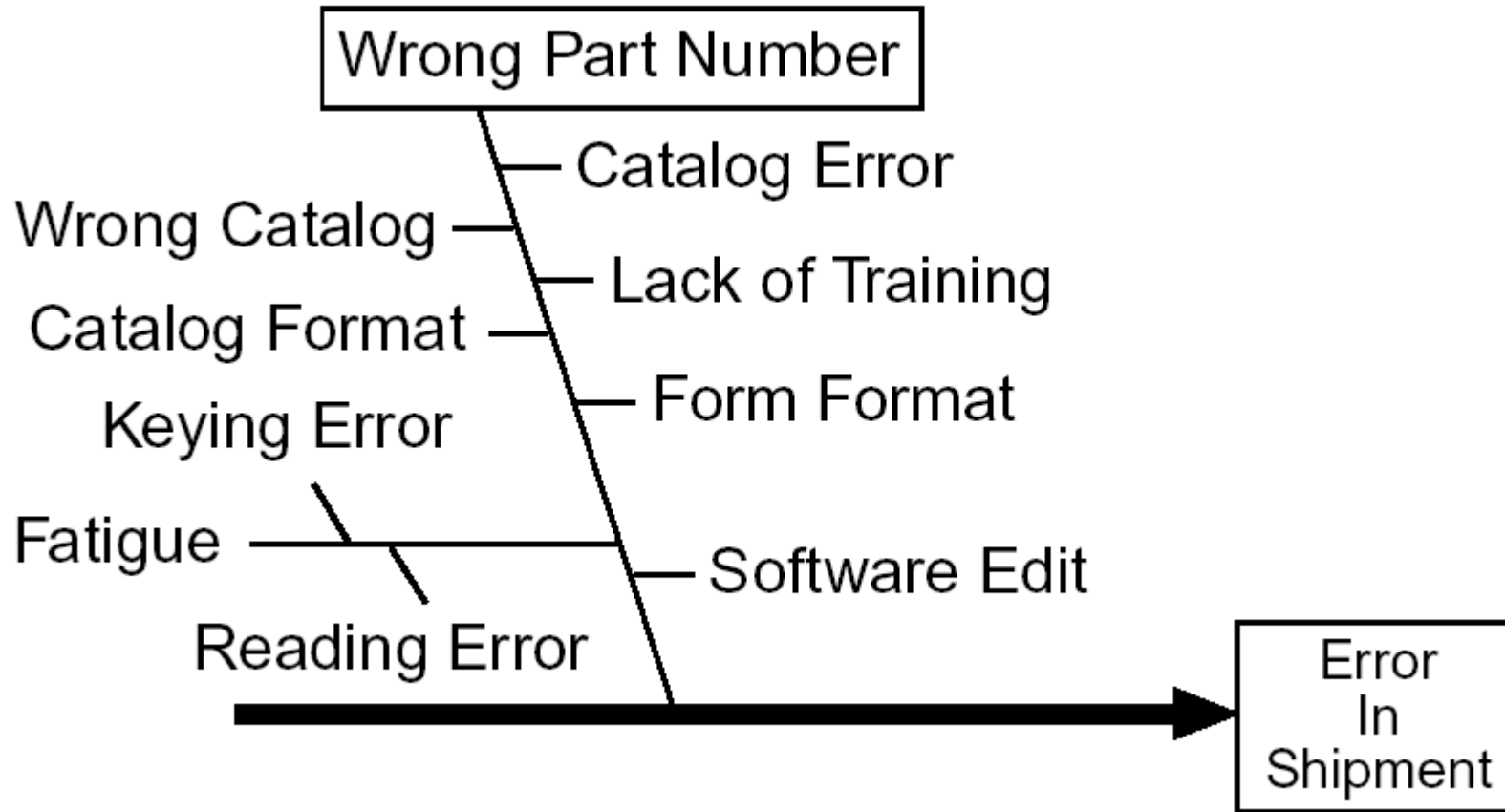
Round robin process



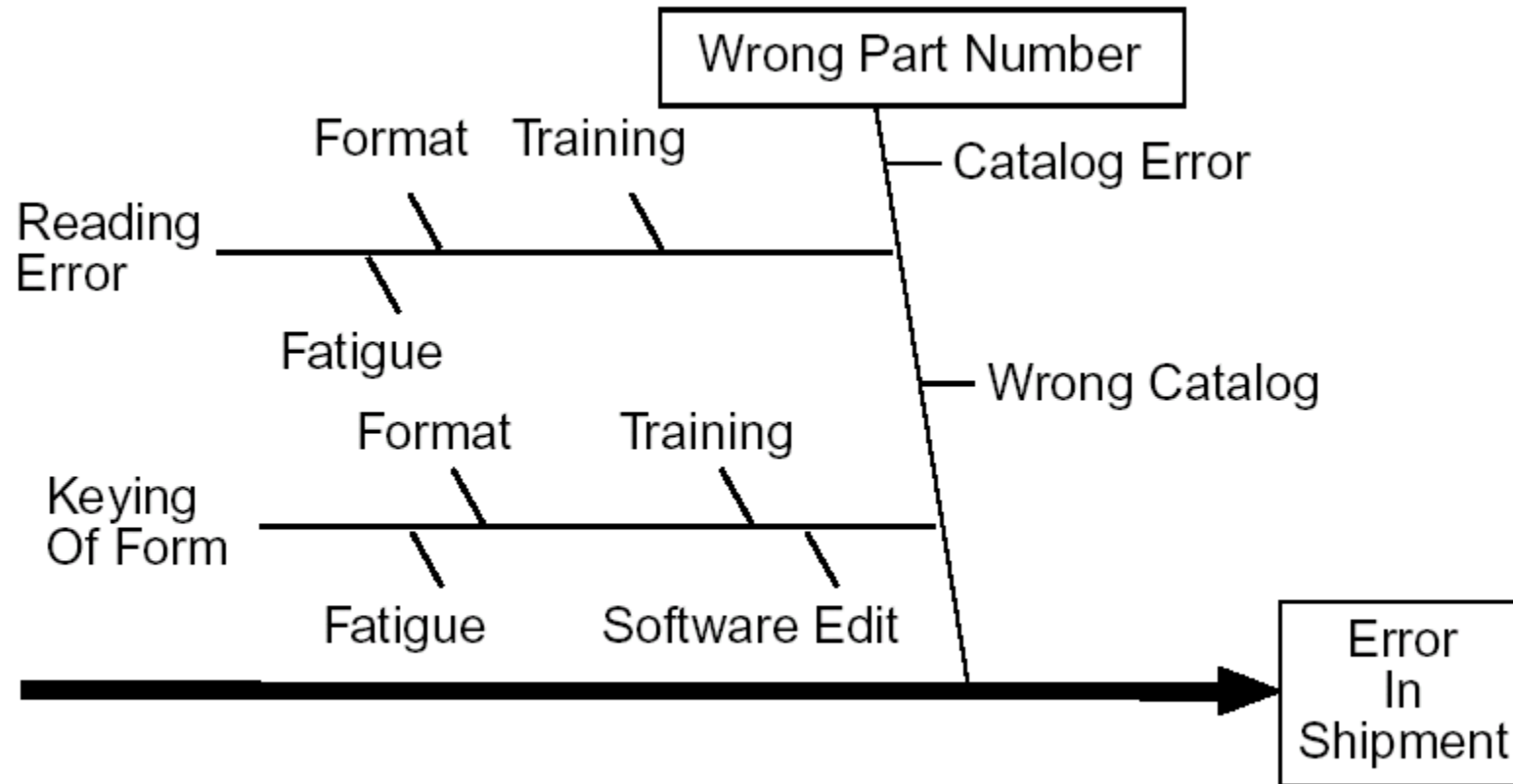
Rules

1. You can say “pass”
2. You can build on other’s ideas
3. No critique allowed (even self)
4. Indicate where to note the idea on the fishbone diagram

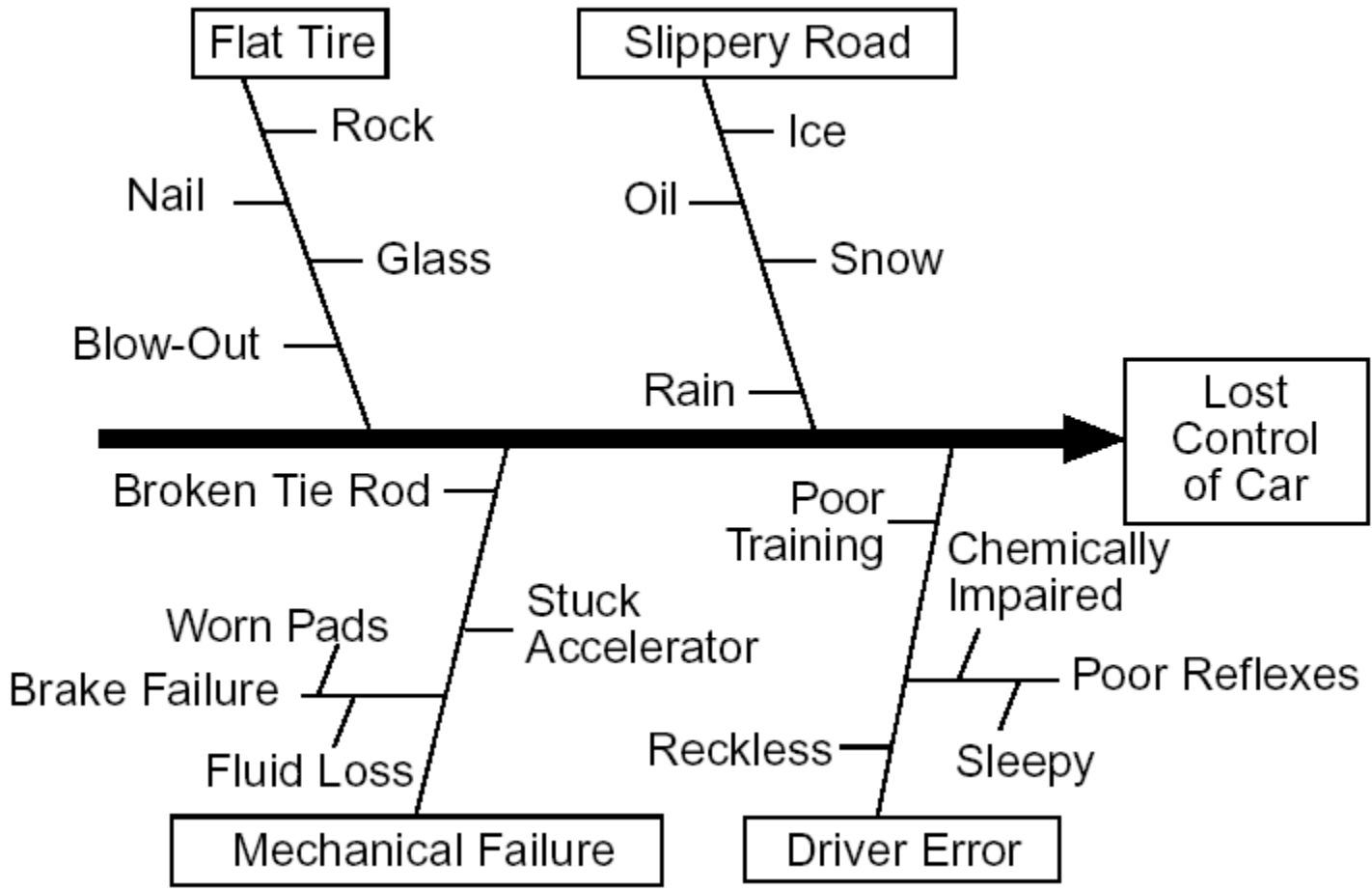
Why?



Why? Why? Why?

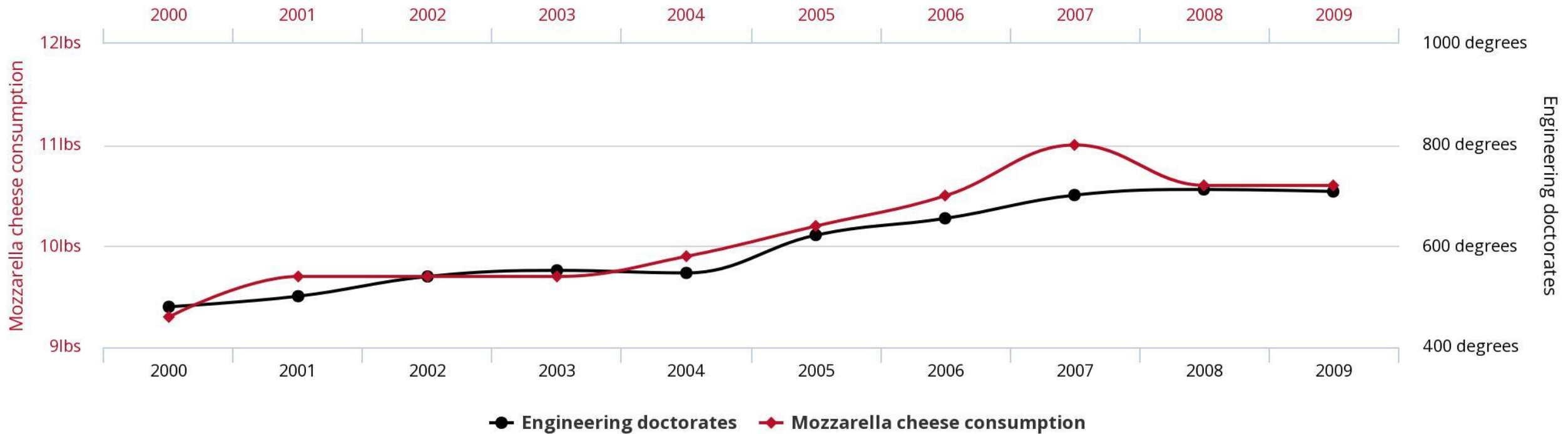


Lost control of a car



Correlation is not causation...

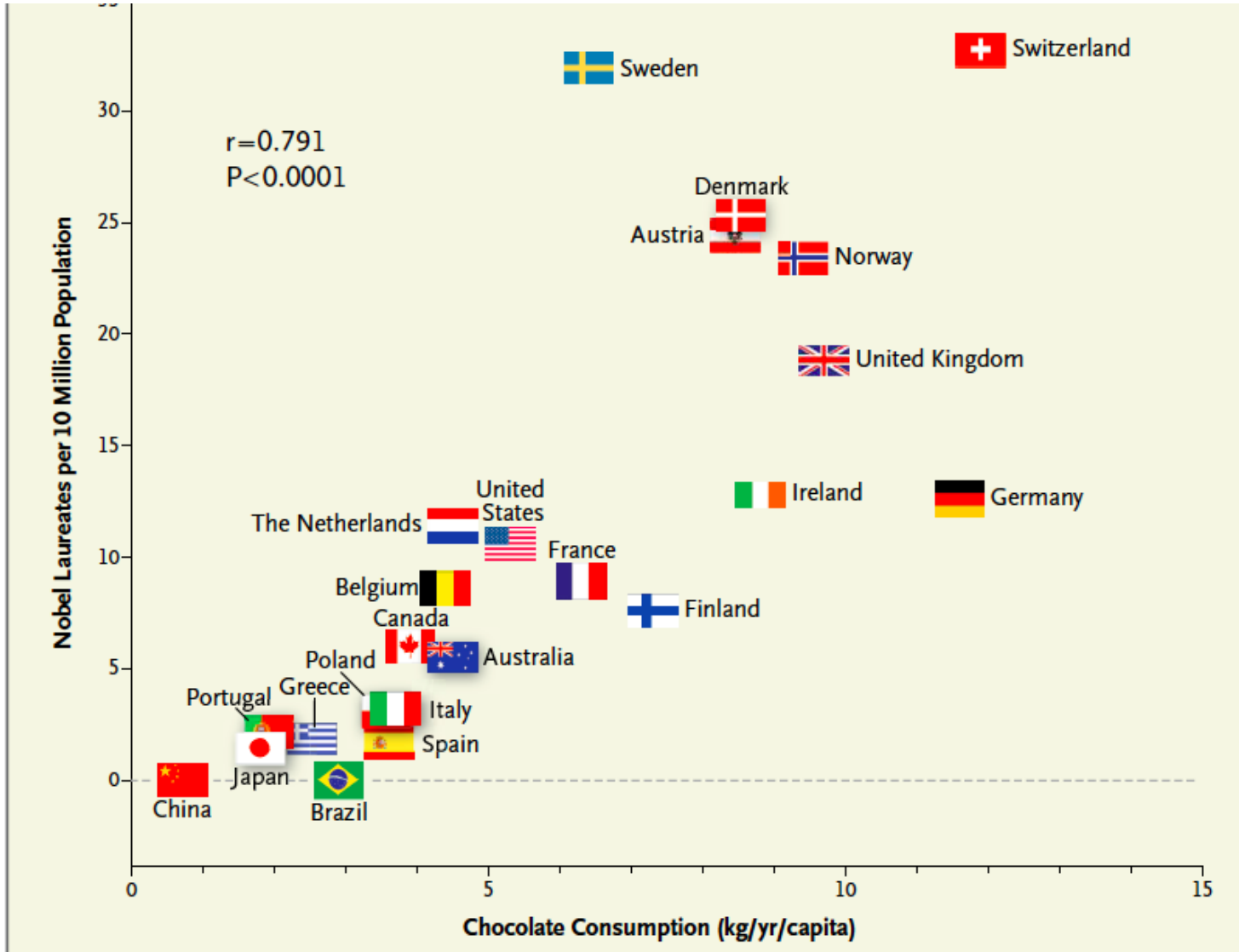
Per capita consumption of mozzarella cheese
correlates with
Civil engineering doctorates awarded



tylervigen.com

<http://www.tylervigen.com/spurious-correlations>

Correlation is not causation...

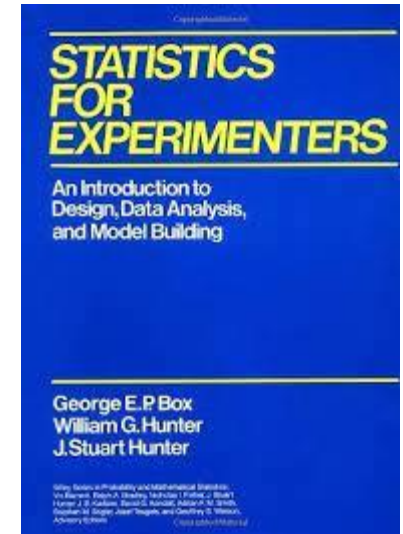


Correlation is not causation...

The population of Oldenburg in Germany and the number of observed storks in 1930-1936*

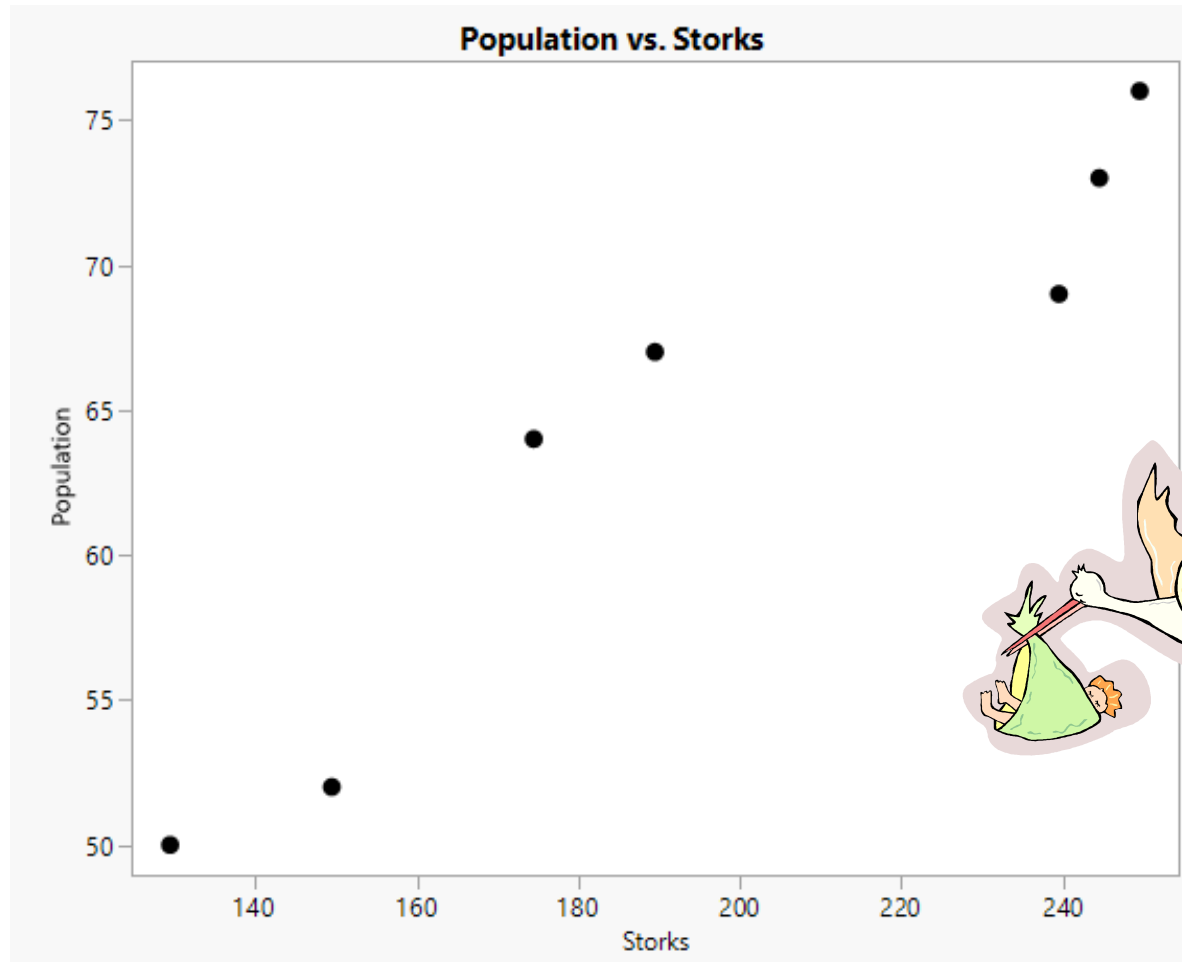
year	1930	1931	1932	1933	1934	1935	1936
Population in thousands	50	52	64	67	69	73	76
Number of storks	130	150	175	190	240	245	250

* Box, Hunter and Hunter, *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, J. Wiley, 1978



Spurious correlation

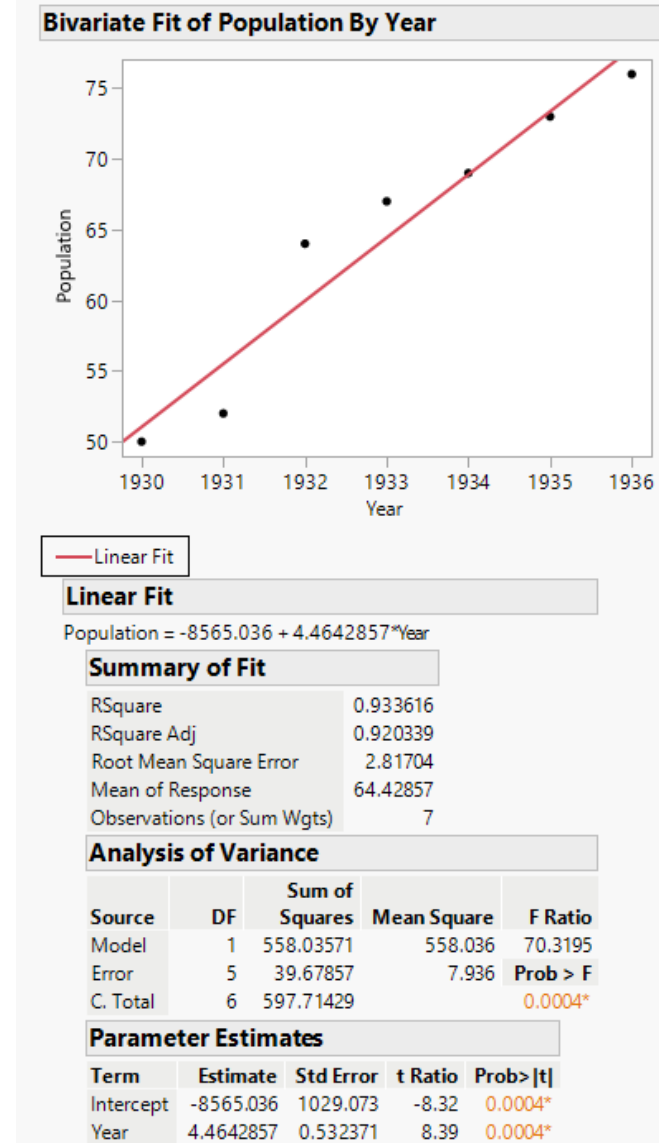
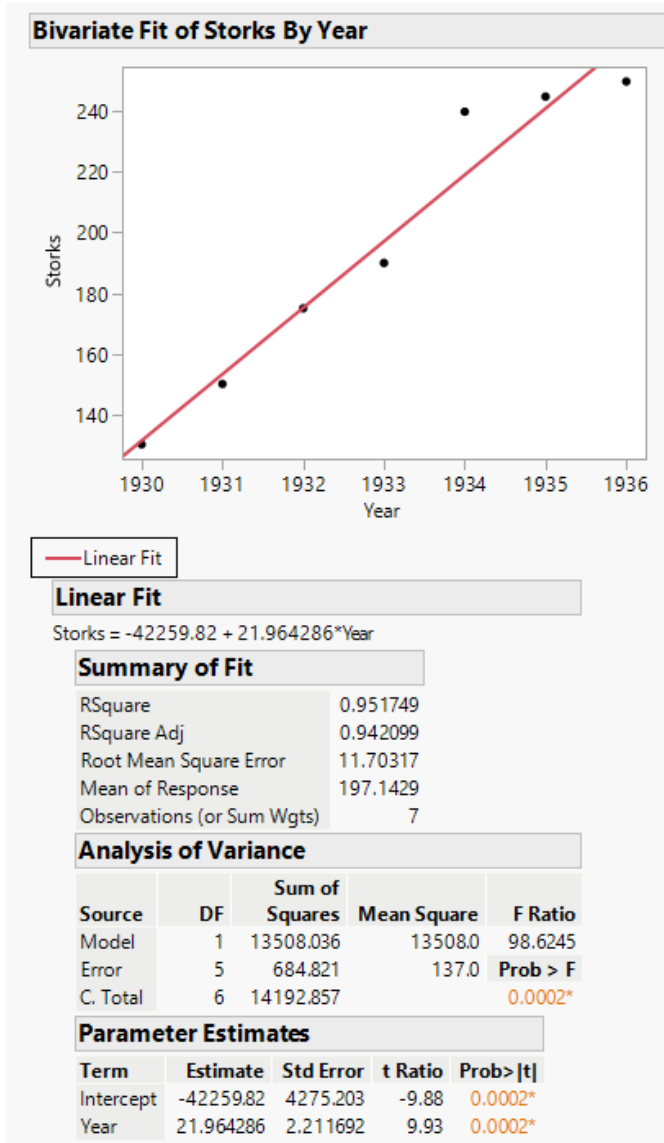
Time is a confounding variable



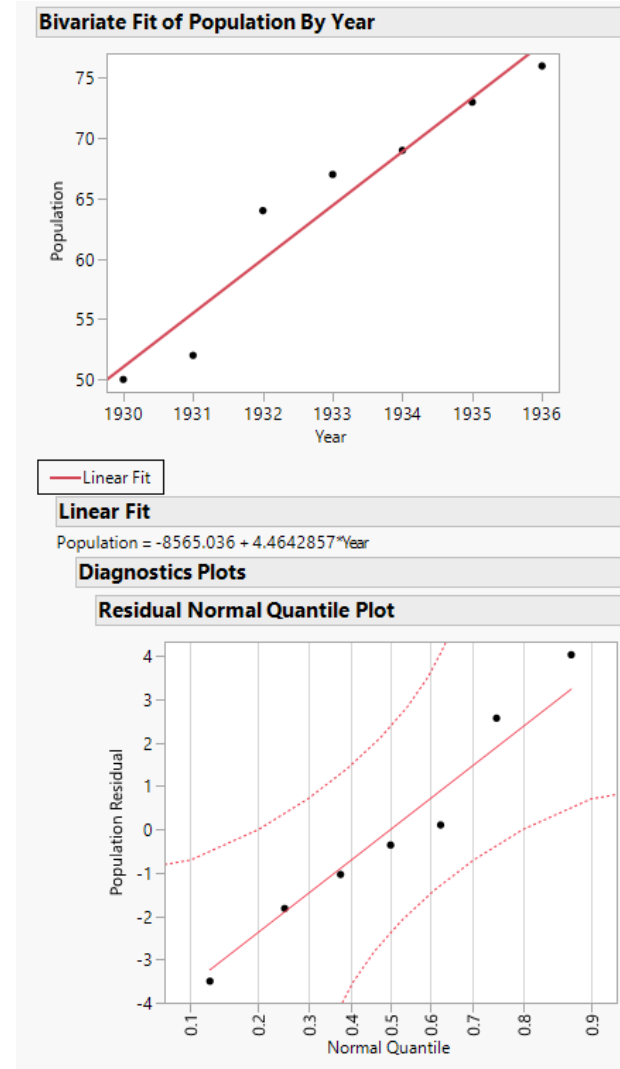
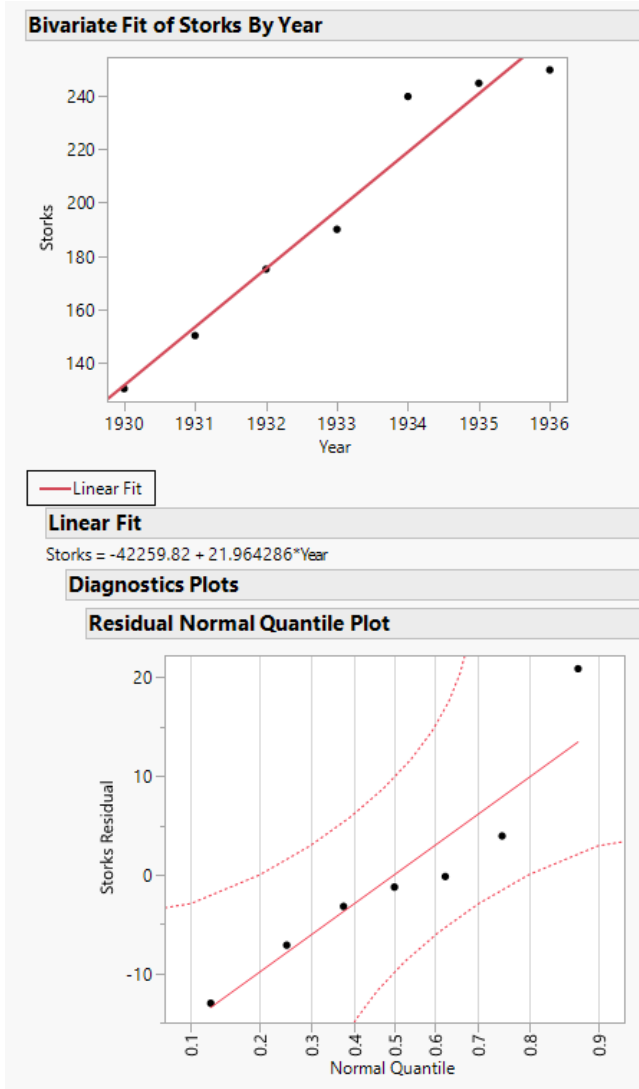
Correlations

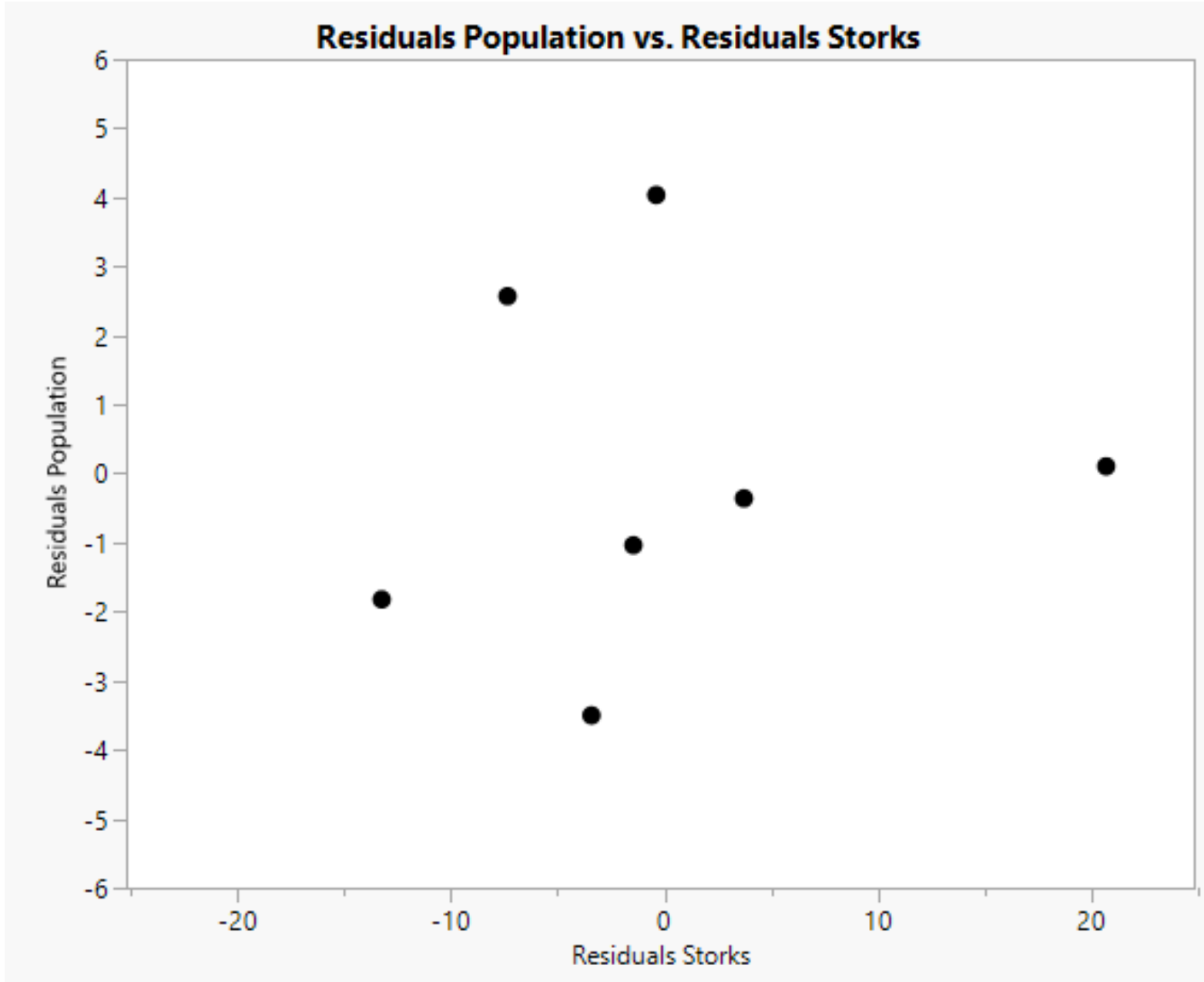
Pearson correlation 0.931
P-value 0.002

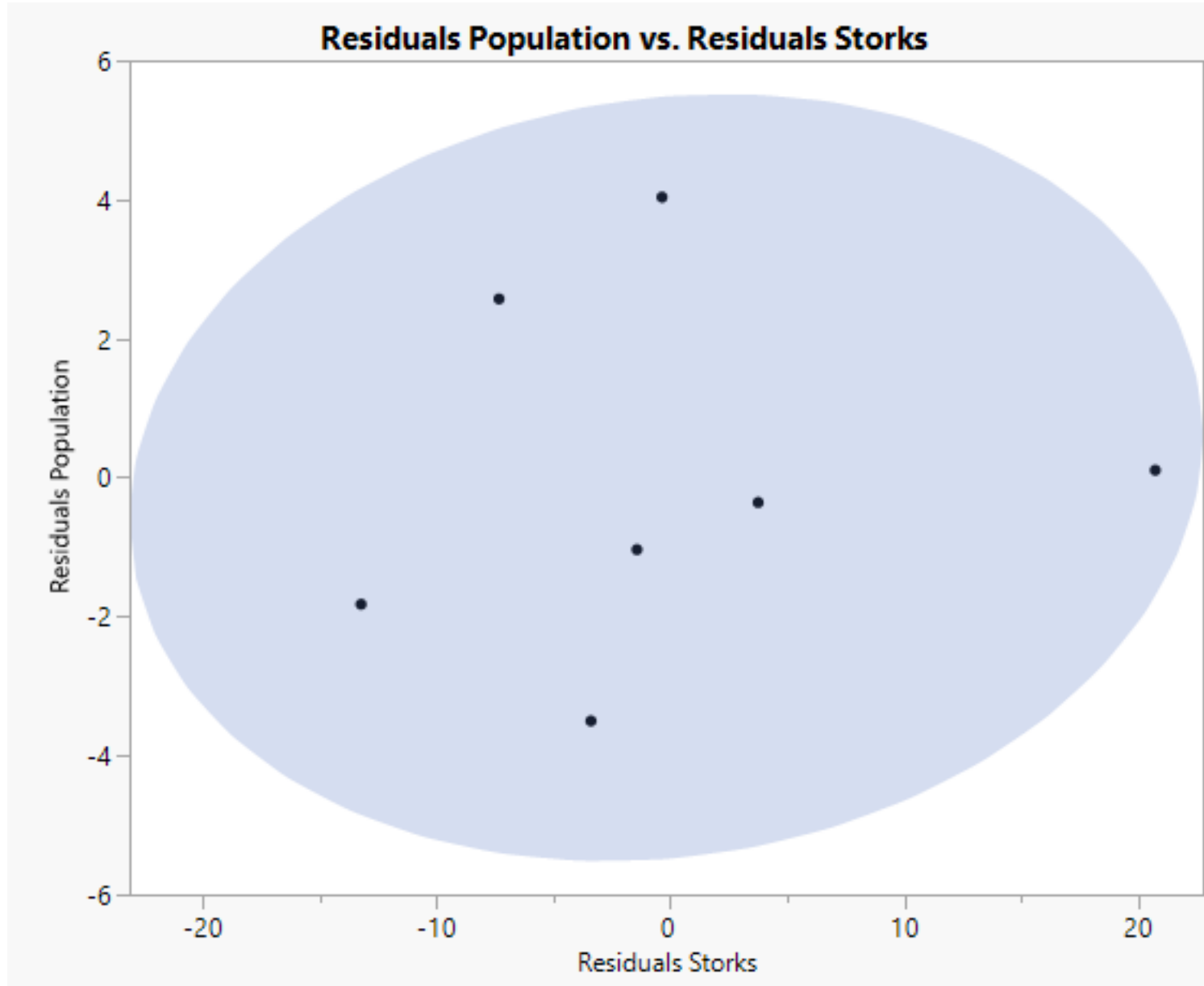
Spurious correlation



Spurious correlation







Correlations

Pearson correlation 0.931
P-value 0.002



Correlations

Pearson correlation -0.163
P-value 0.727



A demonstrable phenomenon
a.k.a. “a research claim”

“We may say that a phenomenon is experimentally demonstrable when we know how to conduct an experiment which will rarely fail to give us statistically significant results.”

R.A. Fisher (1935) The Design of Experiments, Macmillan Pub Co.



A demonstrable phenomenon
a.k.a. “a research claim”

“Personally, the writer prefers to set a low standard of significance at the 5 percent point... A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance.”

R.A. Fisher (1926) The arrangement of field experiments, the journal of the Ministry of Agriculture, 33: 504.



“No aphorism is more frequently repeated in connection with field trials, than that we must ask Nature few questions, or, ideally, one question, at a time. The writer is convinced that this view is wholly mistaken. Nature, he suggests, will best respond to a logical and carefully thought-out questionnaire. A factorial design allows the effect of several factors and interactions between them, to be determined with the same number of trials as are necessary to determine any one of the effects by itself with the same degree of accuracy.”

R.A. Fisher (1926). The arrangement of field experiments, *Journal of the Ministry of Agriculture of Great Britain* 33, 503–513.

Bradford Hill Criteria for Causation



1897 –1991

Strength (effect size): A small association does not mean that there is not a causal effect, though the larger the association, the more likely that it is causal.

Consistency (reproducibility): Consistent findings observed by different persons in different places with different samples strengthens the likelihood of an effect.

Specificity: Causation is likely if there is a very specific population at a specific site and disease with no other likely explanation. The more specific an association between a factor and an effect is, the bigger the probability of a causal relationship.

Temporality: The effect has to occur after the cause (and if there is an expected delay between the cause and expected effect, then the effect must occur after that delay).

Biological gradient: Greater exposure should generally lead to greater incidence of the effect. However, in some cases, the mere presence of the factor can trigger the effect. In other cases, an inverse proportion is observed: greater exposure leads to lower incidence.[

Plausibility: A plausible mechanism between cause and effect is helpful (but Hill noted that knowledge of the mechanism is limited by current knowledge).

Coherence: Coherence between epidemiological and laboratory findings increases the likelihood of an effect. However, Hill noted that "... lack of such [laboratory] evidence cannot nullify the epidemiological effect on associations".

Experiment: "Occasionally it is possible to appeal to experimental evidence".

Analogy: The effect of similar factors may be considered.

Bradford Hill, A. (1953). Observation and experiment. *New England Journal of Medicine* 248:995-1001

Bradford Hill, A. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine* 58:295-300



Austin Bradford Hill
(1897-1991)

The Environment and Disease: Association or Causation?

by Sir Austin Bradford Hill CBE DSC FRCP(hon) FRS
(*Professor Emeritus of Medical Statistics,
University of London*)

Amongst the objects of this newly-founded Section of Occupational Medicine are firstly 'to provide a means, not readily afforded elsewhere, whereby physicians and surgeons with a special knowledge of the relationship between sickness and injury and conditions of work may discuss their problems, not only with each other, but also with colleagues in other fields, by holding joint meetings with other Sections of the Society'; and, secondly, 'to make available information about the physical, chemical and psychological hazards of occupation, and in particular about those that are rare or not easily recognized'.



Richard Doll
(1912 – 2005)

BRITISH MEDICAL JOURNAL

LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG PRELIMINARY REPORT

BY
RICHARD DOLL, M.D., M.R.C.P.
Member of the Statistical Research Unit of the Medical Research Council

AND
A. BRADFORD HILL, Ph.D., D.Sc.
Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

Br Med J. 1950 Sep 30; 2(4682):
739–748. Smoking and
Carcinoma of the Lung, Richard
Doll and A. Bradford Hill

U.S. Surgeon General Luther Terry holds a copy of the 387 page report of the Advisory Committee to the Surgeon General of the Public Health Service on the relationship of smoking to health Jan 11, 1964. He spoke at a Washington news conference at which the study was released. It termed smoking a health hazard calling for corrective action.”
(AP Photo/hwg)



US report ties smoking to cancer

Bridging observational studies and randomized experiments by embedding the former in the latter

Marie-Abele C Bind and Donald B Rubin

Abstract

Consider a statistical analysis that draws causal inferences from a being valid in the standard frequentist senses; i.e. the analysis pro valid in the sense of rejecting true null hypotheses at the nom which are presented as having at least their nominal coverage fo statements, the analysis must embed the observational study in observed data, or a subset of that hypothetical randomized data involves: (1) a purely conceptual stage that precisely formulate th experiment where the exposure is assigned to units; (2) a de before any outcome data are observed, (3) a statistical analysis st and non-exposed units of the hypothetical randomized experime statistical evidence for the sizes of possible causal effects. Stage the effort, whereas Stage I demands careful scientific argume readers of the proffered statistical analysis. Otherwise, the resu a presentation of scientifically meaningless arithmetic calculation most scientifically interesting to the dedicated researcher a perspective is rarely implemented with any rigor, for example, approach using an example examining the effect of parental smol in East Boston in the 1970s.

CAUSAL INFERENCE IN RETROSPECTIVE STUDIES

PAUL W. HOLLAND

*Research Statistics Group
Educational Testing Service*

DONALD B. RUBIN

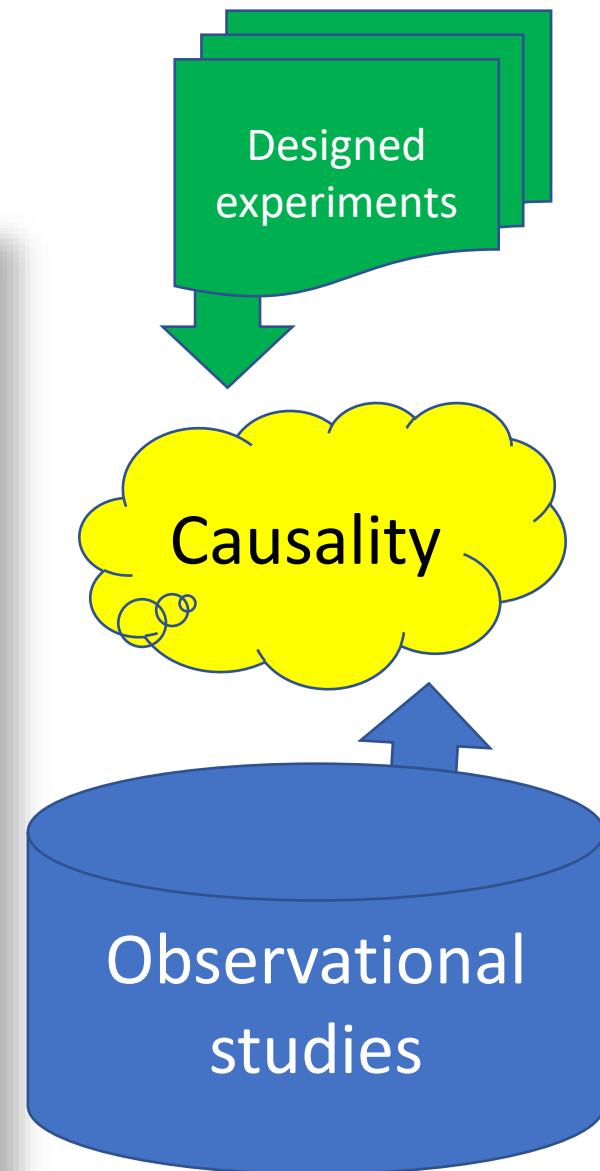
Harvard University

Philosophical discussions of causality often emphasize the *meaning* of causation. Scientists are usually concerned with *understanding* causal mechanisms. Purely statistical discussions of causality are substantially more limited in scope, because the unique contribution of statistics is to *measuring* causal effects and not to the understanding of causal mechanisms or to the meaning of causation. This distinction is sometimes expressed as “statistics can establish correlation, but not causation.” We feel our emphasis on *measurement* is more appropriate, because it focuses on what statistical theory *can* contribute to discussions of causality. Measuring causal effects accurately without any understanding whatsoever of the causal mechanisms

AUTHORS' NOTE: A version of this article titled “Causal Inference in Prospective and Retrospective Studies” was delivered at the Jerome Cornfield Memorial Session of the American Statistical Association, August 1980, in Houston. The topic of the article was especially appropriate for that session since many important contributions to the study of health effects from prospective and retrospective studies were made by Jerome Cornfield.

EVALUATION REVIEW, Vol. 12 No. 3, June 1988 203-231

© 1988 Sage Publications, Inc.



Imagining

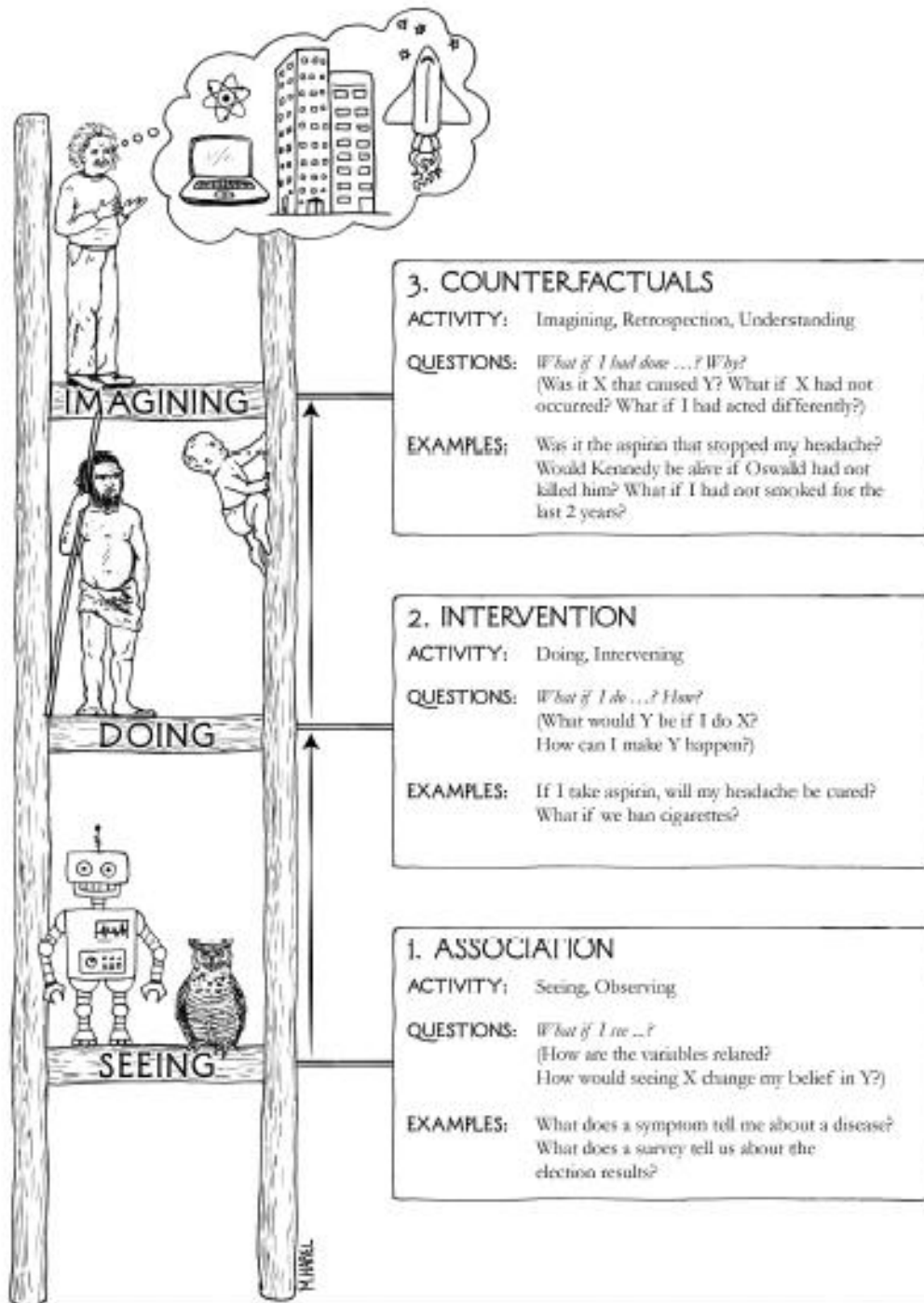
3

Doing

2

Seeing

1



JUDEA PEARL
WINNER OF THE TURING AWARD
AND DANA MACKENZIE

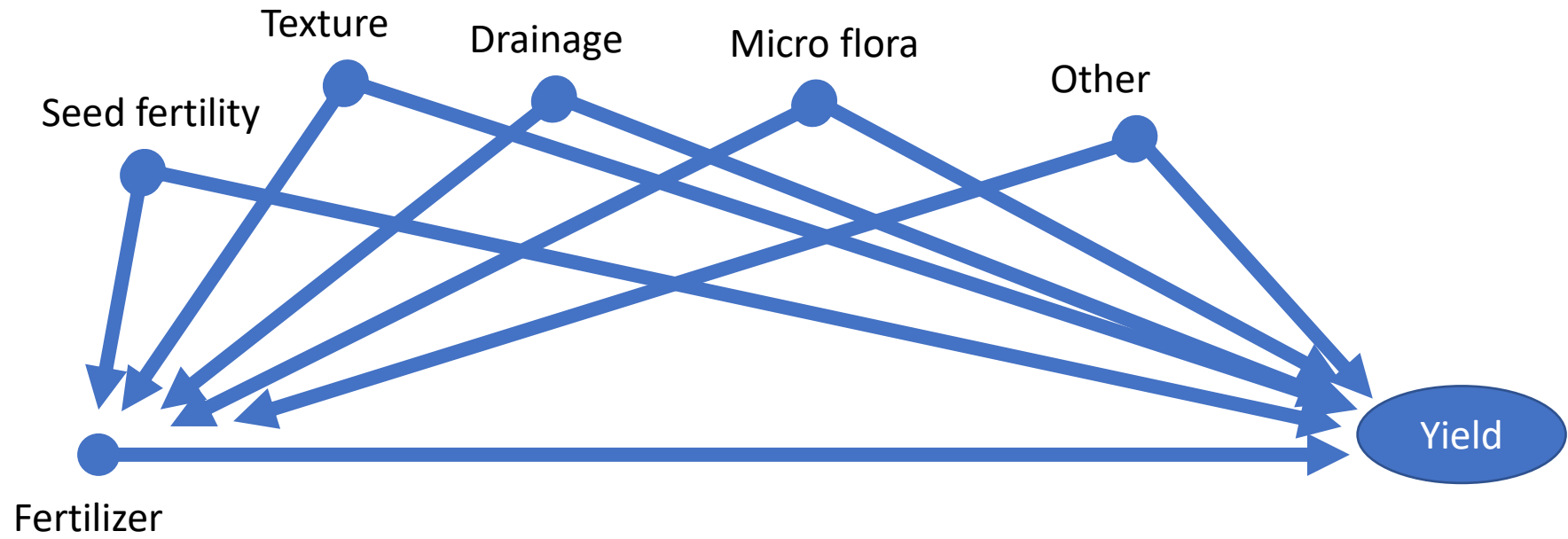
THE BOOK OF WHY



THE NEW SCIENCE
OF CAUSE AND EFFECT

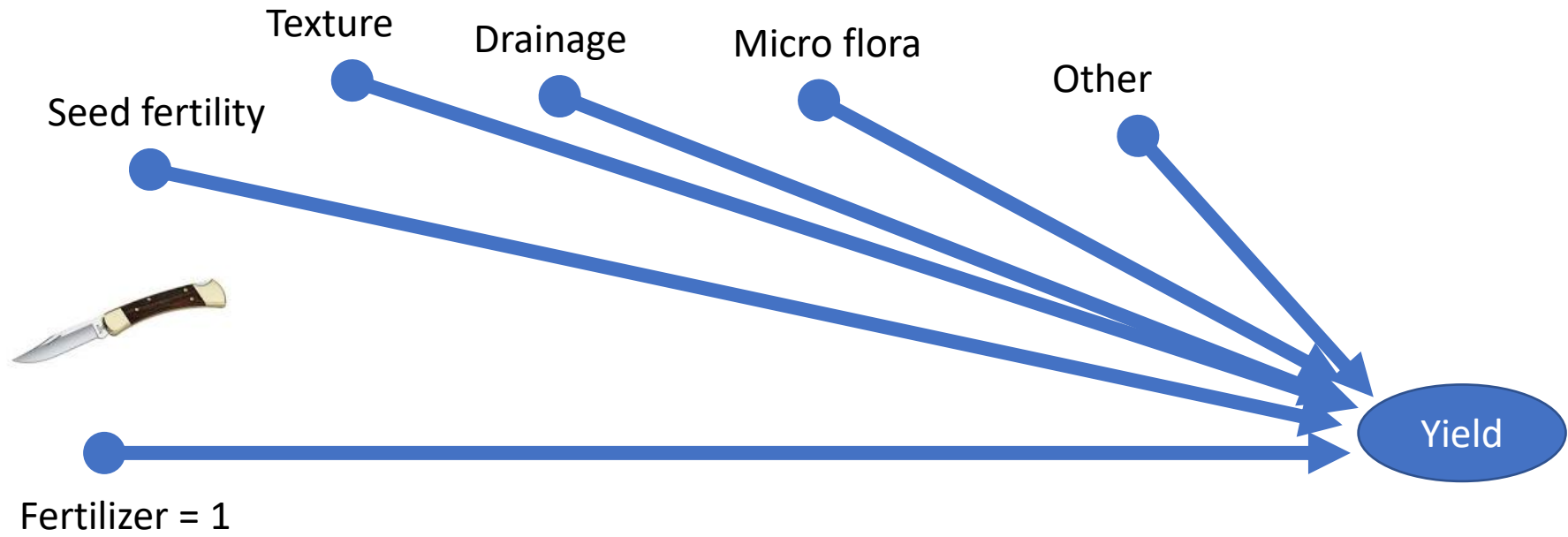


What we have



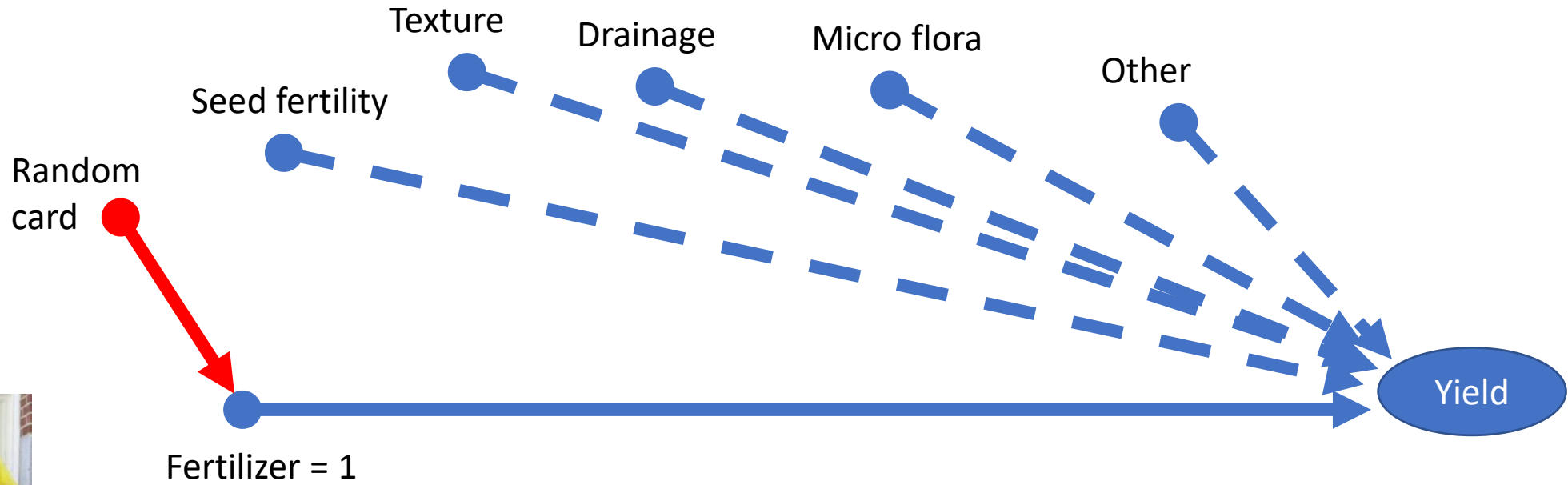


What we want





What we get with randomization



2019 RCT in economics



Impact of economic reforms

Abhijit Banerjee and Esther Duflo: The 2019 Nobel couple fighting poverty. They pioneered “randomized controlled trials” in Economics.

<https://foreignpolicy.com/2019/10/22/economics-development-rcts-esther-duflo-abhijit-banerjee-michael-kremer-nobel/>

MIT News
ON CAMPUS AND AROUND THE WORLD

https://dash.harvard.edu/bitstream/handle/1/37373527/RWP22_021_Alsan.pdf?sequence=1&isAllowed=y

Social media messages from health care workers help reduce travel-related Covid-19 spread

Large-scale video campaign allowed physicians and public health messengers to encourage staying home over the 2020 holidays.

J-PAL North America
August 23, 2021

<https://news.mit.edu/2021/social-media-messages-health-care-workers-reduce-travel-covid-spread-0823>

<https://www.bbc.com/news/world-asia-india-50048519>

nature medicine

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature medicine](#) > [articles](#) > [article](#)

Article | [Open Access](#) | Published: 19 August 2021

Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: a cluster randomized controlled trial

Emily Breza, Fatima Cody Stanford, Marcella Alsan, Burak Alsan, Abhijit Banerjee, Arun G. Chandrasekhar, Sarah Eichmeyer, Traci Glushko, Paul Goldsmith-Pinkham, Kelly Holland, Emily Hoppe, Mohit Karnani, Sarah Liegl, Tristan Loisel, Lucy Ogbu-Nwobodo, Benjamin A. Olken, Carlos Torres, Pierre-Luc Vautrey, Erica T. Warner, Susan Wootton & Esther Duflo [✉](#)

<https://www.nature.com/articles/s41591-021-01487-3>

A new large-scale randomized evaluation has found that messages delivered by physicians increased knowledge about Covid-19 and use of preventative health measures, like mask-wearing and social distancing, regardless of recipients' race or political beliefs. This research shows that information campaigns delivered by trusted experts can be effective in changing people's health-related beliefs and behaviors.

Impact of communication media during COVID-19

The evaluation tested the effectiveness of three video messages about Covid-19, recorded by physicians of different ages, genders, and races. One message defined Covid-19 and discussed common symptoms associated with the virus and asymptomatic transmission. A second message reminded the viewer that Covid-19 was actively circulating in the United States. The final message described U.S. Centers for Disease Control and Prevention social distancing guidelines. The study included over 18,000 Black and white adults of modest incomes (the majority below \$60,000) in the United States.

Granger Causality

$$Y_t = a_0 + \sum_{j=1}^p a_j Y_{t-j} + e_t$$

$$Y_t = a_0 + \sum_{j=1}^p a_j Y_{t-j} + \sum_{j=1}^p b_j X_{t-j} + e_t.$$

In the second model, the contributed lagged terms b_j of X_{t-j} are tested for significance. If some of the b_j are significant and the addition of the X_{t-j} adds explanatory power, X_t Granger causes Y_t . The assumptions here are that:

1. The cause happens prior to its effect.
2. The cause has unique information about the future values of its effect.

Impact of an educational reform

Effects of Free Choice Among Public Schools

VICTOR LAVY

Hebrew University, Royal Holloway University of London, CEPR and NBER

First version received March 2008; final version accepted September 2009 (Eds.)

In this paper, I investigate the impact of a programme in Tel-Aviv, Israel, that terminated an existing inter-district busing integration programme and allowed students free choice among public schools. The identification is based on difference-in-differences and regression discontinuity designs that yield various alternative comparison groups drawn from untreated tangent neighbourhoods and adjacent cities. Across identification methods and comparison groups, the results consistently suggest that choice significantly reduces the drop-out rate and increases the cognitive achievements of high-school students. It also improves behavioural outcomes such as teacher–student relationships and students’ social acclimation and satisfaction at school, and reduces the level of violence and classroom disruption.



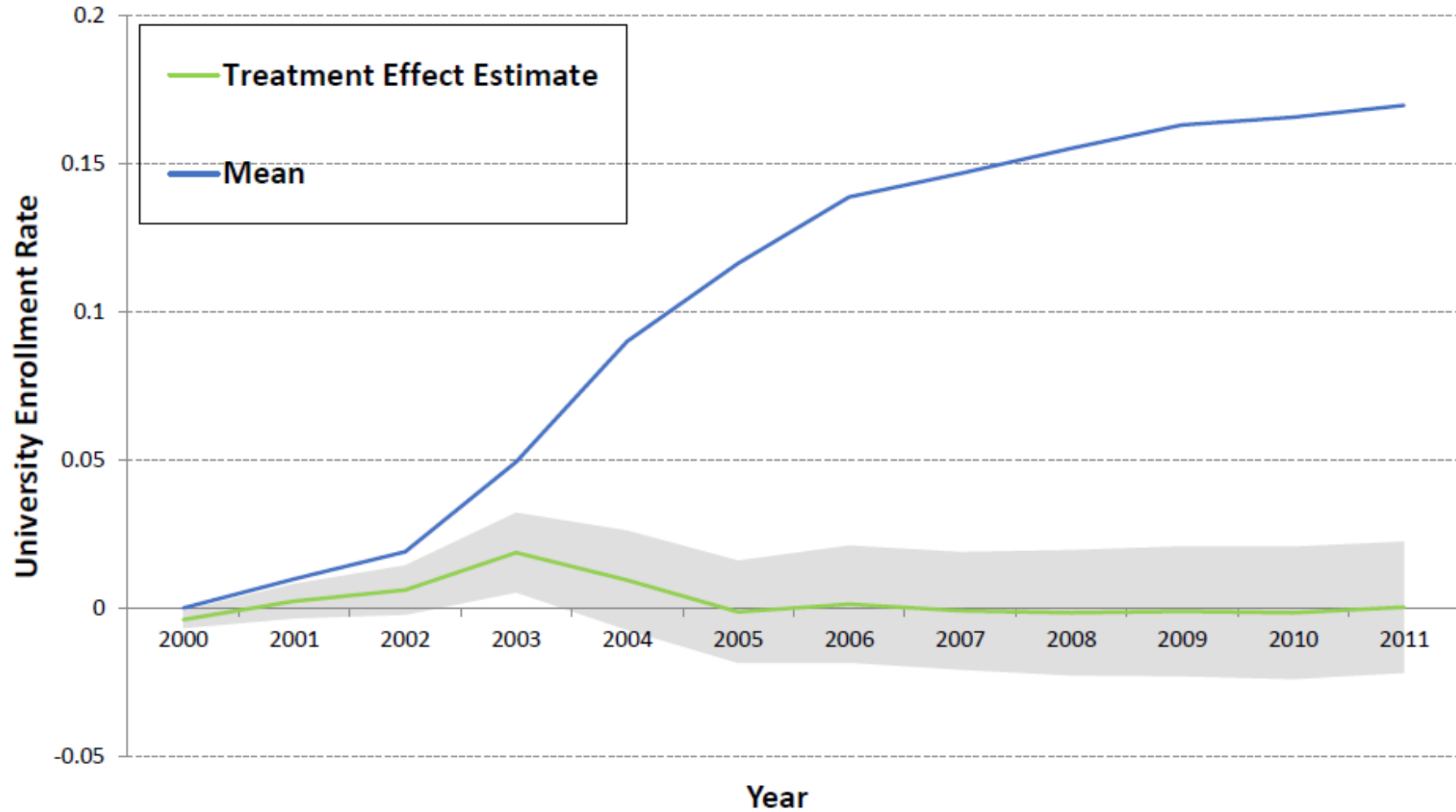
Joshua Angrist



2021

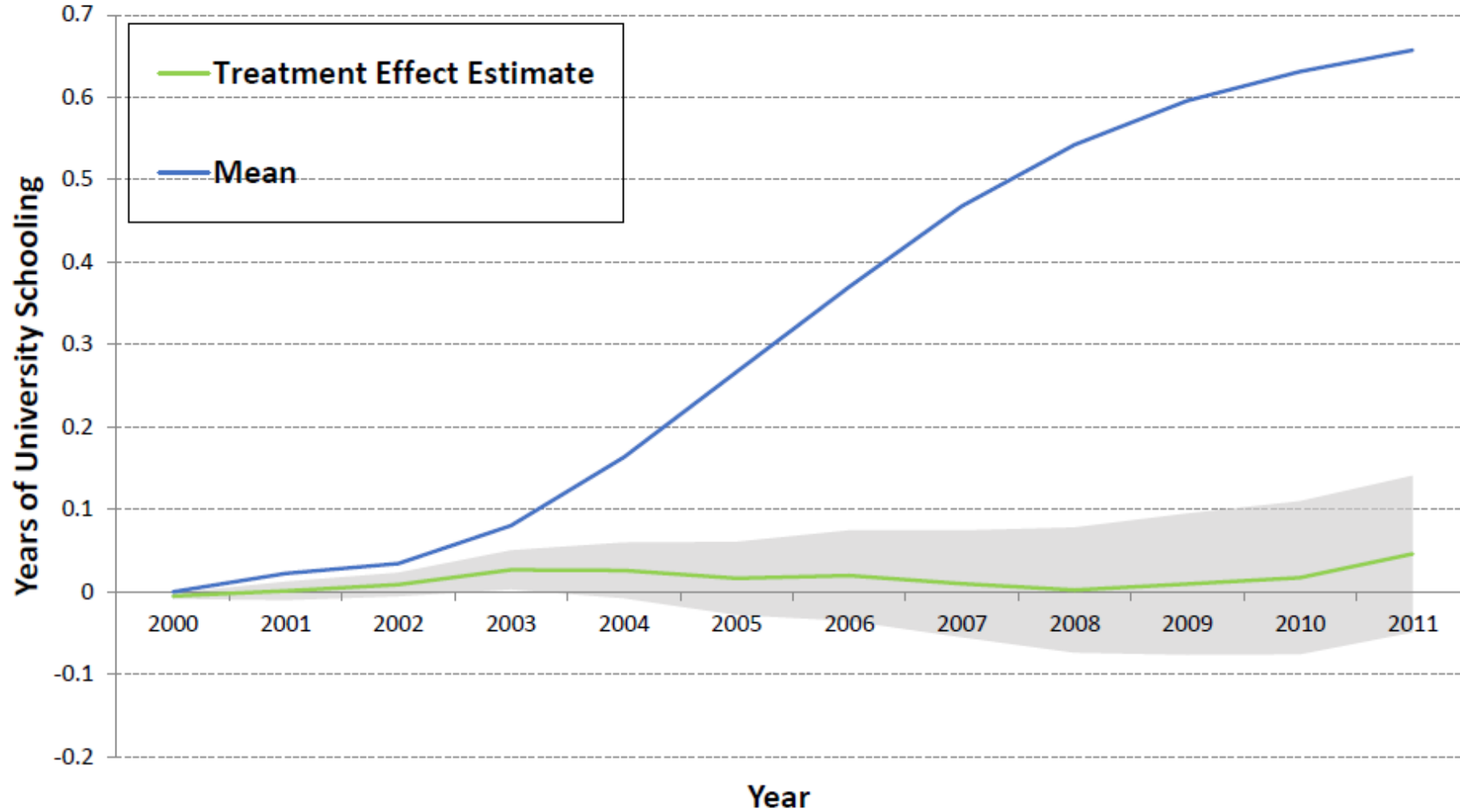


Mean and Treatment Effect: Enrollment Rate in Universities



*Shaded area indicates two sided confidence intervals, 10% significance level.

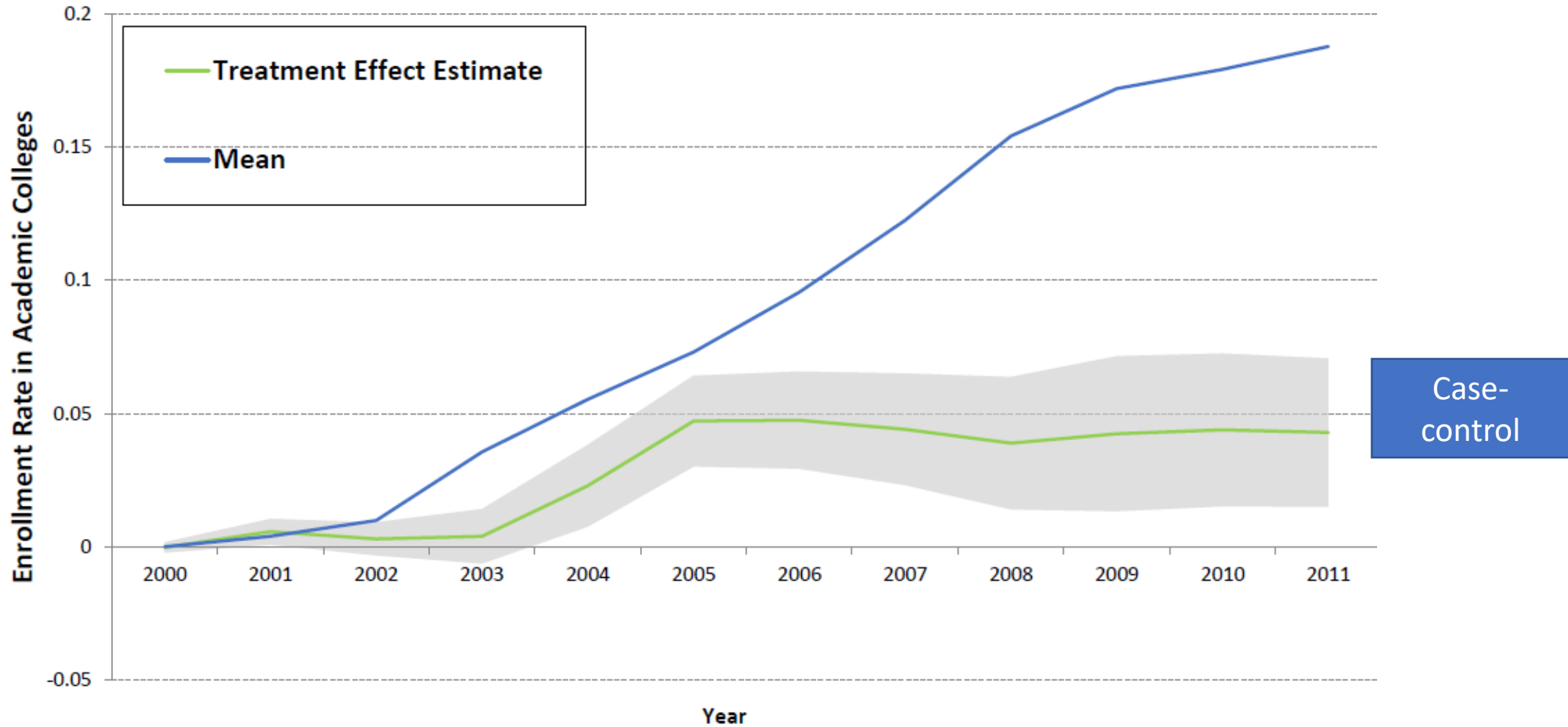
Mean and Treatment Effect: Years of University Schooling



Case-control

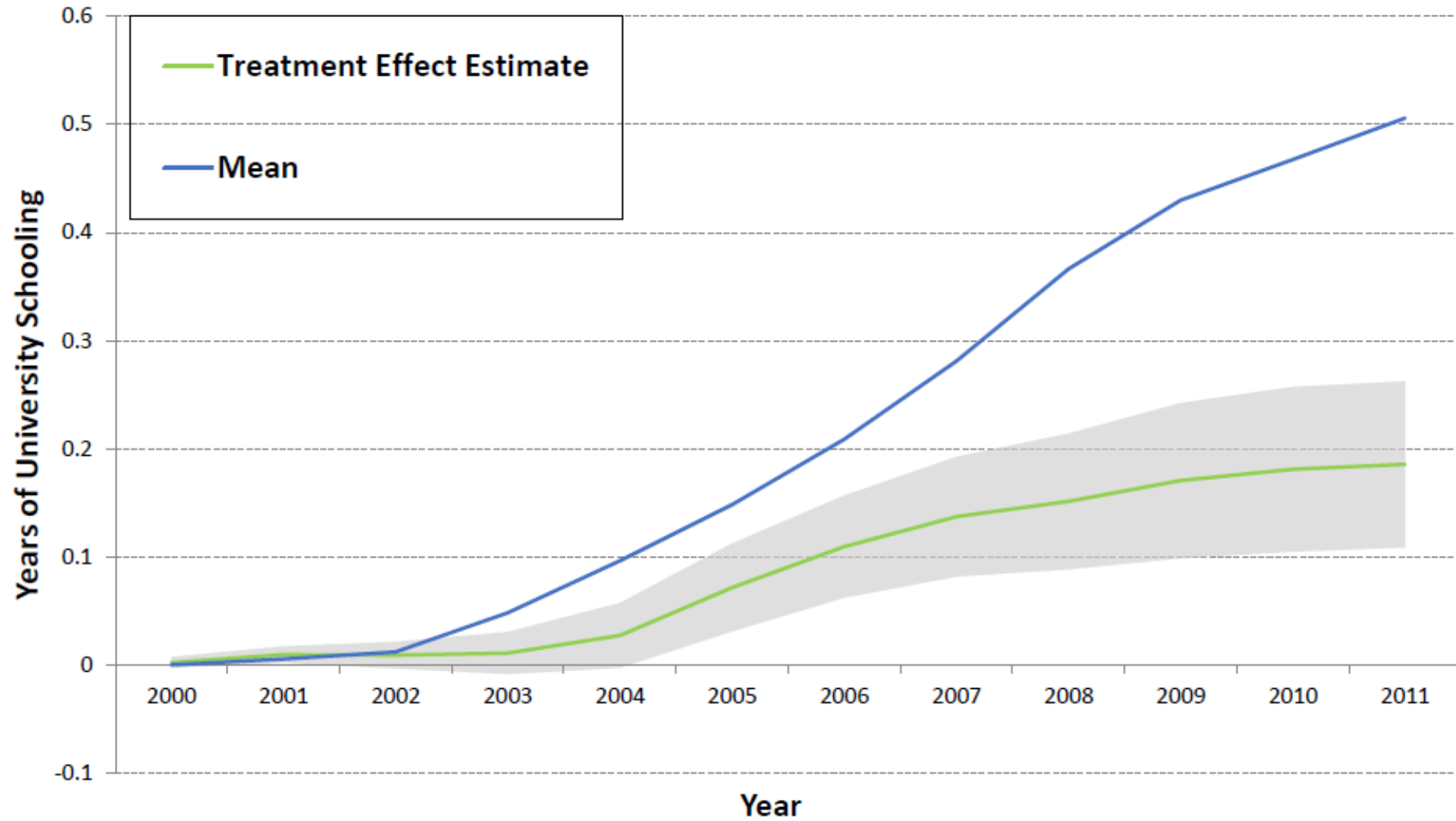
*Shaded area indicates two sided confidence intervals, 10% significance level.

Mean and Treatment Effect: Enrollment Rate in Academic Colleges)



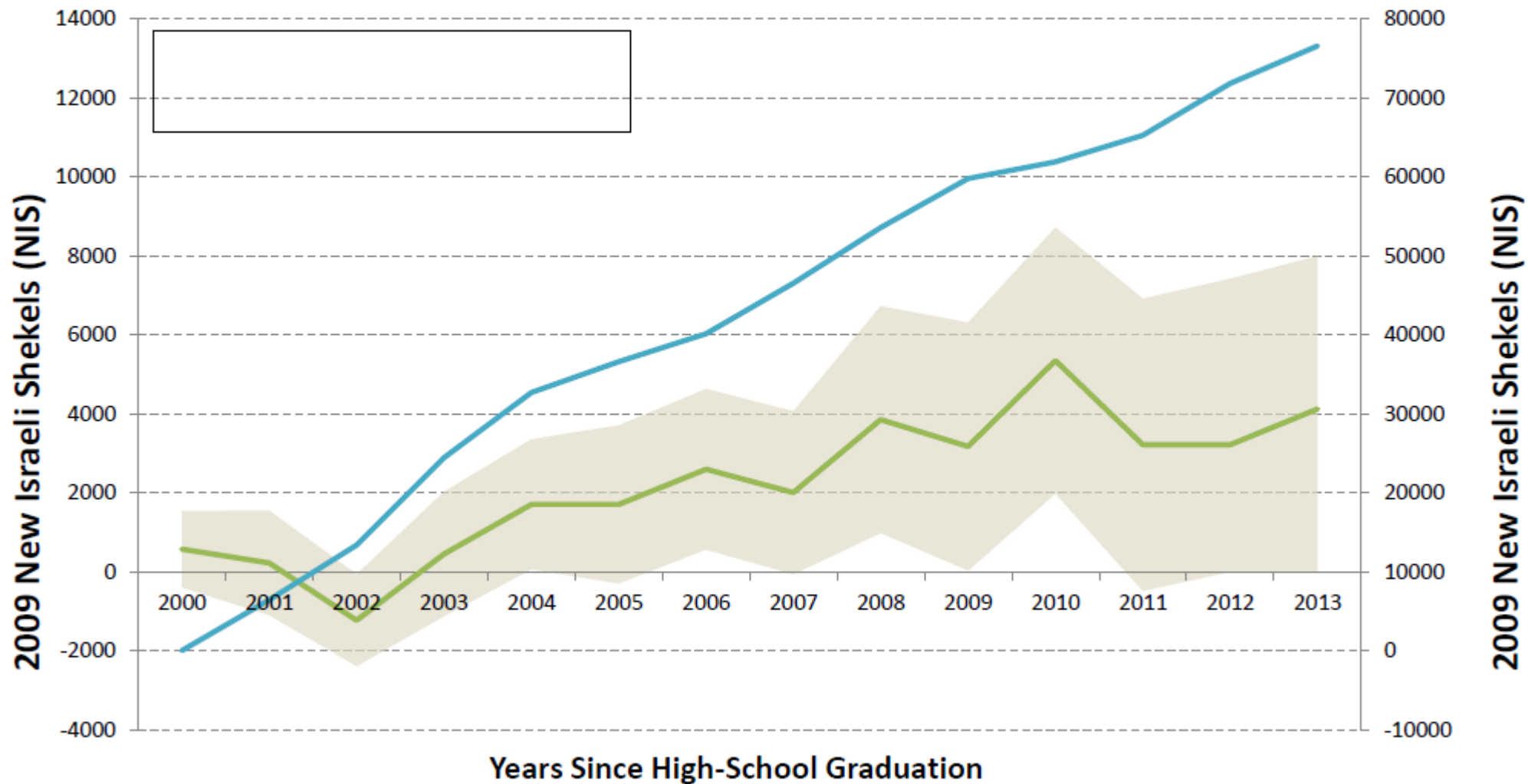
*Shaded area indicates two sided confidence intervals, 10% significance level.

Mean and Treatment Effect: Years of Academic Colleges Schooling



*shaded area indicates two sided confidence intervals, 10% significance level.

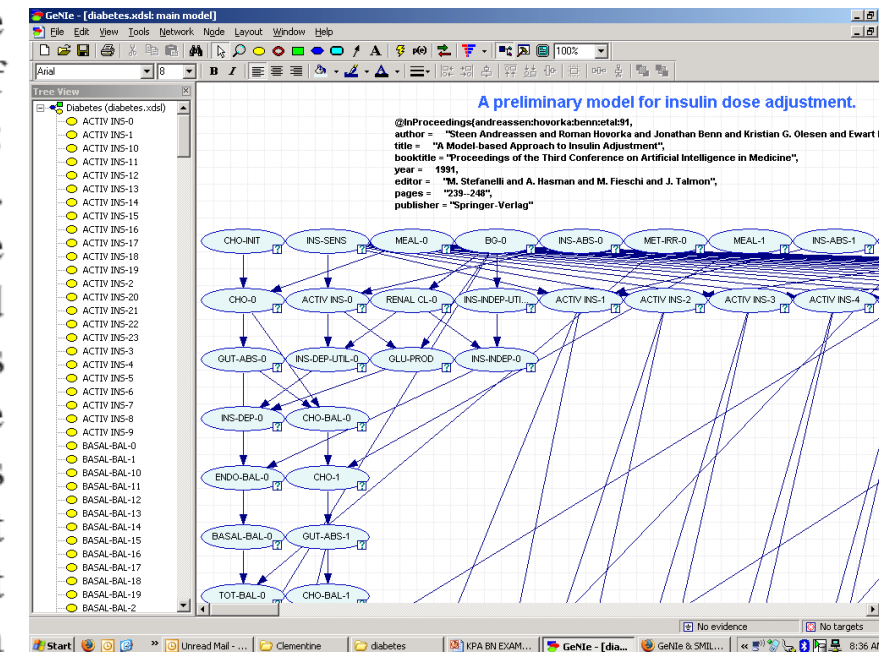
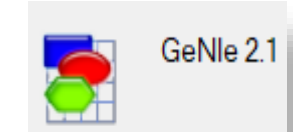
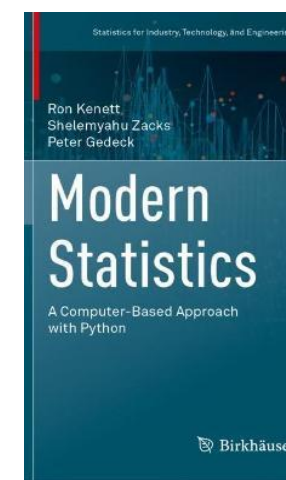
Mean and Treatment Effect: Annual Earnings - 2009 Prices NIS



*Shaded area indicates two sided confidence intervals, 10% significance level.

8.3 Bayesian Networks

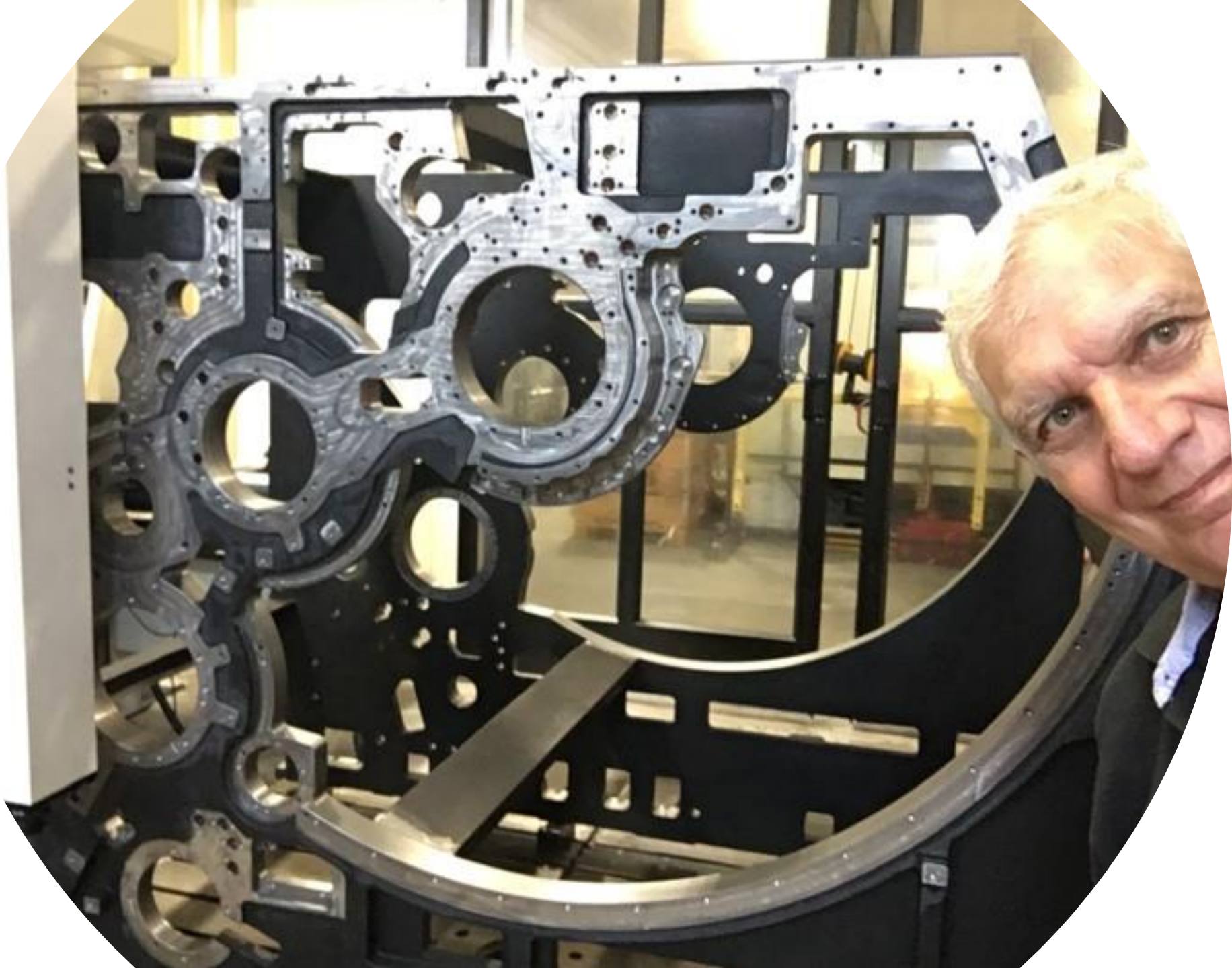
Bayesian networks (BNs) were introduced in Sect. 2.1.6. They implement a graphical model structure known as a directed acyclic graph (DAG) that is popular in statistics, machine learning, and artificial intelligence. BNs enable an effective representation and computation of the joint probability distribution over a set of random variables (Pearl 1985). The structure of a DAG is defined by two sets: the set of nodes and the set of directed arcs; arcs are often also called edges. The nodes represent random variables and are drawn as circles labeled by the variable names. The arcs represent links among the variables and are represented by arrows between nodes. In particular, an arc from node X_i to node X_j represents a relation between the corresponding variables. Thus, an arrow indicates that a value taken by variable X_j depends on the value taken by variable X_i . This property is used to reduce the number of parameters that are required to characterize the joint probability distribution (JPD) of the variables. This reduction provides an efficient way to compute the posterior probabilities given the evidence present in the data



<https://www.mdpi.com/1660-4601/19/8/4859/pdf>

<https://libri.unimi.it/index.php/milanoup/catalog/download/73/155/653-1>

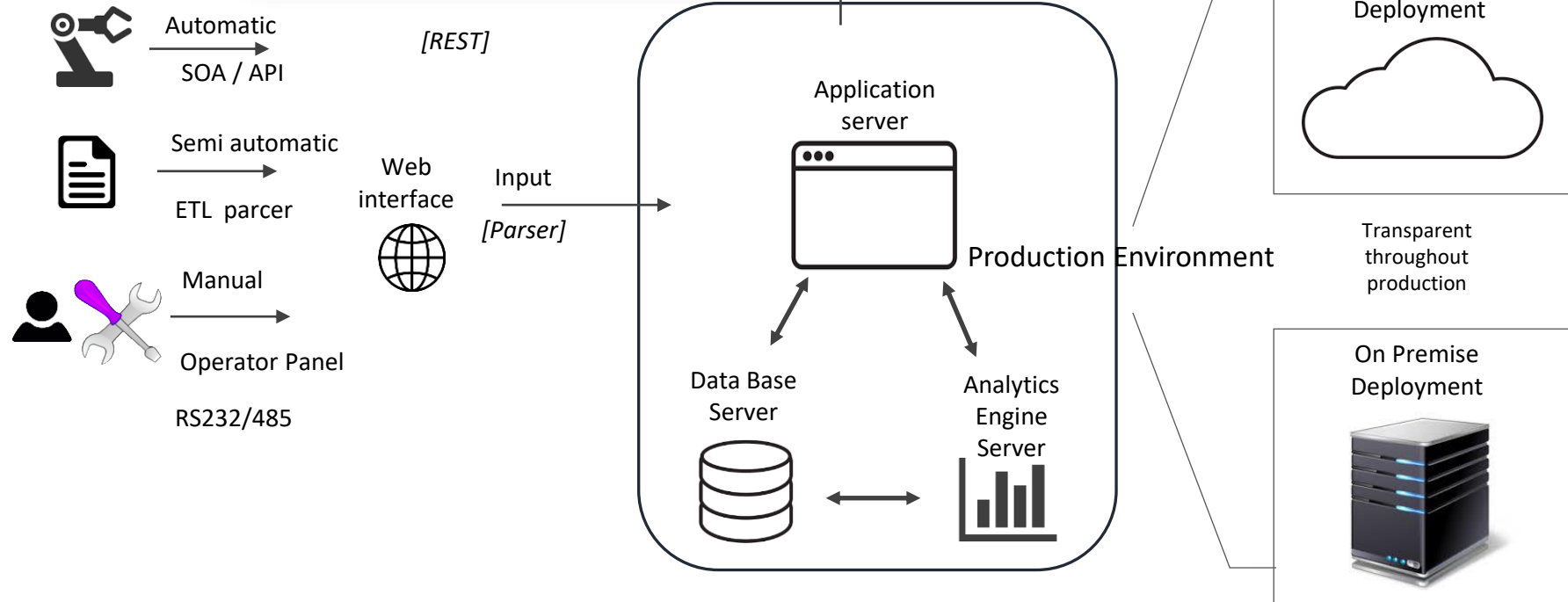
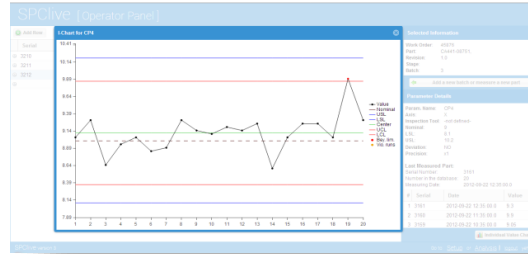
Industrial
applications



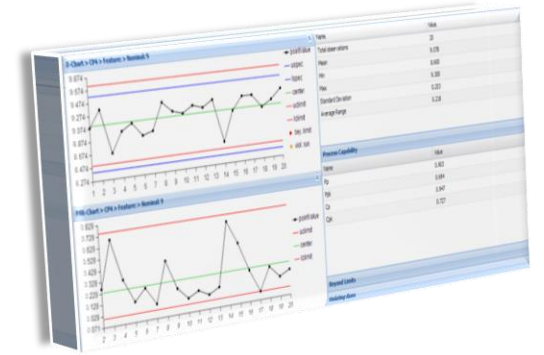
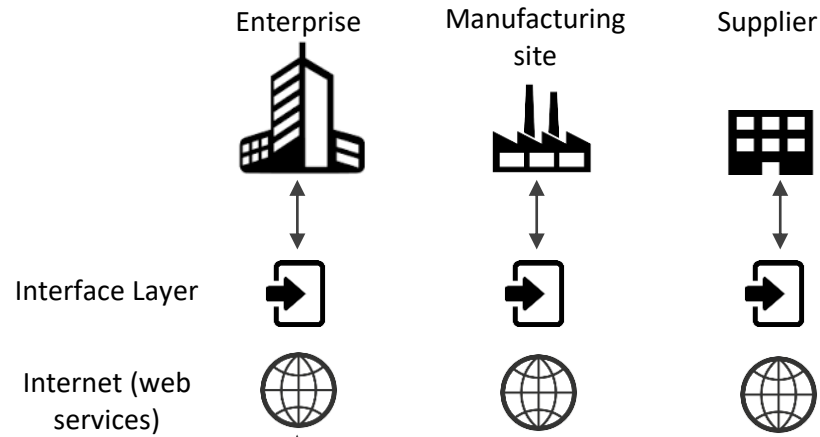
Data flow and core components



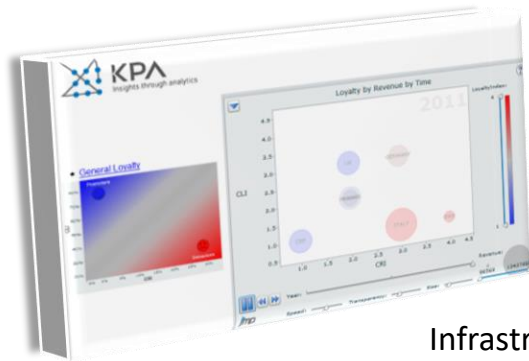
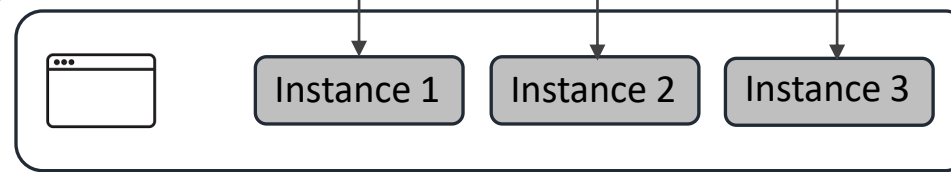
ET.3120.A.038-P_REV.03_long		
5/9/2012 7:14		
(mm)	ACTUAL	NOMINAL LO-TOL
Temperature Compensation: OFF		
Temperature Compensation: OFF		
Temperature Compensation: OFF		
=====CAGE CRITICAL DIMENSIONS=====		
===== Datum Planes Definition =====		
Plane B =====		
Plane:PLNB		
Flatness	0.0484	0.1
Perpendicular	0.0489	0.1 DAT(datA)
Plane C =====		
Plane:PLNC		
Perpendicular	0.002	0.05 DAT(datA)
Flatness	0.0019	0.03
Perpendicular	0.0055	0.05 DAT(datB)
Plane D =====		
Plane:PLND		
Flatness	0.0586	0.1
Perpendicular	0.0592	0.1 DAT(datA)
Parallelism	0.0903	0.15 DAT(datB)



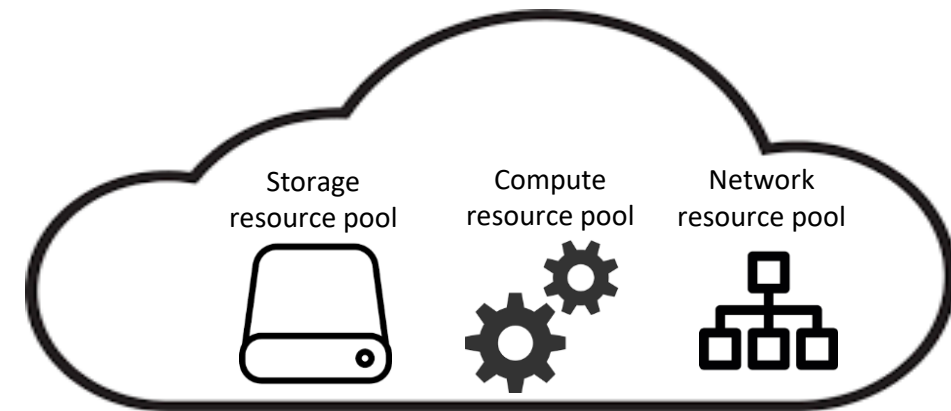
Cloud hosted SaaS



Application Layer



Infrastructure Layer



Security Paradigm

SPC Operator Panel

Work Order / Part Number Selection

Work Order:

Part Number:

Details selection

Batch:

Manufacturing Machine:

Operator:

Shift:

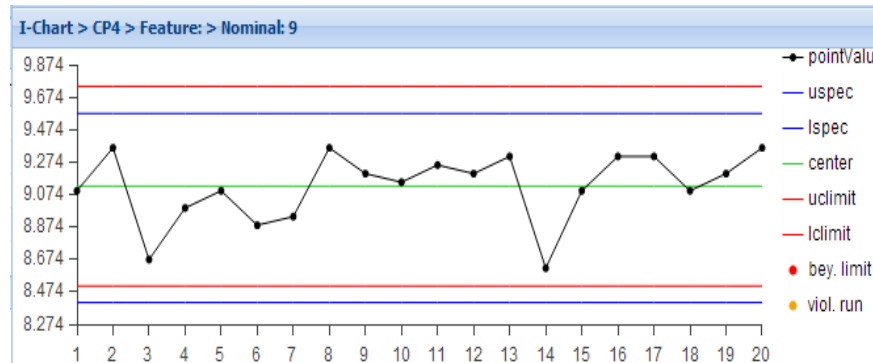
Inspection Tool:

Production:

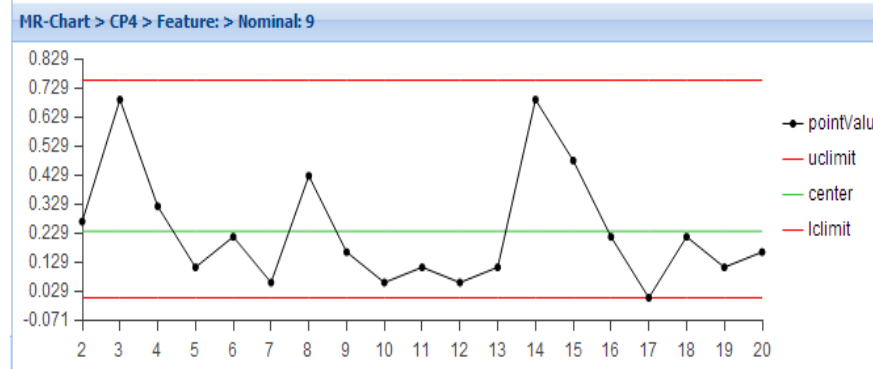


Serial	CP1 - Z	CP2 - X	CP2 - Y	CP2 - Z
⊖	0.5	0.03	0.03	0.03
⊖	0.55	0.025	0.026	0.028
⊖				

Monitoring, Diagnostics and Prognostics



Name	Value
Total observations	20
Mean	9.078
Min	8.600
Max	9.300
Standard Deviation	0.203
Average Range	0.218

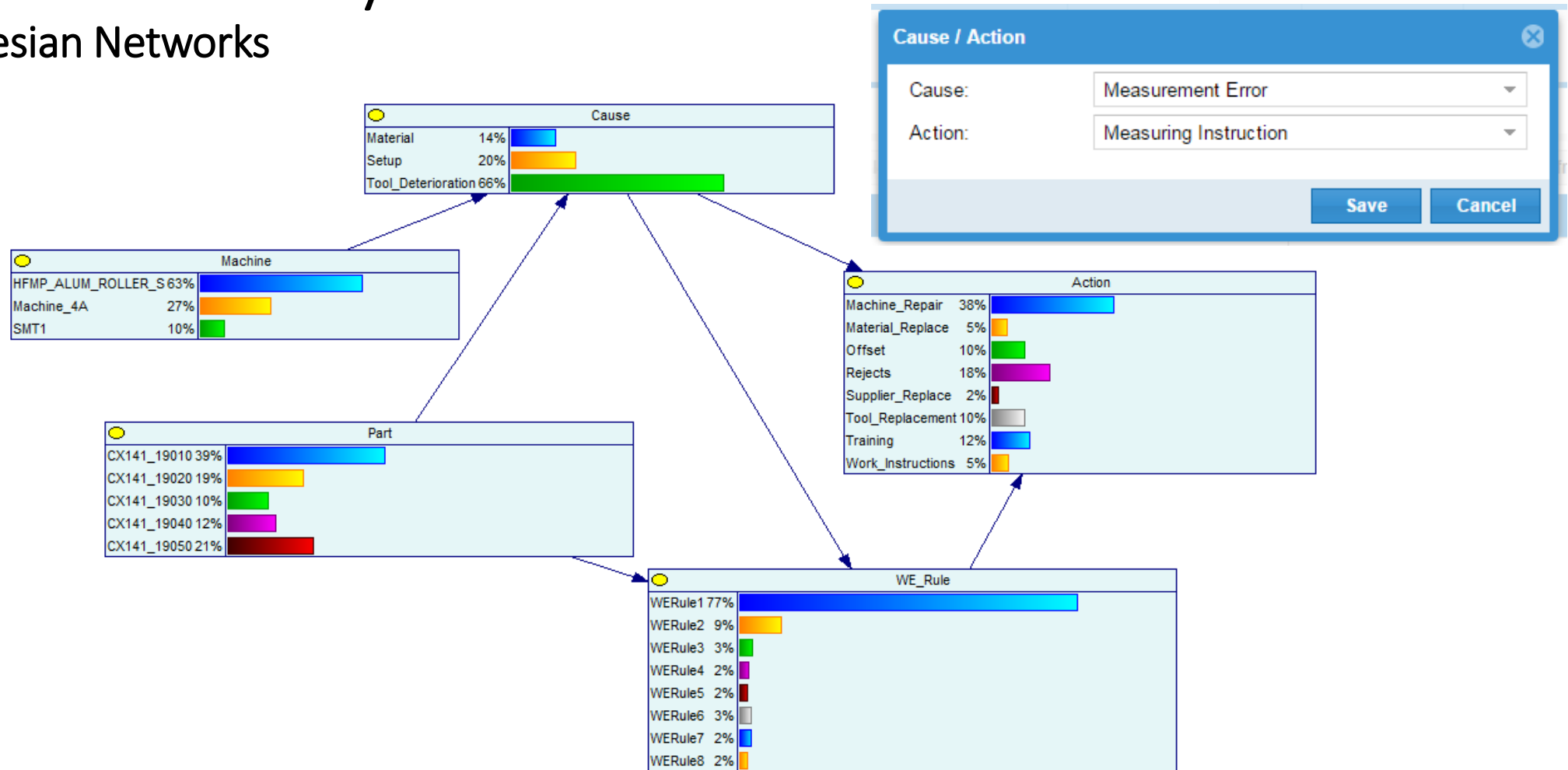


Process Capability	
Name	Value
Pp	0.903
Ppk	0.694
Cp	0.947
Cpk	0.727

Beyond Limits	
Violating Runs	

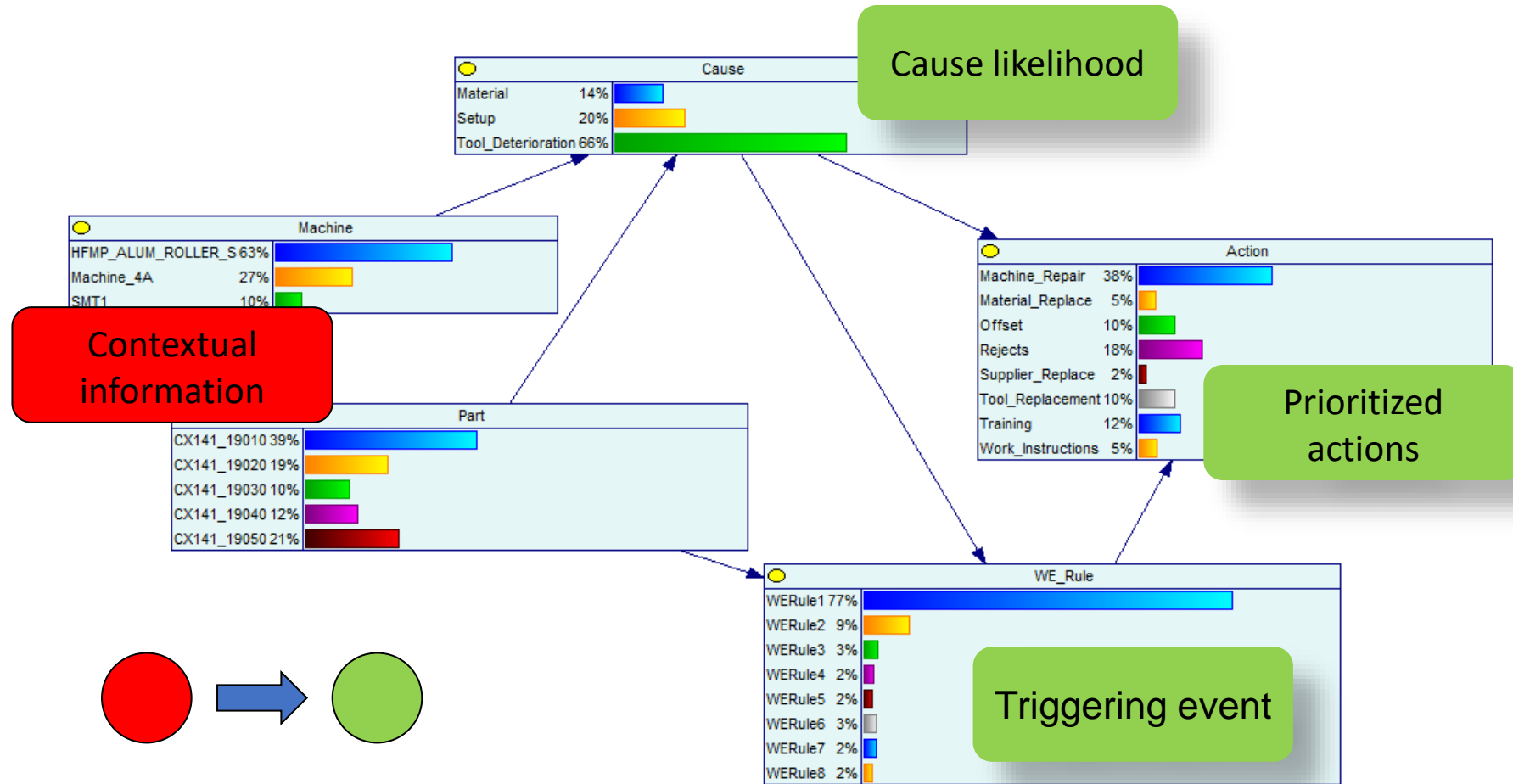
Predictive Analytics

Using Bayesian Networks



Predictive Analytics

Using Bayesian Networks

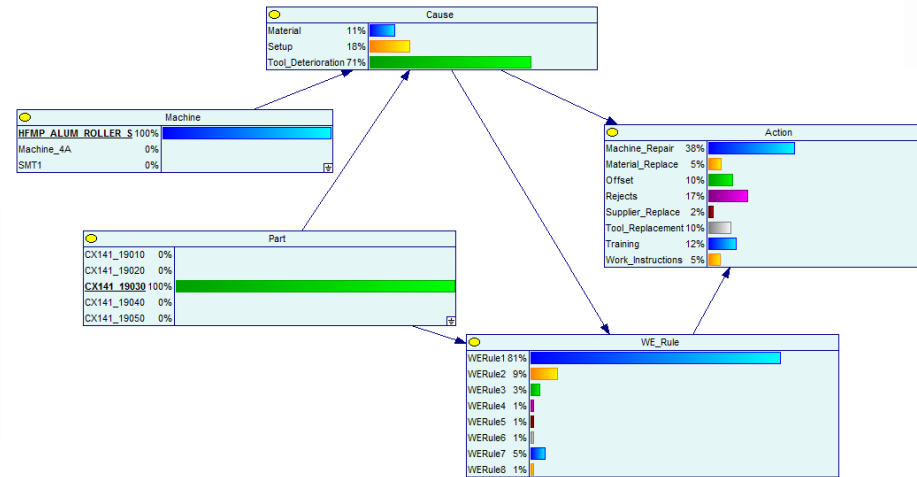


Predictive Analytics

Using Bayesian Networks

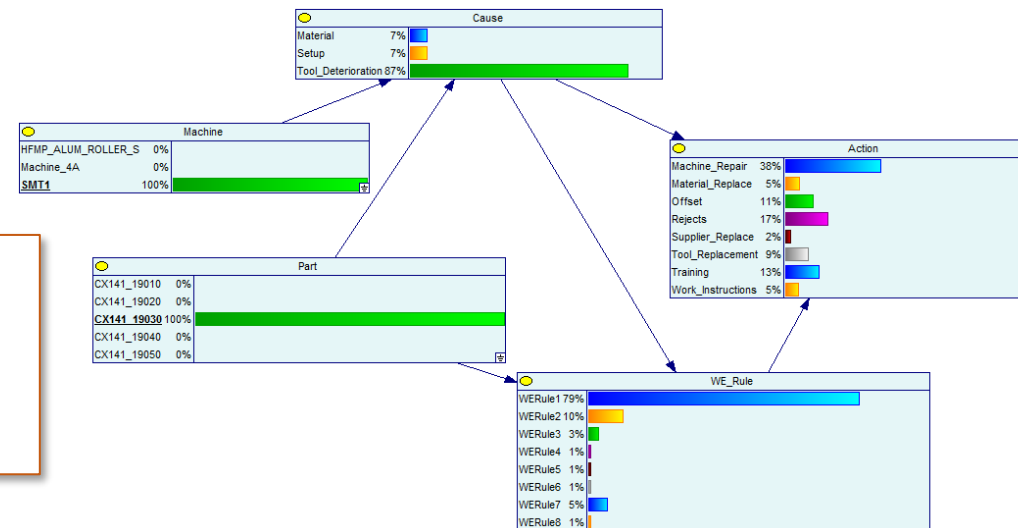
Machine **HFMP ALUM ROLLER S**
Part **CX141 19030**

- Setup issues 18% probability
- Tool issues 71% probability



Machine **SMT1**
Part **CX141 19030**

- Setup issues 7% probability
- Tool issues 87% probability



Chapter 9 Reliability Analysis

Reliability

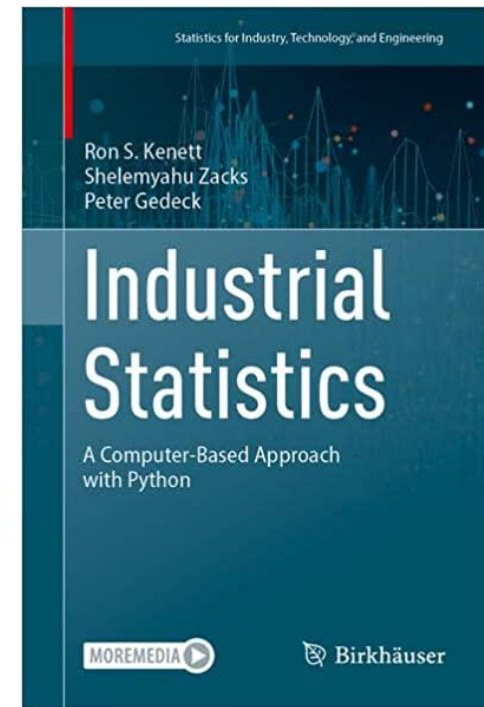
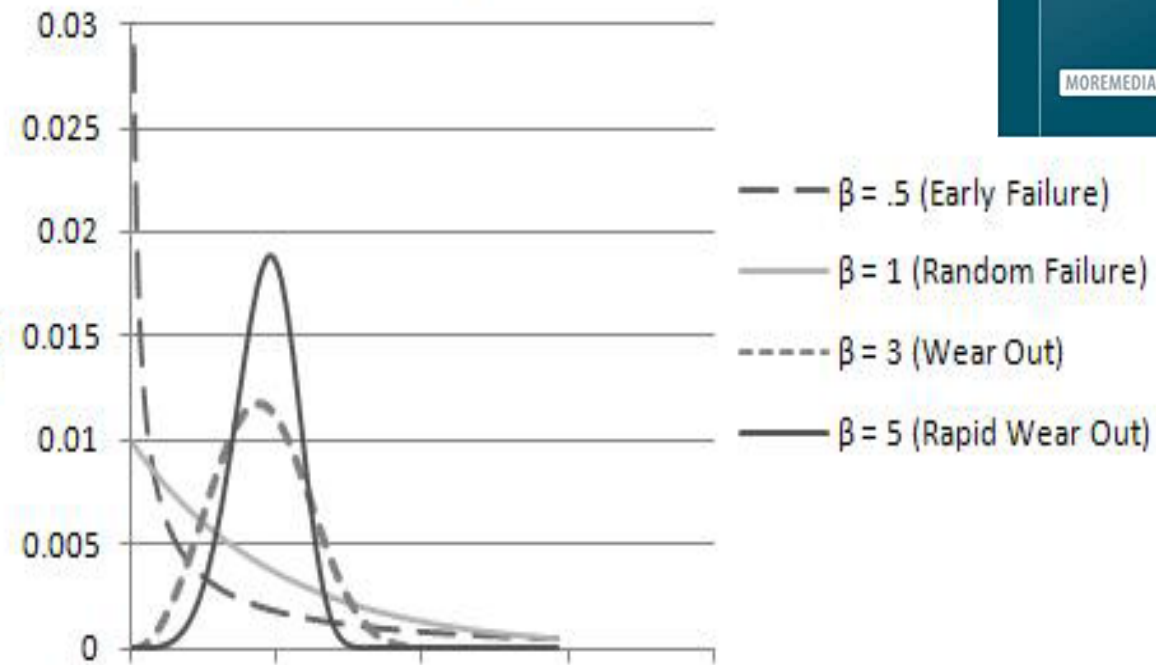
Preview The previous chapter dwelled on design decisions of product and process developers that are aimed at optimizing the quality and robustness of products and processes. This chapter is looking and performance over time and discusses basic notions of repairable and non repairable systems. Graphical and non parametric techniques are presented together with classical parametric techniques for estimating life distributions. Special sections cover reliability demonstration procedures, sequential reliability testing, burn-in procedures and accelerated life testing. Design and testing of reliability is a crucial activity of organizations adopting advanced quality and industrial standards discussed in Chapter 1.

Systems and products are considered to be of high quality, if they conform to their design specifications and appeal to the customer. However, products can fail, due to degradation over time or due to some instantaneous shock. A system or a component of a system is said to be **reliable** if it continues to function, according to specifications, for a long time. Reliability of a product is a dynamic notion, over time. We say that a product is highly reliable if the probability that it will function properly for a specified period, is close to 1. As will be defined later, the reliability function, $R(t)$, is the probability that a product will function in at least t units of time.

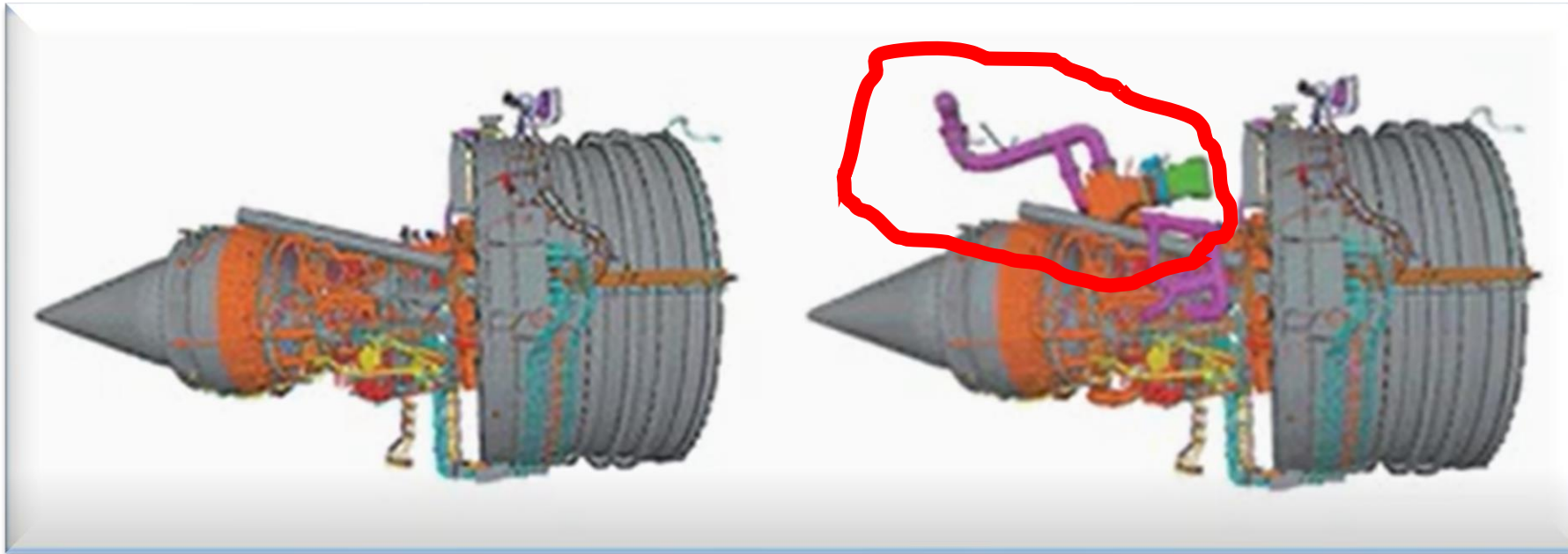
We distinguish between the reliability of systems which are unrepairable and that of repairable systems. A repairable system, after failure, goes through a period of repair and then returns to function normally. Highly reliable systems need less repair. Repairable systems which need less repair are more available to operate, and are therefore more desirable. **Availability** of a system at time t is the probability that the system will be up and running at time t . To increase the availability of repairable systems, maintenance procedures are devised. Maintenance schedules are designed to prevent failures of a system by periodic replacement of parts, tuning, cleaning, etc. It is very important to develop maintenance procedures, based on the reliability properties of the components of systems, which are cost effective and helpful to the availability of the systems. An alternative to scheduled maintenance is condition based maintenance (CBM) introduced in Chapters 1 and 8. Both sched-

Weibull Distribution

Effect of the Shape Parameter for $\eta = 100$



Engine with Bleed System

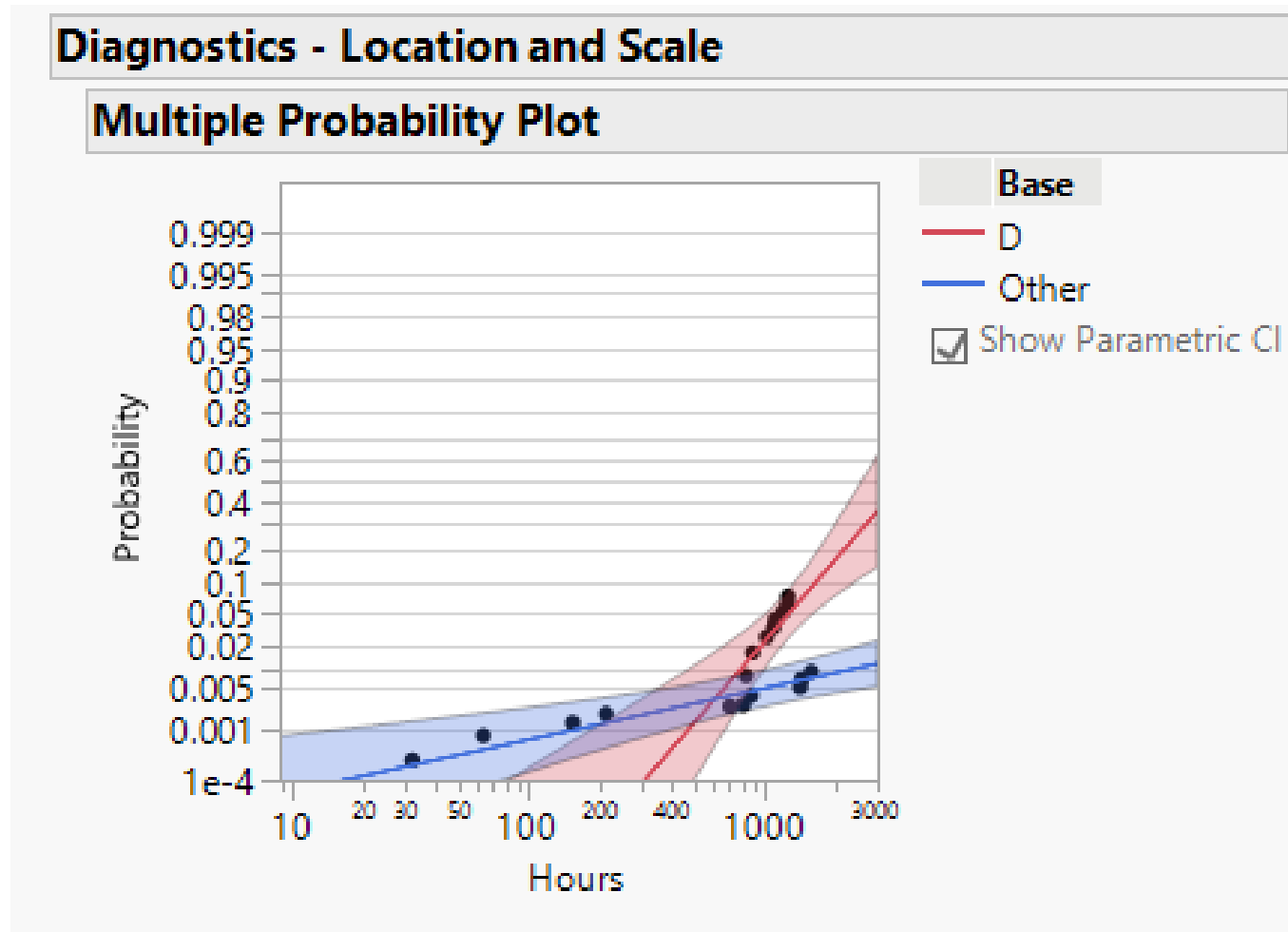


Engine Maintenance Data

The screenshot shows the JMP Pro interface with a data table titled 'Bleed'. The table has columns for 'Hours', 'Status', 'Frequency', and 'Base'. The data is organized into 19 rows. The 'Status' column contains values like 'Censored', 'Failed', and 'D'. The 'Base' column contains values like 'Other' and 'D'. The interface also shows a sidebar with project navigation options and a status bar at the bottom indicating 'evaluations done'.

	Hours	Status	Frequency	Base
1	12	Censored	39	Other
2	20	Censored	52	Other
3	30	Censored	46	Other
4	32	Failed	1	Other
5	50	Censored	31	Other
6	64	Failed	1	Other
7	85	Censored	48	Other
8	150	Censored	102	Other
9	153	Failed	1	Other
10	212	Failed	1	Other
11	250	Censored	158	Other
12	250	Censored	2	D
13	400	Censored	312	Other
14	550	Censored	101	Other
15	550	Censored	2	D
16	650	Censored	101	Other
17	650	Censored	2	D
18	708	Failed	1	D
19	750	Censored	100	Other

Engine Reliability Analysis



Engine Reliability Analysis

Statistics

Estimates

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
μ_0	8.255611	0.2978409	7.671853	8.839368
σ_0	0.679083	0.1675767	0.350639	1.007527
μ_1	16.149882	2.5560796	11.140058	21.159706
σ_1	3.594191	0.9726201	1.687890	5.500491

Nested Model Tests

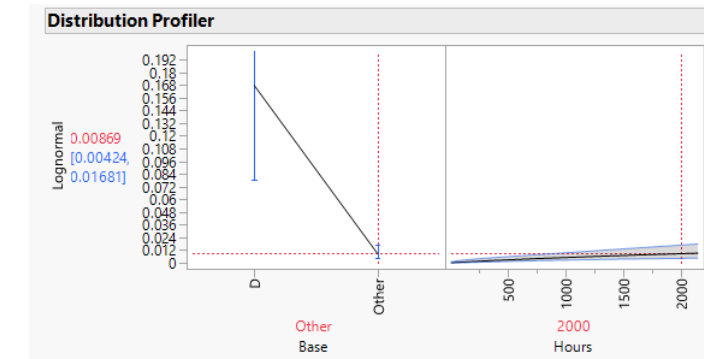
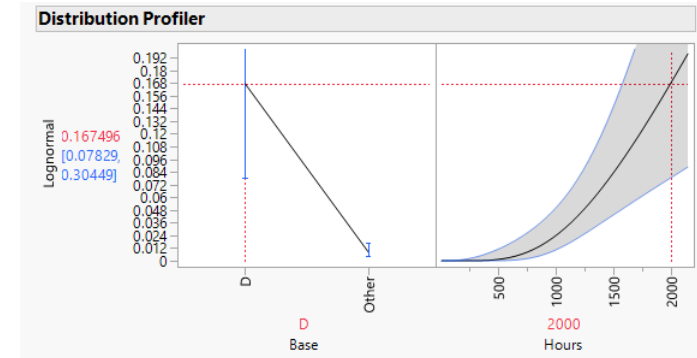
Models

Diagnostics	Model	-2 LogLikelihood	AICc	BIC	Number of Parameters
<input type="checkbox"/>	No Effect	478.8877	482.893	494.3304	2
<input checked="" type="checkbox"/>	Location	463.8969	469.9075	487.0609	3
	Location and Scale	446.898	454.9157	477.7834	4

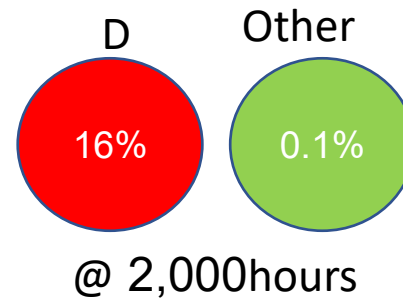
Diagnostics - Location and Scale

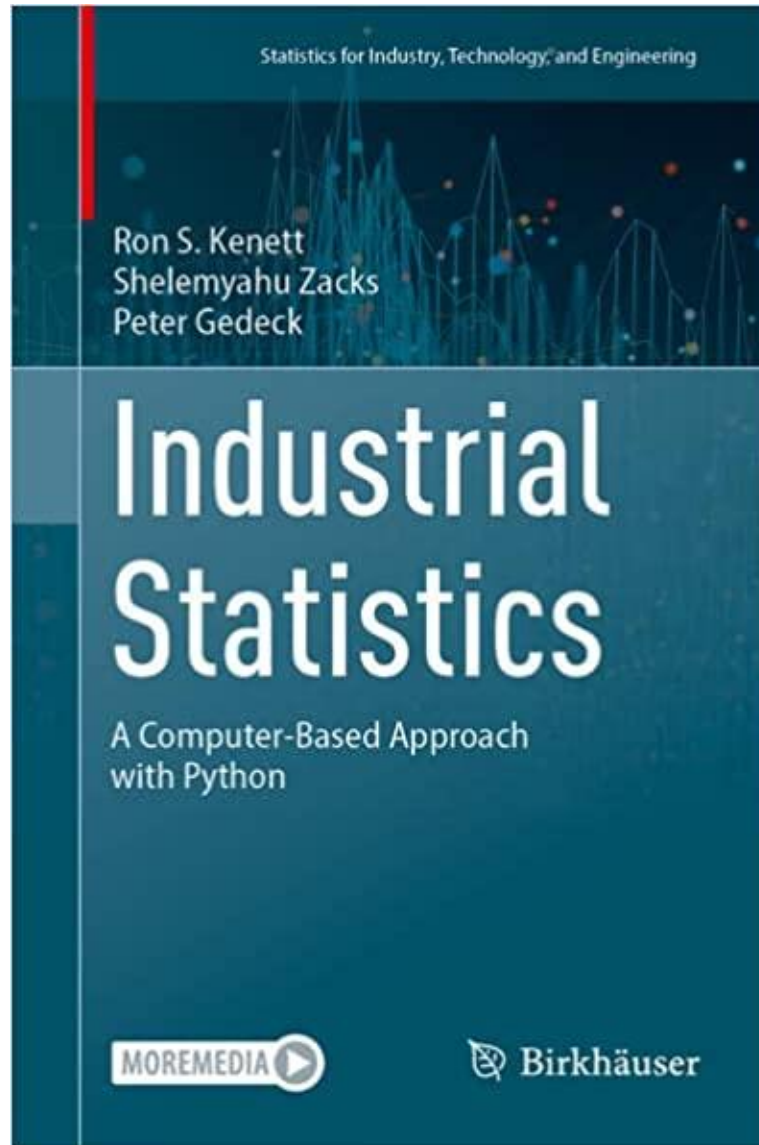
Diagnostics - Location

Engine Maintenance Improvement



Fix



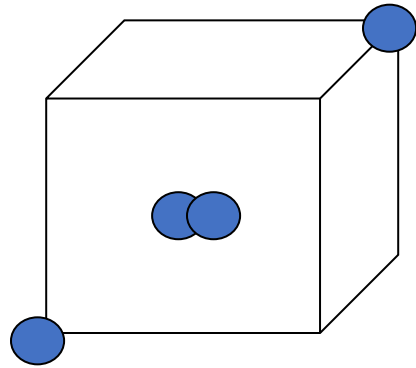


Chapter 5 Classical Design and Analysis of Experiments

Design of Experiments

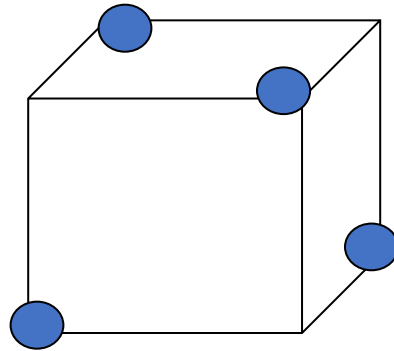
Preview Experiments are used in industry to improve productivity, reduce variability, enhance quality and obtain robust products and manufacturing processes. In this chapter we study how to design and analyze experiments which are aimed at testing scientific or technological hypotheses. These hypotheses are concerned with the effects of procedures or treatments on quality and productivity; or the general relationship between variables. Designed experiments help determine the conditions under which a production process yields maximum output or other optimum results, etc. The chapter presents the classical methods of design of experiments. It starts with an introductory section with examples and discusses guiding principles in designing experiments. The chapter covers the range of classical experimental designs including complete block designs, Latin squares, full and fractional factorial designs with factors at two and three levels. The basic approach to the analysis is through modeling the response variable and computing ANOVA tables. Particular attention is given to the generation of designs using Python.

Overview of DOE strategy



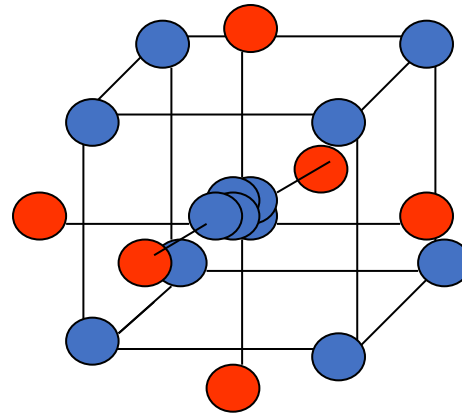
Scoping

Initial
assessment



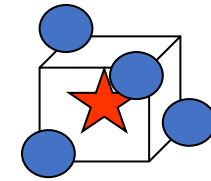
Screening

Fractional
designs



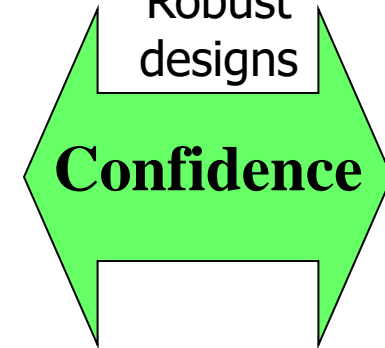
Optimizing

Response
surfaces

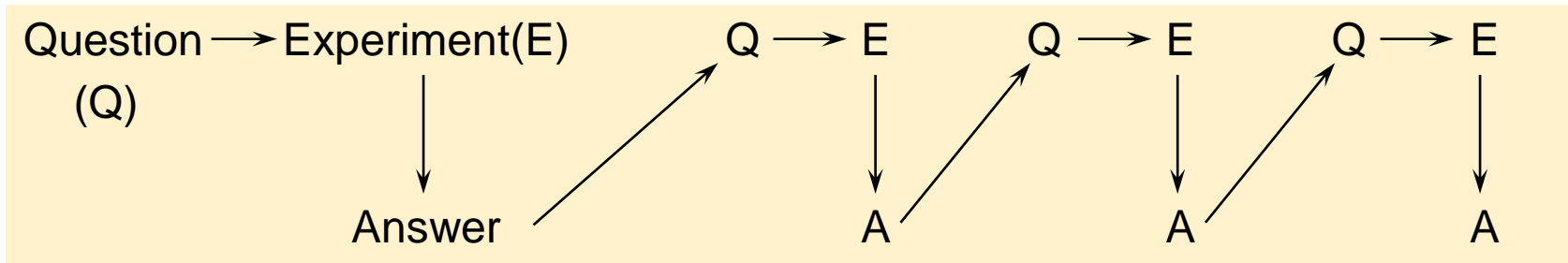


Robustness

Robust
designs



Ask a Series of Questions



Screen: Which are the biggest factors?

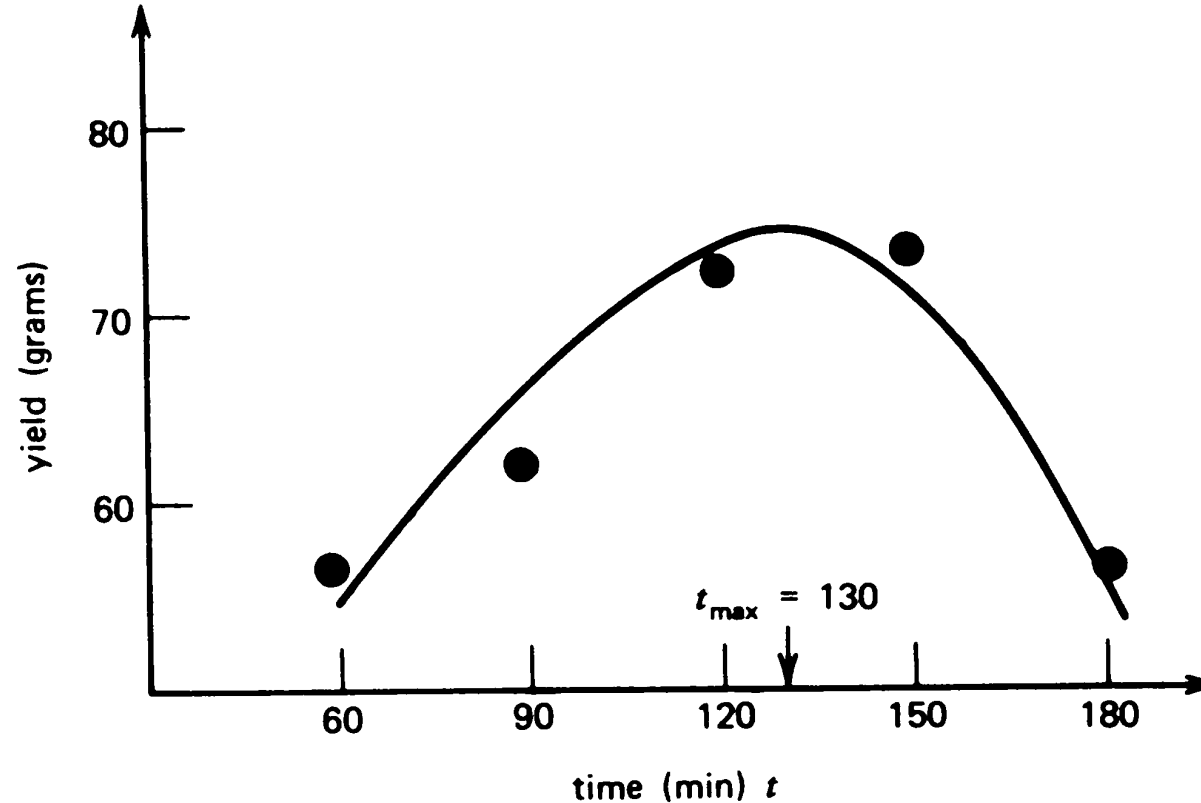
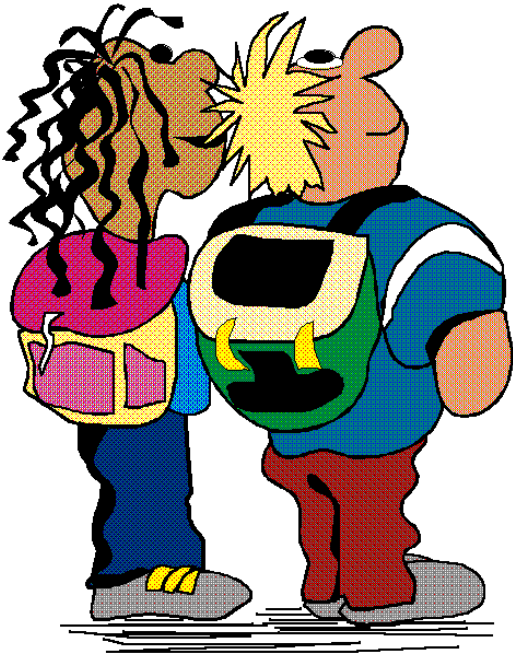
Characterize: What is the relationship between responses and factors?

Optimize: What is the best setting of the factors?

Verify: Are the results repeatable?

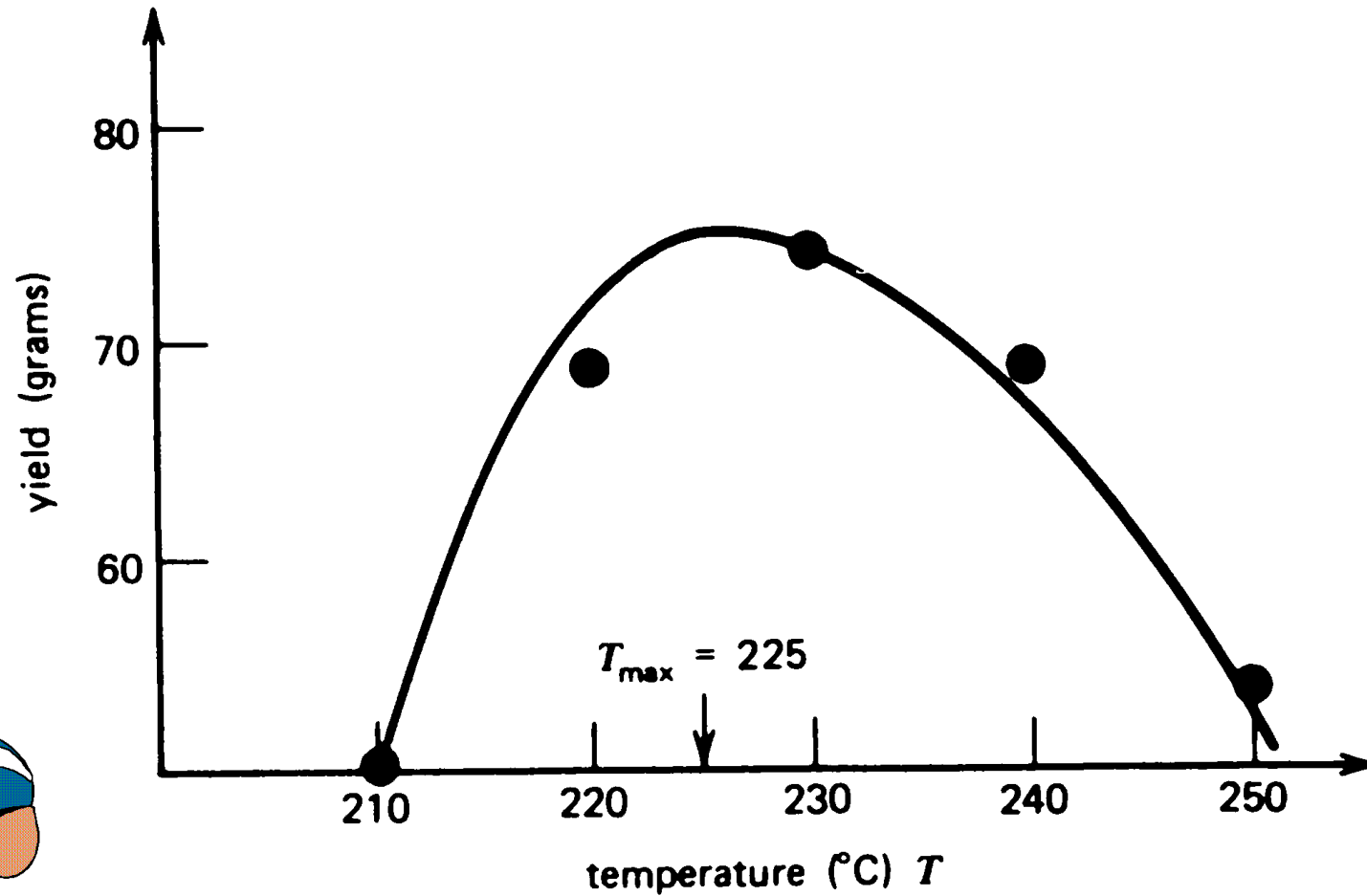
One Factor at a Time Experiments

OFAT



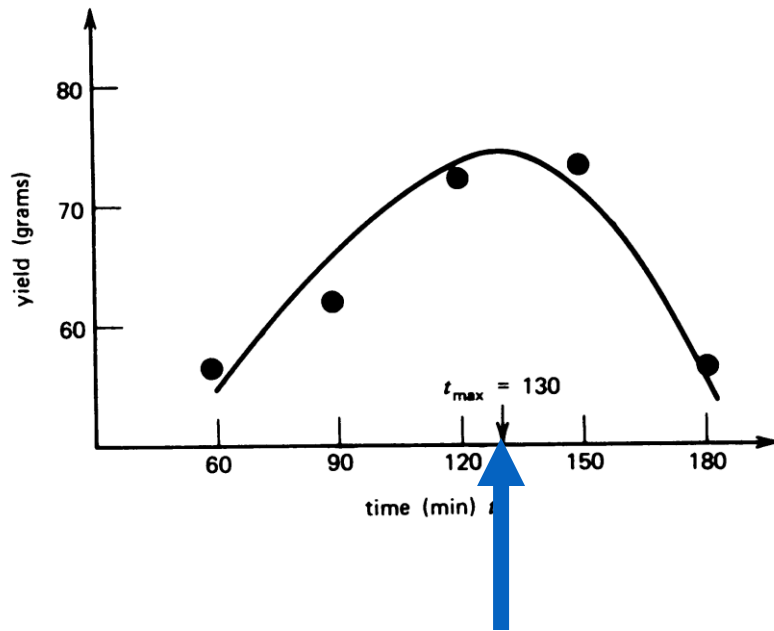
Effects of Reaction Time on Yield
(Reaction Temperature held fixed at 225° C)

One Factor at a Time Experiments

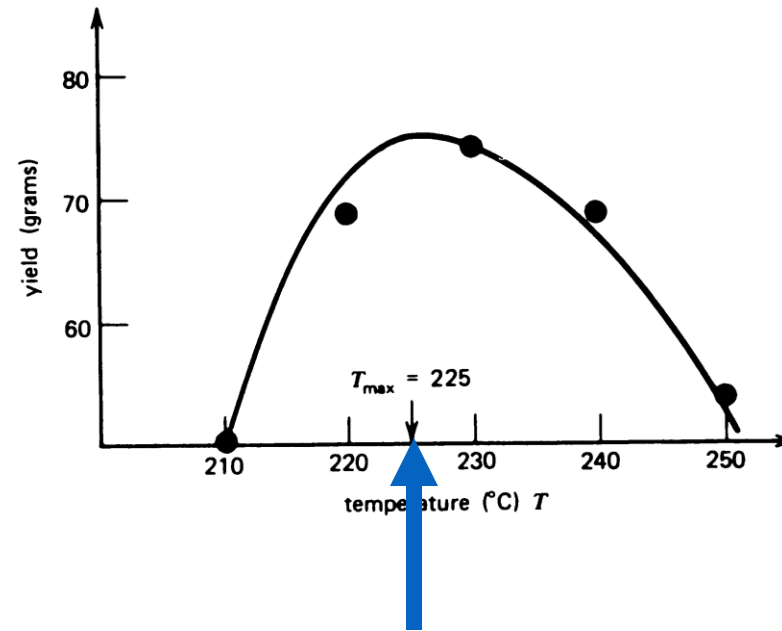


Effects of Reaction Temperature on Yield
(Reaction Time held fixed at 130 minutes)

One Factor at a Time Experiments

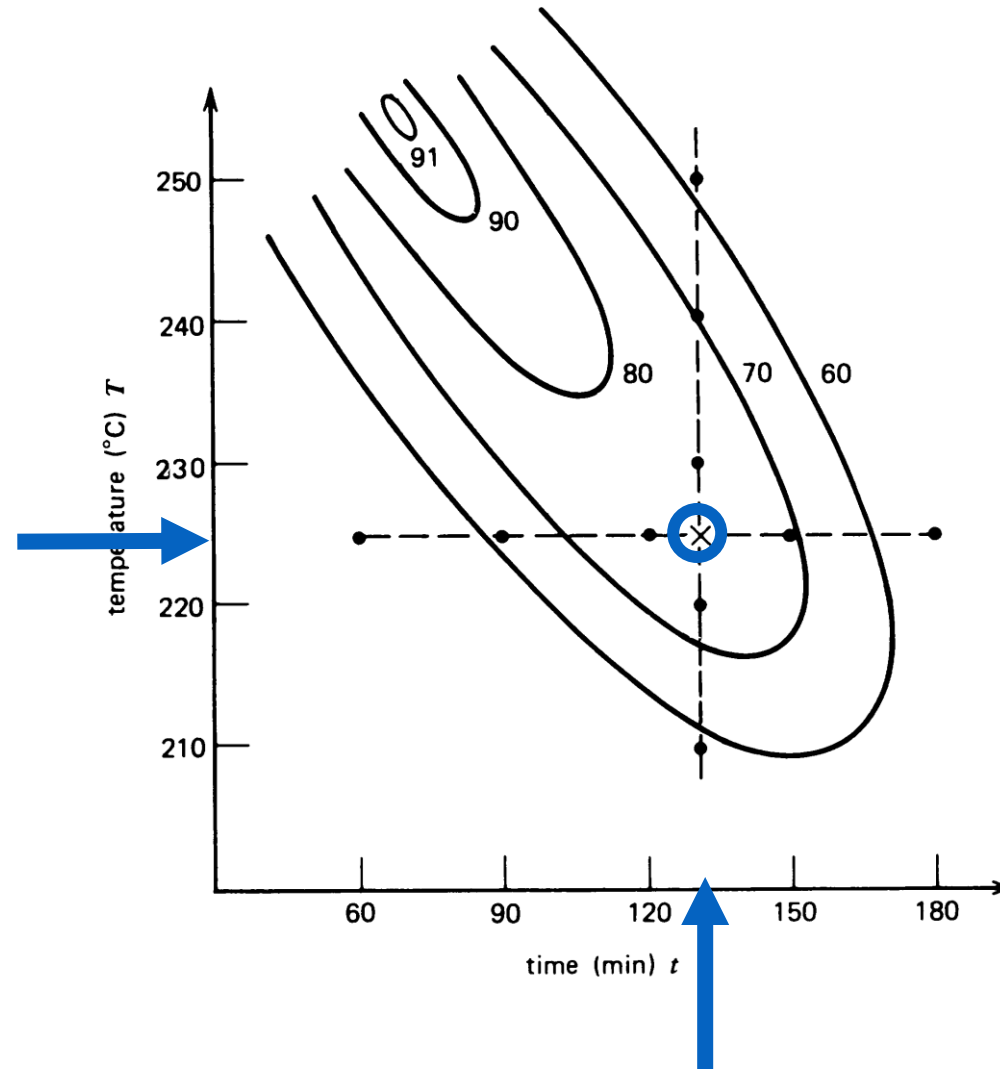


Effects of Reaction Time on Yield
(Reaction Temperature held fixed at 225° C)



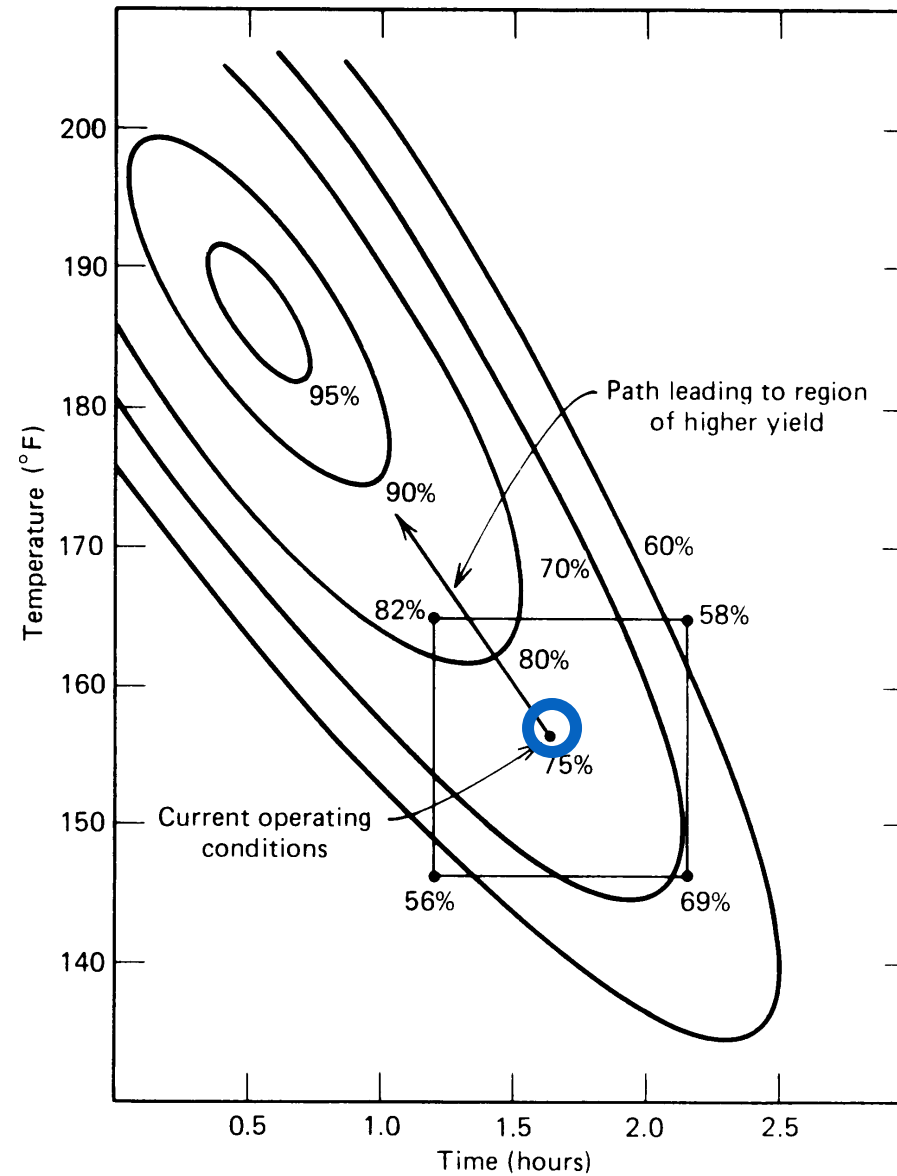
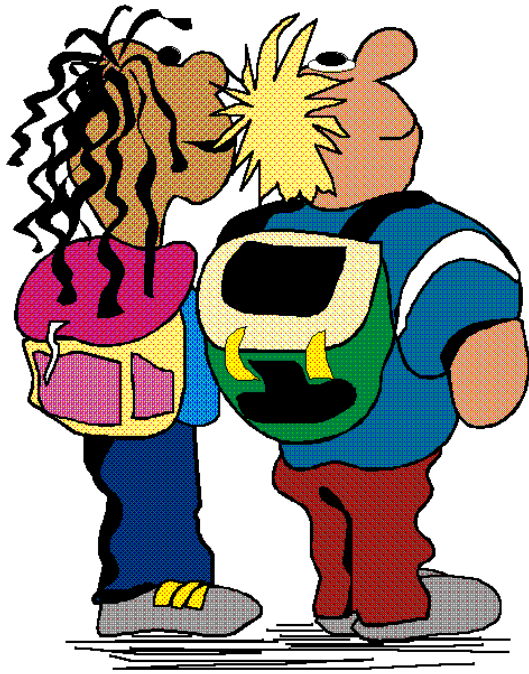
Study Effects of Reaction Temperature on Yield
(Reaction Time held fixed at 130 minutes)

One Factor at a Time Experiments



Reaching the Top

DoE



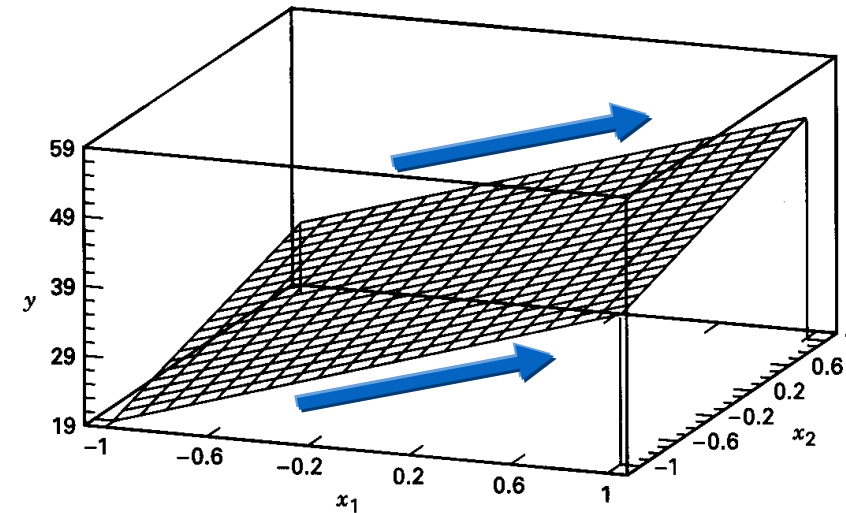
Regression Models

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

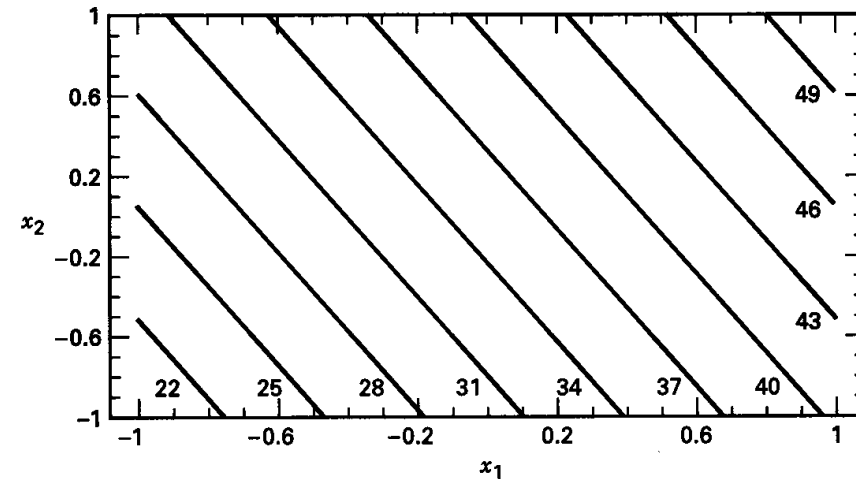
The least squares fit is

$$\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 0.5x_1x_2$$

$$\cong 35.5 + 10.5x_1 + 5.5x_2$$



(a) The response surface



(b) The contour plot

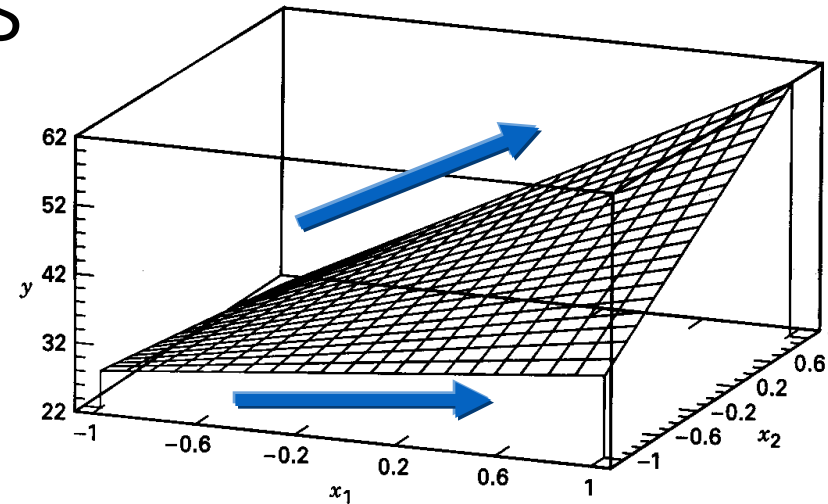
Figure 5-5 Response surface and contour plot for the model $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 0.5x_1x_2$.

The Effect of Interactions

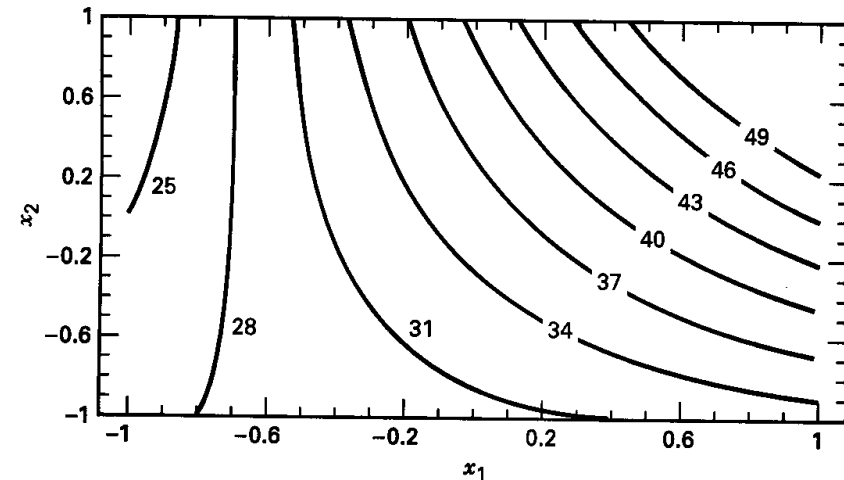
Suppose that we add an interaction term to the model:

$$\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 8x_1x_2$$

Interaction is a form of **curvature**



(a) The response surface

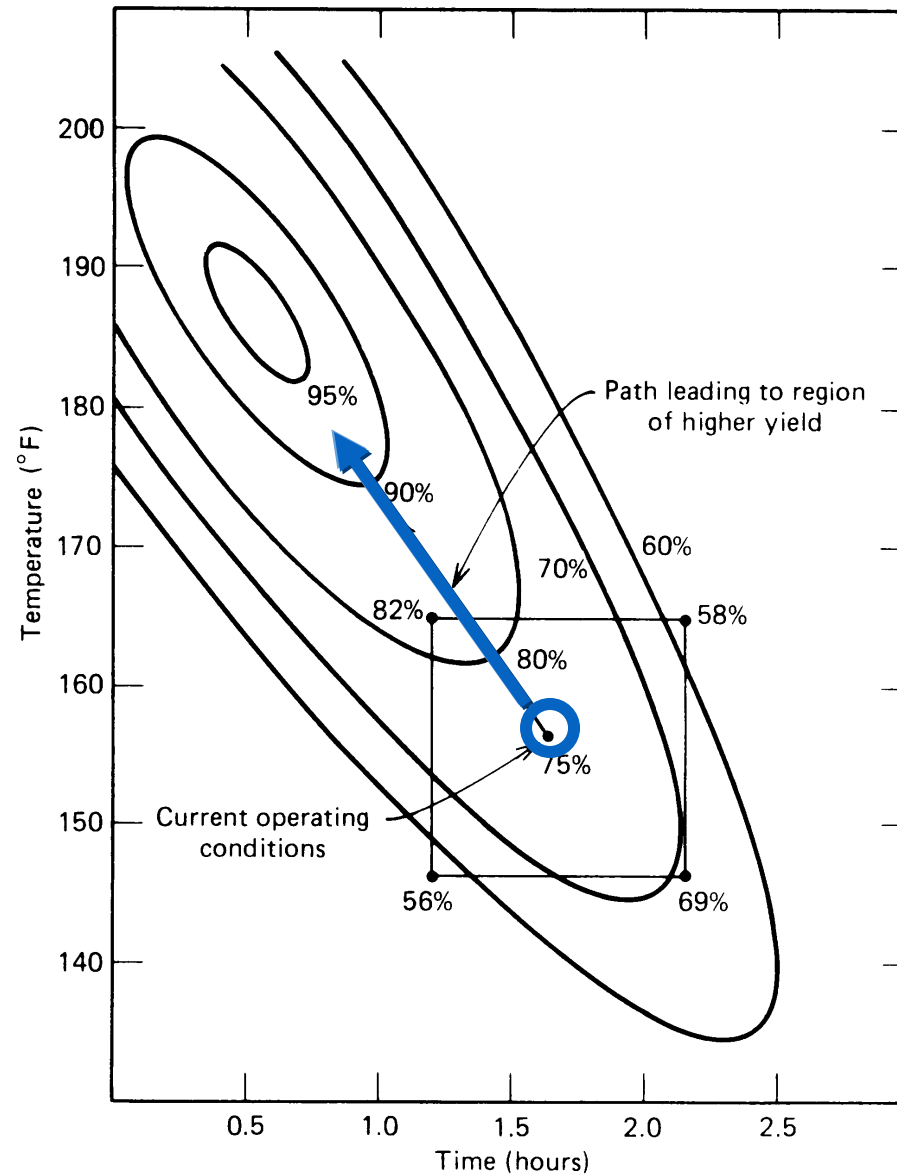
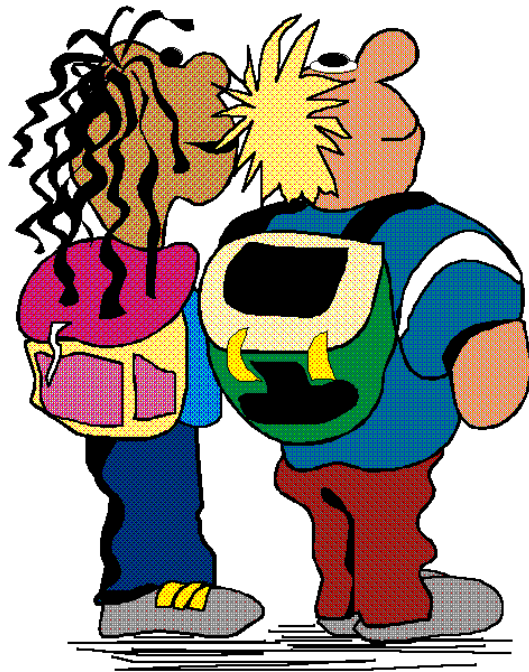


(b) The contour plot

Figure 5-6 Response surface and contour plot for the model $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 8x_1x_2$.

Steepest Ascent

DoE

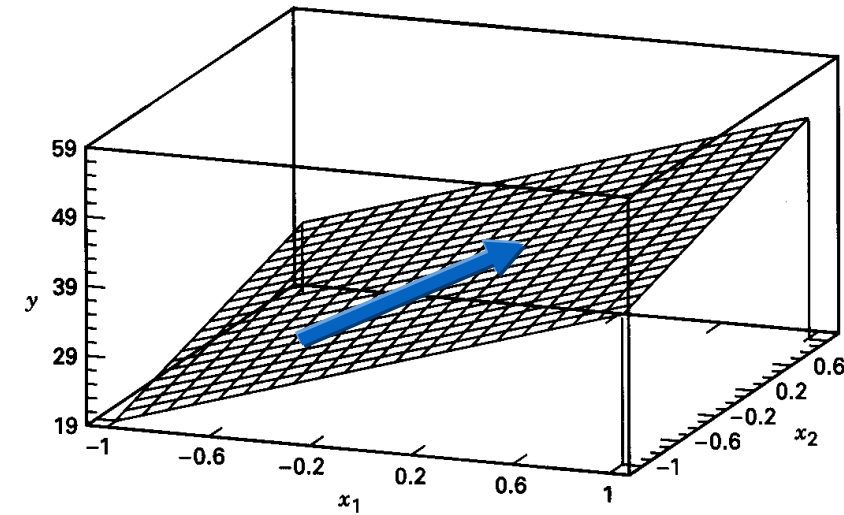


Steepest Ascent

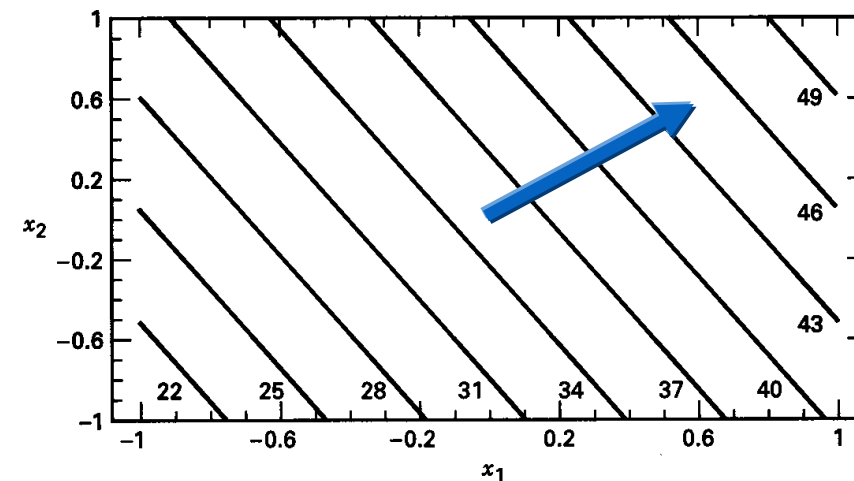
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

The least squares fit is

$$\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 0.5x_1x_2$$
$$\cong 35.5 + 10.5x_1 + 5.5x_2$$



(a) The response surface



(b) The contour plot

Figure 5-5 Response surface and contour plot for the model $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2$.

Basic Steps and Guiding Principles

1. The **objectives** of a study should be well stated, and criteria established to test whether these objectives have been met.
2. The **response** variable(s) should be clearly defined so that the study objectives are properly translated to measurable variables. At this stage measurement uncertainty should be established.
3. All **factors** which might affect the response variable(s) should be listed, and specified. We call these the controllable factors. This requires interactive brainstorming with content experts.
4. The type of **measurements** or observations on all variables should be specified.
5. The **levels** of the controllable factors to be tested should be determined.

6. A statistical **model** should be formulated concerning the relationship between the pertinent variables, and their error distributions. This can rely on prior knowledge or literature search.
7. An experimental **layout** or experimental array should be designed so that the inference from the gathered data will be: a. valid; b. precise; c. generalizable; d. easy to obtain.
8. The trials should be performed, if possible, in a **random** order, to avoid bias by factors which are not taken into consideration.
9. A **protocol** of execution should be prepared, including the method of analysis. The method of analysis and data collection depends on the design.
10. The **execution** of the experiment should carefully follow the protocol with proper documentation.
11. The results of the experiments should be carefully analyzed and reported ensuring proper documentation and traceability. Modern technology can ensure that data, analysis and conclusions are fully integrated and reproducible.
12. **Confirmatory** experiments should be conducted, to validate the inference (conclusions) of the experiments.

and $T_2 T_1$, giving each equal probability. The full discussion of this process of randomization is deferred to Chapter 5.

A typical arrangement of treatments resulting from such a randomization is shown in Table 3.1 together with some fictitious observations. For each pair of units the difference between the observation on T_2 and the observation on T_1 is calculated. The treatment effect is estimated by \bar{d} , the mean of these differences, and the estimated standard error of \bar{d} , and a test of the statistical significance of \bar{d} can be obtained by simple standard statistical calculations (Goulden, 1952, p. 51), the amount of the uncontrolled variation being estimated from the observed dispersion of the differences in the last column of Table 3.1.

Cox, D.R. (1992) Planning of Experiments, Wiley

TABLE 3.1

PAIRED COMPARISON EXPERIMENT

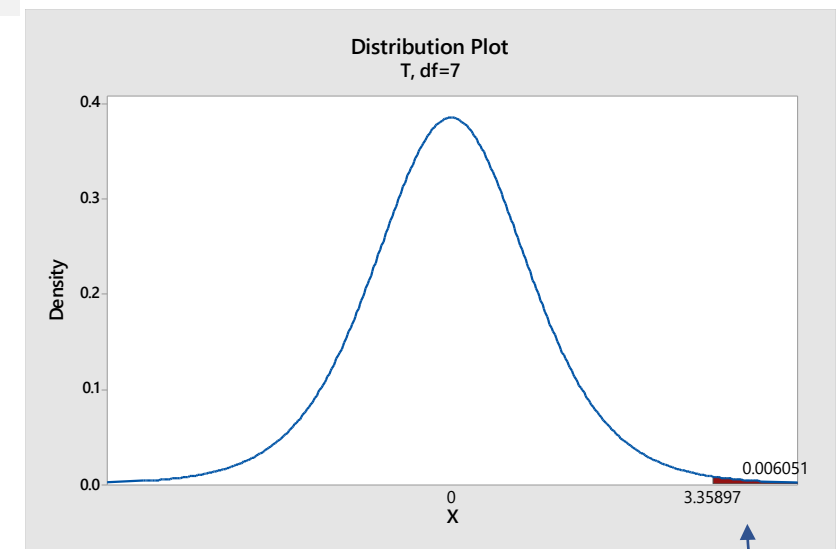
Day	First Unit	Second Unit	Difference, d
1	$T_1:2.8$	$T_2:3.2$	0.4
2	$T_2:3.1$	$T_1:3.1$	0.0
3	$T_2:3.4$	$T_1:2.9$	0.5
4	$T_1:3.0$	$T_2:3.5$	0.5
5	$T_2:2.7$	$T_1:2.4$	0.3
6	$T_2:2.9$	$T_1:3.0$	-0.1
7	$T_2:3.5$	$T_1:3.2$	0.3
8	$T_1:2.6$	$T_2:2.8$	0.2

Mean, $\bar{d} = 0.262$

Estimated standard error = 0.078

A treatment is applied
as T_1 or T_2 .

What is the treatment effect?
Is the effect at T_2 greater
than the effect at T_1 ?



5.4.1.2 Randomization Tests

A randomization test for paired comparison, constructs a **reference distribution** of all possible averages of the differences that can be obtained by randomly assigning the sign + or – to the value of D_i . It computes then an average difference \bar{D} for each one of the 2^n sign assignments.

The P -value of the test, for the two sided alternative, is determined according to this reference distribution, by

$$P = \Pr\{\bar{Y} \geq \text{Observed } \bar{D}\}.$$

For example, suppose we have four differences, with values 1.1, 0.3, –0.7, –0.1. The mean is $\bar{D}_4 = 0.15$. There are $2^4 = 16$ possible ways of assigning a sign to $|D_i|$. Let $X_i = \pm 1$ and $\bar{Y} = \frac{1}{4} \sum_{i=1}^4 X |D_i|$. The possible combinations are listed in Table 5.1

Table 5.1: Sign Assignments and Values of \bar{Y}

Signs				D
-1	-1	-1	-1	-0.55
1	-1	-1	-1	0
-1	1	-1	-1	-0.4
1	1	-1	-1	0.15
-1	-1	1	-1	-0.20
1	-1	1	-1	0.35
-1	1	1	-1	-0.05
1	1	1	-1	0.50
-1	-1	-1	1	-0.50
1	-1	-1	1	0.05
-1	1	-1	1	-0.35
1	1	-1	1	0.2
-1	-1	1	1	-0.15
1	-1	1	1	0.40
-1	1	1	1	0
1	1	1	1	0.55

Under the reference distribution, all these possible means are equally probable. The P -value associated with the observed $\bar{D} = 0.15$ is $P = \frac{7}{15} = 0.47$. If the number of pairs (blocks) n is large the procedure becomes cumbersome, since we have to

determine all the 2^n sign assignments. If $n = 20$ there are $2^{20} = 1,048,576$ such assignments. We can however estimate the P -value by taking a RSWR from this reference distribution. In Python this is performed with the following commands:

```
random.seed(1)
X = [1.1, 0.3, -0.7, -0.1]
m = 20000

Di = pd.DataFrame([random.choices((-1, 1), k=len(X)) for _ in range(m)])
DiX = (Di * X)

np.mean(DiX.mean(axis=1) > np.mean(X))
```

| 0.31425

5.7 Full Factorial Experiments

5.7.1 The Structure of Factorial Experiments

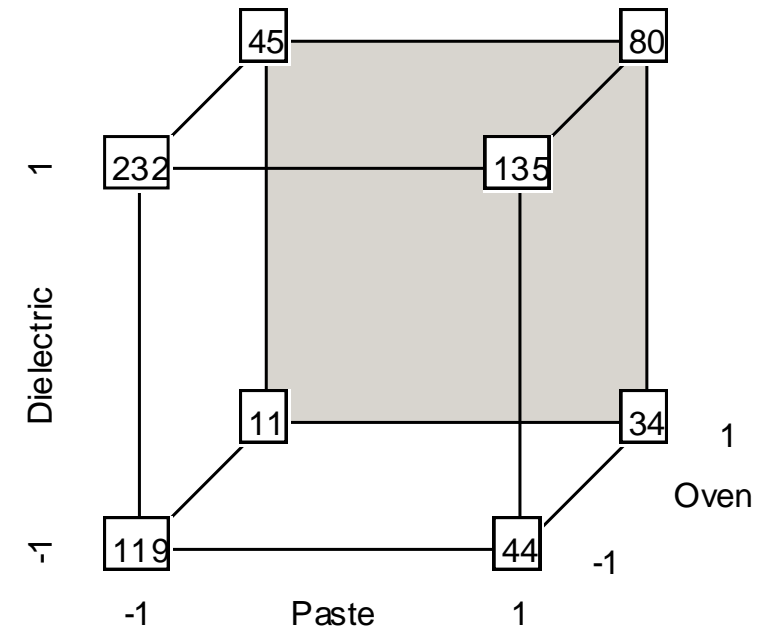
Full factorial experiments are those in which complete trials are performed of all the combinations of the various factors at all their levels. For example, if there are five factors, each one tested at three levels, there are altogether $3^5 = 243$ treatment combinations. All these 243 treatment combinations are tested. The full factorial experiment may also be replicated several times. The order of performing the trials is random.

In full factorial experiments, the number of levels of different factors do not have to be the same. Some factors might be tested at two levels and others at three or four levels. Full factorial, or certain fractional factorials which will be discussed later, are necessary, if the statistical model is not additive. In order to estimate or test the **effects** of **interactions**, one needs to perform factorial experiments, full or fractional. In a full factorial experiment, all the main effects and interactions can be tested or estimated. Recall that if there are p factors A, B, C, \dots there are p types of main effects, $\binom{p}{2}$ types of pairwise interactions AB, AC, BC, \dots , $\binom{p}{3}$ interactions between three factors, ABC, ABD, \dots and so on. On the whole there are, together with the grand mean μ , 2^p types of parameters.

In the following section we discuss the structure of the ANOVA for testing the significance of main effects and interaction. This is followed by a section on the estimation problem. In Sections [5.7.4](#) and [5.7.5](#) we discuss the structure of full factorial experiments with 2 and 3 levels per factor, respectively.

Response Resistance

Cube Plot



5.7.4 2^m Factorial Designs

2^m factorial designs are full factorials of m factors, each one at two levels. The levels of the factors are labelled as “Low” and “High” or 1 and 2. If the factors are categorical then the labelling of the levels is arbitrary and the values of the main effects and interaction parameters depend on this arbitrary labeling. We will discuss here experiments in which the levels of the factors are measured on a continuous scale, like in the case of the factors effecting the piston cycle time. The levels of the i -th factor ($i = 1, \dots, m$) are fixed at x_{i1} and x_{i2} , where $x_{i1} < x_{i2}$.

By simple transformation all factor levels can be reduced to

$$c_i = \begin{cases} +1, & \text{if } x = x_{i2} \\ -1, & \text{if } x = x_{i1} \end{cases}, \quad i = 1, \dots, m.$$

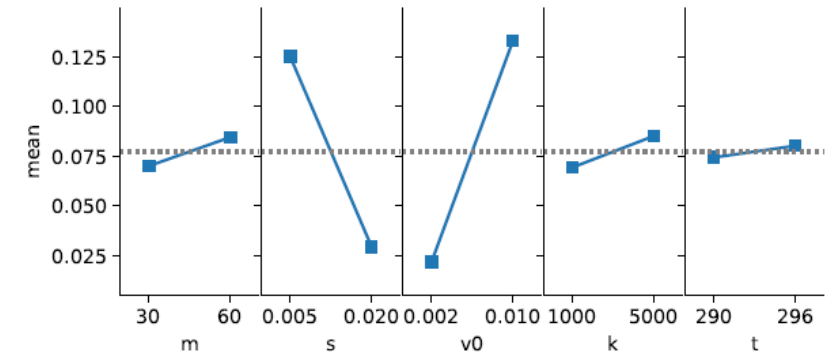
In such a factorial experiment there are 2^m possible treatment combinations. Let (i_1, \dots, i_m) denote a treatment combination, where i_1, \dots, i_m are indices, such that

$$i_j = \begin{cases} 0, & \text{if } c_i = -1 \\ 1, & \text{if } c_i = 1. \end{cases}$$

Thus, if there are $m = 3$ factors, the number of possible treatment combinations is $2^3 = 8$. These are given in Table [5.17](#)

Table 5.21: LSE of Main Effects and Interactions

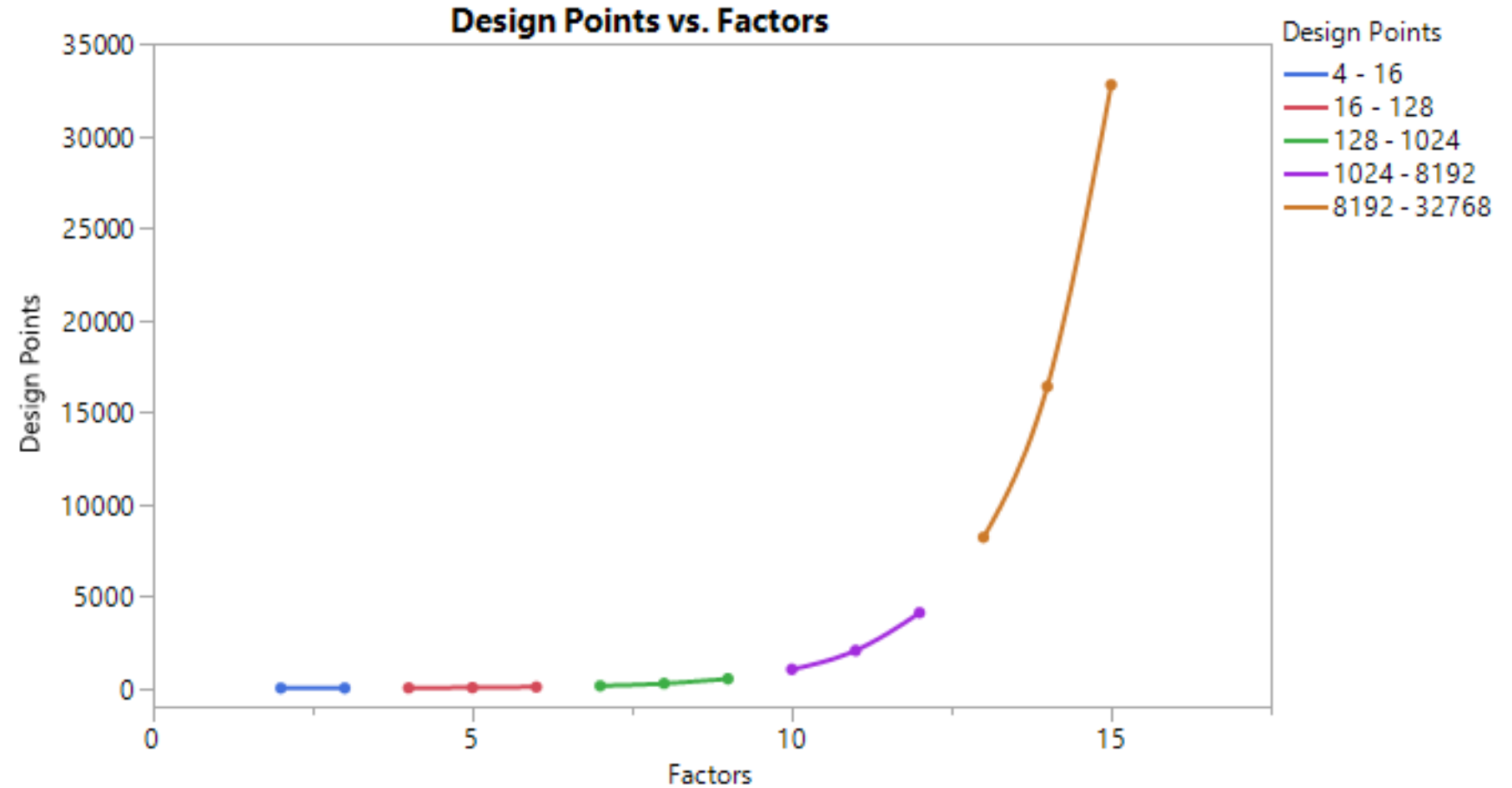
	LSE	S.E.	t	
m	-0.0054	0.00222	-2.44	
s	37.9277	0.00222	17097.52	**
v0	-61.9511	0.00222	-27927.08	**
k	-0.0000	0.00222	-0.00	
t	-0.0001	0.00222	-0.07	
m:s	-0.0413	0.00222	-18.62	**
m:v0	0.1332	0.00222	60.03	**
m:k	0.0000	0.00222	0.00	
m:t	0.0000	0.00222	0.01	
s:v0	-1165.8813	0.00222	-525569.93	**
s:k	-0.0004	0.00222	-0.17	
s:t	-0.1173	0.00222	-52.86	**
v0:k	0.0005	0.00222	0.21	
v0:t	0.2834	0.00222	127.76	**
k:t	0.0000	0.00222	0.00	



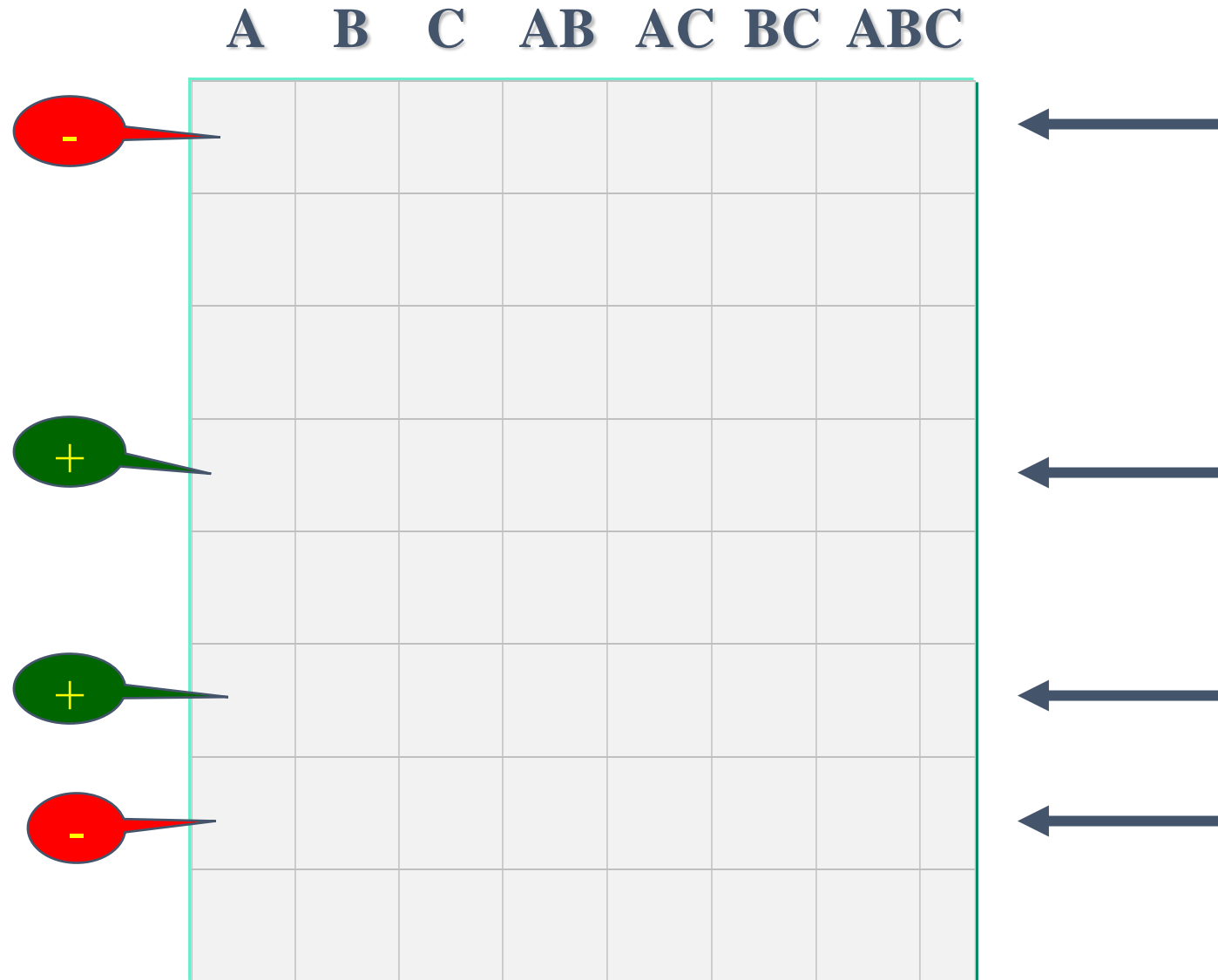
Size of Full Factorial Designs



Graph Builder



Fractional Factorial Design



is in practice not any way.

A full factorial experiment is a combination of fractional factorial designs. In Python we obtain a fraction of a full factorial design with the `mistat` package.

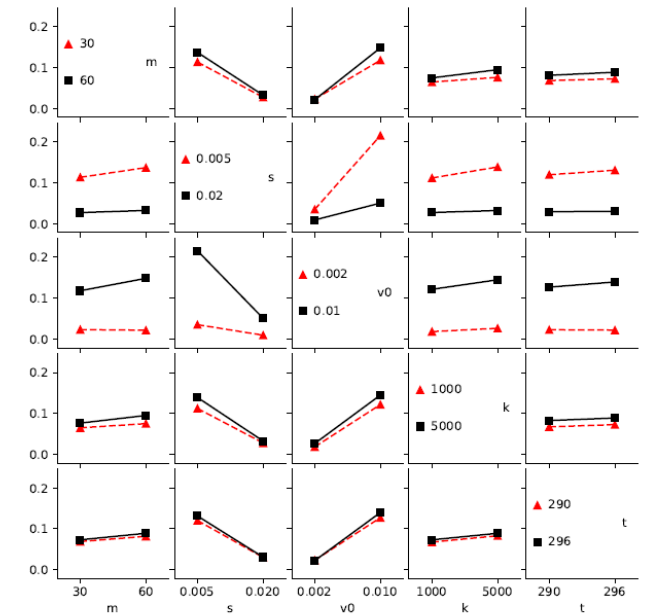
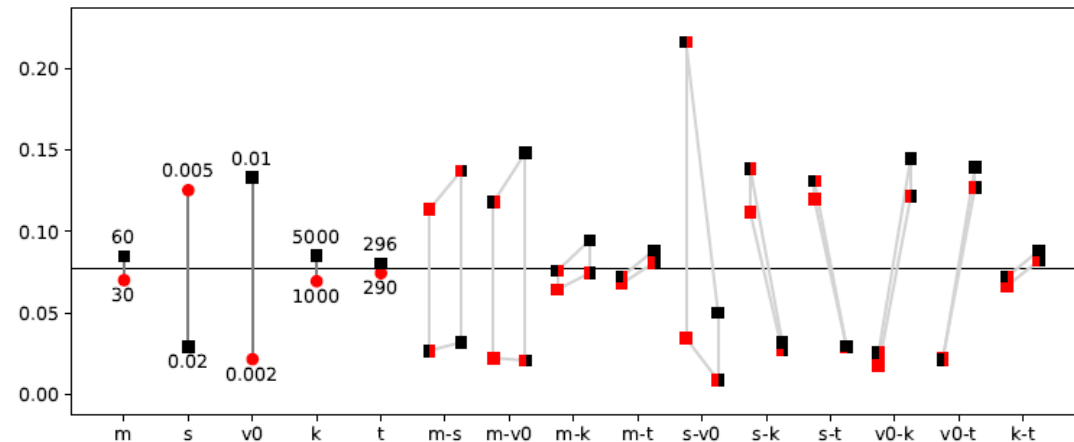
```

dl = {
    'A': [-1, 1],
    'B': [-1, 1],
    'C': [-1, 1],
    'D': [-1, 1],
    'E': [-1, 1],
}
mistat.addTreatments(doe.frac_fact_res(dl, 4), mainEffects=['A', 'B', 'C', 'D', 'E'])

```

Treatments	A	B	C	D	E
0	(1)	-1	-1	-1	-1
1	AE	1	-1	-1	1
2	BE	-1	1	-1	1
3	AB	1	1	-1	-1
4	CE	-1	-1	1	1
5	AC	1	-1	1	-1
6	BC	-1	1	1	-1
7	ABCE	1	1	1	-1
8	D	-1	-1	-1	1
9	ADE	1	-1	-1	1
10	BDE	-1	1	-1	1
11	ABD	1	1	-1	-1
12	CDE	-1	-1	1	1
13	ACD	1	-1	1	-1
14	BCD	-1	1	1	-1
15	ABCDE	1	1	1	1

Fig. 5.6: Two-Way Interaction Plots



This is a half fractional replications of a 2^5 designs as will be explained in Section 5.8. In Table 5.18 we present the design of a 2^5 full factorial experiment derived using Python.

5.7.5 3^m Factorial Designs

We discuss here the estimation and testing of model parameters, when the design is full factorial, of m factors each one at $p = 3$ levels. We assume that the levels are measured on a continuous scale, and are labelled Low, Medium and High. We introduce the indices i_j ($j = 1, \dots, m$), with values 0, 1, 2 for the Low, Medium and High levels, correspondingly, of each factor. Thus, we have 3^m treatment combinations, represented by vectors of indices (i_1, i_2, \dots, i_m) . The index ν of the **standard order** of treatment combination is

$$\nu = \sum_{j=1}^m i_j 3^{j-1}. \quad (5.7.42)$$

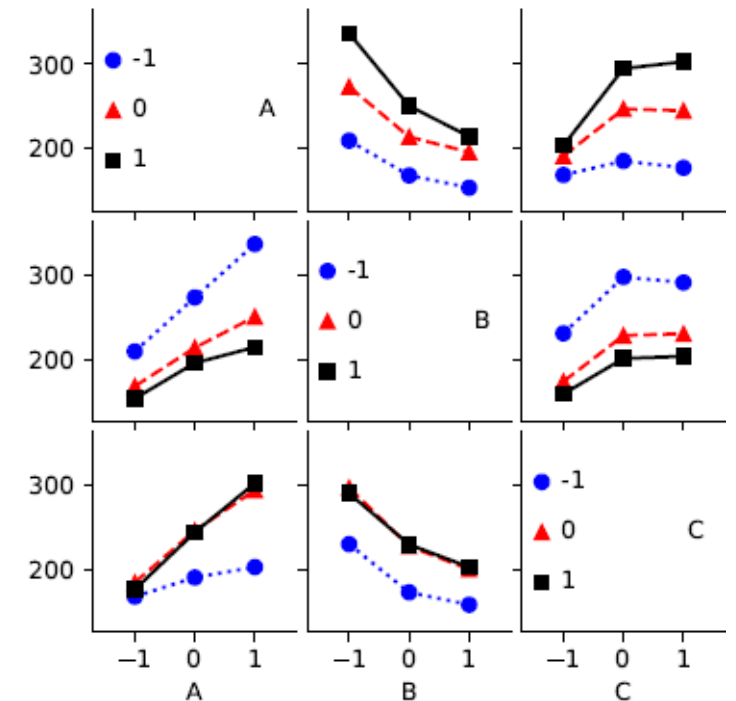
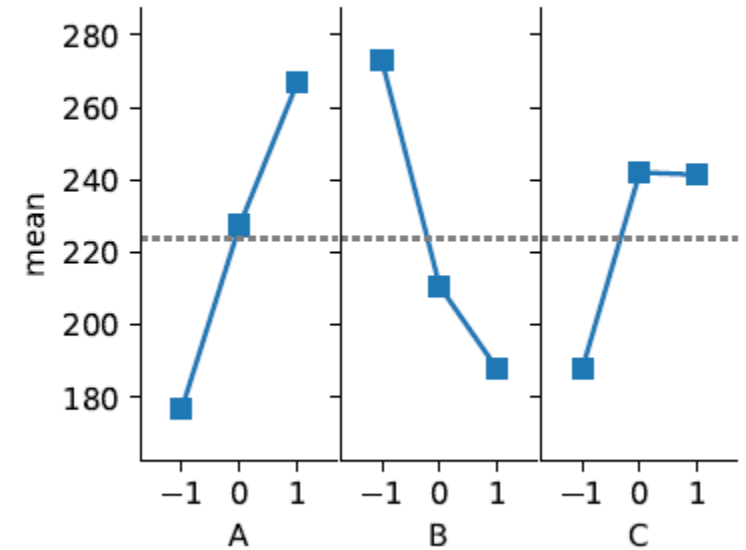
This index ranges from 0 to $3^m - 1$. Let \bar{Y}_ν denote the yield of n replicas of the ν -th treatment combination, $n \geq 1$.

Since we obtain the yield at three levels of each factor we can, in addition to the linear effects estimate also the quadratic effects of each factor. For example, if we have $m = 2$ factors, we can use a multiple regression method to fit the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_1 x_2 + \beta_5 x_1^2 x_2 + \beta_6 x_2^2 + \beta_7 x_1 x_2^2 + \beta_8 x_1^2 x_2^2 + e. \quad (5.7.43)$$

This is a quadratic model in two variables. β_1 and β_3 represent the linear effects of x_1 and x_2 . β_2 and β_6 represent the quadratic effects of x_1 and x_2 . The other coefficients represent interaction effects. β_4 represents the linear \times linear interaction, β_5 represents the quadratic \times linear interaction, etc. We have two main effects for each factor (linear and quadratic) and 4 interaction effects.

Generally, if there are m factors we have, in addition to β_0 , $2m$ parameters for main effects (linear and quadratic) $2^2 \binom{m}{2}$ parameters for interactions between 2 factors, $2^3 \binom{m}{3}$ interactions between 3 factors, etc. Generally, we have 3^m parameters, where



5.9 Exploration of Response Surfaces

The functional relationship between the yield variable Y and the experimental variables (x_1, \dots, x_k) is modeled as

$$Y = f(x_1, \dots, x_k) + e,$$

where e is a random variable with zero mean and a finite variance, σ^2 . The set of points $\{f(x_1, \dots, x_k), x_i \in D_i, i = 1, \dots, k\}$, where (D_1, \dots, D_k) is the experimental domain of the x -variables, is called a **response surface**. Two types of response surfaces were discussed before, the **linear**

$$f(x_1, \dots, x_k) = \beta_0 + \sum_{i=1}^k \beta_i x_i \quad (5.9.1)$$

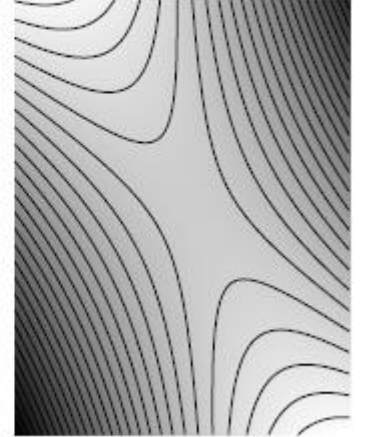
and the **quadratic**

$$f(x_1, \dots, x_k) = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i \neq j} \beta_{ij} x_i x_j. \quad (5.9.2)$$

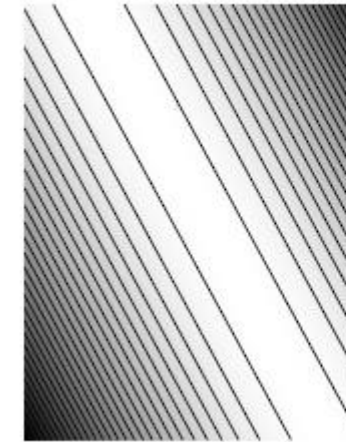
Simple maximum



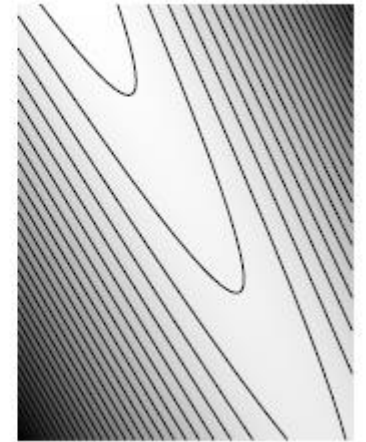
Minimax



Stationary ridge



Rising ridge



5.9.2.2 Central Composite Designs

A Central Composite Design is one in which we start with $n_c = 2^k$ points of a factorial design, in which each factor is at levels -1 and $+1$. To these points we add $n_a = 2k$ axial points which are at a fixed distance α from the origin. These are the points

$$(\pm\alpha, 0, \dots, 0), (0, \pm\alpha, 0, \dots, 0), \dots, (0, 0, \dots, 0, \pm\alpha).$$

Finally, put n_0 points at the origin. These n_0 observations yield an estimate of the variance σ^2 . Thus, the total number of points is $N = 2^k + 2k + n_0$. In such a design,

$$\begin{aligned} b &= 2^k + 2\alpha^2, \\ c &= 2^k, \\ c + d &= 2^k + 2\alpha^4, \quad \text{or} \\ d &= 2\alpha^4. \end{aligned} \tag{5.9.15}$$

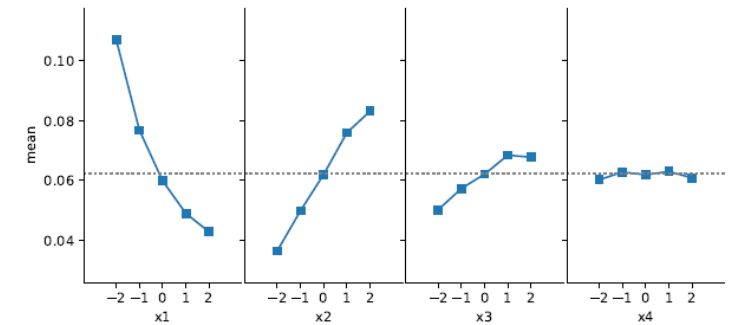
The rotatability condition is $d = 2c$. Thus, the design is rotatable if

$$\alpha^4 = 2^k \quad \text{or} \quad \alpha = 2^{k/4}. \tag{5.9.16}$$

For this reason, in central composite designs, with $k = 2$ factors we use $\alpha = \sqrt{2} = 1.414$. For $k = 3$ factors we use $\alpha = 2^{3/4} = 1.6818$. For rotatability and

Table 5.30: Factors and Level in Piston Simulator Experiment

Factor	Code	Levels				
Piston Surface Area s	x_1	.0075	.01	.0125	.015	.0175
Initial Gas Volume v_0	x_2	.0050	.00625	.0075	.00875	.0100
Spring Coefficient k	x_3	1000	2000	3000	4000	5000
Filling Gas Temperature t_0	x_4	340	345	350	355	360
	code	-2	-1	0	1	2



Optimal designs are constructed using several approaches. Their optimality is determined by different criteria that lead to different solutions. The **D-optimality** criterion minimizes the determinant of the covariance matrix of the model coefficient estimates, the information matrix $X'X$ of the design. D-optimality aims at deriving precise estimates of effects, i.e. main effects, quadratic effects and interactions. This assumes a known pre-specified model and is fully determined by the experimental design before the experiment is conducted. D-optimal designs are used in experiments conducted to test for significance of effects in order to best interpret the fitted model. Their main application is in designs aimed at distinguishing active from inert factors. A design is **A-optimal** if it minimizes the sum of the variances of the regression coefficients, the trace of the inverse of the information matrix. This is another approach

to obtain precise estimates of the effects. **I-optimal** designs minimize the average variance of prediction over the design space. If the primary experimental goal is to predict a response or determine regions in the design space where the response falls within an acceptable range, the I-optimality criterion is more appropriate than the D-optimality criterion. In these cases, precise prediction of the response takes precedence over precise estimation of the parameters. A related approach is **G-optimal** designs, which minimize the maximum prediction variance over the design region. The maximum entry in the diagonal of the hat matrix $X^{-1}(X'X)^{-1}X$. These designs are calculated using Monte Carlo experiments of the design space. The minimum G aberration criterion is useful when selecting good, regular two-level fractional factorial designs and is used to discriminate among regular fractional factorial designs with the same resolution. Minimum aberration compares the frequency of aliases of regular designs at different levels. Regular designs with the smallest frequency of worst aliases are considered the best. The minimum G-aberration criterion can also handle irregular design spaces (Tang and Deng, 1999; Goos, 2011).

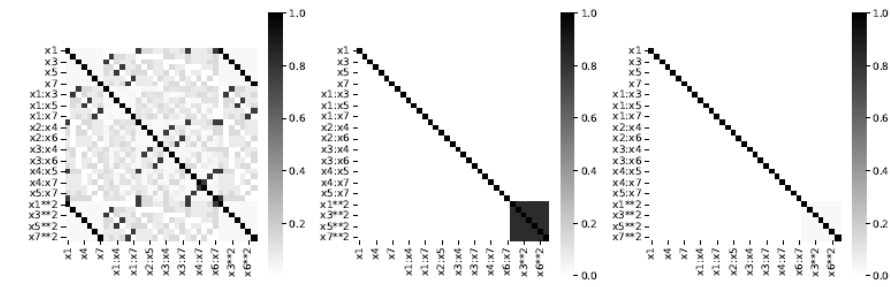


Fig. 5.14: Correlation plots of experimental arrays with 35, 80 and 169 experimental runs

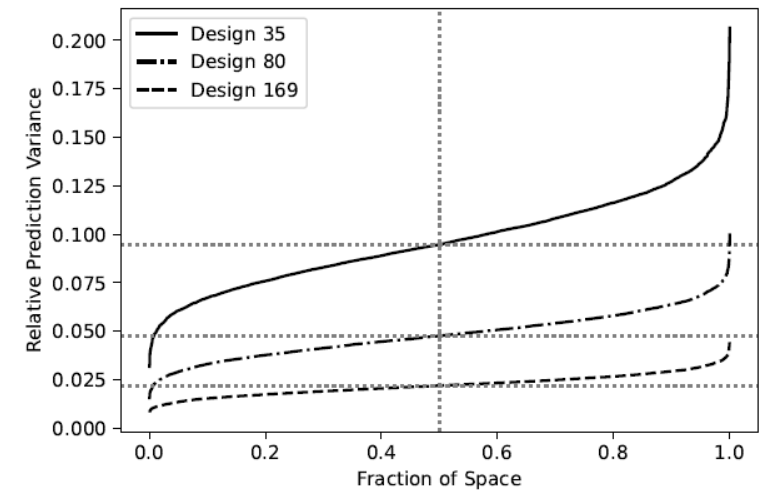
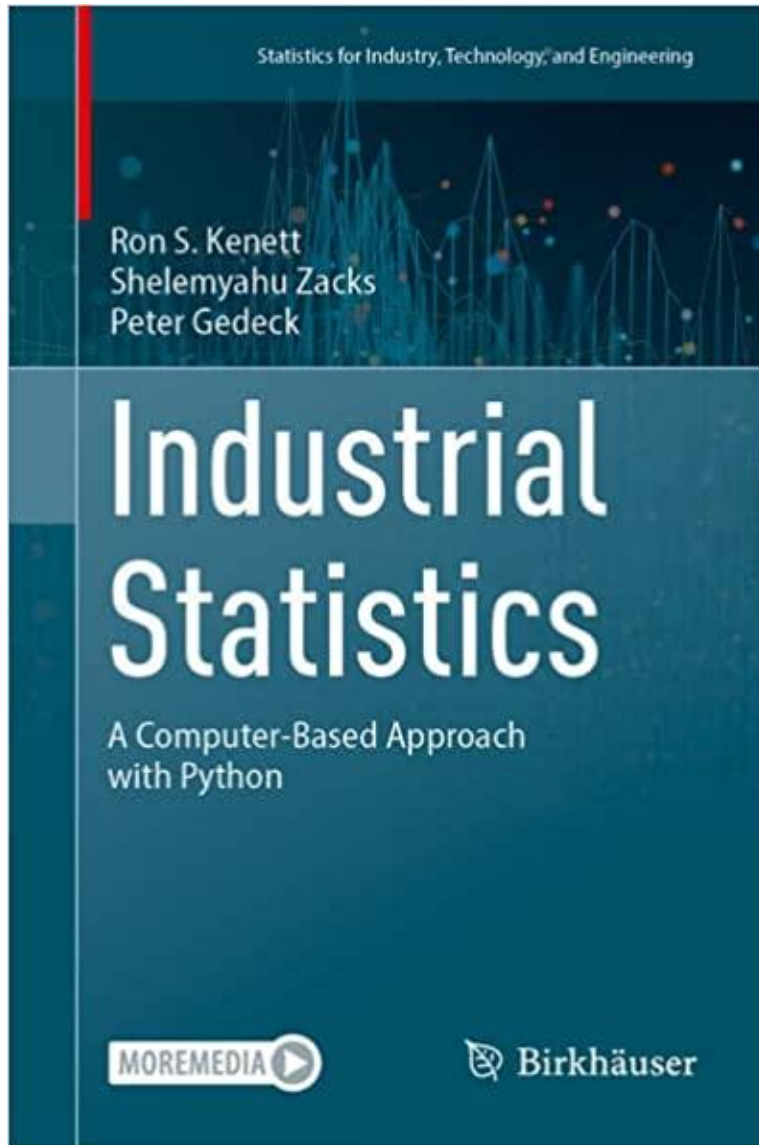


Fig. 5.15: Fraction of design space plot for the three designs



Chapter 6 Quality by Design

Quality by Design (QbD)

Preview Quality is largely determined by decisions made in the early planning phases of products and processes. A particularly powerful technique for making optimal design decisions is the statistically designed experiment introduced in Chapter 5. This chapter covers the basics of experimental designs in the context of engineering and economic optimization problems. Taguchi's loss function, signal to noise ratios, factorial models and orthogonal arrays are discussed using case studies and simple examples. A special section is dedicated to the application of Quality by Design (QbD) in the pharmaceutical industry. QbD is supported internationally by the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) and by the Food and Drug Administration (FDA).

The QbD Column: Applying QbD to make analytic methods robust

JMP Blog

The QbD Column: Overview of Quality by Design

JMP Blog

The QbD Column: A QbD factorial experiment

JMP Blog

The QbD Column: A QbD fractional factorial experiment

JMP Blog

The QbD Column: Achieving robustness with stochastic emulators

JMP Blog

The QbD Column: Response surface methods and sequential exploration

JMP Blog

The QbD Column: Split-plot experiments

JMP Blog

0 ronkenett 2 10-21-2016

A QbD update: Current and future trends in Quality by Design

<https://community.jmp.com/t5/JMP-Blog/A-QbD-update-Current-and-future-trends-in-Quality-by-Design/ba-p/258425>

Created: APR 16, 2020 09:48 AM

Quality by Design (QbD) is about processes, products and clinical understanding. Five years ago, my colleagues and I launched a [QbD column](#), which consisted of nine entries. I encourage you to review the original posts in the QbD column since they are still very relevant today.

Before you revisit the column, take a moment to watch this four-minute [video](#) of the Sanofi factory of the future to see an advanced implementation of biotechnology, chemistry, manufacturing and controls (CMC). To appropriately manufacture a drug product, a specific manufacturing process, product characteristics, and product testing must all be defined to ensure that the product is safe, effective and consistent between batches.

The QbD column series provides a peek under the hood on how the pharmaceutical industry develops and optimizes products and processes, so that they meet the requirements of the factory of the future. Let me give you an overview of each column:

paper poster jsj
statistics graph builder



QbD update: a look back and to the future to see how QbD can help pharma and other industries streamline and optimize processes for state-of-the-art results.

Translation is in progress. Please retry in a few minutes.



arati_mejda STAFF

The QbD Column: Response surface methods and sequential exploration

Created: OCT 28, 2015 06:08 PM

Editor's note: This post is by @ronkenett, @david_s and Benny Yoskovich of the [KPA Group](#). It is part of a series of blog posts called [The QbD Column](#). For more information on the authors, click the authors' community names above.

<https://community.jmp.com/t5/JMP-Blog/The-QbD-Column-Response-surface-methods-and-sequential/ba-p/30654>



Ron Kenett



David Steinberg



Benny Yoskovich



Blog 6 Xu_pb12.jmp



Blog 6 Xu_ccd.jmp

George Box and K.B. Wilson introduced the idea of response surface methodology in a famous article[1] in 1951. There were several novel and extremely useful ideas in the article:

- Designed experiments can be a great tool in experimentally optimizing conditions.
- When feedback is rapid, there are great benefits to breaking up the experimental effort into a sequence of experiments, rather than trying to "learn everything at once".
- The results of one experiment will often stimulate changes in strategy: new factors may be added, old ones may be dropped, factor ranges may move.
- The results may indicate that a more complex regression model is needed to adequately reflect the relationship between the factors and the

The QbD Column: Achieving robustness with stochastic emulators

Created: OCT 5, 2015 12:51 PM

Editor's note: This post is by [@ronkenett](#), Anat Reiner-Benaim and [@david_s](#) of the [KPA Group](#). It is part of a series of blog posts called [The QbD Column](#). For more information on the authors, click the authors' community names above.



Ron Kenett



David Steinberg



Anat Reiner-Benaim



Blog 4 experiment with specs.jmp

In an earlier installment of [The QbD Column](#) titled [A QbD factorial experiment](#), we described a case study where the focus was on modeling the effect of three process factors on one response, viscosity. Here, we expand on that case study to show how to optimize process parameters of a product by considering eight responses and considering both their target values and the variability in achieving these targets. We apply a stochastic emulator approach, first proposed in [Bates et al., 2006](#), to achieve robust on target performance. This provides additional insights to the setup of product and process factors, within a design space.

A case study with eight responses and three process factors

The case study refers to a formulation of a generic product designed to match the properties of an existing brand using in vitro tests. In vitro release is one of several standard methods used to characterize performance of a finished topical dosage form (for details see SUPAC, 1977).

The in vitro release testing apparatus has six cells where the tested generic product is compared to the brand product. A 90% confidence interval for

<https://community.jmp.com/t5/JMP-Blog/The-QbD-Column-Achieving-robustness-with-stochastic-emulators/ba-p/30644>

Optimize Spore Production

An example

Goal: Optimize solid medium for the production of spores of *Coniothyrium minutans*.

Response: Spore production after 31 days (in units of 10^9).

Results: After optimization, production increased by a factor of 7.

Sequential Design of
Experiments

Optimize Spore Production

Experimental Strategy

- 1. Two-level screening experiment to find the key factors.**
2. Extend the design to estimate all terms of second-degree polynomial in the key factors.
3. Use the fitted model to find the optimum conditions.

Optimize Spore Production

Experimental Factors and Levels

Factor	Units	Low	Middle	High
Starch	g/L	18	54	90
Urea	g/L	2	6	10
KH ₂ PO ₄	g/L	1	2.5	4
CaCl +	mg/L	0.5	1	2
MgCl +	g/L	0.05	0.1	0.2
Thiamin	mg/L	1	2	4
Tr. El.	mL/L	5	10	20

Optimize Spore Production

- The design was a full factorial two-level 16 run design.
- Only 8 runs per day were possible. “Day” was added as a “block” factor.
- Middle levels were used to add two “center points” (with all factors at middle levels), one on each day. Center points give a check against pure non-linear effects.
- Some factors have their levels equally spaced on a log scale.

Optimize Spore Production

For the experimental code, use the following labels:

A – KH_2PO_4

E – Starch

B – $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$

F – Urea

C – Trace Elem.

G – $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$

D – Day

H – Thiamin

E=ABC F=ABD G=ACD H=BCD

Optimize Spore Production

	A	B	C	D	E	F	G	H	Y
1	-	-	-	-	-	-	-	-	6.2
2	+	-	-	-	+	+	+	-	3.8
3	-	+	-	-	+	+	-	+	3.8
4	+	+	-	-	-	-	+	+	6.2
5	-	-	+	-	+	-	+	+	8.8
6	+	-	+	-	-	+	-	+	5.6
7	-	+	+	-	-	+	+	-	4.8
8	+	+	+	-	+	-	-	-	8.4
CP	0	0	0	-	0	0	0	0	8.8

Optimize Spore Production

	A	B	C	D	E	F	G	H	Y
9	-	-	-	+	-	+	+	+	5.6
10	+	-	-	+	+	-	-	+	7.6
11	-	+	-	+	+	-	+	-	7.2
12	+	+	-	+	-	+	-	-	6.0
13	-	-	+	+	+	+	-	-	12.2
14	+	-	+	+	-	-	+	-	7.4
15	-	+	+	+	-	-	-	+	7.2
16	+	+	+	+	+	+	+	+	10.8
CP	0	0	0	+	0	0	0	0	9.0

Optimize Spore Production (Reviewed)

Experimental Strategy

1. Two-level screening experiment to find the key factors.
- 2. Extend the design to estimate all terms of second-degree polynomial in the key factors.**
- 3. Use the fitted model to find the optimum conditions.**

Optimize Spore Production (Reviewed)

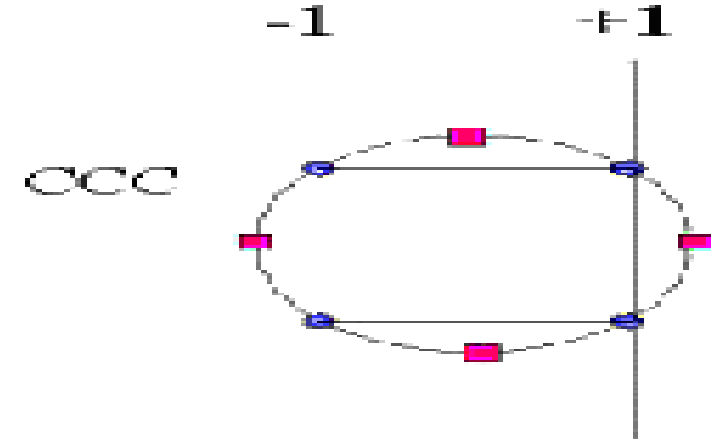
CCD factors and levels

		Coded	Level		
Factor	-1.633	-1	0	1	1.633
Starch	18.35	50	100	150	181.65
Urea	0.53	1.00	1.75	2.50	2.97
Tr. El.	3.7	10	20	30	36.3

Optimize Spore Production (Reviewed)

Three
factors CCD
With 6 center
points

	Starch	Urea	Trace	#Spores
---	1-	-1	-1	5.8
+--	1-	-1	1	7.5
-+-	1-	1	-1	9.1
++-	1-	1	1	13.5
--+	1	-1	-1	5.8
+-+	1	-1	1	4.8
-++	1	1	-1	5.8
+++	1	1	1	11.4
a00	1.633-	0	0	6.1
A00	1.633	0	0	6.1
0a0	0	-1.633	0	3.1
0A0	0	1.633	0	8.1
00a	0	0	-1.633	4.5
00A	0	0	1.633	8.8
000	0	0	0	9.1
000	0	0	0	8.4
000	0	0	0	7.9
000	0	0	0	8.1
000	0	0	0	6.1
000	0	0	0	6.2



Optimize Spore Production (Reviewed)

Response Y

Effect Screening

The parameter estimates below were transformed to be uncorrelated.

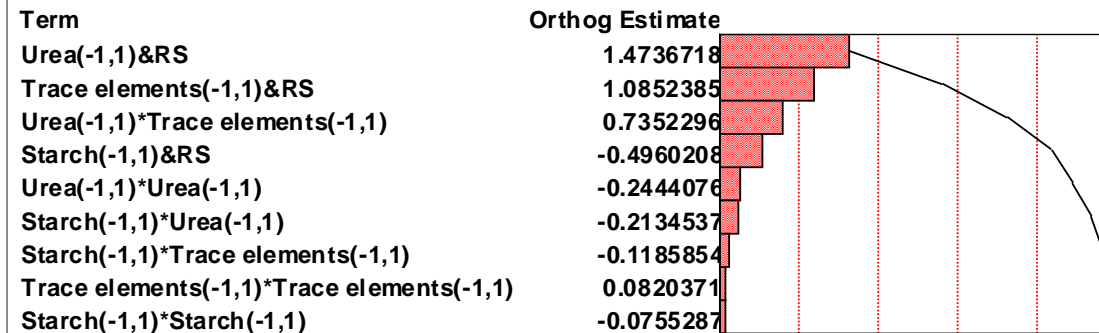
The parameter estimates below were transformed to have equal variances.

Lenth PSE
 t-Test Scale 0.8963144
 Coded Scale 0.3201806

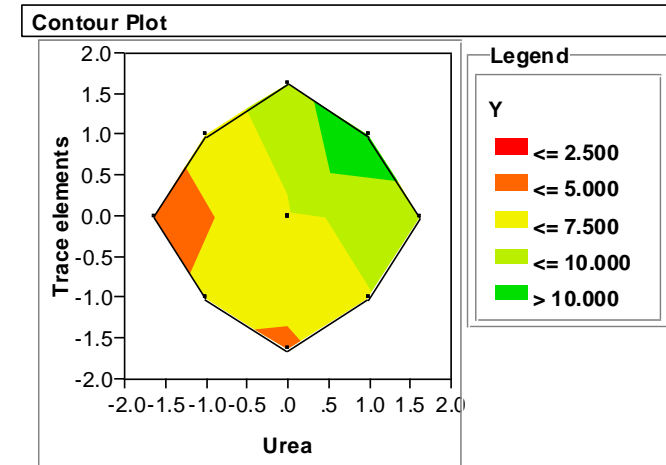
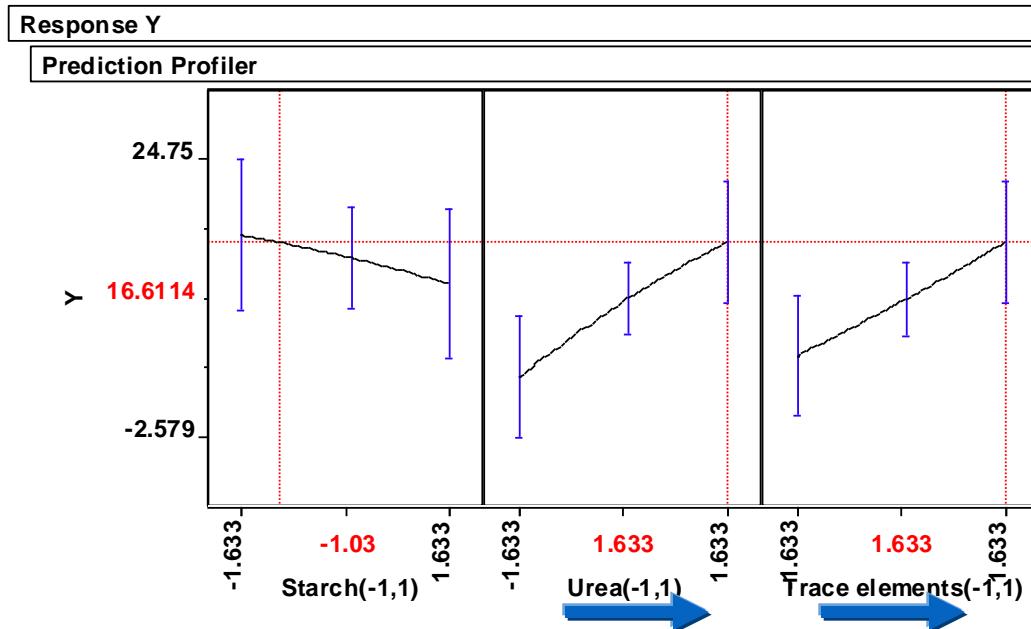
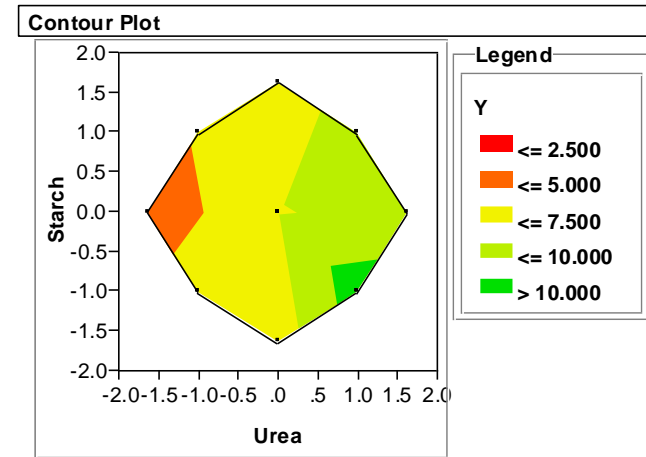
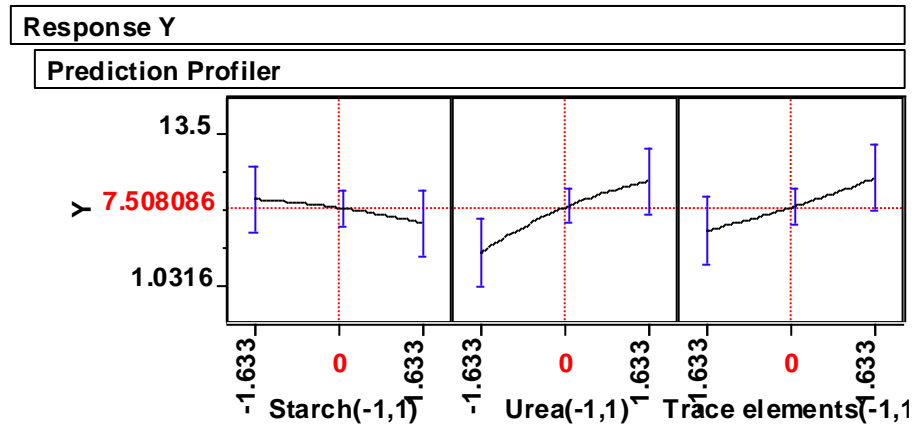
Parameter Estimate Population

Term	Original	Orthog Coded	Orthog t-Test	Prob> t
Intercept	7.508086	7.310000	20.4636	<.0001
Starch(-1,1)&RS	-0.607498	-0.496021	-1.3886	0.1951
Urea(-1,1)&RS	1.804869	1.473672	4.1254	0.0021
Trace elements(-1,1)&RS	1.329138	1.085238	3.0380	0.0125
	-0.337500	-0.213454	-0.5975	0.5634
	-0.187500	-0.118585	-0.3320	0.7468
Urea(-1,1)*Trace elements(-1,1)	1.162500	0.735230	2.0582	0.0666
	-0.105293	-0.075529	-0.2114	0.8368
	-0.292791	-0.244408	-0.6842	0.5094
	0.100956	0.082037	0.2297	0.8230

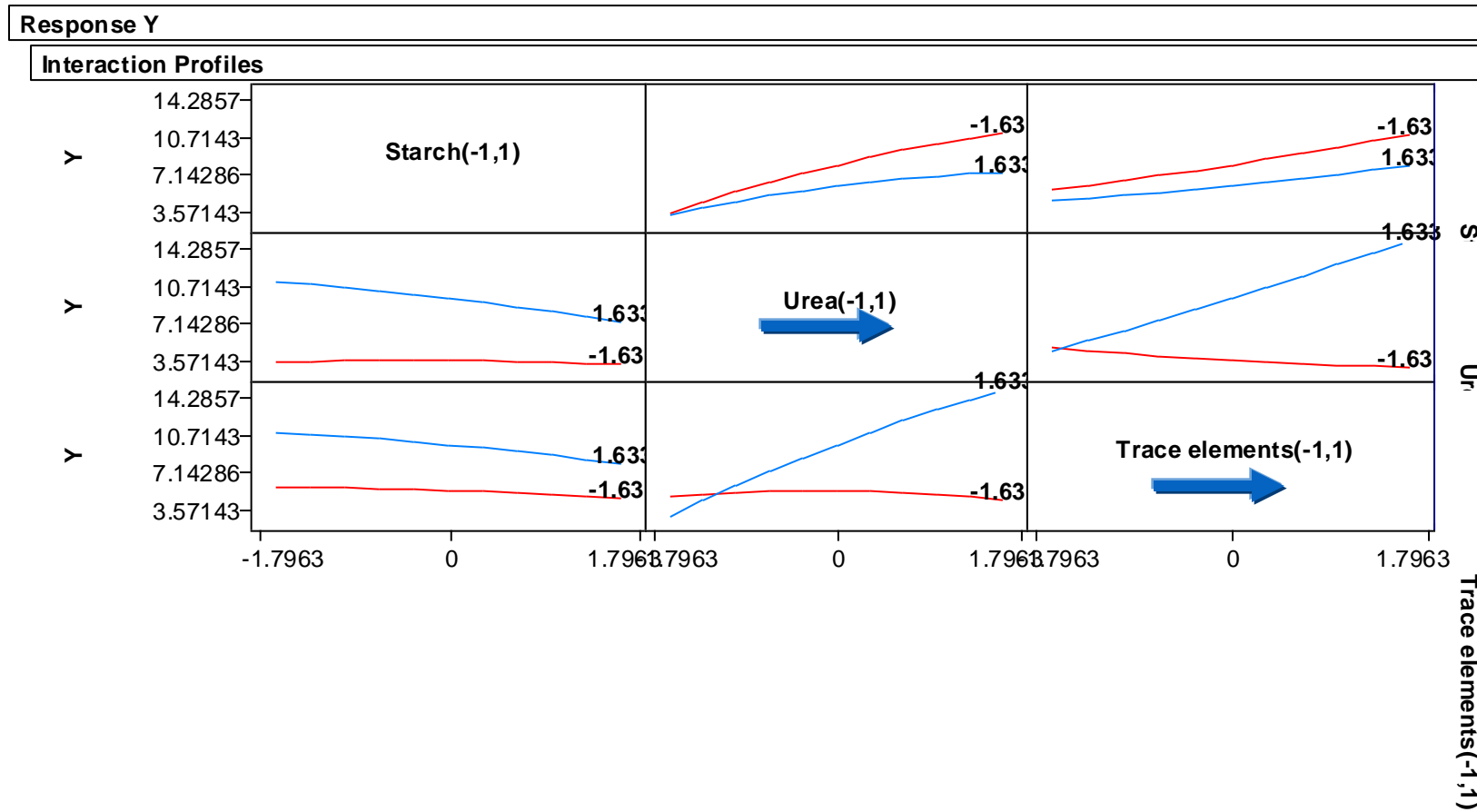
Pareto Plot of Transformed Estimates



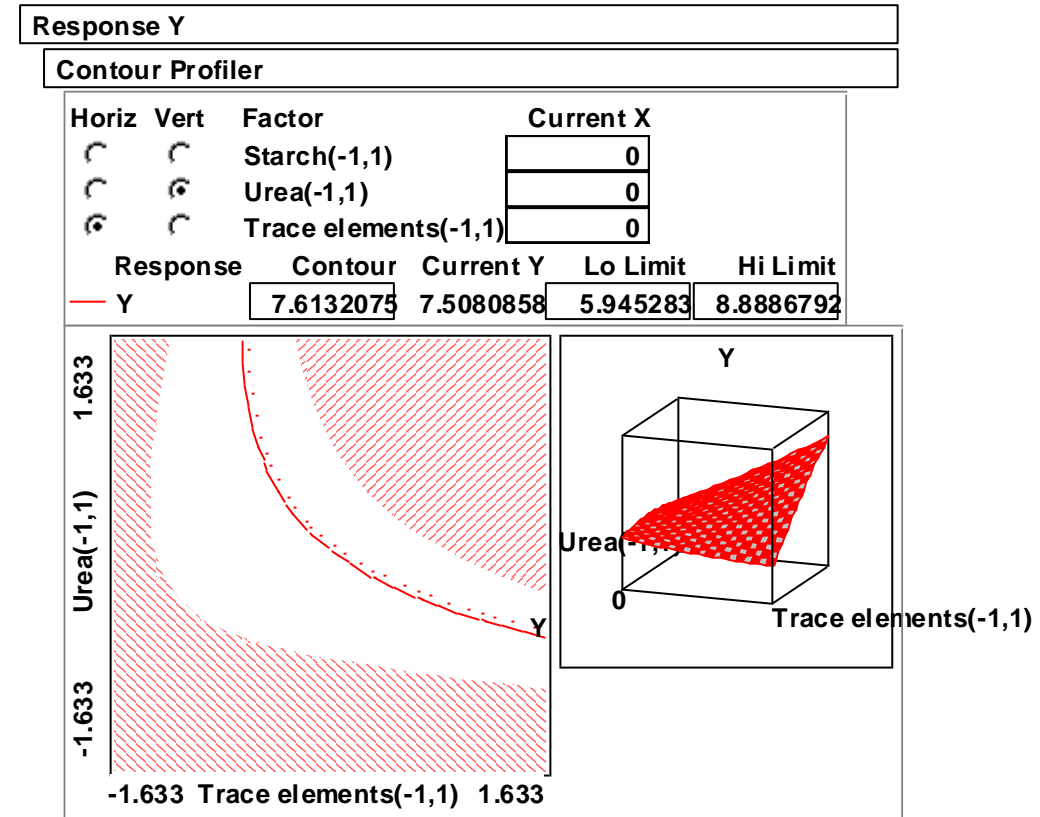
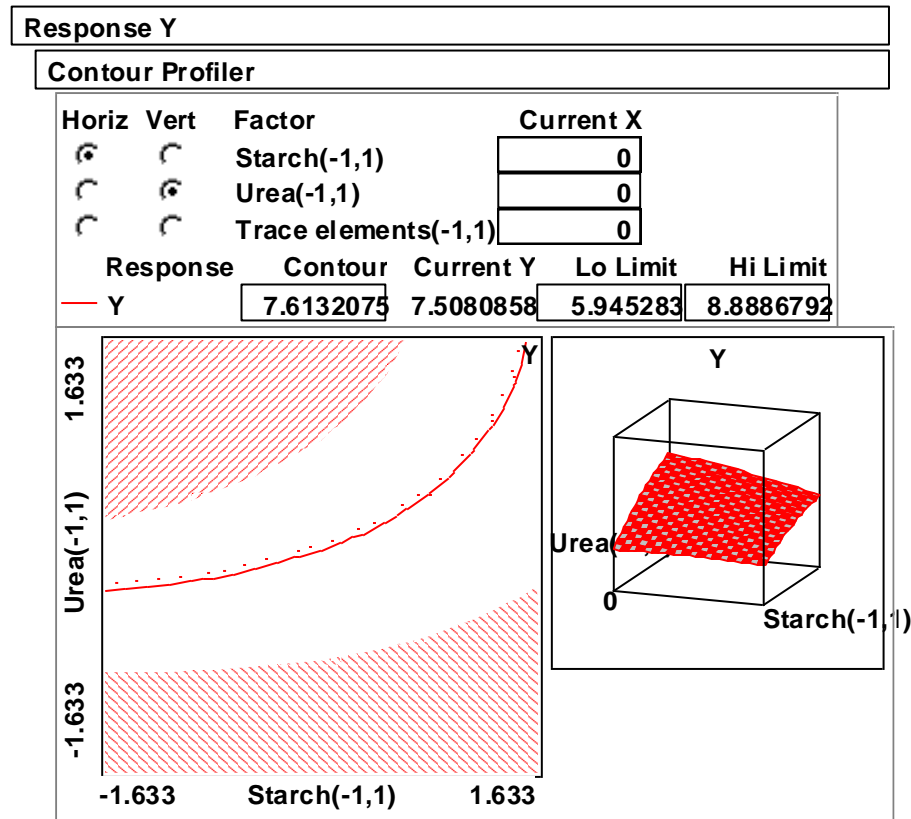
Optimize Spore Production (Reviewed)



Optimize Spore Production (Reviewed)



Optimize Spore Production (Reviewed)

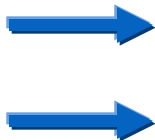


Optimize Spore Production (Reviewed)

		Coded	Level		
Factor	-1.633	-1	0	1	1.633
Starch	18.35	50	100	150	181.65
Urea	0.53	1.00	1.75	2.50	2.97
Tr. El.	3.7	10	20	30	36.3

Second CCD

		Coded	Level		
Factor	-1.633	-1	0	1	1.633
Starch	18.35	50	100	150	181.65
Urea	2.96	3.75	5.00	6.25	7.04
Tr. El.	27.3	40	60	80	92.6



<https://plato.stanford.edu/entries/aristotle-causality/>

In the *Posterior Analytics*, Aristotle places the following crucial condition on proper knowledge: we think we have knowledge of a thing only when we have grasped its cause (*APost.* I 2, 71 b 9–11. Cf. *APost.* II 11, 94 a 20). That proper knowledge is knowledge of the cause is repeated in the *Physics*: we think we do not have knowledge of a thing until we have grasped its why, that is to say, its cause (*Phys.* II 3, 194 b 17–20). Since Aristotle obviously conceives of a causal investigation as the search for an answer to a why-question, and a why-question is a request for an explanation, it can be useful to think of a cause as a certain type of explanation.

<https://www.youtube.com/watch?v=ADs7fWlVuVk&t=310s>

Kenett QSR ENBIS 2021

Quality Statistics Reliability

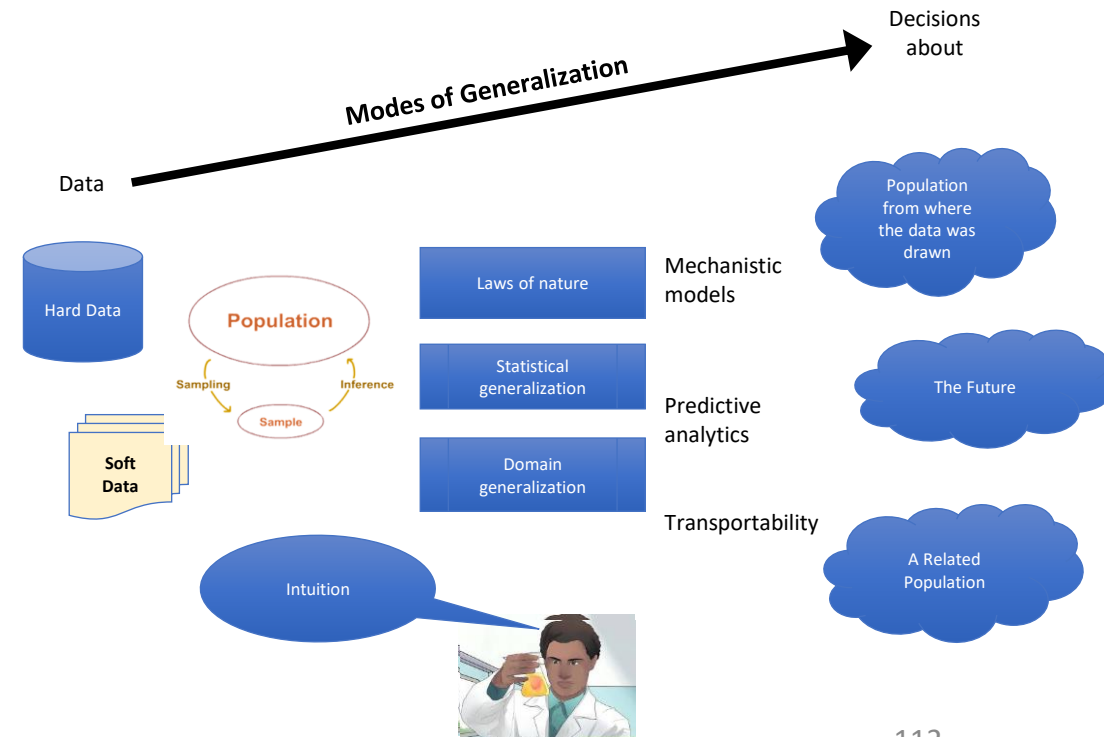
informis ANNUAL MEETING 2022 ANAHEIM, CALIFORNIA

enbis European Network of Industrial Statisticians

“You can generalize it if you can interpret it, and vice versa”

Prof Ron S. Kenett
ron@kpa-group.com

Play (k) 0:15 / 14:48 • Introduction >



https://www.linkedin.com/pulse/reproducibility-applied-research-big-picture-ron-s-kenett/?fbclid=IwAR1Gal5ShBdoh2gOAV_hHdLIynX6Who01VcdhHqZBj8SiyNQaGRb_CCAHM

On statistical significance
and reproducibility

https://shepherd.com/best-books/how-numbers-turn-into-information?fbclid=IwAR00jliS41NcldvGxYaY_5AETXckZtscz1Lr8qE8OR5XQQ4FHmyKwhi8TFc

https://ceeds.unimi.it/wp-content/uploads/2020/02/Kenett_Causality_2020.pdf

<https://www.youtube.com/watch?v=nWaM6XmQEmU>



**AR ANNUAL
REVIEWS**

Annual Review of Statistics and Its Application
Statistical Significance

D.R. Cox

Nuffield College, Oxford University, Oxford OX1 1NF, United Kingdom;
email: david.cox@nuffield.ox.ac.uk

Sir David Cox - In gentle praise of Significance Tests - Significance tests (part 4)

https://www.youtube.com/watch?v=txLj_P9UICQ

Recap

1. Introduction
2. Data types and data integration
3. Supervised learning
4. Model performance
5. Time series
6. Data visualization
7. Causality and experimental design

<https://www.youtube.com/watch?v=gHoeuuwPc>

