“A fib, a lie and statistics”

Ernst Wit
University of Groningen
e.c.wit@rug.nl
http://www.math.rug.nl/∼ernst

23 January 2014

rijksuniversiteit groningen

A fib, a lie and statistics – outline

I. Light in ancient Greece?

II. Period of Muslim Error

III. Gambling's moral high ground

IV. Florence Nightingale's treason

V. How the thermonuclear bomb saved statistics

VI. Death of Truth & Truth of Death in Iraq.

PART I: Greek $\sigma\tau o\chi o\varsigma$?

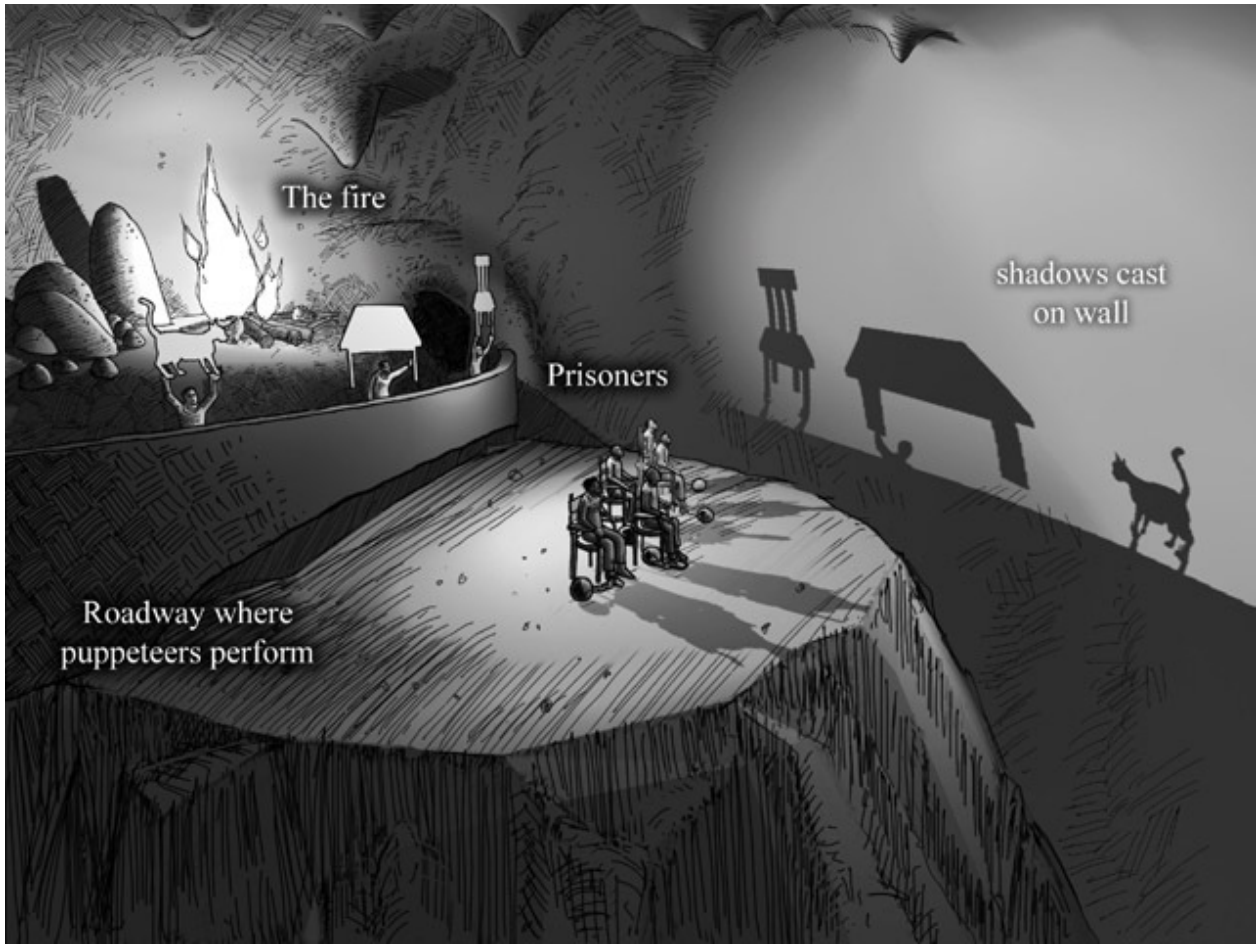
How Plato invented the Pure Form of Statistics

Early Monte Carlo Simulations: Astragali

- Sheep ankle bone
- used in groups of 5
- Used in religious ceremonies:
  - divination rites (Asia Minor)
  - Greece: Zeus = (1,3,3,4,4), Cronos = (4,4,4,6,6)
- Used in games:
  - Greeks, Romans, Early Christian's (knucklebones: mentioned by Homer, Sophocles and Plato)
  - Medieval gamblers liked the game Hazard.

rijksuniversiteit groningen

Early Statistics: Plato's Cave Allegory



shadows = data, **flickering fire** = random noise,
cat/table/chair = true parameters

rijksuniversiteit groningen

PART II: The Muslim Error

How Al-Gebra terrorized the West

Islamic prohibition of Gambling

Qur'an is quite specific:

> *It is forbidden for two or more participants to put up prizes that all go to the winner.*

With nothing better to do, **Al-Hajjaj ibn Yusuf ibn Matar** in translated Euclid's *Elements* into Arabic in 808.

Soon, Arabic mathematicians were inventing...

- algebra (by Al-Khwarizmi)
- arithmetic
- trigonometry
- non-Euclidean geometry
- integral and differential calculus
- astronomy
- ... cryptography.

Cryptography. A flickering of hope...?

International diplomacy and warfare used cryptography extensively.

Abu Yusuf Yaqub ibn Ishaq al-Kindi (801-873)

- wrote a book "On Deciphering Cryptographic Messages"
- pioneered a frequency analysis of cryptic texts, using a statistical analysis of letters and letter combinations in Arabic.
- Girolamo Cardano considered him one of the twelve greatest minds of the Middle Ages.

PART III: Girolamo Cardano

Gambling's moral high ground

rijksuniversiteit groningen

Girolamo Cardano: Love of Gambling

- born 1501 in Pavia
- quarrelsome man
- doctor saving the life of Hamilton, Archbishop of St Andrews.
- dying wife, troublesome children
- devoted gambler
- Wrote in 1525 "Liber de Ludo Aleae:"
  - manual for playing dice and cards;
  - how to cheat;
  - identifying a sample space with equally likely outcomes

Gerolamo Cardano, 1501–76

rijksuniversiteit groningen

Problem of Points: Tartaglia, Fermat, Pascal

- Game of chance between two people.
- Each person has equal chance to win each round.
- Person who first wins n rounds get stake.
- Game is interrupted before the end.

Question:
how to divide the stakes fairly?

rijksuniversiteit groningen

Christiaan Huygens: De Ratiociniis in Ludo Aleae

Christiaan Huygens (1629-1695) stood at the beginning of the development of probability

- His 1656/7 tract deals with modelling games of chance.
- Jacob Bernoulli, *Ars Conjectandi*, 1713
- De Moivre, *Doctrine of Chances*, 1718
- Laplace, *Memoir*, 1774.

rijksuniversiteit groningen

Difficulty of combining observations

- **2003**: One of my students in an exam:

  "The standard error goes up, because with more observations there is more confusion and therefore less accuracy."

- **1749**: Euler fitting an equation with 8 unknowns to 75 obs:

  "... from these equations we can conclude nothing; and the reason, perhaps, is that I have tried to satisfy several observations exactly, whereas I should have only satisfied them approximately; and this error has then multiplied itself."

- **1750**: Tobias Mayer fits equation with 3 variables and 27 obs.
- **1769/1778**: Daniel Bernoulli struggles with the difference between "value" and "probability" of an observation, while proposing robust and ML methods, respectively.

rijksuniversiteit groningen

PART IV: Interlude

Florence Nightingale's treason

Statistics: a science without a name?

- **Antiquity**: 'statisticum' meaning 'regarding the state'.
- **1749**: 'Statistik' introduced by Gottfriech Achenwall as qualitative state science.
- **1791**: 'statistics' appeared in Sir John Sinclair of Ulbster's *Statistical Account of Scotland*.
- **1833**: founding of a 'Statistics Section' of the BA in the UK.
- **1834**: founding of 'Statistical Societies' in the UK.
- **1850s**: Crimean War.

Florence Nightingale: Saint .....?

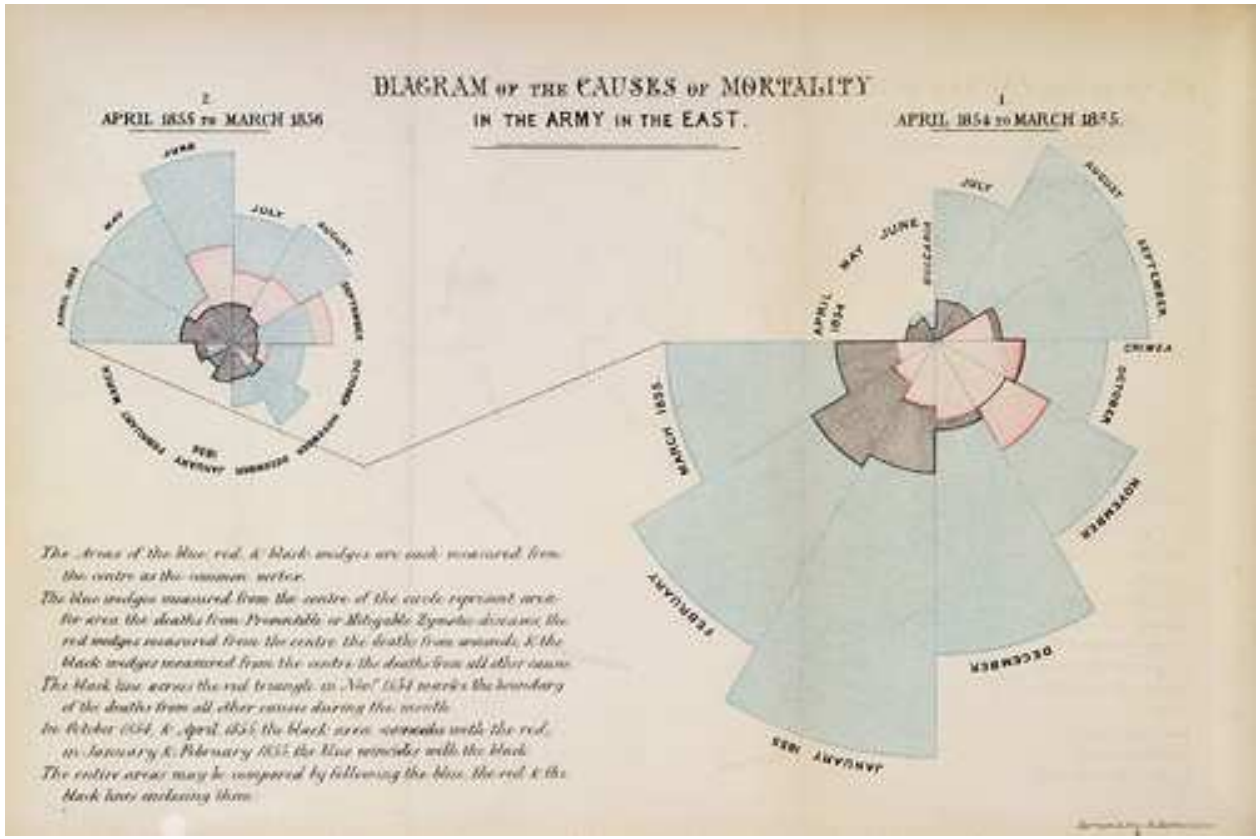
Probably how many of you think of Florence Nightingale...

Florence Nightingale: ..... or villain!

... how you probably should be thinking of her.

Statistics: science of certainty?

- 1850s: Florence Nightingale (1820-1910) introduces



Coxcomb diagram: rotation invariant representation of data.

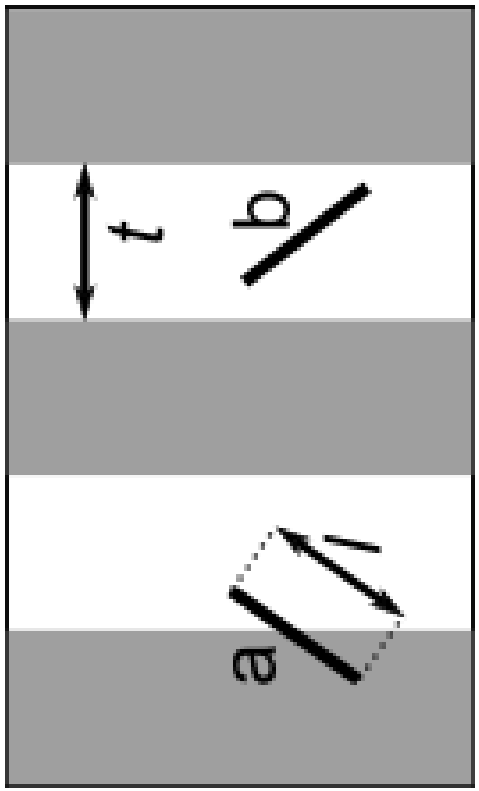
- 1860s: Nightingale becomes a lobbyist in parliament.
- 1870s: Appearance of the phrase:

  A fib, a lie and statistics.

PART VI: The Bomb

How gambling saved statistics — again

Buffon's needle problem (1735)



Buffon suggested a method to determine π:

- Probability that a randomly dropped needle crosses a line (if $l \leq t$): $\frac{2l}{t\pi}$.
- Experiment with n trials & x crossings:

$$\hat{\pi} = \frac{2ln}{tx}.$$

- Mario Lazzarini (1902): $l = 5$, $t = 6$, $n = 3,408$, $x = 1,808$,

$$\hat{\pi} = \frac{34080}{10848} = \frac{355}{113} = 3.1415929$$

rijksuniversiteit groningen

Mario Lazzarini: the second coming of Girolamo Cardano

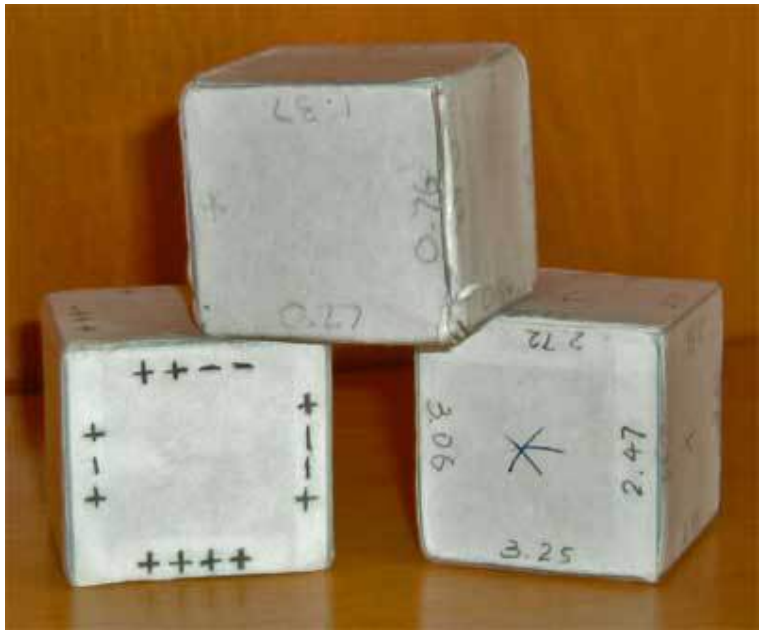
- In fact, $355/113$ is a very good rational approximation of π, as shown by an article of ... Lazzarini (1901).
- Since $\hat{\pi} = 5n/3x$, we get that

$$\begin{aligned} \frac{x}{n} &= \frac{5}{3}\frac{113}{355} \\ x &= \frac{113n}{213} \end{aligned}$$

- If n is a multiple of 213, the "right" x would be an integer.
- So, we could perform an experiment:
  perform multiples of 213 needle throws and stop as soon as $x = 113n/213$.
- Probability that we need at most 16 multiples of 213 is 0.28...
- ... Or you could just cheat! Cardano's back!

rijksuniversiteit groningen

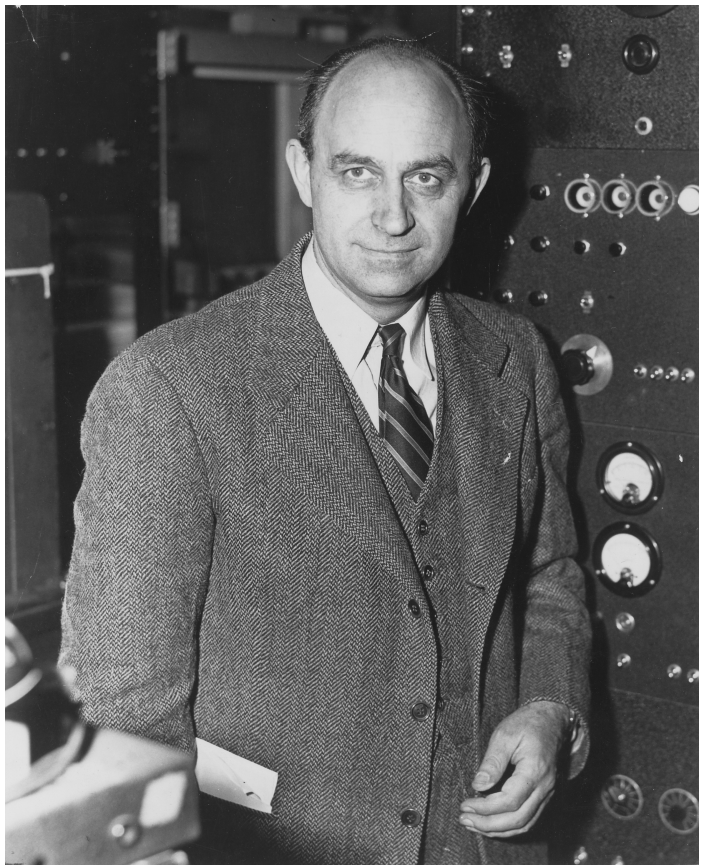
Galton's dice



Galton continued Cardano's tradition...

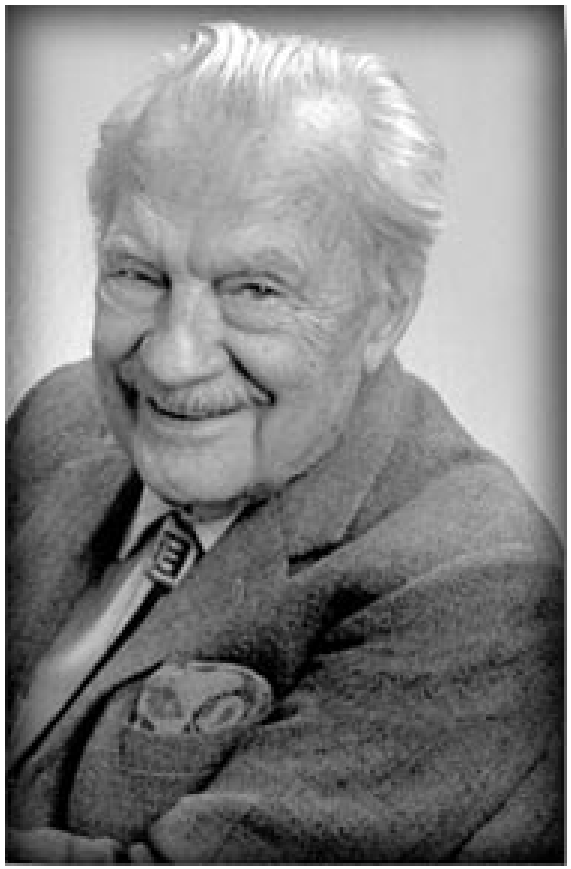
- 3 dice with one roll generate four samples of a standard normal distribution:
  - **Top die:** order statistics of a half standard normal.
  - **Bottom right:** order statistics of half standard normal, conditional on top die exceeding 2.27.
  - **Bottom left:** all possible sign configurations of 4 samples.

Fermi: Monte Carlo FERMIAC and the Chicago Pile-1



- 1932: Discovery of neutron.
- 1933: Juliot and Curie discovered radioactivity induced by α particles.
- Fermi combined the two, but got strange results.
- Using hand-calculated Monte Carlo simulations of neutron slowing down process (Gaussian, exponential or other) for physical insight.
- 1938: Nobel Prize
- 1942: self-sustained nuclear chain reaction

The Bomb: the Eniac, Fermiac and Maniac



- 1943: Establishment of Los Alamos
- Fall 1943: IBM punched-card machines perform Monte Carlo simulations of 12 hard spheres.
- Los Alamos repairman: Nick Metropolis
- 1952: Metropolis' MANIAC operational.
- 1953: Metropolis Algorithm (discovered by M. Rosenbluth)
  "Instead of choosing the (liquid) configurations randomly then weighing them with $\exp(-E/kT)$*, we choose configurations with a probability* $\exp(-E/kT)$ *and weight them evenly."*

rijksuniversiteit groningen

PART VI

Death of Truth & Truth of Death in Iraq

The First Gulf War, 1991

After Iraq invade Kuwait in 1990, strict economic **sanctions** were imposed. They lasted until 2003.

Malnutrition, lack of clean water and medical supplies were **claimed** to have caused excess child deaths (< 5 yrs). Estimates:

- 500,000 until 1999 (UNICEF).
- 239,000 until 1998 (Denis Halliday, UN)
- 1.5 million (Iraqi Baathist government)
- 170,000 (Project on Defense Alternatives)
- 350,000 "even using conservative estimates" (Slate Explainer)
- 345K-530K for 1990-2002 period (Prof Garfield, Columbia)
- 1.5 million (Former U.S. Attorney General Ramsey Clark)
- **"very likely [less than] 0.5 million." (Spagat, economist)**

We present a recent analysis by Spagat (2010).

rijksuniversiteit groningen

Where do all these claims come from?

In 1995 Zaidi & Fawzi (UN-FAO) did a child mortality survey:

- They interviewed 64 households in Bagdad
- using interviewers from Iraqi ministry
- asking:
  - births in period 1985-1990 and 1990-1995
  - child deaths (< 5 yrs) in those 2 periods.

Zaidi and Fawzi estimated:

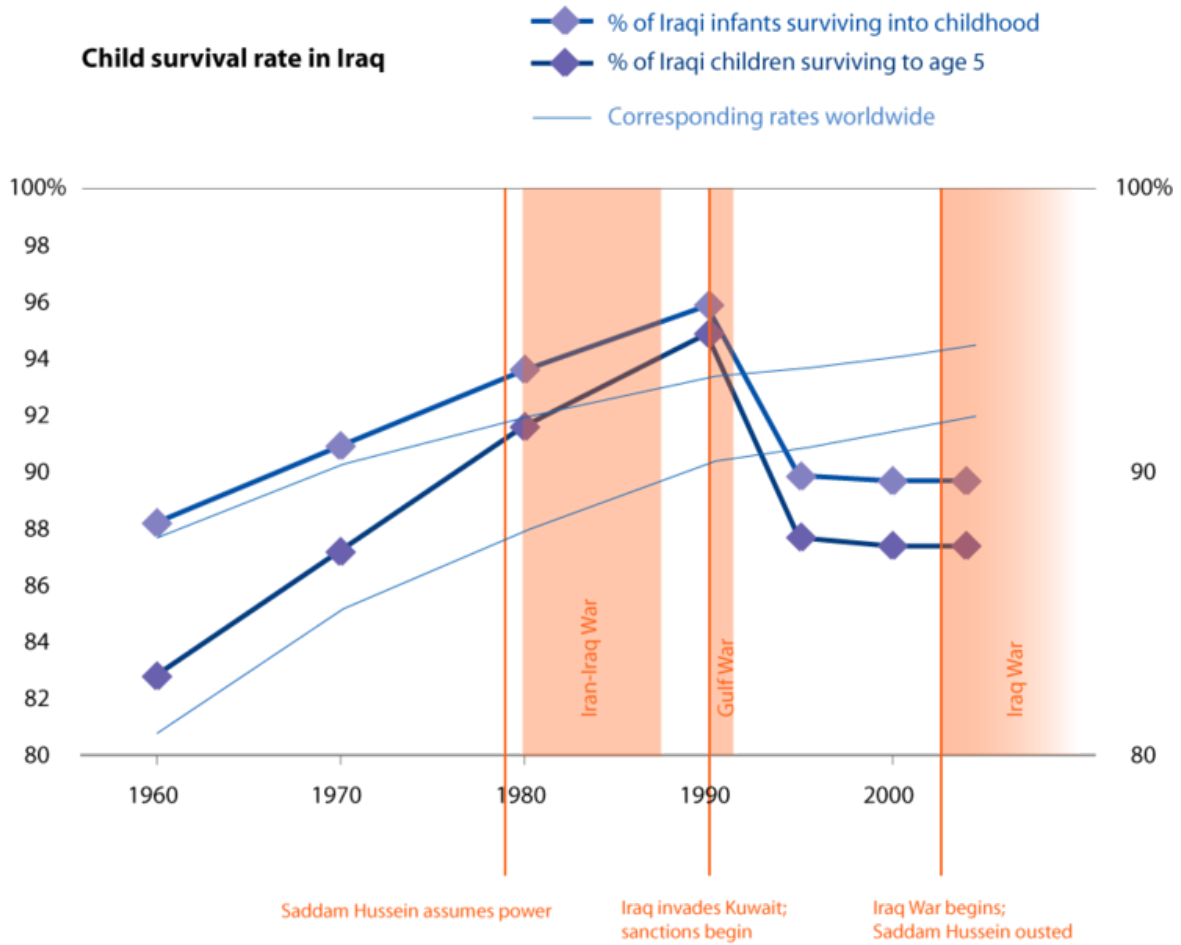
- 1985-1990 infant mortality: 4,06%.
- 1990-1995 infant mortality: 19.82%.

Between 1990-1995 in all of Iraq 3.6 mln children were born.

Excess child deaths: $(0.1982 - 0.0406) \times 3.6\text{mln} = 567\text{K}$

rijksuniversiteit groningen

Picture from the Zaidi and Fawzi study



"Iraq's infant and child survival rates fell after sanctions were imposed."

rijksuniversiteit groningen

Doubts about Zaidi and Fawzi study

In 1996 Zaidi and Fawzi reinterviewed 20 of the 64 households with Jordanian interviewers:

- Of the totally 74 deaths only 9 could be confirmed.
- An additional 9 deaths were reported.

1996 estimate of the 1990-1995 infant mortality rate: 3.8%.

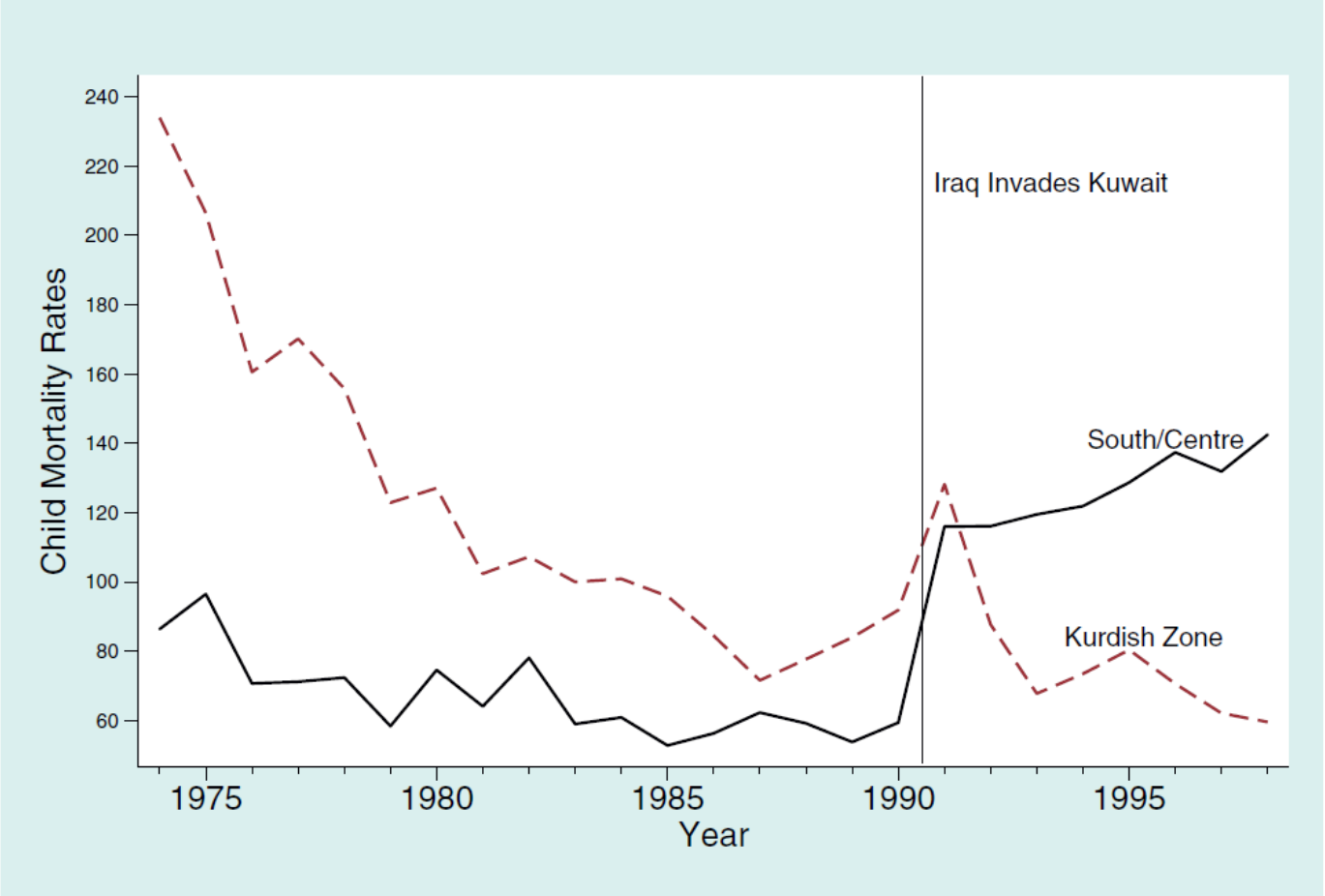
In 1997 Zaidi and Fawzi reinterviewed 26 "switchers" from 1996:

- 17 stuck with their 1996 version.
- 9 switched back to their 1995 version.

1997 estimate of the 1990-1995 infant mortality rate: 5.7%.

In 1997 Zaidi & Fawzi retracted their 567,000 excess child deaths.

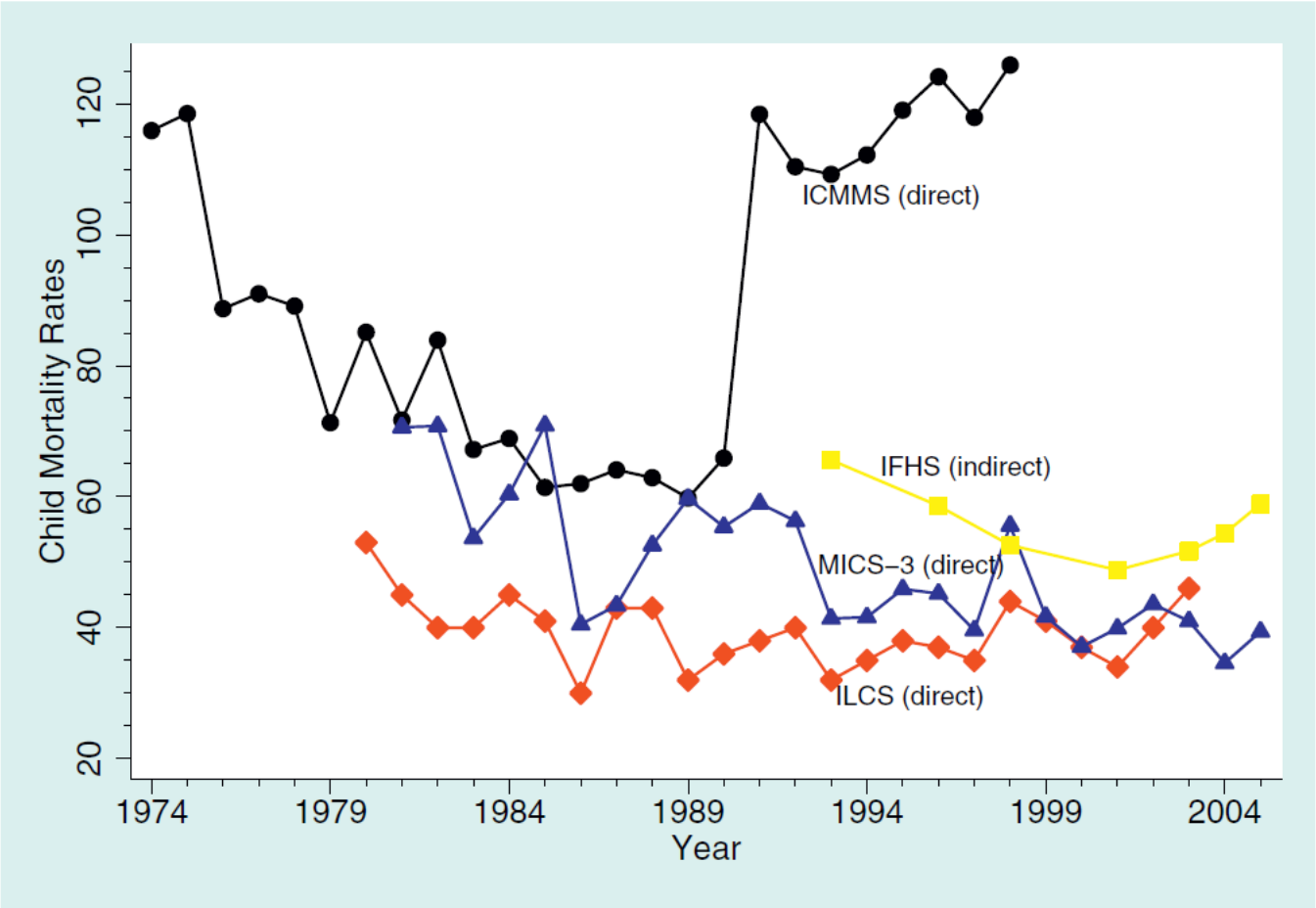
Iraq Child and Maternal Mortality Survey: 500,000 again!



Comparing Saddam's South/Centre with independent Kurdish zone, both under economic sanctions.

rijksuniversiteit groningen

Three further studies...



Comparison of the ICMMS and three later surveys. ICMMS is clearly in spectacular disagreement.

rijksuniversiteit groningen

So what about Statistics?

- Formal mathematics rarely inspired developments in probability and statistics;
- Gambling did!
- Gambling associations (through Huygens/De Moivre/LaPlace) allowed Bernoulli to define the first notion of likelihood.
- Formal mathematical methods facilitated statistical development in the early 19th century
- Modern Monte Carlo techniques connected gambling and sampling.
- **Good sampling can make the difference between**
  - **a fib and a lie ON THE ONE HAND**
  - **... and statistics ON THE OTHER.**