The Role of Statistical Science in Modern Society

Il ruolo delle scienze statistiche nella societa` moderna

Elvezio Ronchetti

Department of Econometrics
University of Geneva
Switzerland

http://www.unige.ch/ses/metri/ronchetti/

Outline

- Introduction
- What is statistics?
- Ex. 1 Has # of accidents decreased in TI?
- Ex. 2 The Vietnam lottery draft
- Ex. 3 The Challenger disaster
- Conclusions

---

Introduction

- From latin « status » (state)
- ---> statistica, statistics
- Army, taxes, census reported in the Bible
- ---> enumeration, data

Introduction

- 17th – 18th century: calculus of probability; Pascal-Fermat(1654), J. Bernoulli(1713), deMoivre(1718), Laplace, Euler, Gauss, Lagrange, Legendre
- 19th century: statistics; Quetelet
- 20th century: axiomatic basis of probability; Hilbert's(1900) sixth problem: Kolmogorov(1933)
- 1920 ---> statistical models, statistical inference Pearson, Fisher, Neyman, Tukey
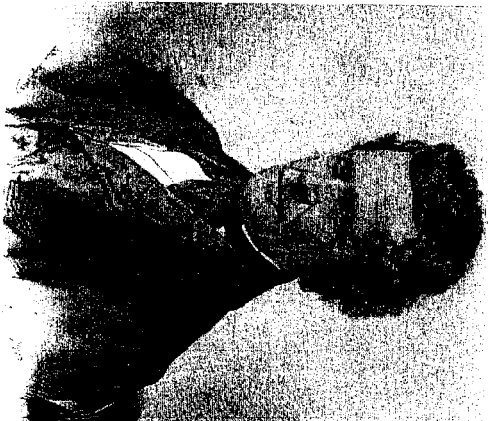
Stefano Franscini (1796 – 1857) (Wikipedia)

Introduction

- « La Statistica della Svizzera e` il manifesto politico di Stefano Franscini. Ma e` anche un trattato di scienza dell'amministrazione, una guida al buon governo, ... »

- « ... (Melchiorre) Gioia rivendicava alla statistica lo statuto di scienza interpretatrice e non meramente descrittiva, perche` collegava i dati, si fondava sul confronto e stabiliva nessi causali. »

- « ... Franscini si formo` a questa scuola e ne condivise i riferimenti ideologici. »

- Da: S. Franscini, Statistica della Svizzera, a cura di Raffaello Ceschi, Armando Dado` editore, Locarno (1991), p. XXIV, XXVII
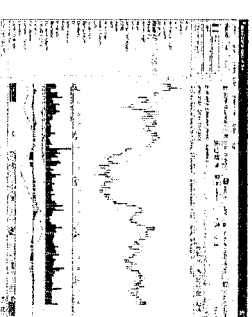

What is statistics?

- Statistics is the science of the research process

➢ Market research (surveys, polls, ...)
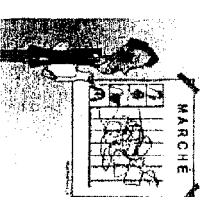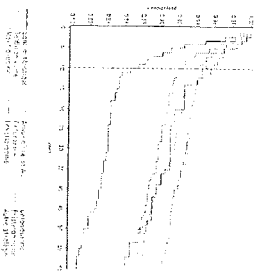

What is statistics?

- Statistics is the science of the research process

➢ Portfolio management
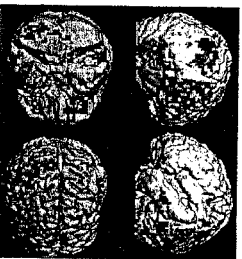
What is statistics?

- Statistics is the science of the research process
  - Clinical trials

What is statistics?

- Statistics is the science of the research process
  - Brain research and neurosciences

What is statistics?

- Statistics is the science of the research process
  - Environmental studies

The statisticians of the « Intergovernmental Panel on Climate Change » Received in 2007 the Peace Nobel Prize!
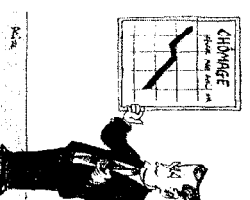
What is statistics?

- Statistics is the science of the research process
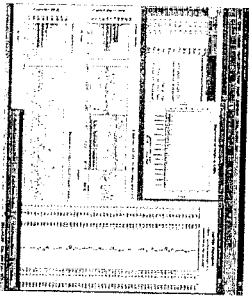 - Economic studies

The 2000 and 2003 Nobel Prizes in Economics were awarded for work in statistics.
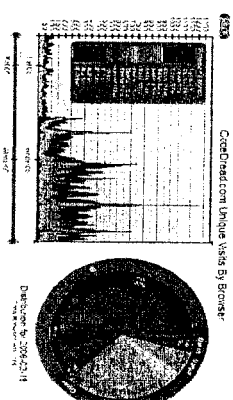
What is statistics?

- Statistics is the science of the research process
  - ⊳ Quality control in industrial production
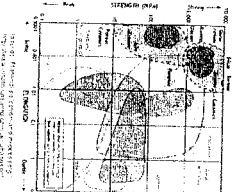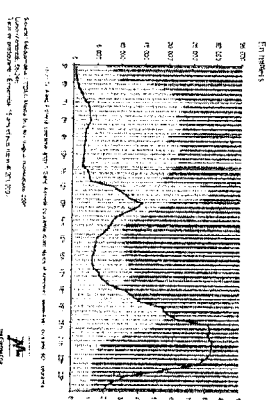
What is statistics?

- Statistics is the science of the research process
  - ⊳ Development of new technologies

What is statistics?

- Statistics is the science of the research process
  - ⊳ Web management

CoreDirect.com Unique Visits By Browser

What is statistics?

- Statistics is the science of the research process
  - ⊳ Measures of audience shares (media)

What is statistics?

- Statistics is the science of the research process

  ➢ Etc.

What is statistics?

- Statistics is the science of the research process
  ➢ As mathematics is the language of science,
  ➢ statistics is the logic of science.

"In God we trust. All others must bring data."

W. Edwards Deming

What is statistics?

- Statistics allows to determine which informations are useful and which predictions are reasonables.
- It helps to find indicators for solving problems.
- The knowledge obtained by means of statistical methods helps to do planning, to determine the resources allocation and to improve quality.

What is statistics?

- It provides a more reliable evaluation of risk.
- Example: in 1986 according to the engineers, the US space shuttle Challenger had one chance out of 200 to have an accident. A posteriori (with data available before the launch), it was shown that in reality the probability of having an accident was higher than 50%... Using statistical methods, NASA would have perhaps avoided the catastrophe!

The statistician

- The statistician takes part in research projects from the collection of data to data analysis and decision making.
- This implies working in an interdisciplinary and international environment.

The statistician

- Examples in industry and business
  - ➢ **Production**: develop products and services for customers satisfaction
  - ➢ **Marketing**: design studies for new products
  - ➢ **IT**: develop software and tools for quality control
  - ➢ **Finance**: model and understand financial markets to provide better portfolio allocations

The statistician

- Exemples in institutions and foundations
  - ➢ **Economics**: study the mechanisms of unemployment, inflation, growth; determine tax brackets
  - ➢ **Statistical offices**: produce economic and social indicators at the national and international level (international organizations)
  - ➢ **Media**: determine radio, TV and other media audience shares to establish advertisement prices
  - ➢ **Politics**: study economic and social factors to support political decisions

The statistician

- Examples in health care and medicine

  ➢ **Epidemiology**: study the development of diseases and epidemics, determine health risk factors

  ➢ **Public health**: prevention of diseases, health costs

  ➢ **Pharmacology**: new treatments (development, control, studies)

  ➢ **Medical research**: image analysis in neurosciences

The statistician

- Examples in natural sciences

  ➢ **Agriculture**: study natural and chemical pesticides, select vegetable species for better adaptation to the environment

  ➢ **Wheather predictions**: study wheather trends, winds, avalanches for better predictions

  ➢ **Hydrology**: evaluate flooding risks for better land management

  ➢ **Climatologie**: determine the factors that influence the evolution of the climate at the planetary level
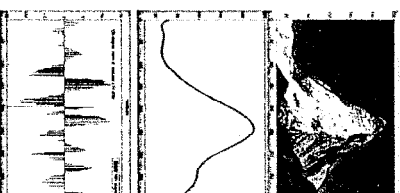
Master of Science
in Statistics

www.masterstat.unige.ch

Ex. 1
Has # of accidents decreased in TI?

Venti morti sulle strade sono ancora troppi

I dati statistici sulle vittime nel 2008 confermano una diminuzione pluriennale

Data: #accidents with deaths between 2000-08

2000	01		02	03	04	05	06	07		08
34	45*		25	26	29	24	17	23		20

* 11 deaths in the Gotthard tunnel accident

Use data form 2000 until 2006 to estimate the average # accidents:

$$\bar{Y}_6 = 24.$$

Assume a Poisson model $Y_i \sim P(\lambda)$,

$i = 1, \cdots, 7$ and test the hypothesis:

$$H_0 : \lambda = 24.$$

p-value:

$$P_{H_0}[Y_7 < 20] = P_{H_0}\left[\frac{Y_7 - 24}{\sqrt{24}} < \frac{20 - 24}{\sqrt{24}}\right]$$
$$\approx \Phi(-0.81) = 21\% \ n.s.$$

Extension: Generalized Linear Models

$$log\lambda_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip},$$

where the $x's$ are covariates, e.g.

$$x_{i1} = 1 \quad \text{if new campaign at time } i,$$
$$= 0 \quad \text{otherwise}$$

and test e.g. the hypothesis $\beta_1 = 0$.

Ex. 2

The Vietnam lottery draft

- Until 1973, American men were drafted to fill the needs of the US armed forces that could not be met by volunteers.

- On December 1, 1969, the first draft lottery was held. It determined the "order of call" for 1970 for all men of draft age, which included all men born in the years 1944 through 1950. Approximately 850000 men were affected by the 1969 draft lottery.

- From: http://www.sss.gov/lotter1.htm

The Vietnam lottery draft

- For the lottery, 366 blue plastic capsules, each containing one date of the calendar year, were dumped in a large glass container.

- The capsules were then drawn out and opened, one by one, and assigned sequentially rising numbers. Congressman Alexander Pirnie (R-NY) drew the first capsule, which contained the date September 14. Thus, all men born on that date, from 1944 through 1950, received the first priority for call to duty. The last date drawn was June 8, which was assigned draft number 366.
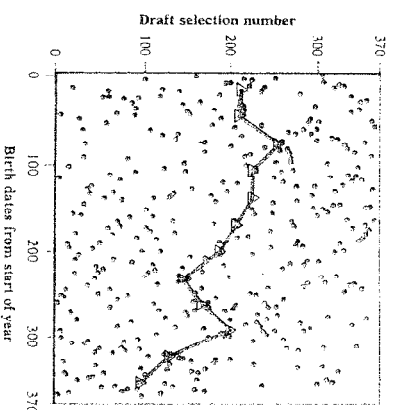
Draft selection vs. birth date
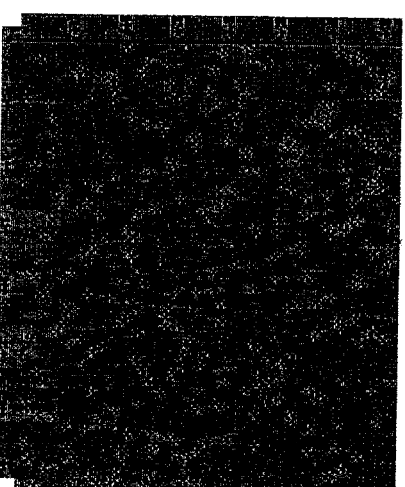
- Monthly medians; Moore &McCabe(1993)
- Lottery random?

Draft selection number / Birth dates from start of year

- Estimate correlation coefficient between draft selection number and birth date
- Estimated correlation = - 0.226
- p-value: Probability that a random lottery would produce a correlation at least as far from 0 as the observed value is 0.1%.
- --> Strong evidence that the lottery was not random

---

Ex. 3
The Challenger disaster

- On January 28, 1986, *Challenger*, a Space Shuttle operated by NASA, broke apart 73 seconds into its flight leading to the deaths of its 7 crew members. The spacecraft disintegrated over the Atlantic, off the coast of Florida at 11:39 a.m. EST.

- Disintegration of the shuttle began after an O-ring seal in its right solid rocket booster (SRB) failed at liftoff. The O-ring failure caused a breach in the SRB joint it sealed, allowing a flare of pressurized hot gas from within the solid rocket motor to reach the outside. (Wikipedia)
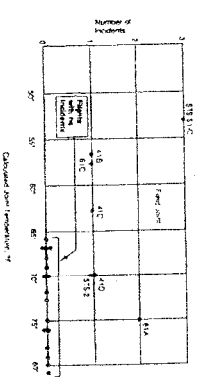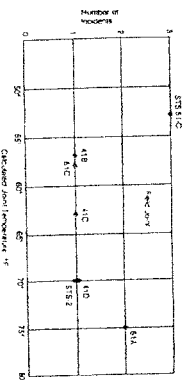

Ex. 3
The Challenger disaster

Ex. 3
The Challenger disaster
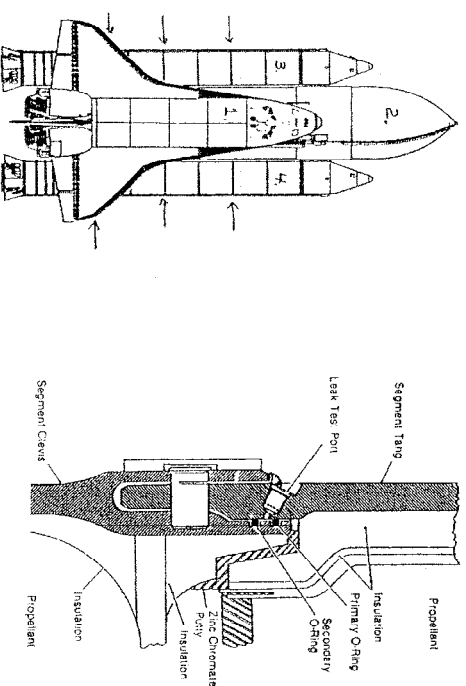
Conclusions

- Statistical reasoning is necessary in the information society.

- Simple statistical models can help to critically evaluate conclusions based on data.

- Some of these aspects can be taught in high school by means of simple examples.

Ex. 3
The Challenger disaster

Prob. of incident = p(Temp)

Logistic regression model:

from Dalal, Fowlkes, Hoadley(1989), *J. Am. Stat. Ass.*

$$log \frac{p}{1-p} = \beta_0 + \beta_1 \cdot Temp$$

Estimation: $\hat{\beta}_0 = 1.38$, $\hat{\beta}_1 = -0.21$

$$\hat{p}(-0.6°C) = [1 + exp(-\hat{\beta}_0 - \hat{\beta}_1 \cdot (-0.6))]^{-1}$$
$$= 0.8$$