Common statistical tools

Hossein Jorjani
hossein.jorjani@slu.se
018 – 67 1964

A word of caution

Don't believe whatever you read in Wikipedia!

I have my own opinion, and I don't give a damn to what Wikipedia or any other individual source says.

Form your own opinion

And, please remember that …

Scientists are by no means bound to follow philosophers' advice on [conducting research], and they don't have such habits.

Jorjani (1995; *cf* Molander (1983))

Induction

Webster

An instance of reasoning	from a part to a whole, or
A conclusion arrived at by reasoning	from particulars to generals, or
Reasoning	from the individual to the universal

Wiktionary

(logic) the derivation of general principles from specific instances

(mathematics) a general proof of a theorem by first proving it for a specific integer (for example) and showing that, if it is true for one integer then it must be true for the next

Deduction

Webster

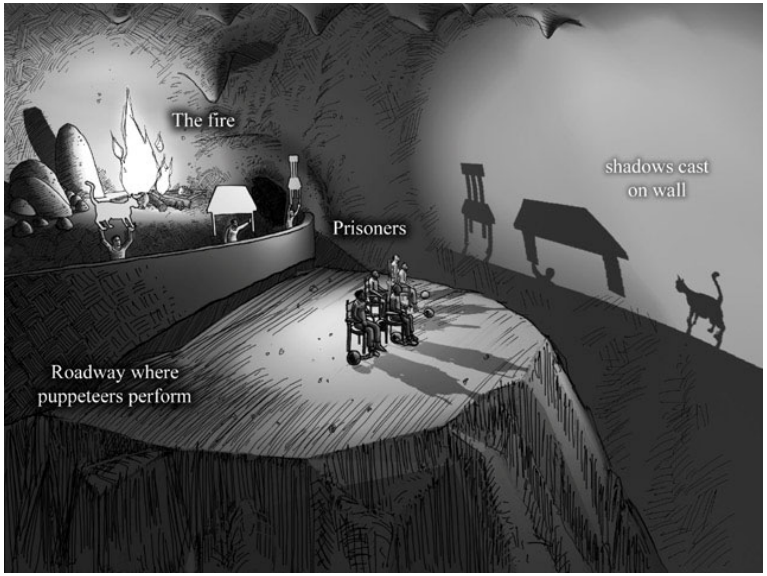
To deduce by logical process:

To draw (a conclusion) necessary from given premises;

To infer (something) about a particular case from a general principle that holds for all such cases.

Wiktionary

A process of reasoning that moves from the general to the specific, in which a conclusion follows necessarily from the premises presented, so that the conclusion cannot be false if the premises are true.



Induction-deduction

Every phenomenon has a true type (archetype=idea).

The archetype can manifest itself in many forms.

Every phenomenon has an essence.

Essence will be known by examining many observations.

No more deduction, no more induction

Francis Bacon 1561-1626

We have suffered enough from deduction!

David Hume 1739-1740

Good science cannot be based on induction!

What the hell are we going to do now?

We have suffered enough from deduction!

Good science cannot be based on induction!

?

Summary

Two schools of statistics

- Frequentist statistics
- Bayesian statistics

Frequentist statistics

Is this an inductive mode of reasoning?

Pearson, 1903

... probably the most correct way of looking at any fraternal correlation table would be to suppose it a random sample of all pairs of brothers which would be obtained by giving a large, or even indefinitely large, fertility to each pair, for what we actually do is to take families of varying size and take as many pairs of brothers as they provide.

Is this an inductive mode of reasoning?

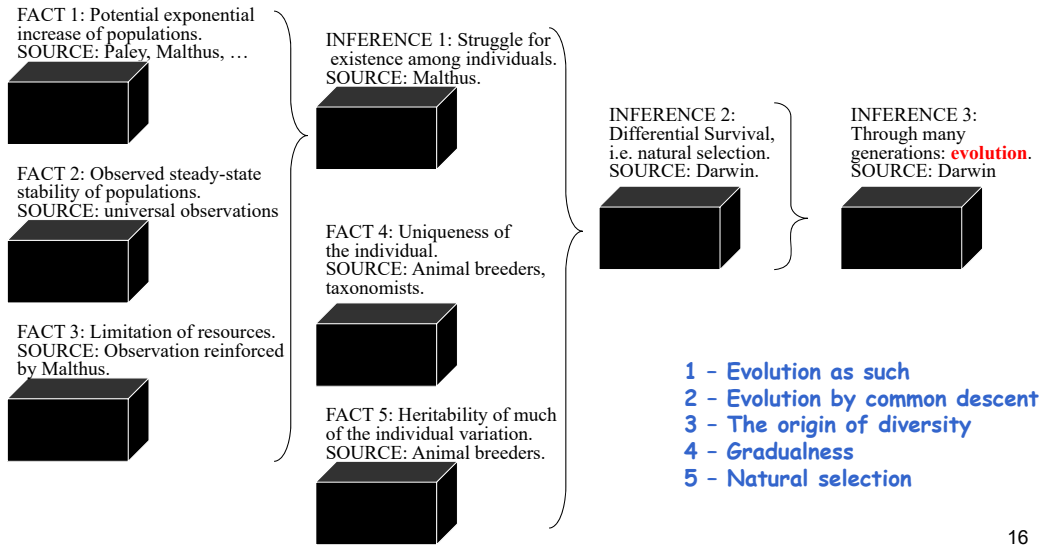
Fisher, 1918

... each pair of brothers is a random sample of two from an infinite fraternity, that is to say from all the sons which a pair of parents might conceivably have produced, ...

Modern (frequentist) statistics

Regression	Francis Galton	1886
Correlation	(Francis Galton	1886)
	Carl Pearson	1903
Analysis of variance	Ronald Fisher	1918

Darwin's theories of evolution



Before Fisher (1918)

Regression & Correlation

- No genetic theory
- Based on phenotypic observations
- Obviously normally distributed

Mendel and evolution

Discrete distribution based on one single gene with large effect: “Hopeful Monster” theory

Continuous distribution with small gradual changes

Biometrician-Mendelian controversy

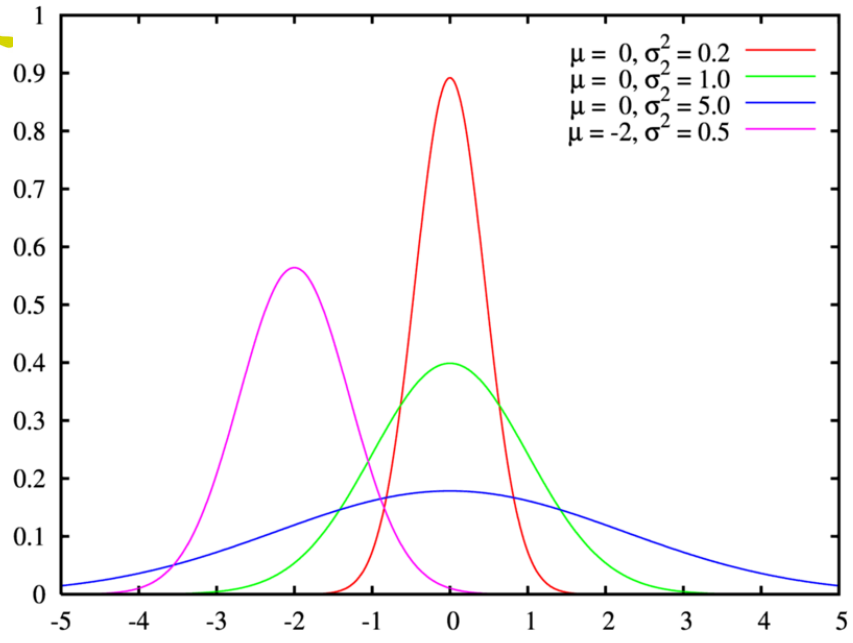
Ronald Aylmer Fisher
(1890-1962)

Geneticist & Statistician

Fisher (1918)

Analysis of variance

- Based on Mendelian genetic theory
- Postulating the aggregate behavior of many genes
- Invoking normal distribution



Inductive statistical Inference theory

Sir R. A. Fisher

(Repeated sampling)

Estimation of unknown parameters

Hypothesis testing

Let's estimate some (sample) parameters

Mean $\bar{X} = \frac{\sum X}{n}$

Variance $V_X = \frac{\sum (X - \bar{X})^2}{n-1}$

Standard deviation $SD = \sqrt{V}$

Properties of variance

For independent X and Y

$$V_{(X+Y)} = V_X + V_Y$$

$$V_{(X-Y)} = V_X + V_Y$$

$$V_{(2X)} = 4V_X \qquad V_{(kX)} = k^2 V_X$$

$$V_{(X+k)} = V_X$$

Two equations for variance (actually SS)

$$\sum (X-\bar{X})^2 = \sum (X^2 + 2X\bar{X} + \bar{X}^2)$$

$$= \sum X^2 - 2\bar{X}\sum X + n\bar{X}^2$$

$$= \sum X^2 - 2\frac{\sum X}{n}\sum X + n\frac{\sum X}{n}\frac{\sum X}{n}$$

$$= \sum X^2 - \frac{(\sum X)^2}{n}$$

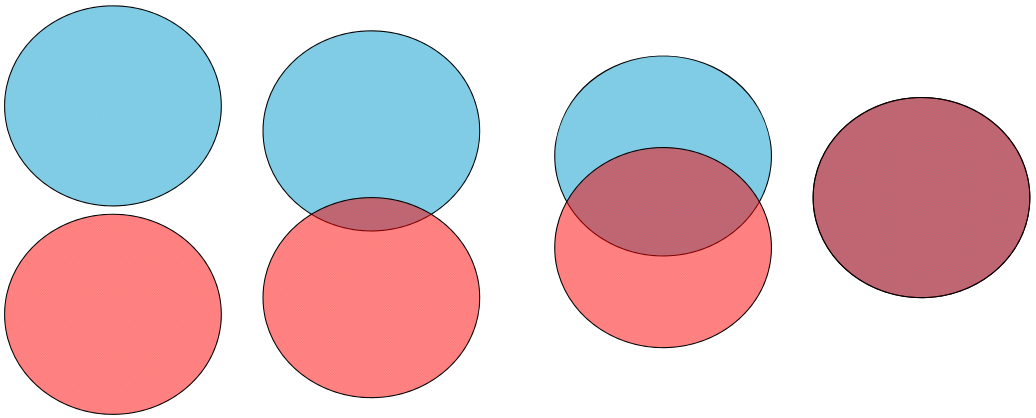
Regression and correlation

$$COV_{Y\square X} = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{n-1}$$

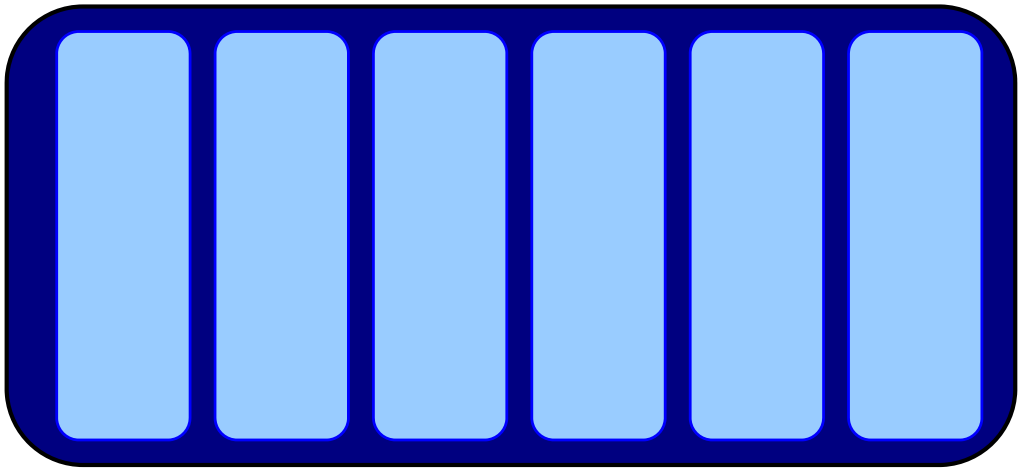
$$b_{Y\square X} = \frac{COV_{Y\square X}}{V_X}$$

$$r_{Y\square X} = \frac{COV_{Y\square X}}{\sqrt{V_X V_Y}}$$

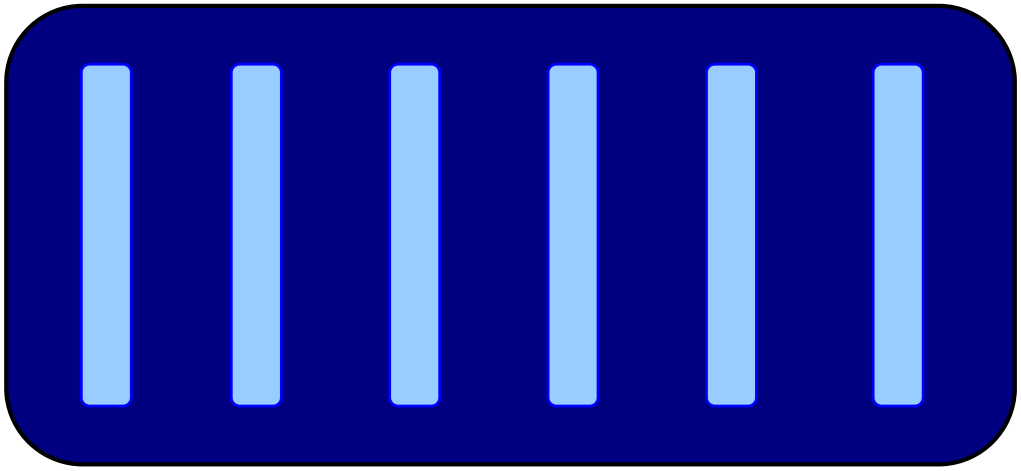
COV = 0 COV=small COV=large COV=1



Within & Between group (co)variance



Within & Between group (co)variance



Standard what?

$$SE_{\bar{X}} = \sqrt{\frac{V_X}{n}}$$

A simple hypothesis test

$$t = \frac{\bar{X} - \mu}{SE_{\bar{X}}}$$

Modeling reality

Observations =

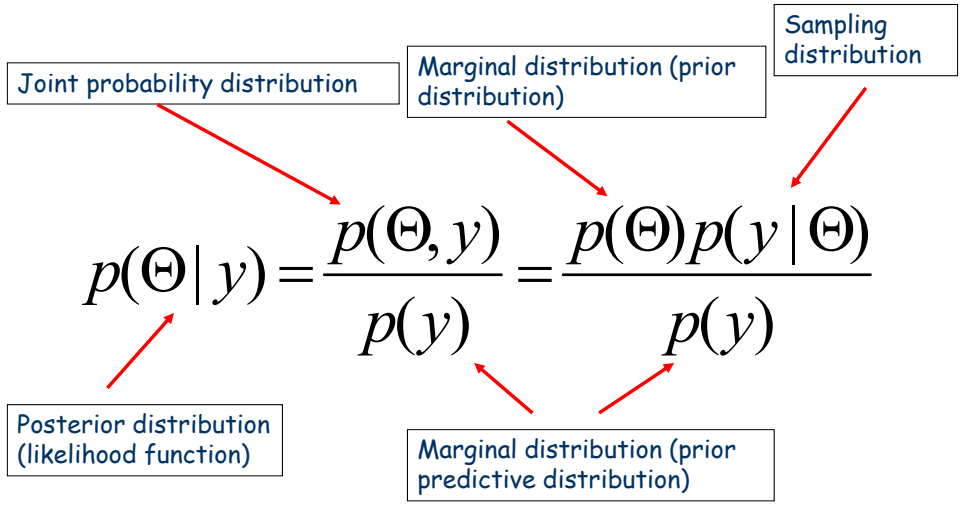
Nuisance parameters +

Parameters of interest +

Unexplained part

?

Bayes' Rule, 1763 (1702-1761)



Hossein Jorjani

Department of Animal Breeding & Genetics
Swedish University of Agricultural Sciences
Box 7023
S – 750 07 Uppsala, Sweden