

1

# Aggregation

- Combination of Observations
- Gain information by throwing away information
- Earliest documented use is 1635

Value of targeted reduction or compression of data

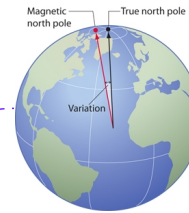
Strong temptation ——— Select ONE observation considered to be the BEST

When was the word, **MEAN** first used ?

Variations of Needle

Difference between True North and Magnetic North

William Borough in 1581



*In Limbouch's the firsteent of*  
Olibee. Anno. 1581.

Longitude		Latitude	
Dep.	Min.	Dep.	Min.
17	50	34	17
18	50	35	17
19	47	30	10
20	46	0	30
21	45	15	31
22	38	0	23
23	34	40	19
24	30	35	14
25	25	30	9

Gellibrand Gresham college professor

Used the word **meane** in 1635

Late 16th century Mean - officially recognized as a method for combining observations

When was the averaging first used, in a practical setting ?

428 BCE Calculating length of ladder using MODE

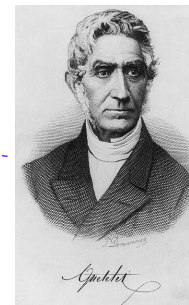
EARLY 1500's To measure length of foot

Take 16 observations and take an average



1800's used in Astronomy + Geodesy

Mid 1700's Shape of Earth



Average Man

Adolphe Quetelet

$$A = z + y \cdot \sin^2 L$$

# 2

## Information Measurement

Information could be measured

Accuracy is related to amount of data

Trial of Pyx

1696-1727  
Issac Newton  
Master of the Mint

Abhraham de Moivre

Variations of sums did not increase proportionately with the number of independent terms

Proportional to square-root of number of independent terms

Averages of samples followed Gaussian if the random variables are Binomially distributed

$$\sigma_{mean} = \frac{\sigma_{sample}}{\sqrt{n}}$$

Pierre Simon Laplace

CLT

Average follow Gaussian even when the individual obs are not Gaussian

Fischer Information

Expected value of the square of score function

Charles S Pierce

$$\mathcal{I}(\theta) = \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta} \log f(X; \theta) \right)^2 \middle| \theta \right] = \int \left( \frac{\partial}{\partial \theta} \log f(x; \theta) \right)^2 f(x; \theta) dx,$$

Correlation

Francis Edgeworth

Class of error distributions that exhibited a paradox

If the probability of the data changes rapidly with theta, then this derivative of score function will be large -- the more information the data will likely be

3

Likelihood

Calibration of inferences using probability

John Arbuthnot

Translated a probability piece from Latin to English

Wrote An Argument against Divine providence

Getting exact  $n/2$  heads in  $n$  tosses is exceedingly small for large  $n$

Question should be about approximately fair

Daniel Bernoulli

Surprising closeness of planetary orbit planes

Cannot attribute to random distribution

David Hume  
Richard Price  
Thomas Bayes

Laplace usage of p-values

Simon Newcomb

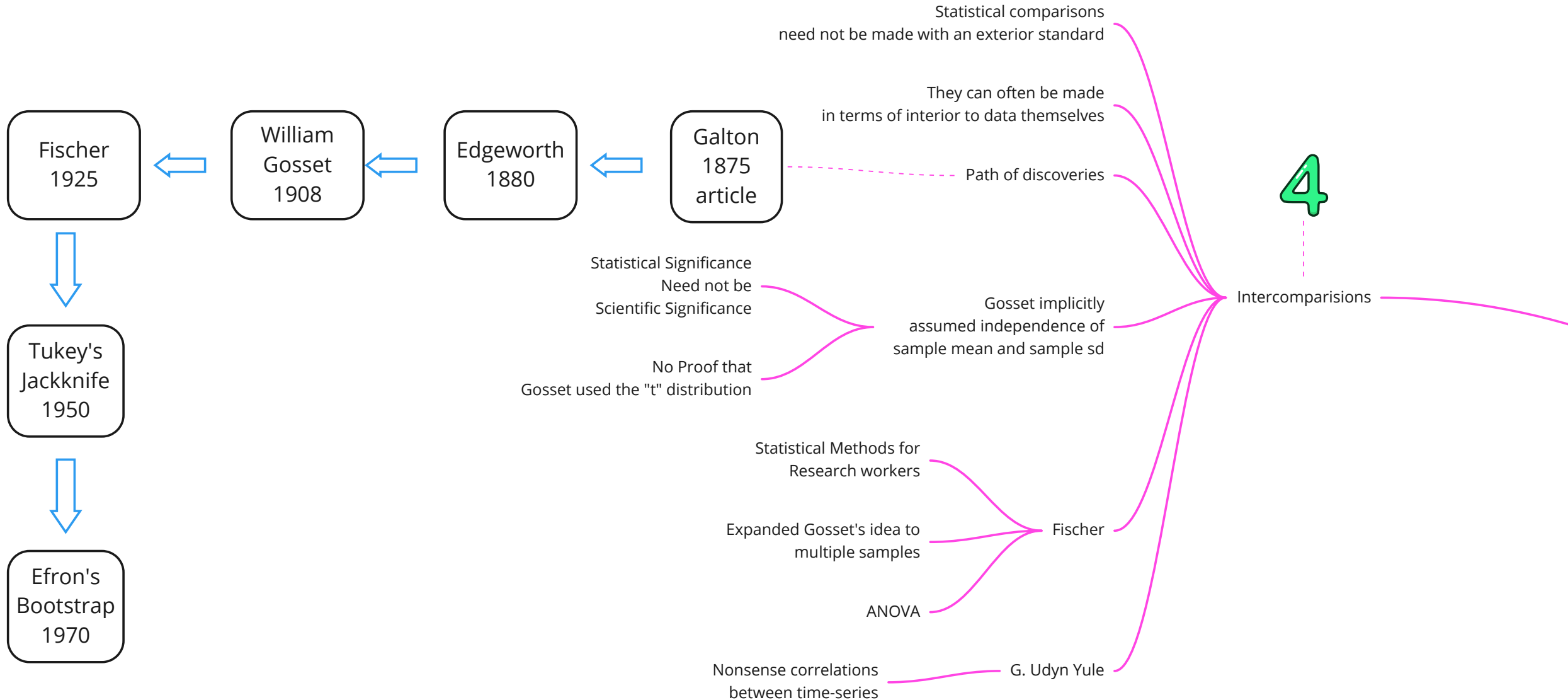
Six bright stars in one single small square

Fischer

1900 Likelihood theory

MLE and Standard Error of the estimate can be found from the likelihood function

Trouble with high-dimensional problems



$a/b = c/d$

Systematically biased

Errors may be quite large

Origin of Species

Darwin's Rule of Three



Half of Century of Enlightenment (1885-1935)



Regression

Rule of Three works when variables are perfectly correlated

Darwin's Predicament

Quincux and its Evolution

Regression towards mean

Solving Darwin Puzzle

Darwin's Cousin

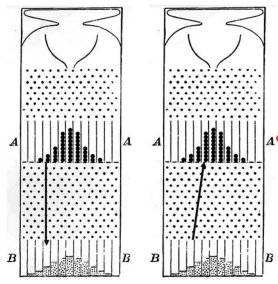
5

Regression

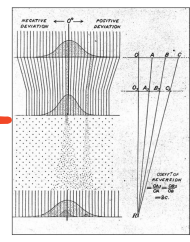
Francis Galton

Why doesn't the variance explode over many generations?

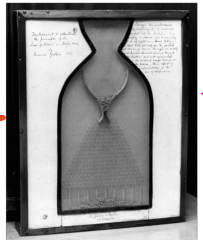
Reconcile population stability over short term scales with diversity in the long run



1889



1877



1873

Developments

Causal Inference

Shrinkage Estimation

Bayesian Inference

Analyzing data via bivariate normal distributions

