DATA ANALYSIS (DA_2022)
A Course for the Master in Genetics and Molecular Biology (English track)

Andrea Giansanti

Dipartimento di Fisica, Sapienza Università di Roma

DA_2022 L1 INAUGURAL LECTURE, Rome march 2, 2022

DIPARTIMENTO DI FISICA

SAPIENZA
UNIVERSITÀ DI ROMA

WHO AM I?

my courses: COMPUTATIONAL BIOPHYSICS (LM physics, sem 1)
PLATFORMS AND ALGORITHMS FOR CORPORATE DECISIONS (LM management, sem 1)

DATA ANALYSIS (DA_2022)
(LM GMB, Neurobiology , sem 2)

CV at: http://www2.phys.uniroma1.it/doc/giansanti/CV/

Room 211 Physics Department Marconi Building

0649914367 / 3385075611

andrea.giansanti@roma1.infn.it **X preferred e-mail**
andrea.giansanti@uniroma1.it

L1 OUTLINE

- Presentation of the course program, logistics, first readings, data collection
- data metadata ontologies: Wittgenstein redux
- elements of scientific method: from Galilei's "simplicity" to Parisi's "complexity"

data, metadata, ontologies: Wittgenstein redux
facts not things the unbearable lightness and of interpretations

1*	The world is all that is the case.
1.1	The world is the totality of facts, not of things.
1.11	The world is determined by the facts, and by their being all the facts.
1.12	For the totality of facts determines what is the case, and also whatever is not the case.
1.13	The facts in logical space are the world.
1.2	The world divides into facts.
1.21	Each item can be the case or not the case while everything else remains the same.
2	What is the case—a fact—is the existence of states of affairs.
2.01	A state of affairs (a state of things) is a combination of objects (things).

* The decimal numbers assigned to the individual propositions indicate the logical importance of the propositions, the stress laid on them in my exposition. The propositions n.1, n.2, n.3, etc. are comments on proposition no. n; the propositions n.m1, n.m2, etc. are comments on proposition no. n.m; and so on.

Basic outline of the 2022 course

- Logic, statistics (decriptive/inferential), probability
- Experiment, event (outcome), probability
- Scientific method: testing hypotheses, making decisions, numeracy
- Inference, machine learning
- The evolutionary space of biological sequences
- Case studies (suggested by colleagues and myself)
- Programming environments (Python, R, Matlab, Excell, shell programming)

DA_2022 TO DO

- Send and e-mail to andrea.giansanti@roma1.infn.it

Subject: DA_2022 ENROLLMENT

Enroll to the DA_2022 Course on the Sapienza Moodle Platform

https://elearning.uniroma1.it/course/view.php?id=14921#section-0

Let us collect a set of original data

- We want to collect real data about the sample of the human population constituted by this class: the people in this course, then please let us discuss NOW how to build up a small database collecting 3 info per person:
- BODY MASS (kg, with one decimal)
- BODY HEIGHT (metres, with 2 decimals)
- SEX (M/F)

I propose to have two volunteers…

Let us make our own dataset: Body Mass Index (BMI)

- The ‘average man’ of Adolphe Quetelet (1796-1874)
- BMI =[Weight(Kg)/(height(m))^2]

Another experiment to be done on the internet

- Collect the properties of several mineral waters available on the Italian Market…
- http://acqueminerali.it/

again: volunteers…

G. GALILEI AND THE REMOUVAL OF THE ANIMAL. PUTTING PHYSICS AGAINST BIOLOGY?

Per tanto io dico che ben sento tirarmi dalla necessità, subito che concepisco una materia o sostanza corporea, a concepire insieme ch'ella è terminata e figurata di questa o di quella figura, ch'ella in relazione ad altre è grande o piccola, ch'ella è in questo o quel luogo, in questo o quel tempo, ch'ella si muove o sta ferma, ch'ella tocca o non tocca un altro corpo, ch'ella è una, poche o molte, né per veruna imaginazione posso separarla da queste condizioni; ma ch'ella debba essere bianca o rossa, amara o dolce, sonora o muta, di grato o ingrato odore, non sento farmi forza alla mente di doverla apprendere da cotali condizioni necessariamente accompagnata: anzi, se i sensi non ci fussero scorta, forse il discorso o l'imaginazione per se stessa non v'arriverebbe già mai. Per lo che vo io pensando che questi sapori, odori, colori, etc., per la parte del suggetto nel quale ci par che riseggano, non sieno altro che puri nomi, ma tengano solamente lor residenza nel corpo sensitivo, sì che rimosso l'animale, sieno levate ed annichilate tutte queste qualità;

[...] Ma che ne' corpi esterni, per eccitare in noi i sapori, gli odori e i suoni, si richiegga altro che grandezze, figure, moltitudini e movimenti tardi o veloci, io non lo credo; e stimo che, tolti via gli orecchi le lingue e i nasi, restino bene le figure i numeri e i moti, ma non già gli odori né i sapori né i suoni, li quali fuor dell'animal vivente non credo che sieno altro che nomi, come a punto altro che nome non è il solletico e la titillazione, rimosse l'ascelle e la pelle intorno al naso [...]

G. Galilei, Il Saggiatore, cap. 47. In varie edizioni.

English Translation

Therefore I say that upon conceiving of a material or corporeal substance, I immediately feel the need to conceive simultaneously that it is bounded and has this or that shape; that it is in this place or that at any given time; that it moves or stays still; that it does or does not touch another body; and that it is one, few, or many. I cannot separate it from these conditions by any stretch of my imagination. But that it must be white or red, bitter or sweet, noisy or silent, of sweet or foul odor, my mind feel no compulsion to understand as necessary accompaniment. Indeed, without the senses to guide us, reason or imagination alone would perhaps never arrive at such qualities. For that reason, I think that tastes, odors, colors, and so forth are no more than mere names so far as pertains to the subject wherein they reside, and that they have their habitation only in the sensorium. Thus, if the living creature were removed, all these qualities would be removed and annihilated. [. . .]

[. . .] I do not believe that, for exciting in us tastes, odors, and sounds there are required in external bodies anything but sizes, shapes, numbers, and slow or fast movements; and I think that if ears, tongues, and noses were taken away, shapes and numbers and motions would remain but not odors or tastes or sounds. These, I believe, are nothing but names, apart from the living animal—just as tickling and titillation are nothing but names when armpits and the skin around the nose are absent [. . .]

A Galilean style: (scientific method vs. …stamp collection)

(Question for you: is a galilean medicine/biology possible? Yes/no ? Why?)
In a nutshell, the galilean method is based on:,

- **Reduction of complexity: remouval of the animal**
- **Measurements (reproducible, a big issue), operational definition**
[Protocols vs intuitive suggestions]
- **Sensible experiences**
- Space /time localizations made **operational**.
- **Use of mathematical models**
- **A programmatic, built-in pessimism:**

[Scientific epistemology is pessimist:
the mind is often wrong: it dynamically requires corrections. (Pascal, Bacone, Galilei)]

- **Prediction: control**. (where data analysis come out)

- **The universal structure of a scientific paper:**

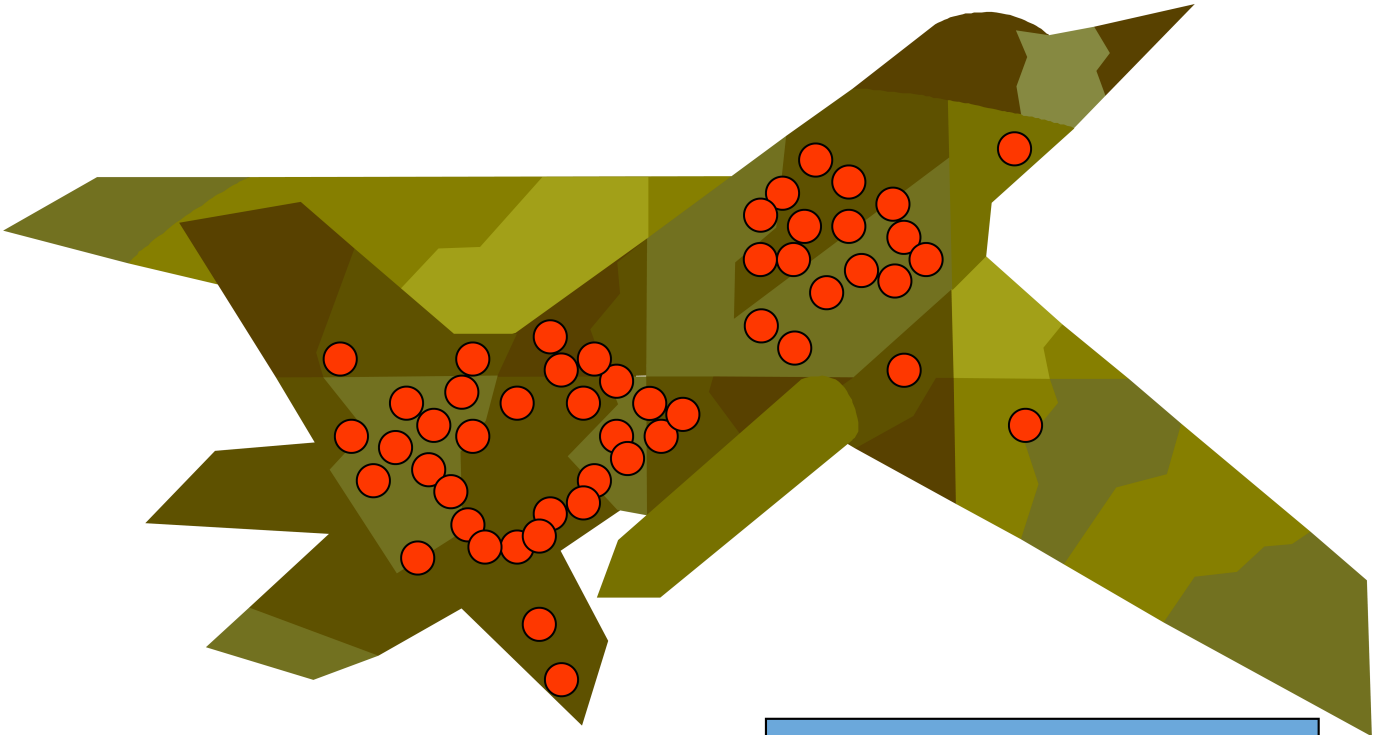
1. Introduction (what is the problem, where it comes from: what is the scientific question, reference to previous works)
2. Materials and methods: written in the style of a cooking recipe (with the aim of sharing a pleasure, willing to be reproduced)
3. Results (Which facts we are proposing to the attention of the community, are they valid ? DA)
4. Discussion (validity of the results, DA)

- The galilean paradigm: based on the “removal of the animal”
- What is physics: the study of material bodies, localized in space and time
- Reference frames + clocks (newtonian time, not percolating, it uniformly flows, always at the same rate
- Biology is based on the careful observation of single cases, then correlated in a qualitative way, based on senses, into classes (classification) (e.g. species)
- **The darwinian shift**: biological time vs physical time
- **The molecular revolution** (Watson & Crick) macroscopic genetic laws can be explained by looking at molecular materials and information

EVOLUTION HAS TO WITH RANDOMNESS

A machine – now obsolete to generate random events (numbers)

Evolution in a nutshell (remembering Anna Tramontano)



Manguel M, Samaniego F.J.,
Abraham Wald's Work on Aircraft Suvivability,
J. American Statistical Association. 79, 259-270, (1984)