Biology 583 – Advanced biometry

Brian McGill

Purpose of course

1. To take the last step from "cookbook" statistics to scientific inference
2. To survey other statistical techniques you may need
3. To familiarize you with the basic tools of the trade

Approach

- To know what is out there and how to learn what you need to know
- Not to be an expert in software, mathematics, advanced statistics
- Survey or capstone course?
- This is an intensive course

General flow:

- Due on Mondays
  - Two (or 3) primary literature papers
    - Download, read
    - Submit comments online BEFORE class
    - Comments are very brief
- Due on Fridays 5:00 PM
  - Some weeks will be a computer exercise
    - Download online
    - Submit online BEFORE class
  - All weeks additional readings
    - Some from textbooks, some are "standard" references
    - Rest – I ask you to skim for 30-60 minutes

Online demonstration

Collaboration policy

- Paper readings on your own – will discuss in class
- Computer exercises
  - May work in a group if:
    - You list collaborators
    - You write your own computer code (without basing it on other peoples)
    - Submit your own computer code and writeup

Grading

- Grading will be based:
  - 15% class discussion (including readings)
  - 25% writeups on readings (plus/check/minus)
  - 40% homeworks (computer exercises) (0-10)
  - 20% take-home "final"
- I will throw out your lowest scores for
  - 3 readings
  - 1 homework
  - No excuses for late/missed – use low scores

What is science? What is statistics? Does science need statistics?

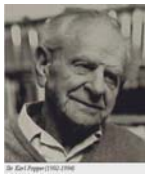
Which of these are scientific

- I know the earth is round
- I know gravity attracts bodies proportional to their mass
- I know God exists
- I know God doesn't exist
- I know a mass extinction is happening
- I know species originated through evolution
- I know species are so perfectly adapted they required intelligent design
- I know aliens exist
- I know aliens have visited Earth
- I know ESP works

Which panel do you trust

- Government sets up three blue ribbon panels to evaluate three aspects of a problem.
- Which one do you trust the most?
  - Scientists
  - Economists
  - Politicians

The demarcation question:



Karl Popper (1902-1994)

- Science vs. pseudoscience
- Is there a distinction?
- What defines the distinction?

Imre Lakatos (1922-1974)

What is the distinction?

- Is there a distinction?
- What is it?

Brief history of physics

- Aristotle (~350's BCE, Greece)
  - Rate of falling (velocity) constant
  - Heavier bodies fall faster
- Euclid (~300 BCE, Greece)
  - Deductive axiomatic approach to geometry
- Ptolemy (~100 AD, Alexandria)
  - Earth at center
  - Planets follow epicycles (circles on circles)

Age of enlightment

- Copernicus (Germany)
  - *De revolutionibus orbium coelestium* (1549)
    - All planets *including Earth* revolve around the sun
- Brahe (Hapsburg empire)
  - Incredibly accurate measures of planet's positions in sky (~1600)
- Kepler (Hapsburg)
  - Kepler's 3 laws (The Harmony of the World 1619):
    - Each planet follows an ellipse with the sun at 1 foci
    - Radius vector from planet to sun sweeps equal areas in equal time (fastest at perigee)
    - Period^2 proportional to mean distance^3

Galileo (Pisa)

- On the motion of bodies (1590)
  - Never published
  - Argued that experiments important
  - Used incline plane as study system for falling bodies
  - Observed regular law: distance intervals increased by odd numbers: 1,3,5,7, or equivalently $x=c\ t^2$
  - Velocity NOT constant, is acceleration
  - Independent of mass of object
- Turned telescope on heavenly bodies, observed moons Jupiter, phases of Venus→ Copernican view of world

Newton

- Calculates that an inverse square law would lead to Kepler's laws
  - $F=M_1M_2/r^2$
  - Moves on to other work until 20 years later Halley asks Newton if he can solve this problem
- Principia (1687)
  - Gravity as a universal force
  - Explains Galileo's 1:3:5 rule & Kepler's Ellipse
  - Less accurate than epicycles
  - Presented in deductive fashion

19th century & electromagnetism

- Maxwell (Scottland)
  - *Electricity & Magnetism* (1873)
  - Maxwell's 4 equations – unification of light, electricity, magnetism, ultimately radio waves
- Consensus
  - Light is a wave
  - It is a wave in the "ether"
- Michelson-Morley
  - Detect the ether by variation in speed of light
  - Failed!
  - Lorenz hypothesized a "contraction" of the ether

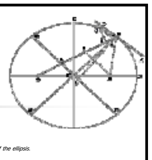
Einstein

- Three papers in 1905
  - Brownian motion described statistically
  - Quantum effect of light (photoelectric effect→solar cells)
    - Challenges wave idea
  - Special relativity
    - No special absolute frame of reference (ether)
    - Space & time behave bizarrely to make this true (contract with speed)
    - Purely deductive – predictions seemed bizarrely out of touch with reality
    - 1919 measurement of stars near sun at eclipse showed light bends – empirical confirmation!

Philosophy of science

- Greco-Roman period (500BCE-300AD)
  - Deduction – reasoning from first principles
  - Euclid's *Elements*
    - 5 axioms → all of geometry via thinking
  - Plato's cave – our sensory reality is a distortion of the underlying beauty of the world
- Middle Ages
  - Authority – "Aristotle said so"

Newton's *Principia*



OF THE MOTION OF BODIES.

SECTION III.

Of the motion of bodies in eccentric conic sections.

PROPOSITION XI. PROBLEM VI.

If a body revolves in an ellipsis; it is required to find the law of the centripetal force tending to the focus of the ellipsis.

Let S be the focus of the ellipsis. Draw SP cutting the diameter DK of the ellipsis in E, and the ordinate Q*v* in *x*; and complete the parallelogram Q*x*PR. It is evident that EP is equal to the greater semi-axis AC: for drawing HI from the other focus H of the ellipsis parallel to EC, because CS, CH are equal, ES, EI will be also equal; so that EP is the half sum of PS, PI, that is (because of the parallels HI, PR, and the equal angles IPR, HPZ), of PS, PH, which taken together, are equal to the whole axis 2AC. Draw QT perpendicular to SP, and putting L for the principal latus rectum of the ellipsis (or for $2BC^2/AC$), we shall have L × QR to L × P*v* as QR to P*v*, that is, as PE or AC to PC; and L × P*v* to G*v*P as L to G*v*; and G*v*P to Qv^2 as PC^2 to CD^2; and by (Corol. 2, Lem. VII) the points Q and P coinciding, Qv^2 is to Qx^2 in the ratio of equality; and Qx^2 or Qv^2 is to QT^2 as EP^2 to PF^2, that is, as CA^2 to PF^2, or (by Lem. XII) as CD^2 to CB^2. And compounding all those ratios together, we shall have L × QR to QT^2 as AC × L × PC^2 × CD^2, or $2CB^2$ × PC^2 × CD^2 to PC × G*v* × CD^2 × CB^2, or as, 2PC to G*v*. But the points Q and P coinciding, 2PC and G*r* are equal. And therefore the quantities L × QR and QT^2, proportional to these, will be also equal. Let those equals be drawn into SP^2/QR, and L × SP^2 will become equal to SP^2× QT^2 / QR. And therefore (by Corol. 1 and 5, Prop. VI) the centripetal force is reciprocally as L × SP^2, that is, reciprocally in the duplicate ratio of the distance SP. Q.E.D.

The same otherwise.

Since the force tending to the centre of the ellipsis, by which the body P may revolve in that ellipsis, is (by Corol. 1, Prop. X.) as the distance CP of the body from the centre C of the ellipsis; let CE be drawn parallel to the tangent PR of the ellipsis; and the force by which the same body P may revolve about any other point S of the ellipsis, if CE and PS intersect in E, will be as PE^2/SP^2 (by Cor. 3, Prop. VII.); that is, if the point S is the focus of the ellipsis, and therefore PE be given as SP^2 reciprocally. Q.E.D.

With the same brevity with which we reduced the fifth Problem to the parabola, and hyperbola, we might do like here: but because of the dignity of the Problem and its use in what follows, I shall confirm the other cases by particular demonstrations.

PROPOSITION XII. PROBLEM VII.

Suppose a body to move in an hyperbola; it is required to find the law of the centripetal force tending to the focus of that figure.

Let CA, CB be the semi-axes of the hyperbola; PG, KD other conjugate diameters; PF a perpendicular to the diameter KD; and Q*v* an ordinate to the diameter GP. Draw SP cutting the diameter DK in E, and the ordinate Q*v* in *x*, and complete the parallelogram QRP*x*. It is evident that EP is equal to the semi-transverse axis AC; for drawing HI, from the other focus H of the hyperbola, parallel to EC, because CS, CH are equal, ES, EI will be also equal; so that EP is the half difference of PS, PI; that is (because of the parallels IH, PR, and the equal angles IPR, HPZ), of PS, PH, the difference of which is equal to the whole axis 2AC. Draw QT perpendicular to SP; and putting L for the principal latus rectum of the hyperbola (that is, for $2BC^2/AC$), we shall have L × QR to L × P*v* as QR to P*v*, or P*x* to P*v*, that is (because of the similar triangles P*xv*, PEC), as PE to PC, or AC to PC. And L × P*v* will be to G*v* × P*v*, as L to G*v*; and (by the properties of the conic sections) the rectangle G*v*P is to Qv^2 as K, PC^2 to CD^2; and by (Cor. 2, Lem. VII.), Qv^2 to Qx^2, the points Q and P coinciding, becomes a ratio of equality; and Qx^2 or Qv^2 is to QT^2 as EP^2 to PF^2, that is, as CA^2 to PF^2, or (by Lem. XII.) as CD^2 to CB^2: and, compounding all those ratios together, we shall have L × QR to QT^2 as AC × L × PC^2 × CD^2, or $2CB^2$ × PC^2 × CD^2 to PC × G*v* × CD^2 × CB^2, or as 2PC to G*v*. But the points P and Q coinciding, 2PC and G*v* are equal. And therefore the quantities L × QR and QT^2, proportional to them, will be also equal. Let those equals be drawn into SP^2/QR, and we shall have L × SP^2 equal to SP^2× QT^2 / QR. And therefore (by Cor. 1 and 5, Prop. VI.) the centripetal force is reciprocally as L × SP^2, that is, reciprocally in the duplicate ratio of the distance SP. Q.E.D.

The same otherwise.

Find out the force tending from the centre C of the hyperbola. This will be proportional to the distance CP. But from thence (by Cor. 3, Prop. VII.) the force tending to the focus S will be as PE^3/SP^2 that is, because PE is given reciprocally as SP^2. Q.E.D.

And the same way may it be demonstrated, that the body having its centripetal changed into a centrifugal force, will move in the conjugate hyperbola.

LEMMA XIII.

The latus rectum of a parabola belonging to any vertex is quadruple the distance of that vertex from the focus of the figure.

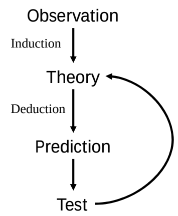
This is demonstrated by the writers on the conic sections.

Enlightment

- Simultaneous with Galileo/Newton
- Bacon (England)
  - *Novum Organum* (1620)
    - IX. --- The cause and root of nearly all evils in the sciences is this -- that while we falsely admire and extol the powers of the human mind we neglect to seek for its true helps.
    - XII. --- The logic now in use serves rather to fix and give stability to the errors which have their foundation in commonly received notions than to help the search after truth. So it does more harm than good.
  - Argued for "induction" based on real world vs. deduction
  - "Father" of the scientific method

What is scientific method (grade school)

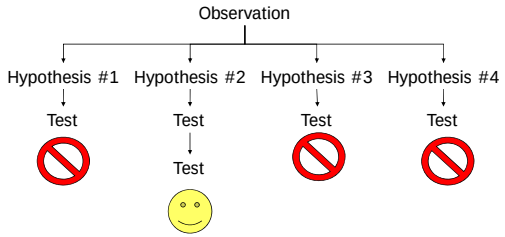
Scientific method (as per grade school)



Fast forward

- Karl Popper (1902-1994)
  - The demarcation problem
    - What demarcates science from non-science (e.g. Freudian psychotherapy)?
    - Rejects induction – can always find data that supports a theory – only corroborative if the prediction is "risky" – possibly false
    - A scientific theory is "prohibitive" – i.e. makes statements that things will not happen and can be falsified
  - → "Falsifiability" is the centerpiece of science

Hypothetico-deductive "NSF" view of science



Problems with this approach?

Practical problems with this view?

- For any data, there are an infinite # of theories explaining this data
- What if your trial set of hypotheses omits the "correct" one

Kreb's enclosures

- Spent $7,000,000 of Canadian taxpayers money
- Observation – hare cycles
- Hypotheses:
  - H1: Predator-prey cycles
  - H2: Hare-food cycle
- Two factors: fertilized/unfertilized, predator exclosure/unenclosed
- Plot size 1 km^2!!! – no replicates
- Results – all 4 treatment combinations cycled!
  - Both hypotheses disproved!

Winfree's fungi

- Observed a spatial pattern in fungi grown on a Petri dish
- Developed 18 hypotheses and a set of 7? Experiments that could potentially falsify each one of them
- Performed experiments
- Falsified all 18!
- Left ecology, went on to become world famous mathematical biologist studying heart defibrillations

Philsophical problems

- Thomas Kuhn (American)
  - *The structure of scientific revolution* 1962
  - Scientists are people and resist falsification of their ideas
  - Falsfication more often results in "touching-up" theories than rejection
  - Progress occurs as periods of:
    - "normal science" where core assumptions are taken as true, falsification just results in decoration
    - Brief interludes of "scientific revolution" when core theories are rejected
    - Example of Ether and Einstein
    - Revolutions often depend on people dying and new people entering the field
  - Falsification is a "rivet-popping" occurrence and applies to whole paradigms at once

Imre Lakatos

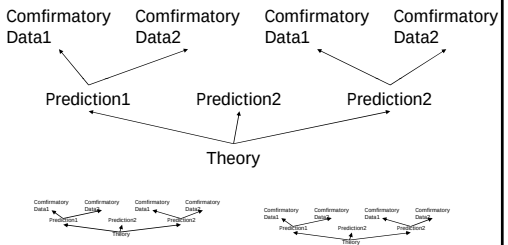
- Hungarian fled to England
  - *The Methodology of Scientific Research Programmes* (1977)
- "Research programmes"
  - "Hard-core" of assumptions that are not falsifiable by its adopters (need some starting point)
  - "Protective-belt" of elucidations that are falsifiable
  - All theories face challenges (appear partly falsified)
  - Research programmes compete
  - Two fates of research programme:
    - Predictive extension – protective belt grows successfully by making "bold
    - "degenerate" – protective belt is post hoc attempt to save hard core
- Reconciled Popper & Kuhn?
- Builds on "Duhem-Quine" idea that hypothesis are intertwined and can't isolate just one for testing

Lakatos redux

- Competition among programmes
- Competition decided by groups of people – gradual "vote" with feet
- Falsification not central
- Positive prediction the measure of success

"'refutations' are not the hallmark of empirical failure, as Popper has preached, since all programmes grow in a permanent ocean of anomalies. What really counts are dramatic, unexpected, stunning predictions: a few of them are enough to tilt the balance; where theory lags behind the facts, we are dealing with miserable degenerating research programmes."

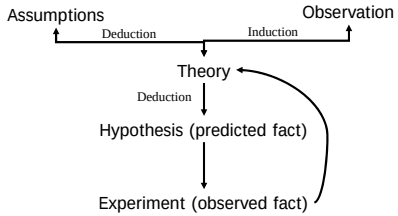
Lakatos



Contrast

- One theory-many predictions (unified theory)
- One observation-many hypotheses

Scheiner (modified)



My key elements of science

- Competition among theories
- Superior theories displace inferior theories
- Ultimate decider is empirical reality

Continuum I

- Hypothetico-deductive
  - Deduction of alternative hypotheses
  - Falsification of hypotheses
- Lakatosian
  - Competition of theories
  - Successful predictions

Causality

- Idea complex
- Descartes advanced idea
  - God as watchmaker – design universe and rules – then set loose – entirely predictable
- More often called “mechanism” in ecology
- First principles vs. relative mechanism

Hierarchy

Ecosystem ecology, macroecology

↑

Community ecology

↑

Population biology

↑

Functional ecology: physiology, behavior, evolutionary ecology

↑

Cellular & Molecular Biology

↑

Chemistry and physics

EEBB is multifactored

- For any given pattern, there are usually several “mechanisms” causing this same pattern
  - Some add to the pattern, some detract
  - The ones that are biggest and go in the majority direction “create” the patttern?
- E.G. Species diversity increases with area
  - Sampling (finite packing of organisms)
  - Habitat heterogeneity
  - Environmental heterogeneity
  - Evolutionary scope
- What does “mechanism” or “causality” mean in a multicausal world?

The elucidation of mechanism

- Correlation is NOT causation
  - Shoe size is correlated with test scores in school children

Koch’s postulates (1884)

- Causation of disease
  1. The bacteria must be present in every case of the disease.
  2. The bacteria must be isolated from the host with the disease and grown in pure culture.
  3. The specific disease must be reproduced when a pure culture of the bacteria is inoculated into a healthy susceptible host.
  4. The bacteria must be recoverable from the experimentally infected host.
- Much more demanding than a controlled experiment
  - What is analog for EEBB?
  - Is it feasible?

Experiments→Mechanism?

- Controlled, replicated, manipulative experiments
  - Hold everything constant
  - Change one thing
  - If area without manipulation (control) differs from area with manipulation, then manipulation must be the cause
- But several challenges to this
  - Must truly hold everything else constant
  - Causality still slippery

Should science care about mechanism?

- NSF answer: Yes
- Robert Peters (A Critique of Ecology 1993)
  - Mechanism is a waste of time and hopeless
    - Ecology is extraordinarily complex, does mechanism even make sense?
  - Ecology would be much better to worry about prediction
- Physics
  - Gravity is a phenomenological concept – it has absolutely no mechanism

Does science *require* experiment?

- Yes:
  - NSF: Yes?
  - Hairston (1989) *Ecological Experiments*
- No:
  - Peters
  - Rajan (Darwin wouldn't get funded today)
  - MacArthur:
    - "Astronomy was a respected science long before ecology, and Copernicus and Galileo never moved a star"

Continuum II: experimentation

- Natural History
- Observational data
- Natural experiment
- Before/After field experiment
- Parallel controlled field experiment
- Laboratory experiment (microcosm)

What fits where on spectrum?

Pros & Cons

- Pros natural history/observation
  - Induction
  - Darwin, etc.
  - Especially appropriate in immature sciences
- Cons natural history
  - No test
    - But observation is a test if hypothesis driven
- Pros natural experiments
  - Realistic
  - Large-scale
  - Cheap
- Cons natural experiments
  - Controls weaker
- Cons controlled, manipulative experiments
  - Expensive
  - Progressively less realistic
  - Limited in scope

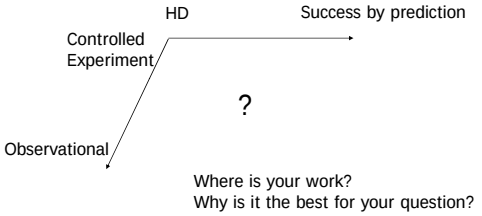
My answer

- Context matters
  - Post hoc test with same data as hypothesis formation is a no-no
  - But hypothesis driven observational data can test theories
- Bottom line – hypothesis driven is vital! Where on spectrum depends on reality of tradeoffs

Reality

- Scientific inference is many faceted
- There is not one right approach
- There are several axes/spectra of approach
- Key is to:
  - Understand the axes
  - Know the history & social context of axes
  - Know the strengths/we

Continua of scientific inference



Where is your work?
Why is it the best for your question?

Ecology

- A strong debate over general principles vs. specific details (Kingsland)
- Community ecology
  - Lawton
    - There are not 10,000,000 kinds of population dynamics
    - WIWACs (World is infinitely wonderful and complicated)
  - Lawton (2000) – failed to come up with general principles, abandon it
  - Simberloff (2004) – will only ever come up with specific examples
- Platt: Szilard debate:
  - "Biology is the science of heterogeneous systems"
  - "There are two kinds of biologists, those looking to see if there is one thing that can be understood, and those who keep saying it is very complicated and that nothing can be understood"

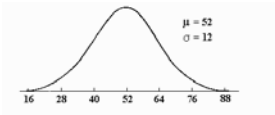
What should science accomplish?

- NSF Criteria:
  - Attention should be given to the significance of research objectives, feasibility of experimental design, rigor of hypotheses, and adequacy of data analysis.
  - evaluation of the scientific quality and importance of the proposed research, and your assessment of the applicant's ability to make original and creative contributions to the scientific area of the proposed research
- "There are worse sins for a scientist than to be wrong. One is to be trivial"
  - Robert MacArthur

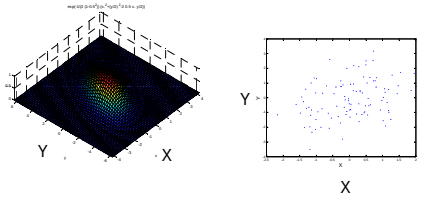
What is statistics?

Statistics

- Statistics is the study of distributions (spreads of measured data)



Multivariate distributions as well



Does science need statistics?

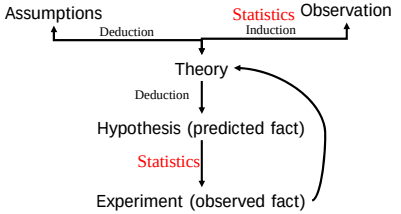
Why are there spreads in data?

- Traditional science posits:
  - Any measurement gets X+ε
    - X the underlying "truth"
    - ε the error – two types:
      - Measurement
      - Process (unmodelled causal factors)
        - What does this mean in EEB where everything
- In my opinion, more useful to look at sources of variation
  - Measurement is the least in EEBB
  - Individual variation (Genetically based)
  - Phenotypic plasticity (Environmental history)
  - Ontogeny
  - Temporal heterogeneity
  - Spatial heterogeneity
- Either way, ***variation is fact of life in EEBB***

Even if data is noisy, why does science need statistics?

- Humans are terrible at probability
- Our brains are wired to convolute probability with cost
  - Low probability, serious (eaten by tiger) treated same as high probability not serious (stub toe)
- What is probability that 2 people in this room have same birthday?

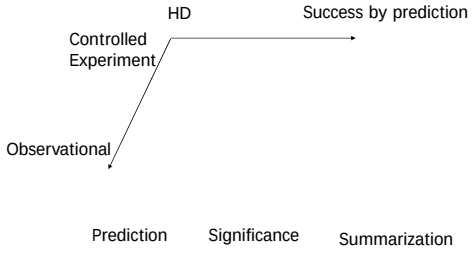
Scheiner (modified)



What can we do with distributions?

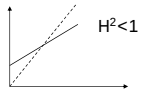
- Summarize data
  - Distributions are hard to get your head around
  - Location (mean, median); spread (variance), skew etc.
  - Especially useful for multivariate data
- Significance (Assign probability that data comes from a specified distribution)
  - $p<0.05$
- Predict (Guess what future samples will look like)
  - Effect size: $Y_{est}=f(x)$, d=3.4
  - Accuracy: r^2 goodness of fit (e.g. χ^2)
- Ties to scientific inference

Continua of scientific inference



A brief history of statistics

- Started with study of probability
  - Guess the motivation?
  - Pascal & Fermat correspondence (1654)
- De Moivre 1733
  - Standard error
  - Normal limit of binomial
- Legendre 1805
  - Least square fit astronomical data
- Guass 1809
  - Residuals of least square are normal
- Galton 1870-1880's
  - Correlation
  - Regression to the mean



History continued

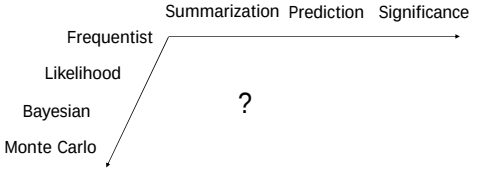
- Around 1900
  - Pearson discovers $\chi 2$
  - Lognormal discovered
  - "Student's" t
- 1919 Fisher – ANOVA
- 1930's Pearson confidence intervals

Philosophical approaches

- Frequentist (p-value) (normal statistics)
- Likelihood (probability based)
- Bayesian (probability with prior expectation)
- Monte Carlo (current data only)

- Each approach has a different answer on how to form confidence intervals, etc.
- Frequentist is:
  - most traditional, "NSFy"
  - Strongest assumptions about nature of data/reality
  - Strongly related to falsification

Continua of statistics



Where is your work?
Why is it the best for your question?

Summary I

- Scientific inference is a mix of:
  - Deduction and induction
  - Falsification and successful prediction
  - Observation and experiment
  - All have pros/cons
  - "Right" depends on question/context
- Statistics help us deal with the variation in data
  - Multiple goals (summarization, significance, prediction)
  - Multiple approaches (frequentist, etc.)

Summary II

- Popper was right:
  - Science vs. pseudoscience important distinction
  - Falsification is key – scientific predictions are falsifiable
- Lakatos & Kuhn were right:
  - Falsification applies to whole bodies of theories and occurs slowly as a sociological process
- Lakatos, Platt were right:
  - Multiple theories important source of objectivity
- Lakatos was right:
  - And Popper, Platt half right
  - Successful prediction ultimate test of theories

Summary III – aka ½ takehome final