

Significance tests (cont.)

For technology significance test with overlap is ok; we're discussing testing theories

Ways to improve theory matters:

A. Investigators

1. Expected size of trend, on theory?
2. Power (Cohen), despite its dangers
3. Campbell-Fiske matrix. Put in main suspected nuisance variables.
4. Pilot study (despite dangers)

B. Editors, referees, journal policy

1. Require replication in one paper
2. Tables: *Separations*, standard deviations, sigmas, overlap, not mere significance levels.
Give p value, not "n.s.," ".05," ".01". Confidence intervals always preferable (translatable to significance but not reverse). Several overlap cuts (e.g., 10, 25, 50, 75, 90%) passed by experimental group; so reader can choose which to emphasize for his purpose. Percent variance accounted for, in some designs. β -weight \neq causal influence, usually.
3. Journal section for short reports of negative pilot studies.

C. Reviewers of literature

Stress power (usually never mention it)
Emphasize that Box Score doesn't tell much about verisimilitude. Delusion that box score (+) over (-) is "pretty good." If negatives accepted, it's terrible (modus tollens)

Psychologists must first grasp Popper point, then water it down like Lakatos. Most in "soft" psychology haven't reached the Popper stage.

D. Theoreticians

Optimism and pessimism both: 1) We're doing fine with present methods;
2) Soft fields cannot aspire to stronger tests
Numerical point predictions only one kind of strong Popperian test. Ex: Wien's Law, no parameters predicted; "Some function of λ^5 ," look at graph, beautiful fit.
Intermediate theory strength

E. PhD educational practices

Require math for research psychologists: More *math* rather than more *statistics* (much of Fisher statistics of little use to us)
Read experiments in other sciences, especially derivation chains with text;
Ex: Millikan oil-drop experiment; but *physics* is not the only model of good science.
History of science? I don't know. I think it has helped me.
Publish or perish pathology is terrible for people and for science. Proxmire had a point. Evaluation of publications should rely more on *Science Citation Index* than mere paper count.

Nevitt Sanford study of achievers: Syndrome of self-doubt

Some important, meaningful scientific questions can't be answered at a given time, lack auxiliary theories or good instruments; Crick & Watson needed (a) exact weights, (b) X-ray, (c) quantitative details of hydrogen bonds, molecular distances.

Probability concept

Important in social sciences, our laws are often stochastic

Epistemic versus object-language use of 'probability'

Carnap: probability₁ (confirmation) and probability₂ (relative frequency):
physical or social; relative frequency of event or attribute in a specified physical class.
Stated in object language. One we use in statistics courses.

Theory of probability began with gambling problems

Ratio of favorable to total "equally likely" ways; Pascal and Fermat invented it;
limit of relative frequency as definition introduced by von Mises, Reichenbach

von Mises theory

Collective: random sequence, relative frequency converges to a limit;
all "place selections" yield same limit

Not "limit" in usual sense of math, rather "stochastic limit"

Relative frequency can't be defined by "equally likely ways"

(e.g., insurance tables don't list equally likely causes of death)

Probability₂ always leads to a number

Probability₁ is metalinguistic concept; epistemic; logical relation between hypothesis
and evidence; relation between propositions, beliefs, statements; prima facie, seems
not to be about relative frequencies.