Options, options and more options

smörgåsbord
buffet
potpourri
bouillabaisse
minestrone
cornucopia

Basic experimental designs

Categorical treatments

- pre- & post- (BACI)
- 1 factor **& controls**
- 2+ factors **& controls**
- subsets & spatial arrays
  - randomized blocks
  - Latin square
  - split plots
- covariates
- repeated measures

Continuous predictors

- pre- & post- (BACI)
- 1 predictor
- multiple predictors
- covariates
  - spatial autocorrelation
  - conditions
- repeated measures

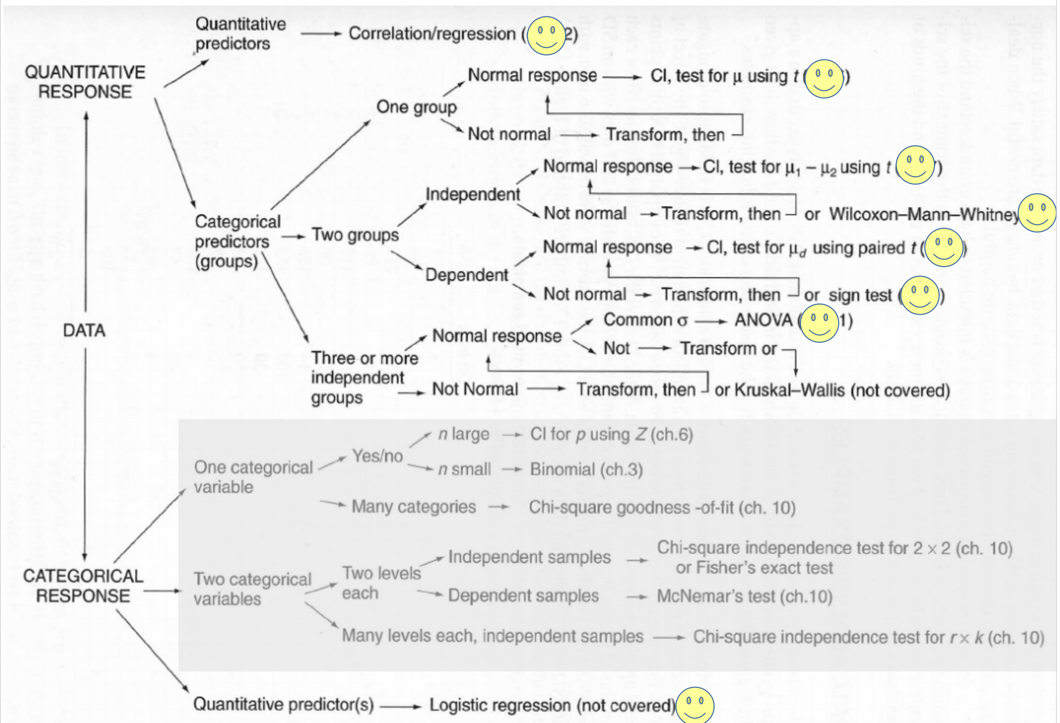

Figure 13.1
A flowchart of inference methods

Samuels & Witmer. 1999. Statistics for the Life Sciences. Prentice-Hall.

Let us count the ways..

1. Z test
2. *t* test (2 independent means)
3. paired *t* test
4. Chi-square
5. ANOVAs (aov, lme)
   a) additive
   b) factorial
   c) randomized block
   d) split plot
   e) repeated measures
6. Regressions
   a) bivariate OLS (lm)
   b) multiple OLS (lm)
   c) bivariate SMA (smatr)
   d) logistic (glm)
7. ANCOVAs
8. Mixed-effects models (lme4, glmmTMB)

AIC-based model selection

fixed and random effects
(a contentious distinction)

<u>Fixed</u>

- effects are constant across observations
- specifically chose treatment levels, may be only options available
- the main point is to compare effects of different levels (e.g. lo vs. hi)
- observations are independent

<u>Random</u>

- effects vary across observations
- treatments are subsamples of all possible options
- effects of different levels is not important
- observations are *inter*dependent

Null hypothesis world view

- I'd be surprised if anything happens in my study system
- No really important processes should occur
- The view from my window is boring
- But I publish something exciting while using stats based on the above

Model comparison world view

- “all models are wrong but some are useful: (G. Box, 1976)

Parsimony

Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so over-elaboration and over-parameterization is often the mark of mediocrity.

Worrying Selectively

Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.

Model comparison world view

- “all models are wrong but some are useful” (G. Box)
- compare models based on “most efficient” or “most plausible”
- parsimony is valued
- more complex models often “best” [but not always]
- models less efficient than the null are bunk [always include a null in the list!]

Increasing Levels of Inference, based on McGill et al. (2006); also see Platt (1964)

A. Single theory test. Are empirical data consistent with or contradictory to a theory?
   1. Correct shape approach. The model displays the same general shape or relation as empirical data.
   2. Curve-fitting approach.—The model fits the empirical data well when the parameters are chosen via curve fitting.

B. Null hypothesis test. Does the theory fit empirical data better than a null hypothesis (H_0) representing a simple scenario.
   1. Hypothetico-deductive approach.—The model fits the empirical data significantly better than H_0 after penalization for number of parameters.
   2. *A priori* parameters approach. The model fits the empirical data significantly better than H_0 after penalization for number of parameters and when the model parameters are chosen independently of the empirical data.

classic frequentist null hypothesis test

C. Multiple, complex predictions approach.—A single model is tested using multiple *a priori* predictions. These predictions are more complex than the data-fitting predictions tested in levels A and B. Two examples of complex predictions are predictions of correlations and predictions about dynamic processes. Each of the complex predictions is then tested at least at level B (i.e., against an appropriate H_0).

D. Model comparison test.—Realistic alternate models (as opposed to null hypotheses) are contrasted against each other, consistent with Platt (1964).
   1. Best theory approach.—The alternative models are ranked according to their match to empirical reality according to some score (such as r^2), and the best model is selected.
   2. Last standing approach.—Rigorous attempts are made to falsify all models in a Popperian fashion until only one model remains unfalsified, which is then accepted as the best model (Platt 1964).
   3. Model weighting approach.—This level not only involves multiple realistic theories, but assigns weights to them according to their explanatory or predictive power. The classical analysis of variance and partitioning of sums of squares is a linear example of such a technique. Akaike weights achieve a similar but not identical result (Burnham and Anderson 1998).
   4. Conditional weights approach.—This approach identifies how the model weights of D3 depend on the scale or context. This level of model comparison answers the question, ''under what conditions is one model better than another in explaining the data?''

maximum likelihood AIC
&
Bayesian inference

McGill, BJ, BA Maurer & MD Weiser. 2006. Empirical evaluation of neutral theory. Ecology 87:1411–1423.

Platt, JR. 1964. Strong inference. Science 146:347–353.

Design & Analysis

"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he only may be able to say what the experiment died of."

- Sir Ronald Fisher

<u>Potential Problems</u>

design-analysis mismatch

confounded treatments

lack of controls

insufficient replication

Soooo… how many samples?

That depends on the tradeoff between
what is possible and what is needed

A solution to minimum sample size for regressions