Paper presented at the TEST conference in Madrid, 15–16 Dec. 2011

Statistics without hypothesis testing

N. T. Longford, SNTL and UPF Barcelona, Spain

NTL@sntl.co.uk

The thesis:

Hypothesis testing has *absolutely no role* in statistical research or practice because it is *logically inconsistent* and poorly suited to *making decisions*

An adaptation of this statement to model selection:

Do not select — combine instead

Model-selection based estimators vs. composite estimators

What's wrong with hypothesis testing

A. The *asymmetry* of the hypothesis and alternative.

a, Failure to reject the hypothesis, but

subsequently acting as if the hypothesis were valid

— *logical inconsistency*

b, Rejecting the hypothesis, but

acting as if the alternative were valid (for certain)

— denial of uncertainty

B. Ignorance of the consequences of the two kinds of bad decisions

The verdict/statement should depend on the inferential agenda

should incorporate the preferences/priorities of the client

Example — ANOVA

Suppose all the assumptions of one-way ANOVA are satisfied,

and the design is balanced: 8 groups $\times$ 7 observations

Q1. The expectation of group 1 (μ_1)

Q2. The within-group variance (σ^2)

A1. & A2. Test the null-hypothesis $(\mu_1 = \ldots = \mu_8)$

a1, Use $\hat{\mu}_1$ if 'reject' and $\hat{\mu}$ (pooled) otherwise

d1, 7 vs. 56 observations used — potential gain: **700%**

a2, Pool within-group SSQs (χ^2_{48}) vs. correct by $\hat{\mu}$ (χ^2_{55})

d2, 48 vs. 55 deg. freedom — potential gain: *14.5%*

No hypothesis test can be good for both inferential tasks

Composition

Using $\hat{\mu}_1$ or $\hat{\mu}$?

Selection:

Attempt to match the performance of the more efficient estimator

Logical basis: Validity $\equiv$ efficiency (**??**)

Composition:

Convex combination $\tilde{\mu} = (1 - b)\hat{\mu}_1 + b\hat{\mu}$

with b (estimated by $\hat{b}$) set to minimise MSE (or another crit.)

Composition: Greater ambition, flexibility, and no glaring weaknesses

Coefficient b depends on the target — combining estimators not models

If you **have to** select: select estimators, not models

Model selection

Models $\mathcal{M}_0\,,\ \mathcal{M}_1\,,\ \ldots,\ \mathcal{M}_K$

model $\mathcal{M}_0$ a priori valid

Single-model-based estimators $\hat{\theta}_0\,,\ \hat{\theta}_1\,,\ \ldots,\ \hat{\theta}_K$

$\hat{\theta}_k$ is unbiased *conditionally* on $\mathcal{M}_k$

Let $\mathcal{I}_k$ be the indicator of selecting $\mathcal{M}_k$

Selected-model-based estimator

$$\hat{\theta}^{\dagger} = \mathcal{I}_0\,\hat{\theta}_0 + \mathcal{I}_1\,\hat{\theta}_1 + \cdots + \mathcal{I}_K\,\hat{\theta}_K$$

— a mixture, in which $\mathcal{I}_k$ correlated with $\hat{\theta}_k$

Confusion of $(\hat{\theta}^{\dagger}\,|\,\mathcal{I}_k = 1)$, $(\hat{\theta}^{\dagger}\,|\,\mathcal{M}_k)$, $(\hat{\theta}^{\dagger}\,|\,\mathcal{M}_k\,\&\,\mathcal{I}_k)$, and $(\hat{\theta}^{\dagger}\,|\,\mathcal{M}_0)$

Making decisions

Interested in θ and have a class of estimators $\hat{\theta}_c$, $c \in \mathcal{R}$

Declare the loss function(s)

e.g., piecewise linear:

$\theta - \hat{\theta}_c$ when $\hat{\theta}_c > \theta$ and $R\left(\hat{\theta}_c - \theta\right)$ when $\hat{\theta}_c < \theta$

a plausible range of penalty ratios R

Choose c that minimises the expected loss.

Bayesian: use the posterior distribution of θ

Variation on the theme:

Equilibrium priors: Priors for which the choice is immaterial

Are all the plausible priors in one subspace?

Examples

Decision theory applied to elementary statistical problems:

NTL, TAS 2010: Inference about the Poisson rate

NTL, SJOS 2011: Comparing two normal random samples

NTL, sbmd 2012: Comparing two variances (indep. normal samples)

A predecessor example: NTL, *Stat. Med.* (2001).

Carryover in crossover trials.

$\hat{\theta}_{\rm A}$ — estimator with the carryover effect absent

$\hat{\theta}_{\rm B}$ — estimator with the carryover effect present

P.R. Freeman (*Stat. Med.*, 1989): Model selection is a bad idea. Use $\hat{\theta}_{\rm A}$

N.T.L. (*Stat. Med.*, 2001): $\tilde{\theta} = (1-b)\hat{\theta}_{\rm A} + b\hat{\theta}_{\rm B}$, with b ... minimax

Conclusion

Hypothesis testing is an invention from the pre-computing age.
It has an amateurish nature not suited to modern statistical practice
that involves decisions related to non-trivial resources and activities.

Hypothesis testing is poorly suited to making intelligent decisions,
because it takes no account of the consequences of the two kinds of errors.

NTL: I do use hypothesis testing in my lectures, and in my work,
but never for any serious business, when getting it right really matters.

Never test (the patience of) *any hippos!!*

THANK YOU