‘Which model ...?’ is the wrong question

N. T. Longford, SNTL and Univ. Pompeu Fabra, Barcelona, Spain

(NTL@sntl.co.uk)

Statistics: making decisions in the presence of uncertainty (analysis) and with limited resources (design)

Model as a conduit:
If I knew the model, then the analysis/inference would be efficient
— absolutely not true, and sometimes not relevant

Select a model and use it for all related inferences (a bad idea) vs.
Combine estimators for a particular purpose (e.g., min MSE for a target)
The false rationale for model selection:

Let’s find a model that looks ‘good’, and then ...

— such a model is a *random* entity — *model uncertainty*

Examples in which the search for a model is/would be a distraction:

- Textbook ANOVA (one-way, with homoscedasticity and normality)
- Clinical trials for comparing two treatments (randomisation)
- Small-area estimation (inference about districts of a country)

What is ‘good’ inference (estimation, hypothesis test, confidence interval)?

Integrity: Adhere to *this* criterion, without any conditioning

A bad (*circular*) criterion: ‘Good’ means: based on a well selected model
One-way ANOVA

(Longford, 2005, JRSS A; 2008, SORT):

Textbook:
Test the hypothesis of equal means — use the selected-model estimator
This estimator is extremely inefficient in some common settings

Use the same model for estimating $\sigma^2$
— a poor strategy (look at the degrees of freedom)

The problem is not with hypothesis testing, but with model choice in general
Bayes factors — no relief/no solution

Combine estimators, with weights specific to the target/estimand
The goal: small MSE
Gross inefficiency of the selected-model based estimator in one-way ANOVA
(Longford, 2008; SORT)
Root-MSE of alternative estimators as functions of the deviation $\mu_1 - \mu$

(Longford, 2008; SORT)
Model selection

Elementary estimators: $\hat{\mu}_1 \sim \mathcal{N}(\mu_1, \frac{1}{n_1}\sigma^2)$ and $\hat{\mu} \sim \mathcal{N}(\mu, \frac{1}{n}\sigma^2)$

Model selection: $\mathcal{I}$ — indicator of selecting model A

$$\hat{\mu}_1^\dagger = (1 - \mathcal{I})\hat{\mu}_1 + \mathcal{I}\hat{\mu}$$

$$\mathbb{E}(\hat{\mu}_1^\dagger) = \mu_1 + p_B \{\mathbb{E}(\hat{\mu} | \mathcal{I} = 1) - \mathbb{E}(\hat{\mu}_1 | \mathcal{I} = 1)\}$$

$$\text{MSE}(\hat{\mu}_1^\dagger; \mu_1) = p_A \text{var}(\hat{\mu}_1 | \mathcal{I} = 0) + p_B \text{var}(\hat{\mu} | \mathcal{I} = 1)$$

$$+ p_A \{\mathbb{E}(\hat{\mu}_1 | \mathcal{I} = 0) - \mu_1\}^2 + p_B \{\mathbb{E}(\hat{\mu} | \mathcal{I} = 1) - \mu_1\}^2$$

$p_A = P(\mathcal{I} = 0); \ p_B = 1 - p_A$. Note: $\mathcal{I}$ and $\hat{\mu}_1$ are correlated

Bias and large MSE are (almost) guaranteed
Combination of estimators

\[ \tilde{\mu}_1 = (1 - b_1)\hat{\mu}_1 + b_1\hat{\mu}, \]

\[ \text{MSE} (\tilde{\mu}_1 ; \mu_1 \mid b_1) = b_1^2 \left\{ g_1 \sigma^2 + (\mu_1 - \mu)^2 \right\} - 2b_1 g_1 \sigma^2 + \frac{\sigma^2}{n_1} \]

\[ b_1^* = \frac{g_1}{g_1 + \frac{(\mu_1 - \mu)^2}{\sigma^2}} \]

where \( g_1 = \frac{1}{n_1} - \frac{1}{n} \)

Substitute \( \hat{b}_1^* \) for \( b_1^* \)

Assess the consequences of over/under-estimating \( (\mu_1 - \mu)^2 / \sigma^2 \)

Scope for incorporating prior information, not necessarily Bayesian

(Longford, 2008, Chapter 1)
Clinical trials

*Randomised* allocation of subjects to two treatments
Estimation of the (constant) treatment effect

Including ‘important’ covariates in (a regression) analysis
— wasting degrees of freedom (Better model — Worse inference)

*Crossover trials* (within-subject contrasts) with design ‘AB and BA’

Freeman (1989, *Stat. Med.*): Do not estimate the *carryover*
— waste of the data from the 2nd period

— reduce the ‘weight’ given to the 2nd period

Do not choose! — combine!!
Small-area estimation

A country with districts \( d = 1, \ldots, D \) and quantities \( \theta_d \); ‘national’ value \( \theta \)

Notation: \( \hat{\theta}_d \sim Z(\theta_d, v_d), \hat{\theta} \sim Z(\theta, v) \) and \( c_d = \text{cov}(\hat{\theta}_d, \hat{\theta}) \)

A setting similar to ANOVA, except that \( D \gg \) — random effects (??)

Sample size sufficient for estimating \( \theta \), but not for \( \theta_d \) for some \( d \)

ANOVA irrelevant — composition of (unbiased) estimators \( \hat{\theta}_d \) and \( \hat{\theta} \):

\[
\tilde{\theta}_d = (1 - b_d) \hat{\theta}_d + b_d \hat{\theta}
\]

\[
b_d^* = \frac{v_d - c_d}{v_d + v - 2c_d + \sigma_B^2} = \frac{v_d}{v_d + \sigma_B^2}
\]

\[
\sigma_B^2 = \text{var}_D(\theta_d) \quad \text{— estimate} \ \sigma_B^2 \ \text{and study sensitivity} \ (\hat{v}_d)
\]

Extensions for auxiliary information (Longford, 2005)
Conclusion

The importance of model selection is vastly over-rated because of not appreciating the pervasiveness of uncertainty and ignoring the basics of conditional probabilities and distributions.

Asymptotic theory (for AIC, BIC, u&IC) is questionable for an essentially small-sample problem.

Hypothesis testing (and intermediate decision, incl. model selection) — a *steam engine* in the age of the *iPod* because it is oblivious to the consequences of the errors I and II.

*Examples*: 1. The Albanian long jumper Shenki Xhadni (2044);
2. Crossing the road in uptown Bendery during a Euro game.


