# Some models are useful — for confidence intervals?

John Copas

University of Warwick, UK

*jbc@stats.warwick.ac.uk*

1. Introduction
    - parameter of interest
    - empirical models
    - editor's dilemna
2. Examples
    - contingency table
    - subset selection in regression
3. General theory
4. Discussion

- Confidence Intervals (CI) — significance level $\alpha$

  $\phi$ = parameter of interest

  $\phi_0 \in CI$: not significant, "$\phi_0$ is consistent with the data"

  $\phi_0 \notin CI$ : significant, "$\phi_0$ is not supported by the data"

- Model (M)

  $CI = CI(data, M) = CI_M$

- Assumptions

  scientific model : based on knowledge or shared understanding of the science

  empirical model : based on assumptions which "fit the data"

*Texts books imply that ...*

- we should treat $M$ as fixed.
- but we should check that $M$ fits the data by a test of goodness of fit or by graphical diagnostics

If $M$ is a scientific model, then OK.

But if $M$ is an empirical model, can we really treat $M$ as fixed ??

*An editor's dilemna*

A paper is submitted for publication ...

- Author claims significance by showing that $\phi_0 \notin CI_M$
- Referee does an independent re-analysis of the data and disagrees : he finds that $\phi_0 \in CI_{M'}$
- Suppose that $M$ and $M'$ fit the data equally well

What will the editor decide?

**Example 1 : Heart Disease and Blood Pressure**

| | BP=1 | BP=2 | BP=3 | BP=4 |
|---|---|---|---|---|
| $D$ | 20 | 28 | 20 | 24 |
| $\bar{D}$ | 388 | 527 | 204 | 118 |
| | 408 | 555 | 224 | 142 |

Problem : calculate a CI for $\phi = P(\mathrm{D}|\mathrm{BP} = 2)$

CI for $\phi = P(\mathrm{D}|\mathrm{BP} = 2)$

1. *saturated model* $M_S$

   $CI_{M_S} = (0.035, 0.072)$

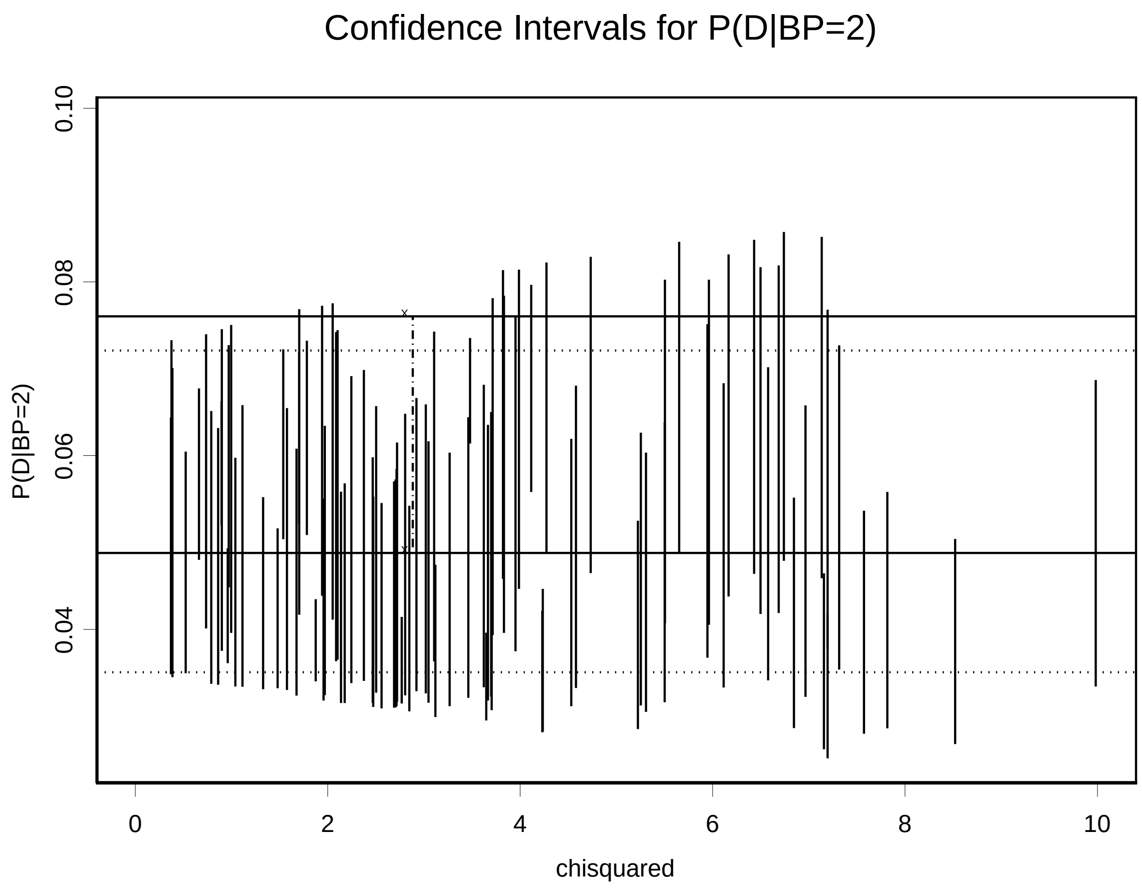
2. *logistics model* $M_L$

   $CI_{M_L} = (0.049, 0.076)$

But why logistic?

- probit GLM
- complementary log-log GLM
- BP on a log scale
- ....

all give different CI's, and all can be checked by the $\chi^2$ test




vertical lines = CI's for 100 different models

## Example 2 : Multiple Regression

$y$ = oxygen uptake per body mass

$x$=covariates

| $x$ | $\hat{\beta}$ | se | $t$ | P |
|---|---|---|---|---|
| age | -0.200 | 0.094 | -2.14 | .04 |
| running time | -2.872 | 0.342 | -8.41 | $< 10^{-4}$ |
| running pulse | -0.354 | 0.115 | -3.07 | .01 |
| max pulse | 0.206 | 0.139 | 1.49 | .15 |
| sex | 1.919 | 1.338 | 1.43 | .16 |

Problem : calculate a CI for $\phi = \mathrm{E}(y|x = \xi)$

CI for prediction $\phi = \mathrm{E}(y|x = \xi)$

Can calculate CI from regression model using

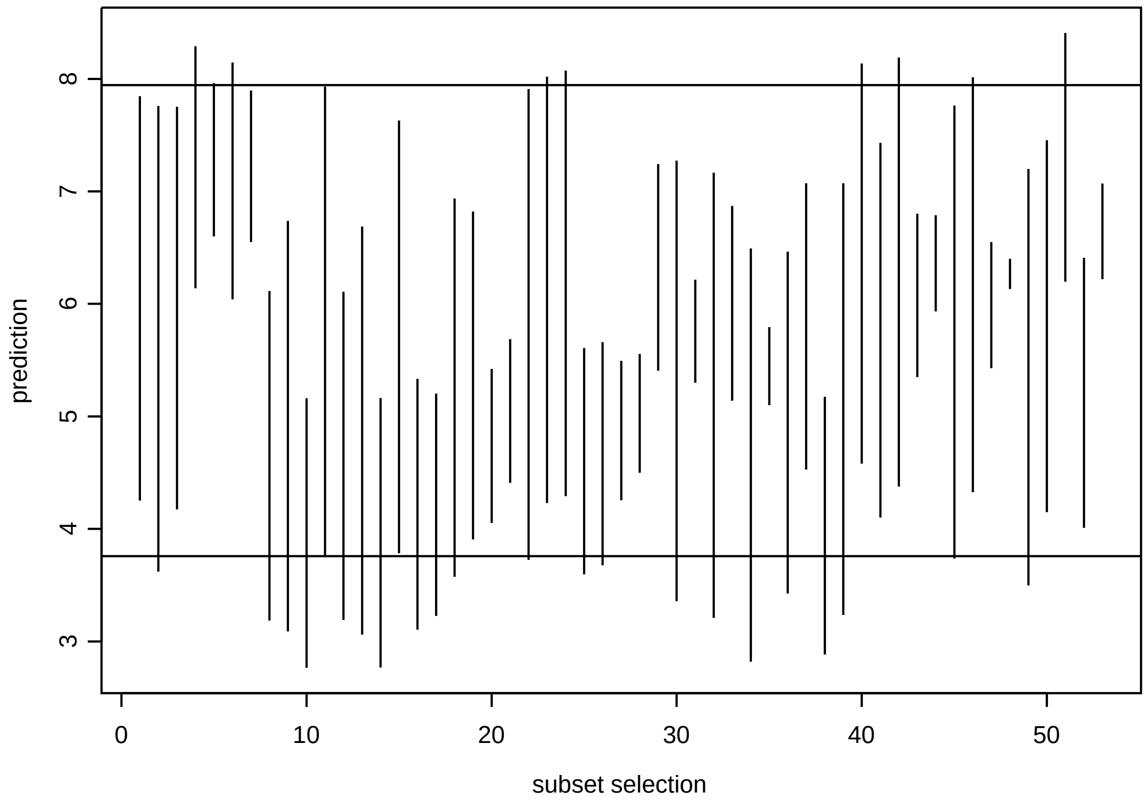
- all $x$'s
- significant $x$'s
- (linear) age-standardized $x$'s
- ....

More generally

- regress on $Ax$ for some matrix A

All models can be checked by the F-test on the residual sums of squares

Prediction confidence intervals




vertical lines = CI's for subsets/transformed subsets with non-significant $F$

**Asymptotic theory**

Scientific assumptions $\rightarrow$ model $M_0$

*Ex1* saturated model

*Ex2* full regression

Empirical assumptions $\rightarrow$ model $M$

*Ex1* logistic regression

*Ex2* regress $y$ on $Ax$ (e.g. subset regression)

$\phi$ = parameter of interest

*Ex1* $\phi = P(D|BP = 2)$

*Ex2* $\phi = \mathrm{E}(y|x = \xi)$

Under $M_0$

- MLE $= \hat{\phi}_{M_0} = \hat{\phi}$
- $\mathrm{Var}(\hat{\phi}) = \sigma^2_{M_0} = \sigma^2$
- $CI_{M_0} = CI = \hat{\phi} \pm d_\alpha \sigma$

Under $M$

- MLE $= \hat{\phi}_M$
- $\mathrm{Var}(\hat{\phi}_M) = \sigma^2_M$
- $CI_M = \hat{\phi}_M \pm d_\alpha \sigma_M$

Assume $M \subseteq M_0$ ($M$ is a special case of $M_0$)

Then

$$\begin{aligned} \sigma_M^2 &\leq \sigma^2 \\ \mathrm{Var}_M(\hat{\phi}_M - \hat{\phi}) &= \sigma^2 - \sigma_M^2 \\ z_M &= \frac{\hat{\phi}_M - \hat{\phi}}{\sqrt{\sigma^2 - \sigma_M^2}} \sim_M N(0,1) \end{aligned}$$

Hence

$$\begin{aligned} CI_M &= \hat{\phi}_M \pm d_\alpha \sigma_M \\ &= \hat{\phi} + (\hat{\phi}_M - \hat{\phi}) \pm d_\alpha \sigma_M \\ &= \hat{\phi} + \{z_M \sqrt{\sigma^2 - \sigma_M^2} \pm d_\alpha \sigma_M\} \end{aligned}$$

$$CI_M = \hat{\phi} + \{z_M \sqrt{\sigma^2 - \sigma_M^2} \pm d_\alpha \sigma_M\}$$

Let $G_M$ be a goodness of fit test so we accept model $M$ as an empirical model if $G_M \leq g_M$ ($g_M = \alpha$-level threshold)

$$\textit{Ex1} \quad G_M = \chi^2$$

$$\textit{Ex2} \quad G_M = F$$

Then we can show that

- $$\max_{M \subseteq M_0} \{|z_M| \big| G_M \leq g_M\} \geq d_\alpha$$

  (attained when $G_M = |z_M|$)

- $$\{\sqrt{\sigma^2 - \sigma_M^2} + \sigma_M\} \leq \sqrt{2}\sigma$$

  (attained when $\sigma_M^2 = \frac{1}{2}\sigma^2$)

Hence

$$\begin{aligned}\hat{\phi} + d_\alpha \sigma &\leq \min_{G_M} \max_{M} CI_M^{(upper)} \leq \hat{\phi} + \sqrt{2} d_\alpha \sigma \\ \hat{\phi} - \sqrt{2} d_\alpha \sigma &\leq \max_{G_M} \min_{M} CI_M^{(lower)} \leq \hat{\phi} - d_\alpha \sigma\end{aligned}$$

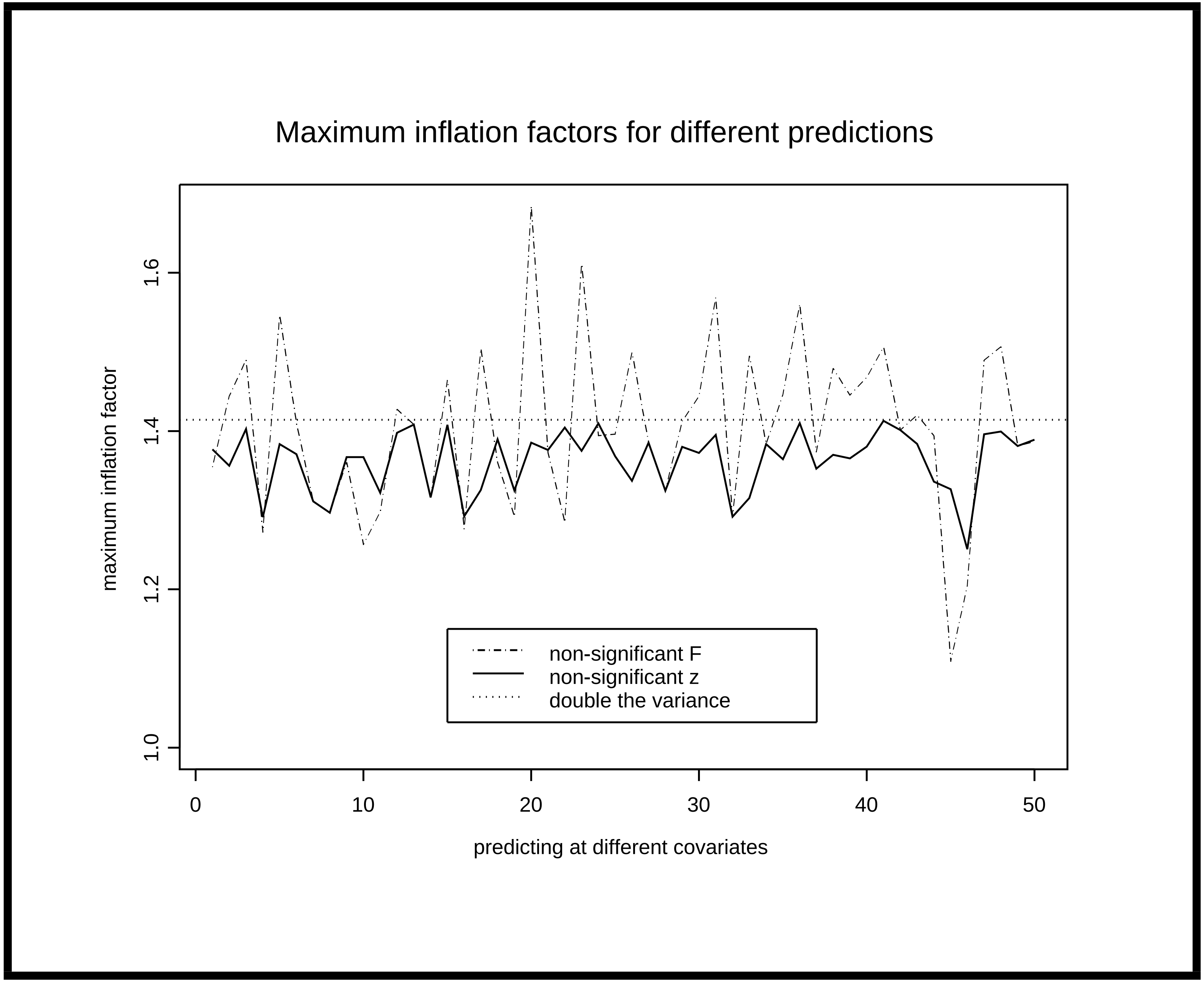
Equivalently:

$$\text{inflation factor } = f_M = \frac{CI_M^{upper} - \hat{\phi}}{CI^{upper} - \hat{\phi}}$$

then

$$1 \leq \min_{G_M} \max_{M} f_M \leq \sqrt{2}$$

$f_M = \sqrt{2} \Rightarrow$ assume model $M_0$ but "double the variance"


Maximum inflation factors for different predictions
maximum inflation factor
1.0
1.2
1.4
1.6
0
10
20
30
40
50
predicting at different covariates
non-significant F
non-significant z
double the variance

**Example 2 : Multiple Regression**

$M$: regress $y$ on $Ax$ where $A$ is $k_M \times k$, $k_M \leq k$

$$X^T X = I_k \ , \ AA^T = I_{k_M} \ , \ \xi^T \xi = 1$$

$$\hat{\phi}_M = \hat{\beta}^T A^T A \xi \quad , \quad \mathrm{Var}(\hat{\phi}) = n^{-1} \xi^T A^T A \xi$$

$$z_M = \frac{n^{\frac{1}{2}} \hat{\beta}^T (I_k - A^T A)\xi}{\{\xi^T (I_k - A^T A)\xi\}^{\frac{1}{2}}}$$

$$f_M = \frac{d_\alpha n^{-\frac{1}{2}} (\xi^T A^T A \xi)^{\frac{1}{2}} - \hat{\beta}^T (I_k - A^T A)\xi}{d_\alpha}$$

Then

$$\max_M \{ f_M \,\big|\, |z_M| \leq d_\alpha \}$$

$$= \begin{cases} \sqrt{2} & \text{if} \quad |r| \leq 1/\sqrt{2} \\ |r| + (1 - r^2)^{\frac{1}{2}} & \text{if} \quad |r| > 1/\sqrt{2} \end{cases}$$

where

$$r = \frac{\hat{\beta}^T \xi}{(\hat{\beta}^T \hat{\beta})^{\frac{1}{2}} (\xi^T \xi)^{\frac{1}{2}}}$$

(angle between $\hat{\beta}$ and $\xi$)

**Discussion**

- Two equally well-fitting models can give surprisingly different confidence intervals
- If $G_M \leq g_M$ then
  - length $(CI_M) \leq$ length $(CI)$
  - $CI_M$ may overlap $CI$
- If $\widetilde{CI} = \hat{\phi} \pm 2.77\sigma$, then for any $\phi \in \widetilde{CI}$ and for any 5% goodness of fit test, $\phi \in CI_M$ for some well-fitting model $M$
- If $\phi_0$ is significant at the 5% level for *all* well-fitting models $M$, then $\phi_0$ has to be significant under model $M_0$ at *at least* the 0.5% level

- *Some models are useful — for confidence intervals?*

  If

  "useful model" $= \{CI_M \subset CI\}$

  then

  "useful" $\neq \{G_M \leq g_M\}$

- it is impossible to check from the data that a model will be "useful" for all possible $\phi$

- statistical inference needs *external* evidence that a model (or class of models) is sensible