

LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE

28 March 2017

D.R.Cox
Nuffield College, Oxford

Some statistical themes

- Formulation of research questions
- Planning of investigation

Some statistical themes

- Formulation of research questions
- Planning of investigation
- Metrology

Some statistical themes

- Formulation of research questions
- Planning of investigation
- Metrology
- Collection of data

Some statistical themes

- Formulation of research questions
- Planning of investigation
- Metrology
- Collection of data
- Preliminary analysis

Some statistical themes

- Formulation of research questions
- Planning of investigation
- Metrology
- Collection of data
- Preliminary analysis
- Formal analysis

Some statistical themes

- Formulation of research questions
- Planning of investigation
- Metrology
- Collection of data
- Preliminary analysis
- Formal analysis
- Conclusions of analysis

Some statistical themes

- Formulation of research questions
- Planning of investigation
- Metrology
- Collection of data
- Preliminary analysis
- Formal analysis
- Conclusions of analysis
- Interpretation

Some types of dependence analysis of a single outcome on v explanatory variables measured on n individuals

- Small v , modest n

Some types of dependence analysis of a single outcome on v explanatory variables measured on n individuals

- Small v , modest n
- Appreciable v , $n > 5v$, say

Some types of dependence analysis of a single outcome on v explanatory variables measured on n individuals

- Small v , modest n
- Appreciable v , $n > 5v$, say
- Small n and large v , for example $n = 40$, $v = 50,000$
- Joint work with Heather Battey, Department of Mathematics, Imperial College

- Small n and large v , for example $n = 40, v = 50,000$
- Joint work with Heather Battey, Department of Mathematics, Imperial College

- Small n and large v , for example $n = 40, v = 50,000$
- Joint work with Heather Battey, Department of Mathematics, Imperial College
- Need to formalize the notion of sparsity
- Lasso (Tibshirani, 1996): penalized least squares (or maximum likelihood)

- Small n and large v , for example $n = 40, v = 50,000$
- Joint work with Heather Battey, Department of Mathematics, Imperial College
- Need to formalize the notion of sparsity
- Lasso (Tibshirani, 1996): penalized least squares (or maximum likelihood)

Minimize

$$\| y - \beta^T x \|^2 - \lambda \| \beta \|^2$$

Ridge regression

Minimize

$$\| y - \beta^T x \|^2 - \lambda \| \beta \|^2$$

Ridge regression

Minimize (Tibshirani, 1996)

$$\| y - \beta^T x \|^2 - \lambda \| \beta \|_{L_1}$$

Lasso

Minimize

$$\| y - \beta^T x \|^2 - \lambda \| \beta \|^2$$

Ridge regression

Minimize (Tibshirani, 1996)

$$\| y - \beta^T x \|^2 - \lambda \| \beta \|_{L_1}$$

Lasso

New approach

Aim for confidence set of possible models

Minimize

$$\| y - \beta^T x \|^2 - \lambda \| \beta \|^2$$

Ridge regression

Minimize (Tibshirani, 1996)

$$\| y - \beta^T x \|^2 - \lambda \| \beta \|_{L_1}$$

Lasso

New approach

Aim for confidence set of possible models

Represent variables in say a $10 \times 10 \times 10 \dots$ hypercube and test in sets of 10

Repeat

Minimize

$$\| y - \beta^T x \|^2 - \lambda \| \beta \|^2$$

Ridge regression

Minimize (Tibshirani, 1996)

$$\| y - \beta^T x \|^2 - \lambda \| \beta \|_{L_1}$$

Lasso

New approach

Aim for confidence set of possible models

Represent variables in say a $10 \times 10 \times 10 \dots$ hypercube and test in sets of 10

Repeat

Informal checks

Minimize

$$\| y - \beta^T x \|^2 - \lambda \| \beta \|^2$$

Ridge regression

Minimize (Tibshirani, 1996)

$$\| y - \beta^T x \|^2 - \lambda \| \beta \|_{L_1}$$

Lasso

New approach

Aim for confidence set of possible models

Represent variables in say a $10 \times 10 \times 10 \dots$ hypercube and test in sets of 10

Repeat

Informal checks

Parallel with earlier work

Implications