Some Recent Progress in Simple Statistical Methods

Andrew Gelman
Department of Statistics and Department of Political Science
Columbia University

24 Jan 2008

Simple statistical methods

- Much of statistics is simple methods (more precisely, methods that seem simple to us):
  - $0.5/\sqrt{n}$, linear regression, logistic regression, ordered logit, mean and variance for stratified sampling, ...
- There are some simple methods out there that I don't like so much

Simple statistical methods

- Much of statistics is simple methods (more precisely, methods that seem simple to us):
  - $0.5/\sqrt{n}$, linear regression, logistic regression, ordered logit, mean and variance for stratified sampling, ...
- There are some simple methods out there that I don't like so much
  - Multiple comparisons corrections, stepwise regression, rules for detecting outliers, ...

Simple statistical methods

- Much of statistics is simple methods (more precisely, methods that seem simple to us):
  - $0.5/\sqrt{n}$, linear regression, logistic regression, ordered logit, mean and variance for stratified sampling, ...
- There are some simple methods out there that I don't like so much
  - Multiple comparisons corrections, stepwise regression, rules for detecting outliers, ...

Simple statistical methods

- Much of statistics is simple methods (more precisely, methods that seem simple to us):
  - $0.5/\sqrt{n}$, linear regression, logistic regression, ordered logit, mean and variance for stratified sampling, ...
- There are some simple methods out there that I don't like so much
  - Multiple comparisons corrections, stepwise regression, rules for detecting outliers, ...

Simple statistical methods

- Much of statistics is simple methods (more precisely, methods that seem simple to us):
  - $0.5/\sqrt{n}$, linear regression, logistic regression, ordered logit, mean and variance for stratified sampling, ...
- There are some simple methods out there that I don't like so much
  - Multiple comparisons corrections, stepwise regression, rules for detecting outliers, ...

New simple statistical methods

- Why new simple methods?
  - Want to be able to interpret results more easily
  - Want more robust inferences
  - Want to explain results more easily to others
- Properties of "simple methods"

New simple statistical methods

- Why new simple methods?
  - Want to be able to interpret results more easily
  - Want more robust inferences
  - Want to explain results more easily to others
- Properties of "simple methods"

New simple statistical methods

- Why new simple methods?
  - Want to be able to interpret results more easily
  - Want more robust inferences
  - Want to explain results more easily to others
- Properties of "simple methods"

New simple statistical methods

- Why new simple methods?
  - Want to be able to interpret results more easily
  - Want more robust inferences
  - Want to explain results more easily to others
- Properties of "simple methods"
  - Automatic
  - Do not use all available information

New simple statistical methods

- Why new simple methods?
  - Want to be able to interpret results more easily
  - Want more robust inferences
  - Want to explain results more easily to others
- Properties of "simple methods"
  - Automatic
  - Do not use all available information

New simple statistical methods

- Why new simple methods?
  - Want to be able to interpret results more easily
  - Want more robust inferences
  - Want to explain results more easily to others
- Properties of "simple methods"
  - Automatic
  - Do not use all available information

New simple statistical methods

- Why new simple methods?
  - Want to be able to interpret results more easily
  - Want more robust inferences
  - Want to explain results more easily to others
- Properties of "simple methods"
  - Automatic
  - Do not use all available information

New simple statistical methods

- Why new simple methods?
  - Want to be able to interpret results more easily
  - Want more robust inferences
  - Want to explain results more easily to others
- Properties of "simple methods"
  - Automatic
  - Do not use all available information

Typical regression output

```
> M1 <- lm (formula = partyid ~ female + black + age + I(age^2) +
    parents.party + education + income + ideology + income:ideology)
> display (M1)
                coef.est  coef.se
black            -0.98     0.17
parents.party     0.49     0.03
income           -0.43     0.15
education         0.18     0.06
ideology          0.20     0.11
income:ideology   0.15     0.03
  n = 989, k = 10
  residual sd = 1.58, R-Squared = 0.49
```

Standardized regression output

```
> display (standardize (M1))
                   coef.est coef.se
c.black            -0.98     0.17
z.parents.party     1.66     0.11
z.income            0.41     0.12
z.education         0.34     0.12
z.ideology          1.84     0.10
z.income:z.ideology 0.94     0.22
  n = 989, k = 10
  residual sd = 1.58, R-Squared = 0.49
```

Separation in logistic regression

```
glm (vote ~ female + black + income, family=binomial(link="logit"))
```

1960

	coef.est	coef.se
(Intercept)	-0.14	0.23
female	0.24	0.14
black	-1.03	0.36
income	0.03	0.06

1964

	coef.est	coef.se
(Intercept)	-1.15	0.22
female	-0.09	0.14
black	-16.83	420.40
income	0.19	0.06

1968

	coef.est	coef.se
(Intercept)	0.47	0.24
female	-0.01	0.15
black	-3.64	0.59
income	-0.03	0.07

1972

	coef.est	coef.se
(Intercept)	0.67	0.18
female	-0.25	0.12
black	-2.63	0.27
income	0.09	0.05

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
    Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`
- Performs well in cross-validation on a corpus of datasets

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`
- Performs well in cross-validation on a corpus of datasets

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`
- Performs well in cross-validation on a corpus of datasets

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`
- Performs well in cross-validation on a corpus of datasets

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`
- Performs well in cross-validation on a corpus of datasets

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`
- Performs well in cross-validation on a corpus of datasets

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`
- Performs well in cross-validation on a corpus of datasets

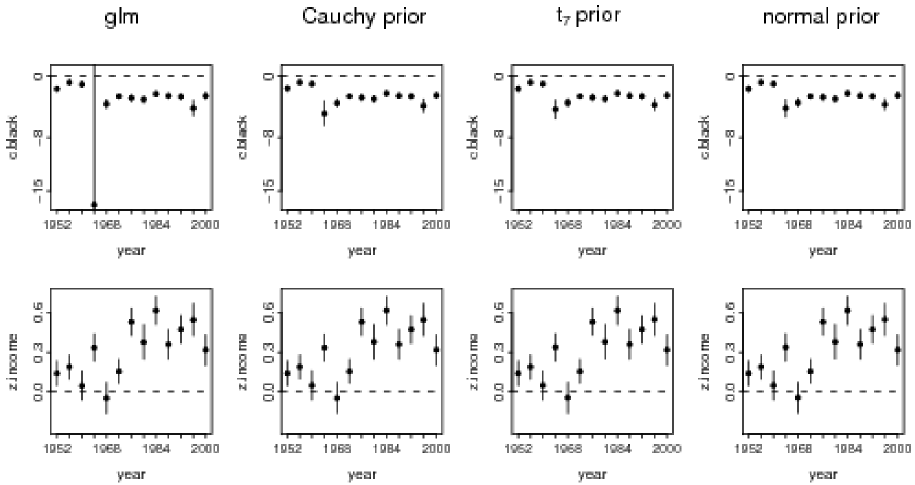
Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`
- Performs well in cross-validation on a corpus of datasets

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- Some prior info: logistic regression coefs are almost always between -5 and 5:
  - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
  - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of `glm`
- Performs well in cross-validation on a corpus of datasets

Regularization in action!



Another example of conservatism

Dose	#deaths / #animals
−0.86	0/5
−0.30	1/5
−0.05	3/5
0.73	5/5

- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is 7.8 ± 4.9
  - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

Another example of conservatism

Dose	#deaths / #animals
−0.86	0/5
−0.30	1/5
−0.05	3/5
0.73	5/5

- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is 7.8 ± 4.9
  - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

Another example of conservatism

Dose	#deaths / #animals
−0.86	0/5
−0.30	1/5
−0.05	3/5
0.73	5/5

- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is 7.8 ± 4.9
  - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

Another example of conservatism

Dose	#deaths / #animals
-0.86	$0/5$
-0.30	$1/5$
-0.05	$3/5$
0.73	$5/5$

- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is 7.8 ± 4.9
  - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

Another example of conservatism

Dose	#deaths / #animals
-0.86	$0/5$
-0.30	$1/5$
-0.05	$3/5$
0.73	$5/5$

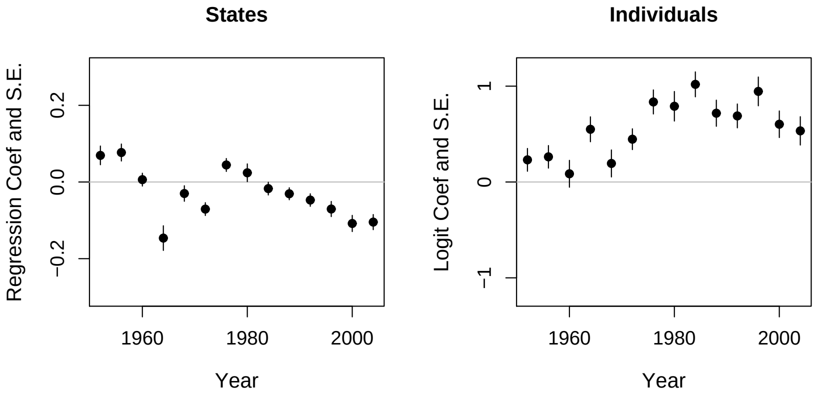
- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is 7.8 ± 4.9
  - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

Another example of conservatism

Dose	#deaths / #animals
-0.86	$0/5$
-0.30	$1/5$
-0.05	$3/5$
0.73	$5/5$

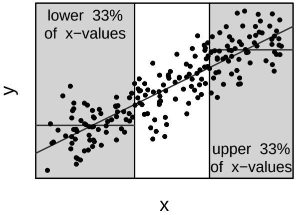
- Slope of a logistic regression of Pr(death) on dose:
  - Maximum likelihood est is 7.8 ± 4.9
  - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

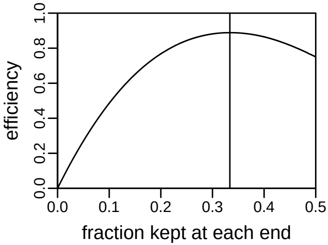
Income and voting, for states and for individuals



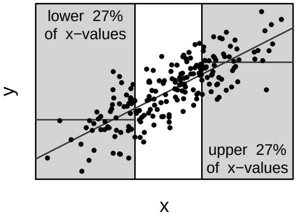
Regression coefficients or direct comparisons?

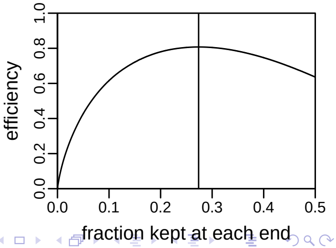
Uniform



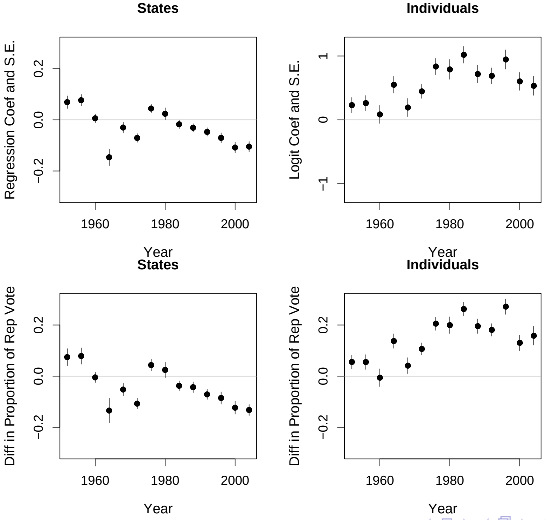


Normal





Regression coefficients or direct comparisons?



Research on simple statistical methods

- It starts with an *annoyance*, for example:
  - Interpreting a table of regression coefficients
  - Unstable logistic regression estimates
  - Explaining results to a general audience
- The role of theory

Research on simple statistical methods

- It starts with an *annoyance*, for example:
  - Interpreting a table of regression coefficients
  - Unstable logistic regression estimates
  - Explaining results to a general audience
- The role of theory

Research on simple statistical methods

- It starts with an *annoyance*, for example:
  - Interpreting a table of regression coefficients
  - Unstable logistic regression estimates
  - Explaining results to a general audience
- The role of theory

Research on simple statistical methods

- It starts with an *annoyance*, for example:
  - Interpreting a table of regression coefficients
  - Unstable logistic regression estimates
  - Explaining results to a general audience
- The role of theory
  - sd of a binary variable
  - Bayesian inference
  - Statistical efficiency

Research on simple statistical methods

- It starts with an *annoyance*, for example:
  - Interpreting a table of regression coefficients
  - Unstable logistic regression estimates
  - Explaining results to a general audience
- The role of theory
  - sd of a binary variable
  - Bayesian inference
  - Statistical efficiency

Research on simple statistical methods

- It starts with an *annoyance*, for example:
  - Interpreting a table of regression coefficients
  - Unstable logistic regression estimates
  - Explaining results to a general audience
- The role of theory
  - sd of a binary variable
  - Bayesian inference
  - Statistical efficiency

Research on simple statistical methods

- It starts with an *annoyance*, for example:
  - Interpreting a table of regression coefficients
  - Unstable logistic regression estimates
  - Explaining results to a general audience
- The role of theory
  - sd of a binary variable
  - Bayesian inference
  - Statistical efficiency

Research on simple statistical methods

- It starts with an *annoyance*, for example:
  - Interpreting a table of regression coefficients
  - Unstable logistic regression estimates
  - Explaining results to a general audience
- The role of theory
  - sd of a binary variable
  - Bayesian inference
  - Statistical efficiency

Research on simple statistical methods

- It starts with an *annoyance*, for example:
  - Interpreting a table of regression coefficients
  - Unstable logistic regression estimates
  - Explaining results to a general audience
- The role of theory
  - sd of a binary variable
  - Bayesian inference
  - Statistical efficiency

Generalizable principles

- Scaling of variables
- Weakly informative prior distributions
- Expressing estimates as comparisons
- All statistics is "consulting"

Generalizable principles

- Scaling of variables
- Weakly informative prior distributions
- Expressing estimates as comparisons
- All statistics is "consulting"

Generalizable principles

- Scaling of variables
- Weakly informative prior distributions
- Expressing estimates as comparisons
- All statistics is "consulting"

Generalizable principles

- Scaling of variables
- Weakly informative prior distributions
- Expressing estimates as comparisons
- All statistics is "consulting"

Generalizable principles

- Scaling of variables
- Weakly informative prior distributions
- Expressing estimates as comparisons
- All statistics is "consulting"