A Statistical Analysis of Speed Camera Data

Lee Fawcett and Neil Thorpe

Newcastle University

ASP seminar: Monday 17th October 2011

Structure of this talk

1. Background

- Speed cameras in the U.K.
- Regression to the mean (RTM)
- Empirical Bayes approach

2. Application

- Empirical Bayes versus Full Bayes
- Healthcare implications
- Further modelling

3. Reviewers' comments

- All help welcome!

Background

- **1996**: Government report: road safety cameras effective weapon in reducing casualty figures
- High implementation/running costs
- **1998**: Government allowed traffic authorities to recover these costs via speeding fines
- **2000** paper: Speed cameras an important part of the government's 2010 casualty reduction targets
- **April 2000**: two year pilot programme involving eight road safety camera partnerships (SCPs)
- Results at the end of 2000 prompted an earlier–than–expected national roll–out of SCPs

Northumbria SCP

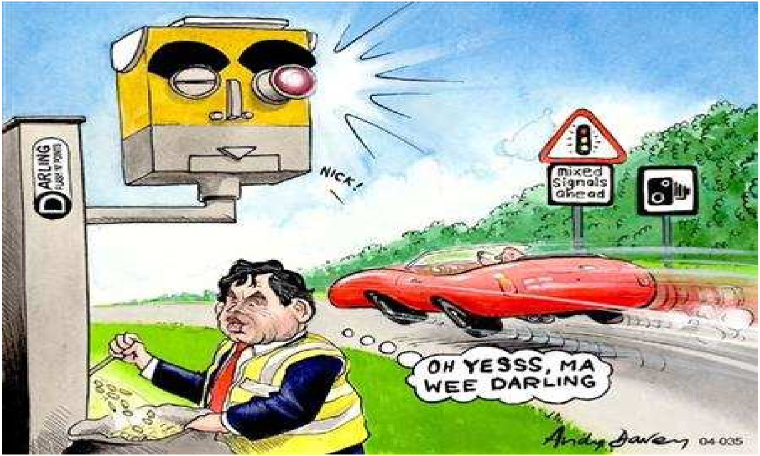
NSCP

- Joined the national programme in April 2003
- 56 mobile speed camera sites
- ‘before’ period (April 2001–March 2003) vs ‘after’ period (April 2004–March 2006)

Aims:

- To investigate changes in the number/severity of casualties
- To investigate changes in cost–of–treatment estimates

Speed cameras: Friend or foe?



Speed cameras: Friend or foe?

Speed cameras: Friend or foe?

Speed cameras: Friend or foe?

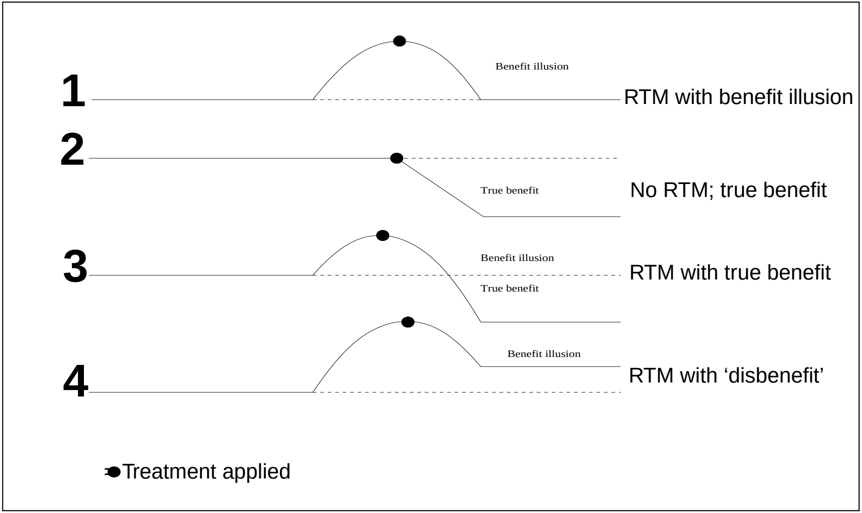
Regression To *[the]* Mean (RTM)

The 56 sites were chosen because of their unusually high casualty history (“blackspots”).

- The number of casualties is bound to decrease in any ‘after’ period, “...even if a garden gnome is used instead of a speed camera” (*Paul Smith, SafeSpeed*)

- Main consequence: ‘before’ versus ’after’ will probably exaggerate the treatment effect
- Studies have shown that a reduction owing to RTM of between 20–30% is common

Regression To *[the]* Mean (RTM)



Solution: Empirical Bayes approach

[The modelling framework I am about to describe was suggested in the early 1980s (e.g. Hauer, 1980) and has become the 'gold standard' in the road safety literature]

Let $y_{j,\text{before}}$ be the casualty frequency at site j in the before period. Then let

$$y_{j,\text{before}} | m_j \sim \text{Poisson}(m_j),$$

where m_j itself has a gamma distribution with mean μ_j and variance μ_j^2/γ.

This Poisson–Gamma specification gives a posterior for $m_j | y_{j\text{before}}$ that is also of gamma form:

$$m_j | y_{j,\text{before}} \sim \text{Gamma}(\gamma + y_{j,\text{before}}, \gamma/\mu_j + 1).$$

Solution: Empirical Bayes approach

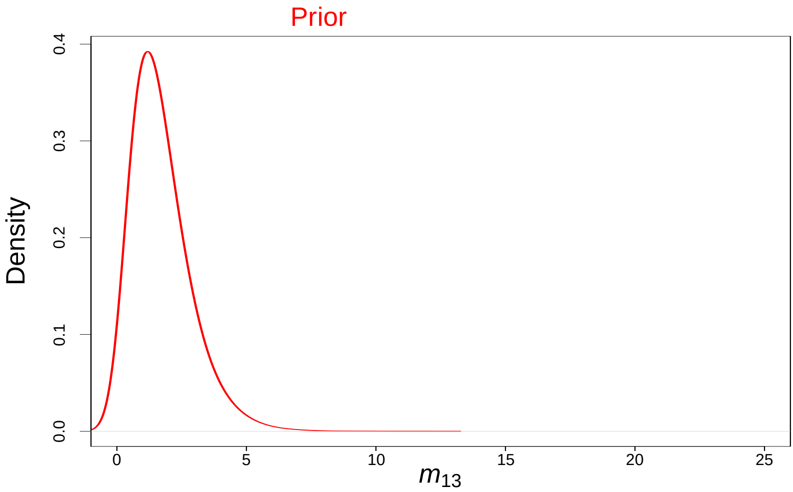
The mean of this posterior is then

$$E[m_j | y_{j,\text{before}}] = \frac{\gamma + y_{j,\text{before}}}{\gamma/\mu_j + 1}$$
$$= \alpha_j \mu_j + (1 - \alpha_j) y_{j,\text{before}}, \qquad (1)$$

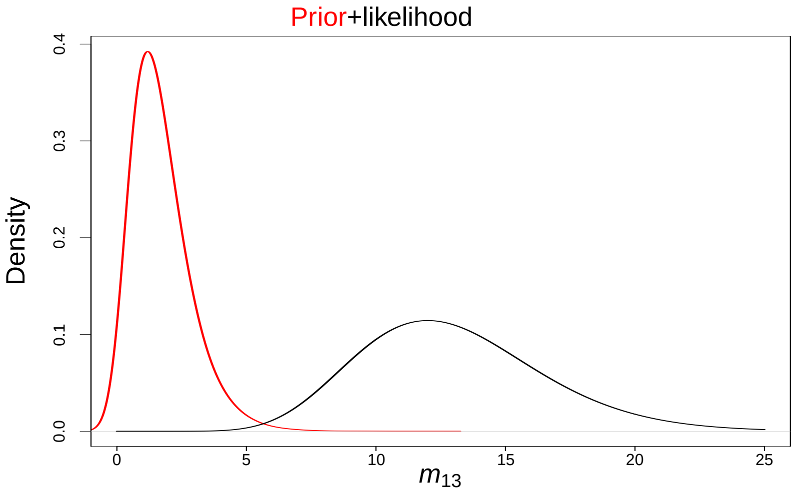
where

$$\alpha_j = \gamma/(\gamma + \mu_j), \qquad 0 \leq \alpha_j \leq 1.$$

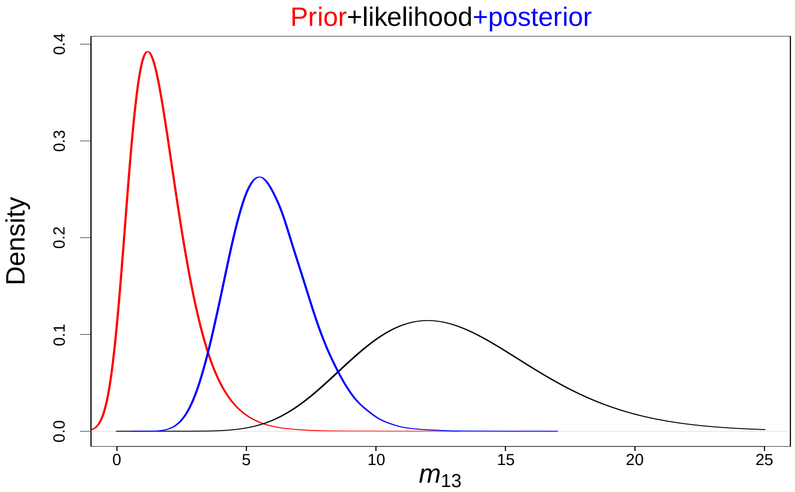
Solution: Empirical Bayes approach



Solution: Empirical Bayes approach



Solution: Empirical Bayes approach



Estimating μ_j

A generalised linear modelling approach can be used to build a predictive accident model (PAM) for the prior mean μ_j, where data from non–treatment sites are used to estimate the regression coefficients:

$$\hat{\mu}_j = \exp\left\{\hat{\beta}_0 + \sum_{p=1}^{n_p} \hat{\beta}_p x_{p_j}\right\}.$$

Problem: are treatment/non–treatment sites exchangeable?

Estimating γ

The unconditional distribution of $y_{j,\text{before}}$, found by integrating the posterior with respect to m_j, is negative binomial with

- mean μ_j;
- variance $\mu_j + \kappa\mu_j^{2*}$.

The variance can be estimated by the squared residuals from the regression model, and thus an estimate of γ obtained.

From this, we can get an estimate of the weight α_j and thus the EB estimate of casualty frequency via Equation (1).

$^*\kappa = 1/\gamma$ is the negative binomial 'over-dispersion' parameter

Application of Empirical Bayes

From 67 (non–speed camera) sites in Northumbria, we were given data relating to:

- x_1: Speed limit (mph)
- x_2: Average observed speed (mph)
- x_3: 85th percentile speed (mph)
- x_4: % of drivers over the speed limit
- x_5: % of drivers at least 15mph over the speed limit
- x_6: Daily traffic flow
- x_7: Road classification (A, B, C, U)
- x_8: Road type (single/dual/mixed)

Standard regression techniques were used to obtain the PAM.

Application of Empirical Bayes

- x_2: Average observed speed (mph)
- x_4: % of drivers over the speed limit
- x_6: Daily traffic flow
- x_7: Road classification (A, B, C, U)

This gives:

$$\hat{\mu}_j = \exp\left\{1.93 - 0.04x_{2_j} - 0.01x_{4_j} + 0.44x_{6_j} + 0.67I_{1_j} + 0.85I_{2_j} + 1.06I_{3_j}\right\}$$

EB estimates of casualty frequency

This PAM was then used to estimate μ_j, $j = 1, \ldots, 56$, for each of our speed camera sites...

... and hence we obtain the EB estimate of casualty frequency for each of these sites, giving results like:

	$y_{j,\text{before}}$	μ_j	α_j	EB	$y_{j,\text{after}}$
⋮	⋮	⋮	⋮	⋮	⋮
Site 13	12	1.71	0.59	5.95	2
⋮	⋮	⋮	⋮	⋮	⋮
Site 47	16	7.84	0.24	14.06	5
⋮	⋮	⋮	⋮	⋮	⋮
Total	**436**			**321**	**298**

Site 13: Observed change: –10; after RTM: –4
Total: Observed change: –138; after RTM: –23 $\rightarrow$ 26.4% RTM.

Fully Bayesian analysis

Initially, exactly the same model structure as the EB analysis.

However, we now unify the entire modelling procedure by assigning independent prior distributions to the regression coefficients:

$$\beta_i \sim N(0, v_{\beta_i}), \qquad i = 0, \ldots, n_p,$$

and

$$\log(\kappa) \sim N(0, v_\kappa),$$

using large v_- to represent non–informative priors.

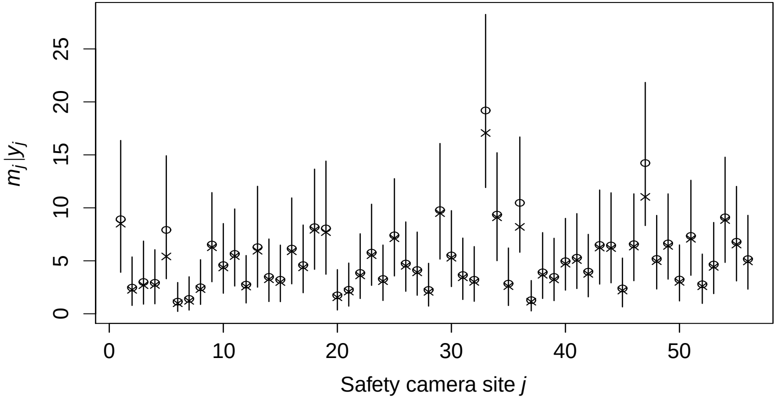
MCMC sampling scheme

- Initialise β_i and $\log(\kappa)$ at their prior means
- Use random walk Metropolis–Hastings scheme to update
- At each iteration R:
  1. use $\beta_i^{(R)}$ to estimate $\mu_j^{(R)}$ at each speed camera site j;
  2. use the current values $\mu_j^{(R)}$ and $\gamma^{(R)} = 1/\kappa^{(R)}$ as the mean and shape of the gamma prior for m_j;
  3. now use Gibbs sampling for straightforward sampling from the full conditional distribution for m_j
- Run for a million iterations (from many starting points to check convergence)

Results

	Posterior for m_j			
	Mean	St. dev.	Median	95% credible interval
⋮	⋮	⋮	⋮	⋮
Site 2	2.47	1.19	2.26	(0.78, 5.37)
	2.38	*0.936*	—	—
⋮	⋮	⋮	⋮	⋮
Site 13	6.28	2.45	5.945	(2.50, 12.04)
	5.95	*1.56*	—	—
⋮	⋮	⋮	⋮	⋮
Site 47	14.23	3.47	11.03	(8.32, 21.84)
	14.06	*3.27*	—	—
⋮	⋮	⋮	⋮	⋮
Total T	322	23.83	308	(289.92, 369.97)

Results



Implications for healthcare demand

If the speed cameras 'saved' $(T - \sum_{\forall j} y_{j,\text{after}})$ casualties, what *would* these have cost the NHS, in terms of treatment?

- Each A&E admission falls into one of **8** Health Resource Group (HRG) tarrifs:
  - 'High cost' – e.g. patients requiring CT or MRI scans, to
  - 'Low cost' – e.g. routine urine/bacteriological investigations
  - Each HRG has an associated financial tarrif
- If an A&E admission then becomes an inpatient admission:
  - there are over **700** inpatient HRGs
  - total inpatient costs are mainly a function of time
  - we break inpatient costs into financial groups of £500

Implications for healthcare demand

1. Consider each A&E HRG/Inpatient tarrif category combination τ as a multinomial outcome with associated financial tarrif £C_τ
2. probabilities p_τ are just the observed proportions falling into each τ in the ‘before’ period (found via a difficult data linkage exercise)

Implications for healthcare demand

3. Estimate of the number of casualties that *would have* fallen into each category τ obtained by multiplying p_τ by the total change in casualty frequency after RTM

4. Overall financial saving £S:

$$\mathbb{E}(S) = \left(\sum_{j=1}^{56} \mathbb{E}(m_j|y_j) - y_{j,\text{after}} \right) \sum_{\forall \tau} p_\tau C_\tau ,$$

in the EB analysis; in the FB analysis, at each iteration R we find

$$S^{(R)} = \left(T^{(R)} - \sum_{j=1}^{56} y_{j,\text{after}} \right) \sum_{\forall \tau} p_\tau C_\tau$$

Results

Thousand £		Empirical Bayes	Posterior Mean	Posterior St. dev.	Posterior Median	Posterior 95% credible interval
	Midpoint	25.6	24.9	13.2	24.4	(0.3, 57.5)
S	Minimum	23.5	22.8	12.1	22.3	(0.1, 52.5)
	Maximum	27.7	27.1	14.4	26.5	(0.6, 62.5)
$S*$		1215.6	1529.8	786.3	1479.8	(45.6, 4122.3)

Results

Message to the road safety people

- The standard (EB) approach to account for RTM is over–optimistic in its estimation of the variability in casualty frequency
- A fully Bayesian analysis gives a more complete inferential procedure...
- ... providing an easy, convenient way of summarising the posterior
- A fully Bayesian analysis also allows us to:
  - easily try out other (possibly more realistic) non–conjugate priors for m_j;
  - consider more complex model structures.

Sensitivity to other priors

We examined the sensitivity of our results to the choice of prior for m_j by considering

- $m_j \sim \text{lognormal}(\text{mean} = \lambda_j, \text{variance} = \sigma^2)$, and
- $m_j \sim \text{Weibull}(\text{shape} = \omega, \text{scale} = \nu_j)$,

choosing (λ_j, σ^2) and (ω, ν_j) so as to allow relative comparisons with the original gamma prior.

Results

	EB	Gamma Mean	Gamma Median	Lognormal Mean	Lognormal Median	Weibull Mean	Weibull Median
		(95% CI)		(95% CI)		(95% CI)	
T	321	322	308	355	338	317	303
		(290, 370)		(309, 394)		(296, 371)	
RTM (%)	–26.4	–26.5	–29.7	–18.9	–22.8	–27.6	–30.8
		(−35.6, −14.2)		(−26.3, −9.0)		(−39.3, −15.3)	
S (thousand £)	25.6	24.9	24.4	29.3	29.3	25.3	24.9
		(0.3, 57.5)		(6.1, 73.5)		(0.7, 70.9)	
S^* (thousand £)	1215.6	1529.8	1479.8	2803.0	2801.0	986.3	951.3
		(45.6, 4122.3)		(581.4, 5126.5)		(69.1, 4910.9)	

- Some agreement between Gamma and Weibull priors
- Lognormal prior: less reduction due to RTM $\rightarrow$ greater treatment effect $\rightarrow$ greater financial savings due to the cameras
- DIC suggests Weibull most appropriate
- Need for more careful prior elicitation!

Choice of prior for the regression coefficients

Independent Normal priors, i.e.

$$\beta_i \sim N(0, v_{\beta_i}), \qquad i = 0, \ldots, n_p,$$

probably not the best! How can we improve on this? Difficult.

1. Data augmentation prior? Use

$$\boldsymbol{\beta}_{\backslash 0, n_p} \sim N_{n_p}\left(\mathbf{0}, n(\mathbf{X}_{n_p}^T \mathbf{X}_{n_p})^{-1}\right),$$

and a vague prior for β_0 (as before).

Choice of prior for the regression coefficients

2. Conditional mean prior?

- Elicit a prior on $\tilde{\mathbf{M}} = (\tilde{M}_1, \ldots, \tilde{M}_{n_p})$, where the $\tilde{M}_p$'s are mean responses at covariates $\mathbf{x}_p$, $p = 1, \ldots, n_p$;
- Denote by $\tilde{\mathbf{X}}$ the matrix with $\mathbf{x}_p^T$ in the ith row;
- Following the notation of Bedrick *et al.* (1996), $\mathbf{G}$ and $\mathbf{G}^{-1}$ are vector transformations that apply g and g^{-1} to each element – e.g. $g(\cdot) = \log(\cdot)$ or $g(\cdot) = \mathrm{logit}(\cdot)$;
- Assessing the $\tilde{M}_p$'s independently, the conditional mean prior is

$$\pi_0(\tilde{\mathbf{M}}) = \prod_{p=1}^{n_p} \pi_{0_p}(\tilde{M}_p).$$

- Writing

$$\tilde{\mathbf{M}} = \mathbf{G}^{-1}(\tilde{\mathbf{X}}\beta_{\setminus 0,n_p}) \qquad \text{and} \qquad \beta_{\setminus 0,n_p} = \tilde{\mathbf{X}}^{-1}\mathbf{G}(\tilde{\mathbf{M}})$$

induces a prior on β of the form

$$\pi(\beta_{\setminus 0,n_p}) = \prod_{p=1}^{n_p} \pi_{0_p} g^{-1}(\tilde{\mathbf{x}}_p^T \beta_{\setminus 0,n_p}) \Big/ |\tilde{\mathbf{X}}^{-1}| \prod_{p=1}^{n_p} \dot{g}(\tilde{M}_p).$$

Choice of prior for the regression coefficients

To implement the conditional mean prior, we need means a_i and variances b_p for each mean response $\tilde{M}_p$ at covariate $\mathbf{x}_p$, $p = 1, \ldots, n_p$.

A regression analysis from a previous study of casualty frequencies at another group of sites in the Northumbria region gives a regression equation of the form

$$\mu = \exp\left\{\beta_0 + \sum_{p=1}^{n_p} \beta_p x_{p_j}\right\}.$$

Covariates at n_p of these sites can then be used to suggest means a_p and variances b_p, and suitable priors for $\tilde{M}_p$ proposed around these.

Choice of prior for the regression coefficients

	a_i	b_i	Gamma Shape	Gamma Scale	Lognormal Mean	Lognormal Variance	Weibull Shape	Weibull Scale
$\tilde{M}_1$	9.93	11.63	8.48	0.85	2.24	0.11	3.20	11.09
$\tilde{M}_2$	2.77	1.64	4.68	1.69	0.92	0.19	2.29	3.13
$\tilde{M}_3$	3.59	3.10	4.16	1.16	1.17	0.22	2.15	4.05
$\tilde{M}_4$	3.12	1.87	5.21	1.67	1.05	0.18	2.43	3.52
$\tilde{M}_5$	8.19	6.25	10.73	1.31	2.06	0.09	3.64	9.08
$\tilde{M}_6$	5.42	3.79	7.75	1.43	1.63	0.12	3.04	6.07

Effect of using more informed priors: Greater posterior precision.

Trend?

Casualty figures for the Northumbria region reveal that:

- Since the mid–1970s, overall casualty figures have fallen by around 2% per year;
- Since 2005, the number of (reported) 'slight' casualties has increased by about 0.5% per year.

Thus, we now specify the following modified form for μ_j:

$$\mu_j = \xi \exp\left\{\beta_0 + \sum_{p=1}^{n_p} \beta_p x_{p_j}\right\},$$

where ξ is a trend effect constant across all sites j.

Since the difference between the mid–points of the before and after periods is 3 years ($2002 \to 2005$), we use:

$$\xi \sim U(0.94, 1.015).$$

Results

		Mean	St. dev.	Posterior Median	95% credible interval
T	*without trend*	327	25.528	313	(285, 354)
	with trend	323	25.709	309	(273, 340)
S	*without trend*	30.9	14.092	30.9	(0.9, 58.5)
	with trend	28.7	14.192	28.8	(0.9, 56.6)
$S*$	*without trend*	1541.5	349.136	1541.7	(72.6, 4183.2)
	with trend	1318.5	358.725	1324.5	(68.5, 3979.6)

Reviewers' comments

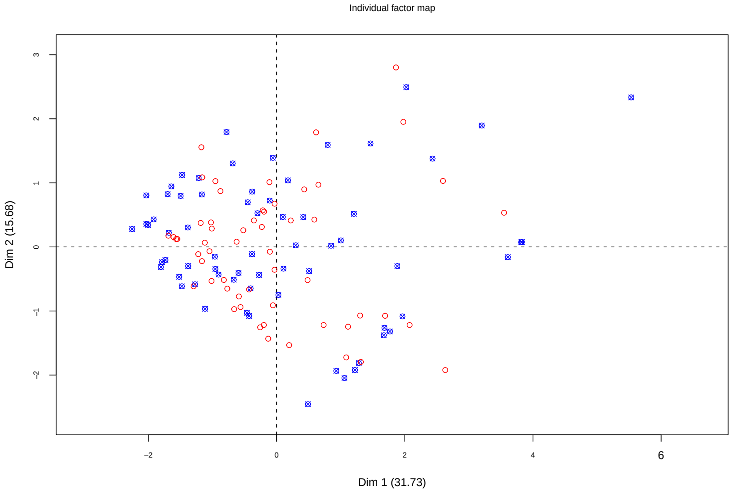
"Technique used only valid if reference sites and treatment sites can be considered exchangeable...

- Principal Components Analysis/Multiple Factor Analysis on

		x_1	x_2	$\cdots$	x_8
	Site 1				
Reference	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	Site 67				
	Site 1				
Treated	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	Site 56				

Plot scores for first PC against those for second PC, using different plotting character for "reference" and "treated".

Reviewers' comments



Reviewers' comments

- Permutation tests – e.g., we can compare individual variables between treatment and reference sites using:

$$\delta_p = |\bar{x}_p^{\mathrm{TRT}} - \bar{x}_p^{\mathrm{REF}}|, \qquad p = 1, \ldots, 8.$$

If the treatment and reference sites *are* exchangeable with respect to the explanatory variables, then the values of δ_p would not be significantly different to those obtained after a random allocation of sites to each group.

Reviewers' comments

1. Randomly choose N permutations of "reference" and "treated" allocations
2. For each permutation Π_k, $k = 1, \ldots, N$, find $\delta_p^{(\Pi_k)}$
3. Compare δ_p for the "real" allocation to the permutation distribution for δ_p
4. A p–value for the hypothesis H_0 : sites are exchangeable can be estimated as the proportion of permutations for which $\delta_p^{(\Pi_k)} \geq \delta_p$

Result: H_0 retained for all covariates.

Reviewers' comments

Can also perform a permutation test on the mean Mahalanobis distance of each site in the treatment set to sites in the reference set:

$$\bar{D} = \frac{1}{56}\sum_{j=1}^{56}\sqrt{(\mathbf{X}_j^{\mathrm{TRT}} - \bar{\mathbf{M}}^{\mathrm{REF}})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X}_j^{\mathrm{TRT}} - \bar{\mathbf{M}}^{\mathrm{REF}})},$$

where $\bar{\mathbf{M}}^{\mathrm{REF}} = (\bar{x}_1^{\mathrm{REF}}, \ldots, \bar{x}_1^{\mathrm{REF}})$ and the covariance matrix $\boldsymbol{\Sigma}$ has (s,t)–th entry given by $\mathrm{cov}(x_s^{\mathrm{REF}}, x_t^{\mathrm{REF}})$, $s, t = 1, \ldots, 8$.

Result: H_0 retained!

Reviewers' comments

"The authors account for RTM without explaining the terms of the controversy and the reasons why they are favourable to accept RTM...

- Not sure about this one...
- Perhaps look at historical casualty figures for the speed camera sites to check for "blips"?

"... little in the way of methodological novelties..."

- Not necessary for Series A?

References

Bedrick, E.J., Christensen, R. and Johnson, W. (1996). A new perspective on priors for generalized linear models. *J. Am. Statist. Ass.*, **51**, pp. 211—218.

Department for Transport (2009). Benefits to Society Arising from Prevention of Road Accidents and Casualties. Department for Transport, London.

Fawcett, L. and Thorpe, N. (2011). Mobile safety cameras: Estimating casualty reductions and the demand for secondary healthcare. *J. Roy. Statist. Soc., Series A: Statistics in Society*. Submitted.

Hauer, E. (1980). Bias–by–selection: overestimation of the effectiveness of safety countermeasures caused by the process of selection for treatment.*Acci. Anal. & Prev.*, **12**, 2, pp. 113—117.