The next weeks

March 16	§10.7 Semiparametric models
March 23	Generalized additive models and lasso
March 30	Finishing pieces, + review

Homework 3: due April 2, 5 pm

(updated March 20) Final Test: April 17, 1 - 3 pm

STA 2201S Assignment 3. *due Monday, April 2 before 5 pm*

When answering questions requiring numerical work, the results are to be reported in a narrative summary, in your own words. Tables and Figures may be included, but must be formatted along with the text. **Do not include in this summary printouts of computer code.** Analysis of variance/deviance tables, tables of coefficients and their estimated standard errors, and other output should be formatted separately and reported only to the relevant number of significant digits. All computer code used to obtain the results summarized in the response should be provided as an appendix.

1. (Faraway *Extending the Linear Model with* `R`, Ch. 11): The dataset `teengamb` in the package `faraway` gives data on annual gambling expenditure per year (in pounds) (gamble), with several covariates: sex (0 = M, 1 = F), status (a score reflecting socio-economic status), income (pounds per week), verbal (a score from 0 -12 on a test of verbal ability). Of interest is which covariates are associated with gambling expenditure.

   (a) Using an appropriate parametric model, investigate the relationship between gambling and other factors, and summarize your conclusions in non-technical language, accompanied by no more than 3 tables and 3 figures.

   (b) Investigate the use of non-parametric smoothing techniques on the data; do any insights emerge from this approach that were missed in the analysis in part (a)? Summarize your results for this part of the question by describing which methods you used, what information they provided, and whether or not they altered the conclusions from part (a). Your text should not be more than two pages, and you may include up to four figures.

2. (a) Show that if y_{ij} are independently distributed as as Poisson distribution with means μ_{ij}, $i = 1, \ldots, I; j = 1 \ldots, J$, that y_{ij} given y_{i+} are distributed as multinomial, with sample size y_{i+} and probability vector $\pi_{ij} = \mu_{ij}/\mu_{i+}$.

   (b) If $\log \mu_{ij} = \mu + \alpha_i + \beta_j$, where $\alpha_1 = 0$ and $\beta_1 = 0$, show that the residual deviance from this model is the same as the log-likelihood ratio statistic for testing independence in a multinomial model. Your task is to verify it algebraically; it has been verified numerically for HW2Q4 by Wei Lin, who showed that the observed and (fitted) values for the 2×2 table of breathlessness and wheeze, ignoring age, are as follows, whether computed using the multinomial model or the Poisson glm.

the Poisson product

an $I \times J$ table

	Wheeze	
Breathlessness	N	Y
N	14022 (12680.9)	1833 (3174.1)
Y	600 (1941.1)	1827 (485.0)

Kernel smoothing

- regression smoothing $y_j = g(x_j) + \epsilon_j$

K

$= w_j(x_0)$

$\frac{1}{h} w\left(\frac{x_j - x_0}{h}\right)$

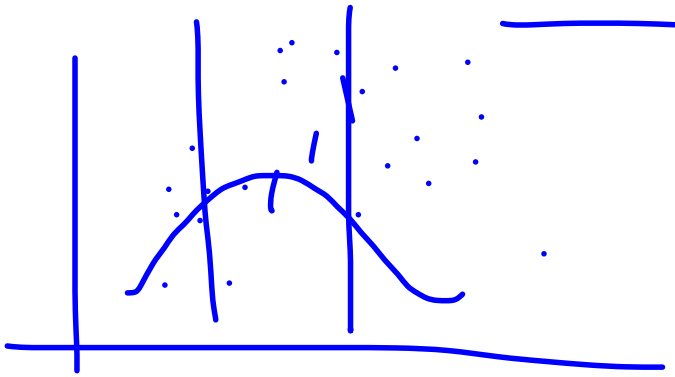
- $\hat{g}(x_0) = \sum_{j=1}^{n} S(x_0; x_j, h) y_j$
  $S(x_0; x_j, h)$ from $(X^T W X)^{-1} X^T W$

density

- local likelihood: $y_j \sim f(\cdot; \beta, x_j)$

f glm $\rightarrow$ mle $\equiv$ IRWLS

- $$\max_\beta \sum \log f(y_j; \beta, x_j) \longrightarrow \max_\beta \sum \frac{1}{h} w\left(\frac{x_j - x_0}{h}\right) \log f(y_j; \beta, x_j)$$



Example 10.32

528 10 · Nonlinear Regression Models

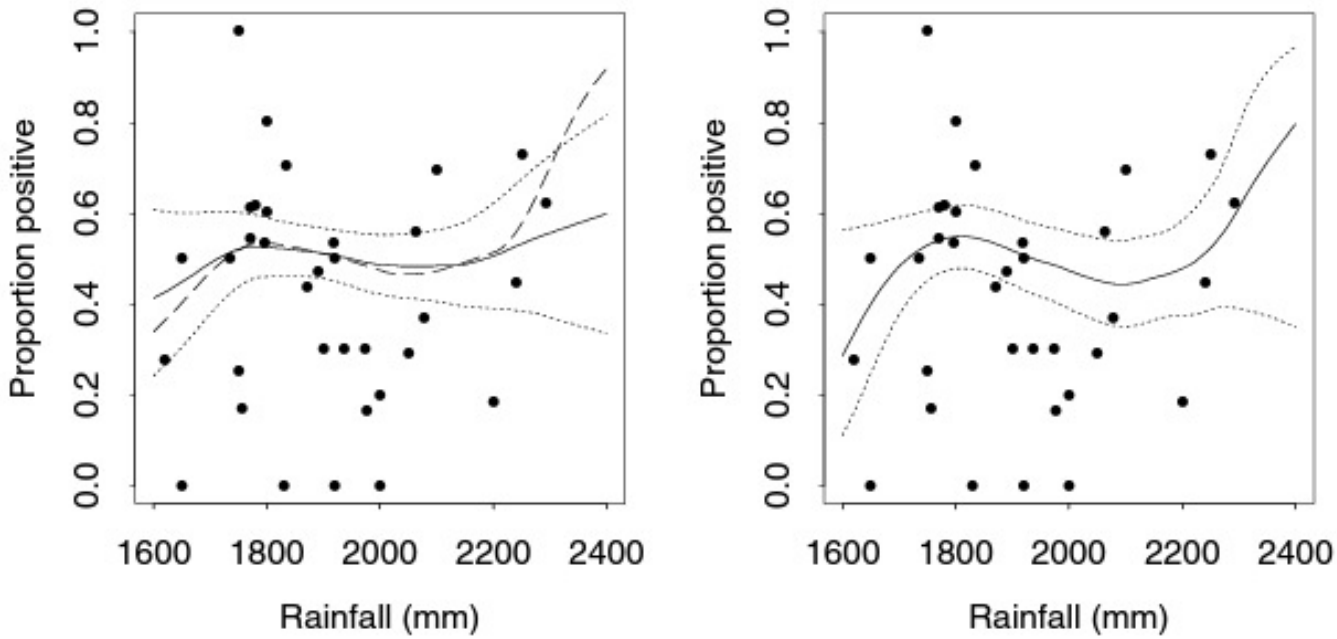


Figure 10.17 Local fits to the toxoplasmosis data. The left panel shows fitted probabilities $\hat{\pi}(x)$, with the fit of local linear logistic model with $h = 400$ (solid) and 0.95 pointwise confidence bands (dots). Also shown is the local linear fit with $h = 300$ (dashes). The right panel shows the local quadratic fit with $h = 400$ and its 0.95 confidence band. Note the increased variability due to the quadratic fit, and its stronger curvature at the boundaries.

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

$$= g(x) \leftarrow \text{not specified}$$

Flexible modelling using basis expansions

(§10.7.2)

HTF Elem St Learn Ch 5

- $y_j = g(x_j) + \epsilon_j$
- Flexible linear modelling

$$g(x) = \Sigma_{m=1}^{M} \beta_m h_m(x)$$

- This is called a linear basis expansion, and h_m is the mth basis function
- For example if X is one-dimensional:
  $g(x) = \beta_0 + \beta_1 x + \beta_2 x^2$, or
  $g(x) = \beta_0 + \beta_1 \sin(x) + \beta_2 \cos(x)$, etc.
- Simple linear regression has $h_1(x) = 1$, $h_2(x) = x$

Piecewise polynomials

- piecewise constant basis functions
  $h_1(x) = I(x < \xi_1), \quad h_2(x) = I(\xi_1 \leq x < \xi_2),$
  $h_3(x) = I(\xi_2 \leq x)$
- equivalent to fitting by local averaging

- piecewise linear basis functions , with constraints
  $h_1(x) = 1, \quad h_2(x) = x$
  $h_3(x) = (x - \xi_1)_+, \quad h_4(x) = (x - \xi_2)_+$
- windows defined by knots $\xi_1, \xi_2, \ldots$

- piecewise cubic basis functions
  $h_1(x) = 1, h_2(x) = x, h_3(x) = x^2, h_4(x) = x^3$

- continuity $h_5(x) = (x - \xi_1)_+^3, \quad h_6(x) = (x - \xi_2)_+^3$

- continuous function, continuous first and second derivatives

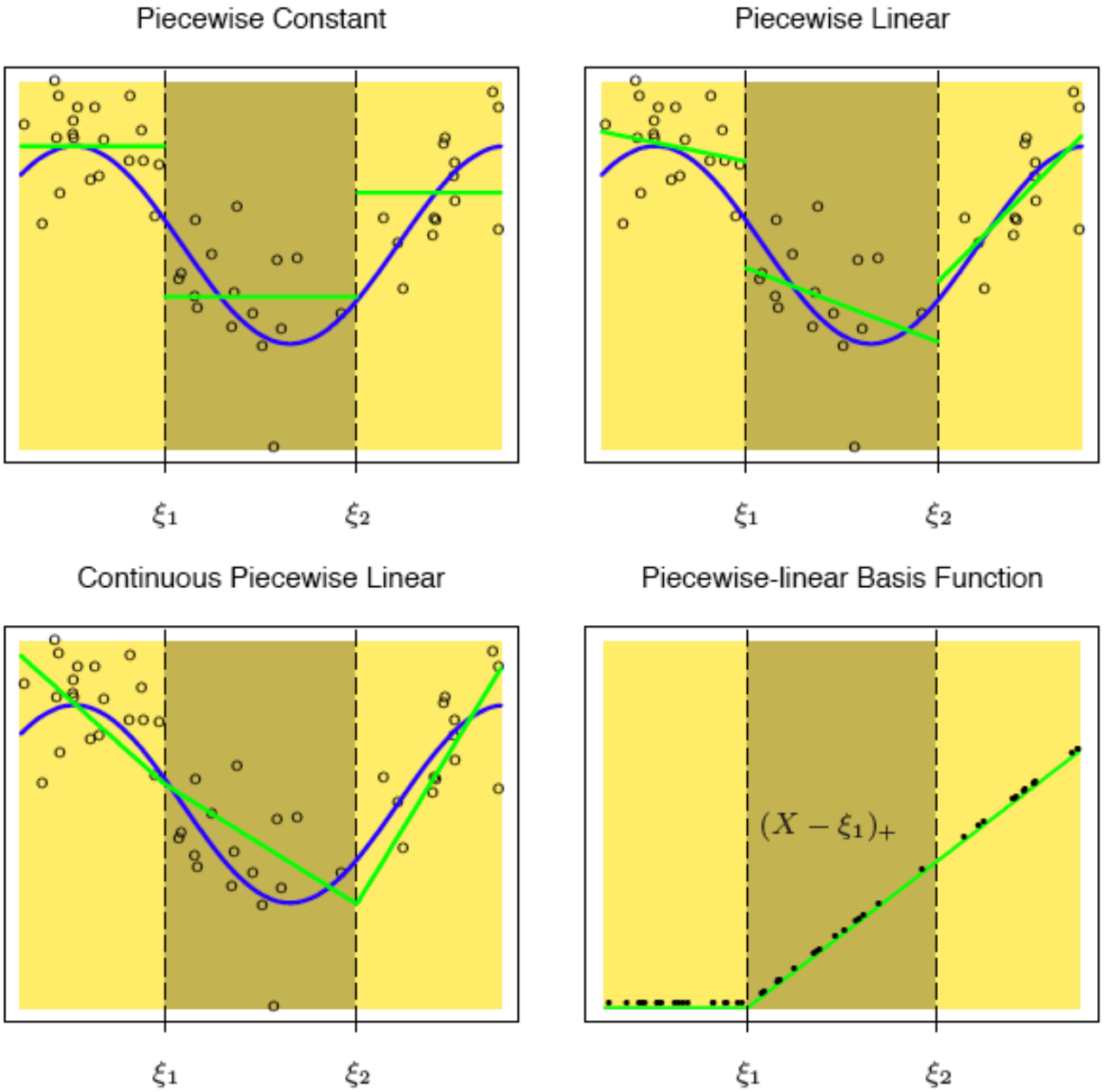


FIGURE 5.1. *The top left panel shows a piecewise constant function fit to some artificial data. The broken vertical lines indicate the positions of the two* knots ξ_1 *and* ξ_2*. The blue curve represents the true function, from which the data were*

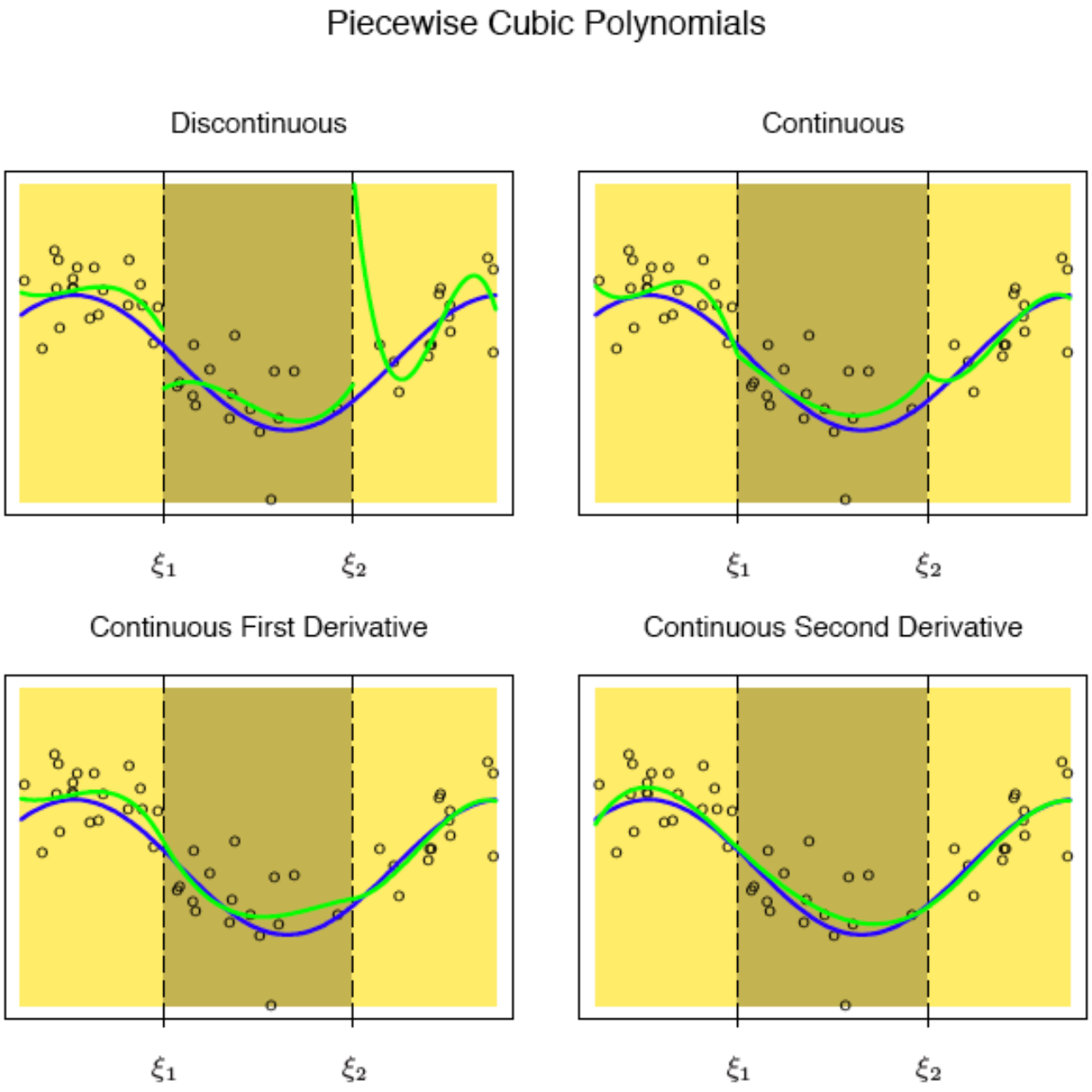
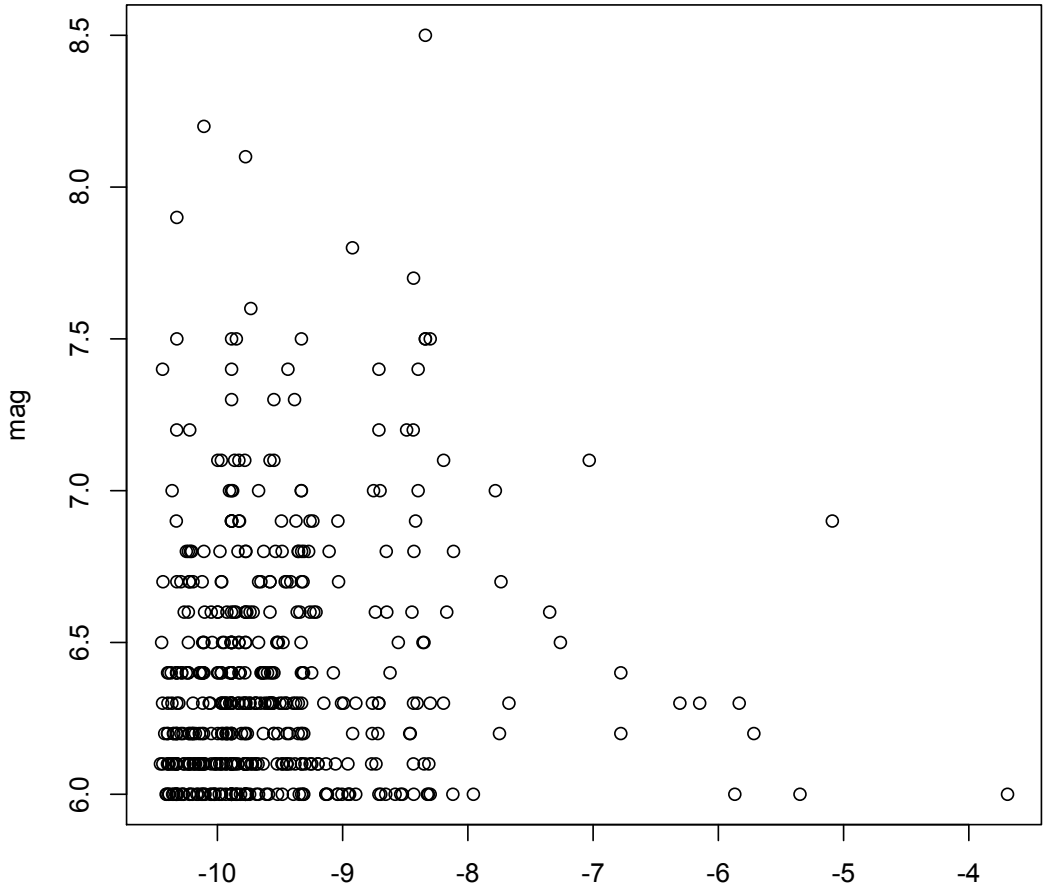


FIGURE 5.2. *A series of piecewise-cubic polynomials, with increasing orders of continuity.*

Example: earthquake data

Ex 10.31

```
> data(quake,package="SMPracticals")
> quake
          time mag
1     40.08333 6.0
2    162.38889 6.9
3    210.22917 6.0
4    303.85417 6.2
> with(quake, plot(log(1/time),mag))
```



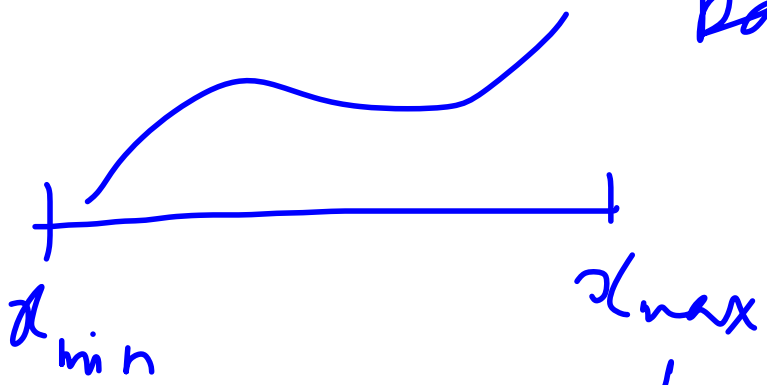
frequency of rep.'d

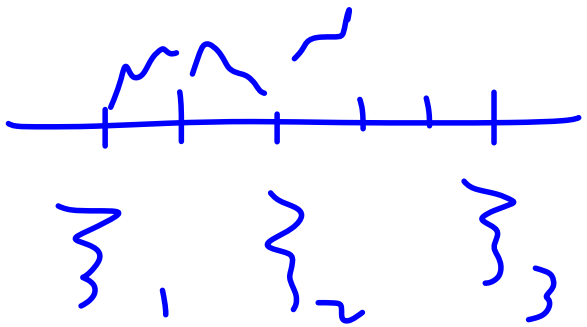
← intensity

Cubic splines $n = 483$

- truncated power basis of degree 3 $(X-\xi_1)_+^3$
- need to choose number of knots K and placement of knots $\xi_1, \ldots \xi_K$ design $h_j(X)$ terms
- construct features matrix using truncated power basis set
- use constructed matrix as set of predictors B-spline basis

```
> with(quake, bs(log(1/time))[1:10,])
#bs(x) with no other arguments just gives a single cubic polynomial
              1         2         3
 [1,] 0.0000000 0.0000000 1.0000000
 [2,] 0.1018013 0.3903714 0.4989780
 [3,] 0.1359705 0.4189773 0.4303434
 [4,] 0.1884790 0.4408886 0.3437743
 [5,] 0.2056632 0.4436068 0.3189471
 [6,] 0.2108533 0.4440520 0.3117209
 [7,] 0.2522139 0.4418128 0.2579802
 [8,] 0.2752334 0.4363260 0.2305684
 [9,] 0.3398063 0.4045238 0.1605223
[10,] 0.3398083 0.4045224 0.1605203
...
attr(,"degree")
[1] 3
attr(,"knots")
numeric(0)
attr(,"Boundary.knots")
[1] -10.454784  -3.690961
attr(,"intercept")
[1] FALSE
attr(,"class")
[1] "bs"      "basis"   "matrix"
```





... cubic splines

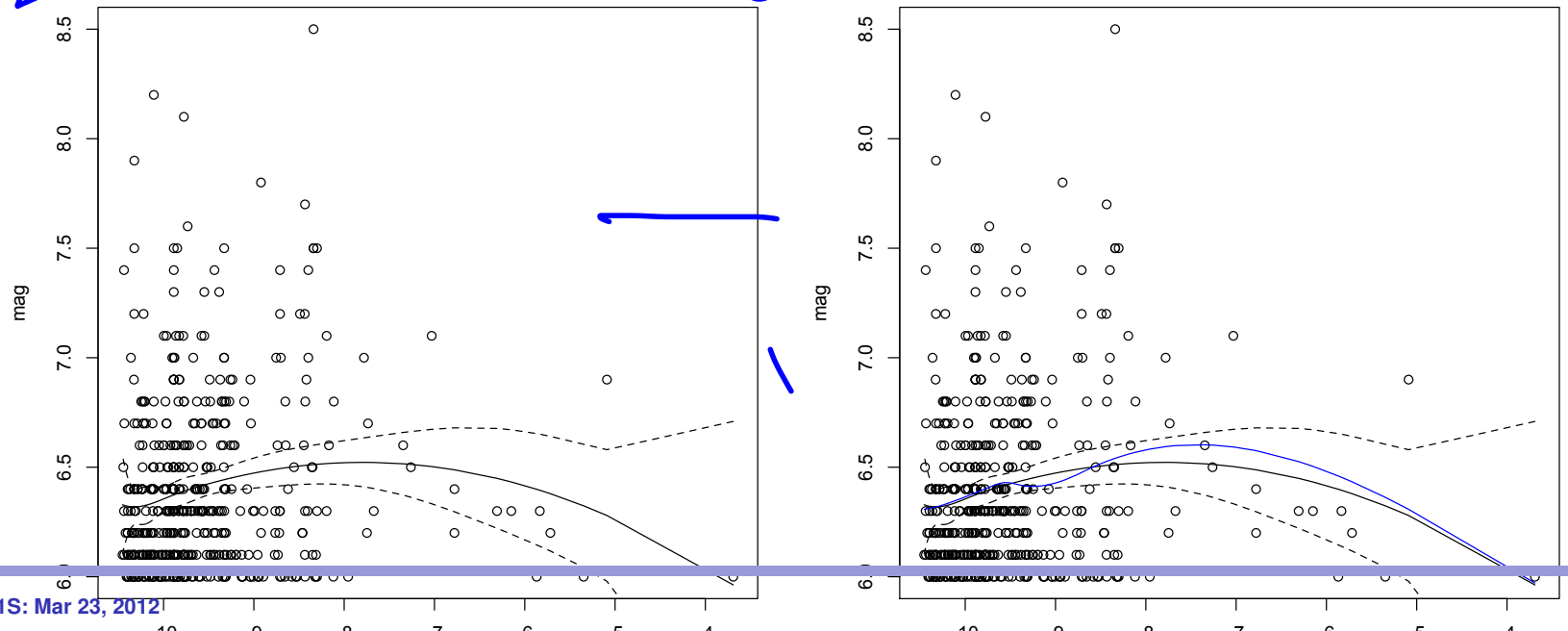
```
 > with(quake,bs(log(1/time), df=5)[1:10,])
# gives a proper cubic spline basis, here with 5 df
      1          2         3         4         5
 [1,] 0 0.00000000 0.0000000 0.0000000 1.0000000
 [2,] 0 0.01110655 0.1250814 0.4247847 0.4390274
 [3,] 0 0.01846075 0.1661869 0.4486889 0.3666635
 [4,] 0 0.03370916 0.2283997 0.4600092 0.2778819
 [5,] 0 0.03989014 0.2484715 0.4585984 0.2530400
 [6,] 0 0.04188686 0.2545024 0.4577416 0.2458691
 [7,] 0 0.06023519 0.3019733 0.4443033 0.1934881
 [8,] 0 0.07263434 0.3278645 0.4319962 0.1675050
 [9,] 0 0.11941791 0.3975881 0.3789378 0.1040562
[10,] 0 0.11941975 0.3975902 0.3789357 0.1040544
...
attr(,"degree")
[1] 3
attr(,"knots")
33.33333% 66.66667%
-9.943294 -9.520987
attr(,"Boundary.knots")
[1] -10.454784  -3.690961
```

$\beta_m h_m(x) + \beta_0$

... earthquake data

```
> quake.bs = lm(mag ~ bs(log(1/time),df=5),data = quake)
> quake.pred = predict(quake.bs, se.fit = TRUE, interval = "confidence")
> quake.pred
$fit
         fit      lwr      upr
1   5.962665 5.216283 6.709047
2   6.279641 5.979190 6.580092
3   6.323859 6.042772 6.604946
> lines(log(1/quake$time),quake.pred[[1]][,1])
> lines(log(1/quake$time),quake.pred[[1]][,2], lty=2)
> lines(log(1/quake$time),quake.pred[[1]][,3], lty=2)
> quake.lo = loess(mag ~ log(1/time), data = quake)
> quake.lopred = predict(quake.lo, se=T)
```

quake.ns = lm(mag ~ ns(log(1/time), 5)

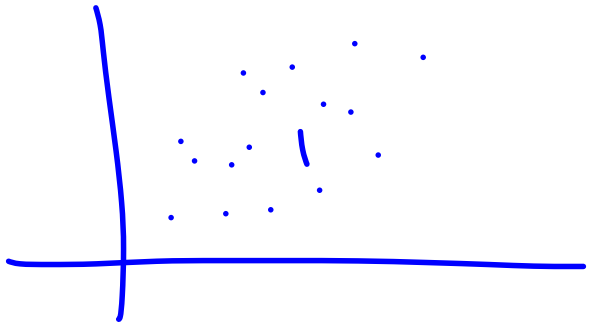


B-splines and N-splines

- The *B*-spline basis equivalent to the truncated power basis
- In R `library(splines)`:
  `bs(x, df=NULL, knots=NULL, degree=3, intercept=FALSE, Boundary.knots=range(x))`
- Must specify either `df` or `knots`. For the *B*-spline basis, `# knots = df - degree` and degree is usually 3
- Natural cubic splines are linear at the end of the range
- `ns(x, df=NULL, knots=NULL, degree=3, intercept=FALSE, Boundary.knots=range(x))`
- For natural cubic splines, `# knots = df - 1`

$$y_i = g(x_i) + \varepsilon_i$$

$$\beta_0 + g(x_i) + \varepsilon_i$$



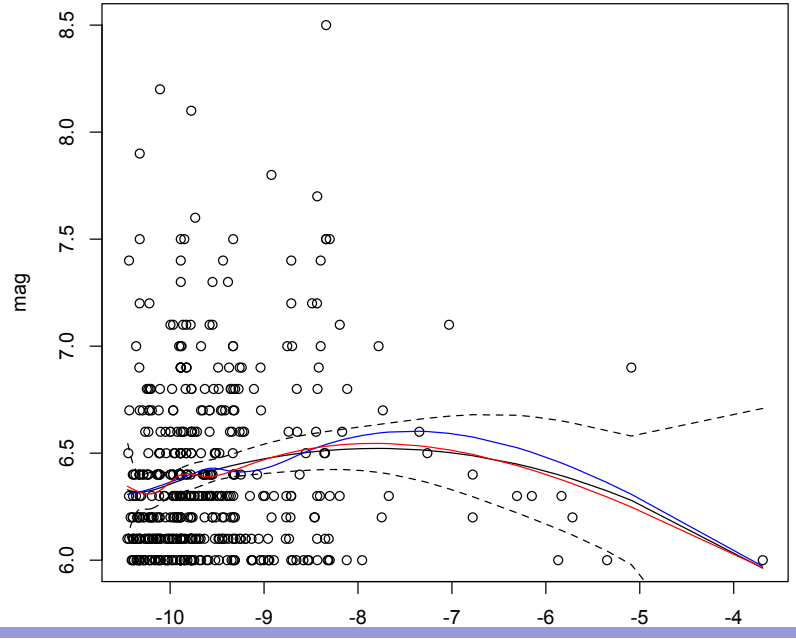
... regression splines

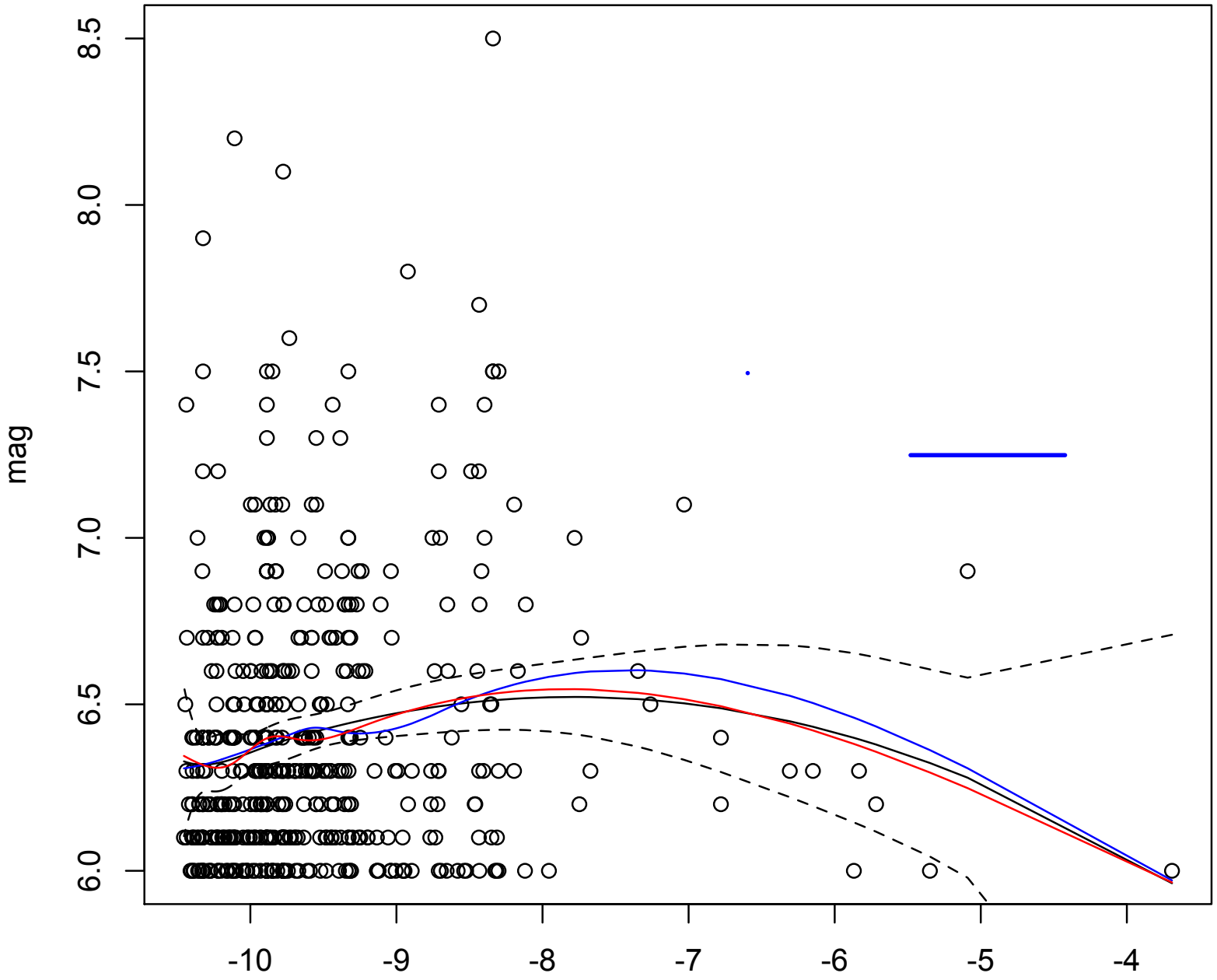
The individual coefficients don't mean anything, we need to evaluate groups of coefficients. For example

step( )

works too

```
> library(MASS)
> stepAIC(quake.ns)
Start:  AIC=-876.54
mag ~ ns(log(1/time), df = 5)

                          Df Sum of Sq    RSS     AIC
<none>                                 76.737 -876.54
- ns(log(1/time), df = 5)  5    2.1534 78.890 -873.18
```




mag
8.5
8.0
7.5
7.0
6.5
6.0
-10
-9
-8
-7
-6
-5
-4

... regression splines

- easily extended to multiple regression, and generalized linear models
- example: `data(heart, package = "ElemStatLearn")`

```
> heart[1:5,]
  row.names sbp tobacco  ldl adiposity famhist typea obesity
1         1 160   12.00 5.73     23.11 Present    49   25.30
2         2 144    0.01 4.41     28.61  Absent    55   28.87
3         3 118    0.08 3.48     32.28 Present    52   29.14
4         4 170    7.50 6.41     38.03 Present    51   31.99
5         5 134   13.60 3.50     27.78 Present    60   25.99
  alcohol age chd
1   97.20  52   1
2    2.06  63   1
3    3.81  46   0
4   24.26  58   1
5   57.34  49   1
```

... heart data

```
> heart.ns = glm (chd ~ ns(sbp,4)+ ns(tobacco,4) + ns(ldl,4) + famhist + ns(obesity, 4) +
+  ns(age,4), family=binomial)
> summary(heart.ns)

Call:
glm(formula = chd ~ ns(sbp, 4) + ns(tobacco, 4) + ns(ldl, 4) +
    famhist + ns(obesity, 4) + ns(age, 4), family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7216  -0.8322  -0.3777   0.8870   2.9694

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.265534   2.367227  -0.957 0.338547
ns(sbp, 4)1     -1.474172   0.843870  -1.747 0.080652 .
ns(sbp, 4)2     -1.351182   0.759548  -1.779 0.075251 .
ns(sbp, 4)3     -3.729348   2.021064  -1.845 0.065003 .
ns(sbp, 4)4      1.381701   0.995268   1.388 0.165055
ns(tobacco, 4)1  0.654109   0.453248   1.443 0.148975
ns(tobacco, 4)2  0.392582   0.892628   0.440 0.660079
ns(tobacco, 4)3  3.335170   1.179656   2.827 0.004695 **
ns(tobacco, 4)4  3.845611   2.386584   1.611 0.107104
ns(ldl, 4)1      1.921215   1.311052   1.465 0.142812
ns(ldl, 4)2      1.783272   1.014883   1.757 0.078897 .
ns(ldl, 4)3      4.623680   2.972938   1.555 0.119885
ns(ldl, 4)4      3.354692   1.447217   2.318 0.020448 *
famhistPresent   1.078507   0.237685   4.538 5.69e-06 ***
ns(obesity, 4)1 -3.089393   1.707207  -1.810 0.070355 .
ns(obesity, 4)2 -2.385045   1.200450  -1.987 0.046945 *
ns(obesity, 4)3 -4.998882   3.796264  -1.317 0.187909
ns(obesity, 4)4  0.009109   1.751127   0.005 0.995850
ns(age, 4)1      2.628298   1.116674   2.354 0.018588 *
```

```
> update(heart.ns, . ~ . - ns(sbp,4))

Call:  glm(formula = chd ~ ns(tobacco, 4) + ns(ldl, 4) + famhist + ns(obesity,      4) + ns(a

Coefficients:
    (Intercept)  ns(tobacco, 4)1  ns(tobacco, 4)2  ns(tobacco, 4)3
       -3.91758          0.61696          0.46188          3.51363
ns(tobacco, 4)4      ns(ldl, 4)1      ns(ldl, 4)2      ns(ldl, 4)3
        3.82464          1.70945          1.70659          4.19515
    ns(ldl, 4)4   famhistPresent  ns(obesity, 4)1  ns(obesity, 4)2
        2.90793          0.99053         -2.93143         -2.32793
ns(obesity, 4)3  ns(obesity, 4)4      ns(age, 4)1      ns(age, 4)2
       -4.87074         -0.01103          2.52772          3.12963
    ns(age, 4)3      ns(age, 4)4
        7.34899          1.53433

Degrees of Freedom: 461 Total (i.e. Null);  444 Residual
Null Deviance:     596.1
Residual Deviance: 467.2  AIC: 503.2
> 467.2 - 458.1
[1] 9.1
> pchisq(9.1,df=4)
[1] 0.941352
> 1-.Last.value
[1] 0.05864798  # compare Table 5.1
```

The function `step` does all this for you:

```
> step(heart.ns)
Start:  AIC=502.09
chd ~ ns(sbp, 4) + ns(tobacco, 4) + ns(ldl, 4) + famhist + ns(obesity,
    4) + ns(age, 4)

                 Df Deviance    AIC
<none>                458.09 502.09
- ns(obesity, 4)  4   466.24 502.24
- ns(sbp, 4)      4   467.16 503.16
- ns(tobacco, 4)  4   470.48 506.48
- ns(ldl, 4)      4   472.39 508.39
- ns(age, 4)      4   481.86 517.86
- famhist         1   479.44 521.44
 ...
 > anova(heart.ns)
Analysis of Deviance Table

Model: binomial, link: logit

Response: chd

Terms added sequentially (first to last)

                Df Deviance Resid. Df Resid. Dev
NULL                              461     596.11
ns(sbp, 4)       4    19.26       457     576.85
ns(tobacco, 4)   4    46.90       453     529.95
ns(ldl, 4)       4    19.08       449     510.87
famhist          1    25.29       448     485.58
ns(obesity, 4)   4     3.73       444     481.86
ns(age, 4)       4    23.77       440     458.09
```

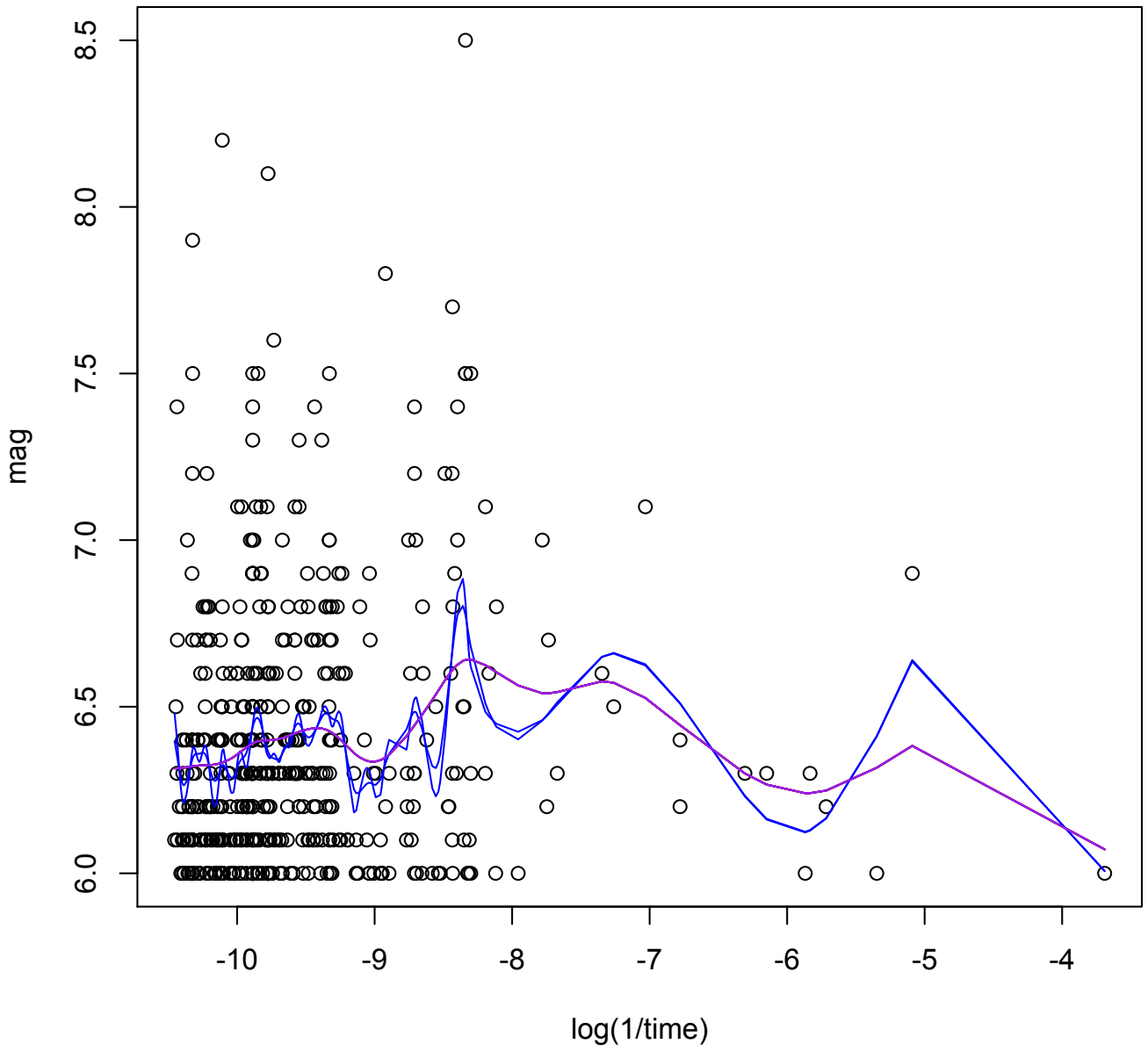
Smoothing splines §10.7.2

- $y_j = g(t_j) + \epsilon_j, \quad j = 1, \ldots, n$
- choose $g(\cdot)$ to solve

$$\min_g \sum_{j=1}^{n} \frac{\{y - g(t_j)\}^2}{2\sigma^2} - \frac{\lambda}{2\sigma^2} \int_a^b \{g''(t)\}^2 dt, \quad , \lambda > 0$$

- solution is a cubic spline, with knots at each observed x_i value
- see Figure 10.18 for a non-regularized solution

- has an explicit, finite dimensional solution
- $\hat{g} = \{\hat{g}(t_1), \ldots, \hat{g}(t_n)\} = (I + \lambda K)^{-1} y$
- K is a symmetric $n \times n$ matrix of rank $n - 2$


mag
8.5
8.0
7.5
7.0
6.5
6.0
-10
-9
-8
-7
-6
-5
-4
log(1/time)

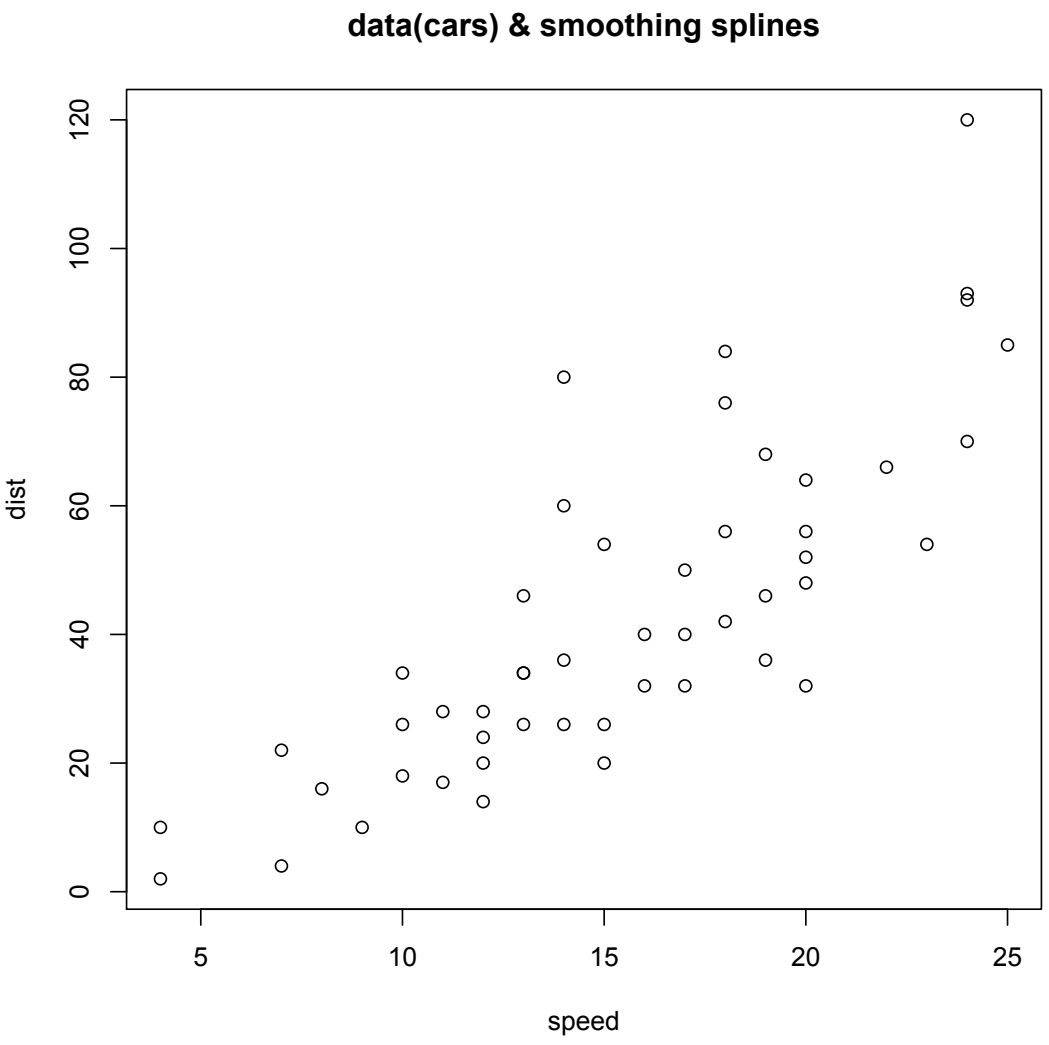
... smoothing splines

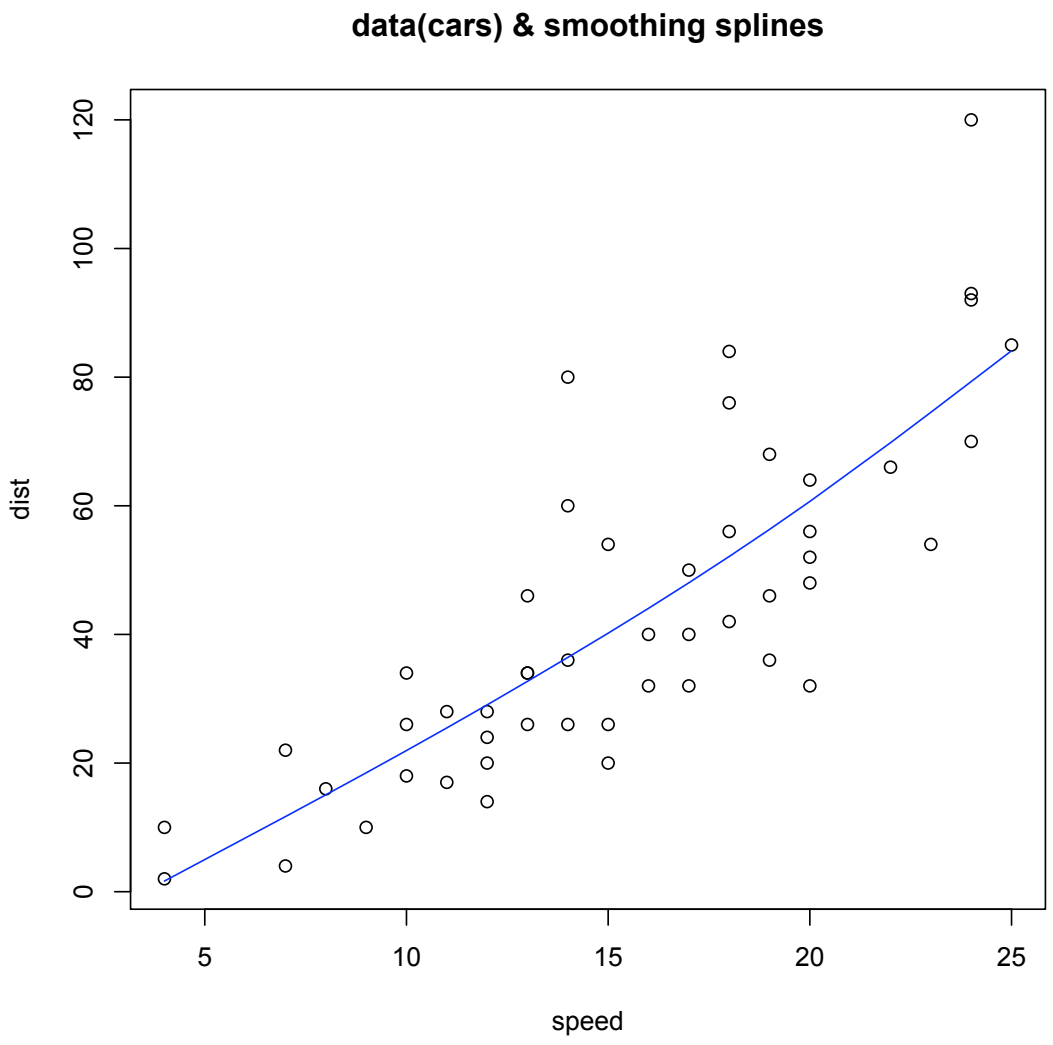
```
> quake$int = log(1/quake$time)
> quake[1:4,]
       time mag       int
1  40.08333 6.0 -3.690961
2 162.38889 6.9 -5.089994
3 210.22917 6.0 -5.348198
4 303.85417 6.2 -5.716548

> attach(quake)
> plot(int,mag)
> quake.ss2 = smooth.spline(x = int, y = mag, df = 5)
> lines(quake.ss2, col="red")
> quake.ss3
Call:
smooth.spline(x = int, y = mag, cv = TRUE)

Smoothing Parameter  spar= 1.499945  lambda= 0.0001340604 (25 iterations)
Equivalent Degrees of Freedom (Df): 11.35023
Penalized Criterion: 64.57512
PRESS: 0.1730025
> lines(quake.ss3, col="blue")
```

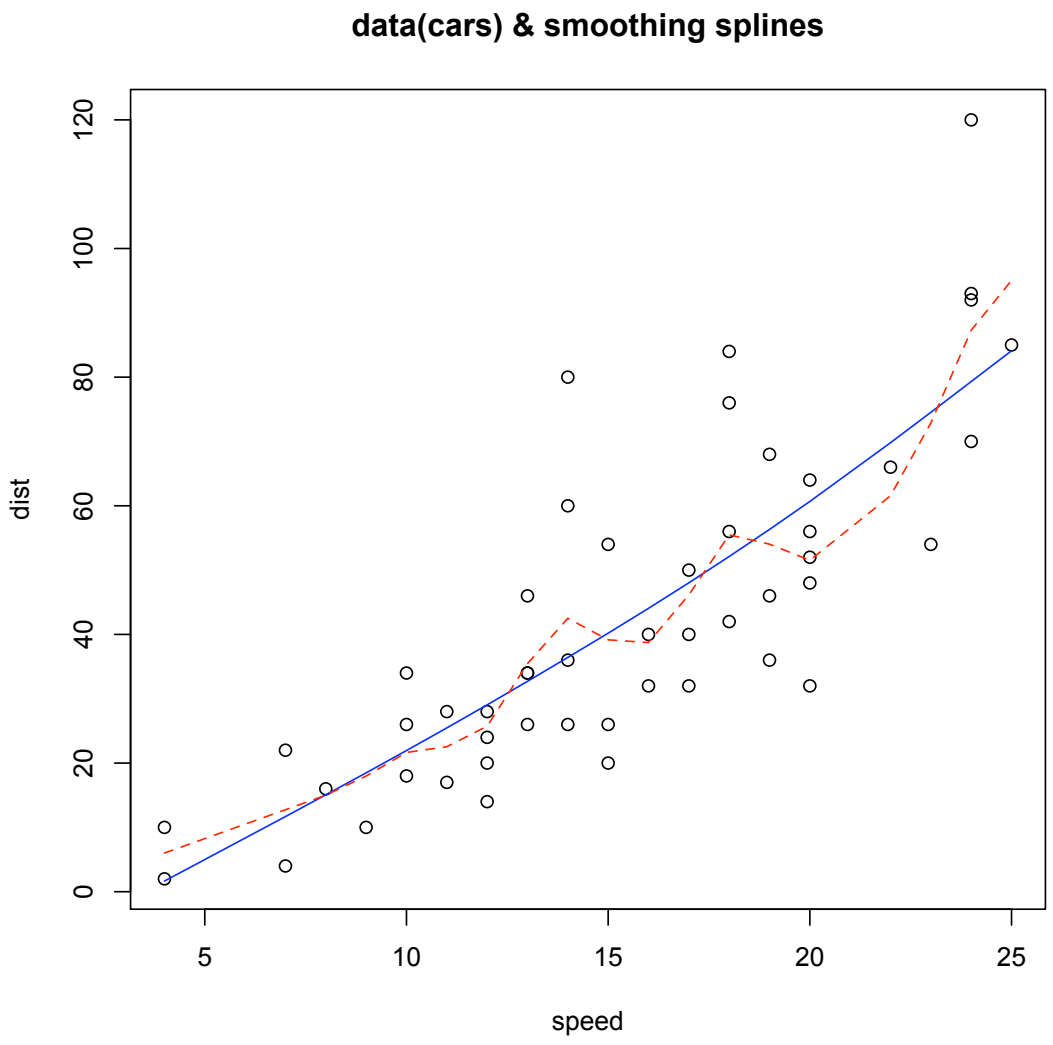
... smoothing splines

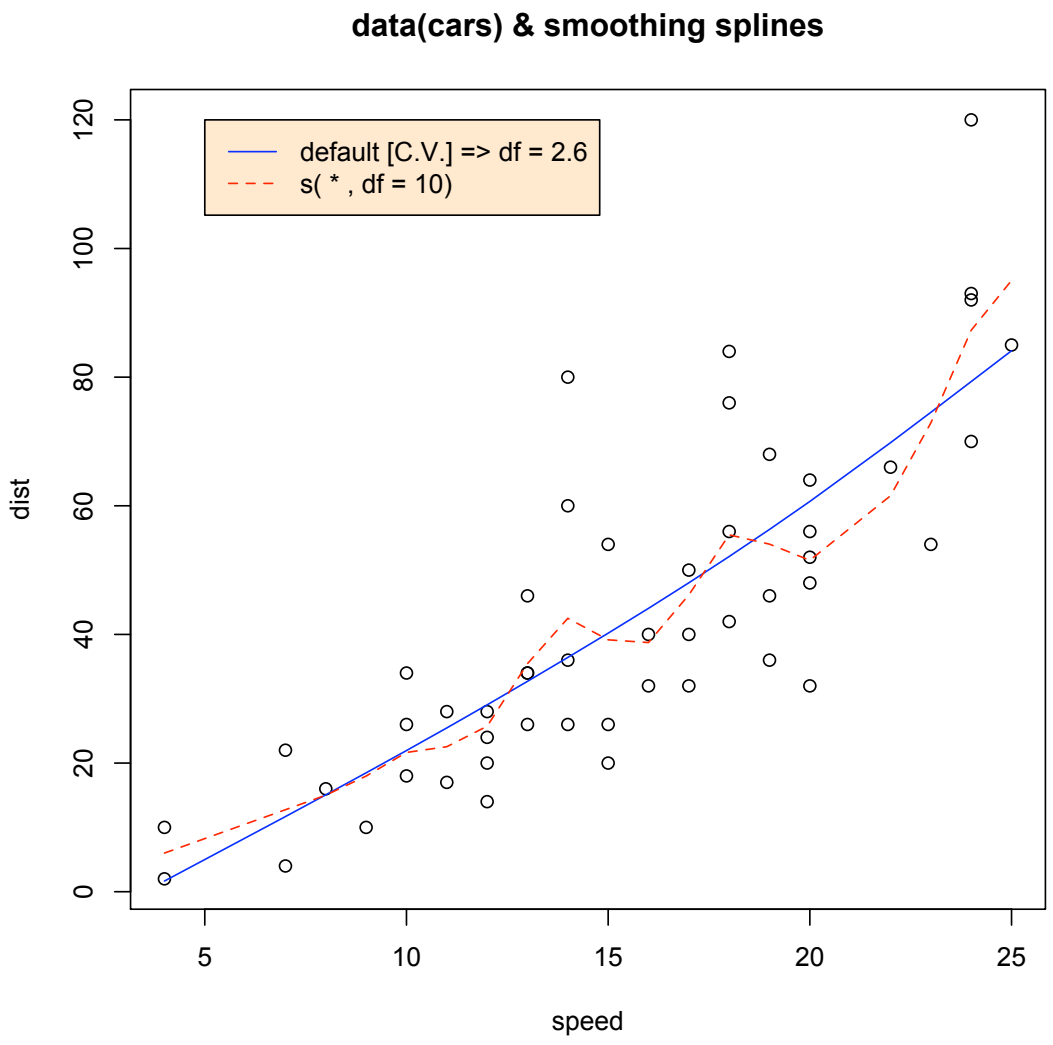
An example from the R help file for `smooth.spline`:

```
> data(cars)
> attach(cars)
> plot(speed, dist, main = "data(cars)  &  smoothing splines")
> cars.spl <- smooth.spline(speed, dist)
> (cars.spl)
Call:
smooth.spline(x = speed, y = dist)

Smoothing Parameter  spar= 0.7801305  lambda= 0.1112206 (11 iterations)
Equivalent Degrees of Freedom (Df): 2.635278
Penalized Criterion: 4337.638
GCV: 244.1044
> lines(cars.spl, col = "blue")
>      lines(smooth.spline(speed, dist, df=10), lty=2, col = "red")
>      legend(5,120,c(paste("default [C.V.] => df =",round(cars.spl$df,1)),
+                     "s( * , df = 10)"), col = c("blue","red"), lty = 1:2,
+             bg='bisque')
> detach()
```


data(cars) & smoothing splines
dist
speed
0
20
40
60
80
100
120
5
10
15
20
25


data(cars) & smoothing splines
dist
speed
0
20
40
60
80
100
120
5
10
15
20
25


data(cars) & smoothing splines
dist
speed
0
20
40
60
80
100
120
5
10
15
20
25


data(cars) & smoothing splines
default [C.V.] => df = 2.6
s(* , df = 10)
dist
speed
0
20
40
60
80
100
120
5
10
15
20
25

Multidimensional splines

- so far we are considering just 1 X at a time
- for regression splines we replace each X by the new columns of the basis matrix
- for smoothing splines we get a univariate regression
- it is possible to construct smoothing splines for two or more inputs simultaneously, but computational difficulty increases rapidly
- these are called thin plate splines
- alternative:
  $E(Y \mid X_1, \ldots, X_p) = f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p)$
  additive models
- binary response:
  $\text{logit}\{E(Y \mid X_1, \ldots, X_p)\} = f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p)$
  generalized additive models

Which smoothing method?

- basis functions: natural splines, Fourier, wavelet bases
- regularization via cubic smoothing splines
- kernel smoothers: locally constant/linear/polynomial
- adaptive bandwidth, running medians, running *M*-estimates
- Dantzig selector, elastic net, rodeo (Lafferty & Wasserman, 2008)
- Faraway (2006) Extending the Linear Model:
  - with very little noise, a small amount of local smoothing (e.g. nearest neighbours)
  - with moderate amounts of noise, kernel and spline methods are effective
  - with large amounts of noise, parametric methods are more attractive
- "It is not reasonable to claim that any one smoother is better than the rest"
  - `loess` is robust to outliers, and provides smooth fits
  - spline smoothers are more efficient, but potentially sensitive to outliers

Ethics and Statistics

Chance Magazine, 2011 # 4 and 2012 # 1

... ethics

Open Data and Open Methods

- **Columns,Ethics and Statistics**

An ethics problem arises when you are considering an action that (a) benefits you or some cause you support, (b) hurts or reduces benefits to others, and (c) violates some rule. Other definitions are possible; there is a vast literature on professional ethics that I will not discuss, instead focusing here on my own perspective as a statistician.

"In future columns, I would like to explore many dimensions of ethics, including those that arise in clinical research and statistical analysis, to problems involving probability and uncertainty, as well as more general concerns such as plagiarism and misrepresentation of research findings"

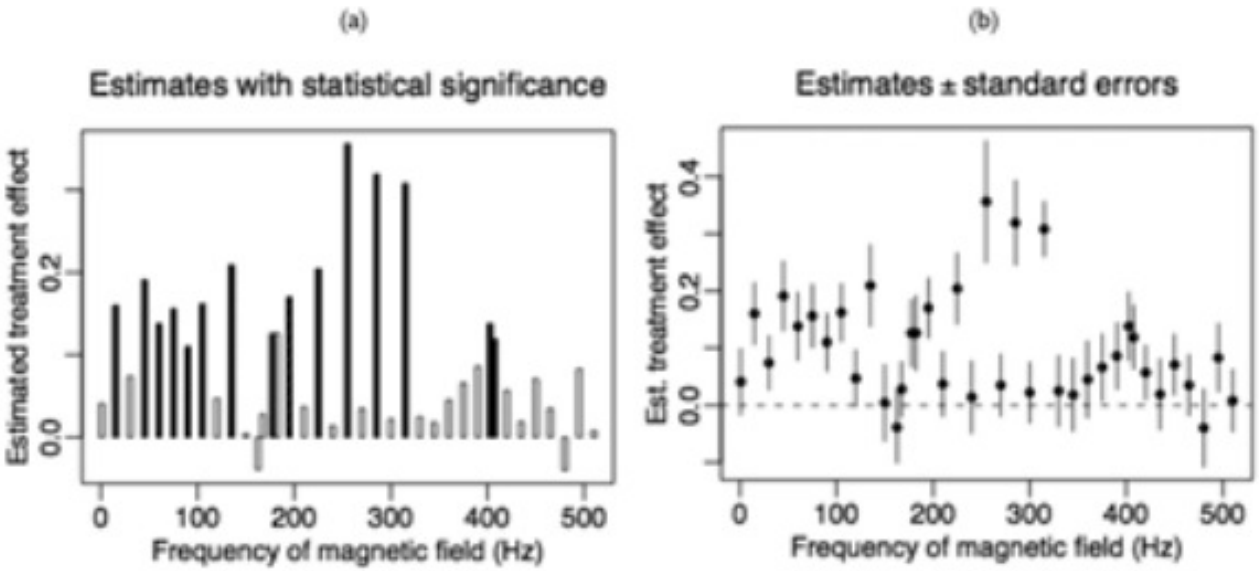
... ethics

Example:

An Unethical Refusal to Share Data

"Before attempting any sort of quantitative treatment, however, I will tell some stories. The story for the present column concerns the ethical imperative to share data. ... A bit more than 20 years ago, I attended – as a PhD student – a statistics conference on the health effects of low-frequency electromagnetic fields."

"The treatment appeared to have an effect, and it varied by frequency, not in any obvious way, but perhaps in some manner that made sense given the underlying biophysics. Figure 1a shows the basic findings of Blackman et al., in which they summarized their results based on the statistical significance level of their estimate at each frequency."



"From my statistical training, I was suspicious of using significance levels in this way – indeed, several years later, Hal Stern and I wrote a paper, "The Difference Between 'Significant' and 'Not Significant' Is Not Itself Statistically Significant" – and so I made a new graph showing estimates and confidence intervals, shown here as Figure 1b."

... ethics

"I need to respond to the column by Andrew Gelman about ethics (Vol. 24, No. 4). Most of the column is about a paper published by the principal investigator, Carl Blackman, and me, as the statistician on the project. There are basically two parts to his column. The first is a claim of us being unethical and the second is his assertion of a flawed statistical analysis."

"Gelman says the analysis was flawed and, as he pointed out several times, his "proof" seems to be that he had a PhD (although not at the time) and I only had a master's degree."

"Gelman is correct that ethics is important. We should all be ethical in our research, and so too should we be ethical in our complaints about ethics." ... Dennis House

... ethics

"Dr. Gelman levels the charge, 20 years after the fact, that I violated the principle of openness in scientific research by denying his request to send him copies of my logbooks and that I designed experiments and data analyses that led to a "waste of effort," presumably because I and my coworker misapplied statistical principles in the analysis of the experimental findings. Both assertions are based on misleading and incomplete information, and in my view, are groundless."

"The speculative use of p-values to highlight features of the data was far from "a waste of effort"; rather, it led ... to scientific discovery that has had substantial, beneficial consequences for expanding the understanding of how electromagnetic fields can influence biological systems and processes."

"Perhaps there are even good reasons why the statistically sophisticated neuroscience research community, in some cases, still draws conclusions from the differences between significance levels."... Carl Blackman

60 Minutes

Anil Potti, Duke University

from Wikipedia, “Potti is alleged to have engaged in scientific misconduct while a cancer researcher at both Duke University’s Medical Center and School of Medicine. He resigned in November 2010 after Duke suspended him, terminated the clinical trials based on his research and retracted his published data. A scientific misconduct investigation is ongoing.”

from Eric, “Kevin Baggerly and Kevin Coombes from the University of Texas MD Anderson Cancer Center were the researchers who made significant contributions in recognizing this fraud by unsuccessfully trying to reconstruct Potti’s results with his data.”

... 60 Minutes

The Annals of Applied Statistics
2009, Vol. 3, No. 4, 1309–1334
DOI: 10.1214/09-AOAS291

DERIVING CHEMOSENSITIVITY FROM CELL LINES: FORENSIC BIOINFORMATICS AND REPRODUCIBLE RESEARCH IN HIGH-THROUGHPUT BIOLOGY

BY KEITH A. BAGGERLY[1] AND KEVIN R. COOMBES[2]

University of Texas

High-throughput biological assays such as microarrays let us ask very detailed questions about how diseases operate, and promise to let us personalize therapy. Data processing, however, is often not described well enough to allow for exact reproduction of the results, leading to exercises in "forensic bioinformatics" where aspects of raw data and reported results are used to in-

"In this report we examine several related papers purporting to use microarray-based signatures of drug sensitivity derived from cell lines to predict patient response. Patients in clinical trials are currently being allocated to treatment arms on the basis of these results. However, we show in five case studies that the results incorporate several simple errors that may be putting patients at risk."

NY Times Link

HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | Edition: U.S. / Global

The New York Times

Business Day

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION | ARTS | STYL

Search | Global | DealBook | Markets | Economy | Energy | Media | Pe

Justices Back Mayo Clinic Argument on Patents

By ADAM LIPTAK
Published: March 20, 2012

WASHINGTON — The Supreme Court unanimously ruled on Tuesday that medical tests that rely on correlations between drug dosages and treatment are not eligible for patent protection.

Writing for the court, Justice Stephen G. Breyer said natural laws may not be patented standing alone or in connection with processes that involve "well-understood, routine, conventional activity."

RECOMMEND
TWITTER
LINKEDIN
SIGN IN TO E-MAIL
PRINT
REPRINTS

Log in to see
are sharing c
Privacy Polic

What's Po

The Benefit
Bilingualism

Interpretation of results

```
> model3=lm(accrate~herd+country+incnt_inst+incnt_indiv);
> summary(model3)

Call:
lm(formula = accrate ~ herd + country + incnt_inst + incnt_indiv)

Residuals:
    Min      1Q   Median      3Q     Max
-1.32868 -0.24326  0.02753  0.24041  1.66754

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.058861   0.795847  -1.330 0.184494
herd             -0.172756   0.081466  -2.121 0.034880 *
countryAUSTRIA    0.247744   0.190419   1.301 0.194364
countryBELGIUM   -0.014519   0.186436  -0.078 0.937983
countryCANADA    -0.004316   0.203303  -0.021 0.983079
countryCHINA     -1.066666   0.228357  -4.671 4.75e-06 ***
countryDENMARK    0.258505   0.204457   1.264 0.207207
countryFINLAND   -0.734713   0.199212  -3.688 0.000274 ***
countryFRANCE     0.377458   0.214447   1.760 0.079529 .
countryGERMANY    0.581863   0.234003   2.487 0.013510 *
countryGREECE    -0.013584   0.212099  -0.064 0.948983
countryHUNGARY   -0.182257   0.255822  -0.712 0.476817
countryICELAND    0.909629   0.375108   2.425 0.015973 *
countryIRELAND   -0.105535   0.252915  -0.417 0.676813
countryISRAEL    -0.579917   0.209803  -2.764 0.006105 **
countryITALY     -0.086864   0.201099  -0.432 0.666129
countryJAPAN     -0.041103   0.257750  -0.159 0.873422
countryKOREA     -1.037880   0.197248  -5.262 2.93e-07 ***
countryNETHERLANDS  0.165324   0.189812   0.871 0.384543
countryNEW ZEALAND -0.184874   0.252181  -0.733 0.464139
countryNORWAY    -0.105921   0.205659  -0.515 0.606956
countryPOLAND    -0.413966   0.201613  -2.053 0.041021 *
countryPORTUGAL  -0.156725   0.224211  -0.699 0.485155
countryRUSSIA    -0.524473   0.238484  -2.199 0.028722 *
countrySINGAPORE -0.843727   0.232082  -3.635 0.000333 ***
countrySPAIN     -0.246486   0.228118  -1.081 0.280883
countrySWEDEN    -0.127983   0.188354  -0.679 0.497421
```