Section 2: The Classical Linear Model

<u>Regression analysis</u> - a process by which a descriptive or predictive relationship is derived from data consisting of values of a response variable, y, and corresponding values of the predictor or regressor variables $x_1, x_2, \ldots$.

<u>Aims</u> - 1. ~~estimate~~ parameters
2. ~~test~~ whether a subset of the parameters is 0
3. to develop a descriptive relationship
4. to develop a predictive relationship
5. to predict future observations
6. to make decisions
7.

31 Aug 01

<u>Skills</u> - ① recognize when regression analysis is appropriate
② be able to perform the analysis
③ know when the assumptions necessary for (LS) regression are appropriate and when they are not
④ able to use the computer programs
⑤ provide useful summary information
⑥ know practical importance of the assumptions
⑦ explain anova as it relates to regression
⑧ interpret the anova to determine if a substantial relationship has been discovered

31 Aug 01

Linear Models

Gauss-Markov Theorem

- a basic result of ~~statistics~~/science
- leads to regression analysis which is the workhorse of ~~statistics~~
- has many variants and extensions
- relates to least squares analysis

Simple beginning

$Y_1, \ldots, Y_n$ i.i.d. $EY = \mu$

$$Y_j = \mu + \varepsilon_j \qquad E\varepsilon_j = 0$$

To estimate μ, OLS

$$\min \sum_j (Y_j - \mu)^2$$

$$\hat{\mu} = \overline{Y}$$

$$E\overline{Y} = \mu, \quad \text{var}\, \overline{Y} = \sigma^2/n$$

Will see that $\overline{Y}$ is BLUE of μ

31 Aug, 01

But <u>care</u> is needed.

$$Y_j = \alpha + \beta X_j + \varepsilon_j$$

OLS $\quad \hat{\alpha} + \hat{\beta}\bar{X} = \bar{Y}$

$$\hat{\beta} = \sum (Y_j - \bar{Y})(X_j - \bar{X}) \Big/ \sum (X_j - \bar{X})^2$$

Suppose $X_j \equiv C$, trouble.
But can still learn something.

$$Y_j = \alpha + C\beta + \varepsilon_j$$

So can get and use BLUE of $\mu = \alpha + C\beta$

On occasion extra parameters are introduced for simplification

eg. one way array

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad j = 1, .., J;\ i = 1, .., I$$

singular $\mu \Rightarrow \mu + c$, $\alpha_i \Rightarrow \alpha_i - c$ model unchanged

but the α_i have a simple interpretation - <u>effects</u>

One approach, $\sum_i \alpha_i = 0$. - a constraint

31 Aug 01

G-M
Model: $\underset{n\times 1}{\underset{\sim}{Y}} = \underset{n\times m}{\underset{\sim}{X}}\,\underset{m\times 1}{\underset{\sim}{\beta}} + \underset{n\times 1}{\underset{\sim}{\epsilon}}$

$E\underset{\sim}{\epsilon} = 0$, $\text{var}\,\underset{\sim}{\epsilon} = \sigma^2 \underset{\sim}{I}$, $\underset{\sim}{X}$ fixed

$E\underset{\sim}{Y} = E\{\underset{\sim}{Y} \mid \underset{\sim}{X}\} = \underset{\sim}{X}\underset{\sim}{\beta}$, $\text{var}\,\underset{\sim}{Y} = \sigma^2 \underset{\sim}{I}$

Interested in "effective" estimation of $\underset{\sim}{\beta}$
Examples: Nile-wavelets, electron micrographs
At this point BLUE

Properties of vector rv's. E, var, $\text{cov}\{\underset{\sim}{C}\underset{\sim}{Y}, \underset{\sim}{D}\underset{\sim}{Y}\} = \underset{\sim}{C}\,\underset{\sim}{\Sigma}_{YY}\underset{\sim}{D}'$

$\text{var}\,\underset{\sim}{B}\underset{\sim}{Y} = \underset{\sim}{B}\,\Sigma_{YY}\underset{\sim}{B}'$

Matrices: $\underset{\sim}{I}$, inverse, trace, latents, transpose, rank, ..

svd $\underset{m\times n}{\underset{\sim}{A}} = \underset{m\times m}{\underset{\sim}{U}}\,\underset{m\times n}{\underset{\sim}{\Lambda}}\,\underset{n\times n}{\underset{\sim}{V}'}$ $\underset{\sim}{U}, \underset{\sim}{V}$ orthogonal

$\underset{\sim}{\Lambda} = \text{diag}\{\lambda_j\}$, $\lambda_j \geq 0$

$r(\underset{\sim}{A}) = \#\{\lambda_j \neq 0\}$

$\underset{\sim}{A}^- = \underset{\sim}{V}\,\underset{\sim}{\Lambda}^-\,\underset{\sim}{U}'$ $\underset{\sim}{\Lambda}^- = \text{diag}\{1/\lambda_j \mid \lambda_j \neq 0\}$

Will follow notation of Rao (7.3) as much as possible

$\underset{\sim}{A}\underset{\sim}{A}^-\underset{\sim}{A} = \underset{\sim}{U}\underset{\sim}{\Lambda}\underset{\sim}{V}'\underset{\sim}{V}\underset{\sim}{\Lambda}^-\underset{\sim}{U}'\underset{\sim}{U}\underset{\sim}{\Lambda}\underset{\sim}{V}' = \underset{\sim}{U}\underset{\sim}{\Lambda}\underset{\sim}{V}' = \underset{\sim}{A}$

31 Aug 01

Some linear algebra background

$\underset{\sim}{X}^{n \times m}$: matrix

$\mathcal{M}(\underset{\sim}{X})$: manifold/space generated by columns of $\underset{\sim}{X}$

$\underset{\sim}{X} = [\underset{\sim}{X}_1 \ldots \underset{\sim}{X}_m]$

$\sum_{j=1}^{m} \alpha_j \underset{\sim}{X}_j$ $\qquad \alpha_j$: scalars

a.k.a. range (space) of $\underset{\sim}{X}$, column space of $\underset{\sim}{X}$

rank of $\underset{\sim}{X}$ $\quad r(\underset{\sim}{X})$ $\quad \dim \mathcal{M}(\underset{\sim}{X})$
: # linearly indep. columns

$r(X) = r(\underset{\sim}{X}') = r(X'X)$ Will prove this later

31 Aug 01

Remarks:

$\mathcal{V}$: all column vectors with n entries

$\mathcal{V} = \mathcal{S} + \mathcal{S}^{\perp}$ $\qquad$ $\mathcal{S}^{\perp}$: <u>null space of $(\mathcal{S})$</u>

$\mathcal{S}^{\perp}$: vectors orthogonal to vectors of $\mathcal{S}$

$\mathcal{V} = \mathcal{M}(\underset{\sim}{X}') + \mathcal{N}(\underset{\sim}{X}')$

$\mathcal{V} = \mathcal{M}(\underset{\sim}{X}'\underset{\sim}{X}) + \mathcal{N}(\underset{\sim}{X}'\underset{\sim}{X})$

and so $\mathcal{M}(\underset{\sim}{X}') \equiv \mathcal{M}(\underset{\sim}{X}'\underset{\sim}{X})$ iff $\mathcal{N}(\underset{\sim}{X}') = \mathcal{N}(\underset{\sim}{X}'\underset{\sim}{X})$

31 Aug 01

Lemma. $\mathcal{M}(X') = \mathcal{M}(X'X)$, i.e. spaces generated by columns of X' and $X'X$ are the same.

Proof. If α column vector such that

$$\alpha' X' = 0 \qquad \text{i.e.} \quad \alpha \perp X'$$

then $$\alpha' X' X = 0 \qquad \text{i.e.} \quad \alpha \perp X'X$$

Next if

$$\alpha' X' X = 0 \qquad \text{i.e.} \quad \alpha \perp X'X$$

then $$\alpha' X' X \alpha = 0 \qquad \text{i.e.} \quad \alpha' X' = 0$$

i.e. $$\alpha \perp X'$$

I.e. every vector orthogonal to X' is also orthogonal to $X'X$ and $\mathcal{M}(X') = \mathcal{M}(X'X)$.

and so $\mathcal{M}(X') = \mathcal{M}(X'X)$

rank $\quad r(X) = \dim \mathcal{M}(X) \equiv$ number of linearly independent columns

$$r(X') = r(X'X) \qquad (\text{Also} = r(X))$$

31 Aug 01

The G-M model is the basis for so much.

→ Extends OLS

Definition. Given $\underset{\sim}{P}^{m\times 1}$, the expression $\underset{\sim}{P}'\underset{\sim}{\beta}$ is <u>estimable</u> if $\exists$ a linear function of $\underset{\sim}{Y}$ with expectation $\underset{\sim}{P}'\underset{\sim}{\beta}$ for all $\underset{\sim}{\beta}$.

i.e. $\exists\ \underset{\sim}{L} \ni E\,\underset{\sim}{L}'\underset{\sim}{Y} = \underset{\sim}{P}'\underset{\sim}{\beta} \quad \forall \underset{\sim}{\beta}$

Lemma. If $\underset{\sim}{P}'\underset{\sim}{\beta}$ is estimable, $\underset{\sim}{P} \in \mathcal{M}(\underset{\sim}{X}') = \mathcal{M}(\underset{\sim}{X}'\underset{\sim}{X})$

Proof. $E(\underset{\sim}{L}'\underset{\sim}{Y}) = \underset{\sim}{L}'\underset{\sim}{X}\underset{\sim}{\beta} = \underset{\sim}{P}'\underset{\sim}{\beta} \quad \forall \underset{\sim}{\beta}$

$$\Longrightarrow \underset{\sim}{L}'\underset{\sim}{X} = \underset{\sim}{P}'$$

$$\underset{\sim}{P} = \underset{\sim}{X}'\underset{\sim}{L} \in \mathcal{M}(\underset{\sim}{X}')$$

31 Aug 01

Previous example $y_j = \beta_1 + C\beta_2 + \epsilon_j$

$$\underset{\sim}{X} = \begin{bmatrix} 1 & c \\ \vdots & \vdots \\ 1 & c \end{bmatrix} \qquad \underset{\sim}{X}'\underset{\sim}{X} = \begin{bmatrix} n & cn \\ cn & c^2 n \end{bmatrix} = n \begin{bmatrix} 1 & c \\ c & c^2 \end{bmatrix}$$

$$\mathcal{M}(\underset{\sim}{X}') = \alpha_1 \begin{bmatrix} 1 \\ c \end{bmatrix} + \dots + \alpha_n \begin{bmatrix} 1 \\ c \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ c \end{bmatrix} = \mathcal{M}(X'X)$$

$$\beta_1 + C\beta_2 = [1 \quad C] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$ clearly estimable

$$\underset{\sim}{P} = \begin{bmatrix} 1 \\ C \end{bmatrix}$$

Not a big deal

31 Aug 01

<u>Gauss Markov Theorem</u>. Suppose $E\,\underset{\sim}{Y} = \underset{\sim}{X}\,\underset{\sim}{\beta}$, $\text{var}\,\underset{\sim}{Y} = \sigma^2 \underset{\sim}{I}$. Let $\underset{\sim}{\rho}'\underset{\sim}{\beta}$ be an estimable function. Then in the class of all unbiased linear estimates of $\underset{\sim}{\rho}'\underset{\sim}{\beta}$, $\underset{\sim}{\rho}'\hat{\underset{\sim}{\beta}}$ has minimum variance and is unique, where $\hat{\underset{\sim}{\beta}}$ satisfies

$$\underset{\sim}{X}'\underset{\sim}{X}\hat{\underset{\sim}{\beta}} = \underset{\sim}{X}'\underset{\sim}{Y}$$

<u>Proof</u>.

$\underset{\sim}{\rho}'\underset{\sim}{\beta}$ is estimable, so $\underset{\sim}{\rho} \in \mathcal{M}(\underset{\sim}{X}'\underset{\sim}{X})$, i.e.

① $\underset{\sim}{\rho}_{m\times 1} = \underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\lambda}$ for some $\underset{\sim}{\lambda}_{m\times 1}$

Suppose $\underset{\sim}{L}'\underset{\sim}{Y}$ is some unbiased estimate of $\underset{\sim}{\rho}'\underset{\sim}{\beta}$

so $\underset{\sim}{L}'\underset{\sim}{X}\underset{\sim}{\beta} = \underset{\sim}{\rho}'\underset{\sim}{\beta} \quad \forall\, \underset{\sim}{\beta}$

② or $\underset{\sim}{L}'\underset{\sim}{X} = \underset{\sim}{\rho}'$

→ Consider $\text{var}\,\underset{\sim}{L}'\underset{\sim}{Y}$

$$= \text{var}\{\underset{\sim}{L}'\underset{\sim}{Y} - \underset{\sim}{\lambda}'\underset{\sim}{X}'\underset{\sim}{Y} + \underset{\sim}{\lambda}'\underset{\sim}{X}'\underset{\sim}{Y}\}$$

$$= \text{var}\{\underset{\sim}{L}'\underset{\sim}{Y} - \underset{\sim}{\lambda}'\underset{\sim}{X}'\underset{\sim}{Y}\} + \text{var}\{\underset{\sim}{\lambda}'\underset{\sim}{X}'\underset{\sim}{Y}\}$$

31 Aug 01

since

$$\mathrm{cov}\{(\underset{\sim}{L}'-\underset{\sim}{\lambda}'\underset{\sim}{X}')\underset{\sim}{Y},\ \underset{\sim}{\lambda}'\underset{\sim}{X}'\underset{\sim}{Y}\} = (\underset{\sim}{L}'-\underset{\sim}{\lambda}'\underset{\sim}{X}')\sigma^2 \underset{\sim}{I}\,\underset{\sim}{X}\,\underset{\sim}{\lambda}$$

$$= (\underset{\sim}{L}'\underset{\sim}{X} - \underset{\sim}{\lambda}'\underset{\sim}{X}'\underset{\sim}{X})\underset{\sim}{\lambda}\sigma^2$$

$$= (\underset{\sim}{P}' - \underset{\sim}{P}')\underset{\sim}{\lambda}\sigma^2 \qquad \text{from } ①, ②$$

So have

$$\mathrm{var}\{\underset{\sim}{L}'\underset{\sim}{Y}\} \geq \mathrm{var}\{\underset{\sim}{\lambda}'\underset{\sim}{X}'\underset{\sim}{Y}\}$$

$$= \mathrm{var}\{\underset{\sim}{\lambda}'\underset{\sim}{X}'\underset{\sim}{X}\hat{\beta}\}$$

$$= \mathrm{var}\ \underset{\sim}{P}'\hat{\underset{\sim}{\beta}}$$

with equality iff

$$0 = \mathrm{var}\{(\underset{\sim}{L}'-\underset{\sim}{\lambda}'\underset{\sim}{X}')\underset{\sim}{Y}\} = (\underset{\sim}{L}'-\underset{\sim}{\lambda}'\underset{\sim}{X}')(\underset{\sim}{L}-\underset{\sim}{X}\underset{\sim}{\lambda})\sigma^2$$

or $\underset{\sim}{L}' = \underset{\sim}{\lambda}'\underset{\sim}{X}'$

Implying $\underset{\sim}{L}'\underset{\sim}{Y} = \underset{\sim}{\lambda}'\underset{\sim}{X}'\underset{\sim}{Y} = \underset{\sim}{P}'\hat{\underset{\sim}{\beta}}$

See also that $\underset{\sim}{P}'\hat{\underset{\sim}{\beta}}$ has a unique value.

31 Aug 01

Notes.

1. $\underset{\sim}{X}'\underset{\sim}{X}\hat{\underset{\sim}{\beta}} = \underset{\sim}{X}'\underset{\sim}{Y}$ normal equations

$\underset{\sim}{X}'(\underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}}) = 0$

$\underset{\sim}{X} \perp \underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}} = \hat{\underset{\sim}{\epsilon}}$

2. If seek $\min_{\beta} (\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta}) = \|\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta}\|^2$

led to normal equations

$\|\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta}\|^2 = \|\underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}}\|^2 + \|\underset{\sim}{X}(\underset{\sim}{\beta} - \hat{\beta})\|^2$

1 Sept. 01

Examples of regression analysis.

a) Electron micrographs

Electron microscopy

Image $Y(x,y) = V(x,y) + \text{noise}$ $\quad 0<x<X$, $0<y<Y$

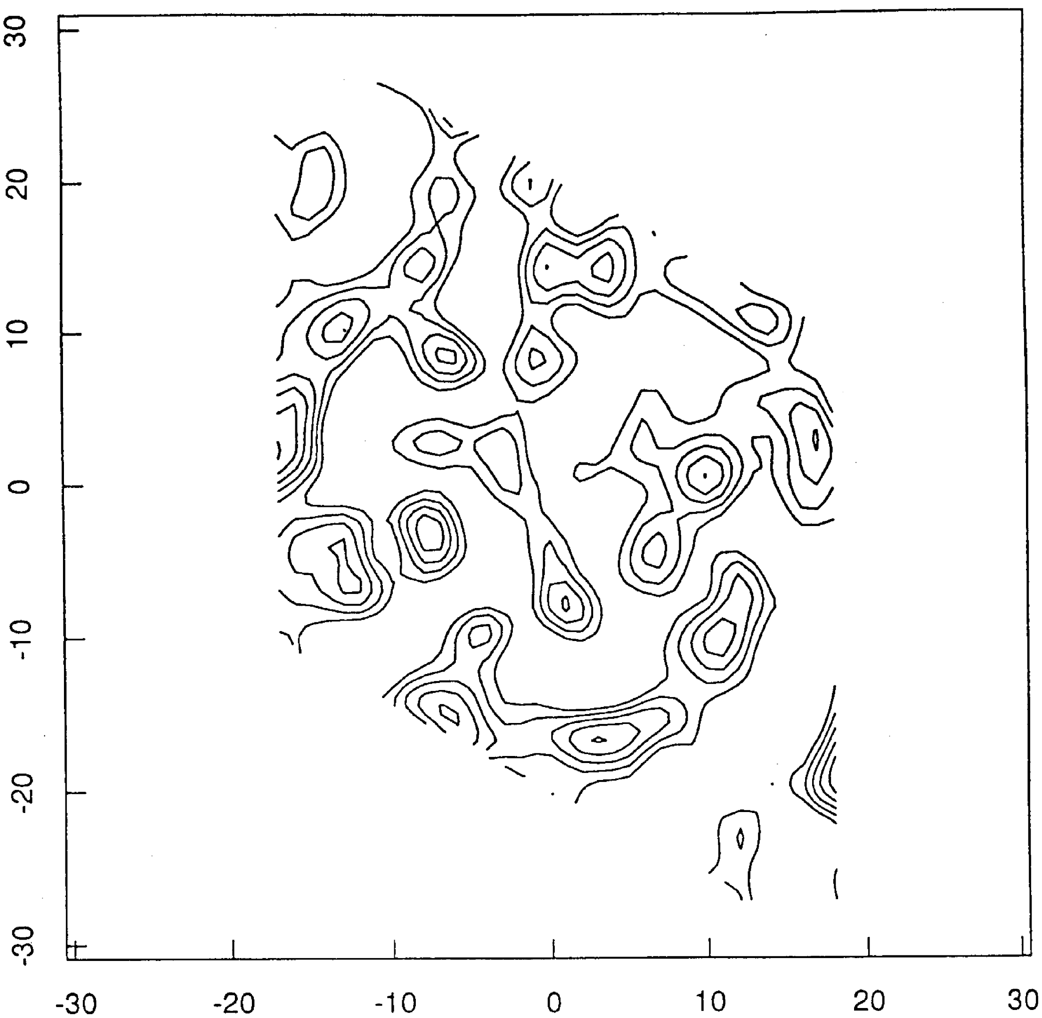
Crystal, periodic

$$V(x,y) = \sum_{h,k} F_{h,k}\, e^{2\pi i(hx+ky)/\Delta}$$

OLS $$\hat{F}_{h,k} \sim \int_0^Y \int_0^X Y(x,y)\, e^{-2\pi i(hx+ky)/\Delta}\, dx\, dy$$

$$\hat{V}(x,y) = \sum_{h,k} \hat{F}_{h,k}\, e^{2\pi i(hx+ky)/\Delta}$$

through electron microscopy effectively by averaging images of 160 ; cells. It is difficult to determine essential features of the substance from poor quality of such figures led electron microscopists to seek improved

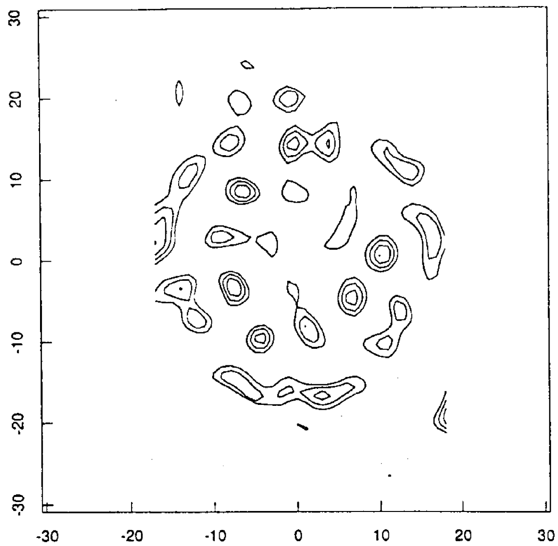


ge of purple membrane obtained by simple averaging of 160 neighbouring unit cells.

ıre of a crystal being periodic, it is convenient to represent V' by a . Supposing the period to be Δ along each of the axes, one can write

$$V'(x, y) = \sum_{h,k} F_{h,k} e^{2\pi i(hx+ky)/\Delta} \qquad (1.1)$$

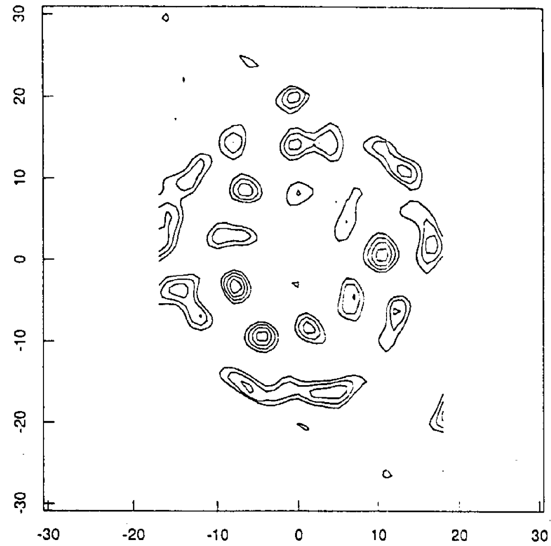
Estimated Purple Membrane Projection Structure - t1



J = 1

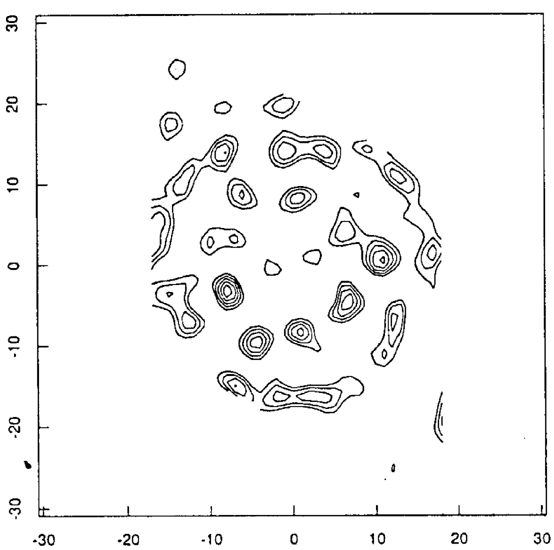
(b)

Estimated Purple Membrane Projection Structure - t1



J = 2

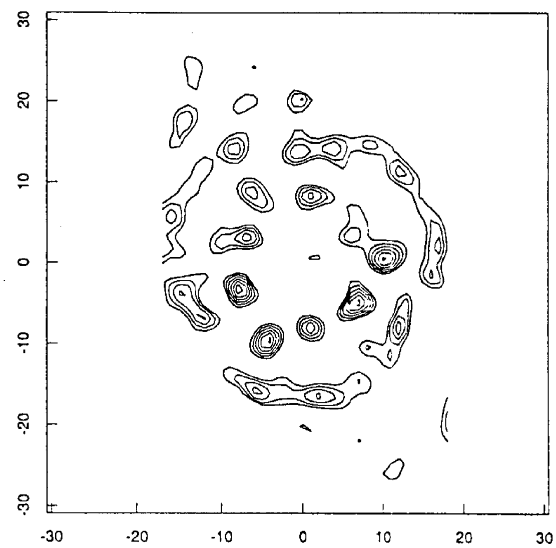
Estimated Purple Membrane Projection Structure - t1



(d)

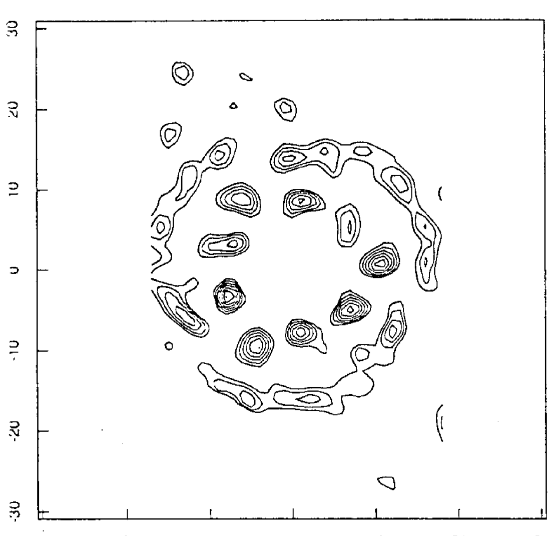
J = 4

Estimated Purple Membrane Projection Structure - t1



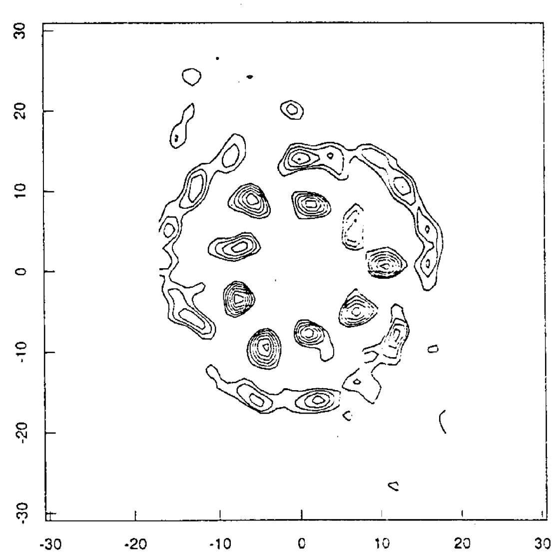
J = 8

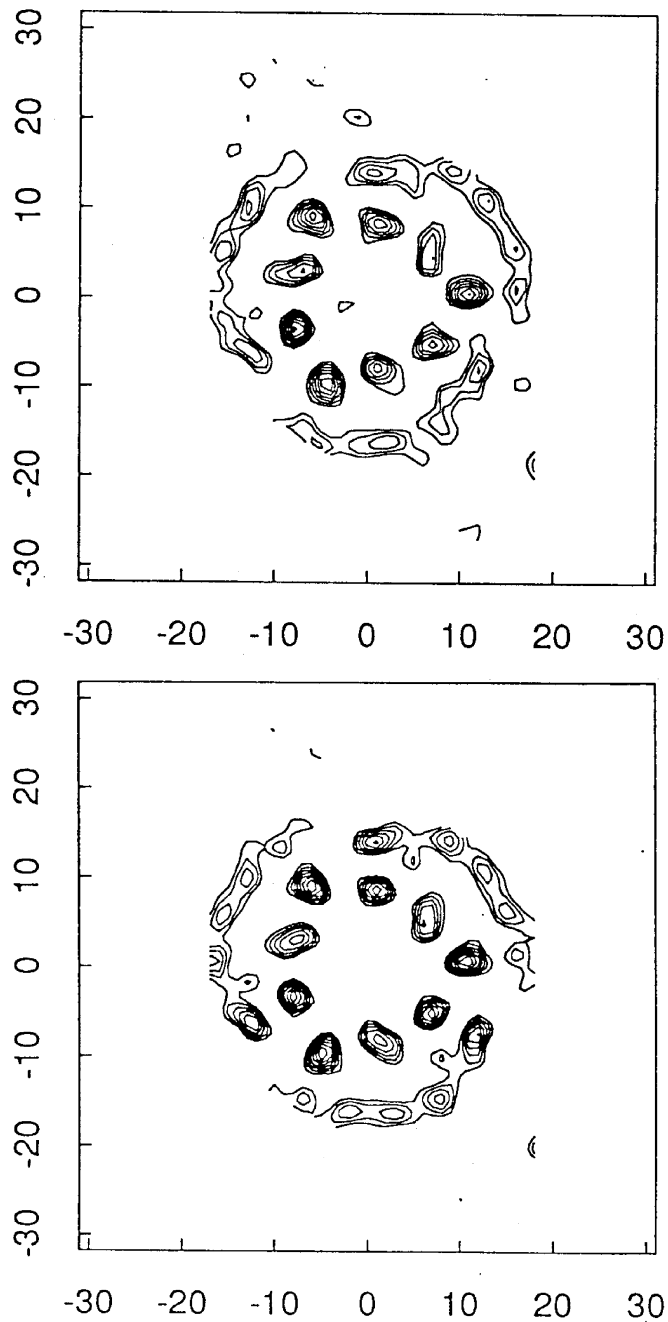
Estimated Purple Membrane Projection Structure - t1



(f)

Estimated Purple Membrane Projection Structure - t1





(a)

(b)

e final estimated image obtained by combining 42 individual images. The contour levels are the same as in Figs 2 and 4.
ghly accurate estimate based on values out to 5 Angstroms given in Henderson *et al.* (1986).

tively images to the highly accurate, 3-fold symmetrized image of Hender-
1986).

1 Sept 01

b) <u>Hydrology / Wavelet analysis</u>

$$g(x) = \sum_{jk} \beta_{jk} \psi_{jk}(x)$$

$$\psi_{jk}(x) = \psi(2^j x - k)$$ location & scale

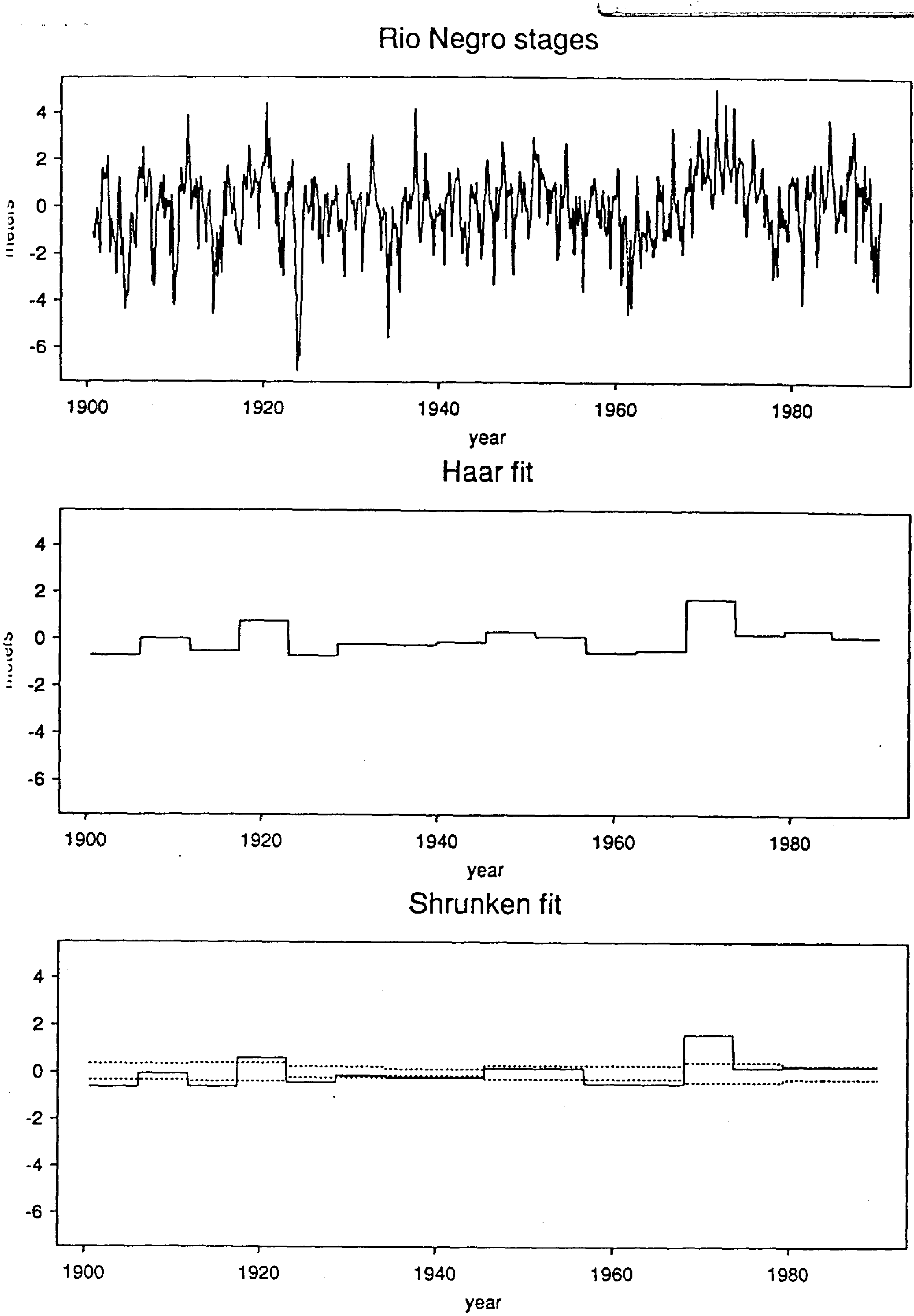
Time series $Y(t) = g(t/T) + \text{noise}$

$\hat{\beta}_{jk}$ from OLS

$\hat{g}(t)$

Haar function $\psi(x) = \begin{cases} -1 & 0 < x < \frac{1}{2} \\ 1 & \frac{1}{2} < x < 1 \end{cases}$

<u>Note</u>. In the examples shrinking was employed to improve the estimates.



ɜl is obtained from daily Rio Negro stages by computing the monthly averages, then removing the ; to get a seasonally adjusted monthly series. The middle panel is the naive Haar estimate (15) with ıel provides the wavelet estimate (19) employing the multiplier function (17). The dashed lines give approximate ± 2 standard error limits about the overall mean level

6 Sept. 2001

"Solving" for $\hat{\beta}$.

Generalized inverses.

$\underset{\sim}{A}^{m \times n}$

$\underset{\sim}{A}^{-}$ ($n \times m$) is a generalized inverse of $\underset{\sim}{A}$ if

$$\boxed{\underset{\sim}{A}\,\underset{\sim}{A}^{-}\underset{\sim}{A} = \underset{\sim}{A}}$$

Not necessarily unique

Example of use:

Theorem. If $\underset{\sim}{A}\,\underset{\sim}{X} = \underset{\sim}{Y}$ is a consistent equation, then $\underset{\sim}{X} = \underset{\sim}{A}^{-}\underset{\sim}{Y}$ is a solution.

Proof. $\underset{\sim}{Y} \in \mathcal{M}(\underset{\sim}{A})$ so $\underset{\sim}{Y} = \underset{\sim}{A}\,\underset{\sim}{\lambda}$ for some $\underset{\sim}{\lambda}$

$$\underset{\sim}{A}\,\underset{\sim}{X} = \underset{\sim}{A}\,(\underset{\sim}{A}^{-}\underset{\sim}{Y}) = \underset{\sim}{A}\,\underset{\sim}{A}^{-}\underset{\sim}{A}\,\underset{\sim}{\lambda} = \underset{\sim}{A}\,\underset{\sim}{\lambda} = \underset{\sim}{Y}$$

Corollary. $\underset{\sim}{X} = \underset{\sim}{A}^{-}\underset{\sim}{Y} + (\underset{\sim}{I} - \underset{\sim}{A}^{-}\underset{\sim}{A})\underset{\sim}{Z}$ is a general solution

6 Sept. 2001

Forms of $\underset{\sim}{A}^-$

1. Suppose $\underset{\sim}{A}$ has rank r

Let the columns of $\underset{\sim}{L}^{m\times r}$ form a basis for $\mathcal{M}(\underset{\sim}{A})$

so

$$\underset{m\times n}{\underset{\sim}{A}} = \underset{m\times r}{\underset{\sim}{L}}\ \underset{r\times n}{\underset{\sim}{R}}$$

for some $\underset{\sim}{R}$.

$m, n \geq r$ inverses below exist

$$\underset{n\times m}{\underset{\sim}{A}^-} = \underset{n\times r}{\underset{\sim}{R}'}\ \underset{r\times r}{(\underset{\sim}{R}\underset{\sim}{R}')^{-1}}\ \underset{r\times r}{(\underset{\sim}{L}'\underset{\sim}{L})^{-1}}\ \underset{r\times m}{\underset{\sim}{L}'}$$

Check

$$\underset{\sim}{A}\underset{\sim}{A}^-\underset{\sim}{A} = \underset{\sim}{L}\underset{\sim}{R}\underset{\sim}{R}'(\underset{\sim}{R}\underset{\sim}{R}')^{-1}(\underset{\sim}{L}'\underset{\sim}{L})^{-1}\underset{\sim}{L}'\underset{\sim}{L}\underset{\sim}{R}$$

$$= \underset{\sim}{L}\underset{\sim}{R}$$

2. Singular value decomposition

$$\underset{m\times n}{\underset{\sim}{A}} = \underset{m\times r}{\underset{\sim}{U}}\ \underset{r\times r}{\underset{\sim}{\Lambda}}\ \underset{r\times n}{\underset{\sim}{V}'}$$

with $\underset{\sim}{U}'\underset{\sim}{U} = \underset{\sim}{I}$, $\underset{\sim}{V}\underset{\sim}{V}' = \underset{\sim}{I}$, $\underset{\sim}{\Lambda} = \text{diag}\{\lambda_j\}$, $\lambda_j > 0$
$r = r(\underset{\sim}{A})$

$$\underset{\sim}{\Lambda}^- = \{1/\lambda_j\}$$

$$\underset{\sim}{A}^- = \underset{\sim}{V}\underset{\sim}{\Lambda}^-\underset{\sim}{U}'$$

This provides the solution of $\underset{\sim}{A}\underset{\sim}{X} = \underset{\sim}{Y}$ with min $\|\underset{\sim}{X}\|^2$

Sometimes $\underset{m\times m}{\underset{\sim}{U}}$, $\underset{m\times n}{\underset{\sim}{\Lambda}}$, $\underset{n\times n}{\underset{\sim}{V}}$

6 Sept. 2001

Cholesky decomposition chol()

$\underset{\sim}{A} = \underset{\sim}{U}^T \underset{\sim}{U}$ $\underset{\sim}{A}$ symm $\geq \underset{\sim}{0}$

$\underset{\sim}{U}$ upper-triangular

OR

$\underset{\sim}{A} = \underset{\sim}{L}\underset{\sim}{L}^T$

$\underset{\sim}{L}$ lower-triangular

QR decomposition qr()

$\underset{\sim}{M} = \underset{\sim}{Q}\underset{\sim}{R}$

$\underset{\sim}{Q}$: orthonormal columns

$\underset{\sim}{R}$: upper triangular

faster than svd

6 Sept 2001

3. Splus uses QR decomposition

Procedures in Splus

solve( )
chol( )
eigen( )
svd( )
qr( )

6 Sept. 2001

Back to the G-M setup.

$$X'X\hat{\beta} = X'Y$$

$$C = (X'X)^{-}$$

$\hat{\beta} = CX'Y$ is a solution

Theorem. Under G-M conditions ($p'\beta$ estimable)

i) $E\, p'\hat{\beta} = p'\beta$

ii) $\text{var}\{p'\hat{\beta}\} = \sigma^2 p'Cp$

iii) $\text{cov}\{p'\hat{\beta}, R'\hat{\beta}\} = \sigma^2 p'CR$ $\quad$ p, R estimable

Note. Can view $\sigma^2(X'X)^{-}$ as var $\hat{\beta}$ provided just consider estimable functions. $\text{cov}\{\hat{\beta}_i, \hat{\beta}_j\} =$

Proof. $p'\beta$ is estimable, so $p = X'X\lambda$ for some λ

$$E\, p'\hat{\beta} = p'CX'X\beta = \lambda X'XCX'X\beta$$

$$= \lambda X'X\beta$$

$$= p\beta$$

6 Sept. 2001

$$\text{var}\{\underset{\sim}{\rho}'\,\underset{\sim}{\hat\beta}\} = \text{var}\{\underset{\sim}{\lambda}'\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\sim}{C}\,\underset{\sim}{X}'\underset{\sim}{Y}\}$$

$$= \underset{\sim}{\lambda}'\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\sim}{C}\,\underset{\sim}{X}'\,\sigma^2\underset{\sim}{I}\,\underset{\sim}{X}\,\underset{\sim}{C}'\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\sim}{\lambda}$$

$$= \underset{\sim}{\lambda}'\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{C}\,\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\sim}{\lambda}\;\sigma^2$$

$\underset{\sim}{C}'$ also a gen inverse

$\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\sim}{C}\,\underset{\sim}{X}'\underset{\sim}{X} = \underset{\sim}{X}'\underset{\sim}{X}$

$\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\sim}{C}'\underset{\sim}{X}'\underset{\sim}{X} = \underset{\sim}{X}'\underset{\sim}{X}$

$$= \underset{\sim}{\rho}'\underset{\sim}{C}\,\underset{\sim}{\rho}\;\sigma^2$$

Note. If $\underset{\sim}{X}'\underset{\sim}{X}$ is invertible, $\underset{\sim}{C} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}$

$$\text{var}\,\hat{\underset{\sim}{\beta}}_i = \sigma^2 C_{ii}$$

$$\text{cov}\{\hat\beta_i, \hat\beta_j\} = \sigma^2 C_{ij}$$

---

8 Sept. 2001

Further aspects of the GM setup.

$$E\,\underset{\sim}{Y} = \underset{\sim}{X}\,\underset{\sim}{\beta}$$

so $\underset{\sim}{X}\underset{\sim}{\beta}$ is estimable by

$$\underset{\sim}{X}\hat{\underset{\sim}{\beta}} = \underset{\sim}{X}\,\underset{\sim}{C}\,\underset{\sim}{X}'\underset{\sim}{Y}$$

$$= \underset{\sim}{H}\,\underset{\sim}{Y} = \hat{\underset{\sim}{Y}} \qquad \text{fitted values}$$

where $\underset{\sim}{H} = \underset{\sim}{X}\,\underset{\sim}{C}\,\underset{\sim}{X}'$

$$= \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'$$

The hat matrix "Puts the hat on $\underset{\sim}{Y}$"
H is square

Residuals. $\hat{\underset{\sim}{\epsilon}} = \underset{\sim}{Y} - \hat{\underset{\sim}{Y}}$ (cp. $\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta}$)

$$= (\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{Y}$$

Theorem. Under GM setup

a) $E\hat{\underset{\sim}{Y}} = \underset{\sim}{X}\underset{\sim}{\beta}$, $E\hat{\underset{\sim}{\epsilon}} = 0$

b) $\text{var}\,\hat{\underset{\sim}{Y}} = \underset{\sim}{H}\sigma^2$

$\text{var}\,\hat{\underset{\sim}{\epsilon}} = (\underset{\sim}{I} - \underset{\sim}{H})\sigma^2$ vs. $\text{var}\,\underset{\sim}{\epsilon} = \underset{\sim}{I}\sigma^2$

$\text{cov}\{\hat{\underset{\sim}{Y}}, \hat{\underset{\sim}{\epsilon}}\} = 0$

(II)

8 Sept 2001

Scholium i) $\underset{\sim}{H}\,\underset{\sim}{X} = \underset{\sim}{X}$

ii) $\underset{\sim}{H}^2 = \underset{\sim}{H}$ (idempotent)

iii) $r(\underset{\sim}{H}) = r = r(\underset{\sim}{X})$

iv) $r = \text{tr}(H)$

v) $(\underset{\sim}{I} - \underset{\sim}{H})^2 = \underset{\sim}{I} - \underset{\sim}{H}$ (idempotent)

vi) $r(\underset{\sim}{I} - \underset{\sim}{H}) = n - r$

Proof. i) $E\hat{\underset{\sim}{Y}} = \underset{\sim}{X}\,\underset{\sim}{\beta} = E\,\underset{\sim}{H}\,\underset{\sim}{Y} = \underset{\sim}{H}\,\underset{\sim}{X}\,\underset{\sim}{\beta} \quad \forall\, \underset{\sim}{\beta}$

ii) $\underset{\sim}{H}\,\underset{\sim}{X} = \underset{\sim}{X}$

$\underset{\sim}{H}^2 = \underset{\sim}{H}\,\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}' = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}' = \underset{\sim}{H}$

iii) $\underset{\sim}{H} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}'$, so $r(\underset{\sim}{H}) \le r(\underset{\sim}{X}) = r$

Also $\underset{\sim}{X} = \underset{\sim}{H}\,\underset{\sim}{X}$, so $r(\underset{\sim}{X}) \le r(\underset{\sim}{H})$

iv)

Sublemma. If $\underset{\sim}{A}$ is idempotent, $r(\underset{\sim}{A}) = \text{tr}(\underset{\sim}{A})$

Proof. $\underset{\sim}{A}$ is square (trace defined)

$\underset{m\times m}{\underset{\sim}{A}} = \underset{m\times r}{\underset{\sim}{U}}\;\underset{r\times r}{\underset{\sim}{\Lambda}}\;\underset{r\times m}{\underset{\sim}{V}'}$ and

Number of $\lambda_j > 0$ give the rank

$\underset{\sim}{A}^2 = \underset{\sim}{A}$

$\underset{\sim}{U}\,\underset{\sim}{\Lambda}\,\underset{\sim}{V}'\,\underset{\sim}{U}\,\underset{\sim}{\Lambda}\,\underset{\sim}{V}' = \underset{\sim}{U}\,\underset{\sim}{\Lambda}\,\underset{\sim}{V}'\underset{\sim}{U}\,\underset{\sim}{\Lambda}\,\underset{\sim}{V}'$

6 Sept 2001

$$\Lambda V' U \Lambda = \Lambda$$

$$B \Lambda = \Lambda$$

$$\sum_i B_{ii} \lambda_i = \lambda_i , \quad B_{ij} \lambda_j = 0 \quad i \neq j$$

$B_{ii} = 1$ for nonzero λ_i

$$tr(A) = tr(U \Lambda V') = tr(\Lambda V' U)$$

$$= tr(B) = r$$

v) $(I - H)^2 = I - 2H + H^2$

vi) $r(I - H) = tr(I - H) = n - r$

6 Sept. 2001

Back to the Theorem

Proof. a) $E\hat{Y} = X\beta$, as estimable

$E\hat{\epsilon} = E(Y - \hat{Y})$

b) Saw that $\text{var}\, p'\hat{\beta} = p' C p \sigma^2$

$\text{var}\, \hat{Y} = \text{var}\, X\hat{\beta}$

$= X C X' \sigma^2$

$= H \sigma^2$

(Also shows H is symmetric)

$\text{var}\,(\hat{\epsilon} = \text{var}(I - H)Y$

$= (I - H)\sigma^2 I (I - H)'$

$= (I - H)\sigma^2$

$\text{cov}\{\hat{Y}, \hat{\epsilon}\} = H \sigma^2 I (I - H)'$

$= 0$

Residuals are uncorrelated with the fitted values.

9 Sept. 01

Some review,

GM assumptions $\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$

$$E\underset{\sim}{\epsilon} = 0, \quad \text{var}\,\underset{\sim}{\epsilon} = \sigma^2 \underset{\sim}{I} = \text{var}\,\underset{\sim}{Y}$$

$\underset{\sim}{p}'\underset{\sim}{\beta}$ **estimable**

Normal equations

$$\underset{\sim}{X}'\underset{\sim}{X}\hat{\underset{\sim}{\beta}} = \underset{\sim}{X}'\underset{\sim}{Y}$$

BLUE (GM Theorem)

$$\underset{\sim}{p}'\hat{\underset{\sim}{\beta}} \qquad \text{(unique)}$$

Residuals

$$\hat{\underset{\sim}{\epsilon}} = \underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}} \qquad \text{(unique)}$$

$$\underset{\sim}{X} \perp \hat{\epsilon}, \quad \underset{\sim}{X}\hat{\underset{\sim}{\beta}} \perp \hat{\epsilon}$$

Minimum SS

$$\min_{\beta} \|\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta}\|^2 = \|\underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}}\|^2$$

Fitted values

$$\hat{\underset{\sim}{Y}} = \underset{\sim}{X}\hat{\underset{\sim}{\beta}} \qquad \text{(unique)}$$

$$\hat{\sigma}^2 = \hat{\underset{\sim}{\epsilon}}'\hat{\underset{\sim}{\epsilon}}/(n-r)$$

9 Sept. 01

Hat matrix

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' \qquad \text{idempotent}$$

$$\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}, \quad \hat{\boldsymbol{\epsilon}} = (\mathbf{I}-\mathbf{H})\mathbf{Y} = \mathbf{Y} - \hat{\mathbf{Y}}$$

$$\mathbf{Y} = \hat{\mathbf{Y}} + \hat{\boldsymbol{\epsilon}}$$

Means and variances

$$E\hat{\mathbf{Y}} = \mathbf{X}\boldsymbol{\beta} \qquad \text{var}\,\hat{\mathbf{Y}} = \mathbf{H}\sigma^2$$

$$E\hat{\boldsymbol{\epsilon}} = \mathbf{0} \qquad \text{var}\,\hat{\boldsymbol{\epsilon}} = (\mathbf{I}-\mathbf{H})\sigma^2 \quad \neq \mathbf{I}\sigma^2$$

$$\text{cov}\{\hat{\mathbf{Y}}, \hat{\boldsymbol{\epsilon}}\} = \mathbf{0}$$

215a — 10th Sept. 2001

Beginnings of world of residuals.

Lemma i) $\hat{Y}'\hat{\epsilon} = 0$ $\qquad \hat{\epsilon} \perp Y$

ii) $X'\hat{\epsilon} = 0$

iii) $\|Y\|^2 = \|\hat{Y}\|^2 + \|\hat{\epsilon}\|^2$ $\qquad$ Figure below

iv) $\|Y - X\beta\|^2 = \|Y - X\hat{\beta}\|^2 + \|X(\hat{\beta}-\beta)\|^2$

Proof. $\hat{Y} = X\hat{\beta}$, $\hat{\epsilon} = Y - \hat{Y} = Y - X\hat{\beta}$, $X'X\hat{\beta} = X'Y$

$\qquad = HY \qquad\qquad = (I - H)Y$

i) From normal equations

$$X'(Y - X\hat{\beta}) = 0 \quad \text{giving ii)}$$

$$\hat{Y}'\hat{\epsilon} = \hat{\beta}'X'(Y - X\hat{\beta}) = 0 \quad \text{giving i)}$$

iii) is iv) with $\beta = 0$

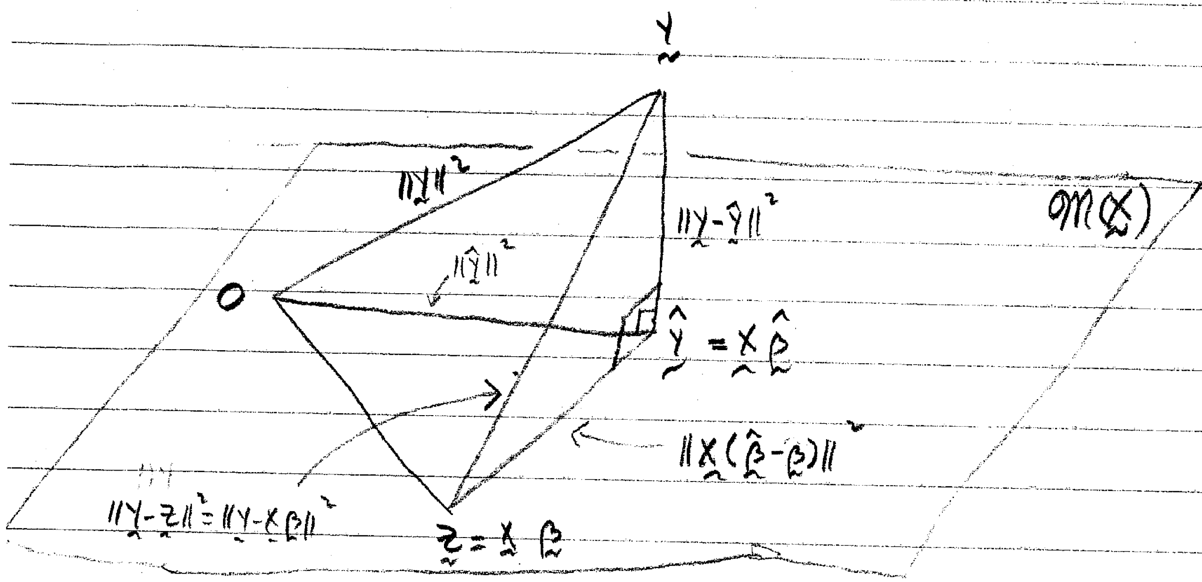
iv) $\|Y - X\beta\|^2 = \|Y - X\hat{\beta} + X(\hat{\beta}-\beta)\|^2$

$$= \|Y - X\hat{\beta}\|^2 + (Y - X\hat{\beta})'X(\hat{\beta}-\beta) + (\hat{\beta}-\beta)'\underbrace{X'(Y - X\hat{\beta})}_{0}$$

$$+ \|X(\hat{\beta}-\beta)\|^2$$

10 Sept. 2001

Display of iii), iv) and first ANOVA Table.



Want to minimize $\|Y - X\beta\|^2$

ANOVA Table.

$r = r(X)$

Source	SS	df	ESS
Regression	$\|\hat{Y}\|^2 = \|X\hat{\beta}\|^2$	r	$\beta'X'X\beta + r\sigma^2$
Error	$\|Y-\hat{Y}\|^2 = \|Y - X\hat{\beta}\|^2$	$n-r$	$(n-r)\sigma^2$
Total	$\|Y\|^2$	n	

Saw: $E\hat{Y} = X\beta$, $\text{var}\,\hat{Y} = H\sigma^2$, $\text{var}\,\hat{\epsilon} = (I-H)\sigma^2$

$$E\hat{\epsilon}'\hat{\epsilon} = tr\{(I-H)\sigma^2\} = (n-r)\sigma^2$$

$$E\hat{Y}'\hat{Y} = (E\hat{Y})'(E\hat{Y}) + tr\{H\sigma^2\} = \beta'X'X\beta + r\sigma^2$$

10 Sept. 2001

Estimation of σ^2

$$E\,\hat{\epsilon}'\hat{\epsilon} = (n-r)\sigma^2$$

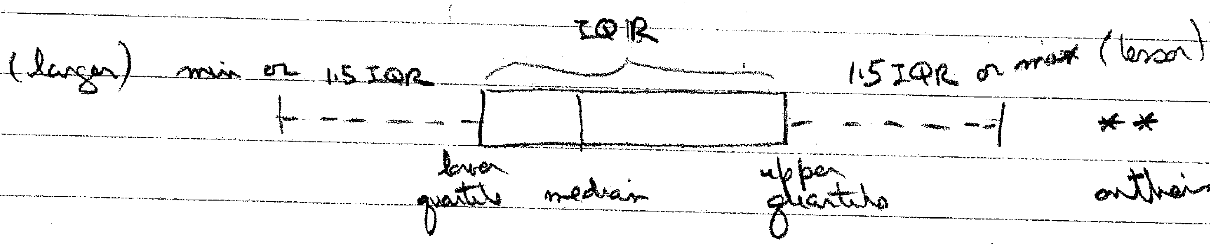
$$\hat{\sigma}^2 = \hat{\epsilon}'\hat{\epsilon}/(n-r) \qquad r = r(X)$$

Outlier. An observation strikingly far from some central value (eg. median)

Look for via examination of the $\hat{\epsilon}$

eg. histogram, boxplot, probplot (to come), indexplot

Boxplot



Need to think hard about outliers

10 Sept. 2001

$\operatorname{var} \hat{\underset{\sim}{\epsilon}} = (\underset{\sim}{I} - \underset{\sim}{H})\sigma^2$

Sometimes use <u>standardized residuals</u>

$$\hat{\epsilon}_i^s = \hat{\epsilon}_i \,/\, \hat{\sigma}\sqrt{1-H_{ii}}$$

Both residuals output of lm( )

11 Sept 2001

$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$ $\quad E(\epsilon) = 0 \quad \text{var}\,\epsilon = \sigma^2 \underset{\sim}{I}$

$X'X\hat{\beta} = X'Y$

$\hat{Y} = X\hat{\beta}$ <u>fitted values</u>

$\hat{\underset{\sim}{\epsilon}} = \underset{\sim}{Y} - \hat{\underset{\sim}{Y}}$ <u>residuals</u>

$= (I-H)Y \qquad H = X(X'X)^{-}X'$

$= (I-H)\epsilon \qquad H^2 = H$

<u>hat matrix</u>

$E\hat{\epsilon} = 0$

$\text{var}\,\hat{\epsilon} = (I-H)\sigma^2$

$s^2 = \hat{\sigma}^2 = \hat{\epsilon}'\hat{\epsilon}/(n-r) \qquad r = r(X)$

<u>Standardized residuals</u>

$\hat{\epsilon}_i^* = \hat{\epsilon}_i / s\sqrt{1-H_{ii}}$

H_{ii}: <u>leverage</u>

11 Sept 2001

Some useful plots

1. Prob plot of residuals
2. Response vs. explanatories — curvilinear? dependence?
3. Residuals vs explanatories in model — curvature?
4. " " " not in model — include?
5. " " predicted — wedging variance increases.
6. Residuals vs. index (cp. 4)

mother (24, 13)

11 Sept 2001

Probability plots

1. Normal

$u_1, \ldots, u_n$ IN (μ, σ^2)

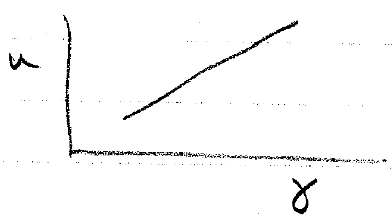
$u_{(1)} \le u_{(2)} \le \ldots \le u_{(n)}$

$u_{(i)} = \mu + \sigma z_{(i)}$

Plot $u_{(i)}$ vs. $\Phi^{-1}\left(\frac{i}{n+1}\right) = \gamma_i$

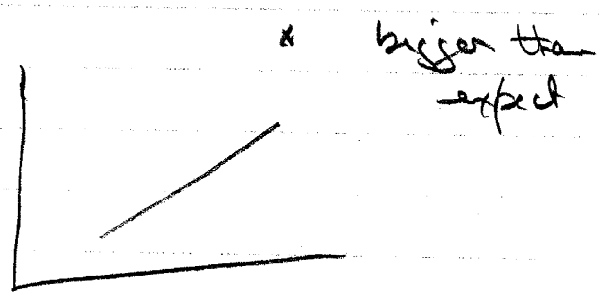
Possibilities include

1.



OK
Can estimate μ, σ
Do some simulations

2.



outlier

examine/drop

3+

17 Sept. 06

<u>Outlier</u> - an observation strikingly far from some central value (like a mean or median)

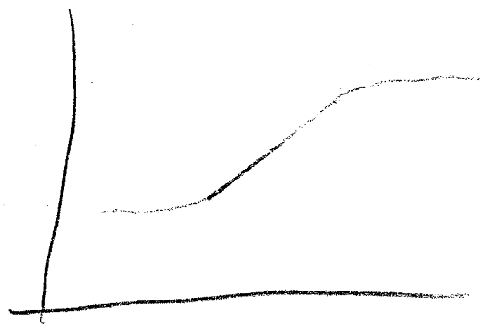
May not be influential

11 Sept 2001

3.

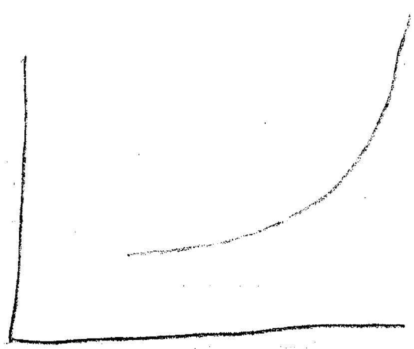
long-tailed
non-normal

4



short-tailed
non-normal

5.



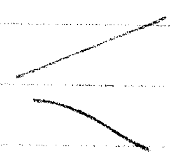
asymmetric,
non-normal

11 Sept 2001

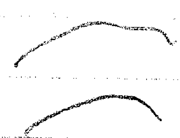
Regression:

Use $\hat{\underset{\sim}{\epsilon}}$ or $\hat{\underset{\sim}{\epsilon}}^{\wedge}$

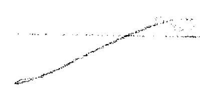
Residual plots:



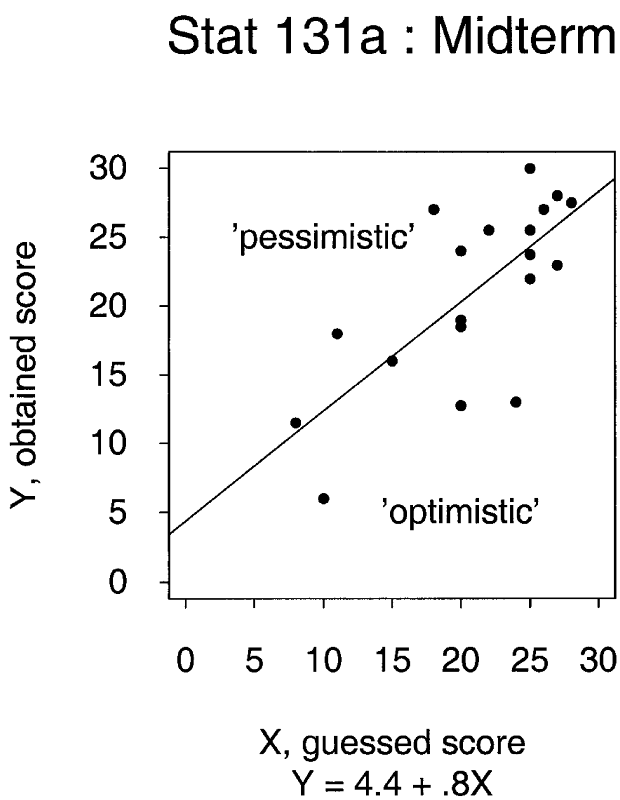
wedging

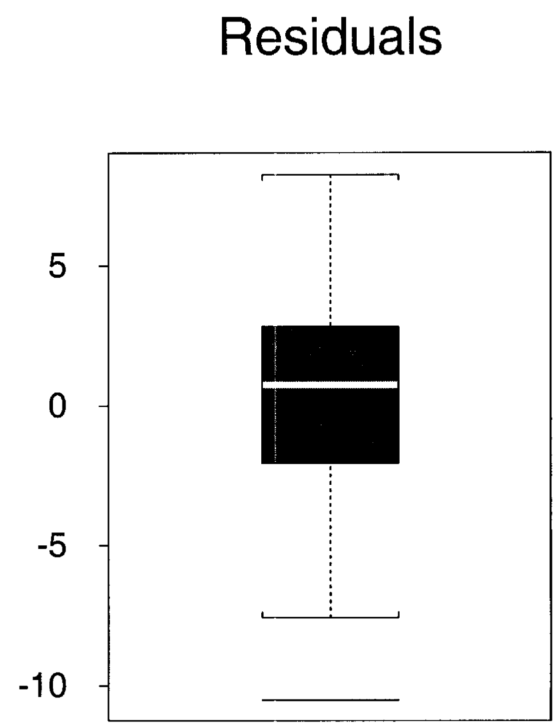


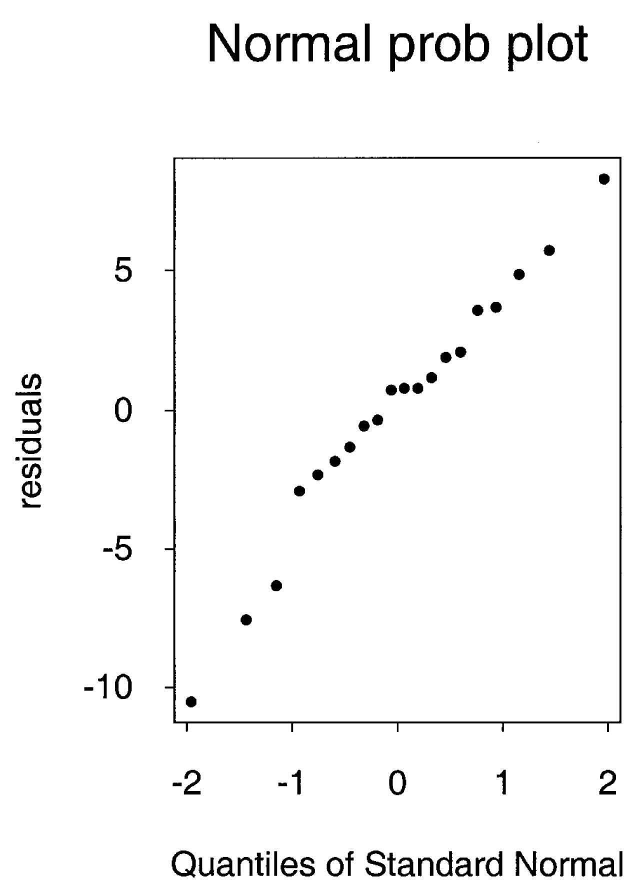
extra terms or transformation

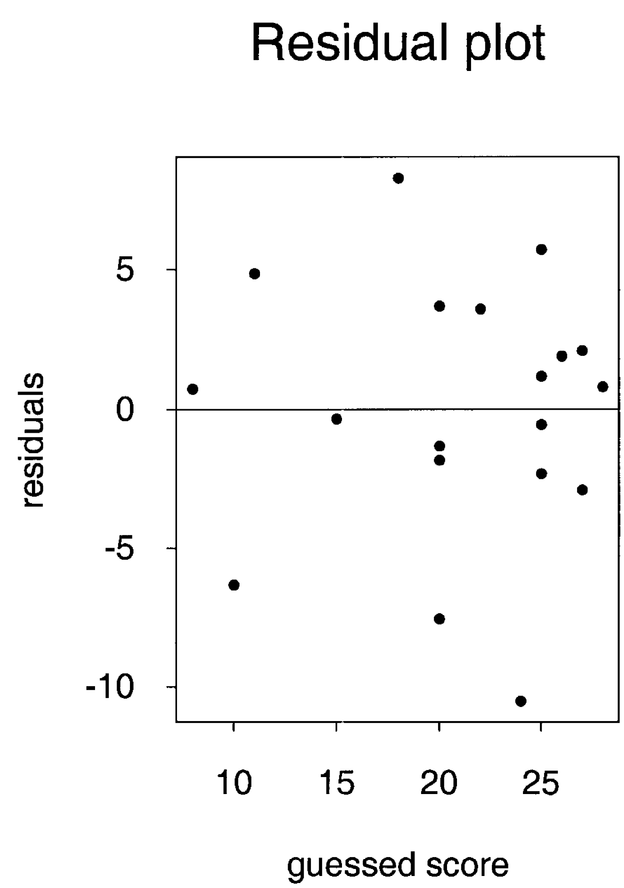


omitted variable


Stat 131a : Midterm
Y, obtained score
'pessimistic'
'optimistic'
0
5
10
15
20
25
30
X, guessed score
Y = 4.4 + .8X


Residuals
5
0
-5
-10


Normal prob plot
residuals
5
0
-5
-10
-2
-1
0
1
2
Quantiles of Standard Normal


Residual plot
residuals
5
0
-5
-10
10
15
20
25
guessed score