5) The classical estimator is the least squares (OLS) estimator $\hat{\beta}_{OLS} \equiv \hat{\underline{\beta}} = (X^TX)^{-1}X^T\underline{Y}$.

$\hat{\underline{\beta}}$ is the value of $\underline{b}$ that minimizes

$\sum_{i=1}^{n} r_i^2(\underline{b})$, the sum of squared residuals.

$r_i(\underline{b}) \equiv Y_i - \hat{Y}_i(\underline{b}) = Y_i - \underline{x}_i^T\underline{b}$.

6) p. Usually $x_1 \equiv 1$ so

$Y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i = \underline{x}_i^T\underline{\beta} + e_i$

For OLS, $\hat{\underline{\beta}} = (X^TX)^{-1}X^TY$ and

$\hat{\underline{Y}} = X\hat{\underline{\beta}} = H\underline{Y}$

7) p. computer output

label	Estimate or coef
constant	$\hat{\beta}_1$
x_2	$\hat{\beta}_2$
$\vdots$	
x_p	$\hat{\beta}_p$

R^2

$\hat{\sigma}$

n

$df = n-p$

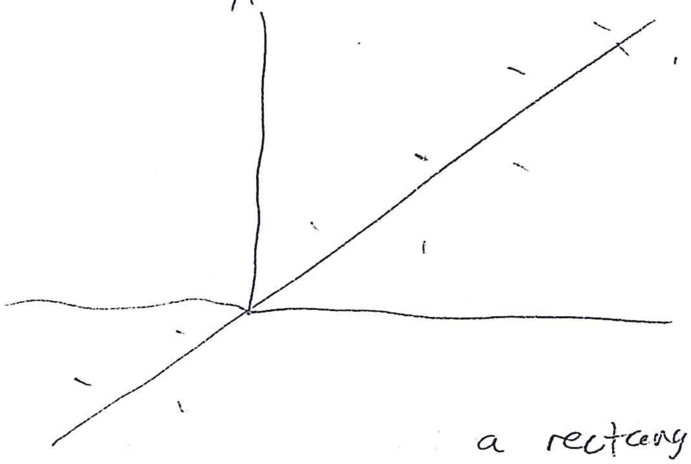
ANOVA Table

Source	df	SS	MS	F	pval
reg	$p-1$	SSR	MSR	$F_0 = \frac{MSR}{MSE}$	for H_0 $\beta_2 = \cdots = \beta_p = 0$
residual (or error)	$n-p$	SSE	MSE		

is on the model. 37.5

$R^2 = [\mathrm{corr}(\hat{Y}_i, Y_i)]^2$

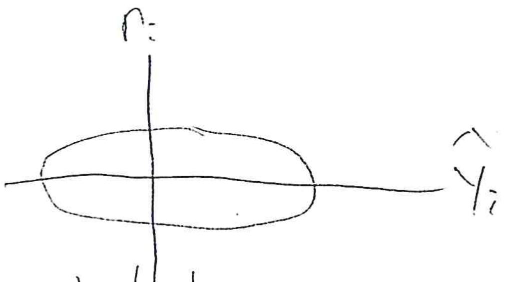
93) Know for exams. The response plot is a plot of $\hat{Y}_i = \underline{x}_i^T \hat{\underline{\beta}} = ESP$ vs Y_i



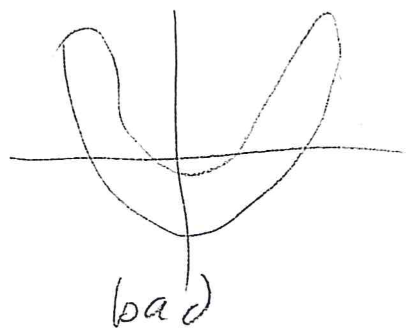
If the unimodal MLR model holds, the plotted points should scatter about the identity line in a rectangular or ellipsoidal band.

The vertical deviation $Y_i - \hat{Y}_i = \underbrace{r_i}_{i\text{th residual}}$.

MLR 0) Know for exams. The residual plot of $\hat{Y}_i$ vs r_i is used to check the lack of fit of the MLR model.

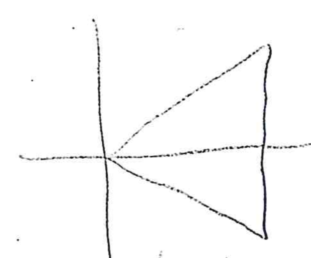


ideal box or ellipsoid centered about $r_i = 0$ unimodal MLR model



bad

x_i^2 needed



bad

σ^2 not constant

1st thing to check

bad σ^2 not constant

11) OLS CLT (central limit theorem).

Consider the MLR model $Y_i = \underline{x}_i^T \underline{\beta} + e_i$.

Assume the e_i are iid with $E(e_i) = 0$, $V(e_i) = \sigma^2$; and that $\max(h_1, \ldots, h_n) \xrightarrow{P} 0$ as $n \to \infty$ and

$\frac{X^T X}{n} \to W^{-1}$ as $n \to \infty$. Then

$$\sqrt{n}\,(\hat{\underline{\beta}} - \underline{\beta}) \xrightarrow{D} N_p(\underline{0}, \sigma^2 W) \text{ and}$$

$$(X^T X)^{\frac{1}{2}} (\hat{\underline{\beta}} - \underline{\beta}) \xrightarrow{D} N_p(0, \sigma^2 I_p).$$

12) $\hat{\sigma}^2 = MSE = \frac{1}{n-p} \sum_{i=1}^{n} r_i^2$

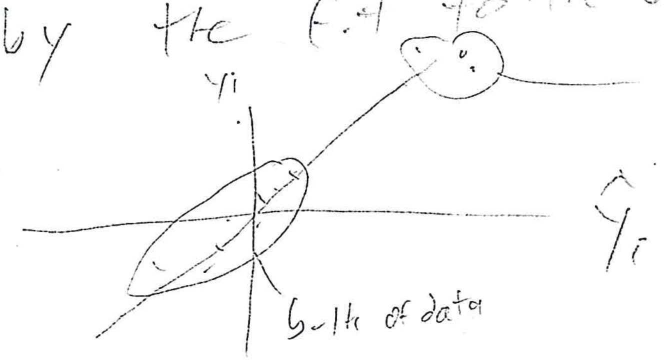
and $\hat{W} = n(X^T X)^{-1}$

$$\hat{\underline{\beta}} \sim AN_p\left[\underline{\beta},\ MSE\,(X^T X)^{-1}\right].$$

$$SE(\hat{\beta}_i) = \sqrt{MSE\,(X^T X)^{-1}_{ii}}\ , \quad i = 1, \ldots, p.$$

13) A leverage point has outlying $\underline{x}_i$.

A <u>good leverage point</u> is a case that is fit well by the fit to the bulk of the data.



good leverage points if $Y_i - \hat{Y}_i(\underline{b})$ is small

$\underline{b}$ OLS from cases in bulk of data

14) $Y_i = x_i^T \beta = \beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p$

$\hat{Y}_i(b) = x_i^T b$. So $\hat{Y}_i = \hat{Y}_i(\hat{\beta})$. The residual $r_i = Y_i - \hat{Y}_i$.

ex) $Y = \log(M)$, M = Mussel muscle mass

$x_2 = \log(H)$ H = ht of shell

$x_3 = \log(S)$, S = shell mass

$Y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + e$ MLR model

output

more table output

label	Est or coef
intercept constant	$-5.07459 = \hat{\beta}_1$
log(H)	$1.12399 = \hat{\beta}_3$
log(S)	$0.573167 = \hat{\beta}_3$

predict Y if $x_2 = 4$ and $x_3 = 5$

soln) $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 =$

$-5.07459 + 1.12399(4) + 0.573167(5)$

$= \boxed{2.2869}$

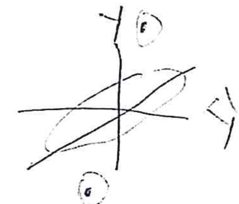
today's QUIZ

§5.6, 5.7

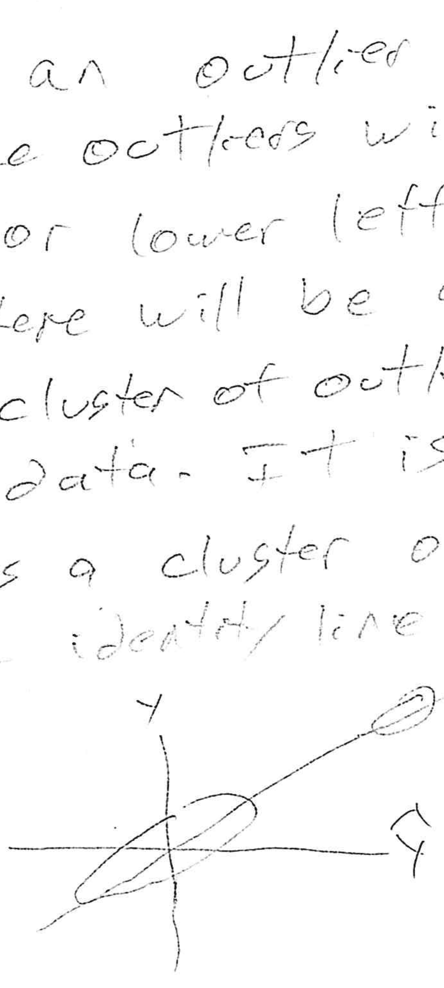
15) The response plot and residual plot are good for detecting outliers.

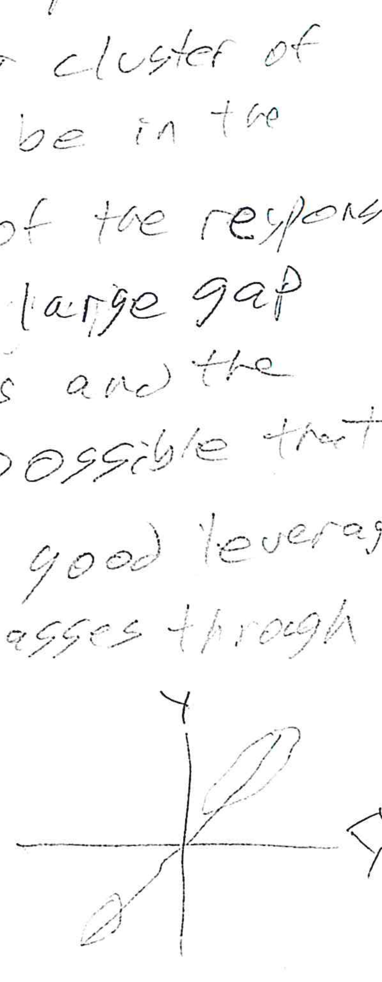
Zemen, Samuel G.

16) Beginners often label too many   R5&9 54)
points as outliers. Mentally draw a
box about the bulk of the data ignoring
any outliers (about the identity line
for the response plot and about the
horiz line r=0 for the residual plot.)
Double the width of the box. Alternatively
visually estimate the SD of the residuals
in both plots. Case more than 5 SD's
from the r=0 or identity line and cases
outside the doubled box are potential
outliers.

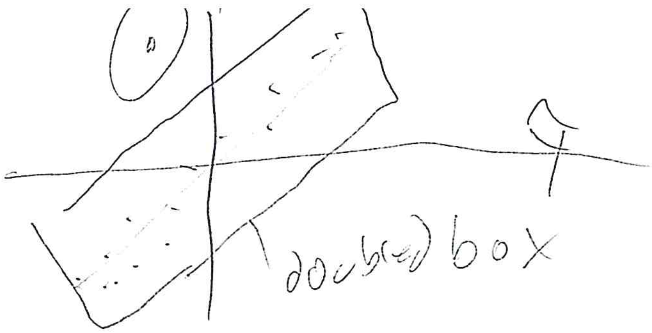


The identity line can also pass near
or through an outlier or cluster of
outliers. The outliers will be in the
upper right or lower left of the response
plot, and there will be a large gap
between the cluster of outliers and the
bulk of the data. It is possible that
the cluster is a cluster of good leverage
points if the identity line passes through
the cluster.





39.5



can get a plot like this if # outliers << n

see HW6 5.12

16) Robust and resistant MLR estimators should make $|r_i(\hat{\beta}_R)|$ large for certain outlier configurations, but tend not to work very well.

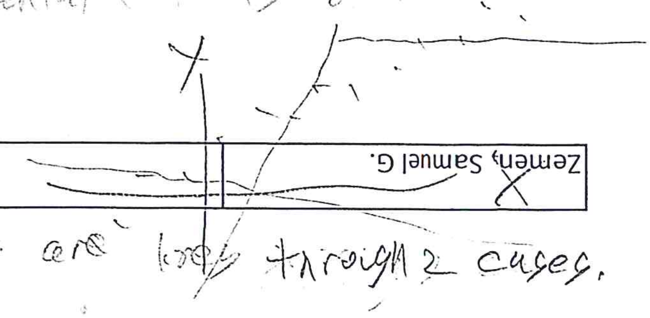
17) An MLR <u>elemental set</u> is a set J of p cases $(Y_i, \underline{x}_i)$ chosen without replacement. This set is just large enough to produce an estimator $\hat{\underline{\beta}}_J$ of $\underline{\beta}$.

$$\hat{\underline{\beta}}_J = (X_J^T X_J)^{-1} X_J^T \underline{Y}_J = X_J^{-1} \underset{\sim}{Y}_J$$

$p \times p$ so square

$X_J \hat{\underline{\beta}}_J = \underline{Y}_J$ so $\hat{\underline{Y}} = X \hat{\beta}_J$ is such that $\hat{Y}_i = Y_i$ for $i \in J$.

So the hyperplane for the elemental fit is determined by the p cases.



Zemen, Samuel G.

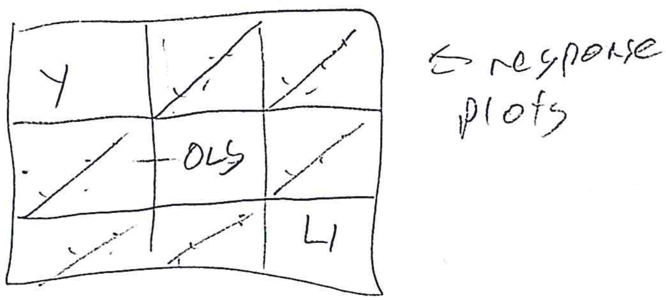
$p=2$ elemental hyperplanes are lines through 2 cases.

RSB3401

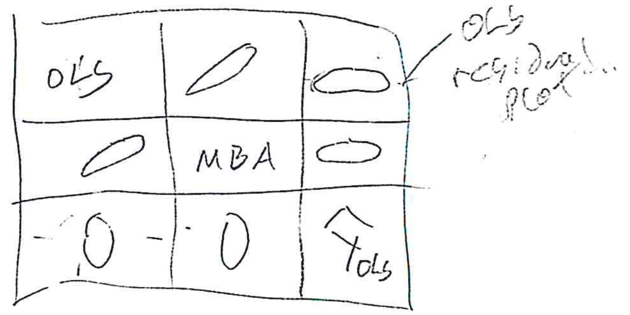
18) Let $\underline{b}_1, \ldots, \underline{b}_J$ be estimators of $\underline{\beta}$.

An FF plot is a scatterplot matrix of $\hat{\underset{\sim}{Y}}(\underline{b}_1), \ldots, \hat{\underline{Y}}(\underline{b}_J)$ with $\underline{Y}$ on the top or bottom row.

An RR plot is a scatterplot matrix of $\underline{r}(\underline{b}_1), \ldots, \underline{r}(\underline{b}_J)$ with $\hat{Y}_{OLS}$ on the top or bottom row.



FF plot



RR plot

19) Let $c_n = c \approx \frac{n}{2}$. The LMS(c) (= LQS(c)) criterion is $Q_{LMS}(\underline{b}) = r^2_{(c)}(\underline{b})$ where $r^2_{(1)} \le \cdots \le r^2_{(n)}$ are the ordered squared residuals. The LTS(c) criterion is

$$Q_{LTS}(\underline{b}) = \sum_{i=1}^{c} r^2_{(i)}.$$

The LTA(c) criterion is

$$Q_{LTA}(\underline{b}) = \sum_{i=1}^{c} |r(\underline{b})|_{(i)}$$

where $|r(\underline{b})|_{(i)}$ is the ith ordered absolute residual.

20) The least median of squares = LMS = LQS = least quantile of squares estimator is $\hat{\underline{\beta}}_{LMS}$. The least trimmed sum of squares LTS estimator

is $\hat{\beta}_{LTS}$. The least trimmed sum of absolute deviations LTA estimator is $\hat{\beta}_{LTA}$. These estimators correspond to the $\hat{\beta}_L \in \mathbb{R}^p$ that minimizes the corresponding criterion. 405

21) LMS → get chebyshev estimator for all $\binom{n}{c}$ subsets of size c

or $\binom{n}{p+1}$ subsets of size p+1

LTS → get OLS estimator "

LTA → get L_1 "

or all $\binom{n}{p}$ elemental sets.

LTA, LMS, LTS complexity $\geq O(n^p)$ and $O(n^4)$ complexity takes too long to compute.

22) LTS concentration algorithm. A start is an initial trial fit and an attractor is the fit after refinement. Let b_{0j} be the jth start. Compute the n residuals $r_i(b_{0j})$. Then b_{1j} is the OLS estimator computed for the c_n cases with the smallest squared residuals $r_i^2(b_{0j})$. Continue the iteration for k steps resulting in $b_{0j}, b_{1j}, \ldots, b_{kj}$ = jth attractor, $j = 1, \ldots, K$. The FLTS estimator uses the attractor that minimizes Q_{LTS}.

23) The elemental basic resampling algorithm uses K elemental starts = attractors $b_{01}, \ldots, b_{0K}$ is a special case.

24) Let $1 \le d_n \le n$ be the number of contaminated cases. Let W be the clean data and W_d^n be the contaminated data. The breakdown value for $\hat{\beta}$ is $\min\left(\frac{d_n}{n} : \sup_{W_d^n} \|\hat{\beta}(W_d^n)\| = \infty\right)$ where the sup is over all corrupted samples.

ex) OLS has breakdown value $= \frac{1}{n}$ and is zero breakdown.

25) Let $\gamma_n = \frac{d_n}{n}$. High breakdown regression estimators have $\gamma_n \to 0.5$ as $n \to \infty$ if the clean data W are in general position: any p clean cases give a unique estimate of β. Estimators are zero breakdown if $\gamma_n \to 0$, as $n \to \infty$.

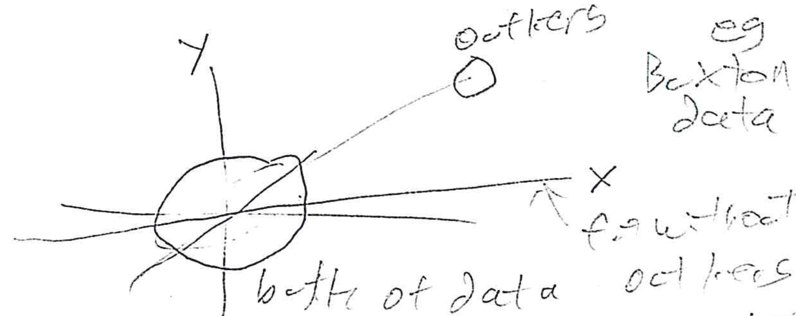
26) Need the number of randomly drawn elemental sets $K \to \infty$ to get a consistent estimator of β for the elemental basic resampling algorithm. You can't get a consistent estimator by applying an estimator to a fixed number of randomly drawn cases (Kp cases). The number of cases needs $\to \infty$.

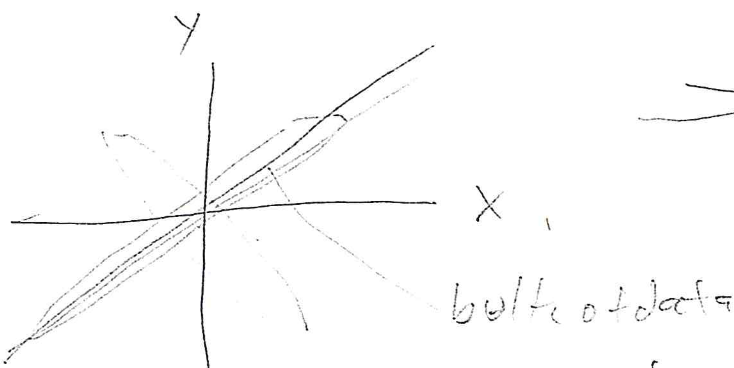
27) When n is odd $MED(n) = Y_{\left(\frac{n+1}{2}\right)}$ is from an elemental set and $MED(n)$ is $\sqrt{n}$ consistent for the pop median. But $MED(n)$ is not a randomly drawn elemental set; all n cases were used to select $Y_{\left(\frac{n+1}{2}\right)}$.

[illegible]
with $\bar{K}$ = 500 or 3000 are very good at detecting certain outlier configurations and even mixture distributions, with the LMS, LTA or LTS criterion. This outlier resistance decreases rapidly as p increases. 41.5

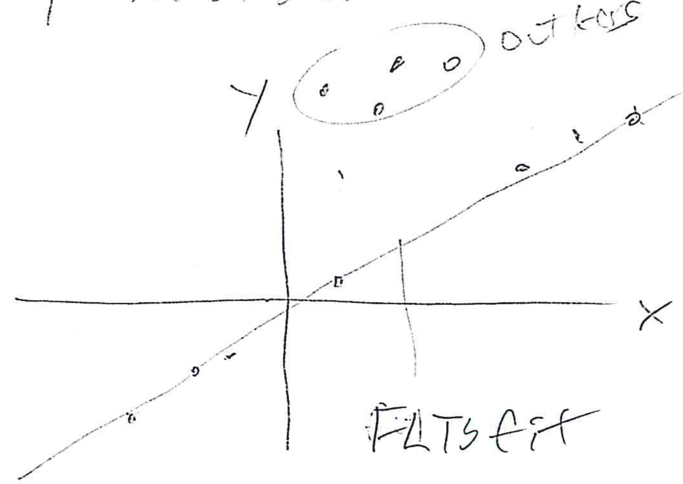
28) These algorithms have trouble if the MLR relationship for the bulk of the data is weak, and there is a tight cluster of outliers.

eg Buxton data





X w/ry tighter mixture dist



For $p=2$, ch6 estimators have trouble with data sets like this
eg Belgian telephone data

29) Masking occurs if the analysis suggests one or more outliers are good cases.
Swamping occurs if the analysis suggests one or more good cases are outliers.