3) Know! In a 1D regression model

Y is conditionally independent of $\underline{x}$ given $h(x)$ a single linear combination (often $h(x) = \underline{x}^T \underline{\beta}$) of the predictors, written $Y \perp\!\!\!\perp \underline{x} \mid \underline{\beta}^T \underline{x}$.

4) Know p 302 If $Y \perp\!\!\!\perp \underline{x} \mid \underline{\beta}^T \underline{x}$, then $Y \perp\!\!\!\perp \underline{x} \mid (a + c\underline{\beta}^T \underline{x})$ for any constants a and $c \neq 0$. The quantity $a + c\underline{\beta}^T \underline{x}$ is called a sufficient predictor (SP). A sufficient summary plot (SSP) is a plot of SP vs Y. An estimated sufficient predictor (ESP) is $\tilde{\alpha} + \tilde{\underline{\beta}}^T \underline{x}$ where $\tilde{\underline{\beta}}$ is an estimator of $c\underline{\beta}$ for some $c \neq 0$. An estimated sufficient summary plot (ESSP) or response plot is a plot of ESP vs Y.

5) * p 448 The most used statistical regression models are 1D models.

ex) $y = g(\alpha + \underline{\beta}^T \underline{x}, e)$

~~Single index model~~

$y = m(\alpha + \underline{\beta}^T \underline{x}) + e$ (60.5)

(often the function m is unknown)

ex) $y = \alpha + \underline{\beta}^T \underline{x} + e$ (MLR)

ex) $y = t^{-1}(\alpha + \underline{\beta}^T \underline{x} + e)$ (transformation model)

ex) logistic regression in ch 10

ex) poisson regression in ch 13

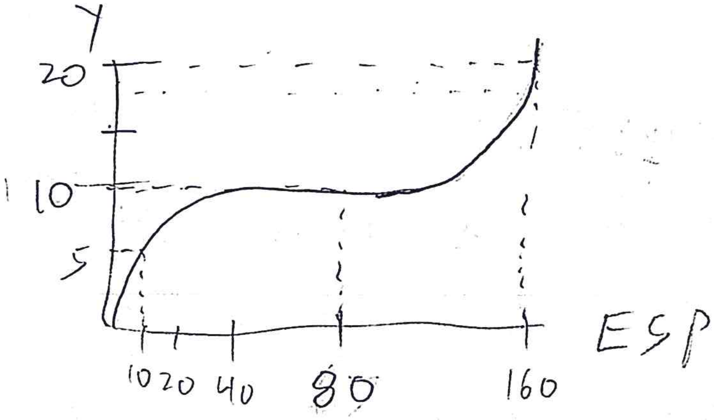
6) Any model with a single (nontrivial) predictor x $(p-1=1)$ is a 1D model. In this case x is both a SP and ESP and a plot of x vs y is both a SSP and a response plot (ESSP).

7) If $p-1 > 1$ the SP is unknown and an SSP can not be made (except for simulated data). For the MLR model $Y = \alpha + \underline{\beta}^T \underline{x} + e$ the ESP $= \hat{Y} = \hat{\alpha} + \hat{\underline{\beta}}^T \underline{x}$, and the ESSP = forward response plot = FY plot.

8) Know! The response plot is used to visualize the conditional distribution $Y|SP$ and should be made for any 1D regression.

9) Know predict Y for a given value of the ESP with "up and over lines".

ex)



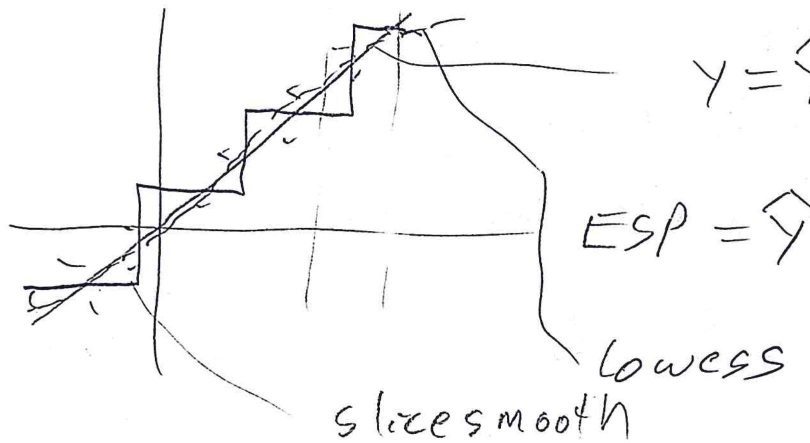
If ESP = 10, $\hat{Y} = 5$
80, $\hat{Y} = 10$
160, $\hat{Y} = 20$.

See HW11.

10) Adding an estimator of $\hat{E}(Y|ESP)$ to the plot can be useful.

Slice smooth divides the ESP into J "slices" containing roughly the same number of cases ($\frac{n}{J}$) and computes the sample mean of Y in each slice. The resulting step function is added to the plot. **Lowess** is similar but smooths the step function.

(a)

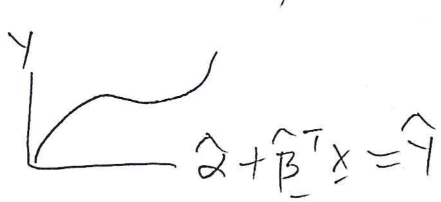


$Y = \hat{Y}$ = identity line
= parametric MLR estimator of mean function

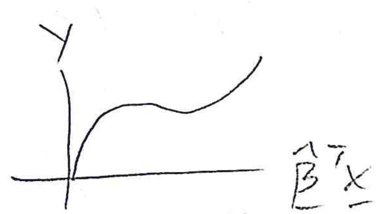
see HW A.

11) For simulated data, $E(Y|\underline{x})$ " = " $E(Y|SP)$
$\approx \hat{E}(Y|ESP)$ may be known.
Let $\hat{m}$ = lowess fit. Then a plot of $\hat{m}$ vs $E(Y|\underline{x})$ should be linear,
or $\hat{m}$ vs $\hat{E}(Y|ESP)$
See HW

12) ~~know~~ For 1D data $Y \perp\!\!\!\perp \underline{x} \mid \underline{\beta}^T\underline{x}$, the OLS estimator
$\hat{\alpha} + \hat{\underline{\beta}}^T\underline{x}$ obtained from the MLR regression of Y on $\underline{x}$ (even though the MLR model is not appropriate) is often such that $\hat{\underline{\beta}} \approx c\underline{\beta}$, $c \neq 0$. Hence $\hat{\underline{\beta}}^T\underline{x}$ (or $\hat{\alpha} + \hat{\underline{\beta}}^T\underline{x}$) is a good ESP.



response plot



response plot = OLS view

13) p 432 $Y \perp\!\!\!\perp \underline{x} \mid \underline{\beta}^T \underline{x}$

Suppose $\hat{\underset{\sim}{b}}$ is a consistent estimator of $\underline{\beta}_i^*$.
Then $\underline{\beta}_i^* = c_x \underline{\beta} + \underline{u}_g$ where the (population) bias vector $\underline{u}_g = \underline{\beta}_i^* - c_x \underline{\beta}$.

14) * p 432 Often if there are no strong nonlinearities present among the predictors, $\underset{\sim}{u_g}$ is "small" compared to $c_x \underline{\beta}$ and $\hat{\underline{b}}^T \underline{x}$ is a useful ESP.

15) Let $(\hat{\alpha}, \hat{\underline{\beta}})$ be the OLS estimator.
If $\underline{x} \sim EC$ and $Cov(\underline{x})$ is positive definite then $\hat{\underline{\beta}}$ is a consistent estimator of $c_x \underline{\beta}$ for some constant c_x and the (population) bias vector $\underline{u}_g = \underset{\sim}{0}$

16) p 435 Let $\begin{pmatrix} Y_i \\ \underline{x}_i \end{pmatrix}$ be iid, $Cov(\underline{x}) = \underset{(p-1)\times(p-1)}{\Sigma_x}$

$Cov(\underline{x}, y) = E\left[(\underline{x} - E(\underline{x}))\underbrace{(y - EY)}_{\text{scalar}} \right] = \underset{(p-1)\times 1}{\Sigma_{xy}}$.

The OLS estimator $(\hat{\alpha}, \hat{\underline{\beta}})$ estimates $(\alpha_{OLS}, \underline{\beta}_{OLS})$ where $\underline{\beta}_{OLS} = \Sigma_x^{-1} \Sigma_{xy}$.

$Y_i = m(\beta^T x_i) + e_i$ (6.2)

17) a) If $Y_i = m(\beta^T x_i) + e_i$

then $\beta_{OLS} = \Sigma_x^{-1} \Sigma_{xY} = c_{m,x}\, \beta + u_{m,x}$

where the scalar $c_{mx} = E[\beta^T (x - Ex)\, m(\beta^T x)]$

and the bias vector $\underset{\sim}{u}_{mx} = \Sigma_x^{-1} E[m(\beta^T x)\, r]$.

Here $r = x - E(x) - (\Sigma_x \beta)\, \beta^T (x - Ex)$.

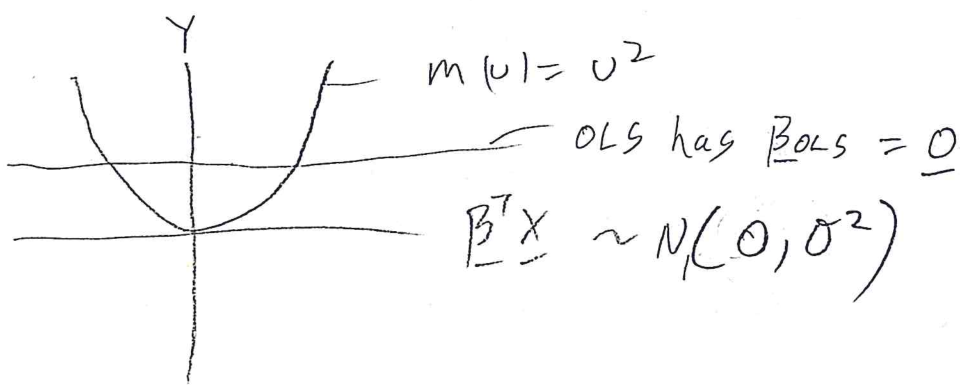
b) $u_{m,x} = 0$ if x is EC with 2nd moments.

$c_{mx} \neq 0$ unless $\mathrm{Cov}(x, Y) = 0$.

c) If the MLR model $Y_i = \alpha + x_i^T \beta + e_i$ holds, then $c_{mx} = 1$ and $u_{mx} = 0$,

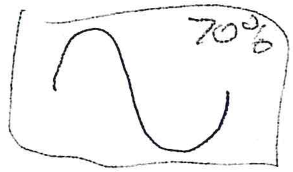
proof see HW Problem 9.2

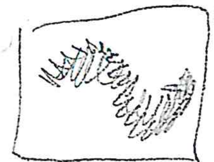
18) Typically $\mathrm{Cov}(x, Y) \neq \underset{\sim}{0}$ unless i) $\beta^T x$ follows a symmetric distribution and ii) m is symmetric about the median of $\beta^T x$.



p 4.2 19) the OLS view = OLS response plot is often useful if there are no strong nonlinearities among the predictors (make a DD plot and scatterplot matrix of $x_1, \ldots, x_{p-1}$ to check this assumption).

20) The <u>trimmed views estimator</u> plots $\hat{\beta}_M^T \underline{x}$ or $\hat{\alpha}_M + \hat{\beta}_M^T x$ vs Y but is otherwise identical to the TV MLR estimator that plots $\hat{\alpha}_M + \hat{\beta}^T x$ vs Y.





21) p 39 The trimmed views estimator uses ellipsoidal trimming. M% of the cases with the largest Mahalanobis distances are trimmed. Eg $(\hat{\alpha}_M, \hat{\beta}_M)$ is the OLS estimator applied to the cases not trimmed. A trimmed view is a plot of $\hat{\alpha}_M + \hat{\beta}_M^T \underline{x}$ vs Y for $M = 0, 10, \ldots, 90$. The best trimmed view $\hat{\alpha}_E + \hat{\beta}_E^T x$ vs Y corresponds to the view with a smooth mean function and the smallest variance function. The best trimmed view is used as the response plot.

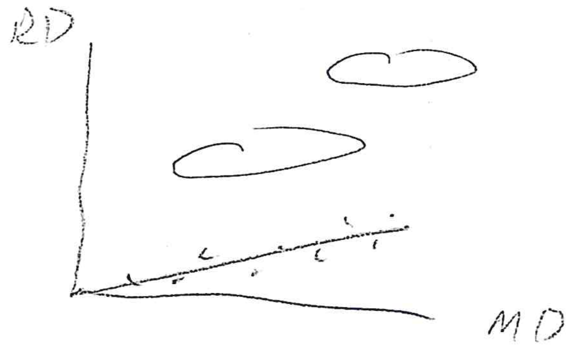
22) The tvreg estimator is a trimmed views estimator, the plot uses $\hat{\alpha}_M + \hat{\beta}_M^T x$ vs Y) and the identity line is

added to the plot. The tvreg estimator can be used to check whether an MLR model or some other single index model is appropriate.

23] p 437 Ellipsoidal trimming can result in a better response plot than the OLS view (0% trimming) for at least 3 reasons.

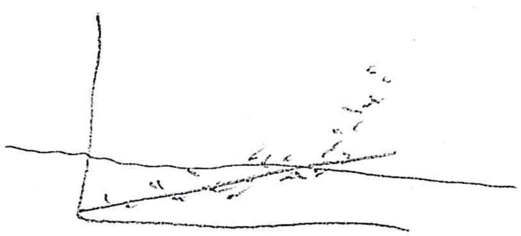
a) trimming often removes strong nonlinearities from the predictors and the distribution of the untrimmed cases is closer to EC than the distribution of all of the cases.

See Fig 11.3 for the effect of trimming on a scatterplot matrix.



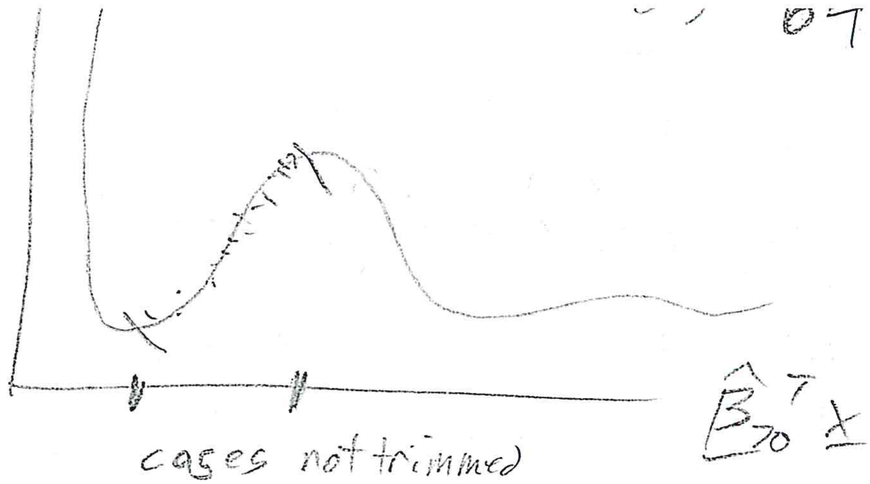
DD plot

get rid of the 2 clusters of outliers by trimming, then the remaining cases are much more EC



← not EC but trimming will cause the untrimmed cases to have greater clustering about the identity line than one gets by using all of the data.

b) Under heavy trimming the mean function of the untrimmed cases may be more linear than the mean function for the entire data set see fig 4.7 07



c) $\hat{\beta}_M = c_M \beta + \underset{\sim}{u}_M$ and for some M

$\|c_M \beta\|$ might be large compared to $\|\underset{\sim}{u}_M\|$.

2. b) The trimmed views estimator depends on the regression estimator (often MLR estimator) and the MLD estimator

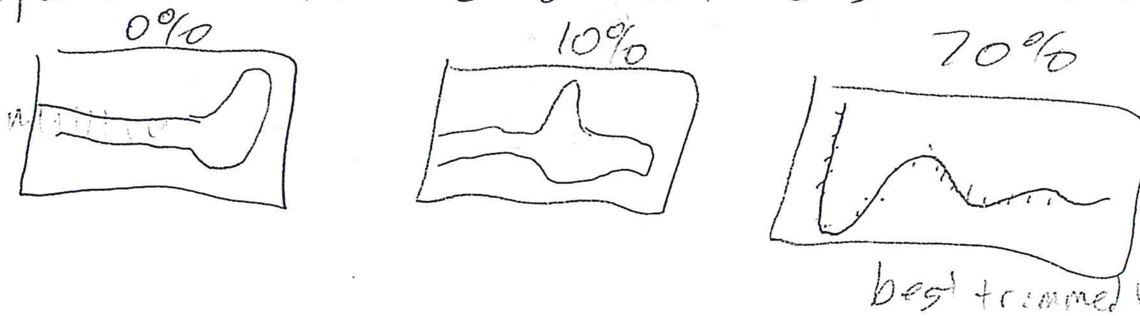
possible choices

reg	MLD
OLS	FCH RFCH
lmsreg	cov.mcd
sir	$(\bar{X}, S)$
	etc

See HW 1 for a preview.

As a project take a regression-MLD combo and compare it to the default OLS − FCH combo.

ex)



iid with pdf f and assume that the regression model $y_i = g(\underline{x}_i, \underline{\eta}, e_i)$ has a unique solution for e_i given by $e_i = h(\underline{x}_i, \underline{\eta}, y_i)$. Then the ith <u>residual</u> is $\hat{e}_i = h(\underline{x}_i, \hat{\underline{\eta}}, y_i)$ where $\hat{\underline{\eta}}$ is an estimator of $\underline{\eta}$.

Note) could have $\underline{\eta} = \mu$, $\underline{\eta} = \underline{\beta}$, $\underline{\eta} = \binom{\alpha}{\underline{\beta}}$ etc.

ex) $\eta = \binom{\alpha}{\underline{\beta}}$. $y = m(\alpha + \underline{\beta}^T \underline{x}) + e$

where m is known. Then

$$e = y - m(\alpha + \underline{\beta}^T \underline{x})$$

so $\hat{e}_i = y_i - m(\hat{\alpha} + \hat{\underline{\beta}}^T \underline{x}_i)$

is the residual for e_i.

If m was unknown, then $\hat{e}_i = y_i - \hat{m}(\hat{\alpha} + \hat{\underline{\beta}}^T \underline{x}_i)$ where $\hat{m}$ is an estimator of m (such as lowess).

See [illegible]

1) *p49 The <u>binomial regression model</u> states that $Y_1, \ldots, Y_n$ are independent binomial$(m_i, \rho(\underset{\sim}{x}_i))$. The <u>logistic regression</u> (LR) model is a special case with

$$\rho(\underline{x}) = \frac{\exp(\underline{\beta}^T \underline{x})}{1 + \exp(\underline{\beta}^T \underline{x})}.$$

The model is <u>binary</u> if $m_i \equiv 1$ for $i = 1, \ldots, n$.

2) <u>Know for E3</u>. Instead of predicting Y, predict $\rho(\underline{x}) = P(\text{"success"} \mid \underline{x}) = P(Y=1 \mid \underline{x})$

if Y is binary

$$\hat{\rho}(\underline{x}) = \frac{\exp(\hat{\underline{\beta}}^T \underline{x})}{1 + \exp(\hat{\underline{\beta}}^T \underline{x})} = \frac{\exp(ESP)}{1 + \exp(ESP)}.$$

see Quiz 11

ex] banknote data $Y = 0$ genuine $Y = 1$ counterfeit

label	est	SE	est/SE	pval
constant	−389.806	104.224	−3.740	.0002
bottom	2.26423	0.333233	6.795	.0000
left	2.83356	0.79601	3.562	.0004

predict $\hat{\rho}(\underline{x})$ if x_2 = bottom = 9.4 and x_3 = left = 130.1.

so $ESP = \hat\beta_1 + \hat\beta_2 x_2 + \hat\beta_3 x_3 =$ (6.2)

$-389.806 + 2.26423(9.4) + 2.83356(130.1)$

$= 0.123918$

so $\exp(ESP) = \exp(.123918) = 1.131923$

and $\hat\rho(\underline{x}) = \dfrac{\exp(ESP)}{1+\exp(ESP)} = \dfrac{1.131923}{1+1.131923} = 0.5309$

3] * p. 24 The <u>Poisson regression (PR) model</u> states that $Y_1, \ldots, Y_n$ are independent Poisson$(\mu(\underline{x}_i))$. The <u>loglinear regression</u> (LLR) model is the special case with $\mu(\underline{x}) = \exp(\underline{\beta}^T \underline{x})$.

4] <u>know for E3</u> Instead of predicting Y, predict $\mu(\underline{x})$ with $\hat{Y} = \hat\mu(\underline{x}) = \exp(\hat{\underline{\beta}}^T \underline{x}) = \exp(ESP)$

2x] Y = # of possums in area see Quiz 11

x_2 = bark x_3 = habitat x_4 = stags (hollow trees + 1)

label	est	stderr	Est/SE	pval
constant	−0.8994	.2135	−4.213	0.000
bark	0.0336	.0121	2.773	0.0055
habitat	0.1069	.0297	3.603	0.0003
stags	0.0302	.0094	3.210	0.0013

predict $\hat{\mu}(\underline{x})$ if $x_2 = 8.9$, $x_3 = 5.8$ and $x_4 = 8.2$

Soln $ESP = -0.4994 + 0.0336(8.9) + (.1069)(5.8) + (.0302)8.2$

$= \underline{0.2673}$

So $\hat{\mu}(\underline{x}) = \exp(ESP) = \exp(0.2673) = \boxed{1.3064}$

5) $ESP = \underline{x}^T \underline{\hat{\beta}}$ just like for MLR.

6) A parametric 1D regression model

is $Y|\underline{x} = Y \sim D(h(\underline{x}), \underset{\sim}{\gamma})$

$\leftarrow$ $q \times 1$ vector of parameters

where D is a parametric distribution.

7) A generalized linear model (GLM) and 1D generalized additive model (GAM) use D from a 1 parameter exponential family. The poisson(μ) and binomial(m, ρ) (with m known) are special cases. The GLM uses $SP = \underline{x}^T \underline{\beta} = h(\underline{x})$. The GAM uses $SP = AP = \alpha + \sum_{i=2}^{p} S_i(x_i)$.

9) PIs for Y_f given $\underline{x}_f$ and $(\underline{x}_1, Y_1), \ldots, (\underline{x}_n, Y_n)$:
use the parametric bootstrap;
Fit the model to get $Y|\underline{x} \sim D(\hat{h}(\underline{x}), \hat{\gamma})$.

Generate $Y_1^*, \ldots, Y_B^* \overset{iid}{\sim} D(\hat{h}(\underline{x}_f), \hat{\gamma})$.

Find the shorth of $Y_1^*, \ldots, Y_B^*$,

10) Variable selection is like that for MLR for PR and LR. Use AIC, or BIC instead of C_p. There are lasso estimators for β for PR and LR, forward selection and backward elimination too.

Parametric bootstrap $\underline{Y}^* = (Y_i^*)$,

$Y_i^* \sim D(\underline{x}_i^T \hat{\underline{\beta}}, \hat{\gamma})$, $i = 1, \ldots, n$. Do variable selection of $\underline{Y}^*$ on $\underline{X}$ to get $\hat{\underline{\beta}}_1^*$. Repeat to get the bootstrap sample $\hat{\underline{\beta}}_1^*, \ldots, \hat{\underline{\beta}}_B^*$.
The nonparametric bootstrap may work, too.

11) A response plot is a plot of ESP $= \hat{h}(\underline{x})$ vs Y.
GLM ESP $= x^T\hat{\underline{\beta}}$ vs Y
GAM ESP = EAP $= \hat{\alpha} + \sum_{j=2}^{p} \hat{S}_j(x_j)$.