Chapter 1 is abstract page #s from 2017 text Math 4841

1] p2 Regression is the study of the conditional distribution $Y|\underline{x}$ or $Y|\underline{\beta}^T\underline{x}$ of a response variable Y given a vector of predictors $\underline{x}$. In a 1D regression, Y is independent of $\underline{x}$ given a single linear combination of the predictors $SP = \underline{\beta}^T\underline{x}$, written $Y \perp\!\!\!\perp \underline{x} \mid SP$.

2] A linear model is $Y = \underline{\beta}^T\underline{x} + e$.

Multiple linear regression has at least one continuous predictor. In experimental design, all of the predictors are categorical.

3] p3 iid = independent and identically distributed

4] p3 response variable = dependent variable
predictor variables = carriers, covariates, explanatory variables or independent variables

5] skim or skip most of Ch 1.

Ch 2

1]* p17 The response variable Y is the variable that you want to predict. The predictor variables $x_1, \ldots, x_p$ are the

variables used to predict Y.

ex] Predict brain weight from size of head
Y X

2] Let $\underset{\sim}{x} = (x_1, \ldots, x_p)^T$. $\underset{\sim}{x}$ is a column vector, $\underset{\sim}{x}^T$ is a row vector.

3] p17 A quantitative variable takes on numerical values while a qualitative variable takes on categorical values.

ex] brain weight in grams is quantitative
sex = F or M is qualitative
0 or 1

4] p18 know Suppose the response variable Y is quantitative and at least one predictor variable x_i is, too. Then the multiple linear regression (MLR) model is

$$Y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + e_i$$
$$= \underset{\sim}{x}_i^T \underset{\sim}{\beta} + e_i \quad \text{for } i = 1, \ldots, n.$$

Here n is the sample size and e_i is a random variable called the ith error.

5] Notation Often the subscript i will be suppressed, especially when describing models. So the MLR model is

$$Y = \underset{\sim}{x}^T \underset{\sim}{\beta} + e.$$

6] * p 18 In matrix notation, the n equations become $\underset{n\times 1}{\underset{\sim}{Y}} = \underset{n\times p}{\underset{\sim}{\mathbf{X}}}\ \underset{p\times 1}{\underset{\sim}{\beta}} + \underset{n\times 1}{\underset{\sim}{e}}$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \overbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}}^{\mathbf{X}} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_p \end{bmatrix}$$

$$\underbrace{X_1 \quad X_2 \quad \cdots \quad X_p}_{\text{columns of } \mathbf{X}}$$

7] Often the 1st column of $\mathbf{X}$ is $X_1 = \underset{\sim}{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, the $n\times 1$ vector of 1's.

8] The ith <u>case</u> $(\underset{\sim}{x}_i^T, Y_i)$ corresponds to the ith row $\underset{\sim}{x}_i^T$ of $\mathbf{X}$ and the ith element Y_i of $\underset{\sim}{Y}$.

9] For the MLR model $Y = \underset{\sim}{x}^T \underset{\sim}{\beta} + e$,

Y and e are random variables but we only

have the observed values Y_i and x_i. (2.5

If the e_i are iid, $E(e_i)=0$, $V(e_i)=\sigma^2$, then the unknown parameters are $\underline{\beta} = (\beta_1, \ldots, \beta_p)^T$ and σ^2.

10] p18 The constant variance MLR model has $e_1, \ldots, e_n$ iid with $E(e_i)=0$ $V(e_i)=\sigma^2$ and the errors are independent of the predictor variables. Then the cases $(\underline{x}_i^T, Y_i)$ are independent for $i=1, \ldots, n$.

11] p19 know The unimodal MLR model has the same assumptions as 10], but also the iid errors come from a unimodal distribution that is not highly skewed.

12] p19 The normal (or Gaussian) MLR model adds the assumption that the e_i are iid $N(0, \sigma^2)$ to 10]. 12] is also a special case of 11]

Since the normal distribution is symmetric,

13] The normality assumption is too strong, but the MLR model works best if the error distribution is roughly symmetric and not too far from normal. Larger sample sizes n are needed if the error distribution is not roughly "mound shaped."

14] p19 Notation $A \equiv B = f(c)$ means A and B are equivalent and equal and $f(c)$ is the formula used to compute A and B.

15] p17 MLR is used to study the conditional distribution of $Y|\underset{\sim}{x}$ (Y given $\underset{\sim}{x}$), or $Y|\underset{\sim}{\beta}^T\underset{\sim}{x}$, to summarize the relationship between Y and $\underset{\sim}{x}$, and for prediction.

16] p19 Know Let $\underset{\sim}{b}$ be an estimate of $\underset{\sim}{\beta}$, then the ith fitted value

$$\hat{Y}_i \equiv \hat{Y}_i(\underset{\sim}{b}) = \underset{\sim}{x}_i^T \underset{\sim}{b} = x_{i1}b_1 + x_{i2}b_2 + \cdots + x_{ip}b_p.$$

The ith residual $r_i \equiv r_i(\underset{\sim}{b}) = Y_i - \hat{Y}_i(\underset{\sim}{b})$.

17] The (Ordinary) least squares (OLS) estimator $\hat{\underset{\sim}{\beta}} \equiv \hat{\underset{\sim}{\beta}}_{OLS}$ minimizes

$$Q_{OLS}(\underset{\sim}{b}) = \sum_{i=1}^{n} r_i^2(\underset{\sim}{b}) \quad \text{and} \quad \hat{\underset{\sim}{\beta}}_{OLS} = (\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T\underset{\sim}{Y}.$$

18] Know for final The ith fitted value

$$\hat{Y}_i \equiv \hat{Y}_i(\hat{\underset{\sim}{\beta}}) = \underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}} = x_{i1}\hat{\beta}_1 + \cdots + x_{ip}\hat{\beta}_p$$

and the ith residual $r_i = Y_i - \hat{Y}_i$.

19] We almost always use OLS for ch2&3 so r_i, $\hat{Y}_i$, $\hat{\underset{\sim}{\beta}}$ refer to OLS.

20] p19 Let $\hat{\underset{\sim}{Y}} = (\hat{Y}_1, \ldots, \hat{Y}_n)^T$. Then

$\hat{\underset{\sim}{Y}} = \underset{\sim}{X}\hat{\beta} = H\underset{\sim}{Y}$ where the hat matrix $H = \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T$ provided the inverse exists.

21] p20 An MLR model is linear in $\underset{\sim}{\beta}$. If there is an MLR model in Y and $\underset{\sim}{\beta}$, then $Y = \underset{\sim}{x}^T\underset{\sim}{\beta} + e$ or $Y = \underset{\sim}{w}^T\underset{\sim}{\beta} + \underset{\sim}{e}$;

the symbol for the vector $\underset{\sim}{x}$ or $\underset{\sim}{w}$ is not important.

22] If $Y = \underset{\sim}{x}^T\underset{\sim}{\beta} + e = x_1\beta_1 + \cdots + x_p\beta_p + e$,

then $\frac{\partial Y}{\partial \beta_i} = x_i$. If $Y = \underset{\sim}{w}^T\underset{\sim}{\beta} + e$

then $\frac{\partial Y}{\partial \beta_i} = w_i$.

ex] $Y_i = \beta_1 + \beta_2 w_i + \beta_3 w_i^2 + e_i$

is an MLR model in Y and $\underset{\sim}{\beta}$:
take $x_{i1} = 1$, $x_{i2} = w_i$, $x_{i3} = w_i^2$.

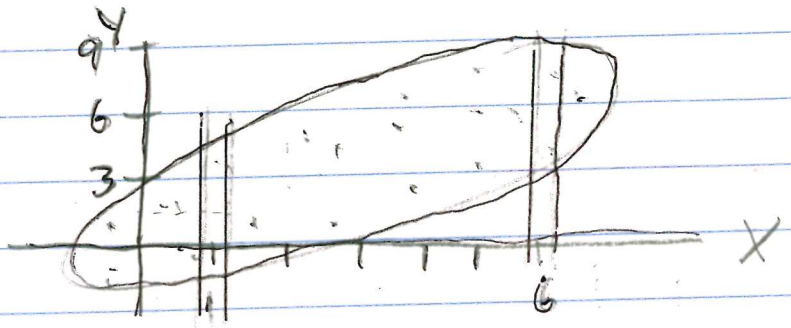
Then $Y_i = \underset{\sim}{x}_i^T\underset{\sim}{\beta} + e_i$.

ex] $Y_i = \frac{\beta_1}{1 + \beta_2 e^{\beta_3 x_i}} + e_i$ is not an MLR model. Note that $\frac{\partial Y}{\partial \beta_1} = \frac{1}{1 + \beta_2 e^{\beta_3 x_i}}$ depends on β_2 & β_3.

§2.2

23] p20 A scatterplot is a plot of X vs Y and is used to visualize the conditional distribution Y|X

ex



If X=1, $E(Y|X=1) \approx 1.5$ and $P(-2<Y<5|X=1)$ is high.
If X=6, $E(Y|X=6) \approx 4.5$ and $P(3<Y<9|X=6)$ is high but $P(-2<Y<5|X=6)$ is not high.

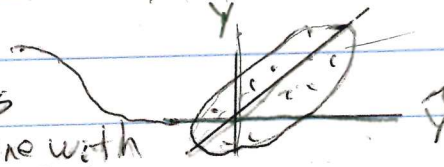
24] p21 know for final The response plot is a plot of $\hat{Y} = \underline{x}^T \hat{\underline{\beta}}$ vs Y.

It is used to visualize the MLR model in the background of the data if $\underline{x}^T\hat{\underline{\beta}}$ is a good estimator of $\underline{x}^T\underline{\beta}$ ($n \geq 5p$).

25] $Y = \hat{Y} + r = \hat{Y} + Y - \hat{Y}$

26] * Ignoring r, $Y = \hat{Y}$ is the identity line with unit slope and zero intercept.

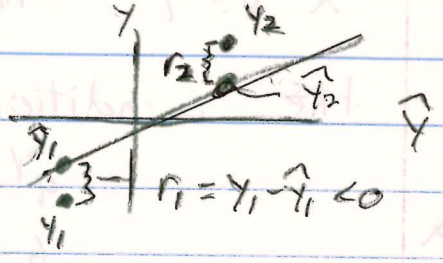
Ideal shape: plotted points scatter about the identity line with



response plot
no other pattern.

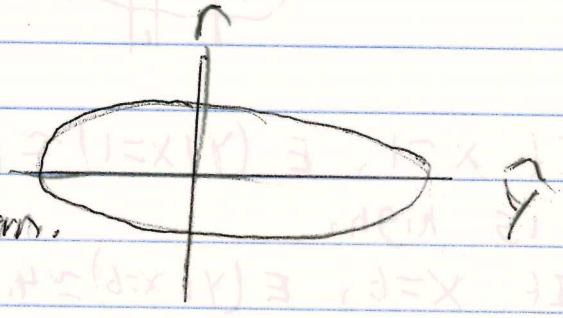
27] The vertical deviations from the identity line are the residuals

$Y_i - \hat{Y}_i = r_i$.



28] p21 Know for final A residual plot is a plot of $\hat{Y} = \underline{x}^T\hat{\beta}$ vs r

Ideal shape: plotted points scatter about the $r = 0$ line with no other pattern.



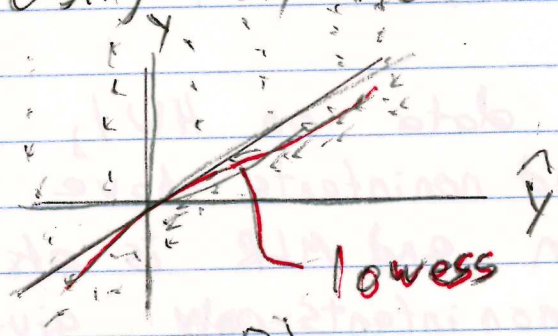
29] p21 Know for final For any MLR analysis, always make the response and residual plots.

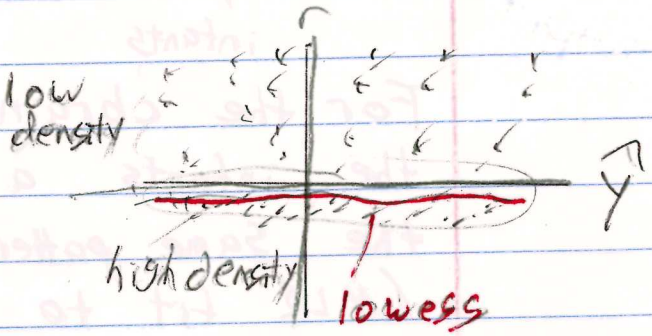
See Figures 1.2–1.4, 2.1, 2.3, 2.5a, 3.6, 3.7, 3.11

Remark: Plotting $\hat{Y}$ vs Y is a good idea for any model that predicts Y with $\hat{Y}$ MLR, DOE, time series, neural networks, lasso, Econometric models.

30] The two ideal shapes are for the iid unimodal error model and $n \geq 5p$. If the plots are near the ideal shape and $n \geq 5p$ then expect inference to be approximately correct except for prediction intervals.

31] If the response and residual plots suggest iid skewed errors and lowess added to the plot is close to the identity or r=0 lines, then except that much larger values of n ($n >> 5p$) may be needed before inference is approx correct. The scatterplot smoother lowess tries to estimate $E(V|\hat{Y})$ without using any model where $V=Y$ or $V=r$.

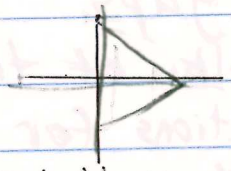


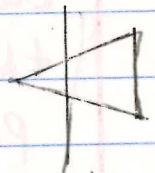


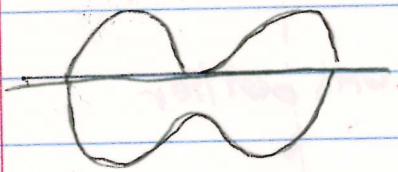
see HW2 D)

32] Do not look at a plot of r_i vs Y_i because Y_i and r_i are correlated and you could see a pattern.

33] Residual plots


left opening megaphone


right opening megaphone

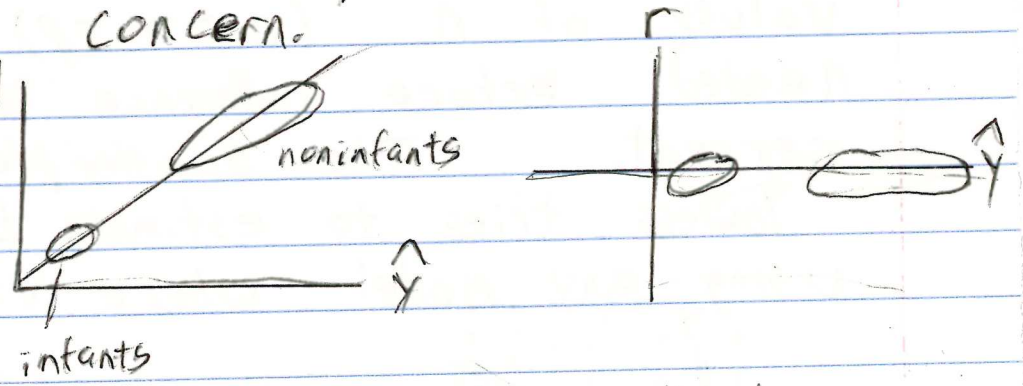
and  all suggest that the constant variance assumption should be checked.

34] A parabolic residual plot  suggests that one or more squared predictors X_i^2 needs to be added to the model.

Skip to Point 41 now to get $\hat{Y} = X^T\hat{\beta}$ from output

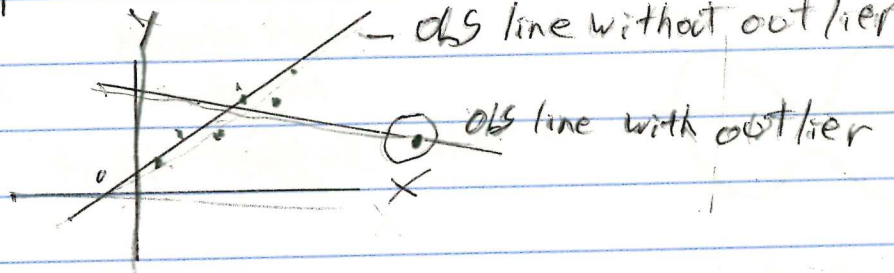
Read ex 2.3 carefully.

35] Gaps in the plots are cause for concern.

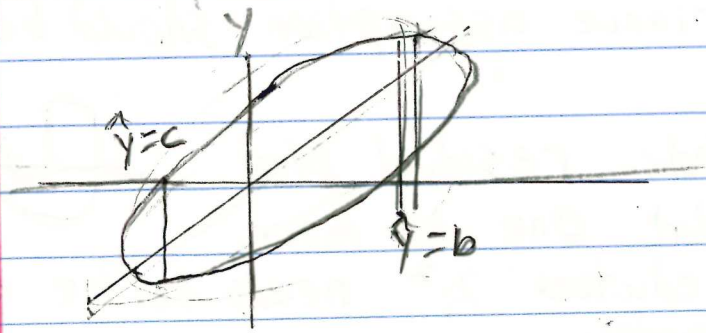


For the cbrain data on HW 1, the infants and nonintants have the same pattern and MLR is ok (MLR fit to nonintants only gives good predictions for intants).

But often outliers, observations far from the bulk of the data, cause gaps. Here the MLR fit to the bulk of the data gives poor predictions for outliers



36]



$\hat{Y} = \underset{\sim}{x}^T\hat{\beta} \approx \underset{\sim}{x}^T\underset{\sim}{\beta}$

For $\underset{\sim}{x}$ such that $\hat{Y} = \underset{\sim}{x}^T\hat{\underset{\sim}{\beta}} = b \approx \underset{\sim}{x}^T\underset{\sim}{\beta}$,

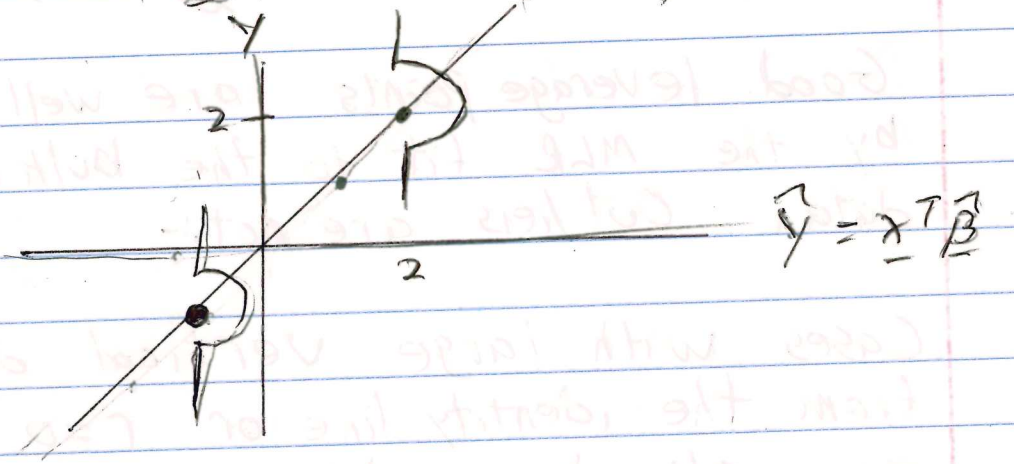
$E(Y|\underset{\sim}{x}^T\underset{\sim}{\beta}) \approx E(Y|\underset{\sim}{x}^T\hat{\underset{\sim}{\beta}}) \approx \underset{\sim}{x}^T\hat{\underset{\sim}{\beta}} = \hat{Y} = b.$

So $\underset{\sim}{x}$ with $\hat{Y} = 10 \Rightarrow E(Y|\underset{\sim}{x}^T\underset{\sim}{\beta}) \approx E(Y|\underset{\sim}{x}^T\hat{\underset{\sim}{\beta}}) = 10$

$\hat{Y} = -5 \Rightarrow E(Y|\underset{\sim}{x}^T\underset{\sim}{\beta}) \approx E(Y|\underset{\sim}{x}^T\hat{\underset{\sim}{\beta}}) = 5$

3.7) The iid normal MLR model is especially simple: $Y|\underset{\sim}{x}^T\underset{\sim}{\beta} \sim N(\underset{\sim}{x}^T\underset{\sim}{\beta}, \sigma^2)$

and $Y|\underset{\sim}{x}^T\underset{\sim}{\beta} \approx N(\underset{\sim}{x}^T\hat{\underset{\sim}{\beta}}, \hat{\sigma}^2) = N(\hat{Y}, MSE)$



$Y|\underset{\sim}{x}^T\hat{\underset{\sim}{\beta}} = 2 \approx N(2, \hat{\sigma}^2)$

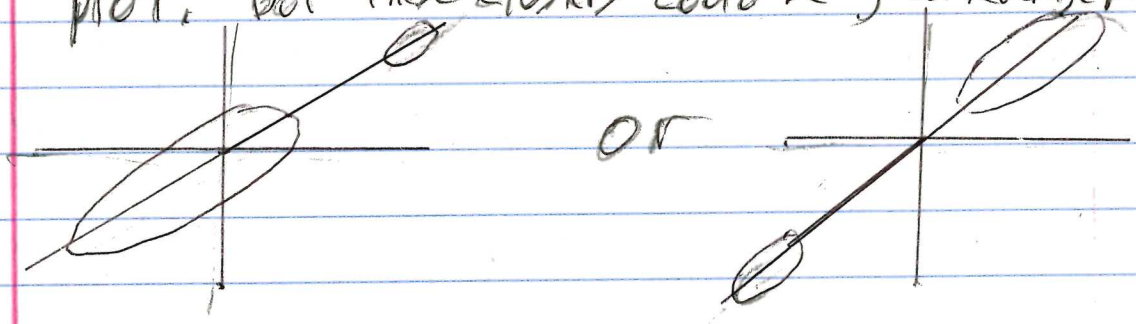
$Y|\underset{\sim}{x}^T\hat{\underset{\sim}{\beta}} = -1 \approx N(-1, \hat{\sigma}^2)$

$Y|\underset{\sim}{x}^T\hat{\underset{\sim}{\beta}} = a \approx N(a, \hat{\sigma}^2)$ for any a in the range of $\hat{Y}$.

Think of the picture as a 3 dimensional tunnel: the road is $\hat{E}(Y|\underset{\sim}{x}^T\underset{\sim}{\beta}) = \underset{\sim}{x}^T\hat{\underset{\sim}{\beta}} = \hat{Y}$ and the different normal curves are $\approx N(\underset{\sim}{x}^T\hat{\underset{\sim}{\beta}}, \hat{\sigma}^2)$

(Replace $\hat{\underset{\sim}{\beta}}$ by $\underset{\sim}{\beta}$, $\underset{\sim}{x}^T\hat{\underset{\sim}{\beta}}$ by $\underset{\sim}{x}^T\underset{\sim}{\beta}$, $\hat{\sigma}^2$ by σ^2 gives a similar picture)

38] p22-23 Outliers lie far from the bulk of the data. If there are gaps in the plots, outliers tend to be in the upper right or lower left corner of the response plot. But these clusters could be "good leverage points"

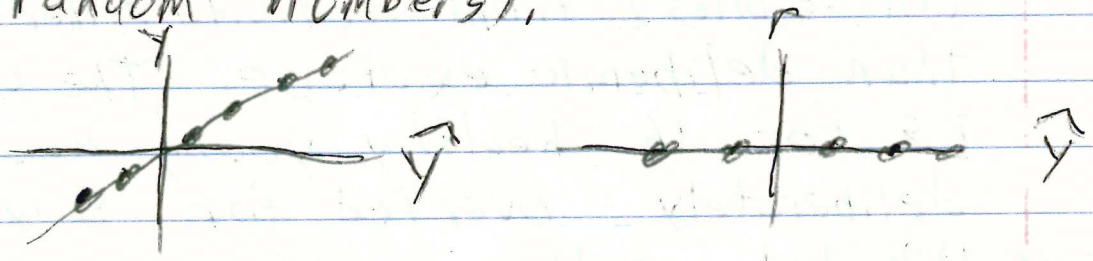


Good leverage points are well predicted by the MLR fit to the bulk of the data. Outliers are not.

Cases with large vertical deviations from the identity line or r=0 line may also be outliers. But beginners tend to label too many cases as outliers. Mentally put a box about the bulk of the data ignoring any potential outliers. Then double the width (about the identity line or r=0 line). Cases outside the doubled box are outliers. Or mentally estimate the standard deviation of the residuals in both plots. Look for cases with residuals more than 5 standard deviations from the r=0 line.

39] Want $n \geq 5p$. If $n < 5p$, try to get rid of unnecessary predictors. If this is not possible, do not expect your model to work well for inferences, including prediction.

extreme case: $n = p$ then $\hat{Y} = H Y = I_n Y = Y$. So plotted points fall exactly on the identity line and $r = 0$ line regardless of the value of the predictors (the $x_2, \ldots, x_p$ could be random numbers).



40] A common mistake is to use an MLR model where $n \approx p$ or $n = 2p$ or $n < 5p$. These models "fit noise" and do not work well.

Do 41 – ex's before 35]

41] computer output p 42

label or predictor	estimate or coef	stderr	tval	pval
intercept or constant	$\hat{\beta}_1$			
x_2	$\hat{\beta}_2$			
x_3	$\hat{\beta}_3$			
⋮	⋮			
x_p	$\hat{\beta}_p$			

model contains constant unless told otherwise

42] Know for final Assume $Y = \beta_1 + \beta_2 x_2 + \cdots + \beta_p x_p + e$.

Given x, find $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$

or given $x_2, \ldots, x_p$

find $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p$

ex]

label	coef	stderr	tvalue	pvalue
constant	3.5051	.3036	11.54	0.0
nea	-0.00344	0.000741	-4.64	0.0

Suppose it is desired to predict weight gain (in kg) from NEA change, the change in energy use (in calories) from activity other than deliberate exercise. The output is for 16 healthy young adults deliberately overfed for 8 weeks.

a) Find regeq. b) Predict weight gain if nea = 485.

soln a) $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x = 3.5051 - 0.00344 x$

b) $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x = 3.5051 - 0.00344(485)$

$= \boxed{1.8367}$

ex) n = 36 gestation time, days; body wt kg; brain wt g; Y = log brain wt for different animals

label	coef	stderr	tvalue	pvalue
constant	-0.457			
log(body)	0.551			
log(gest)	0.668			

allomet.isp

Predict Y = log(brain wt) if log(body) = 2 & log(gest) = 5

$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 = -0.457 + .551(2) + .668(5) = \boxed{3.985}$