Unit - II

Residual analysis :

① Two the error analysis for the following data

x	1	2	2.5	4	5	5.5	6	7	7.5
y	2	4.5	3	6.5	6	10	9	11	5

Calculate R^2 and Σe_i^2 for the given data and also the same without considering the last value of x and y.

Sol: (with value)

x	y	$(x-\bar{x})$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
1	2	-3.5	12.25	-4.33	18.7489	15.155
2	4.5	-2.5	6.25	-1.83	3.3489	4.575
2.5	3	-2	4	-3.33	11.0889	6.66
4	6.5	-0.5	0.25	0.17	0.0289	-0.085
5	6	0.5	0.25	-0.33	0.1089	-0.165
5.5	10	1	1	3.67	13.4689	3.67
6	9	1.5	2.25	2.67	7.1289	4.005
7	11	2.5	6.25	4.67	21.8089	11.675
7.5	5	3	9	-1.33	1.7689	-3.99
40.5	57		41.5		77.5001	41.5

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{40.5}{9} = 4.5$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{57}{9} = 6.333$$

$$r = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma (x_i-\bar{x})^2}\sqrt{\Sigma (y_i-\bar{y})^2}} = \frac{41.5}{\sqrt{41.5}\sqrt{77.5001}} = \frac{41.5}{56.71202}$$

$$r = 0.73176$$

$$r^2 = 0.53548$$

(without last value)

$x = 33, \quad y = 52$

(2)

$$\bar{x} = \frac{\sum x_i}{n} = \frac{33}{8} = 4.125$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{52}{8} = 6.5$$

x	y	$(x-\bar{x})$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
1	2	-3.125	9.7656	-4.5	20.25	14.0625
2	4.5	-2.125	4.5156	-2	4	4.25
2.5	3	-1.625	2.6406	-3.5	12.25	5.6875
4	6.8	-0.125	0.0156	0	0	0
5	6	0.875	0.7656	-0.5	0.25	-0.4375
5.5	10	1.375	1.8906	3.5	12.25	4.8125
6	9	1.875	3.5156	2.5	6.25	4.6875
7	11	2.875	8.2656	4.5	20.25	12.9375
			31.3748		75.5	4.6

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}} = \frac{46}{\sqrt{31.3748}\sqrt{75.5}} = \frac{46}{48.67026}$$

$$r = 0.9451$$

$$r^2 = 0.89328.$$

calculate $\sum e_i^2$:

x	y	x^2	y^2	xy
1	2	1	4	2
2	4.5	4	20.25	9
2.5	3	6.25	9	7.5
4	6.5	16	42.25	26
5	6	25	36	30
5.5	10	30.25	100	55
6	9	36	81	54
7	11	49	121	77
7.5	5	56.25	25	37.5
40.5	57	223.75	438.5	298

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - (\sum y_i)(\sum_{i=1}^{n} x_i)/n}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i^2)}{n}}$$

$$= \frac{298 - (57)(40.5)/9}{223.75 - \frac{(40.5)^2}{9}}$$

$$= \frac{298 - 256.5}{223.75 - 182.25} = \frac{41.5}{41.5}$$

$$\boxed{\hat{\beta}_1 = 1}$$

$\bar{x} = 4.5$, $\bar{y} = 6.33$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$= 6.33 - 1(4.5) = 6.33 - 4.5$

(3)

$\hat{\beta}_0 = 1.83$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

$\hat{y}_1 = 1.83 + 1(1) = 2.83$

$\hat{y}_2 = 1.83 + 1(2) = 3.83$

$\hat{y}_3 = 1.83 + 1(2.5) = 4.33$

$\hat{y}_4 = 1.83 + 1(4) = 5.83$

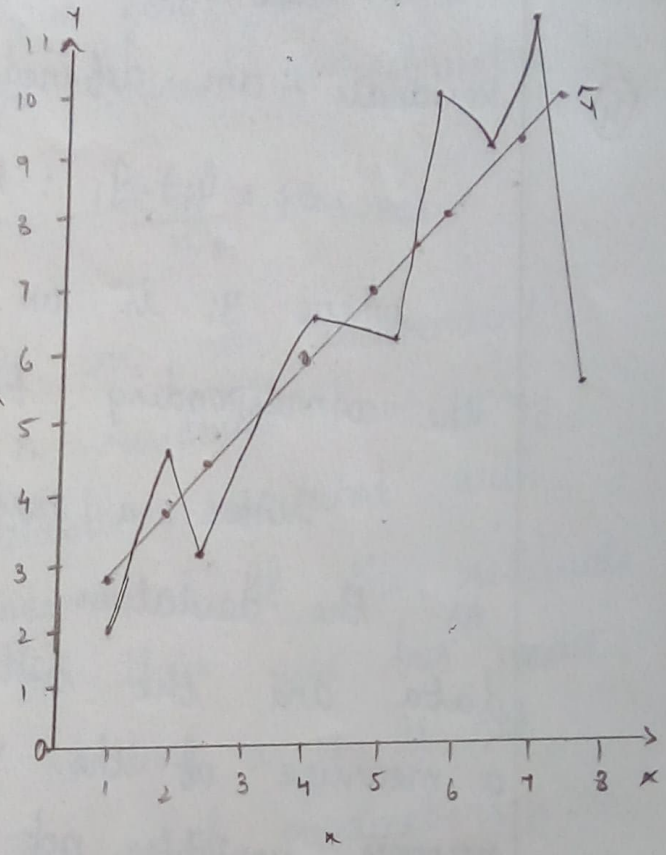
$\hat{y}_5 = 1.83 + 1(5) = 6.83$

$\hat{y}_6 = 1.83 + 1(5.5) = 7.33$

$\hat{y}_7 = 1.83 + 1(6) = 7.83$

$\hat{y}_8 = 1.83 + 1(7) = 8.83$

$\hat{y}_9 = 1.83 + 1(7.5) = 9.33$



Error :

	ϵ_i^2
$\epsilon_1 = y_1 - \hat{y}_1 = 2 - 2.83 = -0.83$	0.6889
$\epsilon_2 = y_2 - \hat{y}_2 = 4.5 - 3.83 = 0.67$	0.4489
$\epsilon_3 = y_3 - \hat{y}_3 = 3 - 4.33 = -1.33$	1.7689
$\epsilon_4 = y_4 - \hat{y}_4 = 6.5 - 5.83 = 0.67$	0.4489
$\epsilon_5 = y_5 - \hat{y}_5 = 6 - 6.83 = -0.83$	0.6889
$\epsilon_6 = y_6 - \hat{y}_6 = 10 - 7.33 = 2.67$	7.1289
$\epsilon_7 = y_7 - \hat{y}_7 = 9 - 7.83 = 1.17$	1.3689
$\epsilon_8 = y_8 - \hat{y}_8 = 11 - 8.83 = 2.17$	4.7089
$\epsilon_9 = y_9 - \hat{y}_9 = 5 - 9.33 = -4.33$	18.7489
0.03	56.0001

Definition of Residual :

(4)

We have already known that the Residuals are defined as,

$$e_i = y_i - \hat{y}_i \; ; \; i = 1, 2 \cdots n \longrightarrow (1)$$

where y_i is an observation and $\hat{y}_i$ is the corresponding fitted value.

Since, a residual may be viewed as the deviation between the observed data and the fitted data, It is also a measure of the variability in the response variable not explained by the regression model. It is also convinent to think of the residuals as the observed values of the assumptions on the errors should show up in the residuals.

Analysis of residuals is an effective way to discover several types of model in adequacies. plotting residuals is a very effective way to investigate how will the regression model fit the data and also to check the assumptions are regression model.

The residuals have several important properties. They have zero mean and their approximate average variance is estimated by.

$$\frac{\sum (e_i - \bar{e})^2}{n-p} = \frac{\sum e_i^2}{n-p} = MS_{Res}$$

The residuals are not independent however, as the n residuals have only n-p degrees of freedom associated with them. This non independent of the residuals has little effect on their use for model adequacy checking as long as n is not small relative to the no. of parameters p.

Properties of Residuals:

We can consider the properties we have seen in the topics of properties of least square fit.

Methods of scaling residuals:

There are four popular methods for scalling residuals.

There scaled residuals are help ful in finding observations that are outliers or extreme values. that is the observations that are separated in some fashion from the rest of the data the methods are

(i) Standarised residuals.

(ii) Studentized residuals.

(iii) Press residuals (Prediction error sum of squares).

(iv) R-student residuals.

(b)

(i) Standardized Residuals:

Since the approximate average variance of a residual is estimated by MS_{Res}, a logical scaling for the residuals would be the standardized residuals.

$$d_i = \frac{R_i}{\sqrt{MS_{RES}}}, \quad i = 1, 2 \dots n$$

The standardized residuals have mean zero and approximately unit variance consequently a large standardized residuals ($d_i > 5$, say) potentially indicates an outlier.

ii) Studentized Residuals:

Using MS_{Res} as the variance of the i^{th} residual e_i is only an approximation. We can improve the residual scaling by dividing e_i by the exact standard deviation of the i^{th} residuals. Recall from Eq. that we may write the vector of residuals as

$$e = (I-H)\,y$$

where $H = x(x'x)^{-1}x'$ is the hat matrix. The hat matrix has several useful properties. It is symmetric ($H' = H$) and idempotent ($HH = H$). Similary the matrix $I-H$ is symmetric and idempotent. Substituting $y = x\beta + \epsilon$ into yields.

$$e = (I-H)(x\beta+\epsilon) = x\beta - Hx\beta + (I-H)\epsilon$$

$$= x\beta - x(x'x)^{-1}x'x\beta + (I-H)\epsilon = (I-H)\epsilon$$

Thus, the residuals are the same linear transformation of the observations y and the errors ϵ. The covariance matrix of the residual is

$$\text{var}(e) = \text{var}[(I-H)\epsilon] = (I-H)\,\text{var}(\epsilon)(I-H)' = \sigma^2(I-H)$$

Since $\text{var}(\epsilon) = \sigma^2 I$ and $I-H$ is symmetric and idempotent. The matrix $I-H$ is generally not diagonal, so the residuals have different various and they are correlated.

The variance of the ith residual is

$$\text{var}(e_i) = \sigma^2(1-h_{ii}),$$

where h_{ii} is the ith diagonal element of the hat matrix H. The covariance between residuals e_i and e_j is

$$\text{cov}(e_i, e_j) = -\sigma^2 h_{ij}$$

where h_{ij} is the jth element of the hat matrix. Now since $0 \le h_{ii} \le 1$,

(8)

using the residual mean square MS_{Res} to estimate the variance of the residuals actually overestimates $\text{var}(e_i)$. Furthermore, since h_{ii} is a measure of the location of the e_i depends on where the point x_i lies. Generally points near the centre of the x-space have larger variance than residuals at more remote locations. Violation of model assumptions are more likely at remote points, and these violations may be hard to detect from inspection of the ordinary residuals e_i because their residuals will usually be smaller.

A logical procedure, then, is to examine the studentized residuals

$$r_i = \frac{e_i}{\sqrt{MS_{Res}(1-h_{ii})}} \quad ; \quad i = 1, 2 \dots n$$

instead of e_i (or d_i). The studentized residuals have constant variance $\text{var}(r_i) = 1$ regardless of the location of x_i when the form of the model is correct. In many situations the variance of the residuals stabilizes, particularly for large data sets. In these case there may be little difference between the studardized and studentized residuals. Thus, standardized and studentized residuals often convey equivalent

information However, since any point with a large residuals and a large hii is potentially highly influential on the least-squares fit, examation of the studentized residuals is generally recommended.

(9)

Some of these points are very easy to see by examining the studentized residuals for a simple linear regression. it is easy to show that the studentized residuals are

$$r_i = \frac{e_i}{\sqrt{MS_{RES}\left[1 - \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right)\right]}} \qquad i = 1, 2 \cdots n$$

Notice that when the observation x_i is close to the midpoint of the x-data, $x_i - \bar{x}$ will be small, and the estimated standard deviation of e_i [The denominator of equ 4.9 will be large. conversely, when x_i is near the extreamends of the range of the x-data, $x_i - \bar{x}$ will be large, and the estimated standard deviation of e_i will be small. Also, when the sample size n is really large, the so in big data sets, studentized residuals may not differ dramatically from standardize residuals.

Press Residuals :

The standardized and studentized residuals are effective in detecting outliers. Another approach to making residuals useful in finding outliers is to examine the quantity that is computed from $y_i - \hat{y}_{(i)}$, where $\hat{y}_{(i)}$ is the fitted value of the i^{th} response based on all observations except the i^{th} one. The logic behind this is that if the i^{th} observation y_i is really unusual, the regression model based on all observation may be overly influenced by this observation. This could procedure a fitted value $\hat{y}_i$ that is very similar to the observed value y_i, and consequently, the ordinary residuals e_i will be small. Therefore, it will be hard to detect the outlier. However, if the i^{th} observation is deleted, then $\hat{y}_{(i)}$ cannot be influenced by that observation, so the resulting residuals should be likely to indicate the presence of the outlier.

If we delete the i^{th} observation, fit the regression model to the remaining $n-1$ observations, and calculate the predicted value of y_i corresponding to the deleted observation, the corresponding

Prediction error is

$$e_{(i)} = y_i - \hat{y}_{(i)}$$

(11) . This prediction error calculation is repeated for each observation $i = 1, 2 \dots n$. These prediction errors are usually called PRESS residuals. (because of their use in computing the prediction error sum of squares discussed) some authors call the $e_{(i)}$ deleted residuals.

It would initially seem that calculating the PRESS residuals requires fitting n different regression. However, it is possible to calculate PRESS residuals from the results of a single least squares fit to all n observations we shown. How this is acomplished. It turns out that the i^{th} PRESS residual is

$$e_i = \frac{e_i}{1 - h_{ii}} \; ; \; i = 1, 2 \dots n$$

It is easy to see that the PRESS residuals is just the ordinary residual weighted according to the diagonal elements of the hat matrix h_{ii} - Residuals associated with points for which h_{ii} is large will have large PRESS residuals. These points will generally be high influence points. Generally a large difference between the ordinary residual and the

PRESS residuals will indicate a point where the model fits the data well, but a model built without that points predicts poorly.

Finally, the variance of the i^{th} Press residual is

$$\text{var}[e_{(i)}] = \text{var}\left[\frac{e_i}{1-h_{ii}}\right] = \frac{1}{(1-h_{ii})^2}\left[\sigma^2(1-h_{ii})\right]$$

$$= \sigma^2/1-h_{ii}$$

So that a standardized PRESS residual is

$$\frac{e_{(i)}}{\sqrt{\text{var}(e_{(i)})}} = \frac{e/(1-h_{ii})}{\sqrt{\sigma^2(1-h_{ii})}} = \frac{e}{\sqrt{\sigma^2(1-h_{ii})}}$$

which, if we use MS_{Res} to estimate σ^2 is just the studentized residuals discussed previously.

R-Student :

The studentized residual r_i discussed above is often considered an outlier diagnostic. It is customary to use M_{Res} as an estimate of σ^2 in computing r_i. This is referred to as internal scaling of the residual because MS_{Res} is an internally generated estimated of σ^2 obtained from fitting the model to all n observations. Another approach would be to use an estimate of σ^2 based on a

data set with the i^{th} observation removed. Denote the estimate of σ^2 so obtained by $S^2_{(i)}$. We can show that

$$S_i^{(2)} = \frac{(n-p)\,MS_{Res} - e_i^2(1-h_{ii})}{n-p-1}$$

The estimate of σ^2 is used instead of MS_{Res} to produce an externally studentized residual, usually called R-student given by

$$t_i = \frac{e_i}{\sqrt{S^2_{(i)}(1-h_{ii})}} \; ; \; i = 1, 2 \cdots n$$

In many situations t_i will differ little from the studentized residual r_i. However, if the i^{th} observation is influential, then $S^2_{(i)}$ can differ significantly from MS_{Res}, and thus the R-student statistics will be more sensitive to this point.

It turns out that under the usual regression assumption t_i will follow the t_{n-p-1} distribution. Appendix 9 establishes a formal hypothesis-testing procedure for outlier detection based on R student. one could use a Bonferroni type approach and compare all n values of $|t_i|$ to $t_{(\alpha/2n),\, n-p-1}$ to provide guidance regarding outliers. However it is our view that a formal approach is usually not necessary and that

only relatively crude cutoff values need be considered. In general, a diagnostic view as opposed to a strict statistical hypothesis-testing view is best. Further more, detection of outliers often needs to be considered simultaneously with detection of influential observations.

(14)

Residual Plots:

Graphical analysis of residuals is a very effective way to investigate the adequacy of the fit of a regression model and to check the underlying assumption. Here we introduce and illustrate the basic residual plots. They are namely.

i) Normal probability plot.
ii) plot of residuals against the fitted values $\hat{y}$.
iii) plot of residuals against the regressor.
iv) plot of residuals in time squence.

These plots are diffically generated by regression computer software packages. They should be examined rountinely in all regression modeling problems. It is often a good idea to plot both the original residuals and one or more of the scaled residuals. We mainly plot studentized residuals because they have

constant variance.

(i) Normal probability plot :

(15)

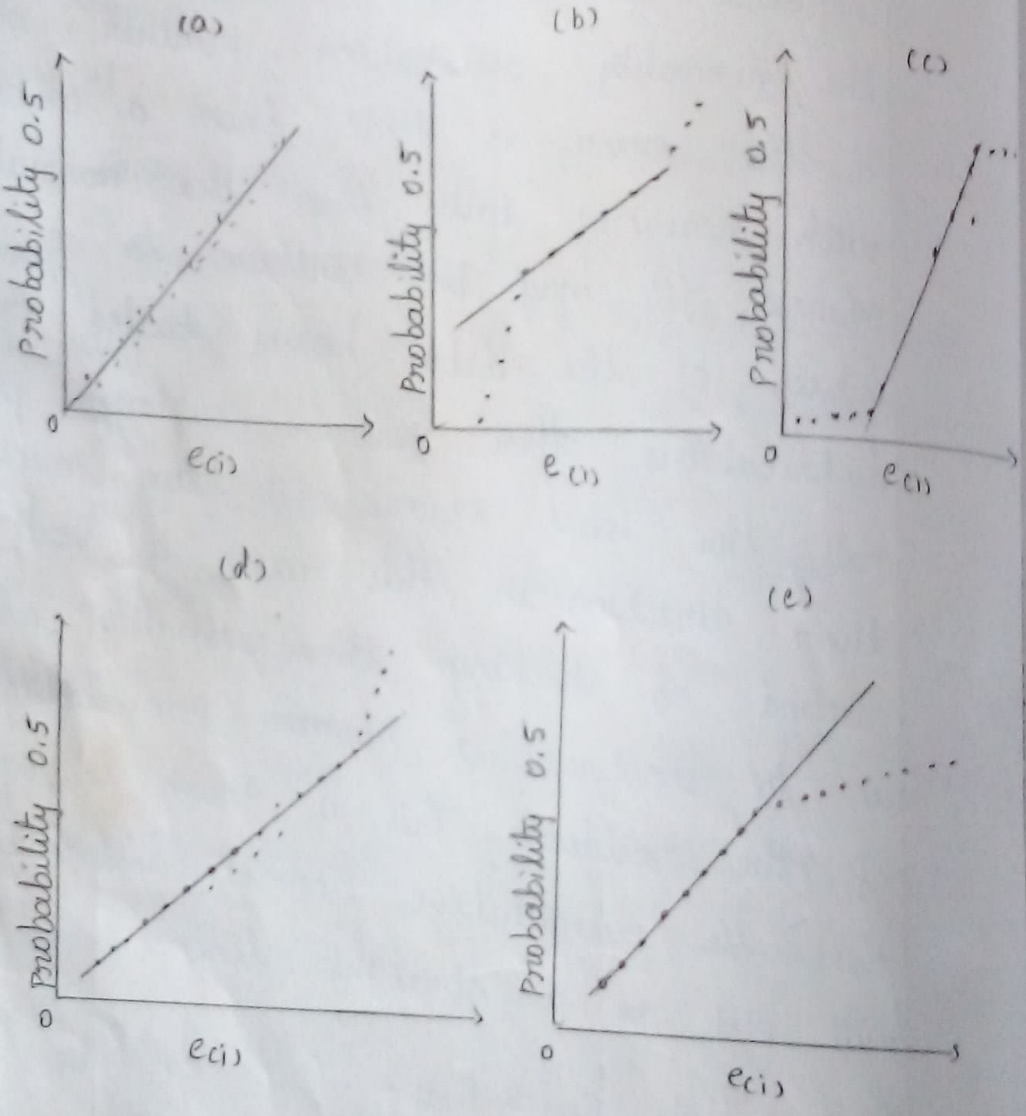
Small depature from the normality assumption do not affect the model greatly, but gross non normality is potentially more serious as the t or F statistic and confidence and Prediction intervals depend on the normality assumption. Further, more if the error is come from a distribution with heaviour tails then the normal least square fit may be senstive to a small subset of the data heavy tailed error distributions often generate outlayers that pull the least square fit too much in their direction in this cases a very simple method of checking the normality assumption is to construct a normal probability plot of the residuals. This a graph designed to that the cumulative normal distribution will plot as a straight line.

Let $e_{[1]} \leq e_{[2]} < \cdots < e_{[n]}$ be the residuals ranked in the increasing order if we plot $e_{[i]}$ against the cumulative probability

$$P_i = \frac{(i - \frac{1}{2})}{n} \quad , i = 1, 2 \cdots n$$

On the normal probability plot. The resulting points should lie approximately on a

straight line. The straight line is usually determined, with exphasis on the central values rather then the extremes [The 0.33 and 0.67 cumulative probability points]. substeuntial depatures from a straight line indicates that the distribution is not normal.



Normal Probability plots

a) Ideal, b) heavy tailed distribution.

c) Light - tailed distribution.

d) positive skew.

e) negative skew.

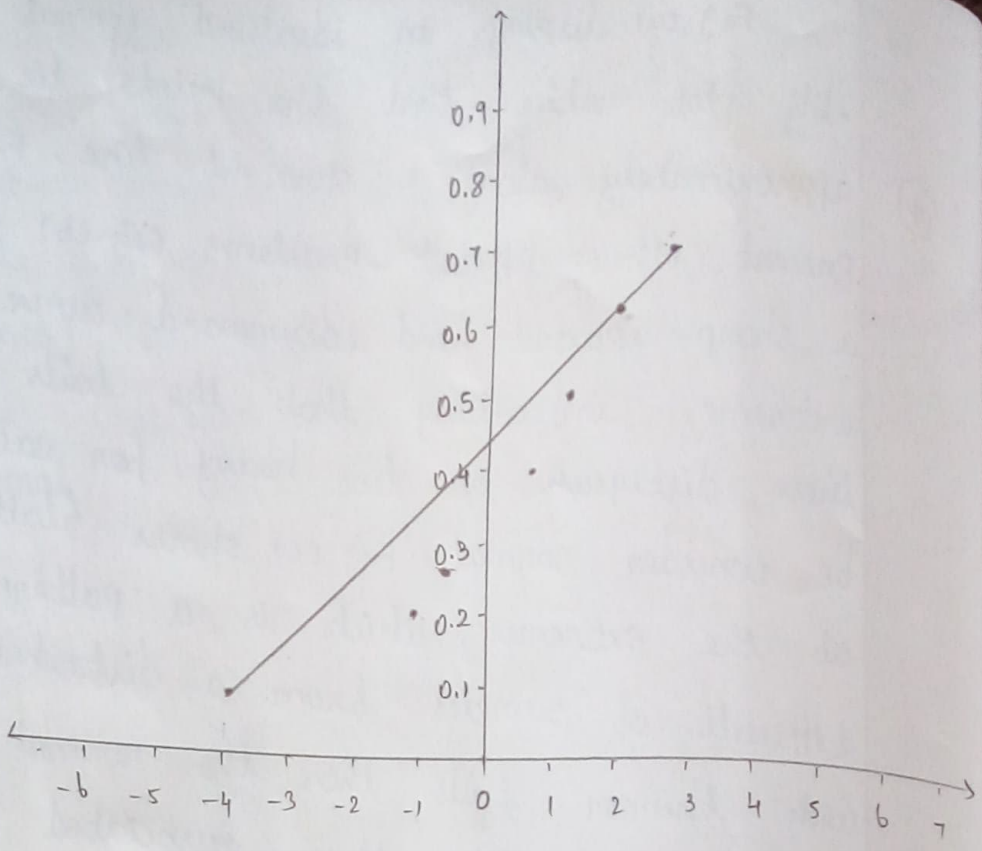
Fig (a) displays an idealised normal probabi-lity plot notice that the points lie approximately along a straight line. Fig (b-e) Present other difficult problems. Fig (b) shows a sharp upward and downward curve at both extremes, indicating that the tails of these distiqush on two heavy for each to be consider normal. Fig (c) shows flattening at the extremes, which is a pattern tiffiault of samples from a distribution with thinner tails then the normal. Fig (d and e) exhibits patterns associated with possitive and negative skew respectively.

17

Problem:

1) construct a normal probability plot for the example we have discussed in the earlier in residual analysis.

assending order:	$P(i-1/2)/n$
$e_1 = -4.333$	$(1-1/2)/7 = 0.07142$
$e_2 = -1.333$	$(2-1/2)/7 = 0.21428$
$e_3 = -0.833$	$(3-1/2)/7 = 0.3571$
$e_4 = 0.667$	$(4-1/2)/7 = 0.5$
$e_5 = 1.167$	$(5-1/2)/7 = 0.6428$
$e_6 = 2.167$	$(6-1/2)/7 = 0.7858$
$e_7 = 2.67$	$(7-1/2)/7 = 0.9285$



ii) Plot of residuals against the fitted value $\hat{y}$:

A plot of residuals e_i versus the corresponding fitted values $\hat{y}$ is usefull for detecting several common types of model inadequacies.

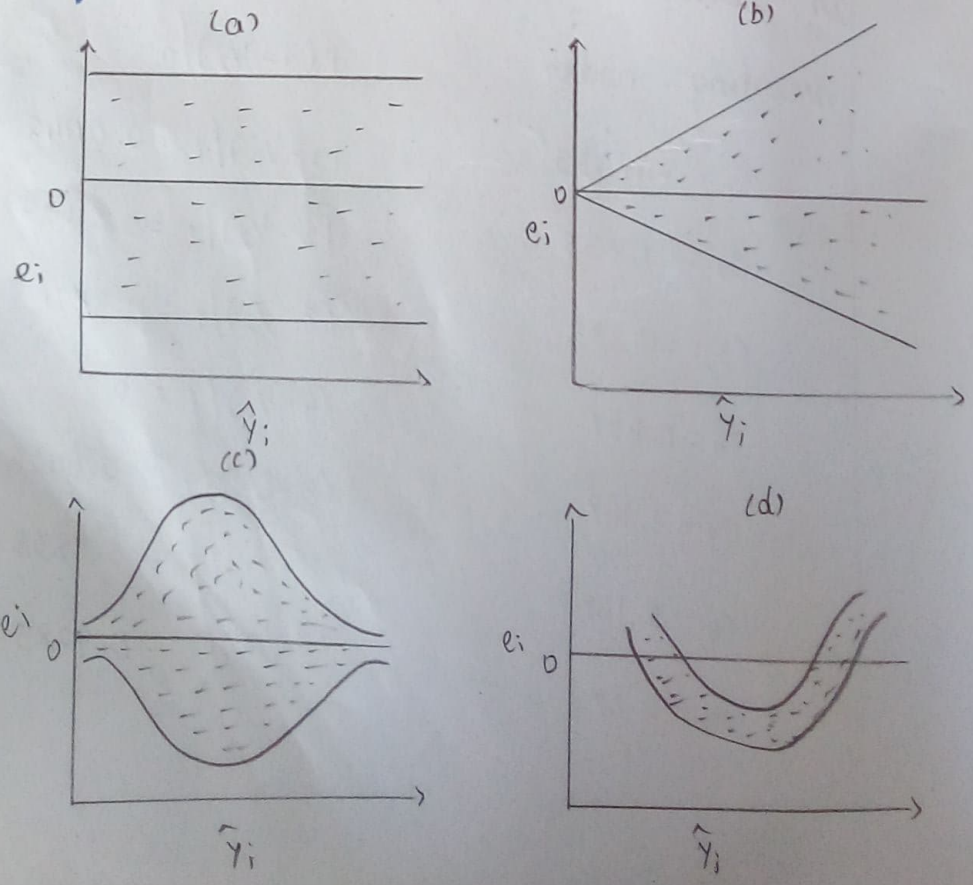


Fig (a) which indicates that the residuals can be contained in a horizontal line (band). then there are no obivious model defects. plots of e_i versus (or) against $\hat{y}_i$ that resembles any of the patterns. In figures (b-d) are symptomatic of model deficiences.

The patterns in figure (b) and (c) indicate that the variance of the errors is not constant. The outword opening funnel. pattern in fig (b) implies that the variance is an increasing function of y [some time an inword opening funnel is also possible].

The double bow pattern in fig (c) often occurs when y is a proportion between 0 and 1. The variance of a binomial proportion near 0.5 is greater than near 0 or 1. The usual approach for dealing with inequality of variance is to apply a suitable transformation in either the regressor variable or the response variable or both. In practice, transformation on the response variable are generally employed to stamilize variance.

A curved plot such as in fig (d) indicates non-linearity. This could mean that other regressor variables are needed in the model. For example, a squared term may be necessary. Transformation on the

to ŷ

regressor on the response variable may also be helpful in these cases.

A plot of the residuals against $\hat{y}_i$ may also reveal one or more usually large residuals. These points are of course, potential outlayers. Large residuals that occur at the extreme $\hat{y}_i$ values could also indicate that either the variance is not constant or the true relationship between x and y is not linear. These possibilities should be investigated before the points are considered as outliers.

(iii) Plot of residuals against the regressor (x_i):

Plotting the residuals against corresponding value of each regressor variable (x_i), can also be helpful. For known the inadquacies. These plots often exhibits pattern such as shows in fig (a, b, c, d) in fitted value of $\hat{y}$. Except that the horizontal scale is x_{ij} for the j^{th} regressor rather than $\hat{y}_i$. Once again an impression of a horizontal band containing the residuals is desirable. The funner and double bow patterns in figures (b) and (c) indicate non constant variance. The curved

band in figure (d) on a non-linear pattern in general implies that the assume relation -ship between the y and the regressor x_i is non correct. Thus either higher order terms x_i [such as x_i^2] on the transformation should be considered.

In the simple linear regressor case it is not necessary to plot residuals versus both $\hat{y}_i$ and the regressor variable x_i. The reason is that the fitted values $\hat{y}_i$ are linear combination of the regressor value x_i, Show that plots only would differ in the scale for the abscissa.

(iv) Plot of residuals in time sequence:

If the time series in which the data where collected is known. It is a good idea to plot the residual against time order. Idealy, this plot will resambles in figures (a, b, c, d) that is a horizantal band will enclosed all of the residuals, and the residuals will fluctuate in a more or less random fashion within this band. However, if this plot resambles the pattern in (b)-(d) This may indicate that the variances is changing with time, or that linear are quadratic terms in time should be added to the model. The

time sequence plots of residuals may indicate that the error at one time period are correlated with these at other time periods. The correlation between model errors at different time period is called auto correlation. A plot such as figure (a) indicates +ve auto correlation, while fig (b) is typical of negative auto correlation. The presentise of auto correlation is a potential series violence of the basic regression assumption. More discussion about methods for detecting auto correlation and rematial measures are discussed.

Date 20.01.20

Diagnostics tests:

When we compute a sample average each observation in the sample has the same weight in determining the outcome. In the regression suituation this is not the case. For example, we have considered that the location of observation is x can play unimportant role indetermining the regression co-efficient. we have also focused attention (on) outliers or observation that have unusual y value.

On the other hand we some times find that a small subset of data exerts a dis-proposanet influence on the model

co-efficient and properties. In extreme case, the parameter estimates may depend more on the influential subset of points than on the majority of the data. Consequently we would like to find influencial points asses infact on the model if these influential points are indeed bad values then they should be eliminated from the samples. ~~Here~~ ~~we~~

Here we present several diagnostic test for leverage and influence. These diagnostic are available in most multiple regression computer packages. It is important to use the diagnostic in conjuction with residuals ~~tec~~ analysis techniques.

Sometimes we find that a regression co-efficient may have a sign that does not make engeneering (or) scientific sence, a regressor know to be important may be satistically in significant. These situation may be the result of one (or) fue influential observation residuals Finding the observations we can use any of the diago -nastic test.

PRESS STATISTIC :

(24)

We defined the PRESS residuals as $e_{(i)} = y_i - \hat{y}_{(i)}$, where $\hat{y}_{(i)}$ is the predicted value of the i^{th} observed response based on a model fit to the remaining $n-1$ sample points. We noted that large PRESS residuals are potentially useful in identifying observations where the model does not fit the data well, or observations for which the model is likely to provide poor future predictions.

Allen [1971, 1974] has suggested using the Prediction Error sum of squares (or the PRESS statistic), defined as the sum of the squared PRESS residuals, as a measure of model quality. The PRESS statistic is

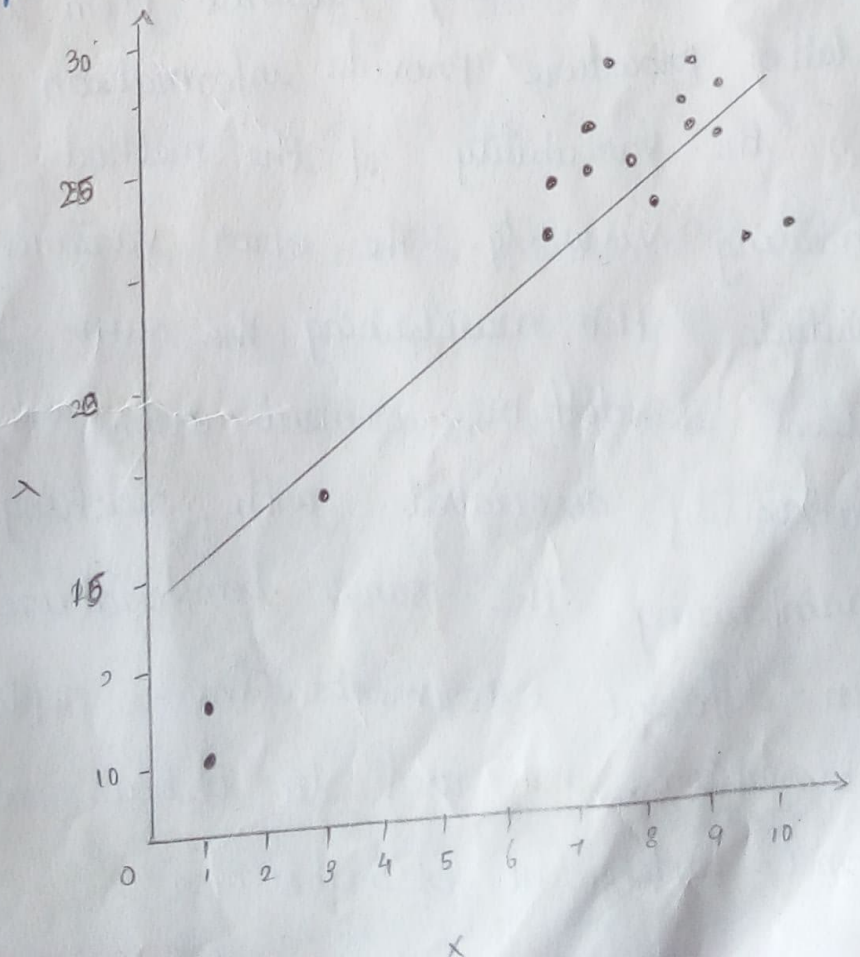
$$PRESS = \sum_{i=1}^{n} [y_i - \hat{y}_{(i)}]^2$$

$$= \sum_{i=1}^{n} \left(\frac{e_i}{1 - h_{ii}} \right)^2$$

PRESS is generally regarded as a measure of how well a regression model will perform in predicting new data. A model with a small value of PRESS is desired. □

A Formal test of lack of fit :

We will now present a formal statistical test for the lack of fit of a regression model. The Produce assumptes that the normality independence, and constant variance requirements are met and that only the first-order or straight-line character of the relationship is in doubt. For example, consider the data in figure. There is some indication that the straight-line fit is not very satisfactory, and it would be helpful to have a test procedure to determine if there is systematic curvature present.



The lack-of fit requires that we have replicate observations on the response y for at least one level of x.

We emphasize that these should be true replications, not just duplicate readings or measurements of y.

For example,

Suppose that y is product viscosity and x is temperature. True replication consists of running n_i separate experiments at $x = x_i$ and observating viscosity, not just running a single experiment at x_i and measuring viscosity n_i times.

The readings obtained from the latter procedure provide information only on the variability of the method of measing viscosity. The error variance σ^2 includes this maintaining ~~the same temperatu~~ level ~~in~~ ~~diff~~ measurement error and the variability associated with reaching and maintaining the same temperature level in different experiments. These replicated observations and used to obtain a model-independent estimate of σ^2.

Suppose the we have n_i observations on the response at the i^{th} level of the regressor $x_i = i = 1, 2 \dots m$. Let Y_{ij} denote the j^{th} observation on the response at x_i, $i = 1, 2 \dots m$ and $j = 1, 2 \dots n$. There are $n = \sum_{i=1}^{m} n_i$ total observations. The test procedure involves partitioning the residual sum of squares in two components.

$$SS_{Res} = SS_{PE} + SS_{LOF}$$

where SS_{PE} is the sum of squares due to pure error and SS_{LOF} is the sum of squares due to lack of fit.

To develop this partitioning of SS_{Res}, note that the (ij) th residual is

$$Y_{ij} - \hat{Y}_i = (Y_{ij} - \bar{Y}_i) + (\bar{Y}_i - \hat{Y}_i) \longrightarrow (1)$$

where $\bar{Y}_i$ is the average of the n_i observation at x_i. squaring both sides of Eq(1) and summing over i and j yields.

$$\sum_{i=1}^{m} \sum_{i=1}^{n_i} (Y_{ij} - \hat{Y}_i)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^{m} n_i (\bar{Y}_i - \hat{Y}_i)^2 \longrightarrow (2)$$

since the cross-product term equals zero.

The left hand side of equation (3) is usual residual sum of squares.

SS_{PE}

The two components on the right hand side of the equ (3) measure Pure error and lack of fit.

We see that the pure sum of squares

$$SS_{PE} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \longrightarrow (4)$$

we consider the equation (4) is obtained by computing the corrected sum of square of the reapeat observation at each level of x and then pooling over the m level of x. Since there are $n_i - 1$ d.f for pure error at each level of x then the total no. of d.f associated with pure error sum of square is

$$\sum_{i=1}^{m} (n_i - 1) = n - m \longrightarrow (5)$$

The sum of square for lack of fit

$$\sum_{i=1}^{m} n_i (\bar{y}_i - \hat{y})^2 \longrightarrow (6)$$

is a weighted sum of squared deviatio between the mean response $\bar{y}_i$ at each x level and the corresponding the fitted value. If the fitted values $\hat{y}_i$ are close to the corresponding average response $\bar{y}$ then there is a strong indication that the regression function is linear. There are $m-2$ d.f associated with SS_{LOF}, since their are m levels of x and two d.f

are lost because two parameters ~~estimated~~ must be estimated to obtained the $\bar{y}_i$.

(29) Computationally we usually obtain SS_{LOF} by subtracting SS_{PE} from SS_{RES}.

The test statistics for lack of fit is

$$F_0 = \frac{(SS_{LOF} / m-2)}{(SS_{PE} / n-m)}$$

$$F_0 = \frac{SS_{LOF}/(m-2)}{SS_{PE}/(n-m)}$$

i.e MS_{LOF} / MS_{PE}. → (7)

variance stabilizing transformation:

The assumption of constant variance is a basic requirement of regression analysis a common reason for a vailasion of this assumption is for the response variable y to follow a probability distribution in which the variance is furtionaly realeated to the mean For example, If y is a poisson r.v in a simple linear regression model then the variance y is equal to the mean. Since the mean of y is rebated to the regressor variable x then the variance of y will be proposonal to x. variance stabilizing transformation are often useful in this case. Thus if the distribution of y is

poisson we could regress against x since the variance of the square root of a poisson r.variable is independent of mean.

(30)

Several commonly used variance stabilizing used on summarised in the following table.

The strength of a transformation depends upon the amount of curvature that indicates

useful ~~of~~ variance stabilizing transformatio

Relationship of σ^2 to expectation of Y	Transformation
1. $\sigma^2 \propto$ constant.	$Y' = Y$ (no transformation)
2. $\sigma^2 \propto E(Y)$	$Y' = \sqrt{Y}$ (Square root, poisson data).
3. $\sigma^2 \propto E(Y)[1 - E(Y)]$	$Y' = \sin^{-1}(\sqrt{Y})$ (arcsin binomial proportions. $0 \leq y_i \leq 1$)
4. $\sigma^2 \propto (E(y))^2$	$Y' = \ln(Y)$ (log)
5. $\sigma^2 \propto (E(y))^3$	~~$Y' = Y^{-1/2} = \frac{1}{\sqrt{Y}}$~~
6.	$Y' = Y^{-1/2} = \frac{1}{\sqrt{Y}}$ (Reci-Procal square root).
6. $\sigma^2 \propto (E(y))^4$	$Y' = Y^{-1} = \frac{1}{Y}$ (Reci-Procal)

Sometimes we can use prion experiance on theoretical consideration to guid

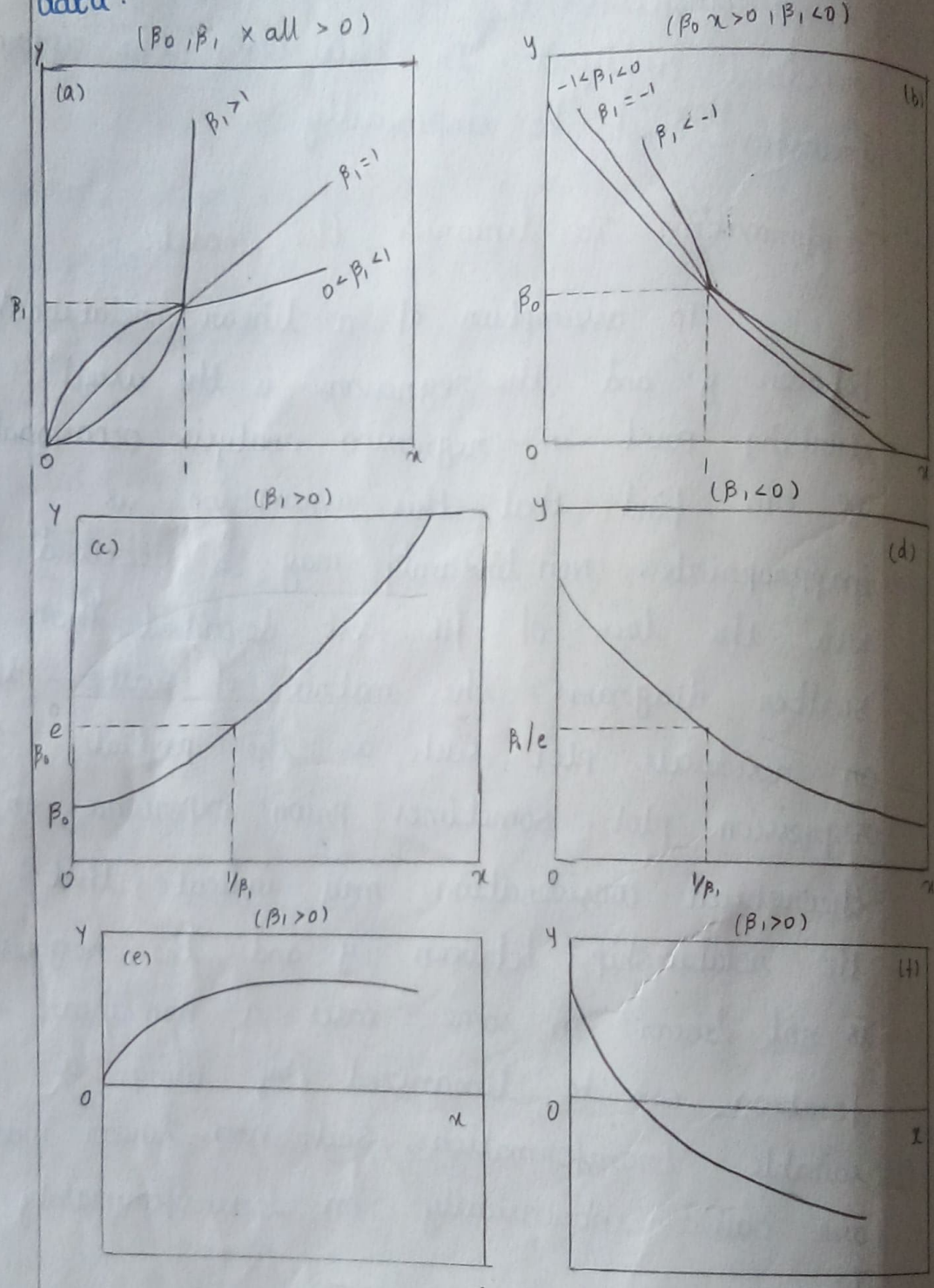
to appropriate transformation. However in many case we have no a Priory reason to suspect the error is not constant. Out first indication of the problem is from inspection of the scatter diagram or residual analysis. In this case the appropriate transformation must be emparically.

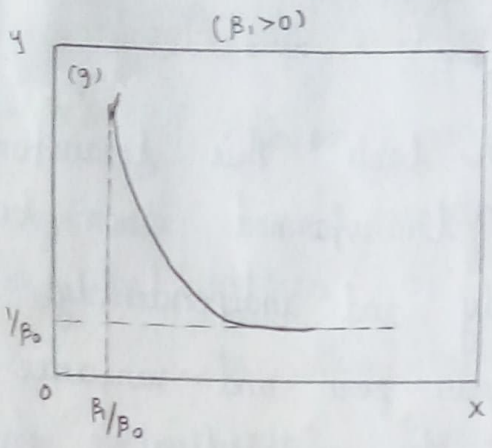
Tranformation To linearize the model:

To assumption of a linear relationship between y and the regressors is the usual starting point in regression analysis occasionally we can find that this assumption is inappropriate. Non linearity may be detected via the lack-of-fit test described from scatter diagrams. the matrix of scatter-plots or residuals plot such as the partial regression plot. Sometimes Prion experience or theoretical consideration may indicate that the relationship between y and the regressors is not linear. In some cases a non linear function can be linearized by using a suitable transformation. Such non-linear models are called intrinsically or transformably linear.

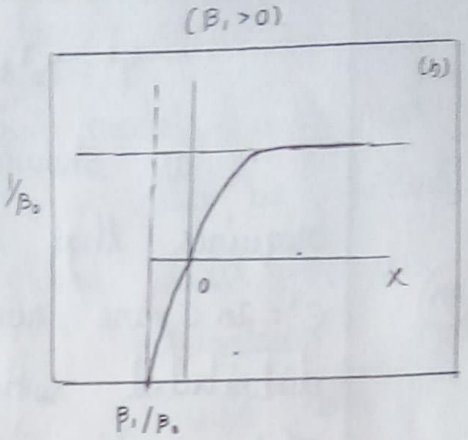
Several linearizable functions are shown in figure. The corresponding non-linear functions, transformations, and resulting linear forms are shown in table. which the scatter

diagram of y against x indicates curvature we may be able to match the observed of the plot to one of the curves in fig and use the linear form of the function to represent the data.





positive curvature



Negative curvature

Linearizable function (From Daniel and Wood [1980], used with permission of the publishes)

Linearizable function and corresponding linear form

Figure	Linearizable function	Transformation	Linear form
a, b	$Y = \beta_0 x^{\beta_1}$	$y' = \log Y_0$, $x' = \log x$	$y' = \log \beta_0 + \beta_1 x'$
c, d	$Y = \beta_0 e^{\beta_1 x}$	$y' = \ln Y$.	$y' = \ln \beta_0 + \beta_1 x$
e, f	$Y = \beta_0 + \beta_1 \log x$	$x' = \log x$	$y' = \beta_0 + \beta_1 x'$
g, h	$Y = \frac{x}{\beta_0 x - \beta_1}$	$y' = \frac{1}{y}$, $x' = 1/x$	$y' = \beta_0 - \beta_1 x'$.

To illustrate a non-linear model that is intrinsically linear, consider the exponential function.

$$Y = \beta_0 e^{\beta_1 x} \epsilon$$

This function is intrinsically linear since it can be transformed to a straight line by a logarithmic transformation.

$$\ln y = \ln \beta_0 + \beta_1 x + \ln \epsilon$$

(con)

$$y' = \beta_0' + \beta_1 x + \epsilon'.$$

as shown in table. This transformation requires that the transformed error terms $\epsilon' = \ln \epsilon$ are normally and independently distributed with mean zero and variance σ^2. This implies that the multiplicative error ϵ in the original model is log normally distributed. we should look at the residuals from the transformed model to see if the assumptions are valid. Generally if x and y are in the proper matrix, the usual least squares assumptions are more likely to be satisfied, although it is not unusual to discover at this stage that a non-linear model is preferable.

various types of reciprocal transformations are also useful. For example, the model.

$$y = \beta_0 + \beta_1 (1/x) + \epsilon$$

can be linearized by using the reciprocal transformation $x' = \frac{1}{x}$. The resulting linearized model is

$$y = \beta_0 + \beta_1 x' + \epsilon$$

Other models that can be linearized by reciprocal transformations are

$$\frac{1}{y} = \beta_0 + \beta_1 x + \epsilon$$

$$y = \frac{x}{\beta_0 x - \beta_1 + \epsilon}$$

Generalized and weighted least squares:

Linear regression models with non constant error variance can also be fitted by the method of weighted least squares

In this method of estimation the deviation between the observed and the expected values of y_i is multiplied by a weight w_i, choosen inversly Proporso -nal to the variance of y_i. For the case of simple linear regression, the weighted least squares function is

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} w_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \longrightarrow (1)$$

The resulting least squares normal equation are

$$\left.\begin{aligned} \hat{\beta}_0 \Sigma w_i + \hat{\beta}_1 \Sigma w_i x_i &= \Sigma w_i y_i \\ \hat{\beta}_0 \Sigma w_i x_i + \hat{\beta}_1 w_i x_i^2 &= \Sigma w_i x_i y_i \end{aligned}\right\} \longrightarrow (2)$$

Solving the equation (2) will Produce weighted least square estimates are $\hat{\beta}_0, \hat{\beta}_1$

Thus we given developement of weighted least square for the multiple regression model, i.e.. generalized weighted least squares. Here we begin by considering a sighiltly more general suituation concering the structure of the model errors.

When the error ϵ are uncorrelated but have unequal variances so that the co-variance matrix ϵ is

$$\sigma^2 V = \sigma^2 \begin{bmatrix} 1/w_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/w_2 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & \cdots & 0 & 0 & \cdots \; 1/w_n \end{bmatrix}$$

The estimation is usually called weighted least squares. Let $W = V^{-1}$. Since V is a diagonal matrix, W is also diagonal with diagonal elements $w_1, w_2 \cdots w_n$.

The weighted least squares normal equations are

$$(x'Wx)\hat{\beta} = x'Wy$$

This is multiple regression of the weighted least squares normal equation therefore

$$\hat{\beta} = (x'Wx)^{-1} x'Wy$$

is the weighted least squared estimator.