Analysis of Residuals in MLR

1 Introduction

When we compute the sample average, a data point that is much larger or smaller than the rest of the data will have a large influence on the value of the average. We usually check to see if the data point was an error (in which case the value is checked for correctness) or if there is something interesting about the data point that warrants a closer look.

Likewise, we are concerned with observations that have high leverage, are outlying, or are influential in a regression model. In this module, you will learn various measures to detect these observations. A lot of these measures are based on residuals. Generally speaking, we are most concerned with influential observations, as their presence (or removal) significantly alter the estimated regression equation.

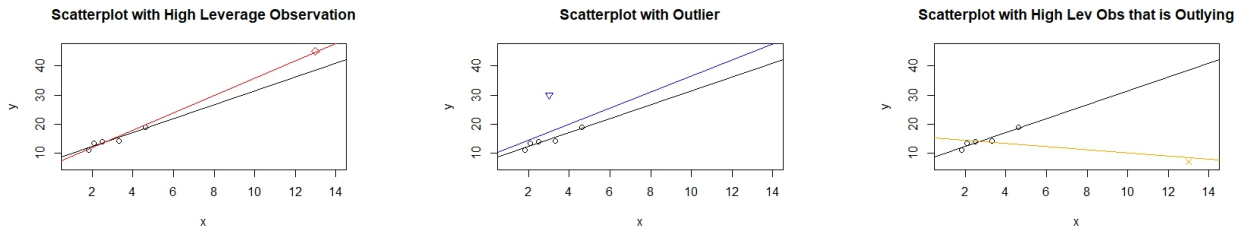
Lastly, we will be using residuals from MLR to help us assess if any of the predictor variables need to be transformed, and if so, how to transform them, in order to meet the assumptions for MLR.

1.1 Terminology

You may have noticed in the earlier paragraphs that we are differentiating between observations that have high leverage, are outlying, or are influential. These observations have slightly different definitions:

- **High leverage**: an observation whose predictor(s) is extreme.
- **Outlier**: an observation whose response variable does not follow the general trend of the data.
- **Influential**: an observation whose presence (or removal) unduly affects any part of the regression analysis, usually in terms of unduly affecting the predicted response, the estimated coefficients, or the results from hypothesis tests and confidence intervals.

The scatterplots below in Figure 1 display three examples of these types of observations. We are using just one predictor for ease of visualization. In each example, there are 6 observations.



(a) With High Leverage Obs (b) With Outlying Obs (c) With Influential Observation

Figure 1: Scatterplot of Data with high leverage, Outlying, and Influential Observation

Figure 1a displays the scatterplot of an observation that is considered high leverage, denoted by the red diamond in the top right of the plot.

- The general pattern of the observations is a positive linear association between both variables. As x gets larger, y gets larger.
- The high leverage observation is a high leverage observation, as its value of x is a lot larger than the values of x for the other observations.

- The high leverage observation has a large value of x, and its response has a value that is consistent with the general pattern of the observations. Therefore, it not considered an outlier.
- In the figure, two estimated regression lines are overlaid. The line in black is the regression line for the observations excluding the high leverage observation, and the line in red is the regression line when the high leverage observation is included. The lines are almost indistinguishable. Therefore, the high leverage observation is not influential.

Figure 1b displays the scatterplot of an observation that is considered an outlier, denoted by the blue triangle in the top left of the plot.

- The general pattern of the observations is a positive linear association between both variables. As x gets larger, y gets larger.
- The outlier is not a high leverage observation, as its value of x is not a lot larger or smaller than the values of x for the other observations.
- The outlier has a small value of x, but its response has a value that is a lot larger than what the general pattern would suggest. Therefore, it is considered an outlier.
- In the figure, two estimated regression lines are overlaid. The line in black is the regression line for the observations excluding the outlier, and the line in blue is the regression line when the outlier is included. The lines are almost indistinguishable. Therefore, the outlier is not influential.

Figure 1c displays the scatterplot of an observation that is considered influential, denoted by the orange cross in the bottom right of the plot.

- The general pattern of the observations is a positive linear association between both variables. As x gets larger, y gets larger.
- The influential observation is also a high leverage observation, as its value of x is a lot larger than the values of x for the other observations.
- The influential observation has a large value of x, but its response has a value that is a lot smaller than what the general pattern would suggest. Therefore, it is also considered an outlier.
- In the figure, two estimated regression lines are overlaid. The line in black is the regression line for the observations excluding the influential observation and the line in orange is the regression line when the influential observation is included. The lines are very different. Therefore, the observation is influential.

From Figure 1, we can see that an observation that has high leverage or is outlying is not guaranteed to be influential. As a general rule , observations that have high leverage and/or are outlying have the potential to be influential. Observations that are both outlying and have high leverage are very likely to be influential.

We will now look at measures to detect these observations. These measures typically involve the residuals, or variation of residuals, from our regression.

2 Detecting High Leverage Observations

2.1 Limitation of residuals in detecting high leverage observations

An intuitive way to detect observations that have high leverage, are outlying, or are influential, is to use the residuals, defined as:

$$e_i = y_i - \hat{y}_i. \tag{1}$$

However, there are certain limitations in using residuals to detect these observations. To illustrate this, let us go back to Figure 1. Visually, the residual is the vertical distance of an observation from the regression equation. In Figures 1a and 1c, notice that the high leverage observation (red diamond) and influential observation (orange cross) are very close to the estimated regression equation in red and orange respectively. So these observations will have small residuals. However, in Figure 1b , the outlier is far away from their respective estimated regression equations, and so have a large residual. So **residuals may be unable to detect high leverage observations**.

2.2 Hat matrix

Recall in MLR that the predicted response (or fitted values), can be written in matrix form as

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}. \tag{2}$$

Using the method of least squares, the estimated coefficients can be found using

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}. \tag{3}$$

Subbing (3) into (2), we have

$$\begin{aligned} \hat{\boldsymbol{y}} &= \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= \boldsymbol{H}\boldsymbol{y}, \end{aligned} \tag{4}$$

where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is called the **hat matrix**. If we write (4) in scalar form, we have

$$\begin{aligned} \hat{y_1} &= h_{11}y_1 + h_{12}y_2 + \cdots + h_{1n}y_n \\ \hat{y_2} &= h_{21}y_1 + h_{22}y_2 + \cdots + h_{2n}y_n \\ &= \vdots \\ \hat{y_n} &= h_{n1}y_1 + h_{n2}y_2 + \cdots + h_{nn}y_n \end{aligned}$$

where h_{ij} denotes the (i, j)th entry in the hat matrix.

2.3 Leverage

It turns out we use the diagonal entries of the hat matrix, h_{ii}, to define **leverage**, which measures the distance of the predictors for observation i and the center of the predictors for all observations.

Using (4) in scalar form, we can see that the predicted response for each observation is a linear combination of observed responses $y_1, y_2, \cdots, y_n$. The leverage h_{ii} measures the impact that observation y_i has in predicting its response. If the leverage is high, observation i has a high impact on its prediction. Leverages have the following properties:

- $h_{ii} = \boldsymbol{X_i}'\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X_i}$.
- $0 \leq h_{ii} \leq 1$.
- $\sum_{i=1}^{n} h_{ii} = p$, where p is number of parameters. Therefore the average of the leverages is $\frac{p}{n}$.

There are various recommendations of how to determine if an observation has high leverage. We will use $h_{ii} > \frac{2p}{n}$ as a general rule, or twice the average of the leverages. Note that we should use this as a guide, and remember that the higher the leverage, the further away the predictors for observation i are from the center of the predictors for all observations.

Let us find the residual and leverage of the observations that make up the scatterplot in Figure 1a. Note that there are 6 observations, and the high leverage observation has index 6. First, we take a look at the residuals. The absolute values of the residuals are reported below and are sorted in increasing order:

```
##absolute value of residuals, sorted in increasing order
sort(abs(result.lev$res))
```

```
##         5         6         2         4         3         1
## 0.1561530 0.3052037 0.5633585 0.7479269 1.3792798 1.4147975
```

Notice that the high leverage observation has a residual with absolute value of 0.3052. There are a few other observations with larger residuals. So residuals will fail to identify observation 6 has having high leverage. We now look at leverages, reported below and sorted in increasing order:

```
##leverages, sorted in increasing order
sort(lm.influence(result.lev)$hat)
```

```
##         4         3         2         1         5         6
## 0.1667478 0.1839964 0.2132494 0.2345507 0.2492167 0.9522391
```

```
##criteria
p<-2
n<-dim(Data.lev)[1]
2*p/n
```

```
## [1] 0.6666667
```

Note that observation 6 has leverage of 0.9522, which is a lot larger relative to the leverages of the other five observations. Indeed, it is larger than the suggested criteria of $\frac{2p}{n} = 0.6667$. We can see how leverage, and not residuals, is measure that should be used to identify high leverage observations.

3 Analysis of Residuals: Detecting Outliers

In the previous section, we noted that residuals should not be used to detect high leverage observations. Another limitation is that the numerical value of residuals is based on the unit of the response variable. So what makes a residual large depends on the unit of the response variable. In view of this limitation, we consider standardizing residuals to make them unitless. There are a few ways of standardizing residuals, some of which are useful in detecting outliers.

3.1 Standardized residuals

Recall that the errors in our regression model, denoted by ϵ, have variance denoted by σ^2, which in turn is estimated by the $MS_{res} = \frac{\sum(y_i - \hat{y}_i)^2}{n-p}$. Therefore, the **standardized residuals**, d_i, are found by

$$d_i = \frac{e_i}{\sqrt{MS_{res}}}. \tag{5}$$

We are dividing the residuals, e_i, by the standard error of the errors.

3.2 Studentized residuals

However, MS_{res} estimates the variance of the errors, not the residuals. It turns out the variance of the residuals is

$$Var(e_i) = MS_{res}(1 - h_{ii}). \tag{6}$$

So we should be standardizing the residual by using the standard error of the residuals instead, and we get **studentized residuals**

$$r_i = \frac{e_i}{\sqrt{MS_{res}(1-h_{ii})}}. \tag{7}$$

Let us find the residuals and studentized residuals for the observations given in Figure 1b. Note that there are 6 observations, and the outlying observation has index 6. First, we take a look at the residuals. The absolute values of the residuals are reported below and are sorted in increasing order:

```
##absolute value of residuals, sorted in increasing order
sort(abs(result.out$res))
```

```
##         1         2         4         5         3         6
##  1.259631  2.007224  2.796350  2.892751  3.754347 12.710304
```

Residuals identify the outlying observation, index 6. However, the numerical values for these residuals depend on the unit of the response variable, so it can be hard to ascertain what value of residual is considered large. So we can look at the studentized residuals instead:

```
##absolute value of studentized residuals, sorted in increasing order
sort(abs(rstandard(result.out)))
```

```
##         1         2         5         3         4         6
## 0.2134751 0.3186190 0.5258773 0.5970578 0.8023354 1.9850133
```

The studentized residual estimates how many standard deviations its predicted response is from the actual response. So the predicted response for the outlying observation is estimated to be 2 standard deviations away from its true value. Typically, studentized residuals with magnitude larger than 2 are flagged. In comparison with the other five studentized residuals, the outlier has a studentized residual quite a lot larger than the others.

Let us take a look at the studentized residual for the observations shown in Fig 1c. Again, note that there are 6 observations, and the influential observation has index 6.

```
sort(abs(rstandard(result.both)))
```

```
##         2         3         1         5         4         6
## 0.1180102 0.1618282 0.3298582 1.1348254 1.8125713 1.9409466
```

The value of the studentized residual for the influential observation is 1.9409, pretty close to 2. However, notice that in relationship with the other studentized residuals, it is not a lot larger. So we may not deem this observation as outlying, or deem that observation 4 is also outlying, which is incorrect. This can happen when the observation is influential. So, we see some limitation in using studentized residuals to flag outliers. We will further refine the residual so we can identify outliers more clearly.

3.3 Deleted residuals

We have seen an example in the previous subsection where an influential observation that is also outlying has studentized residuals that are not a lot larger than the other studentized residuals. This is because if the observation also has high leverage (for example in Fig 1c), the observation is likely to pull the estimated regression equation towards itself, resulting in a small residual and studentized residual that is not a lot larger than the rest.

To address this issue, we remove the observation in question from estimating the regression model, so the observation will not pull the regression equation towards itself.

We can use the **deleted residual** to measure this. It is defined as the difference in the value of the actual response for an observation and its prediction when the observation is not used to estimate the regression equation. The deleted residual for observation i is denoted as

$$e_{(i)} = y_i - \hat{y}_{i(i)}. \tag{8}$$

There are two values in the subscript for $\hat{y}_{i(i)}$. The value outside the subscript denotes the observation we are making the prediction for. The subscript surrounded by the parenthesis indicates the observation has been removed from estimating the regression equation.

Let us use Figure 1c as an example to demonstrate. The influential observation is denoted by the orange cross, and is observation 6:

- y_6 denotes the value of the response variable for this observation.
- $\hat{y}_{6(6)}$ denotes the predicted response for this observation, if it was removed from estimating the regression equation. Visually, on Figure 1c, $\hat{y}_{6(6)}$ is the predicted response based on the black regression line, since the black regression line was estimated without observation 6.
- $e_{(6)}$ denotes the vertical distance of observation 6 (the orange cross) from the black regression line.

A mathematically equivalent form for (8) is

$$e_{(i)} = \frac{e_i}{1 - h_{ii}}. \tag{9}$$

Note that the larger h_{ii} is, the deleted residual is larger compared to the residual. Thus deleted residuals will at times identify outliers when residuals would not, especially if it has high leverage.

Just like residuals, deleted residuals depend on the unit of the response variable, so we should scale deleted residuals.

3.4 Externally studentized residuals

We can scale the deleted residuals to obtain **externally studentized** residuals, which are

$$t_i = \frac{e_i}{\sqrt{MS_{res(i)}(1 - h_{ii})}}. \tag{10}$$

Note that $MS_{res(i)}$ denotes the MS_{res} of the regression model that is estimated without observation i. From a computational standpoint, calculating t_is this way requires us to fit n regression models, in order to find $MS_{res(i)}$ for each t_i.

A computationally more efficient to compute t_i is

$$t_i = e_i \left[\frac{n - 1 - p}{SS_{res}(1 - h_{ii}) - e_i^2} \right]^{1/2}. \tag{11}$$

The formulation in (11) only requires us to fit one estimated regression model. The reason why (10) and (11) are equivalent is based on the relationship between MS_{res} and $MS_{res(i)}$:

$$(n - p)MS_{res} = (n - 1 - p)MS_{res(i)} + \frac{e_i^2}{(1 - h_{ii})}. \tag{12}$$

Let us take a look at the externally studentized residuals for the scatterplots shown in in Figures 1b and 1c. Again, we are sorting them based on increasing order based on absolute values:

```
##from Fig 1b
sort(abs(rstudent(result.out)))
```

```
##          1          2          5          3          4          6
##  0.1859371  0.2795018  0.4720327  0.5417717  0.7585583 14.0687652
```

```
##from Fig 1c
sort(abs(rstudent(result.both)))
```

```
##         2         3         1         5         4         6
## 0.1023782 0.1406084 0.2896319 1.1935239 3.7138867 6.9686966
```

Notice how the externally studentized residuals for observation 6 in both plots are a lot larger than for the other observations. It is now a lot more obvious that these are outliers, compared to using studentized residuals earlier on.

Generally speaking, $|t_i| > 3$ is a decent rule to use to flag outliers. But this rule should be used in relation with all the t_is.

We have covered measures to detect high leverage observations and outliers. These observations usually have something interesting that make them "stand out" from the other observations. But they are not necessarily influential in a regression setting.

4 Influential Observations

In a regression setting, an observation is influential if its presence, or removal, drastically alters the regression analysis. We usually quantify this alteration in terms of how much the predicted response and / or the estimated coefficients change with and without the observation in question. Generally speaking, observations that have high leverage and are outlying have the most potential to be influential. As an example from Figure 1, we see that the estimated regression line is drastically altered only in Figure 1c.

4.1 DFBETAs

We can look at how the estimated coefficients change when the observation in question is removed from estimating the model. This is the motivation behind **DFBETAs**:

$$DFBETAS_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{MS_{res(i)} c_{jj}}} \tag{13}$$

where c_{jj} is the jth diagonal element of $(\boldsymbol{X'X})^{-1}$, since $Var(\hat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X'X})^{-1}$.

Notice there are two subscripts in $DFBETA_{j,i}$. The first subscript denotes the coefficient, and the second denotes the observation. The numerator in (13) measures the change in the estimated coefficient, and then we standardize this change by dividing with its standard error.

A few other things to note with (13):

- The sign for $DFBETAS_{j,i}$ indicates whether excluding an observation leads to an increase or decrease in that estimated regression coefficient.
- $DFBETAS$ can be interpreted as the number of standard errors the estimated coefficient changes when observation i is removed from building the model.
- Suggested rule for influential observation: $|(\text{DFBETAS})_{j,i}| > \frac{2}{\sqrt{n}}$. Again, this should be used as a guide, and in conjunction with the values of $DFBETAS$ for all coefficients and observations.

Let us go back to the scatterplot in Fig 1c and compute the $DFBETAs$:

```
dfbetas(result.both)
```

```
##   (Intercept)           x
## 1 -0.15341916   0.08625257
## 2 -0.04957068   0.02491157
## 3  0.05689633  -0.02049090
## 4  1.05250538   0.03665073
## 5 -0.66636733   0.39576044
## 6 13.00009691 -28.26234465
```

Notice the information is presented in a 6×2 matrix. The dimension will always by $n \times p$, where each row index corresponds to the observation, and the column index corresponds to the coefficient. The value in row 6 column 1 informs us that removing observation 6 from estimating the model will change $\hat{\beta}_0$ by about 13 standard errors, and change $\hat{\beta}_1$ by about 28 standard errors. In comparison to the $DFBETAS$ for other observations, these values are huge, so we flag observation 6 as influential.

Checking in with the suggested rule for influential observations:

```
##criteria for DFBETAs
2/sqrt(n)
```

```
## [1] 0.8164966
```

So clearly observation 6 is influential, since its $DFBETAs$ are much larger than 0.8165.

4.2 DFFITs

Another measure for influential observations is based on the change in the predicted response when the model is estimated with and without the observation in question. One such measure is $DFFITs$, defined as

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{i(i)}}{\sqrt{S^2_{(i)} h_{ii}}} \tag{14}$$

$$= t_i \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{1/2}. \tag{15}$$

A few notes about $DFFITs$:

- Using (14), $DFFITs$ can be interpreted as the number of standard errors $\hat{y}_i$ changes if observation i is removed when estimating the model.
- $DFFITs$ measures the influence of observation i on its own fitted value.
- As a guide, $|DFFITS_i| > 2\sqrt{p/n}$ considered influential. Again, this should be used in conjunction with the $DFFITs$ for all observations.
- Using (15), we can see that observations with high leverage and are outlying are likely to have high values for $DFFITs$.

Let us go back to the scatterplot in Fig 1c and compute the $DFFITs$:

```
dffits(result.both)
```

```
##           1           2           3           4           5           6
##  -0.16032698  -0.05330067   0.06676822   1.66138579  -0.68764196 -31.11630918
```

The $DFFITs$ for observation 6 is around -31, which is a lot larger in magnitude compared to the other $DFFITs$. We can safely say that it is influential, since its presence or removal changes its predicted response by about 31 standard errors.

Checking in with the suggested rule for influential observations:

```
##criteria for DFFITs
2*sqrt(p/n)
```

```
## [1] 1.154701
```

The magnitude of $DFFITS_6$ is clearly larger than this criteria, so it is influential based on this criteria. Interestingly, the criteria will also flag observation 4 as influential, even though visually it does not appear to be, based on Fig 1c.

4.3 Cook's distance

Another measure that is motivated by the change in fitted values is **Cook's Distance**

$$D_i = \frac{\left(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{(\boldsymbol{i})}\right)' \left(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{(\boldsymbol{i})}\right)}{pMS_{res}} \tag{16}$$

$$= \frac{r_i^2}{p} \frac{h_{ii}}{1 - h_{ii}}. \tag{17}$$

$\hat{\boldsymbol{y}}_{(\boldsymbol{i})}$ denotes the vector of fitted values with observation i removed. A few comments about Cook's distance:

- From (16), we can see that the numerator of Cook's distance measures the Squared Euclidean distance between the vector of fitted values with all observations, $\hat{\boldsymbol{y}}$, and the vector of fitted values with observation i removed, $\hat{\boldsymbol{y}}_{(\boldsymbol{i})}$. So it measures how the fitted values for all observations change, if observation i is removed. This is unlike $DFFITs$, which only measures the change in fitted value for observation i.
- From (17), we can see that observations with high leverage and that are outlying are likely to have large Cook's distances.
- A suggested guideline is that observations with Cook's distance larger than 1 will be flagged as influential.

Let us go back to the scatterplot in Fig 1c and compute the Cook's distances:

```
cooks.distance(result.both)
```

```
##           1           2           3           4           5           6
##  0.016670353  0.001887380  0.002952539  0.328733436  0.213742357 37.555213339
```

The Cook's distance for observation 6 is about 37.5552, which is a lot larger than 1, and a lot larger than the Cook's distance for other observations. So we flag it as an influential observation.

4.4 Some comments about flagged observations

A tendency is to remove observations that have been flagged using any of the measures in this module. Be **EXTREMELY CAREFUL** with deleting such observations. Many times, these observations provide interesting case studies and should always be identified and discussed. Something is making them different from the other observations.

A few other comments:

- Do not get too caught up with the guidelines for flagging these observations. Use these guidelines in relation to the values for all observations as well. Also use context to guide you.
- If it is clear that these observations were data entry errors, fix them.
- If these observations represent unusual circumstances that do not meet the **objectives** of the study, or if there is something fundamentally wrong with the observation, you may delete them, but make sure to clearly explain why. For example, you may be measuring the weights of newborn babies, but a 20 week old baby's weight was entered into the study. Clearly, the 20 week old is not part of the study, so we can remove that baby's weight.
- Sometimes, these observations are flagged due to a predictor or response variable being a lot larger than the rest. Log transforming the corresponding variable can help make the value a lot closer to the rest.
- Fit the model with and without the influential observations and see how differently the models answer our questions of interest.
- You should always address the removal of any observations in your report. Also provide the justifications for their removal.

5 Partial Regression Plots

Another use of residuals is to use them to inform us if we need to transform predictor variables in MLR.

In SLR, we learned how to use scatterplots and residual plots to assess the various regression assumptions, and how to transform the predictor or response variable as needed. For MLR, we still use residual plots in the same way to assess a couple of the regression assumptions:

1. The errors have mean 0.
2. The errors have constant variance.

A challenge in MLR is that if assumption 1 is violated, we know we can remedy this by transforming at least one of the predictors, but we cannot use scatterplots to decide which predictor to transform, and how to transform. The main reason is that scatterplots only take into account one of the predictors and ignoring the other predictors, whereas MLR fits all the predictors into the model at the same time.

A **partial regression plot** illustrates the marginal effect of adding a predictor when the others are already in the model. We can use partial regression plots to decide if a predictor should be added into our MLR model, or if needed, how to transform the predictor. Note that there are a number of other commonly used names for partial regression plots such as added variable plots and marginal effects plots.

Partial regression plots are created in the following manner. Suppose we have predictors $x_1, x_2, \cdots, x_{k-1}$ in the model, and want to decide if we need to add, drop, or transform predictor x_k:

- We regress y against $x_1, x_2, \cdots, x_{k-1}$ and obtain the residuals. Denote these residuals by $e(y|x_1, x_2, \cdots, x_{k-1})$.
- We regress the predictor in question, x_k, against the other predictors in the model and obtain the residuals. Denote these residuals by $e(x_k|x_1, x_2, \cdots, x_{k-1})$.
- Then, we plot $e(x_k|x_1, x_2, \cdots, x_{k-1})$ against $e(y|x_1, x_2, \cdots, x_{k-1})$. This plot is the partial regression plot of x_k.

Usually, we assess the partial regression plot for the following patterns: (1) Random horizontal band (no pattern); or (2) Linear pattern; or (3) Nonlinear pattern.

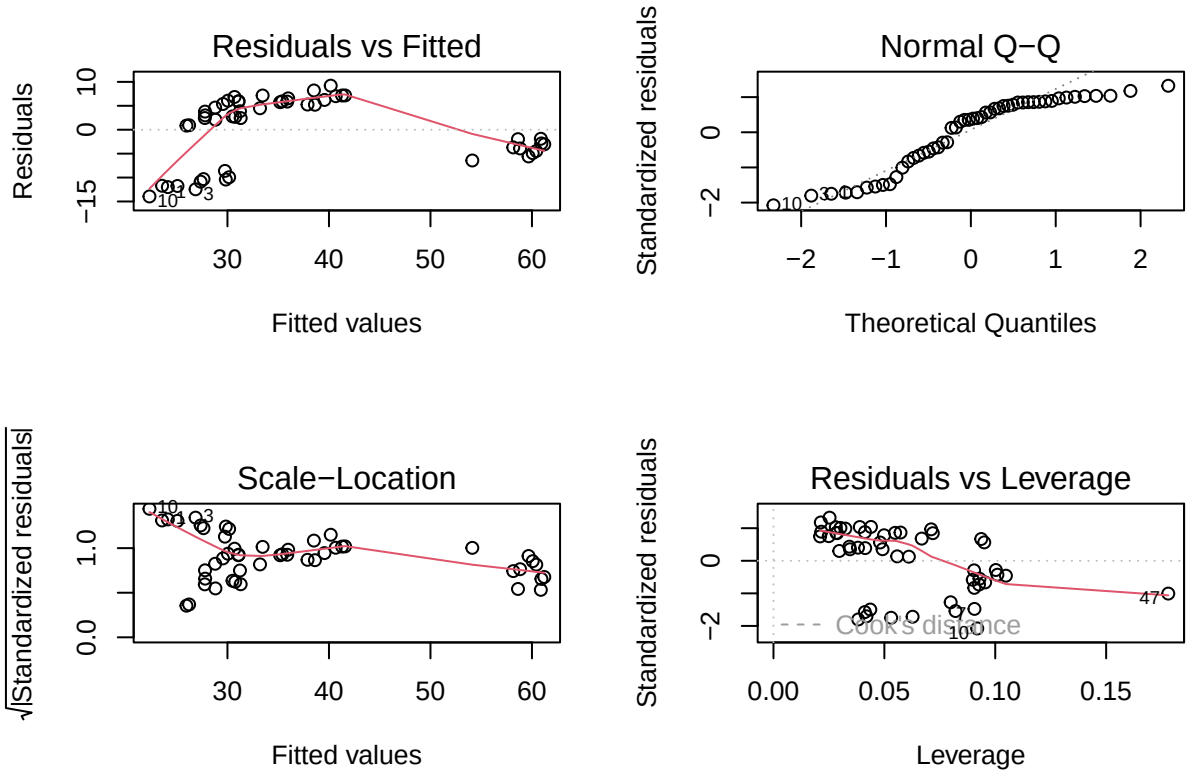
- If we see a random horizontal band, it means we can drop x_k from the model.
- If we see a linear pattern, keep x_k as in the model without transformation.
- If we see a nonlinear pattern, we will need x_k as a predictor, but it needs to be transformed. We use the shape of the partial regression plot to aid us in transforming the predictor.

Some properties of partial regression plots:

- The estimated intercept is 0.
- The estimated slope is the estimated coefficient of x_k in the model with $x_1, x_2, \cdots, x_k$ as predictors.

Let us take a look at an example based on simulated data. For simplicity, we have a response variable and two predictors, x_1, x_2. We consider a MLR model here, and create the corresponding diagnostic plots:

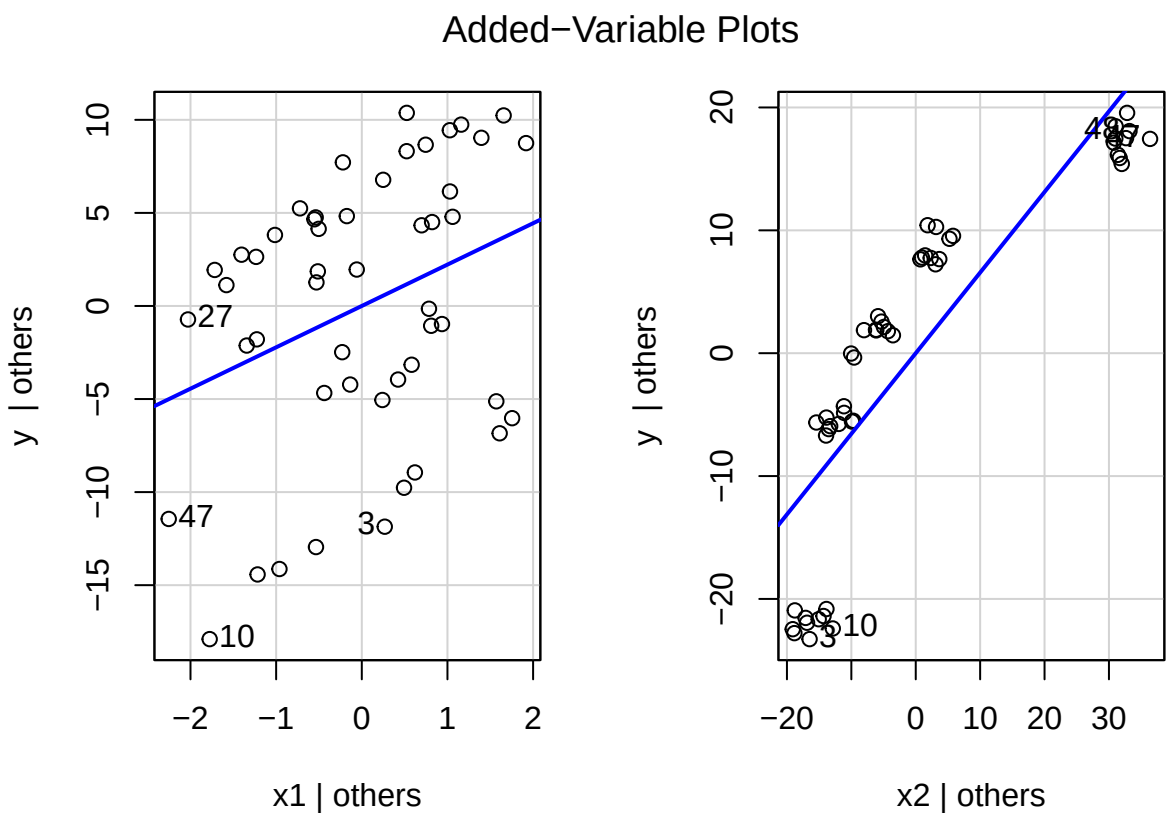
```
##create dataframe
Data<-data.frame(x1,x2,y)

##MLR
result<-lm(y~x1+x2, data=Data)

##residual plot
par(mfrow=c(2,2))
plot(result)
```



Looking at the residual plot at the top left, the assumption that the variance of the errors is constant is reasonably met, as the vertical spread of the residuals is similar throughout the plot. Assumption 1, that the errors have mean 0, is not met. We can see a curved pattern in the residuals and they are not evenly scattered across the horizontal axis. For small and large values on the horizontal axis, the residuals are almost exclusively negative and so do not have a mean of 0. For moderate values on the horizontal axis, the residuals are almost exclusively positive and so do not have a mean of 0.

So we know that we need to transform at least one predictor, since assumption 1 is not met. However, the residual plot does not inform us which predictor to transform, and how. So we create partial regression plots:

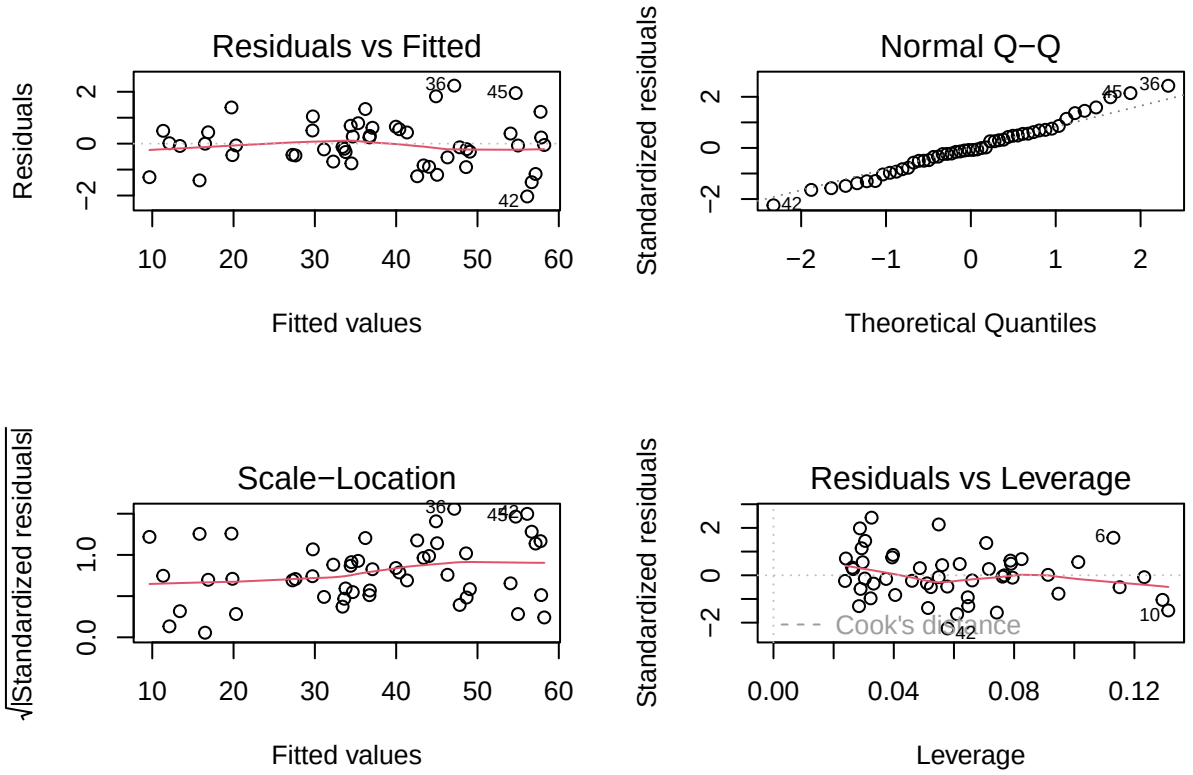
```
library(car)
car::avPlots(result)
```



The first plot on the left is the partial regression plot for x_1, and the second plot on the right is the partial regression plot for x_2.

- Based on the partial regression plot for x_1, we see a clear linear pattern, as the plots are evenly scattered across the blue line. So x_1 does not need to be transformed.
- Based on the partial regression plot for x_2, we see a non linear pattern. For values on the horizontal axis that are small and large, the plots are all below the blue line; however, for values on the horizontal axis that are moderate, the plots are all above the blue line. The shape of this pattern resembles a logarithmic function, we will log transform x_2, refit the regression, and reassess the residual plot.

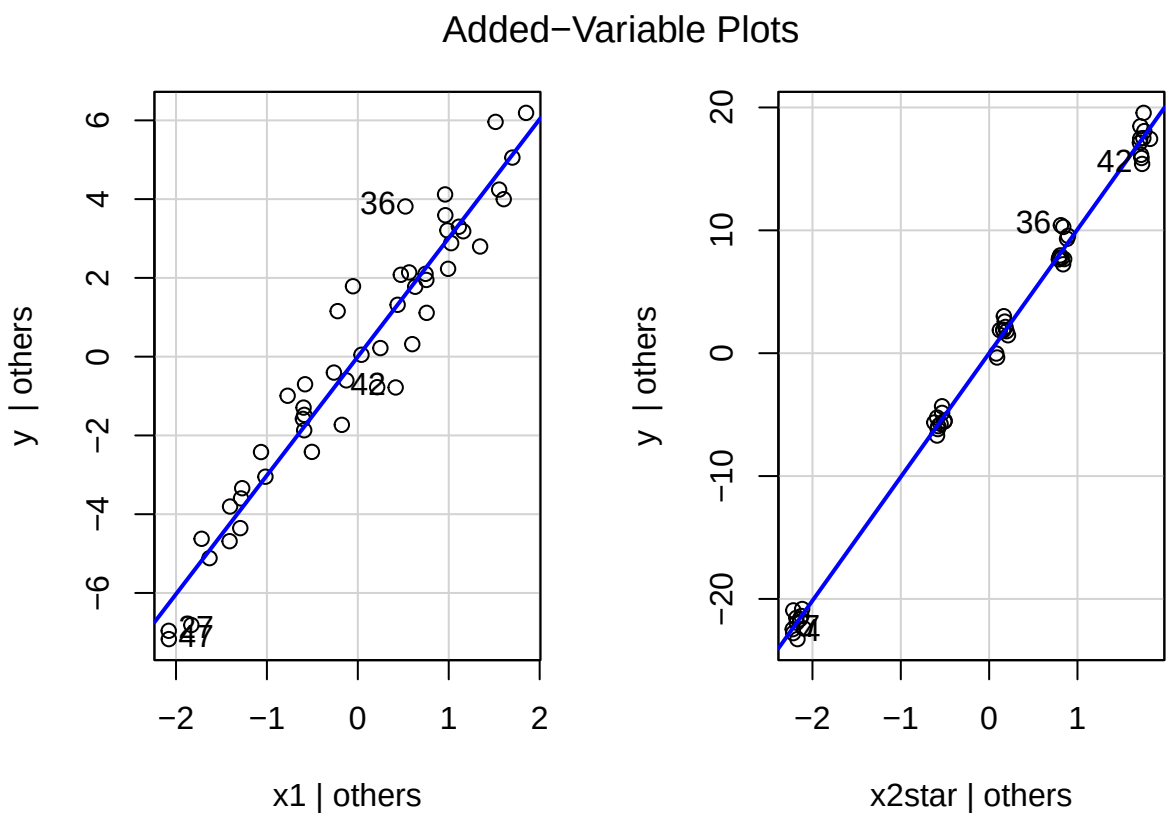
```
##need to transform x2
x2star<-log(x2)
Data<-data.frame(Data,x2star)

##regress using x2star
result2<-lm(y~x1+x2star, data=Data)

##diagnostic plots
par(mfrow=c(2,2))
plot(result2)
```



The residual plot now looks a lot more reasonable. The residuals are evenly scattered across the horizontal axis with constant vertical spread, so assumptions 1 and 2 are met. The transformation of x_2 was successful.

Out of curiosity, we can look at the resulting partial regression plots:

```
##partial regression plots
car::avPlots(result2)
```



It is not surprising to see that for both partial regression plots, the plots are evenly scattered across the blue lines. In fact, the slopes of the blue lines will be the same value as the corresponding estimated coefficient:

```
result2
```

```
##
## Call:
## lm(formula = y ~ x1 + x2star, data = Data)
##
## Coefficients:
## (Intercept)           x1       x2star
##       5.898        3.014       10.072
```

The estimated coefficients for x_1 and x_2 in the MLR are about 3 and 10 respectively, which match with what we see in their corresponding partial regression plots.