Lecture 17

October 16, 2018

Partial Regression and Partial Residual Plots

Partial Regression and Partial Residual Plots

- These plots are often useful for detecting departures from linearity (and also for identifying unusual observations).

Partial Regression and Partial Residual Plots

- These plots are often useful for detecting departures from linearity (and also for identifying unusual observations).
- The simplest plots for detecting departures from linearity are:

Partial Regression and Partial Residual Plots

- These plots are often useful for detecting departures from linearity (and also for identifying unusual observations).
- The simplest plots for detecting departures from linearity are:
- (a) plot the response values y against each explanatory variable values, and

Partial Regression and Partial Residual Plots

- These plots are often useful for detecting departures from linearity (and also for identifying unusual observations).
- The simplest plots for detecting departures from linearity are:
- (a) plot the response values y against each explanatory variable values, and
- (b) plot the residuals $\hat{e}$ against each explanatory variable values.

Partial Regression and Partial Residual Plots

- These plots are often useful for detecting departures from linearity (and also for identifying unusual observations).
- The simplest plots for detecting departures from linearity are:
- (a) plot the response values y against each explanatory variable values, and
- (b) plot the residuals $\hat{e}$ against each explanatory variable values.
- The problem with these plots however is that they only look at the marginal effect of the ith explanatory variable on y and ignore the presence of the other explanatory variables.

Partial Regression and Partial Residual Plots

- These plots are often useful for detecting departures from linearity (and also for identifying unusual observations).
- The simplest plots for detecting departures from linearity are:
- (a) plot the response values y against each explanatory variable values, and
- (b) plot the residuals $\hat{e}$ against each explanatory variable values.
- The problem with these plots however is that they only look at the marginal effect of the ith explanatory variable on y and ignore the presence of the other explanatory variables.
- To correct this, one often looks at partial regression plots (also called added variable plots) and partial residual plots.

- Let $Res(y, X^{-i})$ denote the vector of residuals obtained by regressing y on all the explanatory variables *except* the ith one.

- Let $Res(y, X^{-i})$ denote the vector of residuals obtained by regressing y on all the explanatory variables *except* the ith one.
- Also let $Res(x_i, X^{-i})$ denote the vector of residuals obtained by regressing x_i (x_i is the column of the X-matrix corresponding to the ith explanatory variable) on all the explanatory variables *except* the ith one.

- Let $Res(y, X^{-i})$ denote the vector of residuals obtained by regressing y on all the explanatory variables *except* the ith one.
- Also let $Res(x_i, X^{-i})$ denote the vector of residuals obtained by regressing x_i (x_i is the column of the X-matrix corresponding to the ith explanatory variable) on all the explanatory variables *except* the ith one.
- In the ith partial regression plot, one plots $Res(y, X^{-i})$ against $Res(x_i, X^{-i})$.

- Let $Res(y, X^{-i})$ denote the vector of residuals obtained by regressing y on all the explanatory variables *except* the ith one.
- Also let $Res(x_i, X^{-i})$ denote the vector of residuals obtained by regressing x_i (x_i is the column of the X-matrix corresponding to the ith explanatory variable) on all the explanatory variables *except* the ith one.
- In the ith partial regression plot, one plots $Res(y, X^{-i})$ against $Res(x_i, X^{-i})$.
- This plot therefore looks at the relationship between y and x_i in the presence of all the other explanatory variables.

- Let $Res(y, X^{-i})$ denote the vector of residuals obtained by regressing y on all the explanatory variables *except* the ith one.
- Also let $Res(x_i, X^{-i})$ denote the vector of residuals obtained by regressing x_i (x_i is the column of the X-matrix corresponding to the ith explanatory variable) on all the explanatory variables *except* the ith one.
- In the ith partial regression plot, one plots $Res(y, X^{-i})$ against $Res(x_i, X^{-i})$.
- This plot therefore looks at the relationship between y and x_i in the presence of all the other explanatory variables.
- A remarkable feature of the ith partial regression plot is that if one performs a regression of $Res(y, X^{-i})$ against $Res(x_i, X^{-i})$, one gets the intercept estimate to be zero and the slope estimate will exactly equal $\hat{\beta}_i$.

- Let $Res(y, X^{-i})$ denote the vector of residuals obtained by regressing y on all the explanatory variables *except* the ith one.
- Also let $Res(x_i, X^{-i})$ denote the vector of residuals obtained by regressing x_i (x_i is the column of the X-matrix corresponding to the ith explanatory variable) on all the explanatory variables *except* the ith one.
- In the ith partial regression plot, one plots $Res(y, X^{-i})$ against $Res(x_i, X^{-i})$.
- This plot therefore looks at the relationship between y and x_i in the presence of all the other explanatory variables.
- A remarkable feature of the ith partial regression plot is that if one performs a regression of $Res(y, X^{-i})$ against $Res(x_i, X^{-i})$, one gets the intercept estimate to be zero and the slope estimate will exactly equal $\hat{\beta}_i$.
- This fact can be proved for instance using the block matrix inverse formula (see wikipedia for this formula).

- The partial residual plot is another plot for viewing the relationship between y and x_j in the presence of the other variables.

- The partial residual plot is another plot for viewing the relationship between y and x_j in the presence of the other variables.
- This plot is motivated by the following. Suppose we focus on the j^{th} explanatory variable and are interested in finding out what function $f(\cdot)$ to use in the model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{j-1} x_{i,j-1} + f(x_{ij}) + \beta_{j+1} x_{i,j+1} + \cdots + \beta_p x_{ip} + \epsilon_i.$$

- The partial residual plot is another plot for viewing the relationship between y and x_j in the presence of the other variables.
- This plot is motivated by the following. Suppose we focus on the j^{th} explanatory variable and are interested in finding out what function $f(\cdot)$ to use in the model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{j-1} x_{i,j-1} + f(x_{ij}) + \beta_{j+1} x_{i,j+1} + \cdots + \beta_p x_{ip} + \epsilon_i.$$

- In other words, we are okay with using linear functions for all the variables except for x_j for which we might think that a non-linear function might provide a better model for the response.

- The partial residual plot is another plot for viewing the relationship between y and x_j in the presence of the other variables.
- This plot is motivated by the following. Suppose we focus on the j^{th} explanatory variable and are interested in finding out what function $f(\cdot)$ to use in the model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{j-1} x_{i,j-1} + f(x_{ij}) + \beta_{j+1} x_{i,j+1} + \cdots + \beta_p x_{ip} + \epsilon_i.$$

- In other words, we are okay with using linear functions for all the variables except for x_j for which we might think that a non-linear function might provide a better model for the response.
- In order to find the correct function f, the most natural idea is to look at a scatter plot of $y_i - \sum_{k \neq j} \beta_k x_{ik}$ against x_{ij} for $i = 1, \ldots, n$. This plot would of course reveal the form of the function $f(\cdot)$.

- The problem of course is that we do not know the values $\beta_k, k \neq j$ so we cannot really make this plot.

- The problem of course is that we do not know the values $\beta_k, k \neq j$ so we cannot really make this plot.
- A natural fix is to replace $\beta_k, k \neq j$ by $\hat{\beta}_k, k \neq j$ (here $\hat{\beta}_k$ denotes the estimated regression coefficient of x_k in the full regression of y on all of $x_1, \ldots, x_p$).

- The problem of course is that we do not know the values $\beta_k, k \neq j$ so we cannot really make this plot.
- A natural fix is to replace $\beta_k, k \neq j$ by $\hat{\beta}_k, k \neq j$ (here $\hat{\beta}_k$ denotes the estimated regression coefficient of x_k in the full regression of y on all of $x_1, \dots, x_p$).
- This leads to the plot of $y_i - \sum_{k \neq j} \hat{\beta}_k x_{ik}$ against x_{ij} for $i = 1, \dots, n$. This plot is called the partial residual plot.

- The problem of course is that we do not know the values $\beta_k, k \neq j$ so we cannot really make this plot.
- A natural fix is to replace $\beta_k, k \neq j$ by $\hat{\beta}_k, k \neq j$ (here $\hat{\beta}_k$ denotes the estimated regression coefficient of x_k in the full regression of y on all of $x_1, \dots, x_p$).
- This leads to the plot of $y_i - \sum_{k \neq j} \hat{\beta}_k x_{ik}$ against x_{ij} for $i = 1, \dots, n$. This plot is called the partial residual plot.
- Partial residual because $y_i - \sum_{k \neq j} \hat{\beta}_k x_{ik}$ looks like a partial residual (if we replace $\sum_{k \neq j}$ by $\sum_k$, then we would get the residual).

- The problem of course is that we do not know the values $\beta_k, k \neq j$ so we cannot really make this plot.
- A natural fix is to replace $\beta_k, k \neq j$ by $\hat{\beta}_k, k \neq j$ (here $\hat{\beta}_k$ denotes the estimated regression coefficient of x_k in the full regression of y on all of $x_1, \ldots, x_p$).
- This leads to the plot of $y_i - \sum_{k \neq j} \hat{\beta}_k x_{ik}$ against x_{ij} for $i = 1, \ldots, n$. This plot is called the partial residual plot.
- Partial residual because $y_i - \sum_{k \neq j} \hat{\beta}_k x_{ik}$ looks like a partial residual (if we replace $\sum_{k \neq j}$ by $\sum_k$, then we would get the residual).
- Note also that $y_i - \sum_{k \neq j} \hat{\beta}_k x_{ik} = \hat{e}_i + \hat{\beta}_j x_{ij}$ so this is also a plot of $\hat{e}_i + \hat{\beta}_j x_{ij}$ against x_{ij}. This is why these are also called *component + residual* plots.

- One hopes that the partial residual (or component + residual) plots reveal some natural functions $f(\cdot)$ for use in regression.

- One hopes that the partial residual (or component + residual) plots reveal some natural functions $f(\cdot)$ for use in regression.
- It may be noted that if we fit a regression line to the scatterplot in a partial residual plot, then the slope of the fitted regression like would precisely equal the slope of $\hat{\beta}_j$ in the full multiple regression (do you see why??).

Checking Assumptions on $e_1, \ldots, e_n$

Checking Assumptions on $e_1, \ldots, e_n$

- The confidence intervals and hypothesis tests for $\hat{\beta}_j, j = 0, \ldots, p$ in the linear model depend crucially on the assumption that the errors $e_1, \ldots, e_n$ are independent, have equal variance σ^2 and are normally distributed.

Checking Assumptions on $e_1, \ldots, e_n$

- The confidence intervals and hypothesis tests for $\hat{\beta}_j, j = 0, \ldots, p$ in the linear model depend crucially on the assumption that the errors $e_1, \ldots, e_n$ are independent, have equal variance σ^2 and are normally distributed.
- The errors $e_1, \ldots, e_n$ are of course unobservable. How does one then check the assumptions of independence, constant variance and normality of the errors?

Checking Assumptions on $e_1, \dots, e_n$

- The confidence intervals and hypothesis tests for $\hat{\beta}_j, j = 0, \dots, p$ in the linear model depend crucially on the assumption that the errors $e_1, \dots, e_n$ are independent, have equal variance σ^2 and are normally distributed.
- The errors $e_1, \dots, e_n$ are of course unobservable. How does one then check the assumptions of independence, constant variance and normality of the errors?
- The idea is to use the residuals $\hat{e}_1, \dots, \hat{e}_n$ which act as proxies for the errors. It is important to note that the residuals are not exactly interchangeable with the errors however.

Checking Assumptions on $e_1, \dots, e_n$

- The confidence intervals and hypothesis tests for $\hat{\beta}_j, j = 0, \dots, p$ in the linear model depend crucially on the assumption that the errors $e_1, \dots, e_n$ are independent, have equal variance σ^2 and are normally distributed.
- The errors $e_1, \dots, e_n$ are of course unobservable. How does one then check the assumptions of independence, constant variance and normality of the errors?
- The idea is to use the residuals $\hat{e}_1, \dots, \hat{e}_n$ which act as proxies for the errors. It is important to note that the residuals are not exactly interchangeable with the errors however.
- For example, $var(\hat{e}_i) = \sigma^2(1 - h_{ii})$ where h_{ii} is the ith leverage and $cov(\hat{e}_i, \hat{e}_j) = -\sigma^2 h_{ij}$ where h_{ij} is the (i, j)th entry of the hat matrix.

- Because each h_{ii} lies between 0 and 1 and their average is $(1+p)/n$ which is usually small, in most of the cases, it turns out that each h_{ii} is small.

- Because each h_{ii} lies between 0 and 1 and their average is $(1+p)/n$ which is usually small, in most of the cases, it turns out that each h_{ii} is small.
- Hence each $\hat{e}_i$ has roughly variance equal to σ^2.

- Because each h_{ii} lies between 0 and 1 and their average is $(1 + p)/n$ which is usually small, in most of the cases, it turns out that each h_{ii} is small.
- Hence each $\hat{e}_i$ has roughly variance equal to σ^2.
- Similarly, because $\sum_{j=1}^{n} h_{ij}^2 = h_{ii}$ for each i, it also turns out that h_{ij} is typically close to zero for most i and j.

- Because each h_{ii} lies between 0 and 1 and their average is $(1+p)/n$ which is usually small, in most of the cases, it turns out that each h_{ii} is small.
- Hence each $\hat{e}_i$ has roughly variance equal to σ^2.
- Similarly, because $\sum_{j=1}^{n} h_{ij}^2 = h_{ii}$ for each i, it also turns out that h_{ij} is typically close to zero for most i and j.
- Thus the residuals $\hat{e}_1, \dots, \hat{e}_n$ have variance roughly equal to σ^2 and correlation roughly equal to zero.

- Because each h_{ii} lies between 0 and 1 and their average is $(1+p)/n$ which is usually small, in most of the cases, it turns out that each h_{ii} is small.
- Hence each $\hat{e}_i$ has roughly variance equal to σ^2.
- Similarly, because $\sum_{j=1}^{n} h_{ij}^2 = h_{ii}$ for each i, it also turns out that h_{ij} is typically close to zero for most i and j.
- Thus the residuals $\hat{e}_1, \dots, \hat{e}_n$ have variance roughly equal to σ^2 and correlation roughly equal to zero.
- This is true under the assumption that $e_1, \dots, e_n$ are independent with variance σ^2. The residuals can therefore be used to test assumptions on $e_1, \dots, e_n$. Alternately, one might use standardized residuals.

Constant Variance Assumption

Constant Variance Assumption

- Plot residuals (y-axis) against the fitted values (x-axis).

Constant Variance Assumption

- Plot residuals (y-axis) against the fitted values (x-axis).
- If all is well, you should see constant variance in the vertical direction and the scatter should be symmetric vertically about zero.

Constant Variance Assumption

- Plot residuals (y-axis) against the fitted values (x-axis).
- If all is well, you should see constant variance in the vertical direction and the scatter should be symmetric vertically about zero.
- Things to look for are heteroscedasticity (nonconstant variance) and nonlinearity (which indicates that some change in the model is necessary).

Constant Variance Assumption

- Plot residuals (y-axis) against the fitted values (x-axis).
- If all is well, you should see constant variance in the vertical direction and the scatter should be symmetric vertically about zero.
- Things to look for are heteroscedasticity (nonconstant variance) and nonlinearity (which indicates that some change in the model is necessary).
- Also plot the residuals (y-axis) against each explanatory variable values (for explanatory variables that are both in and out of the model; we will be looking at variable selection methods later).

Constant Variance Assumption

- Plot residuals (y-axis) against the fitted values (x-axis).
- If all is well, you should see constant variance in the vertical direction and the scatter should be symmetric vertically about zero.
- Things to look for are heteroscedasticity (nonconstant variance) and nonlinearity (which indicates that some change in the model is necessary).
- Also plot the residuals (y-axis) against each explanatory variable values (for explanatory variables that are both in and out of the model; we will be looking at variable selection methods later).
- Look for the same things as the residuals against fitted values plot; except that in the case of plots against explanatory variables that are not in the model, look for any relationship that might indicate that this explanatory variable should be included.

- If indeed there is some evidence of nonconstant variance, two common ways of dealing with it are (a) using weighted least squares and (b) using variable transformations.

- If indeed there is some evidence of nonconstant variance, two common ways of dealing with it are (a) using weighted least squares and (b) using variable transformations.
- We will look at weighted least squares later. The most common variable transformations are taking powers (most common power is square root) and logarithms. We shall look at these in the next class.