Checking Assumptions

Merlise Clyde

STA721 Linear Models

Duke University

October 8, 2012

Linear Model

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Linear Model

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

Linear Model

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\boldsymbol{\mu} \in C(\mathbf{X}) \quad \Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\beta$$

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\begin{aligned} \boldsymbol{\mu} \in C(\mathbf{X}) \quad &\Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\beta \\ \boldsymbol{\epsilon} \quad &\sim \quad \mathsf{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n) \end{aligned}$$

Linear Model

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\begin{aligned} \boldsymbol{\mu} \in C(\mathbf{X}) \quad &\Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\beta \\ \boldsymbol{\epsilon} \quad &\sim \quad \mathsf{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n) \end{aligned}$$

Focus on

- Wrong mean for a case or cases

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\begin{aligned} \boldsymbol{\mu} \in C(\mathbf{X}) \quad &\Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\beta \\ \boldsymbol{\epsilon} \quad &\sim \quad \mathsf{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n) \end{aligned}$$

Focus on

- Wrong mean for a case or cases
- Wrong distribution for $\boldsymbol{\epsilon}$

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\begin{aligned} \boldsymbol{\mu} \in C(\mathbf{X}) \quad &\Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\beta \\ \boldsymbol{\epsilon} \quad &\sim \quad \mathsf{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n) \end{aligned}$$

Focus on

- Wrong mean for a case or cases
- Wrong distribution for $\boldsymbol{\epsilon}$
- Cases that influence the mean

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\begin{aligned} \boldsymbol{\mu} \in C(\mathbf{X}) \quad &\Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\beta \\ \boldsymbol{\epsilon} \quad &\sim \quad \mathrm{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n) \end{aligned}$$

Focus on

- Wrong mean for a case or cases
- Wrong distribution for $\boldsymbol{\epsilon}$
- Cases that influence the mean

If $\mu_i \neq \mathbf{x}_i^T \beta$ then expected value of $e_i = Y_i - \hat{Y}_i$ is not zero;

Standardized residuals

standardized residuals

$$r_i = e_i / \sqrt{\sigma^2 (1 - h_{ii})}$$

- Under correct model have mean 0 and scale 1

Standardized residuals

standardized residuals

$$r_i = e_i / \sqrt{\sigma^2(1 - h_{ii})}$$

- Under correct model have mean 0 and scale 1
- Use estimate of σ^2

Standardized residuals

standardized residuals

$$r_i = e_i / \sqrt{\sigma^2(1 - h_{ii})}$$

- Under correct model have mean 0 and scale 1
- Use estimate of σ^2
- The distribution of r_i is not a t

Standardized residuals

standardized residuals

$$r_i = e_i / \sqrt{\sigma^2(1 - h_{ii})}$$

- Under correct model have mean 0 and scale 1
- Use estimate of σ^2
- The distribution of r_i is not a t
- if h_{ii} (leverage) is close to 1, then $\hat{Y}_i$ is close to Y_i so e_i is approximately 0

Standardized residuals

standardized residuals

$$r_i = e_i / \sqrt{\sigma^2 (1 - h_{ii})}$$

- Under correct model have mean 0 and scale 1
- Use estimate of σ^2
- The distribution of r_i is not a t
- if h_{ii} (leverage) is close to 1, then $\hat{Y}_i$ is close to Y_i so e_i is approximately 0
- Variance is also almost 0, so standardize value may not flag "outliers"

Predicted Residuals

Estimates without Case (i):

$$\hat{\beta}_{(i)} = (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^T \mathbf{Y}_{(i)}$$

Predicted Residuals

Estimates without Case (i):

$$
\begin{aligned}
\hat{\beta}_{(i)} &= (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^T \mathbf{Y}_{(i)} \\
&= \hat{\beta} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}}
\end{aligned}
$$

Predicted Residuals

Estimates without Case (i):

$$\begin{aligned}\hat{\beta}_{(i)} &= (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^T \mathbf{Y}_{(i)} \\ &= \hat{\beta} - \frac{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i e_i}{1 - h_{ii}}\end{aligned}$$

Predicted residual

$$e_{(i)} = y_i - \mathbf{x}_i^T \hat{\beta}(i) = \frac{e_i}{1 - h_{ii}}$$

Predicted Residuals

Estimates without Case (i):

$$
\begin{aligned}
\hat{\beta}_{(i)} &= (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^T \mathbf{Y}_{(i)} \\
&= \hat{\beta} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}}
\end{aligned}
$$

Predicted residual

$$
e_{(i)} = y_i - \mathbf{x}_i^T \hat{\beta}(i) = \frac{e_i}{1 - h_{ii}}
$$

with variance

$$
\mathsf{var}(e_{(i)}) = \frac{\sigma^2}{1 - h_{ii}}
$$

Predicted Residuals

Estimates without Case (i):

$$
\begin{aligned}
\hat{\beta}_{(i)} &= (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^T \mathbf{Y}_{(i)} \\
&= \hat{\beta} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}}
\end{aligned}
$$

Predicted residual

$$
e_{(i)} = y_i - \mathbf{x}_i^T \hat{\beta}(i) = \frac{e_i}{1 - h_{ii}}
$$

with variance

$$
\mathrm{var}(e_{(i)}) = \frac{\sigma^2}{1 - h_{ii}}
$$

Standardized predicted residual is

$$
\frac{e_{(i)}}{\sqrt{\mathrm{var}(e_{(i)})}} = \frac{e_i/(1 - h_{ii})}{\sigma/\sqrt{1 - h_{ii}}} = \frac{e_i}{\sigma\sqrt{1 - h_{ii}}}
$$

Predicted Residuals

Estimates without Case (i):

$$\begin{aligned}\hat{\beta}_{(i)} &= (\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}\mathbf{X}_{(i)}^T\mathbf{Y}_{(i)} \\ &= \hat{\beta} - \frac{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i e_i}{1 - h_{ii}}\end{aligned}$$

Predicted residual

$$e_{(i)} = y_i - \mathbf{x}_i^T\hat{\beta}(i) = \frac{e_i}{1 - h_{ii}}$$

with variance

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

Standardized predicted residual is

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

Can show that these are the same as standardized residual!

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

$$\text{SSE}_{(i)} = \text{SSE} - \frac{e_i^2}{1 - h_{ii}}$$

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

$$\begin{aligned} \text{SSE}_{(i)} &= \text{SSE} - \frac{e_i^2}{1 - h_{ii}} \\ \hat{\sigma}^2_{(i)} = \text{MSE}_{(i)} &= \frac{\text{SSE}_{(i)}}{n - p - 1} \end{aligned}$$

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

$$
\begin{aligned}
\text{SSE}_{(i)} &= \text{SSE} - \frac{e_i^2}{1 - h_{ii}} \\
\hat{\sigma}^2_{(i)} = \text{MSE}_{(i)} &= \frac{\text{SSE}_{(i)}}{n - p - 1}
\end{aligned}
$$

Externally Standardized residuals

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

$$\begin{aligned} \text{SSE}_{(i)} &= \text{SSE} - \frac{e_i^2}{1 - h_{ii}} \\ \hat{\sigma}^2_{(i)} = \text{MSE}_{(i)} &= \frac{\text{SSE}_{(i)}}{n - p - 1} \end{aligned}$$

Externally Standardized residuals

$$t_i = \frac{e_{(i)}}{\sqrt{\hat{\sigma}^2_{(i)}/(1 - h_{ii})}} = \frac{y_i - \mathbf{x}_i^T \hat{\beta}_{(i)}}{\sqrt{\hat{\sigma}^2_{(i)}/(1 - h_{ii})}}$$

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

$$\begin{aligned} \text{SSE}_{(i)} &= \text{SSE} - \frac{e_i^2}{1 - h_{ii}} \\ \hat{\sigma}^2_{(i)} = \text{MSE}_{(i)} &= \frac{\text{SSE}_{(i)}}{n - p - 1} \end{aligned}$$

Externally Standardized residuals

$$t_i = \frac{e_{(i)}}{\sqrt{\hat{\sigma}^2_{(i)}/(1 - h_{ii})}} = \frac{y_i - \mathbf{x}_i^T \hat{\beta}_{(i)}}{\sqrt{\hat{\sigma}^2_{(i)}/(1 - h_{ii})}} = r_i \left(\frac{n - p - 1}{n - p - r_i^2} \right)^{1/2}$$

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

$$\begin{aligned}
\mathrm{SSE}_{(i)} &= \mathrm{SSE} - \frac{e_i^2}{1-h_{ii}} \\
\hat{\sigma}^2_{(i)} = \mathrm{MSE}_{(i)} &= \frac{\mathrm{SSE}_{(i)}}{n-p-1}
\end{aligned}$$

Externally Standardized residuals

$$t_i = \frac{e_{(i)}}{\sqrt{\hat{\sigma}^2_{(i)}/(1-h_{ii})}} = \frac{y_i - \mathbf{x}_i^T \hat{\beta}_{(i)}}{\sqrt{\hat{\sigma}^2_{(i)}/(1-h_{ii})}} = r_i \left(\frac{n-p-1}{n-p-r_i^2} \right)^{1/2}$$

May still miss extreme points with high leverage, but will pick up unusual y_is

Distribution of Externally Studentized Residual

$$t_i = \frac{e_{(i)}}{\sqrt{\hat{\sigma}^2_{(i)}/(1-h_{ii})}} = \frac{y_i - \mathbf{x}_i^T \hat{\beta}_{(i)}}{\sqrt{\hat{\sigma}^2_{(i)}/(1-h_{ii})}} \sim \mathrm{St}(n-p-1)$$

Outlier Test

H_0: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H_a: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

Outlier Test

H_0: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H_a: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

- Show that t-test for testing H_0: $\alpha_i = 0$ is equal to t_i

Outlier Test

H$_0$: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H$_a$: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

- Show that t-test for testing H$_0$: $\alpha_i = 0$ is equal to t_i
- if p-value is small declare the ith case to be an outlier: E$[Y_i]$ not given by $\mathbf{X}\beta$ but $\mathbf{X}\beta + \delta_i \alpha_i$

Outlier Test

H_0: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H_a: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

- Show that t-test for testing H_0: $\alpha_i = 0$ is equal to t_i
- if p-value is small declare the ith case to be an outlier: $E[Y_i]$ not given by $\mathbf{X}\beta$ but $\mathbf{X}\beta + \delta_i \alpha_i$
- Can extend to include multiple δ_i and δ_j to test that case i and j are both outliers

Outlier Test

H$_0$: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H$_a$: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

- Show that t-test for testing H$_0$: $\alpha_i = 0$ is equal to t_i
- if p-value is small declare the ith case to be an outlier: E$[Y_i]$ not given by $\mathbf{X}\boldsymbol{\beta}$ but $\mathbf{X}\boldsymbol{\beta} + \delta_i \alpha_i$
- Can extend to include multiple δ_i and δ_j to test that case i and j are both outliers
- Extreme case $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{I}_n \boldsymbol{\alpha}$ all points have their own mean!

Outlier Test

H_0: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H_a: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

- Show that t-test for testing H_0: $\alpha_i = 0$ is equal to t_i
- if p-value is small declare the ith case to be an outlier: $\mathrm{E}[Y_i]$ not given by $\mathbf{X}\beta$ but $\mathbf{X}\beta + \delta_i \alpha_i$
- Can extend to include multiple δ_i and δ_j to test that case i and j are both outliers
- Extreme case $\boldsymbol{\mu} = \mathbf{X}\beta + \mathbf{I}_n \boldsymbol{\alpha}$ all points have their own mean!
- Control for multiple testing

Outlier Test

H_0: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H_a: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

- Show that t-test for testing H_0: $\alpha_i = 0$ is equal to t_i
- if p-value is small declare the ith case to be an outlier: $E[Y_i]$ not given by $\mathbf{X}\beta$ but $\mathbf{X}\beta + \delta_i \alpha_i$
- Can extend to include multiple δ_i and δ_j to test that case i and j are both outliers
- Extreme case $\boldsymbol{\mu} = \mathbf{X}\beta + \mathbf{I}_n \boldsymbol{\alpha}$ all points have their own mean!
- Control for multiple testing

Multiple Testing

look to see if the $\max |t_i|$ is larger than expected under the null

Multiple Testing

look to see if the max $|t_i|$ is larger than expected under the null

- Need distribution of the max of Student t random variables (simulate)

Multiple Testing

look to see if the max $|t_i|$ is larger than expected under the null

- Need distribution of the max of Student t random variables (simulate)
- Conservative is Bonferroni Correction: For n tests of size α the probability of falsely labeling at least one case as an outlier is no greater than $n\alpha$; e.g. with 21 cases and $\alpha = 0.05$, the probability is no greater than 1.05!

Multiple Testing

look to see if the max $|t_i|$ is larger than expected under the null

- Need distribution of the max of Student t random variables (simulate)
- Conservative is Bonferroni Correction: For n tests of size α the probability of falsely labeling at least one case as an outlier is no greater than $n\alpha$; e.g. with 21 cases and $\alpha = 0.05$, the probability is no greater than 1.05!
- Adjust $\alpha^* = \alpha/n$ so that the probability of falsely labeling at least one point an outlier is α $\alpha/n = .00238$ with $n = 21$

Multiple Testing

look to see if the max $|t_i|$ is larger than expected under the null

- Need distribution of the max of Student t random variables (simulate)
- Conservative is Bonferroni Correction: For n tests of size α the probability of falsely labeling at least one case as an outlier is no greater than $n\alpha$; e.g. with 21 cases and $\alpha = 0.05$, the probability is no greater than 1.05!
- Adjust $\alpha^* = \alpha/n$ so that the probability of falsely labeling at least one point an outlier is α $\alpha/n = .00238$ with $n = 21$

Cook's Distance

Cook's Distance measure of how much predictions change with ith case deleted

Cook's Distance

Cook's Distance measure of how much predictions change with ith case deleted

$$D_i = \frac{\|\mathbf{Y}_{(i)} - \mathbf{Y}\|^2}{p\hat{\sigma}^2} = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\beta}_{(i)} - \hat{\beta})}{p\hat{\sigma}^2}$$

Cook's Distance

Cook's Distance measure of how much predictions change with ith case deleted

$$
\begin{aligned}
D_i &= \frac{\|\mathbf{Y}_{(i)} - \mathbf{Y}\|^2}{p\hat{\sigma}^2} = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\beta}_{(i)} - \hat{\beta})}{p\hat{\sigma}^2} \\
&= \frac{r_i^2}{p} \frac{h_{ii}}{1 - h_{ii}}
\end{aligned}
$$

Cook's Distance

Cook's Distance measure of how much predictions change with ith case deleted

$$
\begin{aligned}
D_i &= \frac{\|\mathbf{Y}_{(i)} - \mathbf{Y}\|^2}{p\hat{\sigma}^2} = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\beta}_{(i)} - \hat{\beta})}{p\hat{\sigma}^2} \\
&= \frac{r_i^2}{p} \frac{h_{ii}}{1 - h_{ii}}
\end{aligned}
$$

Flag cases where $D_i > 1$ or large relative to other cases

Cook's Distance

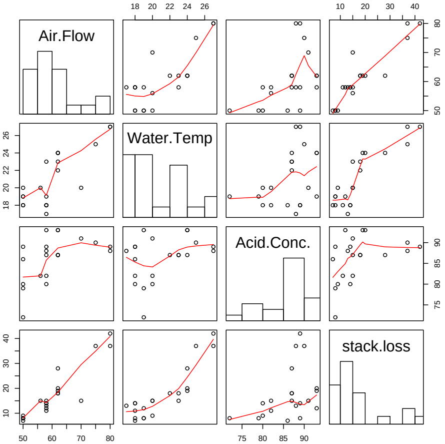
Cook's Distance measure of how much predictions change with ith case deleted

$$\begin{aligned} D_i &= \frac{\|\mathbf{Y}_{(i)} - \mathbf{Y}\|^2}{p\hat{\sigma}^2} = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\beta}_{(i)} - \hat{\beta})}{p\hat{\sigma}^2} \\ &= \frac{r_i^2}{p} \frac{h_{ii}}{1 - h_{ii}} \end{aligned}$$

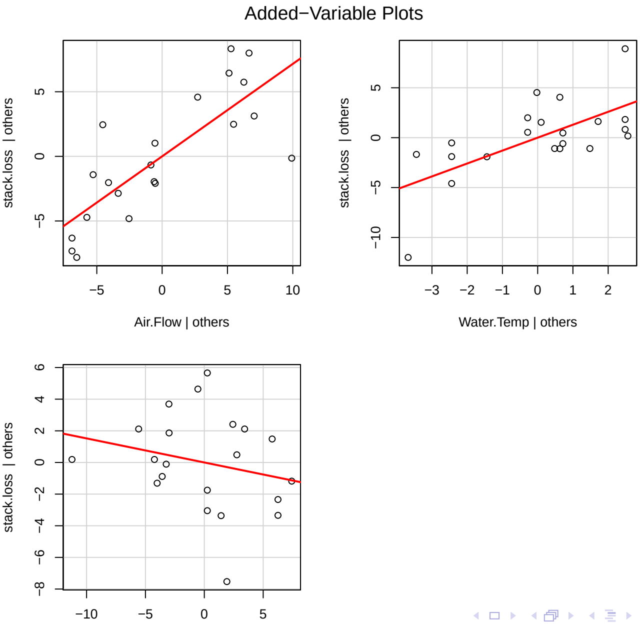
Flag cases where $D_i > 1$ or large relative to other cases

Influential Cases

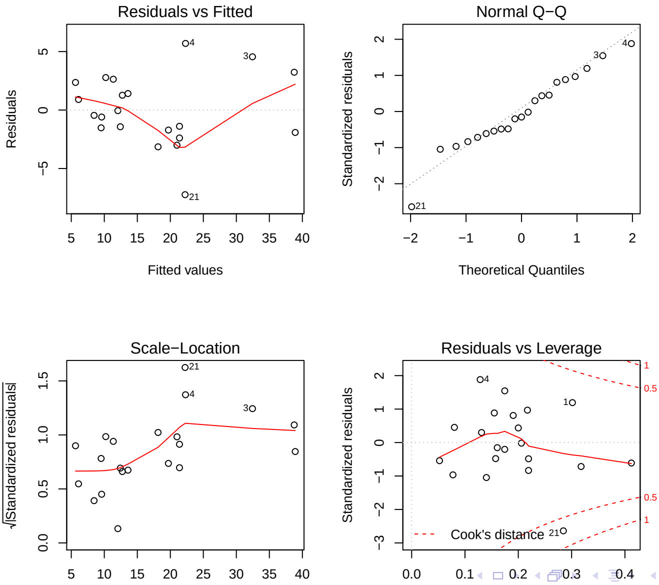
Stackloss Data



Stackloss Added Variable Plot



Stackloss Data Again



Case 21

- Leverage 0.285 (compare to $p/n = .19$)

Case 21

- Leverage 0.285 (compare to $p/n = .19$)
- p-value t_{21} is 0.0042

Case 21

- Leverage 0.285 (compare to $p/n = .19$)
- p-value t_{21} is 0.0042
- Bonferroni adjusted p-value is 0.0024

Case 21

- Leverage 0.285 (compare to $p/n = .19$)
- p-value t_{21} is 0.0042
- Bonferroni adjusted p-value is 0.0024
- Cooks' Distance .69

Case 21

- Leverage 0.285 (compare to $p/n = .19$)
- p-value t_{21} is 0.0042
- Bonferroni adjusted p-value is 0.0024
- Cooks' Distance .69
- Other points? Masking?

Case 21

- Leverage 0.285 (compare to $p/n = .19$)
- p-value t_{21} is 0.0042
- Bonferroni adjusted p-value is 0.0024
- Cooks' Distance .69
- Other points? Masking?
- Refit without Case 21

Case 21

- Leverage 0.285 (compare to $p/n = .19$)
- p-value t_{21} is 0.0042
- Bonferroni adjusted p-value is 0.0024
- Cooks' Distance .69
- Other points? Masking?
- Refit without Case 21

Other analyses have suggested that cases (1, 2, 3, 4, 21) are outliers

Multiple Outliers

- Hoeting, Madigan and Raftery (in various permutations) consider the problem of simultaneous variable selection and outlier identification.

Multiple Outliers

- Hoeting, Madigan and Raftery (in various permutations) consider the problem of simultaneous variable selection and outlier identification.
- This is implemented in the library(BMA) in the function MC3.REG. This has the advantage that more than 2 points may be considered as outliers at the same time.

Multiple Outliers

- Hoeting, Madigan and Raftery (in various permutations) consider the problem of simultaneous variable selection and outlier identification.
- This is implemented in the library(BMA) in the function MC3.REG. This has the advantage that more than 2 points may be considered as outliers at the same time.
- The function uses a Markov chain to identify both important variables and potential outliers, but is coded in Fortran so should run reasonably quickly.

Multiple Outliers

- Hoeting, Madigan and Raftery (in various permutations) consider the problem of simultaneous variable selection and outlier identification.
- This is implemented in the library(BMA) in the function MC3.REG. This has the advantage that more than 2 points may be considered as outliers at the same time.
- The function uses a Markov chain to identify both important variables and potential outliers, but is coded in Fortran so should run reasonably quickly.
- Can also use BAS or other variable selection programs

Multiple Outliers

- Hoeting, Madigan and Raftery (in various permutations) consider the problem of simultaneous variable selection and outlier identification.
- This is implemented in the library(BMA) in the function MC3.REG. This has the advantage that more than 2 points may be considered as outliers at the same time.
- The function uses a Markov chain to identify both important variables and potential outliers, but is coded in Fortran so should run reasonably quickly.
- Can also use BAS or other variable selection programs

Using BAS

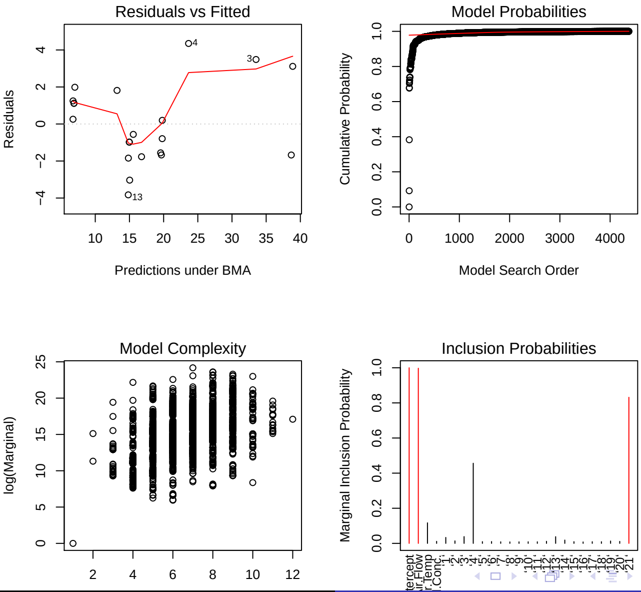
```
library(MASS)
data(stackloss)
n = nrow(stackloss)
stack.out = cbind(stackloss, diag(n))

library(BAS)
BAS.stack = bas.lm(stack.loss ~ ., data=stack.out,
                   prior="hyper-g-n", a=3,
                   modelprior=beta.binomial(1, ncol(stack.c
                   n.models=2^20,
                   method="MCMC", MCMC.it=2^21)
```

Output

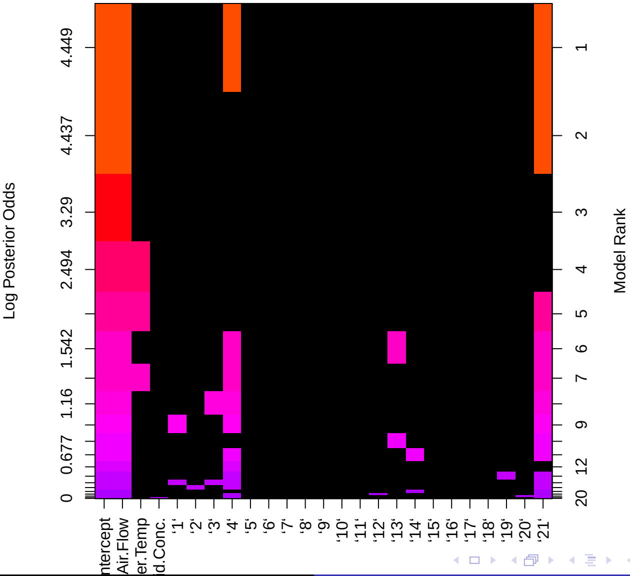
```
> summary(BAS.stack)
     Intercept Air.Flow Water.Temp Acid.Conc. `1` `2` `3` `4` `5` `6` `7` `8`
[1,]         1        1          0          0   0   0   0   1   0   0   0   0
[2,]         1        1          0          0   0   0   0   0   0   0   0   0
[3,]         1        1          0          0   0   0   0   0   0   0   0   0
[4,]         1        1          1          0   0   0   0   0   0   0   0   0
[5,]         1        1          1          0   0   0   0   0   0   0   0   0
     `9` `10` `11` `12` `13` `14` `15` `16` `17` `18` `19` `20` `21`
[1,]   0    0    0    0    0    0    0    0    0    0    0    0    1
[2,]   0    0    0    0    0    0    0    0    0    0    0    0    1
[3,]   0    0    0    0    0    0    0    0    0    0    0    0    0
[4,]   0    0    0    0    0    0    0    0    0    0    0    0    0
[5,]   0    0    0    0    0    0    0    0    0    0    0    0    1
               BF PostProbs     R2 dim  logmarg
[1,] 1.0000000000    0.2940 0.9605   4 22.16893
[2,] 0.0644223528    0.2904 0.9271   3 19.42664
[3,] 0.0008704905    0.0922 0.8458   2 15.12248
[4,] 0.0092248619    0.0416 0.9088   3 17.48308
[5,] 0.0845728129    0.0249 0.9466   4 19.69879
>
```

BAS



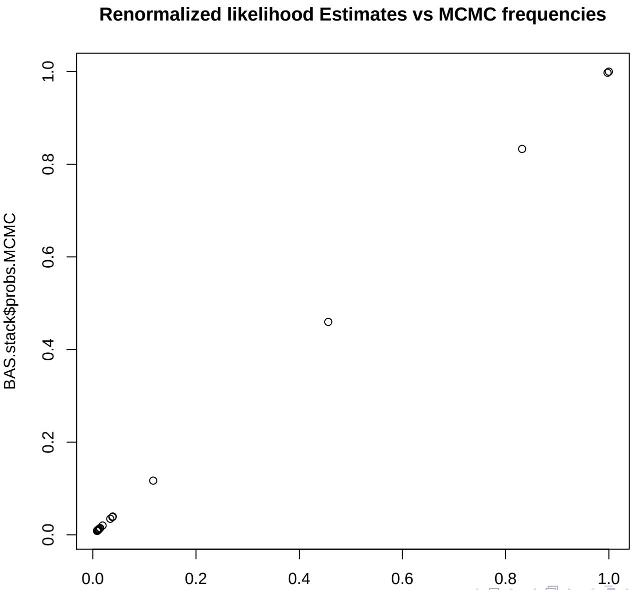
Residuals vs Fitted
Residuals
Predictions under BMA
4
3
13
Model Probabilities
Cumulative Probability
Model Search Order
Model Complexity
log(Marginal)
Inclusion Probabilities
Marginal Inclusion Probability

BAS



Log Posterior Odds
4.449
4.437
3.29
2.494
1.542
1.16
0.677
0
Model Rank
1
2
3
4
5
6
7
9
12
20
'1'
'2'
'3'
'4'
'5'
'6'
'7'
'8'
'9'
'10'
'11'
'12'
'13'
'14'
'15'
'16'
'17'
'18'
'19'
'20'
'21'

Diagnostics



To Remove or Not?

- For suspicious cases, check data sources for errors

To Remove or Not?

- For suspicious cases, check data sources for errors
- Check that points are not outliers because of wrong mean function or distributional assumptions

To Remove or Not?

- For suspicious cases, check data sources for errors
- Check that points are not outliers because of wrong mean function or distributional assumptions
- Investigate need for transformations (use EDA at several stages)
- Influential cases - report results with and without cases (results may change - are differences meaningful?)

To Remove or Not?

- For suspicious cases, check data sources for errors
- Check that points are not outliers because of wrong mean function or distributional assumptions
- Investigate need for transformations (use EDA at several stages)
- Influential cases - report results with and without cases (results may change - are differences meaningful?)
- Outlier test - suggests alternative population for the case(s); if not influential may in keep analysis, but will inflate $\hat{\sigma}^2$ and interval estimates

To Remove or Not?

- For suspicious cases, check data sources for errors
- Check that points are not outliers because of wrong mean function or distributional assumptions
- Investigate need for transformations (use EDA at several stages)
- Influential cases - report results with and without cases (results may change - are differences meaningful?)
- Outlier test - suggests alternative population for the case(s); if not influential may in keep analysis, but will inflate $\hat{\sigma}^2$ and interval estimates
- Document how you handle any case deletions - reproducibility!

To Remove or Not?

- For suspicious cases, check data sources for errors
- Check that points are not outliers because of wrong mean function or distributional assumptions
- Investigate need for transformations (use EDA at several stages)
- Influential cases - report results with and without cases (results may change - are differences meaningful?)
- Outlier test - suggests alternative population for the case(s); if not influential may in keep analysis, but will inflate $\hat{\sigma}^2$ and interval estimates
- Document how you handle any case deletions - reproducibility!
- Robust Regression Methods

To Remove or Not?

- For suspicious cases, check data sources for errors
- Check that points are not outliers because of wrong mean function or distributional assumptions
- Investigate need for transformations (use EDA at several stages)
- Influential cases - report results with and without cases (results may change - are differences meaningful?)
- Outlier test - suggests alternative population for the case(s); if not influential may in keep analysis, but will inflate $\hat{\sigma}^2$ and interval estimates
- Document how you handle any case deletions - reproducibility!
- Robust Regression Methods