Linear Regression Models

STATS 762 – Lecture Slides 5

March 20, 2019

Model Assumptions

For ordinary regression the model assumptions are:

1. A linear regression surface that relates the explanatory variables to the expected value of the response.
2. Errors are independent.
3. Errors have constant variance.
4. Errors are Normally distributed.

It is important to remember that these assumptions are not essential in order to have a useful model.

Model Assumptions (cont.)

The linear regression surface is the most important of these assumptions as it impacts the interpretation and accuracy of the fitted model.

The other three assumptions do not affect the validity of the fitted model but they will affect the sampling distributions for the fitted coefficients and thus statistical inference that depends on these sampling distributions.

- The fitted surface is still a valid approximation of the relationship between the expected value of the response and the regressors. Further $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.
- The sampling distribution of $\hat{\boldsymbol{\beta}}$ is not necessarily (exactly) Normal and its covariance matrix is not given by $\sigma^2 \left(\mathbf{X}^t\mathbf{X}\right)^{-1}$.

Impact of the Error Assumptions

Much of the statistical inference we do is based on the sampling distribution of $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} \sim \mathrm{N}\left(\boldsymbol{\beta},\ \ \sigma^2 \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\right).$$

This distribution was derived from:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{Y}$$

which leads to

$$\mathrm{E}\left(\hat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\boldsymbol{\mu}_{\mathbf{Y}} \ \text{ and } \ \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{X}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}.$$

Impact of the Error Assumptions (cont.)

If $\mathbf{\Sigma_Y} = \sigma^2\mathbf{I}$ then $\mathbf{\Sigma}_{\boldsymbol{\beta}} = \sigma^2 \left(\mathbf{X}^t\mathbf{X}\right)^{-1}$.

- However, if the observations are correlated or have non-constant variance then $\mathbf{\Sigma_Y} \neq \sigma^2\mathbf{I}$ and the expression $\mathbf{\Sigma}_{\boldsymbol{\beta}} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{\Sigma_Y}\mathbf{X}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}$ applies.
- $\hat{\boldsymbol{\beta}}$ has exactly a Normal distribution if $\mathbf{Y}$ has a Normal distribution but the central limit theorem ensures it will be asymptotically Normal regardless.

Impact of the Error Assumptions (cont.)

- If $\mathrm{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ then $\mathrm{E}\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$ for any $\boldsymbol{\Sigma}_{\mathbf{Y}}$ and any distribution of $\mathbf{Y}$.
- Inferences concerning $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ are insensitive to the distribution of $\mathbf{Y}$ due to the central limit theorem – i.e. the sampling distributions of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\mu}}$ will almost always be Normal to a very good approximation.
- The only situation where the distribution of $\mathbf{Y}$ is of real concern is if you want to form prediction intervals as these are predictions about an individual observation.

Non-Planar Regression Surface

Diagnostics for a non-linear regression surface are:

- Preliminary plots used to explore the data often give the first indication of non-linear relationships between the response and the regressors.
- Plots of residuals versus fitted values or residuals versus individual regressors.
- `gam` plots or partial residual plots.

Possible remedies are:

- Transform the response,
- Transform one or more of the regressors,
- Add polynomial terms for one or more of the regressors,
- Add interaction terms.

Correlated Data

Diagnostics for correlated data are:

- A careful assessment of how the data was collected.
- For data collected over time a plot of residuals versus time order or a correlogram.

Possible remedies are:

- If the amount of correlation is small – ignore.
- Adopt a more appropriate analysis to take into account the correlation – e.g. time series models, mixed models ...

Non-constant Error Variance

Diagnostics for non-constant error variance:

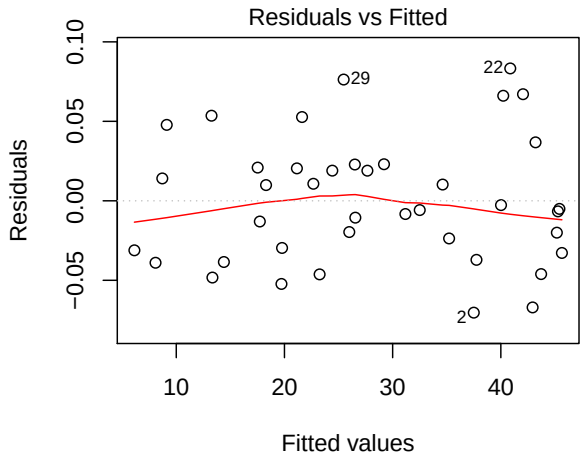
- Plot of residuals versus fitted values,
- A scale location plot.

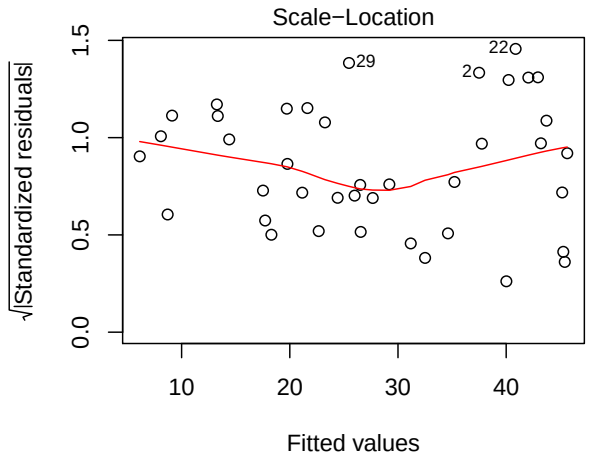
Possible remedies are:

- Transform the response,
- Use weighted least squares.

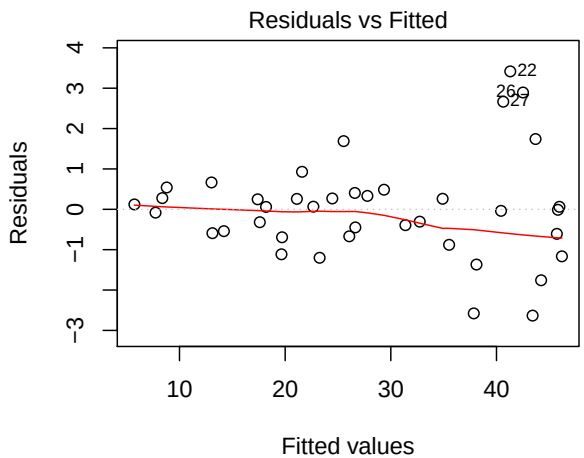
Detecting Non-constant Error Variance

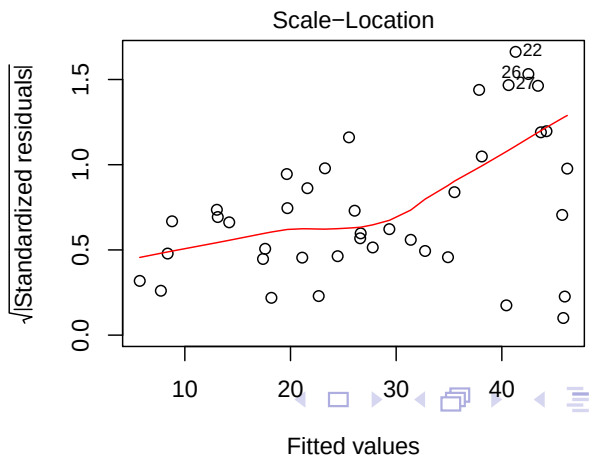
Constant error variance:





Non-constant error variance:





Errors are not Normal

Diagnostics for a non-Normal error distribution:

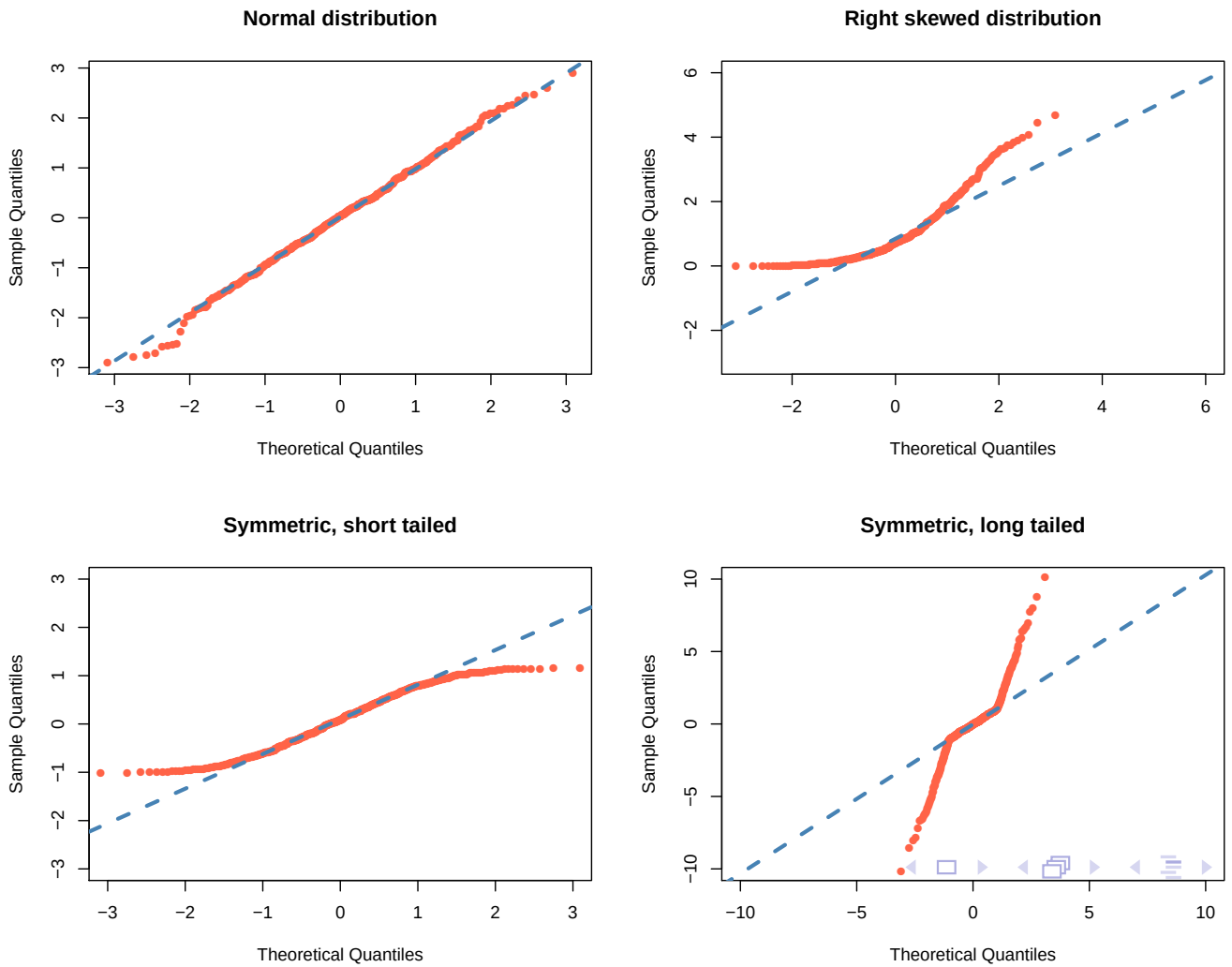
- qq-plot of residuals.

Possible remedies are:

- Ignore – don't worry, be happy. In most cases, the sampling distribution of the $\widehat{\beta}$'s will be adequately approximated by a Normal distribution (bless you central limit theorem),
- Use a glm with a more appropriate distributional assumption,
- Transform the response.

Detecting Non-Normality

```
qqnorm(residuals(xyz.lm))
```

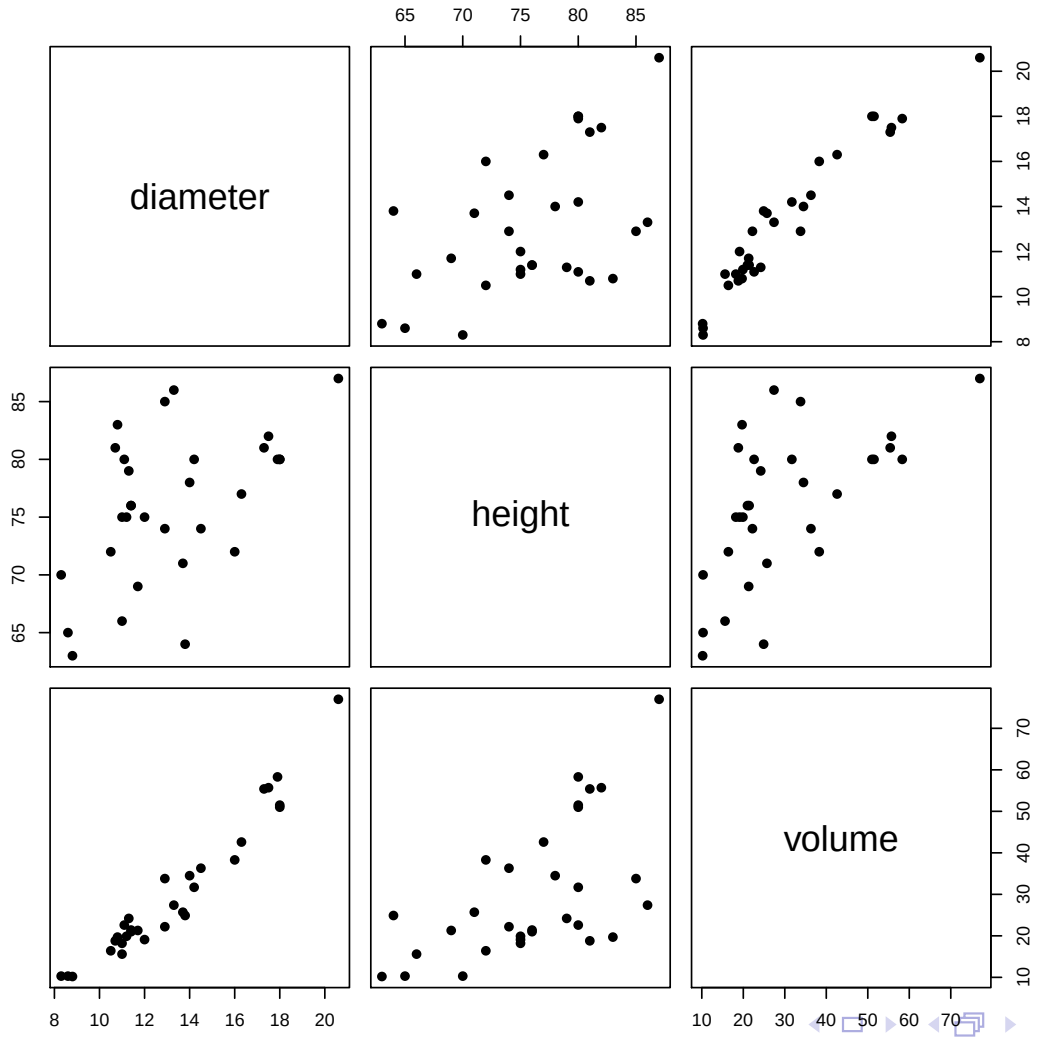


Cherry Tree Data

Measurements are given for 31 cherry trees. We want to predict timber volume (cubic feet) from height (feet) and diameter (inches) of the tree.

Diameter	Height	Volume
8.3	70	10.3
8.6	65	10.3
8.8	63	10.2
10.5	72	16.4
10.7	81	18.8
10.8	83	19.7
----------	----------	----------
18.0	80	51.0
20.6	87	77.0

Cherry Tree Pairs Plot

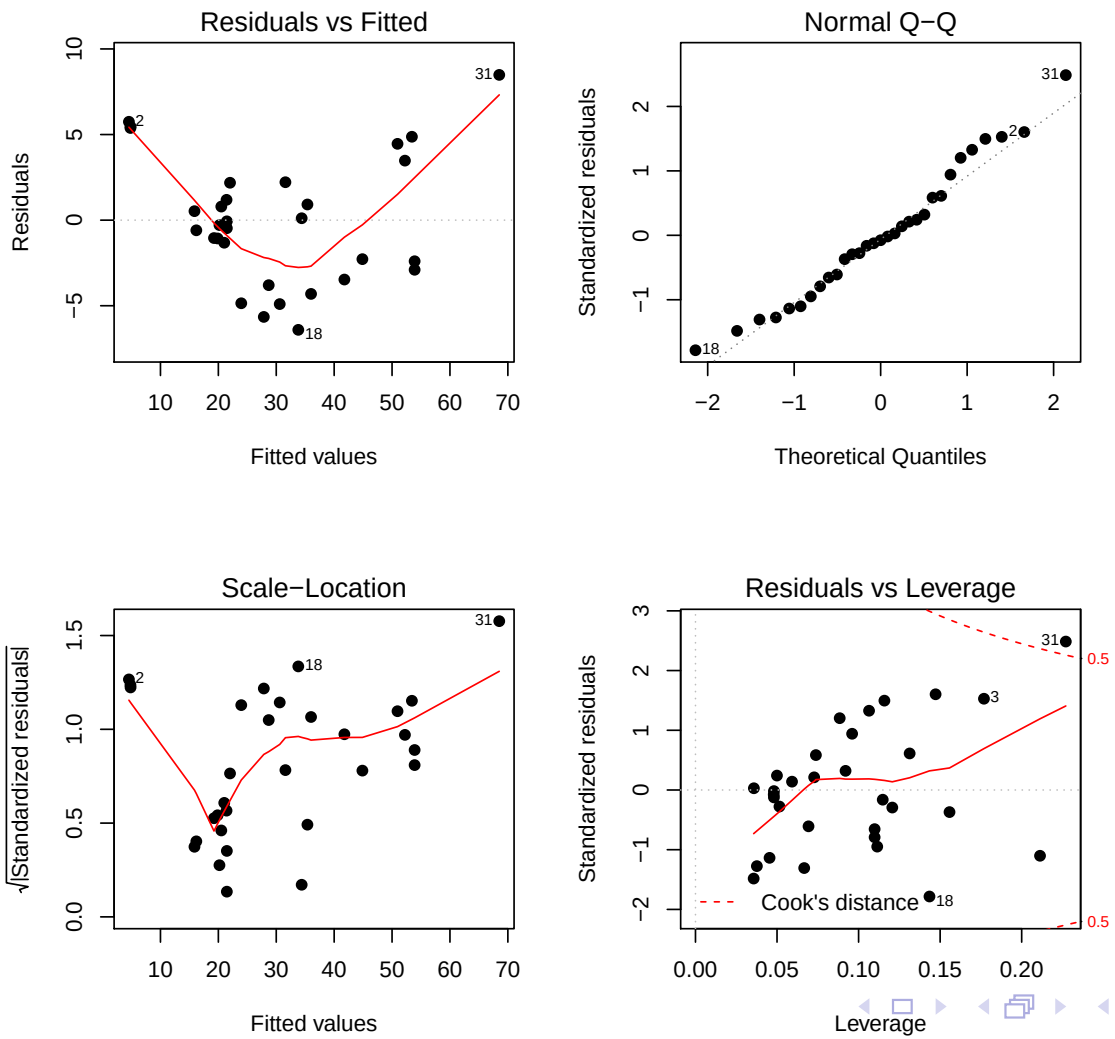


Cherry Tree Regression Model

Start by fitting the simple ordinary regression model:

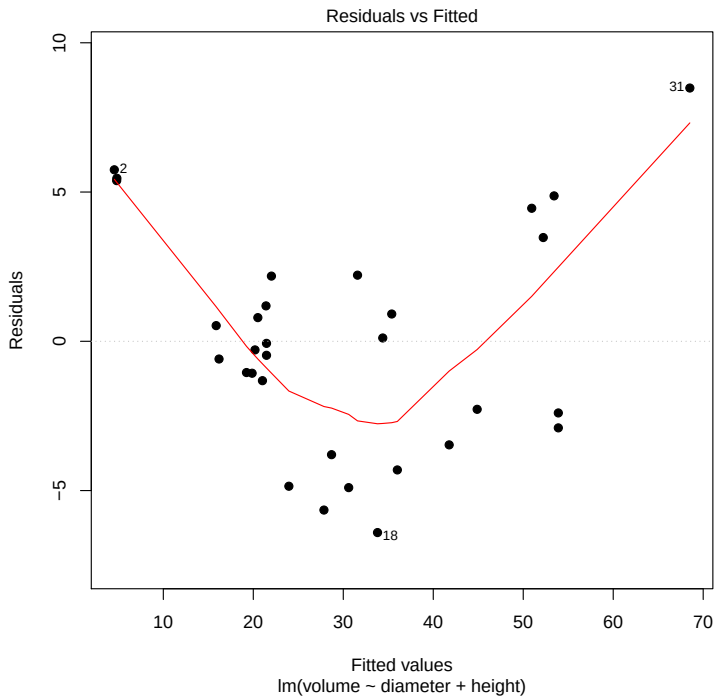
```
> cherry.lm <- lm(volume~diameter+height,data=cherry.df)
> summary(cherry.lm)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -57.9877     8.6382  -6.713 2.75e-07 ***
diameter      4.7082     0.2643  17.816  < 2e-16 ***
height        0.3393     0.1302   2.607   0.0145 *
---
Residual standard error: 3.882 on 28 degrees of freedom
Multiple R-squared:  0.948,Adjusted R-squared:  0.9442
F-statistic:   255 on 2 and 28 DF,  p-value: < 2.2e-16
```

Diagnostic Plots



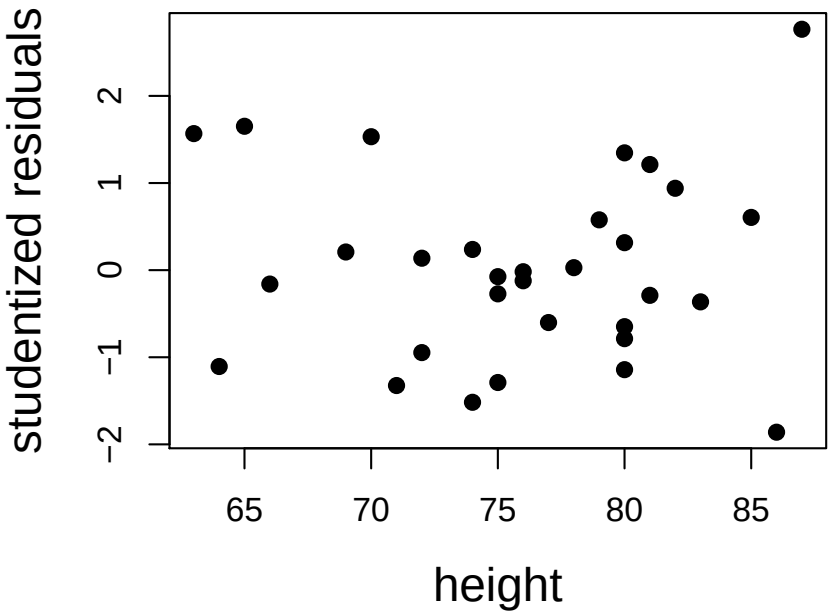
Residuals versus Fitted Values

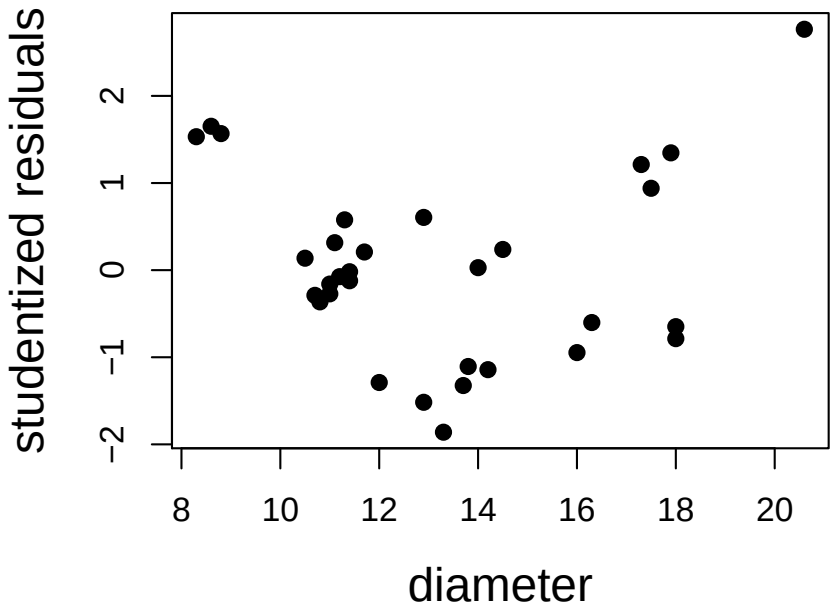
The plot of residuals versus fitted values suggests that the regression surface may be curved.



Residuals versus Regressors

Plots of the studentized residuals versus each regressor:





Additive Models

Additive models can be used to explore the nature of the curvature in the regression surface.

- These are models of the form

$$Y = g_1(x_1) + g_2(x_2) + \cdots + g_k(x_k) + \varepsilon$$

  where $g_1, \ldots g_k$ are transformations.

- Fitted using the `gam` function from the `mgcv` package.
- The transformations can be set to be "smoothers" that are estimated by the function.
- Plots of the smoother transformations suggest suitable transformations for the explanatory variables.

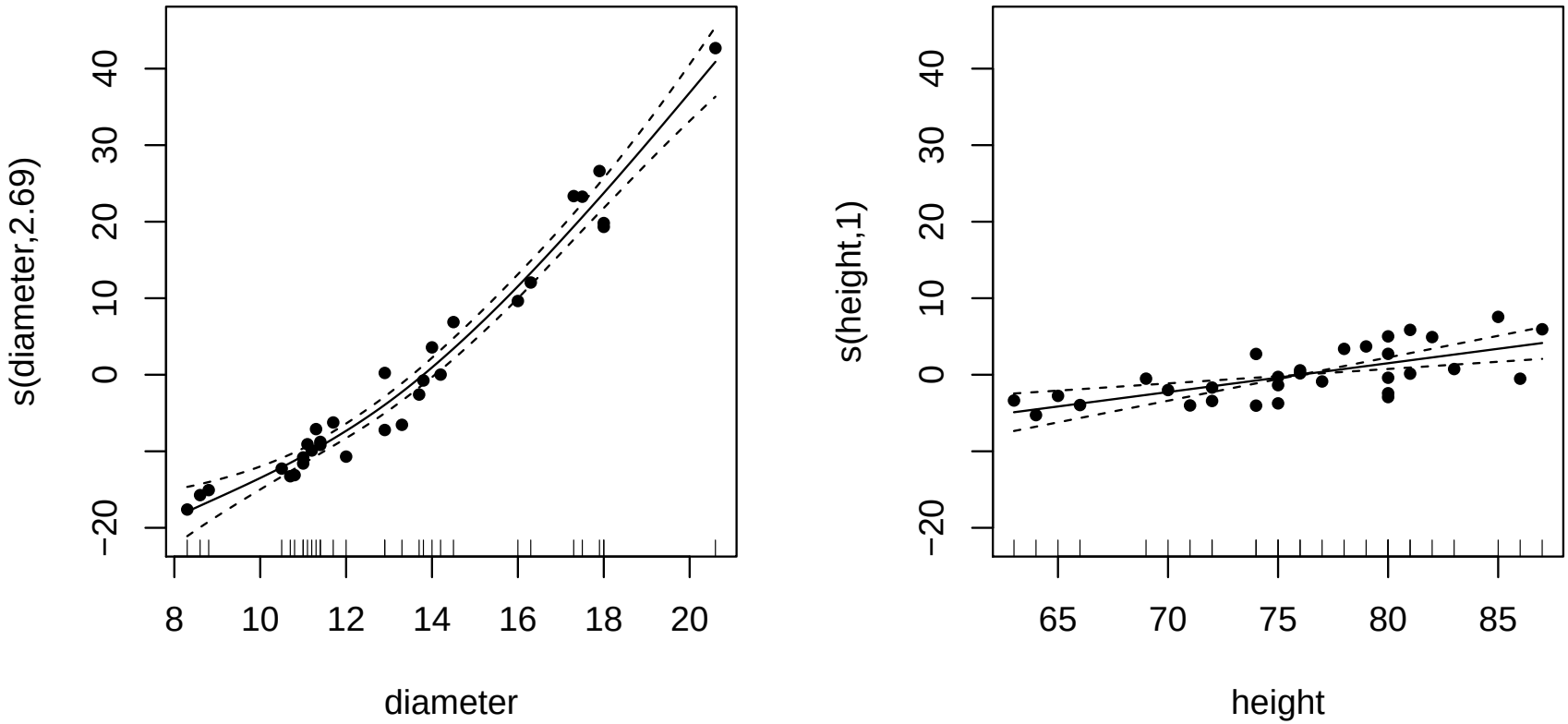
Generating a gam Plot

To produce the gam plot:

```
library(mgcv)
cherry.gam <- gam(volume~s(diameter)+s(height),
                                        data=cherry.df)

plot(cherry.gam,residuals=T,pages=1,pch=20)
```

The gam Plot



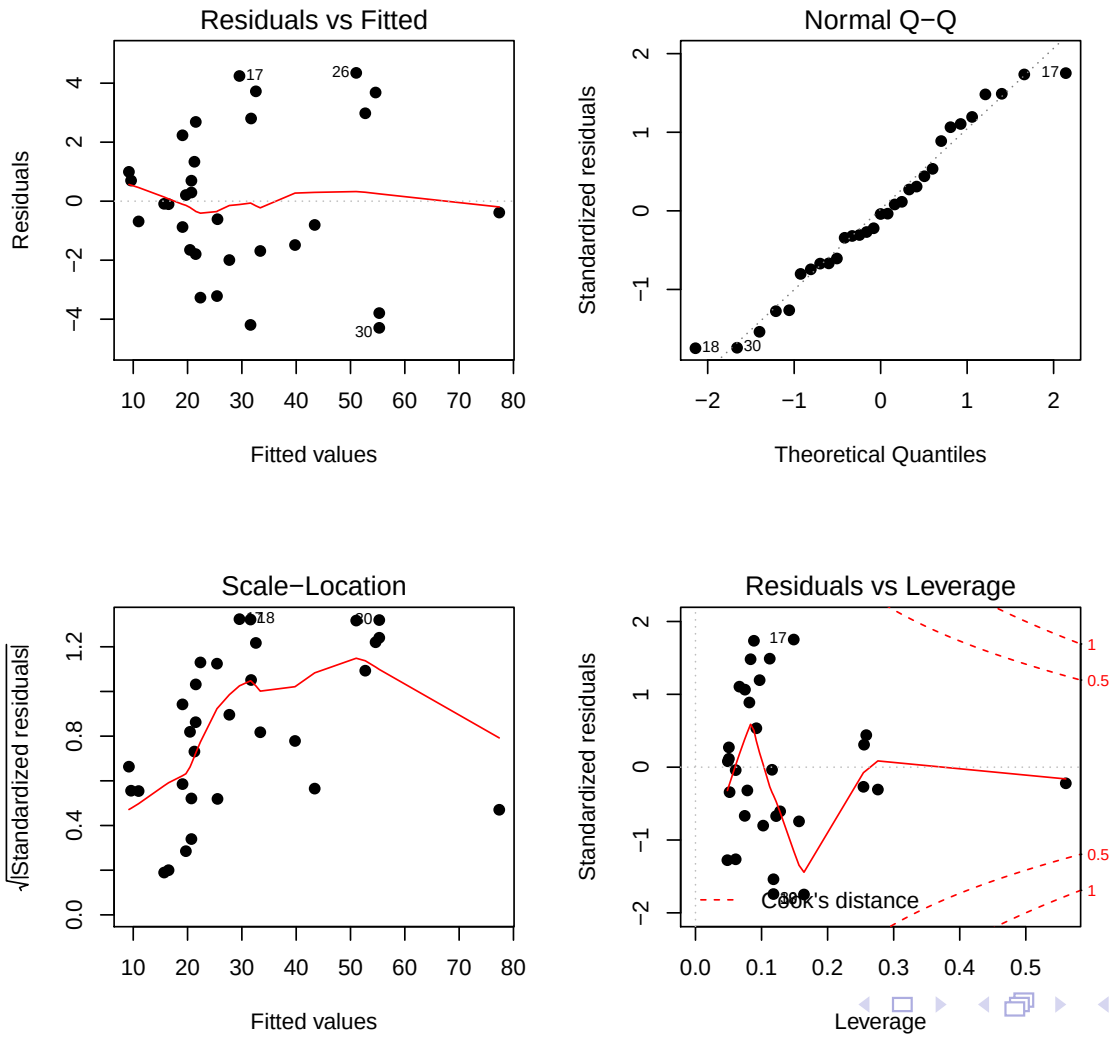
- Just a linear term is fine for `height`.
- Possibly add a quadratic term for `diameter`?

Quadratic Model

Try adding the quadratic term for diameter to our model:

```
cherryA.lm <- lm(volume~diameter+I(diameter^2)+height,
                                                data=cherry.df)
summary(cherryA.lm)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -9.92041   10.07911  -0.984 0.333729
diameter     -2.88508    1.30985  -2.203 0.036343 *
I(diameter^2)  0.26862    0.04590   5.852 3.13e-06 ***
height         0.37639    0.08823   4.266 0.000218 ***
---
Residual standard error: 2.625 on 27 degrees of freedom
Multiple R-squared:  0.9771,Adjusted R-squared:  0.9745
F-statistic: 383.2 on 3 and 27 DF,  p-value: < 2.2e-16
```

Diagnostic Plots for the Quadratic Model



Using Theory: Cherry Trees

A tree trunk is a bit like a cone. For a cone volume, height and diameter are related by

$$\text{volume} = \frac{\pi}{3} \times \frac{\text{diameter}^2}{4} \times \text{height}.$$

Taking the log of a multiplicative relationship produces a additive relationship:

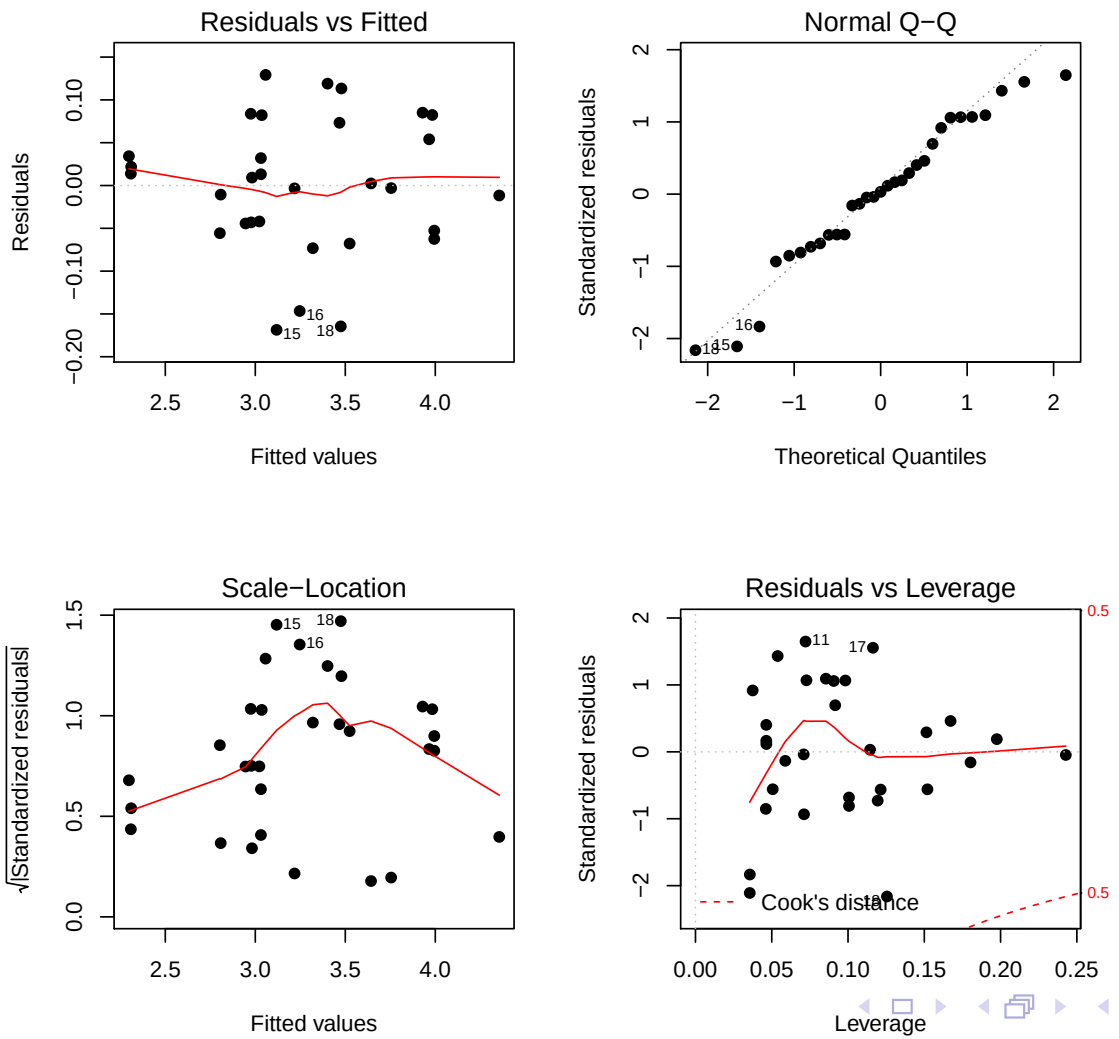
$$\log(\text{volume}) = \log(\pi/12) + 2 \times \log(\text{diameter}) + \log(\text{height}).$$

- Linear regression using the logged variables is worth trying.

Logged Model

```
> cherryB.lm <- lm(log(volume)~log(diameter)+log(height),
                                                data=cherry.df)
> summary(cherryB.lm)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -6.63162    0.79979  -8.292 5.06e-09 ***
log(diameter)  1.98265    0.07501  26.432  < 2e-16 ***
log(height)    1.11712    0.20444   5.464 7.81e-06 ***
---
Residual standard error: 0.08139 on 28 degrees of freedom
Multiple R-squared:  0.9777,Adjusted R-squared:  0.9761
F-statistic: 613.2 on 2 and 28 DF,  p-value: < 2.2e-16
```

Diagnostic Plots for the Logged Model



Which Model?

So which model is better?

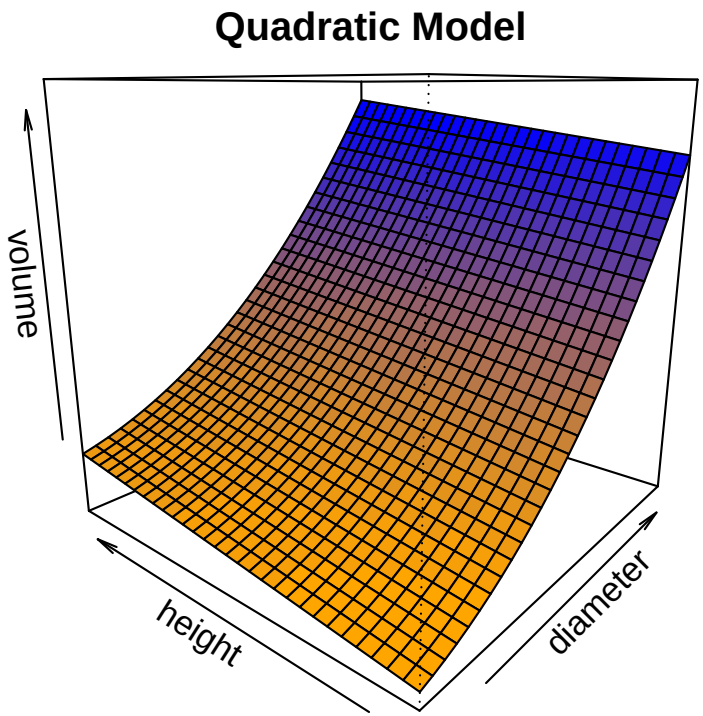
- Both model provide a good fit to the data – remember that all models are just approximations and there is no "true model".
- The quadratic model doesn't require transformations of either the response or the regressors.
- The logged model is (to some extent) supported by theory and may better satisfy the constant variance assumption.

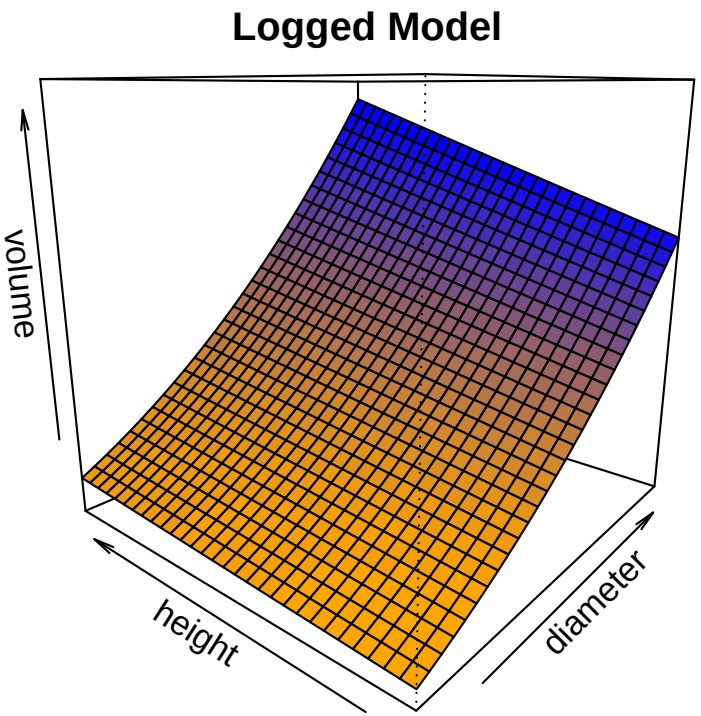
Comparing the Two Models

Let's compare the fitted regression surfaces for the two models.

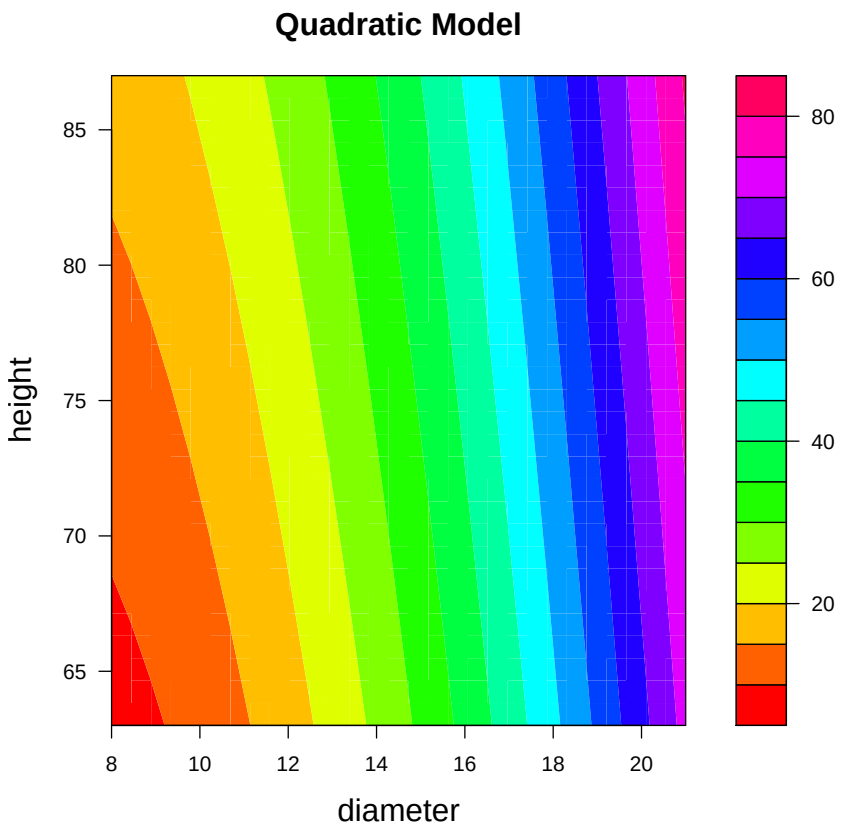
- As there are only two regressors, this gives us a chance to look at some different graphical options to represent 3-dimensional surfaces:
  - wireframe plots,
  - filled contour plots,
  - ordinary contour plots.

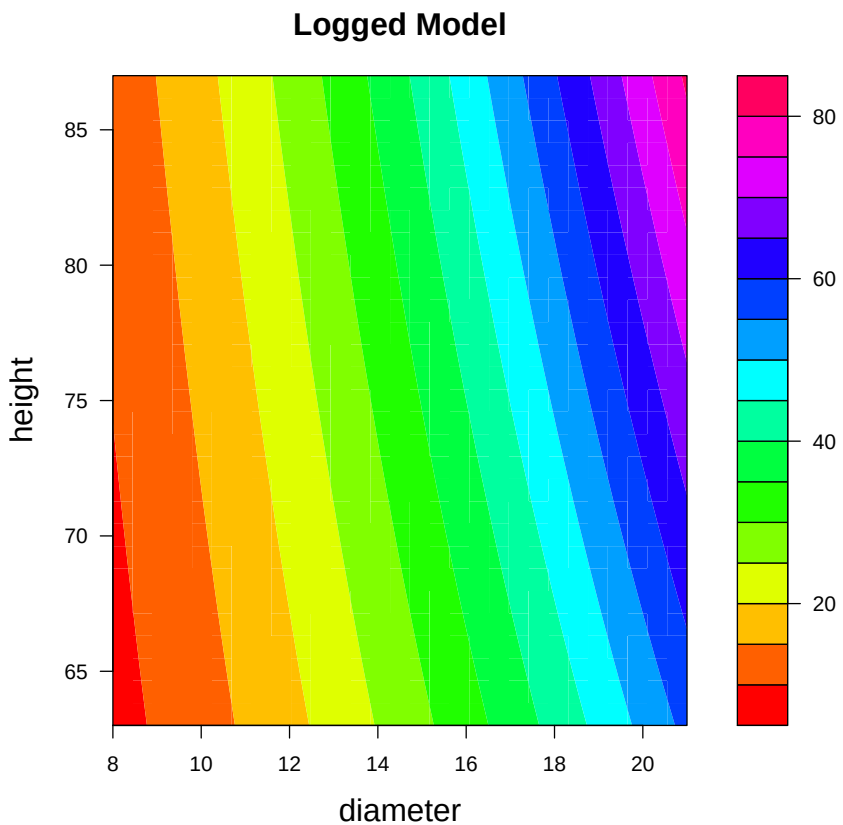
Wireframe Plots



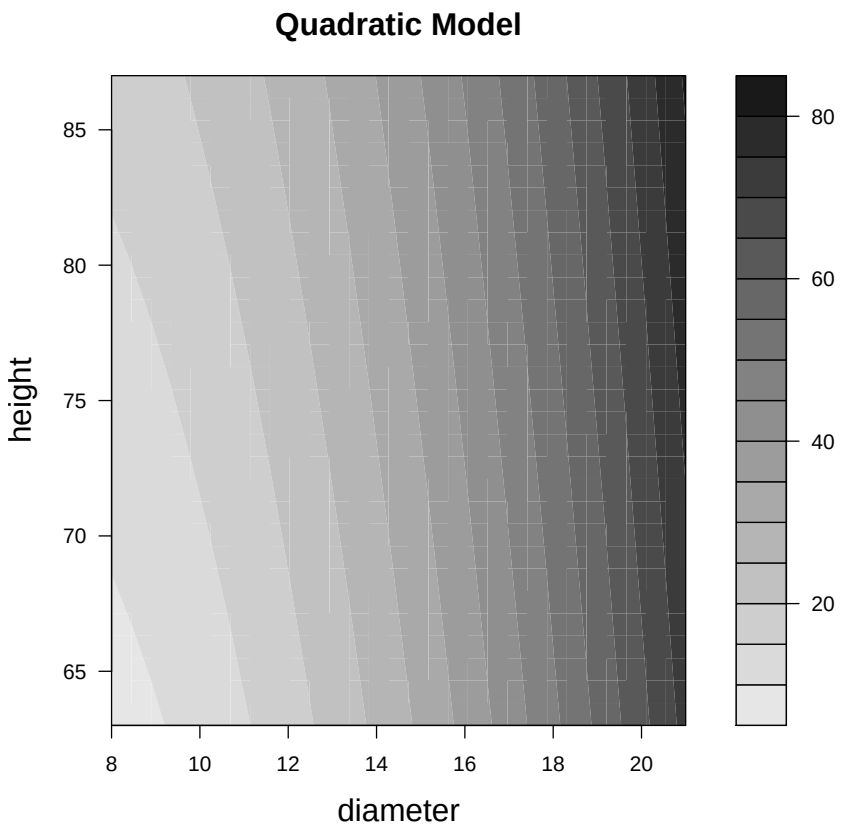


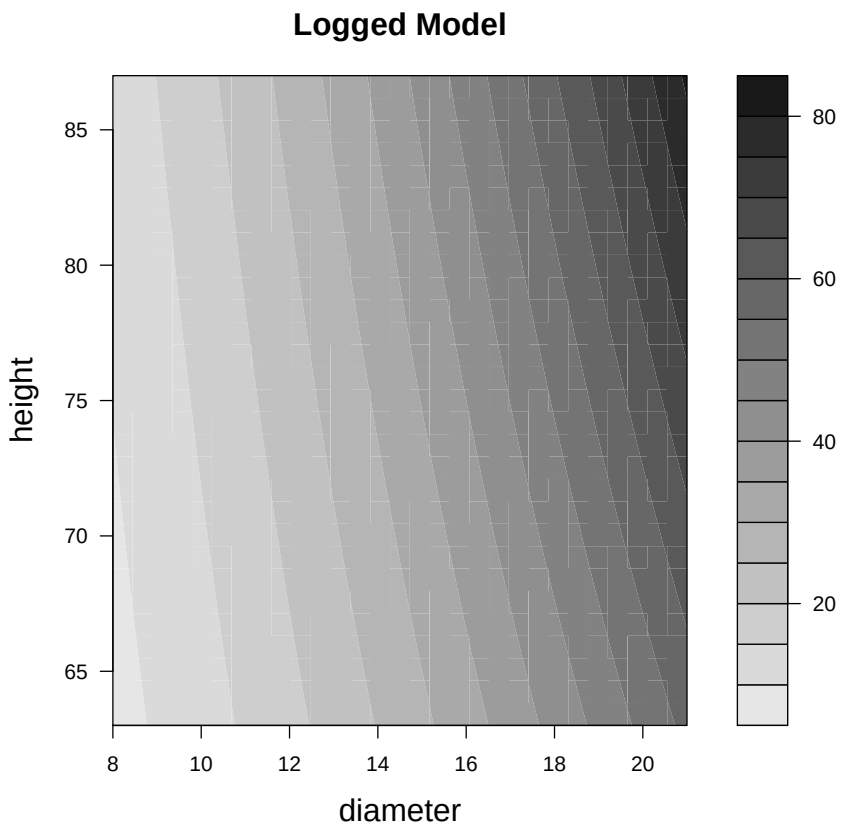
Filled Contour Plots



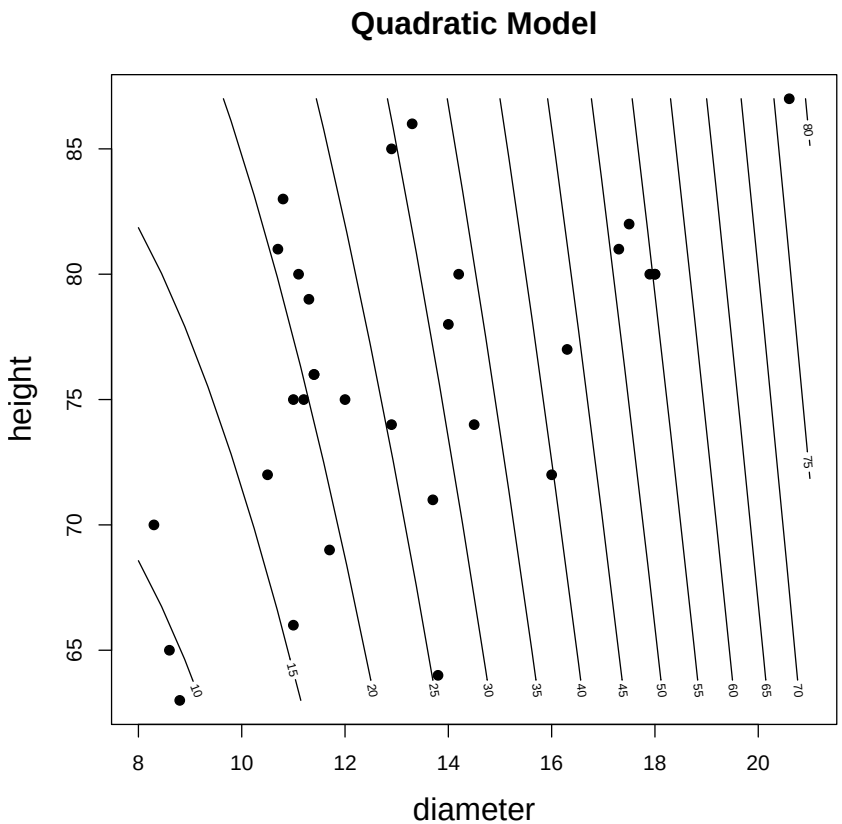


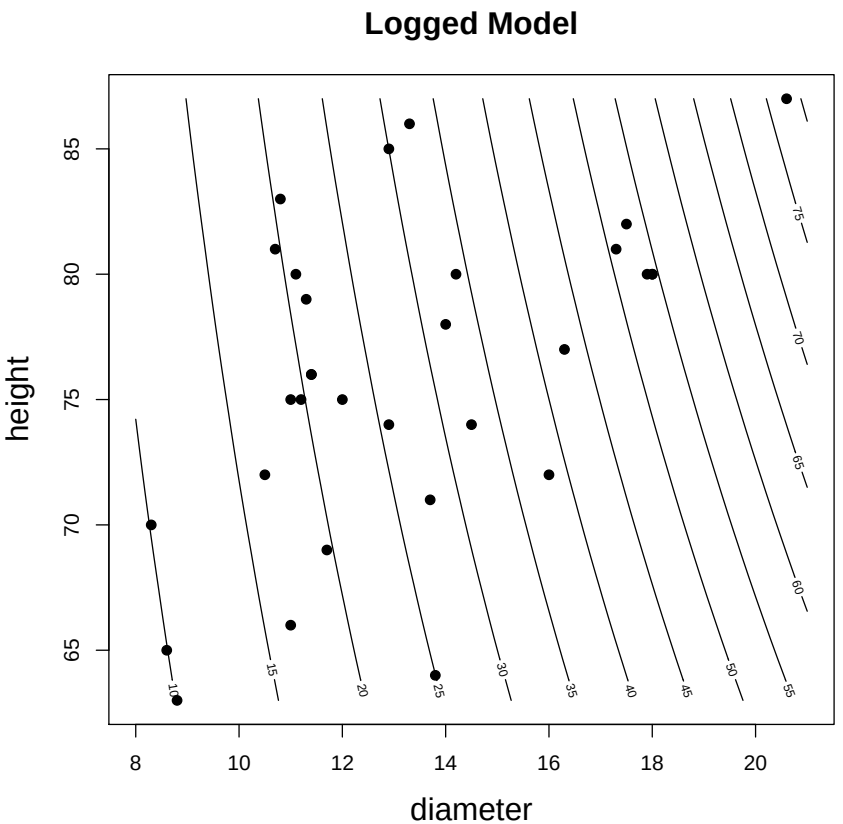
Filled Contour Plots in Grey



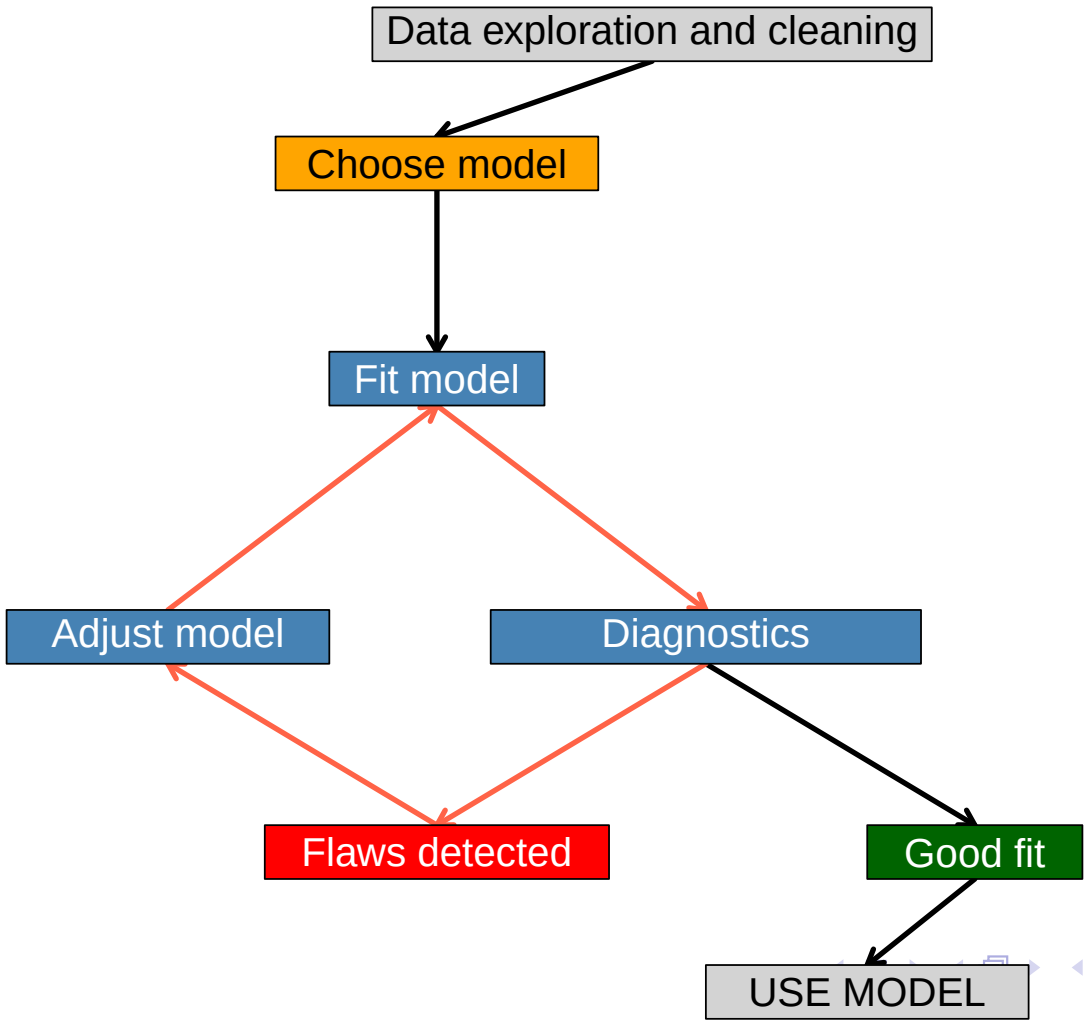


Ordinary Contour Plots





Model Building Cycle



Comments

There is no right way or even best way to implement the model building cycle.

- Usually a good idea to start with data exploration and cleaning.
- Consider the purpose of the model a that may require that certain terms be included.
- Then I concentrate on finding a model that satisfies the linearity assumption.
- Next look for outliers, high leverage and influential points.
- Finally, consider non-constant variance and Normality.

Model Building for GLM's

The model building cycle is the same for GLM's as for ordinary regression.

- Start with data exploration and data cleaning.
- First concentrate on getting the relationship between the linear combination of predictors and the mean response sorted out.
- Plots of studentized residuals should be random scatter between -3 and $+3$.
- Under certain circumstances, a lack-of-fit test can be performed using the residual deviance (more on this later).

Model Assumptions for GLM's

For GLM's we have the following set of assumptions:

1. The link function is a reasonable approximation of the relationship between the mean of the response and the linear combination of predictors.
2. The observations are independent.
3. The specified response distribution is suitable for the data.

The Link Function

For a GLM, we don't expect the response to be a linear function of the regressors. Instead, the link function applied to the response should be a linear function of the regressors.

For logistic regression:

$$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k.$$

For Poisson regression:

$$\log \mu = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k.$$

Link Function Diagnostics

Diagnostics for a poor fitting link function.

- Preliminary plots used to explore the data where the response has been transformed using the inverse link function.
- Plots of residuals versus fitted values or residuals versus individual regressors.
- `gam` plots.

Possible remedies are:

- Transform one or more of the regressors,
- Add polynomial terms for one or more of the regressors,
- Add interaction terms.
- Use a different link function.

Correlated Data

Diagnostics for correlated data are:

- A careful assessment of how the data was collected.
- For data collected over time a plot of (studentized) residuals versus time order or a correlogram.

Possible remedies are:

- If the amount of correlation is small – ignore.
- Adopt a more appropriate analysis to take into account the correlation – e.g. time series models, mixed models ...

Response Distribution

The nature of the response often indicates what distribution should be used – the binomial is often used for proportions and the Poisson is often used for counts. Diagnostics:

- Try fitting models with alternative distributions (the quasi-binomial for proportions and the quasi-Poisson distribution or negative binomial distribution for counts).
- Plots of studentized residuals may have unusually large values.

Remedies:

- Use a more suitable response distribution.

Tobacco Budworm Example

Tobacco budworms cause much damage to cotton crops

- Due to intensive cropping and the use (misuse) of pesticides the insect has become resistant to pyrethroids (pesticide)
- We will look at data from an experiment that examined the level of resistance in adult moths

The Experiment

Batches of 20 pyrethroid resistant moths (either all male or all female) were exposed to a range of doses of cypermethrin two days after emergence from pupation.

- Number of dead moths (does not respond when poked with a blunt instrument) were recorded 72 hours after treatment.
- Goals are to assess the effect of increasing dose of cypermethrin on toxicity and determine if there is a difference in resistance between female and male moths.
- Data is grouped: for each of 6 doses (1.0, 2.0, 4.0, 8.0, 16.0, 32.0 mg) and each of female (sex= 1) and male (sex= 0), the number of deaths (s) out of $n = 20$ moths was recorded.

The Data

```
> budworm.df
   sex dose  s  n
1    0    1  1 20
2    0    2  4 20
3    0    4  9 20
4    0    8 13 20
5    0   16 18 20
6    0   32 20 20
7    1    1  0 20
8    1    2  2 20
9    1    4  6 20
10   1    8 10 20
11   1   16 12 20
12   1   32 16 20
```

Some Background

It is known from experience that fitting a model using log dose rather than dose often works works better for toxicity data. Some possible strategies are:

- Fit both models and compare results of summary.
- Plot the logits for the observed proportions to see which fits better.

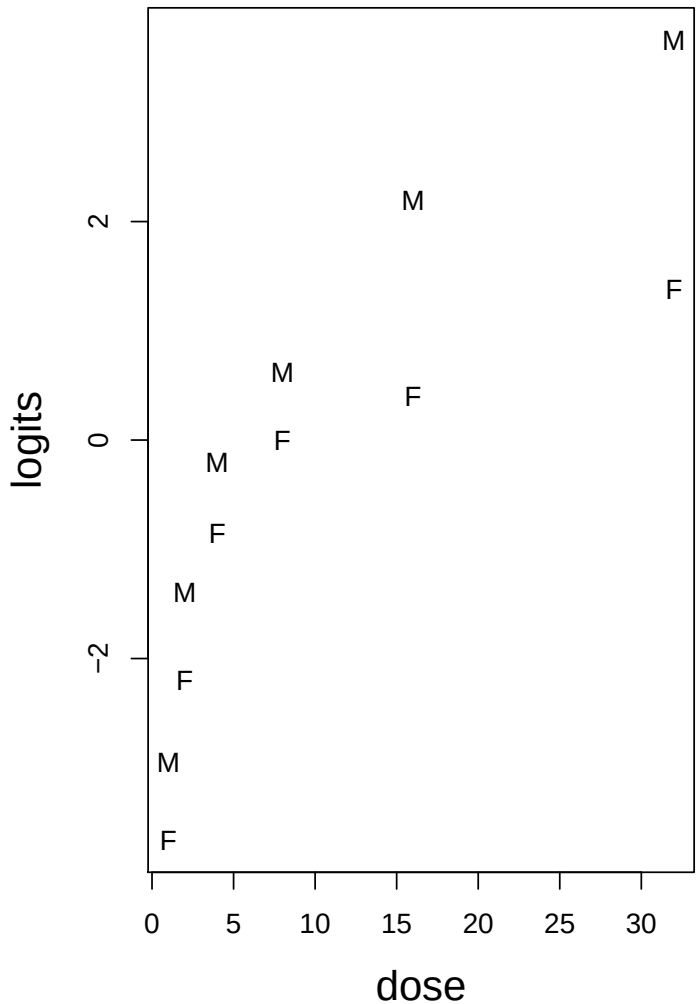
The Dose Model

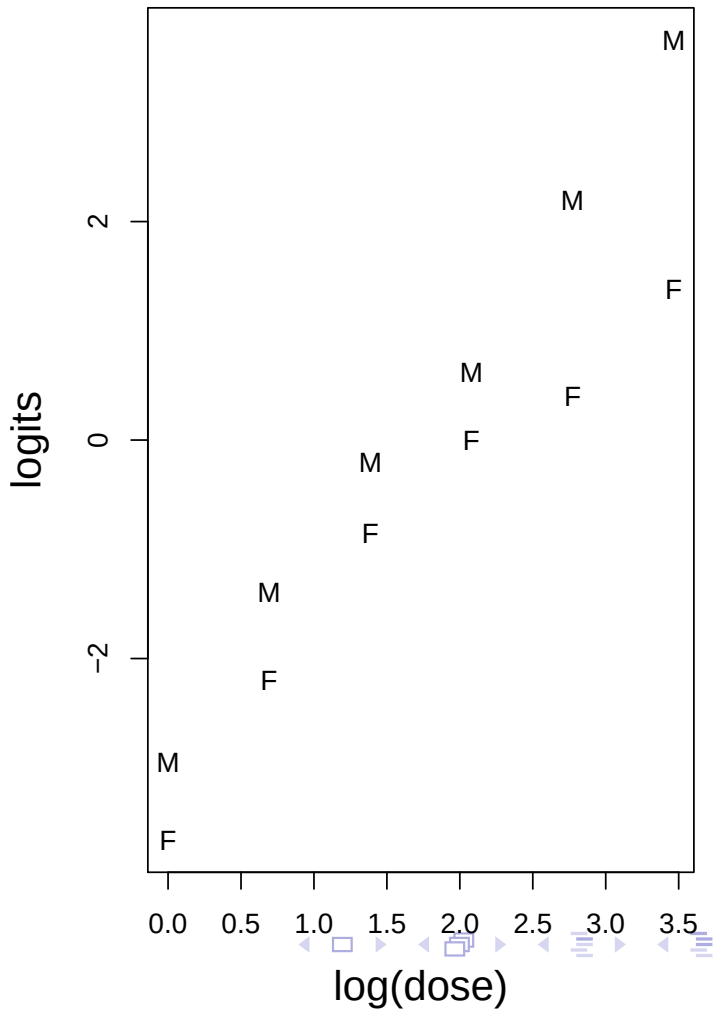
```
> dose.glm<-glm(cbind(s, n-s) ~ sex + dose,
                family=binomial, data = budworm.df)
> summary(dose.glm)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.16607    0.26155  -4.458 8.26e-06 ***
sex         -0.96855    0.32954  -2.939  0.00329 **
dose         0.15996    0.02341   6.832 8.39e-12 ***
---
    Null deviance: 124.876  on 11  degrees of freedom
Residual deviance:  27.968  on  9  degrees of freedom
AIC: 64.078
```

The Log Dose Model

```
> logdose.glm<-glm( cbind(s, n-s) ~ sex + log(dose),
                    family=binomial, data = budworm.df)
> summary(logdose.glm)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3724     0.3855  -6.154 7.56e-10 ***
sex          -1.1007     0.3558  -3.093  0.00198 **
log(dose)     1.5353     0.1891   8.119 4.70e-16 ***
---
    Null deviance: 124.8756  on 11  degrees of freedom
Residual deviance:   6.7571  on  9  degrees of freedom
AIC: 42.867
```

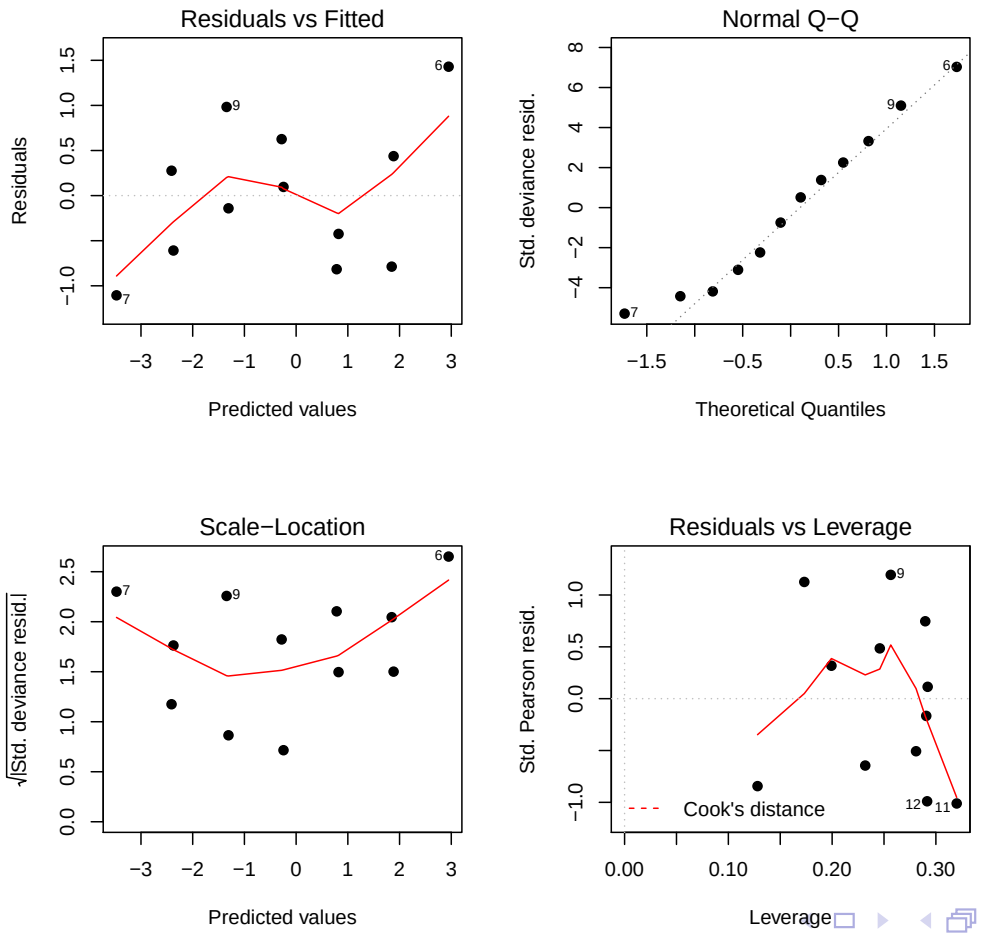
Logits Plots





Summary

The log dose model fits a lot better. Some diagnostic plots:



Interaction?

It makes sense to try the model that includes an interaction between log dose and sex.

- Test to see whether the effect of increasing dose is the same on male moths as on female moths.
- For the logits versus log dose plot on slide 47, the no interaction model has parallel lines for female and male moths – for the model that includes an interaction these lines can have different slopes.

The Log Dose with Interaction Model

```
> logdoseI.glm<-glm( cbind(s, n-s) ~ sex * log(dose),
                     family=binomial, data = budworm.df)
> summary(logdoseI.glm)
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -2.8186     0.5480  -5.143 2.70e-07 ***
sex            -0.1750     0.7783  -0.225    0.822
log(dose)       1.8163     0.3059   5.937 2.91e-09 ***
sex:log(dose)  -0.5091     0.3895  -1.307    0.191
---
    Null deviance: 124.8756  on 11  degrees of freedom
Residual deviance:   4.9937  on  8  degrees of freedom
AIC: 43.104
```

The Log Dose with Interaction Model

```
> anova(logdoseI.glm,test="Chisq")

              Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                             11    124.876
sex            1    6.077        10    118.799   0.0137 *
log(dose)      1  112.042         9      6.757   <2e-16 ***
sex:log(dose)  1    1.763         8      4.994   0.1842
```

- No evidence that the interaction is needed in the model.

Interpreting the Model

For the log dose model:

```
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3724     0.3855  -6.154 7.56e-10 ***
sex          -1.1007     0.3558  -3.093  0.00198 **
log(dose)     1.5353     0.1891   8.119 4.70e-16 ***
```

From the signs of the estimated coefficients we conclude that the probability of mortality is less for female moths than for male moths and that the probability of mortality increases as dose increases.

Interpreting the Model (cont.)

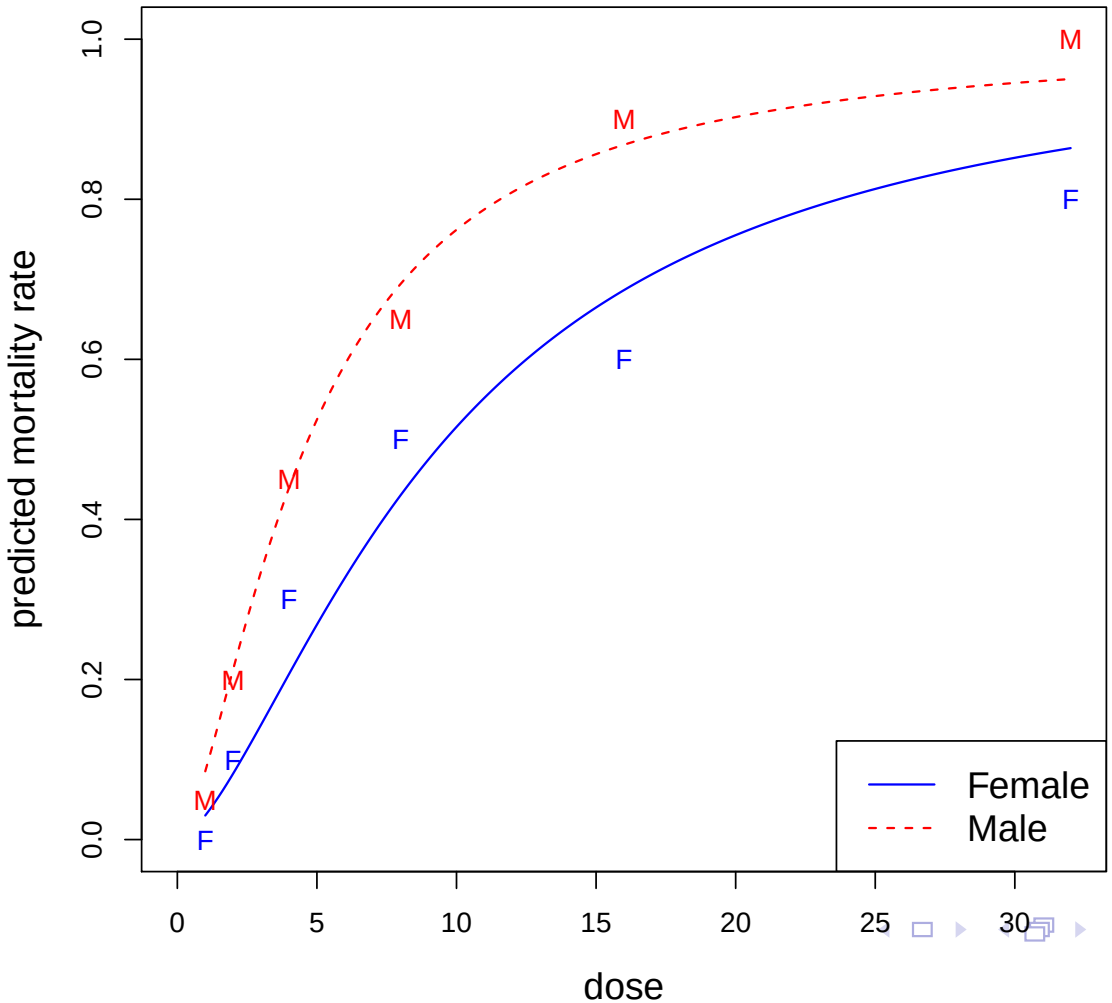
If we wish to be more precise we can talk about the impact on the log odds of mortality.

$$\log \frac{\hat{\pi}}{1-\hat{\pi}} = -2.37 - 1.10 \text{ sex} + 1.53 \log(\text{dose}).$$

- Estimated log odds is 1.10 less for females than for males.
- Estimated log odds increases by 1.53 for each unit increase in log(dose).
- Probably only meaningful to people really familiar with working with logits.

Interpreting the Model (cont.)

In this case a graph is extremely useful.



R Commands to Create Graph

```
> ds<-seq(1,32,length=200)
> newF.df<-data.frame(sex=1,dose=ds)
> newM.df<-data.frame(sex=0,dose=ds)
> estsF<- predict(logdose.glm,newF.df,type="response")
> estsM<- predict(logdose.glm,newM.df,type="response")

> plot(c(0,32),c(0,1),xlab="dose",ylab="predicted
                    mortality rate",type="n",cex.lab=1.3)
> lines(ds,estsF,lty=1,col="blue",lwd=1.3)
> lines(ds,estsM,lty=2,col="red",lwd=1.3)
> points(budworm.df$dose,budworm.df$s/20, pch=c(rep("M",6),
              rep("F",6)), col=c(rep("red",6),rep("blue",6)))
> legend("bottomright",legend=c("Female","Male"),lty=1:2,
                         col=c("blue","red"),lwd=1.3,cex=1.3)
```

A Lack of Fit Test for GLM's

For GLM's the residual deviance measures how well the model fits the data relative to the maximal model. For many GLM's (including Poisson regression and logistic regression for grouped data) the deviance will have approximately a χ^2 and can be used to test for lack of fit in the model.

- Tests for general lack of fit. Evidence of lack of fit may be due to the response distribution being mis-specified, one or more important regressors being omitted, an unsuitable link function or problems with the data (outliers and influential points).
- Reference distribution is χ^2_{n-k-1} for Poisson regression and χ^2_{m-k-1} for grouped binomial regression where m is the number of covariate patterns.

A Lack of Fit Test for GLM's (cont)

- A small p-value indicates lack of fit.
- Does not work well for ungrouped binomial data – in this case the χ^2 distribution is not a good approximation of the sampling distribution of the residual deviance. For grouped binomial data the bigger the n_i's (number of observations for covariate pattern i) the better.
- Also doesn't work for GLM's which have an estimated scale parameter (e.g. quasibinomial and quasipoisson), since in these cases the residual deviance is used to estimate the scale parameter.
- Also called a goodness of fit test.

Lack of Fit Tests for Budworm Models

Let's calculate the lack of fit p-values for the 3 models we tried for the budworm data.

The dose model:

```
> 1-pchisq(27.968,9)
[1] 0.0009656815
```

The log dose model:

```
> 1-pchisq(6.7571,9)
[1] 0.662392
```

The log dose model with interaction:

```
> 1-pchisq(4.9937,8)
[1] 0.7582493
```

West Virginia Mining Accident Data

This example features the number of accidents per mine in a three month period in 44 coal mines in West Virginia. The variables are:

COUNT: the number of accidents (response),

INB: inner burden thickness,

EXTRP: percentage of coal extracted from mine,

AHS: the average height of the coal seam in the mine,

AGE: the age of the mine.

Mining Accident Data

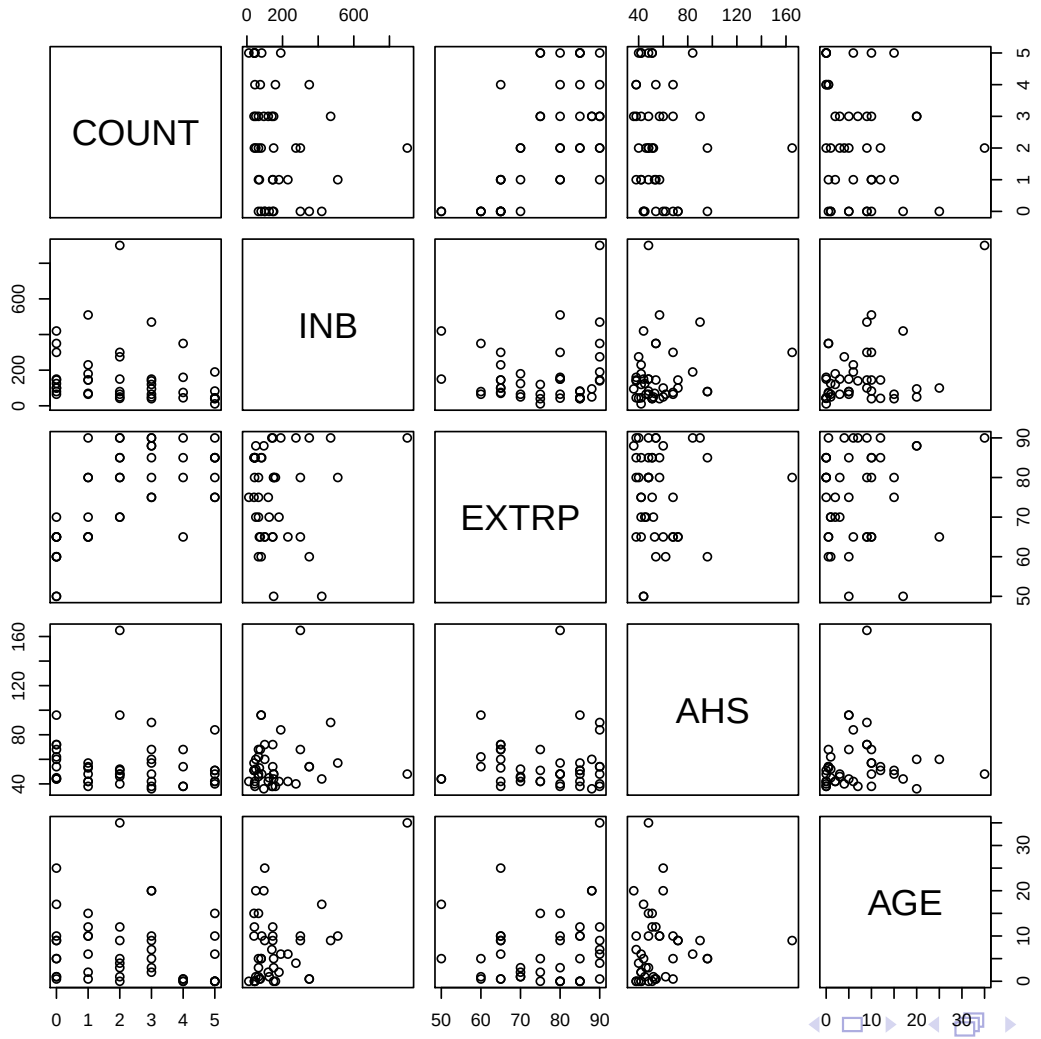
COUNT	INB	EXTRP	AHS	AGE
2	50	70	52	1.0
1	230	65	42	6.0
0	125	70	45	1.0
4	75	65	68	0.5
1	70	65	53	0.5
2	65	70	46	3.0
0	65	60	62	1.0
0	350	60	54	0.5
4	350	90	54	0.5
4	160	80	38	0.0

. . . 44 lines in all

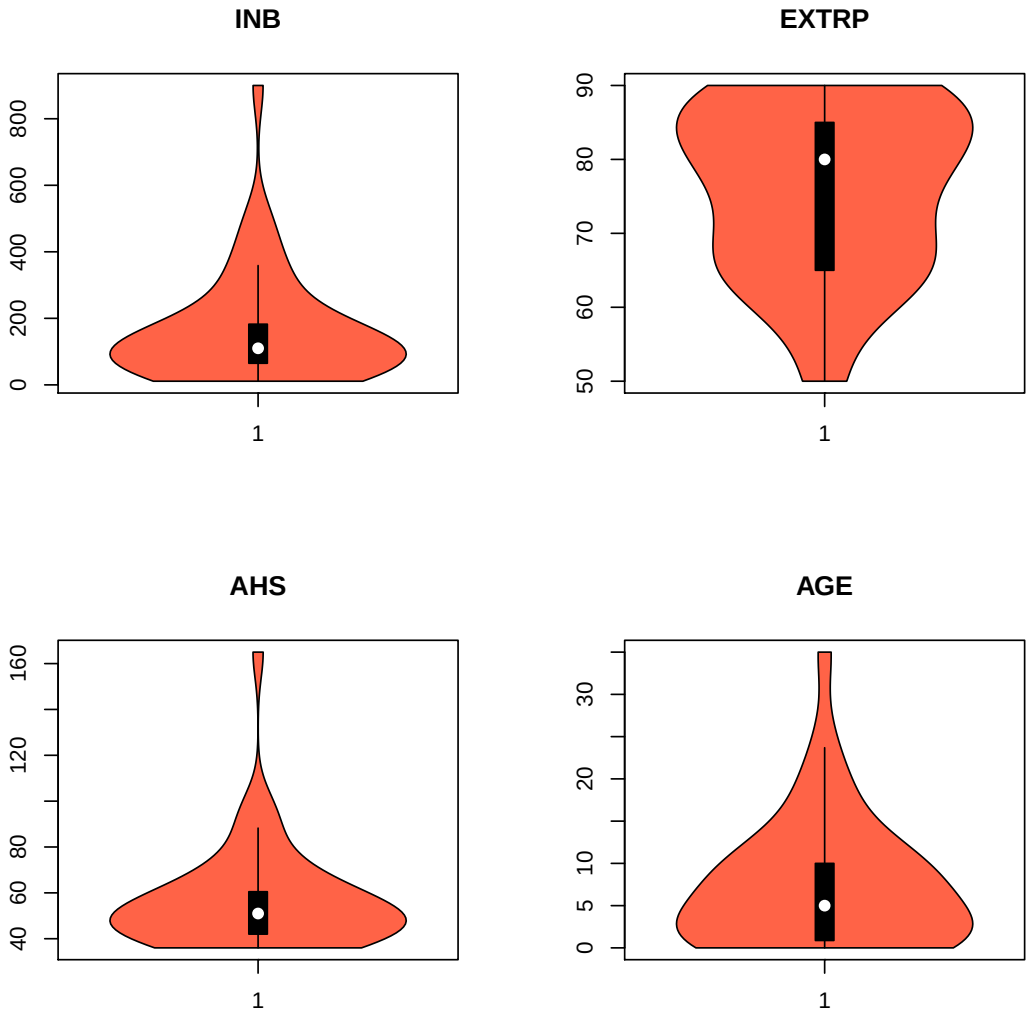
Mining Accident Data

```
> summary(mines.df)
     COUNT             INB              EXTRP
 Min.   :0.000   Min.   : 11.0   Min.   :50.00
 1st Qu.:1.000   1st Qu.: 65.0   1st Qu.:65.00
 Median :2.000   Median :110.0   Median :80.00
 Mean   :2.227   Mean   :164.9   Mean   :75.93
 3rd Qu.:3.250   3rd Qu.:182.5   3rd Qu.:85.00
 Max.   :5.000   Max.   :900.0   Max.   :90.00
      AHS              AGE
 Min.   : 36.00   Min.   : 0.000
 1st Qu.: 42.00   1st Qu.: 0.875
 Median : 51.00   Median : 5.000
 Mean   : 56.64   Mean   : 7.159
 3rd Qu.: 60.50   3rd Qu.:10.000
 Max.   :165.00   Max.   :35.000
```

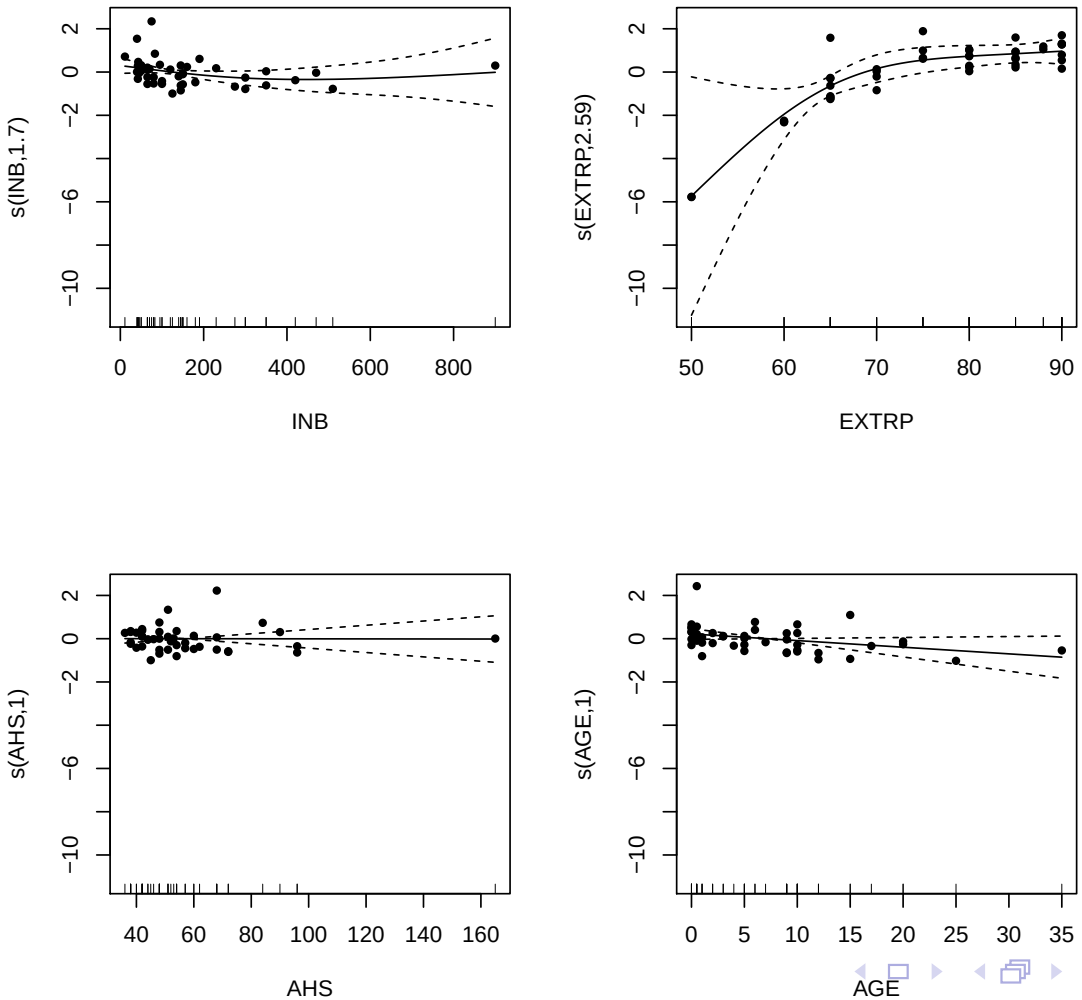
Pairs Plot of Data



Violin Plots of Regressors



gam Plots


s(INB,1.7)
2
0
−2
−6
−10
0
200
400
600
800
INB
s(EXTRP,2.59)
50
60
70
80
90
EXTRP
s(AHS,1)
40
60
80
100
120
140
160
AHS
s(AGE,1)
0
5
10
15
20
25
30
35
AGE

gam Plots (cont.)

The gam plot suggests:

- INB and EXTRP both appear to have non-linear effects on COUNTS,
- AHS and AGE both appear to have linear effects on COUNTS,
- INB and AHS appear to have small effects and may not be needed in the model.

Models to Try

Some models that we can try are:

- Linear in all regressors.
- Quadratic in `INB` and `EXTRP` and linear in `AHS` and `AGE`.
- Full second order model in `INB` and `EXTRP` and linear in `AHS` and `AGE`.
  - Previous model plus `INB:EXTRP` interaction.

The All Linear Effects Model

```
> mines1.glm<-glm(COUNT~.,family=poisson,data=mines.df)
> summary(mines.glm)
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.6097078  1.0284740  -3.510 0.000448 ***
INB         -0.0014441  0.0008415  -1.716 0.086145 .
EXTRP        0.0622011  0.0122872   5.062 4.14e-07 ***
AHS         -0.0017578  0.0050737  -0.346 0.729003
AGE         -0.0296244  0.0163143  -1.816 0.069394 .
---
    Null deviance: 74.984  on 43  degrees of freedom
Residual deviance: 37.717  on 39  degrees of freedom
AIC: 143.99

> 1-pchisq(37.717,39)
[1] 0.5283455
```

Model 2

```
> mines2.glm<-glm(COUNT~ INB + I(INB^2)+ EXTRP +I(EXTRP^2)
                  + AHS +AGE,family=poisson,data=mines.df)
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.281e+01  1.050e+01  -2.172   0.0299 *
INB         -3.741e-03  2.009e-03  -1.862   0.0626 .
I(INB^2)     4.396e-06  2.615e-06   1.681   0.0928 .
EXTRP        5.672e-01  2.711e-01   2.092   0.0364 *
I(EXTRP^2)  -3.265e-03  1.746e-03  -1.871   0.0614 .
AHS          4.241e-04  5.023e-03   0.084   0.9327
AGE         -3.618e-02  1.846e-02  -1.960   0.0500 *
    Null deviance: 74.984  on 43  degrees of freedom
Residual deviance: 28.728  on 37  degrees of freedom
AIC: 139

> 1-pchisq(28.728,37)
[1] 0.8327824
```

Model 3

```
mines3.glm<-glm(COUNT~ INB*EXTRP + I(INB^2)+ I(EXTRP^2)
                + AHS +AGE,family=poisson,data=mines.df)
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.654e+01  1.149e+01  -2.310   0.0209 *
INB         -3.046e-02  1.521e-02  -2.002   0.0453 *
EXTRP        7.091e-01  3.017e-01   2.351   0.0187 *
I(INB^2)     5.960e-07  3.285e-06   0.181   0.8560
I(EXTRP^2)  -4.453e-03  1.984e-03  -2.244   0.0248 *
AHS          2.108e-04  5.226e-03   0.040   0.9678
AGE         -2.747e-02  1.889e-02  -1.454   0.1460
INB:EXTRP    3.381e-04  1.878e-04   1.800   0.0718 .
    Null deviance: 74.984  on 43  degrees of freedom
Residual deviance: 24.952  on 36  degrees of freedom
AIC: 137.22

> 1-pchisq(24.952,36)
[1] 0.9169485
```

Model 3A

```
mines3A.glm<-glm(COUNT~ INB*EXTRP + I(INB^2)+ I(EXTRP^2)
                      + AGE,family=poisson,data=mines.df)
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.654e+01  1.150e+01  -2.309   0.0210 *
INB         -3.042e-02  1.517e-02  -2.005   0.0450 *
EXTRP        7.096e-01  3.016e-01   2.352   0.0186 *
I(INB^2)     5.600e-07  3.161e-06   0.177   0.8594
I(EXTRP^2)  -4.456e-03  1.984e-03  -2.246   0.0247 *
AGE         -2.731e-02  1.850e-02  -1.476   0.1398
INB:EXTRP    3.380e-04  1.877e-04   1.801   0.0717 .
    Null deviance: 74.984  on 43  degrees of freedom
Residual deviance: 24.953  on 37  degrees of freedom
AIC: 135.22
```

Model 3B

```
mines3B.glm<-glm(COUNT~ INB*EXTRP +  I(EXTRP^2) +
                    AGE,family=poisson,data=mines.df)
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.723e+01  1.086e+01  -2.507  0.01219 *
INB         -3.177e-02  1.321e-02  -2.405  0.01618 *
EXTRP        7.296e-01  2.806e-01   2.600  0.00931 **
I(EXTRP^2)  -4.602e-03  1.810e-03  -2.543  0.01100 *
AGE         -2.564e-02  1.584e-02  -1.618  0.10562
INB:EXTRP    3.580e-04  1.509e-04   2.372  0.01770 *

    Null deviance: 74.984  on 43  degrees of freedom
Residual deviance: 24.984  on 38  degrees of freedom
AIC: 133.26
```

Remove AGE?

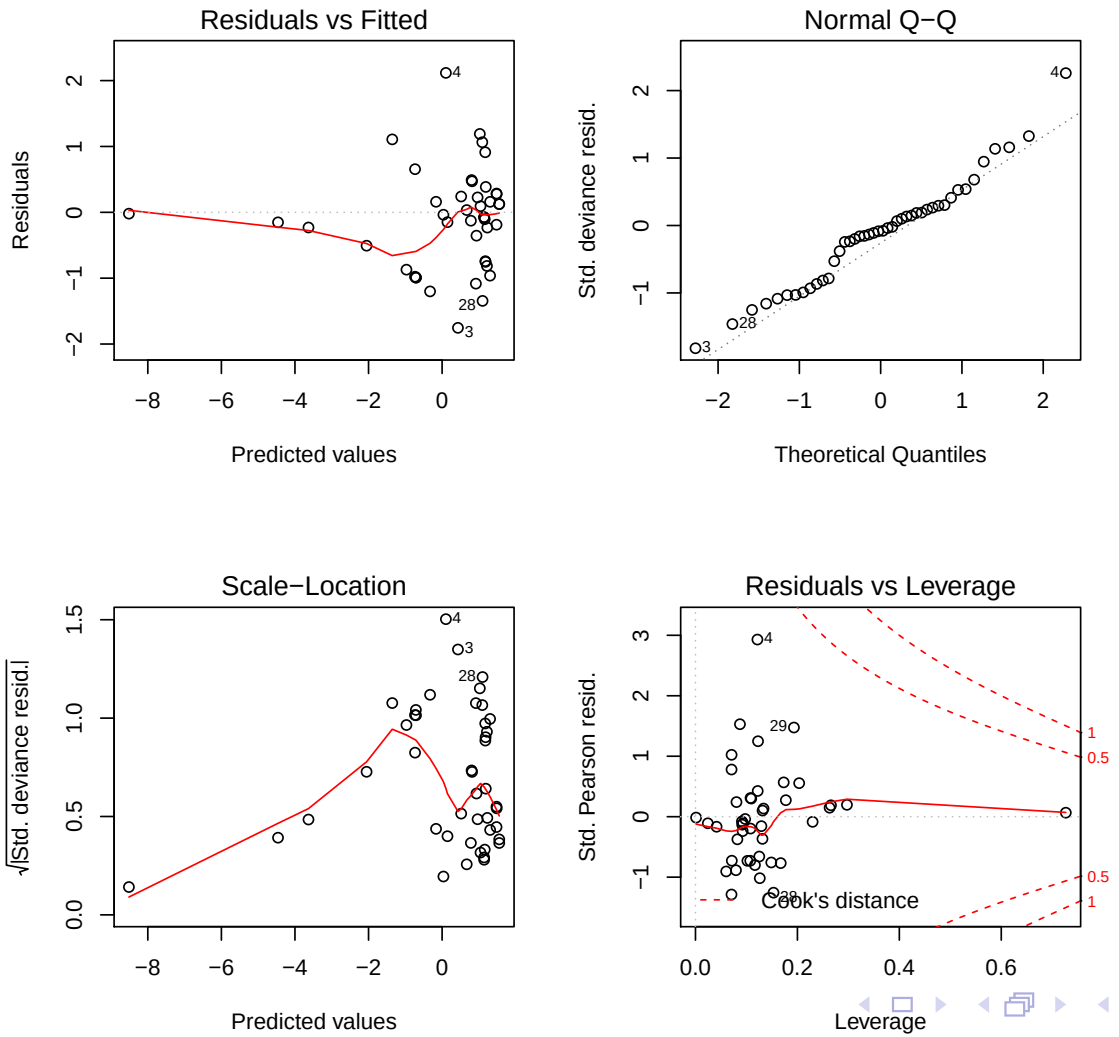
Let's do an added variable test:

```
> mines3C.glm<-glm(COUNT~ INB*EXTRP +  I(EXTRP^2),
                       family=poisson,data=mines.df)
> anova(mines3C.glm,mines3B.glm,test="Chisq")
Analysis of Deviance Table

Model 1: COUNT ~ INB * EXTRP + I(EXTRP^2)
Model 2: COUNT ~ INB * EXTRP + I(EXTRP^2) + AGE
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        39     27.734
2        38     24.984  1   2.7496  0.09728 .
```

- Weak evidence that AGE should be kept.

Diagnostic Plots



Estimated Coefficients

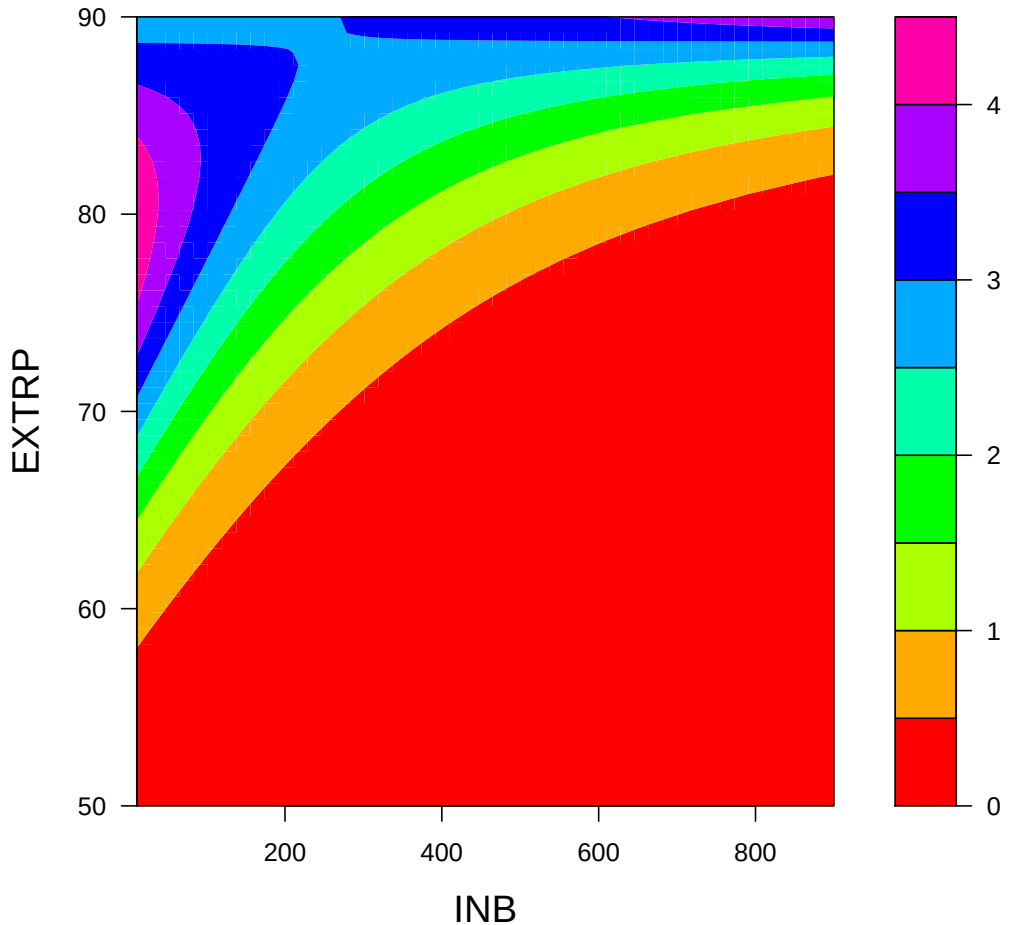
So our selected model has:

```
                Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.723e+01  1.086e+01  -2.507  0.01219 *
INB         -3.177e-02  1.321e-02  -2.405  0.01618 *
EXTRP        7.296e-01  2.806e-01   2.600  0.00931 **
I(EXTRP^2)  -4.602e-03  1.810e-03  -2.543  0.01100 *
AGE         -2.564e-02  1.584e-02  -1.618  0.10562
INB:EXTRP    3.580e-04  1.509e-04   2.372  0.01770 *
```

- As `AGE` increases number of accidents tends to decrease.
- Hard to tell impact of `INB` and `EXTRP` from this.

Filled Contour Plot

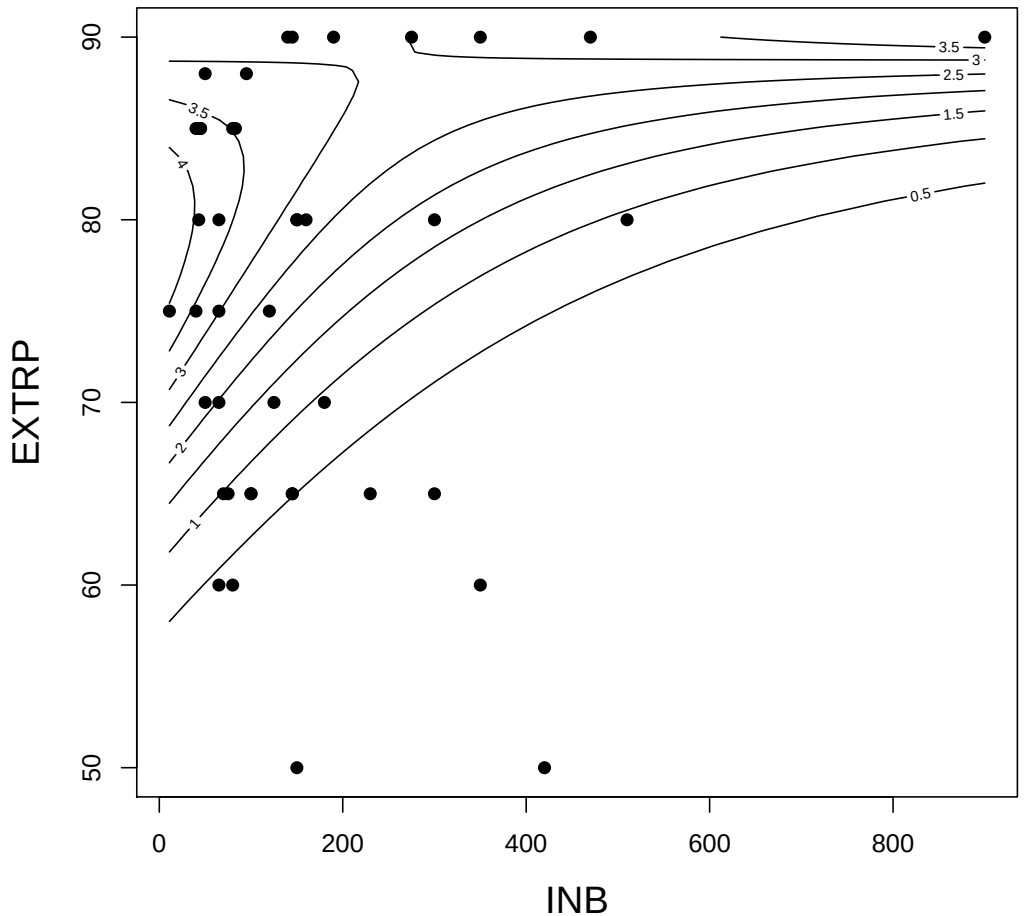
Filled contour plot for `COUNTS`.



- `AGE` is set to its mean value.

Filled Contour Plot

Contour plot for `COUNTS`.



- `AGE` is set to its mean value.

Conclusions

Conclusion based on the fitted model.

- The impact of `AGE` (age of mine) can be evaluated on its own as it doesn't depend on the levels of `INB` and `EXTRP`. For each unit increase in `AGE` the predicted number of accidents is multiplied by $\exp(-.02564) = .975$.
- The impact of `INB` (inner burden thickness) and `EXTRP` (percentage of coal extracted from mine) need to be evaluated jointly since the presence of an interaction indicates that the effect of one depends on the level of the other. Our plots indicate that the predicted number of accidents are the highest for `EXTRP` between 75 and 85 and for `INB` between 10 and 100.
- `AHS` (average height of the coal seam) does not have an appreciable impact on the predicted number of accidents.