

9

Checking the models II: the other three assumptions

9.1 Homogeneity of variance

Plotting the residuals against the fitted values is one of the standard techniques introduced in the main text to check the assumption of homogeneity of variance. It is illustrated here with the *diets* dataset. We use the standardised residuals rather than raw residuals for these model checking plots, as they are easier to interpret (see main text). SAS calls these ‘Studentized residuals’. This is illustrated first with the *diets* dataset.

SAS COMMANDS FOR BOX 9.1 **Weight change with diet**

```
Commands  proc glm data=gandh.Chapter9;
           class DIET;
           model WGHTCH = DIET / solution;
           output p=Predictd Student=SResid;

           proc gplot;
             plot SResid * Predictd;
           run;
```

Menu route Statistics > Anova > Linear Models...

WGHTCH → Dependent

DIET → Class

Plots with Residual tab

Plot residuals vs variables

Residuals: Studentized

Variables: Predicted Y

Statistics

Type I

Parameter estimates

96 Checking the models II: the other three assumptions

SAS OUTPUT FOR BOX 9.1 **Weight change with diet**

The GLM Procedure

Dependent Variable: WGHTCH

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	173.5632400	86.7816200	4.97	0.0267
Error	12	209.4237200	17.4519767		
Corrected Total	14	382.9869600			

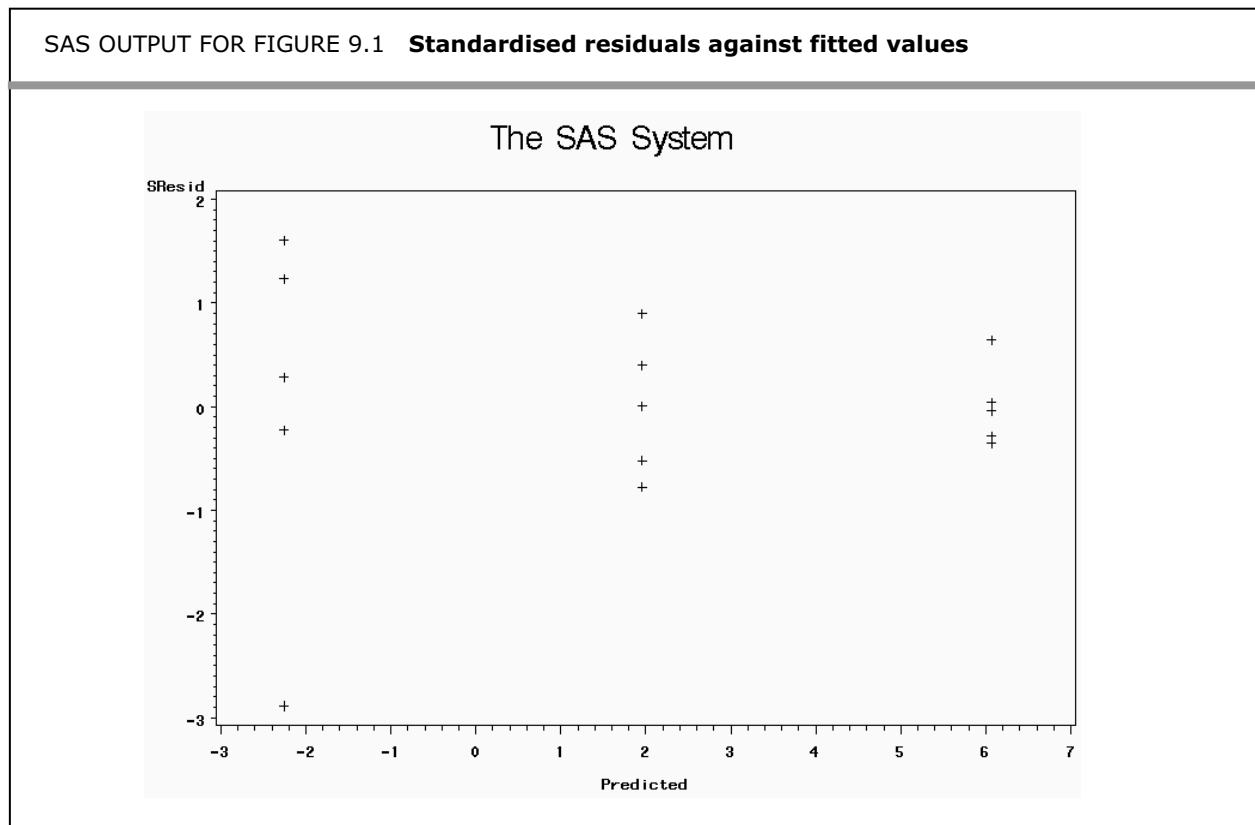
R-Square Coeff Var Root MSE WGHTCH Mean
 0.453183 217.1287 4.177556 1.924000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
DIET	2	173.5632400	86.7816200	4.97	0.0267

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DIET	2	173.5632400	86.7816200	4.97	0.0267

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-2.258000000 B	1.86825997	-1.21	0.2501
DIET 1	4.214000000 B	2.64211859	1.59	0.1367
DIET 2	8.332000000 B	2.64211859	3.15	0.0083
DIET 3	0.000000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.



If you are using the menu route, only the analysis comes up automatically after the commands have been submitted. You need to navigate in the left hand pane of the Analyst environment, to the part of the tree labelled 'Residual of...'. Double clicking on this will reveal the residual graph. In the command route, the residual graph appears automatically.

9.4 Model criticism and solutions

The Normality of error may be examined by plotting a histogram of residuals or a Normal Order plot. Normal Order plots allow a more quantitative assessment of the Normality of the distribution of residuals. This method will first of all be described with some Normally distributed residuals. Using the *plantlets* data set, the height of two species are compared. SAS's 'PROC CAPABILITY' can deal with the fact that your dataset has columns of different lengths. However, the menu route we describe here uses 'PROC UNIVARIATE', and this assumes the variables used are as long as the longest variable of the dataset. To avoid this complication, we read in the *plantlets* dataset separately for the next exercise.

SAS COMMANDS FOR BOX 9.2 AND FIGURE 9.6 **Analysis of height in the plantlets dataset**

```

Commands  proc glm data=gandh.plantlets;
           class SPECIES;
           model HGHT = SPECIES / SS1;
           output STUDENT=HSResids;

           proc capability;
           var HSResids;
           histogram / normal;
           probplot / normal(mu=est sigma=est);
           run;

```

Menu route *Import and open the plantlets dataset*

Step 1

Statistics > Anova > Linear Models...

HGHT → Dependent

SPECIES → Class

Save Data

Create and save diagnostics data

STUDENT → Add

Statistics

Type I

Step 2

Close the output window that will have been produced by executing the commands above. Double-click on the most recent 'Diagnostics Table' in left hand pane of Analyst window. You should be able to see the output dataset. It will look just like the original dataset, with the added column of residuals on the right hand side. With the diagnostics table still open, move on to step 3.

Step 3

File > Save as by SAS Name...

and save it in the WORK library (click on WORK in left hand pane, then type a name in the Member Name box, and click on Save). Remember the name you used! (*Contd.*)

(Contd.) *Step 4*

Close the Diagnostics Table window. This should return you to the Analyst environment.

Step 5

File > Open by SAS Name...
and open the file you have just saved

Step 6

Graphs > Histogram...
_STUDENT → Analysis

Fit
 Normal

At this point the histogram of residuals will be produced.

Step 7

Graphs > Probability Plot...
_STUDENT → Analysis

At this point the Normal probability plot will be produced.

SAS OUTPUT FOR BOX 9.2 Analysis of height in the plantlets dataset

The GLM Procedure

Dependent Variable: HGHT

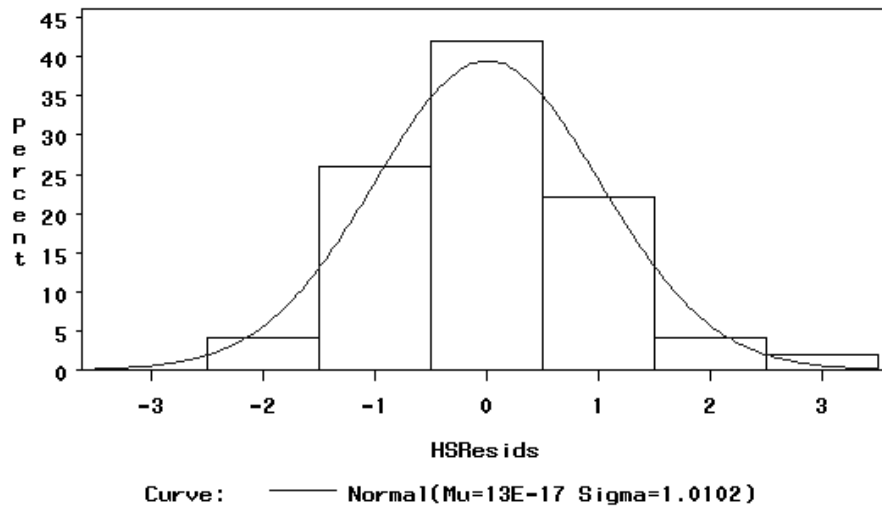
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	260.8328000	260.8328000	46.28	<.0001
Error	48	270.5464000	5.6363833		
Corrected Total	49	531.3792000			

R-Square	Coeff Var	Root MSE	HGHT Mean
0.490860	20.28458	2.374107	11.70400

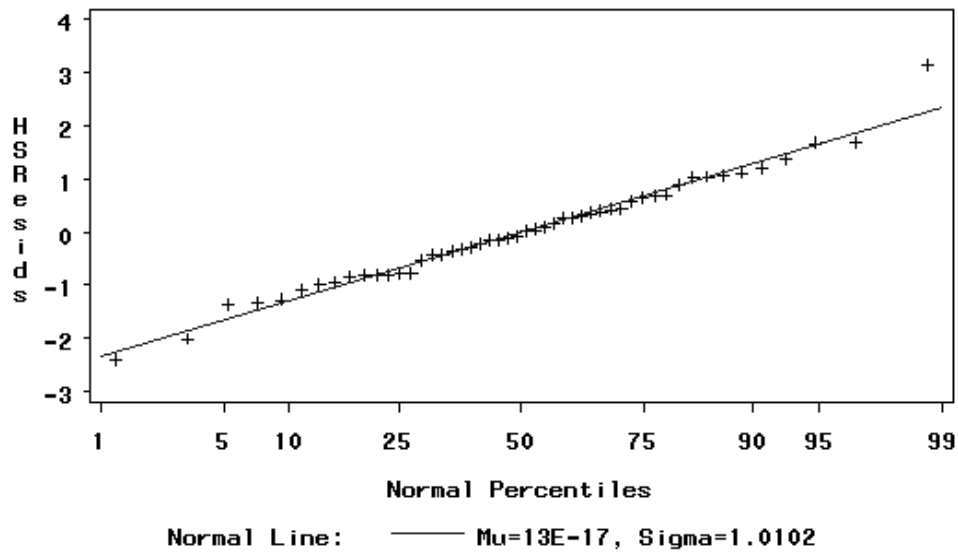
Source	DF	Type I SS	Mean Square	F Value	Pr > F
SPECIES	1	260.8328000	260.8328000	46.28	<.0001

SAS OUTPUT FOR FIGURE 9.6 **Normally distributed residual plots**

The SAS System



The SAS System



To see the plots produced by SAS, you may need to find the relevant icons. If you used the command route, the relevant icons will be in the Results pane, to the left of the screen. Double-click to navigate your way through the tree structure. If you have graphs already showing, you may need to close them first. If you used the menu route, the relevant icons will be in the left hand pane of the Analyst environment.

The residuals in Figure 9.6 appear to be Normally distributed and originate from a data set size 50. Imagine these standardised residuals being ranked. If these 50 values have been drawn at random from a Standard Normal Distribution (see revision section), then we can calculate the expected value of the lowest residual—which, in this case, we would expect to be -2.24 standard deviations below the mean of zero. Corresponding expected residuals can be calculated for all 50 points. This can be calculated in SAS by the use of the NORMAL option to the PROC RANK.

SAS can also store the actual standardised residuals produced by the fit of the data to the model you have specified. To do this, repeat the previous analysis, then use PROC RANK to put the normal scores in 'normscor'. The option 'BLM' refers to a mathematical rule by which SAS deals with the extremes of $-\infty$ and $+\infty$ of the Normal distribution. Finally, plot them against each other, as we now illustrate:

ALTERNATIVE SAS COMMANDS FOR FIGURE 9.6B	
Commands	<pre>proc glm data=gandh.plantlets; class SPECIES; model HGHT = SPECIES / SS1; output STUDENT=HSResids; proc rank normal=BLOM; var HSResids; ranks normscor; proc gplot; plot HSResids*normscor; run;</pre>
Menu route	<p><i>Steps 1–5</i> Exactly as in previous Box ‘SAS Commands for Box 9.2 and Figure 9.6’</p> <p><i>Step 6</i> Edit > Mode > Edit Data > Transform > Rank... _STUDENT → Rank Options <input checked="" type="radio"/> Normal Scores (This puts the normal scores of _STUDENT into a variable called _STUDENT_ranks.)</p> <p><i>Step 7</i> Graphs > Scatter Plot > Two-Dimensional... _STUDENT → Y Axis _STUDENT_rank → X Axis</p>

This produces a Normal probability plot similar to that in Figure 9.6. This exercise illustrates precisely where this plot comes from. For the other two examples in this section, we return to requesting the Normal probability plot as part of the analysis.

As the process is rather arduous, particularly in the menu route, we show how SAS can produce all the residuals and variables simultaneously for three response variables, provided the model is the same. The residuals are called `_STUDENT1`, `_STUDENT2` and `_STUDENT3` in the menu route, provided there are no existing residuals columns, but you will find them in any event at the end of the new dataset. We choose to call the residuals `HGHTres`, `LONGres` and `LSres` in the command route. This command essentially recreates the `HGHT` output as shown above, as well as the output for the two new variables (`LONGEV` and `LEAFSIZE`), so we will show only the new output. The `capability` procedure also produces a number of summary statistics which we do not include here.

Finally, note that SAS produces Normal order plots with the Normal percentiles on the x axis, whereas the plots in the main text have the equivalent Normal score on the y axis.

SAS COMMANDS FOR BOX 9.3 AND FIGURE 9.7 **Analysing longevity in the plantlets dataset**

```

Commands  proc glm data=gandh.plantlets;
           class SPECIES;
           model HGHT LONGEV LEAFSIZE = SPECIES / SS1;
           output STUDENT=HGHTres LONGres LSres;

           proc capability;
           var HGHTres LONGres LSres;
           histogram / normal;
           probplot / normal(mu=est sigma=est);
           run;

```

Menu route *Import and open the plantlets dataset*

Step 1

Statistics > Anova > Linear Models...

HGHT LONGEV LEAFSIZE → Dependent

SPECIES → Class

Create and save diagnostics data

STUDENT → Add

Step 2

Double-click on the most recent 'Diagnostics Table' in left hand pane of Analyst window. You should be able to see the output dataset. It will look just like the original dataset, as the added columns of residuals are way off on the right hand side.

Step 3

File > Save as by SAS Name...

and save it in the WORK library (click on WORK in left hand pane, then type a name in the Member Name box, and click on Save). Remember the name you used!

Step 4

Close the Diagnostics Table window. This should return you to the Analyst environment.

(Contd.)

(Contd.)

Step 5

File > Open by SAS Name...

and open the file you have just saved

Step 6

Graphs > Histogram...

_STUDENT1 _STUDENT2 _STUDENT3 → Analysis

Fit

Normal

This will produce three histograms—you may need to navigate in the left hand pane of the Analyst environment to view them.

Step 7

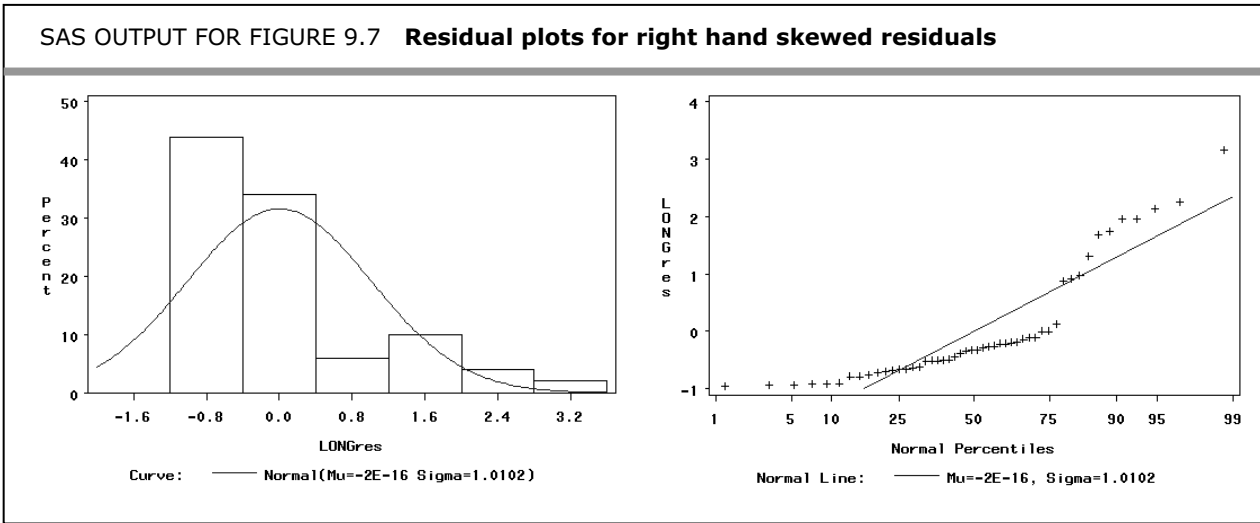
Graphs > Probability Plot...

_STUDENT1 _STUDENT2 _STUDENT3 → Analysis

This will produce three probability plots – again navigate in the left hand pane of the Analyst environment to view them.

SAS OUTPUT FOR BOX 9.3 Analysing longevity in the plantlets dataset

The GLM Procedure					
Dependent Variable: LONGEV					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3299.9688	3299.9688	0.30	0.5854
Error	48	525102.6200	10939.6379		
Corrected Total	49	528402.5888			
	R-Square	Coeff Var	Root MSE	LONGEV Mean	
	0.006245	116.5145	104.5927	89.76800	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
SPECIES	1	3299.968800	3299.968800	0.30	0.5854



SAS OUTPUT FOR BOX 9.4 **Analysing leaf size in the plantlets dataset**

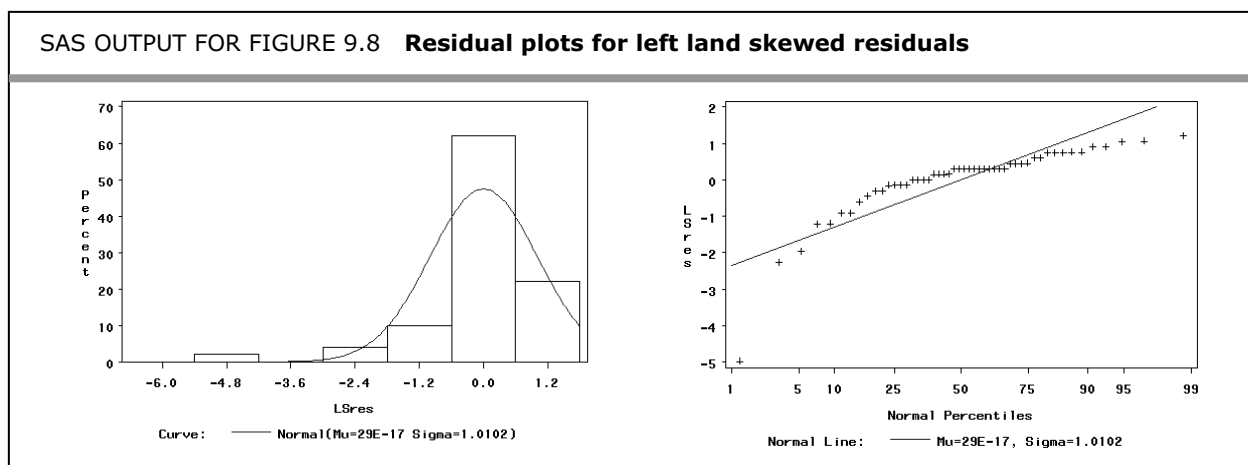
The GLM Procedure

Dependent Variable: LEAFSIZE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.11520000	0.11520000	0.25	0.6180
Error	48	21.94960000	0.45728333		
Corrected Total	49	22.06480000			

R-Square	Coeff Var	Root MSE	LEAFSIZE Mean
0.005221	18.02312	0.676227	3.752000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SPECIES	1	0.11520000	0.11520000	0.25	0.6180



Plotting residuals against fitted values

This technique was introduced earlier in this chapter as a means of checking for homogeneity of variance. It can also illustrate problems with linearity, as illustrated with the 3Ys dataset.

SAS COMMANDS FOR BOX 9.5 AND FIGURE 9.9 **Dataset 3Ys, Y1**

```
Commands  proc glm data=gandh.Chapter9;
           model Y1 = X;
           output p=FV1 r=R1;

           proc gplot;
             plot R1 * FV1;
           run;
```

Menu route Statistics > Anova > Linear Models...

Y1 → Dependent

X → Quantitative

Plots with Residual tab

Plot residuals vs variables

Residuals: Ordinary

Variables: Predicted Y

108 Checking the models II: the other three assumptions

SAS OUTPUT FOR BOX 9.5 Dataset 3Ys, Y1

The GLM Procedure

Dependent Variable: Y1

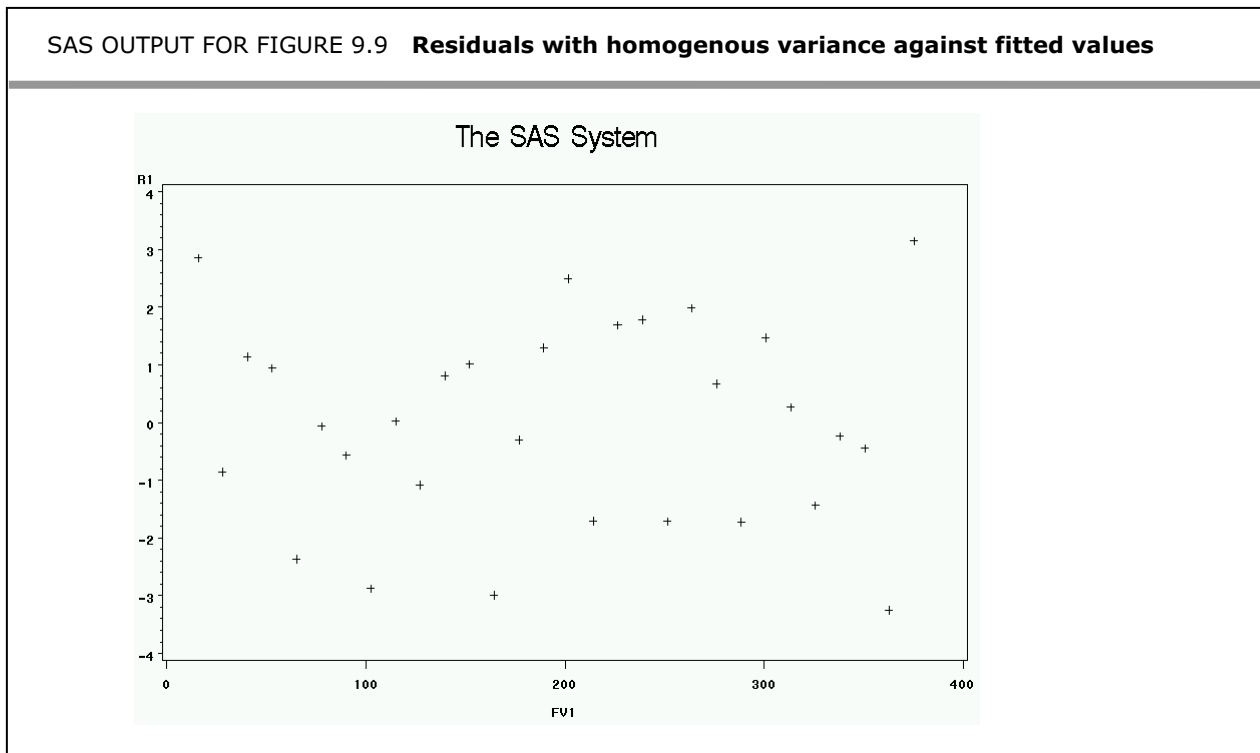
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	345771.5478	345771.5478	108377	<.0001
Error	28	89.3322	3.1904		
Corrected Total	29	345860.8800			

R-Square	Coeff Var	Root MSE	Y1 Mean
0.999742	0.913647	1.786179	195.5000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
X	1	345771.5478	345771.5478	108377	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X	1	345771.5478	345771.5478	108377	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	3.24551724	0.66887562	4.85	<.0001
X	12.40351502	0.03767691	329.21	<.0001



The same commands may be repeated for the other two variables in this dataset, in which the residuals show increasing and decreasing variance with the fitted values (see Figure 9.10 and 9.11 in the main text).

Plotting the residuals against each explanatory variable

We will illustrate this with the second response variable in the *3Ys* dataset. In this instance, plotting against the fitted values, and plotting against the explanatory variable (*X*) will result in very much the same graph, as there is only one explanatory variable. However, in models with several explanatory variables, the same technique may allow you to pinpoint which explanatory variable may be responsible for the non-linearity. Note that we ask to plot ‘Standardized’ residuals in the menu route as the equivalent of ‘STUDENT’ in the command line route. (The menu’s ‘Studentized residuals’ correspond to the ‘RSTUDENT’ residuals, which are calculated by omitting the datapoint in question from the analysis.) This can be confusing because when saving (as opposed to plotting) residuals, the menu route uses the same terms as the command line route. Thus to save what the menu plot subdialog box calls ‘Standardized’ residuals, you must ask the command save subdialog to save ‘STUDENT’. To summarise this rather confusing situation, see the Table below.

NAMES USED IN SAS FOR THE STANDARDIZED RESIDUALS OF THE MAIN TEXT		
Route	Saving	Plotting
Menu	Use ‘Save Data’ button and choose STUDENT as in SAS Command Box 9.3	Use ‘plots’ button and choose STANDARDIZED as in SAS Commands for Figure 9.13
Command	Use output subcommand and call them STUDENT—save to a new variable with your chosen name.	You need to save them to plot them—by which time they are called by your own name. In SAS Commands for Figure 9.13 we called them SResid.

110 Checking the models II: the other three assumptions

SAS COMMANDS FOR AN EXAMPLE OF FIGURE 9.13

```
Commands  proc glm data=gandh.Chapter9;
           model Y2 = X;
           output Student=SResid;

           proc gplot;
             plot SResid * X;
           run;
```

Menu route Statistics > Anova > Linear Models...

Y2 → Dependent

X → Quantitative

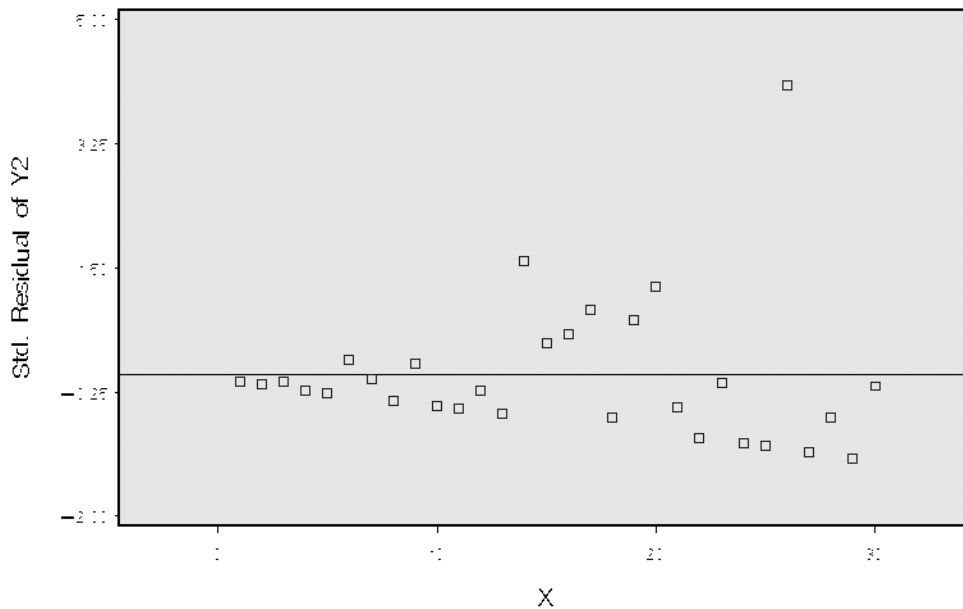
Plots with Residual tab

Plot residuals vs variables

Residuals: Standardized

Variables: Independents

SAS OUTPUT ILLUSTRATING PROBLEMS WITH LINEARITY (AS IN FIGURE 9.13)



These techniques may then be applied to the *bacteria* dataset, to resolve whether interactions or transformations are the better way of tackling non-linearity and heterogeneity of variance, as illustrated in the subsequent section of the main text.

9.5 Predicting the volume of merchantable wood: an example of model criticism

This example brings together both model criticism and basic data manipulation techniques, with an analysis based on the *merchantable timber* dataset. This case study starts off with an analysis of the untransformed data.

SAS COMMANDS FOR BOX 9.11 AND FIGURE 9.19 **Untransformed analysis of tree volume**

```
Commands  proc glm data=gandh.Chapter9;
           model VOLUME = DIAM;
           output STUDENT=SResids p=Predicted;

           proc gplot;
             plot SResids * Predicted;
           run;
```

Menu route Statistics > Anova > Linear Models...

VOLUME → Dependent

DIAM → Quantitative

Plots with Residual tab

- Plot residuals vs variables
- Residuals: Standardized
- Variables: Predicted Y

112 Checking the models II: the other three assumptions

SAS OUTPUT FOR BOX 9.11 Untransformed analysis of tree volume

The GLM Procedure

Dependent Variable: VOLUME

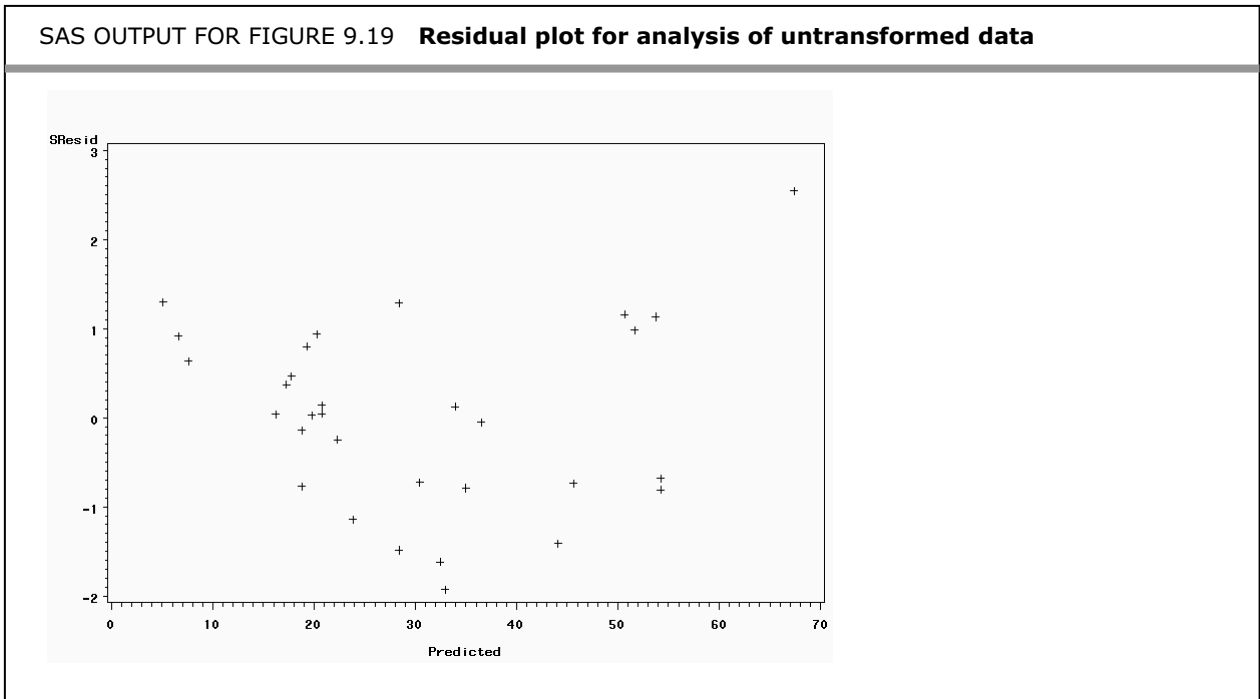
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	7581.910472	7581.910472	419.47	<.0001
Error	29	524.173399	18.074945		
Corrected Total	30	8106.083871			

R-Square	Coeff Var	Root MSE	VOLUME Mean
0.935336	14.09124	4.251464	30.17097

Source	DF	Type I SS	Mean Square	F Value	Pr > F
DIAM	1	7581.910472	7581.910472	419.47	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DIAM	1	7581.910472	7581.910472	419.47	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-36.94487574	3.36477068	-10.98	<.0001
DIAM	60.79138258	2.96818649	20.48	<.0001



The second analysis transforms the data and includes two explanatory variables. To take the natural logs of VOLUME, HEIGHT and DIAM, follow the commands or menus in the next box. The second analysis is then performed on the transformed data. In the menu route, the new variables will remain available until you quit SAS, when you will be given the choice of saving the changes to the dataset. In the command route, the DATA step creates a temporary dataset which we call WORK.LOGTREES, which we can refer to until we quit SAS. If we wanted to have the new dataset LOGTREES available afterwards, we should save it to a different 'library'. For example, to save it into the same library as the datasets for the book, we would save it as GANDH.LOGTREES, simply by substituting WORK.LOGTREES with GANDH.LOGTREES in the following commands.

SAS COMMANDS TO TRANSFORM THE DATA	
Commands	<pre>data WORK.LOGTREES; set gandh.Chapter9; LVOL =log(VOLUME); LHGHT =log(HEIGHT); LDIAM =log(DIAM); run;</pre>
Menu route	<p>Edit > Mode > Edit Data > Transform > Compute...</p> <p>Type "LVOL" into top left box log (VOLUME) → Main pane</p> <p><i>Repeat with (LHGHT, HEIGHT) and then (LDIAM, DIAM) in place of (LVOL, VOLUME).</i></p>

We now analyse the transformed data.

114 Checking the models II: the other three assumptions

SAS COMMANDS TO ANALYSE THE DATA FOR BOX 9.12 AND FIGURE 9.20 Comparing trees to cylinders

```

Commands  proc glm data= WORK.LOGTREES;
           model LVOL = LHGHT LDIAM;
           output STUDENT=SResids p=Predicted;

           proc gplot;
             plot SResids * Predicted;
           run;
    
```

Menu route Statistics > Anova > Linear Models...

LVOL → Dependent

LHGHT LDIAM → Quantitative

Plots with Residual tab

- Plot residuals vs variables
- Residuals: Standardized
- Variables: Predicted Y

SAS OUTPUT FOR BOX 9.12 Comparing trees to cylinders

The GLM Procedure

Dependent Variable: LVOL

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	8.12325410	4.06162705	613.29	<.0001
Error	28	0.18543543	0.00662269		
Corrected Total	30	8.30868953			

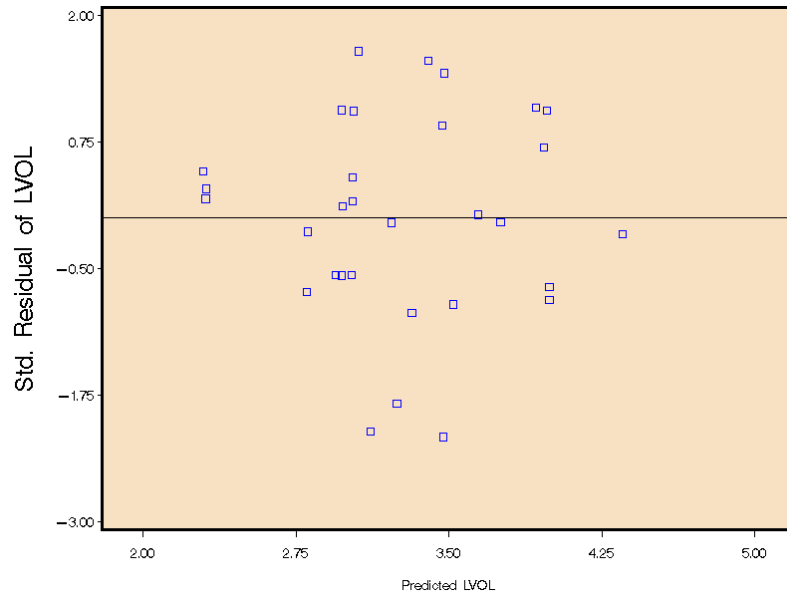
	R-Square	Coeff Var	Root MSE	LVOL Mean
	0.977682	2.486606	0.081380	3.272732

Source	DF	Type I SS	Mean Square	F Value	Pr > F
LHGHT	1	3.49572253	3.49572253	527.84	<.0001
LDIAM	1	4.62753157	4.62753157	698.74	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
LHGHT	1	0.19776393	0.19776393	29.86	<.0001
LDIAM	1	4.62753157	4.62753157	698.74	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-1.704745012	0.88183561	-1.93	0.0634
LHGHT	1.117080728	0.20442234	5.46	<.0001
LDIAM	1.982709240	0.07500698	26.43	<.0001

SAS OUTPUT FOR FIGURE 9.20 **Residual plot for tree volume treated as a cylinder**



The third step in this case study is to estimate the mean deviation of the intercept from the expected cylinder intercept. This is done by creating a new variable, and requesting its mean and standard error, using the following commands:

SAS COMMANDS FOR BOX 9.13

Estimating the mean deviation of the intercept from the expected cylinder intercept

```

Commands  data;
           set WORK.LOGTREES;
           LDEVINT = LVOL - LHGHT - 2*LDIAM;

           proc means mean clm;
           var LDEVINT;
           run;
    
```

Menu route *Step 1*

Data > Transform > Compute...

Type "LDEVINT" into top left box

LVOL - LHGHT - 2*LDIAM → Main pane

Step 2

Statistics > Hypothesis Tests > One-Sample t-test for a Mean...

LDEVINT → Variable

Tests

with Confidence Intervals tab

Interval

SAS OUTPUT FOR BOX 9.13

Estimating the mean deviation of the intercept from the expected cylinder intercept

The MEANS Procedure

Analysis Variable : LDEVINT

Mean	Lower 95% CL for Mean	Upper 95% CL for Mean
-1.1993619	-1.2283724	-1.1703514

Finally, we estimate the proportion of the cylinder that represents useable timber.

SAS COMMANDS FOR BOX 9.14

Estimating the proportion of the cylinder that represents useable timber

```

Commands  data;
           set WORK.LOGTREES;
           LUSEFRAC = LVOL - (-0.242 + LHGHT + 2*LDIAM) ;

           proc means mean clm;
           var LUSEFRAC;
           run;
    
```

Menu route *Step 1*

Data > Transform > Compute...

Type "LUSEFRAC" into top left box

LVOL - (-0.242 + LHGHT + 2*LDIAM) → Main pane

Step 2

Statistics > Hypothesis Tests > One-Sample t-test for a Mean...

LUSEFRAC → Variable

Tests with Confidence Intervals tab

Interval

SAS OUTPUT FOR BOX 9.14

Estimating the proportion of the cylinder that represents useable timber

The MEANS Procedure

Analysis Variable : LUSEFRAC

Mean	Lower 95% CL for Mean	Upper 95% CL for Mean
-0.9573619	-0.9863724	-0.9283514

9.8 Exercises

Stabilising the variance

The SAS commands used in this exercise fall into three parts.

1. Transforming the data

SAS COMMANDS TO TRANSFORM DATA	
Commands	<pre>data WORK.COTYLS; set gandh.Chapter9; SRDAM =sqrt (DAMAGE) ; LOGDAM=log (DAMAGE) ; INVDAM =-1/ DAMAGE; run;</pre>
Menu route	<p>Edit > Mode > Edit</p> <p>Data > Transform > Compute...</p> <p>Type "SRDAM" into top left box</p> <p>sqrt(DAMAGE) → Main pane</p> <p>Data > Transform > Compute...</p> <p>Type "LOGDAM" into top left box</p> <p>log(DAMAGE) → Main pane</p> <p>Data > Transform > Compute...</p> <p>Type "INVDAM" into top left box</p> <p>-1/ DAMAGE → Main pane</p>

2. Plotting the graphs

SAS COMMANDS FOR FIGURE 9.21

```

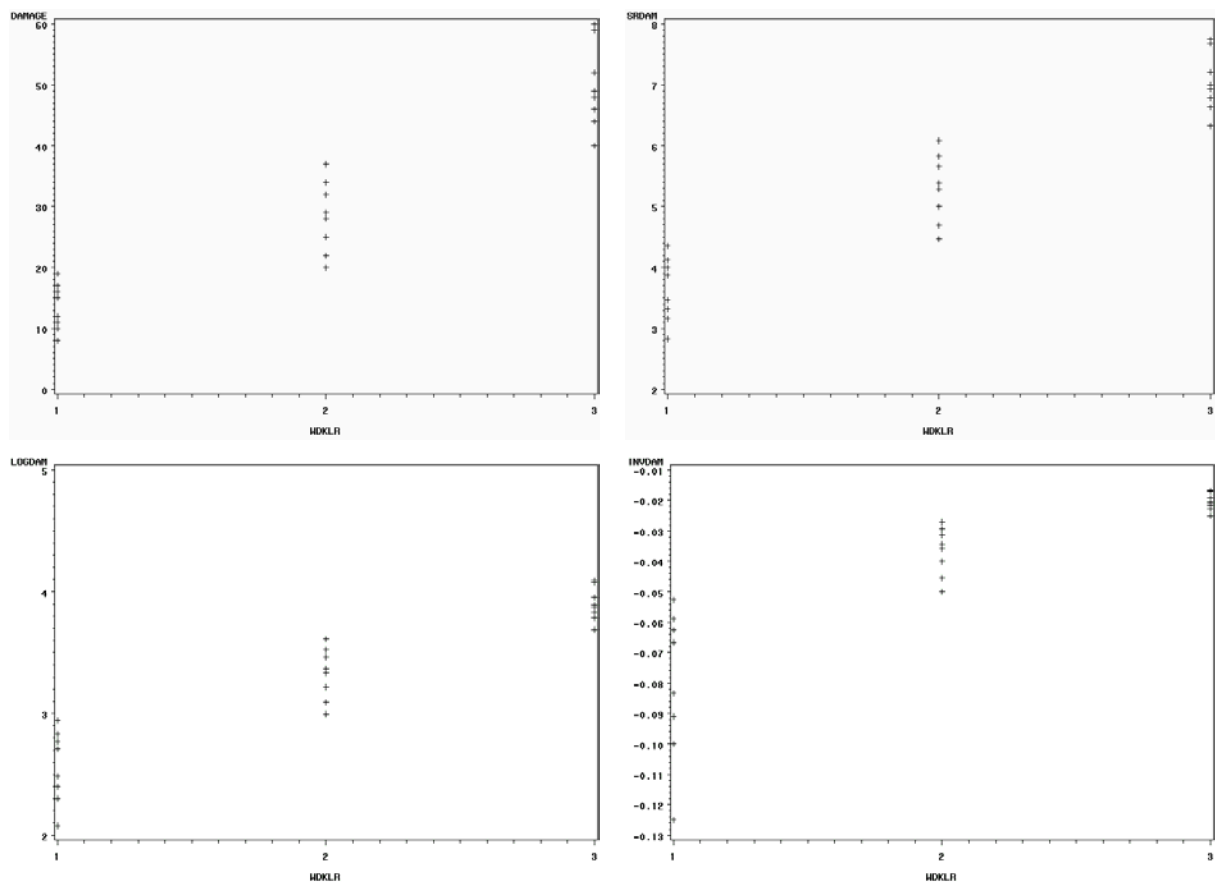
Commands  data;
          set WORK.COTYLS;
          proc gplot;
            plot DAMAGE*WDKLR;
            plot SRDAM*WDKLR;
            plot LOGDAM*WDKLR;
            plot INV DAM*WDKLR;
          run;

```

Menu route Graphs > Scatter Plot > Two-Dimensional

WDKLR → X Axis

DAMAGE SRDAM LOGDAM INV DAM → Y Axis

SAS OUTPUT FOR FIGURE 9.21 **Plots for the Soyabean dataset**

120 Checking the models II: the other three assumptions

3. Obtaining descriptive statistics

SAS COMMANDS FOR BOX 9.15	
Commands	<pre>proc means data=WORK.COTYLS; by WDKLR; var DAMAGE SRDAM LOGDAM INVDAM; run;</pre>
Menu route	Statistics > Descriptive > Summary Statistics... DAMAGE SRDAM LOGDAM INVDAM → Analysis WDKLR → Class

SAS OUTPUT FOR BOX 9.15 Basic statistics for the soyabean dataset								
The MEANS Procedure								
WDKLR	N	Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
1	8	DAMAGE	8	13.5000000	3.8172541	8.0000000	19.0000000	
		SRDAM	8	3.6408024	0.5286727	2.8284271	4.3588989	
		LOGDAM	8	2.5653900	0.2972337	2.0794415	2.9444390	
		INVDAM	8	-0.0799830	0.0246014	-0.1250000	-0.0526316	
2	8	DAMAGE	8	28.3750000	5.8782286	20.0000000	37.0000000	
		SRDAM	8	5.3012235	0.5575760	4.4721360	6.0827625	
		LOGDAM	8	3.3260207	0.2132570	2.9957323	3.6109179	
		INVDAM	8	-0.0366675	0.0079883	-0.0500000	-0.0270270	
3	8	DAMAGE	8	49.7500000	6.9846771	40.0000000	60.0000000	
		SRDAM	8	7.0383191	0.4922992	6.3245553	7.7459667	
		LOGDAM	8	3.8984822	0.1393193	3.6888795	4.0943446	
		INVDAM	8	-0.0204443	0.0028222	-0.0250000	-0.0166667	

Stabilising the variance in a blocked experiment

The SAS output for this exercise may be found in the answers to exercises.

Lizard skulls

The first analysis is with the raw data.

SAS COMMANDS FOR BOX 9.16 AND FIGURE 9.22

```
Commands  proc glm data=gandh.Chapter9;
           model BVOL = JAWL / solution;
           output STUDENT=JSRes p=JPred;

           proc gplot;
             plot JSRes * JPred;
           run;
```

Menu route Statistics > Anova > Linear Models...

BVOL → Dependent

JAWL → Quantitative

Parameter estimates

with Residual tab

Plot residuals vs variables

Residuals: Standardized

Variables: Predicted Y

Which gives the following output:

122 Checking the models II: the other three assumptions

SAS OUTPUT FOR BOX 9.16 Analysis of lizard skulls

The GLM Procedure

Dependent Variable: BVOL

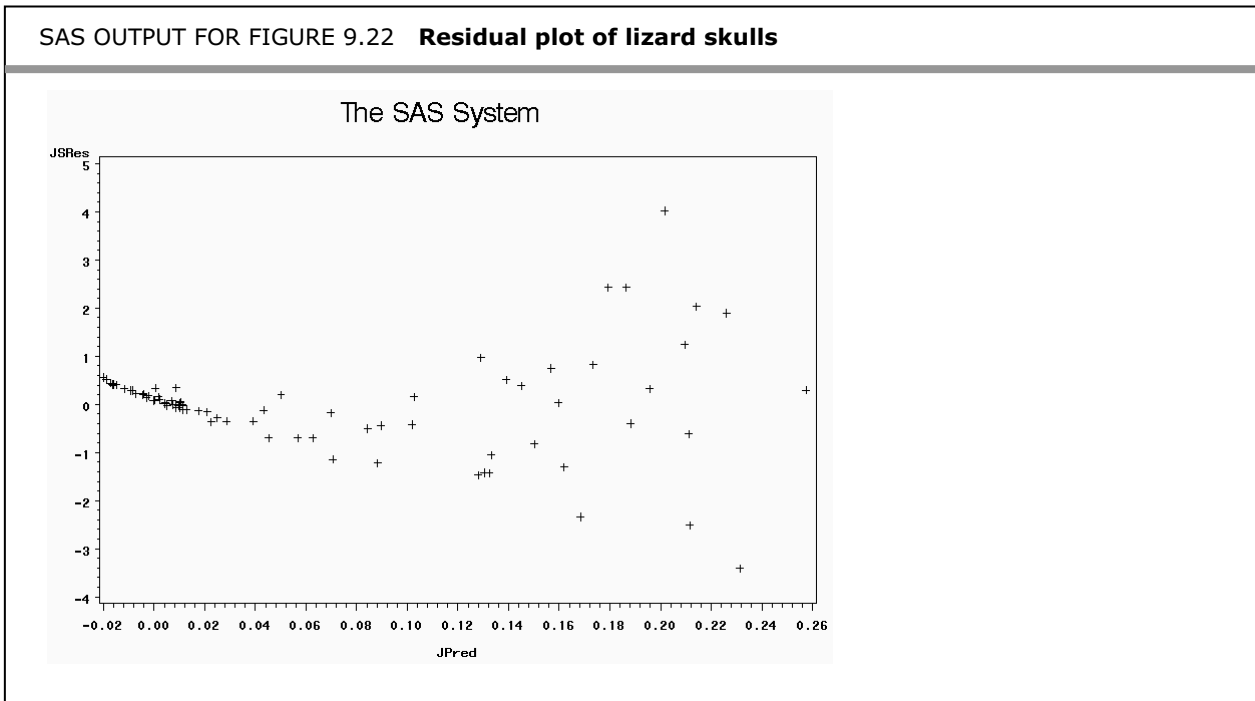
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.52277903	0.52277903	284.65	<.0001
Error	78	0.14324990	0.00183654		
Corrected Total	79	0.66602894			

R-Square	Coeff Var	Root MSE	BVOL Mean
0.784919	63.23804	0.042855	0.067768

Source	DF	Type I SS	Mean Square	F Value	Pr > F
JAWL	1	0.52277903	0.52277903	284.65	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
JAWL	1	0.52277903	0.52277903	284.65	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-.0415829798	0.00806001	-5.16	<.0001
JAWL	0.5245633683	0.03109129	16.87	<.0001



The second analysis is performed on the transformed allometric measurements, and includes SITE as an extra explanatory variable.

SAS COMMANDS FOR BOX 9.17 AND FIGURE 9.23 **Alternative analysis for lizard skulls**

```

Commands  data;
           set gandh.Chapter9;
           LBVOL =log(BVOL);
           LJAWL =log(JAWL);

           proc glm;
             class SITE;
             model LBVOL = SITE LJAWL / solution;
             output p=LJPred STUDENT=LJres;

           proc gplot;
             plot LJPred * LJres;
           run;

```

Menu route Edit > Mode > Edit
 Data > Transform > Compute...
 Type "LBVOL" into top left box
 log(BVOL) → Main pane

Data > Transform > Compute...
 Type "LJAWL" into top left box
 log(JAWL) → Main pane

Statistics > Anova > Linear Models...
 LBVOL → Dependent
 SITE → Class
 LJAWL → Quantitative

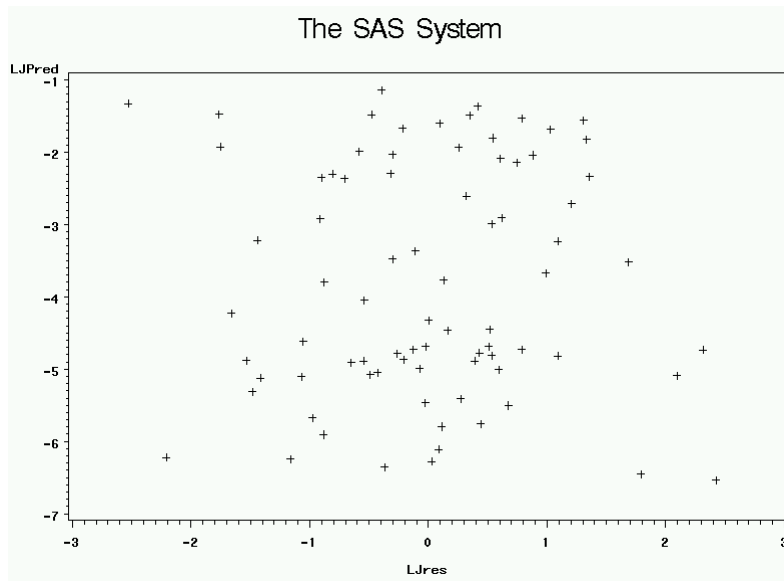
Parameter estimates

with Residual tab

Plot residuals vs variables
 Residuals: Standardized
 Variables: Predicted Y

124 Checking the models II: the other three assumptions

SAS OUTPUT FOR BOX 9.17 Alternative analysis for lizard skulls					
The GLM Procedure					
Dependent Variable: LBVOL					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	203.8496484	22.6499609	114.33	<.0001
Error	70	13.8674233	0.1981060		
Corrected Total	79	217.7170716			
	R-Square	Coeff Var	Root MSE	LBVOL Mean	
	0.936305	-11.64867	0.445091	-3.820960	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
SITE	8	193.4450991	24.1806374	122.06	<.0001
LJAWL	1	10.4045493	10.4045493	52.52	<.0001
Source	DF	Type III SS	Mean Square	F Value	Pr > F
SITE	8	0.30129326	0.03766166	0.19	0.9915
LJAWL	1	10.40454929	10.40454929	52.52	<.0001
Parameter	Estimate	Standard Error	t Value	Pr > t	
Intercept	-0.194509289 B	0.83651078	-0.23	0.8168	
SITE 1	0.091062068 B	0.59273817	0.15	0.8783	
SITE 2	0.128745869 B	0.27097206	0.48	0.6362	
SITE 3	0.147941590 B	0.33413989	0.44	0.6593	
SITE 4	-0.022697369 B	0.29941546	-0.08	0.9398	
SITE 5	0.048001363 B	0.47863989	0.10	0.9204	
SITE 6	0.074226379 B	0.32878683	0.23	0.8220	
SITE 7	-0.056013253 B	0.23989853	-0.23	0.8161	
SITE 8	0.167927931 B	0.62051121	0.27	0.7875	
SITE 9	0.000000000 B	.	.	.	
LJAWL	1.990412653	0.27465047	7.25	<.0001	

SAS OUTPUT FOR FIGURE 9.23 **Residual plot for alternative analysis of lizard skulls****Checking the perfect model**

The SAS output for this exercise may be found in the answers to exercises.