Chapter 1 Solutions

1.1 Part (c) is false. The predicted value of Y when $X = 2$ is $\hat{Y} = 100 + 15(2) = 130$, not 110. Parts (a), (b), and (d) are true.

1.2 A residual plot does not help assess (c) the condition of independence of the residuals. It does help assess (a) linearity, (b) constant variance, and (d) zero mean.

1.3 The slope is given in the output under the heading `Coef` for the predictor $WingLength$. The estimate is $\hat{\beta}_1 = 0.4674$.

1.4 The slope is given in the output under the heading `Coef` for the predictor $Year$. The estimate is $\hat{\beta}_1 = 0.01251$.

1.5 The intercept is given in the output under the heading `Coef` for the `Constant`. The estimate is $\hat{\beta}_0 = 1.3655$.

1.6 The intercept is given in the output under the heading `Coef` for the `Constant`. The estimate is $\hat{\beta}_0 = -16.47$.

1.7 As wing length increases by 1 mm, the weight increases by 0.4674 g, on average.

1.8 As year increases by 1, the length of the winning long jump increases by 0.01251 m, on average.

1.9 The regression standard error is given in the output as `S = 1.39959`. We can also compute this from the information given in the `Error` row of the Analysis of Variance:

$$\hat{\sigma}_\epsilon = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{223.31}{116-2}} = \sqrt{1.959} = 1.39959$$

A typical deviation of a sparrow weight from the line predicted by its wing length might be about 1.4 grams.

1.10 The regression standard error is given in the output as `S = 0.259522`. We can also compute this from the information given in the `Error` row of the Analysis of Variance:

$$\hat{\sigma}_\epsilon = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{1.751}{28-2}} = \sqrt{0.06735} = 0.2595$$

A typical deviation of a winning Olympic long jump length from the line predicted by its year might be about 0.26 meters.

1.11 The degrees of freedom for the regression standard error are $n - 2 = 116 - 2 = 114$. The value also appears in the `DF` column of the Analysis of Variance section of the output.

1.12 The degrees of freedom for the regression standard error are $n - 2 = 28 - 2 = 26$. The value also appears in the `DF` column of the Analysis of Variance section of the output.

1.13 The predicted value is $\hat{y}_1 = 25 + 7(10) = 95$. The residual is $y_1 - \hat{y}_1 = 100 - 95 = 5$.

1.14 The predicted value is $\hat{y}_1 = 78 - 0.5(30) = 63$. The residual is $y_1 - \hat{y}_1 = 60 - 63 = -3$.

1.15 a. Computer output gives the fitted regression model as $\hat{Width} = 37.72 - 0.01756Year$

b. As *Year* increases by 1, *Width* decreases by 0.01756 mm, on average.

c. Plugging 1966 into the fitted regression equation, we get $37.72 - 0.01756(1966) = 3.197$ mm.

1.16 a. The computer output gives the fitted regression model as $\hat{Eggs} = -8.98 + 7.33Lantern$.

b. As lantern size increases by 1 mm, the predicted number of eggs laid increases by 7.3 on average.

c. Plugging 14 into the fitted regression equation, we get $-8.98 + 7.33(14) = 93.6$ eggs.

1.17 a. The computer output gives the fitted regression equation as $\hat{MaxGripStrength} = 36.16 + 4.705Attractive$.

b. As *Attractive* increases by 1, *MaxGripStrength* increases by 4.7 kg, on average.

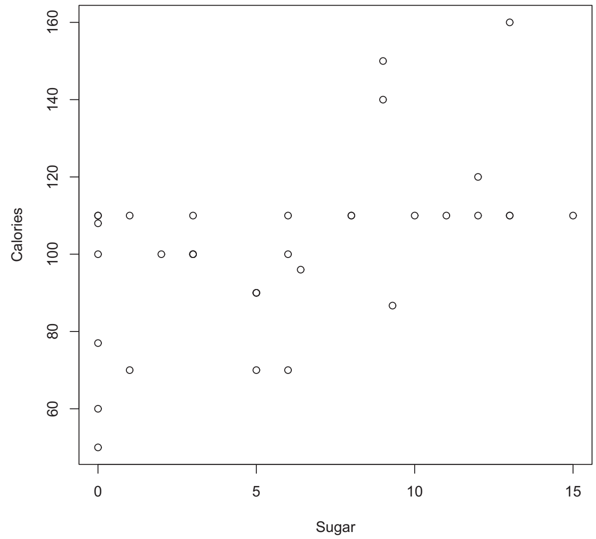
c. Plugging 3 into the fitted equation from part (a) we get a predicted $\hat{MaxGripStrength} = 36.16 + 4.705(3) = 50.3$ kg.

1.18 a. The computer output gives the fitted regression equation as $\hat{MaxGripStrength} = 9.3 + 29.0SHR$.

b. As *SHR* increases by 1, *MaxGripStrength* increases by 29 kg, on average.

c. Plugging 1.5 into the fitted equation from part (a) we get a predicted $\hat{MaxGripStrength} = 9.3 + 29.0(1.5) = 52.8$ kg.

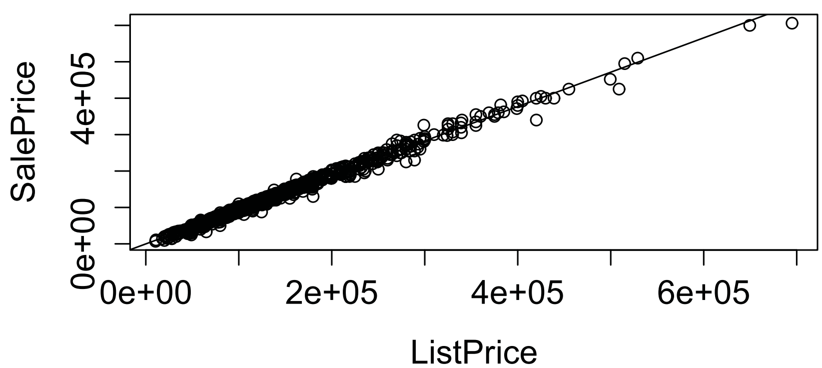
1.19 a. The scatterplot shows a moderate positive association between $Calories$ and $Sugar$.



b. Based on regression output, the prediction equation is $\widehat{Calories} = 87.43 + 2.48 Sugar$.

c. For every additional gram of sugar in a serving of cereal, the expected calories increase by 2.48 calories.

1.20 a. There is a clear, linear, and strong relationship between list price and sale price, as the plot indicates.



b. The regression summary is given below.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.448e+02  5.236e+02  -0.277    0.782
ListPrice    9.431e-01  3.201e-03 294.578   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8019 on 927 degrees of freedom
Multiple R-squared:  0.9894,Adjusted R-squared:  0.9894
```

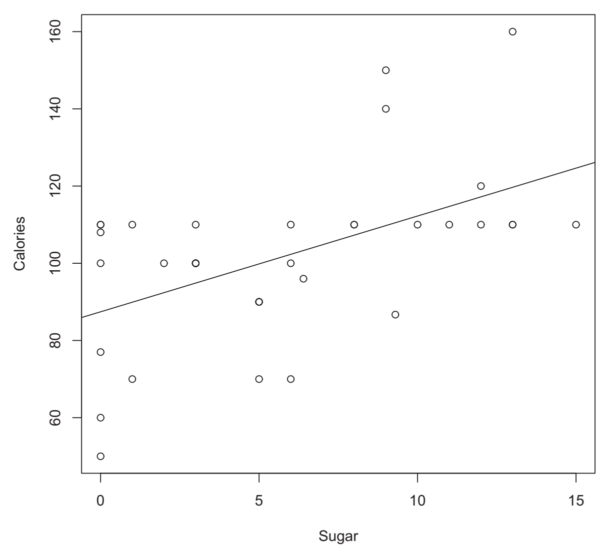
This shows us that the regression equation is $\widehat{SalePrice} = -144.8 + 0.943 ListPrice$.

c. Each increase of a dollar to the list price corresponds to a \$0.94 increase in sales price.

1.21 a. The prediction equation is $\widehat{Calories} = 87.43 + 2.48 Sugar$, so when $Sugar = 10$, the predicted $Calories$ is $\widehat{Calories} = 87.43 + 2.48(10) = 112.23$ calories.

b. For Cheerios, $\widehat{Calories} = 87.43 + 2.48(1) = 89.91$, so the residual is $110 - 89.91 = 20.09$ calories.

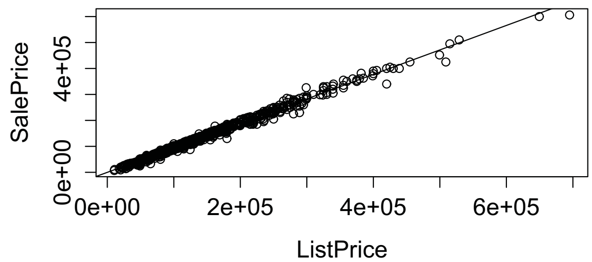
c. Although there is a somewhat positive association, there is still quite a bit of scatter away from the line.



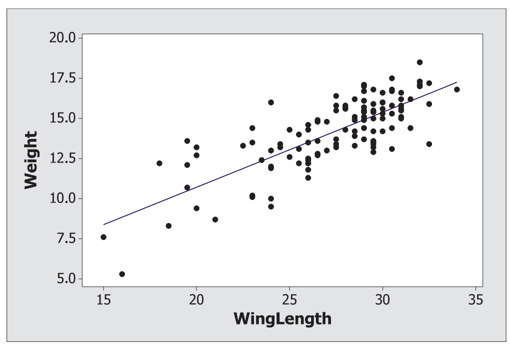
1.22 a. The prediction equation is $\widehat{SalePrice} = -144.8 + 0.943 ListPrice$, so when $ListPrice = 99,500$, the predicted $SalePrice$ is $\widehat{SalePrice} = -144.8 + 0.943(99,500) = 93,683.7$ dollars.

b. For the house at 1317 Prince St, $\widehat{SalePrice} = 93,683.7$, so the residual is $95,000 - 93,683.7 = 1316.3$ dollars.

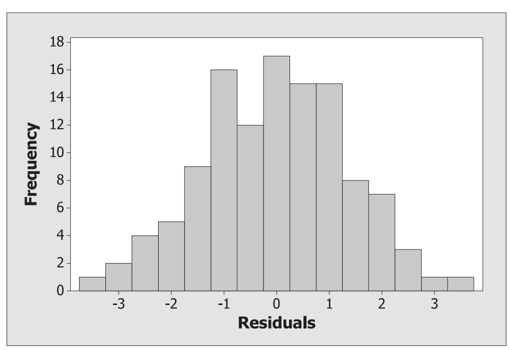
c. The relationship between list price and sales price is very strong and linear for the sample of houses from Grinnell, Iowa.



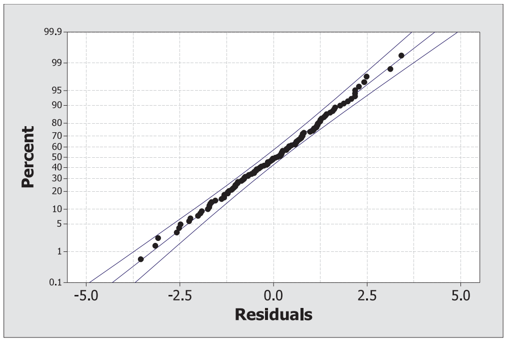
1.23 a. The scatterplot with the least squares line illustrates a very good fit and does not suggest any outliers or influential points.



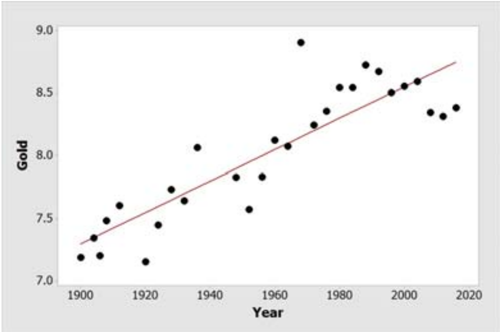
b. A histogram of the residuals shows a nice bell-shaped pattern centered at zero. Thus, the histogram does not reveal any problems with the conditions for this simple linear model.



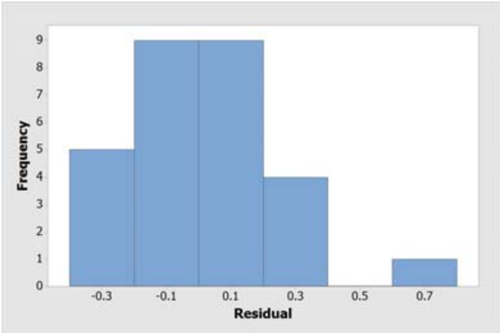
c. A normal probability plot shows a clear linear pattern. Thus, the residuals appear to follow a normal distribution.



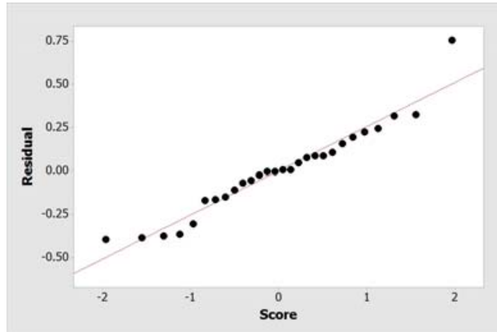
1.24 a. The scatterplot with the least squares line illustrates a good fit. There is one point that is higher than expected in 1968.



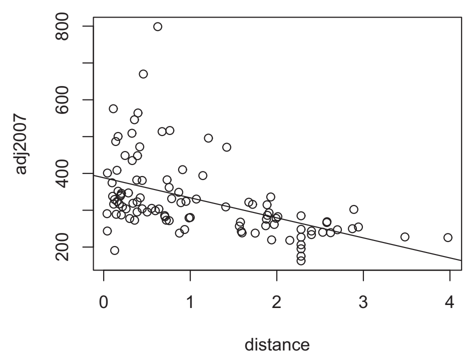
b. A histogram of the residuals shows a mostly nice bell-shaped pattern centered at zero. There is one larger residual that might cause one to worry.



c. A normal probability plot shows a clear linear pattern. Once again there is one residual that is larger and does not fit the pattern.



1.25 a. The scatterplot shows a weak downward trend; homes farther away from the bike trail tend to sell for less. The scatter about the trend line is great for homes near the trail and much smaller for homes far away from the trail.



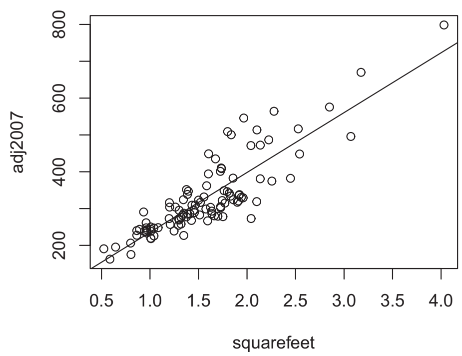
b. The equation of the best-fit line is $\widehat{Adj2007} = 388.204 - 54.427 Distance$. Each mile farther from a trail reduces, on average, the selling price by about 54,000 dollars.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  388.204     14.052  27.626  < 2e-16 ***
Distance     -54.427      9.659  -5.635 1.56e-07 ***
---
Residual standard error: 92.13 on 102 degrees of freedom
Multiple R-squared: 0.2374,Adjusted R-squared: 0.2299
F-statistic: 31.75 on 1 and 102 DF,  p-value: 1.562e-07
```

c. The regression standard error is 92.13. If model conditions are met, then the average deviation from the line is about 92,000 dollars. Such a simple interpretation is compromised here because of the lack of consistent scatter about the line.

d. The model conditions are violated here because of the lack of consistent scatter about the line, as mentioned in part (a).

1.26 a. The scatterplot shows a fairly strong, positive, linear trend between *SquareFeet* and *Adj2007*.



b. The equation of the simple linear regression line is:

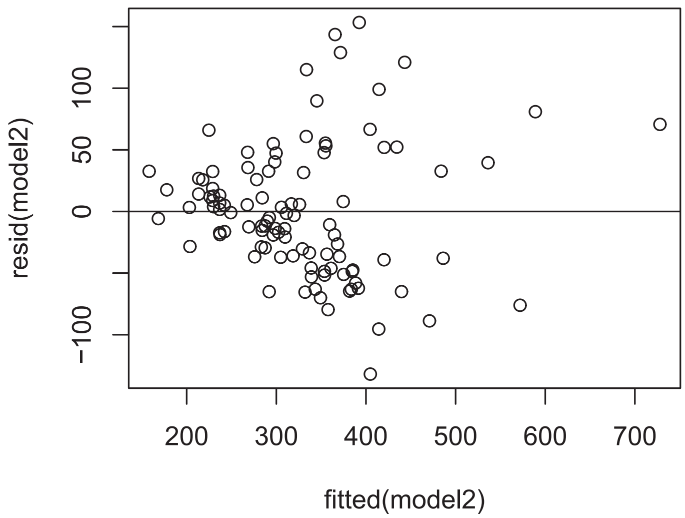
$\hat{Adj2007} = 72.973 + 162.526 SquareFeet.$

Each additional thousand squarefeet of floorspace is associated with an approximate added \$162,000 in selling price.

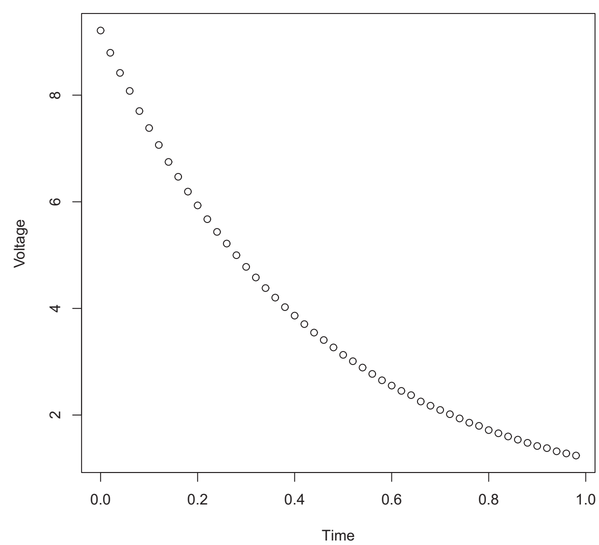
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   72.973     15.541   4.695 8.32e-06 ***
squarefeet   162.526      9.351  17.381  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 53 on 102 degrees of freedom
Multiple R-squared: 0.7476,Adjusted R-squared: 0.7451
F-statistic: 302.1 on 1 and 102 DF,  p-value: < 2.2e-16
```

c. The regression standard error for this model is 53,000; on average, the line predicts selling price to within about 53,000 dollars of reality.

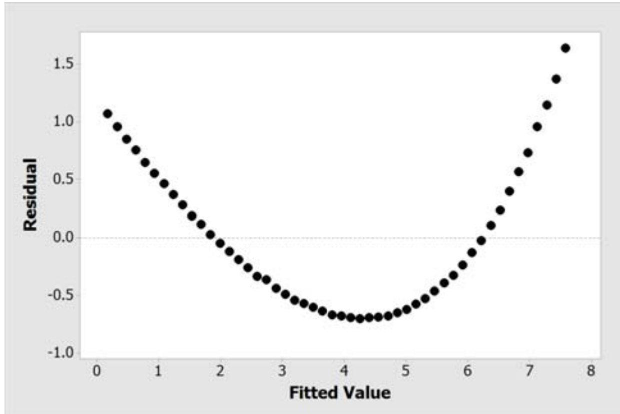
d. There is a slight nonconstancy of variance, as evidenced by the residual-versus-fit plot; larger homes are associated with larger residuals from the line.



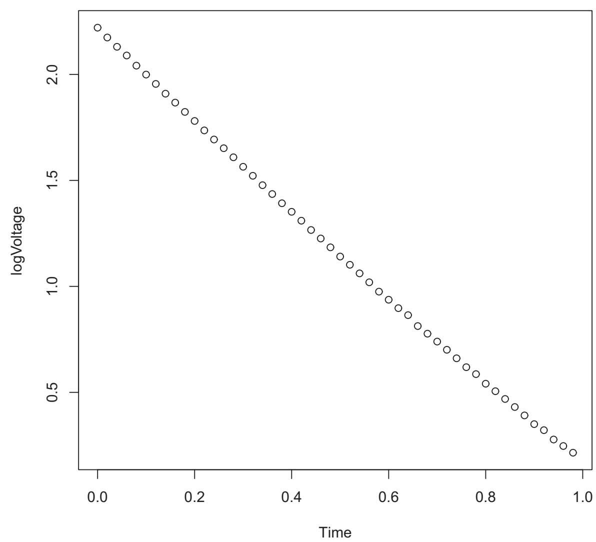
1.27 a. The scatterplot shows that as *Time* increases, *Voltage* goes down sharply. However, the decrease shows a nonlinear (curved) pattern.



b. The residual versus fits plot shows a clear curved pattern.



c. After creating a new variable, $logVoltage$, the scatterplot with $Time$ (below) is much more linear.

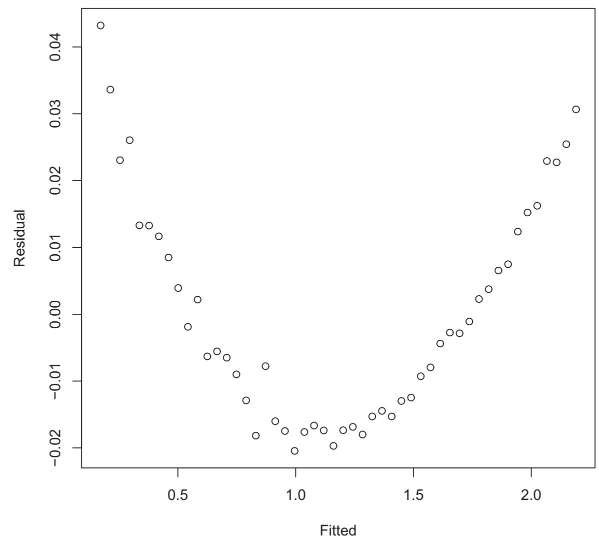


d. Fitting the regression line with technology gives the output

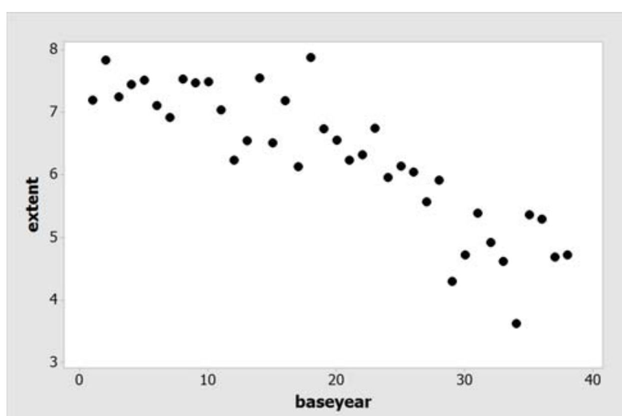
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.189945   0.004637   472.3   <2e-16 ***
Time        -2.059065   0.008154  -252.5   <2e-16 ***
```

This yields the prediction equation $\widehat{logVoltage} = 2.19 - 2.059Time$.

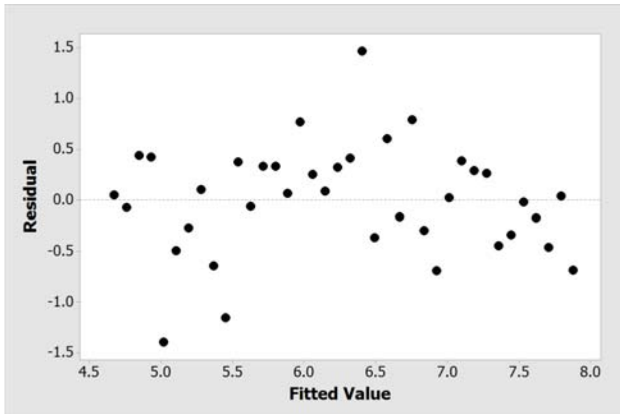
e. The plot of residuals versus fitted values for the model to predict $logVoltage$ shows a striking curved pattern in the residuals. The original (transformed) data have a mostly linear relationship, but some curvature remains after the dominant linear trend is removed. Using the regression model will give predictions that are too high in the middle and too low at the extremes of the $Time$ range.



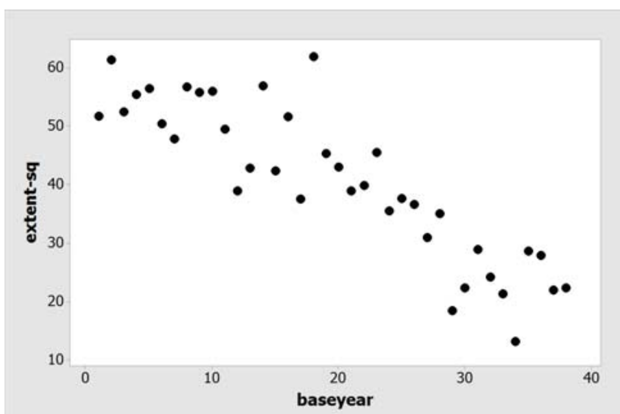
1.28 a. The scatterplot is below. There is clearly a reasonably strong, negative trend to this data. As the years go on, in general, the *Extent* of the sea ice is decreasing. This trend is not, however, linear. There is curvature to it that suggests that as time goes on the amount of decrease is increasing.



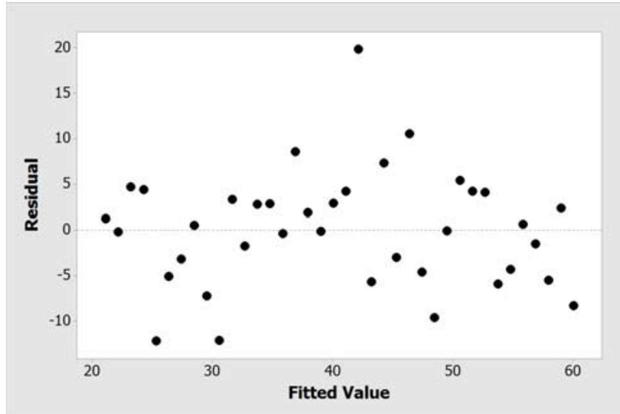
b. The residuals versus fits graph which follows also shows the curvature. In fact, it is somewhat easier to see in this plot. Given this amount of curvature, we should not fit a linear model to this data.



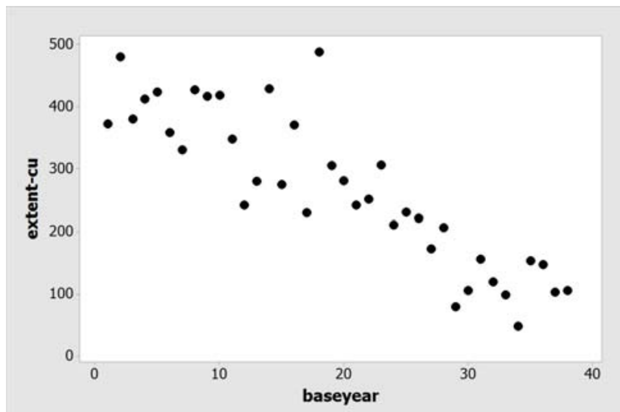
c. The scatterplot is below. While there is still some curvature, it is much less than in the scatterplot from part (a).

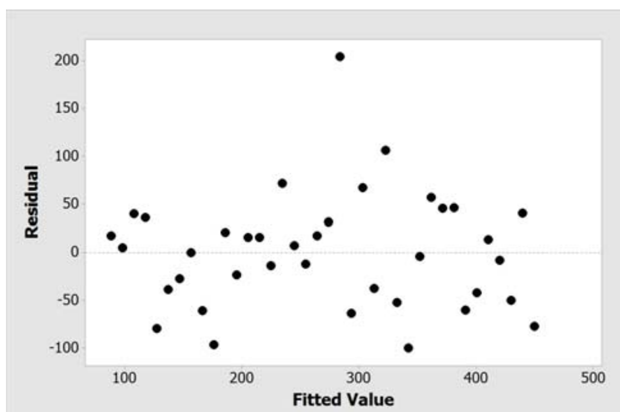


d. The residuals versus fits graph (given below) also shows that there is less curvature in this relationship.



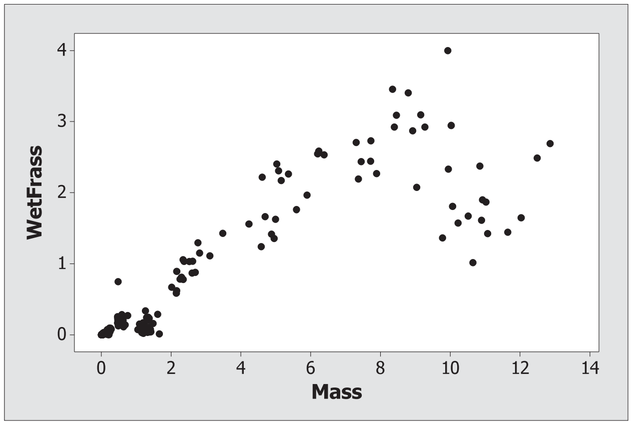
e. The scatterplot and residual plot are given below. In this case there seems to be a decent linear relationship. Very little curvature is evident in either plot.



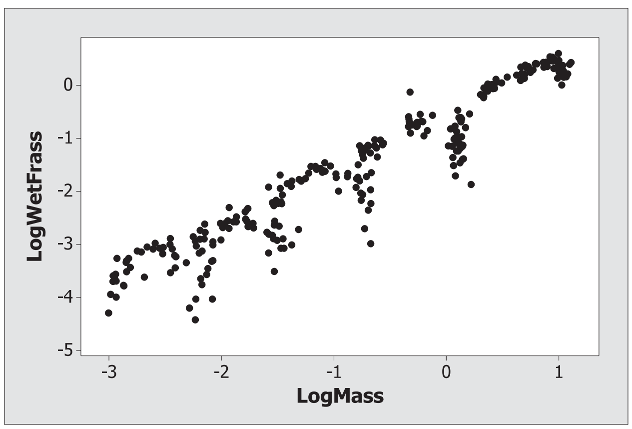


f. The most linear model is the one with the cube of *Extent*. The data is sufficiently linear that this model would be appropriate.

1.29 a. The scatterplot of $WetFrass$ versus $Mass$ shows clear curvature with more variability in the amount of wet frass for the larger caterpillars.



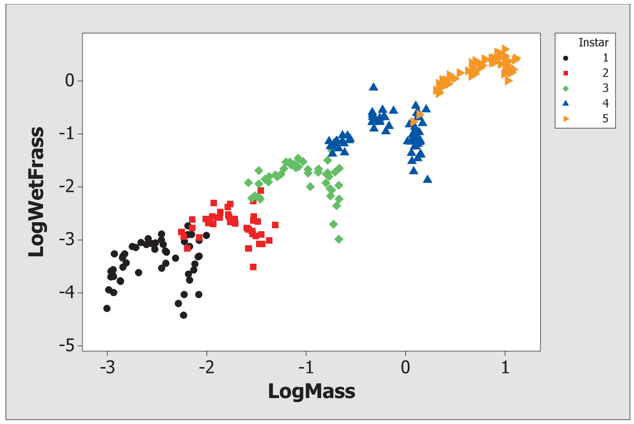
b. The scatterplot of $LogWetFrass$ versus $LogMass$ shows a strong positive association between the transformed variables, with intermittent periods of increased variability.



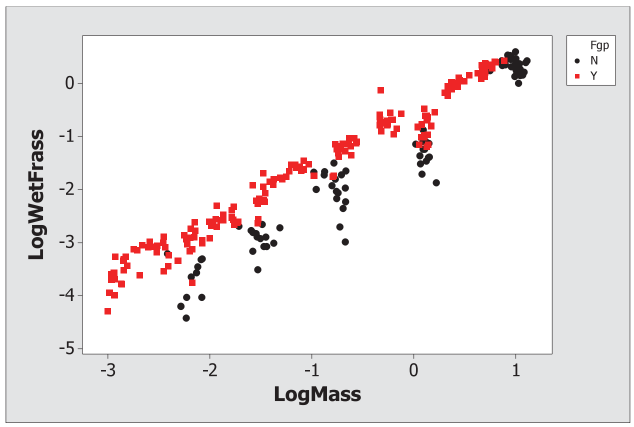
c. The log transformed variables show a more linear pattern. The fitted regression line for these variables is

$$Log\widehat{WetFrass} = -0.739 + 1.054LogMass$$

d. Here is a plot for the relationship with different symbols/colors for the five $Instar$ groups. There is curvature within the $Instars$, especially for the larger caterpillars in each group, but the linear model provides a good summary of the overall pattern for each $Instar$.



e. Here is a plot for the relationship with different symbols/colors for the free-growth and no-free-growth periods. Yes, the overall pattern is definitely more linear when the caterpillars are in a free-growth period.



1.30 a. The scatterplot of $Nassim$ versus $Mass$ shows clear curvature (perhaps quadratic) with more variability in nitrogen assimilation for the larger caterpillars.