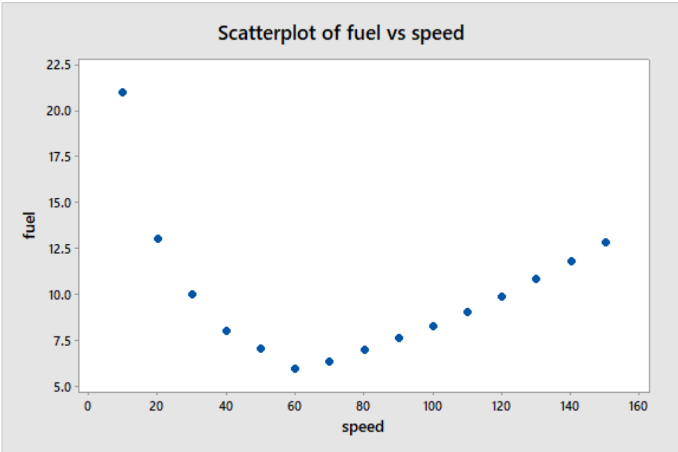
Supplementary exercises 2.12, 2.27, 2.28 and 2.60 of IPS7e

Data: Measurements of a car's (a British Ford Escort) speed and fuel. The fuel consumption is a response variable, whereas the speed is a controlled variable (set at 10 to 150 km/h in steps of 10 km/h) and is therefore an explanatory variable.

Exercise 2.12

(a) We should put the speed on the x-axis because it is an explanatory variable.



(b) The curve is composed by a decreasing part and an increasing part, and therefore not linear. The fuel consumption is minimal for a speed around 60 km/h. The increase in fuel when speeding more than 60 km/h is almost linear, but the increase when driving slower than 60 km/h is quite curved. The relationship probably makes sense because car engines were constructed to perform optimally at a certain speed.

(c) The association is positive above 60 km/h and negative below.

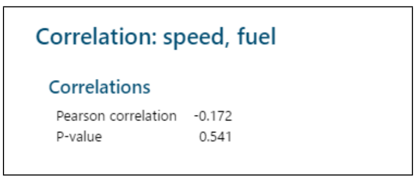
(d) The association seems quite strong because the points seem to follow a very regular pattern (which is just not linear).

Exercise 2.27

Before carrying out the requested computation, a note about the interpretation of a correlation when one variable is an explanatory variable. Strictly speaking, the correlation only makes sense for a pair of response variables (because the explanatory variable, in this case the speed, does not have a sampling distribution). For one explanatory and one response variable it is more appropriate to compute the linear regression equation. The correlation has strong links with that equation and may therefore be interpreted indirectly from the linear regression equation. The P-value is the same for correlation and linear regression, and therefore also valid in this sense.

For the sake of this exercise, we ignore the problem with interpreting the correlation when one variable is explanatory (controlled), but in a real example this use of the correlation is not recommended (though not uncommon). At the very least, one would need to be careful in wording of conclusions about a correlation coefficient determined from a controlled variable.

```
MTB > Correlation 'speed' 'fuel'.
```



Comments:
The correlation is negative and quite close to zero, the latter because there is no strong linear association (as discussed, the association is clearly non-linear). In this case, the correlation does not give an adequate summary of the strength and direction of the association.

Exercise 2.28

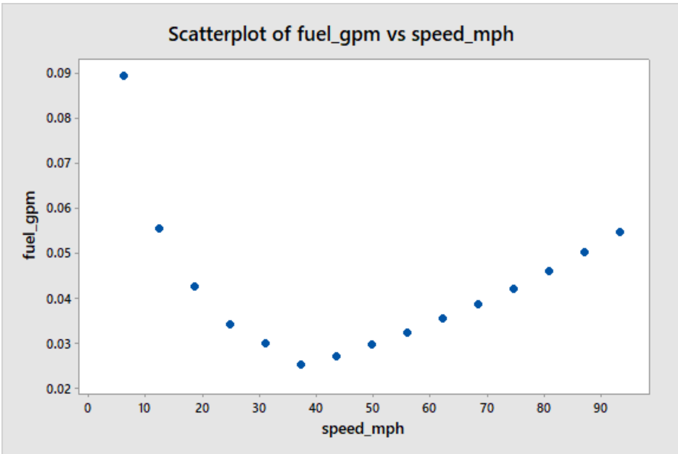
(a) To measure the speed in miles per hour, divide the values by 1.609. To measure the fuel consumption in gallons per km, divide the values by $3.785 \cdot 100 = 378.5$. To measure the fuel consumption in gallons per mile, multiply the resulting values by 1.609, i.e. multiply the original fuel consumption values by 1.609/378.5. The pattern in the graph is unchanged (only the axes have changed), and the correlation is unchanged from above.

```
MTB > Name C3 'speed_mph'
MTB > Let 'speed_mph' = 'speed'/1.609
MTB > Name C4 'fuel_gpm'
MTB > Let 'fuel_gpm' = 'fuel'*1.609/378.5
MTB > Correlation  'speed_mph' 'fuel_gpm'.
```

Correlation: speed_mph, fuel_gpm

Correlations

Pearson correlation	-0.172
P-value	0.541



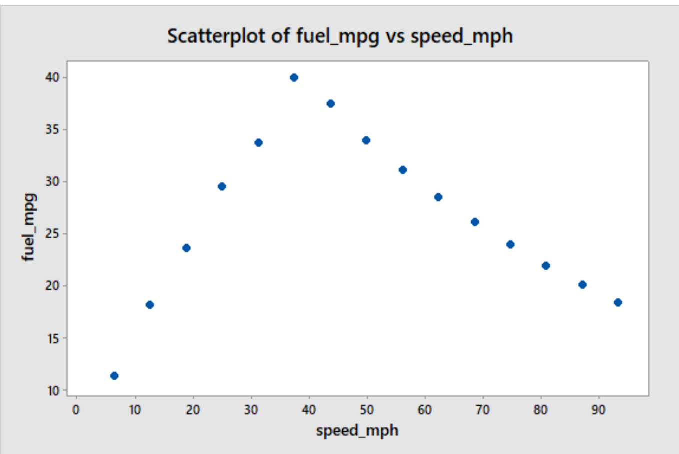
(b) We continue with the requested calculation in Minitab.

```
MTB > Name C5 'fuel_mpg'
MTB > Let 'fuel_mpg' = 1/'fuel_gpm'
MTB > Correlation  'speed_mph' 'fuel_mpg'.
```

Correlation: speed_mph, fuel_mpg

Correlations

Pearson correlation	-0.043
P-value	0.879



Comments:
The correlation is now -0.043, and has changed from above (there is no easy formula to know what it would be without actually doing the transformation and computing the value directly). However, the correlation still does not give an adequate impression of the strength and direction of the association.

Exercise 2.60

We will, as requested, fit a linear regression model (for demonstration purposes, because we have already demonstrated that it is not an appropriate model).

```
MTB > Name c6 "RESI1"
MTB > Fitline 'fuel' 'speed';
SUBC>  GVars 'speed';
SUBC>   RType 1;
SUBC>   Confidence 95.0;
SUBC>   Resid  'RESI1'.
```

Regression Analysis: fuel versus speed

The regression equation is
fuel = 11.06 - 0.01466 speed

Model Summary

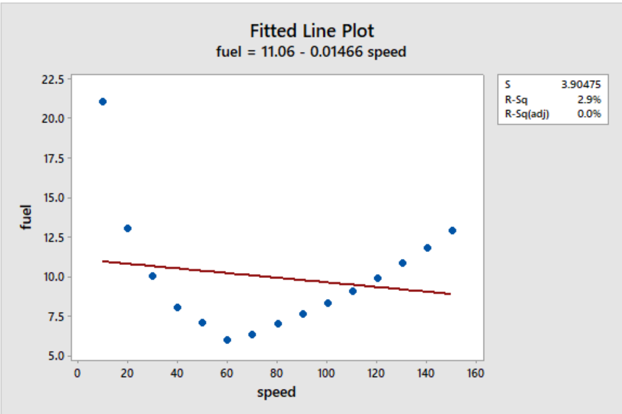
S	R-sq	R-sq(adj)
3.90475	2.95%	0.00%

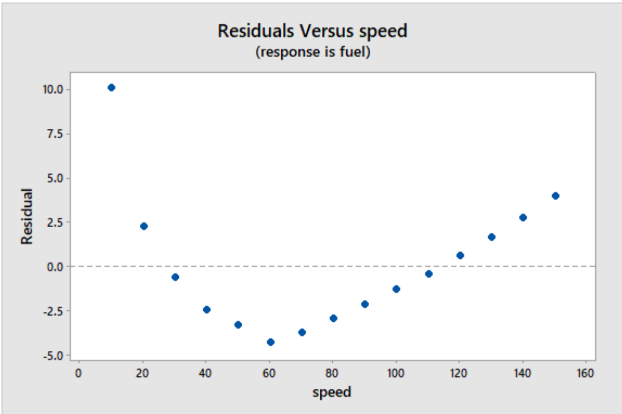
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	6.015	6.0153	0.39	0.541
Error	13	198.211	15.2470		
Total	14	204.227			

Fitted Line: fuel versus speed

Residuals from fuel vs speed





(a) We used the Fitted Line Plot menu to plot the observations with the regression line overlaid.

(b) We stored the residuals in the `RESI1` column and can now use the `Sum` function (available in the `Calc-Column Statistics` menu) to compute the sum for the column.

```
Sum 'RESI1'.
```

Sum of RESI1

Sum of RESI1 = -3.01981E-14

It is seen that the sum is essentially zero (up to the numerical precision of the software).

(c) The residual plot (requested in the `Graphs` submenu of the Fitted Line Plot menu) shows a similar pattern. Note that when the residuals are plotted against the fitted values, the pattern is reversed along the x-axis from the original data, due to the estimated negative slope.