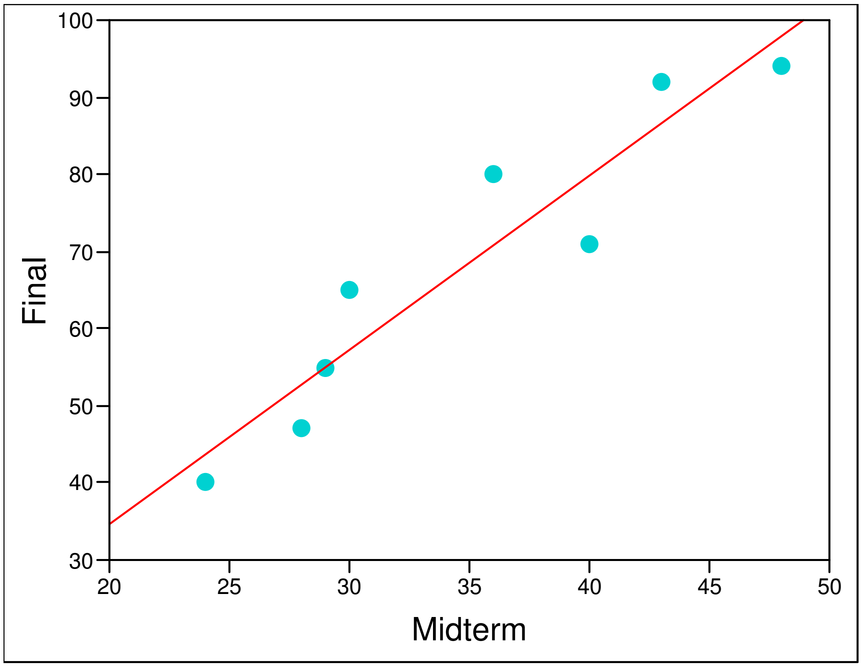
Unit 2 Assignment Solutions

1. (a)



There appears to be a strong positive linear relationship between Midterm Score and Final Exam Score.

(b)

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
43	92	8.25	24	198.00
28	47	-6.75	-21	141.75
36	80	1.25	12	15.00
30	65	-4.75	-3	14.25
24	40	-10.75	-28	301.00
48	94	13.25	26	344.50
29	55	-5.75	-13	74.75
40	71	5.25	3	15.75
				sum = 1105

$$r = \frac{1}{(n-1)s_x s_y}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{7(8.36)(20.03)}(1105) = 0.9427$$

(c) $b_1 = r\dfrac{s_y}{s_x} = (0.9427)\left(\dfrac{20.03}{8.36}\right) = 2.259$ $\quad b_0 = \bar{y} - b_1\bar{x} = 68.00 - (2.259)(34.75) = -10.5$

The equation of the least squares regression line is thus $\hat{y} = -10.5 + 2.259\,x$. The line is shown on the scatterplot above.

(d) The slope is $b_1 = 2.259$. This means that, for every additional mark a student receives on their midterm, we predict the Final Grade percentage to increase by 2.259.

(e) $r^2 = (0.9427)^2 = 0.8887$. This means that 88.87% of the variation in a student's Final Exam Score can be explained by his or her Midterm Score.

(f) The predicted Final Exam Score for a student with a Midterm Score of 15 is

$$\hat{y} = -10.5 + 2.259(15) = 23.385.$$

The predicted Final Exam Score for a student with a Midterm Score of 32 is

$$\hat{y} = -10.5 + 2.259(32) = 61.788.$$

The prediction for a midterm score of 32 is more reliable. This is because 15 is outside our range of data for X, and so by making this prediction, we are extrapolating. We know there is a strong linear relationship within our range of data, but we have no idea what happens to this relationship outside our data range.

(g) The residual for the second student is

$$y_i - \hat{y} = 47 - (-10.5 + 2.259(28)) = 47 - 52.752 = -5.752\,.$$

The negative sign of the residual tells us that this point falls below the least squares regression line, i.e. the Final Exam Score for this student is lower than we would predict it to be based on his or her Midterm Score.

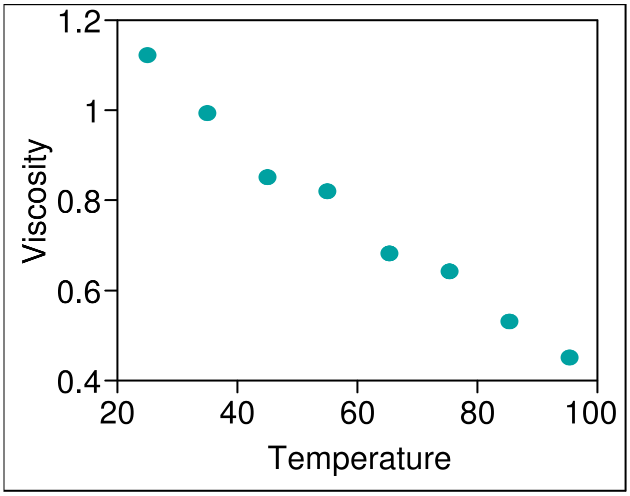
(h) The correlation does not depend on the units of measurement. Therefore, if Midterm Score had been measured as a percentage, we would still have $r = 0.9427$.

The slope gives us the predicted increase (or decrease) in Y when X increases by one unit. We know that, when Midterm Score increases by one mark, we predict the Final Exam Score to increase by 2.259%. A 1-mark increase corresponds to a 2% increase in Midterm Score. So we know that when Midterm Score increases by 2%, the predicted Final Exam Score increases by 2.259%. Therefore, if the Midterm Score increases by 1%, we predict the Final Exam Score will increase by 2.259/2 = 1.1295. This would be the value of the slope of the least squares regression line if we measured the Midterm Score as a percentage.

The intercept is the predicted Final Exam Score when the Midterm Score is zero. A Midterm Score of zero is the same, whether we measure it out of 50 or out of 100, so the intercept does not change, $b_0 = -10.5$. Therefore, if the Midterm Score had been measured as a percentage, the equation of the least squares regression line would be

$$\hat{y} = -10.5 + 1.1295\,x..$$

2. (a)



There appears to be a strong negative linear relationship between Temperature and Viscosity.

(b)

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
24.9	1.12	-35.2	0.36	-12.672
35.0	0.99	-25.1	0.23	-5.773
44.9	0.85	-15.2	0.09	-1.368
55.1	0.82	-5	0.06	-0.300
65.2	0.68	5.1	-0.08	-0.408
75.2	0.64	15.1	-0.12	-1.812
85.2	0.53	25.1	-0.23	-5.773
95.3	0.45	35.2	-0.31	-10.912
				sum = – 39.018

$$r = \frac{1}{(n-1)s_x s_y}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{7(24.635)(0.228)}(-39.018) = -0.9924$$

(c) $b_1 = r\frac{s_y}{s_x} = (-0.9924)\left(\frac{0.228}{24.635}\right) = -0.009185$ $\quad b_0 = \bar{y} - b_1\bar{x} = 0.76 - (-0.009185)(60.1) = 1.312$

The equation of the least squares regression line is thus $\hat{y} = 1.312 - 0.009185\,x$.

(d) The slope, $b_1 = -0.009185$. This means that, for every one-degree Celsius increase in the Temperature, we predict the Viscosity to decrease by 0.009185 mPa·s.

(e) $r^2 = (-0.9924)^2 = 0.9849$. This means that 98.49% of the variation in Viscosity can be explained by its regression on Temperature.

(f) When the Temperature is 15°C, we predict the Viscosity to be

$$\hat{y} = 1.312 - 0.009185(15) = 1.174225 \text{ mPa·s.}$$

When the Temperature is 60°C, we predict the Viscosity to be

$$\hat{y} = 1.312 - 0.009185(60) = 0.7609 \text{ mPa·s.}$$

The prediction for 60°C is more reliable than that for 15°C. This is because 15 is outside our range of data for X, and so by making this prediction, we are extrapolating. We know there is a strong linear relationship without our range of data, but we have no idea what happens to this relationship outside our data range.

(g) The residual for the third observation is

$$y_i - \hat{y} = 0.85 - (1.312 - 0.009185(44.9)) = 0.85 - 0.8996 = -0.0496 .$$

The negative sign of the residual tells us that this point falls below the least squares regression line, i.e. the Viscosity for this observation is lower than we would predict it to be based on its Temperature.

(h) The correlation does not depend on the units of measurement. Therefore, if Temperature had been measured in degrees Fahrenheit, we would still have r = – 0.9924.

The slope gives us the predicted increase (or decrease) in Y when X increases by one unit. We know that, when Temperature increases by 1°C, we predict the Viscosity to decrease by 0.009185 mPa·s. A 1°C increase corresponds to a 1.8°F increase. So we know that when Temperature increases by 1.8°F, the predicted Viscosity decreases by 0.009185 mPa·s. Therefore, if the Temperature increases by 1°F, we predict the Viscosity will decrease by 0.009185/1.8 = 0.005102. This would be the value of the slope of the least squares regression line if we measured Temperature in degrees Fahrenheit.

The intercept of our original least squares regression line is 1.312. This is the predicted Viscosity when the Temperature is 0°C. But we now want the predicted Viscosity when the Temperature is 0°F. A Temperature of 0°F corresponds to a Temperature of – 17.78°C. Using the original least squares regression line, the predicted Viscosity for this Temperature is

$$\hat{y} = 1.312 - 0.009185(-17.78) = 1.4753$$

This would be the value of the intercept of the least squares regression line if Temperature had been measured in degrees Fahrenheit. (Note that, as a prediction, this value would not be trustworthy, as this Temperature falls far outside our range of Temperatures. However, we still need to do this to find the intercept). Therefore, if the Temperature had been measured in °F, the equation of the least squares regression line would be

$$\hat{y} = 1.4753 - 0.005102\,x .$$

3. (a) The equation of the least squares regression line is $\hat{y} = b_0 + b_1 x$. When $x = \bar{x}$, the predicted value of y is

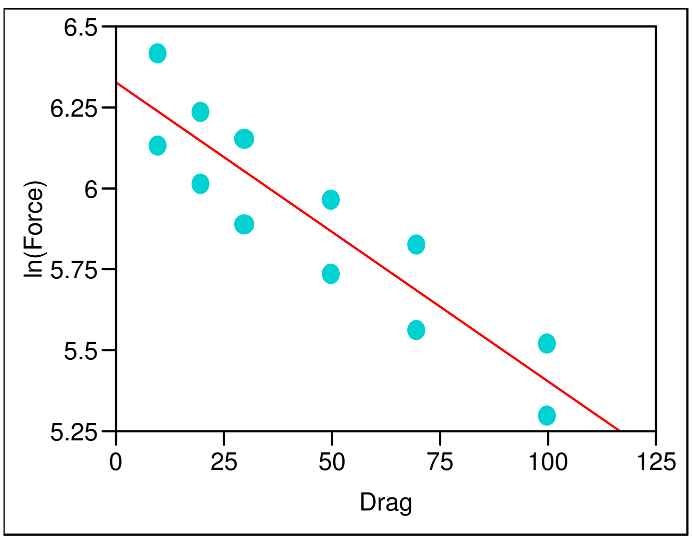
$$\hat{y} = b_0 + b_1\bar{x} = (\bar{y} - b_1\bar{x}) + b_1\bar{x} = \bar{y}.$$

Therefore, we have shown that the least squares regression line must pass through the point $(\bar{x}, \bar{y})$.

(b) The sum of residuals is equal to

$$\begin{aligned}\sum_{i=1}^{n}(y_i - \hat{y}) &= \sum_{i=1}^{n}\left(y_i - (b_0 + b_1 x_i)\right) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} b_0 - \sum_{i=1}^{n} b_1 x_i \\ &= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}(\bar{y} - b_1\bar{x}) - \sum_{i=1}^{n} b_1 x_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\bar{y} + b_1\sum_{i=1}^{n}\bar{x} - b_1\sum_{i=1}^{n} x_i \\ &= n\bar{y} - n\bar{y} + b_1 n\bar{x} - b_1 n\bar{x} = 0.\end{aligned}$$

4. (a)



The relationship appears to be more linear.

(b) $b_1 = r^* \dfrac{s_{y^*}}{s_x} = (-0.913)\left(\dfrac{0.325}{32.287}\right) = -0.0092$ $\qquad b_0^* = \bar{y}^* - b_1\bar{x} = 5.896 - (-0.0092)(46.667) = 6.325$

The equation of the least squares regression line is thus $\hat{y}^* = 6.325 - 0.0092x$. The line is shown on the scatterplot above.

(c) We know that the predicted value of ln(Force) when Drag = 60 is

$$\hat{y}^* = 6.325 - 0.0092(60) = 5.773,$$

and so the predicted Drag is $e^{5.773} = 321.5$ pounds.