Introductory Statistics Tutorial Answers
Chapter 12 – Relationships between Quantitative Variables: Regression and Correlation

Section A: The Straight Line Graph

1. **(a)** $\beta_0 = 5,\ \beta_1 = 3$
(b) $\beta_0 = 10,\ \beta_1 = -14$

2. **(a)** $y = -3 + 2x$
(b) $y = 7 - 4x$

3. **(a)** 6
(b) 12.5
(c) 25
(d) 7.5

Section B: Regression

1. **(a)** $\hat{y} = 11.238 + 1.309x$

(b) Predicted lung capacity = 11.238 + 1.309 x 30 = 50.5

(c) Predicted lung capacity = 11.238 + 1.309 x 25 = 44.0
Residual = Observed value – predicted value = 55 – 44.0 = 11

(d) **'**Years smoking' is used to predict lung capacity.
'Years smoking' is a quantitative variable and 'Lung capacity' is continuous and random.
There is a possible linear trend but the observations (28, 30) and (33, 35) are possible outliers which cause concern with the appropriateness of the model.
The residuals versus 'Years smoking' plot along with the *P-value* for the *W*-test for Normality indicates some concern with the assumption that the errors are Normally distributed.

(e) H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$
P-value = 0.0086
There is strong evidence that an increase in years of smoking is associated with an increase in lung capacity (i.e. an effect of years of smoking on lung capacity).
With 95% confidence, we estimate that for every additional year of smoking an emphysema patient's lung capacity increases by between 0.44 and 2.18 units.

(f) **(i)** $r = 0.774$
(ii) *Excel* calls it Multiple R.

2. **(a)** For $x = 1.46$, $\hat{y}$ = -29.86 + 37.72 x 1.46 = 25.2
Residual = Observed value – predicted value = 11.6 – 25.2 = –13.6

(b) The lactic acid concentration is used to predict the taste score.
The lactic acid concentration is quantitative, and the taste score is continuous and random.
The scatter plot shows a linear trend with scatter about that trend.
From the plot of residuals versus lactic acid concentration there is no concern with the assumption that the errors are Normally distributed with mean 0 and with the same standard deviation for each value of *X*.

(c) H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$
P-value = 0.000
There is strong evidence that an increase in lactic acid concentration is associated with an increase in taste score (i.e. an effect of lactic acid concentration on taste score).
95% confidence interval for β_1 is:
37.720 ± 2.048 x 7.186 = (23.0, 52.4)
With 95% confidence, we estimate that for every increase of one unit in the lactic acid concentration the taste score increases by between 23.0 and 52.4 units.

(d) **(i)** We predict that, on average, cheddar cheese with a lactic acid concentration of 1.8 will have a taste score of 38.04.
(ii) With 95% confidence, we estimate that the mean taste score for cheddar cheese with a lactic acid concentration of 1.8 will be somewhere between 31.2 and 44.9.
(iii) With 95% confidence, we predict that the taste score of the next piece of cheddar cheese with a lactic acid concentration of 1.8 will be somewhere between 13.0 and 63.1.

(e) Estimated slope = 37.72
Estimated increase in taste score for a 1 unit change in lactic acid concentration is 37.72.
Estimated increase in taste score for a 0.05 unit change in lactic acid concentration is 0.05 x 37.72 = 1.886.
(1) is the correct response.

(f) (2)

Section C: Old Exam Questions

1. (3)
2. (5)
3. (2)
4. (3)
5. (2)