



## 6.6

# Correlation & linear regression

## Questions

1. Consider the data set and its corresponding scatter diagram as output from a GC:

| x     | y     | L3 |
|-------|-------|----|
| -1.23 | 5.88  |    |
| 0.0   | -0.89 |    |
| 1.52  |       |    |
| 1.98  | -1.44 |    |
| 3.12  | 5.21  |    |
| 4.23  | 11.02 |    |

  

- (a) Give a reason why it might be unwise to use either of the regression lines  $y$  on  $x$  or  $x$  on  $y$  to estimate the value of  $y$  when  $x = 1.5$ .
- (b) The equation of the regression line of  $y$  on  $x$  is  $y = 1.069x + 1.083$ . Find the observed value of  $y$  when  $x = 1.52$ , and hence deduce the linear (product moment) correlation coefficient,  $r$ , between  $x$  and  $y$ . Adding more points  $(x, y)$  to the data may change the value of  $r$ . Give an example of a pair of points that would not cause such a change.
- (c) Find the equation of the regression line of  $(x - 1.108)^2$  on  $y$ . Use this equation to estimate the value(s) of  $x$  when  $y = 3$ .

Can we use the same equation to estimate the value of when  $x = 3$ ?



2. (a) The marks of two short quizzes in Mathematics,  $x$  and  $y$ , of 10 students are shown in

| Student                                | 1 | 2 | 3 | 4  | 5 | 6 | 7  | 8 | 9 | 10 |
|--|---|---|---|----|---|---|----|---|---|----|
| Marks for 1 <sup>st</sup> quiz ( $x$ ) | 6 | 5 | 8 | 8  | 9 | 6 | 10 | 4 | 9 | 6  |
| Marks for 2 <sup>nd</sup> quiz ( $y$ ) | 8 | 7 | 8 | 10 | 1 | 8 | 10 | 5 | 8 | 7  |

- (i) Find the product-moment correlation coefficient between  $x$  and  $y$  and comment on the relationship between  $x$  and  $y$ .

- (ii) Plot the data on a scatter diagram of  $y$  against  $x$ .
- (iii) By using (ii), state with a reason whether or not your interpretation in (i) should be amended. Justify your answer.

- (b) Drums of hair shampoo are kept in storage for a number of weeks before being rebottled for retail sale.

To investigate the relationship between the number of weeks ( $x$ ),  $0 \leq x \leq 15$ , that each drum is kept in storage and the amount of water content loss ( $y$  ml) due to evaporation, 6 drums are examined and the following results are obtained:

$$\Sigma x = 48, \Sigma x^2 = 472, \Sigma y = 408, \Sigma y^2 = 29584, \Sigma xy = 3664$$

- (i) Find the least squares line of regression of  $y$  on  $x$ .
- (ii) Estimate to the nearest integer, the water content loss due to evaporation for a drum kept in storage for eight weeks.
- (iii) Explain why you would not expect to get good estimates for evaporation loss from the line of regression when the storage time is more than a year.



3. A PE teacher, Miss Melon, wants to determine if there is any correlation between the weight of and the mean amount of money spent on canteen food per month by an 18-year-old college student. A group of eight 18-year-old college students was interviewed and had their weights taken. The results are tabulated as follows:

|   |      |      |      |      |      |     |      |      |
|---|------|------|------|------|------|-----|------|------|
| Weight, $x$ / kg                              | 34   | 57   | 43   | 65   | 72   | 46  | 51   | 54   |
| Mean amount spent on food per month, $y$ / \$ | 21.6 | 58.1 | 32.8 | 59.3 | 60.3 | $k$ | 56.8 | 49.5 |

- (i) Given that the equation of the least squares regression line of  $y$  on  $x$  is  $y = 1.0031x - 4.3018$ , find the value of  $k$ , leaving your answer to 1 decimal place.
- (ii) Calculate the value of the linear product moment correlation coefficient for these 8 college students.
- (iii) The following models for the above data are suggested for  $b > 0$ :

(A)  $y = a + bx$ ; (B)  $y = a + bx^2$ ;

(B)  $y = a + b \ln x$ ; (D)  $y = a + \frac{b}{x}$ .

State, with a reason, which model is most appropriate.

- (iv) With the choice of the model indicated in (iii), calculate the least squares estimates of  $a$  and  $b$  and calculate the value of the linear product moment correlation coefficient for the transformed data.
- (v) Suppose we have an 18-year-old college student who spends an average of \$80 per month on food, determine the weight of this student using the least squares regression line obtained in (iv). Comment on the suitability of your answer.

5. The table below gives the observed values of bivariate  $x$  and  $y$ .

|     |    |    |     |    |    |    |    |
|-----|----|----|-----|----|----|----|----|
| $x$ | 20 | 30 | 34  | 35 | 36 | 40 | 42 |
| $y$ | 32 | 25 | $a$ | 22 | 26 | 18 | 19 |

It is given that the equation of the regression line  $y$  on  $x$  is  $y = 43.5 - 0.602x$ .

- (i) Find the value of  $a$  correct to the nearest integer.
- (ii) Write down the equation of the regression line  $x$  on  $y$  and the value of product moment correlation coefficient between  $x$  and  $y$ .

4. Measurements of the relative humidity and moisture content of samples of a certain type of raw material on 10 days yielded the following results:

|                        |    |    |    |    |    |    |    |    |    |    |
|------------------------|----|----|----|----|----|----|----|----|----|----|
| Relative humidity (x%) | 46 | 53 | 37 | 42 | 34 | 29 | 60 | 44 | 41 | 48 |
| Moisture content (y%)  | 12 | 14 | 11 | 13 | 10 | 8  | 17 | 12 | 10 | 15 |

- (i) Draw a scatter diagram for the data. Calculate the product-moment correlation coefficient. Explain what they reveal about the relationship between relative humidity and moisture content.
- (ii) It is required to estimate the moisture content given that the relative humidity is 58%. Find the equation of a suitable least squares regression line. Use your equation to obtain the required estimate.
- (iii) Use your equation in (ii) to estimate the moisture content when relative humidity is 10%. Comment on the reliability of the estimation.
- (iv) It was found out that the moisture content of the raw material was measured wrongly due to a fault in the measuring instrument. The actual moisture content is 10% more than the measured values. Will this affect the product-moment correlation coefficient? Explain your answer.

6. The following are the summary of 6 pairs of values of the variable of  $x$  and  $y$ :

$\Sigma x = 78$ ,

$\Sigma y = 234$ ,

$\Sigma x^2 = 1462$  and  $\Sigma y^2 = 9274$ .

It is given further that when  $x = 9$  the estimated value of  $y$  is 37.

Find the equation of the regression line  $x$  on  $y$ .

7. It is believed that the probability  $p$  of a randomly chosen pregnant woman giving birth to a Down's Syndrome child is related to the women's age  $x$ , in years. The table gives the observed values of  $p$  for 5 different values of  $x$ .

|     |         |         |         |         |         |
|-----|---------|---------|---------|---------|---------|
| $x$ | 25      | 30      | 35      | 40      | 45      |
| $p$ | 0.00067 | 0.00125 | 0.00333 | 0.01000 | 0.03330 |

- (i) Give a sketch of the scatter diagram for the data.

- (ii) State, with a reason, which of the following would be an appropriate model to represent the above data.

(A)  $p = a + \frac{b}{x}$ ,

(B)  $p = a + be^{-x}$ ,

(C)  $\ln p = a + bx$

where  $a$  and  $b$  are constants and  $b > 0$ .

- (iii) For the appropriate model, calculate the values of  $a$  and  $b$ , and find the product-moment correlation coefficient.

- (iv) Obtain an estimate of the probability of a 19 year-old woman giving birth to a Down's Syndrome child. Comment on the reliability of your answer.



8. Explain why it is advisable to plot a scatter diagram before interpreting a correlation coefficient calculated for a sample drawn from a bivariate distribution. Sketch scatter diagrams indicating the following:

- (i) a linear product moment correlation coefficient close to zero, but with an obvious relation between the variables.
- (ii) a linear product moment correlation coefficient close to +1, but without a high degree of correlation between the variables.



9. The yield of a particular crop on a farm depends primarily on the amount of rainfall in the growing season. The values of the yield  $Y$ , in tons per acre, and the rainfall  $X$ , in centimeters, for 7 successive years are given in the table below.

|     |      |      |      |      |      |      |      |
|-----|------|------|------|------|------|------|------|
| $X$ | 12.3 | 13.7 | 14.5 | 11.2 | 13.2 | 14.1 | 12.0 |
| $Y$ | 6.25 | 8.02 | 8.42 | 5.27 | 7.21 | 8.71 | 5.68 |

- (i) Find the linear product moment correlation coefficient between  $X$  and  $Y$ . Hence, comment on the value and its implication on the data.
- (ii) If the values of the yield and rainfall are re-recorded in kilograms per acre and metres respectively, do you expect any change in the value of the linear product moment correlation coefficient and why?



10. For a given set of data for  $X$  and  $Y$ , it was found that the regression line of  $X$  on  $Y$  is  $y = -\frac{9}{8}x + 5.5$  and the linear product moment correlation coefficient  $r$  between  $X$  and  $Y$ , is given by  $r^2 = \frac{16}{25}$ . It is also known that  $(\bar{x}, \bar{y}) = (4, 1)$ . Find the regression line of  $Y$  on  $X$ .



11. A medical statistician wishes to carry out a test to see if there is any correlation between the head circumference and body length of newly-born babies. A random sample of 10 newly-born babies has their head circumference,  $c$  cm, and body length,  $l$  cm measured. This bivariate sample is illustrated in the table below.

|     |    |    |      |    |    |    |    |    |      |      |
|-----|----|----|------|----|----|----|----|----|------|------|
| $c$ | 31 | 32 | 33.5 | 34 | 34 | 51 | 35 | 36 | 36.5 | 37.5 |
| $l$ | 45 | 49 | 49   | 47 | 50 | 34 | 50 | 53 | 51   | 51   |

One particular point has been recorded incorrectly with its values of  $c$  and  $l$  interchanged. Identify this point.

- (i) Make the necessary correction and use a suitable regression line to estimate the length of the baby whose head has the circumference of
  - (a) 34.5 cm,
  - (b) 45 cm.
- (ii) Give a reason why the estimation found in (i) part (b) may not be a good one.



12. An investigation on the number of hours,  $t$  (correct to the nearest hour), spent studying for an examination by 8 students and the marks,  $m$ , they obtained was done. However, a value was missing from the report and this was indicated by  $p$  below.

| Student | A  | B  | C  | D  | E   | F  | G  | H  |
|---------|----|----|----|----|-----|----|----|----|
| $t$     | 5  | 6  | 7  | 10 | 11  | 12 | 13 | 16 |
| $m$     | 68 | 68 | 69 | 72 | $p$ | 72 | 76 | 78 |

The equation of the regression line of  $m$  on  $t$  is  
 $m = 0.94t + 62.6$ .

Find the value of  $p$  and write down the equation of the regression line of  $t$  on  $m$ .

Determine the difference in marks obtained by a student if he had studied for 2 more hours.

Another 3 students were surveyed and the results were as follows.

| Student | A  | B  | C  |
|---------|----|----|----|
| $t$     | 15 | 16 | 17 |
| $m$     | 83 | 87 | 93 |

Give a sketch of the scatter diagram for the data of the 11 students and state, with a reason, which of the following models is most appropriate.

(A)  $m = a + bt^2$

(B)  $m = a + \frac{b}{t}$

(C)  $m = a + blnt$

For the most appropriate model, determine least squares estimates of  $a$  and  $b$ , and state the product moment correlation coefficient for the transformed data.



- 13.** The number of people applying for the course in Biomedical Science in a particular university in each of 6 semesters is given below:

| Semester (x)             | 1  | 2  | 3  | 4   | 5   | 6   |
|--------------------------|----|----|----|-----|-----|-----|
| Number of applicants (y) | 11 | 29 | 68 | 138 | 215 | 560 |

The admissions office believes that the number of applicants ( $y$ ), and the semester ( $x$ ) are related by the equation  $y = A(B)^x$ , where  $A$  and  $B$  are constants.

- (i) Using a suitable transformation involving  $Y = \log y$ , give a sketch of the scatter diagram. Explain whether the scatter diagram provides evidence that the relation is a reasonable model.
- (ii) Find the equation of the estimated line of regression of  $Y$  on  $x$ , and the least squares estimate of  $A$ .
- (iii) Explain whether the correlation coefficient supports the reasonability of the model.

The admission office realized that they had accidentally left out the data for a “special semester” when the course was also open to applications.

- (iv) After inserting the data for “special semester”, it is found that the equation of regression line of  $Y$  on  $x$  found in part (ii) remains the same. The equation of regression line of  $x$  on  $Y$  is given by  $x = 2.00Y - 0.297$ . Both the line of regression of  $x$  on  $Y$  and the line of regression of  $Y$  on  $x$  pass through the same point  $(\bar{x}, \bar{Y})$ . Show that  $\bar{Y} = 1.98$ , correct to 3 significant figures. Hence, find the number of applicants in the “special semester”, correct to the nearest whole number.

- 14.** A study comparing the amount of advertising time on TV per week for a product and the number of sales per week for the same product was conducted. The results over 8 weeks are given below:

|                                 |     |     |     |     |     |     |     |     |
|---------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| Advertising time (minutes), $x$ | 10  | 12  | 15  | 14  | 17  | 16  | 22  | 20  |
| Sales (thousands), $y$          | 2.3 | 2.8 | $k$ | 3.1 | 3.2 | 2.9 | 5.0 | 4.0 |

- (i) Find the coordinates of the point through which the regression line  $y$  on  $x$  and that of  $x$  on  $y$  both pass. Give your answer in terms of  $k$ .

- (ii) Given that the regression line of  $y$  on  $x$  is  $y = 0.197x + 0.184$ , find  $k$ . Hence find the linear product moment correlation coefficient  $r$  between advertising time and sales per week.
- (iii) Plot a scatter diagram of  $y$  against  $x$ .
- (iv) State with a reason, the effect on  $r$  if the advertising time was in hours instead of minutes.
- (v) Give 2 reasons why it is reasonable to use the regression line of  $y$  on  $x$  to estimate the value of  $x$  when  $y = 3.4$ .



- 15.** In an investigation of a shrub, research workers measured the average width  $y$  (in cm) and stem density  $x$  (number of stems per  $m^2$ ) at 10 sites with the following results:

|     |     |   |      |     |    |      |      |      |      |      |
|-----|-----|---|------|-----|----|------|------|------|------|------|
| $x$ | 4   | 5 | 6    | 9   | 11 | 14   | 15   | 19   | 21   | 22   |
| $y$ | 0.8 | 2 | 0.65 | 0.6 | 0  | 0.65 | 0.55 | 0.35 | 0.40 | 0.38 |

- (i) Calculate the correlation coefficient for the data.
- (ii) Give a sketch of the scatter diagram for the data, as shown on your calculator, and comment on your answer in (i).
- (iii) Identify the 2 data pairs which should be removed, and calculate the correlation coefficient, and the regression line of  $y$  on  $x$ , for the revised data.
- (iv) Use the regression line of  $y$  on  $x$  in (iii) to demonstrate that it is unwise to extrapolate beyond the range of the data.
- (v) A new data pair  $(a, b)$  is now obtained at an 11<sup>th</sup> site, where  $4 \leq a \leq 22$ . Using the revised data in (iii) and the new data pair  $(a, b)$ , the regression equation of  $y$  on  $x$  is calculated and found to be identical to that using only the revised data. Find a possible data pair  $(a, b)$ .



- 16.** The heart rate ( $x$ ) and diastolic blood pressure ( $y$ ), both in suitable units, were measured for each of 10 hospital patients after being given a certain drug. Unfortunately, the nurse who was in charge of recording the measurements did not label the data sets, and the results, hence labeled “Data Set A” and “Data Set B”, are given in the table below:

|     |    |    |    |    |    |    |    |    |    |    |
|-----|----|----|----|----|----|----|----|----|----|----|
| $A$ | 51 | 75 | 54 | 58 | 63 | 78 | 49 | 70 | 64 | 68 |
| $B$ | 88 | 70 | 58 | 91 | 82 | 71 | 90 | 77 | 85 | 76 |



- (i) With the aid of a suitable diagram, comment on the correlation between the 2 data sets, pointing out any irregularities, if applicable.
- (ii) A doctor, on closer look, discovered that not all data points are reliable and that the value of the linear product moment correlation coefficient is only approximately  $-0.4$ . Suggest which pair of data points the doctor should ignore so as to obtain a better value of the linear product moment correlation coefficient, and calculate the new linear product moment correlation coefficient.
- (iii) With the pair of data points dropped in part (ii), the estimated least squares regression line of  $y$  on  $x$  is given as  $y = 128.5 - 0.74x$ .
- State what data set B represents.
  - Estimate the heart rate when the diastolic blood pressure is 80, correct to the nearest whole number.
- (iv) Give a reason why it might be unwise to use either of the regression lines to establish diastolic blood pressure when the heart rate is 90.



- 17.** Water in a reservoir undergoes a purification process before it can be consumed. The effectiveness ( $y \%$ ) of the process for various flow rates ( $x \text{ m}^3 \text{s}^{-1}$ ) is shown below.

|     |    |    |    |    |    |    |    |    |    |
|-----|----|----|----|----|----|----|----|----|----|
| $x$ | 1  | 2  | 4  | 6  | 8  | 10 | 20 | 30 | 40 |
| $y$ | 80 | 60 | 45 | 40 | 30 | 25 | 18 | 15 | 10 |

The variables  $x$  and  $y$  are thought to be related by the equation  $e^y = ax^b$ , where  $a$  and  $b$  are constants.

- Sketch a scatter diagram of  $y$  against  $\ln x$ .
- Find the least squares regression line of  $y$  on  $\ln x$ .
- Describe how you could use parts (i) and (ii) to assess how well the data fitted the calculated equation.
- Estimate the values of  $a$  and  $b$ .
- Predict the effectiveness of the process when water flows at  $49 \text{ m}^3 \text{s}^{-1}$ . Comment on the reliability of your prediction.
- Comment on the validity of the above model for large values of  $x$ .



**Answer keys:****6.1**

2017(Stat.VI).Q1

(a)(i) 48 (ii) 150 (b)(i) 3060 (ii) 0.939

2013(Stat.VI).Q6

(i) 13580 (ii) 288 (iii) 240

2013(Stat.VI).Q4

(ii) 957, 986

2013(Stat.VI).Q7

(iii)  $\frac{5}{32}, \frac{7}{32}, \frac{3}{32}$  (iv) exclusive

2013(Stat.VI).Q5

(i)  $c = 9.14$  (ii) 0.159

2012(Stat.VI).Q1

 $\frac{210}{720}, \frac{378}{720}, \frac{126}{720}, \frac{1}{120}$ 

2012(Stat.VI).Q7

(a) 1260 (b)(i) 1680 (ii) 360 (c) 4920

2012(Stat.VI).Q5

(i) 0.264 (ii) 2830

2012(Stat.VI).Q3

(i)  $m = 926, \sigma = 105$ (ii)  $m = 66150, \sigma = 257$ 

1. 24 ; 12

2. 44,100; 28 ; 36

3. 900 ; 6

4. 186 ; 12

5. 23,460 ; 900 ; 184,756

6. 8640 ; 2880

7. 70 ; 60480

8. 1260 ; 1728 ; 151200

9. 360 ; 1320

10. 10080 ; 100800

11. 119 ; 1485120

12. 24 ; 12 ; 72

13. 60 ; 4

14. 86400

15. 234

16. 3600; 2400

17. 42

18. 900 ; 252

19. 576 ; 1440 ; 144

20. 180 ; 900 ; 900

21.  $5; \frac{n(n-3)}{2}$ 

22. 387600 ; 21840

23. 600

24. 238 ; 2880

25. 277134 ; 261118

27. 8; 4

28.  $\frac{16}{33}; 0.978$ 29.  $\frac{10}{9}; \frac{109}{294}; \frac{185}{588}x - \frac{185}{588}(42-x)$ 

30. 1365; 0.476

31. 17280; 31680

32. 7200; 247

33. 564480; 6720

34. 17280; 43200

35. 64;  $\frac{21}{32}$ 

36. 604800; 10886400

37. 4989600; 32400

38. 57600

39. 3628800; 806400; 80640;

302400

40. 126; 280; 25920

41. 33600

42. 720; 720

43. 39916800; 1008000; 86400

44. 86400; 10886400; 4233600

45. 18; 10

46. 75; 456; 120; 12

47. 87516; 240240

48. 48; 288

49. 672; 595

50. 24; 12; 60

51. 151200

52. 110880

53. 12

54. 26; 79

55. 1680; 1693440; 27720

56. 3628800; 2016000; 645120; 280

57. 259200

58. 2880

59. 5040; 288; 1248

60. 3628800; 2903040; 768

62. 3003; 69854400

63. 3840

64. 11520

65. 60; 4

66.  $\frac{1}{50}; \frac{1}{10}$ 

67. 2 min ; 27

68. 0.365 ; 0.269 ; 0.1

69.  $\frac{21}{2}k; k = \frac{1}{21}; \frac{1}{3}; \frac{3}{32}$ 70.  $\frac{1}{42}; \frac{1}{3}$ 

71. 0.285 ; 0.038

72. no ; no ;  $\frac{1}{6}$ 73.  $2\frac{7}{64}; \frac{2}{13}$ 

74. 0.410

75.  $\text{Var}(X) = 1.6875$  ; Least m = 7 ;  $\frac{4}{21}$ 76.  $a+b-ab$  ;  $a-ab$  ;  $(1-a)(1-b)$  ; 0.877.  $\frac{20}{7}$  ;  $\$3\frac{1}{2}$ 

78. 0.758

79. 0.185 ; 6

80. 0.8 ; 0.404

81.  $\frac{4}{49}; \frac{1}{4}$ 82.  $\frac{144}{205}$  ; 0.530 ; 0.0891

83. 0.65 ; 0.903

84. 0.42 ;  $\frac{24}{29}$ 

85. 0.0139 ; 0.208 ; 0.119 ; 0.111 ; 0.2

86. 0.70 ; 0.6

87.  $\frac{23}{60}; \frac{5}{12}; \frac{7}{23}$ 

88. 0.222

89. 0.0699

90. 0.0316 ; 0.0650 ; 0.268 ; 0.414

91.  $\frac{1}{16}; \frac{1}{50}; \frac{9}{50}$  ; 0.28492.  $\frac{1}{5}; \frac{2}{7}; \frac{1}{2}$ 93.  $\frac{27}{40}; \frac{37}{70}$ 94.  $\frac{3}{4}$ 95.  $k = \frac{5}{2}$  ; 125

96. 0.584 ; 3.33

97.  $\frac{1}{8}; \frac{19}{40}; \frac{19}{30}; \frac{11}{40}; \frac{7}{8}$ 98.  $\frac{1}{5}; \frac{4}{5}; \frac{4}{9}; \frac{216}{11515}$ 99.  $\frac{1}{36}; \frac{209}{1800}; \frac{16}{75}$ 100.  $\frac{7}{328}; \frac{34}{95}; \frac{9}{34}$ 101. 0.4 ; 0.8 ;  $\frac{3}{7}$ 

102. 0.7 ; 0.45 ; 0.166

103.  $\frac{109}{150}$ 104.  $\frac{5}{36}; \frac{79}{216}$  ; 65.6 ; 0.273

105. 0.168 ; 0.900

106.  $\frac{9}{10}; \frac{1}{5}$ 

107. 0.30184

108.  $\frac{1}{3}; \frac{7}{60}$ 109. 2;  $\frac{368}{729}$ 

110. 0.648 ; A = 3 ; B = 3

111.  $\frac{72}{595}; \frac{7}{68}; \frac{36}{245}; \frac{3}{2092}$ 

112. 35 ; 18 ; 53

113.  $\frac{1}{10}, \frac{3}{10}, \frac{4}{25}$  ; 0.203



114.  $\frac{1}{2a}(e^a - e^{-a})$ ;  $e^{-a} \leq y \leq e^a$

115.  $\frac{3}{7}$ ;  $\frac{37}{90}$ ;  $\frac{22}{3}$ ; 0.0649

116. 2.5; 1.25;  $\frac{1}{2}$

117.  $\frac{0.25p}{0.4-0.15p}$ ; 0.429

118. 0.55

119. 0.857; 0.139; 0.603

120. 0.165

121.  $\frac{14}{81}$ ; 0.617; 3.10

122.  $\frac{8}{9}$ ;  $\frac{1}{9}$ ;  $\frac{1}{9}$ ;  $\frac{1}{63}$ ;  $\frac{13}{63}$ ;  $\frac{50}{63}$

123.  $3\frac{1}{4}$ ;  $\frac{1}{2}$

124.  $\frac{11}{14}$

125.  $\frac{32}{195}$ ;  $\frac{17}{30}$

126. 2.45; 0.0433

127. 0.331; 0.439; 0.488; 0.262

128. 1.75

129.  $\frac{1}{2}$ , A and B not independent

130. 0.227; \$4.3125

131. 0.0163

132.  $\frac{14}{325}$ ;  $\frac{21}{325}$ ;  $\frac{63}{650}$ ;  $\frac{89}{208}$

133.  $-3 + a$ ; -2

134. 0.26; 0.63

135.  $\frac{54}{11}$ ,  $\frac{336}{605}$ ; 0.629

136.  $\frac{1}{7}$ ; 19;  $\frac{1}{6}$

137.  $\frac{41}{132}$ , -\$0.62

138.  $\frac{1}{33}$ ; 0.0107; 0.376;  $4.84 \times 10^{-4}$

139. 0.348; 0.00188

140. 0.0315;  $\frac{3}{8}$

141. 0.076; 0.0683; 0.236

142. 0.0625; 0.00764

143. 0.0054; 0.8118; 0.667

144.  $\frac{1}{4}$ ;  $\frac{1}{4}$ ;  $\frac{1}{16}$ ;  $\frac{1}{4}$ ;  $\frac{3}{4}$

145.  $\frac{1}{2}$

146.  $1 - p_B$ ;  $p_A p_B$

147.  $\frac{4}{34}$

148. 0.0500

149.  $\frac{5}{18}$ ;  $\frac{5}{7}$ ;  $\frac{36}{91}$

150.  $(\frac{5}{6})^{n-1} (\frac{1}{6})$ ;  $(\frac{5}{6})^{n-1}$ ;  $\frac{1}{6}$

151. 0.112; 0.461; 0.671

152. 9; 13860; 0.176; 0.827

153.  $\frac{49}{144}$

154.  $\frac{1}{3}$ ;  $\frac{18}{23}$ ;  $\frac{3}{16}$

155. switch doors

156.  $\frac{1}{720}$ ;  $\frac{1}{48}$ ;  $\frac{1}{2(n-2)!}$

157.  $\frac{5}{28}$ ;  $\frac{125}{1568}$

158.  $\frac{34}{45}$ ;  $\frac{135}{374}$

## 6.2

2017(Stat.VI).Q2

(i) constant probability (of completing) and Independent trials/events

(ii) 0.647 (iii) 0.251

2013(Stat.VI).Q3

$$\Sigma(x - 5) = 26, \Sigma(x - 5)^2 = 257$$

2013(Stat.VI).Q4

(i) 0.469 (ii) 2.15 mins

2013(Stat.VI).Q5

(ii) 0.830 (iii)  $\frac{1}{2}$

2012(Stat.VI).Q2

$$\Sigma x = 804, \Sigma x^2 = 27011.76 = 27000$$

2012(Stat.VI).Q4

(ii) LQ = 15400 (iii)  $\frac{4}{33}$

2012(Stat.VI).Q6

(i) 0.531 (ii) 0.136

2012(Stat.VI).Q1

$\sqrt{2}$

2012(Stat.VI).Q5

(ii)  $a = 4.36$

2012(Stat.VI).Q7

(i) 0.294 (ii) 5 (iii) 0.0753 to 0.0754

1. 70

2. 0.0144; 0.0192; 0.428

3. 0.857; 0.0204; 1

4. 0.015; 0.0035

5.  $n = 39$ ;  $\sim 0.0189$ ; 0.876

6.  $\lambda = 5$ ; 0.000251; 0.141; 0.821

7. 0.139; 0.547; 0.868

8.  $n \geq 2.63$ ; 0.986; 0.946

9. 0.558; 0.008; 0.30

10. 0.0121; 0.932

11. 0.264; 19; 0.168; 4

12. 0.143; 0.00376

13. 0.911; 0.0340; 0.765

14.  $E(Y) = \frac{5}{2}$ ,  $\text{Var}(Y) = \frac{5}{4}$ ;

$\text{Var}(W) = \frac{25}{6}$ ;

$P(W=4) = 0.0139$

15. 2; 0.9932

16. 0.911; 12.3; 0.3505; 0.2627

17.  $1.47 \times 10^{-3}$

18.  $\frac{1}{4}$

19. 0.184; 0.0414

20.  $\frac{5}{3}$ ;  $\frac{1}{3}x$ ;  $\frac{1}{3} + \frac{1}{6}x$

21.  $k \geq 92$ ; 0.9772

22.  $k = \frac{1}{10}$ ; 0.104

23. 0.146; 0.5498

24. 0.135; 0.323; 0.1429; 0.9515

25. 6

26. 0.0148; 3; 0.647

27. 0.323; 0.999; 0.158

28. 0.143; 0.0338

29. 0.0186; 15.1; 17.8; 2 vans

30. 0.423; 0.9378; 0.045

31. 0.633; 0.8515; 0.1508

32. 0.064; 0.7386

33. 0.594

34. 0.0226; 0.848; 0.386

35. 0.109; 0.478; 0.824

36. 0.423; 5; 0.0207

37. 0.359; 0.0425

38. 0.515

39. 0.16, 0.18; 0.996; 0.0797

40. 3, 4.8; 0.1; 0.260

41.  $Y \sim \text{Bin}(50, \frac{3}{4})$ ; 36; 0.743

42. 0.170; 21; 0.9913

43. 0.380; 0.0305; 3; 0.0291

44. 0.974; 42

45. 0.000239; 0.323; 0.809

46. 0.265; 0.109; 0.371

47. 2; 0.231

48. 0.081; 0.964

49. 0.528; 0.451

50. 0.986; 0.516; 0.0464

51. 0.99144; 0.331; 0.238

52. 0.541; 0.449

53. 331; 0.484

54. 0.585

55. 0.303; 0.485

56. 0.0964; 5

57. 0.0547; 0.0538

58. 0.404; 3

59. 0.323

60. 0.329; 0.932

61. 0.738; 0.173

62. 0.130; 0.216; 0.762

63. 1; 0.7386;  $\{n \in \mathbb{Z}^+ : n \geq 7\}$

64. 0.315; 9.42 am

65. 0.235; 78

66. 15; 0.9458

|                                       |  |   |
|---------------------------------------|--|---|
| 67. 1.10; 0.652; 0.166                | 9. $\sigma = 0.150 ; 0.650 ; 0.11 \text{ cm}$                    | 57. 1.28; 0.0185                          |
| 68. 0.990; 15                         | 10. $\mu = 3 ; 0.0116 ; 0.885$                                   | 58. 0.0619; 0.109; 0.906                  |
| 69. 0.0949; 10                        | 11. 0.200 ; 0.038  | 59. 0.0142                                |
| 70. 0.769; 0.933; 0.737               | 12. 0.131 ; 0.504  | 60. 0.7; 5, 33; $232\pi, 221184\pi^2$     |
| 71. 0.240                             | 13. $n = 8.36$   | 61. (289.465, 310.735)                    |
| 72. 0.762; 0.258; 0.0963              | 14. 0.922 ; 0.916 ; $n = 155$                                    | 62. 0.8999                                |
| 73. 0.363                             | 15. 0.0765 ; 0.0745 ; 0.334                                      | 63. 5; 0.1304                             |
| 74. 0.75, 44; 0.2                     | 16. 0.1047   | 64. 0.0559; 0.3814; 0.2892                |
| 75. 0.991; 0.0959                     | 17. 0.0484 ; 0.083 ; $n = 5$                                     | 65. (121, 134); 0.573; 0.266              |
| 76. 1.22; 0.0802                      | 18. 0.252 ; 0.884 ; 1692   | 66. 5.88; 0.588; 0.3275; 0.2562           |
| 77. 0.326; 0.266                      | 19. $\frac{3}{8} ; 54.1 ; 0.650$                                 | 67. 0.224; 0.168                          |
| 78. 0.753; 0.985; 9; 0.0241           | 20. 0.00156 ; 0.570  | 68. $\frac{275}{12} ; 0.00253$            |
| 79. 0.363                             | 21. 0.0038   | 69. 0.0401; 0.929; 0.0346                 |
| 80. 0.0198; 0.0571; 0.441             | 22. 0.337 ; 0.4637   | 70. 302                                   |
| 81. 0.0391; 0.0965; 0.865             | 23. 0.3700 ; $k = 98$  | 71. 0.0653; 0.978; 350                    |
| 82. 1.50; 0.957; 8                    | 24. 0.0383   | 72. 16; 80.3; 0.673                       |
| 83. 0.3324; 0.295                     | 25. $p = 0.08 ; q = 0.30 ; 0.377 ; n = 3$                        | 73. 0.0514; 296; $Y \sim N(0.9, 0.052^2)$ |
| 84. 1.82; 0.791                       | 26. $n = 9 ; 2, 3$   | 74. 0.0668; 0.748                         |
| 85. 22.1%; 0.0844                     | 27. $\mu = 48.6 ; (25.1, 72.1) ; 0.7405$                         | 75. 0.4753                                |
| 86. 0.841; 0.00181                    | 28. $\bar{x} = 24.75 ; \sigma^2 = 4.15 ; 0.066 ; (0.105, 0.202)$ | 76. 0.1, 0.2                              |
| 87. 0.0273; 0.234; 20; 0.0143         | 29. 0.0913 ; 0.2895 ; 0.1587                                     | 77. 0.5                                   |
| 88. 0.224; 0.167; 0.235               | 30. 25 ; (9.54, 12.46) ; 200                                     | 78. 0.0548; 0.9637; 0.158; 0.9574         |
| 89. 1; 0.0212; 0.0170; 0.753          | 31. 0.987 ; 0.567  | 79. 11.9                                  |
| 90. 0.7; 0763; 0.441                  | 32. 0.904 ; 0.702  | 80. 208.2; 6.78                           |
| 91. 0.0735; 27                        | 33. 75 ; 0.3182 ; 0.130 ; 0.2306                                 | 81. 344; 0.968; 0.108; 12                 |
| 92. 0.715; 0.999; 0.0156              | 34. $\mu = 23.334; \sigma = 1.979$                               | 82. 0.730; 0.821; 163.49; 0.0652          |
| 93. 0.082; 0.0221; 0.0787             | 35. 0.985 ; 0.288 ; 0.882  | 83. 217; 83                               |
| <b>6.3</b>                            |  |   |
| 2017(Stat.VI).Q3                      | 36. 12   | 84. 0.267; 0.117; 0.788; 727              |
| (a)(i) 5.82 (ii) 0.221 (b) 788 or 789 | 37. 1.34; 0.313 ; 0.253 ; 0.378                                  | 85. 3; 0.614                              |
| 2013(Stat.VI).Q1                      | 38. 0.323 ; 0.523  | 86. 0.252; 0.133                          |
| 0.0328                                | 39. 156.7 ; 0.2819 ; 4   | 87. 330; 0.125; 0.147; 0.0182             |
| 2013(Stat.VI).Q7                      | 40. 0.1295 ; 278   | 88. 172, 1000; 0.16; 0.0222               |
| (i) 0.889 (ii) 0.176                  | 41. 0.0965   | 89. 0.382                                 |
| 2012(Stat.VI).Q3                      | 42. $\mu = 1000 ; \sigma = 172 ; 0.675$                          | 90. 0.341                                 |
| (i) $\sigma = 1.93$ (ii) 0.895        | 43. 0603 hr ; 0.982  | 91. 0.261                                 |
| 1. 106 min                            | 44. $m = 25 ; E(S) = 0 ; \text{Var}(S) = 100$                    | 92. 0.153; 0.721; 0.614                   |
| 2. 0.324 ; 0.0109 ; 5.5 grams         | 45. 0.577 ; 0.817  | 93. 0.0668; 0.819; 266.45; 0              |
| 3. 0.440 ; 0.109 ; 0.444              | 46. $9\frac{1}{3} ; t = 40$                                      | 94. 0.0123; 19.4, 0.789; 0.122; quota     |
| 4. 0.970 ; 0.104                      | 47. 0.198 ; 0.719 ; 0.409  | 95. 0.134; 0.738                          |
| 5. 0.238 ; 0.118                      | 48. narrower; wider.   | 96. 0.962                                 |
| 6. 0.1587 ; 0.334 ; 0.7111            | 49. 0.116; 0.864; 5  | 97. 4.32; 0.742; 0.153                    |
| 7. 0.930 ; 0.226 ; 0.371              | 50. 0.125; 0.987; 0.0855; 5                                      | 98. 0.937; 0.478; 0.0188; $a \geq 15.2$   |
| 8. 0.203 ; 0.420 ; 0.0182             | 51. 0.113; 0.759; 0.655  | 99. 0.0787                                |
|                                       | 52. 0.1562; 0.1136; 0.1864; 0.0022                               | 100. $\sigma = 16.3, \mu = 63.7 ; 0.0021$ |
|                                       | 53. $\mu_0 < 6.42$   | 101. 0.8911; 0.316                        |
|                                       | 54. 0.189; 0.707   | 102. 4.24; 0.966; 0.7275                  |
|                                       | 55. $\frac{8}{11} ; 0.703$                                       |   |
|                                       | 56. 0.9672; 0.09304  |   |

**6.4**

2017(Stat.VI).Q4

- (i) 299.83 to 300.37 or (299.83, 300.37),  
(ii) The confidence interval includes 300 so the claim is supported



2013(Stat.VI).Q2

20

2012(Stat.VI).Q4

(i) 0, 1 (ii) 0.0243 (iii) no

2012(Stat.VI).Q6

(ii) (0.119, 0.261) (iii)  $x = 93$ 

1. 0.311

2. 0.651

3.  $3\frac{7}{32}$ ; 0.0884.  $\frac{1}{2}$ ; 62

6. 0.786

7.  $\mu = 7$ ;  $\frac{4}{7}$ ; 0.119; 0.8319. 0.928;  $b \approx 0.221$ ; 0.977210.  $n=34$ 

11. 0.953

12. 0.283;  $n \geq 637.3$ 13.  $k = \frac{1}{91}$ ; 0.0544; \$56.3714.  $a = 0.2$ ,  $b = 0.4$ ; 0.31; 0.1377

15. 0.0474; 80; 0.356

16.  $0.2182$ ;  $D = 10 \cot \theta$ ;

$$\frac{4}{\pi} \left( \frac{\pi}{2} - \cot^{-1} \frac{d}{10} \right)$$

17. 0.146; 0.0167

18.  $E(X) = 5$ ;  $a = 3$ ,  $b = 5$ ; 0.682

21. 0.0527; 0.1106; 0.0119

22.  $\frac{5}{17}$ ;  $\frac{27}{64}$ ;  $E(X+1) = \frac{11}{10} \ln 11$ ,

$$\text{Var}(X) = 4.04$$

$$\tilde{X} \sim N(1.638, 0.05387)$$

23.  $\frac{3}{32}$ 24.  $n = 120$ ;  $\alpha = 2$ 25.  $\frac{91}{54}$ 

26. (45.5, 49.5); 0.039

27.  $\frac{1}{2}, \frac{1}{8}$ (n.a.); -0.176; 0.071; 128.  $\frac{1}{30}$ ;  $\frac{31}{45}$ ;  $\frac{7}{15}$ 

29. 0.208; 0.534; 85

30. (0.805, 0.945)

31.  $a = 9$ ,  $b = -3$ ,  $\mu = \frac{1}{3}$ ;  
 $a = 0.569$ ; 174

32. 7.92; 0.231

34. (0.556, 0.710); 0.154

35.  $\frac{5}{3}$ ;  $5\frac{1}{18}$ 

36. 0.9453; 0.0661

37. (0.28, 0.36)

38.  $\sqrt{541}$ ;  $2\frac{1}{7}$ 39.  $n = 6$ 

40. 0.594

41. 3.146; 67.3%

42.  $1\frac{13}{162}$ 43.  $k = 3$ ; 0.206; 1.16; 0.0540

44. 10.54; (9.54, 11.81)

45. 9.8, 1.2; (9.55, 10.1); 21, 95; 375

46. 90%;  $39 \leq y \leq 200$ 

47. 0.9772; 136; 80

48. 0.278; 0.05

49. (0.737, 0.823); 264

50. 5; 98; 50

51.  $\alpha > 3.44\%$ ; (1527.086, 1528.914); 60052. 53.8, 24.4;  $52.8 < \mu < 54.8$ ;  
53.253.  $1.5, \frac{1}{12}$ ; 0.989354.  $13, \frac{17}{8}$ ; 0.0728

55. 92.5%; 0.00260

56. 280

57. 47.488, 330.986; (45.6, 49.4);  
45.5;  $\alpha > 4.00$ 

58. 0.48; 150; 0.794

59. 28.47, 13.89; 98.4%

60. 119.9, 0.2625; 0.9347

61.  $1.74 < \mu < 2.06$ ; 1.9662.  $\alpha > 6.75$ 

63. 97; 2.5; 124

64. 165, 1130; 0.991; 182

66. 80,  $82\frac{3}{7}$ ;  $\bar{X} \sim N(80, \frac{115}{7})$ ; 0.109;  
117

68. 0.941

69. stratified; systematic

70. 225; 2.32

72. 111; 0.889; -1.41

74. 139

75. quota

76. 42

77. 31.28, 14.4; 0.0896

80. 190

81. 5.794, 0.882;  $5.607 < \mu < 5.982$ ;

6.52

82. 553

**6.5**

2017(Stat.VI).Q5

Assume standard deviation  
unchanged, No evidence that  
heights are shorter

2013(Stat.VI).Q3

(i)  $\mu_{\text{est}} = 2866$  or  $2870$ ,  $\sigma_{\text{est}}^2 = 4130$ 

2013(Stat.VI).Q6

(i) 0.256 (ii) 0.117 (iii) Type I

2012(Stat.VI).Q2

increased

1. 90.1

2. 0.0384; 0.0768

3. 54

4.  $\frac{13}{3888}$ 5.  $\hat{\mu} = 5.8$ ;  $\hat{\sigma}^2 = 3.638$ ; 10%6.  $0.0479 \leq \alpha \leq 1$ 7.  $\alpha > 5.31\%$ 

8. 0.182; 0.0026

9. 2.6%

10. (5.98, 6.42);  $\mu_0 > 6.47$  or  
 $\mu_0 < 5.93$ 

11. 64

13. 6.58 am

15. 1139;  $\alpha > 3.5\%$ 

18. (3.89, 3.91)

19.  $n = 106$ ;  $100 > \alpha > 2.28$ 

21. (0.18, 0.38); 437

22. 24.68; (24.3, 25.1); 0.05

23. 94

24. 95

25. 364

26. 3.3%

27. 35.8; (26.5, 28.5)

28.  $m > 497$ 

29. 2.37; 250.8 ml

30. 1120; 8181.8; 112

32. 1344; 6.53%

33.  $\bar{X} < 63.3$  or  $\bar{X} > 66.7$ 

34. 0.680; 0.353

35.  $\frac{7850}{149}$ ; 4.58%

36. 13.19%

37. -1; 9900;  $\mu_0 > 1.93$  or  
 $\mu_0 < -3.93$ 

38. (30.6, 31.8)

39. (0.193, 0.357);  $\alpha > 1.0$ 

40. 6

41. 5.13%

42. 63.2

43. 4, 16

44. 78.3%; 8

48. 86.9, 9.27

50. 1178.14, 787.02

51.  $5.50 < \bar{X} < 6.50$ 52.  $a = 15$ ;  $b = 9$ ;  $\alpha = 0.2398$ 

53. 10% significance test

54.  $m < 18.0$ 55.  $H_0: \mu = 23.9$ ,  $H_1: \mu > 23.9$ 

56. 12.7%

57.  $\bar{x} = 5.055$ ,  $s^2 = 0.00187$

58.  $\mu_0 > 150$

60.  $\alpha < 4.01$

62.  $\alpha > 2.28$

## 6.6

1.  $-3.00, 0.392, (1.603, 2.797)$ ;

$(x - 1.108)^2 = 0.662y + 1.718$ ,

3.03 or 0.817

2.  $0.147$ ;  $y = 4.55x + 31.6$ ; 68

3.  $50.2$ ;  $0.863$ ;  $a = -163.14$ ,  
 $b = 53.72$ ,  $r = 0.905$ ; 92.39

4.  $r = 0.932$ ; 16.2%; 3.08

5.  $20$ ; -0.888

6.  $x = 1.514y - 46.03$

7.  $0.994$ ; 0.000165

9.  $r = 0.98$

10.  $y = -\frac{18}{25}x + \frac{97}{25}$

11.  $c = 51$ ,  $l = 34$ ;

$l = 19.722 + 0.86981c$ ; 49.7,

58.9

12.  $73$ ,  $t = m - 62$ ; A;  $r = 0.919$

13.  $Y = 0.781 + 0.327x$ ; 6.04; 190

14.  $\bar{x} = 15.75$ ,  $y = \frac{23.3 + k}{8}$ ;  
 $r = 0.9287$

15.  $r = -0.542$ ; (5,2) and (11,0),

$y = 0.840 - 0.0213x$ ;

(13.75, 0.5475)

16.  $r = -0.928$ ; diastolic blood  
pressure,  $y = 80$

17.  $y = 73.3 - 18.4 \ln x$ ;  $a = 6.89 \times 10^{31}$ ,  
 $b = -18.4$ ; 1.84%