

## Chapter 13 practice questions

1. The producer of a TV dancing show asked a group of 150 viewers their age and the type of Latin dance they preferred. The types of Latin dances in the show were Argentine tango, samba, rumba and cha-cha-cha. The data obtained are shown in the table.

	Dance			
	Argentine tango	Samba	Rumba	Cha-cha-cha
20 years old and younger	35	23	12	10
Older than 20 years old	20	17	18	15

A  $\chi^2$  test was carried out, at the 5% significance level.

- Write down the null hypothesis for this test.
- Write down the observed number of viewers who preferred rumba and were older than 20 years old.

(c) Use your graphic display calculator to find the  $p$ -value for this test.

The producer claims that the type of Latin dance a viewer preferred is independent of their age.

- Decide whether this claim is justified. Give a reason for your decision.
2. Four parties contested an election held two years ago. Conservatives captured 29% of the vote, liberals 41% of the vote, while the greens and the extreme right parties split the rest among themselves equally. A survey of 1500 voters this year yielded the following: 410 conservatives, 590 liberals, 270 greens, and 230 went to the extreme right party. Has the voter support changed since the elections? Test at the 10% level of significance.
- If the voter support has not changed, what is the expected number of voters for each party?
  - Write down the hypotheses for this test.
  - Perform the test in question.
3. In a debate on voting, a survey was conducted. The survey asked people's opinion on whether the minimum voting age should be reduced to 16 years of age. The results are shown as follows.

	Age 18–25	Age 26–40	Age 41+	Total
Oppose the reduction	12	20	48	80
Favour the reduction	18	15	17	50
Total	30	35	65	130

A  $\chi^2$  test at the 1% significance level was conducted. The  $\chi^2$  critical value of the test is 9.21

(a) State:

(i)  $H_0$ , the null hypothesis for the test

(ii)  $H_1$ , the alternative hypothesis for the test.

(b) Write down the number of degrees of freedom.

(c) Show that the expected frequency of those between the ages of 26 and 40 who oppose the reduction in the voting age is 21.5, correct to three significant figures.

(d) Find:

(i) the  $\chi^2$  statistic

(ii) the associated  $p$ -value for the test.

(e) Determine, giving a reason, whether  $H_0$  should be rejected.

4. It has been estimated that employee absenteeism costs US companies more than \$100 billion per year. To understand the trend of absenteeism among its employees, a large corporation recorded the weekdays during which individuals in a random sample of 724 absentees were away over the past 6 months, as shown in the table.

Do these data suggest that absenteeism is higher on some days of the week than on others? Test at the 10% level of significance.

Day of the week	Monday	Tuesday	Wednesday	Thursday	Friday
Number absent	174	124	142	136	148

5. Many restaurants use a device that allows credit card users to pay with their credit card at the table. The device also allows the user to specify an amount for a waiter tip. In an experiment to see how it works, a random sample of credit card users was drawn. Some paid with the device and some paid in cash. The percentage left as a tip was recorded and listed below.

Can we infer that users of the device leave larger tips? Test at the 10% level of significance.

Cash	10.3	15.2	13.0	9.9	12.1	13.4	12.2	14.9	13.2	12.0	
Device	13.6	15.7	12.9	13.2	12.9	13.4	12.1	13.9	15.7	15.4	17.4

- (a) What type of test is this? Give two assumptions you are making.  
 (b) Write down your hypotheses for the test.  
 (c) Draw your conclusion and give a reason for it.

6. The manager of a travel agency surveyed 1200 travellers. She wanted to find out whether there was a relationship between a traveller's age and their preferred destination. The travellers were asked to complete the survey shown.

My age is:

25 or younger	26–40	41–60	61 or older

My preferred destination is:

New York	Tokyo	Melbourne	Dubai	Marrakech

A  $\chi^2$  test was carried out, at the 5% significance level, on the data collected.

- (a) Write down the null hypothesis.  
 (b) Find the number of degrees of freedom.

The critical value of this  $\chi^2$  test is 21.026

- (c) Use this information to write down the values of the  $\chi^2$  statistic for which the null hypothesis is rejected.

From the travellers taking part in the survey, 285 were 61 years or older and 420 preferred Tokyo.

- (d) Calculate the expected number of travellers who preferred Tokyo and were 61 years or older.

7. A survey was conducted among a random sample of people about their favourite TV show. People were classified by gender and by category of favourite TV show (sports, documentary, news and reality TV).

The results are shown in the contingency table.

	Sports	Documentary	News	Reality TV	Total
Male	20	24	32	11	87
Female	18	30	20	25	93
Total	38	54	52	36	180

- (a) Find the expected number of females whose favourite TV shows are documentaries.

A  $\chi^2$  test at the 5% significance level is used to determine whether TV show category is independent of gender.

- (b) Write down the  $p$ -value for the test.  
 (c) State the conclusion of the test. Give a reason for your answer.

8. A psychologist examined whether the content of TV shows influenced the ability of viewers to recall brand names of items featured in commercials. The researcher randomly assigned volunteers to watch one of two programs, each containing the same nine commercials. One of the programs had violent content and the other had neutral content. After the show ended, the subjects were asked to recall the brands of products that were advertised. Results are listed below.

Violent: mean = 3.18, standard deviation = 1.86,  $n = 108$

Neutral: mean = 4.16, standard deviation = 1.75,  $n = 108$

Do these results indicate that viewer memory for commercials may differ depending on program content? Test at the 10% level of significance.

Hours	Mass lost (%)
150	0.761
200	1.44
200	1.15
300	1.65
450	2.56
500	2.44

Table 14.9 Data for question 1

## Exercise 14.3

- Six metal plates are immersed in an acid bath. After some time, they are removed, and the mass measured to determine the mass that was lost. Table 14.9 gives data from this experiment.
  - Find the equation of the least-squares regression line for mass lost,  $L$ , in terms of hours immersed,  $h$ . Interpret your model in context.
  - For each additional 100 hours a metal plate was in this acid bath,  $k$  percent of the mass was lost. Find the value of  $k$ .
  - Use your least-squares regression line to predict the mass loss of a metal plate immersed for 400 hours.
  - Is it appropriate to use your model to predict the number of hours a metal plate with 1.3% mass loss was immersed? Explain.
- Does the acceleration ability of a car affect fuel efficiency? Here are data from 15 cars.

Car	0–60 mph time (s)	Fuel consumption (miles gal <sup>-1</sup> )
Mazda MX-5 Miata Club	6.7	34
Honda Civic Si	7.3	34
Fiat 124 Spider Lusso	7.1	31
Mini Cooper S	7.0	30
Subaru BRZ Premium	7.2	30
Toyota 86	7.2	30
Volkswagen GTI Autobahn	6.6	29
Ford Fiesta ST	7.3	29
Fiat 500 Abarth	8.0	28
Porsche 718 Boxster (base)	4.4	26
Subaru Impreza WRX Premium	6.0	26
Audi TT 2.0T (AT)	6.3	26
Ford Focus ST	6.6	26
BMW M235i	5.2	25
Ford Mustang Premium (2.3T, AT)	6.4	25

- Find the equation of the least-squares regression line for fuel consumption,  $E$ , in terms of acceleration time,  $t$ . Interpret your model in context.
- Use your least-squares regression line to predict the fuel consumption of a car with a 0–60 mph time of 5.7 s. Give a reason why this prediction may not be reliable.
- Is it appropriate to use your model to predict the efficiency of a car with a 0–60 time of 9 seconds? Explain.

3. Do bigger aeroplanes use more fuel? The table contains data on the number of passenger seats and fuel consumption of aeroplanes currently in widespread use.

<b>Seats</b>	405	296	288	258	240	230	193	188	148
<b>Fuel consumption (L min<sup>-1</sup>)</b>	224	140	138	94	100	95	87	62	79

<b>Seats</b>	142	131	122	115	112	103	102	78
<b>Fuel consumption (L min<sup>-1</sup>)</b>	56	46	54	70	51	40	51	48

- (a) Generate a scatter diagram and describe the association between the number of seats and fuel consumption.
- (b) Find the equation of the least-squares regression line for fuel consumption,  $F$ , in terms of the number of seats,  $n$ . Interpret your model in context.
- (c) Use your model to predict the fuel consumption of an aeroplane with 350 seats.
- (d) Give a reason why a linear model may not be the most appropriate model for this data.
4. A computer-based workout app ranks its workouts according to the training stress score. Another training website calculates a separate relative effort value for each workout that is uploaded to the site. Are training stress score and relative effort related? The table shows a sample of one cyclist's last 12 workouts.

<b>Training stress score</b>	82	96	48	79	79	79	88	112	82	74	74	96
<b>Relative effort</b>	59	70	37	60	58	51	67	79	48	44	46	84

- (a) Draw a scatter diagram and describe the association between training stress score and relative effort.
- (b) Find the equation of the least-squares regression line for relative effort,  $E$ , in terms of the training stress score,  $S$ . Interpret your model in context.
- (c) Use your model to predict the relative effort for a workout with a training stress score of 60.
- (d) Write down a valid domain for your model.
- (e) If you removed the possible outlier at (48, 37), how would your answers to (d) and (c) change?
- (f) Remove the outlier at (48,37) and recalculate the least-squares regression line. Did the model change significantly?

Change in NEA (kcal)	Change in mass (kg)
-94	4.2
-57	3
-29	3.7
135	2.7
143	3.2
151	3.6
245	2.4
355	1.3
392	3.8
473	1.7
486	1.6
535	2.2
571	1
580	0.4
620	2.3
690	1.1

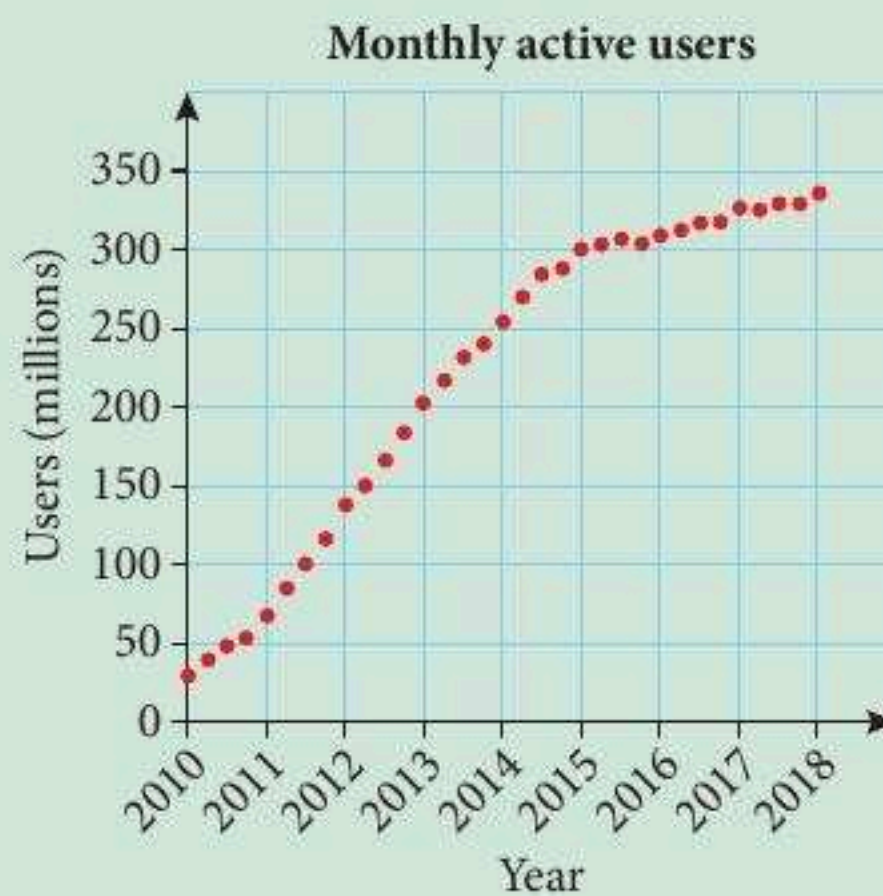
Table 14.10 Data for question 5

Defects	Production rate (units per hour)
20	400
30	450
10	350
20	375
30	400
25	400
30	450
20	300
10	300
40	300

Table 14.11 Data for question 6

5. It is well known that many people who eat more don't gain weight as quickly as others do. Could non-exercise activity (fidgeting, maintaining posture, etc.) explain the difference? In a 1999 study, researchers deliberately overfed 16 non-obese volunteers and measured the change in their calorie expenditure through non-exercise activity (NEA) as well as their weight gain after 8 weeks. The data collected are shown in Table 14.10.
- Generate a scatter diagram and describe the association between change in non-exercise activity and change in mass.
  - Find the equation of the least-squares regression line for change in mass,  $M$ , in terms of the change in non-exercise activity,  $N$ . Interpret your model in context.
  - Use your model to predict the change in mass for an individual with a change in non-exercise activity of 200.
  - Explain what a negative value for change in non-exercise activity means in this context.
  - Write down a valid domain for your model.
  - Making a prediction for change in mass given a change in non-exercise activity of 1100 is an extrapolation. Give another reason why it is particularly nonsensical in this context.
6. A company believes that its workers make more mistakes when they work faster. To test this theory, they collect data on the production rate and the number of mistakes at that rate. The data are given in Table 14.11.
- Identify the explanatory and response variable in this context.
  - Generate a scatter diagram and describe the association between defects and production rate.
  - Find the equation of the least-squares regression line for defects,  $D$ , in terms of the production rate,  $P$ . Interpret your model in context.
  - Find the value of  $r$  and interpret in context.
  - Would you advise this company to lower production rates in order to reduce defects?
  - Remove the outlier from the data and re-do parts (b)–(e).
  - Is it OK to remove this outlier in this case? Explain.

7. To track the number of active users for a website, it is common to count the number of users who have logged in at least once in a given month. This count is called monthly active users. The number of monthly active users for one social media service is shown in the scatter diagram.



- Describe the association shown in the scatter diagram.
- A piecewise linear model may be appropriate. Give the domain for each piece of a piecewise linear model, using years since 2010 ( $y$ ) as the explanatory variable.

8. A dairy farmer would like to investigate whether temperature has an effect on the butterfat content of the milk his dairy cows produce. Over the course of a year, he records the butterfat content of the milk they produce and the air temperature. His data are shown in Table 14.12.

- Identify the explanatory and response variable in this context.
- Generate a scatter diagram and describe the association between temperature and butterfat content.
- Find the equation of the least-squares regression line for butterfat,  $F$ , in terms of the temperature,  $T$ . Interpret your model in context.
- Find the value of  $r$  and interpret.
- If the farmer wants to increase the butterfat content of the cows' milk, would investing in a climate-controlled barn for the cows be a good choice? Explain your answer.

Temperature (°C)	Butterfat (%)
3	4.87
3	5.09
4	4.97
5	4.52
7	4.83
8	4.85
13	4.77
13	4.48
13.5	4.23
13.5	4.85
14	4.51
14.5	4.74
15	4.45
15.5	4.7
16	4.65
16.5	4.45
18	4.63
18	4.65
18.5	4.65
18.5	4.59

Table 14.12 Data for question 8

## Chapter 14 practice questions

- What is wrong with each statement?
  - The value of Pearson's correlation coefficient for the time it takes an athlete to run 5 km and the time it takes them cycle 30 km is  $r = 1.21$
  - The fuel consumption of cars decreases linearly with the mass of the car with Pearson's  $r = 0.78$
  - Among mammals, those with greater average body mass have longer life expectancy; the rank correlation coefficient is  $r_s = -0.85$
  - For a set of  $(x, y)$  data, the least-squares regression line is  $y = 25 - 3.52x$  with  $r = 0.64$