Chapter 10: Statistics

Example 10.1

Here is some data in raw form. Prepare a frequency table starting with 20 and a class interval of 5. Then draw a histogram of the data and a cumulative frequency graph.

29	36	40	44	48	52	54	56	59	60	61	61	62	63	64	66	68	71	72	73	63	64	66	68
31	36	41	44	49	52	54	57	59	60	61	62	62	64	64	66	68	71	72	75	63	64	66	68
33	36	41	44	49	52	55	57	59	60	61	62	62	64	65	66	69	71	72	35	38	43	47	71
34	37	41	45	49	53	55	58	59	60	61	62	63	64	65	66	69	71	73	36	40	44	48	71
34	37	42	45	50	53	55	58	59	60	61	62	63	64	65	67	70	71	73	50	54	56	59	72
35	37	42	45	50	53	55	58	59	60	61	62	63	64	65	67	70	71	73	51	54	56	59	72
35	37	43	46	50	54	55	58	59	60	61	62	63	64	65	67	70	71	73	60	60	61	62	73
35	38	43	46	50	54	55	58	59	60	61	62	63	64	65	67	70	72	73	60	61	61	62	73

Example 10.2

The following are the five closing prices of a stock market index for the first business week in November 2018. This is a sample of size $n = 5$ for the closing prices from the entire population: 2794.83, 2810.38, 2795.18, 2825.18, 2748.76. What is the average closing price for the first business week in November 2018?

Example 10.3

The table lists the frequency distribution of 25 families in Lower Austria who were polled in a marketing survey to state the number of litres of milk consumed during a particular week. Construct a frequency histogram and find the modal class, median and mean.

Number of litres	Frequency	Relative frequency
0	2	0.08
1	5	0.20
2	9	0.36
3	5	0.20
4	3	0.12
5	1	0.04

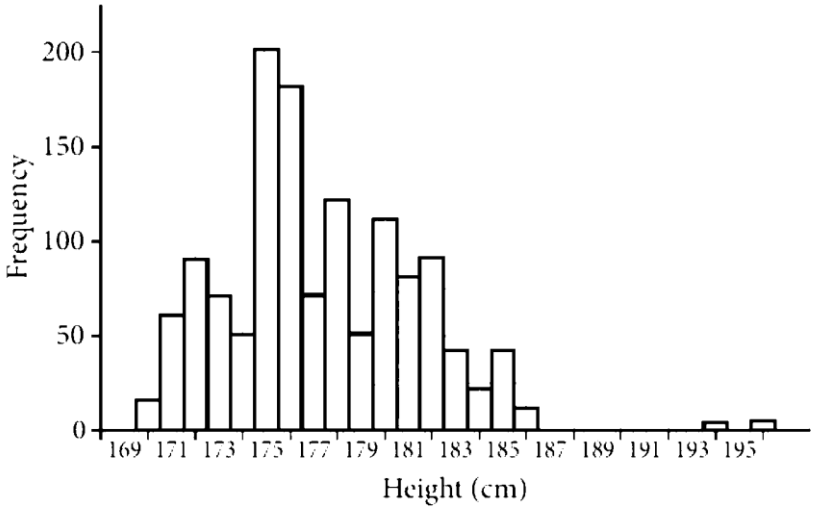
Example 10.4

Speed limits in some European cities are set to 50 km h^{-1}. Drivers in various cities react to such limits differently. The speeds of cars in Vienna, Brussels, and Stockholm are given in the table. Use box-and-whisker plots to compare the results.

Vienna	62	60	59	50	61	63	53	46	58	49	51	37	47	51	63	52	44	50	45	44
Brussels	64	61	63	57	49	49	46	58	45	60	51	36	65	45	47	46				
Stockholm	43	44	34	35	31	34	29	33	36	38	45	47	29	48	51	49	48			

Example 10.5

The records of a large high school show the heights of their students (to the nearest cm) for the year 2016.



(a) Which statistics would best represent the data here? Why?

(b) Estimate the mean and standard deviation.

(c) Develop a cumulative frequency graph of the data.

(d) Use your result from part (c) to estimate the median, Q_1, Q_3, and IQR.

(e) Are there any outliers in the data? Why?

(f) Write a few sentences describing the distribution.

Example 10.6

The table below lists the fuel consumption in km/litre of 34 small cars during both city driving and highway driving. Draw a scatter plot of the data and comment on any patterns you observe.

City	7.3	8.5	8.5	7.3	7.7	5.1	4.7	4.3	7.3	3.8	3.8	6.4
Highway	10.2	11.9	11.9	10.7	10.7	8.5	6.8	6.8	9.8	6.4	5.5	9.4

City	5.1	9.4	6.8	5.5	8.5	8.5	6.4	11.1	5.1	9.0	8.1	8.1
Highway	7.3	11.9	9.8	8.1	11.1	12.4	9.8	13.7	8.1	12.4	11.5	11.9

City	6.8	7.7	6.8	7.7	10.7	9.8	8.5	7.7	6.0	25.6
Highway	9.8	11.1	9.8	9.8	13.7	13.2	12.4	11.1	9.4	28.2

Example 10.7

The table below gives the data for a lab experiment involving the length (mm) of a metal alloy bar used in electronic equipment when it is exposed to heat.

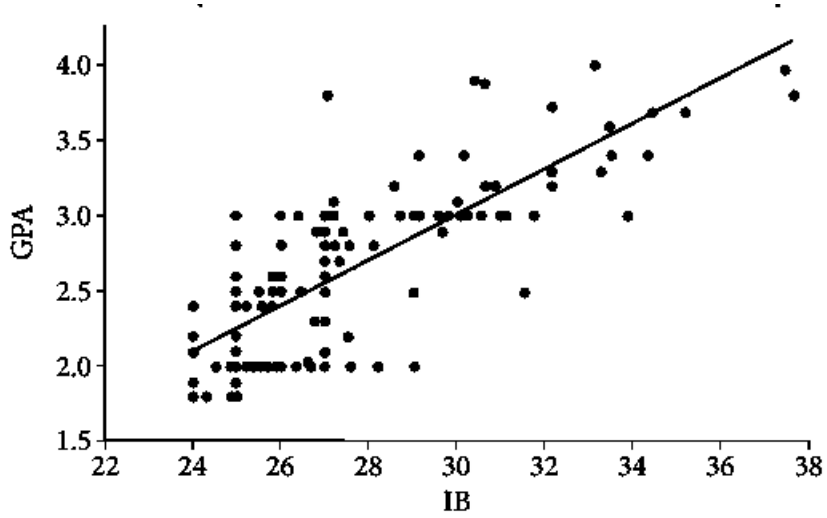
Temp(℃)	40	45	50	55	60	65	70	75	80
Length	20	20.12	20.20	20.21	20.25	20.25	20.34	20.47	20.61

Draw a scatter plot. Comment on the strength of the relationship.

Use both r and r^2.

Example 10.8

This scatter plot represents a random sample of IB students who went through four years of university, and is a comparison of their scores on the IB exams they took and their GPAs in their university studies (scale $1-4$).



There appears to be a linear relationship. When we run a linear regression, the equation is:

$$\text{GPA} = -1.51 + 0.151\ \text{IB}$$

This means that, on average, for every increase of 1 point in the total IB score, we expect an increase of 0.15 points in the GPA. If we want to predict the GPA of a student who scored 30 on an IB diploma, the model predicts, on average, a grade of:

$$\text{GPA} = 21.51 + 0.151(30) = 3.02$$

The correlation coefficient of this relationship is $r = 0.758$, which is a relatively strong correlation. In addition, $r^2 = 75.5\%$. This means that changes in the IB score may help us explain 57.5% of the variation in the GPA.

Does that mean high IB scores cause high university grade averages? The answer is no. They only help predict the future university grade averages.

Example 10.9

Here are data for two variables. Draw the line of regression and indicate the distances, the sum of whose squares is minimised by the choice of the line of regression.

x	11	12	13	14	15	16	17
y	21	43	31	34	29	55	33

Example 10.10

The following data represent the volume in cubic cm and weight in grams of a certain fruit studied by a biologist.

Volume (x)	223	236	242	226	223	221	233	222	222	218	232	223
Weight (y)	165	171	173	170	168	172	168	167	162	166	164	164

Obtain the least-squares regression line of y on x as well as the regression line of x on y. Use the model to predict the weight of a 230-cubic-cm fruit. Also, predict the volume of a 168 g fruit.