

Applied Statistics

Additional Exercises - Descriptive Statistics

Revised edition

Part I contains exercise assignments only. Part II contains checked solutions, frequency tables, numerical results, and all requested graphs.

Calculation conventions

- Raw-data quartiles use the median-of-halves (Tukey hinges) convention.
- Variance and central moments use the population denominator n .
- Grouped location measures use linear interpolation within the relevant class.
- Kurtosis is reported as excess kurtosis (normal distribution = 0).

Part I

Exercise Assignments

Complete each task on paper and verify calculations in RStudio where appropriate.

Exercise 2.1 - Simple Classification and a Complete Descriptive Analysis

The following measured values are given:

105, 95, 100, 98, 105, 110, 95, 98, 112, 120, 105, 105, 110, 100, 98, 100, 112, 110, 110, 105, 105, 98, 100, 100, 110, 105, 110, 112, 98, 100, 105, 105.

- construct the ordered data set;
- construct a frequency table and a frequency distribution table;
- draw the absolute-frequency polygon n_i , relative-frequency line graph f_i , cumulative-frequency bar chart N_i , and cumulative relative-frequency step graph F_i ;
- calculate all principal measures of location, variability, skewness, and kurtosis.

Exercise 2.2 - Quality of Concrete Panels

Numbers of defective panels observed over 50 weeks:

14, 16, 11, 10, 8, 13, 12, 14, 16, 12, 15, 13, 12, 10, 16, 12, 17, 9, 12, 12, 14, 18, 15, 13, 17, 9, 13, 11, 11, 12, 15, 13, 14, 13, 13, 8, 10, 15, 11, 11, 14, 14, 11, 9, 13, 10, 16, 15, 13, 12.

- construct the absolute-frequency distribution and display it as a histogram;
- calculate the mean, mode, median, quartiles, variance, standard deviation, mean deviation, range, interquartile range, coefficient of variation, skewness, and kurtosis.

Exercise 2.3 - Rope Strength

Rope strength measurements [MPa]:

302, 310, 312, 310, 313, 318, 305, 309, 301, 309, 310, 311, 307, 308, 311, 300, 310, 308, 310, 307, 313, 299, 315, 312, 310, 308, 314, 333, 305, 310, 309, 314.

- construct the ordered data set;
- calculate the same descriptive statistics as in Exercise 2.2;
- draw a boxplot.

Exercise 2.4 - Weights of Fir Seed Samples

Weights [g] of 40 fir seed samples:

4.20, 4.54, 4.37, 4.47, 4.29, 4.52, 4.38, 4.21, 4.53, 4.60, 4.56, 4.70, 4.59, 4.73, 4.61, 4.31, 4.42, 4.73, 4.37, 4.64, 4.46, 4.66, 4.50, 4.71, 4.32, 4.69, 4.39, 4.59, 4.34, 4.70, 4.70, 4.28, 4.41, 4.63, 4.80, 4.61, 4.44, 4.85, 4.76, 4.79.

- classify the data into $k = 7$ class intervals;
- construct a frequency distribution table;
- draw a bar chart of N_i and a step graph of F_i .

Exercise 2.5 - Pin Diameters

Diameters of 65 pins, in hundredths of a millimetre:

1002, 1004, 1006, 1011, 1003, 1018, 996, 993, 994, 1014, 1002, 1004, 1017, 1006, 1015, 1005, 1011, 1002, 997, 1006, 1012, 998, 994, 993, 1012, 1014, 1012, 1004, 981, 1018, 1014, 1012, 997, 1014, 996, 1005, 982, 1018, 987, 997, 1020, 1017, 988, 1012, 1003, 1014, 987, 993, 995, 1011, 1012, 1004, 994, 996, 1002, 995, 1003, 1004, 1003, 1006, 1003, 1011, 1002, 996, 994.

- classify the data into $k = 5$ class intervals;
- construct a frequency distribution table;
- draw a step graph of F_i ;
- determine the grouped empirical distribution function.

Exercise 2.6 - Computer Dealers by Number of Units Sold

Grouped classification by systems sold in the last quarter:

Class I_i	50-150	150-250	250-350	350-450	450-550	550-650	650-750	750-850
n_i	8	9	7	13	11	75	12	8

- construct a frequency distribution table;
- calculate the grouped mean, mode, median, quartiles, variance, standard deviation, mean deviation, coefficient of variation, skewness, and kurtosis.

Exercise 2.7 - Schools by Pupils per Class

Grouped distribution of schools by the average number of pupils per class:

Class I_i	16-18	18-20	20-22	22-24	24-26	26-28	28-30
n_i	6	10	22	16	10	4	2

- construct a frequency distribution table;
- draw a bar chart of N_i ;
- calculate the same grouped descriptive statistics as in Exercise 2.6.

Part II

Checked Solutions

Verification summary

All source results were independently recomputed. The only numerical correction is the grouped mean deviation in Exercise 2.6: $d = 137.6498$ (not 137.6478). Minor differences in final digits may result from rounding.

Exercise 2.1 - Solution

Ordered data set

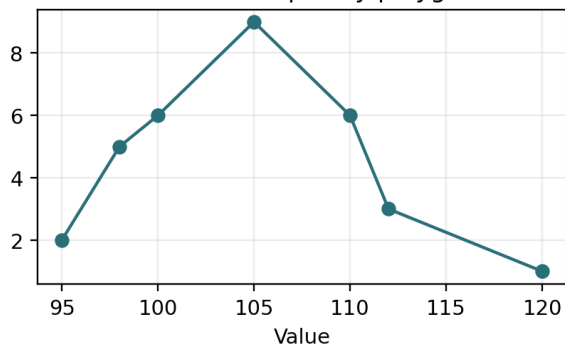
95, 95, 98, 98, 98, 98, 98, 100, 100, 100, 100, 100, 100, 105, 105, 105, 105, 105, 105, 105, 105, 110, 110, 110, 110, 110, 112, 112, 112, 120.

x_i	95	98	100	105	110	112	120
n_i	2	5	6	9	6	3	1
f_i	0.0625	0.1562	0.1875	0.2812	0.1875	0.0938	0.0312
N_i	2	7	13	22	28	31	32
F_i	0.0625	0.2188	0.4062	0.6875	0.8750	0.9688	1.0000

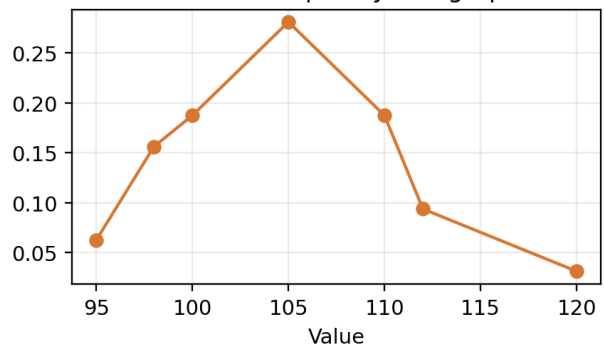
Checked numerical results

$\bar{x} = 104.4062$; $\hat{x} = 105$; $\tilde{x} = 105$; $Q_1 = 100$; $Q_3 = 110$; $S^2 = 34.5537$; $S = 5.8782$; $d = 4.8301$; $R_v = 25$; $RQ = 10$; $V = 0.0563$; $g_3 = 0.4101$; $g_4 = -0.3481$.

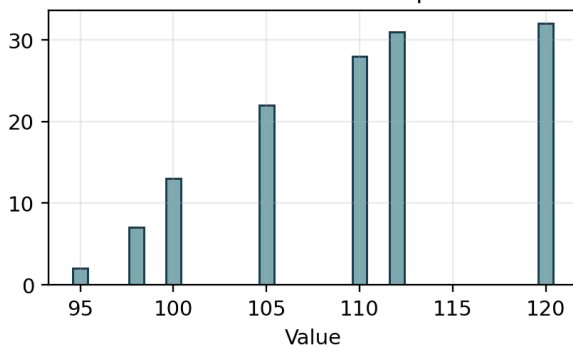
Absolute-frequency polygon



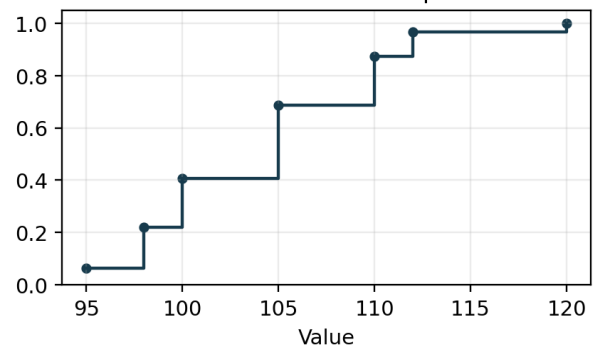
Relative-frequency line graph



Cumulative absolute frequencies



Cumulative relative frequencies



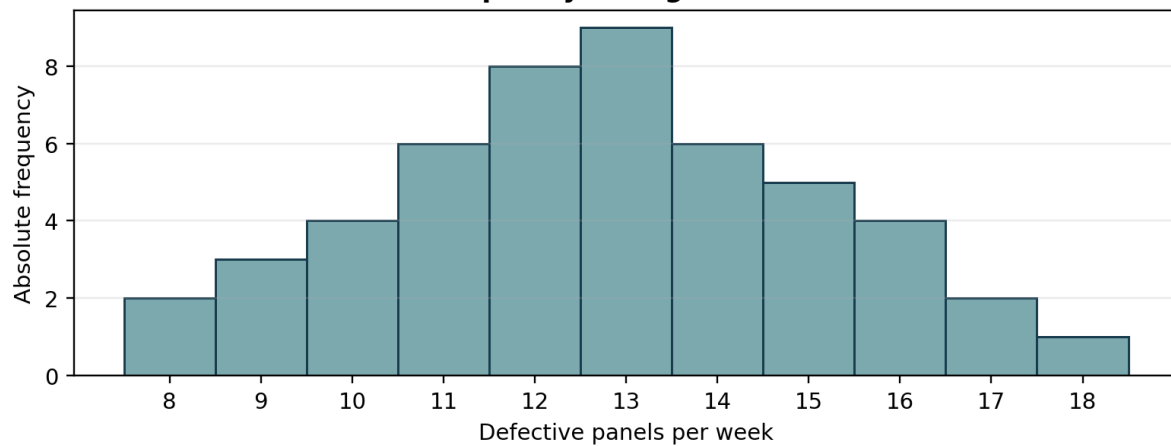
Exercise 2.2 - Solution

x_i	8	9	10	11	12	13	14	15	16	17	18
n_i	2	3	4	6	8	9	6	5	4	2	1

Checked numerical results

$\bar{x} = 12.7400$; $\hat{x} = 13$; $\check{x} = 13$; $Q_1 = 11$; $Q_3 = 14$; $S^2 = 5.6324$; $S = 2.3733$; $d = 1.9208$; $R_v = 10$; $RQ = 3$; $V = 0.1863$; $g_3 = 0.0322$; $g_4 = -0.5407$.

Exercise 2.2 - Absolute-frequency histogram



Exercise 2.3 - Solution

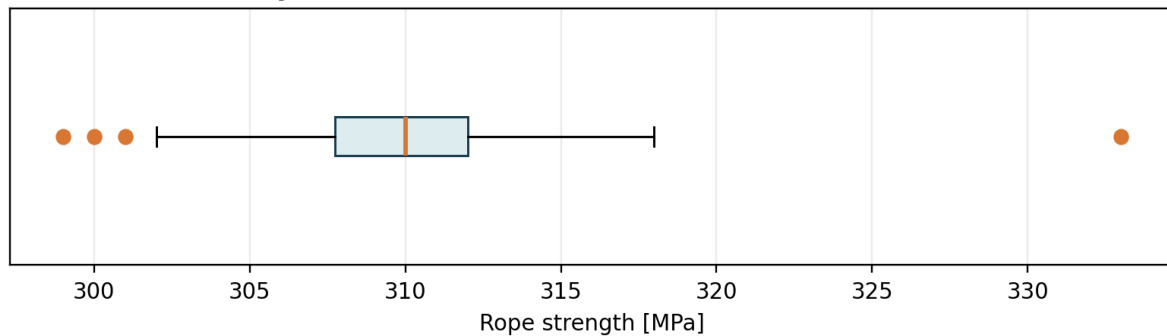
Ordered data set

299, 300, 301, 302, 305, 305, 307, 307, 308, 308, 308, 309, 309, 309, 310, 310, 310, 310, 310, 310, 311, 311, 312, 312, 313, 313, 314, 314, 315, 318, 333.

Checked numerical results

$\bar{x} = 309.7812$; $\hat{x} = 310$; $\tilde{x} = 310$; $Q_1 = 307.5$; $Q_3 = 312$; $S^2 = 35.1709$; $S = 5.9305$; $d = 3.7461$; $R_v = 34$; $RQ = 4.5$; $V = 0.0191$; $g_3 = 1.4742$; $g_4 = 5.3448$.

Exercise 2.3 - Boxplot

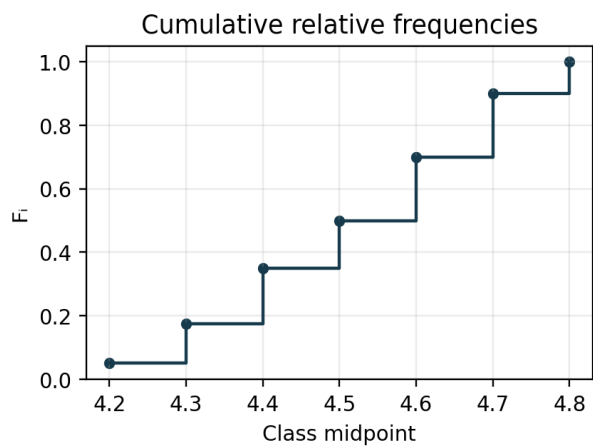
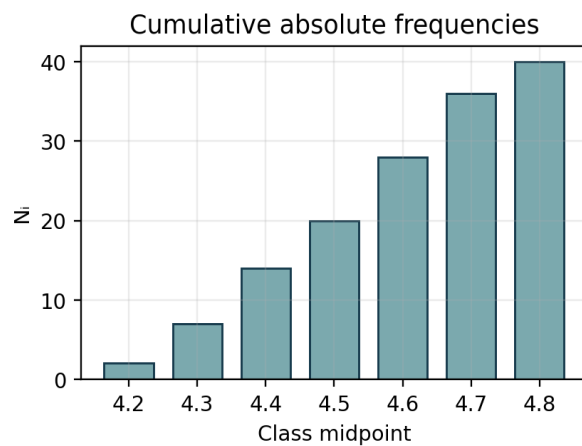


Under the 1.5 IQR rule, 333 MPa is plotted as a high outlier.

Exercise 2.4 - Solution

Using right-closed classes $l_i = (t_{i-1}, t_i]$, the class width is $h = 0.10$.

i	l_i	z_i	n_i	f_i	N_i	F_i
1	(4.15,4.25]	4.2	2	0.050	2	0.050
2	(4.25,4.35]	4.3	5	0.125	7	0.175
3	(4.35,4.45]	4.4	7	0.175	14	0.350
4	(4.45,4.55]	4.5	6	0.150	20	0.500
5	(4.55,4.65]	4.6	8	0.200	28	0.700
6	(4.65,4.75]	4.7	8	0.200	36	0.900
7	(4.75,4.85]	4.8	4	0.100	40	1.000

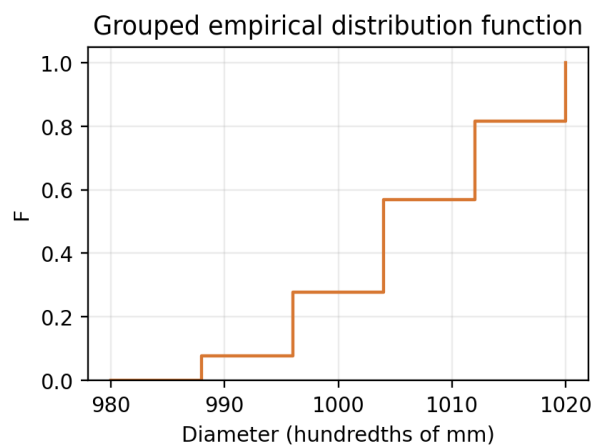
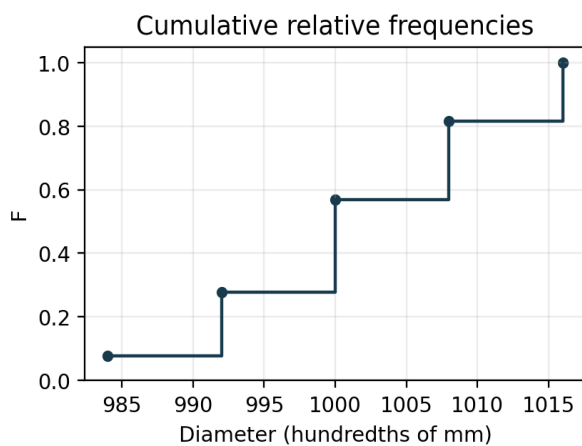


Exercise 2.5 - Solution

i	I_i	z_i	n_i	f_i	N_i	F_i
1	(980,988]	984	5	0.0769	5	0.0769
2	(988,996]	992	13	0.2000	18	0.2769
3	(996,1004]	1000	19	0.2923	37	0.5692
4	(1004,1012]	1008	16	0.2462	53	0.8154
5	(1012,1020]	1016	12	0.1846	65	1.0000

Grouped empirical distribution function

Range of x	F(x)
$x < 984$	0
$984 \leq x < 992$	0.0769
$992 \leq x < 1000$	0.2769
$1000 \leq x < 1008$	0.5692
$1008 \leq x < 1016$	0.8154
$x \geq 1016$	1



Exercise 2.6 - Solution

i	I_i	z_i	n_i	f_i	N_i	F_i
1	50-150	100	8	0.0559	8	0.0559
2	150-250	200	9	0.0629	17	0.1189
3	250-350	300	7	0.0490	24	0.1678
4	350-450	400	13	0.0909	37	0.2587
5	450-550	500	11	0.0769	48	0.3357
6	550-650	600	75	0.5245	123	0.8601
7	650-750	700	12	0.0839	135	0.9441
8	750-850	800	8	0.0559	143	1.0000

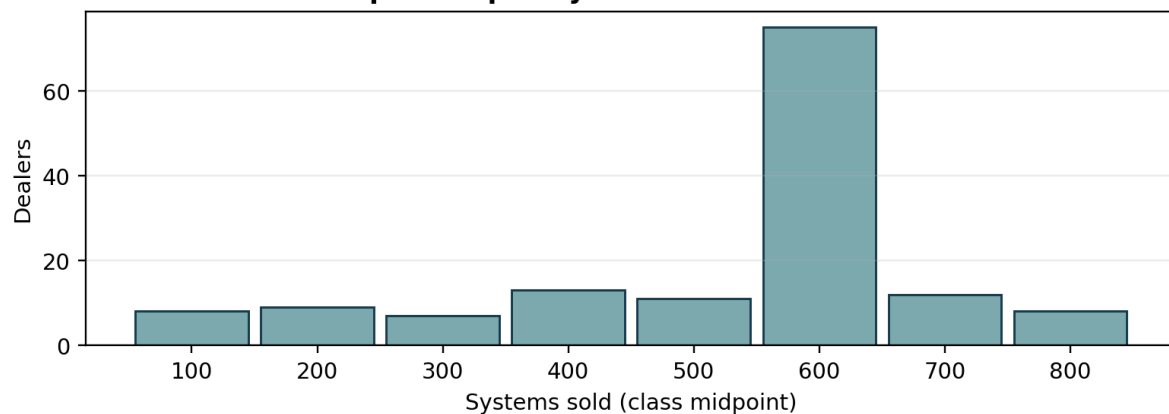
Checked numerical results

$\bar{x} = 525.8741$; $\hat{x} = 600.3937$; $\check{x} = 581.3333$; $Q_1 = 440.3846$; $Q_3 = 629.0000$; $S^2 = 30449.4107$; $S = 174.4976$; $d = 137.6498$; $V = 0.3318$; $g_3 = -1.0232$; $g_4 = 0.3534$.

Correction

The recomputed mean deviation is $d = 137.6498$. The value 137.6478 in the source differs by 0.0020 and is not explained by standard rounding.

Exercise 2.6 - Grouped frequency distribution



Exercise 2.7 - Solution

i	I_i	z_i	n_i	f_i	N_i	F_i
1	16-18	17	6	0.0857	6	0.0857
2	18-20	19	10	0.1429	16	0.2286
3	20-22	21	22	0.3143	38	0.5429
4	22-24	23	16	0.2286	54	0.7714
5	24-26	25	10	0.1429	64	0.9143
6	26-28	27	4	0.0571	68	0.9714
7	28-30	29	2	0.0286	70	1.0000

Checked numerical results

$\bar{x} = 21.9714$; $\hat{x} = 21.3333$; $\check{x} = 21.7273$; $Q_1 = 20.1364$; $Q_3 = 23.8125$; $S^2 = 8.0849$; $S = 2.8434$; $d = 2.3118$; $V = 0.1294$; $g_3 = 0.2974$; $g_4 = -0.2109$.

