Stat 202
Homework #4

1.69 Blood proteins in children from Papua New Guinea. C-reactive protein (CRP) is a substance that can be measured in the blood. Values increase substantially within 6 hours of an infection and reach a peak within 24 to 48 hours after. In adults, chronically high values have been linked to an increased risk of cardiovascular disease. In a study of apparently healthy children aged 6 to 60 months in Papua New Guinea, CRP was measured in 90 children.[28] The units are milligrams per liter (mg/l). Here are the data from a random sample of 40 of these children: CRP

0.00	3.90	5.64	8.22	0.00	5.62	3.92	6.81	30.61	0.00
73.20	0.00	46.70	0.00	0.00	26.41	22.82	0.00	0.00	3.49
0.00	0.00	4.81	9.57	5.36	0.00	5.66	0.00	59.76	12.38
15.74	0.00	0.00	0.00	0.00	9.37	20.78	7.10	7.89	5.53

(a) Find the five-number summary for these data.

(b) Make a boxplot.

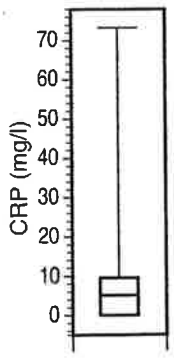
(c) Make a histogram.

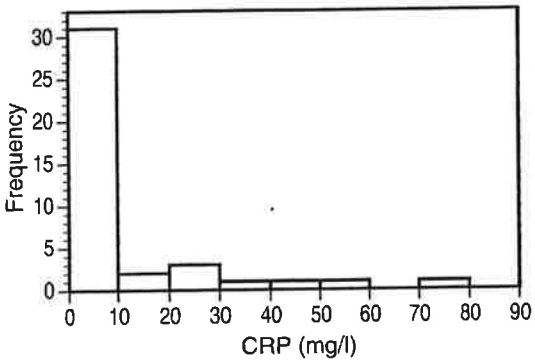
(d) Write a short summary of the major features of this distribution. Do you prefer the boxplot or the histogram for these data?

1.70 Transform the blood proteins values. Refer to the previous exercise. With strongly skewed distributions such as this, we frequently reduce the skewness by taking a log transformation. We have a bit of a problem here, however, because some of the data are recorded as 0.00 and the logarithm of zero is not defined. For this variable, the value 0.00 is recorded whenever the amount of CRP in the blood is below the level that the measuring instrument is capable of detecting. The usual procedure in this circumstance is to add a small number to each observation before taking the logs. Transform these data by adding 1 to each observation and then taking the logarithm. Use the questions in the previous exercise as a guide to your analysis and prepare a summary contrasting this analysis with the one that you performed in the previous exercise. CRP

Answers

1.69. **(a)** The five-number summary is Min = 0 mg/l, $Q_1 = 0$ mg/l, $M = 5.085$ mg/l, $Q_3 = 9.47$ mg/l, Max = 73.2 mg/l. **(b)** & **(c)** The boxplot and histogram are shown below. (Students might choose different interval widths for the histogram.) **(d)** Preferences will vary. Both plots reveal the sharp right-skew of this distribution, but because Min = Q_1, the boxplot looks somewhat strange. The histogram seems to convey the distribution better.



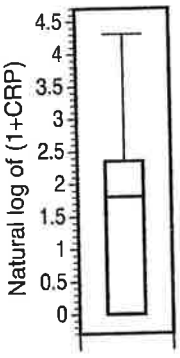


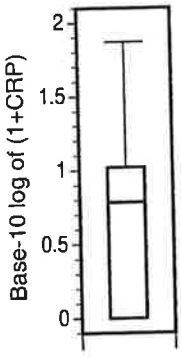
1.70. Answers depend on whether natural (base-e) or common (base-10) logarithms are used. Both sets of answers are shown here. If this exercise is assigned, it would probably be best for the sanity of both instructor and students to specify which logarithm to use.

(a) The five-number summary is:

Logarithm	Min	Q_1	M	Q_3	Max
Natural	0	0	1.8048	2.3485	4.3068
Common	0	0	0.7838	1.0199	1.8704

(The ratio between these answers is roughly $\ln 10 \doteq 2.3$.)





(b) & **(c)** The boxplots and histograms are shown below. (Students might choose different interval widths for the histograms.) **(d)** As for Exercise 1.69, preferences will vary.