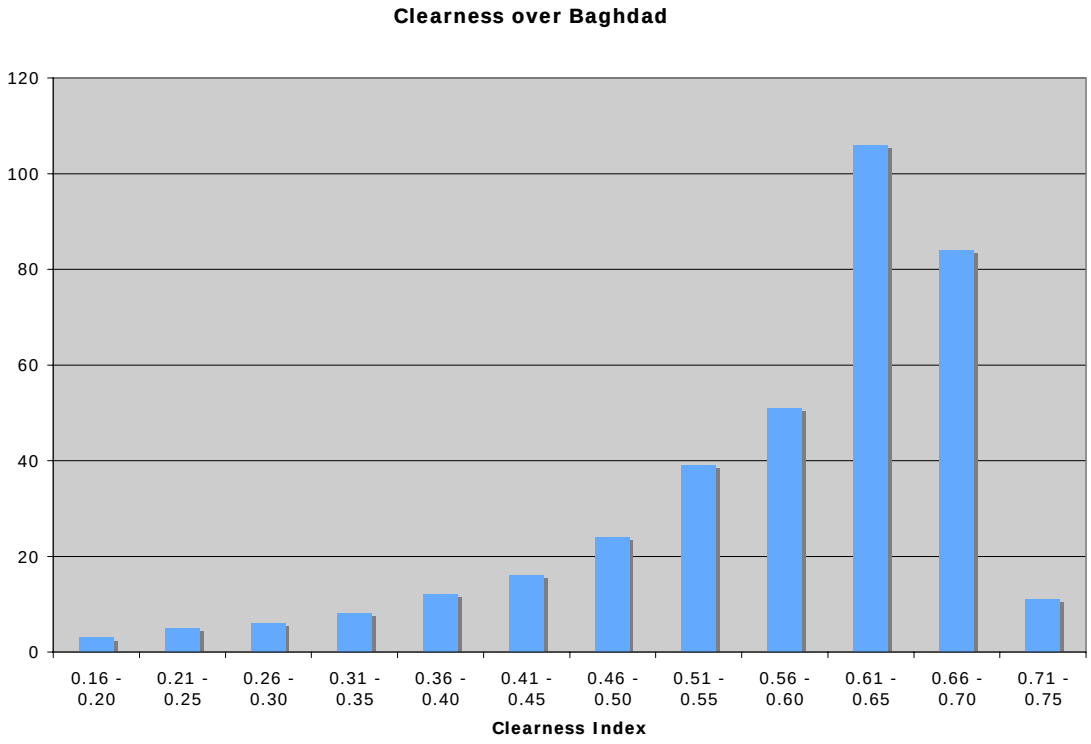
STAT 35, Interactive and Computational Probability
UCLA Statistics

http://www.stat.ucla.edu/~dinov/courses_students.html

SOLUTIONS TO HOMEWORK 1

Question 1_1



The data look as though they approximately follow a Beta distribution, with parameter values alpha=5 and beta=3. This model was used to compute the probabilities in the final column of the table:

Class interval for the clearness index	Number of days	Relative Frequency	Cumulative Relative Frequency	Model Probabilities (beta distribution)
0.16 - 0.20	3	0.0082	0.0082	0.004
0.21 - 0.25	5	0.0137	0.0219	0.01
0.26 - 0.30	6	0.0164	0.0384	0.017
0.31 - 0.35	8	0.0219	0.0603	0.028
0.36 - 0.40	12	0.0329	0.0932	0.052
0.41 - 0.45	16	0.0438	0.1370	0.059
0.46 - 0.50	24	0.0658	0.2027	0.074
0.51 - 0.55	39	0.1068	0.3096	0.136
0.56 - 0.60	51	0.1397	0.4493	0.129
0.61 - 0.65	106	0.2904	0.7397	0.155
0.66 - 0.70	84	0.2301	0.9699	0.16
0.71 - 0.75	11	0.0301	1.0000	0.12

Cloudy days are those with clearness index< 0.35. In our sample, 8 days were like this, which is 6% of our sample. The Beta distribution we are using predict cloudy days 5.5% of the time, which is a very good approximation. Clear days are those with clearness index of at least 0.66. In our sample, 95 days i.e. 26% of them were clear.

Question 1_2

The following statistics may be calculated by Webstat:

Mean: 370.69
Median: 369.50

The two differ by 1.19, which is slight. If one looks at a graph of this data, it does not appear to be very skewed or have any serious outliers. However, if we increase the largest time enough, there will be a large difference between the Mean and the Median. One can use trial-and-error to find the point where the largest observation is so high that the sample median is only half the sample mean … it happens when the largest observation of 424 is changed to 10,000.

Question 1_3

The five-number summary lists the following features of the sample:

- the minimum
- the first quartile
- the median
- the third quartile
- the maximum

(See slide #29 of Chapter 1 of the lecture notes.)

Software gives the follow results for our sample of 7 observations:

14.5 25.6 66.3 69.8 76.2

The following measures of spread may be computed using the Webstat software:

Variance: 582.5
Standard Deviation:24.14
Mean-absolute-deviation:22.61

Question 1_4

-All of the outcomes are equally likely. This means that the long-run relative frequencies should all be equal. But the short-run relative frequencies will differ somewhat from the probability (which is the long-run relative frequency) and they will differ from each other.

-We may underestimate the variability that will occur when we have a small number of observations in our experiment. Some numbers will turn up more than others just by chance.

-Because even and odd numbers are equally likely to turn up in our experiment, in the long run their relative frequencies will be the same. But the results we see in the short run are random, and will differ somewhat from person to person. These differences get smaller as we accumulate more and more observations i.e. as we approach the long run.

-Thus if we did 1000 experiments instead of just 10, the differences between the relative frequencies of the different numbers, and of odd and even numbers from one student to the next, will start to melt away. This is a result of the Law of Large Numbers.