

Descriptions of Data. (Chapter 11, parts I, II, and III)

The purpose of this handout is to illustrate the various ways of summarizing data and to give you practice in thinking what each method tells you about the data and how the various summaries are related.

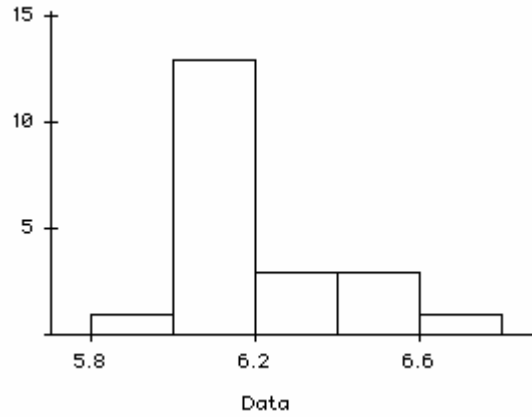
Go through the following steps to investigate each dataset and summaries. Three datasets are provided, with fairly complete summaries and the histograms for six other datasets are also given.

1. From the data, first I sorted it to make some things easier.
2. Then I made the histogram. Can you see that it gives a picture of the spread of the data? (Excel doesn't make histograms. I used an add-on to Excel that is not on our lab computers. If you want a copy of that add-on for Excel ask me for it by email.)
3. Then I computed the
 - a. sample mean,
 - b. sample median,
 - c. sample range, and
 - d. sample standard deviation.You could do any of these by hand, but also with technology. It is particularly convenient to use technology to compute the standard deviation. For these data, check all of these except the standard deviation by computing them by hand.
4. I chose not to have the spreadsheet compute the coefficient of variation because it is so easy to do by hand and you need the practice to remember the definition of it. (Coeff. of variation = standard deviation divided by the mean.)
5. Look in the text to see a picture of a normal distribution. Each of these sets of data on the last page of this handout was generated from a normal distribution, but each is a fairly small dataset, so the shape doesn't look exactly like the theoretical normal distribution.
6. In a "normal distribution" we'd expect
 - a. About 68% of the values to be within 1 standard deviation of the mean,
 - b. About 95% of the values to be within 2 standard deviations of the mean, and
 - c. About 99.7% of the values to be within 3 standard deviations of the mean.
7. For this set of data, compare how many are in each interval with those theoretical values.
 - a. Count how many are in the first interval. Then compute what percent are in that interval.
 - b. Count how many are in the second interval. Then compute what percent are in that interval.
 - c. Count how many are in the third interval. Then compute what percent are in that interval.

Dataset A:

Data	Sorted data
6.1	5.98
6.11	6.02
6.15	6.03
6.54	6.05
6.08	6.05
6.67	6.06
6.02	6.08
6.05	6.08
6.19	6.1
6.11	6.11
6.35	6.11
6.05	6.12
6.03	6.15
6.06	6.19
6.37	6.3
6.08	6.35
6.48	6.37
6.3	6.48
6.51	6.51
6.12	6.54
5.98	6.67

Histogram

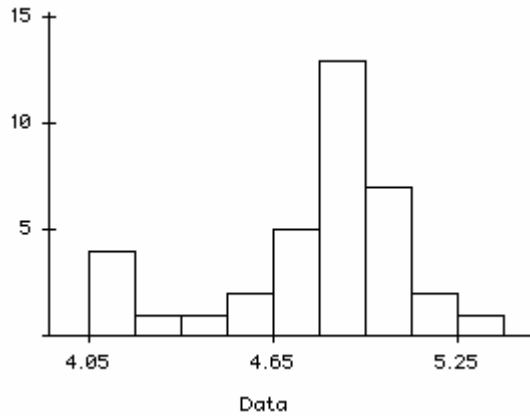


sample mean	6.207143				
sample median	6.11				
sample range	0.69				
sample stdev	0.2012		Coefficient of Variation		
		Lower limit	Upper limit	How many	Percent
within 1 stdev		6.005943	6.408343		
within 2 stdev		5.804743	6.609543		
within 3 stdev		5.603543	6.810743		

Dataset B.

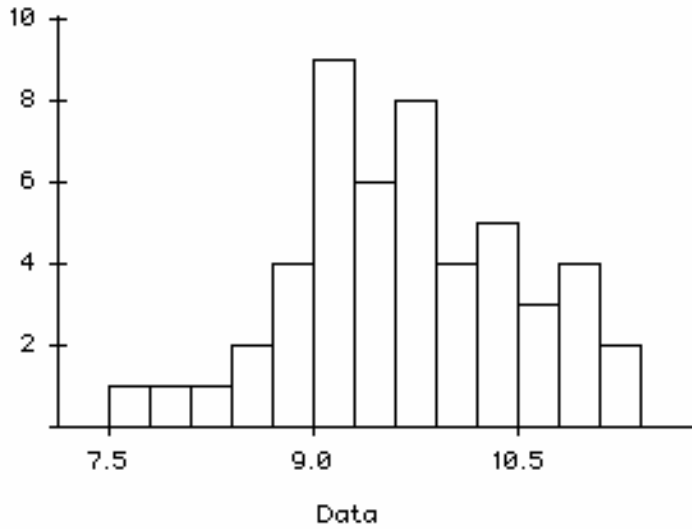
Data	Sorted data
4.569	4.065
4.998	4.115
4.709	4.126
5.122	4.199
4.795	4.281
4.828	4.363
5.028	4.539
4.065	4.569
4.823	4.688
4.199	4.709
4.881	4.776
4.798	4.795
4.99	4.798
5.041	4.8
4.948	4.811
4.115	4.823
4.884	4.828
5.016	4.832
4.87	4.851
4.832	4.87
4.949	4.878
4.851	4.881
4.281	4.884
4.963	4.911
4.539	4.948
4.126	4.949
5.039	4.963
4.811	4.99
4.688	4.998
4.776	5.016
5.271	5.028
4.878	5.039
4.8	5.041
4.911	5.122
4.363	5.209
5.209	5.271

Histogram



sample mean	4.776833				
sample median	4.8415				
sample range	1.206				
sample stdev	0.302282		Coefficient of Variation		
		Lower limit	Upper limit	How many	Percent
within 1 stdev		4.474552	5.079115		
within 2 stdev		4.17227	5.381397		
within 3 stdev		3.869988	5.683679		

Data	Sorted Data
9.452	7.765
10.856	7.992
10.135	8.224
10.633	8.484
8.711	8.659
10.296	8.711
9.795	8.732
9.867	8.814
9.557	8.916
10.602	9.082
9.68	9.086
9.207	9.089
9.957	9.123
9.802	9.196
10.367	9.207
9.082	9.215
10.294	9.277
8.916	9.286
9.406	9.32
9.32	9.406
9.286	9.452
11.017	9.475
11.316	9.545
9.277	9.557
9.089	9.626
7.765	9.68
9.123	9.795
9.626	9.802
9.836	9.808
10.276	9.817
10.285	9.836
8.224	9.867
10.843	9.957
8.814	9.984
7.992	10.135
9.475	10.168
9.086	10.276
9.215	10.285
9.808	10.294
10.997	10.296
9.817	10.367
10.168	10.602
9.196	10.633
9.984	10.642
9.545	10.843
10.642	10.856
8.484	10.997
11.122	11.017
8.732	11.122
8.659	11.316



sample mean	9.67268				
sample median	9.653				
sample range	3.551				
sample stdev	0.825098		Coefficient of Variation		
		Lower limit	Upper limit	How many	Percent
within 1 stdev		8.847582	10.49778		
within 2 stdev		8.022485	11.32288		
within 3 stdev		7.197387	12.14797		

I generated several other sets of data, each from some normally distributed population. Here are the histograms of those datasets.

