Topic 13 - Sample Histogram

Statistics for Managers

June 3, 1999

Given a several observation on a particular random variable, one of the purposes of statistical methods is to discover certain properties of its probability mass (or density) function. The unknown probability-mass function is called a population. Sampling from the population of values provides data to make these inferences.

The most elementary statistical techniques use one or both of the following

- Data Frequency Table
- Sample Histogram

Data Frequency Table

Using a pre-determined set of interval for the values of the observations, the data frequency table lists the number, and percentage of observations that fall in each interval.

Example 1. 1992 Automobiles Available in the US Market

(The actual data of 57 cars is found at the end of this section.)

- Table 1 1992 Automobiles List of Variables

Variables:
Country of Origin
Length
Width
Height
Weight

Objective

We wish to study the population of automobile sizes. In addition, we wish to study how the populations are different for each country of origin.

Class intervals

We will concentrate on car-length. Class intervals will be determined. For this purpose, 6 to 20 intervals can be used – we will start with six since the dataset is small (n = 57).

- Table 2 1992 Automobiles Preliminary Statistics

	Length	Width	Height	Weight
min	158.7	64.8	51.9	2025
max	221	78.1	58.9	4740
range	62.3	13.3	7	2715
range/6	10.38333	2.216667	1.166667	452.5

In the above, we have used EXCEL to determine the minimum, maximum, and the range = maximum - minimum. One sixth of the range guides us in determining the interval widths. Therefore, for car length, a width of 10 inches would be easy to work with. The first interval could start with a convenient value of 155 and end with 165.

- Table 3 Car Length Class Intervals

Midpoint Label	Car Length: Class Intervals
160	155-165
170	165-175
180	175-185
190	185-195
200	195-205
210	205-215
220	215-225

Excel denotes intervals using the BIN values that specify the upper limits. (See the BIN Table.) Note we have used "BIN" to denote the binomial distribution; however, EXCEL uses this term for a set of interval upper limits.

- Table 4 BIN limits for Car Length

Midpoint Label	BIN UPPER LIMIT
NA	155
160	165
170	175
180	185
190	195
200	205
210	215
220	225

Putting the numbers in the intervals

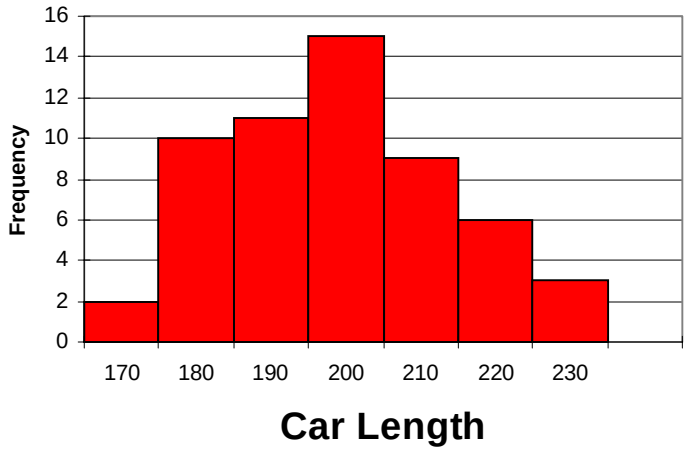
Simply place the number in its appropriate interval. For example, 195.9 would fall in the "200" interval. By convention, the interval includes all observations less than or equal to the upper limit of the interval, so 175 would fall in the "170" interval.

- Table 5 Excel Data Frequency Table

Bin	Frequency	Cumulative %
165	2	3.57%
175	10	21.43%
185	11	41.07%
195	15	67.86%
205	9	83.93%
215	6	94.64%
225	3	100.00%
More	0	100.00%

Sample Histogram

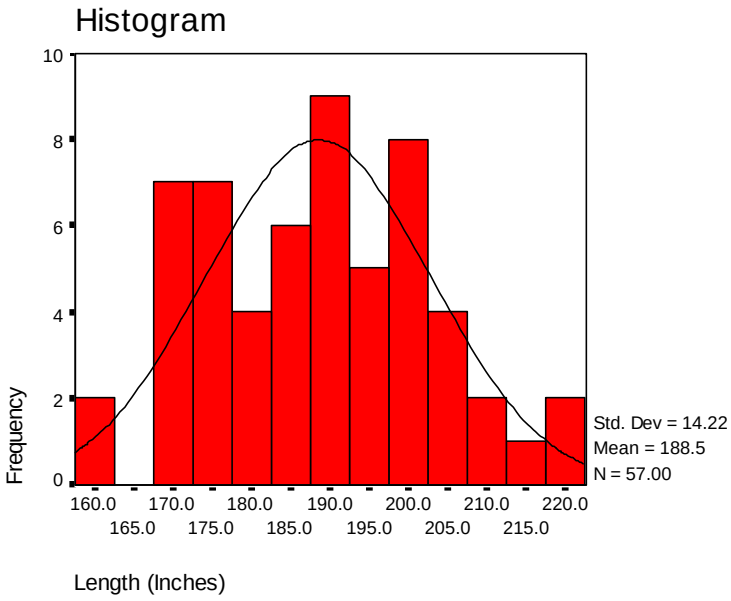
A sample histogram is a bar graph of frequencies. In a histogram, the relative area in the bar is the percentage of data in the corresponding interval.



- Figure 1 Sample Histogram for Car Length

Example 2. SPSS sample histogram

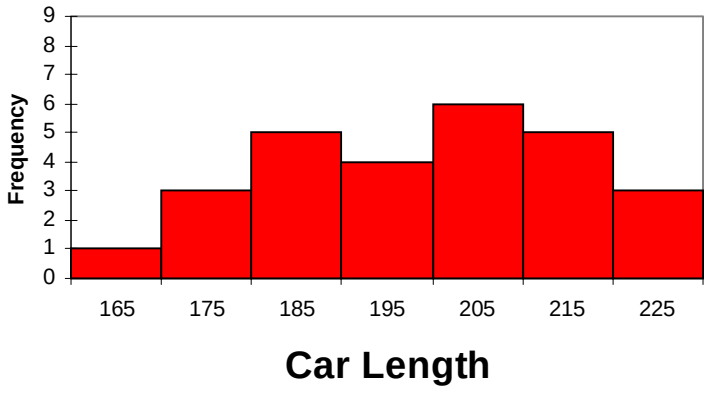
Here is a histogram of the car-length from SPSS. SPSS automatically has used intervals of five units each, and supplied the normal probability-mass function with mean 188.5 and standard deviation 14.22.



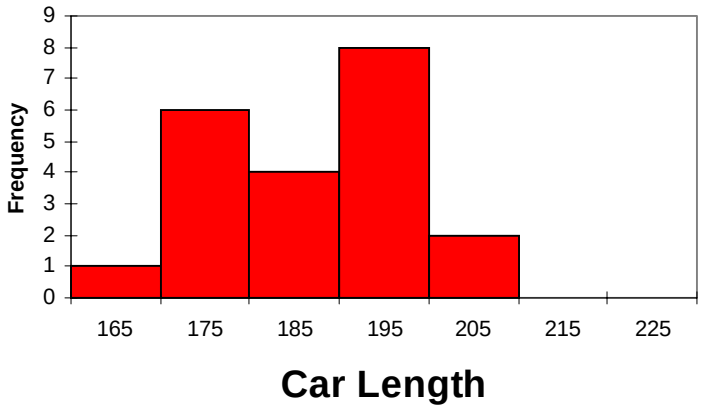
- Figure 2 SPSS Sample Histogram for Car Length with Normal Probability Mass Function

Note that the normal probability density function may provide a convenient probability model for the unknown probability-mass function.

Example 3. Comparison of Car-Lengths



- Figure 3 US Car Length Sample Histogram



- Figure 4 ASIA Car Length Sample Histogram

Table 1. Data: 1992 Automobiles

1992 Autos[a]

	Location of Manufacturer	Height (Inches)	Length (Inches)	Width (Inches)	Weight (Pounds)
1	Asia	55.1	194.9	71.3	3486
2	Asia	53.9	190.4	70.1	3212
3	Europe	56.3	192.6	70.0	3385
4	Europe	55.6	185.8	68.9	3484
5	United States	55.7	200.0	74.9	3457
6	United States	55.1	205.3	73.6	3536
7	United States	54.5	193.9	72.5	3320
8	United States	52.9	198.3	73.1	3497
9	United States	56.9	215.8	78.1	4061
10	United States	52.2	189.2	67.5	2782
11	United States	57.4	221.0	76.5	4277
12	United States	55.0	208.0	73.4	3627
13	United States	56.2	183.4	68.2	2797
14	United States	52.0	182.3	66.3	2509
15	United States	53.3	198.3	71.7	3447
16	United States	55.1	198.6	68.9	3452
17	United States	51.9	158.7	65.7	2205
18	United States	52.5	170.0	66.7	2355
19	United States	54.1	192.0	71.2	3131
20	United States	52.8	176.7	68.3	2532
21	United States	52.7	198.7	72.7	3550
22	United States	52.4	170.7	65.2	2450
23	Asia	54.7	185.2	67.1	2989
24	Asia	54.1	173.0	67.0	2480
25	Asia	54.5	171.6	65.9	2524
26	Asia	55.4	184.3	68.9	2723
27	Asia	54.9	175.0	66.7	2535
28	Asia	54.3	188.8	66.5	3576
29	Asia	56.3	199.8	71.9	3950
30	Asia	53.9	187.8	70.0	3406
31	Asia	55.3	196.7	71.7	3759
32	United States	55.6	205.1	72.7	3623
33	United States	56.9	218.9	76.9	4120
34	Asia	54.1	171.5	65.9	2487
35	Asia	54.9	193.7	70.7	3596
36	Europe	54.1	175.1	66.5	2900
37	Europe	58.9	205.2	74.3	4740
38	United States	56.9	212.4	77.8	3768
39	Asia	55.5	190.2	69.9	3505
40	Asia	55.1	187.6	69.3	3146
41	Asia	53.9	170.3	65.6	2359
42	Asia	54.1	179.9	66.9	2894
43	United States	55.7	200.4	74.1	3468
44	United States	54.8	205.5	74.6	3527
45	United States	53.5	181.2	67.3	2784
46	United States	52.7	171.7	67.3	2617
47	United States	55.5	200.6	73.6	3362
48	United States	53.1	186.9	68.6	2777
49	Europe	56.1	184.5	66.5	2779
50	United States	52.5	176.3	67.6	2437
51	Asia	53.5	178.9	66.5	3155
52	Asia	54.9	176.8	65.4	2635
53	Asia	55.1	187.8	69.7	3285
54	Asia	54.5	171.5	65.2	2343
55	Asia	53.2	161.8	64.8	2025
56	Europe	55.7	172.6	66.1	2330
57	Europe	55.5	191.7	69.3	3460
Total N	57	57	57	57	57

a. Limited to first 100 cases.