

Chapter 1

1.2 Yes. The set of all students who are taking a statistics course in this semester can be viewed as the population in a study whose conclusion is limited to those currently enrolled in the course. However, it also can be viewed as a sample in a study whose conclusion involves an inference about all students currently attending the university.

1.4 a. Frequency and relative frequency distribution of owner occupied housing in Monroe County

Class	$f_i$	$f_i/N$
less than 25,000	513	0.033
25,000- 99,999	12173	0.773
100,000-199,999	2642	0.168
200,000-299,999	331	0.021
300,000-399,999	65	0.004
400,000+	24	0.002
Column Sum	15748	1.000

b. These data appear to be a census.

1.6 Quarterbacks (1-19) tend to weigh less than running backs (20-49) who tend to weigh less than offensive linemen (50-79). Although not exact, the assignment of jersey numbers will tend to correspond to a number line showing weight.

1.8 When we are comparing profitability of U.S. corporations among different major business types, we are conducting a cross sectional analysis. When we are dealing with a comparison of profitability of U.S. corporations from 1983 to 1994, we are involved in a time series analysis.

1.10 Answer depends on the data source used by the student.

1.12 Pension and life insurance needs must be assigned to a living person who has a known age. The average age for deaths in a region in a given year tells us nothing about life expectancies at that age. For example, if the average dead age was 69, we cannot conclude that the expected life of a newborn is 69 years and that the expected life of a 68 year old is one more year. To see the absurdity of this method consider the implied expected life of a 70 year old; it would be a negative one year. If we want to know the

expected life of a 70 year old we need to know of those who were 70 say 40 years ago how many were alive 39 years ago, 38 years ago, 37 years ago, and so to the present. These years lived by this 70 year old cohort can then be averaged. But if life expectancies are rising then even this method of calculation will not be an accurate representation of the life expectancy of some one who is currently 70 years old. There may be no method of sampling that can give an unbiased estimator of expected life at specific ages.

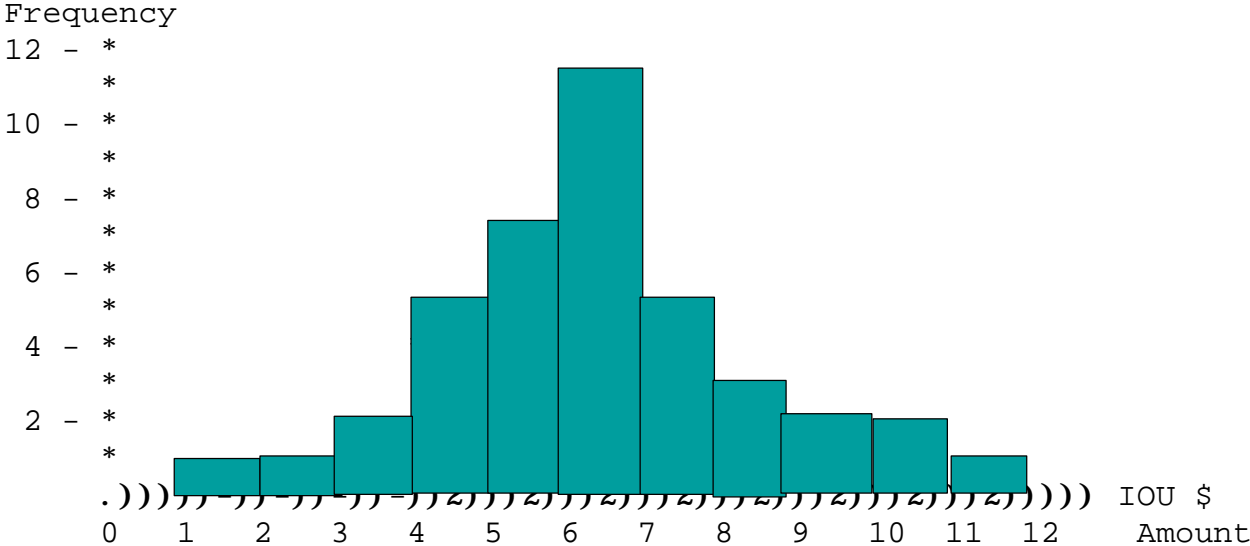
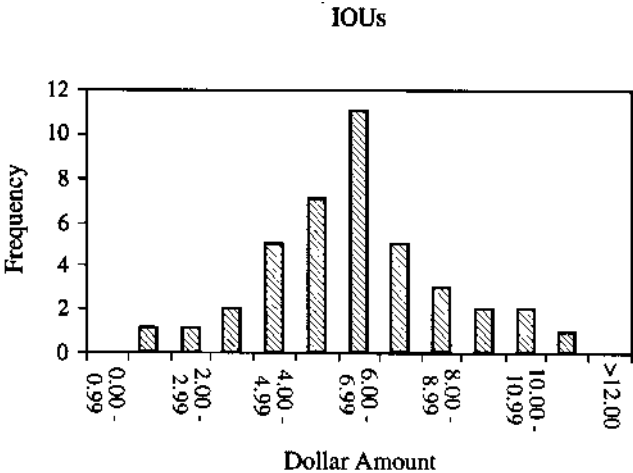
1.14 She was suggesting that more than 1 percent of those 30 percent who refused the interview may have had sex exclusively with other men in the past ten years. Thus, the study by the Battelle Human Affairs Research Center has understated the truth because of this sample selection bias. Also, Kinsey's sample was for 3 years and not exclusive.

1.16 a. The frequency distribution produced with the EXCEL "Histogram" dialog routine is

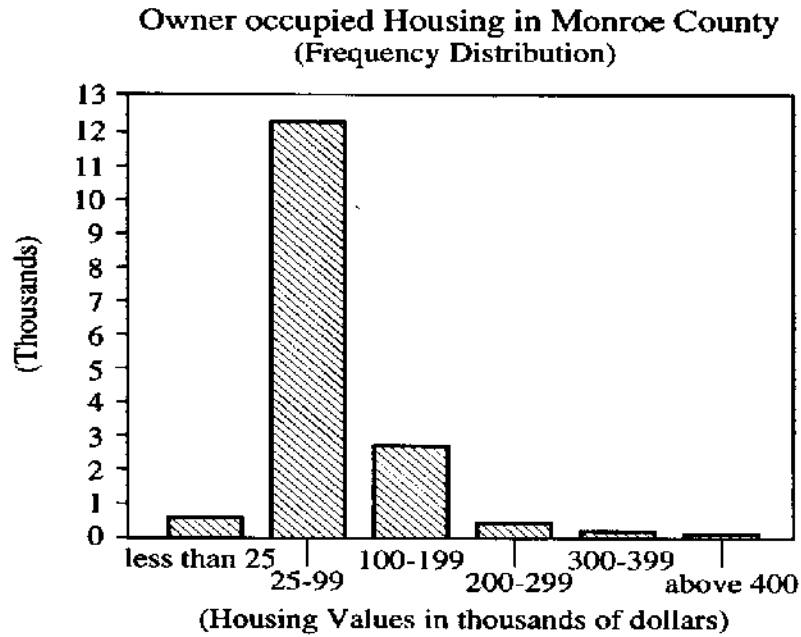
Class	Dollar Amount	Frequency	Relative Frequency	Cumulative Relative Frequency
1	0.00 - 0.99	0	0	0
2	1.00 - 1.99	1	0.0250	0.0250
3	2.00 - 2.99	1	0.0250	0.0500
4	3.00 - 3.99	2	0.0500	0.1000
5	4.00 - 4.99	5	0.1250	0.2250
6	5.00 - 5.99	7	0.1750	0.4000
7	6.00 - 6.99	11	0.2750	0.6750
8	7.00 - 7.99	5	0.1250	0.8000
9	8.00 - 8.99	3	0.0750	0.8750
10	9.00 - 9.99	2	0.0500	0.9250
11	10.0 - 10.99	2	0.0500	0.9750
12	11.0 - 11.99	1	0.0250	1.0000
13	>12.00	0	0	1.0000
Column Sum		40	1	

b. The histogram produced with the EXCEL dialog routine is given immediately below. This is not a good representation of frequency distributions in part a because the "bin ranges" are not accurately portrayed

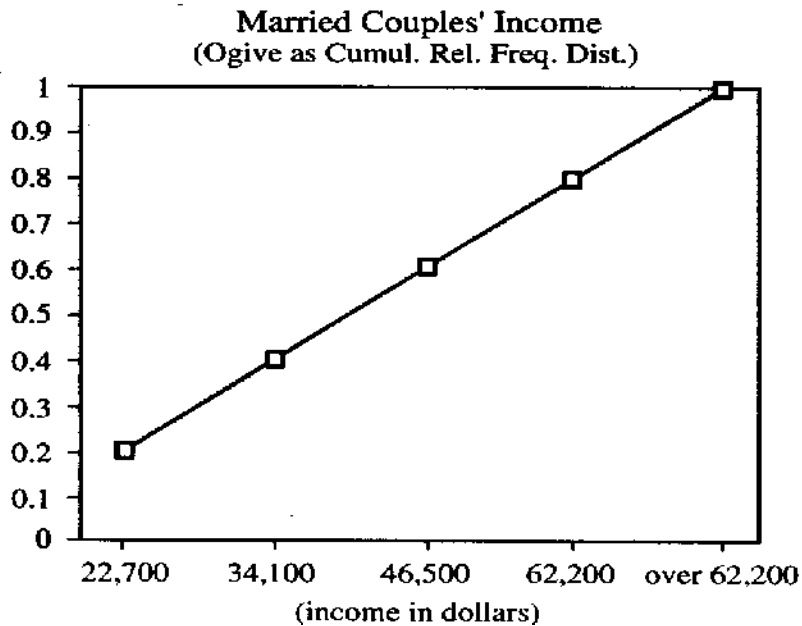
along the horizontal axis. The "frequency bars" make it look as if there are breaks in the data where there are none. (Methods for producing abutting "bars" with EXCEL are demonstrated in Chapter 4. One such graph is shown as the second diagram below.)



1.18 Histogram for owner occupied housing in Monroe County



1.20 Ogive for married couples income



1.22 If the intent is to present the number of owners in each group or class, then the histogram would be better because

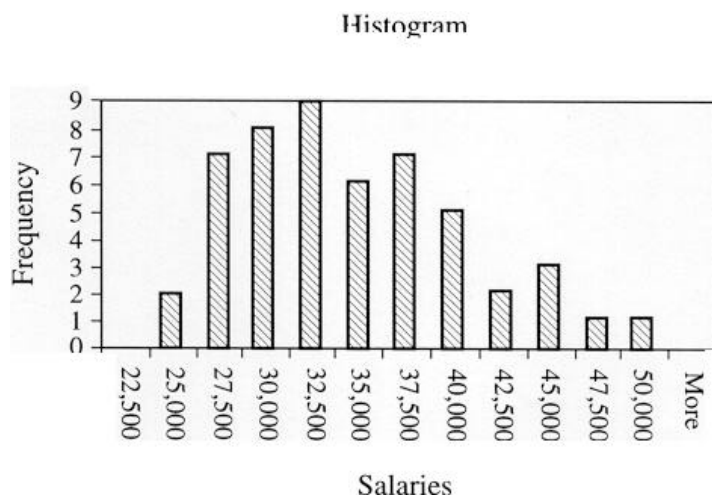
it shows absolute frequency (the number of owners) in each group. The ogive is better for visualizing the portion of the distribution below a given housing value.

1.24 Because of rounding error the relative frequencies below do not sum to exactly 100 percent; thus, the percentage in each classification may differ by a tenth of a percentage point. As calculated below, the "just to get by" group is about 60 percent of the families. The percentage for "live in reasonable comfortable" cannot be determined directly because \$40,000 is within the \$23,000 to \$50,000 class. If data are evenly distributed within this class, however, then about 39.5 percent of the families are living in reasonable comfort. Only 5.8 percent are fulfilling their dreams.

INCOME	REL.FREQ.			
Under 5	4.5			
5<10	9.7			
10<15	9.2	Just to	Live in	
15<25	16.9	get by	reasonable	
25<35	14.7	14.7	comfort	
35<50	16.3	16.3	10.8667	Fullfill
50<75	16.1	16.1	16.1000	all
75<100	6.7	6.7	6.7000	dreams
100+	5.8	5.8	5.8000	5.8
(in \$1,000)	----	----	-----	---
	99.9	59.6	39.4667	5.8

1.26 The frequency distribution and EXCEL histogram for teachers' salaries are as follows:

Salary	Frequency
<22500	0
22500 - 24999	2
25000 - 27499	7
27500 - 29999	8
30000 - 32499	9
32500 - 34999	6
35000 - 37499	7
37500 - 39999	5
40000 - 42499	2
42500 - 44999	3
45000 - 47499	1
47500 - 49999	1
>50000	0
Column Sum	51



1.28 The correct calculation of the "degree of hype" will be as follows:

Ski Area	Calculation	Real Degree of Hype
Killington	$(77 - 70)/70 \times 100 =$	10%
Sugarbush	$(53 - 46)/46 \times 100 =$	15%
Sunday River	$(36 - 29)/29 \times 100 =$	24%
Sugarloaf	$(45 - 36)/36 \times 100 =$	25%
Okemo	$(32 - 25)/25 \times 100 =$	28%
Smuggler's Notch	$(32 - 22)/22 \times 100 =$	45%

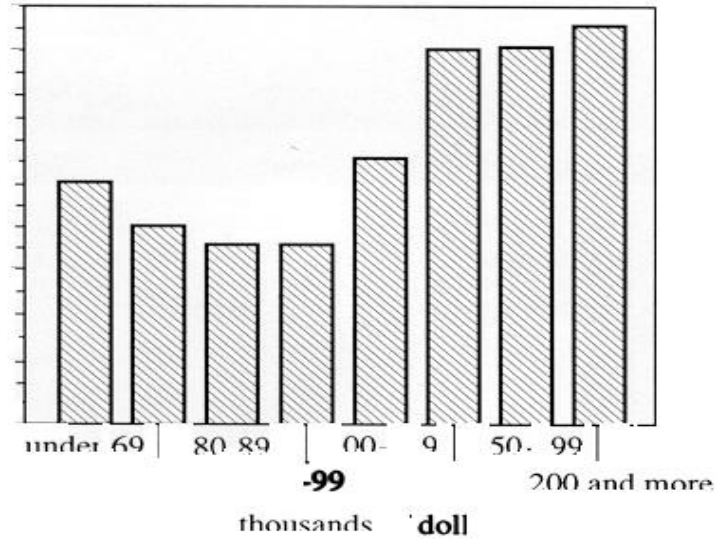
The original "degree of hype" was calculated as the difference between claimed and actual miles divided by the number of claimed miles. To obtain a better measure of "degree of hype," the difference might be divided by the actual amount of miles to obtain "hype" as a percentage of miles more than the resorts actually have.

1.30 Plaintiffs want the preselection pool to be the entire workforce because that will give the highest proportion of women or other minority. Imposing a restriction is a form of sample selection that can be used to decrease the proportion of women or other minority. If a firm does not have a high proportion of women or other minority a highly censored preselection pool might be to its advantage for comparison purposes.

1.32 You might question whether this report collected data for all female-owned businesses. If there was underreporting from the female-owned businesses in the base year, the 20 percent increase in 1992 must be an overestimated figure.

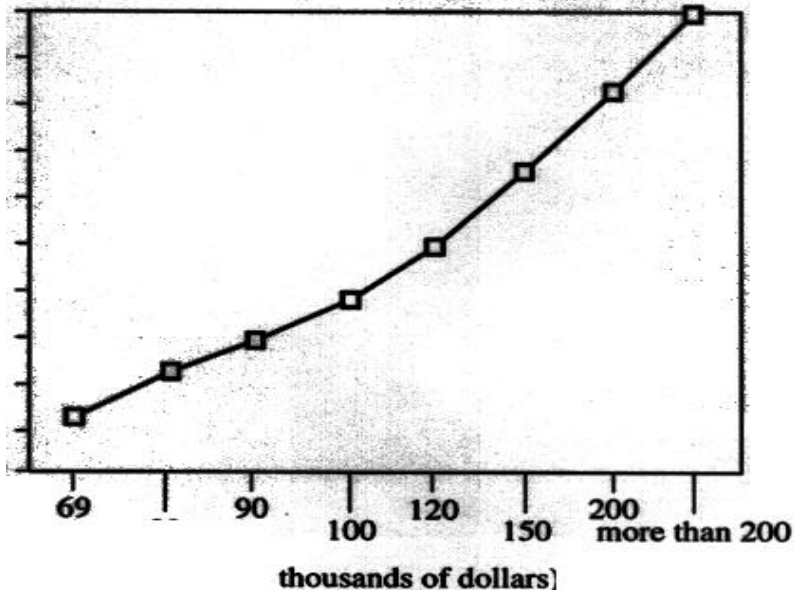
1.34 a. Histogram of housing sales by price

Housing Prices  
Relative Frequency Distribution



B. Cumulative relative frequency for home sales

Housing Prices  
(Ogives)  
Cumulative Frequency



1.36 The police are to stop only suspicious vehicles, that is, those they suspect of being in violation of a law. The trucks that are stopped are not "the typical type of truck" on the road; thus, any measures taken from these trucks could not be used to infer relative frequencies for the population of all trucks.

1.38 a. If those firms that did not respond had all responded in favor of support, the percentage could have been as high as 95 percent  $(=\frac{44}{115}\times 0.86 + \frac{115-44}{115}\times 1.00)$

b. If those firms that did not respond had all responded in disapproval of support, the percentage could have been as low as 33 percent  $(=\frac{44}{115}\times 0.86 + \frac{115-44}{115}\times 0.00)$