Topic 1: Introduction to Statistics Exercises

Q1

For each of the following variables, determine whether the variable is categorical or numerical. If the variable is numerical, determine whether the variable is discrete or continuous.

a) Number of cell phones in a household
b) Length of the longest phone call made in a month
c) Whether the household has a land line
d) Whether there is a high-speed Internet connection in the household

Q2

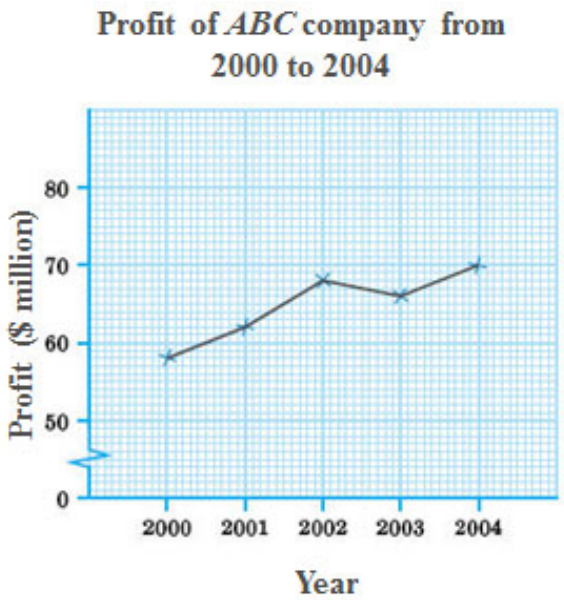
The following data is about the cost of electricity (in $) during July 2014 for a random sample of 50one-bedroom apartments in a large city.

96	171	202	179	147	102	153	197	127	82
157	185	90	116	175	111	148	213	130	165
141	149	206	175	123	128	144	168	109	167
95	163	150	154	130	143	187	166	139	149
108	119	183	151	114	135	191	137	129	158

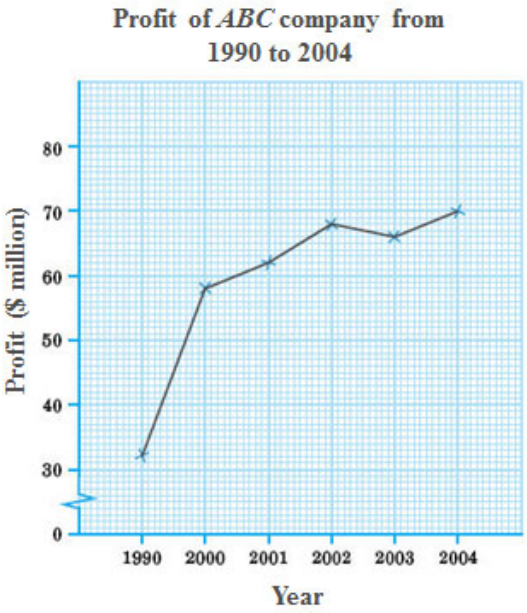
a) Construct a frequency distribution and a percentage distribution that have class intervals with the upper class boundaries $99, $119, and so on.
b) Construct a cumulative percentage distribution.
c) Construct a histogram.
d) What is the total frequency of cost to be at least $120 but less than $180?

Q3

Figure 1 below shows the profits of ABC company from 2000 to 2004. To show the company's profit from 1990-2004 to shareholders, the managing director added the profit of the company in 1990 to the graph (Figure 2).



(Figure 1)

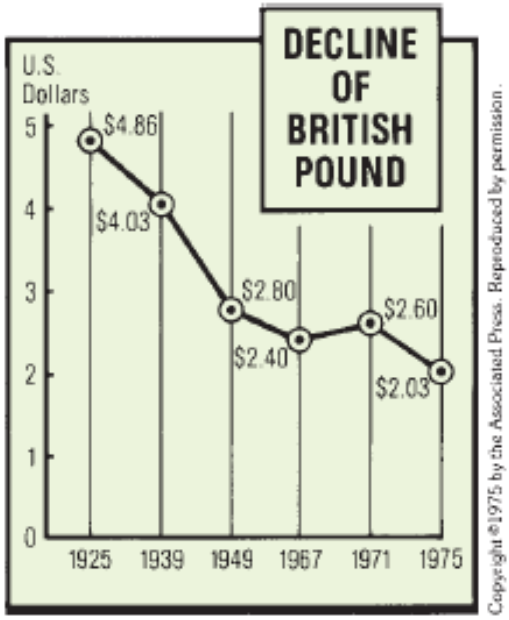


(Figure 2)

Do you think that the managing director is misleading the shareholders? Justify your answer.

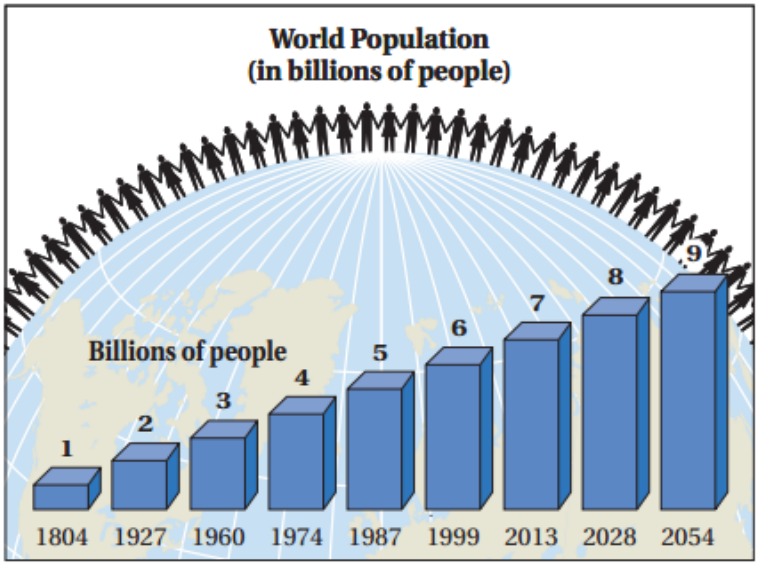
Q4

The graph below appeared in the Lexington Herald-Leader newspaper on 5th October, 1975. Discuss the correctness of this graph.



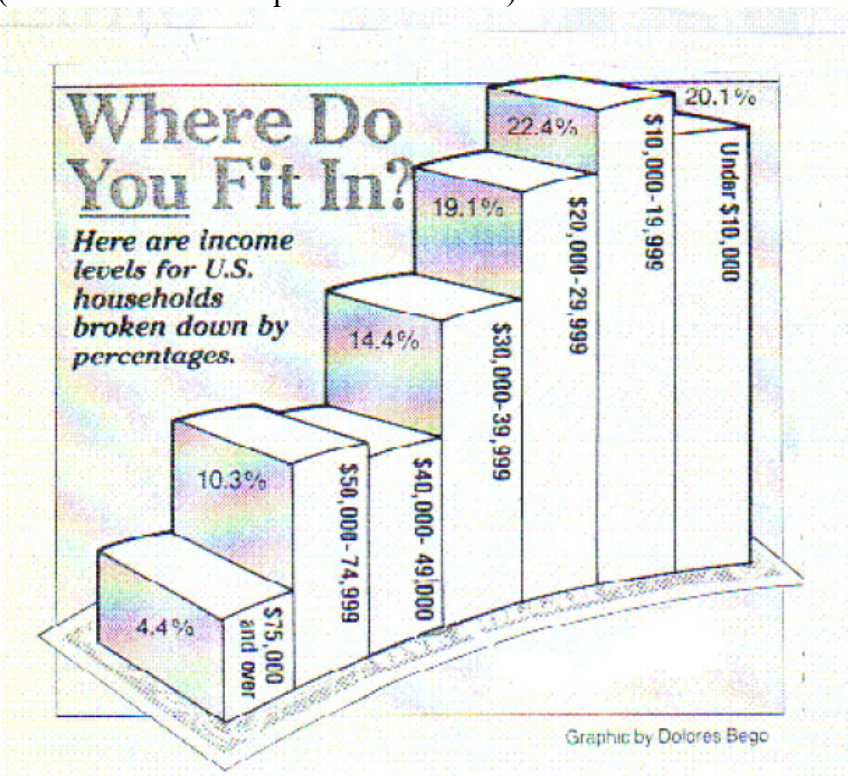
Q5

The following graph shows the world population from 1804-2054 (numbers for future years are based on United Nations projections). Critique the graph in terms of its layout, content and clarity.



Q6

The following graph shows the U.S. household income data in 1985.
(Source: The U.S. Department of Labor)



Critique the graph in terms of its layout, content and clarity.

Q7

The following is a set of data for a population of size N=10:

7 5 11 8 3 5 2 1 10 8

a) Compute the population mean.

b) Compute the population standard deviation

Q8

A food inspector, examining 10 bottles of a certain brand of honey, obtained the following percentages of impurities:

23.5 19.8 21.3 22.6 19.4 18.2 24.7 21.9 20.0 21.1

What are the mean and standard deviation of this sample?

Q9

The data contain the price for two tickets with online service charges, large popcorn, and two medium soft drinks at a sample of six theater chains:

$36.15 $31.00 $35.05 $40.25 $33.75 $43.00

a) Compute the mean and median

b) Compute the variance, standard deviation and range

c) Are the data skewed? If so, how?

d) Based on the results of (a) through (c), describe the data.

Q10

The data contains the total fat, in grams per serving, for a sample of 20 chicken sandwiches from fast-food chains. The data are as follows:

7	8	4	5	16	20	20	24	19	30
23	30	25	19	29	29	30	30	50	56

a) Compute the first quartile (Q_1), the third quartile (Q_3), and the interquartile range.

b) List the five-number summary.

c) Construct a boxplot and describe the shape.

Q11

The following data is the number of vitamin supplements sold by a health food store in a sample of 11 days:

19 19 20 20 20 22 23 25 26 27 30

a) What are the average and standard deviation of daily sale of vitamin supplements of the health food store?

b) Work out a five-number summary of the data in the sample. Comment on the distribution of the sample data.

Q12

A bank branch located in a commercial district of a city has developed an improved process for serving customers during the 12:00 to 1 p.m. peak lunch period. The waiting time in minutes (operationally defined as the time the customer enters the line to the time he or she is served) of all customers during this hour is recorded over a period of a week. A random sample of 15 customers is selected, and the results are as follows:

4.21	5.55	3.02	5.13	4.77	2.34	3.54	3.20
4.50	6.10	0.38	5.12	6.46	6.19	3.79	

Another branch located in a residential area is most concerned with the Friday evening hours from 5 to 7 p.m. The waiting time in minutes (operationally defined as the time the customer enters the line to the time he or she is served) of all customers during these hours is recorded over a period of a week. A random sample of 15 customers is selected, and the results are as follows:

9.66	5.90	8.02	5.79	8.73	3.82	8.01	8.35
10.49	6.68	5.64	4.08	6.17	9.91	5.47	

a) For each bank branch, compute the mean, median and interquartile range.
b) Form the box-and-whisker plot, and describe the shape of the distribution of waiting time at the two bank branches.
c) Compare and contrast the distributions of the waiting time at the two bank branches.