**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.

1. Number of cell phones in a household

Numeric -> discrete

1. Length of the longest phone call made in a month

Numeric -> continuous

1. Whether the household has a land line

Yes/no -> continuous

1. Whether there is a high-speed Internet connection in the household

Yes/no -> continuous

**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 |

* + 1. Construct a frequency distribution and a percentage distribution that have class intervals with the upper class boundaries $99, $119, and so on.
    2. Construct a cumulative percentage distribution.

4 7 9 13 9 5 3

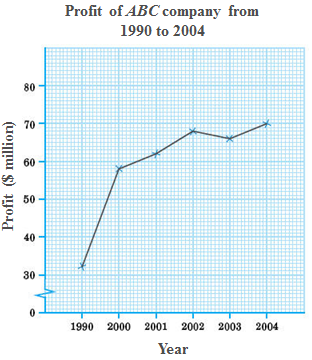
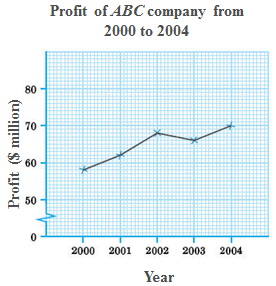
8% 22% 40% 66% 84% 94% 100%

* + 1. Construct a histogram.
    2. What is the total frequency of cost to be at least $120 but less than $180?

8 + 13 + 9 = 31 62%

**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).

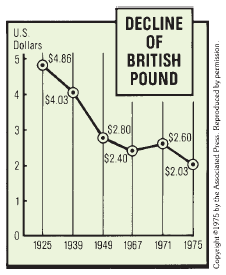


(Figure 1) (Figure 2)

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

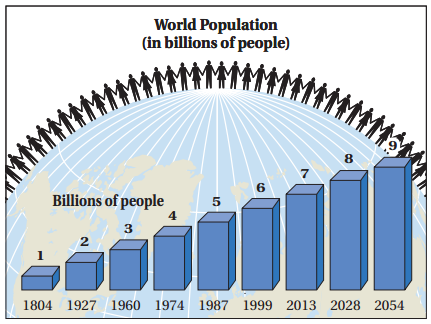
**Q4**

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



**Q5**

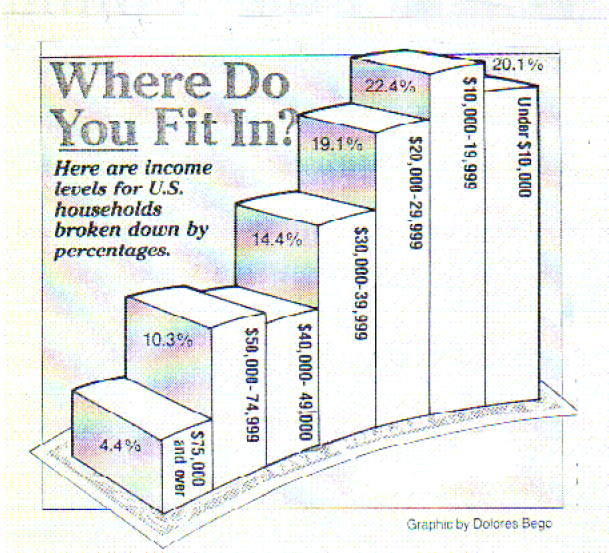
The following graph shows the world population from 1804-2054 (numbers for future years are based on United Nations projections). Critique the graph in terms of its layout, content and clarity.



**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.

Q4

X-axis: The intervals of each two consecutive years are not equal. The bin widths range from 4 years to 18 years. It gives an impression that the British pound declined steadily from year 19711975 compared with year 19391949

The line and the data: The measurement unit is not clearly stated. Say in 1925, the $4.86, does it mean US$4.86 to $1 British pounds.

Title: too subjective.

Q5

The figures of people lining the globe do not give any information about world population. And it may give the impression that future world population will be declining. In the chart, it appears that world population has been raising linearly.

Notice that the time intervals on the horizontal axis are not uniform in size.

You should expect to have the graph showing slow increase in world population at the beginning years and exponential increase after year 1960.

Q6

The 3-D display makes it difficult to read the bars. Focusing at the front of each bar, the side of each bar, or the back of each bar will give different impressions.

The x-axis goes from right to left, instead of the usual direction left to right, thereby giving a misleading perception of the asymmetry. Moreover, the curved and sloped x-axis exaggerates the difference between lower income bars and upper-income bars.

Percentage figure was missed from the bar of $40,000 $49,000. Moreover, the upper boundary of this bar should be $49,999.

The bar widths are not proportional to the interval ranges. For example, it goes by $10,000 then by $25,000, increasing the height of the $50,000-$74,999 bar. • Total number of households was not given.

**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

1. Compute the population mean.
2. 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

1. Compute the mean and median
2. Compute the variance, standard deviation and range
3. Are the data skewed? If so, how?
4. 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

1. Compute the first quartile (Q1), the third quartile (Q3), and the interquartile range.
2. List the five-number summary.
3. 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:

1. 19 20 20 20 22 23 25 26 27 30
2. What are the average and standard deviation of daily sale of vitamin supplements of the health food store?
3. 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

* 1. 6.68 5.64 4.08 6.17 9.91 5.47

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