1. Consider the following output from DataDesk when analyzing the pH values of the 1986 data collected on precipitation events.

Summary stats for 1986
NumNumeric = 55
NumNonNumeric = 0
NumCases = 55
Mean = 6.0673
Median = 6.1000
Std Deviation = 0.47339
Range = 2.4000
Minimum = 4.6000
Maximum = 7
75-th %ile = 6.4000

Which of the following is NOT CORRECT?

(a) The 25th percentile is about 5.9.
(b) Some outliers appear to be present below a pH of 5.4.
(c) About 95% of the observations have pH values in the approximate range 6 ± 1.
(d) About 10% of the values are in the range 5.8 to 6.0.
(e) About 75% of the values are less than 6.4.

2. The following is a histogram showing the actual frequency of the closing prices on the New York exchange of a particular stock.

Based on the above frequency histogram for New York Stock exchange, the class that contains the 80th percentile is:

(a) 20-30
(b) 10-20
(c) 40-50
(d) 50-60
(e) 30-40

3. A histogram of the heights of 39 plants is as follows:
The 75th percentile of the height distribution is approximately:

(a) 9.4
(b) 9.7
(c) 7.7
(d) 7.5
(e) 10.0

4. The weights of the male and female students in a class are summarized in the following boxplots:

Which of the following is NOT correct?

(a) About 50% of the male students have weights between 150 and 185 lbs.
(b) About 25% of female students have weights more than 130 lbs.
(c) The median weight of male students is about 162 lbs.
(d) The mean weight of female students is about 120 because of symmetry.
(e) The male students have less variability than the female students.
5. Consider the following box plots of the grades in a course in statistics for each sex drawn according to the convention that the whiskers reach the 10th and 90th percentiles.

![Box Plots](image)

Which of the following is correct?

(a) The mean grade of the female students is about 72.
(b) The median grade for the male students is about 68.
(c) About 25% of female students get grades above 72.
(d) About 10% of male students get grades below 60.
(e) About 50% of female students get grades between 62 and 82.

6. Consider the following box-plot of the yield of barley drawn using the convention that the whiskers reach the 10 and 90th percentiles.

![Box Plot](image)

Which of the following is NOT correct?

(a) The mean is about 200 g/400 m².
(b) the median is about 180 g/400 m².
(c) About 50% of the yields are between 160 and 220 g/400 m².
(d) About 25% of the yields are below 220 g/400 m².
(e) About 10% of the yields are below 130 g/400 m².

7. Consider the following ogive of the scores of students in an introductory statistics course:

![Ogive graph]

A grade of C or C+ is assigned to a student who scores between 55 and 70. The percentage of students that obtained a grade of C or C+ is:

(a) 25%
(b) 30%
(c) 20%
(d) 50%
(e) 15%

8. Which of the following is NOT correct about constructing histograms?

(a) The approximate number of classes is 1+3.3log(n).
(b) All class intervals should be of equal width.
(c) The bars of the histogram are centred over the class mark (midpoint).
(d) The first and last classes should be open-ended to account for extreme points.
(e) There should be no spaces between bars.

9. Forest companies routinely take samples from tracts that have been replanted to monitor the growth of the trees. Suppose that in a recent sample of two tracts, the diameter of the trees was measured with the following results:

<table>
<thead>
<tr>
<th>Tract</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trees</td>
<td>75</td>
<td>210</td>
</tr>
<tr>
<td>Range</td>
<td>232-315</td>
<td>215-250 (mm)</td>
</tr>
</tbody>
</table>
Histograms to compare the two groups are to be constructed. Which of the following is not recommended.

(a) The number of classes in Group A and Group B should be around 7 or 8.
(b) The class width of both groups will be 10 mm.
(c) The class bounds will be 232-242 mm, 242-252 mm, etc. for Group A and 215-225 mm, 225-235 mm, etc for Group B.
(d) The vertical scale for both groups should be relative frequency (%).
(e) The two histograms will be stacked and aligned so that the vertical axes are the same and the horizontal axes are identical.

10. Consider the following SAS procedure to construct a histogram.

```
PROC CHART DATA=BARLEY;
  VBAR YIELD/ TYPE=PERCENT
  MIDPOINTS=90 TO 300 BY 30;
```

Which of the following statements is correct?

(a) The first bar will contain only values from 75 to 105.
(b) All yields below 90 will not be used.
(c) The vertical axis will be labelled with the actual frequency in each class.
(d) Several charts will be produced, one for each distinct year.
(e) The last class will include values above 315.

11. For each student in a class, the sex and weight (in kilograms) are recorded. Consider the following SAS program:

```
DATA STUDENTS;
  INPUT SEX $ WEIGHT;
  DATALINES;
  F 62
  .
  .  (<-- more data here)
  .
  M 78
  ;
  PROC SORT DATA=STUDENTS;
    BY SEX;
  PROC CHART DATA=STUDENTS;
    VBAR WEIGHT/TYPE=PERCENT MIDPOINTS=55 TO 95 BY 10;
    BY SEX;
```
Which of the following statements is correct?

(a) The first class will contain weights from 55 to 65 kg.
(b) All weights below 55 kilograms will be discarded.
(c) One vertical bar chart will be produced; it will contain the weights of all of the students, and the males and females will be indistinguishable.
(d) The vertical axis will be labelled “frequency”.
(e) Two separate vertical bar charts will be produced – one for males and one for females.

12. A single stem-and-leaf plot is a useful tool because:

(a) it includes the average and the standard deviation.
(b) it shows the percentage distribution of the data values.
(c) it enables us to examine the data values for the presence of trends, cycles, and seasonal variation.
(d) it enables us to locate the centre of the data, see the overall shape of the distribution, and look to marked deviations from the overall shape.
(e) it enables us to compare this dataset against others of a similar kind.

13. Forty students wrote a Statistics examination having a maximum of 50 marks. The mark distribution is given in the following stem-and-leaf plot:

```
0|28
1|2245
2|0133358889
3|001356679
4|22444466788
5|000
```

The third quartile of the mark distribution is equal to:

(a) 75
(b) 44
(c) 32
(d) 37.5
(e) 30

14. Thirty students wrote a statistics examination having a maximum of 50 marks. The mark distribution is given in the following stem-and-leaf plot:
0|9
1|25
2|01335889
3|00136679
4|02244478
5|0

The median mark is equal to:

(a) 30.5
(b) 30.0
(c) 25.0
(d) 28.5
(e) 44.0

15. Rainwater was collected in water collectors at thirty different sites near an industrial basin and the amount of acidity (pH level) was measured. The following stem-and-leaf diagram shows the pH values that ranged from 2.6 to 6.3.

<table>
<thead>
<tr>
<th>Stems</th>
<th>Leaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>679</td>
</tr>
<tr>
<td>3</td>
<td>237789</td>
</tr>
<tr>
<td>4</td>
<td>1222446899</td>
</tr>
<tr>
<td>5</td>
<td>0556788</td>
</tr>
<tr>
<td>6</td>
<td>0233</td>
</tr>
</tbody>
</table>

The median acidity is:

(a) 4.2
(b) 4.4
(c) 4.5
(d) 4.6
(e) Average of 15 and 16.

16. Refer to the previous question. Which of the following box-plots is correct:
17. Refer to the previous question. The interquartile range is:

(a) 7.75
(b) 23.25
(c) 5.625
(d) 3.77
(e) 1.855

18. The following is a stem-plot of the birth weights of male babies born to a group of mothers who smoked during pregnancies. The stems are in units of kg.

<table>
<thead>
<tr>
<th>Stems</th>
<th>Leaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3, 4, 6, 7, 7, 8, 8, 8, 9</td>
</tr>
<tr>
<td>3</td>
<td>2, 2, 3, 4, 6, 7, 8, 9</td>
</tr>
<tr>
<td>4</td>
<td>1, 2, 2, 3, 4</td>
</tr>
<tr>
<td>5</td>
<td>3, 5, 5, 6</td>
</tr>
</tbody>
</table>

The median birth weight is:

(a) 13.5
(b) 3.2
(c) 3.5
(d) 3.7
(e) Average of 13 and 14.

19. Refer to the previous question. The first quartile (25th) percentile of the weights is

(a) 2.3
(b) 2.7
20. Which one of the following statements is **FALSE**?

(a) Pie charts are better than bar graphs for comparing relative sizes.
(b) Data that are nominal scale are presented using frequency tables.
(c) Means and standard deviation of ordinal data are meaningless.
(d) The scatter-plot is the basic graphic tool for investigating relationships between two interval or ratio scaled variables.
(e) Box-plots are a good choice for comparing the distribution of values among groups.

21. Which of the following is **NOT CORRECT**?

(a) The scatterplot is the basic graphical tool for investigating relationships between two continuous interval or ratio scaled variables.
(b) The frequency table is useful for summarizing data from a nominal scaled variable.
(c) Means and standard deviations of nominal or ordinal scaled variables are useful summary measures.
(d) Pie charts don’t perform well because people have difficulty in accurately quantifying angles.
(e) Boxplots perform well for comparing groups because it is relatively straightforward to see how the mean and median change over the groups.

22. The following is a display comparing the favorite TV shows selected from a specified set by gender. Each person had to select one preferred show from the three shows given below:

<table>
<thead>
<tr>
<th>Percent</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Size</td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Male: sssssssssssssssssfffffffbbbbbbbbbbbbbb

Female: sssssssssssssssssfffffffbbbbbbbbbbbb

| Star Trek; | Friends; | Baywatch; |

Which of the following is **FALSE**:

(a) A greater percentage of males have StarTrek as their favorite show than females.
(b) More females (in absolute number) selected Baywatch as their favorite show than males.

(c) About 1/5 of females surveyed selected StarTrek as their favorite show from the three given.

(d) The modal favorite show (from the three specified) for males is StarTrek.

(e) About twice as many females (in relative numbers) enjoy Star Trek than males.

23. An experiment was conducted to investigate the effect of a new weed killer to suppress weed germination in onion crops. Two chemicals were used, the standard weed killer (C) and the new chemical (W). Both chemicals were tested at high and low concentrations. Measurements are made on each of 50 plots for each treatment combinations of the % weed germination. Here are some box-plots of the results where the whiskers extend to the min and max of the data.

Which of the following is NOT a feature of this data:

(a) At either high or low concentrations, the new chemical (W) gives better control of weed germination than the control (C).

(b) Fewer weeds germinate at higher concentrations of both chemicals.

(c) The results from the control chemical are less variable than the new chemical.

(d) High or low concentrations of either chemical have approximately the same effects on weed germination.

(e) Some of the results from the low concentration of weed killer W have fewer weeds germinating than some of the results from the high concentration of W