

Respond to each item, giving sufficient detail. Neatly handwrite your responses. *This should be very helpful to you as you prepare for exams.*

1. The population in a statistical study is the complete set of people or things being studied. The _____ is the subset of the population from which the raw data are actually obtained.

List the 4 “common sampling methods”.

- (1) Simple random sampling
- (2) _____
- (3) _____
- (4) Stratified sampling

There are two major types of bias that can affect sample selection. They are:

- (1) Selection bias, and (2) Participation bias

Identify the type of sampling used.

- (a) 49 students are selected at random from the Sophomore class, 39 from the Junior class, and 48 from the Senior class. _____
 - (b) A tax auditor selects every 1000th income tax return that is received.

 - (c) A pollster uses a computer to generate 500 random numbers, then interviews the voters corresponding to those numbers. _____
 - (d) To avoid working late, a quality control analyst simply inspects the first 100 items produced in a day. _____
2. Identify the type of study that would be most appropriate to answer the given question. Include such terms as observational versus experimental, case-control, with or without blinding, single and double blinding)
- (a) How do lawyers’ salaries compare to doctors’ salaries?
 - (b) Does an hour of meditation per day lower blood pressure?
 - (c) Does the new medication relieve depression?
 - (d) Does caffeine cause birth defects?

3. Gas mileages (measured and rounded to the nearest mile per gallon) are shown below for various company cars.

22 23 16 24 30 17 36 22 25 27
 30 29 38 20 18 15 18 17 20 33
 24 32 20 26 28

Complete the frequency chart below for 5-point bins, starting with the given class.

Mileage (mpg)	Frequency (# of cars)	Relative Frequency	Cumulative Frequency
15 – 19			

4. Match the following statistical graphs with their definitions.

___ histogram

___ bar graph

___ scatterplot

___ line chart

___ pie chart

(A) a graph with rectangular bars of lengths usually proportional to the magnitudes or frequencies of what they represent; used for comparing two or more values; the bars can be horizontally or vertically oriented; sometimes a stretched graphic is used instead of a bar: often the categories are ranked from largest to smallest, and this graph is often used exclusively for categorical or qualitative variables

(B) a circular chart divided into sectors or wedges according to the percentage of frequencies in each category of the distribution; the arc length of each sector (and consequently its central angle and area), is proportional to the quantity it represents.

(C) a graph with data values for each category shown as a dot, and the dots are connected with line segments. For each dot, the horizontal position is the center of the bin it represents and the vertical position is the data value for the bin.

(D) this graph involves plotting points to display the relationship between two variables; the independent variable is placed on the horizontal axis, with the dependent variable on the vertical axis;

(E) a graph that displays the data by using vertical bars of various heights to represent the frequencies of a distribution; the categories (bars) must be adjacent; this graph always involves continuous variables, and either midpoints or boundaries are used on the horizontal axis

5. Match the following statistics with the most appropriate definition/procedure.

- | | |
|---------------------------------|---|
| _____ mean | (A) The middle value after a sort |
| <u> C </u> mode | (B) The maximum value minus the minimum value |
| _____ median | (C) The most commonly occurring value(s) |
| _____ range | (D) A measure of variability involving an average of the differences between each data value and the mean |
| _____ standard deviation | (E) The sum of the data values divided by the number of values |

An _____ is a data value that is much higher or much lower than almost all other values. There is a statistical formula to help determine possible outliers.

An interval with lower and upper bounds for a data set, using the quartile approach, is $(Q_1 - 1.5 \times \text{IQR}, Q_3 + 1.5 \times \text{IQR})$. The interquartile range (IQR) is $Q_3 - Q_1$. Any value outside this range of acceptability is considered **an outlier**.

6. List the 5 components of a box plot (the “five number summary”).

Low value , _____ , _____ , _____ , _____

How do you find the **first quartile**? _____

How do you find the **third quartile**? _____

A distribution is **symmetric** if its left half is a mirror image of its right half.

A distribution is **left-skewed** if its values are more spread out on the left side.

A distribution is _____ - skewed if its values are more spread out on the right side.

7. The box plot for the test scores of 32 students is given below. Use it to answer the following questions.



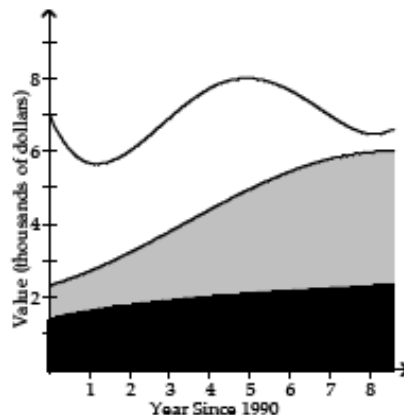
- (a) How many students made below 75? _____
- (b) How many students made below 91? _____
- (c) How many students made above 59? _____
- (d) How many students made between 59 and 91? _____ ($91 - 59 = 32$ is the interquartile range)
- (e) How many students made between 59 and 100? _____

8. Write the formula for finding the **standard deviation** for sample data.

$$\text{Standard deviation} = s = \sqrt{\frac{\sum X^2 - (\sum X)^2/n}{n-1}} \text{ or } s = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$$

Find the sample standard deviation for the following data: 32 cm, 29 cm, 30 cm, 34 cm, 28 cm
Show your work with either formula below.

9. The stack plot below shows the value of each of Danny's investments. The stack plot contains three regions. The uppermost unshaded region represents the value of Danny's investment in individual stocks. The center lightly-shaded region represents the value of Danny's investment in mutual funds. The bottom darkly-shaded region represents the value of Danny's investment in a CD. The thickness of a region at a particular time tells you its value at that time.
- (a) In which year was the total value of Danny's investments the greatest? _____
- (b) In which year was the total value of Danny's investments the least? _____
- (c) In which year was the value of Danny's investment in individual stocks the highest? _____
- (d) In 1998, about what percentage of Danny's total investment was in mutual funds? _____
- (e) In 1990, about what percentage of Danny's total investment was in stocks? _____



10. Write a few sentences describing something you learned that was new for you in class this unit. You may include a favorite activity, an interesting application, a teaching and learning technique, or a specific concept that you better understand as a result of this unit.

Do your best! Rise to the challenge! Live and learn!