Practice Midterm #1

1) A characteristic of interest to be measured for each unit is called: **Variable**

2) The numerical value that is calculated using all the units in a sample is a: **Statistic**

3) What is the systematic tendency on the part of the sampling procedure to include or exclude a certain type of unit? **Selection Bias**

4) When the population is first divided into groups and a simple random sample of units from each group is taken such that all units combined belong to the sample is called:
   a) Stratified Random Sample
   b) 1-in-k Systematic Sample
   c) Cluster Sample
   d) Volunteer sampling
   e) Convenience Sample
   f) other

5) What is the symbol that represents the mean of a population?
   a) $\mu$
   b) $\sigma$
   c) $\bar{x}$

6) What is the symbol that represents the standard deviation of a sample?
   a) $\sigma^2$
   b) $\sigma$
   c) $s$

7) Determine whether the following events are mutually exclusive:

   Imagine that you are tossing two coins.
   
   Event A = “Both coins have the same side”
   Event B = “One coin is tail”

   **No, they are not mutually exclusive since you could have two tails, which is both event A and event B.**

8) A claim is made that 60% of all the undergraduate students at a community college have their own computer. Amy thinks this figure is too high and decides to collect some data to test her theory. She randomly selects a sample and conducts a survey. But most of the students in her sample didn’t respond to the question. What type of bias is this an example of?

   **Non-response Bias**

9) If in a scatterplot, the data points seem to be descending from left to right along a straight line, what type of association would you say they have?

   **Negative Linear Association**

10) A study was conducted to assess the effect of child abuse on crime, researchers recruited people convicted of violent crimes and interviewed them to determine their childhood abuse level. The researchers recruited a comparable group of people who lived in the same neighborhoods the convicts had last lived in, and who were approximately...
the same ages at the convicts. The same interviews were performed with this second
group. What type of study is this an example of?

   a. An experiment
   b. A Prospective Observational Study
   c. A Retrospective Observational Study

Questions 11-13 are based on the following information:

A study was conducted in Juneau, Alaska to assess if gender has an effect on smoking
habit. A simple random sample of size 100 was selected from the total adult male
population of 2050 and a simple random sample of size 120 was selected from the total
adult female population of 1000. The proportion of smokers among 100 males was found
to be .40, while the total numbers of smokers in the overall sample was found to be 70.

11) Give the response variable and the explanatory variable.

   Explanatory Variable: Gender
   Response Variable: Smoking habit

12) The proportion of .40 is a: statistic

13) What type of sampling technique was used to obtain the 220 selected adult
     residents?

   a) Stratified Random Sample
   c) Systematic Sampling
   e) Convenience Sample
   b) Simple Random Sample
   d) Cluster Sampling
   f) Multistage Sampling

14) Determine whether the following are qualitative or quantitative (discrete, continuous):

   a. A person’s religion – Qualitative

   b. Number of students in the class – Quantitative (discrete)

   c. Person’s marital status – Qualitative

   d. Amount of water in a glass – Quantitative (continuous)

15) A manufacturer is interested in the average length of steel rods. He takes a random
    sample of size 20 and records the length measurements:

    35, 47, 53, 41, 23, 54, 42, 45, 32, 47, 57, 46, 26, 51, 45, 50, 37, 48, 54, 41

b. What is the shape of the distribution of the rod lengths? (Symmetric, Skewed to the left, Skewed to the right)

**Skewed to the left**

16) Given the following sample of data: [11, 12, 13, 14, 14, 15, 15, 16, 17, 17, 19, 20, 21, 22, 22, 22, 23, 23, 24, 24, 27]

a. Construct a frequency plot of the data.

```
11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
```

b. Describe the distribution of the data.

**Bimodal**

Use the following information to answer questions 17 & 18:

The New Brew Pub manufactures and distributes three types of beers: a low calorie light beer, a regular beer, and a dark beer. To investigate the relationship between gender and beer preference, a sample of 450 beer drinkers was selected. After taste-testing each of the three beers, the individuals were asked to state their preference, defined as their first choice. The data is summarized in the following 2-way frequency table:

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>Light</th>
<th>Regular</th>
<th>Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td></td>
<td>60</td>
<td>120</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>90</td>
<td>90</td>
<td>30</td>
</tr>
</tbody>
</table>

17) Give the marginal distribution for beer preference.

<table>
<thead>
<tr>
<th>Beer Preference</th>
<th>PERCENTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light</td>
<td>33.3%</td>
</tr>
<tr>
<td>Regular</td>
<td>46.7%</td>
</tr>
<tr>
<td>Dark</td>
<td>20%</td>
</tr>
</tbody>
</table>

18) Give the conditional distribution of beer preference given the gender.
19) Given the following sample of data: [29, 21, 26, 24, 17, 24, 18, 22, 23, 15, 17, 19]
   a) What is the mean (show all work)?
   \[
   \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{255}{12} = 21.25
   \]

   b) What is the standard deviation (show all work)?
   \[
   \sigma = \sqrt{\frac{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}} = \sqrt{\frac{12(5611) - (255)^2}{132}} = \sqrt{\frac{2307}{132}} = 4.18
   \]

   c) What is the median?
   Order the data: 15, 17, 17, 18, 19, 21, 22, 23, 24, 24, 26, 29. Therefore, the median is:
   \[
   \frac{21 + 22}{2} = 21.5
   \]

   d) What is the mode(s)? 17 and 24

   e) What is the range? 29-15=14

20) Given the following sample of data: [10.5, 11.3, 11.9, 12.0, 12.3, 12.3, 12.5, 12.7, 13.4, 13.7, 13.8, 14.2, 14.8, 15.1, 15.3, 16.7, 16.8, 18.8, 20.8]
   a) Find the five-number summary
   Min. Value: 10.5
   Q1: 12.3
   Median: 13.7
   Q3: 15.3
   Max. Value: 20.8

   b) Construct a boxplot
21) In the population of students who took the GRE in a particular year, the "analytical ability" scores were normally distributed with a mean of 520 and a standard deviation of 10, \( X \sim N(520, 10) \)

(a) Draw the distribution of the "analytical ability" score and label the X axis using the 68-95-99.7 (1\(\sigma\)-2\(\sigma\)-3\(\sigma\)) rule often referred to as the Empirical rule.

Fill in the following:
About 68% of students have scores between _510_ and _530_.
About 95% of students have scores between _500_ and _540_.
About 99.7% of students have scores between _490_ and _550_.

(b) Find the probability that a student has a score over 550. In the process of working out your answer, do the following: (i) Draw a picture of the problem in degrees scale...
and shade in the area desired. (ii) Draw a picture of the problem in Z scale and shade in the table area given. (iii) Compute the value of Z corresponding to X=550
(iv) Determine the probability of score of 550 or more.

Work: Do (i) then (iii) then (ii).
(i) Bell shaped curve with a mean at 520 and a line at 550 with area to right shaded in.
(ii) Bell shaped curve with a mean at 0.0 and a line at 3 with the area to the right shaded in.
(iii) Work  \[ Z= \frac{X-\mu}{\sigma} = \frac{550-520}{10} = \frac{30}{10} = 3 \]  (iv) P(Z>3)=1- Area to left of 3 =1- 0.9987=0.0013 ans.

(c) Find the 99th percentile for the score. In the process of doing this, give the following:
(i) The 99th percentile corresponds to what proportion? (ii) Find the 99th percentile for Z the standard normal variable. (iii) Use your answer to (ii) to find the 99th percentile for score of students.

(i) 99% corresponds to 0.99. (ii) Area to left of Z =0.99 then by Table III, Z= 2.326.
(iii) X=\mu +Z\sigma= 520 + (2.326\times10) = 520 + 23.26 = 543.26.

(d) Julie has a score of 550. Is this an unusually high score? Use the answer to (c) or (b) to answer this question. Any time in the top 1% is considered to be an unusually high for a student. Write a sentence below giving your decision and reason.
Using (b): From (b) we see that 0.13% of values exceed 550; so 550 is in the top 1% and is an unusually high score.
Using (c): The value 550 is greater that 543.26, the 99th percentile so 550 is in the top 1% and is an unusually high score.