**Key Terms**

- **central tendency**: those numbers that reflect “middle” or center scores in a distribution of scores. The most common measures being: mean, mode, median.

- **mean**: the most common and widely used measure of central tendency. This is the arithmetic average. It is calculated by taking the total of all the scores and dividing this figure by the number of actual scores. Symbolized by “X” with a line on top (\( \bar{X} \))

- **Median**: the score at the 50th percentile of the distribution. The exact center score in an array of scores. There are equal numbers of scores below (in value) and above in value. Scores must be ordered by numerical value (low to high) to compute the median.

- **Mode**: this is the score that occurs more frequently than any other score. If there are two (2) scores that occur more frequently (the same number of times) then both scores are the mode. This is a “bi-modal distribution”.

- **Range**: the number that represents the dispersion of scores throughout a distribution. It represents the scale (value) difference between the largest and the smallest values in a distribution. It is not very useful because of its inherent instability. It does, however, describe how much dispersion there is in an array of scores. The **crude range** is calculated by subtracting the lowest number from the highest number and adding 1.

- **normal curve**: a “bell-shaped” curve or frequency distribution in which the scores are spread normally throughout the distribution. In this distribution the greatest number of scores tend to be towards the center of the curve with extremely high or extremely low scores being less frequent (see example below).

- **skewed distributions**: non-normal frequency distributions in which extreme scores tend to be more frequent or distributed in other unusual ways. The most common types of skewed distributions are: positively skewed, negatively skewed, platykurtic, leptokurtic, and bimodal (see examples below).

**Abbreviations**

- \( n \) = the number (frequency) of scores, observations or subjects in a study  
- \( Ss \) = subjects  
- \( P \) = population or the total number of people to which a study refers or from which a sample is drawn  
- \( f \) = frequency, how many times a particular score or “data” event occurs  
- \( r \) = correlation  
- \( R \) = multiple correlation  
- \( \bar{X} \) = mean
TYPES OF FREQUENCY DISTRIBUTIONS

Normal Curve

Leptokurtic Curve

Platykurtic Curve

Positive Skew

Negative Skew

Bimodal Curve

Histogram

Frequency Polygon
**CORRELATION**

Correlation refers to the relationship between variables or events. In statistical analysis correlation refers to a quantifiable relationship between sets of numbers that represent data taken from subjects. The Pearson-Product Moment correlation (the most basic) represents a comparison of two or more measurements taken on a person and comparing the relationship with other persons (subjects) from which the same measurements have been taken. The correlation coefficient \( r \) is the number that represents the linear relationship between these variables or measurements.

**RULES OF CORRELATION**

- Two (or more) measurements are obtained on a group of individuals, events, or subjects. For example: height (in inches) and weight (in pounds) are taken from 20 people or IQ scores and grade point average (GPA) are taken from a population of 11th graders.

- The range of the correlation coefficient, the number that represents the power of the relationship between the measurements that were taken (height/weight; IQ score/GPA) can vary between +1.00 and -1.00. These are the extreme limits of correlation and represent a perfect relationship between the variables. The closer the correlation value (\( r \)) approaches .00, the less significant and weaker is the relationship between the variables.

- A positive correlation (or relationship) indicates that high values in one variable tend to be related with high values in the other variable. For example: high scores in weight are related to high scores in height. Most subjects therefore are both greater in height measurement (taller) and greater in weight measurement (heavier) or 11th graders with higher GPAs also tend to have higher IQ scores. This would be represented by a correlation coefficient of \( r = +.87 \) (for example). Positive correlations above +.50 are considered high (depending upon the number of subjects).

- A negative correlation (or relationship) indicates that high values in one variable are related to low values in the other variable. This is known as a negative or inverse relationship. A negative correlation coefficient is indicated by a negative number (\( r = -.87 \), for example). In the above example this would indicate that students with low GPAs would tend to have higher IQ scores and people who are heavier would tend to have lower values in height (be shorter) or visa versa.

- Both positive and negative correlations that exceed ±.50 (depending upon the number of subjects) would indicate that there is a significant relationship between the variables being studied.

- Correlation continuum: +1.00....(.80)....(.50)....(.25)....(.00)....(-.25)....(-.50)....(-.80).... -1.00 (strong correlation) (weak correlation) (strong correlation)

- Correlation means that there is a mathematical relationship between two or more variables it does NOT mean CAUSATION.
PART A: Computations. For the following data compute the mean, mode, median and crude range.

(a) 13, 15, 16, 17, 19, 22, 23, 34, 45
(b) 2, 3, 5, 5, 7, 8, 11, 13, 19
(c) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
(d) 1, 2, 3, 3, 4, 4, 6, 7, 8, 9, 9, 9, 10, 11
(e) 45, 45, 46, 48, 49, 59, 43, 47, 48, 58
(f) 1, 44, 56, 78, 65, 43, 12
(g) 123, 12, 34, 54, 0, 56, 76, 23, 76
(h) 12, 120, 34, 56, 67, 54, 56, 12, 23
(i) 0, 2, 0, 2, 3, 5, 6, 7, 67, 45, 1, 0, 100
(j) 12, 34, 560, 561, 567
(k) 10, 8, 6, 9, 5, 4, 0, 0, 0
(l) 100, 99, 45, 99, 100, 0, 0, 0, 12

PART B: Graphs & Plotting. Either draw a graph or plot according to the problems below.

(a) Draw examples of the following: a normal curve, platykurtic curve, leptokurtic curve, negative skew, positive skew, and bimodal distribution

(b) A examination is based upon 100 total possible points. The class crude range was 19. What does this tell you about the scores on the examination?

(c) There is a high positive correlation between smoking cigarettes and incidences of lung cancer. Does this indicate that smoking causes lung cancer? What does this statistic indicate? How is it used by people in the medical profession?

(d) Plot the following and describe the correlation (positive, negative, no correlation). Susan weighs 100 lb. and is 5’ tall; Jeff is 300 lb. and 6’ tall; Bill is 150 lb. and is 5’7” tall; Fred is 400 lb. and is 6’7” tall; Roberta is 75 lb. and 4’ tall; Jane is 123 lb. and 5’4” tall.

(e) Plot the following and describe the correlation (positive, negative, or no correlation) Susan weighs 100 lb. and is 7’ tall; Jeff is 300 lb. and 4’ 5” tall; Bill is 150 lb. and is 6’ tall; Fred is 124 lb. and is 6’7” tall; Roberta is 475 lb. and 4’ tall; Jane is 250 lb. and 5’4” tall.

(f) A weight distribution chart for Japanese sumo wrestlers and a height chart for racing jockeys at Keystone Race Track. Which is a positive skew? Which represents a negative one?

PART C: Correlation: For the following sets of data determine the correlation (positive, negative or weak/no correlation)

(a) Between test #1 and test #2
(b) Between test #1 and test #3
(c) Between test #2 and test #3

<table>
<thead>
<tr>
<th>Subject</th>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne</td>
<td>100</td>
<td>98</td>
<td>67</td>
</tr>
<tr>
<td>Sam</td>
<td>99</td>
<td>95</td>
<td>69</td>
</tr>
<tr>
<td>Betty</td>
<td>95</td>
<td>92</td>
<td>72</td>
</tr>
<tr>
<td>Fred</td>
<td>56</td>
<td>55</td>
<td>100</td>
</tr>
<tr>
<td>Pete</td>
<td>75</td>
<td>81</td>
<td>87</td>
</tr>
<tr>
<td>Eldridge</td>
<td>89</td>
<td>89</td>
<td>77</td>
</tr>
<tr>
<td>Charles</td>
<td>79</td>
<td>83</td>
<td>85</td>
</tr>
<tr>
<td>Robert</td>
<td>65</td>
<td>67</td>
<td>95</td>
</tr>
<tr>
<td>Jennifer</td>
<td>62</td>
<td>63</td>
<td>97</td>
</tr>
<tr>
<td>Shelly</td>
<td>57</td>
<td>56</td>
<td>99</td>
</tr>
</tbody>
</table>