

## 2.2 Frequency Distributions

A **frequency distribution** is an organized tabulation of the number of individuals or frequency of occurrence in each category on a scale of measurement.

**Frequency Distribution** is a table constructed to show how *many times* a given score or group of scores occurred in a set of data. A **simple frequency** (ungrouped frequency) distribution (see Table 2.2.1) is the ordering of the scores of a distribution from highest to lowest scores in a table with their corresponding frequencies. When scores are grouped into intervals showing how many scores occurred in each interval, this is called a **grouped frequency** distribution (Table 2.2.2).

**Apparent Limits** are the limits displayed in a grouped frequency table (Table 2.2.2, col. 1); these limits give a reasonable range between which groups of data exist.

**Real Limits** (for continuous data – measurements or observations that depend upon the accuracy of the measuring instrument) of any interval extend from  $\frac{1}{2}$  unit below and above the apparent lower and upper limits respectively. The real limits of the “20 – 30” interval are 19.5 and 30.5. The real lower limit is designated  $L$  and the real upper limit is designated  $U$ .

The **Midpoint**,  $MP$ , of an interval is its exact center. The  $MP$  of any interval is found by adding the apparent upper limit to the apparent lower limit and dividing by 2. The  $MP$  of the “20 – 30” interval is 25.

**Interval size** is denoted by the symbol  $i$ ; it is the distance between the real lower limit and the real upper limit. The interval size is determined by subtracting  $L$  from  $U$ . It is recommended that you group data so that there are between 8 and 15 intervals.

**Frequency ( $f$ )** simply indicates how many scores are located or counted in each interval.

**Number,  $N$ ,** is the number of scores in a distribution or the total of all the frequencies for all the intervals.  $N$  is also called the sample size.

The following are some outputs of frequency distributions from various statistical programs and Excel spreadsheets.

Table 2.2.1 *Simple Frequency Distribution of 9th Graders (ODE)*

| 9 <sup>th</sup> Graders | Frequency ( $f$ ) | Percent |
|-------------------------|-------------------|---------|
| 100                     | 1                 | 1.06    |
| 98                      | 1                 | 1.06    |
| 95                      | 1                 | 1.06    |
| 86                      | 1                 | 1.06    |
| 85                      | 1                 | 1.06    |
| 84                      | 2                 | 2.13    |
| 83                      | 2                 | 2.13    |
| 82                      | 1                 | 1.06    |
| 81                      | 2                 | 2.13    |
| 80                      | 1                 | 1.06    |
| 79                      | 1                 | 1.06    |
| 78                      | 3                 | 3.19    |
| 77                      | 2                 | 2.13    |
| 75                      | 4                 | 4.26    |
| 74                      | 3                 | 3.19    |
| 73                      | 5                 | 5.32    |
| 72                      | 3                 | 3.19    |
| 71                      | 3                 | 3.19    |
| 69                      | 4                 | 4.26    |
| 68                      | 5                 | 5.32    |
| 67                      | 3                 | 3.19    |
| 66                      | 3                 | 3.19    |
| 65                      | 3                 | 3.19    |
| 64                      | 3                 | 3.19    |
| 63                      | 1                 | 1.06    |
| 62                      | 2                 | 2.13    |
| 61                      | 4                 | 4.26    |
| 60                      | 2                 | 2.13    |

Table 2.2.1 *Simple Frequency Distribution of 9th Graders (ODE)cont.*

|    |   |      |
|----|---|------|
| 59 | 3 | 3.19 |
| 57 | 1 | 1.06 |
| 56 | 2 | 2.13 |
| 55 | 2 | 2.13 |
| 54 | 2 | 2.13 |
| 53 | 1 | 1.06 |
| 52 | 1 | 1.06 |
| 51 | 2 | 2.13 |
| 50 | 4 | 4.26 |
| 49 | 1 | 1.06 |
| 47 | 3 | 3.19 |
| 40 | 1 | 1.06 |
| 37 | 1 | 1.06 |
| 34 | 1 | 1.06 |
| 33 | 1 | 1.06 |
| 28 | 1 | 1.06 |

*Note.*  $N = 94$ , Percent =  $f/N$

Table 2.2.2 *Grouped Frequency Distribution of 9th Graders (ODE\_New)*

| Apparent Limits<br>(Scores) | Frequency<br>( $f$ ) | Real Limits<br>( $L - U$ ) | Midpoint<br>( $MP$ ) | Interval Size<br>( $i$ ) |
|-----------------------------|----------------------|----------------------------|----------------------|--------------------------|
| 100 – 109                   | 1                    | 99.5 – 109.5               | 104.5                | 10                       |
| 90 – 99                     | 2                    | 89.5 – 99.5                | 94.5                 | 10                       |
| 80 – 89                     | 7                    | 79.5 – 89.5                | 84.5                 | 10                       |
| 70 – 79                     | 8                    | 69.5 – 79.5                | 74.5                 | 10                       |
| 60 – 69                     | 11                   | 59.5 – 69.5                | 64.5                 | 10                       |
| 50 – 59                     | 9                    | 49.5 – 59.5                | 54.5                 | 10                       |
| 40 – 49                     | 3                    | 39.5 – 49.5                | 44.5                 | 10                       |
| 30 – 39                     | 2                    | 29.5 – 39.5                | 34.5                 | 10                       |
| 20 – 29                     | 1                    | 19.5 – 29.5                | 24.5                 | 10                       |

The screenshot shows the SPSS Data Editor interface with the 'Frequencies' dialog box open. The 'Frequencies' dialog box has 'Passed 9th Grade [pass9th]' selected in the 'Variable(s):' list. The 'Display frequency tables' checkbox is checked. Below the dialog box, the 'Frequencies: Charts' sub-dialog box is open, showing 'Histograms' selected as the 'Chart Type' with the 'With normal curve' checkbox checked. The 'Chart Values' section has 'Frequencies' selected. The background data table shows columns for school ID, enrollment, and pass9th status.

| School ID | Enrollment | Pass9th |
|-----------|------------|---------|
| 27        | 77503      | 2       |
| 28        | 151992     | 2       |
| 29        | 30330      | 2       |
| 30        | 2063       | 96.7    |
| 31        | 2584       | 96.4    |
| 32        | 2968       | 95.4    |
| 33        | 2464       | 95.5    |
| 34        | 2766       | 93.0    |
| 35        | 11226      | 95.0    |
| 36        | 2834       | 96.1    |
| 37        | 2252       | 95.3    |
| 38        | 2250       | 95.0    |
| 39        | 2837       | 96.2    |
| 40        | 2309       | 94.8    |
| 41        | 2492       | 94.6    |
| 42        | 2615       | 94.9    |
| 43        | 2205       | 96.5    |
| 44        | 2420       | 96.2    |
| 45        | 2063       | 96.7    |
| 46        | 2584       | 96.4    |

Figure 2.2. 1 SPSS frequency procedure: Analyze -> Charts -> Select Histograms

### Tony and Frequency Distributions

**Tony:** “Given a set of data, how best to construct a grouped frequency distribution?”

**Rose:** “The concept of frequency distribution is not complex; however, the mechanics of construction a grouped and sometimes a simple frequency distribution table

of graph is often complex for most students.” “So it is best illustrated with the statistics tutorials and the Excel spreadsheet program.” The mode is readily computed from a frequency distribution; the most frequent score or the *midpoint* of group interval.

Table 2.2.3

*Grouped Frequency SPSS Output*

| Scores    | Frequency |
|-----------|-----------|
| 20 – 29   | 1         |
| 30 – 39   | 3         |
| 40 – 49   | 5         |
| 50 – 59   | 18        |
| 60 – 69   | 30        |
| 70 – 79   | 24        |
| 80 – 89   | 10        |
| 90 – 100  | 2         |
| 100 – 109 | 1         |

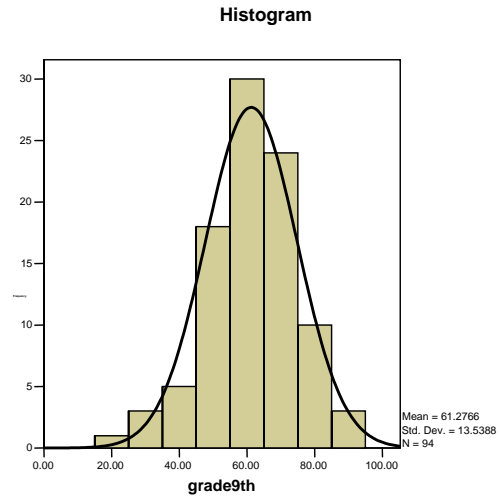


Figure 2.2.2 Histogram of grouped frequency.

Table 2.2.4 *Output from Stat4U after Recoding*

FREQUENCY ANALYSIS BY BILL MILLER (Stats4U)

| Frequency Analysis for passFreq |        |       |      |           |           |           |
|---------------------------------|--------|-------|------|-----------|-----------|-----------|
| MP                              | UP TO  | FREQ. | PCNT | CUM.FREQ. | CUM.PCNT. | %ILE RANK |
| 24.50                           | 34.50  | 1     | 0.01 | 1.00      | 0.00      | 0.01      |
| 34.50                           | 44.50  | 3     | 0.03 | 4.00      | 0.04      | 0.03      |
| 44.50                           | 54.50  | 5     | 0.05 | 9.00      | 0.10      | 0.07      |
| 54.50                           | 64.50  | 18    | 0.19 | 27.00     | 0.29      | 0.19      |
| 64.50                           | 74.50  | 30    | 0.32 | 57.00     | 0.61      | 0.45      |
| 74.50                           | 84.50  | 24    | 0.26 | 81.00     | 0.86      | 0.73      |
| 84.50                           | 94.50  | 10    | 0.11 | 91.00     | 0.97      | 0.91      |
| 94.50                           | 104.50 | 2     | 0.02 | 93.00     | 0.99      | 0.98      |
| 104.50                          | 114.50 | 1     | 0.01 | 94.00     | 1.00      | 0.99      |

### Proportion and Percent Distributions

So far we have dealt with frequency distributions based on raw scores or scores per interval (frequencies of each scores or groups of scores); however, a more common and useful presentation is the proportional or percentage distribution.

The advantage of proportional and percentage distributions is that they allow comparisons between different samples or dataset when the number of cases or sample size varies. The basic question of interest to a proportional or percentage distribution is, “What percentage or proportion of cases falls within a given score?” These types of distributions are manipulated to calculate the mean and median from a frequency distribution.

A proportional or percentage distribution represents the relative frequency of scores rather than their absolute frequency. The procedure to construct proportional,  $p$  or percentage,  $P$  distribution is illustrated below; 14.3% of scores have a value of 33:

#### Construction Guide: Proportional and Percentage Distribution (ungrouped)

The proportion for each score  $p_i$ : ( $i$  is the score or case number)

$$p_i = \frac{f_i}{n}, \text{ where } f_i \text{ is frequency for score } i \text{ and } n \text{ is total frequency}$$

The percentage for each score,  $P_i$ :

$$P_i = \frac{f_i}{n} \cdot 100$$

| $i$ | $X_i$ | $f_i$    | $p_i$              | $P_i(\%)$            |
|-----|-------|----------|--------------------|----------------------|
| 1   | 35    | 6        | 0.286              | 28.6                 |
| 2   | 34    | 2        | 0.095              | 9.5                  |
| 3   | 33    | 3        | 0.143              | 14.3                 |
| 4   | 32    | 1        | 0.048              | 4.8                  |
| 5   | 31    | 4        | 0.190              | 19.0                 |
| 6   | 30    | 5        | 0.238              | 23.8                 |
|     |       | $n = 21$ | $\Sigma p_i = 1.0$ | $\Sigma P_i = 100\%$ |

## Cumulative Distributions

Cumulative distributions, like proportional and percentage distributions, form the basis for comparative interpretation of data. Cumulative frequency distribution represents the number of cases up to and including a specific score or value. There are three forms to the cumulative distributions (listed, ungrouped, grouped).

### Listed and Ungrouped Cumulative Distributions

The dataset of listing of data is such that no range or values of scores occurs more than once. Table 2.2.4 shows the cumulative frequency distribution table. The cumulative frequency,  $C_f$  column shows that for score value of 30 or less, there is only one case. For score value 34, there are 5 cases. All cumulative frequency tables are form by cumulating the frequency for each score from bottom to top (remember that in a frequency table the raw scores or groups intervals are listed highest at top of table and lowest at bottom).

Table 2.2.5 shows the ungrouped cumulative frequency distribution, we follow the same rule as we did for determining the cumulative listed frequency. Observed that the highest cumulative frequency is equal that of the total frequency,  $n$ .

Table 2.1. 4 *Listed Cumulative Distribution*

| $X_i$            | $f_i$ | $C_f$    |
|------------------|-------|----------|
| 35               | 1     | <b>6</b> |
| 34               | 1     | 5        |
| 33               | 1     | 4        |
| 32               | 1     | 3        |
| 31               | 1     | 2        |
| 30               | 1     | 1        |
| $n = \mathbf{6}$ |       |          |

Table 2.1.5 *Ungrouped Cumulative Distribution*

| $X_i$             | $f_i$ | $C_f$     |
|-------------------|-------|-----------|
| 40                | 6     | <b>21</b> |
| 39                | 2     | 15        |
| 38                | 3     | 13        |
| 37                | 1     | 10        |
| 36                | 4     | 9         |
| 35                | 5     | 5         |
| $n = \mathbf{21}$ |       |           |

It is a simple step to go from a cumulative frequency distribution to a cumulative percentage or proportional distribution. To construct a cumulative percentage or proportional distribution, it is necessary to first construct a cumulative frequency distribution. Below is a guide to construct a cumulative percentage or proportional distribution for a listed frequency or ungrouped frequency distribution.

**Construction Guide: Proportional and Percentage Cumulative Distribution (ungrouped)**

The cumulative proportion for each score  $C_p$ :

$$C_p = \frac{C_f}{n}, \text{ where } C_f \text{ is frequency for a given score and } n \text{ is total frequency}$$

The cumulative percentage for each score,  $CP$ :

$$CP = \frac{C_f}{n} \cdot 100$$

| $X_i$ | $f_i$    | $C_f$     | $C_p$                              | $CP(\%)$                           |
|-------|----------|-----------|------------------------------------|------------------------------------|
| 35    | 6        | <b>21</b> | <b>1.00</b>                        | <b>100.0</b>                       |
| 34    | 2        | 15        | 0.714                              | 71.4                               |
| 33    | 3        | 13        | 0.619                              | 61.9                               |
| 32    | 1        | 10        | 0.476                              | 47.6                               |
| 31    | 4        | 9         | 0.429                              | 42.9                               |
| 30    | 5        | 5         | 0.238                              | 23.8                               |
|       | $n = 21$ |           | <b>Max(<math>C_p</math>) = 1.0</b> | <b>Max(<math>CP</math>) = 100%</b> |

We could make the following statement from the cumulative distribution table, 61.9% of scores are less than or equal to ( $\leq$ )  $X = 33$ .

Construction of cumulative percentage or proportional distribution for *grouped* frequency tabulation is similar to constructing one from an ungrouped frequency distribution. When data are grouped into scale intervals or classes, we use the *midpoint* of the class or interval when we reference the frequency, percentage, or a proportion. For



example, for a class or group interval of 20 to 21 with a frequency of five, we would say that five scores have a value of 20.5 [*midpoint*, (20 + 21)/2].

For cumulative data, instead of referring to the *midpoint*, we use the *upper limit*, of the group interval. Below is a guide to construct the cumulative percentage or proportional distribution from a grouped frequency distribution (same procedure as ungrouped and listed; however we phase the cumulative percentage differently relative to the grouped interval).

**Construction Guide: Proportional and Percentage Cumulative Distribution (grouped)**

The cumulative proportion for each score  $C_p$ :

$$C_p = \frac{C_f}{n}, \text{ where } C_f \text{ is frequency for a given score and } n \text{ is total frequency}$$

The cumulative percentage for each score,  $CP$ :

$$CP = \frac{C_f}{n} \cdot 100$$

| $X_i$ | $f_i$    | $C_f$     | $C_p$                              | $CP(\%)$                           |
|-------|----------|-----------|------------------------------------|------------------------------------|
| 50-52 | 6        | <b>21</b> | <b>1.00</b>                        | <b>100.0</b>                       |
| 47-49 | 2        | 15        | 0.714                              | 71.4                               |
| 44-46 | 3        | 13        | 0.619                              | 61.9                               |
| 41-43 | 1        | 10        | 0.476                              | 47.6                               |
| 38-40 | 4        | 9         | 0.429                              | 42.9                               |
| 35-37 | 5        | 5         | 0.238                              | 23.8                               |
|       | $n = 21$ |           | <b>Max(<math>C_p</math>) = 1.0</b> | <b>Max(<math>CP</math>) = 100%</b> |

For group interval 48 – 49 we can make the following statements

| $X_i$ | $f_i$ | $C_f$ | $CP(\%)$ |
|-------|-------|-------|----------|
| 47-49 | 2     | 15    | 71.4     |

For the *apparent upper limit*, we would say that 15 cases or scores, or 71.4% of all cases, fall at or below the score or value of 49 (or 49.5 if we needed to use the exact or *real upper limit* of the group interval).

Although the cumulative percentage distribution gives the percentage of all scores that fall at or below an observed value or score, it is often necessary to compute the score for a specific cumulative percentage. For example, in the previous example, we found that 71.4% of scores were at or below the value of 49. We may need to find the score at which 70%, or 90%, or 50% (median), etc., below which all scores falls. To find the scores at which a given percentage of the distribution fall below for percentage points not readily obvious from the cumulative percentage distribution, we need to do additional computation; we need to find the Percentile.

### Percentile from Cumulative Percentage Distribution

Percentile is the score at which a certain percent (%) of the scores is less than or equal to.

The percentile of a cumulative percentage distribution or a transformed frequency distribution is determined by the following formula:

$$P_{\%} = LL_i + \left[ \left( \frac{n_p - C_f}{f_i} \right) \cdot I \right]$$

where  $P_{\%}$  = any specified percentile point

$LL_i$  = the *exact lower limit* or *real lower limit* of the group interval containing

Percentile point,  $P_{\%}$

$n_p$  = number of scores or cases comprising the specified percentage for  $n$ .

$n_p = (p)(n)$ , where  $p$  is decimal equivalent to % and  $n$  is total frequency

$C_f$  = cumulative frequency up to but not including the percentile interval

$f_i$  = frequency within the percentile interval

$I$  = group interval size or class size

**Construction Guide: Computing Percentile Points**

Percentile from cumulative distribution

$$P_{\%} = LL_i + \left[ \left( \frac{n_p - C_f}{f_i} \right) \cdot I \right] = 43.5 + \left[ \left( \frac{10.5 - 10}{3} \right) \cdot 3 \right] = 43.6$$

 $P_{\%}$  = any specified percentile point $LL_i$  = the *exact lower limit* or *real lower limit* of the group interval containing Percentile point,  $P_{\%}$  $n_p$  = number of scores or cases comprising the specified percentage for  $n$ . $n_p = (p)(n)$ , where  $p$  is decimal equivalent to % and  $n$  is total frequency $C_f$  = cumulative frequency up to but not including the percentile interval $f_i$  = frequency within the percentile interval $I$  = group interval size or class size

| $X_i$        | $f_i$ | $C_f$     | $C_p$        | $CP(\%)$     |
|--------------|-------|-----------|--------------|--------------|
| 50-52        | 6     | 21        | <b>1.00</b>  | <b>100.0</b> |
| 47-49        | 2     | 15        | 0.714        | 71.4         |
| <b>44-46</b> | 3     | <b>13</b> | <b>0.619</b> | <b>61.9</b>  |
| 41-43        | 1     | 10        | 0.476        | 47.6         |
| 38-40        | 4     | 9         | 0.429        | 42.9         |
| 35-37        | 5     | 5         | 0.238        | 23.8         |
| $n = 21$     |       |           |              |              |

Question: Find the score at which 50% of all scores fall (the median)

Step 1: Compute number of cases equal to 50% of  $n$ ,  $n_p = (0.50)(21) = \mathbf{10.5}$ Step 2: Using  $C_f$ , find the interval at which 50% falls, percentile interval or median interval here: 44 – 46Step 3: Find  $LL_i$ , the exact lower limit of the percentile interval,  $LL_i = \mathbf{43.5}$ Step 4: Find the number of cases up to but not including those in the percentile interval  
 $C_f = \mathbf{10}$ Step 5: Subtract  $(n_p - C_f) = 10.5 - 10 = \mathbf{0.5}$ Step 6: Divide step 5 by number of case within percentile interval,  $f_i : 0.5/3 = 0.1667$ Step 7: Multiply Step 6 by interval size,  $I = 3$ ,  $0.1667/3 = 0.0556$ Step 8: Add Step 7 and Step 3:  $43.5 + 0.0556 = 43.556$  or  $P_{50} = \mathbf{43.6}$

### Mean from Frequency Distribution

The arithmetic mean is readily computed from a listed data (frequency of each score is 1). It is the sum of all the scores,  $\sum X$  divided by the sample size,  $n$ .

$$\text{mean} = \bar{X} = \frac{\sum X}{n}$$

To compute the arithmetic mean from either grouped or ungrouped frequency distribution requires computation that considers the frequency,  $f$ , of each score. The formula for the mean for or grouped or ungrouped data is the sum of all the products of the scores (or *midpoint*, for grouped) and their corresponding frequencies divided by the number of scores or cases. The use of this formula requires that each score,  $X_i$  (or *midpoint*, if the data is grouped into intervals) be first multiplied by its associated frequency ( $f_i$ ).

$$\text{mean} = \bar{X} = \frac{\sum f_i X_i}{n}$$

#### Construction Guide: Computing Mean from Ungrouped Data

Mean from frequency:

$$\text{mean} = \bar{X} = \frac{\sum f_i X_i}{n} = \frac{683}{21} = 32.52$$

| $X_i$ | $f_i$    | $f_i X_i$            |
|-------|----------|----------------------|
| 35    | 6        | 210                  |
| 34    | 2        | 68                   |
| 33    | 3        | 99                   |
| 32    | 1        | 32                   |
| 31    | 4        | 124                  |
| 30    | 5        | 150                  |
|       | $n = 21$ | $\sum f_i X_i = 683$ |

Step 1: Multiply ( $f_i$ )( $X_i$ )      Step 2: Add  $f_i X_i$  column,  $\sum f_i X_i = 683$

Step 3: Add  $f_i$  column  $\sum f_i = 21$       Step 4: Divide Step 2 by Step 3,  $683/21 = 32.52$

**Construction Guide: Computing Mean from Grouped Data**

Mean from frequency:

$$\text{mean} = \bar{X} = \frac{\sum f_i X_i}{n} = \frac{915}{21} = 43.57$$

| $X_i$ | <i>Midpoint</i><br>( <i>New X</i> ) | $f$      | $fX$                 |
|-------|-------------------------------------|----------|----------------------|
| 50-52 | 51                                  | 6        | 306                  |
| 47-49 | 48                                  | 2        | 96                   |
| 44-46 | 45                                  | 3        | 135                  |
| 41-43 | 42                                  | 1        | 42                   |
| 38-40 | 39                                  | 4        | 156                  |
| 35-37 | 36                                  | 5        | 180                  |
|       |                                     | $n = 21$ | $\sum f_i X_i = 915$ |

Step 1: List the midpoint for each class interval, new  $X$

Step 2: Multiply ( $f$ )(*Midpoint*) or  $fX$

Step 3: Add  $fX$  column,  $\sum fX = 915$

Step 4: Add  $f$  column  $\sum f = 21$

Step 4: Divide Step 3 by Step 4,  $915/21 = 43.57$