Social Studies 201

January 17, 2005

Presenting data

See text, chapter 4, pp. 87-160.

Data sets

When data are initially obtained from questionnaires, interviews, experiments, administrative sources, or other methods, a statistical analyst encounters a list or array of values of a variable or variables. These values represent characteristics of individuals or objects – the values of a variable or variables taken on by the individuals or objects. If the data are quantitative, these are lists of numbers; if the data are qualitative, these are lists of the words or phrases associated with each response.

In order to analyze such data, it is awkward and time consuming to always have to deal with the list of all possible values. In general, when using these data, the first step of a statistical analyst is to organize them into a form that makes it straightforward to examine and analyze the data set.

This section of the notes examines ways of organizing data. One basic distinction to consider when organizing data is the difference between ungrouped and grouped data. This distinction will be used throughout the course and is as follows (see text, pp. 94-95).

> **Ungrouped data**. A list of all the values of a variable in a data set is referred to as ungrouped data.

That is, ungrouped data comprise simply a list of all the values of one or more variables across the cases of a data set. Examples of ungrouped data are provided in Examples 2.7 and 2.8.

> **Grouped data**. These are data organized into categories or intervals, usually presented in a table or diagram.

Grouped data refer to tables, diagrams, or other presentations of data where values or sets of values are grouped together into categories or intervals.

The following notes give examples of ungrouped and grouped data. Three ways of organizing ungrouped data into grouped form are discussed – a tally or count, using computer programs, and a stem-and-leaf display.

Notation

See text, section 4.4, pp. 97-100.

A large part of learning any new academic discipline or new method is to learn the language of the discipline or method. For statistical analysis, there are conventions about how to label and organize data. While statisticians use various techniques, symbols, and notation, the approach taken in these notes and in the Social Studies 201 text are similar to those used across many statistics texts. Some of the notational conventions used for this course are as follows.

- **Variables** – X, Y, Z. Variables are usually given symbols such as X, Y, or Z. That is, upper-case letters near the end of the alphabet are used to indicate variables. For example X might represent a variable such as age of respondent and Y might represent grade of students. The symbol Z is reserved for a specific variable, the standardized normal variable. In Module 6 and in the second part of the Social Studies 201 text, Z will be used to indicate the standardized normal variable.

- **Frequency of occurrence** – f. The lower case letter f is used to denote frequency or the number of times each value of a variable occurs. In the stem-and-leaf display of Example 2.9, if X is the variable credit hours, there is one student taking 3 credit hours, so $f = 1$ when the variable $X = 3$. For $X = 15$ credit hours, the frequency of occurrence is $f = 8$, that is, eight students took fifteen credit hours. The different numbers taken on by f indicate the frequencies of occurrence of the different categories of credit hours, that is, the number of respondents in each category.

- **Proportion** – p. The lower-case letter p is used to denote the proportion of total cases in a data set taking on a particular value. Examples of how to calculate proportions are provided later in Module 2.

- **Percentage** – P. The upper-case letter P is used to denote the percentage of total cases in the sample that take on a particular value. Examples of how to calculate and work with percentages are provided later in Module 2.

- **Sample size** – n. The lower case letter n is ordinarily the symbol for the sample size, that is, the number of cases in the data set. In example 2.8, there were $n = 15$ students in the sample (see Table **??**).

  This use of n as sample size is almost universal in statistical analysis. Whenever you encounter n in an article dealing with statistics, you can be quite sure that this represents the sample size for the data being discussed. A first question that a researcher might encounter is "What is the n?" That is, those examining the research wish to know how large a sample size a researcher used.

- **Population size**. The upper case letter N is ordinarily the symbol for the size of the population from which a sample of size n is drawn.

  If a census of the population is conducted so that all N members of the population are surveyed, then the sample size equals the population size, that is, $n = N$. It is more common to sample only a subset of the population, so generally the sample size n is much less than the population size N, that is, $n < N$.

- **Indexes**. Sometimes it is necessary to identify each individual value of a variable or frequency. Indexes are subscripts for the frequencies, f, or the variables X or Y, used to denote specific values of the variable. In the following distribution of credit hours for fourteen students, variable is credit hours, labelled as X, and the frequencies of occurrence of each credit hour are labelled as f.

Number of credit hours (X)	Number of respondents (f)
3	1
9	1
12	3
15	8
17	1
Total	$n = 14$

In this example, the values of X could have indexes associated with them, so they could be labelled X_i. There are five different values

for the variable, so $i = 1, 2, 3, 4, 5$ represent the five possible values X_1, X_2, X_3, X_4, and X_5 for X. Similarly, in index notation, the frequencies would be f_i, again with the five values f_1, f_2, f_3, f_4, and f_5.

Using this notation, the first value of X is $X_1 = 3$. The first frequency would be labelled f_1 and there is only one respondent taking three credit hours, so $f_1 = 1$.

Moving to the fourth line, indicating the number of students taking fifteen credit hours, the indexed value of the variable is $X_4 = 15$, and there are eight students taking on this fourth value, so $f_4 = 8$.

The complete set of indexed values for the distribution of credit hours of fourteen students is as follows.

$$
\begin{array}{cc}
X_i & f_i \\
X_1 = 3 & f_1 = 1 \\
X_2 = 9 & f_2 = 1 \\
X_3 = 12 & f_3 = 3 \\
X_4 = 15 & f_4 = 8 \\
X_5 = 17 & f_5 = 1 \\
\text{Total} & n = 14
\end{array}
$$

The complete set of indexes is not ordinarily used in this course. But from time to time, it is necessary to use the indexes, so you should have a general familiarity with the use of indexes.

- **Sum of frequencies**. The sum of all the frequencies of occurrence is equal to the sample size n. If there are k categories for a variable X, then the sample size is

$$n = f_1 + f_2 + f_3 + ... + f_k.$$

In the case of the distribution of credit hours in the previous bullet, there are $k = 5$ categories and sample size $n = 14$.

$$n = f_1 + f_2 + f_3 + f_4 + f_5 = 1 + 1 + 3 + 8 + 2 = 14$$

Distributions

See text, section 4.3, pp. 95-97 and section 4.6, pp. 117-119.

When summarizing data to present to others, a statistical analyst usually organizes these as what is termed a distribution. These distributions may be frequency distributions, proportional distributions, or percentage distributions.

A distribution is a format for presenting data, so the list of values taken on by a variable X, along with the relative occurrence of each value or set of values, is indicated. The occurrence may be in terms of frequencies (f), proportions (p), or percentages (P). The distribution is usually presented as a table.

In a table representing the distribution, the first column contains the list of values of X. Another column, usually the second column, lists the frequencies of occurrence, or number of times, each respective value of the variable occurs. This information gives the reader of the table a good idea of how the members of a sample or population are distributed across values of X.

Frequency Distribution

See text, section 4.3, pp. 95-97.

A frequency distribution lists the values of variable along with the frequency of occurrence of the variable. By convention, the values of the variable, usually represented by X or Y, are listed vertically in the first column of the table. The corresponding frequencies of occurrence, f, are listed in a second column, in the appropriate row. The total of the numbers in the second column, the sum of the frequencies of occurrence, is the sample size. This total is usually reported in the last row of the table.

A generic format for a frequency distribution table is given in Table 1. In this table, the variable is represented by X and the frequencies represented by f. There are k categories of the variable X, and the values of X and f have been included in index notation. In the last row, the total of the frequencies is the sample size.

Table 1: Generic format for a frequency distribution table

X	f
X_1	f_1
X_2	f_2
X_3	f_3
.	.
.	.
X_k	f_k
Total	n

Example 2.12 – Distribution of incomes of clerical workers. Use the data from the stem-and-leaf display of Example 2.10 to obtain a frequency distribution for the incomes of the sample of sixty Saskatchewan clerical workers.

Answer. The incomes from the sample of sixty Saskatchewan clerical workers were originally listed in the stem-and-leaf display handout. Where there are this many cases, one way to organize the values is to use a stem-and-leaf display. Alternatively, the number of values in each category could be counted. Using the groupings by tens from the stem-and-leaf display, a frequency distribution is given in Table 2.

The frequencies for the frequency distribution are the counts of the number of values in each row of the stem-and-leaf display. For example, in the first row of the display there were three values, so $f = 3$ is entered in the first row of the frequency distribution, indicating the number of cases from zero to nine. The second row of the stem-and-leaf display included all the cases with values from ten to nineteen – there were thirteen of these and this is the entry for f in the 10-19 category in the frequency distribution of Table 2.

Consistent with the conventions, values of the variable are placed in the first column of the table. In this example, these values

Table 2: Frequency distribution of income of sixty Saskatchewan clerical workers

Income (X)	Frequency (f)
0-9	3
10-19	13
20-29	15
30-39	17
40-49	9
50-59	2
60-69	1
Total	60

have been grouped into units of 10, as in the stem-and-leaf display. Alternative groupings could have been used, but Table 2 gives a good summary presentation describing how incomes are distributed. The frequencies of occurrence are provided in the second column of the table, labelled either "Frequency" or f. That is, there are three respondents with incomes between zero and nine thousand dollars of income, thirteen respondents between ten and nineteen thousand dollars of income, and so on. In total, there are $n = 60$ cases, the sum of the frequencies, f.

Additional notes on frequency distributions

You will be working with distributions of the type just illustrated for the rest of this module and in Modules 3 and 4. As you proceed through the examples and exercises, you will become familiar with a variety of ways of presenting frequency distributions. Before proceeding to the next sections of these notes, a few items to note are as follows.

- **Extensive use**. A frequency distribution table is the most common way to present a table summarizing the distribution of a variable for members of a sample or population. These will be used extensively through the semester.

- **Labels**. Values of the variable are presented in the first column of the frequency distribution table and the frequencies in the second column. Make sure you label these columns. If you use the X and f symbols, make sure you define these algebraic symbols, especially X. If you use numerical values as codes for X (as was the case with sex of students in Example 2.11), make sure the codes are labelled in the table, or in the text accompanying the table. Otherwise a reader might not be able to determine the meaning of the values of X.

  For a variable such as income, the values in the X column represent incomes. Incomes are meaningful in numerical form and are ordinarily reported that way, so no further labelling of the categories for this variable is necessary.

- **Include all cases**. When constructing a frequency distribution, make sure you include all cases. In order to ensure you do this, total the frequencies in the second column. Check to make sure this total matches the sample size. One exception is as follows.

- **Missing values**. Some respondents may not provide information about all variables, so some values for some variables may be missing. These missing cases are usually not included in a frequency distribution table. As a result, the sum of the frequencies for the same data set may differ from variable to variable and table to table.

  Another reason for missing values is that some questions may not be relevant for each respondent. In the SPSS tables of hours students worked at jobs, not all students had jobs, so some respondets would not be expected to provide information about the variables JOBHOURS and PAY.

- **Grouping**. When constructing a frequency distribution table, where the variable X takes on many values, it is necessary to decide how to group the data. In the above example of incomes, X is grouped by units of ten, but it could have been grouped by units of five (0-4, 5-9, 10-14, etc.), or some other grouping could have been used. Some guidelines concerning grouping are provided in the text, pp. 101-104.

Proportional distribution

See text, section 4.6, pp. 117-119.

Instead of presenting a table with frequencies of occurrence, the distribution may be presented as a proportional distribution. In a proportional distribution table, the values of the variable X are listed, along with the proportion of cases, p, taking on each value of the variable.

For each value of a variable X, the proportion, p, taking on that value of X is the frequency of occurrence (f) divided by the total number of cases (n).

$$\text{Proportion} = p = \frac{f}{n}$$

In index notation, if f_i cases take on value X_i of a variable, and if there are n cases in total, the proportion of cases in category i is p_i.

$$p_i = \frac{f_i}{n}$$

A proportion p represents the share of total cases taking on any particular value and is reported as a decimal. When reported, it is common to round the proportion to two or three decimal places. For example, in the example of the distribution of credit hours, eight of the fourteen students reported taking fifteen credit hours (category 4 or $i = 4$). The proportion of students in category 4, that is, those taking fifteen credit hours is 0.533.

$$p_4 = \frac{f_4}{n} = \frac{8}{15} = 0.533$$

Proportions must be values between 0 and 1, and the sum of all the proportions is 1. That is, for any particular value X_i,

$$0 \le p_i \le 1$$

and across all k possible values of a variable X_i,

$$p_1 + p_2 + p_3 + ... + p_k = 1.$$

For a proportional distribution table, the sum of the proportions must equal exactly 1. That is, all n cases must be included. If you construct a

proportional distribution where the sum of the proportions is very far from exactly 1.0, you may have calculated incorrectly. If it is 0.999 or 1.002, a small difference of this sort may emerge as a result of rounding. In the latter case, adjust one of the proportions, preferably one of the larger proportions, by the necessary amount to produce a total of exactly 1. See pp. 126-127 of the text for guidelines about rounding.

An example of a proportional distributions follow.

Example 2.13 – Proportional distributions for income. Obtain the proportional distribution tables for the incomes of clerical workers (Table 2).

Answer. The distribution and explanation are as follows.

Income of clerical workers. For the distribution of income of a sample of Saskatchewan clerical workers, there are a total of $n = 60$ respondents. The proportion of those with income of 0-9 thousand dollars is the frequency of 3 individuals with these incomes, divided by the sample size, n. That is $3/60 = 0.050$.

There were 13 respondents with incomes of 10-19 thousand dollars, so the proportion is $13/60 = 0.217$, rounded to three decimal places. Other proportions are obtained in a similar manner and the proportional distribution is reported in Table 3.

Interpretation of table. It may be more straightforward to understand the distribution of income by using a proportional distribution, rather than a frequency distribution. A glance down the proportions column quickly yields the information that 30-39 is the most common category, with a larger proportion of individuals (0.283) than in any other category. The 20-29 and 10-19 categories occur almost as frequently (0.250 and 0.217, respectively), and there are relatively few individuals with incomes under ten thousand dollars (0.050) and over fifty thousand dollars (0.033 and 0.017, for a total of 0.050 as well).

Note that the sum of all the proportions equals exactly 1.000, so all sixty individuals have been taken into account in the table.

Table 3: Proportional distribution of income of sixty Saskatchewan clerical workers

Income (X)	Proportion (p)
0-9	0.050
10-19	0.217
20-29	0.250
30-39	0.283
40-49	0.150
50-59	0.033
60-69	0.017
Total	1.000

Note: Sample size for this sample is $n = 60$.

It is useful to report the sample size in a proportional distribution table, as in Table 3. The sample size cannot be determined by examining the proportional distribution itself. Good statistical analysts report the sample size on which the sample is based, and a reader may wonder the same. As a result, it is advisable to report the sample size for the proportional distribution, either in the table or in the text accompanying the table.

Percentage distribution

See text, section 4.6, pp. 117-119.

Instead of summarizing a data set in a frequency or proportional distribution, it is more common to present it as a percentage distribution table. In such a table, the values of the variable X are listed in the first column and the the percentage of cases taking on each value of X are reported in the appropriate row of the second column of the table.

For each value of a variable X, the percentage, P, taking on that value of X is the frequency of occurrence (f) divided by the total number of cases (n) to produce the proportion of cases with value X. This proportion is mutiplied by 100 per cent to produce the percentage of cases, P, with value

X.

$$\text{Percentage} = P = \frac{f}{n} \times 100$$

Multiplying the proportions by one hundred makes these Ps represent the number of cases per one hundred, or per cent.

In index notation, if f_i cases take on value X_i of a variable, and if there are n cases in total, the percentage of cases taking on value X_i is P_i.

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

The percentage represents the share of total cases taking on any particular value, reported as a percentage, that is, per one hundred.

Percentages are usually reported to the nearest integer, or possibly to one or two decimal places. For example, in the case of credit hours in Table **??**, eight of the fourteen students reported taking fifteen credit hours (category $i = 4$). The percentage of students in category $i = 4$, taking fifteen credit hours, is 0.533.

$$P_4 = \frac{f_4}{n} \times 100 = \frac{8}{15} \times 100 = 53.3\%$$

In a percentage distribution, the percentages must be between 0 and 100, and the sum of all the percentages is one hundred per cent, or 100%. That is, for any particular value X_i,

$$0 \leq P_i \leq 100$$

and across all k possible values of a variable X_i, and across all k categories, the sum of the percentages must be one hundred per cent.

$$P_1 + P_2 + P_3 + ... + P_k = 100$$

If you construct a percentage distribution where the sum of the percentages is very far from 100, you may have calculated incorrectly. If it is 99.7% or 100.2%, this small difference from one hundred per cent may emerge as a result of rounding. In the latter case, adjust one of the percentages, preferably one of the larger percentages, by the necessary amount to produce a total of exactly 100%. See pp. 126-127 of the text for guidelines about rounding and p. 133 for an example of how to address this issue.

Percentage distributions are the most common form for reporting data in the popular media and many statistical reports. A table of a percentage distribution provides readers with a quick, summary view of how members of a sample or population are distributed across values of a variable.

For example, on June 25, 2004, the polling firm, Ipsos-Reid reported "Liberals and Conservatives Tied in Voter Support, But Tories Have Edge in Seats." Accompanying the report, the percentages in Table 4 were provided by Ipsos-Reid. This percentage distribution provides a reader with a quick

Table 4: Percentage support for federal political parties, June 25, 2004

Party supported	Percentage (P)
Conservative	31
Liberal	32
NDP	17
Bloc Quebecois	12
Green	6
Other	2
Total	100

Source: http://www.ipsos-na.com/news/pressrelease.cfm?id=2294, accessed August 5, 2004.

summary of the distribution of respondents by party supported. The reader need not be bothered with the actual number of survey respondents who favoured each party – what is important here is the percentage distribution by party. If the poll is representative of the popuation, these percentages can be used to determine which way the electorate is leaning or to predict election results.

When you analyze percentage distributions, take care to report or determine the sample size. While a percentage distribution provides a generally understandable and quick, summary view, much statistical work requires knowledge of frequencies of occurrences. These can always be obtained from a percentage distribution if the sample size n is known. In the case of the Ipsos-Reid poll, two thousand adult Canadians were surveyed between June 21 and June 23, 2004. Thus $n = 2000$ for this poll.

Examples of percentage distributions follow.

Example 2.14 – Percentage distributions for income. Obtain percentage distribution tables for the incomes of clerical workers (Table 2).

Answer. The distributions and explanations follow.

Income of clerical workers. For the distribution of income of a sample of Saskatchewan clerical workers, there are a total of $n = 60$ respondents. The percentage of those with income of 0-9 thousand dollars is the frequency (3) divided by n, and multiplied by one hundred, that is, $(3/60) \times 100 = 5.0\%$. There were 13 respondents with incomes of 10-19 thousand dollars, so the percentage is $(13/60) \times 100 = 21.7\%$, rounded to one decimal place. Other percentages are obtained in a similar manner and the percentage distribution is reported in Table 5.

Table 5: Percentage distribution of income of sixty Saskatchewan clerical workers

Income (X)	Percentage (P)
0-9	5.0
10-19	21.7
20-29	25.0
30-39	28.3
40-49	15.0
50-59	3.3
60-69	1.7
Total	100.0

Note: Sample size for this sample is $n = 60$.

Since the sample size for this percentage distribution of incomes is not apparent in the table, it is worthwhile reporting the sample size separately from the table. This can be done as part of the table, as in Table 5, or in the text accompanying the table.

Class limits

See text, section 4.7.3, pp. 134-142.

When constructing a frequency or percentage distribution for a continuous variable, it is not possible to list all possible values of the variable. When reporting such distributions, values of the variable must be grouped into intervals. The same is true for a discrete variable with many possible values. In this case, it would be inefficient to list all the values for a variable with more than ten or fifteen possible values. When working with data of this type, values of the variable are grouped into intervals, for example, 0-5, 5-10, 10-15, and so on. The the end points of the intervals are termed the **class limits**.

In many such groupings, there is a gap between the end points that are reported for the intervals. For example, a variable might be grouped into categories such as 0-4, 5-9, 10-14, and so on, leaving a gap of one unit between each interval. The end points initially reported are termed the **apparent class limits** of the intervals. For graphical presentation of data and some other types of data analysis, it is worthwhile to determine the real class limits. The following notes describe how to address this issue.

The data from the the stem-and-leaf display illustrate the problem. Prior to organizing the stem-and-leaf presentation, the variable X was rounded to produce values of income in thousands of dollars. While income is a continuous, ratio scale variable, its values were rounded to the nearest thousand dollars for purposes of constructing the stem-and-leaf display and the later distributions.

In these examples, the variable X is income in thousands of dollars, so the unit for X is one thousand dollars. The intervals used for summarizing these data are:

$$
\begin{array}{c}
0\text{-}9 \\
10\text{-}19 \\
20\text{-}29 \\
30\text{-}39 \\
40\text{-}49 \\
50\text{-}59 \\
60\text{-}69
\end{array}
$$

The values 0, 9, 10, 19, 20, 29, 30, 39, and so on, are the apparent class limits. But there is a gap of one unit between the reported end points of the intervals. This apparent gap emerges, not because there is a gap in the values of the variable, but because of the decision to round incomes into thousands of dollars and group these incomes into the intervals 0-9, 10-19, 20-29, and so on. For example, if categories such as 0-4, 5-9, 10-14, etc. had been used instead, the position of the gaps would change. In order to correct for these gaps between intervals of grouping data, real class limits can be used.

> **Real class limits**. The real class limits are the values of the variable midway between the reported end points, or apparent class limits, of adjacent intervals.

The preceding example will be used to illustrate these. The first interval apparently ends at 9 and the next begins at 10; the midpoint between these two is 9.5. Similarly, the midpoint between 19 and 20 is 19.5. The resulting real class limits for the above intervals are shown in the following display.

Apparent class limits	Real class limits
0-9	-0.5 - 9.5
10-19	9.5 - 19.5
20-29	19.5 - 29.5
30-39	29.5 - 39.5
40-49	39.5 - 49.5
50-59	49.5 - 59.5
60-69	59.5 - 69.5

Calculating and using real class limits eliminates the apparent gap between intervals. This gap originally emerged, not because of any inherent gap in the values of the variable, but as a result of the way the statistical analyst organized these data. If real class limits are used when presenting data as a frequency distribution table or histogram (see text, p. 145), there is no gap between the bars. In this case, there should be no gap since the variable is continuous, meaning that any value of the variable X could possibly occur.

Class limits make sense in two other ways – determining interval width and rounding. These are described in the following notes.

- **Interval width**. The width of the interval is the difference between the upper and lower class limits of the interval. If the apparent class limits of 9, 10, 19, 20, etc. had been used in the above example, it appears that the interval width is nine thousand dollars, that is, $19 - 10 = 9$, $29 - 20 = 9$, and so on. But this is misleading, since the interval from 10 to 19 really represents ten, not nine, units of the variable income (in thousands of dollars). By using the real class limits of 9.5-19.5, 19.5-29.5, and so on, the interval widths in the above discussion each represent ten units of income. That is, $19.5 - 9.5 = 10$, $29.5 - 19.5 = 10$, and so on each ten thousand dollars of income. This is the proper interval width, since the data were organized into intervals representing ten units of the variable.

  Note the oddity of the first real class limits, from -0.5 to 9.5. In order to make the first interval ten units wide, it is necessary to make this interval begin at 0.5, rather than at 0. While this may initially seem incorrect, such a procedure preserves the proper interval width of 10, since $9.5 - (-0.5) = 10$ units of income.

- **Rounding**. The values of income in the stem and leaf display of Example 2.10 were rounded to the nearest ten thousand dollars. A case reported as 9 thousand dollars thus might be associated with an income anywhere between $8,500 and $9,500. An income such as $9,521, just above $9,500, would be rounded up to 10 thousand dollars, and included in the 10-19 interval. In contrast, an income such as $9,498 would be rounded down to 9 thousand dollars, and included in the 0-9 interval. Thus the values of $9,500, or 9.5 thousand dollars, is the proper dividing point between the 0-9 and 10-19 intervals. This is the value of the real class limit between these two intervals. Using real class limits preserves the proper dividing point between categories, when conventional rules on rounding are used to group values of a variable. See the text, pp. 139-142 for a fuller explanation and other examples.

  When using data with large values, eg. 100-199, 200-299, and 300-399, or larger values such as 2,000-2,499, 2,500-2,999, and 3,000-3,499, the difference between the apparent and real class limits is too small to have an appreciable effect on calculations or to distinguish in diagrams. In these situations, it is preferable to construct intervals such as 100-200,

200-300, 300-400, etc. and ignore the difference between apparent and real class limits. That is, the gap between intervals in such cases is only 0.5 units, too small to have much effect on values such as 300, 400, 2,500, 3,000, etc.

If the values of the variable have been grouped into intervals where there is no gap between the end points of the intervals, then the real and apparent class limits are identical. In this case, the apparent class limits are the real class limits, and the discussion is not relevant, since there is no gap between the end points of the intervals. For example, in the following grouping, the apparent and real class limits are identical.

<div align="center">

Apparent and real

class limits

0-5

5-10

10-15

15-20

</div>

That is, using this grouping, there is no gap between the intervals so the values 0, 5, 10, 15, 20, etc. are both the apparent and real class limits. In this case, do not attempt to split the difference between the end points of adjacent intervals - the analyst has already taken care by constructing the intervals so that they meet each other and no gap exists.

Last edited January 21, 2005.