

Mathematics 13 : Elementary Statistics

Unit 1 : Section 3

Lesson 9 :

Descriptive Statistics

Descriptive Statistics

Descriptive statistics refers to a collection of statistical methods used to summarize or describe important characteristics of data.

There are several methods available to assist in the description of data. Each method is designed to provide a distinct insight into the available information.

Lesson 9 :

Descriptive Statistics

Not only is it important to understand how to apply the various descriptive statistics methods, it is also essential to realize what specifically each method reveals about the data.

In general, these descriptive statistics methods are either graphical or numerical in nature.

Lesson 9 :

Descriptive Statistics

The purpose of applying graphical methods in descriptive statistics is to organize and present the data in an informative and useful manner.

The overall pattern or distribution of the data as well as any distinguishing features contained in the data can be displayed through the appropriate use of various graphs, charts, and plots.

Lesson 9 :

Descriptive Statistics

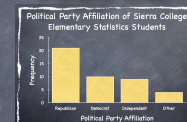
Subsequent lessons will introduce such graphical methods as

Distribution Table

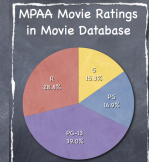
Length of Sentences (in words) from "Harry Potter and the Sorcerer's Stone"

| Length of Sentence | Frequency |
|--------------------|-----------|
| 1 - 5 | 2 |
| 6 - 10 | 4 |
| 11 - 15 | 4 |
| 16 - 20 | 4 |
| 21 - 25 | 3 |
| 26 - 30 | 1 |

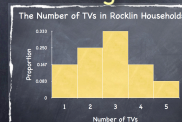
Pareto Chart



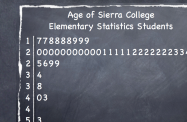
Pie Chart



Histogram



Stem-and-Leaf Plot



Lesson 9 :

Descriptive Statistics

When utilizing graphical methods in descriptive statistics, it is essential that the graphs be clearly labeled.

This would include a title. The title should describe the data as well as the population or sample of interest.

Lesson 9 :

Descriptive Statistics

The purpose of applying numerical methods in descriptive statistics is to clearly convey some crucial characteristic of the data.

In statistics, various numerical methods have been developed to quantitatively measure particular aspects of data.

Some of these aspects include measures of magnitude, central tendency, dispersion, and position within the data.

Lesson 9 :

Descriptive Statistics

Subsequent lessons will introduce such numerical methods as

| | |
|----------------------------|---|
| Frequency f_x | Proportion p_x and \hat{p}_x |
| Mean μ_x and \bar{x} | Standard Deviation σ_x and s_x |
| Median Med | Median Absolute Deviation MAD |
| Mode $Mode$ | Variation Ratio VR |
| Standard Score z and t | Percentile P_k |

Lesson 9 :

Descriptive Statistics

Not only is it important to be capable of calculating the various numerical descriptive statistics methods, it is also essential to understand how to interpret the result.

Lesson 9 :

Lesson 10 :

Distribution Table

Distribution Table

A **distribution table** displays all of the various data values (in one column) along with an associated parameter or statistic of interest (in another column).



Lesson 10 :

Distribution Table

Distribution tables can be constructed using data of any type (qualitative or quantitative, continuous or discrete) and at every level of measurement (nominal, ordinal, interval, or ratio).

Distribution tables describe how the parameter or statistic of interest is allocated (or distributed) amongst all the various data values.

Lesson 10 :

Distribution Table

There are several parameters or statistics available to choose from for inclusion in a distribution table.

In this lesson, we will learn to determine two such results:

the frequency
and
the proportion

Lesson 10 :

Distribution Table

The **frequency** is a count that represents the number of instances a particular data value occurs in the population or sample.

The frequency of the particular data value x is denoted by f_x .

Lesson 10 :

Distribution Table

For instance,

in the sample of milk chocolate plain M&M's shown,



the frequency of green M&M's is 3. Thus, $f_{\text{green}} = 3$.

Lesson 10 :

Distribution Table

The frequency describes exactly how many of that particular data value appear in the population or sample.

When it was determined that there were 3 green M&M's in the previous sample, all we learned was how frequently green M&M's occurred in that sample. The frequency does not provide us sufficient information to determine if this was a small, medium, or large amount of M&M's.

Lesson 10 :

Distribution Table

In some situations, 3 would be considered small. In other situations, 3 would be considered large. Size is a relative notion.

If it is important to the issue under investigation to describe the relative size of various data values, the proportion is the appropriate parameter or statistic to include in a distribution table.

Lesson 10 :

Distribution Table

The **proportion** is a relative frequency that compares the frequency of a particular data value to the overall population size or total sample size.

The proportion is calculated by dividing the frequency by the total of all the frequencies.

$$\text{Proportion} = \frac{\text{Frequency}}{\text{Total of All Frequencies}}$$

Lesson 10 :

Distribution Table

The population proportion for a particular data value X is denoted by

$$P_x = \frac{f_x}{N}$$

The sample proportion for a particular data value x is denoted by

$$\hat{p}_x = \frac{f_x}{n}$$

Lesson 10 :

Distribution Table

For instance,

in the sample of milk chocolate plain M&M's shown,



$$f_{\text{green}} = 3$$

$$n = 24$$

the proportion of green M&M's is

$$\hat{p}_{\text{green}} = \frac{f_{\text{green}}}{n} = \frac{3}{24} = 0.125 = 12.5\%$$

Lesson 10 :

Distribution Table

Results of proportion calculations can be represented as a fraction, decimal, or percent.

It is customary to express the final decimal result to at least the thousandth place value (three places after the decimal). This enables the final percentage result to be expressed to at least the tenth of a percent place value.

Lesson 10 :

Distribution Table

Data values can be listed in the distribution table either individually or in groups.

Either way, the approach used to list the data values must consider the number of rows required to display all of the various data values.

The number of rows in a distribution table must be reasonable in order to clearly reveal the distribution of the data.

Lesson 10 :

Distribution Table

Too many rows can make the table difficult to decipher and typically dilutes the essential elements of the distribution.

Too few rows can overly condense the table which quite often conceals crucial details.

In general, a reasonable number of rows to display in a distribution table ranges somewhere from 4 on the low side up to 24 on the high side.

Lesson 10 :

Distribution Table

Each and every data value must be included in the distribution table and belong to one and only one of the rows.

Lesson 10 :

Example 1

Use the political party affiliation data (Q2) from our Sierra College Elementary Statistics Student Survey to complete the distribution table given below and describe the result.

| Political Party | Frequency | Percentage |
|-----------------|-----------|------------|
| | | |
| | | |
| | | |
| | | |

Lesson 10 : Distribution Table

Example 1

Political Party Affiliation of Sierra College Elementary Statistics Students

| Political Party | Frequency | Percentage |
|-----------------|-----------|------------|
| Democrat | 10 | 22.7% |
| Republican | 21 | 47.7% |
| Independent | 9 | 20.5% |
| Other | 4 | 9.1% |

Lesson 10 : Distribution Table

Example 1

Description :

Nearly half, 47.7%, of Sierra College Elementary Statistics Students surveyed identified themselves as republican. There were about the same percentage of students choosing democrat, 22.7%, as those who selected independent, 20.5%. The remaining 9.1% picked other.

Lesson 10 : Distribution Table

Lesson 11 :

Pareto Chart

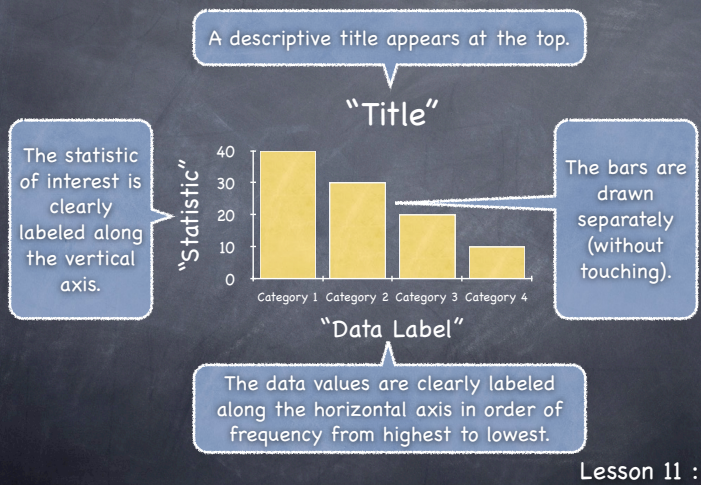
Pareto Chart

A **Pareto chart** is a bar graph constructed from qualitative data where the categories are arranged along the horizontal axis in order, from highest to lowest, of frequency or proportion which is displayed along the vertical axis.

The bars are drawn separately (without touching) in order to convey that the data represents distinct, separate categories.

Lesson 11 :

Pareto Chart



Pareto Chart

A Pareto chart is the preferred graphical method of displaying the distribution of the data whenever the data under examination is qualitative, and it is important to the issue under investigation to easily illustrate which categories are the most frequently occurring.

Lesson 11 :

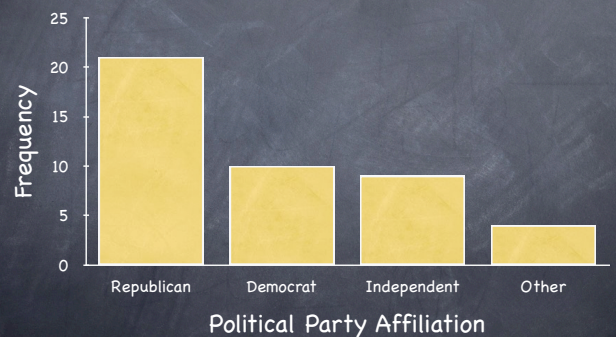
Example 1

Use the political party affiliation data (Q2) from our Sierra College Elementary Statistics Student Survey to construct a frequency Pareto chart and describe the result.

Lesson 11 : Pareto Chart

Example 1

Political Party Affiliation of Sierra College Elementary Statistics Students



Lesson 11 : Pareto Chart

Example 1

Description:

Most of the 44 Sierra College Elementary Statistics Students surveyed, 21, identified themselves as republican. There were about the same number of students choosing democrat, 10, as those who selected independent, 9. The remaining 4 picked other.

Lesson 11 : Pareto Chart

Lesson 12 :

Pie Chart

Pie Chart

A **pie chart** presents the data in the form of a circle that is divided into sections. Each section corresponds to a particular data value. The area of each section is proportional to the proportion of the category that section represents.

Pie charts are primarily constructed for qualitative data where the relative size of each section depicts the percentage.

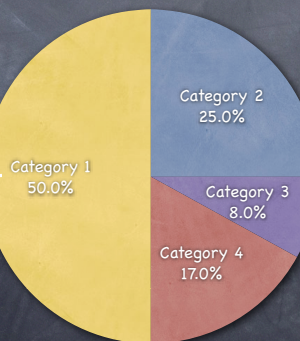
Lesson 12 :

Pie Chart

A descriptive title appears at the top.

"Title"

Each individual section is clearly labeled with the corresponding data value and percentage.



The relative size of each section depicts the percentage of each category.

Lesson 12 :

Pie Chart

In statistics, drawing a pie chart by hand is rarely done. In practice, pie charts are constructed using software applications.

To describe the distribution of qualitative data represented in a pie chart, the various categories observed are merely mentioned along with their associated percentages.

Lesson 12 :

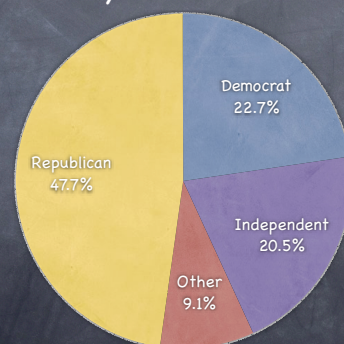
Example 1

Use the political party affiliation data (Q2) from our Sierra College Elementary Statistics Student Survey to construct a percentage pie chart and describe the result.

Lesson 12 : Pie Chart

Example 1

Political Party Affiliation of Sierra College Elementary Statistics Students



Lesson 12 : Pie Chart

Example 1

Description:

Nearly half, 47.7%, of Sierra College Elementary Statistics Students surveyed identified themselves as republican. There were about the same percentage of students choosing democrat, 22.7%, as those who selected independent, 20.5%. The remaining 9.1% picked other.

Lesson 12 : Pie Chart

Copyright © 2012 Supportive Learning Solutions, LLC
All rights reserved.

Your solutions should be clear, complete, and sufficiently detailed in order to demonstrate your understanding and communicate your reasoning and method of solving the problem.

Exercise 1

The final grades received by students in Professor Brown's Experimental Psychology (PSYC 105) course last semester are given below.

B, B, C, B, C, A, A, B, D, C, B, F, B, A, C, D, C, and B

Construct a proportion distribution table with this data.

Exercise 2

University Hospital conducted a study in order to determine the blood type distribution of its patients. During each of the hospital's three laboratory work shifts (day, evening, and overnight), a random sample of fifteen patients who had blood tests performed during that shift was selected. The blood types for these randomly selected patients are given below.

Day Shift : O, A, A, B, O, A, A, O, A, A, O, AB, O, O, A
Evening : A, O, O, O, A, O, O, B, O, AB, O, O, A, A, B
Overnight : O, O, O, O, A, B, A, A, O, A, A, A, B, O, O

Construct a percentage distribution table with this data.

Exercise 3

University Hospital conducted a study in order to determine the blood type distribution of its patients. During each of the hospital's three laboratory work shifts (day, evening, and overnight), a random sample of fifteen patients who had blood tests performed during that shift was selected. The blood types for these randomly selected patients are given below.

Day Shift : O, A, A, B, O, A, A, O, A, A, O, AB, O, O, A
Evening : A, O, O, O, A, O, O, B, O, AB, O, O, A, A, B
Overnight : O, O, O, O, A, B, A, A, O, A, A, A, B, O, O

Construct a proportion Pareto chart with this data and describe the result.

Exercise 4

University Hospital conducted a study in order to determine the blood type distribution of its patients. During each of the hospital's three laboratory work shifts (day, evening, and overnight), a random sample of fifteen patients who had blood tests performed during that shift was selected. The blood types for these randomly selected patients are given below.

Day Shift : O, A, A, B, O, A, A, O, A, A, O, AB, O, O, A
Evening : A, O, O, O, A, O, O, B, O, AB, O, O, A, A, B
Overnight : O, O, O, O, A, B, A, A, O, A, A, A, B, O, O

Construct a proportion pie chart with this data.

Exercise 5

Use the genre data collected in the Movie Database Sample to construct a frequency Pareto chart.

Exercise 6

Use the mathematics ability rating data (Q3) from our Sierra College Elementary Statistics Student Survey to construct a percentage pie chart and describe the result.

Your solutions should be clear, complete, and sufficiently detailed in order to demonstrate your understanding and communicate your reasoning and method of solving the problem.

Exercise 1

The final grades received by students in Professor Brown's Experimental Psychology (PSYC 105) course last semester are given below.

B, B, C, B, C, A, A, B, D, C, B, F, B, A, C, D, C, and B

Construct a proportion distribution table with this data.

**Final Grade Distribution in Professor Brown's Experimental Psychology
(PSYC 105) Course Last Semester**

| Final Grade | Proportion | |
|-------------|------------|----------------------|
| A | 16.7% | = $3/18 \cdot 100\%$ |
| B | 38.9% | = $7/18 \cdot 100\%$ |
| C | 27.8% | = $5/18 \cdot 100\%$ |
| D | 11.1% | = $2/18 \cdot 100\%$ |
| F | 5.6% | = $1/18 \cdot 100\%$ |

Exercise 2

University Hospital conducted a study in order to determine the blood type distribution of its patients. During each of the hospital's three laboratory work shifts (day, evening, and overnight), a random sample of fifteen patients who had blood tests performed during that shift was selected. The blood types for these randomly selected patients are given below.

Day Shift : O, A, A, B, O, A, A, O, A, A, O, AB, O, O, A
 Evening : A, O, O, O, A, O, O, B, O, AB, O, O, A, A, B
 Overnight : O, O, O, O, A, B, A, A, O, A, A, A, B, O, O

Construct a percentage distribution table with this data.

**Blood Type Distribution of Patients at
University Hospital**

| Blood Type | Percentage |
|------------|------------|
| O | 46.7% |
| A | 37.8% |
| B | 11.1% |
| AB | 4.4% |

$$= 21/45 \cdot 100\%$$

$$= 17/45 \cdot 100\%$$

$$= 5/45 \cdot 100\%$$

$$= 2/45 \cdot 100\%$$

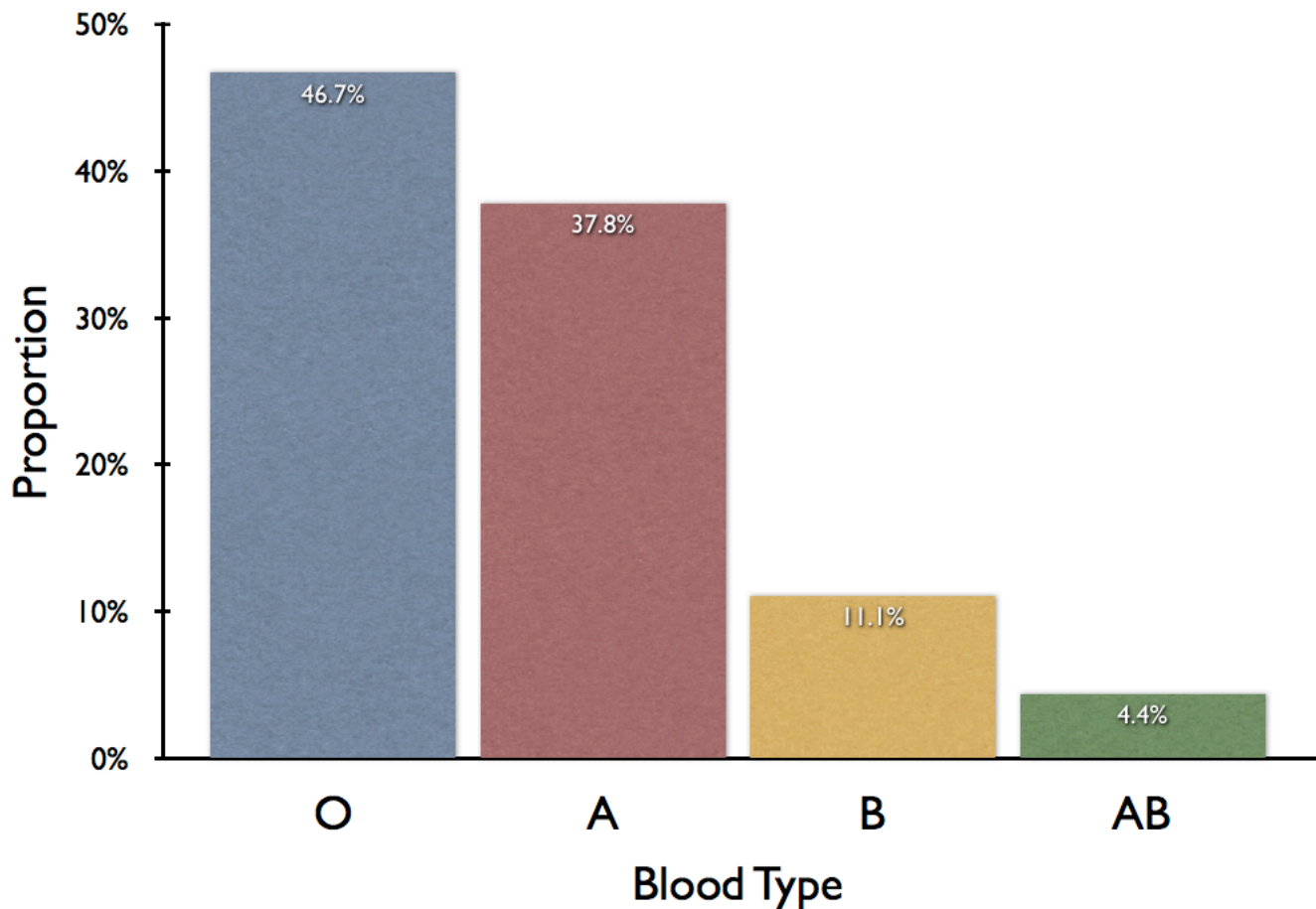
Exercise 3

University Hospital conducted a study in order to determine the blood type distribution of its patients. During each of the hospital's three laboratory work shifts (day, evening, and overnight), a random sample of fifteen patients who had blood tests performed during that shift was selected. The blood types for these randomly selected patients are given below.

Day Shift : O, A, A, B, O, A, A, O, A, A, O, AB, O, O, A
Evening : A, O, O, O, A, O, O, B, O, AB, O, O, A, A, B
Overnight : O, O, O, O, A, B, A, A, O, A, A, A, B, O, O

Construct a proportion Pareto chart with this data and describe the result.

Blood Type Distribution of Patients at University Hospital



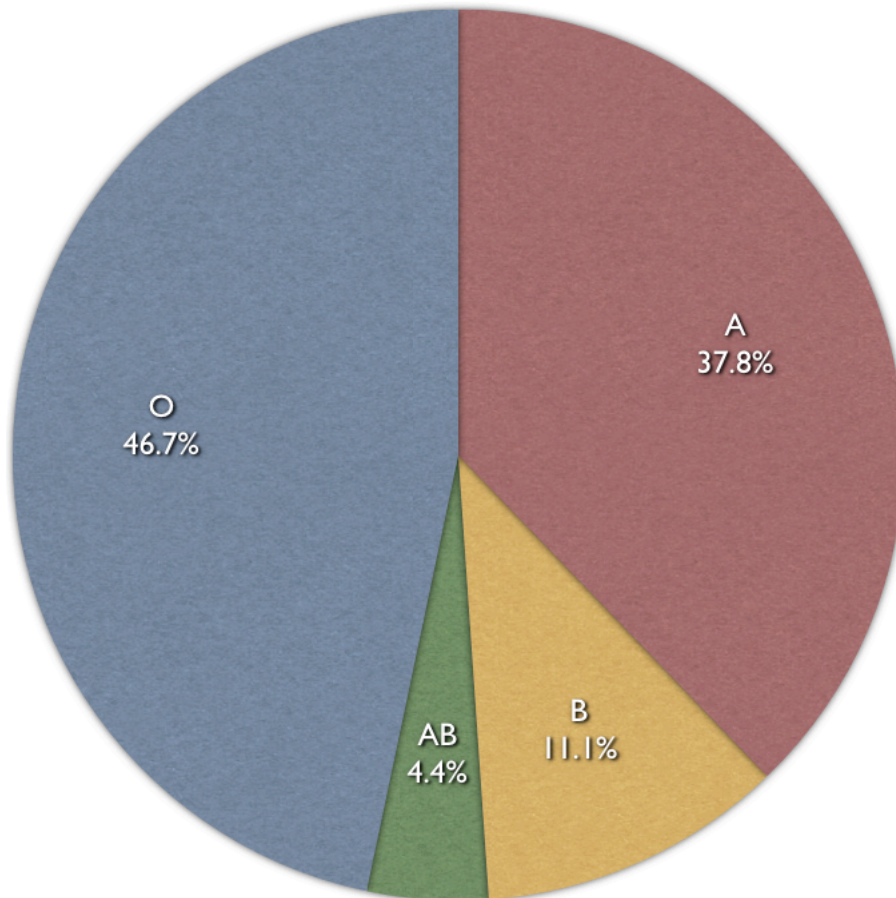
Most of University Hospital's patients, 46.7%, have blood type "O". A little fewer, 37.8%, have blood type "A". A lot fewer, 11.1%, have type "B" blood, and very few, 4.4%, have blood type "AB".

Exercise 4

University Hospital conducted a study in order to determine the blood type distribution of its patients. During each of the hospital's three laboratory work shifts (day, evening, and overnight), a random sample of fifteen patients who had blood tests performed during that shift was selected. The blood types for these randomly selected patients are given below.

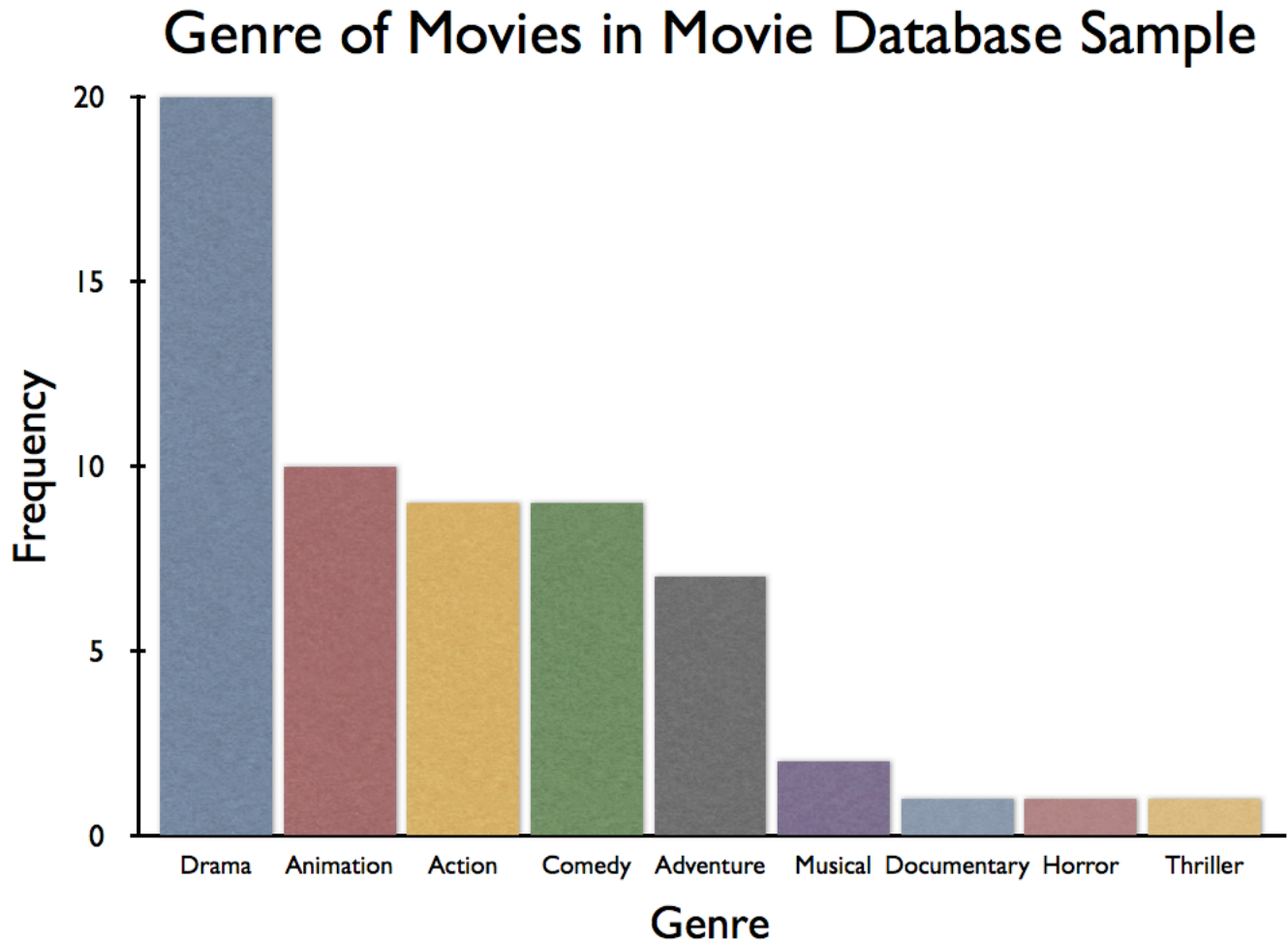
Day Shift : O, A, A, B, O, A, A, O, A, A, O, AB, O, O, A
Evening : A, O, O, O, A, O, O, B, O, AB, O, O, A, A, B
Overnight : O, O, O, O, A, B, A, A, O, A, A, A, B, O, O

Construct a proportion pie chart with this data.

Blood Type Distribution of Patients at University Hospital

Exercise 5

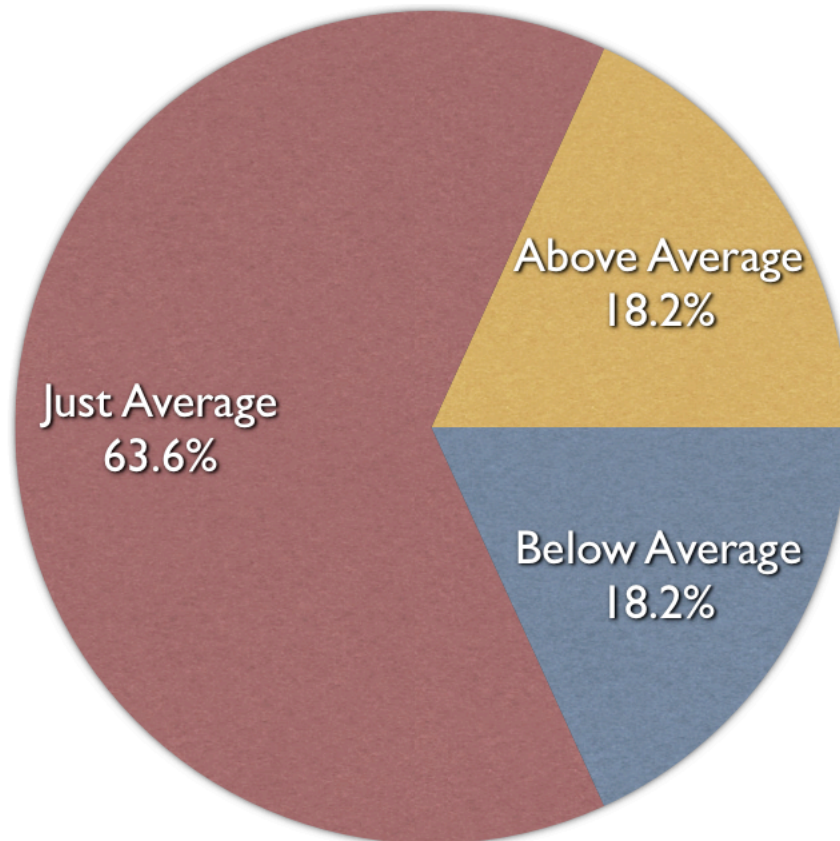
Use the genre data collected in the Movie Database Sample to construct a frequency Pareto chart.



Exercise 6

Use the mathematics ability rating data (Q3) from our Sierra College Elementary Statistics Student Survey to construct a percentage pie chart and describe the result.

Mathematics Ability Rating of Sierra College Elementary Statistics Students Surveyed



The majority of Sierra College Elementary Statistics students surveyed, 63.6%, rated their ability as a mathematics student as just average. A much lower but equal percentage of Sierra College Elementary Statistics students surveyed rated their ability as a mathematics student as above average, 18.2%, and below average, 18.2%.