IB Math SL **Intro to Statistics**

Notes – Unit 3

1. **A bit of History:**

*Visual representations of scientific data have been used for centuries - Copernicus drew schematic sketches of planetary orbits around the sun, for example - but the visual representation of numerical data in the form of graphs is a more recent development. In 1786,* ***William Playfair****, a Scottish economist, published* The Commercial and Political Atlas*, which contained a variety of economic statistics presented in graphs.*

<http://www.visionlearning.com/library/module_viewer.php?mid=156>

1. **Basic Concepts:**



This is often a very large, sometimes infinite, group. In order to collect information, one could collect data from EVERY member of the population, however that is often not feasible.



1. **Discrete Data:**

Data that is an exact number $(x\in N)$

Raw Data: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Example:

IB Math grades (Raw Data) “Processed” into a Frequency Table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 4 | 4 | 5 | 6 | 7 | 4 | 5 |
| 3 | 3 | 3 | 7 | 5 | 3 | 3 |
| 4 | 6 | 4 | 6 | 5 | 3 | 3 |
| 5 | 4 | 3 | 4 | 6 | 3 | 4 |
| 3 | 6 | 7 | 5 | 5 | 5 | 4 |
| 6 | 2 | 7 | 4 | 5 | 4 | 7 |
| 2 | 4 | 5 | 6 | 6 | 4 | 7 |

|  |  |  |
| --- | --- | --- |
| Grade, xi | Tally marks | Frequency, fi |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |
| 7 |  |  |

Looking at the frequency table, f5 = \_\_\_\_\_\_\_.

1. **Continuous Data:**

Data that is not an exact number and is, in fact, rounded $(x\in R)$ due to an inability to get an exact number.

Example: Classify the following as either discrete or continuous data.

1. The number of fish caught by an angler.
2. The length of the fish that were caught.
3. The time taken to catch the fish.
4. The number of friends that angle took with him.
5. **Grouped Data:**

This is a convenient way to look at a wide range of data points. In your frequency table, you will use a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ as opposed to a single value. As a guide, shoot for the number of class intervals to be “approximately $\sqrt{the number of data points}$”1 broken into convenient sizes.

Example:

Weights of 24 tomatoes (to nearest .1g) “Processed” into a Frequency Table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 33.5 | 39.1 | 29.8 | 35.5 | 32.3 | 37.6 |
| 29.4 | 30.6 | 35.0 | 41.2 | 33.4 | 34.9 |
| 37.1 | 36.8 | 28.7 | 32.7 | 38.0 | 30.7 |
| 31.9 | 35.8 | 37.2 | 34.4 | 39.5 | 36.8 |

|  |  |  |
| --- | --- | --- |
| Class Interval | Tally marks | Frequency (tally#) |
| 25.0- 29.9 |  |  |
| 30.0-34.9 |  |  |
| 35.0-39.9 |  |  |
| 40.0-44.9 |  |  |

1. **Frequency Histograms:**

A visual representation of the frequency table (bar chart).

Example:

The following frequency table represents the heights of 53 male employees at a hospital, rounded to the nearest centimeter. *(Is this discrete or continuous data?)*

|  |  |
| --- | --- |
| Class Interval | Frequency |
| 155-159 | 5 |
| 160-164 | 8 |
| 165-169 | 8 |
| 170-174 | 10 |
| 175-179 | 10 |
| 180-184 | 6 |
| 185-189 | 3 |
| 190-194 | 2 |
| 195-199 | 1 |

 **Draw a Frequency Histogram (Bar Chart):**

**Use of the GDC**

The GDC allows us to visually represent our data without having to first arrange it into a frequency table. To produce a frequency histogram in the calculator:

1. Enter the raw data into L1.
2. 2nd STAT PLOT
3. Highlight the bar graph
4. WINDOW
5. Set x-min and x-max to accommodate your range of data
6. Set x-scl to match the size of the class interval
7. Set y-min = -2 and y-max to accommodate your range of frequencies
8. GRAPH
9. TRACE will allow you to see the height of each bar (frequency)

1 Mr. Ledger

1. **The Cumulative Frequency Table:**

This table has an additional column – one where you total up the frequency for each additional class interval.

|  |  |  |
| --- | --- | --- |
| Grade, xi | Frequency, fi | Cumulative Frequency, F1 |
| 1 | 0 |  |
| 2 | 2 |  |
| 3 | 10 |  |
| 4 | 13 |  |
| 5 | 10 |  |
| 6 | 8 |  |
| 7 | 6 |  |

Example: This is the same data from last class (displaying IB exam scores).

|  |  |  |
| --- | --- | --- |
| Upper Class Boundary | Frequency | Cumulative Frequency, F1 |
| ≤154.5 | 0 |  |
| ≤159.5 | 5 |  |
| ≤164.5 | 8 |  |
| ≤169.5 | 8 |  |
| ≤174.5 | 10 |  |
| ≤179.5 | 10 |  |
| ≤184.5 | 6 |  |
| ≤189.5 | 3 |  |
| ≤194.5 | 2 |  |
| ≤199.5 | 1 |  |

Example: This is the same data from earlier (displaying height of male hospital employees). Note: The class intervals have been written to display the upper boundary of the interval, so in addition, we have added a top row to indicate a height at which no employee was under.

1. **The cumulative Frequency Curve:**

A cumulative frequency curve plots, and connects, the cumulative frequency vs. the upper class boundary. Ideally, one would connect the points with a nice curve. When we use our calculator, however, our only option is to use the “line graph” option which connects each with a straight line.

A cumulative frequency curve will allow us to make predictions and calculate percentiles.

1. **Percentiles and Quartiles:**

Percentile: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Example: The 90th percentile is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Quartile: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The Lower Quartile = 25th percentile

The Upper Quartile = 75th percentile

\*\* we will talk more about quartile when we discuss median later on.

The following are Paper 1 sample questions:

**1.** A test marked out of 100 is written by 800 students. The cumulative frequency graph for the marks is given below.



1. Write down the number of students who scored 40 marks or less on the test.

 (b) The middle 50% of test results lie between marks *a* and *b*, where *a* < *b*. Find *a* and *b*.

**2.** A box contains 100 cards. Each card has a number between one and six written on it.

The following table shows the frequencies for each number.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Number** | 1 | 2 | 3 | 4 | 5 | 6 |
| **Frequency** | 26 | 10 | 20 | *k* | 29 | 11 |

(a) Calculate the value of *k*.

(2)

(b) Complete the cumulative frequency table.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Number** | 1 | 2 | 3 | 4 | 5 | 6 |
| **Frequency** |  |  |  |  |  |  |

(c) Produce the cumulative frequency graph.

 

(d) Find

(i) the upper quartile;

(ii) the 10th percentile.

**3.** The cumulative frequency curve shows the marks obtained in an examination by a group of 200 students.

 (a) Use the cumulative frequency curve to complete the frequency table below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mark (*x*)  | 0  *x* < 20 | 20  *x* < 40 | 40  *x* < 60 | 60  *x* < 80 | 80  *x* < 100 |
| Number ofstudents  | 22 |  |  |  | 20 |

(b) Forty percent of the students fail. Find the pass mark.

****