Quantitative Tools - Level 1 - Fall 2015
Louis de Charsonville

Lecture 5

Core Values

1 Mode

Definition 1.1. Mode
The mode is the value that appears most often in a set of data. It is also the most likely value for a variable.

1.1 Qualitative variables & discrete quantitative variable

For these variables, finding the mode is straight-forward : this is the value which has the highest frequency.

Example 1.1. Student's grade in Statistics

Grades	7	8	9	10	11	12	13	14	16	19
Frequencies	1	2	4	6	9	10	8	7	1	2

Table 1: Grades

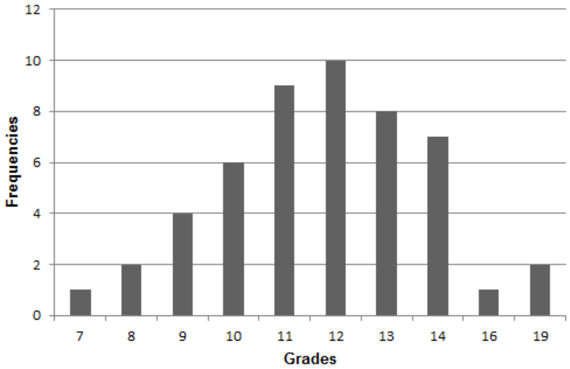

Figure 1: Bar plot of Grades

The Mode is 12.

1.2 Continuous quantitative variables

For continuous variables, it is a bit more tricky to find the mode since the strict definition of the mode does not make sense for continuous variable : two values will never be the same and each value will occur precisely once. In order to compute, the mode we discretize the quantitative variable, that is we divide it into classes. Then, finding the **modal class** is straight-forward : it is the class with the highest frequency.
The mode can then be defined as :

- Approximatively, the midpoint of the modal class

- the estimated mode with the formula :

$$A = B^l + \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})} * w$$

$$with \begin{cases} B^L & \text{is the lower class boundary of the modal class} \\ f_{m-1}, f_m, f_{m+1} & \text{respectively the frequency of the group before the modal class, the frequency} \\ & \text{of the modal class, the frequency of the group after the modal group} \\ w & \text{is the group width} \end{cases}$$

But the mode of discretized continuous variable is class-sensitive in the sense that a different choice of bins could lead to a different mode. See the example below

Example 1.2. Wages in a firm Imagine that the distribution of the wages in a firm is the following

Wages	Frequencies	Corrected frequencies (or Densities)
10000 - 15000	4	4
15000 - 20000	6	6
20000 - 25000	10	10
25000 - 30000	20	20
30000 - 35000	15	15
35000 - 40000	12	12
40000 - 50000	4	2

Table 2: Wages in a firm

The modal class is the class for wages between \$25,000 and \$30,000. An approximative mode is \$27,500. Using the formula to compute the precise mode, we find :

$$\begin{aligned} Mode &= 25000 + 5000 * \frac{20 - 10}{20 - 10 + 20 - 15} \\ &= \$28,333 \end{aligned}$$

The mode can also be found graphically using the intersection of the two black dotted line in the graph below :

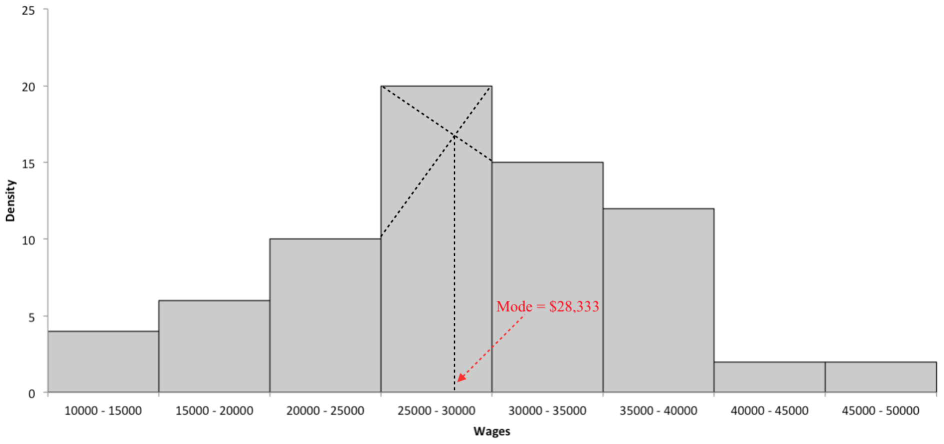

Figure 2: Graphical Determination of the Mode

A drawback of the Mode is that it is sensitive to the classes chosen, see the example below.

Example 1.3. Wages in a firm We use the same distribution of wages but just merge some bins. We got the following wages table :

Wages	Frequencies
10000 - 20000	10
20000 - 30000	30
30000 - 40000	27
40000 - 50000	4

Table 3: Wages in a firm

The modal class is the class for wages between $20,000 and $30,000.
We use the formula to compute the precise mode, and we get :

$$\begin{aligned} Mode &= 20000 + 10000 * \frac{30-10}{30-10+30-27} \\ &= \$28,695 \end{aligned}$$

2 Median

Definition 2.1. Median The median is the value which splits the distribution of a variable in two equal parts. For a distribution of wages, for example, the median is the wage below which 50% of salaries are situated. Equivalently, it is the wage above which 50% of salaries are situated.

2.1 Discrete quantitative variable

Consider a series of N observations sorting in ascending order.

- If N is odd, then N can be written $N = 2 * k + 1$ finding the median is straightforward, the median is the $(k+1)^{th}$ number. Indeed, there is k values before and after the $(k+1)^{th}$ number.
- If N is even, then N can be written $N = 2 * k$, and no values of the series can be considered as the median. By convention we usually consider that the median is the midpoints between the k^{th} and $(k+1)^{th}$ numbers.

However, sometimes, the median may not exist (or cannot be found precisely). Consider the following example

Example 2.1. Student's grade in Statistics

Grades	7	8	9	10	11	12	13	14	16	19
Frequencies	1	2	4	6	9	10	8	7	2	2

Table 4: Grades

There are 51 students. So the median grade that splits the students distribution in two is the grade of the 26^{th} students. Here, the 26^{th} student got 12. But 19 students got more than 12 and 22 students got less than 12.

2.2 Continuous quantitative variables

The median for a continuous variable grouped in classes belongs to a class (or bin). The way to find it is the following :

- Find the class to which the median belong
- Derive the median from the boundaries of the class using the even distribution within the class hypothesis.

Example 2.2. Wages in a firm

Wages bins	Frequencies	Cumulative frequencies
250 - 350	24	24
350 - 400	32	56
400 - 450	51	107
450 - 500	70	177
500 - 525	47	224
525 - 550	41	265
550 - 600	70	335
600 - 650	58	393
650 - 700	40	433
700 - 800	24	457
800 - 950	3	460

The median income is the income so that $\frac{460}{2} = 230$ workers earn less and 230 workers earn more. The median is the mean of the 230^{th} and 231^{th} workers' income The 230^{th} workers' income belong to the $525 - 550$ bin, so does the 231^{th} workers' income. We imagine that the wages are evenly distributed within the class. So that the 230^{th} worker's income is $525 + \frac{6}{41} = 528.66$, and the 230^{th} worker's income is $525 + \frac{7}{41} = 529.27$.
The median income is therefore $\frac{528.66+529.27}{2} = 528.97$

2.3 Properties of the median

The median has two very interesting properties :

- It is indifferent to extreme values.
- The sum of absolute deviation of a series to a constant value is minimal when this constant value is the median.

3 Means

3.1 Arithmetic mean

3.1.1 Definitions

The arithmetic mean is the most known of core values and it is also used very frequently.

Definition 3.1. Simple arithmetic mean
The arithmetic mean of $x_1, x_2, ..., x_n$ values appearing once in a sample is :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Definition 3.2. Simple arithmetic mean
The arithmetic mean of $x_1, x_2, ..., x_n$ values appearing w_i times, called the weighted arithmetic mean is :

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i * x_i}{\sum_{i=1}^{n} w_i}$$

$$\bar{x} = \sum_{i=1}^{n} \alpha_i * x_i$$

$$\text{with } \alpha_i = \frac{w_i}{\sum_{i=1}^{n} w_i}$$

3.1.2 Properties

The arthimetic mean has very interesting properties :

- If all the values are equal then the arithmetic mean equals one of them

$$\text{if } x_1 = x_2 = ... = x_n \quad \text{then } \bar{x} = x_1$$

- The mean of a sum equals the sum of means

$$\bar{x+y} = \bar{x} + \bar{y}$$

- The sum of all deviations from mean equals to zero

$$\bar{x-\bar{x}} = \bar{x} - \bar{x} = 0$$

- If we add a number b to all x_i then the mean is also increased by b.

$$\bar{x+b} = \bar{x} + b$$

- If we multiply all the x_i by a number a then the mean is also multiplied by a

$$\bar{ax+b} = a * \bar{x} + b$$

- The sum of square deviations of the $x_i, i \in [1, n]$ from a number a reachs a minimal value when a is the mean of the $x_i, i \in [1, n]$.

TO BE CONTINUED

3.1.3 Structural effects & common pitfalls of means

3.2 Geometric mean

3.2.1 Definitions

3.2.2 Properties

3.3 Harmonic mean

3.3.1 Definitions

3.3.2 Properties

* * *