

Statistics

Lecture 13: | Histogram

Prof. Dr. Mostafa Elhosseini

<https://youtube.com/drmelhosseini>

| Histogram

- Histogram is an accurate graphical representation of the distribution of numerical data
- It is an estimate of the probability distribution of a continuous variable
- It is a kind of bar plot
- Used for visualizing distribution of images

| Numerical Data

- Some numerical data is obtained by **counting (Discrete)** to determine the value of a variable
 - Number of traffic citations a person received during the last year
 - Number of persons arriving for service during a particular period
- Other data is obtained by taking **measurements – Continuous – entire interval**
 - weight of an individual,
 - reaction time to a particular stimulus).
- The prescription for **drawing a histogram** is generally **different** for these **two cases**

| Histogram for Discrete Data

- A **discrete variable x** almost always results from **counting**, in which case possible values are **0, 1, 2, 3, . . .** or some **subset of these integers**
- The **frequency** of any particular x value is the **number of times that value occurs in the data set**.
- The **relative frequency** of a value is the **fraction or proportion** of times the value occurs

| Histogram for Discrete Data

- First, **determine the frequency and relative frequency** of each x value.
- Then **mark possible x values on a horizontal scale**.
- Above each value, **draw a rectangle** whose **height** is the **relative frequency** (or alternatively, the **frequency**) of that value.

| Histogram for Discrete Data - Example

- 1, 4, 2, 1, 0, 2, 1, 0, 1, 2, 1, 0, 0, 2, 2, 3, 1, 1, 3, 6

| Histogram for Discrete Data - Example

■ 1, 4, 2, 1, 0, 2, 1, 0, 1, 2, 1, 0, 0, 2, 2, 3, 1, 1, 3, 6

Number	Frequency
0	4
1	7
2	5
3	2
4	1
5	0
6	1

| Histogram for Discrete Data

- If numbers are not repeated or not accurate, so it is better to use **ranges // buckets // bins // category**