Statistics

Lecture 15: |Histogram Example

Prof. Dr. Mostafa Elhosseini

https://youtube.com/drmelhosseini

Power companies need information about customer usage to obtain accurate forecasts of demands. Investigators from Wisconsin Power and Light determined energy consumption (BTUs) during a particular period for a sample of 90 gas-heated homes. An adjusted consumption value was calculated as follows:

$$\text{adjusted consumption} = \frac{\text{consumption}}{(\text{weather, in degree days})(\text{house area})}$$

This resulted in the accompanying data (part of the stored data set FURNACE.MTW available in MINITAB), which we have ordered from smallest to largest).

2.97	4.00	5.20	5.56	5.94	5.98	6.35	6.62	6.72	6.78
6.80	6.85	6.94	7.15	7.16	7.23	7.29	7.62	7.62	7.69
7.73	7.87	7.93	8.00	8.26	8.29	8.37	8.47	8.54	8.58
8.61	8.67	8.69	8.81	9.07	9.27	9.37	9.43	9.52	9.58
9.60	9.76	9.82	9.83	9.83	9.84	9.96	10.04	10.21	10.28
10.28	10.30	10.35	10.36	10.40	10.49	10.50	10.64	10.95	11.09
11.12	11.21	11.29	11.43	11.62	11.70	11.70	12.16	12.19	12.28
12.31	12.62	12.69	12.71	12.91	12.92	13.11	13.38	13.42	13.43
13.47	13.60	13.96	14.24	14.35	15.12	15.24	16.06	16.90	18.26

Histogram for Continuous Data

- Determine the **frequency** and **relative frequency** for each class.
- **Mark** the **class boundaries on a horizontal** measurement axis.
- **Above** each **class interval**, **draw a rectangle** whose height is the corresponding **relative frequency** (or **frequency**).

Example

- 1, 3, 27, 32, 5, 63, 26, 25, 18, 16, 4, 45, 29, 19, 22, 51, 58, 9, 42, 6

Bucket	Frequency
0 – 9	6
10 – 19	3
20 – 29	5
30 – 39	1
40 - 49	2
50 - 59	2
60 - 69	1

Example

Bucket	Frequency
0 – 9	6
10 – 19	3
20 – 29	5
30 – 39	1
40 - 49	2
50 - 59	2
60 - 69	1

Bucket	Frequency
0 – 9 **0-<10**	6
10 – 19 **10-<20**	3
20 – 29 **20-<30**	5
30 – 39 **30-<40**	1
40 – 49 **40-<50**	2
50 – 59 **50-<60**	2
60 – 69 **60-<70**	1

Histogram for Discrete Data

- In general, it is good to have more intervals rather than fewer, but it is also good to have large numbers of sample points in the intervals. Striking the **proper balance** is a matter of judgment and of **trial and error**
- There is no hard-and-fast rule as to how to choose the endpoints of the **class intervals**
- **Density** presents the **relative frequency divided by the class width**
- When the **classes are of equal width**, the frequencies, relative frequencies, and densities are **proportional to one another**

Histogram for Discrete Data

- In some cases, **histograms are drawn** with class intervals of **differing widths**. This may be done when it is desired for the histogram to have a **smoother appearance**, or when the **data come** in the form of a frequency table in which the classes have **unequal widths**
- When the **class intervals are of unequal widths**, the heights of the rectangles must be set equal to the **densities**. The **areas** of the rectangles will then be the **relative frequencies**.
- Examples power forecast Devore page 16 BTU
- How to implement with Python