Lecture 03

Dotplots:

- **a graphical descriptive statistic**
- **applicable given univariate data $x_1, \ldots, x_n$**
- **able to observe centrality, dispersion, outliers**
- **not so widely used (histograms are better)**

Histograms:

- **a graphical descriptive statistic**
- **applicable given univariate data $x_1, \ldots, x_n$**
- **able to observe centrality, dispersion, outliers**
- **we encourage intervals of equal length**
- **generated by statistical software**

Histograms (we illustrate by hand):

- **data are weights of students in kg: 47, 55, 79, 63, 64, 67, 54, 59, 58, 84, 70, 61, 65, 59**

Issues in constructing histograms:

- **always label axes and provide a title**
- **how many intervals should be chosen?**
- **be aware of the scale of the vertical axis**
- **handling intervals that are not of equal length**

Sample mean $\bar{x}$:

- **a numerical descriptive statistic of centrality**
- **applicable given univariate data** $x_1, \ldots, x_n$
- $\bar{x} = \frac{x_1 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum x_i}{n}$

Sample median $\tilde{x}$:

- **a numerical descriptive statistic of centrality**
- **applicable given univariate data** $x_1, \ldots, x_n$
- $\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \textbf{if } n \textbf{ odd} \\ \left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)}\right)/2 & \textbf{if } n \textbf{ even} \end{cases}$

Consider a sample of n house prices:

- $\bar{x} = \$850,000$
- $\tilde{x} = \$700,000$
- **Why do the statistics differ?**

The median is more *robust* than the mean wrt outliers:

Know how to approximate the median and mean from a histogram: