# 1 Getting Started

**Definition 1.** ***Frequency Distribution****. A frequency distribution is a tabular summary of a set of data showing the frequency (or number) of items in each of several non-overlapping classes.*

**Definition 2.** ***Relative Frequency Distribution****. A relative frequency distribution is a tabular summary of a set of data showing the relative frequency in each class.*

## 1.1 Histogram

A histogram displays the frequencies or relative frequencies (vertical axis) and the variable of interest (horizontal axis).

## 1.2 Ogive

Ogive depicts the cumulative distribution or cumulative relative frequency distribution.

**Exercise 1.1.** ***Redskins Payroll Again.*** Compute the frequency distribution, relative frequency distribution, and cumulative relative frequency distribution of the defensive players' salaries using the following classes: less than 400 000$, less than 600 000$, less than 900 000$, less than 2 000 000$, less than 4 000 000$, less than 6 000 000$, less than 8 000 000$, less than 9 000 000$. Draw the histogram and ogive.

# 2 Measures of Dispersion

**Definition 3.** ***Dispersion****. Variation or scattering of data around some average or central value.*

**Definition 4.** ***Outlier****. An observation that lies outside some overall pattern of distribution (Moore & McCabe 1999) i.e. is distant from the rest of the sample or population.*

## 2.1 Variance

Variance is a measure of dispersion. Population variance is estimated as:

$$\sigma_P^2 = T^{-1} \sum_{i=1}^{T} (x_i - \mu)^2 , \tag{1}$$

where: $\mu$ population mean, whilst the sample variance as:

$$\sigma^2 = (T-1)^{-1} \sum_{i=1}^{T} (x_i - \bar{x})^2 . \tag{2}$$

**Exercise 2.1.** ***Investment Towards Uncertainty.*** An investor is interested in gambling on currency exchange trading. Suppose that the investor wishes to minimise the risk in the portfolio. What would be the currency the investor add to the portfolio? [Excel file sheet: Forex]

**Exercise 2.2.** ***Regional Disparities in Poland.*** Using the data on regional unemployment rate estimate the variance of of the unemployment rate in Wielkopolska and Mazowsze. Compare the results. Does the variance provide useful information on disparities in investment attractiveness? [Excel file sheet: Regions]

## 2.2 Standard Deviation

Standard deviation is the square root of variance:

$$\sigma_P = \sqrt{\sigma_P^2}, \tag{3}$$

$$\sigma = \sqrt{\sigma^2}. \tag{4}$$

Standard deviation is measured in the same units as the original data.

**Exercise 2.3.** ***Disparities Among the Players.*** Using the data on the Redskins payroll calculate the standard deviation for the salaries. Does the standard deviation provide useful information on selling tickets vs. winning championship dilemma?

## 2.3 The $z$-Score

The $z$-score, estimated as:

$$z_i = \frac{x_i - \bar{x}}{\sigma}, \tag{5}$$

is used to detect the outliers. If $z_i > |3|$, the $i$th observation should be considered to be an outlier. Note, however, that there is a growing number of techniques dealing with detecting outliers.

**Exercise 2.4.** *Stars and Mediocres Again.* Using the data on Washington Redskins' payroll estimate the $z$-score for the defensive players. Are there any outliers among the players?

# 3 Vocabulary

| **English** | **Polish** | **German** |
|---|---|---|
| dispersion | dyspersja | Dispers |
| distribution | rozkład | Häufigkeitsverteilung |
| frequency | częstotliwość | Frequenz |
| histogram | histogram | Histogramm |
| ogive | ogiwa | Ogive |
| outlier | obserwacja odstająca | Ausreißer |
| sample | próba | Stichprobe |
| standard deviation | odchylenie standardowe | Standardabweichung |
| variance | wariancja | Varianz |
| z-score | | Z-Faktor |

Table 1: Vocabulary