where σ_i^2 is the variance in dimension i. The **maximum scaled difference** (used by Maxwell and Buddemeier 2002, for coastal typology) is defined by

$$\max_i \frac{(x_i - y_i)^2}{\sigma_i^2}.$$

17.3 Similarities and distances for binary data

Usually, such similarities s range from 0 to 1 or from -1 to 1; the corresponding distances are usually $1 - s$ or $\frac{1-s}{2}$, respectively.

- **Hamann similarity**
  The **Hamann similarity** 1961, is a similarity on $\{0,1\}^n$, defined by

$$\frac{2|\overline{X\Delta Y}|}{n} - 1 = \frac{n - 2|X\Delta Y|}{n}.$$

- **Rand similarity**
  The **Rand similarity** (or Sokal–Michener's *simple matching*) is a similarity on $\{0,1\}^n$, defined by

$$\frac{\overline{|X\Delta Y|}}{n} = 1 - \frac{|X\Delta Y|}{n}.$$

  Its square root is called the *Euclidean similarity*. The corresponding metric $\frac{|X\Delta Y|}{n}$ is called the *variance* or *Manhattan similarity*; cf. **Penrose size distance**.
- **Sokal–Sneath similarity** 1
  The **Sokal–Sneath similarity** 1 is a similarity on $\{0,1\}^n$, defined by

$$\frac{2|\overline{X\Delta Y}|}{n + |\overline{X\Delta Y}|}.$$

- **Sokal–Sneath similarity** 2
  The **Sokal–Sneath similarity** 2 is a similarity on $\{0,1\}^n$, defined by

$$\frac{|X\cap Y|}{|X\cup Y| + |X\Delta Y|}.$$

- **Sokal–Sneath similarity** 3
  The **Sokal–Sneath similarity** 3 is a similarity on $\{0,1\}^n$, defined by

$$\frac{|X\Delta Y|}{|\overline{X\Delta Y}|}.$$

- **Russel–Rao similarity**
  The **Russel–Rao similarity** is a similarity on $\{0,1\}^n$, defined by

$$\frac{|X \cap Y|}{n}.$$

- **Simpson similarity**
  The **Simpson similarity** (*overlap similarity*) is a similarity on $\{0,1\}^n$, defined by

$$\frac{|X \cap Y|}{\min\{|X|, |Y|\}}.$$

- **Forbes similarity**
  The **Forbes similarity** is a similarity on $\{0,1\}^n$, defined by

$$\frac{n|X \cap Y|}{|X||Y|}.$$

- **Braun–Blanquet similarity**
  The **Braun–Blanquet similarity** is a similarity on $\{0,1\}^n$, defined by

$$\frac{|X \cap Y|}{\max\{|X|, |Y|\}}.$$

  The average between it and the **Simpson similarity** is the **Dice similarity**.
- **Roger–Tanimoto similarity**
  The **Roger–Tanimoto similarity** 1960, is a similarity on $\{0,1\}^n$, defined by

$$\frac{|\overline{X \Delta Y}|}{n + |X \Delta Y|}.$$

- **Faith similarity**
  The **Faith similarity** is a similarity on $\{0,1\}^n$, defined by

$$\frac{|X \cap Y| + |\overline{X \Delta Y}|}{2n}.$$

- **Tversky similarity**
  The **Tversky similarity** is a similarity on $\{0,1\}^n$, defined by

$$\frac{|X \cap Y|}{a|X \Delta Y| + b|X \cap Y|}.$$

  It becomes the **Tanimoto**, **Dice** and (the binary case of) **Kulczynsky** 1 **similarities** for $(a, b) = (1, 1)$, $(\frac{1}{2}, 1)$ and $(1, 0)$, respectively.

- **Mountford similarity**
  The **Mountford similarity** 1962, is a similarity on $\{0,1\}^n$, defined by

$$\frac{2|X \cap Y|}{|X||Y \backslash X| + |Y||X \backslash Y|}.$$

- **Gower–Legendre similarity**
  The **Gower–Legendre similarity** is a similarity on $\{0,1\}^n$, defined by

$$\frac{|\overline{X \Delta Y}|}{a|X \Delta Y| + |\overline{X \Delta Y}|} = \frac{|\overline{X \Delta Y}|}{n + (a-1)|X \Delta Y|}.$$

- **Anderberg similarity**
  The **Anderberg similarity** (or *Sokal–Sneath* 4 *similarity*) is a similarity on $\{0,1\}^n$, defined by

$$\frac{|X \cap Y|}{4}\left(\frac{1}{|X|} + \frac{1}{|Y|}\right) + \frac{|\overline{X \cup Y}|}{4}\left(\frac{1}{|\overline{X}|} + \frac{1}{|\overline{Y}|}\right).$$

- **Yule Q similarity**
  The **Yule Q similarity** (Yule 1900) is a similarity on $\{0,1\}^n$, defined by

$$\frac{|X \cap Y| \cdot |\overline{X \cup Y}| - |X \backslash Y| \cdot |Y \backslash X|}{|X \cap Y| \cdot |\overline{X \cup Y}| + |X \backslash Y| \cdot |Y \backslash X|}.$$

- **Yule Y similarity of colligation**
  The **Yule Y similarity of colligation** (Yule 1912) is a similarity on $\{0,1\}^n$, defined by

$$\frac{\sqrt{|X \cap Y| \cdot |\overline{X \cup Y}|} - \sqrt{|X \backslash Y| \cdot |Y \backslash X|}}{\sqrt{|X \cap Y| \cdot |\overline{X \cup Y}|} + \sqrt{|X \backslash Y| \cdot |Y \backslash X|}}.$$

- **Dispersion similarity**
  The **dispersion similarity** is a similarity on $\{0,1\}^n$, defined by

$$\frac{|X \cap Y| \cdot |\overline{X \cup Y}| - |X \backslash Y| \cdot |Y \backslash X|}{n^2}.$$

- **Pearson ϕ similarity**
  The **Pearson ϕ similarity** is a similarity on $\{0,1\}^n$, defined by

$$\frac{|X \cap Y| \cdot |\overline{X \cup Y}| - |X \backslash Y| \cdot |Y \backslash X|}{\sqrt{|X| \cdot |\overline{X}| \cdot |Y| \cdot |\overline{Y}|}}.$$

- **Gower similarity** 2
  The **Gower similarity** 2 (or *Sokal–Sneath* 5 *similarity*) is a similarity on $\{0,1\}^n$, defined by
$$\frac{|X \cap Y| \cdot |\overline{X \cup Y}|}{\sqrt{|X| \cdot |\overline{X}| \cdot |Y| \cdot |\overline{Y}|}}.$$

- **Pattern difference**
  The **pattern difference** is a distance on $\{0,1\}^n$, defined by
$$\frac{4|X \backslash Y| \cdot |Y \backslash X|}{n^2}.$$

- **Q_0-difference**
  The **Q_0-difference** is a distance on $\{0,1\}^n$, defined by
$$\frac{|X \backslash Y| \cdot |Y \backslash X|}{|X \cap Y| \cdot |\overline{X \cup Y}|}.$$

17.4 Correlation similarities and distances

- **Covariance similarity**
  The **covariance similarity** is a similarity on $\mathbb{R}^n$, defined by
$$\frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{n} = \frac{\sum x_i y_i}{n} - \overline{x} \cdot \overline{y}.$$

- **Correlation similarity**
  The **correlation similarity** (or *Pearson correlation*, or, by its full name, *Pearson product-moment correlation linear coefficient*) s is a similarity on $\mathbb{R}^n$, defined by
$$\frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{(\sum(x_j - \overline{x})^2)(\sum(y_j - \overline{y})^2)}}.$$
  The dissimilarities $1 - s$ and $1 - s^2$ are called the **Pearson correlation distance** and *squared Pearson distance*, respectively. Moreover,
$$\sqrt{2(1-s)} = \sqrt{\sum \left(\frac{x_i - \overline{x}}{\sqrt{\sum(x_j - \overline{x})^2}} - \frac{y_i - \overline{y}}{\sqrt{\sum(y_j - \overline{y})^2}} \right)}$$
  is a normalization of the Euclidean distance (cf., a different one, **normalized l_p-distance** above in this chapter).
  In the case $\overline{x} = \overline{y} = 0$, the correlation similarity becomes $\frac{\langle x,y \rangle}{||x||_2 \cdot ||y||_2}$.