Probability and Statistics

1 – Descriptive Statistics

Stefan Heiss

Technische Hochschule Ostwestfalen-Lippe
Dep. of Electrical Engineering and Computer Science

October 13, 2023

Lemma (1.27)

Let $x = (x_1, x_2, \ldots, x_N)$ *be a data set and*

$$\xi_i \;=\; \alpha + \beta\, x_i \qquad \text{for } i = 1, \ldots, N$$

$\xi = (\xi_1, \ldots, \xi_N)$

$\bar{\xi} = \alpha + \beta \bar{x}$

$s_\xi = |\beta| \cdot s_x$

for some constant values α *and* β. *Then:*

$$\xi_i{}^o \;=\; \operatorname{sgn}(\beta) \cdot x_i{}^o \qquad \text{for } i = 1, \ldots, N$$

$$\xi_i^o = \frac{1}{s_\xi}(\xi_i - \bar{\xi}) = \frac{1}{|\beta| \cdot s_x}(\xi_i - \alpha + \beta \bar{x})$$

$$= \frac{1}{|\beta| \cdot s_x}(\cancel{\alpha} + \beta x_i - \cancel{\alpha} + \beta \bar{x}) = \frac{\beta}{|\beta|} \cdot \frac{1}{s_x}(x_i - \bar{x})$$

$$= \operatorname{sgn}(\beta) \cdot x_i^o \qquad \square$$

Lemma (1.28)

Let $x = (x_1, x_2, \ldots, x_N)$ *and* $y = (y_1, y_2, \ldots, y_N)$ *be non-constant data sets with mean* $\overline{x}$ *and mean* $\overline{y}$*, respectively. Furthermore let* r *be the correlation coefficient of* x *and* y*. Then:*

(i) $r \in [-1, 1]$

(ii) $|r| = 1$ *iff and only if there exist constants* a, b*, such that*

$$y_i = a + bx_i \qquad \text{for all } i \in \{1, \ldots, N\} .$$

(In this case $r = \operatorname{sgn}(b)$*.)*

(iii) *Let* a*,* b*,* c *and* d *constant values with* $b \cdot d > 0$*. Set*

$$\xi_i = a + b \cdot x_i, \qquad \eta_i = c + d \cdot y_i$$

for $i = 1, 2, \ldots, N$*. Then* $\xi = (\xi_1, \xi_2, \ldots, \xi_N)$ *and* $\eta = (\eta_1, \eta_2, \ldots, \eta_N)$ *have the same correlation coefficient as* x *and* y*:*

$$r = r_{x,y} = r_{\xi,\eta}$$

Proof of Lemma (1.28)

(i): Consider the vectors:

$$\mathrm{v} = (x_1 - \overline{x}, x_2 - \overline{x}, \ldots, x_N - \overline{x}) \quad \text{and} \quad \mathrm{w} = (y_1 - \overline{y}, y_2 - \overline{y}, \ldots, y_N - \overline{y})$$

$$r = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2 \sum_{i=1}^{N}(y_i - \overline{y})^2}} = \frac{\mathrm{v} \cdot \mathrm{w}}{|\mathrm{v}| \cdot |\mathrm{w}|} = \cos(\alpha) \in [-1, 1]$$

where α denotes the angle between v and w.

Proof of Lemma (1.28)

(ii): $|r| = 1$, if and only if v and w are collinear, i.e. if and only if there exists some $b \in \mathbb{R}$ such that:

w = b · v

$$
\begin{array}{lrcll}
 & y_i - \overline{y} & = & b(x_i - \overline{x}) & \text{for all } \; i = 1, 2, \ldots, N \\
\Longleftrightarrow & y_i & = & (\overline{y} - b\overline{x}) + b\,x_i & \text{for all } \; i = 1, 2, \ldots, N \\
\Longleftrightarrow \quad \text{for some } \; a: & y_i & = & a + b\,x_i & \text{for all } \; i = 1, 2, \ldots, N
\end{array}
$$

a = ȳ - bx̄

Proof of Lemma (1.28)

(ii): $|r| = 1$, if and only if v and w are collinear, i.e. if and only if there exists some $b \in \mathbb{R}$ such that:

$$
\begin{array}{llrcll}
 & & y_i - \overline{y} & = & b(x_i - \overline{x}) & \text{for all } \; i = 1, 2, \ldots, N \\
\iff & & y_i & = & (\overline{y} - b\overline{x}) + b\, x_i & \text{for all } \; i = 1, 2, \ldots, N \\
\iff & \text{for some } \; a: & y_i & = & a + b\, x_i & \text{for all } \; i = 1, 2, \ldots, N
\end{array}
$$

"$\Leftarrow$" :

For the last equivalence note, that $y_i = a + b\, x_i$ for all i implies:

$$
a \; = \; y_i - b\, x_i \; \text{ for all } \; i \qquad \Longrightarrow \qquad a \; = \; \frac{1}{N} \sum_{i=1}^{N} (y_i - b\, x_i) \; = \; \overline{y} - b\overline{x}
$$

Proof of Lemma (1.28)

(iii):

$$\xi_i{}^o = \operatorname{sgn}(b) \cdot x_i{}^o \quad \text{and} \quad \eta_i{}^o = \operatorname{sgn}(d) \cdot y_i{}^o \qquad \text{for } i = 1, \ldots, N$$

$$r_{\xi,\eta} = \frac{\sum_{i=1}^{N} \xi_i{}^o \, \eta_i{}^o}{(N-1)} = \frac{\operatorname{sgn}(b) \cdot \operatorname{sgn}(d) \cdot \sum_{i=1}^{N} x_i{}^o \, y_i{}^o}{(N-1)} = r_{x,y}$$