Descriptive statistics

Central tendency

sample mean
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
for grouped data, with n_i observations in group i:
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i \bar{x}_i$$

median

$$\tilde{x} = \begin{cases} x_{(n+1)/2} & n \text{ odd} \\ \frac{1}{2}\left(x_{n/2} + x_{n/2+1}\right) & n \text{ even} \end{cases}$$

for continuous data, the median is the value where the cumulative distribution function equals 1/2

mode

$$\hat{x} = \max_i f(x_i)$$

Variability

sample variance
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$
for grouped data, with n_i observations in group i:
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2$$

sample standard deviation
$$s = \sqrt{s^2}$$

sample coefficient of variation
$$c_v = \frac{s}{\bar{x}}$$

sample range
$$r = x_{(n)} - x_{(1)}$$

sample interquartile range
$$\mathrm{IQR} = x_{(3n/4)} - x_{(n/4)}$$

Correlation

sample covariance
$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

sample correlation coefficient
$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

$$-1 \le r_{xy} \le 1$$

$r_{xy} = 1$ if $y_i = a + b x_i$ with $b > 0$

$r_{xy} = -1$ if $y_i = a + b x_i$ with $b < 0$

$r_{xy} = 0$ if x and y are uncorrelated

$r_{xy} \approx 0$ if x and y are independent

$r_{xy} \approx 0$ if x and y are related by a symmetric function

$r_{xy} \approx 1$ if x and y are related by a monotonic function

Regression

$$y_i = a + b x_i + \varepsilon_i$$

ε_i is the error term, a is the intercept, b is the slope

least squares

minimizes the residual

$$\sum_{i=1}^{n} \varepsilon_i^2$$

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{s_{xy}}{s_x^2}$$

$$s_\varepsilon^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - a - b x_i)^2$$

$$s_a^2 = s_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x^2}\right)$$

$$s_b^2 = \frac{s_\varepsilon^2}{s_x^2}$$

$$s_{ab} = -\frac{s_\varepsilon^2 \bar{x}}{s_x^2}$$

$$s_{y|x}^2 = s_\varepsilon^2 \left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{s_x^2}\right)$$

$$s_{y|x} = s_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{s_x^2}}$$

$$s_{y|x} \approx s_\varepsilon \quad \text{if } x \approx \bar{x}$$

$$s_{y|x} \approx s_\varepsilon \sqrt{1 + \frac{1}{n}} \quad \text{if } x \approx \bar{x}$$

$$s_{y|x} \approx s_\varepsilon \sqrt{1 + \frac{(x - \bar{x})^2}{s_x^2}} \quad \text{if } n \gg 1$$

$$s_{y|x} \approx s_\varepsilon \quad \text{if } n \gg 1 \text{ and } x \approx \bar{x}$$

$$s_{y|x} \approx s_\varepsilon \sqrt{1 + \frac{1}{n}} \quad \text{if } n \gg 1$$

$$s_{y|x} \approx s_\varepsilon \sqrt{1 + \frac{(x - \bar{x})^2}{s_x^2}} \quad \text{if } x \approx \bar{x}$$

$$s_{y|x} \approx s_\varepsilon \sqrt{1 + \frac{(x - \bar{x})^2}{s_x^2}} \quad \text{if } n \gg 1 \text{ and } x \approx \bar{x}$$

$$s_{y|x} \approx s_\varepsilon \sqrt{1 + \frac{1}{n}