Statistical Formulas

Descriptive Statistics

Frequencies

Sample size n num of individuals in the sample.
Absolute frequency n_i. (num of x_i in the sample)
Relative frequency $f_i = n_i/n$.
Cumulative absolute freq $N_i = \sum_{k=0}^{i} n_i$.
Cumulative relative freq $F_i = N_i/n$.

Central tendency statistics

Mean $\bar{x} = \dfrac{\sum x_i n_i}{n}$.
Median me The value with cum.rel.freq. $F_{me} = 0.5$.
Mode mo The most frequent value.

Position statistics

Quartiles Q_1, Q_2, Q_3 divide the distribution into 4 equal parts. Their cum.rel.freqs. are $F_{Q_1} = 0.25$, $F_{Q_2} = 0.5$ and $F_{Q_3} = 0.75$.
Percentiles $P_1, P_2, \cdots, P_{99}$ divide the distribution into 100 equal parts.
The cum.rel.freq. is $F_{P_i} = i/100$.

Dispersion statistics

Interquartile range $IQR = Q_3 - Q_1$.
Variance $s^2 = \dfrac{\sum (x_i - \bar{x})^2 n_i}{n} = \dfrac{\sum x_i^2 n_i}{n} - \bar{x}^2$
Standard deviation $s = +\sqrt{s^2}$.
Coefficient of variation $cv = \dfrac{s}{|\bar{x}|}$.

Shape statistics

Coefficient of skewness $g_1 = \dfrac{\sum (x_i - \bar{x})^3 f_i}{s^3}$.

Coefficient of kurtosis $g_2 = \dfrac{\sum (x_i - \bar{x})^4 f_i}{s^4} - 3$.

Standarization

$$z = \frac{x - \bar{x}}{s_x}$$

Regression and correlation

Linear regression

Covariance $s_{xy} = \dfrac{\sum x_i y_j n_{ij}}{n} - \bar{x}\bar{y}$.
Regression lines :
y **on** x**:** $y = \bar{y} + \dfrac{s_{xy}}{s_x^2}(x - \bar{x})$
x **on** y**:** $x = \bar{x} + \dfrac{s_{xy}}{s_y^2}(y - \bar{y})$
Regression coefficients

$$(y \text{ on } x)\ b_{yx} = \frac{s_{xy}}{s_x^2} \qquad (x \text{ on } y)\ b_{xy} = \frac{s_{xy}}{s_y^2}$$

Coefficient of determination

$$r^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} \qquad 0 \leq r^2 \leq 1$$

Correlation coefficient

$$r = \frac{s_{xy}}{s_x s_y}. \qquad -1 \leq r \leq 1$$

Non-linear regression

Exponential model $y = e^{a+bx}$
Apply the logarithm to the dependent variable and compute the line $\log y = a + bx$.
Logarithmic model $y = a + b \log x$
Apply the logarithm to the independent variable and compute the line $y = a + b \log x$.
Potential model $y = ax^b$
Apply the logarithm to both variables and compute the line $\log y = a + b \log x$.