Aprašomoji statistika

Trumpai priminsime kai kurias empirines atsitiktinių dydžių charakteristikas.
Tegu $x = \langle x_1, x_2, ..., x_n \rangle$ yra imtis, gauta stebint atsitiktinio dydžio X reikšmes. Didėjimo tvarka išdėstyta šios imties duomenų eilė

$$x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$$

vadinama *imties variacine eilute*.

- Imties x *vidurkiu* ir *dispersija* vadinami skaičiai

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \,, \quad s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \,.$$

- Imties x q-osios eilės *kvantilis* yra skaičius

$$v_q = \begin{cases} x_{([nq]+1)}, & \textit{jei nq nėra sveikasis skaičius}, \\ (x_{(nq)} + x_{(nq+1)})/2, & \textit{jei nq yra sveikasis skaičius}, \end{cases}$$

  čia $0 < q < 1$. Dažniausiai naudojami $q = \frac{1}{4}, \ \frac{2}{4}, \ \frac{3}{4}$ eilės kvantiliai. Jie vadinami atitinkamai pirmuoju, antruoju ir trečiuoju *kvartiliais*. Antrasis kvartilis dar vadinamas *mediana*.

- Atsitiktinio dydžio X *empirinė pasiskirstymo funkcija* apibrėžiama taip:

$$F_n^*(u) = \frac{n(u)}{n} \,,$$

  čia $n(u)$ - imties x duomenų, mažesnių už u, skaičius.

- Tegu k - skirtingų imties x duomenų skaičius, f_i - i-ojo duomens santykinis dažnis

$$f_i = \frac{n_i}{n} \,, \ i = 1, 2, ..., k.$$

  Imties įvairovės indeksas (Index of Qualitative Variation) yra

$$IQV = \frac{k}{k-1} \left(1 - \sum_{i=1}^{k} f_i^2 \right) \,.$$

- Tegu $\langle (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) \rangle$ yra imtis, gauta stebint dvimačio atsitiktinio dydžio (X, Y) reikšmes. Tada X ir Y *empirinis koreliacijos koeficientas* yra

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{(n-1) s_x s_y} \,.$$