Alapfogalmak a statisztikában

A tömegesen előforduló jelenségek és folyamatok számba vételével, az így nyert adatok vizsgálatával, elemzésével foglalkozik a statisztika.
A statisztikai adatokat célszerűen leírjuk, a lehetséges értékeket gyakoriságukkal együtt táblázatban foglaljuk össze. Ezt nevezzük gyakorisági eloszlásnak.

A gyakoriságok összege megegyezik a sokaságban szereplő elemek számával.
Ha a gyakoriságot elosztjuk a sokaság elemeinek a számával, akkor a relatív gyakoriságot kapjuk meg.

Statisztikai adatok ábrázolása

Ha nagy mennyiségű adat alapján szeretnénk következtetni, lényeges hogy az adatokat könnyen olvashatóvá tegyük. A média igényli is a gyorsan áttekinthető ábrázolásokat. Azonban egy színes ábra, úgy nevezett diagrammra vetett pillantás nagyon sok mindent elárulhat az arányokról, a változás esetleges irányáról. Az ábrázoláshoz használható az oszlopdiagram, a kördiagram, vonaldiagram, a sávdiagram, a grafikon stb.

Középértékek

A számtani közép az átlag.

A számsokaságban legtöbbször előforduló számot a számsokaság móduszának nevezzük.

Rendezzük nagyság szerint sorba a számadatokat, a középsőt nevezzük mediánnak (ha két középső van, akkor ezek átlagát vesszük).

A szórás

A számsokaság legnagyobb és legkisebb számának a különbségét terjedelemnek nevezzük

Átlagos abszolút eltérésnek nevezni. $S_n(\overline{x}) = \dfrac{|x_1 - \overline{x}| + |x_2 - \overline{x}| + \dots + |x_n - \overline{x}|}{n}$

Belátható, hogy négyzetösszegek jobban jellemzik a sokaság szerkezetét, ezért használják szóródás jellemzésére az átlagos négyzetes eltérést a statisztikában.

Az $X_1;X_2;\ \dots;X_n$ számsokaság tetszőleges X számtól vett átlagos négyzetes eltérésnek nevezzük a következőt:

$$D_n^2(x) = \frac{(x_1 - x)^2 + (x_2 - x)^2 + \dots + (x_n - x)^2}{n}$$

Ha az x a sokaság átlaga ($x = \overline{x}$), akkor ezt a kifejezést a sokaság szórásnégyzetének nevezzük.

Az empirikus szórás: $D_n(\overline{x}) = \sqrt{\dfrac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \dots + (x_n - \overline{x})^2}{n}}$