Eloszlások a statisztikában

A statiszt, statisztika szó a latin status szóból ered, amely a középkorban az állam jelentést vette fel. A statiszt szó már Shakespeare-nál megtalálható. Eredetileg az állammal kapcsolatos minden jelentős tényre vonatkozott, nem csak számszerűekre. Későbbi jelentésében például népszámlálásból származó, a népességre vonatkozó számadatok összességét jelenti, másszóval a „politikai aritmetika". Az adatok tudományos igényű vizsgálata a XVII.szd-ban kezdődött. 1666-ban jelent meg J. Graunt kapitány *Natural and political Annotations made upon the Bills of Mortality* c. műve, amely London városának (1592-óta vezetett) halálozási mutatóit elemzi. Követői Sir W. Petty (*Five Essays in politickal Arithmetick*, 1686) és Halley, a neves asztronómus és matematikus, aki 1693-ban közölte *An Estimate of the Degrees of Mortality of Mankind, drawn from curious Tables of the Births and Funerals at the City of Breslaw...* c. cikkét, amelyben az életjáradék kötvények árának meghatározási módszerét alapozta meg elsőként. J.P. Süssmilch, porosz tiszteletes, 1761-ben *Die gottliche Ordnung in den Veränderungen des menschlichen Geschlets aus der Geburt, dem Tode, und der Fortplanzung desselben erweisen* c. művében először vizsgálta szisztematikusan a ma vitálstatisztikának nevezett számadatokat ötvözve a politikai aritmetika és a leíró statisztika módszereit, először próbált meg általános elméletet felállítani megsejtve a nagy számok törvényét. Maga a statisztika szót leíró statisztika jelentésben von Bielfeld használta először 1770-ben, és a XIX.szd. első felében általánosan elterjedt. Ezután jelentése általánosabbá vált, elvesztette szoros kapcsolatát az államügyekkel, egyéb diszciplinákra is vonatkoztatták, másrészről beszűkült, kizárólag a numerikus jellemzőkre utalóvá vált.

A statisztika elméletének megalapítója a belga Quetelet (1796-1874), aki írásaiban (pl. *L'Anthropométrie, ou mesure des differentés facultés de l'homme*, 1871) az átlagember (l'homme moyen) képzetét alakította ki, akinek fizikális, mentális tulajdonságai körül a népesség egészének tulajdonságai ingadoznak az általa kimutatott nagy számok törvényének megfelelően (a valószínűségszámításban ez Bernoulli nevéhez kötődik).

Kísérletek, vizsgálatok során változókat figyelünk meg, amelyek a vizsgált jelenség kvantitatív jellemzőit adják meg. Megkülönböztetünk függő és független változókat, a kutatásokban gyakorta a független változók változását a függetlenek függvényeként határozzuk meg. A változók közötti relációt kétféleképpen vizsgáljuk, az egyik a nagyságrendi – azaz hogy mekkora a függőségi viszonynak a mértéke; a másik a megbízhatósági, azaz hogy a megfigyelt összefüggés mennyire tekinthető véletlennek, illetve mennyire tükrözi a teljes populáció viselkedését.

A változók értékei lehetnek diszkrétek illetve folytonosak, amikor a felvehető értékek elméletileg nem megszámlálható halmazt alkotnak, itt most ez utóbbiakról szólunk. Azt a függvényt, amely annak az eseménynek a valószínűségét veszi fel az x pontban, hogy a változó értéke x-nél kisebb, eloszlásfüggvénynek nevezzük. Általában feltesszük, hogy ez differenciálható, deriváltja a sűrűségfüggvény.

Elvégezve egymástól függetlenül több mérést illetve megfigyelést, vagy másszóval, a statisztikai sokaságból mintát veszünk, a kapott eredményeket osztályokba sorolva, a kumulált relatív gyakoriságokat kiszámítva lépcsős függvényt szerkeszthetünk, amelyet tapasztalati eloszlásfüggvénynek nevezünk. Elméletileg a mintaelemszámot növelve ezek közelítenek egy sima görbét, az elméleti eloszlásfüggvényt. Hasonlóan, a gyakoriságokat kiszámítva lépcsős függvényt szerkeszthetünk, amelyet tapasztalati sűrűségfüggvénynek nevezünk. Elméletileg a mintaelemszámot növelve ezek közelítenek egy sima görbét, az elméleti sűrűségfüggvényt.

A legfontosabb elméleti eloszlásfüggvény a normális eloszlásfüggvény. De Moivre 1730-ban a Stirling-formula segítségével közelítette az (n, p)-paraméterű binomiális eloszlást az $(np, \sqrt{np(1-p)})$-paraméterű normális eloszlással. Később ezeket az eredményeket Gauss és Laplace (1809, 1812) alkalmazta asztrológiai illetve geodéziai mérések hibaelemzéséhez.

Pearson mutatta meg, hogy a Gauss-Laplace féle normális eloszlás messze nem az általános törvénye a megfigyelési hibák eloszlásának vagy élő szervezetekben a típustól való eltérésnek. Több asszimetrikus eloszlásfüggvényt vezetett be, például a χ^2 eloszlás alkalmazása (1900) is Pearson nevéhez fűződik. Ezek nagy többsége mégis a normáleloszlásból származtatható. A statisztikai vizsgálatok nagy része normális eloszlású mintákra vonatkozik. Ezek akkor is alkalmazhatók, ha a mintaeloszlás nem normáleloszlást követ, feltéve, hogy a mintaelemszám megfelelően nagy, aminek az oka a centrális határeloszlástételben rejlik.