

A statisztika diszciplinája

A statisztikai vizsgálatok lényegében háromféle kérdésre keresnek választ. Először, ha előzetesen ismeretes egy jelenség elméleti eloszlása, a jelenségre elvégzett megfigyeléssorozatban milyen eredményeket várhatunk? Másodszor, valamely jelenségre elvégzett megfigyeléssorozat numerikus adatai alapján milyen matematikai modellel (azaz valószínűségeloszlással) magyarázhatjuk a megfigyeléseket? Harmadszor, valamely jelenségre elvégzett megfigyeléssorozat numerikus adatai alapján a jelenségre a jövőben elvégzendő megfigyeléssorozatban milyen eredményeket várhatunk? A statisztika tekinthető végeredményben a matematikai valószínűségszámítás alkalmazásának, amelynek precíz elméleti megalapozása nem is oly régen történt. Ámbár egyesek, mint például COMTE, a matematikai valószínűségszámítás elméletének ilyen alkalmazásáról bizonyos, például szociometriai jelenségekre azt tartotta, hogy „egy légvár, és éppen ennél fogva teljesen hibás”.

Megkülönböztetünk leíró statisztikát és matematikai statisztikát. A leíró statisztika valamely tömeges jelenséget leíró numerikus adatoknak a könnyebb áttekinthetőség végett táblázatba foglalásával, grafikonokkal, diagrammokkal történő szemléltetésével, illetve a primér adatokból származtatott numerikus mutatók kiszámításával foglalkozik. A matematikai statisztika ezzel ellentétben a mintaadatokból történő következtetéseket támasztja alá valószínűségszámítási alapokból kiindulva matematikai bizonyítás révén: azaz, a megfigyelt jelenségekből meg nem figyelt jelenségek jellemzőire történő következtetésre nyújt módszereket.

Pontosabban: Az $(\Omega, \mathcal{A}, \mathcal{P})$ hármast **statisztikai mezőnek** nevezzük, ha (Ω, \mathcal{A}) σ -algebra és $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ valószínűségi mértékek rendszere, ahol θ **paraméter**, Θ a **paramétertér**. Megfigyelünk egy **mintának** nevezett $X = (\xi_1, \dots, \xi_n) : \Omega \rightarrow \Xi$ valószínűségi változót, ahol a ξ_i valószínűségi változók azonos eloszlásúak – ez az ún. **elméleti eloszlás** – és teljesen függetlenek, továbbá Ξ a **mintatér vagy populáció**, \mathbb{R}^n egy részhalmaza. A (Ξ, \mathcal{B}) σ -algebra olyan, hogy X mérhető függvény legyen. Minden paraméterre az X valószínűségi változó indukál egy $P_\theta^X : B \mapsto P_\theta^X(B) = P_\theta(\{\omega \in \Omega | X(\omega) \in B\})$ mértéket. A statisztikai indukció X megfigyelt (x_1, \dots, x_n) értékeiről a θ paraméterre történik: az X egy $t(X)$ mérhető függvényét **statisztikai függvénynek** (röviden statisztikának) nevezzük.

A matematika legerőteljesebben alkalmazott ága éppen a statisztika, nehéz lenne olyan tudományterületet találni, ahol nem használják eredményeit. Ezért hasznos egy olyan feladatgyűjtemény, amely nem csak a matematika szakos hallgatók igényeit elégíti ki, hanem mélyebb matematikai eszköztár nélkül is megérthető, de egyúttal a szakos hallgatók is tanulhatják. Számolásigénye miatt a szá-

mítástechnika és statisztika összefonódott, számtalan nagyteljesítményű szoftver támogatja a statisztikai műveleteket jobbnál jobb szolgáltatásokkal. Az EXCEL^{TM1} táblázatkezelőre azért esett a választásunk, mivel a hallgatók számára jól ismert a használata, a számítógépes tantermek gépein általánosan rendelkezésre áll campusszerte, és – talán első pillantásra furcsának tűnhet – a magas szintű statisztikai szolgáltatásai használata nélkül is alkalmas a számítások elvégzésére. Amennyire lehet, a beépített statisztikai eljárások használatát kerüljük, és megpróbáljuk azokat a számításokat utánózni, amikor még nem voltak számítógépek, maguknak a műveleteknek fáradságos és hosszadalmas kézi elvégzése nélkül. Persze ez nem sikerülhet teljesen, hiszen a számolások hatékony elvégzésének (a logarléc kezelésének, táblázatok használatának, alkalmas számítási formuláknak) az elsajátítása igénybe venné az egy féléves kurzus időtartamának jó részét.

A feldolgozott anyag egy standard BSc statisztika kurzusnak tekinthető, nemcsak matematika alapszakosoknak, hanem egyéb természettudományos képzések hallgatóinak is. Egy átlagos feladat linkeken keresztül elérhetően tartalmazza az Exceles megoldást PDF-fájlként, a megoldáshoz felhasználható XLS sablont a nyers adatokkal, illetve az elméleti útmutatót. Lévén feladatgyűjtemény, a bizonyítások jobbára kimaradnak innen. Minden paragrafusban a megoldáson felül egy feladat megoldásához teljesen részletes útmutatót is biztosítunk, képernyőkivágásokkal együtt. A többi feladatban csak utalunk arra, hogy mely nem közismert beépített függvényt kell használni: a szoftver alapszintű ismeretét feltételezzük. Ha másképpen nem jelezzük, a feladatokban feldolgozott problémák és a közölt nyers adatok fiktívek.

¹Microsoft Inc.