Statisztika1 1. előadás (BSc matematikus)

Paradoxonok (a lenti feladatsorban).

Véletlen mintavételezés, leíró statisztikák: mintaátlag, empirikus variancia, szórás, kovariancia, korreláció, medián, kvartitisek (tankönyv 63-75. old., Glivenko–Cantelli tétel bizonyítása nem kell).

Definíció: Az F eloszlásfüggvényű folytonos eloszlás p-kvantilise (100p-percentilise) a_p, ha $F(a_p) = p$. Ha $p = \frac{1}{2}$, akkor ezt mediánnak, ha $p = \frac{1}{4}$, akkor alsó kvartilisnek, ha $p = \frac{3}{4}$, akkor felső kvartilisnek nevezzük (ezek nem feltétlenül egyértelműek). A felső és alsó kvartilisek különbsége az ún. *interkvartilis tartomány.*

Normalitásvizsgálat heurisztikusan. Legyen $x_1, \ldots, x_n$ egy fae. minta realizáltja (az X háttérváltozóval képviselt eloszlásból), és osszuk fel a számegyenest az $a_1, \ldots, a_k$ osztópontokkal $k+1$ részre úgy, hogy mindegyikben legyen legalább 3 megfigyelés, és ne essen határra megfigyelés. Legyen

$$f_i := |\{j : x_j < a_i\}|, \quad i = 1, \ldots, k.$$

Ha adataink valamely $\mathcal{N}(\mu, \sigma^2)$ normális eloszlásból származnak és n "nagy", akkor a következő relatív gyakoriságok jól közelítik a valószínűséget:

$$r_i = \frac{f_i}{n} \approx p_i = P(X \leq a_i) = \Phi(\frac{a_i - \mu}{\sigma}), \quad i = 1, \ldots k.$$

Alkalmazuk a Φ^{-1} ún. *probit* transzformációt! Ekkor

$$\Phi^{-1}(r_i) \approx \frac{a_i - \mu}{\sigma},$$

azaz az $(a_i, \Phi^{-1}(r_i))$ $(i = 1, \ldots, k)$ pontok közelítőleg egy egyenesen helyezkednek el, aminek meredekségéből és y-tengelymetszetéből következtethetünk μ és σ értékeire.

Megjegyezzük, hogy a probit transzformációra régebben ún. Gauss-papír állt rendelkezésre, míg a lineáris közelítés regressziós módszerekkel ellenőrizhető.