Formule za 5. vajo pri predmetu Naključni pojavi

Vzorčno povprečje za vzorec z n elementi je:

$$\langle X\rangle = \frac{1}{n}\sum_{i=1}^{n} X_i\,. \tag{1}$$

Če je spremenljivka X normalno porazdeljena z gostoto $\mathcal{N}(X; m, \sigma)$, potem je vzorčno povprečje tudi normalno porazdeljeno, in sicer z gostoto $\mathcal{N}(\langle X\rangle; m, \sigma/\sqrt{n})$. Če pa X ni normalno porazdeljena, potem se porazdelitev vzorčnega povprečja z naraščajočo velikostjo vzorca približuje normalni:

$$\lim_{n\to\infty} f_{\langle X\rangle}(\langle x\rangle) = \mathcal{N}\left(\langle X\rangle; m, \sigma/\sqrt{n}\right)\,. \tag{2}$$

Običajno predpostavimo, da je za $n \geq 30$ porazdelitev $\langle X\rangle$ približno normalna, ne glede na porazdelitev spremenljivke X.

Parametre porazdelitev (npr. m, σ, p, λ itn.) lahko ocenjujemo točkovno ali intervalno. Pri točkovni oceni sta na voljo dve metodi: metoda momentov in metoda največje zanesljivosti.

Pri metodi momentov izraze za teoretične momente verjetnostne porazdelitve, katere parametre želimo oceniti, izenačimo z izrazi za ustrezne vzorčne momente. Potrebujemo toliko parov momentov, kolikor parametrov želimo oceniti. V dobljenem sistemu enačb so ocenjevani parametri neznanke. Z rešitvijo sistema enačb dobimo cenilke za parametre. Cenilko parametra θ označimo z $\hat{\theta}$.

Pri metodi največje zanesljivosti tvorimo funkcijo zanesljivosti:

$$L(X;\theta) = f(x_1;\theta)\cdot f(x_2;\theta)\cdot\cdots\cdot f(x_n;\theta)\,, \tag{3}$$

v kateri določimo parameter θ tako, da je vrednost funkcije $L(X;\theta)$ maksimalna. V ta namen $L(X;\theta)$ odvajamo po θ in odvod izenačimo z nič. Iz dobljene enačbe dobimo izraz za parameter θ, ki predstavlja izraz za cenilko $\hat{\theta}$. Če je ocenjevanih parametrov več, potem funkcijo zanesljivosti odvajamo po vsakem parametru posebej, s čimer dobimo sistem enačb za parametre. Rešitve sistema so cenilke za parametre.

Pri intervalnem ocenjevanju parametrov določimo interval zaupanja $[L, U]$, za katerega $(1-\alpha)\cdot 100\,\%$ zaupamo, da pokriva ocenjevani parameter θ:

$$P\,(L < \theta < U) = 1 - \alpha\,. \tag{4}$$

α imenujemo tveganje, $1-\alpha$ pa zaupanje. Intervalne ocene so lahko dvostranske in leve ali desne enostranske:

$$l < \theta < u \qquad \text{ali} \qquad l < \theta \qquad \text{ali} \qquad \theta < u\,. \tag{5}$$

Napaka intervalne ocene je $|l-\theta|$ ali $|u-\theta|$.

Dvostranski interval zaupanja za povprečje m normalno porazdeljene populacije z znano varianco σ^2 je:

$$\langle x\rangle - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < m < \langle x\rangle + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\,. \tag{6}$$

Dvostranski interval zaupanja za povprečje m poljubno porazdeljene populacije z neznano varianco σ^2 pri velikem vzorcu ($n > 30$) je:

$$\langle x\rangle - z_{\alpha/2}\frac{s}{\sqrt{n}} < m < \langle x\rangle + z_{\alpha/2}\frac{s}{\sqrt{n}}\,, \tag{7}$$

kjer je s vzorčna standardna deviacija oziroma s^2 vzorčna varianca:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \langle X\rangle\right)^2\,. \tag{8}$$

Dvostranski interval zaupanja za povprečje m normalno porazdeljene populacije z neznano varianco σ^2 pri majhnem vzorcu ($n < 30$) je:

$$\langle x \rangle - t_{n-1;\alpha/2} \frac{s}{\sqrt{n}} < m < \langle x \rangle + t_{n-1;\alpha/2} \frac{s}{\sqrt{n}} \,. \tag{9}$$

Pri tem predpostavljamo, da je naključna spremenljivka

$$T = \frac{\langle X \rangle - m}{S/\sqrt{n}} \tag{10}$$

Studentovo ali t-porazdeljena z $n-1$ prostostnimi stopnjami. Vrednosti $t_{n-1;\alpha}$ odčitamo iz tabele A.2.
Dvostranski interval zaupanja za varianco σ^2 normalno porazdeljene populacije je:

$$\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}} \,. \tag{11}$$

Pri tem predpostavljamo, da je naključna spremenljivka

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \tag{12}$$

χ^2-porazdeljena z $n-1$ prostostnimi stopnjami. Vrednosti $\chi^2_{n-1;\alpha}$ odčitamo iz tabele A.3.
Približni dvostranski interval zaupanja za delež p populacije pri velikem vzorcu je:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \,. \tag{13}$$

Pri tem predpostavljamo, da lahko binomsko porazdelitev aproksimiramo z normalno.
Dvostranski interval zaupanja za vsoto (razliko) povprečij m_1 in m_2 normalno porazdeljenih populacij z znanima variancama je:

$$\langle x_1 \rangle \pm \langle x_2 \rangle - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < m_1 \pm m_2 < \langle x_1 \rangle \pm \langle x_2 \rangle + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \,, \tag{14}$$

V kolikor sta populaciji poljubno porazdeljeni z neznanima variancama in velikima vzorcema, je interval enak zgornjemu, le da σ_1^2 in σ_2^2 nadomestimo z s_1^2 in s_2^2.
Dvostranski interval zaupanja za vsoto (razliko) povprečij m_1 in m_2 normalno porazdeljenih populacij z neznanima variancama in pri majhnih vzorcih je:

$$\langle x_1 \rangle \pm \langle x_2 \rangle - t_{n_1+n_2-2;\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < m_1 \pm m_2 < \langle x_1 \rangle \pm \langle x_2 \rangle + t_{n_1+n_2-2;\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \,, \tag{15}$$

kjer je s_p kombinirana vzorčna standardna deviacija oziroma s_p^2 varianca:

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \,. \tag{16}$$

Približni dvostranski interval zaupanja za vsoto (razliko) deležev p_1 in p_2 dveh populacij pri velikih vzorcih je:

$$\hat{p_1} \pm \hat{p_2} - z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}} < p_1 \pm p_2 < \hat{p_1} \pm \hat{p_2} + z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}} \,. \tag{17}$$

Pri tem predpostavljamo, da lahko binomski porazdelitvi aproksimiramo z normalnima.