Tilastomatematiikka

Ensimmäinen välikoe 18.02.2022

1. Tarkastellaan erään kuukauden aikana sairaalahoitoa vaativan koronatartuntojen esiintyvyyttä ikäryhmittäin ja rokotussuojan mukaan kokoa $N = 200000$ olevassa sairaudelle alttiissa populaatiossa.

		Ikäryhmät				
		12-29	30-49	50-69	70+	Yhteensä
Sairaalahoitoon	Ei rokotettu	2	5	5	2	
	Rokotettu	1	1	3	5	
Alttiita yhteensä		46221	57365	59061	37353	200000
Rokotusaste		82 %	85 %	92 %	95 %	

Taulukko 1: Sairaalahoitoa vaativan taudin esiintyminen ikäryhmittäin ja rokotussuojan mukaan

Vastaa aineiston perusteella seuraaviin kysymyksiin.

a) Millä todennäköisyydellä henkilö on alle 30-vuotias? (1p)

b) Millä todennäköisyydellä henkilö on rokotettu? (2p)

c) Jos henkilö on rokotettu, niin millä todennäköisyydellä hän saa sairaalahoitoa vaativan koronatartunnan? (3p)

2. a) Diskreetin satunnaismuuttujan X jakauma on

x	1	2	3
$\mathbb{P}(X = x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Laske odotusarvo $\mathbb{E}(X)$ ja varianssi $\mathrm{Var}(X)$.

b) Jatkuvan satunnaismuuttujan Y tiheysfunktio on

$$f(y) = \begin{cases} \frac{1}{3}, & 0 < y < 3, \\ 0, & \text{muulloin.} \end{cases}$$

Laske odotusarvo $\mathbb{E}(Y)$ ja todennäköisyys $\mathbb{P}(Y > 1)$.

3. Tuotteita poimitaan tuotantolinjalta satunnaisesti tarkastusta varten. Pitkän aikavälin kokemus on osoittanut, että keskimäärin 3% tuotteista on tavalla tai toisella rikki.

a) Eräänä maanantaina 110 tuotetta poimittiin tarkastettavaksi. Millä todennäköisyydellä poimittujen tuotteiden joukossa oli ainakin kolme tavalla tai toisella rikki?

b) Laske normaalijakauma-approksimaatiolla todennäköisyys, että 110 tuotteesta korkeintaan 5 on rikki? Käytä jatkuvuuskorjausta.

Tehtävien ratkaisuperiaatteet

1. a) Aineiston perusteella alle 30-vuotiaita on yhteensä 46221, joten henkilö on alle 30-vuotias todennäköisyydellä $p = \frac{46221}{200000} \approx 23\,\%$.

   b) Merkitään tapahtumia

$$\begin{aligned} R &= \text{"henkilö on rokotettu"}, \\ A_i &= \text{"henkilö kuuluu ikäryhmään } i, \quad i = 1,2,3,4, \end{aligned}$$

   missä ikäryhmä viittaa järjestyksessään taulukoissa annettuihin ikäryhmiin. Kokonaistodennäköisyyden mukaan

$$\begin{aligned} \mathbb{P}(R) &= \mathbb{P}(R|A_1)\mathbb{P}(A_1) + \mathbb{P}(R|A_2)\mathbb{P}(A_2) + \mathbb{P}(R|A_3)\mathbb{P}(A_3) + \mathbb{P}(R|A_4)\mathbb{P}(A_4) \\ &= 0.82 \cdot \frac{46221}{200000} + 0.85 \cdot \frac{57365}{200000} + 0.92 \cdot \frac{59061}{200000} + 0.95 \cdot \frac{37353}{200000} \\ &\approx 88\,\%. \end{aligned}$$

   c) Merkitään S = "henkilö saa sairaalahoitoa vaativan koronatartunnan". Kysytään todennäköisyyttä $\mathbb{P}(S|R)$, joka voidaan laskea vaikkapa "lukumäärien" avulla. Rokotettuja joutuu sairaalahoitoon yhteensä $1 + 1 + 3 + 5 = 10$ henkilöä. Edellisen kohdan mukaan rokotettuja on yhteensä $\mathbb{P}(R)N$, joten

$$\mathbb{P}(S|R) = \frac{\mathbb{P}(S \cap R)}{\mathbb{P}(R)} = \frac{10}{200000 \cdot \mathbb{P}(R)} \approx 5.7 \cdot 10^{-5},$$

   eli alle 6 henkilöä 100000:sta.

2. Molemmat satunnaismuuttujat ovat tasajakautuneita, joten odotusarvon voi päätellä "arkijärjellä".

   a) Koska jakauma on tasainen, niin "arkijärjen" mukaan odotusarvo on $\mathbb{E}(X) = 2$, joka voidaan todeta myös laskemalla

$$\mathbb{E}(X) = \sum_{k=1}^{3} k \cdot \mathbb{P}(X = k) = \frac{1+2+3}{3} = 2.$$

   Varianssi sen sijaan joudutaan laskemaan esimerkiksi kaavakokoelmasta löytyvällä kaavalla $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, jonka mukaan

$$\mathrm{Var}(X) = \sum_{k=1}^{3} k^2 \cdot \mathbb{P}(X = k) - 2^2 = \frac{1^2 + 2^2 + 3^2}{3} - 4 = \frac{2}{3}.$$

   c) Tässä molemmat menevät "intuitiolla" tai "arkijärjellä", joten odotusarvo $\mathbb{E}(X) = \frac{3}{2}$ on "puolivälissä" ja $\mathbb{P}(Y > 1) = \frac{2}{3}$, sillä $\{Y > 1\}$ kattaa $\frac{2}{3}$ välistä $[0,3]$. Sama asia voidaan todeta myös laskemalla

$$\mathbb{E}(X) = \int_0^3 y f(y)\mathrm{d}y = \int_0^3 \frac{1}{3} y = \frac{3}{2} \quad \text{ja} \quad \mathbb{P}(Y > 1) = \int_1^3 f(y)\mathrm{d}y = \int_1^3 \frac{1}{3}\mathrm{d}y = \frac{2}{3}.$$

3. a) Merkitään X = "rikki olevien tuotteiden lukumäärä otoksessa", joka noudattaa binomijakaumaa $\mathrm{Bin}(110, 0.03)$. Kysytään todennäköisyyttä $\mathbb{P}(X \geq 3)$. "Komplementin kautta" saadaan

$$\mathbb{P}(X \geq 3) = 1 - \mathbb{P}(X < 3) = 1 - (\mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2)) \approx 64\,\%.$$

b) Normaalijakauma-approksimaatiota varten tarvitaan tunnusluvut $\mathbb{E}(X) = np = 110 \cdot 0.03 = 3.3$ ja $\mathrm{Var}(X) = np(1-p) = 110 \cdot 0.03 \cdot 0.97 = 3.201$. Kysytyksi todennäköisyydeksi saadaan

$$
\begin{aligned}
\mathbb{P}(X \leq 5) &= \mathbb{P}(X \leq 5.5) && \text{(jatkuvuuskorjaus)} \\
&= \mathbb{P}\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}} \leq \frac{5.5 - 3.3}{\sqrt{3.201}}\right) && \text{(standardointi)} \\
&\approx \Phi(1.23) && \text{(normaaliapproksimaatio)} \\
&\approx 89\,\%. && \text{(taulukosta)}
\end{aligned}
$$

Binomijakaumalla laskettu tarkka todennäköisyys $88.6\,\% \approx 89\,\%$.

Kaavoja

Todennäköisyyden ominaisuuksia

$$
\begin{aligned}
&\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), \\
&\mathbb{P}(A \backslash B) = \mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B), \\
&\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A), \\
&\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B), \\
&\mathbb{P}(B|A) = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(A)}
\end{aligned}
$$

Odotusarvoja ja variansseja

Ptnf. tai tf.	$\mu_X := \mathbb{E}(X)$	$\sigma_X^2 := \mathrm{Var}(X)$
$\mathbb{P}(X = x)$	$\sum_x x\mathbb{P}(X = x)$	$\sum_x (x - \mu_X)^2 \mathbb{P}(X = x)$
$f_X(x)$	$\int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x$	$\int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\mathrm{d}x$
$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$
$p(1-p)^{x-1}$	$1/p$	$(1-p)/p^2$
$\frac{a^x}{x!}\mathrm{e}^{-a}$	a	a
$1/(b-a)$	$(a+b)/2$	$(b-a)^2/12$
$\theta \mathrm{e}^{-\theta x}$	$1/\theta$	$1/\theta^2$
$\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2

$$
\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y), \quad \mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2
$$

Eräitä testimuuttujia

$$
\begin{aligned}
&\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ (likimain, kun ”n on suuri”)}, \\
&\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}, \\
&\frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{1}{n} + \frac{1}{m}}\sqrt{\frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}}} \sim t_{n+m-2}, \\
&\sqrt{n-1}s_x \frac{\frac{S_{xY}}{s_{xx}} - \beta}{S_r} \sim t_{n-2}
\end{aligned}
$$

Regressio, korrelaatio ja kovarianssi

$$
\begin{aligned}
&r = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}}; \; s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}); \; s_{xx} = s_x^2; \\
&y = a + bx; \quad b = \frac{s_{xy}}{s_{xx}}; \quad a = \overline{y} - b\overline{x}; \\
&s_r^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - a - bx_i)^2 = \frac{n-1}{n-2}(1 - r^2)s_{yy}; \\
&\sigma_{XY} = \mathrm{Cov}(X,Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))), \quad \sigma_{XX} = \sigma_X^2; \\
&\qquad = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\
&\rho(X,Y) = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}
\end{aligned}
$$