Tilastomatematiikka

Toinen välikoe 14.03.2019

1. Tehtaan laatuinsinöörinä sinun on testattava, täyttääkö tuote tiettyjä laatukriteerejä. Sitä varten tuotannosta otetaan kokoa $n = 10$ otos, jonka perusteella tehdään koko tuotantoa koskevia päätelmiä. Olkoon laatukriteerinä jonkin ominaisuuden odotusarvo. Oletetaan, että ominaisuuteen liittyy satunnaisvaihtelua, joka noudattaa normaalijakaumaa.

   a) Jos hajonta on tuntematon, niin mikä on testimuuttuja?

   b) Jos hajonta tunnetaan, niin mikä on testimuuttuja?

   c) Jos hajonta tunnetaan ja vastahypoteesi on $H_1 : \mu \neq \mu_0$, niin mikä on kynnysarvo riskitasolla 2%?

   d) Jos hajonta on tuntematon ja vastahypoteesi on $H_1 : \mu > \mu_0$, niin mikä on kynnysarvo riskitasolla 1%?

   e) Jos hajonta on tuntematon, vastahypoteesi on $H_1 : \mu > \mu_0$ ja testimuuttujan arvo otoksessa on 3.07, niin mikä on johtopäätös riskitasolla 1%?

   f) Jos hajonta tunnetaan, vastahypoteesi on $H_1 : \mu \neq \mu_0$ ja testimuuttujan arvo otoksessa on 3.07, niin mikä on p-arvo?

2. Tutkittiin viherlipeän natriumsulfidikonsentraation y [g/l] riippuvuutta paperintuotannon volyymistä x yksikkönä tonnia per päivä ja saatiin seuraava havaintoaineisto

x	825	830	890	895	890	910	915	960	990	1010	1012	1030
y	40	42	49	46	44	48	46	43	53	52	54	57

   a) Määrää havaintoja vastaava regressiosuora ja korrelaatiokerroin.

   b) Piirrä havaintoja vastaava sirontakuvio ja regressiosuora samaan koordinaatistoon.

   c) Laske lineaarisen mallin antama ennuste ja residuaali, kun tuotanto on 910 tonnia.

3. Muuttujien X ja Y yhteisjakauman tiheysfunktio on

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{16}xy, & \text{kun } 0 < x < 2 \text{ ja } 0 < y < 4, \\ 0, & \text{muulloin.} \end{cases}$$

   a) Laske odotusarvo $E(X)$. (2p)

   b) Määrää muuttujien X ja Y välinen kovarianssi. Mitä voit kovarianssin perusteella sanoa muuttujien välisestä riippuvuudesta? (4p)

Kaavoja

$$
\begin{aligned}
&\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)\\
&\mathrm{P}(A \backslash B) = \mathrm{P}(A \cap \bar{B}) = \mathrm{P}(A) - \mathrm{P}(A \cap B)\\
&\mathrm{P}(\bar{A}) = 1 - \mathrm{P}(A)\\
&\mathrm{P}(A|B) = \mathrm{P}(A \cap B)/\mathrm{P}(B)\\
&f(x) = 1/(b-a),\ \mathrm{E}(X) = (a+b)/2,\ \mathrm{Var}(X) = (b-a)^2/12\\
&f(x) = \theta e^{-\theta x},\ \mathrm{E}(X) = 1/\theta,\ \mathrm{Var}(X) = 1/\theta^2\\
&\mathrm{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k},\ \mathrm{E}(X) = np,\ \mathrm{Var}(X) = np(1-p)\\
&\mathrm{P}(X=k) = p(1-p)^{k-1},\ \mathrm{E}(X) = 1/p,\ \mathrm{Var}(X) = (1-p)/p^2\\
&\mathrm{P}(X=k) = \frac{a^k}{k!} e^{-a},\ \mathrm{E}(X) = a,\ \mathrm{Var}(X) = a\\
&\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ (likimain, kun ”}n\text{ on suuri”)}\\
&\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}\\
&\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{\frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}}} \sim t_{n+m-2};\\
&r = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}};\ s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y});\ s_{xx} = s_x^2;\\
&y = a + bx; \quad b = \frac{s_{xy}}{s_{xx}}; \quad a = \overline{y} - b\overline{x}\\
&s_r^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \frac{n-1}{n-2}(1-r^2)s_{yy}\\
&\sqrt{n-1}\, s_x \frac{\frac{S_{xY}}{s_{xx}} - \beta}{S_r} \sim t_{n-2};\\
&\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y); \quad \mathrm{Var}(X) = \mathrm{E}((X - \mathrm{E}(X))^2);\\
&\sigma_{XY} = \mathrm{Cov}(X,Y) = \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))), \quad \sigma_{XX} = \sigma_X^2;\\
&\qquad = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)\\
&\rho(X,Y) = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}.
\end{aligned}
$$

Tehtävien ratkaisuperiaatteet

1. a) $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$.

   b) $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

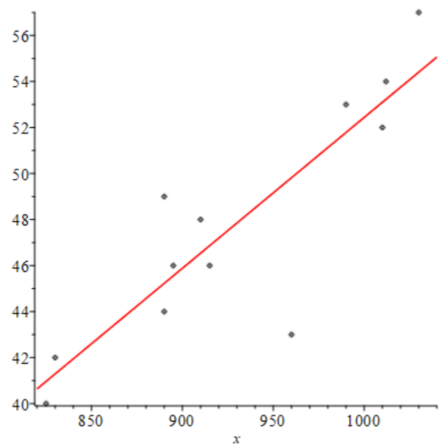
   c) Kynnysarvo c ratkeaa ehdosta $P(|Z| \leq c) = 0.98 \Leftrightarrow \Phi(c) = 0.99 \Rightarrow c = 2.33$.

   d) Kynnysarvoksi c saadaan t-jakauman taulukosta $c = 2.821$.

   e) Koska $3.07 > 2.821$, niin johtopäätös on H_1.

   f) P-arvoksi p saadaan $p = P(|Z| > 3.07) = 2 \cdot (1 - \Phi(3.07)) = 0.0022$.

2. a) Regressiosuora $y = -13.1 + 0.0656x$ ja korrelaatiokerroin $r \approx 0.866$.



Kuva 1: Tehtävän 2 regressiosuora ja sirontakuvio

   c) Mallin antama ennuste on $\hat{y}_i = -13.14 + 0.0656 \cdot 910 \approx 46.6$, joten residuaali on $y_i - \hat{y}_i \approx 48 - 46.6 = 1.4$.

3. a) $E(X) = \iint_{\mathbb{R}^2} x f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = \int_{x=0}^{2}\int_{y=0}^{4} \frac{1}{16}x^2 y \mathrm{d}x\mathrm{d}y = \frac{4}{3}$.

   b) Kuten a)-kohdassa saadaan

$$E(Y) = \iint_{\mathbb{R}^2} y f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = \int_{x=0}^{2}\int_{y=0}^{4} \frac{1}{16}xy^2 \mathrm{d}x\mathrm{d}y = \frac{8}{3}$$

ja

$$E(XY) = \iint_{\mathbb{R}^2} xy f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = \int_{x=0}^{2}\int_{y=0}^{4} \frac{1}{16}x^2y^2 \mathrm{d}x\mathrm{d}y = \frac{32}{9},$$

joten kovarianssi on

$$\sigma_{XY} = E(XY) - E(X)E(Y) = \frac{32}{9} - \frac{4}{3} \cdot \frac{8}{3} = 0.$$

Voidaan osoittaa, että X ja Y ovat riippumattomia, mutta kovarianssin perusteella emme voi sanoa mitään muuta muuttujien välisestä riippuvuudesta kuin sen, etteivät X ja Y riipu *lineaarisesti* toisistaan.