Tilastotieteen perusmenetelmät I Syksy 2011
2. välikoe 13.12.2011, Tehtävien ratkaisut

2. a)

a1) i) P(pääaine kansantaloustiede tai laskentatoimi)

$$= \frac{46+53}{257} = \frac{99}{257} = \underline{0.385}$$

ii) $P(\text{mies} \mid \text{pääaine johtaminen}) = \frac{11}{24} = \underline{0.458}$

tai $P(\text{mies} \mid \text{johtaminen}) = \dfrac{P(\text{mies} \cap \text{johtaminen})}{P(\text{johtaminen})}$

$$= \frac{11/257}{24/257} = \frac{11}{24} = \underline{0.458}$$

a3) X = naisten lkm seitsemän opiskelijan otoksessa

$$X \sim Bin\left(7, \frac{110}{257}\right)$$

$$P(X=k) = \binom{7}{k}\left(\frac{110}{257}\right)^k\left(1-\frac{110}{257}\right)^{7-k} \qquad k=0,1,\dots,7$$

$$P(X=4) = \underbrace{\binom{7}{4}}_{35}\left(\frac{110}{257}\right)^4\left(1-\frac{110}{257}\right)^{7-4} = 0.2198 = \underline{0.220}$$

b) X = taloustieteiden tiedekunnan opiskelijoiden lkm 150 opiskelijan otoksessa

$$X \sim Bin(\underbrace{150}_{n}, \underbrace{0.111}_{p})$$

$$\left.\begin{aligned} n \cdot p &= 150 \cdot 0.111 = 16.65 > 5 \\ n(1-p) &= 150 \cdot 0.889 = 133.35 > 5 \end{aligned}\right\} \Rightarrow$$

normaalijakauma-approksimaatio sallittu

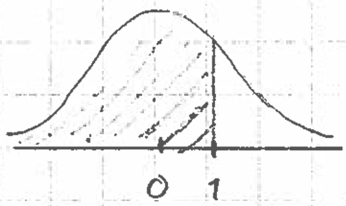
$$X \sim Bin(150,\ 0.111) \approx N(\underbrace{150 \cdot 0.111}_{np},\ \underbrace{150 \cdot 0.111 \cdot 0.889}_{np(1-p)})$$

$$= N(16.65,\ 14.80185)$$

$$P(X \le 20) = P(X \le 20.5) = P\Big(\underbrace{\frac{X-16.65}{\sqrt{14.80185}}}_{Z \sim N(0,1)} \le \frac{20.5-16.65}{\sqrt{14.80185}}\Big)$$

$$= P(Z \le 1.00) = 1 - P(Z \ge 1.00)$$

$$= 1 - 0.1587 = \underline{\underline{0.8413}}$$



$$P(X \le 20) = 0.8419054 \qquad (\text{tarkka tn})$$

1.

A	a2
B	b1
C	c6
D	d5
E	e5
F	f1

$$D^2(X) = \sum_{i=1}^{5} x_i^2 p_i - \mu^2$$

$$P(1 \le X \le 3) = F(3) - F(0)$$

$$E(X) = 0.4 \Rightarrow \alpha = \frac{1}{0.4} = 2.5$$

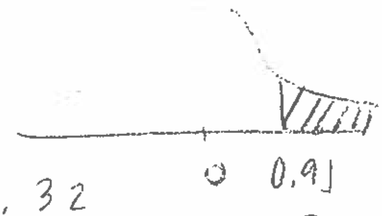
3.

X = kassasaldo $\overset{\text{likim.}}{\sim} N(25, 5.5^2)$

a) $P(X \geq 30) = P\left(\frac{X-25}{5.5} \geq \frac{30-25}{5.5}\right)$

$= P(Z \geq 0.909) \approx P(Z \geq 0.91)$

$= \underline{0.1814}$



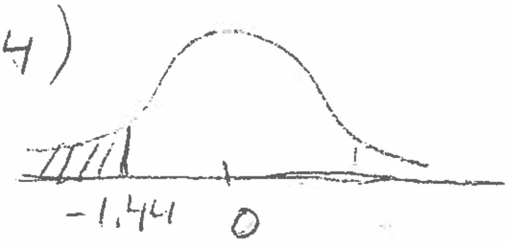
b) $\bar{X} \sim N(25, \frac{5.5^2}{7}) \quad 4.32 = 2.08^2$

$P(\bar{X} < 22) = P\left(\frac{\bar{X}-25}{5.5/\sqrt{7}} < \frac{22-25}{5.5/\sqrt{7}}\right)$

$\approx P(Z < -1.443) \approx P(Z < -1.44)$

$= P(Z > 1.44)$

$= \underline{\underline{0.0749}}$



$\frac{5.5}{\sqrt{7}} \approx 2.078805 = \sigma_{\bar{x}}$, $\sigma_{\bar{x}}^2 = 2.08^2 \approx 4.32$

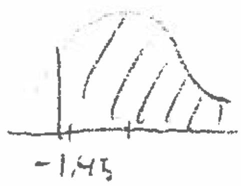
c) Y = Tarjan palvelurahn

$= \frac{X-5}{6} = \frac{1}{6}X - \frac{5}{6}$ $\quad Y \sim N(3.33, 0.917^2)$

$E(Y) = \frac{1}{6}E(X) - \frac{5}{6} = \frac{1}{6} \cdot 25 - \frac{5}{6} = \frac{10}{3} \approx 3.33$

$D^2(Y) = (\frac{1}{6})^2 D^2(X) = \frac{1}{36} \cdot 5.5^2 \approx 0.840$

$\Rightarrow D(Y) = \sqrt{0.840} = 0.91666...$



$P(Y > 2) = P\left(\frac{Y-3.33}{0.917} > \frac{2-3.33}{0.917}\right) = P(Z > -1.4545)$

$\approx P(Z > -1.45) = 1 - P(Z > 1.45) = 1 - 0.0735 = \underline{0.9265}$

4. 1) Oletetaan, että $(X_1, \dots, X_{400})$ on satunnaisotos $Bern(\pi_1)$ -jakaumasta ja $(Y_1, \dots, Y_{540})$ vastaavasti $Bern(\pi_2)$ -jakaumasta.

$$X_i = \begin{cases} 1, & \text{jos kantakaupunkilainen kannattaa ruuhkamaksua} \\ 0, & \text{muulloin} \end{cases}$$

$$Y_j = \begin{cases} 1, & \text{jos esikaupunkialueelle asuva kannattaa ruuhkamaksua} \\ 0, & \text{muulloin} \end{cases}$$

$\pi_1 = P(X_i = 1)$, $i = 1 \dots 400$

$\pi_2 = P(Y_j = 1)$, $j = 1 \dots 540$

2) $\begin{cases} H_0: \pi_1 = \pi_2 \\ H_1: \pi_1 \neq \pi_2 \end{cases}$

3) $$z = \frac{P_1 - P_2}{\sqrt{P(1-P)(\frac{1}{n} + \frac{1}{m})}} \overset{\text{likimain}}{\sim} N(0,1),$$ kun H_0 on tosi

4) $$P = \frac{nP_1 + mP_2}{n+m} = \frac{400 \cdot \frac{176}{400} + 540 \frac{189}{540}}{400 + 540} \approx 0.3883$$

$$z = \frac{0.44 - 0.35}{\sqrt{0.3883(1 - 0.3883)(\frac{1}{400} + \frac{1}{540})}} \approx 2.799 \quad (\approx 2.80)$$

5) $P\text{-arvo} = P(z \leq -2.80 \text{ tai } z \geq 2.80) = 2 \cdot 0.0026 = 0.0052$

6) Tämän tutkimuksen perusteella yksityisautojen ruuhkamaksujen kannatusosuudet poikkeavat toisistaan (ovat erisuuret) kantakaupungissa ja esikaupunkialueella asuvien 18-35 -vuotiaiden keskuudessa.

(5.) popul. 1: ravintolassa A tehdyt tilaukset

$X \xrightarrow{\text{sat. otos}} (X_1, X_2, \ldots, X_{12})$

X = tilauksen odotusaika (min)

Olet.: $X \sim N(\mu_x, \sigma_x^2)$

popul. 2: ravintolassa B tehdyt tilaukset

$Y \xrightarrow{\text{sat. otos}} (Y_1, Y_2, \ldots, Y_{20})$

Y = tilauksen odotusaika (min)

Olet.: $Y \sim N(\mu_y, \sigma_y^2)$

Oletetaan lisäksi, että $\sigma_x^2 = \sigma_y^2 = \sigma^2$

a) Lasketaan luottamusväli odotusarvojen erotukselle $\mu_x - \mu_y$. Kyseinen lv. on muotoa

$$\left((\bar{X}-\bar{Y}) - t_{\alpha/2} \cdot S \cdot \sqrt{1/n + 1/m}\ ,\ (\bar{X}+\bar{Y}) + t_{\alpha/2} \cdot S \cdot \sqrt{1/n + 1/m}\right)$$

$$S^2 = \frac{(n-1)\cdot S_x^2 + (m-1)\cdot S_y^2}{n+m-2} = \frac{(12-1)\cdot 3.762^2 + (20-1)\cdot 4.096^2}{12+20-2}$$

$$\approx 15.8149 \Rightarrow S \approx \sqrt{15.8149}$$

a1) 95 % lv.: $100(1-\alpha) = 95 \Rightarrow \alpha = 0.05$ ja $\alpha/2 = 0.025$

$t(n+m-2) = t(12+20-2) = t(30)$-jakauman taulukko: $t_{\alpha/2} = t_{0.025} = 2.042$

$\Rightarrow$ kysytty lv. on: $(\bar{X}-\bar{Y}) \pm t_{\alpha/2} \cdot S \cdot \sqrt{1/n + 1/m}$

$$\Leftrightarrow (20.17 - 17.40) \pm 2.042 \cdot \sqrt{15.8149} \cdot \sqrt{1/12 + 1/20}$$

$$\Leftrightarrow 2.77 \pm 2.965 \Leftrightarrow \underline{\underline{(-0.195, 5.735)}}$$

Koska nolla kuuluu lasketun lv.:n sisään, ei ravintoloiden A ja B keskimääräisissä tilausten odotusajoissa näyttäisi olevan eroa.

a2) 90 % lv.: $100(1-\alpha) = 90 \Leftrightarrow \alpha = 0.10$ ja $\alpha/2 = 0.05$

$t(30)$-jakauman taulukko: $t_{\alpha/2} = t_{0.05} = 1.697$

$\Rightarrow$ kysytty lv. on $(20.17-17.40) \pm 1.697 \cdot \sqrt{15.8149} \cdot \sqrt{1/12 + 1/20}$

$\Leftrightarrow 2.77 \pm 2.464 \Leftrightarrow \underline{\underline{(0.306, 5.234)}}$

Koska nolla ei kuulu lasketun lv.:n sisään, ravintoloiden A ja B keskimääräiset tilausten odotusajat eivät näyttäisi olevan samat (ravintolassa A keskimääräinen odotusaika on pidempi).

b) Merkitsevyystestauksen hypoteesit: $\begin{cases} H_0: \mu = 15 \text{ min} \\ H_1: \mu \neq 15 \text{ min} \end{cases}$

P-arvo = 0,0005923 $\Rightarrow$ aineisto on ristiriidassa H_0:n kanssa, joten mainos ei näytä havaintojen perusteella uskottavalta.

Parametrin μ 95 % lv. on (17.77650 min, 22.55683 min). Mainoksessa esitetty 15 min. ei kuulu lv:n sisään, joten johtopäätös on sama kuin testauksessa.

c) Tulostuksessa on parametrin μ lv., joka on muotoa

$$\left(\bar{X} - t_{\alpha/2} \cdot s_x/\sqrt{n}\,,\ \bar{X} + t_{\alpha/2} \cdot s_x/\sqrt{n}\right)$$

99 % lv.: $100(1-\alpha) = 99 \Rightarrow \alpha = 0.01$ ja $\alpha/2 = 0.005$

$t(n-1) = t(12-1) = t(11)$-jakauman taulukko: $t_{\alpha/2} = t_{0.005} = 3.106$

$\Rightarrow$ kysytty lv. on $20.17 \pm 3.106 \cdot 3.762/\sqrt{12} \Leftrightarrow 20.1$

$\Leftrightarrow 20.17 \pm 3.373$

$\Leftrightarrow \underline{\underline{(16.797, 23.543)}}$