Aalto-yliopisto

MS-A0502 Todennäköisyyslaskennan ja tilastotieteen peruskurssi — Gripenberg, Kiiski
Harjoitus 4
17–21.11.2014, viikko 47

Vastaa Stack-tehtäviin (`stack3.aalto.fi/course/view.php?id=30`)
viimeistään 24.11.2014 kl. 12.00.

Palauta P-tehtävät viimeistään 24.11.2014 kl. 12.
Muista kirjoittaa nimesi, opiskelijanumerosi ja harjoitusryhmäsi!

P1. Alla on lueteltu joukko muuttujia:

(1) Mansikoiden C-vitamiinipitoisuus: mg/100g.
(2) Alvarin aukiolta löydetyn kasvin laji.
(3) Paine, joka vaaditaan teräksisen säiliön murtumiseen: N/m^2.
(4) Henkilöiden mielipide metropolihallinnosta mitattuna asteikolla ”kannatan”, ”ei mielipidettä”, ”vastustan”.
(5) Espoo Bluesin sijoitus jääkiekkoliigassa.
(6) Opiskelijan koulutusohjelma.
(7) Opiskelijan kurssin MS-A0502 suorittamisvuosi.

Vastaa seuraaviin kysymyksiin

(a) Millä muuttujilla on nominaali- eli laatueroasteikko?
(b) Millä on ordinaali- eli järjestysasteikko?
(c) Millä muutujilla on intervalli- eli välimatka-asteikko?
(d) Millä muuttujilla on ratio- eli suhdeasteikko?

Huom! *Kaikissa tapauksissa ei ole välttämättä olemassa yksi ainoa oikea vastaus!*

P2. Erään talon joillakin asukkailla on seuraavat kuukausitulot (e/kk):

3300	3200	1700	4400	4000	4200	1400	4100	1700
4300	1900	2200	2200	2400	2600	3100	2900	3200
800	1700	2100	2300	2500	2700	1900	1800	1500
1200	700	6400	7500	1600	1800	2000	2200	2500.

Muodosta aineistosta luokiteltu frekvenssijakauma, jonka luokat ovat: $0-1000$, $1001-2000$, $2001-3000$, $3001-4000$, $4001-6000$, ja $6001-8000$. Määritä myös tätä frekvenssijakaumaa vastaavan histogrammikuvion suorakaiteiden korkeudet ottaen huomioon, että pinta-alojen pitää suhtautua toisiinsa kuten vastaavat luokkafrekvenssit. Hahmottele myös ko. histogrammikuvio paperille.

Määritä aineistosta myös seuraavat tunnusluvut:

(a) Minimi ja maksimi.
(b) Vaihteluväli (eli pienin väli $[a, b]$ niin, että kaikki arvot ovat tällä välillä) ja vaihteluvälin pituus.
(c) Mediaani.

P3. Sinulla on otos satunnaismuuttujasta, joka oletetaan olevan normaalijakautunut. Otoksen koko on 30 ja sen keskiarvo on 2.5. Miten iso otoksesta laskettu otosvarianssi voi korkeintaan olla jos saat odotusarvolle 98%:n (symmetrisen) luottamusvälin, jonka pituus on korkeintaan 1.8?

Vastaus: 4.01

P4. Olkoot $X_i, i = 1, 2, \ldots, n$ riippumattomia satunnaismuuttujia siten, että $\mathrm{E}(X_i) = 0$ ja $\mathrm{Var}(X_i) = \sigma^2$ kun $i = 1, 2, \ldots, n$.
Osoita, että $\mathrm{E}(S^2) = \sigma^2$ kun

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

missä $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ eli otosvarianssi on varianssin **harhaton** estimaattori.

Vihje: Laske $\mathrm{E}\big((n-1)S^2\big)$ *ja huomaa, että koska satunnaismuuttujien* X_i *odotusarvot ovat* 0 *niin* $\mathrm{E}(X_i^2) = \mathrm{Var}(X_i)$ *ja* $\mathrm{E}(\overline{X}) = 0$ *joten* $\mathrm{E}(\overline{X}^2) = \mathrm{Var}(\overline{X})$ *ja muista mikä* $\mathrm{Var}(\overline{X})$ *on kun* X_i*:t ovat riippumattomia.*

Huom! *Jos* $\mathrm{E}(X_i) = \mu \neq 0$ *niin voisimme ottaa satunnaismuuttujiksi* $Y_i = X_i - \mu$ *eikä* S^2 *muuttuisi vaikka jokainen* X_i *korvattaisiin* Y_i*:llä joten oletus että odotusarvo on* 0 *ei ole rajoitus, pelkästään yksinkertaistus.*

P5. Satunnaismuuttujan X tiheysfunktio on

$$f(x, \theta) = \begin{cases} (\theta - 1)x^{-\theta}, & x \geq 1, \\ 0, & x < 1, \end{cases}$$

missä parametri $\theta > 1$. Satunnaismuuttujasta on saatu havainnot 2, 5 ja 14.

(a) Estimoi θ momenttimenetelmällä.

(b) Estimoi θ suurimman uskottavuuden menetelmällä.

Vihje: Muista, että $\int_1^\infty x(\theta - 1)x^{-\theta}\,\mathrm{d}x = \frac{\theta-1}{\theta-2}$ *kun* $\theta > 2$ *ja että* $\frac{\mathrm{d}}{\mathrm{d}\theta} a^\theta = a^\theta \ln(a)$*. Kun lasket suurimman uskottavuuden estimaattia, voit menetellä kuten tämän viikon stack-tehtävässä 7, sillä erolla että tässä voit ratkaista ääriarvokohdan analyyttisesti.*

Vastaus: ≈ 2.17 ja ≈ 1.61