MTTTP5

Lisätehtäviä ratkaisuineen luentomonisteen lukuun 5 liittyen

1. Olkoon puolueen A kannatusosuus populaatiossa 30 %. Tarkastellaan tästä populaatiosta tehtyä satunnaisotosta, jonka koko on n. Määritellään satunnaismuuttuja X= puolueen A kannattajien lukumäärä otoksessa. Määritä X:n jakauma sekä jakauman odotusarvo ja varianssi. Tarkastele vielä suhteellista osuutta X/ n. Antaako X/ n keskimäärin oikean arvion kannatusosuudelle? Miksi?

2. Olkoon $X_1, X_2, \ldots , X_n$ satunnaisotos jakaumasta, jonka odotusarvo on μ ja varianssi σ^2. Estimoidaan jakauman odotusarvoa estimaattorilla $\overline{X}=(X_1+ \ldots + X_n)/ n$. Onko estimaattori harhaton? Määritä lisäksi estimaattorin keskivirhe. Mitä voit sanoa estimaattorin jakaumasta? Jos σ^2 on tuntematon, niin miten voisit estimoida keskivirhettä?

3. Auton sytytystulppien valmistaja väittää, että tulpat kestävät keskimäärin 60000 km keskihajonnan ollessa 6000 km sekä vaihtelu luonnehdittavissa normaalijakaumalla. Haluat estimoida tulppien keskimääräistä kestoa 4 satunnaisesti valitun tulpan avulla. Mitä estimaattoria käytät? Onko se harhaton? Määritä estimaattorisi keskivirhe sekä jakauma olettaen valmistajan väite todeksi.

4. Oletetaan, että puolueen A kannatusosuus populaatiossa on 18 %. Tehdään populaatiosta n alkion satunnaisotos. Määritellään kannatusosuuden estimaattori p = 100X/ n, missä X on puolueen A kannattajien lukumäärä otoksessa. Onko p kannatusosuuden harhaton estimaattori? Miksi? Määritä lisäksi estimaattorin keskivirhe.

5. Väliestimoi tehtävän 3 tilanteessa tulppien keskimääräinen kestoa.

6. Erään rikollisen puolustusasianajaja väittää, että kyseisessä tuomioistuimessa valamiehistöt eivät ole edustavia. Yhtenä perusteluna väittämäänsä hänellä on se, että valamiehistö on usein ollut keskipalkaltaan koko maan tasoa korkeampi. Tuorein tilastotieto maan keskipalkasta on \$8500. Selvitettiin 100 viimeisimmän valamiehistöön kuuluvien jäsenten palkat. Saatiin keskipalkaksi \$22890 ja palkan keskihajonnaksi \$7670. Onko perusteltua uskoa asianajajan väite? Tutki asiaa sopivan luottamusvälin avulla.

7. Tutkittiin kahden lisäaineen (A ja B) vaikutusta teräksen kovuuteen. Koska teräksen tuote-erien laatu vaihtelee, poimittiin näytteet 10 tuote-erästä, joista kukin jaettiin edelleen kahtia. Toiseen osaan lisättiin lisäainetta A ja toiseen lisäainetta B. Mitattiin kovuusindeksi:

Tuote-erä	1	2	3	4	5	6	7	8	9	10
Lisäaine A	22	26	29	22	31	34	31	20	33	34
Lisäaine B	27	25	31	27	29	41	32	27	32	34

Onko perusteita pitää toista lisäainetta parempana kuin toista? Koska otokset ovat riippuvia, laske ensin kovuusindeksien erotukset tuote-erittäin ja sitten sopiva luottamusväli.

8. Lentoyhtiön ongelmana on matkustajien varaamien paikkojen käyttämättä jättäminen. Tästä syystä yhtiö ottaa lennoille varauksia enemmän kuin koneessa on paikkoja. Halutaan arvioida sitä, kuinka paljon (prosentteina) varauksia voidaan ottaa yli paikkojen. Tehdään 500 matkustajan satunnaisotos, joiden joukossa todetaan olevan 40 matkustajaa, jotka olivat jääneet tulematta varatulle lennolle. Anna ohjeet yhtiölle varauksien ottamisesta. Käytä sopivaa luottamusväliä.

9. Erään mielipidemittauksen mukaan presidenttiehdokkaan kannatus on 52%. Mielipidettä oli kysytty 1500 äänioikeutetulta siten, että voidaan olettaa olevan kyse satunnaisotannasta. Laske virhemarginaali. Jos otoskoko olisi ollut 300, niin ikä olisi ollut virhemarginaali?

10. Kymmenen vuotta sitten eräässä yliopistossa tehdyn tutkimuksen mukaan 18 % yliopiston opiskelijoista ei uskonut löytävänsä koulutustaan vastaavaa työtä valmistuttuaan. Haluttiin tutkia, oliko tämä prosenttiosuus muuttunut ja kysyttiin mielipidettä 100 satunnaisesti valitulta opiskelunsa aloittaneelta, joista 25 % ei uskonut työllistyvänsä koulutustaan vastaavasti valmistumisen jälkeen. Onko kymmenessä vuodessa tapahtunut muutosta?

11. Ohessa analysointituloksia liittyen Tampereella myynnissä olleisiin kerrostalohuoneistoihin (Aineisto Aamulehti 31.10.99). Tuloksista löytyy neliöhinnan tunnuslukuja laskettuna sekä keskusta- että esikaupunkialueelta. Laske sopiva luottamusväli ja tulkitse tulokset.

Level	**Number**	**Mean**	**Std Dev**
Esikaupunki	26	7250,59	1691,78
Keskusta	30	9613,15	1278,22

Ratkaisuja

1.
$X \sim Bin(n, 0.3)$, jolloin $E(X) = n \cdot 0.3$ ja $Var(X) = n \cdot 0.3 \cdot 0.7$.
$E(X/n) = (1/n)E(X) = (1/n) \cdot n \cdot 0.3 = 0.3$. Koska otoksessa A:n kannattajien suhteellisen osuuden odotusarvo on 0.3, niin antaa X/n keskimäärin oikean arvion kannatusosuudelle. Voidaan tietysti yhtä hyvin tarkastella prosentuaalista osuutta $100X/n$, jonka odotusarvo on 30.

2.
$E(\overline{X}) = E((X_1 + ... + X_n)/n) = \{E(X_1) + ... + E(X_n)\}/n = (\mu + ... + \mu)/n = \mu$, joten $\overline{X}$ on μ:n harhaton estimaattori. $Var(\overline{X}) = Var((X_1 + ... + X_n)/n) =$
$\{Var(X_1) + ... + Var(X_n)\}/n^2 = (\sigma^2 + ... + \sigma^2)/n^2 = \sigma^2/n$, joten estimaattorin keskivirhe on $\sigma/\sqrt{n}$, jota voi estimoida $s/\sqrt{n}$. Jos otos normaalijakaumasta, niin keskiarvon jakauma on myös normaalijakauma. Jos otos ei ole normaalijakaumasta, niin keskeisen raja-arvolauseen perusteella otoskeskiarvo on likimain normaalisti jakautunut, kunhan otoskoko riittävä.

3.

Otoskeskiarvo on odotusarvon harhaton estimaattori. Tässä (olettaen valmistajan väite oikeaksi) $\overline{X} \sim N(60000, 6000^2/4)$, joten $\overline{X}$:n keskivirhe on $6000/2 = 3000$.

4.
$X \sim Bin(n, 0.18)$, jolloin $E(X) = n \cdot 0.18$ ja $Var(X) = n \cdot 0.18 \cdot 0.82$.
$E(p) = E(100X/n) = (100/n)E(X) = (100/n) \cdot n \cdot 0.18 = 18$, joten p on kannastusprosentin harhaton estimaattori. $Var(p) = Var(100X/n) =$
$(100/n)^2 Var(X) = (100/n)^2 \cdot n \cdot 0.18 \cdot 0.82 = 18 \cdot 82/n$, jonka neliöjuuri on estimaattorin hajonta eli keskivirhe.

5.
$100(1-\alpha)$%:n luottamusväli odotusarvolle, kun jakauman varianssi tunnettu, on $\overline{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$. Saadaan $60000 \pm 1{,}96 \cdot 6000/2$.

6.
$100(1-\alpha)$%:n luottamusväli odotusarvolle, kun jakauman varianssi on tuntematon, on $\overline{X} \pm t_{\alpha/2, n-1} s/\sqrt{n}$. Tässä $t_{0.05/2, 100-1} = 1.98$, $\overline{x} = 22890$ ja $s = 7670$, joten 95%:n luottamusväli $22890 \pm 1.98 \cdot 7670/\sqrt{100}$ eli 22890 ± 1519, joka ei sisällä lukua 8500 (maan keskipalkkaa). On siis perusteltua uskoa asianajajan väite.

7.

Tuote-erä	1	2	3	4	5	6	7	8	9	10
Lisäaine A	22	26	29	22	31	34	31	20	33	34
Lisäaine B	27	25	31	27	29	41	32	27	32	34
EROTUS	**-5**	**1**	**-2**	**-5**	**2**	**-7**	**-1**	**-7**	**1**	**0**

Käytetään , kuten tehtävässä 1, luottamusväliä odotusarvolle, kun populaation varianssi on tuntematon. 100(1-α)%:n luottamusväli odotusarvolle, kun jakauman varianssi on tuntematon, on $\overline{X} \pm t_{\alpha/2,\, n-1} s/\sqrt{n}$. Erotuksista laskettu keskiarvo on -2.3 (tai 2.3, jos erotukset laskettu toisin päin) ja keskihajonta (kaavasta 1.2) s = 3.4335, $t_{0.05/2,\, 10-1} = 2.262$, joten 95%:n luottamusväli on -2.3 ± 2.262 · 3.4335/ $\sqrt{10}$ eli -2.3 ± 2.456, joka sisältää nollan. Näin voidaan ajatella erotuksen olevan peräisin jakaumasta, jonka odotusarvo on nolla. Siis lisäaineet saman vertaisia.

8.

Prosenttiosuuden luottamusväli $p \pm z_{\alpha/2}\sqrt{\frac{p(100-p)}{n}}$. Muodostetaan 95%:n luottamusväli ($\alpha = 0.05$, $z_{\alpha/2} = 1.96$), saadaan $8 \pm 1.96\sqrt{\frac{8(100-8)}{500}}$ eli (5.6, 10.4). Voidaan arvioida, että paikan varanneista 5.6% - 10.4% jää tulematta, joten varauksia voidaan ottaa tämän mukaisesti yli.

9.

Virhemarginaali (95%) on $\pm 1.96\sqrt{\frac{p(100-p)}{n}}$ (ks. tehtävän 8 luottamusväli). Nyt p = 52. Jos n = 1500 virhemarginaali on ±2.5%, jos n = 3000 virhemarginaali on ±1.8%.

10.

Prosenttiosuuden luottamusväli $p \pm z_{\alpha/2}\sqrt{\frac{p(100-p)}{n}}$. Muodostetaan 95%:n luottamusväli ($\alpha = 0.05$, $z_{\alpha/2} = 1.96$), saadaan $25 \pm 1.96\sqrt{\frac{25(100-25)}{100}}$ eli (16.5, 33.5). Koska 18 % kuuluu tälle välille voidaan tehdä päätelmä, että tilanne ei ole muuttunut.

11.

Neliöhinta markkoina!!

Level	Number	Mean	Std Dev
Esikaupunki	**n** = 26	$\bar{x}$=7250,59	$\mathbf{s_x}$ =1691,78
Keskusta	**m** = 30	$\bar{y}$=9613,15	$\mathbf{s_y}$ =1278,22

100(1-α)%:n luottamusväli odotusarvojen erotukselle, kun jakauman varianssit ovat tuntemattomia, mutta yhtä suuria, on $\bar{X} - \bar{Y} \pm t_{\alpha/2, n+m-2}\, s\sqrt{\frac{1}{n} + \frac{1}{m}}$, missä s = $\sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}}$. Tässä 95%:n luottamusväli (α = 0.05, $t_{0.05/2,\, 26+30-2} \approx 2$)

$- 2362.56 \pm 2 \cdot 1484.08\sqrt{\frac{1}{26} + \frac{1}{30}}$ eli -2362.56 ± 795.16. Koska nolla ei kuulu luottamusvälille, voidaan sanoa, että neliöhinnat eivät ole keskimäärin samoja vaan keskusta-asunnot ovat 1567.4 mk - 3157.72 mk neliöhinnaltaan kalliimpia.