

# MS-A0502 Todennäköisyysslaskennan ja tilastotieteen peruskurssi

Tentti ja välikokeet 8.1.2014/ Aro

Kirjoita selvästi *jokaiseen koepaperiin* seuraavat tiedot:

- MS-A0502 8.1.2014 ja mitä vaihtoehtoa olet suorittamassa (ks. alla)
- opiskelijanumero + kirjain
- TEKSTATEN sukunimi ja kaikki etunimet
- koulutusohjelma ja vuosikurssi
- nimikirjoitus

Lue *tarkasti* seuraavat ohjeet:

- (1) Jos olet suorittamassa kurssia tentillä, vastaa kysymyksiin 1-5.
- (2) Jos olet suorittamassa 1. välikoetta erikoisluvalla, vastaa kysymyksiin 1, 2, 6 ja 7.
- (3) Jos olet suorittamassa 2. välikoetta erikoisluvalla, vastaa kysymyksiin 3-5 ja 8.

Tentin kesto on 3 tuntia. 1. välikokeen kesto on 2 tuntia ja 2. välikokeen 3 tuntia.

**Merkitse vastauspapereihisi selvästi mitä vaihtoehtoa olet suorittamassa!**

Sallitut apuvälineet: *Laskin* ja Mellinin kaava- ja taulukkokokoelmat.

Vastausohjeet: **Vastaa lyhyesti ja ytimekkäästi, mutta perustele ratkaisusi.**  
**Esimerkiksi pelkkä lukuarvo vastauksena ei anna pisteitä.**

1. (a) Olkoot  $\Pr(A) = 0.5$  ja  $\Pr(B) = 0.1$ . Määritä tapahtuman  $A \cup B$  todennäköisyys, kun  
Tehtävä 1: A ja B ovat toisensa poissulkevia  
Tehtävä 2:  $\Pr(A|B) = 0.5$   
Tehtävä 3: A ja B ovat riippumattomia  
(b) Erään liikeyrityksen työntekijöistä 1 % käyttää huumeita. Pikatesti paljastaa 95 % huumeiden käyttäjistä, mutta toisaalta 2 % niistä, jotka eivät käytä huumeita tulee virheellisesti luokitelluiksi huumeiden käyttäjiksi. Määrää todennäköisyys, että pikatestissä huumeiden käyttäjäksi luokiteltu henkilö ei todellisuudessa käytä huumeita.
2. Tehdas valmistaa tuotetta, jolla on erittäin korkeat laatuksiteerit. Keskimäärin vain 60 % tuotteista täyttää kriteerit. Poimitaan tuotteita tarkastettavaksi satunnaisesti yksi kerrallaan.  
Tehtävä 1: Mikä on todennäköisyys sille, että joudumme tarkastamaan vähintään 4 tuotetta *ensimmäisen* viallisen tuotteen löytämiseksi?  
Tehtävä 2: Mikä on odotettavissa oleva lukumäärä tuotteille, jotka joudumme tarkastamaan *ensimmäisen* viallisen tuotteen löytämiseksi?  
Tehtävä 3: Mikä on todennäköisyys, että joudumme tarkastamaan ainakin 5 tuotetta *kolmannen* viallisen tuotteen löytämiseksi?

3. Mitattaessa neljän henkilön verenpainetta saatiin seuraavat tulokset (X on henkilön ikä ja Y henkilön verenpaine):

x	40	55	60	65
y	125	152	150	158

$$\bar{x} = 55.00 \quad s_x^2 = 116.67$$

$$\bar{y} = 146.25 \quad s_y^2 = 212.25$$

$$s_{xy} = 151.67$$

**Tehtävä 1:** Määrää tavanomaisen yhden selittäjän lineaarisen regressiomallin

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i : N(0, \sigma^2), i = 1, 2, \dots, n$$

regressiokertoimien  $\beta_0$  ja  $\beta_1$  pienimmän neliösumman (PNS-) estimaatit.

**Tehtävä 2:** Jännösvarianssin  $\sigma^2$  harhattomaksi estimaatiksi tutkijat laskivat arvon  $s^2 = 22.625$ . Ennusta muuttujan y keskimääräinen arvo, kun  $x = 50$ . Määrää myös 95%:n luottamusväli ennusteelle.

4. Kokeessa verrattiin menetelmiä A ja B englannin opetuksessa.

Kokeeseen valittiin peruskoulun 1. luokan oppilaiden joukosta 7 oppilasparia siten, että jokaisessa parissa kummankin oppilaan ikä kuukausina, sukupuoli, älykkyys, terveys, sosiaalinen tausta jne. olivat mahdollisimman samankaltaisia. Toinen *sovitetun parin* jäsenistä valittiin arvalla saamaan opetusta menetelmällä A ja toista opetettiin menetelmällä B.

Opetusjakson jälkeen kaikki kokeessa mukana olleet oppilaat joutuivat samaan testiin. Tulokset testistä (testipisteet) on annettu alla (korkeampi pistemäärä osoittaa parempaa englannin taitoa).

	Menetelmä	
Pari	A	B
1	73	74
2	76	71
3	95	91
4	69	71
5	68	71
6	94	81
7	56	55

Testaa 5 %:n merkitsevyystasoa käyttäen nollahypoteesia, että menetelmät A ja B tuottavat ko. testillä mitattuna keskimäärin yhtä hyviä tuloksia, kun vaihtoehtoisena hypoteesina on, että menetelmät eivät tuota yhtä hyviä tuloksia.

Havainnot saa olettaa normaalijakautuneiksi.

5. Valimo valmistaa palkkeja. Palkkien paksuus (tuumaa) vaihtelee satunnaisesti noudattaen normaalijakaumaa odotusarvona  $\mu$  ja varianssina  $\sigma^2$ , jotka ovat tuntemattomia. Palkkien joukosta poimittiin yksinkertainen satunnaisotos jonka koko oli  $n=7$  ja mitattiin palkkien paksuus: 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, 9.6.
- (a) Määrää odotusarvon ja varianssin harhattomat estimaatit.
- (b) Määrää 95 %:n luottamusväli palkkien paksuuden odotusarvolle.

6. Satunnaismuuttujan  $X$  tiheysfunktio on muotoa

$$f(x) = \begin{cases} ax^2 - 2ax, & \text{kun } 0 \leq x \leq 1 \\ 0, & \text{muulloin} \end{cases}$$

- (a) Määrää vakion  $a < 0$  arvo.
- (b) Määrää satunnaismuuttujan  $X$  kertymäfunktio.
- (c) Määrää tapahtuman  $X = 0.5$  todennäköisyys.
- (d) Määrää tapahtuman  $0 \leq X \leq 0.25$  todennäköisyys tiheysfunktion avulla.
- (e) Määrää tapahtuman  $0 \leq X \leq 0.25$  todennäköisyys kertymäfunktion avulla.

7. Heität virheetöntä noppaa 24000 kertaa. Mikä on todennäköisyys, että kuutosten lukumäärä on suljetulla välillä [3920, 4160]? Käytä normaalijakauma-approksimaatiota.

8. Eräs taho on julkaissut lukuja, jotka kertovat kuinka monta kilowattituntia eri kodin sähkölaitteet kuluttavat vuosittain. Se väittää, että pölynimuri kuluttaa keskimäärin 46 kilowattituntia vuodessa.

Valitaan 12 pölynimurin suuruinen yksinkertainen satunnaisotos, ja kirjataan imurien vuosittainen kilowattituntien kulutus. Aritmeettiseksi keskiarvoksi saadaan  $\bar{x} = 42$  ja otosvarianssiksi  $s^2 = 141.61$ . Löytyykö tästä riittävästi todisteita 5% merkitsevyystasolla, että pölynimurit kuluttavat keskimäärin vähemmän kuin 46 kilowattituntia vuosittain? Oleta, että pölynimurin vuosittain kuluttamien kilowattituntien määrä on normaalijakautunut.