

Tilastomenetelmien peruskurssi (TILP150)

Syksy 2006

6. harjoitukset / viikko 42

Näissä harjoituksissa käsitellään Lukuja 8.4, 8.5, 8.5.1 (1. tehtävä), 8.5.2 (2. tehtävä), 8.6, 8.6.1 (3. ja 6. tehtävä) ja 8.6.2 (5. tehtävä).

1. Syksyn 2006 TILP150-kurssilaisilta kysyttiin viikoittaista työssäolotuntien määrää seuraavasti:

TYÖ="Opiskelujen ohella teen palkallista työtä keskimäärin viikossa lukukauden aikana (tunteina)"

Alla on esitetty muuttujan TYÖ frekvenssitaulu:

TYÖ (h)	frekv
0	90
1-10	40
11-20	16
21-30	5
yli 30	4
yht.	155

Joissakin tutkimuksessa on kerrottu, että puolet opiskelijoista käy töissä. Arvioidaan tilastollisella testillä muuttujasta TYÖ tehtyjen havaintojen perusteella, pitääkö väite paikkansa.

Oletetaan, että saatu aineisto on satunnaisotos opiskelijoista ($n = 155$). Merkitään tuntematonta työssä käyvien osuutta π :llä. Aloita laskemalla havaittu työssä käyvien osuus p . Huom. Työssä käyviä ovat sellaiset, joiden ilmoittama tuntimäärä on vähintään yksi (1). Mieti sen jälkeen, mitä tilastollista testiä käyttäisit, ja suorita ko. testi.

Vinkki:

- 1) Kirjaa (sekä sanallisesti että symboleita käyttäen) nollahypoteesi ja kaksisuuntainen vastahypoteesi
- 2) Tutki oletukset: $n \cdot \pi_0$ ja $n \cdot (1 - \pi_0)$
- 3) Kirjaa testisuure ja sen otantajakauma nollahypoteesin vallitessa
- 4) Laske testisuureen arvo
- 5) Laske P -arvo
- 6) Tee johtopäätös

2. Palataan 1. harjoitusten 3. tehtävässä esitettyyn artikkeliin.

Ovatko kuolleisuudet antibioottia ja lumelääkettä saaneilla aisia sairastavilla sekä hiv-positiivisilla lapsilla yhtä suuret vai vähentääkö antibiootti kuolleisuutta?

Artikkelissa sanotaan tulosten olevan hälyttäviä. Tutkitaan tuloksia tarkemmin tilastollisella testillä. Tutkimuksessa oli sekä antibioottia saanut koeryhmä (merk. A) ja vertailuryhmä (merk. B), joista havaitut kuolleisuudet ovat $p_A = 0.28$ ja $p_B = 0.42$. Vastaavat otoskoot ovat $n_A = 267$ ja $n_B = 267$. Minkä nimistä tilastollista testiä käyttäisit tässä yhteydessä?

Testauksen vaiheet:

- 1) *Todetaan, että kyseessä ovat satunnaisotokset ja otoskoot ovat isot (sillä $np_A > 5, n(1 - p_A) > 5, np_B > 5, n(1 - p_B) > 5$).*
- 2) *Kirjaa sekä sanallisesti että symboleita käyttäen nollahypoteesi ja yksisuuntainen vastahypoteesi siten, että antibiootin odotetaan vähentävän kuolleisuutta.*
- 3) *Kirjaa testisuure ja sen otantajakauma nollahypoteesin vallitessa.*
- 4) *Laske testisuureen arvo (piirrä kuvaaja).*
- 5) *Laske testisuureen arvoa vastaava P -arvo.*
- 6) *Tee johtopäätös.*

3. Sähkölamppujen valmistaja on mainoksessaan väittänyt lamppujensa palavan keskimäärin 1 500 h (eli $\mu_0 = 1500$). Paloajan X oletetaan noudattavan normaalijakaumaa $N(\mu, \sigma^2)$. 25:n sähkölampun otoksen perusteella saatiin keskimääräiseksi paloajaksi $\bar{x} = 1470$ h ja paloajan keskihajonnaksi $s = 200$ h. Onko mainoksen tieto oikeutettu otoksen tietojen perusteella yhden otoksen keskiarvotestin perusteella?

Vinkki:

- 1) *Todetaan, että kyseessä on satunnaisotos. OK.*
- 2) *Kirjaa sekä sanallisesti että symboleita käyttäen nollahypoteesi ja kaksisuuntainen vastahypoteesi.*
- 3) *Kirjaa testisuure ja sen otantajakauma nollahypoteesin vallitessa.*
- 4) *Laske testisuureen arvo (piirrä kuvaaja).*
- 5) *Nyt P -arvoa pitää arvioida t -jakaumataulukon avulla. Vrt. luentomonisteen esim. 8.8.*
- 6) *Tee johtopäätös.*

4. Vertaillaan syksyn 2006 TILP150-kurssilaisiin liittyvän aineiston perusteella miesten ja naisten matematiikan arvosanojen jakaumia liitteessä esitettyjen tunnuslukujen avulla. Tarkastelemme muuttujia MATARVPAIN=matematiikan arvosana (asteikko 4-12.5) ja SP=sukupuoli (1=nainen, 2=mies).

- a) Poimi SPSS-tulostuksista jakaumien sijaintia kuvaavia tunnuslukuja.
- b) Poimi SPSS-tulostuksista jakaumien vaihtelua kuvaavia tunnuslukuja.
- c) Vertaile sanallisesti jakaumien sijaintia ja vaihtelua tunnuslukujen ja viikkisilaatikoiden avulla.

(Luvut 5.1, 5.2 ja 5.3)

5. Jatkoa edelliseen tehtävään. Tutkitaan, onko miesten ja naisten matematiikan arvosanojen odotusarvojen välillä eroa tilastollisella testillä. Tarkastelemme muuttujia MATARVPAIN=matematiikan arvosana (asteikko 4-12.5) ja SP=sukupuoli (1=nainen, 2=mies).

Liitteessä on esitetty SPSS-tulostukset. Mistä testistä on kyse?

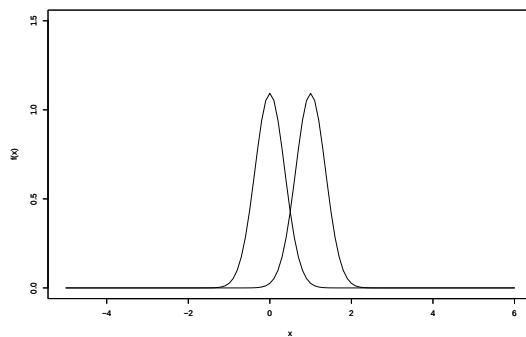
- a) Kirjaa nollahypoteesi ja kaksisuuntainen vastahypoteesi.
 - c) (Isoilla otoskoilla testi on robusti (ei herkkä) normaalisuusoletuksen suhteen. Iso otoskoko Moore & McCabe ohjeistuksen mukaan on $n_A + n_B \geq 40$, ok) Voidaanko varianssit olettaa yhtäsuuriksi Levenen testin mukaan? (Totea P -arvo).
 - d) Poimi SPSS-tulostuksista testisuureen arvo t_{hav} ja vastaava P -arvo (piirrä kuva)
 - e) Tee johtopäätös havaitun eron tilastollisesta merkitsevyydestä P -arvon perusteella.
6. "Teoriatehtävä": Nostaako kuuden kuukauden harjoitusohjelma nuorten naisten luun mineraalisisältöä? Joukko tutkijoita haluaa vastauksen kysymykseen. Tarkastellaan muuttujaa X =”mineraalisisällön muutos prosentteina”.

Aikaisempien tutkimusten perusteella olkoon ko. muuttujan keskihajonta $\sigma = 2$. Oletetaan lisäksi, että $X \sim N(\mu, 4)$, jolloin $\bar{X} \sim N(\mu, \frac{4}{n})$.

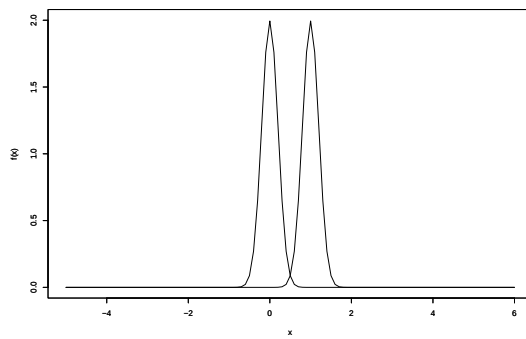
Seuraavassa tarkastellaan, millaisia arvoja keskiarvoa $\bar{X}$ saa, kun $\mu = 0$ (nollahypoteesi on totta) ja vastaavasti, millaisia arvoja keskiarvoa $\bar{X}$ saa, kun $\mu = 1$ (nollahypoteesi ei ole totta).

- a) Olkoon otoskoko $n = 30$. Millä välillä keskiarvoista on 95 %, kun $\mu = 0$. Millä välillä keskiarvoista on 95 %, kun $\mu = 1$. Käytä tarkempaa sääntöä. Ks. kuva 1.

- b) Olkoon otoskoko $n = 100$. Millä välillä keskiarvoista on 95 %, kun $\mu = 0$.
 Millä välillä keskiarvoista on 95 %, kun $\mu = 1$. Käytä tarkempaa sääntöä.
 Ks. kuva 2.



Kuva 1: Keskiarvon jakauma, kun otoskoko n on 30 ja $\mu = 0$ (vasen käyrä) ja kun $\mu = 1$ (oikea käyrä)



Kuva 2: Keskiarvon jakauma, kun otoskoko n on 100 ja $\mu = 0$ (vasen käyrä) ja kun $\mu = 1$ (oikea käyrä)

Demohyvyitys: 5 tehtävää=0.75 pistettä, 3 tehtävää=0.35 pistettä