

## Tilastomenetelmien peruskurssi (TILP150)

Syksy 2006

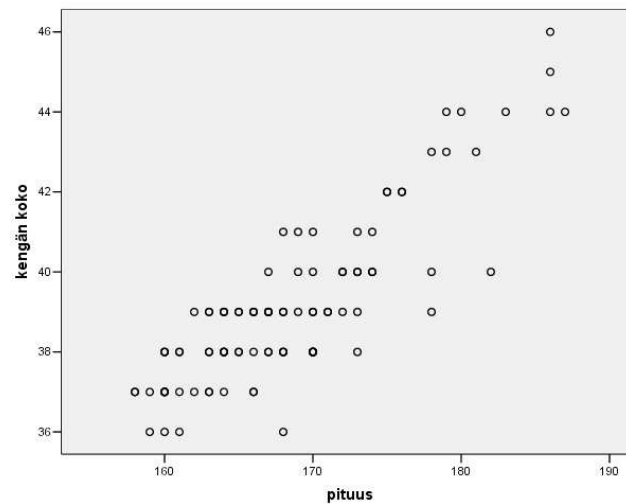
4. harjoitukset/viikko 40.

Näissä harjoituksissa käsittelemme Lukuja 6.2.2 (1.-2. tehtävät) ja 7 (3.-8. tehtävät).

1. 3. harjoituksissa 6. tehtävässä käsitelimme jo kevään 2005 TILA03 kurssilaisten kengännumeroja ja pituuksia. Jatketaan vielä regressiosuoran parissa.

Sirontakuvio on esitetty kuvassa 1. Aineistosta laskettu Pearsonin korrelaatiokerroin on 0.86. Aineistosta lasketut regressiosuoran parametrit ovat  $a = -5.4$  ja  $b = 0.264$ . Selitysaste on  $0.86^2 = 0.74$ .

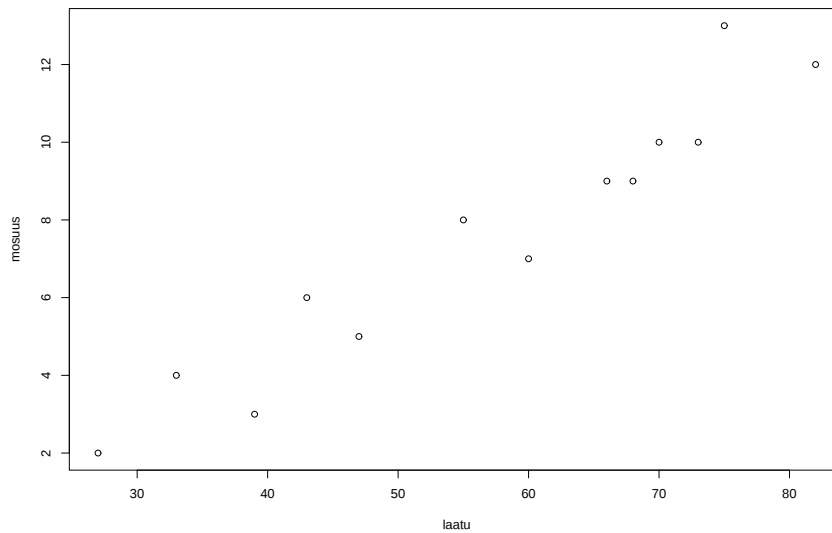
- a) Miten regressiosuoran regressioparametrit tulkitaan tässä tapauksessa?
- b) Mitä selitysaste kuvaa?
- c) Miten regressiokerroin eroaa korrelaatiokertoimesta?



Kuva 1: Kevään 2005 Tilastomenetelmien peruskurssilaisten kengännumeron ja pituuden sirontakuvio.

2. Halutaan tutkia tuotteen laadun vaikutusta valmistajan markkinaosuuteen. Olkoon muuttujat  $X$  = tuotteen laatu (objektiivisen arvioijan arvio asteikolla 0-100) ja  $Y$  = tuotteen markkinaosuus prosentteina (Lähde: Aczel & Sounderpandian, 2002). Seuraava aineisto on esitetty kuvassa 1.

|     |    |    |    |    |    |    |    |    |    |    |    |    |    |
|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| $X$ | 27 | 39 | 73 | 66 | 33 | 43 | 47 | 55 | 60 | 68 | 70 | 75 | 82 |
| $Y$ | 2  | 3  | 10 | 9  | 4  | 6  | 5  | 8  | 7  | 9  | 10 | 13 | 12 |



- a) Sovita aineistoon regressiosuora, ts. määrää regressiosuoran parametreit, regressiokerroin  $b$  ja vakiotermi  $a$ . Ks. alla aputuloksia. Miten ne tulkitaan tässä tapauksessa ?
- b) Mikä on regressiomallin selitysaste? Miten se tulkitaan? (Laske  $r_{xy}$  käyttäen alla olevia aputuloksia.)

Aputuloksia:  $\bar{x} = 56.8$  ja  $s_x = 17.5$  sekä  $\bar{y} = 7.5$  ja  $s_y = 3.4$  ja  $\sum_{i=1}^{13} (x_i - \bar{x})(y_i - \bar{y}) = 687.6$  ja  $\sum_{i=1}^{13} (x_i - \bar{x})^2 = 3684.3$

3. Tarkastellaan satunnaisilmiötä: 1 euron kolikon heitto kerran. Perusjoukko  $\mathcal{E} = \{kruuna, klaava\}$ . Ts. alkeistapaukset ovat kruuna ja klaava.
  - a) Mikä on kruunan saamisen todennäköisyys yhdessä heitossa klassisen todennäköisyyden mukaan?
  - b) Arvio ko. todennäköisyyttä empiirisesti: heitä 1 euron kolikkoa 10 kertaa. Laske kruunien (ei numeroa) määrä sekä suhteellinen frekvenssi.

4. Suuri osa sähköpostiviesteistä on ns. spam-viestejä. Lähteen Moore & McCabe (2006) mukaan niiden aihealueet ja jakauma on seuraava:

| <i>Aihe</i>    | <i>Aikuiset</i> | <i>Rahoitus</i> | <i>Terveys</i> | <i>Vapaa-<br/>aika</i> | <i>Tuotteet</i> | <i>Huijaukset</i> |
|----------------|-----------------|-----------------|----------------|------------------------|-----------------|-------------------|
| <i>Todenn.</i> | 0.145           | 0.162           | 0.073          | 0.078                  | 0.210           | 0.142             |

Valitaan spam-viesteistä satunnaisesti yksi.

- Mikä on todennäköisyys, että arvottu spam-viesti ei sisällä edellisiä aiheita (*vihje: komplementin todennäköisyys*)?
  - Corinne on kyllästynyt aikuisille kohdistettuihin viesteihin sekä huijausviesteihin. Mikä on todennäköisyys, että satunnaisesti valittu spam-viesti osuu jompaan kumpaan edelliseen kategoriaan (*vihje: yhdisteen todennäköisyys*)?
5. Veikkauksen Luontoarvan voitonjako on seuraava: 100 000 kpl 2 euroa, 75 000 kpl 4 euroa, 20 000 kpl 5 euroa, 10 000 kpl 10 euroa, 1800 kpl 15 euroa, 550 kpl 40 euroa, 180 kpl 100 euroa, 50 kpl 1000 euroa ja 5 kpl 15 000 euroa. Arvan painosmäärä on 1 000 000 kpl. Arvan hinta on 2 euroa.
- Merkitään  $X$ ="voiton määrä euroina". Määrää  $X$ :n mahdolliset arvot (ts. mitä rahamääriä voi voittaa) eli  $x_i$ :t (10 kappaletta). Esim.  $x_1 = 0$  jne.
  - Laske näiden arvojen todennäköisyydet eli  $p_i$ :t (klassinen todennäköisyys). Esim.  $p_1 = \mathbb{P}(X = x_1) = \mathbb{P}(X = 0) = \frac{792415}{1000000}$ .
  - Mikä on voiton odotusarvo  $\mu$ , ts. mikä on keskimääräinen voitto yhtä ostettua arpaa kohti (vastaus 0.90 e)? Odotusarvo lasketaan kaavalla  $\mu = \sum_{i=1}^{10} x_i \cdot p_i$ .
6. Tarkastellaan 20-vuotiaiden poikien pituutta. Oletetaan, että pituus  $X$  on normaalijakautunut eli  $X \sim N(\mu, \sigma^2)$ . Kasvukäyrää käyttäen odotusarvoa  $\mu$  arvioidaan keskiarvolla  $\bar{x} = 177$  ja keskihajontaa  $\sigma$  otoskeskihajonnalla  $s = 5.5$ , ts. oletetaan, että  $X \sim N(177, 5.5^2)$ . Käytä hyväksesi 68-95-99.7 sääntöä vastatessasi seuraaviin kysymyksiin (moniste jaettu luennolla, kopioita on MaD-rakennuksen 3. kerroksen aula-lassa olevalla pöydällä). Sääntö on myös nettisivulla <http://www.oswego.edu/srp/stats/6895997.htm>
- Millä välillä 20-vuotiaiden poikien pituuksista on 68 %?
  - Millä välillä 20-vuotiaiden poikien pituuksista on 95 %?

7. Harjoitellaan normaalijakaumataulukon käyttöä. Satunnaismuuttuja  $Z$  noudattaa normaalijakaumaa  $N(0, 1)$ . Mitä on todennäköisyys (piirrä kuvia)
- a)  $\mathbb{P}(Z \leq -1.60)$
  - b)  $\mathbb{P}(Z \leq 2.05)$
  - c)  $\mathbb{P}(Z > 1.78)$
  - d)  $\mathbb{P}(-2.3 < Z \leq 2.3)$
8. Jatkoa 6. tehtävälle. Laske seuraavat todennäköisyydet (olettaen, että  $X \sim N(177, 5.5^2)$ ).
- a) Mikä on 20-vuotiaista pojista sellaisten osuus, jotka ovat korkeintaan 171.5 cm pitkiä. Ts. laske todennäköisyys  $\mathbb{P}(X \leq 171.5)$ . Piirrä kuva.
  - b) Mikä on sellaisten 20-vuotiaista poikien osuus, joiden pituus on välillä 166 – 188. Ts. laske todennäköisyys  $\mathbb{P}(166 \leq X \leq 188)$ . (Jatkuvalla satunnaismuuttujalle tämä on sama kuin  $\mathbb{P}(166 < X \leq 188)$ . Piirrä kuva.
  - c) Mikä on sellaisten 20-vuotiaista poikien osuus, joiden pituus on 196 cm tai enemmän. Ts. laske todennäköisyys  $\mathbb{P}(X \geq 196)$ . Piirrä kuva.

Demohyvitys: 7 tehtävää= 0.75 pistettä, 4 tehtävää= 0.35 pistettä