

## Kvantitatiiviset menetelmät

**HUOM! Tentti pidetään tiistaina 22.2.  
klo 16-18 V20 ls.**

**Uusintamahdollisuus on rästitentissä  
25.5. ke 14–16 PR sali 1. Siihen tulee  
ilmoittautua WebOodissa 9.–18.5.  
välisenä aikana.**

© Mikko Mattila 2005

1

## Soveltuvan menetelmän valinta

| SELITTÄVÄ MUUTTUJA |                                  |                                                        |
|--------------------|----------------------------------|--------------------------------------------------------|
|                    | Laatuero- tai järjestysasteikko  | Välimatka- tai suhdelukuasteikko                       |
| SELITTÄVÄ MUUTTUJA | Laatuero- tai järjestysasteikko  | - Ristiintaulukointi<br>- Log-lineaariset mallit       |
|                    | Välimatka- tai suhdelukuasteikko | - Logistinen regressio<br>- Multinomiaalinen regressio |

© Mikko Mattila 2005

2

## Regressioanalyysi

- Regressioanalyysin avulla tutkitaan yhden tai useamman selittävän muuttujan vaikutusta selitettävään muuttujaan
- Esim. vaikuttaako koulutuksen pituus palkan suuruuteen?
- Regressioanalyysi on yleisimmin käytetty monimuuttujamenetelmä yhteiskuntatieteissä

© Mikko Mattila 2005

3

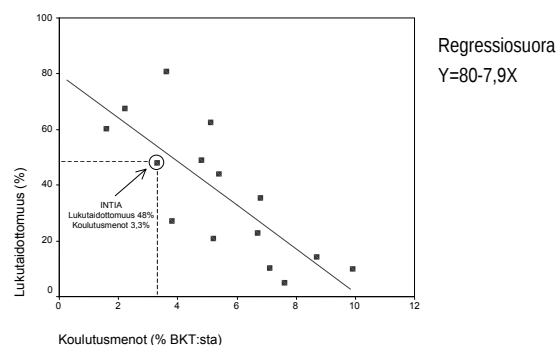
## Regressioanalyysi

- Etu: voidaan tutkia yhtä aikaa monen muuttujan vaikutusta selitettävään muuttujaan → yhteen regressiotaulukkoon mahtuu paljon informaatiota
- Selitettävän muuttujan oltava vähintään välimatka-asteikollinen
- Selittävät muuttujat vähintään välimatka-asteikollisia tai ns. dummy-muuttujia
- Kaikki luokittelu- tai järjestysasteikon muuttujat on mahdollista muuntaa dummy-muuttujiksi

© Mikko Mattila 2005

4

## Koulutusmenot & lukutaidottomuus



© Mikko Mattila 2005

5

## Regressiosuora

- Regressiosuora osoittaa muuttujien välisen riippuvuuden voimakkuuden
- Regressiosuora:  $Y = a + bX$
- Y selitettävä muuttuja
- X selittävä muuttuja
- b on regressiokerroin
- a on vakiotehtäjä
- Huom. regressiokerroin kertoo suoran kulmakertoimen, Pearsonin korrelaatiokerroin sen kuinka lähellä suoraa havainnot ovat
- Korrelaatiokerroin on symmetrinen, regressiokerroin ei

© Mikko Mattila 2005

6

## Regressiokerroin

- Regressiokerroin  $b$  kertoo kuinka paljon  $Y$  muuttuu, kun  $X$  muuttuu yhden yksikön
- Jos  $b < 0$ , yhteys negatiivinen ( $X$ :n kasvaessa  $Y$  pienenee)
- Jos  $b > 0$  yhteys positiivinen ( $X$ :n kasvaessa  $Y$  suurenee)
- Jos  $b = 0$ , ei lineaarista yhteyttä muuttujien välillä

© Mikko Mattila 2005

7

## Regressiomallin ennustekyky

- Regressiomallin ennustekyky riippuu siitä, kuinka lähellä havainnot ovat regressiosuoraa
- Jos ne ovat lähellä → hyvä ennustekyky
- Jos ne ovat kaukana → huono ennustekyky
- Virhetermi (residuaali) on havainnon arvon erotus regressiosuorasta (eli mallin ennustearvosta)

© Mikko Mattila 2005

8

## PNS-menetelmä

- PNS=pienimmän neliösumman menetelmä (engl. ordinary least squares eli OLS)
- regressiosuoran kulmakertoimen ja vakion laskeminen perustuu PNS-menetelmään
- minimoidaan havaintojen ja regressiosuoran etäisyyksien (residuaalien) neliöt

© Mikko Mattila 2005

9

## Mallin hyvyyden arviointi

- $R^2$ -luku kertoo selitysosuuden
  - kuinka suuri osuus  $Y$ :n vaihtelusta voidaan selittää  $X$ :n vaihtelulla?
  - vaihtelee nolla ja yhden välillä
- F-testi kertoo pystyvätkö  $X$ -muuttujat ylipäänsä selittämään  $Y$ :n vaihtelua
- Estimaatin keskivirhe (SEE) kertoo virhetermien keskihajonnan
- t-testi kertoo millä todennäköisyydellä regressiokerroin eroaa nolasta perusjoukossa

© Mikko Mattila 2005

10

## HIV & eliniän odote

- Esimerkki: selittääkö HIVin levinneisyys odotettavissa olevaa elinikää?

|                                | Regressiokerroin | t-arvo | Merkitsevyys |
|--------------------------------|------------------|--------|--------------|
| Vakio                          | 68,4**           | 91,5   | $p < 0,001$  |
| HIV tapaukset (/1000 henkilöä) | -0,27**          | -11,3  | $p < 0,001$  |
| $R^2$                          | 0,44             |        |              |
| Korjattu $R^2$                 | 0,44             |        |              |
| F-testi                        | 128,0**          |        | $p < 0,001$  |
| Estimaatin keskivirhe          | 8,7              |        |              |

© Mikko Mattila 2005

11

## HIV & eliniän odote

- Regressiokaava voidaan käyttää "ennusteiden" tekoon
- Kaava:  $Y = 68,4 - 0,27X$
- Kun HIV määrä kasvaa yhdellä (/1000 henkeä), lyhenee eliniän odote 0,27 vuotta
- Jos Suomessa (0,21) HIV-levinneisyys sama kuin Ranskassa (2,21) – elinikä lyhenisi 0,54 vuotta
  - $(2,21 - 0,21) * 0,27 = 0,54$
- Jos sama kuin Tansaniassa (39,6) –elinikä lyhenisi 10,6 vuotta
  - $(39,6 - 0,21) * 0,27 = 10,6$

© Mikko Mattila 2005

12

## Monimuuttujaregressio

- Kaavana:  $Y = a + b_1X_1 + b_2X_2$
- selittävien muuttujien määrällä ei ole varsinaisesti ylärajaa (käytännössä tutkimusongelma ja havaintojen määrä rajoittavat määrää)
- Regressiokertoimet ilmaisevat kuinka paljon Y muuttuu kun X muuttuu yhden yksikön ja kaikki muut selittävät muuttujat pysyvät vakiona.

© Mikko Mattila 2005

13

## HIV + BKT & eliniän odote

|                                | Regressiokerroin | t-arvo | Merkitsevyys |
|--------------------------------|------------------|--------|--------------|
| Vakio                          | 64,4**           | 87,0   | p<0,001      |
| HIV tapaukset (/1000 henkilöä) | -0,23**          | -11,6  | p<0,001      |
| BKT /henkilö                   | 0,57**           | 9,44   | p<0,001      |
| R <sup>2</sup>                 | 0,64             |        |              |
| Korjattu R <sup>2</sup>        | 0,63             |        |              |
| F-testi                        | 143,2**          |        | p<0,001      |
| Estimaatin keskivirhe          | 7,04             |        |              |

© Mikko Mattila 2005

14

## Dummy-muuttujat

- Luokittelu- tai järjestysasteikon muuttuja voidaan sisällyttää analyysiin tekemällä niistä dummy-muuttujia
- Dummy-muuttuja saa vain kaksi arvoa: 0 tai 1
- Regressiokerroin ilmaisee, kuinka paljon tutkittu ryhmä (dummy=1) eroaa muista (dummy=0)
- Jos luokittelumuuttujassa n vaihtoehtoa, täytyy luoda n-1 dummy-muuttujaa
- Esim: koulutus mitattu kolmella vaihtoehdolla (1=peruskoulu, 2=keskiaste, 3=korkeakoulu)
  - luodaan kaksi dummy-muuttujaa
  - peruskoulu 0/1, keskiaste 0/1 (vertailukohtana korkeakoulu)
  - vertailukohdan valinnalla ei väliä

© Mikko Mattila 2005

15

## HIV + BKT + Afrikka & eliniän odote

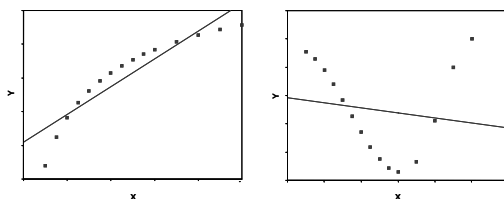
|                                | Regressiokerroin | t-arvo | Merkitsevyys |
|--------------------------------|------------------|--------|--------------|
| Vakio                          | 67,3**           | 98,8   | p<0,001      |
| HIV tapaukset (/1000 henkilöä) | -0,14**          | -7,1   | p<0,001      |
| BKT /henkilö                   | 0,44**           | 8,4    | p<0,001      |
| Afrikka-dummy                  | -11,02**         | -8,76  | p<0,001      |
| R <sup>2</sup>                 | 0,76             |        |              |
| Korjattu R <sup>2</sup>        | 0,75             |        |              |
| F-testi                        | 165,7**          |        | p<0,001      |
| Estimaatin keskivirhe          | 5,81             |        |              |

© Mikko Mattila 2005

16

## Regressioanalyysin rajoitteet

- Lineaarisuusoletus
- Voidaan usein korjata muuttujien muunnoksilla

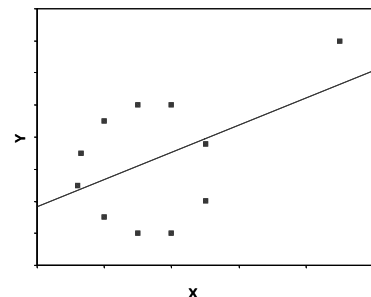


© Mikko Mattila 2005

17

## Regressioanalyysin rajoitteet

- Poikkeavat havainnot eli outlierit voivat vääristää tuloksia



© Mikko Mattila 2005

18

## Regressioesimerkki

- Tutkitaan regressioanalyysin avulla, miten taustatekijät liittyvät poliittisen osallistumisen aktiivisuuteen
- Selittävät muuttujat:
  - ikä
  - sukupuoli
  - koulutus
  - kiinnostus politiikasta

© Mikko Mattila 2005

19

## Regressioesimerkki

- Ensin täytyy tehdä hieman valmisteluja:
  - aineiston sukupuolimuuttuja q1 on koodattu 1=mies, 2=nainen, joten se muutetaan dummy-muuttujaksi uudelleenkoodaamalla nainen=1, mies=0
- Samoin koulutus on mitattu useilla vaihtoehdoilla
  - luodaan kaksi uutta muuttujaa, yliopisto ja keskiaste
- Ikä- ja kiinnostusmuuttujat voidaan lisätä malliin sellaisenaan

© Mikko Mattila 2005

20

## Regressioesimerkki

3. Seuraavassa on jonkko erilaista yhteiskunnallisen osallistumisen muotoja. Jos haluaisitte antaa edustavaa todella tarkalleen, niin mikä näistä olette tehnyt tai saattaisitte tehdä?

|                                                                                                      | Olette tehnyt              | Ette ole tehnyt, mutta saattaisitte tehdä | Ette tekisi missään tapauksessakaan | En osaa sanoa              |
|------------------------------------------------------------------------------------------------------|----------------------------|-------------------------------------------|-------------------------------------|----------------------------|
| [p3_1] Kirjoittaa yleisönoasteon .....                                                               | <input type="checkbox"/> 1 | <input type="checkbox"/> 2                | <input type="checkbox"/> 3          | <input type="checkbox"/> 4 |
| [p3_2] Oittaa yhteyttä poliittisiin päätöksentekijöihin jossakin tärkeässä asiassa .....             | <input type="checkbox"/> 1 | <input type="checkbox"/> 2                | <input type="checkbox"/> 3          | <input type="checkbox"/> 4 |
| [p3_3] Kirjoittaa nimenne vetoamuksiin .....                                                         | <input type="checkbox"/> 1 | <input type="checkbox"/> 2                | <input type="checkbox"/> 3          | <input type="checkbox"/> 4 |
| [p3_4] Osallistuu kokouksiin, maksu- tai ostotilaisuuksiin .....                                     | <input type="checkbox"/> 1 | <input type="checkbox"/> 2                | <input type="checkbox"/> 3          | <input type="checkbox"/> 4 |
| [p3_5] Osallistuu rauhanomaisiin mielenosoituksiin .....                                             | <input type="checkbox"/> 1 | <input type="checkbox"/> 2                | <input type="checkbox"/> 3          | <input type="checkbox"/> 4 |
| [p3_6] Osallistuu kansainvälisiin mielenosoituksiin, joissa ajemmin on väkivaltainen taustansa ..... | <input type="checkbox"/> 1 | <input type="checkbox"/> 2                | <input type="checkbox"/> 3          | <input type="checkbox"/> 4 |
| [p3_7] Osallistuu sellaisiin mielenosoituksiin, joissa ajemmin on ilmennyt väkivaltaisuutta .....    | <input type="checkbox"/> 1 | <input type="checkbox"/> 2                | <input type="checkbox"/> 3          | <input type="checkbox"/> 4 |
| [p3_8] Käyttää väkivaltaa poliittisten päämäärien edistämiseksi .....                                | <input type="checkbox"/> 1 | <input type="checkbox"/> 2                | <input type="checkbox"/> 3          | <input type="checkbox"/> 4 |

© Mikko Mattila 2005

21

## Regressioesimerkki

Ensivaiheessa

tarkastellaan

ainoastaan

ikämuuttujan

vaikutusta

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1     | ,305 <sup>a</sup> | ,093     | ,091              | 2,50705                    |

a. Predictors: (Constant), [q2] Ikä

| Model |            | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|--------|------|
|       |            | B                           | Std. Error | Beta                      |        |      |
| 1     | (Constant) | 16,551                      | ,402       |                           | 41,207 | ,000 |
|       | [q2] Ikä   | 5,091E-02                   | ,008       | ,305                      | 6,292  | ,000 |

a. Dependent Variable: OSALL

© Mikko Mattila 2005

22

## Regressioesimerkki

Lisätään sukupuolta kuvaava muuttuja malliin

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1     | ,311 <sup>a</sup> | ,096     | ,092              | 2,50560                    |

a. Predictors: (Constant), NAINEN, [q2] Ikä

| Model |            | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|--------|------|
|       |            | B                           | Std. Error | Beta                      |        |      |
| 1     | (Constant) | 16,734                      | ,429       |                           | 39,002 | ,000 |
|       | [q2] Ikä   | 4,999E-02                   | ,008       | ,299                      | 6,154  | ,000 |
|       | NAINEN     | -,309                       | ,257       | -,059                     | -1,203 | ,230 |

a. Dependent Variable: OSALL

© Mikko Mattila 2005

23

## Regressioesimerkki

Lisätään koulutus

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1     | ,349 <sup>a</sup> | ,122     | ,112              | 2,47694                    |

a. Predictors: (Constant), KESKIAST, NAINEN, [q2] Ikä, YLIOPIST

| Model |            | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|--------|------|
|       |            | B                           | Std. Error | Beta                      |        |      |
| 1     | (Constant) | 17,109                      | ,464       |                           | 36,889 | ,000 |
|       | [q2] Ikä   | 4,900E-02                   | ,008       | ,294                      | 5,969  | ,000 |
|       | NAINEN     | -,298                       | ,255       | -,057                     | -1,172 | ,242 |
|       | YLIOPIST   | -1,242                      | ,384       | -,162                     | -3,236 | ,001 |
|       | KESKIAST   | -,453                       | ,280       | -,083                     | -1,614 | ,107 |

a. Dependent Variable: OSALL

© Mikko Mattila 2005

24

## Regressioesimerkki

Model Summary

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1     | .393 <sup>a</sup> | .154     | .143              | 2.42507                    |

a. Predictors: (Constant), [q22] Kuinka kiinnostunut olette politiikasta, NAINEN, [q2] Ika, YLIOPIST, KESKIAST

Coefficients<sup>a</sup>

| Model |                                               | Unstandardized Coefficients |            | Standardized Coefficients |  | t      | Sig. |
|-------|-----------------------------------------------|-----------------------------|------------|---------------------------|--|--------|------|
|       |                                               | B                           | Std. Error | Beta                      |  |        |      |
| 1     | (Constant)                                    | 15,599                      | ,597       |                           |  | 26,124 | ,000 |
|       | [q2] Ika                                      | 4,990E-02                   | ,008       | ,300                      |  | 6,194  | ,000 |
|       | NAINEN                                        | -,343                       | ,250       | -,065                     |  | -1,375 | ,170 |
|       | YLIOPIST                                      | -,894                       | ,385       | -,117                     |  | -2,320 | ,021 |
|       | KESKIAST                                      | -,251                       | ,279       | -,046                     |  | -,901  | ,368 |
|       | [q22] Kuinka kiinnostunut olette politiikasta | ,596                        | ,151       | ,192                      |  | 3,958  | ,000 |

a. Dependent Variable: OSALL

Ja lopuksi kiinnostus politiikasta

© Mikko Mattila 2005

25

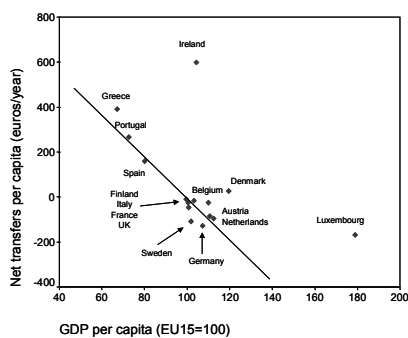
## Esimerkki: EU-rahavirrat

- Mitkä tekijät selittävät EU-budjetin nettorahavirtoja?
- Kuinka paljon EU:n laajentuminen maksaisi, jos uudet jäsenvaltiot saisivat samat edut kuin vanhat?
- Selitettävä muuttuja: EU:n vuosittaiset nettokustannukset henkeä kohden 15 jäsenvaltiossa (mitattu euroissa, vuodet 1995-2000)

© Mikko Mattila 2005

26

## Esimerkki: EU-rahavirrat



Regressioyhtälö:  
NET=896.7-9.06GDP

© Mikko Mattila 2005

27

## Esimerkki: EU-rahavirrat

| Country     | GDP per capita (EU15=100) | Net transfers (euros/capita) | "Fair" net transfers (euros/capita) |
|-------------|---------------------------|------------------------------|-------------------------------------|
| Greece      | 67                        | 391                          | 288                                 |
| Portugal    | 73                        | 266                          | 238                                 |
| Spain       | 80                        | 160                          | 170                                 |
| Finland     | 100                       | -10                          | -7                                  |
| France      | 101                       | -24                          | -15                                 |
| UK          | 101                       | -46                          | -17                                 |
| Sweden      | 102                       | -108                         | -27                                 |
| Italy       | 103                       | -16                          | -38                                 |
| Ireland     | 104                       | 598                          | -49                                 |
| Germany     | 107                       | -128                         | -75                                 |
| Belgium     | 110                       | -24                          | -100                                |
| Austria     | 111                       | -85                          | -107                                |
| Netherlands | 113                       | -95                          | -124                                |
| Denmark     | 120                       | 27                           | -187                                |
| Luxembourg  | 179                       | -169                         | -725                                |

© Mikko Mattila 2005

28

## Esimerkki: EU-rahavirrat

| Country   | GDP per capita (EU15=100) | "Fair" net transfers (euros/capita) | Total costs (billion euros) |
|-----------|---------------------------|-------------------------------------|-----------------------------|
| Romania   | 24                        | 679                                 | 15.24                       |
| Bulgaria  | 27                        | 652                                 | 5.33                        |
| Lithuania | 29                        | 634                                 | 2.34                        |
| Latvia    | 33                        | 598                                 | 1.44                        |
| Estonia   | 37                        | 561                                 | 0.81                        |
| Poland    | 39                        | 543                                 | 21.00                       |
| Slovakia  | 48                        | 462                                 | 2.49                        |
| Hungary   | 50                        | 444                                 | 4.45                        |
| Malta     | 55                        | 398                                 | 0.16                        |
| Czech     | 59                        | 362                                 | 3.72                        |
| Slovenia  | 69                        | 271                                 | 0.54                        |
| Cyprus    | 85                        | 126                                 | 0.08                        |
| Total     |                           |                                     | 57.60                       |

© Mikko Mattila 2005

29