**YFIS200 HARJOITUS 2**

**Jokivuori & Virmasalo; viikko 8**

**Logistinen regressioanalyysi (LRA)**

Logistinen regressioanalyysi on tavanomaisen regressioanalyysin erityistyyppi. Sitä käytetään silloin, kun selitettävä muuttuja (Y) voi saada vain **kaksi arvoa**. Logistinen regressioanalyysi ei pyri ennustamaan määriä, vaan ”todennäköisyyksiä”. Kyse on siis siitä, millä ”todennäköisyydellä” tarkasteltavana oleva asia tapahtuu tai pätee. Tulokset kertovat, vaikuttavatko selittävät muuttujat ”tapahtuman” ”todennäköisyyteen” ja kuinka suuri vaikutus on.

**Nouda aineisto ESSSuomi2008.sav omasta kansiostasi!**

| **[f27] Oletteko koskaan ollut työttömänä työnhakijana yhtäjaksoisesti yli kolmen kuukauden pituisen jakson?** |
| --- |
|  | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Kyllä | 654 | 29,8 | 29,8 | 29,8 |
| Ei | 1541 | 70,2 | 70,2 | 100,0 |
| Total | 2195 | 100,0 | 100,0 |  |

Y 🡪 0 = ei ole ollut työttömänä; 1 = on ollut työttömänä

| **Työttömänä** |
| --- |
|  | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | ,00 | 1541 | 70,2 | 70,2 | 70,2 |
| 1,00 | 654 | 29,8 | 29,8 | 100,0 |
| Total | 2195 | 100,0 | 100,0 |  |

Selittjät (x:t)

**Sukupuoli** 1 =mies; 2 = nainen

**Sukupolvet**

**RECODE f3\_1 (Lowest thru 1949=1) (1950 thru 1959=2) (1960 thru 1969=3) (1970 thru 1979=4) (1980 thru Highest=5) (ELSE=SYSMIS) INTO Sukupolvet.**

**EXECUTE.**

**F5.** Mikä luonnehdinnoista parhaiten kuvaa asuinympäristöänne?

ALLE 20 000 ASUKKAAN KAUPUNGIT KUULUVAT LUOKKAAN 4.

**Suuri kaupunki (yli 100 000 asukasta) 1**

**Suuren kaupungin lähiö tai lähiseutu (suuren kaupungin**

**vaikutusalueella / työssäkäyntialueella) 2**

**Pieni tai keskikokoinen kaupunki tai kunta (20 000-100 000 as.) 3**

**Pienempi taajama tai kunta (alle 20 000 asukasta) 4**

**Maaseutu (haja-asutusalue) 5**

**F6.** Mikä on koulutuksenne. Valitkaa kortista 67 korkein koulutusaste,

jonka olette suorittanut?

Vähemmän kuin peruskoulun ala-aste tai

vähemmän kuin kansakoulu 0

Peruskoulun ala-aste (1-6 luokat), kansakoulu 1

Peruskoulun yläaste (7-9/10 luokat), keskikoulu 2

Lukio, ylioppilas- tai ammatillinen tutkinto 3

Opisto- tai korkeakoulututkinto 5

Lisensiaatin tai tohtorin tutkinto 6

**RECODE f6 (2=2) (3=3) (4=4) (5=5) (6=6) (0 thru 1=1) (ELSE=SYSMIS) INTO f6recoded.**

**EXECUTE.**

**F23.** Minkätyyppisen työnantajan palveluksessa olette tai olitte?

Valitkaa yksi vaihtoehto.

Keskus- tai paikallishallinto 1

Muu julkinen sektori

(esim. koulutuspalvelut tai terveyspalvelut) 2

Julkisen sektorin omistama yritys 3

Yksityinen yritys 4

Yksityisyrittäjä 5

Muu 6

**RECODE f23 (5=3) (1 thru 3=1) (4=2) (ELSE=SYSMIS) INTO f23Recoded.**

**EXECUTE.**

**Työnanatajasektori: 1 = julkinen, 2= yksityinen ja 3 = yksityisyrittäjä**

**Analyze 🡪 Regression …Binary Logistic**



**Kohtaan Dependent valitaan Työttömänä-muuttuja (0,1)**. Kohtaan **Covariates** viedään edellä luetellut selittäjät.

Kun selittäjät ovat listassa, klikataan **Categorial** ja sitten maalataan kaikki muuttujat ja viedään ne kohtaan **Categorial Covariates** ja sitten valitaan referenssiryhmäksi (vertailutyhmäksi muuttujan ensimmäinen luokka) ensimmäinen eli **first**. Painetaan **CHANGE** ja sitten **Continue.**

LOGISTIC REGRESSION VARIABLES Työttömänä

 /METHOD=ENTER Sukupuoli Sukupolvet f5 f6recoded f23Recoded

 /CONTRAST (Sukupuoli)=Indicator(1)

 /CONTRAST (Sukupolvet)=Indicator(1)

 /CONTRAST (f5)=Indicator(1)

 /CONTRAST (f23Recoded)=Indicator(1)

 /CONTRAST (f6recoded)=Indicator(1)

 /PRINT=GOODFIT

 /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Muuttujien perässä on nyt (Cat). **Klikataan Options**:



TULOSTUKSESTA:

| **Case Processing Summary** |
| --- |
| Unweighted Casesa | N | Percent |
| Selected Cases | Included in Analysis | 2009 | 91,5 |
| Missing Cases | 186 | 8,5 |
| Total | 2195 | 100,0 |
| Unselected Cases | 0 | ,0 |
| Total | 2195 | 100,0 |
| a. If weight is in effect, see classification table for the total number of cases. |

| **Classification Tablea,b** |
| --- |
|  | Observed | Predicted |
|  | Työttömänä | Percentage Correct |
|  | ,00 | 1,00 |
| Step 0 | Työttömänä | ,00 | 1401 | 0 | 100,0 |
| 1,00 | 608 | 0 | ,0 |
| Overall Percentage |  |  | 69,7 |
| a. Constant is included in the model. |
| b. The cut value is ,500 |

| **Model Summary** |
| --- |
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 2324,896a | ,067 | ,094 |
| a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001. |

| **Hosmer and Lemeshow Test** |
| --- |
| Step | Chi-square | df | Sig. |
| 1 | 7,745 | 8 | ,459 |

| **Classification Tablea** |
| --- |
|  | Observed | Predicted |
|  | Työttömänä | Percentage Correct |
|  | ,00 | 1,00 |
| Step 1 | Työttömänä | ,00 | 1328 | 73 | 94,8 |
| 1,00 | 502 | 106 | 17,4 |
| Overall Percentage |  |  | 71,4 |
| a. The cut value is ,500 |

Taulusta **Variables in the Equation** lähdetään tulkitsemaan muuttujien yhteyttä työttömyyden kokemiseen. Idea on se, että jokaisen muuttujan ensimmäinen luokka on referenssiryhmä, joihin muiden luokkien tunnuslukua (kohdassa **Exp(B))** verrataan (ensimmäisen luokan arvo on aina tasan 1, joten sitä ei ole merkitty taulukkoon).

Esim. Sukupuoli(1) Exp(B) on naisten saama riskiluku (0,883 ja p=,232), sillä miehet ovat tässä muuttujassa referenssiryhmä. Sukupuoli ei siis ennusta/selitä työttömyyden kokemista, mutta entä ikäryhmä (sukupolvet) ja koulutustaso (f6recoded) …

| **Variables in the Equation** |
| --- |
|  | B | S.E. | Wald | df | Sig. | Exp(B) |
| Step 1a | Sukupuoli(1) | -,124 | ,104 | 1,426 | 1 | ,232 | ,883 |
| Sukupolvet |  |  | 90,117 | 4 | ,000 |  |
| Sukupolvet(1) | ,788 | ,156 | 25,645 | 1 | ,000 | 2,200 |
| Sukupolvet(2) | 1,022 | ,166 | 37,935 | 1 | ,000 | 2,779 |
| Sukupolvet(3) | 1,247 | ,177 | 49,905 | 1 | ,000 | 3,481 |
| Sukupolvet(4) | -,096 | ,179 | ,289 | 1 | ,591 | ,908 |
| f5 |  |  | 5,391 | 4 | ,249 |  |
| f5(1) | ,244 | ,197 | 1,540 | 1 | ,215 | 1,276 |
| f5(2) | ,287 | ,149 | 3,690 | 1 | ,055 | 1,332 |
| f5(3) | ,278 | ,162 | 2,961 | 1 | ,085 | 1,321 |
| f5(4) | ,077 | ,173 | ,200 | 1 | ,655 | 1,080 |
| f6recoded |  |  | 38,707 | 4 | ,000 |  |
| f6recoded(1) | ,024 | ,198 | ,014 | 1 | ,905 | 1,024 |
| f6recoded(2) | -,022 | ,172 | ,016 | 1 | ,899 | ,978 |
| f6recoded(3) | -,737 | ,191 | 14,900 | 1 | ,000 | ,479 |
| f6recoded(4) | -1,182 | ,455 | 6,760 | 1 | ,009 | ,307 |
| f23Recoded |  |  | 9,868 | 2 | ,007 |  |
| f23Recoded(1) | ,075 | ,116 | ,424 | 1 | ,515 | 1,078 |
| f23Recoded(2) | -,454 | ,178 | 6,518 | 1 | ,011 | ,635 |
| Constant | -1,233 | ,205 | 36,256 | 1 | ,000 | ,292 |
| a. Variable(s) entered on step 1: Sukupuoli, Sukupolvet, f5, f6recoded, f23Recoded. |