

### Logistinen regressio esimerkki:

Kaavakokoelmassa yksi kaava, käytännössä neljä eri vaihtoehtoa. Esim. Frekvenssit äänestyskystä, jossa äänestettiin työpaikan taukotilan maalaamisesta vihreäksi.

	Mies	Nainen
Kyllä	12	15
Ei	8	20

Lasketaan aluksi oddsit, että osataan tulkita tuloksia (tätä ei siis tarvitse tehdä enää kun menetelmä on hallussa!):

Miehen todennäköisyys äänestää kyllä:  $12/(12+8) = 0.6$

Naisten todennäköisyys äänestää kyllä:  $15/(15+20) = 0.428$

Odds(Mies-Kyllä):  $(12/20)/(8/20) = 12/8 = 1.5$

Odds(Mies-Ei):  $8/12 \approx 0.66$

Odds(Nainen-Kyllä):  $15/20 = 0.75$

Odds(Nainen-Ei):  $20/15 \approx 1.33$

OR (mies nainen – Kyllä ):  $1.5/0.75 = 2$

OR (mies nainen- Ei):  $0.50$

Tällaisen raportointi olisi hyvin raskasta ja ikävää. Siksi koko informaatio tiivistetään yhteen yhtälöön.

	Mies	Nainen
Kyllä	12	15
Ei	8	20

Esim. referenssiluokka "Mies" ja mallinnetaan Kyllä-vastauksen todennäköisyyttä:

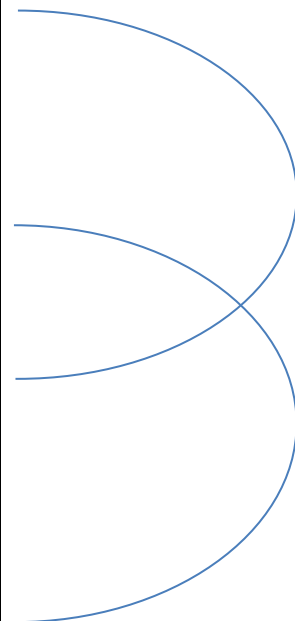
$$b_0 = \ln\left(\frac{12}{8}\right) \approx 0.405$$

$$b_1 = \ln\left(\frac{15/20}{12/8}\right) \approx -0.693$$

$OR = e^{-0.693} \approx 0.50 \rightarrow$  "Kyllä"- vastauksen riski puolet naisilla miehiin verrattuna.

$$\text{Mallin kaava: } \ln\left(\frac{p_1}{1-p_1}\right) = 0.405 - 0.693 * x$$

"Kyllä"-vastauksen riski referenssiluokka "Mies"	
bo	0,405
b1	-0,693
OR	0,5 "Kyllä"- vastauksen riski puolet naisilla miehiin verrattuna.
"Kyllä"-vastauksen riski referenssiluokka "Nainen"	
bo	-0,288
b1	0,693
OR	2 "Kyllä"-vastauksen riski kaksinkertainen miehillä naisiin verrattuna.
"Ei"-vastauksen riski referenssiluokka "Mies"	
bo	-0,405
b1	0,693
OR	2 "Ei"-vastauksen riski kaksinkertainen naisilla miehiin verrattuna.
"Ei"-vastauksen riski referenssiluokka "Nainen"	
bo	0,288
b1	-0,693
OR	0,5 "Ei"-vastauksen riski puolet miehillä naisiin verrattuna.



Symmetria!

Lasketaan yhtälön avulla alkuperäiset todennäköisyydet:

$$p = \frac{1}{1 + e^{-(0.405 - 0.693 * x)}}$$

Nyt täytyy muistaa, että valitsemme referenssiluokaksi miehen eli  $x=1 \Rightarrow$  nainen.

$$p_1 = \frac{1}{1+e^{-(0.405-0.693*1)}} = \frac{1}{1.75} = 0.428$$

Kun  $x=0$  kyseessä on referenssiryhmä eli miehet:

$$p_0 = \ln\left(\frac{1}{1+e^{-(0.405)}}\right) = \frac{1}{1.572} = 0.6$$

### Harjoitellaan:

Laske seuraavalle aineistolle logistisen regression yhtälö. Käytä referenssi ryhmänä Helsinkiläisiä ja mallinna vegaani-ruokavalion todennäköisyyttä.

Määritä logistisen regressioyhtälön avulla Helsinkiläisen todennäköisyys olla vegaani ja Porilaisen todennäköisyys olla ei-vegaani.

Määritä myös helsinkiläisten ja porilaisten välinen OR.

	Helsinki	Pori
Vegaani	19	15
Ei-vegaani	120	200

Vastaus:

	Helsinki	Pori
Vegaani	19	15
Ei-vegaani	120	200

Esim. referenssiluokka "Helsinki" ja mallinnetaan Vegaani-vastauksen todennäköisyyttä:

$$b_0 = \ln\left(\frac{19}{120}\right) \approx -1.84$$

$$b_1 = \ln\left(\frac{15/200}{19/120}\right) \approx -0.747$$

$$\text{Mallin kaava: } \ln\left(\frac{p_1}{1-p_1}\right) = -1.84 - 0.747 * x$$

$$p_{(\text{helsingiläinen on vegaani})} = \frac{1}{1+e^{-(-1.84-0.747*0)}} = 0.137$$

$$p_{(\text{Porilainen on vegaani})} = \frac{1}{1+e^{-(-1.84-0.747*1)}} = 0.070$$

$$p_{(\text{Porilainen on Ei-vegaani})} = 1 - 0.07 = 0.93$$

$$OR = e^{-0.747} \approx 0.47 \rightarrow \text{"vegaani"-elämäntavan riski}$$

0.47-kertainen Helsingissä asuvaan verrattuna. Tai toisinpäin  $1/0.47 = 2.1$  eli Helsingiläisellä on 2.1-kertainen riski olla vegaani porilaiseen verrattuna.

Tarkistetaan:

	Helsinki	Pori
Vegaani	19	15
Ei-vegaani	120	200

Helsingissä vegaanisisuuden todennäköisyys on  $19/139 = 13.7\%$

Porissa Vegaanisisuuden todennäköisyys on  $15/215 = 0.070\%$

Harjoitus 2)

	Rokotettu	Ei-rokotettu
Vesirokko	5	200
Ei-vesirokkoa	230	120

Esim. referenssiluokka "vesirokko" ja mallinnetaan rokotettu-vastauksen todennäköisyyttä:

$$b_0 = \ln\left(\frac{5}{230}\right) \approx -3.83$$

$$b_1 = \ln\left(\frac{200/120}{5/230}\right) \approx 4.34$$

$$\text{Mallin kaava: } \ln\left(\frac{p_1}{1-p_1}\right) = -3.83 + 4.34 * x$$

$$p(\text{rokotettu saa vesirokon}) = \frac{1}{1+e^{-(-3.83+4.34*0)}} = 0.021$$

$$p(\text{rokottamaton saa vesirokon}) = \frac{1}{1+e^{-(-3.83+4.34*1)}} = 0.625$$

$$OR = e^{4.34} \approx 76.7$$

-> Vesirokon riski on 76.7-kertainen jos lasta ei ole rokotettu.