Logistic regression

Statistics II (LIX002X05)

University of Groningen, Faculty of Arts, Information Science
Wilbert Heeringa

Introduction

- **Linear regression**:
  investigate the relationship between a **numerical** response variable and one or more explanatory variables.
- **Logistic regression**:
  investigate the relationship between a **categorical** response variable and one or more explanatory variables.

- Types of logistic regression: **binomial** or **dichotomous** (two possible outcomes) and **multinomial** or **polytomous** (three or more outcomes).

Outline

- We will look at three examples:
  - Example 1:
    Binomial logistic regression with one categorical predictor
  - Example 2:
    Binomial logistic regression with two numerical predictors
  - Example 3:
    Multinomial regression

Example 1

- In New York City the /r/ at the end of a syllable is usually pronounced as [ə] (the schwa).
- In the 50's and 60's of the previous century the pronuncation changed, the [r] was more and more pronounced.
- William Labov investigated whether this change had a sociological basis.
- He visited three department stores: Saks (upper class), Macy's (middle class) and S. Klein (lower class).
- He asked a shop assistent where he could find a particular article. He knew already that it was sold on the fourth floor. Therefore the answer was: "on the *fou**r**th floo**r***".
- He responded like he did not understand the answer, therefore the shop assistent repeated the answer.

Data

- Data for New York City:

	pronunciation		
status	[r]	[ə]	both
upper	30	6	32
middle	20	74	31
lower	4	50	17

Simplification

- We leave out the column both. Now the response variable is dichotomous. Coding in SPSS: 0=[r] en 1=[ə].

	pronunciation		
status	[r]	[ə]	both
upper	30	6	32
middle	20	74	31
lower	4	50	17

Table in SPSS

status	pronunciation	frequency
upper	[r]	30
upper	[ə]	6
middle	[r]	20
middle	[ə]	74
lower	[r]	4
lower	[ə]	50

- The variable *status* should be defined as ‘string’ and ‘nominal’, in that case SPSS will consider this variable as being categorical.
- In SPSS the six cases (3 statuses and 2 pronunciations each) should be weighed by their frequencies.

Odds

- Each observation belongs to one of the two categories: [ə] or [r].
- Probability of having [ə]:
$$\frac{6 + 74 + 50}{184} = 0.71$$
- Probability of having [r]:
$$\frac{30 + 20 + 4}{184} = 0.29$$
- **Odds**:
  The ratio of two fractions of two possible results.
- Odds of [ə] to [r]:
$$\frac{0.71}{0.29} = 2.45$$

Odds

- Interpretation:
  - If odds > 1,
    [ə] is more likely to be pronounced than [r].
  - If odds $= 1$,
    the two pronunciations have the same probability.
  - If odds < 1 and > 0,
    [r] is more likely to be pronounced than [ə].
- Now we calculate the odds for each of the statuses individually.

Odds

- Status=upper:

$$ODDS = \frac{6/(6+30)}{30/(6+30)} = 0.20$$

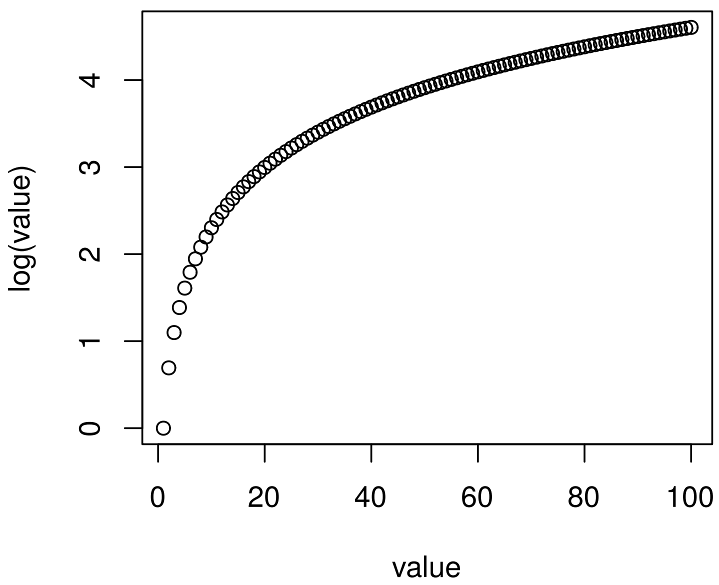
- Status=middle:

$$ODDS = \frac{74/(74+20)}{20/(74+20)} = 3.70$$

- Status=lower:

$$ODDS = \frac{50/(50+4)}{4/(50+4)} = 12.50$$

Logarithm



ln = logaritmus naturalis (natural logarithm)

Log Odds

- **Log odds** are logarithmically transformed **odds**.
- Log odds are centred around 0:
  - If odds > 0,
    [ə] is more likely to be pronounced than [r].
  - If odds $= 0$,
    the two pronunciations have the same probability.
  - If odds < 0,
    [r] is more likely to be pronounced than [ə].

Log Odds

- If p is the probability of having pronunciation [ə], and $1 - p$ the probability of having pronunciation [r], then:

$$log\ odds_{[\text{ə}]/[\text{r}]} = ln\left(\frac{p}{1-p}\right) = logit(p)$$

- Reversely, if $t = \text{logit}(p)$, than:

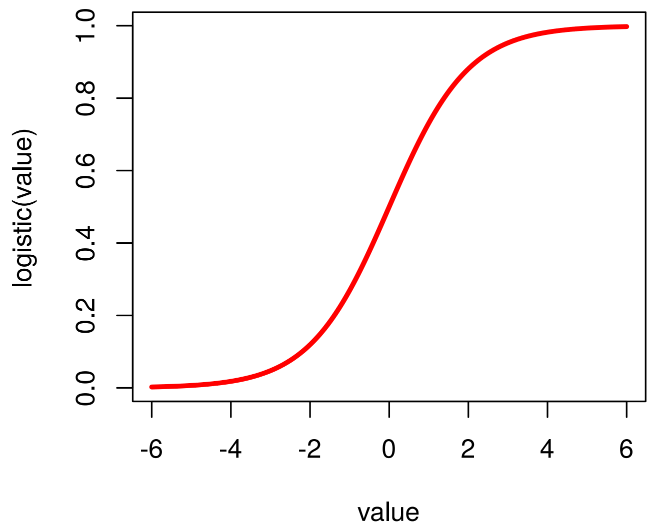
$$p = \frac{1}{1+e^{-t}} = logistic(t)$$

where $e = 2.718281828459...$

- $Logistic(t)$ is known as the **logistic** function.

Logistic function

- The logistic function can take an input with any value from negative to positive infinity, whereas the output always takes values between zero and one (Hosmer & Lemeshow 2000).



Model

- Model assumption for simple **linear** regression with mean μ and dependent variable y:

$$\mu_y = \beta_0 + \beta_1 x$$

- **Logistic** regression: if p is the probability of one outcome (pronunciation [ə] in our example), the mean response variable p in terms of the explanatory variable x is:

$$p = \beta_0 + \beta_1 x$$

- However, especially for high and low values of x it is not guaranteed that $0 \leq p \leq 1$. Solution: use the logistic function.

$$p = logistic(\beta_0 + \beta_1 x)$$

- The parameters of the logistic model are β_0 and β_1.

Model

- Note that the inverse of the model is:

$$logit(p) = \beta_0 + \beta_1 x$$

Estimate of the parameters

- SPSS replaces *status* by two binary **indicator** variables:

	status	stat(1)	stat(2)
upper	1	1	0
middle	2	0	1
lower	3	0	0

  where stat (1): compares *upper* to *lower*, stat (2): compares *middle* to *lower*.

- Parameters and their estimates:

parameter	estimate	
β_0	b_0	intercept
β_{1_1}	b_{1_1}	stat(1)
β_{1_2}	b_{1_2}	stat(2)

Estimate of the parameters

- If status=upper, than:

$$log\ odds_{upper} = ln(0.20) = -1.61 = b_0 + b_{11}1 + b_{12}0$$

- If status=middle, than:

$$log\ odds_{middle} = ln(3.76) = 1.32 = b_0 + b_{11}0 + b_{12}1$$

- If status=lower, than:

$$log\ odds_{lower} = ln(12.50) = 2.53 = b_0 + b_{11}0 + b_{12}0$$

Estimate of the parameters

- If status=upper, than:

$$-1.61 = b_0 + b_{11}1 + b_{12}0$$

- If status=middle, than:

$$1.32 = b_0 + b_{11}0 + b_{12}1$$

- If status=lower, than:

$$2.53 = b_0 + b_{11}0 + b_{12}0$$

- Therefore:
  b_0=2.53, b_{11}=-4.14 (stat(1)), b_{12}=-1.21 (stat(2)).

Confidence intervals

- A level C confidence interval for the slope β_1 is:

$$(b_1 - z^* SE_{b_1}, b_1 + z^* SE_{b_1})$$

- z^* is the value of the standard normal density curve with surface C between $-z^*$ and z^*.

Confidence intervals

- Stat(1):
  b_{1_1}=-4.135, $SE_{b_{1_1}}$=0.686, z^*=1.96 for a 95%-confidence interval
- A 95%-confidence interval for the slope β_{1_1} is:

$$(b_{1_1} - z^* SE_{b_{1_1}}, b_{1_1} + z^* SE_{b_{1_1}})$$

$$(-4.135 - 1.96 \times 0.686, -4.135 + 1.96 \times 0.686)$$

  We are 95% confident that the slope is found between -5.48 and -2.79.

- Similarly a confidence interval for slope β_{1_2} can be found.

Confidence intervals

- Given the negative interval (not including 0), we conclude that upper class people pronouncing the /r/ as [ə] is less likely than lower class people pronouncing the /r/ as [ə].

Odds ratio

- We know that:

if:

$$\begin{aligned} log(ODDS_{upper}) &= b_0 + b_{1_1} \\ log(ODDS_{middle}) &= b_0 + b_{1_2} \\ log(ODDS_{lower}) &= b_0 \end{aligned}$$

then:

$$\begin{aligned} ODDS_{upper} &= e^{b_0 + b_{1_1}} \\ ODDS_{middle} &= e^{b_0 + b_{1_2}} \\ ODDS_{lower} &= e^{b_0} \end{aligned}$$

- Hence, the odds ratios can be rewritten as follows:

$$\frac{ODDS_{upper}}{ODDS_{lower}} = \frac{e^{b_0 + b_{1_1}}}{e^{b_0}} = e^{b_{11}} \qquad \frac{ODDS_{middle}}{ODDS_{lower}} = \frac{e^{b_0 + b_{1_2}}}{e^{b_0}} = e^{b_{12}}$$

Odds ratio

- The ratio of two odds.
- Stat (1):

$$\frac{ODDS_{upper}}{ODDS_{lower}} = \frac{0.20}{12.50} = e^{b_{11}} = e^{-4.14} = 0.016$$

Pronunciation [ə] is 0.016 times more likely for the upper class than for the lower class.

- Stat (2):

$$\frac{ODDS_{middle}}{ODDS_{lower}} = \frac{3.76}{12.50} = e^{b_{12}} = e^{-1.21} = 0.300$$

Pronunciation [ə] is 0.300 times more likely for the middle class than for the lower class.

- If the slope is 0, than the odds ratio is 1.

Confidence intervals

- A level C confidence interval of the odds ratio e^{β_1} is obtained by transforming confidence interval of the slope:

$$(e^{b_1 - z^* SE_{b_1}}, e^{b_1 + z^* SE_{b_1}})$$

- z^* is the value of the standard normal density curve with surface C between $-z^*$ and z^*.

Confidence intervals

- Stat(1):
  b_{1_1}=-4.135, $SE_{b_{1_1}}$=0.686, z^*=1.96 for a 95%-confidence interval

- The 95%-confidence interval of the odds ratio $e^{\beta_{11}}$ is:

$$(e^{b_{1_1} - z^* SE_{b_{1_1}}}, e^{b_{1_1} + z^* SE_{b_{1_1}}})$$

$$(e^{-5.48}, e^{-2.79})$$

  We are 95% confident that the odds ratio is found between 0.004 and 0.061.

- Similarly a confidence interval the odds ratio $e^{\beta_{12}}$ can be found.

Confidence intervals

- Since the interval is smaller than 1, we conclude that the probability that upper class people will pronounce the /r/ as [ə] is smaller than the probability that lower class people pronounce the /r/ as [ə] (odds ratio $=$ 0.016, CI is 0.004 to 0.061).

Significance tests

- Hypotheses significance tests of β_1:

  $H_0 : \beta_1 = 0$ (the slope is 0, or: the odds ratio is 1)
  $H_a : \beta_1 \neq 0$; the p-value is $P(\chi^2 > X^2)$

  where X^2 is a stochastic variable which has approximately a χ^2 distribution with 1 degree of freedom.

- The test statistic X^2 is:

$$X^2 = \left(\frac{b_1}{SE_{b_1}}\right)^2$$

Significance tests

- Hypotheses for Stat(1):

$$H_0 : \beta_{1_1} = 0$$
$$H_a : \beta_{1_1} \neq 0$$

- Test statistic:

$$X^2 = \left(\frac{b_{1_1}}{SE_{b_{1_1}}}\right)^2 = \left(\frac{-4.135}{0.686}\right)^2 = 36.33$$

- There is 1 degree of freedom.

Significance tests

- Go to `http://www.vassarstats.net/` and choose Distributions, Chi-Square Distributions. Enter the number of degrees of freedom (df=1).
- The table on the left does not show an entry for X^2=36.33, the highest value is 14.0 with p-value = 0.000183.
- We report $p < 0.0005$. We reject H_0 and accept H_a.
- A similar procedure is followed for Stat(2).

Results

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	STAT			43,900	2	,000			
	STAT(1)	-4,135	,686	36,382	1	,000	,016	,004	,061
	STAT(2)	-1,217	,578	4,444	1	,035	,296	,095	,918
	Constant	2,526	,520	23,627	1	,000	12,500		

a. Variable(s) entered on step 1: STAT.

- STAT(1): compares upper versus lower, and STAT(2) compares middle versus lower.
- De column B gives b_{1_1}, b_{1_2} and b_0.
- De column $Wald$ gives the X^2 values.
- The column Exp(B) gives the odds ratios, followed by the corresponding confidence intervals.

Reference level

- Reference level: lower class. Therefore stat(1) compares upper to lower and stat(2) compares middle to lower.
- The choice of the reference level depends on your research question.
- Given the fact that for both the middle and lower class most speakers pronounce [ə] while most upper class speakers pronounce [r], it may be more meaningful to choose the upper class as reference level.
- In that case both lower and upper are compared to upper.
- If you choose a reference level of the variable thoughtlessly, you can miss important information from your data.

Likelihood ratio R^2

- How well explains status the variation in the pronunciation of /r/?
- *Log Likelihood* L measures the quality of the model: how well does the model fit the data?
- L is calulated as:

$$L = k \times ln(p) \quad + \quad (n-k) \times ln(1-p)$$

where:
k: number of times of having a particular outcome
n: total number of observations
p: probability of having a particular outcome according to the model

Likelihood ratio R^2

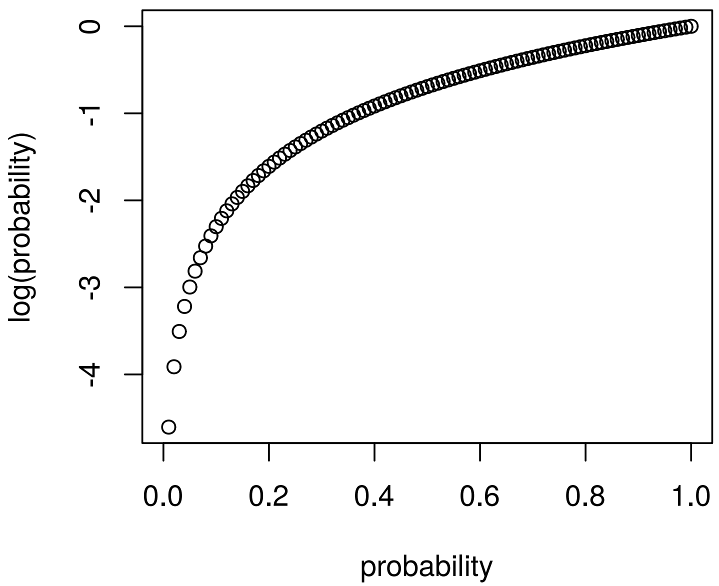
- The log likelihood of the model is:

$$L = \sum_{i=1}^{m} (k_i \; ln(p_i) + (n_i - k_i) \; ln(1 - p_i))$$

where:
m: the number of different values of the explanatory variable

- $-2L$ has a χ^2 distribution with $n - 1$ degrees of freedom.
- The lower $-2L$, the better the model predicts the probabilities of having a particular outcome, i.e. of having pronunciation [ə] in our example.

Likelihood ratio R^2



Logarithmic probabilities.

Likelihood ratio R^2

- First we calculate the -2 Log Likehood in case we do not distinguish social statuses:

	pronunciation	
status	[r]	[ə]
upper	30	6
middle	20	74
lower	4	50
totaal	54	130

In total there are 184 observations, 130 times the /r/ is pronounced as [ə], therefore the estimated $p = 0.707$ and $1 - p$=0.293.

Likelihood ratio R^2

- The log likelihood is:

$$L = k\ ln(p) + (n - k)\ ln(1 - p)$$

therefore:

$$L = 130\ ln(0.707) + 54\ ln(0.293) = -111.5$$

and:

$$-2L = -2 \times -111.5 = 223.0$$

Likelihood ratio R^2

- Now we calculate the -2 Log Likelihood for each value of the explanatory variable $status$: upper, middle, lower.

Likelihood ratio R^2

- Log likelihood 'upper':
36 observations, 6 times pronuncation [ə], $p = 0.167$, $1 - p$=0.833:

$$L_{upper} = 6\ ln(0.167) + 30\ ln(0.833) = -16.2$$

- Log likelihood 'middle':
94 observation, 74 times pronunciation [ə], $p = 0.787$, $1 - p$=0.213:

$$L_{middle} = 74\ ln(0.787) + 20\ ln(0.213) = -48.6$$

- Log likelihood 'lower':
54 observations, 50 times pronunciation [ə], $p = 0.926$, $1 - p$=0.074:

$$L_{lower} = 50\ ln(0.926) + 4\ ln(0.074) = -14.3$$

Likelihood ratio R^2

- Now we calculate the sum of L_{upper}, L_{middle} en L_{lower}:

$$L_{status} = -16.2 - 48.6 - 14.3 = -79.1$$

therefore:

$$-2L_{status} = -2 \times -79.1 = 158.2$$

Likelihood ratio R^2

- Reduction:

$$-2L - \ -2L_{status} = 223.0 - 158.3 = 64.7$$

- Number of degrees of freedom: number of categories according to the explanatory variable - 1. In our example: 3 - 1 = 2.
- Go to `http://www.vassarstats.net/` and choose Distributions, Chi-Square Distributions. Enter the number of degrees of freedom (df=2).
- The table on the left does not show an entry for X^2=64.7, the highest value is 16.0 with p-value = 0.000335. We report $p < 0.0005$.
- What amount of variance in the response variable (pronunciation) is explained by the explanatory variable (status)?

$$R^2_{logistic} = \frac{-2\ L - -2L_{status}}{-2\ L} = \frac{64.7}{223.0} = 0.290 = 29\%$$

Significance and effect size

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	158,267	,296	,421

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	64,461	2	,000
	Block	64,461	2	,000
	Model	64,461	2	,000

Watch out:

Chi-square $= -2\ L - -2L_{status} = 64.461$, **-2 Log likelihood** $= -2\ L_{status} = 158.267$

Therefore: $-2\ L = [-2\ L - -2L_{status}] + [-2\ L_{status}] = 64.461 + 158.267 = 222.728$

Significance and effect size

- Effect size:

$$R^2_{logistic} = \frac{-2\ L - -2L_{status}}{-2\ L} = \frac{64.461}{222.728} = 0.289 = 29\%$$

This is almost the same as SPSS's **Cox & Snell R Square**.

Classification table

- In SPSS:

Classification Table[a]

Observed			Predicted uitspraak 0	Predicted uitspraak 1	Percentage Correct
Step 1	uitspraak	0	30	24	55,6
		1	6	124	95,4
	Overall Percentage				83,7

a. The cut value is ,500

- Encoding: 0=pronunciation [r], 1=pronuncation [ə].
- 24 [r]'s are predicted as [ə]'s, 6 [ə]'s are predicted as [r].
- Percentage of correctly predicted speech segments:

$$\frac{30 + 124}{30 + 24 + 6 + 124} = 83.7\%$$

Example 2

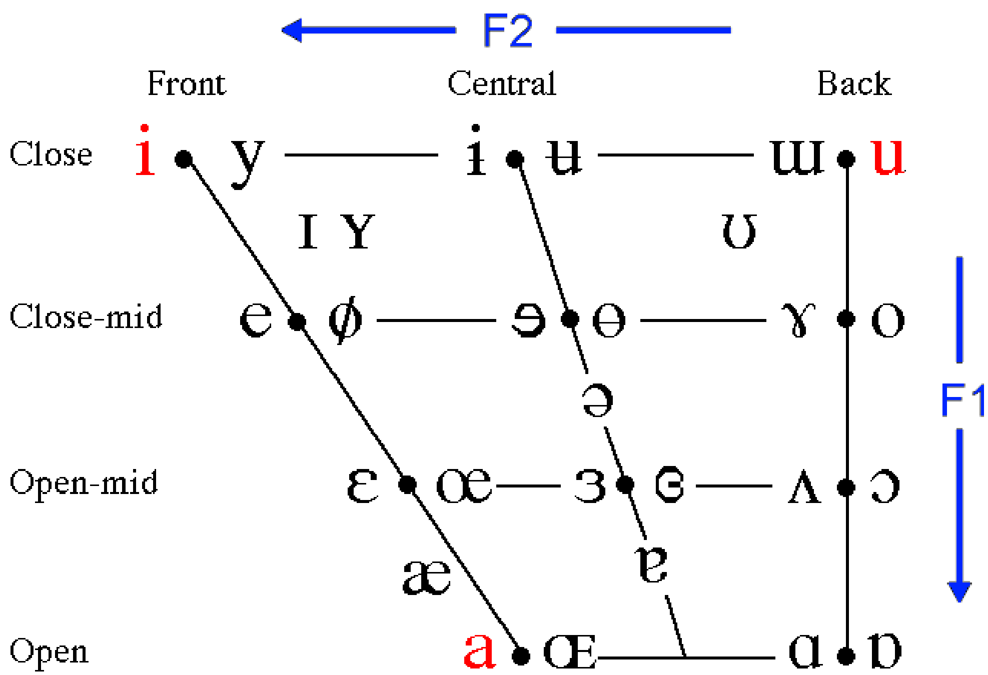
- *The north wind* is translated in Norwegian as *nordavinden*.
- For 15 Norwegian dialects we investigate the pronuncation of /u/ and /i/ in the first and third syllable respectively
- There are 4 male speakers and 11 female speakers.
- Source: recordings made by Jørn Almberg and Kristian Skarbø (Trondheim) which are available at:

  `http://www.ling.hf.ntnu.no.nos`

- Timbre of vowels is determined by the intensities of frequencies. Formants are frequencies which are amplified by the vocal tract. Lowest formants: F1 and F2.

Distribution of 15 dialects in the Norwegian language area.

Vowel space and formants



- F1 runs from 240Hz [i] to 850Hz [a], F2 runs from 595Hz [u] to 2400Hz [i].
- Can gender be predicted by looking at the F2 frequencies in someone's speech?

Variables

- In this example there are multiple (i.e. two) **numerical** (or quantitative) explanatory variables – F2[i] and F2[u] – and a categorical response variable.

Assumptions

- 1. Linearity:
  There should be a linear relationship between any continuous predictor and the logit of the outcome variable. We test whether the interaction between a predictor and its log transformation is significant.
- 2. No perfect multicollinearity:
  Make scatterplots and calculate correlation coefficients for each pair of predictors. The r's should be lower than 0.9.
- Calculate the Variance Inflation Factor (VIF) for the continuous predictors. The VIF is an index which measures how much variance of an estimated regression coefficient is increased because of multicollinearity.

Assumptions

- 3. Independence:
  Cases of data should not be related. Violating this assumption procedeces overdispersion.
- Overdispersion: the presence of greater variability in a data set than would be expected based on the statistical model. The test statistic will be too high, and p-values too small (Type I errors).

1. Linearity

- In SPSS we compute the logarithmic transformations of F2[i] and F2[u]. We call them: lnF2[i] and lnF2[u].
- Perform the regression analysis and enter the following 'covariates': F2[i], F2[u], F2[i]*lnF2[i], F2[u]*lnF2[u].
- Results:

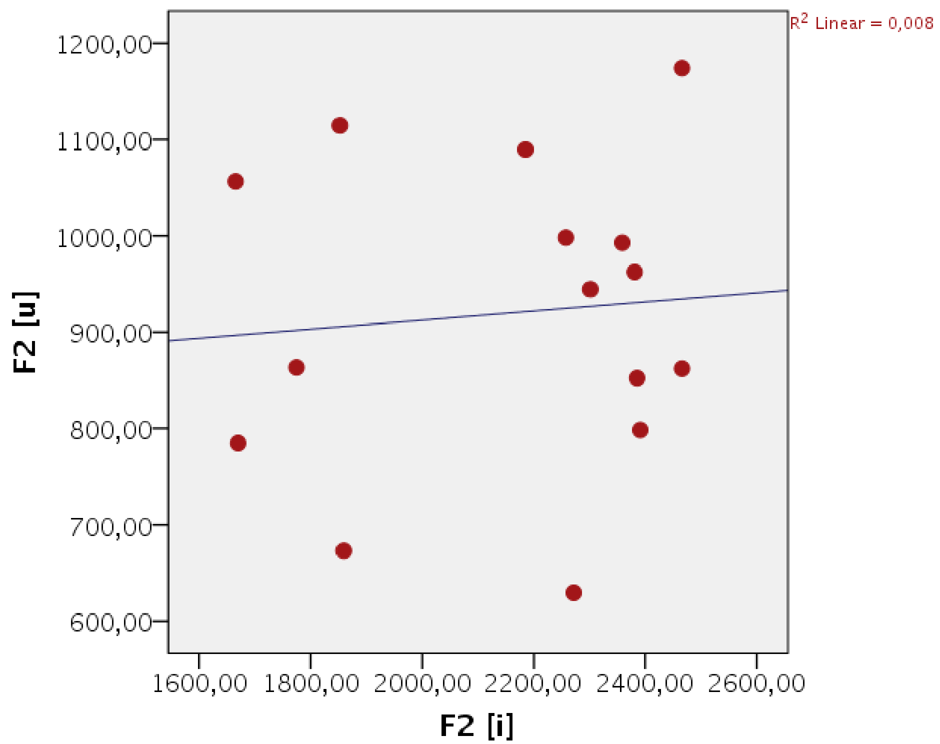
Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	F2u	-,090	,509	,031	1	,860	,914
	F2i	,239	,658	,132	1	,717	1,270
	F2i by lnF2i	-,027	,076	,126	1	,723	,973
	F2u by lnF2u	,012	,065	,034	1	,854	1,012
	Constant	-57,559	187,880	,094	1	,759	,000

a. Variable(s) entered on step 1: F2u, F2i, F2i * lnF2i , F2u * lnF2u .

- Interactions are not significant, the assumption is met.

2. No perfect multicollinearity



Correlation r=0.088 ($p = 0.755$), and R^2=0.008.

2. No perfect multicollinearity

- Another way is to run a linear regression regression with the same predictors and response variable. Under Statistics check Collinearity diagnostics. Switch off all of the default options.
- Look in the second table called ‘Coefficients’ in the column ‘VIF’.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-1,590	,939		-1,693	,116		
	F2 [u]	,000	,001	,166	,721	,485	,992	1,008
	F2 [i]	,001	,000	,569	2,471	,029	,992	1,008

a. Dependent Variable: geslacht

- All VIF values should be smaller than 10, and the average of the VIF values should not substantially be greater than 1.
- We find all values smaller than 10 and not substantially be greater than 1.

3. Independence

- All cases are independent of each other.
- The dispersion parameter $\hat{\o}$ is the test statistic divided by the degrees of freedom. This ratio should be smaller or equal to 1.
- In our case:

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	5,795	2	,055
	Block	5,795	2	,055
	Model	5,795	2	,055

$$\hat{\o} = \frac{\chi^2}{df} = \frac{5.795}{2} = 2.8975 \qquad > 1$$

Model without predictors

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	1,012	,584	3,002	1	,083	2,750

Classification table

Classification Table[a,b]

			Predicted		
			geslacht		Percentage Correct
Observed			0	1	
Step 0	geslacht	0	0	4	,0
		1	0	11	100,0
	Overall Percentage				73,3

a. Constant is included in the model.

b. The cut value is ,500

Encoding: 0=male speaker, 1=female speaker.

Model with predictors

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	F2 [u]	,003	,005	,584	1	,445	1,003	,995	1,012
	F2 [i]	,005	,003	3,754	1	,053	1,005	1,000	1,010
	Constant	-12,773	7,188	3,157	1	,076	,000		

a. Variable(s) entered on step 1: F2U, F2I.

Significance and effect size

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	11,603	,320	,467

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	5,795	2	,055
	Block	5,795	2	,055
	Model	5,795	2	,055

$$R^2_{logistic} = \frac{5.795}{11.603 + 5.796} = \frac{5.795}{17.398} = 0.333 = 33\%$$

Classification table

Classification Table[a]

			Predicted		
			geslacht		Percentage Correct
	Observed		,00	1,00	
Step 1	geslacht	,00	3	1	75,0
		1,00	1	10	90,9
	Overall Percentage				86,7

a. The cut value is ,500

Encoding: 0=male speaker, 1=female speaker. The speaker in Herøy is a male speaker, but is predicted being a female speaker by the model. The speaker in Larvik is a female speaker, but is predicted as being a male speaker by the model.

Example 3

- Entering high school students make program choices among general program, vocational program and academic program.
- Their choice might be modeled using their reading score, math score and their social economic status.
- Example taken from: *R Data Analysis Examples: Multinomial Logistic Regression*, from `http://www.ats.ucla.edu/stat/mult_pkg/faq/general/citingats.htm`, (accessed May 9, 2016).

Example 3

- Predictor variables:
  reading score (read), *math score* (math), *social economic status* (socst)

- Response variable:
  program choice with possible values: vocational, general, academic.

Results

```
Coefficients :
                      Estimate Std. Error t-value  Pr(>|t|)
academic:(intercept) -9.3031628  1.5650170 -5.9444 2.774e-09 ***
general:(intercept)  -4.1165411  1.5181110 -2.7116  0.006696 **
academic:read         0.0235763  0.0289506  0.8144  0.415436
general:read          0.0061777  0.0302050  0.2045  0.837942
academic:math         0.1013774  0.0317736  3.1906  0.001420 **
general:math          0.0364647  0.0334915  1.0888  0.276254
academic:socst        0.0718275  0.0244025  2.9434  0.003246 **
general:socst         0.0410250  0.0246267  1.6659  0.095738 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Results

- Every predictor appears twice, first with *academic*, and then with *general*.
- Comparison to reference level 'vocational'.
- Estimate shows log odds ratios, we convert them to odds ratios.

Results

	odds ratios	sig
academic:(intercept)	0.000091	< 0.001
general:(intercept)	0.016301	< 0.01
academic:read	1.023856	
general:read	1.006197	
academic:math	1.106694	< 0.01
general:math	1.037138	
academic:socst	1.074470	< 0.01
general:socst	1.041878	

- The choice of an academic program is 1.106694 times more likely than the choice of a vocation program for student with higher math scores.
- The choice of an academic program is 1.074470 times more likely than the choice of a vocational program for students with a higher social economic status.

Results

```
Log-Likelihood: -170.98
McFadden R^2:  0.16226
Likelihood ratio test : chisq = 66.233 (p.value = 2.4157e-12)
```

- McFadden's R^2 approximates the Likelihood ratio R^2:

$$R^2 = \frac{66.233}{66.233 + (-\mathbf{2}\times - 170.98)} = 0.16226$$

- Values from 0.2 to 0.4 correspond with 0.7 to 0.9 in linear models. These are considered to indicate a very good fit (Louviere et al. 2000: 55).
- Reduction of unexplained variance from the baseline model (model without predictors): χ^2 is 66.233 which is a significant improvement.