Introduction to Statistics
lab 11

Johan A. Elkink
jos.elkink@ucd.ie

27 November 2017

Data

For this lab we will make use of the Asia Barometer data as prepared on the teaching data page.[1]

Logistic regressions

To perform logistic regression analysis in your software package of choice, you can use the following commands, assuming the dependent variable is **y** and the independent variables are **x1** and **x2**. Note that inserting categorical independent variables in the model works exactly the same as for linear regression.

SPSS: `LOGISTIC REGRESSION VAR = y /METHOD = ENTER x1 x2 /SAVE PRED (pred) PGROUP (predg).` [2]

R: `summary(m <- glm(y ~ x1 + x2, data = ab, family = binomial(link = "logit")))`

Stata: `logit y x1 x2`

Producing a confusion table:

SPSS: `CROSSTABS predg BY y.`

R: `table(predict(m, type = "response") > 0.5, ab$y)`

Stata: `predict pred`
`gen predg = pred > 0.5`
`tab predg y`

Producing ROC curves can be done as follows:

SPSS: `ROC pred BY y (1).`

R: `res <- HMeasure(ab$y, predict(m)); plotROC(res)` [3]

---

[1] http://www.joselkink.net/teaching/teaching-data/

[2] The SAVE subcommand is not necessary for the regression, but required if you want to use the predicted probabilities or groups subsequently.

[3] This requires the hmeasure package.

Stata: `predict linpred, xb`
`roccomp y linpred, graph` [4]

Questions

1. Table 1 contains a series of four logistic regressions explaining turnout in elections (**abstain**). Replicate these models. The variables used are **polinterest**, **urban**, **female**, **age**, and **country**.

2. What do you conclude about the relation between political interest and turnout in elections? Tip: use the "divide by four" rule of thumb and consider the scale on which the independent variable is measured.

3. Based on these regressions, would you conclude that urbanisation is an important predictor?

4. By hand (calculator), based on Model 2, compute the predicted probability for an urban female to vote. Repeat for an urban male.

5. Using Model 3, produce a confusion table.

6. Using Model 3, produce an ROC curve.

---

[4]See `http://www.ats.ucla.edu/stat/stata/faq/roc.htm`. This needs to be done immediately after the `logit` command.

	(1)	(2)	(3)	(4)
Interest in politics	−0.38***	−0.38***	−0.36***	−0.45***
	(0.12)	(0.12)	(0.13)	(0.14)
Urban		0.58**	0.46*	0.13
		(0.24)	(0.25)	(0.30)
Female		−0.13	−0.10	−0.18
		(0.21)	(0.22)	(0.24)
Age			−0.04***	−0.05***
			(0.01)	(0.01)
Japan				0.40
				(0.40)
Korea				−1.88***
				(0.48)
China				−1.19**
				(0.55)
Philippines				−0.21
				(0.39)
Taiwan				−1.47***
				(0.44)
Thailand				−1.99***
				(0.60)
intercept	−1.02***	−1.38***	0.33	1.91***
	(0.19)	(0.30)	(0.45)	(0.62)
N	701	701	602	602

Note: $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 1: Logistic regressions explaining turnout in elections. Countries are relative to Hong Kong.