GLM 1: Practical session

In this practical session you will be given a data set and asked to fit and interpret a series of GLMs.

HINT: use

```
glm <outcome> <explanatory vars>, family() link() eform
```

to return exponentiated parameters and their 95% C.I.s in the model output.

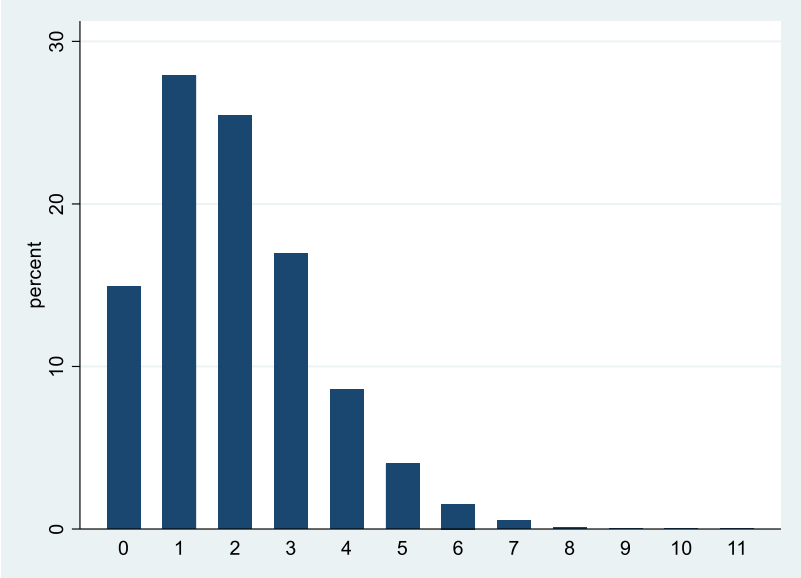
The data can be found on the **Advanced Statistics for Records Research** moodle page and are in asrr_glm_1.dta. The file contains simulated (i.e. made up) data for 5000 people covering a study period of 1 year. This is all data that could be derived from administrative hospital records.

Variable Name	**Variable Label**	**Variable Description**
id	person id	unique person identifier from 1 to 5000
age	age (year)	age in years at beginning of the study period
sex	sex	1 = Female 0 = Male
cvd	History of Cardiovascular Disease	1 = diagnosis of cardiovascular disease in the year before the study period 0 = no diagnosis
imd5	IMD quintiles	Index of multiple deprivation (IMD) quintiles derived from patient addresses (postcodes). 1 = most deprived, 5 = least deprived
ae_atd	A&E attendances	number of A&E attendances over the study period
death	Death	1 = died over the study period 0 = alive at the end of the study period

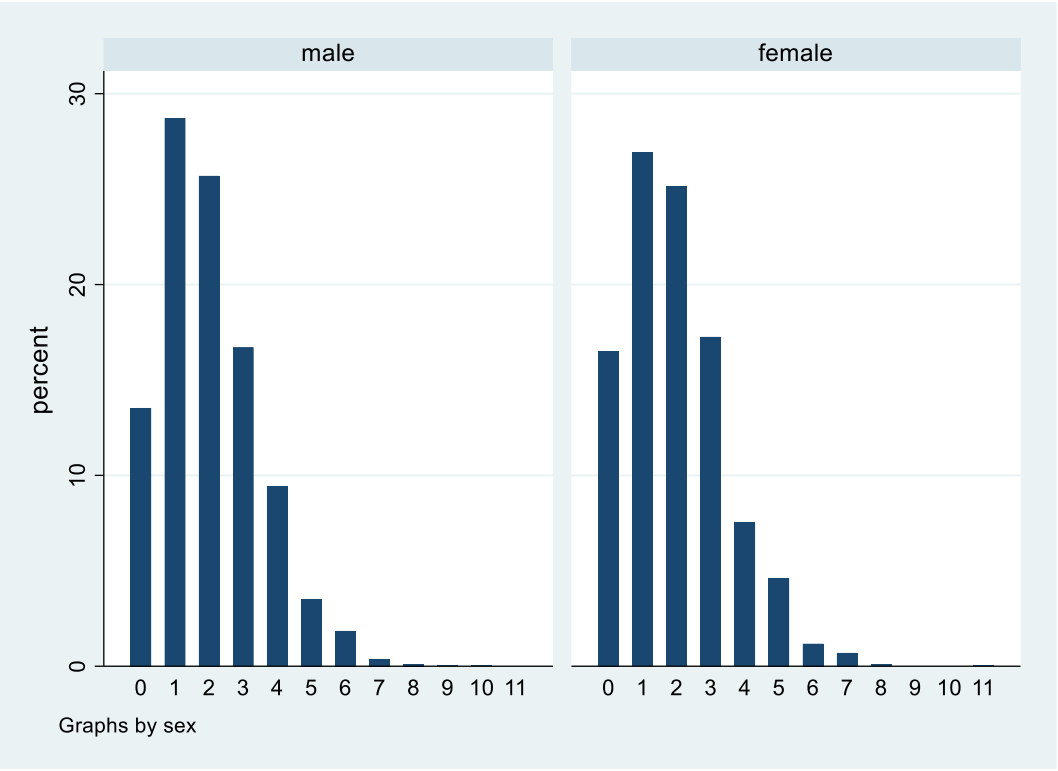
1. Using the `tabulate <var>` (or `tab`) or `summarize <var>, detail` (or `su <var>, d`) command in stata, complete the table below to get a better idea of what is in the data file (e.g. `tab imd5` or `su age, d`) :

Patient characteristic	**Overall (n = 5000)**
Age (years), median (min, max)	60 (37, 84)
Sex (1 = Female) (%)	2295 (45.9)
IMD quintiles (%)	
1	962 (19.2)
2	1019 (20.4)
3	986 (19.7)
4	1040 (20.8)
5	993 (19.9)
History of Cardiovascular Disease (%)	1824 (36.5)
Deaths (%)	2111 (42.2)
A&E attendances (median [range])	2 (0, 11)

2. Now we're going to look at how the risk of A&E attendance over the study period varies by age and sex:
    a. Using `graph bar, over(ae_atd)`, look at the distribution of number A&E attendances over the whole sample.



b. How does its distribution differ by sex? (HINT: add the option `by(<var>)` to your stata command)



c. Use an appropriate GLM to confirm whether the number of A&E attendances differs by sex (remember to add the prefix `i.` when including binary or factor variables in your modelling)

Sex	Rate Ratio	95% C.I.
Female	0.97	0.94-1.01

d. Add age to the model in (c). Interpret the association between number of A&E attendances and age, adjusting for sex:
Holding sex constant, a one year increase in age is associated with a 3.2% increase (95% CI, 2.9 to 3.5%) in the number of A&E attendances over the study period (i.e. increases in age are associated with an increased risk of A&E attendances).

ae_atd	IRR	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
sex						
female	.971626	.0196068	-1.43	0.154	.9339475	1.010825
age	1.032167	.001655	19.75	0.000	1.028928	1.035416
_cons	.294489	.0291912	-12.33	0.000	.24249	.3576385

e. Finally, add cvd to the model in (d). Interpret the association between number of A&E attendances and History of Cardiovascular disease, adjusting for age and sex:
Holding sex and age constant, having a history of cardiovascular disease versus not having a history of cardiovascular disease is associated with a 20.9% increase (95% C.I. 16.2 to 25.9%) in the number of A&E attendances over the study period (i.e. having a history of cardiovascular disease is associated with an increased risk of A&E attendances).

ae_atd	IRR	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
sex						
female	.9755615	.019691	-1.23	0.220	.9377213	1.014929
age	1.030833	.0016599	18.86	0.000	1.027584	1.034091
cvd						
yes	1.209132	.0247201	9.29	0.000	1.16164	1.258567
_cons	.2951683	.0292664	-12.31	0.000	.2430367	.3584822

3. Now we will look at how the likelihood of death varies over the study period by some of the other patient characteristics.

a. How many people died, by IMD quintile?
`tab imd5 death, row`

IMD quintile	Number that died	%
1	433	45.0
2	454	44.6
3	419	42.5
4	439	42.2
5	366	36.9

b. Using a GLM, model the number of deaths by IMD quintile and completed the table below:

IMD quintile (compared to 1)	Odds Ratio	95% C.I.
2	0.98	0.822 to 1.172

3	0.90	0.755 to 1.080
4	0.89	0.748 to 1.065
5	0.71	0.595 to 0.855

c. Interpret these findings (i.e. the parameter estimates and their 95% C.I.s)?

The effect of living in a less deprived neighboured than living in one of the most deprived neighbourhoods is associated with a reduction in the odds of dying. However, the 95% C.I.s for the effects of living in a neighbourhood within the 2nd, 3rd or 4th IMD quintile compared to living in the one of the most deprived neighbourhoods are compatible with no effect (i.e. they cross 1). The effect of living in the least deprived neighbourhoods compared to living in the most deprived neighbourhoods is associated with a 29% decrease (1-0.71 = 0.29, 95% C.I. of a 14 to 40% decrease) in the odds of dying.

d. Add in age and sex to the model. Describe what happens to the parameter estimates for imd5 (in one sentence)?
They pretty much stay the same (i.e. there is still evidence that the odds of dying are decreased among those who live in the least deprived neighbourhoods.)

death	Odds Ratio	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
imd5						
2	.9836556	.0897486	-0.18	0.857	.8225828	1.176269
3	.8908176	.0821875	-1.25	0.210	.7434577	1.067386
4	.8943572	.0814989	-1.23	0.220	.7480742	1.069245
5 - least deprived	.7080062	.0660021	-3.70	0.000	.5897745	.8499398
age	1.045206	.0049217	9.39	0.000	1.035604	1.054897
sex						
female	.9076755	.0527777	-1.67	0.096	.8099098	1.017243
_cons	.0600455	.0175141	-9.64	0.000	.0339002	.1063555

e. Add cvd into the model in (e). Describe what happens to the parameter estimates for imd5 now (in one sentence)?
The parameter estimates have changed and now the confidence interval for the effect of living in one of the least deprive neighbourhoods on the odds of dying is compatible with no effect (i.e. it crosses 1).

death	Odds Ratio	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
imd5						
2	1.064797	.1001809	0.67	0.505	.885486	1.280417
3	.9944426	.0947547	-0.06	0.953	.8250373	1.198632
4	1.0787	.1019433	0.80	0.423	.8963082	1.298207
5 - least deprived	.8849785	.0856053	-1.26	0.207	.732141	1.069722
age	1.040353	.0050364	8.17	0.000	1.030529	1.050271
sex						
female	.9233467	.0552097	-1.33	0.182	.8212377	1.038152
cvd						
yes	2.72559	.168941	16.18	0.000	2.413794	3.077661
_cons	.0478156	.0143726	-10.12	0.000	.0265285	.086184

Practical end.

Solutions and code (stata and R) will be available on the course moodle page at the end of the practical.