Multilevel Models

David Barron

Hilary Term 2017

Introduction

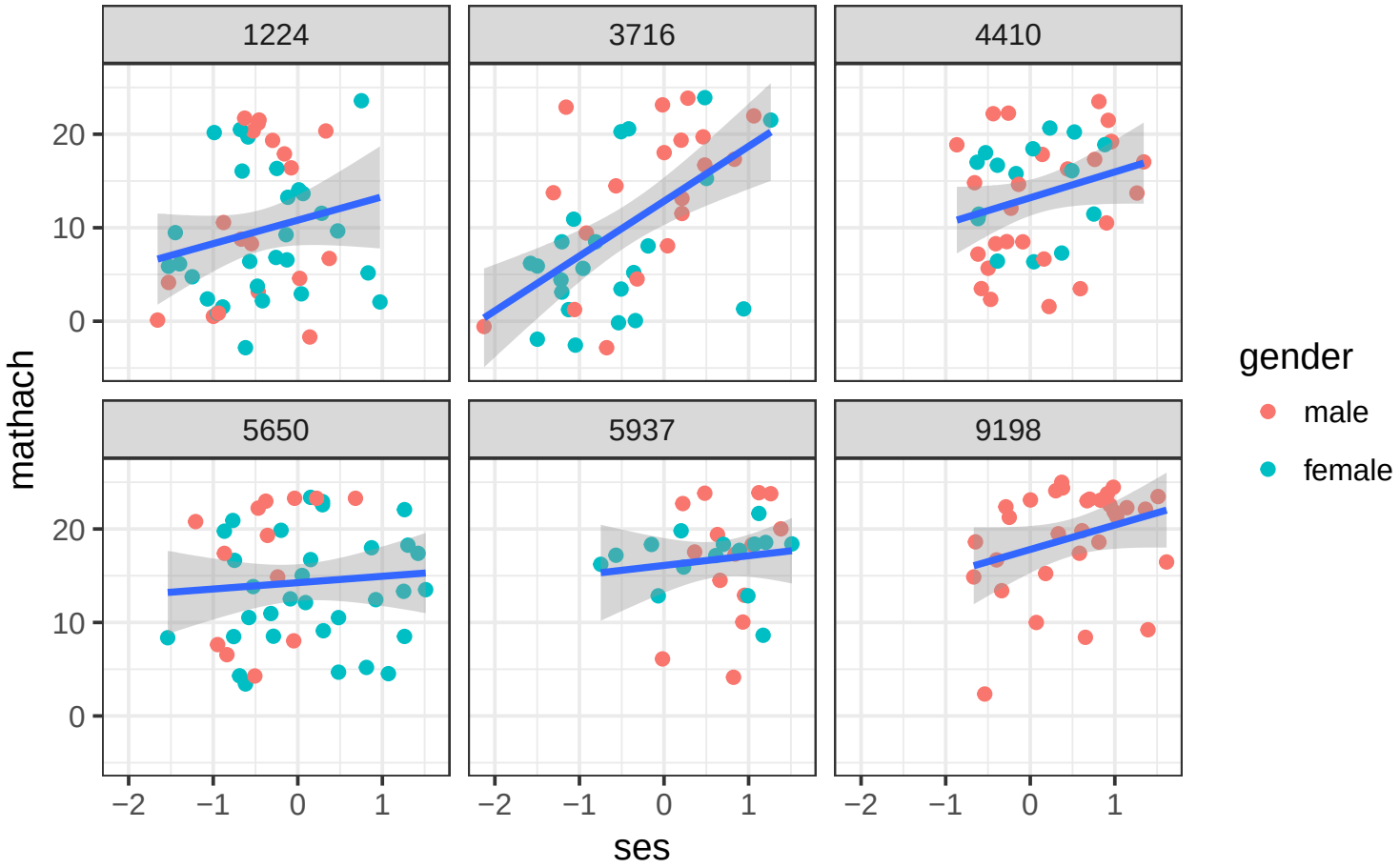
Hierarchical data

Social research often involves investigating relationships between individuals and social groups. The general concept is that individuals are influenced by social groups to which they belong, and that the properties of these groups are in turn influenced by the individuals who make up that group. This gives rise to a hierarchical model, with individuals nested within groups. For example, we might select a sample of schools and then sample the students within each school. There can be more than two levels in the hierarchy (e.g., student, class, school, city, region, etc.), although we'll restrict our attention to the two-level case.

Example

For example, we might be interested in the relationship between the socio-economic status of high school students and their scores in maths tests. In order to investigate this relationship we could randomly sample students, but a more realistic procedure would be to sample schools and then look at all the students of a given age in each school. The following plot shows scatter plots and regression lines for six schools (out of 160 in the study). This is one possible approach to such data: analyse each school separately. At the other extreme, we could pool all the data together and get a single estimate. The methods we are going to be looking at can be thought of as falling between these extremes in the sense that they enable us to *model* the variation in the effect of SES on maths achievement across schools.

Example data



Levels of measurement

Variables can be defined at any level of the hierarchy. In some cases variables are measured directly at their natural level. For example, at the school level we can measure school size, whether it is a private or state school, whether it is in an urban or rural location, etc. At the individual student level we can measure their sex, socio-economic status, etc. We can also move from a lower to higher level by aggregation, for example the mean SES of students in the school.

Why not use standard regression methods?

The main statistical issue is lack of independence of the cases. That means that hypothesis tests are misleading; we are, in effect, treating each level 1 case as providing completely independent information, which they do not do. We might expect a higher degree of similarity among individuals within level 2 groups than between these groups.

Aggregation/disaggregation?

One possible "solution" is to aggregate all level 1 variables to level 2. This wastes information and can also lead to the *ecological fallacy*: drawing conclusions about level 1 based only on measurements at level 2. Robinson (1950) presented data on the percentage of blacks and illiteracy rates in 9 US regions in 1930. Correlation at the aggregate level was .95, but at the individual level (individual race and illiteracy) correlation was only .20 Similar problems can result from disaggregation. In addition, we cannot make inferences about cross-level hypotheses.

Regression model

2-level regression model

We are going to start by looking at normally-distributed outcome variables (or at least where this is a reasonable approximation, such that in a single-level analysis we would use linear regression).

The two-level model assumes that there is a hierarchical data set with a single dependent variable measured at the lowest level (level 1) and explanatory variables at levels 1 and 2. It can be viewed as a system of regression equations. For example, suppose we have collected data in J schools with data from N_j students in each school. On the student level we have the dependent variable "math test score" (Y) and the explanatory variable "socio-economic status," (X) and on the school level we have the explanatory variable "school mean SES" (Z).

Traditional models

There are two "extreme" models that we could use to analyse these data.

- **Full pooling:** All pupils are pooled together and, with the exception of school-level explanatory variables, school is ignored.
- **No pooling:** Perform separate regressions for each school. One problem with this is we are unable to obtain estimates of school-level variables.

Multi-level models can be thought of as being in between these two extremes, sometimes called *partial pooling*.

Level 1 model

We can set up a regression equation in each school to predict Y by X as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \epsilon_{ij}.$$

This means that we assume each school has a different intercept (β_{0j}) and a different slope (β_{1j}). We assume that the random errors have zero mean and usually also that there is a constant variance σ^2. Since the intercept and slope vary across level 2 units, they are often called random coefficients.

Level 2 model

The idea is that we can model the variation in the random coefficients by making them the "dependent variables" in a set of regressions:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j}$$

The first equation implies that the general level of math achievement (β_{0j}) in each school can be explained by the mean SES. The second equation implies the relationship between maths achievement and SES depends on mean SES. If γ_{11} is positive, students in schools with high mean SES tend to see maths achievement increase more rapidly with SES than students in schools with low mean SES.

Error terms

The two level-2 error terms are assumed to have zero mean and constant variances, τ_{00} and τ_{11}. The are assumed to be independent of the errors at level 1. The covariance, τ_{12} between these error terms is *not* assumed to be zero.

Fixed and random parts

Note that γ_{00} and γ_{10} do not vary across schools: they are *fixed coefficients.* All the between school variation left in the β coefficients after predicting these with the level-2 variable is assumed to be residual error variation. We can rearrange the two levels of equations into a single equation:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij} + u_{1j}X_{ij} + u_{0j} + \epsilon_{ij}$$

Note that this consists of fixed and random parts. Note that this equation shows that there will be heteroskedasticity.

Intra-class correlation

The lack of independence among level-1 units within a level-2 unit can be expressed as the intra-class correlation coefficient, ρ. If we estimated a 2-level model without explanatory variables, we would have:

$$Y_{ij} = \gamma_{00} + u_{0j} + \epsilon_{ij}.$$

This decomposes the variance into two independent components: σ^2, the variance of the ϵ_{ij} and τ_{00}, the variance of the u_{0j} errors. We can then estimate the intra-class correlation by using

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2}.$$

This is an estimate of the variance explained by the grouping structure. ICC is just the proportion of group level variance compared to total variance.

Putting it into practice

Example data

Data on 7185 school students in 160 schools. The outcome variable is achievement in a maths test. Explanatory variables include socioeconomic status (a level-1 variable); school mean socioeconomic status and school sector (level-2 variables).

Example: Intercept only

```
lmer(formula = mathach ~ 1 | schid, data = hsb)
coef.est  coef.se
   12.64     0.24

Error terms:
 Groups   Name        Std.Dev.
 schid    (Intercept) 2.93
 Residual             6.26
---
number of obs: 7185, groups: schid, 160
AIC = 47122.8, DIC = 47115
deviance = 47115.8
```

Interpretation

The fixed effect is merely the constant term: it is the mean of all the schools' mean math achievement scores. The two random effects parameters are the variances of the two random effects, so $\tau_{00} = 8.61$ and $\sigma^2 = 39.15$. The intraclass correlation is therefore:

$$\rho = \frac{8.61}{8.61 + 39.15} = .18$$

This shows us that there is more variation in maths achievement within than between schools, and that there is a fair amount of clustering of math achievement scores within schools.

Include a level-2 fixed effect

We add the effect of school mean socioeconomic status as an explanatory variable—in other words, we think that the intercept varies across schools depending on the school's mean SES:

$$Y_{ij} = \beta_{0j} + \epsilon_{ij} \text{ and } \beta_{0j} = \gamma_{00} + \gamma_{01}\text{MEANSES} + u_{0j}$$
$$\text{where } \epsilon_{ij} \sim N(0, \sigma^2) \text{ and } u_{0j} \sim N(0, \tau_{00})$$

Combining these together gives us:

$$Y_{ij} = \{\gamma_{00} + \gamma_{01}\text{MEANSES}\} + \{u_{0j} + \epsilon_{ij}\}$$

Level-2 fixed

```
lmer(formula = mathach ~ meanses + (1 | schid), data = hsb)
            coef.est coef.se
(Intercept) 12.65     0.15
meanses      5.86     0.36

Error terms:
 Groups   Name        Std.Dev.
 schid    (Intercept) 1.62
 Residual             6.26
---
number of obs: 7185, groups: schid, 160
AIC = 46969.3, DIC = 46957
deviance = 46959.1
```

Interpretation

Because MEANSES is centred around the grand mean, the constant (γ_{00}) is the mean math achievement for a school with an average level of MEANSES. The other fixed effect is interpreted as an ordinary regression parameter; it is the effect of school mean socioeconomic status on the math achievement scores of students.

The random effects are interpreted as before. Notice that the estimate for $\tau_{00} = 2.64$ is now much smaller than before (it was 8.61), suggesting that a good part of the observed school-to-school variation in math achievement scores is due to variation in school mean socioeconomic status.

$$\frac{8.61 - 2.64}{8.61} = .69$$

so we have explained .69 of the observed variation in math achievement scores across schools. The remaining intraclass correlation is $2.64/(2.64 + 39.16) = .06$.

Random slope

This time we add an individual (level-1) variable, SES. We allow the effect of SES to vary across schools:

$$
\begin{aligned}
Y_{ij} &= \beta_{0j} + \beta_{1j}(\mathrm{SES}_{ij} - \bar{\mathrm{SES}}_j) + \epsilon_{ij} \\
\beta_{0j} &= \gamma_{00} + u_{0j} \\
\beta_{1j} &= \gamma_{10} + u_{1j}
\end{aligned}
$$

$$
\text{where } \epsilon_{ij} \sim N(0, \sigma^2) \text{ and}
$$

$$
\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right]
$$

This can be written as:

$$
\begin{aligned}
Y_{ij} = &\{\gamma_{00} + \gamma_{10}(\mathrm{SES}_{ij} - \bar{\mathrm{SES}}_j)\} \\
&+ \{u_{0j} + u_{1j}(\mathrm{SES}_{ij} - \bar{\mathrm{SES}}_j) + \epsilon_{ij}\}
\end{aligned}
$$

Random slope

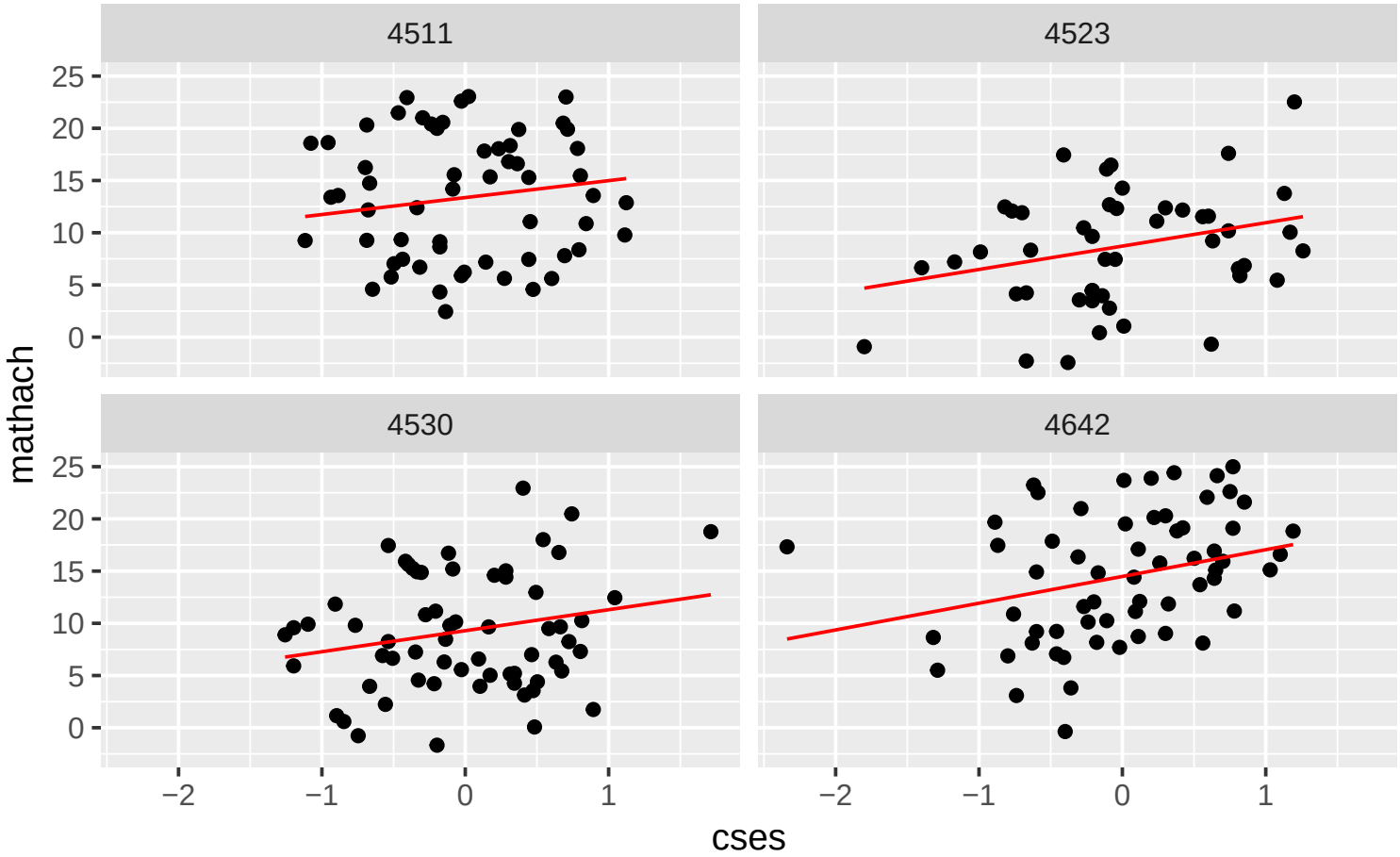
```
lmer(formula = mathach ~ cses + (1 + cses | schid), data = hsb)
             coef.est coef.se
(Intercept) 12.65     0.24
cses         2.19     0.13

Error terms:
 Groups   Name        Std.Dev. Corr
 schid    (Intercept) 2.95
          cses        0.83     0.02
 Residual             6.06
---
number of obs: 7185, groups: schid, 160
AIC = 46726.2, DIC = 46708
deviance = 46711.0
```

Interpretation

The fixed effect for cses is the estimated effect of pupil socioeconomic status on his/her math achievement score. The centering implies that the constant is the mean math achievement of schools for the average pupil (cses $= 0$).

The random effects tell us how much the slopes and intercepts vary across schools. The estimate of $\tau_{00} = 8.68$ is the variability of the intercepts, $\tau_{11} = .69.$ is the variability of the slopes, and $\tau_{01} = \tau_{10} = .051$.

Example for four schools



Level-1 and Level-2 variables

Now we add both level-1 (CSES) and level-2 (MEANSES, SECTOR) variables together:

$$
\begin{aligned}
Y_{ij} &= \beta_{0j} + \beta_{1j}(\mathrm{SES}_{ij} - \bar{\mathrm{SES}}_j) + \epsilon_{ij} \\
\beta_{0j} &= \gamma_{00} + \gamma_{01}\mathrm{MEANSES}_j + \gamma_{02}\mathrm{SECTOR}_j + u_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}\mathrm{MEANSES}_j + \gamma_{12}\mathrm{SECTOR}_j + u_{1j}
\end{aligned}
$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$ and

$$
\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right]
$$

This can be written as:

$$
\begin{aligned}
Y_{ij} = \gamma_{00} &+ \gamma_{01}\mathrm{MEANSES}_j + \gamma_{02}\mathrm{SECTOR}_j \\
&+ \gamma_{10}(\mathrm{SES}_{ij} - \bar{\mathrm{SES}}_j) \\
&+ \gamma_{11}\mathrm{MEANSES}_j(\mathrm{SES}_{ij} - \bar{\mathrm{SES}}_j) \\
&+ \gamma_{12}\mathrm{SECTOR}_j(\mathrm{SES}_{ij} - \bar{\mathrm{SES}}_j) \\
&+ u_{0j} + u_{1j}(\mathrm{SES}_{ij} - \bar{\mathrm{SES}}_j) + \epsilon_{ij}
\end{aligned}
$$

As can be seen, there are five fixed effects and three random effects.

Level-1 and Level-2 variables

```
lmer(formula = mathach ~ meanses * cses + schtype * cses + (1 +
    cses | schid), data = hsb)
                    coef.est coef.se
(Intercept)          12.11     0.20
meanses               5.34     0.37
cses                  2.94     0.16
schtypeprivate        1.22     0.31
meanses:cses          1.04     0.30
cses:schtypeprivate -1.64     0.24

Error terms:
 Groups   Name        Std.Dev. Corr
 schid    (Intercept) 1.54
          cses        0.32     0.39
 Residual             6.06
---
number of obs: 7185, groups: schid, 160
AIC = 46523.7, DIC = 46489
deviance = 46496.4
```

Multilevel GLMs

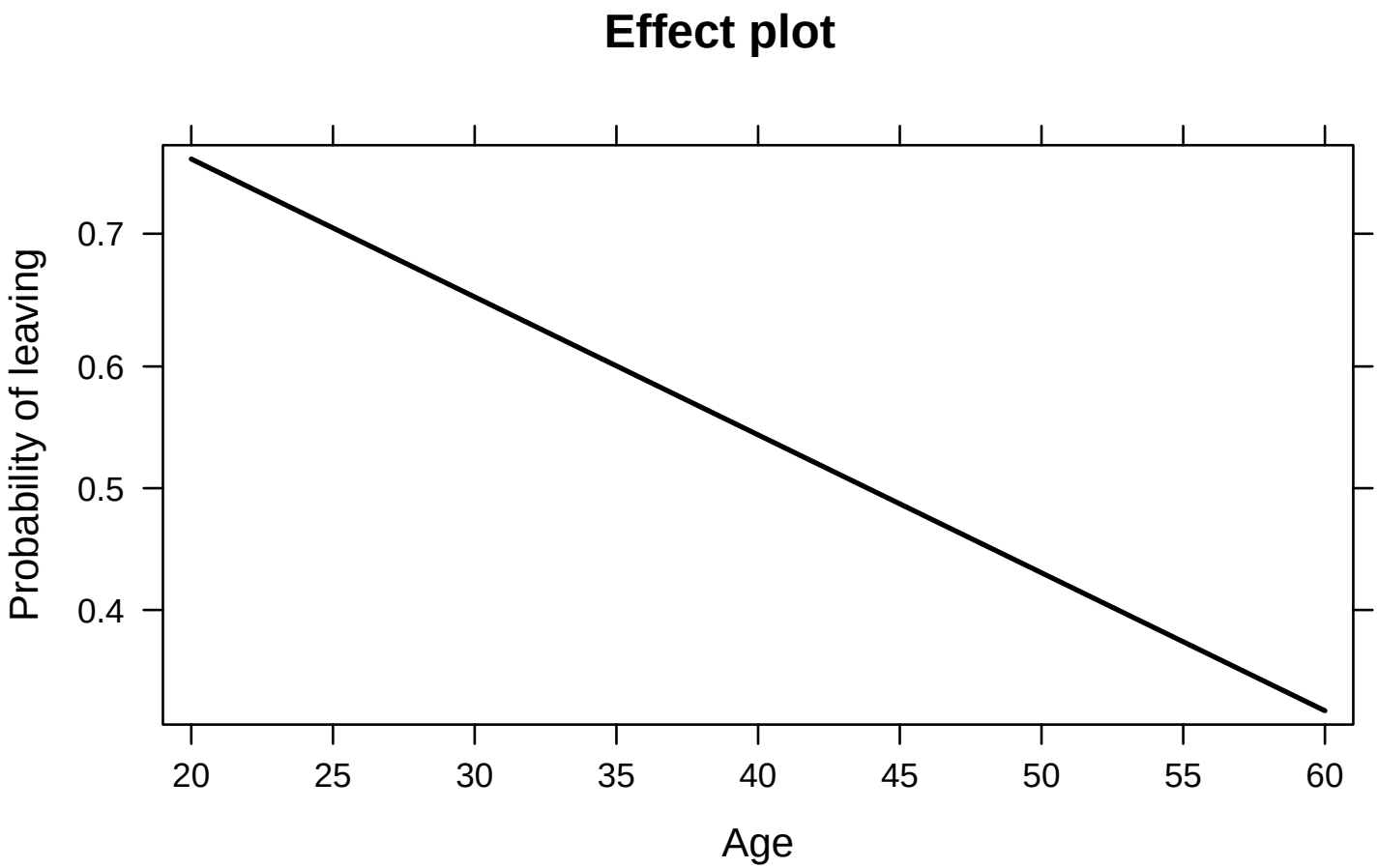
Multilevel GLMs

Multilevel models aren't restricted to linear regression; exactly the same logic applies to logistic, Poisson, etc. In this example the outcome variable comes from a survey of nurses in London hospitals. There are 1,389 nurses in 18 hospitals. The outcome variable is whether the nurse plans to leave their job in the next 3 years.

Example

```
glmer(formula = leave ~ fam + valued + age + sex + INN_OUT +
    (1 | HOSPITAL), data = nurse, family = binomial, subset = age >=
    18)
                    coef.est coef.se
(Intercept)          3.02     0.27
famyes              -0.75     0.13
valuedyes           -0.93     0.13
age                 -0.05     0.01
sexmale              0.28     0.21
INN_OUTOuter London -0.73     0.15

Error terms:
 Groups   Name        Std.Dev.
 HOSPITAL (Intercept) 0.18
 Residual             1.00
---
number of obs: 1385, groups: HOSPITAL, 18
AIC = 1678.9, DIC = 1639
deviance = 1652.0
```

Effect plot


Effect plot
Probability of leaving
0.7
0.6
0.5
0.4
20
25
30
35
40
45
50
55
60
Age