

Random Effects Poisson Models

Examples of random coefficient poisson models

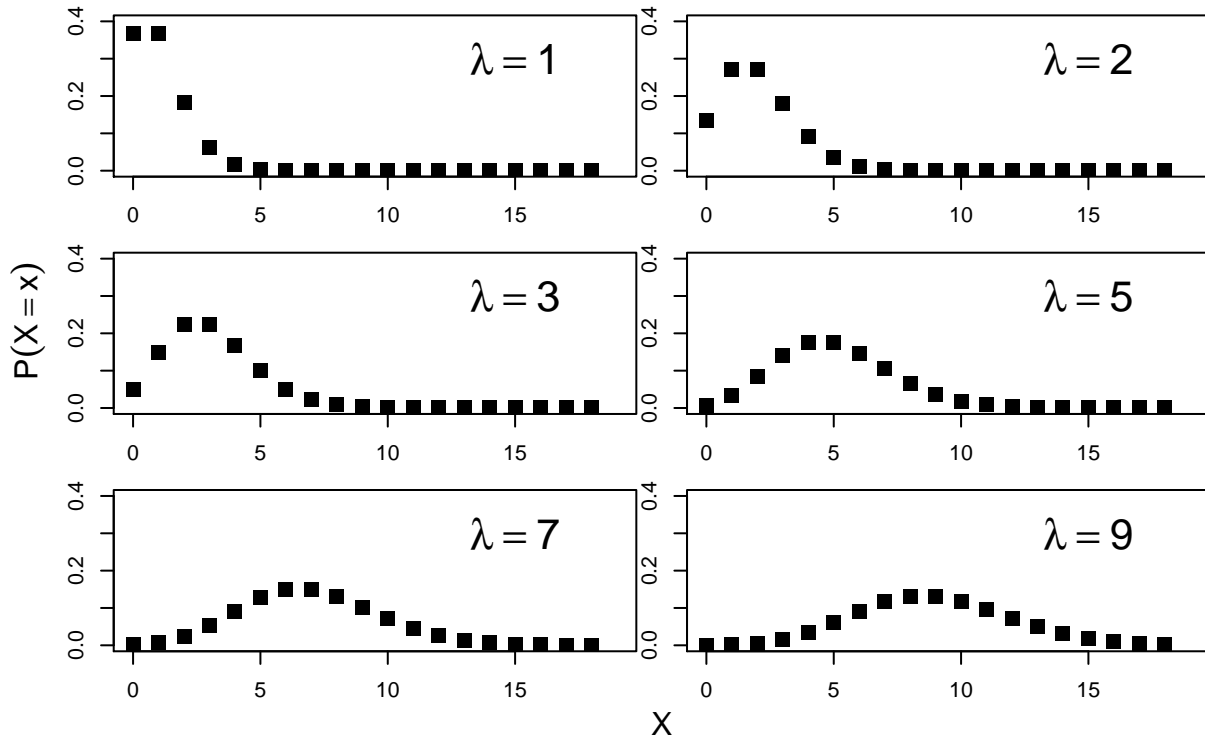
Poisson models are useful for count data. Examples could include number of traffic tickets an individual receives in a year, number of tumor sites in cancer patients, or number of awards received by students. In each of these cases, we might expect most people to have very few, with a relatively small number of individuals having higher numbers.

Background on the Poisson distribution

Unlike the familiar Gaussian distribution which has two parameters (μ , σ^2), the Poisson distribution is described by a single parameter, (λ) that is both the mean and variance. That is, ($\lambda = E(x)$) and ($\lambda = \text{Var}(x) = E(x^2) - E(x)^2$). In many ways, this can be a strong assumption. In some cases, use of the negative binomial distribution which estimates a second scale parameter to allow for overdispersion. Here is what the Poisson distribution looks like for different values of (λ).

```
par(mfrow = c(3, 2), mar = c(1, 1, 2, 1), oma = c(4, 4, 2, 0))
for (i in c(1, 2, 3, 5, 7, 9)) {
  curve(dpois(x, lambda = i), from = 0, to = 18, n = 19, type = "p", pch = 15,
        cex = 1.5, xlim = c(0, 19), ylim = c(0, 0.4), xlab = "", ylab = "")
  text(x = 15, y = 0.3, substitute(lambda == x, list(x = i)), cex = 2)
}
mtext("Probability Mass Function for Poisson Distribution", line = 0.5,
      outer = TRUE, cex = 1.2)
mtext(expression(P(X == x)), side = 2, line = 1.5, outer = TRUE)
mtext("X", side = 1, line = 1.5, outer = TRUE)
```

Probability Mass Function for Poisson Distribution



Without random coefficients, the standard Poisson model is:

$$\log E(y_i) = \alpha + X_i' \beta$$

The log link is the canonical link function for the Poisson distribution, and the expected value of the response is modeled.

With random coefficients, for example a random intercept, the model becomes:

$$\log E(y_{ij}|u_j) = \alpha + X_{ij}' \beta + u_j$$

Where y_{ij} is the observation for individual (i) in group (j) and u_j is the random effect for group (j). Thus the two distributions are:

$$y \sim Pois(\lambda)$$

and

$$u \sim N(0, \sigma^2)$$

The random coefficient model is conditional on the random effect. To show what this means, consider a simple model:

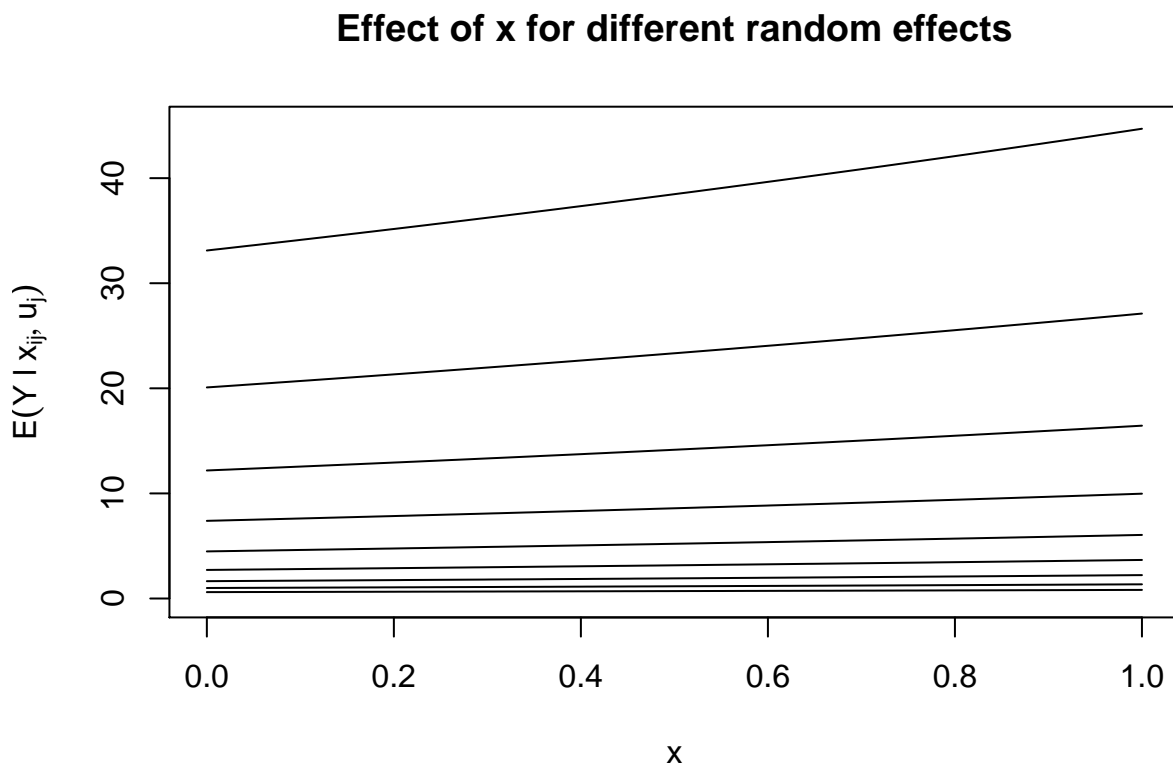
$$\log E(y_{ij}|u_j) = -.5 + .3x_{ij} + u_j$$

In the original units, this becomes:

$$E(y_{ij}|u_j) = \exp(-.5 + .3x_{ij} + u_j)$$

Now look what happens when we graph the estimated change for a 1 unit change in x for values of the random variable (u) ranging from 0 to 4 by increments of .5.

```
par(mfrow=c(1,1))
f <- function(x_ij, u_j) {
  exp(-0.5 + x_ij * 0.3 + u_j)
}
for (i in seq(0, 4, 0.5)) {
  curve(f(x, u_j = i), from = 0, to = 1, n = 200, add = i > 0, ylim = c(0,
    45), ylab = "")
}
title(ylab = bquote(E(Y ~ 1 ~ x[ij], u[j])), main = "Effect of x for different random effects")
```



Clearly, on the scale of the original units, a 1 unit increase in x has different effects depending on the value of u, hence the conditionality of the model. Population ???average??? effects can be obtained by integrating out the random effect or by fitting a marginal model such as using GEEs.

Although the outcome is assumed to have a Poisson distribution, the random effect (in the above example, u) is typically assumed to have a Gaussian distribution.

Description of the data The example data we use comes from a sample of the high school and beyond data set, with a made up variable, number of awards a student receives, awards. Our main predictor will be sex, female, and students are clustered (grouped) within schools, cid. First, we will load all the packages required for these examples.

```

## load foreign package to read Stata data files
require(foreign)

## Loading required package: foreign
## ggplot2 package for graphs
require(ggplot2)

## Loading required package: ggplot2

You need to install the glmmADMB() package via instructions given in http://glmmadmb.r-forge.r-project.org/. Note that if you are not able to install the glmmADMB() package you may need to use Rtools and compile from source, use Linux, or ask the authors of the package for more information.

install.packages("R2admb", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/waves/Documents/R/win-library/3.4'
## (as 'lib' is unspecified)

## package 'R2admb' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\waves\AppData\Local\Temp\RtmpUtlJQG\downloaded_packages
install.packages("glmmADMB", repos="http://glmmadmb.r-forge.r-project.org/repos", type="source")

## Installing package into 'C:/Users/waves/Documents/R/win-library/3.4'
## (as 'lib' is unspecified)

require(glmmADMB)

## Loading required package: glmmADMB
## Loading required package: MASS
##
## Attaching package: 'glmmADMB'

## The following object is masked from 'package:MASS':
##
##     stepAIC

## The following object is masked from 'package:stats':
##
##     step

## load lme4 package
require(lme4)

## Loading required package: lme4
## Loading required package: Matrix
##
## Attaching package: 'lme4'

## The following object is masked from 'package:glmmADMB':
##
##     VarCorr

## read in data
dat <- read.dta("https://stats.idre.ucla.edu/stat/data/hsbdemo.dta")
dat$cid <- factor(dat$cid)

```

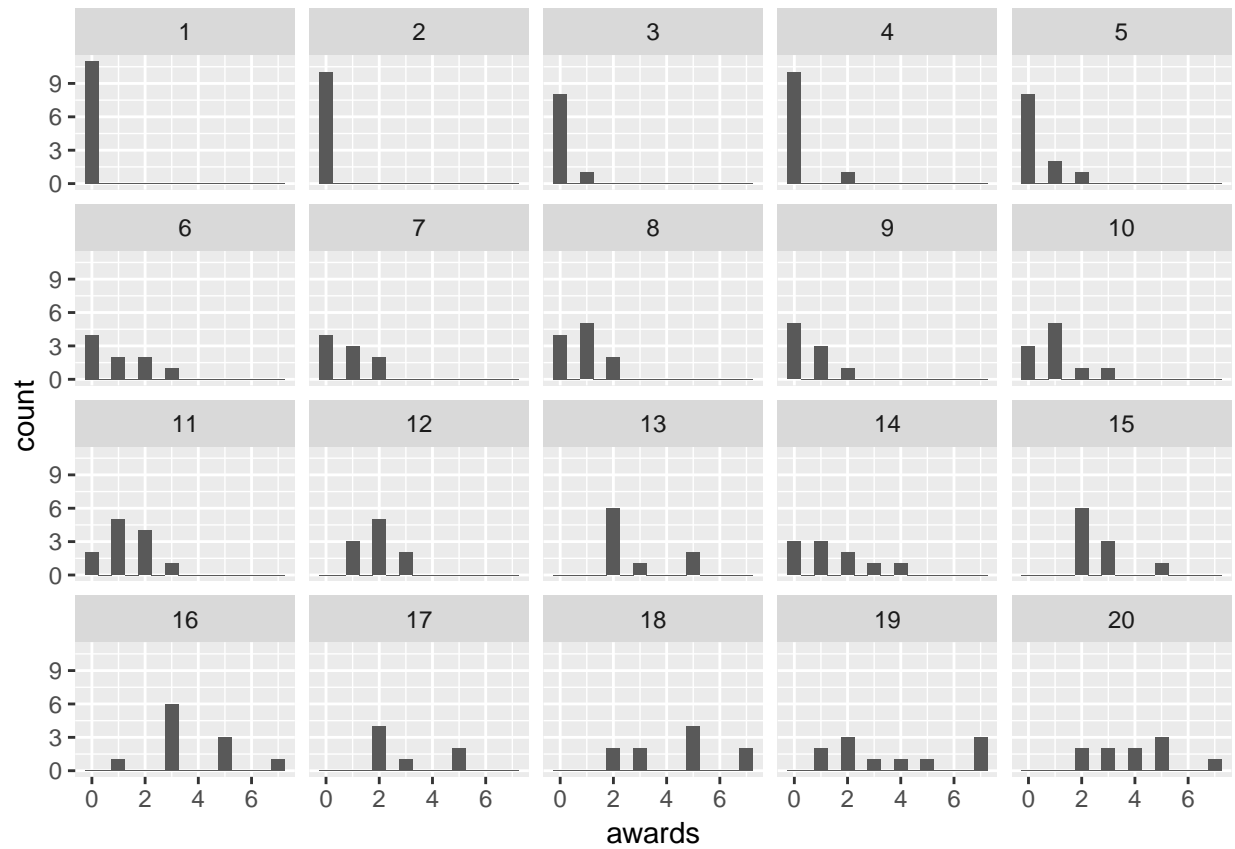
```
## look at the first few rows of the dataset
head(dat)
```

```
##   id female   ses schtyp   prog read write math science socst
## 1  45 female   low public vocation  34   35  41     29    26
## 2 108  male middle public  general  34   33  41     36    36
## 3  15  male   high public vocation  39   39  44     26    42
## 4  67  male   low public vocation  37   37  42     33    32
## 5 153  male middle public vocation  39   31  40     39    51
## 6  51 female   high public  general  42   36  42     31    39
##      honors awards cid
## 1 not enrolled      0  1
## 2 not enrolled      0  1
## 3 not enrolled      0  1
## 4 not enrolled      0  1
## 5 not enrolled      0  1
## 6 not enrolled      0  1
```

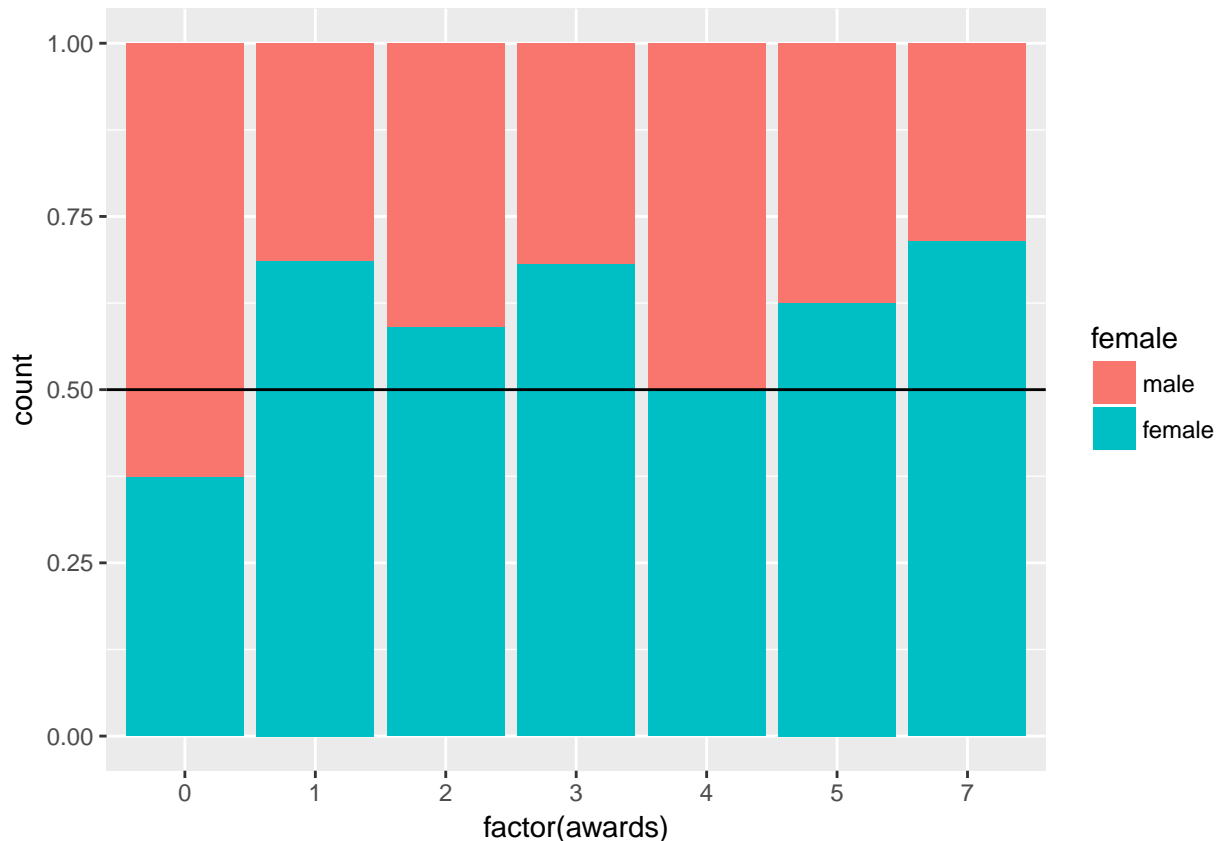
We can get a sense of the distributions in the data using the ggplot2 package. The first plot is just histograms of number of awards for every cid. The second is a filled density plot. The density sums to 1 and the fill shows the distribution of female at every level of awards. If the distribution of female is equal across all awards, they would fall on the horizontal line.

awards by school

```
ggplot(dat, aes(awards)) + geom_histogram(binwidth = 0.5) + facet_wrap(~cid)
```



```
## density of awards by sex, line at .5 is the null of no sex differences
## in number of awards
ggplot(dat, aes(factor(awards))) + geom_bar(aes(fill = female), position = "fill") + geom_hline(yintercept = 0.5)
```



Analysis methods you might consider Random coefficient poisson models, the focus of this page. Poisson regression with robust standard errors Random coefficient poisson model analysis Because generalized linear mixed models (GLMMs) such as random coefficient poisson models are rather difficult to fit, there tends to be some variability in parameter estimates between different programs. We will demonstrate the use of two packages in R that are able to fit these models, lme4 and glmmADMB.

```
## fit a random intercept only model using the Laplace approximation
## (equivalent to 1 point evaluated per axis in Gauss-Hermite
## approximation)
m1a <- glmer(awards ~ 1 + (1 | cid), data = dat, family = poisson(link = "log"))
## fit a random intercept only model using 100 points per axis in the
## adaptive Gauss-Hermite approximation of the log likelihood more points
## improves accuracy but will take longer
m1b <- glmer(awards ~ 1 + (1 | cid), data = dat, family = poisson(link = "log"), nAGQ = 100)

## compare (only slightly different)
rbind(m1a = coef(summary(m1a)), m1b = coef(summary(m1b)))
```

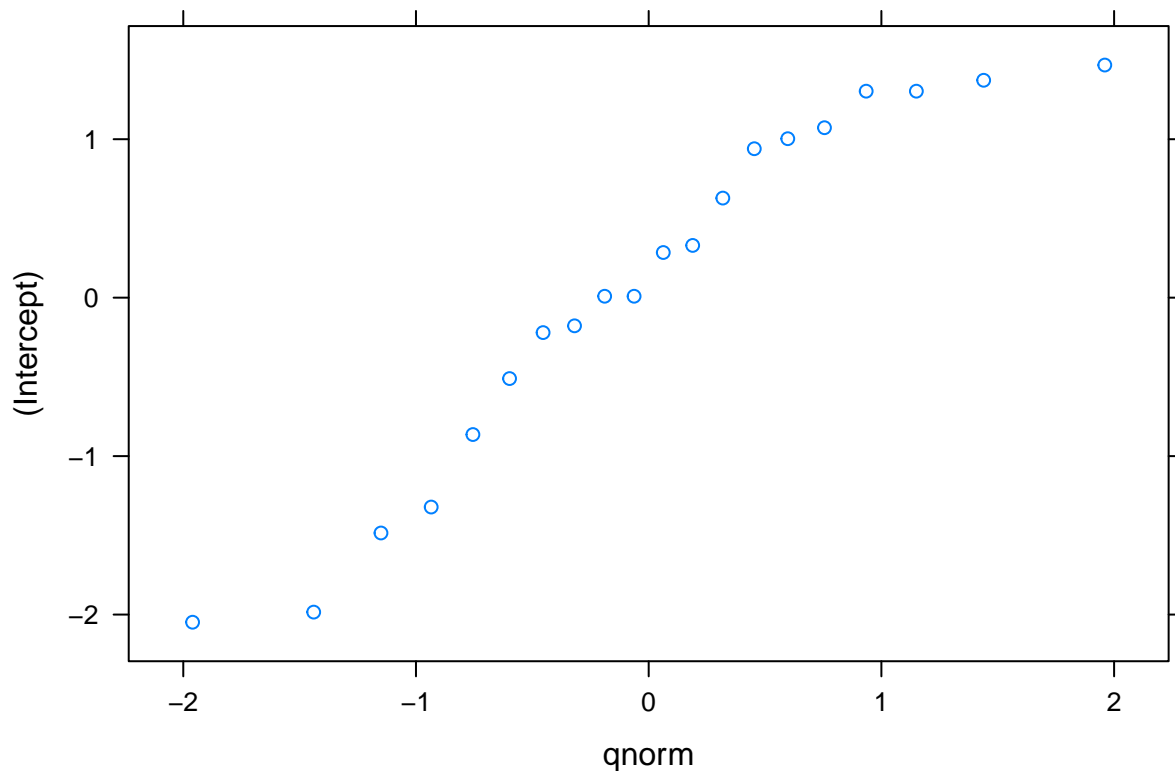
```
##               Estimate Std. Error    z value Pr(>|z|)
## (Intercept) -0.008100387  0.2864037 -0.02828311 0.9774364
## (Intercept) -0.009574892  0.2902939 -0.03298344 0.9736878
```

```
summary(m1b)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 100) [glmerMod]
## Family: poisson ( log )
## Formula: awards ~ 1 + (1 | cid)
```

```
## Data: dat
##
##      AIC      BIC   logLik deviance df.resid
##    228.6    235.2   -112.3    224.6     198
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.3857 -0.5260 -0.3383  0.3379  3.3769
##
## Random effects:
##   Groups Name      Variance Std.Dev.
##   cid      (Intercept) 1.458    1.207
## Number of obs: 200, groups:  cid, 20
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.009575   0.290294  -0.033   0.974
## QQ plot
plot(ranef(m1b))

## $cid
```

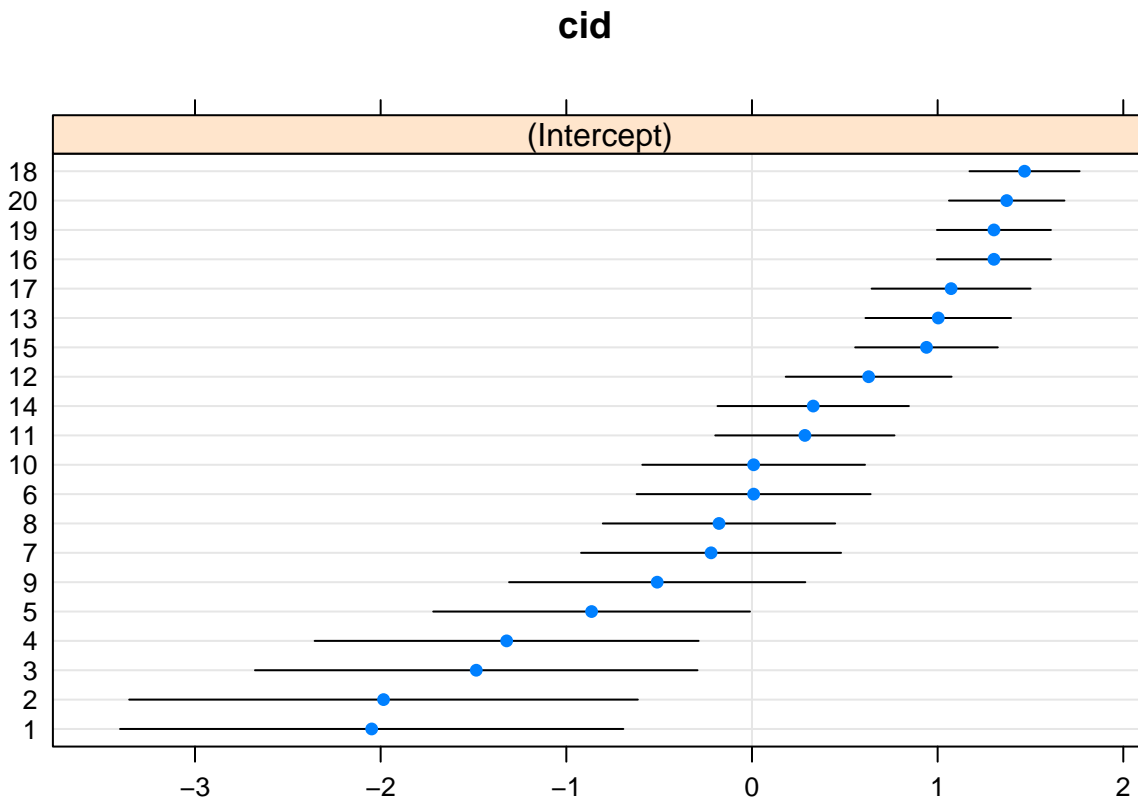


```
## $cid

## Caterpillar plot
lattice::dotplot(ranef(m1b, postVar = TRUE))
```



```
## Warning in ranef.merMod(m1b, postVar = TRUE): 'postVar' is deprecated:
## please use 'condVar' instead
## $cid
```



```
## $cid
```

The estimate for the intercept is essentially 0, although the random effects variance indicates that there is some variability in the intercepts between schools. Now we will add in female as an explanatory variable.

```
m2 <- glmer(awards ~ 1 + female + (1 | cid), data = dat, family = poisson(link = "log"), nAGQ = 100)
summary(m2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 100) [glmerMod]
## Family: poisson ( log )
## Formula: awards ~ 1 + female + (1 | cid)
## Data: dat
##
##      AIC      BIC    logLik deviance df.resid
##    221.1    231.0   -107.6    215.1     197
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.5312 -0.5919 -0.3304  0.2047  2.8806
##
## Random effects:
## Groups Name      Variance Std.Dev.
```

```
## cid      (Intercept) 1.431    1.196
## Number of obs: 200, groups:  cid, 20
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.2229    0.2975  -0.749  0.45370
## femalefemale  0.3632    0.1193   3.044  0.00234 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr)
## femalefemal -0.252
```

There appears to be a fairly strong effect of females such that females tend to get more awards than males. Now we will fit the same models using the glmmADMB package

```
## random intercept only model
library(glmmADMB)
m.alt1 <- glmmadmb(awards ~ 1 + (1 | cid), data = dat, family = "poisson", link = "log")
m.alt2 <- glmmadmb(awards ~ 1 + female + (1 | cid), data = dat, family = "poisson", link = "log")
summary(m.alt1)

##
## Call:
## glmmadmb(formula = awards ~ 1 + (1 | cid), data = dat, family = "poisson",
##          link = "log")
##
## AIC: 575.1
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.00881    0.28902  -0.03    0.98
##
## Number of observations: total=200, cid=20
## Random effect variance(s):
##
## Warning in .local(x, sigma, ...): 'sigma' and 'rdig' arguments are present
## for compatibility only: ignored
##
## Group=cid
##              Variance StdDev
## (Intercept)   1.443  1.201
##
##
## Log-likelihood: -285.528
```

The results from glmmadmb match closely with those from glmer.