Why Linear Mixed Effects Models?
Reproductive output of a rare plant:
random effects of populations

In demography it is habitual to estimate reproductive output of individuals as a component together with other vital rates in population models (Quintana-Ascencio et al., 2003, 2018, 2019). We could use *Hypericum cumulicola* data to evaluate a model to predict number of reproductive structures of individuals with different heights as if individuals were independent across populations. However, we recognize that plants in the same population are likely to be more like to each other than (and consequently different) from those in other populations. These types of random effects should be considered to avoid pseudo-replication, improve our predictions and provide generality to our results. Here, we introduce a method to incorporate hierarchical effects in our models.

Figure 1. *Hypericum cumulicola in the scrub*

McElreath (2016) suggests approaching this problem about predicting from three perspectives:

1. Complete pooling. We assume that the effect of populations on individuals is invariant, same as estimating a common intercept.
2. No pooling. We assume that each population tell us nothing about any other. "This is the model with amnesia".
3. Partial pooling. We use an adaptive regularizing prior. It helps to improve the estimates for populations with small sample size.

We call the libraries of two packages that we will need during the analysis and then continue by preparing the data

```
orig_data <- read.table("hypericum_data_94_07.txt", header=T)
dt <- subset(orig_data, !is.na(ht_init) & rp_init > 0 & year < 1997)
yr <- unique(dt$year)
```

```
dt$height <- dt$ht_init
dt$fruits <- dt$rp_init
dt$id <- dt$tag
```

We estimate the coefficients of the model assuming complete independence of the populations and plot the model (in blue) in Figure 2.

1. ***Complete independence***

$$log(reproductive\ structures)_k = \beta_1 + \beta_2 * log(height)_k \qquad Likelihood{\sim}N(\mu,\sigma)$$

```
sdlnfruits <- sd(dt$lfr)

m1 <- map2stan(
  alist(
    lfr ~ dnorm(mu,sigma),
    mu <- a + b*lgh,
    a ~ dnorm(0,100),
    b ~ dnorm(0,10),
    sigma ~ dunif(0,10)
  ),
  data = dt,chains =3,
  start <- list(a = 10,b= 1, sigma=sdlnfruits)
)
```
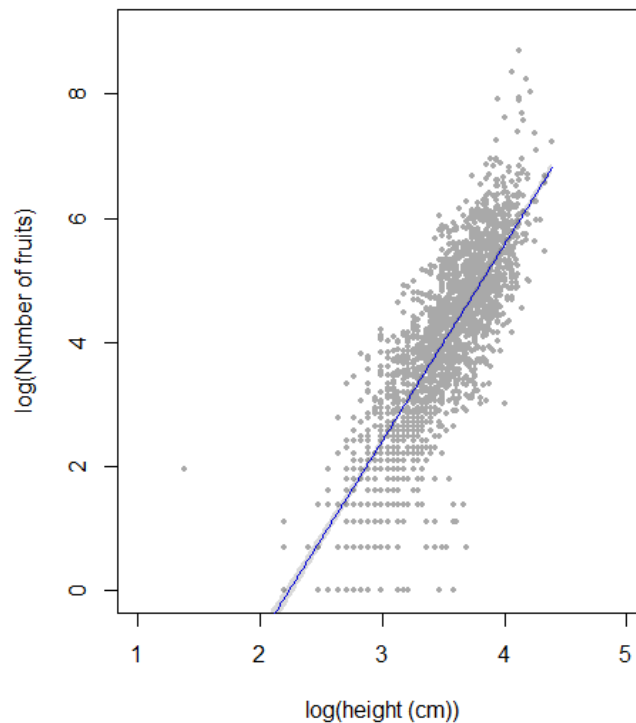
The output of the general model under the assumption of complete independence should be familiar. This model explains approximately 66 % of the variance.

```
> precis(m1,digits=1)
      Mean StdDev lower 0.89 upper 0.89 n_eff Rhat
a     -7.34   0.18      -7.4       -6.8   804    1
b      3.24   0.05       3.1        3.3   802    1
sigma  0.82   0.01       0.8        0.8  1224    1
```
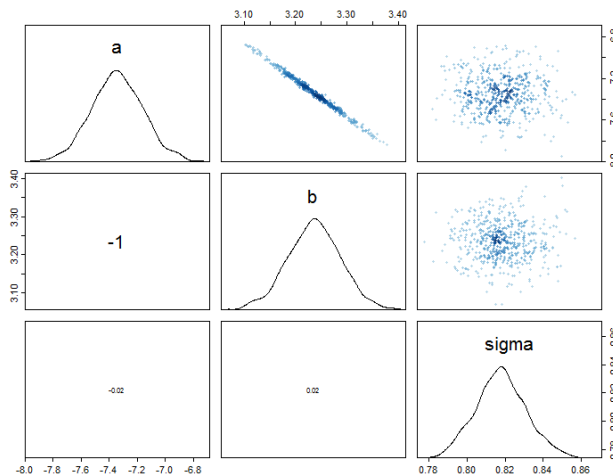

**Figure 2.** Posterior distributions of the parameters and their correlation.

Figure 3. Plot of log(height) vs log(fruits), data as points in grey, overall model assuming independence in blue.

We now estimate the coefficients of the model of each population. We plot the models for each of the 14 populations studied in Figure 4. We notice that there is variation for the models for each of these populations. We could model the log (number of fruits) as a linear function of the log (height) using these specific intercepts (*Beta 1*) and slopes (*Beta 2*) for each population, but this will significantly reduce the degrees of freedom and limit the generality of our interpretation. The coefficients per model by population are listed in Table 1. The code can be found in the file LMM 2019.R

Population specific formulation. Observe that the indices of the coefficients are population specific.

$$log(reproductive\ structures)_i = \beta_{1_i} + \beta_{2_i} * log(height)_k \qquad Likelihood \sim N(\mu, \sigma)$$
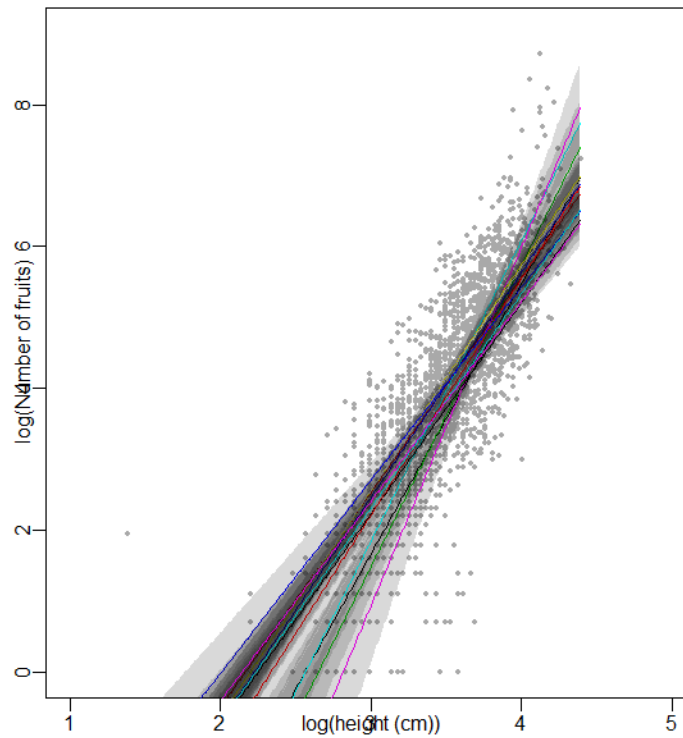
Figure 5. Plot of log(height) vs log(fruits), data as points in grey, models per each population in different colors.

Table 1. Coefficients estimated per population.

Population	1	29	32	42	50	57	59	62	67	87	88	91	93	103
Beta1	-9.8	-6.5	-11.3	-7.0	-10.9	-14.1	-7.0	-6.9	-6.7	-7.8	-9.7	-5.5	-6.8	-6.0
Beta2	3.82	3.02	4.3	3.17	4.2	5.0	3.2	3.1	2.98	3.34	4.18	2.72	3.02	2.80
Sigma	0.73	0.72	0.76	0.65	0.79	1.34	0.67	0.63	0.56	0.82	0.68	0.80	0.76	0.77
N	91	141	179	169	189	94	174	135	122	129	140	183	206	149

Notice that there is one population (57) with extreme deviation probably because peculiar attributes or smaller sample size. If we are interested in more general inference (and better use of the data), we could instead estimate the variation around the coefficients (the intercept or both the intercept and slope) and assume that they are normally distributed. In this way we will inform the estimate of one population with that from other populations. For more details on these assumptions see Zuur et al. (2009) and McElreath (2016).

We start assuming random intercepts by population but a common slope.

$$log(reproductive\ structures)_k = \beta_1 + \alpha_{1_i} + \beta_2 * log(height)_k$$

$$\alpha_{1_i} \sim N(0, \sigma_1)$$

$$Likelihood \sim N(\mu, \sigma)$$

We plot this model (Figure 4). The line in red is the model obtained with the fixed (average) coefficients. The lines in blue represent the variation estimated by population as their displacement from the average curve. The random intercept models are curves that shift by a factor that is normally distributed with a given variance. If the deviation from the average is larger the shift is greater.
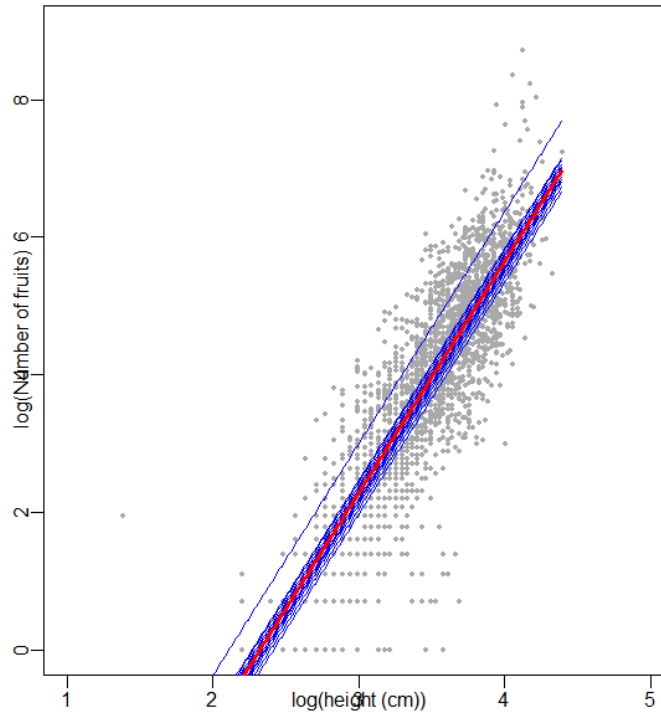
Figure 6. Plot of log(height) vs log(fruits), data as points in grey, models with random intercept per each population in blue. Average model in red.

Table 2. Coefficients estimated for mixed model with variable intercepts. Fixed coefficients: Beta1 = -7.48, Beta2 =3.29, sigma intercept =0.29, sigma = 0.78.

Population	1	29	32	42	50	57	59	62	67	87	88	91	93	103
Beta1	-8.1	-7.9	-7.1	-7.8	-8.0	-7.9	-7.7	-7.9	-7.7	-7.8	-7.9	-7.6	--7.6	-8.0

We can now try a model that estimates random intercepts and slopes. This is specified in the model

$$log(rep\ struct)_k = \beta_1 + \alpha_{1_i} + [\beta_2 + \partial_{2_i}] * log(height)_k$$

$$\alpha_{1i} \sim N(0, \sigma_1)$$

$$\partial_{1i} \sim N(0, \sigma_2)$$

$$Likelihood \sim N(\mu, \sigma)$$

This is the most likely model of the set we evaluate here, and it is presented in Figure 6 and Table 3. Models are compared in Table 4. Notice that this model does not evaluates the possible spatial correlation between populations. We will explore this issue in another demo.
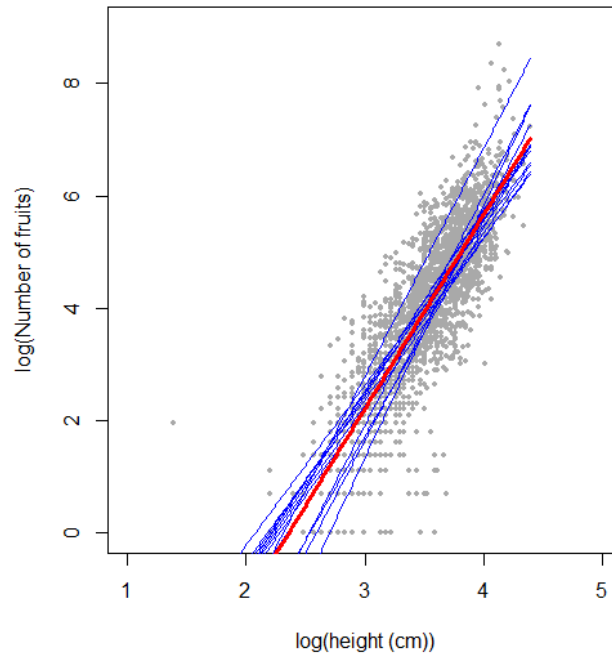
Figure 6. Plot of log(height) vs log(fruits), data as points in grey, models with random intercept per each population in blue. Average model in red.

Table 3. Coefficients estimated for mixed model with variable intercepts and slopes. Fixed coefficients: Beta1 = -8.17, Beta2 =3.46, sigma intercept =2.05, sigma slope = 0.60, sigma = 0.76.

Population	1	29	32	42	50	57	59	62	67	87	88	91	93	103
Beta1	-9.5	-7.9	-9.4	-5.91	-6.9	-6.3	-6.8	-10.5	-7.2	-10.4	-12.3	7.17	-7.1	-7.0
Beta2	3.7	3.4	4.1	2.8	3.1	2.9	3.1	4.0	3.2	4.1	3.5	3.2	3.2	3.1

Table 4. Comparison of the average coefficients of the population level for the three approaches (only fixed effects for mixed models), and their WAICs. The model with random intercept and random slope was the most informative of the set .

	Intercept	s.e	Slope	s.e	WAIC
no mixed	-7.1	0.2	3.2	0.1	5118
random slope	-7.3	0.2	3.3	0.05	4953
random intercept & slope	-8.1	0.6	3.5	0.2	4845

NOTE: all the materials for this demo can be found at:

https://sciences.ucf.edu/biology/d4lab/methods-2/

References

McElreath, R.M. 2016. Statistical Rethinking: a Bayesian course with examples in R and Stan.Chapman and Hall.

Quintana-Ascencio, P. F., E. S. Menges, and C. Weekley. 2003. A fire-explicit population viability analysis of Hypericum cumulicola in Florida rosemary scrub. Conservation Biology, 17: 433-449.

Quintana-Ascencio, P.F. Koontz, S., Smith, V., David, A., Sclater, V. L. & E. S Menges. 2018. Predicting landscape-level distribution and abundance: Integrating demography, fire, elevation, and landscape habitat configuration. Journal of Ecology, 106: 2395-2408

Quintana-Ascencio, P.F. Koontz, S.M., Ochocki, B., Sclater, V. L., López-Borghesi, F., Li, H. & E. S Menges. 2019. Assessing the roles of seed bank, seed dispersal and historical disturbances for metapopulation persistence of a pyrogenic herb. Journal of Ecology, 107: 2760-2771.

Zuur, A, J.M. Hilbe and E N. Leno. 2015. A beginner's guide to GLM and GLMM with R. Highland Statistics, Ltd.