***Random-slope regression***

***Goal:*** To explore situations where individuals may not only differ in their intercepts but also in the slope of their reaction to changes in a given environmental variable. Random slope regression is a great tool to analyse among-individual differences in phenotypic plasticity.

**Step 1: Introduction to random slopes**

**Sub-goal**: to develop understanding of hierarchies in variance when individuals express their traits as a response to a changing environmental factor, but do so differently. In other words, individuals have variable reaction norms in both intercept and slope.

**Introduction**: Modules “*Basic Lessons about Variance”* and “*Non-random Environments”* have explored how among-individual (VI) variance in mean values can be modelled. In Step 3 of module “*Basic Lessons about Variance”*, individuals also varied, to some extent, in how an environmental factor influenced the expression of their phenotypes around this mean, and these deviations produce within-individual variance as a response to environmental fluctuations (Vβx). Throughout all these lessons, we have assumed that if the environment changes, individuals change their phenotype in exactly the same way. This is a dubious assumption and in most biological situations, it is more reasonable and often more interesting to account for the fact that individuals might vary in how the environment affects phenotype. For example, individuals may differ in how they cope with a stressful situation, and less tolerant individuals will alter their activity to a greater degree in the presence of a stress factor (e.g. high temperature, presence of a predator, or shortage of food) than stress-tolerant conspecifics. In general, among-individual differences in plasticity (also called individual by environment interaction; IxE) are necessary if adaptive plasticity is to be under selection, and insofar as such differences have a genetic basis (GxE), such interactions are necessary for the evolution of plasticity. Studies of IxE in natural populations remain relatively rare. Here we introduce random-slope regression, in which the assumption about a common within-individual variance (and plasticity) is relaxed and we explicitly estimate among-individual variance in slope.

**Exercise:**

Here we depart from the random-intercept regression model presented in Step 4 of module “*Basic Lessons about Variance”* by introducing a parameter that allows defining individual deviations from the population mean response to environmental changes. The fundamental difference between the random-intercept regression and the random-slope model, is that while the former uses individual-level “random” effects to model individual-specific intercepts (i.e. that depict the among-individual variance of mean trait values), the latter introduces an additional “random” effect on the slopes (i.e. that describe how individuals respond to changing environments). For the random-intercept regression, we used the following model to recreate our simulated set of effects:

**R code:**

# install.packages("lme4")

LMM1 <- lme4::lmer(Phenotype ~ 1 + X1 + (1|Individual), data = sampled\_data)

We will modify this model as:

**R code:**

LMM2 <- lme4::lmer(Phenotype ~ 1 + X1 + (1|Individual) + (0 + X1|Individual), data = sampled\_data)

in which is the effect of measured environment on the measure of phenotype in the *i*th individual. Accordingly, we can apply the following parameterization regimes:

Number of individuals: …….

Number of trait expressions: …….

Among-individual variance (VI): …….

Residual variance (VR): …….

Note: you can consider residual variance as solely measurement error variance, or you could consider that it might also include some unknown environmental variance in addition to the known environmental variance determined below.

Variance due to population mean effect of an environmental factor ():…….

Variance due to individual-specific responses to an environmental factor (random slopes; VSx) :

…….

Set in the background is the environmental variable and its sampling. Here we should use uniform sampling, where each individual is sampled at the same time so there are no biases. By definition, our environmental variable is also mean centred.

RUNNN

**Results**:

A mixed-effects statistical model estimates these parameters:

|  |  |
| --- | --- |
| True | Estimated |
| Fixed effects | Fixed effects |
| Mean of the trait (β0) = 0 | Sampled mean of the trait (β’0) = ….. |
| Population-specific slope of the environmental effect (β) = ….. | |  | | --- | | Estimated population-specific slope of the environmental effect (β’) = ….. | |
| Random effects | Random effects |
| Among-individual variance in intercept (VI) = …… | Sampled among-individual variance in intercept (V’I) = ….. |
| Variance due to individual-specific responses to an environmental factor (random slopes; VSx) = ….. | Variance due to individual-specific responses to an environmental factor (random slopes; V’Sx) = ….. |
| Residual variance (VR) = ….. | Residual variance of sample (V’R) = ….. |

Individual-specific responses can be best visualised by plotting the individual-specific regression lines.

GRAPHS

Environment vs. Phenotype per individual (fitted reaction norms + individual’s dataset)

**Point:** Again, we performed linear regression approach in a mixed model framework, but in this case we generated multiple regression lines (by defining different intercepts and slopes) for each individual in a single model to describe the individual-specific effect of an explanatory variable. Individual-specific responses to changes that occur along an environmental gradient form “reaction norms”. These reaction norms, if the environmental gradient is centred around zero, can be characterized by their intercept that describes individual mean expression values, and by their slope that expresses the plasticity of traits within individuals. Statistically, one can evaluate whether the random-intercept or the random-slope model fits the data at hand better. However, the consideration of differences in how individuals respond to environmental fluctuations may be straightforward on a biological basis. This model simultaneously accommodates tests for individual personality differences (i.e. the calculation of repeatability makes sense) as well as tests for individual by environment interaction.

**Statistical model:**

Where

Note that is the true variance in *x*, and is the true mean of *x*. Also, in SQuID each environmental variable (x) is standardized (i.e., and )

**R code:**

# install.packages("lme4")

LMM <- lme4::lmer(Phenotype ~ 1 + X1 + (1|Individual) + (0 + X1|Individual), data = sampled\_data)

**Step 2: Intercept-slope correlation**

**Sub-goal**: To introduce and understand correlation (covariance) between intercepts and slopes.

**Introduction**: Up until this point, we have been making the hidden and simplifying assumption that each variable in our variance partitioning is independent of other variables. In other words, we have assumed, for example, that measurement error is not correlated with individual intercept or environment. For the most part, we will continue to make these assumptions, but the introduction of random slopes means we have two variances, in intercepts and slopes, that are part of phenotypic variance and attributes of individuals. Biologically, it becomes very interesting if there is a correlation between intercepts and slopes. That correlation can be either positive or negative, and we will have you simulate data both ways so you can see what the sign of the correlation does to the pattern of variance within and among individuals.

We use the same equation as before, e.g.,

but now we make the formally hidden assumption explicit with some new nomenclature:

This means that the intercept value of the *i* th individual () and the slope of that individual () are distributed as multivariate normal with means for each attribute of 0 and a variance-covariance structure of . We then specify the variance-covariance structure in a 2 by 2 matrix. You have seen the variance terms ( and Var(S)) before, but the new term introduced here is , which is the measure of how much and covary together.

**Exercise**: We will do a very similar simulation as in Step 1, but you will then enter some value for the covariance between the intercept and the slope. You will enter this covariance as a standardized value by expression it as a correlation ranging between -1 and +1:

We will then examine the appearance of the resulting data.

Number of individuals: …….

Number of trait expressions: …….

Among-individual variance (VI): …….

Residual variance (VR): …….

Variance due to population mean response to an environmental factor ():…….

Variance due to individual-specific response to an environmental factor (random slopes) (VSx) :

…….

Correlation between I and S (CorIS ): ……

RUNNN

**Results**:

A mixed statistical model estimates these parameters:

|  |  |
| --- | --- |
| True | Estimated |
| Fixed effects | Fixed effects |
| Mean of the trait (β0) = 0 | Sampled mean of the trait (β’0) = ….. |
| Population-specific slope of the environmental effect (β) = ….. | |  | | --- | | Estimated population-specific slope of the environmental effect (β’ ) = ….. | |
| Random effects | Random effects |
| Among individual variance in intercept (VI) = …… | Sampled among-individual variance in intercept (V’I) = ….. |
| Among-individual variance in slopes (random slopes) (VS) = ….. | Sampled among-individual variance in slope (V’S) = ….. |
| Correlation between individual specific intercepts and slopes () = …. | Correlation between individual specific intercepts and slopes () = …. |
| Residual variance (VR) = ….. | Residual variance of sample (V’R) = ….. |

Note that we have slightly changed what is presented in the table above. Whereas earlier we presented the phenotypic variance due to a particular term (e.g., the variance due to individual differences in response to an environmental factor, or Vsx), above we present what many statistical packages produce, which in the case of random slope, is the variance in slope (Vs) rather than the phenotypic variance due to slope.

The best way to develop an intuition about what the covariance represents is to compare a graph with and without it. Below are two panels of simulated data. The one on the left shows your simulated individuals without the covariance, and the one on the right is the same set with the covariance you entered. The difference will, of course, be more obvious if you entered a large covariance (correlation).

Graph 1 (set on left): Plot of phenotype versus environment in data set without covariance, with each individual color-coded and individual reactions norms plotted

Graph 2 (set on right): Same plot but with covariance added.

Try this step a couple of times to explore in particular what a negative versus positive covariance does to the pattern of reaction norms.

**Point**: Intercept-slope covariance is an additional layer to understanding the phenotypic equation. You will note that it is not a parameter in the equation itself, but is rather describing the relationship between two parameters in the equation—that is, how they behave across individuals. The covariance does normally appear in a partitioning of variance:

where

Note: In SQuID and hence the covariance does not contribute to total phenotypic variance.

The biology of the covariance is just beginning to be explored (e.g., Mathot et al. 2011). Undoubtedly further understanding of it will contribute to understanding the evolution of plasticity, since the covariance links variation in plasticity with variation in mean trait values, and therefore could affect evolution in unexpected ways.

**R code:**

# install.packages("lme4")

LMM <- lme4::lmer(Phenotype ~ 1 + X1 + (X1|Individual), data = sampled\_data)

**Step 3: Sample size and precision**

**Sub-goal**: to develop understanding of how the level of replication in terms of the number of individuals and samples per individual influence the estimation of the among-individual variance in intercepts (VI), slopes (VSx) and their correlation (CorI,S).

**Introduction**:

Several papers have addressed the issue of sample size requirements for random regression models. The general conclusion is that accurate estimates of random regression parameters require relatively high sample sizes, particularly for estimating the correlation between intercepts and slopes. When planning an optimal experimental design, researchers are faced with a trade-off between the number of individuals and the number of observations per individual they can sample.

**Exercises**

We will perform a set of simulations that will address what is the sampling design, in terms of number of individuals and repeats per individual, that will maximize the precision of parameter estimates. Let’s assume that some researchers want to explore the effect of sample size in their estimates. They also want to determine the best decision in terms of allocating effort in sampling more individuals or more repeats per individual. The use of SQuID helps researchers to determine the best sampling design. They will use the same equation as we used in the step 2, e.g.,

**R code:**

# install.packages("lme4")

LMM <- lme4::lmer(Phenotype ~ 1 + X1 + (X1|Individual), data = sampled\_data)

The table below summarises the expected values for each parameter of the model:

|  |  |
| --- | --- |
| Fixed effects |  |
| Mean of the trait (β0) | 0.5 |
| Population-specific slope of the environmental effect (β1) | 0.5 |
| Random effects |  |
| Individual variance (VI) | 0.5 |
| Variance due to individual-specific responses to an environmental factor (random slopes) (VSx) | 0.5 |
| Correlation between individual specific intercepts and slopes () | 0.5 |
| Measurement error variance (Vm) | 0.05 |

For these simulations we are going to use a stochastic environment effect (x). We will also sample individual trait expressions the same number of times but at different instances of time among individuals. The among-individual variance in sampling of timing will be 0.

You should base your decision on the graphic representation of the estimates. The figure below is a summary of parameter estimates of 100 models performed from the 100 simulated datasets.

**Input**

A bar with numbers of measurements that can only be set to 100, 225, 400, 900.

For each sample size there will be three combinations: one with more individuals than repeats, one with equal number of repeats and observations, and one with more repeats than individuals.

|  |  |  |
| --- | --- | --- |
| Observations | Individuals | Repeats per individual |
| 100 | 5 | 20 |
| 100 | 10 | 10 |
| 100 | 20 | 5 |
| 225 | 5 | 45 |
| 225 | 15 | 15 |
| 225 | 45 | 5 |
| 400 | 10 | 40 |
| 400 | 20 | 20 |
| 400 | 40 | 10 |
| 900 | 18 | 50 |
| 900 | 30 | 30 |
| 900 | 50 | 18 |

**Results**

Figures

Each histogram shows the frequency distribution of each parameter based on the 100 simulations that you ran for each of the three study designs. “NI” is the number of individuals and “NR” is the number of repeats per individuals. The red line represents the “true” value that you set for this parameter above. The histograms provide information on the expected bias (inaccuracy) and imprecision that comes with each chosen study design. If the chosen study design would be perfect (i.e., resulting in highly precise and accurate estimates) your parameter estimates should be narrowly distributed around the true value. If your estimates are precise but inaccurate, the distribution should be narrow but peaking away from the true value. If your estimates are accurate but imprecise the distribution should be peaking at the true value but simultaneously very broad. Finally, if your estimates are imprecise and inaccurate the distribution would both wide and peaking away from the true value.