***Module xx, Step 3. Sampling when environments differ in characteristics***

**Sub-goal**: In this step we ask how sampling biases (as explored in Module 3, step 1) affect parameter estimates when the two environments differ in how they change among sampling periods.

**Introduction**: The issue of sampling biases was explored in some detail in Module 3. Here we extend that analysis to ask how parameter values are influenced by sampling issues when two environments are affecting phenotypes and the environments change differently between sampling occurrences. For example, one environment could be random between sampling occurrences but another could be temporally autocorrelated. If the researcher is sampling individuals with some variation in number of measurements and when they occur, then biases will inevitably arise as in Module 3, but here we explore how the more complex model might behave in that circumstance. You perhaps can imagine a wide variety of possible complexities that messy sampling might affect. We will only explore one set of possibilities, but this step will set you up to try many different scenarios.

**Exercise**:

As in step 1, we need to simulate a population and the data we collect from that population.

Number of individuals:

[Enter]

Among-individual variance (VI):

[Enter]

Measurement error variance (Vm):

[Enter]

Mean and variance in trait expressions measured

[Enter]

The environment

Environment 1:

[Enter slope, and whether shared or unshared]

Environment 1 decay function:

[Enter]

Environment 2:

[Enter slope and whether shared or unshared]

Environment 2 decay function:

[Enter]

Interaction:

[Enter magnitude and direction]

< Run simulation>

Results

Statistical output:

|  |  |
| --- | --- |
| Estimated Values | True Value |
|  |  |
|  |  |
|  |  |
| Individual variance (V’I) = ….. | Individual Variance (VI) = …… |
| Environmental variance accounted for = ….. | Environmental variance (VE) = …. |
| Residual variance (V’R) = ….. | Measurement variance (Vm) = ….. |

{To effectively write the end of this, I need to run a bunch of these simulations with the above table as output, as I am not sure other than vague intuition what the results will be.}

Conclusion

To be written

***Module 5, Step 4. When environments are correlated***

**Sub-goal**: In this step we ask how correlations among environments affect parameter estimates.

**Introduction**: Correlations among 2 or more fixed effects are the main reason for doing multiple regressions (after all, there is little point in assessing the independent effect of each factor in a joint model if they are not correlated). Strong correlations (collinearity) create problems with estimating independent effects. This step explores the effect of correlations on both main and interaction effects.

**Exercise**:

As in step 1, we need to simulate a population and the data we collect from that population.

Number of individuals:

[Enter]

Among-individual variance (VI):

[Enter]

Measurement error variance (Vm):

[Enter]

Mean and variance in trait expressions measured

[Enter]

The environment

Environment 1:

[Enter slope and whether shared or unshared]

Environment 2:

[Enter slope and whether shared or unshared]

Interaction:

[Enter magnitude and direction]

Correlation

[Enter magnitude and direction]

< Run simulation with and without correlation>

Results

Statistical output:

|  |  |  |
| --- | --- | --- |
| Estimated Values  with correlation | Estimated values without correlation | True Value |
|  |  |  |
|  |  |  |
|  |  |  |
| Individual variance (V’I) = ….. | Individual variance (V’I) = ….. | Individual Variance (VI) = …… |
| Environmental variance accounted for = ….. | Environmental variance accounted for = ….. | Environmental variance (VE) = …. |
| Residual variance (V’R) = ….. | Residual variance (V’R) = ….. | Measurement variance (Vm) = ….. |

{As above, to write the end of this, I will need to run some of these simulations with the above table as output in order to understand the results.}

Conclusion

To be written after running simulations some

**Sub-goal**: to develop awareness of the replication in samples per individual required to provide reasonable the estimates of the among-individual variance in slopes and the interaction parameter (VS12).

**Introduction**:

MDPP describes a reaction norm plane in n-dimensitonal environmental space. SQuID is limited to two dimensions, yet even in this simplified case, the level of sampling required to define individual variation in that reaction norm plane may be immense. Here we explore the combination of sampling needed.

**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 1, e.g.,

**R code:**

# install.packages("lme4")

LMM <- lme4::lmer(Phenotype ~ 1 + X1\*X2 + (1+X1\*X2|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 |
| Population-specific slope to X1(β1) | 0.5 |
| Population-specific slope to X2 (β2) | 0.5 |
| Population-specific parameter to interaction of X1 and X2 (β12) | -0.5 |
| Random effects |  |
| Individual variance (VI) | 0.5 |
| Variance due to individual-specific responses to X1 (random slopes) (VS1) | 0.5 |
| Variance due to individual-specific responses to X2 (random slopes) (VS2) | 0.5 |
| Variance due to individual-specific responses to the interaction between X1 and X2 (random slopes) (VS12) | 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 as in Step 3 of Random Regression, but with additional betas and sS terms added (omit correlation term as we are ignoring it here).

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.