|  |  |  |  |
| --- | --- | --- | --- |
| Обозначение | Формула | Словесный термин | Примечание. Что показывает, для чего считается |
| a |  | Number *M. trossulus* with T-morphotype in a sample | Лучше избегать слова “frequency”, чтобы не путать с частотой, как оценкой вероятности события |
| b |  | Number *M. trossulus* with E-morphotype in a sample |  |
| c |  | *M. edulis* with T-morphotype |  |
| d |  | *M. edulis* with E-morphotype |  |
| ***MTprev***  ***Или PMT*** | (***a***+***b***)/(***a***+***b***+***c***+***d***) | M.trossulus prevalence in a population | Частота (proportion) одного из видов в поселениях. Характеризует генетическую структуру смешанного поселения. Собственно, разработка метода оценки этой величины с помощью морфологического маркера, и является целью работы.  *NB. Мне кажется, что если мы продаем работу, как подход к диагностике присутствия в акватории плохой троссулус, то лучше как раз везде выпячивать терминологию из клинических исследований.* |
| ***PT*** | (***a***+***c***)/(***a***+***b***+***c***+***d***) | Proportion of T-morphotype in a population | Это тот экспресс-тест, который мы используем для диагностики присутствия в популяции инвазивной M.trossulus. На основе этого “proxy” производится оценка генетической структуры поселения, находящегося в фокусе интереса. |
| ***P(T|MT)*** | ***a***/(***a***+***b***) | Conditional probability of being T-morphotype if a mussel belongs to M.trossulus, or, in terms of ROC-analysis, “sensitivity” | Величина показывает насколько тесно связан T-морфотип с генотипом M.trossulus. |
| ***P(Е|MЕ)*** | ***d***/(***с***+***d***) | Conditional probability of being E-morphotype if a mussel belongs to M.edulis, or, in terms of ROC-analysis, “specificity” | Величина показывает насколько тесно связан E-морфотип с генотипом M.edulis. |
| ***AUC*** | 1/2(Sensitivity+Specificity) | Area under curve, or accuracy | Эта величина характеризует дискриминирующую силу классифицирующего признака (в нашем случае способность морфотипов определять генотипы).  В случае бинарного отклика (T-morphotype vs E-morphotype) величина показывает с какой вероятностью случайно взятая из популяции мидия, имеющая T-морфотип, может быть правильно идентифицирована как M.trossulus. AUC имеет ограничения, связанные с тем, что не подразумевает варьирования MTprev в пределах совокупности, которую она характеризует. Поэтому может рассматриваться только как характеристика дискриминирующей силы для всей выборки в целом.  Один из посылов статьи (причем, я бы не делал его основным!!!) это то, что дискриминирующая сила непостоянна в некоторых типах акваторий она снижается. Это наше предупреждение пользователям. Причины этого снижения, нам на самом деле неизвестны. Мы лишь можем сказать, что есть некоторая корреляция с соленостью.  NB! Я уверен, что соленость здесь хоть и играет какую-то роль, но не ключевую. Что-то мне подсказывает, что здесь важны еще и другие факторы. Возраст гибридизации, например. Но попробуй это донести до ПП... |
| ***PMT|T*** | (***MTprev\*PT***)/***PT\_exp***, where ***PT\_exp***=(***MTprev\*P(T|MT))+((1- MTprev)\* (1-P(Е|MЕ))*** | Conditional probability of being M.trossulus if it is known that a mussel possess T-morphotype | Эта величина нам необходима для оценки видовой принадлежности у конкретной мидии, для которой известен морфотип.  NB! Здесь важно понимать, что в современном мире нет места дискретным видам и видовая принадлежность имеет вероятностную природу. То есть держа в руках какую-то особь мы можем ее отнести к тому или иному виду только с определенной вероятностью, которая у «хороших» видов, да, близка к 1, но у большинства других не совсем. |
| ***PME|E*** | аналогично | Conditional probability of being M.edulis if it is known that a mussel possess E-morphotype | Аналогично. |
| NB при всей полезности знаний об этой вероятности она имеет важное ограничение: согласно теореме Байеса эти две последние вероятности имеют связь с *MTprev (или 1-MTprev). Эту связь мы можно оценить ДВУМЯ способами.*   1. *Первый способ. Мы можем предположить, что возможны любые значения MTprev, но при этом P(T|MT) (Sensitivity) и P(T|ME)=1-P(Е|MЕ) (1-Specificity) есть величины ПОСТОЯННЫЕ, равные оценкам, сделанным по всей изученной совокупности в целом. То есть мы предполагаем, что Specificity и Sensitiity НЕ ВАРЬИРУЮТ в зависимости от MTprev. Для описания этой модели мы строим наши баесовские кривые. Важное замечание: у нас нет никаких оснований для утверждения, что Specificity и Sensitiity не связаны с генетической структурой популяции. Можно предложить довольно много механизмов, когда такая связь появляется. Поэтому нам нужно продублировать оценку связи PMT|T (PME|E) с гентической структрой популяции вторым способом.* 2. *Второй способ. Мы можем сделать очень мягкое допущение: предположить, что что между* ***PMT|T (или PME|E) и*** *MTprev существут линейная связь. Это означает всего лишь, что эти две вероятности возрастают / убывают по мере изменения генетической структуры. Мо можем вывить эту линейную связь с помощью регрессионного анализа. Для этого мы вводим Модель 1.* | | | |
| ***Model-1*** | Logit(Congr) = b0 + b1\*Place + b2\*MTprev + b3\*I\_morphT + Interactions + Random effect of population.  Это я условно написал Надо написать аккуратнее на LaTex | The Model-1 discribes the association between the probability of correct identification of “species” (i.e. M.tossulus-like or M.edulis-like) useing the mussel’s morphotype. | Подбор коэффициентов этой модели позволяет написать в явном виде формулы, по которым можно вычислить вероятность того, что данная конкретная мидия, имеющая данный конкретный морфотип может быть отнесена к соответствующему виду (ну, естественно, с ограничениями, хараткерными для регрессионного анализа).  NB! Это ключевая модель во всей работе и ПП, кажись, этого не понимает.  Мы видим, что логистические кривые очень близки к баесовским кривым. Они идут очень параллельно. Но есть нюанс.  По крайней мере, в случае с Белым морем, баесовская кривая сильно за пределами довертельного интервала, полученного для регрессионной модели для T-морфотипа. Это означает Specificity зависит от генетической структуры смешанного поселения. Почему это так? Надо понимать, что отскок получается за счет b0. То есть, в случае логистической кривой это означает, что при средних значениях MTprev у нас наблюдается некоторый дефицит правильно определенных троссулусов (имеющих T-морфотип). Почему так? Один из вариантов ответа - потому что при среднем значении MTprev, то есть когда приблизительно поровну и троссулусов и эдулисов начинается конкурентное исключение троссулусов (истинных, корректно определяемых) это и приводит к падению уровня корректного определения по сравнению с тем, что предсказывается первым способом. В Баренцевом море это, видимо чуть мягче. |
| Полученная модель позволяет делать очень многое. В частности, если нам известна генетическая структура того или иного поселения в пределах Белого моря, опресненного или неопресненного Баренцева моря мы можем с легкостью дать оценку генетической структуры в некой новой выборке из того же местообитания, где проводились оценки генетической структуры, например, другим исследователем ранее. Но! Эта модель не годится для оценки генетической структуры в абсолютно новом местообитании. Для этого нам нужна Модель-2. | | | |
| ***Model-2*** | Logit(MTprev) = b0 + b1\*PT + b2\*Place + Interaction | The Model-2 discries the association between genetic structure of a population and proportion of T-morphotype. | Эта модель дает конкретные формулы, которые позволяют вычислить каждого из трех регионов долю троссулусов, если нам известна доля язычковых.  Мы видим, что во всех изученных регионах эта связь ЕСТЬ. Она везде сильная. В Белом море вообще почти PT = MTprev. Угловой коэффициент во всех трех случаях очень близок. Отличается сильно b0. То есть в случае PT = 0 (нет язычковых мидий) в Белом море почти не будет и троссулусов, в Баренцевом опресненном, есл нет язычковых, то небольшая доля троссулусов в популяции, таки, будет. А вот в соленом Баренцевом все иначе - там нет таких мест T-морфотип полностью отсутствовал бы. Почему это так? Может это действительно что-то адаптивное, что эдулисы сперли у троссулусов? Но это уже высшая генетика, в которой я не разбираюсь. |
| Теперь, имея в руках Model-1 и Model-2, мы можем дать оценку гентической структуре любой популяции, находящейся в пределах одного из терх регионов, если знаем долю мидий с T-морфотипом. Кроме того, оценив долю миди с Т-морфотипом в локальном поселении, мы можем вычислить вероятность того, что данная конкретная мидия относится к троссулусам, если она имеет Т-морфотип, или к эдулисам, если она имеет Е-морфотип.  Вот он калькулятор! | | | |