Теоретические модели нужны для того, чтобы, используя минимальные данные о генотипах мидий из района интереса, построить наиболее надежную предиктивную модель, позволяющую решить две задачи. Во-первых, оценить по частоте морфотипов частоту генотипов. Во-вторых, исходя из знания о морфотипе конкретной мидии, сделать обоснованный вывод о вероятности ее принадлежности к тому или иному виду.

Задача - выработать стратегию поиска калибровочных выборок, которые будут использоваться для построения теоретических моделей, основанных на теории вероятностей (надо придумать какое-то название для них, не писать же «калькулятор»).

Для выработки стратегии поиска калибровочных выборок мы сделали следующие допущения.

1. Полученные нами эмпирические регрессионные модели достаточно надежно описывают связь Ptros c P\_T (регрессионная модель 5) и Pcorrect c Ptros и Morph (регрессионная модель 4), то есть они могут являться эталоном, с которым будут сравниваться теоретические модели, построенные по калибровочным выборкам.
2. Для настройки теоретических моделей необходимо минимальное количество калибровочных выборок.
3. Минимально необходимым количеством калибровочных выборок для построения каждой теоретической модели является две выборки. *Надо придумать почему две*
4. Приемлемыми калибровочными выборками являются те выборки, на основе которых может быть построена теоретическая модель, дающая предсказания максимально похожие на предсказания соответствующей эмпирической модели.

Для каждого изученного региона (W, BL, BH) были найдены все возможные парные комбинации выборок (population) из Modelling data set. Для каждой пары была вычислена величина, характеризующая степень сходства популяций по величине Ptros.

Данная величина вычислялась по следующей формуле:

Delta = max(Ptros1;Ptros2)(1 - min(Ptros1;Ptros2)

Эта величина стремится к

* 1 если Ptros1 >> Ptros2 или Ptros2 >> Ptros1 Максимально различная генетическая структура калибровочных выборок.
* 0 если Ptros1 = Ptros2 Максимально похожая генетическая структура калибровочных выборок
* 0.25 если Ptros1 = Ptros2= 0.5. Калибровочные выборки содержат равные доли M.trossulus и M. edulis

Материал из каждой пары популяций был объединен и были вычислены P\_T\_MT и P\_T\_ME, необходимые для теоретических моделей.

Используя коэффициенты эмпирической регрессионной Модели 5, были вычислены предсказанные значения Ptrosregr для P\_T, принадлежащих интервалу от 0 до 1. Далее были вычислены значения Ptrosteor, которые предсказывает теоретическая модель, основанная на P\_T\_MT и P\_T\_ME, найденных для данной пары калибровочных выборок. Ptrosteor  были вычислены для допустимого диапазона P\_T (см. ограничения в описании теоретических моделей). Далее на всем допустимом диапазоне P\_T была проведена оценка отклонений Ptrosregr от Ptrosteor по следующей формуле:

Goodness1 = 1/(Σ(Ptrosregr - Ptrosteor)2)

Эта величина стремится к нулю если предсказания теоретической модели сильной отклоняются от предсказаний эмпирической и стермится к бесконечности если предсказания совпадают.

Аналогично было проведено сравнение предсказаний регрессионной модели 4 и соответствующей ей теоретической модели. Для этого были вычислены предсказанные значения Pcorrectregr для Ptros, лежащих в диапазоне 0 до 1 и Pcorrectteor для допустимого диапазона Ptros (см. ограничения в описании теоретических моделей). Далее на всем допустимом диапазоне Ptros была проведена оценка отклонений Pcorrectregr от Pcorrectteor :

Goodness2 = 1/(Σ(Pcorrectregr - Pcorrectteor)2)

Таким образом, для каждой пары популяций были получены три величины: Delta, Goodness1 (для первой теоретической модели) и Goodness2 (для второй теоретической модели). Для каждого региона были построены точечные диаграммы Goodness1 vs Delta и Goodness2 vs Delta. Для каждого региона были подобраны сглаживающие кривые (loess regression), описывающие связь между Goodness и Delta.