

Bioenv

- 对 **vegan** 包的 **bioenv** 函数介绍
 - 目的：挑选与群落距离矩阵相关系数最大的环境因子组合。
 - 方法：计算群落距离矩阵和环境因子距离矩阵，两个距离矩阵之间的相关系数，挑选最佳的环境因子组合方法。其中环境因子组合方法为
 - 对单个的环境因子分析：群落距离矩阵与单个环境因子距离矩阵的相关性，挑选相关系数最高的环境因子。
 - 对任意两个环境因子分析：群落距离矩阵与任意两个环境因子距离矩阵的相关性，挑选相关系数最高的两个环境因子。
 -
 - 对 **N** 个环境因子分析：群落距离矩阵与 **N** 个环境因子距离矩阵的相关性，得到全部环境因子（**N** 个）的相关系数。
 - 问题：
 - 当多个环境因子的组合相关系数差不多时，如果单纯的按照相关系数的值挑选 1 个最高的环境因子的组合，可能漏掉部分关键的环境因子。
 - 当环境因子过多时，计算量非常大。即便是使用并行处理，仍然耗费很长时间。
- 精分项目中的应用---PANSS
 - 目的：找到 PANSS 分量表的组合。作为后续回归分析的 **response**，**predictor** 为 mOTU profile.
 - 数据介绍：
 - 环境因子：PANSS 有 30 个分量表（p1-p7, m1-m7, g1-g15，只有 90 个 case 测量 PANSS），每个分量表代表不同的症状程度。比如 p1 代表妄想，p2 代表联想散漫.....
 - 群落：mOTU profile, 171 个样本，360 个 mOTUs
 - 数据处理：
 - 数据匹配。因为 PANSS 只有 90 个病人评分，所以群落数据匹配为对应的 90 个样本。
 - 数据缩减。因为 30 个分量表全部计算，计算量非常大。所以把 PANSS 拆成三部分：
 - p1-p7：阳性症状，计算得到 p1 p2 p6 p7 的相关系数最大。
 - p1-p7, m1-m7：阳性症状&阴性症状， 计算得到 p1 m2 的相关系数最大。
 - p1, m2, g1-g15：前两部挑出的 p1 m2 和综合病理量表 g1-g15， 计算得到 p1 g3 g4 g7 g8 g11 g12 g13 的相关系数最大。
 - 后续：

- 对上述的组合处理：简单的相加？