**Homework-01**

**The questions**

1. What difficulties does the old paradigm of the hypothesis test encounter?

（1）Ronald Fisher的假说-验证的思想是小概率反证法，即构造一个小概率事件或原假说（H0），同时提出另一个与原假说对立的备择假说（H1），然后设计实验，即从总体中抽出样本，依据样本计算统计量的值，根据预设的概率水平检验和选择的统计量的概率分布，确定原假说成立的可能性大小，做出拒绝或接受H0判断。

按照Fisher的本意，假说不是被证实的，而是采纳“证伪“的结论，即证伪主义。一般而言，任何一个H0被提出时，被认为极有可能是错的，即构造的原假说是一个小概率事件，在一次实验中，这种事件是不可能发生的。就这个意义而言，整个验证过程是试图推翻H0，以支持与原假设对立的H1。在假说-验证体系中，只有当H0被拒绝了，才可明确是否存在因果关系，增加对问题的理解。

与物理化学等学科不同的是生态学问题的答案不是简单“是”或“不是”。通过H0显著性检验来验证生态学假说并不严谨。在生态学上构建单一主导因子的H0是很难的。

另外，生态过程是始终演化的，生态规律（因果关系）在不断变化。在逻辑上，生态学的所有实验验证倾向于“证实”假说（H0）成立，而不是严格的Fisher证伪结论。

（2）另一困境是操控实验。操控实验包括正交实验设计法与析因法，无论哪种实验，都遵循重复、随机、区组化的原则。由于研究对象所处的环境是开放的，其中影响因子及相互关系难以被发现和确定。另外，在大尺度环境下，要采取严谨和一致操作才可能排除无关干扰。可是，Anja Jaeschke等认为无论如何精心设计和操控，都“不可能包括所有多样性和空间”。因此Samuel Scheiner等认为野外实验存在难以重复的困难，不能作为假说的判决性实验。

2. What advances does the paradigm of data science used for ecological data?

（1）相关性分析

数据科学范式寻找相关性，而不是因果律。数据科学范式认为分析变量之间的相关性比探索因果律更重要，基于相关分析的预测是大数据研究的核心。一方面数据规律本身是从过去所积累的数据中挖掘出来的，完全可解释过去的现象或问题，另一方面，大数据包括海量的各种现实数据，通过机器学习过去的经验来推测未来。因此，与因果律相比，基于大数据相关性的预测更准确，而且不易受偏见的影响。

（2）全数据归纳：样本≠总体，避免经验科学范式相关≠因果

数据科学家采用归纳方法，而不是演绎逻。经验科学采用的是不完全归纳。数据科学范式沿袭了经验科学的归纳逻辑，不同的是数据科学家采用的是全数据模式，即“样本=整体”的完全归纳法，克服了小样本不完全归纳法的局限性，利于发现异常值。

（3）利用非参数或半参数模型，例如提升回归树、朴素贝叶斯、聚类分析、多层感知等建立分类与回归模型。

（4）将数据科学范式与理论科学范式融合，避免归纳偏见。