**Homework-01 学号：SA24008220 姓名：熊力樊**

**1. 生态学主要采用的是“假说-验证”范式，面临的困境是什么？**

1. 生态学问题的答案不是简单绝对的“是”或者“不是”，通过小概率事件或原假说显著性检验来验证生态学假说并不严谨。例如，检验群落中竞争的作用，不是通过一个简单实验可以证伪的，群落中不仅存在竞争，还有捕食、寄生和干扰等作用。在生态学中构建单一因素的主导因子是困难的。
2. 生态过程是始终演化的，生态规律（因果关系）在不断变化。生态学的所有实验验证倾向于“证实”假说成立，而不能严格证伪。
3. 操控实验为另一困境。操作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、随机化、区组化的原则。由于实验对象所处环境是开放的，其中影响因子及相互关系难以被发现和确定。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。但生态学实验，例如多样性实验、野外实验都存在无法包括所有的多样性和空间，并且难以重复，不能作为假说的判决性实验。

**2. 何为“数据科学”范式，与“假设-验证”范式主要区别是什么？**

1. 科学发现的逻辑起点不同。经验科学范式认为“科学始于观察”，即在自然和实验观察的基础上，通过归纳提炼出科学理论，假说-验证范式主张科学发现始于科学问题，而数据科学颠覆了原来的科学发现模式，从数据出发，利用数据挖掘方法发现数据中蕴含的规律性，形成了“科学始于数据”的新模式。
2. 寻找相关性，而不是因果律。无论是经验科学范式，还是假说-验证范式，都认为科学研究的目的就是寻找现象之间的因果关系。与其他研究范式不同，数据科学范式认为分析变量之间的相关性比探寻因果律更重要，基于相关分析的预测是大数据研究的核心。相关不能揭示因果关系，那么数据挖掘发现的规律是否是科学规律？按照科学哲学观点，如果一个命题能够解释以往出现的现象，又能预测未来可能出现的新现象和新问题，其科学性就得到了检验。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，完全可以解释过去的现象或问题，另一方面，大数据包含了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。
3. 第三，数据科学采用归纳方法，而不是演绎逻辑。经验科学采用的是不完全归纳。例如，19 世纪德国生物学家 Carl Bergman（1814～1865）通过归纳法提出了“同一物种在越冷的地方，其个体的体积越大，外形越接近球形”这一规律。另外，从达尔文的生物进化论到种群增长的Logistic方程都采用了归纳方法。数据科学范式沿袭了经验科学的归纳逻辑，不同的是数据科学采用的是全数据模式，克服了小样本不完全归纳法的局限性，利于发现异常值。

**3. 什么是数据驱动的方法？与传统上哪种范式采用的方法相似？**

1. 数据驱动是数据科学采用的具体归纳手段和工具
2. 与传统的统计分析相比，数据驱动在很多方法上与其是同源的。例如，数据驱动采用的朴素贝叶斯分类就是统计理论的发展和延伸。再如，常用于数据驱动的主成分分析和回归分析也属于统计学范畴。但数据驱动与传统数据分析存在本质上的区别，传统数据分析需要对数据分布和变量间的关系做假设，确定用什么样的概率函数来描述变量间的关系，然后建立参数模型，并依据统计推断，确定模型的合理性，而数据驱动并不需要对数据做任何假设，而是侧重利用机器学习方法，自动寻找变量间的关系。因此，对于海量数据，数据驱动具有更加强大、更灵活、更高效的特点。