**Homework-01**

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

答：

假说-验证范式的本质逻辑是：先针对科学问题，构思一个假说，界定有关变量，并假设变量之间存在因果关系，然后根据假说推演可验证的预言，再设计实验收集数据，验证预言，最后依据实验结果，拒绝或接受假说。这种范式的思想是小概率反证法，即构造一个小概率事件或假说H0，同时提出另一个与原假说对立的备择假说H1，然后根据实验结果拒绝或接受H0，这属于证伪主义。

假说-验证范式遇到的困难有：首先，生态学问题非常复杂，并非像H0与H1的关系，不是非此即彼，我们更应该证实H0成立，而非严格的Fisher证伪结论，实际上H0不存在或者很难精准地提出来；其次，批判性试验很困难。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，但无论如何精心设计和操控，都“不可能包括所有的多样性空间”。

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

答：

数据科学指的是利用超级计算能力，直接分析海量数据发现相关关系，获得新知识。它在科学发现的逻辑上不同于其他范式，主要有三个方面：第一，科学发现的逻辑起点不同，它从数据出发，利用数据挖掘的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；第二，寻找相关性而非因果规律，分析变量之间的相关性比探寻因果律更重要，因为基于大数据相关性的预测更准确，而且不容易受偏见的影响；第三，数据科学采用归纳方法，而非演绎逻辑，数据科学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，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。

综上，数据科学范式主要有以下四个方面的优势：能够进行相关性分析；全数据归纳：因样本=总体，避免了经验科学范式中相关≠因果的缺陷；利用非参数或半参数模型；将数据科学范式与理论科学范式融合，避免了归纳偏见。