任意初始化Q(s, a)

循环

初始化s为初始状态

循环

从s中选择

执行动作，观察奖励和下一个状态

直到是终止状态