

买卖点选择 - 36.8%原则

- 参考资料1：数学也荒唐 Chapter14
- 参考资料2：秘书问题

问题：作为英明的老板，你决定招个新人。传闻贵司只要优中之优，此言不虚。共有 N（此数目已知）人来参加面试，面试顺序随机。每次面试之后，你只有两个选择：要么聘用此人，面试结束；要么请其回家，老死不相往来。要注意，两个选择必择其一。如果到最后也没有找到满意的人选，则必须聘用最后一人。万万不可让平庸之辈混进公司啊！那么问题来了：想让雇到精英的机会最大，该何时终止甄选呢？

案例剖析

A、B、C、D 这个 4 人看了招聘启事后，前来应聘数字项目主管的职位。他们的资质参差不齐：A 一般，B 良好，C 优秀，D 则是顶级。作为招聘者，你的目标就是把 D 招进来。面试顺序随机，共 24 种可能：

- ABCD、ABDC、ACBD、ACDB、ADBC、ADCB
- BACD、BADC、BCAD、BCDA、BDAC、BDCA
- CABD、CADB、CBAD、CBDA、CDAB、CDBA
- DABC、DACB、DBAC、DBCA、DCAB、DCBA

第一种策略是先到先得，不管其他人（即策略0），选到最优者的概率为 1/4。另一种同样莫名其妙的策略是把所有人都面试一遍，但就要最后一人（策略3），选到最优者的概率也是 1/4。

但成功概率可以高过 1/4，没想到吧。可以先面试 1 人了解一下大致水平，但这人肯定不要，仅供参考，一出现比他水平高的就直接要了（策略1）。

这种方法真的比 策略0 和 策略3 好吗？我们来分情况讨论，看选到最优者 D 的概率有多高。

- 第一个面试的是C：唯一比C强的就是D，D一出现就会被录取，有6/24的可能性。
- 第一个面试的是B：此时，C和D谁先出现，谁被录取。在24种可能中，B为第一有6种：BACD、BADC、BCAD、BCDA、BDAC和BDCA，只有3种情况会选到D，所以最终选D的可能性有3/24。
- 第一个面试的是A：此时，第二个面试者会被选中。经过检验发现，只有2/24的可能性会选到D。
- 第一个面试的是D：那就把他淘汰了，认倒霉吧。
- 结果如下所示

1	2	3	4	5	6
CABD	CADB	CBAD	CBDA	CDAB	CDBA
BACD	BADC	BCAD	BCDA	BDAC	BDCA
ABCD	ABDC	ACBD	ACDB	ADBC	ADCB
DABC	DACB	DBAC	DBCA	DCAB	DCBA

数学推导

对于 N 个候选人，我们可以分别考虑每个策略 k , $k \in (0, N - 1]$ (策略 k 定义为排除前 k 个观测，从第 $k + 1$ 个观测起择优选择的策略)，计算选到最优者的概率 $P(N, k)$ 是多少。设最优者为第 j 个候选人， j 为一个提前确定值的概率为 $\frac{1}{N}$ 。

- 如果 $j \leq k$ ，则最优者被排除；
- 如果 $j = k + 1$ ，那他来得正好，恰好会被选中；
- 如果 $j > k + 1$ ，只有 $k + 1$ 到 $j - 1$ 之间没有合适人选时，他才会被选中，这一概率为 $\frac{k}{j-1}$ 。可以这样理解，选取 $j - 1$ 个数，最大的数值在前 k 位的概率；
- 所以，最优者是第 j 个候选人且通过策略 k 被选中的概率是： $\frac{1}{N} \times \frac{k}{j-1}$ ，且 $j > k$ 。若 $j \leq k$ 时，此概率为 0。若 $j = k + 1$ ，则 $\frac{1}{N} \times \frac{k}{j-1} = \frac{1}{N}$ ；
- 由此可得，选中最优者的概率为 $P(N, k) = \sum_{j>k} \frac{1}{N} \times \frac{k}{j-1}$ ，整理后可得 $P(N, k) = \frac{k}{N} \sum_{j>k} \frac{1}{j-1}$ ，即

$$P(N, k) = \frac{k}{N} \left(\frac{1}{k} + \frac{1}{k+1} + \frac{1}{k+2} + \dots + \frac{1}{N-1} \right)$$

- 当 $N \rightarrow +\infty$ 时，令 $\frac{k}{N} = x$ ， $P(x) = x \int_x^1 \frac{1}{t} dt = -x \log(x)$ ，令 $P(x)$ 对 x 的导数为 0，解出 x ，我们得到最优的 $x = 1/e$ 。

统计模拟

```
library(dplyr)
library(ggplot2)

secretary_function <- function(total_n){
  sp_k_n <- function(k, n){
    num <- k:(n-1)
    return(k/n*sum(1/num))
  }

  df_temp <- data.frame(start_value = 1:(total_n-1)) %>%
    group_by(start_value) %>%
    mutate(final = sp_k_n(start_value, total_n)) %>%
    ungroup()
  df_temp <- rbind(data.frame(start_value = 0, final = 0), df_temp)

  return(df_temp)
}

mydata <- secretary_function(10000)
p <- ggplot(data = mydata, aes(x = start_value, y = final))
p <- p + geom_bar(stat = "identity", aes(fill = start_value)) + geom_line()
```

