**七、费米估计**

数据分析面试过程中，常常会出现一类估算问题，“估计一下星巴克一年能创造多少收入”，“你的家乡每年有多少人乘坐火车”等等，这一类问题统称为费米问题，这类问题看似很难入手，难以获取准确的答案，但面试官的目的并非让你得出一个非常准确的数字，只是通过这样一个问题考察候选人的逻辑拆解能力和结构化思维能力，所以面对这类问题，谨记“化大为小、化难为易”，将难以获取的大数一层层尽可能完整地拆解下来，对小数进行合理估算，再按照之前的拆解思路组装回去便可以得到令人满意的问题答案。

对于费米估计问题，可从估计事务的**需求端**、**供给端**、**供需端和相关事物的数量估算**上进行考虑，下面从实例出发对费米估计问题的解决方法进行说明。

**1、需求端 — 从用户需求的角度估计市场数据量级**

示例问题：全中国一年要消费多少牛奶

STEP1：我们将全中国一年消耗的牛奶量进行初步拆解

**全中国一年消耗的牛奶量=每天人均消耗牛奶量\*全国人数\*一年的天数**

其中全国人数和一年的天数为确定值，剩下每天人均消耗牛奶量需要进行估算。

STEP2：拆解每天人均牛奶消耗量

因不同人对牛奶的喜爱度不同，对牛奶的喜爱程度可能与地区有关，与年龄段有关，认为对牛奶的消耗能力与年龄段的相关性更大，按年龄段进行估计

**每天人均牛奶消耗量 = (3-10岁每天人均牛奶消耗量\*3-10岁人群占比+10-30岁每天人均牛奶消耗量\*10-30岁人群占比+30-50岁每天人均牛奶消耗量\*30-50岁人群占比+50-70岁每天人均牛奶消耗量\*50-70岁人群占比+其余人群每天人均牛奶消耗量\*其余人群占比)**

根据中国人口年龄层占比，可以估计得到不同年龄段人群的数量占比，通过日常生活经验估计不同年龄段人群每天牛奶的消耗量，完成估计后便可得到每天人均牛奶消耗量的估计值。

STEP3：将估计值代入原公式得到最终的每年中国牛奶消耗量

思路优化：以上只进行三步拆解，得到最终的估计值，其中第二步拆解中可在年龄分层下继续进行地区分层，三餐消耗量分层等多个维度进一步拆解。

**2. 供给端 — 从产品供应的角度估计市场数据量级**

示例问题：北京市每年多少人乘坐地铁

STEP1：初步拆解地铁乘坐人数

**北京市地铁乘坐人数 = 工作日每天地铁乘坐人数\*每年工作日天数+普通周末每天地铁乘坐人数\*普通周末天数+法定节假日(认为是旅游高峰)每天地铁乘坐是人\*法定节假日天数**

同样的，不同种类日期的天数是固定的，接下来按种类对每天地铁乘坐人数进行估计

STEP2：拆解不同种类日期地铁每天乘坐人数

以工作日每天地铁乘坐人数为例，地铁每天运行时间是固定的，7-10点为上班高峰期，18-20点为下班高峰期，开始运行和最后运行1h为空闲时段，其余时段为正常时段，故按高峰期、空闲时段、正常时段进行分类估计

**工作日每天地铁乘坐人数=高峰时期地铁乘坐人数+空闲时段地铁乘坐人数+正常时段地铁乘坐人数**

对不同时段地铁乘坐人数进行进一步拆解

**xx时段地铁乘坐人数=北京市地铁线路条数\*(x时段总时间/xx时段相邻列车时间间隔)\*每节车厢可搭乘乘客数\*高峰时段平均实际满载率\*车厢节数**

高峰期、空闲期、正常时段列车相邻列车运行时间间隔可通过相关资料获取，每辆列车高峰期满载率可通过经验进行相关估计，北京市地铁线路条数、车厢节数、每节车厢可搭乘乘客数也为确定数，便可对不同时期地铁乘坐人数进行估计

同理，可对普通周末、节假日的地铁搭乘人数进行估计

STEP3：数据整合，得到结果

经过STEP2中的数据估算，可得到不同类型日期下北京市地铁乘坐人数的估算值，将其根据STEP1中的公式进行组合，便可得到最终的估算结果

**3.供需端/相关物品估算**

利用供需结合的方式便是从估计事物的两端进行综合考虑，供给和需求满足一定比例作为估算的前提条件

例如 估计xx市需要多少个ATM

假设xxx区有20000名常住人口，约10台ATM机，则可认为1台ATM机可满足2000人提款需求，若xx市有300万常住人口，在供需达到平衡时，需要1500台ATM机(以上数据纯属虚构)

除了供需端估算方法，若某些事物的相关物品较容易估算数据，也可将不好估算的预测事物数量转化成相关事物的估算上，如估算某个国家有多少辆车，可以转化为估算汽油消耗量、汽车保养店数量等，进行相关物品数量估算，反推出最终的估算结果。

**面经真题：**

1、估算你的家乡每年有多少人乘坐火车（字节跳动）

2、估算一下北京有多少个垃圾桶(阿里巴巴)

3、长沙一年的燃油税费是多少(网易)

4、估算北京五环实时车流量（[滴滴](https://www.nowcoder.com/jump/super-jump/word?word=%E6%BB%B4%E6%BB%B4" \t "_blank)）

5、每天全国的电商订单量大概是多少?（[滴滴](https://www.nowcoder.com/jump/super-jump/word?word=%E6%BB%B4%E6%BB%B4)）

6、估算上海地铁每天的客运量（[拼多多](file:///C:\jump\super-jump\word%3fword=%25E6%258B%25BC%25E5%25A4%259A%25E5%25A4%259A)）

**7、**估算今年国庆全国的旅游消费总支出？（[神策数据](file:///C:\jump\super-jump\word%3fword=%25E7%25A5%259E%25E7%25AD%2596%25E6%2595%25B0%25E6%258D%25AE)）

8、估算江苏省面积（[快手](file:///C:\jump\super-jump\word%3fword=%25E5%25BF%25AB%25E6%2589%258B)）

9、若[贝壳](https://www.nowcoder.com/jump/super-jump/word?word=%E8%B4%9D%E5%A3%B3" \t "_blank)要进入一个新的城市要如何去估计这个城市的需求量（[贝壳](https://www.nowcoder.com/jump/super-jump/word?word=%E8%B4%9D%E5%A3%B3)找房）

10、怎么估算上海外卖员的数量（[拼多多](https://www.nowcoder.com/jump/super-jump/word?word=%E6%8B%BC%E5%A4%9A%E5%A4%9A" \t "_blank)）

**八、数理统计脑筋急转弯**

数据分析面试过程中有时会进行智力题的考察，平时可通过面经对智力题进行拓展积累，下面整理了几种常考智力题型

**1. 烧绳子问题**

一根不均匀的绳子，全部烧完要1小时，问如何用两根绳子确定15分钟？

答：第一根绳子点燃两头，第二根绳子点燃一头，第一根绳子燃尽时点燃第二根绳子的另一端同时开始计时，至第二根绳子燃尽时，恰为15min

**2. 赛马问题**

五条赛道，25匹马，没有秒表，请问最少比几次才能找到最快的三匹

答：7次

第一轮(5次):将25匹马分成五组进行比赛

第二轮(1次):将上面5组的每组第一名放在一起进行比赛，本轮比赛中取得第一名的马所在的组设为A组、第二名到第五名分别归属于B、C、D、E组，则跑的最快的三匹马只能从A、B、C三组中产生

第三轮(1次)：第二轮比赛中的第一名为最快的一匹，2、3名从A组2/3名、B组1/2名、C组第1名中产生，故将A组2/3名、B组1/2名、C组第1名共五匹马放在一组进行比赛，决出前两名与上轮比赛决出的第一名便是25匹马中的前3名

合计三轮七次

**3. 天平问题**

（1）27个砝码中，只有1个与其他26个砝码重量不同且不知该砝码是轻还是重，问最少需要称多少次能找到该砝码？

答：将27个砝码分为3堆，每堆9个，分别记为A(9)、B(9)、C(9)。  
将A(9)与B(9)比较，此时平衡，则C(9)中包含特殊砝码。  
将C(9)均分成3堆，每堆3个，分别记为Ca’(3)、Cb’(3)、Cc’(3)，将Ca’(3)与Cb’(3)比较，此时平衡，则Cc’(3)中包含特殊砝码。  
将Cc’(3)中的3个砝码分别记为 i、j、k ，将 i 与 j 比较，此时平衡，则 k 为特殊砝码。这就是三分解的最优情况，称了3次。

（2）称出1-60g所有整数重量，最少需要多少砝码？

答：需要6个砝码，分别重1g,2g,4g,8g,16g,32g。

（3）两个砝码分别重7g和2g，如何只称3次将140g的盐分为50g和90g?

答：用7g+2g的砝码称出9g盐,140g-9g=131g。  
用7g+2g砝码把131g盐分成两部分61g,70g (61g盐+2g砝码+7g砝码=70g盐）。  
用2g砝码和9g盐把61g盐分成50g和11g两部分，即可得到一堆50g，一堆90g的盐。

（4）一袋装有每个球质量为45克的袋子，9袋装有每个球质量为50克的袋子,两类袋子混在一起。为了快速跳出45克球的袋子,商贩用标准秤仅称量一次就达到了目的,请简要说明该商贩的操作过程？

答：依次以1到10将袋子编号。  
从1号袋子中取1个球,从2号袋子中取2个球……从10号袋子中取10个球。  
一起将这55个球称重,看重量与2750克少多少,将少的重量除以5,得到的数就是装有45克球的袋子号码。  
如得到的重量为2700克,2750-2700=50,50/5=10,10号袋子就是45g的球。

**4、过河问题**

在漆黑的夜里，四位旅行者来到了一座狭窄而且没有护栏的桥边。如果不借助手电筒的话，大家是无 论如何也不敢过桥去的。

不幸的是，四个人一共只带了一只手电筒，而桥窄得只够让两个人同时过。

如果各自单独过桥的话，四人所需要的时间分别是1、2、5、8分钟；而如果两人同时过桥，所需要的时间就是走得比较慢的那个人单独行动时所需的时间。

如何设计一个方案，让这四人尽快过桥？

答: 此类问题的要点在于将过桥最快的两人放在一起、过桥最慢的两人放在一起，使得总时间最少，给出以下方案：

给耗时1、2、5、8分钟的人分别编号1、2、3、4号。

1、2号同时过桥，耗时2分钟，此时总共耗时2分钟。

1号回来，耗时1分钟，此时总共耗时1+2=3分钟。

3、4号同时过桥，耗时8分钟，此时总共耗时8+3=11分钟。

2号回来，耗时2分钟，此时总共耗时11+2=13分钟。

最后1、2号同时过桥，耗时2分钟，此时总共耗时13+2=15分钟。

**参考资料：**

微信公众号：阿狸和小兔。

2020年5月7日发布文章：什么？数分也有可能考智力题？