抽样

一、抽样的意义与作用
二、概率抽样的原理与程序
三、概率抽样方法
四、非概率抽样方法
五、样本规模与抽样误差

抽样

一、抽样的意义与作用

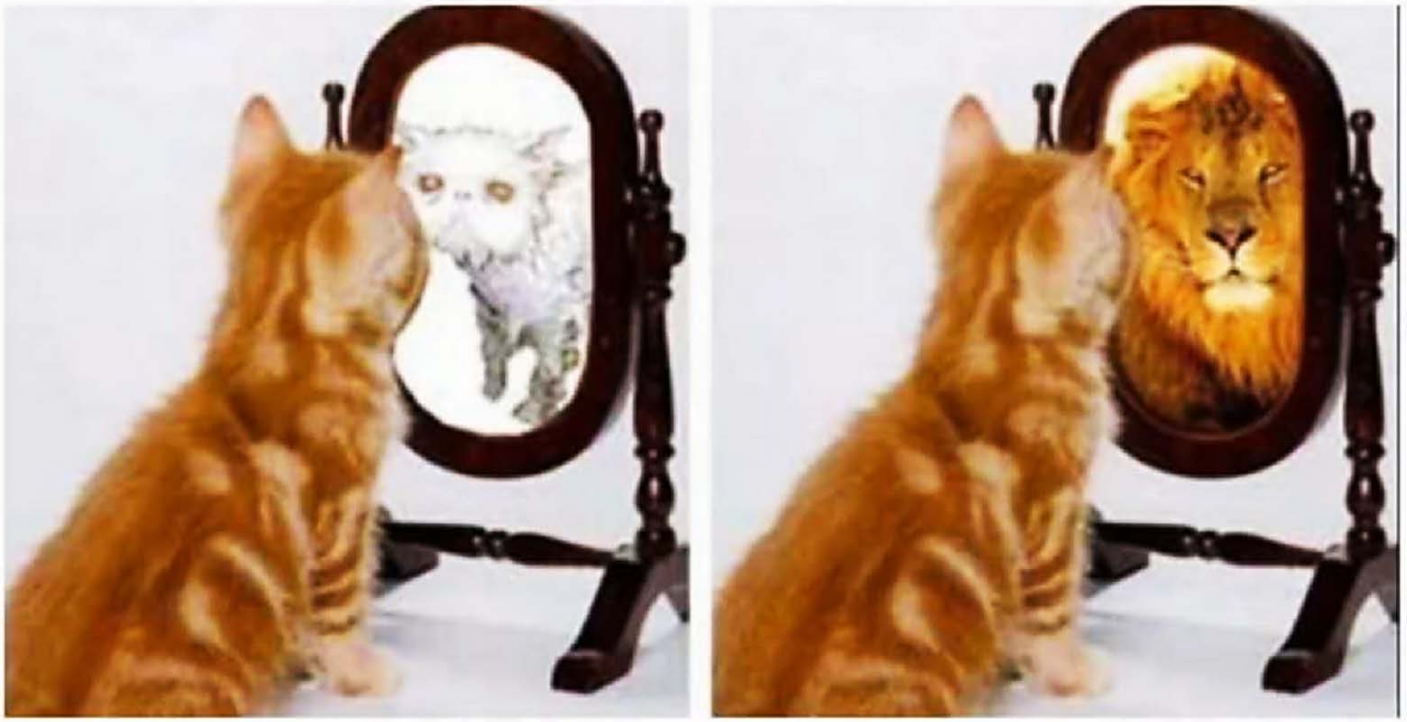
二、概率抽样的原理与程序

三、概率抽样方法

四、非概率抽样方法

五、样本规模与抽样误差

为什么需要研究方法：有数据，就有真相？



错误的分析误人误己

某年的高考作文题:
亚伯拉罕.瓦尔德

- 二战时候美国军方做了一番调查，统计了飞机各部位的中弹数量。
- 因为飞机对重量很敏感，没办法搞成坦克那个铁疙瘩样儿，所以，装甲必须加在最需要的地方。

飞机部位	每平方英尺[①]的平均弹孔数
引擎	1.11
机身	1.73
油料系统	1.55
其余部位	1.80

- A、机身最需要
- B、引擎最需要

选择性偏误

- 在这个简单的例子中，我们使用了一个错误的理论分类器，结果造成了选择偏差(统计上称为内生选择偏差)，最终导致了虚假相关。

- 这样的分析、报告和研究到处可见，常常看到。

案例：肥胖悖论(obesity paradox)

- 2017年12月的《J Neurosurg》，Neha S. Dangayach将患者按体质指数(BMI)分为两组:
  - 132例为BMI ≥ 225kg/m2的超重组
  - 70例为BMI<25kg/m2的非超重组

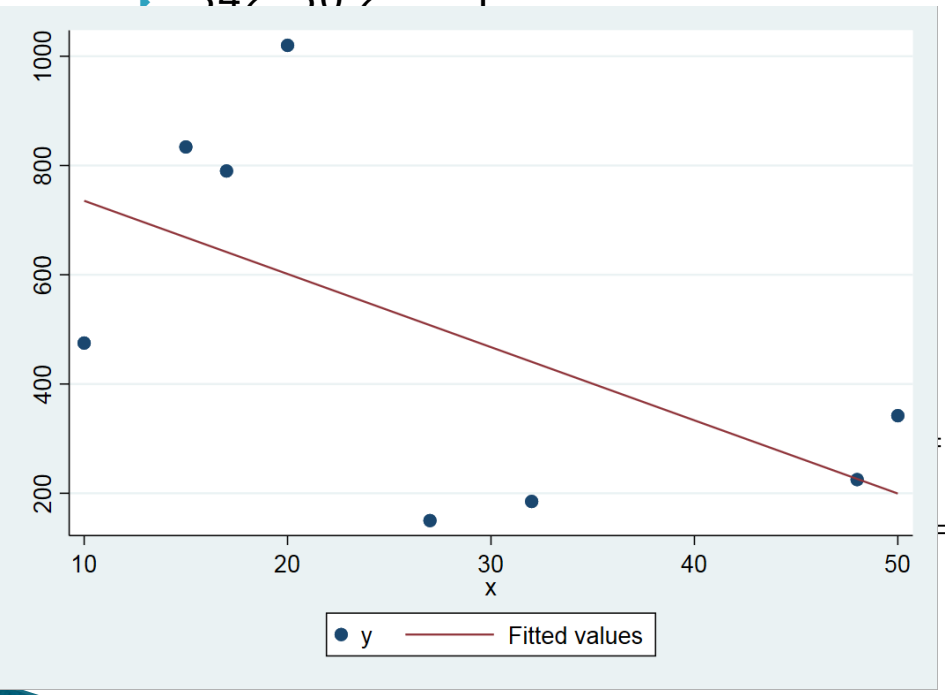
- 入住重症监护病房(ICU)诊治的超重患者的存活期可提高1年。所以，肥胖的人，反而更长寿？

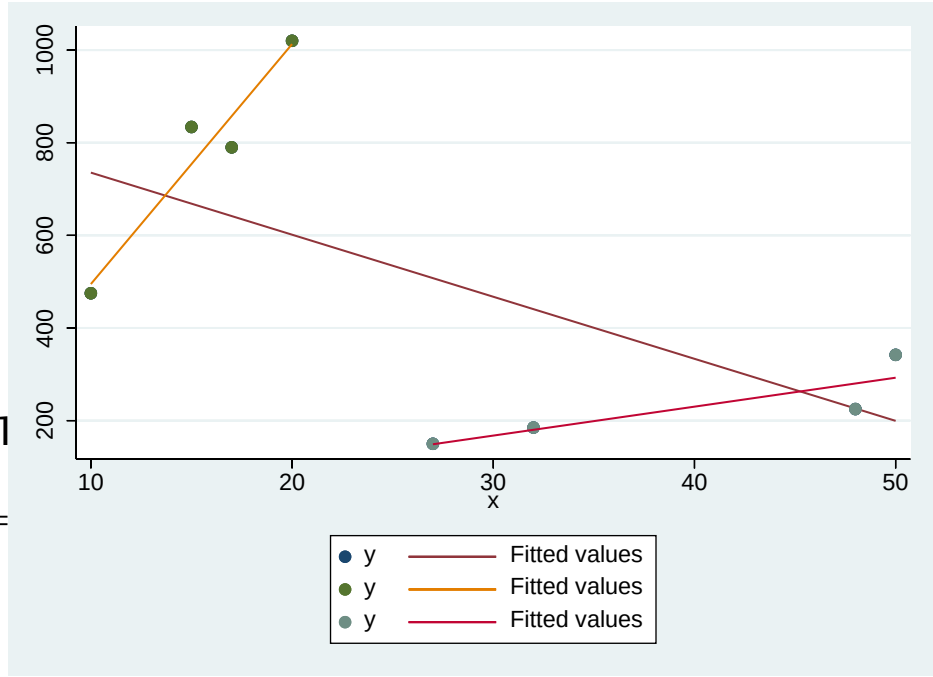
样本选择性偏误+数据分析偏误造成了肥胖悖论

- ●132例为BMI≥225kg/m2的超重组——都是更健康的胖子，那些不健康的胖子已经死了
- ●70例为BMI<25kg/m2的非超重组——都是更不健康的瘦子，那些健康的瘦子进不来
- BMI是不好的分类器，健康水平才是好的分类器。当人们的健康水平被看做是影响因素时，“肥胖悖论”也就消失了。

辛普森悖论的机制模拟

- clear
- input y x I D
- 150 27 2 1
- 225 48 2 1
- 342 50 2 1





一、抽样的意义与作用

（一）抽样的概念

- 例子
- 某省2018年共有12.8万名在校大学生，为了调查大学生的就业倾向，现在从12.8万人里抽取1000人进行调查

（一）抽样的概念

1、总体

- 总体(population)通常与构成它的元素共同定义：总体是构成它的所有元素的集合，而元素(element)则是构成总体的最基本单位。
- 例如，当我们对某省大学生的择业倾向进行研究时，该省12.8万名在校大学生的集合就是我们研究的总体，而每一个在校大学生便是构成总体的元素。