

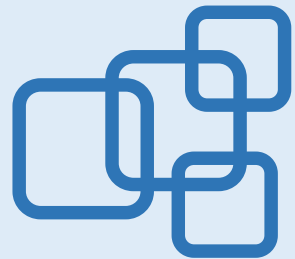
带你走入统计学

统计与数据科学学院 马俊玲



专题五

数据分析的推断思维

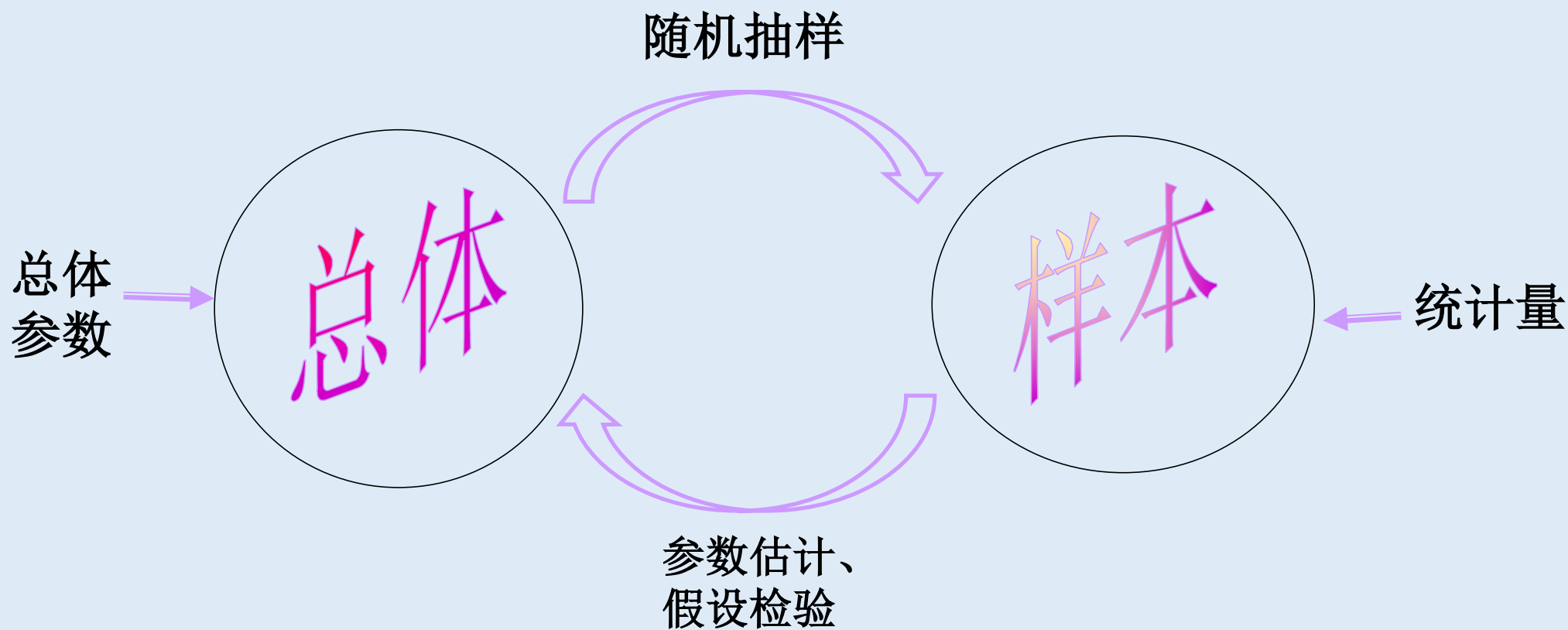


专题五 数据分析的推断思维

问题：你去某地区旅游，看到当地一个成年男子甲的身高为175cm，一个成年女子乙的身高为170cm，你会有什么印象？女的高吗？男的高吗？就这两位比较谁高？



统计推断：利用**样本统计量**对**总体某些性质或数量特征**进行推断。



专题五 数据分析的推断思维

- 参数估计(Estimation)的思维
 - 矩法思维
 - 最可能性（最大似然法）思维
- 假设检验(Hypothesis Testing)的思维
 - 小概率事件思维

一、参数估计思维

- 总体：2024年高考学生身高（性别） 样本：被抽到同学
- 总体均值 $\hat{\mu} = \bar{X}$ （用被抽到同学的平均身高来反映总体平均身高）
- 总体比例 $\hat{p} = \frac{m}{n}$ （用被抽到同学的性别比例来反映总体性别比例）
- 总体方差 $\hat{\sigma}^2 = S^2$ （用被抽到同学的身高方差来反映总体方差）

矩法思维：对总体参数代表的某个特征通过样本数据构建对应统计量直接估计。

矩法思维案例

某地区制衣公司在该地区随机测量了15200个人的身高数据，其均值与标准差如下：

成年男子(5115人)		成年女子(5507)	
均值	标准差	均值	标准差
167.48	6.09	156.58	5.47

基于以上数据，我们假设成年男子与成年女子身高的分布为：

$$X \sim N(167.48, 6.09^2), \quad Y \sim N(156.58, 5.47^2)$$

由此分布，我们有：

$$P\{X \geq 175\} = 0.108, \quad P\{Y \geq 170\} = 0.007.$$

● 矩法思维案例

思考：

基于以上数据，我们假设成年男子与成年女子身高的分布为：

$$X \sim N(167.48, 6.09^2), \quad Y \sim N(156.58, 5.47^2)$$

抽样分布思维

最可能性（最大似然法）思维



根据已知信息推断最可能的情况

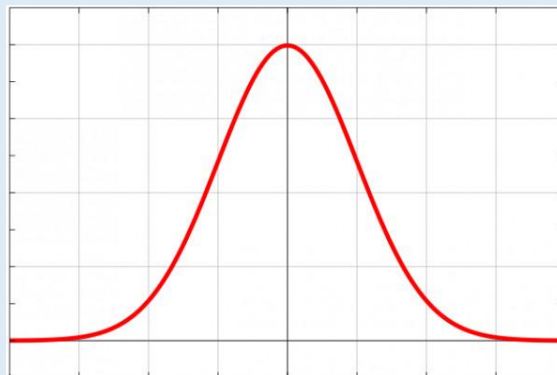
最可能性（最大似然法）思维

回到身高问题：

成年男子(5115人)		成年女子(5507)	
均值	标准差	均值	标准差
167.48	6.09	156.58	5.47

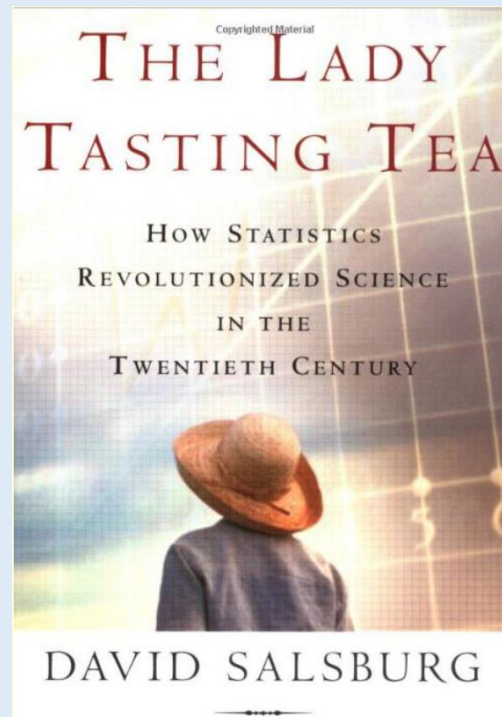
基于以上数据，我们假设成年男子与成年女子身高的分布为：

$$X \sim N(167.48, 6.09^2), \quad Y \sim N(156.58, 5.47^2)$$



二、假设检验的思维（小概率事件思维）

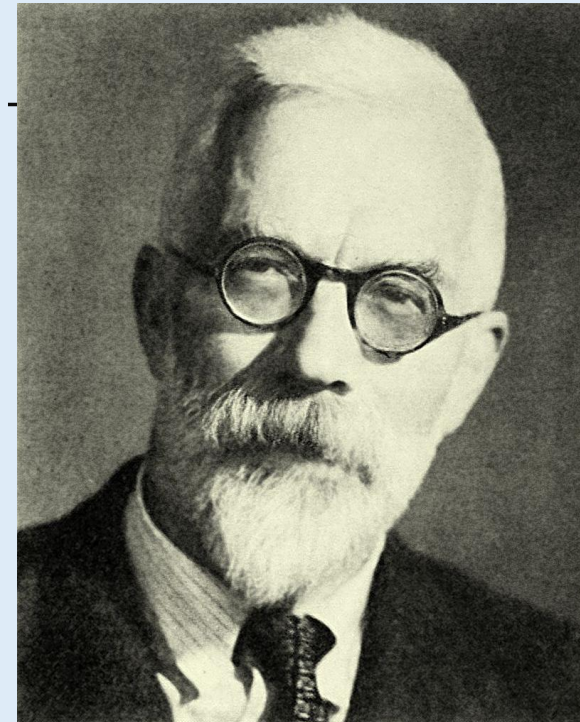
● 案例-女士品茶试验



Salsburg, David (2001). The Lady Tasting Tea, W.H. Freeman and Company

小概率事件思维

● 案例-女士品茶试验



20世纪20年代，在英国剑桥的午后茶，一位女士坚称：把茶加进奶里，或把奶加进茶里，不同的做法，会使茶的味道品起来不同。在场的绝大多数大学教授对这位女士的“胡言乱语”嗤之以鼻。然而，在座的一个身材矮小、戴着厚眼镜、下巴上蓄着的短尖髯开始变灰的先生，却不这么看，他对这个问题很有兴趣。

小概率事件思维

- 第一步：Fisher首先引进一个假设：

H: 此女士无鉴别力

- 第二步：给她8杯(茶奶、奶茶各4杯)。当假设H成立时，以X记此女士说对茶奶的杯数，则由概率论知

$$P\{X = k\} = \frac{\binom{4}{k} \binom{4}{4-k}}{\binom{8}{4}}, \quad k = 1, 2, 3, 4.$$

- 第三步：由上述分布知，全部说对的概率为1/70

数学

统计

实验结果，该女士说对了8杯。

小概率事件思维

小概率事件



刘晓琳
阜阳人民医院儿科主任

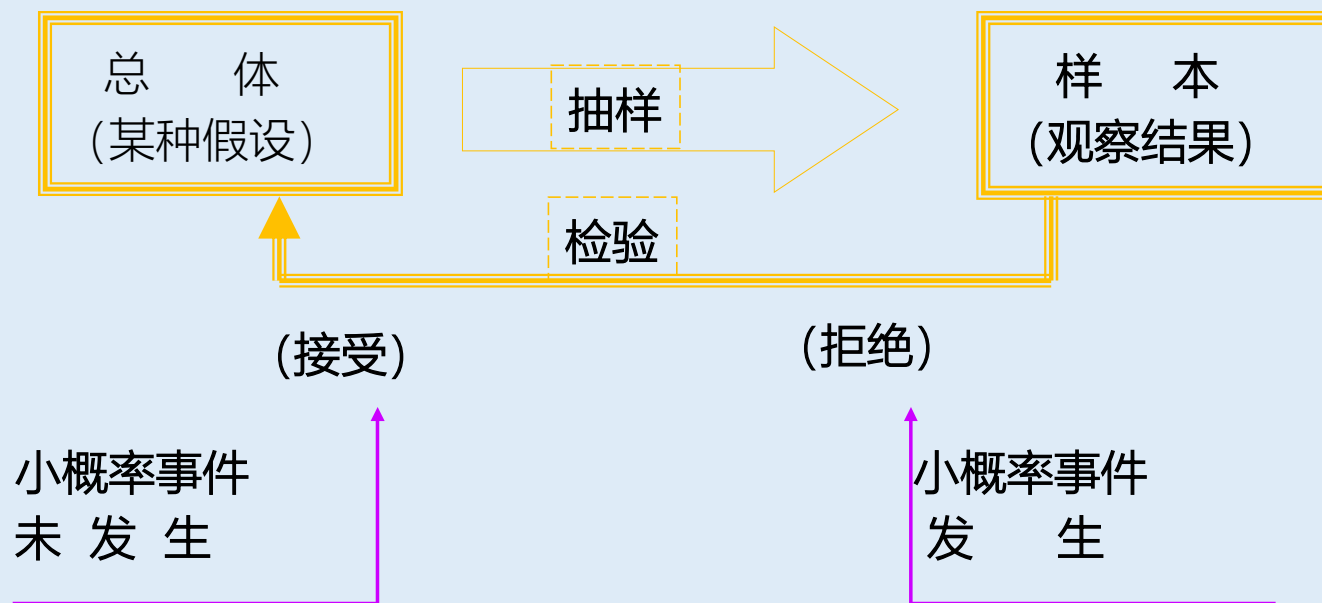
- 2004年上半年发现“大头娃娃”事件，揭开劣质奶粉，挽救了无数婴儿。
- 2008年3月27-28日有三位儿童因“肺炎”去世。虽然症状与肺炎一样，但常规肺炎大多是左心衰竭，而这三个孩子却是右心衰竭。刘主任认为极不正常，及时向上级汇报。4月23日，经专家诊断，确定死因为手足口病。

小概率事件思维

Hypothesis Test

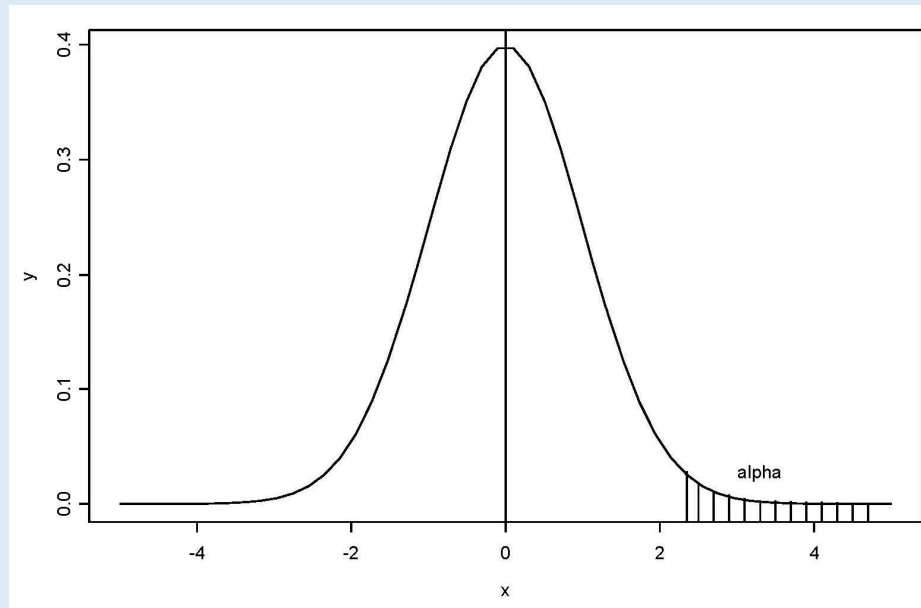
- 对于女士品茶实验，如何由“8杯都说对了”这一实验结果来判断此女士是否具有鉴别力呢？
 - ✓ 如果取小概率临界值为0.01，则由于 $1/70=0.014>0.01$ ，故我们只能说发生了一件概率为 $1/70$ 的随机事件
 - ✓ 如果取小概率临界值为0.05，则由于 $0.014<0.05$ ，则Fisher认为“8杯都说对了”提供了非常不利于前提假设的显著性证据
 - ✓ 上述过程就称为显著性假设检验，且小概率临界值称为显著性水平，此值需要在实验之前确定，而不能有了数据后再给出。

小概率事件思维用于假设检验统计推断



- 小概率原理：认为在随机的一次试验中小概率事件是不可能发生的。

小概率事件



分位数(quantile)

思考与练习

1. 思考小概率事件思维在日常生活中的体现。
2. 下列链接文章列举了小概率事件在法官断案中的误用事实案例。请认真阅读思考。
<https://www.163.com/dy/article/DRIM562L05129C83.html> 统计学家改判糊涂案
3. 第二次世界大战时期，德国物资特别紧缺，为了控制物资，政府把面粉发给指定的面包房，面包师傅每天将面粉烤成重量为500g的面包再发给居民。有居民发现面包经常重量不够500g，怀疑面包师傅克扣面粉，可以构建小概率事件用小概率思维分析这个问题吗？



THANK YOU!

