

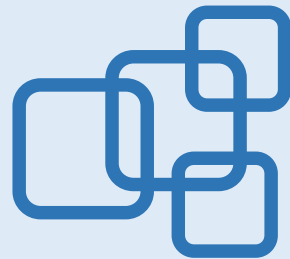
带你走入统计学

统计与数据科学学院 马俊玲



专题三

从实践问题到统计量



专题三 从实践问题到统计量

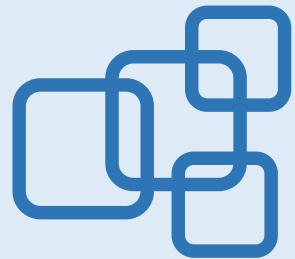
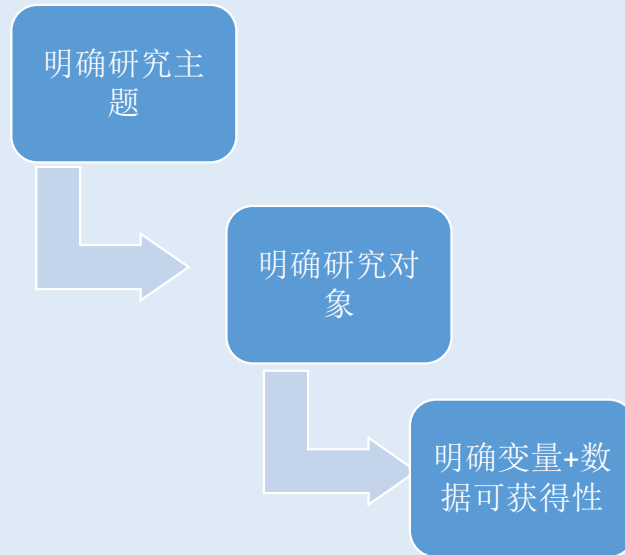
- 一、案例—富人是否变得更富了？
- 二、认识统计量

一、案例—富人是否变得更富了？

（一）案例问题

在“互联网泡沫”的最高峰时期，科技股的价格飙升到了一个前所未有的水平，微软的创始人比尔·盖茨以超过1000亿美元的身家成为世界首富。除了最富裕的18个国家，比尔·盖茨超过了全世界余下国家的国民生产总值之和，不久比尔·盖茨的身家就缩水了，一方面是由于微软股票价格的下跌，另一方面由于他把他所有股份捐给了慈善机构。然而，媒体仍然在持续不断地曝光那些巨富的故事，似乎富人变得更富有了，而我们远远地落在后面。事实真的如此吗？

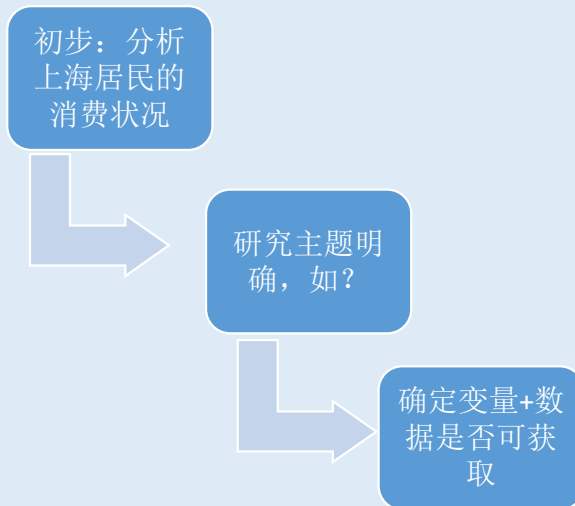
(一) 案例问题



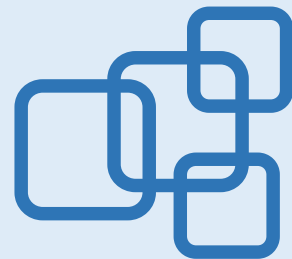


(一) 案例问题

讨论:



<https://tjj.sh.gov.cn/sjfb/index.html>



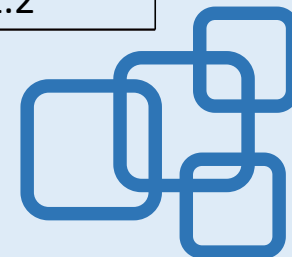


(二) 案例分析

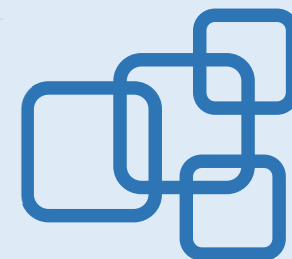
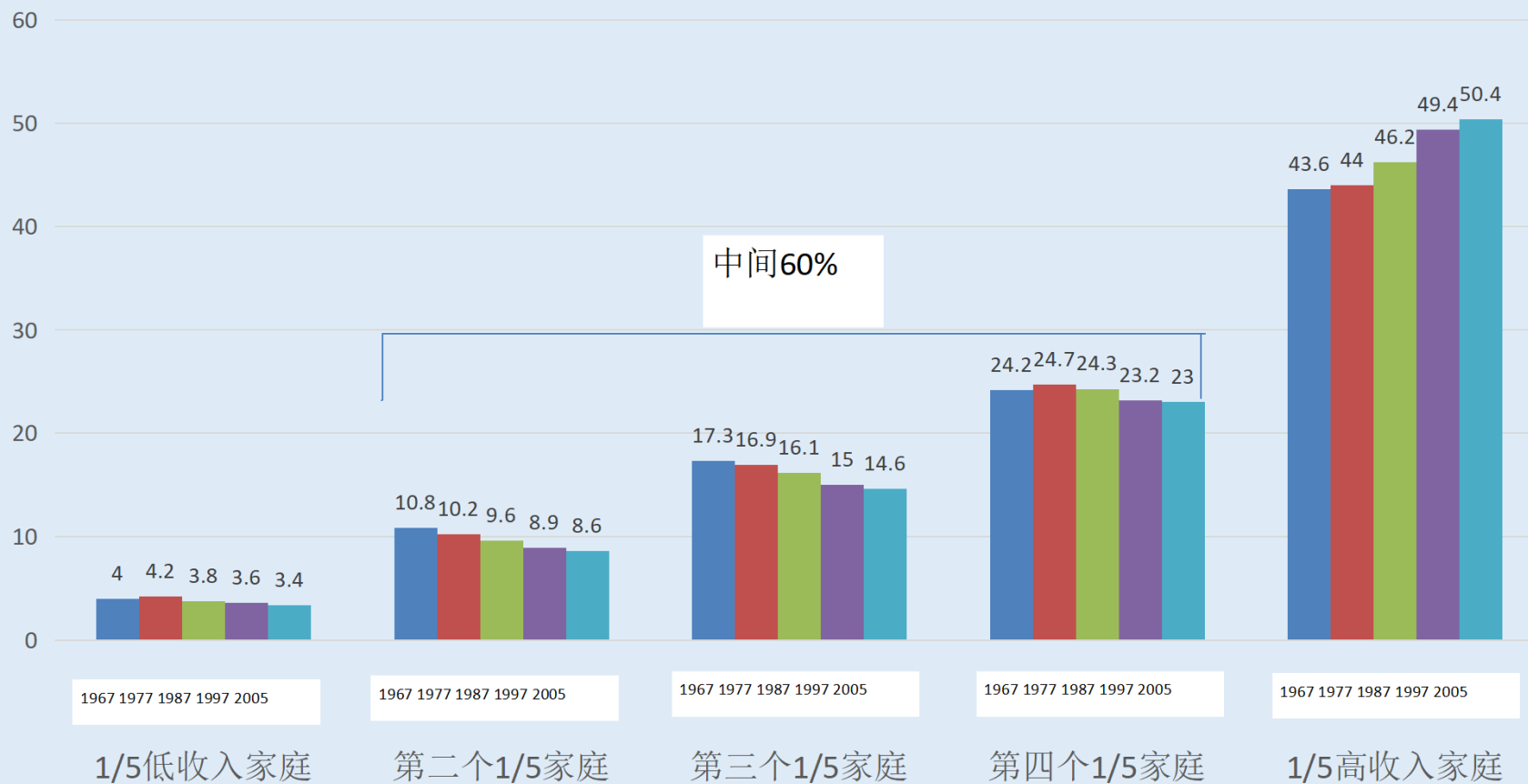
5分法，即按收入高低区分为5组人群。通常在收入最高的一组中进一步区分出收入最高的5%的群体与其他人群进行比较。

表3.1 按收入高低划分的5组人群分别占家庭总收入百分比

	1/5低收入家庭	第二个1/5家庭	第三个1/5家庭	第四个1/5家庭	1/5高收入家庭	5%最高收入家庭
1967	4	10.8	17.3	24.2	43.6	17.2
1977	4.2	10.2	16.9	24.7	44	16.8
1987	3.8	9.6	16.1	24.3	46.2	18.2
1997	3.6	8.9	15	23.2	49.4	21.7
2005	3.4	8.6	14.6	23	50.4	22.2



(二) 案例分析

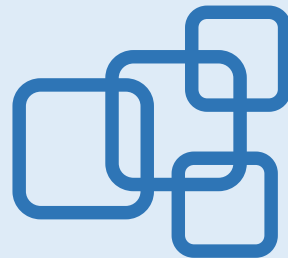


（二）案例分析



数据可视化分析结果：

可以发现，除了最富裕的20%群体之外的全部群体，所享有的社会收入份额从1967年以来一直在下降。与此同时，最富有的20%的群体所占的收入份额却在稳步增长，与那5%的最富有群体一样。

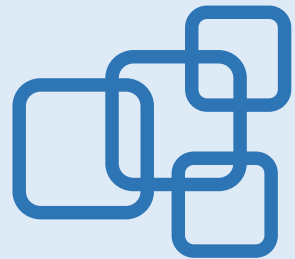


(三) 案例结论



结论：

与普通公众相比，富人的确是变得更富了。



(四) 对案例结论的讨论



Vilfredo Pareto
(1848-1923)
意大利经济学家



$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\bar{X}_w = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i X_i$$

十九世纪末，20世纪初，帕里托发现：社会上20%的人占有80%的社会财富



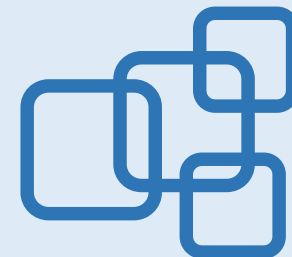
(四) 对案例结论的讨论

帕累托法则——二八定律

1897年，意大利经济学者巴莱多偶然注意到19世纪英国人的财富和收益模式。在调查取样中，发现大部分的财富流向了少数人手里。同时，他还从早期的资料中发现，在其他的国家，都发现有这种微妙关系一再出现，而且在数学上呈现出一种稳定的关系。

于是，帕累托从大量具体的事实中发现：**社会上20%的人占有80%的社会财富**，即：财富在人口中的分配是不平衡的。

同时，人们还发现生活中存在许多不平衡的现象。因此，**二八定律成了这种不平等关系的简称**。习惯上，二八定律讨论的是顶端的20%，而非底部的80%。



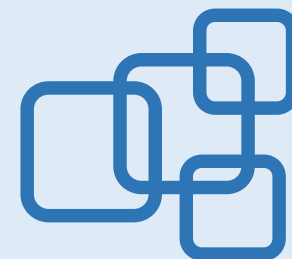


（四）对案例结论的讨论

帕累托二八效率（最优）法则：

只要在让一部分人变得更好的同时并没有让其他人变得更糟，那么这种改变就是好的。

不断增长的美国经济是符合帕累托法则的，因为尽管绝大多数人群在收入份额中所占的比例要比过去要小，但是每个人的绝对财富以及生活条件要比过去要好。

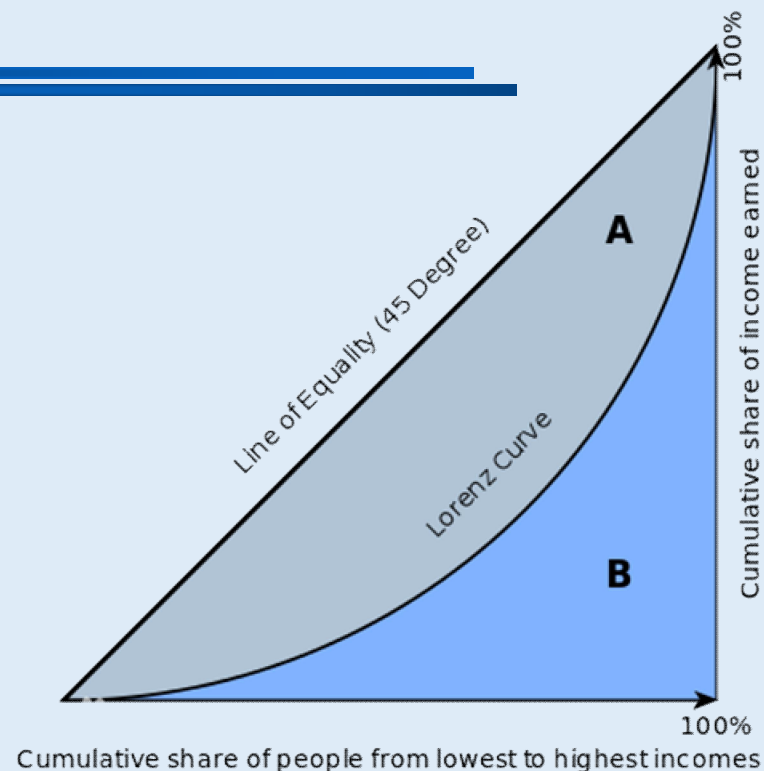


二、认识统计量

(一) 统计指标—以基尼系数为例

基尼系数用来描述收入分配均衡性。

洛伦兹曲线（Lorenz curve），也译为“劳伦兹曲线”。就是，在一个总体（国家、地区）内，以“最贫穷的人口计算起一直到最富有人口”的人口百分比对应各个人口百分比的收入百分比的点组成的曲线。（为了研究国民收入在国民之间的分配问题，美国统计学家（或说奥地利统计学家）M.O.洛伦兹（Max Otto Lorenz, 1876- 1959）1907年（或说1905年）提出了著名的洛伦兹曲线。）



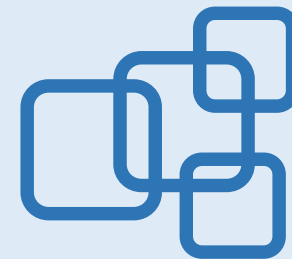
设实际收入分配曲线和收入分配绝对平等曲线之间的面积为A，实际收入分配曲线右下方的面积为B。并以A除以（A+B）的商表示不平等程度。这个数值被称为基尼系数或称洛伦茨系数。

(一) 统计指标—以基尼系数为例



基尼系数，按照联合国有关组织规定：

低于0.2	收入绝对平均
0.2-0.3	收入比较平均
0.3-0.4	收入相对合理
0.4-0.5	收入差距较大
0.5以上	收入差距悬殊

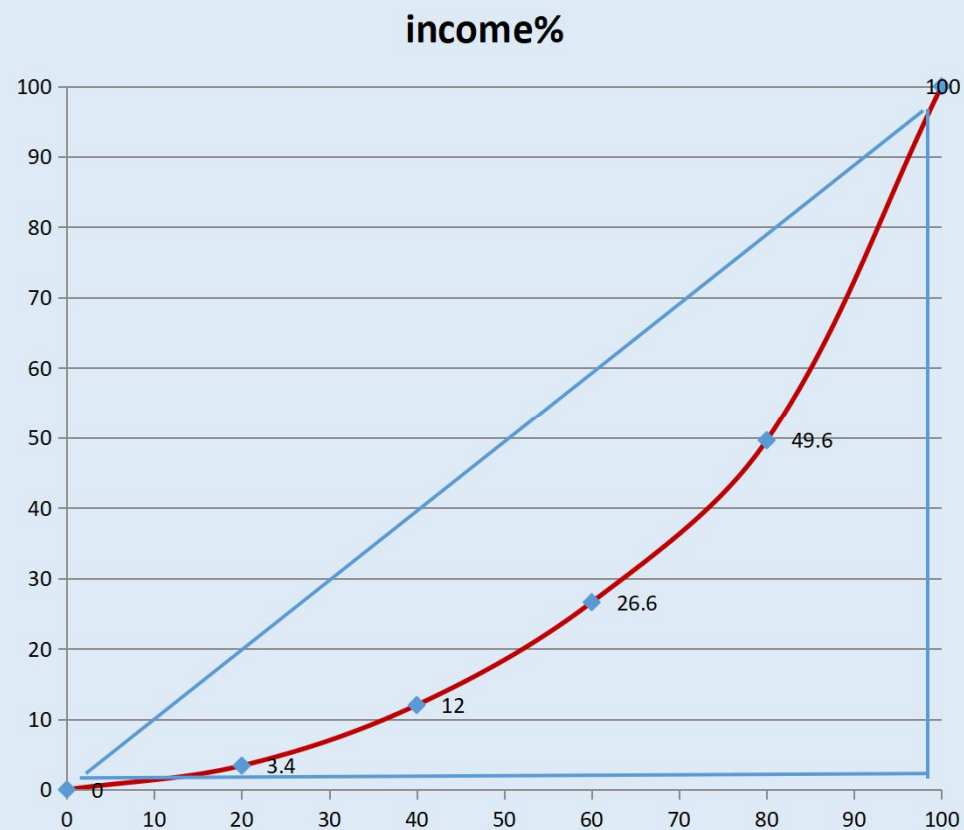


(二) 认识统计量——以基尼系数为例



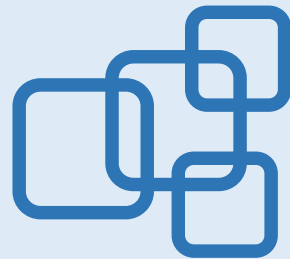
参数——基尼系数

统计量——构造函数估计基尼系数



(三) 更多统计量

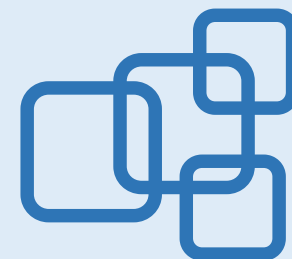
如统计指数, CPI, CCI.





思考与练习

1. 分析我国的收入分配均衡情况，尝试计算基尼系数；
2. 世界上不同发展水平国家的收入分配均衡情况与其发展水平有关吗？可以查阅相关理论研究文献确认可能有关的前提下，再尝试从实证数据去分析。
3. 想想实践中还有哪些统计指数？



3 从实践问题到统计量

<https://www.stats.gov.cn/sj/ndsj/2023/indexch.htm>



6-3 全国居民按东、中、西部及东北地区分组的人均可支配收入

单位：元

组别	2016	2017	2018	2019	2020	2021	2022
东部地区	30654.7	33414.0	36298.2	39438.9	41239.7	44980.3	47026.7
中部地区	20006.2	21833.6	23798.3	26025.3	27152.4	29650.0	31433.7
西部地区	18406.8	20130.3	21935.8	23986.1	25416.0	27798.4	29267.4
东北地区	22351.5	23900.5	25543.2	27370.6	28266.2	30517.7	31405.0

6-2 全国居民按收入五等份分组的人均可支配收入

单位：元

组别	2016	2017	2018	2019	2020	2021	2022
20%低收入组家庭人均可支配收入	5528.7	5958.4	6440.5	7380.4	7868.8	8332.8	8601.1
20%中间偏下收入组家庭人均可支配收入	12898.9	13842.8	14360.5	15777.0	16442.7	18445.5	19302.7
20%中间收入组家庭人均可支配收入	20924.4	22495.3	23188.9	25034.7	26248.9	29053.3	30598.3
20%中间偏上收入组家庭人均可支配收入	31990.4	34546.8	36471.4	39230.5	41171.7	44948.9	47397.4
20%高收入组家庭人均可支配收入	59259.5	64934.0	70639.5	76400.7	80293.8	85835.8	90116.3

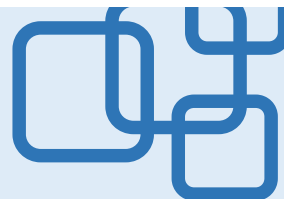


5-5 全市居民人均可支配收入(按收入水平分)(2015-2022年)

单位：元

组别	2015	2016	2017	2018	2019	2020	2021	2022
低收入户 (20%)	18343	20204	22170	23926	25723	25394	27057	27997
中低收入户 (20%)	32968	36277	38452	41886	44971	44855	50226	51844
中等收入户 (20%)	45239	49342	53023	57864	62596	63969	70453	72830
中高收入户 (20%)	60627	65555	71451	77910	85170	88026	94678	97129
高收入户 (20%)	99621	105425	116018	126970	139298	145915	157816	162630

注：全国居民按收入五等份分组是指将所有调查户按人均可支配收入水平从低到高顺序排列，平均分为五个等份，处于最低20%的收入群体为低收入组，依此类推依次为中间偏下收入组、中间收入组、中间偏上收入组、高收入组。本表数据为不同分组家庭的人均可支配收入。6-7、6-12表与此相同。





THANK YOU!

