参数估计

讲师：Jeary

目录

目标

通过本章课程的学习，您将能够：

- 掌握点估计的重要方法
- 掌握区间估计的重要方法

01

点估计

重要方法

- 参数估计
  - 点估计
    - 矩估计法
    - 最大似然估计法
  - 区间估计

原点矩

设k为正整数 (或为0), a 为任何实数, X 为随机变量, 则期望值 $E((X-a)^k)$ 叫做随机变量 X 对 a 的 k 阶矩

如果 $a=0$，则有 $E(X^k)$，叫做 k 阶原点矩，记作:

$$v_k(X)$$

矩估计

矩估计法：

矩估计法的理论依据是大数定律，矩估计是基于一种简单的“替换”思想，即用样本矩估计总体矩

基本思想：

用样本的k阶原点矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 作为总体 $\mu_k = E(X^k)$ 的估计

案例

某炸药厂一天中发生着火的现象的次数X服从参数为λ的而泊松分布，λ未知，有以下样本值:

着火的次数k	0	1	2	3	4	5	6	
发生k次数着火天数n_k	75	90	54	22	6	2	1	$\Sigma=250$

求估计参数λ

最大似然估计

基本思想：

求未知参数使得样本获得取样本值的概率最大

步骤：

设$X_1,\dots,X_n$是来自X的样本， $x_1,\dots,x_n$是对应$X_1,\dots,X_n$的样本值

①写出似然函数：

$$L(x_1,\dots,x_n;\theta_1,\dots,\theta_m)=\prod_{i=1}^{n}P(x_i,\theta_1,\dots,\theta_m)$$ 离散型

$$L(x_1,\dots,x_n;\theta_1,\dots,\theta_m)=\prod_{i=1}^{n}f(x_i,\theta_1,\dots,\theta_m)$$ 连续型

②取对数；

最大似然估计

③对$\theta_1,\ldots,\theta_m$，求偏导数：

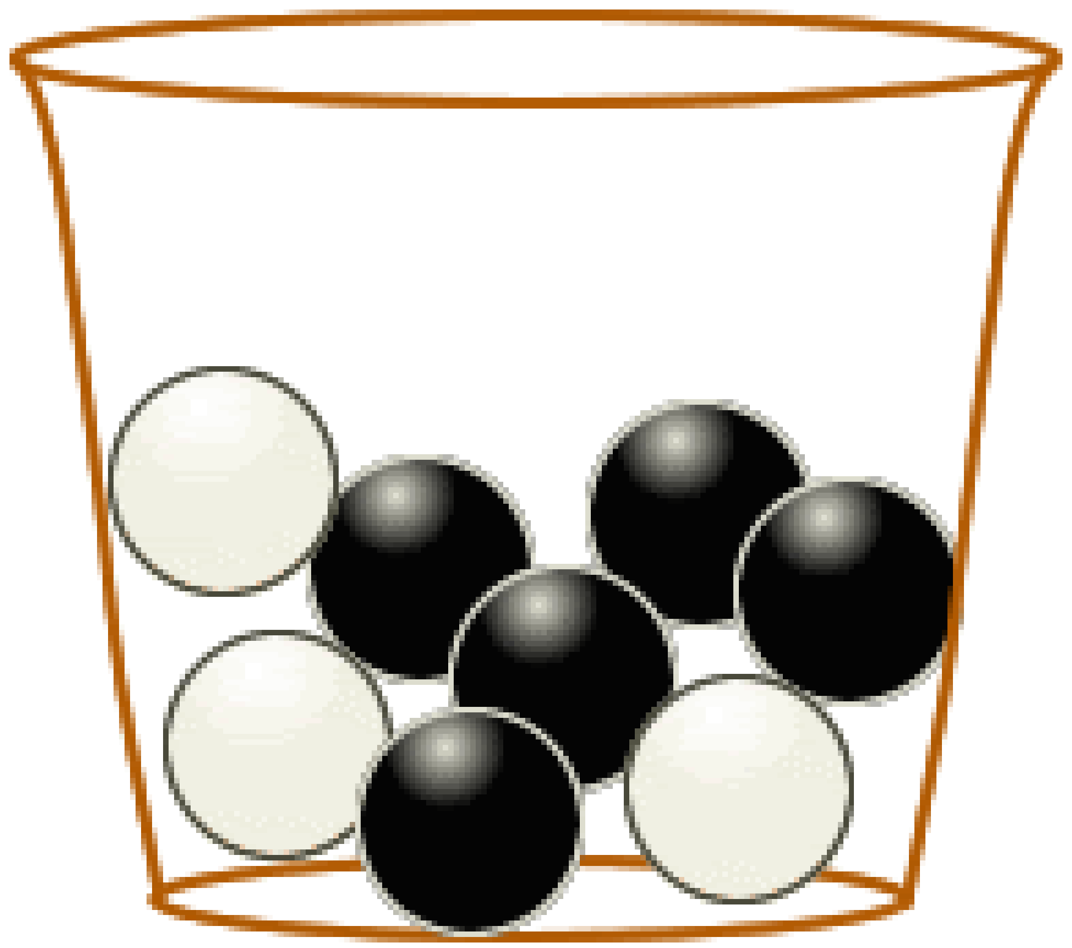
$$\frac{\partial \ln L}{\partial \theta_i}, i=1,\ldots,m$$

④判断方程组$\frac{\partial \ln L}{\partial \theta_i}=0$是否有解；

若有解，则其解即为所求最大似然估计；

若无解，则最大似然估计常在θ_i的边界点上面

案例



有一个罐子装有数目未知、比例未知的黑球和白球，为了搞清楚罐子中黑球和白球的比例，每次任意从摇匀的罐子中拿出球并记录颜色，放回罐中重复该过程。用记录的颜色来估计罐中黑白球比例。若在100次的重复记录中，有70次是白球，则罐子中白球占比最有可能是多少?

案例

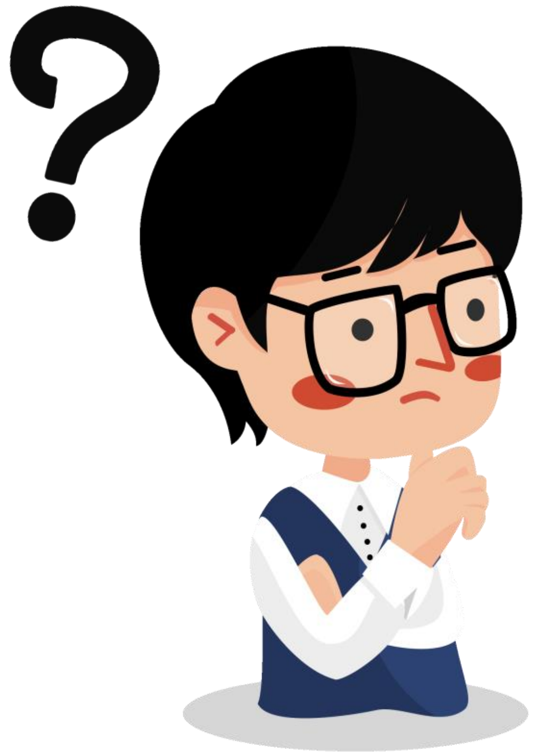
如果第一次抽象的结果记为x_1,第二次抽样的结果记为x_2 ... 那么样本结果为$x_1, x_2, x_3 \ldots, x_{100}$，可以得到如下表达式：

$$\mathrm{P} = (x_1, x_2, x_3 \ldots, x_{100} | \mathrm{Model})$$

$$= \mathrm{P}(x_1 | \mathrm{Mel})\mathrm{P}(x_2 | \mathrm{M})\mathrm{P}(x_{100} | \mathrm{M})$$

$$= (1 - P^{70}(1 - \mathrm{P})^{30}$$

思考题



01 矩估计依据的理论是什么?

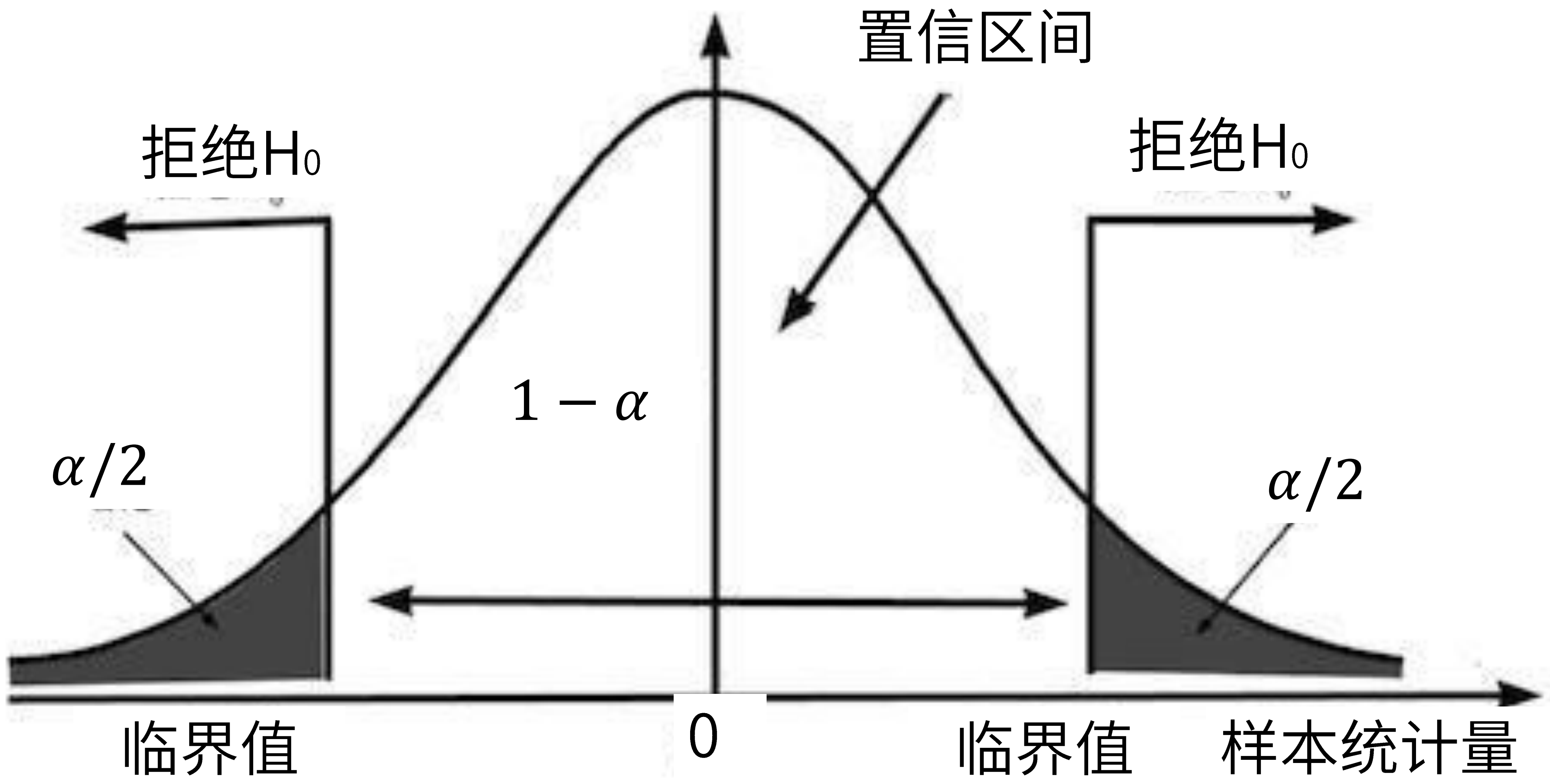
02 最大似然估计的一般步骤是怎样的?

02

区间估计

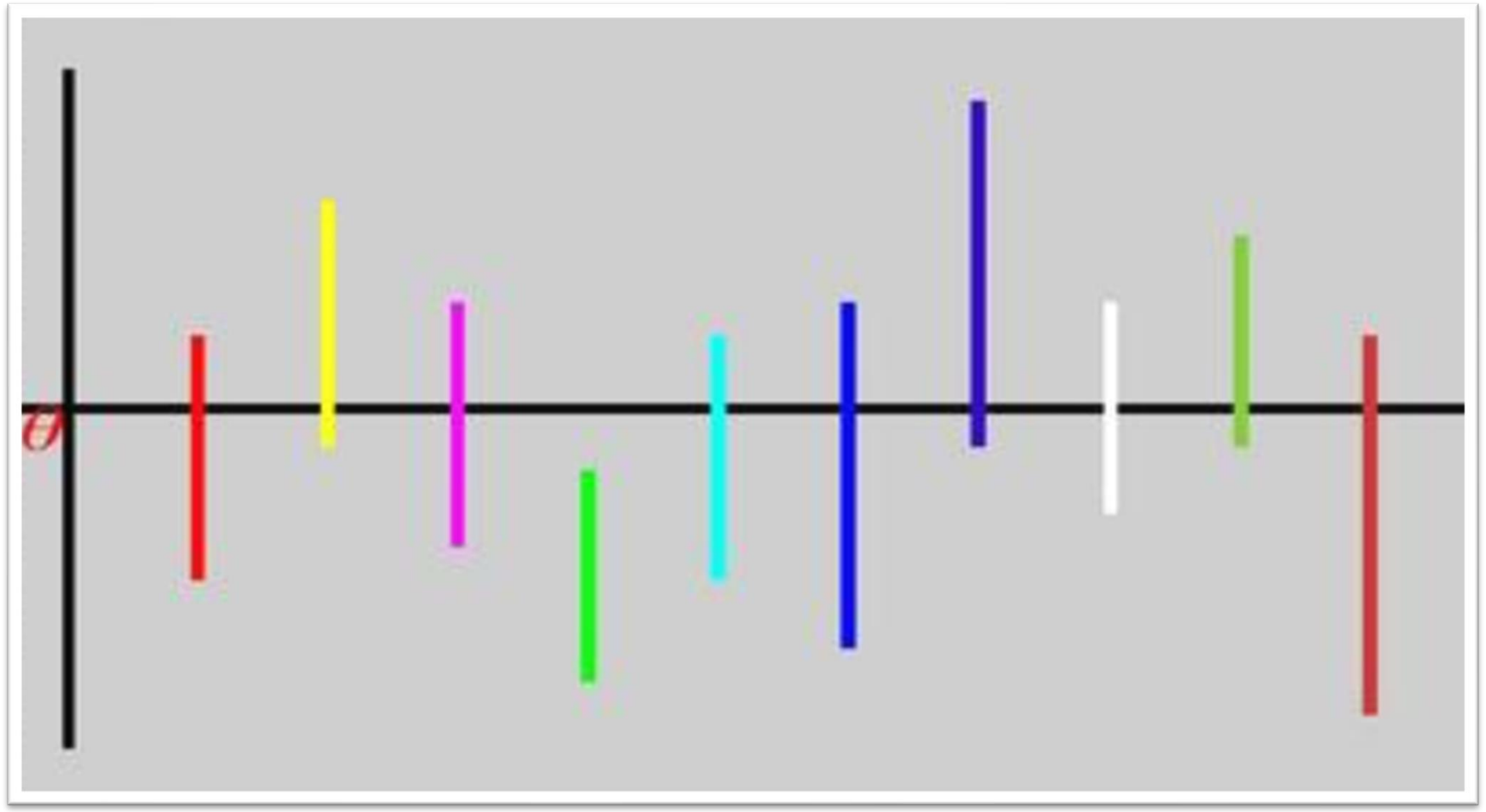
置信区间

由样本统计量所构造的总体参数的估计区间



区间估计

在点估计的基础上，给出总体参数估计的一个区间范围，该区间通常由样本统计量加减估计误差



案例一：刮刮卡

假设只有一个大奖，游戏规则为：

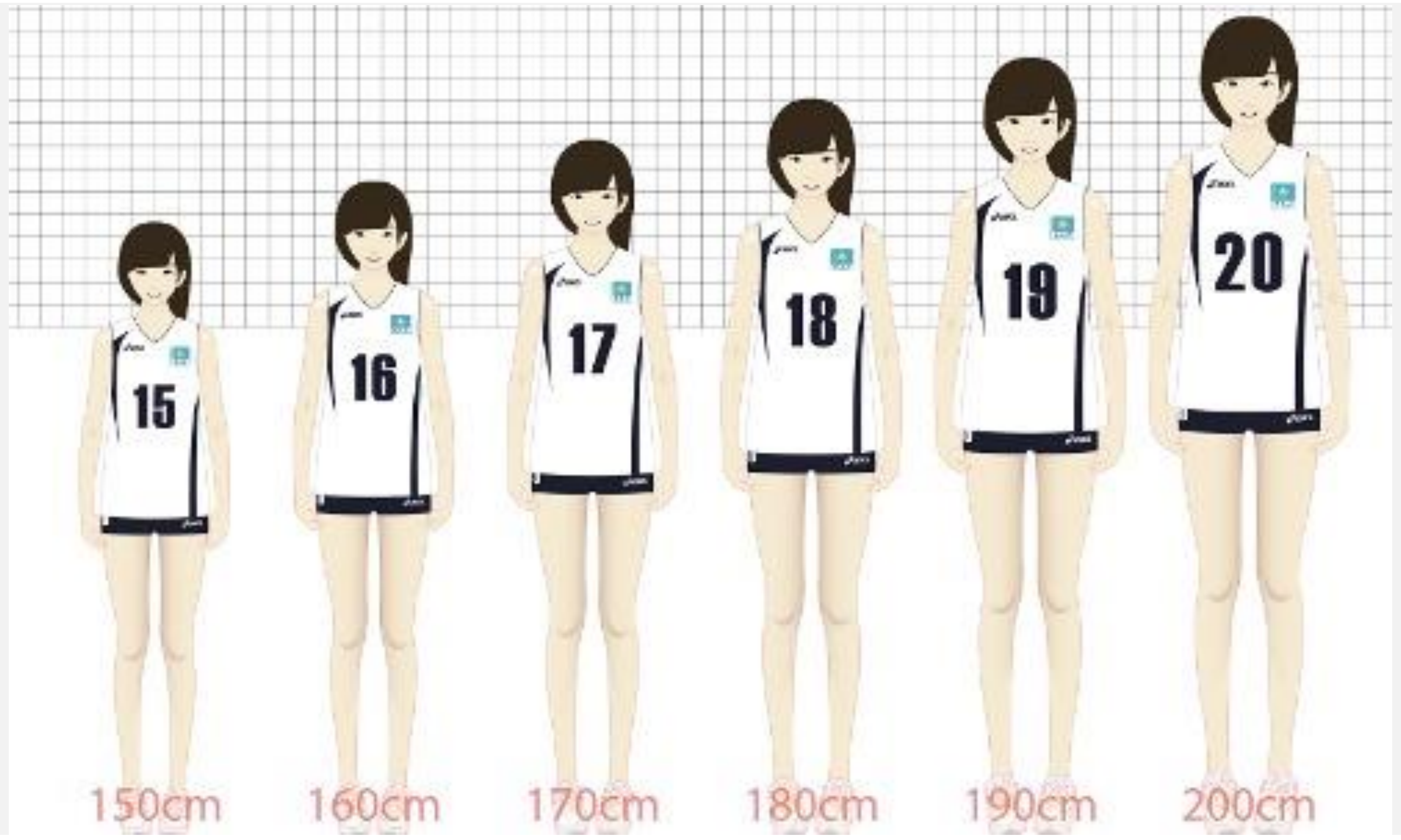
1. 大奖事先固定，一定印在某一张刮刮卡上
2. 购买刮刮卡后，刮开即知道是否中奖

案例一：刮刮卡

点估计：买一张，猜测这一张会中奖

区间估计：买一盒，猜测这一盒里面会有某一张中奖

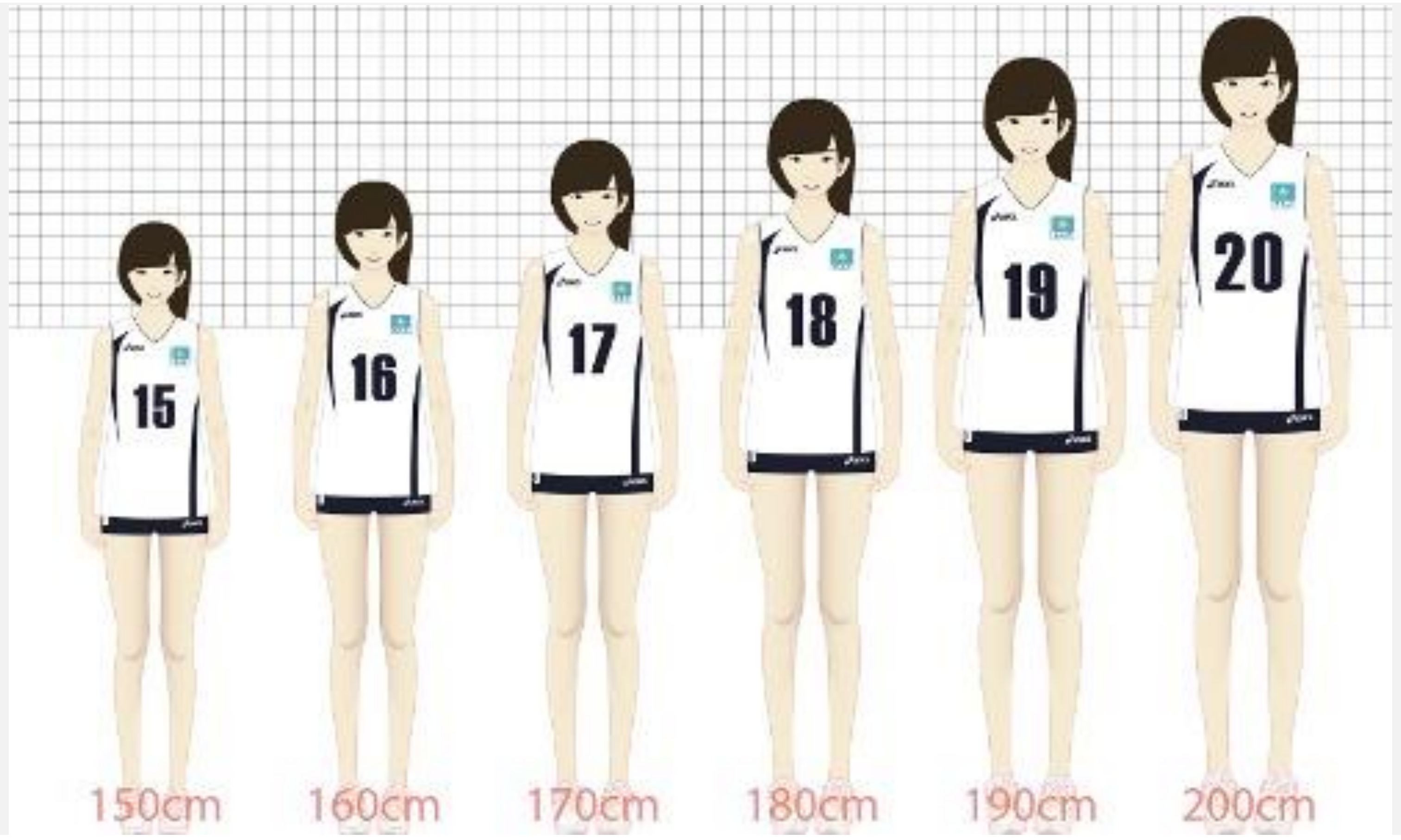
案例二：人的身高

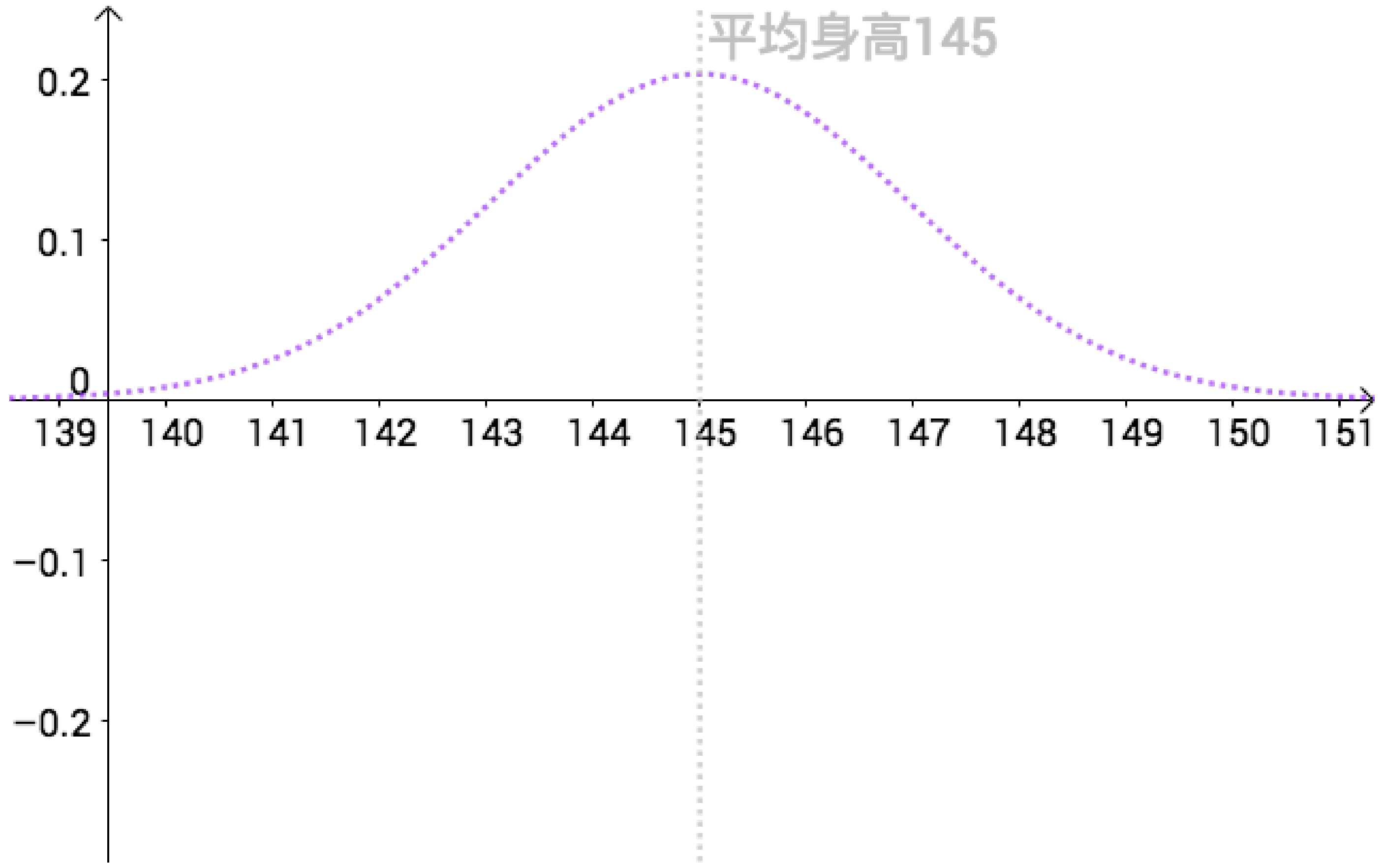


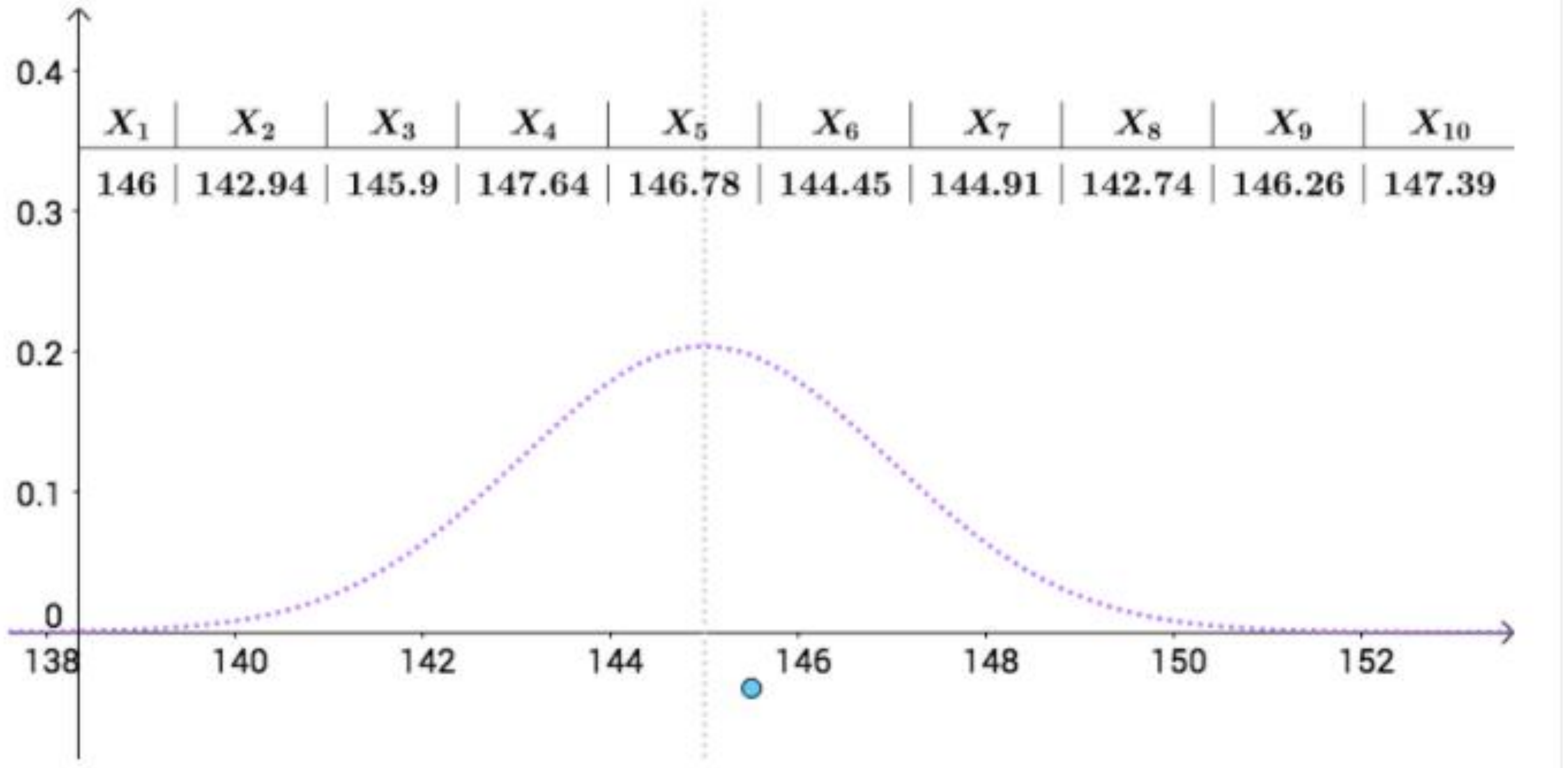
假设人类的身高分布服从正态分布：

$$X \sim N(145, 1.4^2)$$

案例二：人的身高

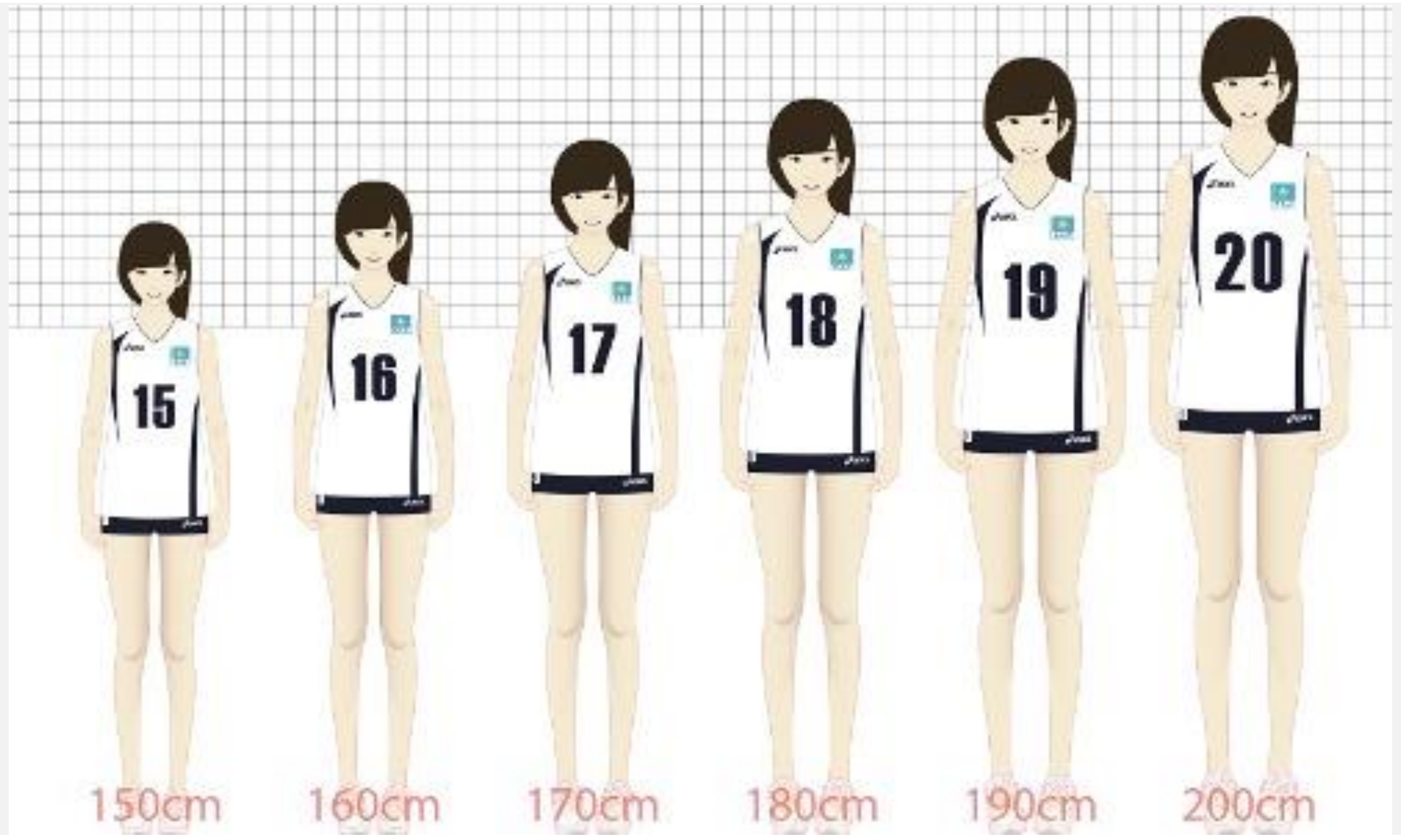


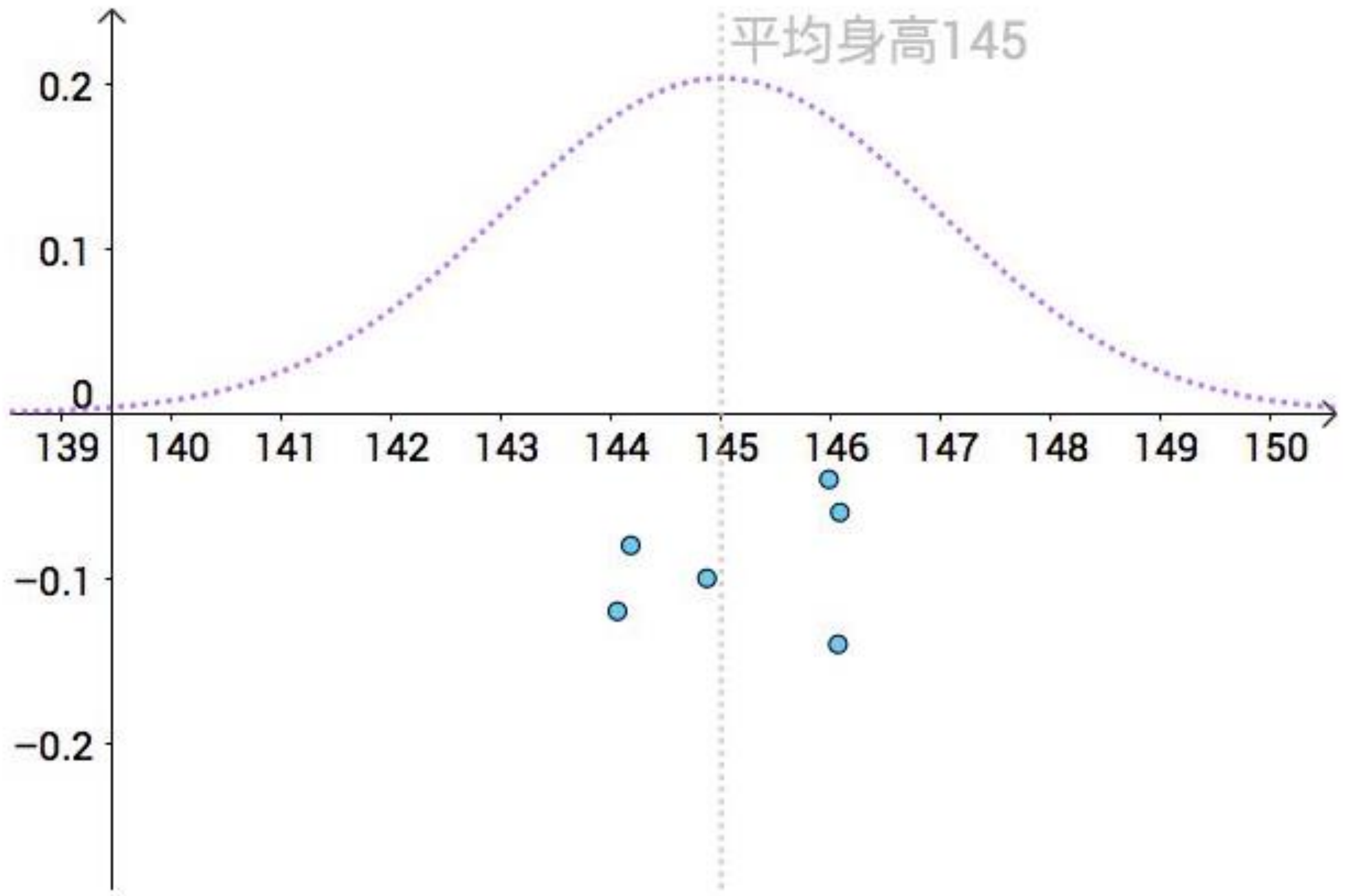


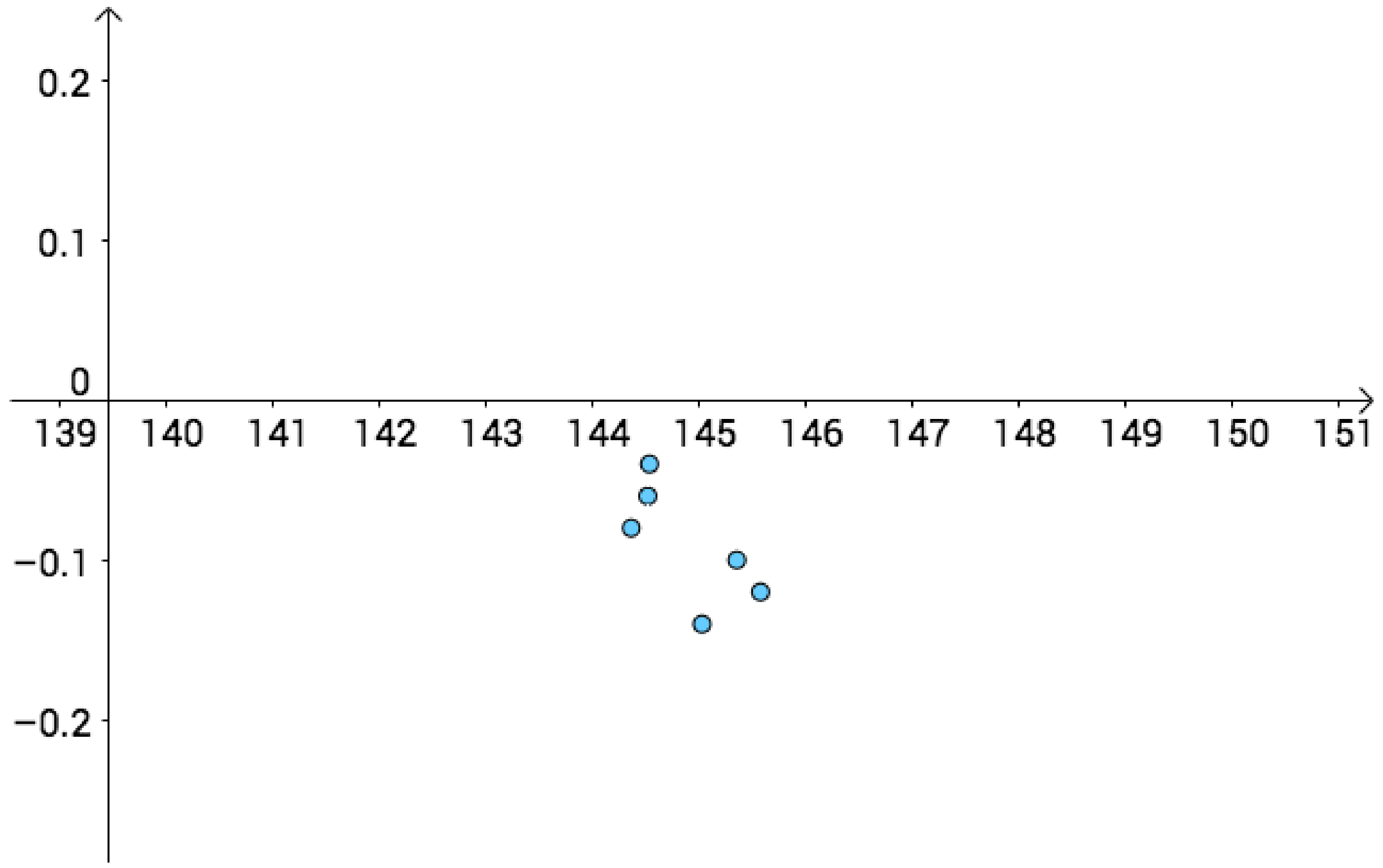


X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
146	142.94	145.9	147.64	146.78	144.45	144.91	142.74	146.26	147.39

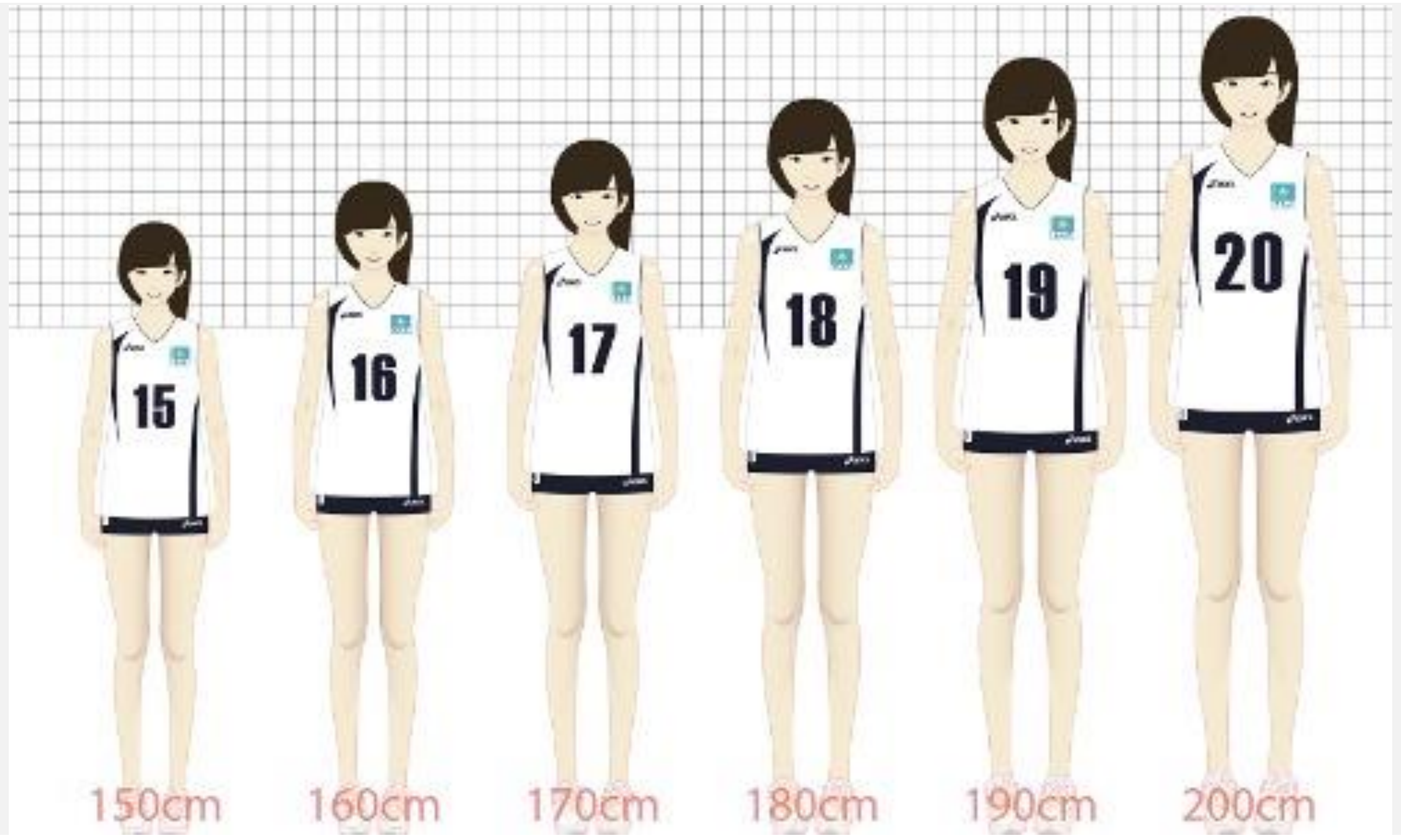
案例二：人的身高


15
16
17
18
19
20
150cm
160cm
170cm
180cm
190cm
200cm

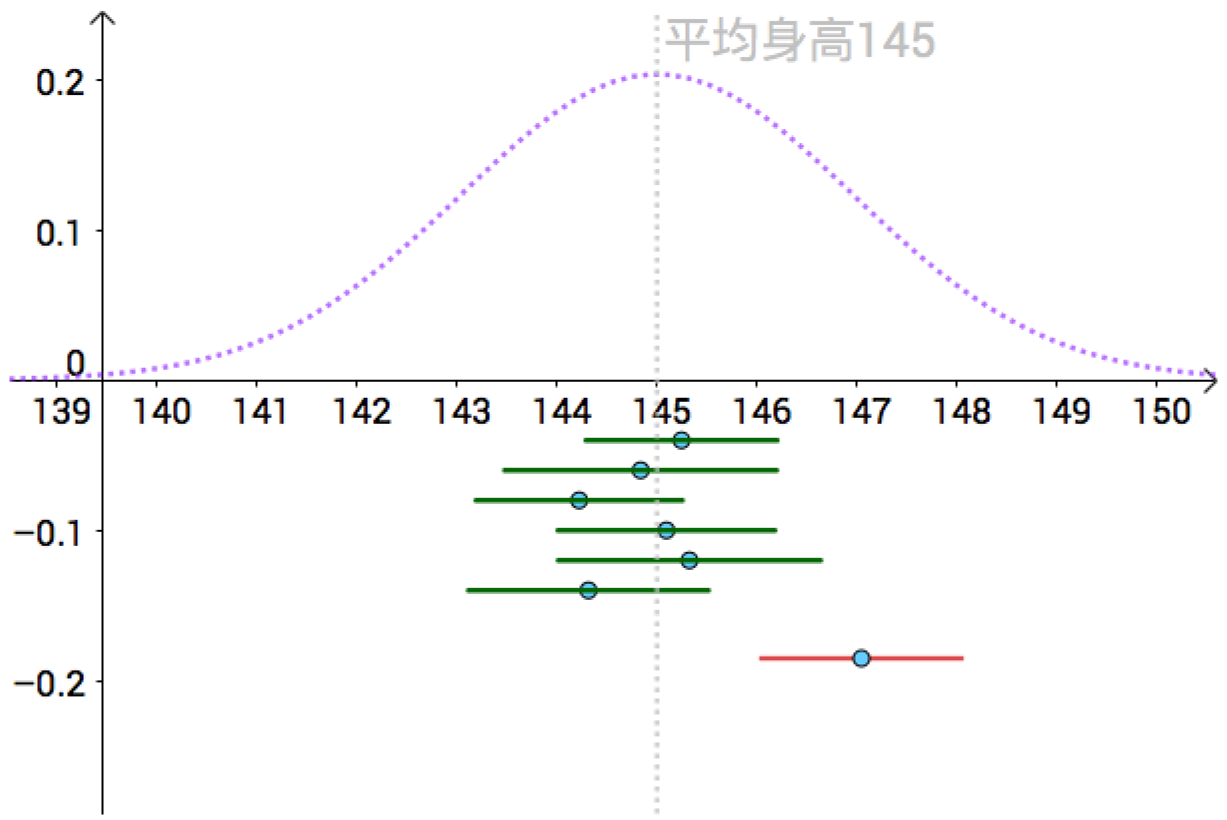

平均身高145
0.2
0.1
0
−0.1
−0.2
139
140
141
142
143
144
145
146
147
148
149
150


0.2
0.1
0
−0.1
−0.2
139
140
141
142
143
144
145
146
147
148
149
150
151

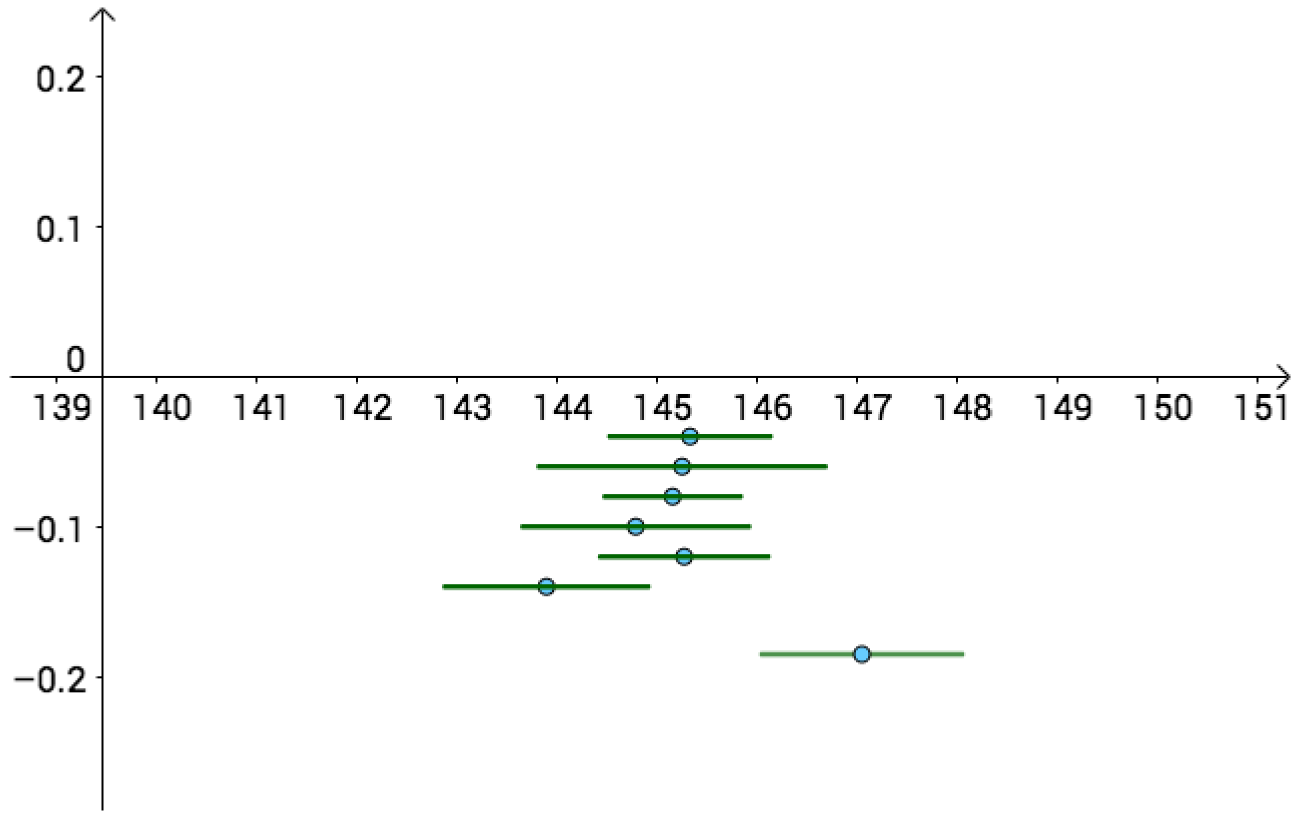
案例二：人的身高



150cm
160cm
170cm
180cm
190cm
200cm

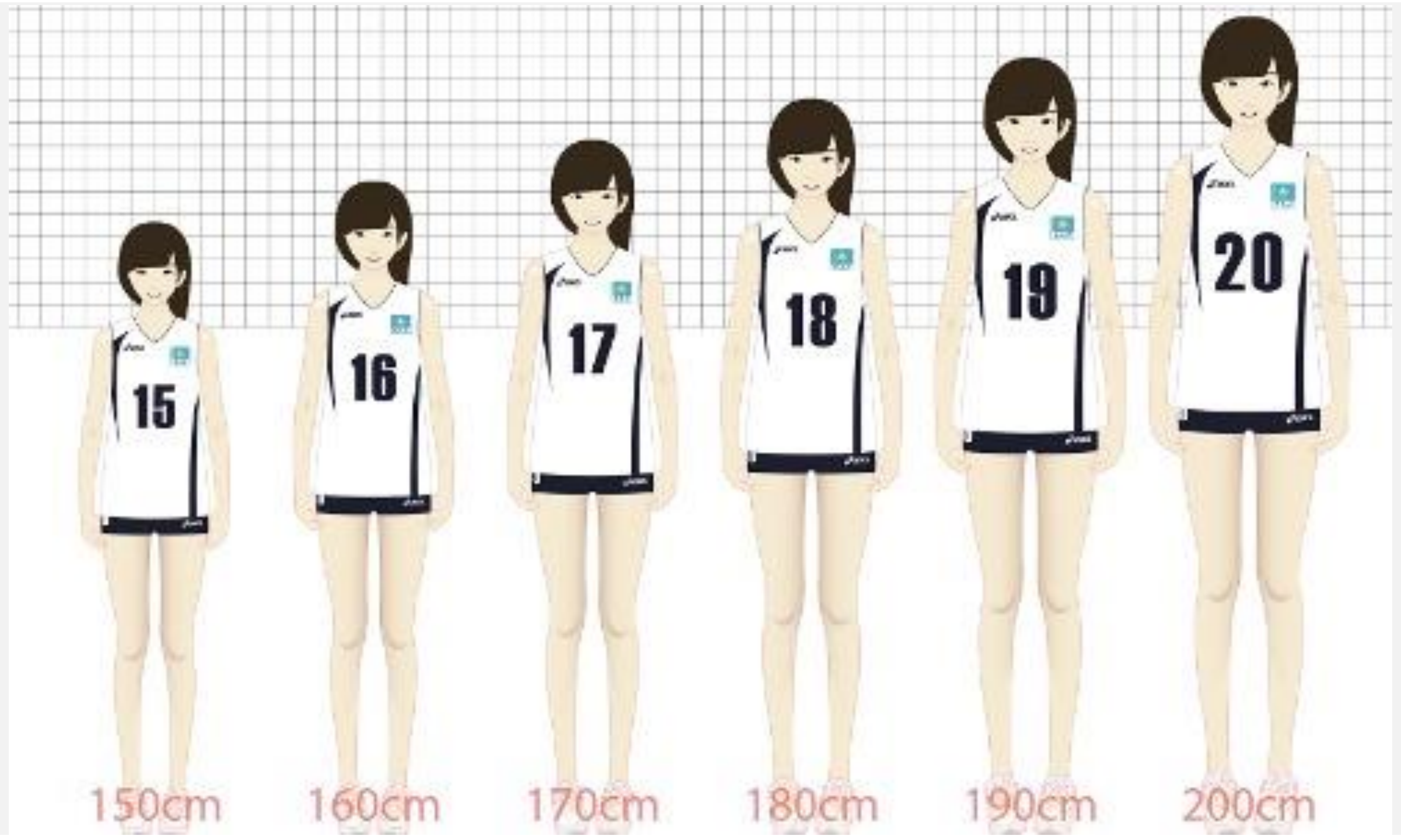


平均身高145
0.2
0.1
0
−0.1
−0.2
139 140 141 142 143 144 145 146 147 148 149 150

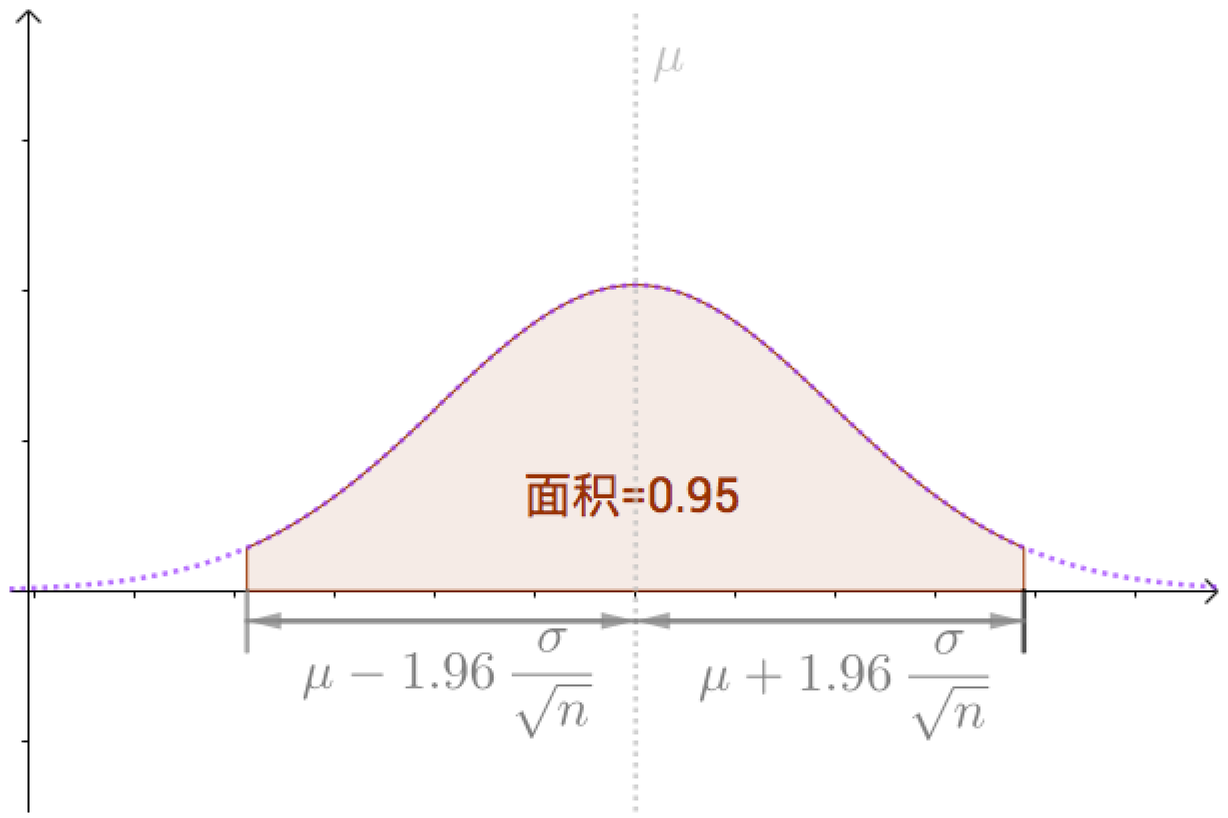


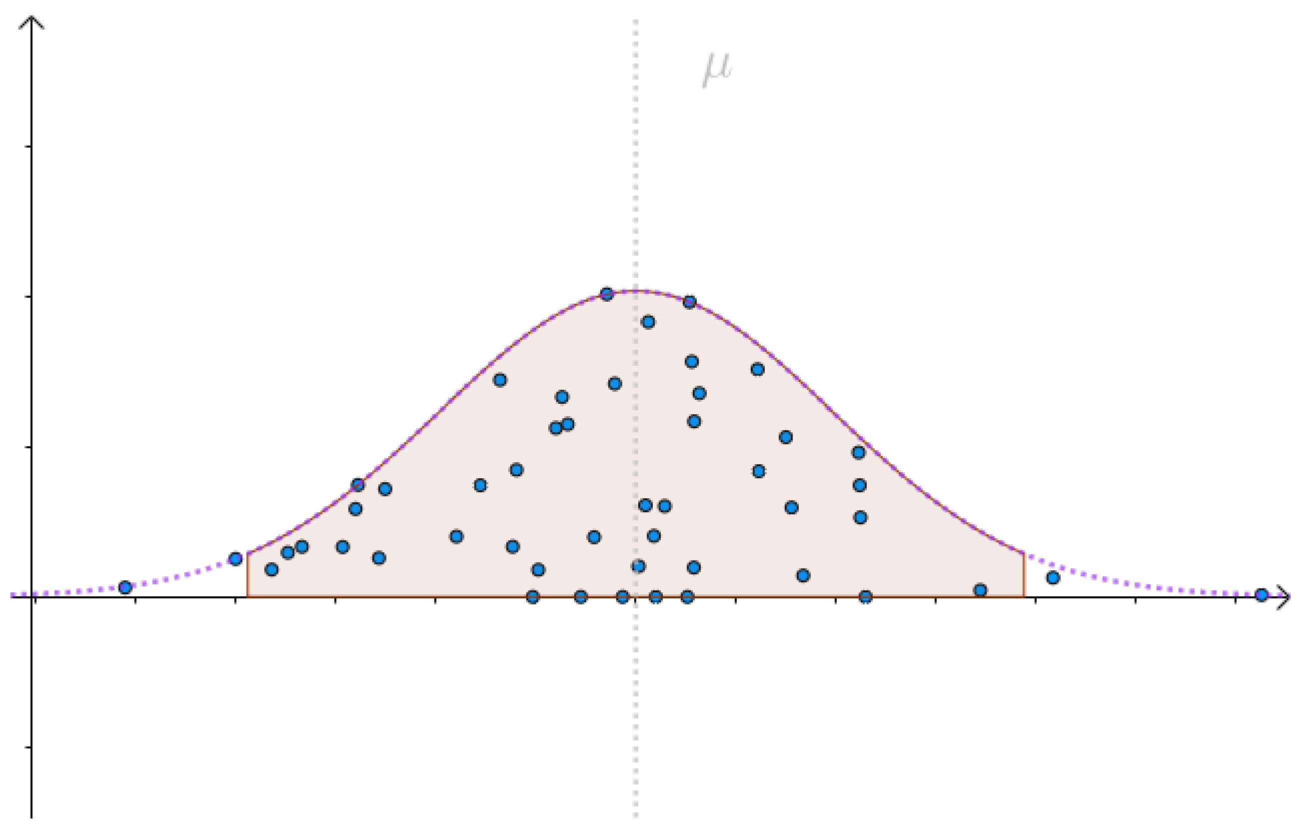
0.2
0.1
0
−0.1
−0.2
139 140 141 142 143 144 145 146 147 148 149 150 151

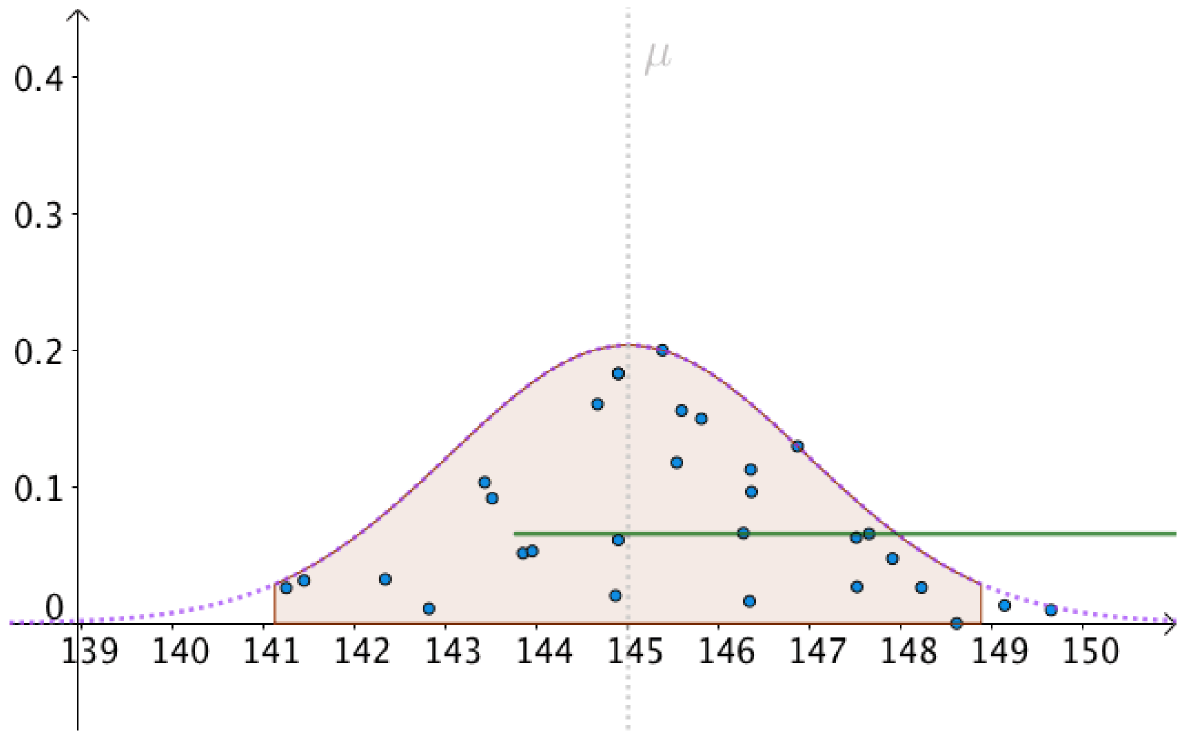
案例二：人的身高



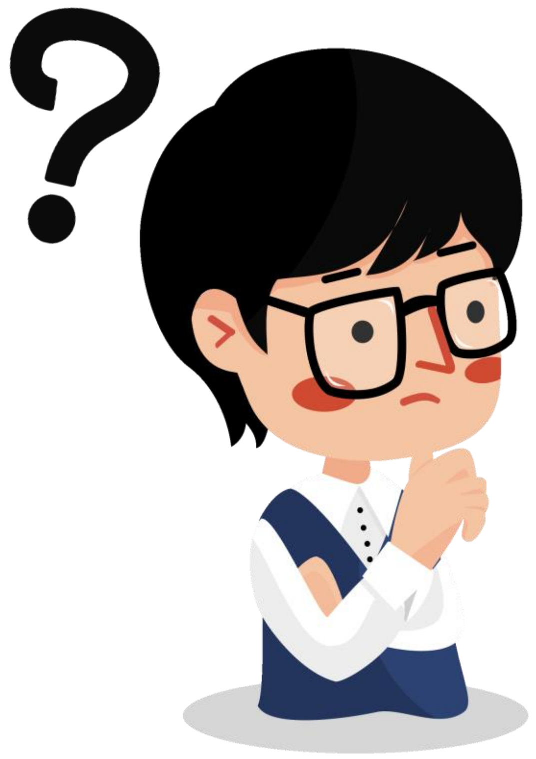
$X \sim N(\mu, \sigma^2)$，其中μ未知，σ未知







思考题



01 置信概率表达了区间估计的：

A、精准性　B、可靠性　C、显著性　D、规范性

02 对一个特定情形的估计来说，置信水平越低，所对应的置信区间：

A、越小　B、越大　C、不变　D、无法判断

03

总结

总结

本章包含两小节内容：

第一节

- 点估计的基本概念和方法

第二节

- 区间估计的基本概念和方法

谢谢观看

参考书目：概率论与数理统计·第四版（浙江大学） 高等教育出版社

网易云课堂 × 微专业