第 4 节 正态分布（★★☆）

内容提要

本节归纳正态分布有关题型，先梳理会用到的一些基础知识.

1．正态分布的概念：设函数 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}(x\in\mathbf{R})$，其中 $\mu\in\mathbf{R}$，$\sigma>0$ 为参数. 若随机变量 X 的概率分布密度函数为 $f(x)$，则称 X 服从正态分布，记作 $X\sim N(\mu,\sigma^2)$，其中 μ 和 σ^2 分别是服从正态分布的随机变量的均值和方差. 特别地，当 $\mu=0$，$\sigma=1$ 时，称随机变量 X 服从标准正态分布. 我们称函数 $f(x)$ 为正态密度函数，称它的图象为正态密度曲线，简称正态曲线，如图 1.

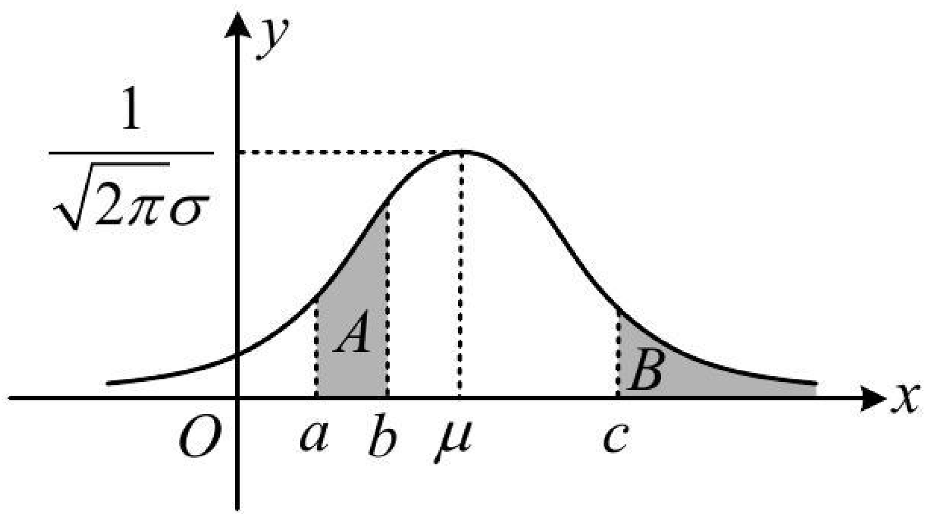


图1

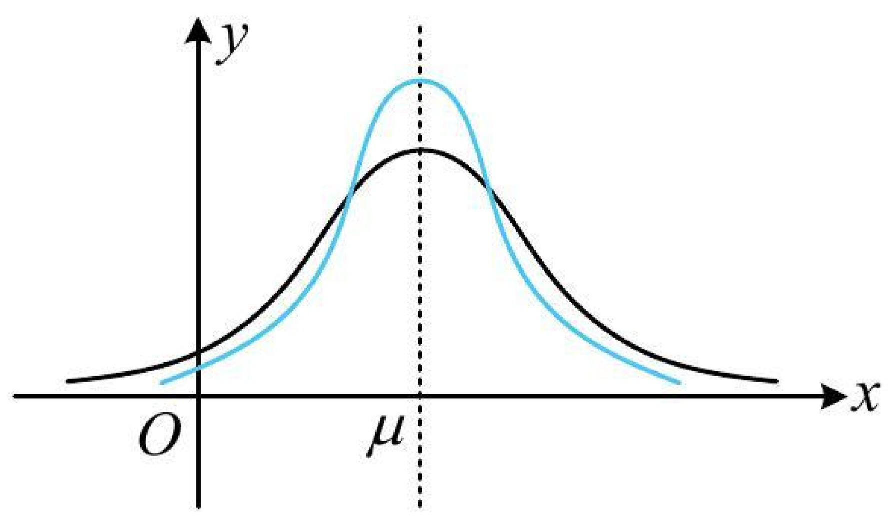


图2

2．取值概率：服从正态分布的随机变量取任何一个值的概率均为 0，我们更关注它在某区间内取值的概率，如上图 1，若 $X\sim N(\mu,\sigma^2)$，则 $P(a\le X\le b)$ 等于区域 A 的面积，$P(X\ge c)$ 等于区域 B 的面积；由于正态曲线关于 $x=\mu$ 对称，所以 $P(X\ge\mu)=0.5$.

3．调整参数对正态曲线的影响：

①取定 σ，调整 μ，则正态曲线的形状不变，但会沿 x 轴方向平移；

②取定 μ，调整 σ，则正态曲线的位置不变，但形状会发生变化. 若 σ 增大，则峰值 $\frac{1}{\sqrt{2\pi}\sigma}$ 减小，结合正态曲线与 x 轴围成的面积始终为 1 可知曲线会变得“矮胖”，X 的取值变得更分散，如上图 2 中黑色曲线；反之，若 σ 减小，则曲线会变得“瘦高”，X 的取值变得更集中，如上图 2 中蓝色曲线.

4．3σ 原则：若 $X\sim N(\mu,\sigma^2)$，则对给定的 $k\in\mathbf{N}^*$，$P(\mu-k\sigma\le X\le\mu+k\sigma)$ 是一个只与 k 有关的定值. 特别地，当 k 取 1，2，3 时的情况在统计中有广泛的应用，尤其重要：

$P(\mu-\sigma\le X\le\mu+\sigma)\approx0.6827$，$P(\mu-2\sigma\le X\le\mu+2\sigma)\approx0.9545$，$P(\mu-3\sigma\le X\le\mu+3\sigma)\approx0.9973$.

可以看到，X 在 $[\mu-3\sigma,\mu+3\sigma]$ 外取值的概率只有 0.0027，在实际应用中，通常认为服从正态分布 $N(\mu,\sigma^2)$ 的随机变量只取 $[\mu-3\sigma,\mu+3\sigma]$ 内的值.

典型例题

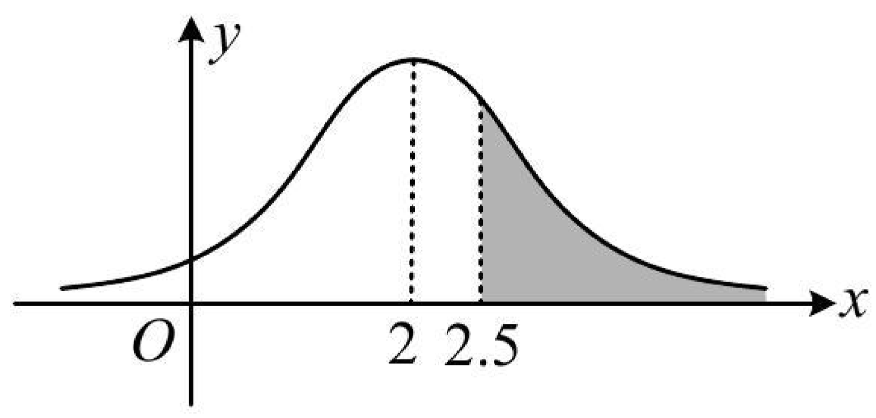
类型Ⅰ：用正态曲线求概率

【例 1】（2022•新高考Ⅱ卷）随机变量 X 服从正态分布 $N(2,\sigma^2)$，若 $P(2<X\le2.5)=0.36$，则 $P(X>2.5)=$ ______.

解析：可画正态曲线来分析，如图，求正态分布在某区间取值的概率，只需求该区间那部分的面积，

由题意，$\mu=2$，如图，$P(X>2.5)=P(X>2)-P(2<X\le2.5)=0.5-0.36=0.14$.

答案：0.14



【变式 1】(多选)已知某批零件的质量指标ξ（单位：毫米）服从正态分布$N(25.4,\sigma^2)$，且$P(\xi\ge 25.45)=0.1$，现从该批零件中随机取3件，用X表示这3件零件中质量指标值ξ落在区间$(25.35,25.45)$外的件数，则(　　)

（A）$P(25.35<\xi<25.45)=0.8$　　（B）$E(X)=2.4$　　（C）$D(X)=0.48$　　（D）$P(X\ge 1)=0.488$

解析：A 项，如图，由对称性，$P(\xi\le 25.35)=P(\xi\ge 25.45)=0.1$，

所以$P(25.35<\xi<25.45)=1-P(\xi\le 25.35)-P(\xi\ge 25.45)=1-0.1-0.1=0.8$，故 A 项正确；

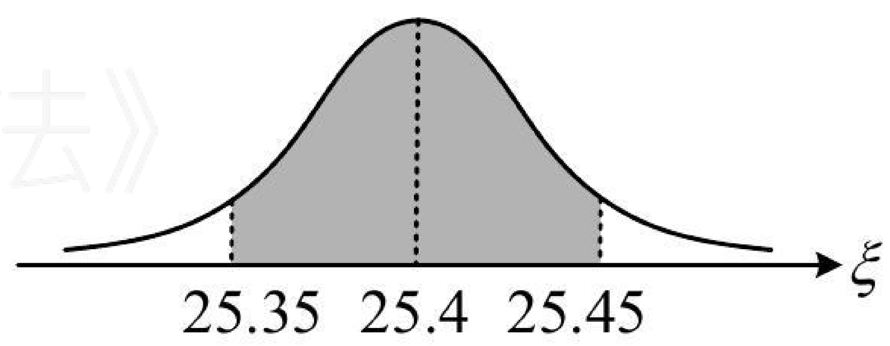
B 项，要分析 3 件产品中落在$(25.35,25.45)$外的件数，需先求出 1 件产品落在该区间外的概率，

随机取 1 件零件，质量指标值位于区间$(25.35,25.45)$外的概率为 0.2，所以$X\sim B(3,0.2)$，

从而$E(X)=3\times 0.2=0.6$，$D(X)=3\times 0.2\times(1-0.2)=0.48$，故 B 项错误，C 项正确；

D 项，因为$X\sim B(3,0.2)$，所以$P(X\ge 1)=1-P(X=0)=1-C_3^0\times(1-0.2)^3=0.488$，故 D 项正确.

答案：ACD



【变式 2】对一个物理量做 n 次测量，并以测量结果的平均值作为该物理量的最后结果. 已知最后结果的误差$\varepsilon_n\sim N(0,\dfrac{2}{n})$，为使误差$\varepsilon_n$在$(-0.5,0.5)$的概率不低于 0.9545，至少要测量_____次.（若$X\sim N(\mu,\sigma^2)$，则$P(|X-\mu|<2\sigma)=0.9545$）

解析：$\varepsilon_n\sim N(0,\dfrac{2}{n})\Rightarrow\mu=0$，$\sigma=\sqrt{\dfrac{2}{n}}$，由题意，$P(-0.5<\varepsilon_n<0.5)\ge 0.9545$ ①，

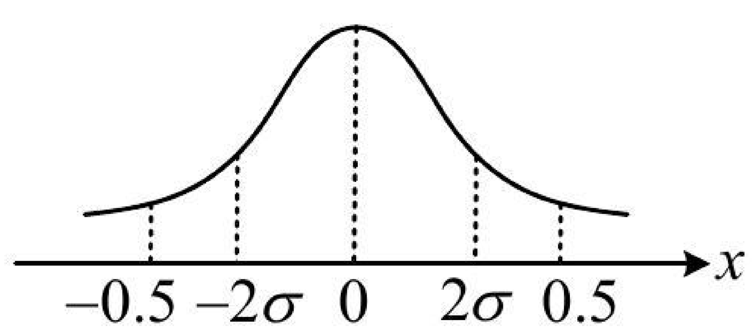
涉及正态分布的区间取值概率，可画正态曲线来看，

由题意，若$X\sim N(\mu,\sigma^2)$，则$P(|X-\mu|<2\sigma)=P(-2\sigma<X-\mu<2\sigma)=P(\mu-2\sigma<X<\mu+2\sigma)=0.9545$，

所以对于ε_n，有$P(-2\sigma<\varepsilon_n<2\sigma)=0.9545$，如图，不等式①等价于$(-2\sigma,2\sigma)\subseteq(-0.5,0.5)$，

从而$2\sigma\le 0.5$，故$2\sqrt{\dfrac{2}{n}}\le\dfrac{1}{2}$，解得：$n\ge 32$，所以至少要测量 32 次.

答案：32



【总结】涉及正态分布算区间概率，可画正态曲线，利用其对称性和3σ区间取值概率来分析即可.

类型Ⅱ：正态分布在实际问题中的综合应用

【例 2】某车间生产一批零件，现从中随机取 10 个，测量其内径（单位：mm）的数据如下：192，192，193，197，200，202，203，204，208，209，设这 10 个数据的均值为μ，标准差为σ.

（1）求μ和σ；

（2）已知这批零件的内径X服从正态分布$N(\mu,\sigma^2)$，若该车间又新购一台设备，安装调试后，试生产了 5 个零件，测量其内径分别为：181，190，198，204，213，如果你是该车间的负责人，以原设备生产性能为标准，试根据3σ原则判断这台设备是否需要进一步调试？并说明你的理由.

附：若$X\sim N(\mu,\sigma^2)$，则$P(\mu-\sigma<X<\mu+\sigma)\approx 0.6827$，$P(\mu-2\sigma<X<\mu+2\sigma)\approx 0.9545$，

$P(\mu-3\sigma<X<\mu+3\sigma)\approx 0.9973$，$0.9973^4\approx 0.99$.

解：（1）由题意，$\mu=\dfrac{192+192+193+197+200+202+203+204+208+209}{10}=200$，

$\sigma^2=\dfrac{1}{10}[(192-200)^2+(192-200)^2+(193-200)^2+(197-200)^2+(200-200)^2+(202-200)^2+(203-200)^2$

$+(204-200)^2+(208-200)^2+(209-200)^2]=36$，所以$\sigma=6$.

（2）（要检验设备是否正常，先看试生产的零件的内径是否有落在区间$[\mu-3\sigma,\mu+3\sigma]$外的）

由（1）可得$X\sim N(200,36)$，所以$[\mu-3\sigma,\mu+3\sigma]$即为$[182,218]$，且$P(182<X<218)\approx 0.9973$，

试生产的 5 个零件中有一个内径为 181mm，落在了$[182,218]$外，

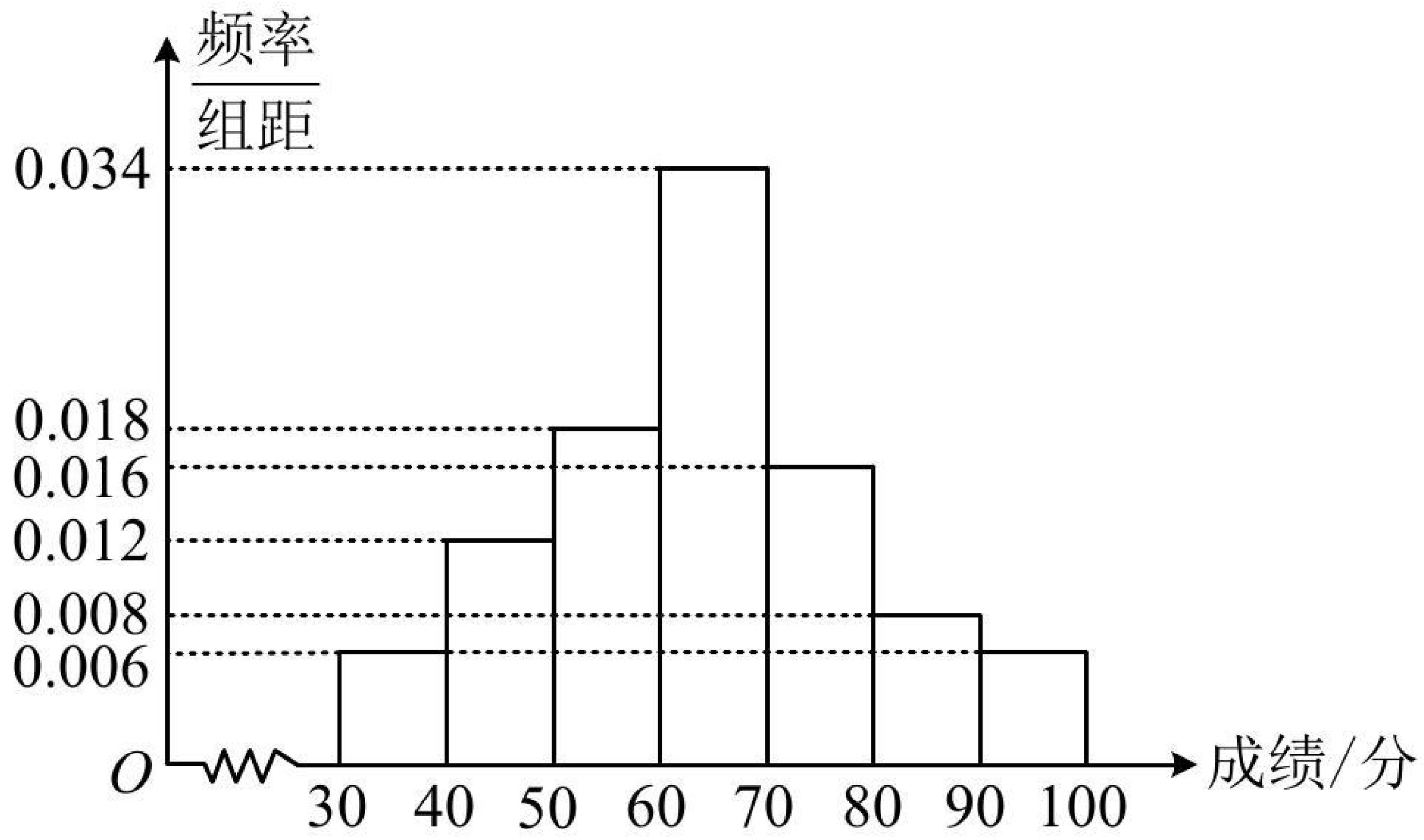
（此时应再计算试生产的 5 个零件，恰有 1 个落在$[182,218]$外的概率，看此概率大不大）

记生产 5 个零件，落在$[182,218]$外的个数为Y，则$Y\sim B(5,0.0027)$，

所以$P(Y=1)=\mathrm{C}_5^1\times 0.0027\times 0.9973^4\approx 0.013$，此概率很小，而小概率事件发生了，故该设备需进一步调试.

【反思】由于服从正态分布的随机变量在区间$[\mu-3\sigma,\mu+3\sigma]$外取值的概率很小，所以试生产零件数不多的情况下，出现该区间外的值的概率也较小，故可按是否出现区间外的值来检验设备是否异常.

【例 3】某市为了传承发展中华优秀传统文化，组织该市中学生进行了一次文化知识有奖竞赛，竞赛奖励规则如下：得分在$[70,80)$内的学生获三等奖，得分在$[80,90)$内的学生获二等奖，得分在$[90,100]$内的学生获一等奖，其他学生不获奖. 为了解学生对相关知识的掌握情况，随机抽取 100 名学生的竞赛成绩，并以此为样本绘制了样本频率分布直方图，如图.



（1）现从该样本中随机抽取两名学生的竞赛成绩，求这两名学生中恰有一名学生获奖的概率；

（2）若该市所有参赛学生的成绩 X 近似服从正态分布 $N(\mu,\sigma^2)$，其中 $\sigma\approx 15$，μ 为样本平均数的估计值，利用所得正态分布模型解决以下问题：

（i）若该市共有 10000 名学生参加了竞赛，试估计参赛学生中成绩超过 79 分的学生数；（结果四舍五入到整数）

（ii）若从所有参赛学生中（参赛学生数大于 10000）随机取 3 名学生进行访谈，设其中竞赛成绩在 64 分以上的学生数为 ξ，求 ξ 的分布列和期望.

附：若随机变量 $X\sim N(\mu,\sigma^2)$，则 $P(\mu-\sigma<X<\mu+\sigma)\approx 0.6827$，$P(\mu-2\sigma<X<\mu+2\sigma)\approx 0.9545$，$P(\mu-3\sigma<X<\mu+3\sigma)\approx 0.9973$.

解：（1）由题意，样本 100 人的成绩中，获奖的有 $100\times(0.016+0.008+0.006)\times 10=30$ 人，

所以抽到的两名学生中恰有一名学生获奖的概率 $P=\dfrac{C_{70}^1C_{30}^1}{C_{100}^2}=\dfrac{14}{33}$.

（2）（i）（要估算成绩超过 79 分的人数，应先求 $P(X>79)$，而要求此概率，需找到 79 与 μ 和 σ 的关系）

由题意，$\mu=35\times 0.06+45\times 0.12+55\times 0.18+65\times 0.34+75\times 0.16+85\times 0.08+95\times 0.06=64$，所以 $X\sim N(64,15^2)$，

如图，$P(X>79)=P(X>\mu+\sigma)=\dfrac{1-P(\mu-\sigma<X<\mu+\sigma)}{2}\approx\dfrac{1-0.6827}{2}=0.15865$，

该市共有 10000 名学生参加了竞赛，所以参赛学生中成绩超过 79 分的学生数约为 $10000\times 0.15865\approx 1587$.

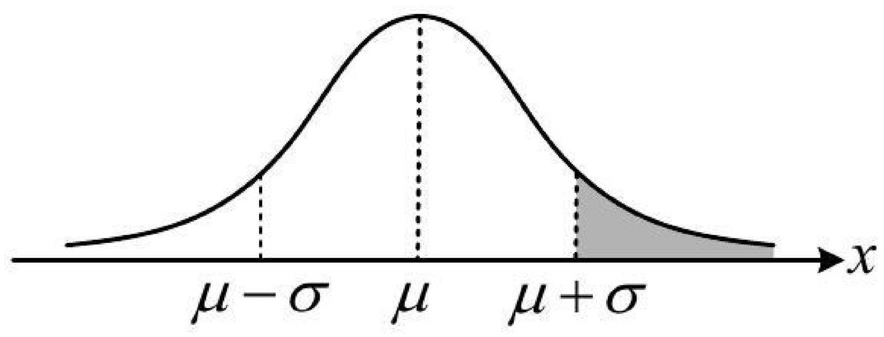
（ii）（抽取人数远小于总人数，故可用二项分布来近似处理）

因为 $\mu=64$，所以 $P(X>64)=\dfrac{1}{2}$，故 $\xi\sim B(3,\dfrac{1}{2})$，所以 $P(\xi=0)=C_3^0\times(1-\dfrac{1}{2})^3=\dfrac{1}{8}$，

$P(\xi=1)=C_3^1\times\dfrac{1}{2}\times(1-\dfrac{1}{2})^2=\dfrac{3}{8}$，$P(\xi=2)=C_3^2\times(\dfrac{1}{2})^2\times(1-\dfrac{1}{2})=\dfrac{3}{8}$，$P(\xi=3)=C_3^3\times(\dfrac{1}{2})^3=\dfrac{1}{8}$，故 ξ 的分布列为：

ξ	0	1	2	3
P	$\dfrac{1}{8}$	$\dfrac{3}{8}$	$\dfrac{3}{8}$	$\dfrac{1}{8}$

因为 $\xi\sim B(3,\dfrac{1}{2})$，所以 ξ 的期望 $E(\xi)=3\times\dfrac{1}{2}=\dfrac{3}{2}$.



【总结】在实际问题中，计算正态分布在某区间取值的概率，关键是找到区间端点与μ和σ的关系，借助正态曲线和$P(\mu-\sigma<X<\mu+\sigma)$，$P(\mu-2\sigma<X<\mu+2\sigma)$，$P(\mu-3\sigma<X<\mu+3\sigma)$这些参考数据来计算，所以凡是涉及正态分布，务必关注μ和σ.

强化训练

1.（2022·江西模拟·★）设随机变量 $X\sim N(3,\sigma^2)$，若 $P(X>5)=0.2$，则 $P(1<X<3)=$_____.

2.（2023·山东济南三模·★）已知随机变量 X，Y，其中 $X\sim B(6,\frac{1}{3})$，$Y\sim N(\mu,\sigma^2)$，$E(X)=E(Y)$，$P(|Y|<2)=0.3$，则 $P(Y>6)=$_____.

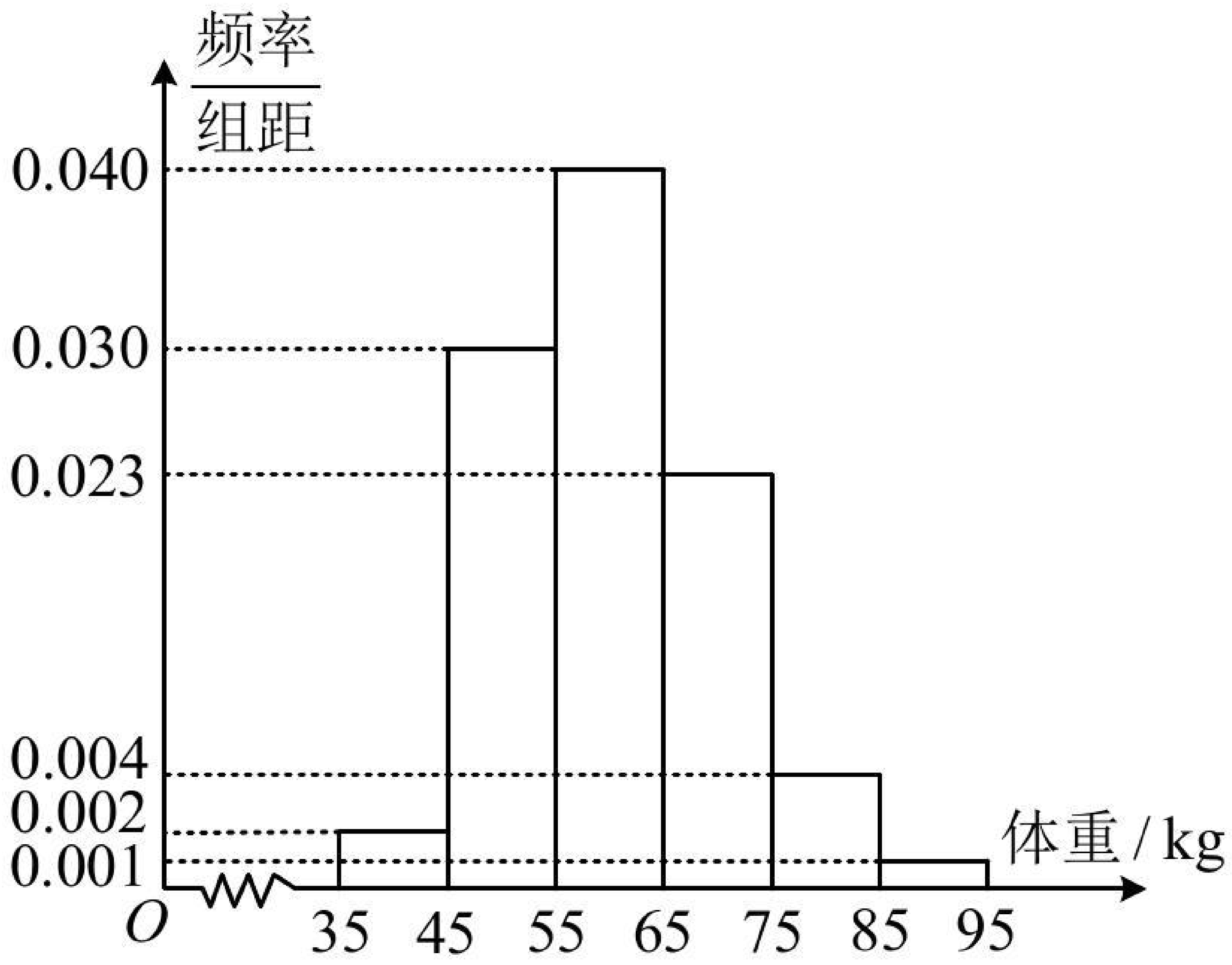
3.（2023·山东潍坊一模·★★）某学校共 1000 人参加数学测验，考试成绩 ξ 近似服从正态分布 $N(100,\sigma^2)$，若 $P(80\le\xi\le100)=0.45$，则估计成绩在 120 分以上的学生人数为（　　）

（A）25　（B）50　（C）75　（D）100

4.（2023·安徽模拟·★★）小明统计了最近一段时间某超市冷饮的销售量 X，根据统计发现 X 近似服从正态分布 $N(50,\sigma^2)$，且 $P(X\ge20)=0.9$，已知该超市冷饮的销售量在区间 $[20,80]$ 内的有 80 天，则可以估计小明一共统计了_____天.

5.（2023·四省联考·★★★）某工厂生产的产品的质量指标服从正态分布 $N(100,\sigma^2)$，质量指标介于 99 至 101 之间的产品为良品，为使这种产品的良品率达到 95.45%，则需调整生产工艺，使得 σ 至多为_____.（附：若 $X\sim N(\mu,\sigma^2)$，则 $P(|X-\mu|<2\sigma)\approx0.9545$）

6．（2023·安徽模拟·★★★）为贯彻落实《健康中国行动（2019~2030 年）》、《关于全面加强和改进新时代学校体育工作的意见》等文件的精神，确保 2030 年学生体质达到规定要求，各地将认真做好学生的体质健康检测. 某市决定对某中学学生的身体健康状况进行调查，现从该校抽取 200 名学生测量他们的体重，得到如下的样本数据的频率分布直方图.



（1）求这 200 名学生体重的平均数 $\bar{x}$ 和方差 s^2；（同一组数据用该区间的中点值作代表）

（2）由频率分布直方图可知，该校学生的体重 Z 服从正态分布 $N(\mu,\sigma^2)$，其中 μ 近似为样本平均数 $\bar{x}$，σ^2 近似为样本方差 s^2.

①利用该正态分布，求 $P(50.73<Z\le 69.27)$；

②若从该校随机抽取 50 名学生，记 X 表示这 50 名学生的体重位于区间 $(50.73,69.27]$ 内的人数，利用①的结果，求 $E(X)$.

参考数据：$\sqrt{86}\approx 9.27$，若 $Z\sim N(\mu,\sigma^2)$，则 $P(\mu-\sigma<Z\le\mu+\sigma)\approx 0.6827$，$P(\mu-2\sigma<Z\le\mu+2\sigma)\approx 0.9545$，$P(\mu-3\sigma<Z\le\mu+3\sigma)\approx 0.9973$.