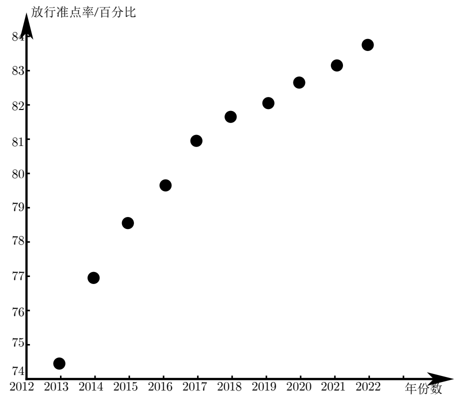
第11讲:概率高考压轴梳理

1. 放行准点率是衡量机场运行效率和服务质量的重要指标之一. 某机场自 2012 年起采取相关策略优化各个服务环节, 运行效率不断提升. 以下是根据近 10 年年份数 x_i 与该机场飞往 A 地航班放行准点率 $y_i(i=1,2,\cdots,10)$ (单位: 百分比) 的统计数据所作的散点图及经过初步处理后得到的一些统计量的值.

$\hat{y}=\hat{c}t+\hat{d}$

$\hat{c}=\dfrac{\sum_{i=1}^{10}t_iy_i-10\bar{t}\cdot\bar{y}}{\sum_{i=1}^{10}t_i^2-10\cdot\bar{t}^2}=\dfrac{1226.8-10\times1.5\times80.4}{27.7-10\times1.5^2}=4$



$\hat{d}=\bar{y}-\hat{c}\cdot\bar{t}=80.4-4\times1.5=74.4$

$\bar{x}$	$\bar{y}$	$\bar{t}$	$\sum_{i=1}^{10}x_i^2$	$\sum_{i=1}^{10}x_iy_i$	$\sum_{i=1}^{10}t_i^2$	$\sum_{i=1}^{10}t_iy_i$
2017.5	80.4	1.5	40703145.0	1621254.2	27.7	1226.8

$\hat{y}=4\cdot\ln(\hat{x}-2012)+74.4$

其中 $t_i=\ln(x_i-2012),\bar{t}=\dfrac{1}{10}\sum_{i=1}^{10}t_i$

当 $x=2023$ 时 $\hat{y}=4\ln11+74.4=4\times2.40+74.4=84$

(1) 根据散点图判断, $y=bx+a$ 与 $y=c\ln(x-2012)+d$ 哪一个适宜作为该机场飞往 A 地航班放行准点率 y 关于年份数 x 的经验回归方程类型 (给出判断即可, 不必说明理由), 并根据表中数据建立经验回归方程, 由此预测 2023 年该机场飞往 A 地的航班放行准点率. ∴预测 2023 年飞往 A 地准点率 84%

赵礼显数学

(2) 已知 2023 年该机场飞往 A 地、B 地和其他地区的航班比例分别为 0.2、0.2 和 0.6. 若以 (1) 中的预测值作为 2023 年该机场飞往 A 地航班放行准点率的估计值, 且 2023 年该机场飞往 B 地及其他地区 (不包含 A、B 两地) 航班放行准点率的估计值分别为 80% 和 75%, 试解决以下问题:

①现从 2023 年在该机场起飞的航班中随机抽取一个,求该航班准点放行的概率;

②若 2023 年某航班在该机场准点放行, 判断该航班飞往 A 地、B 地、其他地区等三种情况中的哪种情况的可能性最大, 说明你的理由.

附:(1) 对于一组数据 $(u_1,v_1),(u_2,v_2),\cdots,(u_n,v_n)$, 其回归直线 $v=\alpha+\beta u$ 的斜率和截距的最小二乘估计分别为 $\beta=\dfrac{\sum\limits_{i=1}^{n}(u_i-\bar{u})(v_i-\bar{v})}{\sum\limits_{i=1}^{n}(u_i-\bar{u})^2}=\dfrac{\sum\limits_{i=1}^{n}u_iv_i-n\bar{u}\bar{v}}{\sum\limits_{i=1}^{n}u_i^2-n\bar{u}^2}$;$\alpha=\bar{v}-\beta\bar{u}$;

(2) 参考数据:$\ln 10\approx 2.30,\ln 11\approx 2.40,\ln 12\approx 2.48$.

① 设 A_1 表示飞往 A.
A_2 表示飞往 B
A_3 表示飞往其他
C 表示准点

$p(C)=p(A_1)\cdot p(C|A_1)+p(A_2)\cdot p(C|A_2)+p(A_3)\cdot p(C|A_3)$

$=0.2\times 0.84+0.2\times 0.8+0.6\times 0.75=0.778.$

② $p(A_1|C)=\dfrac{p(A_1C)}{p(C)}=\dfrac{p(A_1)\cdot p(C|A_1)}{p(C)}=\dfrac{0.2\times 0.84}{0.778}$

$p(A_2|C)=\dfrac{p(A_2)\cdot p(C|A_2)}{p(C)}=\dfrac{0.2\times 0.8}{0.778}$, $p(A_3|C)=\dfrac{p(A_3)\cdot p(C|A_3)}{p(C)}=\dfrac{0.6\times 0.75}{0.778}$ 大.

赵礼显数学

2. 调查问卷中常常涉及到个人隐私或本人不愿正面回答的问题，被访人可能拒绝回答，即使回答，也不能期望答案是真实的．某小区要调查业主对物业工作是否满意的真实情况，现利用“随机化选答抽样”方法制作了具体调查方案，其操作流程如下：在一个箱子里放 3 个红球和 2 个白球，被调查者在摸到球后记住颜色并立即将球放回，如果抽到的是红球，则回答“你的性别是否为男性?”如果抽到的是白球，则回答“你对物业工作现状是否满意?”两个问题均用“是”或“否”回答．

(1) 共收取调查问卷 100 份，其中答案为“是”的问卷为 60 份，求一个业主对物业工作表示满意的概率，已知该小区共有业主 500 人，估计该小区业主对物业工作满意的人数；

$\frac{3}{5}\times\frac{1}{2}+\frac{2}{5}p=\frac{3}{5}$

$p=\frac{3}{4}$

$500\times\frac{3}{4}=375$(人)

(2) 现为了提高对物业工作满意的业主比例，对小区业主进行随机访谈，请表示不满意的业主在访谈中提出两个有待改进的问题．

①若物业对每一个待改进的问题均提出一个相应的解决方案，该方案需要由 5 名业主委员会代表投票决定是否可行．每位代表投赞同票的概率均为 $\frac{1}{3}$，方案需至少 3 人投赞成票，方能予以通过，并最终解决该问题，求某个问题能够被解决的概率 p_0；

$p_0=C_5^3\left(\frac{1}{3}\right)^3\cdot\left(\frac{2}{3}\right)^2+C_5^4\left(\frac{1}{3}\right)^4\cdot\frac{2}{3}+\left(\frac{1}{3}\right)^5=\frac{17}{81}$

②假设业主所提问题各不相同，每一个问题能够被解决的概率都为 p_0，并且都相互独立．物业每解决一个问题，业主满意的比例将提高一个百分点．为了让业主满意的比例提高到 80%，试估计至少要访谈多少位业主?

$n\cdot\left(1-\frac{3}{4}\right)\cdot 2\cdot\frac{17}{81}\geq 5.$

$n\geq\frac{810}{17}\doteq 47.6$

$\therefore$ 至少 48人

赵礼显数学

3. 某单位为患病员工集体筛查新型流感病毒，需要去某医院检验血液是否为阳性，现有 $k(k\in N^{*}$，$k\geqslant 2)$ 份血液样本，有以下两种检验方案，方案一：逐份检验，则需要检验 k 次；方案二：混合检验，将 k 份血液样本分别取样混合在一起检验一次，若检验结果为阴性，则 k 份血液样本均为阴性，若检验结果为阳性，为了确定 k 份血液中的阳性血液样本，则对 k 份血液样本再逐一检验逐份检验和混合检验中的每一次检验费用都是 $a(a>0)$ 元，且 k 份血液样本混合检验一次需要额外收 $\frac{5}{4}a$ 元的材料费和服务费．假设在接受检验的血液样本中，每份样本是否为阳性是相互独立的，且据统计每份血液样本是阳性的概率为 $p(0<p<1)$．

(1) 假设有 5 份血液样本，其中只有 2 份样本为阳性，若采用逐份检验的方式，求恰好经过 3 次检验就能把阳性样本全部检验出来的概率；

(2) 若 $k(k\in N^{*}, k\geqslant 2)$ 份血液样本采用混合检验方案，需要检验的总次数为 X，求 X 分布列及数学期望；

(3) ①若 $k=10$，$0<p<1-\sqrt[10]{0.225}$，以检验总费用为决策依据，试说明该单位选择方案二的合理性；

②若 $p=1-\frac{1}{\sqrt[7]{e}}$，采用方案二总费用的数学期望低于方案一，求 k 的最大值．

$$P=\frac{A_3^3+C_2^2C_3^1C_2^1A_2^2}{A_5^3}=\frac{6+12}{5\times4\times3}=\frac{18}{5\times4\times3}=\frac{3}{10}$$

(2) X 取值 1. $k+1$

$P(X=1)=(1-p)^k$

$P(X=k+1)=1-(1-p)^k$ ∴ $E(X)=(1-p)^k+(k+1)-(k+1)\cdot(1-p)^k=k+1-k\cdot(1-p)^k$

3. 某单位为患病员工集体筛查新型流感病毒，需要去某医院检验血液是否为阳性，现有 $k(k\in N^*, k\geqslant 2)$ 份血液样本，有以下两种检验方案，方案一：逐份检验，则需要检验 k 次；方案二：混合检验，将 k 份血液样本分别取样混合在一起检验一次，若检验结果为阴性，则 k 份血液样本均为阴性，若检验结果为阳性，为了确定 k 份血液中的阳性血液样本，则对 k 份血液样本再逐一检验逐份检验和混合检验中的每一次检验费用都是 $a(a>0)$ 元，且 k 份血液样本混合检验一次需要额外收 $\frac{5}{4}a$ 元的材料费和服务费．假设在接受检验的血液样本中，每份样本是否为阳性是相互独立的，且据统计每份血液样本是阳性的概率为 $p(0<p<1)$．

$\ln 11-\frac{11}{7}-\ln 9+\ln 4=\ln 11-2\ln 3+2\ln 2-\frac{11}{7}>0$

$\ln 8-\frac{8}{7}-\ln 9+\ln 4=5\ln 2-\ln 3-\frac{8}{7}$

$f(\ln 2)<0$

(1) 假设有 5 份血液样本，其中只有 2 份样本为阳性，若采用逐份检验的方式，求恰好经过 3 次检验就能把阳性样本全部检验出来的概率；

② $aE(X)+\frac{5}{4}a<ka \Rightarrow k+1-k\cdot(1-p)^k+\frac{5}{4}-k<0$

k 最大值 11

(2) 若 $k(k\in N^*, k\geqslant 2)$ 份血液样本采用混合检验方案，需要检验的总次数为 X，求 X 分布列及数学期望；

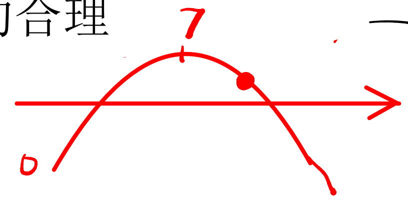
$\frac{9}{4}-k\cdot e^{-\frac{k}{7}}<0 \quad \therefore k\cdot e^{-\frac{k}{7}}>\frac{9}{4}\Rightarrow \ln k-\frac{k}{7}>\ln\frac{9}{4}$

(3) ①若 $k=10$，$0<p<1-\sqrt[10]{0.225}$，以检验总费用为决策依据，试说明该单位选择方案二的合理性；

$\ln 11\approx 1.6$

$f(x)=\ln x-\frac{x}{7}-\ln\frac{9}{4}$，$f'(x)=\frac{1}{x}-\frac{1}{7}=\frac{7-x}{7x}$

$f(x)$ 在 $(0,7)\uparrow$ $(7,+\infty)\downarrow$



②若 $p=1-\frac{1}{\sqrt[7]{e}}$，采用方案二总费用的数学期望低于方案一，求 k 的最大值．

① 方案一：$Z=10a$

方案二：$Z=aX+\frac{5}{4}a$

$\therefore E(Z)=aE(X)+\frac{5}{4}a=a(k+1-k(1-p)^k)+\frac{5}{4}a=a(11-10(1-p)^{10})+\frac{5}{4}a$

$<a(11-10\times 0.225)+\frac{5}{4}a$

$=(8.75+1.25)a=10a$

$\therefore$ 方案二合理.

4. 人工智能是研究用于模拟和延伸人类智能的技术科学，被记为是 21 世纪最重要的尖端科技之一，其理论和技术正在日益成熟，应用领域也在不断扩大，人工智能背后的一个基本原理：首先确定先验概率，然后通过计算得到后验概率，使先验概率得到修正和校对，再根据后验概率做出推理和决策.

基于这一基本原理，我们可以设计如下试验模型：有完全相同的甲、乙两个袋子，袋子有形状和大小完全相同的小球，其中甲袋中有 9 个红球和 1 个白球；乙袋中有 2 个红球和 8 个白球. 从这两个袋子中选择一个袋子，再从该袋子中等可能摸出一个球，称为一次试验. 若多次试验直到摸出红球，则试验结束. 假设首次试验选到甲袋或乙袋的概率均为 $\frac{1}{2}$ (先验概率). $P(B_1)=P(A_1)\cdot P(B_1|A_1)+P(A_2)\cdot P(B_1|A_2)$

$=\frac{1}{2}\times\frac{9}{10}+\frac{1}{2}\times\frac{2}{10}=\frac{11}{20}$

(1) 求首次试验结束的概率；记 A_1 表示选甲，A_2 表示选乙袋. B_i 表示第 i 次摸出红球. $i=1,2$.

(2) 在首次试验摸出白球的条件下，我们对选到甲袋或乙袋的概率 (先验概率) 进行调整.

①求选到的袋子为甲袋的概率；$P(A_1|\overline{B_1})=\frac{P(A_1\overline{B_1})}{P(\overline{B_1})}=\frac{P(A_1)\cdot P(\overline{B_1}|A_1)}{P(\overline{B_1})}=\frac{\frac{1}{2}\times\frac{1}{10}}{\frac{9}{20}}=\frac{1}{9}$.

②将首次试验摸出的白球放回原来袋子，继续进行第二次试验时有如下两种方案：方案一，从原来袋子中摸球；方案二，从另外一个袋子中摸球.

请通过计算，说明选择哪个方案第二次试验结束的概率更大.

方案一：$P(B_2)=P(A_1)\cdot P(B_2|A_1)+P(A_2)\cdot P(B_2|A_2)=\frac{1}{9}\times\frac{9}{10}+\frac{8}{9}\times\frac{2}{10}=\frac{25}{90}=\frac{5}{18}$

方案二：$P(B_2)=\frac{1}{9}\times\frac{2}{10}+\frac{8}{9}\times\frac{9}{10}=\frac{74}{90}=\frac{37}{45}$ 方案二概率大

赵礼显数学

5. 为响应德智体美劳的教育方针，唐徕回民中学一年级举行了由全体学生参加的一分钟跳绳比赛，计分规则如表：

每分钟跳绳个数	[145，155)	[155，165)	[165，175)	[175，185)	185 以上
得分	16	17	18	19	20

年级组为了了解学生的体质，随机抽取了 100 名学生，统计了他们的跳绳个数，并绘制了如图样本频率直方图.

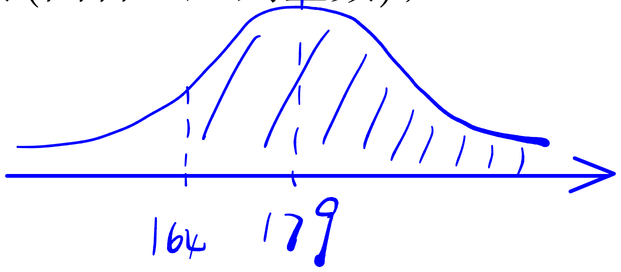
$p=\frac{C_6^2+C_8^2+C_6^1\cdot C_8^1+C_6^1\cdot C_{18}^1}{C_{100}^2}=\frac{29}{550}$

(1) 现从这 100 名学生中，任意抽取 2 人，求两人得分之和小于 35 分的概率（结果用最简分数表示）；

(2) 若该校高二年级 2000 名学生，所有学生的一分钟跳绳个数 X 近似服从正态分布 $N(\mu,\sigma^2)$，其中 $\sigma^2\approx225$，μ 为样本平均数的估计值（同一组中数据以这组数据所在区间的中点值为代表）. 利用所得到的正态分布模型解决以下问题：

①估计每分钟跳绳 164 个以上的人数（四舍五入到整数）；

$X\sim N(179,225)$



$P(X>164)=1-\frac{1-0.6826}{2}$

$=\frac{1}{2}+0.3413=0.8413$

每分钟164以上人数 $2000\times0.8413=1682.6\approx1683$(人)

②若在全年级所有学生中随机抽取 3 人，记每分钟跳绳在 179 个以上的人数为 Y，求 Y 的分布列和数学期望与方差.

（若随机变量 X 服从正态分布 $N(\mu,\sigma^2)$，则 $P(\mu-\sigma<X\leqslant\mu+\sigma)=0.6826$，$P(\mu-2\sigma<X\leqslant\mu+2\sigma)=0.9544$，$P(\mu-3\sigma<X\leqslant\mu+3\sigma)=0.9974$.）

$Y\sim B(3,\frac{1}{2})$

Y 取值 0，1，2，3

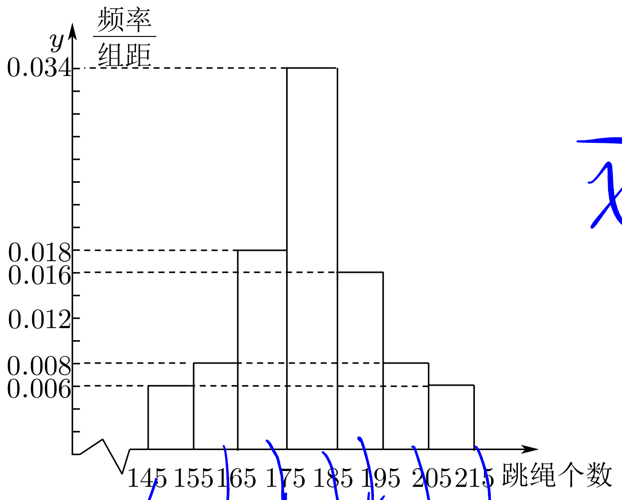
$P(Y=0)=(\frac{1}{2})^3=\frac{1}{8}$

$P(Y=1)=C_3^1\cdot(\frac{1}{2})^1\cdot(1-\frac{1}{2})^2=\frac{3}{8}$

$P(Y=2)=C_3^2(\frac{1}{2})^2\cdot(1-\frac{1}{2})=\frac{3}{8}$

$P(Y=3)=(\frac{1}{2})^3=\frac{1}{8}$

∴ Y 的分布列为



6人 8 18 34 16 8 6

$\bar{x}=150\times0.06+160\times0.08+170\times0.18$

$+180\times0.34+190\times0.16+200\times0.08$

$+210\times0.06=179$(个)

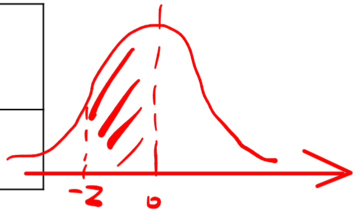
∴ $\mu=179$.

∴ $E(Y)=np=3\times\frac{1}{2}=\frac{3}{2}$

$D(Y)=np(1-p)=3\times\frac{1}{2}\times\frac{1}{2}=\frac{3}{4}$

6. 某地在每周六的晚上 8 点到 10 点半举行灯光展，灯光展涉及到 10000 盏灯，每盏灯在某一时刻亮灯的概率均为 $p(0<p<1)$，并且是否亮灯彼此相互独立. 现统计了其中 100 盏灯在一场灯光展中亮灯的时长 (单位：min)，得到下面的频数表：

亮灯时长 /min	[50, 60)	[60, 70)	[70, 80)	[80, 90)	[90, 100]
频数	10	20	40	20	10



以样本中 100 盏灯的平均亮灯时长作为一盏灯的亮灯时长.

$p=\dfrac{55\times0.1+65\times0.2+75\times0.4+85\times0.2+95\times0.1}{2.5\times60}=\dfrac{1}{2}$

(1) 试估计 p 的值；

(2) 设 X 表示这 10000 盏灯在某一时刻亮灯的数目.

① $X\sim B(10000,\frac{1}{2})$ $E(X)=10000\times\frac{1}{2}=5000$. $D(X)=10000\times\frac{1}{2}\times\frac{1}{2}=2500$

①求 X 的数学期望 $E(X)$ 和方差 $D(X)$；

② $Z=\dfrac{X-5000}{50}\in(-2,0]$ $P(-2<Z\leq0)=\dfrac{0.9545}{2}=0.47725$

②若随机变量 Z 满足 $Z=\dfrac{X-E(X)}{\sqrt{D(X)}}$，则认为 $Z\sim N(0,1)$. 假设当 $4900<X\leqslant5000$ 时，灯光展处于最佳灯光亮度. 试由此估计，在一场灯光展中，处于最佳灯光亮度的时长 (结果保留为整数).

最佳灯光亮度 $2.5\times60\times0.47725=71.60(min)$

附：

①某盏灯在某一时刻亮灯的概率 p 等于亮灯时长与灯光展总时长的商；

②若 $Z\sim N(0,1)$，则 $P(\mu-\sigma<X\leqslant\mu+\sigma)=0.6827$，$P(\mu-2\sigma<X\leqslant\mu+2\sigma)=0.9545$，$P(\mu-3\sigma<X\leqslant\mu+3\sigma)=0.9973$.

7. 2020 年初，新冠肺炎疫情袭击全国，某省由于人员流动性较大，成为湖北省外疫情最严重的省份之一，截至 2 月 29 日，该省已累计确诊 1349 例患者 (无境外输入病例).

(1) 为了解新冠肺炎的相关特征，研究人员从该省随机抽取 100 名确诊患者，统计他们的年龄数据，得如表的频数分布表：

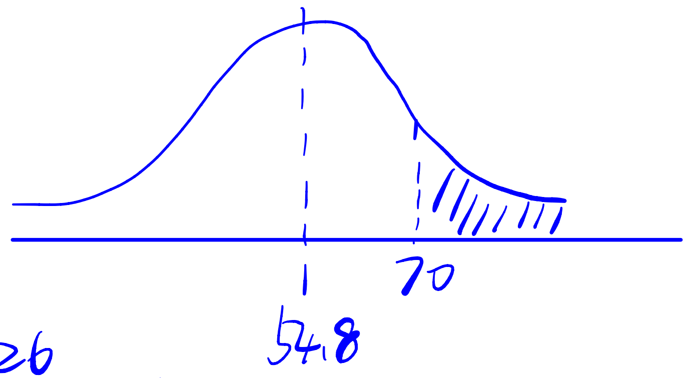
年龄	[10, 20]	(20, 30]	(30, 40]	(40, 50]	(50, 60]	(60, 70]	(70, 80]	(80, 90]	(90, 100]
人数	2	6	12	18	22	22	12	4	2

由频数分布表可以大致认为，该省新冠肺炎患者的年龄服从正态分布 $N(\mu, 15.2^2)$，其中 μ 近似为这 100 名患者年龄的样本平均数 (同一组中的数据用该组区间的中点值作代表). 请估计该省新冠肺炎患者年龄在 70 岁以上 ($\geqslant 70$) 的患者比例；

$\mu = 15\times0.02 + 25\times0.06 + 35\times0.12 + 45\times0.18 + 55\times0.22 + 65\times0.22 + 75\times0.12 + 85\times0.04 + 95\times0.02$

$= 54.8$

$Z \sim N(54.8, 15.2^2)$



$P(Z\geqslant 70) = \dfrac{1-0.6826}{2} = 15.87\%$

赵礼显数学

(2) 截至 2 月 29 日，该省新冠肺炎的密切接触者(均已接受检测)中确诊患者约占 10%，以这些密切接触者确诊的频率代替 1 名密切接触者确诊发生的概率，每名密切接触者是否确诊相互独立. 现有密切接触者 20 人，为检测出所有患者，设计了如下方案：将这 20 名密切接触者随机地按 n ($1<n<20$ 且 n 是 20 的约数) 个人一组平均分组，并将同组的 n 个人每人抽取的一半血液混合在一起化验，若发现新冠病毒，则对该组的 n 个人抽取的另一半血液逐一化验，记 n 个人中患者的人数为 X_n，以化验次数的期望值为决策依据，试确定使得 20 人的化验总次数最少的 n 的值.

参考数据：若 $Z \sim N(\mu,\sigma^2)$，则 $P(\mu-\sigma<Z<\mu+\sigma)=0.6826$，$P(\mu-2\sigma<Z<\mu+2\sigma)=0.9544$，$P(\mu-3\sigma<Z<\mu+3\sigma)=0.9973$，$0.9^4\approx 0.66$，$0.9^5\approx 0.59$，$0.9^{10}\approx 0.35$.

Y_n 表示一组 n 个人化验次数，1，$n+1$，　$n=2, 4, 5, 10$.

$P(Y_n=1)=\left(\frac{9}{10}\right)^n$

$P(Y_n=n+1)=1-\left(\frac{9}{10}\right)^n$

$E(Y_n)=\left(\frac{9}{10}\right)^n+(n+1)-(n+1)\cdot\left(\frac{9}{10}\right)^n=(n+1)-n\cdot\left(\frac{9}{10}\right)^n$

化验总次数 $f(n)=E(Y_n)\cdot\frac{20}{n}=20+\frac{20}{n}-20\cdot\left(\frac{9}{10}\right)^n$

$\because f(2)=20+10-20\times 0.9^2$
$=13.8$

$f(4)=20+5-20\times 0.9^4\approx \underline{11.8}$

$f(5)=12.2$.

$f(10)\approx 15$

当 $n=4$ 时化验次数最少.