目录

1 概率统计 回归教材训练 1

1. 四名同学各抛骰子 5 次，分别记录每次骰子出现的点数。根据四名同学的统计结果，可以判断出一定没有出现点数 6 的是

   A. 平均数为 3，中位数为 2　　B. 中位数为 3，众数为 2

   C. 平均数为 2，方差为 2.4　　D. 中位数为 3，方差为 2.8

2. 已知总体划分为 3 层，通过分层随机抽样，各层抽取的样本、样本均值和样本方差分别为：$l, \bar{x}, s_1^2; m, \bar{y}, s_2^2; n, \bar{z}, s_3^2$，记总体的样本均值为 $\bar{w}$，样本方差为 s^2，证明：

$$\bar{w} = \frac{l}{l+m+n}\bar{x} + \frac{m}{l+m+n}\bar{y} + \frac{n}{l+m+n}\bar{z},$$

$$s^2 = \frac{1}{l+m+n}\left\{l[s_1^2 + (\bar{x}-\bar{w})^2] + m[s_2^2 + (\bar{y}-\bar{w})^2] + n[s_3^2 + (\bar{z}-\bar{w})^2]\right\}.$$

3. 在对树人中学高一年级学生身高的调查中，采用样本量比例分配的分层随机抽样。如果不知道样本数值，只知道抽取了男生 25 人，均值和方差分别为 170 和 10，抽取了女生 25 人，均值和方差分别为 160 和 40。你能由这些数据计算出总样本的方差，并对高一年级全体学生的身高方差作出估计吗？

4. 从两名男生（记为 B_1 和 B_2）、两名女生（记为 G_1 和 G_2）中抽取两人。

   (1) 分别计算有放回简单随机抽样、不放回简单随机抽样和按性别等比例分层随机抽样下，抽到的两人都是男生的概率；

   (2) 比较这三种抽样方式在避免极端样本（例如全是男生的样本）上的优劣。

5. 从 1～20 这 20 个整数中随机选择一个数，设事件 A 表示选到的数能被 2 整除，事件 B 表示选到的数能被 3 整除。求下列事件的概率：

   (1) 这个数既能被 2 整除也能被 3 整除；

   (2) 这个数能被 2 整除或能被 3 整除；

(3) 这个数既不能被 2 整除也不能被 3 整除。

6. 互为对立的两个事件是非常特殊的一种事件关系。请分别证明如果事件 A 与事件 B 相互独立，那么 A 与 $\overline{B}$，$\overline{A}$ 与 B，$\overline{A}$ 与 $\overline{B}$ 是独立的。

7. 一个均匀的正八面体，八个面分别标以数字 1 到 8，任意拨捏一次这个正八面体，观察它与地面接触的面上的数字，得到样本空间为 $\Omega = \{1,2,3,4,5,6,7,8\}$。构造适当的事件 A,B,C，使得 $P(A)P(B)P(C) = P(ABC)$ 成立，但不满足 A,B,C 三个事件是**两两独立**的。

8. “用事件 A 发生的频率 $f_n(A)$ 估计概率 $P(A)$，重复试验次数 n 越大，估计的就越精确”，判断这种说法是否正确，并举例说明。

9. 有两个盒子，其中 1 号盒子中有 95 个红球，5 个白球；2 号盒子中有 95 个红球，5 个白球。现从两个盒子中任意选择一个，再从中任意摸出一个球。如果摸到的是红球，你认为选择的是哪个盒子？做出你的推断，你的推断犯错误的概率是多少？

10. (1) 4 名同学分别报名参加学校的足球队、篮球队、乒乓球队，每人限报其中的一个运动队，不同报法的种数是 3^4 还是 4^3？

(2) 3 个班分别从 5 个景点中选择一处游览，不同选择的种数是 3^5 还是 5^3？

11. 在国庆长假期间，要从 7 人中选择若干人在 7 天假期值班（每天只需 1 人值班），不出现同一人连续值班 2 天，有多少种可能的安排方法？

12. 2160 有多少个不同的正因数？这些正因数的和是多少？

13. 用 $0\sim 9$ 这 10 个数字，可以组成多少个没有重复数字的三位数？

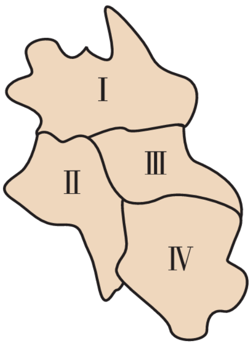
14. 求证：

(1) $1+\sum\limits_{k=1}^{n} k\cdot A_k^k = A_{k+1}^{k+1}$;

(2) $\dfrac{(n+1)!}{k!}-\dfrac{n!}{(k-1)!}=\dfrac{(n-k+1)\cdot n!}{k!} \quad (k\le n)$.

15. 证明等式 $C_n^k\cdot C_{n-k}^{m-k}=C_m^k\cdot C_n^m$ 并构造一个实际背景，说明这个等式的意义。

16. 如图，现要用 5 种不同的颜色对某市的 4 个区县地图进行着色，要要求有公共边的两个地区不能用同一种颜色，共有几种不同的着色方法？



17. (1) 平面内有两组平行线，一组有 m 条，另一组有 n 条，这两组平行线相交，可以构成多少个平行四边形？

(2) 空间有三组平行平面，第一组有 m 个，第二组有 n 个，第三组有 l 个，不同两组的平面都相交，且交线不都平行，可以构成多少个平行六面体？

18. 移动互联网给人们的沟通交流带来了方便。某种移动社交软件平台，既可以供用户彼此添加“好友”单独交流，又可以供多个用户建立一个“群”（“群里”的人彼此不一定是“好友”关系）共同交流。如果某人在平台上发了信息，其他的“好友”都可以看到，但“群”里的非“好友”不能看到。现在这个“群”里有一个 10 人的“群”，其中 1 人在平台上发了一条信息，“群”里有 3 人说看到，不能看到的有 7 人。那么这个“群”里与信息发送人是“好友”关系的情况可能有多少种？

19. 证明：

(1) $(n+1)^n-1$ 能被 n^2 整除；

(2) $99^{10}-1$ 能被 1000 整除；

(3) 55^{55} 除以 8 所得的余数是 7。

20. 求证：

$$2^n-C_n^1\times 2^{n-1}+C_n^2\times 2^{n-2}+\cdots+(-1)^{n-1}C_n^{n-1}\times 2+(-1)^n=1.$$

21. (1) 求 $\left(9x+\dfrac{1}{3\sqrt{x}}\right)^{18}$ 的展开式的常数项；

(2) 已知 $(1+\sqrt{x})^n$ 的展开式中第 9 项、第 10 项、第 11 项的二项式系数构成等差数列，求 n；

(3) 求 $(1+x+x^2)(1-x)^{10}$ 的展开式中 x^4 的系数；

(4) 求 $(x^2+x+y)^5$ 的展开式中 x^5y^2 的系数。

22. 在 $(1+x)^3+(1+x)^4+\cdots+(1+x)^{n+2}$ 的展开式中，含 x^2 项的系数是多少？

2 概率统计 回归教材训练 2

1. 证明：当 $P(AB)>0$ 时，$P(ABC)=P(A)P(B|A)P(C|AB)$。依据你能发现计算 $P(A_1A_2\cdots A_n)$ 的公式吗？

2. 已知 3 张奖券中只有 1 张有奖，甲、乙、丙 3 名同学依次不放回地各随机抽取 1 张。

 (1) 他们中奖的概率与抽奖的顺序有关吗？

 (2) 推广到 n 张奖券和 n 名同学，请你验证 (1) 的结论是否仍然成立。

3. 已知 $P(A)>0, P(B)>0, P(B|A)=P(B)$，证明：$P(A|B)=P(A)$

4. 在 A、B、C 三个地区暴发了流感，这三个地区分别有 6%、5%、4% 的人患了流感。假设这三个地区的人数的比例为 5:7:8，现从这三个地区中任意选取一个人。

 (1) 求这个人患流感的概率；

 (2) 如果此人患流感，求此人选自 A 地区的概率。

5. 甲、乙、丙三人相互做乒乓球训练，第 1 次由甲将乒乓球传出，每次传球时，传球者都可能地将球传给两人中的任何一人。求 n 次乒乓球在甲手中的概率。

6. 证明： $E(aX+b)=aE(X)+b,\ D(aX+b)=a^2D(X).$

7. 设 $E(X)=\mu$， a 是不等于 μ 的常数，探讨 X 相对于 μ 的偏离程度 $E\left[(X-\mu)^2\right]$ 与 X 相对于 a 的偏离程度 $E\left[(X-a)^2\right]$ 的大小关系，并说明结论的意义。

8. 某种资格证考试，每位考生一年内最多有 3 次考试机会。一旦某次考试通过，便可领取资格证书，不再参加以后的考试，否则继续参加考试，直到用完 3 次机会。李明决定参加考试，如他每次参加考试通过的概率依次为 0.6，0.7，0.8，且每次考试是否通过相互独立，试求：

   (1) 李明在一年内参加考试数 X 的分布列；

   (2) 李明在一年内领取到资格证书的概率。

9. 一个数轴上的质点在随机外力的作用下，从原点 O 出发，每隔 1 秒等可能地向左或向右移动一个单位，共移动 6 次。求下列事件的概率：

   (1) 质点回到原点；

   (2) 质点位于 4 的位置。

10. 设 $X \sim B(n, p)$，求证：

   (1) $\sum_{k=0}^{n} P\{X=k\} = 1$;

   (2) 先证明组合恒等式 $kC_n^k = nC_{n-1}^{k-1}$，再证明

$$E(X) = \sum_{k=0}^{n} kP\{X=k\} = np.$$

   (3) $D(X) = E(X^2) - E^2(X) = \sum_{k=0}^{n} k^2 P\{X=k\} - n^2p^2 = np(1-p)$ (Hint: $k^2 = (k+1)k - k$.)

11. 一般地，假设一批产品共有 N 件，其中有 M 件次品。从 N 件产品中随机抽取 n 件（不放回），用 X 表示抽取的 n 件产品中的次品数，则 X 的分布列为：

$$P(X=k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \quad k = m, m+1, m+2, \ldots, r.$$

其中，$n, N, M \in \mathbb{N}^*, M \leq N, n \leq N, m = \max(0, n-N+M), r = \min(n, M)$。即 X 服从超几何分布，求证：

(1) $\sum_{k=m}^{r} C_M^k C_{N-M}^{n-k} = C_N^n$;

(2) 利用组合恒等式 $kC_n^k = nC_{n-1}^{k-1}$ 和 (1) 所证明结论证明

$$E(X) = \sum_{k=m}^{r} kP\{X = k\} = \frac{nM}{N}.$$

(3) $D(X) = E(X^2) - E^2(X) = \sum_{k=0}^{n} k^2 P\{X = k\} - \frac{n^2M^2}{N^2} = \frac{nM(N-M)(N-n)}{N^2(N-1)}$. (Hint: $k^2 = (k+1)k - k$.)

(4) 可以看出，当 $N >> n$ 时，不放回抽取几乎没有影响每次抽到次品的概率，于是可以用二项分布近似超几何分布，也即

$$D(X) \approx n\frac{M}{N}\left(1 - \frac{M}{N}\right)$$

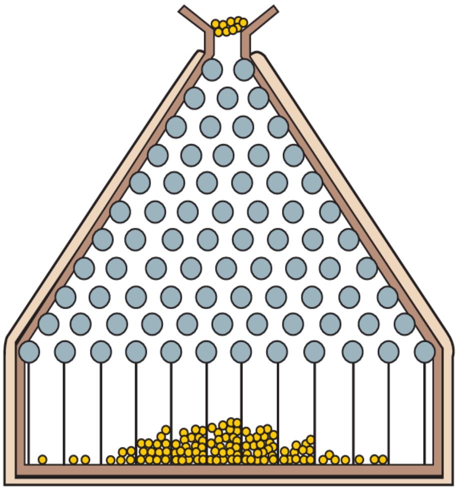
12. 一个袋子中有 10 个大小相同的球，其中有 4 个黄色球，6 个白球，从中随机地抽出 5 个球作为样本。用 X 表示样本中黄色球的个数。

(1) 分别求有放回摸球和不放回摸球，求 X 的分布列；

(2) 分别求有放回摸球和不放回摸球，用样本中黄色球的比例估计总体黄色球的比例，求误差的绝对值不超过 0.1 的概率。

(3) 为了使得用样本中黄色球的比例估计总体黄色球的比例尽可能准确，应该采用哪种摸球方式？

13. 下图是一块高尔顿板的示意图。在一块木板上钉着与排列相互平行但相互错开的圆柱形小木钉，小木钉之间留有适当的空隙作为通道，前面挡有一块玻璃。将小球从顶端放入，小球下落的过程中，每次碰到小木钉后都会等概率地向左或向右落下，最后落入底部的格子中。格子从左到右分别编号为 0，1，2，…，10，用 X 表示小球最后落入格子的编号，求 X 的分布列。



14. 甲、乙两人选手进行象棋比赛，如果每局比赛甲获胜的概率为 0.6，乙获胜的概率为 0.4，那么采用 3 局 2 胜制还是采用 5 局 3 胜制对甲更有利？

15. 设 $X \sim B(n,p)$，求出使得 $P\{n=k\}$ 最大的 k 的值。

16. 某单位有 10000 名职工，想通过检测的方法筛查乙肝病毒携带者。如果对每个人的血样采一化验，就需要化验 10000 次。统计专家提出了一种化验方法：随机地按 5 人一组分组，然后将各组 5 人的血样混合再化验。如果混合血样是阴性，说明这 5 个人全部阴性；如果混合血样是阳性，说明其中至少有一人的血样是阳性，需要再分别化验一次。

    (1) 按照这种化验方法能减少化验次数吗？

    (2) 如果携带病毒的人只有 2%，按照 k 个人一组，k 取多大时化验次数最少？

17. 根据分类变量 X 与 Y 的成对样本数据，计算得到 $\chi^2 = 2.974$。依据 $\alpha = 0.05$ 的独立性检验 $P\{\chi^2 \geq 3.841 = x_\alpha\} = 0.05 = \alpha$，结论为：

A. 变量 X 与 Y 不独立

B. 变量 X 与 Y 不独立，这个结论犯错误的概率不超过 0.05

C. 变量 X 与 Y 独立

D. 变量 X 与 Y 独立，这个结论犯错误的概率不超过 0.05

18. 根据变量 Y 和 x 的成对样本数据，由一元线性回归模型

$$\hat{y} = bx + a + e, \quad E(e) = 0, \quad D(e) = \sigma^2$$

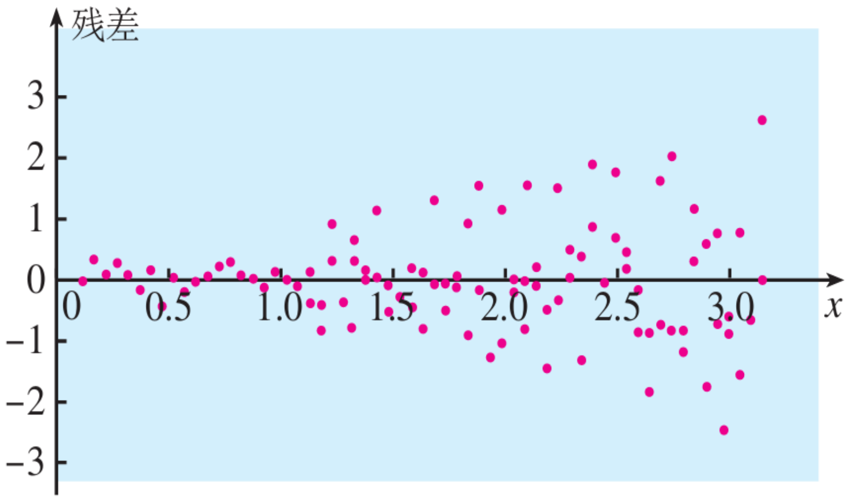
得到经验回归模型，模型误差如图所示。模型误差

A. 满足一元线性回归模型的所有假设

B. 不满足一元线性回归模型的 $E(e) = 0$ 的假设

C. 不满足一元线性回归模型的 $D(e) = \sigma^2$ 的假设

D. 不满足一元线性回归模型的 $E(e) = 0$ 和 $D(e) = \sigma^2$ 的假设



19. 如果散点图中所有的散点都落在一条斜率为非 0 的直线上，请回答下列问题：

(1) 解释变量和响应变量的关系是什么？

(2) R^2 是多少？

20. 设变量 Y 与 x 之间高度线性相关，且满足一元线性回归模型

$$\begin{cases} Y = bx + a + e \\ E(e) = 0,\, D(e) = \sigma^2 \end{cases}$$

$(x_i, y_i)\ (i = 1, 2, \cdots, n)$ 为已知样本数据。用使得残差平方和 $Q(a,b) = \sum\limits_{i=1}^{n}(y_i - bx_i - a)^2$ 取得最小值时的 $\hat{a}, \hat{b}$ 作为 a, b 的估计值，称为 a, b 的最小二乘估计值。证明：

(1) 对任意实数 k 都有 $\sum\limits_{i=1}^{n}[(y_i - \bar{y}) - k(x_i - \bar{x})] = 0$;

(2) $(\bar{x}, \bar{y})$ 在经验回归直线 $\hat{y} = \hat{b}x + \hat{a}$ 上；

(3) $\hat{b} = \dfrac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}$.

21. 设 X 和 Y 为定义在样本空间 Ω 上取值于 $\{0,1\}$ 的成对分类变量。$\{X = 0\}$ 和 $\{X = 1\}$，$\{Y = 0\}$ 和 $\{Y = 1\}$ 都是互为对立事件。记零假设 $H_0 : P(Y = 1 \mid X = 0) = P(Y = 1 \mid X = 1)$，这里，$P(Y = 1 \mid X = 0)$ 表示从 $X = 0$ 的条件下判断 $Y = 1$ 的概率。

(1) 叙述分类变量 X 与 Y 独立的定义；

(2) 证明 H_0 成立等价于分类变量 X 与 Y 独立。

22. 为比较甲、乙两所学校学生的数学水平，采用简单随机抽样的方法抽取88名学生．通过测验得到了如下数据。分别记学校和数学成绩为分类变量 X 与 Y。

学校	数学成绩		合计
	不优秀 $(Y=0)$	优秀 $(Y=1)$	
甲校 $(X=0)$	33	10	43
乙校 $(X=1)$	38	7	45
合计	71	17	88

(1) 若利用频率稳定于概率的原理，分类变量 X 与 Y 是否独立？

(2) 依据 $\alpha=0.1$ 的独立性检验，证明：学校和数学成绩独立；

(3) 如果上表中所有数据都扩大的原来的 10 倍，在相同的检验标准下，再用 $\alpha=0.1$ 的独立性检验推断学校和数学成绩之间的关系，结论还一样吗？请你试着解释其中的原因。

附： χ^2 独立性检验中可能用到的小概率值 α 与对应的临界值 x_α 如下，即 $P\{\chi^2\geq x_\alpha\}=\alpha$:

α	0.1	0.05	0.01	0.005	0.001
x_α	2.706	3.841	6.635	7.879	10.828