共 8 题，前 7 题每题 12 分，第 8 题 16 分

1．设总体 $X \sim N(1,4)$，$Y \sim N(-0.5,2)$，且 $X_1, X_2, \cdots, X_{25}$ 与 $Y_1, Y_2, \cdots, Y_{10}$ 分别为来自总体 X 与 Y 的样本。$\bar{X}$ 与 $\bar{Y}$ 为样本均值，S_X^2 与 S_Y^2 为样本方差。求 $P\{\bar{X} < \bar{Y}\}$，$P\{S_X^2 > 3S_Y^2\}$。

解：$U = \dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} = \dfrac{(\bar{X}-\bar{Y})-1.5}{0.6} \sim N(0,1)$，

$$P\{\bar{X} < \bar{Y}\} = P\left\{\frac{(\bar{X}-\bar{Y})-1.5}{0.6} < -2.5\right\} = \Phi(-2.5) = 0.0062；$$

$$F = \frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} = \frac{1}{2}\frac{S_X^2}{S_Y^2} \sim F(n_1-1, n_2-1) = F(24,9)，$$

$$P\{S_X^2 > 3S_Y^2\} = P\left\{\frac{1}{2}\frac{S_X^2}{S_Y^2} > 1.5\right\} = 0.27。$$

2．设总体 X 的密度函数 $p(x;\theta) = (\theta-1)x^{-\theta} I_{x>1},\ \ (\theta>2)$，$X_1, X_2, \cdots, X_n$ 为样本。

（1）求数学期望 $E(X)$ 以及参数 θ 的矩估计 $\hat{\theta}_1$。

（2）写出似然函数 $L(\theta)$，并求 θ 的最大似然估计 $\hat{\theta}_2$。

解：$E(X) = \int_1^{+\infty} (\theta-1)x^{1-\theta}dx = \dfrac{\theta-1}{2-\theta}x^{2-\theta}\Big|_1^{+\infty} = \dfrac{\theta-1}{\theta-2}$，有 $\theta = \dfrac{2E(X)-1}{E(X)-1}$，$\hat{\theta}_1 = \dfrac{2\bar{X}-1}{\bar{X}-1}$；

$$L(\theta) = (\theta-1)^n (x_1x_2\cdots x_n)^{-\theta} I_{x_1, x_2, \cdots, x_n > 1}，$$

当 $x_1, x_2, \cdots, x_n > 1$ 时，$\ln L(\theta) = n\ln(\theta-1) - \theta\ln(x_1x_2\cdots x_n)$，

令 $\dfrac{d\ln L(\theta)}{d\theta} = \dfrac{n}{\theta-1} - \ln(x_1x_2\cdots x_n) = 0$，得 $\theta = \dfrac{n}{\ln(x_1x_2\cdots x_n)} + 1$，

且 $\dfrac{d^2\ln L(\theta)}{d\theta^2} = -\dfrac{n}{(\theta-1)^2} < 0$，可知 $\hat{\theta}_2 = \dfrac{n}{\ln(X_1X_2\cdots X_n)} + 1$。

3． 设总体 X 服从正态分布 $N(\mu,\sigma^2)$，抽取容量为 16 的样本，测得样本均值 $\overline{x}=37$，样本方差 $s^2=8^2$，求

（1）总体期望 μ 的 95%置信区间。

（2）总体方差 σ^2 的 95%置信区间。

解：总体期望 μ 的 95%置信区间

$$\left[\overline{x}\pm t_{0.975}(15)\frac{s}{\sqrt{n}}\right]=\left[37\pm 2.1314\times\frac{8}{4}\right]=[32.7372, 41.2628]；$$

总体方差 σ^2 的 95%置信区间

$$\left[\frac{(n-1)s^2}{\chi^2_{0.975}(15)},\frac{(n-1)s^2}{\chi^2_{0.025}(15)}\right]=\left[\frac{15\times 8^2}{27.4884},\frac{15\times 8^2}{6.2621}\right]=[34.9238,153.2907]。$$

4． 为了比较甲乙两个班的概率论考试成绩，分别从两个班上随机抽取 11 名和 9 名同学，根据他们的考试成绩计算得 $\overline{x}=78, s_1^2=6^2, \overline{y}=73, s_2^2=5^2$。并设两个班的考试成绩都服从正态分布。在显著水平 $\alpha=0.05$ 下作如下检验。

（1）检验其方差有无显著差异，并计算 p 值。

（2）利用第（1）问的结果，检验甲班是否明显比乙班成绩高，并计算 p 值。

解：$H_0:\sigma_1^2=\sigma_2^2 \quad vs \quad H_1:\sigma_1^2\neq\sigma_2^2$，$F=\dfrac{S_x^2/\sigma_1^2}{S_y^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$，

$$W=\{F\le f_{0.025}(10,8)\text{ 或 }F\ge f_{0.975}(10,8)\}=\{F\le 0.2594\text{ 或 }F\ge 4.2951\}，$$

$f=\dfrac{s_x^2}{s_y^2}=\dfrac{6^2}{5^2}=1.44\notin W$， $p=2P\{F\ge 1.44\}=0.618$，方差无显著差异；

$H_0:\mu_1\le\mu_2 \quad vs \quad H_1:\mu_1>\mu_2$，$T=\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-1)$，

$W=\{t \ge t_{0.95}(18)\}=\{t \ge 1.7341\}$，

$$s_w=\sqrt{\frac{(n_1-1)s_x^2+(n_2-1)s_2^2}{n_1+n_2-2}}=\sqrt{\frac{10\times 6^2+8\times 5^2}{18}}=5.5777,$$

$t=\dfrac{78-73}{5.5777\times\sqrt{\dfrac{1}{11}+\dfrac{1}{9}}}=1.9944\in W$， $p=P\{T\ge 1.9944\}=0.0307$，甲班明显比乙班成绩高。

5. 掷一枚骰子 60 次，结果如下

点数	1	2	3	4	5	6
出现次数	10	5	7	17	6	15

试检验这枚骰子是否均匀？

（$\alpha=0.05$，请写出假设、检验统计量及其分布、拒绝域、决策，并计算 p 值）

解：$H_0: p_i=\dfrac{1}{6},\quad i=1,2,\cdots,6$， $\chi^2=\displaystyle\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}\sim\chi^2(r-1)$，

$W=\{\chi^2\ge\chi^2_{0.95}(5)\}=\{\chi^2\ge 11.0705\}$，

$\chi^2=\dfrac{0^2}{10}+\dfrac{5^2}{10}+\dfrac{3^2}{10}+\dfrac{7^2}{10}+\dfrac{4^2}{10}+\dfrac{5^2}{10}=12.4\in W$， $p=P\{\chi^2\ge 12.4\}=0.03$，可以认为这枚骰子不均匀。

6. 现有三种食品，其蛋白质含量都服从正态分布，且方差相等。为了检验其蛋白质含量是否存在显著差异。从每种食品中各取 5 份，测量蛋白质含量得计算表

食品	蛋白质含量					T_i	T_i^2	Σy_{ij}^2
A_1	19.5	17.9	20	19.8	18.4	95.6	9139.36	1831.26
A_2	16.4	18.4	18.1	17.8	16.4	87.1	7586.41	1520.93
A_3	17.3	18.3	17.6	18.4	18.3	89.9	8082.01	1617.39
Σ						272.6	24807.78	4969.58

（1）检验三种食品蛋白质含量有无显著差异，写出方差分析表（$\alpha=0.05$）。

（2）求 A_1 平均蛋白质含量 μ_1 的 0.95 置信区间。

解：（1） $H_0: a_1=a_2=a_3$， $F=\dfrac{S_A/(r-1)}{S_e/(n-r)}\sim F(2,12)$，

$W=\{F\ge f_{0.95}(2,12)\}=\{F\ge 3.89\}$。

$$S_T=\sum_{i=1}^{3}\sum_{j=1}^{5}y_{ij}^2-\frac{T^2}{15}=4969.58-\frac{272.6^2}{15}=15.5293,$$

$$S_A=\frac{1}{5}\sum_{i=1}^{3}T_i^2-\frac{T^2}{15}=\frac{24807.78}{5}-\frac{272.6^2}{15}=7.5053,\quad S_e=S_T-S_A=8.0240,$$

方差分析表

来源	平方和	自由度	均方和	F 比	p 值
因子	7.5053	2	3.7527	5.6122	0.0190
误差	8.0240	12	0.6687		
总和	15.5293	14			

$f=5.6122\in W$，故拒绝H_0，可以认为蛋白质含量有显著差异；

（2）μ_1 的 0.95 置信区间为$\left[\bar{Y}_{1\cdot}\pm t_{0.975}(12)\cdot\frac{\hat{\sigma}}{\sqrt{5}}\right]=[18.3232,19.9168]$。

7．某企业近 6 年的利润数据如下表

年份 x	15	16	17	18	19	20
利润 y	67	82	91	113	126	136

经计算得$\bar{x}=17.5$，$\bar{y}=102.5$，$l_{xx}=17.5$，$l_{xy}=249.5$，$l_{yy}=3597.5$。

（1）建立利润对年份的一元线性回归方程$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x$。

（2）对建立的回归方程作显著性检验，列出方差分析表。（$\alpha=0.05$）

（3）求今年$x_0=21$时，利润Y的预测值$\hat{Y}_0=\hat{\beta}_0+\hat{\beta}_1x_0$及 0.95 预测区间。

解：（1）$\hat{\beta}_1=\frac{l_{xy}}{l_{xx}}=14.2571$，$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}=-147$，$\hat{Y}=-147+14.2571x$；

（2）$H_0:\beta_1=0$，$F=\frac{S_R}{S_e/4}\sim F(1,4)$，$W=\{F\ge f_{0.95}(1,4)\}=\{F\ge 7.71\}$。

$$S_T=l_{yy}=3597.5,\quad S_R=\frac{l_{xy}^2}{l_{xx}}=3557.1571,\quad S_e=S_T-S_A=40.3429,$$

方差分析表

来源	平方和	自由度	均方和	F 比	p 值
因子	3557.1571	1	3557.1571	352.6926	0.000048
误差	40.3429	4	10.0857		
总和	3597.5	5			

$f=352.6926\in W$，故拒绝H_0，回归方程显著。

（3）$x_0=21$时，$\hat{y}_0=-147+14.2571\times 21=152.4$，其 0.95 预测区间

$$\left[\hat{y}_0\pm t_{0.975}(n-2)\cdot\sqrt{\frac{S_e}{n-2}}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}}\right]=[140.3533,164.4467]。$$

8．总体X服从两点分布，满足$P\{X=1\}=\theta$，$P\{X=0\}=1-\theta$，$(0<\theta<1)$，且$X_1,X_2,\cdots,X_n$为样本。

（1）概率θ的点估计是频率$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}$，由此猜测$g(\theta)=\theta^2$的点估计为$\bar{X}^2$。根据$Y=n\bar{X}=\sum_{i=1}^{n}X_i$服从二项分布$b(n,\theta)$计算$E(\bar{X}^2)$，判断$\bar{X}^2$是否$g(\theta)=\theta^2$的无偏估计？

（2）如果$\bar{X}^2$不是$g(\theta)=\theta^2$的无偏估计，请根据$E(\bar{X}^2)$的结果及$E(\bar{X})=\theta$修偏得到由$\bar{X}$与$\bar{X}^2$构成的估计量$\hat{g}$，使得$\hat{g}$是$g(\theta)=\theta^2$的无偏估计，即$E(\hat{g})=\theta^2$。

（3）由$p(x;\theta)=\theta^x(1-\theta)^{1-x},\ x=0,1$，证明$\theta$的 Fisher 信息量$I(\theta)=\frac{1}{\theta(1-\theta)}$，并求$g(\theta)=\theta^2$的 C-R 下界。

（4）根据$Y=n\bar{X}=\sum_{i=1}^{n}X_i$服从二项分布$b(n,\theta)$，可得

$$\mathrm{Var}(Y^2-Y)=n(n-1)[2\theta^2+(4n-8)\theta^3-(4n-6)\theta^4],$$

根据此结论求出$\mathrm{Var}(\hat{g})$。$\mathrm{Var}(\hat{g})$与$g(\theta)=\theta^2$的 C-R 下界相差多少？$\hat{g}$是否$g(\theta)=\theta^2$的有效估计？

（5）写出样本联合概率函数$p=p(x_1,x_2,\cdots,x_n;\theta)$，求偏导数$\frac{\partial p}{\partial\theta}$并化简得到它与$p$的关系，证明$\hat{g}$是$g(\theta)=\theta^2$的 UMVUE。

解：$E(\bar{X}^2)=\mathrm{Var}(\bar{X})+(E\bar{X})^2=\frac{\theta(1-\theta)}{n}+\theta^2=\frac{1}{n}\theta+\frac{n-1}{n}\theta^2\neq\theta^2$，

故 $\bar{X}^2$ 不是 $g(\theta)=\theta^2$ 的无偏估计；

因 $E\left(\dfrac{n\bar{X}^2-\bar{X}}{n-1}\right)=\theta^2$，即 $\hat{g}=\dfrac{n\bar{X}^2-\bar{X}}{n-1}$ 是 $g(\theta)=\theta^2$ 的无偏估计；

$\ln p(x;\theta)=x\ln\theta+(1-x)\ln(1-\theta)$， $\dfrac{\partial\ln p(x;\theta)}{\partial\theta}=\dfrac{x}{\theta}-\dfrac{1-x}{1-\theta}=\dfrac{x-\theta}{\theta(1-\theta)}$，

$$I(\theta)=E\left[\left(\frac{\partial\ln p(X;\theta)}{\partial\theta}\right)^2\right]=\frac{1}{\theta^2(1-\theta)^2}E[(X-\theta)^2]=\frac{1}{\theta(1-\theta)},$$

$g(\theta)=\theta^2$ 的 C-R 下界为 $\dfrac{[g'(\theta)^2]}{nI(\theta)}=\dfrac{4\theta^3(1-\theta)}{n}$；

$$\mathrm{Var}(\hat{g})=\frac{1}{n^2(n-1)^2}\mathrm{Var}(Y^2-Y)=\frac{1}{n(n-1)}[2\theta^2+(4n-8)\theta^3-(4n-6)\theta^4],$$

$$\mathrm{Var}(\hat{g})-\frac{[g'(\theta)^2]}{nI(\theta)}=\frac{1}{n(n-1)}[2\theta^2+(4n-8)\theta^3-(4n-6)\theta^4]-\frac{4\theta^3(1-\theta)}{n}=\frac{2\theta^2(1-\theta)^2}{n(n-1)},$$

故 $\hat{g}$ 不是 $g(\theta)=\theta^2$ 的有效估计；

$$p(x_1,x_2,\cdots,x_n;\theta)=\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}=\theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}},\quad x_1,x_2,\cdots,x_n=0\text{ 或 }1,$$

$$\frac{\partial p}{\partial\theta}=n\bar{x}\cdot\theta^{n\bar{x}-1}(1-\theta)^{n-n\bar{x}}-\theta^{n\bar{x}}(n-n\bar{x})(1-\theta)^{n-n\bar{x}-1}=\left(\frac{n\bar{x}}{\theta}-\frac{n-n\bar{x}}{1-\theta}\right)p=\frac{n(\bar{x}-\theta)}{\theta(1-\theta)}p,$$

可知有 $E\varphi=0$，得到 $E(\varphi\bar{X})=0$，进一步得到 $E(\varphi\bar{X}^2)=0$，

从而得到 $E(\varphi\hat{g})=\dfrac{nE(\varphi\bar{X}^2)-E(\varphi\bar{X})}{n-1}=0$，可知 $\hat{g}$ 是 $g(\theta)=\theta^2$ 的 UMVUE。