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，并且相互独立。$\bar{X}$ 与 $\bar{Y}$ 为样本均值，S_X^2 与 S_Y^2 为样本方差。求 $P\{\bar{X} < \bar{Y}\}$，$P\{S_X^2 > 3S_Y^2\}$。

2． 设总体 X 的密度函数 $p(x;\theta) = (\theta-1)x^{-\theta} I_{x>1}, \quad (\theta > 2)$，$X_1, X_2, \cdots, X_n$ 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 $\hat{\theta}_2$。

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 $\bar{x} = 37$，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。

4． 为了比较甲乙两个班的概率论考试成绩，分别独立地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 $\bar{x} = 78, s_1^2 = 6^2, \bar{y} = 73, s_2^2 = 5^2$。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。

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）

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。从每种食品中独立地各取 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。

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 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$。

（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 $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ 及 0.95 预测区间。

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 $\hat{\theta} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$，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 θ 的 Fisher 信息量 $I(\theta) = \dfrac{1}{\theta(1-\theta)}$，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$\dfrac{\partial p}{\partial \theta}$并化简得到它与$p$的关系，证明$\hat{g}$是$g(\theta)=\theta^2$的 UMVUE。

附录：$\Phi(2.5)=0.9938$，$\chi^2_{0.95}(5)=11.0705$，$\chi^2_{0.97}(5)=12.3746$，$\chi^2_{0.025}(15)=6.2621$，

$\chi^2_{0.975}(15)=27.4884$，$t_{0.975}(4)=2.7764$，$t_{0.975}(12)=2.1788$，$t_{0.975}(15)=2.1314$，

$t_{0.95}(18)=1.7341$，$t_{0.9693}(18)=1.9944$，$f_{0.95}(1,4)=7.71$，$f_{0.999952}(1,4)=350$，

$f_{0.95}(2,12)=3.89$，$f_{0.981}(2,12)=5.6122$，$f_{0.975}(8,10)=3.8549$，$f_{0.975}(10,8)=4.2951$，

$f_{0.691}(10,8)=1.4401$，$f_{0.73}(24,9)=1.5$，$f_{0.955}(24,9)=3.0$

数理统计常用公式：

1、单个正态总体 $U=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，$T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$，$\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$；两个独立正态总体 $U=\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$；$F=\dfrac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$；当 $\sigma_1^2=\sigma_2^2$ 但未知时，$T=\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2)$，$S_w=\sqrt{\dfrac{(n_1-1)S_X^2+(n_2-1)S_Y^2}{n_1+n_2-2}}$。

2、参数 θ 的费希尔信息量 $I(\theta)=E\left[\dfrac{\partial}{\partial\theta}\ln p(X,\theta)\right]^2$，$g(\theta)$ 无偏估计的 C-R 下界为 $\dfrac{[g'(\theta)]^2}{nI(\theta)}$。

3、分类数据 χ^2 检验 $\chi^2=\sum_{i=1}^{r}\dfrac{(n_i-np_i)^2}{np_i}\,\dot\sim\,\chi^2(r-1)$；若 p_i 的计算与 k 个未知参数有关，有 $\chi^2=\sum_{i=1}^{r}\dfrac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}\,\dot\sim\,\chi^2(r-k-1)$；独立性检验 $\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\dfrac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}\,\dot\sim\,\chi^2((r-1)(c-1))$。

4、方差分析中 $S_T=\sum_{i=1}^{r}\sum_{j=1}^{m_i}(Y_{ij}-\overline{Y})^2=\sum_{i=1}^{r}\sum_{j=1}^{m_i}Y_{ij}^2-\dfrac{1}{n}T^2$，$S_A=\sum_{i=1}^{r}m_i(\overline{Y}_{i\cdot}-\overline{Y})^2=\sum_{i=1}^{r}\dfrac{T_i^2}{m_i}-\dfrac{1}{n}T^2$，$S_e=\sum_{i=1}^{r}\sum_{j=1}^{m_i}(Y_{ij}-\overline{Y}_{i\cdot})^2=\sum_{i=1}^{r}\sum_{j=1}^{m_i}Y_{ij}^2-\sum_{i=1}^{r}\dfrac{T_i^2}{m_i}$；满足 $S_T=S_e+S_A$，以及 $\dfrac{S_e}{\sigma^2}\sim\chi^2(n-r)$，并且当 H_0：$a_1=a_2=\ldots=a_r=0$ 成立时，$\dfrac{S_A}{\sigma^2}\sim\chi^2(r-1)$，且 S_e 与 S_A 相互独立.

此外，$\overline{Y}\sim N(\mu,\dfrac{\sigma^2}{n})$，$\overline{Y}_{i\cdot}\sim N(\mu_i,\dfrac{\sigma^2}{m})$，$\sigma^2$ 的无偏估计 $\hat{\sigma}^2=\dfrac{S_e}{n-r}=MS_e$.

5、回归分析中，$l_{xx}=\sum(x_i-\overline{x})^2=\sum x_i^2-n\overline{x}^2$，$l_{xy}=\sum(x_i-\overline{x})(y_i-\overline{y})=\sum x_iy_i-n\overline{x}\overline{y}$，$l_{yy}=\sum(y_i-\overline{y})^2=\sum y_i^2-n\overline{y}^2$，$\hat{\beta}_1=\dfrac{l_{xy}}{l_{xx}}\sim N\left(\beta_1,\dfrac{\sigma^2}{l_{xx}}\right)$，$\hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{x}\sim N\left(\beta_0,\left(\dfrac{1}{n}+\dfrac{\overline{x}^2}{l_{xx}}\right)\sigma^2\right)$；

$\hat{Y}_0=\hat{\beta}_0+\hat{\beta}_1x_0\sim N\left(\beta_0+\beta_1x_0,\left[\dfrac{1}{n}+\dfrac{(x_0-\overline{x})^2}{l_{xx}}\right]\sigma^2\right)$，$Y_0-\hat{Y}_0\sim N\left(0,\left[1+\dfrac{1}{n}+\dfrac{(x_0-\overline{x})^2}{l_{xx}}\right]\sigma^2\right)$；

$S_T=\sum(Y_i-\overline{Y})^2=l_{YY}$，$S_R=\sum(\hat{Y}_i-\overline{Y})^2=\hat{\beta}_1^2l_{xx}=\dfrac{l_{xY}^2}{l_{xx}}$，$S_e=\sum(Y_i-\hat{Y}_i)^2=l_{YY}-\dfrac{l_{xY}^2}{l_{xx}}$；

$S_T=S_e+S_R$，以及 $\dfrac{S_e}{\sigma^2}\sim\chi^2(n-2)$，当 H_0：$\beta_1=0$ 成立时，$\dfrac{S_R}{\sigma^2}\sim\chi^2(1)$，且 S_e 与 S_R 独立。