说明：1、样本均值与样本方差分别定义为

$$\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k,\quad S^2=\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\overline{X})^2;$$

2、分位点Q_α都取为上分位点，即$P(X>Q_\alpha)=\alpha$。

共六题，满分 100 分；不必抄题，步骤尽量详细写出。

一（15 分）、假定$X_1,X_2,\dots,X_n$是来自正态总体 $N(\mu,4)$的一组简单随机样本。问样本容量 n 至少为多大才能保证$E|\overline{X}-\mu|\le 0.05$？ $X\sim N(u,4)$. $\overline{X}\sim N(u,\frac{4}{n})$

提示：计算标准正态分布的一阶绝对矩。$\frac{\overline{X}-u}{\sqrt{\frac{4}{n}}}\sim N(0,1)$. $E|\overline{X}-u|=\sqrt{\frac{4}{n}}E\left|\frac{\overline{X}-u}{\sqrt{\frac{4}{n}}}\right|$

$=\sqrt{\frac{4}{n}}\int_{-\infty}^{+\infty}|x|\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\sqrt{\frac{4}{n}}\cdot 2\int_0^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$

$=\frac{2\sqrt{2}}{\sqrt{\pi n}}\le 0.05$

二（20 分）、$X_1,X_2,\dots,X_n$是从如下的总体中抽取的n个简单随机样本：

$$f(x)=(\theta+1)x^\theta,\quad 0<x<1,\theta>-1$$

1（14 分）、分别计算参数θ的矩估计与极大似然估计；

2（6 分）、请给出θ的一种不同于矩估计与极大似然的第三种估计方法，并说明你的理由。

三（15 分）、假定$X_1,X_2,\dots,X_n$是来自正态总体 $N(\mu_1,\sigma^2)$的简单随机样本，$Y_1,Y_2,\dots,Y_m$ 是来自另一个独立的正态总体 $N(\mu_2,\sigma^2)$ 的简单随机样本，μ_1,μ_2,σ^2 都未知，具体构造$3\mu_1-2\mu_2$的一个置信水平$1-\alpha$的一个合理的区间估计。

提示：合理的含义是指所有的样本都应该被使用。

四（10 分）、把一枚骰子随机抛掷 120 次，记录下各点出现的次数：

点数	1	2	3	4	5	6
频数	18	22	20	25	16	19

在检验水平 0.05 下能否认为这枚骰子是均匀的？

提示：使用卡方分析。（卡方拟合优度检验）. 检验一个数据是否来源于已知的总体。

……现从中随机分别抽取了 12，15，13 个学生，记录其成绩如下：

甲学院：70，68，89，60，82，45，43，93，83，36，73，77；

乙学院：89，81，78，35，48，78，91，62，51，76，85，66，74，80.5

丙学院：73，45，79，60，56，68，91，53，71，79，71，61，87。

假设各学院学生成绩服从正态分布并且方差相等，在检验水平 0.05 下

1（10 分）、甲学院与乙学院学生的成绩是否无显著差异？

2（10 分）、甲、乙、丙三个学院学生的成绩是否无显著差异？

提示：三个学院的的样本均值、样本方差分别为：

甲：$\overline{x}_1=68.25$，$s_1^2=348.93$；乙：$\overline{x}_2=70$，$s_2^2=269.57$；

丙：$\overline{x}_3=68.77$，$s_3^2=181.53$；

全部 40 名学生的样本均值与样本方差为：$\overline{x}=69.08$，$s^2=251.61$

六.（20 分）随机调查了某单位 12 名科研人员每周的科研时间(x)与平均每年发表的论文数(y)。假定y关于x具有线性回归 $y=\beta_0+\beta_1x+\varepsilon$，$\varepsilon\sim N(0,\sigma^2)$。具体观测数据如下：

x	9	20	18	33	31	13	25	30	5	47	25	11
y	3.5	5.5	5	6	4.5	6.5	6	5.5	3	7.5	4	2

1.（10 分）计算β_0与β_1的最小二乘估计；

2.（10 分）给出σ^2的估计并且在水平 0.05 下检验 H_0：$\beta_1=0\leftrightarrow H_1$：$\beta_1\ne 0$

提示：自变量x与因变量y各自的样本均值与样本方差分别为：

$\overline{x}=22.25$，$s_x^2=144.39$；$\overline{y}=4.92$，$s_y^2=2.49$；

样本的协方差为 $\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=12.10$。

$L_{xx}=S_x^2\times(n-1)$

$L_{yy}=S_y^2\times(n-1)$

$L_{xy}=\sigma n$

17

一.解:由 $X\sim N(\mu,4) \Rightarrow \bar{X}\sim N(\mu,\frac{4}{n}) \Rightarrow \bar{X}-\mu \sim N(0,\frac{4}{n})$

设 $t=\bar{X}-\mu$，则 $t\sim N(0,\frac{4}{n})$

$$E(|\bar{X}-\mu|)=\frac{2}{\sqrt{n}}\cdot E\left(\frac{1}{2/\sqrt{n}}|\bar{X}-\mu|\right)=\frac{2}{\sqrt{n}}\cdot E\left(\left|\frac{\bar{X}-\mu}{2/\sqrt{n}}\right|\right)$$

$$=\frac{2}{\sqrt{n}}\cdot E(|t|)\le 0.05$$

而 $$E(|t|)=\int_{-\infty}^{+\infty}|x|\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}dx=2\int_0^{+\infty}x\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}dx$$

$$=2\cdot\int_0^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}\cdot d\left(\frac{x^2}{2}\right)=\frac{2}{\sqrt{2\pi}}\cdot\int_0^{+\infty}e^{-\frac{x^2}{2}}d\left(\frac{x^2}{2}\right)=\sqrt{\frac{2}{\pi}}$$

$\therefore \frac{2}{\sqrt{n}}\cdot\sqrt{\frac{2}{\pi}}\le 0.05 \Rightarrow n\ge\frac{3200}{\pi}$　　n 取 1020.

$(1019.1083,\ \pi=3.14)$

二.解:1. ①矩估计如下:

$$E(X)=\int_0^1 x\cdot(\theta+1)\cdot x^{\theta}\cdot dx=\frac{\theta+1}{\theta+2}\cdot x^{\theta+2}\Big|_0^1=\frac{\theta+1}{\theta+2}=\bar{X}$$

$\Rightarrow \theta$ 的矩估计为 $\hat{\theta}=\frac{2\cdot\bar{X}-1}{1-\bar{X}}$.

②极大似然估计如下:

$$L(\theta)=\prod_{i=1}^{n}f(x_i)=(\theta+1)^n\cdot(x_1\cdot x_2\cdots x_n)^{\theta}$$

$$\ln L(\theta)=n\ln(\theta+1)+\theta\cdot\sum_{i=1}^{n}\ln(x_i)$$ ~~对θ~~

$$\Rightarrow \frac{\partial\ln L(\theta)}{\partial\theta}=\frac{n}{\theta+1}+\sum_{i=1}^{n}\ln x_i=0$$

$\Rightarrow \theta$ 的极大似然估计 $\hat{\theta}=-\frac{n}{\sum_{i=1}^{n}\ln x_i}-1$

解: 构造 $Z=3X-2Y$, 其中 $X\sim N(\mu_1,\sigma^2)$, $Y\sim N(\mu_2,\sigma^2)$. 由定理1.2.2

$\therefore \bar{X}\sim N(\mu_1,\frac{\sigma^2}{n})$, $\bar{Y}\sim N(\mu_2,\frac{\sigma^2}{m})$, $S_1^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$, $S_2^2=\frac{1}{m}\sum_{i=1}^{m}(y_i-\bar{y})^2$

$\Rightarrow Z=3X-2Y\sim N(3\mu_1-2\mu_2,13\sigma^2)$

$3\bar{X}-2\bar{Y}\sim N(3\mu_1-2\mu_2,\frac{9}{n}\sigma^2+\frac{4}{m}\sigma^2)$.

因为 $$\xi\triangleq\frac{(3\bar{X}-2\bar{Y})-(3\mu_1-2\mu_2)}{\sigma\cdot\sqrt{\frac{9}{n}+\frac{4}{m}}}\sim N(0,1)$$

由定理1.1.1知 $$Q\triangleq\frac{n\cdot S_1^2+m\cdot S_2^2}{\sigma^2}\sim\chi^2(m+n-2)$$

又因 ξ 与 Q 相互独立, 由 t 分布定义知,

$$T=\frac{\xi}{\sqrt{Q/(m+n-2)}}\sim t(m+n-2)$$

即 $$\frac{\dfrac{(3\bar{X}-2\bar{Y})-(3\mu_1-2\mu_2)}{\sigma\cdot\sqrt{\frac{9}{n}+\frac{4}{m}}}}{\sqrt{\dfrac{n\cdot S_1^2+m\cdot S_2^2}{\sigma^2}\cdot\dfrac{1}{m+n-2}}}=\frac{(3\bar{X}-2\bar{Y})-(3\mu_1-2\mu_2)}{\sqrt{n\cdot S_1^2+m\cdot S_2^2}}\cdot\sqrt{\frac{m\cdot n(m+n-2)}{9\cdot m+4n}}\sim t(m+n-2)$$

$\therefore 3\mu_1-2\mu_2$ 的一个置信水平为 $1-\alpha$ 的置信区间为

$$P\left\{\frac{|(3\bar{X}-2\bar{Y})-(3\mu_1-2\mu_2)|}{\sqrt{n\cdot S_1^2+m\cdot S_2^2}}\cdot\sqrt{\frac{mn(m+n-2)}{9m+4n}}<t_{1-\frac{\alpha}{2}}(m+n-2)\right\}=1-\alpha.$$

$$P\left\{3\bar{X}-2\bar{Y}-\sqrt{\frac{(n\cdot S_1^2+m\cdot S_2^2)\cdot(9m+4n)}{m\cdot n\cdot(m+n-2)}}\cdot t_{1-\frac{\alpha}{2}}(m+n-2)<3\mu_1-2\mu_2\right.$$
$$\left.<3\bar{X}-2\bar{Y}+\sqrt{\frac{(n\cdot S_1^2+m\cdot S_2^2)\cdot(9m+4n)}{m\cdot n\cdot(m+n-2)}}\cdot t_{1-\frac{\alpha}{2}}(m+n-2)\right\}=1-\alpha.$$

9.

解：一般取 $m \approx 1.87\cdot(n-1)^{0.4}$，见 P_{82} 例3.3.1

此处取 $m=6$.

$n\cdot p_i = 120\times\frac{1}{6}=20$.

设 H_0：$P(\xi=k)=\frac{1}{6}(k=1,2,\cdots,6)$ 即骰子是均匀的，每次抛掷点数概率为 $\frac{1}{6}$.

ξ = "掷一次出现点数"

由题意 χ^2 检验法：

设总体 ξ 的分布律 H_0：$P(\xi=x_i)=p_i\ (i=1,2\cdots m)$

p_i 中不含未知参数，对 ξ 观测 n 次，其中 $(\xi=x_i)$ 出现 ν_i 次.

$(\sum_{i=1}^{m}\nu_i=n)$，问是否可接受 H_0.

由 P_{87} 公式 3.3.7 $\Rightarrow$ $\chi_n^2=\sum_{i=1}^{m}\frac{(\nu_i-np_i)^2}{np_i}=\frac{1}{20}\left[(18-20)^2+(22-20)^2+(20-20)^2+(15-20)^2+(16-20)^2+(19-20)^2\right]$

$= 2.5$.

取上分位数点时 $C=\chi^2_{\alpha}(m-1-l)$

取下分位数点时 $C=\chi^2_{1-\alpha}(m-1-l)$. $\begin{cases}\text{当 } \chi_n^2 > \chi^2_{1-\alpha}(m-1-l) \text{ 时，拒绝 } H_0.\\ \text{当 } \chi_n^2 < \chi^2_{1-\alpha}(m-1-l) \text{ 时，接受 } H_0.\end{cases}$

l 在这里取 $l=0$，因不含未知参数（l 表示含有未知参数个数）

$C=\chi^2_{0.95}(5)=11.071$.

即 $\chi_n^2 < C=\chi^2_{1-\alpha}(m-1-l)$，因此，接受 H_0，认为骰子是均匀的.

五解 1. 参照 P_{14} 式子(3.2.20) 和 P_{22} 例3.2.9.

此题可归结为下列假设检验问题：

$H_0: a_1=a_2$，$H_1: a_1\neq a_2$.

因为 $\sigma_1^2=\sigma_2^2=\sigma^2$（题中所给条件）由式(3.2.20) H_0 的拒绝域

$\cancel{\mathscr{X}_0=\left\{|\bar{\xi}-\bar{\eta}|>t_{1-\frac{\alpha}{2}}(n_1+n_2-2)\sqrt{\frac{(n_1+n_2)(n_1S_1^2+n_2S_2^2)}{n_1n_2(n_1+n_2-2)}}\right\}}$

又因 $\bar{x}_1=68.25$，$\bar{x}_2=70$，$S_1^2=348.43$，$S_2^2=269.57$

所以 $|\bar{x}_1-\bar{x}_2|=1.75 < t_{0.975}(25)\cdot\sqrt{\frac{27\times(12\times348.43+15\times269.57)}{12\times15\times25}}$

$=\sqrt{44.38956}$

故认为甲学院与乙学院无显著差异.

（其中 $t_{0.975}(25)=2.0595$）

2.

学院 \ 成绩 \ 学生	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
甲学院	70	68	89	60	82	45	43	93	83	36	73	77			
乙学院	89	81	78	35	48	78	91	62	51	76	85	66	74	80	56
丙学院	73	45	79	60	56	68	91	53	71	79	71	61	87		

解 参照 P_{115} 例4.1.1

记 η_1,η_2,η_3 分别为这三个学院学生成绩，即3个总体，$\eta_{i1},\cdots,\eta_{in_i}$ 为 η_i 的样本，

视 $\eta_i\sim N(a_i,\sigma^2)$，我们的问题就归结为判断原假设 $H_0: a_1=a_2=a_3$ 是否成立

解 1. 由题知 $\bar{x}=22.25$，$\bar{y}=4.92$，$S_x^2=144.39$，$S_y^2=2.49$，样本协方差 $\frac{1}{n-1}\sum_{t=1}^{n}(x_t-\bar{x})(y_t-\bar{y})=$

$$L_{xx}=\sum_{t=1}^{n}(x_t-\bar{x})^2=(n-1)\cdot S_x^2=11\times 144.39=1588.29.$$

$$L_{xy}=\sum_{t=1}^{n}(x_t-\bar{x})(y_t-\bar{y})=\underbrace{\sum_{t=1}^{n}x_t y_t-n\cdot\bar{x}\cdot\bar{y}}_{\text{不用}}=145.2$$

$$L_{yy}=\sum_{t=1}^{n}(y_t-\bar{y})^2=(n-1)\cdot S_y^2=11\times 2.49=27.39.$$

$$\Rightarrow \hat{\beta}_1=\frac{L_{xy}}{L_{xx}}=0.0914, \quad \hat{\beta}_0=\bar{y}-\bar{x}\cdot\hat{\beta}_1=2.88635.$$

2. $U_R=\hat{\beta}_1^2\cdot L_{xx}=13.274$　　　(见 P206 式 (5.2.15), (5.2.30)).

$Q_e=L_{yy}-U_R=14.116$

见 P208 式 (5.2.41) $\Rightarrow \hat{\sigma}_e^2=\frac{Q_e}{n-k-1}=\frac{Q_e}{10}=1.4116$　　$(k=1)$

$$F\triangleq\frac{U_R/k}{Q_e/(n-k-1)}=\frac{U_R\cdot 10}{Q_e}=9.4035$$　　(见 P212 Q_e, F …)

对显著性水平 $\alpha=0.05$ 查 F 分布表得临界值 $F_{0.95}(k, n-k-1)=F_{0.95}(1,10)=4.5$

$=4.96$