应用数理统计二轮复习材料

表 1：常用分布一览表

	分布类型	分布形式	分布律 （概率密度）	期望	方差	平方的期望	矩估计	极大似然估计
离散型分布	两点分布 （0-1 分布）	$X \sim B(1,p)$	$P\{X=k\}=p^k(1-p)^{1-k},k=0,1$	p	$p(1-p)$	p	$\hat{p}=\overline{X}$	$\hat{p}=\overline{X}$
	二项分布	$X \sim B(n,p)$	$P\{X=k\}=C_n^k p^k(1-p)^{n-k},k=0,\ldots,n$	np	$np(1-p)$	$n^2p^2+np(1-p)$	$\hat{p}=\dfrac{\overline{X}}{N}$	$\hat{p}=\dfrac{\overline{X}}{N}$
	泊松分布	$X \sim P(\lambda)$	$P\{X=k\}=\dfrac{\lambda^k e^{-\lambda}}{k!},k=0,\ldots,n$	λ	λ	$\lambda^2+\lambda$	$\hat{\lambda}=\overline{X}$	$\hat{\lambda}=\overline{X}$
	超几何分布	$X \sim H(N,M,n)$	$P\{X=k\}=\dfrac{C_M^k C_{N-M}^{n-k}}{C_N^n}$, $k=l_1,l_1+1,\ldots,l_2$, $l_1=\max\{0,n-(N-M)\}$, $l_2=\min\{0,M\}$	$\dfrac{nM}{N}$	$\dfrac{nM}{N}(1-\dfrac{M}{N})(\dfrac{N-n}{N-1})$	--	--	--
	几何分布	$X \sim G(p)$	$P\{X=k\}=(1-p)^{k-1}p,k=1,2,\ldots$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$	$\dfrac{2-p}{p^2}$	--	--
	巴斯卡分布	$X \sim NB(r,p)$	$P(x=k)=C_{k-1}^{r-1}p^r(1-p)^{k-r},p,k=r,r+1,\ldots$	$\dfrac{r}{p}$	$\dfrac{r(1-p)}{p^2}$	$\dfrac{r^2+r(1-p)^2}{p^2}$	--	--

连续型分布	均匀分布	$X\sim U(a,b)$	$f(x)=\begin{cases}\frac{1}{b-a} & ,a<x<b\\ 0 & ,其他\end{cases}$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$	$\frac{(a+b)^2}{4}+\frac{(a-b)^2}{12}$	$\bar{X}\pm\sqrt{3(n-1)/n}S$	$X_{(1)},X_{(n)}$
	指数分布	$X\sim E(\lambda)$	$f(x)=\begin{cases}\lambda e^{-\lambda x} & ,x>0\\ 0 & ,x\le 0\end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{2}{\lambda^2}$	$\hat{\lambda}=1/\bar{X}$	$\hat{\lambda}=1/\bar{X}$
	正态分布	$X\sim N(\mu,\sigma^2)$	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$\mu^2+\sigma^2$	$\hat{\mu}=\bar{X},\hat{\sigma}^2=\frac{n-1}{n}S^2$ $\hat{\sigma}=\sqrt{(n-1)/n}S$	$\hat{\mu}=\bar{X},\hat{\sigma}^2=\frac{n-1}{n}S^2$ $\hat{\sigma}=\sqrt{(n-1)/n}S$
共轭分布	Beta 分布	$X\sim Beta(\alpha,\beta)$	$f(x)=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\frac{\alpha(\beta-\alpha)}{(\alpha+\beta)^2(\alpha+\beta+1)}$	--	--
	Gamma 分布	$X\sim\Gamma(\alpha,\lambda)$	$f(x)=\begin{cases}\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\frac{\alpha(\alpha+1)}{\lambda^2}$	--	--
	逆 Gamma 分布	$X\sim I\Gamma(\alpha,\lambda)$	$f(x)=\begin{cases}\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}e^{-\frac{\lambda}{x}}, & x\ge 0\\ 0, & x<0\end{cases}$	$\frac{\lambda}{\alpha-1}$	$\frac{\lambda^2}{(\alpha-1)^2(\alpha-2)}$	$\frac{\lambda^2+\lambda(\alpha-1)(\alpha-2)}{(\alpha-1)^2(\alpha-2)}$	--	--
	正态分布（共轭）	--	--	$\frac{a_0\sigma^2+n\bar{x}\sigma_0^2}{\sigma^2+n\sigma_0^2}$	$\frac{\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}$	$\frac{(a_0\sigma^2+n\bar{x}\sigma_0^2)^2+\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}$	--	--
备注	共轭分布的一些常用性质： 1. Bete 函数与 Gamma 函数的关系：$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$　　2. Gamma 函数的性质：$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$, $\Gamma(\alpha+1)=\alpha!$ 3. Gamma 分布与逆 Gamma 分布的关系：$X\sim\Gamma(\alpha,\lambda)\Rightarrow\frac{1}{X}\sim\Gamma(\alpha,\lambda)$ 4. （共轭）正态分布的参数说明：$X\sim N(\theta,\sigma^2),\theta\sim N(a_0,\sigma_0^2)\to h(\theta\mid x_1,\dots,x_n)\propto\exp\{-\frac{1}{2\sigma_0^2}(\theta-a_0)^2-\frac{n}{2\sigma^2}(\theta-\bar{x})^2\}$ 5. 常用分布还包括 χ^2 分布、t 分布和 F 分布；其中，χ^2 分布的期望为 n，方差为 2n。							

表 2：常用估计—检验一览表

分布类型	待检验参数	条件	置信区间	检验类型	拒绝域
正态分布	μ	σ^2 已知	$\frac{\sqrt{n}\,\lvert\bar{X}-\mu_0\rvert}{\sigma_0}<u_{\alpha/2}$	正态检验	$\frac{\sqrt{n}\,\lvert\bar{X}-\mu_0\rvert}{\sigma_0}>u_{\alpha/2}$
					$\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma_0}>u_{\alpha}$
					$\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma_0}<-u_{\alpha}$
		σ^2 未知	$\frac{\sqrt{n}\,\lvert\bar{X}-\mu_0\rvert}{S}<t_{\alpha/2}(n-1)$	T 检验	$\frac{\sqrt{n}\,\lvert\bar{X}-\mu_0\rvert}{S}>t_{\alpha/2}(n-1)$
					$\frac{\sqrt{n}(\bar{X}-\mu_0)}{S}>t_{\alpha}(n-1)$
					$\frac{\sqrt{n}(\bar{X}-\mu_0)}{S}<-t_{\alpha}(n-1)$
	σ^2	μ 已知	$\chi^2_{1-\alpha/2}(n)<\sum_{k=1}^{n}\frac{(X_k-\mu_0)^2}{\sigma_0^2}<\chi^2_{\alpha/2}(n)$	χ^2 检验	$\sum_{k=1}^{n}\frac{(X_k-\mu_0)^2}{\sigma_0^2}>\chi^2_{\alpha/2}(n)$ $or\sum_{k=1}^{n}\frac{(X_k-\mu_0)^2}{\sigma_0^2}<\chi^2_{1-\alpha/2}(n)$
					$\sum_{k=1}^{n}\frac{(X_k-\mu_0)^2}{\sigma_0^2}>\chi^2_{\alpha}(n)$
					$\sum_{k=1}^{n}\frac{(X_k-\mu_0)^2}{\sigma_0^2}<\chi^2_{1-\alpha}(n)$

正态分布	σ^2	μ 未知	$\chi^2_{1-\alpha/2}(n-1)<\sum_{k=1}^{n}\frac{(X_k-\bar{X})^2}{\sigma_0^2}<\chi^2_{\alpha/2}(n-1)$	χ^2 检验	$\sum_{k=1}^{n}\frac{(X_k-\bar{X})^2}{\sigma_0^2}>\chi^2_{\alpha/2}(n-1)$ $or\sum_{k=1}^{n}\frac{(X_k-\bar{X})^2}{\sigma_0^2}<\chi^2_{1-\alpha/2}(n\text{-}1)$
					$\sum_{k=1}^{n}\frac{(X_k-\bar{X})^2}{\sigma_0^2}>\chi^2_{\alpha}(n-1)$
					$\sum_{k=1}^{n}\frac{(X_k-\bar{X})^2}{\sigma_0^2}<\chi^2_{1-\alpha}(n\text{-}1)$
	$\mu_1-\mu_2$	$\sigma_1^2\ \sigma_2^2$ 已知	$\frac{\lvert(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\rvert}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}<u_{\alpha/2}$	正态检验	$\frac{\lvert(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\rvert}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}>u_{\alpha/2}$
					$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}>u_{\alpha}$
					$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}<-u_{\alpha}$

正态分布	$\mu_1-\mu_2$	σ_1^2 σ_2^2 未知、相等	$\frac{\lvert(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\rvert}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}<t_{\alpha/2}(n_1+n_2-2)$	T 检验	$\frac{\lvert(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\rvert}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}>t_{\alpha/2}(n_1+n_2-2)$
					$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}>t_{\alpha}(n_1+n_2-2)$
					$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}<-t_{\alpha}(n_1+n_2-2)$
		σ_1^2 σ_2^2 未知、成比例（$\sigma_1^2=a\sigma_2^2$）	$\frac{\lvert(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\rvert}{\sqrt{\frac{(n_1-1)S_1^2+a(n_2-1)S_2^2}{a(n_1+n_2-2)}}\sqrt{\frac{a}{n_1}+\frac{1}{n_2}}}<t_{\alpha/2}(n_1+n_2-2)$	T 检验	$\frac{\lvert(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\rvert}{\sqrt{\frac{(n_1-1)S_1^2+a(n_2-1)S_2^2}{a(n_1+n_2-2)}}\sqrt{\frac{a}{n_1}+\frac{1}{n_2}}}>t_{\alpha/2}(n_1+n_2-2)$
					$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+a(n_2-1)S_2^2}{a(n_1+n_2-2)}}\sqrt{\frac{a}{n_1}+\frac{1}{n_2}}}>t_{\alpha/2}(n_1+n_2-2)$
					$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+a(n_2-1)S_2^2}{a(n_1+n_2-2)}}\sqrt{\frac{a}{n_1}+\frac{1}{n_2}}}<-t_{\alpha/2}(n_1+n_2-2)$
		σ_1^2 σ_2^2 未知、无关系	$\frac{\lvert(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\rvert}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}<u_{\alpha/2}$	近似正态检验	$\frac{\lvert(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\rvert}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}>u_{\alpha/2}$
					$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}>u_{\alpha/2}$
					$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}<-u_{\alpha/2}$

正态分布	σ_1^2/σ_2^2	μ_1 μ_2 已知	$F_{1-\alpha/2}(n_1,n_2) < \frac{n_2}{n_1}\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2} < F_{\alpha/2}(n_1,n_2)$	F 检验	$\frac{n_2}{n_1}\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2} > F_{\alpha/2}(n_1,n_2)$ $or\ \frac{n_2}{n_1}\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2} < F_{1-\alpha/2}(n_1,n_2)$
					$\frac{n_2}{n_1}\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2} > F_{\alpha}(n_1,n_2)$
					$\frac{n_2}{n_1}\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2} < F_{1-\alpha}(n_1,n_2)$
		μ_1 μ_2 未知	$F_{1-\alpha/2}(n_1-1,n_2-1) < S_1^2/S_2^2 < F_{\alpha/2}(n_1-1,n_2-1)$	F 检验	$S_1^2/S_2^2 > F_{\alpha/2}(n_1-1,n_2-1)$ $or S_1^2/S_2^2 < F_{1-\alpha/2}(n_1-1,n_2-1)$
					$S_1^2/S_2^2 > F_{\alpha}(n_1-1,n_2-1)$
					$or S_1^2/S_2^2 < F_{1-\alpha}(n_1-1,n_2-1)$
指数分布	λ	--	$\chi^2_{1-\alpha/2}(2n) < 2n\lambda_0\bar{X} < \chi^2_{\alpha/2}(2n)$	χ^2 检验	$2n\lambda_0\bar{X} < \chi^2_{1-\alpha/2}(2n)$ $or\,2n\lambda_0\bar{X} > \chi^2_{\alpha/2}(2n)$
					$2n\lambda_0\bar{X} < \chi^2_{1-\alpha}(2n)$
					$2n\lambda_0\bar{X} > \chi^2_{\alpha}(2n)$

两点分布	p	--	$\dfrac{\lvert p_s-p_0\rvert}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}<u_{\alpha/2}$	正态检验	$\dfrac{\lvert p_s-p_0\rvert}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}>u_{\alpha/2}$
					$\dfrac{p_s-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}>u_{\alpha/2}$
					$\dfrac{p_s-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}<-u_{\alpha/2}$
	p_1-p_2	--	$\dfrac{\lvert p_{s1}-p_{s2}\rvert}{\sqrt{\dfrac{p_{s1}(1-p_{s1})}{n_1}+\dfrac{p_{s2}(1-p_{s2})}{n_2}}}<u_{\alpha/2}$	正态检验	$\dfrac{\lvert p_{s1}-p_{s2}\rvert}{\sqrt{\dfrac{p_{s1}(1-p_{s1})}{n_1}+\dfrac{p_{s2}(1-p_{s2})}{n_2}}}<u_{\alpha/2}$
					$\dfrac{p_{s1}-p_{s2}}{\sqrt{\dfrac{p_{s1}(1-p_{s1})}{n_1}+\dfrac{p_{s2}(1-p_{s2})}{n_2}}}>u_{\alpha/2}$
					$\dfrac{p_{s1}-p_{s2}}{\sqrt{\dfrac{p_{s1}(1-p_{s1})}{n_1}+\dfrac{p_{s2}(1-p_{s2})}{n_2}}}<-u_{\alpha/2}$
最小二乘	β_0	σ^2 未知	$\dfrac{\lvert\hat{\beta}_0-\beta_0\rvert}{\hat{\sigma}\sqrt{\dfrac{1}{n}+\dfrac{\bar{x}^2}{L_{xx}}}}<t_{\alpha/2}(n-2)$	F 检验	$\dfrac{\hat{\beta}_0^2}{\hat{\sigma}^2}\dfrac{nL_{xx}}{\sum_{i=1}^{n}x_i^2}\sim F(1,n-2)$

<table>
<tr><td rowspan="3">最小二乘</td><td>β_1</td><td rowspan="3">σ^2 未知</td><td>$\frac{|\hat{\beta}_1-\beta_1|}{\hat{\sigma}\sqrt{\frac{1}{L_{xx}}}}<t_{\alpha/2}(n-2)$</td><td>F 检验</td><td>$\frac{\hat{\beta}_1^2}{\hat{\sigma}^2}L_{xx}\sim F(1,n-2)$</td></tr>
<tr><td>$\beta_0-\beta_1$</td><td>$\frac{|(\hat{\beta}_0-\hat{\beta}_1)-(\beta_0-\beta_1)|}{\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(\bar{x}+1)^2}{L_{xx}}}}<t_{\alpha/2}(n-2)$</td><td>--</td><td>--</td></tr>
<tr><td>y</td><td>$\frac{|y-y^*|}{\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}}<t_{\alpha/2}(n-2)$
（预测区间）</td><td>--</td><td>--</td></tr>
<tr><td>备注</td><td colspan="5">1. 符号说明：$S_W=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1+n_2-2)}}$
2. 混合总体的区间估计和假设检验问题，需要分别构造两个总体的 χ^2 检验（枢轴量），二者加和构造新的 χ^2 检验（枢轴量）
3. 表格中拒绝域给出的顺序，按如下假设形式的顺序给出：
$$H_0:a=a_0\leftrightarrow H_1:a\neq a_0\,|\,H_0:a\le a_0\leftrightarrow H_1:a>a_0\,|\,H_0:a\ge a_0\leftrightarrow H_1:a<a_0(a\text{为待检验参数})$$
4. 特殊假设形式的判断： $H_0:a=a_0\leftrightarrow H_1:a=a_1$
当 $a_1<a_0$ 时，将其视为 $H_0:a\ge a_0\leftrightarrow H_1:a<a_0$ 的一种特殊情况；
当 $a_1>a_0$ 时，将其视为 $H_0:a\le a_0\leftrightarrow H_1:a>a_0$ 的一种特殊情况。
5. 最小二乘的假设检验统一使用 F 统计量进行检验，F 统计量的形式为：$F_i=\frac{\hat{\beta}_i^2}{c_{ii}\hat{\sigma}^2}\sim F(1,n-2)$</td></tr>
</table>

表 3：区间估计与假设检验构造过程对比表

（注：以σ^2未知的正态总体为例）

【区间估计】		【假设检验】	
$\hat{\mu}=\bar{X}$	1. 找到 μ 的估计量 $\bar{X}$ ；	$\hat{\mu}=\bar{X}$	1. 找到 μ 的估计量 $\bar{X}$ ；
$\lvert \bar{X}-\mu \rvert < C$	2. 让 $\bar{X}$ 和 μ 做差得到精确解和近似解的距离，并让距离控制在 C 以下；	$\lvert \bar{X}-\mu \rvert > C$	2. 让 $\bar{X}$ 和 μ 做差得到精确解和近似解的距离,并让距离超过某个临界常数 C；
$P\{\lvert \bar{X}-\mu \rvert < C\}=1-\alpha$	3. 让距离在 C 以下的概率达到置信度；	$\beta_I=P\{\lvert \bar{X}-\mu \rvert > C\}=\alpha$	3. 利用功效函数,让反第一类错误的概率达到显著性水平；
$P\{\frac{\sqrt{n}\lvert \bar{X}-\mu \rvert}{S}<\frac{\sqrt{n}C}{S}\}=1-\alpha$	4. 找到一个合适的分布,对原概率形式做变换；	$P\{\frac{\sqrt{n}\lvert \bar{X}-\mu \rvert}{S}>\frac{\sqrt{n}C}{S}\}=\alpha$	4. 找到一个合适的分布,对原概率形式做变换；
$\frac{\sqrt{n}C}{S}=t_{\alpha/2}(n-1)$	5. 令距离的上界为找到的合适分布的分位点；	$\frac{\sqrt{n}C}{S}=t_{\alpha/2}(n-1)$	5. 令距离的上界为找到的合适分布的分位点；
$C=\frac{S\cdot t_{\alpha/2}(n-1)}{\sqrt{n}}$	6. 解出常数 C；	$C=\frac{S\cdot t_{\alpha/2}(n-1)}{\sqrt{n}}$	6. 解出常数 C；
$\bar{X}-\frac{S\cdot t_{\alpha/2}(n-1)}{\sqrt{n}}<\mu<\bar{X}+\frac{S\cdot t_{\alpha/2}(n-1)}{\sqrt{n}}$	7. 化简整理得到置信区间。	$W=\{\mu<\bar{X}-\frac{S\cdot t_{\alpha/2}(n-1)}{\sqrt{n}} or \mu>\bar{X}+\frac{S\cdot t_{\alpha/2}(n-1)}{\sqrt{n}}\}$	7. 化简整理得到拒绝域。

应用数理统计解题模板

题型一：抽样分布【题解】

例：设 $X_1, X_2, \ldots, X_n$ 为来自总体 X 的简单随机样本， $X \sim N(\mu, \sigma^2)$，

1、求出 $\bar{X}$ 与 $X_1 - \bar{X}$ 的联合概率密度函数。

2、若 $\mu = 0, n = 26$，满足 $P(\bar{X}^2 < c(X_1 - \bar{X})^2) = 0.90$，常数 c 的值是多少。

1.解：因为【 $X \sim N(\mu, \sigma^2)$ 】，所以【 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n}), X_1 - \bar{X} \sim N(0, \frac{n-1}{n}\sigma^2)$ 】。

所以，【 $(\bar{X}, X_1 - \bar{X}) = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \end{pmatrix} X$ 】

令【 $A = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \end{pmatrix}$ 】，于是，【 $\vec{\mu} = A\vec{\mu}_0 = \begin{pmatrix} \mu \\ 0 \end{pmatrix}, \Sigma = A\Sigma_0 A^T = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{n-1}{n}\sigma^2 \end{pmatrix}$ 】

由 Σ 可知，【 $\bar{X}, X_1 - \bar{X}$ 】独立。所以，

【联合密度函数为： $f(x,y) = f_x(x) f_y(y) = \frac{1}{\sqrt{2\pi}\sqrt{\frac{1}{n}}\sigma} e^{-\frac{(x-\mu)^2}{2\times\frac{\sigma^2}{n}}} \frac{1}{\sqrt{2\pi}\sqrt{\frac{n-1}{n}}\sigma} e^{-\frac{y^2}{2\times\frac{n-1}{n}\sigma^2}}$ 】。

2.解：因为【 $X \sim N(\mu, \sigma^2)$ 】，所以【 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n}), X_1 - \bar{X} \sim N(\mu, \frac{n-1}{n}\sigma^2)$ 】。

所以【 $\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1), \frac{X_1 - \bar{X}}{\sqrt{\frac{n-1}{n}\sigma^2}} \sim N(0,1)$ 】。

当 $\mu = 0, n = 26$ 时 $\frac{\sqrt{26}\bar{X}}{\sigma} \sim N(0,1), \frac{\sqrt{26}(X_1 - \bar{X})}{5\sigma} \sim N(0,1)$，所以 $\frac{26\bar{X}^2}{26(X_1 - \bar{X})^2/5} \sim F(1,1)$。

所以，【 $P(\bar{X}^2 < c(X_1 - \bar{X})^2) = P(\frac{26\bar{X}^2}{26(X_1 - \bar{X})^2/5} < 5c) = 0.90$ 】。

查表可得， $F_{0.1}(1,1) = 39.9 = 5c$，解得 $c = 7.98$

题型一：抽样分布【模板】

【独立性问题】

解：（利用线性变换处理独立性问题）

因为【样本服从分布】，所以【各随机变量服从分布】。

所以，【随机变量拼接为原总体的线性变换】

令 $A =$【线性变换矩阵】

于是，【 $\vec{\mu} = A\vec{\mu}_0, \Sigma = A\Sigma_0 A^T$ 】

由 Σ 可知，【两个随机变量】独立。

所以，【结论】。

【抽样分布定理与分布构造问题】

解：因为【样本服从分布】，所以【各随机变量服从分布】。

所以【利用随机变量间关系构造合适的分布】。

所以，【结论】。

应用数理统计解题模板

题型二：Bayes 估计【题解】

二、（10 分）设简单随机样本 $\boldsymbol{X_1, X_2, \cdots, X_n}$ 来自两点分布 $\boldsymbol{B(1,\theta)}$，参数 $\boldsymbol{\theta(0<\theta<1)}$ 的先验分布为均匀分布 $\boldsymbol{U(0,1)}$，在损失函数 $\boldsymbol{L(\theta,d)=\dfrac{(\theta-d)^2}{\theta(1-\theta)}}$ 下，求 $\boldsymbol{\theta}$ 的 Bayes 估计 $\boldsymbol{\hat{\theta}_B}$.

解：由题可知，损失函数为【 $L(\theta,d)=\dfrac{(\theta-d)^2}{\theta(1-\theta)}$ 】，所以风险函数为【 $R(\theta,d)=E(L(\theta,d))$ 】。

所以，【 $B(d)=E(L(\theta,d))=E[\dfrac{(\theta-d)^2}{\theta(1-\theta)}|X_1,...,X_n)]$, 记 $Y=\{X_1,...,X_n\}$ 】

则【 $B(d)=E[\dfrac{\theta^2-2\theta d(Y)+d^2(Y)}{\theta(1-\theta)}|Y]=E[\dfrac{\theta}{1-\theta}|Y]-2d(Y)E[\dfrac{1}{1-\theta}|Y]+d^2(Y)E[\dfrac{1}{\theta(1-\theta)}|Y]$ 】

对 $d(Y)$ 求导得【 $B'(d)=-2E[\dfrac{1}{1-\theta}|Y]+2d(Y)E[\dfrac{1}{\theta(1-\theta)}|Y]$ 】.

令 $B'(d)=0$ 得【 $d(\eta)=E[\dfrac{1}{1-\theta}|Y]/E[\dfrac{1}{\theta(1-\theta)}|Y]$ 】

因为【 $B''(d)=2E[\dfrac{1}{\theta(1-\theta)}|Y]>0$ 】，所以【 $d(\eta)=E[\dfrac{1}{1-\theta}|Y]/E[\dfrac{1}{\theta(1-\theta)}|Y]$ 】为极小值点。即【 $\tilde{d}(X_1,...,X_n)=E[\dfrac{1}{1-\theta}|X_1,...,X_n]/E[\dfrac{1}{\theta(1-\theta)}|X_1,...,X_n]$ 】

因为参数 θ 的先验分布为 $\theta\sim U(0,1)$，样本 $X\sim B(1,\theta)$ 。

所以，参数的后验分布为：

$$h(\theta|X_1,...,X_n)=\frac{\pi(\theta)f(X_1,...,X_n|\theta)}{\int_0^1\pi(\theta)f(X_1,...,X_n|\theta)d\theta}\propto\pi(\theta)f(X_1,...,X_n|\theta)$$

所以，$h(\theta|X_1,...,X_n)\propto\theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}}$，所以，$\theta|X_1,...,X_n\sim Beta(n\bar{x}+1,n-n\bar{x}+1)$

所以，参数的贝叶斯估计为 $\hat{\theta}=\dfrac{E[\dfrac{1}{1-\theta}|X_1,...,X_n]}{E[\dfrac{1}{\theta(1-\theta)}|X_1,...,X_n]}=\bar{X}$

题型二：Bayes 估计【模板】

解：

（步骤 1：确定贝叶斯估计形式）

由题可知，损失函数为【 $L(\theta,d)$ 】，所以风险函数为 $R(\theta,d)=E(L(\theta,d))$ 。

所以，【 $B(d)=E(L(\theta,d))\}$ 】

则【 $B(d)$ 展开】

对 $d(Y)$ 求导得【 $B'(d)$ 】.令 $B'(d)=0$ 得【 $d(\eta)$ 】

因为【 $B''(d)>0$ 】，所以【 $d(\eta)$ 】为极小值点。即【 $\tilde{d}(X_1,...,X_n)$ 】

(步骤 2：根据参数的先验分布和样本，推出后验分布，求解参数估计值)

因为【参数】的先验分布为【参数的先验分布】，样本【样本服从分布】。

所以，参数的后验分布为：

$$h(\theta|X_1,...,X_n)=\frac{\pi(\theta)f(X_1,...,X_n|\theta)}{\int_0^1\pi(\theta)f(X_1,...,X_n|\theta)d\theta}\propto\pi(\theta)f(X_1,...,X_n|\theta)$$

所以，【 $h(\theta|X_1,...,X_n)\propto$ 核】，所以，【 $\theta|X_1,...,X_n\sim$ 后验分布】

所以，参数的贝叶斯估计为 $\hat{\theta}=$【结果】。

应用数理统计解题模板

题型三：点估计【题解】

例：假设总体 $\boldsymbol{X}\sim\boldsymbol{U}(\theta,3\theta)$， $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n$ 为来自总体一个的简单随机样本，其中 $\theta>0$，是未知参数。

1、求 θ 的矩估计 $\hat{\theta}_1$，并说明 $\hat{\theta}_1$ 是否是 θ 的无偏估计；

2、求 θ 的极大似然估计 $\hat{\theta}_2$

解：1. 因为【 $\boldsymbol{X}\sim\boldsymbol{U}(\theta,3\theta)$ 】，所以【 $E(X)=\int_{\theta}^{3\theta}x\dfrac{1}{2\theta}dx=\theta=\bar{X}$ 】，即【 $\hat{\theta}_1=\bar{X}$ 】。

$$
\begin{aligned}
&E(\hat{\theta}_1)=E(\bar{X})=E(\frac{1}{n}\sum_{i=1}^{n}X_i)\\
&=\frac{1}{n}E(\sum_{i=1}^{n}X_i)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)\\
&=\frac{1}{n}\sum_{i=1}^{n}E(X)=\frac{1}{n}\cdot nE(X)=E(X)\\
&=\theta
\end{aligned}
$$

，所以 $\hat{\theta}_1$ 是 θ 的无偏估计

2. 构造似然函数为： $L(\theta)=(\dfrac{1}{2\theta})^n$

令 $g(\theta)=\ln L(\theta)=-n\ln 2\theta$，对 $g(\theta)$ 求导数得： $g'(\theta)=-\dfrac{n}{\theta}$

因为 $g'(\theta)=-\dfrac{n}{\theta}\le 0$ 恒成立，所以 $L(\theta)$ 关于 θ 单调递减。

因为 $\theta<x<3\theta$，即 $\dfrac{x}{3}<\theta<x$。

所以当 $\theta=\dfrac{X_{(n)}}{3}$ 时 $L(\theta)$ 最大，即 $\hat{\theta}_2=\dfrac{X_{(n)}}{3}$。

题型三：点估计【模板】

【矩估计】

解：因为【总体服从分布】，所以【 $E(X)=$（取值乘概率求和）】$=\bar{X}$，即【结果】。

（当一阶原点矩失效或估计多个参数时）

【 $E(X^2)=$（取值乘概率求和）】$=\dfrac{1}{n}\sum_{i=1}^{n}X_i^2$

（或）【 $E[(X-E(X))^2]=$（二阶中心矩）】$=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

注：常用样本方差代替样本的二阶中心矩。

【极大似然估计】

解：构造似然函数为： $L(\theta)=f(x_1,x_2,...,x_n\,|\,\theta)$

令 $g(\theta)=\ln L(\theta)=\ln f(x_1,x_2,...,x_n\,|\,\theta)$，对 $g(\theta)$ 求导数得： $g'(\theta)=(\ln L(\theta))'$

（当 $g'(\theta)\ge 0$或$g'(\theta)\le 0$ 恒成立时）

根据 θ 的取值范围，确定使得 $L(\theta)$ 取得最大值的 θ。

令 $g'(\theta)=0$，解得 $\hat{\theta}=$【结果】。

【无偏估计】

解： $E(\hat{\theta})=$【推导和化简过程】$=\theta$，所以 $\hat{\theta}$ 是无偏估计。

（若不等于 θ 则不是无偏估计）

应用数理统计解题模板

题型四：区间估计【题解】

五、（10 分）设 $X_1,X_2,\cdots,X_{10}$ 和 $Y_1,Y_2,\cdots,Y_{10}$ 分别来总体 $X\sim E(\lambda)$ 和总体 $Y\sim E(2\lambda)$ 的两组独立简单随机样本，总体 X 和总体 Y 的密度函数分别为：

$$f_X(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & \text{其他}\end{cases},\quad f_Y(y)=\begin{cases}2\lambda e^{-2\lambda y}, & y>0\\ 0, & \text{其他}\end{cases}.$$

其中 $\lambda>0$ 为未知参数，并已知 $\bar{X}=0.9$ 和 $\bar{Y}=0.8$，求 λ 的置信水平为 0.95 的双侧置信区间.

解：因为【 $X\sim E(\lambda),Y\sim E(2\lambda)$ 】。

【对于 X 总体有 $\hat{\lambda}=\dfrac{1}{\bar{X}}$，对于 Y 总体有 $(2\hat{\lambda})=\dfrac{1}{\bar{Y}}$ 】。

令【 $C_1<\hat{\lambda}\bar{X}<C_2$ 】，令【 $C_3<2\hat{\lambda}\bar{Y}<C_4$ 】。

则有【 $2n\lambda\bar{X}\sim\chi^2(2n)$ 】，【 $2n\cdot 2\lambda\bar{Y}\sim\chi^2(2n)$ 】。

所以有：【 $2n\lambda\bar{X}+2n\cdot 2\lambda\bar{Y}\sim\chi^2(4n)$ 】。

所以，$P\{$【 $D_1<2n\lambda\bar{X}+4n\lambda\bar{Y}<D_2$ 】$\}=1-\alpha$

解得，置信区间为【 $\chi^2_{1-\alpha/2}(4n)<2n\lambda\bar{X}+4n\lambda\bar{Y}<\chi^2_{\alpha/2}(4n)$ 】，

即【 $\dfrac{\chi^2_{1-\alpha/2}(4n)}{2n\bar{X}+4n\bar{Y}}<\lambda<\dfrac{\chi^2_{\alpha/2}(4n)}{2n\bar{X}+4n\bar{Y}}$ 】。

题型四：区间估计【模板】

（非混合总体参考《表三：区间估计与假设检验构造过程对比表》的解题思路求解）

解：因为【总体服从分布】，【条件】。

【参数的点估计=（点估计）】。

（分别将两个总体构造成 χ^2 分布）

令【第一个总体参数与精确值间差距满足条件】，

令【第二个总体参数与精确值间差距满足条件】。

则有【第一个总体参数变换服从 χ^2 分布】，【第二个总体参数变换服从 χ^2 分布】。

所以有：【两个总体混合后服从 χ^2 分布】。

所以，$P\{$【混合总体的表达式】$\}=1-\alpha$

解得，置信区间为【置信区间】，即【结论】。

应用数理统计解题模板

题型五：带参数假设检验【题解】

例：设 $\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n$ 是来自总体 $\boldsymbol{N}(\boldsymbol{\mu}_1,\boldsymbol{\sigma}^2)$ 的简单随机样本，$\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots\boldsymbol{Y}_m$ 是来自总体 $\boldsymbol{N}(\boldsymbol{\mu}_2,\boldsymbol{\sigma}^2)$ 的简单随机样本，且总体 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 相互独立，$\boldsymbol{\mu}_1,\boldsymbol{\mu}_2$ 未知，对于

$$\boldsymbol{H}_0:\boldsymbol{\sigma}^2\geq\boldsymbol{\sigma}_0^2 \quad\Leftrightarrow\quad \boldsymbol{H}_1:\boldsymbol{\sigma}^2<\boldsymbol{\sigma}_0^2$$

利用全部样本在检验水平 $\boldsymbol{\alpha}$ 下构造原假设的拒绝域.

解：因为【$\boldsymbol{X}\sim\boldsymbol{N}(\boldsymbol{\mu}_1,\boldsymbol{\sigma}^2),Y\sim\boldsymbol{N}(\boldsymbol{\mu}_2,\boldsymbol{\sigma}^2)$】，【$\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 相互独立，$\boldsymbol{\mu}_1,\boldsymbol{\mu}_2$ 未知】。

【对于 X 总体有 $\hat{\sigma}^2=S_1^2$，对于 Y 总体有 $\hat{\sigma}^2=S_2^2$】。

（分别将两个总体构造成 χ^2 分布）

令【$\dfrac{S_1^2}{\hat{\sigma}^2}<C_1$】，令【$\dfrac{S_2^2}{\hat{\sigma}^2}<C_2$】。

则有【$\dfrac{(n-1)S_1^2}{\sigma^2}\sim\chi^2(n-1)$】，【$\dfrac{(m-1)S_2^2}{\sigma^2}\sim\chi^2(m-1)$】。

所以有：【$\dfrac{(n-1)S_1^2}{\sigma^2}+\dfrac{(m-1)S_2^2}{\sigma^2}\sim\chi^2(n+m-2)$】。

所以，$\beta_I=P\{$【$\dfrac{(n-1)S_1^2}{\sigma^2}+\dfrac{(m-1)S_2^2}{\sigma^2}<D$】$\}=\alpha$

解得，拒绝域为 $W=\{$【$\dfrac{(n-1)S_1^2}{\sigma^2}+\dfrac{(m-1)S_2^2}{\sigma^2}<\chi^2_{1-\alpha}(m+n-2)$】$\}$，

即 $W=\{$【$\dfrac{(n-1)S_1^2+(m-1)S_2^2}{\chi^2_{1-\alpha}(m+n-2)}<\sigma^2$】$\}$

题型五：带参数假设检验【模板】

（非混合总体参考《表三：区间估计与假设检验构造过程对比表》的解题思路求解）

解：因为【总体服从分布】，【条件】。

【参数的点估计=（点估计）】。

（分别将两个总体构造成 χ^2 分布）

令【第一个总体参数与精确值间差距满足条件】，
令【第二个总体参数与精确值间差距满足条件】。

则有【第一个总体参数变换服从 χ^2 分布】，【第二个总体参数变换服从 χ^2 分布】。

所以有：【两个总体混合后服从 χ^2 分布】。

所以，$\beta_I=P\{$【混合总体的表达式】$\}=\alpha$

解得，拒绝域为 $W=\{$【拒绝域】$\}$，即【结论】。

应用数理统计解题模板

题型六：非参数假设检验——皮尔逊拟合优度检验【题解】

例：在某公路上某处 50 分钟之内，记录每 15 秒路过汽车的辆数，得到数据如下：

辆数	0	1	2	3	4	⩾5
频数	92	68	28	11	1	0

试问这个分布能否看作为泊松分布（$\alpha = 0.05$）？

解：将正半轴分为【[0,1)，[1,2)，[2,3)，[3,4)，[4,5)，[5,+∞)】，【6】个区间，则有如下假设：

$$H_0 : F(x) = F_0(x) \leftrightarrow H_1 : F(x) \neq F_0(x)$$

其中，$F_0(x)$ 为【泊松分布】。

由极大似然估计可得：

【$\hat{\lambda} = \bar{X} = \frac{1}{n}\sum_{i=1}^{k} v_i n_i = \frac{1}{200} \times (0\times 92 + 1\times 68 + 2\times 28 + 3\times 11 + 4\times 1) = 0.805$】

于是得下表：

【样本】i	0	1	2	3	4	⩾5
频数 v_i	92	68	28	11	1	0
理论频率 p_i	0.447	0.3599	0.1449	0.0389	0.0078	0.0015
理论频数 np_i	89.418	71.981	28.972	7.774	1.565	0.3
$v_i^2 / n_i p_i$	94.657	64.239	27.061	15.565	0.6390	0

由皮尔逊定理得：$\chi^2(4) =$【$\sum_{i=1}^{k} v_i^2 / n_i p_i - n = 2.161$】

查表可知：$\chi^2_{0.05}(4) = 9.448$

因为【$\chi^2(4) < \chi^2_{0.05}(4)$】，所以【接受】$H_0$，即【这个分布能看作泊松分布】。

题型六：非参数假设检验——皮尔逊拟合优度检验【模板】

解：将正半轴分为【区间组】，【区间个数】个区间，则有如下假设：

$$H_0 : F(x) = F_0(x) \leftrightarrow H_1 : F(x) \neq F_0(x)$$

其中，$F_0(x)$ 为【分布类型】。

（当分布类型中存在未知参数时）

由极大似然估计可得：【参数=参数估计量=参数估计值】

于是得下表：

【样本】i	
频数 v_i	
理论频率 p_i	
理论频数 np_i	
$v_i^2 / n_i p_i$	

由皮尔逊定理得：$\chi^2(k-r-1) =$【表格最后一行求和$-n$】

查表可知：$\chi^2_{\alpha}(k-r-1)$

因为【比较两值的大小关系】，所以【拒绝/接受】H_0，即【题干】。

【注释】

1. 理论频率需要通过待检验分布去计算；
2. 理论频数=理论频率×样本容量；
3. K 表示区间个数，r 表示未知参数个数。

题型七：非参数假设检验——卡方独立性检验【题解】

例：为了研究慢性气管炎与吸烟量的关系，调查了 385 人，统计数字如下表所示。

	a 支/日	b 支/日	c 支/日	和
患病人数	26	147	37	210
健康者	30	123	22	175
和	56	270	59	385

试问慢性气管炎与吸烟量是否有关系（$\alpha=0.05$）？

解：此题可视为假设检验问题：$H_0: p_{ij}=p_{i\cdot}\times p_{\cdot j} \leftrightarrow H_1: p_{ij}\neq p_{i\cdot}\times p_{\cdot j}$

$$\chi^2=n\sum_{i=1}^{r}\sum_{j=1}^{s}[n_{ij}-\frac{n_{i\cdot}n_{\cdot j}}{n}]^2/n_{i\cdot}n_{\cdot j}=【$$

$$385\times[\frac{(26-210\times56\div385)^2}{210\times56}+\frac{(30-56\times175\div385)^2}{56\times175}+\frac{(147-270\times210\div385)^2}{270\times210}+\frac{(123-270\times175\div385)^2}{270\times175}+\frac{(37-59\times210\div385)^2}{59\times210}+\frac{(22-59\times175\div385)^2}{59\times175}]$$

$$】=【0.676+0.812+0.001+0.001+0.721+0.866=3.077】$$

查表可知：$\chi^2_{0.05}(2)=5.991$

因为【$\chi^2<\chi^2_{0.05}(2)$】，所以【接受】H_0，即【慢性气管炎与吸烟量无关系】。

题型七：非参数假设检验——卡方独立性检验【模板】

解：此题可视为假设检验问题：$H_0: p_{ij}=p_{i\cdot}\times p_{\cdot j} \leftrightarrow H_1: p_{ij}\neq p_{i\cdot}\times p_{\cdot j}$

$\chi^2=n\sum_{i=1}^{r}\sum_{j=1}^{s}[n_{ij}-\frac{n_{i\cdot}n_{\cdot j}}{n}]^2/n_{i\cdot}n_{\cdot j}=$【计算过程表达式】=【计算值】

查表可知：$\chi^2_{\alpha}((r-1)(s-1))$

因为【比较两个值的大小】，所以【拒绝/接受】H_0，即【题干】。

【注释】

1. r 和 s 为两个因素的水平指标个数；
2. 四格表检验也可直接带公式计算。

题型八：非参数假设检验——秩和检验【题解】

例：一种产品可以用 A 、B 两种材料制造，随机抽取了 12 个产品按照性能从低到高排序：

B，B，A，B，B，A，A，B，A，A，A，A

在 0.05 水平下检验这两种材料产品性能是否有差异。（$H_0:\mu_1=\mu_2 \leftrightarrow H_1:\mu_1\neq\mu_2$）

解：【产品 B】的秩分别为：【1，2，4，5，8】。

令$W=$【1+2+4+5+8=20】。

查表可知，在【0.05】下的拒绝域为：$\{W\le 22或W\ge 43\}$

所以，【拒绝】H_0，即【两种产品的性能有差异】。

题型八：非参数假设检验——秩和检验【模板】

解：【第二类数据】的秩分别为：【第二类数据的秩】。

令$W=$【第二类数据的秩求和】。

查表可知，在【检验水平】下的拒绝域为：{【拒绝域】}

所以，【接受/拒绝】H_0，即【题干】。

题型九：单因素方差分析【题解】

例：试判断在检验水平为 0.05 的条件下，是否可以认为灯丝配料对灯泡寿命有显著影响？

灯丝	使用寿命（小时）							
甲	1600	1610	1650	1680	1700	1720	1800	
乙	1580	1640	1640	1700	1750			
丙	1460	1550	1600	1640	1660	1740	1820	1820
丁	1510	1520	1530	1570	1600	1680		

解：此题可看作如下假设检验问题：

H_0:【灯丝配料对灯泡寿命】无显著影响 $\leftrightarrow H_1$:【灯丝配料对灯泡寿命】有显著影响

$n =$【26】，$r =$【4】

$$TSS = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \text{【227250】}$$

$$CSS = \sum_{i=1}^{r} n_i(\bar{y}_i-\bar{y})^2 = \text{【47399.17】}$$

$$RSS = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 = \text{【179850.8】}$$

可列单因素方差分析表如下：

方差来源	平方和	自由度	均方	F-比
分类变量	【47399.17】	【3】	【15799.7】	【1.9327】
残差变量	【179850.8】	【22】	【8175.04】	
总计	【227250】	【25】		

因为【 $F_{0.05}(3,22)=3.05 > F$比 】，所以【可以认为灯丝配料对灯泡寿命】【无】显著影响。

题型九：单因素方差分析【模板】

解：此题可看作如下假设检验问题：

H_0:【题干】无显著影响 $\leftrightarrow H_1$:【题干】有显著影响

$n =$【数据个数】，$r =$【水平个数】

$$TSS = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \text{【总平方和】}$$

$$CSS = \sum_{i=1}^{r} n_i(\bar{y}_i-\bar{y})^2 = \text{【自变量平差和】}$$

$$RSS = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 = \text{【残差平方和】}$$

可列单因素方差分析表如下：

方差来源	平方和	自由度	均方	F-比
分类变量	【 CSS 】	【 $r-1$ 】	【 CMS 】	【 F比 】
残差变量	【 RSS 】	【 $n-r$ 】	【 RMS 】	
总计	【 TSS 】	【 $n-1$ 】		

因为【 F比 和 $F_\alpha(r-1,n-r)$ 比较】，所以【题干】【有/无】显著影响。

【注释】

1.对 TSS 和 RSS 的计算，可以采用样本方差公式计算，可提升计算速度；

2.当出现舍入误差的情况时，采用优先计算 RSS 和 CSS 的方式规避误差；

题型十：最小二乘法与回归分析【题解】

例：设 $y=\beta_0+\beta_1 x+\varepsilon, \varepsilon \sim N(0,\sigma^2)$. 已知变量 x 和 y 的 9 对独立观测数据如下：

x	-4	-3	-2	-1	0	1	2	3	4
y	17	14	20	18	23	25	22	25	34

1、利用所给数据求 y 对 x 的一元线性回归方程；

2、检验假设 $H_0:\beta_1=0, H_1:\beta_1\neq 0$.（检验水平 $a=0.05$）

已知： $\bar{x}=0, \bar{y}=22, \sum_{t=1}^{n} x_t y_t=112, \sum_{t=1}^{n}(x_t-\bar{x})^2=60, \sum_{t=1}^{n}(y_t-\bar{y})^2=272.$

1.解：由最小二乘思想，有如下推导过程：

$$\|Y-X\beta\|^2=\|Y-X\hat{\beta}+X\hat{\beta}-X\beta\|^2$$

$$=\|Y-X\hat{\beta}\|^2+\|X(\hat{\beta}-\beta)\|^2+2(\hat{\beta}-\beta)^T X^T(Y-X\hat{\beta})$$

使 $\|Y-X\beta\|^2 \geq \|Y-X\hat{\beta}\|^2$ 成立的充要条件为 $(\hat{\beta}-\beta)^T X^T(Y-X\hat{\beta})=0$

得到正规方程 $(X^T X)\beta=X^T Y$

由 $\beta=(X^T X)^{-1}X^T Y$ 得【 $\beta_1=\frac{L_{xy}}{L_{xx}}=1.867, \beta_0=\bar{y}-\beta_1\bar{x}=22$ 】

所以，一元线性回归方程为【 $y=22+1.867x$ 】。

2.解：由题中条件可得 $\hat{\sigma}^2=\frac{1}{n-k-1}(L_{yy}-\frac{L_{xy}^2}{L_{xx}})=8.99$， $C=(X^T X)^{-1}=\begin{pmatrix} \frac{\sum_{i=1}^{n} x_i^2}{nL_{xx}} & -\frac{\bar{x}}{L_{xx}} \\ -\frac{\bar{x}}{L_{xx}} & \frac{1}{L_{xx}} \end{pmatrix}$

于是，构造 F 检验统计量 $F_1=\frac{\hat{\beta}_1^2}{c_{11}\hat{\sigma}^2} \sim F(1,7)$

解得【 $F_1=\frac{1.867^2}{\frac{1}{60}\times 8.99}=23.26$ 】，查表可知 $F_{0.05}(1,7)=5.59$ 。

因为【 $F_1>F_{0.05}(1,7)$ 】，所以【拒绝】 H_0 。

题型十：最小二乘法与回归分析【模板】

由最小二乘思想，有如下推导过程：

$$\|Y-X\beta\|^2=\|Y-X\hat{\beta}+X\hat{\beta}-X\beta\|^2$$

$$=\|Y-X\hat{\beta}\|^2+\|X(\hat{\beta}-\beta)\|^2+2(\hat{\beta}-\beta)^T X^T(Y-X\hat{\beta})$$

使 $\|Y-X\beta\|^2 \geq \|Y-X\hat{\beta}\|^2$ 成立的充要条件为 $(\hat{\beta}-\beta)^T X^T(Y-X\hat{\beta})=0$

得到正规方程 $(X^T X)\beta=X^T Y$

【线性回归方程问题】

解：由 $\beta=(X^T X)^{-1}X^T Y$ 得【 $\beta_1=\frac{L_{xy}}{L_{xx}}, \beta_0=\bar{y}-\beta_1\bar{x}$ 】

所以，一元线性回归方程为【一元线性回归方程】。

【假设检验问题】

解：由题中条件可得

$$C=(X^T X)^{-1}=\begin{pmatrix} \frac{\sum_{i=1}^{n} x_i^2}{nL_{xx}} & -\frac{\bar{x}}{L_{xx}} \\ -\frac{\bar{x}}{L_{xx}} & \frac{1}{L_{xx}} \end{pmatrix}$$

$$\hat{\sigma}^2=\frac{1}{n-k-1}(L_{yy}-\frac{L_{xy}^2}{L_{xx}})$$

于是，构造 F 检验统计量 $F_i=\frac{\hat{\beta}_i^2}{c_{ii}\hat{\sigma}^2} \sim F(1,n-2)$

解得【 F_i 】，查表可知 $F(1,n-2)=$【查表数据】。

因为【 F_i 与 $F(1,n-2)$ 比较】，所以【拒绝/接受】 H_0，所以【题干】。