数理统计复习概要

1 统计量与三大分布

1.1 统计量

1.1.1 定义

只依赖于样本的量称为统计量. 统计量不含任何未知参数.

1.1.2 常用统计量

样本均值：$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$；样本方差：$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$.

1.1.3 常用性质

$E\left(\bar{X}\right)=E\left(X\right)$；$D\left(\bar{X}\right)=\frac{1}{n}D\left(X\right)$；$E\left(S^2\right)=D\left(X\right)$；$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N\left(0,1\right)$；$\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t\left(n-1\right)$.

1.2 三大抽样分布

χ^2 分布：$X_1,X_2,\ldots,X_n\sim N\left(0,1\right)$ 且相互独立，则 $\sum_{i=1}^{n}X_i^2\sim\chi^2\left(n\right)$.

t 分布：$X\sim N\left(0,1\right),Y\sim\chi^2\left(n\right)$ 且相互独立，则 $\frac{X}{\sqrt{Y/n}}\sim t\left(n\right)$.

F 分布：$X\sim\chi^2\left(n_1\right),Y\sim\chi^2\left(n_2\right)$ 且相互独立，则 $\frac{X/n_1}{Y/n_2}\sim F\left(n_1,n_2\right)$.

并有：$F\left(n_1,n_2\right)=F^{-1}\left(n_2,n_1\right),F_{1-\alpha}\left(n_1,n_2\right)=F_{\alpha}^{-1}\left(n_2,n_1\right)$.

2 参数估计

2.1 矩估计

矩估计求解方法：

$$\mu_1 = E(X) = f(\theta),\ \theta = g(\mu_1),\ \hat{\theta} = g(\bar{X}).$$

例 设总体 X 具有分布律：

X	1	2	3
p	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

其中θ $(0<\theta<1)$ 是未知参数，已经取得了样本值 $x_1 = 1,\ x_2 = 2,\ x_3 = 1,\ x_4 = 3$，试求$\theta$的矩估计值.

解 由 $E(X) = \bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$，得：

$$\mu_1 = E(X) = \theta^2 + 2\times 2\theta(1-\theta) + 3(1-\theta)^2 = 3-2\theta，\ \theta = \frac{3-\mu_1}{2}，\ \hat{\theta} = \frac{3-\bar{X}}{2} = \frac{3-7/4}{2} = \frac{5}{8}$$ ■

2.2 最大似然估计

最大似然估计求解方法：

$$L(\theta) = \prod_{i=1}^{n} f(x_i,\theta),\ \frac{\mathrm{d}}{\mathrm{d}\theta}\left[L(\theta)\right] = 0,\ \hat{\theta} = ?.$$

例 设总体 X 具有分布律：

X	1	2	3
p	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

其中θ $(0<\theta<1)$ 是未知参数，已经取得了样本值 $x_1 = 1,\ x_2 = 2,\ x_3 = 1,\ x_4 = 3$，试求$\theta$的最大似然估计量.

解 θ的似然函数为：

$$L(\theta) = \prod_{i=1}^{4} p(x_i;\theta) = \theta^2 \times \left[2\theta(1-\theta)\right] \times \theta^2 \times (1-\theta)^2 = 2\theta^5(1-\theta)^3$$

取对数求导：

$$\frac{\mathrm{d}}{\mathrm{d}\theta}[\ln L(\theta)] = \frac{\mathrm{d}}{\mathrm{d}\theta}[\ln 2 + 5\ln\theta + 3\ln(1-\theta)] = \frac{5}{\theta} - \frac{3}{1-\theta} = 0,\ \hat{\theta} = \frac{5}{8}$$ ■

2.3 估计量的评选标准

2.3.1 均方误差

$MSE\left(\hat{\theta}\right)=E\left[\left(\hat{\theta}-\theta\right)^2\right]=D\left(\hat{\theta}\right)+\left[E\left(\hat{\theta}\right)-\theta\right]^2$. 其中$\left[E\left(\hat{\theta}\right)-\theta\right]$决定是否无偏，$D\left(\hat{\theta}\right)$影响有效性.

2.3.2 无偏估计量

若$E\left(\hat{\theta}\right)=\theta$，则$\hat{\theta}$为$\theta$的无偏估计量.

2.3.3 有效性

方差$D\left(\hat{\theta}\right)$较小的无偏估计量更优.

3 区间估计

置信水平为$1-\alpha$时，单正态总体均值、方差的置信区间与单侧置信限：

待估参数		枢轴量及其分布	置信区间
μ	σ^2已知	$\dfrac{\bar{X}-\mu}{\sigma}\sqrt{n}\sim N(0,1)$	$\left(\bar{X}\pm\dfrac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$
	σ^2未知	$\dfrac{\bar{X}-\mu}{S}\sqrt{n}\sim t(n-1)$	$\left(\bar{X}\pm\dfrac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right)$
σ^2		$\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$	$\left(\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$

求解步骤：

① 确定类型：待估参数及条件；

② 根据类型写出枢轴量及其分布；

③ 计算各待求量，代入置信区间公式计算.

例 设某机床加工的零件长度 $X\sim N(\mu,\sigma^2)$，今抽查 16 个零件，测得长度(单位: mm)为：

12.15, 12.12, 12.01, 12.08, 12.09, 12.16, 12.03, 12.01, 12.06, 12.13, 12.07, 12.11, 12.08, 12.01, 12.03, 12.06

试求σ^2的置信度为 95%的置信区间;

解 由条件可知，所求的是σ^2的置信区间，构造枢轴量：

$$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$$

由数据计算得：

$$\bar{X}=12.075,\ S^2=0.00244,\ S=0.049396$$

$1-\alpha=0.95,\ n=16,\ \chi^2_{\alpha/2}(n-1)=\chi^2_{0.025}(15)=27.488,\ \chi^2_{1-\alpha/2}(n-1)=\chi^2_{0.975}(15)=6.262$

则σ^2的置信水平为 0.95 的区间估计为：

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right]=\left[\frac{15\times0.00244}{27.488},\frac{15\times0.00244}{6.262}\right]=[0.0013,\ 0.0058]\qquad\blacksquare$$

4 假设检验

4.1 解题方法

显著性水平为α时，单正态总体均值、方差的检验法：

待检验参数		原假设	备择假设	检验统计量	拒绝域
μ	σ^2 已知	$H_0:\mu=\mu_0$ $H_0:\mu\geq\mu_0$ $H_0:\mu\leq\mu_0$	$H_1:\mu\neq\mu_0$ $H_1:\mu<\mu_0$ $H_1:\mu>\mu_0$	$Z=\frac{x-\mu_0}{\sigma}\sqrt{n}$	$\lvert z\rvert\geq z_{\alpha/2}$ $z\leq-z_\alpha$ $z\geq z_\alpha$
	σ^2 未知	$H_0:\mu=\mu_0$ $H_0:\mu\geq\mu_0$ $H_0:\mu\leq\mu_0$	$H_1:\mu\neq\mu_0$ $H_1:\mu<\mu_0$ $H_1:\mu>\mu_0$	$T=\frac{x-\mu_0}{S}\sqrt{n}$	$\lvert t\rvert\geq t_{\alpha/2}(n-1)$ $t\leq-t_\alpha(n-1)$ $t\geq t_\alpha(n-1)$
σ^2		$H_0:\sigma^2=\sigma_0^2$ $H_0:\sigma^2\geq\sigma_0^2$ $H_0:\sigma^2\leq\sigma_0^2$	$H_1:\sigma^2\neq\sigma_0^2$ $H_1:\sigma^2<\sigma_0^2$ $H_1:\sigma^2>\sigma_0^2$	$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}$	$\begin{cases}\chi^2\geq\chi^2_{\alpha/2}(n-1)\\\chi^2\leq\chi^2_{1-\alpha/2}(n-1)\end{cases}$ $\chi^2\leq\chi^2_{1-\alpha}(n-1)$ $\chi^2\geq\chi^2_\alpha(n-1)$

求解步骤：

① 提出原假设与备择假设；

② 构造检验统计量，明确其分布；

③ 确定临界值，即拒绝域；

④ 做出决策：接受原假设或拒绝原假设.

例 设某机床加工的零件长度 $X\sim N(\mu,\sigma^2)$，今抽查 16 个零件，测得长度(单位: mm)为：

12.15, 12.12, 12.01, 12.08, 12.09, 12.16, 12.03, 12.01, 12.06, 12.13, 12.07, 12.11, 12.08, 12.01, 12.03, 12.06

在 5%的显著性水平下，能否认为该机床加工的零件长度为 12.10mm.

解 本问题是方差未知的条件下，$\mu=12.10$的假设检验，故：

$$H_0\colon\mu=12.10,\ H_1\colon\mu\neq12.10,\ \alpha=0.05,\ n=16,\ \overline{X}=12.075,\ S=0.049396$$

统计量为：

$$T = \frac{\overline{X} - 12.1}{S}\sqrt{n} \sim t(n-1)$$

H_0的拒绝域为：

$$|T| = \left|\frac{\overline{X} - 12.1}{S/\sqrt{16}}\right| \geq t_{0.025}(15) = 2.1314$$

故：

$$|t| = \left|\frac{\overline{X} - 12.1}{S} \times \sqrt{16}\right| = \left|\frac{12.075 - 12.1}{0.049396} \times 4\right| = 2.02444 < 2.1314$$

所以接受H_0，即可以认为该机床加工的零件长度为 12.10mm. ■

4.2 两类错误

第一类错误：$P\left(拒绝H_0 \mid H_0为真\right) = P_{H_0}\left(拒绝H_0\right)$；

第二类错误：$P\left(接受H_0 \mid H_0为假\right) = P_{H_1}\left(接受H_0\right)$.