

X	$P\{X\}$	$E(X)$	$D(X)$
(0-1)分布	$P\{X = k\} = p^k(1-p)^{1-k}$	p	$p(1-p)$
二项分布 $X \sim b(n, p)$	$P\{X = k\} = \binom{n}{k} p^k(1-p)^{n-k}$	np	$np(1-p)$
泊松分布 $X \sim \pi(\lambda)$	$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ
均匀分布 $X \sim U(a, b)$	$f(x) = \frac{1}{b-a}, a < x < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布	$f(x) = \frac{1}{\theta} e^{-x/\theta}, x > 0$	θ	θ^2
正态分布 $X \sim N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty$	μ	σ^2
Γ 分布 $X \sim \Gamma(\alpha, \theta)$	$f(x) = \frac{1}{\theta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\theta}, x > 0$		

1. 设随机变量 X 具有概率密度 $f_X(x)$, $-\infty < x < +\infty$, 又设函数 $g(x)$ 处处可导且恒有 $g'(x) > 0$ (或恒有 $g'(x) < 0$), 则 $Y = g(X)$ 是连续型随机变量, 其概率密度为

$$f_Y(y) = f_X[h(y)]|h'(y)|, \alpha < y < \beta$$

其中 $\alpha = \min\{g(-\infty), g(+\infty)\}$, $\beta = \max\{g(-\infty), g(+\infty)\}$, $h(y)$ 是 $g(x)$ 的反函数。

2. 参数为 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ 的二维正态分布:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]}$$

3. 设二维随机变量 (X, Y) 的概率密度为 $f(x, y)$, (X, Y) 关于 Y 的边缘概率密度为 $f_Y(y)$ 。若对于固定的 y , $f_Y(y) > 0$, 则称 $f(x, y)/f_Y(y)$ 为在 $Y = y$ 的条件下 X 的条件概率密度, 记为

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

4. 有界区域 G (面积为 A) 上的二维均匀分布:

$$f(x, y) = \frac{1}{A}, (x, y) \in G$$

5. 设 (X, Y) 是二维连续型随机变量, 它具有概率密度 $f(x, y)$, 则 $Z = X + Y$ 仍为连续型随机变量, 其概率密度为

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f(z-y, y)dy = \int_{-\infty}^{+\infty} f(x, z-x)dx$$

若 X 和 Y 相互独立, 设 (X, Y) 关于 X, Y 的边缘密度分别为 $f_X(x), f_Y(y)$, 则

$f_{X+Y}(z)$ 满足卷积公式

$$f_X * f_Y = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)dy = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx$$

6. 设随机变量 X, Y 相互独立, 且分别服从参数为 $\alpha, \theta; \beta, \theta$ 的 Γ 分布 (分别记成 $X \sim \Gamma(\alpha, \theta), Y \sim \Gamma(\beta, \theta)$), 则 $Z = X + Y$ 服从参数为 $\alpha + \beta, \theta$ 的 Γ 分布, 即 $X + Y \sim \Gamma(\alpha + \beta, \theta)$ 。
7. 设 (X, Y) 是二维连续型随机变量, 它具有概率密度 $f(x, y)$, 则 $Z = Y/X, Z = XY$ 仍为连续型随机变量, 其概率密度分别为

$$f_{Y/X}(z) = \int_{-\infty}^{+\infty} |x|f(x, xz)dx$$

$$f_{XY}(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|}f(x, \frac{z}{x})dx$$

若 X 和 Y 相互独立, 设 (X, Y) 关于 X, Y 的边缘密度分别为 $f_X(x), f_Y(y)$, 则

$$f_{Y/X}(z) = \int_{-\infty}^{+\infty} |x|f_X(x)f_Y(xz)dx$$

$$f_{XY}(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|}f_X(x)f_Y(\frac{z}{x})dx$$

8. 设 X, Y 是两个相互独立的随机变量, 它们的分布函数分别为 $F_X(x)$ 和 $F_Y(y)$, 则 $M = \max\{X, Y\}$ 及 $N = \min\{X, Y\}$ 的分布函数分别为

$$F_{\max}(z) = F_X(z)F_Y(z)$$

$$F_{\min}(z) = 1 - [1 - F_X(z)][1 - F_Y(z)]$$

9. 设 Y 是随机变量 X 的函数: $Y = g(X)$ (g 是连续函数), 如果 X 是连续型随机变量, 它的概率密度为 $f(x)$, 若 $\int_{-\infty}^{+\infty} g(x)f(x)dx$ 绝对收敛, 则有

$$E(Y) = E(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

10. $D(X + Y) = D(X) + D(Y) - 2E\{[X - E(X)][Y - E(Y)]\} = D(X) + D(Y) - 2Cov(X, Y)$
11. $D(X) = 0$ 的充要条件是 X 以概率 1 取常数 $E(X)$, 即 $P\{X = E(X)\} = 1$ 。
12. 切比雪夫 (Chebyshev) 不等式: 设随机变量 X 具有数学期望 $E(X) = \mu$, 方差 $D(X) = \sigma^2$, 则对于任意正数 ε , 不等式

$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$$

成立。

13. 协方差 $Cov(X, Y) = E\{[X - E(X)][Y - E(Y)]\} = E(XY) - E(X)E(Y)$

$$\text{相关系数 } \rho_{XY} = Cov(X, Y) / [\sqrt{D(X)}\sqrt{D(Y)}]$$

14. 协方差性质:

$$\text{Cov}(X, Y) = \text{Cov}(Y, X), \quad \text{Cov}(X, X) = D(X)$$

$$\text{Cov}(aX, bY) = ab\text{Cov}(X, Y), \quad \text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$$

15. $|\rho_{XY}| = 1$ 的充要条件是, 存在常数 a, b , 使 $P\{Y = a + bX\} = 1$ 。

16. 矩: X 和 Y 是随机变量,

X 的 k 阶原点矩 (k 阶矩): $E(X^k)$

X 的 k 阶中心矩: $E\{[X - E(X)]^k\}$

X 和 Y 的 $k + l$ 阶混合矩: $E(X^k Y^l)$

X 和 Y 的 $k + l$ 阶混合中心矩: $E\{[X - E(X)]^k [Y - E(Y)]^l\}$

17. 协方差矩阵: n 维随机变量 (X_1, X_2, \dots, X_n) 的二阶混合中心矩 $\text{Cov}(X, Y)$ 组成的 $n \times n$ 矩阵 \mathbf{C} 。

18. n 维正态随机变量 (X_1, X_2, \dots, X_n) 的概率密度:

$$f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} (\det \mathbf{C})^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \mathbf{C}^{-1}(x-\mu)}$$

19. 大数定律的一般形式: 设 $\{X_n, n = 1, 2, \dots\}$ 是一个随机变量序列, 而且对每个 n , $E(X_n)$ 存在, 如果对于任意给定的 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k) \right| < \varepsilon \right\} = 1$$

则称随机变量序列 $\{X_n\}$ 服从大数定律。

20. Chebyshev 大数定律: $X_1, X_2, \dots, X_n, \dots$ 两两相互独立, $D(X_n)$ 存在且有界。

21. Chebyshev 大数定律的特殊情况: $X_1, X_2, \dots, X_n, \dots$ 独立同分布, $E(X_k) = \mu, D(X_k) = \sigma^2$ 。

22. 泊松大数定律: 如果在独立试验序列中, 事件 A 在第 n 次试验中出现概率为 p_n , 设 n_A 是前 n 次试验中事件 A 出现的次数, 则对任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{n_A}{n} - \frac{p_1 + p_2 + \dots + p_n}{n} \right| < \varepsilon \right\} = 1$$

23. Markov 大数定律: 不要求独立性, 只要求 Markov 条件:

$$\lim_{n \rightarrow \infty} D \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = 0$$

24. Khinchin 大数定律: $X_1, X_2, \dots, X_n, \dots$ 独立同分布, $E(X_k) = \mu$, 不要求 $D(X_n)$ 存在。

25. Bernoulli 大数定律: 设 f_A 是 n 次独立重复试验中事件 A 发生的次数, p 是事件 A 在每次试验中发生的概率, 则对于任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{f_A}{n} - p \right| < \varepsilon \right\} = 1$$

26. 中心极限定理：

设 $X_1, X_2, \dots, X_n, \dots$ 是独立的随机变量序列，若 $E(X_k), D(X_k) (k = 1, 2, \dots)$ 都存在，令

$$Y_n = \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n E(X_k)}{\sqrt{\sum_{k=1}^n D(X_k)}}$$

若对任意 $x \in \mathbb{R}$ ，有

$$\lim_{x \rightarrow \infty} P\{Y_n \leq x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \Phi(x)$$

则称该随机变量序列服从中心极限定理。

27. 中心极限定理相关的计算方法：

$$P\{a < \bar{X} < b\} = P\left\{\frac{a - \mu}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{b - \mu}{\sigma/\sqrt{n}}\right\} \approx \Phi\left(\frac{b - \mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{a - \mu}{\sigma/\sqrt{n}}\right)$$

28. Lyapunov 定理 ($P_{122-123}$)。

29. De Moivre-Laplace 定理 (P_{123})。

30. 若 X_1, X_2, \dots, X_n 为 F 的一个样本，则 (X_1, X_2, \dots, X_n) 的联合分布函数为

$$F^*(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F(x_i)$$

若 X 具有概率密度 f ，则 (X_1, X_2, \dots, X_n) 的联合概率密度为

$$f^*(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i)$$

31. 设有容量为 n 的样本观察值 x_1, x_2, \dots, x_n ，样本的 p 分位数($0 < p < 1$)记为 x_p ，它具有以下性质：至少有 np 个观察值小于或等于 x_p ；至少有 $n(1-p)$ 个观察值大于或等于 x_p 。

32. 第一四分位数 Q_1 与第三四分位数 Q_3 之间的距离 IQR 称为四分位数间距，若数据小于 $Q_1 - 1.5IQR$ 或大于 $Q_3 + 1.5IQR$ ，就认为它是疑似异常值，从箱线图中去除后得到修正箱线图。

33. 样本均值： $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

$$\text{样本方差：} S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} (\sum_{i=1}^n X_i^2 - n\bar{X}^2)$$

$$\text{样本标准差：} S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\text{样本}k\text{阶原点矩：} A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

$$\text{样本}k\text{阶中心矩：} B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$

34. 经验分布函数：设 X_1, X_2, \dots, X_n 是总体 F 的一个样本，用 $S(x)$ 表示 X_1, X_2, \dots, X_n 中不大于 x

的随机变量的个数, 定义经验分布函数 $F_n(x)$ 为

$$F_n(x) = \frac{1}{n}S(x), -\infty < x < +\infty$$

35. 对于任一实数 x , 当 $n \rightarrow \infty$ 时 $F_n(x)$ 以概率1一致收敛于分布函数 $F(x)$, 即

$$P\left\{\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} |F_n(x) - F(x)| = 0\right\} = 1$$

36. 设 X_1, X_2, \dots, X_n 是来自总体 $N(0,1)$ 的样本, 则称统计量

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

服从自由度为 n 的 χ^2 分布, 记为 $\chi^2 \sim \chi^2(n)$ 。

37. $\chi^2(n)$ 分布的概率密度:

$$\chi^2 = \sum_{i=1}^n X_i^2 \sim \Gamma\left(\frac{n}{2}, 2\right)$$

$$f(y) = \frac{1}{2^{n/2} \Gamma(n/2)} y^{n/2-1} e^{-y/2}, y > 0$$

38. χ^2 分布的性质:

可加性: 设 $\chi_1^2 \sim \chi^2(n_1), \chi_2^2 \sim \chi^2(n_2)$ 且相互独立, 则有 $\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$ 。

数学期望和方差: 若 $\chi^2 \sim \chi^2(n)$, 则有 $E(\chi^2) = n, D(\chi^2) = 2n$ 。

39. χ^2 分布的上分位点: 对于给定的正数 $\alpha(0 < \alpha < 1)$, 满足条件

$$P\{\chi^2 > \chi_{\alpha}^2(n)\} = \int_{\chi_{\alpha}^2(n)}^{\infty} f(y) dy = \alpha$$

的点 $\chi_{\alpha}^2(n)$ 就是 $\chi^2(n)$ 分布的上 α 分位点。 $n > 40$ 时近似地有

$$\chi_{\alpha}^2(n) \approx \frac{1}{2} (z_{\alpha} + \sqrt{2n-1})^2$$

40. 设 $X \sim N(0,1), Y \sim \chi^2(n)$, 且 X, Y 相互独立, 则称随机变量

$$t = \frac{X}{\sqrt{Y/n}}$$

服从自由度为 n 的 t 分布, 记为 $t \sim t(n)$ 。

41. $t(n)$ 分布的概率密度:

$$h(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n} \Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, -\infty < t < +\infty$$

$$\lim_{n \rightarrow \infty} h(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

42. t 分布的上分位点: 对于给定的正数 $\alpha(0 < \alpha < 1)$, 满足条件

$$P\{t > t_{\alpha}(n)\} = \int_{t_{\alpha}(n)}^{\infty} h(t) dt = \alpha$$

的点 $t_\alpha(n)$ 就是 $t(n)$ 分布的上 α 分位点。 $t_{1-\alpha}(n) = -t_\alpha(n)$, 且 $n > 45$ 时近似地有

$$t_\alpha(n) \approx z_\alpha$$

43. 设 $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$, 且 U, V 相互独立, 则称随机变量

$$F = \frac{U/n_1}{V/n_2}$$

服从自由度为 (n_1, n_2) 的 F 分布, 记为 $F \sim F(n_1, n_2)$ 。

44. $F(n_1, n_2)$ 分布的概率密度:

$$\psi(y) = \frac{\Gamma[(n_1 + n_2)/2](n_1/n_2)^{n_1/2} y^{(n_1/2)-1}}{\Gamma(n_1/2)\Gamma(n_2/2)[1 + (n_1 y/n_2)]^{(n_1+n_2)/2}}, y > 0$$

$$\frac{1}{F} \sim F(n_2, n_1)$$

45. F 分布的上分位点: 对于给定的正数 $\alpha (0 < \alpha < 1)$, 满足条件

$$P\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1, n_2)}^{\infty} \psi(y) dy = \alpha$$

的点 $F_\alpha(n_1, n_2)$ 就是 $F(n_1, n_2)$ 分布的上 α 分位点。 $F_{1-\alpha}(n_1, n_2) = 1/F_\alpha(n_2, n_1)$ 。

46. 设总体 X 的均值为 μ , 方差为 σ^2 , X_1, X_2, \dots, X_n 是来自 X 的一个样本, \bar{X}, S^2 分别是样本均值和样本方差, 则有

$$E(\bar{X}) = \mu, \quad D(\bar{X}) = \sigma^2/n, \quad E(S^2) = \sigma^2$$

47. 定理一: 设 X_1, X_2, \dots, X_n 是来自正态总体 $N(\mu, \sigma^2)$ 的样本, \bar{X} 是样本均值, 则有

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

48. 定理二: 设 X_1, X_2, \dots, X_n 是来自正态总体 $N(\mu, \sigma^2)$ 的样本, \bar{X}, S^2 分别是样本均值和样本方差, 则有

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

且 \bar{X} 和 S^2 相互独立。

49. 定理三: 设 X_1, X_2, \dots, X_n 是来自正态总体 $N(\mu, \sigma^2)$ 的样本, \bar{X}, S^2 分别是样本均值和样本方差, 则有

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

50. 定理四: 设 X_1, X_2, \dots, X_{n_1} 与 Y_1, Y_2, \dots, Y_{n_2} 分别是来自正态总体 $N_1(\mu_1, \sigma_1^2)$ 和 $N_2(\mu_2, \sigma_2^2)$ 的两个相互独立的样本, \bar{X}, \bar{Y} 分别是这两个样本的样本均值, S_1^2, S_2^2 分别是这两个样本的样本方差, 则有

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$$

当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 时,

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

其中

$$S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}, \quad S_w = \sqrt{S_w^2}$$

51. 设 X 是连续型随机变量, 其概率密度 $f(x; \theta)$, $\theta \in \Theta$ 的形式已知, θ 为待估参数, Θ 是 θ 可能取值的范围. 设 X_1, X_2, \dots, X_n 是来自 X 的样本, 则 X_1, X_2, \dots, X_n 的联合密度为

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

$L(\theta)$ 称为样本的似然函数. 若 $L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta)$, 则称

$\hat{\theta}(x_1, x_2, \dots, x_n)$ 为 θ 的最大似然估计值.

52. 无偏性: 设 X_1, X_2, \dots, X_n 为总体 X 的一个样本, $\theta \in \Theta$ 是包含在总体 X 的分布中的待估计函数; 若估计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 的数学期望 $E(\hat{\theta})$ 存在且对于任意 $\theta \in \Theta$ 有 $E(\hat{\theta}) = \theta$, 则称 $\hat{\theta}$ 为 θ 的无偏估计量.
53. 有效性: 设 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \dots, X_n)$ 和 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \dots, X_n)$ 都是 θ 的无偏估计量, 若有 $D(\hat{\theta}_1) \leq D(\hat{\theta}_2)$, 则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 有效.
54. 相合性: 设 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 为参数 θ 的估计量, 若对于任意 $\theta \in \Theta$, 当 $n \rightarrow \infty$ 时 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 依概率收敛于 θ , 则称 $\hat{\theta}$ 是 θ 的相合(一致)估计量.
55. 置信区间: 设总体 X 的分布函数 $F(x; \theta)$ 含有一个未知参数 θ , 对于给定值 $\alpha (0 < \alpha < 1)$, 若由样本 X_1, X_2, \dots, X_n 确定的两个统计量 $\underline{\theta} = \underline{\theta}(X_1, X_2, \dots, X_n)$ 和 $\bar{\theta} = \bar{\theta}(X_1, X_2, \dots, X_n)$ 满足

$$P\{\underline{\theta} = \underline{\theta}(X_1, X_2, \dots, X_n) < \theta < \bar{\theta} = \bar{\theta}(X_1, X_2, \dots, X_n)\} = 1 - \alpha$$

则称随机区间 $(\underline{\theta}, \bar{\theta})$ 是 θ 的置信水平为 $1 - \alpha$ 的置信区间.

56. 不同情况下求正态总体均值、方差的置信区间的枢轴量 W (P_{172}).

需要增加如下:

待估参数 σ^2 , μ 已知, 枢轴量 Z :

$$Z = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$$

表 7-1 正态总体均值、方差的置信区间与单侧置信限(置信水平为 $1-\alpha$)

待估参数	其他参数	枢轴量 W 的分布	置信区间	单侧置信限
一个正态总体	μ	$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim N(0,1)$	$(\bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2})$	$\bar{\mu} = \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha}$ $\underline{\mu} = \bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha}$
	μ	$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$	$(\bar{X} \pm \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1))$	$\bar{\mu} = \bar{X} + \frac{S}{\sqrt{n}} t_{\alpha}(n-1)$ $\underline{\mu} = \bar{X} - \frac{S}{\sqrt{n}} t_{\alpha}(n-1)$
	σ^2	$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$(\frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)}, \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)})$	$\bar{\sigma}^2 = \frac{(n-1)S^2}{\chi_{1-\alpha}^2(n-1)}$ $\underline{\sigma}^2 = \frac{(n-1)S^2}{\chi_{\alpha}^2(n-1)}$
两个正态总体	$\mu_1 - \mu_2$	$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$(\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$	$\bar{\mu}_1 - \underline{\mu}_2 = \bar{X} - \bar{Y} + z_{\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ $\underline{\mu}_1 - \bar{\mu}_2 = \bar{X} - \bar{Y} - z_{\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
	$\mu_1 - \mu_2$	$t = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$	$(\bar{X} - \bar{Y} \pm t_{\alpha/2}(n_1 + n_2 - 2) S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$	$\bar{\mu}_1 - \underline{\mu}_2 = \bar{X} - \bar{Y} + t_{\alpha}(n_1 + n_2 - 2) S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ $\underline{\mu}_1 - \bar{\mu}_2 = \bar{X} - \bar{Y} - t_{\alpha}(n_1 + n_2 - 2) S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
	$\frac{\sigma_1^2}{\sigma_2^2}$	$F = \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$	$(\frac{S_1^2}{S_2^2} \frac{F_{\alpha/2}(n_1 - 1, n_2 - 1)}{F_{1-\alpha/2}(n_1 - 1, n_2 - 1)}, \frac{S_1^2}{S_2^2} \frac{F_{1-\alpha/2}(n_1 - 1, n_2 - 1)}{F_{\alpha/2}(n_1 - 1, n_2 - 1)})$	$\bar{\frac{\sigma_1^2}{\sigma_2^2}} = \frac{S_1^2}{S_2^2} \frac{1}{F_{1-\alpha}(n_1 - 1, n_2 - 1)}$ $\underline{\frac{\sigma_1^2}{\sigma_2^2}} = \frac{S_1^2}{S_2^2} \frac{1}{F_{\alpha}(n_1 - 1, n_2 - 1)}$

57. (0-1)分布参数的区间估计：设有一容量 $n > 50$ 的大样本来自(0-1)分布的总体 X ， X 的分布律为 $f(x; p) = p^x(1-p)^{1-x}, x = 0, 1$ ，其中 p 为未知参数，则 p 的置信度为 $1-\alpha$ 的置信区间是

$$\left(\frac{-b - \sqrt{b^2 - 4ac}}{2a}, \frac{-b + \sqrt{b^2 - 4ac}}{2a} \right)$$

其中, $a = n + z_{\alpha/2}^2, b = -(2n\bar{X} + z_{\alpha/2}^2), c = n\bar{X}^2$ 。

58. 单侧置信区间 ($P_{169-170}$)。

59. 正态总体均值和方差的假设检验 ($P_{189-190}$)。

60. 置信区间与假设检验之间的关系：

要求出参数 θ 的置信水平为 $1 - \alpha$ 的置信区间, 先求出显著性水平为 α 的假设检验问题：

$H_0: \theta = \theta_0, H_1: \theta \neq \theta_0$ 的接受域 $\underline{\theta}(X_1, X_2, \dots, X_n) < \theta_0 < \bar{\theta}(X_1, X_2, \dots, X_n)$,

那么 $(\underline{\theta}(X_1, X_2, \dots, X_n), \bar{\theta}(X_1, X_2, \dots, X_n))$ 就是 θ 的置信水平为 $1 - \alpha$ 的置信区间。

61. 假设检验问题的 p 值是由检验统计量的样本观察值得出的原假设可被拒绝的最小显著性水平。对于任意指定的显著性水平 α , 有

1) 若 p 值 $\leq \alpha$, 则在显著性水平 α 下拒绝 H_0 ;

2) 若 p 值 $> \alpha$, 则在显著性水平 α 下接受 H_0 。

62. 设总体 X 的分布未知, x_1, x_2, \dots, x_n 是来自 X 的样本值, 检验假设：

H_0 ：总体 X 的分布函数为 $F(x)$ ； H_1 ：总体 X 的分布函数不是 $F(x)$ 。

其中 $F(x)$ 不含位置参数。(也常以分布律或概率密度代替 $F(x)$)。检验统计量

$$\chi^2 = \sum_{i=1}^k \frac{n}{p_i} \left(\frac{f_i}{n} - p_i \right)^2 = \sum_{i=1}^k \frac{f_i^2}{np_i} - n = \sum_{i=1}^k \frac{(f_i - np_i)^2}{np_i} \sim \chi^2(k-1)$$

拒绝域为 $\chi^2 \geq \chi_{\alpha}^2(k-1)$ 。注意 n 不能小于50, 而且应有 $np_i \geq 5$ 。

63. 若总体 X 的分布函数 $F(x)$ 未知, 即 $F(x; \theta_1, \dots, \theta_r)$, 则需先利用样本求出位置参数的最大似然估计 (在 H_0 下), 以估计值作为参数值, 求出 p_i 的估计值 \hat{p}_i , 在上式中以 \hat{p}_i 代替 p_i , 检验统计量

$$\chi^2 = \sum_{i=1}^k \frac{f_i^2}{n\hat{p}_i} - n \sim \chi^2(k-r-1)$$

其中 r 是 $F(x)$ 中未知参量的个数, 拒绝域为 $\chi^2 \geq \chi_{\alpha}^2(k-r-1)$ 。

64. 置换检验、多重假设检验 (PPT Lec08 - 2 P_{32-42})；秩和检验 ($P_{208-213}$)。

表 8-1 正态总体均值、方差的检验法(显著性水平为 α)

	原假设 H_0	检验统计量	备择假设 H_1	拒绝域
1	$\mu \leq \mu_0$ $\mu \geq \mu_0$ $\mu = \mu_0$ (σ^2 已知)	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$z \geq z_\alpha$ $z \leq -z_\alpha$ $ z \geq z_{\alpha/2}$
2	$\mu \leq \mu_0$ $\mu \geq \mu_0$ $\mu = \mu_0$ (σ^2 未知)	$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$t \geq t_\alpha(n-1)$ $t \leq -t_\alpha(n-1)$ $ t \geq t_{\alpha/2}(n-1)$
3	$\mu_1 - \mu_2 \leq \delta$ $\mu_1 - \mu_2 \geq \delta$ $\mu_1 - \mu_2 = \delta$ (σ_1^2, σ_2^2 已知)	$Z = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$\mu_1 - \mu_2 > \delta$ $\mu_1 - \mu_2 < \delta$ $\mu_1 - \mu_2 \neq \delta$	$z \geq z_\alpha$ $z \leq -z_\alpha$ $ z \geq z_{\alpha/2}$

• 190 •

第八章 假设检验

续表

	原假设 H_0	检验统计量	备择假设 H_1	拒绝域
4	$\mu_1 - \mu_2 \leq \delta$ $\mu_1 - \mu_2 \geq \delta$ $\mu_1 - \mu_2 = \delta$ ($\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知)	$t = \frac{\bar{X} - \bar{Y} - \delta}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$	$\mu_1 - \mu_2 > \delta$ $\mu_1 - \mu_2 < \delta$ $\mu_1 - \mu_2 \neq \delta$	$t \geq t_\alpha(n_1 + n_2 - 2)$ $t \leq -t_\alpha(n_1 + n_2 - 2)$ $ t \geq t_{\alpha/2}(n_1 + n_2 - 2)$
5	$\sigma^2 \leq \sigma_0^2$ $\sigma^2 \geq \sigma_0^2$ $\sigma^2 = \sigma_0^2$ (μ 未知)	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$ $\sigma^2 \neq \sigma_0^2$	$\chi^2 \geq \chi_\alpha^2(n-1)$ $\chi^2 \leq \chi_{1-\alpha}^2(n-1)$ $\chi^2 \geq \chi_{\alpha/2}^2(n-1)$ 或 $\chi^2 \leq \chi_{1-\alpha/2}^2(n-1)$
6	$\sigma_1^2 \leq \sigma_2^2$ $\sigma_1^2 \geq \sigma_2^2$ $\sigma_1^2 = \sigma_2^2$ (μ_1, μ_2 未知)	$F = \frac{S_1^2}{S_2^2}$	$\sigma_1^2 > \sigma_2^2$ $\sigma_1^2 < \sigma_2^2$ $\sigma_1^2 \neq \sigma_2^2$	$F \geq F_\alpha(n_1 - 1, n_2 - 1)$ $F \leq F_{1-\alpha}(n_1 - 1, n_2 - 1)$ $F \geq F_{\alpha/2}(n_1 - 1, n_2 - 1)$ 或 $F \leq F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$
7	$\mu_D \leq 0$ $\mu_D \geq 0$ $\mu_D = 0$ (成对数据)	$t = \frac{\bar{D} - 0}{S_D/\sqrt{n}}$	$\mu_D > 0$ $\mu_D < 0$ $\mu_D \neq 0$	$t \geq t_\alpha(n-1)$ $t \leq -t_\alpha(n-1)$ $ t \geq t_{\alpha/2}(n-1)$

65. 单因素试验的方差分析 ($P_{224-233}$):

性知, 当 H_0 为真时,

$$\frac{S_A/(s-1)}{S_E/(n-s)} = \frac{S_A/\sigma^2}{s-1} / \frac{S_E/\sigma^2}{n-s} \sim F(s-1, n-s).$$

由此得检验问题(1.2)'的拒绝域为

$$F = \frac{S_A/(s-1)}{S_E/(n-s)} \geq F_\alpha(s-1, n-s). \quad (1.20)$$

上述分析的结果可排成表 9-5 的形式, 称为方差分析表.

表 9-5 单因素试验方差分析表

方差来源	平方和	自由度	均方	F 比
因素 A	S_A	$s-1$	$\bar{S}_A = \frac{S_A}{s-1}$	$F = \frac{\bar{S}_A}{\bar{S}_E}$
误差	S_E	$n-s$	$\bar{S}_E = \frac{S_E}{n-s}$	
总和	S_T	$n-1$		

表中 $\bar{S}_A = S_A/(s-1)$, $\bar{S}_E = S_E/(n-s)$ 分别称为 S_A, S_E 的均方. 另外, 因在 S_T 中 n 个变量 $X_{ij} - \bar{X}$ 之间仅满足一个约束条件(1.6), 故 S_T 的自由度为 $n-1$.

在实际中, 我们可以按以下较简便的公式来计算 S_T, S_A 和 S_E .

$$\text{记 } T_{.j} = \sum_{i=1}^{n_j} X_{ij}, j = 1, 2, \dots, s, \quad T_{..} = \sum_{j=1}^s \sum_{i=1}^{n_j} X_{ij},$$

即有

$$\left. \begin{aligned} S_T &= \sum_{j=1}^s \sum_{i=1}^{n_j} X_{ij}^2 - n\bar{X}^2 = \sum_{j=1}^s \sum_{i=1}^{n_j} X_{ij}^2 - \frac{T_{..}^2}{n}, \\ S_A &= \sum_{j=1}^s n_j \bar{X}_{.j}^2 - n\bar{X}^2 = \sum_{j=1}^s \frac{T_{.j}^2}{n_j} - \frac{T_{..}^2}{n}, \\ S_E &= S_T - S_A. \end{aligned} \right\} \quad (1.21)$$

66. 双因素等重复试验的方差分析 ($P_{233-240}$):

取显著性水平为 α , 得假设 H_{01} 的拒绝域为

$$F_A = \frac{S_A/(r-1)}{S_E/(rs(t-1))} \geq F_\alpha(r-1, rs(t-1)). \quad (2.19)$$

类似地, 在显著性水平 α 下, 假设 H_{02} 的拒绝域为

$$F_B = \frac{S_B/(s-1)}{S_E/(rs(t-1))} \geq F_\alpha(s-1, rs(t-1)). \quad (2.20)$$

在显著性水平 α 下, 假设 H_{03} 的拒绝域为

$$F_{A \times B} = \frac{S_{A \times B}/((r-1)(s-1))}{S_E/(rs(t-1))} \geq F_\alpha((r-1)(s-1), rs(t-1)). \quad (2.21)$$

上述结果可汇总成下列的方差分析表:

表 9-9 双因素试验的方差分析表

方差来源	平方和	自由度	均方	F 比
因素 A	S_A	$r-1$	$\bar{S}_A = \frac{S_A}{r-1}$	$F_A = \frac{\bar{S}_A}{\bar{S}_E}$
因素 B	S_B	$s-1$	$\bar{S}_B = \frac{S_B}{s-1}$	$F_B = \frac{\bar{S}_B}{\bar{S}_E}$
交互作用	$S_{A \times B}$	$(r-1)(s-1)$	$\bar{S}_{A \times B} = \frac{S_{A \times B}}{(r-1)(s-1)}$	$F_{A \times B} = \frac{\bar{S}_{A \times B}}{\bar{S}_E}$
误差	S_E	$rs(t-1)$	$\bar{S}_E = \frac{S_E}{rs(t-1)}$	
总和	S_T	$rst-1$		

记

$$T_{\dots} = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t X_{ijk},$$

$$T_{ij\cdot} = \sum_{k=1}^t X_{ijk}, \quad i = 1, 2, \dots, r; j = 1, 2, \dots, s,$$

$$T_{i\cdot\cdot} = \sum_{j=1}^s \sum_{k=1}^t X_{ijk}, \quad i = 1, 2, \dots, r,$$

$$T_{\cdot j\cdot} = \sum_{i=1}^r \sum_{k=1}^t X_{ijk}, \quad j = 1, 2, \dots, s.$$

我们可以按照下述(2.22)式来计算上表中的各个平方和.

$$\left. \begin{aligned} S_T &= \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t X_{ijk}^2 - \frac{T_{\dots}^2}{rst}, \\ S_A &= \frac{1}{st} \sum_{i=1}^r T_{i\cdot\cdot}^2 - \frac{T_{\dots}^2}{rst}, \\ S_B &= \frac{1}{rt} \sum_{j=1}^s T_{\cdot j\cdot}^2 - \frac{T_{\dots}^2}{rst}, \\ S_{A \times B} &= \left(\frac{1}{t} \sum_{i=1}^r \sum_{j=1}^s T_{ij\cdot}^2 - \frac{T_{\dots}^2}{rst} \right) - S_A - S_B, \\ S_E &= S_T - S_A - S_B - S_{A \times B}. \end{aligned} \right\} \quad (2.22)$$

67. 双因素无重复试验的方差分析 ($P_{240-243}$):

• 242 •

第九章 方差分析及回归分析

表 9-15

方差来源	平方和	自由度	均方	F 比
因素 A	S_A	$r-1$	$\bar{S}_A = \frac{S_A}{r-1}$	$F_A = \bar{S}_A / \bar{S}_E$
因素 B	S_B	$s-1$	$\bar{S}_B = \frac{S_B}{s-1}$	$F_B = \bar{S}_B / \bar{S}_E$
误差	S_E	$(r-1)(s-1)$	$\bar{S}_E = \frac{S_E}{(r-1)(s-1)}$	
总和	S_T	$rs-1$		

取显著性水平为 α , 得假设 $H_{01}: \alpha_1 = \alpha_2 = \dots = \alpha_r = 0$ 的拒绝域为

$$F_A = \frac{\bar{S}_A}{\bar{S}_E} \geq F_{\alpha}((r-1), (r-1)(s-1)).$$

假设 $H_{02}: \beta_1 = \beta_2 = \dots = \beta_s = 0$ 的拒绝域为

$$F_B = \frac{\bar{S}_B}{\bar{S}_E} \geq F_{\alpha}((s-1), (r-1)(s-1)).$$

表 9-15 中的平方和可按下述式子来计算:

$$\left. \begin{aligned} S_T &= \sum_{i=1}^r \sum_{j=1}^s X_{ij}^2 - \frac{T_{..}^2}{rs}, \\ S_A &= \frac{1}{s} \sum_{i=1}^r T_{i.}^2 - \frac{T_{..}^2}{rs}, \\ S_B &= \frac{1}{r} \sum_{j=1}^s T_{.j}^2 - \frac{T_{..}^2}{rs}, \\ S_E &= S_T - S_A - S_B, \end{aligned} \right\} \quad (2.27)$$

其中 $T_{..} = \sum_{i=1}^r \sum_{j=1}^s X_{ij}$, $T_{i.} = \sum_{j=1}^s X_{ij}$, $i = 1, 2, \dots, r$,

$$T_{.j} = \sum_{i=1}^r X_{ij}, \quad j = 1, 2, \dots, s.$$

例 3 下面给出了在某 5 个不同地点、不同时间空气中的颗粒状物(以 mg/m^3 计)的含量的数据:

68. 一元线性回归 ($P_{244-249}$) :

引入下述记号 :

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)$$

则 a, b 的估计值 :

$$\hat{b} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

69. σ^2 的无偏估计量 :

$$\hat{\sigma}^2 = \frac{Q_e}{n-2} = \frac{1}{n-2} (S_{yy} - \hat{b}S_{xy})$$

70. 线性回归的显著性检验 :

$$t = \frac{\hat{b}}{\hat{\sigma}} \sqrt{S_{xx}} \sim t(n-2)$$

拒绝域 : $|t| = \frac{|\hat{b}|}{\hat{\sigma}} \sqrt{S_{xx}} \geq t_{\alpha/2}(n-2)$, 当假设 H_0 被拒绝时认为回归效果是显著的。

71. 系数 b 的置信区间 :

$$\left(\hat{b} \pm t_{\alpha/2}(n-2) \times \frac{\hat{\sigma}}{\sqrt{S_{xx}}} \right)$$

72. 回归函数值 $\mu(x_0) = a + bx_0$ 的置信区间 :

$$\left(\hat{a} + \hat{b}x_0 \pm t_{\alpha/2}(n-2) \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right)$$

73. $Y_0 = a + bx_0 + \varepsilon_0$ 的预测区间 :

$$\left(\hat{a} + \hat{b}x_0 \pm t_{\alpha/2}(n-2) \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right)$$

74. 可化为一元线性回归的例子 ($P_{255-257}$)。

75. 多元线性回归 ($P_{257-261}$)。