大话统计学　公式总整理

陈文贤　著作　《大话统计学》　清华大学出版社 2022 年

第 2 章

总体平均数 $\mu = \dfrac{x_1 + x_2 + \cdots + x_N}{N} = \dfrac{\sum_{i=1}^{N} x_i}{N}$

样本平均数 $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

加权平均数：$\bar{x}_w = \dfrac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n} = \sum_{i=1}^{n} x_i \Big/ \sum_{i=1}^{n} w_i$

几何平均数 $G = \mu_G = \bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \cdots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$

调和平均数 $H = \mu_H = \bar{x}_H = \dfrac{n}{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \cdots + \dfrac{1}{x_n}} = \dfrac{n}{\sum_{i=1}^{n} \dfrac{1}{x_i}}$

加权调和平均数 $H = \dfrac{\text{全部数据总值}}{\sum \dfrac{\text{各数据的总值}}{\text{各单位数据}}} = \dfrac{\sum M_i}{\sum \dfrac{M_i}{x_i}}$

中位数 $M_e = \begin{cases} x_{(n+1)/2} & \text{若 } n \text{ 是奇数} \\ \dfrac{x_{n/2} + x_{(n/2)+1}}{2} & \text{若 } n \text{ 是偶数} \end{cases}$

截尾平均数 $\bar{x}_T = \dfrac{\sum_{i=k+1}^{n-k} x_i}{n-2k}$　　$k = \left\lfloor \dfrac{pn}{100} \right\rfloor = (\dfrac{pn}{100}$ 取整数的部分)

百分位数 数据间距　Excel 公式 Percentile.INC(array, k)　R> quantile(x, 0.25, type=7)

(1) $k^* = \left\lfloor 1 + \dfrac{nk}{100} - \dfrac{k}{100} \right\rfloor$ 是等于或小于 $1 + \dfrac{nk}{100} - \dfrac{k}{100}$ 的最大整数，

$r = 1 + \dfrac{nk}{100} - \dfrac{k}{100} - k^*$

(2) $P_k = x_{k^*} + r(x_{k^*+1} - x_{k^*})$

百分位数 数据个数 Excel 公式 Percentile.EXC(array, k) , R> quantile(x, 0.25, type=6)

(1) $k^* = \left\lfloor \dfrac{nk + k}{100} \right\rfloor$ 是等于或小于 $\dfrac{nk+k}{100}$ 的最大整数，

$r = \dfrac{nk + k}{100} - k^*$

(2) $P_k = x_{k^*} + r(x_{k^*+1} - x_{k^*})$

异众比率 $V_r = \dfrac{\sum f_i - f_{M_o}}{\sum f_i} = 1 - \dfrac{f_{M_o}}{\sum f_i}$

$\sum f_i$ = 变量值的总频数，f_{M_o} = 众数组的频数

极差或全距 $R = x_n - x_1$

四分位差或四分位距 $Q = Q_3 - Q_1$

方差 （全部数据） $\sigma^2 = \dfrac{\sum\limits_{i=1}^{N}(x_i-\mu)^2}{N} = \dfrac{\sum\limits_{i=1}^{N}x_i^2 - 2\mu\sum\limits_{i=1}^{N}x_i + \sum\limits_{i=1}^{N}\mu^2}{N} = \dfrac{\sum\limits_{i=1}^{N}x_i^2}{N} - \mu^2 = \dfrac{\sum\limits_{i=1}^{N}x_i^2}{N} - \dfrac{\left(\sum\limits_{i=1}^{N}x_i\right)^2}{N^2}$

方差 样本数据 $s^2 = \dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \dfrac{\sum\limits_{i=1}^{n}x_i^2 - n\bar{x}^2}{n-1} = \dfrac{\sum\limits_{i=1}^{n}x_i^2}{n-1} - \dfrac{\left(\sum\limits_{i=1}^{n}x_i\right)^2}{n(n-1)}$

平均平方差 $MSD = \dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}{n}$

标准差 总体数据 $\sigma = \sqrt{\dfrac{\sum\limits_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\dfrac{\sum\limits_{i=1}^{N}x_i^2}{N} - \mu^2}$

标准差 样本数据 $s = \sqrt{\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\dfrac{\sum\limits_{i=1}^{n}x_i^2 - n\bar{x}^2}{n-1}}$

平均离均差： $MD_\mu = \sum\limits_{i=1}^{n}|x_i - \mu|/n$

样本数据平均绝对差 $MAD = \sum\limits_{i=1}^{n}|x_i - \bar{x}|/n$

以中位数为中心的平均差 平均离中差： $MD_{M_e} = \sum\limits_{i=1}^{n}|x_i - M_e|/n$

变异系数 $VC = \dfrac{\sigma}{\mu}$ 或 $VC = \dfrac{s}{\bar{x}}$

平均差系数 $MC = \dfrac{MD_\mu}{\mu}$

全距系数 $RC = \dfrac{x_{\max} - x_{\min}}{x_{\max} + x_{\min}} = \dfrac{x_n - x_1}{x_n + x_1}$

四分距系数 $QC = \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

总体数据，则，：

三阶原点距 $M_3' = \dfrac{\sum\limits_{i=1}^{N}x_i^3}{N}$

三阶中心距 $M_3 = \dfrac{\sum\limits_{i=1}^{N}(x_i-\mu)^3}{N} = \dfrac{\sum x_i^3 - 3\mu\sum x_i^2 + 3\mu^2\sum x_i - N\mu^3}{N}$

样本数据，则，：

三阶原点距 $m_3' = \dfrac{\sum\limits_{i=1}^{n}x_i^3}{n-1}$

三阶中心距 $m_3 = \dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^3}{n-1} = \dfrac{\sum x_i^3 - 3\bar{x}\sum x_i^2 + 3\bar{x}^2\sum x_i - n\bar{x}^3}{n-1}$

(1) 总体数据，四阶原点距 $M_4^{'}=\frac{\sum_{i=1}^{N}x_i^4}{N}$，四阶中心距 $M_4=\frac{\sum_{i=1}^{N}(x_i-\mu)^4}{N}$

(2) 样本数据，四阶原点距 $m_4^{'}=\frac{\sum_{i=1}^{n}x_i^4}{n-1}$，四阶中心距 $m_4=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^4}{n-1}$

一阶原点距 $M_1^{'}=\mu$，一阶中心距 $M_3=0$

二阶原点距 $M_1^{'}=\sum x^2/N$，二阶中心距 $M_3=\sigma^2$

皮尔生偏度指数 $SK=\frac{3(\mu-M_d)}{\sigma}$ $SK=\frac{3(\bar{x}-M_d)}{s}$

利用三阶距，偏度系数 $SK=\frac{M_3}{\sigma^3}$ $SK=\frac{m_3}{s^3}$

Excel 的公式，偏度系数 $SK=\frac{nM_3}{(n-2)\sigma^3}$ $SK=\frac{nm_3}{(n-2)s^3}$

利用四阶距，计算：

峰度系数 总体数据 $K=\frac{M_4}{\sigma^4}$ 峰度系数=3，正态峰型

峰度系数 样本数据 $K=\frac{m_4}{s^4}$ 峰度系数=3，正态峰型

峰度系数 样本数据 Excel 公式 $K=\frac{n(n+1)m_4}{(n-2)(n-3)s^4}-\frac{3(n-1)^2}{(n-2)(n-3)},\quad s\neq 0,\ n\geq 4$

若峰度系数等于 0，则为正态峰型

第 3 章

样本空间 ＝ S。
事件空间 ＝ Ω
事件 ＝A , B

逆事件 ＝ $\overline{\text{A}}$

概率 $P:\Omega \to [0,1]$

排列公式 $P_r^n = \dfrac{n!}{(n-r)!}$

组合公式 $C_r^n = \dbinom{n}{r} = \dfrac{n!}{r!(n-r)!}$

多项公式 $\dbinom{n}{n_1, n_2, \cdots, n_k} = \dfrac{n!}{n_1! n_2! \cdots n_k!}$

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!} = \binom{n}{r, n-r} = \binom{n}{n-r} = C_{n-r}^n$$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = \sum_{k=0}^{n} \binom{n}{k, n-k} x^k y^{n-k}$$

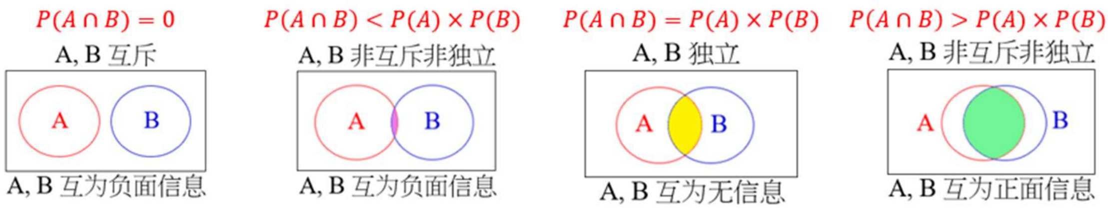
$$(x_1 + x_2 + \ldots + x_k)^n = \sum_{n_1+n_2+\ldots+n_k=n} \binom{n}{n_1, n_2, \ldots, n_m} x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}$$

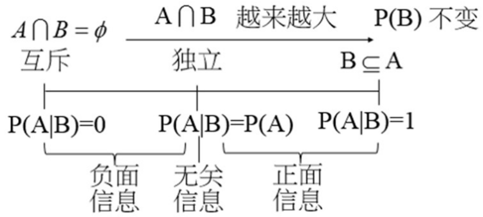
超几何公式 $C_{r_1}^{n_1} C_{r_2}^{n_2} \cdots C_{r_k}^{n_k} = \dbinom{n_1}{r_1}\dbinom{n_2}{r_2}\cdots\dbinom{n_k}{r_k} = \dfrac{n_1! n_2! \cdots n_k!}{r_1! r_2! \cdots r_k! (n_1-r_1)!(n_2-r_2)! \cdots (n_k-r_k)!}$

排列组合的公式

	只有一类(组)	
	不同物品 n 个取出 r 个	
	不放回	放回(重复)
排列	P_r^n R> DescTools CombN(n, r, repl=FALSE, ord=TRUE)	重复 排列 n^r R> DescTools CombN(3, 2, repl=TRUE, ord=TRUE)
组合	C_r^n R> DescTools CombN(3, 2, repl=FALSE, ord=FALSE)	重复组合 C_r^{n+r-1} R> DescTools CombN(3, 2, repl=TRUE, ord=FALSE)

条件概率 $P(A\mid B)=\dfrac{P(A\bigcap B)}{P(B)}$ ， $P(B)\neq 0$





加法律 $P(A\bigcup B)=P(A)+P(B)-P(A\bigcap B)$

$P(A\bigcup B\bigcup C)=P(A)+P(B)+P(C)-P(A\bigcap B)-P(A\bigcap C)-P(B\bigcap C)+P(A\bigcap B\bigcap C)$

乘法律(事件交的概率) $P(A\bigcap B)=P(A\mid B)P(B)=P(B\mid A)P(A)$

$P(A\bigcap B\bigcap C)=P(A)P(B\mid A)P(C\mid A\bigcap B)=P(B)P(C\mid B)P(A\mid B\bigcap C)=P(C)P(A\mid C)P(B\mid A\bigcap C)$

若 $P(A_1\bigcap A_2\bigcap\cdots\bigcap A_{n-1})>0$，则

$P(A_1\bigcap A_2\bigcap\cdots\bigcap A_n)=$

$P(A_1)P(A_2\mid A_1)P(A_3\mid A_1\bigcap A_2)P(A_4\mid A_1\bigcap A_2\bigcap A_3)\cdots P(A_n\mid A_1\bigcap A_2\bigcap\cdots\bigcap A_{n-1})$

概率公式 $P(B)=\sum_{i=1}^{n}P(B\bigcap A_i)=\sum_{i=1}^{n}P(B\mid A_i)P(A_i)$

贝叶斯公式 $P(A\mid B)=\dfrac{P(B\mid A)P(A)}{P(B\mid A)P(A)+P(B\mid\overline{A})P(\overline{A})}$ ， $P(B)\neq 0$

$$P(A_i\mid B)=\frac{P(B\mid A_i)P(A_i)}{\sum_{j=1}^{k}P(B\mid A_j)P(A_j)}\qquad i=1,2,\cdots,k\ ;\quad P(B)\neq 0$$

第 4 章

随机变量 $X: S \to R$

随机变量 X 的值域 $R_X = \{X(E_i) | E_i \in S\}$

随机变量 X 的概率函数 $P: R_X \to [0,1]$

X 是离散型随机变量 $R_X = \{x |\ x = 可数的\}$

X 的期望(均值) $\mu = E(X) = \sum_{i=1}^{k} x_i P(x_i)$

X 的方差 $\sigma^2 = V(X) = E[(X-\mu)^2] = \sum_{i=1}^{k} (x_i - \mu)^2 P(x_i) = E(X^2) - \mu^2$

X 是连续型随机变量 $R_X = \{x | -\infty < x < +\infty\}$

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

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = E(X^2) - \mu^2$$

1. $E(a) = a$
2. $E(bX) = bE(X)$
3. $E(a + bX) = a + bE(X)$
4. $V(a) = 0$
5. $V(bX) = b^2 V(X)$
6. $V(a + bX) = b^2 V(X)$
7. $\sigma_{bX} = |b| \sigma_X$
8. $\sigma_{a+bX} = |b| \sigma_X$

9. $E(aX + bY) = aE(X) + bE(Y)$

10. $V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2abCov(X,Y)$

11. $Cov(X,X) = V(X)$

12. $V(X \pm Y) = V(X) + V(Y) \pm 2Cov(X,Y)$

13. $Cov(X,Y) = Cov(Y,X)$

14. $Cov(a + bX, c + dY) = bdCov(X,Y)$

15. $Cov(a,Y) = 0, \quad \forall a \in R$

16. $Cov(X + Z, Y) = Cov(X,Y) + Cov(Z,Y)$

11. $Cov(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)$

例：$Cov(X+Y, Z-W) = Cov(X,Z) - Cov(X,W) + Cov(Y,Z) - Cov(Y,W)$

12. $V(aX + bY + cZ) = a^2 V(X) + b^2 V(Y) + c^2 V(Z)$

$+ 2abCov(X,Y) + 2acCov(X,Z) + 2bcCov(Y,Z)$

13. $V(\sum_{i=1}^{k} a_i X_i) = \sum_{i=1}^{k} a_i^2 V(X_i) + \sum_{i \neq j} a_i a_j Cov(X_i, X_j) = \sum_{i=1}^{k} a_i^2 V(X_i) + 2\sum_{i=1}^{k} \sum_{j=i+1}^{k} a_i a_j Cov(X_i, X_j)$

$$= \begin{pmatrix} a_1\sigma_1, & \cdots & , a_k\sigma_k \end{pmatrix} \begin{pmatrix} 1 & \cdots & \rho_{1k} \\ \vdots & \ddots & \vdots \\ \rho_{k1} & \cdots & 1 \end{pmatrix} \begin{pmatrix} a_1\sigma_1 \\ \vdots \\ a_k\sigma_k \end{pmatrix}$$

14. 若 X, Y 是独立，则 $V(XY) = [E(X)]^2 V(Y) + [E(Y)]^2 V(X) + V(X)V(Y)$

相关系数

$$\rho_{XY} = corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

总体数据协方差

$$\sigma_{XY} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$$

样本数据协方差

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

样本数据相关系数

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum(x_i - \bar{x})^2)\,(\sum(y_i - \bar{y})^2)}}$$

第 5 章

离散型概率分布

概率分布	$f_X(x)$	$F_X(x)$	$\mathbb{E}[X]$	$\mathbb{V}[X]$
U(a,b)	$\frac{I(a<x<b)}{b-a+1}$	$\begin{cases} 0 & x<a \\ \frac{\lfloor x\rfloor - a+1}{b-a} & a\le x\le b \\ 1 & x>b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Bern(p)	$p^x(1-p)^{1-x}$	$(1-p)^{1-x}$	p	$p(1-p)$
B(n,p)	$\binom{n}{x}p^x(1-p)^{n-x}$	$I_{1-p}(n-x,x+1)$	np	$np(1-p)$
Mult(n,p)	$\frac{n!}{x_1!\dots x_k!}p_1^{x_1}\cdots p_k^{x_k} \quad \sum_{i=1}^{k} x_i = n$		np_i	$np_i(1-p_i)$
HG(N,m,n)	$\frac{\binom{m}{x}\binom{m-x}{n-x}}{\binom{N}{x}}$	$\approx \Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right)$	$\frac{nm}{N}$	$\frac{nm(N-n)(N-m)}{N^2(N-1)}$
NB(r,p)	$\binom{x+r-1}{r-1}p^r(1-p)^x$	$I_p(r,x+1)$	$r\frac{1-p}{p}$	$r\frac{1-p}{p^2}$
G(p)	$p(1-p)^{x-1} \quad x\in\mathbb{N}^+$	$1-(1-p)^x \quad x\in\mathbb{N}^+$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Pois(λ)	$\frac{\lambda^x e^{-\lambda}}{x!}$	$e^{-\lambda}\sum_{i=0}^{x}\frac{\lambda^i}{i!}$	λ	λ

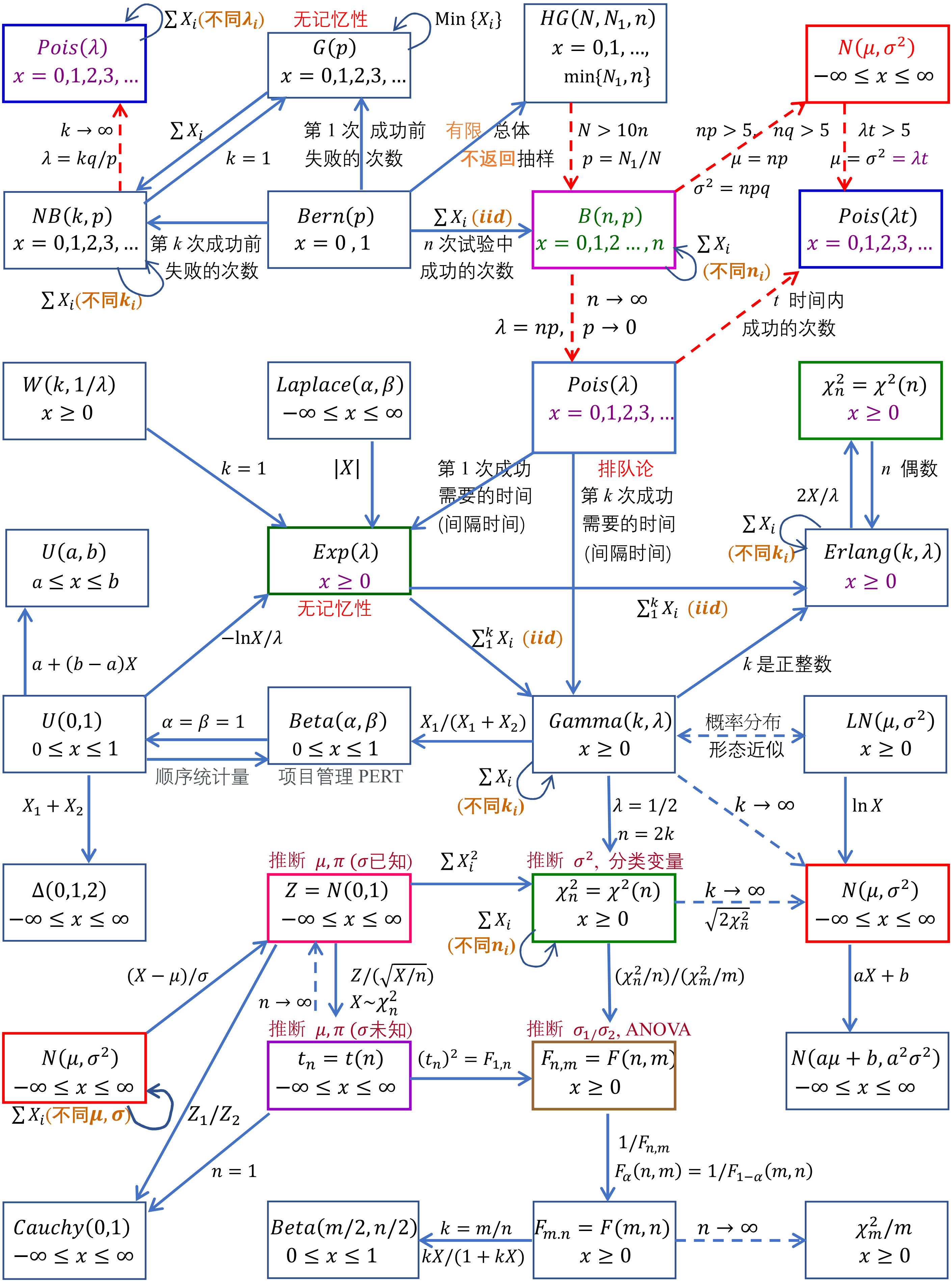
B(n,p) 二项分布； Mult(n,p) 多项分布 Multinomial

连续型概率分布

概率分布	$f_X(x)$	$F_X(x)$	$\mathbb{E}[X]$	$\mathbb{V}[X]$
U(a,b)	$\frac{I(a<x<b)}{b-a}$	$\begin{cases} 0 & x<a \\ \frac{x-a}{b-a} & a<x<b \\ 1 & x>b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
N(μ,σ²)	$\phi(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$	$\Phi(x)=\int_{-\infty}^{x}\phi(t)\,dt$	μ	σ^2
LN(μ,σ²)	$\frac{1}{x\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\right\}$	$\frac{1}{2}+\frac{1}{2}\operatorname{erf}\left[\frac{\ln x-\mu}{\sqrt{2\sigma^2}}\right]$	$e^{\mu+\sigma^2/2}$	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$
MVN(μ,Σ)	$(2\pi)^{-k/2}\|\Sigma\|^{-1/2}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$		μ	Σ
t (n)	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}$	$I_x\left(\frac{\nu}{2},\frac{\nu}{2}\right)$	0	0
χ²(k)	$\frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-x/2}$	$\frac{1}{\Gamma(k/2)}\gamma\left(\frac{k}{2},\frac{x}{2}\right)$	k	$2k$
F(d_1,d_2)	$\frac{\sqrt{\frac{(d_1x)^{d_1}d_2^{d_2}}{(d_1x+d_2)^{d_1+d_2}}}}{x\mathrm{B}\left(\frac{d_1}{2},\frac{d_1}{2}\right)}$	$I_{\frac{d_1x}{d_1x+d_2}}\left(\frac{d_1}{2},\frac{d_1}{2}\right)$	$\frac{d_2}{d_2-2}$	$\frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$
Exp(1/β)	$\frac{1}{\beta}e^{-x/\beta}$	$1-e^{-x/\beta}$	β	β^2
Gamma(α,β)	$\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}$	$\frac{\gamma(\alpha,x/\beta)}{\Gamma(\alpha)}$	$\alpha\beta$	$\alpha\beta^2$
Inverse Gamma(α,β)	$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}e^{-\beta/x}$	$\frac{\Gamma\left(\alpha,\frac{\beta}{x}\right)}{\Gamma(\alpha)}$	$\frac{\beta}{\alpha-1}\ \alpha>1$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)^2}\ \alpha>2$
Dir(αᵢ)	$\frac{\Gamma\left(\sum_{i=1}^k\alpha_i\right)}{\prod_{i=1}^k\Gamma(\alpha_i)}\prod_{i=1}^k x_i^{\alpha_i-1}$		$\frac{\alpha_i}{\sum_{i=1}^k\alpha_i}$	$\frac{\mathbb{E}[X_i](1-\mathbb{E}[X_i])}{\sum_{i=1}^k\alpha_i+1}$
Beta(α,β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$I_x(\alpha,\beta)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Weibull(λ,k)	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}e^{-(x/\lambda)^k}$	$1-e^{-(x/\lambda)^k}$	$\lambda\Gamma\left(1+\frac{1}{k}\right)$	$\lambda^2\Gamma\left(1+\frac{2}{k}\right)-\mu^2$
Pareto(x_m,α)	$\alpha\frac{x_m^\alpha}{x^{\alpha+1}}\quad x\ge x_m$	$1-\left(\frac{x_m}{x}\right)^\alpha\quad x\ge x_m$	$\frac{\alpha x_m}{\alpha-1}\ \alpha>1$	$\frac{x_m^\alpha}{(\alpha-1)^2(\alpha-2)}\ \alpha>2$

Gamma function $\Gamma(s)=\int_0^\infty t^{s-1}e^{-t}dt$

$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$, $\alpha>1$; $\Gamma(n)=(n-1)!$, $n\in N$; $\Gamma(1/2)=\sqrt{\pi}$



概率分布关联图　陈文贤 著作　可供教学，请勿引用出版。

概率符号定义 请见 《大话统计学》， 表示 独立随机变量之线性组合仍为 相同分布。

第 6 章

随机变量 X， X_i 是抽样的随机变量，期望值 $E(X_i)=\mu$，方差 $V(X_i)=\sigma^2$，样本量为 n，：

抽样平均 $\overline{X}=\dfrac{X_1+X_2+\cdots+X_n}{n}$

$$E(\overline{X})=\frac{\sum_{i=1}^{n}E(X_i)}{n}=\frac{n\mu}{n}=\mu$$

$$V(\overline{X})=\frac{1}{n^2}\sum_{i=1}^{n}V(X_i)=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}$$

抽样方差 S^2： $S^2=\dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}$

$$E(S^2)=\sigma^2$$

中心极限定理： $X_1,X_2,\cdots,X_n$ 为随机抽样，总体是任何概率分布，平均数 μ，方差 σ^2。

当抽样数目 $n\to\infty$，则 $\overline{X}\approx N\left(\mu,\dfrac{\sigma^2}{n}\right)$

若 X_i 是独立 且 $X_i\sim N(\mu,\sigma^2)$。则：

1. $\overline{X}$ 与 S^2 是独立。
2. $\overline{X}$ 是正态分布，平均数 μ，方差 σ^2/n： $\overline{X}\sim N(\mu,\sigma^2/n)$
3. $\dfrac{(n-1)S^2}{\sigma^2}$ 是卡方分布，自由度 $n-1$ ： $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$
4. $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$ 是 t 分布，自由度 $n-1$ ： $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

所以： $E(S^2)=\sigma^2$， $V(S^2)=2\sigma^4/(n-1)$

第 7 章

若 $\hat{\theta}$ 为 θ 的估计量

无偏估计量 $E(\hat{\theta})=\theta$

一致性估计量 $\lim\limits_{n\to\infty} P\{|\hat{\theta}-\theta|<\delta\}=1$，任何实数 $\delta>0$

近似无偏估计量 $\lim\limits_{n\to\infty} E(\hat{\theta})=\theta$

平均数 置信区间

条件一	条件二	条件三	置信区间
X_i 正态分布	σ 已知	有限总体 $N>20n$ 或无限总体	$\left[\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
		有限总体 $N<20n$	$\left[\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}},\ \bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right]$
	σ 未知	有限总体 $N>20n$ 或无限总体	$\left[\bar{x}-t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}},\ \bar{x}+t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\right]$
		有限总体 $N<20n$	$\left[\bar{x}-t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}},\ \bar{x}+t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right]$
X_i 非正态分布	σ 已知	$n<30$	不能用 Z 或 t 分布作区间估计，用非参数统计
		$n>30$	$\left[\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
	σ 未知	$n<30$	不能用 Z 或 t 分布作区间估计，用非参数统计
		$30<n<100$	$\left[\bar{x}-t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}},\ \bar{x}+t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\right]$
		$n \geq 100$	$\left[\bar{x}-z_{\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x}+z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$

n 为大样本，即 $t=np\geq 5$ 且 $n-t=n(1-p)\geq 5$，则利用标准正态分布：

p是近似正态分布 $p\sim N(\pi,\frac{\pi(1-\pi)}{n})$，总体比例 π 的$1-\alpha$置信区间为：

$$\left[p-z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}},\ p+z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right]$$

有限总体随机(不投返式)抽样，总体的总体数目 N，则比例 π 的$1-\alpha$置信区间为：

$$\left[p-z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}},\ p+z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}\right]$$

假设$X_1, X_2, \cdots, X_n$为 n 个正态分布的抽样随机变量，平均数 μ 和标准差 σ 未知，s 为 其抽样标准差。

总体方差 σ^2 的$1-\alpha$置信区间为 $\left[\frac{s^2(n-1)}{\chi^2_{\alpha/2}(n-1)}, \frac{s^2(n-1)}{\chi^2_{1-\alpha/2}(n-1)}\right]$

总体标准差 σ 的$1-\alpha$置信区间为 $\left[\sqrt{\frac{s^2(n-1)}{\chi^2_{\alpha/2}(n-1)}}, \sqrt{\frac{s^2(n-1)}{\chi^2_{1-\alpha/2}(n-1)}}\right]$

样本量

1. 如果总体标准差σ 已知，$1-\alpha$置信区间的长度 L (区间全长，置信区间的半径，置信区间为$p \pm E$) 确定，则样本量 n 是：

   样本量 $n = \left[\frac{2z_{\alpha/2}\sigma}{L}\right]^2$ $n = \left[\frac{z_{\alpha/2}\sigma}{E}\right]^2$

2. 如果总体标准差σ 未知，$1-\alpha$置信区间的长度 L(区间全长，有的公式取置信区间的半径，置信区间为 $p \pm E$) 确定，L = 2E，以历史资料或过去经验估计标准差 s。或者随机抽样 30 个样本，计算样本标准差 s。如果数据呈正态分布，则可以用全矩除以 4 概略估计 s。

   样本量 $n = \left[\frac{2z_{\alpha/2}s}{L}\right]^2$ $n = \left[\frac{z_{\alpha/2}s}{E}\right]^2$

3. 如果是「有限总体随机抽样」，总体数目是 N，标准差σ 已知，$1-\alpha$置信区间的长度 L，则样本量 n 是：

   样本量 $n = \frac{4z^2_{\alpha/2}\sigma^2 N}{(N-1)L^2 + 4\sigma^2 z^2_{\alpha/2}}$ $n = \frac{z^2_{\alpha/2}\sigma^2 N}{(N-1)E^2 + \sigma^2 z^2_{\alpha/2}}$

估计总体比例的抽样样本量

1. 如总体比例 π 已知(实际上是不可能的，因为 π 是要估计的参数)，$1-\alpha$置信区间的长度 L(区间全长，有的公式取置信区间的半径)确定，则样本量 n 是：

   $n = \left[\frac{2z_{\alpha/2}}{L}\right]^2 p(1-p)$

2. 如总体比例 π 未知，$1-\alpha$置信区间的长度 L 确定，则随机抽样 30 个样本作事前检验，计算样本比例 p。

   样本量 $n = \left[\frac{2z_{\alpha/2}}{L}\right]^2 p(1-p)$

   如果 n 大于 30，则再抽样 *n - 30* 个样本。

   如果 n 小于 30，则不必再抽样，利用原有的 30 个样本。

3. 如总体比例 π 未知，而不想作事前检验，取 $\pi(1-\pi)$ 的最大值 1/4。$1-\alpha$置信区间的长度 L 确定，则样本量 n 是：$n = \left[\frac{z_{\alpha/2}}{L}\right]^2$

4. 如果是「有限总体随机抽样」，总体数目是 N，$1-\alpha$置信区间的长度 L，则样本量 n 是：

   $n = \frac{4z^2_{\alpha/2}\hat{p}(1-\hat{p})N}{(N-1)L^2 + 4\hat{p}(\hat{p}-1)z^2_{\alpha/2}}$

估计	参数	估计量	概率分配	标准误差
单总体平均数 (σ^2 已知)	μ	$\overline{X}$	Z	$\dfrac{\sigma}{\sqrt{n}}$
单总体平均数 (σ^2 未知)	μ	$\overline{X}$	$t(n-1)$	$\dfrac{s}{\sqrt{n}}$
单总体比例	π	p	Z	$\sqrt{\dfrac{p(1-p)}{n}}$
单总体方差	σ^2	s^2	$\chi^2(n-1)$	$\times$
双总体平均数 (σ_1^2,σ_2^2 已知)	$\mu_1-\mu_2$	$\overline{X}_1-\overline{X}_2$	Z	$\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$
双总体平均数 (σ_1^2,σ_2^2 未知, 相等)	$\mu_1-\mu_2$	$\overline{X}_1-\overline{X}_2$	$t(n-2)$ $n=n_1+n_2$	$s_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$
双总体平均数 (σ_1^2,σ_2^2 未知, 不等)	$\mu_1-\mu_2$	$\overline{X}_1-\overline{X}_2$	$t(v)$	$\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}$
双总体比例	$\pi_1-\pi_2$	p_1-p_2	Z	$\sqrt{\dfrac{p_1(1-p_1)}{n_1}+\dfrac{p_2(1-p_2)}{n_2}}$
双总体方差	$\dfrac{\sigma_1^2}{\sigma_2^2}$	$\dfrac{s_1^2}{s_2^2}$	$F(n_1,n_2)$	$\times$
一因素 ANOVA	请见表 10.1			
二因素 ANOVA	请见表 10.2			
回归与相关	请见表 11.3			

第 8 章

第一类错误(type I error)或弃真错误

α= P{拒绝 H_0 | H_0 为真} = 拒绝「对的 H_0」的概率 =「弃真错误」的概率

第二类错误(type II error)或取伪错误

β= P{接受 H_0 | H_1 为真} = 接受「错的 H_0」的概率=「取伪错误」的概率

第一类错误α是检验的一个重要指标，称作假设检验的**显着性水平**(significance level)。

总体均值检验的样本量

双侧检验的样本量： $n=\dfrac{\left(z_{\alpha/2}+z_\beta\right)^2\sigma^2}{(\mu_0-\mu_1)^2}$

左侧与右侧检验的样本量： $n=\dfrac{(z_\alpha+z_\beta)^2\sigma^2}{(\mu_0-\mu_1)^2}$

$z_\beta=\sqrt{\dfrac{n(\mu_0-\mu_1)^2}{\sigma^2}}-z_\alpha$　　双侧检验接受域的长度　　$L=2\ z_\alpha\dfrac{\sigma}{\sqrt{n}}$

总体比例检验的样本量

双侧检验的样本量： $n=\dfrac{\left(z_{\alpha/2}\sqrt{p_0(1-p_0)}+z_\beta\sqrt{p_1(1-p_1)}\right)^2}{(p_0-p_1)^2}$

左侧检验与右侧检验的样本量： $n=\dfrac{(z_\alpha\sqrt{p_0(1-p_0)}+z_\beta\sqrt{p_1(1-p_1)})^2}{(p_0-p_1)^2}$

功效(power) = 1–β

总体均值检验，方差已知

检验值　$z^*=\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$　　若$\left|z^*\right|\ge z_{\alpha/2}$，则拒绝 H_0

总体均值检验，方差未知

检验值　$t^*=\dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}$　　若$\left|t^*\right|\ge t_{\alpha/2}(n-1)$

总体比例检验

检验值　$z^*=\dfrac{p-\pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$　　若$\left|z^*\right|\ge z_{\alpha/2}$，则拒绝 H_0

总体方差检验

$\chi_*^2=\dfrac{(n-1)s^2}{\sigma_0^2}$　　若 $\chi_*^2\le\chi_{1-\alpha/2}(n-1)$　$\chi_*^2\ge\chi_{\alpha/2}(n-1)$，

第 9 章

两组样本为独立抽样，总体方差σ_1^2及σ_2^2已知，

平均数差$\mu_1-\mu_2$的 $1-\alpha$置信区间为：

$$(\bar{x}_1-\bar{x}_2)-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\le\mu_1-\mu_2\le(\bar{x}_1-\bar{x}_2)+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

如果置信区间的长度订为L，而第一个总体的抽样成本为每个样本 c_1，第二个总体的抽样成本为每个样本c_2。则两个总体的抽样数目 n_1，n_2为：

$$n_1=4\left(\sigma_1^2+\sigma_1\sigma_2\sqrt{\frac{c_2}{c_1}}\right)\left(\frac{z_{\alpha/2}}{L}\right)^2 \quad , \quad n_2=4\left(\sigma_2^2+\sigma_1\sigma_2\sqrt{\frac{c_1}{c_2}}\right)\left(\frac{z_{\alpha/2}}{L}\right)^2$$

如果总体标准差σ_1与σ_2未知，则上述样本量公式，可用预先抽样的样本标准差s_1与s_2代入。

如果是「不重复式抽样」(抽样后不放回)，第一个总体的总体数目N_1，抽样数目为n_1，第二个总体的总体数目N_2，抽样数目为n_2，则

$$V(\bar{X}_1-\bar{X}_2)=\frac{N_1-n_1}{N_1-1}\times\frac{\sigma_1^2}{n_1}+\frac{N_2-n_2}{N_2-1}\times\frac{\sigma_2^2}{n_2}$$

平均数差$\mu_1-\mu_2$的 $1-\alpha$置信区间的上下限为：$(\bar{x}_1-\bar{x}_2)\pm z_{\alpha/2}\sqrt{V(\bar{X}_1-\bar{X}_2)}$

即 $$(\bar{x}_1-\bar{x}_2)\pm z_{\alpha/2}\sqrt{\frac{(N_1-n_1)\sigma_1^2}{(N_1-1)n_1}+\frac{(N_2-n_2)\sigma_2^2}{(N_2-1)n_2}}$$

总体方差σ_1^2及σ_2^2未知，则两总体平均数差的区间估计又分成两种情形计算。

两组样本为独立抽样，总体方差 σ_1^2及σ_2^2未知，但是相等，

则平均数差$\mu_1-\mu_2$的 $1-\alpha$置信区间为：将 $\bar{x}_1$，$\bar{x}_2$ 代入下列式子：

$$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2}(\nu)s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\le\mu_1-\mu_2\le(\bar{x}_1-\bar{x}_2)+t_{\alpha/2}(\nu)s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

合并方差 $s_p^2=\dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ ， t 分布自由度 $\nu=n_1+n_2-2$

两组样本为独立抽样，总体方差σ_1^2及σ_2^2未知，但是不相等，

则平均数差$\mu_1-\mu_2$的$1-\alpha$置信区间为，将$\bar{x}_1$，$\bar{x}_2$ 代入下列式子：

$$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2}(\nu)\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\le\mu_1-\mu_2\le(\bar{x}_1-\bar{x}_2)+t_{\alpha/2}(\nu)\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

t 分布自由度 $$\nu=\frac{\left(s_1^2/n_1+s_2^2/n_2\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1-1}+\dfrac{(s_2^2/n_2)^2}{n_2-1}}$$

两组样本为配对样本

置信区间的估计如下：

令第一个总体的样本数据为 x_{1i}，第二个总体的样本数据为 x_{2i}。两组样本数据的样本量一定相等，样本量为n。

计算 $d_i=x_{1i}-x_{2i}$ $\quad \bar{x}_d=\dfrac{\sum d_i}{n}=\bar{x}_1-\bar{x}_2$ $\quad s_d^2=\dfrac{\sum(d_i-\bar{x}_d)^2}{n-1}$

平均数差$\mu_1-\mu_2$的 $1-\alpha$置信区间为，将 $\bar{x}_d$，$s_d=\sqrt{s_d^2}$ 代入下列式子：

即 $$\bar{x}_d-t_{\alpha/2}(n-1)\frac{s_d}{\sqrt{n}}\le\mu_1-\mu_2\le\bar{x}_d+t_{\alpha/2}(n-1)\frac{s_d}{\sqrt{n}}$$

两总体比例差的区间估计

$$(p_1-p_2)\pm z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$$

「不重复式抽样」

$$(p_1-p_2)\pm z_{\alpha/2}\sqrt{\frac{N_1-n_1}{N_1-1}\times\frac{p_1(1-p_1)}{n_1}+\frac{N_2-n_2}{N_2-1}\times\frac{p_2(1-p_2)}{n_2}}$$

两总体方差比的区间估计

置信区间下限 $\bar{s}_L^2=\frac{s_1^2}{s_2^2}F_{1-\alpha/2}(n_2-1,n_1-1)$

置信区间上限 $\bar{s}_U^2=\frac{s_1^2}{s_2^2}F_{\alpha/2}(n_2-1,n_1-1)$

两个总体平均数检验，方差已知

$$z^*=\frac{\bar{x}_1-\bar{x}_2-d_0}{\sqrt{(\sigma_1^2/n_1)+(\sigma_2^2/n_2)}}$$

若 $\left|z^*\right|\ge z_{\alpha/2}$，则拒绝 H_0

两个总体平均数检验，方差未知但相等

$$t^*=\frac{\bar{x}_1-\bar{x}_2-d_0}{s_p\sqrt{(1/n_1)+(1/n_2)}}\qquad s_p=\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

若 $|t^*|\ge t_{\alpha/2}(\nu)$，则拒绝 H_0

两个总体平均数检验，方差未知且不等

$$t^*=\frac{\bar{x}_1-\bar{x}_2-d_0}{\sqrt{(s_1^2/n_1)+(s_2^2/n_2)}}\qquad \nu=\frac{\left(s_1^2/n_1+s_2^2/n_2\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$$

若 $|t^*|\ge t_{\alpha/2}(\nu)$

两个总体平均数检验，样本是配对样本

$$t^*=\frac{\bar{x}_1-\bar{x}_2-d_0}{s_d/\sqrt{n}}$$

若 $|t^*|\ge t_{\alpha/2}(n-1)$

两个总体方差检验

$$f^*=\frac{s_1^2}{s_2^2}$$

p 值= $2\min\{P(F(n_1-1,n_2-1)\le f^*),\ P(F(n_1-1,n_2-1)\ge f^*)\}$

两个总体比例检验

若 $d_0=0$，则 $s_{p_1-p_2}=\sqrt{\left(\frac{n_1p_1+n_2p_2}{n_1+n_2}\right)\left(1-\frac{n_1p_1+n_2p_2}{n_1+n_2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

若 $d_0\ne0$，则 $s_{p_1-p_2}=\sqrt{\left(\frac{p_1(1-p_1)}{n_1}\right)+\left(\frac{p_2(1-p_2)}{n_2}\right)}$

第 10 章

$$Y_{ij}=\mu_i+\varepsilon_{ij}=\mu+\alpha_i+\varepsilon_{ij} \qquad i=1,\ldots,a;\quad j=1,\ldots,n$$

$$SS_T=\sum_{i,j}(y_{ij}-\bar{\bar{y}})^2$$

$$SS_A=\sum_{i,j}(\bar{y}_i-\bar{\bar{y}})^2$$

$$SS_E=\sum_{i,j}(y_{ij}-\bar{y}_i)^2$$

$$\sum_i\sum_j(y_{ij}-\bar{\bar{y}})^2=\sum_i\sum_j(\bar{y}_i-\bar{\bar{y}})^2+\sum_i\sum_j(y_{ij}-\bar{y}_i)^2$$

$$\Downarrow \qquad\qquad \Downarrow \qquad\qquad \Downarrow$$

$$SS_T= \qquad SS_A \qquad + \qquad SS_E$$

$$MS_A=\frac{SS_A}{a-1}=\frac{n\sum_{i=1}^a(\bar{y}_i-\bar{\bar{y}})^2}{a-1}$$

$$MS_E=\frac{SS_E}{a(n-1)}=\frac{\sum_{i=1}^a\sum_{j=1}^n(y_{ij}-\bar{y}_i)^2}{a(n-1)}$$

$$\Rightarrow\quad F=\frac{MS_A}{MS_E}=\frac{\dfrac{(a-1)MS_A/\sigma^2}{a-1}}{\dfrac{(an-a)MS_E/\sigma^2}{an-a}}\sim F(a-1,an-a)$$

方差分析表：

变异来源	平方和	自由度	均方	F 比值
处理之间(因素)	$SS_A=n\sum_{i=1}^a(\bar{y}_i-\bar{\bar{y}})^2$	$a-1$	$MS_A=\frac{SS_A}{a-1}$	$F=\frac{MS_A}{MS_E}$
处理之内(误差)	$SS_E=\sum_{i=1}^a\sum_{j=1}^n(y_{ij}-\bar{y}_i)^2$	$a(n-1)$	$MS_E=\frac{SS_E}{a(n-1)}$	
总和	$SS_T=\sum_{i=1}^a\sum_{j=1}^n(y_{ij}-\bar{\bar{y}})^2$	$an-1$		

若 $F\ge F_\alpha(a-1,a(n-1))$，则拒绝 H_0。

Bartlett 检验法。假定 $X_i\sim N(\mu_i,\sigma_i^2)$

从第 i 个总体随机抽样 n_i 个样本。总样本量 $N=\sum_{i=1}^k n_i$。

计算每个总体的样本方差 s_i^2， $i=1,\ldots,k$

$$s_i^2=\frac{\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2}{n_i-1}$$

$$s^2=\frac{\sum_{i=1}^k\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2}{N-k}=\frac{\sum_{i=1}^k(n_i-1)s_i^2}{N-k}$$

$$C = 1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{N-k}\right]$$

$$B = \frac{1}{C}\left[(N-k)\log_e s^2 - \sum_{i=1}^{k}(n_i - 1)\log_e s_i^{\,2}\right]$$

$$B = \frac{\log_e 10}{C}\left[(N-k)\log_{10} s^2 - \sum_{i=1}^{k}(n_i - 1)\log_{10} s_i^{\,2}\right] \qquad \log_e 10 = 2.3026$$

当 $B > \chi^2_{\alpha,k-1}$，则 拒绝 H_0

双因素方差分析 ANOVA 方差分析表：

变异来源	平方和	自由度	平均平方和	F 比值
因素 A	$SS_A = b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{\bar{y}})^2$	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F_1 = \frac{MS_A}{MS_E}$
因素 B	$SS_B = a\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{\bar{y}})^2$	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$F_2 = \frac{MS_B}{MS_E}$
误差	$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	$(a-1)\times(b-1)$	$MS_E = \frac{SS_E}{(a-1)(b-1)}$	
总和	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{\bar{y}})^2$	$ab-1$		

若 $F_1 \geq F_\alpha[a-1,(a-1)(b-1)]$，则拒绝 H_0^1。 若 $F_2 \geq F_\alpha[(b-1,(a-1)(b-1)]$，则拒绝 H_0^2。

第 11 章

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i = 1,\ldots,n$$

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

$$Cov(X,Y) = \frac{S_{xy}}{n-1}$$

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

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

$$\hat{\beta}_1 = b_1 = \frac{S_{xy}}{S_{xx}} = \frac{Cov(X,Y)}{S_x^2}$$

$$\hat{\beta}_0 = b_0 = \bar{y} - b_1\bar{x}$$

$$\hat{y}_i = b_0 + b_1 x_i$$

$$SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 = \sum_{i=1}^{n}(y_i - \bar{y} - b_1\bar{x} - b_1 x_i)^2$$

$$= \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2b_1\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) + b_i^2 S_{xx} = S_{yy} - 2b_1 S_{xy} + b_1^2 S_{xx} \quad SS_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = b_1 S_{xy} = r^2 SS_T$$

$$= S_{yy} - b_1 S_{xy} = S_{yy} - b_1^2 S_{xx} = (1 - r^2)SS_T$$

$$MS_E = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{SS_E}{n-2}$$

$$SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2 = SS_E + SS_R = S_{yy}$$

$$\hat{\sigma} = \sqrt{\frac{SS_E}{n-2}}$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = b_1\sqrt{\frac{S_{xx}}{S_{yy}}} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum(x_i - \bar{x})^2)(\sum(y_i - \bar{y})^2)}}$$

$$r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

一元回归的方差分析

变异来源	自由度	平方和	平均平方和	F 比值
回归模型 (已解释变异)	1	SS_R	$MS_R = SS_R$	$F = \frac{MS_R}{MS_E}$
误差 (未解释变异)	$n-2$	SS_E	$MS_E = \frac{SS_E}{n-2}$	
总和	$n-1$	SS_T		

若 $F \geq F_\alpha(1, n-2)$，则拒绝 H_0，即自变量与因变量有显着关系。

3．利用 t 分布检验 $H_0: \rho = 0$ 的假设：

计算 $t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ r 是样本相关系数。

若 $|t|\geq t_{\alpha/2}(n-2)$，则拒绝 H_0。

一元线性回归与相关分析的有关分布

参数	估计量	有关分布
β_0	b_0	$\sqrt{\frac{nS_{xx}}{MS_E\sum_i x_i^2}}(b_0-\beta_0)\sim t(n-2)$
β_1	b_1	$\sqrt{\frac{S_{xx}}{MS_E}}(b_1-\beta_1)\sim t(n-2)$
$\beta_0+\beta_1x_p$	$b_0+b_1x_p$	$\frac{b_0+b_1x_p-\beta_0-\beta_1x_p}{\sqrt{MS_E\left[\frac{1}{n}+\frac{(x_p-\bar{x})^2}{S_{xx}}\right]}}\sim t(n-2)$
y_p	$b_0+b_1x_p$	$\frac{y_p-b_0-b_1x_p}{\sqrt{MS_E\left[1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{S_{xx}}\right]}}\sim t(n-2)$
σ^2	$\frac{SS_E}{n-2}$	$\frac{SS_E}{\sigma^2}\sim\chi^2(n-2)$
ρ	r	$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\sim t(n-2)$ (当 ρ 接近 0)

第 12 章

多项分布卡方检验

1. 计算 $k^* = \sum_{i=1}^{k} \frac{(y_i - e_i)^2}{e_i} = \sum_{i=1}^{k} \frac{(y_i - n\pi_i^0)^2}{n\pi_i^0} = \frac{1}{n}\sum_{i=1}^{k} \frac{y_i^2}{\pi_i^0} - n$ = 检验 H_0 的统计值
2. 以上 K 的分布近似卡方分布，利用卡方检验。
3. 若 $k^* \geq \chi_\alpha^2(k-1)$，则拒绝 H_0
4. p 值 $=P\{\chi^2(k-1)\geq k^*\}$，若 p 值≤$\alpha$，则拒绝 H_0

拟合优度检验

1. 计算 $k^* = \sum_{i=1}^{k} \frac{(y_i - e_i)^2}{e_i} = \sum_{i=1}^{k} \frac{(y_i - nP(A_i))^2}{nP(A_i)}$ = 检验 H_0 的统计值
2. 如果概率分布 F 有 m 个参数，若 $k^* \geq \chi_\alpha^2(k-m-1)$，则拒绝 H_0
3. 若 $k^* \geq \chi_\alpha^2(k-1)\}$，则拒绝 H_0
4. p 值 $=P\{\chi^2(k-1)\geq k^*\}$，若 p 值≤$\alpha$，则拒绝 H_0

独立性卡方检验

1. 计算 $R_i = \sum_{j=1}^{b} y_{ij}$，$C_j = \sum_{i=1}^{a} y_{ij}$，$n = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}$，$e_{ij} = \frac{R_i C_j}{n}$
2. $K = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(Y_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{Y_{ij}^2}{e_{ij}} - n$ 是一个检验 H_0 的统计量。
3. 计算 $k^* = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(y_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{y_{ij}^2}{e_{ij}} - n$ = 检验 H_0 的统计值
4. 若 $k^* \geq \chi_\alpha^2((a-1)\times(b-1))$，则拒绝 H_0。

中位数卡方检验

独立样本	样本 I	样本 II	总和
大于共同中位数的样本量	a	b	$a+b$
小于共同中位数的样本量	c	d	$c+d$
总和	$a+c$	$b+d$	$n = a+b+c+d$

计算 $k^* = \frac{a^2}{(a+b)(a+c)/n} + \frac{b^2}{(a+b)(b+d)/n} + \frac{c^2}{(c+d)(a+c)/n} + \frac{d^2}{(c+d)(b+d)/n} - n$

经过代数运算后 $k^* = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$， 若 $k^* \geq \chi_{\alpha,1}^2$，则拒绝 H_0。

McNemar 检验 -- 两总体配对样本比例检验

配对样本		甲总体 A (录取)	甲总体 B (不录取)	总和
乙总体	A (录取)	a	b	$a+b$
	B (不录取)	c	d	$c+d$
总和		$a+c$	$b+d$	$n = a+b+c+d$

计算 $\chi^* = \frac{(b-c)^2}{b+c}$ 或修正为 $\chi^* = \frac{(|b-c|-1)^2}{b+c}$， 若 $\chi^* \geq \chi_{\alpha,1}^2$，则拒绝 H_0。

列联系数

$$\chi^2=\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{(Y_{ij}-e_{ij})^2}{e_{ij}}=\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij}^2}{e_{ij}}-n$$

1. ϕ 系数

$$\phi=\sqrt{\frac{\chi^2}{n}}$$

R 语言 R> DescTools::Phi(table) # 请见 12.11 节

2. *Phi* 系数

$$Phi=\frac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

		0	1	
属	0	*a*	*b*	*a*+*b*
性	1	*c*	*d*	*c*+*d*
		a+*c*	*b*+*d*	*n* = *a*+*b*+*c*+*d*

R 语言 R> psych::phi(t) # t 是 1×4 向量 或 2×2 矩阵，请见 12.11 节

3. 皮尔逊列联系数(Pearson’s contingency coefficient)

$$C=\sqrt{\frac{\chi^2}{n+\chi^2}}$$

R 语言 R> DescTools::ContCoef(table) # 请见 12.11 节

4. Cramer V 系数

$$V=\sqrt{\frac{\chi^2}{n\cdot(k-1)}}\qquad k=\min\{\mathrm{a},\mathrm{b}\}$$

R 语言 R> DescTools::CramerV(table) # 请见 12.11 节

第 13 章

符号检验

1. 计算 $x_-^* =$ 所有样本数据 $\{x_1, x_2, \cdots, x_n\}$ 中小于 M_0 的数目。
. 计算 $x_+^* =$ 所有样本数据 $\{x_1, x_2, \cdots, x_n\}$ 中大于 M_0 的数目。
2. 检验值 $v^* = \min\{x_-^*, x_+^*\} =$ 双侧检验 H_0^I 的统计值。
$v^* = x_-^* =$ 左侧检验 H_0^{II} 的统计值。 $v^* = x_+^* =$ 右侧检验 H_0^{III} 的统计值。
3. 利用二项分布 $P\{B(n,0.5) \le v^*\} = \sum_{i=0}^{v^*} \binom{n}{i} (0.5)^n$

双侧检验 H_0^I， p 值 $= 2P\{B(n,0.5) \le v^*\}$。
. 左侧检验 H_0^{II}， p 值 $= P\{B(n,0.5) \le v^*\}$。
右侧检验 H_0^{III}， p 值$= P\{B(n,0.5) \ge v^*\}$。
如果 p 值 $\le \alpha$，则拒绝 H_0。
4. 利用正态分布近似，若 $n \ge 10$，则 $B(n,0.5)$ 近似正态分布 $N(0.5n, 0.25n)$：

$$z^* = \frac{2v^* - n}{\sqrt{n}} \quad \text{或} \quad \text{连续性修正} \begin{cases} z^* = \dfrac{(v^* + 0.5) - 0.5n}{0.5\sqrt{n}} & if\ v^* < 0.5n \\ z^* = \dfrac{(v^* - 0.5) - 0.5n}{0.5\sqrt{n}} & if\ v^* > 0.5n \end{cases}$$

双侧检验 H_0^I， p 值 $= 2P\{Z \ge |z^*|\}$。
. 左侧检验 H_0^{II}， p 值 $= P\{Z \le z^*\}$。 . 右侧检验 H_0^{III}， p 值 $= P\{Z \ge z^*\}$。
如果 p 值 $\le \alpha$，则拒绝 H_0。

表 13.3 符号检验步骤

检验方法	双侧检验 $H_0: M = M_0$ $H_1: M \ne M_0$	左侧检验 $H_0: M \ge M_0$ $H_1: M < M_0$	右侧检验 $H_0: M \le M_0$ $H_1: M > M_0$
二项分布	$v^* = \min\{x_-^*, x_+^*\}$ p 值$= 2P\{B(n,0.5) \le v^*\}$ 若 p 值$<\alpha$，则拒绝 H_0	$v^* = x_+^*$ p 值$= P\{B(n,0.5) \le v^*\}$ 若 p 值$<\alpha$，则拒绝 H_0	$v^* = x_+^*$ p 值$= P\{B(n,0.5) \ge v^*\}$ 若 p 值$<\alpha$，则拒绝 H_0
正态分布(近似)	z^* 定义在公式(13.1) p 值$= 2P\{Z \ge \|z^*\|\}$ 若 p 值$<\alpha$，则拒绝 H_0	z^* 定义在公式(13.1) p 值$= P\{Z \le z^*\}$ 若 p 值$<\alpha$，则拒绝 H_0	z^* 定义在公式(13.1) p 值$= P\{Z \ge z^*\}$ 若 p 值$<\alpha$，则拒绝 H_0

检验双总体配对样本
1. 如果样本数据 $\{x_1, x_2, \cdots, x_n\}$ 与样本数据 $\{y_1, y_2, \cdots, y_n\}$ 是「配对/相依性」样本。
2. 计算 $d_i = x_i - y_i$，如果样本数据 $\{d_1, d_2, \cdots, d_n\}$ 中有等于 0，则样本替删除这些等于 $d_i = 0$ 的资料。样本量 n 改为新的样本量目。
3. 计算 $x_-^* =$ 所有数据 $\{d_1, d_2, \cdots, d_n\}$ 中小于 0 的数目。
计算 $x_+^* =$ 所有样本数据 $\{d_1, d_2, \cdots, d_n\}$ 中大于 0 的数目。

符号秩检验 检验单总体中位数
$y_i = x_i - M_0 =$ 所有样本数据 $\{x_1, x_2, \cdots, x_n\}$ 减去 M_0。
2. 如果有 $y_i = 0$，则不计算顺序，样本量减 1。
3. 将 y_i 取绝对值 $|y_i|$，然后按大小，由小到大排列顺序。

计算统计值 $x_+^* =$ 所有样本数据 $\{x_1, x_2, \cdots, x_n\}$ 中，大于 M_0 的顺序加起来。

$$V \sim N(\mu, \sigma^2), \quad \mu = \frac{n(n+1)}{4}, \quad \sigma^2 = \frac{n(n+1)(2n+1)}{24}, \quad z^* = \frac{x_+^* - \mu}{\sigma} \qquad (13.2)$$

表 13.4 符号秩 Wilcoxon 检验步骤

检验方法	双侧检验 $H_0: M = M_0$ $H_1: M \neq M_0$	左侧检验 $H_0: M \geq M_0$ $H_1: M < M_0$	右侧检验 $H_0: M \leq M_0$ $H_1: M > M_0$
递归方程式	补充教材	补充教材	补充教材
正态分布 (近似)	$z^* = \dfrac{x_+^* - \mu}{\sigma}$ 公式(13.2) p 值=$2P\left\{Z \geq \left\lvert z^* \right\rvert\right\}$ 若 p 值$<\alpha$，则拒绝H_0	$z^* = \dfrac{x_+^* - \mu}{\sigma}$ p 值=$P\left\{Z \leq z^*\right\}$ 若 p 值$<\alpha$，则拒绝H_0	$z^* = \dfrac{x_+^* - \mu}{\sigma}$ p 值=$P\left\{Z \geq z^*\right\}$ 若 p 值$<\alpha$，则拒绝H_0
查表法	$t^* = \min\left\{x_-^*, x_+^*\right\}$ 查表 A7，得T_L，T_U。 若 $v^* \leq T_L$ 或 $v^* \geq T_U$， 则拒绝H_0	$t^* = \min\left\{x_-^*, x_+^*\right\}$ 查表 A7，得 T_L 若 $v^* \leq T_L$，则拒绝H_0	$t^* = \max\left\{x_-^*, x_+^*\right\}$ 查表 A7，得 T_U 若 $v^* \geq T_U$，则拒绝H_0

检验双总体配对样本

如果样本数据$\{x_1, x_2, \cdots, x_n\}$与样本数据$\{y_1, y_2, \cdots, y_n\}$是「配对/相依性」样本。

2. 计算 $d_i = x_i - y_i$，如果样本数据$\{d_1, d_2, \cdots, d_n\}$中有等于 0，则样本替删除这些等于$d_i = 0$的资料。样本量 n 改为新的样本量目。

3. 计算x_-^* =所有数据$\{d_1, d_2, \cdots, d_n\}$中小于 0 的数目。
   计算x_+^* =所有样本数据$\{d_1, d_2, \cdots, d_n\}$中大于 0 的数目。

4. 令 $M_0 = 0$，以下步骤同符号秩 Wilcoxon 检验步骤第 5 步。

游程检验

R = 随机变量的游程的数目　　r^* = 样本数据的游程的数目

R 的分布如下：

$$P\{R = 2k\} = 2\frac{\binom{m-1}{k-1}\binom{n-1}{k-1}}{\binom{m+n}{m}} \qquad P\{R = 2k+1\} = \frac{\left[\binom{m-1}{k-1}\binom{n-1}{k} + \binom{m-1}{k}\binom{n-1}{k-1}\right]}{\binom{m+n}{m}} \qquad (13.3)$$

表 13.5 游程检验步骤

利用 R 的分布(用计算机)	正态分布	查表法
r^* = 游程的数目 $P(R = i)$的分布公式(13.3) p 值 = $2\min\left(P\left\{R \leq r^*\right\}, P\left\{R \geq r^*\right\}\right)$	r^* =游程的数目 $z^* = \dfrac{r^* - \mu}{\sigma}$ 公式(13.4)	r^* =游程的数目 查表 A10 得 L_r，U_r 若 $r^* \leq L_r$ 或 $r^* \geq U_r$，则拒绝H_0 查表 A11 得 $P(R \leq r^*)$ p 值=$2\min\left(P\left\{R \leq r^*\right\}, P\left\{R \geq r^*\right\}\right)$

检验双总体独立样本

1. 将两组样本数据混合，从小到大排列顺序。
2. 在排好顺序的数据上，若 x_i，则记作＋；若 y_i，则记作－。
3. r^* = 样本数据的游程的数目，为检验 H_0 的统计值。
4. 以下步骤同上述游程检验步骤第 4 步。

秩和检验 Mann-Whitney 检验

所有样本数据$\{x_1 - d_0, x_2 - d_0, \cdots, x_n - d_0, y_1, y_2, \cdots y_m\}$合并，按大小顺序，由小到大排列。($m \geq n$)。

2. 令 $R_i = X_i - d_0$的排列顺序(随机变量)：$T = \sum_{i=1}^{n} R_i$。

计算 $r_i = x_i - d_0$ 的排列顺序；$t^* = \sum_{i=1}^{n} r_i$

$$T \sim N\left(\mu, \sigma^2\right), \quad \mu = \frac{n(n+m+1)}{2}, \quad \sigma^2 = \frac{nm(n+m+1)}{12} \tag{13.5}$$

表 13.6　Mann-Whitney 检验步骤

	双侧	左侧	右侧
检验方法	$H_0: M_x = M_y$ $H_1: M_x \neq M_y$	$H_0: M_x \geq M_y$ $H_1: M_x < M_y$	$H_0: M_x \leq M_y$ $H_1: M_x > M_y$
递归方程式	补充教材	补充教材	补充教材
正态分布 (近似)	$z^* = \dfrac{t^* - \mu}{\sigma}$ 公式(13.5) p 值 = $2P\left\{Z \geq \left\lvert z^* \right\rvert\right\}$ 若 p 值＜α，则拒绝 H_0	$z^* = \dfrac{t^* - \mu}{\sigma}$ p 值 = $P\left\{Z \leq z^*\right\}$ 若 p 值＜α，则拒绝 H_0	$z^* = \dfrac{t^* - \mu}{\sigma}$ p 值 = $P\left\{Z \geq z^*\right\}$ 若 p 值＜α，则拒绝 H_0
查表法	秩和检验查表 A8 MW U 检验查表 A9	秩和检验查表 A8 MW U 检验查表 A9	秩和检验查表 A8 MW U 检验查表 A9
模拟法	补充教材	补充教材	补充教材

Kruskal-Wallis 检验

1. 将 x_{ij} 「全部」按照大小，由小到大排列。
2. $r_{ij} = x_{ij}$ 在所有 N 个数据的排序，1 是最小，N 是最大，相同大小，排序取平均。
3. $R_i = \sum_{j=1}^{n_i} r_{ij}$，$\bar{R}_i = \frac{R_i}{n_i}$　(R_i 是第 i 组(总体)样本数据之排序的总和)
4. 检验值 $H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left(\bar{R}_i - \frac{N+1}{2}\right)^2 = \frac{12}{N(N+1)} \sum_{i=1}^{k} \left(\frac{R_i^2}{n_i}\right) - 3(N+1)$　(13.6)
5. 若 $H \geq \chi^2_{\alpha,(k-1)}$，则拒绝 H_0。

Friedman 检验

1. 将 x_{ij} 在「每个集区 i」按照大小，由小到大排列。
2. $R_{ij} = x_{ij}$ 在第 i 个集区数据的排序，1 是最小，k 是最大，相同大小，排序取平均。
3. $R_j = \sum_{i=1}^{n} R_{ij} =$ 在第 j 组(总体)样本数据之排序的总和
4. 检验值 $H = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1)$　(13.7)
5. 若 $H \geq \chi^2_\alpha(k-1)$，则拒绝 H_0。

Spearman 秩相关系数

1. 样本数据 $\{x_1, x_2, \cdots, x_n\}$，按大小顺序，由小到大排列。
   样本数据 $\{y_1, y_2, \cdots, y_n\}$，按大小顺序，由小到大排列。
2. $r_i = x_i$ 的排列顺序；$s_i = y_i$ 的排列顺序；　$d_i = r_i - s_i$
3. 以 r_i 和 s_i 计算相关系数：

$$r_{sp} = \frac{S_{rs}}{\sqrt{S_{rr} S_{ss}}} = \frac{\sum \left(r_i - \frac{n+1}{2}\right)\left(s_i - \frac{n+1}{2}\right)}{n\left(n^2 - 1\right)/12} = \frac{12\left[\sum r_i s_i - n\left(\frac{n+1}{2}\right)^2\right]}{n\left(n^2 - 1\right)} \tag{13.8}$$

第 14 章

1. 加法模型： $Y = T + S + C + I$
2. 乘法模型： $Y = T \times S \times C \times I$

平稳序列

平稳序列(Stationary time series)有

严平稳(strictly stationary)序列和

弱平稳(weakly stationary)序列，平稳序列是弱平稳序列。

定义 时间序列满足下列条件，称为平稳序列：

(1). $E(y_t) = E(y_{t-s}) = \mu_y$

(2). $V(y_t) = V(y_{t-s}) = \sigma_y^2$

(3). $Cov(y_t, y_{t-s}) = Cov(y_{t-j}, y_{t-j-s}) = \gamma_s$

以上式子对所有 t , t-s , t-j , t-j-s 都成立。

平稳型序列预测

简单移动平均法

$$F_t = \frac{Y_{t-n} + Y_{t-n+1} + \ldots + Y_{t-2} + Y_{t-1}}{n} = \frac{1}{n}\sum_{i=i}^{n} Y_{t-i} = F_{t-1} + \frac{Y_{t-1} - Y_{t-n-1}}{n}$$

加权移动平均

$$F_t = (W_{t-n}Y_{t-n} + W_{t-n+1}Y_{t-n+1} + \ldots + W_{t-2}Y_{t-2} + W_{t-1}Y_{t-1})/(\sum_{i=1}^{n} W_{t-i}) = \sum_{i=i}^{n} W_{t-i}Y_{t-i} \Big/ \sum_{i=i}^{n} W_{t-i}$$

简单指数平滑法

$$F_t = F_{t-1} + \alpha(Y_{t-1} - F_{t-1}) = \alpha Y_{t-1} + (1-\alpha)F_{t-1}$$

趋势型序列预测

双重移动平均

$$M_t = (Y_{t-n+1} + \ldots + Y_{t-1} + Y_t)/n = M_{t-1} + (Y_t - Y_{t-n})/n \ , \quad t \ge k$$

$$M_t^{(2)} = (M_{t-n+1} + \ldots + M_{t-1} + M_t)/n = M_{t-1}^{(2)} + (M_t - M_{t-n})/n \, , \quad t \ge 2k-1$$

$$a_t = 2M_t - M_t^{(2)}$$

$$b_t = 2(M_t - M_t^{(2)})/(n-1)$$

所以第 $t+k$ 期的预测值 $F_{t+k} = a_t + b_t \times k$

一元线性回归

$\hat{Y}_t = a + bt \qquad t = 1,2,\ldots,n$

$$b = \left(n\sum_{t=1}^{n} tY_t - \sum_{t=1}^{n} t \sum_{t=1}^{n} Y_t\right) \Bigg/ \left(n\sum_{t=1}^{n} t^2 - \left(\sum_{t=1}^{n} t\right)^2\right) , \qquad a = \left(\sum_{t=1}^{n} Y_t - b\sum_{t=1}^{n} t\right) \Bigg/ n$$

$F_t = a + bt$

双重指数平滑

$$S_t = \alpha Y_t + (1-\alpha)S_{t-1} \qquad S_t^{(2)} = \alpha S_t + (1-\alpha)S_{t-1}^{(2)}$$

$$a_t = 2S_t - S_t^{(2)} \qquad b_t = \frac{\alpha}{1-\alpha}(S_t - S_t^{(2)})$$

$F_{t+k} = a_t + b_t \times k$

趋势指数平滑

$$LT_t = \alpha(Y_t - Y_{t-1}) + (1-\alpha)(F_t - F_{t-1})$$

$$ET_{t+1} = \beta(LT_t) + (1-\beta)(ET_t)$$

$F_{t+1} = \alpha Y_t + (1-\alpha)F_t + ET_{t+1}$

第 1 步：决定最初值(Y_0)，估计趋势值(ET_1)，平滑常数 α，趋势平滑常数 β 。

第 2 步：计算第一期的预测值 $F_1 = Y_0 + ET_1$。

第 3 步：假设有 n 个实际值，计算第二期估计趋势值(ET_2)：

$LT_t = \alpha(Y_t - Y_{t-1}) + (1-\alpha)(F_t - F_{t-1})$

$ET_{t+1} = \beta(LT_t) + (1-\beta)(ET_t)$

用趋势指数平滑法来计算未来的预测值(F_t)：

$F_{t+1} = \alpha Y_t + (1-\alpha)F_t + ET_{t+1}$

第 4 步：依此类推，将未来 n - 2 期之估计趋势值(ET_t)与预测值(F_t)。

第 5 步：n 期以后的第 k 期预测值为 $F_{n+k} = \alpha Y_n + (1-\alpha)F_n + k(ET_{n+1})$

季节指数分析

中央移动平均法

趋势方程拟合法

时间序列预测方法：趋势加季节

相加性季节指数

利用多元回归方法

指数平滑模型 Holt Winter models

模型 (T,S)： T 是趋势，S 是季节，N 是无加法或乘法，A 是有加法，M 是有乘法。

1.（N,N）模型: 简单指数平滑(simple exponential smoothing model)

$\hat{y}_{t+h} = L_t$

$L_t = \alpha y_t + (1-\alpha)L_{t-1}$

2.（A,N）模型: Holt 加法模型 Holt's additive Method

$\hat{y}_{t+h} = L_t + hT_t$

$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + T_{t-1})$

$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$

3.（M,N）模型: Holt 乘法模型 (Holt's multiplicative Method)

$\hat{y}_{t+h} = L_t \times hT_t$

$L_t = \alpha y_t + (1-\alpha)(L_{t-1} \times T_{t-1})$

$T_t = \beta(L_t \div L_{t-1}) + (1-\beta)T_{t-1}$

4.（A_d,N）模型: Holt 趋势阻尼(衰减)模型 (Holt's Damped Model)

$\hat{y}_{t+h} = L_t + (\emptyset + \emptyset^2 + \cdots + \emptyset^h)T_t$

$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + \emptyset\, T_{t-1})$

$T_t = \beta(L_t - L_{t-1}) + (1-\beta)\emptyset\, T_{t-1}$

5.（A,A）模型: Holt-Winters 季节加法模型 Holt-Winters Seasonal Method

$\hat{y}_{t+h} = L_t + hT_t + S_{t+h-m(k+1)}$

$L_t = \alpha(y_t - S_{t-m}) + (1-\alpha)(L_{t-1} + T_{t-1})$

$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$

$S_t = \gamma(y_t - L_{t-1} - T_{t-1}) + (1-\gamma)S_{t-m}$

6.（A,M）模型: Holt 季节乘法模型 (Holts Method)

$\hat{y}_{t+h} = (L_t + hT_t) \times S_{t+h-m(k+1)}$

$L_t = \alpha(y_t \div S_{t-m}) + (1-\alpha)(L_{t-1} + T_{t-1})$

$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$

$S_t = \gamma(y_t \div (L_{t-1} + T_{t-1})) + (1-\gamma)S_{t-m}$

7. (A_d,M) 模型: Holt-Winters 趋势阻尼(衰减)季节乘法模型 (Holt-Winter's Seasonal Method with Damped Trend)

$$\hat{y}_{t+h} = [L_t + (\emptyset + \emptyset^2 + \cdots + \emptyset^h)T_t] \times S_{t+h-m(k+1)}$$
$$L_t = \alpha(y_t \div S_{t-m}) + (1-\alpha)(L_{t-1} + \emptyset\, T_{t-1})$$
$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)\emptyset\, T_{t-1}$$
$$S_t = \gamma(y_t \div (L_{t-1} + \emptyset\, T_{t-1})) + (1-\gamma)S_{t-m}$$

其中 y_t = 时间序列 t 期的实际值 ; $\hat{y}_{t+h}$ = 时间序列 t+h 期的预测值

L_t = t 期的平均水平 ; T_t = t 期的增长趋势 ; S_t = t 期的季节系数

α = 水平平滑参数 ; β = 趋势平滑参数 ; γ = 季节平滑参数 ;

$\emptyset$ = 阻尼参数 ; m = 季节频率 ; $k = (h-1)/m$ 的整数部分

ETS 模型

ETS 代表误差,趋势,季节(Error, Trend, Seasonal)

1. ETS(A,N,N) 模型:

$$\hat{y}_{t+h} = L_t$$
$$y_t = L_{t-1} + \epsilon_t$$
$$L_t = L_{t-1} + \alpha\epsilon_t \qquad \epsilon_t \sim N(0,\sigma^2)$$

2. ETS(M,N,N) 模型:

$$y_t = L_{t-1}(1+\epsilon_t)$$
$$L_t = L_{t-1}(1+\alpha\epsilon_t) \qquad \epsilon_t \sim N(0,\sigma^2)$$

3. ETS(A,A,A) 模型: Holt-Winters additive method with additive errors

$$\hat{y}_{t+h} = L_t + hT_t + S_{t+h-m(k+1)}$$
$$y_t = L_{t-1} + T_{t-1} + S_{t-m} + \epsilon_t$$
$$L_t = L_{t-1} + T_{t-1} + \alpha\epsilon_t$$
$$T_t = T_{t-1} + \beta\epsilon_t$$
$$S_t = S_{t-m} + \gamma\epsilon_t \qquad \epsilon_t \sim N(0,\sigma^2)$$

R 语言的 ETS 函数:

```
> forecast :: ets(y , model = "ZZZ", damped , alpha , beta , gamma , phi , … )
```

指数平滑模型与 ETS 模型的 R 语言函数

指数平滑模型 R::forecast 函数		季节因素		
		无 N	加法 A	乘法 M
趋势因素	.无 N	(N,N)模型 15.4.3 节 ses(), ets(y, "ANN") ets(y, "MNN")	(N,A)模型 ets(y, "ANA") ets(y, "MNA")	(N,M)模型 ets(y, "MNM")
	加法 A	(A,N)模型 15.5.4 节 holt(), ets(y, "AAN") ets(y, "MAN")	(A,A)模型 HoltWinters(y,) ets("AAA","MAA")	(A,M)模型 15.8.3 节 HoltWinters(y,) ets("MAM")
	阻尼加法 A_d	(Ad,N)模型 ets("ANN") holt(,damped =T), ets("MNN",damped =T)	(Ad,A)模型 hw() ets(y, "AAA") ets("MAA",damped)	(Ad,M)模型 ets("MAM",damped)
	乘法 M	(M,N)模型 ets("AMN") aTSA::Holt() ets(y, "MMN")	(M,A)模型 ets(y, "AMA") ets(y, "MMA")	(M,M)模型 ets("MMM")

自回归模型 Box-Jenkins (ARIMA) models

1. 平稳序列

平稳序列：

(1). $E(y_t) = E(y_{t-s}) = \mu_y$

(2). $V(y_t) = V(y_{t-s}) = \sigma_y^2$

(3). $Cov(y_t, y_{t-s}) = Cov(y_{t-j}, y_{t-j-s}) = \gamma_s$

以上式子对所有 t , t-s , t-j , t-j-s 都成立。

2. 白噪音

白噪音(white noise, WN) 序列 y_t 是

(1). $E(y_t) = E(y_{t-s}) = 0$

(2). $V(y_t) = V(y_{t-s}) = \sigma_y^2$

(3). $Cov(y_t, y_{t-s}) = Cov(y_{t-j}, y_{t-j-s}) = 0$

白噪音是一个平稳序列。

3. 随机游走

随机游走(random walk)序列 y_t 是：

$y_t = y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim WN(0, \sigma^2)$, ε_t 是白噪音。

4. 单位根检验

单位根检验

随机趋势 <-> 单位根 <-> 非平稳序列

5. R 语言的单位根检验

单位根检验有：ADP 检验(Augmented Dickey-Fuller test)， PP 检验(Phillips–Perron test), KPSS 检验(Kwiatkowski, Phillips, Schmidt, Shin test)。

(1). ADF, PP 检验 H_0 : H0 时间序列 Y_t 有单位根，非平稳序列。

H_1 : 时间序列 Y_t 没有单位根，平稳序列。

若要结果是平稳序列，ADF 检验结果应该是 p 值 < 0.05，拒绝零假设。

(2). KPSS 检验 H_0：时间序列 Y_t 没有单位根，平稳序列

H_1: 时间序列 Y_t 有单位根，非平稳序列

如果要得到的结果是平稳序列，则我们不要拒绝零假设，所以 KPSS 检验结果，应该是 p 值 > 0.05 。

R 语言的单位根检验：

```
> tseries :: adf.test()
> tseries :: kpss.test(x, null = "Level")
> fUnitRoots :: unitrootTest ;  fUnitRoots :: adfTest
> urca :: ur.kpss()
> aTSA :: pp.test()
```

如果单位根检验是非平稳序列，则进行差分。

6. 滞后算子差分算子

滞后算子 B 或延迟算子 L

$By_t = y_{t-1}$, $B^2 y_t = y_{t-2}$, 或 $Ly_t = y_{t-1}$, $L^2 y_t = y_{t-2}$, $L^k y_t = y_{t-\mathrm{k}}$

差分算子 $\nabla y_t = y_t - y_{t-1} = (1 - \mathrm{L})\, y_t$, $\nabla^d y_t = (1-L)^d\, y_t$

随机游走经过差分，成为白噪音平稳序列：

$$y_t = y_{t-1} + \varepsilon_t , \quad \varepsilon_t \sim WN(0, \sigma^2)$$

$$y_t - y_{t-1} = y_{t-1} + \varepsilon_t - y_{t-1}$$

$$\nabla y_t = y_{t-1} + \varepsilon_t - y_{t-1} = \varepsilon_t$$

R 语言的差分函数

```
> base :: diff(y, differences = 2)   #  二阶差分
> forecast :: ndiffs()
> urca :: ur.df(x, type = "trend", selectlags = c("AIC"))
```

自相关函数

$$\rho_k = \rho(Y_t, Y_{t+k}) = \frac{cov(Y_t, Y_{t+k})}{\sqrt{V(Y_t)V(Y_{t+k})}}$$

$$ACF(k):\ r_k = \frac{\sum_{i=1}^{n-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

偏自相关函数

$$PACF(k):\ \frac{Cov(Y_t, Y_{t-k} \mid Y_{t-1}, \dots, Y_{t-k+1})}{\sqrt{Var(Y_t \mid Y_{t-1}, \dots, Y_{t-k+1})Var(Y_{t-k} \mid Y_{t-1}, \dots, Y_{t-k+1})}}$$

Box-Jenkins 模型 ARIMA

1. 自回归模型（Autoregressive model, AR），序列值和过去值的回归

$$AR(p):\ X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_i$$

其中：c 是常数项或飘移(drift)；φ_i 是自回归参数， $\varepsilon_t \sim WN(0, \sigma^2)$ 随机误差。

2. 移动平均模型（Moving average model, MA），序列值和过去误差项的回归

$$MA(q):\ X_t = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

其中 μ 是序列的均值，θ_i 是移动平均参数， ε_{t-i} 是 白噪声。

3. 自回归移动平均模型（Autoregressive moving average model, ARMA）

$$ARMA(p, q):\ X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

ARMA 模型由两部分组成：AR 代表 p 阶自回归过程，MA 代表 q 阶移动平均过程。

4. 自回归整合移动平均模型（Autoregressive Integrated moving average model, ARIMA）

$$ARIMA(p, d, q):\ \left(1 - \sum_{i=1}^{p} \varphi_i L^i\right)(1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t$$

其中：ε_i 是 白噪声；φ_i 是自回归参数；θ_i 是移动平均参数；d 是差分；L 是滞后算子或延迟算子。

5. 季节 ARIMA 模型（Seasonal ARIMA model, SARIMA）

$$\mathrm{ARIMA}:\quad \underbrace{(\mathrm{p}, \mathrm{d}, \mathrm{q})}_{\text{非季节部分}} \qquad \underbrace{(\mathrm{P}, \mathrm{D}, \mathrm{Q})_m}_{\text{季节部分，m是年季节期数}}$$

$$\mathrm{ARIMA}\ (\mathrm{p}, \mathrm{d}, \mathrm{q})(P, D, Q)_m\ :\ \left(1 - \sum_{i=1}^{p} \varphi_i L^i\right)\left(1 - \sum_{i=1}^{P} \emptyset_i L^{im}\right)(1 - L)^d (1 - L^m)^D X_t$$

$$= \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\left(1 + \sum_{i=1}^{Q} \vartheta_i L^{im}\right)\varepsilon_t$$

其中：ε_i 是 白噪声；φ_i 是自回归参数；θ_i 是移动平均参数；d 是差分；L 是滞后算子或延迟算子。$\emptyset_i$ 是季节自回归参数；ϑ_i 是季节移动平均参数；D 是季节差分；m 是年季节期数。

第 15 章

表 15.2　商品物价及数量

商品名称	数量		单价		权数				个体指数	
	基期 q_0	报告期 q_1	基期 p_0	报告期 p_1	q_0p_0	q_1p_1	q_1p_0	q_0p_1	量比 q_1/q_0	价比 p_1/p_0
甲	50	40	\$22	\$30	1100	1200	880	1500	0.8	1.36
乙	2	3	\$20	\$20	40	60	60	40	1.5	1.0
丙	80	100	\$5	\$6	400	600	500	480	1.25	1.2
Σ	132	143	47	56	1540	1860	1440	2020	3.55	3.56

简单综合指数 Dutot 指数

物价指数 $$I_P=\frac{\sum_{i=1}^{n}\frac{p_{1i}}{n}}{\sum_{i=1}^{n}\frac{p_{0i}}{n}}=\frac{\sum_{i=1}^{n}p_{1i}}{\sum_{i=1}^{n}p_{0i}}=\frac{\sum p_1}{\sum p_0}=\frac{56}{47}=1.10$$

物量指数 $$I_Q=\frac{\sum_{i=1}^{n}\frac{q_{1i}}{n}}{\sum_{i=1}^{n}\frac{q_{0i}}{n}}=\frac{\sum_{i=1}^{n}q_{1i}}{\sum_{i=1}^{n}q_{0i}}=\frac{\sum q_1}{\sum q_0}=\frac{143}{132}=1.08$$

$q_{0i}=$ 第 i 商品在基期的数量；　$q_{1i}=$ 第 i 商品在报告期的数量
$p_{0i}=$ 第 i 商品在基期的价格；　$p_{1i}=$ 第 i 商品在报告期的价格；　$n=$ 商品的数目。

简单平均指数
简单算术平均指数 Caeli 指数：

物价指数 $$A=\frac{\sum_{i=1}^{n}\frac{p_{1i}}{p_{0i}}}{n}=\frac{\sum\frac{p_1}{p_0}}{n}=\frac{3.56}{3}=1.19$$

物量指数 $$A=\frac{\sum_{i=1}^{n}\frac{q_{1i}}{q_{0i}}}{n}=\frac{\sum\frac{q_1}{q_0}}{n}=\frac{3.55}{3}=1.18$$

简单几何平均指数 Jevons 指数：

物价指数 $$G=\sqrt[n]{\prod_{i=1}^{n}\left(\frac{p_{1i}}{p_{0i}}\right)}=\sqrt[n]{\prod\frac{p_1}{p_0}}=\sqrt[3]{1.36\times1\times1.2}=1.18$$

物量指数 $$G=\sqrt[n]{\prod_{i=1}^{n}\left(\frac{q_{1i}}{q_{0i}}\right)}=\sqrt[n]{\prod\frac{q_1}{q_0}}=\sqrt[3]{0.8\times1.5\times1.25}=1.14$$

简单调和平均指数 Harmonic 指数：

物价指数 $$H=\frac{n}{\sum_{i=1}^{n}\frac{p_{0i}}{p_{1i}}}=\frac{n}{\sum\frac{p_0}{p_1}}=\frac{3}{\frac{1}{1.36}+\frac{1}{1}+\frac{1}{1.2}}=1.17$$

物量指数 $H=\frac{n}{\sum_{i=1}^{n}\frac{q_{0i}}{q_{1i}}}=\frac{n}{\sum\frac{q_0}{q_1}}=\frac{3}{\frac{50}{40}+\frac{2}{3}+\frac{80}{100}}=1.1$

加权综合指数

1. 拉氏指数(Laspeyres index)：以基期的数量 q_0 为权数。

物价指数 $P_L=\frac{\sum p_1q_0}{\sum p_0q_0}=\frac{2020}{1540}=1.31$

物量指数 $Q_L=\frac{\sum p_0q_1}{\sum p_0q_0}=\frac{1440}{1540}=0.94$

2. 帕氏指数(Paasche index)：以报告期的数量 q_1 或单价 p_1 为权数。

物价指数 $P_P=\frac{\sum p_1q_1}{\sum p_0q_1}=\frac{1860}{1440}=1.29$

物量指数 $Q_p=\frac{\sum p_1q_1}{\sum p_1q_0}=\frac{1860}{2020}=0.92$

3. 罗氏指数(Lowe index)：以基期及报告期以外的固定期(a 期)的数量 q_a 或单价 p_a 为权数。假设三个商品 a 期的数量 q_a 分别是 45, 2.5, 90，单价 p_a 分别是 25, 20, 5.5。

物价指数 $P_{Lo}=\frac{\sum p_1q_a}{\sum p_0q_a}=\frac{30\times45+20\times2.5+6\times90}{22\times45+20\times2.5+5\times90}=\frac{1940}{1490}=1.30$

物量指数 $Q_{Lo}=\frac{\sum p_aq_1}{\sum p_aq_0}=\frac{40\times25+3\times20+100\times5.5}{50\times25+2\times20+80\times5.5}=\frac{1610}{1730}=0.93$

4. 马氏指数或称马埃指数 (Marshall-Edgeworth index)：以基期及报告期的数量的平均 $(q_0+q_1)/2$ 为权数。

物价指数 $P_M=\frac{\sum p_1(q_0+q_1)/2}{\sum p_0(q_0+q_1)/2}=\frac{\sum p_1(q_0+q_1)}{\sum p_0(q_0+q_1)}=1.30$

物量指数 $Q_M=\frac{\sum q_1(p_0+p_1)/2}{\sum q_0(p_0+p_1)/2}=\frac{\sum q_1(p_0+p_1)}{\sum q_0(p_0+p_1)}=0.93$

5. 费氏指数或称费喧指数(Fisher index)：以拉氏及帕氏两个指数，取几何平均，又称理想指数(Ideal index)。

物价指数 $P_F=\sqrt{P_L\times P_P}=\sqrt{1.31\times1.29}=1.30$

物量指数 $Q_F=\sqrt{Q_L\times Q_P}=\sqrt{0.94\times0.92}=0.93$

6. 华氏指数(Walsh index)：以基期及报告期的数量的几何平均 $\sqrt{q_0q_1}$ 为权数。

物价指数 $P_Y=\frac{\sum p_1\sqrt{q_0q_1}}{\sum p_0\sqrt{q_0q_1}}=\frac{30\sqrt{50\cdot40}+20\sqrt{2\cdot3}+6\sqrt{80\cdot100}}{22\sqrt{50\cdot40}+20\sqrt{2\cdot3}+5\sqrt{80\cdot100}}=\frac{1927.24}{1480.04}=1.30$

物量指数 $Q_Y=\frac{\sum q_1\sqrt{p_0p_1}}{\sum q_0\sqrt{p_0p_1}}=\frac{40\sqrt{22\cdot30}+3\sqrt{20\cdot20}+100\sqrt{5\cdot6}}{50\sqrt{22\cdot30}+2\sqrt{20\cdot20}+80\sqrt{5\cdot6}}=\frac{1635.30}{1762.66}=0.93$

加权平均指数

1. 加权平均指数：以固定期 p_0q_0 当权数

物价指数 $P_A=\frac{\sum(\frac{p_1}{p_0})p_0q_0}{\sum p_0q_0}=P_L$ 物量指数 $Q_A=\frac{\sum(\frac{q_1}{q_0})p_0q_0}{\sum p_0q_0}=Q_L$

2. 杨格指数(Young index)：以固定期 $p_a q_a$ 当权数：

假设三个商品 a 期的数量 q_a 分别是 45, 2.5, 90，单价 p_a 分别是 25, 20, 5.5。

物价指数 $P_Y = \dfrac{\sum(\frac{p_1}{p_0})p_a q_a}{\sum p_a q_a} = \dfrac{1.36\times45\times25+1\times2.5\times20+1.2\times90\times5.5}{45\times25+2.5\times20+90\times5.5} = \dfrac{2174}{1670} = 1.3$

物量指数 $Q_Y = \dfrac{\sum(\frac{q_1}{q_0})p_a q_a}{\sum p_a q_a} = \dfrac{0.8\times45\times25+1.5\times2.5\times20+1.25\times90\times5.5}{45\times25+2.5\times20+90\times5.5} = 0.95$

3. 加权几何平均指数：以固定期 $p_0 q_0$ 当权数

$$P_G = \sqrt[\sum p_0 q_0]{\prod\left(\frac{p_1}{p_0}\right)^{p_0 q_0}} \qquad \log P_G = \frac{\sum[p_0 q_0(\log p_1 - \log p_0)]}{\sum p_0 q_0}$$

物值指数

物值指数没有加权指数，只有简单指数，简单综合指数的公式是：

物值指数 ＝ 当期物值摠和 ÷ 基期物值总和