地统计系列

地理探测器

-2-

q统计量的推导

UP：小勇啊哈

q统计量的推导

一、q统计量的推导

假设层h中 $Y_{hi} \sim N(\mu_h, \sigma^2)$, 各层各单元的分布相互独立，则$\overline{Y_h} \sim N(\mu_h, \sigma^2/N_h)$

全区方差和: $$SST = \sum_{h}^{L}\sum_{i=1}^{N_h}(Y_{hi} - \bar{Y})^2$$

层内方差和: $$SSW = \sum_{h}^{L}\sum_{i=1}^{N_h}(Y_{hi} - \overline{Y_h})^2$$

层间方差和: $$SSB = \sum_{h}^{L}\sum_{i=1}^{N_h}(\overline{Y_h} - \bar{Y})^2$$

—L为子区域数

—N为总单元数

—N_h为第h个子区域/层h单元数

—Y_{hi}为层h第i个单元的值

—$\overline{Y_h}$为层h的均值

—$\bar{Y}$为全区的均值

1 $SST = SSW + SSB$

$$SST = \sum_{h}^{L}\sum_{i=1}^{N_h}(Y_{hi} - \bar{Y})^2 = \sum_{h}^{L}\sum_{i=1}^{N_h}(Y_{hi} - \overline{Y_h} + \overline{Y_h} - \bar{Y})^2$$

$$= \sum_{h}^{L}\sum_{i=1}^{N_h}(Y_{hi} - \overline{Y_h})^2 + \sum_{h}^{L}\sum_{i=1}^{N_h}(\overline{Y_h} - \bar{Y})^2 + 2\sum_{h}^{L}(\overline{Y_h} - \bar{Y})\sum_{i=1}^{N_h}(Y_{hi} - \overline{Y_h}) = SSW + SSB$$

一、q统计量的推导

空间分异性度量

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2} = 1 - \frac{SSW}{SST}$$

$$= 1 - \frac{SSW}{SSW + SSB}$$

1

$$= 1 - \frac{1}{1 + F(L-1)/(N-L)}$$

—σ_h^2为第h个子区域的方差

—σ^2为全区的方差

检验q值显著性

4

$$F = \frac{SSB/(L-1)}{SSW/(N-L)}$$

$$= \frac{N-L}{L-1}\frac{q}{1-q} \sim F(L-1, N-L; \lambda)$$

$$\lambda = \frac{1}{\sigma^2}\left[\sum_{h=1}^{L} N_h \overline{Y_h}^2 - \frac{1}{N}\left(\sum_{h=1}^{L} N_h \overline{Y_h}\right)^2\right]$$

—λ为非中心F分布的非中心参数

2

$$SSW = \sum_{h}^{L}\sum_{i=1}^{N_h}(Y_{hi} - \overline{Y_h})^2 \sim \sigma^2 \chi(N-L)$$

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

3

$$SSB = \sum_{h}^{L}\sum_{i=1}^{N_h}(\overline{Y_h} - \overline{Y})^2 \sim \sigma^2 \chi(L-1; \lambda)$$

一、q统计量的推导

5

随机变量$Y_1 \sim Y_n$相互独立，且服从标准正态分布，设$Y = (Y_1, Y_2, \dots, Y_n)^T$,$Q$为正交矩阵，$Z = QY = (Z_1, Z_2, \dots, Z_n)^T$，则$Z_1, Z_2, \dots, Z_n$独立且服从标准正态分布。

设$Q = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{1n} \end{bmatrix}$，因为$Q$是正交矩阵，所以

$$\sum_{k=1}^{n} {a_{ik}}^2 = 1 \qquad \sum_{k=1}^{n} a_{ik} a_{jk} = 0$$

$$Z = QY, Z_i = a_{i1}Y_1 + a_{i2}Y_2 + \cdots + a_{in}Y_n$$

$$E(Z_i) = \sum_{k=1}^{n} a_{ik} E(Y_k) = 0 \qquad D(Z_i) = \sum_{k=1}^{n} {a_{ik}}^2 D(Y_k) = 1$$

所以Z_i服从**标准正态分布**

一、q统计量的推导

5

随机变量$Y_1 \sim Y_n$相互独立，且服从标准正态分布，设$Y=(Y_1,Y_2,\dots,Y_n)^T$,Q为正交矩阵，$Z=QY=(Z_1,Z_2,\dots,Z_n)^T$，则$Z_1,Z_2,\dots,Z_n$独立且服从标准正态分布。

$$Cov\left(Z_i,Z_j\right)=\sum_{t=1}^{n}\sum_{m=1}^{n}a_{it}a_{jm}Cov(Y_t,Y_m)=\sum_{k=1}^{n}a_{ik}a_{jk}Cov\left(Y_k,Y_k\right)=0$$

对于任意满足正态分布的随机变量，独立和不相关等价。

所以$Z_1,Z_2,\dots,Z_n$独立。

一、q统计量的推导

2

$$\sum_{i=1}^{N_h}(Y_{hi}-\overline{Y_h})^2\sim\sigma^2\chi(N_h-1) \qquad SSW=\sum_{h}^{L}\sum_{i=1}^{N_h}(Y_{hi}-\overline{Y_h})^2\sim\sigma^2\chi(N-L)$$

随机变量$Y_1\sim Y_n$相互独立，且服从标准正态分布 $Y_i\sim N(0,1),\bar{Y}=\frac{\bar{Y}-\mu}{\sigma}$

$$\sum_{i=1}^{n}(Y_i-\bar{Y})^2=\sum_{i=1}^{n}(Y_i^2+\bar{Y}^2-2Y_i\bar{Y})=\sum_{i=1}^{n}Y_i^2+\frac{1}{n}\left(\sum_{i=1}^{n}Y_i\right)^2-\frac{2}{n}\left(\sum_{i=1}^{n}Y_i\right)^2$$

$$=\sum_{i=1}^{n}Y_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}Y_i\right)^2=\left(1-\frac{1}{n}\right)\sum_{i=1}^{n}Y_i^2-\frac{2}{n}\sum_{1\le i\le j\le n}Y_iY_j$$

此为二次型，易得 $\sum_{i=1}^{n}(Y_i-\bar{Y})^2=Y^TAY$

其中 $A=\begin{bmatrix}1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n}\\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n}\\ \cdots & \cdots & \cdots & \cdots\\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n}\end{bmatrix},Y=(Y_1,Y_2,\ldots,Y_n)^T$

一、q统计量的推导

2

$$\sum_{i=1}^{N_h}(Y_{hi}-\overline{Y_h})^2\sim\sigma^2\chi(N_h-1) \qquad SSW=\sum_{h}^{L}\sum_{i=1}^{N_h}(Y_{hi}-\overline{Y_h})^2\sim\sigma^2\chi(N-L)$$

由$|\lambda E-A|=0$得，$(\lambda-1)^{n-1}\lambda=0, \lambda_1=\lambda_2=\cdots=\lambda_{n-1}=1, \lambda_n=0$

且A为实对称矩阵(实对称矩阵的正交对角化),则一定存在正交矩阵Q，有

$$Q^TAQ=\begin{bmatrix}1 & 0 & \dots & 0 & 0\\ 0 & 1 & \dots & 0 & 0\\ \dots & \dots & \dots & \dots & \dots\\ 0 & 0 & \dots & 1 & 0\\ 0 & 0 & \dots & 0 & 0\end{bmatrix}=QAQ^T$$

令$Z=QY=(Z_1,Z_2,\dots,Z_n)^T$，$Y=Q^TZ$

$Y^TAY=(Q^TZ)^TAQ^TZ=Z^TQAQ^TZ={Z_1}^2+{Z_2}^2+\cdots+{Z_{n-1}}^2\sim\chi(n-1)$

$$\sum_{i=1}^{N_h}\left(\frac{(Y_{hi}-\mu)}{\sigma}-\frac{(\overline{Y_h}-\mu)}{\sigma}\right)^2\sim\chi(N_h-1) \qquad SSW=\sum_{h}^{L}\sum_{i=1}^{N_h}(Y_{hi}-\overline{Y_h})^2\sim\sigma^2\chi(N-L)$$

一、q统计量的推导

小勇啊哈 bilibili

3

$$SSB=\sum_{h}^{L}\sum_{i=1}^{N_h}(\overline{Y_h}-\bar{Y})^2\sim\sigma^2\chi(L-1;\lambda)\quad \lambda=\frac{1}{\sigma^2}\left[\sum_{h}^{L}N_h{\mu_h}^2-\frac{1}{N}\left(\sum_{h}^{L}N_h\mu_h\right)^2\right]$$

$$SSB=\sum_{h}^{L}\sum_{i=1}^{N_h}(\overline{Y_h}-\bar{Y})^2=\sum_{h}^{L}N_h(\overline{Y_h}-\bar{Y})^2$$

令 $\quad \tilde{Y}=(\sqrt{N_1}\overline{Y_1},\sqrt{N_2}\overline{Y_2},\dots,\sqrt{N_L}\overline{Y_L})^T$

$$\tilde{N}=(\sqrt{N_1},\sqrt{N_2},\dots,\sqrt{N_L})^T$$

$$B=diag(1,1,\dots 1)-\frac{\tilde{N}\tilde{N}^T}{N}$$

易得B是对称矩阵，且$B^TB=B$

$$B^TB=\left(E-\frac{\tilde{N}\tilde{N}^T}{N}\right)^T\left(E-\frac{\tilde{N}\tilde{N}^T}{N}\right)=E-\frac{\tilde{N}\tilde{N}^T}{N}=B$$

$$SSB=\sum_{h}^{L}N_h(\overline{Y_h}-\bar{Y})^2=\left(B\tilde{Y}\right)^TB\tilde{Y}=\tilde{Y}^TB^TB\tilde{Y}=\tilde{Y}^TB\tilde{Y}$$

一、q统计量的推导

3

$$SSB=\sum_{h}^{L}\sum_{i=1}^{N_h}(\overline{Y_h}-\overline{Y})^2 \sim \sigma^2\chi(L-1;\lambda) \quad \lambda=\frac{1}{\sigma^2}\left[\sum_{h}^{L}N_h{\mu_h}^2-\frac{1}{N}\left(\sum_{h}^{L}N_h\mu_h\right)^2\right]$$

由$|\lambda E-B|=0$得，$(\lambda-1)^{n-1}\lambda=0,\lambda_1=\lambda_2=\cdots=\lambda_{n-1}=1,\lambda_n=0$

且B为实对称矩阵(实对称矩阵的正交对角化),则一定存在正交矩阵Q，有

$$Q^TBQ=\begin{bmatrix}1&0&\dots&0&0\\0&1&\dots&0&0\\\dots&\dots&\dots&\dots&\dots\\0&0&\dots&1&0\\0&0&\dots&0&0\end{bmatrix}=QBQ^T$$

因为$\tilde{Y}_h\sim N(\sqrt{N_h}\mu_h,\sigma^2)$,令$T_h=\frac{\tilde{Y}_h}{\sigma}\sim N\left(\frac{\sqrt{N_h\mu_h}}{\sigma},1\right)$,$T=(T_1,T_2,\dots,T_L)^T$

令$Z=QT=(Z_1,Z_2,\dots,Z_L)^T$，$T=Q^TZ$

一、q统计量的推导

3

$$SSB=\sum_{h}^{L}\sum_{i=1}^{N_h}(\overline{Y_h}-\bar{Y})^2\sim\sigma^2\chi(L-1;\lambda)\qquad\lambda=\frac{1}{\sigma^2}\left[\sum_{h}^{L}N_h{\mu_h}^2-\frac{1}{N}\left(\sum_{h}^{L}N_h\mu_h\right)^2\right]$$

令

$$Q=\begin{bmatrix}a_{11}&\cdots&a_{1n}\\ \cdots&\cdots&\cdots\\ a_{n1}&\cdots&a_{1n}\end{bmatrix}$$

所以

$$Z_h=\sum_{k=1}^{L}a_{hk}T_k$$

$$E(Z_h)=E(\sum_{k=1}^{L}a_{hk}T_k)\qquad D(Z_h)=1\qquad Cov(Z_i,Z_j)=0$$

所以$Z_1,Z_2,\dots,Z_L$是期望不一致的，方差为1，相互独立的正态分布

$$SSB=\tilde{Y}^TB\tilde{Y}=\sigma^2T^TBT=\sigma^2(Q^TZ)^TB(Q^TZ)=\sigma^2Z^TQBQ^TZ$$
$$=\sigma^2\left({Z_1}^2+{Z_2}^2+\cdots+{Z_{L-1}}^2\right)$$

一、q统计量的推导

3

$$SSB=\sum_{h}^{L}\sum_{i=1}^{N_h}(\overline{Y_h}-\bar{Y})^2\sim\sigma^2\chi(L-1;\lambda)\qquad \lambda=\frac{1}{\sigma^2}\left[\sum_{h}^{L}N_h{\mu_h}^2-\frac{1}{N}\left(\sum_{h}^{L}N_h\mu_h\right)^2\right]$$

假如有L个服从方差为1的正态分布的随机变量，分别是$Z_h\sim N(\mu_h,1)$，这些随机变量的平方和的分布服从非中心卡方分布，即

$$ {Z_1}^2+{Z_2}^2+\cdots+{Z_L}^2\sim\chi(L-1;\lambda)\qquad \lambda=\sum_{h=1}^{L}{\mu_h}^2$$

所以 $SSB\sim\sigma^2\chi(L-1;\lambda)$

λ 怎么求呢？直接硬求可不好求

$$\lambda=\sum_{h=1}^{L-1}E(Z_i)^2=E(Z^T)QBQ^TE(Z)=E\left((Q^TZ)^T\right)BE(Q^TZ)=E(T^T)BE(T)$$

$$=\frac{1}{\sigma^2}E\left(\tilde{Y}^T\right)BE\left(\tilde{Y}\right)=\frac{1}{\sigma^2}E\left(\tilde{Y}^T\right)B^TBE\left(\tilde{Y}\right)$$

一、q统计量的推导

3

$$SSB=\sum_{h}^{L}\sum_{i=1}^{N_h}(\overline{Y_h}-\bar{Y})^2\sim\sigma^2\chi(L-1;\lambda)\qquad \lambda=\frac{1}{\sigma^2}\left[\sum_{h}^{L}N_h{\mu_h}^2-\frac{1}{N}\left(\sum_{h}^{L}N_h\mu_h\right)^2\right]$$

假如有L个服从方差为1的正态分布的随机变量，分别是$Z_h\sim N(\mu_h,1)$，这些随机变量的平方和的分布服从非中心卡方分布，即

$$ {Z_1}^2+{Z_2}^2+\cdots+{Z_L}^2\sim\chi(L-1;\lambda)\qquad \lambda=\sum_{h=1}^{L}{\mu_h}^2$$

$$E(B\tilde{Y})^TE(B\tilde{Y})=E\left(\tilde{Y}-\frac{\tilde{N}\tilde{N}^T}{N}\tilde{Y}\right)^TE(\tilde{Y}-\frac{\tilde{N}\tilde{N}^T}{N}\tilde{Y})$$

$$=E(\tilde{Y})^TE(\tilde{Y})-2E(\tilde{Y})^T\frac{\tilde{N}\tilde{N}^T}{N}E(\tilde{Y})+E(\tilde{Y})^T\frac{\tilde{N}\tilde{N}^T}{N}\frac{\tilde{N}\tilde{N}^T}{N}E(\tilde{Y})$$

$$=E(\tilde{Y})^TE(\tilde{Y})-E(\tilde{Y})^T\frac{\tilde{N}\tilde{N}^T}{N}E(\tilde{Y})=\sum_{h}^{L}N_h{\mu_h}^2-\frac{1}{N}(\sum_{h}^{L}N_h\mu_h)^2$$

所以 $$\lambda=\frac{1}{\sigma^2}\left[\sum_{h}^{L}N_h{\mu_h}^2-\frac{1}{N}\left(\sum_{h}^{L}N_h\mu_h\right)^2\right]$$

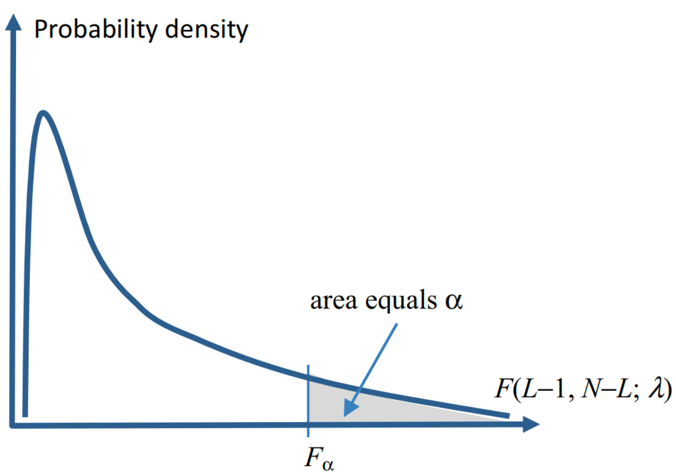
一、q统计量的推导

4

$$F=\frac{N-L}{L-1}\frac{q}{1-q}\sim F(L-1,N-L;\lambda) \quad \lambda=\frac{1}{\sigma^2}[\sum_{h=1}^{L}N_h\overline{Y_h}^2-\frac{1}{N}(\sum_{h=1}^{L}N_h\overline{Y_h})^2]$$

因为 $SSB\sim\sigma^2\chi(L-1;\lambda)$ $SSW\sim\sigma^2\chi(N-L)$

所以 $$F=\frac{\frac{SSB}{(L-1)}}{\frac{SSW}{(N-L)}}\sim F(L-1,N-L;\lambda) \quad \lambda=\frac{1}{\sigma^2}[\sum_{h=1}^{L}N_h\overline{Y_h}^2-\frac{1}{N}(\sum_{h=1}^{L}N_h\overline{Y_h})^2]$$



检验在置信度α下q统计量的显著性：

$$P(q<x)=P\left(F<\frac{N-L}{L-1}\frac{x}{1-x}\right)=1-\alpha$$

置信度α可以理解为$q\geq x$的概率

Thanks