

# 统计方法选用手册，妈妈再也不担心我的选择困难症了！

2017-07-13 量化研究方法

## 一、 两组或多组计量资料的比较

### 1.两组资料：

1)大样本资料或服从正态分布的小样本资料

- 若方差齐性，则作成组t检验
- 若方差不齐，则作t' 检验或用成组的Wilcoxon秩和检验

2)小样本偏态分布资料，则用成组的Wilcoxon秩和检验

### 2.多组资料：

1)若大样本资料或服从正态分布，并且方差齐性，则作完全随机的方差分析。如果方差分析的统计检验为有统计学意义，则进一步作统计分析：选择合适的方法（如：LSD检验，Bonferroni检验等）进行两两比较。

2)如果小样本的偏态分布资料或方差不齐，则作Kruskal Wallis的统计检验。如果Kruskal Wallis的统计检验为有统计学意义，则进一步作统计分析：选择合适的方法（如：用成组的Wilcoxon秩和检验，但用Bonferroni方法校正P值等）进行两两比较。

## 二、 分类资料的统计分析

### 1.单样本资料与总体比较

1)二分类资料：

- 小样本时：用二项分布进行确切概率法检验；
- 大样本时：用U检验。

2)多分类资料：用Pearson  $\chi^2$ 检验（又称拟合优度检验）。

### 2. 四格表资料

1) $n > 40$ 并且所有理论数大于5，则用Pearson  $\chi^2$

2) $n > 40$ 并且所有理论数大于1并且至少存在一个理论数 $< 5$ ，则用校正  $\chi^2$ 或用Fisher's确切概率法检验

3) $n \leq 40$ 或存在理论数 $< 1$ ，则用Fisher's 检验

### 3. $2 \times C$ 表资料的统计分析

1)列变量为效应指标，并且为有序多分类变量，行变量为分组变量，则行评分的CMH  $\chi^2$ 或成组的Wilcoxon秩和检验

2)列变量为效应指标并且为二分类，行变量为有序多分类变量，则用趋势 $\chi^2$ 检验

3)行变量和列变量均为无序分类变量

- $n > 40$ 并且理论数小于5的格子数 < 行列表中格子总数的25%，则用 Pearson  $\chi^2$
- $n \leq 40$ 或理论数小于5的格子数 > 行列表中格子总数的25%，则用 Fisher's 确切概率法检验

#### 4. R×C表资料的统计分析

1)列变量为效应指标，并且为有序多分类变量，行变量为分组变量，则CMH  $\chi^2$ 或Kruskal Wallis的秩和检验

2)列变量为效应指标，并且为无序多分类变量，行变量为有序多分类变量，作none zero correlation analysis的CMH  $\chi^2$

3)列变量和行变量均为有序多分类变量，可以作Spearman相关分析

4)列变量和行变量均为无序多分类变量，

- $n > 40$ 并且理论数小于5的格子数 < 行列表中格子总数的25%，则用 Pearson  $\chi^2$
- $n \leq 40$ 或理论数小于5的格子数 > 行列表中格子总数的25%，则用 Fisher's 确切概率法检验

### 三、Poisson分布资料

### 1.单样本资料与总体比较：

1)观察值较小时：用确切概率法进行检验。

2)观察值较大时：用正态近似的U检验。

### 2.两个样本比较：用正态近似的U检验。

配对设计或随机区组设计

## 四、 两组或多组计量资料的比较

### 1.两组资料：

1)大样本资料或配对差值服从正态分布的小样本资料，作配对t检验

2)小样本并且差值呈偏态分布资料，则用Wilcoxon的符号配对秩检验

### 2.多组资料：

1)若大样本资料或残差服从正态分布，并且方差齐性，则作随机区组的方差分析。如果方差分析的统计检验为有统计学意义，则进一步作统计分析：选择合适的方法（如：LSD检验，Bonferroni检验等）进行两两比较。

2)如果小样本时，差值呈偏态分布资料或方差不齐，则作Fredman的统计检验。如果Fredman的统计检验为有统计学意义，则进一步作统计分析：选择合适的方法（如：用Wilcoxon的符号配对秩检验，但用Bonferroni方法校正P值等）进行两两比较。

## 五、 分类资料的统计分析

### 1.四格表资料

1) $b+c > 40$ ，则用McNemar配对  $\chi^2$ 检验或配对边际 $\chi^2$ 检验

2) $b+c \leq 40$ ，则用二项分布确切概率法检验

### 2.C×C表资料：

1)配对比较：用McNemar配对  $\chi^2$ 检验或配对边际 $\chi^2$ 检验

2)一致性问题（Agreement）：用Kap检验

变量之间的关联性分析

## 六、 两个变量之间的关联性分析

### 1.两个变量均为连续型变量

1)小样本并且两个变量服从双正态分布，则用Pearson相关系数做统计分析

2)大样本或两个变量不服从双正态分布，则用Spearman相关系数进行统计分析

### 2.两个变量均为有序分类变量，可以用Spearman相关系数进行统计分析

### 3.一个变量为有序分类变量，另一个变量为连续型变量，可以用Spearman相关系数进行统计分析

## 七、 回归分析

**1.直线回归：**如果回归分析中的残差服从正态分布（大样本时无需正态性），残差与自变量无趋势变化，则直线回归（单个自变量的线性回归，称为简单回归），否则应作适当的变换，使其满足上述条件。

**2.多重线性回归：**应变量（Y）为连续型变量（即计量资料），自变量（ $X_1, X_2, \dots, X_p$ ）可以为连续型变量、有序分类变量或二分类变量。如果回归分析中的残差服从正态分布（大样本时无需正态性），残差与自变量无趋势变化，可以作多重线性回归。

1)观察性研究：可以用逐步线性回归寻找（拟）主要的影响因素

2)实验性研究：在保持主要研究因素变量（干预变量）外，可以适当地引入一些其它可能的混杂因素变量，以校正这些混杂因素对结果的混杂作用

**3.二分类的Logistic回归：**应变量为二分类变量，自变量（ $X_1, X_2, \dots, X_p$ ）可以为连续型变量、有序分类变量或二分类变量。

1)非配对的情况：用非条件Logistic回归

- 观察性研究：可以用逐步线性回归寻找（拟）主要的影响因素
- 实验性研究：在保持主要研究因素变量（干预变量）外，可以适当地引入一些其它可能的混杂因素变量，以校正这些混杂因素对结果的混杂作用

2)配对的情况：用条件Logistic回归

- 观察性研究：可以用逐步线性回归寻找（拟）主要的影响因素
- 实验性研究：在保持主要研究因素变量（干预变量）外，可以适当地引入一些其它可能的混杂因素变量，以校正这些混杂因素对结果的混杂作用

**4.有序多分类有序的Logistic回归：**应变量为有序多分类变量，自变量（ $X_1, X_2, \dots, X_p$ ）可以为连续型变量、有序分类变量或二分类变量。

1)观察性研究：可以用逐步线性回归寻找（拟）主要的影响因素

2)实验性研究：在保持主要研究因素变量（干预变量）外，可以适当地引入一些其它可能的混杂因素变量，以校正这些混杂因素对结果的混杂作用

**5.无序多分类有序的Logistic回归：**应变量为无序多分类变量，自变量（ $X_1, X_2, \dots, X_p$ ）可以为连续型变量、有序分类变量或二分类变量。

1)观察性研究：可以用逐步线性回归寻找（拟）主要的影响因素

2)实验性研究：在保持主要研究因素变量（干预变量）外，可以适当地引入一些其它可能的混杂因素变量，以校正这些混杂因素对结果的混杂作用

## **八、生存分析资：要求资料记录结局和结局发生的时间**

（如；死亡和死亡发生的时间）

### **1.用Kaplan-Meier方法估计生存曲线**

### **2.大样本时，可以寿命表方法估计**

### **3.单因素可以用Log - rank比较两条或多条生存曲线**

### **4.多个因素时，可以作多重的Cox回归**

1)观察性研究：可以用逐步线性回归寻找（拟）主要的影响因素

2)实验性研究：在保持主要研究因素变量（干预变量）外，可以适当地引入一些其它可能的混杂因素变量，以校正这些混杂因素对结果的混杂作用