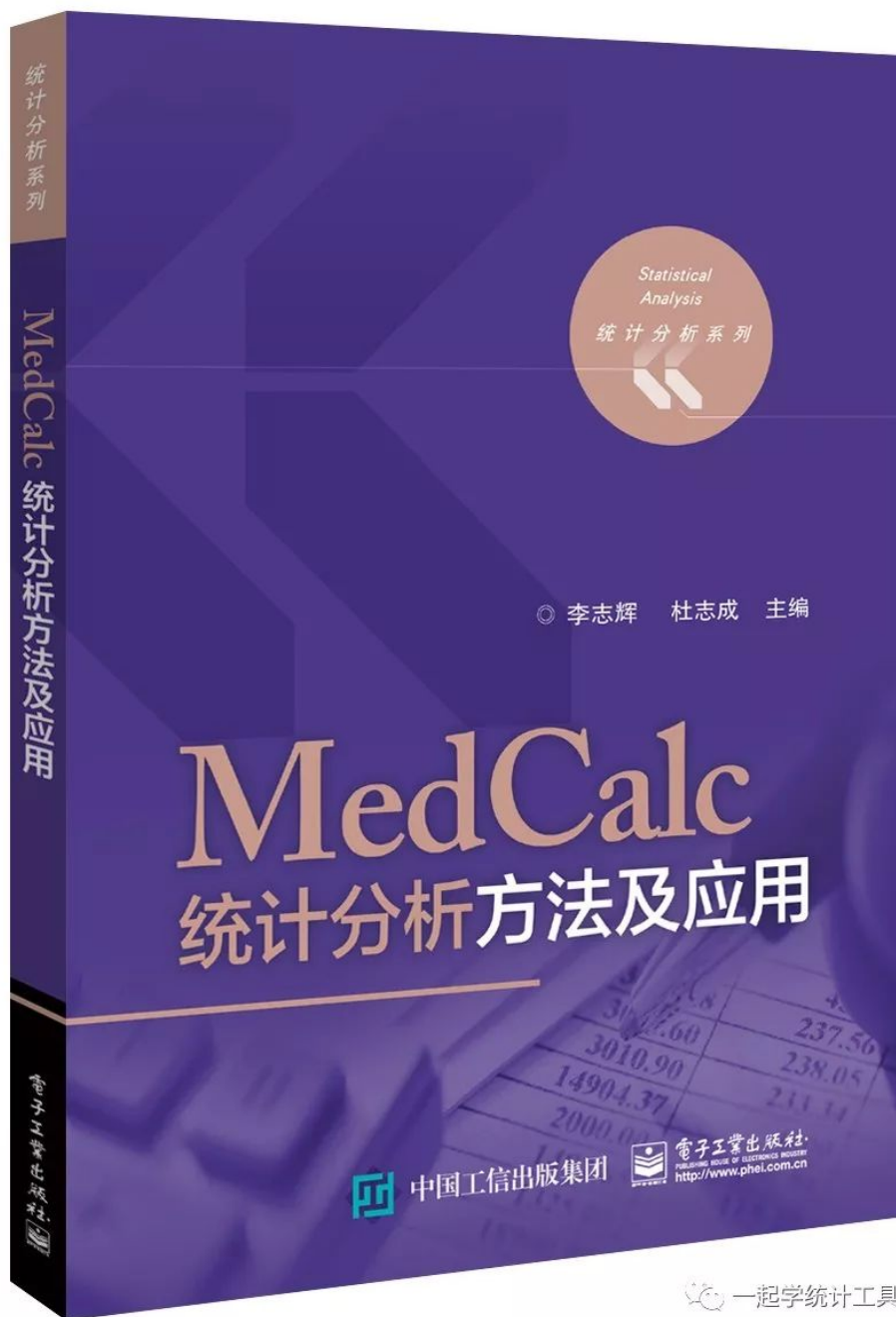

利用MedCalc进行样本量估计

作者：李欣 李志辉 来源：一起学统计工具

本文原地址：<https://www.iikx.com/news/statistics/8410.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

众所周知，PASS是最权威的功能最强大，方法最多(数百种统计方法)的功效和样本含量计算工具，该软件除了可以计算功效和样本含量外，还可以计算各种效应量、置信度或显著性等。MedCalc和Minitab一样，同样也有常用方法的样本含量计算功能，不同的是Minitab可以计算功效、样本含量和效应量，MedCalc只能计算样本含量一个参数。相同的是两者的界面都非常简单，极易上手。



一起学统计工具

17.1 样本含量估计概述

样本含量(Sample Size)又称样本量或样本大小，是指在抽样研究中每个样本所包含的调查对象数或受试对象数。样本含量是试(实)验设计中重复性原则的体现，是试(实)验设计中研究工作者须重点考虑的问题，直接关系到研究工作的规模大小，涉及研究经费、人力物力和时间的安排。一般来说，样本含量过少，所得指标不够稳定，实验结果缺少可靠性，结论缺乏充分依据；样本含量过大，则会增加实际研究的难度，延长研究周期，亦造成时间、人力、经济上不必要的浪费。鉴此，在保证研究结论具有一定可靠性的条件下，对样本含量进行合理估算，确定最少的试验单位数尤为重要。

通过功效和样本含量估计，可以帮助我们在设计实(试)验方案并执行之前(前瞻性研究)，或执行之后(回顾性研究)评估检验功效和样本含量。

- 前瞻性研究：应用在试验资料收集之前以充分考虑研究设计的灵敏度。研究者需要确保研究有足够大的检验功效，能够将你确定为重要变量的差值(效应)检测出来。例如，可以通过增大样本含量或采取措施控制容许误差来提高设计的灵敏度。
- 回顾性研究：应用在试验资料收集之后帮助了解已完成试验的检验功效。例如，假设开展了一项研究，但资料分析显示研究中并没有任何统计学差异的结果。你可以基于期望试验检测到的变量的最小差值(效应)计算出检验功效。如果检验功效较低，那么可能需要修改试验设计后继续评估计算。但如果计算得到的检验功效很高，那么你可能要断定研究中确实不存在有意义的差值(效应)，并终止试验。

通常来说，样本含量大小与检验水准 α 、检验功效 $(1 - \beta)$ 、容许误差 δ 、单双侧检验以及总体的相关信息密切相关。

1. 检验水准

检验水准 α 是指在假设检验中出现第I类错误的可能性，即当相比较的两个事物间原本并不存在差异时，通过假设检验检验出两者存在差异的可能性，一般用概率来度量。用公式表示为第I类错误的概率 $= P(\text{拒绝 } H_0 \mid H_0 \text{ 为真}) \leq \alpha$ 。在假设检验中， α 的数值是事先确定的，一般取 $\alpha = 0.05$ 或 $\alpha = 0.01$ (偶尔可取 0.10)，但 α 的数值不能主观随意确定，必须根据第I类错误的危害性来决定。

例如在做新药与常规药物疗效比较时，第I类错误会将原本等效的新药不恰当地奉为创新，而将原本市场上的常规药物摒弃掉，造成错误和浪费。当然，要想彻底消灭第I类错误是不可能做到的，因此必须结合研究实际情况确定合适的 α 值大小。在样本含量估计中，要求 α 值越小，研究所需的样本含量越大。

2. 检验功效

检验功效 $(1 - \beta)$ 是指在特定的 α 水准下，若总体间确实存在差异，该次研究能发现此差异的能力。 $(1 - \beta)$ 越大，假设检验的效能越高。根据样本假设检验得到的推论结合真实情况，假设检验有四种可能的结果。检验的功效是指当 H_0 为假时准确否定它的概率。这四种可能的结果总结如下：

假设检验的结论	真实情况	
	H_0 为真	H_0 为假
拒绝 H_0	I类错误, $P=\alpha$	正确决策, $p=1-\beta$
不拒绝 H_0	正确决策	II类错误, $P=\beta$

如果假设检验 H_0 实际上是正确的，但由样本资料计算获得的检验统计量得出拒绝 H_0 的结论，统计学上称此类情况为发生了I类错误。发生I类错误的概率称为 α (alpha)，即检验水准。

如果假设检验 H_0 为假，但由样本资料计算获得的检验统计量得出不拒绝 H_0 的结论，就发生了II类错误。犯II类错误的概率称为 β (beta)。

在科研设计时只有确定 β 的数值才能进一步估算样本含量大小。通常取 $\beta=0.1$ 或 $\beta=0.2$ ，相应检验功效为0.9或0.8，一般检验功效最低不宜低于0.75，否则可能出现非真实的阴性结果。在样本含量估计中，要求 $(1-\beta)$ 值越大，研究所需的样本含量越大。

3.差值或容许误差

在实际应用中，差值指样本与总体、两样本之间平均值或率指标之差等。在样本含量大小的计算中，需要根据专业知识判断或预实验结果来预计样本统计量(或 p)与相应总体参数(μ 或 π)的最大差异控制在什么范围，即确定容许误差大小 δ 。

在科研设计中，容许误差和检验功效的大小成正比，而与样本统计量的准确性要求成反比。即，科研设计中要求的容许误差愈小，即调查结果准确性愈高，则把握程度愈低;反之，容许误差范围增大，样本统计量的准确性减小，把握程度随之增大。因此，要求容许误差 δ 愈小，所需样本含量愈大。

4.单、双侧检验

在实际研究工作中何时使用单侧检验，何时使用双侧检验要根据研究的目的、资料性质及样本特征来确定。例如，为了研究某新药的减肥效果，欲评价新药减重的效果和传统药物的减重效果是否相同时，可以使用独立样本双侧t检验的样本含量估计来检测具有足够功效的差异需要多少样本;而在实际生产中，工艺改革后，我们希望能新药的平均减重 μ 不能低于传统药物，因此可将研究目的改为“新药的平均减重 μ 不低于传统药物”，此时，使用独立样本单侧t检验的样本含量估计来检测具有足够功效的差异需要多少样本。同一实验中，用单侧检验比用双侧检验所需例数少。但选用单侧检验需要谨慎，在专业上必须有充分的理由。

5.总体的相关信息

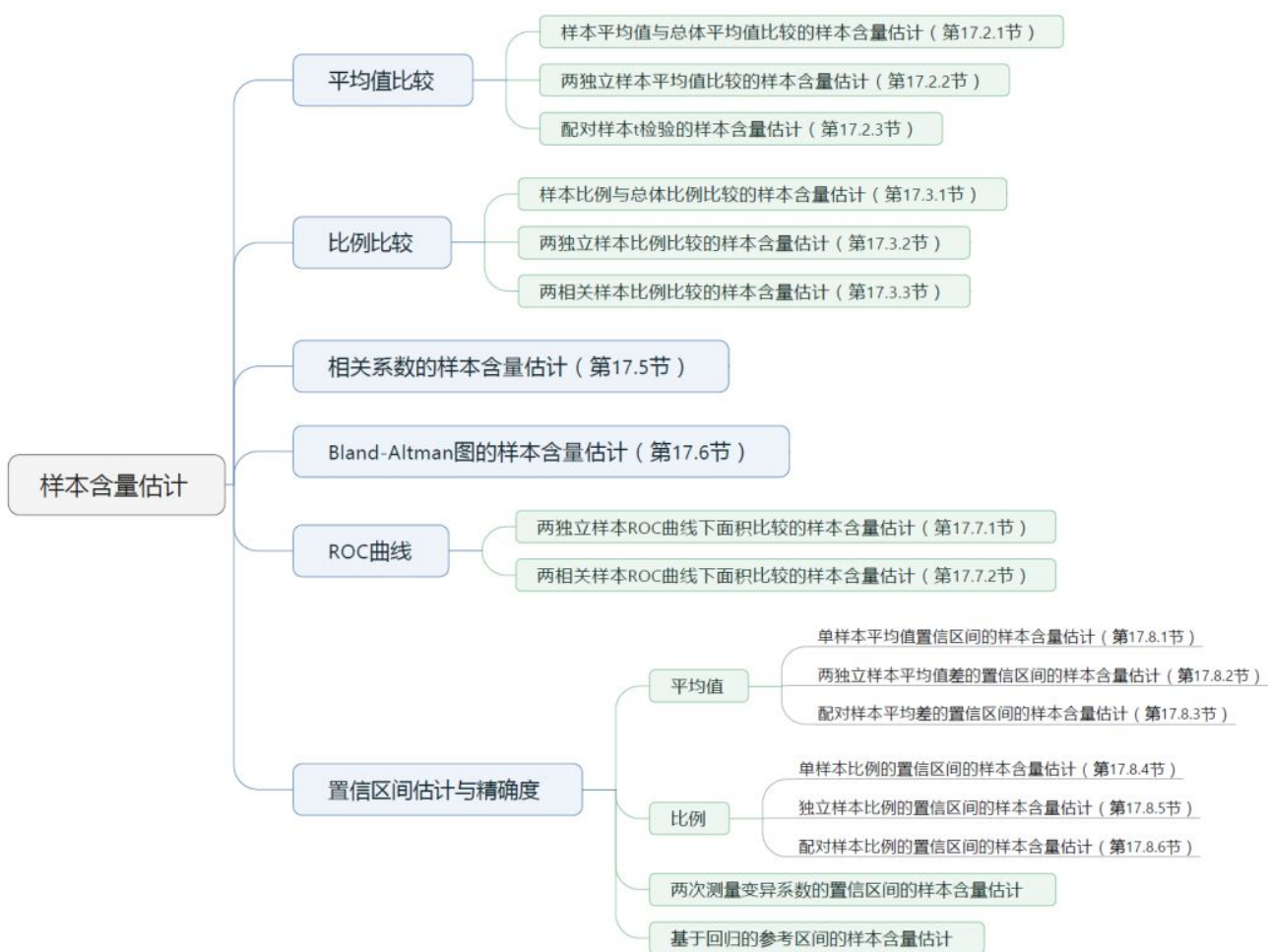
进行平均值比较时，需了解个体变异大小即总体标准差 σ ，率的比较需了解总体率 π 的大小，相关分析时需了解总体相关系数 ρ 的大小。 σ 越大，所需样本含量越多;总体率 π 越接近于0.50，则

所需样本含量越多; ρ 越小, 所需样本含量越多。 σ 、 π 、 ρ 一般未知, 通常以样本的 s 、 p 、 r 作为估计值, 多由预实验、查阅文献、经验估计而获得。

综上所述, 使检验假设的结论具有一定可靠性时所必需的样本含量即指按照一定的检验水准(即假设检验的I类错误概率 α)发现所比较因素间特定的效应差异 δ , 同时又保证检验功效为 $(1-\beta)$ 时所需的最少实验单位数。

MedCalc可为以下类型的研究计算样本含量:

样本平均值与总体平均值比较、两独立样本平均值比较、配对样本t检验、样本比例与总体率比较、样本比例与总体率比较、两相关样本比例比较、相关系数、生存分析对数秩检验、Bland-Altman图、两独立样本ROC曲线下面积比较、两相关样本ROC曲线下面积比较以及置信区间估计与精确度的样本含量估计。



更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归作者所有，请勿用于商业用途，[爱科学iikx.com](http://iikx.com)转发