

公卫医师医学统计学辅导：可信区间的估计 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/16/2021_2022__E5_85_AC_E5_8D_AB_E5_8C_BB_E5_c22_16187.htm

一、参数估计的意义
一组调查或实验数据，如果是计量资料可求得平均数，标准差等统计指标，如果是计数资料则求百分率藉以概括说明这群观察数据的特征，故称特征值。由于样本特征值是通过统计求得的，所以又称为统计量以区别于总体特征值。总体特征值一般称为参数(总体量)。我们进行科研所要探索的是总体特征值即总体参数，而我们得到的却是样本统计量，用样本统计量估计或推论总体参数的过程叫参数估计。本章第一节例6.1通过检查110个健康成人的尿紫质算得阳性率为10%，这是样本率，可用它来估计总体率，说明健康成人的尿紫质阳性率水平，这样的估计叫“点估计”。但由于存在抽样误差，不同样本(如再检查110人)可能得到不同的估计值。因此我们常用“区间估计”总体率(或总体均数)大概在那一个范围内，这个范围就叫可信区间。区间小的一端叫下限，大的一端叫上限。常用的有95%可信区间与99%可信区间。根据同一资料所作95%可信区间比99%可信区间窄些(上、下限较靠近)，但估计错误的概率后者为1%，前者为5%，进行总体参数的区间估计时可根据研究目的与标准误的大小选用95%、或99%。

二、总体均数的估计
为了说明常用的总体均数之区间估计法，我们不妨回顾一下上节所叙的t分布。由求t的基本公式我们看到X与 μ 的距离等于 $t(SX)$ ，又根据X集中分布在 μ 周围的特点，若取t的5%界即 $t_{0.05}$,(或1%界)乘以SX作为X与 μ 的距离范围，就可用式(6.6)或式(6.7)求出区间来估计总体

均数 μ 所在范围，估错的概率仅有5%或1%，因此称95%或99%可信区间。下面用实例说明其求法。95%可信区间 $\bar{X} - t_{0.05} \frac{S}{\sqrt{n}}$ ，99%可信区间 $\bar{X} - t_{0.005} \frac{S}{\sqrt{n}}$ 例6.2 上面抽样实验中第1号样本的均数为488.6，标准差为61.65，例数10，自由度 $=10-1=9$ ，试求95%与99%可信区间。1.求标准误 95%可信区间 $488.6 - 2.262(19.50)$ 样本号 X 95%可信区间 6546.75 15.78 ~ 577.62 7524.55 00.45 ~ 548.55 28476.14 54.91 ~ 497.29 72465.34 47.02 ~ 483.58 75526.65 03.10 ~ 550.10 平时我们并不重复抽取许多样本来一次次估计总体均数而仅是一次，至于算出的均数会类似一百个样本均数中的那一个就很难说了。如果不遇到类似上列那些均数过大或过小的样本，求出可信区间后总体均数真是在该区间内，那么便是一次成功的估计：但是极少数情况下我们也会遇到极端的样本，以至总体均数并不在我们提出的区间内。不过，我们具体所作的这次估计到底属于前种情况还是后一种，这是无法知道的，因为我们不知道 μ 是多少(若已知 μ 便不必估计它了)。然而象后种情况那样作出错估的概率终究很小，只5%或1%，所以用这样的方法估计总体均数还是可行的。三、总体率的估计 上面已经提到，计数资料可以计算相对数(率)。我们若由样本统计量 P 估计总体参数 π ，同样要考虑率的抽样误差，据数理统计研究结果，样本率的分布也近似正态分布，尤其当 π 比较靠近50%且样本较大时。于是对样本，百分率的可信区间可利用正态分布规律估计，公式是：95%可信区间 $P - 1.96Sp$ 99%可信区间 $P - 2.58Sp$ (按正态分布，双侧尾部面积 $\alpha=0.05$ 时的 u 值为1.96， $\alpha=0.01$ 时的 u 值为2.58，故用这两式求可信区间时不必查表找临界 u 值，记住这两数即可。) 例6.3 某医院收

治200例急性菌痢患者，其中粪便细菌培养阳性者共80例，试估计菌痢细菌培养的总阳性率95%与99%可信区间。1.求阳性率 $P = 80/200 \times 100\% = 40\%$ (或0.40) 2. 3.求可信区间 95%可信区间 $40\% - 1.96(3.46\%)$ 99%可信区间 $40\% - 2.58(3.46\%)$ 如果是小样本的百分率，求可信区间可通过查表获得，附表4是n为10、15、20、30时查95%与99%可信区间的简表。此外，统计学专著中还有更详细的表可查。转贴于：100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com