

做诊断指标的诊断价值研究时样本量如何选择？

作者：张华 赵一鸣 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/1937.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！

在前一期文章中我们看了t检验中p值的决定因素和诊断指标的决策价值的决定因素，两者差别在于样本量，即均值和标准差保持不变时，样本量增加时p值变小，但样本量本身不会影响诊断指标的决策价值。

在大样本的临床研究中，如果某指标差异有统计学意义，并不代表其有较好的预测能力。

什么算是大样本呢？我们先来模拟一组数据，我们固定均值差为4和标准差为2(此时灵敏度和特异度在85%左右)，减少样本量，观察p值、灵敏度和特异度的变化，我们可以得到下表：

均值差	标准差	每组样本量	P	灵敏度	特异度
4	2	1000	2.7×10^{-308}	84.6%	85.0%
4	2	500	4.3×10^{-152}	78.2%	90.4%
4	2	100	2.3×10^{-27}	85.0%	82.0%
4	2	10	0.000056	100%	90%
4	2	3	0.025	100%	100%

从上表看出，当样本量为1000时，此时p值小于 10^{-308} ，当样本量减少，p值增大，灵敏度和特异度基本不变(由于抽样误差的原因，灵敏度和特异度有波动)。样本量为10时，p还是小于0.001，每组样本量为3时， $p < 0.05$ 。也即是说预测能力较好的指标在样本量很小时就有统计学差异了。

接着我们再固定p接近0.05

，标准差为2，不同样本量下计算指标灵敏度、特异度和ROC曲线下面积。

均值差	标准差	每组样本量	P	灵敏度	特异度	曲线下面积
4	2	3	0.025	100%	100%	1.000
1.5	2	10	0.049	80%	70%	0.770
0.8	2	30	0.049	83%	43%	0.644
0.4	2	50	0.049	54%	64%	0.598

从上表可以看到指标的**诊断价值随样本量增加而减少**，

当**样本量增加到50**， $p=0.049$

时， $AUC=0.598$ ，**诊断价值已经非常小**

。也即是说如果你的样本量有几十个还未出现统计学差异，此指标单独进行预测的能力已经不太好。看到这个结果你是不是**很震惊**，你还会随意拿指标去做预测吗？

通过上面的数据，我们总结一下，就是在**筛选诊断指标**

时，想看某指标是否有**预诊断价值**，

只需要较小的有代表性的样本就可以了，**建议每组的样本量在10-30之间**，如果样本量大于30还未出现统计学差异，这个指标就可以**放弃了**；但在做某指标的**诊断价值的验证**时，为保证**诊断模型的稳定**，样本量可**不能太小**，一般认为**100-200例以上**可以接受。

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

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