

---

# 抽样方法介绍：简单随机和分层随机抽样

作者：曾琳，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/2112.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！**

在确定了研究对象的纳入、排除标准，即划分好设计人群后，从总体中抽取研究样本。抽样方法的选择和应用非常重要，它决定了我们抽得的样本对总体代表性的优劣，也会影响研究结果的外推。不少读者看了研究中的三种人群后，都表示对抽样方法感兴趣。这次我们就来聊聊抽样方法，由于相关的内容较多，我们分几期慢慢和大家唠唠。这期先讲讲简单随机抽样和分层随机抽样。

无论是简单随机抽样还是分层随机抽样都是较为常见的抽样方法，都属于随机抽样的范畴，应用这两种方法获得的研究对象样本都对总体有较好的代表性。

简单随机抽样(simple random sampling)是把符合要求的每一个个体都作为抽样的对象，通过随机化使每个个体被抽中选入样本的机会是相等的。由于随机化过程可以保证每个个体被抽中入选研究样本的机率相同，因此能保证研究样本对总体的代表性。举个小例子，假设符合我们要求的研究对象总体有20000人(这么整，一看就是假滴)，我们研究需要从中抽取200人作为研究样本，如果采用简单随机抽样的方法来获得研究样本，那么总体中每个人被我们抽中的机率都是1%。因为保证了机率相等，那么我们可以说我们抽得的样本是杠杠的好哇!

看到这儿有人会说了，这种抽样方法看起来很美，但是似乎都是在估计疾病频率的横断面研究里用，我们临床研究应用得很少吧。如果你也深以为然的话，那小编要提醒你，病例-对照设计的临床研究中也常会用到这个方法哦。比如，乳腺癌的预后研究中，如果是病例-对照设计，我们会先获得一组乳腺癌发生了转移的病例，say 100人。当然，我们的对照组一定是患乳腺癌但没有发生转移的病例，由于目前的乳腺癌预后还不错，所以没有发生转移的患者很多，我们如何从中挑选代表性好的对照呢?简单随机抽样啊，我们可以找同期的乳腺癌但没有发生转移的患者，从中随机抽取100、200甚至400例形成对照组(具体几例可以根据研究设计的需要了，当然例数越多，统计效能越高)。病例-对照研究中，对照组的选择是最考验研究者的设计功力的，对照的代表性常受质疑。我们可以在非病例中(在这个研究中是没有发生不良结局的患者中)采用随机抽样的方法获得有良好代表性的对照。

简单随机抽样的优点是能获得良好代表性的研究样本，操作实施也比较容易理解;其缺点是在抽样范围较大时，需要对总体中每个研究对象进行编号并收集基本信息，工作量太大从而影响研究可行性。另一方面，当某一重要研究因素在人群中分布不均匀时，采用简单随机抽样可能会导致在总体中占比例较少的个体被遗漏，从而导致选择偏倚。分层抽样则可以很好地解决这一问题。

分层抽样(Stratified Sampling)是从分布不均匀的研究人群中抽取有代表性样本的方法。先按照研究对象的属性(如年龄、性别、病情、病程、临床亚型、职业、教育程度、民族等)将研究人群分

---

为若干层，然后在每层内再开展随机抽样。还是借用上面的例子，同样是乳腺癌预后的病例-对照研究，乳腺癌的转移与否和病程可能相关，如果仅仅是简单随机抽样的方法获得对照组有可能出现病程短的人比较多(因为病程短的患者出现转移的机率低)，这样会降低对照组和病例组的可比性。这时我们可以根据病程分层再随机抽样，先看看我们的病例组中病程5年以下的有多少例，5年以上的有多少例，假设是2：3好了，那么我们就在乳腺癌没有转移的患者中根据病程是否大于5年分层，参考病例组的构成在不同的层内抽取研究对象构成对照组。

一定要注意，分层抽样要求层内变异越小越好，层间变异越大越好，这样可以提高样本的代表性，便于层间进行比较。分层随机抽样不能保证每个个体被抽中的概率相等，有可能处于不同分层之间的个体被抽中概率是不同的。

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