统计中如何将连续变量转换为分类变量？

作者：石岩岩，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/6379.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！

统计中如何将连续变量转换为分类变量？在做统计分析的时候，类似“年龄”这样的变量，不仅可以按照常规作为连续变量纳入分析，同时，还可以将其合理地转变为分类变量再纳入分析，以成功发掘出那些容易被隐藏的真实世界。

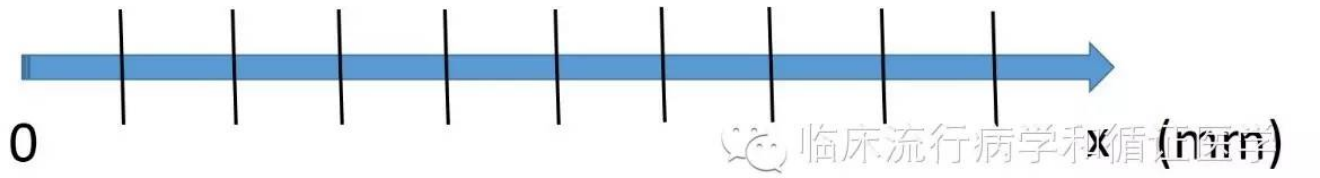
而最近我们又被很多次地问到了相似的问题，似乎不少研究者对此不太熟悉，这里有必要再给大家举一例，以理解
统计中如何将连续变量合理地转换为分类变量
：如某医院某科室的研究生要做一课题，拟分析“孕前子宫内膜厚度”对“妊娠发生与否”的影响。很显然，“孕前子宫内膜厚度”是自变量，“妊娠发生与否”是因变量，需要用到的分析方法当然是Logistic回归。在将“孕前子宫内膜厚度”代入模型的时候，除了可作为连续变量，还可分为若干层，成为分类变量。那么，分层时该如何寻找拐点呢?依据文献报道是一种方法，但是已有文献中的拐点可能并不能最适合您手头的数据。那么，我们用图示直观反映如何寻找数据的关键拐点。

1、连续变量样式



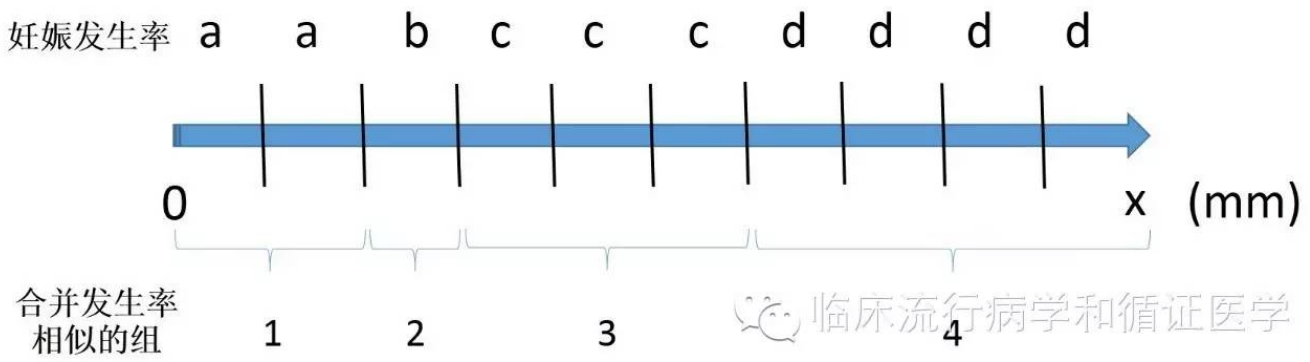
自变量“孕前子宫内膜厚度”本身是一个连续变量，我们用上面这一连续线来表示。如某医生总共收集了100个case。

2、找截点



将所有数据按照大小排列，将排列好的100个case平均分为10组(若想进行更加细致的分析，可以分为更多组)，每组10个case。计算每组的妊娠发生率。

3、合并结局发生率相似的组



将妊娠发生率相似(差别较小)的相邻组合并成一组。根据实际情况决定合并后的组数，一般建议3-4组。由此，将连续变量转换为分类变量，将结局发生率相似的组认为是一类。

4、按照分类变量的要求，设置成哑变量纳入模型

这里涉及的设置哑变量
，具体操作请参考下文：回归分析中的哑变量设置http://www.iikx.com/news/statistics/6380.html。

于是，这样我们就将原本是连续变量的数据转换为分类变量。数据需要进行转换的原因是该连续变量的数据与结局之间并非线性关系，而可能是折线或抛物线等非线性关系。那么，应该直接将连续变量纳入模型进行分析，还是先转换为分类变量?这需要依据数据的实际情况，并结合临床目的进行数据处理和结果解读，以得到对客观世界真实合理的诠释。

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转发