分组数据:潜在分布的参数估计

huhuaping

2021-02-18

Table of Contents

# 1 缘由

**分组数据**（或分配数列）是《统计学原理》课程中描述性统计部分的一个重要内容。课本里面的一些内容大多比较过失，而且很多问题没有说清楚。

这里面涉及到的主要是“有序分类数据”（Ordinal data）。

那么如何来理解两个“有序分类变量”（Ordinal variables）之间的关系？

# 2 参考资料

* Parameter Estimation of Binned Data （[原文网址](https://www.stats-et-al.com/2018/10/parameter-estimation-of-binned-data.html)；配套R代码[地址](https://drive.google.com/file/d/1_b_BoaJ4yhFfD7I0I0LsS_NKWvV0r_tH/view)）
* How does Polychoric Correlation Work? (aka Ordinal-to-Ordinal correlation)（[原文网址](https://www.stats-et-al.com/2021/02/how-does-polychoric-correlation-work.html)）

# 3 数据

source("../R/set-global.R")

df\_bin <- tibble::tribble(  
 ~Group, ~Frequency,  
 "0 to 25", 114,  
 "25 to 50", 76,  
 "50 to 75", 58,  
 "75 to 100", 51,  
 "100 to 250", 140,  
 "250 to 500", 107,  
 "500 to 1000", 77,  
 "1000 to 5000", 124,  
 "5000 or more", 42)

knitr::kable(df\_bin, align = "c", caption = "分组数据")

Table 3.1: 分组数据

|  |  |
| --- | --- |
| Group | Frequency |
| 0 to 25 | 114 |
| 25 to 50 | 76 |
| 50 to 75 | 58 |
| 75 to 100 | 51 |
| 100 to 250 | 140 |
| 250 to 500 | 107 |
| 500 to 1000 | 77 |
| 1000 to 5000 | 124 |
| 5000 or more | 42 |