

R 语言与数据可视化——直方图作业

李柯毅 2100012618

数据背景：该数据库来源于 Kaggle，由 Sushant Chougule、MadhavSeth 等人从 Salary Explorer 网站上抓取并整理。Salary Explorer 统计了全球 222 个国家（地区）的工资情况，其工资数据大部分由网站用户提供，少部分从政府机构、商业公司等处获取。网站根据统计时段内的平均汇率将各国工资统一换算为美元表示，并通过统计学方法整合不同来源的数据以确保准确性。该数据库每月 3 日更新，更新会在 Kaggle 上同步。

数据情况：该组数据统计了 2023 年 9 月各国工资情况。变量包括国家名称、所属大洲、工资时段、中位工资、平均工资、最高工资、最低工资七项，其中国家名称、所属大洲、工资时段属于定性指标，其中工资时段固定为月度。中位工资、平均工资、最高工资、最低工资属于定量指标（数据展示如图 1）。在定量指标中，我认为中位工资最能代表一个国家的工资水平，因为相比于平均值，中位数有效避免了离群值的影响，反映了该国位于 50% 收入水平的民众工资的情况，故本次作业选取其做后续分析。

```
> head(dat)
  country_name continent_name wage_span median_salary average_salary lowest_salary
1  Afghanistan          Asia   Monthly          853.74         1001.15          252.53
2  Aland Islands        Europe   Monthly         3319.24         3858.35          972.52
3    Albania            Europe   Monthly          832.84          956.92          241.22
4    Algeria            Africa   Monthly         1148.84         1308.81          330.11
5 American Samoa      Oceania   Monthly         1390.00         1570.00          400.00
6    Andorra            Europe   Monthly         3668.08         4069.77         1120.51
  highest_salary
1         4460.97
2        17124.74
3         4258.49
4         5824.18
5         6980.00
6        17653.28
```

图 1 数据展示

数据描述：各国中位工资的最小值为 0.261\$，最大值为 9836.070\$，中位数为 1227.46\$，平均数为 1762.632\$（图 2）。

```
> summary(dat$median_salary)
   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
 0.261  567.210 1227.460 1762.632 2389.010 9836.070
> |
```

图 2 数据描述

对中位工资绘制直方图如图 3。可以发现直方图呈现明显的右偏（符合和钱相关则右偏的规律）。全球各国的月薪中位数普遍集中在 0-2000\$ 范围内，中位月薪达 2000\$ 以上后国家数量快速减少，但也可以看到极少数国家中位月薪达到接近 10000\$ 的水平。

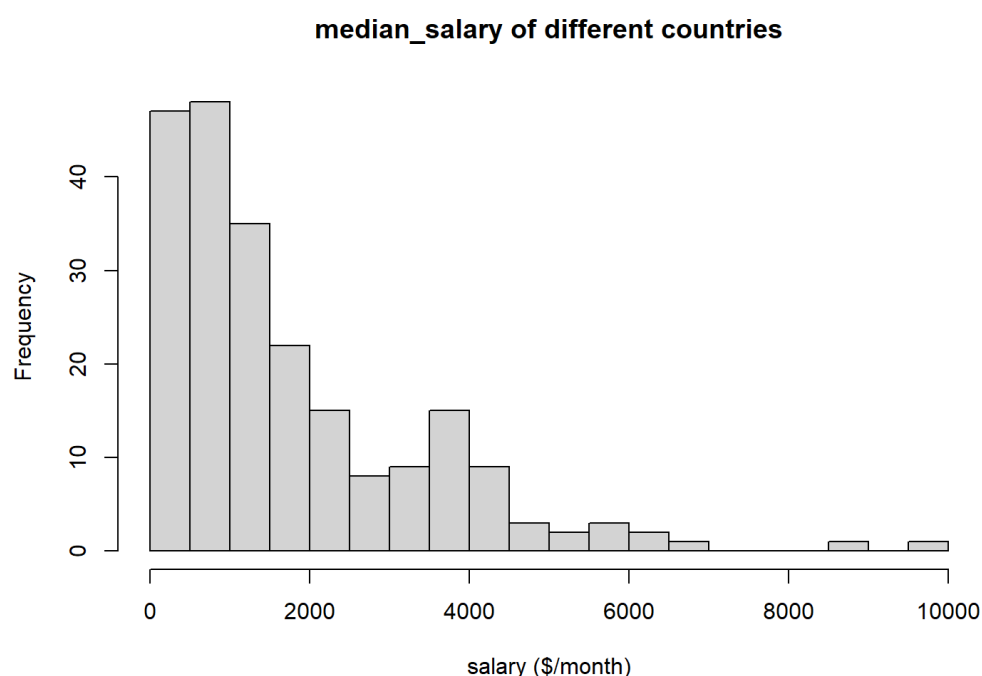


图 3 对中位工资绘制直方图

结果分析：

1. 月薪中位数在 8000\$ 以上的国家有：Guernsey (8689.02\$) 和 Switzerland (9836.07\$)。这两个地区的月薪中位数远远领先于其他国家，以下对其做展开分析：

- a) Switzerland（瑞士）的人均 GDP 一直稳定在全球前五，这里政府稳定、金融业发达且有较高的生产率，其也一直以高生活成本和高工资闻名。综上，瑞士具有全球最高的工资水平不足为奇。

b) Guernsey (根西岛) 取得第二名有些出人意料。根西岛是英国王室的自治属地。该岛不是英国的一部分, 尽管英国政府对辖区负有一定的责任, 英国君主是国家元首。该地区土地面积 62 平方公里, 人口不足 65,000 人。我认为根西岛有如此高的中位月薪数据可能有以下几个原因: ①以金融业为核心: 该地区是欧洲金融中心之一, 金融业拉动了居民收入提升②小面积高度发达: 根西岛土地稀缺, 导致高住房成本与生活成本。③税收极低: 根西岛没有增值税、消费税或资本利得税, 只征收个人所得税, 企业税率也相对低。④统计误差: 因为根西岛人口较少, 统计样本缺乏, 可能导致结果产生偏误。

2. 月薪中位数最低的国家是 Zambia, 只有 0.261\$, 远低于国际赤贫标准 (1.9\$/d)。此外, Syria (10.12\$)、Mauritania (44.62\$) 也在赤贫标准以下。这些地区经济很不发达, 或是连年处于战乱, 人们难以维持基本的生活水平, 应该引起国际的关注。

3. 大部分国家月薪收入中位数仍集中在 2000\$ 以下, 各国月薪极值差距甚至达到了 9836\$, 反映了明显的贫富差距。贫富差距的产生源于各种不平等因素:

- 经济不平等制度: 不平等的经济体制, 如资本主义中的利润追求, 有时会导致贫富差距扩大。资本主义经济通常奖励资本所有者, 而工人可能无法分享相同的财富。
- 教育机会不平等: 缺乏教育机会会使人陷入贫困。许多贫困国家的人口缺乏良好的教育, 这限制了他们获得高薪工作的机会。
- 政治不稳定和腐败: 战争、不稳定的政治环境和腐败可能导致资源分配不公平。腐败政府可能偏向富人或特权阶层, 而不是满足广大民众的需求。
- 国际贸易不平等: 不公平的国际贸易关系可能导致发展中国家的资源被剥夺, 而富裕国家受益。这包括不公平的贸易协定、贸易壁垒和资源剥夺。

延伸: ①该数据库也可以用来解释国家内的贫富差距 (通过每个国家的最高收入和最低收入)。②通过大洲分组后, 也可以比较不同大洲国家的收入分布情况。③同时也可以联系国家 GDP 等经济表现的数据库, 分析人均收入的影响因素。④可以加入时间因素, 分析不同年份各个国家的收入情况, 对比国家的经

济发展速度以及影响收入的因素。

数据来源: <https://www.kaggle.com/datasets/zdataweaver/global-salary-data>