**人口大数据开发及运用的简略报告**

郑惠莉

21937078

目录

[数据描述 2](#_Toc38723446)

[数据分析 3](#_Toc38723447)

[不同省份流入浙江的人口与平均工资 3](#_Toc38723448)

[家庭成员数与平均工资： 4](#_Toc38723449)

[流入人口的文化情况与平均工资： 5](#_Toc38723450)

[务工与经商： 7](#_Toc38723451)

[性别与平均工资： 9](#_Toc38723452)

# 数据描述

数据质量：数据来源浙江流入人口特征数据，一共有10000个样本，16个特征。特征包括：residence：省份；familymembers: 家庭成员个数；gender:性别； birth year: 出生年份； birthmonth： 出生月份；nationality：国家；education：教育程度； hukou：户籍类型；marriage：婚否；hukouplace：户口所在地；flowrange：流入类型；flowyear：来浙年份；flowmonth：来浙月份；flowreasons：来浙原因；salary：工资；employment：是否有职业。其中数值类数据有6个，占全部的37.5%，分类型数据有10个，占全部的62.5%。所有特征都无缺失值。

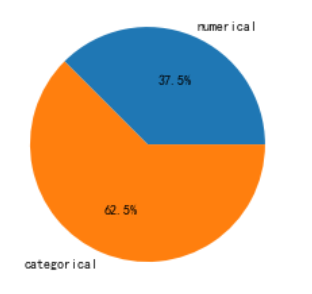


图1.特征类型

数据的总体概况：10000个样本里面，家庭成员平均以及中位数都是有3个，来浙年份的中位数是2015年，工资的中位数是4000元。

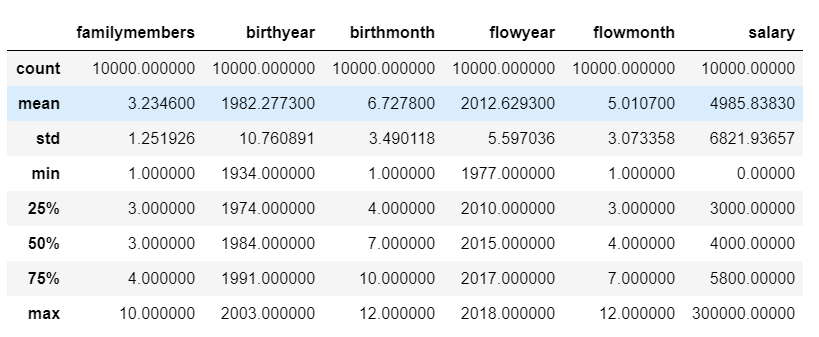


表1.总体概况

# 数据分析

## 不同省份流入浙江的人口与平均工资

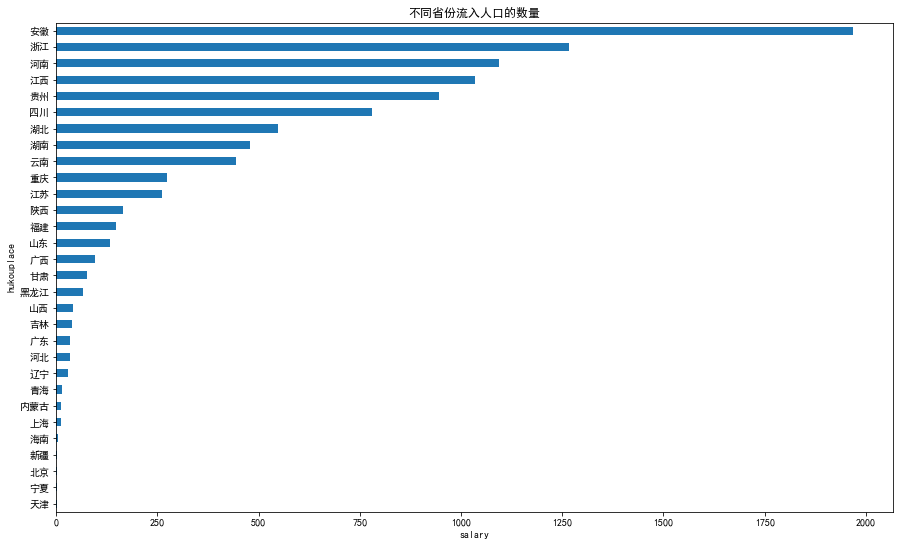


图2.不同省份流入人口的数量

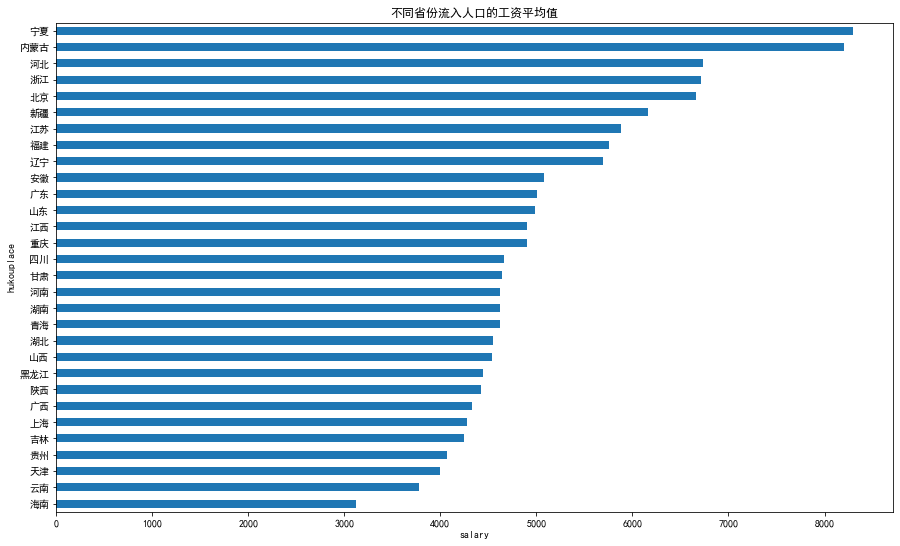


图3.不同省份流入人口的平均工资

数据图3显示，户籍地是宁夏和内蒙古的来浙人口的平均工资排名前二的，但图2显示宁夏，内蒙古来浙江人数很小，工资均值高可能是个别样本提高了平均值。其中安徽，浙江，江苏，福建来浙人口数比较大，对应的省份人口的工资平均值有参考价值。可以看到浙江省内流入的人口平均工资是较高的，可能由于户籍地距离较近，对浙江文化比较熟悉，比较容易融入，人脉较多，相对更稳定，等等其他可能的因素。

## 家庭成员数与平均工资：

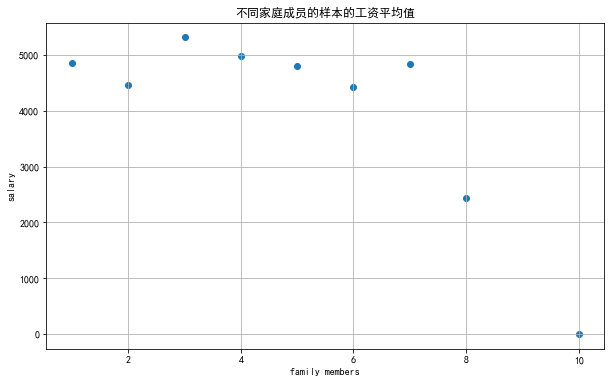


图4.不同家庭成员的人群的平均工资

图四显示，当家庭成员较大时，超过七个人时，平均工资都比较低，可能时因为家庭成员多，生活成本高，平均资源少。当家庭成员是两个时，平均工资比单人和多人家庭低，可能是由于夫妇没有孩子的话，比较年轻，家庭压力不是特别大，没有更大的动力追求更高的工资，也有可能是因为工资较低，所以暂时选择不扩充家庭人员数量。在家庭成员为3到6的范围内时，平均工资逐渐下降。

## 流入人口的文化情况与平均工资：

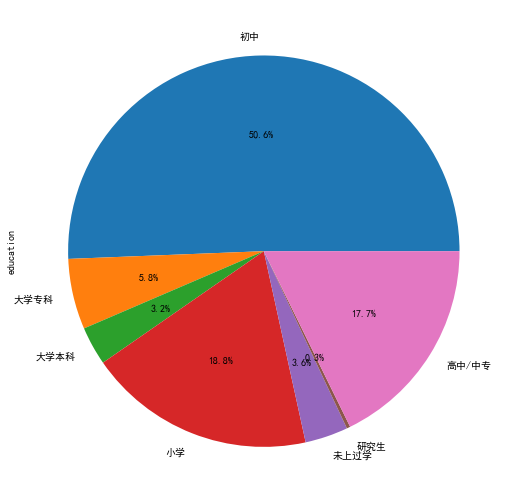


图5.不同学历程度人口的数量占比

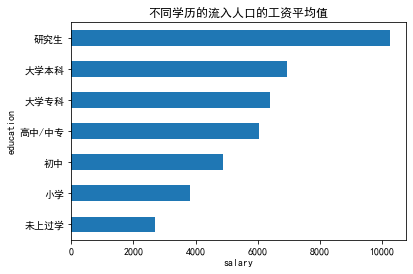


图6.不同学历程度的平均工资

从图五可以看到，初中学历的人数占来浙人口的一半之多，小学以及高中/中专的学历都在17%左右，大学专科为5.8%，本科仅占3.2%，研究生更是仅有0.3%。可以看到大部分人都是受过文化教育的，但是高质量文化的人群还是占少数。但在图六中，受到文化教育的质量越高的人群，平均工资水平也是越高。差距最大的时本科生与研究生的平均工资水平，教育质量只差一级，但平均工资水平差距接近三分之一。而大学本科和大学专科和高中/中专的距离并没有很大。

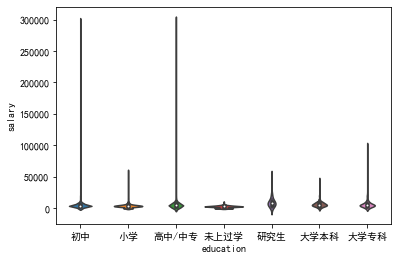


图7.不同学历程度的工资violin图

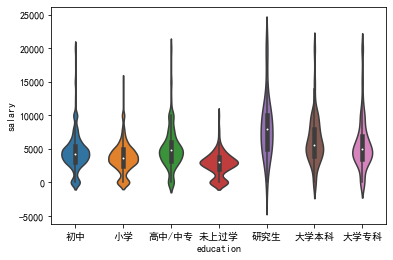


图8.不同学历程度的工资violin图（排除离群值）

在加载所有数据时，初中以及高中水平的最高工资高达三十万，考虑这可能是个别情况，去掉了离群值之后，再重新做的图八，图八显示，研究生的最高工资在25000左右，中位数也是所有人群中最高的，但是频率较小，符合上述图五和图六的情况。未上过学的人群中位数较低，人群频率也较大。上述四张图都可以看出，来浙人口中，受过高质量教育水平的人较少，但高质量教育人群工资都相应较高，也有少部分人教育水平稍低，收入很高。可以看出浙江是一个尊重知识的地方，有知识可以创造财富，也是认可努力的地方。

## 务工与经商：

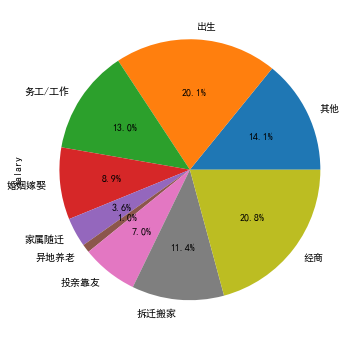


图9.不同来浙理由的占比

图9可以看到来浙人口大部分人选择了务工和经商。

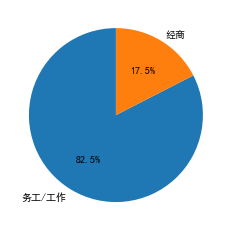


图10.务工与经商的占比

图10显示大部分人选择务工/工作，不到20%的人选择经商。原因可能是经商困难比较大，成本也比务工高，比起工作不稳定性更高，风险更大。

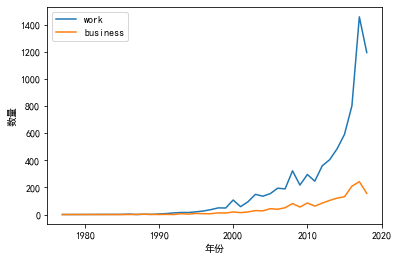


图11.务工与经商的数量随时间变化的曲线图

图11显示在90年代中旬，流入浙江人口数量增加，来浙务工的人数呈指数型增加，经商人数缓慢增加，直到2018，两者人口数都有相应减少，可能是因为当年浙江房价上上涨，对于外来人口有压力。

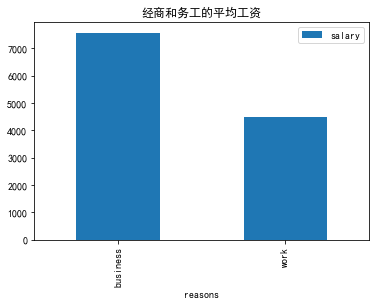


图12.务工与经商的平均工资水平

图12显示经商的平均工资比工作高，对于经商成功的，收入比务工高。但经商的困难是比工作的困难多，风险高的，最后的回报也会比平均务工工资高。

## 性别与平均工资：

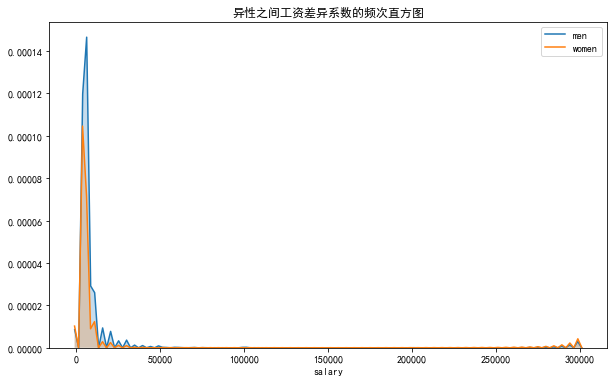


图13. 异性之间工资差异系数的频次直方图

图13显示在平均工资0-50000内，女性的比例低于男性，但在250000-300000之间，女性比例稍微高于男性。原因可能是正常情况下，有更多男性赚钱。