



中国传媒大学
COMMUNICATION UNIVERSITY OF CHINA

题目：国内女性社会地位认知及相关影响因素研究

学年学期：2021-2022 学年 春季学期

课程名称：数据科学导论

课程编号：2131030161

课程序号：01

任课教师：王小宁

姓 名：余程明

学 号：2021218093031

评分区域（由阅卷老师填写）：

结课成绩：

总评成绩：

提交时间： 2022 年 7 月 8 日

国内女性社会地位认知及相关影响因素研究

——基于 CGSS2015 的实证分析

中国传媒大学广告学院

广告学（计算广告双学士学位复合型人才培养项目）

2021218093031 余程明

2022 年 7 月 7 日

目录

一、 问题概述	2
二、 文献综述	2
三、 模型构建	2
基本思想	2
自变量与因变量	3
四、 数据分析	3
基于数据可视化的对比分析	3
数据描述	4
数据可视化	6
五、 总结与反思	10

一、问题概述

当今社会，关于女性社会地位的话题热度激增，甚至许多电视剧将女性在社会中的困顿场景搬上了大荧幕，引起网络上许多的话题讨论，也让更多的人开始关注到女性由于顽固的历史劣根性遗留下来的偏见而遭受的不公平待遇。如果说这一系列社会事件让女性所受到的不平等待遇暴露在大众的视野当中，那么 2022 年 6 月 10 日的“唐山烧烤店打人事件”和 6 月 24 日美国联邦最高法院宣布废除女性堕胎权的轰炸性新闻，无疑将这一话题引爆到又一高峰。

基于此，本文将以 CGSS 2015 中国综合社会调查数据作为研究基础，采取分类对比的研究方法，对影响女性社会地位认知的因素进行探索。本研究采取的数据分析方法包括数据可视化和 logistic 回归分析：通过数据可视化的方法更直观的感受男性和女性对“女性社会地位认知”之间的差异，并通过 logistic 模型检验受教育程度对“女性社会地位认知”的四个方面的表现的影响。

二、文献综述

在进行该研究之前，已经有众多学者对国内外以及古今的女性社会地位对比、演变和弊端等方面做过相关论述及研究。

有文献对女性的生育率进行了数据分析，指出最不发达地区女性的低教育水平和妇幼医疗保障的滞后是这些地区高生育的重要原因，是造成女性为生育而牺牲经济利益表象的主要原因。^[1]该结论论述了女性生育率的相关决定性印象因素，而“女性生育率”可以看作“女性社会地位”的一个衡量指标，于此同时，该结论同样指出了教育水平对女性生育率的影响，这也是本文以 logistic 模型验证受教育水平对女性社会地位认知影响的理论支撑。

另有学者认为，在当前社会，男尊女卑、重男轻女等错误思想不应成为歧视女性、特殊对待女性的原因。女性题材的都市剧很好地向大众展示了女性新形象，传递

着女性主义。^{【2】}从如今女性题材的都市剧的角度剖析了目前社会存在的一些陈旧、错误的思想，以及女性之于社会的重要意义。

从以上文献可知：当前女性的确面临着许多不公平的待遇，并且教育水平在很大程度上影响着思想的开放程度以及社会对于女性社会地位的认知，但可以看到目前还没有利用相关数据验证除生育率以外的女性社会地位表现的相关研究。据此，本文将对 CGSS2015 中四个关于女性看法的问卷进行数据分析，试图得出有数据支撑的严谨结论。

三、模型构建

（一）基本思想

本文主要有两个研究目的：

1. 通过数据可视化以及对比分析的方法直观地感受男性和女性对于“女性社会地位”的四个相关表现的态度之间的差异；
2. 利用 logistic 回归模型对“受教育程度”之于此四个表现的影响进行系数估计和显著性检验。

（二）变量、衡量指标的确定

自变量：受教育程度，以问卷中 a7a（您目前的最高教育程度（包括目前正在读的））作为衡量指标；性别

因变量：女性社会地位认知，以问卷中 a421（您是否同意-男人以事业为重，女人以家庭为重）、a422（您是否同意-男性能力天生比女性强）、a423（您是否同意-干得好不如嫁得好）以及 a424（您是否同意在经济不景气时，应该先解雇女性员工）作为衡量指标。

四、数据分析

（一）基于数据可视化的对比分析

1. 读取数据：

利用 csv 文件筛选方法分别得到男性和女性两个子表。其中 data1 为男性数据，data2 为女性数据。

```
data1 = df1[['a421', 'a422', 'a423', 'a424']]
data1.head(3)
```

	a421	a422	a423	a424
0	4	4	4	2
1	2	1	1	1
4	4	4	4	4

```
data2 = df2[['a421', 'a422', 'a423', 'a424']]
data2.head(3)
```

	a421	a422	a423	a424
2	4	2	4	2
3	4	4	-8	2
6	1	1	2	1

2. 缺失值处理：

考虑到本研究的调查基数比较大，且经过初步观察，缺失值的数量相对较少，因此采用 dropna（）方法直接删除含有缺失值的项目，在对结果影响堆笑地条件下尽量简化程序。

3. 数据描述：

注：a421（您是否同意-男人以事业为重，女人以家庭为重）、a422（您是否同意-男性能力天生比女性强）、a423（您是否同意-干得好不如嫁得好）以及 a424（您是否同意在经济不景气时，应该先解雇女性员工）的回答分数意义为：1 分——完全不同意，2 分——比较不同意，3 分——无所谓同不同意，4 分——比较同意，5 分——完全同意。可以看出，受访者的回答分数越高，则可以认为

其认为女性的社会地位越低。

```
stal=new_new_data1.describe()
print(stal)
```

	a421	a422	a423	a424
count	4952.000000	4952.000000	4952.000000	4952.000000
mean	3.405897	3.034733	3.045032	2.293821
std	1.132519	1.130544	1.105335	0.959226
min	1.000000	1.000000	1.000000	1.000000
25%	2.000000	2.000000	2.000000	2.000000
50%	4.000000	3.000000	3.000000	2.000000
75%	4.000000	4.000000	4.000000	3.000000
max	5.000000	5.000000	5.000000	5.000000

```
sta2=new_new_data2.describe()
print(sta2)
```

	a7a	a421	a422	a423	a424
count	10498.000000	10498.000000	10498.000000	10498.000000	10498.000000
mean	4.946180	3.350257	2.997428	3.09859	2.220994
std	3.135634	1.170967	1.168109	1.13418	0.986297
min	1.000000	1.000000	1.000000	1.000000	1.000000
25%	3.000000	2.000000	2.000000	2.000000	2.000000
50%	4.000000	4.000000	3.000000	3.000000	2.000000
75%	6.000000	4.000000	4.000000	4.000000	3.000000
max	14.000000	5.000000	5.000000	5.000000	5.000000

通过计算相关特征值，由图表所示：男性对于社会地位的认知态度基本保持在 3 分上下，且方差较小，说明男性对于女性社会地位认知的态度相对集中。反观女性，虽然其平均水平仍保持在 3 分上下，但其整体水平低于男性，可以认为女性和男性对于女性社会地位认知的平均态度差异不大。但从整体上看，所有受访者的平均认知水平仍集中在 3 分，说明整个社会的平均态度时保持中立，这也符合正态分布的常规认知。

4. 程序优化：

通过上表可以看出 a421, a422, a423, a424 四个表现基本一致，且考虑到四个变量两两对比的程序比较繁琐，于是选择用 a421 作为代表性元素进行程序编写，以进行数据分析。

5. 由于饼状图可以直观的体现出同一个话题讨论下，受访者的不同意见之

间的显著的差异、对比，因此本文采用饼状图的方式对数据进行可视化。与此同时，将 csv 文件分筛选为两个子文件的目的在于分别研究男性和女性对关于”女性社会地位认知“相关问题的态度及想法，以此从另一个维度进行对比分析，使对比研究更鲜明，结论更具说服力。下面进行饼图生成。

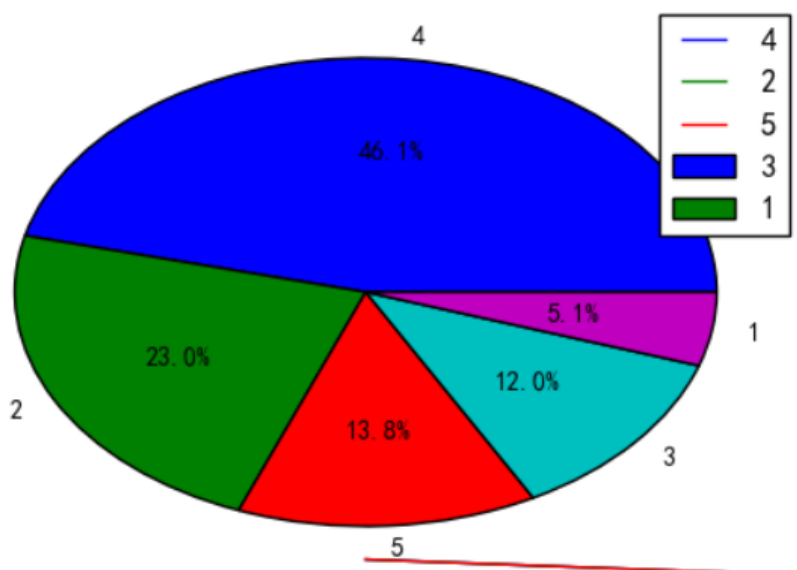
6. 分别生成男性和女性的关于数据 a421（您是否同意-男人以事业为重，女人以家庭为重）的饼状图观察数据 a421 之间的分布情况：

```
data = pd.read_csv('man.csv', encoding='gbk')

plt.rcParams['font.sans-serif'] = 'simhei'
plt.rcParams['axes.unicode_minus']=False

data = data.groupby('认知程度').sum()['人数'].sort_values(ascending=False)
print(data.values)
plt.pie(data.values, labels=data.index, autopct='%1f%%')
plt.legend(data.index)
plt.show()
```

[2284 1138 684 594 252]



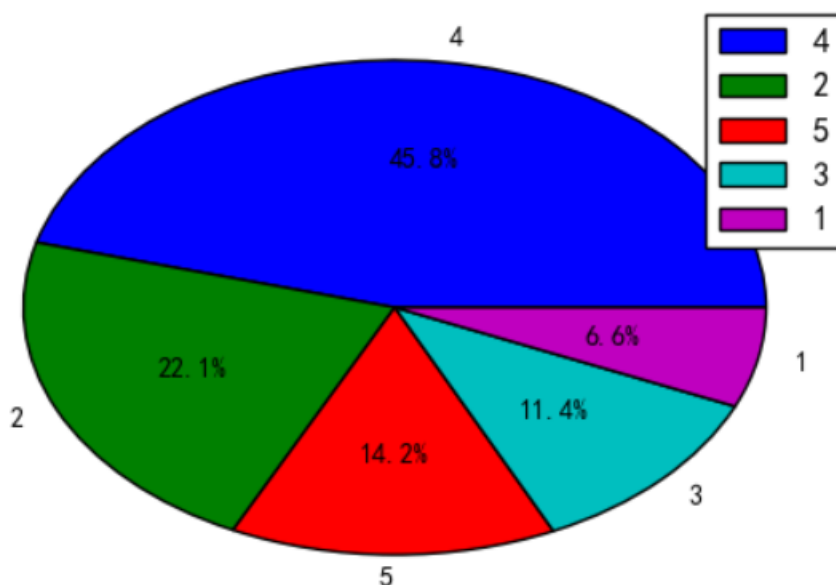
男性对女性相关认知饼状图

```
data = pd.read_csv('woman.csv', encoding='gbk')

plt.rcParams['font.sans-serif'] = 'simhei'
plt.rcParams['axes.unicode_minus'] = False

data = data.groupby('认知程度').sum()['人数'].sort_values(ascending=False)
print(data.values)
plt.pie(data.values, labels=data.index, autopct='%1f%%')
plt.legend(data.index)
plt.show()
```

[4667 2250 1448 1165 668]



女性对女性相关认证饼状图

通过饼状图可以直观地看出：无论男性还是女性，对该话题的不同态度之间的比例是基本一致的，这可以说明社会中男性和女性对于女性社会地位认知基本一致，存在“压迫”、“男权”现象的可能性较小；另外值得注意的一点是打 4 分的受访者都占据了接近半数的受访人数，可能说明国内社会的陈旧思想、历史遗留下来的不良风气仍然存在。

（二）回归模型分析

1. 读取数据并筛选数据：由于本研究的因变量衡量指标有 4 个，因此在筛选数据时分别读取四个形状为 (10969, 2) 的表格信息，以建立四个回归分析模

型。

2. 选定变量并划分数数据集：

选定变量

```
X= data[['a7a']] #选取列名为a7a的列作为X
Y1= data[['a421']] #选取列名为a421的作为Y1.
Y2= data[['a422']]
Y3= data[['a423']]
Y4= data[['a424']]
```

划分数数据集

```
##对上述的X和Y1进行划分为3和7
X_train, X_test, y1_train, y1_test = train_test_split(X, Y1, test_size=0.3, random_state=42)
X_train, X_test, y2_train, y2_test = train_test_split(X, Y2, test_size=0.3, random_state=42)
X_train, X_test, y3_train, y3_test = train_test_split(X, Y3, test_size=0.3, random_state=42)
X_train, X_test, y4_train, y4_test = train_test_split(X, Y4, test_size=0.3, random_state=42)
```

选定四个“对女性社会地位认知”的相关表现作为四个因变量，分别与因变量“受访者的最高教育程度”构建出四个独立的一元线性回归模型。简便起见，利用 `train_test_split()` 方法进行数据集的划分。

3. 打印模型：

打印整个模型

```
: print("Y1 = " + str(model.intercept_[0]) + " + " + str(model.coef_[0][0]) + " * X")
```

```
Y1 = 3.688528018058471 + -0.07882860802086364 * X
```

```
: print("Y2 = " + str(mode2.intercept_[0]) + " + " + str(mode2.coef_[0][0]) + " * X")
```

```
Y2 = 3.2755306200025913 + -0.06831266928532875 * X
```

```
: print("Y3 = " + str(mode3.intercept_[0]) + " + " + str(mode3.coef_[0][0]) + " * X")
```

```
Y3 = 3.2127392117409532 + -0.04993568328874153 * X
```

```
: print("Y4 = " + str(mode4.intercept_[0]) + " + " + str(mode4.coef_[0][0]) + " * X")
```

```
Y4 = 1.8646706538935345 + 0.007230349031911904 * X
```

从一元线性回归方程的拟合结果来看：相关参数的绝对值很小，除此之外还存在这相关系数为负数的情况，由此可以认为，并非学历越高，对于女性社会地位的认知就越公平，甚至还存在两个变量呈负相关的可能。

4. 计算相关系数：

```
R1=model.score(X, Y1)
print(R1)
```

0.03128030862293596

```
R2=model2.score(X, Y2)
print(R2)
```

0.02185909882624626

```
R3=model3.score(X, Y3)
print(R3)
```

0.008264720998341413

```
R4=model4.score(X, Y4)
print(R4)
```

-5.545065710887975e-05

从相关系数的计算结果上看，四个模型的绝对值都远远小于 1，由此进一步说明学历与对女性社会地位认知之间的关系并不强，

五、总结与反思

通过基于 CGSS2015 的数据分析，可以初步得到以下结论：

1. 男性和女性对于女性社会地位的认知并无明显差异，可推测社会基本不存在一方压迫另一方的现象，否则男女对立必然严重，两性之间的态度和认知也大概率会出现明显差异；
2. 一突出特征有关于“您是否同意-男人以事业为重，女人以家庭为重”的调查，无论男女均有接近一般的人数表示“比较同意”，在一定程度上说明传统的男女分工内容思想对当今社会仍发挥和不可忽视的巨大作用；
3. 学历并非对女性社会地位认知产生影响的决定性因素。

针对上文提出的“最不发达地区女性的低教育水平和妇幼医疗保障的滞后是这些地区高生育的重要原因”^[1]的观点，由于该观点中提到“最不发达地区”，具有一定的特例性，同时观点表示除了低教育水平以外，妇幼医疗保障的滞后也是一大影响因素，因此仅凭本研究目前的结论，没有充分的理由否认该观点的合理性。

六、参考文献

- [1] 刘琳. 女性社会地位与总和生育率 [C]南京人口管理干部学院学报, 2006, 22(2): 32.
- [2] 谢玉青. 社会地位分析 ——基于热播剧的思考 [J] 艺术科技, 2021, 34(9):95.