

任课教师	陈丹	考查方式	论文	得分	
------	----	------	----	----	--

云南大学 2023 年秋季学期  
数学与统计 学院 统计学 专业  
 专业选修课 《属性数据分析》 期末考  
 查作业

题目： 阶层地位获取的影响因素探究

——基于 CGSS 混合截面数据的实证分析

姓 名 枫叶

学 号 \_\_\_\_\_

年 级 2021 级

专 业 统计学

2023 年 12 月 22 日

**摘要：**阶层流动问题一向是社会学领域的热点问题，本文基于 CGSS 数据库 2017-2021 年的成果构建了混合截面数据，使用 logistic 模型对影响个体进入某一阶层的因素进行了分析，结果显示父辈的受教育水平越高，则后代进入高阶层的概率越高，但本人的受教育水平发挥着更大的作用，父辈的行政职级和政治面貌影响不显著，除此之外个体的性别、户口、党员身份等都起到了一定作用，异质性分析显示上述结论存在一定地区差异。

## 一、数据来源与变量说明

### 1.数据来源

本文所使用数据来自于中国综合社会调查（CGSS），其按照科学的抽样方法对全国居民展开抽样问卷调查，虽然其调查过程依旧存在一些争议，但依旧不失为当前中国最权威且最具代表性的社会调查数据之一。本文为了扩大样本量，同时考虑到时间跨度过大可能会引入过多噪声，最终选取了 CGSS2017,2018,2021 三个年份形成混合截面数据，共有 33517 个样本观测值。

### 2.变量设计

**被解释变量：**受访者本人的阶层位置，我们使用 EGP 阶层分类来反映阶层位置，并按照侯利明和秦广强（2019）的研究作了修正<sup>1</sup>，最终形成了九分类的有序分类变量。

**解释变量：**受访者 14 岁时父亲和母亲的受教育年限、政治面貌、行政职级，其中受教育年限按照最高学历折算为对应的受教育年限，政治面貌按照群众及共青团员、民主党派、党员重新编码，行政职级按股级及以下、科级、处级、副司局级及以上重新编码。

**控制变量：**受访者本人的受教育年限、政治面貌、性别、年龄、户口，其中受教育年限和政治面貌的处理同上，户口按农业户口和非农业户口的划分重新编码。

### 3.缺失值处理

经过重新编码后的数据中存在较多缺失值，这主要是两个原因导致的：一是 EGP 阶层分类依赖于受访者当前的工作，如果受访者目前没有工作则该项缺失，对这一部分样本直接作删除处理；二是 CGSS 采取了问卷调查的形式收集数据，受访者不知道或拒绝回答时会记录为缺失值，显然前者的缺失机制为完全随机缺失，而后者

---

<sup>1</sup> 本文的研究基础为我个人的公众号文章《EGP 阶层分类框架——以 CGSS 数据为例》，撰写公众号文章时编写了适配 CGSS 所有年份的代码，但由于当时尚不熟悉 purrr 包，故代码较为粗糙和繁复，由于时间所限，这里没有进行优化。

为非随机缺失，Little’s MACR 检验给出的 p 值极小，即不能认为数据为完全随机缺失。对于非随机缺失数据既不能盲目删除，也没有较好的插值方法，限于笔者能力所限，这里只能粗糙地作成列删除处理，删除后的数据集共有 9145 个观测值。

表 1 给出了一个变量的概括性说明，表 2 是对各阶层含义的说明

表 1 变量概览

变量类型	变量	赋值
被解释变量	阶层位置	从高层到低层依次为 1,2,3,4,5,7,8,9,10
	父亲受教育年限	依据最高学历折算为对应受教育年限
	母亲受教育年限	依据最高学历折算为对应受教育年限
解释变量	父亲政治面貌	群众及共青团员=0，民主党派=1，党员=2
	母亲政治面貌	群众及共青团员=0，民主党派=1，党员=2
	父亲行政职级	股级及以下=0，科级=1，处级=2，副司局级及以上=3
	母亲行政职级	股级及以下=0，科级=1，处级=2，副司局级及以上=3
	个体受教育年限	依据最高学历折算为对应受教育年限
控制变量	个体政治面貌	群众及共青团员=0，民主党派=1，党员=2
	个体性别	男性=1，女性=2
	个体年龄	连续变量
	个体户口	农业户口=0，非农户口=1
其他	地区	所在省份名称

表 2 阶层定义

阶层	阶层名称	含义
1	Higher service	高级专业技术人员;行政管理者、大企业的管理者;大资产所有者
2	Lower Service	较低专业技术人员、行政管理者;小型企业管理者
3	Routine clerical/sales	行政、商业部门的非体力雇员;销售人员;其他服务业雇员
4	Small employer	小资产者,自雇雇人
5	Independent	小资产者,自雇不雇人
6	Farmers/Farm managers	自营农民
7	Manual foreman	低级技术人员;体力雇员的监督者
8	Skilled manual	熟练技术型体力工人

阶层	阶层名称	含义
9	Semi-unskilled manual	半技术和非技术的体力工人
10	Farm workers	农业劳动者

## 二、列联表分析

图 1 展示了有序变量之间的 Kendall 相关系数，由于样本量较大，所有相关系数都通过了显著性检验。从图中可以看到所有变量均与个体的 EGP 阶层类别呈现出一定的负相关关系，其中个体的受教育水平具有最强的相关性，由于 EGP 阶层类别越高表示所处阶层越低，这意味着这些因素可能都起到了防止个体进入较低阶层的作用。此外，父亲或母亲的受教育水平与本人受教育程度有一定程度的正相关，这可能表示学历越高的家庭越倾向于让孩子接受更多教育；父亲受教育水平和母亲受教育水平呈现出较强的正相关关系，这可能意味着学历水平属于“门当户对”的一个方面，当然这也可能是因为学历水平相近的一对异性更有可能具有相似的家庭环境、认知环境或是社交环境；父辈的政治面貌和行政职级存在一定程度的正相关，但受教育水平与行政职级的关联并不大，这可能说明学历因素在职级升迁中居于较次的地位。

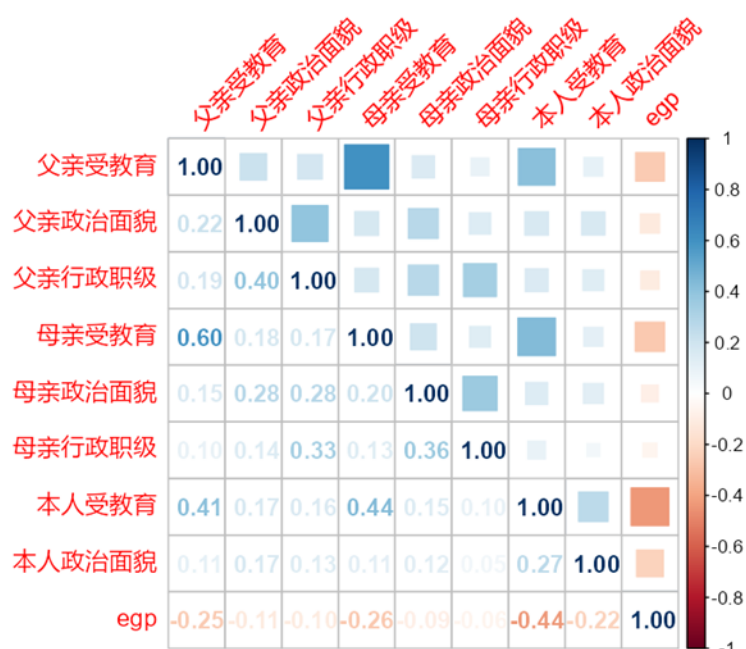


图 1 相关系数热力图

考虑到变量之间的相互作用，进一步做条件独立性的检验，这里使用 Cochran-Mantel-Haenszel 检验，虽然各层之间相合方向不一致可能会导致该检验更容易出现

不显著的情况，但由于没有找到高维列联表的相合一致性检验程序，这里只能接受这一误差，从 CMH 检验的计算方法上来看，如果在不考虑各层相合一致性的情况下出现显著地不独立，那么在相合一致性的情况下也会显著地不独立。

表 3 阶层~父辈行政职级+性别列联表

		egp	1	2	3	4	5	7	8	9	10
性别	父辈行政职级										
1	0		810	686	401	312	620	177	701	1060	56
	1		66	34	19	12	8	6	13	18	0
	2		15	7	8	4	1	3	3	4	0
	3		6	3	0	1	0	1	0	2	0
2	0		496	800	818	194	598	21	269	679	39
	1		34	40	32	4	3	2	4	12	0
	2		13	15	5	1	1	0	0	1	0
	3		2	2	1	0	0	1	0	1	0

表 4 阶层~户口+性别列联表

		egp	1	2	3	4	5	7	8	9	10
性别	户口										
1	0		384	319	208	201	482	116	561	799	42
	1		513	411	220	128	147	71	156	285	14
2	0		236	394	447	147	461	11	228	537	34
	1		309	463	409	52	141	13	45	156	5

在控制了性别的情况下，CMH 检验给出的结果显示阶层和父辈行政职级、户口均非独立关系，但是考虑到影响阶层地位的因素极其复杂，只控制一种因素所得到的结果是较不可靠的，为此需要基于回归分析剥离各因素的贡献。

### 三、Logistic 回归

首先将所有变量放入方程中进行多元有序 logistic 回归（由于结果过长，这里不作展示），为了确定模型的有效性，使用 brant 检验考察平行线假设是否成立，结果显示 p 值趋近于零，这表明平行线假设不成立，考虑到各阶层之间的流动影响因素极其复杂，我们所选取的变量无法产生同等的效应是正常的。为此参考王冉（2022）的做法，将 EGP 阶层分类中的 1,2,3,4,5,7 阶层统一归为非体力劳动者阶层，8,9,10 阶层统一归为体力劳动者阶层，改为使用二元 logistic 模型，这样虽然缺乏了解释上的精细度，但可以提升结果的可靠性。此外考虑到父亲和母亲各项因素的相关性，参考郭未和于瑶（2023）的研究，使用父母平均受教育、父母最高政治面貌，父母最高行政职级来度量父辈因素的影响，回归结果如表 5 所示。

从基准回归结果可以看出，父辈受教育水平越高，子代成为体力劳动者的概率越小，而父辈行政职级和政治面貌的影响不显著；本人的受教育水平可以显著地防止自身成为体力劳动者阶层，且这一作用比父辈受教育水平的影响更强；相对于群众或共青团员来说，具有党员身份的个体更不可能成为体力劳动者，民主党派身份

的影响不显著；相较于男性，女性更不可能成为体力劳动者，相较于农业户口，具有非农户口的个体更不可能成为体力劳动者，年龄的系数显著，但其值较小，可以忽略年龄的影响。

表 5

	基准回归	东部地区	中部地区	西部地区	东北地区
截距项	1.593***	2.093***	1.234***	1.643***	1.260*
父辈受教育	-0.042***	-0.062***	-0.010	-0.061***	-0.033
父辈行政职级 1	0.151	0.196	0.574	-0.918	-0.044
父辈行政职级 2	-0.151	-0.211	1.111	0.501	-13.462
父辈行政职级 3	0.376	0.380	1.340		
父辈政治面貌 1	0.699	0.828	-11.848	2.070	
父辈政治面貌 2	-0.040	0.008	-0.078	-0.156	-0.099
本人受教育	-0.173***	-0.202***	-0.145***	-0.142***	-0.192***
本人政治面貌 1	-0.581	-0.104	-12.250	-0.964	-14.153
本人政治面貌 2	-0.796***	-0.765***	-0.759***	-0.786**	-1.124**
性别 2	-0.764***	-0.624***	-0.870***	-0.938***	-0.741***
户口 1	-0.379***	-0.355***	-0.607***	-0.362*	-0.148
年龄	0.007**	0.007+	0.005	-0.003	0.023*
Num.Obs.	9145	4476	2093	1667	728
Log.Lik.	-4767.640	-2179.667	-1183.009	-892.064	-380.000
RMSE	0.42	0.40	0.44	0.43	0.42

注：+  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , 下同

考虑到中国地区间的差异，我们进一步按照东中西和东北四个地区进行分组回归，结果同样列于表 5，其中西部和东北地区由于设计阵奇异，导致部分系数不可估，这是样本无对应属性导致的。分组回归结果中值得注意的是中部地区和东北地区的父辈受教育水平影响不显著，这可能与对应地区的文化传统或是经济结构有关；父辈行政职级和政治面貌的影响依旧不显著，这可能说明对于是否成为体力劳动者来说，这两个因素的影响不大，当然也存在另一可能，由于样本中父辈具有行政职级或民主党派、党员身份较少，导致了样本选择偏误；户口类型在东北地区的影响

不显著，这可能是东北旧重工业基地的经济结构导致了居民工作选择较为单一，不受户口类型影响。

为了进一步提供更精细的解释，接下来依旧采用二元 logistic 模型，考察各因素对个体落入每一个 EGP 阶层的概率的影响，结果如表 6 所示。可以看到，父辈受教育水平显著提升了子代成为专业技术人员、行政管理者、企业管理者和大资产者的概率（即阶层 1,2），不过本人的受教育水平对此发挥着更大的影响，相反地，父辈和个体的受教育水平降低了个体成为体力劳动者的概率（即阶层 8,9,10），此外本人的受教育水平越高，个体成为小资产者（阶层 4,5）的概率也越低。值得注意的是，父辈的政治面貌对每个阶层的影响都不显著，而相较于没有行政职级的父辈，若父辈职级达到了科级，子代却不太可能成为一个低级技术人员、行政管理者、小企业管理者或小资产者（阶层 2,5），若父辈达到了司局级，个体成为低级技术人员或体力雇员监督者（阶层 7）的可能性会增加；相较于群众或共青团员，具有党员身份的个体更有可能进入阶层 1,2,7，而不太可能进入阶层 4,5,8,9，从党员的推荐机制上来看，这一结论很符合现实；相较于男性，女性进入阶层 1,4,7,8,9 的可能性较低，进入阶层 2,3,5 的可能性较高，这可能是因为阶层 8,9 属于体力劳动，女性较难适应该工作，而阶层 2,3,5 中属于办公室工作人员或服务业的工作类别更适配女性；相较于农业户口，非农户口的个体更容易进入阶层 2,3，而不容易成为阶层 5,8,9。

表 6

	阶层 1	阶层 2	阶层 3	阶层 4	阶层 5	阶层 7	阶层 8	阶层 9	阶层 10
截距项	-6.01***	-3.64***	-1.79***	-1.76***	-1.26***	-2.67***	0.62**	-0.09	-5.33***
父辈受教育	0.02*	0.04***	-0.01	0.00	-0.01	0.02	-0.06***	-0.04***	-0.04
父辈行政职级 1	0.16	-0.27+	-0.05	0.20	-0.81*	0.07	0.29	0.03	-13.51
父辈行政职级 2	-0.12	-0.32	-0.09	0.58	-1.04	0.48	0.18	-0.30	-13.39
父辈行政职级 3	0.06	-0.21	-1.16	0.47	-11.87	1.45+	-10.90	0.70	-13.78
父辈政治面貌 1	-0.40	-0.19	0.01	1.15	-11.91	-13.03	0.08	0.88	-13.91
父辈政治面貌 2	-0.02	-0.04	0.12	-0.08	0.07	-0.08	-0.07	0.05	-0.85
本人受教育	0.26***	0.16***	0.01	-0.03+	-0.08***	-0.01	-0.10***	-0.14***	-0.12***
本人政治面貌 1	0.78	0.59	-12.23	-11.82	-0.22	-12.81	0.08	-0.81	-14.83
本人政治面貌 2	0.51***	0.26**	0.05	-1.02***	-1.05***	0.40*	-0.62***	-0.77***	0.12
性别 2	-0.35***	0.46***	1.03***	-0.36***	0.15*	-1.88***	-0.94***	-0.38***	-0.15
户口 1	-0.01	0.22***	0.45***	-0.05	-0.39***	-0.05	-0.52***	-0.17*	-0.20

	阶层 1	阶层 2	阶层 3	阶层 4	阶层 5	阶层 7	阶层 8	阶层 9	阶层 10
年龄	0.02***	-0.01***	-0.02***	-0.01*	0.01***	-0.02*	-0.02***	0.01***	0.05***
Num.Obs.	9145	9145	9145	9145	9145	9145	9145	9145	9145
RMSE	0.34	0.36	0.34	0.23	0.33	0.15	0.30	0.38	0.10

## 四、研究展望

本文所作分析较为简单粗糙，存在不少值得改进的地方，如 EGP 阶层分类框架依赖于个体被调查时的工作，这就导致无工作的个体无法进入样本，可以考虑使用更为全面的阶层分类框架或是基于 Heckman 选择误差模型改进这一问题。此外，本文在理论分析上有严重欠缺，没有充分利用已有的代际传递理论，这导致本文所选取的变量完全出于直觉，没有深远意义，变量间的关系也未得到全面地研究，就因果推断来说，本文是极为欠缺的。



## 参考文献

- [1] 郭未, 于瑶. 曾梦想仗剑走天涯: 小镇青年人力资本、社会阶层的代际与空间流动[J]. 浙江工商大学学报, 2023(5): 153-165.
- [2] 季洋. 家庭背景对子女教育程度与阶层流动的影响[D]. 华中师范大学, 2021.
- [3] 侯利明, 秦广强. 中国 EGP 阶层分类的操作化过程——以中国综合社会调查(CGSS)数据为例[J]. 社会学评论, 2019, 7(2): 16-26.
- [4] 张宛丽. 中国社会阶级阶层研究二十年[J]. 社会学研究, 2000(1): 24-39.
- [5] 从阶层分类到阶层量化: 社会阶层量化方法新范式的一种探索[J].
- [6] Ganzeboom H B G, Treiman D J. Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations[J]. Social Science Research, 1996, 25(3): 201-239.
- [7] 王冉. 个人职业地位获得影响因素及提升策略研究[D]. 郑州大学, 2022.
- [8] 刘欣. 协调机制、支配结构与收入分配: 中国转型社会的阶层结构[J]. 社会学研究, 2018, 33(1): 89-115+244.

## 附录：代码展示

```
library(dplyr)
library(haven)
library(purrr)
library(ggplot2)
library(naniar)
library(MASS)
library(modelsummary)
library(corrplot)
library(flextable)
#转 EGP
CGSS_to_EGP <- function(CGSSlist,years){
  library(dplyr)
  library(occupar)
  library(haven)
  #预处理
  CGSS_select <- function(CGSSlist,years){
    #occ2008 <- c("c8codeisco","c2","c6","?","c5")
    #2008 年无管理人数变量
    occ2010 <- c("risco881","a58","a59aa","a59fa","a59a")
    occ2011 <- c("iscorp1","a58","a59aa","a59fa","a59a")
    occ2012 <- c("iscorp1","a58","a59aa","a59fa","a59a")
    occ2013 <- c("iscorp1","a58","a59aa","a59fa","a59a")
    occ2015 <- c("a59disco88","a58","a59a1","a59f1","a59a")
    occ2017 <- c("isco08_a59","a58","a59a1","a59f1","a59a")
    occ2018 <- c("isco08_a59d","a58","a59a1","a59f1","a59a")
    occ2021 <- c("isco08_a59d","A58","A59a1","A59f1","A59a")
    #由第一份数据创建数据框
    if (years[1]<2017){
      occ <- paste("occ",years[1],sep = "")
      CGSS <- CGSSlist[[1]] %>%
```

```

      .[,eval(parse(text=eval(parse(text = "occ"))))] %>%
      data.frame(.,year=years[1]) %>%
      `colnames<-`(c("isco","工作性质","雇人数","管人数","自雇","year"))
    }else{
      occ <- paste("occ",years[1],sep = "")
      CGSS <- CGSSlist[[1]] %>%
      .[,eval(parse(text=eval(parse(text = "occ"))))] %>%
      isco08to88() %>%
      data.frame(.,CGSSlist[[1]][,eval(parse(text=eval(parse(text
      "occ"))))] %>%
      `colnames<-`(c("isco","工作性质","雇人数","管人数","自雇","year"))
    }
    #将后续数据并入第一份数据框中
    for (i in 2:length(years)){
      if (years[i]<2017){
        occ <- paste("occ",years[i],sep = "")
        CGSS <- CGSSlist[[i]] %>%
        .[,eval(parse(text=eval(parse(text = "occ"))))] %>%
        data.frame(.,year=years[i]) %>%
        `colnames<-`(c("isco","工作性质","雇人数","管人数","自雇
        ", "year")) %>%
        bind_rows(.,CGSS)
      }else{
        occ <- paste("occ",years[i],sep = "")
        CGSS <- CGSSlist[[i]] %>%
        .[,eval(parse(text=eval(parse(text = "occ"))))] %>%
        .[[1]] %>%
        isco08to88() %>%
        data.frame(.,CGSSlist[[i]][,eval(parse(text=eval(parse(text
        "occ"))))] %>%

```

```

`colnames<-`(c("isco"," 工作性质 "," 雇人数 "," 管人数 "," 自雇
","year")) %>%
  bind_rows(.,CGSS)
}
}
return(CGSS)
}
#转换 EGP
to_EGP <- function(CGSS){
  toEGP_1 <- function(CGSS){
    isco88 <- zap_labels(CGSS[[1]])
    is_farmer <- zap_labels(CGSS[[2]])
    egp <- rep(NA,length(isco88))
    ## CLASS 1 Higher service
    egp[isco88 %in% c(1000, 2000, 1100, 2100, 1110, 2110, 1120, 2111, 2112,
2113, 2114, 2120, 2121, 1200, 2122, 1210, 2130, 1220,
2131, 1222, 1223,
2140, 1224, 2141, 1225, 2142, 1226, 2143, 1227, 2144,
1228, 2145, 1229,
2146, 1230, 2147, 1231, 1232, 2149, 1233, 2200, 3143,
1234, 2210, 3144,
1235, 2211, 1236, 2212, 1237, 2213, 1239, 2220, 2221,
1250, 2222, 1251,
2223, 2224, 2229, 2310, 2350, 2351, 2352, 2400, 2411,
2420, 2421, 2422,
2429, 2440, 2441, 2442, 2443, 2445)] <- 1 ## CLASS 1
    ## CLASS 2 Lower service
    egp[isco88 %in% c(3000, 3100, 3110, 3111, 1130, 3112, 1140, 3113, 1141,
3114, 1142, 3115, 5121, 1143, 3116, 3117, 3118, 3119,
2132, 3120, 2139,

```

```

3121, 3122, 3123, 3130, 3131, 3132, 3133, 3139, 5150,
3140, 5151, 2148,
3141, 5152, 3142, 3145, 3150, 3151, 3152, 1240, 3200,
3210, 3211, 1252,
3212, 1300, 3213, 1310, 2230, 3220, 2300, 3221, 1312,
3222, 1313, 2320,
3223, 1314, 2321, 3224, 1315, 2322, 3225, 1316, 2323,
3226, 1317, 2330,
3227, 1318, 2331, 3228, 1319, 2332, 3229, 2340, 3240,
2359, 3241, 3242,
2410, 2412, 2419, 3400, 3410, 3411, 2430, 3412, 2431,
3413, 2432, 3414,
3415, 3416, 3417, 3419, 2444, 3420, 3421, 2446, 3422,
2450, 3423, 2451,
3429, 2452, 2453, 3431, 2454, 3432, 2455, 2460, 3434,
3440, 3441, 3442,
3443, 3444, 3449, 3450, 3451, 3470, 3471, 3472, 3473,
3474,
3475)] <- 2 ## CLASS 2

## EGP CLASS 3 Routine clerical/sales
egp[isco88 %in% c(4000, 4100, 4122, 9100, 4110, 9110, 4111, 9111, 4112, 9112,
4113, 9113, 4114, 4115, 4120, 4121, 3230, 3231, 3232,
3300, 3310, 3320, 3330,
3340, 3430, 3433, 3439, 3460, 3480, 5000, 5100, 5110,
5111, 5112, 5113, 5120, 4130, 5131,
4131, 4132, 5133, 4133, 4140, 4141, 4143, 4144, 4190,
4200, 4210, 4211,
4212, 4213, 4214, 4215, 4220, 4221, 5200, 4222, 5210,
4223, 5220, 5230)] <- 3

```

```

## CLASS 7 Manual Manual foreman
egp[isco88 %in% c(3452, 7510)] <- 7 ## CLASS 7

## CLASS 8
egp[isco88 %in% c(7000, 7120, 5122, 7124, 7129, 7130, 7132, 5140, 7133, 5141,
7134, 5143, 7136, 8150, 7137,
                        8151, 7140, 8152, 7141, 8153, 8154, 8155, 5161, 7200,
8159,
                        5162, 7210, 8160, 7211, 8161, 5164, 7212, 8162, 7213,
8163, 7214, 8170, 7215, 8171, 7216,
                        8172, 7220, 7221, 7222, 7223, 7224, 7230, 7231, 7232,
7233, 7240, 7241, 7242,
                        7243, 7244, 7245, 7300, 7310, 7311, 7312, 7313, 7323,
7324, 7340, 7341, 7342, 7343, 7344,
                        7345, 7346, 7400, 7410, 7411, 7412, 7413, 7414, 7415,
7416, 7420, 7422, 8311,
                        7423, 7430, 7433, 7434, 7435, 7436, 8332, 7437, 8333,
7440, 7441, 7442, 7500, 7520)] <- 8

## CLASS 9
egp[isco88 %in% c(8000, 9000, 7100, 8100, 7110, 8110, 7111, 8111, 7112, 8112,
7113, 8113, 8120, 9120, 7121,
                        8121, 9130, 7122, 8122, 9131, 5123, 7123, 8123, 9132,
5130,
                        8124, 9133, 8130, 9140, 5132, 8131, 9141, 7131, 8139,
9142, 5139, 8140, 9150, 8141, 9151,
                        8142, 9152, 4142, 5142, 7135, 8143, 9153, 9160, 5149,
9161, 9162, 9200, 7142,
                        5160, 7143, 5163, 9300, 9310, 5169, 9311, 9312, 9313,
9320, 8200, 9321, 8210, 9322, 8211,
                        9330, 8212, 9331, 8220, 9332, 8221, 9333, 8222, 8223,
8224, 7234, 8229, 8230,

```

```

8231, 8232, 8240, 8250, 8251, 8252, 8253, 8260, 8261,
8262, 7320, 8263, 7321, 8264, 7322,
8265, 8266, 8269, 7330, 8270, 7331, 8271, 7332, 8272,
8273, 8274, 8275, 8276,
8277, 8278, 8279, 8280, 8281, 8282, 8283, 8284, 8285,
8286, 8290, 8300, 7421, 8310, 8312,
7424, 8320, 8321, 7431, 8322, 7432, 8323, 8324, 8330,
8334, 8340, 8400,
7530)] <- 9

## CLASS 10 farm workers
egp[isco88 %in% c(6000, 6100, 6110, 6111, 6112, 6113, 6114, 6120, 6121, 6122,
6123, 6124, 6129, 6130, 6134,
6140, 6141, 6142, 6150, 6151, 9210, 6152, 9211, 6153,
9212,
6154, 9213, 8331)] <- 10

## CLASS 11 farmers
egp[isco88 %in% c(1221, 6131, 6132, 6133, 6200, 6210, 1311)] <- 11
egp[is_farmer %in% c(2,3)] <- 11
return(egp)
}#第一阶段转换
toEGP_2 <- function(egp,CGSS){
  isco88 <- zap_labels(CGSS[[1]])
  n_employees <- zap_labels(CGSS[[3]])
  n_manage <- zap_labels(CGSS[[4]])
  self_employed <- zap_labels(CGSS[[5]])
  n_employees[is.na(n_employees)] <- 0
  n_manage[is.na(n_manage)] <- 0
  self_employed[!self_employed %in% c(1,2,8)] <- 0
  self_employed[self_employed %in% c(1,2,8)] <- 1
  #常规非体力劳动者变更为下层服务阶层

```

```

    egp[egp==3&(n_employees>0|n_manage>0)] <- 2
    #自营业主的生成
    egp[(egp %in% 2:3)&(self_employed==1)&(isco88 %in%
c(1310:1319,3400:3439,4000:5230))] <- 4
    #独立自营的生成
    egp[(egp %in% 7:9)&(self_employed==1)&!isco88 %in% c(9300:9333)] <- 5
    #技术体力劳动者变更为体力工人监督者
    egp[(egp==8)&(n_employees>0|n_manage>0)] <- 7
    #农业劳动者变更为农业经营者
    egp[(egp==10)&(self_employed==1)] <- 11
    #自营业主变更为独立自营
    egp[(egp==4)&!(n_employees>0|n_manage>0)] <- 5
    #独立自营变更为自营业主(意义不明)
    egp[(egp==5)&(n_employees>0|n_manage>0)] <- 4
    #上层服务人员的变更
    egp[(egp %in% c(2:4))&(n_employees>10|n_manage>10)] <- 1
    #合并 10 和 11
    egp[egp==11] <- 10
    return(egp)
}#第二阶段修正
EGP <- toEGP_1(CGSS) %>%
  toEGP_2(.,CGSS) %>%
  bind_cols(.,CGSS) %>%
  `colnames<-`(c("EGP","isco","工作性质","雇人数","管人数","自雇","year"))
  return(EGP)
}
EGP <- CGSS_select(CGSSlist,years) %>%
  to_EGP()
  return(EGP)
}
d2017 <- read_dta("D:/数据与信息/社会数据/CGSS/2017/cgss2017.dta")

```



```
d2018 <- read_dta("D:/数据与信息/社会数据/CGSS/2018/cgss2018.dta")
d2021 <- read_dta("D:/数据与信息/社会数据/CGSS/2021/cgss2021.dta") %>%
  rename(a89b=A89b,a89c=A89c,a89f=A89f,a90b=A90b,a90c=A90c,a90f=A90f,
         a7a=A7a,a10=A10,a2=A2,a31=A3_1,a18=A18)
CGSSlist <- list(d2017,d2018,d2021)
years <- c(2017,2018,2021)
EGP <- CGSS_to_EGP(CGSSlist,years)
#获取其他变量
d2017 %>%
  rename(provinces=s41) %>%
  mutate(provinces=as_factor(provinces)) %>%
  list(d2021,d2018,..) %>%
  map_dfr(function(x){ x %>%
    dplyr::select(c("a89b","a89c","a89f",
                    "a90b","a90c","a90f",
                    "a7a","a10","a2","a31",
                    "provinces","a18"))}) %>%
  mutate(egp=EGP$EGP,year=EGP$year,isco=EGP$isco) %>%
  mutate(across(everything(),zap_labels)) %>%
  `colnames<-`(c("父亲受教育","父亲政治面貌","父亲行政职级",
                  "母亲受教育","母亲政治面貌","母亲行政职级",
                  "本人受教育","本人政治面貌","性别","出生日期",
                  "地区","户口","egp","year","isco")) %>%
  mutate(across(contains("受教育"),~case_when(.x<=2~1,
                                                .x<=3~6,
                                                .x<=4~9,
                                                .x<=8~12,
                                                .x<=12~16,
                                                .x<=13~20,
                                                NA~NA,
                                                T~0)),
```

```

    across(contains("政治面貌"),~case_when(.x<=2~0,
                                              .x==3~1,
                                              .x==4~2)),
    across(contains("行政职级"),~case_when(.x<=2~0,
                                              .x<=4~1,
                                              .x<=6~2,
                                              .x==7~3,
                                              .x==NA~NA)),

    年龄=year-出生日期,
    户口=case_when(户口%in%c(1,3)~0,
                   户口%in%c(2,4,5,6,7)~1)) %>%

mutate(父辈受教育=(父亲受教育+母亲受教育)/2,
       父辈行政职级=pmax(父亲行政职级,母亲行政职级),
       父辈政治面貌=pmax(父亲政治面貌,母亲政治面貌)) %>%

mutate(across(-c("父辈受教育","本人受教育","年龄","父亲受教育","母亲受教育"),as.factor)) %>%

dplyr::select(-c("出生日期","year")) ->ori_data
rm(d2017,d2018,d2021,CGSSlist)
gc()
ori_data %>%

dplyr::select(-"父辈受教育") %>%

filter(!is.na(isco)) %>%

mcar_test()

nona_data <- ori_data %>%

na.omit()

nona_data_num <- map_dfr(nona_data,as.numeric)

#列联表分析

ftable(data = nona_data,egp~性别+父辈行政职级)

xtabs(data = nona_data,~父辈行政职级+egp+性别) %>%

mantelhaen.test()

ftable(data = nona_data,egp~性别+户口)

```

```

xtabs(data = nona_data,~户口+egp+性别) %>%
  mantelhaen.test()
#相关系数
corr <- cor(nona_data_num[c(1:8,12)],method = "kendall")
corr_p <- cor.mtest(nona_data_num[c(1:8,12)],method = "kendall")
corrplot(corr,p.mat = corr_p$p,method = "square",type = "upper", tl.pos = 't',tl.srt = 45)
corrplot(corr,p.mat = corr_p$p,method = "number",add = T,type = "lower", tl.pos = 'n',
cl.pos = 'n')
#logistic 回归
model0 <- polr(egp~父亲受教育+父亲行政职级+父亲政治面貌+母亲受教育+母亲
行政职级+母亲政治面貌+
              本人受教育+本人政治面貌+性别+户口+年
龄,data=nona_data,method = "logistic",Hess = T)
brant::brant(model0)
modelsummary(model0,stars = T,output = "flextable")
nona_data2 <- nona_data %>%
  mutate(阶层=case_when(egp%in%c(1,2,3,4,5,7)~1,
                        egp%in%c(8,9,10)~2)) %>%
  mutate(across("阶层",as.factor)) %>%
  arrange(阶层)
model1 <- glm(阶层~父辈受教育+父辈行政职级+父辈政治面貌+
              本人受教育+本人政治面貌+性别+户口+年
龄,data=nona_data2,family = binomial(link = "logit"))
modelsummary(model1,stars = T,statistic = NULL,output = "flextable",gof_omit =
"IC|F") %>%
  align(align = "center",part = "all") %>%
  save_as_docx(path="D:/预删除文件夹/大三上/应用回归分析/临时.docx")
#东中西分组
east <- c("北京市","天津市","河北省","上海市","江苏省","浙江省","福建省","山东
省","广东省","海南省")
center <- c("山西省","安徽省","江西省","河南省","湖北省","湖南省")

```

```

west <- c("内蒙古自治区","广西壮族自治区","重庆市","四川省","贵州省","云南省",
,"西藏自治区","陕西省","甘肃省",
      "青海省","宁夏回族自治区","新疆维吾尔自治区")
northeast <- c("辽宁省","吉林省","黑龙江省")
map2(list(nona_data2,nona_data2,nona_data2,nona_data2),list(east,center,west,northeast),function(x,y){
  x %>%
    filter(地区 %in% y) %>%
    glm(阶层~父辈受教育+父辈行政职级+父辈政治面貌+
      本人受教育+本人政治面貌+性别+户口+年龄,data=.,family =
binomial(link = "logit"))
}) %>%
  modelsummary(stars = T,statistic = NULL,output = "flextable",gof_omit =
"IC|F") %>%
  align(align = "center",part = "all") %>%
  save_as_docx(path="D:/预删除文件夹/大三上/应用回归分析/临时.docx")
table(nona_data2[nona_data2$地区 %in% west,]$父辈行政职级)
#阶层分组
map2(list(nona_data,nona_data,nona_data,nona_data,nona_data,nona_data,nona_data,nona_data,nona_data),
c(1,2,3,4,5,7,8,9,10),function(x,y){
  x %>%
    mutate(阶层=if_else(egp==y,1,0)) %>%
    mutate(across(contains("阶层"),as.factor)) %>%
    glm(阶层~父辈受教育+父辈行政职级+父辈政治面貌+
      本人受教育+本人政治面貌+性别+户口+年龄,data=.,family=binomial(link='logit'))
  }) %>%
  modelsummary(stars = T,statistic = NULL,fmt = fmt_decimal(digits=2),output =
"flextable",gof_omit = "IC|F") %>%
  align(align = "center",part = "all") %>%

```

```
save_as_docx(path="D:/预删除文件夹/大三上/应用回归分析/临时.docx")
```