作业02：二手数据收集

中国家庭追踪调查CFPS

任课教师：胡华平

学生姓名：\_\_\_\_\_\_\_\_\_\_\_\_\_\_；学生学号：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_；专业班级：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# 作业发布

## 作业材料

（1）**作业目标**：二手数据收集；文献资料检索；中国家庭追踪调查CFPS；面板数据。

（2）**作业任务**：

* **题目及答题纸**：.docx文件（点击下载<homework02-data-collect.docx>）

（3）**配套数据**：

* **配套数据1**：CFPS2018年数据集.xlsx文件（点击下载[ex02-data-cfps2018-balance.xlsx](resource/ex02-data-cfps2018-balance.xlsx)）。
* **配套数据2**：CFPS2020年数据集.xlsx文件（点击下载[ex02-data-cfps2020-balance.xlsx](resource/ex02-data-cfps2020-balance.xlsx)）。
* **配套数据2**：变量定义.xlsx文件（点击下载[ex02-label-balance.xlsx](resource/ex02-label-balance.xlsx)）。

**要求**：使用微软Office Word软件，在台式机或笔记本电脑上，在word文档中正确且美观地输入相关步骤及公式。（提示：公式输入可以使用Mathtype软件。）

## 截止时间

**作业发布时间**：2025-03-20（周四）24:00:00

**提交截止时间**：2025-03-31（周一）24:00:00

## 作业提交

（1）根据作业要求，完成Office Word电子文档一份（注意不能是wps文档），提交前请将文件命名为下述格式：homework02\_word\_张三\_2019000001.docx。

（2）根据作业要求，将检索下载到的1篇重要期刊文章(孙, 吴, 和 张 2024)全文的pdf格式，命名为homework02\_pdf\_cfps文献01\_张三\_2019000001.pdf，并提交到作业系统。

（3）根据作业要求，将检索下载到的2篇CFPS实证期刊文章全文的pdf文件分别提交到作业系统。其中，一篇中文CFPS实证期刊pdf文件命名为homework02\_pdf\_cfps文献02\_中文\_张三\_2019000001.pdf；一篇英文CFPS实证期刊pdf文件命名为homework02\_pdf\_cfps文献03\_英文\_张三\_2019000001.pdf。

（4）根据实验要求，完成CFPS面板数据集的.xlsx文件的EXCEL相关操作，保存并提交1份EXCEL文件.xslx文件，提交前请将文件命名为下述格式：homework02\_cfps面板数据\_张三\_2019000001.xlsx。

# 作业任务

## 任务1：下载1篇重要期刊文献

（1）请在校园网环境下，登陆**中国知网**<https://kns.cnki.net/>，查找并下载下列发表的一篇期刊文献(孙, 吴, 和 张 2024)：孙妍, 吴琼, 张春泥. 中国家庭追踪调查:设计理念及数据运用问题[J]. 调研世界, 2024(1): 4–14. 请你下载该文章全文的pdf格式（不能是CAJ格式），并命名为homework02\_pdf\_cfps文献01\_张三\_2019000001.pdf，作业提交时请上传该pdf文件。

**答**：本题不需要作答，完成相关操作，作业提交时按要求上传本pdf文件即可。

（2）请你阅读上述文献(孙, 吴, 和 张 2024)，并回答如下几个问题：

a）中国家庭追踪调查的2020年第六轮调查数据集CFPS2020，是截面数据还是面板数据？

**答：**

b）中国家庭追踪调查的2018年第五轮调查数据集CFPS2018，能否与2020年的第六轮调查数据集CFPS2020一起构造得到一份面板数据？

**答：**

c）通过阅读上述文献(孙, 吴, 和 张 2024)，并观察其中的“表2 CFPS 家庭层面应答及追踪率”（见下面 [图 1](#fig-cfps-respond-ratio)）。请你分析为什么截面应答率和跨轮追踪率为什么呈现逐轮（逐年）下降的总体趋势？

|  |
| --- |
| 图 1: CFPS 家庭层面应答及追踪率 |

**答：**

## 任务2：下载2篇期刊文献

（1）请在校园网环境下，登陆**中国知网**<https://kns.cnki.net/>，查找并下载一篇已经发表的**中文**期刊文献，要求：a)该期刊文献利用CFPS数据进行实证建模分析；b）检索关键词请包含“CFPS”；c）该文献发表所在期刊应该为**北大中文核心来源期刊**或者**CSSCI来源期刊**。请你下载该文章全文的pdf格式（不能是CAJ格式），并命名为homework02\_pdf\_cfps文献02\_中文\_张三\_2019000001.pdf，作业提交时请上传该pdf文件。

**答**：本题不需要作答，完成相关操作，作业提交时按要求上传本pdf文件即可。

（2）请在校园网环境下，登陆**西北农林科技大学图书馆网站**<https://lib.nwsuaf.edu.cn/>，进入【文献资源】》》【电子资源】》》【Wiley-Blackwell电子期刊】（或者使用Springer电子期刊<https://link.springer.com/>），查找并下载一篇已经发表的**外文**期刊文献，要求：a)该期刊文献利用CFPS数据进行实证建模分析；b）检索摘要或关键词请包含“China Family Panel Studies (CFPS)”。请你下载该文章全文的pdf格式，并命名为homework02\_pdf\_cfps文献03\_外文\_张三\_2019000001.pdf，作业提交时请上传该pdf文件。

**答**：本题不需要作答，完成相关操作，作业提交时按要求上传本pdf文件即可。

（3）请你把上述中文文章和英文文章，整理为标准的参考文献格式。

**答**：

[1]

[2]

|  |  |
| --- | --- |
|  | **温馨提示**：参考文献格式有不同规范，请你参照如下格式体例：  [1] 孙妍, 吴琼, 张春泥. 中国家庭追踪调查:设计理念及数据运用问题[J]. 调研世界, 2024(1): 4–14. |

（4）请阅读你下载好的上述中文文章和英文文章，考虑到它们都用到了CFPS数据集进行统计分析，请你找到其中任意一张**描述性统计**的表格，请截图该表格到下面空白处，并简要解释这张表格的基本结论和信息。

**答**：截图并分析如下：

## 任务3：CFPS数据集操作

本此作业材料中的**配套资料1**和**配套资料2**，分别提供了中国家庭追踪调查CFPS在2018年和2020年两个轮次的**家庭经济情况**部分的数据集（.xlsx）、变量定义（.xlsx）和调查问卷（.pdf）。

数据集主要考虑了家庭经济情况的一些变量，包括家庭收入、家庭支出、家庭财产、家庭负债等。为了操作方便，我们仅保留了**陕西省**居民户的数据样本，其中2018年的数据集共有314户，2020年的数据集共有282户。

数据集中的具体变量和标签定义如下：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 表 1: CFPS家庭经济情况的变量定义   | index | name | label | | --- | --- | --- | | 1 | fincome1 | 全部家庭纯收入 | | 2 | fincome1\_per | 家庭人均收入 | | 3 | expense | 家庭总支出 | | 4 | mortage | 房贷支出 | | 5 | total\_asset | 家庭净资产 | | 6 | savings | 现金和存款总值 | | 7 | house\_debts | 总房贷 | |

请你根据上述材料，把两个年度轮次的数据集，整理为一个面板数据集。具体地：

（1）整理出一份**扁数据形式**（wide format）面板数据集，保存为一个xlsx表单SHEET，并命名为wide。

**答**：本题不需要作答，完成相关操作，作业提交时按要求上传本xlsx文件即可。

|  |  |
| --- | --- |
|  | **要求**：变量命名为类似如dt18\_fincome1/dt20\_fincome1等，其中前缀命名dt18和dt20分别代表2018年和2020年的数据集变量。 |

（2）整理出一份**长数据形式**（long format）面板数据集，保存为一个xlsx表单SHEET，并命名为long。

**答**：本题不需要作答，完成相关操作，作业提交时按要求上传本xlsx文件即可。

|  |  |
| --- | --- |
|  | **要求**：合并后面板数据集（long format）的数据年度变量请命名为Year。 |

最终将上述表单SHEET两个保存为一个xlsx数据文件，并命名为homework02\_cfps面板数据\_张三\_2019000001.xlsx，作业提交时请上传该xlsx文件。

|  |  |
| --- | --- |
|  | **温馨提示**：   * 请认真分析两个2018年和2020年两个轮次数据集的变量名和变量标签涵义。 * 上述2018年和2020年两个数据集拥有一些**共有变量**，同时也有不同的变量。 * 由于某些原因，2018年和2020年两个数据集中的居民户可能发生了部分改变。 |