

# CHFS 数据变量命名规则说明

## 一、数据清理

### 1、清理过程

数据采集回来后，中心会对所有数据进行初步处理，主要处理包括：删除由于访员严重臆答、作弊的无效样本，删除无效变量，删除敏感数据；校正人为导致的重复样本编号，校正访员主动报备的人为误操作；合并追踪和新访数据，拆分家庭和个人数据，拆分多选；加注标签，加注问卷类型；清理备注和其他选项，数值题插值及相关规则确定等。

经过初步处理后数据生成可使用的版本。在数据使用过程中如再发现极值或异常值，中心会进行二次录音核查确认。如果没有录音，则通过创建模型处理极值和异常值，从而更新数据版本并告知用户。

### 2、数据存储

数据清理结束，生成的中国家庭金融调查数据存储在以下 3 个数据集里面。

1. 数据集文件名中含有“**hh**”，代表问卷中家庭部分的数据，例如：资产与负债，家庭的支出与收入，金融知识、基层治理与主观评价等；
2. 数据集文件名中含有“**ind**”，代表问卷中个人部分的数据，例如：人口统计特征（部分），个人工作及收入信息，保险与保障，家庭成员教育等；
3. 数据集文件名中含有“**master**”，代表是地区数据，具体包括调查的省份信息和每次调查数据的权重变量。

除此之外，原始数据使用格式对应为 dta 格式（可用 stata13.0 或以上版本打开）。

## 二、变量命名规则

中国家庭金融调查问卷数据的变量名由首位的字母及后面四位数字构成，不同的首字母对应问卷的不同部分（如，b-农业/工商业；d-金融资产），一些特别的标识性变量（ID 变量、城乡、省份等）则根据它们的含义被重新命名以便使用。部分变量在原变量名后加上后缀字

母 it，表示对前一个问题的数值范围追问。问卷中的每个问题前均给出了对应的变量名称，需要特别注意的变量及其命名规则说明如下：

## 1、ID 变量

ID 变量分为家庭变量（hhid）和个体变量（pline）。

hhid 是标识家庭的变量。每一个 hhid 代表一个家庭；hhid 可唯一识别家庭，hhid 后四位是调查年份，例如：hhid\_2011 代表的是 2011 年访问的样本编号。pline 是标识每个家庭中家庭成员的变量。每一个 pline 代表一个家庭成员；hhid 和 pline 可唯一识别个体，同一个家庭同一个家庭成员不同时期的 pline 保持不变。

## 2、权重变量（swgt）

在我们的抽样设计下，由于每户家庭被抽中的概率不同，因此每户家庭代表的中国家庭数量也就不同。在推断总体的时候，需要通过权重的调整来真实准确地反映每户样本家庭代表的家庭数量，以获得对总体的正确推断。中国家庭金融调查的所有计算结果都经过抽样权重的调整。在 master 数据集中含有权重变量：swgt。

其抽样权重的计算方法如下：根据每阶段的抽样分别计算出调查市县被抽中的概率  $P_1$ 、调查社区（村）在所属区县被抽中的概率  $P_2$  以及调查样本在所属社区（村）被抽中的概率  $P_3$ ，分别计算出三阶段的抽样权重  $W_1=1/P_1$ 、 $W_2=1/P_2$ 、 $W_3=1/P_3$ ，最后得到该样本的抽样权重为  $W=W_1 \times W_2 \times W_3$ 。

即  $swgt = \text{这个社区被抽中概率的倒数}$ 。

为得出更准确结论，我们建议数据使用者在分析时使用抽样权重变量“swgt”。

## 3、变量中的其他选项

在数据变量单选题或多选题中，一般情况下最后 1 个选项为“其他（请注明）”。若受访者选择了该选项，对应会多问 1 道填空题。在 2011 年和 2013 年家庭金融调查数据中，该新生成的变量命名格式为：原变量后加上后缀 p，例如：a4002p 表示受访者主要关注的信息是其他相关信息。

## 4、循环问题

对于所有循环询问的问题，命名规则为在原变量名后加上后缀\_#；#代表第#次循环。例如：c2003\_1 表示第一套房子的建筑面积；c2003\_2 则表示第二套房子的建筑面积。

除此之外，问卷中还包含其他由于多选题选项加载而产生的循环，如 c5005\_1 表示 c5002 选项 1 农用土地的加载，c5005\_2 表示选项 2 农村集体建设用地的加载。

## 5、多项选择题

对于多项选择题，处理原则为将每一个选项转换为取值为 0 和 1 的哑变量。多项选择题分为两类：非循环多项选择题、循环多项选择题。

### (1) 非循环多项选择题

非循环多项选择题的命名规则为在原变量名后加上后缀\_\*\_mc；\*代表第\*个选项。例如：b1004\_1\_mc 表示是否勾选 b1004 的第一个选项“粮食作物”；0 表示未选择，1 表示选择。

表 1 非循环多选题处理说明

varname_*_mc	变量信息
0	第*个选项未选择
1	第*个选项已选择

### (2) 循环多项选择题

循环多项选择题的命名规则为在原变量名后加上后缀\_\*\_#\_mc；\*代表第\*个选项，#代表第#次循环。例如，b3002a\_2\_1\_mc 表示在第一次循环时是否勾选第二个选项“估计贷款申请不会被批准”；0 表示未选择，1 表示选择。

表 2 循环多选题处理说明

varname_*_#_mc	变量信息
0	在第#次循环中第*个选项未选择
1	在第#次循环中第*个选项已选择

## 6、插值变量

为了解决数据缺失问题，对受访者没有回答具体数值而回答数值范围的题目（即变量名后缀含 it 的题目）进行了插值处理。对于插值后的变量，命名规则为原变量名加上后缀\_imp。

2011 年和 2013 年家庭金融调查数据具体截尾变量详见附表 1-1 和 2-1。

## 7、截尾处理

根据统计法规定，为保护受访者隐私，我们对收入和资产的极值进行了截尾处理。将收入或资产超过某一规定值的样本替换为该规定值，并同时给出一个截尾处理的哑变量，命名规则为变量名加上前缀 `sensor_`，即 `sensor_varname`。例如，我们将 2011 年家庭金融调查数据的年收入（`hh_income`）在 300 万以上的替换为 300 万，并同时给出一个哑变量 `sensor_hh_income`，0 表示未进行处理；1 表示进行了截尾处理。

表 3 截尾处理说明

<code>sensor_varname</code>	变量信息
0	未处理
1	已处理

2011 年和 2013 年家庭金融调查数据具体截尾变量详见附表 1-2 和 2-2。

## 8、其它特殊变量

### （1）关于 2011 年个别变量说明

为了便于研究，在 2011 年家庭金融调查数据中我们新生成了关于收入的两个变量，即 `labor_inc`（劳动性收入）和 `hh_income`（个人总收入）。

#### 1) `labor_inc`

`labor_inc` 是指个人税后工资薪金收入，即为劳动性收入（包括调查问卷中提到的最主要的工作以及第二职业工作税后工资薪金收入总和）。

#### 2) `hh_income`

`hh_income` 是代表家庭 2010 年的可支配收入总额，包括所有家庭成员全年的税后现金收入和实物收入总和。是根据 `hh data` 里的各项家庭收入以及 `ind data` 里的个人税后工资薪金收入（`labor_inc`）加总得到的。包括五部分：工资薪金收入、农业生产收入、工商业收入、投资性收入和转移性收入。

- 工资薪金收入：**受雇于他人获得的税后现金收入和实物收入。包括：税后工资收入、税后奖金收入、以及从单位获得的税后补贴收入。
- 农业生产净收入：**指从事农业生产获得的扣除农业生产成本以后的全部净收入，包括用于自己消费的农产品的市场价值。
- 工商业净收入：**指从事或参与工商业生产经营的家庭，在扣除生产经营成本和税收后，

从工商业生产经营项目中获得的属于该家庭部分的工商经营净收入。

4. **投资性收入**: 包括受访家庭拥有的房屋、汽车等出租给他人获得的出租性收入、银行存款利息收入、投资金融产品获得的净收入(包括投资股票、债券、基金、期货、黄金等金融产品获得的净收入, 投资保险获得的分红收入)。
5. **转移性收入**: 包括从政府获得的各种补贴性收入、家庭获得的征地拆迁补偿、退休工资、提取或使用的住房公积金、保险性收入(包括赔付的失业保险金、商业人寿保险赔付金、商业健康保险赔付金、商业养老保险赔付金、其他保险赔付金等)以及人情收入。

**说明**: hh\_income 收入为负是因为 2010 年家庭从事农业生产、工商业经营净收入等为负造成的。hh\_income 出现负数代表家庭收入呈亏损状态。在家庭收入中, 有些家庭有可能是入不敷出, 比如做生意亏本了, 所以收入有可能是存在负值的,

## (2) 关于 2013 年个别变量说明

### 1) asset

家庭总资产(asset)包括非金融资产和金融资产。非金融资产包括农业经营资产、工商业经营资产、土地资产、房产、车辆资产、其他非金融资产。金融资产包括社保账户余额、现金、存款、股票、基金、债券、衍生品、理财、外币资产、黄金、其他金融资产和借出款。

### 2) total\_income

家庭总收入(total\_income)包括工资性收入、农业经营收入、工商业经营收入、转移性收入和投资性收入。

### 3) h2001 和 h2002 变量说明

[h2001]去年, 您家是否从政府那里获得以下补贴/补助,这里政府补贴不包括前面提到的农业生产经营补贴?(可多选)

- |           |                 |
|-----------|-----------------|
| 1.没有获得    | 6.救济金、赈灾款       |
| 2.特困户补助金  | 7.食物补贴          |
| 3.独生子女奖励金 | 8.退耕还           |
| 4.五保户补助金  | 9.其他(请注明) _____ |
| 5.抚恤金     |                 |

[h2002]去年,【CAPI自动加载H2001各项】分别领取了多少?(单位: 元)

[h2002 01] 特困户补助金\_\_\_\_\_

[h2002 02] 独生子女奖励金\_\_\_\_

[h2002 03] 五保户补助金\_\_\_\_

[h2002 04] 抚恤金\_\_\_\_

[h2002 05] 救济金、赈灾款\_\_\_\_

[h2002 06] 食物补贴\_\_\_\_

[h2002 07] 退耕还林\_\_\_\_

[h2002 08] 其他（请注明）\_\_\_\_

**请注意：**数据里面的变量 h2002\_01 对应的是[h2001]里的第二个选项：特困户补助金，h2002\_02 对应的是[h2001]里的第三个选项：独生子女奖励金，以此类推。

### （3）关于 2013 年问卷批次说明

2013 年问卷一、二批次调查问卷略有删减，删减变量为：

H3020、H3023、H3024、H3027、H3036、H3039、H3041、H3042

以上变量第一批次调查完成访问 16000 户左右。

## 三、附表

### 1、2011 年变量列表

#### （1）关于 2011 年插值变量列表

附表 1-1 2011 年插值变量列表

变量名称				
c3016_#_imp	d5109_imp	d4103_#_imp	k1101_imp	a3036_imp
c2013_#_imp	d6106a_imp	d8104_#_imp	c7009_#_imp	f1010_imp
c2016_#_imp	d6110_imp	c7052a_imp	b1012_imp	f1031_imp
c2055_#_imp	d6115_imp	c7057_imp	b2013_imp	f2006_imp
d1105_imp	d6116_imp	g1018_imp	b2014_imp	f4008_imp
d2104_imp	d7106a_imp	c7050_#_imp	b2015_imp	f4011_imp
d3103_imp	d7110_imp	c7019_#_imp	b3005_imp	f1005_imp
d3109_imp	d7112_imp	g1014_imp	b1005_imp	f1022_imp
d3116_imp	d8106_imp	g1015_imp	b3033_imp	f1025_imp
d3117_imp	d9103_imp	h2004_imp	a3020_imp	f4005_imp

d5107_imp	d9105_imp	c8007_imp	a3022_imp	g1017_imp
-----------	-----------	-----------	-----------	-----------

## (2) 关于 2011 年截尾变量列表

附表 1-2 2011 年截尾变量列表

变量名	截尾标准	样本量
<b>household data</b>		
hh_income	3000000	7
hh_income	-500000	1
b2003b	2000000	15
b2003c	2000000	4
b2015_imp	2000000	8
b1005_imp	2000000	1
b1013a	2000000	4
c4007	200	2
c4008b	200	21
c6007	200	3
b1014	5000000	1
b2003a	5000000	15
b2013_imp	5000000	33
b2013	5000000	26
b2015	5000000	1
c5010_1	5000000	2
c5010_2	5000000	2
c8005_2	5000000	1
c8005_4	5000000	2
k2102a	5000000	1
c2016_2_imp	500	6
c2016_3_imp	500	2
c2016_1_imp	500	34
c2016_1	500	32
c2016_2	500	5
c2016_3	500	2
<b>individual data</b>		
labor_inc	2000000	1
f1031	2000000	1
f1031_imp	2000000	1

## 2、2013 年变量列表

### (1) 关于 2013 年插值变量列表

附表 2-1 2013 年插值变量列表

变量名称				
b1004a_imp	c2032_#_imp	d1111_imp	d7106a_imp	k1101_imp
b1005_imp	c2041_#_imp	d2104_imp	d7110_imp	k2102a_imp
b1005b_imp	c2055_#_imp	d2106_imp	d7112_imp	k2102b_imp
b1012_imp	c3002_#_imp	d2117_imp	d8104_#_imp	k2109_imp
b1013a_imp	c3002a_#_imp	d3103_imp	d8104b_imp	k2113_imp
b1014_imp	c3003_#_imp	d3109_imp	d8106_imp	a3020_imp
b2003b_imp	c3016_#_imp	d3110_imp	d9102_imp	a3022_imp
b2003c_imp	c3017c_imp	d3116_imp	d9103_imp	a3023_imp
b2013_imp	c3021_#_imp	d3117_imp	d9105_imp	a3024_imp
b2014_imp	c3022_#_imp	d4103_#_imp	e1007b_imp	a3036_imp
b2015_imp	c7000a_#_imp	d4104_#_imp	e1021_imp	a3037_imp
b3003a_#_imp	c7000c_#_imp	d4108_#_imp	e1022_imp	f1005_imp
b3005_#_imp	c7000e_#_imp	d4111_imp	e1027_imp	f1010_imp
b3008_#_imp	c7009_#_imp	d5107_imp	e1034_imp	f1018a_imp
b3031_#_imp	c7015a_#_imp	d5108_imp	e2020_imp	f1022_imp
b3031a_#_imp	c7019_#_imp	d5109_imp	e3004b_imp	f1025_imp
b3044_#_imp	c7047b_#_imp	d6106a_imp	e3004c_imp	f1031_imp
b3051_#_imp	c7047d_#_imp	d6106b_imp	e3007a_imp	f1031a_imp
c2000a_#_imp	c7048b_#_imp	d6110_imp	e3014_imp	f2006_imp
c2000c_#_imp	c7048k_#_imp	d6111_imp	g1014_imp	f4005_imp
c2000f_#_imp	c7051b_imp	d6115_imp	g1015_imp	f4008_imp
c2013_#_imp	c7052a_imp	d6115c_imp	g1017_imp	f4011_imp
c2013a_#_imp	c8007_imp	d6116_imp	g1018_imp	
c2016_#_imp	d1105_imp	d7105_imp	h2004_imp	

### (2) 关于 2013 年截尾变量列表

表 2-2 2013 年截尾变量列表

变量名	截尾标准	样本量
asset	2000w	38
b1004a	300w	7
b1004a_imp	300w	7
b1005_imp	300w	6
b1005b	200w	2



b1005b_imp	200w	2
b1012	200w	7
b1012_imp	200w	8
b1013a	200w	2
b1013a_imp	200w	2
b1014_imp	500w	1
b2003a	300w	46
b2003b_imp	300w	27
b2003c_imp	300w	5
b2013_imp	800w	66
b2014_imp	800w	38
b2015_imp	300w	13
b3003a_1	300w	9
b3003a_1_imp	300w	9
b3003a_2	500w	5
b3003a_2_imp	500w	5
b3005_1	300w	1
b3005_1_imp	300w	1
b3005_2	300w	3
b3005_2_imp	300w	3
b3008_1	200w	1
b3008_1_imp	200w	1
b3031_1	500w	8
b3031_1_imp	500w	10
b3031_2	200w	3
b3031_2_imp	200w	3
b3031a_1	200w	2
b3031a_1_imp	200w	2
b3031a_2	200w	1
b3031a_2_imp	200w	1
b3044_1	500w	21
b3044_1_imp	500w	21
b3044_2	300w	7
b3044_2_imp	300w	7
b3051_1	300w	7
b3051_1_imp	300w	7
b3051_2	300w	1
b3051_2_imp	300w	1
c2000a_1	500w	15
c2000a_1_imp	500w	27
c2000a_2	300w	6
c2000a_2_imp	300w	7
c2000a_3	300w	1
c2000a_3_imp	300w	1
c2000f_1	300w	1
c2000f_1_imp	300w	1
c2000f_2	200w	2
c2000f_2_imp	200w	2
c2013_1	800w	88
c2013_1_imp	800w	87
c2013_2	500w	27

c2013_2_imp	500w	28
c2013_3	300w	3
c2013_3_imp	300w	3
c2013a_1	200w	4
c2013a_1_imp	200w	4
c2016_1_imp	800w	20
c2016_2_imp	500w	24
c2016_3	300w	16
c2016_3_imp	300w	17
c2032_1	300w	1
c2032_1_imp	300w	1
c2032_2	200w	1
c2032_2_imp	200w	1
c2041_1	200w	5
c2041_1_imp	200w	5
c2041_2	300w	1
c2041_2_imp	300w	1
c2055_1	200w	2
c2055_1_imp	200w	2
c2055_2	300w	1
c2055_2_imp	300w	1
c3002_1_imp	200w	5
c3003_1	800w	27
c3003_1_imp	800w	27
c3003_2	200w	2
c3003_2_imp	200w	2
c3016_1	200w	6
c3016_1_imp	200w	6
c3017c	300w	23
c3017c_imp	300w	34
c3021_1	200w	3
c3021_1_imp	200w	3
c3021_3	200w	1
c3021_3_imp	200w	1
c3022_1	300w	11
c3022_1_imp	300w	12
c3022_2	300w	2
c3022_2_imp	300w	2
c4007a	200w	1
c5010_1	300w	44
c5010_2	300w	6
c5010_3	300w	7
c5010_4	300w	1
c7000a_1	300w	1
c7000a_1_imp	300w	1
c7000e_1	200w	2
c7000e_1_imp	200w	2
c7009_2	300w	1
c7009_2_imp	300w	1
c7015a_2	200w	1
c7015a_2_imp	200w	1

c7051b	200w	1
c7051b_imp	200w	1
c7052a	200w	1
c7052a_imp	200w	1
c8005_1	300w	1
c8005_2	300w	2
c8005_6	300w	1
c8007	200w	1
c8007_imp	200w	1
d1105	300w	3
d1105_imp	300w	3
d1111	300w	2
d1111_imp	300w	1
d2104	300w	1
d2104_imp	300w	1
d2117	200w	1
d2117_imp	200w	1
d3103	200w	1
d3103_imp	200w	1
d3109	200w	3
d3109_imp	200w	3
d3110	200w	6
d3110_imp	200w	6
d3116	200w	1
d3116_imp	200w	1
d4103_1	200w	1
d4103_1_imp	200w	1
d4104_1	200w	1
d4104_1_imp	200w	1
d4108_1	200w	1
d4108_1_imp	200w	1
d5108	300w	1
d5108_imp	300w	1
d7105	300w	2
d7105_imp	300w	1
d7106a	300w	1
d7106a_imp	300w	1
d7110	200w	1
d7110_imp	200w	1
d9103	300w	2
d9103_imp	300w	1
e3004b	200w	1
e3004b_imp	200w	2
e3007a	200w	1
e3007a_imp	200w	1
e3014	200w	1
e3014_imp	200w	1
h2004_imp	200w	1
k1101	200w	1
k1101_imp	200w	1
k2102a	300w	7

k2102a_imp	300w	5
k2102c	200w	6
k2102c_imp	200w	6
k2109	300w	1
k2109_imp	300w	1
total_income	500w	8
total_income	-100w	7
total_income_imp	500w	8
total_income_imp	-100w	7

### 3、其它特殊变量列表

附表 3-1 特殊变量列表

变量名称	变量取值	变量取值含义	补充说明
<b>hhead</b>	0	非户主	在 ind 数据集中
	1	户主	
<b>rural</b>	0	城镇	在 master 数据集中
	1	农村	
<b>region</b>	1	东部	北京、天津、河北、辽宁、上海、江苏、浙江、福建、山东、广东、海南
	2	中部	山西、吉林、黑龙江、安徽、江西、河南、湖北、湖南
	3	西部	内蒙、广西、重庆、四川、贵州、云南、陕西、甘肃、青海、宁夏