

China Family Panel Studies



中国家庭追踪调查

技术报告系列: CFPS-27

系列编辑: 谢宇 责任编辑: 张聪

# 中国家庭追踪调查 2012 年家庭收入的调整办法

许琪 张春泥

2014.12.15

2012 年中国家庭追踪调查（以下简称 CFPS）的家庭经济问卷以回顾性提问的方式收集了受访家庭过去一年的各项收入情况。为了方便用户，我们在对数据清理、整合及调整的基础上生成了 2012 年与家庭收入相关的综合变量。本报告将介绍这些综合变量的生成规则。

## 一、 2012 年家庭收入的定义

按照收入的来源，CFPS 的家庭总收入可分为工资性收入、经营性收入、转移性收入、财产性收入和其他收入五个部分。其中，工资性收入是指家庭成员从事农业打工或从事非农受雇工作挣取的税后工资、奖金和实物形式的福利。经营性收入是指家庭从事农、林、牧、副、渔的生产经营扣除成本后的净收入和由自家生产并供自家消费的农林产品的价值，以及家庭从事个体经营或开办私营企业获得的净利润。转移性收入是指家庭通过政府的转移支付（如养老金、补助、救济）和社会捐助获取的收入。财产性收入是指家庭通过出租土地、房屋、生产资料等获得的收入。其他收入包括亲友的经济支持和赠予、礼品礼金等。表 1 提供了 CFPS 2010 年和 2012 年调查中的上述家庭收入项目构成明细。

表 1：2010 年和 2012 年 CFPS 的收入项目构成明细

2010 问卷内容	2012 问卷内容
<b>I. 经营性收入</b> 1. 农业收入 - 农林牧副渔业销售净收入 - 主要农产品自家消费部分收入（调整得到） 2. 开办私营企业的净收入	<b>I. 经营性收入</b> 1. 农业收入（产品净收入和自家吃用） - 种植业和林产品 - 畜牧和水产品 2. 个体经营及开办私营企业的净收入
<b>II. 工资性收入</b> 1. 家庭问卷逐一提问每个同住家庭成员的工资性收入（含工资、奖金、补贴、分到个人名下的红利等），CAPI 系统自动加总	<b>II. 工资性收入</b> 1. 个人农业打工（农活/打散工）收入 2. 个人从事每一份受雇非农工作的收入 - 工作的税后收入（含奖金和现金福利） - 工作的实物形式福利（折算） - 个人接受正式教育期间勤工助学/实习/兼职收入

<p><b>III. 转移性收入</b></p> <p>1. 离退休金/社会保障金/低保等收入</p>	<p><b>III. 转移性收入</b></p> <p>1. 退休退职人员的养老金（退休金）</p> <p>2. 政府补助</p> <ul style="list-style-type: none"> <li>- 低保</li> <li>- 退耕还林补助</li> <li>- 农业补助</li> <li>- 五保户补助</li> <li>- 特困户补助</li> <li>- 工伤人员供养直系亲属抚恤金</li> <li>- 救济金、赈灾款（含实物形式）</li> <li>- 其他政府补助</li> </ul> <p>3. 捐助或补偿</p> <ul style="list-style-type: none"> <li>- 社会捐助（包括现金和实物）</li> <li>- 征地补偿金</li> <li>- 住房拆迁补偿金</li> </ul> <p>4. 接受教育期间的奖学金、各类助学金</p>
<p><b>IV. 财产性收入</b></p> <p>1. 出租房屋的租金总收入</p> <p>2. 出租土地或其他生产资料的租金总收入</p> <p>3. 出租家里其他东西的租金总收入</p> <p>4. 出卖财物（家里东西）的总收入</p>	<p><b>IV. 财产性收入</b></p> <p>1. 出租房屋的租金收入</p> <ul style="list-style-type: none"> <li>- 出租自家正在居住房屋的租金</li> <li>- 出租其他房产的租金</li> </ul> <p>2. 出租土地的收入</p> <ul style="list-style-type: none"> <li>- 出租自家从集体分配到的土地的收入</li> <li>- 转租已租用土地的收入</li> </ul> <p>3. 出租其他家庭资产（如设备、运输工具）的收入</p>
<p><b>V. 其他收入</b></p> <p>1. 收到的礼金/礼品收入及其他收入</p>	<p><b>V. 其他收入</b></p> <p>1. 私人经济支持或赠与</p> <ul style="list-style-type: none"> <li>- 不同住亲戚的经济支持和赠与</li> <li>- 朋友、同事的经济支持和赠与</li> </ul>

注：2010年调查采集的是以2010年调查时间为基点的过去一年的数据，2012年调查采集的是以2012年调查时间为基点的过去一年的数据。

与 2010 年调查相比，2012 年调查对与收入相关的问卷内容和提问方式做了一些调整，主要包括：

1. 2012 年调查在问卷内容上弥补了 2010 年调查对部分收入项目的遗漏。例如，2010 年没有直接提问供自家消费的农产品的价值<sup>1</sup>、个体经营的利润、农业打工收入以及在学者的奖助学金和实习/兼职收入等，这些项目在 2012 年的问卷中均被明确设计为调查问题。

2. 2012 年调查在 2010 年基础上对收入大类别下的具体内容进行了细化。以农业生产收入为例，2010 年调查只是笼统地询问农户从事农、林、牧、副、渔业的全部毛收入和总成本，而 2012 年则分别询问了农户从事种植业和养殖业的销售收入、自家吃用部分的收入以及各生产环节的成本。又如，2010 年在调查家庭的政府转移性收入时只用一道题笼统地询问家庭离退休金/社会保障金/低保收入的总额，而 2012 年调查则对该类收入细化，先让受访者列举收到的政府补助项目，再逐项提问家庭从该补助项目中得到的收入。这些对收入内容的列举与细化有助于受访人回忆和明确需要回答的收入内容，避免遗漏。

3. 2012 年调查对多项收入增设分级展开的逼近式提问法(unfolding brackets)。分级展开提问是指当受访人无法回答或拒绝回答某项重要收入的具体金额时，CFPS 的电子化问卷会提供几级金额的范围请受访人选择其该项收入处在哪一级。这一设计能够降低提问的敏感性，减少数据缺失。CFPS 在 2010 年调查收入时仅对全家工资性总收入这一题使用了分级展开提问，而 2012 年调查对农业总收入、个体经营及办私营企业收入、个人的工资性收入均增设了分级展开提问。我们对分级展开提问数据的处理方法是将对收入区间的上限和下限取均值，以此替代未作答的收入具体金额。

4. 2012 年调查在工资性收入的提问方式上有重大调整。2010 年调查是在家庭层面上提问全家每一位家庭成员的工资性收入，即由家庭经济问卷的回答人先逐一回答每一位家庭成员的工资性收入或收入区间，再回答所有家庭成员的工资性收入的总额或总额的区间。2012 年调查则将工资性收入分散到每一位成员的个人问卷中提问。因此，整个家庭的工资性收入需要通过加总每一份个人问卷的工资性收入来得到。

在以上四项调整中，前三项无疑均有助于提高 2012 年 CFPS 收入数据的质量，而第四项却存在弊端。一般来说，每个人自答的工资性收入会比由他人代答的收入更准确，这是

---

<sup>1</sup> 在 2010 年发布的收入综合变量中，我们根据农户的总产量、销售量和农产品价格对自家消费农产品的价值作了估算，详见：《CFPS-14：中国家庭追踪调查 2010 年农村家庭收入的调整办法》。

2012 年调查将工资性收入分散到个人问卷提问的设计出发点，但是，获得准确的全家工资性收入的前提是每一位有工资性收入的家庭成员都回答了个人问卷，且对个人收入的回答没有缺失值。然而，在实际调查中却难免会出现个体无应答或项目无应答的情况。我们在汇总 2012 年的家庭工资性收入时发现，有一些家庭成员从事受雇工作，却没有回答他们的工资性收入；还有一部分成员因为外出务工，没有回答个人问卷。这些无回答均会造成家庭工资性收入的低估。

为了进一步说明这一问题，我们对比了 CFPS 和国家统计局公布的工资性收入。国家统计局数据显示，从 2009 年至 2011 年，农村人均工资性收入增长了 44%，城镇人均工资性收入增长了 24%。但根据 CFPS 原始数据计算，农村人均工资性收入仅增长了 14%；城镇人均工资性收入不仅没有增长，反而下降了 4%。究其原因，主要是 2012 年调查中一些家庭成员的工资性收入缺失而导致了低估。

**表 2：调整前 CFPS 和国家统计局的人均家庭收入增长率<sup>2, 3</sup>**

		人均工资性收入	
		CFPS	国家统计局
农村	2010	3746	2061
	2012	4270	2963
	比率	1.14	1.44
城镇	2010	8515	12382
	2012	8196	15412
	比率	0.96	1.24

由于工资性收入是城乡家庭收入最重要的组成部分，这些缺失值的存在会严重低估受访家庭的总收入。因此，为了降低数据缺失造成的影响，我们对 2012 年的工资性收入进行了调整。我们接下来将介绍该调整方法。

<sup>2</sup> 此处的工资性收入仅限定在两年可比的部分，即从 2012 年调查的工资性收入中剔除农业打工收入、个人实习和勤工助学收入这几项 2010 年没有涉及的项目。

<sup>3</sup> 此处使用的是 CFPS 全国再抽样样本，且限定在两年都参加调查的家户。

## 二、 工资性收入调整办法

2012年CFPS调查在以下两种情况下会产生个人的工资性收入的缺失。对这两种情况，我们分别有相应的调整方案。

### 1. 项目无应答 (item nonresponse)

项目无应答指受访者虽然回答了个人问卷，但没有回答他们的工资性收入。对此，我们以工资性收入的自然对数为因变量，以受访者的性别、年龄、年龄的平方、教育程度、教育年限与性别的交互项、户口性质、城乡居住地、党员身份、健康自评状况为自变量建立OLS回归模型，用回归模型的预测值对缺失值进行插补。

在2012年CFPS成人样本（长、短问卷合计）中，工作类型为受雇的样本为8970人，其中工资性收入大于0的样本量为7823人，另外1147人的工资性收入缺失或为0。我们首先对工资性收入有效的7823人的工资性收入的对数建立回归模型，然后用该模型估计的回归系数来预测另外1147人的工资性收入。该模型的回归系数及标准误详见表3。

表3：OLS回归模型的估计系数与标准误

	系数	标准误
男性（女性为参照类）	0.526***	0.050
年龄	0.053***	0.005
年龄的平方	-0.001***	0.000
教育程度（文盲为参照类）		
小学	0.141***	0.042
初中	0.305***	0.043
高中、职高或技校	0.352***	0.051
大专	0.515***	0.068
本科	0.813***	0.081
研究生	1.472***	0.247
男性×教育年限	-0.013*	0.005
党员	0.016	0.050
健康状况（非常健康为参照类）		
很健康	-0.020	0.038
比较健康	-0.062†	0.036
一般	-0.094*	0.042
不健康	-0.315***	0.051
非农户口（农业户口为参照类）	0.051	0.037
居住城市（居住农村为参照类）	0.031	0.035
截距	8.361***	0.113

<b>R-square</b>	0.143
<b>样本量</b>	7823

注：因变量为工资性收入的对数。教育年限从教育程度转换得到，转换方法是：文盲为 0 年，小学为 6 年，初中为 9 年，高中、职高或技校为 12 年，大专为 15 年，本科为 16 年，研究生为 19 年。

† p<0.1   \* p<0.05   \*\* p<0.01   \*\*\* p<0.001

## 2. 个体无应答 (unit nonresponse)

个体无应答指受访者没有接受成人问卷的访问。对此，我们先通过家庭成员问卷判断受访者在 2012 年调查时是否正在外出务工，如果是，再使用 CFPS 各省（流出地省）汇总的外出务工人员的工资性收入的中位数进行插补。

2012 年调查没有回答成人问卷的外出务工人员有 1239 人，回答了成人问卷的外出务工人员有 2798 人。我们对 2798 名外出务工者分省统计了他们的年工资性收入的中位数（见表 4），并将这些中位数作为各省缺失的外出务工人员的工资性收入。

表 4：分省计算的外出务工人员年工资性收入的中位数

省份	中位数（元）	样本量（人）
天津	25000	2
河北	20979	106
山西	20000	107
辽宁	21600	163
吉林	22414	22
黑龙江	20000	15
上海	50000	2
江苏	24000	35
浙江	24000	25
安徽	21408	79
福建	24900	72
江西	24000	81
山东	24587	70
河南	24000	410
湖北	20000	40
湖南	22843	64
广东	20000	488
广西	20264	74
重庆	27200	31

四川	20800	80
贵州	20000	131
云南	20000	34
陕西	27000	53
甘肃	20000	614

经过上述对项目无应答和个体无应答收入调整后的农村家庭人均工资性收入为 5077 元，比调整前增加了 807 元，调整以后的 2010-2012 年工资性收入增长率为 36%。调整后城镇家庭人均工资性收入为 8668 元，比调整前增加了 472 元，调整后的收入增长率为 2%（见表 5）。虽然在调整之后，CFPS 的工资性收入增长率与国家统计局公布的比率仍有差距，但是与表 2 相比，调整后的工资性收入的增长率更加接近统计局的结果。

表 5：调整后 CFPS 人均家庭收入增长率

		人均工资性收入	
		CFPS	国家统计局
农村	2010	3746	2061
	2012	5077	2963
	比率	1.36	1.44
城镇	2010	8515	12382
	2012	8668	15412
	比率	1.02	1.24

### 三、 2010 年可比收入

前面提到，2012 年调查与 2010 年调查涵盖的收入项目有差别，即 2012 年涵盖的收入内容更广泛，但这一差别一定程度上造成了 2012 年与 2010 年的家庭收入不可比。CFPS 作为追踪数据，不同期的数据之间的比较非常重要。对此，我们仔细比对了两期调查问卷的收入项目，生成了一组与 2010 年数据可比的 2012 年家庭收入变量。

得到可比收入的具体做法包括：（1）从 2012 年家庭收入中剔除个人农业打工收入、个人实习和勤工俭学收入、奖助学金收入、征地补偿金和住房拆迁款这几项 2010 年收入调查中没有涉及或由于提问方式不一而缺乏可比性的项目；（2）剔除 2012 年家庭非农经营收入，因为 2010 年的家庭非农经营收入只提问了私营企业，没有问及个体经营，而 2012 年同时包含了这两部分收入且无法拆分。

因此，2012 年的家庭收入综合变量包括两个版本：一个是根据 2012 年家庭经济问卷的内容完整计算的家庭收入，另一个版本是与 2010 年保持一致的可比家庭收入。在发布数据中，我们用变量名的尾缀将这两个版本的变量标识出来。

对这两个版本变量的使用，我们建议，当使用 2012 年单期数据时，最好使用完整计算的家庭收入变量，但如果需要比较 2010 年和 2012 年的家庭经济状况时，最好使用与 2010 年可比较的收入变量。

#### 四、 调整后的家庭收入概览

我们将调整后的工资性收入与经营性收入、财产性收入、转移性收入、其他收入加总，生成最终的家庭收入（或调整后的家庭收入）。表 6 描述了全国和五个独立抽样省 2012 年家庭纯收入的水平和结构。2012 年全国家庭纯收入的均值为 45665 元，中位数为 32414 元。从收入的构成上看，工资性收入所占比重最高，为 56.1%；其次是经营性收入，为 26.6%；再次是转移性收入，为 12.0%；财产性收入和其他收入占家庭纯收入的比重较小。

分地区来看，各地家庭纯收入水平相差很大。五个独立抽样省中，收入最高的上海，当地家庭纯收入的均值为 84433 元，而平均家庭收入最低的辽宁为 41668 元。平均来看，一个上海家庭一年的收入约为辽宁家庭的 2 倍，但如果考虑物价的话，两个地区的实际收入差异会低一些。从收入的构成上看，各地差异也较大。

表 6：2012 年各地家庭纯收入水平和结构

地区	家庭纯收入		工资性收入	经营性收入	转移性收入	财产性收入	其他收入	样本量
	均值	中位数						
全国	45665	32414	56.1	26.6	12.0	1.8	3.6	7917
上海	84433	57600	61.2	5.5	24.9	6.4	2.0	977
辽宁	41668	31650	53.3	19.0	22.9	1.3	3.6	1240
河南	47818	36366	64.4	22.0	10.7	1.3	1.5	1336
甘肃	46636	34632	57.5	32.5	7.5	1.4	1.1	1377
广东	53173	39971	65.0	20.5	8.5	1.6	4.5	1062

表 7 分全国、城乡描述了 2010 年和 2012 年的基尼系数和 90/10 比率。这里使用的是可比数据。2012 年全国基尼系数为 0.49，农村为 0.50，城镇为 0.48。可见，城镇内部和农村

内部的收入分化也非常大。从全国看，根据家庭纯收入计算的基尼系数（表 7 的“基尼系数 1”）从 2010 年到 2012 年略有降低；但分城乡看，该基尼系数几乎没有变化。90/10 比率显示，2012 年收入排名在全国 90% 的家庭的收入是排名在 10% 家庭的 13 倍，该倍数在农村样本中为 12.4 倍，城镇样本中为 12.1 倍。无论是全国还是分城乡来看，90/10 比率在这两年间均略有上升，这说明两年间最富的人和最穷的人之间的收入差距有所扩大。

表 7：2010 和 2012 年收入不平等汇总指标及其变化

不平等指标	全国		农村		城镇	
	2010	2012	2010	2012	2010	2012
基尼系数 1	0.51	0.49	0.50	0.50	0.48	0.48
基尼系数 2	0.52	0.48	0.48	0.45	0.50	0.48
90/10 比率	12.9	13.1	11.3	12.4	10.6	12.1

注：“基尼系数 1”根据家庭纯收入计算，“基尼系数 2”根据人均家庭纯收入并用家庭人口数加权计算，下同。

## 五、综合变量名称及标签

表 8 展示了 2012 年发布数据中与家庭收入有关的综合变量。这些变量包括五个收入来源的分项收入、五个来源加总的家庭纯收入、控制了家庭人口规模的人均家庭收入，以及人均家庭收入的分位数。

其中，变量名带“\_1”后缀的是指完整计算的家庭收入，即包含表 1 中所有的收入项目；而带“\_2”后缀的是指与 2010 年可比的收入，仅包含与 2010 年调查一致的收入项目。

变量名带“\_adj”后缀的表示的是经过调整的收入，但我们也保留了根据原始数据生成的收入变量。用户可以自行选择是否采用我们的调整。具体来讲，各变量所做的调整如下：

1. 家庭工资性收入 wage\_1\_adj、wage\_2\_adj：对受雇且工资性收入为 0 或缺失的个人进行了插补；对外出务工没有接受访问的家庭成员的工资性收入进行了插补。

2. 家庭纯收入 fincome1\_adj、fincome2\_adj：对受雇且工资性收入为 0 或缺失的个人进行了插补；对外出务工没有接受访问的家庭成员的工资性收入进行了插补；对家庭纯收入为 0 或缺失的家庭用家庭消费进行替代。

3. 家庭人均纯收入 `fincome1_per_adj`、`fincome2_per_adj`: 由调整后的家庭纯收入除以家庭规模得到。

4. 家庭人均收入分位数 `p_adj`: 由调整后的家庭人均纯收入计算得到的。

最后, 为了方便用户使用, 我们还生成了与 2012 年可比的 2010 年的家庭纯收入和人均家庭纯收入, 即从 2010 年家庭纯收入中减去非农经营和出卖财物这两项与 2012 年不可比的内容。

**表 8: 综合变量及其标签**

变量名	变量标签
<code>wage_1</code>	工资性收入
<code>wage_2</code>	工资性收入 (与 2010 年可比)
<code>wage_1_adj</code>	工资性收入-调整
<code>wage_2_adj</code>	工资性收入-调整 (与 2010 年可比)
<code>foperate_1</code>	经营性收入
<code>foperate_2</code>	经营性收入 (与 2010 年可比)
<code>ftransfer_1</code>	转移性收入
<code>ftransfer_2</code>	转移性收入 (与 2010 年可比)
<code>fproperty_1</code>	财产性收入
<code>fproperty_2</code>	财产性收入 (与 2010 年可比)
<code>felse_1</code>	其他收入
<code>felse_2</code>	其他收入 (与 2010 年可比)
<code>fincome1</code>	2011-2012 全部家庭纯收入
<code>fincome2</code>	2011-2012 家庭纯收入 (2010 可比)
<code>fincome1_adj</code>	2011-2012 全部家庭纯收入-调整
<code>fincome2_adj</code>	2011-2012 家庭纯收入-调整 (2010 可比)
<code>fincome1_per</code>	2011-2012 人均家庭纯收入
<code>fincome2_per</code>	2011-2012 人均家庭纯收入 (2010 可比)
<code>fincome1_per_adj</code>	2011-2012 人均家庭纯收入-调整
<code>fincome2_per_adj</code>	2011-2012 人均家庭纯收入-调整 (2010 可比)
<code>fincper_p</code>	家庭人均收入分位数
<code>fincperadj_p</code>	家庭人均收入分位数-调整