



2017-10-15

CGSS 质量控制

—— 数据核查

llygg6@gmail.com

中国人民大学中国调查与数据中心

前 言

本篇文档是笔者作的关于 CGSS 项目质量控制中的数据核查工作的总结，这些总结来源于笔者在中国人民大学中国调查与数据中心三个月的实践工作。如果让我给自己这三个月的工作打个分，如果总分是 100 的话，我觉得我只能给自己打个 59 分的不及格分数，因为想着准备考研，我没能全身心地投入工作。

这三个月实践工作的顺利展开，离不开胡老师的关心和指导，离不开丹阳姐、锋姐、雨薇、立鑫、欢姐、毛老师、大哥、瑶瑶、兰兰、昕烨、福鑫、贾敏、张钰带给我的帮助、支持和欢笑。感谢在最好的年华遇见你们，未能和你们道别就匆匆离开，因为相信重逢就在不远的前方。

这篇文档实则只能算个粗略的总结，但因时间仓促，只能作罢，希望将来如果有更多的时间可以对其作细致的修订。从中心离职后的几天，又曾回到中心，那天从明德楼走出去的时候，遇到了兰兰。兰兰说：“有时间经常过来坐一坐啊”。我说好，当我转过身去，望着蓝天，嗅到的空气，香甜、感动而无比温暖。

——征帆，2017 年 10 月 15 日 22 时 28 分 00 秒于北京北四环志新村小区

目 录

1.概况	1
2 调查进度	1
2.1 调查进度应包含的指标及其统计规则.....	1
2.2 质量控制为什么需要统计调查进度	1
2.3 调查进度统计的注意事项.....	2
3 问卷逻辑	2
3.1 质量控制为什么要核查问卷逻辑.....	2
3.2 CGSS2017 部分问卷逻辑点	3
4 废卷	4
5 问卷题目答案缺失	4
6 电话核查抽样.....	4
7 其它工作	5
7.1 创建公用大表.....	5
7.2 数据备份.....	5
7.3 输出无联系记录给样本追加小组.....	5
7.4 全流程自动化.....	5
8 附录	6
8.1 程序基本步骤.....	6
8.2 程序代码.....	6

1 概况

在整个中国综合社会调查（CGSS）项目前期的数据采集工作中，调查质量的控制是其中的中心环节。调查质量包括宏观层面的质量和微观层面的质量，宏观层面的质量包括应答情况和电话回收情况，微观层面的质量则指问卷数据本身的质量。宏观层面的质量由调查进度得到，微观层面的质量则主要依靠电话核查得到。无论是调查进度的获取，还是电话核查样本的导出，其依据均是调查从规划设计到具体实施过程中产生的各类数据。因此，与数据打交道的数据核查人员在数据采集工作中扮演着至关重要的角色。

质量控制工作中的数据核查，就是指利用调查进度获知调查宏观层面质量，同时结合问卷逻辑、废卷情况、问卷题目答案缺失情况等相关指标从受访样本中筛选出质量存疑样本、低电话回收率及高缺失率样本，将之反馈给电话核查人员，并汇总电话核查人员通过电话核查得到的调查微观层面质量信息的一系列工作。

2 调查进度

调查进度是获取调查宏观层面质量，从受访样本中筛选出质量存疑样本和低电话回收率样本的重要依据，接下来的 2.1 节将会说明调查进度应当包含的指标及其统计规则，2.2 节将详细解释为什么质量控制需要调查进度，2.3 节则是在统计调查进度的实际操作中，笔者积累的一些注意事项。

2.1 调查进度应包含的指标及其统计规则

调查进度应包括省、县区、村居、访员总进度和每日进度。其中省、县区、村居总进度包括以下指标：下发样本量、上传问卷数、完成问卷数、电话回收数、电核量、废卷数、应答率、完成率、电话回收率；访员总进度包括以下指标：上传问卷数、完成问卷数、电话回收数、电核量、废卷数、完成率、电话回收率；省、县区、村居、访员每日进度表应包括以下指标：上传问卷数、完成问卷数、电话回收数、电核量、废卷数、完成率、电话回收率。以 CGSS2017 为例，各指标的统计规则如下（均针对预处理过后的数据而言）：

1. 下发样本量：样本清单文件的行数
2. 上传问卷数：问卷数据的行数
3. 完成问卷数：lastpage=75 且 Z1 非空的样本数
4. 电话回收数：Z2 或 Z3 中留有电话的样本数
5. 电核量：电话核查记录表上有记录的样本数
6. 废卷数：电话核查记录表中标记为废卷的样本数
7. 应答率：完成问卷数/下发样本量
8. 完成率：完成问卷数/上传问卷数
9. 电话回收率：电话回收数/完成问卷数

2.2 质量控制为什么需要统计调查进度

一方面，就调查质量的宏观层面而言，质量控制需要统计调查进度。这是因为调查进度中的应答率、电话回收率反映了调查的应答情况和电话回收情况，它们从宏观上刻画了调查的总体质量。同时，电话回收率这一指标还直接影响到调查微观层面质量控制工作，因为如果无法回收到受访者的联系方式，则意味着调查微观层面的质量控制工作基本无法展开。

另一方面，就调查质量的微观层面而言，质量控制也需要统计调查进度。从表面上来看，一个调查的进

度似乎与调查在微观层面的质量，即与问卷数据本身是否存在质量问题没有多大联系。但如果我们仔细想一想，就会发现：由调查进度中的相关指标除以调查时长得到的调查速度与问卷质量之间有着千丝万缕的联系。过快的事务处理速度常常伴随着事务处理质量的下降，这是不言而喻的道理。因此一般而言，一名访员或一个地区的调查推进速度过快时，该访员或该地区的问卷出现质量问题的可能性相对较高。

质量控制需要统计调查进度还有另外一个原因，那就是：质量控制工作讲究时效性。对于一个实际存在质量问题的问卷样本而言，通过电话核查确认该问卷样本存在问题的几率将随着时间的推移逐渐减小，而对于一个实际不存在质量问题的问卷样本而言，通过电话核查错误判定该问卷样本存在问题的几率将随着时间的推移逐渐增加。这是因为，随着时间的推移，无论是受访者还是访员，他们大脑中关于当时受访或访问具体情况的部分或全部记忆会不可避免地逐渐消逝或发生偏移。

2.3 调查进度统计的注意事项

以下是笔者在 CGSS2017 调查进度统计的实际工作中积累的一些注意事项：

1. 在进行调查进度统计之前，必须对问卷数据进行预处理。这些预处理包括：删除测试样本编号、合并村居信息表、合并样本清单信息、删除重复样本编号等。

2. 下发样本量的统计由样本清单得到，样本清单为.xlsx 格式的 excel 文件，通常程序语言读取.xlsx 文件非常费时，因此建议先将.xlsx 文件批量转换为.csv 文件，这可以利用 Excel 中的 Visual Basic 开发工具编写程序实现，转换速度为 2-4 个文件/秒；

3. 电话回收数的统计不仅应该考虑联系人留下的手机号码，还应考虑其留下的固定电话号码，它们存于三个不同的变量中，这三个变量是：Z2（联系人移动电话），Z3a（联系人除去区号的固定电话），Z3b（联系人固定电话区号）；

4. 电话回收数是以上几个指标中在逻辑上最复杂的指标。这是因为：不是所有的 7 位数或 8 位数都可以作为固定电话号码，也不是所有的 11 位数都可以作为移动电话号码，过于简单的逻辑会把许多明显无效的号码（如：以 1 开头的固定电话号码，以 144 开头的手机号码等等）统计进去；

5. 在得到调查进度统计后，应将电话回收率过低访员（电话回收率<70%）名单复制到电话核查记录表中，让电话核查人员与这些访员取得联系，督促他们在接下来的访问中注意回收联系人的电话。

3 问卷逻辑

问卷逻辑是指：问卷中不同问题及问题选项之间存在逻辑关系。它是从受访样本中筛选出质量存疑样本的重要依据。3.1 节将解释为什么质量控制需要问卷逻辑，3.2 节将给出 CGSS2017 中的一些问卷逻辑点。

3.1 质量控制为什么要核查问卷逻辑

导致问卷各题目答案间出现逻辑矛盾的原因无外乎两种：受访者自身回答问题时有意或无意地犯了逻辑错误，非受访者（受访者亲人、朋友、访员或其他人员）有意或无意地替受访者回答了问题。而后一种情况，即非受访者替受访者回答问题是调查中明令禁止的行为，一经发现，该份问卷就将作废。因此平均而言，从存在逻辑矛盾的问卷发现质量问题的可能性相较其他问卷而言要高。所以，问卷逻辑是筛选质量存疑样本的重要依据。下面将以 CGSS2017 为例，列出部分存在逻辑矛盾的问卷答案作为参考。

3.2 CGSS2017 部分问卷逻辑点

S37.2. != A2 #抽中受访者的性别与实际受访者的性别不匹配

S37.3. != abs(2017-A3) #抽中受访者的年龄与实际受访者的年龄不匹配

A5.1.==2 & A6==1 #信仰宗教但从来没有参加过宗教活动

A7c < A3 #完成最高学历年份小于出生日期

A8a.1_1.+...+ A8a.1_7. < A8b.1_1.+...+ A8b.1_7. #个人去年全年总收入小于个人去年全年职业/劳动收入

A9==1 & A10==4 #未递交过入党申请书，政治面貌为共产党员

(A14/2)^2/A13>35 & (A14/2)^2/A13<10 #BMI 指数异常值

A15==5 & A16==1 #目前身体很健康，但过去四周，健康问题总是影响到工作或其他日常活动

(A28.1.==1 & A29==1) | (A28.2.==1 & A29==2) | (A28.3.==1 & A29==3) | (A28.4.==1 & A29==4) | (A28.5.==1 & A29==5) | (A28.6.==1 & A29==6) #过去一年从不使用报纸/杂志/广播/电视/互联网/手机定制消息，但最主要信息来源却是报纸/杂志/广播/电视/互联网/手机定制消息

A30.4.==5 & ((A30a>0 & A30a<998) | (A30b>0 & A30b<998)) #过去一年从不在空闲时间从事读书/报纸/杂志活动的，过去一年读书数或读报数却>0

A30c>0 & A30c!=998 & A30c!=999 & A30e==2 #过去 12 个月通过网络阅读资讯和文章时间>0，但在半年内没有上过网

A30e==2 & A30g==1 #最近半年没上过网，却使用过微信支付

A30e==2 & A30h==1 #最近半年没上过网，却使用过支付宝支付

A31.1.==5 & A31a==7 #过去一年在空闲时间频繁社交/串门，却从来不与邻居进行社交娱乐活动

A31.1.==5 & A31b==7 #过去一年在空闲时间频繁社交/串门，却从来不与其他朋友进行社交娱乐活动

A31.1.==5 & A31a==7 & A31b==7 #过去一年在空闲时间频繁社交/串门，却从来不与邻居、也不与其他朋友进行社交娱乐活动

A43.a.==1 & A43e==1 #认为自己位于社会最底层，但社会经济地位属于上层

A43.a.==10 & A43e==5 #认为自己位于社会最高层，但社会经济地位属于下层

A51==1 & A52==5 #觉得自己完全听不懂英语，但觉得自己说英语的能力很好

A49==1 & A50==5 #觉得自己完全听不懂普通话，但觉得自己说普通话的能力很好

A12b>A65 & A12b<998 #受访者拥有房产数大于家庭拥有房产数

A68.1.< A68a.1. | A68.2.< A68a.2. #有子女人数小于亲生子女人数

A69==5 & A69a==1 #分居未离婚，但却住在一起

A69==2 & A69a!=1 & A69a<98 #同居，但却不住在一起

A70-A3.1.<16 #结婚年龄小于 16 岁

(A69==3 | A69==4 | A69==5) & A71b==9997 #目前已婚未丧偶，但与目前配偶结婚年份填的 9997，即从未结婚

A69==3 & A70 != A71b & A70<2018 & A71b <2018 #目前初婚有配偶，但第一次结婚时间不等于与目前配偶结婚时间

A70>A71b & A70<2018 & A71b <2018 #第一次结婚时间大于与目前配偶结婚时间

submitdate-startdate < 1h #系统记录访问时间小于 1 小时

Z8-submitdate > 0.5h #访员记录的完成时间在系统记录的提交时间半个小时以后

4 废卷

废卷是指已完成的问卷中那些因为各种原因而被作废，将来不能再用于统计分析的问卷。废卷情况是从受访样本中筛选出质量存疑样本的重要依据，通常认为：如果某个地区或某个访员存在废卷，该地区或访员的所有问卷样本应作重点核查。

废卷来源有三：一是督导现场废卷、二是电话核查废卷、三是高缺失废卷。下面的第 5 节将具体解释什么是高缺失。

5 问卷题目答案缺失

问卷题目答案缺失情况是从受访样本中筛选出高缺失率样本以及质量存疑样本的重要依据。问卷题目答案缺失是指：受访者拒绝回答该题目或对该题目的回答为不知道。缺失率计算方法为：选择不知道或拒绝回答的题目占有所有答案非空题目的比例。高缺失率的统计标准为：缺失率 $>30\%$ 。缺失率过高的问卷将直接作为废卷，因为对它们的统计分析意义不大。

问卷缺失率的统计规则看似十分简单，但实际操作中，对该部分统计程序的编写是最耗时的。根本原因在于：问卷数据中不知道或拒绝回答对应的数字是不同的，而且有些题目按选项设置变量，即一个题目对应多个变量而不是一个题目对应一个变量。这些复杂情况导致：为了统计的准确性，需要对问卷数据作相当繁冗复杂的处理，这些处理包括：将一个题目对应多个变量的情况整合为一个题目对应一个变量的情况，以及统一不知道或拒绝回答对应的数字。这些处理要求我们将成百上千道问卷题目从头到尾仔细浏览一遍，并且需要处理的题目比较分散，其繁琐枯燥程度可见一斑。因此强烈建议以后 CGSS 问卷系统可以统一不知道、拒绝回答选项对应的数字，并且尽量避免一个题目对应多个变量的情况，以方便数据核查人员进行数据核查，也方便数据分析人员进行数据分析。

6 电话核查抽样

数据核查人员的中心工作有两个，一是调查进度的统计，二就是电话核查抽样。调查进度统计用以获取调查宏观层面的质量，同时也为调查微观层面的质量控制提供依据。电话核查抽样则主要是为电话核查人员提供电话核查的样本，为调查微观层面的质量控制铺平道路。电话核查抽样的依据便是前面 2-5 节提到的调查进度、问卷逻辑、废卷情况和问卷题目答案缺失情况。具体抽样标准依次为：1.抽取有逻辑错误的问卷；2.抽取有废卷访员/地区的全部问卷；3.抽取进度过快地区全部问卷；4.抽取存在高缺失率问卷的访员的全部问卷；5.抽取每位访员的第一份问卷。电话核查占完成问卷的比例控制在 20%-40%，人力财力物力允许情况下可适当调高这个控制比例。

7.其它工作

7.1 创建公用大表

建立一张公用的调查进度统计大表，该大表用于存放调查总进度及每日进度。

建立一张公用的电话核查记录大表，该大表用于存放每日电话核查详情、废卷情况、高缺失率问卷、低电话回收率问卷。其中每日电话核查情况包括更新日期、样本编号、提交日期、电核员姓名、电核员编号、联系受访者情况、联系访员情况、电核处理意见、访员姓名、访员电话等信息。

这两张大表将方便项目组所有质量控制工作人员实时共享有关信息，加速整个质量控制工作的运转。

7.2 数据备份

数据备份应每日进行，以 2017 年 CGSS 项目为例，数据备份共涉及 12 个文件，其中电核问卷数据 4 份，访员信息数据 4 份，调查问卷数据 4 份。这 4 份分别对应：spss 语法文件，spss 数据文件，excel 文件，csv 文件。这 12 份文件应统一存放于以当日系统日期命名的文件夹中，并在本地和云端至少各留有一份。

7.3 输出无联系记录给样本追加小组

所谓无联系记录样本是指：在下发样本清单中有但在后台问卷数据中没有记录的样本编号对应的样本。原则上：所有下发样本都应在后台问卷数据中有记录，无论这个样本是否访问成功。因此匹配下发样本清单文件样本编号与问卷数据中的样本编号，找出差额即为无联系记录样本

7.4 全流程自动化

数据核查人员应致力于实现工作流程的自动化，这里推荐 4 种程序语言：AutoHotkey、Visual Basic、R、Bat. 实践表明，它们的结合使用可以实现上述大部分工作的自动化。有关上述步骤的具体细节参见附录中的程序代码（包括注释及空行在内共 2054 行），为方便查看，可将附录中的程序语言复制到 notepad++ 中，然后选择语言为 AutoIt.

8.附录

8.1 程序基本步骤

1.下载电话核查记录表（简称“电核大表”），该表记录了电话核查人员对所有质量存疑问卷的电话核查情况、废卷情况、高缺失率样本和低电话回收率样本

2.处理电核大表，生成 4 个 csv 文件，这 4 个 csv 文件分别记录电话核查样本详情、废卷情况、高缺失率样本情况和低电话回收率访员情况

3.登录 limesurvey 系统下载问卷数据 result-survey 963159.xlsx 并将其转换为 csv 文件

4.执行 R 语言分析程序

5.上传电核样本数据至 limesurvey 系统

6.上传电核样本、新的高缺失率样本、新的低电话回收率访员到电核大表

7.输出无联系记录样本给追加样本小组

8.更新调查进度大表

9.备份 CGSS 数据

8.2 程序代码

```
#NoEnv ; Recommended for performance and compatibility with future AutoHotkey releases.
```

```
; #Warn ; Enable warnings to assist with detecting common errors.
```

```
SendMode Input ; Recommended for new scripts due to its superior speed and reliability.
```

```
SetWorkingDir %A_ScriptDir% ; Ensures a consistent starting directory.
```

```
/*
```

中国综合社会调查（CGSS）质量控制 AutoHotkey 程序

关于该程序的注意事项如下：

- 1.该程序调用了外部三种程序：Visual Basic 程序、Bat 批处理程序、R 程序；
- 2.使用该程序时请将系统的默认浏览器设置为 Google Chrome，并安装插件 Vimium；
- 3.使用该程序前请确保网络通畅，并开启了 Lantern 一类的软件；
- 4.该程序在使用之前需要针对使用该程序的计算机作调试；
- 5.由于使用未经调试的该程序而造成的一切后果由使用者自行承担；
- 6.该程序最终解释权归编写者所有，允许自由传播，但请注明出处。

作者：llygg6@gmail.com 时间：2017-07-03 —— 2017-09-30

```
*/
```

```
;一、自定义函数部分
```

```
;
```

```
;1.睡眠函数
```

```
;
```

;函数旨在将 Sleep 函数的时间参数单位由 ms 修改为 s

Sm(ti=0.1)

```
{  
    ti := ti*1000  
    Sleep, ti  
}
```

```
;
```

;2.等待窗口函数

```
;
```

;wn 为窗口完整名称, wt 为窗口完整名称的子字符串

Ddck(wn,ti=0.1,wt="")

```
{  
    WinWait, %wn%, %wt%,  
    IfWinNotActive, %wn%, %wt%, WinActivate, %wn%, %wt%,  
    WinWaitActive, %wn%, %wt%,  
    Sm(ti)  
}
```

```
;
```

;3.输入文字函数

```
;
```

;word 为需要输入的文字, wait 为 ctrl+v 后的睡眠时间

Srwz(word,ti=0.2)

```
{  
    Clipboard = %word%  
    Sm()  
    Send,^v  
    Sm(ti)  
}
```

```
;
```

;4.发送按键函数

```
;
```

;函数旨在为所有 Send 命令添加一行 Sleep 命令, 避免手工多次输入 Sleep 的麻烦

;默认休息时间为 0.1 秒, 休息时间由该函数第 2 个参数确定

Fs(na,ti=0.1)

```
{  
    Send,%na%
```

```

        Sm(ti)
    }

;-----
;5.寻找颜色函数
;-----

Sc(a,b,c,d,color,var,ti=0.5)
{
    Loop
    {
        PixelSearch, LogoX, LogoY, a, b, c, d, %color%, var, Fast
        Sleep,100
        If ErrorLevel
            continue
        else
            break
    }
    Sm(ti)
}

;-----
;6.左键单击函数
;-----

Zjdj(a,b,ti=0.5)
{
    MouseClick,left,a,b
    Sm(ti)
}

```

;二、CGSS 质量控制主体程序

```

*:cgsszlkz::

;-----
;1.获取系统当前日期，日期格式为 yyyy-MM-dd 和 yyyyMMdd
;-----

    FormatTime, now_date, %A_Now%, yyyy-MM-dd
    Sm()
    FormatTime, now_date_wg, %A_Now%, yyyyMMdd
    Sm()

```

;2.在 CGSSdata 目录下新建以系统日期命名的文件夹

```
Run, cmd.exe
Ddck("C:\windows\system32\cmd.exe")
Fs("md d:\CGSSdata\")
Send,%now_date_wg%2
Sm()
Fs("{ENTER}",1)
Zjdj(641,9,1)
```

;3.登录与下载

;3.1 打开 limesurvey 登录页面并登录 limesurvey 系统

```
Run,http://101.200.178.132/limesurvey_2017/index.php/admin
Ddck("LimeSurvey_2017 - Google Chrome")
Srww("longzhengfan")
Fs("{TAB}")
Fs("123456")
Fs("{ENTER}")
```

;3.2 判断是否登录了 limesurvey 系统

```
Sc(501,216,553,271,"0x0BEDCB",2)
```

;3.3 下载电核大表

```
Run,https://docs.google.com/spreadsheets/d/1tfRQ_n3-4bTEemeqWZlHvLNIV-
eAjKCv9dpNVs7ETSc/edit?usp=drive_web
```

```
Sc(1345,125,1416,153,"0xF78946",2)
Zjdj(80,144)
Loop,7
{
    Fs("{UP}",0.3)
}
Fs("{RIGHT}")
Fs("{ENTER}")
```

```
Ddck("另存为",1)
Fs("^f",0.4)
Fs("+{TAB}")
Fs("{ENTER}")
```

```

Srwx("D:\CGSSdata\")
Send,%now_date_wg%2
Sm(0.2)
Fs("{ENTER}")
Loop,8
{
    Fs("{TAB}",1)
}
Fs("{ENTER}")

```

;3.4 下载问卷数据

Run,http://101.200.178.132/limesurvey_2017/index.php/admin/export/sa/exportresults/surveyid/96315

9

```
Ddck("LimeSurvey_2017 - Google Chrome",4)
```

```

Fs("f",0.5)
Fs("sj",0.5)
Zjdj(624,427)
Loop,19
{
    Fs("{DOWN}")
}
Fs("f",0.3)
Fs("e",0.3)
Fs("f",0.3)
Fs("k",0.3)
Fs("f",0.3)
Fs("l",0.3)
Fs("f",0.3)
Fs("m",0.3)
Fs("f",0.3)
Fs("sp",0.3)
Fs("^a",0.3)
Fs("{TAB}",0.3)
Fs("f",1)
Fs("sa",0.3)
Ddck("另存为",1)
Fs("{ENTER}",0.3)

```

;4.处理电核大表和样本清单

;4.1 打开 D:\CGSSdata\系统日期\ 目录

Run,D:\CGSSdata\%now_date_wg%2\

Ddck("",0.5,"2017")

;4.2 定位光标到文件：\CGSS2017 电话核查样本更新.xlsx

Fs("^f",0.5)

Srwz("电")

Ddck("",1,"搜索结果")

Loop 5

{

Fs("{TAB}")

}

Fs("{DOWN}")

;4.3 修改文件属性，取消锁定

Fs("!{ENTER}")

Ddck("CGSS2017 电话核查样本更新.xlsx 属性")

Fs("k")

Fs("{ENTER}")

Fs("!{F4}")

;4.4 处理电核大表

Run,D:\CGSSdata\%now_date_wg%2\CGSS2017 电话核查样本更新.xlsx

Ddck("CGSS2017 电话核查样本更新.xlsx - Excel",0.5)

Fs("^+d")

/*

Ctrl+Shift+D 用来执行如下处理电核大表的 Excel Visual Basic 程序

Sub 电核大表处理()

,

' 电核大表处理 宏

,

' 快捷键: Ctrl+Shift+D

,

Application.DisplayAlerts = False

a = "D:\CGSS 调查问卷统计\数据质控测试文档\"

b = Format\$(Now, "yyyy-mm-dd")

Sheets("CGSS2017 电核样本更新及进度管理表").Select

Columns("B:B").Select

Selection.NumberFormatLocal = "0_ "

ActiveWorkbook.SaveAs a & b & " dhdb.csv", xlCSV

```

Sheets("废卷情况汇总").Select
Columns("B:B").Select
Selection.NumberFormatLocal = "0_ "
ActiveWorkbook.SaveAs a & b & " fjbh.csv", xlCSV

Sheets("高缺失率问卷").Select
Columns("F:F").Select
Selection.NumberFormatLocal = "0_ "
ActiveWorkbook.SaveAs a & b & " gqslwj.csv", xlCSV

Sheets("电话回收率低访员").Select
ActiveWorkbook.SaveAs a & b & " hsldfy.csv", xlCSV

ActiveWorkbook.Close SaveChanges:=False
End Sub

```

*/

;4.5 退出电核大表

```

Ddck("",2,"csv")
Fs("!{F4}",1)

```

;4.6 打开 dropbox 共享文件夹

```

Run,C:\Users\xiaozhou\Dropbox\2017CGSS 全部样本 (1)\2017CGSS 全部样本
Ddck("2017CGSS 全部样本",1)
Fs("^f")

```

;4.7 搜索符合条件的样单文件并复制它们

```

Sr wz(".xlsx 作者:Yingfeng")
Ddck(".xlsx 作者:Yingfeng - “2017CGSS 全部样本” 中的搜索结果",4)
Loop 2
{
    Fs("{TAB}",0.5)
}
Fs("{SHIFTDOWN}{PGDN}{PGDN}{PGDN}{PGDN}{PGDN}{SHIFTUP}",4)
Fs("^c",1)
Fs("!{F4}",1)

```

;4.8 在一个特定文件夹中粘贴样单文件

```

Run,D:\CGSS 调查问卷统计\样本清单信息
Ddck("样本清单信息")
Fs("^v",10)
Fs("!{F4}")

```

;4.9 打开空 excel 执行 xlsx 转 csv 的 VB 程序

Run,C:\Program Files (x86)\Microsoft Office\root\Office16\EXCEL.EXE

Ddck("Excel")

Fs("^n")

Ddck("工作簿 1 - Excel")

Fs("^+m")

/*

Ctrl+Shift+M 用来执行如下处理样本清单的 Excel Visual Basic 程序

Sub 样本清单处理()

```
'
' 样本清单处理 宏
'
' 快捷键: Ctrl+Shift+M
'
    Dim fDir As String
    Dim fPath As String
    Dim sPath As String
    fPath = "D:\CGSS 调查问卷统计\样本清单信息\" '输入路径
    sPath = "D:\CGSS 调查问卷统计\样本清单信息\" '输出路径
    fDir = Dir(fPath) '路径下的第一个文件名
    Do While (fDir <> "")
        Workbooks.Open (fPath & fDir) '打开文件
        Sheets("正选样本").Select '定位到需要的 sheet
        ActiveWorkbook.SaveAs sPath & Mid(ActiveWorkbook.Name, 1, 7) & ".csv", xlCSV '另存为与源文件
        同名的 csv 文件
        ActiveWorkbook.Close SaveChanges:=False '关闭文件，不保存
        fDir = Dir '路径下的下一个文件名
    Loop
End Sub
```

*/

Sm(40)

;4.10 将在 D:\CGSS 调查问卷统计\样本清单信息目录 下转好的 csv 文件复制到 D:\CGSS 调查问卷统计\数据质控测试文档\样本清单信息 目录下，并删除 D:\CGSS 调查问卷统计\样本清单信息 目录下所有文件

Run,D:\My bat code\copy.dat

Sm(2)

Run,D:\My bat code\delete.dat

Sm(2)

/*

copy.bat 程序代码如下：

@echo off

xcopy D:\CGSS 调查问卷统计\样本清单信息*.csv D:\CGSS 调查问卷统计\数据质控测试文档\样本清单信息
\ /y

delete.bat 程序代码如下：

@echo off

del /q D:\CGSS 调查问卷统计\样本清单信息\

*/

;5.问卷下载结束后，退出 limesurvey 系统

;5.1 检测问卷数据是否下载完毕

DetectHiddenWindows, On

WinActivate, LimeSurvey_2017 - Google Chrome,

Sm(0.5)

MouseMove, 155, 841

Sc(45,843,45,843,"0xDFDFDF",0)

;5.2 退出 limesurvey 系统

Zj dj(854,728)

Loop,22

{

Fs("{UP}")

}

Zj dj(1351,117)

Loop,2

{

Fs("{DOWN}")

}

Fs("{ENTER}",1)

Fs("!{F4}",1)

;6.处理问卷数据

;

```
;6.1 打开 D:\CGSSdata\系统日期\
Run,D:\CGSSdata\%now_date_wg%2\
Ddck("",0.5,"2017")
```

```
;6.2 定位光标到文件：.\results-survey963159.xlsx
Fs("^f",1)
Srwx("results-survey963159.xlsx")
Ddck("",1,"搜索结果")
Loop 5
{
    Fs("{TAB}")
}
Fs("{DOWN}")
```

```
;6.3 修改文件属性，取消锁定
Fs("!{ENTER}")
Ddck("results-survey963159.xlsx 属性")
Fs("k")
Fs("{ENTER}",1)
Fs("!{F4}")
```

```
;6.4 处理问卷数据
Run,D:\CGSSdata\%now_date_wg%2\results-survey963159.xlsx
Ddck("results-survey963159.xlsx - Excel",3)
Fs("^+w")
```

/*

Ctrl+Shift+W 用来执行如下处理问卷数据的 Excel Visual Basic 程序

```
Sub 问卷数据另存为()
'
' 问卷数据另存为 宏
'
' 快捷键: Ctrl+Shift+W
'

Application.DisplayAlerts = False
a = "D:\CGSS 调查问卷统计\数据质控测试文档\"
b = Format$(Now, "yyyy-mm-dd")

Sheets("2017 年中国综合社会调查 (CGSS) ").Select
```

```

ActiveWorkbook.SaveAs a & b & " wjda.csv", xlCSV
ActiveWorkbook.Close SaveChanges:=False
End Sub

```

```
*/
```

```

;6.5 退出问卷数据
Ddck("",2,"csv")
Fs("!{F4}",1)

```

```

;
;7.运行 R 语言分析程序
;

```

```

Run,D:\Program Files\R\bin\i386\Rgui.exe
Ddck("RGui (32-bit)",2)
Srwz("source('D:\\CGSS 调查问卷统计\\数据质控测试文档\\CGSS 程序 0831 早.r')")
Fs("{ENTER}")

```

/* R 语言程序代码如下：

```

#说明：凡加"_m"的对象名意味着该对象为过渡对象
#所有自编函数以大写字母开头，其余为小写字母
rm(list=ls())
options(scipen=200)
options(warn=-1)
memory.limit(4095)
library("xlsx")
setwd("D:\\CGSS 调查问卷统计\\数据质控测试文档\\")

```

#函数一：时间开头文件路径

```

Sjktwjlj <- function(x,istd=TRUE){
  a <- Sys.Date()
  time <- if(istd==TRUE) a else a-1
  paste(".",time,"",x,sep="")
}

```

#函数二：读取 CSV

#该函数使用了一个外部函数：Sjktwjlj

```

Dqcsv <- function(x,usti=TRUE,istd=TRUE){
  if(usti==TRUE){
    da <- read.csv(Sjktwjlj(x,istd),header=TRUE,as.is=TRUE)
  }else{

```

```

        da <- read.csv(x,header=TRUE,as.is=TRUE)
    }
    iswx <- is.na(da[,2])
    if(any(iswx)){
        da <- da[!iswx,]
    }
    return(da)
}

```

#函数三：输出 CSV

#该函数使用了一个外部函数：Sjktwjlj

```

Sccsv <- function(obj,fn){
    write.csv(obj,Sjktwjlj(fn),row.names=FALSE,na="")
}

```

#函数四：样本清单处理

#该函数使用了一个外部函数：Dqcsv

```

Ybqdcl <- function(file){
    lj <- paste(".\\样本清单信息\\",file,sep="")
    da <- Dqcsv(lj,usti=FALSE)
    colnames(da) <- da[1,]
    da <- da[-1,1:13] #删除无效行列
    da <- da[da[,1]!="",]
    for(mu in 1:7){
        da[,mu] <- as.numeric(da[,mu])
    } #变量类型转换
    da$村居编码 <- as.numeric(substring(file,1,7))
    da$下发样本量 <- length(da$样本编号)
    return(da) #返回处理后的样本清单
}

```

#函数五：单/双/三变量排序

```

Px1 <- function(da,x,isjx=FALSE){
    da[order(da[,x],decreasing=isjx),]
}
Px2 <- function(da,x,y,isjx1=FALSE,isjx2=FALSE){
    da[order(da[,x],da[,y],decreasing=c(isjx1,isjx2)),]
}
Px3 <- function(da,x,y,z,isjx1=FALSE,isjx2=FALSE,isjx3=FALSE){
    da[order(da[,x],da[,y],da[,z],decreasing=c(isjx1,isjx2,isjx3)),]
}

```

#函数六：去重

```
Qc <- function(da,x,isla=FALSE){
  da[!duplicated(da[,x],fromLast=isla),]
}
```

#函数七：数据预处理

#该函数需要六个外部数据：ybqd、dhdb、fjqk、cjxx、ybqd、xfybl

#该函数使用了两个外部函数：Px1、Qc

#该函数内嵌了一个自编函数：Asp

```
Sjycl <- function(da){

  #删除无效列和测试 token
  da <- da[da$token>=1e+13,-c(1466:2101)]

  #替换 15 位 token 为 14 位
  da[da$token>1e+14,"token"] <- as.numeric(substring(da[da$token>1e+14,"token"],2,15))

  #对 attribute_1 缺失的样本添加 attribute_1，如果失败，直接删除
  cbqstj <- is.na(da$attribute_1)
  if(any(cbqstj)==TRUE){
    cbqswz <- which(cbqstj)
    cbqstk <- da[cbqswz,"token"]
    cb <- ybqd[match(cbqstk,ybqd$样本编号),"村居编码"]
    da[cbqswz,"attribute_1"] <- cb
    da <- da[!is.na(da$attribute_1),]
  }

  #部分特殊 token 的处理
  da[da$token==11142773304269,"token"] <- 11427733042692

  #去除重复 token
  da <- Px2(da,"lastpage","id",TRUE,TRUE)
  da <- Qc(da,"token")

  #更改时间变量类型为 POSIXlt
  Asp <- function(x){
    as.POSIXlt(da[,x],format="%Y-%m-%d %H:%M:%S")
  }
  da$submitdate <- Asp("submitdate")
  da$startdate <- Asp("startdate")
  da$datestamp <- Asp("datestamp")
  da$S3 <- Asp("S3")
  da$S7 <- Asp("S7")
}
```

```

da$S11 <- Asp("S11")
da$S15 <- Asp("S15")
da$S19 <- Asp("S19")
da$S23 <- Asp("S23")
da$S39 <- Asp("S39")
da$S44 <- Asp("S44")
da$S49 <- Asp("S49")
da$A00 <- Asp("A00")
da$Z8 <- Asp("Z8")

#标记是否上传
da$是否上传 <- 1

#标记是否完成
con1 <- da$lastpage==75 & da$Z1!="""
da$是否完成 <- ifelse(con1,1,0)

#标记是否提交
con2 <- (!is.na(da$Z7)) & (!is.na(da$submitdate))
da$是否提交 <- ifelse(con2,1,0)

#标记是否回收
mp <- ifelse(is.na(da[, "Z2"]), 0, da[, "Z2"]) #受访者手机
fp <- ifelse(is.na(da[, "Z3a"]), 0, da[, "Z3a"]) #受访者固定电话
con3 <- ((fp >= 2e+6 & fp < 1e+7) | (fp >= 2e+7 & fp < 1e+8) | (mp >= 130e+8 & mp < 140e+8) | (mp >= 145e+8 & mp < 146e+8) | (mp >= 147e+8 & mp < 148e+8) | (mp >= 149e+8 & mp < 154e+8) | (mp >= 155e+8 & mp < 160e+8) | (mp >= 170e+8 & mp < 174e+8) | (mp >= 175e+8 & mp < 179e+8) | (mp >= 180e+8 & mp < 190e+8)) & da$是否完成==1
da$是否回收 <- ifelse(con3,1,0)

#标记是否电核
da$是否电核 <- NULL
dhdb_m <- dhdb[!is.na(dhdb$处理意见.更新.),]
dhwz <- match(dhdb_m$token, da$token)
dhwz <- dhwz[!is.na(dhwz)]
da[dhwz, "是否电核"] <- 1
da[-dhwz, "是否电核"] <- 0

#标记是否废卷
da$是否废卷 <- NULL
fjwz <- match(fjqk$样本编号, da$token)
fjwz <- fjwz[!is.na(fjwz)]
da[fjwz, "是否废卷"] <- 1
da[-fjwz, "是否废卷"] <- 0

```

```

#合并村居信息列表、样本清单、下发样本量
da <- merge(da,cjxx,by.x="attribute_1",by.y="村居编码",all.x=TRUE)
da <- merge(da,ybqd[,1:13],by.x="token",by.y="样本编号",all.x=TRUE)
da <- merge(da,xfybl,by.x="attribute_1",by.y="村居编码",all.x=TRUE)

#按 id 重新排序
da <- Px1(da,"id")

#返回预处理后的数据
return(da)

}

#函数八：指标统计
Zbtj <- function(da,x,class="attribute_1",fuc=function(m) sum(m,na.rm=TRUE)){
  as.vector(tapply(da[,x],da[,class],fuc))
}

#函数九：应答率
Ydl <- function(da,fy=FALSE){
  if(fy==FALSE)round(da$已完成问卷数/da$下发样本量,4) else round(da$已完成问卷数/(da$已上传问卷
数),4)
}

#函数十：电话回收率
Dhhs1 <- function(da){
  round(da$电话回收数/da$已完成问卷数,4)
}

#函数十一：输出 xlsx
Scxlsx <- function(obj,fn,sn,isap=TRUE){
  write.xlsx(obj,Sjktwlj(fn),sheetName=sn,append=isap,row.names=FALSE,showNA=FALSE)
}

#函数十二：访员常去地区
Fycqdq <- function(x){
  a <- table(x)
  names(a)[which.max(a)]
}

#函数十三：筛选未完成省份
Sxwwcsf <- function(da){
  da[da$省!="黑龙江省" & da$省!="江苏省" & da$省!="浙江省" & da$省!="安徽省" & da$省!="湖南省"]
}

```

```
& da$省!="山东省" & da$省!="山西省" & da$省!="江西省" & da$省!="湖北省" & da$省!="河北省" &
da$省!="陕西省" & da$省!="辽宁省" & da$省!="广东省" & da$省!="河南省" & da$省!="重庆市" & da$省!="
四川省" & da$省!="云南省" & da$省!="广西壮族自治区" & da$省!="吉林省" & da$省!="北京市",]
}
```

#函数十四：筛选已完成省份

```
Sxywcsf <- function(da){
  da[da$省=="黑龙江省" | da$省=="江苏省" | da$省=="浙江省" | da$省=="安徽省" | da$省=="湖南省" |
da$省=="山东省" | da$省=="山西省" | da$省=="江西省" | da$省=="湖北省" | da$省=="河北省" | da$省=="
陕西省" | da$省=="辽宁省" | da$省=="广东省" | da$省=="河南省" | da$省=="重庆市" | da$省=="四川省" |
da$省=="云南省" | da$省=="广西壮族自治区" | da$省=="吉林省" | da$省=="北京市",]
}
```

#函数十五：筛选昨日问卷

```
Sxzrwj <- function(da){
  da[da$submitdate>=paste(Sys.Date()-1," 00:00:00",sep="") & da$submitdate<paste(Sys.Date(),"
00:00:00",sep="") & !is.na(da$submitdate),]
}
```

#函数十六：筛选近三日问卷

```
Sxjsrwj <- function(da){
  da[da$submitdate>=paste(Sys.Date()-3," 00:00:00",sep="") & !is.na(da$submitdate),]
}
```

#函数十七：筛选今日问卷

```
Sxjrwj <- function(da){
  da[da$submitdate>=paste(Sys.Date()," 00:00:00",sep="") & !is.na(da$submitdate),]
}
```

#函数十八：创建电核原始样本

```
Cdhr <- function(da){
  da$Z3 <- paste(da$Z3b,da$Z3a,sep="-")
  da$time <- difftime(da$submitdate,da$startdate,units="hours")
  return(da)
}
```

#函数十九：生成变量名

```
myfunc1 <- function(x,n){
  a <- NULL
  for(i in n){
    b <- paste(x,"_",i,sep="") # "name_i" 类型名称
    a <- c(a,b)
  }
  a
}
```



```

}
myfunc2 <- function(x,n){
  a <- NULL
  for(i in 1:n){
    b <- paste(x,".",i,".",sep="") # "name.i." 类型名称
    a <- c(a,b)
  }
  a
}
myfunc3 <- function(x,m,n){
  a <- NULL
  for(i in 1:m){
    for(j in 1:n){
      b <- paste(x,".",i,"_",j,".",sep="") # "name.i.j." 类型名称
      a <- c(a,b)
    }
  }
  a
}

```

#函数二十：电核变量选择

#该函数使用了三个外部函数：myfunc1,myfunc2,myfunc3

```

Dhvs <- function(da){

  #生成 attribute、S37、A3、S35 系列名称
  ATTRIBUTE <- myfunc1("attribute",c(1:11,13:14,16:17))
  S37 <- myfunc2("S37",3)
  A3 <- myfunc2("A3",3)
  S35 <- myfunc3("S35",10,3)
  #所需初始变量名
  var.jlx <- c("token","submitdate","Z2","Z3","Z1",S37,"A2",A3,"A58","A69","A96","A99","S0","S00","time","attribute_12",S3
5,"attribute_15",ATTRIBUTE,"省","县区","街道","村居","S36","Z9")
  var <- var.jlx[-c(2)]
  da_tm <- da[,var.jlx]
  da_tm <- da_tm[order(da_tm$attribute_1),]
  da_tj <- da[,var]
  da_tj <- da_tj[order(da_tj$attribute_1),]
  colnames(da_tj)[c(2:50,70:71)] <- c("attribute_20","attribute_66",myfunc1("attribute",21:28),myfunc1("attribute",41:47),"attribute_12",myfunc1("
attribute",29:40),myfunc1("attribute",48:65),"attribute_15","attribute_67","attribute_68")
  list("da_tj"=da_tj,"da_tm"=da_tm)
}

```

```
}
```

```
###-----
```

```
###程序主体部分
```

```
###-----
```

```
###一、样本清单(ybqd)
```

```
ybqdwjm <- list.files(".\\样本清单信息\\") #样本清单文件名
ybqdxgsj <- file.mtime(paste(".\\样本清单信息\\",ybqdwjm,sep=""))
#样本清单修改时间
ybqd_m <- lapply(ybqdwjm,Ybqdc1) #处理样本清单，输出列表
ybqd <- do.call("rbind",ybqd_m) #列表合并为数据框
ybqd <- Px1(ybqd,"下发样本量",TRUE) #按 下发样本量 降序排列
ybqd <- Qc(ybqd,"样本编号") #去除重复样本编号
Sccsv(ybqd,"ybqd.csv")
#输出样本清单
```

```
###二、下发样本量(xfybl)
```

```
xfybl <- unique(ybqd[,c("村居编码","下发样本量")]) #xfybl 由 村居编码 和 下发样本量 构成
xfybl <- Px2(xfybl,"村居编码","下发样本量",FALSE,TRUE) #xfybl 按 村居编码 升序排列，再按 下发样本量
降序排列
xfybl <- Qc(xfybl,"村居编码") #去除 xfybl 中的重复村居编码
```

```
###三、数据导入与预处理
```

```
wjsj <- Dqcsv("wjda.csv")
dhdb <- Dqcsv("dhdb.csv")
fjqk <- Dqcsv("fjbh.csv")
cjxx <- Dqcsv("cunjuliebiao.csv",usti=FALSE)
cfy.ysc <- Dqcsv("hsldfy.csv")
wjsj <- Sjycl(wjsj)
wjsjzr <- Sxzrwj(wjsj)
```

```
###四、村居统计(cjtj) 和 村居昨日统计(cjzrtj)
```

```
cjtj_m <- list()
cjtj_m$村居编码 <- Zbtj(wjsj,"attribute_1",fuc=function(m) m[1])
```

```

cjtj_m$下发样本量 <- Zbtj(wjsj,"下发样本量",fuc=function(m) m[1])
cjtj_m$已上传问卷数 <- Zbtj(wjsj,"是否上传")
cjtj_m$已完成问卷数 <- Zbtj(wjsj,"是否完成")
cjtj_m$已提交问卷数 <- Zbtj(wjsj,"是否提交")
cjtj_m$电话回收数 <- Zbtj(wjsj,"是否回收")
cjtj_m$电核量 <- Zbtj(wjsj,"是否电核")
cjtj_m$废卷数 <- Zbtj(wjsj,"是否废卷")
cjtj_m$应答率 <- Ydl(cjtj_m)
cjtj_m$电话回收率 <- Dhhs(cjtj_m)
cjtj <- data.frame(cjtj_m,stringsAsFactors=FALSE)
cjtj <- merge(cjxx,cjtj,by.x="村居编码",by.y="村居编码",all.x=TRUE)

cjzrtj_m <- list()
cjzrtj_m$村居编码 <- Zbtj(wjsjzr,"attribute_1",fuc=function(m) m[1])
cjzrtj_m$已上传问卷数 <- Zbtj(wjsjzr,"是否上传")
cjzrtj_m$已完成问卷数 <- Zbtj(wjsjzr,"是否完成")
cjzrtj_m$电话回收数 <- Zbtj(wjsjzr,"是否回收")
cjzrtj_m$电话回收率 <- Dhhs(cjzrtj_m)
cjzrtj <- data.frame(cjzrtj_m,stringsAsFactors=FALSE)
cjzrtj <- merge(cjxx,cjzrtj,by.x="村居编码",by.y="村居编码",all.y=TRUE)

```

###五、县区统计(xqtj) 和 县区昨日统计(xqzrtj)

```

xqtj_m <- list()
xqtj_m$省 <- Zbtj(cjtj,"省","县区编码",fuc=function(m) m[1])
xqtj_m$县区 <- Zbtj(cjtj,"县区","县区编码",fuc=function(m) m[1])
xqtj_m$下发样本量 <- Zbtj(cjtj,"下发样本量","县区编码")
xqtj_m$已上传问卷数 <- Zbtj(cjtj,"已上传问卷数","县区编码")
xqtj_m$已完成问卷数 <- Zbtj(cjtj,"已完成问卷数","县区编码")
xqtj_m$已提交问卷数 <- Zbtj(cjtj,"已提交问卷数","县区编码")
xqtj_m$电话回收数 <- Zbtj(cjtj,"电话回收数","县区编码")
xqtj_m$电核量 <- Zbtj(cjtj,"电核量","县区编码")
xqtj_m$废卷数 <- Zbtj(cjtj,"废卷数","县区编码")
xqtj_m$应答率 <- Ydl(xqtj_m)
xqtj_m$电话回收率 <- Dhhs(xqtj_m)
xqtj <- data.frame(xqtj_m,stringsAsFactors=FALSE)
xqtj[xqtj==0] <- NA

xqzrtj_m <- list()
xqzrtj_m$省 <- Zbtj(cjzrtj,"省","县区编码",fuc=function(m) m[1])
xqzrtj_m$县区 <- Zbtj(cjzrtj,"县区","县区编码",fuc=function(m) m[1])
xqzrtj_m$已上传问卷数 <- Zbtj(cjzrtj,"已上传问卷数","县区编码")
xqzrtj_m$已完成问卷数 <- Zbtj(cjzrtj,"已完成问卷数","县区编码")

```

```
xqzrtj_m$电话回收数 <- Zbtj(cjzrtj,"电话回收数","县区编码")
xqzrtj_m$电话回收率 <- Dhhs(xqzrtj_m)
xqzrtj <- data.frame(xqzrtj_m,stringsAsFactors=FALSE)
```

###六、省统计(stj) 和 省昨日统计(szrtj)

```
#为了与 google 大表保持一致，此部分程序较为繁杂
stj_m <- list()
stj_m$省 <- Zbtj(cjtj,"省","省编码",fuc=function(m) m[1])
stj_m$下发样本量 <- Zbtj(cjtj,"下发样本量","省编码")
stj_m$已上传问卷数 <- Zbtj(cjtj,"已上传问卷数","省编码")
stj_m$已完成问卷数 <- Zbtj(cjtj,"已完成问卷数","省编码")
stj_m$电话回收数 <- Zbtj(cjtj,"电话回收数","省编码")
stj_m$电核量 <- Zbtj(cjtj,"电核量","省编码")
stj_m$废卷数 <- Zbtj(cjtj,"废卷数","省编码")
stj <- data.frame(stj_m,stringsAsFactors=FALSE)
stj <- stj[match(c("北京市","天津市","河北省","山西省","辽宁省","吉林省","黑龙江省","上海市","江苏省","浙江省","安徽省","福建省","江西省","山东省","河南省","湖北省","湖南省","广东省","重庆市","广西壮族自治区","四川省","贵州省","云南省","陕西省","内蒙古","西藏自治区","甘肃省","青海省","宁夏回族自治区","新疆维吾尔自治区"),stj$省),] #按 google 大表上省份顺序重排 stj
stj[match("四川省",stj$省),2:7] <- stj[match("四川省",stj$省),2:7]+stj[match("重庆市",stj$省),2:7] #google 大表上重庆和四川放在一起
stj[match("重庆市",stj$省),2:7] <- NA
stj[match("内蒙古",stj$省),2:7] <- stj[match("内蒙古",stj$省),2:7]+stj[match("西藏自治区",stj$省),2:7]+stj[match("甘肃省",stj$省),2:7]+stj[match("青海省",stj$省),2:7]+stj[match("宁夏回族自治区",stj$省),2:7]+stj[match("新疆维吾尔自治区",stj$省),2:7] #google 大表上西北四省放在一起
stj$省[c(19,21,25)] <- c("深圳市","重庆+四川","西北四省")
stj$地方督导上报数量 <- "" #google 大表上在电话回收数与应答率之间有此列
stj$应答率 <- Ydl(stj)
stj$电话回收率 <- Dhhs(stj)
stj <- stj[1:25,c(1:5,8:10,6:7)] #按 google 大表上列名顺序重排 stj
stj[stj==0] <- NA

szrtj_m <- list()
szrtj_m$省 <- Zbtj(cjzrtj,"省","省编码",fuc=function(m) m[1])
szrtj_m$已上传问卷数 <- Zbtj(cjzrtj,"已上传问卷数","省编码")
szrtj_m$已完成问卷数 <- Zbtj(cjzrtj,"已完成问卷数","省编码")
szrtj_m$电话回收数 <- Zbtj(cjzrtj,"电话回收数","省编码")
szrtj_m$电话回收率 <- Dhhs(szrtj_m)
szrtj <- data.frame(szrtj_m,stringsAsFactors=FALSE)
```

###七、访员统计(fytj) 和 访员昨日统计(fyzrtj)

```

fysj <- wjsj[wjsj[,"S0"]!="",]
fysjzr <- wjsjzr[wjsjzr[,"S0"]!="",]
fytj_m <- list()
fytj_m$访员姓名 <- Zbtj(fysj,"S0","S00",fuc=function(m) m[1])
fytj_m$访员电话 <- Zbtj(fysj,"S00","S00",fuc=function(m) m[1])
fytj_m$常去县区编码 <- as.numeric(Zbtj(fysj,"县区编码","S00",Fycqdq))
fytj_m$省 <- Zbtj(fysj,"省","S00",Fycqdq)
fytj_m$县区 <- Zbtj(fysj,"县区","S00",Fycqdq)
fytj_m$已上传问卷数 <- Zbtj(fysj,"是否上传","S00")
fytj_m$已完成问卷数 <- Zbtj(fysj,"是否完成","S00")
fytj_m$电话回收数 <- Zbtj(fysj,"是否回收","S00")
fytj_m$电核量 <- Zbtj(fysj,"是否电核","S00")
fytj_m$废卷数 <- Zbtj(fysj,"是否废卷","S00")
fytj_m$应答率 <- Ydl(fytj_m,fy=TRUE)
fytj_m$电话回收率 <- Dhysl(fytj_m)
fytj <- data.frame(fytj_m,stringsAsFactors=FALSE)
fytj <- Px2(fytj,"常去县区编码","访员姓名")

fyzrtj_m <- list()
fyzrtj_m$访员姓名 <- Zbtj(fysjzr,"S0","S00",fuc=function(m) m[1])
fyzrtj_m$访员电话 <- Zbtj(fysjzr,"S00","S00",fuc=function(m) m[1])
fyzrtj_m$常去县区编码 <- as.numeric(Zbtj(fysjzr,"县区编码","S00",Fycqdq))
fyzrtj_m$省 <- Zbtj(fysjzr,"省","S00",Fycqdq)
fyzrtj_m$县区 <- Zbtj(fysjzr,"县区","S00",Fycqdq)
fyzrtj_m$已上传问卷数 <- Zbtj(fysjzr,"是否上传","S00")
fyzrtj_m$已完成问卷数 <- Zbtj(fysjzr,"是否提交","S00")
fyzrtj_m$电话回收数 <- Zbtj(fysjzr,"是否回收","S00")
fyzrtj_m$电话回收率 <- Dhysl(fyzrtj_m)
fyzrtj <- data.frame(fyzrtj_m,stringsAsFactors=FALSE)
fyzrtj <- Px2(fyzrtj,"常去县区编码","访员姓名")

```

###八、未完成省份差访员(fytj.cfy.wwc)

```

fytj.cfy <- fytj[fytj$已完成问卷数>3 & fytj$电话回收率<=0.7,]
fytj.cfy <- Px1(fytj.cfy,"电话回收率",TRUE)
fytj.cfy.wwc <- Sxwwcsf(fytj.cfy)
fytj.cfy.wsc <- fytj.cfy.wwc[match(setdiff(fytj.cfy.wwc$访员姓名,cfy.yzc$访员姓名),fytj.cfy.wwc$访员姓名),]
if(nrow(fytj.cfy.wsc)>0){
  fytj.cfy.wsc$更新日期 <- Sys.Date()
  fytj.cfy.wsc <- fytj.cfy.wsc[,c(13,1:12)]
  fytj.cfy.wsc <- Px1(fytj.cfy.wsc,"电话回收率",TRUE)
}

```

###九、进度统计表输出

```

Scxlsx(stj,"调查进度.xlsx","省",FALSE)
Scxlsx(xqtj,"调查进度.xlsx","县区")
Scxlsx(cjtj,"调查进度.xlsx","村居")
Scxlsx(fytj,"调查进度.xlsx","访员")
if(nrow(cjzrtj)>0){
  Scxlsx(szrtj,"调查进度.xlsx","省(昨日)")
  Scxlsx(xqzrtj,"调查进度.xlsx","县区(昨日)")
  Scxlsx(cjzrtj,"调查进度.xlsx","村居(昨日)")
  Scxlsx(fyzrtj,"调查进度.xlsx","访员(昨日)")
}
Sccsv(cjtj,"cjtj.csv")
Sccsv(cjzrtj,"cjzrtj.csv")
Sccsv(fytj.cfy,"完成问卷大于等于 4 且电话回收率小于等于 70%的访员名单.csv")
if(nrow(fytj.cfy.wsc)>0) Sccsv(fytj.cfy.wsc,"低电话回收率未上传访员名单.csv")
Sccsv(wjsj,"原始数据.csv")

```

###十、下发样本与上传问卷差额情况统计

```

cjtj_ysc <- cjtj[cjtj$已上传问卷数>0,]
xfschb <- merge(ybqd[, -ncol(ybqd)],cjtj_ysc,by.x="村居编码",by.y="村居编码",all.y=TRUE)
xfschb <- xfschb[,c("省","县区","街道","村居","样本编号","住户清单列表序号","建筑物编号","下发样本量","村居编码")]
xfschb.xyss <- xfschb[xfschb$下发样本量<=30,]
xfschb.dyss <- xfschb[xfschb$下发样本量>30,]

#差额位置 (cewz)
wjsj.yx <- wjsj[wjsj$S2a!="",]
cewz.xyss <- setdiff(xfschb.xyss$样本编号,wjsj.yx$token)
cewz.dyss <- setdiff(xfschb.dyss$样本编号,wjsj.yx$token)
ceybbh.xyss <- xfschb.xyss[match(cewz.xyss,xfschb.xyss$样本编号),]
ceybbh.dyss <- xfschb.dyss[match(cewz.dyss,xfschb.dyss$样本编号),]
ceybbh.xyss.wwc <- Sxwwcsf(ceybbh.xyss)
ceybbh.dyss.wwc <- Sxwwcsf(ceybbh.dyss)
ceybbh.xyss.ywc <- Sxywcsf(ceybbh.xyss)
ceybbh.dyss.ywc <- Sxywcsf(ceybbh.dyss)
cewz <- setdiff(wjsj$token,xfschb$样本编号)
cewz2 <- setdiff(wjsj$token,wjsj.yx$token)
xgxqd <- wjsj[match(cewz,wjsj$token),c("attribute_1","token")]
xgxqd <- Px1(xgxqd,"attribute_1")
htwxtk <- wjsj[match(cewz2,wjsj$token),c("attribute_1","token")]

```

```
htwxtk <- Px1(htwxtk,"attribute_1")
```

```
Sccsv(ceybbh.xyss.wwc[, -ncol(ceybbh.xyss.wwc)], "未完成省份无联系记录样本（下发样本小于等于 30）.csv")
Sccsv(ceybbh.dyss.wwc[, -ncol(ceybbh.dyss.wwc)], "未完成省份无联系记录样本（下发样本大于 30）.csv")
Sccsv(ceybbh.xyss.ywc[, -ncol(ceybbh.xyss.ywc)], "已完成省份无联系记录样本（下发样本小于等于 30）.csv")
Sccsv(ceybbh.dyss.ywc[, -ncol(ceybbh.dyss.ywc)], "已完成省份无联系记录样本（下发样本大于 30）.csv")
Sccsv(xgxqd, "样本清单中缺失的 token.csv")
Sccsv(htwxtk, "后台无效 token.csv")
```

```
wlxjlsta <- list("无联系记录总数"=nrow(ceybbh.xyss)+nrow(ceybbh.dyss),"无联系记录数（下发样本<=30）"
"=nrow(ceybbh.xyss),"无联系记录数（下发样本>30）"=nrow(ceybbh.dyss),"正在进行省份无联系记录总数"
"=nrow(ceybbh.xyss.wwc)+nrow(ceybbh.dyss.wwc),"正在进行省份无联系记录数（下发样本<=30）"
"=nrow(ceybbh.xyss.wwc),"正在进行省份无联系记录数（下发样本>30）"=nrow(ceybbh.dyss.wwc))
```

```
toyw.1 <- substring(file.mtime(Sjktwjlj("wjda.csv")),1,19)
toyw.2 <- substring(max(ybqdxgsj),1,19)
toyw.3 <- sum(stj$下发样本量,na.rm=TRUE)
toyw.4 <- sum(stj$已上传问卷数,na.rm=TRUE)
toyw.5 <- toyw.3-toyw.4
toyw.6 <- wxjlsta[[1]]
toyw.7 <- wxjlsta[[2]]
toyw.8 <- wxjlsta[[3]]
toyw.9 <- wxjlsta[[4]]
toyw.10 <- wxjlsta[[5]]
toyw.11 <- wxjlsta[[6]]
```

```
cat("雨薇姐：\n","\n","1.根据截止到",toyw.1,"的上传数据以及截止到",toyw.2,"的下发样本清单数据显示：
CGSS 共下发样本",toyw.3,"份，共上传问卷",toyw.4,"份，无联系记录问卷",toyw.5,"份；（备注：上传问卷剔除了
测试问卷、token 为 15 位的问卷以及重复 token 问卷）\n","\n","2.在无联系记录的",toyw.6,"份样本中，
",toyw.7,"份对应村居下发样本量小于等于 30，",toyw.8,"份对应村居下发样本量大于 30；\n","\n","3.正在进
行的省份中，无联系记录样本共计",toyw.9,"份，其中",toyw.10,"份对应村居下发样本量小于等于 30，
",toyw.11,"份对应村居下发样本量大于 30，详情见附件。 \n","\n","隆征帆
\n",substring(Sys.time(),1,19),sep="",file=Sjktwjlj("toyw.txt"))
```

###十一、无电话已完成问卷输出

```
S35 <- myfunc3("S35",10,3)
wdh.ywc_m <- wjsj[wjsj$是否完成==1 & wjsj$是否回收==0,]
wdh.ywc_m <- wdh.ywc_m[,c("attribute_1","submitdate","token","S34",S35)]
wdh.ywc_m <- merge(wdh.ywc_m,ybqd[,c("样本编号","实际门牌号","户主姓名","住宅地址")],by.x="token",by.y="样本编号",all.x=TRUE)
wdh.ywc_m <- merge(wdh.ywc_m,cjxx[,c("村居编码","省","县区","村居")],by.x="attribute_1",by.y="村居编码",all.x=TRUE)
```

```
wdh.ywc <- wdh.ywc_m[,c(1,38:40,2,3,35:37,4:34)]
wdh.ywc <- Px1(wdh.ywc,"attribute_1")
colnames(wdh.ywc)[c(1,5)] <- c("村居编码","样本编号")
Sccsv(wdh.ywc,"无电话已完成样本编号.csv")
```

###十二、高逻辑错误率输出专用程序

```
lg <- wjsj[wjsj$是否完成==1,]
Maer <- function(con){
  ifelse(is.na(ifelse(con,1,0)),0,ifelse(con,1,0))
}
Srjs <- function(da){
  as.numeric(da[,1])*10^6+as.numeric(da[,2])*10^5+as.numeric(da[,3])*10^4+as.numeric(da[,4])*10^3+a
  s.numeric(da[,5])*10^2+as.numeric(da[,6])*10+as.numeric(da[,7])
}

lg$er1 <- Maer(lg$A5.1==2 & lg$A6==1)
lg$er2 <- Maer(lg$A7c < lg$A3.1. & lg$A7c>0)
lg$er3 <- Maer(Srjs(lg[,myfunc3("A8a",1,7)]) < Srjs(lg[,myfunc3("A8b",1,7)]))
lg$er4 <- Maer(lg$A9==1 & lg$A10==4)
lg$er5 <- Maer((lg$A14/2)^2/lg$A13>35 & (lg$A14/2)^2/lg$A13<10)
lg$er6 <- Maer(lg$A15==5 & lg$A16==1)
lg$er7 <- Maer((lg$A28.1==1 & lg$A29==1) | (lg$A28.2==1 & lg$A29==2) | (lg$A28.3==1 & lg$A29==3)
| (lg$A28.4==1 & lg$A29==4) | (lg$A28.5==1 & lg$A29==5) | (lg$A28.6==1 & lg$A29==6))
lg$er8 <- Maer(lg$A30.4==5 & ( (lg$A30a>0 & lg$A30a<998) | (lg$A30b>0 & lg$A30b<998) ))
lg$er9 <- Maer(lg$A30c>0 & lg$A30c!=998 & lg$A30c!=999 & lg$A30e==2)
lg$er10 <- Maer(lg$A30e==2 & lg$A30g==1)
lg$er11 <- Maer(lg$A30e==2 & lg$A30h==1)
lg$er12 <- Maer(lg$A31.1==5 & lg$A31a==7)
lg$er13 <- Maer(lg$A31.1==5 & lg$A31b==7)
lg$er14 <- Maer(lg$A31.1==5 & lg$A31a==7 & lg$A31b==7)
lg$er15 <- Maer(lg$A43.a==1 & lg$A43e==1)
lg$er16 <- Maer(lg$A43.a==10 & lg$A43e==5)
lg$er17 <- Maer(lg$A51==1 & lg$A52==5)
lg$er18 <- Maer(lg$A49==1 & lg$A50==5)
lg$er19 <- Maer(lg$A12b>lg$A65 & lg$A12b<998)
lg$er20 <- Maer(lg$A68.1.< lg$A68a.1. | lg$A68.2.< lg$A68a.2.)
lg$er21 <- Maer(lg$A69==5 & lg$A69a ==1)
lg$er22 <- Maer(lg$A69==2 & lg$A69a!=1 & lg$A69a<98)
lg$er23 <- Maer(lg$A70-lg$A3.1.<16 & lg$A69>2)
lg$er24 <- Maer((lg$A69==3 | lg$A69==4 | lg$A69==5) & lg$A71b==9997)
lg$er25 <- Maer(lg$A69==3 & lg$A70 != lg$A71b & lg$A70<2018 & lg$A71b <2018)
lg$er26 <- Maer(lg$A70>lg$A71b & lg$A70<2018 & lg$A70>0 & lg$A71b <2018 & lg$A71b >0)
```



```

lg$er27 <- Maer(difftime(lg$submitdate,lg$startdate,units='hours')<=1)
lg$er28 <- Maer(difftime(lg$Z8,lg$submitdate,units='hours')>0.5)
lg$S37.2. <- sapply(lg$S37.2.,function(m) if(m=="男") 1 else(if(m=="女") 2 else NA))
lg$er29 <- Maer(lg$S37.2.!=lg$A2)
lg$er30 <- Maer(abs(lg$S37.3.-(2017-lg$A3.1.))>1)
lg$ernu <- apply(lg[,1511:1540],1,sum)
lg$erlc <- as.character(apply(lg[,1511:1540],1,function(m) which(m==1)))
lgtome <- lg[lg$ernu>=3,c("token","submitdate","erlc","省","县区","村居","attribute_1","S0","S00")]
Sccsv(lgtome,"lgtome.csv")

```

###十三、电话核查样本输出

```

dhhccs_m <- wjsj[wjsj$是否提交==1 & wjsj$是否回收==1,] #电话核查抽样初始样本
dhhccs <- Sxjsrwj(dhhccs_m) #已完成、有电话、未完成省份的样本
dhhccs <- rbind(dhhccs,wjsj[match(lg[lg$ernu>=1,"token"],wjsj$token),])
dhhccs <- dhhccs[dhhccs$是否回收==1,]
#合并逻辑错误点大于 2 的问卷
dhhccs <- Qc(dhhccs,"token")

```

#抽样 1：废卷访员

```

dcfyxm <- unique(fjqk$废卷访员) #待查访员姓名
if(length(dcfyxm)>0){
  a <- NULL #存储待查访员姓名在原数据中的行位置
  for(mu in dcfyxm){
    b <- which(dhhccs$S0==mu)
    a <- c(a,b)
  }
  dhhc.sp1 <- dhhccs[a,]
  dhhc.sp1 <- Cdhr(dhhc.sp1)
  dhhc.sp1 <- Sxwwcsf(dhhc.sp1)
}else{
  dhhc.sp1 <- NULL
}

```

#抽样 2：速度过快村居

```

dccjbm <- cjzrtj[cjzrtj$已完成问卷数>=10,"村居编码"]
if(length(dccjbm)>0){
  e <- NULL #存储待查访员姓名在原数据中的行位置
  for(mu in dccjbm){
    d <- which(dhhccs$attribute_1==mu)
    e <- c(e,d)
  }
  dhhc.sp2 <- dhhccs[e,]

```

```

    dhhc.sp2 <- Cdhr(dhhc.sp2)
  }else{
    dhhc.sp2 <- NULL
  }

#抽样 3：逻辑有问题样本
dhhc.sp3 <- rbind(dhhccs[dhhccs$省=="安徽省",],Sxwwcsf(dhhccs))
if(nrow(dhhc.sp3)>0){
  dhhc.sp3 <- Cdhr(dhhc.sp3)
}else{
  dhhc.sp3 <- NULL
}

#抽样 4 来源：昨日及今日新增问卷
dhhc.td <- rbind(Sxzwj(dhhccs),Sxjrwj(dhhccs))
dhhc.td <- Cdhr(dhhc.td)
dhhc.td <- dhhc.td[order(dhhc.td$S0,dhhc.td$startdate),]
dhhc.tdf <- dhhc.td[!duplicated(dhhc.td$S0),]

dhhc <- rbind(dhhc.sp1,dhhc.sp2,dhhc.sp3,dhhc.tdf)
wcftk <- setdiff(dhhc$token,dhdb$token)
dhhc <- dhhc[match(wcftk,dhhc$token),]

#电话核查抽样原始样本
if(nrow(dhhc)<100){
  dhhc <- rbind(dhhc.sp1,dhhc.sp2,dhhc.sp3,dhhc.td)
  #剔除与大表重复 token
  wcftk <- setdiff(dhhc$token,dhdb$token)
  dhhc <- dhhc[match(wcftk,dhhc$token),]
}

Sccsv(dhhc,"电话核查原始样本.csv")

#变量选择
dhyb_to_jm <- Dhvs(dhhc)[[1]]
dhyb_to_me <- Dhvs(dhhc)[[2]]

#适应 google 电核表
dhybcn <- colnames(dhyb_to_me)
dhdbcn <- colnames(dhdb)
xtjlm <- setdiff(dhdbcn,dhybcn) #需添加列名
kdf <- data.frame(matrix(NA,nrow(dhyb_to_me),length(xtjlm)))
kdf[is.na(kdf)] <- ""
colnames(kdf) <- xtjlm

```

```
dhyb_to_me <- cbind(dhyb_to_me,kdf)
dhyb_to_me <- dhyb_to_me[,dhdbcn]
if(nrow(dhyb_to_me)>0) dhyb_to_me$更新日期 <- Sys.Date()
dhyb_to_me <- dhyb_to_me[order(dhyb_to_me$attribute_1,dhyb_to_me$S0),]
```

```
Sccsv(dhyb_to_jm,"dhyb_to_jm.csv")
Sccsv(dhyb_to_me,"dhyb_to_me.csv")
```

###十四、高缺失率问卷输出专用程序

wjsjq <- wjsj[,c(2:1469,1,1470:1510)] #已预处理完的 wjsj,为了适应之前编写的程序, 调整了列位置

#筛选出已完成问卷,去除无效变量

wjsjq_m <- wjsjq[wjsjq\$是否完成==1,]

#剔除 A、B、C、D 部分不可能存在缺失的题目以及对缺失统计无用的变量名

```
wjsjq_m <- wjsjq_m[,c(124:335,338:343,345:350,353,355,357:365,368,372:380,388:389,393:394,396:396,398,400,403,405,407,409,411,413,415,417,419,463:464,488:489,492,494:500,503,506:507,509,511,513,516,528,530,534,537,549,551,558:563,569:577,580,582,584,586,592,595:601,603:608,611,614,616:622,625,629,631,633,636,645,647,650,653,659,661,664,667,673,675,841:845,848:850,852:858,861:865,968,996,998,1013,1015,1017,1029:1036,1039,1042,1063:1065,1067:1069,1071:1073,1075:1097,1099:1107,1109:1117,1119:1127,1129:1137,1139:1147,1149:1153,1155:1158,1160:1163,1165:1168,1170:1173,1175:1178,1180:1183,1188:1194,1197,1206:1251,1283,1320,1322,1344,1346:1351,1353:1358,1360:1365,1367:1372,1374:1379,1381:1386,1388:1393,1395:1400,1402:1407,1409:1414)]
wjsjq2 <- wjsjq_m
```

##替换 A 部分缺失答案最后两位数非 98,99 的变量为 98,99

##A 部分统计规则：最后两位为 98/99

```
Blxg_1 <- function(a,b){
  for(mu in a){
    x <- paste(mu, ".98.", sep="")
    y <- paste(mu, ".99.", sep="")
    cn <- colnames(wjsjq2)
    wjsjq2[,x] <- ifelse(wjsjq2[,x]==1|wjsjq2[,y]==1,b,wjsjq2[,x])
    colnames(wjsjq2)[which(cn==x)] <- mu
    wjsjq2 <- wjsjq2[, -match(y,cn)]
  }
}
Blxg_1(c("A12","A56","A67","A79"),98)
```

```
Blxg_2 <- function(a,b,d,e){
  for(mu in a){
    wjsjq2[,mu] <- ifelse(wjsjq2[,mu]==b|wjsjq2[,mu]==d,e,wjsjq2[,mu])
  }
}
```

```

    }
  }
  Blxg_2(c("A29","A8a.1_7.","A8b.1_7.","A62.1_7.","A75a.1_7.","A75b.1_7."),8,9,98)

##替换 B 部分缺失答案最后一位数非 8 的变量为 8,并处理会引起统计错误的相关变量
##B 部分统计规则：最后一位为 8 或 9
Blxg_1(c("B34","B36"),8)
Blxg_2(c("B9a","B9b","B12"),8,9,0) #将 8,9 对应并非缺失的变量值改为 0

##替换 C 部分缺失答案最后一位数非 8 的变量为 8,并处理会引起统计错误的相关变量
##C 部分统计规则：最后一位为 8 或 9
Blxg_1(c("C57"),8)
Blxg_2(c("C12.1.","C12.2.","C17","C22","C23","C24","C25","C26","C28","C29","C38"),8,9,0)

##替换 D 部分缺失答案最后一位数非 8 的变量为 8,并处理会引起统计错误的相关变量
##D 部分统计规则：最后一位为 8 或 9
Blxg_1(c("D61e"),8)
Blxg_2(c("D10","D22","D23","D39.1.","D39.2.","D39.3.","D39.4.","D39.5.","D39.6."),8,9,0)

cn.wjsjq2 <- colnames(wjsjq2)

#非空题数
Fkts <- function(x,class){
  czs <- which(substring(cn.wjsjq2,1,1)==class)
  a <- length(which(!is.na(x[czs])))
  if(a<15) 0 else a
}

wjsjq2$非空题数 A <- apply(wjsjq2,1,Fkts,"A")
wjsjq2$非空题数 B <- apply(wjsjq2,1,Fkts,"B")
wjsjq2$非空题数 C <- apply(wjsjq2,1,Fkts,"C")
wjsjq2$非空题数 D <- apply(wjsjq2,1,Fkts,"D")
wjsjq2$非空题数 <- wjsjq2$非空题数 A+wjsjq2$非空题数 B+wjsjq2$非空题数 C+wjsjq2$非空题数 D

#缺失题数
Qstj <- function(x,class){
  czs <- which(substring(cn.wjsjq2,1,1)==class)
  a <- x[czs]
  b <- nchar(a)
  if(class=="A"){
    d <- substring(a,b-1,b)
    length(which(d==98|d==99))
  }else{
    d <- substring(a,b,b)
  }
}

```

```

length(which(d==8|d==9))
}
}

wjsjq2$缺失题数 A <- apply(wjsjq2,1,Qstj,"A")
wjsjq2$缺失题数 B <- apply(wjsjq2,1,Qstj,"B")
wjsjq2$缺失题数 C <- apply(wjsjq2,1,Qstj,"C")
wjsjq2$缺失题数 D <- apply(wjsjq2,1,Qstj,"D")
wjsjq2$缺失题数 <- wjsjq2$缺失题数 A+wjsjq2$缺失题数 B+wjsjq2$缺失题数 C+wjsjq2$缺失题数 D

wjsjq2$缺失率 A <- wjsjq2$缺失题数 A/wjsjq2$非空题数 A
wjsjq2$缺失率 B <- wjsjq2$缺失题数 B/wjsjq2$非空题数 B
wjsjq2$缺失率 C <- wjsjq2$缺失题数 C/wjsjq2$非空题数 C
wjsjq2$缺失率 D <- wjsjq2$缺失题数 D/wjsjq2$非空题数 D
wjsjq2$缺失率 <- wjsjq2$缺失题数/wjsjq2$非空题数

wjsjq2 <- wjsjq2[order(wjsjq2$缺失率,wjsjq2$缺失率
A,substring(wjsjq2$attribute_1,1,2),wjsjq2$S0,decreasing=T),]

qstj_m <- wjsjq2
qstj <- qstj_m[order(qstj_m$缺失率,qstj_m$缺失率 A,qstj_m$attribute_1,qstj_m$S0,decreasing=T),c("省","县
区","村居","attribute_1","token","submitdate","S0","S00","缺失率","缺失率 A","缺失率 B","缺失率 C","缺失率
D","非空题数","缺失题数","非空题数 A","缺失题数 A","非空题数 B","缺失题数 B","非空题数 C","缺失题数 C","
非空题数 D","缺失题数 D","Z1","Z2","Z3b","Z3a")]
sbm <- substring(qstj$attribute_1,1,2)
qstj.wwc <- qstj[sbm!="13" & sbm!="14" & sbm!="21" & sbm!="23" & sbm!="32" & sbm!="33" & sbm!="34"
& sbm!="36" & sbm!="36" & sbm!="37" & sbm!="41" & sbm!="42" & sbm!="43" & sbm!="44" & sbm!="50"
& sbm!="51" & sbm!="53" & sbm!="61",#选取正在进行的省份的有电话的已完成问卷]
qstj.wsc <- qstj.wwc[qstj.wwc$缺失率>=0.3|qstj.wwc$缺失率 A>=0.3|qstj.wwc$缺失率 B>=0.3|qstj.wwc$缺失
率 C>=0.3|qstj.wwc$缺失率 D>=0.3,]
qstj.wsc <- qstj.wsc[!is.na(qstj.wsc$attribute_1),]
qstj.wsc <- qstj.wsc[order(qstj.wsc$submitdate,qstj.wsc$缺失率,qstj.wsc$缺失率 B,qstj.wsc$缺失率
C,qstj.wsc$缺失率 D,decreasing=T),]

#剔除已上传的 token
gqslwjbh <- read.csv(Sjktwjlj("gqslwj.csv"),header=T,as.is=T)[,"样本编号"]
gqslwjbh <- gqslwjbh[!is.na(gqslwjbh)]
qstj.wsc <- qstj.wsc[match(setdiff(qstj.wsc$token,gqslwjbh),qstj.wsc$token),]
qstj.wsc$Z3b <- paste(qstj.wsc$Z3b,qstj.wsc$Z3a,sep="-")
qstj.wsc <- qstj.wsc[,c(1:8,24,9:23,25:26)]
if(nrow(qstj.wsc)>0){
  qstj.wsc$更新日期 <- Sys.Date()
  qstj.wsc <- qstj.wsc[,c(ncol(qstj.wsc),1:(ncol(qstj.wsc)-1))]
}else{

```

```
qstj.wsc$更新日期 <- NULL
}
```

```
Sccsv(qstj,"qstj.csv")
Sccsv(qstj.www,"qstjwww.csv")
if(nrow(qstj.wsc)>0) Sccsv(qstj.wsc,"qstjwsc.csv")
```

###十五、周报

```
zb1 <- substring(max(ybqdxgsj),1,19) #下发样本更新截止时间
zb2 <- substring(file.mtime(Sjktwjlj("wjda.csv")),1,19) #上传样本更新截止时间
zb3 <- substring(file.mtime(Sjktwjlj("dhdb.csv")),1,19) #电核样本更新截止时间
zb4 <- length(unique(substring(cjtj_ysc$村居编码,1,2))) #覆盖省份个数
zb5 <- length(unique(substring(cjtj_ysc$村居编码,1,6))) #覆盖县区个数
zb6 <- length(unique(substring(cjtj_ysc$村居编码,1,7))) #覆盖村居个数
zb7 <- sum(stj$下发样本量,na.rm=T) #总下发样本量
zb8 <- sum(stj$已上传问卷数,na.rm=T) #总上传样本量
zb9 <- sum(stj$已完成问卷数,na.rm=T) #总完成样本量
zb10 <- sum(stj$电话回收数,na.rm=T) #总电话回收量
zb11 <- sum(stj$废卷数,na.rm=T) #总废卷数
zwcl <- zb9/(480*25)
zb12 <- paste(round(zwcl*100,2),"%",sep="") #总完成率
zydl <- zb9/zb7
zb13 <- paste(round(zydl*100,2),"%",sep="") #总应答率
zdhsl <- zb10/zb9
zb14 <- paste(round(zdhsl*100,2),"%",sep="") #总电话回收率
zb15 <- length(unique(substring(dhdb$attribute_1,1,2)))-1 #电话核查省份个数
zb16 <- length(unique(substring(dhdb$attribute_1,1,6)))-1 #电话核查区县个数
zb17 <- length(unique(substring(dhdb$attribute_1,1,7)))-1 #电话核查村居个数
zb18 <- sum(stj$电核量,na.rm=T) #总电核量
zb19 <- paste(round(zb18*100/zb10,2),"%",sep="") #电核覆盖率

n <- 1
stjo <- read.xlsx(paste("\\",Sys.Date()-n," 调查进度 .xlsx",sep=""),1,encoding="UTF-8",stringsAsFactors=FALSE)

zb23 <- (zb7-sum(stjo$下发样本量,na.rm=T))/n #本周日均新增下发样本量
zb24 <- (zb8-sum(stjo$已上传问卷数,na.rm=T))/n #本周日均新增上传样本量
zb25 <- (zb9-sum(stjo$已完成问卷数,na.rm=T))/n #本周日均新增完成样本量
zb26 <- (zb10-sum(stjo$电话回收数,na.rm=T))/n #本周日均新增电话回收数
zb27 <- (zb18-sum(stjo$电核量,na.rm=T))/n #本周日均新增电核量
zb28 <- (zb11-sum(stjo$废卷数,na.rm=T))/n #本周日均新增废卷数
zwclj <- zwcl-(zb9-zb25*n)/(480*25)
```

```

zb29 <- round(zwclj*100,2) #总完成率增减百分比
zydlj <- zydl-(zb9-zb25*n)/(zb7-zb23*n)
zb30 <- round(zydlj*100,2) #总应答率增减百分比
zdhhsj <- zdhhs-(zb10-zb26*n)/(zb9-zb25*n)
zb31 <- round(zdhhsj*100,2) #总电话回收率增减百分比

ws <- function(x,n,y){
  paste(x," ",n," ",y,sep="")
}
zb1 <- ws("根据截止到",zb1,"的下发样本数据, ")
zb2 <- ws("截止到",zb2,"的上传样本数据, ")
zb3 <- ws("以及截止到",zb3,"的电话核查数据统计得到：")
zb4 <- ws("CGSS 项目已在全国",zb4,"个省")
zb5 <- ws("、",zb5,"个县区")
zb6 <- ws("和",zb6,"个村居")
zb7 <- ws("下发样本共计",zb7,"份、")
zb8 <- ws("服务器收到上传样本",zb8,"份, ")
zb9 <- ws("其中已完成样本",zb9,"份、")

zb10 <- ws("留有联系方式样本",zb10,"份、")
zb11 <- ws("作废样本",zb11,"份。")
zb12 <- ws("总完成率",zb12,"、")
zb13 <- ws("总应答率",zb13,"、")
zb14 <- ws("总电话回收率",zb14,".")

zb15 <- ws("目前电核小组已核查了全国",zb15,"个省")
zb16 <- ws("、",zb16,"个县区")
zb17 <- ws("和",zb17,"个村居")
zb18 <- ws("的",zb18,"份样本,")
zb19 <- ws("电核覆盖留有联系方式样本的",zb19,".")

zb23 <- ws("与上周数据相比, 本周, CGSS 项目日均新增下发样本",zb23,"份、")
zb24 <- ws("日均新增上传样本",zb24,"份、")
zb25 <- ws("日均新增完成样本",zb25,"份、")
zb26 <- ws("日均新增电话回收数",zb26,"个、")
zb27 <- ws("日均新增电话核查量",zb27,"份、")
zb28 <- ws("日均新增作废样本",zb28,"份、")
zb29 <- ifelse(zb29>=0,ws("总完成率增加",zb29,"个百分点,"),ws("总完成率减少",abs(zb29),"个百分点,"))
zb30 <- ifelse(zb30>=0,ws("总应答率增加",zb30,"个百分点,"),ws("总应答率减少",abs(zb30),"个百分点,"))
zb31 <- ifelse(zb31>=0,paste("总电话回收率增加",zb31,"个百分点。"),paste("总电话回收率减少",abs(zb31),"
个百分点。"))

a1 <- NULL

```

```

for(mu in c(1:14)){
  b <- get(paste("zb",mu,sep=""))
  a1 <- paste(a1,b,sep="")
}

a2 <- NULL
for(mu in c(15:19)){
  b <- get(paste("zb",mu,sep=""))
  a2 <- paste(a2,b,sep="")
}

a3 <- NULL
for(mu in c(23:31)){
  b <- get(paste("zb",mu,sep=""))
  a3 <- paste(a3,b,sep="")
}

#缺失汇报

qsn <- nrow(qstj)
qh1 <- length(which(qstj$缺失率>=0.3)) #总缺失率>=0.3 的问卷数
qh2 <- round(qh1*100/qsn,2)
qh3 <- length(which(qstj$缺失率 A>=0.3)) #缺失率 A>=0.3 的问卷数
qh4 <- round(qh3*100/qsn,2)
qh5 <- length(which(qstj$缺失率 B>=0.3)) #缺失率 B>=0.3 的问卷数
qh6 <- round(qh5*100/qsn,2)
qh7 <- length(which(qstj$缺失率 C>=0.3)) #缺失率 C>=0.3 的问卷数
qh8 <- round(qh7*100/qsn,2)
qh9 <- length(which(qstj$缺失率 D>=0.3)) #缺失率 D>=0.3 的问卷数
qh10 <- round(qh9*100/qsn,2)
qh11 <- length(which(qstj$缺失率 A>=0.3 | qstj$缺失率 B>=0.3 | qstj$缺失率 C>=0.3 | qstj$缺失率 D>=0.3))
qh12 <- round(qh11*100/qsn,2)

qh1 <- ws("缺失率方面，总缺失率大于等于 30%的问卷共计",qh1,"份，")
qh2 <- ws("占比",qh2,"%，")
qh3 <- ws("A 部分缺失率大于等于 30%的问卷共计",qh3,"份，")
qh4 <- ws("占比",qh4,"%，")
qh5 <- ws("B 部分缺失率大于等于 30%的问卷共计",qh5,"份，")
qh6 <- ws("占比",qh6,"%，")
qh7 <- ws("C 部分缺失率大于等于 30%的问卷共计",qh7,"份，")
qh8 <- ws("占比",qh8,"%，")
qh9 <- ws("D 部分缺失率大于等于 30%的问卷共计",qh9,"份，")
qh10 <- ws("占比",qh10,"%.")
qh11 <- ws("必做部分缺失率大于等于 30%或选做部分缺失率大于等于 30%的问卷共计",qh11,"份，")

```



```
qh12 <- ws("占比",qh12,"%")

b1 <- NULL
for(mu in c(1:12)){
  b <- get(paste("qh",mu,sep=""))
  b1 <- paste(b1,b,sep="")
}

cat(a1,"\n\n",a2,"\n\n",a3,"\n\n",b1,sep="",file=Sjktwjlj("周报.txt"))
```

###十六、统计分析

```
wjsj.tjfx <- wjsj
```

##16.1 普查数据读入与格式修改

```
nlsj <- read.xlsx(".\\第六次人口普查数据.xlsx",1,encoding="UTF-8",stringsAsFactors=FALSE)
xbsj <- read.xlsx(".\\第六次人口普查数据.xlsx",2,encoding="UTF-8",stringsAsFactors=FALSE)
nlsj[is.na(nlsj)] <- ""
xbsj[is.na(xbsj)] <- ""
nlsj <- nlsj[,1:2]
xbsj <- xbsj[,1:2]
nlsj <- nlsj[nlsj[,1]!="",]
xbsj <- xbsj[xbsj[,1]!="",]
```

##16.2 问卷数据年龄、性别列加载

```
wjsj.tjfx$nl <- 2017-as.numeric(wjsj.tjfx$A3.1.)
wjsj.tjfx$xb <- wjsj.tjfx$A2
```

##16.3 筛选年龄正常、性别非空的数据

```
wjsj.tjfx <- wjsj.tjfx[wjsj.tjfx$nl>17 & wjsj.tjfx$nl<130 & wjsj.tjfx$xb!="",]
```

##16.4 CGSS 调查年龄数据

```
dcnlsj_m <- table(wjsj.tjfx$nl)
dcnlsj <- list()
dcnlsj$年龄 <- as.numeric(names(dcnlsj_m))
dcnlsj$CGSS 人数 <- as.numeric(as.vector(dcnlsj_m))
dcnlsj <- data.frame(dcnlsj,stringsAsFactors=FALSE)
dcnlsj[dcnlsj$年龄>=100,"年龄"] <- 100
dcnlsj[dcnlsj$年龄==100,"CGSS 人数"] <- sum(dcnlsj[dcnlsj$年龄==100,"CGSS 人数"])
dcnlsj <- dcnlsj[!duplicated(dcnlsj$年龄),]
dcnlsj <- dcnlsj[order(dcnlsj$年龄),]
```

##16.5 CGSS 调查性别数据

```

dcxbsj_m <- table(wjsj.tjfx$xb)
dcxbsj <- list()
dcxbsj$性别 <- ifelse(names(dcxbsj_m)=="1","男","女")
dcxbsj$CGSS 人数 <- as.numeric(as.vector(dcxbsj_m))
dcxbsj <- data.frame(dcxbsj,stringsAsFactors=FALSE)

##16.6 CGSS 调查年龄、性别数据与普查数据的合并
pdnljsj <- merge(nlsj,dcnljsj,by.x="年龄",by.y="年龄",all.x=TRUE)
pdnljsj[is.na(pdnljsj)] <- 0
pdxbsj <- merge(xbsj,dcxbsj,by.x="性别",by.y="性别",all.x=TRUE)

pdnljsj$人数比例 <- round(as.numeric(pdnljsj$人数)/sum(as.numeric(pdnljsj$人数)),4)
pdnljsj$CGSS 人数比例 <- round(pdnljsj$CGSS 人数/sum(pdnljsj$CGSS 人数),4)
pdxbsj$人数比例 <- round(as.numeric(pdxbsj$人数)/sum(as.numeric(pdxbsj$人数)),4)
pdxbsj$CGSS 人数比例 <- round(pdxbsj$CGSS 人数/sum(pdxbsj$CGSS 人数),4)

colnames(pdnljsj) <- c("年龄","普查人数","CGSS 人数","人口普查","CGSS")
colnames(pdxbsj) <- c("性别","普查人数","CGSS 人数","人口普查","CGSS")
pdnljsj <- pdnljsj[,c(1,4:5,2:3)]
pdxbsj <- pdxbsj[,c(1,4:5,2:3)]

Scxlsx(pdnljsj,"CGSS 年龄性别分布.xlsx","年龄分布",FALSE)
Scxlsx(pdxbsj,"CGSS 年龄性别分布.xlsx","性别分布")

```

*/

Sc(20,441,45,457,"0x0000FF",0)

```

;
;8.上传电核数据到 limesurvey 系统
;

```

```

;8.1 用 notepad++更改 dhyb_to_jm.csv 的编码格式
Rjlj = D:\Program Files\Notepad++\notepad++.exe
Wjlj = D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% dhyb_to_jm.csv
Run,%Rjlj% "%Wjlj%"
WinWait, D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% dhyb_to_jm.csv - Notepad++,
IfWinNotActive, D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% dhyb_to_jm.csv - Notepad++,
WinActivate, D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% dhyb_to_jm.csv - Notepad++,
WinWaitActive, D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% dhyb_to_jm.csv - Notepad++,
Fs("{ALT}",0.5)
Fs("{n}",0.5)
Loop,3
{

```

```

    Fs("{UP}",0.2)
}
Fs("{ENTER}")
Fs("^s",1)
;Fs("!{F4}")

```

;8.2 登录 limesurvey 系统

```

Run,http://101.200.178.132/limesurvey_2017/index.php/admin
Ddck("LimeSurvey_2017 - Google Chrome")
Srww("jiamin")
Fs("{TAB}")
Fs("123456")
Fs("{ENTER}")
Sc(501,216,553,271,"0x0BEDCB",2)

```

;8.2 上传电核数据

```

Run,http://101.200.178.132/limesurvey_2017/index.php/admin/tokens/sa/import/surveyid/637954
Ddck("LimeSurvey_2017 - Google Chrome",5)
Zjdj(1142,401,0.5)
Loop 3
{
    Fs("{DOWN}",1)
}
Fs("f",0.5)
Fs("sd")

```

```

Ddck("打开",1)
Fs("^f",1)
Fs("+{TAB}",0.5)
Fs("{ENTER}",0.5)
Srww("D:\CGSS 调查问卷统计\数据质控测试文档")
Fs("{ENTER}",0.5)
Fs("^f",0.5)
Fs("{SHIFT}",0.5)
Send,AND "%now_date% dhyb_to_jm.csv"
Sm(1)
Fs("{ENTER}")
Sm(4)
Loop 3
{
    Fs("{TAB}",0.5)
}
Fs("{DOWN}",0.5)
Fs("{HOME}",0.5)

```

```

Loop,3
{
    Fs("{TAB}",0.5)
}
Fs("{ENTER}",1)

```

```

Fs("f",0.3)
Fs("j",0.3)
Fs("f",0.3)
Fs("k",0.3)
Fs("f",0.3)
Fs("l",0.3)
Fs("f",0.3)
Fs("p",0.3)

```

```

Loop 4
{
    Fs("{DOWN}",0.3)
}
Fs("{TAB}",0.3)
Fs("f",0.3)
Fs("ss",0.3)

```

```

;9.上传数据到电核大表
;

```

;9.1 上传电核数据到电核大表

```

Run,https://docs.google.com/spreadsheets/d/1tfRQ_n3-4bTEemeqWZlHvLNIV-
eAjKCv9dpNVs7ETSc/edit#gid=0

```

```

Sc(1345,125,1416,153,"0xF78946",2)
Zjdj(119,318)
Fs("^ {DOWN}")
Sc(8,759,54,810,"0xDDDDDD",2)
Fs("{DOWN}",2)

```

;9.2 处理电核样本

```

FileGetSize, OutputVar, D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% dhyb_to_me.csv, K
If ErrorLevel
{
    Sm()
}
else
{

```

```
Run,D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% dhyb_to_me.csv
Ddck("",1,"2017")
Fs("^+y",2)
```

```
/*
```

Ctrl+Shift+Y 用来执行如下处理电核样本的 Excel Visual Basic 程序

```
Sub 电核样本处理()
'
' 电核样本处理 宏
'
' 快捷键: Ctrl+Shift+Y
'

Columns("A:A").Select
Columns("A:A").EntireColumn.AutoFit
Columns("B:B").Select
Selection.NumberFormatLocal = "0_ "
Columns("C:C").Select
Columns("C:C").EntireColumn.AutoFit
Columns("T:U").Select
Selection.NumberFormatLocal = "0_ "
Columns("P:P").Select
Selection.NumberFormatLocal = "0_ "
Columns("CC:CC").Select
Selection.NumberFormatLocal = "0_ "

maxrow = ActiveSheet.UsedRange.Rows.Count
Range("A2:CF" & maxrow).Select
Selection.Copy
End Sub
```

```
*/
```

```
Fs("'{SPACE}'")
Fs("{n}")
Ddck("CGSS2017 电话核查样本更新 - Google 表格 - Google Chrome",2)
Fs("^v")
Sc(449,766,811,816,"0xFFF3EC",0)
}
```

;9.3 上传村居进度到电核大表

```
Fs("^+{PGDN}")
Sc(141,262,141,262,"0xA8D7B6",0)
```

```

Fs("^+{PGDN}")
Sc(141,262,141,262,"0xC9C4A2",0)
Fs("^+{PGDN}")
Sc(141,262,141,262,"0xBDA6D5",0)
Fs("^+{HOME}")
Fs("{UP}")
Loop 3
{
    Fs("{LEFT}",1)
}
Send, 省 %now_date%
Sm(2)
Fs("{ENTER}")
Fs("^+{HOME}")

Run,D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% cjtj.csv
Ddck("",1,"cjtj.csv")
Fs("^+c",2)

```

/*

Ctrl+Shift+C 用来执行如下处理村居进度的 Excel Visual Basic 程序

```

Sub 村居进度复制()
'
' 村居进度复制 宏
'
' 快捷键: Ctrl+Shift+C
'
    Range("H2:N499").Select
    Selection.Copy
End Sub

```

*/

```

Fs("!{SPACE}")
Fs("{n}")
Ddck("CGSS2017 电话核查样本更新 - Google 表格 - Google Chrome",1)
Fs("^v")
Sc(449,766,811,816,"0xFFFF3EC",0)

;9.4 上传高缺失率问卷到电核大表
Fs("^+{PGDN}")
Sc(141,262,141,262,"0x99E5FF",0)

```

```

Fs("^ {DOWN}")
Sc(8,759,54,810,"0xDDDDDD",2)
Fs("{DOWN}")
FileGetSize, OutputVar, D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% qstjwsc.csv, K
If ErrorLevel
{
    Sm()
}
else
{
    Run,D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% qstjwsc.csv
    Ddck("",1,"qstjwsc.csv")
    Fs("^+q",2)

/*

```

Ctrl+Shift+Q 用来执行如下处理缺失问卷的 Excel Visual Basic 程序

```

Sub 未上传缺失问卷处理()
'
' 未上传缺失问卷处理 宏
'
' 快捷键: Ctrl+Shift+Q
'

Columns("A:A").Select
Columns("A:A").EntireColumn.AutoFit

Columns("F:F").Select
Selection.NumberFormatLocal = "0_ "

Columns("G:G").Select
Columns("G:G").EntireColumn.AutoFit

Columns("I:I").Select
Columns("I:I").EntireColumn.AutoFit

Columns("K:O").Select
Selection.NumberFormatLocal = "0.00%"

Columns("Z:AA").Select
Selection.NumberFormatLocal = "0_ "

maxrow = ActiveSheet.UsedRange.Rows.Count
Range("A2:AA" & maxrow).Select

```

```

        Selection.Copy
    End Sub

```

```

*/

```

```

    Fs("!{SPACE}")
    Fs("{n}")
    Ddck("CGSS2017 电话核查样本更新 - Google 表格 - Google Chrome",1)
    Fs("^v")
    Sc(449,766,811,816,"0xFF3EC",0)
}

```

;9.5 上传低电话回收访员到电核大表

```

Fs("^+{PGDN}")
Sc(141,262,141,262,"0xA8D7B6",0)
Fs("^DOWN")
Sc(8,759,54,810,"0xDDDDDD",2)
Fs("{DOWN}")
FileGetSize, OutputVar, D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% 低电话回收率未上传访
员名单.csv, K

```

```

    If ErrorLevel

```

```

    {

```

```

        Sm()

```

```

    }

```

```

else

```

```

{

```

```

    Run,D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% 低电话回收率未上传访员名单.csv
    Ddck("",1,"低电话回收率未上传访员名单.csv")
    Fs("^+f",2)

```

```

/*

```

Ctrl+Shift+Q 用来执行如下处理低电话回收率访员的 Excel Visual Basic 程序

```

Sub 低电话回收率访员处理()

```

```

'

```

```

' 低电话回收率访员处理 宏

```

```

'

```

```

'

```

```

' 快捷键: Ctrl+Shift+F

```

```

'

```

```

    Columns("L:M").Select

```

```

    Selection.NumberFormatLocal = "0.00%"

```



```
Columns("C").Select
Selection.NumberFormatLocal = "0_ "

maxrow = ActiveSheet.UsedRange.Rows.Count
Range("A2:M" & maxrow).Select
Selection.Copy
End Sub
```

```
*/
```

```
Fs("{SPACE}")
Fs("{n}")
Ddck("CGSS2017 电话核查样本更新 - Google 表格 - Google Chrome",1)
Fs("^v")
}
```

```
;
;10.更新 google 进度大表
;
```

```
Run,https://docs.google.com/spreadsheets/d/1-pBKYvO8Fd32Rmk76ghDrIcKPC4R5renbWGtE5-DYLG/edit#gid=0
```

```
Ddck("CGSS2017 - Google 表格 - Google Chrome",2)
Fs("{RIGHT}",1)
Fs("^HOME",1)
Loop 12
{
    Fs("{RIGHT}",1)
}
```

```
Run,D:\CGSS 调查问卷统计\数据质控测试文档\%now_date% 调查进度.xlsx
Ddck("",1,"调查进度.xlsx")
Fs("^+j")
```

```
/*
```

Ctrl+Shift+J 用来执行如下处理调查进度的 Excel Visual Basic 程序

```
Sub 调查进度处理()
'
' 调查进度处理 宏
'
' 快捷键: Ctrl+Shift+J
'
```

```

Columns("G:H").Select
Selection.NumberFormatLocal = "0.00%"
Range("B2:J26").Select
Selection.Copy
End Sub

```

```
*/
```

```

Fs("!{SPACE}")
Fs("{n}")
Ddck("CGSS2017 - Google 表格 - Google Chrome",3)
Fs("^v")

```

```
Return
```

```
/*
```

备份中国综合社会调查（CGSS）数据 Autohotkey 程序

关于该程序的注意事项如下：

- 1.该程序在使用之前需要针对使用该程序的计算机作调试；
- 2.使用该程序时请将系统的默认浏览器设置为 Google Chrome，并安装插件 Vimium；
- 3.使用该程序前请确保网络通畅，并开启了 Lantern 一类的软件；
- 4.由于使用未经调试的该程序而造成的一切后果由使用者自行承担；
- 5.该程序最终解释权归编写者所有，允许自由传播，但请注明出处。

作者：llygg6@gmail.com 时间：2017-07-03 —— 2017-09-30

```
*/
```

;三、备份 CGSS 数据主体程序

```
:*:bfcgsssj::
```

```

;_____
;1.获取系统当前日期，日期格式为 yyyy-MM-dd 和 yyyyMMdd
;_____

```

```

FormatTime, now_date, %A_Now%, yyyy-MM-dd
Sm()
FormatTime, now_date_wg, %A_Now%, yyyyMMdd
Sm()

```

;2.在 CGSSdata 目录下新建以系统日期命名的文件夹

;2.2 在 CGSSdata 目录下新建以系统日期命名的文件夹

Run, cmd.exe

Ddck("C:\windows\system32\cmd.exe")

Fs("md d:\CGSSdata\")

Send,%now_date_wg%

Fs("{ENTER}")

;2.3 在 CGSSdata\系统日期\ 目录下新建以访员、问卷、电核命名的三个文件夹

Fs("md d:\CGSSdata\")

Send,%now_date_wg%

Fs("\访员")

Fs("{ENTER}")

Fs("md d:\CGSSdata\")

Send,%now_date_wg%

Fs("\问卷")

Fs("{ENTER}")

Fs("md d:\CGSSdata\")

Send,%now_date_wg%

Fs("\电核")

Fs("{ENTER}")

;2.4 退出 cmd

Zjdj(643,11)

;3.打开登录页面

;3.1 打开 limesurvey 登录页面并登录 limesurvey 系统

Run,http://101.200.178.132/limesurvey_2017/index.php/admin

Ddck("LimeSurvey_2017 - Google Chrome")

Fs("{LWINDOWN}{UP}{LWINUP}")

Sr wz("longzhengfan")

Fs("{TAB}")

Fs("123456")

Fs("{ENTER}")

;3.2 判断是否登录了 limesurvey 系统

Sc(501,216,553,271,"0x0BEDCB",2)

;4.依次打开 6 个下载页面

;4.1 下载地址头部 1、头部 2

sjdz1 = http://101.200.178.132/limesurvey_2017/index.php/admin/export/sa/exportspss/sid/

sjdz2 = http://101.200.178.132/limesurvey_2017/index.php/admin/export/sa/exportresults/surveyid/

;4.2 打开电核数据下载页面 (sps,dta,csv,xlsx)

Run,%sjdz1%637954

Sm()

Run,%sjdz2%637954

Sm()

;4.3 打开访员数据下载页面 (sps,dta,csv,xlsx)

Run,%sjdz1%252672

Sm()

Run,%sjdz2%252672

Sm()

;4.4 打开问卷数据下载页面 (sps,dta,csv,xlsx)

Run,%sjdz1%963159

Sm()

Run,%sjdz2%963159

Sm(2)

;5.下载第 1,2 个文件

;5.1 切换到第 1 个下载页面

Fs("{PGDN}")

Fs("{PGDN}",2)

;5.2 下载第 1 个文件 637954.sps 到 D:\CGSSdata\系统日期\电核

Fs("f",1)

Fs("ss")

Ddck("另存为",1)

Fs("^f",0.4)

Fs("+{TAB}")

Fs("{ENTER}")

Sr wz("D:\CGSSdata\")

Send,%now_date_wg%

Sm(0.2)

```

Srww("\电核")
Fs("{ENTER}")
Loop,8
{
    Fs("{TAB}",0.5)
}
Fs("{ENTER}",2)

```

;5.3 下载第 2 个文件 637954.sps 到 D:\CGSSdata\系统日期\电核

```

Fs("f",1)
Fs("w")
Ddck("另存为",1)
Fs("{ENTER}")

```

;6.下载第 3,4 个文件

;6.1 切换到第 2 个下载页面

```

Fs("^PGDN")

```

;6.2 下载第 3 个文件 637954.csv 到 D:\CGSSdata\系统日期\电核

```

Zjdj(624,427)

```

```

Loop,19

```

```

{
    Fs("{DOWN}")
}

```

```

Fs("f",0.3)

```

```

Fs("e",0.3)

```

```

Fs("f",0.3)

```

```

Fs("k",0.3)

```

```

Fs("f",0.3)

```

```

Fs("l",0.3)

```

```

Fs("f",0.3)

```

```

Fs("m",0.3)

```

```

Fs("f",0.3)

```

```

Fs("sp",0.3)

```

```

Fs("^a",0.3)

```

```

Fs("{TAB}",0.3)

```

```

Fs("f",1)

```

```

Fs("sa",0.3)

```

```

Ddck("另存为",1)

```

```

Fs("{ENTER}",1)

```

;6.3 下载第 4 个文件 637954.xlsx 到 D:\CGSSdata\系统日期\电核

Zjdj(74,663)

Loop,19

{

Fs("{UP}")

}

Fs("f",0.3)

Fs("k",0.3)

Fs("f",0.3)

Fs("sf",0.3)

Ddck("另存为",1)

Fs("{ENTER}",0.3)

;7.下载第 5,6 个文件

;7.1 切换到第 3 个下载页面

Fs("^PGDN")

;7.2 下载第 5 个文件 252672.dat 到 D:\CGSSdata\系统日期\访员

Fs("{f}",1)

Fs("ss")

Ddck("另存为",1)

Fs("^f",0.4)

Fs("+{TAB}")

Fs("{ENTER}")

Srwz("D:\CGSSdata\")

Send,%now_date_wg%

Sm(0.2)

Srwz("\访员")

Fs("{ENTER}")

Loop,8

{

Fs("{TAB}",0.5)

}

Fs("{ENTER}",2)

;7.3 下载第 6 个文件 252672.sps 到 D:\CGSSdata\系统日期\访员

Fs("{f}",1)

Fs("w")

Ddck("另存为",1)

Fs("{ENTER}")

;8.下载第 7,8 个文件

;8.1 切换到第 4 个下载页面

Fs("{PGDN}")

;8.2 下载第 7 个文件 252672.csv 到 D:\CGSSdata\系统日期\访员

Zjdj(624,427)

Loop,19

{

Fs("{DOWN}")

}

Fs("f",0.3)

Fs("e",0.3)

Fs("f",0.3)

Fs("sk",0.3)

Fs("f",0.3)

Fs("sl",0.3)

Fs("f",0.3)

Fs("sm",0.3)

Fs("f",1)

Fs("sa",0.3)

Ddck("另存为",1)

Fs("{ENTER}",1)

;8.3 下载第 8 个文件 252672.xlsx 到 D:\CGSSdata\系统日期\访员

Zjdj(74,663)

Loop,19

{

Fs("{UP}")

}

Fs("f",0.3)

Fs("k",0.3)

Fs("f",0.3)

Fs("sf",0.3)

Ddck("另存为",1)

Fs("{ENTER}",0.3)

;9.下载第 9,10 个文件

;10.1 切换到第 5 个下载页面

```
Fs("^PGDN")
```

```
;10.2 下载第 9 个文件 963159.sps 到 D:\CGSSdata\系统日期\问卷
```

```
Fs("{f}",1)
```

```
Fs("ss")
```

```
Ddck("另存为",1)
```

```
Fs("^f",0.4)
```

```
Fs("+{TAB}")
```

```
Fs("{ENTER}")
```

```
Srwz("D:\CGSSdata\")
```

```
Send,%now_date_wg%
```

```
Sm(0.2)
```

```
Srwz("\问卷")
```

```
Fs("{ENTER}")
```

```
Loop,8
```

```
{
```

```
    Fs("{TAB}",0.5)
```

```
}
```

```
Fs("{ENTER}",2)
```

```
;9.3 下载第 10 个文件 963159.dta 到 D:\CGSSdata\系统日期\问卷
```

```
Fs("{f}",1)
```

```
Fs("w")
```

```
Ddck("另存为",1)
```

```
Fs("{ENTER}")
```

```
;
```

```
;10.下载第 11,12 个文件
```

```
;
```

```
;10.1 切换到第 2 个下载页面
```

```
Fs("^PGDN")
```

```
;10.2 下载第 11 个文件 637954.csv 到 D:\CGSSdata\系统日期\问卷
```

```
Zjdj(624,427)
```

```
Loop,19
```

```
{
```

```
    Fs("{DOWN}")
```

```
}
```

```
Fs("f",0.3)
```

```
Fs("e",0.3)
```

```
Fs("f",0.3)
```

```
Fs("k",0.3)
```

```
Fs("f",0.3)
```



```
Fs("l",0.3)
Fs("f",0.3)
Fs("m",0.3)
Fs("f",0.3)
Fs("sp",0.3)
Fs("^a",0.3)
Fs("{TAB}",0.3)
Fs("f",1)
Fs("sa",0.3)
Ddck("另存为",1)
Fs("{ENTER}",1)
```

```
;10.3 下载第 12 个文件 637954.xlsx 到 D:\CGSSdata\系统日期\问卷
Zjdj(74,663)
Loop,19
{
    Fs("{UP}")
}
Fs("f",0.3)
Fs("k",0.3)
Fs("f",0.3)
Fs("sf",0.3)
Ddck("另存为",1)
Fs("{ENTER}",0.3)
```

Return