

## 附录 第 13 次中国互联网络发展状况统计调查方法

依据统计学理论和国际惯例,在前十二次调查工作基础之上,本次调查采用了计算机网上自动搜寻、网上联机、网下抽样、相关单位上报数据等调查方法。

### 一、域名数、网站数调查

(一) 我国的通用顶级域名数及对应网站数:

通过各通用顶级域名注册单位协助提供。这些数据包括:所有通用顶级域名(gTLD)数、所有通用顶级域名(gTLD)中有网站(即有WWW服务)的域名总数、所有有网站(即有WWW服务)的通用顶级域名(gTLD)按.com、.net、.org分类的数目、所有有网站(即有WWW服务)的通用顶级域名(gTLD)按注册单位所在省份分类的数目。

(二) 我国的CN域名数及对应网站数:

采用计算机网上自动搜索可得到如下数据:CN下的域名数及地域分布情况;CN下WWW站点数及其地域分布情况。

(三) 我国域名总数、网站总数:

将以上(一)(二)两部分的相关数据分别相加,即可得到我国的域名总数、网站总数、域名和网站的地域分布、网站分类数等数据。

### 二、网上联机调查

网上联机调查重在了解网民对网络的使用情况、行为习惯以及对热点问题的看法和倾向。具体方法是将问卷放置在CNNIC的网站上,同时在全国各省的信息港与较大ICP/ISP上设置问卷链接,由互联网用户主动参与填写问卷的方式来获取信息。

CNNIC在2003年12月11日---12月31日进行了网上联机调查。调查得到了国内众多

知名网站、媒体的大力支持,国内许多知名网站均在主页为本次联机调查问卷放置了链接。本次网上联机调查共收到调查问卷 49653 份,经过有效性检查处理得到有效答卷 16760 份。

### 三、网下抽样调查

网下抽样调查侧重于了解中国网民的总量、相关的特征及行为特点等。

#### (一)、调查总体

本调查的目标总体有两个,一是全国有住宅电话的 6 岁以上的人群(总体 A),采用电话调查的方式,样本对全国有代表性;另一个总体是全国所有高等院校中的住校学生(总体 B),采用面访的方式进行调查。在对全国结果进行推断时,将两个子样本的统计量应用加权公式进行汇总。

#### (二)、总体 A 抽样方法

按照科学性和可操作性相结合的原则,我们对目标总体按省进行分层。

##### 抽样指标的确定

从全国的情况来看,各省的城市住宅电话与乡村住宅电话的比例差异很大,由于城市与农村家庭的平均人口数差异很大,所以在确定各省样本量以及用各省数据推断全国时,我们考虑的指标是“拥有住宅电话的人数(或称住宅电话覆盖的人数)”;我们采用地市的“住宅电话数目”作为抽样指标。为了得到地市“住宅电话数目”的近似估计,借助省一级的“住宅电话数目”与有关的经济、人口指标建立的回归预测模型,再利用地市一级的有关经济、人口指标的值来计算。

##### 样本量

为了保证目标比例估计值的精度,在 95%的置信度下,每省的的样本量为 1600,对各省网民人数估算的最大允许绝对误差不超过 3%。

##### 省内各地市的抽样方法

采用 PPS 抽样方法。

第一步:用 PPS 法每省抽取 7 个地市(此处的地市包括地级市和地区行署,每个地

市下都包含城镇和乡村,为不引起歧义,以下简称为地市),其中广东省和四川省由于地市较多,对其抽取8个地市进行调查。在地市多于七个的省中,各省的样本量在抽中各地市中按抽中的次数平均分配,在地市少于或等于七个的省中,各省的样本量在各地市中的分配与各地市的住宅电话成比例。

抽取地市的方法:在各省中抽取地市,根据所确定的入样指标“住宅电话的数目”,按照PPS抽样法,使每个地市被抽中的概率,等于该地市“住宅电话的数目”与该省“住宅电话的数目”之比。利用EXCEL软件产生0-1之间的均匀分布的随机数,根据随机数落在各地市对应累计百分比的范围,抽取7个地市。如果一个地市被抽到两次以上,则该地市样本量相应加倍。例如:某地市被抽中一次,样本量为229个,如果该地市被抽中两次,则样本量为457个。

第二步:获得抽中地市的所有电话局号,根据该地市的局号生成电话号码库。电话号码中除局号外的后四位或后三位数字,由随机数生成。

第三步:确定抽取调查对象,在电话拨通后,把接听电话的人作为被访对象,先询问家庭基本状况和他(她)本人上网(不上网)的有关情况、个人背景资料和家庭其它成员的最简要资料。如果他(她)不上网,但家中有人上网,则再随机抽取一名上网的成员来接听电话,回答有关上网的问题以及自己的个人基本资料。

#### 全国加权方法

对全国的推断采用对各省的调查结果进行事后加权处理的方法。

通过以上方法确定了调查对象后,对有家庭电话的住户进行电话访问,经过事后加权得出总体A;对于总体B(住校的高等学校学生),由于近年来大学生在全国人口中所占比例变化不是很大,而且大学生中网民的比例已经比较高,所以本次调查中涉及大学生的数据是在2000年底进行的大学生面访调查的基础上,结合最新的在校大学生数据建立数学模型推算得到。最后将这两部分调查结果综合加权计算以后即得到中国网民的总量、相关特征、行为特点等数据。

#### (三) 抽样调查成功率

按美国舆论研究协会(AAPOR)的成功率公式三计算,本次抽样调查的成功率为37%。

#### （四）数据预处理

在数据处理之前，对数据中变量的取值、变量之间的逻辑关系等进行检查，对其中的不合格样本进行了核对、删除和补充，并对部分变量进行了事后编码。

在统计报告中有一些平均数（比如每周上网小时数、每周上网天数、邮件帐号数、收发电子邮件数等），在计算这些平均数前，首先采用以大于或小于平均数的三个标准差和检查观测的各变量之间的逻辑关系等方法对数据中的异常值进行排除。