文章编号: 1000-6788(2003)03-0129-05

运用信息熵理论研究条件估值调查中的抽样问题

徐中民[1], 张志强[1,2], 程国栋[1], 苏志勇[3], 张海涛[3]

(1. 中国科学院寒区旱区环境与工程研究所冻土工程国家重点实验室, 甘肃 兰州 730000; 2. 中国科学院资源环境信息中心, 甘肃 兰州 730000; 3. 兰州大学干旱农业生态国家重点实验室, 甘肃 兰州 730000)

摘要: 条件估值研究是当前国际上流行且唯一的衡量环境物品非利用经济价值的方法, 其中样本的数量和问卷的设计是条件估值研究能否成功的关键因素之一. 采用信息熵理论研究了恢复额济纳生态系统条件估值的最佳抽样样本容量问题, 通过比较增加随机抽样样本容量所获得的边际信息和增发问卷的边际成本, 结果发现, 在黑河流域采用支付卡的条件估值研究方法, 发放 400 份有效问卷就能相当近似地得到实证调查中发放 646 份有效问卷的结果. 同时研究发现, 被调查者的相关社会经济信息, 不仅可以减少条件估值研究结果的不确定性, 而且与列联表相关检验结合可以用来揭示被调查者支付意愿的差异. 分析表明, 在恢复额济纳生态系统的问卷调查中, 被调查者的学历、收入、户籍和所居住的区域对支付意愿的结果有较大的影响, 而被调查者的年龄和性别对支付意愿的影响很小.

关键词: 信息熵; 抽样调查; 平均相互信息; 条件估值; 额济纳旗

中图分类号: P062.2　　**文献标识码**: A

The Application of Information Entropy on the Sample Amount and Content in Contingent Valuation Survey

XU Zhong-min[1], ZHANG Zhi-qiang[1,2], CHENG Guo-dong[1], SU Zhi-yong[3], ZHANG Hai-tao[3]

(1. State Key Laboratory of Frozen Soil Engineering, CAREERI, CAS, Lanzhou 730000, China; 2. Scientific information Center for Resources and Environment, CAS, Lanzhou 730000, China; 3. Arid Agriculture Key Laboratory, Lanzhou university, Lanzhou 730000, China)

Abstract: The contingent valuation method is a survey-based approach that directly elicits an individual's willingness to pay for environmental quality. Usually, large cost accompany with contingent valuation survey. In this paper, taking the survey of restoring Ejina banner ecological system as a case, the appropriate sample amount and survey content were discussed by adopting the theory of information entropy.

In this paper, the marginal information gained and marginal cost incurred on increasing sample amount were calculated by taking Ejina banner's survey as a case. The calculated results showed that choosing 400 sample can get information equal distributing 646 sample in our empirical survey. At the same time, the calculated average mutual information and statistical test showed that we should investigated related social economic information with interviewee. The knowledge, income, registered permanent residence and residence site of interviewee should be preferred investigated, the age and sex can be excluded from the survey on interviewee's relating information. Usually, the higher average mutual information of interviewee's social economic information, the higher the influence of social economic information on the willingness to pay.

At last, the empirical experience on sample in contingent valuation survey was summed up.

收稿日期: 2001-10-27

资助项目: 国家自然科学基金重点项目(40235053); 国家自然科学基金(40201019)

作者简介: 徐中民(1973-), 男, 湖南长沙人, 博士, 副研究员, 主要从事生态经济方面的研究

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

Key words: information entropy; sample survey; average mutual information; contingent valuation; Ejina banner

1 引言

条件估值方法(contingent valuation method, 简称, CVM)是一种利用效用最大化原理, 采用问卷调查, 通过模拟市场来揭示消费者对环境物品和服务的偏好, 并推导消费者的支付意愿(willingness to pay, 简称WTP), 从而最终得到公共物品非利用经济价值的一种研究方法, 也是当前用于解决环境物品非利用经济价值的唯一研究方法[1,2].

条件估值的具体操作是通过采访, 要求被调查对象对环境状况的变化直接标出价值或从被调查者对环境物品投标数量'是'或'否'的回答中推导环境物品的总经济价值. 通常, 一项条件估值研究需要解决 4 个主要技术问题, 即调查问卷的设计、采访的方式、提问的方式和数据的统计分析问题. 当前条件估值的研究方法已经在西方国家得到广泛的应用. 由于社会体制, 生活习惯等多种因素的影响, 条件估值的研究方法在发展中国家应用的案例还不多见. 中国的研究只有薛达元对长白山自然保护区生物多样性经济价值的评估研究[3]和本文作者对额济纳生态系统恢复的总经济价值的研究[4]. 从问卷的设计、抽样的随机性、揭露偏好的过程等条件估值的关键过程来考虑, 恢复额济纳生态系统的经济价值研究是我国关于条件估值的首例完整研究案例.

从我们的实证分析来看, 完成条件估值是一项艰巨的任务, 条件估值问卷的设计和调查将碰到各种各样的问题, 由于需要进行大量的问卷调查, 条件估值需要花费很大的成本. 因此, 如何确定抽样调查的规模就是一个值得关注的现实问题. 因此, 结合对恢复额济纳生态系统经济价值的条件估值实证分析, 本文采用信息熵方法分析了样本规模的问题, 并结合列联表分析的方法分析了抽样调查的内容问题.

2 信息熵的研究方法

不确定性是指人们对某事件结果的预计缺乏先验性的程度. 在实际生活中, 某事件结果的不确定性依赖于问题出现的形式和人们理解问题的方式, 某事件开始存在的不确定性依赖于人们开始对该事件可能出现结果的分类, 当然这些分类必须覆盖该事件所有可能出现的结果. 从信息论的角度来看, 给定结果出现的不确定性同该结果出现的可能性密切相关, 因此, 通常采用信息熵来测量事件出现 i 结果的不确定性程度:

$$H_i = \log(p_i) \tag{1}$$

而事件整体的不确定性为[3]

$$H = \sum_i p_i \log(p_i) \tag{2}$$

(1), (2)两式中 H 是某事件结果的不确定性程度, p_i 是某事件出现结果 i 的可能性. 显然某事件结果不确定性程度的上限为: $H_{\max} = \sum_i (1/n)\log(1/n)$, n 是时间可能出现结果的分组数. 由于增加观测能获得信息, 能减少原事件结果出现的不确定性程度. 在观测 m 个样本后原事件出现结果的不确定性程度如(1)式, 在增加观测 n 个样本后(总样本数为 $m+n$), 原问题的不确定性程度为: $H = \sum_i \log(p_i^*)$, 新增加观测 n 个样本后, 原事件结果出现的不确定性程度减少量为 $\sum p_i^* \log(p_i^*/p_i)$. 尽管单项的 $\log(p_i^*/p_i)$ 可能为负, 但由于对数函数的凹性, 经过各分类可能性加权后的总数肯定为非负[3]. 最差的情况是增加观测次数并不能增加任何新的信息, 也就是增加观测的结果并不改变结果的分布状况. 但是不管怎样增加样本观测的次数, 该事件结果出现仍然存在一些残余的不确定性, 这时候可考虑通过观测其它相关的事件来看是否能减少这些残余的不确定性.

知道新的事件 a 后, 原事件 b 结果出现的不确定性程度的减少量用平均相互信息(AMI)指数来度量:

$$AMI = \sum_i \sum_j p(a_j, b_i) \log[p(b_i | a_j)/p(b_i)] \tag{3}$$

式中, $p(a_j, b_i)$ 是原事件和新事件的联合分布概率; $p(b_i | a_j)$ 是原事件和新事件的条件分布概率; $p(b_i)$ 是原事件结果的分布概率.

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

由于知道 a 事件后给 b 事件所带来的信息(不确定性程度的减少)可能为负值, 用联合概率加权是为了得到一个非负数量的平均相互信息量 .

通常对同一事件进行重复观测和对相关事件进行观测能获取增加的信息, 减少事件结果出现的不确定性程度, 但增加观测次数所获得的信息在满足一定样本数量的前提下是边际递减的, 而且增加观测样本数量和观测内容通常有较大的成本, 因此在条件估值的问卷研究中有一个最适样本量和相关调查内容的问题 .

3　实例研究

3.1　黑河流域居民对恢复额济纳生态系统的肯定愿意支付意愿

对恢复额济纳生态系统经济价值问卷调查的 646 份有效问卷, 计算了各投标值肯定愿意支付样本人数的频率分布列于表 1 . 采用期望公式求得黑河流域居民对恢复额济纳生态系统的肯定愿意支付意愿为每户每年 32.18 元 .

表 1　各投标值肯定愿意支付样本人数分布频率

BID	0	2	5	10	20	35	50	75	100	200
频率	0.073	0.085	0.104	0.224	0.172	0.082	0.118	0.023	0.085	0.034

注: BID 是被调查者肯定愿意支付的投标数量(元) .

3.2　最适样本容量

探讨最适样本容量的问题, 需要比较增加样本观测所带来的边际信息量和边际成本 .为辨明最适样本量的问题, 我们分别以随机样本数 100, 200, 300, 400, 500, 600 进行了实验计算, 测算增加不同观测样本数量对条件估值的边际不确定性信息减少量列于表 2, 同时还采用经济学的方法计算了每份问卷的边际成本(见表 3) .

表 2　不同样本数量的不确定性边际减少量和试卷的边际成本

样本数	100	200	300	400	500	600	646
不确定性边际减少量 *	0.1908	0.0209	0.0192	0.0126	0.0056	0.0048	0.0025

* 单位: bit

表 3　每份试卷的边际成本

试卷份数	100	200	300	400	500	600
每枝笔	0.85	0.85	0.8	0.8	0.75	0.75
每份试卷	3	3	3	2.5	2.5	2.5
每份试卷报酬	2	2	2	2	2	2
目的地差旅费 *	206	206	206	206	206	206
往返路途车费 * *	840	840	840	840	840	840
试卷总成本	1631	2422	3198	3784	4495	5226
试卷的边际成本	16.31	7.91	7.76	5.86	7.11	7.31

* 目的地差旅费包括 3 人每天住宿费(72 元), 生活补助费(50 元), 调查交通费(80 元) .

* * 往返是指 3 人来回所需乘车费 .

由表(2)和表(3)的计算结果可知, 发放试卷总数超过 400 份时, 从每份试卷所获取的信息还不足 1% bit, 多增加的每份问卷对开始问题的不确定性的减少能力 0.003, 而超过 400 份问卷后, 每份问卷的边际成本开始上升 .在条件估值的研究中, 需要考虑以最小的成本来获取最大的信息量, 因此, 本次调查如果发

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd.　All rights reserved.

放 400 份问卷也能相当近似地得到当前的分析结果 .因此, 可以认为在黑河流域进行有关生态系统经济价值的调查以发放 400 份问卷为宜, 该数量会随问卷总成本的变动而有所变动 .

3.3 从相关的变量获得所需要的信息

在额济纳生态恢复的调查中, 在发放 646 份有效问卷后, 所得到事件结果的不确定性为 3.09bit(反映不确定性程度的单位) . 如果进一步知道投标者的年龄信息、收入信息、学历等事件后, 就能从新的相关事件中得到信息, 从而减少关于投标数量的不确定性 . 由于采用条件估值的调查方法成本很大, 如果能从新事件中获取信息来减少不确定性, 那将是一种很适用的研究方法 .

在额济纳旗生态恢复的经济价值计算中, 我们分别计算增加已知被调查者的个人信息如学历、年龄、收入、性别、户籍等事件后所获得的平均相互信息 . 并采用列联表相关检验方法分析了被调查者的社会经济信息对其肯定愿意支付数量的影响[5], 结果见表 4 .

被调查者个人的一些社会经济信息的分组情况如下:

年龄分组情况: 1. 18- 24; 2. 25- 35; 3. 35- 50; 4. 50 以上 .

学历分组情况: 1. 小学; 2. 初中; 3. 高中; 4. 大学及大学以上 .

收入分组情况: 1. 1- 5000 元; 2. 5001- 10000; 3. 10001- 20000; 4. 20001 以上 .

不同区域: 黑河干流包括张掖市、临泽县、高台县、肃南县、金塔县的鼎新区, 内蒙古额济纳旗 . 周边地区包括嘉峪关、酒泉市、金塔县、山丹县、民乐县 .

表 4 黑河流域被调查者的社会经济信息、肯定愿意支付数量与平均相互信息

分组	总体	年龄				学历				收入			
		1	2	3	4	1	2	3	4	1	2	3	4
样本数	646	126	324	174	22	37	118	188	303	223	215	169	39
MWTP	32.2	36.01	30.27	32.90	38.98	38.20	35.79	29.81	31.56	29.11	30.47	36.86	39.05
卡方		34.717				49.807				52.019			
自由度		27				27				27			
P		0.20				0.001				0.001			
AMI	3.09	0.04				0.05				0.06			

续表 4

分组	总体	户籍		性别		不同区域	
		城市	农村	男性	女性	黑河干流	周边地区
样本数	646	440	206	451	195	448	198
MWTP	32.2	31.36	33.96	32.95	30.43	35.29	25.17
卡方		42.52		12.03		23.56	
自由度		9		9		9	
P		0.001		0.20		0.005	
AMI	3.09	0.05		0.013		0.027	

从表 4 可以发现, 除年龄与性别对支付意愿的影响不显著外, 其余的因素如学历、收入、户籍和地理区域的不同均显著影响肯定愿意支付数量 这一点除从卡方检验可以看出来外, 还可以从平均相互信息上反映出来 . 进行年龄和性别分组所获得的平均相互信息相对较少, 他们对肯定愿意支付数量的影响也较少 .

文化层次与肯定支付意愿的关系不呈单调变化, 但总体表现为文化层次低的人的支付意愿要高于文化层次高的人 其中原因是因为从整体上来看, 文化层次低的人从事的工作(如农民) 与环境的关系比文化

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

层次高的人密切, 这一点也可以从农村人口与城市人口的支付意愿差别上反应出来 .

收入水平对肯定支付意愿的影响呈单调变化, 这说明黑河流域居民的支付意愿与他们的支付能力有密切的关系 .

城市居民的肯定支付意愿相对要低于农村居民, 与农民直接同生态环境打交道, 环境的恶化直接影响农民的收入等因素有关 .

黑河干流区居民的肯定支付意愿比周边地区的支付意愿要高 10 元钱左右, 这是因为额济纳生态环境问题退化对流域不同地区居民的影响程度有所不同. 尽管流域普遍都受到额济纳生态环境恶化的影响, 但影响程度存在差异, 黑河干流有一个分水问题 .

从调查的角度来看, 增加条件估值调查中关于被调查者社会经济信息的内容肯定能增加新的信息, 减少原问题的不确定性, 同时与列联表分析方法相结合可以辩明被调查者肯定愿意支付数量的差异来源 .从表 4 的计算结果来看, 通常观测对肯定支付意愿影响较大的个人社会经济因素所获得的平均相互信息也大, 也就是减少原问题不确定性程度的能力也相对较大, 如观测被调查者的学历和家庭总收入情况, 分别能减少调查问题不确定性的 1. 6%、1. 9% .对肯定支付意愿影响小的社会经济因素, 对其观测所获得的平均相互信息也小 . 由于增加调查内容会增加调查的成本, 因此, 在了解被调查者的其他相关信息时应优先选择平均相互信息大的指标 .具体到本文所涉及的恢复额济纳生态系统的经济价值调查研究来看, 在被调查者的社会经济因素中, 只需要了解被调查学历、收入、户籍和居住区域, 而被调查的年龄和性别可以不进行调查 .

4　结论

1) 本文采用信息熵理论研究了恢复额济纳生态系统总经济价值的条件估值调查中最佳抽样样本容量问题, 通过比较增加随机抽样样本容量所获得的边际信息和增加问卷的边际成本, 结果发现, 在黑河流域采用支付卡的条件估值研究方法, 发放 400 份有效问卷就能相当近似地得到文献[4]的分析结果 .

2) 在条件估值研究中需要调查被调查者的一些相关信息, 从本文采用的平均相互信息指数度量结果来看, 了解被调查者的相关社会经济信息, 不仅可以减少条件估值研究结果的不确定性, 更重要的是与列联表相关检验结合可以用来揭示被调查者支付意愿的差异 . 在恢复额济纳生态系统的经济价值评估问卷调查中, 可以只调查被调查者的学历、收入、户籍和所居住的区域, 而被调查者的年龄和性别可以不进行调查 .

参考文献:

[1] Loomis J B, Walsh R G. Recreation Economic Decisions, Comparing Benefits and Costs[M]. second edition. Penn-Sylvaniq: Venture Publishing, Inc, 1997.

[2] 薛达元. 生物多样性经济价值评估——长白山自然保护区案例研究[M]. 北京: 中国环境科学出版社. 1997.

[3] Ulanowicz R E. Growth and Development (Ecosystems Phenomenology)[M]. New York: Springer-Verlag, 1986.

[4] 徐中民, 张志强, 程国栋, 等. 额济纳生态系统恢复的总经济价值评估[J]. 地理学报, 2002, 57(1): 107- 116.

[5] 吴喜之, 王兆军. 非参数统计方法[M]. 北京: 高等教育出版社. 1996.

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.