

ExcelでできるCVM

栗山 浩一（北海道大学農学部）

1998年11月

〒060-8589

札幌市北区北9条西9丁目

北海道大学農学部森林科学科

TEL 011-706-3342

FAX 011-706-2527

E-mail: kuri@for.agr.hokudai.ac.jp

*ExcelはMicrosoft®社の製品です。

ExcelでできるCVM

栗山浩一 (北海道大学農学部)

要旨

生態系の価値を評価できる手法として仮想評価法(CVM)が注目を集めている。しかし、CVM で一般的に使われている二肢選択方式は統計的分析が必要であるため、これまでは統計専用のアプリケーションが必要であった。本研究は、初心者でもCVM を体験できるように表計算ソフト(Excel)のみで推定する方法を示すとともに、推定方法の詳細を明らかにすることで統計ソフトがどのように計算しているかを示すことを目的としている。

キーワード: CVM、二肢選択方式、ロジット、最尤推定

1. はじめに

生態系などの非利用価値を評価できる数少ない手法として、近年、仮想評価法 (contingent valuation method: CVM) が世界的に注目を集めている。CVM は環境の仮想的な変化を人々に示して、支払意志額(willingness to pay: WTP)や受入補償額(willingness to accept compensation: WTA)を回答者に直接たずねることで環境の経済価値を評価する手法である。

CVM では、表 1 に示されるように、数種類の質問形式が開発されている。初期の CVM 調査では自由回答形式や付値ゲームが用いられたが、これらの質問形式ではバイアスが生じやすいことが知られており、今日では二肢選択形式が一般的となっている。二肢選択形式とは、回答者にある金額を提示して、回答者に Yes/No のどちらかを答えてもらう方法である。Yes または No のどちらかを答えるだけなので、回答者にとっても答えやすく、またバイアスが比較的少ないことが知られている。

表 1 CVM の質問形式

自由回答形式 (open-end) いくら払うかを自由回答でたずねる。無回答が多くなるという欠点がある。
付値ゲーム形式 (bidding game) 例えば、最初に 1000 円以上払うかをたずねて、Yes ならば 2000 円、3000 円... と提示額を変化させて回答者の支払意志額に到達するまで繰り返し質問する。最初の提示額の影響を受ける可能性があること、および郵送法式では使えないという欠点がある。
支払カード形式 (payment card) (1)100 円、(2)200 円、(3)500 円、... のように金額の選択肢を示して、どれかを選択してもらう。無回答が少なくなるものの、提示した金額の幅が回答に影響する可能性がある。
二肢選択形式 (dichotomous choice) 例えば、1000 円以上払うかをたずねて、Yes/No のどちらかを 1 回だけたずねる。回答者が Yes と答える確率と提示額との関係から統計的に支払意志額を推定する。回答者が非常に答えやすく、信頼性の高い結果が得られるが、多数のサンプルを必要とする。

二肢選択形式には、一回だけ金額を提示するシングルバウンドと二回金額を提示するダブルバウンドがある。ダブルバウンドでは、最初に提示された金額に Yes と答えた場合はさらに高い金額を提示し、最初に No と答えだ場合には低い金額を提示して二回たずねる。Hanemann, Loomis, and Kanninen (1991) が示したように、シングルバウンドよりもダブルバウンドの方が統計的効率性が高く、推定結果の信頼区間が狭まることを示している。ダブルバウンドは、少ないサンプルでも効率的に評価できるという利点がある¹。

ただし、二肢選択形式の Yes/No 回答から環境価値を推定するには統計分析が必要である。分析方法には、ランダム効用モデル、支払意志額関数モデル、生存分析などがあるが、いずれも統計分析専用のアプリケーションを必要とする。シングルバウンドについては既存の統計アプリケーションで推定できるものもあるが、ダブルバウンドについては多くの場合自分でプログラミングを行う必要がある²。このため、CVM に関心を持っているものの、統計アプリケーションを持っていないために評価できないケースや、統計アプリケーションを入手しても統計分析のプログラミング作業に苦しむケースがしばしば見られる。

そこで本研究では、統計アプリケーションを持っていないが CVM を体験してみたいという人や、統計アプリケーションがどのように推定を行っているのかを知りたい人を対象に、一般的に使われている表計算ソフト(Excel)を使って二肢選択形式の CVM による推定方法を紹介する。本研究で紹介するファイル(cvm test.xls)を使えば、表計算ソフトを使える人ならば誰でも CVM による推定を体験することが可能である。

注意事項 (必ずお読みください)

なお、このファイル(cvm test.xls)は初心者が CVM を体験することや CVM の統計分析の詳細を知ることを目的としたものであり、学术论文の作成や環境評価の業務を目的に使うことは想定していない。そのような目的の場合には専用の統計アプリケーションを使うべきである。万が一、学术论文や業務を目的として本ファイル(cvm test.xls)を用いられた場合は、その評価結果に対して著者は責任を取らないので注意されたい。また、本ファイル(cvm test.xls)の著作権は著者が持っているが、個人的利用を目的とする場合は自由に使うことが可能である。

2. cvm test.xls の使い方 (シングルバウンド)

cvm test.xls は非常に簡単に使うことができる。ここでは実際の使い方について解説を行う。まず最初にシングルバウンド二肢選択形式の推定から説明する。ここではランダム効用モデルに基づいた対数線形ロジット分析による推定を行う。なお推定方法の詳細は後述する。

¹ ダブルバウンドによる評価の実際例については栗山(1997)第3章および第4章を参照。

² CVM の評価に使われる代表的な統計アプリケーションについては栗山(1997)付録Bが紹介している。

(1) ファイルを開いて「シングルバウンド」ワークシートを開く

図1 シングルバウンドのワークシート

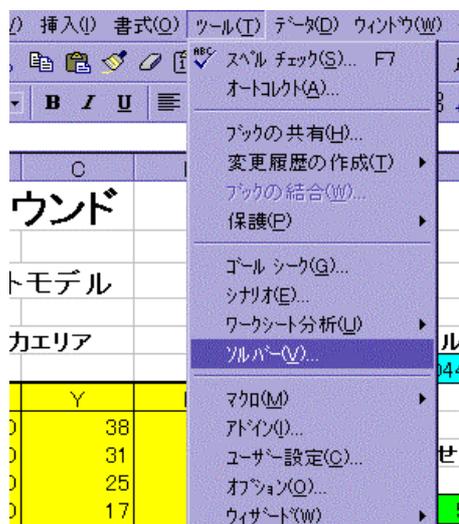
	A	B	C	D	E	F	G
1	シングルバウンド						
2							
3	対数線形ロジットモデル						
4							
5	データ入力エリア					目的セル	
6						-175.044	
7		Bid	Y	N			
8		500	38	8			
9		1000	31	12			
10		2000	25	15			
11		5000	17	23			
12		10000	17	28			
13		20000	8	36			
14							
15							
						変化させるセル	
						a	b
						5	-0.5

(2) データ入力エリアにデータを入力する

CVM のアンケート調査で得られたデータをデータ入力エリアに入力する。提示額を Bid の列に入力し、Yes 回答の数と No 回答の数をそれぞれ Y と N の列に入力する。一般的に低い提示額ほど Yes 回答が多く、提示額が高くなるにつれて No 回答が増える傾向が見られる。CVM がどのようなものを体験するだけならば、現在入力されているデータをそのまま使っても構わない。あるいは架空のデータを入力してみても構わない。

(3) メニューから「ツール/ソルバー」を選択する

図2 ソルバーの選択



ツールの中に「ソルバー」の表示がない場合は「ソルバー」が組み込まれていないので、まず組み込む必要がある。組み込み方は「ヘルプ」メニューから「ソルバーを組み込む」を検索する。

(3)目的セルと変化させるセルを選択する

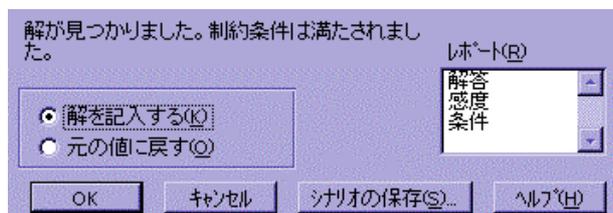
図3 目的セルと変化させるセルの選択



(4)実行ボタンを押すと推定が開始される

推定が完了すると図4が表示されるので、OK ボタンを押す。

図4 推定完了の画面



すると表2のように推定結果が表示される。表2は最初から入力されていたデータをそのまま分析したときの結果である。

表2 推定結果と推定 WTP

推定結果

変数	係数	値	p値
constant	6.300399	6.694	0.000 ***
h(B id)	-0.76547	-6.680	0.000 ***
n			
対数尤度	-151.698		

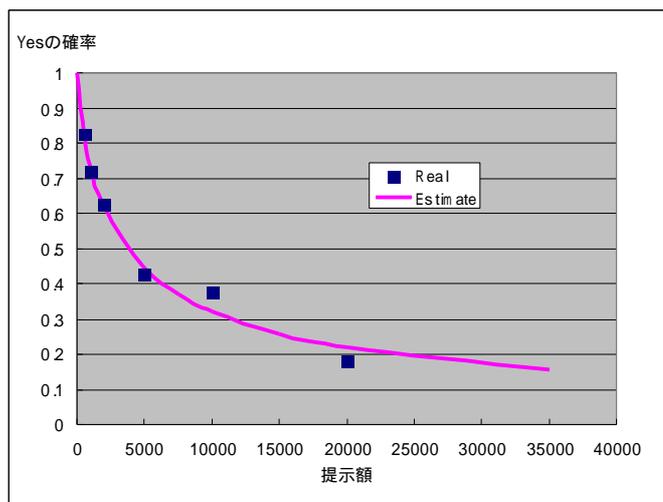
推定WTP

(中央値)	3.755
(平均値)	

変数は constant は定数項、 $h(\beta id)$ は提示額対数値を意味する。係数を見ると $h(\beta id)$ の符号はマイナスなので提示額対数値が高くなると回答者の効用が低下して Yes 回答の確率が低下することを示している。t 値が高いので、どちらの変数も1%水準で有意となっている。有意水準は p 値が示している。***は1%水準、**は5%水準、*は10%水準で有意を意味する。

推定結果は図5のグラフで表示される。横軸は提示額、縦軸はYesと回答する確率を示している。は実際の回答結果、曲線は推定結果を意味する。図5を見ると推定結果は実際の回答結果にフィットしていることがわかる。

図5 推定結果のグラフ



支払意志額の中央値はYesの確率が0.5となるところで、この場合は表2の通り3,755円である。

3. cvm test.xlsの使い方 (ダブルバウンド)

次に二回金額を提示するダブルバウンドの場合について説明する。ここではノンパラメトリック生存分析のTurnbullによる推定を行う

図6 ダブルバウンドのワークシート

	A	B	C	D	E	F	G	H
1	ダブルバウンド							
2								
3	ノンパラメトリック生存分析(Turnbull)							
4								
5	データ入力エリア							
6								
7	T1	TU	TL	YY	YN	NY	NN	
8	1000	3000	500	18	25	3	23	
9	3000	6000	1000	10	19	13	34	
10	6000	15000	3000	6	14	8	49	
11	15000	40000	6000	2	18	5	49	
12								
13								
14	目的セル		変化させるセル					
15	-401.681		dp1	0.15				
16			dp2	0.15				
17			dp3	0.15				
18			dp4	0.15				
19			dp5	0.15				
20			dp6	0.15				
21								

シングルバウンドとの使い方の違いはデータの入力部分のみである。ダブルバウンドの場合は、一回目の提示額に Yes と回答した場合は高い金額を提示し、No と回答した場合には低い金額を提示する。そこで、一回目の提示額を T1 の列、二回目の高い提示額を TU の列、二回目の低い提示額を TL の列に入力する。

回答結果は二回とも Yes の回答数を YY の列、一回目は Yes で二回目は No の回答数を YN の列、一回目は No で二回目は Yes の回答数を NY の列、そして二回とも No の回答数を NN の列に入力する。

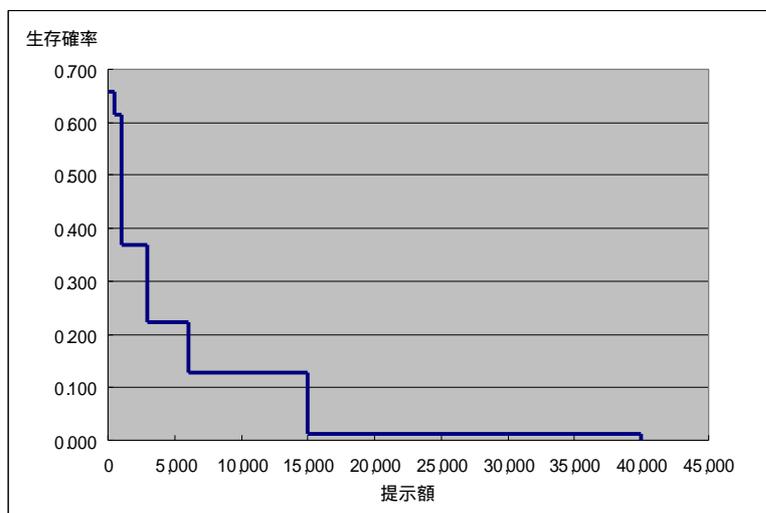
後は、シングルバウンドと同様にソルバーで目的セルと変化させるセルを選択すると推定が開始される。推定結果は表 3 のように表示される。Lower は支払意志額の下限、Upper は上限である。生存確率は支払意志額が Upper 以上である確率を意味する。例えば、支払意志額が 500 円以上の確率は 0.658 となる。金額が高くなるにつれて生存確率は低下する。

表 3 Tumbull 推定結果

推定結果

Lower	Upper	生存確率	値	P 値
0	500	0.658	16.44	0.0000 ***
500	1000	0.614	17.01	0.0000 ***
1000	3000	0.369	12.31	0.0000 ***
3000	6000	0.222	8.76	0.0000 ***
6000	15000	0.129	5.85	0.0000 ***
15000	40000	0.013	1.44	0.1497
40000	+	0.000		
n	296			
対数尤度	-344.36			

図 7 生存曲線のグラフ



またシングルバウンドのときと同様に推定結果がグラフで表示される。グラフの見方はシングルバウンドのときと同様で、横軸は提示額、縦軸は生存確率 (YES の確率) である。シングルバウンドのときと異なり Tumbull はノンパラメトリック法なので推定結果は階段状に表示される。支払意志額の中央値は生存曲線が 0.5 となるところであり表 4 のように区間で表示される。この場合は中央値は 1000 ~ 3000 円となる。平均値は下限値である。

表 4 ダブルバウンドの推定 WTP

推定 WTP

中央値	1000	~	3000
平均値	3,520		

4. 推定方法

ここでは cvm test.xls の推定方法の詳細について説明する。これ以降は統計学の知識を必要とするため、初心者の方は読み飛ばしても構わない。

4.1 シングルバウンド

シングルバウンドでは Hanemann (1984) のランダム効用モデルを用いている。仮想的な環境政策を回答者に提示し、環境政策が実施されたときには負担額が T 円だけかかるとする。一方、環境政策が実施されない場合の負担額は 0 円とする。このとき、負担額 T 円で環境政策が実施されたときの効用関数を U_Y 、負担額 0 円で政策が実施されないときの効用関数を U_N とする。効用関数は観察可能な V_Y, V_N と誤差項 $\varepsilon_Y, \varepsilon_N$ によって構成されるとする。負担額 T 円の環境政策に対して回答者が YES と答える確率は、環境政策が実施されたときの効用が実施されないときの効用より高い確率であるから、

$$\begin{aligned}
 \Pr[\text{Yes}] &= \Pr[U_Y > U_N] \\
 &= \Pr[V_Y + \mathbf{e}_Y > V_N + \mathbf{e}_N] \\
 &= \Pr[\mathbf{e} > -\Delta V]
 \end{aligned}
 \tag{1}$$

ただし、 $\Delta V = V_Y - V_N = \mathbf{b}'\mathbf{x}$ 、 $\mathbf{e} = \mathbf{e}_Y - \mathbf{e}_N$ である。 \mathbf{b} は推定されるパラメータのベクトル、 \mathbf{x} は変数ベクトルである。ここでは、効用差関数として対数線形関数 $\Delta V = \mathbf{b}_1 + \mathbf{b}_2 \ln T$ を用いている。

ここで、 ε_Y および ε_N が第一種極値分布 (Gumbel 分布) に従うと仮定するとロジスティック分布となりロジットモデルが適用できる。ロジットモデルでは、回答者が YES と答える確率は、以下の通りとなる。

$$\Pr[\text{Yes}] = \frac{1}{1 + \exp(-\Delta V)} = \Lambda \quad (2)$$

このとき対数尤度関数は

$$\ln L = \sum_i [d_Y \ln \Pr[\text{Yes}] + d_N \ln(1 - \Pr[\text{Yes}])] \quad (3)$$

となる。ただし、 d_Y は回答者が YES と答えたときに 1 となるダミー変数、 d_N は NO と答えたときに 1 となるダミー変数である。パラメータの推定は最尤法により行われる。つまり(3)式が最大となるようにパラメータ \mathbf{b} が推定される。(3)式の一階の条件は

$$\frac{\partial \ln L}{\partial \mathbf{x}} = \sum_i [d_Y - \Lambda] \mathbf{x} = \mathbf{0} \quad (4)$$

となる。したがって(4)を解くことによりパラメータの推定が行われる。また(3)式の二階の条件は

$$\mathbf{H} = \frac{\partial^2 \ln L}{\partial \mathbf{x} \partial \mathbf{x}'} = - \sum_i \Lambda(1 - \Lambda) \mathbf{x} \mathbf{x}' \quad (5)$$

(5)式より対数尤度関数は大域的に凹関数であり、最大点は唯一であることを示している。

最尤法により推定されたパラメータをもとに支払意志額を算出する。支払意志額には中央値と平均値の2種類がある。中央値は YES と答える確率が 0.5 となるときの提示額に相当する。ロジットモデルのときは(2)式より効用差 V が 0 となるときに相当する。したがって、効用差関数が対数線形の場合、支払意志額の中央値は次式により算出される。

$$\text{中央値 } WTP^* = \exp(-b_1 / b_2) \quad (6)$$

一方の平均値の場合は、 $\int_0^{\infty} \Lambda(T) dT$ によって算出される。効用差関数が対数線形の場合は、Hanemann (1984) が示したように支払意志額の平均値は次の通りとなる。

$$\text{平均値 } WTP^+ = -\exp(-b_1/b_2) \frac{p/b_2}{\sin(-p/b_2)}, \text{ ただし } 0 > \frac{1}{b_2} > -1 \quad (7)$$

4.2 ダブルバウンド

ダブルバウンドでは Turnbull (1976) のノンパラメトリック区間打ち切り生存分析を用いた。ダブルバウンドでは二回たずねているので、WTP はある区間に入ることがわかる。例えば、1000 円では Yes、3000 円では No のときは WTP が 1000 ~ 3000 円の区間の中にある。1000 円で Yes、3000 円でも Yes のときは WTP は 3000 円 ~ + の区間であり、1000 円で No、500 円でも No のときは WTP は 0 ~ 500 円の区間の中にある。

提示額 T_j 円に対して回答者が Yes と答える確率を p_j とする。この確率は生存確率と呼ばれる。WTP が $T_j \sim T_{j+1}$ の区間にあるとすると、対数尤度関数は以下の通りとなる。

$$\ln L = \sum_j N_j \ln(p_j - p_{j+1}) \quad (8)$$

となる。ただし、 N_j は WTP が $T_j \sim T_{j+1}$ の区間にある回答者の人数である。

Turnbull では生存確率が直接推定される。生存確率は(8)式が最大となるように推定される。(8)式の一階の条件は

$$\frac{\partial \ln L}{\partial p_j} = \frac{N_j}{p_j - p_{j+1}} - \frac{N_{j-1}}{p_{j-1} - p_j} = 0 \quad (9)$$

となる。生存確率 p_j は(9)式を解くことで得られる。また(8)式の一階の条件は

$$\frac{\partial^2 \ln L}{\partial p_j \partial p_j} = -\frac{N_{j-1}}{(p_{j-1} + p_j)^2} - \frac{N_j}{(p_j + p_{j+1})^2}, \quad \frac{\partial^2 \ln L}{\partial p_j \partial p_{j+1}} = \frac{N_{j-1}}{(p_j + p_{j+1})^2} \quad (10)$$

となり、対数尤度関数は大域的に凹関数であり、最大点は唯一となる。

支払意志額の算出には、シングルバウンドと同様に中央値と平均値がある。中央値の場合は、生存確率が 0.5 となるところであるが、Turnbull ノンパラメトリック法では生存確率は階段状の曲線となるため中央値は 0.5 を含む区間として推定され、点推定はできない。一方の平均値については、次式により下限値を求めることができる。

$$E[WTP]_{LOWER} = \sum_j T_{j-1}(p_j - p_{j-1}) \leq \int_0^{\infty} T dp(T) = E[WTP]$$

4.3 最尤法

最尤法は、対数尤度関数が最大となるようにパラメータの推定を行う。一般に対数尤度関数は非線型なので、最適解を直ちに求めることが困難であり、試行錯誤により最適解を求める必要がある。繰り返し回数 t 回目のときのパラメータを \mathbf{q}_t とすると、次式によりパラメータの更新を行う。

$$\mathbf{q}_{t+1} = \mathbf{q}_t + \mathbf{I}_t \Delta_t \quad (11)$$

ただし、 Δ_t はステップサイズ、 \mathbf{I}_t は方向ベクトルである。このパラメータの更新方法のアルゴリズムには、Newton 法、BHHH、DFP、BFGS など様々な方法が開発されているが、ここでは最も簡単な Newton 法について解説する。対数尤度関数の一階の条件を \mathbf{q}_t でテーラー展開すると、

$$\frac{\partial \ln L(\mathbf{q}_{t+1})}{\partial \mathbf{q}_{t+1}} = \frac{\partial \ln L(\mathbf{q}_t)}{\partial \mathbf{q}_t} + H(\mathbf{q}_t)(\mathbf{q}_{t+1} - \mathbf{q}_t) = 0 \quad (12)$$

ただし H は対数尤度関数のヘッセ行列である。

$$H(\mathbf{q}_t) = \frac{\partial^2 \ln L(\mathbf{q}_t)}{\partial \mathbf{q}_t \partial \mathbf{q}_t} \quad (13)$$

(12)式を \mathbf{q}_{t+1} について解くと次式が得られる。

$$\mathbf{q}_{t+1} = \mathbf{q}_t + (H(\mathbf{q}_t))^{-1} \frac{\partial \ln L(\mathbf{q}_t)}{\partial \mathbf{q}_t} \quad (14)$$

ロジットの場合はヘッセ行列が(5)式のように簡単な形となるので、このニュートン法により最尤法の最適解が得られる。

最尤法の推定値は漸近的に正規となる性質がある。つまり、 $\hat{\mathbf{q}}_{ML} \rightarrow N[\mathbf{q}, \{I(\mathbf{q})\}^{-1}]$ が成

立する。ただし、 N は正規分布、 $I(q) = -E[\partial^2 \ln L / \partial q \partial q']$ はフィッシャー情報行列である。これよりパラメータの推定値の標準誤差が得られ、推定値の検定が可能となる。

5. おわりに

本研究では、統計アプリケーションを用いずに、表計算ソフト(EXCEL)のみで二肢選択形式のデータを分析する方法を紹介した。二肢選択形式はバイアスが比較的少なく、今日ではもっとも標準的な質問形式となっているが、統計分析が複雑なため初心者には推定が難しいという問題があった。このため、CVM を実際に試してみたいと思っても、統計分析の段階で壁にぶつかるケースがしばしば見られる。このような場合、表計算ソフトのみで二肢選択形式を体験できることは、初心者にとって有用となるであろう。

また統計アプリケーションを用いるにしても、現段階ではシングルバウンドについては比較的簡単に推定できるものの、ダブルバウンドについては自分でプログラムを組むことが必要となる場合がほとんどである。そのため、本格的に CVM を研究するならば、統計アプリケーションをブラックボックスとして使うのではなく、CVM の推定方法について詳細を知っておくことは重要である。このような場合も、cvm test.xls のワークシートを参照することで推定方法の詳細を知ることができるであろう。

このように、本研究で紹介した cvm test.xls を用いることで初心者であっても CVM を体験できるようになる。ただし、実際に CVM で環境価値を評価するときには、統計分析以外にも注意しなければならない点が多いことを忘れてはならない。慎重にサーベイデザインを行わなければバイアスが生じる危険性があるし、二肢選択形式の場合は提示額設計が非常に重要となる。そしてサーベイデザインや提示額設計の問題点を確認するためには、プレテストを繰り返す必要がある。

本研究は、CVM の統計分析を体験して理解することに役立つであろうが、CVM は統計分析だけで結果が出るものではない。なぜなら統計分析は一連の CVM の評価プロセスの一つにすぎないからである。サーベイデザイン、提示額設計、プレテスト、統計分析、評価結果の信頼性の確認など、様々な評価プロセスのすべてが CVM の評価には不可欠であることに注意が必要である。

参考文献

二肢選択形式の推定方法については

Hanemann, W. M. (1984) Welfare Evaluations in Contingent Valuation Experiments with Discrete Response, *American Journal of Agricultural Economics*, 66(3): 332-341

ダブルバウンドの推定方法については

Hanemann, M., Loomis, J., and Kanninen, B. (1991) Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics*, 73(4): 1255-1263.

生存分析については

Nelson, W. (1982) *Applied Life Data Analysis*. New York: Wiley.

最尤法については

Greene, W. H. (1997) *Econometric Analysis*, 3rd ed. Prentice Hall.

CVM については

栗山浩一(1997)公共事業と環境の価値. 築地書館.