

拡張 δ 区間モデルによる準拠集団理論の定式化

SSP 2015 data による検証

浜田 宏
東北大学

2016/03/19

1 問題関心と背景

収入に対する満足度を説明するために考案された δ 区間モデルは、準拠集団および相対的剥奪という古典的社会理論を数学的にフォーマライズする試みだった(浜田 2001)。本稿は δ 区間モデルを 2 次元の場合に拡張して、データにより適合的な準拠集団理論モデルを定式化する。

はじめに、 δ 区間モデルの概要を確認する。基本的な考え方は以下の通りである：

人々は自分の客観的な所得分布上の近隣から準拠集団を選択し、それとの比較によって相対的な満足/不満足度を決定する

ここで近隣という概念を数学的に定義しておく。

定義 1 (δ 区間). 所得額 y の分布上の近隣を閉区間

$$[y - \delta, y + \delta] = \{z | y - \delta \leq z \leq y + \delta\} \quad \text{ただし } \delta \geq 0$$

で表し、これを y の δ 区間と呼ぶ。 δ は y から見た準拠集団の範囲を表している¹。

次にモデルを構成する基本仮定を確認する。

- 所得分布は対数正規分布に従う確率変数 Y で近似できる。確率変数 Y の確率密度関数を $f_Y(t)$ で表す²。
- y で確率変数 Y の実現値 (outcome) を表す。所得 y を中心として、 $\pm\delta$ の範囲に含まれる所得の割合は、確率密度関数 $f_Y(t)$ を $y - \delta$ から $y + \delta$ の範囲で積分した値、つまり、定積分

$$P(y - \delta \leq Y \leq y + \delta) = \int_{y - \delta}^{y + \delta} f_Y(t) dt \quad (2)$$

に等しい。

¹ δ は distance の d を表している。本稿の《 δ 区間》とは、解析学・微積分学の文脈では《 δ 近傍》と呼ばれるものである。数理社会学上 δ 区間という名称が一般的なので、ここでもそう呼ぶ。

²具体的に書けば

$$f_Y(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{t} \exp\left(-\frac{(\log t - \mu)^2}{2\sigma^2}\right) \quad (1)$$

である。ここで μ と σ は分布のパラメータで、確率変数 Y を対数変換した $X = \log Y$ の平均と分散である

- 所得 y である個人は, $[y - \delta, y + \delta]$ の範囲内から準拠集団を選択する.
- 所得 y である個人が所得 $z \in [y - \delta, y + \delta]$ である他者 (準拠集団) を選択する条件付確率密度は (1) を定積分 (2) で除して基準化した関数により定まる³. すなわち,

$$\frac{f_Y(z)}{\int_{y-\delta}^{y+\delta} f_Y(t) dt} \quad \text{ただし } z \in [y - \delta, y + \delta] \quad (3)$$

である. この関数 (3) は閉区間 $[y - \delta, y + \delta]$ で z に関して積分すれば 1 になる.

- 関数 (3) を用いた条件付き期待値は, 準拠集団の平均所得 $E[Y|y - \delta \leq Y \leq y + \delta]$ である⁴.

$$E[Y|y - \delta \leq Y \leq y + \delta] = \int_{y-\delta}^{y+\delta} z \frac{f_Y(z)}{\int_{y-\delta}^{y+\delta} f_Y(t) dt} dz$$

- 準拠集団平均所得を $y^* = E[Y|y - \delta \leq Y \leq y + \delta]$ で表す. ある個人の収入に対する満足度は, 自分の所得 y と準拠集団平均所得 y^* の関数として決まる:

$$u(y, y^*)$$

- 収入に対する満足度は, y^* が大きいほど減少する⁵.

$$\frac{\partial}{\partial y^*} u(y, y^*) < 0$$

以上の基本仮定が《 δ 区間モデル》の要諦であった (浜田 2001).

図 1 に従来の δ 区間モデルのイメージを示す.

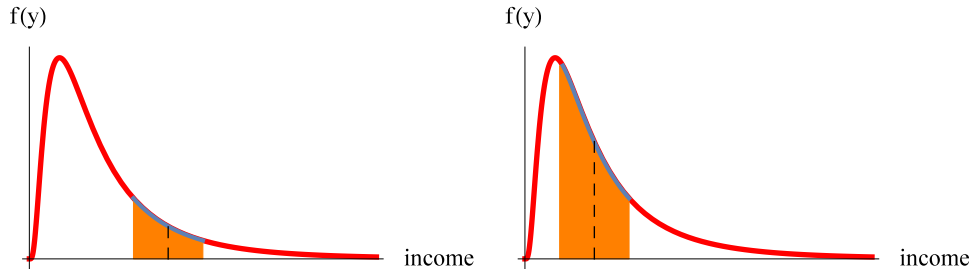


図 1: δ 区間モデルのイメージ. グラフは所得分布の確率密度関数をプロットしたもの. 破線が自己の位置を示し, グレーで色分けした部分が準拠集団所得 (の確率密度) の範囲を表している.

³連続確率変数のなので, (3) は z を選択する確率ではない. Y の実現値である 1 点 z が選ばれる確率は常に 0 である.

⁴条件付き期待値を考える場合には, Y の確率密度関数は, $Y < y - \delta, y + \delta < Y$ の範囲で 0 と考える. よって

$$\begin{aligned} E[Y|y - \delta \leq Y \leq y + \delta] &= \int_{-\infty}^{\infty} z \cdot \frac{f_Y(z)}{\int_{y-\delta}^{y+\delta} f_Y(t) dt} dz \\ &= \int_{-\infty}^{y-\delta} z \cdot 0 dz + \int_{y-\delta}^{y+\delta} z \frac{f_Y(z)}{\int_{y-\delta}^{y+\delta} f_Y(t) dt} dz + \int_{y+\delta}^{\infty} z \cdot 0 dz \\ &= \int_{y-\delta}^{y+\delta} z \frac{f_Y(z)}{\int_{y-\delta}^{y+\delta} f_Y(t) dt} dz \end{aligned}$$

⁵先行研究ではもう少し強い仮定として, $u(y, y^*) = y - y^*$ を用いた (浜田 2001)

2 多次元に拡張された《 δ 区間モデル》

準拠集団がどのような基準で選ばれるか、という問題はマートンが準拠集団理論を提唱して以来、明確な解が定まっていない経験的難問である (Merton 1957=1961). これまでに幸福感や満足度に関する実証研究では、年齢・学歴・居住地域・職業・国全体など種々の基準を用いて準拠集団を操作的に定義してきた (Mcbride 2001; Ferrer-i-Carbonell 2005; Oshio et al 2010; 石田 2011).

近年では準拠集団を回答者に直接尋ねる質問がいくつかの調査で導入されている。以下に示した各調査では、《所得を他者と比較する場合に、その他者をどのようなカテゴリから選ぶか》を回答者に選択させている。

- European Social Survey (2006: wave3) 「同僚」「家族」「友人」「それ以外」「比べなかった」
- German Socio-Economic Panel Survey(2008: pretest) 「近所の人」「友人」「同じ年齢」「同じ性別」「同じ職業」「同僚」
- Wolbring et al (2013)
- Knight et al (2009) 「近所親族」「(同じ)村に住む人」「(同じ)町区に住む人」「(同じ)地域に住む人」「(同じ)市に住む人」「中国のすべての人」「比べない」
- 飯田 (2009) 「所得水準」「性別」「年齢」「職業」「学歴・出身校」
- 飯田 (2011) 「家族」「近所の人」「友人」「同僚」「比較していない」「その他」
- 前田 (2014) 「年齢」「年収」「性別」「仕事」「学歴」「住んでいる地域」「その他」

「準拠集団が誰であったか?」「どのような基準で選択したか?」という質問に共通して多く見られた回答は「年収」「年齢」「仕事」である⁶。

直接測定の特長は、なんとといっても「回答者本人に聞いているので、研究者が基準についてアドホックな仮定を立てなくてもよい」という点である。しかしながら重視される基準が分かっても、その基準にしたがって、どういう範囲から他者が選択されるかについては、結局研究者が仮定しなくてはならない⁷。

以上のことに留意しつつ、以下に2次元に拡張した δ 区間モデルを考える。

2.1 2次元に拡張された δ 区間モデル

1次元の場合に基づき、 δ 区間モデルを次の仮定によって2次元に拡張する。

- 二次元の確率変数 (X, Y) を考え、その同時確率密度関数を $f(x, y)$ とおく。 X が年齢の分布を、 Y が所得の分布を表している。
- (x, y) で確率変数 (X, Y) のそれぞれの実現値を表す。
- 年齢が x である個人は、年齢が $[x - \delta, x + \delta]$ の範囲内にある他者を準拠集団を選択し、準拠集団と自分の《所得》を比較する。

⁶前田 (2014) が述べる通り、年収を比較するのに基準を年収と回答しているのは、質問文のワーディングからいって「年収が近い人」の意味で回答するからだと推測される

⁷前田・石田は比較の方向性バイアスを「上方」「下方」「無」で、認知バイアスを「透滅的」「離散的」「無」の各3種類区分して、計9種類のパターンを仮定してデータにフィットする比較パターンを分析から析出した (Maeada and Ishida 2011). その結果「下方かつ透滅的比較」が相対的にデータにフィットする事が判明した。

- 準拠集団の平均所得は X に対する Y の条件付き期待値, $E[Y|x - \delta \leq X \leq x + \delta]$ で与えられる.
- 収入に対する満足度は, 自分の所得 y と準拠集団平均所得 $y^* = E[Y|x - \delta \leq X \leq x + \delta]$ の関数として決まる:

$$u(y, y^*)$$

- 収入に対する満足度は, y^* が大きいほど減少する.

2次元に拡張された δ 区間モデルは, 1次元の δ 区間モデルの自然な拡張であり, 基本仮定に大きな違いはない. ただし個人の特徴を表す分布として所得だけでなく年齢も考慮しているため, 準拠集団の平均所得が単なる Y の条件付き期待値ではなく, X で条件付けられた Y の期待値となる点に注意しなくてはならない.

総和の極限として積分が定義できることを思い出せば, 連続確率変数の条件付き期待値として準拠集団平均所得を導出できる.

命題. (準拠集団所得平均) 年齢の範囲 $[x - \delta, x + \delta]$ から準拠集団を選択した場合の準拠集団平均所得は, 次の条件付き期待値

$$E[Y|x_1 \leq X \leq x_2] = \int_{-\infty}^{\infty} y \left(\int_{x_1}^{x_2} \frac{f(x, y)}{\int_{x_1}^{x_2} f_X(x) dx} dx \right) dy$$

で与えられる. ただし $f_X(x)$ は X の確率密度関数 (周辺分布) である. この準拠集団平均所得を表す条件付き期待値を以下では δ index と呼ぶ.

証明. Appendix 参照. □

2次元に拡張した δ 区間モデルのイメージを図に示す (図2).

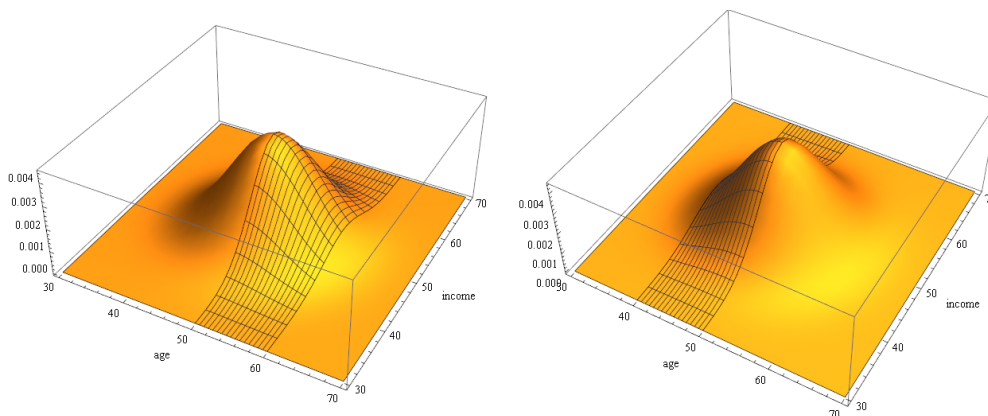


図2: 2次元 δ 区間モデルのイメージ. 曲面のグラフは所得分布と年齢分布の同時確率密度関数をプロットしたもの. メッシュで色分けした部分が準拠集団所得 (の確率密度) の範囲を表している.

モデルの拡張は2次元にとどまらず、一般に n 次元まで可能である。

例 1 (3次元 δ 区間モデル). 3次元の場合の δ 区間モデルの例を示す. 所得を表す確率変数を Y として, その他の次元 (例えば年齢と学歴) の分布を確率変数 X_1, X_2 とおく. 2次元の δ 区間を $D \subset R^2$ とおく. 例えば

$$D = [d_1, d_2] \times [d_3, d_4]$$

という二次元の閉区間は δ 区間である. $f_{X_1 X_2}(x_1, x_2)$ で (X_1, X_2) の同時確率密度関数 (周辺分布) を表し, 添え字の無い $f(x_1, x_2, y)$ で (X_1, X_2, Y) の同時確率密度関数を表す. すなわち

$$f_{X_1 X_2}(x_1, x_2) = \int_{-\infty}^{\infty} y f(x_1, x_2, y) dy$$

という関係が成立している.

このとき, 準拠集団の平均所得は

$$\begin{aligned} E[Y|d_1 \leq X_1 \leq d_2, d_3 \leq X_2 \leq d_4] \\ = \int_{-\infty}^{\infty} y \left(\int_{d_3}^{d_4} \int_{d_1}^{d_2} \frac{f(x_1, x_2, y)}{\int_{d_3}^{d_4} \int_{d_1}^{d_2} f_{X_1 X_2}(x_1, x_2) dx_1 dx_2} dx_1 dx_2 \right) dy \end{aligned}$$

である.

以上の考え方を n 次元にまで拡張した場合に, n 次元 δ 区間モデルが定義できる. ただし先行研究やデータのサンプルサイズの制約を考えれば, 3次元程度が経験的な上限ではないかと推測される.

なお, δ 区間に収入次元そのものを加えるべきではないか, という考え方もあるだろう. 数学的にはそう定義することも可能であり, その場合は

$$D = [d_1, d_2] \times [d_3, d_4] \times [d_5, d_6] \subset R^3$$

として

$$\begin{aligned} E[Y|(x_1, x_2, y) \in D] &= E[Y|d_1 \leq X_1 \leq d_2, d_3 \leq X_2 \leq d_4, d_5 \leq Y \leq d_6] \\ &= \int_{-\infty}^{\infty} y \left(\iiint_D \frac{f(x_1, x_2, y)}{\iint\iint_D f(x_1, x_2, y) dx_1 dx_2 dy} dx_1 dx_2 \right) dy \end{aligned}$$

となる.

3 実証モデル

2次元 δ 区間モデルの妥当性を実証データによって確認する.

- 使用データセット: SSP2015年 (年齢層 25-59, 有職者)
- 説明変数: 準拠集団平均所得 (δ index)

- 統制変数: 年齢, 性別, 学歴, 個人収入, 世帯収入, 婚姻状態, 子供有無, 同居家族人数, EGP6 分類
- 応答変数: 同世代の年収予想 Q30(1), 主観的幸福感 Q

従来の先行研究では, 理論部分がフォーマライズされていないために, 次のような GLM を仮定していた (φ はリンク関数).

$$\varphi(\text{準拠集団収入}) = \beta_0 + \beta_1 \text{個人年収} + \beta_2 \text{年齢} + \beta_3 \text{性別} + \varepsilon$$

しかし上記のようなモデル (関数型) は理論的には不正確である.

準拠集団平均年収は, ある個体にとっての準拠集団という集団レベルの特性である. 個人 i が年齢集団 j に所属している場合の準拠集団平均収入予想値 Y_{ij} は, 個人収入 i_j と集団レベルの特性値である z_j の関数であると考えられる.

したがって, 次のような集団レベルの誤差を考慮した線形混合モデル (以下 LME)

$$Y_{ij} = \gamma_{00} + \underbrace{\gamma_{01} \text{個人収入}_{ij}}_{\text{fixed effect}} + \underbrace{\gamma_{10} z_j}_{\text{fixed effect}} + \underbrace{\beta X}_{\text{統制変数}} + \underbrace{U_j}_{\substack{\text{random effect} \\ \text{集団レベル}}} + \underbrace{R_{ij}}_{\text{個人レベル誤差}}$$

で表すことができるだろう.

ただし, われわれは前節で準拠集団平均所得が本人年齢に基づく準拠集団選択プロセスを経て, どのように決まるのかというメカニズムを数学的に定式化しておいた. 具体的には,

$$z_j = E[Y|x_1 \leq X \leq x_2] = \int_{-\infty}^{\infty} y \left(\int_{x_1}^{x_2} \frac{f(x, y)}{\int_{x_1}^{x_2} f_X(x) dx} dx \right) dy$$

である. ここで確率変数 X が個人の年齢を表している.

つまり説明変数を拡張された δ 区間モデルに基づいて計算すれば, グループレベルの固定効果として明示的に表現できる. よって,

$$Y_{ij} = \gamma_{00} + \underbrace{\gamma_{01} \text{個人収入}_{ij}}_{\text{fixed effect}} + \underbrace{\gamma_{10} E[Y|x - \delta \leq X \leq x + \delta]_j}_{\text{fixed effect}} + \underbrace{\beta X}_{\text{統制変数}} + \underbrace{U_j}_{\substack{\text{random effect} \\ \text{集団レベル}}} + \underbrace{R_{ij}}_{\text{個人レベル誤差}}$$

が δ 区間モデルを検証するための統計モデルとなる. ここで注意すべきは, グループ変数である $E[Y|x - \delta \leq X \leq x + \delta]$ は通常の意味でのグループ平均ではないということである. 単に年齢 j である集団の平均収入なら, グループ変数は $E[Y|X = j]$ である.

しかしわれわれは準拠集団理論 (の数学的表現である δ 区間モデル) に依拠して統計モデルを選択しているので, 平均の計算は命題で明示した条件付き期待値でなくてはならない.

3.1 データによる検証

上で述べた統計モデルを表現するために, 次のような手順で説明変数 (グループレベルの固定効果) をデータから作成する. この手順で作成した説明変数を以下 δ index と呼ぶ.

- 本人年齢 x を基準に年齢が $x \pm \delta$ である人々の収入をデータから取り出す．これが準拠集団の所得分布となる．
- 年齢条件で取り出した準拠集団の所得分布から平均値を計算する．この値が準拠集団平均所得の理論値である
- 準拠集団平均所得の理論値を全ての年齢のレンジで計算する．
- 計算した準拠集団平均所得の理論値をデータ行列に代入する．このとき同じ年齢のユニットは同じ理論値を持つ．
- つまり，準拠集団平均所得理論値は年齢別集団のグループレベルの固定効果である．この固定効果によって表現できないグループレベルの誤差をランダム効果で表す（個人レベルの誤差は個人レベルの統制変数・誤差項で表す）
- 応答変数（回答者が述べた準拠集団平均所得）を理論値（固定効果）に LME で回帰する．統制変数は全て個人レベルで加える．

LME を使う理由は，データ構造の性質というより，理論的に予測した準拠集団平均収入の誤差を準拠集団のレベルで生じた誤差と，個人レベルで生じた誤差に区分して，理論値（固定効果）のパラメータ推定誤差をなるべく小さくするためである⁸．

3.2 記述統計

以下に年齢別の準拠集団所得分布と，その平均である δ index を比較する (図 3-4) ．

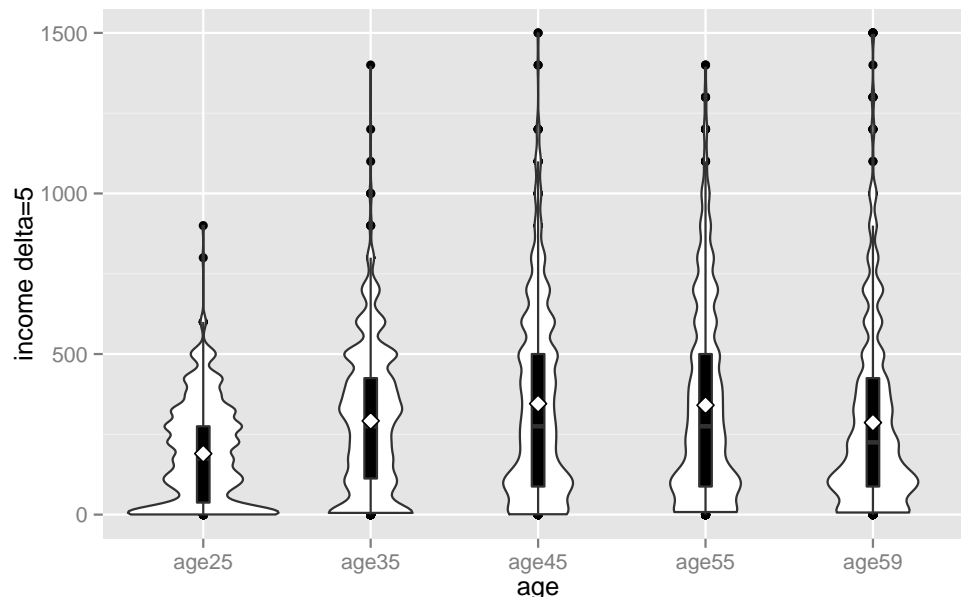


図 3: 年齢 (横軸) と準拠集団所得分布 (縦軸). 本人年齢 ± 5 ($\delta = 5$) 才の所得分布カーネル密度関数と箱ひげ図をプロット. 白い四角は平均 (δ index).

⁸この変数を作成するための R コードは Appendix にて示す

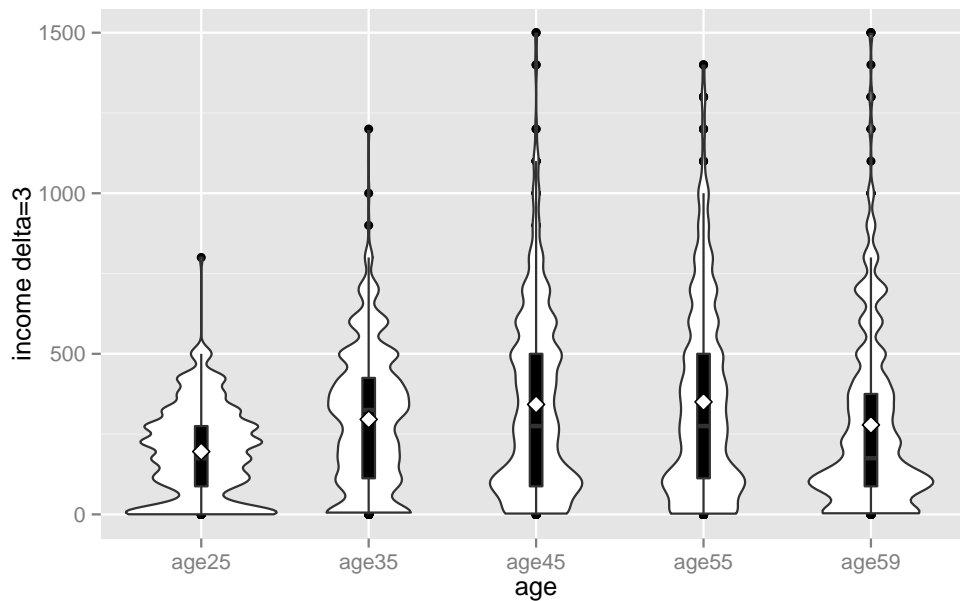


図 4: 年齢 (横軸) と準拠集団所得分布 (縦軸). 本人年齢 $\pm 3(\delta = 3)$ 才の所得分布カーネル密度関数と箱ひげ図をプロット. 白い四角は平均 (δ index).

図 3-4 は、横軸の本人年齢に対応した $\pm\delta$ 才の他者の所得分布を表示している。図からは、 δ の値は共通でも、基準年齢が異なれば、対応する所得分布の形状は異なる様子うかがえる。このデータは、年齢に応じて各個人の主観的世界が異なる、という事実を表現している。図内の白い点は、各所得分布の平均、すなわち δ index である。単純に増加するのではなく、50 歳前後をピークに減少する点に注意する。このことは単純に年齢の一次関数として他者所得予想が表現できないことを示唆している。

次に、応答変数である「同世代の人の年収 $Q30(1)$ 」と δ index の関係を散布図で示す (図 5)。

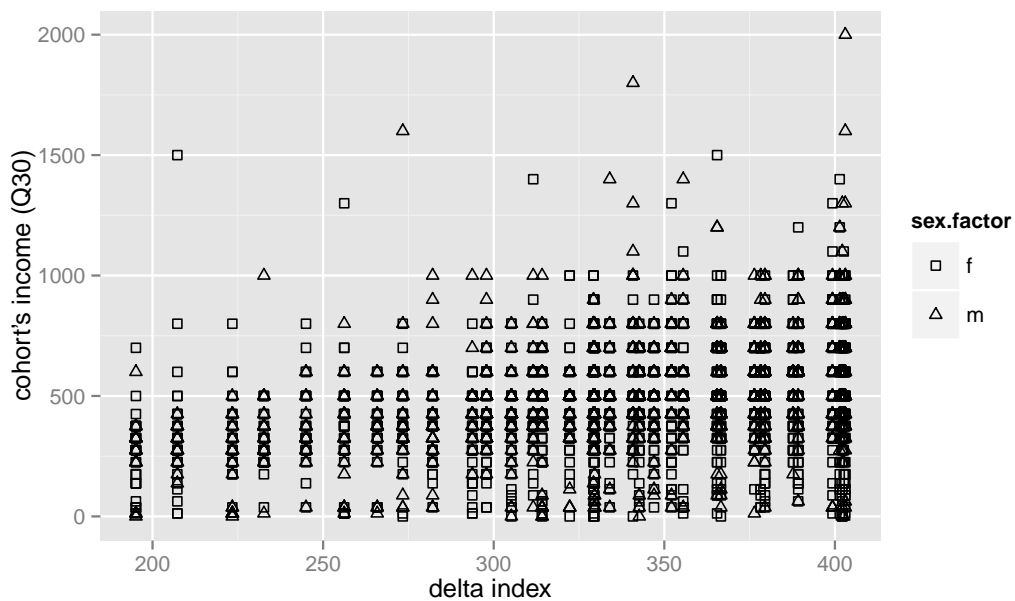


図 5: δ index(横軸) と同世代の人の年収予想 (縦軸) の関係. 縦軸は「Q30(1). あなたと同年代の人の平均的な世帯年収は, いくらぐらいだと思いますか」という質問への回答を金額にリコードした値. 点は \square が女性を, \triangle が男性を表す.

図 5 から直感的に分かるように, 性別を考慮してもなお, 同世代平均収入は, δ index だけでは表現できないばらつきを持っている.

それゆえ δ index 以外の統制変数として, 準拠集団平均収入に影響を及ぼす各条件を LME の右辺に加える. 以下データに使用した変数の記述統計量を示す.

表 1: 記述統計

連続型変数					
	mean	sd	median	min	max
グループレベル					
delta index	335.75	54.12	342.65	195.13	403.04
個人レベル					
同世代所得	445.93	269.13	425	0	4500
年齢	45.27	12.37	46	20	64
個人収入	318.42	349.24	225	0	4500
世帯収入	652.62	464.42	600	0	4500
同居家族人数	3.40	1.42	3	1	10
主観的幸福感	6.78	1.94	7	0	10
離散型変数					
		回答数	%		
性別	男性	1644	0.46		
	女性	1931	0.54		
婚姻状態	結婚	2481	0.69		
	その他	1078	0.30		
	NA	16	0.00		
教育水準	中学校卒業	202	0.06		
	高校卒業	1798	0.50		
	大学卒業	1543	0.43		
	NA	32	0.01		
職業 EGP6	上層ホワイト	878	0.25		
	下層ホワイト	712	0.20		
	自営	243	0.07		
	熟練ブルー	373	0.10		
	非熟練ブルー	533	0.15		
	農業	73	0.02		
	無職	563	0.16		
	学生	88	0.02		
	NA	112	0.03		

4 分析結果

データから推定した準拠集団平均所得理論値 (δ index) を説明変数として含む計量モデルの分析結果を示す。まず GLM ではなく、ランダム効果を仮定した線形混合モデルを用いた方がデータをより予測できることを切片モデル (集団分割だけを仮定した ANOVA) の ICC と、固定効果を含む LME0, LME1 の ICC を比較する (表 2)。ここで LME1 は固定効果として δ index を含むモデルで、LME0 は含まないモデルである。LME0 と LME1 の違いはそれのみで、その他の条件は同じである。

表 2: ランダム効果の分析結果比較. 応答変数は「同年代の年収予想 Q30(1)」

	ANOVA	LME0	LME1
固定効果	Estimate	Estimate	Estimate
切片	433.32 (14.85)	324.85 (26.45)	-8.79 (42.66)
ランダム効果	分散	分散	分散
age (切片)	9005.00 (94.89)	4125.00 (64.23)	29.65 (5.45)
residual	64770.00 (254.50)	48578.00 (220.4)	46236.16 (215.03)
ICC	0.27	0.08	0.00
deviance	48307.39	36157.26	28695.13
AIC	48313.39	36191.26	28731.13

表 2 が示すとおり、切片モデルの ICC は約 0.27 であり、集団間分散は無視できない程度に高い。次に LME の固定効果の詳細を表 3 に示す。

なお統制変数として説明変数に加えた教育水準と職業カテゴリは省略した。それは以下の理由に基づく。準拠集団に選択に教育水準や職業カテゴリが影響を及ぼすことは容易に想像できるし、先行研究によって、その影響は部分的には確かめられている。しかし理論的なメカニズム (δ 区間モデル) を考慮した場合、ダミー変数をただ LME に含めたとしても、理論を表現したことにはならないので、理論検証には役立たない。よってここでは観測できない個人レベルの誤差を統制するという以上の意味はないため、係数および有意性検定の表示を省略する。

表 3: 線形混合モデルによる分析結果

	応答変数:同世代の年収 Q30(1)		
	GLM	LME0	LME1
	(1)	(2)	(3)
δ index	1.070*** (0.092)		1.071*** (0.093)
income	0.034 (0.023)	0.080*** (0.020)	0.035 (0.023)
f.income	0.125*** (0.015)	0.102*** (0.013)	0.125*** (0.015)
male	28.531** (11.662)	8.816 (10.357)	28.450** (11.662)
marriage	37.249*** (11.981)	39.198*** (11.162)	37.046*** (11.988)
family n	-8.091** (3.691)	-3.461 (3.386)	-8.034** (3.692)
Constant	-8.135 (42.317)	324.855*** (26.448)	-8.776 (42.655)
Observations	2,120	2,654	2,120
Log Likelihood	-14,387.720	-18,078.570	-14,347.510
Akaike Inf. Crit.	28,807.430	36,191.140	28,731.030
Bayesian Inf. Crit.		36,291.160	28,832.890

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

分析の結果, δ index を含む LME1 は含まないモデル LME0 よりも AIC が低いことがわかる. このことから δ index は, 理論が予測する通り, 同世代の収入予想を適切に表現していることを示唆している.

次に LME0 と LME1 の逸脱度に関する尤度比検定の結果を示す.

表 4: 尤度比検定.LME0: without δ index; LME1: with δ index

	Df	AIC	BIC	対数尤度	逸脱度	Chisq	有意確率
LME0	17	36268.6	36369	-18117	36234.6		
LME1	18	28809.4	28911	-14387	28773.4	7461.18	0.000

尤度比検定の結果からも, 逸脱度が有意に減少していることが分かる.

4.1 説明変数としての δ index

δ index は説明変数として用いた場合にも、主観的幸福感・世帯収入満足度・個人収入満足度の予測に役立つことが予想される (Eibner and Evans 2005; 石田 2011) . 以下に応答変数が主観的幸福感・世帯収入満足度・個人収入満足度である場合の GLM (誤差分布は正規分布) で、 δ index を含むモデルと含まないモデルを比較する .

はじめに変数間の相関を確認しておく . 個人収入と世帯収入の相関はやや高いが、分析に支障をきたす程度ではないので、統制変数として両方を用いる .

表 5: 変数間の相関係数行列

Variable	1	2	3	δ index	5	6
同世代所得	1.00					
個人収入満足	0.05	1.00				
世帯収入満足	0.05	0.61	1.00			
δ index	0.26	0.03	-0.02	1.00		
個人所得	0.27	0.35	0.19	0.18	1.00	
世帯所得	0.32	0.32	0.42	0.20	0.59	1.00

われわれの理論モデルが正しければ、 δ index を含むモデルの AIC は含まないモデルよりも低下するはずである .

次に応答変数を主観的幸福感、世帯収入満足度、個人収入満足度として、 δ index を説明変数に含む条件と含まない条件を比較する . 用いた統計モデルは誤差関数に正規分布を仮定した恒等リンク関数の GLM である . 実質的には OLS 推定と同じ結果だが、AIC を比較するために最尤推定で対数尤度を計算した .

なお、前節の分析では線形混合モデルを用いたが、応答変数が主観的幸福感、世帯収入満足度、個人収入満足度である場合はいずれの場合も年齢集団をクラスターとした ICC が低いため (全て 5% 以下) , 集団レベルのランダム効果は仮定しない .

表 6: 主観的幸福感・世帯収入満足度・個人収入満足度を応答変数にした分析

	Outcomes:								
	SWB			sa.income			sa.famincome		
	(1)	(2)	(3)	(4)	(5)	(6)	(4)	(5)	(6)
<i>delta index</i>									
income	0.0001 (0.0002)	-0.0006*** (0.001)	0.001*** (0.0001)	-0.001*** (0.0003)	-0.0001* (0.0001)	-0.002*** (0.0003)	0.001*** (0.0001)	-0.0001* (0.0001)	-0.0002*** (0.0001)
f.income	0.0003*** (0.0001)	0.0005*** (0.0001)	0.0002*** (0.00005)	0.0002*** (0.0001)	0.001*** (0.00004)	0.001*** (0.00005)	0.0002*** (0.0001)	0.001*** (0.00004)	0.001*** (0.00005)
male	-0.479*** (0.085)	-0.422*** (0.097)	-0.144*** (0.036)	-0.151*** (0.040)	-0.109*** (0.033)	-0.052 (0.037)			
marriage	0.981*** (0.089)	1.115*** (0.100)	0.172*** (0.037)	0.256*** (0.041)	0.138*** (0.035)	0.172*** (0.038)			
family n	0.002 (0.027)	-0.008 (0.031)	-0.036*** (0.012)	-0.035*** (0.013)	-0.026** (0.011)	-0.033*** (0.012)			
Constant	5.511*** (0.204)	7.123*** (0.367)	1.740*** (0.094)	2.273*** (0.170)	2.044*** (0.079)	2.564*** (0.147)			
Observations	2,685	2,137	2,394	1,904	2,681	2,132			
Log Likelihood	-5,458.133	-4,316.410	-2,712.665	-2,105.995	-2,944.631	-2,247.365			
AIC.	10,950.270	8,668.819	5,459.330	4,247.991	5,923.262	4,530.731			
BIC.	11,050.490	8,770.828	5,557.602	4,347.921	6,023.459	4,632.698			

Note: * p<0.1; ** p<0.05; *** p<0.01

分析の結果、われわれの予想が正しいことが分かった。また尤度比検定の結果、モデルの逸脱度は有意に減少した。

5 結論

本稿では浜田 (2001) が提唱した δ 区間モデルを 2 次元に拡張して、年齢という基準に基づき、準拠集団を選択した場合の、他者の平均収入を予測するモデルを定式化した。そしてこの理論モデルが経験的に正しいかどうかを全国調査データ (SSP2015) を用いて検証した。

結果をまとめておく。

- 新規測定した「同世代の年収」は、拡張した δ 区間モデルの検証に役立つ。
- 拡張した δ 区間モデルに基づき、年齢と個人収入から δ index を簡単に計算でき、説明変数として使用できる。
- δ index は相対的剥奪感と主観的幸福感を説明する変数として、有望である。

δ index はデータに応じて自由に拡張できる。これにより準拠集団理論の定式化と実証がさらに進展することが期待できる。

References

- Bartoszynski, Robert, and Magdalena, Niewiadomska-bugaj, 2008, *Probability and Statistical Inference, Second Edition*, A John Wiley & Sons.
- Bossert, W. and D'Ambrosio, C., 2006, "Reference groups and individual deprivation," *Economics Letters*, 90(3): 421-426.
- Clark, E. Andrew, and Andrew J. Oswald., 1996, "Satisfaction and Comparison Income," *Journal of Public Economics*, 61: 359-81.
- Clark, A. E., Frijters, P., and Shields, M. A., 2008, "Relative income, happiness, and utility: an explanation for the Easterlin paradox and other puzzles," *Journal of Economic Literature*, 46(1): 95-144.
- D'Ambrosio, C. and Frick, J. R., 2007, "Income satisfaction and relative deprivation: an empirical link," *Social Indicators Research*, 81(3): 497-519. Ferrer-i-Carbonell, Ada., 2005, "Income and Well-being: an Empirical Analysis of the Comparison Income Effect," *Journal of Public Economics*, 89(5-6): 997-1019.
- 浜田宏, 2001, 「経済的地位の自己評価と準拠集団 区間モデルによる定式化」『社会学評論』206号 (第52巻・第2号): 283-99.
- 石田淳, 2011, 「相対的剥奪と準拠集団の計量モデル」『理論と方法』26(2): 371-88.
- Kondo, N., Kawachi, I., Hirai, H., Kondo, K., Subramanian, S. V., Hanibuchi, T., and Yamagata, Z., 2009, "Relative deprivation and incident functional disability among older Japanese women and men prospective cohort study," *Journal of Epidemiology and Community Health*, 63: 461-467.

- 前田豊, 2014, 「比較対象選択と所得イメージ」『理論と方法』29(1): 37-58.
- Maeda, Yutaka, and Atsushi, Ishida, 2013, "Income Comparison as a Determining Mechanism of Class Identification: A Quantitative and Simulation Study Using Japanese Survey Data," *International Journal of Japanese Sociology*, 22: 143-59.
- McBride, Michael., 2001, "Relative-income Effects on Subjective Well-being in the Cross-section," *Journal of Economic Behavior & Organization*, 45: 251-278.
- Merton, R. K., 1957, *Social Theory and Social Structure* (revised and enlarged edition). New York: The Free Press.
- Theil, Henri, 1978, *Introduction to Econometrics*, Prentice-Hall.
- Yitzhaki, S, 1979, "Relative deprivation and the Gini coefficient," *Quarterly Journal of Economics*, 93(2): 321-324.

Appendix

条件付き期待値

まず単純な離散的確率変数の具体例で条件付き期待値の計算過程を示す。

例 2 (離散同時確率分布の条件付き期待値). 離散的確率変数の組 (X, Y) の確率分布が次の表で与えられていると仮定する

		Y		行和
		1	2	
X	1	0.05	0.05	0.1
	2	0.2	0.1	0.3
	3	0.3	0.2	0.5
	4	0.05	0.05	0.1
列和		0.6	0.4	

表 7: (X, Y) の確率分布

(X, Y) の同時確率分布をもとにして, X に対する Y の条件付き期待値

$$E[Y|2 \leq X \leq 3]$$

を計算する. X の周辺分布から

$$P(2 \leq X \leq 3) = P(X = 2) + P(X = 3) = 0.3 + 0.5 = 0.8$$

だから, $P(2 \leq X \leq 3) = 0.8$ で基準化した同時確率密度を使って Y の期待値を計算すればよい. 具体的な数値から徐々に一般化する仮定を示す.

$$\begin{aligned}
E[Y|2 \leq X \leq 3] &= 1 \cdot \frac{0.2}{0.8} + 1 \cdot \frac{0.3}{0.8} + 2 \cdot \frac{0.1}{0.8} + 2 \cdot \frac{0.2}{0.8} \\
&= 1 \cdot \frac{P(X=2, Y=1)}{0.8} + 1 \cdot \frac{P(X=3, Y=1)}{0.8} \\
&\quad + 2 \cdot \frac{P(X=2, Y=2)}{0.8} + 2 \cdot \frac{P(X=3, Y=2)}{0.8} \\
&= \sum_{y=1}^2 y \left(\frac{P(X=2, Y=y)}{0.8} + \frac{P(X=3, Y=y)}{0.8} \right) \\
&= \sum_{y=1}^2 y \left(\sum_{x=2}^3 \frac{P(X=x, Y=y)}{0.8} \right)
\end{aligned}$$

以上の例から，離散的確率変数の同時分布の条件付き期待値は，一般に

$$E[Y|x_1 \leq X \leq x_2] = \sum_{y \in Y} y \left(\frac{\sum_{x=x_1}^{x_2} P(X=x, Y=y)}{\sum_{x=x_1}^{x_2} P(X=x)} \right)$$

であることが分かる．

定義 2 (連続確率変数の条件付き期待値)．周辺分布としての Y の確率密度を $f_Y(y)$ とおく．

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

また， $Y = y$ に対する X の条件付き確率密度 $f_{X|Y}(x|y)$ を

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

とおく． $Y = y$ に対する X の条件付き平均 $E(X|Y = y)$ は

$$E(X|Y = y) = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y) dx = \int_{-\infty}^{\infty} x \cdot \frac{f(x, y)}{f_Y(y)} dx.$$

データ上の準拠集団理論値

準拠集団理論値をデータ行列から計算する R コードは以下の通りである．

```
#準拠集団選択用データフレームの作成, ssp$age:年齢; ssp$q31_1R:個人年収;
data1<-data.frame(age=ssp$age,x=ssp$q31_1R)
```

#年齢別準拠集団 (+-delta 才) の収入をベクトル化する関数の定義

```
neighbor<-function(i,delta){
  rg=subset(data1, (age>=i-delta) & (age<=i+delta))
  rg$x }
```

```
#関数 neighbor( ) を使い, 列 rgmean に準拠集団平均収入を外挿する
rgmean<-replicate(3575,NA)#3575 はサンプルサイズ
for(i in 25:59){
  rgmean[data1$age == i] <- mean(neighbor(i,5),na.rm=TRUE)
}
```