

ベトナム家計調査における地域ごとの 働き方の比較とモデリング

Comparison and modeling of working styles by region in
Vietnam Household Living Standard Survey

久保田 貴文*
Takafumi KUBOTA

Keywords : ベトナム、家計調査、空間、自己相関

1. はじめに

本研究の背景としては、統計数理研究所共同利用「アジア諸国世帯統計マイクロデータの二次利用推進に関する研究」によってデータベースが利用可能になったことで、アジアにおけるメンタルヘルスに関連する要因を探ることから取り掛かりを始めた。その中で、ベトナムにおける地域ごとの働き方を比較することを目的とした。

研究の方法としては、ベトナムの家計調査のデータを用いて、デモグラフィックな変数および教育関係、医療関係および働き方に関する変数を抽出し、それらを視覚化するとともに回帰モデルにより分析を進めた。

まず、データについてプログラミング言語 R を用いてクリーニングを行い、さらに R におけるいくつかのパッケージを元にグラフ等を用いて視覚化した。また、インタラクティブに操作を可能とし、かつ一画面に 1 つの結果を表示するために R のパッケージの shinydashboard を用いて性別、地域（プロビンス）ごとに集計されたデータを数値・一覧表・コロプレスマップで視覚化した。

次に、まとめられたデータを元に、線形回帰分析を行い、その結果を考察した。空間自己相関を確かめるために、グローバル Moran 統計量を用いて全体的な空間自己相関の強さを求めるとともに、ローカル Moran 統計量により空間パターンの把握とホットスポットの検出を検討した。最終的に、線形回帰モデルと空間自己回帰モデルにより働き方のモデルを比較した。

2. データについて

本研究では、ベトナムの生活水準調査の匿名標本データである Resampled Micro Data of Vietnamese Household Living Standard Survey (VHLSS) の 2006 年の調査結果 (VHLSS2006)

* 多摩大学経営情報学部 School of Management and Information Sciences, Tama University

を用いて視覚化および解析を行った。

VHLSS2006 のデータのうち、性別 (sex)、年齢 (age)、学歴 (egrade)、続柄 (家長) (relation.h)、1 年以内に怪我もしくは病気 (msuffer.r)、被雇用者率 (rate.work.s) および地域：プロビンス (province) を抽出し (以降、カッコ内の変数名で記述する)、プロビンスごとに集計した。なお、視覚化における地図データについては、Hijmans(2016) より GADM を使用した。

3. 視覚化

VHLSS2006 を元に、R のパッケージ shiny および shinydashboard を用いて視覚化を行った。図 1 では久保田 (2017) により提案された視覚化の方法をさらに拡充させ、性別、続柄、生まれた月などデモグラフィックな変数について、男女別に分布の視覚化を行った。また、図 2 ではプロビンス、Region ごとの続柄の視覚化を行った。



図 1 デモグラフィックな変数の分布 (棒グラフ、ヒストグラム)

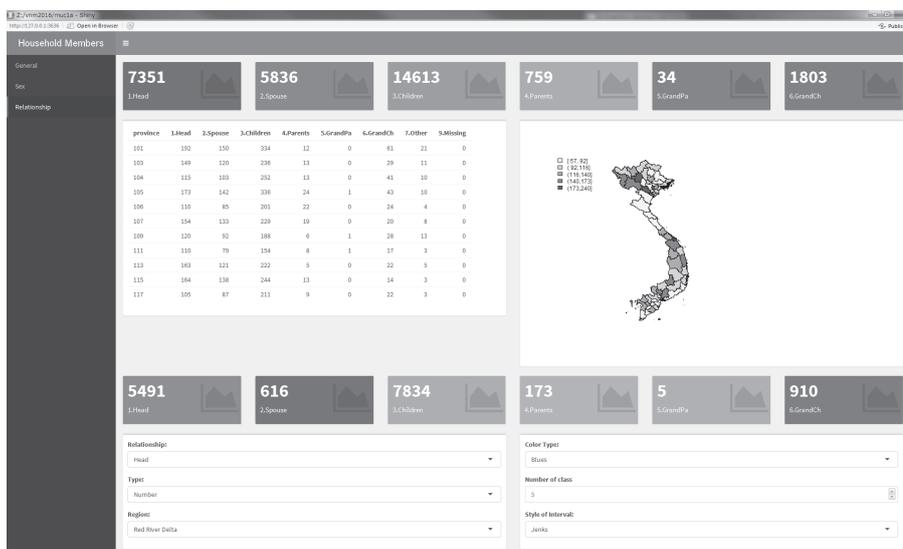


図 2 プロビンス、Region ごとの続柄の視覚化 (表・コロプレスマップ)

4. 空間自己相関と回帰モデルによる解析

地域ごとの働き方を比較するために、給与を貰っている被雇用者の割合を地域ごとすなわち、プロビンスおよび Region ごとにまとめた。図3には、被雇用者の割合をプロビンスごとにコロプレスマップ（塗りわけ地図）によって描画して視覚化した。分割数5で、Jenksの自然なブレイクによって男性および女性のそれぞれ塗りわけした。また、表1、表2にはそれぞれ男性および女性の Region ごとの被雇用者の度数と割合をまとめた。

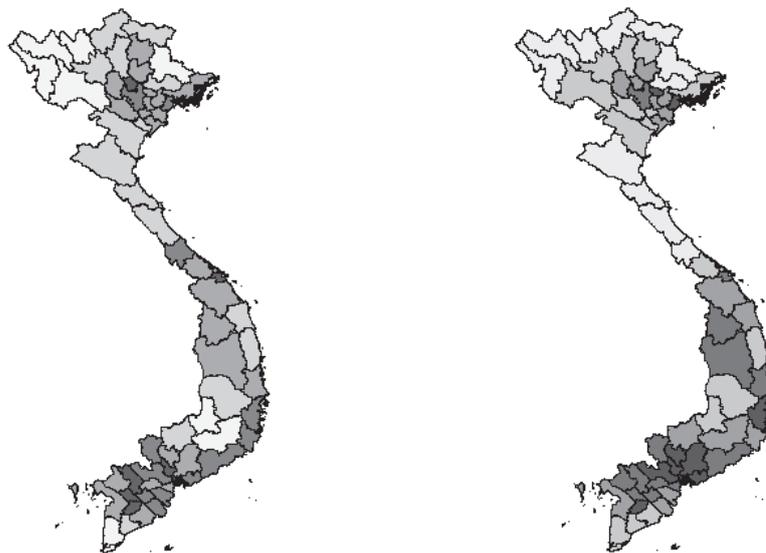


図3 プロビンスにおける被雇用者の割合（左：男性、右：女性）

表1 Region ごとの被雇用者の度数と割合（男性）

Region	n	頻度	回答数	割合
1. Red River Delta	10	830	2510	0.33
2. North East	15	780	3100	0.25
3. North Central	6	470	1720	0.27
4. South Central Coast	6	460	1440	0.32
5. Central Highlands	5	290	1130	0.26
6. South East	8	670	1970	0.34
7. Mekong River Delta	13	1010	3050	0.33

表2 Region ごとの被雇用者の度数と割合（女性）

Region	n	頻度	回答数	割合
1. Red River Delta	10	570	2510	0.23
2. North East	15	390	3100	0.13
3. North Central	6	190	1720	0.11
4. South Central Coast	6	310	1440	0.21
5. Central Highlands	5	210	1130	0.18
6. South East	8	540	1970	0.27
7. Mekong River Delta	13	680	3050	0.22

次に、男女別の rate.work.s において、グローバル Moran 統計量により空間自己相関を指標化した。男性は 0.64、女性は 0.61 とともに高い数値を示した。

また、ローカル Moran 統計量と被雇用率を比較することにより、ホットスポットの候補となるプロビンスを確認した。図 4 には男性の、図 5 には女性の結果を示す。

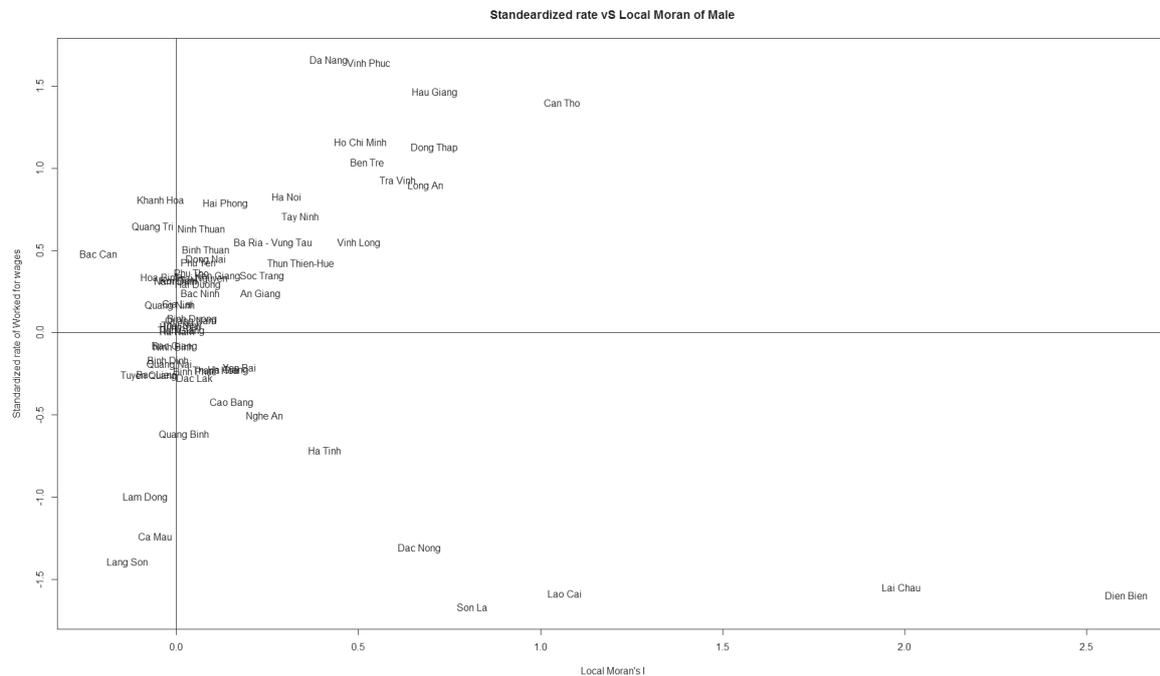


図 4 ローカル Moran 統計量と標準化した被雇用率 (rate.work.s) (男性)

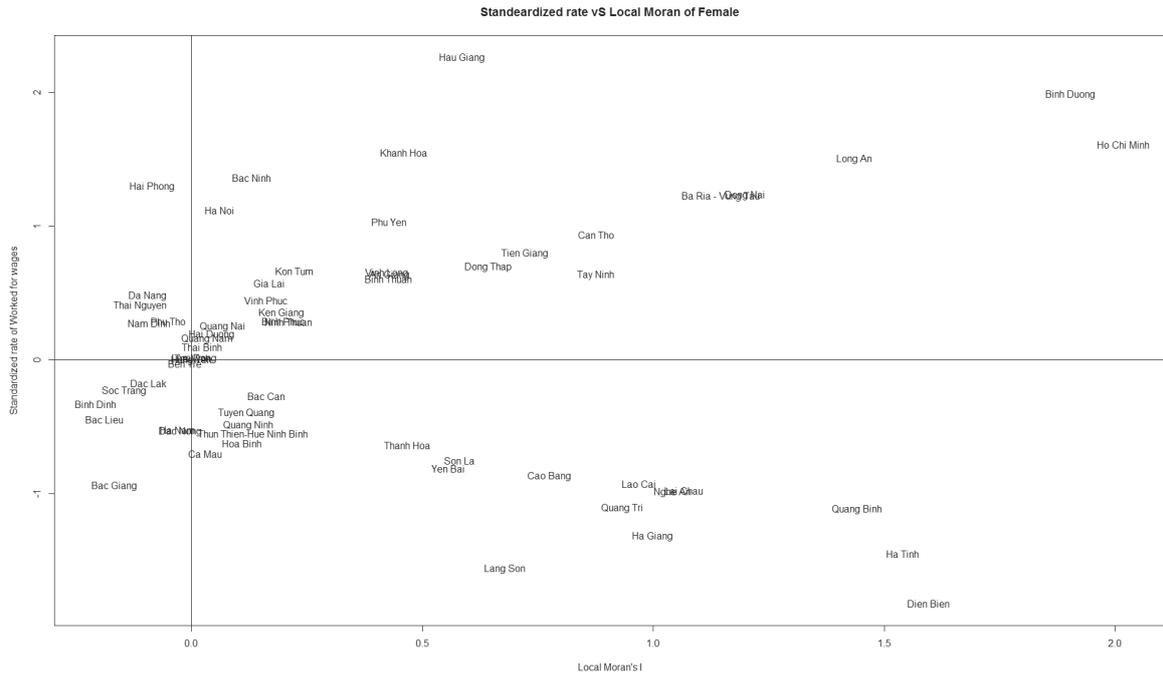


図5 ローカル Moran 統計量と標準化した被雇用率 (rate.work.s) (女性)

次に、Kubota(2016)における決定木分析や、Kubota(2017)における空間自己回帰モデルとの比較のために、まずは、線形回帰モデルにより、rate.wark.sを目的変数として、age, sex, egrade, marriage.r, relation.h, msuffer.rを説明変数として予測を行った。表3にRの出力結果、図6に回帰診断の結果のプロットを示す。

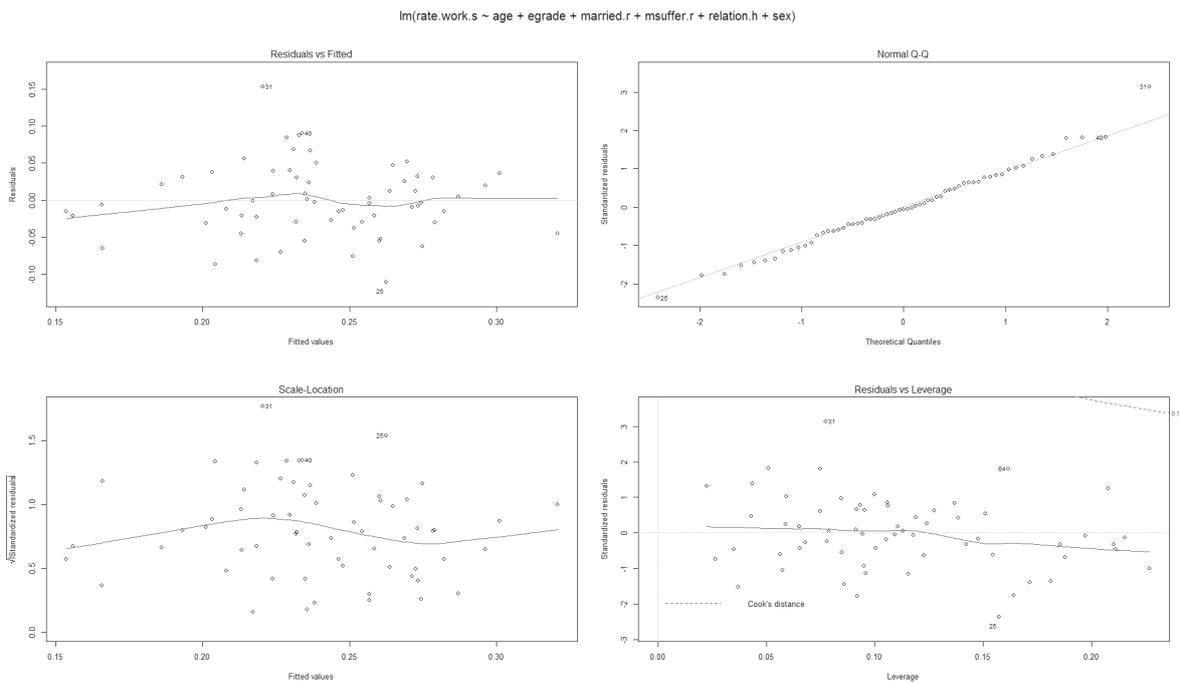


図6 回帰診断の結果

表3 線形回帰の結果 (R の出力結果)

```
lm(formula = rate.work.s ~ age + egrade + married.r + msuffer.r + relation.h + sex, data = tbl.all)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.009536	0.263960	-0.036	0.97131
age	0.012567	0.004485	2.802	0.00697 **
egrade	0.004921	0.007436	0.662	0.51077
married.r	-0.227744	0.203855	-1.117	0.26868
msuffer.r	0.149263	0.063375	2.355	0.02204 *
relation.h	-0.449322	0.473550	-0.949	0.34678
sex	-0.014329	0.454298	-0.032	0.97495

Residual standard error: 0.05091 on 56 degrees of freedom

(2 observations deleted due to missingness)

Multiple R-squared: 0.337, Adjusted R-squared: 0.266

F-statistic: 4.744 on 6 and 56 DF, p-value: 0.0005687

AIC : -187.818

また、MCMC の方法を用いて、空間自己回帰モデルにより、先述の線形回帰モデルと同様の目的変数、説明変数を元に予測を行った。表4にRの出力結果、図7にロビンスごとの実測地と予測値および95%信頼区間を示す。図8には、実測地と予測値の回帰診断を示す。

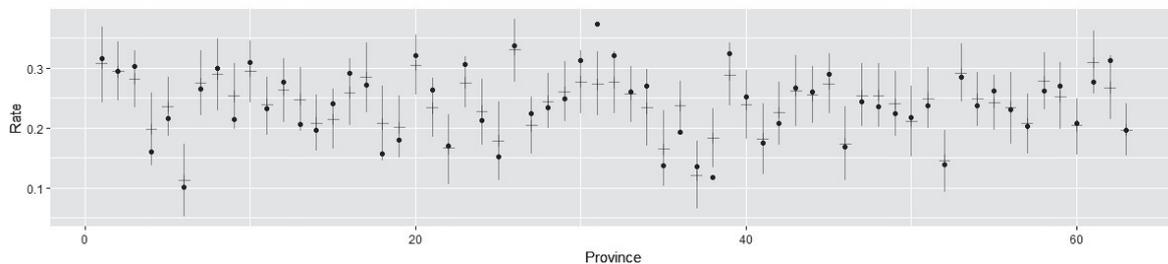


図7 プロビンスごとの実測値と予測値および95%信頼区間

表 4 空間自己回帰モデルを用いた回帰の結果 (R の出力結果)

```
#####
#### Model fitted
#####

Likelihood model - Gaussian (identity link function)
Random effects model - Leroux CAR
Regression equation - rate.work.s ~ age + egrade + married.r + msuffer.r + relation.h + sex
Number of missing observations - 0

#####
#### Results
#####

Posterior quantities and DIC
```

	Median	2.5%	97.5%	n.sample	% accept	n.effective	Geweke.diag
(Intercept)	-0.0384	-0.6028	0.5178	3000	100	3000.0	0.4
age	0.0064	-0.0036	0.0166	3000	100	2763.9	-0.8
egrade	0.0184	-0.0021	0.0387	3000	100	2214.0	1.0
married.r	0.0674	-0.5489	0.6893	3000	100	1902.0	0.1
msuffer.r	0.0593	-0.0692	0.1909	3000	100	3000.0	-1.4
relation.h	-0.4027	-1.4597	0.6594	3000	100	3000.0	-0.6
sex	0.0329	-0.9246	0.9915	3000	100	3000.0	0.1
nu2	0.0017	0.0011	0.0028	3000	100	2833.9	-0.8
tau2	0.0027	0.0013	0.0054	3000	100	2566.9	0.6
rho	1.0000	1.0000	1.0000	NA	NA	NA	NA

```

DIC = -206.097      p.d = 26.65979      LMPL = 124.1584
#####
```

5. 考察

本研究においては、ベトナムの家計調査における地域ごとの働き方の比較をおこなうために、VHLSS2006 のデータを用いてグラフ等により視覚化やインタラクティブに操作を可能とした。また、プロビンスごとにまとめられたデータを元に、線形回帰モデルや空間自己回帰モデルにより分析を行うとともに空間自己相関について確かめた。

視覚化においては、図 1 より、続柄においては、Children の割合が多く、また家長については男性の割合が多い等、また図 2 より、Region 内でのプロビンスの詳細やコロプレスマップでの割合について等 shinydashboard を利活用して視覚化したグラフをインタラクティブにしかも一画面に表示することで全体的な傾向を知ることが可能となった。

また、空間自己相関においては、図3やグローバル Moran 統計量の結果より男性・女性ともに自己相関がみられた。一方で、図4や図5からローカル Moran 統計量と（標準化した）観測地のプロットの第一象限において、ホットスポットの候補のプロビンスを確認したところ、図4より男性では「Can Tho」「Dong Thap」「Hau Giang」、図5より女性では「Ho Chi Ming」「Binh Duong」「Long An」と傾向がやや違うことがうかがえた。

さらに、線形回帰モデルでは説明変数のうち age と msuffer.y が有意になった一方で、空間自己相関モデルでは egrade だけがほぼ0を超えているので関係があり、線形回帰よりも関係のある説明変数が少なくなった。なお、空間自己相関の変数は有意になっているので、その影響が大きいのではと考察された。

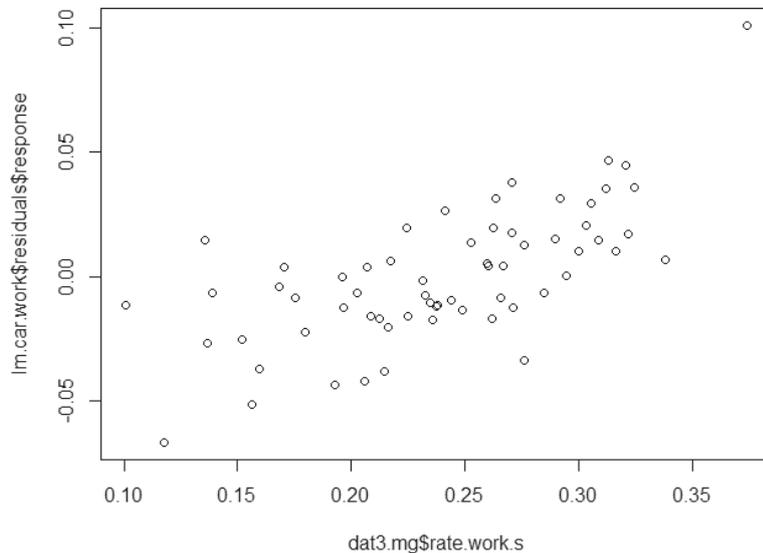


図8 実測値と予測値の回帰診断

謝辞

本研究は統計数理研究所共同利用「アジア諸国世帯統計マイクロデータの二次利用推進に関する研究」(29-共研-2048)の一部である。

参考文献

- Hijmans, R. (2016), GADM database of Global Administrative Areas, URL: <http://www.gadm.org>
- Kubota, T. (2016), Exploratory data analysis of education and medical situation of youth in Vietnam, The 8th International Workshop on Analysis of Micro Data of Official Statistics
- Kubota, T. (2017), Visualization and Spatial Statistical Analysis for Vietnam Household Living Standard Survey, The 2017 conference of the International Federation of Classification Societies
- 久保田貴文 (2017)、ベトナム家計調査データの視覚化、2017年度 統計関連学会連合大会予稿集 PP. 343