

## Rによるデータ分析入門：練習問題・回答例 第5章

1. 労働経済学では、結婚している人は結婚していない人に比べて賃金が高いという事実をマリッジ・プレミアと呼び、その要因を分析する研究がさかんに行われています。このマリッジ・プレミアは主に男性で観察され、結婚することにより仕事に専念できるようになるので賃金があがる、といった解釈もありますが、将来出世しそうな魅力的な男性ほど結婚相手を見つけやすいという事実を反映しているだけという見方もあります。ここでは、個人を追跡したパネル・データを使い、個体固定効果で観察できない個人の属性をコントロールした上でマリッジ・プレミアがどの程度のものなのかを分析してみましょう。

(1) 賃金の対数値 (lwage) を被説明変数、結婚ダミー (married)、教育年数 (educ)、経験年数 (exper)、黒人ダミー (black)、ヒスパニックダミー (hissp) を説明変数としてプーリング回帰と固定効果モデルで賃金関数を推定せよ。なお、固定効果モデルは個体固定効果のみ、と個体固定効果と年固定効果の両方を入れた2つの式を推定せよ。

推計結果は以下の通り。result1 がプーリング回帰、result2 が個体固定効果のみの固定効果モデル、result3 は個体固定効果と年固定効果の両方を入れた推計式である。

```
> fixest::etable(result1, result2, result3, se.below=TRUE)
```

	result1	result2	result3
Dependent Var.:	lwage	lwage	lwage
Constant	0.0592 (0.0638)		
married	0.1226*** (0.0158)	0.0632** (0.0214)	0.0606** (0.0215)
educ	0.1038*** (0.0046)		
exper	0.0498*** (0.0029)	0.0595*** (0.0034)	
black	-0.1120*** (0.0237)		
hisp	0.0236 (0.0211)		
Fixed-Effects:	-----	-----	-----

nr	No	Yes	Yes
year	No	No	Yes
<hr/>			
S. E. type	IID	by: nr	by: nr
Observations	4,360	4,360	4,360
R2	0.16199	0.61282	0.61366
Within R2	--	0.16308	0.00284

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(2) married の係数は、プーリング回帰と固定効果モデルでどの程度異なるか？なぜ、係数が変わってしまったかを説明せよ。

married の係数はプーリング回帰の 0.123 から、2 つの固定効果を入れた result3 では 0.061 まで低下している。これは、プーリング回帰の場合、結婚にも賃金にも影響するが分析者には観察されない要因、たとえば社交性といった要因が考慮されておらず、そのため結婚ダミー married と賃金 lwage の間に強い相関を生じさせている。一方、固定効果モデルの場合、社交性といった観察されない要因は個体固定効果で考慮されているため、結婚ダミー married の係数は、結婚が賃金に及ぼす因果効果として解釈できる。

(3) 固定効果モデルでは一部の変数の係数が得られない。係数が得られない変数はどのような変数で、なぜ係数が得られないのかを説明せよ。

固定効果モデルでは、時間を通じて変化しない変数は個体固定効果と完全に相関してしまうため人種ダミーといった変数は説明変数に追加してもその係数が得られない。

2. アメリカでは州ごとに異なる司法機構があることが知られているが、州内であっても、地域ごとに逮捕されたのち有罪になる比率や、有罪になったのち実刑判決になり収監される比率、平均の刑期も異なる。また警察予算も異なるため、人口当たりの警察官や逮捕される確率も異なる。Cornwell and Trumbull (1994) は州よりも細かい郡レベルのデータを用いて、こうした地域ごとの「罪を犯すコスト」の違いが犯罪発生率に及ぼす影響を分析している。具体的には、ノースカロライナ州の 90 の郡のデータを用いて、被説明変数に犯罪発生率 (lcrmrte)、説明変数には逮捕される確率 (lprobarr)、逮捕されたのち有罪になる確率 (lprobconv)、有罪になったのち収監される確率 (lprobpris)、刑期の平均 (lavgsen)、人口当たり警察官の数 (lpolpc)、人口密度 (ldensity) を説明変数としています。なお、こ

これらの変数はすべて対数変換されています。これに加えて州内の地域を west, centra, other に区別するカテゴリー変数 region から作成した地域ダミーも用意されています。

(1) 各々の説明変数の係数はプラスとマイナス、どちらになると考えられるか予測し、その理由を述べよ。

- ・逮捕される確率 (lprobarr) : 犯罪を犯しても逃げ切れない可能性が高いため犯罪は減少する、係数はマイナス

- ・逮捕されたのち有罪になる確率 (lprobconv) : 逮捕されると有罪になる可能性高いことが認知されると犯罪は減少すると考えられる、係数はマイナス

- ・有罪になったのち収監される確率 (lprobpris)、刑期の平均日数 (lavgsen) : いずれも犯罪のコストを示す変数であり、コストが高まれば犯罪は減少すると考えられる、係数はマイナス

- ・人口当たり警察官の数 (lpolpc) : 警察官が多いと犯罪が摘発される可能性が高まり、やはり犯罪のコストが高まるので犯罪は減少する、係数はマイナス

- ・人口密度 (ldensity) : 人口密集地域の方がスリ・暴行や強盗が発生しやすいと考えれば犯罪は増加する、この場合、係数はマイナス

(2) 上記の回帰式を、①固定効果を一切含まないモデル、②年次固定効果のみを含むモデル、③郡固定効果と年次固定効果の両方を含むモデルの3つを推定せよ。なお、郡と年を示す変数はそれぞれ county と year である。結果は etable () で出力すること。

推計結果は以下の通り。result1 は①、result2 は②、result3 は③に対応する。

```
> fixest::etable(result1, result2, result3, se.below=TRUE)
```

	result1	result2	result3
Dependent Var. :	lcrmte	lcrmte	lcrmte
Constant	-2.035*** (0.1763)		
lprobarr	-0.5227*** (0.0307)	-0.5163*** (0.0313)	-0.3560*** (0.0591)
lprobconv	-0.4023*** (0.0223)	-0.3939*** (0.0283)	-0.2825*** (0.0504)
lprobpris	-0.0223 (0.0499)	-0.0205 (0.0469)	-0.1802*** (0.0452)

lavgsen	-0.1042**	-0.1398**	-0.0044
	(0.0403)	(0.0256)	(0.0329)
lpolpc	0.3221***	0.3250***	0.4214***
	(0.0227)	(0.0273)	(0.0846)
ldensity	0.2521***	0.2585***	0.4073
	(0.0201)	(0.0255)	(0.3848)
factor(region)west	-0.5670***	-0.5680***	
	(0.0291)	(0.0102)	
factor(region)central	-0.2317***	-0.2352***	
	(0.0273)	(0.0141)	
Fixed-Effects:	-----	-----	-----
year	No	Yes	Yes
county	No	No	Yes
S. E. type	IID	by: year	by: county
Observations	630	630	630
R2	0.77138	0.77824	0.95088
Within R2	--	0.77622	0.37123
---			
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

(3) ①～③で説明変数として加えたものの、係数が得られなかった変数があるか？もし、あるならば①～③式のどの変数か、また、なぜ係数が得られなかったかを説明せよ。

west ダミー（西部地区ダミー）と central（中部地区ダミー）の係数が郡固定効果と年次固定効果の両方を含むモデル③（result3）で得られなかった。3つの推計式のうち、この推計式のみ郡固定効果が含まれているが、west と central は時間を通じて変化しない変数であり、固定効果と完全相関してしまうため係数が得られなかったと考えられる。

(4) ①～③で期待通りの符号の係数が得られたか、また、①から③にかけて、各々の説明変数の係数がどのような変化したかを説明せよ。

推定式①では、一人当たりの警察官の数以外は(1)の予想通りの結果が得られている。年次固定効果を追加した推計式②では、やや係数に変化があるものの①の結果と概ね同じであった。一方、年次固定効果と郡固定効果を追加した推計式③では、人口密度、刑期の平均の有意性が失われていることがわかる。この理由として、これらの変数は分析者に観察できな

い郡固有の要素と相関をもっており、そのため郡固定効果を考慮すると有意性が失われたと考えられる。