計量分析演習

第6回

岡島　成治

傾きパラメターをどう解釈するか？

被説明変数	説明変数	解釈
Y(レベル)	X(レベル)	Xが１単位増えたとき、Yがβ_1単位増える。
lnY(ログ)	X(レベル)	Xが１単位増えたとき、Yが$100 \times \beta_1\%$単位増える。
Y(レベル)	lnX(ログ)	Xが１％増えたとき、Yがβ_1/100単位増える。
lnY(ログ)	lnX(ログ)	Xが１％増えたとき、Yが$\beta_1\%$増える。

傾きパラメターをどう解釈するか？

- ログレベルモデル
- 所得が教育年数に関係があるのかの式

$$ln賃金 = \beta_0 + \beta_1職業年数 + u$$

を回帰分析した結果

Source	SS	df	MS			
Model	.999640572	1	.999640572	Number of obs	=	545
Residual	168.079453	543	.309538587	F(1, 543)	=	3.23
				Prob > F	=	0.0729
Total	169.079093	544	.310807157	R-squared	=	0.0059
				Adj R-squared	=	0.0041
				Root MSE	=	.55636

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
exper	.0259028	.0144139	1.80	0.073	-.0024111	.0542166
_cons	1.315388	.0495595	26.54	0.000	1.218037	1.41274

Stataの結果

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.31539    0.04956  26.542   <2e-16 ***
exper        0.02590    0.01441   1.797   0.0729 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Rの結果

職業年数が１年増えると2.5%（推定値は0.0025）賃金が増える

傾きパラメターをどう解釈するか？

- レベルログモデル

$$\text{所得} = -515.47 + 297.53 ln\text{修学年数} + u$$

修学年数が１％伸びると、年収が2.97万円増える。

- ログログモデル

$$ln\text{所得} = 3.15 + 0.81 ln\text{修学年数} + u$$

修学年数が１％伸びると、年収が0.81%万円増える。

もう一度、政策の「効果」とは？

- 「朝ごはんを毎日食べている生徒はテストの点が高い」というのは、朝ごはんを食べている生徒のほうが、そうでない生徒に比べてテストの点が高い「傾向」（相関関係）があると言っているに過ぎない。
- 家庭環境がテストの点に影響している可能性がある

朝ごはんと食べさせている親は子供の教育に熱心

なので、いままで朝ごはんを食べなかった家庭の生徒に朝ごはんを食べさせてもテストの点はきっと変わらない。

よってきちんと因果関係を推定しなければいけない。

重回帰モデル

家庭環境等の外的条件がそろっていない場合には、外的条件に関する情報を制御する確率変数Cのとりうる値 c を使い外的条件をそろえる。(重回帰分析)

$$\mathrm{E}[Y|X=1, C=c] - E[Y|X=0, C=c]$$

重回帰モデル

- 関数のモデル化

因果関係をみるには、政策変数X=xと家庭環境を条件付けした成果変数Yの期待値を考えればよい。

$$E[Y|X,C] = \beta_0 + \beta_1 X + \beta_2 C$$

一般化

$$E[Y|X_i] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots + \beta_k X_k$$

重回帰モデル

この式はあくまでも平均的な値

$$E[Y|X,C] = \beta_0 + \beta_1 X + \beta_2 C$$

しかしたとえ朝ごはんを毎日食べていてもテストの日の調子によって実際のテストの点数は上下する。この「揺らぎ」を誤差項として書き直す。

$$Y = \beta_0 + \beta_1 X + \beta_2 C + U$$

セテリス・パリブス・アプローチ

- 共変数Cを使って外的条件を制御しながら、政策変数の効果を見るアプローチは、「他の要件を一定とする」という意味のラテン語からセテリス・パリブス・アプローチと言われている。

- 「家計所得という他の要因は一定として、朝ごはんを毎日食べるとテストの点数がβ_1だけ上がる」
- 「朝ごはんを毎日食べるという他の要因は一定として、家計所得を増やすとテストの点数がβ_2だけ上がる」

重回帰分析の推定

- 復習（単回帰分析の場合）
- モーメント法を使った
- 誤差項の平均独立と誤差項の期待値が0から

$$\mathrm{E}[\mathrm{U}|\mathrm{X}] = \mathrm{E}[\mathrm{U}] = 0$$

さらに$\mathrm{E}[\mathrm{U}] = 0$という仮定の下では$\mathrm{E}[X\mathrm{U}] = 0$

重回帰分析の推定

- モーメント法を使かう
- 誤差項の平均独立と誤差項の期待値が0から

$$\mathrm{E}[\mathrm{U}|\mathrm{X}] = \mathrm{E}[\mathrm{U}|\mathrm{C}] = \mathrm{E}[\mathrm{U}] = 0$$

さらに$\mathrm{E}[\mathrm{U}] = 0$という仮定の下では$\mathrm{E}[X\mathrm{U}] = \mathrm{E}[\mathrm{CU}] = 0$

決定係数

- 最小２乗法によって求めた一次関数が、どの程度データを説明してくれているのかの指標

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \widehat{u_i^2}}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

もし被説明変数を完全に説明してくれる一次関数を見つけることができれば決定係数は１になる。

自由度調整済み決定係数

- 決定係数は説明変数の数が多くなれば大きくなる。よって説明変数の数が違う重回帰分析のモデルを比べるのは不公平なので説明変数の数に影響を受けない自由度調整済み決定係数を用いる。

$$R_a^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1}$$

k: 説明変数の数

n: 標本サイズ

ビンゴゲーム

- このクラスの人をどれだけ知っていますか？
- 縦3マス横3マスの計9マスのマス目を作成してください。
- そのマス目にこのクラスの知っている人の名前を記入して下さい。
- 縦横斜めのラインが3本早くそろった人には景品があります。

問題 1

データファイル「IceCream2」をインポートして下さい。

このデータは以下の項目を含んでいます。

- Icecream:世帯当たりのアイスクリーム年間消費額
- Income:年間所得
- U15:世帯当たり１５歳以下の子供の平均人数

1. 被説明変数をIcecream説明変数をIncomeにして散布図と書いて回帰分析してください。
2. アイスクリームの年間消費額を年間所得と１５歳以下の子供の人数に回帰してください。

問題 2

データファイル「Males」をインポートして下さい。
このデータは以下の項目を含んでいます。

- school:教育を受けた年数
- exper:職業経験の年数
- wage:賃金の対数

1. 変数wageをschoolに回帰し、散布図中に回帰曲線を図示してください。
2. １の結果から教育年数が増えると賃金が何％増加するか読み取って下さい。
3. 変数wageをschoolとexperに回帰し、その結果を２と比較してください。