

実証レポートの書き方

岡島 成治

研究者になって3年が経ちましたが、未だにレポートを書くことが凄く難しいと感じています。それどころかレポートを書く難しさが最近になってしみじみと分かってきました。シカゴ大学のCochrane教授によると研究の80%を書くことに費やし、たった残りの20%をデータ収集やモデルの試行、その分析等に費やすそうです。また私のアドバイザーでオハイオ州立大学のHaab教授からは「素晴らしい分析でも解かりやすく書いていない論文と、つまらない分析でも解かりやすく書いてある論文では後者の方がより素晴らしい論文だよ」と留学中に言われました。そこで今回、私は実証レポートの書き方について取り上げます。

1. イントロダクション

レポートの構成で最初の部分にあたるのがイントロダクションです。イントロダクションは単にレポートの最初にくるだけでなく、そのレポートの質を表しています。イントロダクションが面白くなければ大抵の読み手はレポートを読むことをやめます。というのは、読み手は我々書き手が考えているほど暇ではないからです。だからこそ、どんなに忙しい読み手に出会っても自分のレポートを最後まで読んでもらえるような興味深いイントロダクションを書いて下さい。そのようなイントロダクションを書くには以下の4つの項目を必ず述べて下さい。

- (1) あなたが疑問に思ったことは何か？
- (2) なぜ、あなたはこの問題に疑問を抱いたか？

- (3) あなたの研究の貢献は何か？どのようにあなたはこの疑問に答えたか？そしてあなたはどのようなデータを利用して政策を定量分析したのか？
- (4) あなたの研究の主な結果は何か？その結果は以前の文献とどう違うのか？

上の4つの項目を、例えば私が現在取り組んでいる研究論文にあてはめると、以下ようになります。

- (1) 家庭の電力消費量を抑えるには、どのような政策が良いのか？
- (2) 日本政府は東日本大震災後に電力消費量を削減するためにマスメディアを利用して電力消費量を抑えて下さいとお願いした。その政策は実際に効果があったのか？
- (3) 電力需要と工業生産指数のデータは「エネルギー消費統計調査」を資源エネルギー庁から、気温のデータは気象庁から入手して、日本政府が東日本大震災後に電力消費量を削減するためにマスメディアを利用して電力消費量を抑えてくださいとお願いした政策の効果を差の差推定法を用いてきちんと因果関係を考慮して分析した。
- (4) 今までの研究では、短期的には電力消費量を削減するためにマスメディアを利用して電力消費量を抑えて下さいとお願いした政策の効果がある (Reiss and White 2005)、と指摘されていたが、長期的には効果があるかどうかを分析されていなかった。私の研究では長期的にも効果があるということが分かった。

2. データ

次に自分の研究で用いるデータについて説明します。その時には以下の4つの点に必ず触れて下さい。

- (1) データの入手方法：研究で使ったデータはどこから入手したのかを述べて下さい。読者はデータが信頼されている所から入手しているのかどうかに注目しています。
- (2) データの種類とサイズ：研究で使ったデータは横断面データ、時系列データ、パネルデータのどれなのか、そしてその標本サイズはどのくらいの大きさなのかを述べて下さい。また、時系列データであるなら、そのデータが年次、月次、日次、秒次のどれであるかを示して下さい。横断面データもしくはパネルデータならば、個人データなのか家計データなのか、企業データなのかを述べて下さい。読者はデータの種類や標本の大きさでどの様な計量経済学の手法を使えばよいのかを想像できます。
- (3) データの観測単位と項目：データの項目とその観測の単位について明示して下さい。そのデータにはどのような項目があるのか、そしてどのような単位が含まれるのかを示して下さい。例えば、項目が自動車の走行距離の場合、計測単位は mile なのか km なのかを説明して下さい。本文中に説明するだけでなく表をつけると読者の助けになります。
- (4) データの基本統計：各データに関して、平均値、最大値、最小値、標準偏差等の統計量について表を用いて記述して下さい。読者はその記述統計表を見てデータの性質を判断することが出来ます。

以下にデータの説明の例を記述します。

例：私は仕事場が禁煙の場合どのような人が喫煙者になるかを分析しました。そこでアメリカの” the National Health Interview Survey in 1991” からデータを入手してこの事柄を定量分析しました。このデータは横断面データの個人データで、その標本サイズは 1000 個です。表 1 にはそのデータにある変数の定義が示されています。また表 2 にはその変数の記述統計が示されています。

表 1 : smoker データにおける変数とその定義 (n=1000)

変数	定義
<i>smoker</i>	=1 喫煙者なら, =0 非喫煙者なら
<i>smkban</i>	=1 仕事場が禁煙なら, =0 そうではない
<i>age</i>	年齢
<i>hsdrop</i>	=1 高校中退なら, =0 そうではない
<i>hsgrad</i>	=1 高校卒業なら, =0 そうではない
<i>colgrad</i>	=1 大学卒業なら, =0 そうではない
<i>black</i>	=1 黒人なら, =0 そうではない
<i>hispanic</i>	=1 ヒスパニック系なら =0 そうではない
<i>female</i>	=1 女性, =0 そうではない

表 2 : 記述統計

変数	平均値	標準偏差	最小値	最大値
<i>smoker</i>	.242	.428	0	1
<i>smkban</i>	.609	.487	0	1
<i>age</i>	38.693	12.113	18	88
<i>hsdrop</i>	.091	.287	0	1
<i>hsgrad</i>	.326	.468	0	1
<i>colgrad</i>	.197	.397	0	1
<i>black</i>	.076	.266	0	1
<i>hispanic</i>	.113	.317	0	1
<i>female</i>	.563	.495	0	1

表 2 のような記述統計の良い点は、例えば全サンプルに占める女性の

割合は56%であることや黒人の割合は7%であること、平均年齢は40歳付近であることなど、読者が簡単におおよそのデータの背景をイメージできることにあります。

3. モデル

分析するにあたって計量経済学のモデルを書く必要があります。その際には、必ず次の2点を明示して下さい。1点目はあなたがどのような手法を使ったのか、その理由はなぜか。2点目はどのような変数を使ったのか、そしてその変数を入れた理由はなぜか、です。

例) 最小2乗法の代わりにプロビットモデルを利用しました。なぜならば被説明変数がダミー変数だからです。そして私が分析に用いたモデルは以下の通りです。

$$\Pr(Y = 1|X) = \Phi(\alpha + X'\beta + u)$$

被説明変数はsmokerで喫煙者なら1, 非喫煙者なら0となるダミー変数です。説明変数には、まずこの研究の疑問点である仕事場が禁煙しているかどうかのダミー変数smkbanを用いました。さらに高学歴の人はタバコを吸わないと言われているので、大卒かそうでないかのダミー変数colgrad、高卒かそうでないかのダミー変数hsgrad、高校中退かそうでないかのダミー変数hsdropも利用しました。

4. 結果

回帰係数の推定値だけでなく、標準誤差やt値p値などの統計的な有意性を示す値についてどれか一つ選んで報告します。また回帰式の推定は必ず数本行って下さい。なぜならば他の回帰式の推定と比べても中心となる回帰式の推定値がそれほど変わらないことを示す必要があるからです。3本から5本ぐらいの回帰式を推定し、それぞれの結

果を報告するようにするのが理想です。

また推定結果の表を必ず報告して下さい。異なる説明変数の回帰式の推定値と標準誤差か t 値あるいは p 値のどれか一つを報告します。

表の形式は、表 3 を参考にして下さい。

表 3：プロビットモデルにおける喫煙行動の影響

被説明変数：smoker			
説明変数	(1)	(2)	(3)
smkban	-0.183*** (0.028)	-0.168*** (0.028)	-0.167*** (0.028)
colgrad	-0.265*** (0.042)	-0.270*** (0.042)	-0.282*** (0.042)
hsdrop	0.480*** (0.048)	0.485*** (0.048)	0.581*** (0.051)
hsgrad	0.328*** (0.032)	0.338*** (0.032)	0.342*** (0.032)
age		-0.004*** (0.001)	-0.005*** (0.001)
female		-0.079*** (0.028)	-0.083*** (0.028)
black			-0.049 (0.052)
hispanic			-0.304*** (0.047)
Constant	-0.712*** (0.029)	-0.503*** (0.053)	-0.436*** (0.054)
サンプルサイズ	10,000	10,000	10,000

(n)

()内は標準誤差

*** は 0.01%, ** は0.05%, * は0.1%で有意であることを表す

また研究テーマとして選ばれた説明変数の中で、最初に一番重要な変数について推定値の大きさを評価して、それがどのような意味を持つか、そしてその変数の統計的有意性を読者に説明して下さい。もし統計的に有意ではなければ、なぜ有意にならなかったのかを説明して下さい。次にその他の興味のある変数についても同様に説明して下さい。

例えば表3のような結果が得られたならば、列(1)では仕事場が禁煙している場合18%の人が禁煙すると分かったが、列(2)において年齢と性別を加えて分析すると仕事場が禁煙している場合17%の人が禁煙すると分かった。さらに列(3)において年齢と性別と人種(黒人かどうかとヒスパニック系かどうか)を加えて分析すると仕事場が禁煙している場合17%の人が禁煙すると分かった。ここで重要なことは列(1)の推定値は他の変数を加えても変化しておらず、統計的にも5%有意水準で信頼できうる結果である。

5. 結論

あなたのレポートの内容をもう一度要約して下さい。そして今までの既存の研究では発見されていないことを発見したか、あるいは既存の研究と同じ結果になったのを再確認したのか等を既存研究と比較して述べて下さい。最後にレポート作成に用いた文献リストを付け足します。文献リストにある文献は本文中に引用した文献に限ります。これでレポートは完成です。

参考文献

Cochrane H. John. 2005. “ Writing Tips for Ph.D. Students.”

https://faculty.chicagobooth.edu/john.cochrane/research/papers/phd_paper_writing.pdf (最終検索日:2017 年 6 月 5 日)

National Health Interview Survey

<https://www.cdc.gov/nchs/nhis/> (最終検索日:2017 年 6 月 5 日)

Reiss, Peter C., and Matthew W. White. 2008. “What changes energy consumption? Prices and public pressures.” *RAND Journal of Economics*, 39(3): 636–663.