

平成25年度卒業論文

浪人と留年

所属ゼミ 村澤ゼミ

学籍番号 1100401107

氏名 中司 雄也

大阪府立大学

経済学部

要約

本稿では、大学の現役入学生と浪人入学生でどちらが留年しやすいか比較する。2005年「社会階層と社会移動(SSM)」全国調査の個票データを利用して、回答者の大学入学時と卒業時の年齢から浪人ダミーと留年年齢を作成し、そして、ポアソン回帰分析を行った結果、浪人・男性・国公立大学生は現役合格生・女性・私立大学に比べ、それぞれ平均的に約0.0804年・約0.0912年・約0.1536年多く留年することがわかった。また、最も留年しやすい年代は大学に在学期間が1970年代であるとわかった。

目次

第 1 章	はじめに	1
第 2 章	先行研究	2
第 3 章	データ	3
1.	SSM 調査.....	3
2.	変数の作成	3
3.	要約統計量.....	4
第 4 章	分析手法	6
1.	ポアソン分布.....	6
2.	ポアソン回帰モデル.....	6
3.	最尤推定.....	6
4.	限界効果.....	7
第 5 章	分析結果	9
1.	gretl の手順.....	9
2.	gretl による分析結果	9
3.	ポアソン回帰分析による結果.....	10
4.	各変数の観測結果	10
5.	OLS 回帰分析とポアソン回帰分析の結果を比較	11
第 6 章	おわりに	13
謝辞	14
参考文献	15

第1章 はじめに

本稿では、大学の現役入学生と浪人入学生でどちらが留年しやすいか比較する。2005年「社会階層と社会移動(SSM)」全国調査の個票データを利用して、大学入学時と卒業時の年齢から浪人ダミーと留年年齢を作成し、そして、ポアソン回帰分析を行った結果、浪人・男性・国公立大学生は現役合格生・女性・私立大学に比べ、それぞれ平均的に約0.0804年・約0.0912年・約0.1536年多く留年することがわかった。また、最も留年しやすい年代は大学に在学期間が1970年代であるとわかった。

本稿の構成は以下の通りである。まず第2章で先行研究を紹介する。次に第3章で使用するデータについて、第4章でポアソン回帰分析について説明する。そして、第5章で分析結果を示し、第6章でその結果について考察する。

第 2 章 先行研究

浪人と留年の関係进行分析した文献は見当たらない。しかし、浪人と留年は個々にある。

浪人について研究した Ono(2007)の文献がある。Ono(2007)では、ハイレベルな大学に行くほど所得は高く、浪人の影響が所得には関係ないと結論がでた。つまり、現役合格でも浪人でも偏差値の高い大学に行くことができれば所得が高くなる可能性があるので、浪人して偏差値の高い大学を目指すのは良い投資である。

そして、留年について分析した内田(2013)の文献がある。留年率は男性が 6 年連続減少したが、前年度より 2 年連続で上昇しており、女子が今年度よりやや減少した。そして、上昇していた留年率が 2000 年ごろより横ばいから減少傾向になったことの原因は明らかでない。

第3章 データ

1. SSM 調査

本稿では 2005 年社会階層と社会移動 (Social Stratification and Social Mobility, 以下 SSM) 全国調査のデータを使用する。SSM 調査は日本で最も伝統のある大規模な社会調査の 1 つで、1955 年の第 1 回調査から 10 年ごとに実施されている。

調査票は面接票と留置票に分かれており、留置票には A 票と B 票の 2 種類がある。調査対象は、2005 年 9 月 30 日現在満 20 歳～69 歳の男女である。

分析結果に間違った影響を与えるので「回答したくない」、「わからない」、「無回答」は除いてある。有効抽出票数は 13031、有効票数は 5742 である (回収率は 44.1%)。

2. 変数の作成

本稿は大学卒業者を対象として分析している。浪人は入学時年齢が 19 歳以上の人とする。19 歳以上でも浪人ではない状況があるが、すべて浪人とする。また、留年は大学の卒業年齢と入学年齢の差が 5 年以上の人とする。留学や休学など留年以外の可能性があるが、全て留年とする。以下に大学の入学者と卒業者の年齢と人数を度数分布表として表 1 に表す。

表 1 大学入学年齢と卒業年齢の度数分布

年齢	入学 (度数)	卒業 (相対度数)
17	1	0
18	647	0
19	207	0
20	59	0
21	16	1
22	5	600
23	4	217
24	0	75
25	1	24
26	0	15
27	0	7
28	0	0
29	0	1
30 以上	4	4
合計	944	944

分析に使用する変数の作成方法を説明する。全ての変数は留年について差異がある可能性を考慮する。

① 浪人ダミー

高校を卒業して大学に入学するまでに、1年以上期間が空いた人である。変数名は「ronin」とする。

なお、ダミー変数は gretl の「Dummies for selected discrete variables」という操作で作成することができる。本稿で用いる説明変数はすべてこの作業を行っている。

② 女性ダミー

変数名は「woman」とする。

③ 国公立ダミー

国立大学と公立大学は国公立大学として一つにまとめた。変数名は「national」とする。

④ 年齢

回答者の年齢である。この年齢は回答者の時代として解釈する。変数名は「age」とする。

⑤ 年齢 2 乗

回答者の年齢を 2 乗にする。変数名は「age2」とする。

3. 要約統計量

(1) 被説明変数

留年に関する変数名は「plus」とする。「卒業年齢－入学年齢－4」を留年年数とする。表 2 は留年年数の度数分布である。

表 2 留年年数の度数分布

留年回数	人数 (度数)	割合 (相対度数)
0	862	91.31
1	54	5.72
2	23	2.43
3	3	0.32
4	2	0.21
合計	944	100

(2) 説明変数

表 3 は説明変数の平均値である。この時の浪人ダミー、女性ダミー、国公立ダミーにおいて平均値は割合の意味を成している。

表 3 説明変数の平均値

説明変数	平均値
浪人ダミー	0.31
女性ダミー	0.28
国公立ダミー	0.29
年齢	44.45

すなわち、944 人中、浪人が 31% (246 人)、女性が 28% (266 人)、国公立大学出身者 29% (270 人) である。

第4章 分析手法

この章は北村（2009、第11章）を参考にしている。

本稿の分析のテーマである留年は非負の整数が発生しないことから、カウントデータ分析を採用する。カウントデータ分析とは、一般にある事象が決まった時間内に起こった回数を数え上げたことで集めた非負の整数を指し、その発生頻度を調べ、分布関数を特定化し、それに基づいて回帰分析することである。カウントデータの特徴として事象は稀にしか起こらず、多くの期間では事象が起こらない。

1. ポアソン分布

本稿の発生確率はポアソン分布で表す。ポアソン分布は、統計学上、稀にしか起こらない時に表すことが多い。ポアソン分布は未知のパラメータ λ が決まればすべての分布が決まる。ポアソン分布は①のように定義できる。

$$P(y=j) = \frac{e^{-\lambda} \lambda^j}{j!} \quad \lambda > 0, j=0,1,2,\dots \quad \textcircled{1}$$

また、この分布は $E(y) = \text{Var}(y) = \lambda$ といったように平均値と分散値が等しい。

2. ポアソン回帰モデル

ポアソン回帰モデルでは、 y の条件付き分布は、説明変数 x とポアソン分布のパラメータ $\lambda_i = \exp(x_i' \beta)$ によって決定されると考え、ポアソン回帰モデルは③のように定義される。

$$f(y_i | x_i' \beta) = \frac{\exp(-\exp(x_i' \beta)) \exp(y_i x_i' \beta)}{y_i!}, y_i = 0, 1, 2, \dots \quad \textcircled{3}$$

また、平均値と分散値は $E(y | x) = \text{Var}(y | x) = \exp(x' \beta)$ になる。ちなみに

$\lambda_i = \exp(x_i' \beta)$ の展開の方程式は $\exp(x_i' \beta) = e^{\beta_1 x_{i1} + \dots + \beta_k x_{ik}}$ になる。

3. 最尤推定

最尤推定とは、尤度を最大化するパラメータ値を推定する方法である。③の式を④の式のような対数尤度関数に変換し、それを最尤推定する。

$$f(y_i | x_i; \beta) = \frac{\exp(-\exp(x_i; \beta)) \exp(y_i x_i; \beta)}{y_i!}, y_i = 0, 1, 2, \dots$$

$$\Leftrightarrow \log L(\beta; y, x) = \sum_{i=1}^n \{y_i x_i; \beta - \exp(x_i; \beta) - \ln y_i!\} \quad \text{④}$$

対数尤度関数が大域的に凸であれば、最適解は一意的に決まる。そして、④の式の1階条件は $\sum_{i=1}^n (y_i - \exp(x_i; \beta)) x_i = 0$ である。これは最適化になる。

4. 限界効果

次に、推定されたパラメータ $\hat{\beta}$ の解釈をする。パラメータ $\hat{\beta}$ を求めることができると λ_i を求めることができる。そして、パラメータ $\hat{\beta}$ の解釈には2通りある。1つ目は、限界平均効果 (marginal mean effect) を用いて行う方法である。限界平均効果とは、各観測値において算出した限界効果の平均値である。限界平均効果を用いた方法は⑤の通りである。

$$\frac{\partial E(y_i | \bar{x}_i)}{\partial x_{ii}} = \exp(x_i; \beta) \beta_i = E(y_i | \bar{x}_i) \beta_i \quad \text{⑤}$$

ここで \bar{x} は x の平均値を表す。限界効果の評価は平均値でなくとも、特定の値で行うことができる。

この関係を次のように書き換えることができる。次の⑥の式では、 x_{ii} の限界的な変化に対して等しいものである。

$$\frac{\partial E(y_i | \bar{x}_i) / E(y_i | \bar{x}_i)}{\partial x_{ii}} = \beta_i \quad \text{⑥}$$

次にパラメータ $\hat{\beta}$ のもう1つの解釈を説明する。第2の方法は限界確率効果 (marginal probability effect) を用いることである。限界確率効果とは、個人 i について、変数 x_{ij} の (実現値から微小な変化の) 選択行動への影響を見るもの

である。これは、 $P=(y_i = j | x_i)$ を x_{ij} で微分する。

$$\frac{\partial P(y_i = j | x_i)}{\partial x_{ij}} = P(y_i = j | x_i) [j - \exp(x_i' \beta)] \beta_i \quad \text{⑦}$$

よって、推定されたパラメータ $\hat{\beta}$ を出すことで λ_i がわかり、ポアソン分布を抽出できる。

第5章 分析結果

1. gretl の手順

ポアソン回帰分析の gretl の手順は次の通りである。

- (a) Model を選択し、Nonlinear models を選択する
- (b) Count data model のタグを選択
- (c) Dependent variable は被説明変数
- (d) Independent variable は説明変数
- (e) Distribution は poisson を選択する
- (f) データを入力し、OK を押すと結果がでる

2. gretlによる分析結果

以下はgretlに表示されるポアソン回帰分析の結果である。

Model 3: Poisson, using observations 1-944
Dependent variable: plus

	<i>Coefficient</i>	<i>Std. Error</i>	<i>z</i>	<i>p-value</i>
const	-6.2133	1.50876	-4.1181	0.00004 ***
ronin	0.663315	0.187934	3.5295	0.00042 ***
woman	-0.764393	0.283776	-2.6937	0.00707 ***
national	1.28442	0.189	6.7958	<0.00001 ***
age	0.146821	0.0646961	2.2694	0.02324 **
age2	-0.00146675	0.00067634	-2.1687	0.03011 **

Mean dependent var	0.123941	S.D. dependent var	0.449459
Sum squared resid	176.8010	S.E. of regression	0.434151
McFadden R-squared	0.105531	Adjusted R-squared	0.090106
Log-likelihood	-347.9162	Akaike criterion	707.8325
Schwarz criterion	736.9332	Hannan-Quinn	718.9237

Overdispersion test: Chi-square(1) = 22.9934 [0.0000]

次に結果の読みとり方を解説していく。

まず、右端にアスタリスクがあるかどうか確認する。アスタリスクが 1 つの

ときは有意水準（p 値）が 10%であることを示している。アスタリスクが 2 つのときは有意水準が 5%、3 つの時は有意水準が 1%以下であることを示している。有意水準とは統計的仮説検定を行う場合に、帰無仮説を棄却するかどうかの判断基準である。有意水準 5%で検定を行うということは、帰無仮説が正しいにも関わらず帰無仮説を棄却する危険率が 5%であることを意味する。本稿の分析では、浪人ダミー、女性ダミー、国公立ダミーは有意水準が 1%以下であり、年齢と年齢 2 乗は有意水準が 5%以下なので「留年に影響しない」という帰無仮説は棄却される。

次に、「*Coefficient*」を見る。これは係数である。係数が正の値が出ると正の関係性を示していると考え。また、係数が負の値が出ると負の関係性を示していると考え。

そして、「*Std. Error*」は標準誤差である。標準偏差をデータの個数の平方根で割ることで算出される誤差のことである。

最後に、「z」は正規確率変数 $X \sim N(\mu, \sigma^2)$ を標準化したものである。

3. ポアソン回帰分析による結果

gretl の結果をわかりやすく表にする。

表 4 ポアソン回帰分析による結果

	係数	標準誤差	z 値	p 値	
浪人ダミー	0.6633	0.1879	3.5295	0.0004	***
女性ダミー	-0.7644	0.2838	-2.6937	0.0071	***
国公立ダミー	1.2844	0.189	6.7958	<0.00001	***
年齢	0.1468	0.0647	2.2694	0.0232	**
年齢 2 乗	-0.0015	0.0007	-2.1687	0.0301	**

注： * = 有意水準 10% * * = 有意水準 5% * * * = 有意水準 1%

4. 各変数の観測結果

以下は各変数の得られた結果を示している。

① 浪人ダミー

有意に正の影響を与えることが観測された。係数を見れば明らかとおり、浪人は現役合格者に比べて留年しやすいことがわかった。そして、浪人は約 0.0804 年多く留年することがわかった。この求め方は「*Coefficient*×*Mean dependent var*」である。これは女性ダミーと国公立ダミーも同様である。

② 女性ダミー

有意に負の影響を与えることが観測された。係数を見れば明らかとおり、男性は女性に比べて留年しやすいことがわかった。そして、男性は約 0.0912 年多く留年することがわかった。

③ 国公立ダミー

有意に正の影響を与えることが観測された。係数を見れば明らかとおり、国公立大学は私立大学に比べて留年しやすいことがわかった。国公立大学生は約 0.1536 年多く留年することがわかった。

④ 年齢

有意に正の影響を与えることが観測された。年齢が高くなるにつれて留年しやすいとわかった。

⑤ 年齢 2 乗

有意に負の影響を与えることが観測された。今回のデータでは約 49 歳が最も留年しやすいとわかった。この求め方は $f(x) = -0.0015x^2 + 0.1468x$ の最大値である。

5. OLS 回帰分析とポアソン回帰分析の結果を比較

回帰分析の *gretl* の手順は以下の通りである。

- (a) **Model** を選択し、**Ordinary Least Squares** を選択する
- (b) **Dependent variable** は被説明変数
- (c) **Independent variable** は説明変数
- (d) **OK** を押すと以下の結果が出る。

ポアソン回帰分析と同様に表5で表す。

表5 OLS回帰分析による結果

	係数	標準誤差	t 値	p 値	
浪人ダミー	0.0932	0.0314	2.9648	0.0031	***
女性ダミー	-0.0687	0.0332	-2.07	0.0387	**
国公立ダミー	0.1834	0.0317	5.7893	<0.00001	***
年齢	0.0138	0.0087	1.5882	0.1126	
年齢 2 乗	-0.0001	9.4865	-1.4711	0.1416	

この結果とポアソン回帰分析を比較すると、浪人ダミー、女性ダミー、国公立ダミーは有意であり、正と負の関係が同じという結果が出た。しかし、年齢に関しては有意ではないと観測された。

第6章 おわりに

留年する人は、浪人で国公立大学かつ男性である人物という結論に至った。本稿の分析では、留年と浪人の関連性の調査を目的で、浪人することで留年の可能性が高くなることがわかった。仮説では、浪人することで1年以上遅れているということでこれ以上社会に出るのが遅くならないようにするため留年しないと仮説をたてたが、本稿の分析で否定された。しかし、その原因が分かっていない。そこで、原因を推測する。浪人は入試という目的に向かって1年間勉強をし続けなければならない。しかし、この1年で勉強に対しての集中力が切れてしまい、大学に入学するとサークルやアルバイトなど課外活動に没頭してしまうことが原因だと推測する。

謝辞

〔二次分析〕に当たり、東京大学社会科学研究所附属社会調査・データアーカイブ研究センターSSJ データアーカイブから〔「2005 年社会階層と社会移動（SSM）全国調査」（2005SSM 研究会データ管理委員会）〕の個票データの提供を受けました。

参考文献

内田千代子 (2013) 『大学における休・退学, 留年学生に関する調査』第 3 報、2013 年 3 月発行。

北村行伸 (2009、第 11 章) 『ミクロ計量経済学入門』日本評論社。

Hiroshi Ono (2006) “Does examination hell pay off ? A cost-benefit analysis of “ronin” and college education in Japan, *“Economics of Education Review”* Vol.26, pp.271-284.