

第11章 カウントデータ分析：

医院への通院回数問題

1 はじめに

カウントデータとは、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ることで集めた統計（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をカウントデータ分析と呼んでいる。従ってここでのカウントとは事象のカウントを指しており、その事象は主体的に選択された結果、発生することもあれば、全く外生的に発生することもある。例えば、医者に診てもらいに病院に行く回数というのは、主体的に決められる事象であるが、地震の発生件数や台風の上陸件数は人智の及ばないところで決まる事象であろう。交通事故や火災はその中間に位置する事象であるが、保険のモラルハザード効果を加味すれば、ある程度内生的に発生するとも考えられる。

カウントデータの特徴としては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、いわゆるゼロ事象だということである。その結果、カウントデータの分布はゼロ周辺に集中し、右に裾を引いたような形をしている。また、当然予測されるように、事象の発生は個別の事情にも強く依存しているが、この個別事情は一般には観察できないので、回帰分析では誤差項の取り扱いが重要になってくる。

統計学上、稀にしか起こらない事象の発生確率はポアソン分布で表すことが多い。とりわけ、ポアソン分布が当てはまるような事象としてよく取り上げられるのは交通事故件数、大量生産の不良品数、倒産件数、火事発生件数などリスクや安全性に関する現象である¹。カウントデータ分析の基礎にもポアソン分布がある。

ちなみに、カウントデータ分析に関してはCameron and Trivedi (1998,2005) および Winkelmann(1997) が標準的な参考文献である。さらに勉強したい方は、これらの文献にあたってほしい。

¹最初のポアソン分布の適用例はBortkiewicz(1898)の「プロシア陸軍において馬に蹴られて死んだ兵士数」であるとされている。サッカーのワールドカップにおける各チームの得点もポアソン分布に従うことが知られている。

2 カウントデータ分析の枠組み

まず、ポアソン分布を理解するところから始めよう²。ポアソン分布 (Poisson Distribution) は次のように定義できる。

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

ここで $j!$ は j の階乗を意味する。この分布は平均値と分散値が等しい。すなわち、

$$E(y) = Var(y) = \lambda$$

ポアソン分布はこのように未知のパラメータ λ が決まれば全ての分布が決まる極めて簡便な分布である。このポアソン分布を用いたポアソン回帰モデルは次のように定義できる。すなわち、 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 $\lambda = \lambda(x'_i \beta)$ によって決定されると考え、つぎのように定義する。

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

ここで $E(y_i | x_i) = \exp(x'_i \beta)$ 、 $Var(y_i | x_i) = \exp(x'_i \beta)$ 。ポアソン回帰モデルは分散不均一となることを排除していない。

この式を次のような対数尤度関数に変換し、それを最尤法推定する。

$$\log L(\beta; y, x) = \sum_{i=1}^n \{y_i x'_i \beta - \exp(x'_i \beta) - \ln y_i!\}$$

これを最適化するための 1 階条件は次のようになる。

$$\sum_{i=1}^n (y_i - \exp(x'_i \beta)) x_i = 0$$

対数尤度関数が大局的に凸であれば、最適解は一意的に決まってくる。推定されたパラメータ $\hat{\beta}$ は一致推定量であり、漸近的に有効推定となる。

推定されたパラメータ $\hat{\beta}$ の解釈はいくつかできる。第一に、限界平均効果 (marginal mean effect) を用いて行う方法は次の通りである。

$$\frac{\partial E(y_i | \bar{x}_i)}{\partial x_{il}} = \exp(x'_i \beta) \beta_l = E(y_i | \bar{x}_i) \beta_l$$

²本節の議論は Cameron and Trivedi (1998, 2005; Chapter 20.)、Winkelmann and Bose (2006, pp.279-294) を参照している。

ここで \bar{x} は x の平均値を表している。もちろん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可能である³。この関係は次ぎように書き換えることができる。

$$\frac{\partial E(y_i | \bar{x}_i) / E(y_i | \bar{x}_i)}{\partial x_{il}} = \beta_l$$

これは x_{il} の限界的な変化に対して y の期待値が相対的にどれくらい変化するかを見たもので、すべての i に対して等しいものである。もし x_{il} が対数表示されているとすれば、 β_l は弾力性を表していることになる。ポアソン回帰モデルの特徴として、モデルの中に交叉項 $x_l x_m$ が含まれていないとしても、説明変数の交叉効果は残るといことがある。

$$\frac{\partial^2 E(y_i | x_i)}{\partial x_{il} \partial x_{im}} = \exp(x_i' \beta) \beta_l \beta_m = E(y_i | x_i) \beta_l \beta_m \neq 0$$

このことは、一般線形モデルではモデルの交叉項が含まれていない限り、ある説明変数の限界的な変化が他の説明変数の限界効果に影響を与えることは無いが、ポアソン回帰モデルではそれが起こることを示している。

第二の方法は限界確率効果 (marginal probability effect) を用いることである。

$$\frac{\partial P(y_i = j | x_i)}{\partial x_{il}} = P(y_i = j | x_i) [j - \exp(x_i' \beta)] \beta_l$$

限界確率効果の符号条件は $j - \exp(x_i' \beta)$ の符号条件に依存している⁴。

当然ながら、平均と分散が一致するような分布は現実のデータではそれほど見られないので、ポアソン分布を当てはめるケースは限定されている。具体的に、ポアソン回帰モデルの問題点としては次の点が指摘されている。第一に、ポアソン分布によるカウントデータ分析ではゼロ値を実際よりも少なめに推定してしまう傾向がある。第二に、カウントデータでは、実際には分散が平均より大きい (overdispersion) ことが多いことが知られている。

これらの問題点はポアソン分布では観測出来ない多様性 (unobserved heterogeneity) を扱っていないからであると判断されている。そこで、確率関数をポアソン分布と誤差項の積であると仮定し、さらに誤差項がガンマ分布に従っていると特定化すると、 y の条件付き分布は次のように表せる。

$$f(y_i | x_i) = \frac{\lambda_i^{y_i}}{y_i!} \int_0^\infty e^{-\lambda_i u_i} u_i^{y_i} \frac{\gamma^\theta}{\Gamma(\theta)} u_i^{\theta-1} e^{-\gamma u_i} du_i$$

ここで $\lambda_i = \exp(x_i' \beta)$ である。ここで誤差項を正規化し ($E(u_i | x_i) = 1$)、パラメータに $\gamma = \theta$ という制約をかけると上式は次のように転換できる。

³例えば、中位値 (median: 50%) や最頻値 (mode) で評価することも出来る。特に分布が対称分布でない場合には、どの値で評価をするかということが重要である。

⁴この条件の下では、限界確率効果は負から正へ単調に変化するか、逆に正から負に単調に変化する。これは single crossing property と呼ばれている。

$$\begin{aligned}
f(y_i|x_i) &= \frac{\lambda_i^{y_i}}{y_i!} \int_0^\infty e^{-\lambda_i u_i} u_i^{y_i} \frac{\theta^\theta}{\Gamma(\theta)} u_i^{\theta-1} e^{-\theta u_i} du_i \\
&= \frac{\lambda_i^{y_i}}{y_i!} \frac{\theta^\theta}{\Gamma(\theta)} \int_0^\infty e^{-(\lambda_i+\theta)u_i} u_i^{y_i+\theta-1} du_i \\
&= \frac{\lambda_i^{y_i}}{\Gamma(y_i+1)} \frac{\theta^\theta}{\Gamma(\theta)} \frac{\Gamma(y_i+\theta)}{(\lambda_i+\theta)^{y_i+\theta}} \\
&= \frac{\Gamma(y_i+\theta)}{\Gamma(y_i+1)\Gamma(\theta)} \left(\frac{\lambda_i}{\lambda_i+\theta}\right)^{y_i} \left(\frac{\theta}{\lambda_i+\theta}\right)^\theta
\end{aligned}$$

この確率関数を負の二項分布 (Negative Binominal Distribution)⁵関数とよび、期待値 $E(y_i|x_i) = \lambda_i$ で、分散 $Var(y_i|x_i) = \lambda_i(1 + \theta^{-1}\lambda_i)$ である。

負の二項分布 (NB) 回帰モデルは上式にさらに制約をかけて導かれる。まず、 $Var = (1 + \delta)\lambda_i$ 、 $\delta = \theta^{-1}\lambda_i$ と書き直し、主体 i 毎にパラメータ θ_i が変動することを許すようなモデルを NB1 モデルと呼び、分散関数は次のように与えられる。

$$Var(y_i|x_i) = (1 + \sigma^2) \exp(x_i'\beta)$$

このモデルを次のように対数尤度関数に変換し、それを最尤法推定すればいい。

$$\ln L(\theta, \beta) = \sum_{i=1}^n \left\{ \left(\sum_{j=0}^{y_i-1} \ln(j + \theta \exp(x_i'\beta)) \right) - \ln y_i! - (y_i + \theta \exp(x_i'\beta)) \ln(1 + \theta^{-1}) - y_i \ln \theta \right\}$$

最適化の 1 階条件は次のように表される。

$$\sum_{i=1}^n \left\{ \left(\sum_{j=0}^{y_i-1} \frac{\theta \lambda_i}{j + \theta \lambda_i} \right) x_i + \theta \lambda_i x_i \right\} = 0$$

$$\sum_{i=1}^n \theta^2 \left\{ - \left(\sum_{j=0}^{y_i-1} \frac{\lambda_i}{(j + \theta)} \right) - \theta^2 \lambda_i \ln(1 + \theta^{-1}) - \frac{\theta^{-1}}{1 + \theta^{-1}} + y_i \theta^{-1} \right\} = 0$$

ここで $\delta = 0$ を検定し、それが棄却できなければポアソン分布に従っていると判断され、棄却されればポアソン分布は採用できないことを意味する。

また、パラメータ θ は各主体共通で、 $\theta^{-1} = \sigma^2$ で定数であると仮定したモデルを NB2 モデルとよび、分散関数は次のように与えられる。

$$Var(y_i|x_i) = \exp(x_i'\beta) + \sigma^2 [\exp(x_i'\beta)]^2$$

⁵これは Greenwood and Yule (1920) によって初めて定式化された。

このモデルを対数尤度関数に変換し、それを最尤法推定すればいい。

$$\ln L(\theta, \beta) = \sum_{i=1}^n \left\{ \left(\sum_{j=0}^{y_i-1} \ln(j + \theta) \right) - \ln y_i! - (y_i + \theta) \ln(1 + \exp(x'_i \beta)) - y_i \ln \theta + y_i x'_i \beta \right\}$$

1階条件は次のように与えられる。

$$\sum_{i=1}^n \frac{y_i - \lambda_i}{1 + \lambda_i/\theta} x_i = 0$$

$$\sum_{i=1}^n \left\{ \theta^2 \left(\ln(1 + \lambda_i/\theta) - \sum_{j=0}^{y_i-1} \frac{1}{(j + \theta)} \right) + \frac{y_i - \lambda_i}{(1 + \lambda_i/\theta)/\theta} \right\} = 0$$

ここで $\sigma^2 \rightarrow 0$ であれば時、NB1 モデルも NB2 モデルもポアソンモデルに収束する。

NB モデルがポアソン回帰モデルで近似できるかどうかを検定するひとつの方法は次の式を OLS 推定して係数 α がゼロであるかどうかを見るというものである。

$$\frac{(y_i - \hat{\lambda}_i)^2 - y_i}{\hat{\lambda}_i} = \alpha \frac{g(\hat{\lambda}_i)}{\hat{\lambda}_i} + u_i$$

ここで $g(\hat{\lambda}) = \hat{\lambda}^2$ あるいは $g(\hat{\lambda}) = \hat{\lambda}$ で表される関数であり、 $\hat{\lambda}_i = \exp(x'_i \hat{\beta})$ とする。帰無仮説 $\alpha = 0$ では $Var(y_i | x_i) = \lambda_i$ となり、ポアソン分布が棄却できないことになる。これは先に述べた overdispersion かどうかを検定していることになる。

ポアソン回帰モデル、NB1 モデル、NB2 モデルの3つの推定方法のうち、ポアソン回帰モデルが選択されるかどうかは、上で述べた $\delta = 0$ が $\alpha = 0$ を検定すればよい。NB1 モデルと NB2 モデルの間の選択に関しては通常のカイ二乗分布に基づく尤度比検定を行うか、同じくカイ二乗分布に基づくワルド検定を行うのが一般的である。その他、カウントデータではゼロ値や1などの値の小さな数値に分布が集中するので、これがどれくらい推定できているかによって判断することも重要である。

3 医院への通院回数に関するカウントデータ分析

カウントデータを使った実証研究としては様々なものがあるが、本節では健康保険のタイプによって病人がどれくらいの頻度で医者に診てもらうかを調べた RAND 研究所が行った社会実験データを用いてみよう⁶。この実験は

⁶このデータに関しては Manning, Newhouse, Duan, Keeler and Leibowitz (1987)、Newhouse and the Insurance Experiment Group (1993)、Deb and Trivedi (2002) などが詳しいので参照されたい。

1974-1982年にかけて行われた医療に関する最大最長の社会実験であったと言われている。具体的には全米から4つの国勢調査区から都市サイズに応じて選ばれた6都市⁷に在住する8000人、2823家族に対して、14種類の健康保険プランをランダムに割り当てて⁸、そのプランの下で3-4年間継続的に調査を行ったものである。これらのプランは自己負担率が0%、25%、50%、95%、100% (individual deductible plan と呼ぶ) に分かれており⁹、負担額に関して、それぞれ上限が設定されている (最大1000ドルを絶対上限とし、家計総年入の5%、10%、15%)。上限を超える分に関しては全て医療保険が支払う¹⁰。また、追加的に40%の家計には金銭的刺激が医療処置にどのような影響を与えるかを見るために、想定外のボーナスを終了2年前に与えた。

ここで、健康保険プランをランダムに割り当てることによって、個人が自発的にプランを選ぶ場合、医療処置の必要に応じてプランを選ぶという内生的選択バイアスの問題を回避している。

この調査に含まれている変数は、医院への通院回数、医療費、個人属性、健康保険プランなどである。具体的には、MDU は医院への通院回数 (平均 2.861、標準偏差 4.505)。LC は $\ln(1 + \text{自己負担率})$ を表し、自己負担率は% 表示された数字を使う (平均 1.710、標準偏差 1.962)。IDP は自己負担率 100% の個人支払控除プラン (individual deductible plan) に属していれば 1、そうでなければ 0 のダミー変数 (平均 0.220、標準偏差 0.414)、LPI は $\ln(\max(1, \text{年間参加インセンティブ支払額}))$ (平均 4.709、標準偏差 2.6NB1NB⁹⁷)。FMDE はもし IDP=1 であれば 0 をとり、そうでなければ $\ln(\max(1, \text{MDE}/(0.01 \text{ 自己負担率})))$ 、ここで MDE とは最大医療負担額を表している (平均 3.153、標準偏差 3.641)。PHYSLIM は肉体的ハンディキャップを負っていれば 1 をとるダミー変数 (平均 0.124、標準偏差 0.322)。NDISEASE は慢性疾患数 (平均 11.244、標準偏差 6.742)。HLTHG は健康状態が良好 (最良に次ぐ分類) を表すダミー (平均 0.362、標準偏差 0.481)。HLTHF は健康状態が普通を表すダミー (平均 0.077、標準偏差 0.267)。HLTHP は健康状態が悪いことを表すダミーである (平均 0.015、標準偏差 0.121)。LINC は $\ln(\text{家計総所得})$ (平均 8.708、標準偏差 1.228)。LFAM は $\ln(\text{家族員数})$ (平均 1.248、標準偏

⁷Dayton (オハイオ州)、Seattle (ワシントン州)、Fitchburge (マサチューセッツ州)、Franklin County (マサチューセッツ州)、Charleston (南カロライナ州)、Georgetown County (南カロライナ州) の 6 都市である。

⁸このランダムな割り当てにおいては、健康保険プランの間に不平等がないように注意して分配された。つまり、実際にうける医療処置と健康状態、年齢、所得などの間にある相関をできるだけ小さくしようと心がけている。

⁹支払われる医療費はほとんど全ての内容をカバーしているが、自己負担率は医療内容に応じて変化する。例えば、入院や緊急医療の場合には 25%、歯科・精神医療の場合には 50% となる。あるプランでは一人当たり年間 150 ドル (家族あたり 450 ドル) 以下の医療支出に関して 95% の自己負担率を求めるものもある。

¹⁰日本では公的健康保険制度の下で、基本的には国民皆保険となっており、参加者によって制度的には大きな違いはないが、アメリカの場合、かなり複雑な健康保険プランに分散しており、それぞれが非線形価格付けを行っている。健康保険制度を考える際の問題意識も日本とアメリカではかなり違っていることに注意されたい。また、既存の健康保険を改革しようとした場合、アメリカの方が遥かに複雑な移行措置が必要になることが予想される。

差 0.539)。EDUCDEC は家計主の教育年数（平均 11.967、標準偏差 2.806）。AGE は年齢（平均 25.718、標準偏差 16.768）。FEMALE は女性ダミー（平均 0.517、標準偏差 0.500）。CHILD は 18 歳以下ダミー（平均 0.402、標準偏差 0.490）。FEMCHILD は FEMALE*CHILD で表されるダミー（平均 0.194、標準偏差 0.395）。BLACK は家計主が黒人であれば 1 をとるダミーを表している（平均 0.182、標準偏差 0.383）。

MDU で表される通院回数の分布は表 1 および図 1 で示されている。ここで明らかなのは 31% の人が一度も医者に診てもらっていないということである。それに対して 10 回を超えて通院している人は合計でも 5% に満たない。このことは通院頻度で表されている医療需要が一部の人に集中して発生しているということである。原則的には、たまたま不運な人が、病気にかかり通院を余儀なくされているのであって、健康保険に入っていることをもってモラルハザードを発生させて病気にかかった人や健康保険による自己負担率が低いから、必要なくとも通院するという行動は最小限に抑えられていると考えられる。そのことを確認する意味で、回帰分析を行った結果が表 2 にまとめられている。推定方法は前節で解説したポアソン回帰モデル、NB1 モデル、NB2 モデルの 3 種類である。

まず、全体の推定結果を概観すると、健康保険プランに関しては、自己負担率の影響を見た LC 変数は有意に負の係数になっており、自己負担率が高くなるほど通院頻度は低下することがわかる。しかし、係数の絶対値は小さいので、この効果はかなり低いと考えられる。IDP も同様に有意な負の効果を持っており、医療需要への強い価格効果があると解釈できる。健康状態に関する変数は最良の健康状態を基準にしているため、健康状態良好のケースを除いて、全て有意に正の効果を持っており、医療需要が高くなることを示している。個人属性に関しては、所得、教育、年齢、女性、年少者などは全て有意に正の効果を持っている。それに対して、家族員数、女性の年少者、黒人家計主では有意に負の効果を持つことが示されている。個人属性に関する結果の全てに対して経済学的な解釈を与えることは難しいが、ここでは個人属性をコントロールすることで、関心のある健康保険プランの効果や健康状態の効果により厳密に推定できていると考えていただきたい。

ポアソン回帰モデルと NB1 モデル、NB2 モデルの間の選択に関しては表 2 の下段にあるように尤度比検定の結果 $\alpha = 0$ 、 $\delta = 0$ はともに棄却され、ポアソン回帰モデルは、ここでは採択できないことがわかった。また、NB1 モデルと NB2 モデルの選択に関しては尤度比検定統計量を比べる限り、やや NB2 モデルの方が望ましいと言えそうである。しかし、先に述べたように、推定結果のフィットを比べる必要がある。フィットの結果はヒストグラムの形で図 2-4 に表示した。もともとのデータは図 1 のような分布をしており、ゼロ値に 31% も集中しているが、ポアソン回帰モデル、NB1 モデル、NB2 モデル全てがゼロ値を過小推定していることがわかる。また、ポアソン回帰モデルや

NB1 モデルでは最頻値は通院 4-5 回を予測しており、これも実際のデータからすると過大評価となっている。その中で、NB2 モデルでは最頻値を 3 回と予測しており、いちおう実際のデータに最も近い推定となっている。私の判断ではこのデータに関しては NB2 モデルを用いるのが適切であると思う。

4 おわりに

高齢化社会の中で確実に起こることは老人医療需要の急激な拡大である。これに対応するためには、老人医療の専門医を養成し、またその分野で働く看護師も大量に採用しなければならない。それと同時に、健康保険制度を見直して、現行で 3 歳から 69 歳までは医療費の自己負担率 30%、70 歳以上で 10%（一定の所得以上は 20%）となっている負担の構造を見直さなければならないだろう。これらの緊急の政策問題に対して、医療経済学の世界では一応の実証結果を出している。最近までの日本における老人医療に関する研究を手際よくまとめている湯田（2007）によると、患者が主体的に選択する医療需要については、総費用の増加は受診回数を低下させるが、その中で重要なのは、通院費用、待ち時間の機会費用、通院時間といった間接的な費用が需要を抑制しており、自己負担率の増加自体は医療需要に大きな影響を与えていないということである。従って、老人医療需要の抑制には、自己負担率の増加による金銭的な負担よりも、間接費用を通して受診を抑制するような施策が効果的であるということになる。緊急性のある医療を優先し、いつでもどこでもだれでも医療を受けることができるというフリーアクセス制の下で行われている非緊急の医療や患者と医師のミスマッチなどを改善すべきであるという主張がなされている。

アメリカではクリントン政権下での医療保険改革が実現しておらず、そもそも健康保険に加入していない人が 4500 万人もいると言われている。すでに書いたようにアメリカの健康保険は日本の制度よりはるかに複雑で国民皆保険を達成するのは極めて難しいと思われるが、大統領選挙の争点の一つになることは確実であろう。

このように、医療経済学ではミクロ計量経済学の手法が広範に使われており、分析結果が実際に人の命に関わるような極めて刺激的かつ実り多い分野である。多くの経済学徒がこの分野に参入されることを期待している。

今回はカウントデータ分析の基礎的な部分しか講義できなかった。Hurdle and Zero-Inflated カウントデータ・モデルやノンパラメトリック・カウントデータ分析に関しては、Cameron and Trivedi (1998)などを参照してほしい。

5 STATA コード

本章で用いたデータは P. Deb and P.K. Trivedi (2002) "The Structure of Demand for Medical Care: Latent Class versus Two-Part Models", *Journal of Health Economics*, 21, 601-625 で用いられたものである。データは www.econ.ucdavis.edu/faculty/cameron/に入っている `randdata.dta` である。また、本章で用いたプログラムは Cameron and Trivedi (2005, Chapter 20) で使われたものを踏襲している。

```
set more off
use randdata.dta, clear
/* educdec is missing for some observations*/
drop if educdec==.
/* rename variables*/
rename mdvis MDU
rename meddol MED
rename binexp DMED
rename lnmeddol LNMED
rename linc LINC
rename lfam LFAM
rename educdec EDUCDEC
rename xage AGE
rename female FEMALE
rename child CHILD
rename fchild FEMCHILD
rename black BLACK
rename disea NDISEASE
rename physlm PHYSLIM
rename hlthg HLTHG
rename hlthf HLTHF
rename hlthp HLTHP
rename idp IDP
rename logc LC
rename lpi LPI
rename fmde FMDE

/* Define the regressor list which in commands can refer to as $XLIST*/
global XLIST LC IDP LPI FMDE PHYSLIM NDISEASE HLTHG HLTHF
HLTHP LINC LFAM EDUCDEC AGE FEMALE CHILD FEMCHILD
BLACK
sum MDU $XLIST
```

```
/* 表 1 頻度分布*/
tabulate MDU
/* 図 1 ヒストグラム*/
hist MDU

/*カウントデータ回帰分析 表 2*/
/*ポアソン回帰*/
poisson MDU $XLIST
estimates store poisml
poisson MDU $XLIST, robust
estimates store poisrobust
predict MDUhat

poisson MDU $XLIST, cluster(zper)
estimates store poiscluster

/*NB 回帰分析 表 2*/
nbreg MDU $XLIST, dispersion(mean) /*NB2*/
nbreg MDU $XLIST, dispersion(mean) robust /*NB2 with robust z*/
estimates store nb2
predict MDUhat2

nbreg MDU $XLIST, cluster(zper)
estimates store nbcluster

nbreg MDU $XLIST, dispersion(constant) /*NB1*/
nbreg MDU $XLIST, dispersion(constant) robust /*NB1 with robust z*/
estimates store nb1
predict MDUhat1

hist MDUhat /*図 2*/
graph save "MDUhat.gph", replace
hist MDUhat1/*図 3*/
graph save "MDUhat1.gph", replace
hist MDUhat2/*図 4*/
graph save "MDUhat2.gph", replace
```

参考文献

- [1] 北村行伸 (2005) 『パネルデータ分析』、岩波書店
- [2] 湯田道生 (2007) 「高齢者医療需要における総費用弾力性の計測」、『日本経済研究』、No.57.(2007年7月)、pp.23-52.
- [3] Amemiya, T.(1985) *Advanced Econometrics*, Harvard University Press.
- [4] Bortkiewicz,L. von.(1898) *Das Gesetz de Kleinen Zahlen*, Leipzig, Teubner.
- [5] Cameron, A.C. and Trivedi, P.K.(1998) *Regression Analysis of Count Data*, Cambridge University Press.
- [6] Cameron, A.C. and Trivedi, P.K.(2005) *Microeconometrics: Methods and Applications*, Cambridge University Press.
- [7] Deb, Partha and Trivedi, Pravin K.(2002) “The Structure of Demand for Health Care: Latent Class versus Two-Part Models”, *Journal of Health Economics*, 21, pp.601-625.
- [8] Gourieroux, Christian and Jasiak, Joann.(2007) *The Econometrics of Individual Risk*, Princeton University Press.
- [9] Greenwood, M. and Yule, G.U.(1920) “An Inquiry into the Nature of Frequency Distributions of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or Repeated Accidents”, *Journal of the Royal Statistical Society A.*, 83, pp.255-279.
- [10] Maddala, G.S.(1983) *Limited-Dependent and Qualitative Variables in Economics*, Cambridge University Press.
- [11] Manning, Willard G., Newhouse, Joseph P., Duan, Naihua., Keeler, Emmett B., and Leibowitz, Arleen. (1987) “Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment”, *American Economic Review*, 77(3), pp.251-277.
- [12] Newhouse, Joseph P., and the Insurance Experiment Group (1993) *Free for All ? Lessons from the RAND Health Insurance Experiment*, Harvard University Press.
- [13] Winkelmann, R.(1997) *Count Data Models: Econometric Theory and Application to Labor Mobility*, Springer-Verlag.
- [14] Winkelmann, Rainer and Boes, Stefan.(2006) *Analysis of Microdata*, Springer.

-
- [15] Wooldridge, Jeffrey. M.(2003a) *Introductory Econometrics*, Thomson.
- [16] Wooldridge, Jeffrey. M.(2003b) *Econometric Analysis of Cross Section and Panel Data*, The MIT Press

表 1 医院への通院回数の頻度分布

通院回数	頻 度	パーセント	累積値
0	6,308	31.25	31.25
1	3,815	18.90	50.15
2	2,795	13.85	63.99
3	1,884	9.33	73.33
4	1,345	6.66	79.99
5	968	4.80	84.79
6	689	3.41	88.20
7	531	2.63	90.83
8	408	2.02	92.85
9	287	1.42	94.27
10	206	1.02	95.29
11	190	0.94	96.24
12	118	0.58	96.82
13	109	0.54	97.36
14	82	0.41	97.77
15	59	0.29	98.06
16	56	0.28	98.34
17	33	0.16	98.50
18	37	0.18	98.68
19	35	0.17	98.86
20	26	0.13	98.98
21	22	0.11	99.09
22	19	0.09	99.19
23	19	0.09	99.28
24	13	0.06	99.35
25	8	0.04	99.39
26	10	0.05	99.44
27	6	0.03	99.46
28	12	0.06	99.52
29	6	0.03	99.55
30	8	0.04	99.59
31	8	0.04	99.63
32	4	0.02	99.65
33	5	0.02	99.68
34	9	0.04	99.72
35	5	0.02	99.75
37	5	0.02	99.77
38	9	0.04	99.82
39	1	0.00	99.82
40	3	0.01	99.84
41	5	0.02	99.86
44	6	0.03	99.89
45	2	0.01	99.90
46	2	0.01	99.91
48	2	0.01	99.92
51	1	0.00	99.93
52	3	0.01	99.94
55	1	0	99.95
56	1	0	99.95
57	1	0	99.96
58	1	0	99.96
62	1	0	99.97
63	1	0	99.97
65	1	0	99.98
69	1	0	99.98
72	1	0	99.99
74	1	0	99.99
76	1	0	100.00
77	1	0.00	100.00
合 計	20,186	100	

表 2 カウントデータ分析 医院への通院確率(1)

Dependent Variable: MDU	Poisson		NB1		NB 2	
	Coefficient	Robust z-ratio	Coefficient	Robust z-ratio	Coefficient	Robust z-ratio
LC	-0.043	-2.84	-0.050	-3.23	-0.057	-5.27
IDP	-0.161	-5.77	-0.148	-4.86	-0.179	-8.66
LPI	0.013	2.91	0.016	3.57	0.014	4.43
FMDE	-0.021	-2.32	-0.021	-2.35	-0.013	-2.14
PHYSLIM	0.268	8.24	0.275	8.07	0.201	8.34
NDISEASE	0.023	13.49	0.026	15.32	0.020	16.13
HLTHG	0.039	1.70	0.007	0.27	0.038	2.32
HLTHF	0.253	5.89	0.237	5.43	0.207	6.43
HLTHP	0.522	6.97	0.426	6.20	0.520	8.34
LINC	0.083	5.99	0.085	7.42	0.075	7.27
LFAM	-0.130	-5.72	-0.123	-5.30	-0.097	-5.98
EDUCDEC	0.018	4.36	0.016	4.03	0.022	7.48
AGE	0.002	2.12	0.003	2.33	0.002	2.22
FEMALE	0.349	12.30	0.367	12.85	0.371	18.17
CHILD	0.336	8.32	0.306	7.13	0.323	10.71
FEMCHILD	-0.363	-8.21	-0.376	-8.40	-0.385	-12.58
BLACK	-0.680	-18.44	-0.710	-19.76	-0.721	-25.69
_cons	-0.190	-1.49	-0.207	-1.83	-0.129	-1.36
Number of observations	20186		20186		20186	
LR chi2(17)	13106.07		2828.01		3404.09	
Pseudo R2	0.098		0.032		0.039	
Log Likelihood	-60087.622		-42777.611		-42489.57	
alpha			1.182			
delta					3.460	
LR test of alpha=0:			chi2(01)=3.5e+04 Prob>chi2=0.000			
LR test of delta=0:					chi2(01)=3.5e+04 Prob>chi2=0.000	

図 1 医院への実際の通院回数のヒストグラム

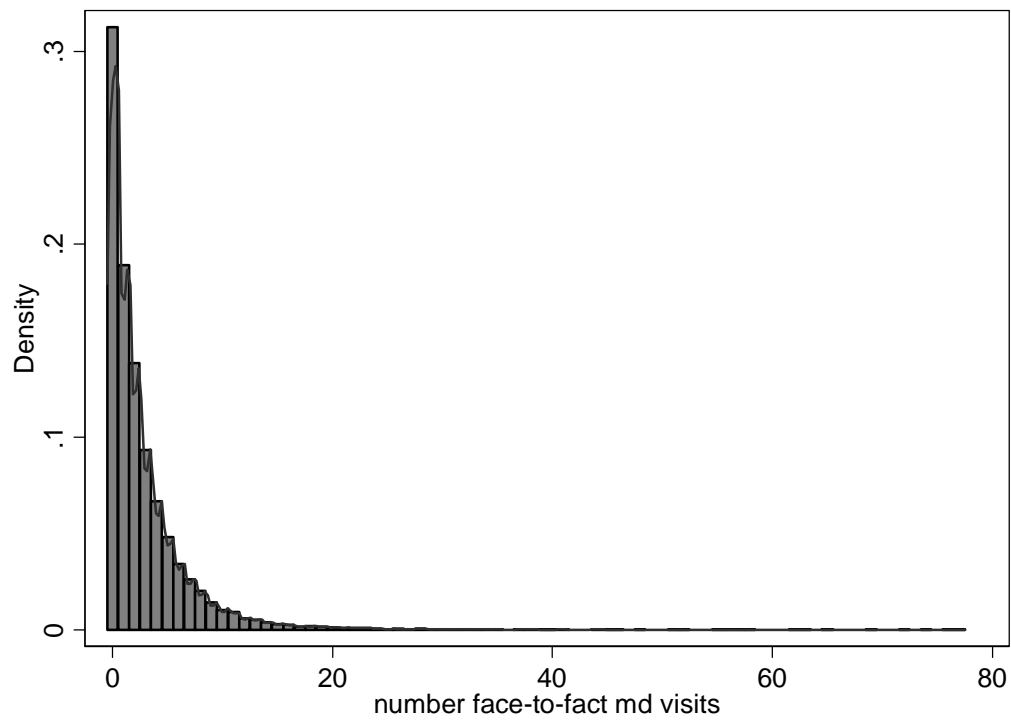


図2 ポアソン推定に基づく予測通院回数のヒストグラム

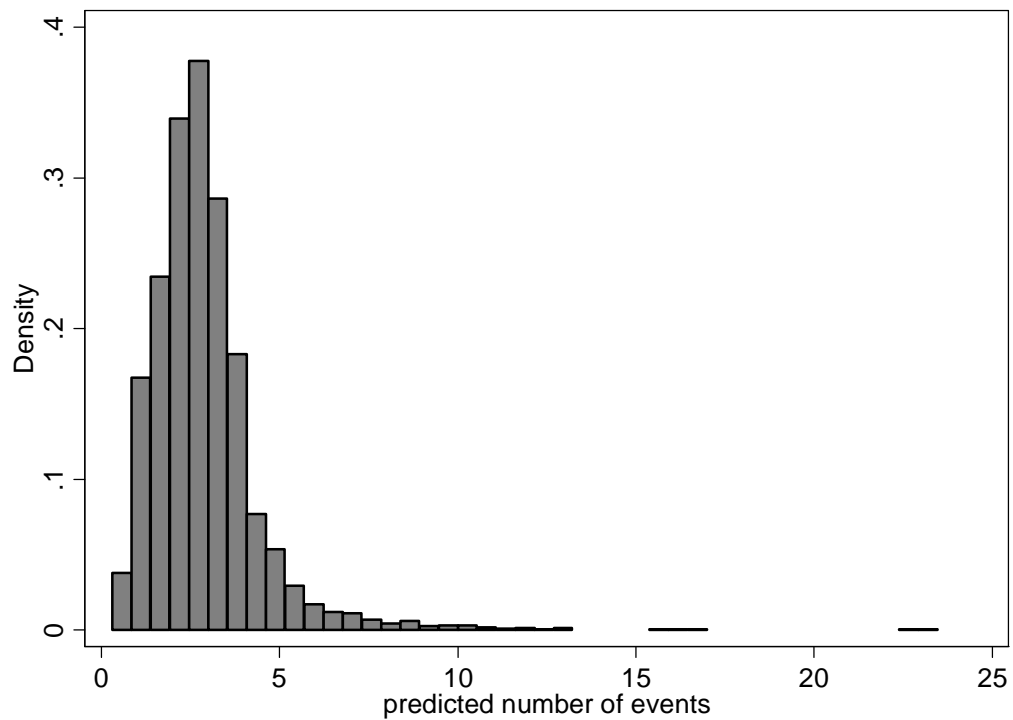


図 3 NB1 推定に基づく予測通院回数のヒストグラム

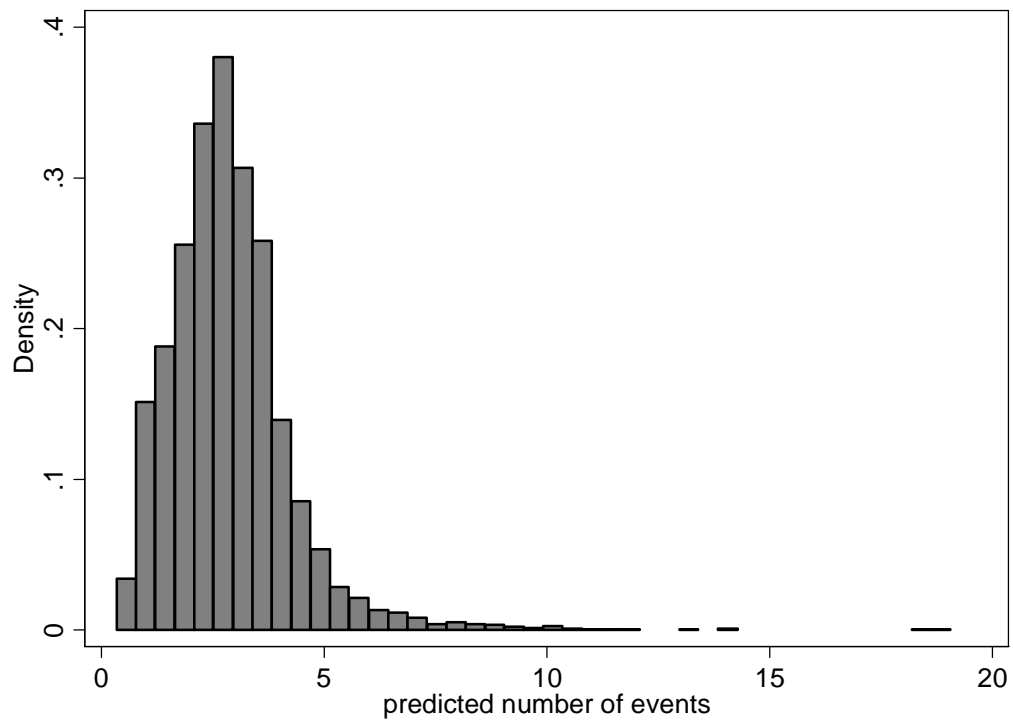


図 4 NB2 推定に基づく予測通院回数のヒストグラム

