

ロジスティック回帰分析の概要

2023年7月(作成)
倉谷 隆博

1. ロジスティック回帰式とシグモイド関数

ロジスティック回帰(Logistic Regression)分析は、重回帰分析では取り扱わない目的変数が 2 値の値をとる場合の分析を行う。結果として、0 ～ 1 の値をとる確率、割合などを予測する分析方法である。ロジスティック回帰式を、(式1)に示す。

$$\ln(p(Y=1)/(1-p(Y=1))) = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r \quad \dots (式1)$$

ln : 自然対数

Y : 2 値の値(例えば、0 か 1 の値)しかとらない確率変数(目的変数)

p : ある事象の発生確率、 $p(Y=1)$ とは、 $Y=1$ となる確率

β_i : 偏回帰係数 ($i = 1, 2, \dots, r$)

x_i : 説明変数

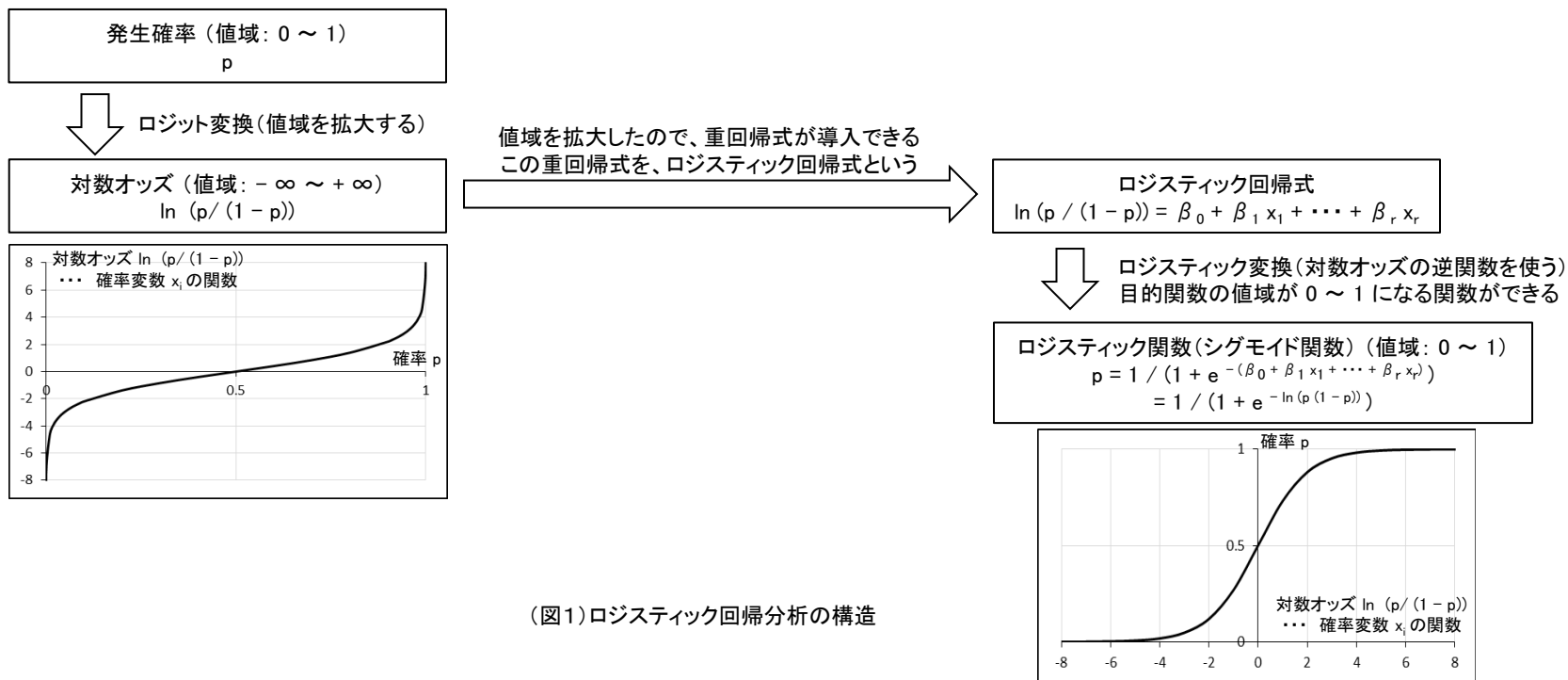
なお、 $p/(1-p)$ をオッズ(odds)、 $\ln(p/(1-p))$ を対数オッズと言う(ロジットとも言う)(付録1)。また、確率 p (値域: 0 ～ 1) から、対数オッズ $\ln(p/(1-p))$ (値域: $-\infty \sim +\infty$) への変換を、ロジット変換という。値域 0 ～ 1 の確率 p を、対数オッズの形にする(ロジット変換する)ことで、(式1)の値域が $-\infty \sim +\infty$ になり、(式1)を重回帰式として取り扱うことができるようになる。

(式1)を確率 p に関して解いて、(式2)を得る。この式をロジスティック関数(シグモイド関数)という。ロジスティック関数は、(図1)から分かるように 0 から 1 の間の値をとる関数になっている。従って、(式2)を使用することで、ある事象の発生確率を説明変数を使用して求めることができる。

$$(式1)より、p/(1-p) = e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_r x_r)} \rightarrow p = (1-p) e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_r x_r)} \rightarrow (1 + e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_r x_r)}) p = e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_r x_r)}$$

$$従って、p = e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_r x_r)} / (1 + e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_r x_r)}) = 1 / (1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_r x_r)}) \quad \dots (式2)$$

(式1)から(式2)への変換を、ロジスティック変換という。これら一連の関係を、(図1)にまとめておく。



(図1)ロジスティック回帰分析の構造

2. ロジスティック回帰式の偏回帰係数の推定

ロジスティック回帰式の偏回帰係数の推定には最尤法(別途資料「統計的推定の手順」を参照)を用いる。最尤法を使って、偏回帰係数の最尤推定値を求める。

1) 対数尤度関数を求める。

何かを実施したときに起こる結果が2つしかない試行をベルヌーイ試行と言う(別途資料「いろいろな確率分布」を参照)。

このような試行を n 回(独立であるとする)繰り返した場合、 $y_i = 1$ となる確率を p_i とすれば、 $y_i = 0$ となる確率は $1 - p_i$ である。従って、尤度は、

$$\text{尤度 } L(\beta | X) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1 - y_i} \quad \dots \text{ (式3)}$$

積を和の形にして計算しやすくするために対数尤度に変換する。

$$\text{対数尤度 } \ell(\beta | X) = \ln L(\beta | X) = \ln \left(\prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1 - y_i} \right) = \sum_{i=1}^n (y_i \ln p_i + (1 - y_i) \ln (1 - p_i)) \quad \dots \text{ (式4)}$$

2) 最尤推定量を求める。

この(式4)に、(式2) $p_i = 1 / (1 + e^{-(\beta_0 + \beta_1 x_{i1} + \dots + \beta_r x_{ir})})$ を代入して、 $\ln L(\beta | X)$ の値が最大になる β を求める。つまり、

$$\partial / \partial \beta [\ell(\beta | X)] = 0 \quad \dots \text{ (式5)}$$

を満足する β を求める。こうして求めた β が最尤推定量 $\hat{\beta}$ になる。

なお、(式2)に最尤推定量を代入して求めた、目的変数 y_i が 1 となる確率 p_i の推定値は、(式6)で表される。

$$\hat{p}_i = 1 / (1 + \exp(-(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_r x_{ir}))) \quad \dots \text{ (式6)}$$

(式6)がロジスティック回帰分析のコアになる式である。

(式6)で求まる $\hat{p}_i$ は、閾値を設けて、例えば、合格/不合格に仕分けるなどのための**判定スコア**として使用される。例えば、合格/不合格を推定する場合、閾値を 0.5 であるとすれば、推定確率 $\hat{p}_i = 0.7$ は $0.7 > 0.5$ なので「合格」とであると推定することができる。

しかし、(式2)を代入した(式5)の偏微分方程式は複雑になり、最尤推定量 $\hat{\beta}$ を解析的に求めることができない。そこで、微分方程式を近似的に解くニュートン・ラプソン法が使用される。ただ、実務的には、Excel の「ソルバー」機能を使用して(ニュートン・ラプソン法を使わないで)、最尤推定量 $\hat{\beta}$ を求めることができる。

偏回帰係数の算出例(後述)に、ソルバー機能を使用した例を載せる。

3) 偏回帰係数の解釈

(式1)において、 x_i を 1 増加した $x_i + 1$ の場合の確率を q とすれば、 $\ln(q(y=1) / (1 - q(y=1))) = \beta_0 + \beta_1 x_i + \dots + \beta_i (x_i + 1) + \dots + \beta_r x_r \quad \dots \text{ (式7)}$

(式7) - (式2)を求めると、

$$\ln(q / (1 - q)) - \ln(p / (1 - p)) = (\beta_0 + \beta_1 x_i + \dots + \beta_i (x_i + 1) + \dots + \beta_r x_r) - (\beta_0 + \beta_1 x_i + \dots + \beta_i x_i + \dots + \beta_r x_r)$$

$$\ln\{(q / (1 - q)) / (p / (1 - p))\} = \beta_i$$

$$(q / (1 - q)) / (p / (1 - p)) = e^{\beta_i} \quad \dots \text{ (式8)}$$

従って、説明変数 x_i を 1 増加させる前のオッズに対する、 x_i を 1 増加させた後のオッズの比が、(式8)で表されることが分かる(オッズについては、付録1)。

説明変数が2値を取る変数、例えば性別である場合、女性 = 0、男性 = 1 であれば、男性の場合は女性の場合に対して、性別の変数が、女性 = 0 $\rightarrow$ 男性 = 1 で 1 増加になり、この場合、オッズ比は、 e^{β_i} 倍になる。

3. 偏回帰係数の推定(算出例)

(表1)の色付きの部分のデータを使用して、ロジスティック回帰式(式1)の偏回帰係数を算出する。

$n = 15$ 、目的変数 $\mathbf{y} = (y_1 \ y_2 \ \cdots \ y_n)^T$ は、2値(ここでは、0 or 1)の値を取る確率変数である。説明変数は、 $\mathbf{x}_1 = (x_{11} \ x_{21} \ \cdots \ x_{n1})^T$ 、 $\mathbf{x}_2 = (x_{12} \ x_{22} \ \cdots \ x_{n2})^T$ であり、 $\mathbf{x}_1$ は量的変数、 $\mathbf{x}_2$ は2値(ここでは、0 or 1)の値をとる質的変数とした。

この表の色付きの部分に、(式1)と(式2)、そして対数尤度の計算に使用する(式4)の各項を計算する部分を追加して表を完成させる。

この表の(4)の各項の総和が、最適化計算の目標値である(式4)の対数尤度になる。

この対数尤度を最大にする偏回係数 $\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2]$ を、(付録2)に示す方法で、Excel のソルバーを使用して求める。

結果は、偏回帰係数の最尤推定量は、 $\hat{\beta}_0 = -8.42004$ ， $\hat{\beta}_1 = 0.49825$ ， $\hat{\beta}_2 = -1.56435$ で、このとき、最大対数尤度は、 -4.18496 であった。

(式1)のロジスティック回帰式は、 $\ln \hat{p}_i / (1 - \hat{p}_i) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} = -8.42004 + 0.49825 x_{i1} - 1.56435 x_{i2}$ で表される。

また、(式6)の判定スコアは、 $\hat{p}_i = 1 / (1 + \exp(-(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}))) = 1 / (1 + \exp(-(-8.42004 + 0.49825 x_{i1} - 1.56435 x_{i2})))$ で表される。

例えば、 $x_{i1} = 23$, $x_{i2} = 1$ のとき、 $y_i = 1$ になる確率は、 $\hat{p}_i = 1 / (1 + \exp(-(-8.42004 + 0.49825 * 23 - 1.56435 * 1))) = 0.81387$ となる。

また、それぞれの説明変数 x_1, x_2 が1増加した場合のオッズ比は、以下の通りである。

$$\exp(\hat{\beta}_1) = e^{0.49825} = 1.64584, \quad \exp(\hat{\beta}_2) = e^{-1.56435} = 0.20922$$

なお、(表1)には、偏回帰係数の検定(後述)に使用する下記の(式9)の計算結果も併せて載せておく。

$$\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) / (1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}))^2 \quad \dots \quad (式9)$$

$$= \exp(-8.42004 + 0.49825 x_{i1} - 1.56435 x_{i2}) / (1 + \exp(-8.42004 + 0.49825 x_{i1} - 1.56435 x_{i2}))^2$$

(表1)ロジスティック回帰分析で使用するデータ

番号	目的変数 y_i	説明変数 x_{i1}	説明変数 x_{i2}	(式1)	(式2)	(式4)の各項	(式9)
1	1	25	1	2.47193	0.92215	-0.08105	0.07179
2	1	22	0	2.54153	0.92700	-0.07580	0.06767
3	1	26	1	2.97019	0.95121	-0.05002	0.04641
4	1	14	0	-1.44450	0.19085	-1.65627	0.15443
5	0	6	0	-5.43052	0.00436	-0.00437	0.00434
6	0	2	1	-8.98789	0.00012	-0.00012	0.00012
7	0	5	0	-5.92878	0.00265	-0.00266	0.00265
8	0	10	0	-3.43751	0.03114	-0.03164	0.03017
9	0	19	0	1.04677	0.74015	-1.34767	0.19233
10	1	20	0	1.54502	0.82419	-0.19335	0.14490
11	0	15	0	-0.94624	0.27964	-0.32801	0.20144
12	1	30	1	4.96320	0.99306	-0.00697	0.00689
13	1	24	1	1.97368	0.87801	-0.13010	0.10711
14	0	15	1	-2.51060	0.07512	-0.07809	0.06948
15	0	17	1	-1.51409	0.18033	-0.19886	0.14781
対数尤度：(式4)の各項の和						-4.18496	

4. 個別の偏回帰係数の検定と信頼区間

(1) 個別の偏回帰係数の検定

(付録5)に示すように、個別の偏回帰係数の検定には偏回帰係数の分散を使用する。この偏回帰係数の分散(分散共分散行列)は、フィッシャー情報量(付録3)を使って表すことができる(付式12)。

そこで、まずロジスティック回帰式のフィッシャー情報量を求める。

(式4)より、対数尤度 $\ell(\beta | X) = \ln L(\beta | X) = \ln \left(\prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1 - y_i} \right) = \sum_{i=1}^n (y_i \ln p_i + (1 - y_i) \ln (1 - p_i)) = \sum_{i=1}^n (\ln (1 - p_i) + y_i \ln (p_i / (1 - p_i)))$

ここで、表記を簡便化するために、 $\Sigma = \sum_{i=1}^n (\beta_0 + \beta_1 x_{i1} + \dots + \beta_r x_{ir})$ とおき、

上式に、(式2) $p_i = \exp \Sigma / (1 + \exp \Sigma)$ を代入する。

$$\ell(\beta | X) = \sum_{i=1}^n \{ \ln (1 - (\exp \Sigma / (1 + \exp \Sigma))) + y_i \ln ((\exp \Sigma / (1 + \exp \Sigma)) / (1 - (\exp \Sigma / (1 + \exp \Sigma)))) \}$$
$$= \sum_{i=1}^n \{ \ln (1 / (1 + \exp \Sigma)) + y_i \ln ((\exp \Sigma / (1 + \exp \Sigma)) / (1 / (1 + \exp \Sigma))) \} = \sum_{i=1}^n \{ \ln (1 / (1 + \exp \Sigma)) + y_i \ln (\exp \Sigma) \} = \sum_{i=1}^n \{ -\ln (1 + \exp \Sigma) + y_i \Sigma \}$$

また、以下の式変形では、関数の微分の公式を使用する。(公式) $\cdot(f(x) g(x))' = f'(x) g(x) + f(x) g'(x)$ $\cdot(\ln f(x))' = f'(x) / f(x)$ 但し、' は微分 d/dx を表す。

$(\Sigma)' = \partial (\Sigma) / \partial \beta_m = \partial (\sum_{i=1}^n (\beta_0 + \beta_1 x_{i1} + \dots + \beta_r x_{ir})) / \partial \beta_m = x_{im}$ である。

従って、 $\partial / \partial \beta_m (\ell(\beta | X)) = \sum_{i=1}^n \{ -(1 + \exp \Sigma)' / (1 + \exp \Sigma) + y_i (\Sigma)' \} = \sum_{i=1}^n \{ -(\Sigma)' \exp \Sigma / (1 + \exp \Sigma) + y_i (\Sigma)' \} = \sum_{i=1}^n \{ -x_{im} \exp \Sigma / (1 + \exp \Sigma) + y_i x_{im} \}$
$$= \sum_{i=1}^n x_{im} (y_i - \exp \Sigma / (1 + \exp \Sigma))$$

さらに、 $\partial^2 / \partial \beta_d \partial \beta_m (\ell(\beta | X)) = \partial / \partial \beta_d (\sum_{i=1}^n x_{im} (y_i - \exp \Sigma / (1 + \exp \Sigma))) = 0 - \sum_{i=1}^n \partial / \partial \beta_d (x_{im} (\exp \Sigma) / (1 + \exp \Sigma))$
$$= -(\sum_{i=1}^n x_{im} ((\exp \Sigma)' / (1 + \exp \Sigma) - (\exp \Sigma) (1 / (1 + \exp \Sigma))')) = -\sum_{i=1}^n x_{im} (x_{id} / (1 + \exp \Sigma) + (\exp \Sigma) (\exp \Sigma)' / (1 + \exp \Sigma)^2)$$
$$= -\sum_{i=1}^n x_{im} (x_{id} / (1 + \exp \Sigma) + x_{id} (\exp \Sigma) / (1 + \exp \Sigma)^2) = -\sum_{i=1}^n x_{im} x_{id} (1 / (1 + \exp \Sigma) + \exp \Sigma / (1 + \exp \Sigma)^2) = -\sum_{i=1}^n x_{im} x_{id} (\exp \Sigma) / (1 + \exp \Sigma)^2$$

ここで、 $X = [1 \ x_1 \ x_2 \ \dots \ x_m \ \dots \ x_d \ \dots \ x_r]$, $x_m = [x_{1m} \ x_{2m} \ \dots \ x_{nm}]^T$, $x_d = [x_{1d} \ x_{2d} \ \dots \ x_{nd}]^T$ と置けば、 $x_m^T x_d = \sum_{i=1}^n x_{im} x_{id}$ となる。

また、対角行列 $W = \text{diag} (\exp \sum_{j=1}^r x_{1j} / (1 + \exp \sum_{j=1}^r x_{1j})^2, \exp \sum_{j=1}^r x_{2j} / (1 + \exp \sum_{j=1}^r x_{2j})^2, \dots, \exp \sum_{j=1}^r x_{nj} / (1 + \exp \sum_{j=1}^r x_{nj})^2)$ を新たに導入する(diag は対角行列を表す)。

フィッシャー情報量は、(付式7)より、 $I_r(\beta) = -E [\partial^2 / \partial \beta^2 (\ln L(\beta | X))]$ であり、また、 $x_m^T W x_d$ がスカラーで $E [x_m^T W x_d] = x_m^T W x_d$ であることを利用して、

$I_r(\beta_j) = -E [\partial^2 / \partial \beta_j^2 (\ln L(\beta_j | X))] = \partial^2 / \partial \beta_j \partial \beta_j (\ell(\beta | X)) = -E [-\sum_{i=1}^n x_{ij} x_{ij} (\exp \Sigma) / (1 + \exp \Sigma)^2] = E [x_j^T W x_j] = x_j^T W x_j \quad \dots (式10)$

よって、このフィッシャー情報量を使用して、(付式16)より偏回帰係数ごとの標準誤差(Standard Error)が求まる。

$SE(\hat{\beta}_j) = (I_r(\hat{\beta})^{-1}_{jj})^{1/2} = (I_r(\beta_j)^{-1})^{1/2} = ((x_j^T W x_j)^{-1})^{1/2} = ((X^T W X)^{-1}_{jj})^{1/2} \quad \dots (式11)$

(付式15)に(式11)の標準誤差を代入した(式12)を使用して、ロジスティック回帰分析のワルド検定(自由度 1 のカイニ乗検定)を行う。

$T = (\hat{\beta}_j / SE(\hat{\beta}_j))^2 = (\hat{\beta}_j / ((X^T W X)^{-1}_{jj})^{1/2})^2 = \hat{\beta}_j^2 / (X^T W X)^{-1}_{jj} \sim \chi^2(1) \quad \dots (式12)$

(2) 個別の偏回帰係数の信頼区間(別途資料「統計的推定の手順」を参照)

(付式13)と(付式16)より、 $(\hat{\theta}_j - \hat{\theta}_j^0) / \{ Q(\hat{\theta})_{jj} \}^{1/2} = (\hat{\theta}_j - \hat{\theta}_j^0) / SE(\hat{\theta}_j) \sim N(0, 1)$ であり、

$\hat{\theta}_j = \hat{\beta}_j, \hat{\theta}_j^0 = \beta_j$ において、 $(\hat{\beta}_j - \beta_j) / SE(\hat{\beta}_j) \sim N(0, 1) \quad \dots (式13)$

確率 $1 - \alpha =$ 信頼度 $100(1 - \alpha) \%$ であり、信頼度 95% とした場合には、 $\alpha = 0.05$ である。

偏回帰係数 β_j が入る信頼区間の上限は標準正規分布の上側 $\alpha/2$ 点 $= z_{\alpha/2}$, 下限は標準正規分布の下側 $\alpha/2$ 点 $= z_{1-\alpha/2} = -z_{\alpha/2}$ である。

$z_{\alpha/2} = z_{0.025} = \text{NORM.INV}(1 - 0.025, 0, 1) = 1.96$ であるので、上側 2.5 % 点 $= z_{0.025} = 1.96$ また、下側 2.5 % 点 $= -z_{0.025} = -1.96$ である。

従って、(式13)を使用して、偏回帰係数 β_j の 95% 信頼区間が得られる。

$-1.96 \leq (\hat{\beta}_j - \beta_j) / SE(\hat{\beta}_j) \leq 1.96$ より、 $\hat{\beta}_j - 1.96 SE(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + 1.96 SE(\hat{\beta}_j) \quad \dots (式14)$

なお、オッズ比 e^{β_j} の 95% 信頼区間は、 $\exp(\hat{\beta}_j - 1.96 SE(\hat{\beta}_j)) \leq \exp(\beta_j) \leq \exp(\hat{\beta}_j + 1.96 SE(\hat{\beta}_j)) \quad \dots (式15)$

5. 個別の偏回帰係数の検定と信頼区間(算出例)

(表1)のデータを使用したロジスティック回帰式の偏回帰係数の検定を行う。

(1) 偏回帰係数の標準誤差を求める

(式11)を使用して、標準誤差を求める。 行列計算には、Excel の、TRANSPOSE 関数、MMULT 関数、MINVERSE 関数を使用する。

① 行列 X (15 × 3) を示す。

	x_1	x_2
1	25	1
1	22	0
1	26	1
1	14	0
1	6	0
1	2	1
1	5	0
1	10	0
1	19	0
1	20	0
1	15	0
1	30	1
1	24	1
1	15	1
1	17	1

② 対角要素以外の要素は全て 0 で、対角要素が(表1)の(式9)の値になる対角行列 W (15 × 15) を用意する。

	対角要素
w_{11}	0.07179
w_{22}	0.06767
w_{33}	0.04641
w_{44}	0.15443
w_{55}	0.00434
w_{66}	0.00012
w_{77}	0.00265
w_{88}	0.03017
w_{99}	0.19233
$w_{10\ 10}$	0.14490
$w_{11\ 11}$	0.20144
$w_{12\ 12}$	0.00689
$w_{13\ 13}$	0.10711
$w_{14\ 14}$	0.06948
$w_{15\ 15}$	0.14781

③ (式11)の $X^T W X$ (15 × 15) を計算する。

$X^T W X$	1.24754	22.89960	0.44962
	22.89960	441.46551	9.33409
	0.44962	9.33409	0.44962

④ (式11)を使って、偏回帰係数ごとの標準誤差を求める。

$(X^T W X)^{-1}$	18.19286	-0.99639	2.49225
	-0.99639	0.05861	-0.22031
	2.49225	-0.22031	4.30552

$SE(\hat{\beta}_0)$	$(18.19286)^{1/2} = 4.26531$
$SE(\hat{\beta}_1)$	$(0.05861)^{1/2} = 0.24209$
$SE(\hat{\beta}_2)$	$(4.30552)^{1/2} = 2.07497$

(2) ワルド検定のための検定統計量を求める。

ワルド検定のための検定統計量は、(式12)で与えられる。

$\hat{\beta}_0 = -8.42002$, $\hat{\beta}_1 = 0.49825$, $\hat{\beta}_2 = -1.56435$ が得られているので、

① $\hat{\beta}_0$ の検定統計量 = $(\hat{\beta}_0 / SE(\hat{\beta}_0))^2 = (-8.42004 / 4.26531)^2 = 3.89697$

② $\hat{\beta}_1$ の検定統計量 = $(\hat{\beta}_1 / SE(\hat{\beta}_1))^2 = (0.49825 / 0.24209)^2 = 4.23589$

③ $\hat{\beta}_2$ の検定統計量 = $(\hat{\beta}_2 / SE(\hat{\beta}_2))^2 = (-1.56435 / 2.07497)^2 = 0.56839$

(4) 採択された偏回帰係数 β_0 と β_1 の信頼区間を求める。

(式14)を使用して、

β_0 の 95 % 信頼区間は、 $-8.42004 - 1.96 * 4.26531 \leq \beta_0 \leq -8.42004 + 1.96 * 4.26531 \rightarrow -16.78005 \leq \beta_0 \leq -0.06003$ (β_0 の値は負の範囲にある)

β_1 の 95 % 信頼区間は、 $0.49825 - 1.96 * 0.24209 \leq \beta_1 \leq 0.49825 + 1.96 * 0.24209 \rightarrow 0.02375 \leq \beta_1 \leq 0.97275$ (β_1 の値は正の範囲にある)

なお、 β_2 の 95 % 信頼区間は、 $-1.56435 - 1.96 * 2.07497 \leq \beta_2 \leq -1.56435 + 1.96 * 2.07497 \rightarrow -5.63129 \leq \beta_2 \leq 2.502591$ (β_2 の値は正負をまたぐ、棄却できない)

また、オッズ比 $e^{\hat{\beta}_1}$ の 95 % 信頼区間は(式15)より、 $\exp(0.49825 - 1.96 * 0.24209) \leq e^{\beta_1} \leq \exp(0.49825 + 1.96 * 0.24209) \rightarrow 1.02404 \leq e^{\beta_1} \leq 2.64520$

(3) ワルド検定を行う(別途資料「統計的検定の手順」を参照、Excel 関数を使う)。

① $\hat{\beta}_0$ の検定統計量の p 値 = CHDIST(3.89697, 1) = 0.04837 < 0.05

② $\hat{\beta}_1$ の検定統計量の p 値 = CHDIST(4.23589, 1) = 0.03958 < 0.05

③ $\hat{\beta}_2$ の検定統計量の p 値 = CHDIST(0.56839, 1) = 0.45090

従って、①の帰無仮説($\hat{\beta}_0 = 0$)と②の帰無仮説($\hat{\beta}_1 = 0$)は棄却され、対立仮説

($\hat{\beta}_0 \neq 0$, $\hat{\beta}_1 \neq 0$)、つまり、 $\hat{\beta}_0 = -8.42004$, $\hat{\beta}_1 = 0.49825$ が各々採択される。

一方、③の帰無仮説($\hat{\beta}_2 = 0$)は棄却できない。 $\hat{\beta}_2 = -1.56435$ は採択されない。

6. 説明変数が1つときのロジスティック回帰分析

ロジスティック分析を俯瞰するため、説明変数が1つときのロジスティック回帰分析の例を載せる。
(表2)のデータを対象にロジスティック回帰式の偏回帰係数の値を Excel のソルバーの機能を使用して求めると、 $\beta_0 = -7.81528$, $\beta_1 = 0.44895$ であった。

(表2)ロジスティック回帰分析で使用するデータ

番号	目的変数 y_i	説明変数 x_{i1}	(式1)	(式2)	(式4)の各項	(式9)
1	0	11	-2.87684	0.05331	-0.05478	0.05631
2	0	12	-2.42789	0.08107	-0.08455	0.08822
3	0	13	-1.97894	0.12143	-0.12946	0.13822
4	0	14	-1.52999	0.17800	-0.19601	0.21654
5	1	15	-1.08104	0.25331	-1.37314	0.33924
6	1	17	-0.18314	0.45434	-0.78890	0.83265
7	0	18	0.26581	0.56606	-0.83486	1.30449
8	0	20	1.16371	0.76201	-1.43551	3.20179
9	1	21	1.61266	0.83378	-0.18179	5.01614
10	1	22	2.06161	0.88712	-0.11978	7.85861
11	1	23	2.51056	0.92488	-0.07809	12.31181
12	1	24	2.95951	0.95071	-0.05055	19.28850
13	1	25	3.40846	0.96797	-0.03256	30.21863
14	1	26	3.85741	0.97931	-0.02090	47.34249
15	1	27	4.30636	0.98670	-0.01339	74.16986
対数尤度: (式4)の各項の和					-5.39427	

対角要素

w_{11}	0.05047
w_{22}	0.07450
w_{33}	0.10669
w_{44}	0.14631
w_{55}	0.18914
w_{66}	0.24792
w_{77}	0.24564
w_{88}	0.18135
w_{99}	0.13859
$w_{10\ 10}$	0.10014
$w_{11\ 11}$	0.06948
$w_{12\ 12}$	0.04686
$w_{13\ 13}$	0.03101
$w_{14\ 14}$	0.02026
$w_{15\ 15}$	0.01313

$W =$

以下、5. と同じ手順で偏回帰係数ごとの標準誤差を求める。

$X^T W X =$

1.66147	29.47706
29.47706	545.84792

$(X^T W X)^{-1} =$

14.35882	-0.77541
-0.77541	0.04371

$SE(\hat{\beta}_j) =$

$SE(\hat{\beta}_0)$	$(14.35882)^{1/2} = 3.78930$
$SE(\hat{\beta}_1)$	$(0.04371)^{1/2} = 0.20906$

次いで、ワルド検定のための検定統計量を求める。

$\hat{\beta}_0$ の検定統計量 $= (\hat{\beta}_0 / SE(\hat{\beta}_0))^2 = (-7.81528 / 3.78930)^2 = 4.25374$

$\hat{\beta}_1$ の検定統計量 $= (\hat{\beta}_1 / SE(\hat{\beta}_1))^2 = (0.44895 / 0.20906)^2 = 4.61163$

p 値 を求め、検定を行う。

$\hat{\beta}_0$ の検定統計量の p 値 $= \text{CHDIST}(4.25374, 1) = 0.03916 < 0.05$

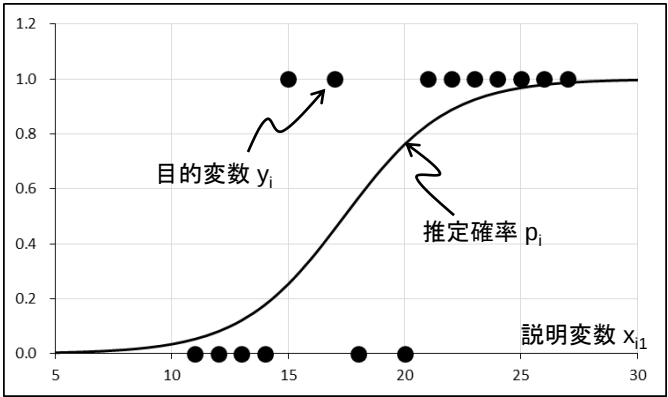
$\hat{\beta}_1$ の検定統計量の p 値 $= \text{CHDIST}(4.23589, 1) = 0.03176 < 0.05$

いずれも、帰無仮説(偏回帰係数の値 = 0)を棄却でき、求めた偏回帰係数の値を採用することになる。

(図2)は、説明変数と目的変数をプロットしたグラフに、求めた偏回帰係数を使用して推定した確率を合わせて載せたグラフである。

(式6)より、推定した確率 $p_i = 1 / (1 + \exp(-(-7.81528 + 0.44895 * x_{i1})))$

(図2)より、0 と 1 の2値の値しかとらない目的変数の値が共存する説明変数の区間において、推定した確率が 0 と 1 の間の滑らかな値を取っていることが分かる。



(図2)実データと推定確率

付録1. 確率(Probability)とオッズ(Odds)

(1) 割合(確率)とオッズの定義

全体の数 = 対象の数 + 対象以外の数 としたとき(付図1)、

割合の定義式 = 対象の数 / 全体の数 ... (付式1)

オッズの定義式 = 対象の数 / 対象以外の数 ... (付式2)

(2) 確率とオッズの関係

対象以外の数 = 全体の数 - 対象の数

全体の数 = 対象の数 / 割合

であるので、

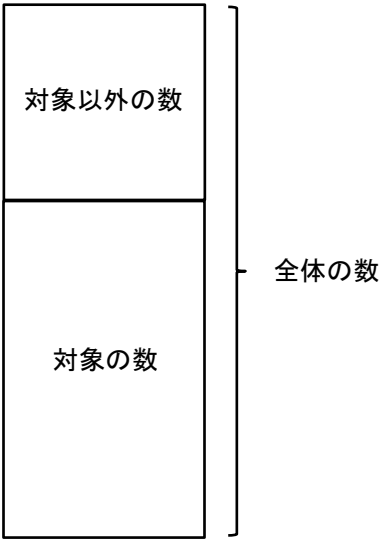
$$\begin{aligned} \text{オッズ} &= \text{対象の数} / \text{対象以外の数} \\ &= \text{対象の数} / (\text{全体の数} - \text{対象の数}) \\ &= \text{対象の数} / (\text{対象の数} / \text{割合} - \text{対象の数}) \\ &= 1 / (1 / \text{割合} - 1) \\ &= \text{割合} / (1 - \text{割合}) \quad \dots \text{(付式3)} \end{aligned}$$

逆に、割合を求める式に書き直せば、 $(1 - \text{割合}) * \text{オッズ} = \text{割合}$ であるので、

$\text{オッズ} - \text{割合} * \text{オッズ} = \text{割合} \rightarrow \text{オッズ} = (1 + \text{オッズ}) \text{割合}$ から、

$\text{割合} = \text{オッズ} / (1 + \text{オッズ}) \quad \dots \text{(付式4)}$

なお、全体の数が大きくなると、割合を確率と見なすことができる。



(付図1) 割合とオッズの説明

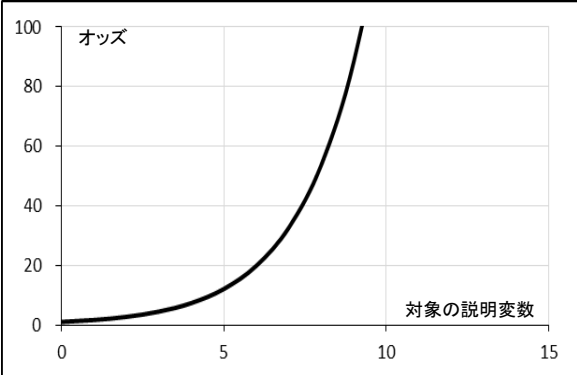
(3) ロジスティック回帰式のオッズ比

(式8)で示すとおり、ロジスティック回帰式のオッズ比は、説明変数の値が 1 増加するとき、オッズが高くなる率を示す。

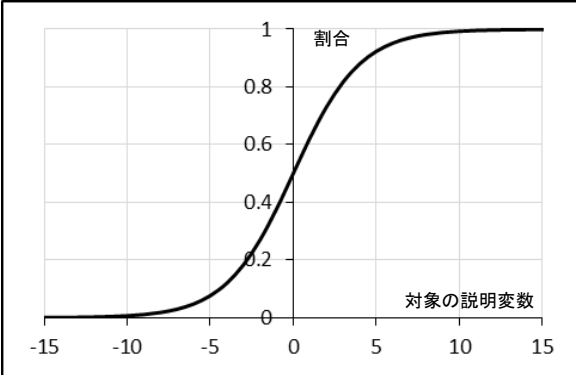
(例) 例えば、オッズ比が、 $e^{0.49825} = 1.64584$ であるとき、説明変数の増加がオッズ、もしくは割合(確率)へ及ぼす影響の度合いを確認する。

(付図2)はオッズへの影響、(図3)は割合への影響を示す。

この例では、オッズ比が 1 より大きいので、説明変数の変化量が大きくなればなるほどオッズへの影響が加速的に大きくなる事が分かる。一方、割合(確率)の変化の視点から見れば 1 に近づいていくことを確認できる。割合(確率)の変化であるので 0 ~ 1 の間に収まっている。



(付図2) オッズ



(付図3) 割合(確率)

付録2. Excel のソルバー機能の使い方

(表1)のデータを例に、ロジスティック回帰分析でのソルバーの使い方を示す。
Excel へのソルバー機能の実装方法は、別途資料「主成分分析の手順」を参照のこと。

(1) (付図4)のデータを準備する(対数尤度のセルと偏回帰係数のセルを設定する)。

	y	x1	x2	(式1)	(式2)	(式4)の各項
1	1	25	1	0.00000	0.50000	-0.69315
2	1	22	0	0.00000	0.50000	-0.69315
3	1	26	1	0.00000	0.50000	-0.69315
4	1	14	0	0.00000	0.50000	-0.69315
5	0	6	0	0.00000	0.50000	-0.69315
6	0	2	1	0.00000	0.50000	-0.69315
7	0	5	0	0.00000	0.50000	-0.69315
8	0	10	0	0.00000	0.50000	-0.69315
9	0	19	0	0.00000	0.50000	-0.69315
10	1	20	0	0.00000	0.50000	-0.69315
11	0	15	0	0.00000	0.50000	-0.69315
12	1	30	1	0.00000	0.50000	-0.69315
13	1	24	1	0.00000	0.50000	-0.69315
14	0	15	1	0.00000	0.50000	-0.69315
15	0	17	1	0.00000	0.50000	-0.69315
(式4)対数尤度						-10.39721
$\beta_0 =$		0.00000	偏回帰係数			
$\beta_1 =$		0.00000	(初期値は、0)		(付図4)	
$\beta_2 =$		0.00000			対数尤度	

- (2) ソルバーを起動し、(付図5)に示す設定を行う。
- ① 最大化したい変数のセル(対数尤度の値のセル)を指定する。
 - ② 最大化するために値を変化させるセル(偏回帰係数の値が入るセル)を指定する。
- ①、②の設定が終わったら、ソルバー起動のため、解決 ③ のボタンをクリックする。

(3) 対数尤度を最大にする偏回帰係数が求まったとのメッセージが表示される。

(4) 対数尤度を最大にする偏回帰係数の値が表示される。

	y	x1	x2	(式1)	(式2)	(式4)の各項
1	1	25	1	2.47193	0.92215	-0.08105
2	1	22	0	2.54153	0.92700	-0.07580
3	1	26	1	2.97019	0.95121	-0.05002
4	1	14	0	-1.44450	0.19085	-1.65627
5	0	6	0	-5.43052	0.00436	-0.00437
6	0	2	1	-8.98789	0.00012	-0.00012
7	0	5	0	-5.92878	0.00265	-0.00266
8	0	10	0	-3.43751	0.03114	-0.03164
9	0	19	0	1.04677	0.74015	-1.34767
10	1	20	0	1.54502	0.82419	-0.19335
11	0	15	0	-0.94624	0.27964	-0.32801
12	1	30	1	4.96320	0.99306	-0.00697
13	1	24	1	1.97368	0.87801	-0.13010
14	0	15	1	-2.51060	0.07512	-0.07809
15	0	17	1	-1.51409	0.18033	-0.19886
(式4)対数尤度						-4.18496
$\beta_0 =$		-8.42004	ソルバーによって算出された、		対数尤度を最大にする偏回帰係数の値	
$\beta_1 =$		0.49825			(付図7)	
$\beta_2 =$		-1.56435				

なお、対数尤度を最大にする偏回帰係数の値が見つからない場合は下記のようなエラーメッセージが表示される。

付録3. スコア関数とフィッシャー情報量

確率密度関数 $f(\mathbf{X})$ の形を決める母数 θ は確率変数 $\mathbf{X} = (x_1, x_2, \dots, x_r)$ のデータを使って、つまり確率変数 $\mathbf{X}$ が持っている情報を使って推定できる。この**確率変数 $\mathbf{X}$ が持っている情報量をフィッシャー情報量 (Fisher information) $I_r(\theta)$ として定義する**。母数 θ の不偏推定量の分散が小さければ小さいほど推定精度が高いことを示す。従ってこのとき、母数 θ を推定するときに使う確率変数 $\mathbf{X}$ の持っている情報量 $I_r(\theta)$ は大きいと考えられる。

フィッシャー情報量 $I_r(\theta)$ の定義式は以下の通りである(尤度関数に関しては、別途資料「統計的推定の手順」を参照)。

フィッシャー情報量 定義式 $I_r(\theta) = E[S_r(\theta)^2] = E[(\partial/\partial\theta \ln L(\theta|\mathbf{X}))^2]$

但し、同時確率密度関数 $f(\mathbf{X}|\theta) = f(x_1|\theta)f(x_2|\theta) \cdots f(x_r|\theta)$

尤度関数 $L(\theta|\mathbf{X}) = L(\theta|x_1)L(\theta|x_2) \cdots L(\theta|x_r)$, なお、 $L(\theta|\mathbf{X}) = f(\mathbf{X}|\theta)$, $\int_{-\infty}^{\infty} L(\theta|\mathbf{X}) d\theta = \int_{-\infty}^{\infty} f(\mathbf{X}|\theta) d\mathbf{X} = 1$

対数尤度関数 $\ell(\theta|\mathbf{X}) = \ln L(\theta|\mathbf{X})$ 但し、 $\ln$ は自然対数

スコア関数 (Score function) 定義式 $S_r(\theta) = \partial/\partial\theta \ln L(\theta|\mathbf{X})$

以下の式変形では、関数の微分の公式を使用する。 (公式) ① $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$ ② $(\ln f(x))' = f'(x)/f(x)$ 但し、' は微分 d/dx を表す。

(1) スコア関数の期待値

$$\begin{aligned} E[S_r(\theta)] &= E\left[\frac{\partial}{\partial\theta} \ln L(\theta|\mathbf{X})\right] = \int_{-\infty}^{\infty} f(\mathbf{X}|\theta) \left(\frac{\partial}{\partial\theta} \ln L(\theta|\mathbf{X})\right) d\mathbf{X} = \int_{-\infty}^{\infty} f(\mathbf{X}|\theta) \left(\frac{\partial}{\partial\theta} \ln f(\mathbf{X}|\theta)\right) d\mathbf{X} \\ &= \int_{-\infty}^{\infty} f(\mathbf{X}|\theta) \left(\frac{\partial}{\partial\theta} \ln f(\mathbf{X}|\theta)\right) / f(\mathbf{X}|\theta) d\mathbf{X} = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(\mathbf{X}|\theta) d\mathbf{X} = \frac{\partial}{\partial\theta} \left(\int_{-\infty}^{\infty} f(\mathbf{X}|\theta) d\mathbf{X}\right) = \frac{\partial}{\partial\theta} 1 = 0 \quad \cdots \text{(付式5)} \end{aligned}$$

(2) フィッシャー情報量の性質

(2-1) フィッシャー情報量はスコア関数の分散に等しい。式変形には、別途資料「いろいろな確率分布」の付録21と(付式5)を利用する。

$$\text{(付式5)より、} V[S_r(\theta)] = E[S_r(\theta)^2] - E[S_r(\theta)]^2 = E[S_r(\theta)^2] - 0 = E[S_r(\theta)^2] = I_r(\theta) \quad \cdots \text{(付式6)}$$

(2-2) フィッシャー情報量は、対数尤度関数の2次導関数の期待値を使って表すこともできる。

$$\begin{aligned} \partial^2/\partial\theta^2 \ln L(\theta|\mathbf{X}) &= \partial/\partial\theta \left(\frac{\partial}{\partial\theta} \ln L(\theta|\mathbf{X})\right) = \partial/\partial\theta \left(\left(\frac{\partial}{\partial\theta} L(\theta|\mathbf{X})\right)/L(\theta|\mathbf{X})\right) = \partial/\partial\theta \left((1/L(\theta|\mathbf{X}))(\partial/\partial\theta L(\theta|\mathbf{X}))\right) \\ &= \left(\frac{\partial}{\partial\theta} (1/L(\theta|\mathbf{X}))\right) \left(\frac{\partial}{\partial\theta} L(\theta|\mathbf{X})\right) + (1/L(\theta|\mathbf{X})) \left(\frac{\partial}{\partial\theta} \left(\frac{\partial}{\partial\theta} L(\theta|\mathbf{X})\right)\right) \\ &= -\left((1/L(\theta|\mathbf{X}))^2 \left(\frac{\partial}{\partial\theta} L(\theta|\mathbf{X})\right)\right) \left(\frac{\partial}{\partial\theta} L(\theta|\mathbf{X})\right) + (1/L(\theta|\mathbf{X})) \left(\frac{\partial^2}{\partial\theta^2} L(\theta|\mathbf{X})\right) \\ &= -\left(\left(\frac{\partial}{\partial\theta} L(\theta|\mathbf{X})\right)/L(\theta|\mathbf{X})\right)^2 + \left(\frac{\partial^2}{\partial\theta^2} L(\theta|\mathbf{X})\right)/L(\theta|\mathbf{X}) \\ &= -\left(\frac{\partial}{\partial\theta} \ln L(\theta|\mathbf{X})\right)^2 + \left(\frac{\partial^2}{\partial\theta^2} \ln L(\theta|\mathbf{X})\right) \end{aligned}$$

よって、 $(\partial/\partial\theta \ln L(\theta|\mathbf{X}))^2 = (\partial^2/\partial\theta^2 \ln L(\theta|\mathbf{X})) / L(\theta|\mathbf{X}) - \partial^2/\partial\theta^2 \ln L(\theta|\mathbf{X})$

$$\text{両辺の期待値をとると、} E[(\partial/\partial\theta \ln L(\theta|\mathbf{X}))^2] = E[(\partial^2/\partial\theta^2 \ln L(\theta|\mathbf{X})) / L(\theta|\mathbf{X})] - E[\partial^2/\partial\theta^2 \ln L(\theta|\mathbf{X})]$$

$$\begin{aligned} \text{ここで、} E[(\partial^2/\partial\theta^2 \ln L(\theta|\mathbf{X})) / L(\theta|\mathbf{X})] &= \int_{-\infty}^{\infty} \left((\partial^2/\partial\theta^2 \ln L(\theta|\mathbf{X})) / L(\theta|\mathbf{X})\right) f(\mathbf{X}|\theta) d\mathbf{X} = \int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta^2} L(\theta|\mathbf{X}) d\mathbf{X} \\ &= \frac{\partial^2}{\partial\theta^2} \int_{-\infty}^{\infty} L(\theta|\mathbf{X}) d\mathbf{X} = \frac{\partial^2}{\partial\theta^2} 1 = 0 \end{aligned}$$

$$\text{よって、} E[(\partial/\partial\theta \ln L(\theta|\mathbf{X}))^2] = 0 - E[\partial^2/\partial\theta^2 \ln L(\theta|\mathbf{X})] = -E[\partial^2/\partial\theta^2 \ln L(\theta|\mathbf{X})]$$

$$\text{従って、} I_r(\theta) = E[S_r(\theta)^2] = E[(\partial/\partial\theta \ln L(\theta|\mathbf{X}))^2] = -E[\partial^2/\partial\theta^2 \ln L(\theta|\mathbf{X})] = -E[\partial/\partial\theta S_r(\theta)] \quad \cdots \text{(付式7)}$$

(2-3) 同じ確率分布に従う独立した複数の確率変数のフィッシャー情報量は個々の確率変数のフィッシャー情報量の和になる。

$$\begin{aligned} I_r(\theta) &= E[S_r(\theta)^2] = E[(\partial/\partial\theta \ln L(\theta|\mathbf{X}))^2] = E[(\partial/\partial\theta \ln(L(\theta|x_1)L(\theta|x_2) \cdots L(\theta|x_r)))^2] \\ &= E[(\partial/\partial\theta (\ln L(\theta|x_1) + \ln L(\theta|x_2) + \cdots + \ln L(\theta|x_r)))^2] = E[(\partial/\partial\theta \ln L(\theta|x_1) + \partial/\partial\theta \ln L(\theta|x_2) + \cdots + \partial/\partial\theta \ln L(\theta|x_r))^2] \\ &= E[(\partial/\partial\theta \ln L(\theta|x_1))^2 + (\partial/\partial\theta \ln L(\theta|x_2))^2 + \cdots + (\partial/\partial\theta \ln L(\theta|x_r))^2 \\ &\quad + (\partial/\partial\theta \ln L(\theta|x_1))(\partial/\partial\theta \ln L(\theta|x_2)) + \cdots + (\partial/\partial\theta \ln L(\theta|x_{r-1}))(\partial/\partial\theta \ln L(\theta|x_r))] \end{aligned}$$

$$\text{ここで、(付式5)の導出と同じ手順で、} i \neq j \text{ のとき、} E[\partial/\partial\theta \ln L(\theta|x_i)](\partial/\partial\theta \ln L(\theta|x_j)) = E[\partial/\partial\theta \ln f(x_i|\theta)] E[\partial/\partial\theta \ln f(x_j|\theta)] = 0$$

$$\begin{aligned} \text{従って、} I_r(\theta) &= E[\partial/\partial\theta \ln L(\theta|x_1))^2 + \partial/\partial\theta \ln L(\theta|x_2))^2 + \cdots + \partial/\partial\theta \ln L(\theta|x_r))^2 + 0 \\ &= E[\partial/\partial\theta \ln L(\theta|x_1))^2] + E[\partial/\partial\theta \ln L(\theta|x_2))^2] + \cdots + E[\partial/\partial\theta \ln L(\theta|x_r))^2] \\ &= E[S_1(\theta)^2] + E[S_1(\theta)^2] + \cdots + E[S_1(\theta)^2] = I_1(\theta) + I_1(\theta) + \cdots + I_1(\theta) = r I_1(\theta) \quad \cdots \text{(付式8)} \end{aligned}$$

付録4. クラメール・ラオの不等式 (Cramer – Rao's Inequality)

シュワルツの不等式(別途資料「周波数解析(ウェーブレット解析)の基礎知識」の付録を参照)より、

$$E[(\hat{\theta} - \theta) S_r(\theta)]^2 \leq E[(\hat{\theta} - \theta)^2] E[S_r(\theta)^2]$$

ここで、 $V[\hat{\theta}] = E[(\hat{\theta} - \theta)^2]$ 、 $I_r(\theta) = E[S_r(\theta)^2]$ であるので、

$$E[(\hat{\theta} - \theta) S_r(\theta)]^2 \leq V[\hat{\theta}] I_r(\theta)$$

ところで、 $E[(\hat{\theta} - \theta) S_r(\theta)]^2 = (E[(\hat{\theta} - \theta) S_r(\theta)])^2 = (E[\hat{\theta} S_r(\theta)] - E[\theta S_r(\theta)])^2 = (E[\hat{\theta} S_r(\theta)] - \theta E[S_r(\theta)])^2$

ここで、(付式5)より、 $E[S_r(\theta)] = 0$ であるので、

$$\begin{aligned} E[(\hat{\theta} - \theta) S_r(\theta)]^2 &= (E[\hat{\theta} S_r(\theta)])^2 = (E[\hat{\theta} (\partial / \partial \theta \ln L(\theta | X))])^2 = (E[\hat{\theta} (\partial / \partial \theta f(X | \theta)) / f(X | \theta)])^2 \\ &= \{ \int_{-\infty}^{\infty} (\hat{\theta} (\partial / \partial \theta f(X | \theta)) / f(X | \theta)) f(X | \theta) dX \}^2 = \{ \int_{-\infty}^{\infty} \hat{\theta} (\partial / \partial \theta f(X | \theta)) dX \}^2 = \{ \partial / \partial \theta \int_{-\infty}^{\infty} \hat{\theta} f(X | \theta) dX \}^2 \end{aligned}$$

ここで、 $\hat{\theta}$ は θ の不偏推定量であるので、 $E[\hat{\theta}] = \theta$ である。つまり、 $E[\hat{\theta}] = \int_{-\infty}^{\infty} \hat{\theta} f(X | \theta) dX = \theta$ である。

よって、 $E[(\hat{\theta} - \theta) S_r(\theta)]^2 = \{ \partial / \partial \theta (\theta) \}^2 = 1$

従って、 $1 \leq V[\hat{\theta}] I_r(\theta)$ が成り立つので、(付式9)が導かれる。

$$V[\hat{\theta}] \geq 1 / I_r(\theta) = 1 / (r I_1(\theta)) \quad \dots \text{(付式9)}$$

この不等式が、クラメール・ラオの不等式である。 **クラメール・ラオの不等式は不偏推定量の分散の下限値を与える不等式である。**

不等式の右辺は、フィッシャーの情報量を使って表され、これを、クラメール・ラオの下限と言う。

フィッシャー情報量の値が大きいほど不偏推定量(最尤推定量)の分散の下限値が小さくなり、推定精度が上がる事が分かる。

なお、データ数の増加に伴って、推定値がある分布に収束するときその分布を「漸近分布」といい、その分布の分散を「漸近分散」という。

そして、この漸近分散がクラメール・ラオの下限であるとき、推定量は「漸近有効」とであるという。

付録5. ワルド検定 (Wald test) の導出

ワルド検定は2つのモデルそれぞれの最尤推定量の差を利用した検定である。

r 個の説明変数からなるモデル(母数 θ^0)と、少ない数 q 個($q < r$)の説明変数からなるモデル(母数 θ^1)を考える。前者のモデルの母数の最尤推定量を $\hat{\theta}^0 = [\hat{\theta}_0^0, \hat{\theta}_1^0, \dots, \hat{\theta}_r^0]$ 後者のモデルの母数の最尤推定量を $\hat{\theta}^1 = [\hat{\theta}_0^1, \hat{\theta}_1^1, \dots, \hat{\theta}_q^1]$ であるとする。モデル検定の帰無仮説と対立仮説を以下のように設定する。

帰無仮説: $\theta^1 = \theta^0$ つまり、2つのモデルに差はない。言い換えれば、後者のモデルの母数は前者のモデルとの違いを生むような母数ではない。

対立仮説: $\theta^1 \neq \theta^0$ つまり、前者の母数と後者の母数にはモデルに差を生むような違いがある

以下、 $\hat{\theta}^1 - \hat{\theta}^0$ の期待値と分散共分散行列を求める。導出にあたり、使用する関数を整理しておく(付録3. 参照)。

確率密度関数: $f(X | \theta)$ 母数 θ の確率分布に従う確率変数 X の確率密度関数であることを表す。

尤度関数: $L(\theta | X)$ 確率変数 X のデータから推定される母数 θ の確率分布の確率密度関数を表す。 $L(\theta | X) = f(X | \theta)$ である(視点が異なるだけ)。

対数尤度関数: $\ell(\theta | X) = \ln L(\theta | X)$ $\ln$ は、自然対数

スコア関数: $S_r(\theta) = \partial / \partial \theta (\ell(\theta | X)) = \partial / \partial \theta (\ln L(\theta | X))$

フィッシャー情報行列: (付式6)と(付式7)より、 $I_r(\theta) = V[S_r(\theta)] = -E[\partial / \partial \theta S_r(\theta)]$

(1) $\hat{\theta}^1 - \hat{\theta}^0$ の期待値を求める。

スコア関数 $S_r(\theta)$ を $\theta = \hat{\theta}^0$ の周りでテイラー展開し(別途資料「いろいろな確率分布」を参照)、最初の2項を使って関数近似する。

$$S_r(\theta) = S_r(\hat{\theta}^0) + \partial / \partial \theta (S_r(\hat{\theta}^0)) / 1! (\theta - \hat{\theta}^0) = S_r(\hat{\theta}^0) + (\partial / \partial \theta (S_r(\hat{\theta}^0))) (\theta - \hat{\theta}^0) = S_r(\hat{\theta}^0) - I_r(\hat{\theta}^0) (\theta - \hat{\theta}^0)$$

この式で、 $\theta = \hat{\theta}^0$, $\hat{\theta} = \hat{\theta}^1$ とおくと、 $S_r(\hat{\theta}^0) = S_r(\hat{\theta}^1) - I_r(\hat{\theta}^1) (\hat{\theta}^0 - \hat{\theta}^1)$

ここで、帰無仮説が成立すれば、 $\hat{\theta}^1$ を $\hat{\theta}^0$ の最尤推定量であると思ふことができる。

このとき、対数尤度 $\ell(\theta | X)$ の偏微分は $\theta = \hat{\theta}^1$ のとき 0 になる。つまり、 $\partial \ell(\theta | X) / \partial \theta |_{\theta = \hat{\theta}^1} = S_r(\theta) |_{\theta = \hat{\theta}^1} = S_r(\hat{\theta}^1) = 0$ になる。従って、

$$S_r(\hat{\theta}^0) = -I_r(\hat{\theta}^1) (\hat{\theta}^0 - \hat{\theta}^1)$$

フィッシャー情報行列 $I_r(\hat{\theta}^1)$ が逆行列を持つならば、 $\hat{\theta}^1 - \hat{\theta}^0 = I_r(\hat{\theta}^1)^{-1} S_r(\hat{\theta}^0)$... (付式10)

従って、 $\hat{\theta}^1 - \hat{\theta}^0$ の期待値は、(付式5)より $E[S_r(\hat{\theta}^0)] = 0$ なので、 $E[\hat{\theta}^1 - \hat{\theta}^0] = E[I_r(\hat{\theta}^1)^{-1} S_r(\hat{\theta}^0)] = I_r(\hat{\theta}^1)^{-1} E[S_r(\hat{\theta}^0)] = 0$... (付式11)

(2) $\hat{\theta}^1 - \hat{\theta}^0$ の分散共分散行列を求める。(付式10)を使って、また、フィッシャー情報量の定義より、また、 $(I_r(\hat{\theta}^1))$ は対称行列で、 $(I_r(\hat{\theta}^1))^{-1} = I_r(\hat{\theta}^1)^{-1}$ であるので

$$\begin{aligned} E[(\hat{\theta}^1 - \hat{\theta}^0)(\hat{\theta}^1 - \hat{\theta}^0)^T] &= E[I_r(\hat{\theta}^1)^{-1} S_r(\hat{\theta}^0) (I_r(\hat{\theta}^1)^{-1} S_r(\hat{\theta}^0))^T] = E[I_r(\hat{\theta}^1)^{-1} S_r(\hat{\theta}^0) S_r(\hat{\theta}^0)^T (I_r(\hat{\theta}^1)^{-1})^T] = E[I_r(\hat{\theta}^1)^{-1} S_r(\hat{\theta}^0) S_r(\hat{\theta}^0)^T I_r(\hat{\theta}^1)^{-1}] \\ &= I_r(\hat{\theta}^1)^{-1} E[S_r(\hat{\theta}^0) S_r(\hat{\theta}^0)^T] I_r(\hat{\theta}^1)^{-1} = I_r(\hat{\theta}^1)^{-1} I_r(\hat{\theta}^0) I_r(\hat{\theta}^1)^{-1} \end{aligned}$$

ここで、帰無仮説が成立すれば、 $\hat{\theta}^1$ を $\hat{\theta}^0$ の最尤推定量であると思ふことができるので、 $\hat{\theta}^0 = \hat{\theta}^1$ と置き換えることができる。

$$E[(\hat{\theta}^1 - \hat{\theta}^0)(\hat{\theta}^1 - \hat{\theta}^0)^T] = I_r(\hat{\theta}^1)^{-1} I_r(\hat{\theta}^1) I_r(\hat{\theta}^1)^{-1} = I_r(\hat{\theta}^1)^{-1} \dots (付式12)$$

$\hat{\theta}^1 - \hat{\theta}^0$ の分散共分散行列は、フィッシャー情報行列の逆行列、つまり、クラメール・ラオの不等式(付式9)で示す不偏推定量の分散の下限値になることが分かる。

(3) $\hat{\theta}^1 - \hat{\theta}^0$ を使った検定統計量を求める。

$\hat{\theta}^1 - \hat{\theta}^0$ の分散共分散行列の分散(対角要素)からなる対角行列(なお、対角要素以外の要素は全て 0)を、 $Q(\hat{\theta}^1)$ とおき、

$\hat{\theta}^1 - \hat{\theta}^0$ を標準化すると、 $(\hat{\theta}^1 - \hat{\theta}^0) - E[\hat{\theta}^1 - \hat{\theta}^0] / Q(\hat{\theta}^1)^{1/2}$ は、標準正規分布 $N(0, 1)$ に従う(別途資料「いろいろな確率分布」参照)。(式11)を使って、

$$((\hat{\theta}^1 - \hat{\theta}^0) - E[\hat{\theta}^1 - \hat{\theta}^0]) / Q(\hat{\theta}^1)^{1/2} = ((\hat{\theta}^1 - \hat{\theta}^0) - 0) / Q(\hat{\theta}^1)^{1/2} = (\hat{\theta}^1 - \hat{\theta}^0) / Q(\hat{\theta}^1)^{1/2} \sim N(0, 1) \dots (付式13)$$

そして、標準正規分布を二乗した統計量は、自由度 1 のカイ二乗分布に従う(付録A)。つまり、

$$(\hat{\theta}^1 - \hat{\theta}^0) Q(\hat{\theta}^1)^{-1} (\hat{\theta}^1 - \hat{\theta}^0)^T \sim \chi^2(1) \dots (付式14)$$

2つのモデルのパラメータの数は $q < r$ であり、パラメータ $\hat{\theta}^0$ のなかには 0 のものがいくつかある。今、 j 番目のパラメータが 0 ($\hat{\theta}_j^0 = 0$) であったとすると、帰無仮説 $\hat{\theta}_j^1 = \hat{\theta}_j^0$ を棄却できなければ、 $\hat{\theta}_j^1 = 0$ ($\hat{\theta}_j^1$ の値に意味はない)となる。従って、(付式14)に j 番目のパラメータ $\hat{\theta}_j^0 = 0$ を代入して、 $\hat{\theta}_j^1$ を検定する統計量 T を設定することができる。

$$T = (\hat{\theta}_j^1)^2 (Q(\hat{\theta}^1)^{-1})_{jj} \sim \chi^2(1)$$

ここで、標準誤差(Standard Error) $SE(\hat{\theta}_j^1) = (不偏推定量の分散 Q(\hat{\theta}^1)_{jj})^{1/2}$ であるから($_{jj}$ は対角要素を示す)、

$$T = (\hat{\theta}_j^1)^2 / (SE(\hat{\theta}_j^1))^2 = (\hat{\theta}_j^1 / SE(\hat{\theta}_j^1))^2 \sim \chi^2(1) \dots (付式15)$$

なお、 $SE(\hat{\theta}_j^1) = (Q(\hat{\theta}^1)_{jj})^{1/2} = (I_r(\hat{\theta}^1)^{-1})_{jj}^{1/2} \dots (付式16)$

と、標準誤差はフィッシャー情報量を使って表すことができる。(付式15)で示す統計量を使った自由度 1 のカイ二乗検定をワルド検定という。

なお、対数尤度の検定方法としては、差を使ったこのワルド検定の他にも、比を使った尤度比検定、傾きを使ったスコア検定がある。

付録A. カイニ乗分布の要約 (別途資料「いろいろな確率分布」参照)

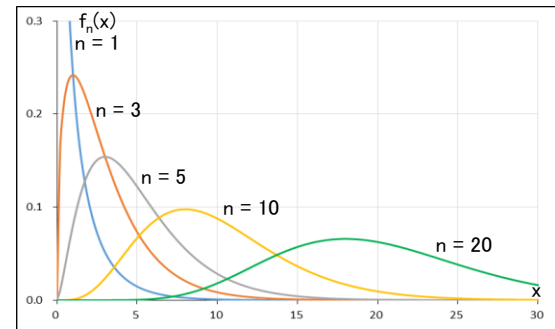
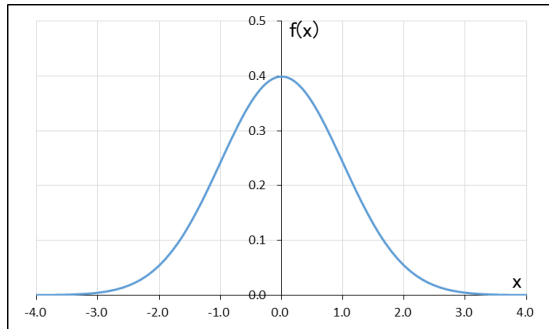
■概要

互いに独立な確率変数 $X_1, X_2, \dots, X_n$ が標準正規分布 ($\mu = 0, \sigma^2 = 1$) に従うとき、
確率変数 $X_1^2 + X_2^2 + \dots + X_n^2$ が従う確率分布を、自由度 n (n は和の数に相当) のカイニ乗分布という。

確率変数 X_i が、標準正規分布に従う

ならば、

$\sum_{i=1}^n X_i^2$ は、自由度 n のカイニ乗分布に従う



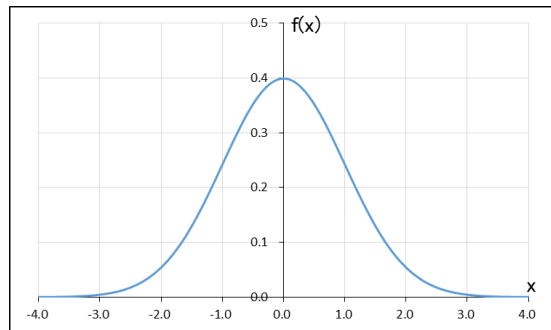
(付図A)カイニ乗分布の導出

付録B. t 分布の要約 (別途資料「いろいろな確率分布」参照)

■概要

t 分布は、標準正規分布を自由度を加味したカイニ乗分布でスケーリングした確率分布である。

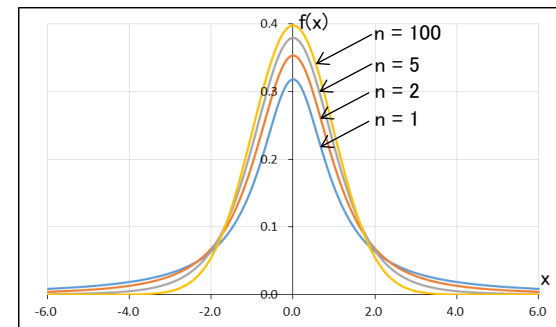
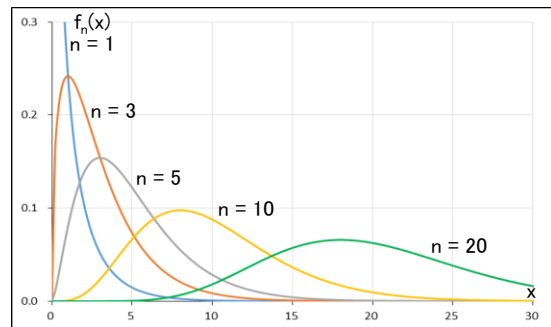
確率変数 X が、標準正規分布に従う



ならば、

$X / (Y / n)^{1/2}$ は、自由度 n の t 分布に従う

確率変数 Y が、自由度 n のカイニ乗分布に従う



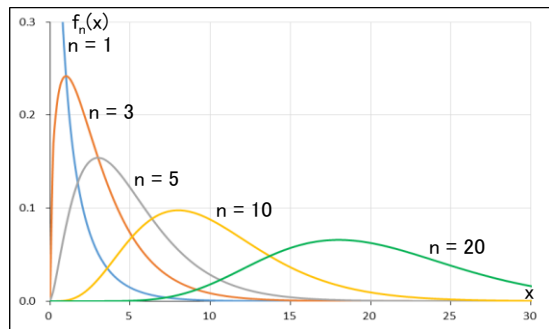
(付図B)t 分布の導出

付録C. F 分布の要約 (別途資料「いろいろな確率分布」参照)

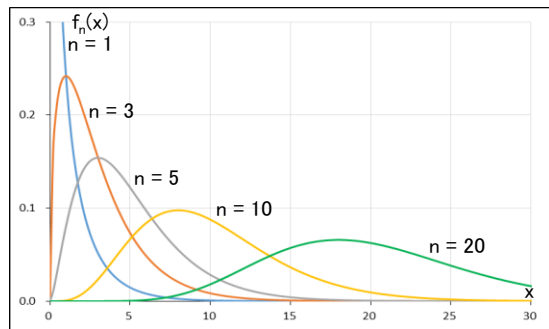
■概要

自由度 m のカイニ乗分布に従う確率変数と自由度 n のカイニ乗分布に従う確率変数が互いに独立であるとき、自由度を加味した確率変数の比が従う確率分布を、パラメータとして自由度 (m, n) の F 分布という。

確率変数 X が、自由度 m のカイニ乗分布に従う

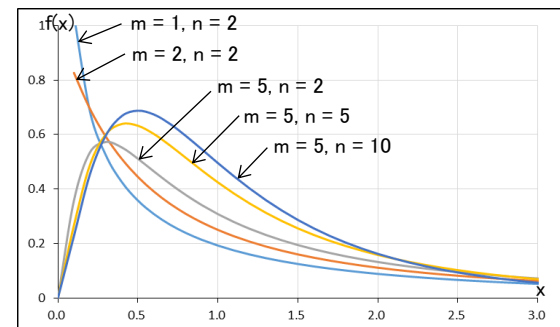


確率変数 Y が、自由度 n のカイニ乗分布に従う



ならば、

$(X / m) / (Y / n)$ は、自由度 (m, n) の F 分布に従う



参考にした資料

阿部真人：「データ分析に必須の知識・考え方 統計学入門」、ソシム

西内啓：「統計学が最強の学問である【実践編】」、ダイヤモンド社

高橋敬子：「線形代数から始める多変量解析」、プレアデス出版

いちばんやさしい医療統計：「オッズ比とリスク比(相対危険度)の違いは？」、[オッズ比とリスク比\(相対危険度\)の違いは？ - YouTube](#)

田中嘉博：「What's オッズ比？ ロジスティック回帰分析の結果の読み方」、[What's オッズ比？: ロジスティック回帰分析の結果の読み方 - YouTube](#)

統計研究疎(株)アイスタット：「多変量解析の手法解説 ロジスティック回帰(1/4)」、[ロジスティック回分析とは？具体例で分かりやすく解説 :: 【公式】株式会社アイスタット | 統計分析研究所 \(istat.co.jp\)](#)

Bell Curve 統計WEB：「ロジスティック回帰1」、[7-1. ロジスティック回帰分析1 | 統計学の時間 | 統計WEB \(bellcurve.jp\)](#)

biostatics：「フィッシャー情報量」、[フィッシャー情報量 | 最尤法,スコア関数,フィッシャー情報量の関係 \(biopapyrus.jp\)](#)

はじめての統計学：「フィッシャー情報量」、[【フィッシャー情報量完全版】フィッシャー情報量の意味、求め方、証明、例題を解説！ - YouTube](#)

はじめての統計学：「フィッシャー情報量とクラメルラオの不等式の下限」、[フィッシャー情報量とクラメルラオの不等式の下限 - YouTube](#)

脳内ライブラリアン：「尤度検定・ワルド検定・スコア検定」、[尤度比検定、ワルド検定、スコア検定をできるだけ分かりやすくまとめる【統計検定1級対策】 - 脳内ライブラリアン \(nounai-librarian.com\)](#)

Qiita：「最尤推定量の一致性と漸近正規性の証明」、[最尤推定量の一致性と漸近正規性の証明 - Qiita](#)

Biostatics：「Wald検定」、[Wald 検定 | 一般化線形モデルの最尤推定量の差を利用した検定 \(biopapyrus.jp\)](#)

Zhou Fan (Stanford University)：「Lecture 26 – Logistic regression」、<https://web.stanford.edu/class/archive/stats/stats200/stats200.1172/Lecture26.pdf>