統計解析

古谷知之

授業概要

* 履修者の状況に応じて変更される場合がありますが、
  全体としては以下のような授業構成となります。
* 講義の中でR演習を行うこともあります。

第1回　ガイダンス・単回帰分析
第2回　重回帰分析(1)
第3回　重回帰分析(2)
第4回　一般化線形回帰モデル(1)
第5回　一般化線形回帰モデル(2)
第6回　一般化線形回帰モデル(3)
第7回　一般化線形回帰モデル(4)
第8回　一般化線形回帰モデル(5)
第9回　一般化線形回帰モデル(6)
第10回　一般化線形混合モデル
第11回　状態空間モデル
第12回　R演習(1)
第13回　R演習(2)
第14回　R演習(3)

統計モデルの種類

	主な推定方法	データ分布	回帰係数
線形回帰モデル（単回帰・重回帰など）	最小二乗法	正規分布	一変数に一つ
一般化線形モデル	最尤推定法	正規分布以外の分布も可能	一変数に一つ
一般化線形混合モデル			変数の個体差に応じて推定可能
階層ベイズモデル	ベイズ推定		

本授業で扱う統計モデル

- 線形回帰モデル
    - 単回帰モデル、重回帰モデル
- 一般化線形回帰モデル
    - 離散：ポアソン回帰モデル、二項反応モデル（ロジスティック回帰モデル、プロビット回帰モデル、補対数対数モデル）、負の二項分布モデル、ゼロ過剰ポアソン回帰モデル、ゼロ過剰負の二項分布モデル
    - 連続：ガンマ回帰モデル、ベータ回帰モデル、指数-ガウス回帰モデル
    - スパース：Lasso回帰モデル、Ridge回帰モデル
- 一般化線形混合モデル
    - マルチレベルモデル
- 状態空間モデル

代表的な一般化線形回帰モデル

- 被説明変数が離散変数
  - 0or1の２値：(二項)ロジスティック回帰モデル、 (二項)プロビット回帰モデル、補対数対数モデル
  - ０以上の整数
    - 発生頻度が少ない：ポアソン回帰モデル、負の二項分布モデル
    - 発生頻度０が非常に多い：Hurdleモデル、ゼロ過剰モデル
- 被説明変数が連続変数
  - [0, 1]の確率値：ベータ回帰モデル
  - ０より大きい値：ガンマ回帰モデル、指数-ガウス回帰モデル
- 被説明変数がスパース
  - Lasso回帰モデル、Ridge回帰モデル

一般化線形モデルの確率分布とリンク関数

モデル	被説明変数	確率分布	リンク関数
線形回帰モデル	実数	正規分布	恒等リンク
ロジスティック回帰モデル	0/1の二値	二項分布	logitリンク
プロビット回帰モデル	0/1の二値	二項分布	probitリンク
補対数対数モデル	0/1の二値	二項分布	cloglogリンク
ポアソン回帰モデル	非負の整数	ポアソン分布	logリンク
負の二項分布モデル	非負の整数	負の二項分布	logitリンク
ベータ回帰モデル	[0,1]の実数	ベータ分布	logitリンク
ガンマ回帰モデル	非負の実数	ガンマ分布	逆logリンク
指数-ガウス回帰モデル	裾の長い実数	指数-ガウス分布	logリンクor恒等リンク

授業内容

- 一般線形回帰モデルのモデル選択（ポアソン回帰モデルを例に）

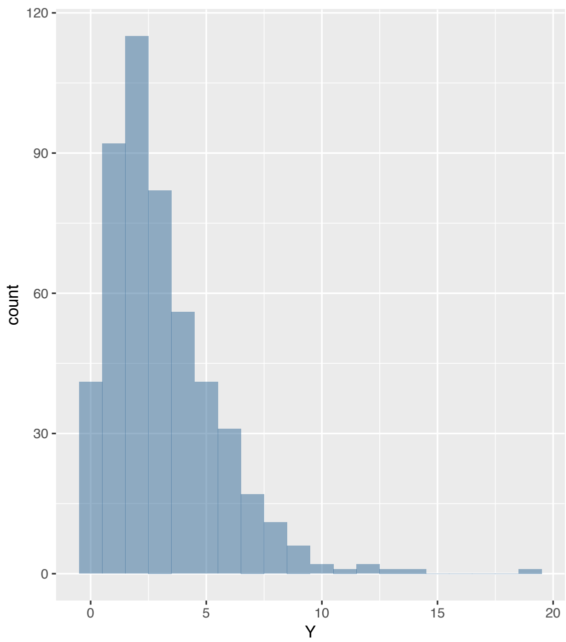
- AIC
  - 最大対数尤度、逸脱度
- 尤度比検定
- Wald検定
- スコア検定

ポアソン分布

- 非常に多くの観測回数が繰り返されるが、観測ケースの発生頻度が非常に低い場合に用いられる確率分布
- 一定範囲内（時間、回数、空間）である事象が発生する平均=λをもちいて計算される
- 試行回数n、発生頻度pとすると、$\lambda = np$
- ポアソン分布は次式で表される

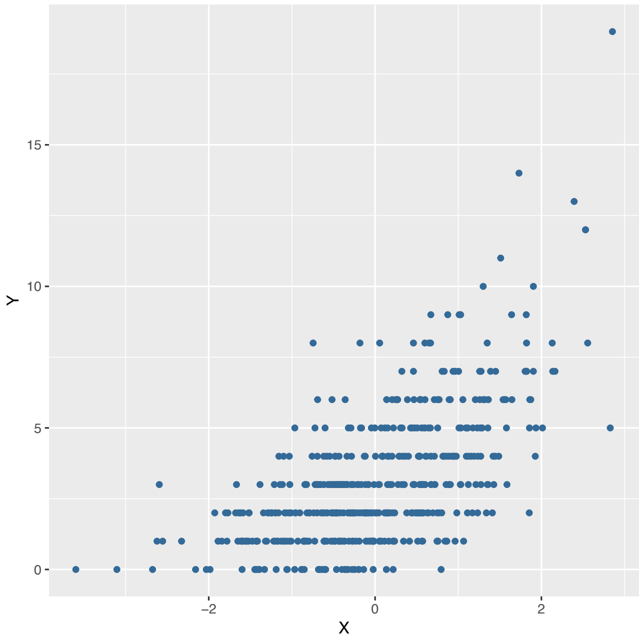
$$P(y|\lambda) = \frac{\lambda^y \cdot \exp(-\lambda)}{y!}$$

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



ポアソン回帰モデル

- ポアソン回帰モデルは、右図のような関係にあるデータに適用される
- 被説明変数は 0 以上の整数
- 適用事例
  - スポーツの得点分布
  - 感染症の感染者数
  - 事故発生件数
  - イベント来訪者数



ポアソン回帰モデル

- 被説明変数がポアソン分布に従うとするポアソン回帰モデルは、例えば次式のように表される。ここでy_iは被説明変数、X_iは説明変数、$\boldsymbol{\beta}$は未知パラメータである

$$y_i = Po(\lambda_i)$$
$$f(y_i|\lambda_i) = \frac{{\lambda_i}^{y_i} \cdot \exp(-\lambda_i)}{y_i!}$$
$$g\big(E(y_i)\big) = \lambda_i = \exp(\eta_i)$$

- このときlogリンク(対数リンク)$\log(\lambda_i)$と線形予測子$\eta_i = X_i\boldsymbol{\beta}$との関係は、次式のように対数リンク関数で表せる

$$\log(\lambda_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} = X_i^T\boldsymbol{\beta}$$

ポアソン回帰モデルの最尤推定

- 尤度関数Lは

$$L=\prod_{i=1}^{n}\frac{{\lambda_i}^{y_i}\cdot\exp(-\lambda_i)}{y_i!}=\prod_{i=1}^{n}\frac{\left(\exp(X_i\boldsymbol{\beta})^{y_i}\exp\left(-exp(X_i\boldsymbol{\beta})\right)\right)}{y_i!}$$

- 対数尤度関数Lは

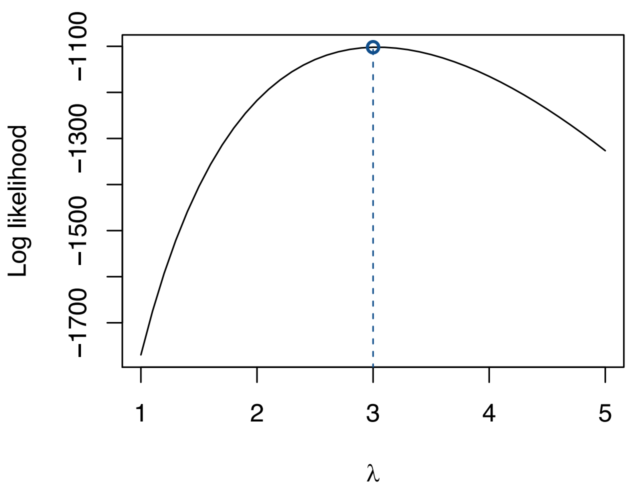
$$\begin{aligned}\ln L&=\sum_{i=1}^{n}\left(y_i\cdot\ln\lambda_i-\lambda_i-\sum_{k=1}^{y_i}\ln k\right)\\&=\sum_{i=1}^{n}\left(y_i\cdot\ln\left(X_i^T\boldsymbol{\beta}\right)-\left(X_i^T\boldsymbol{\beta}\right)-\sum_{k=1}^{y_i}\ln k\right)\end{aligned}$$

- 対数尤度関数を最大化する$\hat{\lambda}$は

$$\frac{d\ln L}{d\lambda}=0$$

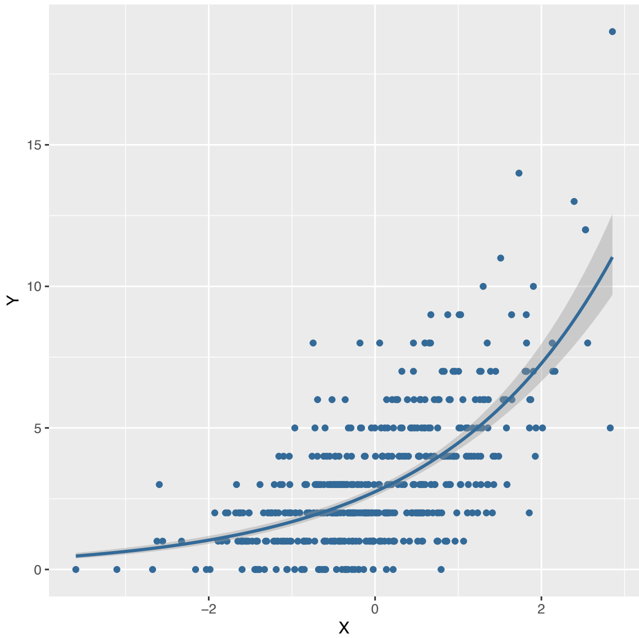
ポアソン回帰モデルの尤度関数

- ポアソン回帰モデルの尤度関数は右図のよう上に凸な関数となる
- この例では、$\lambda=3.0$が最尤推定値$\hat{\lambda}$となる
- 最尤推定値$\hat{\lambda}$を代入した最大対数尤度は$\ln\hat{L}=-1101.96$



ポアソン回帰モデルの推定結果の例

- 500個のランダムなポアソン分布に従う被説明変数と正規分布に従う説明変数を用いてポアソン回帰モデルを推定
- 得られたモデルの曲線と95%信頼区間は右図のとおり



モデル間の比較

- ここで、以下のモデルを考えることにする

- 定数項のみのモデルをモデル0（null model）
- 定数項と変数x_1を加えたモデル1
- $x_1, \cdots, x_p$のp個の変数を持つモデルp
- $x_1, \cdots, x_p, x_q$の$q(> p)$個の変数を持つモデルq
- すべての変数$x_1, \cdots, x_k$を加えたモデルk（full model）

- モデルk以外はreduced modelなどとも呼ばれる

モデル間の比較

- ポアソン回帰モデルを例に挙げると、考えうる説明変数がk個あるとき、比較対象となる対数リンク関数は以下のようになる

- モデル0： $\log(\lambda_i) = \beta_0$
- モデル1： $\log(\lambda_i) = \beta_0 + \beta_1 x_{i1}$

  ：
- モデルp： $\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$
- モデルq： $\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \beta_q x_{iq}$

  ：
- モデルk： $\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ip} + \beta_k x_{iq} \cdots + \beta_k x_{ik}$

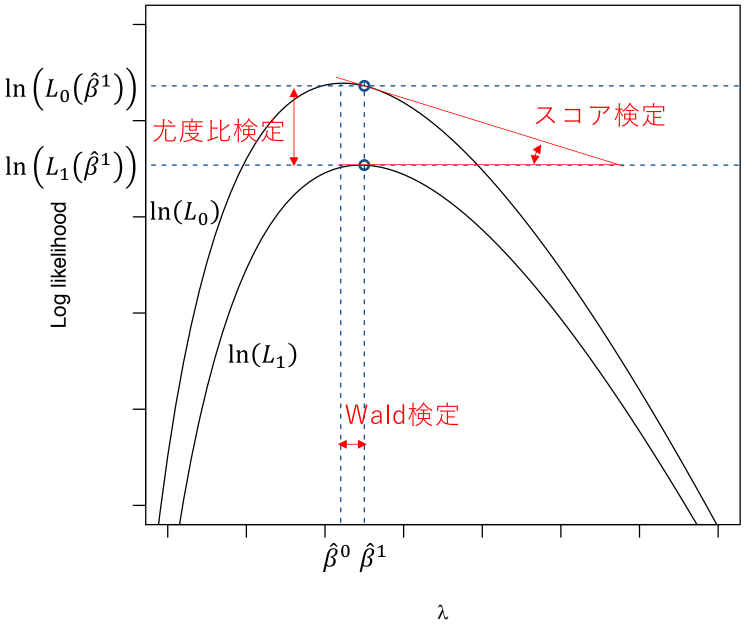
ネスト関係にあるモデル間の比較

- モデル１の線形予測子はモデル０の線形予測子を含有する
- モデルqの線形予測子はモデルpの線形予測子を含有する
- このとき、モデル１はモデル０をネスト化している、モデルqはモデルpをネスト化している、という

- ネスト関係にあるモデルを比較するときには、赤池の情報量基準（AIC）を用いることが有効であることが知られている

一般化織系モデルの比較方法

- モデル間の比較には、尤度比検定、Wald検定、スコア検定も用いられる
- それぞれの比較方法をモデル０とモデル１で示すと、およそ右図のとおり
- 尤度比検定：モデル０とモデル１の尤度比を検定
- Wald検定：回帰係数の差のWald統計量を検定
- スコア検定：対数尤度関数の一次導関数（スコア関数）のスコア統計量を検定



赤池の情報量基準（AIC）

- 赤池の情報量基準AICは、対数尤度関数を最大化する解が得られた場合、対数尤度$\ln\hat{L}$と変数の数kを用いて、次式で表される

$$AIC = -2\ln\hat{L} + 2k$$

- ポアソン回帰モデルの場合、最尤推定値$\hat{\lambda}$が得られた場合の最大対数尤度$\ln\hat{L}$は次式のようになる

$$\ln\hat{L} = \sum_{i=1}^{n}\left(y_i \cdot \ln\hat{\lambda} - \hat{\lambda} - \sum_{k=1}^{y_i}\ln k\right)$$

赤池の情報量基準（AIC）

- AICはどの値が良いというわけではなく、 AICが（相対的に）より小さい値を取るモデルが良いとされる
- 最大対数尤度は「観測データへの当てはまりが最も良いモデルを推定する」という発想
- AICは「良い予測をするモデルが良いモデル」という発想
  - AICは「当てはまりが良いモデル」や「真のモデル」を選ぶというわけではない

逸脱度 (deviance)

- 最大対数尤度$\ln\hat{L}$が得られたとき、最大対数尤度に-2を乗じたものを逸脱度Dという

$$D = -2\ln\hat{L}$$

- モデル0 (null model)のとき最大逸脱度
- モデルk (full model)のとき最小逸脱度
- モデルpとモデルqのを比較する際、各モデルの逸脱度の差$D_{p,q}$を考えることで、「どちらのモデルがより当てはまりが良いか」を把握することができる

逸脱度の差

- モデルqの方がモデルpよりデータへの当てはまりが良いとき、$\ln\hat{L}^p < \ln\hat{L}^q$となり、逸脱度の差$D_{p,q}$は次式であらわされる

$$D_{p,q} = -2\phi\left(\ln\hat{L}^p - \ln\hat{L}^q\right) = 2\phi\left(\ln\hat{L}^q - \ln\hat{L}^p\right)$$

- 当てはまりの良さを意味する対数尤度の差に-2ϕをかけている
- ϕはdispersion parameterと呼ばれるばらつきに関するパラメータであり、確率分布ごとに想定された値がある
  - 正規分布： $\phi = \sigma^2$、二項分布・ポアソン分布： $\phi = 1$

- パラメータを増やすほどモデルの当てはまりは良くなる

逸脱度の差

- 逸脱度の差$D_{p,q}$は「当てはまりの悪さの差」なので、 $\ln\hat{L}^p < \ln\hat{L}^q$のときには、モデル$p$よりモデル$q$の方が、$D_{p,q}$の値の分だけ当てはまりの悪さが改善されているといえる

- 自身が提案したモデルpとフルモデルkの逸脱度との差$D_{p,k}$を残差逸脱度 (residual deviance)という。これは最小逸脱度を基準とする当てはまりの悪さを意味する

$$D_{p,k} = 2\phi\left(\ln\hat{L}^k - \ln\hat{L}^p\right)$$

- 最大逸脱度と最小逸脱度の差$D_{1,k}$をNull逸脱度 (null deviance)という

$$D_{1,k} = 2\phi\left(\ln\hat{L}^k - \ln\hat{L}^1\right)$$

過分散 (over dispersion)

- ばらつきに関するdispersion parameter ϕは、以下の方法により計算できる

- 残差自由度($Residual\ D.F.$)と残差逸脱度の比

$$\phi = \frac{Residual\ Deviance}{Residual\ D.F.}$$

- $\mathrm{Peason's}\ \chi^2$値と残差逸脱度の比

$$\phi = \frac{\mathrm{Peason's}\ \chi^2}{Residual\ D.F.}$$

- 前者は計算が簡単だが安定性にかけることがあるため、後者のほうが正確との意見も多い

過大分散と過小分散

- 過大分散
  - Dispersion parameter ϕが想定している値より大きい場合のこと
  - 本来統計的に有意でない説明変数が有意であると判断される場合（第１種の過誤）が多い
  - ϕがいくつ以上だと過大分散という明確な区分はない（1.5程度？）
  - 変数に異なる種類のデータが混合している場合などに生じる

- 過小分散
  - ϕが想定している値より小さい場合のこと
  - サンプル数を増やすなどする？

尤度比検定

- 尤度比検定は、２つのモデル間の尤度比を用いた検定
- 例えば、モデル０のパラメータベクトルをβ^0、モデル１のパラメータベクトルをβ^1とすると、帰無仮説$\mathcal{H}_0$と対立仮説$\mathcal{H}_1$は以下のようになる

$$\mathcal{H}_0 : \beta^0 = \beta^1$$
$$\mathcal{H}_1 : \beta^0 \neq \beta^1$$

- 帰無仮説$\mathcal{H}_0$は、「モデル１とモデル０は同じ変数の組み合わせになるので、モデル１で追加した変数のパラメータは０である」ことと同じ意味
- 帰無仮説$\mathcal{H}_0$が棄却されれば、「モデル０は採用すべきでない」といえる→「モデル０ではなくモデル１を採用すべきだ」というわけではない（後ほど説明）

尤度比と逸脱度

- モデル０とモデル１の尤度関数をそれぞれL_0、L_1とする
- また各モデルの最大尤度を$\hat{L}^0$・$\hat{L}^1$、最尤推定量を$\hat{\beta}^0$・$\hat{\beta}^1$とする

$$\hat{L}^0 = L_0\left(\hat{\beta}^1\right), \hat{L}^1 = L_1\left(\hat{\beta}^1\right)$$

- 任意の点において尤度関数L_1はL_0より大きくなることから、$\hat{\beta}^1$に関する尤度比LRは次式のようになる

$$LR = \frac{\hat{L}^0}{\hat{L}^1} = \frac{L_0\left(\hat{\beta}^1\right)}{L_1\left(\hat{\beta}^1\right)}$$

尤度比と逸脱度

- 尤度比$LR = \frac{\hat{L}^0}{\hat{L}^1} = \frac{L_0(\hat{\beta}^1)}{L_1(\hat{\beta}^1)} = 1$のとき、２つのモデルは同じになる
- このとき、モデル１(reduced model)に含まれる変数はモデル構築の際に有効でないことを意味する
- 検定を行う際には、尤度比を対数変換した次式の逸脱度の差$D_{0,1}$を用いる

$$D_{0,1} = -2\ln LR = -2\left(\ln\hat{L}^0 - \ln\hat{L}^1\right)$$
$$= -2\left(\ln\left(L_0\left(\hat{\beta}^1\right)\right) - \ln\left(L_1\left(\hat{\beta}^1\right)\right)\right)$$

- モデル０とモデル１とのパラメータ数の差は1個である。このとき、逸脱度Dは自由度1のχ^2分布に従う

尤度比と逸脱度

- 逸脱度の差$D_{0,1}$はどう解釈すればよいか？

$$D_{0,1} = -2\log\lambda = 2\left(\ln\hat{L}^0 - \ln\hat{L}^1\right)$$
$$= -2\left(\log\left(L_0\left(\hat{\beta}^1\right)\right) - \log\left(L_1\left(\hat{\beta}^1\right)\right)\right)$$

- 逸脱度は「当てはまりの悪さ」を意味する（≠当てはまりの良さ）→当てはまりの良さを意味する対数尤度に-2をかけている
- 逸脱度の差は「当てはまりの悪さの差」なので、 $D_{0,1}$の値が大きいほどモデル１の方が当てはまりの悪さが（$D_{0,1}$の値の分だけ）改善されているといえる

尤度比と逸脱度

- モデルqの方がモデルpよりデータへの当てはまりが良い（$\ln\hat{L}^p < \ln\hat{L}^q$）とする
- 変数がp個のモデルpと変数が$q(> p)$個のモデルqを比較する場合、その逸脱度の差$D_{p,q}$は自由度$q-p$のχ^2分布に従う

$$
\begin{aligned}
D_{p,q} &= -2\log\lambda = -2\left(\ln\hat{L}^p - \ln\hat{L}^q\right) \\
&= -2\left(\log\left(L_p\left(\hat{\beta}^1\right)\right) - \log\left(L_q\left(\hat{\beta}^1\right)\right)\right) \sim \chi^2(q-p)
\end{aligned}
$$

- χ^2検定を用いてモデル間の差が統計的に有意であるかどうかを把握することができる（ただし、サンプル数が多い場合）
- χ^2検定の結果、p値< 0.05だと95%水準で有意となり帰無仮説$\mathcal{H}_0$は棄却できる→モデル 0 を採用しない（モデル 1 を採用すべきかはさらに検討が必要）

分散分析 (ANOVA)

- 尤度比を用いて説明変数を抜く場合、分散分析ANOVAを用いることがある
- TYPE I
- TYPE II
- TYPE III

統計的過誤

- 第一種の過誤（偽陽性）は、有意水準αにより検出可能
- 対立仮説$\mathcal{H}_1$が正しいときに第二種の過誤（偽陰性）を犯さない確率を検出力$1-\beta$という
- 第二種の過誤を犯す確率βは直接計算できないが、有意水準αと帰無仮説$\mathcal{H}_0$が主張する仮定と真実との解離度に依存する

	真実	
仮説検定の結果	$\mathcal{H}_0$が正しい	$\mathcal{H}_1$が正しい
$\mathcal{H}_0$を棄却する	第一種の過誤 確率：有意水準α	正しい検出結果 確率：検出力$1-\beta$
$\mathcal{H}_0$を棄却しない	正しい検出結果 確率：$1-\alpha$	第二種の過誤 確率：β

有意水準αの設定と統計的判断

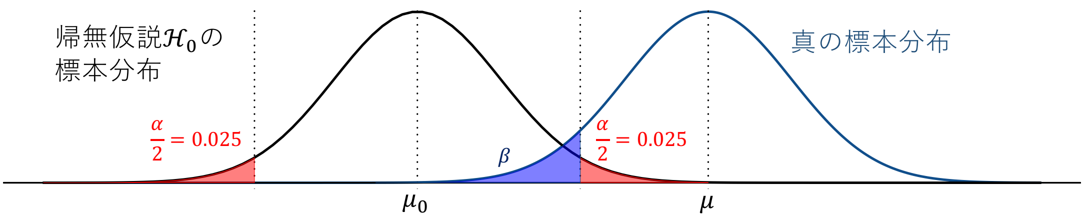
- 有意水準αは、分析者が分析の精度に応じて任意に設定できる
- この授業では$\alpha=0.05$を多く使うが、それは信頼区間を95%に設定しておけば、そこから外れるような事象はほとんど起きないだろうと「勝手に」想定しているから
- 社会科学分野では便宜上$\alpha=0.05$を多用するが、 $\alpha=0.1$や$\alpha=0.001$などであってもよい。
- 医療・薬学分野などでは、非常に小さい有意水準（$\alpha=0.001$など）が採用されることがある
- 有意水準を小さくすれば、統計的判断が厳しくなり、例えば冤罪や偽陽性などの誤検出の可能性を少なくできる。他方、完全犯罪や偽陰性（検出失敗）を増やしてしまう。

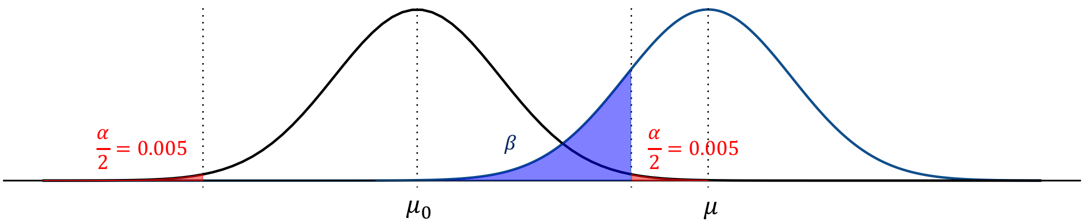
統計的過誤

- 第一種の過誤（偽陽性）：帰無仮説$\mathcal{H}_0$が正しいにもかかわらず帰無仮説$\mathcal{H}_0$を棄却してしまう誤り
- 第二種の過誤（偽陰性）：帰無仮説$\mathcal{H}_0$が正しくないにもかかわらず帰無仮説$\mathcal{H}_0$を棄却できない誤り

	真実	
仮説検定の結果	$\mathcal{H}_0$が正しい	$\mathcal{H}_1$が正しい
$\mathcal{H}_0$を棄却する	第一種の過誤 確率：有意水準α	正しい検出結果 確率：検出力$1-\beta$
$\mathcal{H}_0$を棄却しない	正しい検出結果 確率：$1-\alpha$	第二種の過誤 確率：β

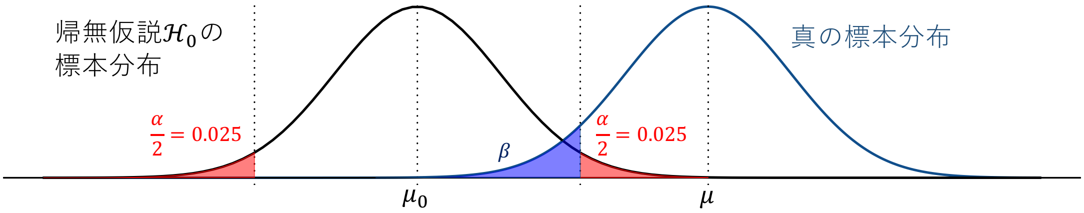
統計的過誤と有意水準α

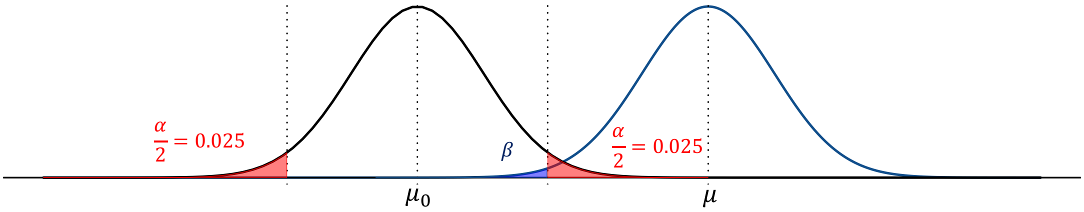




- 有意水準が厳しくなると、第一種の過誤の確率αは小さくなるが、第二種の過誤の確率βは大きくなる（トレードオフの関係）

統計的過誤と標本サイズ





- 帰無仮説の標本分布を知ることはできないのでを動かせないが、標本数を増やし分散を小さくすることで、αは一定でもβを小さくできる

Wald統計量

- パラメータを採用すべきかどうかは、帰無仮説$\mathcal{H}_0$: $\hat{\beta} = 0$を棄却すればよい
- パラメータの最尤推定量$\hat{\beta}$をその標準誤差$SE_{\hat{\beta}}$で割った$z = \frac{\hat{\beta}}{SE_{\hat{\beta}}}$値は平均$z$、分散1の正規分布に従う
- 帰無仮説$\mathcal{H}_0$: $\hat{\beta} = 0$のもとでは、 Z値は平均0、分散1の標準正規分布に従う
- Z値を2乗したZ^2値をWald統計量Wといい自由度1のχ^2分布に従う

$$W = Z^2 = \left(\frac{\hat{\beta}}{SE_{\hat{\beta}}}\right)^2 \sim \chi^2(1)$$

Wald統計量

- z値の絶対値またはWald統計量は、0より大きければ多いほど、帰無仮説を棄却できると言える

- z値の確率分布について、0以下の区間$Pr(< |z|)$は、 z値（あるいはWald統計量の）信頼区間と近似的に考えることができる

- $Pr(< |z|)$が小さいほど、帰無仮説$\mathcal{H}_0$: $\hat{\beta} = 0$を棄却することができ、z値（あるいはWald統計量）は統計的に有意であると見立てることができる

Wald統計量

- 一般に説明変数p個のモデルpのパラメータ$\hat{\beta}^p$ のWald統計量Wは、次式のように表される

$$W = \left(\hat{\beta}^p\right)'\left[Cov\left(\hat{\beta}^p\right)\right]^{-1}\hat{\beta}^p = \left(\hat{\beta}^p\right)'(XV'X)\hat{\beta}^p$$

- このとき、 Wald統計量Wは自由度$p+1$のχ^2分布に従う

Wald統計量

- $p=1$のとき、 $\left[Var\left(\hat{\beta}^{p=1}\right)\right]^{-1}=\begin{bmatrix}\phi_{00} & \phi_{01}\\ \phi_{10} & \phi_{11}\end{bmatrix}$とすると、Wald統計量$W$は以下のようになる

$$W=\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{pmatrix}'\begin{bmatrix}\phi_{00} & \phi_{01}\\ \phi_{10} & \phi_{11}\end{bmatrix}\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{pmatrix}$$

$$=\hat{\beta}_0^2\phi_{00}+\hat{\beta}_0\ \hat{\beta}_1\ \phi_{01}+\hat{\beta}_0\ \hat{\beta}_1\ \phi_{10}+\hat{\beta}_1^2\phi_{11}$$

$$=\frac{Var\left(\hat{\beta}_1\right)\hat{\beta}_0^2+Var\left(\hat{\beta}_0\right)\hat{\beta}_1^2-Cov\left(\hat{\beta}_0,\hat{\beta}_1\right)\hat{\beta}_0\ \hat{\beta}_1}{Var\left(\hat{\beta}_0\right)Var\left(\hat{\beta}_1\right)-Cov^2\left(\hat{\beta}_0,\hat{\beta}_1\right)}$$

Wald統計量

- fullモデルkのパラメータ$\hat{\beta}^k$に対してreduced モデルqのパラメータ$\hat{\beta}^q$を比較したいとき
- 帰無仮説$\mathcal{H}_0$: $\hat{\beta}^q=\hat{\beta}^k$が成り立つなら$E\left[\hat{\beta}^q\right]=\hat{\beta}^k$となる
- Wald統計量Wは、次式のように表される

$$W=\left(\hat{\beta}^p-\hat{\beta}^k\right)'\left[Cov\left(\hat{\beta}^p\right)\right]^{-1}\left(\hat{\beta}^p-\hat{\beta}^k\right)$$

- このとき、 Wald統計量Wは自由度$k-p$のχ^2分布に従う

スコア検定

- 対数尤度関数の一次導関数をスコア関数Uという
- 任意の説明変数変数$j=(1,\dots,k)$の回帰係数β_jのスコア関数$U_j(\boldsymbol{\beta})$は、次式のように表せる

$$U_j(\boldsymbol{\beta})=\frac{\partial \mathrm{ln}L}{\partial \beta_j}=0$$

スコア検定

- ポアソン回帰モデルの対数尤度関数Lは

$$\mathrm{ln}L=\sum_{i=1}^{n}\left(y_i\cdot \mathrm{ln}\lambda_i-\lambda_i-\sum_{k=1}^{y_i}\mathrm{ln}k\right)$$
$$=\sum_{i=1}^{n}\left(y_i\cdot \mathrm{ln}\left(X_i^T\boldsymbol{\beta}\right)-\left(X_i^T\boldsymbol{\beta}\right)-\sum_{k=1}^{y_i}\mathrm{ln}k\right)$$

- スコア関数は次式のようになる

$$U_j(\boldsymbol{\beta})=\frac{\partial \mathrm{ln}L}{\partial \beta_j}=\sum_{i=1}^{n}\left(\frac{(y_i-\lambda_i)X_{ij}}{Var(y_i)}\cdot\frac{\partial \lambda_j}{\partial \eta_j}\right)$$

スコア検定

- スコア統計量を用いてパラメータを採用すべきかどうかを仮説検定するとき、棄却すべき帰無仮説は$\mathcal{H}_0: E\left[U_j\left(\hat{\beta}\right)\right] = 0$となる

- 説明変数p個のモデルpのパラメータ$\hat{\beta}^p$に関するスコア統計量は次式のように表される

$$U_j\left(\hat{\beta}^p\right)'\left[Cov\left(U_j\left(\hat{\beta}^p\right)\right)\right]^{-1}U_j\left(\hat{\beta}^p\right) = U_j\left(\hat{\beta}^p\right)'(XV'X)U_j\left(\hat{\beta}^p\right)$$

- このとき、スコア統計量は自由度$p+1$のχ^2分布に従う