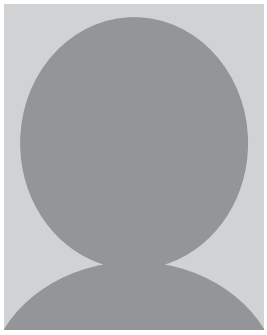
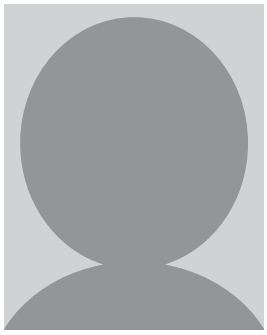
問題解決の数理（’１７）

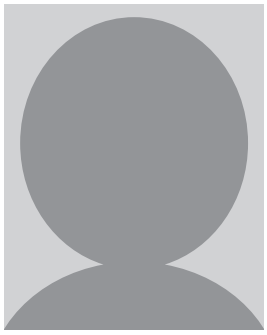
- 収録本番とは多少異なっていることがあります.
- 内容の間違いのご指摘は歓迎します.
- 「完全に無保証」です.



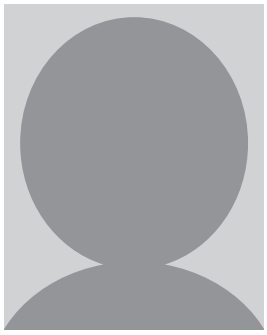
- 今回の講義では，統計モデルについてお話します.
- 統計，あるいは統計学と何らかの形で関わった経験をお持ちの方は多いかと思います.
- 放送大学に限らず、大学の授業では、心理統計、社会統計、経済統計、医学統計、生物統計などなど、多くの分野で統計学の授業が行われています。
- また、本屋で数学関連書籍の売り場に行きますと、統計学関連の本の割合が非常に大きいことに気づかれると思います。



- それだけではなく、心理学関連書籍のコーナーには心理統計の本、社会学関連書籍のコーナーには社会統計の本といった具合に、統計学に関する書籍は非常に多く出版されています.
- これらの事実は，統計学は様々な分野において重要であると認識されていること，
- そして，好き嫌いは別にして，統計学に関する関心の高さを示していると言えるでしょう.



- さらに，近年では，従来の統計学に留まらず，多くの学問分野との協力により，様々な対象のデータから有益な知見を引き出す，総合的な学問という意味で,「統計科学」という呼び名も使われるようになりました.
- 情報工学に限っても，信号処理やパタン認識，近年では機械学習やデータマイニング等々，多くの分野で統計学的手法が用いられています.



- もしこれまでに，調査や実験のデータを検定して有意差が出たと出ないとか，因子分析をして何々因子が抽出されたとか，…，そういった経験だけから「統計学は面白くない」「もう統計学とは縁を切る」と考えるのは非常にもったいないことです．
- 今回お話するのは，古典的な手法についてですが，現代的な手法の基礎になりますので，そのつもりで聞いていただきたいと思います．

統計モデル

統計モデル

- ばらつきや誤差を含むデータの背後にある規則性，データを発生させる仕組みを数式で表したもの
- 観測誤差を含む観測データから背後にある現象を分析・予測

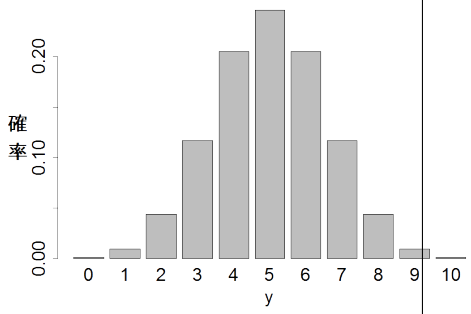
- まず，今回お話する統計モデルをやや大雑把に定義しておきます．
- 統計モデルとは，ばらつきや誤差を含むデータの背後にある規則性や，そのようなデータを発生させる仕組みを数式で表したものと定義されます．
- 統計モデルを利用することにより，観測誤差を含む観測データから，データの背後にある現象を分析したり，現象を予測することができます．

統計モデル

- コインを投げて表が出る確率を θ として，
  N 回コインを投げたときに y 回表が出る確率

$$P(y|N,\theta) = {}_N C_y \theta^y (1-\theta)^{N-y}$$

- 二項分布モデル
- $\theta = 0.5$
- $N = 10$



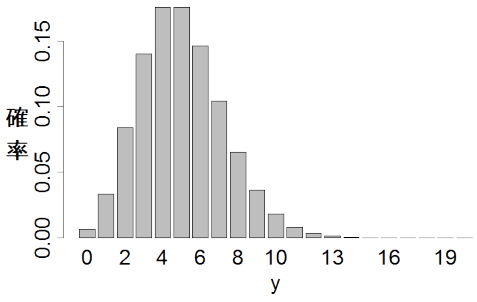
- 例えば，コインを投げて表が出る確率を θ として，N 回コインを投げたときに，y 回表が出る確率は，ここに示すような「二項分布モデル」して表わすことができます．
- ここでは，コインを投げて表が出る確率を 0.5，すなわち $\theta = 0.5$ として，コインを 10 回投げたとします．
- すなわち，$N = 10$ とします．
- このとき，コインの表が出る回数 y は 0 回から 10 回までのいずれかになりますが，それぞれの確率は図のようになります．
- 表が出る確率が 0.5 でコインを 10 回投げるのですから，表が出るのが 5 回前後になる確率が高くなる... ということは直観的かと思います．

統計モデル

- 1 時間に平均 λ 人の客が来る窓口に，
  1 時間に y 人の客が来る確率

$$P(y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

- ポアソン分布モデル
- $\lambda = 5$



- 次の例です．
- 1 時間当たり平均 λ 人の客が来る窓口に，1 時間の間に y 人の客が来る確率を考えます．
- もちろん，二項分布モデルで表わすことができますが，二項分布モデルの θ が小さい時には，このような式で表わされるポアソン分布で近似することができます．
- ポアソン分布については，第 11 回の待ち行列理論で紹介しました．
- ポアソン分布は，このようにヘンテコな式ですが，数学的に便利な性質を持っているため，待ち行列理論以外でもよく利用されます．
- ここでは，1 時間に平均 5 人の客が窓口を訪れるとします．
- すなわち $\lambda = 5$ とします．
- このとき，1 時間に客が y 人来る確率は図のようになります．
- 当然のことながら，平均である 5 人前後になる確率が高くなっています．

統計モデル

- 単位面積当たりの給水量 (x) に対して，植物の収穫量 (y) が，平均 $\beta_0 + \beta_1 x$，分散 σ^2 (標準偏差 σ) の正規分布にしたがう
- 給水量 x のときに収穫量 y の確率密度

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\{y - (\beta_0 + \beta_1 x)\}^2}{2\sigma^2}\right]$$

- 次の例です.
- 単位面積当たりの給水量 x に対して，植物の収穫量 y が，平均が $\beta_0 + \beta_1 x$，分散が σ^2 の正規分布にしたがうとします.
- 給水量 x のときに収穫量 y の確率密度はこの式により表わされます.

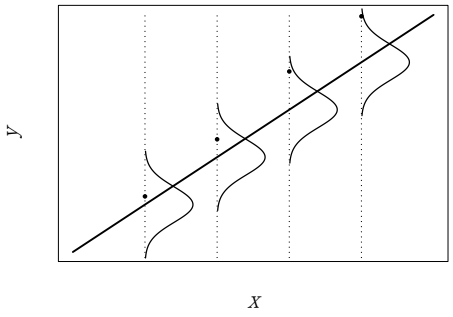
- y は連続量ですので，限りなく高い精度で測定すれば，厳密に特定の値をとる確率は限りなく 0 に近づきますので，確率ではなく確率密度で考えます.
- 確率は，特定の値をとる確率ではなく，特定値の範囲に入る確率として考えます.

統計モデル

- y が，平均 $\beta_0 + \beta_1 x$，分散 σ^2 (標準偏差 σ) の正規分布にしたがう

$$y \simeq \beta_0 + \beta_1 x$$

と近似



- y が，平均 $\beta_0+\beta_1 x$，分散 σ^2 の正規分布にしたがうということは，y は $\beta_0+\beta_1 x$ で近似できることを意味します．
- このことを示したのがこの図です．
- 右上がりの直線は $y = \beta_0 + \beta_1 x$ です．
- 点は特定の x における y の実測値です．
- 横を向いた山のような曲線は，分散が σ^2 の正規分布の確率密度関数で，点 y が直線付近の値になりやすいことを示しています．

統計モデル

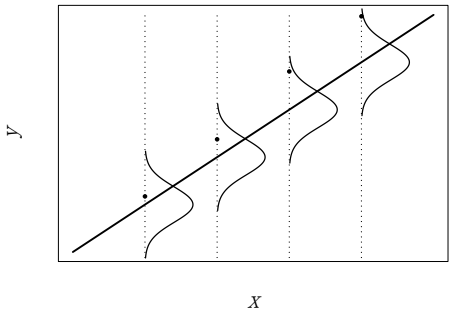
モデルの構築

1. データの背後にある規則性，データを発生させる仕組みを関数で表す
   - 関数の型式は決まる／パラメタの値はわからない
2. 観測データと照らし合わせて最も適切なパラメタの決定 (推定)
   - 非線形の最適化
3. モデルの評価，修正

- 統計モデルを構築するには，まずデータの背後にある規則性，そのようなデータを発生させる仕組みを関数で表します.
- 通常，この時点では，関数の型式は決まるが，パラメタの値はわかりません.

統計モデル

- y が，平均 $\beta_0 + \beta_1 x$，分散 σ^2 (標準偏差 σ) の正規分布にしたがう



$$y \simeq \beta_0 + \beta_1 x$$

と近似

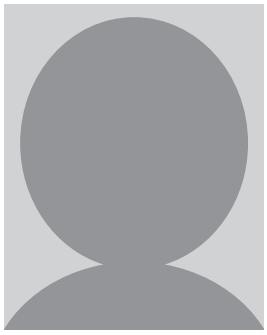
- 直前の例では，データを得た時点で，x と y をプロットしてみれば，x と y の間に近似的に直線関係が成り立つのはすぐに分かるでしょうから，y を $\beta_0 + \beta_1 x$ という 1 次式で近似すること，また，y は $\beta_0 + \beta_1 x$ を平均とする正規分布にしたがうとすることは比較的簡単に決めることができます．
- その意味で，関数の型式は決まります．

統計モデル

モデルの構築

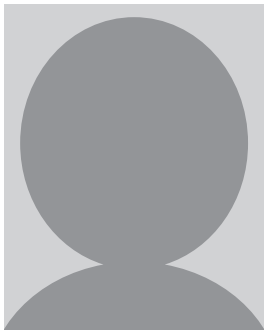
1. データの背後にある規則性，データを発生させる仕組みを関数で表す
   - 関数の型式は決まる／パラメタの値はわからない
2. 観測データと照らし合わせて最も適切なパラメタの決定 (推定)
   - 非線形の最適化
3. モデルの評価，修正

- 一方，パラメタ $\beta_0, \beta 1, \sigma^2$ は，データと照らし合わせて最も適切なパラメタの値を決定，…推定する必要があります．
- パラメタ推定における適切さの基準の考え方はいくつかありますが，いずれの考え方においても非線形の最適化問題になることがほとんどです．



- この科目では，非線形の最適化問題に関しては，線形の最適化問題ほどは取り上げていませんので，今回は，代表的な統計モデルの紹介とともに，パラメタの推定法についてお話します．

最小二乗法によるパラメタ推定

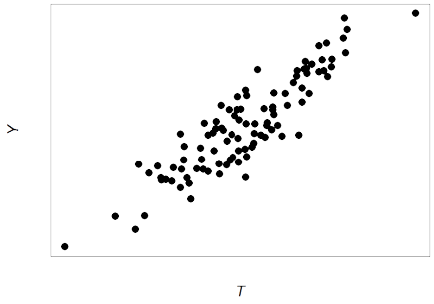


- ここでは，幾つかの統計モデルを紹介し，最小二乗法によりパラメタを推定する方法について説明します.

最小二乗法によるパラメタ推定

線形単回帰モデル

- 1 日の最高気温 T とビールの販売量 Y の関係

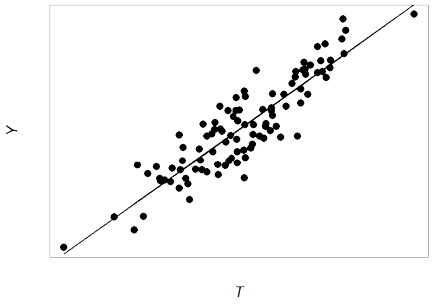


- 第 5 回の在庫管理の講義で，定期発注方式では，需要を予測して発注量を決定し，きめ細かく在庫管理を行うということをお話しました．
- 在庫管理を効率的に行うには，需要を精度良く予測することが重要です．
- そこで，ビールの需要の予測を例に基本的な統計モデルについて説明します．
- 1 日の最高気温 T，以降は単に気温と呼びますが，気温 T とビールの販売量 Y の関係を調べたところ，図に示すような結果が得られました．
- T と Y の間にどのような関係があるでしょうか…という問題です．
- グラフを見れば，気温 T と販売量 Y は直線関係があると考えるのが自然でしょう．

最小二乗法によるパラメタ推定

線形単回帰モデル

- 1 日の最高気温 T とビールの販売量 Y の関係



- こんな感じでしょうか
- しかし，すべての点が 1 本の直線上に乗っている訳ではなく，T と Y の間に厳密な直線関係は成り立たないことも明らかです．
- それでも，誤差があることを認めた上で，T と Y は近似的に直線関係にあると考えるほうが，販売量を予測する上で有効であると思われます．
- このような直線ならデータに良く当てはまっているように見えます．
- しかし，別の傾きや切片，すなわち別の直線の引き方も考えられます．
- T と Y は近似的に直線関係にあると仮定して，直線の傾きと切片はどのように決めるべきか，本当に直線関係があると考えるのが適切か，直線よりふさわしい曲線関係にあるのではないかといった問題が浮上してきます．
- 統計モデルによる解析はこれらの問題を扱います．

最小二乗法によるパラメタ推定

線形単回帰モデル

- 1 日の最高気温 T とビールの販売量 Y の間に近似的に直線関係が成り立ちそう

$$Y = \beta_0 + \beta_T T$$

- 原因となる変数 … **説明変数**
- 結果となる変数 … **目的変数**
- 説明変数と目的変数を結びつけるモデル … **回帰モデル**

- さて，この例では，気温 T とビールの販売量 Y の間に近似的に直線関係

$$y = \beta_0 + \beta_T \text{かける } T$$

  が成り立ちそうです．
- 変数のうち，原因となる変数，ここでは気温 T を説明変数と呼びます．
- 結果となる変数，ここでは販売量 Y を目的変数と呼びます．
- 説明変数や目的変数には，それぞれに色々な別名があります．
- ここではいちいち挙げませんが，この授業以外で別名が出てきても驚かないで下さい．
- そして，説明変数と目的変数を結びつけるモデルを回帰モデルと呼びます．

最小二乗法によるパラメタ推定

線形単回帰モデル

- 説明変数 T と目的変数 Y の間に近似的に

$$Y = \beta_0 + \beta_T T$$

が成り立つと仮定

- n 個の Y の測定値を $Y_1, Y_2, \cdots, Y_n$，
対応する T の値を各々 $T_1, T_2, \cdots, T_n$

$$Y_i = \beta_0 + \beta_T T_i + \varepsilon_i$$

- ε_i は直線近似した時の誤差（残差）

- 説明変数である，気温 T と，目的変数である，販売量 Y の間に近似的に

$$y = \beta_0 + \beta_T \text{かける } T$$

が成り立つと仮定します．

- n 個の Y の測定値を $Y_1, Y_2, \cdots, Y_n$，対応する T の値を各々
$T_1, T_2, \cdots, T_n$ と書くことにしますと，各 i において

$$Y_i = \beta_0 + \beta_T \text{かける } T_i + \varepsilon_i$$

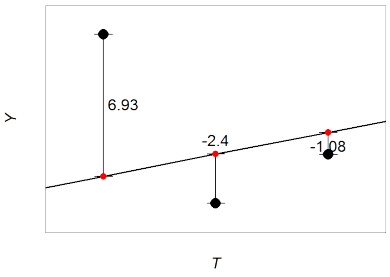
と書くことができます．

- ここで，ε_i は直線近似した時の誤差で残差と呼ばれます．
- 説明変数の 1 次式で表されるモデルは線形回帰モデルと呼ばれ，特に説明変数が 1 個の線形回帰モデルは，線形単回帰モデルと呼ばれることもあります．

最小二乗法によるパラメタ推定

線形単回帰モデル

- 各 ε_i が 0 に近いほど直線がよく当てはまっている



- それでは，いよいよ線形単回帰モデルのパラメタ，β_0, β_T を推定します．
- 各残差 ε_i が 0 に近いほどデータに良く当てはまっている…というのは直観的でしょう．
- 黒い点…黒い丸と言った方がよいかもしれませんが，これらは気温 T_i における Y の測定値 Y_i を表しています．
- 直線は $Y = \beta_0 + \beta_T$かける T で，直線上の赤い小さな点は，T_i における Y_i のモデルによる推定値で，$\beta_0 + \beta_T$かける T_i です．
- Y_i の測定値から推定値を引いたものが残差 ε_i です．
- 図を見るまでもなく，残差は正の値になる場合も負の値になる場合もあります．
- そのため，各残差 ε_i が 0 に近い事の基準を，ε_i の和とすると，正の残差と負の残差が相殺されてしまい，各 ε_i が 0 に近くなることは保証されません．

最小二乗法によるパラメタ推定

線形単回帰モデル

- 各 ε_i が 0 に近いほど直線がよく当てはまっている

$$J(\beta_0, \beta_T) = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} \{Y_i - (\beta_0 + \beta_T T_i)\}^2$$

が最小になるようにパラメタ β_0, β_T を決定
… **最小二乗法**

- そこで，ε_i ではなく ε_i の二乗の和を基準とします．
- ε_i の二乗は非負ですから，ε_i の二乗は足し合わせても相殺されることはありません．
- ε_i の二乗の和である J が最小になるようにパラメタ β_0, β_T を決定して，データに当てはまる直線を得ようというのが最小二乗法です．

最小二乗法によるパラメタ推定

線形単回帰モデル

$$J(\beta_0, \beta_T) = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} \{Y_i - (\beta_0 + \beta_T T_i)\}^2$$

を最小とする β_0, β_T を決定

- J が最小になるためには，

$$\nabla J(\beta_0, \beta_T) = \mathbf{0}$$

- この J を最小化するパラメタ β_0, β_T を求めるのは非線形最適化問題です．
- J が最小になるためには，J の勾配が零ベクトルになる必要があります．
- また，この問題においては，J の勾配ベクトルが零ベクトルになることは J が最小になるための十分条件でもあります．
- J を最小化するパラメタ β_0, β_T を求める問題は非線形最適化問題ですので，ニュートン法などの非線形最適化法を用いて計算する…というのが，一般の回帰モデルにおけるパラメタ推定法になりますが，線形回帰モデルの場合は，そのような繰り返し計算を行わなくても解を求めることができます．

最小二乗法によるパラメタ推定

線形単回帰モデル

$$\begin{aligned}\frac{\partial J}{\partial \beta_0} &= 2n\beta_0 - 2\sum_{i=1}^{n} Y_i + 2\beta_T \sum_{i=1}^{n} T_i = 0 \\ \frac{\partial J}{\partial \beta_T} &= 2\beta_T \sum_{i=1}^{n} T_i^2 - 2\sum_{i=1}^{n} T_i Y_i + 2\beta_0 \sum_{i=1}^{n} T_i = 0\end{aligned}$$

$$\beta_0 = \frac{\sum_{i=1}^{n} T_i^2 \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} T_i Y_i \sum_{i=1}^{n} T_i}{n\sum_{i=1}^{n} T_i^2 - (\sum_{i=1}^{n} T_i)^2}, \quad \beta_T = \frac{n\sum_{i=1}^{n} T_i Y_i - \sum_{i=1}^{n} T_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} T_i^2 - (\sum_{i=1}^{n} T_i)^2}$$

- J をパラメタ β_0, β_T で偏微分すると，上の 2 つの式になります．
- これら 2 つの式の値が 0 になる β_0, β_T の値は，この連立方程式を解けば求めることができ，下の式になります．
- 和の記号が多く，ギョっとするかもしれませんが，計算機を用いれば簡単に計算できます．
- これらの式自体を覚える必要はありません．
- 重要なのは最小二乗法の考え方を理解することです．

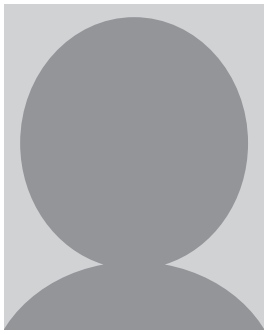
最小二乗法によるパラメタ推定

線形単回帰モデル

$$J(\beta_0, \beta_T) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \{Y_i - (\beta_0 + \beta_T T_i)\}^2$$

を最小とする β_0, β_T を決定

- この式のことです.
- これは，最も単純な線形単回帰モデルの場合ですが，もっと複雑なモデルでも考え方は同様です.



- 今度は，説明変数が複数ある場合について考えます.
- 次のような問題を考えます.

最小二乗法によるパラメタ推定

線形重回帰モデル

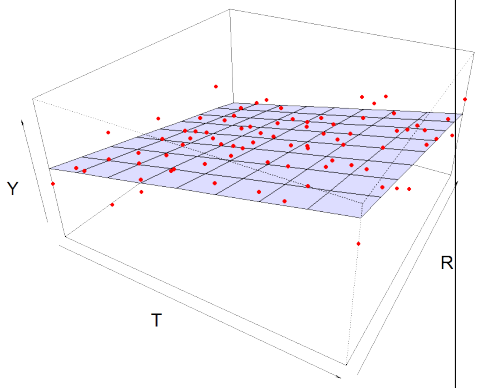
- 1 日の最高気温 T，降水量 R とビールの販売量 Y の間に近似的に
$$Y = \beta_0 + \beta_T T + \beta_R R$$
が成り立つ

- 先ほどの例では，1 日の最高気温とビールの販売量の関係を調べましたが，今回は気温だけでなく，降水量も合わせて，ビールの販売量との関係を調べました．
- その結果，降水量と販売量の間に近似的に直線関係が成り立ちそうであることが分かりました．
- そこで，1 日の最高気温 T，降水量 R を説明変数，ビールの販売量 Y を目的変数として，近似的に
$$Y = \beta_0 + \beta_T \text{かける } T + \beta_R \text{かける } R$$
が成り立つと仮定して，パラメタ $\beta_0, \beta_T, \beta_R$ を推定します．
-

最小二乗法によるパラメタ推定

線形重回帰モデル

- 1 日の最高気温 T,
  降水量 R とビールの
  販売量 Y との関係



- T, R と Y の間に近似的に $Y = \beta_0 + \beta_T$かける $T + \beta_R$かける R が成り立つというのは，幾何的にはこの図のようになります.
- このように方程式 $Y = \beta_0 + \beta_T$かける $T + \beta_R$かける R は平面になります.
- Y の測定値は赤い点で表わされています.

最小二乗法によるパラメタ推定

線形重回帰モデル

- p 個の説明変数 $x_1, x_2, \cdots, x_p$ を用いて，近似的に

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

が成り立つと仮定

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

- 複数の説明変数からなる線形回帰モデル
  … (線形) 重回帰モデル

- さて，パラメタの推定ですが，ここでは一般化して，p 個の説明変数 x_1 から x_p と目的変数 y の間に近似的に

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \text{途中飛ばして} + \beta_p x_p$$

が成り立つと仮定します.

- 複数の説明変数からなる線形回帰モデルのことを重回帰モデルと呼びます.

最小二乗法によるパラメタ推定

線形重回帰モデル

- 重回帰モデルのパラメタ $\beta_0, \beta_1, \cdots, \beta_p$ は

$$J(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})\}^2$$

  を最小化する最小二乗法で求められる

- $\nabla J(\boldsymbol{\beta}) = \mathbf{0}$ を解くと，$\beta_0, \beta_1, \cdots, \beta_p$ が求められる

- 説明変数が p 個になっても考え方は同じです.
- 残差 ε の 2 乗の和を最小にするパラメタ $\boldsymbol{\beta}$ を求めます.
- そのために J の勾配ベクトルが零ベクトルになるパラメタ $\boldsymbol{\beta}$ を求めればいいわけです.
- 変数の数が増えただけで，考え方は単回帰モデルの場合と同じです.
- ただし，変数が多いと見づらいので，ここでは行列を用いて考えます.

最小二乗法によるパラメタ推定

線形重回帰モデル

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ & & & \ddots & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$$\boldsymbol{y} = [y_1, y_2, \cdots, y_n]^T, \quad \boldsymbol{\beta} = [\beta_0, \beta_1, \cdots, \beta_p]^T$$

$$\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]^T$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- y の n 個の観測値を並べた列ベクトル，縦ベクトルを $\boldsymbol{y}$ とします．
- スペースの関係で，行ベクトルの転置というひねった記述になっています．
- このような表記法はこの講義に限らず広く使われますので，覚えておいて下さい．
- 次の行列ラージ $\boldsymbol{X}$ は，第 i 行に y の観測値 y_i に対応する，説明変数 x_{i1} から x_{ip} と 1 を並べた n 行，$p+1$ 列の行列です．
- パラメタ β_0 から β_p を並べた列ベクトルを $\boldsymbol{\beta}$ とします．
- y の n 個の観測値と推定値の誤差，残差を並べた列ベクトルを $\boldsymbol{\varepsilon}$ とします．
- そうしますと，n 個の観測値に関して，$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ と非常にすっきりした形で書くことができます．
- $\boldsymbol{\beta}\boldsymbol{X}$ でなく $\boldsymbol{X}\boldsymbol{\beta}$ となっているのは，ベクトルは列ベクトルを基本とするという記述に関する慣習，行列の掛け算の制約，それから記述の簡潔性の都合です．
- $\boldsymbol{\beta}$ を $\boldsymbol{X}$ の左に書くこともできますが，そうすると転置が何度も出てきて，ややこしくなってしまいます．
- この式では，切片にあたる β_0 も $\boldsymbol{X}\boldsymbol{\beta}$ に組み込まれています．
- 行列 $\boldsymbol{X}$ の左端の列に 1 が並んでいるのは，このためで，非常にスマートな表現になっています．

最小二乗法によるパラメタ推定

線形重回帰モデル

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$J(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

$$\nabla J(\boldsymbol{\beta}) = -2\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{0}$$

を解くと

$$\boldsymbol{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

- 行列による表記ができましたので，いよいよ最小二乗法によりパラメタ $\boldsymbol{\beta}$ の推定を行います．
- 重回帰モデルになっても，残差 ε の二乗の総和を最小にするという考え方は同じです．
- ベクトル $\boldsymbol{\varepsilon}$ を用いれば，残差の総和 J は $\boldsymbol{\varepsilon}$ の転置と $\boldsymbol{\varepsilon}$ をかけたもの，すなわちベクトル $\boldsymbol{\varepsilon}$ 同士の内積になります．
- 内積ですから J の値はスカラーになります．
- J を $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{\beta}$ で表わすと，ベクトル $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$ 同士の内積となります．
- 和の記号が消えて，J もすっきりした形で表わすことができます．
- J の勾配ベクトルは次のような式になります．
- 大雑把な解釈ですが，J におけるベクトル $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$ 同士の内積は，スカラー y,x, β における $y - \beta x$ の 2 乗のようなものと思えば，その β による偏微分らしい式になっていることが分かります．
- あとは，行列の簡単な演算によりパラメタ $\boldsymbol{\beta}$ を求めることができます．
- 行列による表現により，先ほどの単回帰モデルのパラメタの式より分かりやすくなりました．
- もちろん，先ほどの単回帰モデルも行列表現にすれば，同じ式によりパラメタ $\boldsymbol{\beta}$ を求めることができます．

最小二乗法によるパラメタ推定

モデルの評価

$$\boldsymbol{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

- $\boldsymbol{X}$ を構成するベクトルが線形従属
  ＝ 1 つの説明変数が他の説明変数の線形和
  … $\boldsymbol{X}^T\boldsymbol{X}$ が逆行列を持たない
- 線形従属に近い状態 … パラメタが不安定
- 説明変数が冗長
- 分散拡大要因 (variance inflation factor; VIF),
  トレランス (tolerance)

- この $\boldsymbol{\beta}$ の式には $\boldsymbol{X}$ の転置と $\boldsymbol{X}$ をかけた行列の逆行列が含まれます.
- 行列 $\boldsymbol{X}$ を構成するベクトルが線形従属になる場合，つまり 1 つの説明変数が他の説明変数の線形和で表わされる場合，逆行列は存在しません.
- 通常，1 つの説明変数が他の説明変数の線形和で完全に表わされることはほとんどなく，多少の誤差を伴います.

- そうすると，逆行列が存在することになり，$\boldsymbol{\beta}$ として何らかの値が得られますが，データを少し追加したり削除したりすると $\boldsymbol{\beta}$ の値が大きく変化する，不安定な値になってしまいます.
- こういうことが起きるのは，説明変数が冗長になっているためですので，説明変数を削除するなど，見直す必要があります.
- その指標として，分散拡大要因やトレランスがありますが，ここではこれ以上深入りしません.

最小二乗法によるパラメタ推定

モデルの評価

- 測定値 $\boldsymbol{y}$ と推定値 $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}$ の相関係数 (重相関係数)

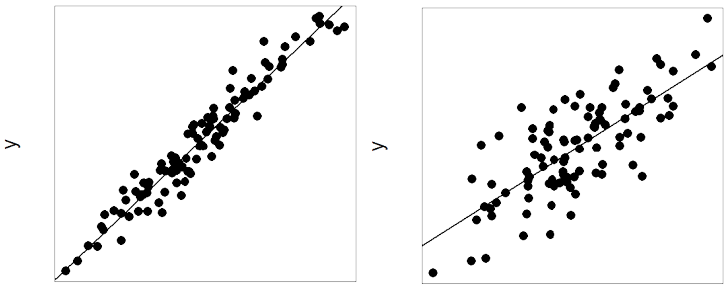
$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

- $-1 \leq R \leq 1$ $\cdots$ 増減をともにする程度
- 決定係数 R^2 $\cdots$ y の変動をモデルで説明できる割合

- モデルのデータへ当てはまりの評価のために，y の測定値とモデルによる y の推定値ラージ $\boldsymbol{Y}$ の相関係数，特に重回帰モデルの推定値と測定値の相関係数については重相関係数と呼びますが，相関係数を考えます．
- 相関係数はこの式により求めることができます．
- ここで，y バーとラージ Y バーは y およびラージ Y の平均値を表します．
- 式から分かるように，相関係数は $\boldsymbol{y}$ とラージ $\boldsymbol{Y}$ が増減を共にする程度を表します．
- 重相関係数が 1 に近ければ推定値が測定値によく当てはまっていると言えます．
- 重相関係数の 2 乗を決定係数と呼びます．
- 決定係数は y の変動をモデルで説明できる割合という意味を持っています．

最小二乗法によるパラメタ推定

モデルの評価



$R^2 = 0.94$ $R^2 = 0.64$

- 重相関係数ないしは，決定係数とモデルの当てはまりの例を見てみましょう．
- これらの図において，横軸はモデルによる y の 推定値，縦軸は y の測定値を表しています．
- 左の例では，ほとんどの点が直線付近に並んでいます．
- 決定係数は 0.94 で y の変動の 94%がモデルにより説明されています．
- 一方，右の例では，直線から大きく逸脱した点も多く，モデルによる推定値があまり当てはまっていないことが分かります．
- 決定変数は 0.64 で左の例よりはだいぶ小さくなっています．

最小二乗法によるパラメタ推定

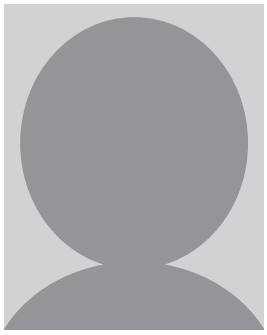
モデルの評価

- $-1 \leq R \leq 1$ … 増減をともにする程度
- 決定係数 R^2 … y の変動をモデルで説明できる割合
- 説明変数の数が多いほど R^2 は大きい
  … 説明変数が多過ぎると予測精度低下
- 自由度調整済み決定係数

$$\bar{R}^2 = 1 - \frac{n-1}{n-p-1}(1-R^2)$$

説明変数の数をペナルティ

- 決定係数は説明変数の数を増やせば増加します.
- しかし，説明変数が多過ぎるとモデルの予測精度が低下してしまいます.
- そのため，モデルの選択指標としては，決定係数そのものではなく，説明変数の数 p をペナルティとした自由度調整済み決定係数を用います.
- モデルの評価指標は他にも色々ありますが，ここではこれくらいにしておきます.

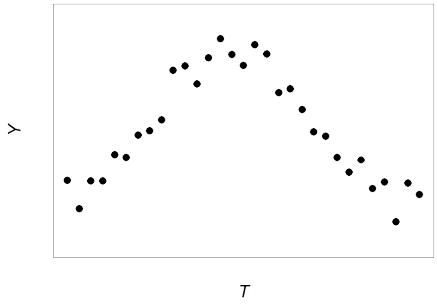


- 今度は，非線形回帰モデルを取り上げます.
- 非線形の回帰モデルでも考え方は，線形回帰モデルの場合と同じです.
- それでは，早速，例を見てみましょう

最小二乗法によるパラメタ推定

非線形回帰モデル

- 1 日の最高気温 T とビールの販売量 Y の関係

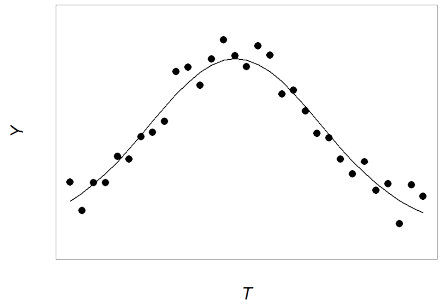


- また，1 日の最高気温 T とビールの販売量 Y の関係について考えます．
- 最初の例では，気温 T と販売量 Y は直線関係にあり，1 日の最高気温 T が高くなるほど，ビールの販売量 Y が多くなるというモノでした．
- 現実的には，気温 T と販売量 Y の間に直線関係が成り立つのは限られた範囲であり，その範囲で線形回帰モデルで予測を行うのは何の問題もありません．
- しかし，より広い範囲で考える場合，気温 T が高すぎると，ビールの販売量 Y は増加しなくなり，さらに気温が高くなると販売量が減少するということもあるでしょう．
- このような非線形な関係を扱うには，非線形回帰モデルを用います．

最小二乗法によるパラメタ推定

非線形回帰モデル

- 1 日の最高気温 T とビールの販売量 Y の関係



- 1 日の最高気温 T とビールの販売量 Y の関係としては，このような曲線関係が成り立ちそうです.

最小二乗法によるパラメタ推定

非線形回帰モデル

- Y と T の間に近似的に

$$Y = f(T|\boldsymbol{\beta})$$

  が成り立つと仮定
- $\boldsymbol{\beta}$ はパラメタのベクトル
- n 個の Y の測定値を $Y_1, Y_2, \cdots, Y_n$,
  対応する T の値を各々 $T_1, T_2, \cdots, T_n$

$$Y_i = f(T_i|\boldsymbol{\beta}) + \varepsilon_i$$

- 非線形回帰モデルでも線形回帰モデルと考え方は同じです.
- 目的変数 y と説明変数 T の間に近似的に関数関係 $y = f(T|\boldsymbol{\beta})$ が成り立つとします.
- ここでは，関数を具体的に定めず $f(T|\boldsymbol{\beta})$ としていますが，実際には具体的な関数の型式を定めます.

最小二乗法によるパラメタ推定

非線形回帰モデル

$$Y_i = f(T_i|\boldsymbol{\beta}) + \varepsilon_i$$

- パラメタ $\boldsymbol{\beta}$ は線形回帰モデルの場合と同様に

$$J(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{Y_i - f(T_i|\boldsymbol{\beta})\}^2$$

  を最小化する最小二乗法で求めることができる
- 一般にはニュートン法などの非線形最適化法を用いて数値的に解を求める

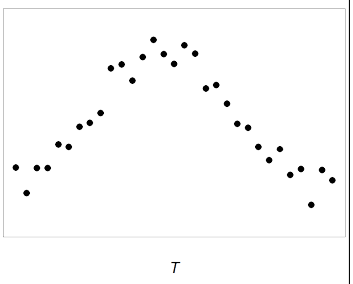
- モデルの推定誤差 ε_i の 2 乗の和を J として，J を最小化するパラメタ $\boldsymbol{\beta}$ を決定するのが，非線形回帰モデルの最小二乗法によるパラメタ推定法です．
- ここまでは線形回帰モデルのパラメタ推定と同じですが，非線形回帰モデルのパラメタ推定では，線形回帰モデルのようにパラメタを $\boldsymbol{T}$ と $\boldsymbol{Y}$ の式として求めることは一般にはできません．
- パラメタは，ニュートン法などの非線形最適化法を用いて数値的に求める必要があります．

最小二乗法によるパラメタ推定

非線形回帰モデル

$$Y = f(T|\boldsymbol{\beta}) = \beta_1 \exp\left(-\frac{(T-\beta_2)^2}{\beta_3^2}\right)$$

$$Y_i = \beta_1 \exp\left(-\frac{(T_i-\beta_2)^2}{\beta_3^2}\right) + \varepsilon_i$$



- このグラフはほぼ左右対称で釣鐘型の形状をしています.
- そこで，今回は上の式に示すガウス関数で近似することを考えます.
- ガウス関数は正規分布の密度関数に似ていますが，正規分布の密度関数はガウス関数の一種です.
- ちなみに，正規分布のことをガウス分布とかガウシアン分布と呼ぶこともあります.
- Y の各測定値を Y_i，それに対応する T を T_i，残差を ε_i として，下の式が成り立つとします.

最小二乗法によるパラメタ推定

$$J(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} \left\{ Y_i - \beta_1 \exp\left(-\frac{(T_i - \beta_2)^2}{\beta_3^2} \right) \right\}^2$$

$$\nabla J(\boldsymbol{\beta}) = \left[\frac{\partial J}{\partial \beta_1} \quad \frac{\partial J}{\partial \beta_2} \quad \frac{\partial J}{\partial \beta_3} \right]^T$$

$$\nabla^2 J(\boldsymbol{\beta}) = \begin{bmatrix} \frac{\partial^2 J}{\partial \beta_1^2} & \frac{\partial^2 J}{\partial \beta_1 \partial \beta_2} & \frac{\partial^2 J}{\partial \beta_1 \partial \beta_3} \\ \frac{\partial^2 J}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 J}{\partial \beta_2^2} & \frac{\partial^2 J}{\partial \beta_2 \partial \beta_3} \\ \frac{\partial^2 J}{\partial \beta_3 \partial \beta_1} & \frac{\partial^2 J}{\partial \beta_3 \partial \beta_2} & \frac{\partial^2 J}{\partial \beta_3^2} \end{bmatrix}$$

- パラメタ $\boldsymbol{\beta}$ は，残差の 2 乗の和である J が最小になるように決定します．
- J の式はちょっと複雑です．
- J の最小化をニュートン法で行う場合，J の β_1, β_2, β_3 による偏微分である勾配ベクトル．
- それから，2 回の偏微分であるヘッセ行列を計算する必要があります．
- この J であれば，ちょっと頑張れば，偏微分の計算はできますが，それでも面倒です．
- そういう場合，偏微分は数値的に求めることができます．

最小二乗法によるパラメタ推定

非線形回帰モデル

- $\nabla J(\boldsymbol{\beta})$ の計算

$$\frac{\partial}{\partial \beta_1} J(\beta_1, \beta_2, \beta_3) = \lim_{\Delta \to 0} \frac{J(\beta_1 + \Delta, \beta_2, \beta_3) - J(\beta_1, \beta_2, \beta_3)}{\Delta}$$

- Δ を 0 に近い数として差分近似

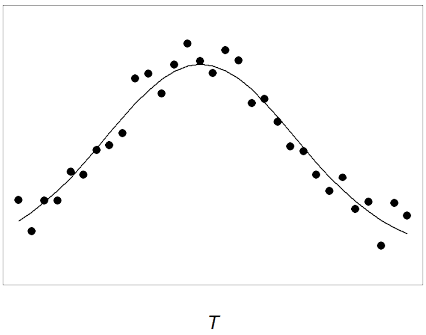
$$\frac{\partial}{\partial \beta_1} J(\beta_1, \beta_2, \beta_3) \simeq \frac{J(\beta_1 + \Delta, \beta_2, \beta_3) - J(\beta_1, \beta_2, \beta_3)}{\Delta}$$

- J の β_1 による偏微分は上の式のように定義されます.
- Δ は限りなく 0 に近づけますが，0 に近い定数として，下の式のような差分を用いることで偏微分を近似します.
- β_2, β_3 による偏微分も同様の方法で近似できます.
- ヘッセ行列の成分である 2 階の偏微分についても，差分で近似することができますし，準ニュートン法を用いて直接ヘッセ行列を計算しないで最適化を図ることもできます.

最小二乗法によるパラメタ推定

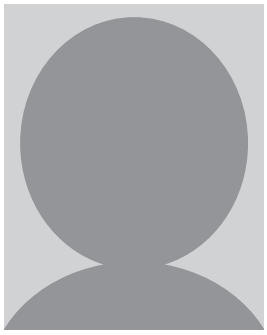
非線形回帰モデル

$$y = \beta_1 \exp\left(-\frac{(T-\beta_2)^2}{\beta_3^2}\right)$$



- このガウス関数モデルのパラメタを推定した結果，図のような曲線が得られました．
- 今回は説明変数が一つの例を取り上げましたが，説明変数が複数ある非線形回帰モデルでも同様の方法でパラメタを推定することができます．

最尤推定法によるパラメタ推定



- ここまで最小二乗法によるパラメタ推定法についてお話しましたが，最小二乗法が唯一のパラメタ推定法…と言う訳ではありません．
- ここでは最尤推定法と呼ばれる推定法についてお話します．

最尤推定法によるパラメタ推定

尤度 (尤度関数)

- ある統計モデル $f(y|\boldsymbol{\theta})$ を仮定しているときに，特定の観測データが得られる確率 (確率密度)

$$p(y_1, y_2, \cdots, y_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})$$

- パラメタ $\boldsymbol{\theta}$ の関数

$$L(\boldsymbol{\theta}|y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})$$

- パラメタを $\boldsymbol{\theta}$ とする確率関数，y が連続値をとるなら，確率密度関数をモデル $f(y|\boldsymbol{\theta})$ とします．
- このモデルに従うデータを n 個発生させたとします．
- 各データは独立に発生するとします．
- つまり，発生されたデータの値によって，次以降に発生されるデータの値が影響を受けないということです．
- 発生された n 個のデータの値が，$y_1, y_2, \cdots, y_n$ となる確率あるいは確率密度は，データが独立に発生されることから，$f(y_1|\boldsymbol{\theta})$ から $f(y_n|\boldsymbol{\theta})$ までかけ合わせた p となります．
- この値が尤度です．
- モデルのパラメタ $\boldsymbol{\theta}$ の値を変えると，当然，尤度の値は変化します．
- それでは，尤度が最も大きくなる $\boldsymbol{\theta}$ の値は何かと考えると，y_1 から y_n のデータを発生したモデルに設定されていたパラメタ $\boldsymbol{\theta}$ の値であろう…ということになります．
- そこで，発生された n 個のデータの値が，$y_1, y_2, \cdots, y_n$ となる確率あるいは確率密度をパラメタ $\boldsymbol{\theta}$ の関数としたものを尤度関数，あるいは単に尤度と呼びます．

最尤推定法によるパラメタ推定

最尤推定法

- 尤度（尤度関数）

$$L(\boldsymbol{\theta}|y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})$$

- 観測データに統計モデルが「どれぐらい
  あてはまっているか」を表す尺度
- 最尤推定法では，観測データのもとで尤度が最大になるようにパラメタ値を決定

- 尤度は，観測データに，特定のパラメタのもとで，統計モデルが「どれぐらいあてはまっているか」を表す尺度と捉えることができます．
- そこで，データを観測して，そのデータを発生させた統計モデルの未知のパラメタを推定するには，尤度が最大になるようにパラメタの値を決定すればよい…というのが，最尤推定法と呼ばれるパラメタ推定法です．
- 尤度という概念は最初難しく感じるかもしれませんが，最尤推定法による推定値は統計学的に良い性質を持っているので，最尤推定法は理論的に優れていると言えます．
- それでは，次の問題を考えてみましょう．

最尤推定法によるパラメタ推定

- 材料は温度を上げると化学的特性が変化
- 加熱温度 x (°C)，サンプル数 N (個)，そのうち特性が変化したサンプル数 y (個)
- 加熱温度と特性が変化する確率の関係を求めよ

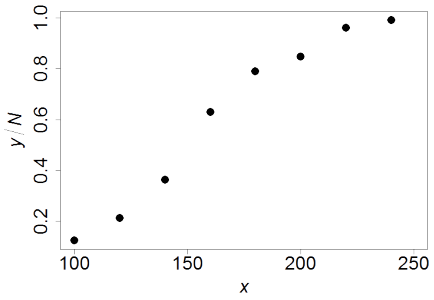
i	1	2	3	4	5	6	7	8
x_i	100	120	140	160	180	200	220	240
N_i	96	99	99	97	100	98	99	100
y_i	12	21	36	61	79	83	95	99

- ある化学会社で新材料を開発しています.
- その材料は温度を上げると化学的特性が変化します.
- 加熱温度と化学的特性が変化する確率の関係を調べるために実験を行いました.
- 加熱温度を x，サンプル数を N ，そのうち化学的特性が変化したサンプル数を y とします.
- 実験結果は表に示す通りです.
- 加熱温度と化学的特性が変化する確率の関係を求めよ…という問題です.
- 温度と化学的特性が変化した割合をプロットしてみます.

最尤推定法によるパラメタ推定

ロジスティック回帰モデル

- 加熱温度 x (°C)
- サンプル数 N (個)
- 特性が変化した
  サンプル数 y (個)

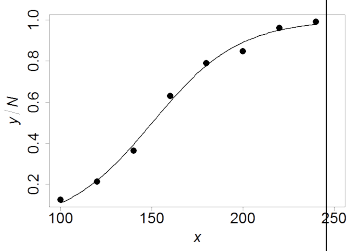


- この図において，横軸は温度，縦軸は化学的特性が変化した割合を表しています．
- 温度を上げると化学的特性が変化した割合が増加していますが，増加の割合が変化して，アルファベットのＳの字のような曲線になっています．
- また，化学的特性が変化した割合は 0 と 1 の間の値を取ります．

最尤推定法によるパラメタ推定

ロジスティック回帰モデル

$$
\begin{aligned}
p(x|\beta_0, \beta_1) &= \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \\
&= \frac{1}{1 + \exp\{-(\beta_0 + \beta_1 x)\}}
\end{aligned}
$$



- このような曲線を表すのに便利なのがここに示すロジスティック関数です.
- ロジスティック関数はパラメタ β_1 が正であれば，x の単調増加関数となり，S 字型の曲線を描きます.
- また，0 から 1 の値をとります.

最尤推定法によるパラメタ推定

ロジスティック回帰モデル

- 加熱温度が x_i (°C) のときに N_i 個のサンプル中 y_i 個の特性が変化する確率は二項分布にしたがう

$$P(y_i|x_i, Ni, P_i) =_{N_i} \mathrm{C}_{y_i} P_i^{y_i} (1 - P_i)^{N_i - y}$$
$$P_i = p(x_i|\beta_0, \beta_1) = \frac{1}{1+\exp\{-(\beta_0+\beta_1 x)\}}$$

- N_i 個の中から y_i 個を選ぶ組み合わせの数

$${}_{N_i}\mathrm{C}_{y_i} = \left(\begin{array}{c} N_i \\ y_i \end{array} \right) = \frac{N_i!}{(N_i - y_i)! y_i!}$$

- それでは，モデル化に移りましょう．
- 加熱温度が x_i のときに N_i 個のサンプル中 y_i 個の化学的特性が変化する確率は，最初の式により表わされる 2 項分布にしたがうとします．
- ただし，右辺のラージ P_i は温度が x_i の時に化学的特性が変化する確率で，2 番目の式により表わされます．
- また，最初の式で，ラージ C に N_i と y_i がついた記号は，N_i 個の中から y_i 個を選ぶ組み合わせの数を表し，一番下の式で求めることができます．
- N_i と y_i が縦に並んで括弧に囲まれた記号も組み合わせの数を表す記号です．
- ベクトルみたいな記号でまぎらわしいのですが，統計学ではこちらの記号がよく用いられますので，注意して下さい．
- このモデルでは，加熱温度が x_i のときに N_i 個のサンプル中 y_i 個の化学的特性が変化する確率が 2 項分布に従い，加熱温度が x のときに化学的特性が変化する確率が x のロジスティック関数で表わされることからロジスティック回帰モデルと呼ばれます．

最尤推定法によるパラメタ推定

ロジスティック回帰モデル

- $i = 1, 2, \cdots, n \quad (n = 8)$ において，
  加熱温度が x_i (°C) のときに N_i 個のサンプル中
  y_i 個の特性が変化する確率

$$f(y_1, y_2, \cdots, y_n | \beta_0, \beta_1) = \prod_{i=1}^{n} {}_{N_i}\mathrm{C}_{y_i} P_i^{y_i} (1 - P_i)^{N_i - y_i}$$

- $f(y_1, y_2, \cdots, y_n | \beta_0, \beta_1)$ をパラメタ β_0, β_1 の関数 $L(\beta_0, \beta_1)$
  $\cdots$ 尤度

- モデル化ができましたので，次に尤度関数を求めます．
- 温度を x_1 から x_8 まで変化させて実験を行いました．
- このときサンプル中で化学的特性が変化したサンプル数が，各々 y_1 から y_8 となる確率は，各値をとる確率をかけ合わせたもので，この式のようになります．
- スペースの都合で，ラージ P_i は x_i の式に展開していないため，式中に β_0 と β_1 が現れていませんが，パラメタは β_0 と β_1 です．
- これをパラメタ β_0，β_1 の関数と見なしたものが尤度関数となります．

最尤推定法によるパラメタ推定

ロジスティック回帰モデル

- 対数尤度関数

$$\log L(\beta_0, \beta_1) = \sum_{i=1}^{n} \log {}_{N_i}\mathrm{C}_{y_i} + \sum_{i=1}^{n} (y_i \log P_i) + \sum_{i=1}^{n} \{(N_i - y_i) \log(1 - P_i)\}$$

- 尤度を最大化するパラメタ β_0, β_1 を求めればいいのですが，最大化の計算を行う時には，尤度の対数をとった対数尤度を最大化する方が計算精度が高くなるので，対数尤度がよく用いられます．
- また，対数尤度は情報量と密接な関係がありますので，尤度の対数という二次的な概念ではなく，対数尤度自体に理論的な意味があります．
- ここではこれ以上述べませんが，興味のある方は調べてみて下さい．

最尤推定法によるパラメタ推定

正規分布モデルと最小二乗法

- 回帰モデル $y = f(\boldsymbol{x}|\boldsymbol{\beta})$ を当てはめるとき，
  y が平均 $f(\boldsymbol{x}|\boldsymbol{\beta})$，分散 σ^2 の正規分布にしたがう

$$g(y|\boldsymbol{x}, \boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\{y - f(\boldsymbol{x}|\boldsymbol{\beta})\}^2}{2\sigma^2}\right]$$

- 最尤推定法によるパラメタの推定値と
  最小二乗法によるパラメタの推定値は一致する

- 最尤推定法が理論的に優れているとお話しました．
- それでは長々と話してきた最小二乗法は何だったんだ…ということになりますが，そのことについてお話します．
- 回帰モデル $y = f(\boldsymbol{x}|\boldsymbol{\beta})$ を当てはめるとき，y が平均 $f(\boldsymbol{x}|\boldsymbol{\beta})$，分散 σ^2 の正規分布にしたがうとします．
-
- y の確率密度関数はこの式になります．
- この時，最尤推定法によるパラメタの推定値は最小二乗法によるパラメタの推定値と一致します．
- この性質は尤度関数を見れば分かります．

最尤推定法によるパラメタ推定

正規分布モデルと最小二乗法

$$
\begin{aligned}
L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\{y_i - f(\boldsymbol{x}_i|\boldsymbol{\beta})\}^2}{2\sigma^2}\right] \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{\sum_{i=1}^{n}\{y_i - f(\boldsymbol{x}_i|\boldsymbol{\beta})\}^2}{2\sigma^2}\right] \\
J &= \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n}\{y_i - f(\boldsymbol{x}_i|\boldsymbol{\beta})\}^2
\end{aligned}
$$

- 尤度関数はこの式のようになります.
- ここで，尤度の指数関数内の分子に注目しますと，これは最小二乗法における残差 ε_i の二乗の和と一致します.
- そのため，最尤推定法によるパラメタの推定値は最小二乗法によるパラメタの推定値と一致します.

最尤推定法によるパラメタ推定

モデル選択

- パラメタ数が多いほど良いモデルという訳ではない
- 予測精度の高いモデル，本質を捉えたモデル
- 様々なモデル選択基準
- 赤池の情報量基準 (AIC)

$$AIC(k) = -2 \text{最大対数尤度} + 2k$$

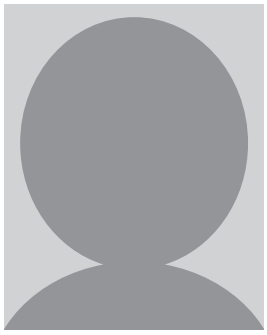
k はパラメタ数
AIC が小さいほど良いモデル

- 最後に再びモデルの評価についてお話します.
- 一般に，同じ型式のモデルであれば，パラメタを増やしていくほど与えられたデータへの当てはまりは良くなっていきます.
- しかし，パラメタ数が多いほど良いモデルという訳ではありません.
- 例えば予測精度の高いモデルや，データの背後にある本質を捉えたモデルというのは，典型的な良いモデルです.
- モデルの目的に合った評価基準に基づき，最適なモデルを選択することになりますが，様々なモデル選択基準が提案されています.
- 有名なものに，予測精度を基準とした赤池の情報量基準 AIC があります.
- AIC はこの式に示しますように，最大対数尤度とモデルのパラメタ数により決まり，AIC が小さいほど良いモデルと言うことになります.
- AIC でもパラメタ数がペナルティとなっており，パラメタ数を増やせば良いという訳ではないことが分かります.

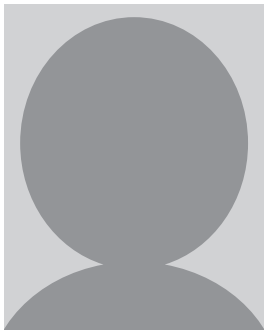
最尤推定法によるパラメタ推定

- コンピュータを内蔵した機器同士がネットワークにより接続され，自律的に連係動作することにより，人の生活を支援する技術と環境
  … **ユビキタスネットワーク**，
  IoT (Internet of Things)
- 多様で大量のデータ
- 役に立つ変数はほんの一部
- モデル選択，正則化

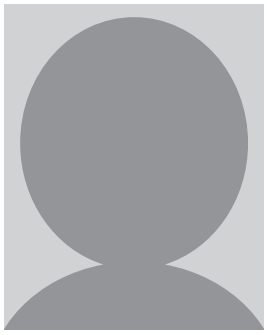
- 現在では，計測機器や通信ネットワークの発展・普及により，多様で大量のデータを簡単に収集できるようになりました．
- 家電製品等，コンピュータを内蔵した機器同士がネットワークにより接続され，自律的に連係動作することにより，人の生活を支援する技術と環境をユビキタスネットワークと呼びますが，2010 年代に入り，物のインターネット，IoT という言葉が使われるようになっています．
- このような多様で大量のデータから有用な情報を引き出そうという試みが盛んになっていますが，興味ある事柄の予測に本当に役立つ変数はほんの一部で，ほとんどの変数はむしろ予測精度を低下させる原因になります．
- そこで，超多数の変数から予測に役立つ変数を効率的に選択して，良いモデルを選択することが重要な問題になります．
- 同じ目的ですが，モデルの複雑さがペナルティになるようにして，役に立たないパラメタが 0 になる仕掛けは正則化と呼ばれており，近年盛んに研究されています．
- モデル選択や正則化については，これ以上お話しできませんが，有効な解析を行うためには重要であるということは知っておいて下さい．



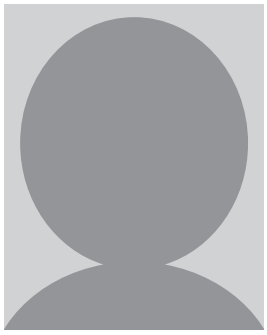
- 今回の講義では，統計モデル，特に回帰モデルとそのパラメタ推定についてお話しました．
- 今回は説明のために，目的変数と説明変数の関係が一目瞭然のデータを用いましたが，統計モデルによる解析の醍醐味は，適切なモデルで解析しないと特性が分からないデータを相手に格闘して，その特性を明らかにしていくことです．



- 近年，計算技術が著しく向上し，複雑なモデルでも容易に計算ができるようになり，以前では手が出なかった複雑なデータでも解析ができるようになっています．
- また，そういった高度な解析を行うためのソフトウエアも，無料かつオープンソースで利用できるものも含めて充実しています．



- 一方，統計学では，こういう場合はこの方法は使えないといったことが多く，実際難しいところもありますが，同じようなことで苦労している人は多く，インターネットを通じて，解決法あるいはヒントとなる情報を得ることもできます．
- ですから，まずは一歩を踏み出して，これらのリソースを用いて，試行錯誤を経て妥当な結論に至るのが，多くの人にとって良い方法であると，私は考えています．



- 最後に一つアドバイスをしておきます.
- それは統計解析に先だって，データを様々な方法で作図することです.
- その際，平均値のみを示すのではなく，できるだけデータの情報を失わないように，生データ，あるいはそれに近い形で表示して，データの分布，条件による変化，外れ値の存在などの情報を読み取ることが重要です.
- このような準備は，適切なモデルを構築するヒントになります.
-