農林統計学

１）確率分布

２）推定論

３）検定論

４）相関と回帰

５）重回帰分析と多変量解析

６）時系列解析

７）スペクトル分析

８）数理計画法

教科書･参考書の紹介

第１限目（10:30〜12:00）

１. 森棟公夫「統計学入門」新世社
２. 水本久夫「統計の基礎」培風館
３. 蓑谷千凰彦「統計学入門」東京図書
４. 宮川公男「基本統計学」有斐閣
５. P.G.ﾎｰｴﾙ著、浅井晃･村上正康訳「初等統計学」培風館

統計学の類型

数理統計　vs　社会統計
記述統計　vs　推測統計
非ベイズ統計　vs　ベイズ統計

統計学と計量経済学

データの整理

データの種類

時系列データ、クロスセクションデータ、パネルデータ

量的情報　と　質的情報（カテゴリカル・データ）

ダミー変数、LOGIT分析、PROBIT分析、TOBIT分析

母集団　と　標本

データの代表値

位置　　平均（mean）・・・・・・算術平均、幾何平均、調和平均

移動平均、加重平均

中位数（median）、

最頻値（mode）

拡がり　分散・・・・・・不偏分散、標準偏差

散布度・・・・　四分位範囲、

四分位分散係数＝四分位範囲／中央値x2

変動係数・・・・四分位範囲CV＝標準偏差／平均

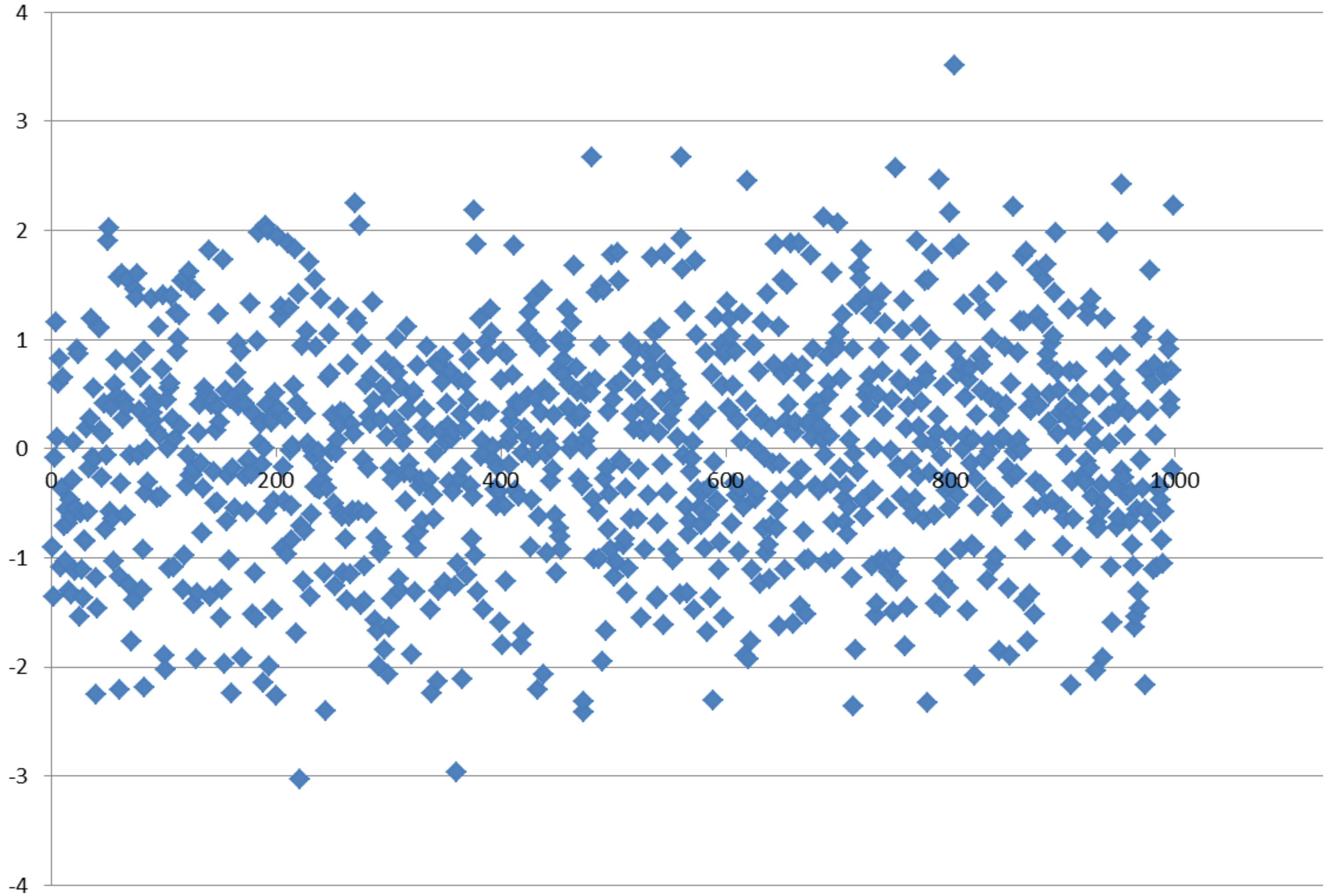
度数分布表（ヒストグラム）

階級、階級値、
度数、相対度数、
累積度数、累積相対度数

・分布の平等度

ローレンツ曲線とジニ係数

ランダムなデータの散布図



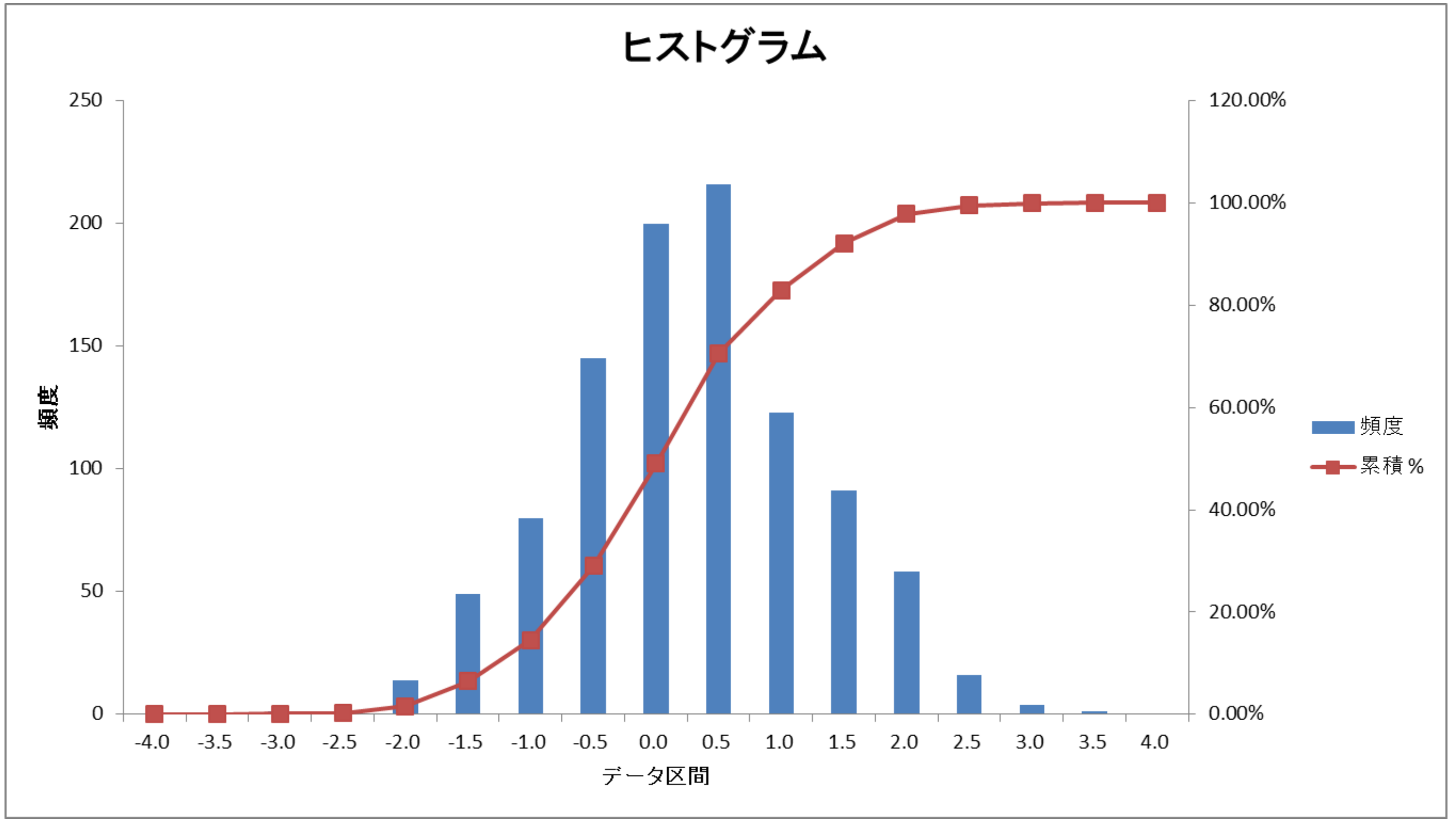
データの整理と統計情報特徴

図によりデータの特徴を視覚的に把握

‐‐度数分布図（ヒストグラム）や散布図など‐‐

0.5℃刻みの階級を横軸にとり、各階級に属する気温変化の標本個数を縦軸に測った1000個の気温変化データの柱状グラフ

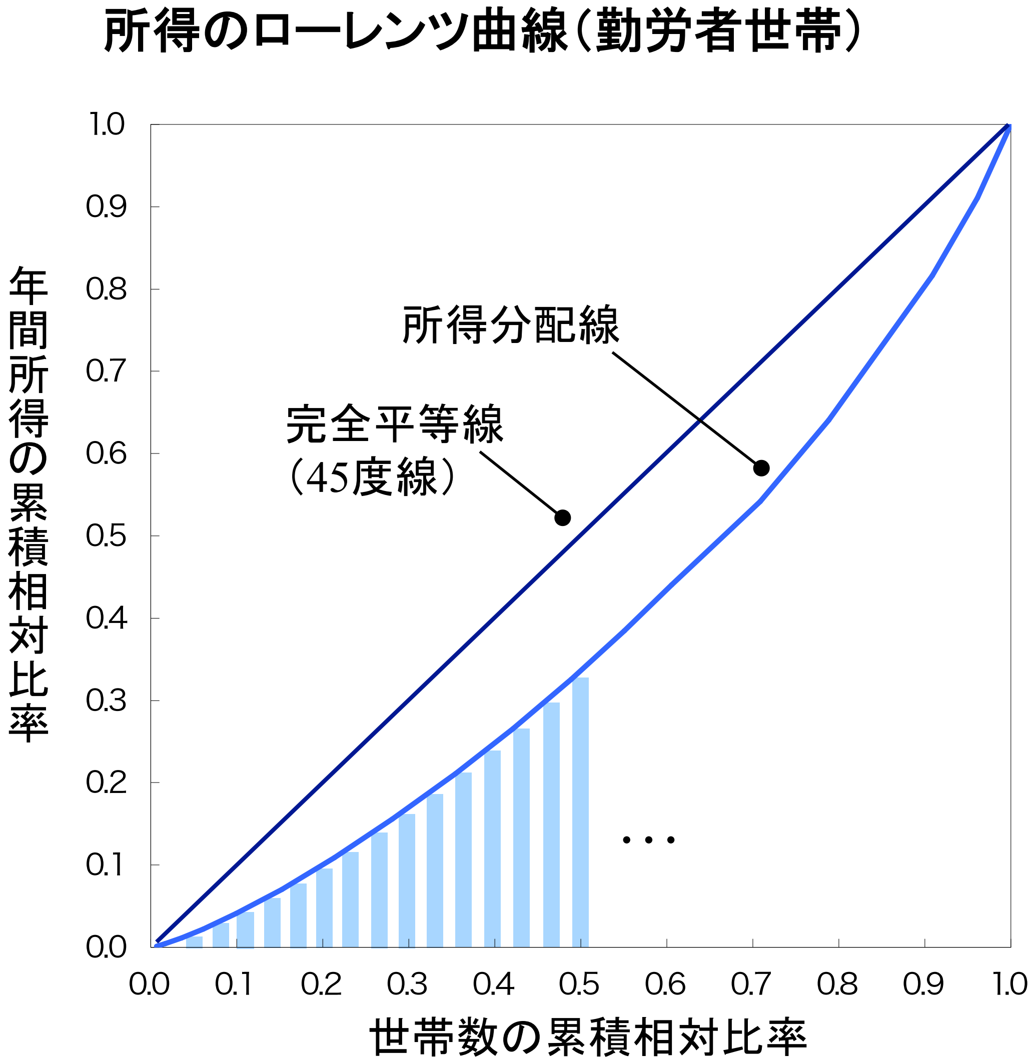
気温変化の度数分布と累積頻度



データ区間	頻度	相対頻度	累積頻度	累積相対頻度
-4.4~-4.0	0	0	0	0.000
-3.9~-3.5	0	0	0	0.000
-3.4~-3.0	1	0	1	0.001
-2.9~-2.5	1	0	2	0.002
-2.4~-2.0	14	0.014	16	0.016
-1.9~-1.5	49	0.049	65	0.065
-1.4~-1.0	80	0.080	145	0.145
-0.9~-0.5	145	0.145	290	0.290
-0.4~0.0	200	0.200	490	0.490
0.1~0.5	216	0.216	706	0.707
0.6~1.0	123	0.123	829	0.830
1.1~1.5	91	0.091	920	0.921
1.6~2.0	58	0.058	978	0.979
2.1~2.5	16	0.016	994	0.995
2.6~3.0	4	0.004	998	0.999
3.1~3.5	1	0.001	999	1.000
3.6~4.0	0	0	1000	1.000

ローレンツ曲線

ーー分布の平等性の検討ーー



独立な事象と条件付確率

- 確率の独立性と従属性
  - 確率の加法定理　$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - 条件付確率　$P(B|A)=P(A \cap B)/P(A)$
  - 確率の乗法定理　$P(A \cap B)=P(B|A)\,P(A)$

↓

  - 独立性
    - $P(A|B) = P(A)$
    - $P(B|A) = P(B)$
    - $P(A \cap B) = P(A)\,P(B)$
- ベイズの公式
  - 条件つき確率を用いて、条件と結果を逆転した確率を求める方法

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

確率変数とその分布

離散変数　と　連続変数

- 確率関数・密度関数と確率分布関数
  - 確率関数（離散型）
  - 密度関数（連続型）
  - 分布関数（確率関数と密度関数の積分形）

基本的な分布関数

１．二項分布

２．ポアソン分布

　　二項分布のポアソン近似

３．一様分布

４．正規分布

– 　標準正規分布

１．二項分布

ベルヌーイ試行
＝試行の結果が2種類しかない実験
P({成功})＝p、
P({失敗})＝1-p

独立なベルヌーイ試行をn回繰り返す時、

「n回の試行のうちで成功した回数」を確率変数Xとすると、

$P(\{X=x\})=p(x)={}_nC_x p^x q^{n-x}$

特定のnとpを伴った二項分布をB(n,p)と表記する。

資料）森棟公夫「統計学入門」新世社

ポアソン分布

１）時間を細かく細分すれば、細分された時間内にその事象が2回以上起きる確率はゼロ、その事象が起きるか起きないかのベルヌーイ試行であると見なせる
２）個々の細分化された時間内に起きる事象は、他の時間内に起きる事象とは独立している。
３）ある細分化された時間内に事象が起きる確率は時間の長さに比例している

全時間s中に事象が起きる回数を確率変数Xとした場合、
n区間のうちx区間で事象が起きる確率は二項分布になり、
λを単位時間内に事象が起きる確率とすると

$$p(x) = {}_nC_x (\lambda s / n)^x (1 - \lambda s / n)^{n-x}$$

ここで、細分化された区間の数nを極限まで増加させると

$$p(x) = \lambda^x \exp(-\lambda) / x! \qquad X = 0, 1, \ldots, \infty$$

出所）森棟公夫「統計学入門」新世社

一様分布

- 確率変数xの密度関数がある連続な区間（a<x<b）で一定、区間外では密度は0であるとする。このような分布を一様分布と呼ぶ。分布のパラメータは区間の両端の値aとbである。密度関数の面積は1になるべきであるから、密度は1/（b-a）となる

$$E(x)= 1/(b-a)\int_a^b x\,dx$$

$$=(a+b)/2$$

$$V(x)=(b-a)^2/12$$

出所）森棟公夫「統計学入門」新世社

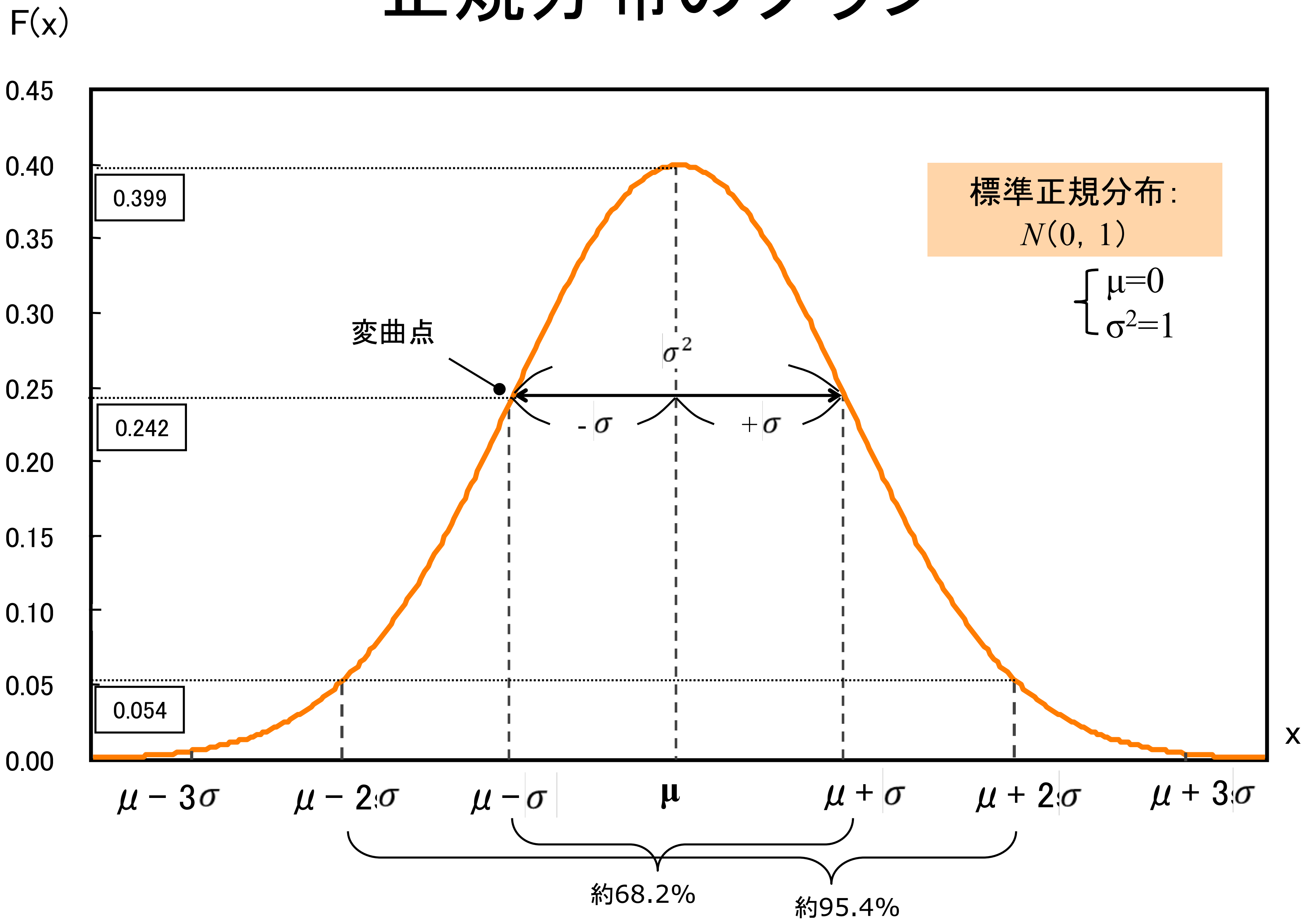
正規分布

- 正規分布（normal distribution）
  - 連続分布の中でも特によく用いられる分布。$N(\mu, \sigma^2)$で表される。
  - とくに、平均0、分散1の正規分布を標準正規分布といい、$N(0, 1)$で表す 。
  - 正規分布は平均値μと分散 σ^2 によって規定される。

正規分布の確率密度関数

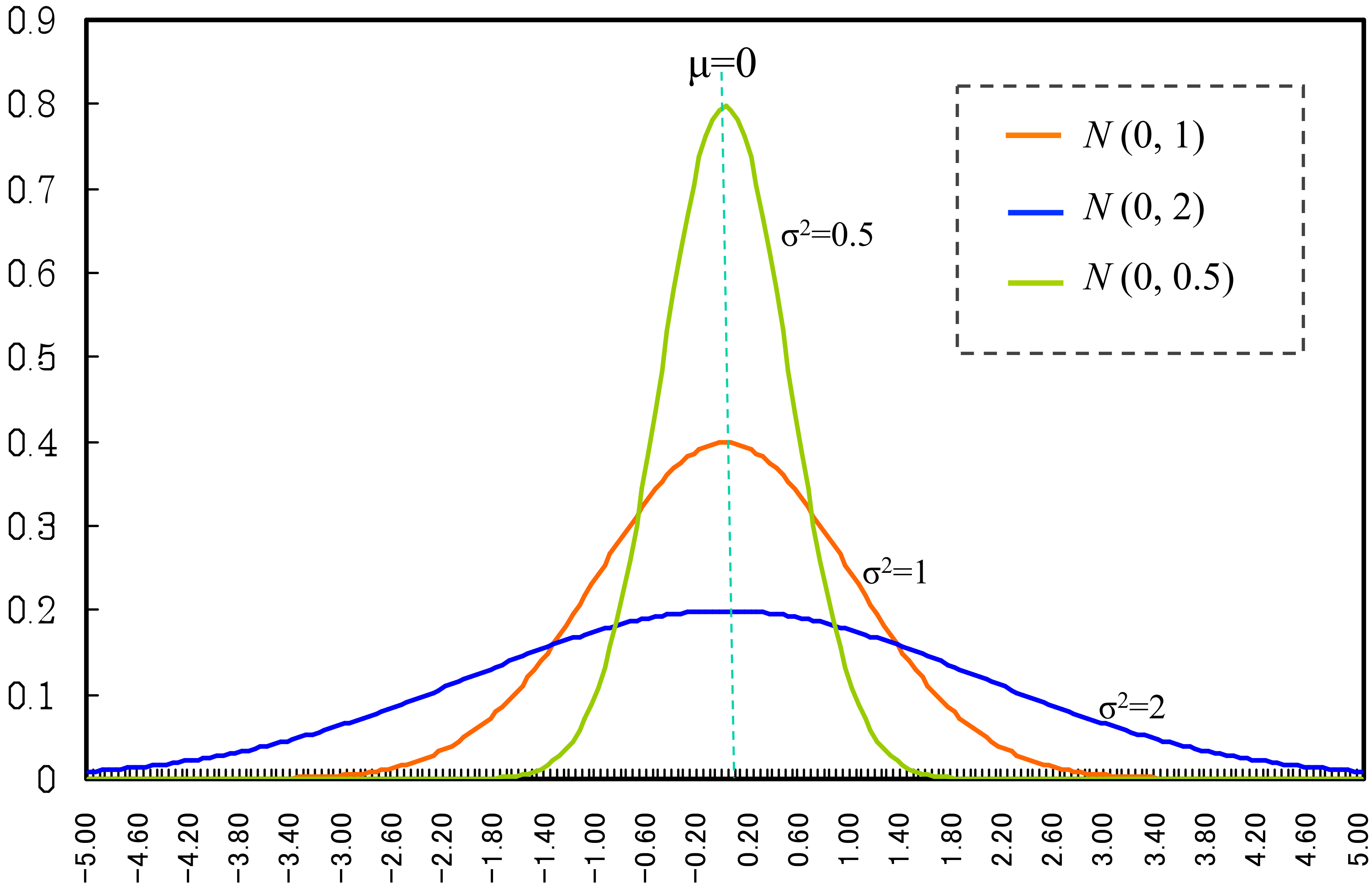
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

正規分布のグラフ

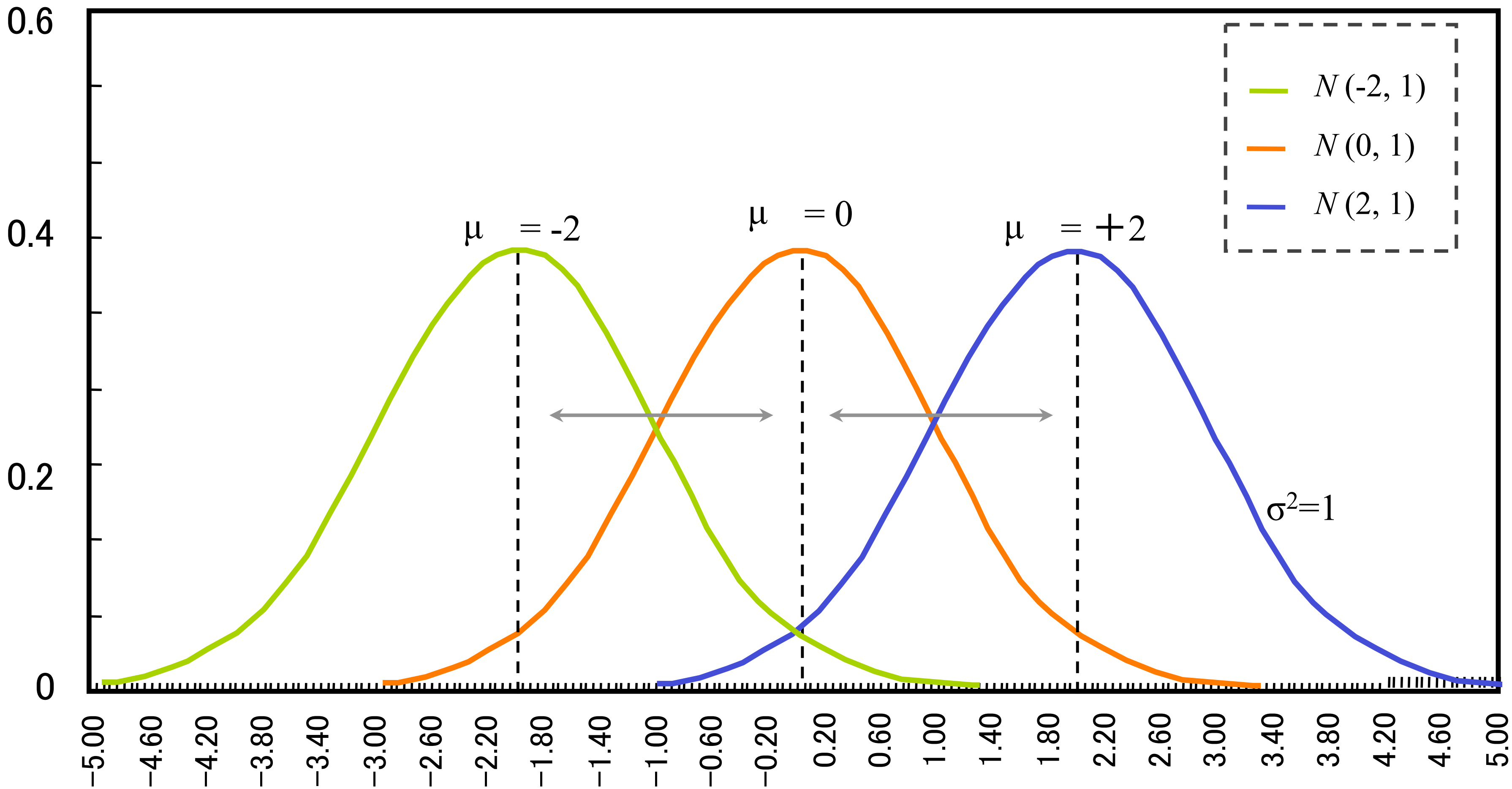


F(x)
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.399
0.242
0.054
変曲点
σ^2
$-\sigma$
$+\sigma$
標準正規分布：
$N(0,\ 1)$
$\mu=0$
$\sigma^2=1$
x
$\mu-3\sigma$
$\mu-2\sigma$
$\mu-\sigma$
μ
$\mu+\sigma$
$\mu+2\sigma$
$\mu+3\sigma$
約68.2%
約95.4%

分散の大きさによる違い



平均値の大きさによる違い



同時確率分布と周辺確率分布

- 共分散　と　相関係数

$$\mathrm{COV}(X、Y) = E\{(X - \mu_x)(Y - \mu_y)\}$$

ここで、μ_x、μ_yは確率変数X,Yの平均値

$$R = \frac{COV(X,Y)}{\sqrt{V(X)V(Y)}}$$

分布の独立性

標本分布

１. 無作為抽出と無作為標本

２. 標本平均の分布

（定理１）無作為標本からの標本平均の分布

（定理２）正規分布の再生性

（定理３）チェビシェフの不等式

（定理４）大数の弱法則

（定理５）大数の強法則

（定理６）中心極限定理

（ポアソン確率の正規近似）

（二項確率の正規近似）

（定理７）標本分散の分布

資料）森棟公夫「統計学入門」新世社

定理１　標本平均の分布

確率変数Xの平均はμ、分散はσ^2であるとする。Xに関する大きさnの無作為標本 $\{X_1, X_2, \ldots, X_n\}$ より、

標本平均

$$\bar{X} = \left\{X_1 + X_2 + \cdots + X_n\right\} / n$$
$$= \sum_{i=1}^{n} X_i / n$$

を求めると、 $\bar{X}$の平均値はμ、分散は σ^2/nとなる。

資料）森棟公夫「統計学入門」新世社

定理 2　正規分布の再生性

独立な確率変数 $X_i,\ i = 1, 2, \ldots, m$が、各 i について平均値μ、分散が σ_i^2の正規分布に従って分布しているとする。 そのとき、各々の確率変数の和 $\sum_{i=1}^{m} X_i$ は、もとの平均値の和 $\sum_{i=1}^{m} \mu_i$ を平均とし、もとの分散の和$\sum_{i=1}^{m} \sigma^2{}_i$ を分散とする正規分布に従って分布する。

資料）森棟公夫「統計学入門」新世社

定理 3　チェビシェフの不等式

標準化された確率変数Zは、次の不等式を満たす。

$$P\left(\left\{\left|Z\right| \geq \lambda\right\}\right) \leq 1/\lambda^2$$

出所）森棟公夫「統計学入門」新世社

定理 4　大数の弱法則

確率変数 Xに関する大きさがnの無作為標本$\{X_1, X_2, \ldots, X_n\}$があり、$X$の平均はμ、分散は σ^2とする。任意の正の定数 cについて、

$$\lim_{n \to \infty} P\left(\left\{\left|\bar{X} - \mu\right| \geq c\right\}\right) = 0$$

となる。ただし $\bar{X}$ は標本平均とする。

資料）森棟公夫「統計学入門」新世社

定理 5　大数の強法則

確率変数 Xに関する大きさがnの無作為標本$\{X_1,\ X_2, \ldots,$ $X_n\}$があり、$E(x) = \mu$ とする。任意の正の定数 cについて、

$$\lim_{n \to \infty} P\left(\left\{\left|\bar{X} - \mu\right| \geq c\right\}\right) = 0$$

となる。ただし $\bar{X}$ は標本平均とする。

出所）森棟公夫「統計学入門」新世社

定理 6　中心極限定理

大きさがnの無作為標本 $\{X_1, X_2, \ldots, X_n\}$ があり、各X_i，$i = 1, 2, \ldots, n$ の平均はμ、分散が σ^2 であるとする。このとき、標本平均 $\bar{X} = \sum_{i=1}^{n} X_i / n$ を標準化した統計量

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2 / n}}$$

の分布は、nが大であれば標準正規分布で近似できる。

出所）森棟公夫「統計学入門」新世社

定理 7　標本分散の分布

正規母集団 $n(\mu, \sigma^2)$ からの大きさ n の無作為標本を $\{X_1, X_2, \ldots, X_n\}$ とすると、

$$Y = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$$

は自由度が $n-1$ の χ^2 分布（カイ2乗分布）に従って分布する。

資料）森棟公夫「統計学入門」新世社

検定用の複合的な確率分布

１）χ^2分布

　　標準正規変量の二乗和の分布

２） t 分布

　　標準正規変量とχ^2変量の比の分布

３）F分布

　　χ^2変量同士の比の分布

資料）森棟公夫「統計学入門」新世社

χ^2分布

χ^2分布は、互いに独立な標準正規確率変数の2乗和の分布で、そのパラメーターは、独立な確率変数の数、あるいは自由度である。Z_1からZ_kを独立に分布する標準正規確率変数とすると、

$$W = Z_1^2 + Z_2^2 + \ldots + Z_k^2$$

が自由度kの χ^2確率変数で、kの値によって次の図のように分布している。

資料）森棟公夫「統計学入門」新世社

χ^2 分布　つづき

Wの平均値は、Wの右辺の個々の平均値を求めて足し合わせればよいが、たとえばZ_1の項の平均値は、Z_1の分散だから1である。したがってWの平均値は自由度kとなる。Wの分散は各Zが独立に分布しているので、Zの2乗の分散をk倍して求まる。2乗の分散は、標準正規確率変数の性質［尖度E(z^4)=3］によって、

$$E(Z_1^2-1)^2 = 2$$

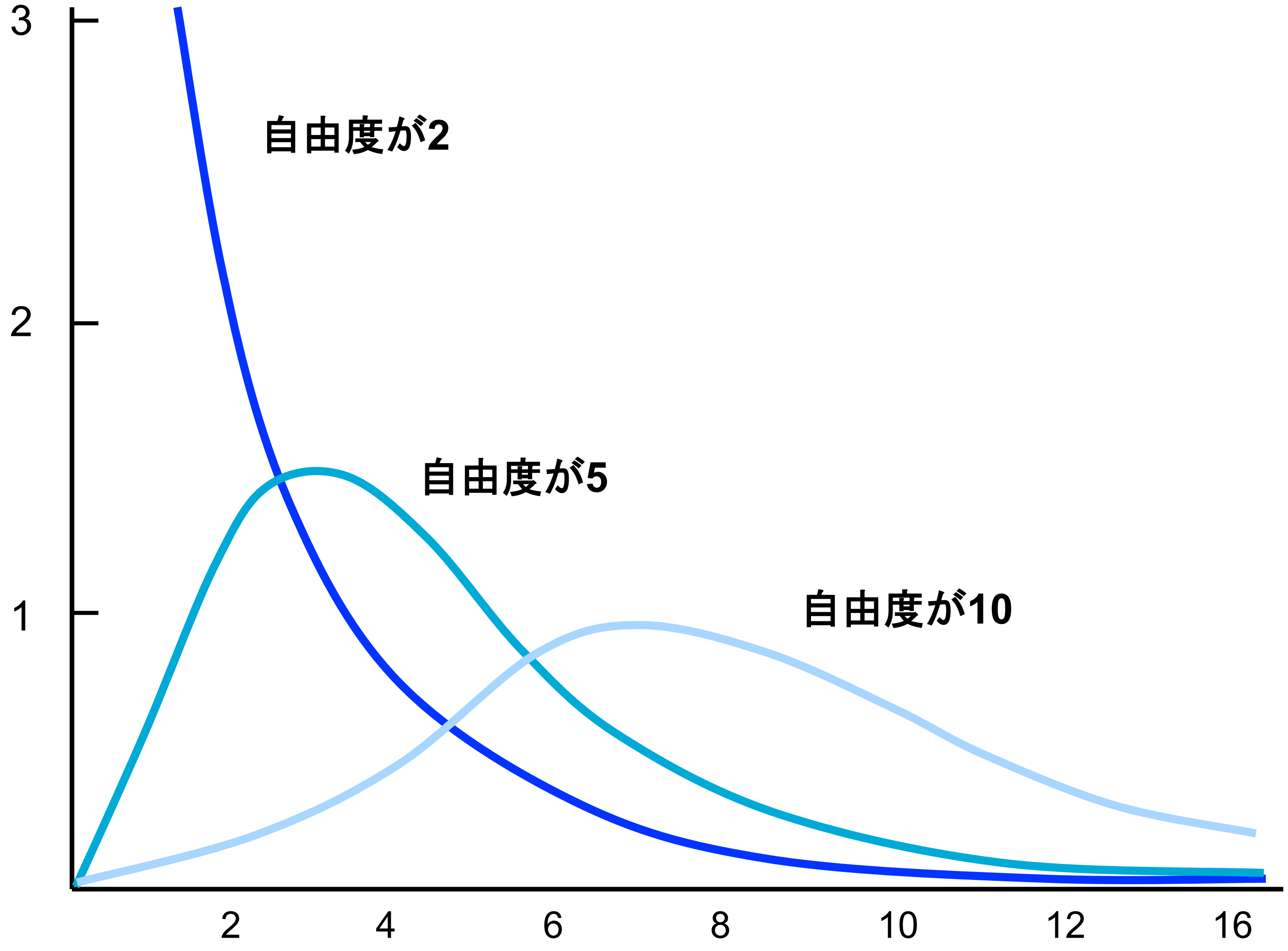
となる。したがって$V(W)$は$2k$、つまり自由度の2倍である。
標準正規確率変数ではなく、X_i が平均μ、分散 σ^2の正規確率変数だと、

$$\frac{\sum_{i=1}^{k}(X_i-\mu)^2}{\sigma^2}$$

が自由度kのχ^2確率変数になる。

資料）森棟公夫「統計学入門」新世社

図:χ^2分布



資料）森棟公夫「統計学入門」新世社

t 分布

Zを標準正規確率変数、Wを自由度がkのχ^2 確率変数とする。さらに、Zと W が独立に分布していれば

$$t = \frac{Z}{\sqrt{W / k}}$$

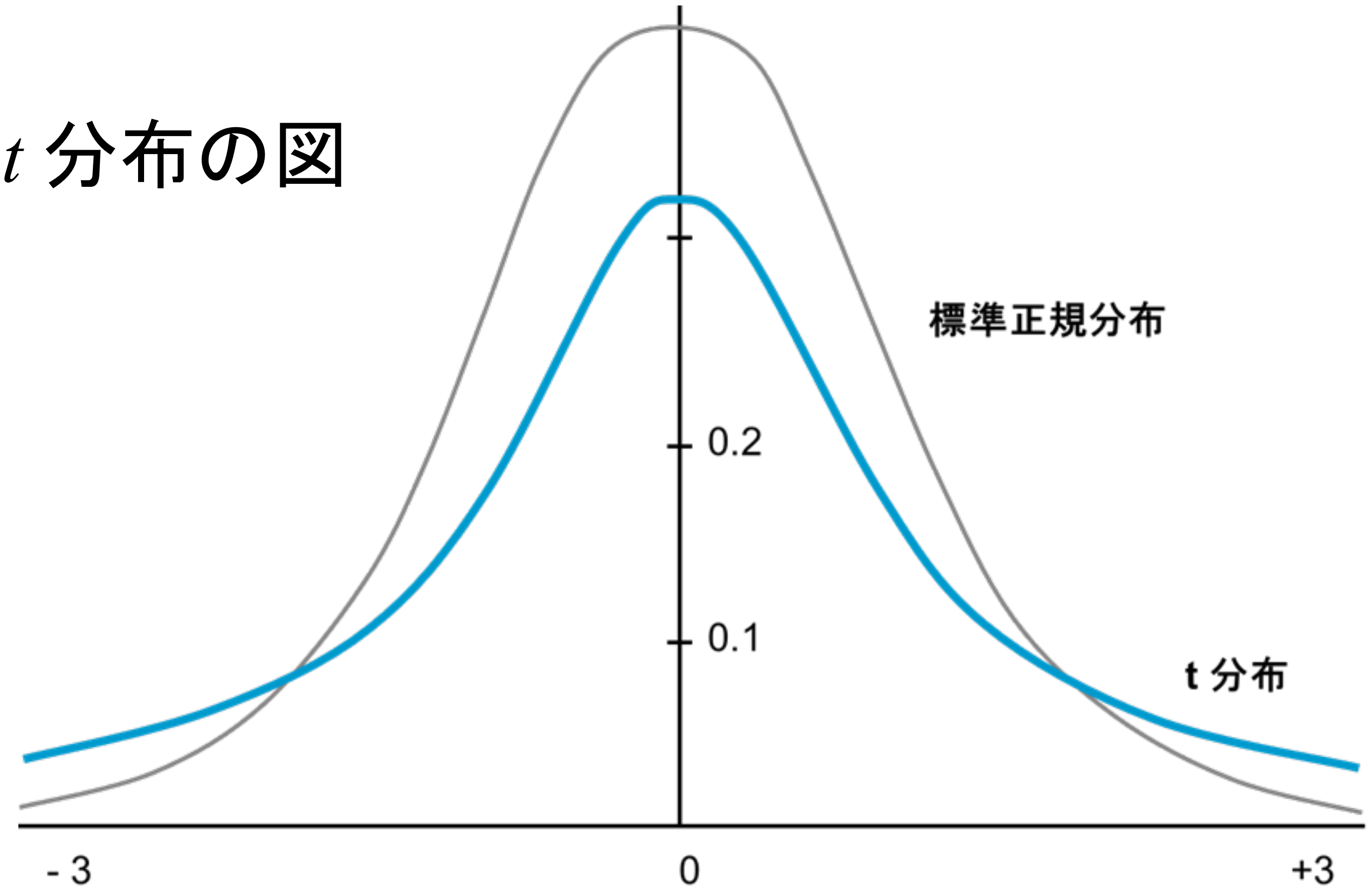
は自由度がkの t 確率変数であり、その分布を「自由度kの **t 分布**」という。上の式を **t 比** と呼ぶことがある。

t分布は原点に関して対称な分布になっている。自由度が1のときは平均値も分散も存在しない。自由度が2のときは平均値は存在するが分散が存在しない。自由度が3以上になって初めて平均値も分散も存在する。

資料）森棟公夫「統計学入門」新世社

t 分布の平均値は0、分散は $k/(k-2)$となるから、分散は1よりも多少大きい。自由度がある程度大きければ、分散はほとんど1と変わらない。図からも自由度が大きくなれば分布の裾が薄くなり、散らばりが小さくなることが理解できる。

t 分布の図



資料）森棟公夫「統計学入門」新世社

t分布の正規近似

- t比の分母は対数の法則によりnが大きければ1に近づくから、nが大なら比は分子の標準正規確率変数と殆ど変らなくなる。従ってt比も分布は自由度が大きければ標準正規分布で近似できる。Zを標準正規確率変数とすれば、

$$P(\{t \leqq c\}) \fallingdotseq P(\{Z \leqq c\})$$

となる。

資料）森棟公夫「統計学入門」新世社

F 分布

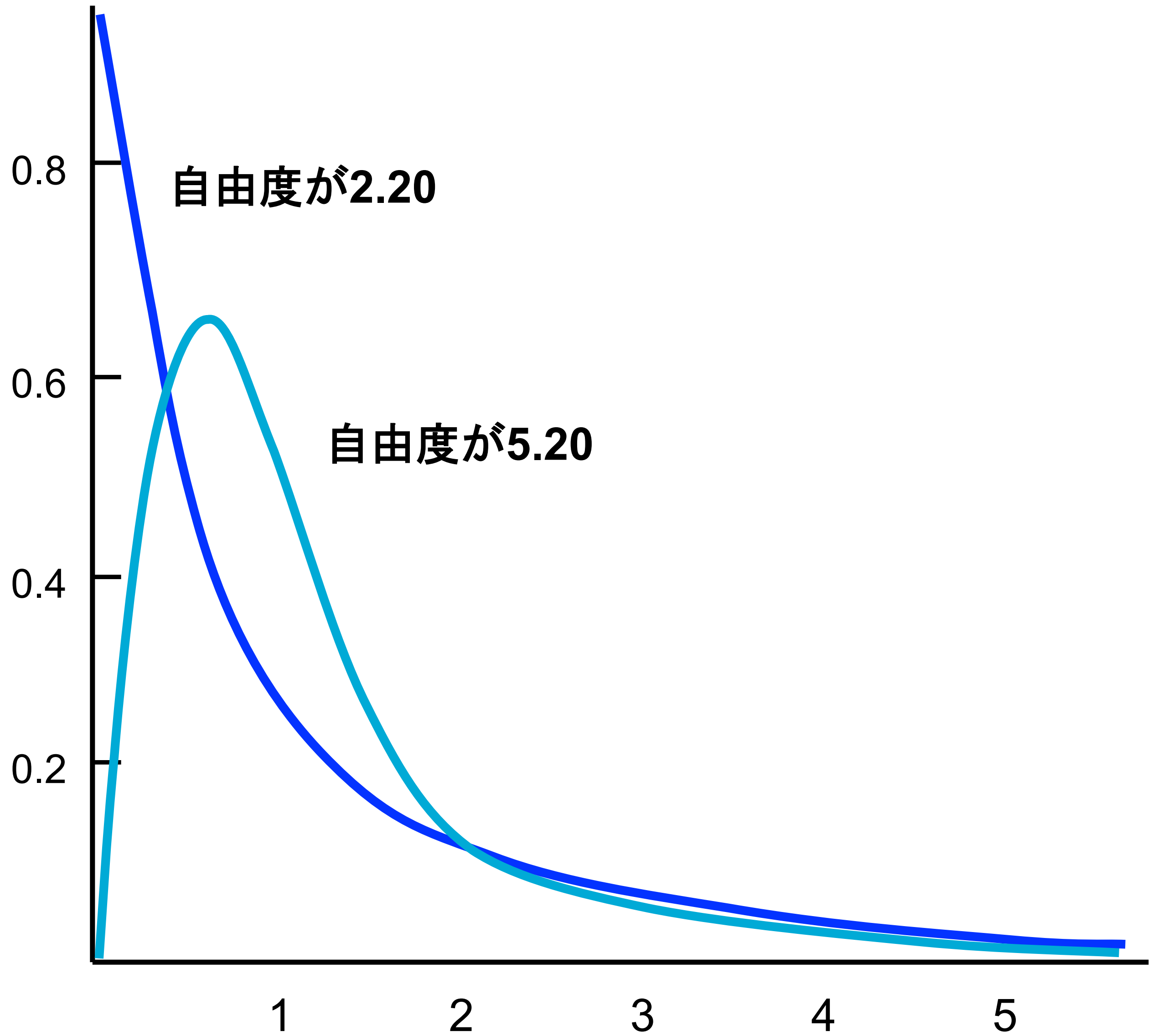
Vを自由度がmの χ^2 確率変数とし、Wを自由度がkの同じく χ^2 確率変数とする。ここで、VとWが独立に分布していればVとW間の比率

$$F = \frac{V/m}{W/k}$$

は「自由度がmとkの **F分布**」に従って分布している。比率Fを **F比** と呼ぶ。F比の平均値は $k/(k-2)$である。またkが3以上でないと平均値は存在しない。kが大きければ、平均値はほぼ1である。

資料）森棟公夫「統計学入門」新世社

図：F 分布



資料）森棟公夫「統計学入門」新世社

推定論

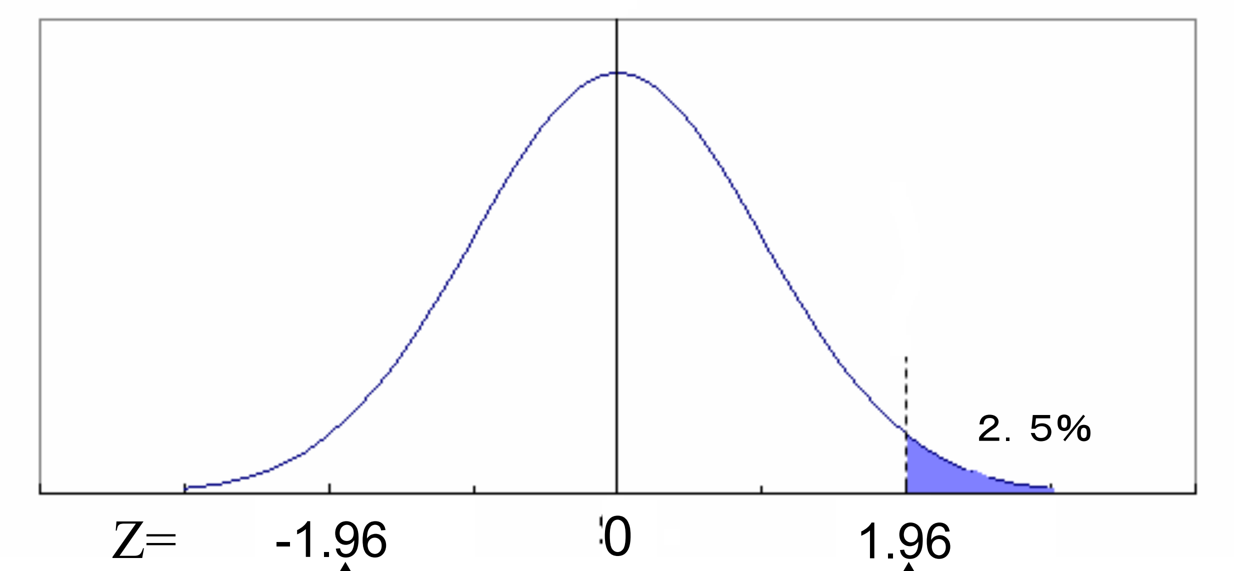
- 母数（パラメータ）
  - 推定量（estimator） vs 推定値（estimate）
  - 統計量 vs 統計値

- 1）点推定 vs 区間推定
  - 信頼区間、
  - 信頼係数、
  - 信頼限界

資料）森棟公夫「統計学入門」新世社

区間推定の例題

ある農産物の価格は月ごとに変動する。平均μは未知だが、標準偏差は12円の正規分布からの無作為標本と見なせるとする。最近6か月の価格が76円, 63円, 83円, 86円, 53円, 71円であった。この時、平均価格の推定値を95%の信頼係数で区間推定するとその信頼限界はどうなるか



$\overline{X} =$ 64.2 81.8

中心極限定理より、標本平均を標準化した変数の分布は、平均μ、分散$\sigma^2/n = 144/6 = 24$の標準正規分布に従うから

標本平均の標準化されたZは $Z = \frac{(\overline{X} - \mu)}{\sqrt{24}}$ となり、標準正規分布表より

95%の信頼区間に入る確率はP({－1.96 ≦ Z ≦ 1.96})＝0.95、従って、

$$\overline{X} - 1.96\sqrt{24} \le \mu \le \overline{X} + 1.96\sqrt{24}$$

これから、$\overline{X}$ に標本平均の実現値72を代入すると、μの区間推定は、 64.2 ≦μ≦81.8となる

- ２）推定量の特性
  - 不偏性(unbiasness) ・・・・・・小標本特性
  - 一致性（consistency）・・・・・大標本特性
  - 効率性（efficiency）
  - 充分性（ sufficiency ）

- ３）平均の推定
- ４）分散の推定
- ５）平均と分散の推定
- ６）成功確率の推定
- ７）標本規模の決定
- ８）推定法
  - 積率法（moment method）
  - 最尤法(maximum likelyhood method)

資料）森棟公夫「統計学入門」新世社

統計的仮説検定

- 1）検定の諸概念

  - 対立仮説(alternative) hypothesis

$$H_a : p < P_0$$

  - 帰無仮説(null hypothesis)

$$H_0 : p \geqq P_0$$

ここで、p は新品質管理の導入後の欠陥品の発生確率
p_0は従来の品質管理の下での欠陥品の発生確率

  - 検定統計量（test statistics)

ある工場で生産される電気部品の標本欠陥比率
古い品質管理法の下での欠陥比率が10%の時
25個中欠陥品が0個→帰無仮説を棄却
25個中欠陥品が5個→帰無仮説を棄却できず

資料）森棟公夫「統計学入門」新世社

棄却域、採択域と臨界値（critical value)

- P値（確率値）　到達された有意水準
  - 帰無仮説の下で検定統計量を超える確率
- 二種類の過誤
  - 第一種の過誤（有意水準＝α）

$$\alpha = 有意水準 = 第1種の過誤$$
$$= P(\{H_0 を棄却 | H_0 が正しい\})$$

  - 第二種の過誤（β）

$$\beta = 第2種の過誤 = P(\{H_0 を選択 | H_a が正しい\})$$
$$= 1 - P(\{H_0 を棄却 | H_a が正しい\})$$

表6.1

真の状態＼判断	H_0採択（H_a棄却）	H_0棄却（H_a採択）
H_0が正しい	正しい判断	第1種の過誤
Haが正しい	第2種の過誤	正しい判断

資料）森棟公夫「統計学入門」新世社

棄却域の決め方

- １）検定統計量を選ぶ
- ２）第１種の過誤（＝有意水準）の大きさを
  決める（通常は1%,5%,10%）。
- ３）有意水準にあった臨界値を決める。
  帰無仮説から見て、臨界値よりも遠く
  対立仮説に近い領域を棄却域とする。

↓

採択域＝棄却域の排反事象となる領域

資料）森棟公夫「統計学入門」新世社

工場で品質管理方法を改善:

検定統計量:　25個の標本のうちで検出される不良品の比率

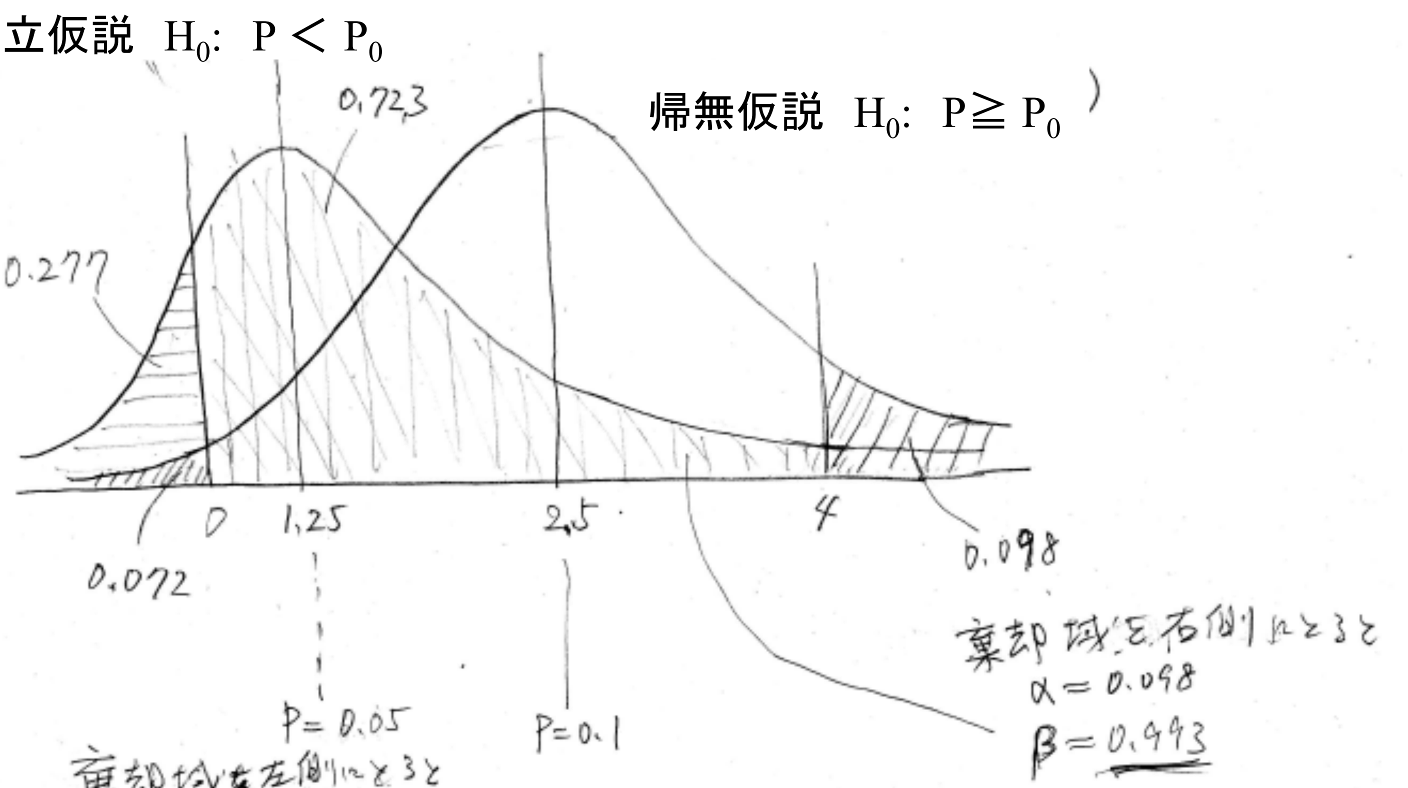
P:　改善後の不良品発生比率

P_0:　改善前の不良品発生比率

帰無仮説　H_0:　$P \geqq P_0$

対立仮説　Ha:　$P < P_0$

二項分布表により仮説検定



棄却域を左側にとると

$\alpha \doteq 0.072$

$\beta = (1-0.277) = 0.723$

資料)森棟公夫「統計学入門」新世社

２）平均値の検定（分散は既知）

- 新生児の体重の分布は正規分布で近似できて、その平均体重はほぼ3.1kg、標準偏差は0.2kg。

  25の無作為標本をとる。

  ＜標本平均の平均値の実現地が3.0であった＞

  ↓

- 対立仮説「平均体重は3.1kgより少ないのではないか」

  帰無仮説「平均体重は3.1kgである」

  標準偏差はどちらの仮説の下でも、0.2/5＝0.04

  検定の有意水準＝５％

資料）森棟公夫「統計学入門」新世社

$$H_0 : \mu = 3.1, \qquad H_a : \mu < 3.1$$

棄却域と臨界値

$$Z_0 = (\overline{X} - 3.1) / 0.04$$

有意水準＝第1種の過誤α＝0.05
第2種の過誤β=0.198

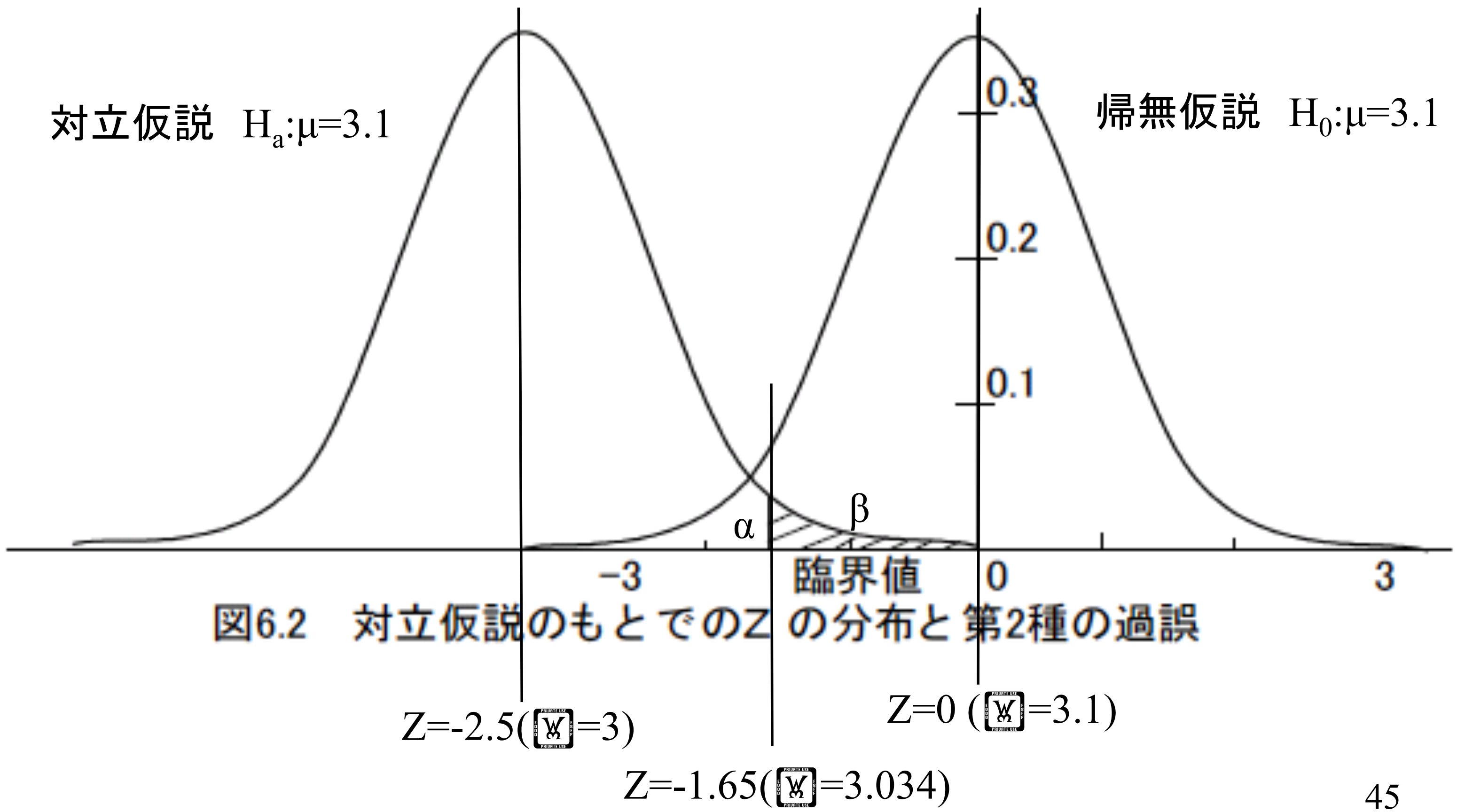


図6.2 対立仮説のもとでのZの分布と第2種の過誤

資料）森棟公夫「統計学入門」新世社

第2種の過誤

$$\beta = P(\{\overline{X} \geqq 3.034 | H_a\}) = P\left(\left\{\frac{\overline{X} - 3.0}{0.04} \geqq \frac{0.034}{0.04} | H_a\right\}\right)$$

$$= P(\{Z_a \geqq 0.85 | H_a\}) \fallingdotseq 0.198$$

- 異なる棄却域

$$P(\{-0.68 \leqq Z_0 \leqq -0.53 | H_0\}) \fallingdotseq 0.05$$

$$\beta = P(\{\overline{X} \leqq 3.073, 3.079 \leqq \overline{X} | H_a\})$$

$$= P\left(\left\{Z_a \leqq \frac{3.073 - 3.0}{0.04}, \frac{3.079 - 3.0}{0.04} \leqq Z_a | H_a\right\}\right)$$

$$= 1 - P(\{1.825 \leqq Z_a \leqq 1.975 | H_a\}) \fallingdotseq 0.99$$

異なる検定統計量

$$\beta = (\{\mu \geqq 2.867 | H_a\})$$

$$= P\left(\left\{\frac{\hat{\mu} - 3.0}{0.141} \geqq \frac{2.867 - 3.0}{0.141} | H_a\right\}\right)$$

$$= P(\{Z_a \leqq -0.943 | H_a\}) \fallingdotseq 0.826$$

資料）森棟公夫「統計学入門」新世社

有意水準と第2種の過誤

表6.2

有意水準α	0.01	0.05	0.10	0.25
β	0.40	0.20	0.11	0.04

例1、1986年の滋賀県の17歳児の平均身長、男子171.5cm、女子158.5cm、男女とも標本の大きさ450、1949年の17歳児の全国平均身長は、男子161.2cm、女子152.2cm。平均身長＝μ、身長の分布は正規分布で近似できる。標準偏差は1986年の全国値より、男子5.75cm、女子5.07cm。　　＜37年間に平均身長が伸びたかどうか＞

帰無仮説：男子μ＝161.2、女子μ＝152.2

対立仮説：男子μ＞161.2、女子μ＞152.2

男子の場合　$(171.5-161.2)/\sqrt{33.1/450}=38.0$

女子の場合　$(158.5-152.2)/\sqrt{25.7/450}=26.4$

男女とも帰無仮説は棄却、P値は殆どゼロ

資料）森棟公夫「統計学入門」新世社

母分布が未知の場合

母集団における分布が未知の場合、検定統計量の分布を正規分布で近似して検定。

$$H_0 : \mu = \mu_0, \qquad H_a : \mu \neq \mu_0$$

母分布は未知だが、母分散C_2は既知の場合

$$Z_0 = \frac{\overline{X} - \mu_0}{c / \sqrt{n}}$$

:中心極限定理により標準正規分布で近似

$$\alpha \fallingdotseq P(\{Z_0 \leqq -a, a \leqq Z_0 | H_0\})$$

$$= P\left(\left\{\overline{X} \leqq \mu_0 - \frac{ca}{\sqrt{n}}, \mu_0 + \frac{ca}{\sqrt{n}} \leqq \overline{X} \middle| H_0\right\}\right)$$

$$= P(\{\overline{X} \leqq b, c \leqq \overline{X} | H_0\})$$

$$\beta = 1 - P(\{b \leqq \overline{X} \leqq c | H_a\})$$

$$= 1 - P(\{\sqrt{n}(b - \mu_a)/c \leqq \sqrt{n}(\overline{X} - \mu_a)/c \leqq \sqrt{n}(c - \mu_a)/c | H_a\})$$

資料）森棟公夫「統計学入門」新世社

3）平均値の検定（分散は未知）

正規母集団の場合

母平均と母分散が共に未知の時、母分散を標本分散で置き換え、統計量を標準化

$$Z = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

各々の仮説の下で自由度n-1のt分布に従う。

$$P(\{a \leqq Z_0 | H_0\}) = \alpha$$

$$\mu_0 + (Sa / \sqrt{n}) \leqq \overline{X}$$

$$a + \frac{\mu_0 - \mu_a}{S / \sqrt{n}} \leqq Z_a$$

資料）森棟公夫「統計学入門」新世社

母分布が未知の場合

母集団における分布が未知であれば、検定は中心極限定理に依存して行う

（例）ある工場で生産される電球の平均寿命は1200時間。設備更新した新工程で生産される電球を100個無作為抽出し、電球の平均寿命が延びたかどうか検定する。帰無仮説はμ=1200、対立仮説はμ＞1200である。得られた標本から計算された標本平均は1230時間、標準偏差は標本観察値より120と推定された。母集団分布は分からないが、nが100と大きいので、標本平均の分布は正規分布により近似できる。標準正規分布表より、有意水準1%に対する臨界値はz_0=2.33、標本平均X'の臨界値に変換すると1200+(120/10)2.33=1228。有意水準1%で対立仮説が採択される。P値もほとんど0.01.

仮説検定と信頼区間

仮説検定における採択域の不等式と、信頼区間の不等式はまったく同じ形式を持つ。

有意水準がαの両側検定では、採択域に関するZ_0の不等式はξをt分布の上側100α%点とすると

$$\overline{X} - \xi(S/\sqrt{n}) \leqq \mu_0 \leqq \overline{X} + \xi(S/\sqrt{n})$$

母数μの100α%信頼区間は、同じξを使って

$$\overline{X} - \xi(S/\sqrt{n}) \leqq \mu \leqq \overline{X} + \xi(S/\sqrt{n})$$

「有意水準αに対して母数μ の100(1- α)%の信頼区間を作成し、この信頼区間にμ_0が入っていれば帰無仮説は採択され、逆にこの信頼区間にμ_0 が入っていなければ帰無仮説は棄却される」という記述の妥当性？

資料）森棟公夫「統計学入門」新世社

4）平均値の差の検定

学歴の差によって所得格差ができるかどうか

$$\overline{X} = \overline{X_1} - \overline{X_2}$$

高卒グループと大卒グループの間で平均所得に差があるかどうかの検定

$$\mu = E\left(\overline{X}\right) = E\left(\overline{X_1}\right) - E\left(\overline{X_2}\right) = \mu_1 - \mu_2$$

$$V\left(\overline{X}\right) = V\left(\overline{X_1}\right) + V\left(\overline{X_2}\right)$$
$$= (\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)$$

$$H_0 : \mu_1 = \mu_2$$

資料）森棟公夫「統計学入門」新世社

正規母集団の場合

1)母分散が既知の場合　$Z_0 = \overline{X} / \sqrt{V\left(\overline{X}\right)}$　:標準正規分布に従う

2)母分散は等しいが未知の場合（t検定）

$$Z_0 = \frac{\overline{X}}{S\sqrt{(1/n_1)+(1/n_2)}}$$

:自由度（n_1+n_2-2）のt分布に従う

$$S^2 = \frac{1}{n_1+n_2-2}\left\{\sum_{i=1}^{n_1}\left(X_{1i}-\overline{X_1}\right)^2 + \sum_{i=1}^{n_2}\left(X_{2i}-\overline{X_2}\right)^2\right\}$$

$$= \frac{1}{n_1+n_2-2}\left\{(n_1-1)S_1^2 + (n_2-1)S_2^2\right\}$$

3)母分散が異なり、かつ未知の場合

上記の1）の検定統計量を使うが、分母のV（X）が未知分散σ_1^2とσ_2^2を含むので、 未知分散を各々の標本から得られる標本分散S_1^2とS_2^2で置換える。検定統計量の帰無仮説の下での分布は標準正規分布で近似できる

資料）森棟公夫「統計学入門」新世社

分布が未知の場合

検定統計量は、正規母集団の3)のケースと同じ

帰無仮説のもとでの分布は中心極限定理により標準正規分布で近似

例題6.3 $\left(29.5-29.3\right)/\sqrt{\left(23.4/470\right)+\left(23.3/470\right)}=0.63$

例題6.4 $\left(171.5-168.5\right)/\sqrt{\left(30.9/430\right)+\left(33.1/430\right)}=7.8$

資料）森棟公夫「統計学入門」新世社

5）成功確率の検定 （二項確率に関する仮説検定の近似的方法）

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

が帰無仮説の下で標準正規分布に従う

$$P(\{-b \leqq Z_0 \leqq b\}) = 1 - \alpha$$

を満たす下側100（α/2）％点bを検定統計量Z0の臨界値にする

$$p_0 \pm \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}} b$$

サイコロを25回投げて、1の目の出る確率を検定、

$$H_0 : p_1 = 1/6 \qquad H_a : p_1 \leqq 1/6$$

標準正規分布表より下側100（α/2）％点の検定統計量Z_0の臨界値はb=1.65,
成功比率の臨界点は、p=1/6-1.65{1/6(1-1/6)/25}$^{0.5}$=0.044,、これに25をかけて1.1、つまり目の出る回数が0か1ならば帰無仮設は棄却される。

6）成功確率の差の検定 （二項確率に関する仮説検定の近似的方法）

$$H_0 : p_1 = p_2 \qquad H_a : p_1 \neq p_2 \qquad V = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

$$Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{V}}$$

は、近似的な標準正規分布になる。結局、平均の差の検定と同じ手続きになる

資料）森棟公夫「統計学入門」新世社

裸眼視力0.3以下の高校生の比率

高校	1年	2年	3年
男子	22.74	24.22	25.21%
女子	28.3	30.43	31.74%

	東大(工)			東工大(工)			早稲田大(工)		
年度	1978	1983	1988	1978	1983	1988	1978	1983	1988
製造業	68	73	43%	69	73	38%	66	79	58%
非製造業	32	27	57%	31	27	63%	34	20	42%
総人数	445	484	453	248	242	184	1040	934	875
Z_0値		−1.9	9.8		−0.90	7.2		−6.9	10
P値		3.10%	0		18%	0		0	0

	東大(工)		東工大(工)		早稲田大(工)	
年度	1983	1988	1983	1988	1983	1988
金融	4	59	5	32	3	59
保険	5	22	0	7	10	31
証券	0	22	0	10	5	35
通信報道出版	3	37	6	21	15	59

資料)森棟公夫「統計学入門」新世社

7）独立性の検定

母集団からの2組の正規確率変数X,Yの独立性を検定

$$\{(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)\}$$

$$H_0 : \rho = 0 \qquad H_a : \rho \neq 0$$

母集団相関係数ρの推定に標本相関係数rを使用して、検定統計量：

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

は、帰無仮説の下で、自由度がn-2のt分布に従う。nが十分大きければTの分布は標準正規分布で近似される

表6.5　在学率と出生率

	日本	イスラエル	イラク	インドネシア	韓国	クウェート	シンガポール
在学率	97	83	32	37	92	79	73
出生率	12	24	44	32	20	35	17

	タイ	中国	トルコ	パキスタン	フィリピン	香港	マレーシア
在学率	29	35	35	10	53	68	59
出生率	28	19	30	43	33	14	31

標本相関係数　r=-0.655　から、　検定統計量Tは−3.0となる。t分布表より、自由度n=12、有意水準5%の臨界値は-1.782なので、在学率と4出生率の間に相関はないという帰無仮説は棄却される。

資料）森棟公夫「統計学入門」新世社

8）尤度比検定法

尤度：母集団における確率変数の確率密度f(x,θ)からの、無作為標本の密度関数

$$L = \prod_{i=1}^{n} f(x_i, \theta)$$

帰無仮説の下での尤度の値 $$L_0 = \prod_{i=1}^{n} f(x_i, \theta_0)$$

対立仮説の下での尤度の値 $$L_a = \prod_{i=1}^{n} f(x_i, \hat{\theta})$$

尤度比 $$\lambda = L_0 / L_a$$

L_0はL_aよりも制約が強いので、Laよりも小さい。
従って、$0 < \lambda < 1$

臨界値の決定 $$P(\{\lambda \leqq c\}) = \alpha$$

定理6.1　確率変数の分布が未知母数を含み、かつその分布関数が幾つかの条件を満たすとする。Xの無作為標本をX_1、・・・・、X_n、帰無仮説はθの一部θ_1・・・θ_kについて、　H_0:　$\theta_1=c_1$、・・・・、$\theta_k=c_{nk}$とする。対立仮説は帰無仮説を否定するものとする。標本の大きさnが十分大きければ、帰無仮説の下で$-2\log(\lambda)$の分布は自由度がkのχ^2分布により、近似できる。棄却域は$\chi 2$分布の右裾に取れば良い。

資料）森棟公夫「統計学入門」新世社

検定の際の比較検討対象となる母集団の数が多数の場合

1）分散分析

$$x_{ij} - \bar{x} = (x_{ij} - \bar{x_i}) + (\bar{x_i} - \bar{x})$$

観測値の総平均からの偏差　観測値のグループ平均からの偏差　グループ平均の総平均からの偏差

$$TSS = \sum_{i=1}^{2}\sum_{j}(x_{ij} - \bar{x})^2$$

TSS = 級間変動＋級内変動

級内 変動＝$\sum_{i=1}^{2}\sum_j (x_{ij} - \bar{x_i})^2$

級間変動＝$\sum_{i=1}^{2} \boldsymbol{n_i}(\bar{\boldsymbol{x_i}} - \bar{\boldsymbol{x}})_2$

$$F = \frac{級間変動/a}{級内変動/b}$$

２）分割表 （ χ^{2}検定）

χ^{2}検定統計量

$$\chi^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i}$$

適合度検定

独立性の検定　　E = n x 行の周辺確率 x 列の周辺確率

同等性の検定

３）ノンパラメトリック検定

順位和検定

$$T_1 + T_2 = (n_1 + n_2)(n_1 + n_2 + 1)/2$$

$$E(T_1) = n_1(n_1 + n_2 + 1)/2 \qquad \sigma^2 = V(T_1) = n_1 n_2 (n_1 + n_2 + 1)/12$$

$$Z_1 = \frac{T_1 - E(T_1)}{\sigma}$$

符号検定

$$Z = \frac{X - (n/2)}{\sqrt{n}/2}$$

符号付順位和検定

$$Z = \frac{T - n(n+1)/4}{n(n+1)(2n+1)/24}$$

nが大の下で標準正規分布

クラスカル・ワリス検定

スピアマン順位相関係数

$$r_s = \frac{\Sigma(R_i - m)(Q_i - m)}{\sqrt{\Sigma(R_i - m)^2 \Sigma(Q_i - m)^2}} \qquad r_s = 1 - \frac{6\Sigma(R_i - Q_i)^2}{n(n^2 - 1)} \qquad t = \frac{r_s\sqrt{n-2}}{\sqrt{1 - r_s^2}}$$

資料）森棟公夫「統計学入門」新世社

付表２． 二項確率関数（その１）

成功回数が x 回以下の確率

$n=5$

x	0.01	0.05	0.10	0.20	0.30	0.40	0.50
0	.951	.774	.590	.328	.168	.078	.031
1	.999	.977	.919	.737	.528	.337	.188
2	1.000	.999	.991	.942	.837	.683	.500
3	1.000	1.000	1.000	.993	.969	.913	.813
4	1.000	1.000	1.000	1.000	.998	.990	.969
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$n=10$

x	0.01	0.05	0.10	0.20	0.30	0.40	0.50
0	.904	.599	.349	.107	.028	.006	.001
1	.996	.914	.736	.376	.149	.046	.011
2	1.000	.989	.930	.678	.383	.167	.055
3	1.000	.999	.987	.879	.650	.382	.172
4	1.000	1.000	.998	.967	.850	.633	.377
5	1.000	1.000	1.000	.994	.953	.834	.623
6	1.000	1.000	1.000	.999	.989	.945	.828
7	1.000	1.000	1.000	1.000	.998	.988	.945
8	1.000	1.000	1.000	1.000	1.000	.998	.989
9	1.000	1.000	1.000	1.000	1.000	1.000	.999
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$n=15$

x	0.01	0.05	0.10	0.20	0.30	0.40	0.50
0	.860	.463	.206	.035	.005	.000	.000
1	.990	.829	.549	.167	.035	.005	.000
2	1.000	.964	.816	.398	.127	.027	.004
3	1.000	.995	.944	.648	.297	.091	.018
4	1.000	.999	.987	.836	.515	.217	.059
5	1.000	1.000	.998	.939	.722	.403	.151
6	1.000	1.000	1.000	.982	.869	.610	.304
7	1.000	1.000	1.000	.996	.950	.787	.500
8	1.000	1.000	1.000	.999	.985	.905	.696
9	1.000	1.000	1.000	1.000	.996	.966	.849
10	1.000	1.000	1.000	1.000	.999	.991	.941
11	1.000	1.000	1.000	1.000	1.000	.998	.982
12	1.000	1.000	1.000	1.000	1.000	1.000	.996
13	1.000	1.000	1.000	1.000	1.000	1.000	1.000
15	1.000	1.000	1.000	1.000	1.000	1.000	1.000

資料）森棟公夫「統計学入門」新世社

付表２． 二項確率関数（その２）

成功回数が x 回以下の確率

n=20

x	0.01	0.05	0.10	0.20	0.30	0.40	0.50
0	.818	.359	.122	.012	.001	.000	.000
1	.983	.736	.392	.069	.008	.001	.000
2	.999	.925	.677	.206	.035	.004	.000
3	1.000	.984	.867	.411	.107	.016	.001
4	1.000	.997	.957	.630	.238	.051	.006
5	1.000	1.000	.989	.804	.416	.126	.021
6	1.000	1.000	.998	.913	.608	.250	.058
7	1.000	1.000	1.000	.968	.772	.416	.132
8	1.000	1.000	1.000	.990	.887	.596	.252
9	1.000	1.000	1.000	.997	.952	.755	.412
10	1.000	1.000	1.000	.999	.983	.872	.588
11	1.000	1.000	1.000	1.000	.995	.943	.748
12	1.000	1.000	1.000	1.000	.999	.979	.868
13	1.000	1.000	1.000	1.000	1.000	.994	.942
14	1.000	1.000	1.000	1.000	1.000	.998	.979
15	1.000	1.000	1.000	1.000	1.000	1.000	.994
16	1.000	1.000	1.000	1.000	1.000	1.000	.999
17	1.000	1.000	1.000	1.000	1.000	1.000	1.000
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000

n=25

x	0.01	0.05	0.10	0.20	0.30	0.40	0.50
0	.778	.277	.072	.004	.000	.000	.000
1	.974	.642	.271	.027	.002	.000	.000
2	.998	.873	.537	.098	.009	.000	.000
3	1.000	.966	.764	.234	.033	.002	.000
4	1.000	.993	.902	.421	.090	.009	.000
5	1.000	.999	.967	.617	.193	.029	.002
6	1.000	1.000	.991	.780	.341	.074	.007
7	1.000	1.000	.998	.891	.512	.154	.022
8	1.000	1.000	1.000	.953	.677	.274	.054
9	1.000	1.000	1.000	.983	.811	.425	.115
10	1.000	1.000	1.000	.994	.902	.586	.212
11	1.000	1.000	1.000	.998	.956	.732	.345
12	1.000	1.000	1.000	1.000	.983	.846	.500
13	1.000	1.000	1.000	1.000	.994	.922	.655
14	1.000	1.000	1.000	1.000	.998	.966	.788
15	1.000	1.000	1.000	1.000	1.000	.987	.885
16	1.000	1.000	1.000	1.000	1.000	.996	.946
17	1.000	1.000	1.000	1.000	1.000	.999	.978
18	1.000	1.000	1.000	1.000	1.000	1.000	.993
19	1.000	1.000	1.000	1.000	1.000	1.000	.998
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
25	1.000	1.000	1.000	1.000	1.000	1.000	1.000

資料）森棟公夫「統計学入門」新世社

付表3． ポアソン分布

事象が x 回起きる確率

$x \backslash \lambda$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	.905	.819	.741	.670	.607	.549	.497	.449	.407
1	.090	.164	.222	.268	.303	.329	.348	.359	.366
2	.005	.016	.033	.054	.076	.099	.122	.144	.165
3	.000	.001	.003	.007	.013	.020	.028	.038	.049
4	.000	.000	.000	.001	.002	.003	.005	.008	.011
5	.000	.000	.000	.000	.000	.000	.001	.001	.002
6	.000	.000	.000	.000	.000	.000	.000	.000	.000

$x \backslash \lambda$	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0	.368	.223	.135	.082	.050	.030	.018	.011	.007
1	.368	.335	.271	.205	.150	.106	.073	.050	.034
2	.184	.251	.271	.257	.224	.185	.147	.113	.084
3	.061	.126	.180	.214	.224	.216	.195	.169	.140
4	.015	.047	.090	.134	.168	.189	.195	.190	.175
5	.003	.014	.036	.067	.101	.132	.156	.171	.175
6	.001	.004	.012	.028	.050	.077	.104	.128	.146
7	.000	.001	.003	.010	.022	.039	.060	.082	.104
8	.000	.000	.001	.003	.008	.017	.030	.046	.065
9	.000	.000	.000	.001	.003	.007	.013	.023	.036
10	.000	.000	.000	.000	.001	.002	.005	.010	.018
11	.000	.000	.000	.000	.000	.001	.002	.004	.008
12	.000	.000	.000	.000	.000	.000	.001	.001	.003
13	.000	.000	.000	.000	.000	.000	.000	.001	.001
14	.000	.000	.000	.000	.000	.000	.000	.000	.000
15	.000	.000	.000	.000	.000	.000	.000	.000	.000

$x \backslash \lambda$	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0
0	.002	.001	.000	.000	.000	.000	.000	.000	.000
1	.015	.006	.003	.001	.000	.000	.000	.000	.000
2	.045	.022	.011	.005	.002	.001	.000	.000	.000
3	.089	.052	.029	.015	.008	.004	.002	.001	.000
4	.134	.091	.057	.034	.019	.010	.005	.003	.001
5	.161	.128	.092	.061	.038	.022	.013	.007	.004
6	.161	.149	.122	.091	.063	.041	.026	.015	.009
7	.138	.149	.140	.117	.090	.065	.044	.028	.017
8	.103	.130	.140	.132	.113	.089	.066	.046	.030
9	.068	.101	.124	.132	.125	.109	.087	.066	.047
10	.041	.071	.099	.119	.125	.119	.105	.086	.066
11	.023	.045	.072	.097	.114	.119	.114	.102	.084
12	.011	.026	.048	.073	.095	.109	.114	.110	.098
13	.005	.014	.030	.050	.073	.093	.105	.110	.106
14	.002	.007	.017	.032	.052	.073	.091	.102	.106
15	.001	.003	.009	.019	.035	.053	.072	.089	.099
16	.000	.001	.005	.011	.022	.037	.055	.072	.087
17	.000	.001	.002	.006	.013	.024	.038	.055	.071
18	.000	.000	.001	.003	.007	.015	.026	.040	.055
19	.000	.000	.000	.001	.004	.008	.016	.027	.041
20	.000	.000	.000	.001	.002	.005	.010	.018	.029
21	.000	.000	.000	.000	.001	.002	.006	.011	.019
22	.000	.000	.000	.000	.000	.001	.003	.007	.012
23	.000	.000	.000	.000	.000	.001	.002	.004	.007
24	.000	.000	.000	.000	.000	.000	.001	.002	.004
25	.000	.000	.000	.000	.000	.000	.000	.001	.002
26	.000	.000	.000	.000	.000	.000	.000	.001	.001
27	.000	.000	.000	.000	.000	.000	.000	.000	.001

資料）森棟公夫「統計学入門」新世社

付表４．　標準正規分布表

負の無限大から x までの確率を与える。座標値の小数第１位までは縦軸に，小数第２位は横軸に示されている。

小数第２位

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7612	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9773	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9983	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

資料）森棟公夫「統計学入門」新世社

付表５．　カイ２乗分布表

自由度が m，確率が p の座標を与える。ただし p は分布の右裾の確率である。

$m \backslash p$	0.995	0.990	0.975	0.950	0.900	0.750
1	0.000	0.000	0.000	0.000	0.016	0.102
2	0.010	0.020	0.051	0.103	0.211	0.575
3	0.072	0.115	0.216	0.352	0.584	1.213
4	0.207	0.297	0.484	0.711	1.064	1.923
5	0.412	0.554	0.831	1.145	1.610	2.675
6	0.676	0.872	1.237	1.635	2.204	3.455
7	0.989	1.239	1.690	2.167	2.833	4.255
8	1.344	1.646	2.180	2.733	3.490	5.071
9	1.735	2.088	2.700	3.325	4.168	5.899
10	2.156	2.558	3.247	3.940	4.865	6.737
11	2.603	3.053	3.816	4.575	5.578	7.584
12	3.074	3.571	4.404	5.226	6.304	8.438
13	3.565	4.107	5.009	5.892	7.042	9.299
14	4.075	4.660	5.629	6.571	7.790	10.17
15	4.601	5.229	6.262	7.261	8.547	11.04
16	5.142	5.812	6.908	7.962	9.312	11.91
17	5.697	6.408	7.564	8.672	10.09	12.79
18	6.265	7.015	8.231	9.390	10.86	13.68
19	6.844	7.633	8.907	10.12	11.66	14.56
20	7.434	8.260	9.591	10.85	12.44	15.45
22	8.643	9.542	10.98	12.34	14.04	17.24
24	9.886	10.86	12.40	13.85	15.66	19.04
26	11.16	12.20	13.84	15.38	17.29	20.84
28	12.46	13.57	15.31	16.93	18.94	22.66
30	13.79	14.95	16.79	18.49	20.60	24.48

$m \backslash p$	0.500	0.250	0.100	0.050	0.025	0.010	0.005
1	0.455	1.323	2.706	3.841	5.024	6.635	7.879
2	1.386	2.773	4.605	5.991	7.378	9.210	10.60
3	2.366	4.108	6.251	7.815	9.348	11.34	12.84
4	3.357	5.385	7.779	9.488	11.14	13.28	14.86
5	4.351	6.626	9.236	11.07	12.83	15.09	16.75
6	5.348	7.841	10.64	12.59	14.45	16.81	18.55
7	6.346	9.037	12.02	14.07	16.01	18.48	20.28
8	7.344	10.22	13.36	15.51	17.53	20.09	21.95
9	8.343	11.39	14.68	16.92	19.02	21.67	23.59
10	9.342	12.55	15.99	18.31	20.48	23.21	25.19
11	10.34	13.70	17.28	19.68	21.92	24.72	26.76
12	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	19.34	23.83	28.41	31.41	34.17	37.57	40.00
22	21.34	26.04	30.81	33.92	36.78	40.29	42.80
24	23.34	28.24	33.20	36.42	39.36	42.98	45.56
26	25.34	30.43	35.56	38.89	41.92	45.64	48.29
28	27.34	32.62	37.92	41.34	44.46	48.28	50.99
30	29.34	34.80	40.26	43.77	46.98	50.89	53.67

資料）森棟公夫「統計学入門」新世社

付表 6． t 分布表

自由度が df，確率が p の座標を与える。ただし p は分布の右裾の確率である。

df \ p	.35	.30	.25	.20	.15	.10	.05	.025	.010
1	.510	.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821
2	.445	.617	.816	1.061	1.386	1.886	2.920	4.303	6.965
3	.424	.584	.765	.978	1.250	1.638	2.353	3.182	4.541
4	.414	.569	.741	.941	1.190	1.533	2.132	2.776	3.747
5	.408	.559	.727	.920	1.156	1.476	2.015	2.571	3.365
6	.404	.553	.718	.906	1.134	1.440	1.943	2.447	3.143
7	.402	.549	.711	.896	1.119	1.415	1.895	2.365	2.998
8	.399	.546	.706	.889	1.108	1.397	1.860	2.306	2.896
9	.398	.543	.703	.883	1.100	1.383	1.833	2.262	2.821
10	.397	.542	.700	.879	1.093	1.372	1.812	2.228	2.764
12	.395	.539	.695	.873	1.083	1.356	1.782	2.179	2.681
14	.393	.537	.692	.868	1.076	1.345	1.761	2.145	2.624
16	.392	.535	.690	.865	1.071	1.337	1.746	2.120	2.583
18	.392	.534	.688	.862	1.067	1.330	1.734	2.101	2.552
20	.391	.533	.687	.860	1.064	1.325	1.725	2.086	2.528
22	.390	.532	.686	.858	1.061	1.321	1.717	2.074	2.508
24	.390	.531	.685	.857	1.059	1.318	1.711	2.064	2.492
26	.390	.531	.684	.856	1.058	1.315	1.706	2.056	2.479
28	.389	.530	.683	.855	1.056	1.313	1.701	2.048	2.467
30	.389	.530	.683	.854	1.055	1.310	1.697	2.042	2.457
∞	.385	.524	.674	.842	1.036	1.282	1.645	1.960	2.326

自由度が無限大の t 分布は標準正規分布に一致する。

資料）森棟公夫「統計学入門」新世社

付表７． *F* 分布表

分子の自由度が *m*，分母の自由度が*k* の *F* 分布表の上側５％点（太字）と１％点（細字）を与える。

k\m	1	2	3	4	5	6	8	10	12	14	16	20	50
5	**6.61**	**5.79**	**5.41**	**5.19**	**5.05**	**4.95**	**4.82**	**4.74**	**4.68**	**4.64**	**4.60**	**4.56**	**4.44**
	16.26	13.27	12.06	11.39	10.97	10.67	10.29	10.05	9.89	9.77	9.68	9.55	9.24
6	**5.99**	**5.14**	**4.76**	**4.53**	**4.39**	**4.28**	**4.15**	**4.06**	**4.00**	**3.96**	**3.92**	**3.87**	**3.75**
	13.75	10.93	9.78	9.15	8.75	8.47	8.10	7.87	7.72	7.60	7.52	7.40	7.09
7	**5.59**	**4.74**	**4.35**	**4.12**	**3.97**	**3.87**	**3.73**	**3.64**	**3.58**	**3.52**	**3.49**	**3.45**	**3.32**
	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.62	6.47	6.35	6.27	6.16	5.85
8	**5.32**	**4.46**	**4.07**	**3.84**	**3.69**	**3.58**	**3.44**	**3.35**	**3.28**	**3.23**	**3.20**	**3.15**	**3.03**
	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.81	5.67	5.56	5.48	5.36	5.06
9	**5.12**	**4.26**	**3.86**	**3.63**	**3.48**	**3.37**	**3.23**	**3.14**	**3.07**	**3.02**	**2.98**	**2.94**	**2.80**
	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.26	5.11	5.00	4.92	4.81	4.51
10	**4.96**	**4.10**	**3.71**	**3.48**	**3.33**	**3.22**	**3.07**	**2.98**	**2.91**	**2.86**	**2.82**	**2.77**	**2.64**
	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.85	4.71	4.60	4.52	4.41	4.12
12	**4.75**	**3.89**	**3.49**	**3.26**	**3.11**	**3.00**	**2.85**	**2.75**	**2.69**	**2.64**	**2.60**	**2.54**	**2.40**
	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.30	4.16	4.05	3.98	3.86	3.56
14	**4.60**	**3.74**	**3.34**	**3.11**	**2.96**	**2.85**	**2.70**	**2.60**	**2.53**	**2.48**	**2.44**	**2.39**	**2.24**
	8.86	6.52	5.56	5.04	4.70	4.46	4.14	3.94	3.80	3.70	3.62	3.51	3.21
16	**4.49**	**3.63**	**3.24**	**3.01**	**2.85**	**2.74**	**2.59**	**2.49**	**2.43**	**2.37**	**2.33**	**2.28**	**2.13**
	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.69	3.55	3.45	3.37	3.26	2.96
18	**4.41**	**3.56**	**3.16**	**2.93**	**2.77**	**2.66**	**2.51**	**2.41**	**2.34**	**2.29**	**2.25**	**2.19**	**2.04**
	8.29	6.01	5.09	4.58	4.25	4.02	3.71	3.51	3.37	3.27	3.19	3.08	2.78
20	**4.35**	**3.49**	**3.10**	**2.87**	**2.71**	**2.60**	**2.45**	**2.35**	**2.28**	**2.23**	**2.18**	**2.12**	**1.96**
	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.37	3.23	3.13	3.05	2.94	2.63
25	**4.24**	**3.39**	**2.99**	**2.76**	**2.60**	**2.49**	**2.34**	**2.24**	**2.17**	**2.11**	**2.06**	**2.01**	**1.84**
	7.77	5.57	4.68	4.18	3.86	3.63	3.32	3.13	2.99	2.89	2.81	2.70	2.40
30	**4.17**	**3.32**	**2.92**	**2.69**	**2.53**	**2.42**	**2.27**	**2.17**	**2.09**	**2.04**	**1.99**	**1.93**	**1.76**
	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.98	2.84	2.74	2.66	2.55	2.24
50	**4.03**	**3.18**	**2.79**	**2.56**	**2.40**	**2.29**	**2.13**	**2.03**	**1.95**	**1.90**	**1.85**	**1.78**	**1.60**
	7.17	5.06	4.20	3.72	3.41	3.19	2.89	2.70	2.56	2.46	2.39	2.27	1.94
100	**3.94**	**3.09**	**2.70**	**2.46**	**2.30**	**2.19**	**2.03**	**1.92**	**1.85**	**1.79**	**1.75**	**1.68**	**1.48**
	6.90	4.82	3.98	3.51	3.20	2.99	2.69	2.51	2.36	2.26	2.19	2.06	1.73

資料）森棟公夫「統計学入門」新世社

相関と回帰

相関係数

- 相関係数（correlation coefficient）
  - 2 つのデータ列の間の相関（類似性の度合い）を示す指標。$-1 \leq R \leq 1$ の間の実数値をとる。

$$
\text{相関係数} = \frac{\text{共分散}}{\sqrt{(x\text{の分散} \times y\text{の分散})}}
$$

$$
R = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot \sum_{i=1}^{n} (y_i - \overline{y})^2}}
$$

相関係数は、データ間の因果関係を説明するものではなく、あくまでもデータ間の線形関係を計測しているに過ぎない。

回帰分析

元データ

	身長 (cm)	体重 (kg)
1	167	66
2	158	59
3	177	74
4	165	65
5	162	67
6	174	66
7	170	69
8	161	65
9	165	65
10	179	74
11	176	73
12	170	68
13	157	72
14	170	69
15	187	77
16	180	74
17	184	75
18	166	67
19	160	61
20	165	66

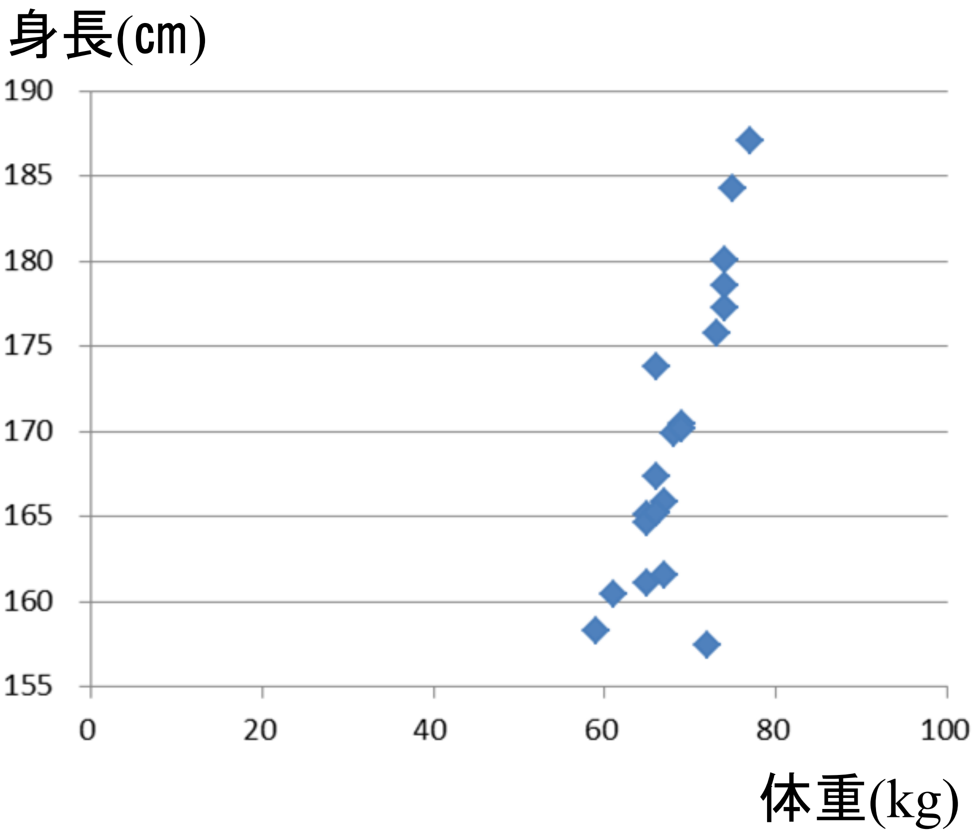
あるクラスの学生20名の身長は各自の体重によりどの程度説明されるか検討して見る。つまり学生の身長を体重に規定される項とそうではない定数項とで説明する回帰分析を適用する。

表は、クラスの学生の身長と体重のデータを示したものである。つまり

(身長) = $\alpha + \beta \times$(体重) + 誤差項

という関係に回帰式を適用する

線形推定



この場合の重相関係数はR=0.811と計算される。このことから、各学生の体重が分かれば、ある程度身長も推測できることが類推される。
この図を見る限り

(身長)=α + β ×(体重)

という計算式で求められそうである。

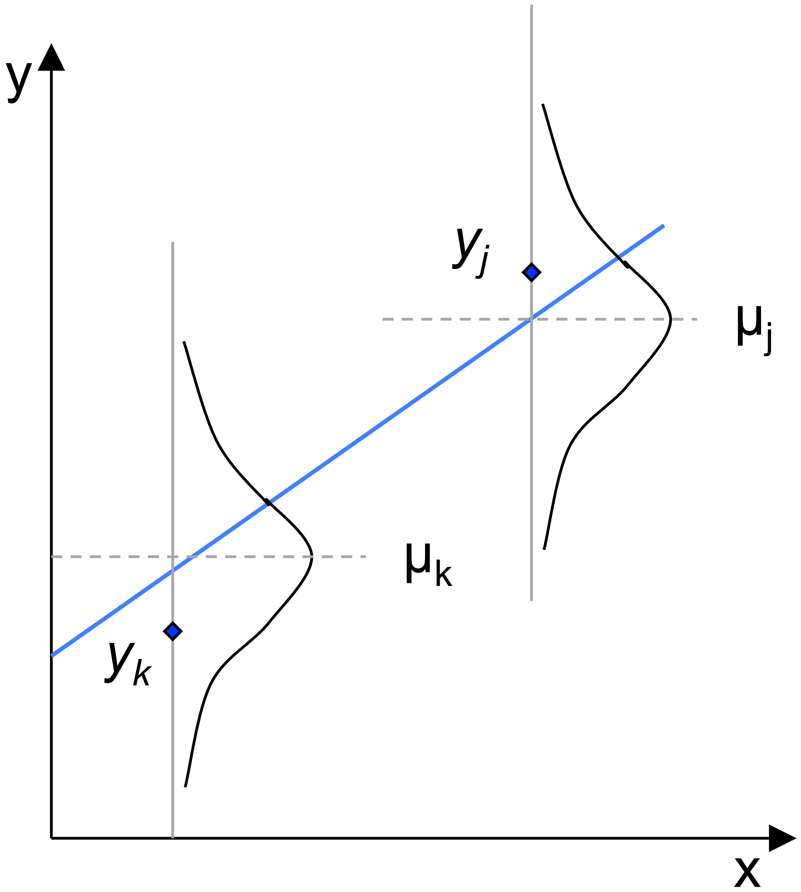
回帰係数

- まず、確率変数 Y_i (平均 μ_i 、分散 σ^2)が身長を表わすとする。
  Y_i の期待値は $E(Y_i) = \mu_i$ となり、この平均値 μ_i が x_i によって決まるという式を考える。

$$E(Y_i) = \mu_i = \alpha + \beta x_i$$

  言い換えれば、Y_i が平均値 μ_i の周りで分布すると仮定することになる。
- つまり、各学生の体重が原因で、身長は結果であると把握する訳である。
- この場合、体重を「**説明変数**」、身長を「**被説明変数**」とよぶ。また、α，β を「**回帰係数**」と呼ぶ。

誤差項の分布

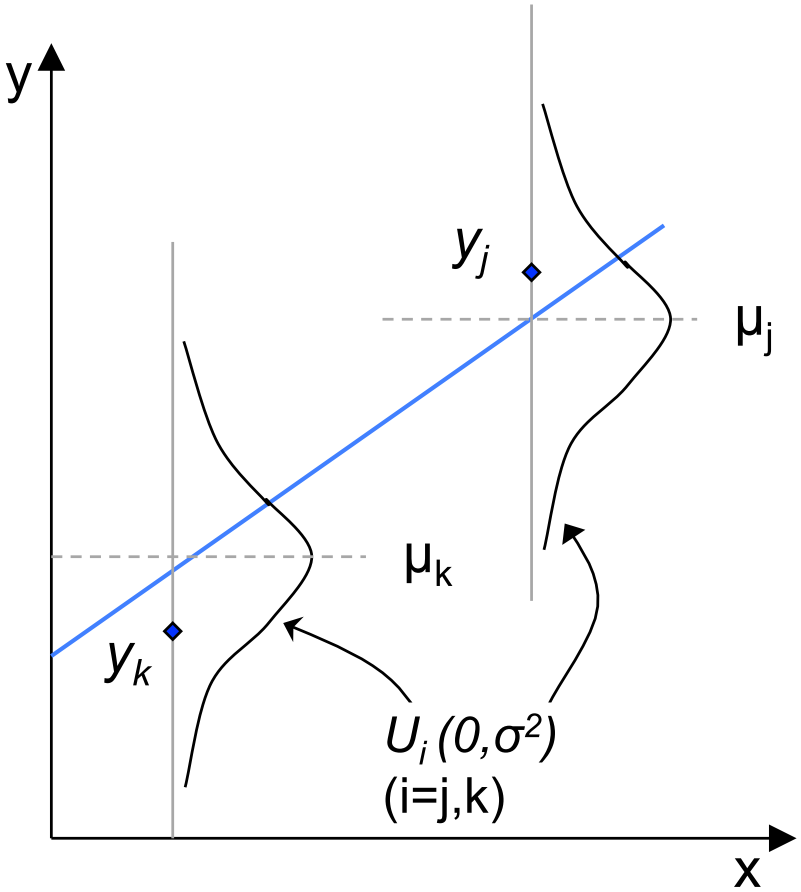


- この式はデータ毎の関係として次のようにも書ける。

$$Y_i = \alpha + \beta x_i + Ui$$
$$(i = 1, ..., n)$$

- U_iは「**<u>誤差項</u>**」であり、その推定値が残差項である
- 誤差項は Y_i の平均値μ_iの周りの散らばりを示す。 U_i は Y_i の平均値からの散らばりなので U_i の期待値 $E(U_i) = 0$ である。

誤差項と未知パラメータ



この場合、分散は、

$$
\begin{aligned}
V(Y_i) &= \sum_{i=1}^{n}(Y_i-\mu_i)^2/(n-1)\\
&= \sum_{i=1}^{n}(Y_i - \alpha - \beta x_i)^2/(n-1)\\
&= \sum_{i=1}^{n}(U_i)^2/(n-1)\\
&= V(U_i)=\sigma^2
\end{aligned}
$$

となる。

ここで、推定すべき未知の値はα、β、σ^2である。（n−1)で割るのは、n個のデータが一つの式で縛られているので、その不偏分散を求めるためである。

残差二乗和

- 残差は正負両方の値をとり得るので、残差の大きさの総和を正確に把握するために二乗して正の値に直して、すべての標本について集計する。
こうして導かれるのが「**残差二乗和**」（RSS：Residual sum of squares）である。
残差二乗和を最小にすることにより導かれる推定方法を最小二乗法という。

最小二乗法

- 残差二乗和をψで表すと、式の上では

$$\psi = \sum_{i=1}^{n}(y_i - a - bx_i)^2$$

- となる。a, bの両方に関してこの残差二乗和ψを最小化にする値が求めるパラメーターα, βになる。ψが最小になる1階の条件は、a, bについて偏微分してゼロと置くことにより求められるので、

$$\sum_{i=1}^{n}(y_i - a - bx_i) = 0 \quad \cdots (1)$$

$$\sum_{i=1}^{n}((y_i - a - bx_i)x_i) = 0 \quad \cdots (2)$$

となり、この連立方程式を解くことによりa, bの値が推定される。

正規方程式

（１）、（２）式の和記号部分を計算して、

$$(1) = \sum_{i=1}^{n} y_i - na - b\sum_{i=1}^{n} x_i = 0$$

$$na = \sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i \quad \cdots(3)$$

$$a = \sum_{i=1}^{n}(y_i)/n - b\sum_{i=1}^{n}(x_i)/n$$

$$a = \bar{y} - b\bar{x} \quad \cdots(4)$$

ここで

$$\sum_{i=1}^{n}(x_i)/n = \bar{x}$$

$$\sum_{i=1}^{n}(y_i)/n = \bar{y}$$

$$(2) = \sum_{i=1}^{n}(x_i y_i) - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n}(x_i)^2 = 0$$

$$a\Big(\sum_{i=1}^{n}(x_i)\Big) + b\Big(\sum_{i=1}^{n}(x_i^2)\Big) = \sum_{i=1}^{n}(x_i y_i) \quad \cdots(5)$$

（３）、（５）式をあわせて、正規方程式と呼ぶ

定数項と回帰係数の解法

$$b\sum_{i=1}^{n}(x_i^2) + \bar{y}\sum_{i=1}^{n}x_i - b\bar{x}\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}(x_iy_i)$$

$$\Big(\sum_{i=1}^{n}(x_i^2) - n\bar{x}\sum_{i=1}^{n}(x_i)\Big)b = \sum_{i=1}^{n}(x_iy_i) - \bar{y}\sum_{i=1}^{n}x_i$$

$$b = \frac{\Big(\sum_{i=1}^{n}(x_iy_i)\Big) - n\bar{x}\bar{y}}{\sum_{i=1}^{n}(x_i^2) - n\bar{x}^2} \quad \cdots(6)$$

＊注意＊

$\sum_{i=1}^{n}(x_i)/n = \bar{x}$

$\sum_{i=1}^{n}(y_i)/n = \bar{y}$

（4）、（6）式より、$\bar{x}$ $\bar{y}$ $\sum_{i=1}^{n}(x_iy_i)$ $\sum_{i=1}^{n}(x_i^2)$を計算するとα、βを計算することができる。

逐次代入による解法

- 実際に計算すると

$$\bar{x} = 68.60 \quad \sum_{i=1}^{n}(x_i{}^2) = 94564 \quad \sum_{i=1}^{n}(x_i y_i) = 233491.8$$

$$\bar{y} = 169.72$$

これを（４）、（６）式に代入して計算すると、

$$\mathrm{b} = \frac{233491.8 - 20 \text{ x } 68.60 \text{ x } 169.72}{94564 - 20 \text{ x } 68.60^2} = 1.440$$

$$\mathrm{a} = 169.72 - 1.440 \text{ x } \ 68.60 = 70.88$$

と計算できる。

ここでは、先ず一つの式から一つの未知パラメーターを解いて、それを他の式に代入して別の未知パラメータを解いていったが、これでは変数の数が増えれば大変である。これを変数の数に関係なく一気に解く方法がある。それがクラメルの公式である。

クラメルの公式による正規方程式の解法

(1),(2)式をaとbを未知数とみなして、行列とベクトル形式で整理すると、

$$\Sigma a \;+\; (\Sigma x_i)\mathrm{b} = \Sigma\, y_i$$
$$(\Sigma x_i)\, a \;+ (\Sigma {x_i}^2)\, \mathrm{b} = \Sigma\, x_i y_i$$

これを係数行列および未知数ベクトルと定数項ベクトルの形に整理すると

$$\begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma {x_i}^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \Sigma\, y_i \\ \Sigma\, x_i y_i \end{bmatrix}$$

クラメルの公式により、この連立方程式の解は次のように解かれる。

$$\mathrm{a} = \frac{\begin{bmatrix} \Sigma\, y_i & \Sigma\, x_i \\ \Sigma\, x_i y_i & \Sigma\, {x_i}^2 \end{bmatrix}}{\begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma {x_i}^2 \end{bmatrix}} = \frac{(\Sigma\, y_i)(\Sigma\, {x_i}^2) - (\Sigma\, x_i)(\Sigma\, x_i y_i)}{n(\Sigma {x_i}^2) - (\Sigma x_i)^2} = 70.883$$

$$b = \frac{\begin{bmatrix} n & \Sigma\, y_i \\ \Sigma x_i & \Sigma\, x_i y_i \end{bmatrix}}{\begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma {x_i}^2 \end{bmatrix}} = \frac{n(\Sigma\, x_i y_i) - (\Sigma x_i)(\Sigma\, y_i)}{n(\Sigma {x_i}^2) - (\Sigma x_i)^2} = 1.440$$

ここでは、2変数からなる単回帰の2次元の例で説明したが、この解法はn変数からなる多次元の係数行列の場合にも適用可能である。

推定結果の応用

・ 以上の計算から、つぎのように推計できた。

$$y_i = 70.883 + 1.440\ x_i + e_t \quad \cdots（7）$$

例えば、体重が60kgである人の期待される身長は、

x_i に60を代入して、167㎝

$$70.883 + 1.440 \text{ x } 60 \fallingdotseq 167\text{㎝}$$

と推定できる。

一般に、（7）式のように特定化された関数形とパラメータの組み合わせで表現される関係式を「<u>**モデル**</u>」と呼ぶ。

モデルの説明力

- 以上で推定したモデルにより、体重と身長の関係は、どの程度説明されるのか？
  ⇒モデルの説明力を示す指標として
  「**決定係数**」がある(R^2 と表記)。
  決定係数は被説明変数の全変動TSS($\Sigma(y_i - \bar{y})$)のうち、モデルの推定式の数値の変動・ESS$\Sigma(\hat{y}_i - \bar{y})$)がどの程度の割合を占めるかを示し、0～1の間の値で示される。
  残 R^2二乗和RSSを用いても R^2 を計算でき、
  =ESS/TSS=1−RSS/TSS
  と示される。
  RSS,TSS,ESSの間にはつぎの関係が成り立つ。
  ESS+RSS=TSS

モデルの説明力と決定係数

- ここで、実際には、RSSの方を先に計算して

$$\mathrm{TSS} = \Sigma(y_i - \bar{y})^2 = 1402.40$$

$$\begin{aligned}\mathrm{RSS} &= \Sigma\left(e_i^2\right) = \Sigma(e_i(y_i - \hat{y}_i)) \\ &= \Sigma(e_i y_i) = \Sigma(y_i^2 - \mathrm{a}y_i - bx_i y_i) \\ &= 479.15\end{aligned}$$

$R^2 = 1 - RSS/TSS = 0.6583$ である。

また $\mathrm{ESS} = \Sigma(\hat{y}_i - \bar{y})^2 = 923.2$ となり、

$\mathrm{ESS} + \mathrm{RSS} = 1402.4 = \mathrm{TSS}$ という関係が成立している。

- 決定係数 R^2=0.658ということは、身長の変動のうち、この推定式では65.8%の変動が説明されることになる。

仮説検定

- 以上のモデルでは、身長は体重の増加関数として推定されたが、その信頼性を検討する。
  各々の変数の回帰係数の有意さを確認する検定の方法として「**<u>t検定</u>**」がある。

統計的仮説検定

- 統計学で、仮説を検定する場合は、
  「**<u>対立仮説</u>**」: H_a
  「**<u>帰無仮説</u>**」: H_0
  を設定し、示したい仮説を対立仮説に、否定したい仮説を帰無仮説にする。これは、ある命題が正しいことを示すには、全てのケースで成り立つことを証明する必要があるが、ある命題を否定するには例外を一つ示せば良いため、背反事象の命題を否定する方が簡単だからである。つまり、背理法を適用することになる。
  　この帰無仮説を一定の「**<u>有意水準</u>**」に対して否定できるかどうかを調べるのが統計的仮説検定の目的である。

仮説検定の例題

例）京都市の高校で36人のクラスで理科の試験を実施した結果、その平均点が65点であった。他方、同一試験の京都府全体の平均点は60点、標準偏差が12であった。点数の分布が正規分布に従っている場合、このクラスは京都府全体の平均点より高いといえるかどうかを検定する？

標本平均を $\bar{X}$

対立仮説を H_a: μ >60

帰無仮説を H_0: μ =60　として設定する。

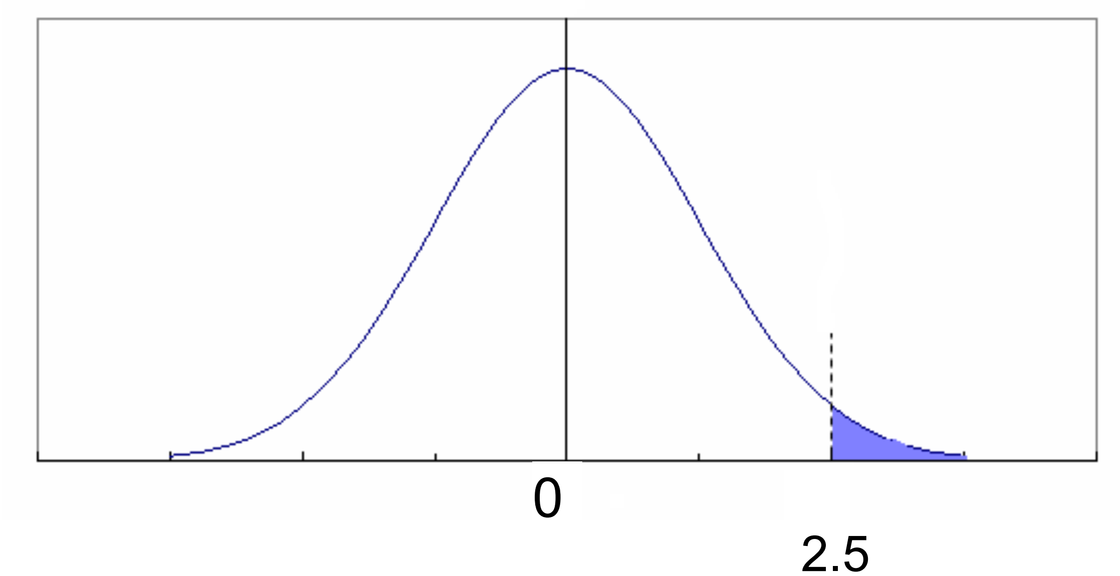
〇標準正規分布を利用するために、このサンプルを標準化して、標準化統計量に変換する。

帰無仮説の下での標準化検定統計量は

$$\mathrm{Z}_0 = (\bar{X} - 60)/(12/\sqrt{36}) = 2.5$$

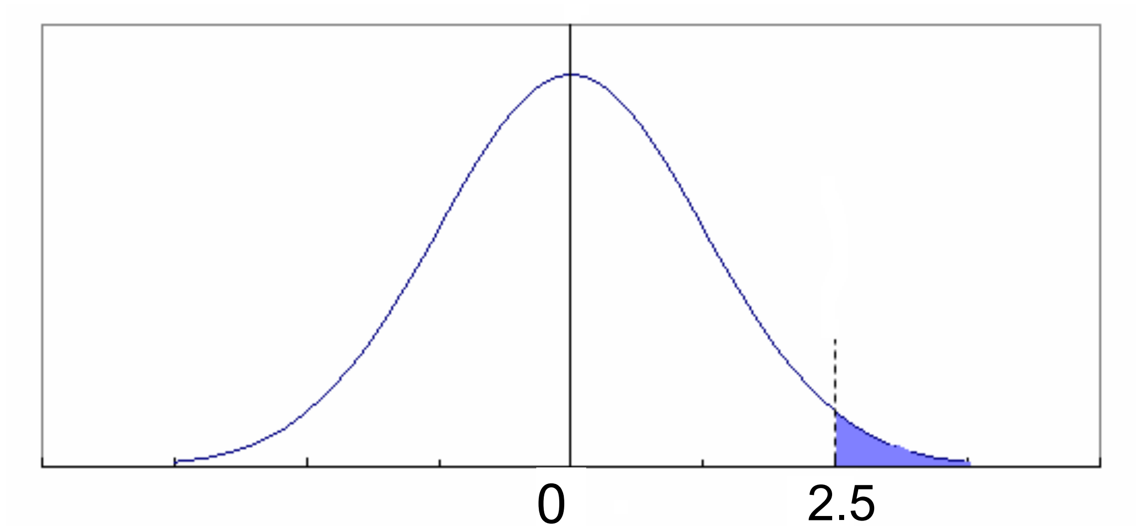
と変換される。

仮説検定と有意水準の設定



- 前述したよう標準化統計量の実現値は2.5であった。標準化統計量は帰無仮説の下で標準正規分布に従って分布している。
  統計書で用意されている標準正規分布表に従って、－∞から2.5までの確率は0.9938と分かる。このことから、帰無仮説のもとで実現値が2.5以上になる確率（P値）は　1−0.9938=0.0062となり、極めて生じ難いことが生じたことになる。ということは設定した帰無仮説が現実的ではなかったからだと判断して、これを棄却する。しかし、この帰無仮説が正しかったにも拘らず間違って棄却する確率は0.0062である。この場合、有意水準1%で帰無仮説は棄却できる。

仮説検定と棄却域



- 従って、京都市の高校の当該クラスの平均値は京都府全体の平均と同じという仮説は棄却できる。
  この時、1%有意の値は、
  $\bar{X}$≧60+2×2.33=64.66
  5%有意では、
  $\bar{X}$≧60+2×1.65=63.3　となり
  平均点が63.3点以上なら5%の有意水準で帰無仮説が棄却できるといえる。

t 検定―t 分布による検定―

- t統計量は次のように定義される。

$$T = (\bar{X} - \mu_0)/(S/\sqrt{n})$$

この場合のt検定では、対立仮説の $\mu = \mu_a$ が帰無 $\mu = \mu_0$ という特定の値と有意に異なるかどうかを調べる。
t統計量は、自由度が標本数n-1のt分布に従って分布することが知られている。

t分布の図は既に示したよう、自由度の小さい段階では正規分布からずれるが、自由度を大きくしていけば次第に正規分布に近づいていく。

係数推定値の有意性の検定

前述の身長推定モデルにおいて、体重の回帰係数は1.440、標本数は20である。ここで、検定したいのは、回帰係数が0と有意に異なるかどうかである。この場合の検定統計量としてのt統計量を求める。

$$T = (\bar{X} - \mu_0)/(S/\sqrt{n})$$

ここで、検定条件は $\mu_0 = 0,\ \bar{X} = \mathrm{b}$ であるから、

$$t = b/(S/\sqrt{n})$$

$S/\sqrt{n} = S_b$ とすると、検定に使う t 値は

$$t = b/S_b$$

となる。

t検定と自由度

これから　S^2 =RSS/(n-2)= $479.15/18 = 26.61$

$$S_b^2 = S^2/\sigma(x_i - \bar{x})^2 = 26.61/444.98 = 0.0598$$

$$S_b = \sqrt{S_b^2} = \sqrt{S^2/\sigma(x_i - \bar{x})^2} = \sqrt{26.61/444.98} = \sqrt{0.0598} = 0.2446$$

従って、　$\mathrm{t} = \mathrm{b}/S_b = 1.440/0.2446 = 5.889$　となる。

自由度は、{推定に用いたデータの数－変数（またはパラメーター）の数}である。この場合、推定した係数が2個、標本数が20なので自由度は18である。そこで、自由度18のt分布において右裾の確率が1%になる確率変数の値は2.552なので、上記のt値はこれを超えていることが分かる。従って、回帰係数βについて、帰無仮説は1%の有意水準で棄却できる。つまり、体重は身長を規定する関係にあるといえる。

系列相関の検定　----ダービン･ワトソン検定----

次に、誤差項の系列相関の有無についての検定には、ダービン・ワトソン検定が適用される。この検定は誤差項の自己相関係数を用いて次の式により検定される。

$$\mathrm{DW} = \sum_{j=2}^{n} (e_j - e_{j-1})^2 \Big/ \sum_{j=1}^{n} e_j^2$$

DW値は残差項の自己相関係数ｒと次の近似式の関係にあり、２に近いほど系列相関は無いと判断される。

$$\mathrm{DW} \fallingdotseq 2(1-\mathrm{r})$$

ここで推定した身長と体重との関係に関するモデルの場合の系列相関の検定結果は、DW＝2.190422となり、誤差項の系列相関は殆ど無いと判断される。

系列相関の検定は特に時系列データやパネルデータで問題になり、クロスセクションデータでは余り大きな意味は持たない。また、系列相関の存在は、一定期間の構造の分析よりも推定された結果に基づいて将来予測をするときに特に大きな問題になる。

系列相関の有無に関する検定は、ダービンとワトソンが提唱するDW検定表において、データの数と説明変数の数の組み合わせで決まる２つの値dLとdUに基づいて、以下のように判定される。

dw<dLの場合	正の自己相関があり、
dU<dw<4-dUの場合	自己相関が無いという仮説を棄却できない
4-dL<dw の場合、	負の自己相関がある。

DL<dw<dU または　4-dU<dw<4-dL の場合、結論は未定。

系列相関の検定を行う場合、説明変数にラグ付き従属変数がない場合にはDW比を使うが、説明変数にラグ付き従属変数がある場合には、次のダービンのｈ統計量を用いる。

$$h = \hat{\rho} \sqrt{\frac{T}{1 - T\, s_{\hat{\gamma}}^2}}$$

ここで、Tはデータ数、$s_{\hat{\gamma}}$はラグ付き従属変数の係数推定値γの標準偏差、$\hat{\rho}$はもとの最小二乗推定における残差項の１次の自己回帰係数である。

ダービンのｈ統計量は近似的に標準正規分布に従うことを利用して検定を行う。有意水準が10%の場合、

h<-645の場合、　　　　負の自己相関あり、

-1.65<h<1.645の場合、自己相関無しという仮説が棄却できない、

1.645<hの場合、　　　　正の自己相関がある。

1-nv≦0の場合には、e_jをe_{j-1}と元の説明変数とに回帰し、e_{j-1}の係数の有意性で判断する。

以上の単回帰式の推定と検定の結果をまとめると次の表のようになる。つまり、体重の係数推定値は1%水準で有意に0と異なっている。その点推定値は1.441であり、またその区間推定結果によると、その推定値は95%の信頼限界で0.927から1.955の範囲に収まる。

	係数	標準誤差	t 値	P-値	下限 95%	上限 95%
切 片	70.883	16.822	4.21	0.000522	35.543	106.224
体 重	1.441	0.245	5.89	0.000014	0.927	1.955

回帰統計	
重相関 R	0.811
重決定 R2	0.658
補正 R2	0.639
標準誤差	5.159
観測数	20
DW比	2.19

さらに、その推定式全体としての説明力に関しては、決定係数が0.65であり、生徒の身長の分布の約65%は、体重により説明できることになる。また、DW比は2.19であり、誤差項の系列相関は殆ど無い、良好な推定結果であることが知られる。

多重回帰式の推定と検定

- ここまでで、身長の推定モデルは説明力も大きく、係数の信頼性も十分高いということが確認された。
しかし、身長は体重だけで決まるわけではない。そこで、以上の最小二乗法による推定を拡充して、説明変数の数を増やすことを試みる。

多重回帰式の推定

	身長(cm)	体重(kg)	父親の身長(cm)
1	167	66	164
2	158	59	155
3	177	74	172
4	165	65	162
5	162	67	159
6	174	66	170
7	170	69	166
8	161	65	159
9	165	65	160
10	179	74	174
11	176	73	172
12	170	68	164
13	157	72	155
14	170	69	166
15	187	77	181
16	180	74	176
17	184	75	178
18	166	67	162
19	160	61	157
20	165	66	164

- 追加する説明変数として、学生の父親の身長を考える。当然、先天的には遺伝的な要素により、強い関係が考えられる。しかし、実際には母親の体格にも影響される筈であり、また、後天的な食習慣や運動経験によっても影響されるので、父親の身長だけに大きく関係するとは限らない。
- この場合、本人の体重と並んで、父親の身長はどの程度強く、子供の身長に関係するかを推定する。

多重回帰式と最小二乗法

- 父親の身長を示す説明変数を z_i としたとき、回帰式は下のようになる。

$$Y_i = \alpha + \beta x_i + \gamma z_i + U_i$$

- 単回帰の場合と同様に、残差二乗和を最小にすることにより、各々の係数を推定するので、それぞれの係数について偏微分して0とおく。

$$\Sigma(y_i - a - bx_i - cz_i) = 0 \quad \cdots (8)$$

$$\Sigma((y_i - a - bx_i - cz_i)x_i) = 0 \quad \cdots (9)$$

$$\Sigma((y_i - a - bx_i - cz_i)z_i) = 0 \quad \cdots (10)$$

この（8）～（10）式からなる連立方程式を解くことにより、各々の回帰係数を推定することになる。

以上に示したように、学生の身長を本人の体重及び父親の身長で説明する重回帰分析の推定結果は以下のようにまとめられる。父親の身長に関する係数の点推定値は1.037であり、また、95%の信頼限界での区間推定は0.819から1.256の範囲となる。

回帰統計	
重相関 R	0.975
重決定 R2	0.951
補正 R2	0.945
標準誤差	2.020
観測数	20
DW比	2.31

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	-23.786	11.517	-2.07	0.05449176	-48.084	0.513
体重	0.314	0.148	2.12	0.04869822	0.002	0.625
父親の身長	1.037	0.104	10.02	0.00000002	0.819	1.256

父親の身長を説明変数に加えることによって、決定係数は大幅に改善された。また、父親の身長の係数推定値は1%水準で有意であり、そのP値は殆どゼロに近い。他方、もとからの説明変数である体重の係数推定値の有意性は若干低くなっていることが分かる。単回帰の時には1%水準で有意であったが、この重回帰分析では1%水準では有意ではなくなり、5%水準で辛うじて有意になっている。誤差項の系列相関に関しては、ダービン・ワトソン比（DW比）が2に近いのでほぼ系列相関も小さいと言えるが、単回帰の結果に比べて2を上回る度合いがやや大きくなっている。
また、学生の身長は、本人の体重よりも、父親の身長により規定される度合いが強いことがわかる。

定常性の検定と時系列解析

通常の回帰分析では、決定係数による説明変数全体としての説明力の程度、各係数推定値の有意性の検定、誤差項の系列相関の有無の検定で十分であるが、サンプル期間の長い時系列データの場合には、その定常性が問題になる。定常性を満たす時系列データとは、変数の自己相関係数を、もとの変数に関してラグをゼロから順番にずらしてとっていった場合に、ラグの数だけラグ付き自己相関係数（時差相関係数）ができる。これをラグに関して図示したものがコレログラムである。

一般的には、このラグ付き自己相関係数はもとの時系列変数の開始時点とラグの両方の関数となる。この時、データの開始時点には依存せず、ラグのみの関数になる場合にその時系列データは定常性を満たしていることになる。

定常性と単位根検定

つまり、もとの時系列データの自己相関係数の系列が

$$R_{t_0,\tau} = f(t_0, \tau) = f(\tau)$$

を満たす時、もとの時系列データは定常であると定義される。回帰分析に使用される各変数がこの定常性を満たさない非定常な変数を含む場合、その他の諸検定をパスしても、それは「見せかけの回帰」である恐れがある。

　この定常性は単位根検定によって検討される。単位根を持つ時系列データは定常性を満たさない。そこで、一定の有意水準の下で、単位根が有るという帰無仮説が棄却できなかった場合には、そのデータ系列は非定常とみなされる。

共和分検定

一般に、重回帰分析において、幾つかの変数が非定常であったとしても、それらの変数を線形結合したものが定常性を満たしている場合、これらの変数は共和分関係にあると言われる。この場合には、幾つかの変数が非定常であっても、その重回帰分析の結果は「見せかけの回帰」ではなく、長期的に均衡した安定的な関係として推定される。ということは、通常の重回帰式において、誤差項を左辺に移項して従属変数として変形した式の右辺は各説明変数の線形結合になっている。それ故、誤差項に関して、単位根検定を適用して、それが定常性を満たしていれば、これらの変数は共和分関係にあることになり、そのまま重回帰分析を適用できることになる。この検定方法は共和分検定と言われる。

＜質的選択モデルによる接近＞

質的選択モデルには、プロビットモデル、ロジットモデル、トービットモデルが良く知られている。
通常の重回帰モデルと比較して、被説明変数（従属変数）が離散変数であるため、誤差項に通常の連続確率分布を想定できない。
それ故、特殊な分析手法が必要となる。

こうした分析には、観察不可能な潜在変数という概念を導入する。
ここでは、海外直接投資予定額FI*を考察する。
潜在変数FI*と説明変数（例えば企業の資本金や従業員の規模、進出地域など）との間に以下の関係を仮定する。

$$FI^* = \beta_1 + \beta_2 X_i + e_i$$

FIとFI*との関係は

$$FI = \begin{cases} 1 & FI^* > 0 \\ 0 & FI^* < 0 \end{cases}$$

によって関係づけられる。　ここで、FI*は特定地域への海外直接投資予定額である。これがプラスの企業は、アンケートにおいてその地域へ進出して海外事業展開をすることに関心がある（FI＝１）と答えることになり、特定地域への海外直接投資予定額がマイナスの企業は、アンケートにおいてその地域へ進出して海外事業展開をすることに関心がない（FI＝０）と答えることになる。

　このモデルでは、

$$
\begin{aligned}
Pi &= Pr(FI=1) = Pr(FI^* > 0) \\
&= Pr(\varepsilon_i > -(\beta_1 + \beta_2 X_i)) \\
&= Pr(\varepsilon_i < (\beta_1 + \beta_2 X_i)) \qquad \cdots\cdots\cdots\cdots\cdots (1)
\end{aligned}
$$

となる。

　この式を用いて分析するためには、誤差項ε_iに関する分布関数を特定化する必要がある。これに標準正規分布を仮定するのがプロビット・モデルである。
　標準正規分布の場合の累積分布関数を使用すると、(1)式は

$$
\begin{aligned}
P_i &= Pr(\varepsilon_i < (\beta_1 + \beta_2 X_i)) \\
&= \Phi(\beta_1 + \beta_2 X_i) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta 1 + \beta 2 Xi} \exp(-\frac{z^2}{2}) \ dz
\end{aligned}
$$

この分布関数のパラメータβ_1、β_2をデータに対して最も当てはまりが良くなるように決定する訳である。

　このようにプロビット・モデルは、数学的に扱い難い標準正規分布の累積分布関数を使用することになるため、多くの実証分析では、標準ロジスチック分布の累積密度関数、

$$\Delta\ (\beta_1+\beta_2 X_i) = \frac{1}{1+e^{-\beta_1-\beta_2 X i}}$$

を想定したロジット・モデルが多用される。

＜トービット・モデル＞

従属変数に制約がある場合の第2タイプのモデルがトービット・モデルであり、
これは、従属変数の値の取り得る範囲に対して何らかの制限が課されているケース（打ち切りデータ・モデル）である。
プロビット・モデルやロジット・モデルの従属変数が離散的データであるのに対して、トービット・モデルの場合は、従属変数は連続的な実数が分析対象となる。
このトービット・モデルには、閲覧モデル（Censored model）と　切断モデル　（Truncated model）の２種類がある。

閲覧モデルの場合は、従属変数のデータがサンプルの分析対象となる範囲でのみ収集されるが、説明変数に関しては各サンプルの全ての範囲の属性データが収集されているケースである。

他方、切断モデルの場合は、従属変数も説明変数も共に各サンプルの分析対象となる範囲のみのデータが収集されており、分析対象とならない範囲のデータはどの変数（従属変数も説明変数も）についても存在しないケースである。
閲覧モデルの場合にも、質的選択モデルと同じく潜在変数が導入される。本稿の分析対象の場合、例えば、潜在的な直接投資予定額FI*は、

$$FI_i^* = \beta_1 + \beta_2 X_i + e_i$$

ここで、モデルの被説明変数FIは、海外直接投資を計画する企業については直接投資予定額そのものFI*であり、海外直接投資を計画しない企業の場合には０となる。

この背景には、次の判断がある。ある地域に進出することによって得られる期待利潤をπとすると、

$$\pi_i = \gamma_0 + \gamma_1 Z_i$$

ここで、Ziは特定地域に海外進出した企業の期待利潤を規定する要因となる種々の変数である。この期待利潤がプラスならば、この企業は潜在変数としての海外直接投資を実施しようとするが（$FI_i = FI_i*$）、期待利潤πがマイナスならば潜在的な海外直接投資計画を控えることになる（$FI_i = 0$）。

$$\text{If} \quad \pi_i > 0 \quad \text{then} \quad FI_i^* > 0$$
$$\text{If} \quad \pi_i \leqq 0 \quad \text{then} \quad FI_i^* \leqq 0$$

つまり、まず各企業は海外進出するか否かを意志決定し、つぎに進出決定した企業に関してその海外直接投資額の大きさを分析することになる。従って、潜在変数FI*と現実の海外直接投資FIとの関係は以下の通りとなる。

$$FI = \begin{cases} FI^* & FI^* > 0 \\ 0 & FI^* \leq 0 \end{cases}$$

つまり、質的選択モデル（プロビット・モデルやロジット・モデル）とトービット・モデルの違いは、潜在変数の値がプラス（**FI*＞０**）の場合に前者では従属変数は１に対応するのに対して、後者では従属変数が潜在変数の値そのものに対応することである。また、両モデルの場合とも、潜在変数の値がゼロ以下（**FI*≦０**）の場合には、従属変数は0に対応する。

＜トービット・モデルの推定結果＞

　この推定に関するデータは、東洋経済「海外進出企業総覧」（会社別編）の最新（2009年）版である。この統計書から、農林水産業、食品業、食品卸売業、飲食・外食業の各業種部門の欄を抽出した。これらの業種部門ごとに、海外直接投資、資本金、従業員、売上高、進出分野、進出年次に関する情報を変数として使用し、この中で、海外直接投資額を従属変数とするトービット・モデルを推定した。

モデル１ ＜農林水産業＞従属変数　海外直接投資額

FDIB

説明変数	係数	z-値	p -値
定数項	-1.195	-1.74	0.08
SB（販売額）	0.022	17.2	0.00
ET（従業員数販売額）	1.319	3.47	0.00
APCC（アジア太平洋地域）	-2.438	-2.25	0.02
NAFTA（北米自由貿易地域）	0.733	1.067	0.28
EU（欧州共同体）	-2.142	-0.838	0.40
_89（1980年代進出ダミー）	0.077	0.195	0.84
対数尤度＝22.3　　AIC=-0.872　　サンプル数＝33			

モデル１は、農林水産業関連企業に関して、その海外直接投資額を従属変数としたトービットモデルの推定結果である。この結果は、売上額が大きく雇用規模の大きい農林水産業関連企業ほど、海外直接投資は大きくなる傾向にあることを示している。また、進出先別に見るとNAFTA地域への進出企業ほど海外投資額は大きいが、EUやアジア太平洋地域への進出企業の場合にはその海外投資額は比較的小さい傾向にあることが示される。また進出年に関しては、1980年代にそれ以前に比べてより大きく海外直接投資が拡大したことを示している。

モデル２ ＜食品産業＞従属変数　海外直接投資額 FDIB

説明変数	係数	z-値	p -値
C（定数項）	-30.837	-2.78	0.00
SB（売上額）	0.030	0.15	0.87
ET（従業員数）	8.608	1.55	0.11
APCC（アジア太平洋地域）	-18.149	-1.90	0.05
対数尤度＝-122.9　AIC=1.96　サンプル数＝130			

モデル２は、日系食品企業に関して、その海外直接投資額を従属変数としたトービットモデルの推定結果である。この結果は、売上額が大きく、雇用者数の規模が大きい日系食品企業ほど、その海外直接投資額も大きくなる傾向にあることを示している。また、進出先別に見ると、アジア太平洋地域に進出している日系食品企業は他の地域への進出企業に比較して海外直接投資額が小さい傾向にあることが分る。

モデル３ ＜食品卸売企業＞ 従属変数　海外直接投資額 FDIB

説明変数	係数	z-値	p -値
C（定数項）	-0.052	-0.31	0.75
SB（売上額）	-0.008	-0.17	0.86
ET（従業員数）	1.159	4.99	0.00
APCC（アジア太平洋地域）	-0.231	-1.55	0.11
_89（販売・輸出入業）	0.232	1.39	0.16
対数尤度＝35.6　AIC=-2.12　サンプル数＝28			

モデル３は、日系食品卸売り企業に関して、その海外直接投資額を従属変数としたトービットモデルの推定結果である。この結果は、見かけ上は売上規模の大きい企業の場合に海外直接投資額を小さくする傾向のあることを示しているが、この係数の推定値は有意ではない。また、従業員規模の大きい企業ほど、その海外直接投資額も大きくなる傾向にあることを示している。さらに、進出地域別に見ると、アジア太平洋地域へ進出している企業の場合には、他の地域に進出している日系食品卸売企業に比較して、その海外直接投資額が小さい傾向にあることが示される。加えて、進出年次別に見ると、1980年代に進出した日系食品卸売企業はそれ以前の年

モデル4　＜飲食・外食企業＞従属変数　海外直接投資額 FDIB

説明変数	係数	z-値	p -値
C（定数項）	0.338	2.21	0.02
SB（売上額）	0.038	1.23	0.21
ET（従業員）	-0.388	-0.64	0.52
APCC（アジア太平洋地域）	-0.259	-1.38	0.16
_89(80年代進出ダミー)	0.298	1.22	0.21
対数尤度＝8.85　AIC=-0.38　サンプル数＝15			

モデル４は、日系の飲食・外食企業に関して、その海外直接投資額を従属変数としたトービットモデルの推定結果である。この結果は、売上規模の大きい日系外食・飲食企業ほど、海外直接投資額も大きくなる傾向にあることを示している。しかし、従業員規模の大きい企業の場合には海外直接投資額が逆に小さくなる傾向にあることが示される。これは外食・飲食産業のようなサービス産業の場合には従業員規模は必ずしも企業の経営規模を反映しないという状況を示唆しているとも考えられるが、この係数推定値は統計的に有意ではないので余り断定的に解釈することはできない。また、進出地域別に見ると、アジア太平洋地域へ進出している日系飲食・外食企業の場合には、他の地域に進出している企業に比較して海外直接投資額は小さい傾向にあることが示される。加えて、進出年次別にみると、1980年代に進出した日系飲食・外食企業は、概してそれ以前の時期に進出した企業に比較して海外直接投資額が拡大していることがわかる。これは、海外に進出している他の業種の日系食品企業の場合にも共通している。

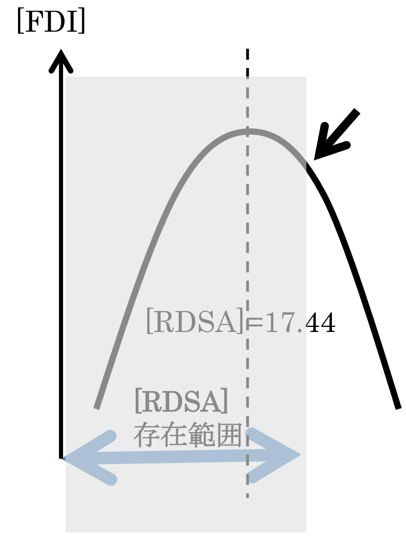
海外直接投資の規定要因に関するトービット・モデル

係数
t値

		(1)	(2)	(3)	(4)	(5)	(6)
為替レート	RATIO	−0.001		−0.002			
		−0.110		−0.140			
売上高	LSALE	2.655	2.630	3.638	2.499		9.719
		1.130	1.120	1.570	1.050		2.95
二乗	LSALEsq	−0.092	−0.090	−0.186	−0.076	−3.520	−0.782
		−0.38	−0.37	−0.77	−0.31	−1.60	−2.46
収益性	PFSA	0.069	0.067	0.074	0.072		
		1.56	1.53	1.71	1.60		
二乗	PFSAsq	−0.001	−0.001	−0.001	−0.001		
		−0.84	−0.82	−0.90	−0.92		
成長性	GROWTH	−0.176				−0.252	
		−0.58				−0.81	
二乗	GROWTHsq	0.002					
		0.22					
労働力	L	6.13E−05	6.07E−05	0.000116	5.61E−05		
		0.43	0.43	0.84	0.39		
二乗	Lsq	−4.1E−09	−4.1E−09	−6.7E−09	−4.2E−09		
		−0.65	−0.64	−1.07	−0.64		
マーケティング・ノウハウ	ADSA	0.188	0.189		0.184	0.053	
		1.48	1.5		1.43	2.17	
二乗	ADSAsq	−0.017	−0.017		−0.016		
		−1.05	−1.06		−0.99		
研究開発力	RDSA	0.339	0.341	0.360	0.346	0.558	40.482
		3.15	3.23	3.36	3.27	4.86	5.28
二乗	RDSAsq	−0.003	−0.004	−0.004	−0.005	−0.016	
		−0.63	−0.67	−0.73	−0.85	−2.86	
ダミーアジア通貨危機	dumASIAcri	−1.040	−1.067	−1.036			
		−3.57	−3.96	−3.58			
ダミーリーマンショック	dumriesk	−1.365	−1.333	−1.399			
		−2.59	−2.92	−2.67			
定数項	Cons.	−14.751	−14.842	−17.117	−14.952	−4.619	−32.450
		−2.47	−2.57	−2.89	−2.55	−11.28	−3.74
対数尤度	Log Likelihood	−600.03	−600.26	−602.19	−613.37	−679.58	−308.21
観測値	Obs.	2796	2796	2796	2930	2931	1231

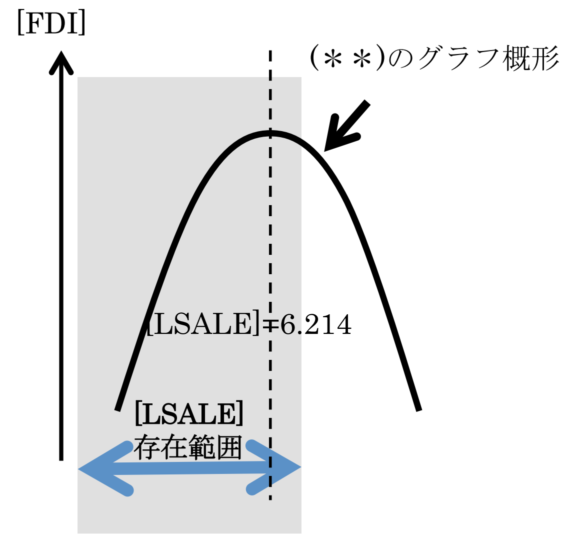
研究開発費RDSAと海外直接投資FDI

$$[FDI] = -0.016[RDSA]^2 + 0.558[RDSA] + \ldots + [Cons.]$$



研究開発費LSALEと海外直接投資FDI

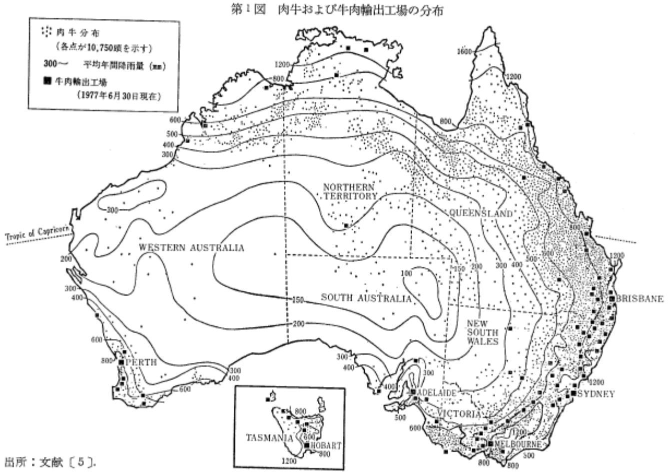
$$[FDI] = -0.782[LSALE]^2 + 9.719[LSALE] + \ldots + [Cons.]$$



スペクトル分析

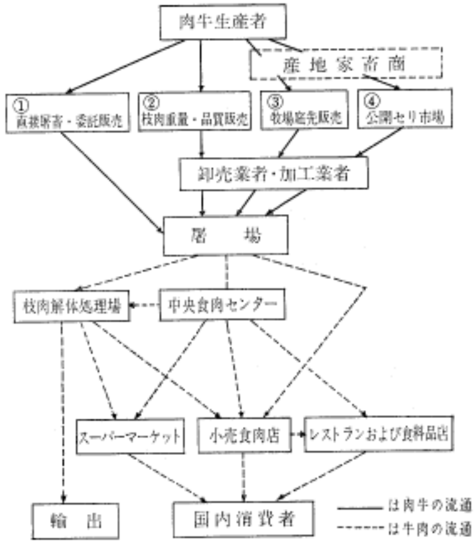
加賀爪　優「オーストラリアにおけるビーフサイクルのスペクトル分析」『農業総合研究』、第33巻第 4号、p149～183、1979年10月

第１図　肉牛および牛肉輸出工場の分布

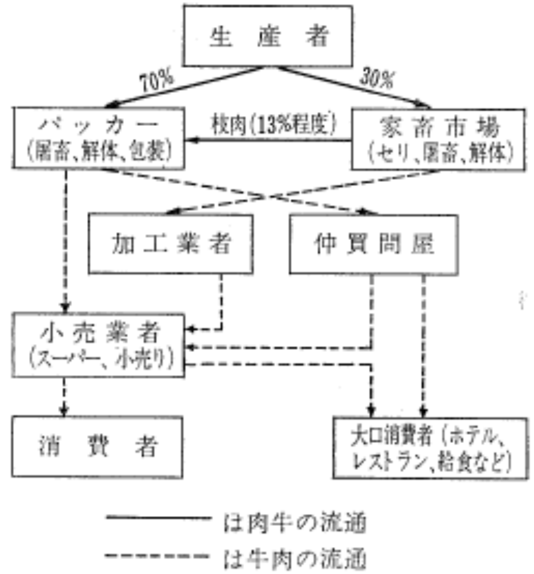


出所：文献〔5〕.

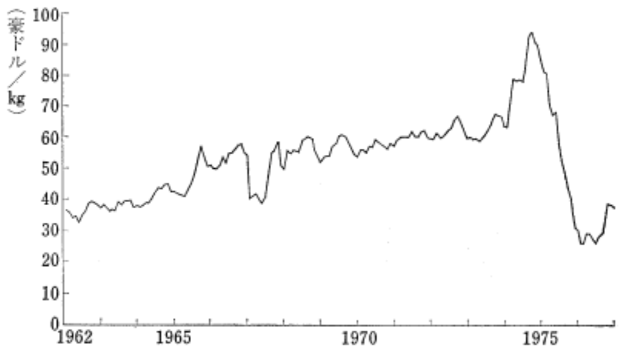
第２—ａ図　オーストラリアにおける牛肉流通経路



第２—ｂ図　アメリカにおける牛肉流通経路

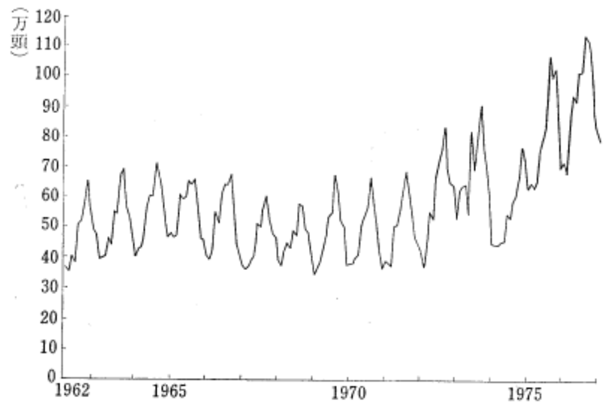


第３図　ＭＢＰ系列（牛肉卸売価格）



第４図　ＡＯＰ系列（雄壮齢牛価格）

第５図　ＳＴＮ系列（総屠殺頭数）



（1）コレログラムと定常時系列

$$R(\tau)=C(\tau)/C(o)$$
$$=\frac{E[(X(t)-\overline{X(t)})(X(t+\tau)-\overline{X(t+\tau)})]}{C(o)} \cdots\cdots(1)$$

$$X(t)=X(t\pm nT) \cdots\cdots(2)$$

（2）自己パワースペクトル分析

<時間領域>　　<周波数領域>

$$X(t)=a_0+\sum_{j=1}^{n}(a_j\cos\theta_j t+b_j\sin\theta_j t) \cdots\cdots(3)$$

$$=a_0+\sum_{j=1}^{n-1}d_j\cos(\theta_j t+\phi_j)+a_n\cos\pi t \cdots\cdots(4)$$

ここで

$\theta_j=2\pi j/n$：角周波数

$d_j=\sqrt{a_j{}^2+b_j{}^2}$：振幅

$\phi_j=\tan^{-1}\left(-\frac{b_j}{a_j}\right)$：位相角

a_j, b_j はフーリエ係数：$a_0=E[X(t)]$

$$a_n=\frac{1}{n}\sum_{t=1}^{n}(-1)^t X(t)$$

$$a_j=\frac{1}{n}\sum_{t=1}^{2n}X(t)\cos\theta_j t$$

$$b_j=\frac{1}{n}\sum_{t=1}^{2n}X(t)\sin\theta_j t$$

周波数領域での分散のペリオドグラム

$$P(\theta_j)=\frac{1}{\pi}\left\{1+2\sum_{\tau=1}^{\infty}R(\tau)\cos\theta_j\tau\right\}$$

$$=\frac{1}{\pi}\sum_{\tau=-\infty}^{\infty}R(\tau)\cos\theta_j\tau \quad \cdots\cdots\cdots\cdots\cdots\cdots(5)$$

$$\begin{cases}\bar{P}_0=0.5\,P_0+0.5\,P_1\\ \bar{P}_\tau=0.25\,P_{\tau-1}+0.5\,P_\tau+0.25\,P_{\tau+1}\\ \qquad\qquad (\tau=1,\cdots,m-1) \quad \cdots\cdots\cdots(6)\\ \bar{P}_m=0.5\,P_{m-1}+0.5\,P_m\end{cases}$$

クロススペクトル分析

$$\begin{cases}X(t)=a_{x0}+\sum_{j=1}^{n}(a_{xj}\cos\theta_j t+b_{xj}\sin\theta_j t)\\ Y(t)=a_{y0}+\sum_{j=1}^{n}(a_{yj}\cos\theta_j t+b_{yj}\sin\theta_j t)\end{cases} \quad \cdots\cdots(7)$$

$$P_{xy}(\theta_j)=\frac{1}{\pi}\sum_{\tau=-\infty}^{\infty}R_{xy}(\tau)\,e^{-\theta_j\tau i} \cdots\cdots\cdots\cdots\cdots\cdots(8)$$

（ここで、Rxy(τ）はラグτの下でのクロス相関係数を示している）

$$
\begin{cases}
P_{xy}(\theta_j) = S_{xy}(\theta_j) - Q_{xy}(\theta_j) \cdot i \\
\qquad\quad = D_{xy}(\theta_j) e^{-i\phi_{12}(\theta_j)} \\
D_{xy}(\theta_j) = \sqrt{S_{xy}{}^2(\theta_j) + Q_{xy}{}^2(\theta_j)} \\
\qquad\qquad \text{：クロス振幅スペクトル}
\end{cases}
\quad \cdots\cdots\cdots\cdots\cdots (9)
$$

＜コヒアレンス＞
2系列間の各周波数ごとの関連度を示す尺度で、回帰分析でいう決定係数に相当する。

$$
CHR^2(\theta_j) = \frac{D_{xy}{}^2(\theta_j)}{P_x(\theta_j) \cdot P_y(\theta_j)}; \quad 0 \leqq CHR^2(\theta_j) \leqq 1 \cdots (10)
$$

＜ゲイン＞
クロス振幅スペクトルDxy(θj)を入力系列のパワースペクトルPx (θj)で除したものはゲインと言われ、回帰分析でいう回帰係数（入力系列Xに対する出力系列Yの回帰係数）に相当する。

$$
G(\theta_j) = \frac{D_{xy}(\theta_j)}{P_x(\theta_j)} \cdots\cdots\cdots\cdots\cdots (11)
$$

＜フェイズ＝位相＞
入列系列X(t)と出列系列Y(t)との位相差を示すもので、次式によりラジアン単位（弧度法）で測られる。

$$
\phi_{xy}(\theta_j) = \tan^{-1} \frac{Q_{xy}(\theta_j)}{S_{xy}(\theta_j)} \cdots\cdots\cdots\cdots\cdots (12)
$$

＜タウ統計量＝タイムラグ＞
ラジアン単位で示された位相差は角周波数2πθjで時間単位へ変換できる。

$$
\tau(\theta_j) = \frac{\phi_{xy}(\theta_j)}{2\pi\theta_j} \cdots\cdots\cdots\cdots\cdots (13)
$$

この統計量τ（θj）がプラスならば入力系列X(t)が出力系列Y(t)に先行していることが示され、τ（θj）がマイナスならば、入力系列X(t)が出力系列Y(t)に遅行していると解釈される。

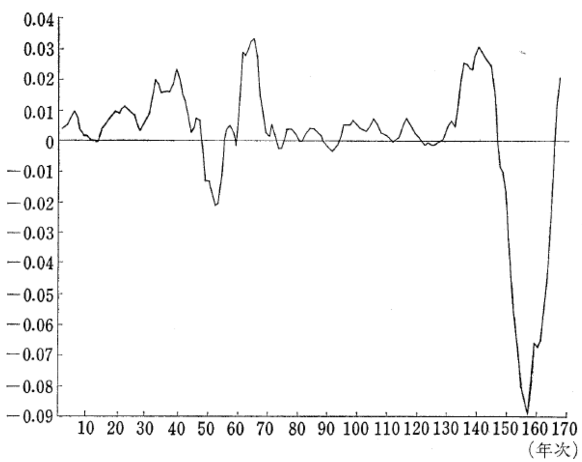
データ系列とその定常化

［１］MBP系列（牛肉卸売市場価格、kg当り豪ドル、1962年1月～1976年12月の180ヶ月）
［２］AOP系列（（壮齢肉牛市場価格、kg当り豪ドル、1962年1月～1976年12月の180ヶ月）
［３］STN系列（総屠殺頭数、単位は万頭、1962年1月～1976年12月の180ヶ月）

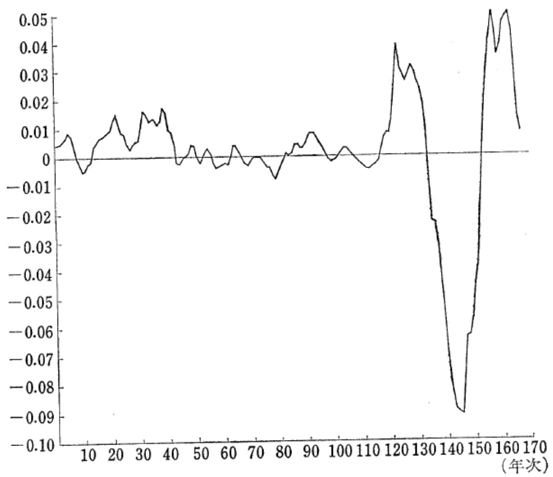
$$x(t)=\{X(t)-X(t-1)\}/X(t-1) \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (14)$$

$$x(t)=\frac{1}{12}\sum_{\mu=-5}^{5}X_{t+\mu}+\frac{1}{24}(X_{t+6}+X_{t-6}) \cdots\cdots\cdots\cdots\cdots (15)$$

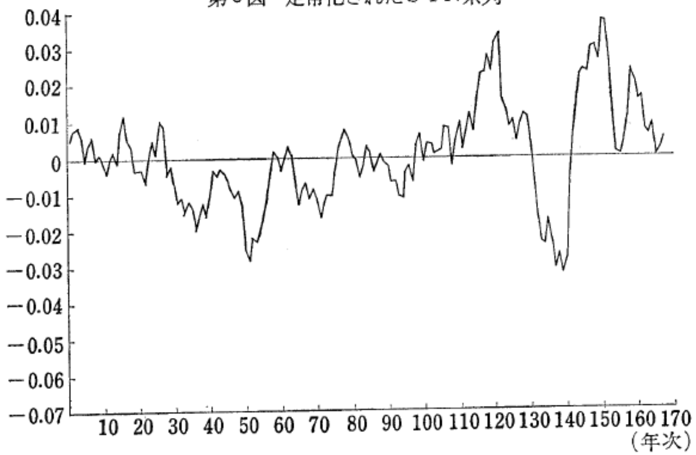
第６図　定常化されたＭＢＰ系列



第７図　定常化されたＡＯＰ系列

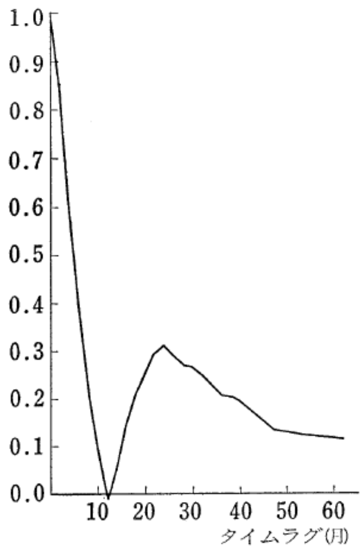


第８図　定常化されたＳＴＮ系列

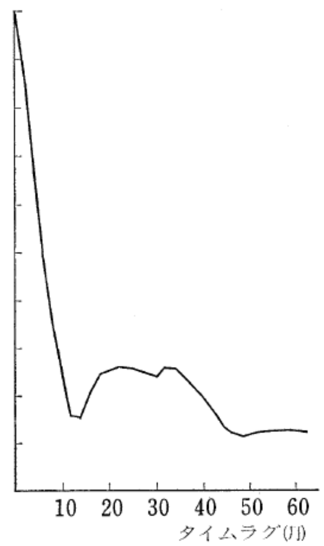


（1）コレログラムによる分析

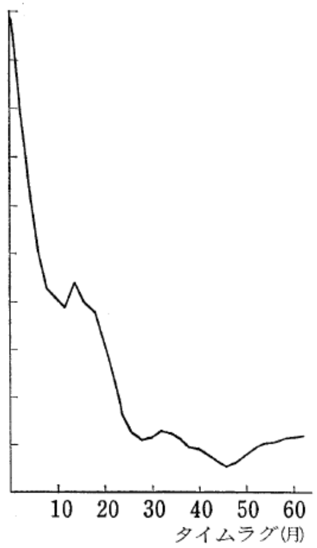
第9図　MBP系列のコレログラム



第10図　AOP系列のコレログラム

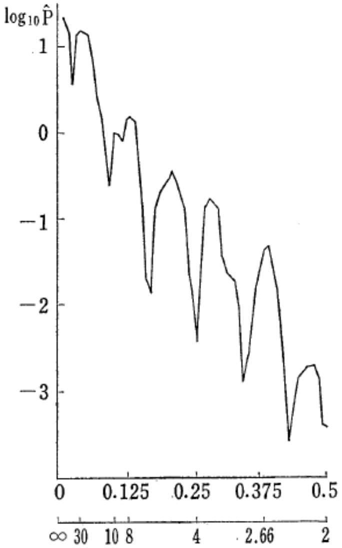


第11図　STN系列のコレログラム

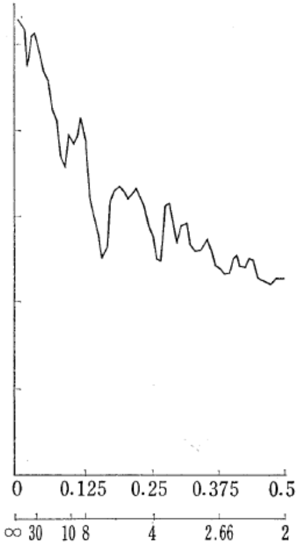


（2）自己パワースペクトル分析

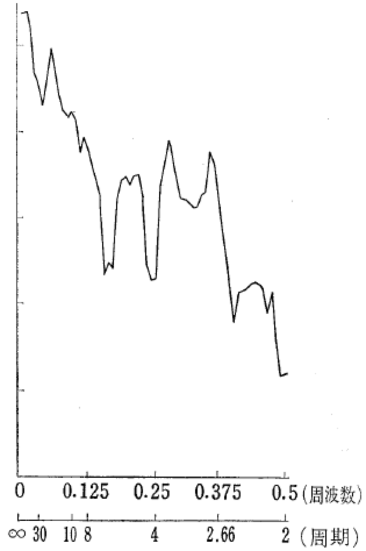
第12図　MBP系列のスペクトル密度



第13図　AOP系列のスペクトル密度



第14図　STN系列のスペクトル密度



相対的ピークの信頼性に関する有意性検定

$$P\left[\chi_{\alpha_1}{}^2(V) \leqq \frac{V\hat{P}(\theta_j)}{P(\theta_j)} \leqq \chi_{\alpha_2}{}^2(V)\right]=(\alpha_1-\alpha_2)=\alpha_3$$

ここで、χα22(V)はカイ二乗分布の上側α%点、α1とα2はχ2分布の上側累積確率、
P(θj) は正規分布に従うランダム変量のスペクトル密度の理論値、
$\hat{P}(\theta_j)$ はその推定値である。ここで、Vは等価自由度と言われる近似式で

$$V=a\cdot\frac{N}{M}$$

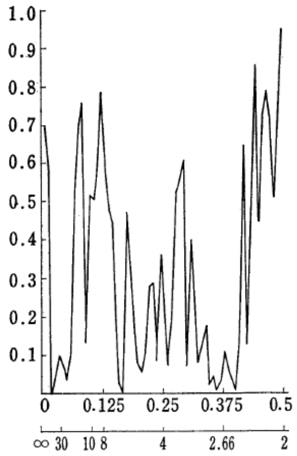
で計算している。このVは、元の系列が必ずしも正確に世紀部ぷにした
がっているとは限らないという事情を考慮して選ばれる。

第1表 スペクトルピークの仮説検定
（サンプルサイズ＝167 自己相関関数のラグ＝32 等価自由度＝19.31）

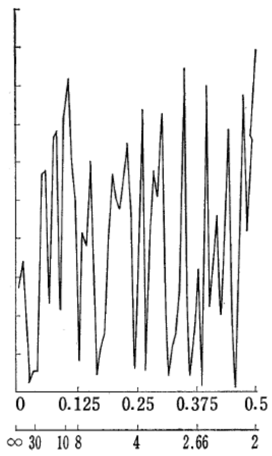
系 列	周 期（周波数）	相対的ピークのスペクトル密度	ホワイトノイズの95%信頼区間	その他のスペクトルピークの周期帯	
	カ月 サイクル/月			カ月	カ月
MBP	31.74 (0.0315)	17.9	15.2～38.8	7.93	4.87
	10.58 (0.0945)	0.993	0.844～2.154	2.59	2.11
	3.62 (0.276)	0.226	0.192～0.490		
STN	127.06 (0.007)	38.5	32.7～83.5	5.07	4.54
	18.14 (0.055)	9.64	8.19～20.91	2.82	2.26
	10.58 (0.094)	2.17	1.84～4.70	2.11	
	3.62 (0.276)	0.901	0.765～1.955		
AOP	31.74 (0.0315)	12.8	10.8～27.7	8.47	5.29
	10.58 (0.0945)	0.943	0.801～2.046	4.54	3.53
	3.17 (0.315)	0.0892	0.0758～0.1935	2.82	

クロススペクトル分析

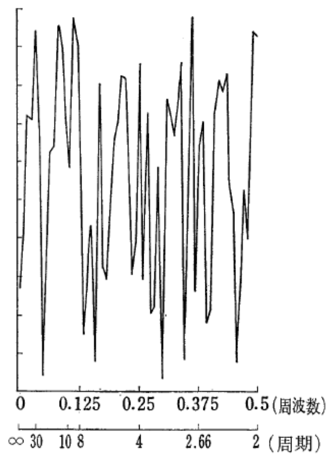
第15図　MBP—AOPのコヒアレンスダイアグラム



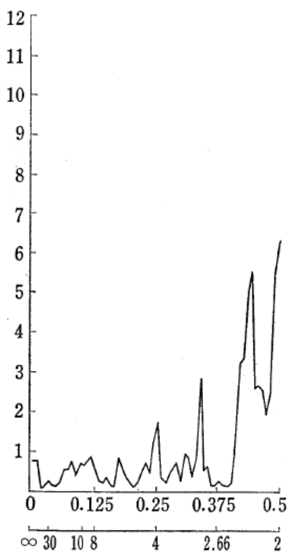
第16図　MBP—STNのコヒアレンスダイアグラム



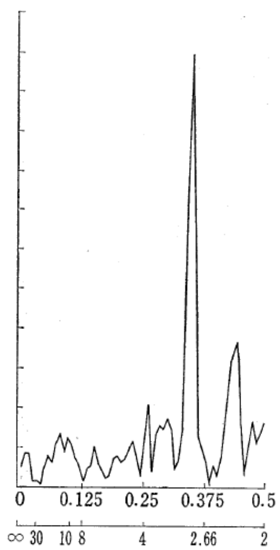
第17図　STN—AOPのコヒアレンスダイアグラム



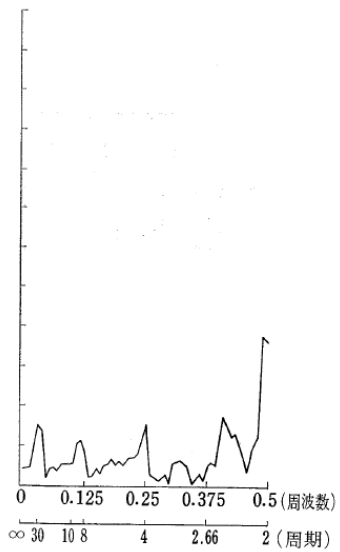
第18図　MBP—AOPのゲインダイアグラム



第19図　MBP—STNのゲインダイアグラム



第20図　STN—AOPのゲインダイアグラム



第2表　クロススペクトル推定値

（コヒアレンスの主な極大値のみ列挙）

［A］ MBP—AOP系列

周　期（周波数）	コヒアレンス	ゲ　イ　ン	フ　ェ　イ　ス	タイムラグ
カ月 サイクル/月				カ月
31.74 (0.0315)	0.0930	0.258	0.610	3.08
8.47 (0.118)	0.785	0.844	−0.708	−0.954
3.43 (0.291)	0.615	0.703	1.44	0.785
2.26 (0.441)	0.85	5.56	1.49	0.538

［B］ MBP—STN系列

12.7 (0.0787)	0.682	1.32	1.06	2.15
10.58 (0.0945)	0.716	1.25	−1.43	−2.41
9.8 (0.102)	0.813	1.09	−1.55	−2.41
2.89 (0.346)	0.843	10.9	−0.459	−0.211
2.11 (0.472)	0.772	1.68	1.48	0.498

［C］ STN—AOP系列

31.74 (0.0315)	0.940	1.52	−1.41	−7.11
12.70 (0.0787)	0.952	0.552	−0.265	−5.36
10.58 (0.0945)	0.705	0.553	−3.14	−5.28
9.09 (0.110)	0.973	1.07	1.86	2.69
2.94 (0.339)	0.854	0.445	−0.189	−0.0889
2.76 (0.362)	0.968	0.308	−2.98	−1.31
2.04 (0.488)	0.934	3.79	0.227	0.074

(i) 約1年周期の成分変動間の先行遅行関係

$$[\mathrm{AOP}] \xrightarrow[\text{拡大変動的（×1/0.553）}]{\text{5.28ヵ月}} [\mathrm{STN}]$$

$$CHR^2=0.705$$

$$\xrightarrow[\text{縮小変動的（×1/1.25）}]{\text{2.41ヵ月}} [\mathrm{MBP}]$$

$$CHR^2=0.716$$

(ii) 約3ヶ月周期の成分変動間の先行遅行関係

$$[\mathrm{MBP}] \xrightarrow[\text{縮小変動的（×0.703）}]{\text{0.785ヵ月}} [\mathrm{AOP}]$$

$$CHR^2=0.615$$

$$\xrightarrow[\text{拡大変動的（×1/0.445）}]{\text{0.0889ヵ月}} [\mathrm{STN}]$$

$$CHR^2=0.854$$

(iii) 約3年周期の成分変動間の先行遅行関係

$$[\mathrm{MBP}] \xrightarrow[\text{縮小変動的（×0.258）}]{\text{3.08ヵ月}} [\mathrm{AOP}]$$

$$CHR^2=0.0930$$

$$\xrightarrow[\text{縮小変動的(×1/1.52)}]{\text{7.11ヵ月}} [\mathrm{STN}]$$

$$CHR^2=0.940$$

数理計画法

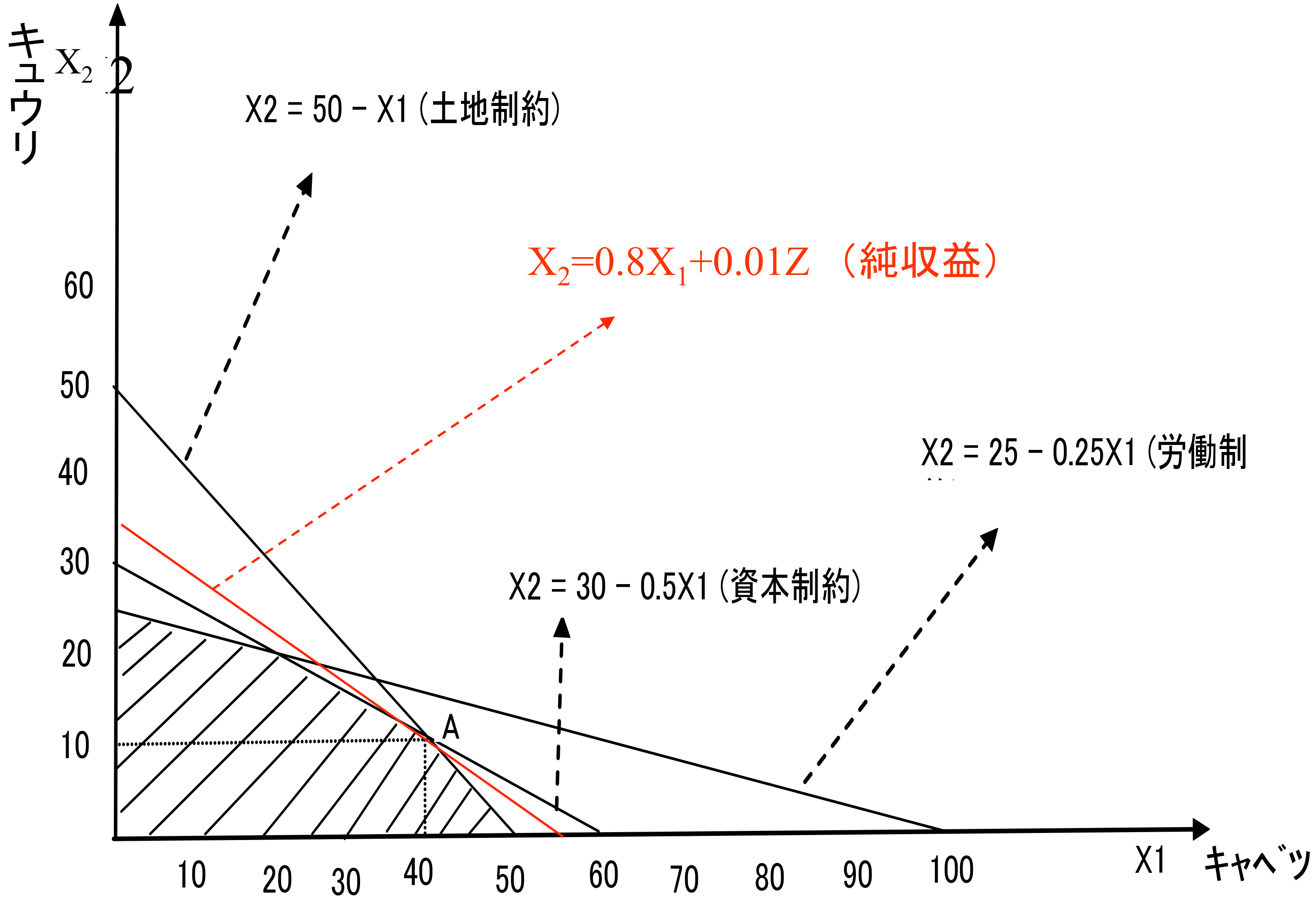
線形計画問題

ある農村に50haの畑があり、キャベツ(X_1 ha)とキュウリ(X_2 ha)を作付する。
キャベツ1ha作るのに労働力1人、キュウリ1ha作るのに4人を要するが、
全部で100人まで動員できる
キャベツ1ha当り50万円、キュウリ1ha当り100万円必要であり、
合計3000万円まで使用できる。
キャベツ1ha当り80万円、キュウリ1ha当り100万円の純収益が得られるとする。
全純収益を最大にするには、キャベツとキュウリを各々、何ha作れば良いか
また、その時、最大収益は何万円になるか？

土地制約条件	$X_1 + X_2 \leqq 50$
労働制約条件	$X_1 + 4X_2 \leqq 100$
資金制約条件	$50X_1 + 100X_2 \leqq 3000$
純収益	$80X_1 + 100X_2 = Z$

資料）農林統計情報部「地域統計分析の理論と実際」農林統計協会

図形による解法



資料）農林統計情報部「地域統計分析の理論と実際」農林統計協会

ｼﾝﾌﾟﾚｯｸｽ表による解法

＜第一段階＞

(i)変数名x_1, x_2, x_3, x_4を表頭に書き入れ、その下に制限式の係数を書き入れる。制限量は左方S欄に書き入れ、その左の基底欄にy_1, y_2, y_3と書く。 基底欄の意味は、第一段階の計画では、遊ばせておく土地y_1＝50、遊ばせておく労働y_2=100、遊ばせておく資本y_3=3000とする計画を示している。

(ii)表の左辺と上辺に各変数の収益係数を書く。遊ばせておく変数y_1, y_2, y_3 の収益係数は各々0である。

(iii)Z_j-C_jの行には各変数の収益係数を負にして書き入れる。なおS列には0を入れる。

＜第１段階から第２段階へ移るための計算＞

(i) Z_j-C_j（ｼﾝﾌﾟﾚｯｸｽ基準）の値のうち負で絶対値が最大の数に注目し、その上の3欄を太線で囲む。この変数は第2段階で計画に採用するものである。

(ii)S欄の値を太線で囲んだ数値で割り、右辺のR欄に書き入れる。それらのうちで正で最小な数に注目し、その行の左の各欄を太線で囲む。この変数は、第2段階で外すべきものである。

(iii)第一段階の表で太線で囲まれた行の各値を太線列との交点の値で割り、第2段階の表の対応する行に書き入れる。

(iv)第１段階の表で太線で囲まれた行以外の各行については、各欄の値からその行の太線で囲まれた値を(iii)で計算した第２段階の表に書き入れた値にかけた値を引いて、第2段階の対応する欄に書き入れる。

(v)第２段階の表のZ_j-C_j（ｼﾝﾌﾟﾚｯｸｽ基準）は、次のように計算される。Zjは左端のCjと各列の対応する欄の数値を掛け算して足したものである。Z_jから表の最上段のC_jを引いて第2段階のZ_j-C_j 行の各欄に記入する。

＜第２段階から第3段階に移るための計算＞

第２段階のZ_j-C_j（ｼﾝﾌﾟﾚｯｸｽ基準）の値のうち負で絶対値が最大の数に注目し、最右欄のRを第１段階と同様に計算して正で最小な数に注目し、これらの列と行を太線で囲んで以上の方法を繰り返す。

＜最適解の判定＞

Z_j-C_j（ｼﾝﾌﾟﾚｯｸｽ基準）に負数がなくなれば、それはもはや改善の余地が無いことを表すので、その段階が最適解である。

この段階の表の基底欄とS欄は以下のことを示している。

y2=20 （労働は20人余る）、 x1=40（キャベツを40ha作る）、 x2=10（キュウリを10ha作る）、 Z_j-C_j=4200（全純収益は4200万円である）

資料）農林統計情報部「地域統計分析の理論と実際」農林統計協会

シンプレクス法による解法

段階	Cj		80	100	0	0	0	R
		基底 S	x1	x2	y1	y2	y3	
1	0	y1 50	1	1	1	0	0	50
	0	y2 100	1	4	0	1	0	25
	0	y3 3000	50	100	0	0	1	30
		Zj-Cj 0	-80	-100	0	0	0	
2	0	y1 25	0.75	0	1	-0.25	0	33
	100	x2 25	0.25	1	0	0.25	0	100
	0	y3 500	25	0	0	-25	1	20
		Zj-Cj 2500	-55	0	0	25	0	
3	0	y1 10	0	0	0	0.5	-0.03	20
	100	x2 20	0	1	0	0.5	-0.01	40
	80	x1 20	1	0	0	-1	0.04	-20
		Zj-Cj 3600	0	0	0	-30	2.2	
4	0	y2 20	0	0	0	1	-0.06	
	100	x2 10	0	1	0	0	0.02	
	80	x1 40	1	0	0	0	-0.02	
		Zj-Cj 4200	0	0	0	0	0.4	

線形計画法　例題

			生産プロセス	
			白菜	大根
利益係数(千円)			120	90
制約要素	土地	10(10a)	1	1
	9月労働	360(時間)	20	40
	11月労働	360(時間)	45	22.5

土地制約　　$X1 + X2 \leqq 10$

11月労働制約　　$20\,X1 + 40\,X2 \leqq 360$

９月労働制約　　$45\,X1 + 22.5\,X2 \leqq 360$

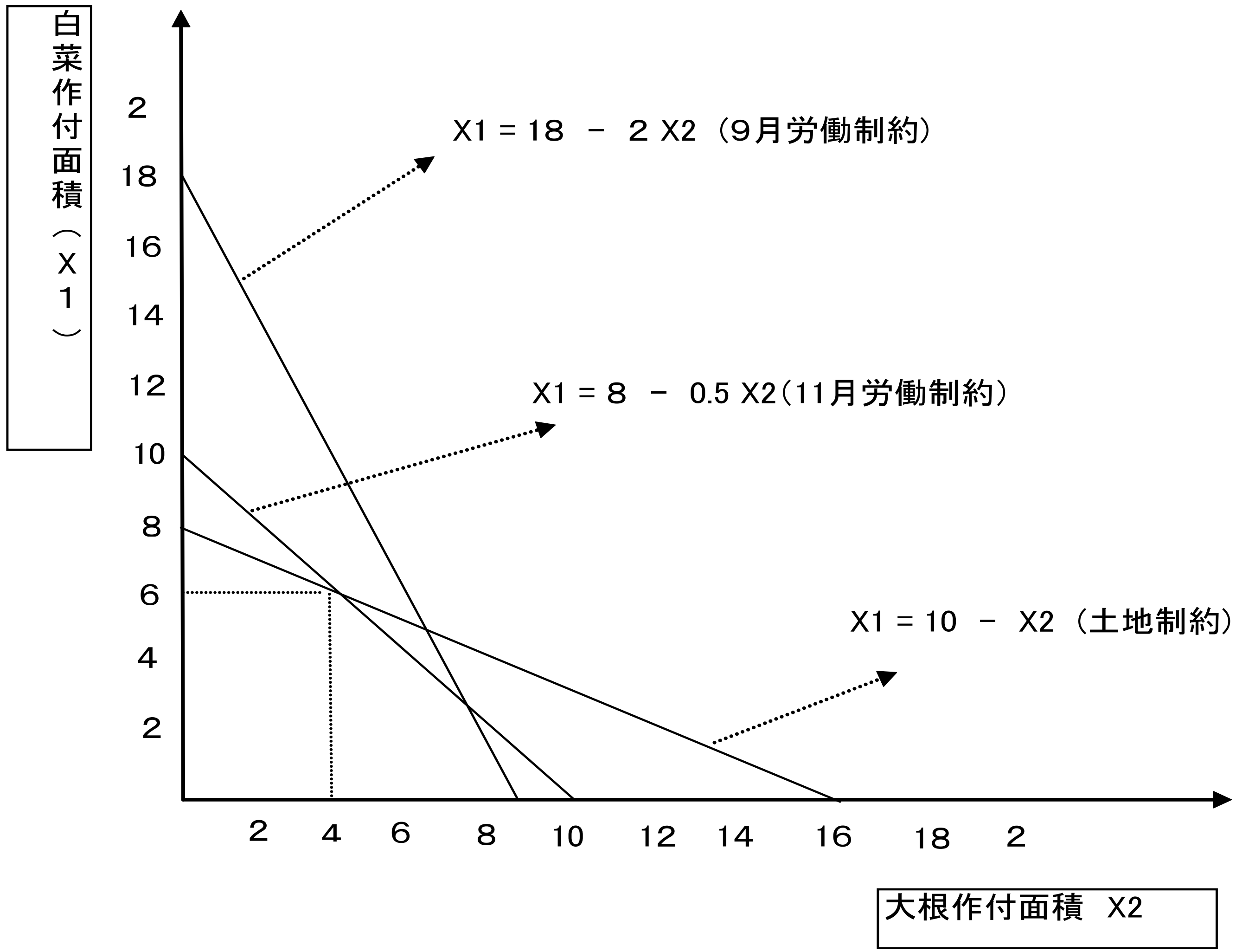
の下で　目的関数

　　利潤関数　　$Z = 12\,X1 + 9\,X2$

を最大化する問題

資料）農林水産省農業研究センター「地域農業の計画手法」農林統計協会

図形による解法



白菜作付面積（X１）
２
18
16
14
12
10
８
６
４
２
X1 = 18 － ２ X2（９月労働制約）
X1 = ８ － 0.5 X2（11月労働制約）
X1 = 10 － X2（土地制約）
２ ４ ６ ８ 10 12 14 16 18 ２
大根作付面積 X2

シンプレクス法による解法

段階	利益係数Cj				120	90	0	0	0	R
	基底	制約資源	制約量	関係	白菜(X1)	大根(X2)	土地(y1)	9月労働(y2)	11月労働(y3)	
1	0	土地(y1)	10	≧	1	1	1	0	0	10
	0	9月労働(y2)	360	≧	20	40	0	1	0	18
	0	11月労働(y3)	360	≧	45	22.5	0	0	1	8
		Z-C(潜在価格)	0		-120	-90	0	0	0	
2	0	土地(y1)	2		0	0.5	1	0	-0.02	4
	0	9月労働(y2)	200		0	30	0	1	-0.4	6.67
	120	白菜(X1)	8		1	0.5	0	0	0.02	16
		Z-C(潜在価格)	960		0	-30	0	0	2.4	
3	90	大根(X2)	4		0	1	2	0	-0.04	
	0	9月労働(y2)	80		0	0	-60	1	0.8	
	120	白菜(X1)	6		1	0	-1	0	0.04	
		Z-C(潜在価格)	1080		0	0	60	0	1.2	