

データリテラシー

第8回 推定

メディア情報コース
平居 悠（ひらい ゆたか）

到達目標

- データサイエンス・AIが社会でどのように活用されているのかを理解する。
- データを適切に読み解き理解する力をつける。
- データを適切な可視化手法で他者に説明できるようになる。
- 表計算ソフトを使ってデータを集計・加工できるようになる。

前回

第 1 回	社会におけるデータ・AI利活用
第 2 回	表計算ソフトの基本操作
第 3 回	セルの参照と集計
第 4 回	度数分布表とヒストグラム
第 5 回	代表値と散らばりの尺度
第 6 回	データのグラフによる表現とデータ比較
第 7 回	データの分布と標本抽出
第 8 回	推定
第 9 回	分割表とクロス集計
第 1 0 回	2 変量データと相関
第 1 1 回	時系列データ
第 1 2 回	仮説検定
第 1 3 回	回帰分析

前回の目標

正規分布のグラフを描いてデータの分布を分析できるようになる。

前回学んだこと

1. 確率と期待値
2. 正規分布
3. 母集団と標本抽出

推測統計

ある集団全体を調べることが難しい場合、その集団から一部分だけを選び出して調べることで、集団について推測を行うこと。

事象と確率

今考えている全ての事象に同程度に起こりやすい場合の数が n 通りあるとする。その中で、ある事象 A に含まれる場合の数が k 通りあるとき、 A の起こる確率 $P(A)$ は、

$$P(A) = \frac{k}{n}$$

確率変数

大文字で表す変数 X は、単なる変数ではなく、**確率を伴う変数**という意味で**確率変数**という。

確率変数

確率変数でとびとびの値しかとらないものを**離散型確率変数**、連続的な値をとるものを**連続型確率変数**という。

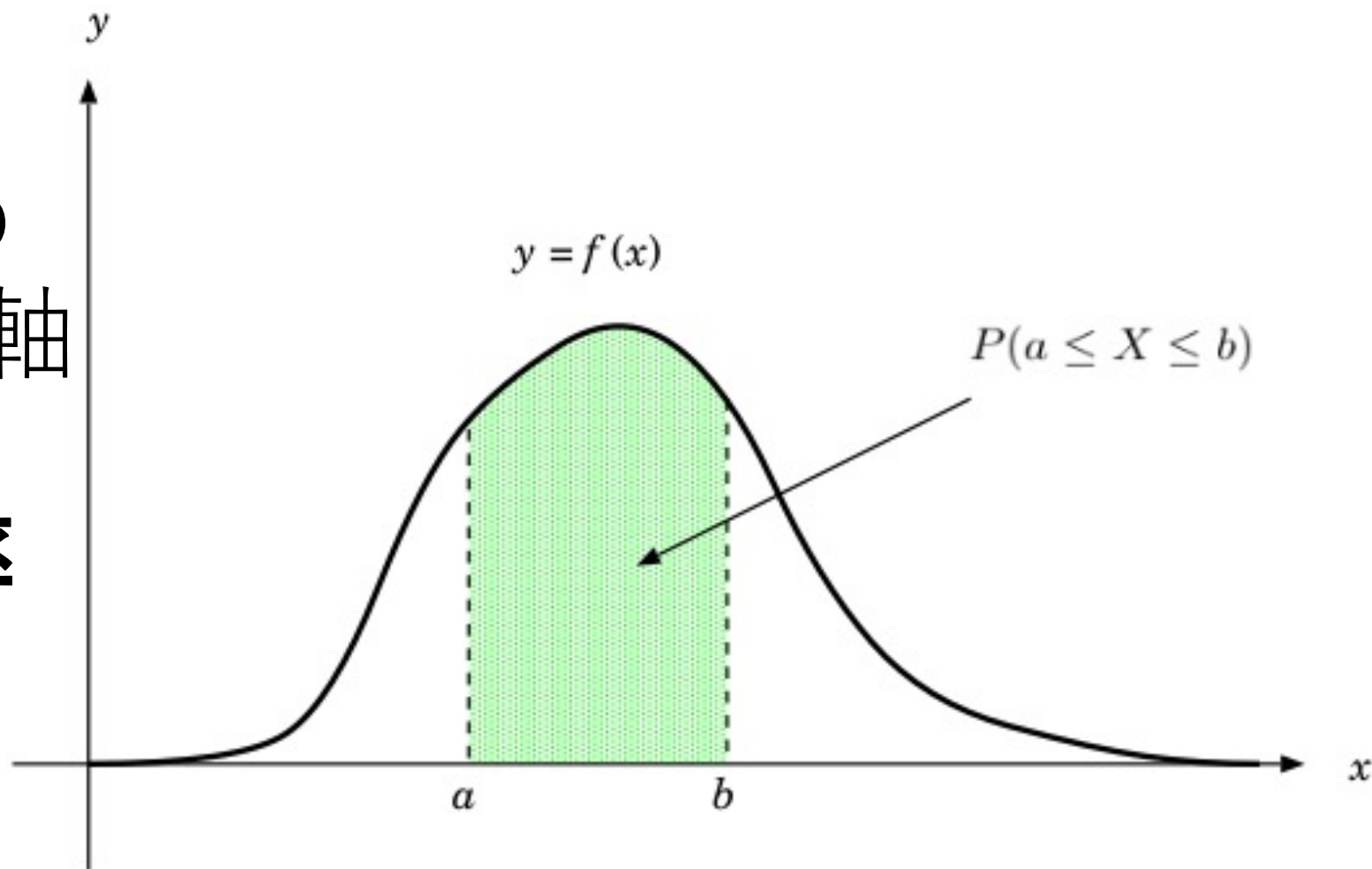
確率分布

確率変数 X は観測値 x によって様々な確率をとる。この確率の分布を**確率分布**といい、確率分布を関数として表したものを**確率密度関数** $f(x)$ という。

$$f(x) = P(X = x)$$

確率分布

確率密度関数の
グラフでは、 x 軸
と $f(x)$ ではさま
れた **面積が確率**
になる。



確率密度関数の性質

確率密度関数 $f(x)$ は次の性質を持たなくてはならない。

- $0 \leq f(x) \leq 1$ (負の確率や1より大きい確率はない)
- x 軸と $f(x)$ で挟まれた全面積は1 (全確率は1)

確率分布の代表値と散らばりの尺度

確率分布の
平均は μ （ミュー）、
分散は σ^2 （シグマ二乗）、
標準偏差は σ （シグマ）
という記号で表されることが多い。

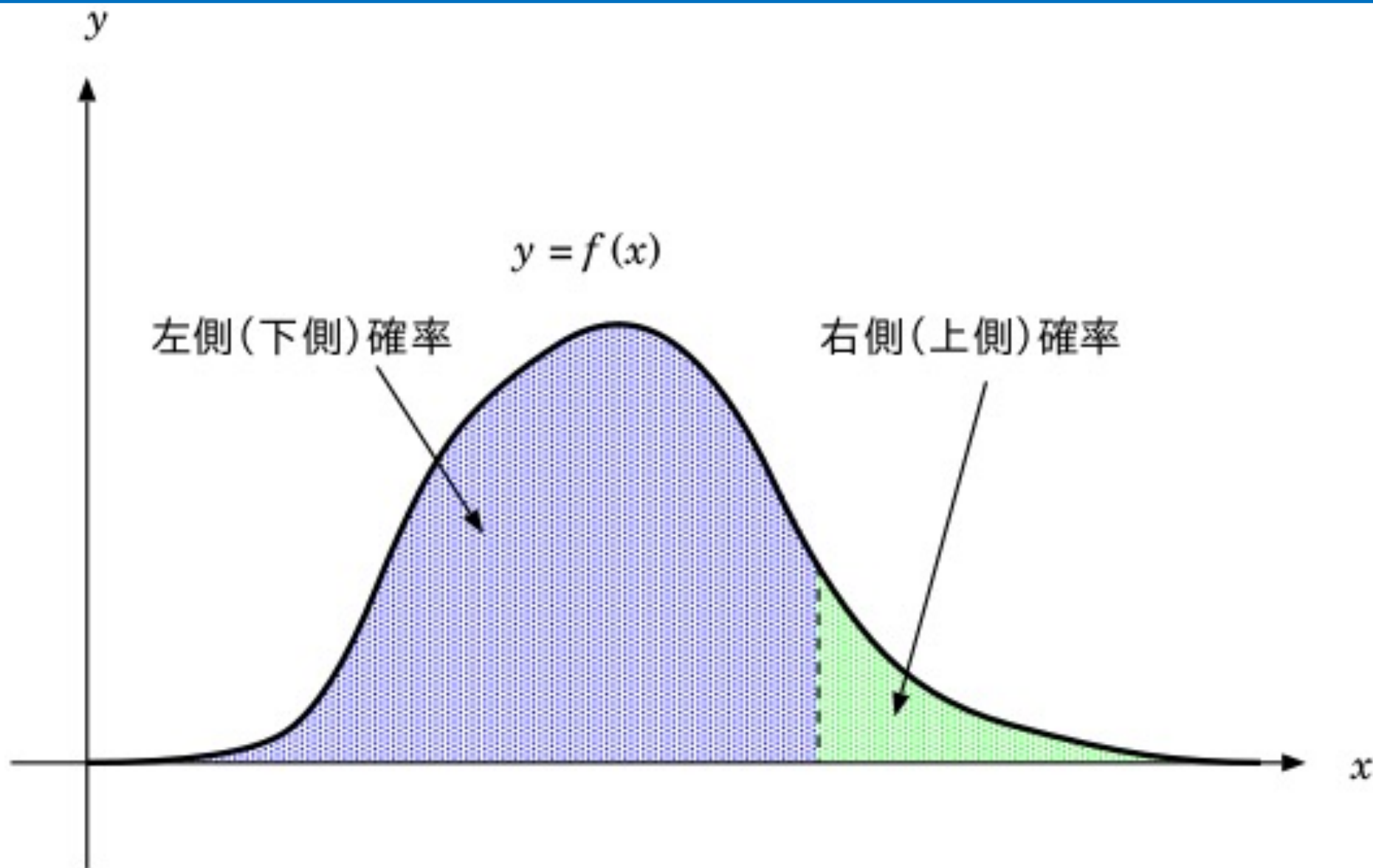
期待値とは

確率変数の平均値

より正確には、

確率変数 X に対して、それがとる値
の重みつき平均

右側確率と左側確率



パーセント点

右側（上側）確率が α （アルファ）となる点を**右側（上側）確率 α のパーセント点**という。 α は確率なので、 $0 \leq \alpha \leq 1$ である。

正規分布の定義

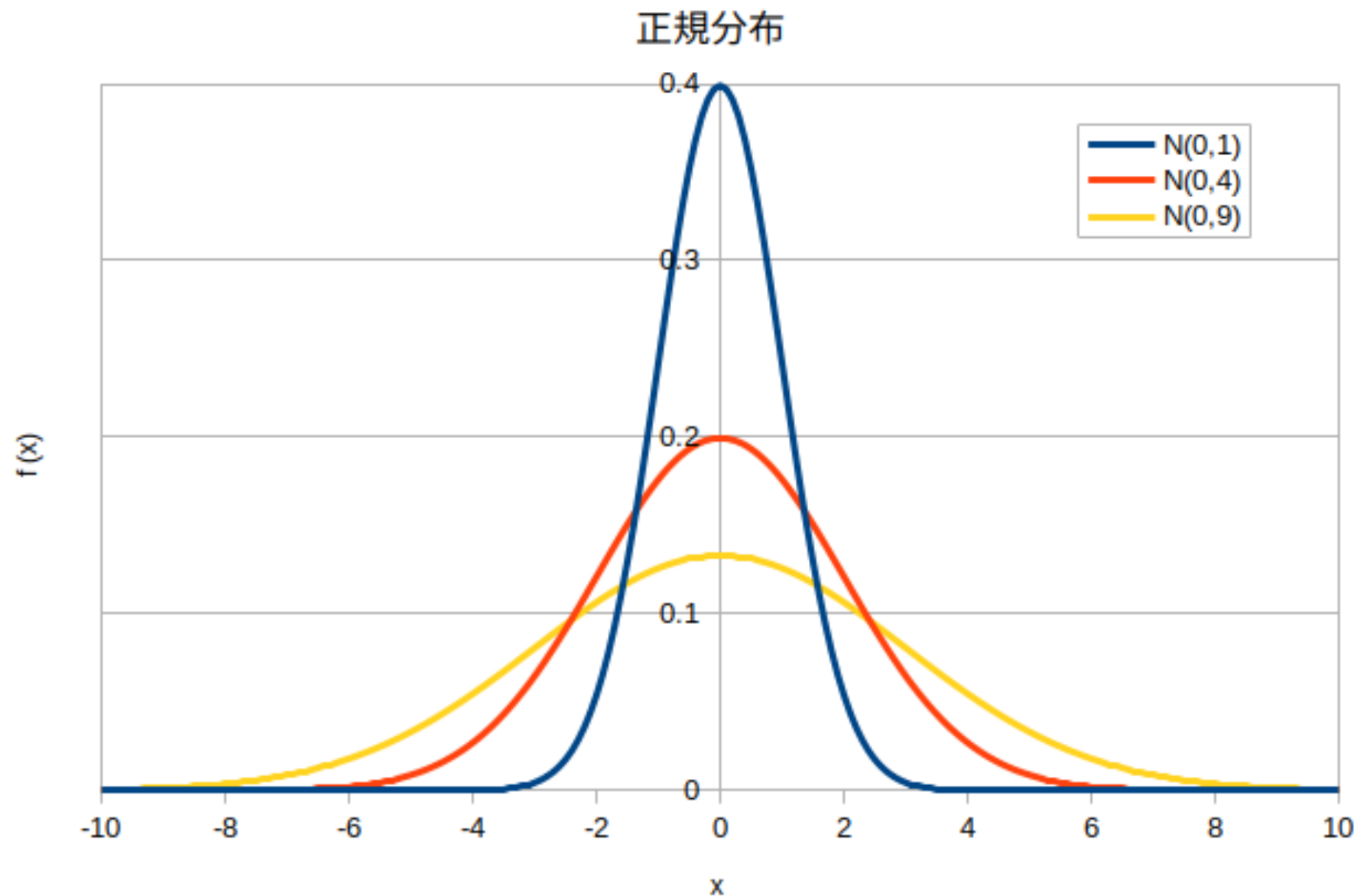
正規分布の確率密度関数は以下で定義される。

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

ここで $e=2.71828\dots$ は**ネイピア数**といい、無限に続く無理数。 μ は平均、 σ は標準偏差。

これを全範囲で積分すると1になる。証明：<https://manabitimes.jp/math/754>

正規分布のグラフ



正規分布のグラフの特徴

- 釣鐘のような形をしている。
- $x = \mu$ を中心として左右対称の形をしている。
- σ が大きいほど頂点が低くなり裾野が広がる。

正規分布のグラフの特徴

分布の形は μ 、 σ という二つのパラメータで決まるため、正規分布のことを記号 $N(\mu, \sigma^2)$ で表す。

特に $\mu = 0$ 、 $\sigma^2 = 1$ の正規分布 $N(0, 1)$ を**標準正規分布**という。

母集団

調査したい集団全体のこと。母集団は要素の集まり。

標本調査

取り出した要素を**標本（サンプル）**といい、要素を取り出すことを**標本抽出（サンプリング）**という。

無作為標本抽出

各要素を全て等しい確率で無作為に取り出すこと。

偏りがない標本抽出によって標本は母集団の良い縮図となり、確率的に母集団を推測することが可能となる。

前回の問い

- 確率密度関数 $f(x)$ のグラフの意味は何か？
 - x 軸と $f(x)$ で挟まれた面積が確率になる。
- 正規分布とはどのような分布か？
 - 平均値を中心に釣鐘のような左右対称な形をした分布
- 母集団と標本抽出とは何か？
 - 調査したい集団全体のことを母集団といい、母集団から要素を取り出すことを標本抽出という。

タイピング練習スケジュール

第1回	ホームポジション
第2回	ホームポジション
第3回	ローマ字
第4回	英語初級
第5回	日本国憲法（trr試験）
第6回	ホームポジション
第7回	ローマ字
第8回	英語初級
第9回	日本国憲法（trr試験）
第10回	ホームポジション
第11回	ローマ字
第12回	英語初級
第13回	日本国憲法（trr試験）

タイピングの練習 (英語初級)

1. ブラウザを起動し、<https://www.koeki-prj.org/trr/>に繋ぐ。
2. 学籍番号（Cは大文字、省略なし8桁）を入力する。
3. Koeki MAILに届いたパスコードをPasscode: 欄に入力する。

ホームポジション

左手でタイプするキー

右手でタイプするキー



左手の人差指から小指までの
ホームポジション

両手の親指の
ホームポジション

右手の人差指から小指までの
ホームポジション

今回

第1回	社会におけるデータ・AI利活用
第2回	表計算ソフトの基本操作
第3回	セルの参照と集計
第4回	度数分布表とヒストグラム
第5回	代表値と散らばりの尺度
第6回	データのグラフによる表現とデータ比較
第7回	データの分布と標本抽出
第8回	推定
第9回	分割表とクロス集計
第10回	2変量データと相関
第11回	時系列データ
第12回	仮説検定
第13回	回帰分析

今回の目標

表計算ソフトを用いて母平均を推定できるように
なる。

今回学ぶこと

1. 母数と統計量

2. 母平均の推定

今回の問い

- 母数とは何か？
- 母数の推定方法には何があるか？

今回学ぶこと

1. 母数と統計量

2. 母平均の推定

母数（パラメーター）

母集団分布の特徴を表す
代表値や散らばりの尺度

よく使われる母数

- **母平均**（母集団の平均） μ
- **母分散**（母集団の分散） σ^2
- **母標準偏差**（母集団の標準偏差）

$$\sigma = \sqrt{\sigma^2}$$

標本から推定できるもの

- 標本は母集団の一部でしかないので、標本だけで母集団分布の全てを知ることにはできない。
- 標本から推定できるのは母集団の**母数**。

統計量

母数の推定に使われる標本の指標

標本平均 $\bar{X}$

母平均 μ の推定に使われる統計量。
母集団から n 個の標本 $X_1, X_2, \dots, X_n$
を無作為抽出したとすると、

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

不偏分散 s^2

母分散 σ^2 の推定に使われる統計量。

$$s^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n - 1} = \frac{1}{n - 1} \sum_{i=1}^n (X_i - \bar{X})^2$$

標本平均と不偏分散の期待値

標本平均 $\bar{X}$ の期待値は母平均 μ 、
不偏分散 s^2 の期待値は母分散 σ^2
になる。

$$E(\bar{X}) = \mu$$

$$E(s^2) = \sigma^2$$

標本平均の期待値

$E(X_i) = \mu$ とする。

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n E(X_i) \\ &= \frac{1}{n} \sum_{i=1}^n \mu \\ &= \mu \end{aligned}$$

不偏分散の期待値

$E(X_i) = \mu, E((X_i - \mu)^2) = \sigma^2, E((X_i - \mu)(X_j - \mu)) = 0 \quad (i \neq j)$ とする。

$$\begin{aligned} E(s^2) &= E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\ &= \frac{1}{n-1} \sum_{i=1}^n E((X_i - \bar{X})^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n E((Y_i - \bar{Y})^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n \{E(Y_i^2) + E(\bar{Y}^2) - 2E(Y_i \bar{Y})\} \\ &= \frac{1}{n-1} \sum_{i=1}^n \left(\sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n}\right) \\ &= \frac{1}{n-1} \sum_{i=1}^n \frac{(n-1)\sigma^2}{n} \\ &= \sigma^2 \end{aligned}$$

ここで $Y_i = X_i - \mu$ とすると、

$$\begin{aligned} E(Y_i) &= 0 \\ E(Y_i^2) &= \sigma^2 \\ E(Y_i Y_j) &= 0 \quad (i \neq j) \\ E(Y_i \bar{Y}) &= \frac{1}{n} \sum_{j=1}^n E(Y_i Y_j) = \frac{\sigma^2}{n} \\ E(\bar{Y}^2) &= \frac{1}{n} \sum_{i=1}^n E(Y_i \bar{Y}) = \frac{\sigma^2}{n} \end{aligned}$$

演習：標本抽出

500人の身長データを母集団とみなす。
この母集団から20人の標本 ($n = 20$) を無作為抽出する。

RANDBETWEEN関数

関数	説明
RANDBETWEEN(最小値,最大値)	最小値以上、最大値以下の範囲の整数の乱数を発生する。

シートで何か作業をしたり、再計算（F9を押す）たびに乱数発生が行われる。

VLOOKUP関数

関数	説明
VLOOKUP(検索値,範囲,列番号,検索の型)	範囲の1列目から検索値を検索し、マッチした行の列番号の列から値を取り出す。 検索の型が0なら完全に一致するもの、1なら近似的に一致するものを検索する。

演習：標本抽出手順

1. dataset1.odsのシートを開く。
2. 新たにシートを追加し、セルA1からE501をコピーして貼り付ける。
3. 新しく追加したセルG1に「標本番号」、H1に「身長[cm]」と入力する。
4. セルG2に「**=RANDBETWEEN(1, 500)**」と入力する。
5. セルH2に「**=VLOOKUP(G2, A\$2:E\$501, 4, 0)**」と入力する。
ここでセルA2からE501の範囲で1列目からG2の番号を検索し、マッチした行の4列目から値を取り出している。

演習：標本抽出手順

6. セルG2とH2を3行目から21行目までコピー&ペーストする。
7. セルG22に「標本平均」と入力する。
8. セルH22に「**=AVERAGE(H2:H21)**」と入力する。
9. セルG23に「不偏標準偏差」と入力する。
10. セルH23に「**=STDEV.S(H2:H21)**」と入力する。

これで列Hに20人の身長データが無作為標本抽出され、セルH22に標本平均 $\bar{x}$ 、セルH23に不偏標準偏差 s が求められる。

今回の問い

- 母数とは何か？
 - 母集団分布の特徴を表す代表値や散らばりの尺度
- 母数の推定方法には何かあるか？

今回学ぶこと

1. 母数と統計量

2. 母平均の推定

母平均の推定方法

- 点推定
- 区間推定

点推定

単純に標本平均 $\bar{X}$ の値を母平均 μ の推定値とすること。

点推定は**大数の法則**を根拠としている。

大数の法則

標本平均 $\bar{X}$ は n が大きくなるにつれ、母平均 μ に近づく。

点推定：サイコロの例

AさんとBさんがそれぞれサイコロ1個を4回転がした結果、以下の通りだったとする。

実験者	1回目 X_1	2回目 X_2	3回目 X_3	4回目 X_4	標本平均 $\bar{X}$
Aさん	5	1	3	1	2.5
Bさん	6	6	3	5	4.0

点推定：サイコロの例

これより、点推定を行うと、

Aさんによれば、**推定値 = 2.5**

Bさんによれば、**推定値 = 4.0**

となる。

点推定：サイコロの例

同じ条件で同じ回数だけ標本抽出を行っても、AさんとBさんで得られる標本は異なったので、それによって推定値も異なっている。

点推定：サイコロの例

- サイコロ1個を転がして出る目の平均は理論的には $\mu = 3.5$
- 上の実験では、AさんもBさんも標本平均は3.5からズレた値
→ サイコロを転がす回数が少ないと母平均からズレやすい
- 転がす回数 n を増やせば増やすほど $\mu = 3.5$ に近い $\bar{X}$ が得やすくなる（**大数の法則**）。

点推定：サイコロの例

点推定は標本から $\bar{X}$ を求めるだけなのでお手軽だが、 $\bar{X}$ の値がどのくらい μ の真の値に近いのかは分からない。

区間推定

ある確率 $1 - \alpha$ で μ が含まれる
区間を推定すること。

区間推定

具体的には

$$P(L \leq \mu \leq U) = 1 - \alpha$$

を満たす区間[L, U]を求めること。
[L, U]を $1 - \alpha$ の信頼区間という。

区間推定

確率 $1 - \alpha$ は信頼係数といい、好きなように設定して良い。信頼係数には、0.9 (90%), 0.95 (95%), 0.99 (99%) がよく使われる。

区間推定

信頼区間は標本から確率的に推定するので、標本の統計量の確率分布（**標本分布**）が必要になる。

区間推定

- 標本分布を知るにはその元となる母集団分布を知らなくてはならない。
- しかし、そもそも母集団を知らないから標本を使うので、あらかじめ母集団分布を知ることはいできない。

区間推定

母集団分布を知らなくても n が大きければ **中心極限定理** によって $\bar{X}$ の標本分布は近似的に正規分布になる。

中心極限定理

母平均 μ 、母分散 σ^2 の母集団があるとする。
この母集団がどのような分布でも、 $\bar{X}$ の標本
分布は近似的に $N(\mu, \sigma^2/n)$ になる。

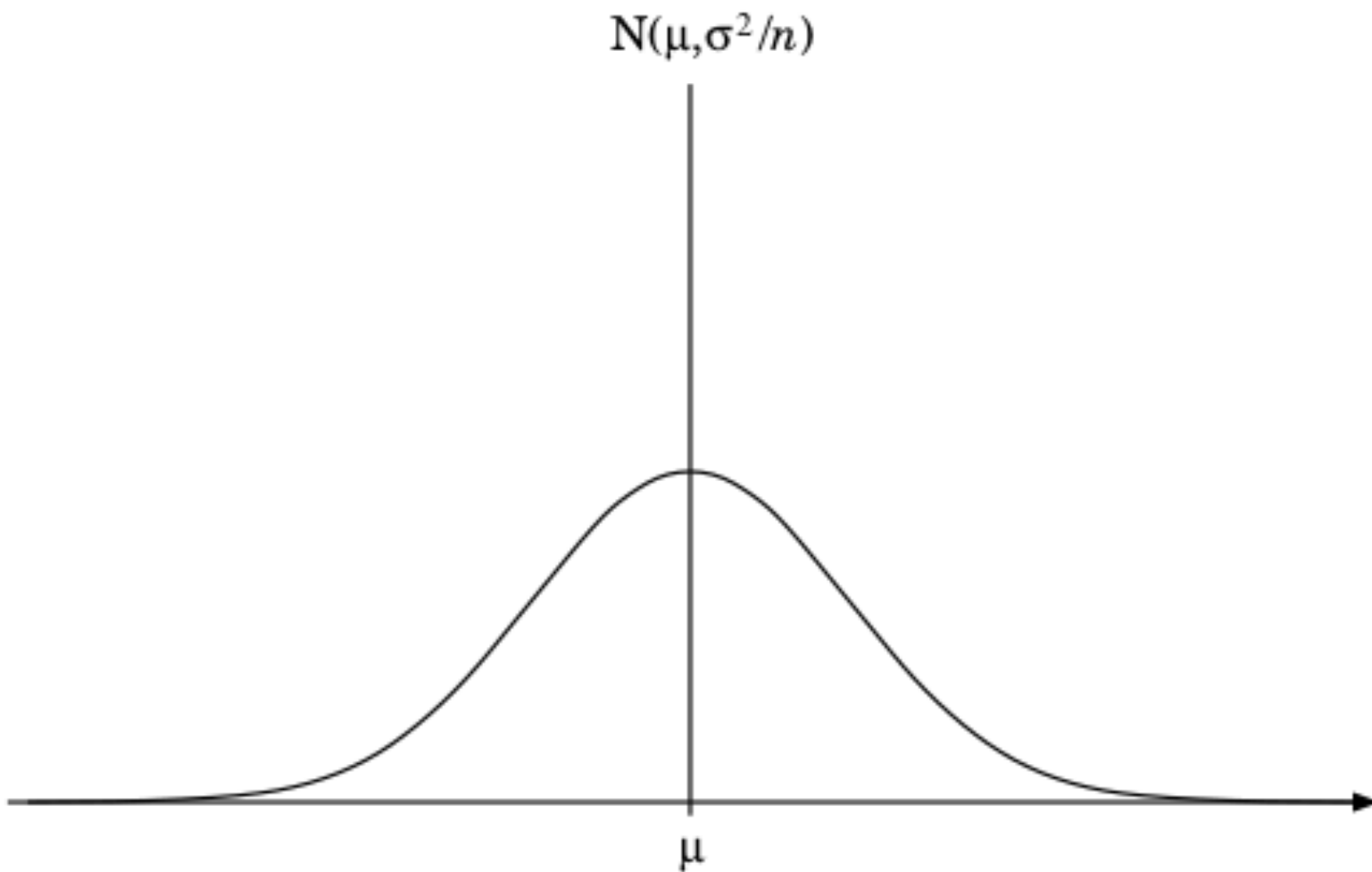
母集団が正規分布の場合は n に関わらず厳密に
いえるが、それ以外の分布の場合は n が大
きいほど近似は良くなる。

中心極限定理

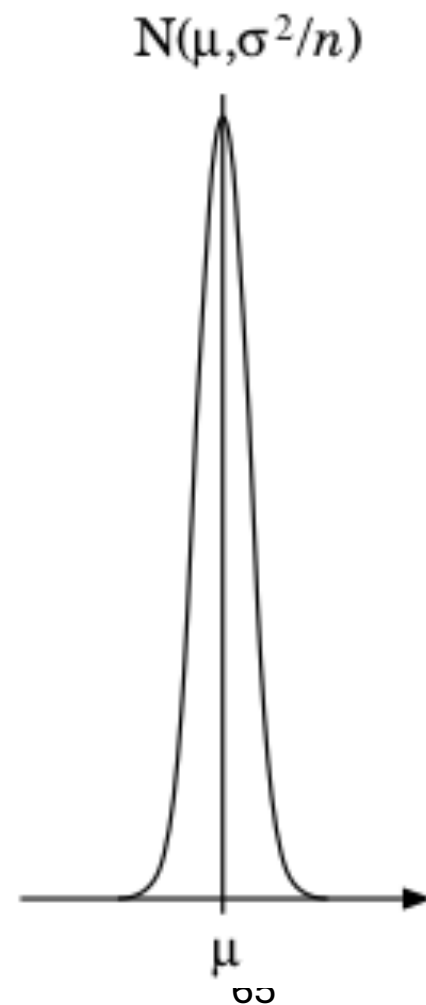
中心極限定理では、 $\bar{X}$ の標本分布は近似的に $N(\mu, \sigma^2/n)$ になる。

分散が σ^2 ではなく、 σ^2/n のため、 n が大きくなるにつれて分散が小さくなり、 $\bar{X} = \mu$ となる確率が高くなる。これは大数の法則で述べていることでもある。

中心極限定理



n が大きくなると



母分散が既知の場合

母分散 σ^2 があらかじめ分かっている（既知）とする。

中心極限定理から、 n が大きければ $\bar{X}$ は近似的に $N(\mu, \sigma^2/n)$ に従う。

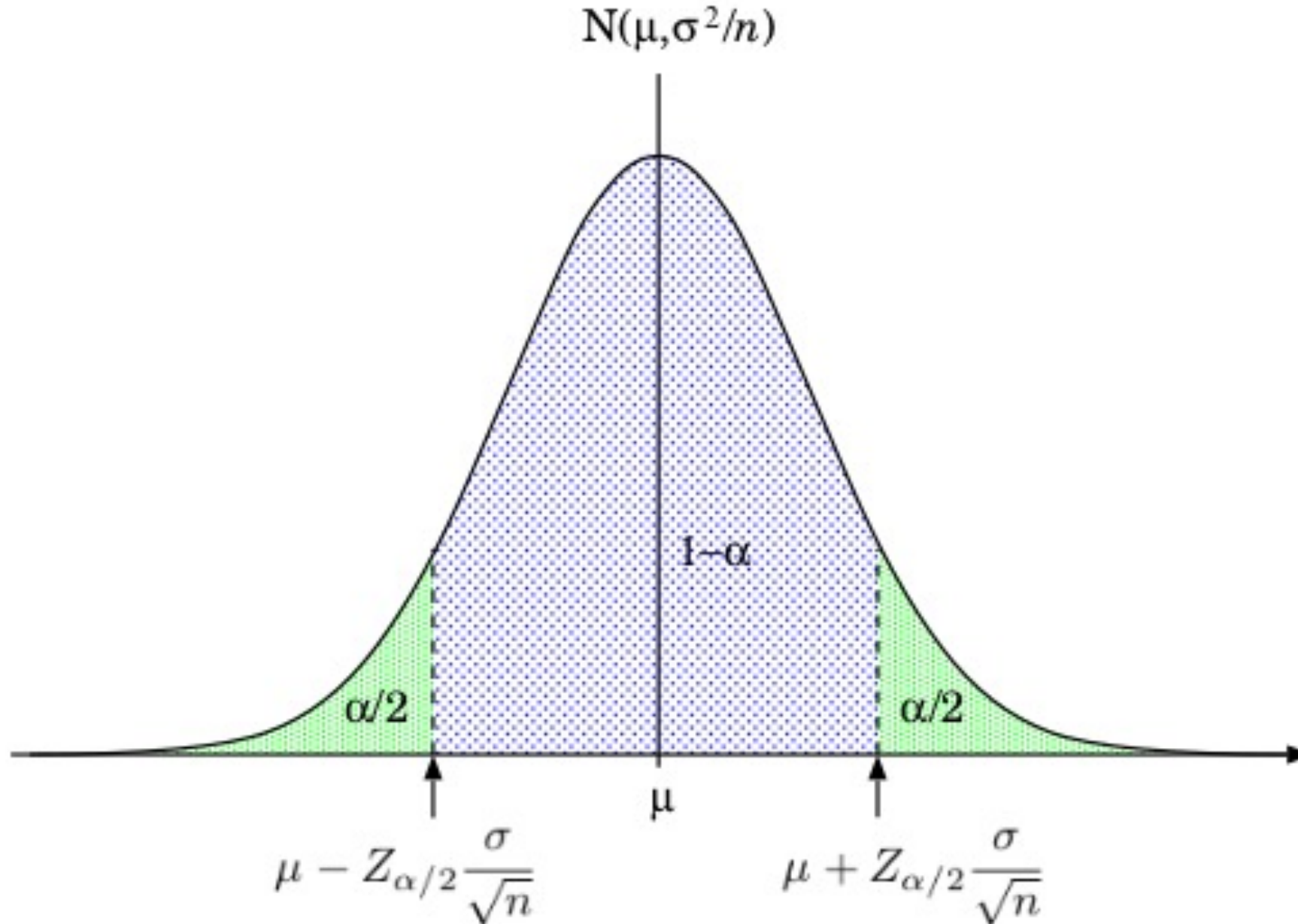
母分散が既知の場合

$N(\mu, \sigma^2/n)$ の中心に $1 - \alpha$ の確率をとり、その両端の確率を $\alpha/2$ ずつに分けると確率の式は以下のようなになる。

$$P\left(\mu - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

ここで $Z_{\alpha/2}$ は $N(0, 1)$ における右側（上側）確率 $\alpha/2$ のパーセント点。

母分散が既知の場合



母分散が既知の場合

括弧の中の不等式を変形して μ を挟む式にすると、

$$P\left(\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

となる。

母分散が既知の場合

したがって信頼区間は、

$$[L, U] = \left[\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

で求められる。

区間推定の導出

$$P\left(\mu - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

を導出する。

区間推定の導出

$N(\mu, \sigma^2/n)$ に従う $\bar{X}$ を標準化すると

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

は $N(0,1)$ に従う。

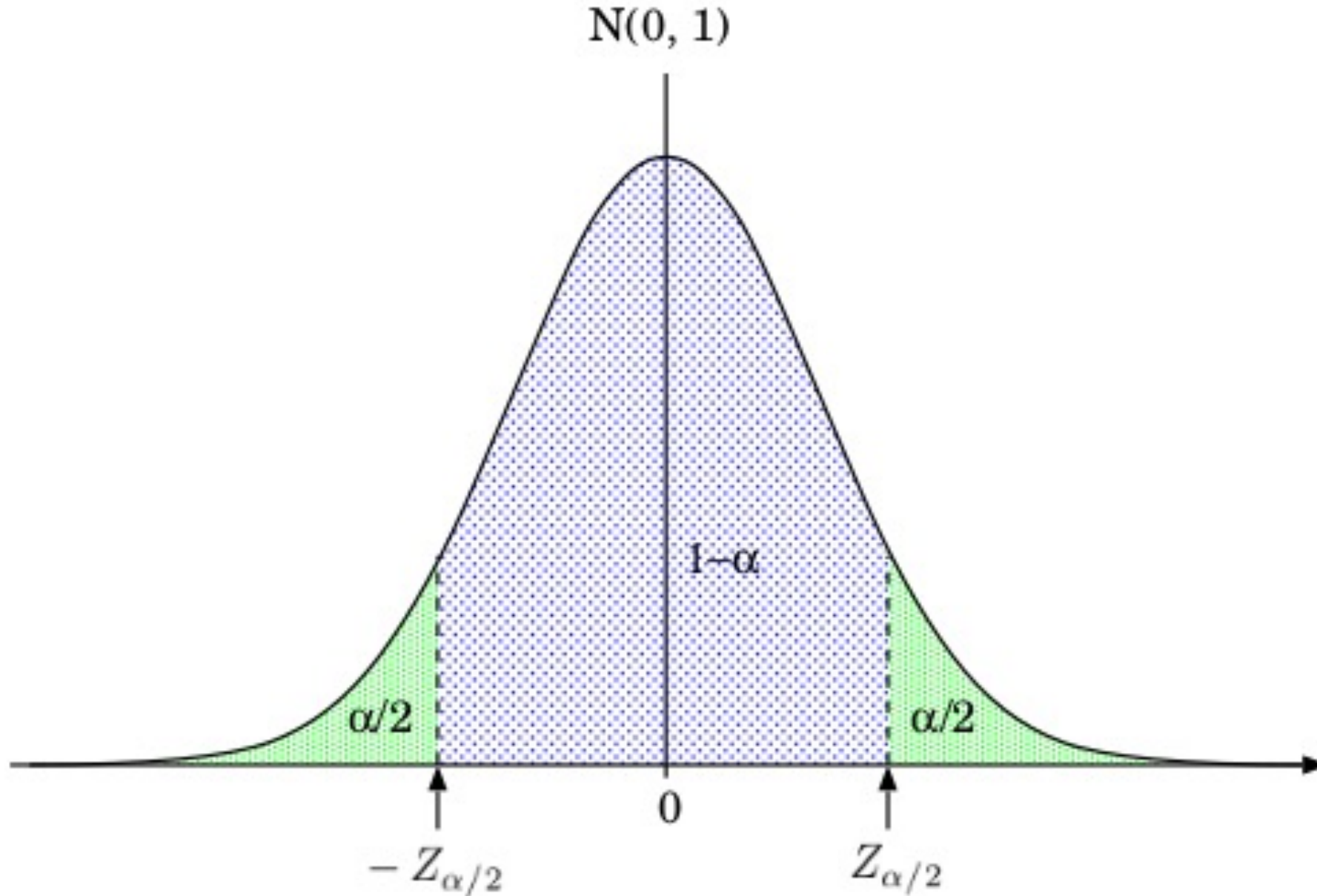
区間推定の導出

$Z_{\alpha/2}$ は確率と

$$P(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) = 1 - \alpha$$

という関係にある。

区間推定の導出



区間推定の導出

$$P(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) = 1 - \alpha$$

この括弧内の不等式は標準化の式を使うと

$$-Z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq Z_{\alpha/2}$$

となり、これを変形すると、

$$\mu - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

区間推定の導出

したがって、

$$P\left(\mu - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

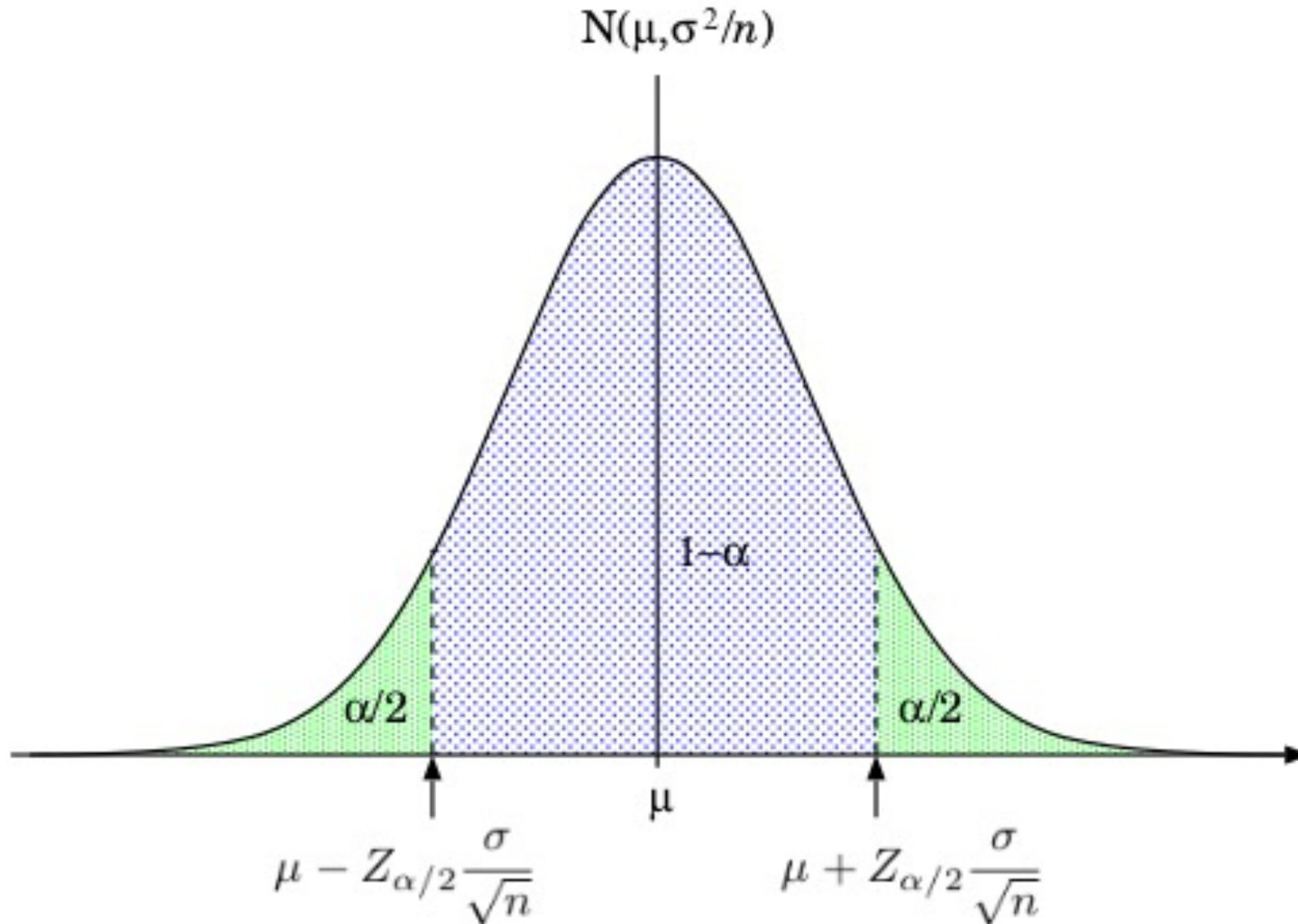
が成り立つ。

区間推定の導出

$$P\left(\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

を導出する。

区間推定の導出



区間推定の導出

この括弧内の不等式は二つの不等式に分解できる。

$$\mu - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{X}$$

$$\bar{X} \leq \mu + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

これをそれぞれ移項すると、

$$\mu \leq \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu$$

区間推定の導出

以上より μ を挟む形の不等式にできる。

$$\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

したがって、

$$P \left(\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

が成り立つ。

母分散が未知の場合

母分散 σ^2 が分かっていない場合は、代わりに標本から求められる不偏分散 s^2 を使う。

母分散が未知の場合

n が大きい場合は、大数の法則により σ^2 は不偏分散 s^2 とほぼ等しいとみなせるため、 $\sigma^2 \simeq s^2$ ($\sigma \simeq s$)と近似して母集団が既知の場合の信頼区間を使うことができる。

$$[L, U] = \left[\bar{X} - Z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + Z_{\alpha/2} \frac{s}{\sqrt{n}} \right]$$

ここで $s = \sqrt{s^2}$ は不偏標準偏差。

区間推定：サイコロの例

サイコロ1個を4回転がして得られた標本から母平均 μ の信頼係数90% ($1 - \alpha = 0.9$)の信頼区間の推定を行ってみよう。

簡単のために母分散 $\sigma^2 = \frac{35}{12}$ (母標準偏差 $\sigma = \sqrt{\frac{35}{12}}$)として、母分散は既知とする。

区間推定：サイコロの例

区間推定には標本が必要である。そこで、AさんとBさんがそれぞれサイコロ1個を4回転がした結果、以下の通りだったとする（無作為標本抽出）。

実験者	1回目 X_1	2回目 X_2	3回目 X_3	4回目 X_4	標本平均 $\bar{X}$
Aさん	5	1	3	1	2.5
Bさん	6	6	3	5	4.0

区間推定：サイコロの例

信頼係数90%の信頼区間を求める。 $1 - \alpha = 0.90$ より、 $\alpha/2 = 0.05$ なので、

$$Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \simeq 1.64 \times \frac{\sqrt{\frac{35}{12}}}{\sqrt{4}} \simeq 1.4$$


であるから、

標準正規分布表 : https://unit.aist.go.jp/mcml/rg-orgp/uncertainty_lecture/normsdist.html

区間推定：サイコロの例

信頼係数90%の信頼区間は、

Aさんの場合：

$$[L, U] = [2.5 - 1.4, 2.5 + 1.4] = [1.1, 3.9]$$

Bさんの場合：

$$[L, U] = [4.0 - 1.4, 4.0 + 1.4] = [2.6, 5.4]$$

区間推定：サイコロの例

LibreOffice Calcでは「**=NORM.INV(1-0.05, 0, 1)*SQRT(35/12)/SQRT(4)**」で計算できる。

数式	LibreOffice Calc
$Z_{0.05}$	<code>NORM.INV(1-0.05,0,1)</code>
$\frac{\sqrt{\frac{35}{12}}}{\sqrt{4}}$	<code>SQRT(35/12)/SQRT(4)</code>

演習：推定

500人の身長から無作為抽出した標本から信頼係数90% ($1 - \alpha = 0.9$)の信頼区間を求める。母分散は未知とする。

演習：推定

n が大きいので s を σ とみなして区間推定を行う。 $1 - \alpha = 0.9$ より $\alpha/2 = 0.05$ である。

$$[L, U] = \left[\bar{X} - Z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + Z_{\alpha/2} \frac{s}{\sqrt{n}} \right] = \left[\bar{X} - Z_{0.05} \frac{s}{\sqrt{20}}, \bar{X} + Z_{0.05} \frac{s}{\sqrt{20}} \right]$$

演習：推定手順

1. セルG25に「信頼区間」、G26に「L」、G27に「U」と入力する。
2. セルH26に「**=H22-NORM.INV(1-0.05, 0, 1)*H23/SQRT(20)**」と入力するとLが求められる。
3. セルH27に「**=H22+NORM.INV(1-0.05, 0, 1)*H23/SQRT(20)**」と入力するとUが求められる。

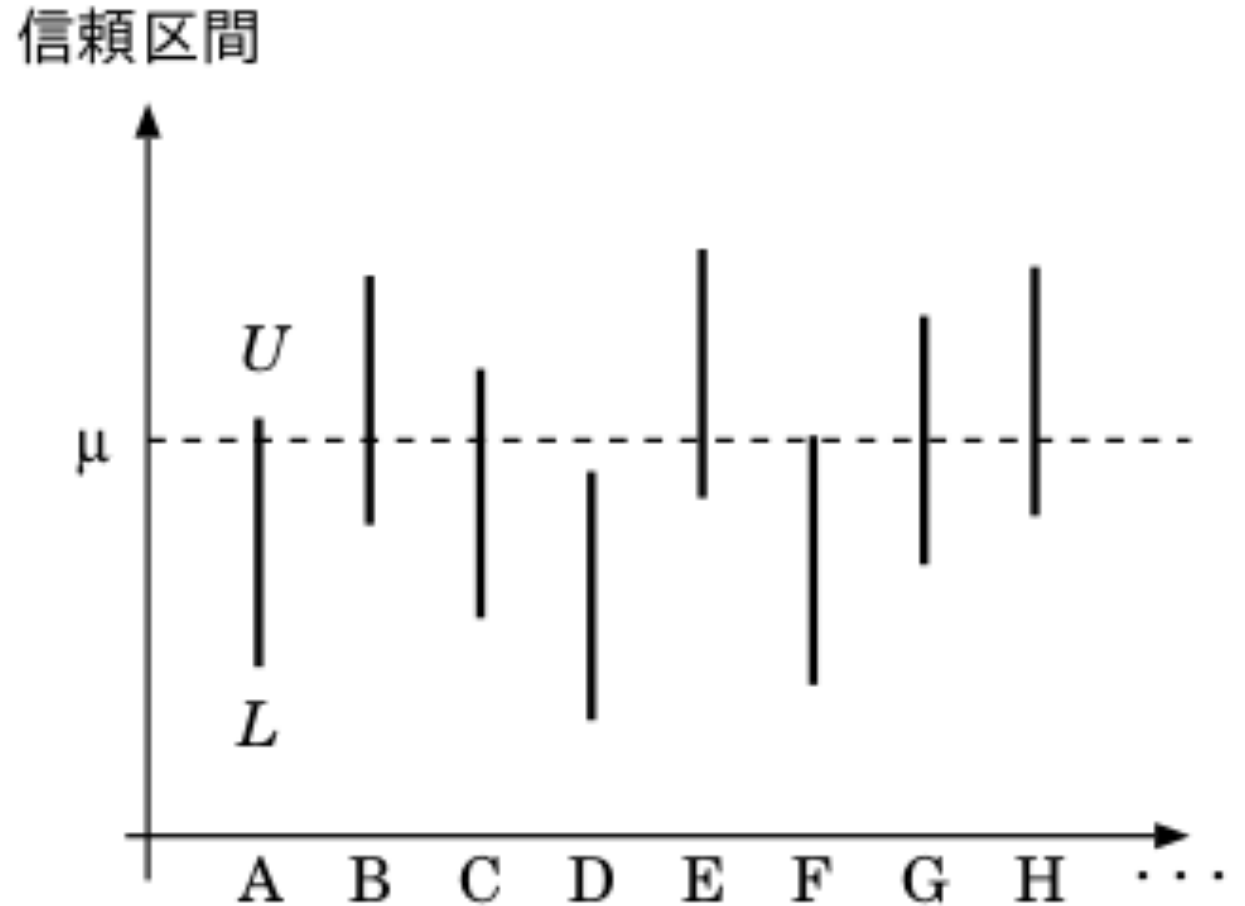
空のセルに「**=AVERAGE(D2:D501)**」と入力して身長之母平均 μ を計算し、求めた信頼区間と比較してみよう。

信頼係数の意味

信頼区間は、信頼係数 $1 - \alpha$ と n と抽出した標本によって求まる。

信頼係数の意味

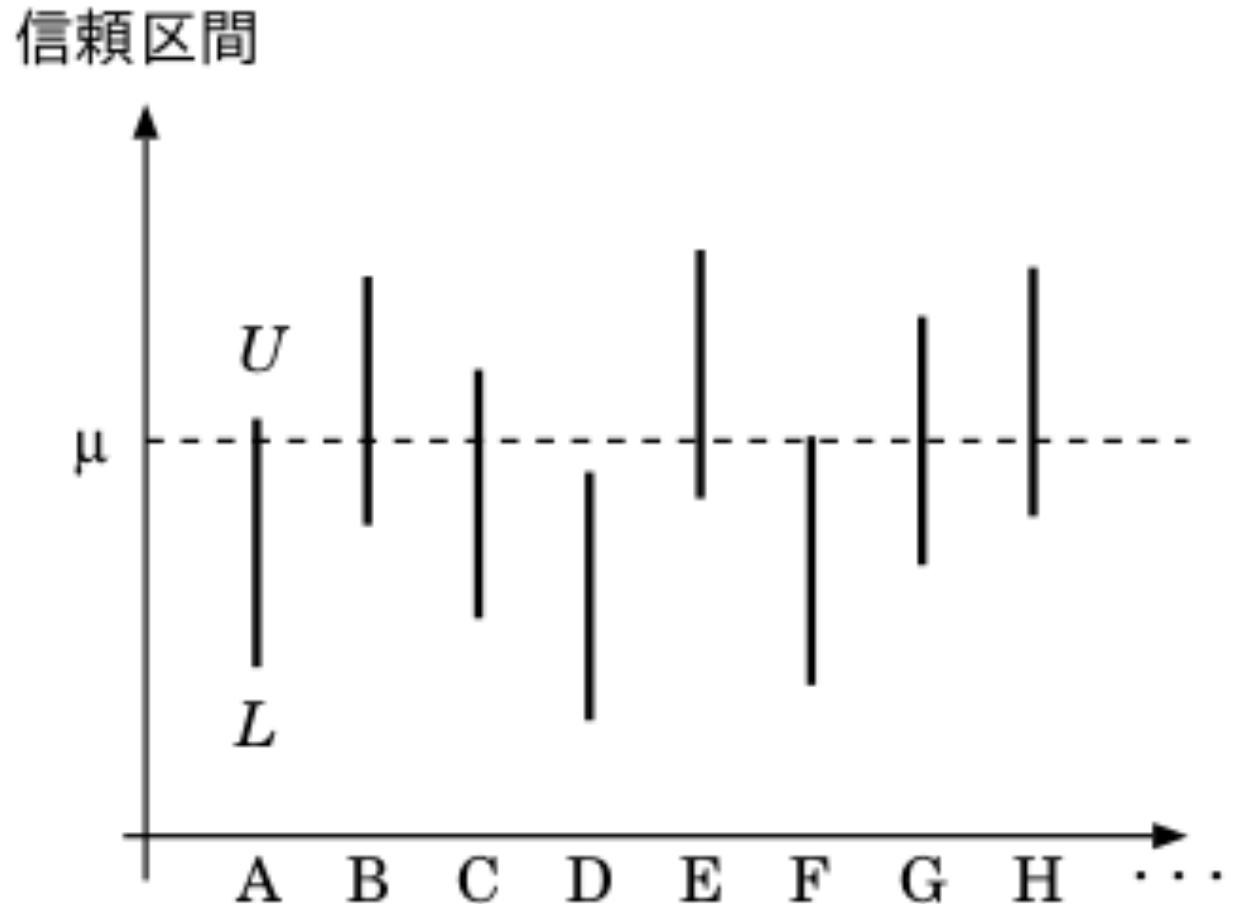
右の図はAさん、Bさん、Cさん、...
がそれぞれ同じ条件で標本を無作為
抽出して得られた
信頼区間を縦棒で
表たものである。



信頼係数の意味

無作為抽出なので、人によって標本が異なり、それによって信頼区間も異なる。

右図の場合、Dさん以外は μ を含む信頼区間が推定できている。



信頼係数の意味

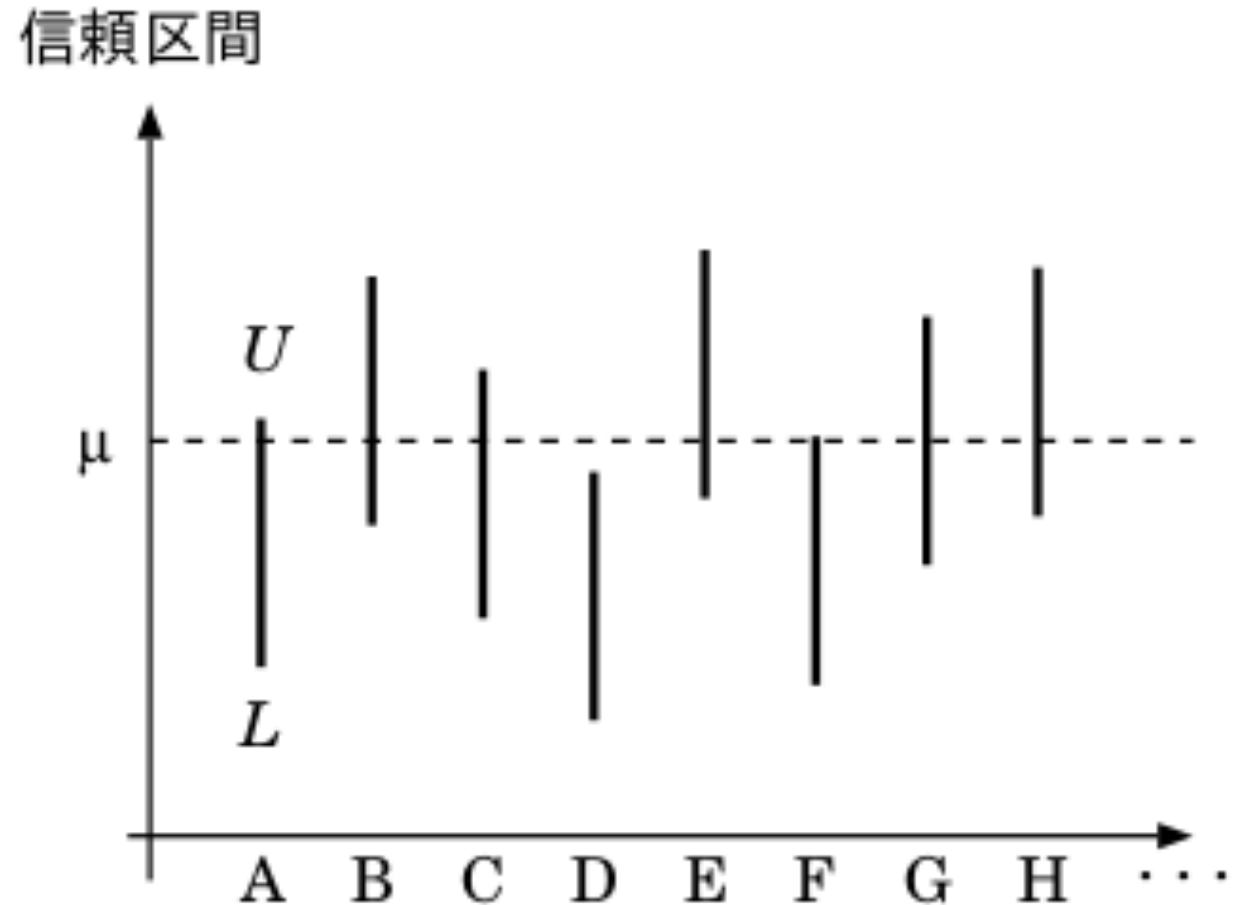
信頼区間は、どれくらいの人々が μ を含む信頼区間を標本から得られるかを決める確率といえる。

信頼係数の意味

例えば、100人がそれぞれ同じ条件で標本を無作為抽出して95%の信頼区間を求めたとすると、95人は μ を含む信頼区間が推定できても、5人は外れる可能性がある。

信頼係数の意味

信頼係数が大きいほど信頼区間は広く
(右図の縦棒が長くなり)、 μ から外れる可能性は低くなる。



信頼係数の意味

しかし、信頼区間が広がると μ の推定の誤差が大きくなるので、信頼係数をむやみに大きくすれば良いというものではない。

信頼係数の意味

- 信頼区間を狭くして推定の誤差を小さくするには、標本を多く抽出するのが良い。
- 例えば、標本サイズ n を9倍にすると、信頼区間の幅は $1/\sqrt{9} = 1/3$ 倍になり、推定の誤差が小さくなる。

今回の問い

- 母数とは何か？
 - 母集団分布の特徴を表す代表値や散らばりの尺度
- 母数の推定方法には何があるか？
 - 母数の推定方法には点推定と区間推定がある。

今回学んだこと

1.母数と統計量

2.母平均の推定

今回の目標

表計算ソフトを用いて母平均を推定できるように
なる。

課題

dataset1.odsにおいて、500人の体重のデータを母集団とみなす。演習と同様にこの母集団から30人の標本 ($n = 30$) を無作為抽出し、その標本から信頼係数95% ($1 - \alpha = 0.95$) の信頼区間を求めなさい。母分散は未知とする。作成したファイルはODF表計算ドキュメント(拡張子.ods)で提出すること。

提出先 : yutaka.hirai@koeki-u.ac.jp

件名 : データリテラシー第8回課題 [学籍番号]

締め切り : 12月4日 (金)

次回

第 1 回	社会におけるデータ・AI利活用
第 2 回	表計算ソフトの基本操作
第 3 回	セルの参照と集計
第 4 回	度数分布表とヒストグラム
第 5 回	代表値と散らばりの尺度
第 6 回	データのグラフによる表現とデータ比較
第 7 回	データの分布と標本抽出
第 8 回	推定
第 9 回	分割表とクロス集計
第 1 0 回	2 変量データと相関
第 1 1 回	時系列データ
第 1 2 回	仮説検定
第 1 3 回	回帰分析