

統計工学 講義資料(4)

~母集団と標本~

(C) Sadami SUZUKI

標準正規分布と確率

標準正規分布・・・数値的評価による**正規分布表**と呼ばれる数値表が与えられており、その任意の値に対する**確率密度の値**、および**任意の区間の間の値をとる確率**などを求めることが可能

確率密度関数

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \quad z \sim N(0, 1^2)$$

$$P(k) = \int_k^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$$

(C) Sadami SUZUKI

標準正規分布と確率

・ **正規分布表**とを活用した**任意の区間の間の値をとる確率の求め方**

任意の確率変数  $X$  に対して・・・

$$P[a \leq X \leq b] = P\left[\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right]$$

(C) Sadami SUZUKI

標準正規分布と基準化

平均が0、分散が1の正規分布を標準正規分布という

基準化

$$\text{変数変換 } Z_i = \frac{X_i - \mu_i}{\sigma_i}$$

あらゆる正規分布が基準化により**標準正規分布に変換可能**

異なる集団を同一土俵上で**比較することが可能となる!**

(C) Sadami SUZUKI

正規分布を前提として、これを活用したものが「**偏差値**」

全体の分布がどのようなものであるかは平均値と分散(標準偏差)により把握可能

基準化したZ値を用いれば、**異なる科目、母集団を同一土俵上で比較可能**

基準化

$$\text{変数変換 } Z_i = \frac{X_i - \mu_i}{\sigma_i}$$

偏差値

$$\text{偏差値} = 50 + 10Z_i$$

$$\sim N(50, 10^2)$$

(C) Sadami SUZUKI

母集団と標本

母集団

抽出 (サンプリング)

標本 (サンプル)

統計的記述

統計的推測

母集団

標本

平均:  $\mu$ , 分散:  $\sigma^2$

(C) Sadami SUZUKI

統計工学 Statistical Engineering

無作為抽出

番号	$x_i$
1	50.8
2	50.3
3	49.6
4	50.6
5	51.5
6	51.0
7	50.8
8	49.2
9	50.7

$\bar{x} = 50.5$

平均:  $\mu$ , 分散:  $\sigma^2$

・ サンプルの抽出法は・・・?

(C) Sadami SUZUKI

統計工学 Statistical Engineering

統計値と統計量

平均:  $\mu$ , 分散:  $\sigma^2$

抽出

サンプル1  
サンプル2  
サンプル3  
サンプル4  
...  
サンプルn

$\bar{x}_1 = \bigcirc \bigcirc$   
 $\bar{x}_2 = \bigcirc \bigcirc$   
 $\bar{x}_3 = \bigcirc \bigcirc$   
 $\bar{x}_4 = \bigcirc \bigcirc$   
 $\vdots$   
 $\bar{x}_n = \bigcirc \bigcirc$

・ 統計値... 標本(サンプル)によって求められた値  
- サンプルが異なればその標本平均は異なる!

(C) Sadami SUZUKI

統計工学 Statistical Engineering

統計値と統計量

平均:  $\mu$ , 分散:  $\sigma^2$

・ サンプルを **母集団分布** に従う **確率変数** として捉え, その関数 **統計量** を考える

観測値:  $x_1, x_2, \dots, x_n$

確率変数:  $X_1, X_2, \dots, X_n$

統計的記述

統計量

・ 統計量... サンプルに対応する **確率変数** の関数:  $g(X) = \bar{x} = \sum_{i=1}^n \frac{x_i}{n}$

・ 統計値... 統計量の実現値

(C) Sadami SUZUKI

統計工学 Statistical Engineering

標本平均  $\bar{X}$  の標本分布

平均値  $E(\bar{X}) = E\left[\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right] = \frac{1}{n}E(X_1 + X_2 + \dots + X_n)$   
 $= \frac{1}{n}[E(X_1) + E(X_2) + \dots + E(X_n)]$   
 $= \mu \quad \because E(X_1) = E(X_2) = \dots = E(X_n) = \mu$

分散  $V(\bar{X}) = V\left[\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right] = \frac{1}{n^2}V(X_1 + X_2 + \dots + X_n)$   
 $= \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n} \quad \because V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n)$

平均値  $\mu$ , 分散  $\sigma^2$  の母集団からとられた大きさ  $n$  の標本の平均値  $\bar{X}$  の期待値は  $\mu$ , 分散は  $\sigma^2/n$  である。

(C) Sadami SUZUKI

統計工学 Statistical Engineering

標本平均  $\bar{X}$  の標本分布

・ 母集団の分布が正規分布であるときは・・・  
標本平均の分布もまた正規分布となる

・ 【定理】いくつかの独立な確率変数のそれぞれが正規分布に従うとき, それらの変数の任意の一次結合は正規分布に従う

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

母集団の分布が正規分布  $N(\mu, \sigma^2)$  であるとき, そこからとられた大きさ  $n$  の無作為標本  $x_1, x_2, \dots, x_n$  の平均値  $\bar{x}$  の標本分布は  $N(\mu, \sigma^2/n)$  である。

したがって

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

の分布は標準正規分布  $N(0,1)$  である。

(C) Sadami SUZUKI

統計工学 Statistical Engineering

・ では, 母集団が正規分布に従うとは限らない場合は・・・

中心極限定理

分布がどのようなものであっても, 平均値  $\mu$ , 分散  $\sigma^2$  をもつ母集団からとられた大きさ  $n$  の標本の平均値  $\bar{x}$  の分布は,  $n$  が大きいとき正規分布  $N(\mu, \sigma^2/n)$  に近づく。したがって

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

の分布は,  $n$  が大きいとき標準正規分布  $N(0,1)$  に近づく。

「正規分布が多くの場合に高い適用可能性を有することを示す定理」と位置づけられる

(C) Sadami SUZUKI