確率と統計　第 9 回

正規分布を使いこなす

6 月 15 日

Table of contents

ありふれた変数の分布

この世でもっとありふれた分布とはどのような分布だろうか.

ありふれた変数の分布

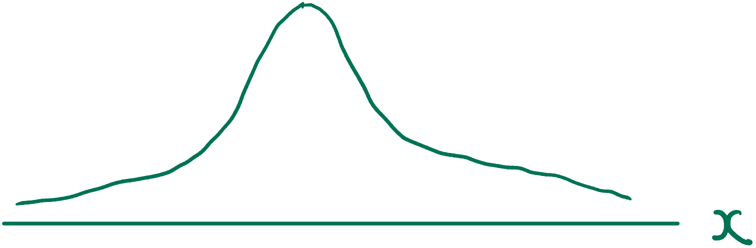
この世でもっとありふれた分布とはどのような分布だろうか.

例えばある県の成人男性の身長を変数 X としてその分布を記録する. このときその分布を表す図は以下のようになるだろう:

ありふれた変数の分布

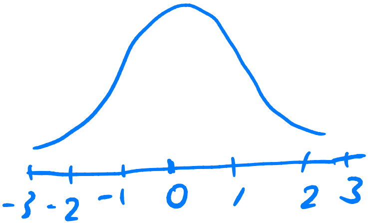
この世でもっとありふれた分布とはどのような分布だろうか.

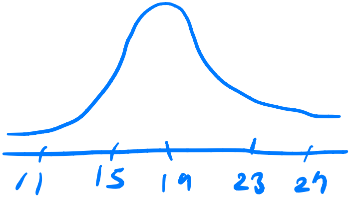
例えばある県の成人男性の身長を変数 X としてその分布を記録する. このときその分布を表す図は以下のようになるだろう:



正規分布モデル

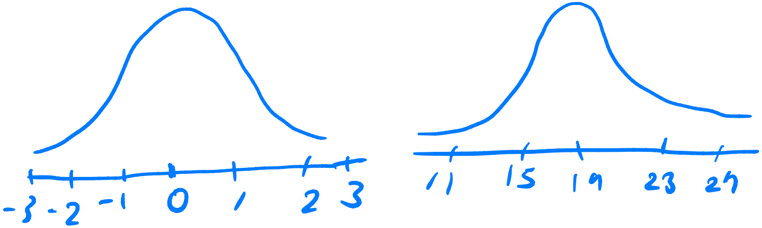
上記のような分布を正規分布という．正規分布は常に左右対称で，単峰で，釣鐘型の曲線だ．





正規分布モデル

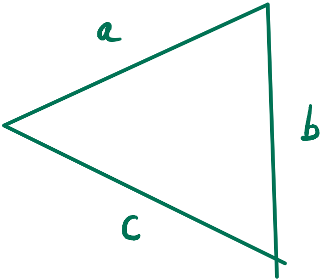
上記のような分布を正規分布という．正規分布は常に左右対称で，単峰で，釣鐘型の曲線だ．

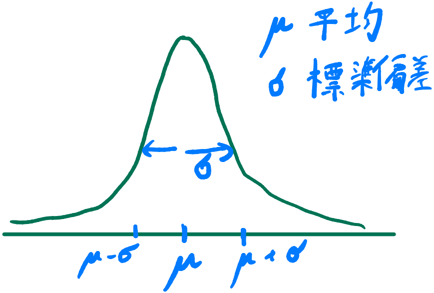


現実にある多くの変数はほぼ正規分布に従っているが，厳密に正規分布に従っている変数は存在しない．これは完全に厳密な三角形が現実に存在しないことと似ている．

正規分布モデル

三角形が三辺の長さで一つに定まるように, 正規分布も平均と標準偏差という 2 つのパラメータを決めると一つに定まる.





正規分布モデル

三角形が三辺の長さで一つに定まるように, 正規分布も平均と標準偏差という 2 つのパラメータを決めると一つに定まる.

そこで平均が μ で標準偏差が σ の正規分布を $N(\mu, \sigma)$ と表記する.

正規分布モデル

三角形が三辺の長さで一つに定まるように, 正規分布も平均と標準偏差という 2 つのパラメータを決めると一つに定まる.

そこで平均が μ で標準偏差が σ の正規分布を $N(\mu, \sigma)$ と表記する.

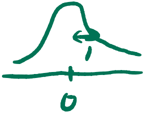
例えば, 平均が 0 で標準偏差が 1 の標準正規分布を以下のように表記する:

正規分布モデル

三角形が三辺の長さで一つに定まるように, 正規分布も平均と標準偏差という 2 つのパラメータを決めると一つに定まる.

そこで平均が μ で標準偏差が σ の正規分布を $N(\mu,\sigma)$ と表記する.

例えば, 平均が 0 で標準偏差が 1 の標準正規分布を以下のように表記する:



$$N(0,1) \quad \text{または} \quad N(\mu=0,\sigma=1).$$

例題: 正規分布の表記

次の正規分布を短縮形で書き記しなさい.

例題: 正規分布の表記

次の正規分布を短縮形で書き記しなさい.

(a) 平均 5 および標準偏差 3
(b) 平均-100 および標準偏差 100
(c) 平均-100 および <u>分散</u> 100

例題: SAT と ACT

下の表は SAT と ACT の総合スコアの平均と標準偏差を示している. SAT と ACT のスコアの分布はどちらも正規分布に近い. アンのスコアが 1300 でトムのスコアが 24 だったとすると, どちらの成績が良いだろうか.

sat_and_act

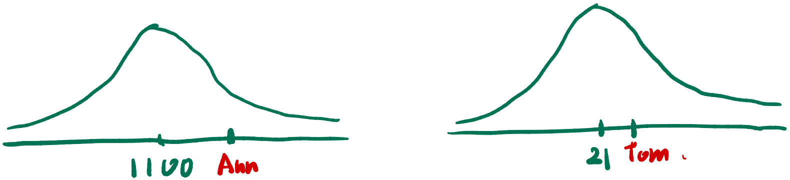
	SAT	ACT
平均	1100	21
標準偏差	200	6

例題: SAT と ACT

アンとトムのスコアの位置はそれぞれ次のようになる:

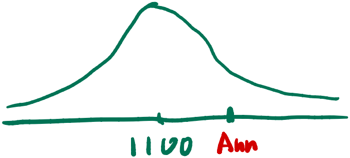
例題: SAT と ACT

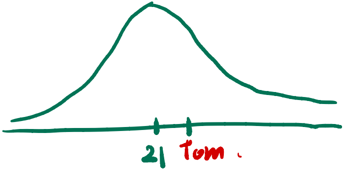
アンとトムのスコアの位置はそれぞれ次のようになる:



例題: SAT と ACT

アンとトムのスコアの位置はそれぞれ次のようになる:

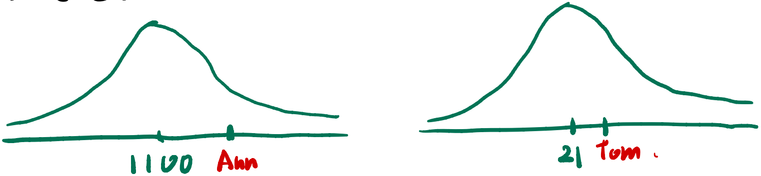




アンとトムのスコアがそれぞれどれだけ平均から離れているか, 平均との差と標準偏差との割合で計算すると...

例題: SAT と ACT

アントムのスコアの位置はそれぞれ次のようになる:



アントムのスコアがそれぞれどれだけ平均から離れているか, 平均との差と標準偏差との割合で計算すると...

$$\frac{1300-1100}{200} = 1 > \frac{24-21}{6} = 0.5$$

よって Ann の成績の方がよい.

Z スコアを用いた標準化

平均 μ, 標準偏差 σ の分布に従う観測値 x の Z スコアを以下の式で定義する:

Z スコアを用いた標準化

平均 μ, 標準偏差 σ の分布に従う観測値 x の Z スコアを以下の式で定義する:

$$Z = \frac{x - \mu}{\sigma}.$$

Z スコアを用いた標準化

平均 μ, 標準偏差 σ の分布に従う観測値 x の Z スコアを以下の式で定義する:

$$Z = \frac{x - \mu}{\sigma}.$$

つまり Z スコアは観測値 x が平均 μ から標準偏差 σ の何倍離れているかを表す.

Zスコアを用いた標準化

平均 μ, 標準偏差 σ の分布に従う観測値 x のZスコアを以下の式で定義する:

$$Z = \frac{x - \mu}{\sigma}.$$

つまり Zスコアは観測値 x が平均 μ から標準偏差 σ の何倍離れているかを表す.

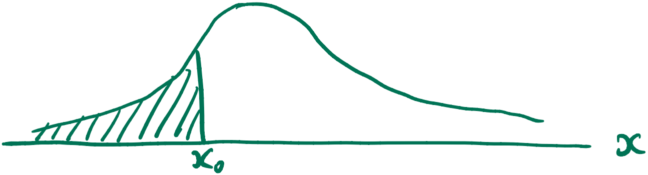
Zスコアを計算することによって素点 x の分布における相対的な位置が分かる.

例題: フクロギツネの頭長

フクロギツネの頭長は平均 92.6mm, 標準偏差 3.6mm の正規分布に従っている. 頭長が 95.4mm と 85.8mm のフクロギツネの Z スコアを求めよ. またどちらの値がより観測されにくいか.

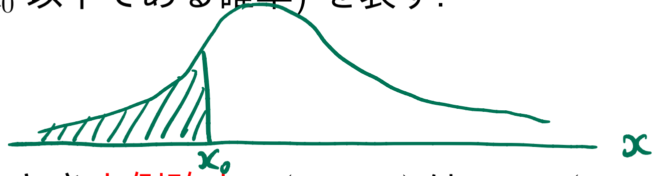
片側確率とは

下図のように色塗りした部分のグラフ全体の面積に占める割合は下側確率 $P(X \leq x_0)$ (観測値 x が x_0 以下である確率) を表す.

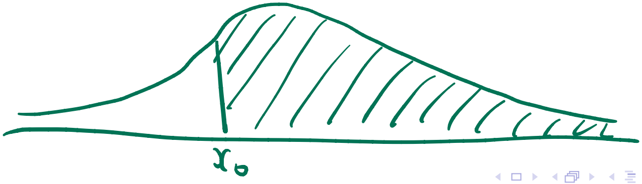


片側確率とは

下図のように色塗りした部分のグラフ全体の面積に占める割合は下側確率 $P(X \leq x_0)$ (観測値 x が x_0 以下である確率) を表す.



このとき上側確率 $P(X \geq x_0)$ は $1 - P(X \leq x_0)$ によって求まる.



片側確率の求め方 (正規分布の場合)

いつでもまずは分布の図を描き, 値を入れ, 対応する領域を塗りつぶす. 次にその図から求める確率を確認する. そして対応する Z スコアを計算して確率を求める.

片側確率の求め方 (正規分布の場合)

いつでもまずは分布の図を描き, 値を入れ, 対応する領域を塗りつぶす. 次にその図から求める確率を確認する. そして対応する Z スコアを計算して確率を求める.

Z スコアから片側確率を求める方法には以下がある:

(a) Python, R, Excel (Google Sheets) のどれかを使う.

(b) 確率分布表を参照する.

例題: SAT スコア

SAT スコアは正規モデル $N(\mu = 1100, \sigma = 200)$ によって近似される. 全受験者の中からシャノンが無作為に抽出された. シャノンのスコアが 1190 以上である確率はいくらか.

例題: SAT スコア

SAT スコアは $N(1100, 200)$ に従っている.

(a) 95 パーセンタイルにある人の SAT スコアを求めなさい.

(b) 97.5 パーセンタイルにある人の SAT スコアを求めなさい.

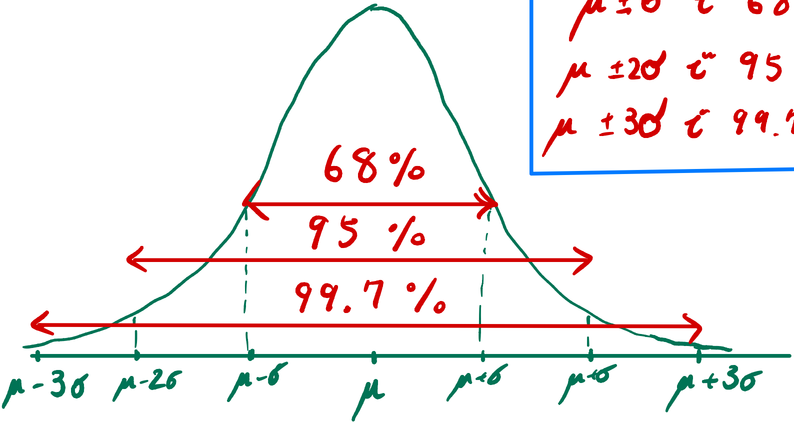
68:95:99.7 ルール

正規分布において, 平均 ± 1 標準偏差, 2 標準偏差, 3 標準偏差の範囲の確率は以下の図のようになる:

$\mu \pm \sigma$ で 68%

$\mu \pm 2\sigma$ で 95%

$\mu \pm 3\sigma$ で 99.7%



例題: 割合の概算

SAT スコアは平均 $\mu = 1100$ と標準偏差 $\sigma = 200$ の正規分布に従っていると近似できる.

(a) スコアが 700 から 1500 である受験者は約何パーセントか.

(b) スコアが 1100 と 1500 の間であるのは約何パーセントか.