統計基礎　目次

1.　量的データと質的データ

データは情報を表現する値のことで、統計で扱うデータは量的データと質的データとに分かれる。

量的データ…体重や身長のように、その値自体に量として意味を持つデータ
量的データはさらに比例尺度データと間隔尺度データとに分かれる
比例尺度データ…数値の差、数値の比例にも意味のあるデータ
（身長や体重などのデータ）
間隔尺度データ…数値の差には意味があるが、数値の比例には意味のなデータ（温度や試験の成績などのデータ）

質的データ…血液型や試験の順番のように量的意味を持たないデータ
質的データはさらに順序尺度データと名義尺度データとに分けられる
順序尺度データ…順序として意味を持つデータ
（試験の成績結果の順番やアンケートの結果など）
名義尺度データ…血液型や男女の分類などのように区分するためのデータ

量的データはその値自体の分布状況は何等かの分布状況を示すが、質的データはその値自体特別な分布状況は示さない

2.　確率分布

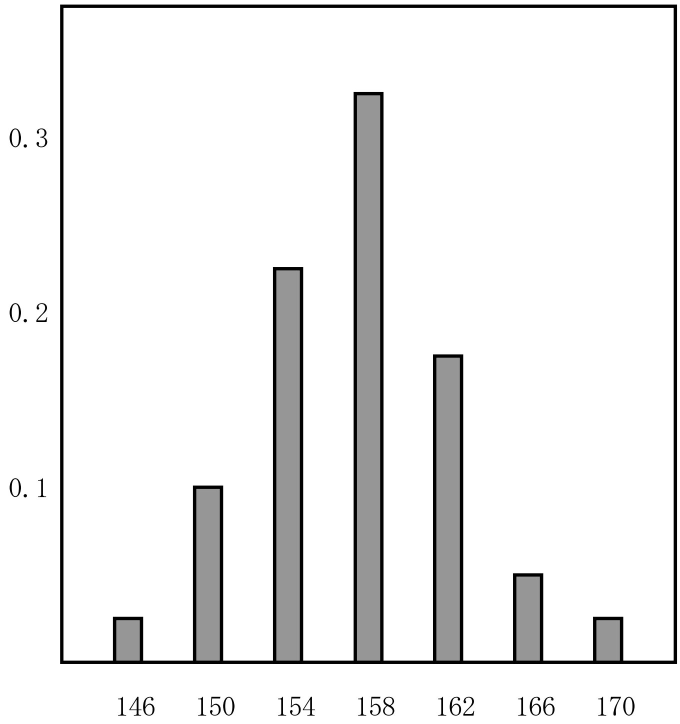
2.1　度数分布

いま、２０才の女性５０人の身長を測定し、その測定した値をもとにして下表のような度数分布表が得られたとする。

［　度 数 分 布 表　］

ＮＯ	階級値	度数	相対度数	累積度数	累積相対度
1	146	2	4%	2	4%
2	150	6	12%	8	16%
3	154	12	24%	20	40%
4	158	17	34%	37	74%
5	162	8	16%	45	90%
6	166	3	6%	48	96%
7	170	2	4%	50	100%
	合計	50	100%		

この度数分布表の相対度数をグラフ化してみると下図のようになる。



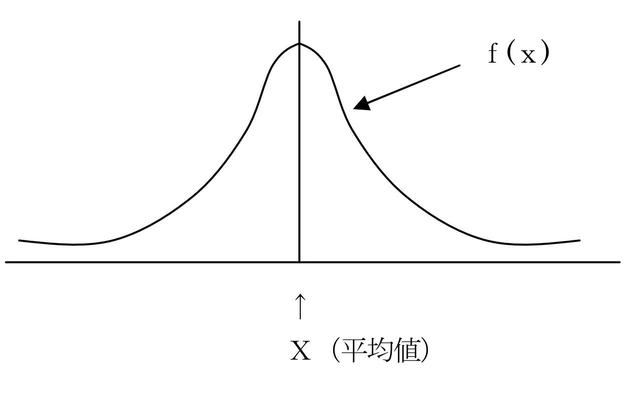
相対度数は、各階級の度数を度数合計の値によって割ったものであるから、各相対度数の合計は１（１００％）になる。また各相対度数は各階級の出現率を示している。

身長１４４～１４８cm（階級値＝１４６cm）の範囲に全体の４％の人が属しており、１４８～１５２cm（階級値＝１５０cm）の範囲に全体の１２％の人が属している。このことは、発生（出現）する確率が０．０４、０．１２あるといえる。サンプル数をもっと増やしていくと、このグラフは中央部が高く左右にすその広がった滑らかな曲線になってくる。このような確率分布の曲線を確率分布曲線といい､身長の出現率のように中央部が高く左 右にすそが広がった左右対象のグラフを正規分布曲線という。

正規分布をするものは自然界に広く見いだすことができる。

2.2　正規分布…連続分布（連続した値をとる分布）

正規分布のグラフは、下のように連続した曲線のグラフになる。



正規分布は一般に下の式で表現される。

$$f(x) = \frac{1}{\delta\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)^2}$$

μ ：母平均　　　σ ：母標準偏差　　e ：対数の底（＝2.1828）

正規分布では　母平均＝ μ 　母分散＝ σ^2 　である。

正規分布は　　N（μ，σ^2）で表現することが多い。

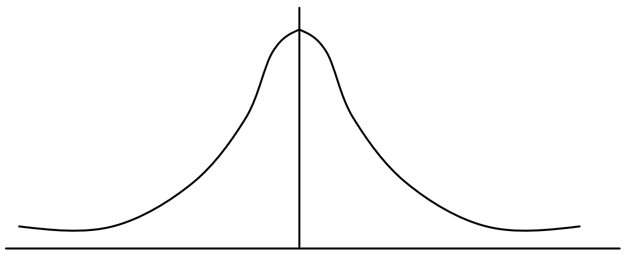
正規分布　母平均　母分散

正規分布の式を、データの標準化をすると、母平均＝０，母分散＝１^2 となる。このようにして　得られた曲線を標準正規分布曲線という。

標準正規分布曲線の式は

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

この式をグラフに書くと



↑

0

母平均（＝０）を中心にして左右対象のグラフとなる。

通常標準正規分布は、N（０，１^2）で表現される。

正規分布を標準正規分布にするには、正規分布のデータ（x）を $\frac{x-\mu}{\delta}$ で変換することにより母平均＝０、母分散＝１^{2}の標準正規分布に変えることができる。

2.3　標準正規分布

正規分布の式を、データの標準化を行い、平均値＝０：分散＝1^2 になるようにしたものが標準正規分布と呼ばれる分布である。

正規分布

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

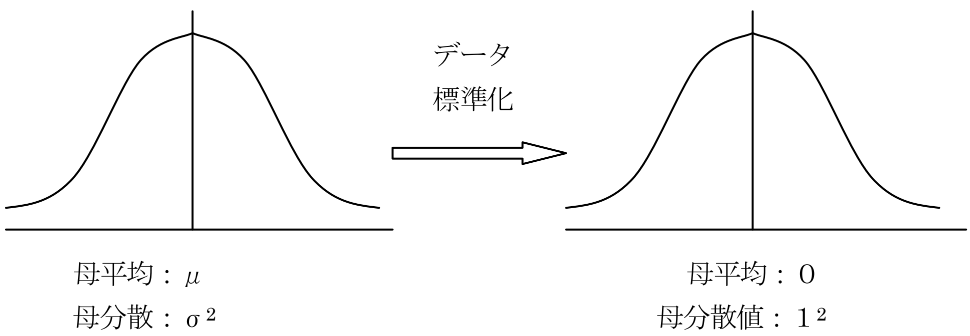
ｘは正規分布に従っている。この時

$$Z = \frac{x-\mu}{\delta}$$　（μ：母平均　σ：母標準偏差）

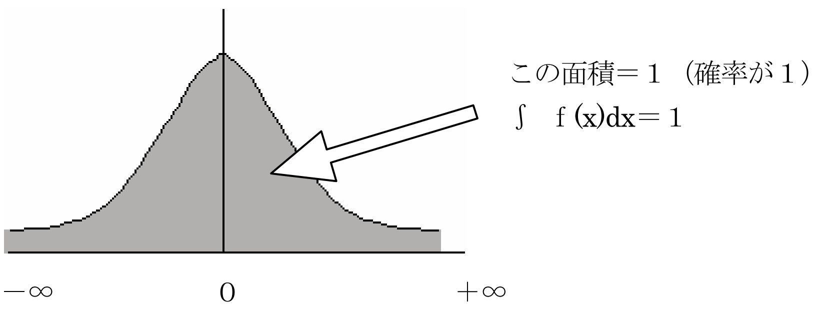
とすると、Ｚの分布は標準正規分布に従う。

標準正規分布の式は、

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



標準正規分布については、その数表が求められており、数表から分布の値（確率値）を得ることができる。



標準正規分布においては、その面積値が確率値を示している。標準正規分布では、母平均＝０・母分散＝１で平均値を中心とした左右対象の分布である。

通常、標準正規分布の数表で表示されている値は、０から右側（＋側）の面積値が表示されている。つまり $\int f(x)dx$ の値が表示されている。

2.4　正規分布と標準正規分布の特徴

［正規分布の特徴］

①母平均：μ を中心とした左右対象の分布である。

②母平均：μ　母分散値：σ^2　で表す。

③正規分布は　N（μ，σ^2）で表す。

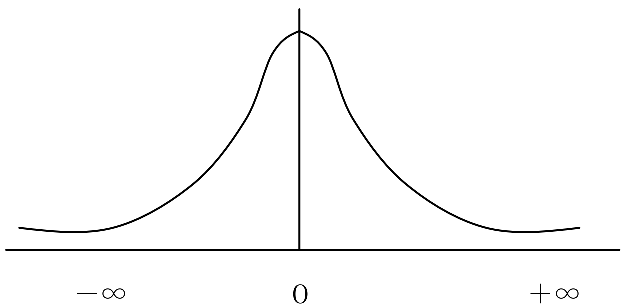
④２つの確率変数ｘとｙが、別々にN（μ_1，σ_1^2）、N（μ_2，σ_2^2）に従う時

ｘ＋ｙの分布は、N（$\mu_1+\mu_2$，$\sigma_1^2+\sigma_2^2$）に従う

ｘ－ｙの分布は、N（$\mu_1-\mu_2$，$\sigma_1^2+\sigma_2^2$）に従う

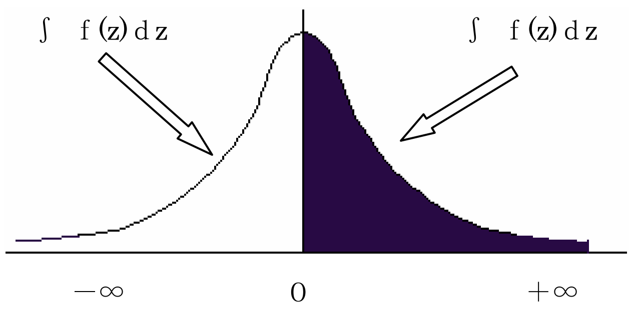
［標準正規分布の特徴］

①母平均＝０を中心とした左右対象のグラフ



左右対象の偶関数であるから、ｚが－∞から０までの面積と０から＋∞までの面積は等しい。

また－∞から＋∞までの面積は１である。

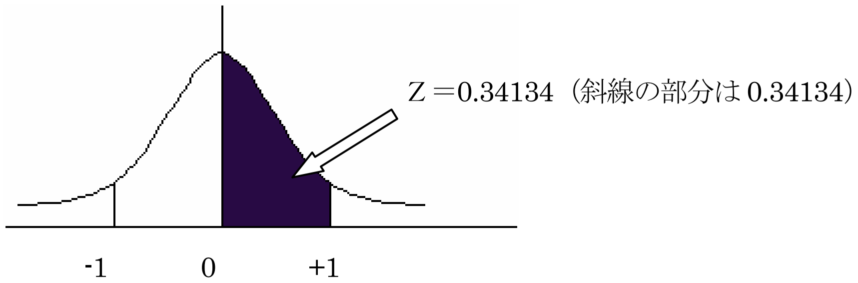


$\int f(z)dz = \int f(z)dz$ であり、また　$\int f(z)dz = 1$　である。

②確率分布

$Z=\int f(z)dz$　とする

(1)z＝１の時



標準正規分布は左右対象の偶関数であるから

－１≦z≦１の間のＺの値は

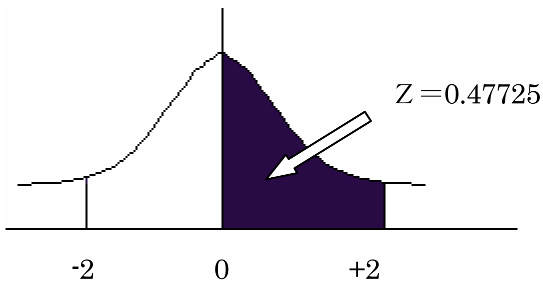
0.34134 × 2＝ 0.68268　となる

このことは、標準正規分布に従う分布においては、－１≦z≦１の間に全データの

６８．２６８％が入る

(2) z ＝２の時

同様に －２≦ z ≦２の間には全データの９５．４５％がはいる。



(3) z ＝３の時

同様に －３≦ z ≦３の間には全データの９９．７３％がはいる。