

統計基礎 目次

1.	量的データと質的データ	1
2.	確率分布	1
2.1	度数分布	1
2.2	正規分布…連続分布（連続した値をとる分布）	2
2.3	標準正規分布	4
2.4	正規分布と標準正規分布の特徴	5
2.5	正規分布以外の確率分布	7
2.5.1	2項分布…離散分布（とびとびの値をとる分布）	7
2.5.2	χ^2 分布・t 分布・F 分布…標本分布	8
3.	母集団と標本について	11
3.1	母集団と標本	11
3.2	母数について	11
3.3	推定	11
3.4	中心極限定理	12
4.	検定	13
4.1	検定の方法	13
4.2	両側検定と片側検定	14
4.3	検定における2種類の誤り	15
5.	母平均の差の検定	16
5.1	2群の母平均の差に関する検定	16
5.2	3群以上の母平均の差に関する検定	16
5.3	母平均の差の検定手順	18
6.	基本統計量を求める	20
6.1	基本統計量	20
6.2	基本統計量を求める	20
6.2.1	基本統計量を関数を使用し求める	20
6.2.2	統計ツールを使用して、基本統計量を求める	21
6.2.3	基本統計量のデータの意味	23
7.	度数分布表・ヒストグラムの作成	24
7.1	ツールを使用した度数分布表・ヒストグラムの作成	24
7.2	データの標準化	25
7.3	偏差値	26
8.	対応のある2群の母平均の差の検定	27
8.1	2群母集団がほぼ正規分布に従いかつ対応のある時	27

8.2	t 値に関する関数	28
9.	対応のない2群の母平均の差の検定	30
9.1	等分散の検定を実施して2群の分散を調べる	30
9.1.1	2群は等分散であるか検定し確認する	30
9.1.2	F 値に関する関数	33
9.2	対応のない等分散2群母集団の母平均の差の検定	34
9.2.1	等分散である2群母集団の「母平均の差の検定」を実施	34
9.2.2	t 検定に関する関数を使用して検定を実施	36
10.	大標本での母平均の差の検定…Z 検定	39
10.1	分析ツールを使用して検定を実施	40
11.	1元配置分散分析	43
11.1	1元配置分散分析	43
11.2	1元配置分散分析の実施	43
12.	繰り返しのない2元配置分散分析	45
12.2	繰り返しのない2元配置分散分析の実施	45
12.3	交互作用の検討	48
13.	繰返しのある2元配置分散分析	49
13.1	繰返しのある2元配置分散分析	49
14.	2変量間に関する分析	53
14.1	2変量間の相関	53
14.1.1	散布図の作成	53
14.1.2	相関係数	54
15.	単回帰分析	56
15.1	単回帰式を求める	56
15.2	説明変量と目的変量	57
15.3	単回帰分析を実施	57
15.3.1	分析ツールを使用し単回帰分析実施	57
15.3.2	回帰に関する関数	65
16.	重回帰分析	68
16.1	重回帰式を求める	68
16.2	重回帰分析を実施	69
16.2.1	分析ツールを使用して重回帰分析実施	69
16.2.2	重回帰に関する関数を使用して重回帰分析を実施	73
16.2.3	重回帰分析の多重共線性	77
16.2.4	標準偏回帰係数	78

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いデータ (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

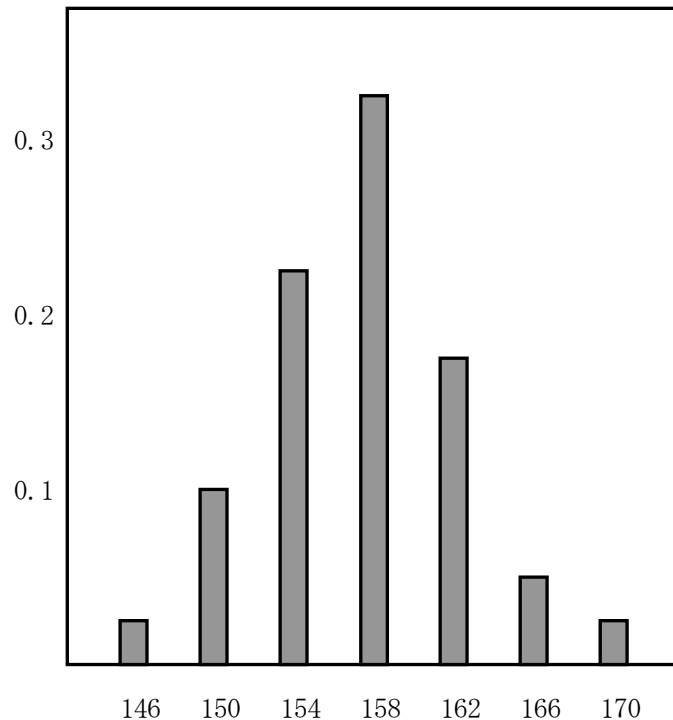
2.1 度数分布

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。



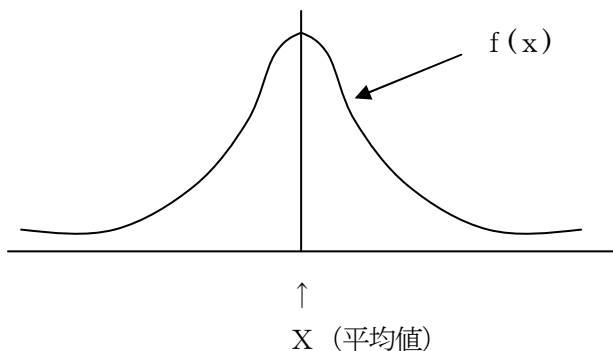
相対度数は、各階級の度数を度数合計の値によって割ったものであるから、各相対度数の合計は1（100%）になる。また各相対度数は各階級の出現率を示している。

身長144～148 cm（階級値＝146 cm）の範囲に全体の4%の人が属しており、148～152 cm（階級値＝150 cm）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の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 $N(\mu, \sigma^2)$ で表現することが多い。

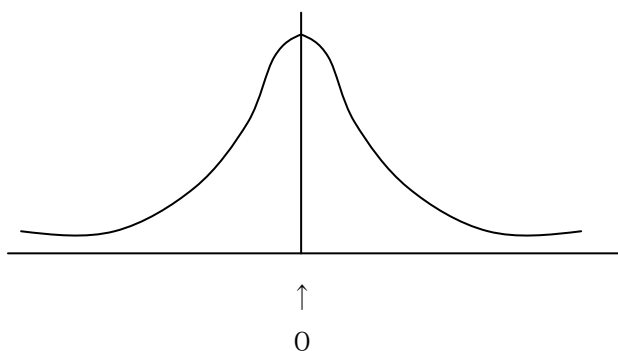
↑ ↑ ↑
正規分布 母平均 母分散

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=1²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



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=1²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=1² 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

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

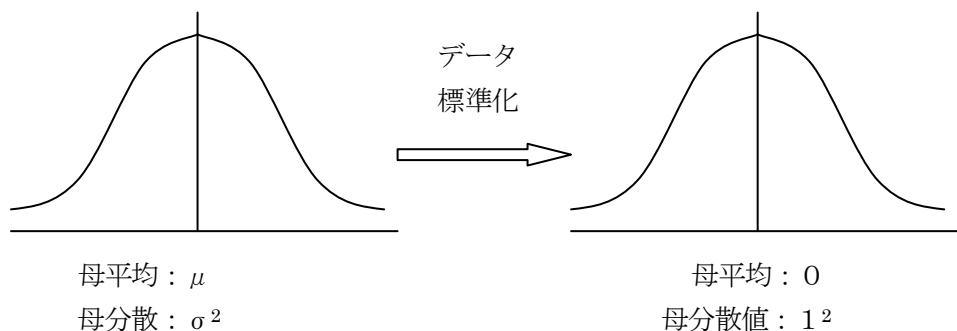
xは正規分布に従っている。この時

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

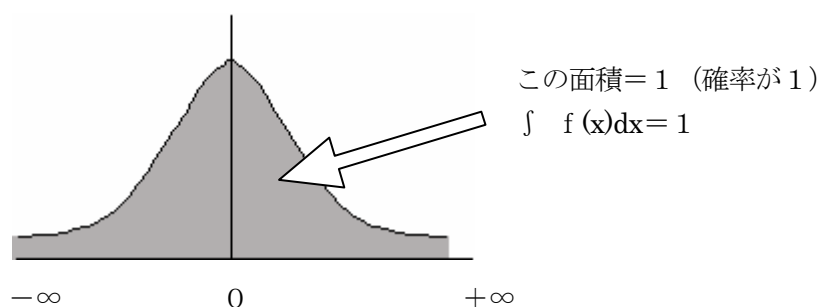
とすると、Zの分布は標準正規分布に従う。

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。

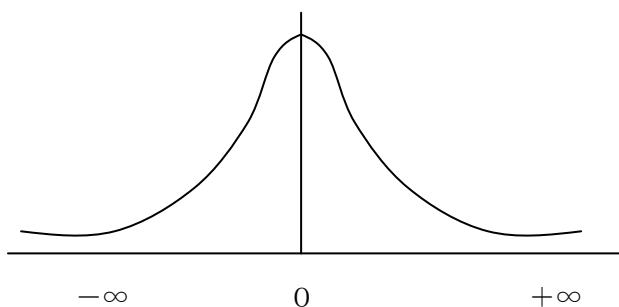
2.4 正規分布と標準正規分布の特徴

[正規分布の特徴]

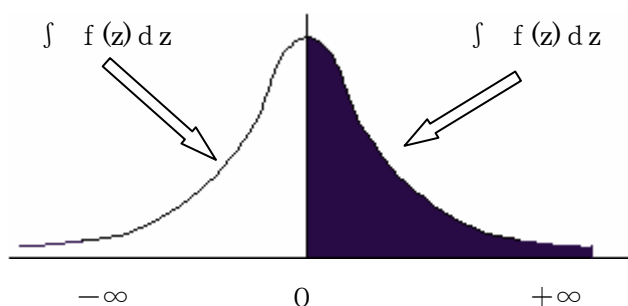
- ①母平均： μ を中心とした左右対象の分布である。
- ②母平均： μ 母分散値： σ^2 で表す。
- ③正規分布は $N(\mu, \sigma^2)$ で表す。
- ④2つの確率変数 x と y が、別々に $N(\mu_1, \sigma_1^2)$ 、 $N(\mu_2, \sigma_2^2)$ に従う時
 $x + y$ の分布は、 $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ に従う
 $x - y$ の分布は、 $N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$ に従う

[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 $-\infty$ から0までの面積と0から $+\infty$ までの面積は等しい。
また $-\infty$ から $+\infty$ までの面積は1である。

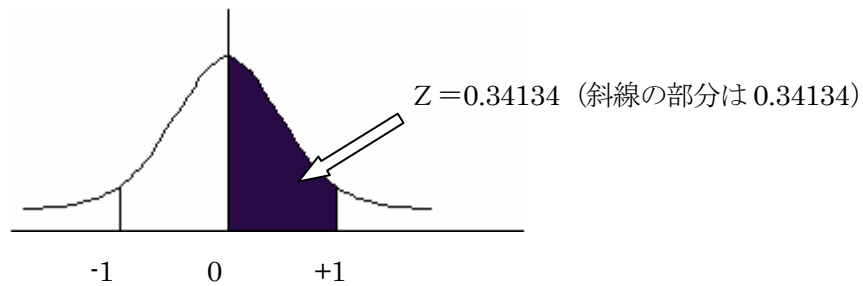


$\int_{-\infty}^0 f(z) dz = \int_0^{+\infty} f(z) dz$ であり、また $\int_{-\infty}^{+\infty} f(z) dz = 1$ である。

②確率分布

$$Z = \int f(z)dz \text{ とする}$$

(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

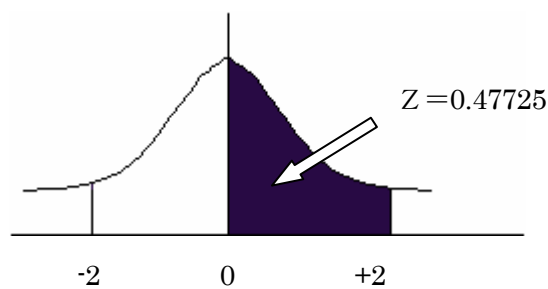
$-1 \leq z \leq 1$ の間の Z の値は

$$0.34134 \times 2 = 0.68268 \text{ となる}$$

このことは、標準正規分布に従う分布においては、 $-1 \leq z \leq 1$ の間に全データの 68.268%が入る

(2) $z=2$ の時

同様に $-2 \leq z \leq 2$ の間には全データの 95.45%がはいる。



(3) $z=3$ の時

同様に $-3 \leq z \leq 3$ の間には全データの 99.73%がはいる。

