研究員の眼

統計分析を理解しよう：正規分布、標準化、標準正規分布の概念

生活研究部 主任研究員　金　明中
(03)3512-1825　kim@nli-research.co.jp

正規分布とは？

今回は正規分布について説明したい。正規分布（normal distribution）とは、連続確率分布の一種である。まず、確率とは、ある出来事（事象、event）が起こる割合のことである。例えば、サイコロを投げると、6 種類の目の内どれか 1 つは必ず出てくるので、1 から 6 までの目が出る割合はどれも同じである。従って、それぞれの目が出る確率は、すべて 1/6 である（式 1））。

式 1）$P(1)=\frac{1}{6},\quad P(2)=\frac{1}{6},\quad P(3)=\frac{1}{6},\quad P(4)=\frac{1}{6},\quad P(5)=\frac{1}{6},\quad P(6)=\frac{1}{6}$

※*P* は probability で、確率を意味する。

また、 分布とは「あちこち分かれて広がること」という意味で、確率分布とはあるできごとが起こる確率の一覧（確率の集合）であり、上述したサイコロの確率分布は、式 2)のようになる。

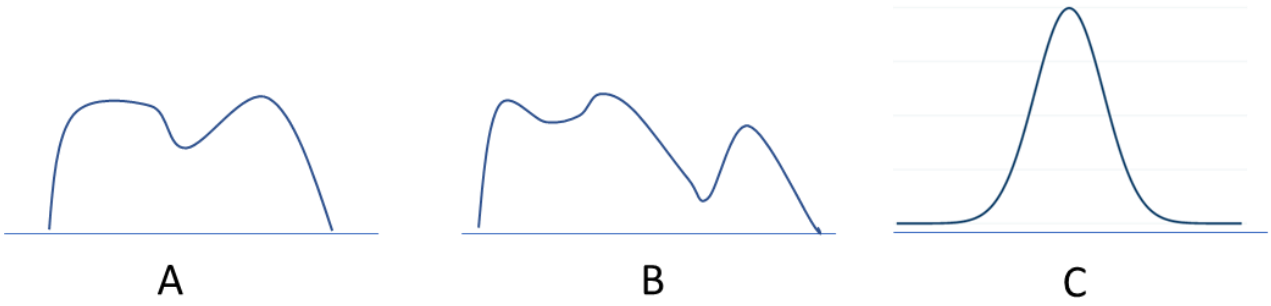
式 2）サイコロの確率分布　$\left\{\frac{1}{6},\quad \frac{1}{6},\quad \frac{1}{6},\quad \frac{1}{6},\quad \frac{1}{6},\quad \frac{1}{6}\right\}$

さらに、確率分布は離散確率分布と連続確率分布に区分することができる。まず、離散確率分布とは、アンケートなどで男性＝１、女性＝２といったように数値そのものには意味がなく、四則演算ができないなどデータを区分するためのデータ（このようなデータを「質的データ」あるいは「離散データ」という）で、確率変数が連続しておらず、離散的である場合の確率分布である。

一方、連続確率分布とは、あるクラスにおける学生の体重、身長など、数値そのものに意味があり、四則演算ができるデータ（このようなデータを「量的データ」あるいは「連続データ」という）で、

確率変数が連続的な場合の確率分布である[1]。そして、連続確率分布をグラフで描いたものが確率密度関数である。確率密度関数は図表１のように多様な形があり得るものの、Ｃのように真ん中に山が来て左右対称の形をしているのが「正規分布」の一般的な形である。

図表１ 正規分布の多様な形

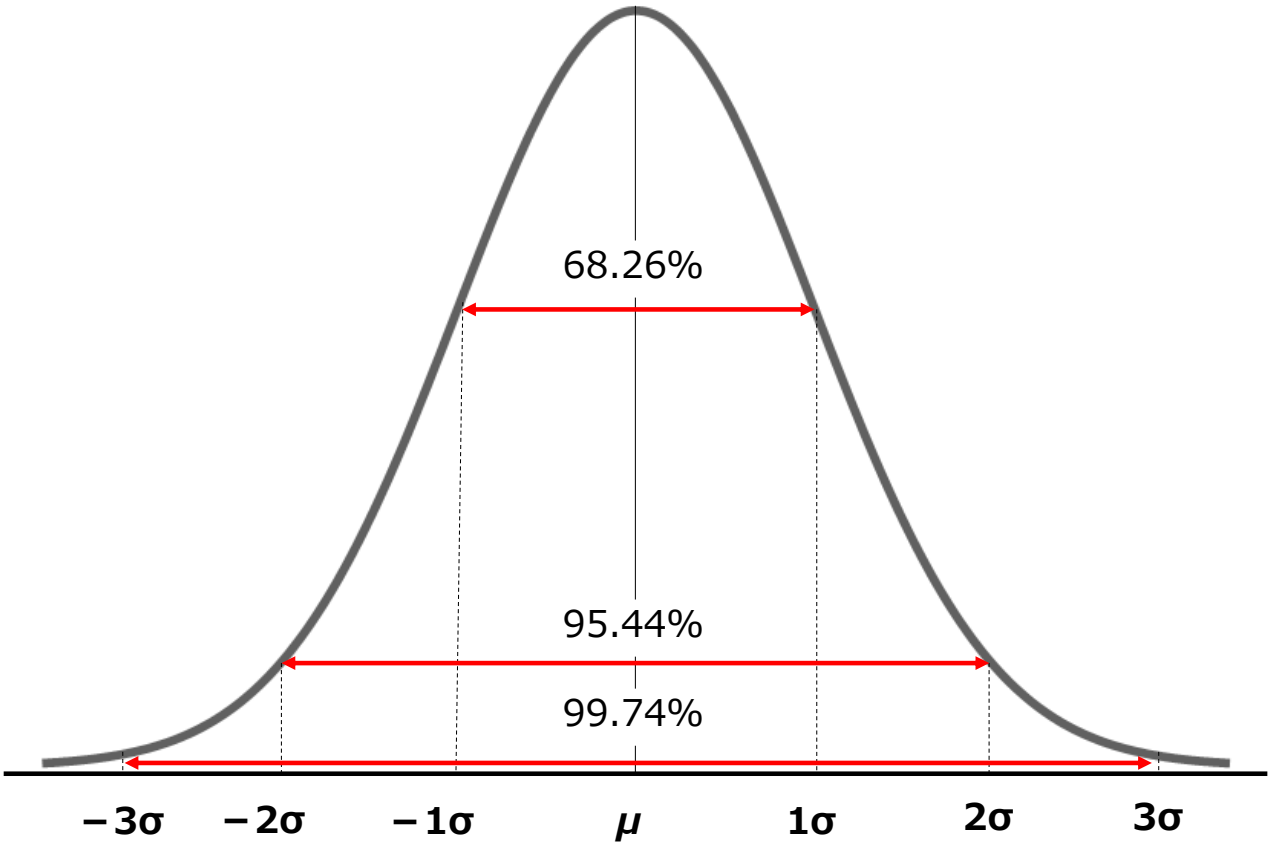


一般的に正規分布は、次のような特徴がある。
①-∞～∞の実数値をとる。
②山が一つで平均値（μ、以下、平均）付近の確率密度が最も大きく、平均と中央値、最頻値が一致する。
③平均を中心として左右対称の釣鐘型の分布である。
④平均から離れるほど、確率密度が小さくなる。
⑤正規分布のカーブの下の面積は形にかかわらず、どれも”１”になっており、分布のカーブの下の面積は確率を示している。

つまり、
⇒　平均から左右に標準偏差１つ分（平均（μ）±標準偏差（σ）×1）の区間にデータが入る確率は68.26%　　$P(\mu-\sigma \le x \le \mu+\sigma)=0.6826$
⇒　平均から左右に標準偏差２つ分（平均（μ）±標準偏差（σ）×2）の区間にデータが入る確率は95.44%　　$P(\mu-2\sigma \le x \le \mu+2\sigma)=0.9544$
⇒　平均から左右に標準偏差３つ分（平均（μ）±標準偏差（σ）×3）の区間にデータが入る確率は99.74%　　$P(\mu-3\sigma \le x \le \mu+3\sigma)=0.9974$
である（図表２）。

[1] 変数（variables、変量とも言う）とは、調査対象により異なり、ある調査を行って得られた結果（データ）に名前を付けたものである。また、確率変数とは、標本空間にある全ての要素を実数に対応させたものだと言える。

図表 2 正規分布の特徴



例えば東京都の中学校 1 年生男子が 10 万人いて、彼らの身長の平均が 160cm、標準偏差が 5.0cm だと仮定しよう。すると、平均から左右に標準偏差 1 つ分の区間、つまり、身長が 155cm から 165cm の間に 68,260 人（10 万人×0.6826）が含まれていることが推測できる。

→ 160cm−5.0cm≦ x ≦160cm+5.0cm

→ 155cm ≦ x ≦ 165cm

確率変数 X が、平均μ、分散σ^2の正規分布に従うとき（式（3））、その確率密度関数は式（4）のようになる。

式（3） $X \sim N(u, \sigma^2)$

式（4） $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$

π=円周率=3.14159…

e=ネイピアの数＝2.71828…（eは、自然対数の底として使われる定数）

式（4）を見ると、かなり難しい式のように見えるものの、πは 3.14159…、e は 2.71828…という

値がすでに決まっているので、平均（μ）と標準偏差（σ）さえ分かれば正規分布の形が決まることになる。つまり、平均は確率密度関数のグラフの位置を決め、標準偏差はグラフの形を決定する。標準偏差が小さいと、平均付近にデータが集まり、標準偏差が大きいと、データが平均から大きく離れることになる。

標準化と標準正規分布

平均と標準偏差により決まる正規分布は世の中に数多く存在し、その形も確率変数により異なるため、世の中のすべての正規分布を分布表として用いることはできない。そこで、ある確率変数のデータが正規分布に従う（$X \sim N(u,\sigma^2)$と仮定できる場合、このデータを標準化した「標準正規分布表」を用いて一定区間の確率（面積）を求める方法が利用されている。

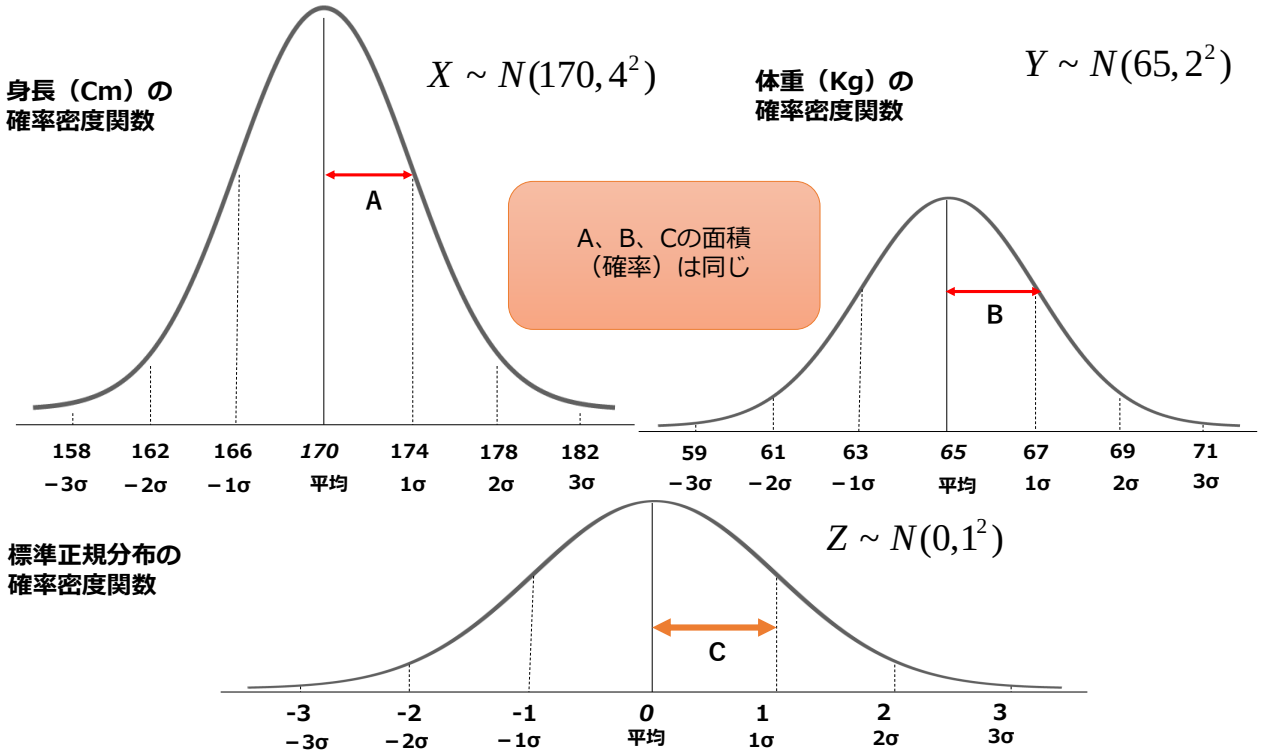
標準化とは、世の中の無数の確率変数が同じ平均と標準偏差を持つように確率変数を変換することである。確率変数Xを標準化するには、該当する値（x_i）から平均（μ）を引き、標準偏差（σ）で割ればよい。すると、確率変数は標準化確率変数に変わり、確率変数の単位に関係なく平均0、標準偏差1の値を持つことになる。標準化したZ_iは、ある値x_iが平均から離れた距離が標準偏差の何倍であるかを意味する。

ある値x_iが平均から離れた距離

式（5）　正規分布　$X \sim N(u,\sigma^2)$　⇒　$Z_i = \dfrac{x_i - \mu}{\sigma}$　⇒　標準正規分布　$Z \sim N(0,1^2)$

確率密度関数の全体の面積は常に1であり、身長でも体重でも平均（μ）±1×標準偏差（σ），平均（μ）±2×標準偏差（σ）、平均（μ）±3×標準偏差（σ）の面積は同じである。従って、身長や体重のように単位が異なっても、標準化して標準正規分布表を利用すると、一定区間の確率（面積）を求めることができる。

図表３ 身長（$X \sim N(170,4^2)$）、体重（$Y \sim N(65,2^2)$）、標準正規分布（$Z \sim N(0,1^2)$）の確率密度関数

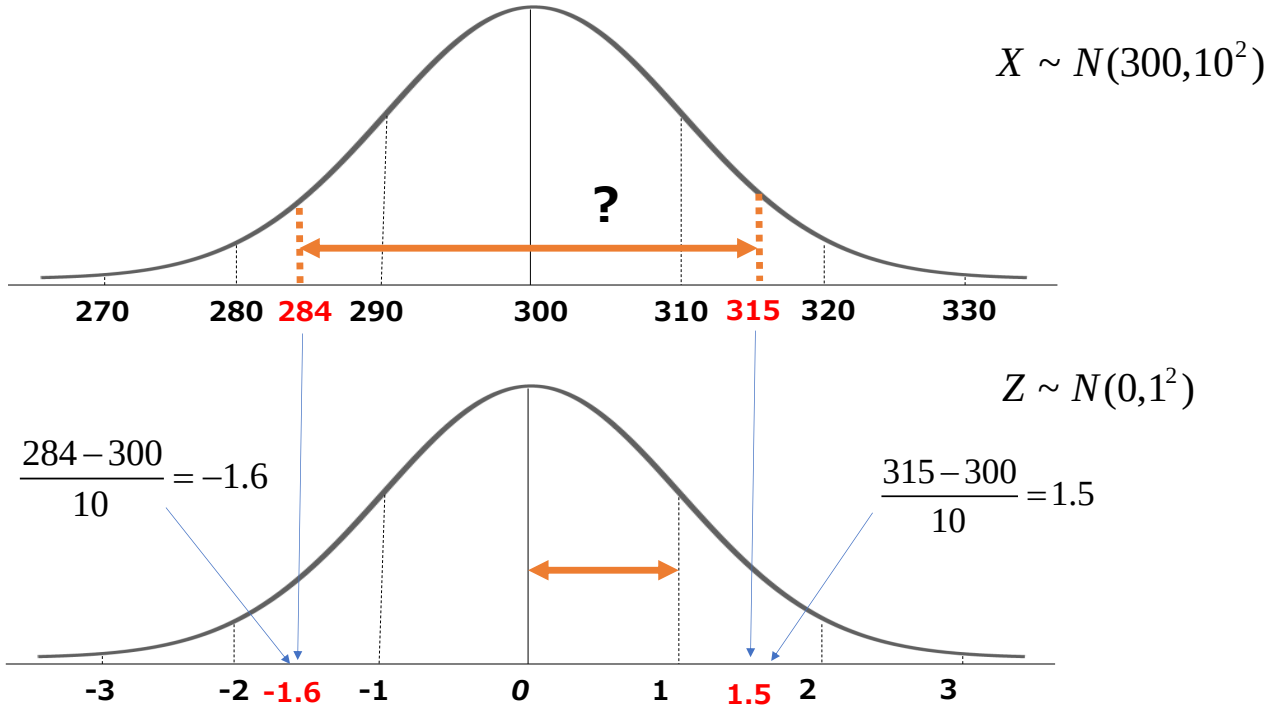


例えばある工場で生産される「さば 缶詰」の内容量が平均 300 g で、標準偏差は 10 g の正規分布を従うと仮定しよう。そこで、ある日この工場で生産された「さば 缶詰」をランダムに抽出し、その内容量が 284 g 以上 315g 以下である確率を求めたい時には、与えられた数値を「標準化」し、標準正規分布表を利用し、確率（面積）を求めることができる。

まず、式（5）を利用して 284 g と 315g を標準化すると、標準化した値はそれぞれ−1.6 と 1.5 になる。つまり、確率変数 X が 284 g から 315g の間に入る確率と、標準正規分布の変数である Z が−1.6 と 1.5 の間に入る確率は同じである（図表 4）。また、標準正規分布の確率密度関数は左右対称であるので、Z の値が 0 から−1.6 の間に入る確率は、Z の値が 0 から 1.6 の間に入る確率を標準正規分布表から確認すればよい（式（6））。そこで、標準正規分布表を利用してその確率を求めると、確率変数 X が 284 g から 315g の間に入る確率は、87.84%（0.4452+0.4332=0.8784）であることが分かる（図表 5）。

式（6）
$$P(284 \le X \le 315) = P(-1.6 \le Z \le 1.5)$$
$$= P(0 \le Z \le 1.6) + P(0 \le Z \le 1.5)$$
$$= 0.4452 + 0.4332 = 0.8784$$

図表４　標準正規分布表で確認した確率変数Ｘが 284ｇから 315g の間に入る確率



図表５ 標準正規分布表で確認した確率変数 X が 284ｇから 315g の間に入る確率

Z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505

本稿の内容が正規分布、標準化、標準正規分布の概念を理解するにおいて少しでも参考になれば幸いである。

付表　標準正規分布表

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

