

第 5 回 正規分布の特徴

A. 正規分布

1. 正規分布の特徴

- ① 母平均 μ と母分散 σ^2 を与えると形が決まる．これを $N(\mu, \sigma^2)$ と書いて表現する．
- ② 平均 μ を中心にして左右対称である．よって，平均より大きい値あるいは小さい値を取る確率はそれぞれ（ ， ）である．
- ③ 曲線は平均 μ の近傍で高く，両側に行くにしたがって単調に低くなる．
- ④ 平均 μ は曲線の位置を決める．平均 μ の異なる 2 つの曲線は左右に移動させれば重ねることができる（図 2）．
- ⑤ 標準偏差 σ は曲線の形を決める． σ が大きければ曲線は扁平になる（図 3）．
- ⑥ (a) $\mu - \sigma$ と $\mu + \sigma$ の間の確率変数を取る確率は約 0.683 である（図 4）．
 (b) $\mu - 2\sigma$ と $\mu + 2\sigma$ の間の確率変数を取る確率は約 0.954 である．
 (c) $\mu - 3\sigma$ と $\mu + 3\sigma$ の間の確率変数を取る確率は約 0.997 である．
- ⑦ 0.95（95%）の確率で $\mu - 1.96\sigma$ と $\mu + 1.96\sigma$ の間の確率変数を取る
 0.99（99%）の確率で $\mu - 2.576\sigma$ と $\mu + 2.576\sigma$ の間の確率変数を取る

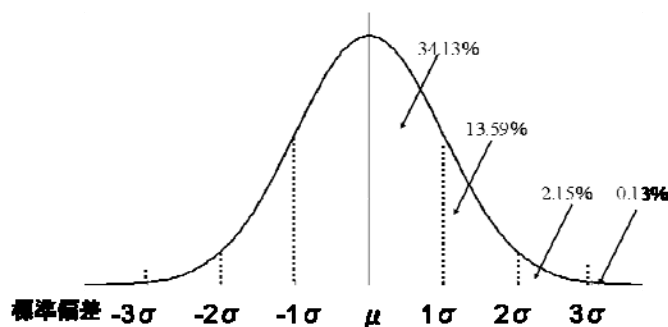


図 1 正規分布

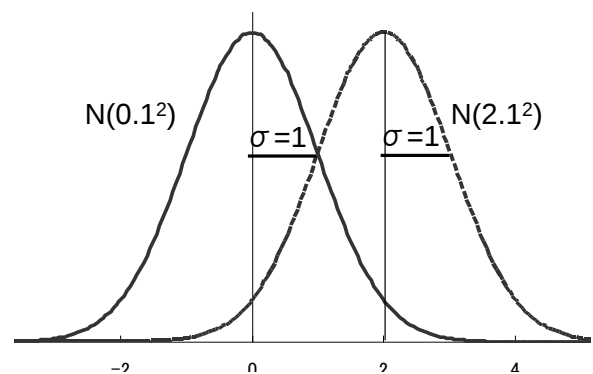


図 2 正規分布 $N(\mu, 1^2)$ の確率密度関数

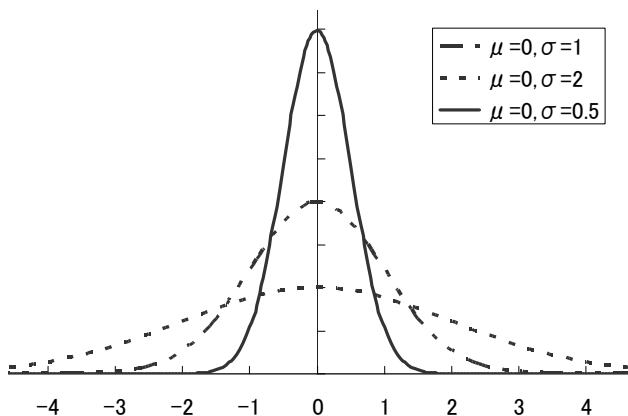


図 3 正規分布 $N(0, \sigma^2)$ の確率密度関数

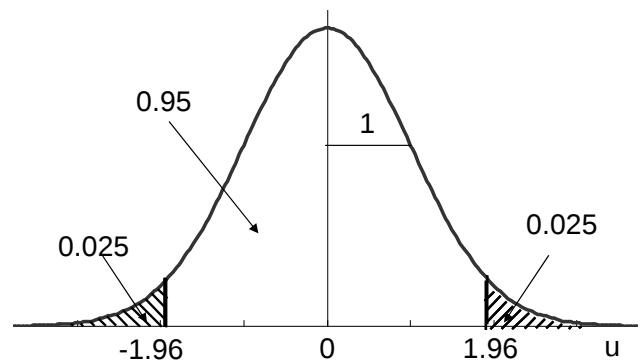
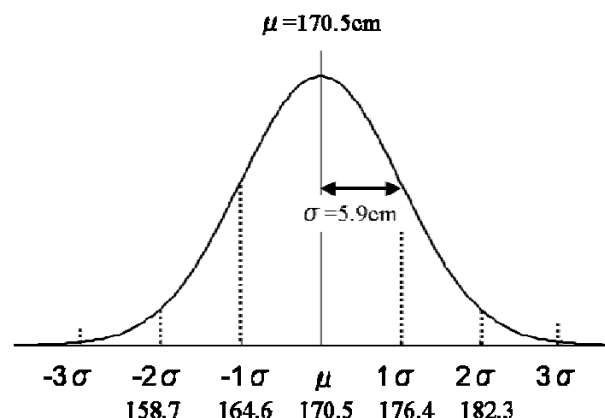


図 4 u の分布, $N(0, 1^2)$

正規分布の例：20～24 歳の男性の身長は人間生活工学研究センターの調査（1992-1994）によると平均 170.5cm，標準偏差 5.9cm であった。身長の分布が正規分布するなら， 2σ 以上平均より背の高い人，すなわち 182.3cm 以上は全体の 2.28%である。平均から標準偏差以内，すなわち 164.6～176.4cm に全体の約 68%が属する。全体の 95%は 158.9～182.1cm に属する。



練習：30 歳代の男性の身長の平均は 169.5cm，標準偏差は 5.8cm であった。身長の分布が正規分布するなら，平均から標準偏差以内，すなわち（ ）～（ ）cm に全体の約（ ）%が属する。 2σ 以下平均より背の低い人，すなわち（ ）cm 以下は全体の（ ）%である。全体の 95%は（ ）～（ ）cm に属する。

2. 正規分布において任意の値と任意の値の間の範囲をとる確率をエクセルから計算する方法

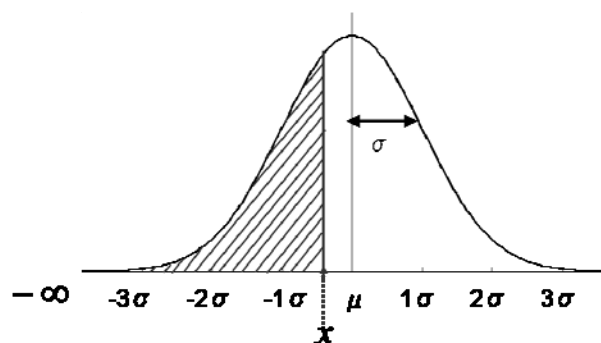
連続分布であるから，正規分布において任意の値を取る確率は 0 である。任意の値と任意の値の間の範囲を取る確率を計算するにはエクセルの関数を利用するのが簡単である。

エクセルの正規分布に関する関数はいくつかある。

今回，利用するのは NORMDIST 関数である。

NORMDIST 関数は平均 μ ，標準偏差 σ の正規分布において， $-\infty$ （無限大）から x までの値を取る確率を以下のように入力することで計算する。

$$= \text{NORMDIST}(x, \mu, \sigma, \text{true})$$



例えば，前述の身長に関する正規分布の例（平均 170.5cm，標準偏差 5.9cm）で，164.6cm 以下の身長の割合は以下の式で求められる。

<code>=NORMDIST(164.6,170.5,5.9,TRUE)</code>	0.158655
--	----------

あるいは平均，標準偏差さえ入力すれば計算するように作られたセルを利用してもよい（なおできるだけエクセルの関数に習熟するためには自分で数式を入れてみる方がよい）。

1. ある値以下となる確率を計算したいとき				平均	170.5
				標準偏差	5.9
平均				ある値	164.6
標準偏差					
ある値				確率	0.158655
確率		#NUM!			

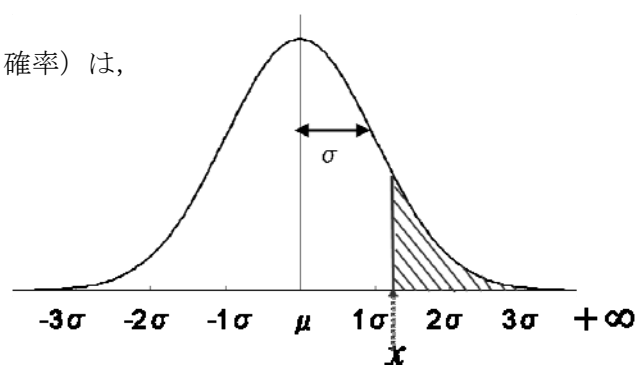
練習 20 歳代の男性の身長は平均は 170.5cm，標準偏差は 5.9cm であった。身長が正規分布するならば，160 cm 以下には全体の約（ ）%が属する。175 cm 以下には全体の約（ ）%が属する。

エクセルの関数

160cm 以下

175cm 以下

ある値より大きくなる確率を計算するには，正規分布全体の確率は 1 となることから，下の図のように考えて，1 から下の図の斜線部分の確率を引き算すると
 $= \text{NORMDIST}(x, \mu, \sigma, \text{true})$ であるから，
 すなわち，斜線部に属する確率（ x より大きくなる確率）は，
 $= 1 - \text{NORMDIST}(x, \mu, \sigma, \text{true})$
 として，計算する。



あるいは平均，標準偏差さえ入力すれば計算するように作られたセルを利用してもよい。

2. ある値以上となる確率を計算したいとき			
平均			
標準偏差			
ある値			
確率		#NUM!	

練習 20 歳代の男性の身長は平均は 170.5cm, 標準偏差は 5.9cm であった. 身長が正規分布するなら, 173 cm 以上には全体の約 () %が属する. 162 cm 以上には全体の約 () %が属する.

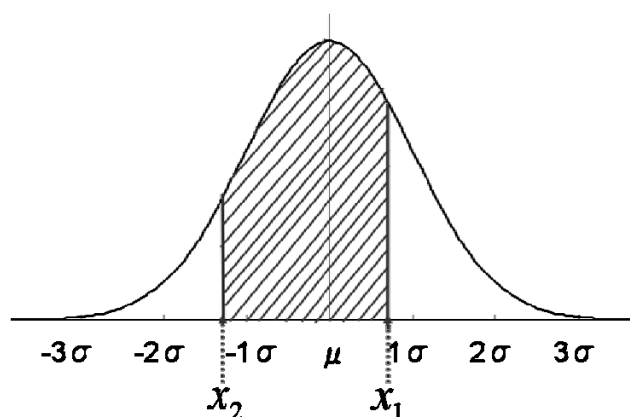
エクセルの関数

173cm 以上

162cm 以上

ある値 (x_2) からある値 (x_1) をとる確率を計算するには, $-\infty$ から x_1 までを取る確率から $-\infty$ から x_2 までを取る確率の差を取る. すなわち下の図のように計算する. エクセルでは

$$= \text{NORMDIST}(x_1, \mu, \sigma, \text{true}) - \text{NORMDIST}(x_2, \mu, \sigma, \text{true})$$
 として, 計算する.



3. AからBの間となる確率を計算したいとき

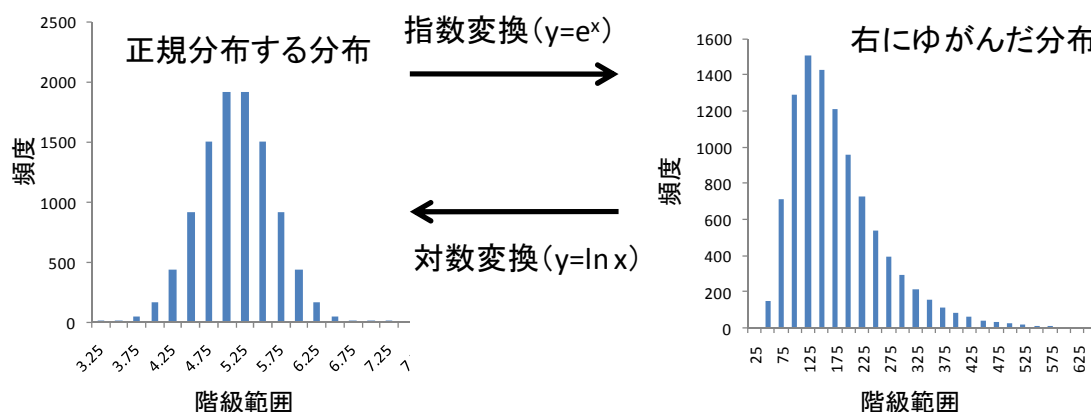
平均			
標準偏差			
A			
B			
確率	#NUM!		

練習 20 歳代の男性の身長は平均は 170.5cm, 標準偏差は 5.9cm であった. 身長が正規分布するなら, 身長が 160~175cm の間にある人は全体の約 () %である.

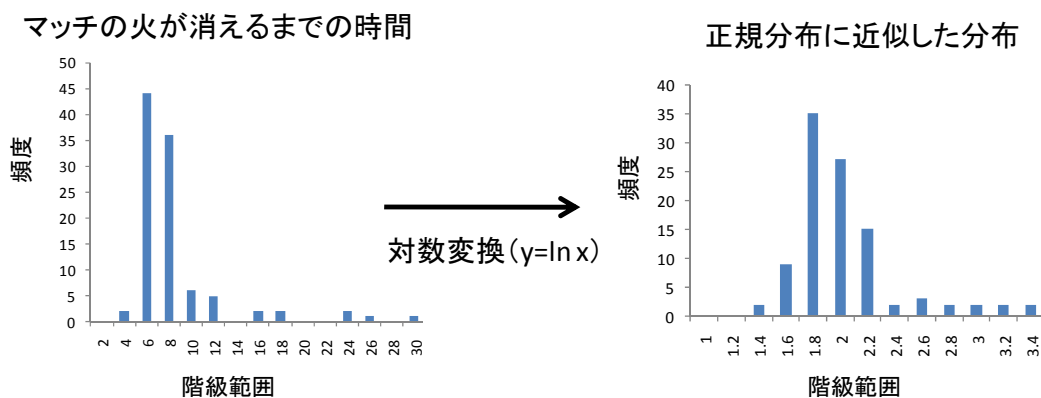
エクセルの関数

B. 正規分布への変換

母集団が正規分布しない場合でも、数値を変換することによって正規分布に近似できることがよくある。正規分布する分布について、その個々の数値を指数変換 ($y=e^x$) すると右にゆがんだ分布を得ることができ、このような右にゆがんだ分布は年収、貯蓄などの分布にみられる。したがって、年収のように右にゆがんだ分布では、対数変換してやることによって、正規分布に近似できることがある。



例：マッチの火が消えるまでの時間を測定した結果のヒストグラム（左）であり、ここから対数変換すると正規分布に近似できそうであると考えた。各測定値を対数変換した結果、右のヒストグラムを得た。ほぼ正規分布しているとみなし、この対数変換後の値について平均、標準偏差を計算したところ、平均=1.92、標準偏差=0.39 となった。



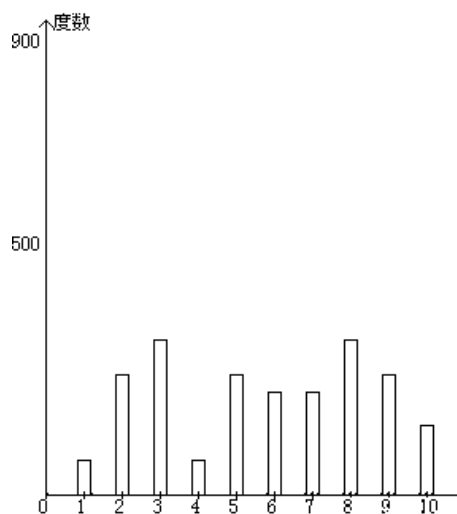
以上の結果を利用すると、マッチの火が 10 秒以内に消える割合は、10 を対数変換し、2.30 を得る。ここからその割合は 84% であると計算できる。

平均、標準偏差、指定された範囲を入力するとその範囲をとる確率を計算するように以下のシートは作ってあります。 適当なセルへコピーして使用してください。				
1. ある値以下となる確率を計算したいとき				
平均	1.92			
標準偏差	0.39			
ある値	2.302585			
確率	0.836701			

C. 標本平均の分布（中心極限定理）

1. 母集団から無作為に取り出した標本の性質（サンプリング変動）

標本を 5 つ無作為に取るという試行を 2000 回繰り返すと 2000 個のデータからなる標本平均はどんな分布となるだろうか？下のグラフに標本平均の分布を想像して書いてみよう。



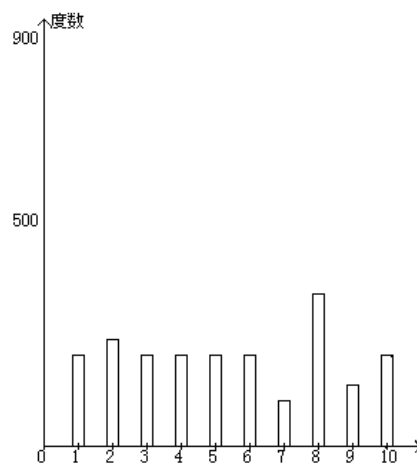
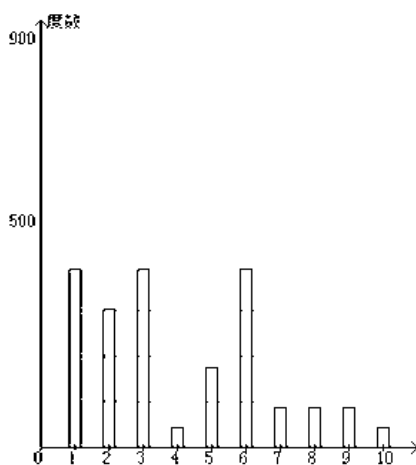
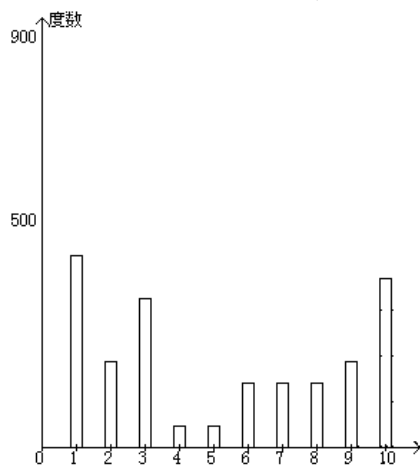
このような標本平均の分布の特徴を考えてみよう。

① 母集団の分布と標本の分布の形はどのように対応するだろうか？

② 標本の平均は母集団の平均に対してどのように位置づけられるだろうか？

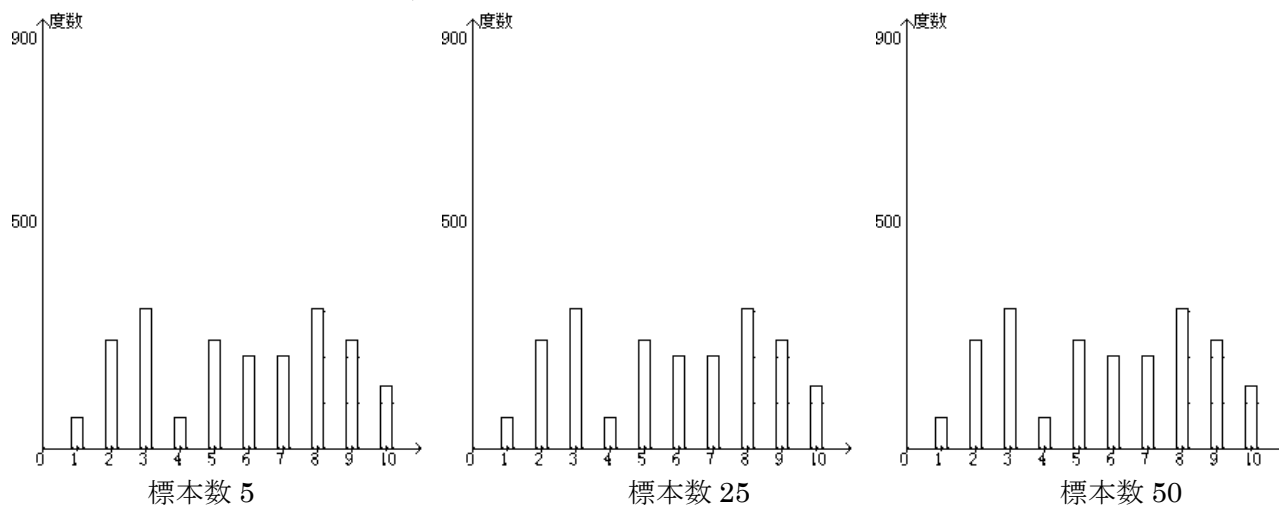
③ 標本の平均はどのくらいばらつくだろうか？

母集団の形と標本平均の分布の形にはどのような関係があるだろうか？下の 3 つの異なる分布をする母集団について、その標本平均がどのように分布するかを書き込んでみよう。



★ 母集団の分布の形にかかわらず、標本平均の分布は（ ）に近似できる。

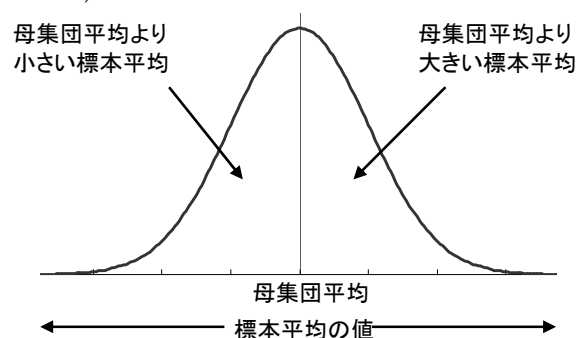
次に標本を取る数を増やしてやるとどうなるだろうか？ すなわち標本を 5 つ無作為にとると
いう試行に代えて、標本を 25, 50 と増やすと標本平均はどのように分布するか？



★ 標本数を増やすと、標本平均の分布の分散が（大きく・小さく）なる。

以上のことから、標本数が十分に多ければ、母集団の分布によらず、標本平均の分布について、次のことがいえる。

- ① 母集団の分布にかかわらず、標本分布は（ ）に近い形となった。
- ② 標本の平均は母集団の平均とほぼ（ ）。
- ③ 標本平均の分布は、標本数が（多い・少ない）と大きくばらつき、標本数が（多い・少ない）とあまりばらつかなかった。



わたしたちは統計的な調査をするときに、個々のデータそのものに関心を持つことはあまりない。さらに（母集団 ・ 標本）ではなく、（母集団 ・ 標本）に関心を持つ。そして（母集団 ・ 標本）の（平均 ・ ばらつき ・ ヒストグラム）に関心を持つことが一般的である。

前述したように正規分布は母集団が正規分布に近似しない場合でも、標本数を十分にとれば、その標本平均に関する分布は正規分布に近似できるという特徴がある。わたしたちは母集団の標本に一般的に関心を持つために、正規分布はたとえ母集団そのものが正規分布していなくても、有用性が大きいことがわかる。

D. 宿題

宿題はホームページ <http://www.ipc.shimane-u.ac.jp/food/kobayasi/biometricsyukudai.html> を見てください。