確率と確率変数

サイコロを投げ，

　３か６の目が出たら出欠を取る

　１，２，４，５の目が出たら出欠は取らない

サイコロの目の全ての出方は

　$\Omega = \{1, 2, 3, 4, 5, 6\}$ ← 標本空間 (sample space)

出欠を取るという事象　　：E_1　$= \{3, 6\}$

出欠を取らないという事象：E_2　$= \{1, 2, 4, 5\}$

$$\Omega = E_1 + E_2 = E_1 + \overline{E_1} = \overline{E_2} + E_2$$

E_1 (E_2) は E_2 (E_1) の余事象

E_1とE_2は排反

サイコロの目の出方が同様に確からしければ

$P(E_1) = 2/6$　　: {3, 6}

$P(E_2) = 4/6$　　: {1, 2, 4, 5}

$P(E_1 + E_2) = 1$　　: {1, 2, 3, 4, 5, 6}

これより

$P(\Omega) = 1$　　$0 \leqq P(E_i) \leqq 1$

すなわち

全事象の確率は1 ，　事象 E_i の起こる確率は0と1の間

確率が 1 $\Leftrightarrow$ 必ずその事象は起こる

確率が 0 $\Leftrightarrow$ その事象は決して起こらない

確率の加法定理

E_1 か E_2 のいずれかが起こる確率 $P(E_1 \cup E_2)$ は,

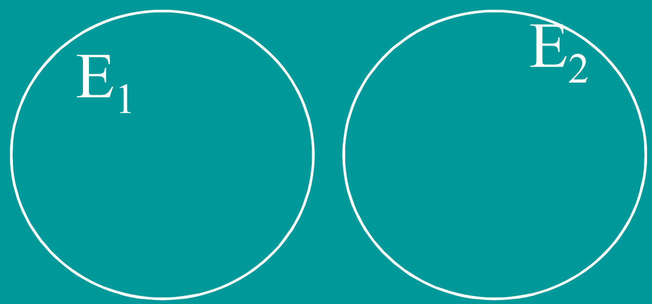
E_1 と E_2 が排反事象のとき,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

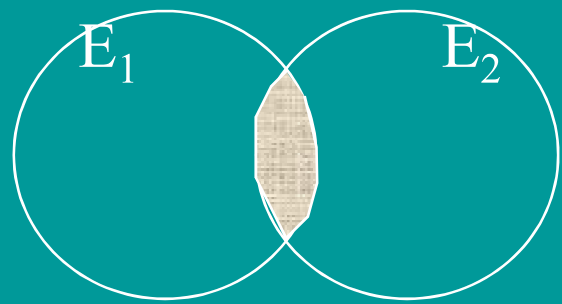
E_1 と E_2 が 排反でないとき ,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

排反事象の場合



排反事象でない場合



出欠の場合

出欠を取る回数をXで表すと,

$X=0 \quad (E_2 \cap E_2)$ ： 1通り

$X=1 \quad (E_1 \cap E_2)$, $(E_2 \cap E_1)$ ： 2通り

$X=2 \quad (E_1 \cap E_1)$ ： 1通り

2回のサイコロの結果は互いに独立である ⇒ 独立事象

確率の乗法定理

記号を次のように改める,

E_{11} :　１回目の講義で出欠を取る

E_{12} :　１回目の講義で出欠を取らない

E_{21} :　２回目の講義で出欠を取る

E_{22} :　２回目の講義で出欠を取らない

E_{ij}　i : 回目

　　j : 出欠

このとき, X＝0 となるためには,

　２回とも出欠を取らないという事象が起きる必要がある.

その確率は, E_{12} の確率$P(E_{12})$と,

E_{12} が起こったという条件の下で E_{22} が起こる確率 $P(E_{22} | E_{12})$ の積である.

$$P(E_{12} \cap E_{22}) = P(E_{12}) \cdot P(E_{22} | E_{12})$$

E_{12} と E_{22} が独立事象の場合,

$$P(E_{12} \cap E_{22}) = P(E_{12}) \cdot P(E_{22})$$

出欠の確率

X＝0 となる確率 $P(E_{12} \cap E_{22})$ は,

$P(X=0) = (4/6)\times(4/6) = 16/36 \fallingdotseq 0.44$ ・・・ ①

X＝1 となる確率 $P(E_{11} \cap E_{22})+(E_{12} \cap E_{21})$ は,

$P(X=1) = (2/6)\times(4/6)+(4/6)\times(2/6) = 16/36 \fallingdotseq 0.44$ ・・・ ②

X＝2 となる確率 $P(E_{11} \cap E_{21})$ は,

$P(X=2) = (2/6)\times(2/6) = 4/36 \fallingdotseq 0.11$ ・・・ ③

これより，以下のことが分かる

2回続けて出欠を取る確率は1割強

1回は出欠を取る確率は5割弱

なお

① ＋ ② ＋ ③ ＝ 1　　　この3種類の事象が全て

確率変数 (random variable)

試行の結果，変数 x が x_1，x_2，・・・，x_n のいずれかの値をとり，それぞれに対する確率 p_1，p_2，・・・，p_n が得られるとき x を離散型確率変数という.

たとえば，

①１～6の目が出るサイコロ投げの場合，

　１～6の目の出る確率はそれぞれ1/6.

②コインをトスした場合，

　表が出る確率と裏が出る確率はともに1/2.

確率分布 (probability distribution)

確率変数 x のとる値 x_i と, それに対する確率 p_i との対応関係を $\mathrm{P}(x = x_i) = p(x_i)$ と表す

確率変数	x_1	x_2	・・・	x_n	計
確　　率	$p(x_1)$	$p(x_2)$	・・・	$p(x_n)$	1

$$\sum_{i=1}^{n} p_i = 1$$

度数分布と確率分布

確率変数	x_1	x_2	・・・	x_n	計
度　　数	n_1	n_2	・・・	n_n	N
相対度数	n_1 / N	n_2 / N	・・・	n_n / N	1
確　　率	$p(x_1)$	$p(x_2)$	・・・	$p(x_n)$	1

統計量の確率分布は標本分布 (sampling distribution) とも呼ばれる

分布関数

離散型確率変数 X が x という値をとる確率 $p(x)$ が数式 $f(x)$ で表されるとき，
$f(x)$ を(離散型確率)**分布関数**という.

離散型確率分布関数の性質

① $0 \leqq f(x) \leqq 1$

② $\Sigma f(x) = 1$

③ $f(\pm\infty) = 0$

確率の求め方

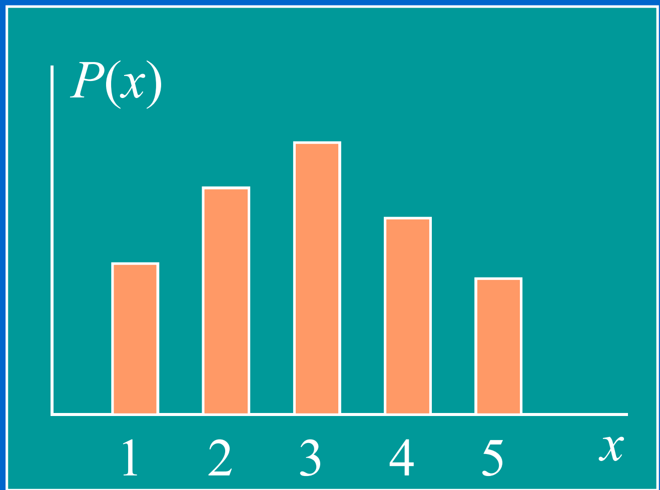
離散型確率分布(二項分布の場合)

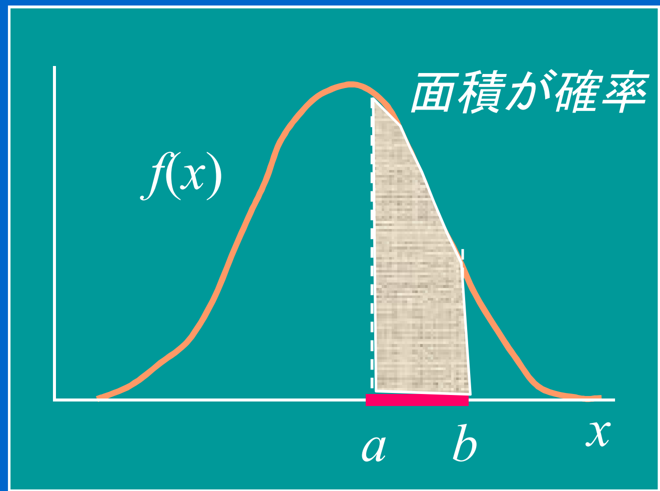
$$P(x) = {}_nC_x\, p^x (1-p)^{n-x}$$

連続型確率分布

$$P(a \leqq x \leqq b) = \int_a^b f(x)\,dx$$

$$\boxed{f(x=a)=0}$$





離散型確率変数の期待値と分散

期待値(平均値)　(expectation , expected value)

$$\mu = E(x) = \Sigma x_i\, p(x_i) = \Sigma x_i p_i$$

分散

$$\sigma^2 = V(x) = E((x - \mu)^2)$$

$$= \Sigma (x_i - \mu)^2\, p(x_i)$$

$$= \Sigma x_i^2\, p(x_i) - \mu^2$$

標準偏差

$$\sigma = \sqrt{\sigma^2}$$

例題

サイコロを1回投げたときの期待値と分散

サイコロの目　x	1	2	3	4	5	6	計
確　　率　$p(x)$	1/6	1/6	1/6	1/6	1/6	1/6	1

【解答】

サイコロの目を x とする

$$\mu = \sum_{i=1}^{6} x_i \times p(x_i) = (1+2+3+4+5+6) \times 1/6 = 3.5 \quad /\!\!/$$

$$\sigma^2 = \Sigma (x_i - \mu)^2 \, p(x_i) = ((1-3.5)^2 + \cdots (6-3.5)^2) \times 1/6$$

$$\fallingdotseq 2.916667 \quad /\!\!/$$

二項分布 (binomial dist.)

$x_i\,(\,i = 1, 2, \cdots, n\,)$ は互いに独立な離散型確率変数で，
x_i は排反的に 0 または 1 をとる　⇒　ベルヌーイ試行 (Bernoulli trial)

すなわち

$$x_i = \begin{cases} 0 & P(\,x_i = 0\,) = p \\ 1 & P(\,x_i = 1\,) = q \end{cases} \qquad p + q = 1$$

結果が2種類

表・裏

勝ち・負け

当たり・はずれ

Yes ・ No

偶数・奇数

1・その他

このとき

$S_n = x_1 + x_2 + \cdots + x_n$　として

$$P(\,S_n = x\,) = f(\,x\,) = {}_n C_x\, p^{\,x} q^{\,n-x}$$

$B(\,n\,,\,p\,)$

例題

ある工場では，製品を1個取り出して，それが良品である確率は0.85である．8個の製品を無作為に取り出したとき以下の確率を求めよ．

①良品が3個である確率．

②良品が7個以上である確率．

【解】

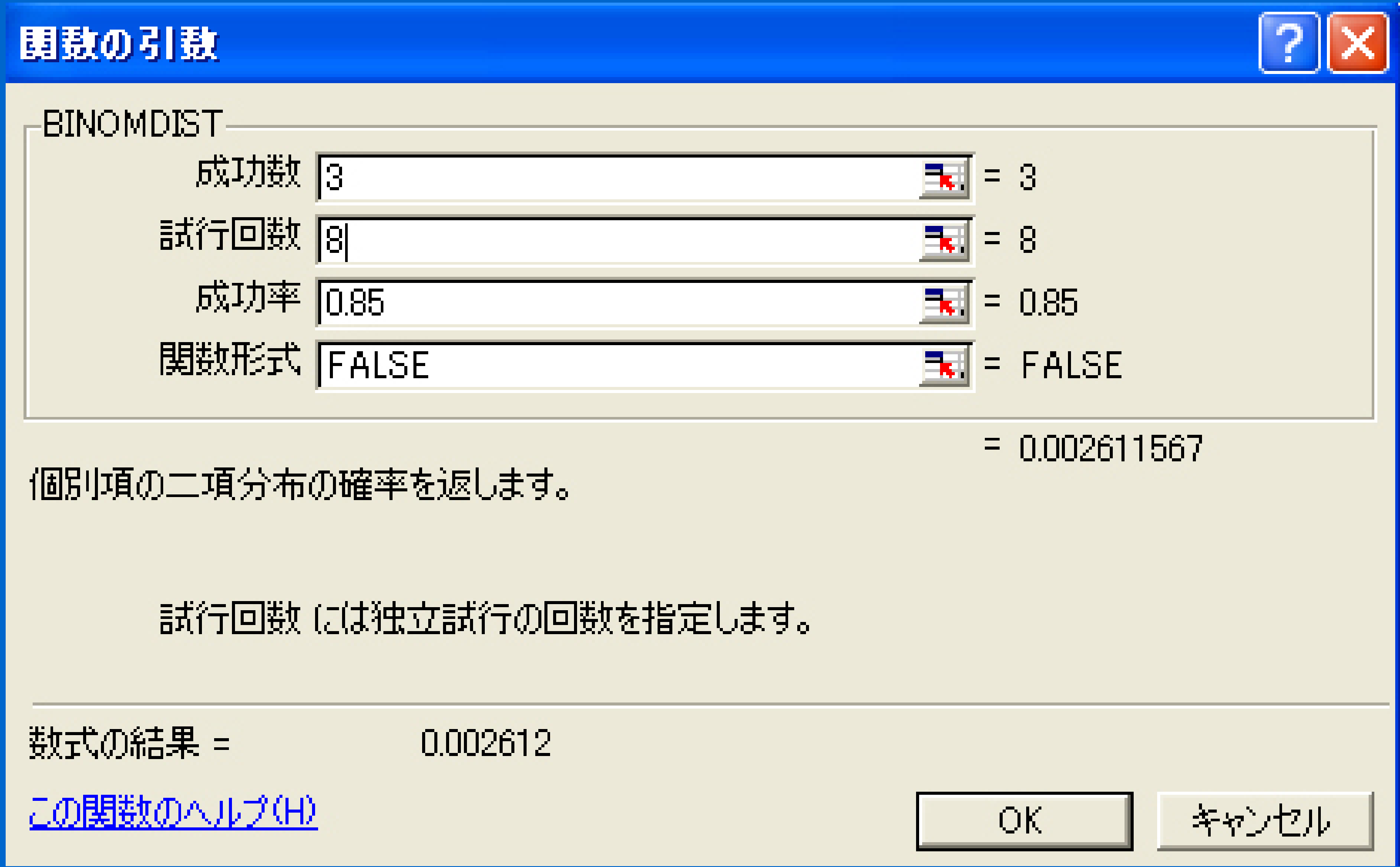
① $P(x=3) = {}_8C_3(0.85)^3(1-0.85)^{8-3} = 0.002612$

② $P(x \geqq 7) = P(x=7) + P(x=8)$

$= {}_8C_7(0.85)^7(1-0.85) + {}_8C_8(0.85)^8(1-0.85)^0$

$= 0.384093 + 0.272491$

$= 0.657183$


関数の引数
BINOMDIST
成功数 3 = 3
試行回数 8 = 8
成功率 0.85 = 0.85
関数形式 FALSE = FALSE
= 0.002611567
個別項の二項分布の確率を返します。
試行回数 には独立試行の回数を指定します。
数式の結果 = 0.002612
この関数のヘルプ(H)
OK
キャンセル

Microsoft Excel - Book1

B2 =BINOMDIST(3,8,0.85,FALSE)

	A	B	C	D
1				
2	①	0.002612	=BINOMDIST(3,8,0.85,FALSE)	
3				
4				
5	②	良品が8個	0.272491	=BINOMDIST(8,8,0.85,FALSE)
6		良品が7個	0.384693	=BINOMDIST(7,8,0.85,FALSE)
7			0.657183	
8				
9		良品が6個	0.237604	=BINOMDIST(6,8,0.85,FALSE)
10		良品が5個	0.083860	=BINOMDIST(5,8,0.85,FALSE)
11		良品が4個	0.018499	=BINOMDIST(4,8,0.85,FALSE)
12		良品が3個	0.002612	=BINOMDIST(3,8,0.85,FALSE)
13		良品が2個	0.000230	=BINOMDIST(2,8,0.85,FALSE)
14		良品が1個	0.000012	=BINOMDIST(1,8,0.85,FALSE)
15		良品が0個	0.000000	=BINOMDIST(0,8,0.85,FALSE)
16			1.000000	

二項分布の期待値と分散

$\mu = E(x) = np \qquad \sigma^2 = V(x) = npq$

$\qquad\qquad\qquad\qquad\qquad = np(1-p)$

標本比率 $\hat{p} = X / n$ を考えると

$E(\hat{p}) = p \qquad D(\hat{p}) = p(1-p)/n$

n が大きくなると ($np \geqq 5$) , X と p の分布は, それぞれ

$N(np, np(1-p))$,　$N(p, p(1-p)/n)$ に近似する.

例題　９．１

10枚のコインを同時に投げて，そのうち4枚表が出る確率.

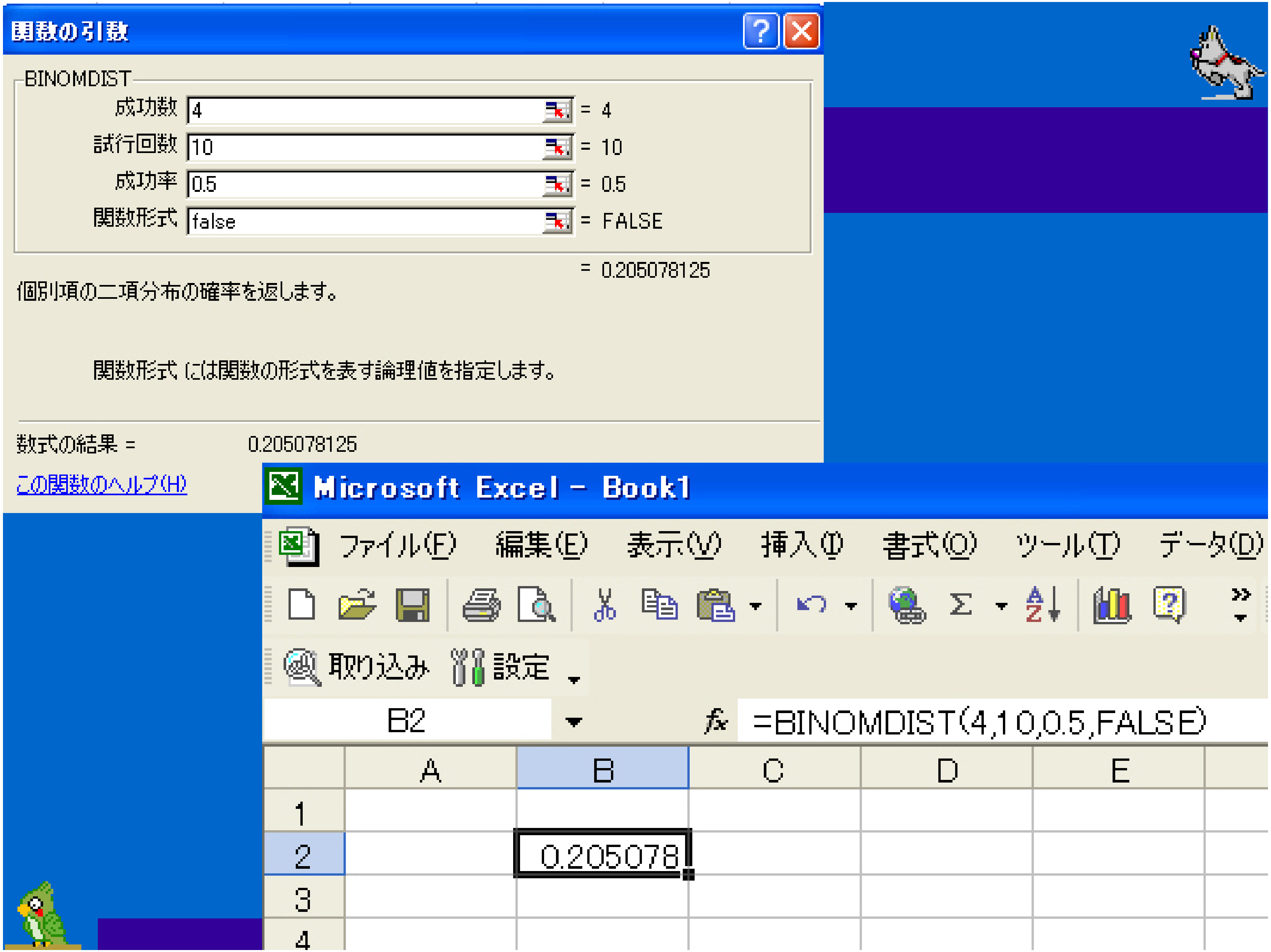
(1枚のコインを10回投げても同じ)

【解答例】

二項分布の公式より,

$$ {}_{10}C_4\,(0.5)^4\,(0.5)^6 = 210 \times (0.5)^{10} = 0.205078 $$

Excleの関数を利用すると,


関数の引数
BINOMDIST
成功数 4 = 4
試行回数 10 = 10
成功率 0.5 = 0.5
関数形式 false = FALSE
= 0.205078125
個別項の二項分布の確率を返します。
関数形式 には関数の形式を表す論理値を指定します。
数式の結果 = 0.205078125
この関数のヘルプ(H)
Microsoft Excel - Book1
ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D)
取り込み 設定
B2
=BINOMDIST(4,10,0.5,FALSE)
A B C D E
1
2 0.205078
3
4

大数の法則 (law of large numbers)

n回試行を繰り返したとき，事象Eがa回起こったとする．
Eの起こる相対度数a/nは，nが大きくなればなるほど，
ある一定値に近づく．
すなわち

$$\overline{x} = 1/n\Sigma x_i f_i = \Sigma x_i (f_i / n) \fallingdotseq \mathrm{E}(X)$$

一般に

$$\lim_{n\to\infty} \overline{x} = \mathrm{E}(X)$$

超幾何分布 (hypergeometric dist.)

N(＝ N_1 ＋ N_2) 個の母集団から,

n個の標本を元に戻さずに(非復元)抽出する.

ただし, N個のうち第１のタイプが N_1 個,

第２のタイプが N_2 個である.

このとき第１のタイプが k 個取り出される確率は,

$$P(k) = \frac{{}_{N1}C_k \times {}_{N2}C_{n-k}}{{}_{N}C_n}$$

例題

男子15名，女子10名から成るグループから，
ランダムに12名を選ぶ．
このとき，男女6名ずつが選ばれる確率は？
【解】

$$p(6) = \frac{{}_{15}C_6 \times {}_{10}C_6}{{}_{25}C_{12}} = 0.202$$ //

幾何分布

$p > 0,\ q > 0$ かつ $p + q = 1$ とし,

離散型確率変数 X に対し,

$x = 0,\ 1,\ 2,\ \cdots,$ のとき

$$P(X = x) = f(x) = pq^x$$

ポアソン分布

離散型確率変数 X に対し，

$x = 0,\ 1,\ 2,\ \cdots,$ のとき

$e \fallingdotseq 2.718\cdots$ とし，

λ を正の定数とすると，

$$P(X = x) = f(x) = \frac{\lambda^{x}}{x\,!}\,e^{-\lambda}$$

連続型分布の期待値と分散

期待値

$$\mu = E(x) = \int_{-\infty}^{\infty} f(x)\,dx$$

分散

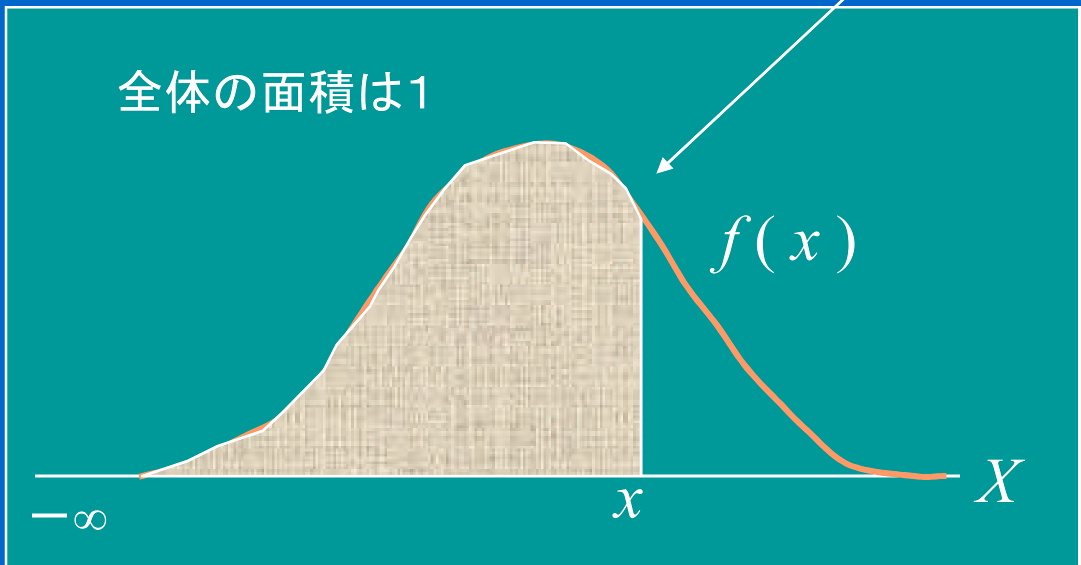
$$\sigma^2 = V(x) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

確率密度関数 (probability density function)

連続型確率変数 X の分布関数を F(x) とし,

$\frac{d}{dx}\mathrm{F}(x) = f(x)$

なる $f(x)$ を確率密度関数という.

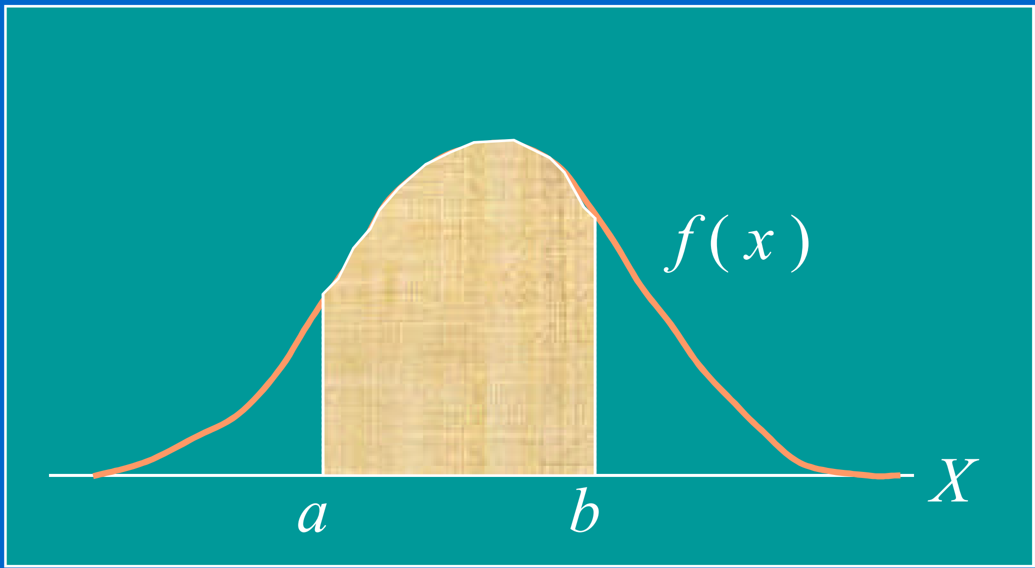


$$P(X \leqq x) = \mathrm{F}(x) = \int_{-\infty}^{x} f(x)\,dx$$

確率密度関数 (cont.)

$$P(a \leqq X \leqq b)$$

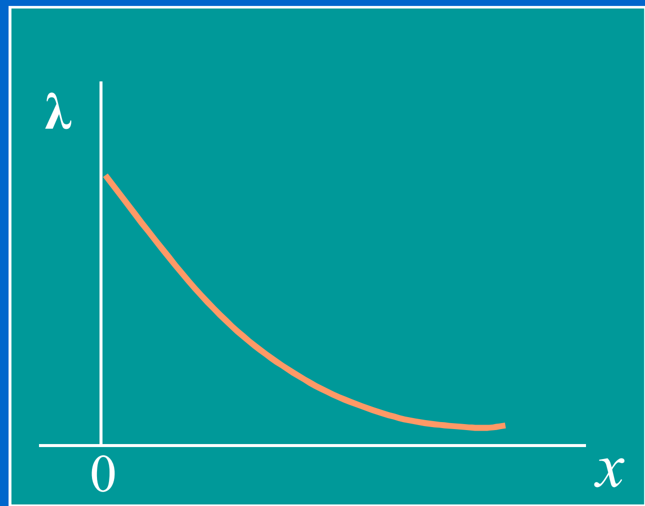
$$= \int_a^b f(x)\,dx$$



指数分布(exponential dist.)

eを自然対数の底とし，λを正の定数としたとき，

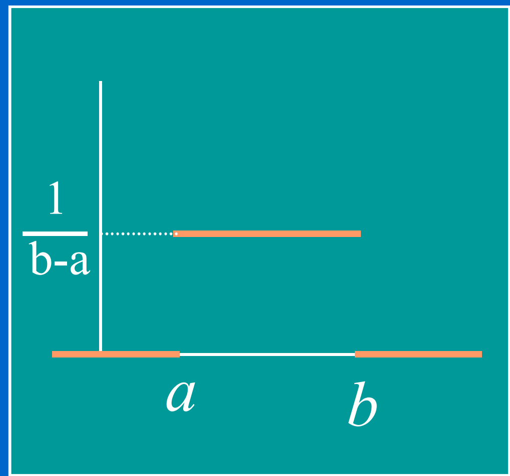
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (x > 0) \\ 0 & (x \leqq 0) \end{cases}$$



一様分布(uniform dist.)

a, bを $a < b$ の定数としたとき,

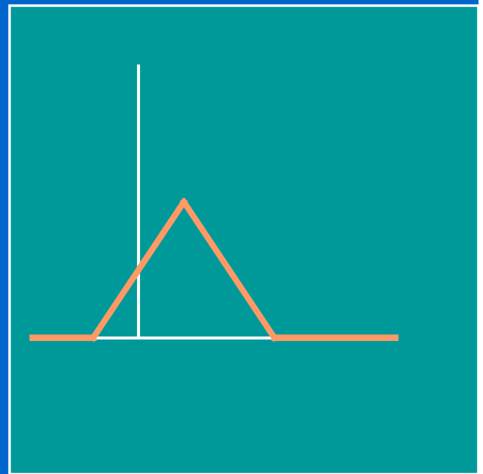
$$f(x) = \begin{cases} 1/(b-a) & (a < x < b) \\ 0 & (\text{その他の区間}) \end{cases}$$



三角分布(triangle dist.)

a, bを $a < 0$, $0 < b$ なる定数とし, $a = -b^2$ とき,

$$f(x) = \begin{cases} a|x| + b & (|x| \leqq -b/a) \\ 0 & (\text{その他の区間}) \end{cases}$$



正規分布(normal dist.)

確率変数 x の確率密度関数が

$$f(x) = \frac{1}{\sqrt{2\pi}\ \sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}} \qquad (-\infty < x < \infty)$$

で与えられるとき，

x は平均μ，分散σ^2の正規分布に従う．

これを，$N(\mu,\ \sigma^2)$ と表す．

正規分布(cont.)

N(μ, σ^2) のμとσが異なることにより,
正規分布は様々な形を呈す.

注意

いずれの場合も,

グラフと横軸 (x軸) で囲まれた部分の面積は1である.

正規分布の性質

① 単峰，左右対称，ベル型

② 平均(μ)は中央，標準偏差(σ)は変曲点

③ $\mu \pm \sigma$の両翼で囲まれた面積は68.26%

すなわち　$P(\mu - \sigma < x < \mu + \sigma) = 0.6826$

同様に　$P(\mu - 2\sigma < x < \mu + 2\sigma) = 0.9544$

さらに　$P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9972$

④ 再生性 (reproductive property)

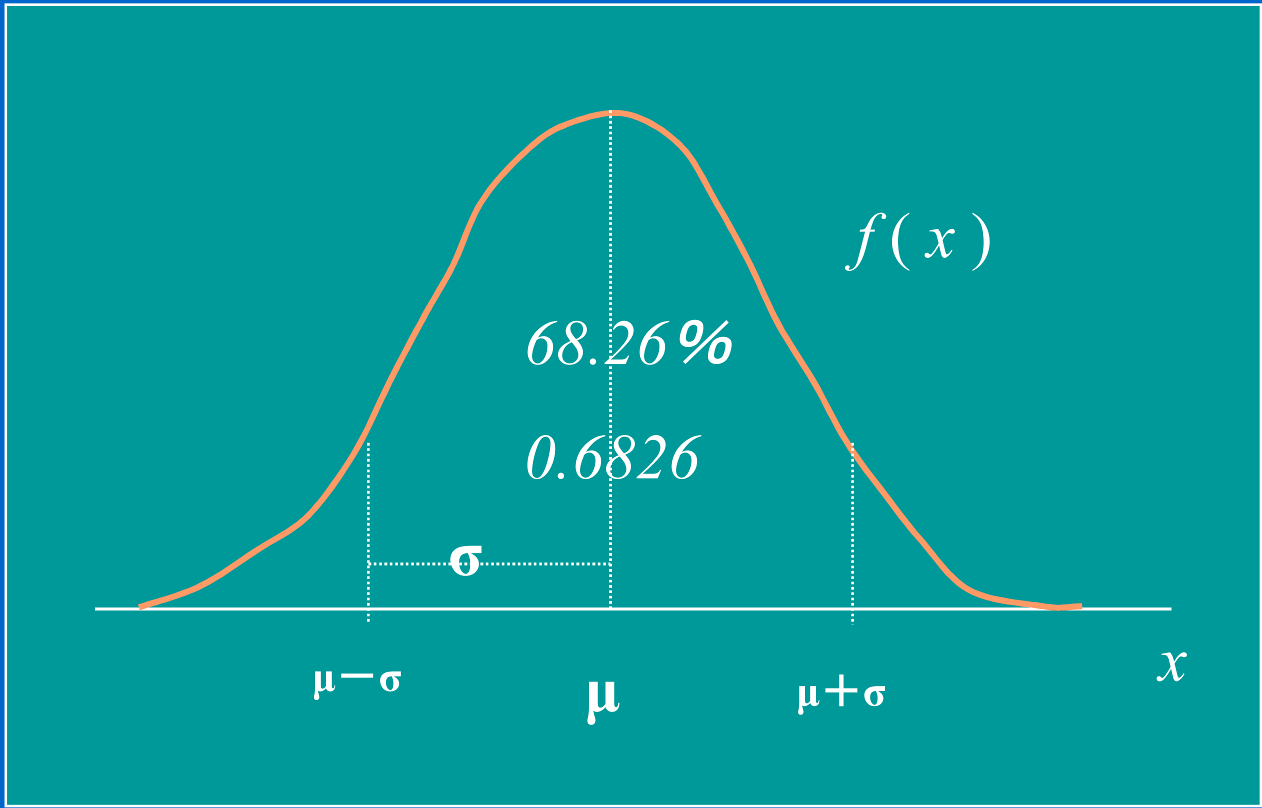
独立な確率変数 x_i ($i = 1,2$) が平均μ_i ($i = 1,2$)，分散σ_i^2 ($i = 1,2$) に従うならば，

和および差は，それぞれ平均$\mu_1 + \mu_2$，分散$\sigma_1^2 + \sigma_2^2$

および平均$\mu_1 - \mu_2$，分散$\sigma_1^2 + \sigma_2^2$に従う．

符号に注意

(μ−σ)と(μ+σ)に囲まれた面積



標準正規分布(standard normal dist.)

正規分布の x を以下の変換によって,
標準(基準)正規分布に変換できる.

$$z = \frac{x - \mu}{\sigma}$$

パラメターが定数

$$N(\mu, \sigma^2) \Rightarrow N(0, 1^2)$$

このグラフは1種類のみである

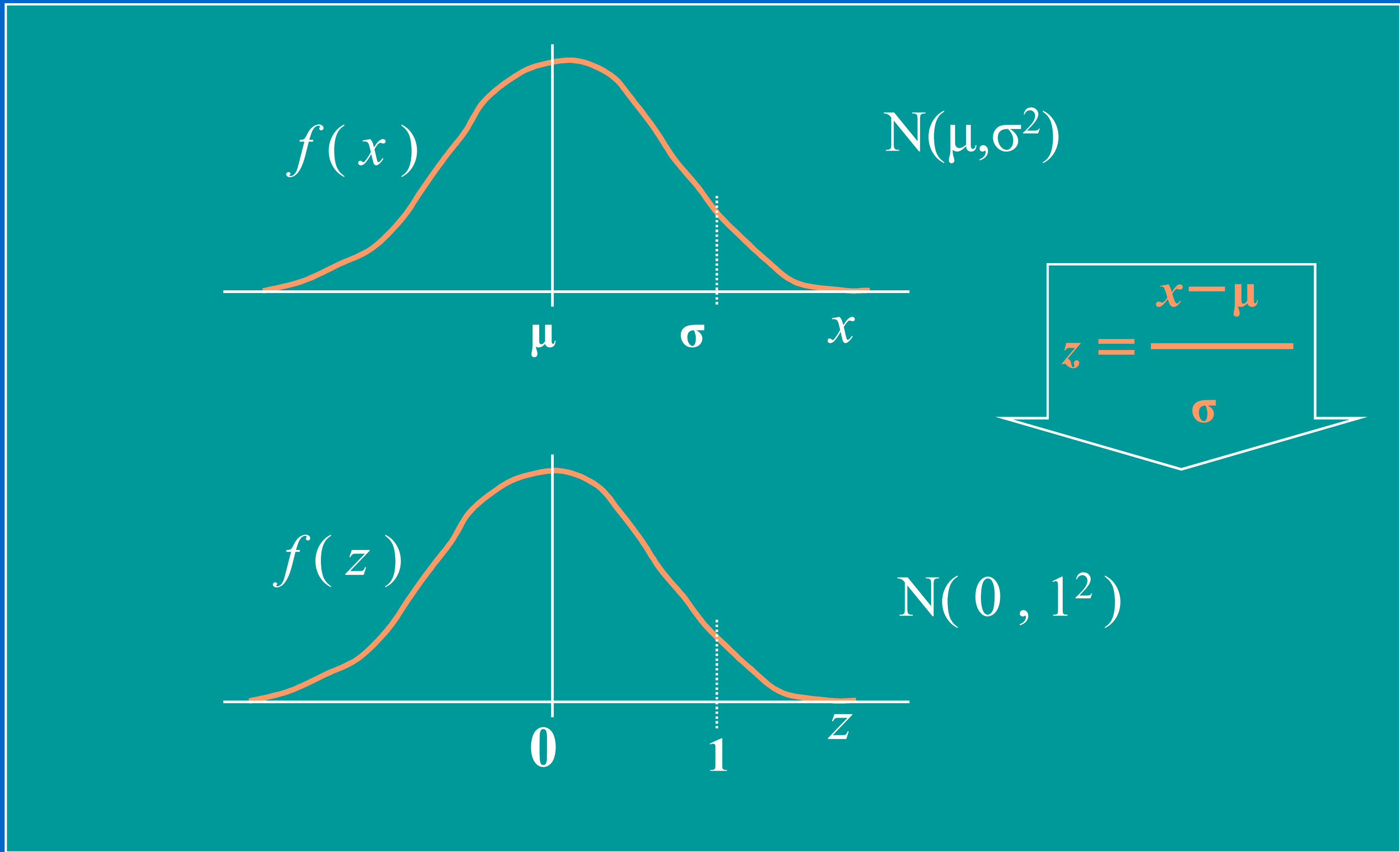
標準正規分布の密度関数

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

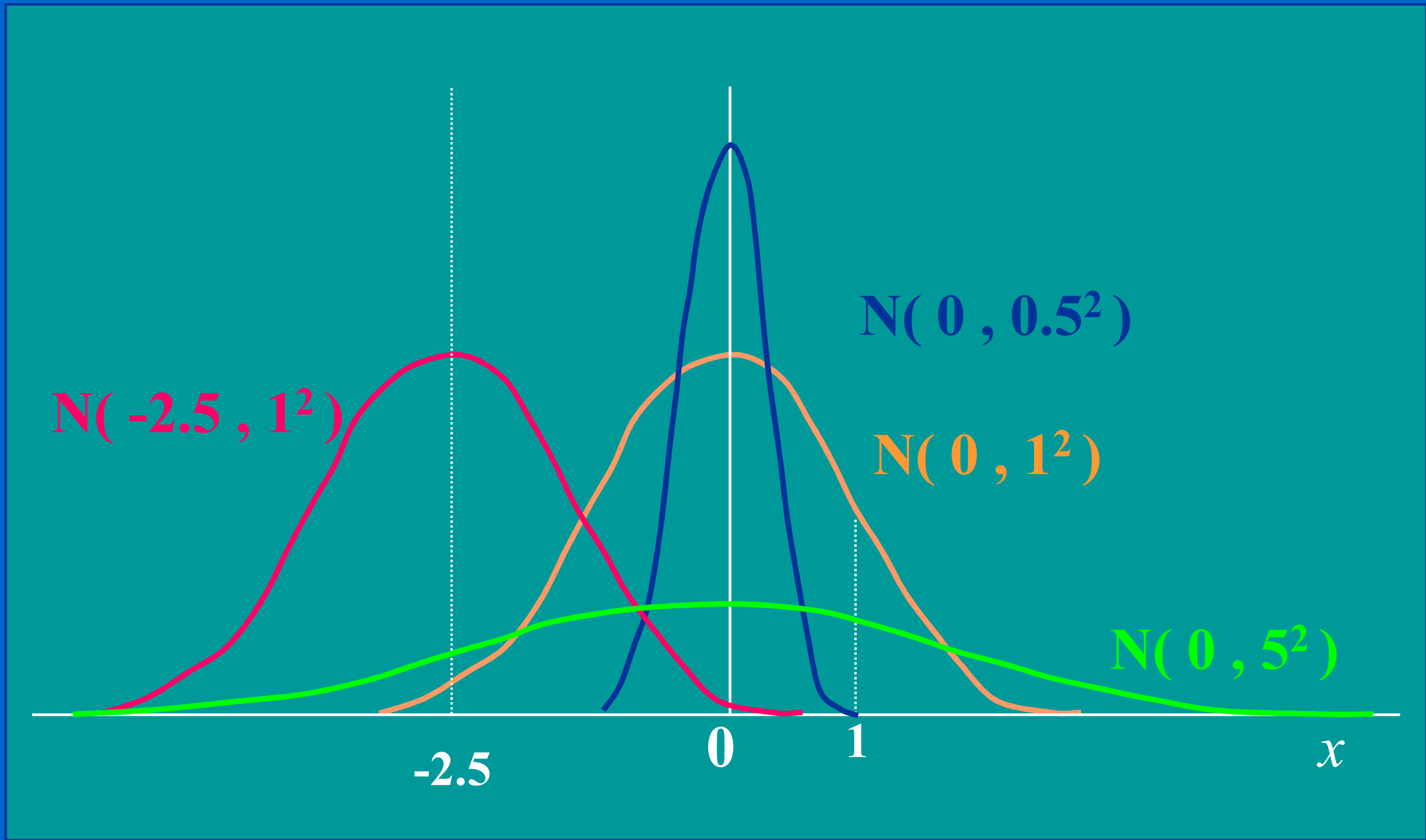
z は平均 0, 分散 1^2 の標準正規分布に従う.

これを, $N(0,\ 1^2)$ と表す.

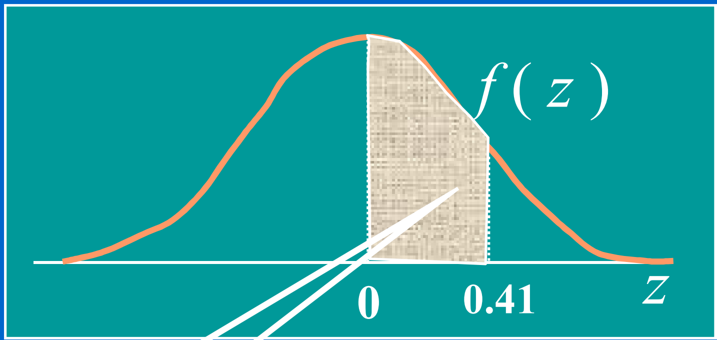
正規分布と標準正規分布



標準正規分布は正規分布の特殊型



標準正規分布表の読み方



15.91％

(0.1591)

$$P(0 \leqq Z \leqq 0.41)$$

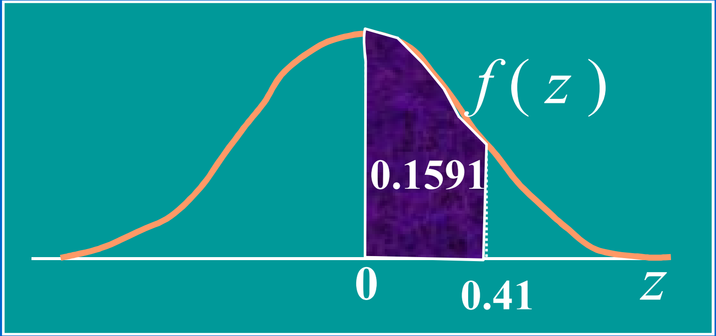
$$= \int_0^{0.41} f(z)\,dz$$

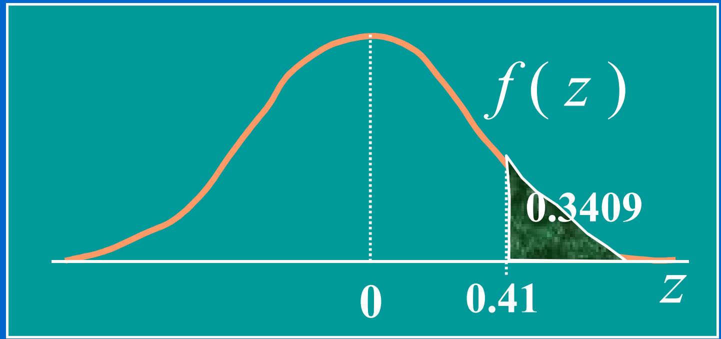
$$= \frac{1}{\sqrt{2\pi}}\, e^{\frac{z^2}{2}} \quad (0 < z < 0.41)$$

標準正規分布表（抜粋）

z	0.00	0.01	0.02	・・・	0.09
0.0	.0000	.0040	.0080	・・・	.0359
0.1	.0398	.0438	.0478	・・・	.0753
0.2	.0793	.0832	.0871	・・・	.1141
0.3	.1179	.1217	.1255	・・・	.1517
0.4	.1554	.1591	.1628	・・・	.1879
0.5	.1915	.1950	.1985	・・・	.2224
・					

異なる標準正規分布表の表現





標準正規分布表 A（抜粋）

z	0.00	**0.01**	0.02	・・・	0.09
0.0	.0000	.0040	.0080	・・・	.0359
0.1	.0398	.0438	.0478	・・・	.0753
0.2	.0793	.0832	.0871	・・・	.1141
0.3	.1179	.1217	.1255	・・・	.1517
0.4	.1554	**.1591**	.1628	・・・	.1879

標準正規分布表 B（抜粋）

z	.00	**.01**	.02	・・・	.09
0.0	.5000	.4960	.4920	・・・	.4641
0.1	.4602	.4562	.4522	・・・	.4247
0.2	.4207	.4168	.4129	・・・	.3859
0.3	.3821	.3783	.3745	・・・	.3483
0.4	.3446	**.3409**	.3372	・・・	.3121

例題　９．２

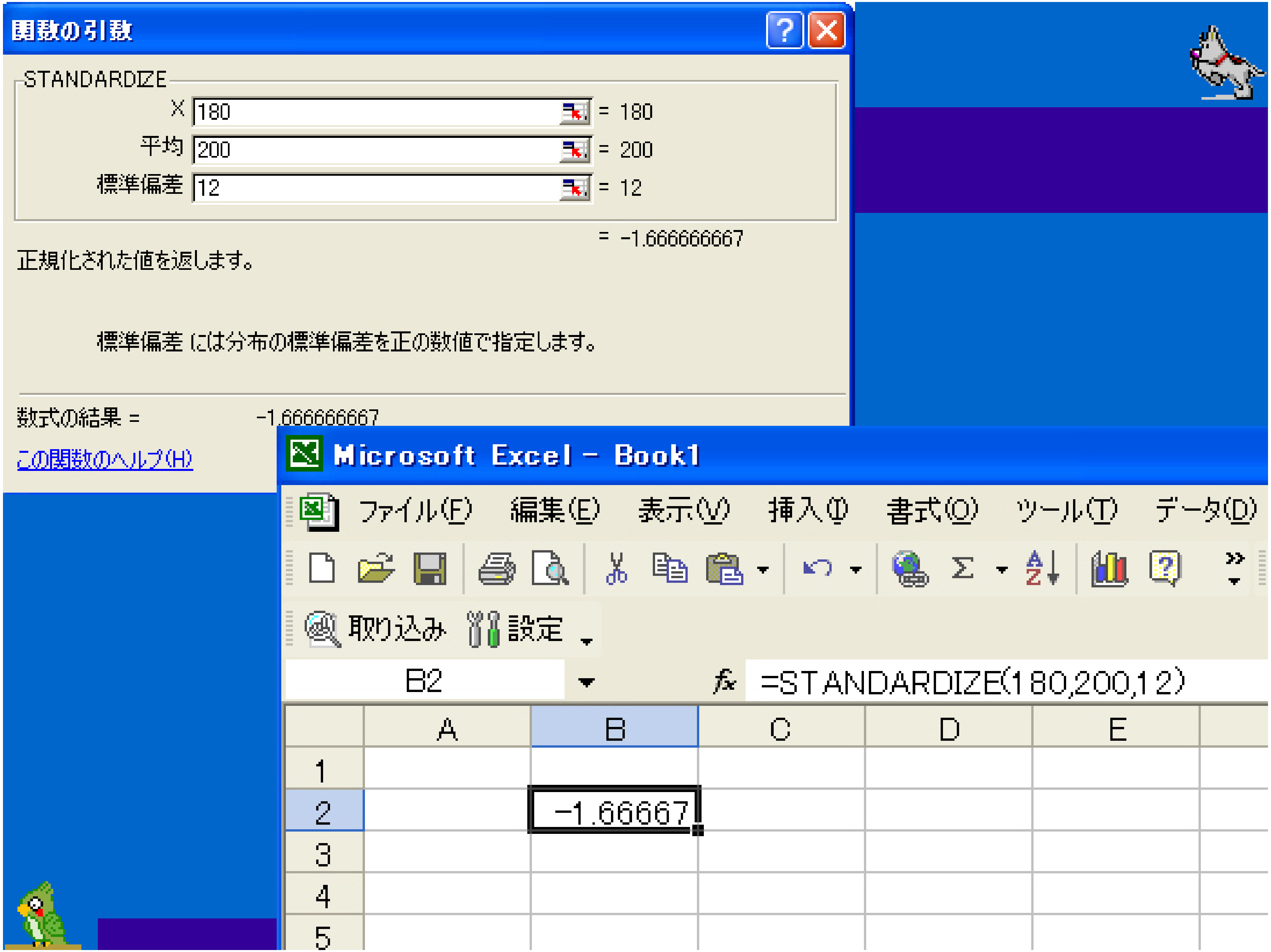
ビデオテープの録画時間は，表示よりも200(秒)長く設定されており，その標準偏差は12(秒)である．

このとき，180(秒)を標準化するといくつになるかを計算せよ．

【解答例】

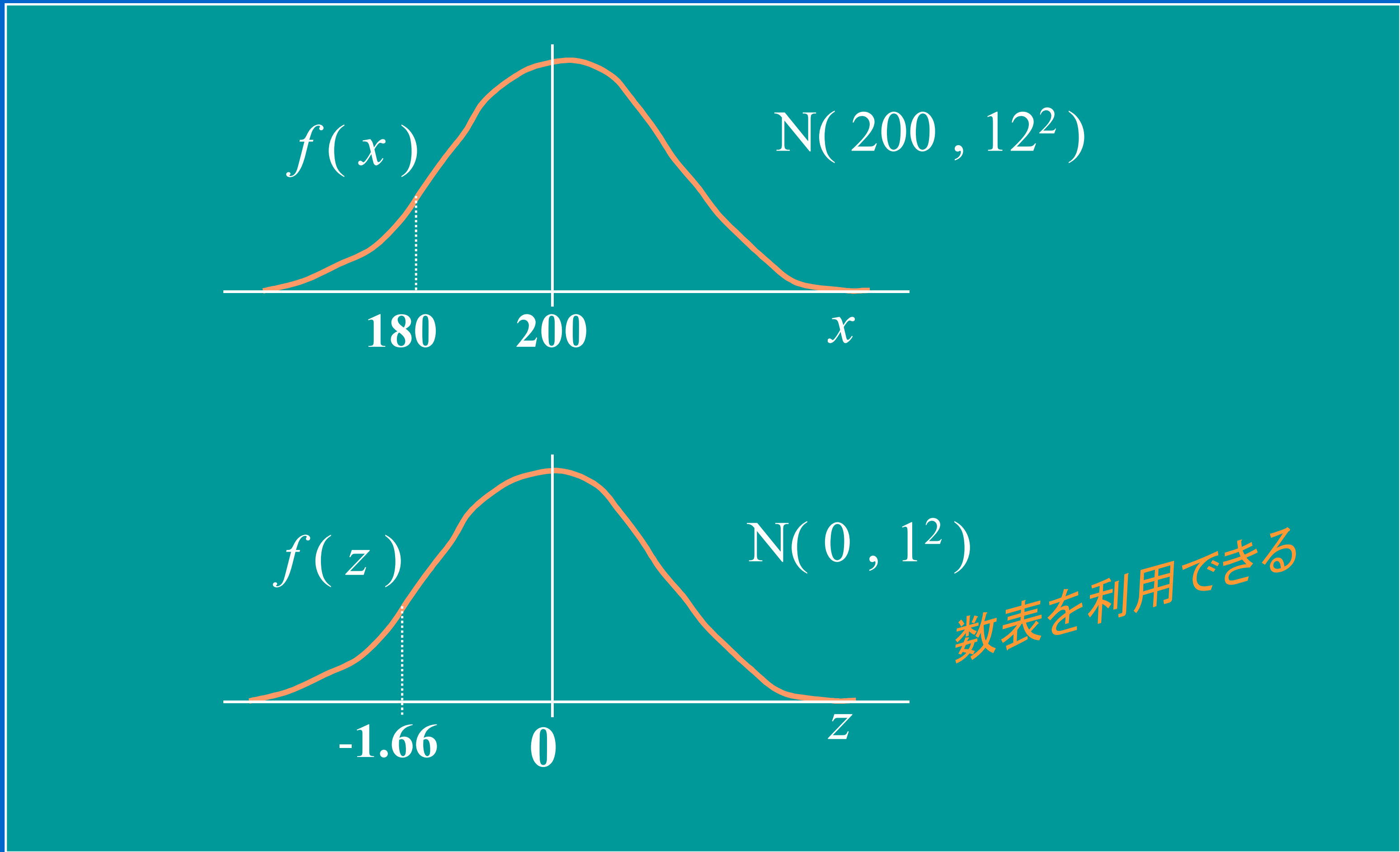
① $Z = (x - \mu) / \sigma = (180 - 200)/12 \fallingdotseq -1.66$

② Excel関数を利用すると，



関数の引数
STANDARDIZE
X 180 = 180
平均 200 = 200
標準偏差 12 = 12
= -1.666666667
正規化された値を返します。
標準偏差 には分布の標準偏差を正の数値で指定します。
数式の結果 = -1.666666667
この関数のヘルプ(H)
Microsoft Excel - Book1
ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D)
取り込み 設定
B2
=STANDARDIZE(180,200,12)
A B C D E
1
2 -1.66667
3
4
5

問題の意味



例題

x が N(20, 4^2)の正規分布に従うとき，以下の確率および値を求めよ．

① $P\,(\,20 \leqq x \leqq 24\,)$　　② $P\,(\,16 \leqq x \leqq 32\,)$

③ $P\,(\,32 \leqq x\,)$　　④ $P\,(\,0 \leqq x \leqq \alpha\,) = 17\,\%$ となる α の値

【解答例】

① 与式$= P\,(\,(20-20)/4 \leqq z \leqq (24-20)/4\,) = P\,(\,0 \leqq z \leqq 1\,) = 0.3413$ //

② 与式$= P\,(\,(16-20)/4 \leqq z \leqq (32-20)/4\,) = P\,(\,-1 \leqq z \leqq 3\,)$

$= P\,(\,-1 \leqq z \leqq 0\,) + P\,(\,0 \leqq z \leqq 3\,)$

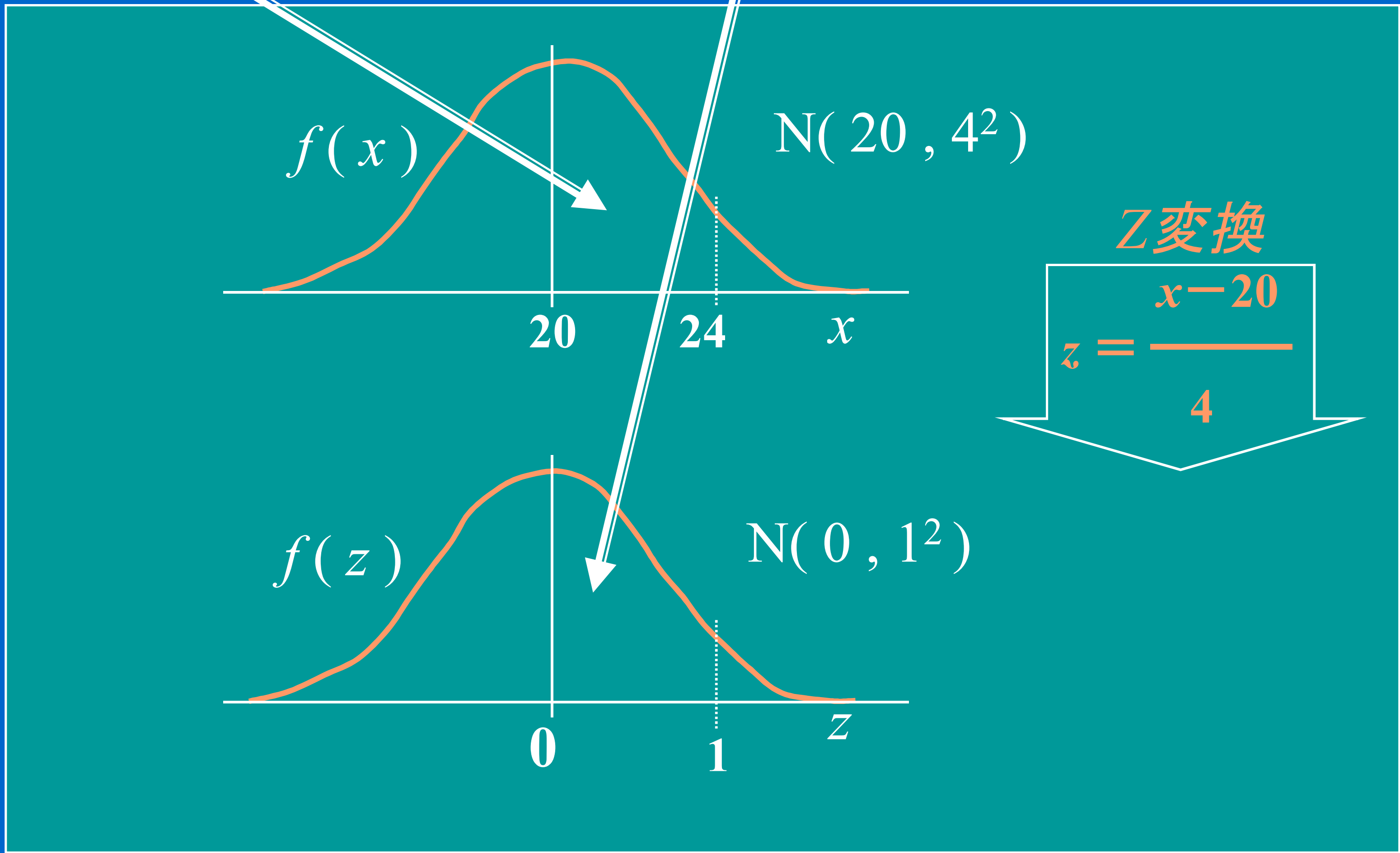
$= P\,(\,0 \leqq z \leqq 1\,) + P\,(\,0 \leqq z \leqq 3\,) = 0.3413 + 0.4986$

$= 0.8399$ //

③ 与式$= P\,(\,(32-20)/4 \leqq z\,) = P\,(\,3 \leqq z\,) = 0.5 - P\,(\,0 \leqq z \leqq 3\,)$

$= 0.5 - 0.4986 = 0.0014$ //

$P(20 \leqq x \leqq 24) = P(0 \leqq z \leqq 1)$ の意味

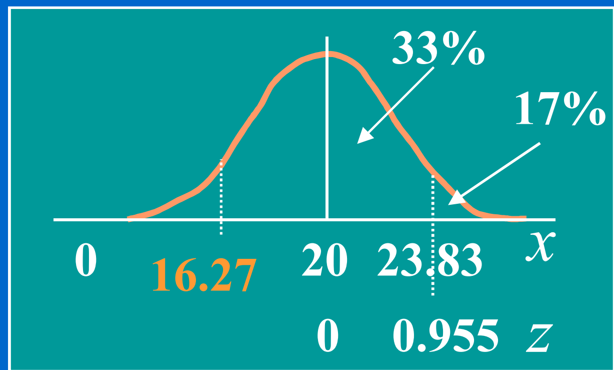


例題 (cont.)

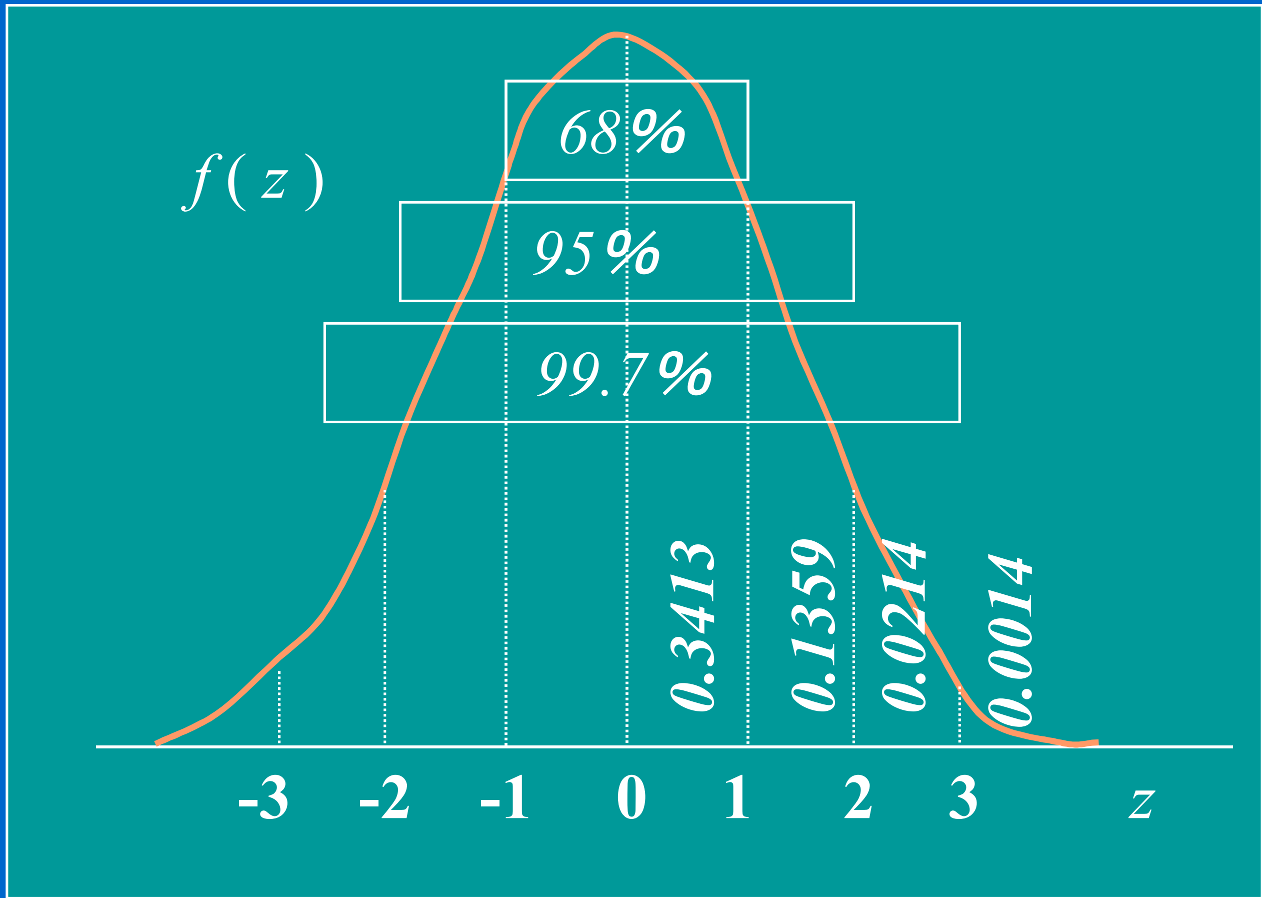
x が N(20, 4^2)の正規分布に従うとき，以下の確率および値を求めよ．
① $P\,(\,20 \leqq x \leqq 24\,)$　② $P\,(\,16 \leqq x \leqq 32\,)$
③ $P\,(\,32 \leqq x\,)$　④ $P\,(\,0 \leqq x \leqq \alpha\,) = 17$ %となるαの値

【解答例】
① 与式＝ 0.3413　② 与式＝ 0.8399　③ 与式＝ 0.0014

④ 与式＝$P\,(\,0 \leqq x \leqq \alpha\,) = 0.17$ となる **α(平均の左側)** を求める
標準正規分布表より，
$0.5-0.17=0.33$ となる z は $z \fallingdotseq 0.955$
$\therefore\ (\,\alpha-20\,)\,/\,4 = 0.955$ ・・・ 平均の右側
$\therefore\ \alpha = 23.83$
求めるαは平均の左側だから
$\alpha = 20-3.83$
$= 16.27$ //



標準正規分布の面積



例題　９．３

ビデオテープの録画時間は，表示よりも200(秒)長く設定されており，その標準偏差は12(秒)であり，N(200，12^2)に従う．

このとき，180(秒)以下となるビデオテープは，全体の何%となるか．

【解答例】

$z=(180-200)/12=-1.667$

標準正規分布表より

$z=1.66$ のとき $Q(z)=0.4515$

$z=1.67$ のとき $Q(z)=0.4525$

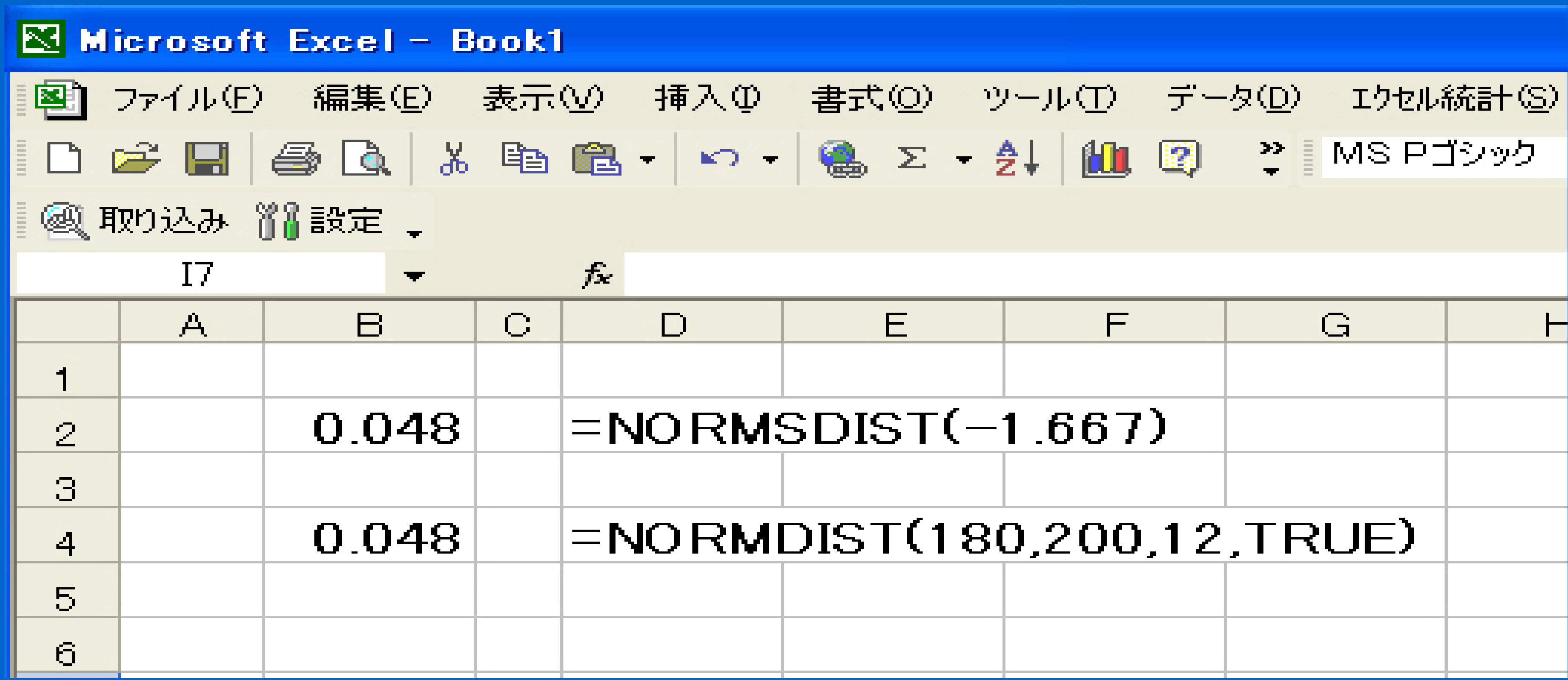
これより

$z=1.667$ のとき $Q(z)=0.4522$　とする

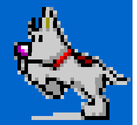
∴

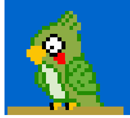
$P(x\leqq 180)=P(z\leqq -1.667)=0.5-0.4522=0.0478$ //

例題 9．3



	A	B	C	D	E	F	G	H
1								
2		0.048		=NORMSDIST(−1.667)				
3								
4		0.048		=NORMDIST(180,200,12,TRUE)				
5								
6								





標本と母集団(復習)

母集団 (population)

n ：標本の大きさ (sample size)

θ

抜き取り，抽出 (sampling)

標本，試料 (sample)

$x_1 , x_2 , x_3 , \cdots , x_n$

母数

- 母平均 (μ, m)
- 母分散 (σ^2)
- 母標準偏差 (σ)
- 母比率 (p)
- 母集団の範囲 (R)
- 母集団の相関係数 (ρ)

$\hat{\theta}$

母数の推定

標本の性質を示す数値

- 標本平均 ($\bar{x}$)
- 標本分散 (s^2)
- 標本の標準偏差 (s)
- 標本比率 ($\hat{p}$)
- 標本の範囲 (R)
- 標本の相関係数 (ρ)　など

N：母集団の大きさ (population size)

標本分布 (sampling dist.)

標本平均($\bar{x}$), 標本分散(s^2), 不偏標本分散(u^2) 等の統計量は確率変数だから確率分布を持つ.

統計量の確率分布は標本分布と呼ばれる.

標本の分布を知ることは, 元となる母集団の分布を知る上で重要である.

標本の分布と母集団の分布の関係が重要

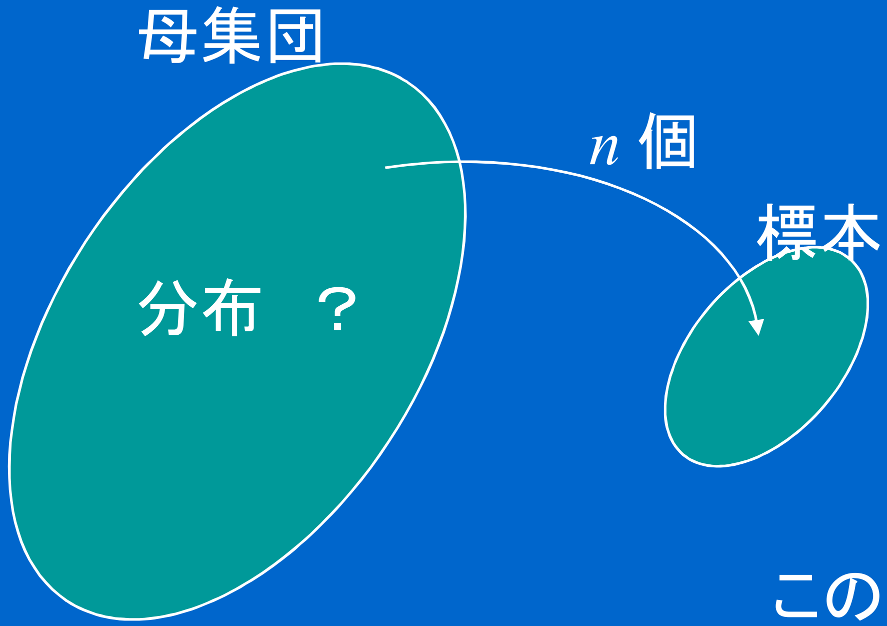
たとえば, 標本平均から母平均を知る.

中心極限定理 (central limit theorem)

母集団の(分布の)型に関係なく,
標本平均 ($\bar{x}$) の分布は,
標本数 (n) が大きければ,

正規分布 $\mathrm{N}\left(\mu,\ \dfrac{\sigma^2}{n}\right)$ に従う.

中心極限定理



$x_1, x_2, x_3, \cdots, x_n$ ： $\bar{x}$

この操作を繰り返す

$\bar{x}_1, \bar{x}_2, \bar{x}_3, \cdots$

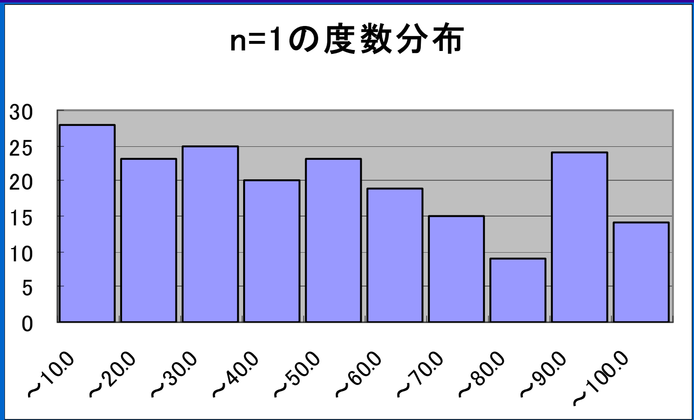
この $\bar{x}$ の分布は n が大きくなると

$\left(\mu, \dfrac{\sigma^2}{n}\right)$ に従う

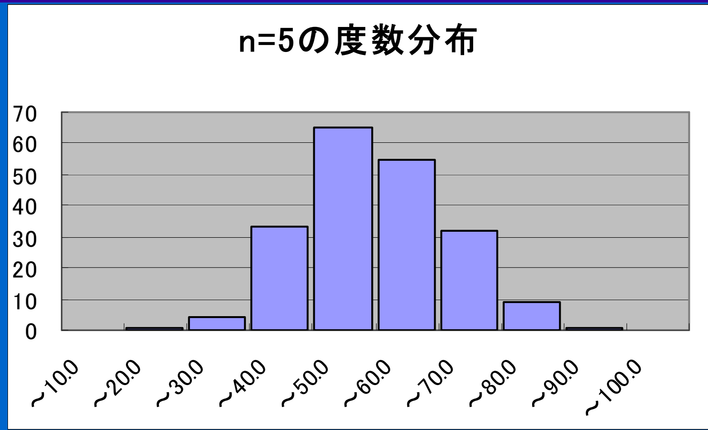
n が大きくなるほど

その精度は高くなる

n個の平均値の分布



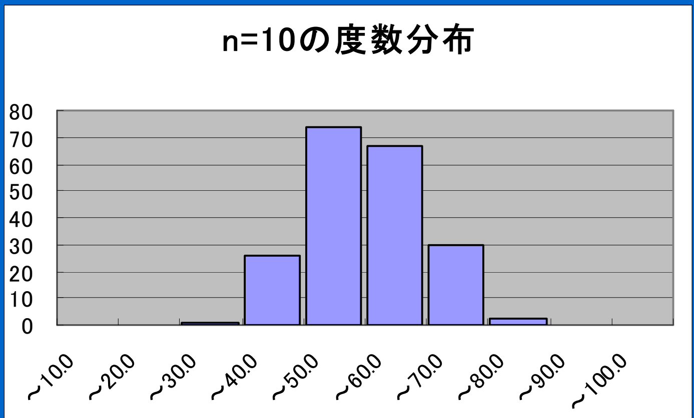
平　　均　44.32
標準偏差　29.46

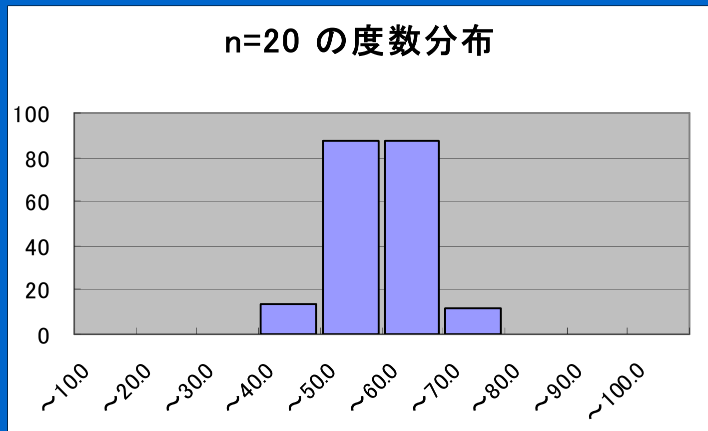


平　　均　50.22
標準偏差　8.75

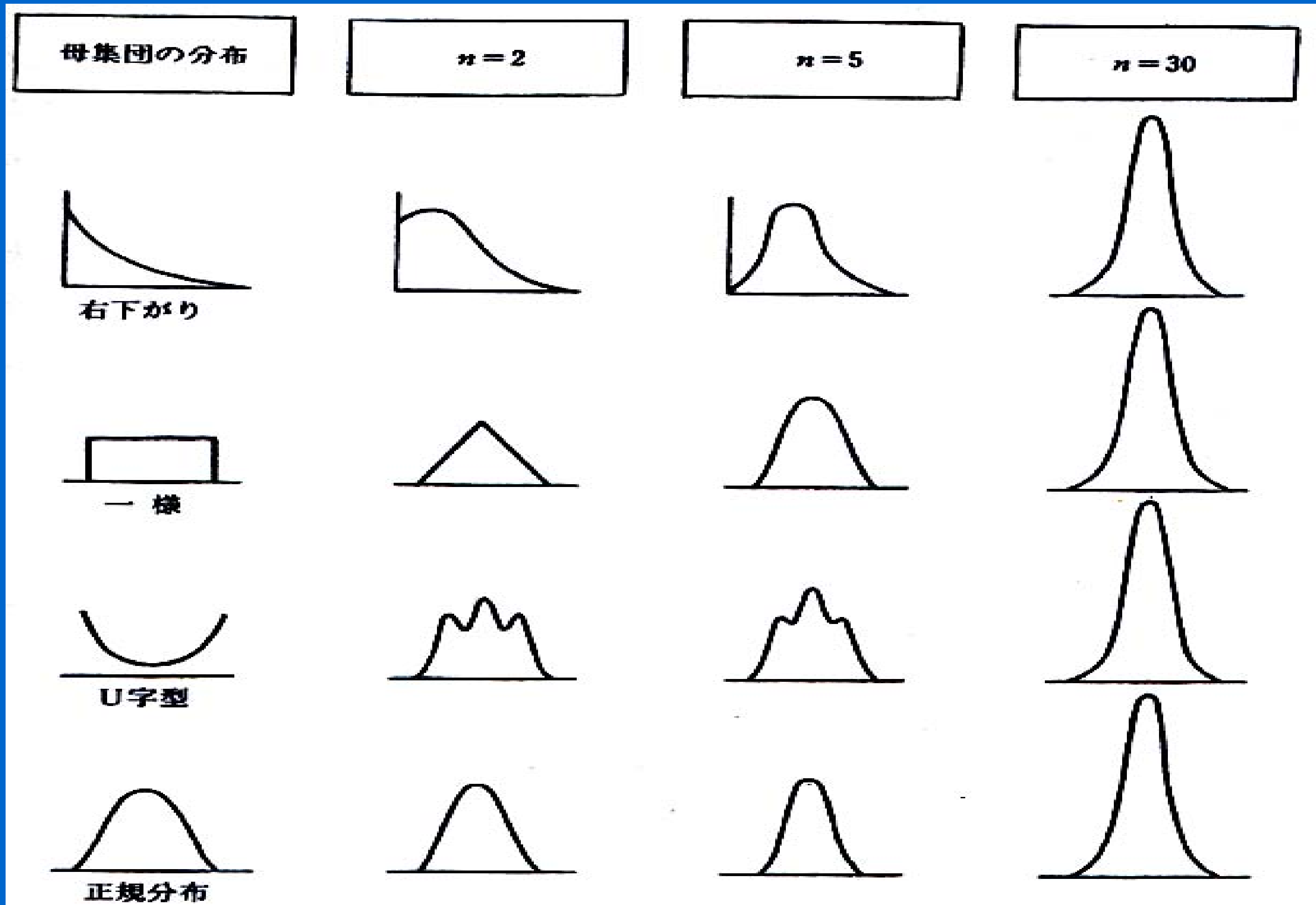
平　　均　50.32
標準偏差　11.72

平　　均　50.01
標準偏差　6.65





中心極限定理のイメージ



中心極限定理の応用

ランダム標本 $X = (X_1, X_2, \cdots, X_n)$ について，
各 $\overline{X_i}$ の平均値が$\mu(\infty < \mu < \infty)$，分散が$\sigma^2 (0 < \sigma^2 < \infty)$であるとき，
標本平均 X の標本分布は，標本の大きさ n が十分大きければ，
正規分布 $N(\mu, \sigma^2 / n)$ に近似する．
すなわち，

$$\overline{X} \sim N(\mu, \sigma^2/n) \quad (n \to \infty)$$

二項分布の正規近似

$B(np, npq)$ において $P(a \leqq S_n \leqq b) \fallingdotseq \int_{\frac{a-np}{\sqrt{npq}}}^{\frac{b-np}{\sqrt{npq}}} \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}} dx$

(a，bは定数)

例題

ある町では，A社のB電化製品の普及率(販売シェア)は24%といわれている．いま，250世帯に電話調査したとき，B電化製品が65世帯以上に普及している確率を求めよ．

【解答例】

① 二項分布を利用した場合．

$$P(x \geqq 65) = \sum_{x=65}^{250} {}_{250}C_x (0.24)^x (0.76)^{250-x}$$ の計算が必要．

② 正規近似を利用した場合．

$p=0.24$，$q=1-p=0.76$，$n=250$ で $P(65 \leqq S_{250} \leqq 250)$を求めればよい．

ここで，$E(S_{250})=np=60$ ，$V(S_n)=npq=45.6$

ところが，$(65-60)/\sqrt{45.6} \fallingdotseq 0.74$ ， $(250-60)/\sqrt{45.6} \fallingdotseq 28.14$

$P(65 \leqq S_{250} \leqq 250) = P(0.74 \leqq x \leqq 28.14) = 0.5-0.27035=0.22965$ //

t 分布 (student's t distribution)

母分散σ^2未知の場合の平均値の分布

$\bar{x}$の分布は正規分布 N(μ, σ^2 / n) に従う

その標準化したものは，N(0, 1^2) に従う

$$z = \frac{\bar{x} - \mu}{\sqrt{\sigma^2 / n}}$$

このとき，σ^2 は未知なのでs^2を使う

s^2 ：標本の分散

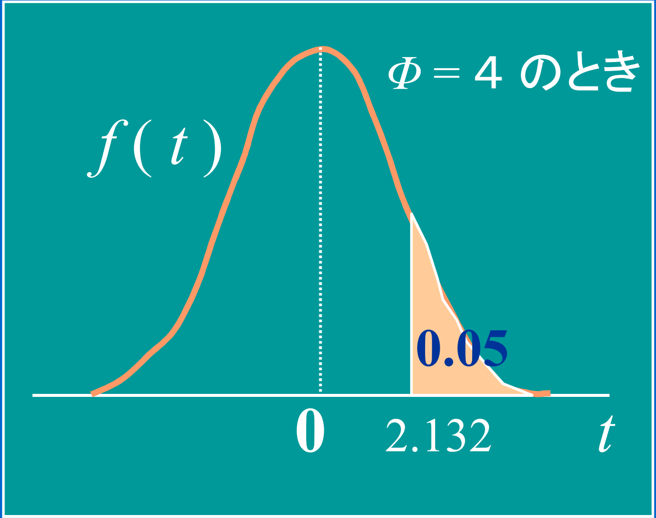
$$t = \frac{\bar{x} - \mu}{\sqrt{s^2 / n}}$$

t は正規分布に従わない

$t(\varphi, \alpha)$

t 分布は小標本の場合(n ＜ 30) の理論

t 分布表の読み方



t 分布表（抜粋）

φ ＼ α	.100	**.050**	.025	・・・	.001	.0005
1	3.078	6.314	12.706		318.31	636.62
2	1.886	2.920	4.303		22.326	31.598
3	1.638	2.353	3.182		10.213	12.924
4	1.533	**2.132**	2.776		7.173	8.610
5	1.476	2.015	2.571		5.893	6.869
・	・					
・	・					

例題　９．６

ビデオテープの問題

n ＝ 10 の標本を取ったとき，

$\bar{x}$ ＝ 215 (秒), s ＝ 13 (秒) であった.

標本平均より大きな値が得られる確率を求めよ.

【解答例】

自由度(φ)＝9　，　μ＝203　(*cf. p.75*)

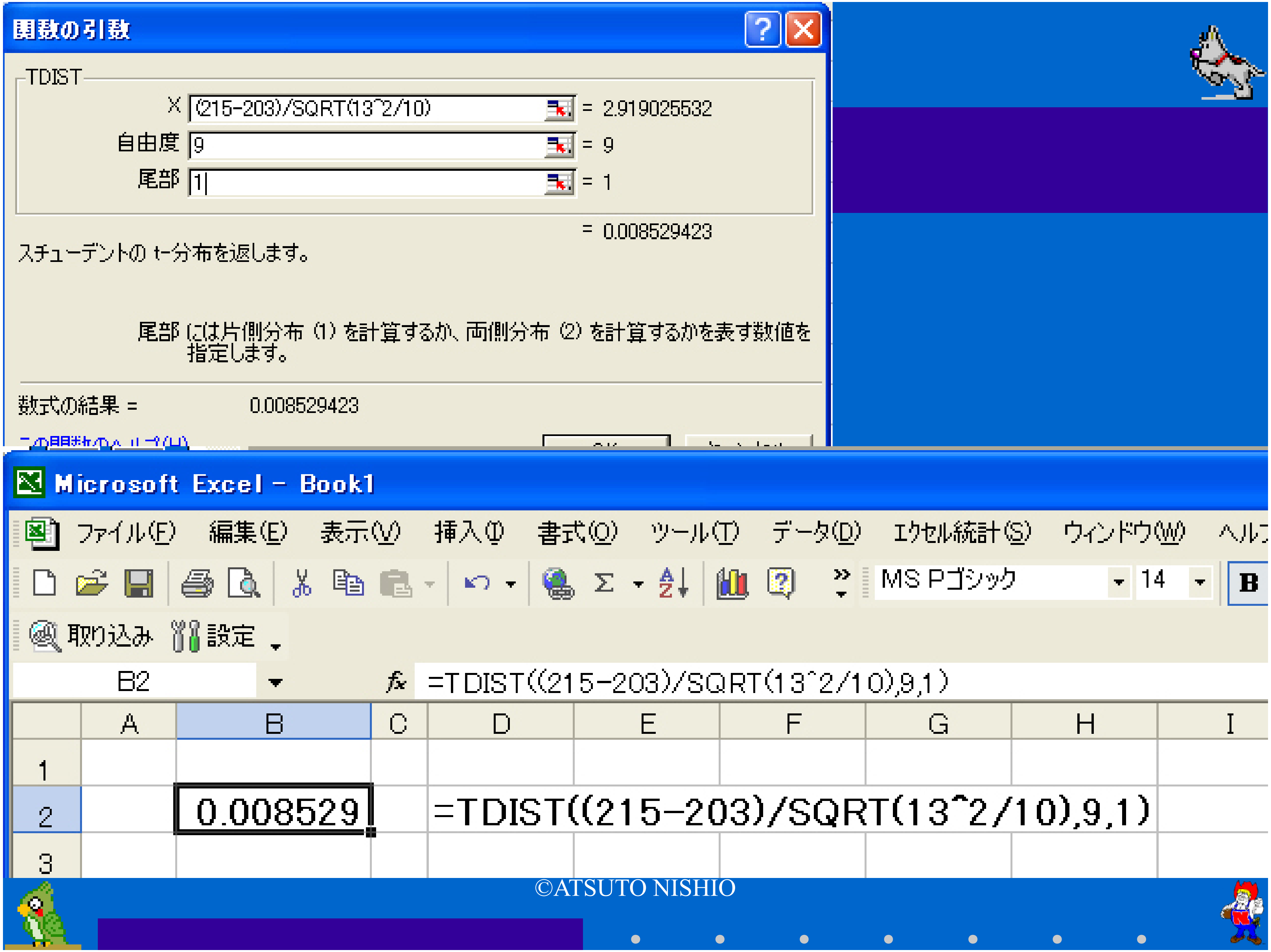
$t = (215 - 203) / \sqrt{13^2 / 10} = 2.92$

【 t 分布表 】

自由度	.01	.005
9	2.821	3.250

2.92はこの間にある

答えは0.01と0.005の間を比例配分する


関数の引数
TDIST
X (215-203)/SQRT(13^2/10) = 2.919025532
自由度 9 = 9
尾部 1 = 1
= 0.008529423
スチューデントの t-分布を返します。
尾部 には片側分布 (1) を計算するか、両側分布 (2) を計算するかを表す数値を指定します。
数式の結果 = 0.008529423
Microsoft Excel - Book1
ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) エクセル統計(S) ウィンドウ(W)
MS Pゴシック 14
取り込み 設定
B2
=TDIST((215-203)/SQRT(13^2/10),9,1)
0.008529
=TDIST((215-203)/SQRT(13^2/10),9,1)
©ATSUTO NISHIO

χ^2 分布

標本分散の分布

$z_i = (x_i - \mu) / \sigma \quad (i = 1, 2, \cdots, n)$ とし，

z_i が互いに独立ならば，N(μ, σ^2)に従う.

このとき

$$\chi^2 = z_1^2 + z_2^2 + \cdots + z_n^2$$

$$= \sum_{i=1}^{n} z_i^2$$

$\chi^2(\varphi, \alpha)$

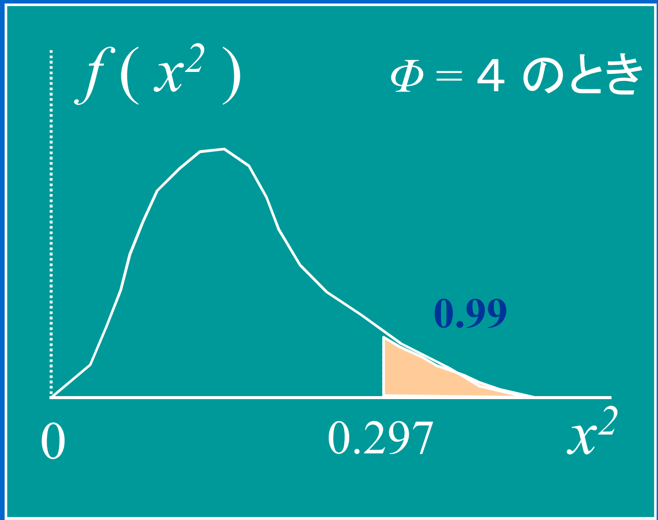
母分散と標本分散

σ^2 は未知だから,代わりに s^2 を用いる

$$\chi^2 \sim \frac{(n-1)s^2}{\sigma^2}$$

$\chi2$は$\varphi = n-1$の$\chi2$分布に従う

χ^2 分布表の読み方



χ^2分布表（抜粋）

φ \ α	.995	.99	.0975	・・・	.01
1	$0.0^{4}393$	$0.0^{3}157$	$0.0^{3}982$		6.63
2	0.0100	0.0201	0.0506		9.21
3	0.0717	0.115	0.216		11.34
4	0.207	0.297	0.484		13.28
5	0.412	0.554	0.831		15.09
・	・				
・	・				

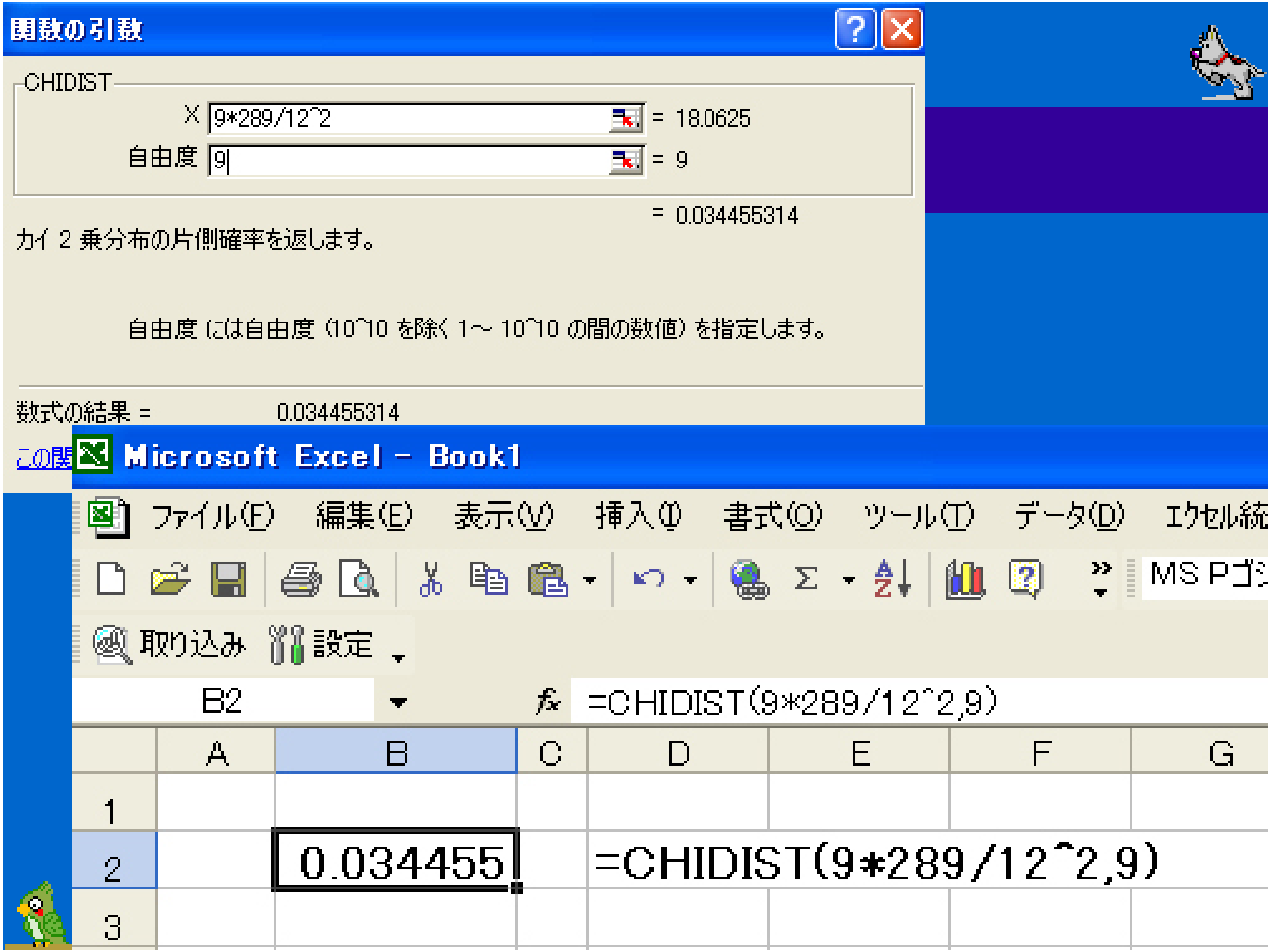
例題　9．5

母集団であるビデオテープ 100 本の録画時間を調べたところ,
$\mu = 203$ (秒) , $\sigma = 12$ (秒) であった.
この母集団より $n = 10$ の標本を取った.
標本の分散 (s^2)が289秒を超える確率を求めよ.

【解答例】

自由度(φ)＝9

$$P(s^2 > 289) = P(\chi^2 > (9 \times 289) / 12^2)$$
$$= P(18.0625)$$
$$= 0.03517$$


関数の引数
CHIDIST
X 9*289/12^2 = 18.0625
自由度 9 = 9
= 0.034455314
カイ 2 乗分布の片側確率を返します。
自由度 には自由度 (10^10 を除く 1～ 10^10 の間の数値) を指定します。
数式の結果 = 0.034455314
Microsoft Excel - Book1
ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D)
取り込み 設定
B2
=CHIDIST(9*289/12^2,9)
A B C D E F G
1
2 0.034455 =CHIDIST(9*289/12^2,9)
3

F 分布

不偏分散比の分布

χ^2 分布に従う2つの標本分散の比

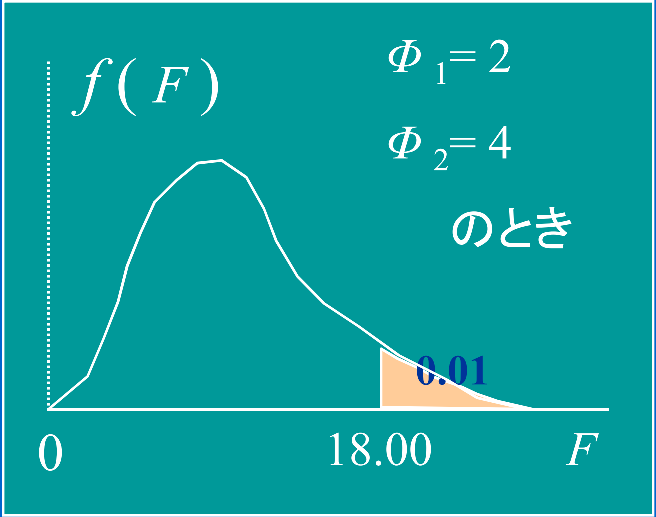
すなわち

２つの確率変数 $X \sim \chi^2_{\varphi 1}$, $Y \sim \chi^2_{\varphi 2}$ が独立であるとき,

$$F = \frac{X / \varphi_1}{Y / \varphi_2}$$

は $F(\varphi_1, \varphi_2, \alpha)$ に従う.

F 分布表の読み方



F 分布表（抜粋）　α＝0.01

Φ_2 \ Φ_1	1	2	3	・・・	9
1	4052	4999	5403		6106
2	98.49	99.00	99.17		99.42
3	34.12	30.81	29.46		27.05
4	21.20	18.00	16.69		14.37
5	16.26	13.27	12.06		9.89
・	・				
・	・				