統計的な推測

以下，根元事象はすべて同様に確からしいとする。

1　確率変数と確率分布

赤玉 2 個と白玉 4 個が入っている袋から，3 個の玉を取り出すとき，取り出した白玉の個数 X の確率分布を求めよ。また，確率 $P(X \geqq 2)$を求めよ。

要　点

袋から取り出したある玉の個数のように，試行の結果によってある値をとり，その値をとる確率が定まる変数 X を **確率変数** という。確率変数 X のとる値 x_1，x_2，x_3，……，x_n と，X がそれらの値をとる確率 p_1，p_2，p_3，……，p_n の対応関係を，X の **確率分布** ，または単に **分布** といい，確率変数 X はこの分布に **従う** という。

確率分布は，右のような表で表すことが多い。

X	x_1	x_2	x_3	……	x_n	計
確率 P	p_1	p_2	p_3	……	p_n	1

このとき，次のことが成り立つ。

$$\boldsymbol{p_1 \geqq 0,\ p_2 \geqq 0,\ p_3 \geqq 0,\ \cdots\cdots,\ p_n \geqq 0,\quad p_1+p_2+p_3+\cdots\cdots+p_n=1}$$

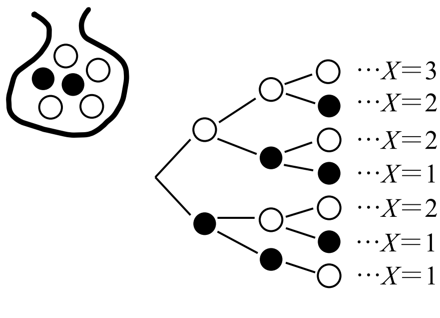
確率変数 X が 1 つの値 a をとる確率を $\boldsymbol{P(x=a)}$ で表す。

また，X が a 以上 b 以下の値をとる確率を $\boldsymbol{P(a \leqq x \leqq b)}$ で表す。

解答

玉の出方は右の樹形図のようになり，

X のとり得る値は 1，2，3 である。



○：白玉　●：赤玉

$$P(X=1)=\frac{{}_2\mathrm{C}_2\ \cdot\ {}_4\mathrm{C}_1}{{}_6\mathrm{C}_3}=\frac{1\cdot 4}{\frac{6\cdot 5\cdot 4}{3\cdot 2\cdot 1}}=\frac{1}{5}$$

$$P(X=2)=\frac{{}_2\mathrm{C}_1\ \cdot\ {}_4\mathrm{C}_2}{{}_6\mathrm{C}_3}=\frac{2\cdot\frac{4\cdot 3}{2}}{\frac{6\cdot 5\cdot 4}{3\cdot 2\cdot 1}}=\frac{3}{5}$$

$$P(X=3)=\frac{{}_4\mathrm{C}_3}{{}_6\mathrm{C}_3}=\frac{\frac{4\cdot 3\cdot 2}{3\cdot 2\cdot 1}}{\frac{6\cdot 5\cdot 4}{3\cdot 2\cdot 1}}=\frac{1}{5}$$

樹形図は X のとり得る値を調べただけで，確率は別途調べる必要がある。例えば，$P(X=1)$ は $\frac{3}{7}$ とはならない。

X の確率分布は，次の表のようになる。

X	1	2	3	計
P	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$	1

また　$P(X \geqq 2)=\frac{3}{5}+\frac{1}{5}=\boldsymbol{\frac{4}{5}}$

2 **確率変数の平均**

右の表のような賞金がついている
20 本のくじがある。
このくじを 1 本引くとき，賞金の
平均を求めよ。

	賞金	本数
1 等	10000 円	1 本
2 等	1000 円	3 本
3 等	100 円	16 本

要　点

確率変数 X が，右の表のような
確率分布に従うとき，

$$x_1p_1+x_2p_2+x_3p_3+\cdots\cdots+x_np_n$$

を確率変数 X の **平均** または **期待値** と
いい，$E(X)$ で表す。

X	x_1	x_2	x_3	……	x_n	計
P	p_1	p_2	p_3	……	p_n	1

確率変数 X の平均

$$E(X)=x_1p_1+x_2p_2+x_3p_3+\cdots\cdots+x_np_n$$

これを，「数列」で学習する Σ を用いて表すと

$$E(X)=\sum_{k=1}^{n}x_kp_k$$

解答

このくじを 1 本引いたときの賞金を X 円とすると，
確率変数 X の確率分布は右の表のようになる。
よって，賞金の平均 $E(X)$ は

X	10000	1000	100	計
P	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{16}{20}$	1

$$E(X)=10000\times\frac{1}{20}+1000\times\frac{3}{20}+100\times\frac{16}{20}=\mathbf{730}\text{（円）}$$

コメント

統計的な推測のために集めたデータは，一般的にはただの数字の羅列にしか見えません。そこで
① グラフ化してその特徴を捉える
② データを特徴付ける値を算出する
などして，意味のある情報を抽出します。このような方法は「縮約」と呼ばれます。
〈注意〉「平均」は，統計において最も基本的な値（統計量）の 1 つです。

3 **確率変数の分散と標準偏差**

赤玉 2 個と白玉 4 個が入っている袋から，3 個の玉を取り出すとき，取り出した白玉の個数 X の
平均，分散，標準偏差を求めよ。

要　点

X	x_1	x_2	x_3	……	x_n	計
P	p_1	p_2	p_3	……	p_n	1

確率変数 X が，右の表のような確率分布に従い，平均が m であるとき，$\boldsymbol{X-m}$ を **偏差** という。

このとき，偏差の 2 乗$(X-m)^2$も確率変数とみることができる。その平均

$$\boldsymbol{E((X-m)^2)=(x_1-m)^2p_1+(x_2-m)^2p_2+(x_3-m)^2p_3+\cdots\cdots+(x_n-m)^2p_n}$$

を，確率変数 X の **分散** といい，$\boldsymbol{V(X)}$ で表す。

〈注意〉確率変数 X の値が平均 m の値に近いほど，$(X-m)^2$ の値も小さくなる。すなわち，分散 $V(X)$の値が小さいほど確率変数 X は平均 m の値の近くに分布するので，散らばり具合は小さいといえる。

分散 $V(X)$は 0 以上の値となるので，正の平方根 $\sqrt{V(X)}$ を考えることができる。

これを X の **標準偏差** といい，$\boldsymbol{\sigma(X)}$ で表す。

確率変数 X の分散と標準偏差

$$\boldsymbol{V(X)=E((X-m)^2)=\sum_{k=1}^{n}(x_k-m)^2p_k}$$

$$\boldsymbol{\sigma(X)=\sqrt{V(X)}}$$

〈注意〉確率変数 X の単位に対して，分散 $V(X)$は単位も 2 乗になってしまうが，標準偏差 $\sigma(X)$は確率変数 X と同じ単位をもつ。

$V(X)=\displaystyle\sum_{k=1}^{n}(x_k-m)^2p_k$ の右辺を，「数列」で学習する$\sum$の性質を用いて変形すると

$$\sum_{k=1}^{n}(x_k-m)^2p_k=\sum_{k=1}^{n}(x_k{}^2-2mx_k+m^2)p_k=\sum_{k=1}^{n}x_k{}^2p_k-2m\sum_{k=1}^{n}x_kp_k+m^2\sum_{k=1}^{n}p_k$$

ここで，$\displaystyle\sum_{k=1}^{n}x_k{}^2p_k$ は，X^2を確率変数とみたときの平均と考えることができ，$E(X^2)$で表される。

また，$\displaystyle\sum_{k=1}^{n}x_kp_k=E(X)=m$，$\displaystyle\sum_{k=1}^{n}p_k=1$ であるから

$$\sum_{k=1}^{n}(x_k-m)^2p_k=E(X^2)-2m\cdot m+m^2\cdot 1=E(X^2)-m^2=E(X^2)-\{E(X)\}^2$$

すなわち，$\boldsymbol{V(X)=E(X^2)-\{E(X)\}^2}$ と表すこともできる。

解答

$\boxed{1}$より，X の確率分布は，右の表のようになる。

X	1	2	3	計
P	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$	1

よって，**X の平均 $E(X)$は**

$$E(X)=1\times\frac{1}{5}+2\times\frac{3}{5}+3\times\frac{1}{5}=\frac{1+6+3}{5}=\mathbf{2}\text{（個）}$$

X の分散 $V(X)$は

$$V(X)=E(X^2)-\{E(X)\}^2=1^2\times\frac{1}{5}+2^2\times\frac{3}{5}+3^2\times\frac{1}{5}-2^2=\frac{1+12+9}{5}-4=\boldsymbol{\frac{2}{5}}$$

Xの標準偏差$\sigma(X)$は　$\sigma(X)=\sqrt{V(X)}=\sqrt{\dfrac{2}{5}}=\dfrac{\sqrt{10}}{5}$（個）

分散 $V(X)$の別解

$$V(X)=E((X-m)^2)=(1-2)^2\times\frac{1}{5}+(2-2)^2\times\frac{3}{5}+(3-2)^2\times\frac{1}{5}=\frac{1+0+1}{5}=\frac{2}{5}$$

コメント

標準偏差は，平均が同じ 2 つのデータ A，B を比べるとき，役に立ちます。A の標準偏差が B の標準偏差より小さいとき，データ A では平均に近い値が出やすく，データ B では平均から離れた値が出る場合も結構あるということになり，データ A，B を選ぶ際の重要な値になります。

4　確率変数の変換

1 個のさいころを 1 回投げるときの出る目を X とする。確率変数 X の平均，分散，標準偏差を求めよ。
また，$Y=2X-1$ で定められる確率変数 Y の平均，分散，標準偏差を求めよ。

要　点

確率変数 $aX+b$ の平均，分散，標準偏差

a，b を定数とする。

$E(aX+b)=aE(X)+b$　　$V(aX+b)=a^2V(X)$　　$\sigma(aX+b)=|a|\sigma(X)$

証明

確率変数 X は，右の表のような確率分布に従うものとし，$Y=aX+b$（a，b は定数）とする。

X	x_1	x_2	x_3	……	x_n	計
P	p_1	p_2	p_3	……	p_n	1

［ $E(aX+b)=aE(X)+b$ ］

$$E(aX+b)=(ax_1+b)p_1+(ax_2+b)p_2+(ax_3+b)p_3+\cdots\cdots+(ax_n+b)p_n$$
$$=a(x_1p_1+x_2p_2+x_3p_3+\cdots\cdots+x_np_n)+b(p_1+p_2+p_3+\cdots\cdots+p_n)$$

ここで，$x_1p_1+x_2p_2+x_3p_3+\cdots\cdots+x_np_n=E(X)$，　$p_1+p_2+p_3+\cdots\cdots+p_n=1$ であるから　$E(aX+b)=aE(X)+b$

［ $V(aX+b)=a^2V(X)$ ］

$V(Y)=E((Y-E(Y))^2)$であり，$Y=aX+b$，$E(Y)=aE(X)+b$ であるから

$$Y-E(Y)=aX+b-(aE(X)+b)=aX-aE(X)$$

よって，$E(X)=m$ とおくと

$$V(Y)=E((aX-am)^2)=(ax_1-am)^2p_1+(ax_2-am)^2p_2+(ax_3-am)^2p_3+\cdots\cdots+(ax_n-am)^2p_n$$
$$=a^2\{(x_1-m)^2p_1+(x_2-m)^2p_2+(x_3-m)^2p_3+\cdots\cdots+(x_n-m)^2p_n\}=a^2V(X)$$

［ $\sigma(aX+b)=|a|\sigma(X)$ ］

$$\sigma(aX+b)=\sqrt{V(aX+b)}=\sqrt{a^2V(X)}=|a|\sqrt{V(X)}=|a|\sigma(X)$$

解答

確率変数 X は，右の表のような確率分布に従う。

X	1	2	3	4	5	6	計
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

よって，**X の平均 $E(X)$ は**

$$E(X)=1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=\frac{21}{6}=\boldsymbol{\frac{7}{2}}$$

X の分散 $V(X)$ は

$$V(X)=E(X^2)-\{E(X)\}^2=1^2\times\frac{1}{6}+2^2\times\frac{1}{6}+3^2\times\frac{1}{6}+4^2\times\frac{1}{6}+5^2\times\frac{1}{6}+6^2\times\frac{1}{6}-\left(\frac{7}{2}\right)^2$$

$$=\frac{1+4+9+16+25+36}{6}-\frac{49}{4}=\frac{91}{6}-\frac{49}{4}=\frac{182-147}{12}=\boldsymbol{\frac{35}{12}}$$

X の標準偏差 $\sigma(X)$ は　$\sigma(X)=\sqrt{V(X)}=\sqrt{\dfrac{35}{12}}=\dfrac{\sqrt{35}\cdot 2\sqrt{3}}{12}=\boldsymbol{\dfrac{\sqrt{105}}{6}}$

また，**Y の平均 $E(Y)$ は**　$E(Y)=E(2X-1)=2E(X)-1=2\cdot\dfrac{7}{2}-1=\boldsymbol{6}$

Y の分散 $V(Y)$ は　$V(Y)=V(2X-1)=2^2V(X)=4\cdot\dfrac{35}{12}=\boldsymbol{\dfrac{35}{3}}$

Y の標準偏差 $\sigma(Y)$ は　$\sigma(Y)=\sigma(2X-1)=|2|\,\sigma(X)=2\cdot\dfrac{\sqrt{105}}{6}=\boldsymbol{\dfrac{\sqrt{105}}{3}}$

5 **二項分布**

1 枚の硬貨を 5 回繰り返し投げるとき，表が出る回数を X とする。X の確率分布を求めよ。
また，表が 4 回以上出る確率を求めよ。

要　点

1 回の試行で事象 A の起こる確率を p とする。この試行を n 回繰り返す反復試行において，事象 A の起こる回数を X とすると，X は確率変数で，$X=r$ となる確率は

$P(X=r)={}_n\mathrm{C}_r p^r q^{n-r}$　（ただし　$q=1-p$，$r=0$，1，2，……，n）

である。よって，X の確率分布は次の表のようになる。

X	0	1	2	…	r	…	n	計
P	${}_n\mathrm{C}_0\, q^n$	${}_n\mathrm{C}_1\, pq^{n-1}$	${}_n\mathrm{C}_2\, p^2q^{n-2}$	…	${}_n\mathrm{C}_r\, p^rq^{n-r}$	…	${}_n\mathrm{C}_n\, p^n$	1

このような確率分布を **二項分布** といい，**$B(n,\ p)$** で表す。
また，確率変数 X は **二項分布 $B(n,\ p)$ に従う** という。
〈注意〉二項定理により，確率 P の総和は 1 であることが確認できる。

$${}_n\mathrm{C}_0\, q^n+{}_n\mathrm{C}_1\, pq^{n-1}+{}_n\mathrm{C}_2\, p^2q^{n-2}+\cdots\cdots+{}_n\mathrm{C}_r\, p^rq^{n-r}+\cdots\cdots+{}_n\mathrm{C}_n\, p^n=(q+p)^n=1^n=1$$

解答

1 回の試行で表が出る確率は$\dfrac{1}{2}$である。よって，確率変数Xは二項分布$B\left(5,\ \dfrac{1}{2}\right)$に従う。

硬貨の表がr回出る確率は　　$P(X=r)={}_5\mathrm{C}_r\left(\dfrac{1}{2}\right)^r\left(\dfrac{1}{2}\right)^{5-r}$　　（$r=0,\ 1,\ 2,\ \cdots\cdots,\ 5$）

であるから，X の確率分布は次の表のようになる。

X	0	1	2	3	4	5	計
P	$\dfrac{1}{32}$	$\dfrac{5}{32}$	$\dfrac{10}{32}$	$\dfrac{10}{32}$	$\dfrac{5}{32}$	$\dfrac{1}{32}$	1

また，表が 4 回以上出る確率 $P(X\geqq 4)$は

$$P(X\geqq 4)=\frac{5}{32}+\frac{1}{32}=\frac{6}{32}=\boldsymbol{\frac{3}{16}}$$

6　二項分布の平均・分散・標準偏差

(1)　1 枚の硬貨を 100 回投げるとき，表が出る回数 X の平均，分散，標準偏差を求めよ。

(2)　プロ野球において，打率 3 割のバッターがいるとする。すなわち，このバッターは全打席において 3 割の確率でヒットを打つものとする。このバッターが 500 回打席に立つとき，打つヒットの本数 X の平均，分散，標準偏差を求めよ。

要　点

確率変数 X が二項分布 $B(n,\ p)$に従うとき

$\boldsymbol{E(X)=np,}$　　$\boldsymbol{V(X)=npq,}$　　$\boldsymbol{\sigma(X)=\sqrt{npq}}$　　ただし　$\boldsymbol{q=1-p}$

解答

(1)　確率変数Xは，二項分布$B\left(100,\ \dfrac{1}{2}\right)$に従う。

よって，**Xの平均$E(X)$は**　$E(X)=100\cdot\dfrac{1}{2}=\mathbf{50}$(回)

Xの分散$V(X)$は　$V(X)=100\cdot\dfrac{1}{2}\cdot\dfrac{1}{2}=\mathbf{25}$

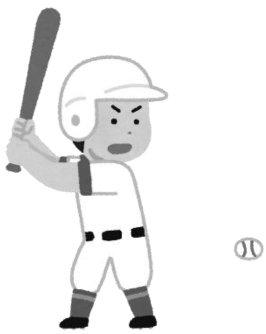
Xの標準偏差$\sigma(X)$は　$\sigma(X)=\sqrt{25}=\mathbf{5}$（回）

(2)　確率変数Xは，二項分布$B\left(500,\ \dfrac{3}{10}\right)$に従う。

よって，**Xの平均$E(X)$は**　$E(X)=500\cdot\dfrac{3}{10}=\mathbf{150}$(本)

Xの分散$V(X)$は　$V(X)=500\cdot\dfrac{3}{10}\cdot\dfrac{7}{10}=\mathbf{105}$

Xの標準偏差$\sigma(X)$は　$\sigma(X)=\boldsymbol{\sqrt{105}}$（本）



$\boxed{7}$ **連続型確率変数**

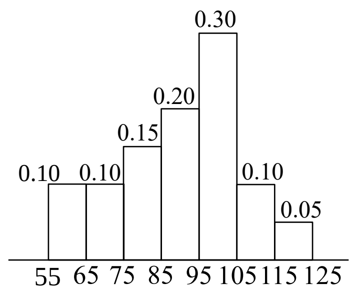
確率変数 X のとり得る値 x の範囲が $0 \leqq x \leqq 1$ で，その確率密度関数が $f(x)=2x\ (0 \leqq x \leqq 1)$ で表されるとき，確率 $P\left(\dfrac{1}{2} \leqq X \leqq 1\right)$ を求めよ。

要　点

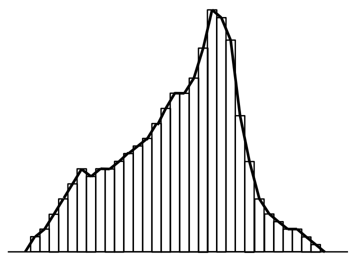
右の度数分布表は，高校生 20 人の 1 日のテレビの視聴時間を調べたものである。この 20 人の中から 1 人を選び，その生徒の視聴時間を X 分とする。
このとき，例えば X の属する階級の階級値が 80 となる確率は，$75 \leqq X < 85$ となる確率 $P(75 \leqq X < 85)$ に等しい。これは階級 75～85 の相対度数 0.15 と一致する。つまり，X は X の属する階級の階級値となる確率変数であるといえる。

階級（分）	階級値（分）	度数（人）	相対度数
55 以上～65 未満	60	2	0.10
65　～75	70	2	0.10
75　～85	80	3	0.15
85　～95	90	4	0.20
95　～105	100	6	0.30
105　～115	110	2	0.10
115　～125	120	1	0.05
合計		20	1.00

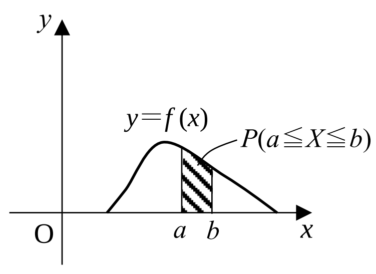
右の図は，上の表で示した各階級の相対度数を，長方形の面積で表したヒストグラムであるとする。
各長方形の面積の和は 1 である。



データを増やし，各長方形の面積の総和は 1 としたままで階級の幅を細かく分けたとき，長方形の上辺の中点を結んだ折れ線は，
1 つの曲線に近づいていくと考えられる。
ただし，両端は度数 0 の階級があるものとして線で結ぶ。



一般に，連続した値をとる確率変数 X を **連続型確率変数** という。
このような連続型確率変数 X の確率分布を考える場合は，1 つの曲線 $y=f(x)$ を対応させ，$a \leqq X \leqq b$ となる確率 $P(a \leqq X \leqq b)$ が，右の図の斜線部分の面積で表されるようにする。
このとき，関数 $f(x)$ を X の **確率密度関数** といい，曲線 $y=f(x)$ を X の **分布曲線** という。



確率密度関数 $f(x)$ は，次のような性質をもつ。

・$\boldsymbol{f(x) \geqq 0}$

・$\boldsymbol{P(a \leqq X \leqq b) = \int_a^b f(x)\,dx}$

・**X のとる値の範囲が $\alpha \leqq X \leqq \beta$ のとき** $\boldsymbol{\int_\alpha^\beta f(x)\,dx = 1}$

連続型確率変数 X が特定の値をとる確率，例えば $P(X=k)$ は，$P(X=k)=\int_k^k f(x)\,dx=0$ である。

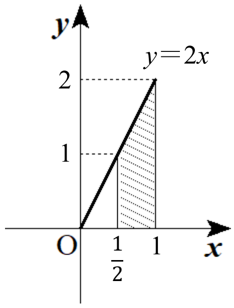
したがって，例えば $P(a \leqq X \leqq b)=P(a \leqq X < b)$ である。

解答

右の図から

$$P\left(\frac{1}{2} \leqq X \leqq 1\right) = \frac{1}{2}(1+2)\cdot\left(1-\frac{1}{2}\right)$$

$$= \boldsymbol{\frac{3}{4}}$$



$\boxed{8}$ **正規分布**

次の問いに答えよ。

(1) 確率変数 Z が標準正規分布 $N(0,\ 1)$ に従うとき，次の確率を求めよ。

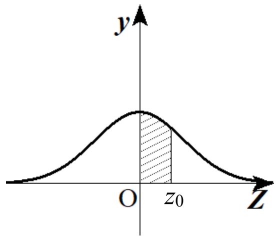
① $P(1 \leqq Z \leqq 2)$

② $P(Z \leqq -0.7)$

(2) 確率変数 X が正規分布 $N(6,\ 3^2)$ に従うとき，$P(3 \leqq X \leqq 6)$ を求めよ。

(3) 高校 2 年生 200 人が，目を閉じてストップウォッチをちょうど 1 分で止めるというゲームを 1 人 1 回したところ，平均タイムは 61.5 秒，標準偏差は 10.0 秒であった。

このタイムは正規分布に従うものとするとき，59 秒以上 61 秒以下で止めた生徒はおよそ何人いるか。



正規分布表

z_0	0	～	3	～	5	6	～	8
0.0	0.0000		0.0120		0.0199	0.0239		0.0319
～								
0.2	0.0793		0.0910		0.0987	0.1026		0.1103
～								
0.5	0.1915		0.2109		0.2088	0.2123		0.2190
～								
0.7	0.2580		0.2673		0.2734	0.2764		0.2823
～								
1.0	0.3413		0.3485		0.3531	0.3554		0.3599
～								
1.9	0.4713		0.4732		0.4744	0.4750		0.4761
2.0	0.4772		0.4788		0.4798	0.4803		0.4812
～								
2.3	0.4893		0.4901		0.4906	0.4909		0.4913
～								
2.5	0.4938		0.4943		0.4946	0.4948		0.4951

当該ファイルに関連のある部分を抜粋しています。

要　点

正規分布

連続型確率変数 X の確率密度関数 $f(x)$ が

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$$　（πは円周率，eは無理数でその値は 2.71828……，m，σは実数，$\sigma>0$）

で表されるとき，**X は正規分布 $N(m,\ \sigma^2)$ に従う** といい，$y=f(x)$のグラフを **正規分布曲線** という。

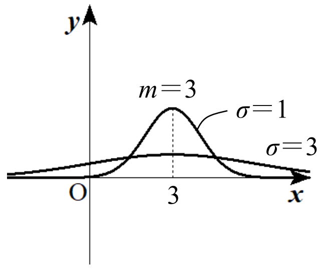
正規分布の平均・標準偏差

確率変数 X が正規分布 $N(m,\ \sigma^2)$に従うとき　　平均は　$\boldsymbol{E(X)=m}$，　　標準偏差は　$\boldsymbol{\sigma(X)=\sigma}$

正規分布曲線の性質

「$\boxed{7}$　**連続型確率変数**」で学習した分布曲線の性質のほかに，次の性質をもつ。

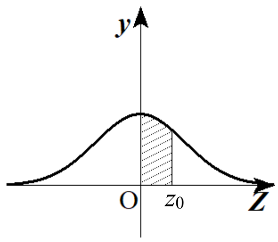
・直線 $x=m$（期待値）に関して左右対称な曲線であり，
　$f(x)$は $x=m$ で最大で，$x=m$ から遠ざかるにつれて
　減少し 0 に近づく。（x 軸が漸近線）
・標準偏差 σ が大きくなるほど，山が低くなって横に
　広がる。



〈注意〉 正規分布(normal distribution)は，偶然生じる誤差を伴う
　　　事象にみられる。例えば，工場で作られたねじの長さ
　　　やある果物の重さなどは正規分布に従うと考えてよい。
　　　正規分布は，ガウス分布と呼ばれることもある。

標準正規分布

平均 0，標準偏差 1 の正規分布 $N(0,\ 1)$を **標準正規分布** という。
確率変数 Z が標準正規分布 $N(0,\ 1)$に従うとき，Z が 0 と z_0 の間の値をとる確率 $P(0\leqq Z\leqq z_0)$は，右の図の斜線部分の面積に等しい。
問題文の横の正規分布表は，$P(0\leqq Z\leqq z_0)$の値をまとめたものである。



確率変数の標準化

確率変数Xが正規分布$N(m,\ \sigma^2)$に従うとき，$Z=\dfrac{X-m}{\sigma}$ とおくと，

確率変数 Z は標準正規分布 $N(0,\ 1)$に従う。

証明

$E(aX+b)=aE(X)+b$，　$V(aX+b)=a^2V(X)$から

$$E(Z)=E\left(\frac{X-m}{\sigma}\right)=E\left(\frac{X}{\sigma}-\frac{m}{\sigma}\right)=\frac{1}{\sigma}E(X)-\frac{m}{\sigma}=0$$　$\boxed{E(X)=m}$

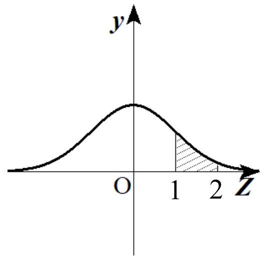
$$V(Z)=V\left(\frac{X-m}{\sigma}\right)=V\left(\frac{X}{\sigma}-\frac{m}{\sigma}\right)=\left(\frac{1}{\sigma}\right)^2V(X)=1$$　$\boxed{V(X)=\sigma^2}$

解答

(1)　①　$P(1 \leqq Z \leqq 2) = P(0 \leqq Z \leqq 2) - P(0 \leqq Z \leqq 1)$

$= 0.4772 - 0.3413$

$= \mathbf{0.1359}$

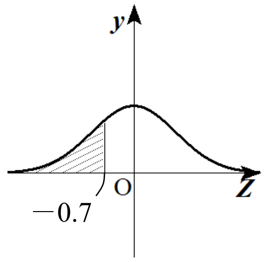


②　$P(Z \leqq -0.7) = P(Z \geqq 0.7)$

$= P(Z \geqq 0) - P(0 \leqq Z \leqq 0.7)$

$= 0.5 - 0.2580$

$= \mathbf{0.2420}$



(2)　$Z = \dfrac{X-6}{3}$ とおくと，確率変数Zは$N(0,\ 1)$に従う。

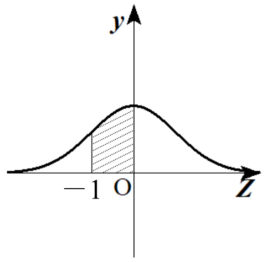
$X = 3$ のとき$Z = \dfrac{3-6}{3} = -1$，$X = 6$ のとき$Z = \dfrac{6-6}{3} = 0$

であるから

$P(3 \leqq X \leqq 6) = P(-1 \leqq Z \leqq 0)$

$= P(0 \leqq Z \leqq 1)$

$= \mathbf{0.3413}$



(3)　タイムをX秒とすると，Xは$N(61.5,\ 10.0^2)$に従うから　$Z = \dfrac{X-61.5}{10.0}$ とおくと，

Zは$N(0,\ 1)$に従う。

$X = 59$ のとき$Z = \dfrac{59-61.5}{10.0} = -0.25$，$X = 61$ のとき$Z = \dfrac{61-61.5}{10.0} = -0.05$

であるから

$P(59 \leqq X \leqq 61) = P(-0.25 \leqq Z \leqq -0.05)$

$= P(0.05 \leqq Z \leqq 0.25)$

$= P(0 \leqq Z \leqq 0.25) - P(0 \leqq Z \leqq 0.05)$

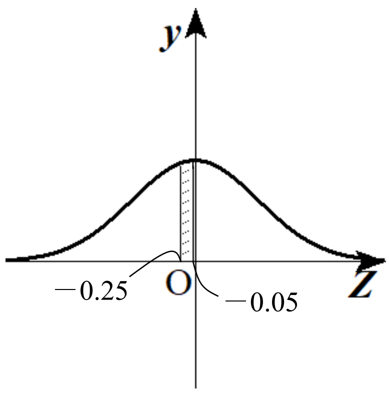
$= 0.0987 - 0.0199$

$= 0.0788$

よって，59 秒以上 61 秒以下で止めた生徒の人数は

$200 \times 0.0788 = 15.76$

したがって，およそ　**16 人**



$\boxed{9}$ **二項分布の正規分布による近似**

あるアイテムに対して, 排出確率が 10%に設定されたガチャがある。このガチャを 400 回まわしたとき, あるアイテムが 43 個以上得られる確率を求めよ。

要 点

確率変数 X が二項分布 $B(n,\ p)$に従っているとすると，n が十分大きいとき，X は正規分布に従うとみなしてよいことが知られている。

〈注意〉二項分布の極限分布は正規分布であるという定理は，ド・モアブル-ラプラスの定理と呼ばれている。

一般に，確率変数 X が二項分布 $B(n,\ p)$に従うとき，

$$\boldsymbol{E(X) = np,\quad V(X) = npq,\quad \sigma(X) = \sqrt{npq}}\qquad \text{ただし}\quad \boldsymbol{q = 1 - p}$$

であるから，次のことが成り立つ。

二項分布の正規分布による近似

確率変数 X が二項分布 $B(n,\ p)$に従っているとすると，n が十分大きいとき，X は正規分布 $N(np,\ npq)$に従うとみなしてよい。ただし，$q=1-p$ である。

さらに，この確率変数を標準化すると，次のようになる。

二項分布の標準正規分布による近似

確率変数 X が二項分布 $B(n,\ p)$に従い n が十分大きいとき，

$$\boldsymbol{Z = \frac{X - np}{\sqrt{np(1 - p)}}}$$

とおくと，Z は標準正規分布 $N(0,\ 1)$に従うとみなしてよい。

解答

あるアイテムが得られる個数をXとすると，Xは二項分布$B\left(400,\ \frac{1}{10}\right)$に従う。

よって，Xの平均mは　$m = 400 \cdot \frac{1}{10} = 40$(個)，　　標準偏差$\sigma$は　$\sigma = \sqrt{400 \cdot \frac{1}{10}\left(1 - \frac{1}{10}\right)} = 6$(個)

400 は十分大きいから，X は正規分布 $N(40,\ 6^2)$に従うとみなしてよい。

これより，$Z = \frac{X - 40}{6}$ とおくと，Zは標準正規分布$N(0,\ 1)$に従うとみなしてよい。

以上から，求める確率は

$P(X \geqq 43) = P(Z \geqq 0.5)$

$= P(Z \geqq 0) - P(0 \leqq Z \leqq 0.5)$

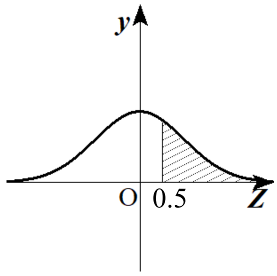
$= 0.5 - 0.1915$

$= \mathbf{0.3085}$

$X=43$ のとき

$$Z = \frac{43 - 40}{6} = 0.5$$

「$\boxed{8}$　正規分布」の正規分布表参照



9．5　母集団と標本

この項目は，例題，解答はなく，要点のみとする。

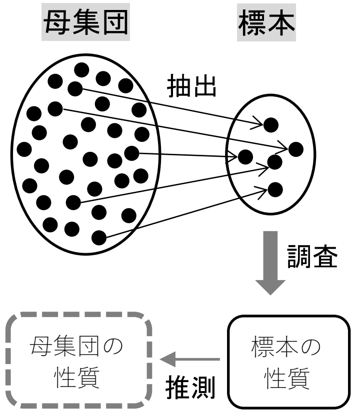
要　点

全数調査と標本調査

集団に対して統計調査をするとき，集団全体をもれなく調べる方法を **全数調査** といい，集団の一部を調べ，その結果から集団全体の性質を推測する方法を **標本調査** という。

一般的には，集団全体をもれなく調べることは難しいので，統計調査では主に標本調査が用いられる。

標本調査では，調査の対象となる集団全体を **母集団** といい，
母集団に含まれる要素の個数を **母集団の大きさ** という。
母集団から取り出された要素の集まりを **標本** といい，
標本に含まれる要素の個数を **標本の大きさ** という。
また，標本を取り出すことを **抽出する** という。



標本の抽出

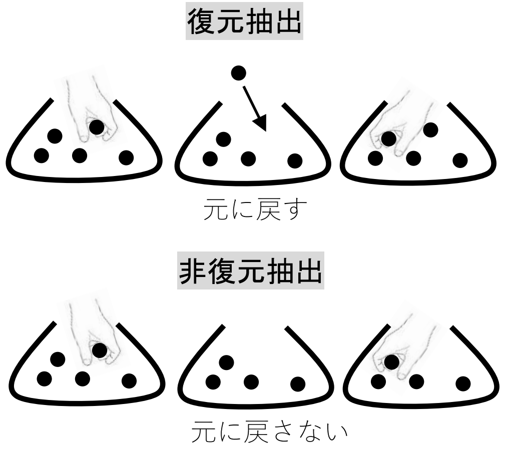
母集団の性質を正しく推測するには，母集団からかたよりなく，
すなわち等しい確率で標本を抽出する必要がある。
母集団の各要素を等しい確率で抽出する方法を **無作為抽出** といい，無作為抽出によって選ばれた標本を **無作為標本** という。

〈注意〉無作為抽出は，ランダムサンプリングともいう。

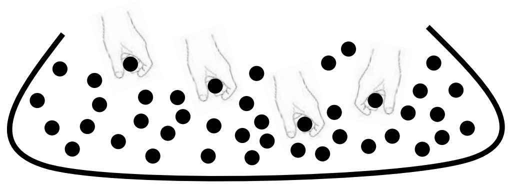
無作為抽出を行うには，母集団の要素に番号を付け，コンピュータで乱数を発生させ，その乱数に応じた要素を抽出すればよい。

復元抽出と非復元抽出

母集団から要素を 1 個取り出したら元に戻し，改めて
また 1 個取り出すことを繰り返す方法を **復元抽出**という。
これに対して，一度取り出した要素は元に戻さないで，
続けて取り出す方法を **非復元抽出** という。



標本の大きさ n に比べて母集団の大きさ N が十分大きいときは，復元抽出による無作為標本であっても，母集団から同じ要素が再び取り出されることは実際にはほとんど起こらないと考えられる。したがって，復元抽出による標本も非復元抽出による標本も同じとみなしてよい。
今後は，復元抽出する場合について考えることにする。



１０　**標本平均の分布**

次の問いに答えよ。

(1)　2 枚の硬貨を投げて表の出た枚数を記録することを 4 回繰り返す。この 4 回の試行の表の出た枚数をX_1，X_2，X_3，X_4とし，その平均を$\bar{X}=\dfrac{X_1+X_2+X_3+X_4}{4}$とする。$\bar{X}$の平均$E(\bar{X})$，標準偏差$\sigma(\bar{X})$を求めよ。

(2)　ある県の 17 歳女子の身長は，平均 158cm，標準偏差 5cm の正規分布に従うものとする。無作為に 25 人を抽出したとき，その標本平均$\bar{X}$が 160cm 以上である確率を求めよ。

要　点

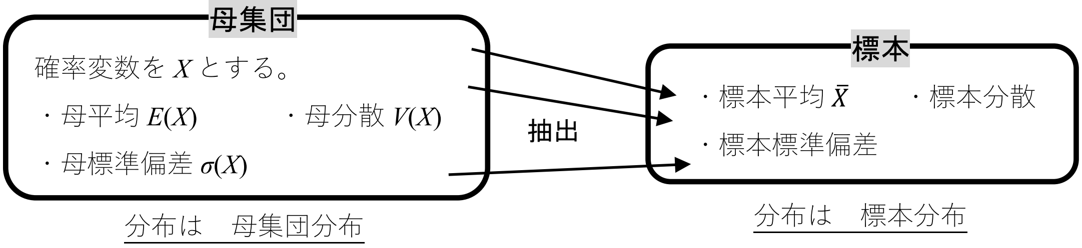
母集団の性質を調べるために標本調査を行う。母集団の平均が知りたいとき，大きさ 1 の標本から母集団の平均を推測することは難しいが，大きさ 4 の標本をとりその平均を求めると，それは母集団の平均と近いといえる。例えばサッカーのリフティングをするとき，1 回の結果が真の実力かどうかは推測しづらいが，4 回の結果の平均をとると，真の実力の値に近いと推測できる。

標本平均

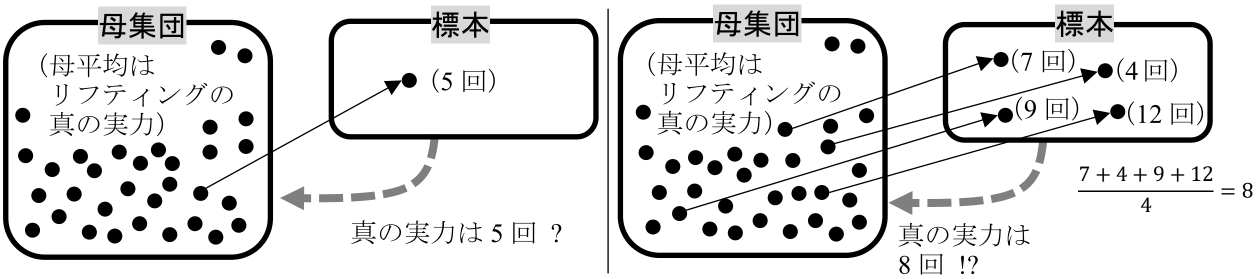
復元抽出によって母集団から無作為抽出した大きさ n の標本の要素の値を，X_1，X_2，X_3，……，X_n とする。この平均を **標本平均** といい，$\boldsymbol{\bar{X}}$ で表す。

$$\bar{X}=\frac{X_1+X_2+X_3+\cdots\cdots+X_n}{n}$$

母集団と標本で同じ用語を使うと不便なので，次のように使い分ける。



標本調査では，母平均や母標準偏差を推測することが目的となる。



〈注意〉標本調査は，全数調査が難しいときに少ない標本のデータから母集団の性質を推測する方法であるが， 標本のデータが多い方が母集団の性質はより正確に推測できるようになる。

標本平均 $\bar{X}$ の分布

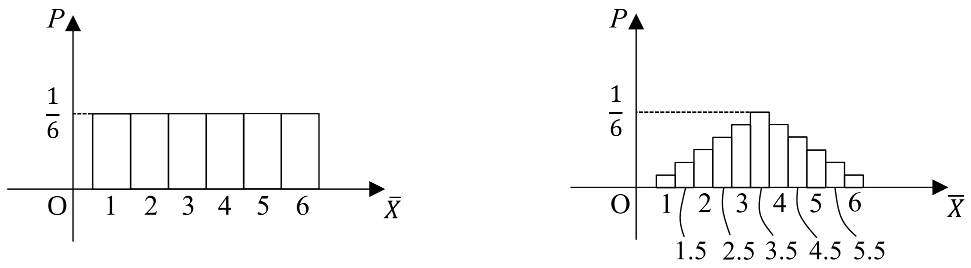
さいころを 1 回投げるときの出る目を X とすると，確率変数 X の確率分布は右のようになる。
標本平均$\bar{X}$は，抽出ごとに値が定まるから，確率変数と考えられる。

X	1	2	3	4	5	6	計
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

さいころを 2 回投げるときの出る目X_1，X_2の平均を$\bar{X}=\dfrac{X_1+X_2}{2}$とすると，確率変数$\bar{X}$の確率分布は次のようになる。

$\bar{X}$	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	計
P	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

大きさ 1 の標本平均$\bar{X}$の確率分布は母集団分布と一致する。
大きさ 1 の標本平均と大きさ 2 の標本平均の分布をグラフにすると次のようになる。



〈注意〉標本分布は 1 つの標本の中の要素の値の分布であるので，標本分布と標本平均$\bar{X}$の分布は違うものである。

このように，標本の大きさ n を大きくすると山形のグラフになり，母集団の平均をとる確率が相対的に高くなっている。また，散らばり具合も，標本の大きさ n が大きくなるほど母集団の平均付近に集まるようになるので，小さくなる。

一般に，標本平均$\bar{X}$について，次のことが成り立つ。

母平均 m，母標準偏差 σ の母集団から，大きさ n の標本を復元抽出するとき，
標本平均$\bar{X}$の平均$E(\bar{X})$，標準偏差$\sigma(\bar{X})$は

$$\boldsymbol{E(\bar{X})=m,\qquad \sigma(\bar{X})=\frac{\sigma}{\sqrt{n}}}$$

また一般に，母平均と母標準偏差をもつどのような母集団についても，標本平均$\bar{X}$の確率分布は，標本の大きさ n が大きくなると，正規分布に近づくことが知られている。

母平均 m，母標準偏差 σ の母集団から，大きさ n の標本を無作為抽出するとき，
nが大きいならば，標本平均$\bar{X}$の分布は，**正規分布$\boldsymbol{N\left(m,\ \frac{\sigma^2}{n}\right)}$**で近似できる。

標本平均$\bar{X}$を標準化した確率変数$Z=\dfrac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}}$は，近似的に**標準正規分布**$\boldsymbol{N(0,\ 1)}$に従う。

なお，母平均 m，母標準偏差 σ の母集団分布が正規分布 $N(m,\ \sigma^2)$のときは，この母集団から無作為に抽出された大きさnの標本平均$\bar{X}$は，nが大きくなくても，つねに正規分布$N\left(m,\ \dfrac{\sigma^2}{n}\right)$に従う。

解答

(1)　2 枚の硬貨を投げて表の出た枚数 X の確率分布は，右の表のようになる。

X	0	1	2	計
P	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	1

よって，X の平均 $E(X)$は

$$E(X)=0\times\frac{1}{4}+1\times\frac{2}{4}+2\times\frac{1}{4}=\frac{0+2+2}{4}=1\ （枚）$$

X の標準偏差 $\sigma(X)$は

$$\sigma(X)=\sqrt{E(X^2)-\{E(X)\}^2}=\sqrt{0^2\times\frac{1}{4}+1^2\times\frac{2}{4}+2^2\times\frac{1}{4}-1^2}=\sqrt{\frac{2}{4}}=\frac{\sqrt{2}}{2}\ （枚）$$

したがって，大きさ 4 の標本の標本平均$\bar{X}$の平均$E(\bar{X})$は　$\boldsymbol{E(\bar{X})}=$(母平均)$=$**1（枚）**

標準偏差$\sigma(\bar{X})$は　$\boldsymbol{\sigma(\bar{X})}=\dfrac{\sigma(X)}{\sqrt{4}}=\dfrac{1}{2}\cdot\dfrac{\sqrt{2}}{2}=\boldsymbol{\dfrac{\sqrt{2}}{4}}$ **（枚）**

(2)　母集団分布が正規分布$N(158,\ 5^2)$であるから，標本平均$\bar{X}$は正規分布$N\left(158,\ \dfrac{5^2}{25}\right)$に従う。

よって，標準化した確率変数$Z=\dfrac{\bar{X}-158}{\frac{5}{\sqrt{25}}}=\bar{X}-158$ は標準正規分布$N(0,\ 1)$に従う。

$\bar{X}=160$ のとき $Z=2$ であるから，求める確率は

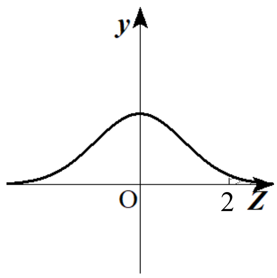
$P(\bar{X}\geqq160)=P(Z\geqq2)$

$=P(Z\geqq0)-P(0\leqq Z\leqq2)$

$=0.5-0.4772$

$=$**0.0228**

「$\boxed{8}$　正規分布」の
正規分布表参照



１１　母平均の推定

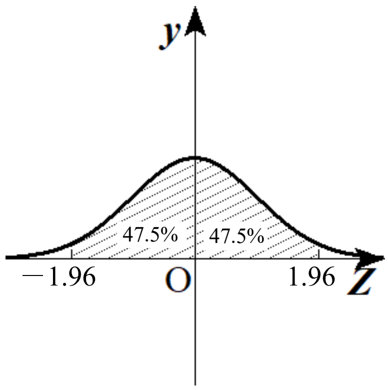
次の問いに答えよ。

(1)　スマートフォンのカメラ機能で，対象の物体の長さを表示するアプリがある。このアプリで測定した全データの実際の長さと表示された長さの母標準偏差は 10.0cm であり，正規分布に従うことが分かっているものとする。このアプリであるイスの高さをいろいろなアングルで 4 回表示させたところ，その標本平均は 70.0cm であった。このイスの実際の高さを信頼度 95%で区間推定せよ。

(2)　ある国家試験受験者の中から，100 人を無作為抽出して休日の勉強時間を調査したところ，標本平均が 10.0 時間，標本標準偏差が 4.0 時間であった。この国家試験受験者の平均勉強時間を，信頼度 95%で区間推定せよ。

要　点

母集団の要素の個数が多すぎて調べることが困難なときなどに，母平均がわからない場合がある。
そのような場合，標本平均から推定することを考える。
一般に，母平均 m，母標準偏差 σ をもつ母集団から，大きさ n の標本を無作為抽出するとき，その標本平均$\bar{X}$の分布は，nが大きいとき正規分布$N\left(m,\ \dfrac{\sigma^2}{n}\right)$で近似できる。$\bar{X}$を$Z=\dfrac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}}$により標準化すると，その分布は標準正規分布 $N(0,\ 1)$で近似できる。
ここで，「８　正規分布」の正規分布表により，
$P(-1.96 \leqq Z \leqq 1.96)=0.95$ が成り立つことがわかる。
これは，確率変数 Z が-1.96 から 1.96 の間の値をとる確率が 0.95 であることを意味する。



すなわち，$P\left(-1.96 \leqq \dfrac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} \leqq 1.96\right)=0.95$ であり

$$P\left(-1.96\times\frac{\sigma}{\sqrt{n}} \leqq \bar{X}-m \leqq 1.96\times\frac{\sigma}{\sqrt{n}}\right)=0.95,$$

よって，　$P\left(m-1.96\times\dfrac{\sigma}{\sqrt{n}} \leqq \bar{X} \leqq m+1.96\times\dfrac{\sigma}{\sqrt{n}}\right)=0.95$ である。

これは，母平均 m がわかっている母集団から無作為抽出したある 1 つの標本平均 $\bar{x}$ について，

$\bar{x}$ が$m-1.96\times\dfrac{\sigma}{\sqrt{n}}$から$m+1.96\times\dfrac{\sigma}{\sqrt{n}}$の間の値をとる確率が 0.95 であることを意味する。

「母平均の推定」では，観測された標本平均 $\bar{x}$ から未知の母平均 m を推定するので，

$\bar{x}$ が 95%の確率で入るような $m-1.96\times\dfrac{\sigma}{\sqrt{n}}$ から $m+1.96\times\dfrac{\sigma}{\sqrt{n}}$ の範囲となる m は，

母平均として 95%の妥当性があると考える。

$m-1.96\times\dfrac{\sigma}{\sqrt{n}}\leqq\bar{x}$ を m について解くと　$m\leqq\bar{x}+1.96\times\dfrac{\sigma}{\sqrt{n}}$

$\bar{x}\leqq m+1.96\times\dfrac{\sigma}{\sqrt{n}}$ を m について解くと　$m\geqq\bar{x}-1.96\times\dfrac{\sigma}{\sqrt{n}}$

である。このとき，$\boldsymbol{\bar{x}-1.96\times\dfrac{\sigma}{\sqrt{n}}\leqq m\leqq\bar{x}+1.96\times\dfrac{\sigma}{\sqrt{n}}}$　を母平均mに対する **信頼度95%の信頼区間** という。ある信頼度の信頼区間を求めることを **区間推定する** という。

〈注意〉$P(-2.58\leqq Z\leqq 2.58)=0.99$ が成り立つから，母平均 m の信頼度 99%の信頼区間は次のようになる。

$$\bar{x}-2.58\times\frac{\sigma}{\sqrt{n}}\leqq m\leqq\bar{x}+2.58\times\frac{\sigma}{\sqrt{n}}$$

(2)　抽出した 1 つの標本の標準偏差を **標本標準偏差** といい，$\boldsymbol{s}$ で表す。
実際の標本調査では，母集団の母標準偏差 σ は不明なことが多いが，一般に標本の大きさ n が大きいときには，母標準偏差 σ のかわりに，標本標準偏差 s を用いてもよいことが知られている。

解答

(1)　イスの実際の高さを m とする。
母標準偏差 $\sigma=10.0$，標本の大きさ $n=4$，標本平均 $\bar{x}=70.0$ であるから，m に対する信頼度 95% の信頼区間は　$70.0-1.96\times\dfrac{10.0}{\sqrt{4}}\leqq m\leqq 70.0+1.96\times\dfrac{10.0}{\sqrt{4}}$

すなわち　$60.2\leqq m\leqq 79.8$
したがって，イスの実際の高さは **60.2cm 以上 79.8cm 以下** と推定できる。

(2)　平均勉強時間を m とする。
標本の大きさ 100 は大きいので，母標準偏差 σ のかわりに標本標準偏差 $s=4.0$ を用いる。
標本平均 $\bar{x}=10.0$ であるから，m に対する信頼度 95% の信頼区間は

$$10.0-1.96\times\frac{4.0}{\sqrt{100}}\leqq m\leqq 10.0+1.96\times\frac{4.0}{\sqrt{100}}$$　すなわち　$9.216\leqq m\leqq 10.784$

したがって，平均勉強時間は **9.2 時間以上 10.8 時間以下** と推定できる。

１２　母比率の推定

ある県で無作為に 100 人抽出したとき，魚釣りの経験者は 10 人であった。ある県における魚釣り経験者の比率を，信頼度 95%で区間推定せよ。

要　点

ある性質をもつものの母集団全体に対する比率を **母比率**，標本全体に対する比率を **標本比率** という。
ある性質をもつ母比率が p である母集団から，大きさ n の標本を無作為抽出するとき，ある性質をもつ標本の要素の個数 X は二項分布 $B(n,\ p)$に従う確率変数である。

「$\boxed{9}$ **二項分布の正規分布による近似**」で学習したように，n が大きいとき，X は正規分布 $N(np,\ n(1-p))$ に従うとみなしてよい。

さらに，標準化した確率変数$Z=\dfrac{X-np}{\sqrt{np(1-p)}}$は，標準正規分布$N(0,\ 1)$に従うとみなしてよい。

ここで，標本比率 $\bar{p}$ は $\bar{p}=\dfrac{X}{n}$ である。

「$\boxed{11}$ **母平均の推定**」と同じように，標本比率 $\bar{p}$ から母比率 p を推定することを考える。
Z は標準正規分布 $N(0,\ 1)$ に従うとみなしてよいから，$P(-1.96 \leqq Z \leqq 1.96)=0.95$ ……①
より，p に対する信頼度 95%の信頼区間を考えることができる。

$Z=\dfrac{X-np}{\sqrt{np(1-p)}}$ であるが，この右辺の分母，分子をnで割ると $Z=\dfrac{\dfrac{X}{n}-p}{\sqrt{\dfrac{p(1-p)}{n}}}$

すなわち $Z=\dfrac{\bar{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}}$ これを①に代入すると $P\left(-1.96 \leqq \dfrac{\bar{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \leqq 1.96\right)=0.95$

ここで，ひとまず $\sqrt{\dfrac{p(1-p)}{n}}$ の中のpは無視して，$\dfrac{\bar{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}}$ の分子の p について不等式を解くことを考える。（理由は後述）

$-1.96 \leqq \dfrac{\bar{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}}$ より $-1.96\sqrt{\dfrac{p(1-p)}{n}} \leqq \bar{p}-p$ よって $p \leqq \bar{p}+1.96\sqrt{\dfrac{p(1-p)}{n}}$

$\dfrac{\bar{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \leqq 1.96$ より $\bar{p}-p \leqq 1.96\sqrt{\dfrac{p(1-p)}{n}}$ よって $p \geqq \bar{p}-1.96\sqrt{\dfrac{p(1-p)}{n}}$

したがって，①は次のように変形できる。 $P\left(\bar{p}-1.96\sqrt{\dfrac{p(1-p)}{n}} \leqq p \leqq \bar{p}+1.96\sqrt{\dfrac{p(1-p)}{n}}\right)=0.95$

「母比率の推定」では，ここから特殊な見方をするので先に結論を示すと，n が十分大きいとき，

かっこの中の不等式の最左辺と最右辺の $\sqrt{\dfrac{p(1-p)}{n}}$ の p を $\bar{p}$ に置き換えてよいことが知られている。

すなわち $P\left(\bar{p}-1.96\sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}} \leqq p \leqq \bar{p}+1.96\sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}}\right)=0.95$

以上から，母集団から大きさ n の標本を無作為抽出し，その標本比率が $\bar{p}$ であり，n が十分大きいとき，

母比率 p の信頼度 95%の信頼区間は　$\boldsymbol{\bar{p}-1.96\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \leqq p \leqq \bar{p}+1.96\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}}$

〈注意〉$P(-2.58 \leqq Z \leqq 2.58)=0.99$ が成り立つから，母比率 p の信頼度 99%の信頼区間は次のようになる。

$$\bar{p}-2.58\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \leqq p \leqq \bar{p}+2.58\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

［特殊な見方をしてよい理由］

$\bar{p}-1.96\sqrt{\frac{p(1-p)}{n}} \leqq p \leqq \bar{p}+1.96\sqrt{\frac{p(1-p)}{n}}$ という不等式は，$\bar{p}$ を定数として

連立不等式 $\begin{cases} \bar{p}-1.96\sqrt{\frac{p(1-p)}{n}} \leqq p \\ p \leqq \bar{p}+1.96\sqrt{\frac{p(1-p)}{n}} \end{cases}$ を解いて p の範囲を求めることができるが，nが十分大きい

とき，この求めた p の範囲と　$\bar{p}-1.96\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \leqq p \leqq \bar{p}+1.96\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$　は大差ないことが知られ

ている。よって，要点のような特殊な見方をして，標本比率 $\bar{p}$ から母比率 p の推定ができる。

解答

標本比率 $\bar{p}$ は　$\frac{10}{100}=0.1$　であるので，母比率 p の信頼度 95%の信頼区間は

$$0.1-1.96\sqrt{\frac{0.1\times0.9}{100}} \leqq p \leqq 0.1+1.96\sqrt{\frac{0.1\times0.9}{100}}$$

すなわち　$0.0412 \leqq p \leqq 0.1588$

したがって，ある県における魚釣り経験者の比率は，

4%以上 16%以下　と推定できる。

コメント

母比率の推定では，信頼区間を求める際，n が十分大きいとき，母比率 p を標本比率$\bar{p}$で代用する場面があった。（母平均の推定でも，母標準偏差 σ を標本標準偏差 s で代用する場面があった。）
これは特殊な見方であることを説明したが，本問のような場合には標本比率 10%から母比率も 10%とすればよく，母比率の信頼度 95%の信頼区間を 4%以上 16%以下と推定することに意味を見出せない，と考える人もいるかもしれない。
標本の大きさ n がいくら大きいとはいえ，母集団の大きさ N に比べればはるかに小さいことが一般的である。このような標本調査において，信頼区間の端の方の比率になることも十分起こり得る。
本問でいえば，例えばある県において 3000 人が参加するイベントを開催するとき，釣り堀コーナーに参加者の 10%にあたる 300 人分のエサを用意するか，念のため 15%にあたる 450 人分のエサを用意するかの判断に影響する。300 人分では足りない可能性もあるが，1000 人分を用意すればほぼ確実に余ることが予想される。このような場面で母比率の推定は役に立つ。

１３　仮説検定の方法

ある工場で作られている高級ジャムの内容量は，平均 300g，標準偏差 5.5g の正規分布に従うことがわかっている。ある日，100 個のジャムを無作為に選んで調べると，内容量の平均が 298.7g であった。このとき，次の問いに答えよ。

(1)　この日のジャムは正常に作られていると判断してよいか。有意水準 5%で検定せよ。また，有意水準 1%で検定せよ。

(2)　この工場関係者は，ジャムが 300g より少ないかどうかに，より強い関心がある。この日のジャムの内容量は 300g より少ないと判断してよいか。有意水準 1%で検定せよ。

要　点

ある標本調査において，極端な結果が出たとする。このとき，偶然に起こったと考えることもできるが，母集団がもっていると考えられている性質が誤りではないか，と疑うことも考えられる。
母集団がもっていると考えられている性質に対して，疑わしいと考えられる事柄を **対立仮説** といい，対立仮説に対して，母集団がもっていると考えられる性質を **帰無仮説** または単に **仮説** という。
あらかじめ定めておく，めったに起こらないと判断する確率を **有意水準** という。有意水準は 5%または 1%とすることが多い。
〈注意〉「有意」水準とは，この基準を下回ると偶然ではなく「意味が有る」と考える基準のことである。
帰無仮説が成り立っていると仮定した状況で，極端な結果が起こる確率を p とする。p と有意水準を比較して，仮説が成り立っていると考えられるかどうかを判断することを， **仮説検定** または単に **検定** という。p が有意水準を下回り，仮説が成り立っていないと判断することを **棄却する** という。（仮説が棄却できないとき，仮説を採択するというが，設定した有意水準を上回ったにすぎないので仮説が成り立っていると断定するまでには至らず，単に棄却できないという状態であることに留意する。）

仮説検定において，確率変数が正規分布に従う連続的な値をとるとき，有意水準 p_0 に対して
立てた仮説のもとでは実現しにくい確率変数の値の範囲を，その範囲の確率が p_0 になるように定める。
この範囲を有意水準 p_0 の **棄却域** といい，実現した確率変数の値が棄却域に入れば仮説を棄却する。
以上をまとめると，一般に仮説検定の手順は次のようになる。

$\boxed{1}$ **あるめったに起こらない事象 A が起こった場合，対立仮説 H_1 を立てる。**

$\boxed{2}$ **H_1 に対する帰無仮説 H_0 を立てる。**

$\boxed{3}$ **あらかじめ定めている有意水準 p_0 と H_0 から棄却域を求める。**

$\boxed{4}$ **標本から得られた確率変数の値が棄却域に入るかどうかを調べ，仮説 H_0 が棄却できるかどうかを判断する。**

解答

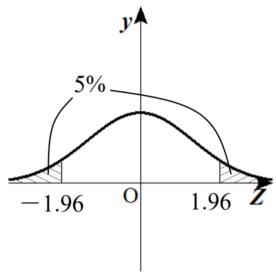
(1)　仮説を「ジャムは正常に作られている。」とする。
ジャムの内容量を Xg とおくと，X は正規分布
$N(300,\ 5.5^2)$に従うので，標本平均$\bar{X}$は
$N\left(300,\ \dfrac{5.5^2}{100}\right)$に従う。ここで，$Z=\dfrac{\bar{X}-300}{\dfrac{5.5}{\sqrt{100}}}$ と
おくと，確率変数 Z は標準正規分布 $N(0,\ 1)$に従う。
有意水準 5%の棄却域は $Z\leqq-1.96$，$1.96\leqq Z$ であるから

$$\bar{X}\leqq 300-1.96\times\frac{5.5}{\sqrt{100}},\qquad 300+1.96\times\frac{5.5}{\sqrt{100}}\leqq\bar{X}$$

すなわち　$\bar{X}\leqq 298.922$，$301.078\leqq\bar{X}$
観測された標本平均$\bar{x}=298.7$ はこの区間に入るので，
仮説は棄却される。
したがって，**有意水準 5%では，** この日のジャムは
正常に作られていないと判断できる。

詳しくは
対立仮説「ジャムは正常に作られていない。」，
帰無仮説「ジャムは正常に作られている。」
である。

「$\boxed{8}$　正規分布」の
正規分布表参照

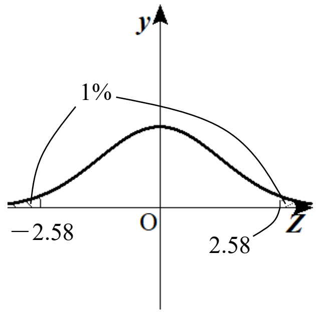


また，有意水準 1%の棄却域は $Z\leqq-2.58$，$2.58\leqq Z$ であるから

$$\bar{X}\leqq 300-2.58\times\frac{5.5}{\sqrt{100}},\qquad 300+2.58\times\frac{5.5}{\sqrt{100}}\leqq\bar{X}$$

すなわち　$\bar{X}\leqq 298.581$，$301.419\leqq\bar{X}$
観測された標本平均$\bar{x}=298.7$ はこの区間に入らないので，
仮説は棄却されない。
したがって，**有意水準 1%では，** この日のジャムは
正常に作られていないとはいえない。

「$\boxed{8}$　正規分布」の
正規分布表参照



(2) 仮説を「ジャムの内容量は 300g 以上である。」とする。

ジャムの内容量を Xg とおくと，X は正規分布 $N(300,\ 5.5^2)$に従うときを調べればよい。

このとき，標本平均$\bar{X}$は $N\left(300,\ \dfrac{5.5^2}{100}\right)$に従う。

ここで，$Z=\dfrac{\bar{X}-300}{\frac{5.5}{\sqrt{100}}}$とおくと，確率変数$Z$は

標準正規分布 $N(0,\ 1)$に従う。

「ジャムの内容量は 300g である。」

の有意水準 1%の棄却域は

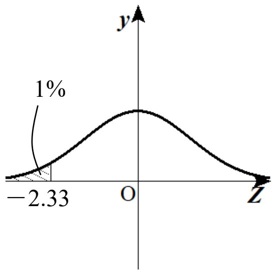
$Z\leqq-2.33$

であるから

「$\boxed{8}$ 正規分布」の
正規分布表参照

$$\bar{X}\leqq 300-2.33\times\frac{5.5}{\sqrt{100}}$$

すなわち $\bar{X}\leqq298.7185$



観測された標本平均 $\bar{x}=298.7$ はこの区間に入るので，仮説は棄却される。

したがって，有意水準 1%では，この日のジャムは **300g より少ないと判断できる。**

詳しくは

対立仮説「ジャムの内容量は 300g より少ない。」，

帰無仮説「ジャムの内容量は 300g 以上である。」

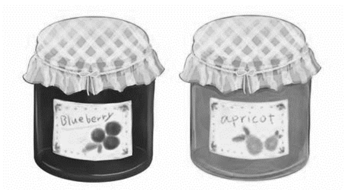
である。

300g 以上という仮説が棄却できるかどうかは，平均を 300g としたときの分布を調べればよい。

（例えば，帰無仮説が「ジャムの内容量は 302g である。」のもとで平均が 298.7g になる確率は，仮説「ジャムの内容量は 300g である。」のもとで 298.7g になる確率より低い。

すなわち，仮説「ジャムの内容量は 300g である。」のもとで得られた確率変数の値が棄却域に入れば，仮説「ジャムの内容量は 300g 以上である。」のもとで得られた同じ値はすべて棄却域に入る。

よって，300g 以上という仮説が棄却できるかどうかは，平均を 300g としたときの分布を調べればよい。）



〈注意〉 仮説検定において，確率分布の両側に棄却域があるものを **両側検定** ，片側に棄却域があるものを **片側検定** という。