

# 第4章 統計

## I. 二項分布・正規分布

基礎問

67.

▶二項分布◀

二項分布  $B(n, p)$  ( $n \geq 1, 0 < p < 1$ ) に従う確率変数  $X$  の期待値(平均)  $E(X)$  は  $np$  に等しい。このことの証明を述べよ。

(奈良県医大)

精講

1° 同じ条件のもとでくり返し行われる  $n$  回の独立試行において、事象  $A$  の起こる確率を  $p$ 、 $A$  が起こる回数を  $X$  とすれば、 $X$  は確率変数となり、その確率分布は

$$P(X=r) = {}_n C_r p^r (1-p)^{n-r} \quad (\text{独立試行の定理})$$

となる(→基礎問 41)。

このとき、確率変数  $X$  は“二項分布  $B(n, p)$  に従う”といった言い方をする。 $B(n, p)$  の  $B$  は Binomial Distribution(二項分布)の頭文字である。

2° 二項分布の確率  ${}_n C_r p^r q^{n-r}$  ( $q=1-p$ ) は  $(p+q)^n$  の展開式(二項展開)  ${}_n C_0 q^n + {}_n C_1 p q^{n-1} + \cdots + {}_n C_r p^r q^{n-r} + \cdots + {}_n C_n p^n$  の項であって、 $p+q=1$  であるから

$$\text{上式} = \sum_{r=0}^n {}_n C_r p^r q^{n-r} = (p+q)^n = 1$$

また、 $r {}_n C_r = n {}_{n-1} C_{r-1}$  ( $r \geq 1$ ) についても、一度確認しておく(→基礎問 23 研究)。

これらは本問の証明でも使われる重要な関係式である。

3° 確率変数  $X$  が二項分布  $B(n, p)$  に従うとき

$$\text{期待値} \quad E(X) = np$$

$$\text{分散} \quad V(X) = npq, \quad \text{ただし, } q = 1-p$$

$$\text{標準偏差} \quad \sigma = \sqrt{npq}$$

となることは覚えておくこと。このまま使ってもよい公式である。研究も一度は目を通しておくこと。

**解答** 確率変数  $X$  は二項分布  $B(n, p)$  に従うから、 $X=r$  となる確率  $P_r$  は

$$P_r = {}_n C_r p^r q^{n-r} \quad (q=1-p)$$

$$\begin{aligned} \text{期待値 } E(X) &= \sum_{r=0}^n r P_r = \sum_{r=0}^n r {}_n C_r p^r q^{n-r} \\ &= \sum_{r=1}^n n {}_{n-1} C_{r-1} p^r q^{n-r} \\ &= np \sum_{k=0}^{n-1} {}_{n-1} C_k p^k q^{n-1-k} \quad (k=r-1 \text{ と置いた}) \\ &= np(p+q)^{n-1} \quad (\text{二項定理}) \\ &= np \end{aligned}$$

**研究** 1° 二項分布  $B(n, p)$  の期待値は次のように考えると簡明である。

確率変数  $X$  は事象  $A$  の起こる回数であるから、 $r$  回の試行で事象  $A$  が起こるか起らないかを

$$X_r = \begin{cases} 1 & (A \text{ が起こる}) \\ 0 & (A \text{ が起らない}) \end{cases} \quad (1 \leq r \leq n)$$

で表すことにすると

$$X = X_1 + X_2 + \cdots + X_n$$

となる。これに  $E(X+Y) = E(X) + E(Y)$ ，“和の平均は平均の和”という性質を使うと  $E(X_r) = 1 \cdot p + 0 \cdot q = p$  より

$$\begin{aligned} E(X) &= E(X_1 + X_2 + \cdots + X_n) \\ &= E(X_1) + E(X_2) + \cdots + E(X_n) = np \end{aligned}$$

2° 分散  $V(X)$  についても調べよう。 $n \geq 2$  のとき

$$\begin{aligned} r^2 {}_n C_r &= \{r(r-1) + r\} {}_n C_r \\ &= n(n-1) {}_{n-2} C_{r-2} + n {}_{n-1} C_{r-1} \end{aligned}$$

が成り立つ。 $r(r-1) {}_n C_r = n(n-1) {}_{n-2} C_{r-2}$  は計算でも示されるが、 $r {}_n C_r = n {}_{n-1} C_{r-1}$  と同じように組合せの意味から説明することもできる。基礎問 23 の研究をみるとよい。これを使うと

$$\begin{aligned} E(X^2) &= \sum_{r=0}^n r^2 P_r = \sum_{r=0}^n \{r(r-1) + r\} {}_n C_r p^r q^{n-r} \\ &= \sum_{r=2}^n n(n-1) {}_{n-2} C_{r-2} p^r q^{n-r} + \sum_{r=1}^n n {}_{n-1} C_{r-1} p^r q^{n-r} \end{aligned}$$

$$\begin{aligned}
 &= n(n-1)p^2(p+q)^{n-2} + np(p+q)^{n-1} \\
 &= n(n-1)p^2 + np \\
 &= n^2p^2 + np(1-p) \\
 &= n^2p^2 + npq
 \end{aligned}$$

$$\begin{aligned}
 \therefore V(X) &= E(X^2) - \{E(X)\}^2 \\
 &= (n^2p^2 + npq) - (np)^2 = npq
 \end{aligned}$$

3° 1の方法をとるなら

$$\begin{aligned}
 E(X_r^2) &= 1^2 \cdot p + 0^2 \cdot q = p \\
 V(X_r) &= E(X_r^2) - \{E(X_r)\}^2 = p - p^2 \\
 &= p(1-p) = pq
 \end{aligned}$$

各  $X_r$  は互いに独立であるから

$$\begin{aligned}
 V(X) &= V(X_1 + X_2 + \cdots + X_n) \\
 &= V(X_1) + V(X_2) + \cdots + V(X_n) = npq
 \end{aligned}$$

4° 他に,  $(q+px)^n = \sum_{r=0}^n {}_n C_r p^r q^{n-r} x^r$  の微分を利用する方法もある。

$$np(q+px)^{n-1} = \sum_{r=0}^n r {}_n C_r p^r q^{n-r} x^{r-1} \quad \cdots (*)$$

$$x=1 \text{ とおくと } np = \sum_{r=0}^n r {}_n C_r p^r q^{n-r} = E(X)$$

(\*)  $\times x$  を微分すると

$$n(n-1)p^2(q+px)^{n-2}x + np(q+px)^{n-1} = \sum_{r=0}^n r^2 {}_n C_r p^r q^{n-r} x^{r-1}$$

$x=1$  とおくと

$$n(n-1)p^2 + np = \sum_{r=0}^n r^2 {}_n C_r p^r q^{n-r} = E(X^2)$$

$$\therefore V(X) = E(X^2) - (E(X))^2 = npq$$

▶ 演習 ◀

[55] あるサイコロを180回ふったら, 1の目が40回出た。この回数から, 1の目が出る回数の期待値を引き, 標準偏差で割った値を求めよ。

(自治医大)

[56] 平均値が6, 分散が2の二項分布に従う確率変数を  $X$  とする。  $X$

$=k$  となる確率  $P_k$  とおく。  $\frac{P_4}{P_3}$  の値を求めよ。

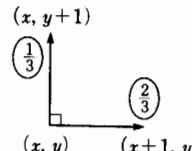
(弘前大)

基礎問 **68.** ▶ 二項分布 ◀

座標平面上の原点から出発する動点  $P(x, y)$  はサイコロを投げて 1, 2, 3, 4 の目が出ると  $x$  軸の正の方向に 1 だけ、5, 6 の目が出ると  $y$  軸の正の方向に 1 だけ動くものとする。サイコロを  $n$  回投げるとき、次の問いに答えよ。

- (1)  $X=x$  を確率変数とする。  $X$  の確率分布、平均値および分散をそれぞれ求めよ。
- (2) (i)  $Z=x-y$  を確率変数とする。(1)を用いて、  $Z$  の確率分布を求めよ。
- (ii)  $Z$  の平均値を求めよ。(iii)  $Z$  の分散を求めよ。(島根医大)

**精講** (1)  $X$  は二項分布  $B\left(n, \frac{2}{3}\right)$  に従う。(2)  $x+y=n$  より  $Z=x-y=2x-n$  となる。あとは、平均、分散の公式を使う。



**解答** (1) 確率分布  $P(X=x) = {}_n C_x \left(\frac{2}{3}\right)^x \left(\frac{1}{3}\right)^{n-x} = {}_n C_x \frac{2^x}{3^n}$  (答)

平均値  $E(X) = \frac{2}{3}n$ , 分散  $V(X) = \frac{2}{9}n$  ……(答)

(2) (i)  $x+y=n$  より  $Z=x-y=2x-n$

$P(Z=2x-n) = P(X=x) = {}_n C_x \frac{2^x}{3^n}$  ……(答)

(ii)  $E(Z) = 2E(X) - n = \frac{n}{3}$  (答) (iii)  $V(Z) = 2^2 V(X) = \frac{8}{9}n$  (答)

## ▶ 演習 ◀

[57] 確率変数  $X$  は二項分布  $B(n, p)$  に従い、その分散は  $\frac{8}{9}$  で、  $X$  が  $n-1$  をとる確率は  $X$  が  $n$  をとる確率の 8 倍である。

- (1) この二項分布の式を求めよ。
- (2) この二項分布の平均値を  $m$ , 標準偏差を  $\sigma$  とするとき、確率  $P(|X-m| > \sigma)$  を求めよ。
- (3)  $X^2 - 3X + 2$  の期待値を求めよ。(千葉大)

基礎問

69.

▶ 正規分布曲線 ◀

- (1) 標準正規分布の曲線  $y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  の増減を調べ、かつ変曲点を求め、このグラフの概形をかけ。
- (2)  $X$  は、平均値が  $m$ 、標準偏差が  $\sigma$  の正規分布に従う確率変数とする。確率  $P(\alpha \leq X \leq \beta)$  を、関数  $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  の定積分の形で表せ。ただし、 $\alpha < \beta$  とする。 (岩手大)

連続的確率変数  $X$  の確率密度関数  $f(x)$  が

精講

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

であるとき、確率変数  $X$  は“正規分布  $N(m, \sigma^2)$  に従う”という。 $N(m, \sigma^2)$  の  $N$  は Normal Distribution (正規分布) の頭文字である。ここで、 $m, \sigma (> 0)$  は定数、 $e$  は自然対数である。

このとき、

$$f(x) > 0, \int_{-\infty}^{\infty} f(x) dx = 1$$

が確かめられ(これは  $f(x)$  が確率密度関数であるための条件、第2の積分計算は高校の範囲を越える。)

$$\text{平均値 } E(X) = \int_{-\infty}^{\infty} xf(x) dx = m$$

$$\text{分散 } V(X) = \int_{-\infty}^{\infty} (x-m)^2 f(x) dx = \sigma^2$$

も示される(→ 確率・統計標準問題精講)。これらは認めることにしよう。

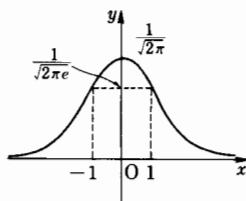
**解答** (1)  $y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  は  $y$  軸に関して対称であるから、 $x \geq 0$

で考えてよい。

$$y' = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot (-x) = -\frac{1}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}}$$

$$y'' = -\frac{1}{\sqrt{2\pi}} (1-x^2) e^{-\frac{x^2}{2}}$$

$x$	0		1		$(+\infty)$
$y'$	0	-		-	
$y''$		-	0	+	
$y$	$\frac{1}{\sqrt{2\pi}}$	$\searrow$	$\frac{1}{\sqrt{2\pi e}}$	$\searrow$	(0)



グラフの概形は右のようになる。

(2) 平均値  $m$ , 標準偏差  $\sigma$  の正規分布曲線は

$$y = f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$$\therefore P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

$\frac{X-m}{\sigma} = x$  とおき, 置換積分すると,  $dX = \sigma dx$  ゆえ,

$$P(\alpha \leq X \leq \beta) = \int_{\frac{\alpha-m}{\sigma}}^{\frac{\beta-m}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad \dots\dots(\text{答})$$

### 1° 標準正規分布曲線

## 研究

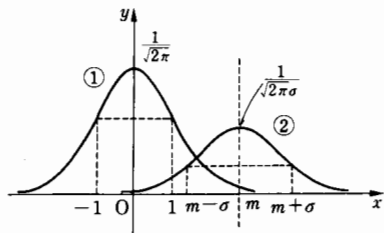
$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \dots\dots\textcircled{1}$$

は, 正規分布曲線

$$y = f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad \dots\dots\textcircled{2}$$

の  $m=0, \sigma=1$  の特別な場合である。

①を  $x$  軸方向に  $\sigma$  倍,  $y$  軸方向に  $\frac{1}{\sigma}$  倍し, さらに  $x$  軸方向に  $m$  だけ平行移動すると②が得られる。このことより①で知ったグラフの性質を②に言いかえると, ②は



1) 直線  $x=m$  に関して対称。

2)  $x$  軸を漸近線にもつ。

3)  $x=m$  で最大となり、最大値は  $\frac{1}{\sqrt{2\pi}\sigma}$  である。

4)  $x=m\pm\sigma$  で変曲点となる。

2° 本問(2)は、確率変数  $X$  が正規分布  $N(m, \sigma^2)$  に従うとき、

$$Z = \frac{X-m}{\sigma} \quad \text{とおくと、} Z \text{ は標準正規分布 } N(0, 1)$$

に従う確率変数となることを示している。確かに

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

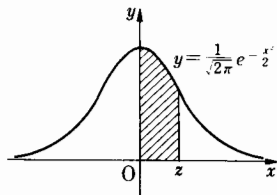
$$V(Z) = V\left(\frac{X-m}{\sigma}\right) = \frac{1}{\sigma^2}V(X) = 1$$

であり、この置き換えを“**確率変数の標準化**”という。

3° したがって、標準正規分布  $N(0, 1)$  のすべての値がわかれば、どんな正規分布  $N(m, \sigma^2)$  でも、標準化することにより、その確率を計算することができる。

右図の斜線部分の面積

$$P(0 \leq Z \leq z) = \int_0^z f(x) dx$$



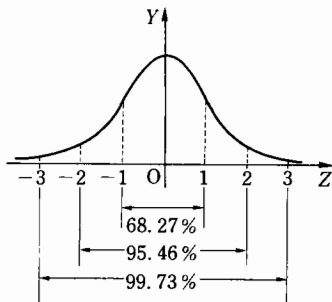
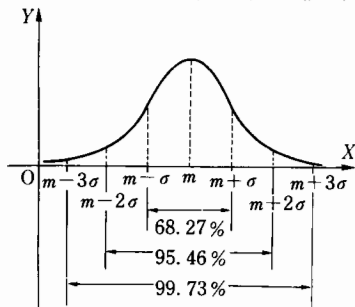
を計算してまとめたものが**正規分布表**(→ 167 ページ)である。

とくに、次の確率はよく使われる。

$$P(|X-m| \leq \sigma) = P(|Z| \leq 1) = 0.6827$$

$$P(|X-m| \leq 2\sigma) = P(|Z| \leq 2) = 0.9546$$

$$P(|X-m| \leq 3\sigma) = P(|Z| \leq 3) = 0.9973$$



## 基礎問

## 70.

## 標準化

- (1)  $Z$  を標準正規分布  $N(0, 1)$  に従う確率変数とすると、  
 $-1.96 \leq Z \leq 1.96$  となる確率は 0.95 である。 $X$  を平均値が 50、  
 分散が 25 の正規分布  $N(50, 25)$  に従う確率変数とすると、  
 $-a \leq X - 50 \leq a$  となる確率が 0.95 となるような  $a$  の値を求めよ。  
 (弘前大)

- (2) 平均 0、標準偏差 1 の正規分布 [0 から  $x$  までの確率を  $F(x)$  で示す] の一部を右に示す。1000 人の集団の身長は正規分布することが知られているとする。しかも平均が 167 cm で標準偏差が 5 cm であるという。この集団において、175 cm 以上の人はほぼ何人ぐらいいるか。また、157 cm 以上の人はほぼ何人ぐらいいるか。  
 (和歌山県医大)

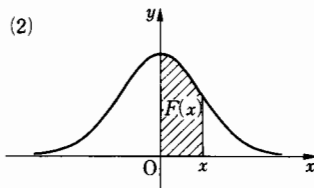
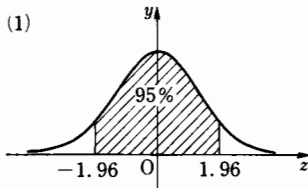
$x$	$F(x)$
1.4	0.4192
1.5	0.4332
1.6	0.4452
1.7	0.4554
1.8	0.4641
1.9	0.4713
2.0	0.4773

前問の研究 2° で説明したように、確率変数  $X$  が正規分布  $N(m, \sigma^2)$  に従うとき、置き換え

$$Z = \frac{X - m}{\sigma}$$

をおこなうことにより、標準正規分布  $N(0, 1)$  に従う確率変数に変形することができる。あとは正規分布表を使うだけである。

(1) の確率 0.95, (2) の確率  $F(x)$  はそれぞれ下の面積を表す。



解答 (1)  $X$  が  $N(50, 5^2)$  に従うから

$$Z = \frac{X - 50}{5}$$

と置くと、 $Z$  は  $N(0, 1)$  に従う。



$$P(-1.96 \leq Z \leq 1.96) = 0.95$$

$$P(-1.96 \times 5 \leq X - 50 \leq 1.96 \times 5) = 0.95$$

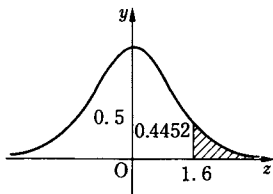
$$\therefore a = 1.96 \times 5 = 9.8 \quad \dots\dots(\text{答})$$

(2) 身長  $X$  の確率分布は  $N(167, 5^2)$  に従う。

$$Z = \frac{X - 167}{5}$$

と置くと、 $Z$  は  $N(0, 1)$  に従うから

$$\begin{aligned} P(X \geq 175) &= P\left(Z \geq \frac{175 - 167}{5}\right) \\ &= P(Z \geq 1.6) \\ &= 0.5 - 0.4452 = 0.0548 \end{aligned}$$

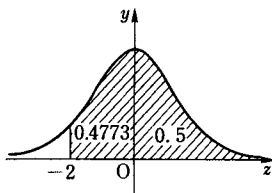


175 cm 以上の人、1000 人のうちの 5.48 % であるから

ほぼ 55 人 \dots\dots(\text{答})

また、157 cm 以上の人

$$\begin{aligned} P(X \geq 157) &= P\left(Z \geq \frac{157 - 167}{5}\right) \\ &= P(Z \geq -2) \\ &= 0.5 + 0.4773 \\ &= 0.9773 \end{aligned}$$



ほぼ 977 人 \dots\dots(\text{答})

▶ 演習 ◀

[58] A と B が、跳んだ距離を競う競技会に参加している。B が跳ぶ距離の確率分布はほぼ正規分布をなすものとする。B が最近参加した 20 回の競技会で跳んだ距離の記録  $x_i (i=1, 2, \dots, 20; \text{単位は m})$  を調べたら、 $\sum_{i=1}^{20} x_i = 107.00$ ,  $\sum_{i=1}^{20} x_i^2 = 572.90$

であった。いま先に跳んだ A の記録が 5.65 m であったとき、A が B に勝つ確率はおよそいくらか。(山梨医大)

[59] ある製品の長さは平均 69 cm, 標準偏差 0.4 cm の正規分布にしたがうことがわかっている。長さが 70 cm 以上の製品は不良品とされるとき、1 万個の製品の中には何個の不良品が含まれると予想されるか。正規分布表(省略)を利用せよ。(琉球大)

## 基礎問

## 71.

二項分布の正規近似

ある国では、その国民の血液型の割合は、O型30%、A型35%、B型25%、AB型10%であるといわれている。いま無作為に400人を選ぶとき、AB型の人が37人以上49人以下となる確率を求めよ。ただし、 $N(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$  の値は、次の表で与えられる。

$x$	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
$N(x)$	.1554	.1915	.2257	.2580	.2881	.3159	.3413	.3643	.3849	.4032	.4192	.4332

(旭川医大)

確率変数 $X$ が二項分布  $B(n, p)$ 

## 精講

$$P_r = {}_n C_r p^r q^{n-r} \quad (q=1-p)$$

に従うとき、

$$\text{平均値 } m=np, \text{ 分散 } \sigma^2=npq$$

であった。このとき、 $n$ を十分大きくとると、 $X$ は正規分布

$$N(m, \sigma^2) = N(np, npq)$$

に近似することができる(ラプラスの定理)。正規分布の確率密度関数はかなり複雑な形をしているが、どこからこの式が生まれたのか。実は二項分布の極限、すなわち

$${}_n C_r p^r q^{n-r} \longrightarrow \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (n \rightarrow \infty)$$

であり、上の近似は明らかである(もちろん、この計算は範囲外)。

**解答** AB型の人数 $X$ は二項分布  $B(400, 0.1)$  に従う。

$$m=np=40, \sigma^2=npq=36$$

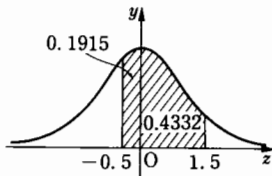
$n$ は十分大きいから、 $Z = \frac{X-40}{6}$  は

$N(0, 1)$  に従う。

$$P(37 \leq X \leq 49)$$

$$= P(-0.5 \leq Z \leq 1.5)$$

$$= 0.1915 + 0.4332 = 0.6247$$



.....(答)

## II. 統計的推測

基礎問

72.

▶ 度数分布表 ◀

あるクラス40人の生徒の体重を測って、次のような度数分布表ができた。これについて体重の平均値  $\bar{x}$  と分散  $s^2$  (標準偏差の平方) を、小数第1位まで求めよ。

体 重	40	41	42	43	44	45	46	47	48	49
人 数	1	0	2	5	7	10	8	4	1	2

(東北学院大)

精講

度数分布表から平均値  $\bar{x}$ , 分散  $s^2$  を計算したい。階級値, 度数がそれぞれ  $x_i, f_i (1 \leq i \leq k)$  で

階級値	$x_1$	$x_2$	……	$x_k$
度 数	$f_1$	$f_2$	……	$f_k$

$n = f_1 + f_2 + \dots + f_k$  のとき,

$$\text{平均値 } \bar{x} = \frac{1}{n} \sum_{i=1}^k x_i f_i, \quad \text{分散 } s^2 = \frac{1}{n} \sum_{i=1}^k x_i^2 f_i - \bar{x}^2$$

であるが、本問のように、 $x_i f_i, x_i^2 f_i$  の値が大きくなり、また、多くなると計算が大変である。簡便計算法として仮平均  $x_0$  を使う方法がある。

$u_i = x_i - x_0$  と置き、 $u_i$  の平均値を  $\bar{u}$  とすると

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^k (x_0 + u_i) f_i = x_0 + \frac{1}{n} \sum_{i=1}^k u_i f_i \\ &= x_0 + \bar{u} \\ s^2 &= \frac{1}{n} \sum_{i=1}^k (x_0 + u_i)^2 f_i - \bar{x}^2 \\ &= \left( x_0^2 + 2x_0 \frac{1}{n} \sum_{i=1}^k u_i f_i + \frac{1}{n} \sum_{i=1}^k u_i^2 f_i \right) - \bar{x}^2 \\ &= \left( x_0^2 + 2x_0 \bar{u} + \frac{1}{n} \sum_{i=1}^k u_i^2 f_i \right) - (x_0 + \bar{u})^2 \\ &= \frac{1}{n} \sum_{i=1}^k u_i^2 f_i - \bar{u}^2 \end{aligned}$$

**解答** 体重の仮平均を45にとり、 $u_i = x_i - 45$  とすると

階数 $i$	0	1	2	3	4	5	6	7	8	9	計
体重 $x_i$	40	41	42	43	44	45	46	47	48	49	
人数 $f_i$	1	0	2	5	7	10	8	4	1	2	40
$u_i$	-5	-4	-3	-2	-1	0	1	2	3	4	
$u_i f_i$	-5	0	-6	-10	-7	0	8	8	3	8	-1
$u_i^2 f_i$	25	0	18	20	7	0	8	16	9	32	135

これより

$$\text{平均値 } \bar{x} = 45 + \frac{1}{40} \sum_{i=0}^9 u_i f_i = 45 + \frac{-1}{40} = 44.975 \quad \cdots \cdots (\text{答})$$

$$\text{分散 } s^2 = \frac{1}{40} \sum_{i=0}^9 u_i^2 f_i - \bar{x}^2 = \frac{135}{40} - \left(\frac{-1}{40}\right)^2 \approx 3.374 \quad \cdots \cdots (\text{答})$$

## 研究

資料が与えられたときの代表値としては平均値  $\bar{x}$  のほかに、中央値  $M_e$  (メデアン, Median), 最頻値  $M_0$  (モード, Mode) などとも使われる。

**中央値  $M_e$**  ……変量を大きさの順に並べたときの中央の値

すなわち、 $x_1 \leq x_2 \leq \cdots \leq x_n$  のとき

$$n \text{ が奇数なら, } M_e = x_{\frac{n+1}{2}}, \quad n \text{ が偶数なら, } M_e = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

**最頻値  $M_0$**  ……度数の最も多い変量の値

## 演習

**[60]** 3つのかごA, B, Cに卵がそれぞれ2個, 4個, 4個入っている。かごAの卵の重さの平均値は50g, 標準偏差は1gで, かごCの卵の重さの平均値は60g, 標準偏差は2gである。かごBの卵はどれも, かごAの卵よりも重く, かごCの卵よりも軽い。そして, その重さは55, 56, 54, 55(単位はg)である。また, 3つのかごの重さをそれぞれ  $a, b, c$  (単位はg) とするとき  $a+b+c=270$ ,  $a^2+b^2+c^2=24306$  である。このとき, 次の値を求めよ。

(1) 10個の卵の重さの平均値, 中央値(メジアン)。

(2) 3つのかごの重さの標準偏差

(3) 10個の卵の重さの標準偏差

(千葉大)

## 基礎問

## 73.

## ▶ 標本平均 ◀

袋の中に1, 2, 3, 4, 5の数字が記入されたカードが1枚ずつ入っている。この袋から復元抽出によって大きさ2の標本をとりだすとき、標本平均  $\bar{X}$  の平均値と分散を求めよ。(弘前大)

## 精講

1° ある母集団から抽出された  $n$  個の変量  $X_1, X_2, \dots, X_n$  に対し  $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$  を標本平均という。

したがって、本問は、とり出される数字を  $X_1, X_2$  とすると、標本の選び方は  $5^2=25$  通りある。標本平均  $\bar{X} = \frac{X_1 + X_2}{2}$  の確率分布は

$\bar{X}$	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	計
$P(\bar{X})$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{5}{25}$	$\frac{4}{25}$	$\frac{3}{25}$	$\frac{2}{25}$	$\frac{1}{25}$	1

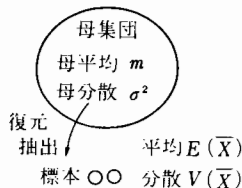
これより、 $E(\bar{X})=3$ 、 $V(\bar{X})=1$  を得る。

2° 途中の計算は省略したが、前問のように仮平均をたてて度数分布表を完成していけばよい。しかし、標本が多くなるとこの計算は大変である。母集団の平均  $m$ 、分散  $\sigma^2$  が分かったとき、これを  $\bar{X}$  の方に反映できないものだろうか。次の関係が知られている。

$$E(\bar{X}) = m$$

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

これは覚えておこう。



解答 母平均  $m = \frac{1+2+3+4+5}{5} = 3$

$$\text{母分散 } \sigma^2 = \frac{1^2+2^2+3^2+4^2+5^2}{5} - 3^2 = 11 - 9 = 2$$

よって、平均値  $E(\bar{X}) = m = 3$ 、分散  $V(\bar{X}) = \frac{\sigma^2}{n} = \frac{2}{2} = 1$

……(答)

研究

1°  $E(\bar{X})=m, V(\bar{X})=\frac{\sigma^2}{n}$  の証明

$$E(\bar{X})=E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)=\frac{m+m+\cdots+m}{n}=m$$

各  $X_i$  は独立ゆえ

$$V(\bar{X})=V\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)=\frac{\sigma^2+\sigma^2+\cdots+\sigma^2}{n^2}=\frac{\sigma^2}{n}$$

2° 母平均  $m$ , 母分散  $\sigma^2$  の母集団(正規分布ではなくてもよい)から抽出された大きさ  $n$  の標本平均  $\bar{X}$  は,  $n$  が十分大きければ

$$\text{正規分布 } N\left(m, \frac{\sigma^2}{n}\right)$$

に近似してよい(これを中心極限定理という)。

演習

[61] 母標準偏差が  $\frac{1}{2}$  の母集団から大きさ  $n$  の無作為標本を復元抽出するとき, その標本平均  $\bar{X}$  の標準偏差が  $\frac{1}{40}$  以下となるような  $n$  の最小値を求めよ。(弘前大)

[62] 体長が平均 50 cm, 標準偏差 3 cm の正規分布に従う生物集団があるとする。正規分布の密度関数  $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$  について,

$F(z)=\int_0^z f(x)dx$  の値が表のように与えられているとき

- (1) 体長 47 cm から 56 cm までのものは全体の何%であるか。
- (2) 4 つの個体を無作為に取り出したとき, 標本平均が 53 cm 以上となる確率を求めよ。ここで, 変数  $x$  が正規分布をしていれば,  $n$  の値にかかわらず, 標本平均  $\bar{x}$  の分布はまた正規分布になることを利用せよ。(福岡大)

$z$	$F(z)$
1	0.3413
2	0.4772
3	0.4987

## 基礎問

## 74.

## ▶ 母平均の推定 ◀

ある動物用の新しい飼料を試作し、任意抽出された100匹にこの新しい飼料を毎日与えて1週間後に体重の変化を調べた。増加量の平均は2.57 kg、標準偏差0.35 kgであった。この増加量について、

- (1) 母平均を信頼度95%で推定せよ。(信頼区間を求めよ。)
- (2) 標本平均と母平均の違いを95%の確率で0.05 kg以下にするには標本数をいくらにすればよいか。(山梨医大)

**精講** 1° 標本平均  $\bar{X}$  から母平均  $m$  を推定したい。“信頼度95%で推定せよ”ということは、95%の確率で  $m$  はこの範囲にあるといいきることである。

母分散が  $\sigma^2$  であるとき  $\bar{X}$  は正規分布

$N\left(m, \frac{\sigma^2}{n}\right)$  に従うから、

$$m - 1.96 \frac{\sigma}{\sqrt{n}} < \bar{X} < m + 1.96 \frac{\sigma}{\sqrt{n}}$$

であることが、95%の確率でいえる。

これを  $m$  について整理すると

$$\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} < m < \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}$$

となる。この区間を母平均  $m$  の信頼度95%の信頼区間という。

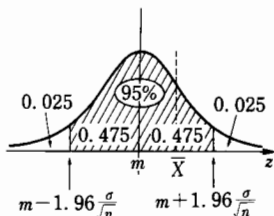
$$\text{信頼度99\%のときは、} \bar{X} - 2.58 \frac{\sigma}{\sqrt{n}} < m < \bar{X} + 2.58 \frac{\sigma}{\sqrt{n}}$$

となる。ここで出てくる“1.96”、“2.58”という数字は覚えておくこと。

また、標本から母集団を評価するわけだから、母平均ばかりか母分散  $\sigma^2$  も未知なことが多い。このときは、 $n$  を十分大きくとると標本分散で代用することができる。

2° 標本の大きさ  $n$  を決めるには、信頼区間の幅に着目すればよい。たとえば、

“ $\bar{X}$  と  $m$  との誤差を、信頼度95%で  $\pm a$  以内(信頼区間  $2a$ )”にするためには、信頼度95%の信頼区間



$\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} < m < \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}$  の幅は  $2 \times 1.96 \frac{\sigma}{\sqrt{n}}$  であるから、これを  $2\alpha$  でおさえて

$$2 \times 1.96 \frac{\sigma}{\sqrt{n}} \leq 2\alpha \quad \therefore 1.96 \frac{\sigma}{\alpha} \leq \sqrt{n}$$

すなわち、 $n \geq \left(1.96 \frac{\sigma}{\alpha}\right)^2$  ということになる。

**解答** 母分散  $\sigma^2$  は標本分散  $0.35^2$  で代用する。

(1) 母平均  $m$  の信頼度 95% の信頼区間は

$$\begin{aligned} & 2.57 \pm 1.96 \frac{0.35}{\sqrt{100}} \\ & = \begin{cases} 2.6386 \\ 2.5014 \end{cases} \end{aligned}$$

であるから

$$2.50 < m < 2.64 \quad \dots\dots(\text{答})$$

(2)  $\bar{X}$  と  $m$  の違い  $|\bar{X} - m|$  の 95% 信頼区間は

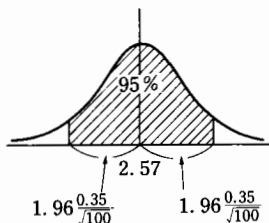
$$|\bar{X} - m| < 1.96 \frac{0.35}{\sqrt{n}}$$

であるから、これを 0.05 以下にするためには

$$1.96 \frac{0.35}{\sqrt{n}} \leq 0.05$$

$$\therefore n \geq \left(\frac{1.96 \times 0.35}{0.05}\right)^2 = 13.72^2 = 188.2384$$

したがって、標本数は 189 本以上にすればよい。  $\dots\dots(\text{答})$



► 演習 ◀

[63] 過去の資料から、18歳の男子の身長標準偏差は5.8 cmであることが知られている。いま、18歳の男子の身長の平均値を信頼度95%で区間推定するためには、何人かを抽出して調査したい。信頼区間の長さを2 cm 以下にするためには、何人以上調査する必要があるか。

(和歌山県医大)



## 基礎問

## 75.

## ▶母比率の推定◀

新しい薬を作っているある工場で、大量の製品全体の中から任意に1000個を抽出して検査を行ったところ、20個の不良品があった。この製品全体について不良率を、二項分布の計算には正規分布を用い、95%の信頼度で推定せよ。(山梨医大)

## 精講

1° 母集団のなかで性質Aをもつものの比率(割合) $p$ を母比率といい、大きさが $n$ の標本のなかで性質Aをもつものが $x$ 個あるときの比率  $\bar{p} = \frac{x}{n}$  を標本比率という。本問

では、標本比率  $\bar{p}$  から母比率  $p$  を推定したいわけである。

考え方の基本は母平均の推定のとおりと同じである。標本が性質Aをもつかたないかの判定であるから、性質Aをもつ標本の個数

$x$  は二項分布  $B(n, p)$  に従う。

したがって、 $n$  が十分大きければ

$x$  は正規分布  $N(np, npq)$  に近似される。

ここで、 $q=1-p$  である。これより、母集団において性質Aをもつものの個数 $X$ の信頼度95%の信頼区間は

$$x - 1.96\sqrt{npq} < X < x + 1.96\sqrt{npq}$$

であり、この式を $n$ で割ると、母比率 $p$ の信頼度95%の信頼区間は

$$\bar{p} - 1.96\sqrt{\frac{p(1-p)}{n}} < p < \bar{p} + 1.96\sqrt{\frac{p(1-p)}{n}}$$

となる。 $n$  が十分大きいときは、母分散を標本分散で代用できるので、

$$\bar{p} - 1.96\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} < p < \bar{p} + 1.96\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

としてよい。

信頼度が99%のときは1.96を2.58に代えればよい。

2° 標本の大きさ $n$ の決め方も、母平均の場合と同じである。

$\bar{p}$  と  $p$  の誤差を、信頼度95%で $\pm\alpha$ (信頼区間 $2\alpha$ )にするためには、

$$\frac{\sigma}{\sqrt{n}} \text{ を } \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \text{ に代えて}$$

$$2 \times 1.96 \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \leq 2\alpha$$

となる  $n$  を解けばよい。

**解答** 標本比率  $\bar{p} = \frac{20}{1000} = \frac{1}{50}$

$$\begin{aligned} \bar{p} \pm 1.96 \sqrt{\frac{\bar{p}(1-\bar{p})}{1000}} &= \frac{1}{50} \pm 1.96 \times \frac{7\sqrt{10}}{5000} \\ &= 0.02 \pm 0.00867 \cdots \\ &= \begin{cases} 0.0287 \\ 0.0113 \end{cases} \end{aligned}$$

したがって、不良率は 1.13% から 2.87% である。……(答)

▶ 演習 ◀

[64] 2つの地域A, B産の大豆をある比率で混ぜ合わせたものがある。A産の大豆の全体に対する比率は、0.2位であると考えられている。いまこのような大豆の集まりから無作為に何粒かを選び出すことにする。A産の大豆の比率を信頼度95%で区間推定するとき、信頼区間の幅を0.02以下にするには、何粒位を選び出せばよいか。(旭川医大)

基礎問 76. ▶ 母平均の検定 ◀

ある母集団の分布が、標準偏差1.2の正規分布であることがわかっているとき、この母集団から標本1個を抽出して、値2.5を得た。

- (1) 母平均が4.5であるという仮説を、有意水準5%で棄却できるか。
- (2) 母平均が5.0であるという仮説を、有意水準5%で棄却できるか。1%ではどちらか。
- (3) 母平均が6.0であるという仮説を、有意水準5%で棄却できるか。1%ではどうか。

以上に、理由を付して答えよ。

(図書館情報大)

### 精講

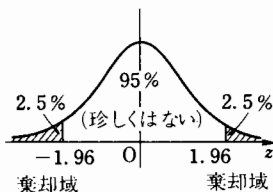
1° 母集団に関してある種の仮説をたてて、標本の結果から、この仮説が正しいかどうかを調べることを検定という。検定の手順をまとめておこう。

- ① 母平均が  $m$  であるという仮説をたてる。  
 ② 標本平均  $\bar{X}$  は正規分布  $N\left(m, \frac{\sigma^2}{n}\right)$  に従うから、これを

$$Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \text{ と標準化する。}$$

- ③ 有意水準 5% のとき、 $|Z| \geq 1.96$  なら仮説を捨てる。  
 有意水準 1% のとき、 $|Z| \geq 2.58$  なら仮説を捨てる。

2°  $|Z| \leq 1.96$  となる確率は右図のよ  
 うに 95% もあるのだから、そう間違っ  
 た仮説とはいえない。しかし、 $|Z| \geq 1.96$   
 になったとすると、その確率は 5%、こ  
 れはそう起こることではない。したがっ  
 て、仮定は正しいとは判断できないとい  
 わけである。



逆に言うと、 $|Z| \geq 1.96$  のとき、正しい仮説であるにもかかわらず捨てる可能性から 5% 残っているということにもなる。この意味から“危険率 5%”という言い方もある。“有意水準 5%”と同じ意味である。

**解答** 標本値  $X$  は  $N(m, 1.2^2)$  に従うから、 $Z = \frac{X - m}{1.2}$  は

$N(0, 1)$  に従う。 $X = 2.5$  ゆえ

$$|Z| = \left| \frac{2.5 - m}{1.2} \right| \leq 1.96 \text{ となるのは、} 0.148 \leq m \leq 4.852$$

$$|Z| = \left| \frac{2.5 - m}{1.2} \right| \leq 2.58 \text{ となるのは } -0.596 \leq m \leq 5.596$$

- (1) 有意水準 5% では棄却できない。……(答)  
 (2) 有意水準 5% では棄却はできるが、1% では棄却できない。(答)  
 (3) 有意水準 5% でも 1% でも棄却できる。……(答)

## 基礎問

## 77.

## 母比率の検定

3種類の品物A, B, Cがある。Aを3個, Bを2個, Cを1個任意に選んで1つにまとめて1個の商品とする。

- (1) 「Aには, A全体の $\frac{1}{16}$ の不良品が含まれ, Bには, B全体の $\frac{1}{9}$ , Cには, C全体の $\frac{1}{25}$ の不良品が含まれている。」という仮説のもとで, 全商品の中から, 無作為に1個の商品を取り出したとき, それが完全な商品である確率を求めよ。ここで, 完全な商品とは不良品が含まれていない商品のことである。
- (2) 商品960個を無作為に抽出したところ, 完全な商品は640個であった。このことから, (1)の仮説は正しいと判断してよいかどうか, 有意水準(危険率)5%で検定(両側検定)せよ。(東北大)

## 精講

1° 原文には, 正規分布表がついていたが, 有意水準5%であるから, 棄却域 $|Z| \geq 1.96$ は覚えたものとして省略した。

2° 母比率の推定のとくと同じく,  $\frac{\sigma}{\sqrt{n}}$ を $\sqrt{\frac{p(1-p)}{n}}$ に代えたと

$$\frac{\bar{X} - np}{\frac{\sigma}{\sqrt{n}}} \rightarrow \frac{\frac{x}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{x - np}{\sqrt{np(1-p)}} \text{ であり,}$$

有意水準5%の棄却域は,  $\frac{|x - np|}{\sqrt{np(1-p)}} \geq 1.96$ である。

解答 (1)  $p = \left(\frac{15}{16}\right)^3 \left(\frac{8}{9}\right)^2 \left(\frac{24}{25}\right) = \frac{5}{8}$  .....(答)

(2) (1)の仮説は「母比率 $=p$ 」ということである。

$$\frac{|x - np|}{\sqrt{np(1-p)}} = \frac{\left|640 - 960 \cdot \frac{5}{8}\right|}{\sqrt{960 \cdot \frac{5}{8} \cdot \frac{3}{8}}} = \frac{8}{3} = 2.66\cdots > 1.96$$

この仮説は棄却され, 正しいとは判断できない。 .....(答)

## 基礎問

## 78.

## ▶ 片側検定 ◀

これまでの生産工程では平均9%の不良率であったが、新しく改良された機械を導入して生産工程を改善した。この新工程による製品から900個を任意抽出して調べたところ、不良品は63個であった。この新工程により不良率は従来のもより下ったといえるか。有意水準5%で片側検定せよ。

## 精講

1° 仮説「 $m=m_1$ 」を検定するとき、仮説からのずれが  $m > m_1$ ,  $m < m_1$  の両方が考えられるときは今までのように両側検定をおこなうが、なんらかの理由で、 $m < m_1$  (あるいは  $m > m_1$ ) となり得ないときは片側検定をおこなう。本問では、改良したのであるから  $m > m_1$  となり得ない。

2° 片側検定の棄却域は、今までの棄却域とは当然違ってくる。

右図のような右片側検定では

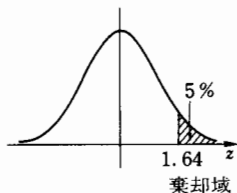
有意水準5%のとき、

$Z \geq 1.64$  なら仮説を捨てる。

有意水準1%のとき、

$Z \geq 2.33$  なら仮説を捨てる。

本問のように、左片側検定のときは、不等号の向きを変えて、 $Z \leq -1.64$ ,  $Z \leq -2.33$  となれば仮説を捨てる。

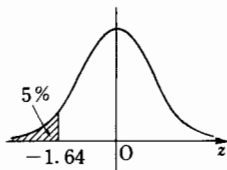


解答 新工程における製品の不良率も9%という仮説をたてる。

$$n=900, x=63, p=0.09$$

であるから

$$\begin{aligned} & \frac{x - np}{\sqrt{np(1-p)}} \\ &= \frac{63 - 900 \cdot 0.09}{\sqrt{900 \cdot 0.09 \cdot 0.91}} \\ &= -2.096 \cdots < -1.64 \end{aligned}$$



この仮説は棄却されたので、不良率は下がったといえる。

……(答)