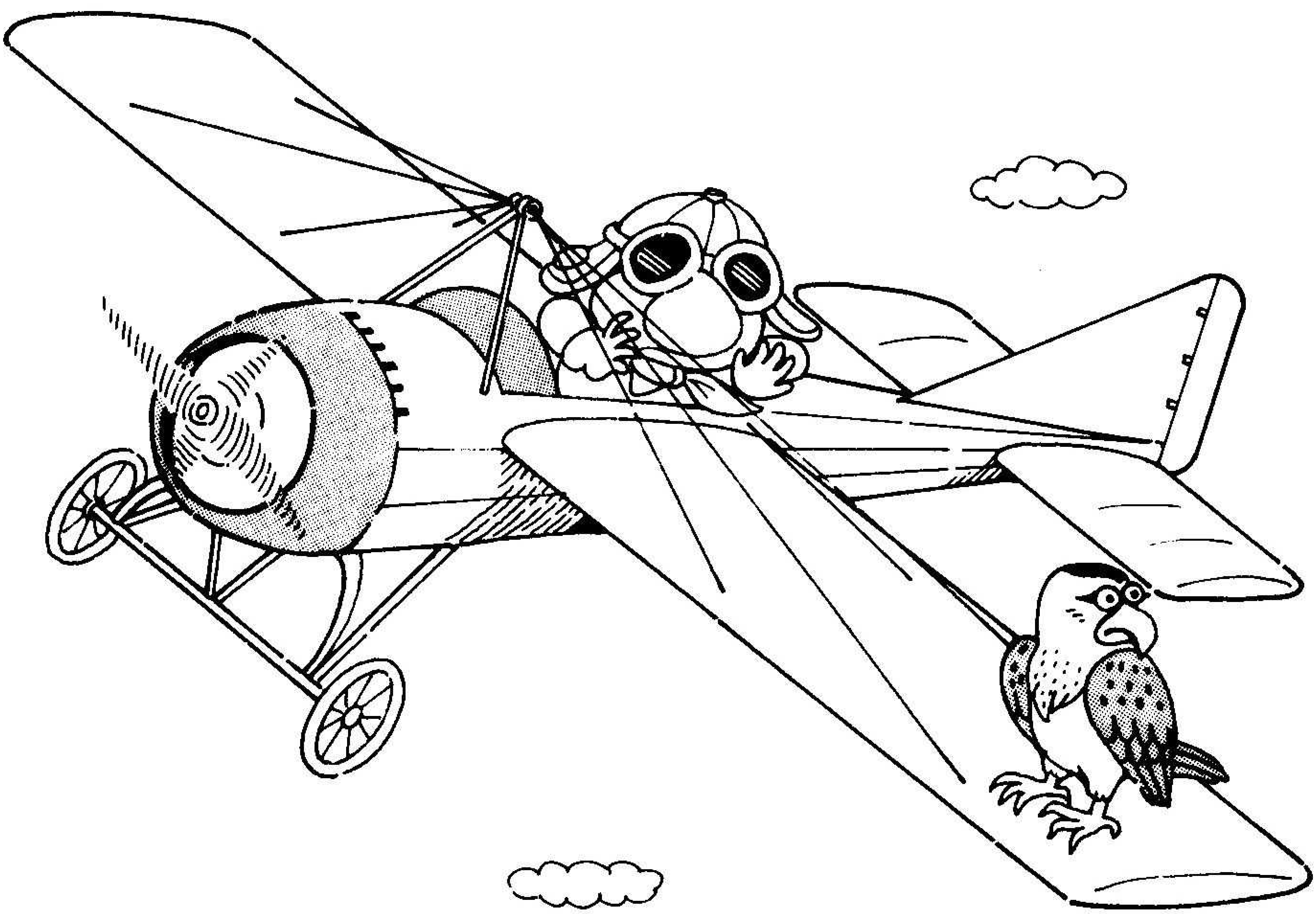
第 4 章

統計的推測



○母集団と標本とは何か

1回目　年　月　日
2回目　年　月　日
3回目　年　月　日

◆母集団はボシュウダンと読む。ハハシュウダンと読んではいけない。いわんやハハナル集団などと読んではいけせんよ。

◆ まず具体的な例から出発しましょう。

例 1. ある工場で製品の不良率を調べたいとしましょう。もちろん，全製品について調べればよいのですが，時間的に不可能なこともありましょう。あるいは，カンヅメのように，開けて調べたのではもはや売物にならない，という場合もありましょう。

このような場合にはその一部について調査し，全体のもようを推定する必要が起こってきます。このような場合に，製品全体を **母集団** といい，調べられたものを **標本** といいます。

例 2. ある学校で，任意に10人を選び，ある事柄について意見を聞いたところ8人が賛成したとしましょう。しかし，このことからこの学校の生徒の80%が賛成だと結論してはいけません。べつの10人を選べば，べつな結果が得られるにちがいありません。このとき，その学校の全員のもつ意見が母集団の意見で，選び出された10人は標本なのです。そして，この 10(10人) を **標本の大きさ** といいます。

*　*　*

◆ さて，統計学で問題となるのは，標本を調べて，母集団の統計的性質を求めることなのです。そして，それには大きく分けて推定と検定の問題とすることができます。

さて，まず第1は，これです。

母集団の平均（これ母平均ということがあります）をm，母集団の標準偏差（これを母標準偏差ということがあります）をσとし，標本の平均を$\overline{x}$，標本の標準偏差をs，標本の大きさをnとするとき，$\overline{x}$の分布について次の関係があります。すなわち，

$\overline{x}$の平均は　m

$\overline{x}$の標準偏差は　$\dfrac{\sigma}{\sqrt{n}} \fallingdotseq \dfrac{s}{\sqrt{n}}$

で与えられます。では，まず，これを：――

練習 1. ある試験の成績が平均70点，標準偏差10点の正規分布にしたがうものとみなされるとき，これから大きさ5の標本をとれば，その平均$\overline{x}$の平均と標準偏差を求めよ。

ヒント　例えば，全国の高校生について統一試験をやった場合平均点は70で，標準偏差が10の正規分布をしていたとします。

その場合に，かってに5人選んでみたら

30, 50, 50, 80, 90（点）

だったとしましょう。この平均は60点で，標準偏差sは21.91です。しかし，べつに勝手に5人選んだのでは，こうはならないでしょう。しかし，かってに5人を選んで，数多くそれぞれの平均$\overline{x}$を求めてみると，その平均はもとの平均 $m=70$ になるし，$\overline{x}$の標準偏差は $\sigma=10$ を $\sqrt{n}=\sqrt{5}$ で割ったもの，つまり

$$\frac{10}{\sqrt{5}}=\frac{10\sqrt{5}}{5}=4.47$$

になるというのです。

答　70, 4.47

練習 2. ある地方で20歳の男子の身長を測ったら平均 165.2cm で，標準偏差 5.7cm であった。この中から任意に10人選んで平均をとると，その平均と標準偏差はどうなるか。

（解）　求める平均は165.2cm，標準偏差は $\frac{5.7}{\sqrt{10}}=1.80$　である。

■練習 3. ある地方で，ある試験の平均が40点で，標準偏差は15点であった。いま大きさ n の標本をとって，その平均を $\bar{x}$ とするとき，$\bar{x}$ の標準偏差を5点を越えないようにするには標本の大きさをなにほどにすればよいか。

（解）　$\bar{x}$ の標準偏差は $\frac{15}{\sqrt{n}}$ であるから

$$\frac{15}{\sqrt{n}}\leqq 5 \qquad \therefore \quad \sqrt{n}\geqq 3$$

$$\therefore \quad n\geqq 9$$

ゆえに，標本の大きさが9以上であればよい。

答　9以上

*　*　*

◆ 一般には，標本の統計的性質から，母集団の統計的性質を知りたいところです。これについて，次の関係があります。

標本平均を $\bar{x}$ とするとき，$\bar{x}$ の分布が平均 m，標準偏差 $\frac{s}{\sqrt{n}}$ の正規分布とみなせるならば，母平均 m は

95%の信頼度で

$$\bar{x}-1.96\frac{s}{\sqrt{n}}<m<\bar{x}+1.96\frac{s}{\sqrt{n}}$$

99%の信頼度で

$$\bar{x}-2.58\frac{s}{\sqrt{n}}<m<\bar{x}+2.58\frac{s}{\sqrt{n}}$$

で与えられます。1.96の代わりに 2；2.58の代わりに3にしてある本もあります。そのときは

95%の信頼度で

$$\bar{x}-\frac{2s}{\sqrt{n}}<m<\bar{x}+\frac{2s}{\sqrt{n}}$$

99%の信頼度で

$$\bar{x}+\frac{3s}{\sqrt{n}}<m<\bar{x}+\frac{3s}{\sqrt{n}}$$

となります。少し正確さは落ちるというものの覚えやすい，という利点があります。ここでは 1.96 と 2.58 を使うことにしましょう。なお，信頼度については（☞ p.184）を参照してください。

■練習 4. ある工場の電球の寿命の標準偏差は 100 時間であることがわかっている。

100個の電球の平均寿命 1800 時間のとき次の問に答えよ。

(1)　母集団の平均を95%の信頼度で推定せよ。

(2)　母集団の平均を99%の信頼度で推定しようとするとき，信頼区間の幅を95%の場合と同程度におさえるためには，何個以上の電球の平均を求めればよいか。

（ヒント）　(1)　母集団の平均を m，標準偏差を σ，標本の大きさを n，標本の平均を $\bar{x}$ としますと，$\bar{x}$ の値から m の値を推定する95%の信頼区間は

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

で与えられます。そこで：——

$$\sigma=100,\ n=100,\ \bar{x}=1800$$

を代入しますと

$$1800-1.96\times\frac{100}{\sqrt{100}}<m<1800+1.96\times\frac{100}{\sqrt{100}}$$

$$\therefore \quad 1780<m<1820$$

となります。

次に，(2)は

$$\frac{2.58\sigma}{\sqrt{n}}\leqq\frac{2\sigma}{\sqrt{100}}$$

より

$$\frac{2.58\times 100}{\sqrt{n}}\leqq\frac{2\times 100}{\sqrt{100}}$$

$$\therefore \quad 12.9\leqq\sqrt{n}$$

$$\therefore \quad 166.4\leqq n$$

ゆえに 167 個以上の平均をとればいいことがわかります。

*　*　*

◎抜きとり検査の扱い方

1回目 年 月 日
2回目 年 月 日
3回目 年 月 日

◆抜きとり検査は統計の華(はな)である。割合スッキリしている。入試によく出るのも当然といっていいでしょう。

◆ さっそくながら具体的な練習にとりかかるとしよう。そして，それをやりながら抜きとり検査の意味がわかればいいでしょう。では，まずこれです：——

練習 1. きわめて多数の製品の一山があるときその一山の全体としての合格，不合格を次のようにして判定するものとする。すなわち，全体の中から5個を抜きとって検査し，その5個中の不良品が1個以下ならば合格とし，2個以上ならば不合格とすることに定める。この規則にしたがって，不良率(不良品の個数と全体の個数との比)が p である製品の一山が，検査に合格する確率を $f(p)$ とするとき，次の各問に答えよ。

(1) $f(p)$ を p の式で表せ。

(2) $f(p)$ は p の単調減少関数であることを証明せよ。

(3) $f(0)$，$f(0.3)$，$f(1)$，$f'(0)$，$f'(1)$ を求めよ。

(4) 以上の結果を利用して $f(p)$ のグラフの概形をえがけ。 (神戸大)

ヒント (1) べつにめんどうはないでしょう。製品の中から任意に1個を抜きとったとき，それが不良品である確率は p で，良品である確率は $1-p$ ですから，5個全部良品である確率は $(1-p)^5$，1個だけが不良品である確率は ${}_5C_4(1-p)^4p$ ですから，合格の確率 $f(p)$ は

$$f(p)=(1-p)^5+{}_5C_4(1-p)^4p$$
$$=(1-p)^4(1+4p)$$

となります。

(2) $0\leqq p\leqq 1$ ですから，

$$f'(p)=-4(1-p)^3(1+4p)+(1-p)^4\cdot 4$$
$$=-20p(1-p)^3\leqq 0$$

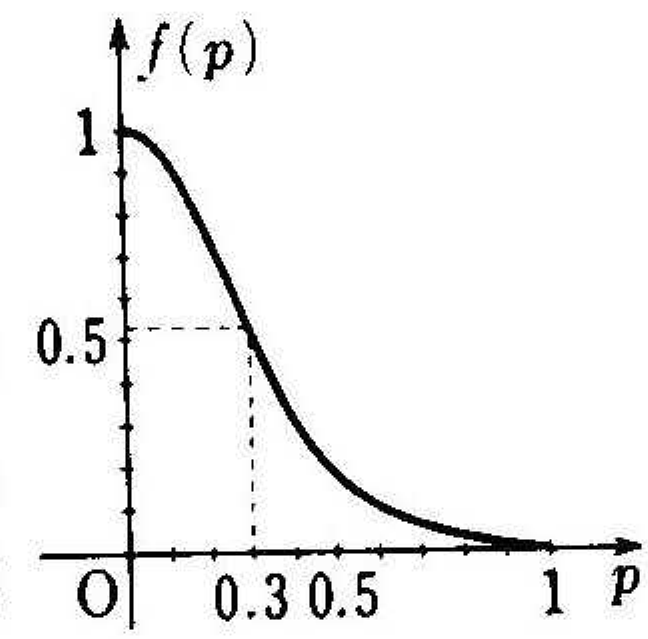
ゆえに，$f(p)$ は単調減少関数です。

(3) $f(0)=1$

$f(0.3)=(0.7)^4(1+1.2)\fallingdotseq 0.53$

$f(1)=0$，$f'(0)=0$，$f'(1)=0$

(4) 以上のことからグラフは右のようになります。



練習 2. ある製品10個の中に3個の不良品がある。この10個の中から4個をとり出すとき，この中に少なくとも1個の不良品が含まれる確率を求めよ。 (岡山大)

解 10個の中から4個とり出す仕方は

$${}_{10}C_4=\frac{10\cdot 9\cdot 8\cdot 7}{1\cdot 2\cdot 3\cdot 4}=210 \text{ (通り)}$$

不良品でない7個の中から4個とり出す仕方は

$${}_7C_4=\frac{7\cdot 6\cdot 5\cdot 4}{1\cdot 2\cdot 3\cdot 4}=35 \text{ (通り)}$$

ゆえに，とり出した4個がすべて不良品でない確率は $\frac{35}{210}=\frac{1}{6}$ である。

よって，求める確率は

$$1-\frac{1}{6}=\frac{5}{6}$$

答 $\frac{5}{6}$

である。

注 こうしてみると，べつに特に変わったことはありませんね。要するに確率の計算問題にすぎません。

* * *

◆ では，やや総合的な問題にいきましょう。

練習 3. ある製品の検査員は製品1個につき，良品を誤って不良品と判定する確率は0.1で，不良品を誤って良品と判定する確率は0.2であるという。この検査員に4個の製品を検査させるとき，実際は3個が良品で，1個が不良品であるのに，良品，不良品はそれぞれ2個であると判定する確率を求めよ。　（熊本大）

（解） 不良品を不良品と判定して，残り3個の良品のうち2個を良品，他の2個を不良品と判定する確率は

$$0.8\times{}_3C_2\times(0.9)^2(0.1)=0.1944$$

また，不良品を良品と判定し，残り3個の良品のうち1個を良品，他の2個を不良品と判定する確率は

$$0.2\times{}_3C_1\times0.9\times(0.1)^2=0.0054$$

である。したがって，求める確率は

$$0.1944+0.0054=0.1998$$

である。

答　0.1998

練習 4. きわめて多数の製品の一山をロットという。同一個数の製品からなるロットがきわめて多数あるとき，これらのロットに対して次のような抜きとり検査を行う。すなわち，各ロットから大きさ5の標本を抜きとり，その中の不良品個数が0であれば標本はもとに戻してそのロットはそのまま出荷し，不良品個数が1個以上であればそのロットは全数検査にかけロット中の不良品を全部良品にとりかえて出荷する。もとのロットの不良率がすべてpであるとした場合，出荷される全ロット中に含まれる製品の不良率$f(p)$はどうなるだろうか。次の問に答えよ。

(1)　$p=0$, $p=1$ のとき，$f(p)$ の値はそれぞれどうか。

(2)　$f(p)$ をpの式で表せ。

(3)　$f(p)$ のグラフをかき，$f(p)$ を最大とするpの値を示せ。　（慶大）

（ヒント） (1)　$p=0$ のときには不良品がないのですから，出荷されるロットの中にも不良品がないハズ。

$$\therefore\quad f(0)=0$$

です。$p=1$ のときには全部不良品ですから，抜きとり検査の結果全部良品にとりかえられます。だから

$$f(1)=0$$

(2)　各ロットがn個の製品からなるとしますと，その中にnp個の不良品があるハズ。あるロットがそのまま出荷されるのは，大きさ5の標本を抜きとったとき全部良品である場合ですから，その確率は $(1-p)^5$ です。

そして，標本の中に不良品があるときは全部良品ととりかえられるのですから，結局出荷されるロットの中の不良品の期待値は

$$np(1-p)^5+0\times\{1-(1-p)^5\}$$
$$=np(1-p)^5$$

となります。

$$\therefore\quad f(p)=\frac{np(1-p)^5}{n}=p(1-p)^5\ \cdots\boxed{答}$$

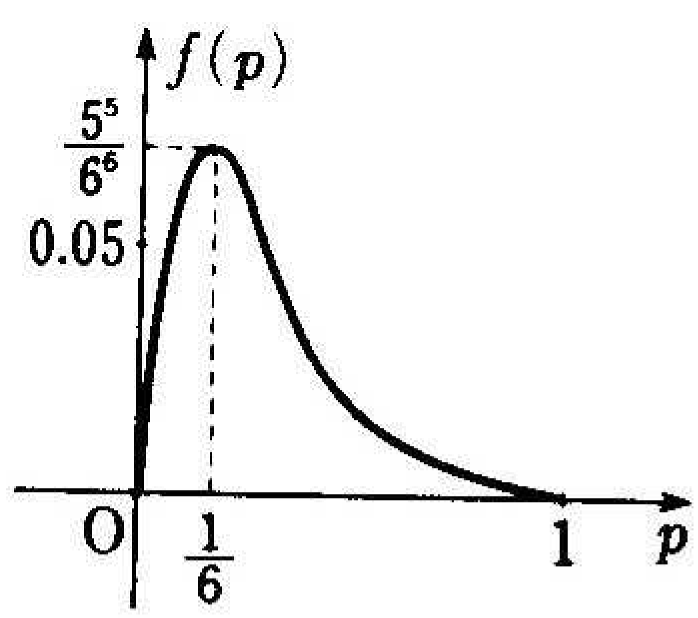
(3)　$f(p)=p(1-p)^5$

のグラフをかくために微分してみますと，

$$f'(p)=(1-p)^4(1-6p)$$

p	0		$\frac{1}{6}$		1
$f'(p)$	1	+	0	−	0
$f(p)$	0	↗	$\frac{5^5}{6^6}$ (約0.07)	↘	0

ゆえに，$f(p)$ が最大になるのは $p=\frac{1}{6}$ であることがわかります。なお，$f(p)$ の最大値は約0.07です。



＊　＊　＊

○母平均の推定の扱い方

1回目 年 月 日
2回目 年 月 日
3回目 年 月 日

◆統計の積極的側面は推定と検定です。では，そもそも推定とは何か。いつか本気でやらないといつも敬遠して通りすぎることになる。

◆ ともあれ，具体的な問題から出発しましょう。

■練習 1. ある事業場において，月給の階級別に，従業員の百分率を計算したら，次の表のようになった。この表から，従業員の平均月給を推定せよ。

月給	百分率
5000 円以上 10000 円未満	24
10000 円以上 15000 円未満	38
15000 円以上 20000 円未満	20
20000 円以上 25000 円未満	12
25000 円以上 30000 円未満	6
計	100

(広島大)

(ヒント) 度数分布表があったとき，各区間の中央の値を x_1, x_2, ……, x_k とし，それぞれの度数を n_1, n_2, ……, n_k としますと，その平均 $\overline{x}$ は

$$\overline{x}=\frac{n_1x_1+n_2x_2+\cdots\cdots+n_kx_k}{n_1+n_2+\cdots\cdots+n_k}$$

で与えられます。そこで，この値を全体の平均として採用すればよいでしょう。本問では $n_1=24$, $n_2=38$, ……, $x_1=7500$円, $x_2=12500$円, ……として計算しますと14400円となります。解答としては表にまとめておくほうがいいでしょう。

(解)

級間	中央の値	百分率	中央値×百分率
5000～10000	7500	24	1800×100
10000～15000	12500	38	4750
15000～20000	17500	20	3500
20000～25000	22500	12	2700
25000～30000	27500	6	1650
			14400

(答) 14,400円

■練習 2. 魚の年令を調べたところ，次表の結果が得られた。

年令	4	5	6	7	8
百分率	15	22	28	22	13

平均年令を推定せよ。

(解) 平均値をもって，平均年令と推定してよい。

$$\frac{4\times15+5\times22+6\times28+7\times22+8\times13}{15+22+28+22+13}$$

$=5.96$

(答) 5.96歳

(注) このように標本の平均値を母集団の平均の推定値とするのは，もっとも単純な場合です。では，もう1つ：――

■練習 3. 袋の中に赤玉と黒玉が合計4個入っている。無作為に2個とり出したら赤玉1個，黒玉1個出た。袋の中には赤玉，黒玉が何個あると推定されるか。

(ヒント) 1個ずつ出たのですから，3つの場合が考えられます。

(1) 赤玉1個，黒玉3個

(2) 赤玉2個，黒玉2個

(3) 赤玉3個，黒玉1個

です。そして，この場合に2個とり出して赤と黒が1個ずつ出る確率はそれぞれ

$$\frac{{}_1C_1\cdot{}_3C_1}{{}_4C_2}=\frac{1}{2},\quad \frac{{}_2C_1\cdot{}_2C_1}{{}_4C_2}=\frac{2}{3},\quad \frac{{}_3C_1\cdot{}_1C_1}{{}_4C_2}=\frac{1}{2}$$

となります。してみると，もっとも確率の大きい2個ずつの場合と考えるのが妥当でしょう。

(答) 赤玉2個，黒玉2個

*　*　*

◆ このように，ある1つの値を推定値として採用するのを点推定（てんすいてい）ということがあります。それに対し，推定値がある区間の中にあるというように幅をもたせて推定するのを区間推定といいます。では，それをやってみましょう。

練習 4. ある県で，20歳の男子100人の身長を測って，平均165.5cm，標準偏差 6 cm を得た。母集団の平均を，信頼度95%で推定せよ。

ヒント 信頼度については（☞ p.184）を参照してください。さて：——

ある母集団から n 個の任意標本をとり出して調べたところ

標本の平均が $\overline{x}$，標本の標準偏差が s

であったとしますと，母集団の平均値 m は信頼度95%で，

$$\overline{x}-1.96\frac{s}{\sqrt{n}}\leqq m\leqq\overline{x}+1.96\frac{s}{\sqrt{n}}$$

とすることができます。ところが

$$\overline{x}=165.5,\quad s=6,\quad n=100$$

ですから

$$1.96\frac{s}{\sqrt{n}}=1.96\times\frac{6}{\sqrt{100}}=1.176\fallingdotseq1.2$$

ゆえに

$$164.3\leqq m\leqq166.7 \qquad \cdots\cdots\boxed{答}$$

練習 5. ある田の稲の穂を100本とってきて，その穂の粒数を調べてみたら，1穂の平均粒数71.67，標準偏差19.43を得た。母集団の平均を，信頼度95%で推定せよ。

解 母集団の平均を m とすると，95%の信頼度で

$$\overline{x}-1.96\frac{s}{\sqrt{n}}\leqq m\leqq\overline{x}+1.96\frac{s}{\sqrt{n}}$$

で与えられる。ここに n は標本の大きさであるから100，s は標本の標準偏差で19.43，$\overline{x}$ は標本の平均71.67である。そして，

$$1.96\frac{s}{\sqrt{n}}=1.96\times\frac{19.43}{\sqrt{100}}\fallingdotseq3.81$$

ゆえに

$$67.86\leqq m\leqq75.48 \qquad \cdots\cdots\boxed{答}$$

*　*　*

◆ 次のような問題もあります。

練習 6. 球の直径を200回測って平均値24.62 cm，標準偏差0.2 cm が得られた。球の直径を信頼度99%で推定せよ。

ヒント 信頼度99%では

$$|\overline{x}-m|\leqq2.58\frac{s}{\sqrt{n}}$$

つまり

$$\overline{x}-2.58\frac{s}{\sqrt{n}}\leqq m\leqq\overline{x}+2.58\frac{s}{\sqrt{n}}$$

です。そして，

$$2.58\times\frac{0.2}{\sqrt{200}}\fallingdotseq0.04$$

$$\therefore\quad 24.58\leqq m\leqq24.66 \qquad \cdots\cdots\boxed{答}$$

練習 7. 平均 m，分散4の正規母集団から大きさ n の任意標本を抽出して，その標本平均を $\overline{x}$ とするとき，次の問に答えよ。

(1) $n=100$，$\overline{x}=10$ のとき，信頼度95%で m の信頼区間を求めよ。

(2) $|\overline{x}-m|\leqq\frac{1}{2}$ となる確率が95%以上であるようにしたい。n をどのようにすればよいか。（神戸大）

ヒント (1) 分散が4ですから，標準偏差は2であり，

$$\overline{x}-1.96\frac{s}{\sqrt{n}}\leqq m\leqq\overline{x}+1.96\frac{s}{\sqrt{n}}$$

より

$$9.6\leqq m\leqq10.4$$

(2) $|\overline{x}-m|\leqq1.96\frac{s}{\sqrt{n}}\leqq\frac{1}{2}$

$s=2$，$n=100$ であるから

$$7.84\leqq\sqrt{n}$$

$$61.4\cdots\leqq n$$

$$\therefore\quad 62\leqq n$$

$\boxed{答}$ 62以上

⑪ 母比率の推定の扱い方

1回目 年 月 日
2回目 年 月 日
3回目 年 月 日

◆母集団の比率が母比率です。ボヒリツと読みます。ハハヒリツなどと読んではいけませんよ。

◆ 二項分布をなす母集団から任意に抽出した大きさnの標本において，ある性質をもつものの比率が$\overline{p}$であったとき，母比率pの信頼区間は次のとおりになります。

信頼度95%以上では

$$\overline{p}-1.96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\leqq p\leqq\overline{p}+1.96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

信頼度99%以上では

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

* * *

◆ では，具体的な問題へ：——

■練習 1. ある原野にはA，B 2種の野ねずみが生息しているという。任意に300匹の野ねずみを捕らえたところ，A種が90匹いた。A種の野ねずみは，この原野全体で何%生息していると考えられるか。信頼度95%で推定せよ。 (旭川医大)

ヒント もちろん二項分布です。そして，

標本の大きさ $n=300$

標本比率 $\overline{p}=\frac{90}{300}=\frac{3}{10}$

ですから

$$\overline{p}-1.96\times\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$
$$=\frac{3}{10}-1.96\times\sqrt{\frac{0.3\times0.7}{300}}$$
$$=0.3-1.96\times\sqrt{0.0007}$$
$$\fallingdotseq0.3-1.96\times0.02646$$
$$\fallingdotseq0.2481$$

同じく

$$\overline{p}+1.96\times\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}=\frac{3}{10}+1.96\sqrt{\frac{0.3\times0.7}{300}}$$
$$\fallingdotseq0.3519$$

$$\therefore\ 0.2481\leqq p\leqq0.3519$$ ……答

では，これを：——

≪ある植物の発芽率は50%であるという。この種子を100粒まいたときの発芽率をpとするとき，

(1) pはどんな分布をするか。

(2) pが40%以上60%以下である確率を求めよ。≫

ヒント (1) 平均0.5，分散 $\frac{0.5\times0.5}{100}=0.0025$ の二項分布をします。また，nは十分大きいので，これは平均0.5，分散$0.0025=0.05^2$の正規分布とみなすことができます。

(2) $\frac{0.4-0.5}{0.05}=-2$，$\frac{0.6-0.5}{0.05}=2$

ですから

$$P(0.4\leqq p\leqq0.6)\fallingdotseq P(-2\leqq Z\leqq2)$$
$$=0.9544$$ ……答

そこで，次の問題にいきましょう。

■練習 2. ある植物の種子を試験的に100粒まいたところ，40粒が発芽したという。この種子を大量にまいたときの発芽率を，99%で推定せよ。

ヒント $\overline{p}=\frac{40}{100}=0.4$

$\overline{q}=1-0.4=0.6$

$n=100$

ですから

$$2.58\sqrt{\frac{\overline{p}\overline{q}}{n}}=2.58\sqrt{\frac{0.4\times0.6}{100}}$$
$$\fallingdotseq0.1264$$
$$\therefore\ 0.2736\leqq p\leqq0.5264$$

となります。

* * *

■**練習 3.** ある都市で2000人の有権者を無作為抽出し，A政党の支持者を調べたところ，420人であった。この都市におけるA政党の支持率を，信頼度95%で推定せよ。

ヒント　$\overline{p}=\dfrac{420}{2000}=0.21$

$n=2000$

ですから

$$0.21-1.96\sqrt{\frac{0.21\times 0.79}{2000}}\leqq p$$
$$\leqq 0.21+1.96\sqrt{\frac{0.21\times 0.79}{2000}}$$

ところが

$$1.96\sqrt{\frac{0.21\times 0.79}{2000}}=\frac{1.96}{100}\sqrt{0.21\times 0.79\times 5}$$

$=0.0196\times\sqrt{0.83}=0.0196\times 0.91$

$\fallingdotseq 0.01\overset{2}{\cancel{7}}$

$$\therefore\quad 0.21-0.02\leqq p\leqq 0.21+0.02$$
$$0.19\leqq p\leqq 0.23$$

■**練習 4.** 新しい薬を作っている工場で，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，95%の信頼度で推定せよ。

（山梨医大）

ヒント　$\overline{p}=\dfrac{20}{1000}=0.02,\ n=1000$

ですから

$$1.96\times\sqrt{\frac{0.02\times 0.98}{1000}}=\frac{1.96}{100}\times\sqrt{1960}\times 10^{-2}$$

$=0.0196\times 44\times 0.01\fallingdotseq 0.0086$

$$\therefore\quad 0.02-0.0086\leqq p\leqq 0.02+0.0086$$
$$0.011\leqq p\leqq 0.029$$

これが求める範囲です。

こうして，いくつかやってみますと，べつにめんどうはないでしょう。要するになれるまで，いくつもやってみる，というにつきるのです。

◆ 本質的には同じことですが，次のような問題をやってみませんか。

■**練習 5.** ある条例に対する市民の賛成者の比率は，だいたい60%前後と予想されている。賛成者の真の比率を2%以下の誤差で推定したい。何人以上について調べる必要があるか。信頼度95%以上とする場合と，99%以上とする場合について検討せよ。

ヒント　95%の場合には

$$\overline{p}-1\ 96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\leqq p\leqq\overline{p}+1.96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

つまり

$$|\overline{p}-p|\leqq 1.96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

ですが，誤差が2%以下にしたいというのですから $\overline{p}=0.60$ を代入して

$$1.96\sqrt{\frac{0.60(1-0.60)}{n}}\leqq 0.02$$
$$\therefore\quad \sqrt{n}\geqq\frac{1.96}{0.02}\sqrt{0.60\times 0.40}$$
$$\therefore\quad n\geqq 98^2\times 0.24$$
$$\therefore\quad n\geqq 2305$$

となりましょう。

次に，99%の場合は，上と同様にして，

$$|\overline{p}-p|\leqq 2.58\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

ですから，同じく $\overline{p}=0.60$ を代入して

$$2.58\sqrt{\frac{0.60(1-0.60)}{n}}\leqq 0.02$$
$$\sqrt{n}\geqq\frac{2.58}{0.02}\sqrt{0.60\times 0.40}$$
$$\therefore\quad n\geqq 129^2\times 0.24$$
$$\therefore\quad n\geqq 3994$$

とすればよいのです。

* * *

◆ 要するに，標本の大きさが n で，そのうちある性質をもったものの個数が r のとき，その比率 $\dfrac{r}{n}=\overline{p}$ は，**n が大きい**とき，**正規分布に従う**ことが大切な点です。

①（推定における）信頼度とは何か

1回目　年　月　日
2回目　年　月　日
3回目　年　月　日

◆信頼度だの，信頼限界だの，信頼区間だの，といろいろウルサイことだ!! しかし，数学を使う段になると，もはやさけては通れない。

◆ 日本人全体について調べることは不可能でも，その中から一部を抜き出して調べることによって，全体を推定することができます。このとき，もとの全体を **母集団**（ぼしゅうだん）といい，抜き出したものを **標本** といいます。ところで，

平均 m，標準偏差 s の母集団からとった n 個の標本の平均を $\overline{x}$ としますと

$$E(\overline{x})=m,\quad s(\overline{x})=\frac{s}{\sqrt{n}}$$

なる関係のあることがわかっています。ここに $E(\overline{x})$ は $\overline{x}$ の平均，$s(\overline{x})$ は $\overline{x}$ の標準偏差を表しています。

すなわち，母集団の平均をm，標準偏差をsとしますと，標本平均 $\overline{x}$ は平均m，標準偏差 $\frac{s}{\sqrt{n}}$ の正規分布をするのです。このため，

95%の確率で

$$\overline{x}-1.96\frac{s}{\sqrt{n}}\leqq m\leqq\overline{x}+1.96\frac{s}{\sqrt{n}}$$

が成り立つことが証明できます。ここに95%の確率で，というのは，この区間にmが含まれる確率が95%だという意味ではなくて，大きさnの標本を何回もとると，この区間にmを含むこともあろうし，含まないこともあろうが，含む確率は95%だということなのです。

*　*　*

◆ ともあれ，具体的な問題をやってみましょう。

■練習1. 大きさ100の標本の平均が56.3で，標準偏差が10.2であった。このとき母集団の平均mを信頼度95%で推定せよ。

（ヒント） 題意により

$$\overline{x}-1.96\frac{s}{\sqrt{n}}\leqq m\leqq\overline{x}+1.96\frac{s}{\sqrt{n}}$$

において

$$\overline{x}=56.3,\quad n=100,\quad s=10.2$$

ですから

$$1.96\times\frac{10.2}{\sqrt{100}}\fallingdotseq 2.0$$

$$\therefore\quad 54.2\leqq m\leqq 58.3$$

（注） この区間 [54.2, 58.3] を信頼区間といいます。

■練習2. ある学年の抽出試験の結果下の表のようになった。各クラスの母平均を信頼度95%で求めよ。

組	人数	平均	標準偏差
A	20	75	5.0
B	30	60	6.0

（解） 信頼度95%で母平均mは

$$\overline{x}-1.96\frac{s}{\sqrt{n}}\leqq m\leqq\overline{x}+1.96\frac{s}{\sqrt{n}}$$

で与えられる。そして $1.96\frac{s}{\sqrt{n}}$ の値は，A, Bクラスでそれぞれ2.2, 2.1であるから

Aクラス：　75±2.2
Bクラス：　60±2.1

となる。

答 $\begin{cases}\text{A}: 75\pm2.2\\ \text{B}: 60\pm2.1\end{cases}$

（注） もちろん答は75±2.2と書いても $72.8\leqq m_A\leqq77.2$ と書いてもかまいません。なお，**母平均**（ぼへいきん）とは母集団の平均という意味ですよ。

*　*　*

◆ 次に，やや理論的な問題をやってみませんか。

■練習 3. 正規母集団からとった大きさ n の標本の標準偏差を s とするとき，母平均を95%の信頼度で推定したら，精度が h であった。

(1) 精度を $\frac{h}{3}$ にするには，標本の個数をいくらにすればよいか。

(2) 99%の信頼度で母平均を推定すれば，精度はいくらになるか。

ヒント (1) まず精度というのは，ある量の推定値が $a-h \leqq x \leqq a+h$ であるとき，h を精度というのです。だから95%なら

$$h=\frac{1.96s}{\sqrt{n}}$$

なのでしょう。してみると，精度を $\frac{h}{3}$ にするには

$$\frac{1.96s}{\sqrt{n'}}=\frac{1}{3}\cdot\frac{1.96s}{\sqrt{n}}$$

$$\therefore \quad \sqrt{n'}=3\sqrt{n} \quad \therefore \quad n'=9n$$

(2) 次に99%の信頼度では，

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

です（☞ p.177）から

$$\bar{x}-\frac{2.58}{1.96}h \leqq m \leqq \bar{x}+\frac{2.58}{1.96}h$$

つまり

$$\bar{x}-1.32h \leqq m \leqq \bar{x}+1.32h$$

となりますから，精度は $1.32h$ となります。

答 $9n$，$1.32h$

注 $\bar{x}-1.96\frac{s}{\sqrt{n}} \leqq m \leqq \bar{x}+1.96\frac{s}{\sqrt{n}}$

や，

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

の代わりに

$$\bar{x}-2\frac{s}{\sqrt{n}} \leqq m \leqq \bar{x}+2\frac{s}{\sqrt{n}}$$

や，

$$\bar{x}-3\frac{s}{\sqrt{n}} \leqq m \leqq \bar{x}+3\frac{s}{\sqrt{n}}$$

を使うことも多いのです。というのも，2と3でオボエヤスイからなんですが。

そのときには上の(2)の解は $\frac{3}{2}h$ となるわけです。もちろん，指定されていない限り，いずれを使ってもかまいません。

■練習 4. 不良率（不良品の数の全数に対する比）が p である製品の山がある。この山から n 個とった標本の不良率を $\bar{p}$ とすると，

$$\bar{p}-2\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \leqq p \leqq \bar{p}+2\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

であることを証明せよ。

ヒント この場合，平均値や標準偏差とは何かつかみにくいのが困った点です。しかし，各製品は不良であるか不良でないかのいずれかですから，不良なら1，不良でないなら0という値をとることにしますと，母集団の平均値は p となります。

そして，大きさ n の標本については不良品の個数を r としますと

$$\bar{p}=\frac{r}{n}$$

で，標準偏差を s としますと

$$s^2=\frac{1}{n}\left\{\left(1-\frac{r}{n}\right)^2\times r+\left(0-\frac{r}{n}\right)^2\times(n-r)\right\}$$

$$=\frac{r}{n}\left(1-\frac{r}{n}\right)=\bar{p}(1-\bar{p})$$

$$\therefore \quad \bar{p}-2\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \leqq p \leqq \bar{p}+2\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Q. E. D.

注 もちろん，次のようでもいいはず。

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

*　　*　　*

◆ 信頼度95%，99%がよく使われますから，1.96とか2.58とかの数字はオボエテおく必要があります(教科書によってはそれぞれ2，3としてあるのもありますが)。それでは信頼度90%ならどうするか？ それは**正規分布表**（☞ p.164）から求めなければなりません。

○（仮設）検定の問題の扱い方

1回目　年　月　日
2回目　年　月　日
3回目　年　月　日

◆統計学が記録の手段から行動の前提となるにおよんで，大きく質的に変化した。検定はまさに，その焦点にある。

◆ まず具体的な問題を考えてみましょう。

ある法律案の可否を世論調査によって調べたとしましょう。10人に質問して6人が賛成だったことから世論は賛否相半しているといってよいだろうか，といった問題が出てきます。では，どうするか？

(i) まず，「世論は賛否相半している」という仮設をたてます。いいかえると，「任意に1人をとり出したときに賛成である確率は$\frac{1}{2}$である」と仮定するのです。

(ii) 次に，そのような仮設のもとで，10人とり出したとき，6人以上が賛成する確率を計算してみます。それは

$$
\begin{aligned}
&{}_{10}C_6\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^4+{}_{10}C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3\\
&\quad+{}_{10}C_8\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2+{}_{10}C_9\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)\\
&\quad+{}_{10}C_{10}\left(\frac{1}{2}\right)^{10}\\
&=({}_{10}C_6+{}_{10}C_7+{}_{10}C_8+{}_{10}C_9+{}_{10}C_{10})\left(\frac{1}{2}\right)^{10}
\end{aligned}
$$

で与えられます。これを計算してみますと0.37……となります。

つまり，10人にアンケートをとれば6人以上賛成という結果が出る確率は約0.37というのですから当然起こりうることです。つまり，この結果から，賛否相半しているという仮設が正しくないとはいいがたいでしょう。

しかし，6人以上賛成という結果が出る確率が0.2になったらどうか，0.1になったらどうか，0.05になったらどうか，となると，返答に窮するにちがいありません。そこで，次のようにするのです。

Aという仮設のもとでは，Bという現象が起こる確率が5％より小であるとき，5％という危険率のもとで仮設を棄却できる，といういい方をします。

＊　＊　＊

◆ では，具体的な問題をとりあげてみましょう。

練習 1. ある町で3000人のうち1000人が流感にかかった。しかし，毎日冷水まさつをしていた10人の学生のうちで，流感にかかったものは2人だけであった。冷水まさつの効果があるといえるか。危険率5％で検定せよ。

（ヒント）「冷水まさつの効果がない」という仮設をおくと，10人の学生のうち流感にかかったものが2人以下である確率は

$$
{}_{10}C_2\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^8+{}_{10}C_1\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^9+\left(\frac{2}{3}\right)^{10}
=\frac{17664}{3^{10}}\fallingdotseq 0.299\cdots\cdots>0.05
$$

つまり「冷水まさつの効果がない」という仮設を棄却できない。いいかえると，冷水まさつの効果はない，ということになります。

練習 2. さいころを4回振ったら1の目が3回出た。このさいころは1の目が出やすいと判断してよいか。危険率5％で検定せよ。

（解）「さいころが正しい」と仮定すると，4回中3回1が出る確率は

$$
{}_4C_3\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^1\fallingdotseq 0.015\cdots\cdots<0.05
$$

ゆえにさいころが正しいという仮設は棄却される。　答　1が出やすい

◆ 検定の問題にもいろいろあります。例えば，

≪銅貨を 10000 回投げたところ5400回表が出た。この銅貨は正しく作られているといえるか≫

というのであれば，「銅貨は正しく作られている」という仮説を設け，10000 回中5400回以上表が出る確率を計算しますと

$$\sum_{k=5400}^{10000} {}_{10000}\mathrm{C}_k\left(\frac{1}{2}\right)^k\left(1-\frac{1}{2}\right)^{10000-k}$$

$$=\left(\frac{1}{2}\right)^{10000}\sum_{k=5400}^{10000} {}_{10000}\mathrm{C}_k$$

となります。この値を計算するのはたいへんです。そこで，これを計算する近似式がほしい，ということになります。

実はこれは近似的に

$$\frac{1}{\sqrt{2\pi}}\int_{7.99}^{100.01} e^{-\frac{t^2}{2}}dt$$

に等しいし，これを計算する表もできていますので問題は解決されます。しかし，この証明は高校の範囲ではありません。

また，

≪6人の学生の体重を測ったら，

52, 54, 61, 59, 61, 55（単位：kg）

なる値を得た。これは平均54 kg の正規母集団からの無作為標本とみなせるか≫

というのであれば，平均54の正規母集団から6個ずつ無作為にとり出したものの性質を知ってないと扱いようがないわけ。

実は，これを扱う理論はできてはいるのですが，高校では扱うことができません。こんなわけで，具体的な問題を扱うことは困難ですが，検定の意味はすでにやった練習でわかったでしょう。

ところで，次のチェビシェフの定理は有用です。

練習 3.（チェビシェフの定理） n 個の数 x_1, x_2, ……, x_n の相加平均を m，標準偏差を σ とするとき，この n 個の数のうち

$$m-k\sigma\leqq x_i\leqq m+k\sigma \qquad \cdots\cdots(*)$$

（k は定数で，$k>1$）

であるものの個数は $n\left(1-\frac{1}{k^2}\right)$ より大きいことを示せ。

（神戸大）

（解） x_1, x_2, ……, x_n の中で不等式 (*) を満足しないものが n' 個あるとし，それを

$$x_\alpha,\ x_\beta,\ \cdots\cdots,\ x_\lambda$$

とすると，

$$\sum_{i=1}^{n}(x_i-m)^2\geqq(x_\alpha-m)^2+(x_\beta-m)^2+\cdots$$
$$\cdots+(x_\lambda-m)^2$$

そして

$$(x_\alpha-m)^2>(k\sigma)^2, \cdots, (x_\lambda-m)^2>(k\sigma)^2$$

$$\therefore\ \sum_{i=1}^{n}(x_i-m)^2>(k\sigma)^2\times n'=n'k^2\sigma^2$$

$$\therefore\ n\sigma^2>n'k^2\sigma^2$$

$$\therefore\ \frac{n}{k^2}>n'$$

$$\therefore\ n-n'>n-\frac{n}{k^2}=n\left(1-\frac{1}{k^2}\right)$$

Q. E. D.

（注） これをいいかえると，

≪n 個の数 x_1, x_2, ……, x_n の相加平均を m，標準偏差を σ とするとき，正の数 k に対して

$$P_r\{|x-m|\leqq k\sigma\}\geqq1-\frac{1}{k^2}$$

である≫

となります。左の銅貨の問題であれば x は

$$p=\frac{1}{2}$$

の二項分布ですから，

$$m=10000\times\frac{1}{2}=5000$$

$$\sigma=\sqrt{npq}=\sqrt{10000\times\frac{1}{2}\times\frac{1}{2}}=50$$

ゆえに　$k\sigma=400$　より　$k=8$

$$P_r\{|x-5000|\geqq400\}\leqq\frac{1}{8^2}\fallingdotseq0.016$$

となります。つまり，5400回表の出る確率は0.016より小ですから，危険率5％はおろか2％でも棄却されることがわかります。

*　*　*

① 母平均の検定

1回目 年 月 日
2回目 年 月 日
3回目 年 月 日

◆推定と検定はまぎらわしい。そこにはいわば，きびしさとやさしさのちがいがゲンゼンとして横たわっている。

◆ 母集団（標準偏差σ）から大きさnの無作為標本をとって，標本の平均を$\bar{x}$であったとします。そこで母集団の平均，つまり母平均はm'であるという仮設をたてて

$$Z=\frac{\bar{x}-m'}{\frac{\sigma}{\sqrt{n}}} \qquad \cdots\cdots(*)$$

とおいたときに

有意水準5％で
$|Z|\geqq 1.96$ であれば，仮設を棄却する
有意水準1％で
$|Z|\geqq 2.58$ であれば，仮設を棄却する

これが母平均の検定の手順なのです。

では，ともあれ，具体的な練習にとりかかるとしようではありませんか。

*　*　*

■練習 1. ある種類のねずみは，生まれてから3か月後の体重が平均65g，標準偏差4.8gの正規分布に従うという。いまこの種類のねずみ10匹を特別な飼料で飼育し，3か月後に体重を測定したところ，下の結果を得た。この飼料はねずみの体重に異常な変化を与えたと考えられるか。有意水準5％で検定せよ。

67, 71, 63, 74, 68, 61, 64, 80,
71, 73 （旭川医大）

ヒント 標本平均は

$$\frac{1}{10}(67+71+63+74+68+61+64+80+71+73)$$
$$=69.2$$

ですから，上の(*)において

$\bar{x}=69.2$, $m'=65$, $n=10$, $\sigma=4.8$

を代入してZを求めてみると

$$|Z|=\left|\frac{69.2-65}{\frac{4.8}{\sqrt{10}}}\right|=\frac{4.2\sqrt{10}}{4.8}$$
$$=\frac{7\sqrt{10}}{8}=\frac{7\times 3.16\cdots}{8}=2.767\cdots\cdots$$

これは1.96より大ですから≪異常な変化を与えない≫という仮設は棄却されます。つまり，異常な変化を与えたといえます。

■練習 2. Y市における全高校3年生男子のうちからランダムに100人を抽出して身長を測ったところ，100人の身長の平均は164 cm，標準偏差は6 cmであった。Y市の3年生男子全員についての平均身長は，全国平均162cmよりも高いといえるか。有意水準0.05で検定せよ。

ヒント (*) において，

$\bar{x}=164$,
$m'=162$, $\sigma=6$
$n=100$

を代入してみますと

$$|Z|=\frac{164-162}{\frac{6}{\sqrt{100}}}=3.3\cdots\cdots$$

この値は1.96より大です。

だから，≪Y市の全高校3年生男子の身長の平均が全国平均と同じ≫という仮設は棄却されます。

つまり，Y市の高校生の身長が全国平均より大きいと判断されるのです。

注 母標準偏差σが不明のときには，標本の大きさnが大きければ，標本の標準偏差を使ってよいのです。練習2.では，このことを使ってあります。

* * *

■練習 3. ある大学の入学試験は，入学定員400名に対し，志願者が2600名あり，500点満点に対し，平均点は285点，標準偏差は72点という結果であった。得点分布が正規分布であるとみなされるとき，次の問いに答えよ。

A高校からの志願者は64名あり，この64名についての平均点は300点であった。A高校の志願者は，志願者全体に比べて優秀であるということができるか。危険率5％で検定せよ。（北海学園大）

ヒント 仮設は≪A高校の志願者は，入学者全体の水準と同じである≫として，この仮設が棄却されるかどうかを検定するわけです。

そうしますと64名の平均点$\overline{X}$の分布は$N\left(285,\ \frac{72^2}{64}\right)$に従い，$\overline{X}=300$ のときの標準化した値は

$$Z=\frac{300-285}{\sqrt{\frac{72^2}{64}}}=1.66\cdots\cdots$$

となります。ところが正規分布表から

$$P(Z \geqq Z_0)=0.05$$

となる Z_0 の値は $Z_0=1.64$ となります。

したがって

$$P(Z \geqq 1.66\cdots)<0.05$$

となり，上の仮設は棄てられます。

注 ここで注意しおきたいことは，**母集団**自身が**正規分布**をしているときには，**標本**の大きさnに関係なく，**標本平均の分布は正規分布**に従うということです。

* * *

◆ では，やや，総合的な問題にいきましょう。

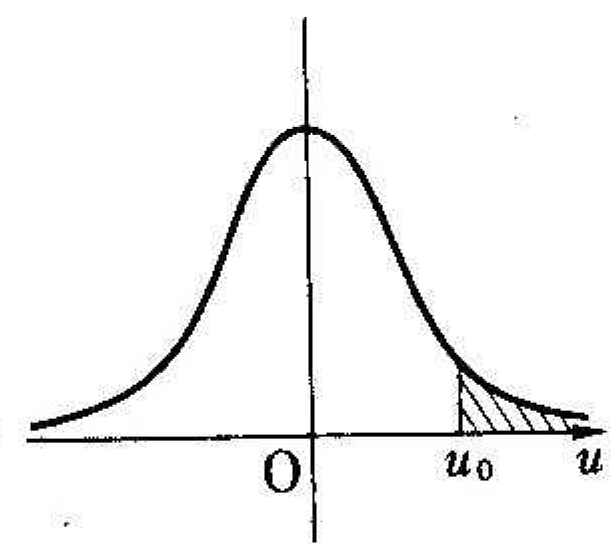
■練習 4. ある工場で生産される製品の中には平均1％の不良品がある。ある日，この工場で作った製品の中から100個抽出したら2個の不良品があった。右の表は標準正規分布 N(0, 1) の表の一部分を抜すいしたものである。これを参照して，次の問いに答えよ。

正規分布表

u_0	$P(u \geqq u_0)$
0.70	0.2420
0.71	0.2389
0.72	0.2358
0.73	0.2327
0.74	0.2296
0.75	0.2266

(1) 合格品に対して1，不良品に対して0なる確率変数を考えたときの標本平均および標本標準偏差を求めよ。

(2) ある危険率で検定すると，この日の状態に異常があったとは断定できない。この危険率の最大値を求めるためにu_0（図参照）の値を小数第3位まで求めよ。（ただし，小数第4位は切り上げること）



(3) この最大値を％で小数第1位まで求めよ。（ただし，小数第2位は切り捨てること）（日本大）

ヒント (1) $x'=E(X)$

$=0\times P(X=0)+1\times P(X=1)$

$=0\times 0.02+1\times 0.98=0.98$ ……答

$\sigma=\sqrt{V(X)}=\sqrt{E(X^2)-\{E(X)\}^2}$

$=\sqrt{0.98-0.98^2}=\sqrt{0.0196}=0.14$ …答

(2) $|Z|=\left|\dfrac{\overline{x}-x'}{\frac{0.14}{\sqrt{n}}}\right|=\left|\dfrac{0.98-0.99}{\frac{0.14}{10}}\right|$

$=0.7142\cdots$ より

$u_0=0.715$ ……答

(3) $P(u \geqq 0.71)=0.2389$

$P(u \geqq 0.72)=0.2358$

から比例部分を使って

$$P(u \geqq 0.715)=\frac{1}{2}(0.2389+0.2358)$$
$$=0.23735$$

よって危険率は

$$2\times P(u \geqq 0.715)=0.4747$$

答 47.4％

○母比率の検定の扱い方

1回目 年 月 日
2回目 年 月 日
3回目 年 月 日

◈検定の問題は多くはない。しかし，比率は比較的よく出るのですよ。これくらいはやっておかなくっちゃ。

◆ 検定についてのだいたいは（☞p.186）を参照してください。ここでは，**比率の検定**について練習することにしましょう。

では，さっそく，これを：——

練習1. 甲は，これまで数学のテストで3題中1題くらいしか解けなかったが，今度のテストでは3題中2題解くことができた。このことから，甲は実力が上がったと判定してよいか，危険率5％で検定せよ。

(ヒント) 甲はこれまで，問題を解く確率が $\dfrac{1}{3}$，解けない確率が $\dfrac{2}{3}$ であったわけです。そこで，実力が上がっていないと仮定すると，3題中2題以上できる確率はどうなるか，と考えるのでしたね。

さて，その確率は

$$ {}_3C_3\left(\frac{1}{3}\right)^3+{}_3C_2\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)=\frac{1}{27}+\frac{6}{27}=\frac{7}{27} $$
$$ =0.259\cdots\fallingdotseq 0.26 $$

です。つまり，実力が上がっていなくても4回に1回くらいは起こりうることなんです。実力が上がったとはいえない，ということになります。解答は次のように書きます。

(解) 甲の実力が上がっていないと仮定すると，問題を解く確率は $\dfrac{1}{3}$ であるから，3題中2題以上解ける確率は

$$ {}_3C_3\left(\frac{1}{3}\right)^3+{}_3C_2\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)\fallingdotseq 0.259>0.05 $$

ゆえに，実力が上がったとは断定できない。

答 上がったといえない。

(注) しかし，2回続けて，2題できたらどうだろう。このときには

$$ \left\{{}_3C_3\left(\frac{1}{3}\right)^3+{}_3C_2\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)\right\}^2\fallingdotseq 0.067\cdots $$

これでも0.05より大ですね。しかし，3回続けて2題できるなら，その確率は0.017…で確かに0.05より小。成績の上がったのは確かといえましょう。

練習2. ある都市で，ある法案に対する賛成者を調べたら，調査人員100人中60人が賛成であった。このことから，この都市では過半数が賛成者であるといえるか。ただし，標準正規分布にしたがう確率変数 u では $p(u>1.645)=0.05$ である。

(ヒント) 賛成，反対が五分五分であったとしますと，賛成者数 x は平均50，標準偏差 $\sqrt{100\cdot\dfrac{1}{2}\cdot\dfrac{1}{2}}=5$ の**二項分布**にしたがうでしょう（☞p.140）。これを正規分布とみると $u=\dfrac{x-50}{5}$ が $N(0,\ 1)$ にしたがうことになります。ゆえに，100人中60人以上賛成する確率は

$$ p\left(u>\frac{60-50}{5}\right)=p(u>2) $$
$$ <p(u>1.645)=0.05 $$

ゆえに，この仮設は捨てられる。つまり過半数の賛成があると認められることになります。

(注) しかし，100人中56人が賛成ならどうでしょう。このときは

$$ u=\frac{56-50}{5}=1.2<1.645 $$

で，もはや過半数の賛成があるとはみられません。

なお，上の解では二項分布の性質を使ってあります。つまり，平均 $m(=np)$，標準偏差

$\sigma(=\sqrt{npq})$ の二項分布は，nが大きいとき正規分布 N(0, 1) として計算してよい，ということです。念のため。

では，もう1つ：——

＊　＊　＊

◆ こんな問題はどうでしょう。

練習 3. ある商社がその見本を1,000人からなる婦人団体に送っているが，返信用に白封筒を入れた従来のやり方では返信は10％であった。今度，返信用に青封筒を入れたところ13％の返信があった。このことから商売上青のほうが役立つといえるか。ただし，送った先の婦人にはなんら差異はないものとする。ただし，N(0, 1) で $p(u>1.645)=0.05$ である。

(ヒント) 青色の封筒と白色の封筒では差がないと仮定して，その仮定が捨てられるか否かを調べるわけ。もっと具体的にいうと，青色のものも返信を得る確率が0.1であるとして，それが0.13返信される確率を調べてみればよいでしょう。では：——

(解) 青色の封筒に効果がなく返信を得る確率が0.1であるとすると，1,000通のうち返信を得る数xは，平均100，標準偏差 $\sqrt{90}(=\sqrt{1000\times0.1\times0.9})$ の二項分布にしたがう。これは正規分布とみなせるから

$$u=\frac{x-100}{\sqrt{90}}$$

は標準正規分布にしたがうと考えてよい。この場合 $x=130$ であったから

$$u=\frac{130-100}{\sqrt{90}}=\sqrt{10}=3.16\cdots>1.645$$

であるから，このことは起こりにくいことである。すなわち，危険率0.05で効果があったと判断してよい。

練習 4. A，B 2つのサイコロがある。どちらも100回振ってみたら，1の目の出た回数がAでは20回，Bでは10回であった。このサイコロには差異があると判断してよいか。危険率5％で検定せよ。ただし，N(0, 1) では $p(u>1.645)=0.05$ とする。

(ヒント) 正しいサイコロとしてA，Bをまず検定してみませんか。

正しいサイコロについては1の目の出る回数は平均 $\frac{100}{6}$，標準偏差 $\sqrt{100\cdot\frac{1}{6}\cdot\frac{5}{6}}$ の二項分布をする。これを正規分布とみますと，

$$u=\frac{x-\frac{100}{6}}{\sqrt{100\cdot\frac{1}{6}\cdot\frac{5}{6}}}=\frac{x-\frac{50}{3}}{\frac{5}{3}\sqrt{5}}=\frac{3x-50}{5\sqrt{5}}$$

が N(0, 1) にしたがうことになります。

Aについて，1が20回以上出る確率は

$$p\left(u>\frac{3\times20-50}{5\sqrt{5}}\right)=p(u>0.89)$$

Bについて，1が10回以下しかでない確率は

$$p\left(u<\frac{3\times10-50}{5\sqrt{5}}\right)=p(u<-1.78\cdots)$$

であるから，Aのほうが正しいという仮設は棄却できませんが，Bのほうは正しいという仮設は棄却できます。つまり，この2つのサイコロは差異があると判断することができます。

しかし，この場合，幸い，一方は正しいという仮設が棄却され，一方は正しいという仮設が棄却されなかったので差異があると判断できましたが，両方とも棄却されたり，両方とも棄却できなかったときに，差異があるかどうかの判断はできません。これは，2つの比率のちがいを調べる問題となって，扱い方もきまっているのですが，高校の範囲ではありません。

ともかく，ここでは，比率の検定とはこういうものだとわかればいいでしょう。詳しいことは専門のほうでやるのです。

＊　＊　＊

◆ くどいけれども最後にもう一度!! 比率の検定の問題は，これを一度 **確率の問題にホンヤクしてしまう** のがコツなのです。

危険率とは何か

1回目　年　月　日
2回目　年　月　日
3回目　年　月　日

◆危険率とは何か，だって!! 答えてもさしつかえないだろうか。その危険率がわからないうちは答えられません，だってさ。

◆ **危険率** というコトバは受験生のもっともキラウものの1つです。具体的な例から出発することにしましょう。

練習 1. 硬貨を5回投げて5回とも表が出た。これは偶然ではないとみるべきか。危険率5％で判定せよ。

ヒント 硬貨を100回投げて100回とも表が出たら，どうもシカケがあるらしい。両面とも表のもようがついているのではあるまいか，と考えるのがふつうでしょう。しかし，5回となると，アヤフヤです。それでは数学的にどう扱うか，それが問題点です。

さて，硬貨を投げて5回とも表が出る確率は $\left(\frac{1}{2}\right)^5=\frac{1}{32}\fallingdotseq 0.031$ です。つまり，このような実験（5回1組の）を100回やると約3回は全部表ということがありうるのです。そこで起こるのは100回中何回くらいを目安にして稀（まれ）なこと，と判定するのか，という問題が起こってきます。この判定の基準を危険率というのです。危険率を5％にとると，3％は小さい。つまり，偶然とは考えられない，ということになりますが，危険率を1％にとると，当然起こりうることが起こっただけだということになります。

解 硬貨がウラとオモテが同様の確からしさで出るとすれば，5回のうち5回ともオモテの出る確率は

$$\left(\frac{1}{2}\right)^5=\frac{1}{32}\fallingdotseq 0.031<0.05$$

であるから，この仮設は棄却される。ゆえに偶然とはみられない。

注 「つまり，硬貨がウラとオモテが同様に出る」という仮設をおいたが100回中5回も起きないなら，この仮設は採用できない，というわけです。では，もう1つ：——

練習 2. Aは，数学のテストで5題中2題解ける実力のあることがわかっている。あるテストで5題中4題を解いた。これは偶然ではないとみるべきか，危険率5％で検定せよ。

解 Aが1問解く確率は $\frac{2}{5}$ であるから，5題中4題解ける確率は

$${}_5C_1\left(\frac{2}{5}\right)^4\left(\frac{3}{5}\right)\fallingdotseq 0.076$$

である。すなわち，Aの実力に変化なく，問題の程度も同じと仮定すると，Aが5問中4問を解く確率は

$$0.076>0.05$$

であるから，この仮設は棄却されない。偶然とみるべきである。

答 偶然である。

注 いつも5題中2題しかできないA君がせっかく4題できても実力向上とは認められないとあっては気の毒であるが，危険率10％とすると，もちろんこの良い点は意味がある。有意である，ということになります。危険率のことを **有意水準**（ゆういすいじゅん）ともいえます。こうして，与えられた危険率，与えられた有意水準をもとにある仮設を判定することを **検定** するといいます。では，これをやってみませんか。

練習 3. サイコロを投げたら5回とも1が出た。このサイコロは正しいサイコロか。危険率5％で検定せよ。

(解) サイコロが正しく作られていたとすると，5回投げて5回とも1が出る確率は

$$\left(\frac{1}{6}\right)^5 \fallingdotseq 0.00012 < 0.05$$

であるから，この仮設は棄却される。つまりこのサイコロは正しく作られていないと判定される。

*　　*　　*

◆ では，やや総合的な問題をやってみましょう。

練習 4. 袋の中に色未知の球が3個入っている。いま，よくかきまぜて2個を同時にとり出し，色を調べた上でもとに戻すという実験を3回行ったところ，とり出された球の色は3回とも赤ばかりであった。袋の中の3球のうち異色の球が含まれているという仮設を，危険率5％で検定せよ。

（高崎経大）

(ヒント) 3球のうち異色の球が含まれているとしますと，例えば

赤　赤　白

ということになりましょう。この中から3個とり出すのですから全体の場合の数は，$_3C_2$，その中でアタリ（つまり赤赤）の場合は1通りですから，2つとも赤である確率は$\frac{1}{3}$となります。3回やって3回とも赤ばかり出る確率は

$$\left(\frac{1}{3}\right)^3 = 0.037\cdots\cdots < 0.05$$

となります。つまり，異色の球が含まれているという仮設は棄却されます。

[答] 異色の球が含まれているとはいえない。

練習 5. 日本人の血液型は，10人に3人の割合で，O型である。南方系のある人種から5人を任意に選んだところ，そのうち4人がO型であった。日本人よりO型の割合が大きいと判定したときに，まちがいとなる可能性を考えてみる。割合がまったく同じときに4人以上O型となる確率は，まちがった判定をする危険率といえる。その危険率を求めよ。

（日本医大）

(ヒント) いささかゴタゴタしていて，文意をとるに手間どったにちがいありません。要するにO型である確率が$\frac{3}{10}$のとき，任意にとった5人のうち4人がO型である確率はどうか，ということなのであろう。それは

$$_5C_1\left(\frac{3}{10}\right)^4\left(\frac{7}{10}\right) = 0.02835$$

答としては 0.02835 では無意味である。0.03 つまり3％というところだ。

[答] 3％

(注) つまり危険率を3％にとれば，日本人よりO型の割合が多いといっていいわけだ。

*　　*　　*

◆ 最後にもう1つ：——

練習 6. ある学校で任意に選んだ10人の学生中8人がある事に賛成した。全体の半数以上が賛成とみてよいか。有意水準5％で検定せよ。また，10人中9人ならばどうか。

(解) 半数が賛成であると仮定すると，任意に選んだ10人中8人以上が賛成である確率は

$$_{10}C_{10}\left(\frac{1}{2}\right)^{10} + {}_{10}C_9\left(\frac{1}{2}\right)^9\cdot\frac{1}{2} + {}_{10}C_8\left(\frac{1}{2}\right)^8\cdot\left(\frac{1}{2}\right)^2$$
$$\fallingdotseq 0.055 > 0.05$$

となるから，半数以上が賛成であるとは断定できない。

しかし，10人中9人以上が賛成である確率は

$$_{10}C_{10}\left(\frac{1}{2}\right)^{10} + {}_{10}C_9\left(\frac{1}{2}\right)^9\cdot\frac{1}{2}$$
$$= \frac{11}{1024} \fallingdotseq 0.010 < 0.05$$

となるから，半数以上が賛成であると断定できる。

(注) 要するに，危険率の問題は危険率というコトバを使わないで考えてみるとよいのだ。

○乱数表とは何か

1回目 年 月 日
2回目 年 月 日
3回目 年 月 日

◈乱数表とは何か，これをどのように使うか，具体的にやってみると，確率の本質がわかってくるだろう。

◆ **乱数表**（らんすうひょう）とは0から9までの数をデタラメに（ランダムに）並べたものです。では，これをどのように使うのか，具体的な例をあげてみましょう。

■練習1. クラス53人の中から，3人を選びたい。どうするか。

ヒント 右の乱数表の上に，目をつぶってエンピツを立てる。当たった点にいちばん近い点を出発点にとるとしましょう。例えば上から10行目，左から11番の3がそれだとしますと

31 42 45 82 39 62 ……

と並んでいますから，クラスの31，42，45の3人を選ぶというわけです。もし，4人なら次は82ですが，82番はいないからとばして次の39を選ぶ，といったぐあいです。

■練習2. 1年365日の中から，勝手に4日を選ぶにはどうするか。

ヒント 1月1日を起点として1, 2, ……, 365まで番号をふってみましょう。そして，上と同じ3からはじめるとすれば

31 42 45 82 39 62 ……

を314, 245, 823, 962, ……と3つずつに切って第314日目，第245日目，823はとばして，……というふうにやってもいいし，823を365で割って得る余り93を採用してもいいでしょう。つまり **剰余類**（☞「数I」p.164）を使うのです。

■練習3. サイコロを振ってやることになっているが，サイコロが見当たらない。乱数表を代用するにはどうするか。

ヒント 乱数表の数をサイコロの目と同じと考えればいいのです。ただし，0と7，8，9をとばして進めばいいのです。

* * *

◆ では，やや，高級な例を：——

■練習4. x軸上原点から出発し，貨幣を投げて表が出たら右へ1だけ進み，裏が出たら左へ1だけ進むことにする。これを4回くり返したとき $x=0$ にいる確率を実験的に求めたい。乱数表をどのように使えばよいか。

ヒント 右の乱数表の左上から使うとしましょう。偶数(0, 2, 4, 6, 8)が出たら右へ1だけ，奇数(1, 3, 5, 7, 9)が出たら左へ1だけ進むことにしてやってみます。そうすると，0116は 右左左右 となりましょう。4055なら 右右左左 といったぐあい。

このようなやり方で，面積や体積を求めたり，めんどうな微分方程式を解いたりすることもできます。これを **モンテカルロ法** と呼んでいます。

右の乱数表は数は少ないのですが，それでも使い方をくふうすると，かなり有効に使えます。まず，横の数字を使って，全部使い終わったら，次は左上から下へ

0 0 6 8 3 3 2 8 1 9 ……

と使うこともできましょう。

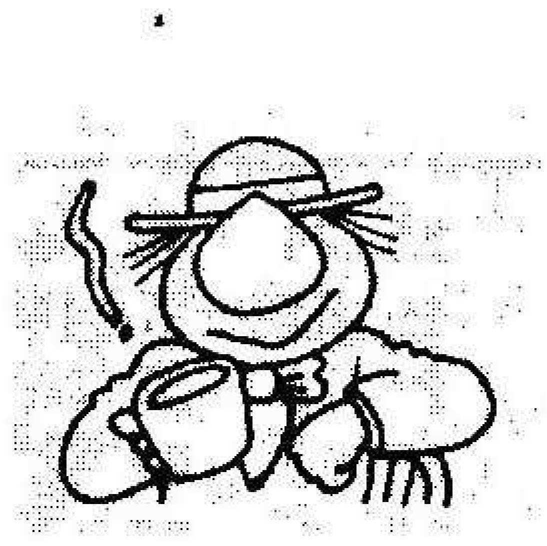
* * *

◆ なお，乱数表というのは数字をデタラメに並べたと申しましたが，デタラメに並んでいることを確かめるにはどうしたらいいでしょうか？

こう考えてみると，この表もいろいろおもしろい問題をはらんでいるといえます。

◆ 乱 数 表 ◆

01 16	40 55	80 94	88 73	24 94	23 20	10 86	33 14	86 24	48 81
08 71	07 30	85 05	38 47	51 54	13 18	31 43	43 11	19 61	97 45
69 86	11 05	48 20	90 41	23 42	61 13	31 52	75 12	27 78	37 66
87 02	32 43	04 15	81 90	46 09	37 21	34 12	26 46	35 69	90 95
34 00	74 49	67 48	93 30	65 64	97 77	01 02	20 93	72 23	22 64
37 51	62 45	50 39	76 84	64 91	23 98	41 30	61 32	91 67	20 40
27 01	33 79	19 75	76 14	68 54	81 77	67 86	51 04	73 45	70 52
83 65	65 83	09 60	87 51	24 76	01 37	00 29	03 31	31 67	86 74
17 88	16 98	75 74	44 29	83 35	79 40	27 86	96 77	46 48	11 21
95 15	90 92	56 31	42 45	82 39	62 99	29 86	31 32	31 15	11 80
49 68	05 14	81 42	89 32	08 06	14 17	16 12	28 37	64 08	28 51
43 78	91 92	09 75	59 35	49 83	45 75	73 32	15 74	42 39	79 45
20 61	73 15	10 78	08 22	35 33	77 41	72 38	42 14	74 50	69 68
63 14	77 30	47 17	17 32	49 58	67 83	33 96	78 48	62 51	35 27
38 69	63 90	45 33	49 20	32 30	06 29	02 21	71 34	68 01	83 53
56 33	34 49	32 64	56 05	62 06	29 67	63 70	14 22	21 13	97 84
27 92	94 83	15 71	14 99	66 65	18 56	69 64	47 00	29 35	38 80
97 32	42 89	89 30	14 62	74 91	76 83	37 70	22 19	24 13	32 05
22 40	72 30	06 85	60 80	41 59	40 14	74 68	86 31	63 09	97 97
18 16	99 16	23 32	31 47	79 91	92 73	66 58	83 18	25 55	82 42
99 38	91 31	43 21	32 55	29 46	80 14	11 63	49 60	62 28	31 97
84 32	96 36	04 31	81 45	76 63	15 92	37 82	03 16	23 60	33 04
39 97	67 49	95 09	88 40	66 85	06 32	09 45	80 60	18 09	92 57
20 92	84 09	47 30	62 36	05 55	32 80	88 22	97 36	97 45	33 29
06 02	23 83	41 35	06 35	87 41	73 08	14 31	83 06	91 02	39 23
87 54	91 30	32 39	84 41	74 68	60 13	74 33	07 31	64 67	57 86
08 14	35 09	76 50	83 72	30 03	35 48	55 44	43 47	67 06	83 95
66 08	44 91	79 24	90 74	03 32	86 29	79 02	10 09	46 54	68 66
30 60	81 01	88 80	08 12	97 34	33 33	27 87	83 54	40 36	60 33
00 11	04 18	10 93	06 75	46 62	07 18	19 25	04 96	53 58	93 31
31 10	53 75	44 42	65 24	84 88	11 64	46 59	14 44	32 74	85 99
25 07	66 12	68 98	52 83	74 23	64 83	84 80	12 93	28 15	63 29
07 12	52 48	79 83	48 62	37 13	67 88	16 62	62 21	14 54	44 60
05 53	16 97	72 54	47 01	49 33	85 84	59 54	72 12	52 71	97 60
96 75	89 62	09 62	77 88	15 94	07 85	79 33	98 33	08 09	13 80
32 19	56 55	15 98	09 83	49 33	73 25	95 31	11 85	46 95	22 66
74 64	19 24	14 56	76 29	84 64	93 67	10 18	17 88	91 32	11 69
31 64	54 38	32 99	71 16	49 79	89 45	20 48	30 28	31 53	38 49
56 92	94 13	57 21	07 14	28 83	70 52	56 72	52 82	71 07	69 09
28 64	08 46	33 27	78 32	55 90	98 44	83 87	99 95	17 78	83 34
10 97	10 07	98 99	48 23	12 63	57 43	65 53	99 29	95 64	88 99
48 67	52 92	33 43	95 09	06 94	64 83	25 91	32 54	22 61	81 19
87 13	17 11	16 76	04 27	36 43	16 43	08 51	52 38	08 97	47 39
04 51	56 77	27 94	17 35	50 73	75 14	38 49	50 71	81 09	66 96
35 17	51 07	38 80	83 03	93 82	42 66	48 31	55 77	87 01	26 16
16 46	18 69	90 05	70 40	32 24	79 41	93 03	98 19	50 87	02 65
67 55	52 30	76 32	48 25	59 19	75 36	62 76	87 53	26 39	38 66
67 65	61 42	75 05	50 45	66 98	86 63	06 98	01 05	50 91	04 66
20 07	24 15	25 32	15 75	40 85	51 85	55 17	27 89	52 18	54 03
77 62	78 88	62 44	30 07	44 68	02 44	01 64	56 13	35 03	96 00



片側検定と蛇足

◆数学には数学的世界観がある。しかし，そこにはどうしても世俗的なものが入ってくる。たとえば，片側検定などはどうかな。

◆ ある科目のテストにおいて，大部分の人は，40点から70点の範囲に入った。しかし，おれは，「残念ながら入らなかった」というのと「幸いに入らなかった」というのでは意味が大いにちがうでしょう。このときには，0点以上39点以下と71点以上100点以下と，はっきりわかれるからです。このことは統計においてももちろん起こってきます。では，次はどうです。

練習 1. 製品にはほぼ10％の不良品が含まれていることがわかっている。新しい製法に変えたところ，100個の無作為標本のうち不良品が3個あらわれた。新しい製法は，今までよりもすぐれているといえるか。有意水準5％で検定せよ。

ヒント　仮設として≪新しい製法でも $p=0.1$ とする≫をとります。

そして，この仮定のもとに100個のうちにあらわれる不良品の個数をXとしますと，Xは二項分布B(100, 0.1) に従うのですから，nが大きいことから，Xは正規分布とみなしてよいでしょう。そこで，

$$m=100\times0.1=10,$$
$$\sigma^2=100\times0.1\times0.9=3^2$$

から，Xは正規分布

$$N(10,\ 3^2)$$

に従うとみてよいでしょう。

標準化しますと

$$Z=\frac{X-10}{3}\ \text{は}\ N(0,\ 1)\ \text{に従う}$$

のですから，$X=3$ のとき

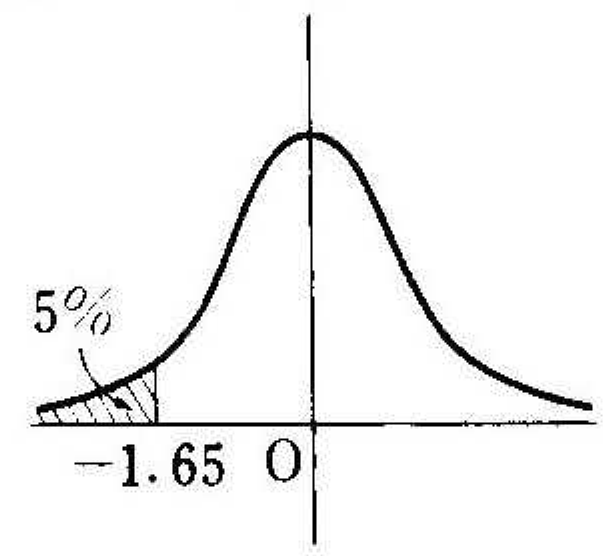
$$Z=\frac{3-10}{3}=-2.3\cdots<-1.65$$

となります。

したがって，はじめの仮設は成立しないことがわかります。したがって

答　新しい製法は，従来よりすぐれているといってよい。

注　上の場合には右の図の左の方の5％だけを考えに入れているわけです。これが**片側検定**というものです。



練習 2. A市内の高校3年生男子の，昨年度の平均身長は168.2cmであった。今年は，100人を無作為抽出して平均身長を調べたところ169.3cmであった。また，標準偏差は5.8cmであった。今年は去年にくらべて身長が大きくなったといえるか。有意水準5％で検定せよ。

ヒント　この場合，身長が小さくなった，ということは考える必要はないでしょう。

さて，仮設≪今年も昨年と同じく身長は168.2cmである≫を採用し，これが棄却されるかどうか調べてみましょう。

nが十分大きいから $\sigma=5.8$

として

$$Z=\frac{\overline{X}-168.2}{\dfrac{5.8}{\sqrt{100}}}\ \text{は}\ N(0,\ 1)\ \text{に従う}$$

$\overline{X}=169.3$ ですから

$$Z=\frac{10(169.3-168.2)}{5.8}=1.90>1.65$$

つまり，この仮設は棄却される。つまり，今年は昨年より平均身長は伸びている!!

このような片側検定は問題の性格から来ていることを忘れてはいけません。