

昭和58年度（問題）

1. 死亡率が $\frac{6}{1,000}$ であった1,000人からなる集団において、3人がある1年間で死亡した。死亡の発生がポアソン分布にしたがっているものとして、死亡率が改善されたと結論できるか、有意水準5%で検定せよ。

(注) $e \approx 2.7$

2. 品質基準として純度98%以上の原料の購入を望んでいる会社が、抜取検査として、納入ロットから10個の標本を抽出してその純度を調べ、この平均が98.8%以上なら、そのロットを合格させている。

- (1) 純度の平均値が99%であるロットの合格する確率を求めよ。
 (2) 純度の平均値が品質基準に達しないロットの合格する確率を求めよ。

ただし、一つのロットを構成する各原料の純度の標準偏差は $\sigma = 0.6\%$ であることがわかっているものとする。

また、解答にあたっては次の正規分布表を用いよ。

ϵ	0.25	0.159	0.15	0.10	0.05	0.025	0.0228	0.01	0.005	0.00135	0.001	0.00003
$u(\epsilon)$	0.674	1.00	1.036	1.282	1.645	1.960	2.00	2.326	2.576	3.000	3.09	4.00

3. 1,000人を性別と色盲か否かについて調べたところ、表1の分割表を得た。色盲の遺伝に関するモデルによれば、それらの確率は表2のようなになるはずである。表1のデータは表2の確率分布に合っていると考えるよいか。ただし、表2で $q = 1 - p$ であって、 q ($0 < q < 1$)は色盲の遺伝子の比率を表わしている。

ここで、 $\chi^2_1(0.05) = 5.99$ 、 $\chi^2_2(0.05) = 7.81$ 、 $\chi^2_3(0.05) = 9.49$ とする。

表1

	男	女
正 常	442	514
色 盲	38	6

表2

	男	女
正 常	$p/2$	$p^2/2 + pq$
色 盲	$q/2$	$q^2/2$

4. 正規分布 $N(\mu, \sigma^2)$ の μ の推定に t 分布を用いて作った90%信頼区間の長さが、 σ より小さくなる確率が0.95より大であるためには、どの位の大きさの標本が必要か。

必要があれば次の数値を用いよ。

n	1	2	3	4	5	6	7	8	9	10
$t_{\alpha}(0.05)$	6.31	2.92	2.35	2.13	2.01	1.94	1.89	1.86	1.83	1.81
$\chi^2_{\alpha}(0.05)$	3.84	5.99	7.81	9.49	11.1	12.6	14.1	15.5	16.9	18.3

n	11	12	13	14	15	16	17	18	19	20
$t_{\alpha}(0.05)$	1.80	1.78	1.77	1.76	1.75	1.75	1.74	1.73	1.73	1.72
$\chi^2_{\alpha}(0.05)$	19.7	21.0	22.4	23.7	25.0	26.3	27.6	28.9	30.1	31.4

5. 非負の区間 $[0, \infty)$ で定義されている確率変数 X が、連続な確率密度関数 $f(x)$ をもち、すべての t, t' に対して次の2条件

$$(1) \quad P(X > t) > 0$$

$$(2) \quad P(X \leq t + t' \mid X > t) = P(X \leq t')$$

をみたすときは、 X は指数分布にしたがうことを証明せよ。逆に、 X が指数分布にしたがうときは、上の2条件をみたすことを証明せよ。

昭和58年度（解答例）

1. $n = 1,000$ の集団の死亡率 P について、

$$\text{帰無仮説： } P = \frac{6}{1,000}$$

$$\text{対立仮説： } P < \frac{6}{1,000}$$

の検定を有意水準 5% で行う。

今、 n 人の集団の死亡率を p とするとき、死亡の発生がポアソン分布に従っていることから、この集団の死亡者数は平均 np のポアソン分布に従う（ポアソン分布の再生性）。

従って、この集団の死亡者数を表わす確率変数を X とするとき、

$$P_r(X \leq k | \lambda = np) = \sum_{i=0}^k e^{-np} \frac{(np)^i}{i!} < 0.05 \quad \text{--- (*)}$$

を満す最大の k を k' とし、 $0 \leq x \leq k'$ がこの検定の棄却域となる（逆に、 $x > k'$ であれば採択域に入る）。

$$k = 3, \quad n = 1,000, \quad p = \frac{6}{1,000} \quad \text{を用いて、}$$

$$\sum_{i=0}^3 e^{-6} \frac{6^i}{i!} = e^{-6} \times \left(\frac{1}{1} + \frac{6}{1} + \frac{36}{2} + \frac{216}{6} \right)$$

$$\approx 2.7^{-6} \times 61$$

$$\approx 0.157 > 0.05$$

即ち、 $k = 3$ は (*) を満す最大の k' を超えている。

よって、仮説は棄却できないから、死亡率が改善されたとは有意水準 5% で結論できない。

(注1) 二項分布で直接計算することも出来るが、 np が小さいので、正規分布で近似することは好ましくない。

(注2) $\sum_{i=0}^k e^{-np} \frac{(np)^i}{i!}$ が自由度 $\phi = 2(k+1)$ の χ^2 -分布によって表わされること

を用いて、死亡者数 $k (= 3)$ に対して、自由度 $\phi = 2(k + 1) = 8$ の χ^2 -分布の上側 5% 点 $\chi^2_{2(k+1)}(0.05)$ が $2np (= 12)$ を超えていれば H_0 を採択としても良いが、この間では $\chi^2_{2(k+1)}(0.05) = 15.5$ の値は与えられていない。

2. 納入ロットの原料の純度を μ 、標準偏差を σ とすれば、納入ロットから 10 個の標本を抽出したとき、その純度の標本変量平均 $\bar{X} = \sum_{i=1}^{10} \frac{X_i}{10}$ (X_i は各原料の純度) は正規分布 $N\left(\mu, \frac{\sigma^2}{10}\right)$ に従うと考えてよい。

従って、純度 μ から $n = 10$ の標本をとったとき、このロットの合格する確率は、

$$P_r(\bar{X} > 98.8) = P_r\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{98.8 - \mu}{\frac{\sigma}{\sqrt{n}}}\right) \quad \text{--- ①}$$

となり、 $U = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ が正規分布 $N(0, 1)$ に従うことから、①式のカッコ内右辺

$\frac{98.8 - \mu}{\frac{\sigma}{\sqrt{n}}}$ を計算して正規分布表から確率を求めればよい。

- (1) $\mu = 99\%$ の合格する確率

$$\begin{aligned} P_r\left(U > \frac{98.8 - 99}{\frac{0.6}{\sqrt{10}}}\right) &\Rightarrow P_r(U > -1.054) \\ &= 1 - P_r(U > 1.054) \quad \text{--- ②} \end{aligned}$$

ここで $P_r(U > 1.054)$ を直線近似で求めれば、

$$\begin{aligned} P_r(U > 1.054) &\Rightarrow 0.15 + \frac{1.054 - 1.036}{1.282 - 1.036} \times (0.10 - 0.15) \\ &\Rightarrow 0.1463 \end{aligned}$$

$$\therefore \text{②式} \Rightarrow 1 - 0.1463 = 0.8537$$

純度の平均値が 99% であるロットの合格する確率はほぼ 85.4%

- (2) $\mu < 98\%$ の合格する確率

$$P_r\left(U > \frac{98.8 - \mu}{\frac{0.6}{\sqrt{10}}}\right) < P_r\left(U > \frac{98.8 - 98}{\frac{0.6}{\sqrt{10}}}\right) \quad \text{--- ③}$$

ここで、 $\mu < 98\%$ より $98.8 - \mu > 98.8 - 98$ であること、 $u(\epsilon)$ に対して確率 ϵ が単調減少であることを用いた。

$$\begin{aligned} \text{今, } P_r\left(U > \frac{98.8 - 98}{\frac{0.6}{\sqrt{10}}}\right) &= P_r(U > 4.216) \\ &< P_r(U > 4) \\ &= 0.00003 \end{aligned} \quad \text{--- ④}$$

$$\text{③, ④より } P_r\left(U > \frac{98.8 - \mu}{\frac{0.6}{\sqrt{10}}}\right) < 0.00003$$

従って、品質基準に達しないロットの合格する確率は0.003%未満で、ほぼ0%と考えてよい。

3. 実際のデータ(表1)がモデル(表2)に合っているかを調べるのであるから適合度の検定となるが、適合度の検定に用いる理論期待値を求めなければならない。

そこで、まず p (および q) の推定量 \hat{p} (および \hat{q}) を最尤法を用いて求める。

$$\text{尤度関数 } L = \left(\frac{p}{2}\right)^{442} \left(\frac{p^2}{2} + pq\right)^{514} \left(\frac{q}{2}\right)^{38} \left(\frac{q^2}{2}\right)^6$$

$q = 1 - p$ を用いて

$$\begin{aligned} \frac{\partial \log L}{\partial p} &= \frac{956}{p} - \frac{50}{1-p} - \frac{514}{2-p} \\ &= \frac{1912 - 3482p + 1520p^2}{p(1-p)(2-p)} = 0 \quad (0 < p < 1) \end{aligned}$$

よって、 $\hat{p} = 0.91294$, $\hat{q} = 0.08706$

これより、442, 514, 38, 6 に対する理論期待値は各々、

$$\begin{aligned} n \frac{\hat{p}}{2} &= 456.47, & n \left(\frac{\hat{p}^2}{2} + \hat{p}\hat{q}\right) &= 496.21 \\ n \frac{\hat{q}}{2} &= 43.53, & n \frac{\hat{q}^2}{2} &= 3.79 \end{aligned}$$

ここで、 $n = 442 + 514 + 38 + 6 = 1000$ 。

次に、

H_0 : 表1のデータは表2の確率分布に合っている。

として、表1の実現値と上記理論期待値を用いて χ^2 -検定を行う。なお、自由度は、理論期待値を求めるとき1つの母数 p を推定しているため、 $4 - 1 - 1 = 2$ となる。

$$\begin{aligned}\chi^2 &= \sum \frac{(\text{実現値} - \text{理論期待値})^2}{\text{理論期待値}} \\ &= \frac{(442 - 456.47)^2}{456.47} + \frac{(514 - 496.21)^2}{496.21} + \frac{(38 - 43.53)^2}{43.53} + \frac{(6 - 3.79)^2}{3.79} \\ &\approx 3.09 < 5.99 = \chi_2^2(0.05)\end{aligned}$$

即ち、表1のデータは表2の理論分布に矛盾するものではない。

4. 正規分布 $N(\mu, \sigma^2)$ からの標本を $X_i (i = 1, 2, \dots, n)$ とするとき、 $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ は

$N(0, 1)$ に従う。ここで、 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 。

また、 $\frac{ns^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$ は自由度 $\phi = n - 1$ の χ^2 -分布に従う。

したがって、 $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \bigg/ \sqrt{\frac{ns^2}{\sigma^2} / (n-1)} = \frac{\bar{X} - \mu}{s/\sqrt{n-1}}$ は自由度 $\phi = n - 1$ の

t -分布となる。

$$\text{よって、} Pr \left\{ \frac{|\bar{X} - \mu|}{s/\sqrt{n-1}} < t_{n-1}(0.05) \right\} = 0.9$$

より、 μ の90%信頼区間は t -分布を用いて

$$\left(\bar{X} - t_{n-1}(0.05) \frac{s}{\sqrt{n-1}}, \bar{X} + t_{n-1}(0.05) \frac{s}{\sqrt{n-1}} \right)$$

この信頼区間の長さ L は、 $L = 2 t_{n-1}(0.05) \frac{s}{\sqrt{n-1}}$ となる。

題意より、この L が σ より小さくなる確率が0.95より大であるから、

$$Pr\{L < \sigma\} = Pr\left\{2t_{n-1}(0.05) \frac{s}{\sqrt{n-1}} < \sigma\right\} \quad \text{--- ①}$$

$t_{n-1}(0.05)$, s とともに正だから,

$$\begin{aligned} \text{①} &= Pr\left\{4(t_{n-1}(0.05))^2 \frac{s^2}{n-1} < \sigma^2\right\} \\ &= Pr\left\{\frac{ns^2}{\sigma^2} < \frac{n(n-1)}{4(t_{n-1}(0.05))^2}\right\} > 0.95 \quad \text{--- ②} \end{aligned}$$

ここで, $\frac{ns^2}{\sigma^2}$ は前記のとおり自由度 $\phi = n-1$ の χ^2 -分布に従うから, ②式を満す

ためには,

$$\frac{n(n-1)}{4(t_{n-1}(0.05))^2} > \chi^2_{n-1}(0.05) \quad \text{--- ③}$$

であればよい。

与えられた表より $n = 2, 3, \dots, 20$ を計算すれば, ③式の左辺, 右辺について,

$$\text{左辺} < \text{右辺} \quad (n = 2, 3, \dots, 19)$$

$$\text{左辺} > \text{右辺} \quad (n = 20)$$

となり, $n = 20$ で③式が初めて成立。

よって, 標本数は20以上必要である。

5. (1) $P(X > t) > 0$
 (2) $P(X \leq t+t' | X > t) = P(X \leq t')$ \iff X は指数分布に従う。

[証明]

(\Rightarrow) $G(t) = P(X > t)$ とする。

$$(1) \text{より } G(t) > 0 \quad (t \geq 0)$$

$$\begin{aligned} (2) \text{の左辺 } P(X \leq t+t' | X > t) &= \frac{P(X > t, X \leq t+t')}{P(X > t)} \\ &= \frac{P(t < X \leq t+t')}{P(X > t)} \\ &= \frac{P(X > t) - P(X > t+t')}{P(X > t)} \quad \text{--- ①} \end{aligned}$$

$$= \frac{G(t) - G(t+t')}{G(t)}$$

(2)の右辺 $P(X \leq t') = 1 - P(X > t')$
 $= 1 - G(t')$

よって, $\frac{G(t) - G(t+t')}{G(t)} = 1 - G(t')$

これを整理して, $G(t+t') = G(t) \cdot G(t')$ ($t, t' \geq 0$) — ②

ここで, $G(t) = P(X > t) = 1 - P(X \leq t) = 1 - \int_0^t f(x) dx$ — ③

より, $G(0) = 1 - \int_0^0 f(x) dx = 1$

②より,

$$\begin{aligned} \lim_{t' \rightarrow +0} \frac{G(t+t') - G(t)}{t'} &= \lim_{t' \rightarrow +0} \frac{G(t) \cdot G(t') - G(t)}{t'} \\ &= \lim_{t' \rightarrow +0} G(t) \frac{G(t') - 1}{t'} \\ &= G(t) \lim_{t' \rightarrow +0} \frac{G(t') - 1}{t'} \\ &= G(t) G'(0) \end{aligned}$$

よって, $G'(t) = G(t) \cdot G'(0)$ — ④

ここで, $t' \geq 0$ より $G'(0)$ は右側微分係数

さて, ③より $G'(t) = \frac{d}{dt} \left\{ 1 - \int_0^t f(x) dx \right\}$
 $= -f(t)$ — ⑤

であるから $G'(0) = -f(0)$ となり, $f(x)$ は連続な確率密度関数だから
 $\lambda = f(0)$ とすると, $\lambda \geq 0$ 。

④より $G'(t) = -\lambda G(t)$ ($\lambda \geq 0, t \geq 0$)

これより $\frac{G'(t)}{G(t)} = -\lambda$

$$\therefore \log G(t) = -\lambda t + c$$

$$G(t) = e^{-\lambda t} \cdot e^c = C e^{-\lambda t}$$

$$G(0) = 1 \text{ より, } C = 1 \text{ となるから } G(t) = e^{-\lambda t} \quad (\lambda \geq 0, t \geq 0)$$

$$\textcircled{5} \text{ より } f(t) = -G'(t) = \lambda e^{-\lambda t} \quad (\lambda \geq 0, t \geq 0)$$

これは指数分布の確率密度関数に他ならない。

$$(\Leftarrow) \quad f(x) = \lambda e^{-\lambda x} \quad (\lambda \geq 0, x \geq 0) \text{ とすると,}$$

$$1. \quad P(X > t) = \int_t^{\infty} f(x) dx = \int_t^{\infty} \lambda e^{-\lambda x} dx$$

$$= \left[-e^{-\lambda x} \right]_t^{\infty} = e^{-\lambda t} > 0$$

$$\square. \quad P(X \leq t+t' | X > t) = \frac{P(X > t) - P(X > t+t')}{P(X > t)} \quad (\textcircled{1} \text{ より})$$

$$= \frac{e^{-\lambda t} - e^{-\lambda(t+t')}}{e^{-\lambda t}} \quad (\text{上記イ. より})$$

$$= 1 - e^{-\lambda t'} \quad \text{--- } \textcircled{6}$$

$$P(X \leq t') = 1 - P(X > t')$$

$$= 1 - e^{-\lambda t'} \quad (\text{上記イ. より}) \quad \text{--- } \textcircled{7}$$

$$\textcircled{6}, \textcircled{7} \text{ より } P(X \leq t+t' | X > t) = P(X \leq t')$$

(証明終)