数理統計学
(第15回)

機械工学科
塩幡　宏規

5.5　検定−母分散の検定

5.5.1　母平均 μ が未知の場合

母集団～正規分布N(μ, σ^2)

無作為標本(X_1, ‥‥, X_n)に対する統計量S

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

仮説: H_0: $\sigma^2 = \sigma_0^2$

〔対立仮説 H_1: $\sigma^2 \neq \sigma_0^2$ 〕

とおくと，次の統計量 χ^2 は自由度 $(n-1)$ の χ^2 分布に従う

$$\chi^2 = \frac{nS^2}{\sigma^2} \sim \chi^2_{n\text{-}1}$$

有意水準αのとき

$$P\left(\chi^2_{n-1} \leqq k_1,\ \chi^2_{n-1} \geqq k_2\right) = \alpha$$

$$\Rightarrow \quad k_1 = \chi^2_{n-1}(1-\alpha/2)$$
$$k_2 = \chi^2_{n-1}(\alpha/2)$$

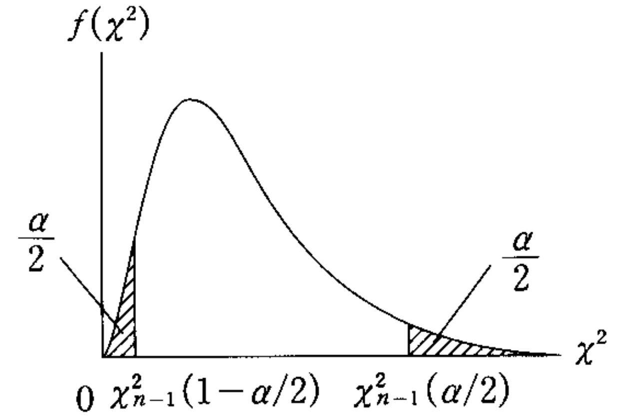


図 5.5

有意水準αに対する棄却領域

$$W = (0, k_1) \cup (k_2, \infty)$$

χ^2の実現値 : $\chi_0{}^2$
に対して

$$\chi_0^2 \in W \ : H_0 \text{を棄却する}$$
$$\chi_0^2 \notin W : H_0 \text{を採択する}$$

◆例5.7　ある食品工場で生産されている製品の重量の標準偏差は0.245で管理されている. この工場の製品から無作為に10個を抽出し重量を測定したところ標準偏差は0.268であった. 工場の管理水準が維持されているがどうかを, 有意水準5%で検定せよ.

解　仮説　: ばらつきに差がない

$H_0: \sigma^2 = 0.245^2 = 0.060$　〔対立仮説$H_1: \sigma^2 \neq 0.245^2$〕

$n=10$,　$s^2 = 0.268^2 = 0.072$

⇨　$\chi_0^2 = 10 \times 0.072/0.060 = 12.0$

有意水準$\alpha = 0.05$の時, $\chi_9^2(0.975) = 2.700$,　$\chi_9^2(0.025) = 19.02$

$$W = (0, 2.700) \cup (19.02, \infty)$$
$$\therefore \chi_0^2 = 12.0 \notin W$$

仮説は有意水準5%で採択される.

5.5.2　母平均 μ が既知の場合

母集団～正規分布N(μ, σ^2)

無作為標本(X_1, ・・・・, X_n)に対する統計量S

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

仮説: H_0: $\sigma^2 = \sigma_0^2$〔対立仮説H_1: $\sigma^2 \neq \sigma_0^2$〕

$$統計量\chi^2 = \frac{nS_0^2}{\sigma^2} \sim \chi_n^2$$

有意水準αのとき　$P\left(\chi_n^2 \leqq k_1',\ \chi_n^2 \geqq k_2'\right) = \alpha$ ⇨ $k_1' = \chi_n^2(1-\alpha/2)$, $k_2' = \chi_n^2(\alpha/2)$

有意水準αに対する棄却領域

$$W = \left(0, k_1'\right) \cup \left(k_2', \infty\right)$$

χ^2の実現値: χ_0^2に対して

$\chi_0^2 \in W$: H_0を棄却する

$\chi_0^2 \notin W$: H_0を採択する

◆**例5.8** 平均身長167cmの学生の集団から無作為に6人抽出して次のデータを得た．身長は正規分布をするものとして，この母分散が1.5^2であるという仮説を有意水準5%で検定せよ．

167cm　173cm　163cm　170cm　166cm　165cm

解　仮説H_0: $\sigma^2 = 1.5^2$ (cm^2)〔対立仮説H_1:$\sigma^2 \neq 1.5^2$(cm^2)〕

分散$S_0{}^2$の実現値$s_0{}^2 = 11.17$(cm^2)

χ^2の実現値$\chi_0{}^2 = 6 \times 11.7 / 1.5^2 = 29.79$

有意水準$\alpha = 0.05$の時，

$\chi_6{}^2(0.975) = 1.237$，$\chi_6{}^2(0.025) = 14.45$

$$W = (0, 1.237) \cup (14.45, \infty)$$
$$\therefore \chi_0^2 = 29.79 \in W$$

仮説は有意水準5%で棄却される．

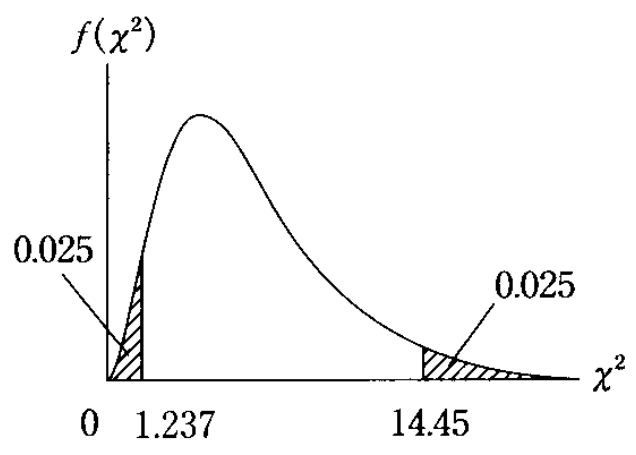


図 5.6

5.6　検定−母相関係数の検定

2次元の確率変数(X,Y) ～ 2次元正規分布 $\mathrm{N}(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$

$\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho$：すべて未知

母相関係数ρが0であることの検定

統計量：

$$R=\frac{\frac{1}{n}\sum_{i=1}^{m}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}\qquad\left(\begin{aligned}\bar{X}&=\frac{1}{n}\sum_{i=1}^{n}X_i\\ \bar{Y}&=\frac{1}{n}\sum_{i=1}^{n}Y_i\end{aligned}\right)$$

仮説：H_0：ρ＝0〔対立仮説H_1：ρ ≠ 0〕

$$T=\frac{\sqrt{n-2}R}{\sqrt{1-R^2}}\ (n\geqq 3)\quad\sim t_{n-2}$$

$$\mathrm{P}(|T|\geqq t_{n-2}(\alpha))=\alpha \Rightarrow t_{n-2}(\alpha)$$

$$T\text{の実現値}t_0=\frac{\sqrt{n-2}\,r}{\sqrt{1-r^2}}$$

$|t_0|\geqq t_{n-2}(\alpha)$：$H_0$を棄却する

$|t_0|<t_{n-2}(\alpha)$：H_0を採択する

◆例5.9 乳幼児30人について身長と体重の相関係数を調べたところ0.74であった．身長と体重の値が2次元正規分布をしているとして母相関係数ρ＝0であるという仮説を有意水準5%で検定せよ．

解　仮説H_0: $\rho = 0$〔対立仮説$\rho \neq 0$〕

$n = 30$, $r = 0.74$　Tの実現値$t_0 = 5.822$

有意水準$\alpha = 0.05$の時，自由度28のt分布から

$t_{n-2}(0.05) = 2.048$

$$\therefore |t_0| = 5.822 > 2.048$$

仮説は有意水準5%で棄却される．

5.7　検定−母比率の検定

5.7.1　母比率$p=p_0$の検定

母集団比率：p　　標本（X_1，‥‥，X_n）　$S_n=\Sigma X_i$

S_n ～2項分布B(n, np)

nが大きいときには正規分布N(np, np(1-p))に従う

仮説H_0: $p=p_0$〔対立仮説H_1: $p\neq p_0$〕

$$\text{統計量}Z=\frac{S_n-np_0}{\sqrt{np_0(1-p_0)}}=\frac{S/n-p_0}{\sqrt{p_0(1-p_0)/n}}\left(\frac{S}{n}\text{は標本比率}\right)$$

nが大のとき，Z～正規分布N(0, 1^2)に従う

$$\mathrm{P}(|Z|>\lambda)=\alpha$$

⇒　λが求められる

$$Z\text{の実現値}z_0=\frac{S_n/n-p_0}{\sqrt{p_0(1-p_0)/n}}$$

$|z_0|\geqq\lambda$: H_0を棄却する

$|z_0|<\lambda$: H_0を採択する

◆例5.10　ある選挙区で選挙人名簿の中から無作為に抽出した300人について面接したところ165人がA候補者を支持していた. この選挙区で支持率が半数より大きいと見てよいか

解　A候補の支持率が半数であるという仮説

H_0: p＝0.5〔対立仮説H_1 : $p \neq 0.5$〕

Zの実現値 $\therefore \left|z_0\right| = \dfrac{1.65/300 - 0.5}{\sqrt{0.5(1-0.5)/300}} = \dfrac{0.05}{0.0289} = 1.73$

有意水準α＝0.05とすると, λ＝1.96

$\therefore \left|z_0\right| = 1.73 > 1.96$

仮説は有意水準5%で採択される.

5.7.2　2つの比率の差の検定

2つの2項母集団A,B→標本の大きさn_A, n_B

→1つの事象に注目した結果，それに属する個数をk_1, k_2のとき

→標本比率を$p_A{}^* = k_1 / n_A$，$p_B{}^* = k_2 / n_B$ の差の検定

母集団比率p_A，p_Bとし，仮説

H_0: $p_A = p_B$　〔対立仮説H_1: $p_A \neq p_B$〕のもとで

$p_A{}^* = k_1 / n_A$，$p_B{}^* = k_2 / n_B$及び共通な $p^* = \dfrac{n_A p_A^* + n_B p_B^*}{n_A + n_B}$ を計算

$$\text{統計量} Z = \frac{p_A^* - p_B^*}{\sqrt{p^*(1-p^*)\left(\dfrac{1}{n_A}+\dfrac{1}{n_B}\right)}} \sim \mathrm{N}\left(0,1^2\right)$$

$$\mathrm{P}\left(|Z| > \lambda\right) = \alpha$$

λが求められる

Zの実現値z_0

$|z_0| \geqq \lambda$：H_0を棄却する

$|z_0| < \lambda$: H_0を採択する

◆例5.11　年賀状350通について,宛名の縦書きと横書きのものを男女別に調べたところ表5.4のとおりであった.性別により書き方に相違があるか.有意水準5%で検定せよ

解　男女に差がないという仮説

$p_A = p_B$　〔対立仮説H_1: $p_A \neq p_B$〕とする

$p_A^* = 179/282 = 0.635$　$p_B^* = 53/63 = 0.779$

A, Bともに共通な比率　$p^* = (179+53)/(282+68) = 0.633$

$$Zの実現値 z_0 = \frac{-0.144}{0.06386} = -2.26$$

有意水準$\alpha = 0.05$とすると, $\lambda = 1.96$

$$\therefore |z_0| = 2.26 > 1.96$$

仮説は有意水準5%で棄却される.すなわち男女別で差が見られる.

5.8　検定−適合度の検定

5.8.1　単純仮説のとき

実験などで得られた結果が，ある型の母集団からの観測値とみなせるかどうかを調べたいときに用いられる

母集団

k個の事象E_1, ･･･ E_k
発生確率p_1, ･･･ p_k

n個抽出し事象E_1, ･･･ E_kに分類しその
発生度数をf_1, ･･･ f_k（f_1, +･･ ･ + f_k ＝n）

仮説H_0: E_i起こる確率P（E_i）＝ p_i　表5.5を作成する

表 5.5

事　象	E_1	E_2	$\cdots$	E_k	計
度　数　f_i	f_1	f_2	$\cdots$	f_k	n
確　率　p_i	p_1	p_2	$\cdots$	p_k	1
期待度数 np_i	np_1	np_2	$\cdots$	np_k	n
$\frac{(f_i-np_i)^2}{np_i}$	$\frac{(f_1-np_1)^2}{np_1}$	$\frac{(f_2-np_2)^2}{np_2}$	$\cdots$	$\frac{(f_k-np_k)^2}{np_k}$	$\sum\frac{(f_i-np_i)^2}{np_i}=\chi^2$

$$\text{統計量}\chi^2 = \sum_{i=1}^{k}(\text{度数 - 期待度数})^2 = \sum_{i=1}^{k}\frac{(f_i - np_i)^2}{np_i} \sim \chi_{k-1}^2$$

有意水準α→$\chi_{k-1}{}^2(\alpha)$

χ^2の実現値χ_0^2

$\chi_0^2 \geqq \chi_{k-1}^2(\alpha)$: 仮説は棄却する

$\chi_0^2 < \chi_{k-1}^2(\alpha)$: 仮説は採択する

◆例5.12　植物の交配によってA, B2つの遺伝子が現われる割合は，非A，非Bを $\overline{A}, \overline{B}$ で表せば $AB : A\overline{B} : \overline{A}B : \overline{AB} = 9:3:3:1$ となることが期待されていたが，実験の結果は表5.6のとおりであった．この理論は当てはまるといえるか．有意水準5%で検定せよ

表 5.6

現れ方	AB	$A\overline{B}$	$\overline{A}B$	$\overline{AB}$	計
出現度数	1260	625	610	5	2500

解　仮説　実験の結果は期待する理論分布と一致する．すなわち H_0:

$$AB : A\overline{B} : \overline{A}B : \overline{AB} = \frac{9}{16} : \frac{3}{16} : \frac{3}{16} : \frac{1}{16}$$

表 5.7

事象	AB	$A\overline{B}$	$\overline{A}B$	$\overline{AB}$	計
度数 f_i	1260	625	610	5	2500
確率 p_i	9/16	3/16	3/16	1/16	1
期待度数 np_i	1406	469	469	156	2500
$\frac{(f_i - np_i)^2}{np_i}$	15.16	51.89	42.39	146.16	$\chi^2 = 255.60$

統計量 $\chi_0^2 = 255.60$

(自由度 $n = 4 - 1 = 3$)

$\chi_3^2(0.05) = 7.815$

$\therefore\ \chi_0^2 > 7.815$

仮説は棄却される

5.8.2　複合仮説のとき

母集団分布がm個の未知の母数(たとえば，正規母集団ではμ，σ^2の2個)を含んでいる場合は，標本から得られる推定値を用いて理論度数(期待度数)求める．

$$\text{検定統計量}\chi^2 = \sum_{i=1}^{k} \frac{(\text{度数} - \text{期待度数})^2}{\text{期待度数}} \sim \chi^2_{k-m-1}$$

（注）度数f_iまたは期待度数np_iが5より小さいものがあれば隣の級と合併して，5以上の大きい度数の級ばかりにする．このときは合併した級の度数だけ自由度は減少する．

◆例5.13　ある野菜の種を1列につき10粒ずつまいたもの80列について，一定の日数後に発芽数を調べて表5.8の結果を得た．この分布は2項分布に従っているとみなせるか．

表 5.8

発芽数	0	1	2	3	4	5	6	7	8	9	10	計
出現度数	6	20	28	12	8	6	0	0	0	0	0	80

解　仮説　H_0：2項分布に従うものとする．
標本平均を求めると $\bar{x}=2.2$ （$=np$：2項分布の平均値）
$\therefore 10p=2.2$,　$p=0.22$,　$1-p=0.78$
このとき2項分布$\mathrm{P}(X=k)={}_{10}C_k\cdot 0.22\cdot 0.78^{10-k}$ を利用して
$\mathrm{P}(X=0)$，$\mathrm{P}(X=1)$，・・・・，$\mathrm{P}(X=6)$を求め，度数表を作成し，期待度数，統計量を計算する

表 5.9

発芽数 k	出現度数 f	確　率 p	期待度数 np	$\frac{(f-np)^2}{np}$
0	6	0.0834	6.7	0.07313
1	20	0.2351	18.8	0.07660
2	28	0.2984	23.9	0.70335
3	12	0.2244	18.0	2.0
4	8	0.1108	8.9	0.09101
5	6	0.0375	3.0	3.0
計	80			$5.94408 = \chi_0{}^2$

統計量$\chi_0^2 = 5.94408$, 自由度$6-1-1=4$

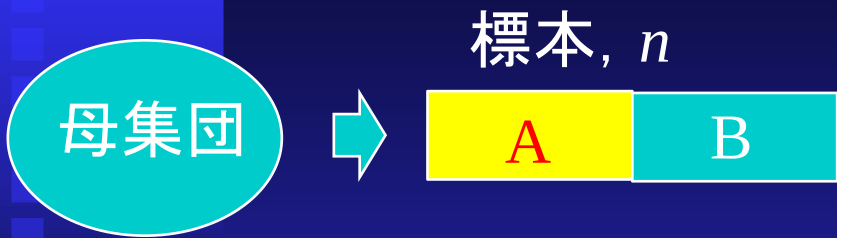
有意水準5%として$\chi_4^2(0.05) = 9.488$

$\therefore \chi_0^2 = 5.944 < 9.488$ ：仮説は棄却されない

この分布は2項分布に従うとみなせる

5.9 検定−独立性の検定　$l \times m$分割表

5.9.1 $l \times m$分割表



仮説H_0: 2つの属性A, Bが独立である. すなわち
$P(A_i \cap B_j) = P(A_i) \cdot P(B_j)$
$= (f_i/n) \cdot (f_j/n)$

表 5.10 f_{ij} 表

A \ B	B_1	$\cdots$	B_j	$\cdots$	B_m	計
A_1	f_{11}	$\cdots$	f_{1j}	$\cdots$	f_{1m}	$f_{1.}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	f_{i1}	$\cdots$	f_{ij}	$\cdots$	f_{im}	$f_{i.}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_l	f_{l1}	$\cdots$	f_{lj}	$\cdots$	f_{lm}	$f_{l.}$
計	$f_{.1}$	$\cdots$	$f_{.j}$	$\cdots$	$f_{.m}$	n

$$\left(\sum_{i=1}^{l} f_{ij} = f_{.j},\quad \sum_{j=1}^{m} f_{ij} = f_{i.},\quad \sum_{i=1}^{l} f_{i.} = \sum_{j=1}^{m} f_{.j} = n \right)$$

実現値f_{ij}に対する期待値

$$nP(A_i \cap B_j) = n \times \frac{f_{i.}}{n} \times \frac{f_{j.}}{n} = \frac{f_{i.} f_{j.}}{n}$$

$$\text{統計量}\chi^2 = \sum_{i=1}^{l} \sum_{j=1}^{m} \left\{ \left(f_{ij} - \frac{f_{i.} f_{j.}}{n} \right) \Big/ \frac{f_{i.} f_{j.}}{n} \right\} \sim \chi^2_{(l-1)(m-1)}$$

◆例5.14　ある製薬会社で3種の薬品A, B, Cについて，その効力を調べるために，それぞれ50人を選び試したところ，表5.11の結果を得た．薬品の間に違いがあるといえるか．有意水準5%で検定せよ．

表5.11

	A	B	C	計
効力あり	31	35	29	95
効力なし	19	15	21	55
計	50	50	50	150

仮説H_0：各薬品と効力とは関係ない

統計量の実現値$\chi_0^2 = 1.609$

表5.12　期待度数

	A	B	C	計
効力あり	31.7	31.7	31.7	95
効力jなし	18.3	18.3	18.3	55
計	50	50	50	150

有意水準5%として，自由度

(3-1)×(2-1)＝2から

$\chi_2^2(0.05) = 5.991$

$\therefore \chi_0^2 = 1.609 < 5.991$

仮説は棄却されない

薬品間に効力の差があるとはいえない

5.9.2　2×2分割表

表5.10にから，表5.13を得る.これより

$$\text{統計量}\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi_1^2$$

表5.13

A＼B	B_1	B_2	計
A_1	a	b	$a+c$
A_2	c	d	$b+d$
計	$a+c$	$b+d$	n

◆例5.15　表5.14は，ビタミンB不足が胎児の性決定に影響があるか否かを，ねずみについて試験した結果である.ビタミンBと胎児の雌雄に関係があるといえるか．有意水準5%で検定せよ.

表5.13

ビタミン＼性	雌	雄	計
ビタミン不足	123	153	276
ビタミン十分	145	150	295
計	268	303	571

仮説H_0:**ビタミンBと胎児の雌雄とは独立である**

統計量の実現値$\chi_0^2 = 1.200$

有意水準5%として，自由度1から

$\chi_1^2(0.05) = 3.841$

$\therefore \chi_0^2 = 1.200 < 3.841$

仮説は棄却されない