教養科目／**B** 自然の摂理の探求

2026 年度 統計学（集中１期）

第１４回

4.5. 適合度検定

表4.4 血液型調査の結果

A型	B型	O型	AB型	合計
30	18	27	5	80

日本人全体の血液型

A型	B型	O型	AB型
38%	22%	30%	10%

表 4.4 の血液型のデータが日本人全体の血液型の比に合っている(適合している)かどうかを検定する.

設定

・$A_1, A_2, \ldots, A_k$: k 個の事象

・データは $A_1, A_2, \ldots, A_k$ のうちのどれか1つだけに分類される.

・$p_i = \Pr(A_i) \quad (i=1,2,\ldots,k) \qquad p_1 + p_2 + \ldots + p_k = 1$

このとき,

帰無仮説 $H_0 : p_i = p_{i0} \ (i=1,2,\ldots,k)$ 対立仮説 $H_1 : H_0$ の否定

を検定する問題を考える.

・p_{i0} (i=1,2,…,k) はある決まった値とする.

・H_0 の否定は「$p_i \neq p_{i0}$ となる i がある」となる.

血液型の例では

帰無仮説 H_0 : p_1=0.38, p_2=0.22, p_3=0.30, p_4=0.10

対立仮説 H_1 : H_0 の否定

ただし，p_1, p_2, p_3, p_4 は A型，B型，O型，AB型の確率

データの個数を n とし，
それぞれの事象に属する度数(**観測度数**という)を

$O_1, O_2, \ldots, O_k$

とおく. このとき，$O_1 + O_2 + \ldots + O_k = n$ が成り立つ.

・検定統計量

$$\chi^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} + \dots + \frac{(O_k-E_k)^2}{E_k}$$

ここで,

$$E_i = n \times p_{i0} \qquad (i=1,2,\dots,k)$$

は**期待度数**と呼ばれ, H_0 が正しいときに期待される各事象に含まれる度数である.

・帰無仮説のときの χ^2 の近似分布

データの個数 n が大きいとき，H_0 が正しいと仮定すると

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \cdots + \frac{(O_k - E_k)^2}{E_k}$$

は近似的に自由度 k-1 の χ^2 分布に従うことが知られている.

・対立仮説のときの χ^2 の取る値の傾向

大きい値を取る傾向がある

・棄却域（有意水準を 0.05 ）

$$W = [\chi^2_{[k-1]}(0.05), \infty)$$

・検定における近似の条件

近似の条件は

$$E_i \geq 5 \quad (i=1,2,\ldots,k)$$

が成り立つことである.

・注意

近似の条件 $E_i \geq 5$ が満たされていない，すなわち，

ある j について $E_j < 5$ である

ときは，それを含むいくつかの事象を1つにまとめて，近似の条件が満たされるようにする.

公式 4.11

帰無仮説 H_0: $p_i = p_{i0} \ (i = 1, 2, \ldots, k)$ 対立仮説 H_1: H_0 の否定

を有意水準 0.05 で検定する.

検定統計量 χ^2 および棄却域 W は

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \cdots + \frac{(O_k - E_k)^2}{E_k},$$

$$W = [\chi^2_{[k-1]}(0.05), \infty).$$

・検定方式

χ^2 の実現値が W に含まれる

⇩

H_0 を棄却

(H_1 であると判断する.)

χ^2 の実現値が W に含まれない

⇩

H_0 を棄却しない

例 4.14. 血液型の例について，実際に検定してみましょう.

帰無仮説 H_0：$p_1=0.38,\ p_2=0.22,\ p_3=0.30,\ p_4=0.10$
対立仮説 H_1：H_0 の否定

・データの個数：n=80
⓪ 近似の条件のチェック.

$$E_i \geq 5 \quad (i=1,2,3,4).$$

・期待度数(E_i)：$80 \times 0.38 = 30.4 \geq 5,\ 80 \times 0.22 = 17.6 \geq 5$
$80 \times 0.30 = 24.0 \geq 5,\ 80 \times 0.10 = 8.0 \geq 5$

・期待度数と観測度数（表 4.5）：

	A型	B型	O型	AB型
観測度数	30	18	27	5
期待度数	30.4	17.6	24.0	8.0

① データから χ^2 の実現値を計算する.

$$\chi^2 = \frac{(30-30.4)^2}{30.4} + \frac{(18-17.6)^2}{17.6} + \frac{(27-24.0)^2}{24.0} + \frac{(5-8.0)^2}{8.0} \fallingdotseq 1.51$$

② 事象の個数 k より, 棄却域を求める.

$$W = [\chi^2_{[4-1]}(0.05), \infty) = [7.815, \infty)$$

③ 結論を出す.

χ^2 の実現値 1.51 は棄却域に含まれないことから,
H_0 は棄却されない.

つまり, 血液型の人数の比が

A : B : O : AB = 38 : 22 : 30 : 10

といえないこともないと判断される.

例 4.15. 手作りのサイコロを 50 回投げて，次の結果を得ました.

出た目	1	2	3	4	5	6	合計
観測度数	4	13	3	11	7	12	50

このサイコロは正常なもの（どの目が出やすいとか出にくいということはないという意味）といえるか，有意水準 0.05 で検定してみましょう.

各目の出る確率は同じ 1/6 かどうか

H_0 : 各目の出る確率は 1/6　　H_1 : H_0 の否定

の適合度検定を行う.

・データの個数：n=50　　・事象の個数：k=6

⓪ 近似の条件のチェック.

$$E_i \geq 5 \quad (i=1,2,\ldots,6).$$

・期待度数 (E_i)：

$50 \times 1/6 = 25/3 = 8.33\cdots \geq 5$, $50 \times 1/6 = 25/3 = 8.33\cdots \geq 5$

$50 \times 1/6 = 25/3 = 8.33\cdots \geq 5$, $50 \times 1/6 = 25/3 = 8.33\cdots \geq 5$

$50 \times 1/6 = 25/3 = 8.33\cdots \geq 5$, $50 \times 1/6 = 25/3 = 8.33\cdots \geq 5$

・期待度数と観測度数（表 4.6）：

出た目	1	2	3	4	5	6	合計
観測度数	4	13	3	11	7	12	50
期待度数	$\frac{25}{3}$	$\frac{25}{3}$	$\frac{25}{3}$	$\frac{25}{3}$	$\frac{25}{3}$	$\frac{25}{3}$	50

① データから χ^2 の実現値を計算する.

$$\chi^2 = \frac{(4-\frac{25}{3})^2}{\frac{25}{3}} + \frac{(13-\frac{25}{3})^2}{\frac{25}{3}} + \frac{(3-\frac{25}{3})^2}{\frac{25}{3}} + \frac{(11-\frac{25}{3})^2}{\frac{25}{3}} + \frac{(7-\frac{25}{3})^2}{\frac{25}{3}} + \frac{(12-\frac{25}{3})^2}{\frac{25}{3}} = 10.96$$

② 事象の個数 k より, 棄却域を定める.

$$W = [\chi^2_{[6-1]}(0.05), \infty) = [11.070, \infty)$$

③ 結論を出す.

χ^2 の実現値 10.96 は棄却域に含まれないことから,
H_0 は棄却されない.

つまり, このサイコロは正常なものといえないこともない
と判断される.

例 4.16. ある大学の 1 年生全体に数学の試験（100 点満点）を行い，試験を受けた人の中からランダムに 100 人選んだ結果，次のような表を得ました.

事象	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	合計
観測度数	9	6	10	16	19	17	12	11	100

A_1 : 20点未満， A_2 : 20点以上30点未満，
A_3 : 30点以上40点未満， A_4 : 40点以上50点未満，
A_5 : 50点以上60点未満， A_6 : 60点以上70点未満，
A_7 : 70点以上80点未満， A_8 : 80点以上

母集団分布は正規分布であるかどうか，検定してみましょう.

過去の試験の結果などから，母集団分布の母平均 50, 母分散 20^2 がわかっているとします．n=100，k=8 として，

H_0 : 母集団分布は正規分布 $N(50,20^2)$ である
H_1 : H_0 の否定

の適合度検定を行う.

帰無仮説 H_0 が正しいと仮定する.

つまり, 試験の点数 X は正規分布 $N(50,20^2)$ に従うと仮定する.

このとき, $\frac{X-50}{20}$ は標準正規分布 N(0,1) に従うので,

$$\Pr(X < 20) = \Pr\left(\frac{X-50}{20} < \frac{20-50}{20}\right)$$

$$= \Pr\left(\frac{X-50}{20} < -1.5\right) = 0.0668 = \Pr(80 \leq X)$$

$$\Pr(X < 20) = 0.0668 = \Pr(80 \leq X)$$

同様に,

$$\Pr(20 \leq X < 30) = 0.0919 = \Pr(70 \leq X < 80)$$

$$\Pr(30 \leq X < 40) = 0.1498 = \Pr(60 \leq X < 70)$$

$$\Pr(40 \leq X < 50) = 0.1915 = \Pr(50 \leq X < 60)$$

事象	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
観測度数	9	6	10	16	19	17	12	11
期待度数	6.68	9.19	14.98	19.15	19.15	14.98	9.19	6.68

⓪ 近似の条件のチェック.

表より、$E_i \geq 5$ (i=1,2,…,8).

① データから χ^2 の実現値を計算する.

$$\chi^2 = \frac{(9-6.68)^2}{6.68} + \frac{(6-9.19)^2}{9.19} + \frac{(10-14.98)^2}{14.98} + \cdots + \frac{(11-6.68)^2}{6.68} \fallingdotseq 7.95$$

② 事象の個数 k より, 棄却域を求める.

$$W = [\chi^2_{[8-1]}(0.05), \infty) = [14.067, \infty)$$

③ 結論を出す.

χ^2 の実現値 7.95 は棄却域に含まれないことから,
H_0 は棄却されない.

この試験の得点分布が正規分布 $N(50,20^2)$ であるといえないこともないと判断される.

注意. 例 4.16 では, 母数(母平均 μ と母分散 σ^2)の値がわかっている場合を扱った. 実際には, 母平均や母分散はわからない場合がほとんどである.

母数の値がわからない場合は, その推定値で代用する.

母数に推定値を用いた場合は, 近似に用いられる χ^2 分布の自由度は推定した母数の数だけ下がる.

したがって, 例 4.16 で母平均と母分散にそれぞれ標本平均, 不偏分散を用いた場合には, 自由度は 8-1-2=5 になる.

4.6. 独立性の検定

ある食堂で値段と味についてアンケートをとりました.
以下の表はその結果です.

値段＼味	美味しい	普通	まずい	計
安い	7	10	7	24
普通	15	23	7	45
高い	5	11	15	31
計	27	44	29	100

このとき，値段と味の間には関係があると判断されるでしょうか.

２つの特性 A と B の関係を考える.

n 回の独立な実験を行い，以下の表が得られたとする.

$A \diagdown B$	B_1	B_2	$\cdots$	B_s	計	
A_1	n_{11}	n_{12}	$\cdots$	n_{1s}	$n_{1\cdot}$	$\sum_{j=1}^{s} n_{1j}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2s}	$n_{2\cdot}$	$\sum_{j=1}^{s} n_{2j}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	
A_r	n_{r1}	n_{r2}	$\cdots$	n_{rs}	$n_{r\cdot}$	$\sum_{j=1}^{s} n_{rj}$
計	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot s}$	n	
	$\sum_{i=1}^{r} n_{i1}$	$\sum_{i=1}^{r} n_{i2}$		$\sum_{i=1}^{r} n_{is}$		

n_{ij} は, 事象 $A_i \cap B_j$ に含まれる観測度数.

このとき，知りたいのは，2つの特性 A と B に関係があるかどうか，すなわち，A と B は独立でないかどうかである．
検定したい仮説として，

帰無仮説 H_0: A と B は独立である
対立仮説 H_1: A と B は独立でない

を考える．

$p_{ij} = \Pr(A_i \cap B_j)$ とし，$p_{i\cdot} = \Pr(A_i)$, $\quad p_{\cdot j} = \Pr(B_j)$ とすると，

$$\text{帰無仮説 } \mathrm{H}_0 : p_{ij} = p_{i\cdot} p_{\cdot j} \ (i = 1, 2, \cdots, r; j = 1, 2, \cdots, s),$$

$$\text{対立仮説 } \mathrm{H}_1 : \mathrm{H}_0 \text{ の否定}$$

と書くことができる．

・データの個数 n が大きいとき，帰無仮説 H_0 のもとで

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(n_{ij}-e_{ij})^2}{e_{ij}}$$

は近似的に自由度 $(r-1)(s-1)$ の χ^2 分布に従うことが知られている.

・ただし， $e_{ij} = n\cdot\frac{n_{i\cdot}}{n}\cdot\frac{n_{\cdot j}}{n} = \frac{n_{i\cdot}n_{\cdot j}}{n}$ 独立期待度数.

・検定における近似の条件は,

$$e_{ij} \geq 5 \qquad (i=1,2,\cdots,r; j=1,2,\cdots,s)$$

が成り立つこと.

・一方，対立仮説 H_1 のもとで χ^2 は大きくなる傾向があることが知られています．

・このことから，有意水準 0.05 の棄却域は

$$[\chi^2_{[(r-1)(s-1)]}(0.05), \infty).$$

・注意 4.3

特に $r = s = 2$ のとき，χ^2 および棄却域 W は

$$\chi^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}, \qquad W = [3.841, \infty).$$

1.9 節で出てきた基準値

公式 4.12

・仮説

帰無仮説 H_0: A と B は独立である
対立仮説 H_1: A と B は独立でない

・検定統計量と棄却域

有意水準 0.05 で検定する場合，検定統計量 χ^2 および棄却域 W は

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(n_{ij}-e_{ij})^2}{e_{ij}}, \qquad W = [\chi^2_{[(r-1)(s-1)]}(0.05), \infty)$$

・検定方式

χ^2 の実現値が W に含まれる
⇩
H_0 を棄却
(H_1 であると判断する．)

χ^2 の実現値が W に含まれない
⇩
H_0 を棄却しない

例 4.17. 食堂の値段と味について実際に検定してみましょう.

値段＼味	美味しい (B_1)	普通 (B_2)	まずい (B_3)	計
安い (A_1)	7	10	7	24
普通 (A_2)	15	23	7	45
高い (A_3)	5	11	15	31
計	27	44	29	100

仮説

帰無仮説 H_0: A と B は独立である
対立仮説 H_1: A と B は独立でない

を,有意水準 0.05 で検定する.

・近似を用いてよいかを確認する.

値段＼味	美味しい (B_1)	普通 (B_2)	まずい (B_3)	計
安い (A_1)	7	10	7	24
普通 (A_2)	15	23	7	45
高い (A_3)	5	11	15	31
計	27	44	29	100

$$e_{11} = \frac{27 \times 24}{100} = 6.48, \quad e_{12} = \frac{44 \times 24}{100} = 10.56, \quad e_{13} = \frac{29 \times 24}{100} = 6.96,$$

$$e_{21} = \frac{27 \times 45}{100} = 12.15, \quad e_{22} = \frac{44 \times 45}{100} = 19.8, \quad e_{23} = \frac{29 \times 45}{100} = 13.05,$$

$$e_{31} = \frac{27 \times 31}{100} = 8.37, \quad e_{32} = \frac{44 \times 31}{100} = 13.64, \quad e_{33} = \frac{29 \times 31}{100} = 8.99$$

すべて, 5 以上なので, 近似を用いてよいかの条件をクリア.

・検定統計量 χ^2 の実現値を計算する.

値段＼味	美味しい (B_1)	普通 (B_2)	まずい (B_3)
安い (A_1)	7	10	7
普通 (A_2)	15	23	7
高い (A_3)	5	11	15

値段＼味	美味しい (B_1)	普通 (B_2)	まずい (B_3)
安い (A_1)	6.48	10.56	6.96
普通 (A_2)	12.15	19.8	13.05
高い (A_3)	8.37	13.64	8.99

⇨ $\dfrac{(n_{ij}-e_{ij})^2}{e_{ij}}$ を計算.

i, j に関して和をとる.

$$\chi^2 = \frac{(7-6.48)^2}{6.48} + \frac{(10-10.56)^2}{10.56} + \cdots + \frac{(15-8.99)^2}{8.99} \fallingdotseq 9.9$$

・棄却域を計算する.

自由度は $(3-1)\times(3-1)=4$ となるので，有意水準 0.05 の棄却域 W は χ^2 分布表（数表 4）より

$$W=[9.488,\infty)$$

となる.

・棄却域に χ^2 の実現値が含まれるかを確かめる.

χ^2 の実現値 9.9 は棄却域 W に含まれる. (**棄却**)

⇩

つまり，値段と味の間には関係があると判断されます.

例 4.18. あるサッカーチームのホームとアウェイの勝敗表である.

	勝	負	計
ホーム	9	3	12
アウェイ	4	4	8
計	13	7	20

ホームのほうがアウェイより勝ちやすいかどうか, すなわち,

帰無仮説 H_0 :ホーム・アウェイと勝敗とは関係がない,
対立仮説 H_1 :ホームのほうがアウェイより勝ちやすい

を有意水準 0.05 で検定してみましょう.

データの個数 nが小さく，近似の条件を満たしていない場合-----フィッシャーの直接確率法

・対応する4つの度数を比較し，独立期待度数より大きい（小さい）観測度数に注目する．

観測度数

	勝	負	計
ホーム	9	3	12
アウェイ	4	4	8
計	13	7	20

独立期待度数

	勝	負	計
ホーム	7.8	4.2	12
アウェイ	5.2	2.8	8
計	13	7	20

ここでは，赤い丸のセル（観測度数が 9，独立期待度数が 7.8）に注目する．

・赤い丸のセルの観測度数よりも大きく(小さく)なる観測度数を求め,以下の表における観測度数のパターンを求める.

	勝	負	計
ホーム	n_{11}	n_{12}	12
アウェイ	n_{21}	n_{22}	8
計	13	7	20

赤い丸のセルの観測度数 9 よりも大きく(小さく)なる観測度数:
10, 11, 12

$$\begin{array}{|cc|}\hline n_{11} & n_{12} \\ n_{21} & n_{22} \\ \hline\end{array} \Rightarrow \underset{\text{(a)}}{\begin{array}{|cc|}\hline 9 & 3 \\ 4 & 4 \\ \hline\end{array}} \quad \underset{\text{(b)}}{\begin{array}{|cc|}\hline 10 & 2 \\ 3 & 5 \\ \hline\end{array}} \quad \underset{\text{(c)}}{\begin{array}{|cc|}\hline 11 & 1 \\ 2 & 6 \\ \hline\end{array}} \quad \underset{\text{(d)}}{\begin{array}{|cc|}\hline 12 & 0 \\ 1 & 7 \\ \hline\end{array}}$$

$$\begin{array}{|cc|}\hline n_{11} & n_{12} \\ n_{21} & n_{22} \\ \hline\end{array} \Rightarrow \underset{(a)}{\begin{array}{|cc|}\hline 9 & 3 \\ 4 & 4 \\ \hline\end{array}} \quad \underset{(b)}{\begin{array}{|cc|}\hline 10 & 2 \\ 3 & 5 \\ \hline\end{array}} \quad \underset{(c)}{\begin{array}{|cc|}\hline 11 & 1 \\ 2 & 6 \\ \hline\end{array}} \quad \underset{(d)}{\begin{array}{|cc|}\hline 12 & 0 \\ 1 & 7 \\ \hline\end{array}}$$

(b), (c), (d) は (a) よりホームのほうが勝ちやすい状況であると考えられる.

・仮説の下で, (a), (b), (c), (d) の起こる確率を考える.

$$P = \frac{12!8!13!7!}{20!}\left(\frac{1}{9!3!4!4!} + \frac{1}{10!2!3!5!} + \frac{1}{11!1!2!6!} + \frac{1}{12!0!1!7!}\right) \fallingdotseq 0.251$$

・検定方式

P が有意水準 0.05 より小さい → 棄却

この例の場合であれば, P は 0.05 より大きいため棄却できない.

・検定方式

P が有意水準 0.05 より小さい → 棄却

この例の場合であれば，P は 0.05 より大きいため棄却できない．

つまり，このデータからは
「ホーム・アウェイと勝敗とは関係がないといえないこともない」
と判断されます．

・χ^2 値との関係

	(a)	(b)	(c)	(d)
χ^2 の値	1.32	4.43	9.38	16.15

まとめ

χ^2 適合度検定

観測度数，期待度数

独立性の検定

公式 4.12

データの個数が小さい場合は，

フィッシャーの直接確率法