統計学 2018 春　　　　　　　　　独立性と適合度検定　　　　　　　　Prof.Hwang

1. 独立性検定の例題

a) 仮説設定

H0：二つの変数は**独立である。**
H1：二つの変数は**独立ではない**（何らかの関連がある。）

ある病気 A に関して、感染経験について調査したところ、男女別に以下のようなデータが得られた。病気 A の感染経験と、性別の間に関連性は有ると言えるか？有意水準α=0.05 として、独立性のカイ二乗検定によって、結論づけろ。

	病気 A に感染 infected with disease A	病気 A に感染せず not infected with disease A	合計
男性 male	76	42	118
女性 female	57	82	139
合計 total	133	124	257

帰無仮説 H0：性別と病気 A の感染経験は独立である（関連性はない）

対立仮説 H1：性別と病気 A の感染経験は独立ではない

b) 期待度数計算

検定に用いる期待度数　$E_{ij} = \frac{n_{i.}n_{.j}}{N}$

期待度数（Expected Frequency）

	病気 A に感染	病気 A に感染していない
男性	61.1 $(=\frac{133*118}{257})$	56.9 $(=\frac{124*118}{257})$
女性	71.9 $(=\frac{133*139}{257})$	67.1 $(=\frac{133*139}{257})$

c）検定統計量

検定統計量は $\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - E_{ij})^2}{E_{ij}} = \chi^2 \ ((r-1)(c-1))$

と書くことができ、カイ二乗分布を用いた検定。

検定統計値の計算：

$$\chi^2 = + \frac{(76-61.1)^2}{257} + \frac{(42-56.9)^2}{257} + \frac{(57-71.9)^2}{257} + \frac{(82-67.1)^2}{257} = 13.1 \sim \chi^2(1)$$

d）P-value の計算

この検定統計値による P-value は、0.0003 であり、5%より小さいので帰無仮説を棄却する。結論は、病気 A の感染経験と性別との間には何らかの関係があることである。

自由度 1 の場合、上側 5%のカイ二乗分布の臨界値(critical value)は、7.879 であり、検定統計値の 13.1 は、それを超えるので帰無仮説は棄却できる。有意な結果である。

２．適合度検定の例題

適合度の検定

（１）各グループに属するデータの数の比率が理論通りかどうかを検定する

（２）観察されたデータが、特定の分布に一致しているといえるか？ 分布の種類を問わない

（３）２つの母集団の確率分布が異なるものであるかどうか？ （ノンパラメトリック ー コルモゴロフ・スミルノフ検定)

１のケースだけここでは考える。あるエンドウマメの交配実験の結果は，メンデルの法則によれば「黄色・丸」「黄色・しわ」「緑色・丸」「緑色・しわ」の４種類の形質のマメが 9 : 3 : 3 : 1 の割合で現れるはずだという。実際にこの実験を行ったところ，それぞれの形質をもったマメの数は 447, 131, 152, 38 であった。この実験においてメンデルの法則が成り立っているかどうかを有意水準 5% で検定せよ。

しかし，実際には，得られたマメは標本でしかありません。たとえ，世界全体で，各形質のマメが理論 通りに 9 ： 3 ： 3 ： 1 の比率で実っているとしても，この実験で得られたマメは，偶然のために，9 ： 3 ： 3 ： 1 にはならないかもしれません。

そこで，実際に観測されたデータをカテゴリに分類したときの比率と，今仮定している理論（モデル） から予想される比率とのずれが，偶然起こりうる程度のものか，それともモデルの方が間違っていると 考えたほうがいいのかを検定するのが，適合度検定です。ここでは，よく用いられる χ 2（カイ２乗）適 合度検定を紹介します。

a)仮説設定

H_0: ４つのグループに属する個体数の比率は$\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$, $\frac{1}{16}$である。

H_1: 「４つのグループに属する個体数の比率は$\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$, $\frac{1}{16}$ではない。

b) 理論からの期待度数の計算：

期待確率（理論確率）から期待度数の計算。各属性の期待度数 E_i はその属性の期待確率(理論確率) を P_i を利用して、

$$E_i = n_i \times P_i$$

この問題で，もしも各形質をもつマメの個数が，正確に理論通り(理論確率・期待確率)に 9 ： 3 ： 3 ： 1 の比率になっていると すれば，得られたマメの総数は 768 ですから，それぞれの個数は 432，144，144，48 になるはずです。

c) カイ二乗値の計算：$\chi^2=\sum_{i=1}^{r}\frac{(n_i-E_i)^2}{E_i}=\chi^2\ (r-1)$

d) 帰無仮説の元での検定統計量：

検定統計値は$\chi^2(3)$= $(447-432)^2/432+(131-144)^2/144+(152-144)^2/144+(38-48)^2/48=4.22$

４つのグループに分かれるので自由度は 3 で，有意水準が 5%のとき，χ^2 ＞ χ^2 (0.05(k=3)) のと き，H_0 は棄却されます。数表より，χ^2 (0.05(3)) = 7.8147 ですから，$\chi^2(3)$=4.22 ＜χ^2 (0.05(3)) = 7.8147 なので H_0 は棄却さ れません。

これまでに説明した検定の考え方では，帰無仮説が「モデルは実測データに適合している」となって いますから，帰無仮説が棄却されたときは「モデルは実測データに適合していないと言える」と結論され，したがって，今回の例の結論は「メンデルの法則が成り立っていないとまでは言えない」 ということになるはずです。