

Rで学ぶ確率統計

第8回 統計的仮説検定②

1

Rで学ぶ確率統計

第8回

統計的検定②

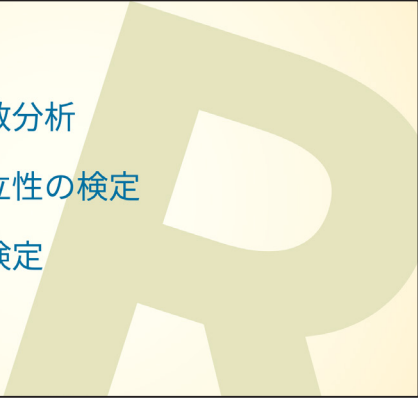
講師 秋光 淳生



2

第8回 統計的検定②

- 1 元配置分散分析
- 2 適合度と独立性の検定
- 3 相関係数の検定



3

①-1 3標本以上の比較

3標本以上の問題

3つのクラスA、B、Cに同じ国語のテストを行い、その結果を比較したい。

A	58	59	61	64	67	67	67	67	68	68	70	71	72	73		
B	54	56	56	57	58	59	59	60	61	62	63	66	68	70	70	
C	52	53	56	59	60	60	61	62	64	65	65	66	68	70	75	76

このとき、クラスによる違いがあると考えられることはできるか。
 テストの点数は正規分布に従い、クラスごとの母分散は等しいとする。
 帰無仮説は「A、B、Cで平均点は等しい ($H_0: \mu_A = \mu_B = \mu_C$)」
 それぞれの得点を
 $X_{Ai} (i = 1, 2, \dots, n_A)$, $X_{Bj} (j = 1, 2, \dots, n_B)$, $X_{Ck} (k = 1, 2, \dots, n_C)$
 と表すとする。

4

①-2 分散分析(1)

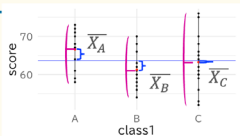
平方和の分解

全体の標本平均を $\bar{X}$ 、各グループの標本平均を $\bar{X}_A, \bar{X}_B, \bar{X}_C$ とすると

$$\sum_{i=1}^{n_A} (X_{Ai} - \bar{X})^2 + \sum_{j=1}^{n_B} (X_{Bj} - \bar{X})^2 + \sum_{k=1}^{n_C} (X_{Ck} - \bar{X})^2$$

$$= \sum_{i=1}^{n_A} (X_{Ai} - \bar{X}_A)^2 + \sum_{j=1}^{n_B} (X_{Bj} - \bar{X}_B)^2 + \sum_{k=1}^{n_C} (X_{Ck} - \bar{X}_C)^2 + n_A(\bar{X}_A - \bar{X})^2 + n_B(\bar{X}_B - \bar{X})^2 + n_C(\bar{X}_C - \bar{X})^2$$

(総平方和) = (群内平方和) + (群間平方和)



Rで学ぶ確率統計

第8回 統計的仮説検定②

5

①-2 分散分析(2)

平方和の分解

自由度: $n_A - 1 + n_B - 1 + n_C - 1 = n - a$ の χ^2 分布

$$\frac{\sum_{i=1}^{n_A} (X_{Ai} - \bar{X}_A)^2}{\sigma^2} + \frac{\sum_{j=1}^{n_B} (X_{Bj} - \bar{X}_B)^2}{\sigma^2} + \frac{\sum_{k=1}^{n_C} (X_{Ck} - \bar{X}_C)^2}{\sigma^2} + \frac{n_A(\bar{X}_A - \bar{X})^2}{\sigma^2} + \frac{n_B(\bar{X}_B - \bar{X})^2}{\sigma^2} + \frac{n_C(\bar{X}_C - \bar{X})^2}{\sigma^2}$$

自由度: $a - 1$ の χ^2 分布

から

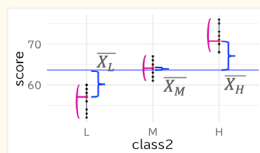
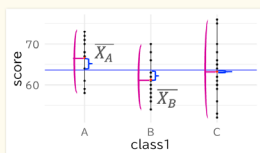
$$F = \frac{(S_A/\sigma^2)/(a-1)}{(S_E/\sigma^2)/(n-a)} = \frac{S_A/(a-1)}{S_E/(n-a)} \sim F(a-1, n-a)$$

6

①-2 分散分析(3)

平方和の分解

3つのクラスA、B、Cを点数順に並べ替えて別のクラスL、M、Hにする



変わらない

(総平方和) = (群内平方和) + (群間平方和)

クラスによって 差がなければ群間平方和の影響が小さい

7

①-2 分散分析(4)

平方和の分解

自由度: $n_A - 1 + n_B - 1 + n_C - 1 = n - a$ の χ^2 分布

$$\frac{\sum_{i=1}^{n_A} (X_{Ai} - \bar{X}_A)^2}{\sigma^2} + \frac{\sum_{j=1}^{n_B} (X_{Bj} - \bar{X}_B)^2}{\sigma^2} + \frac{\sum_{k=1}^{n_C} (X_{Ck} - \bar{X}_C)^2}{\sigma^2} + \frac{n_A(\bar{X}_A - \bar{X})^2}{\sigma^2} + \frac{n_B(\bar{X}_B - \bar{X})^2}{\sigma^2} + \frac{n_C(\bar{X}_C - \bar{X})^2}{\sigma^2}$$

自由度: $a - 1$ の χ^2 分布

から

$$F = \frac{(S_A/\sigma^2)/(a-1)}{(S_E/\sigma^2)/(n-a)} = \frac{S_A/(a-1)}{S_E/(n-a)} \sim F(a-1, n-a)$$

8

①-3 Rによるシミュレーション(1)

```
x <- c(58,59,61,64,67,67,67,68,68,70,71,72,73)
y <- c(54,56,56,57,58,59,59,60,61,62,63,66,68,70,70)
z <- c(52,53,56,59,60,60,61,62,64,65,65,66,68,70,75,76)
na <- length(x); nb <- length(y); nc <- length(z)
h1 <- tibble(score=c(x,y,z),class1=c(rep("A",na),rep("B",nb),rep("C",nc)))
h1 <- h1 %>% mutate(class2 =
  ifelse(score<68,ifelse(score<61,"L","M"),"H")) %>%
  mutate(class2=factor(class2,levels=c("L","M","H")))
h3 <- aov(score~class1,data=h1);summary(h3)
h4 <- aov(score~class2,data=h1) summary(h4) # aov は分散が等しい場合
oneway.test(score~class1,data=h1,var.equal=TRUE) # oneway.test は
oneway.test(score~class1,data=h1) # 分散が等しくない場合も
```

Rで学ぶ確率統計

第8回 統計的仮説検定②

9

②-1 適合度検定(1)

適合度検定

	A_1	A_2	...	A_k	計
観測度数	O_1	O_2		O_k	n
期待度数	E_1	E_2		E_k	n

において

$$\sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

は近似的に χ^2 分布 $\chi^2(k-1)$ に従う。
検定は片側検定で行う。

 X_i が正規分布に従うとき

$$W = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2}$$

とすると

$$W \sim \chi^2(n-1)$$

10

②-1 適合度検定(2)

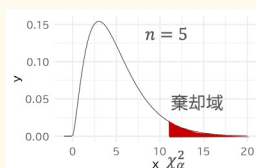
適合度検定

	A_1	A_2	...	A_k	計
観測度数	O_1	O_2		O_k	n
期待度数	E_1	E_2		E_k	n

において

$$\sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

は近似的に χ^2 分布 $\chi^2(k-1)$ に従う。
検定は片側検定で行う。



11

②-2 独立性検定(1)

独立性検定 項目AとCが独立 (関連性がない)

観測度数	A_1	A_2	...	A_k	計
C_1	O_{11}	O_{12}	...	O_{1k}	n_1
C_2	O_{21}	O_{22}	...	O_{2k}	n_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_r	O_{r1}	O_{r2}	...	O_{rk}	n_r
計	m_1	m_2	...	m_k	n
割合	$p_1 = m_1/n$	$p_2 = m_2/n$	...	$p_k = m_k/n$	

$$E_{ij} = n \times \frac{n_i}{n} \times \frac{m_j}{n}$$

12

②-2 独立性検定(2)

独立性検定

観測度数	A_1	A_2	...	A_k	計
C_1	O_{11}	O_{12}	...	O_{1k}	n_1
C_2	O_{21}	O_{22}	...	O_{2k}	n_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_r	O_{r1}	O_{r2}	...	O_{rk}	n_r
計	m_1	m_2	...	m_k	n

において

$$\sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

は近似的に χ^2 分布 $\chi^2((r-1) \times (k-1))$ に従う。片側検定で行う。

Rで学ぶ確率統計

第8回 統計的仮説検定②

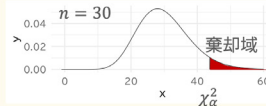
13

②-2 独立性検定(3)

独立性検定

観測度数	A_1	A_2	...	A_k	計
C_1	O_{11}	O_{12}	...	O_{1k}	n_1
C_2	O_{21}	O_{22}	...	O_{2k}	n_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_r	O_{r1}	O_{r2}	...	O_{rk}	n_r
計	m_1	m_2	...	m_k	n

$$\sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2((r-1) \times (k-1))$$



14

②-2 独立性検定(4)

独立性検定 - $n=2$ のとき

	A_1	A_2	計
観測度数	O_{11}	O_{12}	$O_{11} + O_{12}$
期待度数	O_{21}	O_{22}	$O_{21} + O_{22}$
期待値	$p = (O_{11} + O_{21})/n$	$p = (O_{12} + O_{22})/n$	$O_{11} + O_{12} + O_{21} + O_{22}$

自由度1の χ^2 分布

$$O_{ij} - E_{ij} = \frac{O_{11}O_{22} - O_{12}O_{21}}{O_{11} + O_{12} + O_{21} + O_{22}}$$

分子は
どれも等しい

イエーツの補正

$$\sum_{i=1}^r \sum_{j=1}^k \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}} \text{ を用いる}$$

15

②-3 Rによるシミュレーション

```
> p1 <- c(0.18, 0.19, 0.2, 0.21, 0.22)
> p2 <- c(0.2, 0.2, 0.2, 0.2, 0.2)
> n <- 100
> i1 <- rmultinom(1, n, p1)
> chisq.test(i1, p=p2)
```

```
# chisq.test(x,p)
#x:ベクトル
#p=c(p1,p2,...)
#適合度検定
```

```
> b01 <- read_csv("A02L01.csv")
> b02 <- table(b01$study, b01$result)
> summary(b02)
> chisq.test(b02)
> chisq.test(b02, correct=FALSE)
```

```
# chisq.test(x)
#x:ベクトル
#pがなければ独立性検定
```

16

③-1 相関係数の検定(1)

相関性の検定

2種類の確率変数 X_i ($i = 1, 2, \dots, n$), Y_j ($j = 1, 2, \dots, n$) に対し X_i, Y_j が正規分布にしたがうとすると

$$r = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

によって計算される標本相関係数に対し

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

は自由度 $df = n - 2$ の t 分布に従う

Rで学ぶ確率統計

第8回 統計的仮説検定②

17

③-1 相関係数の検定(2)

相関性の検定

また母相関係数を ρ とするとき $\zeta = \frac{1}{2} \log \frac{1+\rho}{1-\rho}$ とする

$$z = \frac{1}{2} \log \frac{1+r}{1-r}$$

は近似的に平均 ζ 、分散 $\frac{1}{n-3}$ の正規分布に従う

ζ の信頼区間は $\left[z - \frac{z_L}{2} \times \frac{1}{\sqrt{n-3}}, z + \frac{z_U}{2} \times \frac{1}{\sqrt{n-3}} \right]$

ρ の信頼区間は $\left[\frac{e^{2z_L} - 1}{e^{2z_L} + 1}, \frac{e^{2z_U} - 1}{e^{2z_U} + 1} \right]$

18

③-2 相関の検定

```
n <- 5; alpha <- 0.05
x <- c(1.275, 0.592, -1.714, -1.888, 1.641)
y <- c(-0.112, -0.212, 0.575, 0.088, 0.890)
r <- cor(x, y)
t <- r * sqrt(n-2) / sqrt(1-r^2)
z <- 1/2 * log((1+r)/(1-r))
zL <- z - qnorm(1-alpha/2, 0, 1) / sqrt(n-3)
zU <- z + qnorm(1-alpha/2, 0, 1) / sqrt(n-3)
rhoL <- (exp(2*zL)-1) / (exp(2*zL)+1)
rhoU <- (exp(2*zU)-1) / (exp(2*zU)+1)
cor.test(x, y)
```

```
# cor.test(x,y)
# xとyは同じ長さのベクトル
# 相関係数を計算する

# r <- cor(x,y)
# rhoL, rhoU の計算は
# 結果を確認するため
```