Сем 13

тест Пирсона + свойства рангов

Пример. таблица сопр-сти: $n=100$

	латте	кап	эспр	Σ
M	10	10	10	30
F	20	30	20	70
Σ	30	40	30	100

- наблюд. незав-мы
- набл. один. распр.

H_0: пол и кофе незав И $p_{\cdot 1} = p_{\cdot 2}$

H_1: не H_0

а) протестир-ть H_0 на ур-не значимости $\alpha = 0.01$

б) p-value

$$T = \sum_{ij} \frac{(N_{ij} - \hat{N}_{ij})^2}{\hat{N}_{ij}} = \sum_{ij} \frac{N_{ij}^2}{\hat{N}_{ij}} - n$$

$\hat{N}_{ij}$?

$$\hat{N}_{11} = E(N_{11} \mid H_0, N_{1\cdot}, N_{2\cdot}, N_{\cdot 1}, N_{\cdot 2}, N_{\cdot 3}) \ ?$$

$T \xrightarrow[n\to\infty]{dist} \ ? \ \chi^2_d$ $\quad d = df_{UR} - df_R$

UR-модель: необр. модель (H_0 или H_1)

p_{11}	p_{12}	p_{13}
p_{21}	p_{22}	p_{23}

$df_{UR} = 6 - 1 = 5$ (своб. пар)

$p_{11} + p_{12} + \ldots + p_{23} = 1$ ур-ие

R-модель: огр. модель

df_R

	β_1	β_2	β_3
α_1			
α_2			

$\beta_1 = \beta_2$

$\alpha_1 + \alpha_2 = 1$

$\beta_1 + \beta_2 + \beta_3 = 1$

уравнения

$$df_R = \underset{\alpha_1, \alpha_2}{2} + \underset{\beta_1, \beta_2, \beta_3}{3} - 3 = 2 \quad (\text{своб-х пар-ра})$$

вер-сти клеток при H_0

	β	β	$1-2\beta$
α	$\alpha\beta$	$\alpha\beta$	$\alpha\cdot(1-2\beta)$
$1-\alpha$	$(1-\alpha)\beta$	$(1-\alpha)\beta$	$(1-\alpha)\cdot(1-2\beta)$

$$T \to \chi^2_d \qquad d = df_{UR} - df_R = 5 - 2 = 3$$

при верной H_0:

$$\hat{\alpha} = \frac{N_{1\cdot}}{N}$$

$$2\hat{\beta} = \frac{N_{\cdot1} + N_{\cdot2}}{N} \qquad \hat{\beta} = \frac{N_{\cdot1} + N_{\cdot2}}{2N}$$

оценки вер-стей клеток при верной H_0.

	$\hat{\beta}$	$\hat{\beta}$	$1-2\hat{\beta}$
$\hat{\alpha}$	$\hat{\alpha}\hat{\beta}$	$\hat{\alpha}\hat{\beta}$	$\hat{\alpha}(1-2\hat{\beta})$
$1-\hat{\alpha}$	$(1-\hat{\alpha})\hat{\beta}$	$(1-\hat{\alpha})\hat{\beta}$	$(1-\hat{\alpha})(1-2\hat{\beta})$

$$\frac{21}{2} = \hat{N}_{12} = \hat{N}_{11} = N\cdot\hat{\alpha}\cdot\hat{\beta} = N\cdot\frac{N_{1\cdot}}{N}\cdot\frac{N_{\cdot1}+N_{\cdot2}}{2N} = \frac{1}{2}\,\frac{N_{1\cdot}\cdot(N_{\cdot1}+N_{\cdot2})}{N}$$

$$\hat{N}_{21} = \hat{N}_{22} = \ldots = \frac{1}{2} N_{2\cdot}\cdot\frac{(N_{\cdot1}+N_{\cdot2})}{N} = \frac{70\cdot70}{100\cdot2} = \frac{49}{2}$$

$$\hat{N}_{13} = \frac{N_{1\cdot}\cdot N_{\cdot3}}{N} = 9 \qquad \hat{N}_{23} = \frac{N_{2\cdot}\cdot N_{\cdot3}}{N} = 21$$

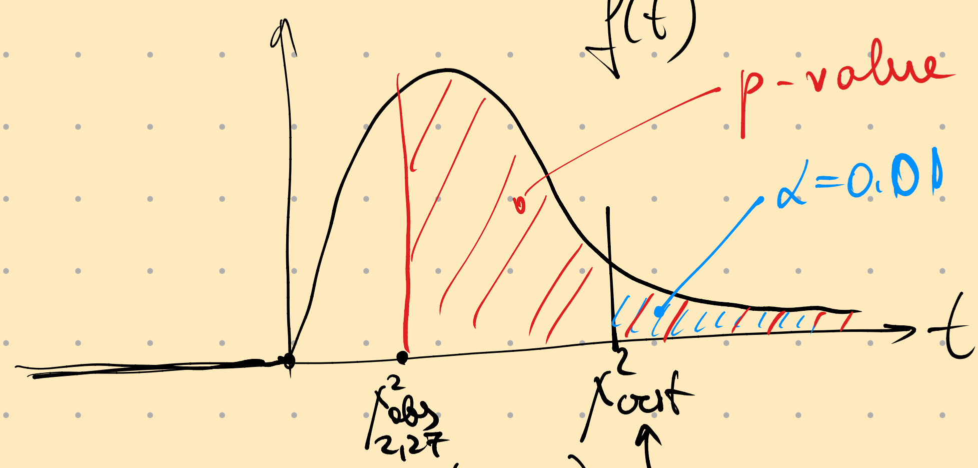
$\hat{}$

N_{ij}

10.5	10.5	9
24.5	24.5	21

$$T = \sum_{ij} \frac{N_{ij}^2}{\hat{N}_{ij}} - n = \ldots = 2.27$$

при H_0: $\quad T \xrightarrow[\text{dist}]{n\to\infty} \chi^2_3$



$\chi^2_{crit} = 11.34$ (табл/компь)

$\chi^2_{obs} = 2.27$

Способ 1:

Если $\chi^2_{obs} > \chi^2_{crit}$, то H_0 отвергается (иначе — не отвергается)

Вывод: H_0 не отвергается
(данные не противоречат H_0)

Способ 2: через p-значение

$$p\text{-value} = P(T_{new} > T \mid H_0, T) =$$

$$= P(\chi^2_3 > 2.27) = 1 - P(\chi^2_3 \le 2.27) =$$

$$= 1 - F_{\chi^2_3}(2.27) =$$

$$= 0.52$$

0.52 $\quad$ 0.01

Если p-value < α, то H_0 отвергается (иначе — не отвергается)

H_0 не отвергается

Упр.

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \longrightarrow \begin{pmatrix} R_1^x \\ R_2^x \\ \vdots \\ R_n^x \end{pmatrix}$$

$X_i \sim$ незав, один. распр. $P(X_i = X_j) = 0$ при $i \neq j$

ранг — номер в числе по возрастанию.

!! а) $E(R_1^x) \stackrel{?}{=}$

б) $Var(R_1^x) \stackrel{?}{=}$

!! в) $\overline{R} = ?$

!! г) $E(\overline{R}) = ?$ $\quad Var\,\overline{R} = ?$

в)

$$\overline{R} = \frac{R_1 + R_2 + \ldots + R_n}{n} = \frac{2+1+5+\ldots+6}{n} = \frac{n+1}{2}$$

все номера от 1 до n

!!

г) $\overline{R}$ — это вырожд. СВ

это конст.

$$E(\overline{R}) = \overline{R} = \frac{n+1}{2} \qquad Var(\overline{R}) = 0$$

а)

R_1 — СВ

значен.	1	2	3	. . . .	n
вер-сти	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	. . . .	$\frac{1}{n}$

$$E(R_1) = \frac{n+1}{2}$$

б) путь крестьянина

$$Var(R_1) = E(R_1^2) - \Big(\underbrace{E(R_1)}_{=\frac{n+1}{2}}\Big)^2$$

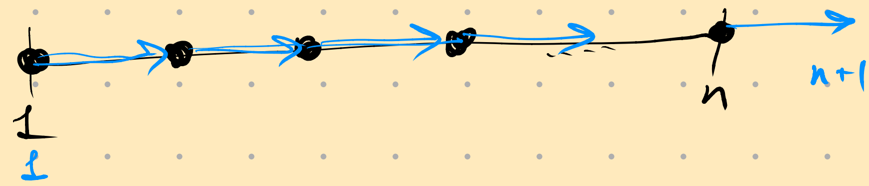
$$E(R_1^2) = \sum_{k=1}^{n} k^2 \cdot P(R_1 = k) = \sum_{k=1}^{n} k^2 \cdot \frac{1}{n} =$$

$$= \frac{1}{n} \cdot \frac{n\cdot(n+1)\cdot(2n+1)}{6}$$

путь мога

$U \sim Unif[0;1]$ незав от R_1

$$R_1 + U = S \sim Unif[1; n+1]$$



$E(U) = \frac{1}{2}$ $\quad Var(U) = \frac{1}{12}$

$Var\, R_1 + Var\, U = Var(S)$

$Var\, R_1 + \frac{1}{12} = \frac{n^2}{12}$

$Var\, R_1 = \frac{n^2 - 1}{12}$

$S \sim 1 + nU$

$Var(S) = n^2 \cdot Var\, U = \frac{n^2}{12}$

Задача. $X_1, X_2, \ldots, X_{n_x}$ $\quad Y_1, \ldots, Y_{n_Y}$

предп:
- Все независимы.
- $X_i \sim Y_j \quad \forall i,j$
- $P(X_i = X_j) = 0$ при $i \neq j$

- расположим набл по возраст-ию.

$$X_5 < Y_3 < X_4 < Y_6 \ldots\ldots < X_{17}$$

№1 №2 №3 №4 ...... №$(n_x + n_Y)$ $\quad n = n_x + n_Y$

- присвоим ранги

S_x - сумма рангов у X_i

S_Y - сумма рангов у Y_i

а) $E(S_x)$ $\quad E(S_Y)$?

б) $Var(S_x)$ $\quad Var(S_Y)$ $\quad Cov(S_x, S_Y)$?

а) !! $E(S_X)$? $= E(R_1^X + R_2^X + R_3^X + \dots + R_{n_X}^X) =$

гоб !

$$= \underbrace{E(R_1^X)} + E(R_2^X) + \dots + E(R_{n_X}^X) =$$

$$= \frac{n+1}{2} + \frac{n+1}{2} + \dots + \frac{n+1}{2} = n_X \cdot \frac{n+1}{2}$$

$$n = n_X + n_Y$$

$$E(S_Y) = (\text{по аналогии}) = n_Y \cdot \frac{n+1}{2}$$

б) $R_1^X + \dots + R_{n_X}^X + R_1^Y + \dots + R_{n_Y}^Y = 1 + 2 + \dots + n =$

$$= \frac{1+n}{2} \cdot n$$

пример

X: 2.7, 3.5, 4.9

Y: 3.1, 6.2

$$\underbrace{2.7}_{X_1}, \underbrace{3.1}_{Y_1}, \underbrace{3.5}_{X_2}, \underbrace{4.9}_{X_3}, \underbrace{6.2}_{Y_2}$$

$R_1^X = 1 + R_1^Y = 2 + R_2^X = 3 + R_3^X = 4 + R_2^Y = 5$ $= 1+2+3+4+5 = 15$

$$S_X + S_Y = n \cdot \frac{1+n}{2}$$

$$S_Y = \text{const} - S_X$$

$$Var(S_Y) = Var(\text{const} - S_X) = Var(S_X)$$

$$Cov(S_X, S_Y) = Cov(S_X, \cancel{\text{const}} - S_X) =$$

$$= -Cov(S_X, S_X) = -Var(S_X)$$

$Var(S_X)$?

- путь крестьяна

$$Var(S_x) = \underbrace{E(S_x^2)} - (E(S_x))^2$$

знаем

$$E(S_x^2) = E\left((R_1^x + \ldots + R_{n_x}^x)^2\right) = \text{раскрыть скобки} = \ldots$$

= наборотьсч с суммами

- хитрость. вар-т 1 как можно дальше работать с дисперсией

$$Var(S_x) = Var(R_1^x + R_2^x + \ldots + R_{n_x}^x) =$$

$$= Var(R_1^x) + Var(R_2^x) + \ldots + Var(R_{n_x}^x) +$$

$$+ \sum_{i>j} 2\,Cov(R_i^x, R_j^x) \;;$$

$\boxed{\text{идея } 1}$ → наобл X-иксов, один. распр.

$$R_1^x \sim R_2^x \sim R_3^x \ldots$$

[ранги зависимы, но один. распр]

$$Var(R_1^x) = Var(R_2^x) = \ldots$$

$$Var(R_1^x) = \frac{n^2-1}{12}$$

→ пары рангов

$$(R_1^x, R_2^x) \sim (R_2^x, R_7^x)$$

один. распр.

$$P(R_1^x = 7, R_2^x = 8) = P(R_2^x = 7, R_7^x = 8)$$

$$cov(R_1^x, R_2^x) = c \quad (?)$$

$$Var(S_x) = n_x \cdot \frac{n^2-1}{12} + n_x \cdot (n_x - 1) \cdot c \quad (?)$$

$\boxed{\text{идея } 2}$ $R_1^x \sim R_2^y$

$$\mathrm{Var}(S_x + S_y) \to 0 \qquad S_x + S_y = 1+2+\ldots+n = \text{const}$$

$$n \cdot \mathrm{Var}(R_1^x) + n\cdot(n-1)\cdot c = 0$$

$$c = -\frac{\mathrm{Var}(R_1^x)}{n-1} = -\frac{n^2-1}{12(n-1)} = -\frac{n+1}{12}$$

error: $\mathrm{Var}(S_x) = n_x \cdot \frac{n^2-1}{12} + n_x\cdot(n_x-1)\cdot\left(-\frac{n+1}{12}\right)$!!

$$\mathrm{Var}(S_x) = \frac{n_x}{12}\cdot(n+1)\cdot\left[n-1-(n_x-1)\right]$$

$$\mathrm{Var}(S_x) = \frac{n_x}{12}\cdot(n+1)\cdot n_y$$

!!