# Таблицы сопряженности

Грауэр Л.В.

# Таблицы сопряженности

$(O_1, \ldots, O_n)$

Качественные признаки:

$$A : A_1, A_2, ..., A_r$$

$$B : B_1, B_2, ..., B_s$$

| | $B_1$ | $B_2$ | $\ldots$ | $B_s$ | |
|---|---|---|---|---|---|
| $A_1$ | $n_{11}$ | $n_{12}$ | $\ldots$ | $n_{1s}$ | $m_1$ |
| $A_2$ | $n_{21}$ | $n_{22}$ | $\ldots$ | $n_{2s}$ | $m_2$ |
| $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | |
| $A_r$ | $n_{r1}$ | $n_{r2}$ | $\ldots$ | $n_{rs}$ | $m_r$ |
| | $n_1$ | $n_2$ | $\ldots$ | $n_s$ | $n$ |

# Независимость признаков

$p_i = P(A_i)$, $i = 1, \ldots, r$
$q_j = P(B_j)$, $j = 1, \ldots, s$

$A$ и $B$ называются *независимыми*, если при любых $i$ и $j$ :

$$p_{ij} = P(A_i \cap B_j) = p_i q_j.$$

$H_0$: $P(A_i \cap B_j) = p_i q_j$ для любых $i, j$.

$H_1$: $\exists (i,j) : P(A_i \cap B_j) \neq p_i q_j$.

# Статистика критерия

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

$$L = \frac{n!}{\prod\limits_{\substack{i=\overline{1,r} \\ j=\overline{1,s}}} n_{ij}!} \prod_{\substack{i=\overline{1,r} \\ j=\overline{1,s}}} (p_i q_j)^{n_{ij}} \qquad \sum_{i=1}^{r} p_i = 1 \text{ и } \sum_{j=1}^{s} q_j = 1$$

$$\hat{p}_i = \frac{m_i}{n}, \ i = 1, \ldots, r, \qquad \hat{q}_j = \frac{n_j}{n}, \ j = 1, \ldots, s$$

# Критическая область

Если $H_0$ верна,

$$\chi^2(X_{[n]}) = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(n_{ij} - \frac{m_i n_j}{n})^2}{\frac{m_i n_j}{n}} \xrightarrow{d} \zeta \sim \chi^2((s-1)(r-1)).$$

$$\alpha \in (0,1) \quad , \; V_k = \{\chi^2(X_{[n]}) > \chi^2_{1-\alpha}((r-1)(s-1))\}$$

В случае таблицы 2x2 используется поправка Йейтса:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(|n_{ij} - \frac{m_i n_j}{n}| - 0.5)^2}{\frac{m_i n_j}{n}}.$$

# Таблицы сопряженности 2x2

$(O_1, \ldots, O_n)$

$A : A_1, A_2,\ B : B_1, B_2$

| | $B_1$ | $B_2$ | сумма |
|---|---|---|---|
| $A_1$ | $n_{11}$ | $n_{12}$ | $m_1$ |
| $A_2$ | $n_{21}$ | $n_{22}$ | $m_2$ |
| сумма | $n_1$ | $n_2$ | $n$ |

Зависимы ли признаки $A$ и $B$?

# Точный критерий Фишера

$T_i$ при зафикисированных значениях $m_1$, $m_2$, $n_1$, $n_2$:

| | $B_1$ | $B_2$ | сумма |
|---|---|---|---|
| $A_1$ | $n^i_{11}$ | $n^i_{12}$ | $m_1$ |
| $A_2$ | $n^i_{21}$ | $n^i_{22}$ | $m_2$ |
| сумма | $n_1$ | $n_2$ | $n$ |

Вероятность получить $T_i$

$$P_i =$$

# Статистика критерия Фишера

Статистика двустороннего критерия

$$P = \sum_{T_i : P_i \leq P_0, i \neq 0} P_i + P_0.$$

$$P < \alpha$$

# Пример

| № улыбки | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| exp1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| exp2 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |

| | 1 | 0 | сумма |
|---|---|---|---|
| exp1 | 6 | 4 | 10 |
| exp 2 | 6 | 4 | 10 |
| сумма | 12 | 8 | 20 |

| exp1 / exp2 | 0 | 1 | сумма |
|---|---|---|---|
| 0 | 2 | 2 | 4 |
| 1 | 2 | 4 | 6 |
| сумма | 4 | 6 | 10 |

| exp1 / exp2 | 0 | 1 | сумма |
|---|---|---|---|
| 0 | 0 | 4 | 4 |
| 1 | 4 | 2 | 6 |
| сумма | 4 | 6 | 10 |

| exp1 / exp2 | 0 | 1 | сумма |
|---|---|---|---|
| 0 | 1 | 3 | 4 |
| 1 | 3 | 3 | 6 |
| сумма | 4 | 6 | 10 |

| exp1 / exp2 | 0 | 1 | сумма |
|---|---|---|---|
| 0 | 3 | 1 | 4 |
| 1 | 1 | 5 | 6 |
| сумма | 4 | 6 | 10 |

| exp1 / exp2 | 0 | 1 | сумма |
|---|---|---|---|
| 0 | 4 | 0 | 4 |
| 1 | 0 | 6 | 6 |
| сумма | 4 | 6 | 10 |

$P = 1$