Казахский национальный университет имени аль-Фараби
Факультет медицины и здравоохранения
Кафедра политики и организации здравоохранения

Тема: «Анализ таблиц сопряженности. Критерий хи-квадрат Пирсона для проверки гипотезы о статистической связи между двумя категориальными признаками. Условия применимости (ожидаемые частоты). Точный тест Фишера как альтернатива для малых выборок и таблиц 2 х 2»

Лектор: к.м.н., доцент Калмаханов С.Б.

Кафедра политики и организации здравоохранения

Ключевые термины

Термин	Значение
Таблица сопряженности	Таблица для анализа двух категориальных признаков
χ^2 Пирсона	Статистический тест для проверки независимости
Ожидаемая частота	Количество случаев, если признаки независимы
Точный тест Фишера	Альтернатива χ^2 для малых выборок, точное p-value

Введение

В медицинском менеджменте часто необходимо анализировать взаимосвязь **двух категориальных признаков**. Например:

- Пол пациента (м/ж) и наличие хронического заболевания (да/нет)
- Тип медицинского учреждения (государственное/частное) и внедрение электронных медицинских записей (есть/нет)

Анализ таблиц сопряженности

Для этого используют **таблицы сопряженности (контингентные таблицы)** и **статистические тесты**:

- **χ^2 Пирсона (chi-square test)** – для проверки гипотезы о независимости двух категориальных признаков.
- **Точный тест Фишера (Fisher's exact test)** – альтернатива χ^2 для малых выборок или таблиц 2×2 с низкими ожидаемыми частотами.

Таблицы сопряженности

Таблица сопряженности – это таблица, в которой строки и столбцы представляют категории двух признаков.

	Заболевание есть	**Заболевания нет**	**Всего**
Мужчины	30	70	100
Женщины	20	80	100
Всего	50	150	200

Строки – категории одного признака (например, пол)
Столбцы – категории второго признака (например, наличие заболевания)
Ячейки – количество наблюдений в каждой комбинации

Критерий χ^2 Пирсона

Цель – проверить гипотезу:

- **H_0 (нулевая гипотеза):** признаки независимы
- **H_1 (альтернативная гипотеза):** признаки зависимы (связаны)

Формула

Для таблицы r×c:

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

где:

O_{ij} – наблюдаемое количество в ячейке (i,j)

E_{ij} – ожидаемое количество, если признаки независимы, рассчитывается как:

$$E_{ij} = \frac{(\textit{сумма по}i) \cdot (\textit{сумма по}j)}{n}$$

n – общий размер выборки

Пример использования критерий χ^2 Пирсона

$$E_{11} = \frac{(100)\cdot(50)}{200} = 25$$

$$E_{12} = \frac{100\cdot150}{200} = 75$$

$$E_{21} = \frac{100\cdot50}{200} = 25$$

$$E_{22} = \frac{100\cdot150}{200} = 75$$

$$\chi^2 = \frac{(30-25)^2}{25} + \frac{(70-75)^2}{75} + \frac{(20-25)^2}{25} + \frac{(80-75)^2}{75} = 6.67/?$$

(При расчетах шаг за шагом получаем $\chi^2 = 6.67$)

Сравниваем с критическим значением χ^2 с df=(r-1)(c-1)=1 и уровнем значимости α=0.05:

χ^2 critical = 3.841 → χ^2 calculated > χ^2 critical → отвергаем H_0, признаки связаны

Условия применимости χ^2

Для корректного применения χ^2 Пирсона:

1. Все наблюдаемые частоты $O_{ij} \geq 1$
2. Не более 20% ячеек с ожидаемыми частотами $Eij < 5$
3. Размер выборки достаточно большой (обычно n > 40-50)

Если эти условия не соблюдаются, результаты χ^2 могут быть некорректными.

Точный тест Фишера

Когда использовать

- Малые выборки
- Таблицы 2×2
- Ожидаемые частоты $E_{ij} < 5$

Идея теста

- Рассчитывается **точная вероятность наблюдаемого распределения** или более экстремального при условии независимости
- Не требует приближений, как χ^2

Пример использования точного теста Фишера

	Событие есть	События нет	Всего
Группа А	1	4	5
Группа В	4	1	5
Всего	**5**	**5**	**10**

χ^2 здесь применять нельзя ($Eij < 5$)
Точный тест Фишера даст корректный p-value

Практические рекомендации для менеджера здравоохранения

1. Использовать χ^2 для анализа взаимосвязи категориальных показателей при **достаточно больших выборках;**

2. Проверять условия применимости ($Eij \geq 5$);

3. При малых выборках использовать **точный тест Фишера;**

4. Отчет должен содержать:

- Таблицу сопряженности
- Расчет ожидаемых частот
- χ^2 или p-value
- Интерпретацию: «Есть/нет статистически значимая связь»

Заключение

1. Малые выборки или ячейки с <5 → **только точный тест Фишера;**

2. Большие таблицы и все $Eij \geq 5$ → χ^2 Пирсона.

В практическом анализе важно не только рассчитать статистику и p-value, но и правильно интерпретировать результат: **статистическая значимость ≠ практическая значимость.**

Спасибо за внимание!