Направление подготовки
31.05.02 Педиатрия (врач педиатр)

Учебный План утвержден решениями Ученого совета НГМУ
Протокол №3 от 17.04.2018 г.:

Учебная дисциплина
Б1.Б.12 МАТЕМАТИКА

КАФЕДРА МАТЕМАТИКИ
НГМУ

ЛЕКЦИОННОЕ ЗАНЯТИЕ

Тема:

Основные понятия теории статистических гипотез.

Проверка статистических гипотез.

Основы теории корреляции.

КАФЕДРА МАТЕМАТИКИ
НГМУ

Раздел 1. Теория вероятностей — 8 час

1	Тема-1.1 Введение в теорию вероятностей. Логические операции над множествами. Элементы комбинаторики. Вероятность события - определения, основные свойства и формулы вычисления.	2 час
2	Тема-1.2 Алгебра событий. Теоремы сложения и умножения вероятностей. Формула полной вероятности и формула Байеса. Формула Бернулли и формула Пуассона.	2 час
3	Тема-1.3 Случайные величины и их числовые характеристики. Основные законы распределения дискретных случайных величин.	2 час
4	Тема-1.4 Непрерывные случайные величины. Числовые характеристики НСВ. Основные законы распределения НСВ. Правило трех сигм.	2час

Раздел 2. Математическая статистика — 4 час

5	Тема-2.1 Основные понятия математической статистики. Статистические оценки параметров распределения.	2 час
6	Тема-2.2 Основные понятия теории статистических гипотез. Проверка статистических гипотез. Основы теории корреляции.	2 час

Тема - 2.2

План лекционного занятия

(лекционное занятие)

1	**Основные понятия теории статистических гипотез.** • Понятия - Статистическая гипотеза, основная и альтернативная. • Проверка статистической гипотезы. • Классификация ошибок при проверке гипотезы • Статистический критерий проверки гипотезы	20 мин
2	**Проверка статистических гипотез.**	35 мин
3	**Основы теории корреляции.**	35мин

Статистическая гипотеза – это любое предположение о виде неизвестного распределения или о параметрах известных распределений.

Статистическая гипотеза – это всякое высказывание о генеральной совокупности, проверяемое по выборке.

Гипотезы принято обозначать буквой Н с индексами.

Будем предполагать, что имеется 2 непересекающиеся гипотезы H_0 и H_1.

H_0 – нулевая гипотеза (или основная).

H_1 – альтернативная или конкурирующая гипотеза.

Проверка статистической гипотезы – это процедура сопоставления высказанного предположения (гипотезы) с выборочными данными.

Выдвинутая гипотеза может быть *правильной или неправильной,* поэтому возникает необходимость ее проверки.

Задача проверки статистических гипотез состоит в том, чтобы на основе выборки $x_1, x_2, x_3, \ldots, xn$ принять (т. е. считать справедливой) либо нулевую гипотезу , либо конкурирующую гипотезу .

Классификация ошибок при проверке правильности гипотезы

Ошибка первого рода состоит в том, что отвергается нулевая гипотеза H_0, когда на самом деле она верна.

Ошибка второго рода состоит в том, что отвергается альтернативная гипотеза H_1, когда на самом деле она верна.

Гипотеза H_0	Отвергается	Принимается
верна	ошибка 1-го рода	правильное решение
неверна	правильное решение	ошибка 2-го рода

Уровень значимости критерия – это вероятность ошибки 1-го рода

Статистический критерий проверки гипотезы – это правило, позволяющее, основываясь только на выборке $x_1, x_2, x_3, \ldots, x_n$ принять, либо отвергнуть нулевую гипотезу.

Виды статистических критериев

Параметрические	Непараметрические
представляют собой функции параметров данной совокупности и используются, если совокупности, из которых взяты выборки, подчиняются нормальному закону распределения.	применяются, если нет подчинения распределения нормальному закону.

Тема - 2.2

План лекционного занятия

(лекционное занятие)

1	**Основные понятия теории статистических гипотез.**	20 мин
2	**Проверка статистических гипотез.** • Общая постановка задачи • Проверка гипотез относительно средних • Проверка гипотез о законах распределения	35 мин
3	**Основы теории корреляции.**	35мин

2. Проверка статистических гипотез.

• Общая постановка задачи

Постановка задачи проверки гипотезы

1. Формулируют (выдвигают) нулевую гипотезу
2. Формулируют противоположную нулевой альтернативную гипотезу .
3. Задают уровень значимости ?
4. Выбирают критерии для проверки выдвинутой гипотезы.
5. По таблице определяют критическое значение критерия.
6. Сравнивают наблюдаемый и критический критерии.
7. Вывод о значимости или незначимости различий критериев.

2. Проверка статистических гипотез.
• Общая постановка задачи

1

Формулируют (выдвигают) нулевую гипотезу:

- ***об отсутствии различий между группами;***
- ***об отсутствии существенного отличия фактического распределения от некоторого заданного, например, нормального, экспоненциального и др;***

Сущность нулевой гипотезы:

- ***разница между сравниваемыми генеральными параметрами равна нулю;***
- ***различия, наблюдаемые между выборочными характеристиками, носят случайный характер, то есть эти выборки принадлежат одной генеральной совокупности.***

2

Формулируют противоположную нулевой альтернативную гипотезу.

3

Задают уровень значимости ?.

Уровень значимости - *это вероятность ошибки отвергнуть нулевую гипотезу , если на самом деле эта гипотеза верна.*

При ? ? 0,05 ошибка возможна в 5% случаев.

4

Выбирают критерии для проверки выдвинутой гипотезы.

Критерий – *это случайная величина К, которая служит для проверки H_0.*

Эти функции распределения известны и табулированы.

Критерий зависит от двух параметров:

- *от числа степеней свободы*
- *от уровня значимости.*

Фактическую величину критерия $K_{НАБЛ}$ получают по данным наблюдения.

5

По таблице определяют критическое значение критерия.

Превышение критического критерия $K_{КРИТ}(?, f)$ при справедливости гипотезы маловероятно.

6

Сравнивают *наблюдаемый $K_{НАБЛ}$ и критический $K_{КРИТ}(?, f)$ критерии.*

- *Если $K_{НАБЛ} > K_{КРИТ}(?, f)$ то отвергают H_0 и принимают H_1;*
- *Если $K_{НАБЛ} > K_{КРИТ}(?, f)$ то отвергают H_1 и принимают H_0;*

7

Вывод о значимости или незначимости различий критериев.

Вывод: Различие статистически значимо при ? > 0,05 или незначимо при ? ? 0,05 .

Проверка гипотез относительно средних

Предполагается, что неизвестные *генеральные дисперсии равны между собой*.

Две независимые выборки объемов n_1 и n_2, взятые из нормально распределенных совокупностей с параметрами $M(X_1)$ и $M(X_2)$, сравнивают друг с другом.

По этим выборкам находят соответствующие выборочные средние $\overline{x_1}$ и $\overline{x_2}$ и исправленные дисперсии S_1^2 и S_2^2.

Уровень значимости задан.

1. Нулевая гипотеза H_0: $M(X_1) = M(X_2)$;
2. Альтернативная гипотеза H_1: $M(X_1)$? $M(X_2)$;
3. ? ? 0,05;
4. Для проверки нулевой гипотезы в этом случае можно использовать критерий Стьюдента сравнения средних.

Величину критерия находим по формуле:

$$t_{НАБЛ} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}} \cdot \sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

Доказано, что величина $t_{НАБЛ}$ при справедливости нулевой гипотезы имеет *t – распределение* Стьюдента с $f = n_1 + n_2 - 2$ степенями свободы.

5. По таблице находим $t_{КРИТ}(?, f = n_1 + n_2 - 2)$

6. Сравниваем $t_{КРИТ}$ и $t_{НАБЛ}$

Если $\left|t_{НАБЛ}\right| < t_{КРИТ}(\alpha, f) \Rightarrow H_0$

Если $\left|t_{НАБЛ}\right| > t_{КРИТ}(\alpha, f) \Rightarrow H_1$

то различие достоверно.

Пример: По двум независимым малым выборкам объемов $n_1=5$ и $n_2=6$, извлеченным из нормальных генеральных совокупностей X_1 и X_2, вычислены выборочные средние: $\overline{x_1} = 3{,}3$ и $\overline{x_2} = 2{,}48$.

Известно, что генеральные дисперсии примерно равны, т. е. $D_{\text{ГЕН}_1} = D_{\text{ГЕН}_2}$

При уровне значимости ? ? 0,05 проверить нулевую гипотезу H_0: $M(X_1) = M(X_2)$, если $t_{НАБЛ} = 3{,}27$.

Решение:

$$t_{КРИТ}(\alpha \le 0{,}05, f = n_1 + n_2 - 2 = 5 + 6 - 2 = 9) = 2{,}26.$$

$$t_{НАБЛ} > t_{КРИТ}(\alpha, f) \Rightarrow отвергаем H_0$$

Вывод: выборочные средние различаются значимо ? ? 0,05

Проверка гипотез о законах распределения

Во многих практических задачах *закон распределения случайных величин заранее не известен, и надо выбрать модель, согласующуюся с результатами наблюдений.*

Выдвигают нулевую гипотезу: неизвестная функция распределения исследуемой случайной величины **X** распределена по некоторому теоретическому закону, например, по нормальному закону.

В качестве теоретической модели может быть рассмотрен любой закон, например, экспоненциальный или биномиальное распределение.

Выбор теоретической модели $F_{ТЕОР}(X)$ определяется сущностью изучаемого явления, а также результатами предварительной обработки наблюдений: формой графика распределения, соотношениями между выборочными данными.

Выдвигается *альтернативная гипотеза*, что данная генеральная совокупность не распределена по закону $F_{ТЕОР}(X)$

Задается уровень значимости, например, ? ? 0,05

Для проверки, *согласуются или нет эмпирические данные с гипотетическим предположением,* относительно теоретической функции распределения, используется *критерий согласия*.

2. Проверкастатистических гипотез.

• Проверкагипотез о законах распределения

Критерий согласия – это критерий проверки гипотезы о предполагаемом законе неизвестного распределения.

Рассмотрим один из них, использующий распределение $?^2$ и получивший название **критерий согласия Пирсона**.

Применим критерий $?^2$ к проверке нулевой гипотезы , что генеральная совокупность распределена нормально.

Критерий предполагает, что результаты наблюдений сгруппированы в вариационный ряд и разбиты на классы.

По выборке объема n построим эмпирическое распределение $F'_{эмп}(x)$:

варианты: $x_1, x_2, \ldots, x_k$

эмпирические частоты: $n_1, n_2, \ldots, n_k$

и сравним его с предполагаемым теоретическим распределением, вычисленным в предположении нормального закона распределения.

Теоретические частоты: $n'_1, n'_2, \ldots, n'_k$

То есть фактически: $H_0: \; n_{ЭМП} = n'_{ТЕОР}$

В качестве критерия проверки нулевой гипотезы примем случайную величину:

где k – число классов.

Из таблиц находим

Сравниваем, если

- расхождение теоретических и эмпирических частот незначимое. Следовательно, данные наблюдений согласуются с гипотезой о нормальном законе распределения генеральной совокупности.

2. Проверкастатистических гипотез.

• Проверкагипотез о законах распределения

Пример: При уровне значимости $? ? 0,05$ проверить гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические и теоретические частоты.

эмпирические частоты: *6 13 38 74 106 85 30 14*;

теоретические частоты: *3 14 42 82 99 76 37 13.*

Решение: $?^2_{набл} = 7,19$

Сравниваем: $?^2_{набл} < ?^2_{крит}(?, f)$? H_0

- расхождение теоретических и эмпирических частот незначимое.

Следовательно, данные наблюдений согласуются с гипотезой о нормальном законе распределения генеральной совокупности.

Тема - 2.2

План лекционного занятия

(лекционное занятие)

1	Основные понятия теории статистических гипотез.	20 мин
2	Проверка статистических гипотез.	35 мин
3	Основы теории корреляции. · Основные понятия теории корреляции · Коэффициент линейной корреляции и его свойства ·Проверка гипотезы о значимости выборочного коэффициента корреляции	35 мин

Корреляционный анализ – это статистический метод, изучающий связь между явлениями, если одно из них входит в число причин, определяющих другое или, если имеются общие причины, воздействующие на эти явления.

Основная задача – выявление связи между случайными величинами.

Функциональная зависимость – это зависимость вида

$$y = f(x)$$

когда каждому возможному значению случайной величины X соответствует одно возможное значение случайной величины Y.

Например, рост и масса. При одном и том же росте - масса различных индивидуумов может быть различна, но между средними значениями этих показателей имеется определенная зависимость.

Установление взаимосвязи между различными признаками и показателями функционирования организма позволяют по изменениям одних судить о состоянии других.

Схема эксперимента: имеется выборка объема n из генеральной совокупности N.

На каждом объекте выборки определяют числовые значения признаков, между которыми требуется установить наличие или отсутствие связи. Таким образом, получают два ряда числовых значений.

X	x_1	x_2	…	x_n
Y	y_1	y_2	…	y_n

Для наглядности, каждую пару можно представить в виде 2-х точек на координатной плоскости.

По оси абсцисс откладывают значения одного вариационного ряда x_i а по оси ординат другого y_i

Такое изображение статистической зависимости называется *полем корреляции* или *корреляционным полем* точек. Оно создает общую картину корреляции.

- *Если точки группируются вдоль некоторого направления, то это говорит о наличии линейной корреляционной связи между признаками.*
- *Если точки распределены равномерно, то линейная корреляционная связь отсутствует.*

• Поле корреляции

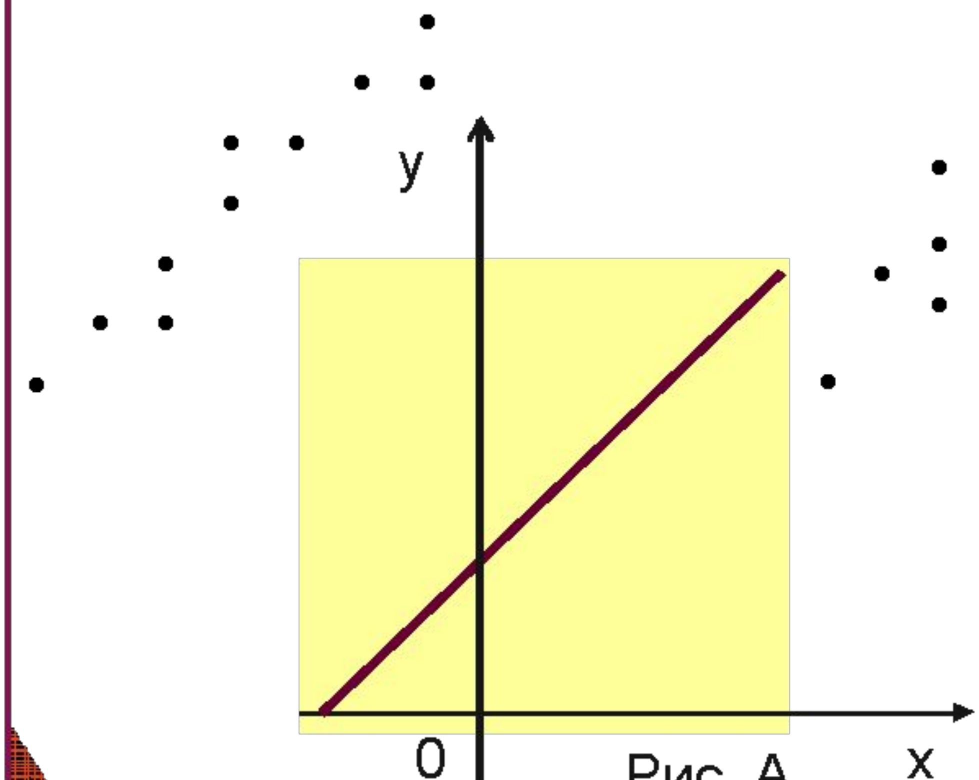

Рис. А

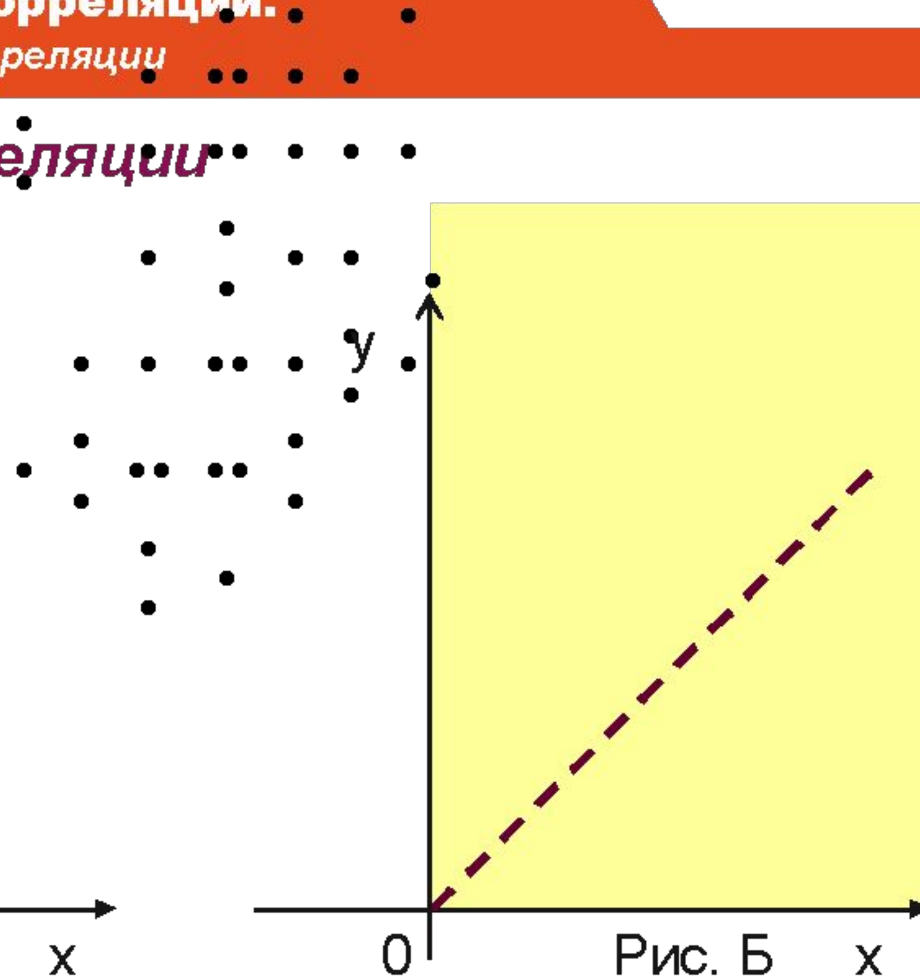

Рис. Б

3. Основы теории корреляции.

• Коэффициент линейной корреляции и его свойства

Корреляционный анализ – это метод, когда данные можно считать случайными и выбранными из совокупности, распределенной по нормальному закону.

Выборочный коэффициент линейной корреляции r – это характеристика тесноты связи между переменными функциональной зависимости, которую можно выразить одним числом. Он характеризует тесноту линейной связи между количественными признаками в выборке.

Свойства коэффициента линейной корреляции:

1. Значения коэффициента r заключены на отрезке [-1; 1].

Если значение коэффициента линейной корреляции $r > 0$, то корреляционная связь между переменными *прямая*.

Если значение коэффициента линейной корреляции $r < 0$, то корреляционная связь между переменными *обратная*.

2. Различают следующие *типы корреляционных связей*, в зависимости от того, насколько модуль r приближается к 1:

r < 0,3	*слабая*
0,3 ? r < 0,5	*умеренная*
0,5 ? r <0,7	*значительная*
0,7 ? r < 0,8	*достаточно тесная*
0,8 ? r < 0,9	*тесная (сильная)*
r ? 0,9	*очень тесная*

Свойства коэффициента линейной корреляции:

3. При $r = 1$ – полная линейная функциональная зависимость.

4. Чем ближе r к 0, тем слабее линейная корреляционная связь.

5. При $r = 0$ линейная корреляционная связь отсутствует.

6. *Значение величины коэффициента корреляции не изменится*, если все значения переменных увеличить (уменьшить) на одно и то же число или в одно и то же число раз.

Эмпирический (опытный) коэффициент корреляции, как и любой другой выборочный показатель, служит оценкой своего генерального параметра.

Выборочный коэффициент линейной корреляции $r_в$ - величина случайная, так как он вычисляется по значениям переменных, случайно попавших в выборку из генеральной совокупности, а значит, как и любая случайная величина имеет ошибку m_r

• Проверка гипотезы о значимости выборочного коэффициента корреляции

Чтобы выяснить, находятся ли случайные величины X и Y генеральной совокупности в линейно корреляционной зависимости, надо проверить значимость $r_в$.

Для этого *проверяют нулевую гипотезу* о равенстве нулю коэффициента корреляции генеральной совокупности $H_0: r_{ген} = 0$, т.е. гипотезу о том, что линейная корреляционная связь между признаками X и Y случайна.

Выдвигается альтернативная гипотеза $H_1: r_{ген}$? 0 т.е. о том, что линейная корреляционная связь не случайна.

Задается уровень значимости, например, ? ? $0,05$.

3. Основы теории корреляции.

• Проверка гипотезы о значимости выборочного коэффициента корреляции

Критерием для проверки нулевой гипотезы является отношение выборочного коэффициента корреляции к своей ошибке

где m_r - ошибка коэффициента корреляции.

Если объем выборки $n < 100$, то

Если объем выборки n ? 100, то

Число степеней свободы для проверки критерия равно $f = n-2$.

Гипотезу проверяют по таблицам распределения Стьюдента в соответствии с выбранным уровнем значимости.

По таблице критических точек распределения Стьюдента находим $t_{крит}$ (?, *f*), определенное на уровне значимости ? ? 0,05 при числе степеней свободы $f = n - 2$, где n – объем двумерной выборки.

Если $t_{набл}$? $t_{крит}$? H_1 отвергают нулевую гипотезу и принимают альтернативную H_1: $r_{ген}$? 0 т.е. о том, что имеется линейная корреляционная связь между признаками.

Если $t_{набл} < t_{крит}$ то нет оснований отвергать нулевую гипотезу, а $r_в$ - статистически незначим. Эта связь случайна.

Пример-1: Проверить значимость коэффициента корреляции $r = 0,74$ между переменными X и Y для выборки объема $n = 50$, при уровне значимости ? ? 0,05.

Проверяется нулевая гипотеза об отсутствии линейной корреляционной связи между переменными X и Y в генеральной совокупности H_0: $r_{ген} = 0$.

При справедливости этой гипотезы

где и

имеет распределение Стьюдента с $f = n - 2$ степенями свободы.

Находим значения:

Поскольку $t_{набл} > t_{крит}$ коэффициент корреляции значимо отличается от нуля, а значит корреляционная зависимость - не случайна.

Пример-2: По выборке объема $n = 122$, извлеченной из нормальной двумерной совокупности (X, Y) найден выборочный коэффициент линейной корреляции $r = 0{,}4$. При уровне значимости ? ? 0,05. проверить нулевую гипотезу, которая заключается в том, что связь между признаками случайна.

H_0: $r_{ген} = 0$.

Решение: При справедливости этой нулевой гипотезы H_1: $r_{ген}$? 0

имеет распределение Стьюдента с $f = n - 2$ степенями свободы.

Находим значения:

Поскольку $t_{набл} > t_{крит}$ то нулевая гипотеза отвергается и принимается альтернативная гипотеза $H_1: r_{ген}\ ?\ 0$

Вывод: между признаками имеется *умеренная линейная корреляционная связь $r = 0,4$.*

СПАСИБО ЗА ВНИМАНИЕ