КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

ИНСТИТУТ МАТЕМАТИКИ И МЕХАНИКИ

Кафедра математического анализа

И.Р. КАЮМОВ, А.Ю. ХАСАНОВА

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Конспект лекций

Казань – 2017

УДК 517.54
ББК 22.161.5

Принято на заседании кафедры математического анализа
Протокол № 2 от 11 октября 2017 года

Рецензент:
доктор физико-математических наук,
профессор кафедры математического анализа КФУ **А.Н. Чупрунов**

Каюмов И.Р., Хасанова А.Ю.
Теория вероятностей и математическая статистика / И.Р. Каюмов, А.Ю. Хасанова – Казань: Казан. ун-т, 2017. – 44 с.

В курсе «Теория вероятностей и математическая статистика» дается представление о случайных событиях и их вероятностях, случайных величинах и законах их распределения, закономерностях массовых однородных случайных процессов. Рассматриваются методы сбора, обработки и интерпретации статистических данных для получения научных и практических результатов.

Оглавление

Тема 1. Элементы комбинаторики
Лекция 1.
Аннотация. Данная тема раскрывает основные понятия комбинаторики.
Ключевые слова. Комбинации, перестановки, сочетания, размещения.
Методические рекомендации по изучению темы.

- Тема содержит лекционную часть, где даются общие представления по теме;
- В дополнение к лекции есть презентация, которую необходимо изучить, и ответить на вопросы;
- В качестве самостоятельной работы предлагается выполнить задания [2]: №№ 11.34-11.46;

Рекомендуемые информационные ресурсы:
1. http://bars.kpfu.ru/course/view.php?id=729.
2. Сборник задач по математике для экономистов: учебное пособие для экономических специальностей вузов./ Р. Ш. Марданов, А. Ю. Хасанова, Р.А. Султанов, А. Г. Фатыхов; под научной редакцией проф. Р. Ш. Марданова.- Казань: Казан. Гос. Ун.-т, 2009. – 576 с.
3. Презентация.

Глоссарий.
Перестановки – комбинации, состоящие из одних и тех же **n** различных элементов, отличающиеся только порядком их расположения.
Размещения – комбинации, составленные по m элементов из n различных элементов, которые отличаются либо составом, либо порядком расположения элементов.
Сочетания – комбинации, составленные по **m** элементов из **n** различных элементов, отличающиеся хотя бы одним элементом.

Вопросы для изучения:
1. Основные понятия комбинаторики.
2. Перестановки. Формула числа перестановок.
3. Сочетания. Формула числа сочетаний.
4. Размещения. Формула числа размещений.
5. Принцип суммы и произведения.

1. **Комбинаторика** изучает дискретные объекты, множества (сочетания, перестановки, размещения и перечисления элементов).

2. **Перестановки.** Перестановками называются комбинации, состоящие из одних и тех же **n** различных элементов, отличающиеся только порядком их расположения. Найдем число возможных перестановок из n элементов. Возьмем последний элемент. Его можно поставить на n позиций. В

результате останется n-1 элементов. Проделаем ту же самую операцию, в результате останется n-2 элементов и n-2 свободные позиции. Следовательно число различных перестановок из **n** элементов равно

$$n(n-1)(n-2)\cdots 3\cdot 2\cdot 1 = n!$$

3. **Размещения.** Размещениями называются комбинации, составленные по m элементов из n различных элементов, которые отличаются либо составом, либо порядком расположения элементов.

Число размещений из **n** элементов по **m** равно

$$A_n^m = \frac{n!}{(n-m)!} = n(n-1)(n-2)...(n-m+1) \quad .$$

4. **Сочетания**. Сочетаниями называются комбинации, составленные по **m** элементов из **n** различных элементов, отличающиеся хотя бы одним элементом.

В отличие от размещений в сочетаниях порядок расположения элементов не имеет значения.

Число сочетаний из n элементов по m:

$$C_n^m = \frac{n!}{m!\,(n-m)!}$$

При решении задач комбинаторики используют следующие правила:

1. Принцип суммы. Если некоторый объект А может быть выбран из совокупности объектов **m** способами, а другой объект В может быть выбран **n** способами, то выбрать либо А, либо В можно **(m+n)** способами.

2. Принцип произведения. Если объект А можно выбрать из совокупности объектов **m** способами и после каждого такого выбора объект В можно выбрать **n** способами, то пара объектов (А,В) в указанном порядке может быть выбрана **m·n** способами.

Пример. В футбольной команде имеется 20 игроков: 3 вратаря, 5 защитников, 6 полузащитников и 6 нападающих. Необходимо определить состав на матч, при условии, что в нем будут играть 1 вратарь, 4 защитника, 4 полузащитника и 2 нападающих. Найти число вариантов комплектования состава.

Решение.

$$N = C_3^1\ C_5^4\ C_6^4\ C_6^2 = 3\cdot 5\cdot 15\cdot 15 = 3375$$

Тема 2. Основные понятия и теоремы теории вероятностей

Лекция 1.

Аннотация. В данной теме вводится понятие события, дается классификация событий, вводятся действия над событиями, даются определения классической вероятности и относительной частоты наступлений события, формулируются теоремы сложения и умножения вероятностей.

Ключевые слова. Событие, несовместные, единственно возможные, равновозможные, полная группа событий, вероятность, относительная частота, сложение, умножение вероятностей.

Методические рекомендации по изучению темы.

- Тема содержит лекционную часть, где даются общие представления по теме;
- В дополнение к лекции есть презентация, которую необходимо изучить, и ответить на вопросы;
- В качестве самостоятельной работы предлагается выполнить задания [2]: №№ 11.13-11.23; 11.62-11.67; 11.127-11.137.

Рекомендуемые информационные ресурсы:

1. http://bars.kpfu.ru/course/view.php?id=729.
2. Сборник задач по математике для экономистов: учебное пособие для экономических специальностей вузов./ Р. Ш. Марданов, А. Ю. Хасанова, Р.А. Султанов, А. Г. Фатыхов; под научной редакцией проф. Р. Ш. Марданова.- Казань: Казан. Гос. Ун.-т, 2009. – 576 с.
3. Презентация.

Список сокращений.

ТВ – теория вероятностей.

Глоссарий.

Событие – качественная характеристика результата опыта или испытания.

Пространство элементарных событий – множество всех различных исходов случайного эксперимента.

Благоприятствующие события – элементарные события (исходы), при которых интересующее нас событие наступает.

Единственно возможные события – появление одного из них в одном испытании является достоверным событием.

Зависимые события. События А и В называются *зависимыми*, если вероятность наступления одного события зависит от того, наступило или нет другое событие.

Классическая вероятность события A – это отношение числа элементарных событий M, благоприятствующих событию A, к общему числу событий N.

Множество элементарных событий – полная группа равновозможных попарно несовместных событий.

Несовместные события – появление одного из них исключает появление других событий в одном испытании.

Относительная частота события *А* – это отношение числа опытов (испытаний) *m*, в которых событие *А* наступило, к общему числу фактически проведенных испытаний *n*.

Полная группа событий – совокупность единственно возможных событий.

Произведением двух или нескольких событий называется событие, состоящее в совместном наступлении этих событий.

Противоположные события – это два несовместных события, образующих полную группу событий.

Равновозможные события – появление одного из них в одном испытании не является объективно более возможным, чем появление других событий.

Событие – это результат опыта или эксперимента.

Совместные события. Два события называются *совместными*, если наступление одного из них не исключает наступление другого в одном испытании.

Сумма двух событий А и В – это событие *С*, состоящее в появлении или *А*, или *В*, или обоих событий: $C = A + B$.

Вопросы для изучения:

1. Сущность и условия применимости теории вероятностей. Основные понятия теории вероятностей.
2. Случайные события, их виды. Полная группа событий. Пространство элементарных событий. Алгебра событий.
3. Классическое определение вероятности. Относительная частота наступления события.
4. Теоремы сложения вероятностей несовместных и совместных событий

1. Теория вероятностей изучает закономерности массовых однородных случайных процессов и является теоретическим обоснованием математической статистики. Одним из основных понятий ТВ является **событие.**

2. Событие – это качественная характеристика результата опыта или испытания. События обозначаются заглавными буквами латинского алфавита: А, В, С или B_1, B_2, B_3 и т.д.

Множество всех различных исходов случайного эксперимента называется **пространством элементарных событий.**

Классификация событий :

достоверные

События → случайные

невозможные

Событие называется **достоверным,** если при выполнении определенных условий оно обязательно произойдет.

Событие называется **невозможным,** если при выполнении определенных условий оно заведомо не произойдет.

Событие называется **случайным,** если при выполнении определенных условий оно может либо произойти, либо не произойти.

Основным объектом внимания в теории вероятностей являются случайные события.

Классификация случайных событий

несовместные

Случайные события → равновозможные

единственно возможные.

События называются **несовместными**, если появление одного из них исключает возможность появления другого события.

События A= {попал } и B = {промахнулся} при одном выстреле несовместны.

Если появление одного из двух событий не исключает возможность появления другого, то события совместны.

События:

A= {Петров – первокурсник} и B= {Петров – отличник} являются совместными.

События называются **равновозможными**, если появление одного из них не более возможно, чем появление остальных.

События называются **единственно возможными**, если при испытании одно из них обязательно произойдет, то есть является достоверным.

Несколько событий образуют **полную группу событий**, если они несовместны и единственно возможны.

Два события несовместных и единственно возможных называются **противоположными**. Например, события «экзамен сдан» и «экзамен не сдан» являются противоположными при одном испытании для одного студента, т.к. образуют полную группу. А события «холодно» и «жарко» не являются противоположными, т.к. не образуют полной группы.

Суммой двух событий А и В называется событие С = А + В, которое состоит в наступлении либо А, либо В, либо и А, и В одновременно.

Произведением двух событий А и В называется событие С=А•В, которое состоит в совместном наступлении и А, и В.

Аналогично выполняются сложение и умножение **n** событий, например,

$$A = B_1 \bullet B_2 \bullet B_3.$$

3. Событие В называется **благоприятствующим** событию А, если из появления события В следует появление события А.

Численной мерой возможности появления события А является вероятность этого события.

Отношение числа М событий, благоприятствующих событию А, к общему числу N событий этой группы называется **классической вероятностью** события А:

.

$$P(A) = \frac{M}{N}$$

Для любого события А справедливо: $0 \leq \mathrm{P}(\mathrm{A}) \leq 1.$

Кроме классической вероятности применяется статистическая вероятность, которая называется относительной частотой наступления события.

Пусть в **n** испытаниях событие А наступило **m** раз. Тогда **относительная частота** наступления события (частость) равна

$$w(A) = \frac{m}{n},$$

причем $0 \leq w(A) \leq 1.$

4. **Теорема.** Вероятность появления одного из двух несовместных событий, безразлично какого, равна сумме вероятностей этих событий:

$$P(A+B) = P(A \text{ или } B) = P(A) + P(B).$$

Следствие 1. Вероятность появления одного из нескольких попарно несовместных событий равна сумме вероятностей этих событий

$$P(B_1+B_2+\ldots+B_n)=P(B_1)+P(B_2)+\ldots+P(B_n).$$

Следствие 2. Сумма вероятностей событий, образующих полную группу, равна 1.

$$P(B_1)+P(B_2)+\ldots+P(B_n)=1.$$

Следствие 3. Сумма вероятностей противоположных событий равна 1.

Теорема. Вероятность появления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий минус вероятность их совместного появления:

$$P(A+B) = P(A) + P(B) - P(AB).$$

Лекция 2.

Аннотация. В данной теме вводится понятие зависимых и независимых событий, дается определение условной вероятности, формулируются теоремы умножения вероятностей, даются формулы полной вероятности и Байеса.

Ключевые слова. Зависимые, условная вероятность, умножение вероятностей, хотя бы одно событие, гипотезы, полная группа, полная вероятность.

Методические рекомендации по изучению темы.

- Тема содержит лекционную часть, где даются общие представления по теме;
- В дополнение к лекции есть презентация, которую необходимо изучить, и ответить на вопросы;
- В качестве самостоятельной работы предлагается выполнить задания [2]: №№ 11.105-11-115; 11.157-11.165.

Рекомендуемые информационные ресурсы:

1. http://bars.kpfu.ru/course/view.php?id=729.

2. Сборник задач по математике для экономистов: учебное пособие для экономических специальностей вузов./ Р. Ш. Марданов, А. Ю. Хасанова, Р.А. Султанов, А. Г. Фатыхов; под научной редакцией проф. Р. Ш. Марданова.- Казань: Казан. Гос. Ун.-т, 2009. – 576 с.

3. Презентация.

Глоссарий.

Независимые события. События А и В называются *независимыми*, если вероятность наступления одного события не зависит от того, наступило или нет другое событие.

Условная вероятность. Если события A и B зависимы, то вероятность события B, вычисленная при условии, что наступило событие A, называется *условной вероятностью В*.

Вопросы для изучения:

1. Независимые и зависимые события. Условная вероятность.
2. Теоремы умножения вероятностей зависимых и независимых событий.
3. Вероятность появления хотя бы одного события.
4. Формула полной вероятности. Формулы Байеса.

Теорема (о вероятности произведения двух независимых событий). Вероятность произведения двух независимых событий (совместного наступления) равна произведению вероятностей этих событий: $P(A \cdot B) = P(A) \cdot P(B)$. Для трех независимых событий: $P(A \cdot B \cdot C) = P(A) \cdot P(B) \cdot P(C)$.

Теорема умножения вероятностей зависимых событий. Вероятность совместного наступления двух зависимых событий (произведение двух зависимых событий) равна произведению вероятности одного из них на условную вероятность другого, вычисленную при условии, что первое событие наступило: $P(A \cdot B) = P(A) \cdot P_A(B)$ или $P(A \cdot B) = P(B) \cdot P_B(A)$.

Теорема сложения вероятностей совместных событий. $P(A+B) = P(A) + P(B) - P(A \cdot B)$. Причем, если события A и B – независимы, то $P(A+B) = P(A) + P(B) - P(A) \cdot P(B)$ если A и B – зависимы, то $P(A+B) = P(A) + P(B) - P(A) \cdot P_A(B)$ или $P(A+B) = P(A) + P(B) - P(B) \cdot P_B(A)$.

Вероятность наступления хотя бы одного из независимых событий $A_1, A_2, ..., A_n$ равна разности между единицей и произведением вероятностей противоположных событий:

$$P(B) = 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot ... \cdot P(\overline{A_n}) = 1 - q_1 q_2 ... q_n.$$

Формула полной вероятности. Формулы Байеса. Пусть требуется определить вероятность события A, которое может наступить при условии наступления одного из событий $B_1, B_2, ..., B_n$, образующих полную группу несовместных событий, называемых гипотезами. Тогда вероятность события A определяется формулой полной вероятности:

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + ... + P(B_n) \cdot P_{B_n}(A) = \sum_{i=1}^{n} P(B_i) \cdot P_{B_i}(A).$$

Формула Байеса (теорема о вероятности гипотез). Имеется полная группа несовместных гипотез $B_1, B_2, ..., B_n$, вероятности которых до опыта известны. Произведен опыт, в результате которого наступило событие A. Нужно определить, как изменятся вероятности гипотез $B_1, B_2, ..., B_n$ в связи с наступлением события A. Эти вероятности определяются с помощью формул Бейеса:

$$P_A(B_j) = \frac{P(B_j) \cdot P_{B_j}(A)}{P(A)} = \frac{P(B_j) \cdot P_{B_j}(A)}{\sum_{i=1}^{n} P(B_i) \cdot P_{B_i}(A)}, j = \overline{1, n}.$$

Пример. Одну и ту же операцию выполняют рабочие 5-го, 4-го и 3-го разрядов. В бригаде из 10 человек двое имеют 5-й разряд, пятеро – 4-й и трое – 3-й разряд. Рабочие 5–го разряда допускают 1% брака, 4-го – 2%, 3–го разряда – 4%. Пусть наудачу взятая деталь оказалась годной (т.е. событие А наступило). Какова вероятность того, что эта деталь изготовлена рабочим 3-го разряда?

Решение. Через B_i обозначим событие, означающее, что деталь изготовлена рабочим i-того разряда. Искомая вероятность равна

$$P_A(B_3) = \frac{P(B_3) \cdot P_{B_3}(A)}{\sum_{i=3}^{5} P(B_i) \cdot P_{B_i}(A)} = \frac{0{,}3 \cdot 0{,}96}{0{,}3 \cdot 0{,}96 + 0{,}5 \cdot 0{,}98 + 0{,}2 \cdot 0{,}99} = 0{,}295$$

Тема 3. Повторные независимые испытания

Аннотация. В данной теме даются понятия повторных независимых испытаний и наивероятнейшего числа наступления события, а также приводятся формулы расчета вероятности наступления события m раз или от a до b раз в n испытаниях и вероятности отклонения относительной частоты события от вероятности его наступления в одном испытании.

Ключевые слова. Повторные независимые испытания, формула Бернулли, наивероятнейшее число наступления событий, функции Лапласа

Методические рекомендации по изучению темы

- тема содержит лекционную часть, в ходе изучения которой занятия необходимо отмечать наиболее существенную информацию, новые термины и понятия, сравнивать новый материал с изученным и усвоенным ранее, устанавливать их взаимосвязь и укладывать новую информацию в уже имеющуюся систему знаний;
- в дополнение к лекционной части приведены примеры решения практических заданий по рассматриваемой теме;
- в качестве самостоятельной работы предлагается решить контрольные задания;
- для самоконтроля приводятся вопросы к теме и тесты.

Рекомендуемые информационные ресурсы

1. http://bars.kpfu.ru/course/view.php?id=783
2. http://tvims.wordpress.com/
3. http://ru.wikibooks.org/wiki/Математика_случая
4. http://teorver-online.narod.ru/
5. http://statistica.ru/theory/

Список сокращений

ТВ – теория вероятностей
ПНИ – повторные независимые испытания

Глоссарий

Наивероятнейшее число наступлений события A или наивероятнейшая частота наступления события A – это число $m=m_0$, при котором вероятность $P_{n,m}$ наступления события A в n независимых испытаниях m раз будет наибольшей.

Повторные независимые испытания – испытания, проводящиеся в одних и тех же условиях с одинаковой вероятностью наступления события A в каждом отдельном испытании.

Вопросы для изучения

1. Формула Бернулли.
2. Наивероятнейшее число наступлений событий.

3. Локальная теорема Лапласа.
4. Интегральная теорема Лапласа.

Формула Бернулли. Пусть проводится n независимых испытаний, в каждом отдельном испытании событие A может появится с вероятностью p и не появиться с вероятностью $q = 1 - p$. При этом нас будет интересовать не исход отдельного испытания, а вероятность того, что в n испытаниях событие A появится m раз. Вероятность появления события A в n испытаниях m раз обозначается $P_{n,m}$. Получаем *формулу Бернулли*, которая используется для вычисления $P_{n,m}$ при $n \le 10$:

$$P_{n,m} = C_n^m p^m q^{n-m}, \text{ где } C_n^m = \frac{n!}{m!(n-m)!}.$$

Пример. Из n аккумуляторов за год хранения k выходит из строя. Наудачу выбирают m аккумуляторов. Определить вероятность того, что среди них s исправных.

Решение. Через p обозначим вероятность того, что выбранный аккумулятор исправен. Очевидно, что $p = 1 - k/n$. Следовательно, вероятность того, что исправны s аккумуляторов равна

$$P_{m,s} = C_m^s \left(1 - \frac{k}{n}\right)^s (k/n)^{1-s} = C_m^s\, (n-k)^s k^{1-s}/n$$

Наивероятнейшее число наступлений событий. *Наивероятнейшее число наступления события A* m=m_0 вычисляется по следующей формуле: $np - q \le m_0 \le np + p$. Причем число m=m_0 должно быть целым.

1. Если np –целое, то m_0 – единственное.
2. Если $np+p$ – нецелое, то m_0 – единственное.
3. Если $np+p$ – целое, то будут два наивероятнейших числа наступления события m_{01}, m_{02}.

Локальная теорема Лапласа. Если $n>10$, то расчеты по формуле Бернулли становятся очень громоздкими, поэтому вычислить позволяет *локальная теорема Лапласа*. Если вероятность появления события A в каждом отдельном испытании постоянна и отлична от нуля и единицы, то вероятность наступления события A в n испытаниях m раз приближенно вычисляется по формуле: $P_{n,m} \approx \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$, где

$x = \frac{m-np}{\sqrt{npq}}$. Для удобства была введена функция $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$ – малая функция Лапласа, значения которой вычислены и приведены в таблице приложения 1. Свойства функции $\varphi(x)$;

1. Функция $\varphi(x)$ четная, т.е. $\varphi(-x) = \varphi(x)$.
2. $\lim\limits_{x\to\infty} \varphi(x) = 0$.
3. $\varphi_{\max}(x) = \varphi(0) = 0{,}3989$.

Интегральная теорема Лапласа. Если вероятность наступления события A в каждом отдельном испытании постоянна и отлична от нуля и единицы, то вероятность того, что в n независимых испытаниях событие A появится не менее a и не более b раз приближенно вычисляется по формуле: $P(a \le m \le b) \approx \frac{1}{\sqrt{2\pi}} \int\limits_{\alpha}^{\beta} e^{-\frac{t^2}{2}} dt$, где $\alpha = \frac{a-np}{\sqrt{npq}}$, $\beta = \frac{b-np}{\sqrt{npq}}$. Для удобства была введена функция $\Phi(x) = \frac{2}{\sqrt{2\pi}} \int\limits_{0}^{x} e^{-\frac{t^2}{2}} dt$ – большая функция Лапласа, значения которой вычислены и приведены в таблице приложения 2. Подставляя выражение $\Phi(x)$ в формулу, получаем рабочую формулу для вычисления вероятности появления события A в испытаниях от a до b раз: $P(a \le m \le b) \approx \frac{1}{2}\left[\Phi(\beta) - \Phi(\alpha)\right]$.

Свойства функции $\Phi(x)$:

1. Функция $\Phi(x)$ нечетная, т.е. $\Phi(-x) = -\Phi(x)$.
2. Функция $\Phi(x)$ монотонно возрастающая.
3. $\lim\limits_{x\to+\infty} \Phi(x) = 1$.
4. При $x > 4.5$ можно считать, что $\Phi(x) \approx 1$.

Следствие из интегральной теоремы Лапласа. С вероятностью, близкой $\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)$, можно утверждать, что при достаточно большом числе испытаний абсолютная величина отклонения относительной частоты события A от его вероятности не превзойдет достаточно малого положительного числа ε:

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) \approx \Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right).$$

Тема 4. Дискретная случайная величина

Аннотация. В данной теме даются понятия дискретной случайной величины, закона распределения, математического ожидания, дисперсии и среднего квадратического отклонения, приводятся формулы расчета числовых характеристик СВ.

Ключевые слова. Случайная величина, закон распределения, числовые характеристики СВ, индикатор события

Методические рекомендации по изучению темы

- тема содержит лекционную часть, в ходе изучения которой занятия необходимо отмечать наиболее существенную информацию, новые термины и понятия, сравнивать новый материал с изученным и усвоенным ранее, устанавливать их взаимосвязь и укладывать новую информацию в уже имеющуюся систему знаний;
- в дополнение к лекционной части приведены примеры решения практических заданий по рассматриваемой теме;
- в качестве самостоятельной работы предлагается решить контрольные задания;
- для самоконтроля приводятся вопросы к теме и тесты.

Рекомендуемые информационные ресурсы

1. http://bars.kpfu.ru/course/view.php?id=783
2. http://tvims.wordpress.com/
3. http://ru.wikibooks.org/wiki/Математика_случая
4. http://teorver-online.narod.ru/
5. http://statistica.ru/theory/

Список сокращений

ТВ – теория вероятностей
СВ – случайная величина
ЗР – закон распределения
ДСВ – дискретная случайная величина

Глоссарий

Дискретная случайная величина – это СВ, которая принимает конечное или счетное множество значений.

Дисперсия D(X) – это математическое ожидание квадрата отклонения значений случайной величины от ее математического ожидания.

Закон распределения ДСВ – это соответствие между возможными значениями СВ x_i и вероятностями их появления p_i.

Индикатор события A – это СВ X, равная единице, если в результате испытания событие A происходит, и равная нулю, если событие A не происходит.

Математическое ожидание M(X) ДСВ – есть сумма произведений всех возможных значений случайной величины x_i на соответствующие вероятности p_i, интерпретируется как среднее значение СВ.

Среднее квадратическое отклонение $\sigma(X)$ – это корень квадратный из дисперсии.

Вопросы для изучения

1. Определения случайной величины, дискретной случайной величины.
2. Закон распределения дискретной случайной величины.
3. Числовые характеристики дискретной случайной величины

Определения случайной величины, дискретной случайной величины. *Дискретной случайной величиной* называется СВ, которая принимает не более чем счетное множество значений.

СВ обозначаются заглавными буквами латинского алфавита: X, Y, Z, а возможные значения СВ соответствующими строчными буквами латинского алфавита. Каждое значение ДСВ наступает с определенной вероятностью: x_1 наступает с вероятностью p_1, x_2 с вероятностью p_2 и т.д.

Закон распределения дискретной случайной величины. Законом распределения ДСВ называется соответствие между возможными значениями СВ x_i и вероятностями их появления p_i.

Существуют две формы задания закона распределения: табличная и графическая.

Графическая форма представляет собой ломаную линию, соединяющую точки прямоугольной системы координат и называется *полигоном распределения вероятностей*.

Табличная форма называется рядом распределения и представляется в виде двух строк. В первой строке даны в порядке возрастания все значения СВ, во второй строке – вероятности их появления. Причем, $\Sigma p_i = 1$.

Числовые характеристики дискретной случайной величины. Закон распределения дает полную характеристику ДСВ с вероятностной точки зрения. Но часто бывают необходимы количественные показатели СВ, которые могли бы выразить существенные особенности заданного распределения. К такими показателям относятся *числовые характеристики СВ* – математическое ожидание, дисперсия и среднее квадратическое отклонение.

Математическое ожидание $M(X)$ – есть сумма произведений всех возможных значений случайной величины x_i на соответствующие вероятности p_i.

$$M(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 + \ldots + x_n \cdot p_n$$

Математическое ожидание $M(X)$ интерпретируется как среднее значение СВ, вокруг которого распределены все возможные значения СВ.

Дисперсией D(X) ДСВ называется математическое ожидание квадрата отклонения значений случайной величины от ее математического ожидания: $D(X) = M[X - M(X)]^2$.

$$D(X) = \sum_{i=1}^{n} (x_i - M(X))^2 p_i = \sum_{i=1}^{n} x_i^2 p_i - (M(X))^2.$$

Дисперсия иногда обозначается как $\sigma^2(X)$.

Дисперсия означает разброс или рассеяние значений СВ около ее среднего значения (математического ожидания).

Средним квадратическим отклонением $\sigma(X)$ ДСВ называется корень квадратный из дисперсии: $\sigma(X) = \sqrt{D(X)}$.

Теоремы о $M(X)$ и $D(X)$ числа появлений события в n независимых испытаниях.

Т1. Математическое ожидание индикатора события A равно вероятности появления события A, т.е. $M(X) = p$.

Т2. Дисперсия индикатора события A равна произведению pq, т.е. $D(X) = pq$.

Т3. Математическое ожидание СВ – числа появлений события A в n независимых испытаниях равно np, т.е. $M(X)= np$.

Т4. Дисперсия СВ – числа появлений события A в n независимых испытаниях равно npq, т.е. $D(X)= npq$.

Тема 5. Непрерывная случайная величина

Аннотация. В данной теме дается понятие непрерывной случайной величины, рассматриваются два способа ее задания, приводятся свойства интегральной и дифференциальной функций распределения, а также формулы расчета числовых характеристик НСВ и вероятности попадания НСВ в заданный интервал.

Ключевые слова. Функция распределения, плотность распределения

Методические рекомендации по изучению темы

- тема содержит лекционную часть, в ходе изучения которой занятия необходимо отмечать наиболее существенную информацию, новые термины и понятия, сравнивать новый материал с изученным и усвоенным ранее, устанавливать их взаимосвязь и укладывать новую информацию в уже имеющуюся систему знаний;
- в дополнение к лекционной части приведены примеры решения практических заданий по рассматриваемой теме;
- в качестве самостоятельной работы предлагается решить контрольные задания;
- для самоконтроля приводятся вопросы к теме и тесты.

Рекомендуемые информационные ресурсы

1. http://bars.kpfu.ru/course/view.php?id=783
2. http://tvims.wordpress.com/
3. http://ru.wikibooks.org/wiki/Математика_случая
4. http://teorver-online.narod.ru/
5. http://statistica.ru/theory/

Список сокращений

ТВ – теория вероятностей

СВ – случайная величина
НСВ – непрерывная случайная величина

Глоссарий

Абсолютно непрерывная функция. Функция f, заданная на отрезке $[a,b]$ называется *абсолютно непрерывной*, если для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что для произвольной конечной системы попрано непересекающихся интервалов (a_k, b_k), $k = 1,2,\dots,n$ с суммой длин, меньшей δ, выполнено неравенство $\sum_{k=1}^{n}|f(b_k) - f(a_k)| < \varepsilon$.

Функция распределения СВ X – это функция, которая определяет для каждого значения аргумента x вероятность того, что величина X примет значение меньшее x.

Непрерывная случайная величина – это СВ, у которой функция распределения является абсолютно непрерывной на любом отрезке числовой оси.

Плотностью распределения $f(x)$ называется предел отношения вероятности попадания непрерывной СВ X на элементарный участок $(x;\ x+\Delta x)$ к длине участка Δx при $\Delta x \to 0$.

Вопросы для изучения

1. Функция распределения и ее свойства.
2. Плотность распределения вероятностей и ее свойства.
3. Математическое ожидание и дисперсия.

Функция распределения и ее свойства. Непрерывная случайная величина имеет бесконечное несчетное множество значений и задается двумя способами: с помощью функции распределения (интегральная функция распределения) и плотности распределения (дифференциальная функция распределения).

Функцией распределения СВ X называют функцию, которая определяет для каждого значения аргумента x вероятность того, что величина X примет значение меньшее x: $F(x) = P(X<x)$.

Свойства функции $F(x)$:

1. $0 \le F(x) \le 1$.
2. Вероятность того, что СВ примет значение, удовлетворяющее одному из двойных неравенств, равна приращению $F(x)$ на этом интервале:

$P(a \le X < b) = P(a < X \le b) = P(a \le X \le b) = P(a < X < b) = F(b) - F(a)$.
3. Функция распределения является неубывающей функцией: $F(x_2) \ge F(x_1)$ при $x_2 \ge x_1$.
4. $\lim\limits_{x \to -\infty} F(x) = 0$ и $\lim\limits_{x \to +\infty} F(x) = 1$.

Плотность распределения вероятностей и ее свойства. *Плотностью распределения f(x)* называется предел отношения вероятности попадания непрерывной СВ X на элементарный участок $(x; x+\Delta x)$ к длине участка Δx при $\Delta x \to 0$.

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x < X < x + \Delta x)}{\Delta x} = F'(x).$$

Свойства плотности распределения:

1. Плотность распределения $f(x)$ неотрицательна.
2. Вероятность того, что непрерывная СВ примет значения на интервале $(a; b)$, равна:

$$P(a < X < b) = \int_a^b f(x)dx.$$

3. Функция распределения $F(x)$ равна:

$$F(x) = \int_{-\infty}^{x} f(x)dx.$$

4. Интеграл от плотности распределения $f(x)$ на всей числовой оси равен 1:

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

Математическое ожидание и дисперсия. Математическое ожидание $M(X)$ непрерывной СВ X, возможные значения которой находятся в интервале (a, b) определяется как

$$M(X) = \int_a^b x \cdot f(x)dx.$$

Если возможные значения непрерывной СВ X принадлежат всей числовой оси, то математическое ожидание $M(X)$ определяется по формуле:

$$M(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx.$$

Дисперсия $D(X)$ непрерывной СВ, заданной на интервале (a, b) вычисляется как:

$$D(X)=\int_a^b (x-M(X))^2\cdot f(x)dx=\int_a^b x^2\cdot f(x)dx-(M(X))^2.$$

Если возможные значения непрерывной СВ X принадлежат всей числовой оси, то дисперсия $D(X)$ определяется по формуле:

$$D(X)=\int_{-\infty}^{+\infty} (x-M(X))^2\cdot f(x)dx=\int_{-\infty}^{+\infty} x^2\cdot f(x)dx-(M(X))^2.$$

Средним квадратическим отклонением $\sigma(X)$ НСВ называется корень квадратный из дисперсии:

$$\sigma(X)=\sqrt{D(X)}.$$

Следует отметить, что дискретные и непрерывные случайные величины не исчерпывают весь класс случайных величин. Встречаются еще так называемые сингулярные случайные величины, для которых не существует (в классическом понимании) плотности распределения. В этом случае

$$M(X)=\int_{-\infty}^{\infty} x\cdot dF(x),$$

$$D(X)=\int_{-\infty}^{+\infty} (x-M(X))^2\, dF(x)=\int_{-\infty}^{+\infty} x^2 dF(x)-(M(X))^2.$$

Отметим, что последние формулы носят универсальный характер и годятся для произвольной случайной величины.

Тема 6. Основные законы распределения случайной величины

Аннотация. В данной теме дается понятие нормального закона распределения, приводятся его параметры, рассматривается плотность распределения и ее график, приводятся свойства кривой Гаусса, формулы расчета вероятности попадания нормально распределенной СВ в заданный интервал, вероятности ее отклонения от среднего значения и формулируется «правило трех сигм».

Ключевые слова. Нормальное распределение, кривая Гаусса, правило трех сигм

Методические рекомендации по изучению темы

- тема содержит лекционную часть, в ходе изучения которой занятия необходимо отмечать наиболее существенную информацию, новые термины и понятия, сравнивать новый материал с изученным и

усвоенным ранее, устанавливать их взаимосвязь и укладывать новую информацию в уже имеющуюся систему знаний;
- в дополнение к лекционной части приведены примеры решения практических заданий по рассматриваемой теме;
- в качестве самостоятельной работы предлагается решить контрольные задания;
- для самоконтроля приводятся вопросы к теме и тесты.

Рекомендуемые информационные ресурсы

1. http://bars.kpfu.ru/course/view.php?id=783
2. http://tvims.wordpress.com/
3. http://ru.wikibooks.org/wiki/Математика_случая
4. http://teorver-online.narod.ru/
5. http://statistica.ru/theory/

Список сокращений

ТВ – теория вероятностей
СВ – случайная величина
ЗР – закон распределения
НР – нормальное распределение

Глоссарий

Нормальным (или распределением по закону Гаусса) называется распределение непрерывной случайной величины, плотность распределения которой определяется формулой:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}},x\in R.$$

Вопросы для изучения

1. Закон нормального распределения. Влияние параметров нормального распределения на форму кривой нормального распределения.
2. Теоремы о нормально распределенной случайной величине. Правило трех сигм. Функция Лапласа.

Закон нормального распределения. Влияние параметров нормального распределения на форму кривой нормального распределения. Интегральная функция распределения $F(x)$ принимает вид:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt.$$

Нормальное распределение определяется двумя параметрами a и σ, что символически обозначается следующим образом $N(a,\sigma)$.

$$a = M(X), \quad \sigma^2 = D(X)..$$

Выясним геометрический смысл параметров распределения a и σ. График функции $f(x)$ называется *нормальной кривой* или *кривой Гаусса*.

При изменении параметра a форма нормальной кривой не изменяется. В этом случае, если математическое ожидание (параметр a) уменьшилось или увеличилось, график нормальной кривой сдвигается влево или вправо .

При изменении параметра σ изменяется форма нормальной кривой. Если этот параметр увеличивается, то максимальное значение функции $f(x)$ убывает, и наоборот. Так как площадь, ограниченная кривой распределения и осью *Ox,* должна быть постоянной и равной 1, то с увеличением параметра кривая приближается к оси *Ox* и растягивается вдоль нее, а с уменьшением σ кривая стягивается к прямой $x = a$.

Теоремы о нормально распределенной случайной величине. Правило трех сигм. Функция Лапласа.

1. Вероятность того, что нормально распределенная случайная величина X примет значения из интервала $(c;d)$, равна

$$P(c < X < d) = \frac{1}{2}\left[\Phi\left(\frac{d-a}{\sigma}\right) - \Phi\left(\frac{c-a}{\sigma}\right)\right]..$$

2. Вероятность того, что отклонение нормально распределенной случайной величины X от ее математического ожидания не превзойдет по абсолютной величине $a > 0$, равна

$$P(|X - a| < \alpha) = \Phi\left(\frac{\alpha}{\sigma}\right)..$$

3. Нормально распределенная случайная величина X с большой вероятностью принимает значения, близкие к своему математическому ожиданию, что описывается «правилом трех сигм»:

$$P(|X-a|<3\sigma)=\Phi\left(\frac{3\sigma}{\sigma}\right)=\Phi(3)=0{,}9973.$$

Тема 7. Закон больших чисел

Аннотация. В данной теме приводятся теоремы и связанные с ними оценочные неравенства закона больших чисел.

Ключевые слова. Сходимость по вероятности.

Методические рекомендации по изучению темы.

- Тема содержит лекционную часть, где даются общие представления по теме;
- В дополнение к лекции есть презентация, которую необходимо изучить, и ответить на вопросы;
- В качестве самостоятельной работы предлагается выполнить задания [2]: №№ 11.34-11.46;

Рекомендуемые информационные ресурсы:

1. http://bars.kpfu.ru/course/view.php?id=729.
2. Сборник задач по математике для экономистов: учебное пособие для экономических специальностей вузов./ Р. Ш. Марданов, А. Ю. Хасанова, Р.А. Султанов, А. Г. Фатыхов; под научной редакцией проф. Р. Ш. Марданова.- Казань: Казан. Гос. Ун.-т, 2009. – 576 с.
3. Колмогоров А.Н. Фомин С.В., Элементы теории функций и функционального анализа, ФИЗМАТЛИТ, 2004

Список сокращений

ЗБЧ – закон больших чисел;

СВ – случайная величина

Вопросы для изучения:

1. Неравенство Чебышева, лемма Маркова.
2. Обобщенная теорема Чебышева. Сущность теоремы Чебышева и ее значение для экономической практики.
3. Закон больших чисел и его следствия.
4. Теоремы Бернулли и Пуассона, их экономический смысл.
5. Особая роль нормального распределения. Центральная предельная теорема.

ЗБЧ – это совокупность теорем и связанных с ними неравенств, дающих ответ на следующие важные вопросы:

– каким должно быть число испытаний для обеспечения заданной точности;

– каковы границы возможного разброса значений СВ при заданном числе испытаний и заданной надежности;

– с какой вероятностью (надежностью) можно доверять результатам, полученным при известном числе испытаний и заданной величине разброса СВ?

Суть ЗБЧ заключается в следующем: при большом числе СВ X_i с вероятностью, близкой к 1, можно утверждать, что абсолютная величина разности между средней арифметической этих СВ и константой – средней арифметической их математических ожиданий $M(X_i)$ сколь угодно мала. Отсюда следует, что **при большом числе испытаний с большой вероятностью средние характеристики СВ стремятся к постоянным неслучайным величинам.**

Смысл ЗБЧ в том, что разброс в воздействии отдельного явления мало сказывается на среднем результате большого числа явлений.

ЗБЧ является теоретической основой выборочного метода.

Определение 1. Последовательность СВ X_1, X_2, …, X_n называется **сходящейся по вероятности** при $n \to \infty$ к СВ X, если для любого сколь угодно малого числа $\varepsilon > 0$ предел

$$\lim_{n \to \infty} P(|X_n - X| < \varepsilon) = 1.$$

Неравенство Чебышёва. С вероятностью, большей, чем $1-\frac{D(X)}{\alpha^2}$, можно утверждать, что при достаточно большом числе испытаний абсолютная величина отклонения СВ X от ее математического ожидания не превзойдет положительного числа $\alpha > 0$, т.е.

$$P(|X - M(X)| \le \alpha) > 1-\frac{D(X)}{\alpha^2}.$$

Здесь D(X) – дисперсия СВ X.

Пусть СВ X принимает положительные значения.

Лемма Маркова. С вероятностью, большей чем $\mathbf{1-\frac{1}{t^2}}$, можно утверждать, что при достаточно большом числе независимых испытаний СВ X не превзойдет $\mathbf{t^2}$ – кратного математического ожидания, то есть:

$$P(X \le t^2 M(X)) > 1 - \frac{1}{t^2}.$$

Если обозначить $a = t^2M(X)$, то $\frac{1}{t^2} = \frac{M(X)}{a}$, тогда $P(X \le a) > 1 - \frac{M(X)}{a}$.

Обобщенная теорема Чебышёва. Если дисперсии независимых СВ X_i ограничены сверху числом C = const, то среднее арифметическое этих случайных величин сходится по вероятности к среднему арифметическому их математических ожиданий, т. е. для любого сколь угодно малого наперед заданного числа $\varepsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| \le \varepsilon\right) = 1.$$

Здесь $\frac{\mathbf{1}}{\boldsymbol{n}}\sum_{i=1}^{n} X_i$ - среднее арифметическое СВ X_i, а $\frac{1}{n}\sum_{i=1}^{n} M(X_i)$ – среднее арифметическое их математических ожиданий. Если **n** – число СВ, а их дисперсии $D(X_i) \le C$, то **оценочное неравенство** обобщенной теоремы Чебышёва имеет вид:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| \le \varepsilon\right) > 1 - \frac{C}{n\varepsilon^2}.$$

Теорема Бернулли. При бесконечно большом числе независимых испытаний с одинаковой вероятностью **p** наступления события А в каждом испытании относительная частота $\frac{m}{n}$ наступлений события А сходится по вероятности к вероятности **p** этого события, т. е. для любого, сколь угодно малого числа $\varepsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = 1.$$

Если **n** – число независимых испытаний, а **p** – вероятность события А в каждом испытании, то **оценочное неравенство теоремы Бернулли:**

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) > 1 - \frac{pq}{n\varepsilon^2}.$$

Теорема Пуассона. При неограниченном увеличении числа независимых испытаний с вероятностью $\mathbf{p_i}$ появления события А в i –м испытании (i =1; n) относительная частота $\frac{m}{n}$ события А сходится по вероятности к $\overline{\boldsymbol{p}}$ - среднему взвешенному вероятностей $\overline{\boldsymbol{p}} = \frac{1}{n}\sum_{i=1}^{k} p_i\, m_i$:

$$\mathbf{lim}_{n\to\infty} \mathbf{P}\left(\left|\frac{m}{n} - \mathrm{p}\right| \le \boldsymbol{\varepsilon}\right) = \mathbf{1}.$$

Оценочное неравенство теоремы Пуассона:

$$\mathrm{P}\left(\left|\frac{m}{n} - \mathrm{p}\right| \le \varepsilon\right) > 1 - \frac{\overline{pq}}{n\varepsilon^2},$$

где $\overline{pq} = \frac{1}{n}\sum_{i=1}^{k} p_i\, q_i m_i$ - среднее взвешенное дисперсий.

Теорема Чебышева (закон больших чисел). При неограниченном увеличении числа независимых испытаний **n** над СВ **X,** имеющей конечную дисперсию, среднее арифметическое наблюдаемых значений этой СВ сходится по вероятности к ее математическому ожиданию, т.е.:

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} x_i - M(X)\right| \le \varepsilon\right) = 1.$$

Оценочное неравенство закона больших чисел: Если D(X) **≤ C,** то

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} x_i - M(X)\right| \le \varepsilon\right) > 1 - \frac{C}{n\varepsilon^2}.$$

Тема 8. Выборочный метод

Аннотация. В данной теме приводятся понятия выборки, вариационного ряда, статистического распределения выборки, полигона и гистограммы, ошибок репрезентативности, выборочных числовых характеристик, доверительного интервала, рассматриваются различные способы и типы отбора, приводятся формулы расчета выборочной средней, дисперсии и доли, методы оценки числовых характеристик генеральной совокупности и определения необходимой численности выборки.

Ключевые слова. Выборочная и генеральная совокупность, доверительный интервал, репрезентативность, ошибки репрезентативности

Методические рекомендации по изучению темы

- тема содержит лекционную часть, в ходе изучения которой занятия необходимо отмечать наиболее существенную информацию, новые термины и понятия, сравнивать новый материал с изученным и усвоенным ранее, устанавливать их взаимосвязь и укладывать новую информацию в уже имеющуюся систему знаний;
- в дополнение к лекционной части приведены примеры решения практических заданий по рассматриваемой теме;
- в качестве самостоятельной работы предлагается решить контрольные задания;
- для самоконтроля приводятся вопросы к теме и тесты.

Рекомендуемые информационные ресурсы

1. http://bars.kpfu.ru/course/view.php?id=783
2. http://tvims.wordpress.com/
3. http://ru.wikibooks.org/wiki/Математика_случая
4. http://teorver-online.narod.ru/
5. http://statistica.ru/theory/

Список сокращений

МС – математическая статистика
СВ – случайная величина

Глоссарий

Бесповторный способ отбора. При выборке единица совокупности, попавшая в выборку, в генеральную совокупность не возвращается и в дальнейшем в выборке не участвует.

Варианты – наблюдаемые значения.

Вариационный ряд – последовательность вариант, записанных в возрастающем порядке.

Выборочное наблюдение – это такой вид несплошного наблюдения, при котором характеристика всей генеральной совокупности дается по некоторой ее части (по выборке), отобранной в случайном порядке.

Гистограмма – ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат интервалы, длиной h, а высотой – величины m_i/h.

Доверительным интервалом называется интервал $(\theta^* - \Delta;\, \theta^* + \Delta)$, содержащий неизвестный параметр θ с надежностью β.

Механический отбор – это когда упорядоченно расположенные единицы совокупности отбирают по одной через определенный интервал, называемый шагом выборки.

Надежностью (доверительной вероятностью) оценки θ по θ***** называется вероятность β, с которой осуществляется неравенство $\left|\theta - \theta^*\right| < \Delta$.

Несмещенность оценки означает, что ее математическое ожидание равно значению оцениваемого параметра генеральной совокупности.

Относительная частота наблюдений – отношение числа наблюдений к объему выборки.

Оценкой неизвестного параметра θ называется случайная величина X, с помощью которой делаются выводы о неизвестном значении данного параметра.

Повторный способ отбора. Единицу, попавшую в выборку, после регистрации снова возвращают в генеральную совокупность, и она сохраняет равную возможность со всеми прочими единицами при повторном отборе единиц вновь попасть в выборку.

Полигон – ломаная линия, соединяющая точки (x_i; m_i).

Предельная ошибка выборки – абсолютная величина отклонения выборочной характеристики от генеральной характеристики.

Серийный отбор – из генеральной совокупности отбирают не отдельные единицы, а целые их группы. Внутри отобранной серии производят сплошное наблюдение.

Собственно-случайный отбор – лотерея, жеребьевка, отбор на основе таблицы случайных чисел.

Состоятельность оценки означает, что по мере увеличения объема выборки ее значение приближается к значению оцениваемого параметра генеральной совокупности.

Средняя ошибка выборки – средняя из возможных ошибок.

Статистическим распределением называют перечень вариант и соответствующих им частот или относительных частот.

Типический отбор используется в случае, когда генеральная совокупность неоднородна. Из каждой выделенной группы в случайном порядке отираются отдельные единицы (собственно-случайным отбором), как правило, в объеме, пропорциональном численности единиц по группам в генеральной совокупности.

Частота – число наблюдений.

Эффективность оценки. Состоятельная оценка называется *эффективной*, если она имеет наименьшую дисперсию на всех выборках данного объема n.

Вопросы для изучения

1. Сущность выборочного наблюдения.
2. Выборочная и генеральная совокупности. Способы отбора: повторный и бесповторный. Типы отбора.

3. Статистическое распределение выборки.
4. Характеристики выборочной и генеральной совокупностей.
5. Ошибки репрезентативности.
6. Теорема Чебышева-Ляпунова. Доверительный интервал для генеральной средней и генеральной доли.

Сущность выборочного наблюдения. *Сплошное наблюдение* предусматривает обследование всех единиц изучаемой совокупности и связано с большими трудовыми и материальными затратами. Изучение не всех единиц совокупности, а только некоторой части, по которой следует судить о свойствах всей совокупности в целом, можно осуществить *несплошным* наблюдением. В статистической практике самым распространенным является выборочное наблюдение. *Выборочное наблюдение* – это такой вид несплошного наблюдения, при котором характеристика всей генеральной совокупности дается по некоторой ее части (по выборке), отобранной в случайном порядке. Обращение к выборкам обеспечивает экономию материальных, трудовых и финансовых ресурсов и времени.

Выборочная и генеральная совокупности. Способы отбора: повторный и бесповторный. Типы отбора. Множество элементов, составляющих объект исследования, называют генеральной совокупностью. Суть выборочного метода заключена в том, что обследованию подвергается только часть элементов генеральной совокупности, которая называется выборочной совокупностью.

По методу отбора различают повторную и бесповторную выборки. При повторной выборке ту или иную единицу, попавшую в выборку, после регистрации снова возвращают в генеральную совокупность, и она сохраняет равную возможность со всеми прочими единицами при повторном отборе единиц вновь попасть в выборку. При бесповторной выборке единица совокупности, попавшая в выборку, в генеральную совокупность не возвращается и в дальнейшем в выборке не участвует.

Основные способы отбора: 1. Собственно-случайный отбор. 2. Механический отбор. 3. Типический отбор 4. Серийный отбор.

Статистическое распределение выборки. Пусть из генеральной совокупности извлечена выборка, причем значение x_1 встречалось m_1 и т.д., x_k – m_k. Наблюдаемые значения x_i называется *вариантами,* а последовательность вариант, записанных в возрастающем порядке –

вариационным рядом. Число наблюдений называется *частотами. Статистическим распределением* называют перечень вариант и соответствующих им частот или относительных частот. Если значения признака не дискретны, т.е. заполняют некоторый интервал, то этот интервал разбивают на несколько возрастающих интервалов и получают так называемый интервальный вариационный ряд. Любой интервальный ряд можно превратить в дискретный, используя вместо x_i середины интервалов. Наглядным изображением статистического ряда распределения служат *полигон* и *гистограмма.*

Характеристики выборочной и генеральной совокупностей. Объем выборки n, объем генеральной совокупности N. Средняя величина количественного признака: выборочная средняя $\tilde{x} = \frac{\sum_{i=1}^{k} x_i m_i}{n}$; генеральная средняя $\bar{x} = \frac{\sum_{i=1}^{l} x_i m_i}{N}$. Дисперсия количественного признака: выборочная дисперсия $\sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \tilde{x})^2 m_i}{n}$; генеральная дисперсия $D = \frac{\sum_{i=1}^{l} (x_i - \bar{x})^2 m_i}{N}$. Доля единиц обладающих изучаемым признаком: выборочная доля $w = \frac{m}{n}$; генеральная доля $p = \frac{M}{N}$.

Ошибки репрезентативности. Отклонение числовых характеристик выборочной совокупности от соответствующих характеристик генеральной совокупности называется *ошибкой репрезентативности* (или ошибкой выборки). Случайная (средняя) ошибка обозначается μ (математическое ожидание средней ошибки выборки) и представляет собой среднюю квадратическую из всех ошибок выборки. а) для среднего значения: $\mu_x = \sqrt{\frac{\sigma^2}{n}}$ – при повторном отборе; $\mu_x = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}$ – при бесповторном отборе; б) для доли:

$\mu_w = \sqrt{\dfrac{w(1-w)}{n}}$ при повторном отборе; $\mu_w = \sqrt{\dfrac{w(1-w)}{n}\left(1-\dfrac{n}{N}\right)}$ – при бесповторном отборе.

Абсолютную величину этого отклонения обозначают Δ и называют предельной ошибкой выборки. а) для среднего значения: $\Delta_x = t\mu_x = t\sqrt{\dfrac{\sigma^2}{n}}$ – при повторном отборе; $\Delta_x = t\mu_x = t\sqrt{\dfrac{\sigma^2}{n}\left(1-\dfrac{n}{N}\right)}$ - при бесповторном отборе, где t – коэффициент надежности; б) для доли: $\Delta_w = t\mu_w = t\sqrt{\dfrac{w(1-w)}{n}}$ – при повторном отборе; $\Delta_w = t\mu_w = t\sqrt{\dfrac{w(1-w)}{n}\left(1-\dfrac{n}{N}\right)}$ – при бесповторном отборе.

Теорема Чебышева-Ляпунова. Доверительный интервал для генеральной средней и генеральной доли. На основании теоремы П.Л. Чебышева (с уточнениями А.М. Ляпунова) с вероятностью, сколь угодно близкой к единице, можно утверждать, что при достаточно большом объеме выборки и ограниченной генеральной дисперсии выборочные обобщающие показатели (средняя, доля) будут сколь угодно мало отличаться от соответствующих генеральных показателей.

Применительно к нахождению среднего значения признака теорема Чебышева-Ляпунова может быть записана так: $P\left(\left|\bar{x}-\tilde{x}\right| \le \Delta_x\right) = \Phi(t)$;

для доли: $P\left(\left|p-w\right| \le \Delta_w\right) = \Phi(t)$.

Предельная ошибка выборки позволяет определить предельные значения характеристик генеральной совокупности и их *доверительные интервалы*:

а) для генеральной средней $\tilde{x} - \Delta_x \le \bar{x} \le \tilde{x} + \Delta_x$;

б) для генеральной доли $w - \Delta_w \le p \le w + \Delta_w$.

Тема 9. Статистическая проверка гипотез

Аннотация. В данной теме даются понятия статистических гипотез, статистического критерия, ошибок первого и второго рода, критической области, приводится методика статистической проверки параметрических гипотез, формулы расчета наблюдаемых значений критериев Фишера-Снедекора, нормального и Стьюдента.

Ключевые слова. Статистический критерий, нулевая гипотеза, конкурирующая гипотеза, критическая область, уровень значимости

Методические рекомендации по изучению темы

- тема содержит лекционную часть, в ходе изучения которой занятия необходимо отмечать наиболее существенную информацию, новые термины и понятия, сравнивать новый материал с изученным и усвоенным ранее, устанавливать их взаимосвязь и укладывать новую информацию в уже имеющуюся систему знаний;
- в дополнение к лекционной части приведены примеры решения практических заданий по рассматриваемой теме;
- в качестве самостоятельной работы предлагается решить контрольные задания;
- для самоконтроля приводятся вопросы к теме и тесты.

Рекомендуемые информационные ресурсы

1. http://bars.kpfu.ru/course/view.php?id=783
2. http://tvims.wordpress.com/
3. http://ru.wikibooks.org/wiki/Математика_случая
4. http://teorver-online.narod.ru/
5. http://statistica.ru/theory/

Список сокращений

МС – математическая статистика
СПГ – статистическая проверка гипотез
СВ – случайная величина

Глоссарий

Конкурирующей (альтернативной) гипотезой называется гипотеза H_1, которая противоречит нулевой гипотезе H_0.

Критическая область – область, при попадании в которую критерия $K_{набл}$ отвергается основная гипотеза.

Критической точкой (границей), называют точку $K_{кр}$, отделяющую критическую область от области принятия гипотез. Для каждого критерия имеются таблицы, по которым находят критическую точку.

Нулевой (основной) гипотезой называется выдвинутая гипотеза H_0.

Областью принятия гипотезы (областью допустимых значений), называют совокупность значений критерия при которой H_0 принимают.

Ошибка первого рода – отвергнуть гипотезу H_0 при её правильности. Вероятность этой ошибки называют *уровнем значимости* проверки гипотезы, и обычно обозначают α.

Ошибка второго рода – принятие ложной гипотезы H_0 при правильности альтернативной гипотезы H_1. Вероятность этой ошибки обычно обозначают β, а величину $1-\beta$ называют *мощностью критерия.*

Статистической называется гипотеза о неизвестном законе распределения случайной величины или о параметрах закона распределения, вид которого известен. В первом случае гипотеза называется *непараметрической*, во втором – *параметрической*.

Вопросы для изучения

1. Виды статистических гипотез. Статистический критерий.
2. Критические точки. Критические области. Уровень значимости.
3. Сравнение двух дисперсий.
4. Сравнение двух средних нормальных генеральных совокупностей (большие и малые независимые выборки).
5. Сравнение генеральной средней со «стандартом» (большие и малые независимые выборки).

Виды статистических гипотез. Статистический критерий. *Статистической* называется гипотеза о неизвестном законе распределения случайной величины или о параметрах закона распределения, вид которого известен. В первом случае гипотеза называется *непараметрической*, во втором – *параметрической*.

Нулевой (основной) гипотезой называется выдвинутая гипотеза H_0.

Конкурирующей (альтернативной) гипотезой называется гипотеза H_1, которая противоречит нулевой гипотезе H_0.

Всегда имеется некоторая величина *К*, которую называют *статистическим критерием* или просто *критерием* проверки гипотезы – это случайная величина с известным распределением. Эту величину обозначают Z, если она распределена нормально, F – по закону Фишера-Снедекора, T – по распределению Стьюдента.

Критические точки. Критические области. Уровень значимости. Область, при попадании в которую критерия $K_{\text{набл}}$ отвергается основная гипотеза, называется *критической областью*.

Основной принцип СПГ: если наблюдаемое значение критерия принадлежит критической области, то гипотезу отвергают; если наблюдаемое значение критерия принадлежит области покрытия гипотезы, то гипотезу принимают.

Критической точкой (границей), называют точку $K_{\text{кр}}$, отделяющую критическую область от области принятия гипотез. Для каждого критерия имеются таблицы, по которым находят критическую точку.

Правосторонней называется критическая область, описываемая неравенством $K > K_{\text{кр}}$. *Левосторонней* называется критическая область, описываемая неравенством $K < K_{\text{кр}}$. *Двусторонней* называется критическая область, описываемая неравенством $|K| > K_{\text{кр}}$. Тип критической области задается знаком альтернативной гипотезы H_1: $<$ – левосторонняя (H_1: $\bar{x}_1 < \bar{x}_2$); $>$ – правосторонняя (H_1: $\bar{x} > a$); $\neq$ – двусторонняя (H_1: $D_1 \neq D_2$). При проверке гипотезы H_0 возможны следующие ошибки: *Ошибка первого рода* – отвергнуть гипотезу H_0 при её правильности. Вероятность этой ошибки называют *уровнем значимости* проверки гипотезы, и обычно обозначают α. *Ошибка второго рода* – принятие ложной гипотезы H_0 при правильности альтернативной гипотезы H_1. Вероятность этой ошибки обычно обозначают β, а величину $1-\beta$ называют *мощностью критерия.*

Сравнение двух дисперсий. При заданном уровне значимости α проверить нулевую гипотезу о равенстве генеральных дисперсий Для проверки гипотезы $H_0: D_1 = D_2$ применяется критерий Фишера-Снедекора (F). При этом

$$F_{набл} = \frac{S_б^2}{S_м^2}, \quad S_i^2 = \frac{n_i}{n_i - 1}\sigma_i^2, \quad (i = 1{,}2).$$

Число степеней свободы: $k_1 = n_1^* - 1,\quad k_2 = n_2^* - 1,$ где n_1^* – объем выборки, по которой вычисляется большая исправленная дисперсия, n_2^* – объем выборки, по которой вычисляется меньшая исправленная дисперсия. По таблице распределения критических точек Фишера-

Снедекора находится $F_{кр}(\alpha;\, k_1;\, k_2)$ при односторонней критической области (возможна только правосторонняя критическая область при конкурирующей гипотезе $H_1: D_1 > D_2$) и $F_{кр}(\alpha/2;\, k_1;\, k_2)$ при двусторонней критической области (при конкурирующей гипотезе $H_1: D_1 \neq D_2$).

Сравнение двух средних нормальных генеральных совокупностей (большие и малые независимые выборки). Для больших независимых выборок $(n_1 > 30, n_2 > 30)$ применяется критерий *Z* нормального распределения. При этом

$$Z_{набл} = \frac{\tilde{x}_1 - \tilde{x}_2}{\sqrt{\dfrac{D_1}{n_1} + \dfrac{D_2}{n_2}}}.$$

Значение $Z_{кр}$ находится из условий: $\Phi(Z_{кр}) = 1 - \alpha$ при двусторонней критической области, $\Phi(Z_{кр}) = 1 - 2\alpha$ при односторонней критической области.

При малых независимых выборках $(n_1 \le 30, n_2 \le 30)$ используется критерий *T* распределения Стьюдента. При этом

$$T = \frac{\tilde{x}_1 - \tilde{x}_2}{\sqrt{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}.$$

Число степеней свободы: $k = n_1 + n_2 - 2$. Значение $T_{кр}(\alpha; k)$ находится по таблице распределения критических точек Стьюдента.

Сравнение генеральной средней со «стандартом» (большие и малые независимые выборки). Если дисперсия генеральной совокупности известна (большая выборка: *n* > 30), применяется критерий *Z* нормального распределения. При этом

$$Z_{набл} = \frac{(\tilde{x} - a)\sqrt{n}}{\sqrt{D}}.$$

Значение $Z_{кр}$ находится из условия: $\Phi(Z_{кр}) = 1 - \alpha$ при двусторонней критической области, $\Phi(Z_{кр}) = 1 - 2\alpha$ при односторонней критической области.

Если дисперсия генеральной совокупности неизвестна (малая выборка: $n \le 30$), то применяется критерий Стьюдента *T*. При этом

$$T_{набл} = \frac{(\tilde{x} - a)\sqrt{n}}{S}.$$

Число степеней свободы: $k = n - 1$. Значение $T_{кр}(\alpha; k)$ находится по таблице распределения критических точек Стьюдента.

Тема 10. Регрессионный анализ

Аннотация. В данной теме даются понятия уравнения регрессии, рассматривается методика построения выравнивающей линии по эмпирическим данным, приводится суть метода наименьших квадратов, а также формулы расчета коэффициентов уравнения регрессии.

Ключевые слова. Факторный признак, результативный признак, уравнение регрессии, метод наименьших квадратов

Методические рекомендации по изучению темы

- тема содержит лекционную часть, в ходе изучения которой занятия необходимо отмечать наиболее существенную информацию, новые термины и понятия, сравнивать новый материал с изученным и усвоенным ранее, устанавливать их взаимосвязь и укладывать новую информацию в уже имеющуюся систему знаний;
- в дополнение к лекционной части приведены примеры решения практических заданий по рассматриваемой теме;
- в качестве самостоятельной работы предлагается решить контрольные задания;
- для самоконтроля приводятся вопросы к теме и тесты.

Рекомендуемые информационные ресурсы

1. http://bars.kpfu.ru/course/view.php?id=783
2. http://tvims.wordpress.com/
3. http://ru.wikibooks.org/wiki/Математика_случая
4. http://teorver-online.narod.ru/
5. http://statistica.ru/theory/

Список сокращений

МС – математическая статистика
МНК – метод наименьших квадратов
СВ – случайная величина

Глоссарий

Корреляционной зависимостью называется статистическая зависимость, при которой значению СВ X ставится в соответствие какая-либо характеристика соответствующего распределения Y.

Результативный признак – зависимая случайная величина Y.

Статистикой называется связь, при которой каждому значению СВ X соответствует распределение другой случайной величины Y.

Уравнение регрессии – функциональная зависимость, устанавливающая зависимость Y от X, т.е. наилучшим образом согласующийся со всеми опытными точками.

Факторный признак – независимая СВ X.

Вопросы для изучения

1. Метод наименьших квадратов .
2. Нахождение уравнения регрессии для линейной и нелинейной зависимостей. Экономические выводы.

Метод наименьших квадратов. Рассматриваются два взаимосвязанных признака X и Y, точная зависимость между которыми заранее неизвестна, причем X – независимая переменная или факторный признак, Y – зависимая переменная или результативный признак. Нужно найти уравнение (функциональную зависимость), устанавливающее закон изменения Y от X, т.е. наилучшим образом согласующийся со всеми опытными точками $(x_i;y_i)$ таблицы значений. Для определения уравнения регрессии применяют метод наименьших квадратов. Суть метода наименьших квадратов состоит в том, что наилучшим приближением (аппроксимацией) искомой функциональной зависимостью будет то, для которого сумма квадратов отклонений ординат эмпирических точек от соответствующих теоретических (выровненных) будет минимальной

Нахождение уравнения регрессии для линейной и нелинейной зависимостей. Экономические выводы. Опытные точки $(x_i;y_i)$ могут располагаться вдоль прямой, гиперболы, параболы, т.е. уравнением, устанавливающим закон изменения результативного признака Y от факторного признака X (уравнением регрессии) будет уравнение прямой

$y = ax + b$ или уравнение гиперболы $y = \frac{a}{x} + b$, или параболы $y = ax^2 + bx + c$. Если опытные точки $A_i(x_i; y_i)$ расположены вдоль некоторой прямой $y = ax + b$, то коэффициенты a, b определяются из следующей системы нормальных уравнений:

$$\begin{cases} \overline{x^2} \cdot a + \bar{x} \cdot b = \overline{xy}, \\ \bar{x} \cdot a + b = \bar{y}, \end{cases}$$

где $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$, по формулам

$$\begin{cases} a = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}, \\ b = \bar{y} - \bar{x} \cdot a. \end{cases}$$

На основе полученного уравнения регрессии делаются прогнозы значений результативного признака.

Тема 11. Корреляционный анализ

Аннотация. В данной теме дается понятие тесноты связи, приводятся формулы расчета показателей тесноты связи, рассматривается методика оценки показателей тесноты связи в генеральной совокупности.

Ключевые слова. Теснота связи, корреляционное отношение, линейный коэффициент корреляции

Методические рекомендации по изучению темы

- тема содержит лекционную часть, в ходе изучения которой занятия необходимо отмечать наиболее существенную информацию, новые термины и понятия, сравнивать новый материал с изученным и усвоенным ранее, устанавливать их взаимосвязь и укладывать новую информацию в уже имеющуюся систему знаний;
- в дополнение к лекционной части приведены примеры решения практических заданий по рассматриваемой теме;
- в качестве самостоятельной работы предлагается решить контрольные задания;
- для самоконтроля приводятся вопросы к теме и тесты.

Рекомендуемые информационные ресурсы

1. http://bars.kpfu.ru/course/view.php?id=783
2. http://tvims.wordpress.com/
3. http://ru.wikibooks.org/wiki/Математика_случая
4. http://teorver-online.narod.ru/
5. http://statistica.ru/theory/

Список сокращений

МС – математическая статистика
КРА – корреляционно-регрессионный анализ

Глоссарий

Корреляционное отношение – универсальный показатель тесноты связи:

$$\eta = \sqrt{\frac{\sigma_y^2 - \sigma_{\overline{y}_x}^2}{\sigma_y^2}}.$$

Линейный коэффициент корреляции – показатель тесноты связи для линейной зависимости:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}.$$

Общая дисперсия признака Y – это дисперсия, характеризующая колеблемость результативного признака Y под воздействием всех факторных признаков:

$$\sigma_y^2 = \overline{y^2} - \bar{y}^2.$$

Остаточная дисперсия признака Y – это дисперсия, характеризующая колеблемость результативного признака Y под воздействием всех факторов, кроме изучаемого признака X, т. е. под воздействием случайных факторов:

$$\sigma_{\overline{y}_x}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y_{x_i}})^2.$$

Теснота связи – оценка рассеяния значений признака *Y* около линии регрессии для различных значений признака *X*.

Вопросы для изучения

1. Определение параметров тесноты связи (линейный коэффициент корреляции, корреляционное отношение).
2. Доверительный интервал для линейного коэффициента корреляции.
3. Проверка гипотезы о значимости выборочного коэффициента корреляции.

Определение параметров тесноты связи (линейный коэффициент корреляции, корреляционное отношение). Корреляционный анализ устанавливает силу или тесноту связи между результативным признаком *Y* и факторным признаком *X*. Причем, корреляционная связь тем теснее, чем меньше рассеяние между *X* и *Y* под влиянием неучтенных факторов, т.е., если влияние фактора *X* мало осложняется действием других факторов, то связь между *X* и *Y* является тесной.

Универсальной мерой тесноты связи является корреляционное отношение η. При парной корреляции $\eta = \sqrt{\dfrac{\sigma_y^2 - \sigma_{y_x}^2}{\sigma_y^2}}$. Корреляционное отношение η принимает значения от 0 до 1, при этом если $\eta \to 1$, то говорят о тесной связи между признаками *X* и *Y*, если же $\eta \to 0$, то факторный признак *X* мало влияет на результативный признак *Y*.

Для линейной зависимости существует дополнительный показатель тесноты связи – линейный коэффициент корреляции $r = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}$, где $\sigma_x^2 = \overline{x^2} - \bar{x}^2$ – общая дисперсия факторного признака *X*. Коэффициент корреляции r принимает значения от -1 до 1. Если r = 1, то имеем возрастающую линейную функциональную зависимость, если r = -1, то убывающую линейную функциональную зависимость, если r = 0, то линейная связь отсутствует, если $r \to 1$, то связь между *X* и *Y* тесная. Близкое совпадение $|r| \approx \eta$ свидетельствует о линейной связи между признаками.

Доверительный интервал для линейного коэффициента корреляции. Для удобства обозначим параметр тесноты связи в генеральной совокупности $\eta_{\text{ген}}$ или $r_{\text{ген}}$ через $\theta_{\text{ген}}$ и соответствующие параметры в выборочной совокупности $\eta_{\text{выб}}$ и $r_{\text{выб}}$ через $\theta_{\text{выб.}}$

Доверительный интервал для параметра тесноты связи в генеральной совокупности находится как:

$$\theta_{выб} - \Delta_{\theta} \le \theta_{ген} \le \theta_{выб} + \Delta_{\theta},$$

где предельная ошибка:

$$\Delta_{\theta} = t \cdot \frac{1-\theta_{выб}^2}{\sqrt{N}}.$$

При этом параметр t определяется из равенства $\gamma = \Phi(t) = 1-\alpha$ и таблицы значений $\Phi(t)$.

Проверка гипотезы о значимости выборочного коэффициента корреляции. Для решения задачи существования корреляционной зависимости в генеральной совокупности выдвигается гипотеза $H_0 : \theta_{ген} = 0$ при конкурирующей гипотезе $H_1 : \theta_{ген} \ne 0$ для заданного уровня значимости α.

В качестве критерия выбирается СВ, имеющая распределение Стьюдента с $k = N$-2 степенями свободы. Наблюдаемая величина вычисляется по формуле:

$$T_{набл} = \frac{\theta_{выб}\sqrt{N-2}}{\sqrt{1-\theta_{выб}^2}}.$$

Значение $T_{кр}(\alpha; k)$ находится по таблице распределения критических точек Стьюдента.