Задание:

- Изучить теорию;
- Написать краткий конспект;
- Разобрать примеры решения;
- Ответить на вопросы;
- По вопросам обращаться 072-1098278 или hvastov@rambler.ru
- Фотоотчёт конспекта прислать в течении 3 дней со дня получения задания на hvastov@rambler.ru

Лекция 11

Тема: Задачи математической статистики. Основные понятия математической статистики.

Цель: Ввести понятие математической статистики

План

1.Предмет и основные задачи математической статистики

2.Основные понятия математической статистики

3.Основыне виды выборок

1.При изучении теории вероятностей мы рассмотрели два различных типа случайностей: случайные события и случайные величины.

Выяснили, что во многих случаях можно теоретическим путем рассчитывать вероятности случайных событий (классическое определение вероятности). В то же время существуют ситуации (и таких, вообще говоря, гораздо больше), когда классическое определение применить невозможно. Тогда единственным путем для определения вероятности случайных событий остается эксперимент, наблюдения.

Замечание: О содержании понятия опыт, эксперимент.

В инженерных науках, в технике, в физике и т. д. это специально создаваемые исследователем условия для появления интересующего нас явления, для измерений интересующей нас величины. Такой эксперимент (опыт) называют активным экспериментом.

В экономике, в социологии и в ряде других наук гораздо чаще приходится добывать интересующую нас информацию, не вмешиваясь активно в условия, а только наблюдая происходящие независимо от исследователя явления. Это иногда называют пассивными экспериментом. Здесь правильнее использовать термин «наблюдение».

Статистика - это наука, изучающая количественные показатели развития общества и общественного производства.

Предметом математической статистики является разработка методов отбора, обработки и анализа опытных данных для получения информации о случайных величинах и их взаимосвязи.

Основные задачи математической статистики.

1. Построение закона распределения по опытным данным.
2. Статистическая оценка параметров распределения.
3. Статистическая проверка статистических гипотез.

Математическая статистика решает также многие другие задачи, но они уже выходят за пределы нашего курса.

2. Математическая статистика изучает методы сбора, обработки и интерпретации результатов опытов (экспериментов).

Генеральной совокупностью называют множество однородных объектов с характерными для них признаками.

Выборочной совокупностью (выборкой) называют подмножество объектов генеральной совокупности, извлеченных из нее случайным образом.

Случайная величина X – наблюдаемые (полученные экспериментально) значения некоторого признака, характерного для всех объектов совокупности.

Статистические ряды

Различают дискретные и интервальные статистические ряды.

Рассмотрим выборку CBX x_1, x_2, …, x_n объема n.

1. Дискретный вариационный ряд распределения случайной величины X имеет вид:

Таблица 1

x_i	x_1	x_2	x_3	x_4	…	x_k
m_i	m_1	m_2	m_3	m_4	…	m_k
ω_i	$\frac{m_1}{n}$	$\frac{m_2}{n}$	$\frac{m_3}{n}$	$\frac{m_4}{n}$	…	$\frac{m_k}{n}$

Чтобы построить дискретный вариационный ряд, необходимо:

1) расположить значения признака X (варианты) в порядке возрастания: $x_1 < x_2 < x_3 < ... < x_k$;

2) найти частоты m_i вариант (количество значений вариант x_i);

3) найти относительные частоты $\omega_i = \dfrac{m_i}{n}$ вариант.

2. Интервальный вариационный ряд распределения случайной величины X имеет вид:

Таблица 2

$x_i - x_{i+1}$	$[x_1; x_2)$	$[x_2; x_3)$	$[x_3; x_4)$	...	$[x_{k-1}; x_k]$
m_i	m_1	m_2	m_3	...	m_k
ω_i	$\frac{m_1}{n}$	$\frac{m_2}{n}$	$\frac{m_3}{n}$	...	$\frac{m_k}{n}$

Чтобы построить интервальный вариационный ряд, необходимо

1) найти размах вариации: $R = x_{max} - x_{min}$, где $x_1 = x_{min}$, $x_k = x_{max}$;

2) определить количество интервалов: k;

3) найти длину интервала: $h = \frac{R}{k}$;

4) найти частоты вариант на интервалах: m_i;

5) найти относительные частоты вариант на интервалах: $\omega_i = \frac{m_i}{n}$

Функция распределения выборки

Эмпирической функцией распределения выборки, представленной в таблице 1, называют функцию вида:

$$F_n(x) = \sum_{i:x_i' < x} \frac{m_i}{n} = \begin{cases} 0, \ x \le x_1; \\ \omega_1, \ x_1 < x \le x_2; \\ \omega_1 + \omega_2, \ x_2 < x \le x_3; \\ \dots \\ \sum_{i=1}^{k-1} \omega_1, \ x_{k-1} < x \le x_k; \\ 1, \ x > x_k. \end{cases} \quad (11.1)$$

Графическое представление выборки

1. Полигоном частот выборки (таблица 1) называют ломаную линию, соединяющую на координатной плоскости точки вида $(x_i; m_i)$, $i = \overline{1, k}$, а полигоном относительных частот – ломаную линию, соединяющую на координатной плоскости точки вида $(x_i; \omega_i)$, $i = \overline{1, k}$.

2. Гистограммой частот выборки (таблица 2) называют столбчатую диаграмму, состоящую из прямоугольников, основаниями которых являются длины интервалов, которые содержат значения вариант, высотами – частоты данных

интервалов, а гистограммой относительных частот – диаграмму, состоящую из прямоугольников, основаниями которых являются длины интервалов, которые содержат значения вариант, высотами – относительные частоты данных интервалов.

3. В статистике применяется несколько видов выборки. Вид выборочного наблюдения определяется способом отбора. Из генеральной совокупности можно отбирать единицы в индивидуальном порядке. При индивидуальном отборе выборочная совокупность образуется путем последовательного отбора отдельных единиц. Индивидуальный отбор организуется в порядке случайного отбора, типического и механического. Случайным отбором принято называть отбор единиц, проводимый в случайном порядке. Выборка, проводимая в порядке индивидуального случайного отбора из генеральной совокупности, принято называть случайной выборкой.

Отбор единиц из генеральной совокупности, разбитой на однородные типические группы, называются типическим отбором, а выборка, основанная на таком отборе – типической выборкой Отбор единиц из генеральной совокупности может производиться механически через определённый интервал, а выборка в таком случае носит название механической.

Вместе с индивидуальным отбором в статистике имеет место серийный (гнездовой) отбор, когда из генеральной совокупности для выборочного исследования отбираются не отдельные единицы, а целые группы.

Существуют повторный и бесповторный отборы. Отбор принято называть повторным, если единица или серия, попавшая в выборку, при одном извлечении из жребия не устанавливается из дальнейшей жеребьевки, т.е. каждый раз жеребьевка производиться из всей массы генеральной совокупности. При таком отборе каждая единица может попасть в выборку несколько раз. Повторный отбор еще называют возвратным.

Бесповторным называют отбор, при котором отобранная и зарегистрированная один раз единица из дальнейшего отбора устраняется.

Повторный и бесповторный отборы производят при проведении случайной, типической и серийной выборок. При механической выборке производят только бесповторный отбор.

Рассмотрим более подробно каждый вид выборки.

Случайная выборка – отбор единиц производиться случайным образом в порядке жеребьевки. Оценка точности выборки осуществляется по формулам, приведенным в таблице.

Формулы ошибок механической выборки (бесповторный отбор):

-при определении среднего значения признака: $\Delta \chi = t\sqrt{(\frac{\delta^2}{n})\cdot(1-\frac{n}{N})}$;

- при определении доли: $\Delta w = t\sqrt{\frac{w(1-w)}{n}\cdot\frac{1-n}{N}}$.

Таблица 34

Формулы предельной ошибки случайной выборки

Способ отбора	Предельная ошибка выборки при определении	
среднего значения признака	доли	
Повторный, если, известны δ^2 и р	$\Delta x = t\sqrt{\frac{\delta^2}{n}}$	$\Delta p = t\sqrt{\frac{pq}{n}}$
Повторный, если, известны δ^2 и W	$\Delta x = t\sqrt{\frac{\delta_0^2}{n}}$	$\Delta w = t\sqrt{\frac{w(1-w)}{n}}$
Бесповторный, в случае если известны δ^2 и р	$\Delta x = t\sqrt{(\frac{\delta^2}{n}) \cdot (1 - \frac{n}{N})}$	$\Delta p = t\sqrt{(\frac{pq}{n}) \cdot (1 - \frac{n}{N})}$
Бесповторный, в случае если известны δ^2 и W	$\Delta x = t\sqrt{(\frac{\delta_0^2}{n}) \cdot (1 - \frac{n}{N})}$	$\Delta w = t\sqrt{\frac{w(1-w)}{n} \cdot \frac{1-n}{N}}$

Типическая выборка - генеральная совокупность разбивается на однородные типические группы по какому – либо признаку, а затем из каждой типической группы производится отбор единиц в порядке случайной или механической выборки. При этом если число единиц, которое должно попасть в выборку от каждой типической группы, определяется по численности единиц в каждой группе, то такая выборка принято называть пропорциональной.

Так, в случае если число единиц генеральной совокупности – N, число единиц в каждой типической группе – соответственно $N_1, N_2 \ldots N_n$, объём выборочной совокупности – n, то число единиц, попавших в выборку от каждой типической группы – $n_1, n_2, \ldots n_n$, - определяется по формуле

$$n_x = \frac{n \cdot N_x}{N} = \frac{n}{N} \cdot N_x,$$

где $\frac{N_x}{N}$ - удельный вес каждой типической группы в генеральной совокупности; $\frac{n}{N}$ - пропорция отбора.

Типическая выборка дает более точные результаты по сравнению со случайной или механической вследствие того, что она обеспечивает представительство выборочной совокупности различных типов единиц, имеющихся в генеральной совокупности. Оценка точности типической выборки осуществляется по формулам, приведенным таблице.

Серийная выборка в отличие от других видов предполагает отбор единиц сериями или гнездами. Серии состоят из единиц, связанных между собой территориально (к примеру, населённый пункт, район, область, и т.п.) или во времени (к примеру, производство продукции за данный период времени).

Серии отбирают в случайном порядке или механически. Точность серийной выборки зависит от того, насколько хорошо средние показатели серии будут представлять генеральную совокупность. Оценка точности серийной выборки осуществляется по формулам, приведенным таблице.

Таблица 35

Формулы предельной ошибки типической выборки

Способ отбора	Предельная ошибка выборки при определении	
среднего значения признака	доли	
Повторный	$\Delta\chi = t\sqrt{\frac{\bar{\delta}_i^2}{n}}$	$\Delta w = t\sqrt{\frac{w_i(1-w_i)}{n}}$
Бесповторный	$\Delta\chi = t\sqrt{(\frac{\bar{\delta}_i^2}{n})\cdot(1-\frac{n}{N})}$	$\Delta w = t\sqrt{\frac{w_i(1-w_i)}{n}\cdot\frac{1-n}{N}}$

Введем обозначения:

r – число серий выборочной совокупности;

R – число генеральной совокупности;

φ^2 - межгрупповая (меж серийная) дисперсия.

Таблица 36

Формулы предельной ошибки серийной выборки

Способ отбора	Предельная ошибка при определении	
среднего значения признака	Доли	
Повторный	$\Delta x = t\sqrt{\frac{\varphi_x^2}{r}}$	$\Delta p = t\sqrt{\frac{\delta_p^2}{r}}$
Бесповторный	$\Delta x = t\sqrt{\frac{\varphi_x^2}{r}\cdot(1-\frac{r}{R})}$	$\Delta w = t\sqrt{\frac{\delta_p^2}{r}\cdot\frac{1-r}{R}}$

При планировании выборочного наблюдения необходима численность выборки. Она определяется по формулам, приведенным в таблице 37.

Таблица 37

Формулы определения крайне важной численности выборки

Способ отбора	Предельная ошибка выборки при определении	
среднего значения признака	доли	
Повторный	$n = \frac{t^2\delta^2}{\Delta_x^2}$	$n = \frac{t^2 pq}{\Delta_p^2}$
Бесповторный	$n = \frac{t^2\delta^2 N}{(\Delta_x^2 N + t^2\delta^2)}$	$n = \frac{t^2 pq\ N}{(\Delta_p^2 N + t^2 pq)}$

Расчет по данным формулам нередко затрудняется из–за отсутствия значения генеральной совокупности. В этом случае используют данные пробного выборочного наблюдения, на базе которого определяется приближенные размеры дисперсии.

Контрольные вопросы

1.Что изучает математическая статистика?

2.Привести пример статических данных

3.Что называется генеральной совокупностью?

4.Что такое выборочная совокупность?

5.Какие основные виды выборок вы знаете?