05.06.20.

Представление данных. Генеральная совокупность, выборка. Понятие о задачах математической статистики.

Видео для усвоения материала:
https://youtu.be/lvJyQJZfgRk
https://youtu.be/yhP_zD7rA_k

Понятие о математической статистике.

Статистикой называется наука о сборе, классификации, обработке и анализе всевозможных качественных и количественных данных, о получении фактов из обобщающих выводов. Одна из задач статистики – сделать имеющуюся информацию наглядной. Здесь помощь оказывают как математические приёмы, так и диаграммы, таблицы, графики. Например, если вы изучаете данные о заболеваемости 1000 человек, то статистика рекомендует разбить эту тысячу на возрастные группы: 100 человек моложе 20 лет, 170 – в возрасте от 20 до 25 лет и т.д. По этим данным можно составить таблицу, а по таблице построить гистограмму возраста и график заболеваемости. В настоящее время использование ЭВМ значительно облегчает составление отчётов, построение таблиц, графиков и других документов. Но применение специальных компьютерных программ мало помогает при выборе методов обработки и анализе результатов. Здесь исследователям помогает изучение методов, разработанных статистикой как наукой. В статистике применяют два основных подхода: метод сплошных наблюдений (описательная статистика) и выборочный метод. Метод сплошных наблюдений предполагает изучение всех элементов совокупности. Он применяется, если надо изучить успеваемость в группе или на факультете, работу предприятия и его филиалов и т.д., когда количество изучаемых объектов не слишком велико. Когда количество объектов велико или сплошное обследование невозможно в силу того, что обследование может привести к уничтожению объекта (например, чтобы узнать качество консервов, банку надо вскрыть), то есть когда не хотят проводить полное обследование объекта, пользуются выборочным методом, при котором из общей совокупности выбирают ограниченное число объектов и их подвергают изучению. Тогда возникает вопрос, насколько результаты такого обследования будут справедливы для всей совокупности. Разрешить этот вопрос помогает математическая статистика. Математическая статистика – раздел математики, посвящённый математическим методам систематизации, обработки и использования статистических данных для научных и практических выводов. При этом статистическими данными называются сведения о числе объектов в какой-либо более или менее обширной совокупности, обладающих теми или иными признаками.

Историческая справка

Математическая статистика возникла в XVII веке и создавалась параллельно с теорией вероятностей. В России методы математической статистики применялись к демографии и страховому делу В.Я. Буняковским ещё в середине прошлого века. В СССР значительные результаты в области математической статистики получены В.И. Романовским, А.Н. Колмогоровым, Е.Е. Слуцким, Н.В. Смирновым, Ю.В. Линником. Большой вклад в математическую статистику внесли английские (Стьюдент, Р. Фишер, Э. Пирсон), а также и американские (Ю. Нейман, А. Вальд) учёные. Решающее значение для математической статистики имели работы русской классической школы теории вероятностей 2-й половины XIX – начала XX века (П.Л. Чебышев, А.А. Марков, А.М. Ляпунов, С.Н. Бернштейн) и работы немецких и английских математиков (К.Ф. Гаусс, К. Пирсон, Ф. Гальтон). Нельзя не отметить вклад советских математиков в составление таблиц функций (Е.Е. Слуцкий, Н.В. Смирнов, Л.Н. Большев).

Математическая статистика для решения своих задач активно привлекает теорию вероятностей. Но, в отличие от теории вероятностей, которая занимается исчислением вероятностей, когда из каких-либо соображений распределение вероятностей известно, статистика решает обратную задачу: отыскивание вероятностных характеристик случайных величин по наблюдаемым реализациям и частотам их появления. Ответ на поставленный выше вопрос о том, насколько результаты выборочного обследования будут справедливы для всей совокупности, математическая статистика формулирует в вероятностных терминах, вводя понятие «уровень доверия» – вероятность, с которой мы не ошибёмся, если поверим выводам, сделанным на основе анализа выборки. Применение вероятностного подхода к изучению выборки очень естественно: данные выборки, не охватывая всей совокупности целиком, являются случайными значениями. По результатам эксперимента необходимо сделать вывод о некотором признаке объекта, когда закон распределения получаемой в результате эксперимента случайной величины (или его параметры), вообще говоря, неизвестен, а эксперимент проводится именно для получения информации о законе распределения случайной величины (что и позволяет сделать вывод о признаке объекта). В вероятностных терминах эксперимент состоит в проведении п независимых испытаний над некоторой случайной величиной; по результатам испытаний надо сделать вывод относительно параметров распределения этой случайной величины (или связанной с ней). При этом изучаемый признак объекта может быть как количественным, так и качественным. Например, если объект – это совокупность всех выпущенных ампул с новокаином, то количественным признаком может быть завод изготовитель, а качественным – бездефектность. На языке теории вероятностей: завод изготовитель – непрерывная случайная величина (случайная величина, возможные значения которой заполняют некоторый интервал), бездефектность – дискретная (случайная величина, принимающая конечное или счётное число значений). В частности, бездефектность – случайная величина, принимающая значение 1 для ампул без дефекта и 0 – для ампул с дефектом. С помощью эксперимента получаются данные о распределении вероятностей случайной величины – например, о проценте бракованных ампул в партии. Методы математической статистики используются как при исследовании непрерывных случайных величин, так и при вычислении их вероятностей

Выборочный метод

Совокупность всех исследуемых объектов называется генеральной совокупностью. Выборкой называется, отобранное некоторым образом из генеральной совокупности число объектов необходимых для исследования и соответствующих определённым признакам. Количество объектов называется объёмом выборки. Например, если из 1000 студентов для контрольной флюорографии отобрано 100, то все студенты – это генеральная совокупность и её объём равен 1000, а отобранные – это выборка с объёмом равным 100. Разность между наибольшим и наименьшим значением числовой выборки называется размахом выборки. Для исследования все данные числовой выборки систематизируются от min до max значения. После этого выборка представляет собой неубывающую последовательность чисел, которая называется вариационным рядом. Выборка и вариационный ряд несут одну и ту же информацию, но с вариационным рядом легче работать в силу его упорядоченности. Если изучается величина, имеющая непрерывное распределение вероятностей, то, скорее всего, вариационный ряд не будет содержать повторяющихся значений. Если же изучается дискретная случайная величина, то при достаточно большом объёме данных в выборке будут повторяющиеся значения. Число повторений значения п1, п2, …, пk называется частотой значений выборки. Относительной (эмпирической) частотой значения называется п1/п, п2/п, …,пk – отношение числа встречаемости значения к общему объёму выборки. Сумма частот значений выборки

Точечные оценки.

Одной из основных задач математической статистики является нахождение приближённого значения некоторого неизвестного параметра а случайной величины х по выборке её значений, полученной в результате п-измерений (х1, х2, х3, …, хп). Таким параметром может быть, например, математическое ожидание случайной величины или её дисперсия. Приближённое значение параметра а, вычисленное каким-либо способом по значению выборки в статистике называют точечной оценкой этого параметра и обозначают an .

Статистические таблицы.

На основе собранных статистических данных составляются статистические таблицы. Таблицы бывают 3-х видов: простые таблицы, групповые таблицы и комбинационных таблицы.

Каждая таблица имеет:

а) перечень характеризуемых в таблице объектов;

б) перечень характеризующих объект признаков.

Таблица, которая содержит перечень объектов, и кроме того имеет итоговую графу, носит название простой таблицы.

Пример.

Число стационарных учреждений и коек на 10 тыс. жителей, по профилю стационара в России, 1994 год.

	ЧИСЛО БОЛЬНИЦ	ЧИСЛО КОЕК НА 10 ТЫС ЖИТЕЛЕЙ
Центральные районные больницы	1778	27,8
Районные больницы	299	2,1
Участковые больницы	4522	9,0
……		
Все стационарн.учрежд.	12272	127,6

В групповой таблице объекты характеризуются по нескольким группам признаков.

Пример.

Сведения о родившихся по возрасту матери и по очерёдности рождения (в % к общему числу родившихся) по Хабаровскому краю.

ГОД	ПО ВОЗРАСТУ МАТЕРИ					ПО ОЧЕРЕДНОСТИ РОЖДЕНИЯ					
	до 20 лет	20 – 29	30 – 39	40 – 49	Всего	1	2	3	4	≥ 5	Всего
1992	21,1	59,2	18,2	1,34	99,84	55,6	28,8	8,9	3,4	3,05	99,75
1993	21,4	60,8	16,5	1,23	99,93	58,7	28,5	8,0	2,6	2,05	99,85
1994	21,0	62,0	15,9	1,01	99,91	59,1	28,7	7,8	2,4	1,93	99,93
1995	20,0	62,9	16,1	0,96	99,96	60,1	27,8	7,6	2,5	1,94	99,94
1996	18,8	64,3	16,0	0,63	99,73	60,1	27,8	7,7	2,4	1,83	99,83

Сравнение объектов по одинаковым признакам осуществляется с помощью комбинационной таблицы.

Пример.

Средние показатели работы службы анестезиологии в г. Хабаровске и Хабаровском крае в 1993 – 1995 гг.

ВИД АНЕСТЕЗИИ	КОЛИЧЕСТВО АНЕСТЕЗИЙ					
	по городу			по краю и городу		
	1993	1994	1995	1993	1994	1995
Ингаляционный наркоз	1034	999	736	3821	4846	3431
Эндотрахеальный наркоз	4590	5365	4911	11669	12665	12639
Внутривенный наркоз	19836	25748	25200	44646	46177	45507
Проводниковая анестезия	290	331	266	1039	1950	1756

Практическая часть.

Написать кроткое – ½ - 2/3 страницы эссе о том, как вы поняли задачи математической статистики.
Тест можно написать от руки, либо набрать его на компьютере и прислать преподавателю.