ЛАБОРАТОРНАЯ РАБОТА № 3. ПОНЯТИЕ ВЕРОЯТНОСТЬ И ПРЕДСТАВЛЕНИЕ ВЕРОЯТНОСТИ В ЭВМ

Вероятность -- это математическая теория, которая дает описание экспериментов со случайными исходами, обладающих свойством статистической устойчивости. Теория вероятностей строится как аксиоматическая теория, то есть в ее основу положена система аксиом. В свою очередь аксиомы сформулированы на основе экспериментальных данных, а именно на свойствах частоты и, в частности, на факте статистической устойчивости, состоящем в тенденции частоты $\nu(A)$ появления события A стать постоянной и равной некоторому числу $P(A)$ при большом числе повторений N эксперимента G.

Таким образом, при построении теории необходимо ввести число $P(A)$ называемое вероятностью события A, что реализуется с помощью одной из аксиом, которая называется аксиомой существования вероятности. Далее необходимо рассмотреть основные свойства частот и выразить эти свойства как утверждения относительно свойств вероятностей. Эти утверждения вместе с постулатом существования вероятности образуют систему аксиом теории вероятностей.

Частоту $\nu(A)$ можно рассматривать как результат измерения (оценивания) вероятности $P(A)$ по экспериментальным данным. Таким образом, равенство $P(A) = p$ означает, что при большом числе N опытов $\nu(A) \approx p$, а ошибка $|\nu(A) - p|$ имеет тенденцию снижаться с увеличением N. Поскольку $0 \le n \le N$, то частота $\nu(A) = \dfrac{n}{N}$ появления события A в серии из N опытов удовлетворяет условию $0 \le \nu(A) \le 1$. Аналогичному условию должна удовлетворять и вероятность: $0 \le P(A) \le 1$.

Рассмотрим значения вероятности на границах интервала $|0,1|$.Пусть $P(A) = 0$, тогда событие A называется невозможным и обозначается символом $\varnothing$. Для невозможного события его частота $\nu(A) \approx 0$ и имеет тенденцию приближаться к нулю с увеличением числа N опытов. Если $P(A) = 1$, то событие A называется достоверным и обозначается символом E. Частота достоверного события $\nu(E) \approx 1$ и с увеличением числа N опытов имеет тенденцию приближаться к единице.

М

Случайная (вероятностная) выборка — это такая выборка, для которой каждый элемент генеральной

совокупности имеет определенную, заранее заданную вероятность быть отобранным. Это позволяет исследователю рассчитать, насколько правильно выборка отражает генеральную совокупность, из которой она выделена (спроектирована). Такую выборку иногда называют еще случайной.

Если генеральная совокупность имеет небольшой объем, единицы отбора удается пронумеровать, и каждая из них получает равную вероятность попасть в выборку, то применяется вероятностный тип выборки, основанный на использовании таблицы случайных чисел.

Вероятностные методы включают:

- простой случайный отбор;
- систематический отбор;
- кластерный отбор;
- стратифицированный отбор.

Реализовать случайную выборку можно двумя приемами: лотерейный метод и с помощью (таблицы)

случайных чисел. С помощью случайной выборки строится подавляющее большинство телефонных опросов и опросов на основе избирательных списков. Для построения такой выборки необходимо иметь полный список всех элементов генеральной совокупности. Простой случайный отбор предполагает, что вероятность быть включенным в выборку известна и является одинаковой для всех единиц совокупности. Он реализуется двумя методами:

- Отбор вслепую (другое название — метод лотереи или жребия).
- Отбор не вслепую (происходит с помощью таблицы случайных чисел).

Итак, в одном случае вы осуществляете свой выбор не глядя, в другом — все осознавая, но для того, чтобы самому не вмешаться и ничего не испортить, обращаетесь к специальным таблицам. Но кроме того, простой случайный отбор подразделяется на две разновидности уже по другому критерию, а именно — возвращению или невозвращению лотерейного шара (вместо него может быть фамилия

респондента) обратно в корзину. В этом случае выделяют:

♦ Случайный повторный (с возвращением) отбор.

♦ Случайный бесповторный (без возвращения) отбор.

В чем сходство и различие двух классификаций? В первом случае — вслепую/не вслепую — ученый мог смотреть на то, как осуществляется отбор, хотя никак не мог ему помешать. И во втором случае он ничего не мог поделать, но выбор осуществляли не его руки, вынимающие из корзины шар, а таблица случайных чисел. Во втором случае — повторный/бесповторный — дело заключается не в исследователе, а в лотерейном шаре: его либо возвращают для нового выбора, либо не возвращают и продолжают процесс без него.

При использовании метода выборки вслепую единицы совокупности в соответствии с их фамилиями,

названиями или другими признаками можно вносить в карточки, которые в перемешанном виде будут помещены в какую-то непрозрачную емкость (ящик, коробку). Из этой емкости кто-то случайным образом вытягивает число карточек, определяемое объемом выборки. Карточки можно возвращать или не возвращать. В первом случае говорят о повторном, во втором — о бесповторном отборе. Их комбинация дает два квадрата, имеющих реальное содержание: можно вслепую выбирать из корзины шары и возвращать их для нового выбора, а можно этого не делать. Однако выборка не вслепую предполагает использование таблицы случайных чисел. Возвращать в нее выбранный номер невозможно, стало быть образуемые вдоль этой оси квадраты не являются реальными. Предлагаемая схема выполняет скорее мнемоническую форму, помогая лучше запомнить материал. Можно также считать, что она имеет демонстративный смысл, но никак не логический. Она придумана для того, чтобы внести какую-то ясность в типологию разновидностей простого случайного отбора.

Первое условие осуществления вероятностной выборки — наличие полного списка всех элементов генеральной совокупности (отсутствие или недоступность которого чаще всего и препятствует ее

реализации) от 1 до N, где N — общее число всех элементов. Если же он имеется, то производится нумерация, после чего можно использовать несколько методик. При использовании лотерейного метода (или метода жребия) жетоны с номерами всех элементов помещают в урну, тщательно перемешивают и извлекают последовательно n жетонов, где n — число элементов выборочной совокупности. Элементы генеральной совокупности, имеющие номера, оказавшиеся на извлеченных

жетонах, будут составлять выборочную совокупность. Это довольно трудоемкая и продолжительная (при больших размерах выборки) операция, к тому же достаточно трудоемкая, поскольку «для обеспечения равного шанса выбора требуется тщательное перемешивание жетонов»1 после каждой выемки очередного номера.

Второе условие вероятностной выборки — хорошая перемешанность элементов генеральной совокупности. Если выборка элементов производится из ящика, то его содержимое следует тщательно

перемешать и уже после этого брать карточки случайным образом. Только при таких условиях все они имеют одинаковую вероятность попасть в выборку. Часто для образования случайной выборки элементы генеральной совокупности предварительно номеруются, а каждый номер записывается на отдельной карточке. В результате получается пачка карточек, число которых совпадает с объемом генеральной совокупности. После тщательного перемешивания из этой пачки берут по одной карточке. Объект (респондент), имеющий одинаковый номер с карточкой считается попавшим в выборку. При этом возможны два принципиально различных способа образования выборочной

совокупности.

Первый способ — вынутая карточка после фиксации ее номера возвращается в пачку, после чего

карточки снова тщательно перемешиваются. Повторяя такие выборки по одной карточке, можно образовать выборочную совокупность любого объема. Выборочная совокупность, образованная по такой схеме, получила название случайной возвратной выборки.

Второй способ — каждая вынутая карточка после ее записи обратно не возвращается. Повторяя по

такой схеме выборки по одной карточке, можно получить выборочную совокупность любого заданного объема. Выборочную совокупность, образованную по данной схеме, называют случайной безвозвратной выборкой. Она возможна лишь в том случае, если из тщательно перемешанной пачки сразу берут нужное число карточек. Заметим, что различие между случайными выборками с возвратом и без возврата стирается, если они составляют незначительную часть большой генеральной совокупности. Однако при большом объеме генеральной совокупности возвратный и безвозвратный методы оказываются очень трудоемкими. В этом случае пользуются таблицами случайных чисел. Она доказала свою эффективность при формировании равновероятностной выборки из больших совокупностей.

В таблицах случайных чисел содержатся числа, порядок включения которых в таблицу осуществлен случайным образом. Единицам совокупности присваивают порядковые номера. В этой таблице выбирают любую начальную точку и, двигаясь в произвольном направлении и произвольно меняя направление движения, выбирают необходимое количество номеров из числа присвоенных, равное заранее установленному объему выборки. Если мы имеем, скажем, популяцию (т. е. генеральную совокупность) из 1507 элементов и хотим спроектировать выборку из 150, мы можем выбирать любые

четыре смежных столбца в таблице случайных чисел. Каждый раз, когда будет появляться число от 0001 до 1507, мы будем считать, что оно обозначает номер отбираемого элемента. Если число появляется более чем один раз, этот номер игнорируется после первого раза. Если мы начнем с первых четырех столбцов в табл. 1, спускаясь по столбцам, то в выборку будут включены элементы под номерами 0799, 1016, 0084, 480 и 1306. Поскольку мы не стремимся умышленно отыскать определенное число, мы можем начать с любого места таблицы и использовать любую систему для движения по таблице. Сегодня таблицу случайных чисел могут заменить машинные устройства, например, компьютер, снабженный специальной программой. Их называют генераторами случайных чисел. Например, при телефонном интервьюировании компьютер, имеющий генератор случайных чисел, может подавать на экран случайным образом отобранные телефонные номера. Систематический отбор является вторым по научной значимости, но первым по популярности

употребления видом простого случайного отбора. Его называют еще механическим отбором и считают упрощенным вариантом простого случайного отбора.

Систематическая выборка (отбор) — процедура отбора каждого к-го элемента из списка элементов

генеральной совокупности. Примером служат разного рода квартирные выборки: выбираются улицы, на которых интервьюер проводит квартирный опрос. Квартиры выбираются по определенной схеме (крайняя квартира справа от лестницы на последнем этаже первого подъезда и т. д.). Если под рукой таблицы случайных чисел нет, а генсовокупность относительно невелика, то можно воспользоваться алфавитным списком, например, персонала предприятия (картотека всегда есть в отделе кадров) или избирательного участка (при опросе по месту жительства). Процедура систематического отбора проста: количество единиц генеральной совокупности, предположим 2000 работников предприятия, делится на количество анкет, скажем 200, и определяется шаг выборки. Он предполагает, что, начиная с любого номера из списка, опрашивается каждый десятый (2000:200= 10). В формализованном виде данная процедура выглядит так: из пронумерованного списка через равные интервалы к отбирается заданное число респондентов. При этом шаг выборки к рассчитывается по простой формуле:

п

Где N — численность генеральной совокупности,

п — численность выборочной совокупности.

Таким образом, шаг выборки, а его еще называют «интервалом скачка» или просто «интервалом», — это математический показатель, рассчитанный как отношение объема генеральной совокупности к объему выборки. Он показывает, сколько номеров в списке фамилий людей, вошедших в генеральную совокупность, надо пропустить (через сколько перешагнуть), чтобы в итоге получить список выборочной совокупности. Буквально шаг выборки означает расстояние между соседними фамилиями респондентов, измеренное количеством отбракованных фамилий из списка генеральной

Совокупности.

Итак, в основу систематической выборки положены не вероятностные процедуры, а алфавитные списки, картотеки, схемы, которые обеспечивают равновероятное попадание в выборку всех единиц генеральной совокупности.

Несмотря на свои преимущества, систематическая выборка может иногда иметь своим результатом предубежденную выборку. Такая ситуация возникает, например, когда элементы размещены в списке,

ранжированном по каким-то характеристикам. В этой ситуации определение места начала случайного отбора будет влиять на средние характеристики всей выборки. Например, если студенты расставлены в списке в соответствии со средним оценочным баллом от высшего к низшему, систематическая выборка, включающая студентов, стоящих в списке под номерами 1, 51, 101 будет иметь более низкий средний балл, чем выборка, включающая студентов под номерами 50, 100 и 150. Каждая новая

выборка будет давать другой средний балл, который представляет собой предубежденную картину студенческой популяции.

В ЖИВОТНОВОДСТВЕ

Специалисты животноводства постоянно сталкиваются с вопросами выбора: той или иной породы для разведения в условиях хозяйства, типа и уровня кормления, лекарства, дающего наилучший эффект при лечении и т. д. Для решения этих вопросов обычно проводят опыт, то есть формируют две сопоставимые группы животных (опытную и контрольную). На опытной проверяют изучаемый рацион кормления, новый ветеринарный препарат, тип скрещивания и т.д. По учтенным результатам опыта проводят сравнение с контрольной группой. Но поскольку каждая выборочная величина отражает в генеральной совокупности эту же величину с ошибкой, возникает вопрос, не является ли случайной и разница между ними?

Решить вопрос, случайна разница или достоверна (устойчива, постоянна) можно, вычислив критерий достоверности разницы - td по формуле:

$$td = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$ где td - показатель достоверности разницы,

M - средние значения признаков в

сравниваемых группах,

m - статистические ошибки средних.

При этом для большой выборки разность считается достоверной при $t = 2$. Величина критерия достоверности разницы td связана с величиной вероятности P. Обычно разница считается достоверной при уровне вероятности $P > 0{,}95$. Это самый низкий уровень вероятности. В биологии применяют три уровня вероятности: первый $P > 0{,}95$, второй $P > 0{,}99$ и третий, самый высокий, -

$P > 0{,}999$. Уровень вероятности $P > 0{,}99$ указывает на то, что из 100 случаев наблюдений изучаемая закономерность будет проявляться в 99. Уровень вероятности в каждом конкретном случае зависит от величины критерия достоверности разницы и числа степеней свободы (ν). При сравнении двух групп степени свободы определяются по формуле:

$\nu = n_1 + n_2 - 2$ где, ν - степени свободы,

n_1 и n_2 - число наблюдений в группах.

В каждом конкретном случае достоверна разница, и при каком уровне вероятности она достоверна решают с помощью таблицы Стъюдента(табл. 8). В этой таблице приводятся минимальные значения критерия достоверности разницы для разного числа степеней свободы.

Для примера определим, достоверна ли разница между крупным рогатым скотом и овцами по содержанию эритроцитов в крови. Для этого определим содержание эритроцитов у 10 голов овец и 10 голов крупного рогатого скота. В результате подсчетов мы получим:

Крупный рогатый скот M = 11,0 млн/мм^3; m = 0,3 млн/мм^3

Овцы M = 7,3 млн/мм^3; m = 0,2 млн/мм^3

Вычислим критерий достоверности разницы:

$$td = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} = \frac{11{,}0 - 7{,}3}{\sqrt{0{,}3^2 + 0{,}2^2}} = 10{,}3;$$

Число степеней свободы в данном случае будет:

$\nu = n_1 + n_2 - 2 = 10 + 10 - 2 = 18$

8. Критические значения t-критерия Стъюдента для трех уровней вероятности

Степени свободы (ν)	Уровни вероятности (Р)			Степени свободы (ν)	Уровни вероятности (Р)		
	0,95	0,99	0,999		0,95	0,99	0,999
1	12,71	63,66	–	18	2,10	2,88	3,92
2	4,30	9,92	31,60	19	2,09	2,86	3,88
3	3,18	5,84	12,92	20	2,09	2,85	3,85
4	2,78	4,60	8,61	21	2,08	2,83	3,82
5	2,57	4,03	6,87	22	2,07	2,82	3,79
6	2,45	3,71	5,96	23	2,07	2,81	3,77
7	2,37	3,50	5,41	24	2,06	2,80	3,75
8	2,31	3,36	5,04	25	2,06	2,79	3,73
9	2,26	3,25	4,78	26	2,06	2,78	3,71
10	2,23	3,17	4,59	27	2,05	2,77	3,69
11	2,20	3,11	4,44	28	2,05	2,76	3,67
12	2,18	3,05	4,32	29	2,05	2,76	3,66
13	2,16	3,01	4,22	30	2,04	2,75	3,65
14	2,14	2,98	4,14	40	2,02	2,70	3,55
15	2,13	2,95	4,07	60	2,00	2,66	3,46
16	2,12	2,92	4,02	120	1,98	2,62	3,37
17	2,11	2,90	3,97		1,96	2,58	3,29

По таблице находим, что для 18 степеней свободы минимальное значение критерия достоверности разницы 2,10 при уровне вероятности Р > 0,95 и 3,92 при уровне вероятности Р > 0,999. В нашем случае критерий достоверности равен 10,3, что значительно превышает минимальное значение для самого высокого уровня вероятности. Таким образом, мы можем сделать заключение, что содержание эритроцитов в крови овец выше, чем у крупного рогатого скота на 3,7 млн. Эта разница высокодостоверна при Р > 0,999.

1. Определить, достоверно ли различаются чистопородные симменталь-ские коровы и их помеси с красно-пестрыми голштинами по основным показателям продуктивности?

Симменталы Помеси

Живая Удой за Жирность Живая Удой за Жирность

масса,кглактац.кг молока,% масса,кглактац.кг молока,%

575 2110 3,59 510 2050 3,86

520 2500 3,64 525 2660 3,97

545 3050 3,77 465 3090 3,85

530 3560 3,83 540 4130 4,15

565 2000 3,65 490 4500 3,80

525 3480 3,70 485 5800 3,75

500 4010 3,78 510 3340 4,01

485 3140 3,85 520 2860 3,89

550 3350 3,80 550 3160 3,92

575 2650 3,75 525 3640 3,85

545 2780 3,85 485 4240 3,76

535 2230 3,70 515 2930 3,95

580 2490 3,98 470 3350 3,73

490 3620 3,50 510 3780 3,81

505 3340 3,66 535 2500 3,95

540 3150 3,92 510 3400 3,85

535 2900 3,75 505 3820 3,77

500 2840 3,78 490 3620 3,82

485 2040 3,85 515 3430 3,82

2. Установить достоверность различий в химическом составе трав по данным таблицы.

Содержание протеина, жира и клетчатки в зеленой массе трав, %

Люцерна			Клевер			Тимофеевка		
протеин	жир	клетчатка	протеин	жир	клетчатка	протеин	жир	клетчатка
5,1	0,7	6,2	4,1	0,9	4,9	3,0	0,6	7,8
5,0	0,6	6,8	3,5	0,8	5,0	3,4	0,7	8,5
4,6	0,7	6,6	3,8	0,8	5,5	3,2	0,7	6,3
4,8	0,6	6,5	4,3	0,9	5,2	3,3	0,8	7,7
4,7	0,8	6,1	3,3	0,8	5,3	3,3	0,8	8,4
5,1	1,0	5,8	3,7	1,1	4,8	3,5	0,9	9,2
4,9	0,8	5,7	3,6	1,2	4,9	3,1	1,0	8,1
4,5	0,6	6,0	3,9	0,6	5,0	3,2	0,8	7,4
4,8	0,7	6,1	4,0	1,2	4,8	3,1	0,9	8,0
4,9	0,5	5,9	3,4	0,7	4,4	3,2	0,8	7,2
5,2	0,8	6,3	4,2	1,0	5,0	3,1	0,7	7,9
5,1	0,7	6,9	3,6	0,9	5,1	3,5	0,8	8,6
4,7	0,8	6,7	3,9	0,9	5,3	3,3	0,8	6,4
4,7	0,5	6,4	4,2	0,8	5,1	3,2	0,7	7,6
4,6	0,7	6,0	3,2	0,7	5,2	3,2	0,7	8,3

3. Установить, достоверно ли отличаются по живой массе при рождении

бычки от телочек?

Масса бычков, кг Масса телочек, кг

32 33 31 29 35 34 30 24 28 27 29 26

32 35 36 34 30 31 27 25 28 29 29 23

31 33 32 37 34 33 26 27 29 27 24 25

32 33 30 31 33 34 26 27 26 24 30 28

32 33 34 35 33 29 25 28 23 25 26 27

4. Определить достоверно ли различаются крупный рогатый скот и овцы по основным клиническим показателям?

Крупный рогатый скот Овцы

Частота Частота Темпера- Частота Частота Темпера-

пульса дыхания тура, ^{0}C пульса дыхания тура, ^{0}C

63 16 38,5 71 23 39,5

68 18 38,7 74 21 39,7

57 23 38,5 76 19 39,5

69 20 38,4 75 22 39,4

66 14 38,2 78 24 39,3

62 17 38,6 75 26 39,6

51 13 38,7 76 20 39,7

72 17 38,5 73 21 39,4

70 18 38,0 76 20 38,8

75 19 37,9 75 26 39,7

66 17 38,8 77 22 39,5

64 21 39,0 79 21 39,6

54 17 38,6 77 17 39,4

63 16 38,5 72 20 39,3

67 20 38,6 70 22 39,7

71 18 38,7 74 23 39,4

6. В кормлении свиней большую роль играют незаменимые аминокислоты, которых не всегда достаточно в рационах. Для обогащения рационов применяют синтетические аминокислоты. Для изучения эффективности применения аминокислот был проведен опыт на свиньях. В контрольной группе животные получали основной рацион, а в опытной дополнительно 0,3% синтетических аминокислот (лизин + метионин). Установить, достоверно ли отличаются по живой массе в конце откорма опытные и контрольные подсвинки.

Живая масса свиней, кг

Опытная группа Контрольная группа

108 121 106 125 124 114 103 100 110 118

132 112 125 122 116 114 96 111 110 113

115 123 126 127 129 112 95 103 120 108

114 120 110 116 106 121 105 112 116 90

108 106 127 129 131 110 118 100 111 110

108 104 109 114 116 113 98 115 112 91