Методы обработки экспериментальных данных начали разрабатываться более двух веков тому назад в связи с необходимостью решения практических задач по агробиологии, медицине, экономике, социологии. Полученные при этом результаты составили фундамент такой научной дисциплины, как математическая статистика. В научный оборот слово «статистика» ввел немецкий ученый, профессор философии и права Готфрид Ахенваль (1719–1772), определяя этим термином круг вопросов, относящихся к **государствоведению:** описание государства, его устройства, быта населения, естественных условиях климата и др. Близким к современному пониманию сущности статистки было направление **политической арифметики,** основоположниками которой являлись английские исследователи: Джон Граунт (1620–1674), торговец, член Лондонского королевского общества; Уильям Петти (1623–1687), экономист, врач, доктор физики, профессор анатомии, изобретатель копировальной машины. Граунт Д. на основе обработки данных о естественном движении населения сделал вывод о существовании тесной связи между интенсивностью демографических процессов и особенностями социальной жизни людей, открыл некоторые закономерности массовых общественных явлений, используя собственный метод обработки и анализа первичных данных. Он попытался также построить таблицы смертности (или таблицы «дожития», показывающие процент доживших до определенного возраста), используемые для целей страхования. Исследования Петти относятся к политической экономии. Петти У. показал, что можно реконструировать информацию при отсутствии достаточного объема исходных данных (то есть с помощью различных расчетов найти нужные характеристики).

Возникновению математической статистики как науки способствовало появление теории вероятностей, развитие ее методов, использование в практических приложениях. Становление статистики связано с именами выдающихся ученых, в числе которых Пьер Симон Лаплас (1749–1827), Симеон Дени Пуассон (1781–1840), Жан Батист Фурье (1768–1830), Ламбер Адольф ﻿ Кетле (1796–1874). Они заложили основы современной статистической методологии, активно применяли полученные методы для установления закономерностей в общественных явлениях.

Современный уровень естественнонаучного эксперимента характеризуется большими потоками информации. При этом визуальный просмотр данных, не говоря уже об анализе, невозможен без применения ЭВМ. Обработка результатов экспериментов предполагает знание основных понятий и методов теории вероятностей и математической статистики. Выявление характерных классов задач в обработке экспериментальных данных и стандартных методов их решения позволяет выделить обработку результатов экспериментов из многообразия задач прикладной статистики. Знакомство со всеми этими методами расширяет арсенал средств, находящихся в распоряжении обработчика, что особенно важно в сложных случаях, например, когда измерения производятся при воздействии большого числа факторов, мешающих их проведению. Появление электронных таблиц (табличных процессоров) привело к тому, что статистические методы, ранее доступные лишь узкому кругу математиков, стали использоваться широким кругом специалистов разных областей. Дальнейшее развитие программного обеспечения привело к созданию большого количества прикладных пакетов по статистике. Удобной универсальной вычислительной средой для решения задач обработки экспериментальных данных является табличный процессор Microsoft Excel.

**Статистика изучает методы сбора и анализа результатов наблюдений массовых случайных явлений в целях выявления существующих закономерностей. Типичная задача математической статистики — на основании результатов наблюдений оценить вероятность случайного события или характеристики случайной величины. При решении любой задачи математической статистики имеется два источника информации. Первый источник — результаты наблюдений** (экспериментов), причем процесс наблюдений может корректироваться на основании предварительных результатов (так называемый последовательный анализ). **Второй источник — априорная (до опытная)** информация о свойствах изучаемого объекта, накопленная к текущему моменту. Эта информация отражается в статистической модели, выбираемой при решении задачи. Следует заметить, что степень обоснованности применения априорной информации зависит от компетентности и добросовестности конкретного исследователя и неверные исходные

 **Математическая статистика — раздел математики, в котором разрабатываются математические методы систематизации и обработки экспериментальных данных с целью изучения закономерностей массовых случайных явлений и использования их для научных и практических выводов. Выделяют: а) описательную статистику; б) теорию оценивания; в) теорию проверки гипотез.**

Обычно исследуют не всю совокупность объектов, а отбирают из неё некоторое количество объектов и исследуют только их. В этом и заключается выборочный метод.

**Генеральной совокупностью** называют совокупность всех объектов, над которыми производят наблюдение.

**Выборочной совокупностью (выборкой)** называют часть отобранных из **генеральной совокупности объектов.**

**Объёмом совокупности** называют **количество объектов в ней**. По выборке судят о генеральной совокупности. Выборка должна правильно представлять генеральную совокупность, то есть быть **репрезентативной**. Это обеспечивается способом отбора и увеличением объёма выборки.

Выборка, в которой меньше 30-ти элементов, называется малой. В противном случае, выборка называется большой. Для выборок малого объема необходимо выбирать специально разработанные методы. Выборочные данные делятся на: а) качественные; б) количественные. Качественные данные представляются (кодируются) определенным числом в соответствии с некоторым свойством.

В дальнейшем будет предполагаться, что наблюдения будут представляться количественной информацией. При этом выделяют: а) данные непрерывного типа – возможно появление любого значения из некоторого интервала; б) данные дискретного типа – возможны лишь изолированные значения из некоторого интервала.

**Первичная обработка** результатов наблюдений. Что такое наблюдаемые данные? **Большой массив беспорядочно расположенных чисел**. Для работы с данными удобно их группировать. Пример. 0 1 2 2 1 2 0 0 0 0 –выборка. Объём выборки: n =10

|  |  |  |  |
| --- | --- | --- | --- |
| xi | 0 | 1 | 2 |
| ni | 5 | 2 | 3 |
|  wi | 0,5 | 0,2 | 0,3 |

Наблюдаемые значения xi называют **вариантами**. Последовательность вариант, записанных в возрастающем порядке, называют **вариационным** **рядом**. Частотой варианты называют число ni , показывающее сколько раз встречается данная варианта. Относительной частотой варианты называют отношение частоты к объёму выборки: wi = ni / n. Статистическим распределением выборки (дискретным статистическим рядом) называется перечень вариант и соответствующих им частот или относительных частот.

Замечание: сумма всех частот равна объёму выборки;

сумма всех относительных частот равна 1;

относительная частота варианты даёт приближённое значение вероятности этой варианты.

**Интервальный статистический ряд**: 1) разбивают весь интервал, в который попадают варианты, на частичные интервалы; 2) в верхнюю строку записывают полученные интервалы; 3) в нижнюю строку записывают частоту попадания в соответствующий интервал.

 27 3,5 21,1 0,8 12,3 18 11 3,4 1,2 5,2 22 17,2

18,1 11,1 0,7 7,9 19 3,2 4,9 25,4 6,1 21,6 22,3 3,4

 18,4 3,4 23,2 13,1 6,5 2,4 18,4 14,1 2,1 24,8 17,4 15,1

 4,8 19,8 10,4 16,1 3,7 29,4 3,1 28,7 16,4 22,2 1,7 12,4

 17 15,3 3,3 14 16,8 10,1 2,4 20 14,1 19 19,8 5,4

 2,5 4,1 24,4 0,4 24,7 1,3 13,7 0,1 28 24 17,1 15

 3,1 19 0,4 23,1 6,7 4,6 14,8 20,7 16,2 9,4 21,3 13,4

16,1 15,7 11,3 5,1 1,9 2,8 17 2 20,8 3,4 16,7 9,3

15,2 8,7 10,7 2,8

Xmax = 28.7 Xmin = 0.1 n = 100. Можно взять интервал от 0 до 30.

1. разбивают весь интервал, в который попадают варианты, на частичные

интервалы: [0, 5); [5, 10); [10, 15); [15, 20); [20, 25); [25, 30). Распределяем данные числа по интервалам.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  [ 0; 5) |  [ 5; 10) |  [ 10; 15) |  [15; 20) |  [ 20; 25) | [ 25; 30] |
| ® ® ® ® ®® ® ® ® ®® ® ® ® ®® ® ® ® ®® ® ® ® ®® ® ® ® ® | ® ® ® ® ®® ® ® ® ® | ® ® ® ® ®® ® ® ® ®® ® ® ® ® | ® ® ® ® ®® ® ® ® ®® ® ® ® ®® ® ® ® ®® ® ® ® ® | ® ® ® ® ®® ® ® ® ®® ® ® ® ® | ® ® ® ® ® |
| 30 | 10 | 15 | 25 | 15 | 5 |

1. в верхнюю строку записывают полученные интервалы; 3) в нижнюю строку записывают частоту попадания в соответствующий интервал

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| xi | 0 - 5 | 5 - 10 | 10 - 15 | 15 - 20 | 20 - 25 | 25 - 30 |
| ni | 30 | 10 | 15 | 25 | 15 | 5 |
| wi | 0.3 | 0.1 | 0.15 | 0.25 | 0.15 | 0.05 |

На сколько интервалов разбивать выборку?

k = 1 +3.332 ⋅lg n или k ≤ 5⋅lgn п – объём выборки.

 Замечание: первая из этих формул носит название формула Старджеса.

Таблица значений q = q($γ;n$)

|  |  |  |  |
| --- | --- | --- | --- |
| n | $$γ$$ | n | $$γ$$ |
| 0.95 | 0.99 | 0.999 | 0.95 | 0.99 | 0.999 |
| 5 | 1.37 | 2.67 | 5.64 | 20 | 0.37 | 0.58 | 0.88 |
| 6 | 1.09 | 2.01 | 3.88 | 25 | 0.32 | 0.49 | 0.73 |
| 7 | 0.92 | 1.62 | 2.98 | 30 | 0.28 | 0.43 | 0.63 |
| 8 | 0.80 | 1.38 | 2.42 | 35 | 0.26 | 0.38 | 0.56 |
| 9 | 0.71 | 1.20 | 2.06 | 40 | 0.24 | 0.35 | 0.50 |
| 10 | 0.65 | 1.08 | 1.80 | 45 | 0.22 | 0.32 | 0.46 |
| 11 | 0.59 | 0.98 | 1.60 | 50 | 0.21 | 0.30 | 0.43 |
| 12 | 0.55 | 0.90 | 1.45 | 60 | 0.188 | 0.269 | 0.38 |
| 13 | 0.52 | 0.83 | 1.33 | 70 | 0.174 | 0.245 | 0.34 |
| 14 | 0.48 | 0.78 | 1.23 | 80 | 0.161 | 0.226 | 0.31 |
| 15 | 0.46 | 0.73 | 1.15 | 90 | 0.151 | 0.211 | 0.29 |
| 16 | 0.44 | 0.70 | 1.07 | 100 | 0.143 | 0.198 | 0.27 |
| 17 | 0.42 | 0.66 | 1.01 | 150 | 0.115 | 0.160 | 0.211 |
| 18 | 0.40 | 0.63 | 0.96 | 200 | 0.099 | 0.136 | 0.185 |
| 19 | 0.39 | 0.60 | 0.92 | 250 | 0.089 | 0.120 | 0.162 |

Пример нахождения промежуточного значения:

65 = $\frac{60+70}{2} \rightarrow q\left(γ,n\right)=q\left(0.95;65\right)= \frac{0.188+0.174}{2}=0.181$

**Задания для самостоятельного решения:**

**№1 (Бобров В) Провести первоначальную обработку статистических данных (найти выборочную среднюю, выборочную дисперсию, среднеквадратическое отклонение).**

**20 15 17 19 23 18 21 15 16 13**

**20 16 19 20 14 20 16 14 20 19**

**15 19 17 16 15 22 21 12 10 21**

**18 14 17 18 19 13 19 18 20 23**

**16 20 18 14 19 17 21 17 19 13 Первый интервал [9.5; 11.5)**

**№ 5 ( Морозов Л) Провести первоначальную обработку статистических данных (найти выборочную среднюю, выборочную дисперсию, среднеквадратическое отклонение).**

**43 73 82 66 72 65 57 52 40 70**

**46 52 43 52 44 56 51 53 77 53**

**38 54 63 46 48 47 54 37 50 65**

**50 29 47 49 59 26 62 63 60 47**

**58 59 66 35 64 65 35 45 55 29 Первый интервал [26; 34)**

**№ 20 (Рой А) Провести первоначальную обработку статистических данных (найти выборочную среднюю, выборочную дисперсию, среднеквадратическое отклонение).**

**15,8 16,0 15,7 16,0 15,7 15,8 15,8 15,9 16,1 15,5**

**15,7 15,9 16,0 15,7 15,7 15,5 16,2 15,7 15,9 15,8**

**15,7 15,4 16,0 15,7 15,5 15,8 16,0 15,8 15,9 16,2**

**15,9 15,6 16,0 15,7 16,1 15,7 16,1 15,9 15,8 15,7**

**15,9 15,6 15,6 15,8 15,6 15,7 15,6 15,8 15,8 15,9**

**15,8 15,5 15,9 15,6 15,7 16,0 15,6 15,8 16,0 16,1**

**15,7 15,5 15,7 15,6 15,5 15,8 16,0 15,4 15,6 15,9**

**Первый интервал [15.4; 15.5)**

**№ 22** **(Аверьянова А)** **Провести первоначальную обработку статистических данных (найти выборочную среднюю, выборочную дисперсию, среднеквадратическое отклонение).**

**2,4 2,5 2,4 2,7 2,4 2,7 2,2 2,3 2,6 2,5 2,3 2,5 2,4 2,5**

**2,7 2,4 2,6 2,4 2,5 2,4 2,5 2,4 2,2 2,4 2,5 2,4 2,9 2,4**

**2,5 2,4 2,6 2,4 2,5 2,6 2,5 2,6 2,4 2,6 2,7 2,6 2,5 2,7**

**2,1 2,5 2,7 2,6 2,5 2,2 2,6 2,5 2,1 2,5 2,8 2,5 2,6 2,5**

**2,4 2,6 2,4 2,5 2,4 2,7 2,4 2,7 2,5 2,7 2,6 2,5 2,8 2,5**

 **Первый интервал [2.1; 2.2)**

**№24 (Булатова Д) Провести первоначальную обработку статистических данных (найти выборочную среднюю, выборочную дисперсию, среднеквадратическое отклонение).**

**136 122 132 128 123 133 130 131 134 149 138 127 119 137**

**133 130 143 134 128 131 118 133 131 132 118 128 122 130**

**139 145 122 130 128 136 132 126 124 117 139 132 141 144**

**138 133 127 150 144 133 134 125 140 135 129 138 147 138**

**150 126 136 135 150 135 138 140 122 142 127 132 127 145**

**140 133 127 142 144 125 132 145 137 132**

**Первый интервал [116; 121)**

**№19 (Андреев В) Провести первоначальную обработку статистических данных (найти выборочную среднюю, выборочную дисперсию, среднеквадратическое отклонение).**

**70 65 76 78 74 56 70 67 60 71 64 75 79 75 71 77**

**62 69 67 73 67 75 76 71 77 66 72 66 74 63 69 70**

**71 65 70 82 67 68 75 77 67 74 81 67 73 69 73 69**

**73 69 64 74 80 66 68 75 73 69 65 61 79 84 68 72**

**73 71 68 69 79 81 76 72 71 65 71 76 78 77 76 70**

**Первый интервал [56; 60)**

**№15 (Машуков И) Провести первоначальную обработку статистических данных (найти выборочную среднюю, выборочную дисперсию, среднеквадратическое отклонение).**

**17 21 8 20 23 18 22 20 17 12**

**20 11 9 19 20 9 19 17 21 13**

**17 22 10 22 20 15 20 19 20 13**

**21 20 21 9 14 11 19 18 23 19 Первый интервал [7.5; 9.5)**

**№17 (Крютченко Н) Провести первоначальную обработку статистических данных (найти выборочную среднюю, выборочную дисперсию, среднеквадратическое отклонение).**

**8,5 7,1 6,7 6,2 2,9 4,4 6,0 5,8 5,4**

**8,2 6,9 6,5 6,1 3,8 6,0 5,6 6,0 5,3**

**7,7 6,8 6,5 6,1 4,2 4,7 5,6 5,4 5,3**

**7,4 6,7 6,4 6,1 4,5 6,0 5,8 5,6 5,1 Первый интервал [2,9; 3,7)**

**№10 (Карась В) Провести первоначальную обработку статистических данных (найти выборочную среднюю, выборочную дисперсию, среднеквадратическое отклонение).**

**58 110 91 101 83 92 95 71 103 110**

**118 102 97 96 99 64 99 107 104 109**

**117 97 78 98 108 82 92 108 111 80**

**111 85 97 112 91 95 94 105 80 90**

**117 115 89 98 93 110 94 92 109 100 Первый интервал [57,9; 66.5)**

**№9 (Верещагин Н) Провести первоначальную обработку статистических данных (найти выборочную среднюю, выборочную дисперсию, среднеквадратическое отклонение).**

**128 158 167 157 150 157 142 137 125 155**

**131 137 118 137 129 141 136 138 162 138**

**123 159 148 131 133 132 139 122 135 150**

**135 104 132 134 144 111 147 148 155 132**

**143 144 151 120 149 150 149 130 140 155**

**Первый интервал [104; 113).**