Уважаемые студенты!

Вам необходимо изучить теоретический материал лекции, рассмотреть решение примеров, ответить на контрольные вопросы.

Ответ присылать на электронную почту

Электронная почта: *hvastov@rambler.ru*

Лекция

Понятие о законе больших чисел.

План

1. Методы статистических исследований
2. Выборка наблюдений. Этапы статистического исследования.
3. Основные характеристики
4. Дисперсия. Алгоритм ее вычисления.

При статистических методах используются специфические методы, такие как массовое наблюдение, группировка, вычисление характеристик и построение графиков.

Рассмотрим статистическое наблюдение на примере практической задачи.

Задача. Представьте, что вы – мэр города и вам необходимо построить в городе кинотеатр, стадион и большой торговый центр. Построить всё сразу нет возможности, поэтому необходимо сначала выбрать один из представленных объектов. Какой объект выбрать первым? Решение. Нам необходимо учесть желание как можно большего количества горожан. Мы могли бы опросить всех горожан, но для большого города это будет сложно. В таком случае опрашивают часть горожан города в надежде на то, что их мнение с той или иной степенью достоверности отражает мнение горожан. Здесь мы имеем дело с делением наблюдений по степени полноты охвата. В первом случае – это сплошное наблюдение, во втором – несплошное. Выборка наблюдений. Наиболее распространённое несплошное наблюдение – выборочное наблюдение. При выборочном наблюдении всё население города называется генеральной совокупностью, а опрашиваемые люди – выборкой. Этапы статистического исследования.

1. Данные измерений упорядочивают и группируют.
2. Составляют таблицы распределения данных.
3. Таблицы распределения данных позволяют построить графики распределения данных в виде многоугольника, гистограммы или круговой диаграммы. Получения паспорта данных измерений, который состоит из небольшого количества основных числовых характеристик полученной информации. Основные характеристики. Объём измерения – количество

источников. Размах измерения – разница между наибольшим и наименьшим значениями результатов измерения. Мода измерения – наиболее часто встречающийся результат. Среднее значение – частное от деления суммы всех результатов измерения на объём измерения.

Медиана – значение варианты, находящееся в середине упорядоченного ряда данных измерения. Если количество вариантов чётное, то медиана равна полусумме вариант, стоящих в середине упорядоченного ряда данных.

Среднее, мода и медиана относятся к одному и тому же типу числовых характеристик данных измерений. Иногда их называют мерами центральных тенденций.

Рассмотрим пример. При поиске нового места работы, как правило, естественно поинтересоваться средним уровнем зарплаты, однако может так случиться, что информация о высокой средней зарплате введёт вас в заблуждение. Например, пусть медиана или мода значительно меньше, чем величина средней зарплаты (см. рисунок 1).

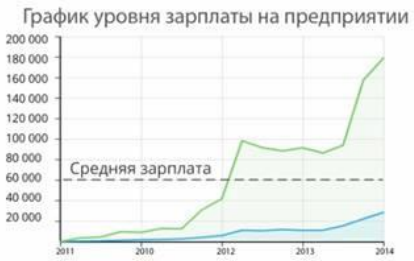

Рисунок 1 – График уровня зарплаты на предприятии

Это значит, что на предприятии есть малая часть сотрудников, которые получают очень высокую зарплату, и есть большинство сотрудников, которые получают низкую зарплату. В таком случае логичнее интересоваться про зарплату в том секторе должностей, на которую претендует соискатель (см. рисунок 2).

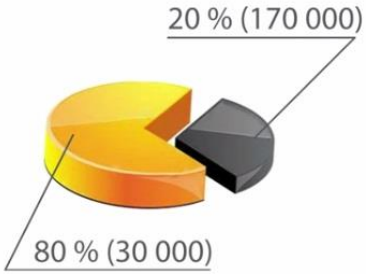

Рисунок 2 – Секторы должностей

Для дальнейшего статистического анализа важно знать не только центральные значения результатов измерений, но и то, насколько эти значения кучно (тесно) расположены вокруг некоторых центральных или средних значений.

Пример – «Пристрелка ружей». На испытательном стенде оружейного завода пристреливают готовые ружья (уточняют и корректируют их прицел). В таблице приведены измерения горизонтальных отклонений от цели при стрельбе из трёх ружей (в сантиметрах).

Для каждого ружья есть десять выстрелов, и в каждой ячейке приведены значения отклонения от цели. Если отклонение слева от центра цели, то оно записано со знаком «-».

	Выстрел									
	1	2	3	4	5	6	7	8	9	10
Ружьё *А*	+1,0	+1,0	+2,0	+1,5	+2,0	+2,0	+1,5	+1,5	+0,5	+1,0
Ружьё *Б*	+1,0	0	−1,5	+1,5	−0,5	−1,5	+2,0	+1,0	−1,0	+2,0
Ружьё *В*	−0,5	−1,0	0	−1,5	−1,0	+1,0	+1,0	+1,5	+1,0	+3,0

Рисунок 3 – Таблица выстрелов

Необходимо вычислить средние значения результатов испытаний.

Решение

1. Среднее для ружья А равно:

$$\frac{1{,}0+1{,}0+2{,}0+1{,}5+2{,}0+2{,}0+1{,}5+1{,}5+0{,}5+1{,}0}{10}=\frac{14{,}0}{10}=1{,}4$$

2. Среднее для ружья Б равно:

$$\frac{1{,}0+0-1{,}5+1{,}5-0{,}5-1{,}5+2{,}0+1{,}0-1{,}0+2{,}0}{10}=\frac{3{,}0}{10}=0{,}3$$

3. Среднее для ружья В равно:

$$\frac{-0{,}5-1{,}0+0-1{,}5-1{,}0+1{,}0+1{,}0+1{,}5+1{,}0+3{,}0}{10}=\frac{3{,}5}{10}=0{,}35$$

$M_{\text{А}}=1{,}4;\ M_{\text{Б}}=0{,}3;\ M_{\text{В}}=0{,}35$

Какое же из ружей более точное? Если для оценки точности пользоваться только средними значениями отклонений, то наиболее точным является ружьё Б, поскольку его среднее число отклонений является наименьшим. За ним следует ружьё В, а последнее по точности – ружьё А.

Ответ: 1,4; 0,3; 0,35.

Однако если внимательно проанализировать таблицу, то мы увидим, что практически все попадания из ружья А расположены очень близко возле его среднего значения 1,4. Скорее всего, имеет место некая систематическая ошибка, и если найти способ её исправить, то после исправления ружьё будет стрелять более точно. С ружьём Б дело обстоит хуже, его выстрелы сильно разбросаны слева и справа от нуля, в среднем ошибка небольшая, 0,3 см, но для каждого отдельного выстрела отклонение может достигать высоких значений, а знак отклонения меняется случайным образом. Вывод: такое ружьё совершенно ненадёжно. Проанализируем ружьё В. Первые 5 выстрелов ружьё стреляло довольно стабильно, и все выстрелы

располагались левее цели, следующие выстрелы правее, а десятый выстрел дал заметную величину ошибки. Можно сделать вывод, что отклонения с ружьём В произошли в процессе испытания и для проведения более точного испытания необходимо вернуть первоначальные настройки ружья. После этого результаты, скорее всего, будут более точными.

Введём количественный параметр, который характеризует разброс данных.

Меры разброса данных – характеристики выборки или генеральной совокупности, предназначенные для описания степени разброса данных:

1. Дисперсия.

2. Среднее квадратическое отклонение.

Дисперсия D– числовая характеристика данных измерения, отвечающая за разброс, рассеивание данных вокруг их среднего значения.

Среднее квадратическое отклонение σ – число, равное квадратному корню из дисперсии.

$$\boldsymbol{\sigma = \sqrt{D}}$$

Чем меньшую величину имеет дисперсия или среднее квадратическое отклонение, тем более тесно (более кучно) расположены результаты измерений вокруг среднего отклонения.

Алгоритм вычисления дисперсии:

Для нахождения дисперсии D данных $x_1, x_2, \dots, x_n$ измерения следует вычислить:

1. Среднее значение $M = \dfrac{x_1 + x_2 + \dots + x_n}{n}$

2. Отклонения данных от M, т. е. $x_1 - M, x_2 - M, \dots, x_n - M;$

3. Квадраты $(x_i - M)^2$ отклонений $x_i - M\ (i = 1,2, \dots, n)$, найденных на предыдущем шаге;

4. Среднее значение всех квадратов отклонений – дисперсия:

$D = \dfrac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n}$ $\boldsymbol{\sigma = \sqrt{D}}$ – среднее квадратическое отклонение.

Подсчитайте дисперсию для результатов измерений ружей А и Б из предыдущего примера.

Решение. Проведём подсчёты для испытания ружья А. Удобно собрать все значения в таблицу.

	Выстрел									
	1	2	3	4	5	6	7	8	9	10
Результат	+1,0	+1,0	+2,0	+1,5	+2,0	+2,0	+1,5	+1,5	+0,5	+1,0
Отклонение	–0,4	–0,4	0,6	0,1	0,6	0,6	0,1	0,1	–0,0	–0,4
Квадрат отклонения	0,16	0,16	0,36	0,01	0,36	0,36	0,01	0,01	0,81	0,16

Рисунок 4 – Таблица подсчета

- строка «результат» содержит результат измерения;

- строка «отклонение» содержит подсчёт отклонения от среднего отклонения, подсчитанного в начале примера. Оно равнялось 1,4;

- строка «квадрат отклонения» – необходима для подсчёта дисперсии.

$$D_A = \frac{0{,}16 \cdot 3 + 0{,}36 \cdot 3 + 0{,}01 \cdot 3 + 0{,}81}{10} = 0{,}24$$

$$\sigma_A = \sqrt{D_A} \approx 0{,}5$$

Таким же образом подсчитаем значение ружья Б. В этом случае среднее значение отклонения равно 0,3.

	Выстрел									
	1	2	3	4	5	6	7	8	9	10
Результат	+1,0	0	−1,5	+1,5	−0,5	−1,5	+2,0	+1,0	−1,0	+2,0
Отклонение	0,7	−0,3	−1,8	1,2	−0,8	−1,8	1,7	0,7	−1,3	1,7
Квадрат отклонения	0,49	0,09	3,24	1,44	0,64	3,24	2,89	0,49	1,69	2,89

Рисунок 5 – Таблица подсчета

$$D_A = \frac{0{,}49 \cdot 2 + 3{,}24 \cdot 2 + 2{,}89 \cdot 2 + 0{,}09 + 1{,}44 + 0{,}64 + 1{,}69}{10} = 1{,}71 \quad \sigma_A = \sqrt{D_A} \approx 1{,}31$$

После подсчёта дисперсии мы меняем своё мнение о качестве ружей.

Конечная задача статистического исследования состоит в том, чтобы, основываясь на свойствах выборки, сделать выводы о всей генеральной совокупности. Такие выводы формулируются на основе методов теории вероятности. В любом случае значение любого признака для выборки в общем случае будет отличаться от такового для всей генеральной совокупности.

Возвращаясь к примеру, описанному в начале занятия, можно утверждать, что результаты опроса любой части горожан меньше ста процентов будут, скорее всего, отличаться от результатов опроса всех горожан. Мы употребляем словосочетание «скорее всего», так как из теории вероятности мы знаем, что любое событие, которое не является достоверным, не обязательно является невозможным.

Контрольные вопросы:

1. Методы статистических исследований.
2. Выборка наблюдений. Этапы статистического исследования.
3. Основные характеристики.