

## ГЛОССАРИЙ

### А

- Автокорреляция - статистическое определение: Коэффициент корреляции, рассчитанный для двух групп данных во временном ряду.

Пример: Отмеченное количество скота в течение двух последовательных лет обычно характеризуется высокой автокорреляцией в том случае, если продолжительность жизни скота значительно превышает два года.

- Анализ неопределенности - статистическое определение: Анализ неопределенности модели имеет целью обеспечить количественные измерения неопределенности выходных значений, вызванной неопределенностями в самой модели и в ее входных величинах, а также исследовать относительное значение этих факторов.
- Асимметрия - отношение центрального момента третьего порядка к кубу среднеквадратического отклонения.

### Б

- Бесповторная выборка - выборка, при которой отобранный объект после проведения обследований не возвращается в генеральную совокупность.

### В

- Вероятность - это отношение числа благоприятных исходов к общему числу исходов при равенстве событийной ценности (веса) исходов.
- Внутригрупповая дисперсия - средняя арифметическая групповых дисперсий, взвешенная по объемам групп.
- Временной ряд - это ряд значений, затронутых случайными процессами и наблюдаемых в последовательные (но обычно равноудаленные) моменты времени.
- Выборка - совокупность случайно отобранных из изучаемой совокупности объектов (генеральной выборки).
- Выборочное среднее - частное от деления суммы значений всех элементов выборки на число элементов выборки.

### Г

- Гистограмма - ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат интервалы длиной  $h$ , а высоты  $n$ .
- Групповая дисперсия - дисперсия значений признака, принадлежащих группе, относительно групповой средней.
- Групповая средняя - среднее арифметическое значений признака, принадлежащих группе.

## Д

- Двумерная случайная величина - величина, имеющая два аргумента.
- Дискретная случайная величина - величина, принимающая отдельные значения с определенными вероятностями.
- Дисперсия (дисперсия генеральной совокупности) - это параметр плотности распределения вероятностей, который выражает изменчивость данной совокупности. Это второй центральный момент случайной величины. Выборочная дисперсия определяется как мера дисперсии, которая представляет собой сумму квадратов отклонений данных наблюдений от их среднего значения, деленную на количество наблюдений минус единица  $s^2 = \frac{1}{n-1} \sum_i^n (x_i - \bar{x})^2$ .
- Дисперсия случайной величины - математическое ожидание квадрата отклонения случайной величины от ее математического ожидания.
- Доверительный интервал - интервал, который покрывает неизвестный параметр  $\theta$  с заданной надежностью (вероятностью)  $p$ . Доверительный интервал обладает тем свойством, что, во-первых, его границы вычисляются исключительно по выборке (и, следовательно, не зависят от неизвестного параметра), и, во-вторых, он накрывает неизвестный параметр с вероятностью  $p$ .
- Достоверное событие - событие, которое обязательно произойдет, если будет осуществлена определенная совокупность условий.

## З

- Закон больших чисел - математическая теорема, которая формализует общеизвестную мудрость, согласно которой усредненное значение становится более точной аппроксимацией к среднему значению по мере увеличения количества наблюдений.
- Закон распределения случайной величины - соответствие между возможными значениями случайной величины и их вероятностями.

## И

- Интервальная оценка - оценка, которая определяется концами интервала.

## К

- Конкурирующая гипотеза - гипотеза противоречащая основной.
- Корреляционная зависимость - зависимость, при которой при изменении одной из величин изменяется среднее значение другой.

- Корреляционный момент - характеристика связи между двумя случайными величинами.
- Коэффициент вариации - выраженное в процентах отношение выборочного среднеквадратического отклонения к выборочной средней.
- Коэффициент корреляции - отношение ковариации к произведению среднеквадратических отклонений двух случайных величин.
- Критерий Стьюдента - направлен на оценку различий величин средних и двух выборок  $X$  и  $Y$ , которые распределены по нормальному закону. Одним из главных достоинств критерия является широта его применения. Он может быть использован для сопоставления средних у связанных и несвязанных выборок, причем выборки могут быть не равны по величине.
- Критическая область - совокупность значений критерия, при которых нулевую гипотезу отвергают.

## Л

- Линейная модель - переменная величина  $y$  считается линейной (или линейной функцией) в зависимости от переменных величин  $x_1, x_2, \dots$  если  $y$  может быть выражена формулой  $y = b_0 + b_1x_1 + b_2x_2 + \dots$ , где члены  $b$  являются постоянными числами. Мнение о том, считается ли функция линейной или нет, может меняться в зависимости от того контекста, в котором она применяется.

Пример: Выброс  $E$  обычно выражается как произведение коэффициента выброса  $F$  и показателя уровня деятельности  $A$ . В том случае, если  $F$  является фиксированной константой, а  $E$  меняется только когда меняется  $A$ ,  $E$  находится в линейной зависимости от  $A$ . В тоже время, когда как  $F$ , так и  $A$  считаются переменными величинами (например, когда применяется уравнение распространения ошибок для оценки дисперсии  $E$  как функции дисперсий и ковариации  $A$  и  $F$ ),  $E$  не является линейной функцией  $F$  и  $A$ .

- Линейная регрессия - обеспечивает возможность вычерчивания прямой линии по точкам комплекта данных наблюдений с учетом последствий изменчивости данных наблюдений.

Пример: если данные наблюдений за выбросами наносятся на график по точкам, соответствующим уровням деятельности, наклон линии, построенной при помощи линейной регрессии, отражает оценку соответствующего коэффициента выброса. Этот метод может быть также использован для оценки прямолинейной тенденции для количества, которое изменяется во времени.

- Логарифмически нормальное распределение - это асимметричное распределение, которое начинается с нуля, возрастает до максимума и затем более плавно убывает к бесконечности.

## М

- Математическое ожидание - число, относительно которого стабилизируется среднее арифметическое возможных значений случайной величины при достаточно большом количестве испытаний.

- Медиана (медиана совокупности) - это значение, которое делит интеграл плотности распределения вероятностей (ПРВ) на две половины. Для симметричных ПРВ она равна среднему значению. Медиана  $\tilde{n}$  это 50-й процентиль совокупности.
- Выборочная медиана - это формула оценки медианы совокупности. Это значение, которое делит упорядоченную выборку на две равные половины. Если имеется  $2n + 1$  наблюдений, медиана принимается в качестве  $(n + 1)$ -го члена упорядоченной выборки. Если имеется  $2n$ , то считается, что она проходит посередине между  $n$ -м и  $(n + 1)$ -м членами.
- Межгрупповая дисперсия - дисперсия групповых средних относительно общей средней.
- Метод наименьших квадратов - Задача заключается в нахождении коэффициентов функциональной зависимости исследуемых переменных величин, при которых обеспечивается минимальная дисперсия разницы выборочных значений и функции, которой аппроксимируют стохастическую зависимость исследуемых переменных. То есть, при данных  $a$  и  $b$  сумма квадратов отклонений экспериментальных данных от найденной прямой будет наименьшей.
- Мода - варианта ряда, которая имеет наибольшую частоту.
- Модель - это количественно определенная абстракция реальной ситуации, которая может упрощать или отбрасывать некоторые характеристики для лучшей концентрации на ее более важных элементах.

Пример: соотношение, определяющее выбросы как произведение коэффициента выбросов и объема деятельности, является простой моделью. Термин "модель" также часто используется в смысле реализации модельной абстракции при помощи компьютерного программного обеспечения, которое рассчитывает комплект выходных значений для данного комплекта входных значений, например, численные модели глобального климата.

- Моменты случайных величин - характеристики случайных величин, определяющие математическое ожидание  $k$ -й степени отклонения случайной величины.

## Н

- Неопределенность - это параметр, связанный с результатом измерения, который характеризует дисперсию значений, которые могли бы быть разумно приписаны измеренному количеству.
- Непрерывная случайная величина - величина, принимающая значения, сколь угодно мало отличающиеся друг от друга.
- Несмещенная оценка - оценка  $x$ , математическое ожидание которой равно оцениваемому параметру  $x$ .
- Нормальное распределение (или распределение Гаусса) - имеет плотность распределения вероятностей, приведенную в нижеследующем уравнении, и определяется двумя параметрами (средним значением  $\mu$  и среднеквадратическим отклонением  $\sigma$ ):  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ , для  $-\infty \leq x \leq +\infty$ .
- Нулевая гипотеза - основная выдвинутая гипотеза.

## О

- Общая дисперсия - дисперсия значений признака всей совокупности относительно общей средней.
- Остаток - для наблюдаемой величины, поведение которой моделируется при помощи статистической модели, остаток представляет собой разницу между наблюдаемой величиной и величиной, спрогнозированной при помощи модели, например, линейная регрессия. Таким образом, остаток  $\hat{\epsilon}$  это компонент наблюдения, который невозможно объяснить при помощи модели.
- Оценка - это определение значения количества или его неопределенности посредством подстановки численных значений наблюдений в формулу для расчета или оценки. Результаты оценки могут быть выражены следующим образом:
  - точечная оценка, которая дает число, которое может быть использовано в качестве аппроксимации к параметру (такому как среднеквадратическое отклонение выборки, при помощи которого оценивается совокупное среднеквадратическое отклонение), или
  - интервальная оценка, точно определяющая доверительный уровень.

Пример: Такое заявление, как "Общий выброс оценивается в 100 кг, а его коэффициент вариации составляет 5%", основано на точечных оценках выборочного среднего и среднеквадратического отклонения, в то время как такое заявление, как "Общий выброс составляет от 90 до 110 кг при вероятности 95%", является выражением результатов оценки в виде доверительного интервала.

## П

- Плотность распределения вероятностей - вероятность того, что непрерывная случайная величина примет значение на указанном интервале.
- Повторная выборка - выборка, при которой отобранный объект возвращается после проведения обследования обратно в генеральную совокупность.
- Полигон частот - ломаная линия, отрезки которой соединяют точки  $(x_1, n_1)$ .
- Производящая функция - функция, определяющая вероятность наступления события при различных вероятностях появления в каждом испытании.
- Простая случайная выборка - выборка  $n$  элементов, взятых из совокупности таким образом, что каждая возможная выборка имеет равную вероятность быть отобранной.
- Процентиль,  $k$ -й процентиль или процентиль совокупности - это значение, которое отделяет самую нижнюю  $k$ -ю часть интеграла плотности распределения вероятностей, т.е. интеграл хвоста плотности распределения вероятностей от  $k$ -го процентиля в направлении более низких плотностей вероятностей.

$k$ -й процентиль совокупности ( $0 \leq k \leq 100$ ) с функцией распределения  $F(x)$  равен  $z$ , где  $z$  удовлетворяет уравнению  $F(z) = k/100$ .

Выборочный  $k$ -й процентиль является аппроксимацией совокупного процентиля, который выводится из выборки. Это значение, ниже которого находятся  $k$  процентов наблюдений.

## Р

- Размах варьирования  $R$  - разность между наибольшей и наименьшей вариантой.
- Регрессия - представление одной случайной величины как функции другой.
- Распределение вероятностей - функция, устанавливающая вероятность того, что случайная величина принимает любое данное значение или относится к данному множеству значений. Вероятность всего множества значений случайной величины равняется 1.

## С

- Систематическая ошибка - это разность между истинным, но обычно неизвестным, значением измеряемого количества и средним значением наблюдаемой величины, которая оценивалась бы посредством выборочного среднего значения бесконечного множества наблюдений.
- Случайная ошибка отдельного измерения - это разность между отдельным измерением и вышеуказанной предельной величиной выборочного среднего значения.
- Случайная величина - величина, которая имеет неизвестное значение до испытания (множество альтернатив), а в результате информативного испытания может принять какое-либо определенное или более ограниченное в альтернативах значение.
- Совокупное среднеквадратическое отклонение - это положительный квадратный корень дисперсии. Оно определяется выборочным среднеквадратическим отклонением, которое представляет собой положительный квадратный корень выборочной дисперсии.
- Состоятельная оценка - оценка, которая при  $n > n_0$  стремится по вероятности к оцениваемому параметру.
- Среднеарифметическое значение - сумма значений, разделенная на количество значений.
- Среднее значение, среднее по совокупности, математическое ожидание или ожидаемое значение — это в широком смысле слова мера центрального значения, к которому стремятся значения, выбранные из распределения вероятности. Среднее значение выборки или средняя арифметическая величина — это формула оценки для среднего значения. Это - несмещенная и согласованная формула оценки среднего по совокупности (предполагаемой величины), и сама по себе является случайной переменной величиной со своим собственным значением дисперсии.
- Статистика - это функция выборочных случайных переменных величин.
- Статистическая гипотеза - гипотеза (предположение) о виде неизвестного распределения, или параметрах неизвестного распределения.
- Статистический критерий - случайная величина, служащая для проверки нулевой гипотезы.
- Статистическое распределение выборки - перечень вариантов и соответствующих им частот или относительных частот.
- Стохастическая зависимость - зависимость, при которой изменение одной из величин влечет изменение другой.

## Т

- Теорема Лапласа - определение вероятности наступления события в  $k$  измерениях из  $n$  (при больших  $k$  и  $n$ ).
- Теория вероятностей - наука, изучающая общие закономерности случайных явлений массового характера.
- Точечная оценка - оценка, которая определяется одним числом.

## У

- Условная вероятность - вероятность наступления интересующего нас события, связанная с дополнительными условиями.

## Ф

- Формула Байеса - определение апостериорной (послеопытной) вероятности на основе априорной (доопытной) на основе проведения эксперимента.
- Формула Бернулли - определение вероятности наступления события в измерениях из  $n$ .
- Формула оценки - это формула, точно определяющая то, каким образом осуществлять расчет оценочного значения выборки совокупного параметра на основе выборочных данных. Например, коэффициенты выбросов часто оцениваются в виде выборочных средних комплектов измерений. Может существовать несколько формул оценки для совокупного параметра, и каждая формула в целом обладает своими собственными выборочными свойствами, при этом согласованность и несмещенность относятся к числу наиболее важных.
- Примеры точечных формул оценок включают среднеарифметическое  $\bar{x}$ , которое является обычно используемой формулой оценки для предполагаемого значения (среднего), и дисперсию выборки  $s^2$ , которая является обычно используемой оценкой для дисперсии.
- Функция распределения - функция, определяющая вероятность того, что  $X$  примет значение меньше  $x$ .

## Х

- Характеристики положения - характеристики, определяющие наиболее возможные значения случайной величины.
- Характеристики рассеивания - характеристики, определяющие разброс возможных значений случайной величины.

## Ц

- Центральная предельная теорема - теорема, доказывающая, что суммирование большого числа случайных величин с различными законами распределения приводит в итоге к нормальному распределению.

## Ч

- Чувствительность - это мера того, насколько одно количество реагирует на изменение другого связанного с ним количества. Чувствительность количества  $Y$ , которое зависит от изменений другого количества  $X$ , определяется как изменение  $Y$ , деленное на величину изменения  $X$ , которое вызвало изменения в  $Y$ .

## Э

- Экстремальные значения выборки - это максимальные и минимальные значения этой выборки. Статистическая теория экстремальных значений связана с оценкой распределений этих экстремальных значений для больших значений  $n$ .
- Эксцесс распределения - мера острровершинности распределения, величина, определяемая отношением центрального момента четвертого порядка к четвертой степени среднего квадратического отклонения за вычетом тройки. Эксцесс показывает, как быстро уменьшается плотность распределения вблизи её максимального значения. Для нормального распределения Гаусса эксцесс равен нулю.
- Эластичность (или нормализованная чувствительность) - это мера того, насколько одно количество реагирует на изменение другого, связанного с ним количества. Эластичность количества  $Y$ , на которое воздействуют изменения в другом количестве  $X$ , определяется в виде процентного изменения  $Y$ , деленного на процентное изменение  $X$ , которое является причиной изменения  $Y$ .
- Эффективная оценка - такая оценка, которая при заданном объеме выборки  $n$  имеет наименьшую возможную дисперсию.