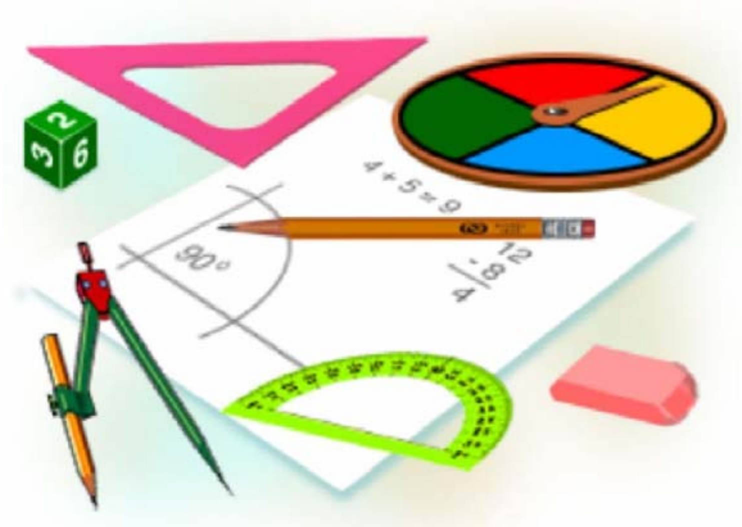
Математическая статистика

Тема: «Статистическое оценивание параметров распределения»

Введение

Математическая статистика – наука, занимающаяся методами обработки экспериментальных данных, полученных в результате наблюдений над случайными явлениями.



При решении своих задач опирается на размышляющий, оценивающий, составляющий аппарат экспериментатора.

Задачи

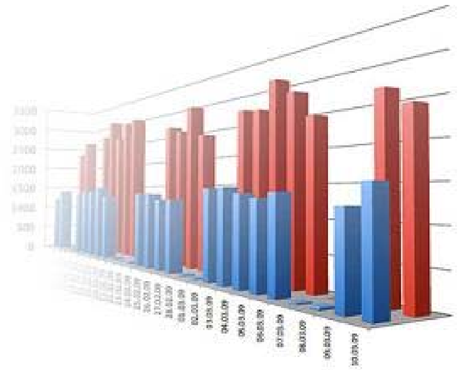
- ✓ <u>описание явлений</u> – упорядочить статистический материал, представить в удобном для экспериментатора виде (таблица, график, диаграмма);

- ✓ <u>анализ и прогноз</u> – приближенная оценка интересующих числовых событий (средняя, дисперсия) и погрешности этих величин;

- ✓ <u>выработка оптимальных решений</u> – в результате возникает задача проверки правдоподобности гипотез, решением которой является принятие или непринятие выдвинутой гипотезы.

Статистические оценки параметров распределения

- Точечные оценки
- Интервальные оценки
- Точность и надежность
- Доверительный интервал для мат.ожидания
- Доверительный интервал для оценки дисперсии



Условия

Для того, чтобы статистические оценки давали хорошее приближение оценивающих параметров, они должны удовлетворять условиям:

- объем выборки должен быть достаточным для оценивания;
- оценка интересующего нас параметра есть случайная величина.

Точечные оценки

Точечной ***называют оценку, определяющую одним числом.***

Пусть требуется изучить количественный признак генеральной совокупности. Допустим, удалось установить, какое имеется распределение. Тогда возникает задача оценки параметров данного распределения.

Пример

Однако чаще всего экспериментатору не известен вид распределения, т.к. он обладает только данными выборки и тогда для оценки параметров нужно найти зависимость этих параметров от наблюдаемых величин.

Генеральная и выборочная средняя

Генеральная и выборочная дисперсии

Статистические оценки:

- *Несмещенные* – есть оценка математического ожидания, которая равна оценивающему параметру;
- *Смещенные* – оценка $M(x)\neq$ оценивающему параметру;
- *Эффективные* – оценка, имеющая при заданном объеме выборки ***n*** наименьшую дисперсию;
- *Состоятельные* – оценка, стремящаяся при $\boldsymbol{n}\to\infty$ по вероятности к оцениваемому параметру.

Генеральная и выборочная средняя

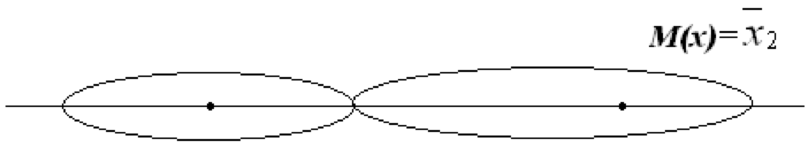
Генеральная средняя – *среднее арифметическое значений генеральной совокупности* $\bar{x}_2$

$$\bar{x}_2 = \frac{1}{N}\sum_{i=1}^{n} x_i$$

$$\bar{x}_2 = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \ldots}{N}$$ – с повторениями

Генеральная средняя есть среднее взвешенное значений генеральной совокупности с их весами, равными соответствующим частотам.

Если рассматривать ***x*** генеральной совокупности как СВ, то $M(x) = \bar{x}_2$



Генеральная и выборочная средняя

Выборочная средняя – *среднее арифметическое значений выборки.*

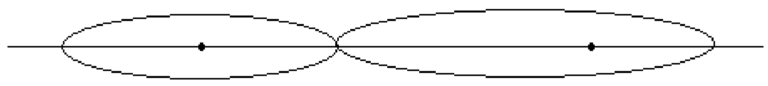
Пусть имеется выборка объема ***n***. Тогда выборочная средняя равна:

$$\bar{x}_{в} = \frac{1}{N}\sum_{i=1}^{n} x_i$$

Выборочная средняя по данным одной выборки есть определенное число.

Выборочная средняя есть *несмещенная* оценка генеральной средней.

При увеличении объема



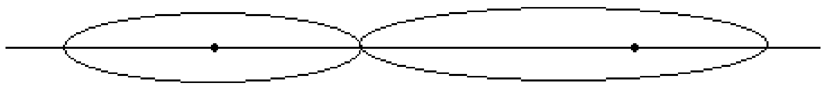
выборки ***n*** выборочная средняя стремится к генеральной средней.

Генеральная и выборочная дисперсии

Генеральной дисперсией *называют среднее арифметическое квадратов отклонений значений генеральной совокупности от их среднего значения.*

$$D_г = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x}_г)^2; \quad D_г = \sum_{i=1}^{N} N_i (x_i - \bar{x}_г)^2 \Big/ N; \quad D_г = \sum_{i=1}^{N} N(x_i - \bar{x}_г)^2 \cdot p_i;$$

Кроме дисперсий для характеристики рассеивания значений генеральной совокупности вокруг своего среднего пользуются другой характеристикой – *средним квадратическим отклонением.*



Генеральная и выборочная дисперсии

Выборочной дисперсией *называют среднее арифметическое квадратов*

$$D = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_г)^2 = S^2$$

отклонений наблюдаемых значений выборки от их среднего значения.

$S^2 = \sum_{i=1}^{n} n_i (x_i - \overline{x}_г)^2 \Big/ n$ – с повторениями

Для оценки рассеивания выборки служит выборочное среднеквадратическое отклонение.

Интервальные оценки

***Интервальной оценкой** называют оценку, определяющуюся двумя концами интервала.*

При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, что приводит к грубым ошибкам. По этой причине при небольшом объеме выборки следует пользоваться другими оценками.

Интервальные оценки позволяют определить точность и надежность оценок.

Точность и надежность

Пусть найденная по данной выборке статистическая характеристика θ^* служит оценкой неизвестного параметра θ генеральной совокупности. Будем считать θ постоянным числом. θ^* будет тем точнее определять параметр θ, чем меньше абсолютная величина разности $|\theta-\theta^*|<\varepsilon$. Чем меньше ε, тем точнее оценка.

Однако статистические методы не позволяют категорически утверждать, что θ^* удовлетворяет условию $|\theta-\theta^*|<\varepsilon$, а можно лишь говорить о вероятности, с которой это неравенство осуществляется: $P(|\theta-\theta^*|<\varepsilon)=\beta$

Точность и надежность

Надежностью *(доверительной вероятностью) оценки* θ *по* θ^* *называется вероятность* β*, с которой осуществляется неравенство* $|\theta-\theta^*|<\varepsilon$.

Обычно надежность оказывается заранее заданным числом, близким к 1. Наиболее частые значения β: 0,95; 0,98; 0,99; 0,999.

Соотношение $P(|\theta-\theta^*|<\varepsilon)=\beta$ означает вероятность того, что интервал $(\theta^*-\varepsilon;\ \theta^*+\varepsilon)$ заключает в себя (покрывает) неизвестный параметр θ, равна доверительной вероятности β.

Точность и надежность

Доверительным интервалом *называется интервал*

($\theta^*-\varepsilon$; $\theta^*+\varepsilon$), покрывающий неизвестный параметр θ с надежностью β.

Иногда вместо доверительной вероятности β используют обратную величину – уровень значимости α = 1–β. Если β – вероятность, что оцениваемый параметр попадет в интервал, α – вероятность, что не попадет. В статистических таблицах указывается именно α.

Метод доверительных интервалов разработал американский статистик Ю. Нейман, исходя из идей английского статистика Р. Фишера.

Доверительный интервал для мат.ожидания

Рассмотрим нахождение доверительного интервала для M(X) нормально распределенной СВ, т.е. нужно найти такой интервал, чтобы выполнялось следующее неравенство $\overline{X} - \varepsilon < M(X) < \overline{X} + \varepsilon$, т.е. данный интервал ширину 2ε. Он обладает симметрией.

Вероятность того, что $\left|X - \overline{X}\right| < \varepsilon$ определяется:

- законом нормального распределения, если известна $D(x)=\sigma^2$
- или распределения Стьюдента, если D(x) неизвестна, а подсчитана ее несмещенная оценка S^2.

Критерий Стьюдента определяется таким параметром как степень свободы $\upsilon=n-1$.

Для расчетов доверительных интервалов для M(X) используют два подхода:

- *когда D(x) известна*
- *когда D(x) неизвестна*

Расчет доверительных интервалов при известной дисперсии

Будем рассматривать выборочную среднюю как случайную величину. Примем без доказательств, что если СВ ***X*** распределена нормально, то и выборочная средняя по независимым наблюдениям распределена нормально.

$$P(|X–M(x)|<\varepsilon)=2\Phi(x)=\beta.$$

Таким образом для нормального распределения $2\Phi(x)=\beta$.

Параметр ***t*** в функции Лапласа: $t = \varepsilon\sqrt{n}/\sigma$

Таким образом для отыскания границ доверительного интервала:

1. по таблицам функции Лапласа находим значение аргумента ***t***, для которого $\Phi(t)=\beta/2$.
2. зная значение ***t*** из условия $t = \varepsilon\sqrt{n}/\sigma$ находим $\boldsymbol{\varepsilon}$ (граница интервала): $\varepsilon = t\sigma/\sqrt{n}$
3. Записываем доверительный интервал: $(\bar{X} - \varepsilon; \bar{X} + \varepsilon)$

Пример

Расчет доверительных интервалов при неизвестной дисперсии

Если D(x) неизвестна, а ее несмещенная оценка S^2, то в этом случае β покрытия M(x) интервалом $(\bar{X}-\varepsilon; \bar{X}+\varepsilon)$ вычисляют по закону распределения Стьюдента со степенью свободы ***υ=n–1*** и ***α=1–β***.

Имеются таблицы, которые по заданным уровням значимости и степеням определяют значения критерия Стьюдента $t_{\alpha,\upsilon}$ по формуле $t_{\alpha,\nu} = \varepsilon\sqrt{n}/S$

Таким образом доверительный интервал для *M(x)* при неизвестной дисперсии строится в виде $(\bar{X} - \frac{t_{\alpha,\nu}\cdot S}{\sqrt{n}}; \bar{X} + \frac{t_{\alpha,\nu}\cdot S}{\sqrt{n}})$

Данный интервал определяет, что с доверительной вероятностью β он покрывает истинное значение *M(x)*.

Пример

Доверительный интервал для оценки дисперсии

Доверительный интервал строится на основании того, что величина $(n-1)S^2/\sigma$ распределена по закону «хи-квадрат»(χ^2) со степенями $\upsilon = n-1$.

Выборочная дисперсия D(x) и нормального распределения связаны следующим соотношением:

$$\chi = (n-1)S^2/\sigma$$

$$\chi^2 \sigma^2 = (n-1)S^2$$

Для заданной доверительной вероятности β или, что тождественно, для заданного уровня значимости $\alpha = 1-\beta$. Потребуем, чтобы выполнялось следующее соотношение

$P(\chi^2_{\alpha/2,\upsilon} < \delta) = P(\chi^2_{1-\alpha/2} < \delta) = \alpha/2;\ \ \delta_2 < \chi^2_{\alpha/2,\upsilon} < \delta_2$

Из этого соотношения следует, что границы доверительного интервала в явном виде выглядят следующим образом:

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,\nu}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},\nu}}$$

Пример

Спасибо за внимание!

Пример 1:Пусть имеется нормальное распределение. Тогда нужно оценить, найти *M(x)* и σ. Для показательного распределения нужно оценить параметр λ. $f(x)= \lambda \cdot e^{-\lambda x}$.

- Пример 2:Случайная величина X имеет нормальное распределение с известным средним квадратическим отклонением $\sigma = 3$.

Найти доверительные интервалы для оценки неизвестного математического ожидания α по выборочным средним $\bar{x}$, если объем выборки $n = 36$ и задана надежность оценки $\gamma = 0{,}95$.

Решение. Найдем t. Из соотношения $2\Phi(t) = 0{,}95$ получим $\Phi(t) = 0{,}475$. По таблице приложения 2 находим $t=1{,}96$.
Найдем точность оценки: $\delta = {}^{t\cdot\sigma}/_{\sqrt{n}}=(1{,}96*3)/\sqrt{36}=0{,}98$
Доверительный интервал таков: ($\bar{x}$-0,98; $\bar{x}$ +0,98). Например, если $\bar{x}$= 4,1, то доверительный интервал имеет следующие доверительные границы: $\bar{x}$-0,98 = 4,1-0,98 = 3,12;
$\bar{x}$ +0,98 = 4,1+0,98 = 5,08.

- Пример 3:Количественный признак X генеральной совокупности распределен нормально. По выборке объема n = 16 найдены выборочная средняя $\bar{x}$ = 20,2 и «исправленное» среднее квадратическое отклонение s = 0,8. Оценить неизвестное математическое ожидание при помощи доверительного интервала с надежностью 0,95.

Решение. Найдем t_y. Пользуясь таблицей приложения 3, по y = 0,95 и n=16 находим t_y=2,13.

Найдем доверительные границы:

$$\bar{x} - t_y \cdot {}^{s}/_{\sqrt{n}} = 20{,}2 - 2{,}13 \cdot {}^{0{,}8}/_{\sqrt{16}} = 19{,}774$$

$$\bar{x} + t_y \cdot {}^{s}/_{\sqrt{n}} = 20{,}2 + 2{,}13 \cdot {}^{0{,}8}/_{\sqrt{16}} = 20{,}626$$

Итак, с надежностью 0,95 неизвестный параметр α заключен в доверительном интервале 19,774 < α < 20,626.

Пример 4:При доверительной вероятности 90% найти доверительный интервал для D(x), если для выборки, объемом 5 выборочная D(x)=6.6, а выборочная средняя 0.4.

По таблицам распределения χ^2 найдем значение критерия χ^2 для уровня значимости $\upsilon=4$ и $\alpha=/2=0.05$

$\chi^2_{0.05,4}=9.5$

$\chi^2_{\alpha/2,\upsilon} = \chi^2_{0.95,4}=0.711$

$4\cdot6.6/9.5 < \sigma^2 < 4\cdot6.6/0.711$

$2.78 < \sigma^2 < 37.13$