

Методы научных исследований

10 Лекция. Основы теории случайных ошибок

Исполнитель: Байболов Асан Ерболатович

Электронный адрес: asan.baibolov@kaznaru.edu.kz

ПЛАН ЛЕКЦИИ

- 1) Основные статистические характеристики;
- 2) Анализ однородности выборки

СПИСОК ЛИТЕРАТУРЫ:

1. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Множественная регрессия. Applied Regression Analysis. - 3-е изд. - М.: «Диалектика», 2007. - 912 с.
2. Радченко С. Г. Методология регрессионного анализа. — К.: «Корнийчук», 2011.
3. Сборник задач по математике: Учеб. пособие для втузов : В 4 ч. Ч. 4: Теория вероятностей. Математическая статистика / Под общ. ред. А. В. Ефимова, А. С. Пospelова. - 3-е изд., перераб. и доп. - М.: Физматлит, 2004. - 432 с.

Основные статистические характеристики

Выборка характеризуется вариационным рядом, под которым в статистике понимают ряд, построенный по количественному признаку. Любой вариационный ряд состоит из двух элементов - вариантов и частот. Вариантами считаются отдельные значения признака, т.е. конкретное значение варьирующего признака. Частоты – это числа, показывающие, как часто встречаются варианты в ряду распределения. Сумма всех частот определяет объем выборки.

Замена теоретической функции распределения на ее выборочный аналог в определении математического ожидания, дисперсии, стандартного отклонения и других характеристик приводит нас к понятиям выборочного среднего, выборочной дисперсии, выборочному стандартному отклонению и т.д. Выборочные характеристики являются оценками соответствующих характеристик генеральной совокупности, поэтому эти оценки должны удовлетворять определенным требованиям.

Среди выборочных характеристик выделяют:

- показатели, относящиеся к центру распределения;
- показатели рассеяния;
- показатели, характеризующие форму распределения.

К показателям, характеризующим центр распределения, относятся среднее значение, мода и медиана.

Мода - это элемент выборки с наиболее часто встречающимся значением.

Медиана – это число, которое является серединой выборки.

Средним значением выборки называется величина

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

Основными показателями рассеяния вариант являются: интервал, дисперсия выборки, стандартное отклонение и стандартная ошибка.

Интервал - это разница между максимальным и минимальным значениями элементов выборки. Интервал является простейшей и наименее надежной мерой вариации или рассеяния элементов в выборке..

Дисперсия выборки - это параметр, характеризующий степень разброса элементов выборки относительно среднего значения. Чем больше дисперсия, тем дальше отклоняются значения элементов выборки от среднего значения.

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

Выборочным стандартным отклонением или средним квадратичным отклонением называется величина:

$$s = \sqrt{s^2}$$

Стандартная ошибка или ошибка среднего находится из выражения:

$$\sigma_0 = \frac{s}{\sqrt{n}}$$

Стандартная ошибка - это параметр, характеризующий степень возможного отклонения среднего значения исследуемой выборки от истинного среднего значения генеральной совокупности элементов. С помощью стандартной ошибки задается так называемый доверительный интервал.

Анализ однородности выборки

Одним из важных вопросов, возникающих при анализе выборки, является вопрос: относится та или иная варианта к данной статистической совокупности? Решение вопроса не представляет сложности, если распределение в этой совокупности является нормальным. Для этого достаточно использовать правило трех сигм. Согласно этому правилу, в пределах интервала $\Delta = \bar{x} \pm 3 \cdot \sigma$ находится 99,7% всех вариантов. Поэтому, если варианта попадает в этот интервал, то она считается принадлежащей к данной совокупности. Если не попадает, то она может быть отброшена..

При числе элементов в выборке $n < 30$ границы доверительного интервала выполняются по формуле:

$$\Delta = \bar{x} \pm \sigma_0 \cdot t_{np}$$

где

$\bar{x}$ - среднее значение выборки;

σ_0 - стандартная ошибка для $n < 30$ определяется по формуле:

$$\sigma_0 = \frac{s}{\sqrt{n-1}}$$

t_{np} - табличное значение распределения Стьюдента с числом степеней свободы n и доверительной вероятностью p .

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Основные статистические характеристики;
2. Анализ однородности выборки

Спасибо за внимание !