Элементы математической статистики. Первичная обработка результатов измерений

- Генеральной совокупностью наблюдаемого признака X называется множество значений, которые может принимать наблюдаемый признак.
- Часть наблюдений, отобранных случайным образом для изучения из генеральной совокупности, называется выборкой.
- Число наблюдений n в выборке называется ее объемом.
- Первичная обработка результатов измерений осуществляется по-разному в зависимости от типа наблюдаемого признака.

- ***Ранжирование***. Если наблюдаемый признак *X* является ДСВ, то первичная обработка результатов наблюдений заключается в ранжировании, т.е. в расположении упомянутых в выборке значений в порядке возрастания с указанием количества повторов каждого значения.
- ***Интервальная обработка выборки***. Если генеральная совокупность не является дискретным множеством, то построение вариационного ряда производится при помощи интервальной обработки выборки.

Число интервалов рассчитывается по эмпирической формуле Стерджеса:

$$k \approx 1 + 3{,}322 \lg n$$

шаг h определяется по формуле $h = \dfrac{R}{k}$

где $R = x_{\max} - x_{\min}$ - размах варьирования,

$x_{\max}$ – наибольшая, а $x_{\min}$ – наименьшая варианта ряда.

В качестве x_0 выбирают $x_0 = x_{\min} - \dfrac{h}{2}$.

Графическое изображение результатов измерений. Полигон и гистограмма распределения

- Полигон служит для изображения вариационного ряда и представляет собой ломаную линию, отрезки которой последовательно соединяют точки

  где x_i - варианты, а n_i - соответствующие им частоты.

- Полигоном относительных частот называют ломаную, отрезки которой последовательно соединяют точки $(x_1; w_1)$, $(x_2; w_2)$,..., $(x_k; w_k)$

  где w_i - относительные частоты.

- Гистограмма служит для графического изображения интервальных рядов и представляет собой ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною h, а высоты равны отношению $\frac{n_i}{h}$. Площадь *i*-го прямоугольника равна $\frac{hn_i}{h} = n_i$.

  - сумме частот варианты *i*-го интервала, а площадь гистограммы частот равна сумме всех частот, т.е. объему выборки *n*.

- В случае гистограммы относительных частот высоты прямоугольников равны $\frac{w_i}{h}$, площадь i-го прямоугольника равна w_i – относительной частоте варианты i-го интервала, а площадь всей гистограммы равна сумме всех относительных частот, т. е. единице.

Эмпирическая функция распределения

- Эмпирической функцией распределения (функцией распределения выборки) называется относительная частота того, что СВ X примет значение меньше заданного *x:*

$$F_n(x) = \frac{n_x}{n}$$

где n_x - число вариант, меньших x ($x \in \mathbb{R}$), n объем выборки.

Точечное оценивание параметров

- ***Точечной оценкой параметра*** θ_i называют всякую формулу, которая по результатам выборки позволяет рассчитывать приближенное значение параметра: $\tilde{\theta}_i = \tilde{\theta}_i(X_1, X_2, ..., X_k)$

Свойства точечных оценок

- ***Несмещенность***. Точечная оценка параметра называется несмещенной, если математическое ожидание оценки совпадает с истинным значением этого параметра: $M[\tilde{\theta}_i(X_1, X_2, ..., X_k)] = \theta_i$
- ***Состоятельность***. Точечная оценка параметра называется состоятельной, если при неограниченном увеличении объема выборки СВ $\tilde{\theta}_i(X_1, X_2, ..., X_k)$ сходится по вероятности к истинному значению этого параметра.
- ***Эффективность***. Несмещенная точечная оценка параметра называется эффективной, если она имеет наименьшую дисперсию среди всех несмещенных оценок рассматриваемого параметра.

- Точечной оценкой математического ожидания является выборочная средняя:

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i \quad \text{или, если все } n_i = 1, \text{ то } \overline{x}_в = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

- Точечной оценкой дисперсии является выборочная дисперсия:

$$D_в = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x}_в\right)^2 \cdot n_i \qquad \text{или } D_в = \frac{1}{n}\sum_{i=1}^{k} x_i^{\,2} n_i - \overline{x}_в^{\,2}$$

- Несмещенной оценкой дисперсии является:

$$s^2 = \frac{n}{n-1} D_в = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_в)^2 \cdot n_i$$

- Выборочное среднее квадратическое отклонение:

$$\sigma_в = \sqrt{D_в}$$

- Исправленное среднее квадратическое отклонение:

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

- ***Модой*** *Mo* унимодального распределения вариационного ряда называется варианта, имеющая наибольшую частоту.

- ***Медианой*** *Me* вариационного ряда называется значение признака, приходящееся на середину ряда. Если $n = 2k$ (ряд имеет четное число членов), то $Me = \frac{x_k + x_{k+1}}{2}$, и если $n = 2k + 1$ (число членов ряда нечетное), то $Me = x_{k+1}$

Интервальное оценивание параметров

- Интервал $\left(\underline{\Theta}, \overline{\Theta}\right)$ называется доверительным интервалом, соответствующим доверительной вероятности γ, если вероятность того, что истинное значение параметра Θ находится на этом интервале, равна γ: $P\left(\Theta \in \left(\underline{\Theta}, \overline{\Theta}\right)\right) = \gamma$
- Доверительный интервал для математического ожидания нормального распределения при известном среднем квадратическом отклонении σ имеет вид:

$$\left(\overline{x}_в - \frac{z_\gamma}{\sqrt{n}}\sigma, \overline{x}_в + \frac{z_\gamma}{\sqrt{n}}\sigma\right)$$

z_γ определяется из уравнения: $2\Phi_0\left(z_\gamma\right) = \gamma,$ где $\Phi_0(x) = \frac{2}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$

- Доверительный интервал для математического ожидания нормального распределения при неизвестном среднем квадратическом отклонении σ имеет вид:

$$\left(\overline{x}_{в} - \frac{t_{k,\gamma}\overline{s}_{в}}{\sqrt{n}}, \overline{x}_{в} + \frac{t_{k,\gamma}\overline{s}_{в}}{\sqrt{n}}\right)$$

где $\overline{s}_{в}^{2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - a\right)^2$, $t_{k,\gamma}$ определяется по таблицам t – распределения Стьюдента при числе степеней свободы $k=n-1$ и надежности γ .