

Лабораторная работа 2. Наглядное представление выборочных данных. Оценка параметров распределения

Задание

В файле **varK.csv** находится выборка из нормального распределения $N(a, \sigma)$ (**K** - номер вашего варианта). Необходимо создать Rscript, который позволит найти основные численные характеристики выборки, построить гистограмму и функцию распределения, найти точечные и интервальные оценки параметров распределения.

Ход работы

1. Из выборки, находящейся в файле, необходимо извлечь 3 выборки меньшего объема:
 - первая выборка должна содержать элементы 1-10;
 - вторая - элементы 1-50;
 - третья - все элементы исходной выборки.
2. Для этих трех выборок необходимо:
 - вычислить среднее выборочное, выборочную дисперсию и несмещенную выборочную дисперсию, выборочные квантили;
 - найти точечные оценки параметров нормального распределения a^* и $(\sigma^2)^*$;
 - в одной системе координат построить гистограмму относительных частот и график функции плотности нормального распределения, имеющего параметры a^* и $(\sigma^2)^*$;
 - в одной системе координат построить графики эмпирической функции распределения и теоретической функции распределения (т.е. функции нормального распределения с параметрами a^* и $(\sigma^2)^*$)
3. (устно) Сравнить получившиеся графики для трех выборок и обосновать на их примере справедливость теорем
 - о связи эмпирической и теоретической функций распределения;
 - о связи гистограммы и функции плотности распределения.
4. Для всех трех выборок найдите доверительные интервалы для математического ожидания и дисперсии уровня доверия $q = 0.95$.
5. (устно) Определите взаимное расположение точечных оценок a^* и $(\sigma^2)^*$ и доверительных интервалов для параметров a и σ^2 .

6. В одной системе координат постройте для каждой из выборок графики изменения длины доверительного интервала **для a** при изменении уровня доверия q от 0.9 до 0.99999 с достаточно мелким шагом (т.е. разделив интервал изменения q не менее, чем на 100 частей)
7. В одной системе координат постройте для каждой из выборок графики изменения длины доверительного интервала **для σ^2** при изменении уровня доверия q от 0.9 до 0.99999 с достаточно мелким шагом (т.е. разделив интервал изменения q не менее, чем на 100 частей)
8. (устно) Опишите, что происходит с доверительными интервалами для параметров a и σ^2
 - при изменении объема выборки;
 - при изменении уровня доверия.

Рекомендации по выполнению работы

Для написания программы, вам могут пригодиться следующие встроенные функции:

read.csv(x) - считывает из csv файла и именем x объект типа data.frame

length(x) - возвращает длину вектора x.

range(x) - возвращает вектор из минимального и максимального значения объекта x.

seq(from, to, length.out=) - создает числовую последовательность от from до to, разделив интервал [from, to] на length.out частей.

mean(x) - возвращает среднее арифметическое значение элементов вектора x.

sd(x) - возвращает среднеквадратическое отклонение элементов вектора x.

var(x) - возвращает несмещенную выборочную дисперсию элементов вектора x.

qt(p, df) - возвращает квантиль распределения Стьюдента с df степенями свободы уровня p.

qchisq(p, df) - возвращает квантиль распределения хи-квадрат с df степенями свободы уровня p.

quantile(x, p) - возвращает выборочный квантиль для вектора x уровня p.

dnorm(x, a, s) - возвращает значение плотности нормального распределения с матожиданием a и среднеквадратическим отклонением s в точке x;

pnorm(x, a, s) - возвращает значение функции нормального распределения с матожиданием a и среднеквадратическим отклонением s в точке x;

ecdf(x) - возвращает эмпирическую функцию распределения для выборки x;

plot(obj, pch=, xlim=, ylim=, main=, xlab=, ylab=, col=) - построение графика для объекта obj на плоскости. Объект obj может быть представлен как одной переменной типа matrix или data.frame с двумя столбцами, так и двумя переменными: в этом случае первая переменная - это вектор значений абсцисс точек данных, а вторая переменная - вектор ординат. Аргументы pch, xlim, ylim, main, xlab, ylab, col задают, соответственно, символ, которым изображается точка данных, диапазон построения по оси абсцисс и ординат, название графика, подписи по осям абсцисс и ординат, цвет точек данных.

lines(x, y) - добавляет в уже созданную функцией plot() системы координат линию, соединяющую точки данных с координатами (x,y)

`hist(x, breaks=, freq=, xlim=, ylim=, col=, main=, xlab=, ylab=)` - построение гистограммы для выборки `x`. Аргумент `breaks` - это либо число столбцов гистограммы, либо вектор концов интервалов группировки, либо названия алгоритма для вычисления необходимого числа интервалов группировки. Параметр `freq` принимает значение `TRUE`, если строится гистограмма частот, `FALSE` - гистограмма относительных частот. Относительно остальных параметров см. описание функции `plot()`.

Более подробно об этих функциях читайте во встроенной справке *RStudio*. Для того, чтобы в окне *Help* появилась информация о функции из вашего кода, достаточно выделить в тексте название функции и нажать *F1*.

Рассмотрим пример построения гистограммы и функции плотности в одной системе координат:

```
x2=seq(mean(x)-5*sd(x), mean(x)+5*sd(x), length.out=200)
f1 <- dnorm(x2, a1, sqrt(s1));
hist(x, breaks= "Sturges", xlim=range(x2), ylim=c(0, 1.4*max(f1)), col='5',
     freq=FALSE, main="Гистограмма относительных частот", xlab="x", ylab="h_i")
lines(x2, f1, col='blue', lwd=2);
```

Рис. 1.1: Фрагмент кода для построения гистограммы и функции плотности нормально распределенной выборки



Рис. 1.2: Результат работы кода