Лабораторная работа № 2 «Интервальное оценивание статистических данных»
Выполнил: Саратовцев Артем 18Пи-1
Цель работы: Изучение интервального оценивания статистических данных, в том числе с использованием встроенных в базовую версию пакета R функций.
ТЕОРИЯ:
Доверительные интервалы - один из типов интервальных оценок, используемых в статистике, которые рассчитываются для заданного уровня значимости. Они позволяют сделать утверждение, что истинное значение неизвестного статистического параметра генеральной совокупности находится в полученном диапазоне значений с вероятностью, которая задана выбранным уровнем статистической значимости.

Доверительный интервал для оценки математического ожидания при известной дисперсии:

$$P\left\{\bar{x}_n - t_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le m \le \bar{x}_n + t_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha\ ,$$

где $t_{1-\alpha/2}$ - квантиль уровня 1 – α/2 стандартного нормального распределения N (0,1).

Доверительный интервал для оценки математического ожидания при неизвестной дисперсии:

$$P\left\{\bar{x}_n - t_{1-\frac{\alpha}{2},n-1}\frac{\hat{s}(\vec{x}_n)}{\sqrt{n}} \le m \le \bar{x}_n + t_{1-\frac{\alpha}{2},n-1}\frac{\hat{s}(\vec{x}_n)}{\sqrt{n}}\right\} = 1-\alpha\ ,$$

где $t_{1-\alpha/2}$ - квантиль уровня 1 – α/2 распределения Стьюдента с n – 1 степенями свободы.

Доверительный интервал для оценки дисперсии при неизвестной математическом ожидании:

$$P\left\{\frac{(n-1)\,\hat{s}^2(\vec{x}_n)}{\mathrm{X}^2_{1-\frac{\alpha}{2},n-1}} \le \sigma^2 \le \frac{(n-1)\,\hat{s}^2(\vec{x}_n)}{\mathrm{X}^2_{\frac{\alpha}{2},n-1}}\right\} = 1-\alpha\ ,$$

где $X^2_{1-\frac{\alpha}{2},n-1}$ и $X^2_{\frac{\alpha}{2},n-1}$ соответственно квантили уровня 1 – α/2 и α/2 распределения X^2 с n – 1 степенями свободы.

ХОД РАБОТЫ:
Задание 1:
Построить функции для вычисления доверительного интервала по заданной выборке и величине доверительного интервала для 1) математического ожидания при известной и неизвестной дисперсии; 2) дисперсии при неизвестном математическом ожидании.

Функция confIntMeanVar:
Значения:
arr - выборка,
S - дисперсия,
G - доверительная вероятность,
находит доверительный интервал при известных математическом ожидании и дисперсии.
Функция qnorm:
нужна для вычисления квантилей нормального распределения.
Функция sqrt:
находит квадратный корень аргумента.
Функция confIntMeanwVar:
вычисляет доверительный интервал для математического ожидания при неизвестной генеральной дисперсии.

Функция qt:

нужна для вычисления квантилей распределения Стьюдента.

Функция confIntVarwMean:

вычисляет доверительный интервал для дисперсии при неизвестном математическом среднем.

Функция qchisq:

нужна для вычисления квантилей распределения хи-квадрат.

```
confIntMeanVar <- function(arr, S, G){
# Confidence interval for mean variance
# доверительный интервал для математического ожидания при известной дисперсии
# S - Sigma (variance)
# G - Gamma
  arr=na.omit(arr)
  N = length(arr)
  d = qnorm(1-(1-G)/2) * S /sqrt(N); d # Delta

  L = mean(sample1) - d; L # Left
  R = mean(sample1) + d; R # Right
  #res = c(L, R); res
  return(c(L, R))
}
```

```
confIntMeanwVar <- function(arr, G){
# Confidence interval for mean without variance
# доверительный интервал для математического ожидания при неизвестной дисперсии
# G - Gamma
  arr=na.omit(arr)
  N = length(arr)
  t = qt(1-(1-G)/2, N-1); t
  D = var(arr)          # Sample Variance
  S = sqrt((N/(N-1))*D) # Corrected Sigma
  d = t * S /sqrt(N); d # Delta

  L = mean(sample1) - d; L # Left
  R = mean(sample1) + d; R # Right
  #res = c(L, R); res
  return(c(L, R))
}
```

```
confIntVarwMean <- function(arr, G){
# Confidence interval for variance without mean
# вычисляет доверительный интервал для дисперсии при неизвестном математическом ожидании
# G - Gamma
  arr=na.omit(arr)
  N  = length(arr)
  a1 = (1 - G)/2
  a2 = (1 + G)/2

  D = var(arr)          # Sample Variance
  S = sqrt((N/(N-1))*D) # Corrected Sigma

  L = sqrt(N-1)*S / sqrt(qchisq(a1, N-1)); L # Left
  R = sqrt(N-1)*S / sqrt(qchisq(a2, N-1)); R # Right
  #res = c(L, R); res
  return(c(L, R))
}
```

Задание 2:
Получить две выборки объема n 1 = 100 и n 2 = 200 из нормального распределения N (m, σ). Для обеих выборок найти доверительные интервалы с заданной доверительной вероятностью γ = 0. 90 с помощью полученных функций.

```
> interval1 <- confIntMeanVar(sample1, S, G); interval1
[1] -2.155536 -1.333109
> interval2 <- confIntMeanVar(sample2, S, G); interval2
[1] -2.035095 -1.453551
> interval1 <- confIntMeanwVar(sample1, G); interval1
[1] -2.158991 -1.329655
> interval2 <- confIntMeanwVar(sample2, G); interval2
[1] -2.058573 -1.430072
> interval1 <- confIntVarwMean(sample1, G); interval1
[1] 2.830948 2.238505
> interval2 <- confIntVarwMean(sample2, G); interval2
[1] 2.932492 2.485811
```

В выводе первой идёт левая граница интервала потом правая. Как можно заметить интервалы отличаются в десятых долях. Это происходит из-за разницы в способе вычисления.

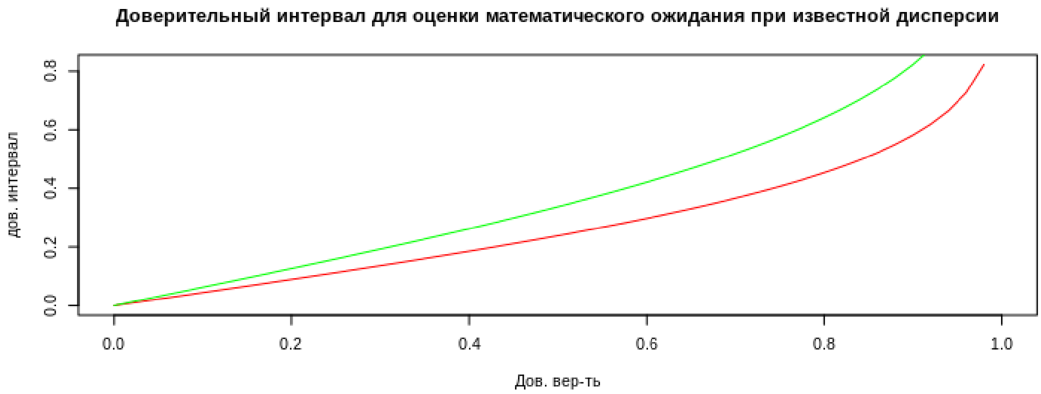
Задание 3:
Для всех трёх случаев в одной системе координат построить графики зависимости длины доверительного интервала от величины доверительной вероятности. При этом γ придать минимум 50 разных значений.
Проанализировать взаимное расположение полученных графиков и объяснить его.

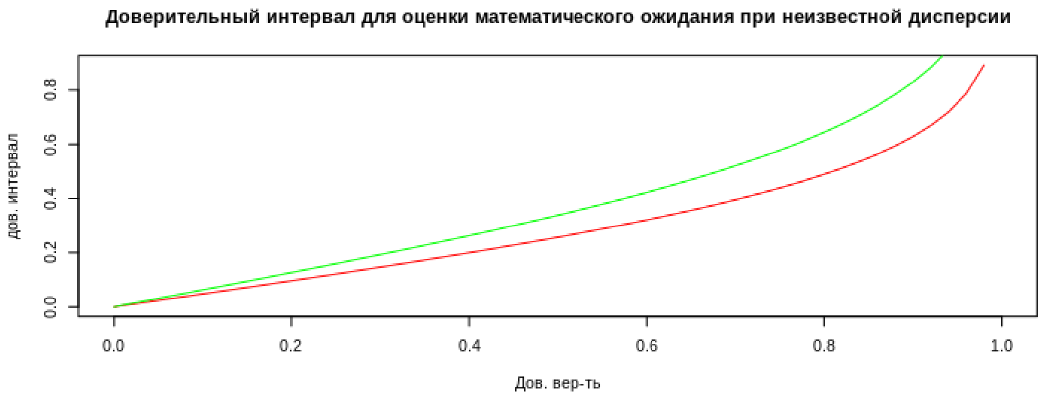
Процесс построения графиков (перед выполнением необходимо достаточно широко открыть окно plot):

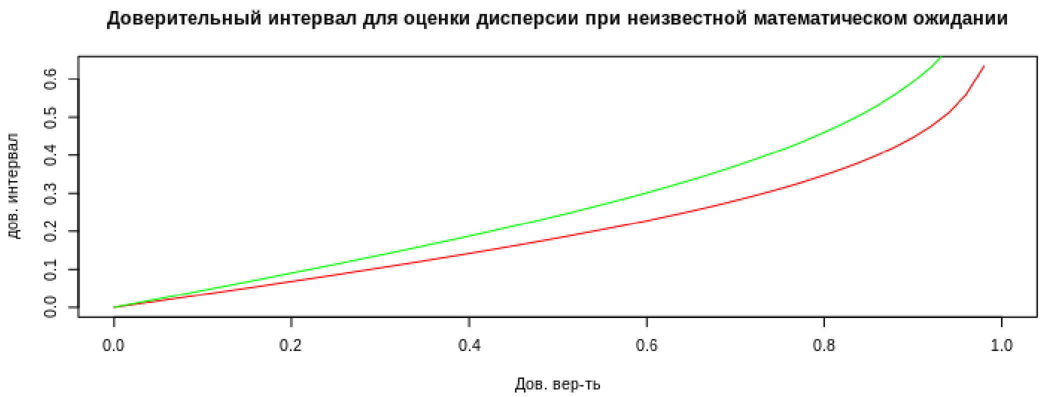
```
intLen <- function(int){
#Длинна интервала
  len <- abs(max(int) - min(int))
  return (len)
}


plot1 <- function(arr1, arr2, S){
  arr1=na.omit(arr1); arr2=na.omit(arr2)
  confProps <- seq(from=0, to=1, by=0.02); confProps
  lengths1 <- c(1:length(confProps))
  lengths2 <- c(1:length(confProps))
  for(i in 1: length(confProps)){
    lengths1[i] <- intLen(confIntMeanVar(arr1, S, confProps[i]))
    lengths2[i] <- intLen(confIntMeanVar(arr2, S, confProps[i]))
  }
  plot(confProps, lengths2,  col = "red", type = "l",
       xlab = "Дов. вер-ть",
       ylab = "дов. интервал",
       main = "Доверительный интервал для оценки математического ожидания при известной дисперсии")
  lines(confProps, lengths1,  col = "green")
}


plot2 <- function(arr1, arr2, S){
  arr1=na.omit(arr1); arr2=na.omit(arr2)
  confProps <- seq(from=0, to=1, by=0.02); confProps
  lengths1 <- c(1:length(confProps))
  lengths2 <- c(1:length(confProps))
  for(i in 1: length(confProps)){
    lengths1[i] <- intLen(confIntMeanwVar(arr1, confProps[i]))
    lengths2[i] <- intLen(confIntMeanwVar(arr2, confProps[i]))
  }
  plot(confProps, lengths2,  col = "red", type = "l",
       xlab = "Дов. вер-ть",
       ylab = "дов. интервал",
       main = "Доверительный интервал для оценки математического ожидания при неизвестной дисперсии")
  lines(confProps, lengths1,  col = "green")
}
```

```
plot3 <- function(arr1, arr2, S){
  arr1=na.omit(arr1); arr2=na.omit(arr2)
  confProps <- seq(from=0, to=1, by=0.02); confProps
  lengths1 <- c(1:length(confProps))
  lengths2 <- c(1:length(confProps))
  for(i in 1: length(confProps)){
    lengths1[i] <- intLen(confIntVarwMean(arr1, confProps[i]))
    lengths2[i] <- intLen(confIntVarwMean(arr2, confProps[i]))
  }
  plot(confProps, lengths2,  col = "red", type = "l",
       xlab = "Дов. вер-ть",
       ylab = "дов. интервал",
       main = "Доверительный интервал для оценки дисперсии при неизвестной математическом ожидании")
  lines(confProps, lengths1,  col = "green")
}
> par(mfcol=c(3, 1))
> plot1(sample1,sample2, S)
> plot2(sample1,sample2, S)
> plot3(sample1,sample2, S)
```

Результат выполнения кода:







Анализ графиков:
Зеленая функция — результат работы функций для выборки объемом 100 элементов.
Красная функция — результат работы функций для выборки объемом 200 элементов.
Как видно из графиков при увеличении объёма выборки длины уменьшились. Это показывает, что при увеличении объема выборки можно получить меньший по длине интервал. Это происходит из-за того, что при расчётах в знаменателе стоит корень из длины выборки.