Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Кафедра «Прикладная математика»

ОТЧЁТ ПО ЛАБОРАТОРНЫМ РАБОТАМ ПО ДИСЦИПЛИНЕ «МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

Выполнил
студент группы 3630102/70201

Крупкина Дарья

Проверил
к. ф.-м. н., доцент

Баженов Александр Николаевич

Санкт-Петербург
2020

Содержание

Список таблиц

1 Постановка задачи

1.1 Задание

Для двух выборок размерами 20 и 100 элементов, сгенерированных согласно нормальному закону $N(x, 0, 1)$, для параметров положения и масштаба построить асимптотически нормальные интервальные оценки на основе точечных оценок метода максимального правдоподобия и классические интервальные оценки на основе статистик χ^2 и Стьюдента. В качестве параметра надёжности взять $\gamma = 0.95$.

2 Теория

2.1 Доверительные интервалы для параметров нормального распределения

2.1.1 Доверительный интервал для математического ожидания m нормального распределения

Дана выборка $(x_1, x_2, ..., x_n)$ объёма n из нормальной генеральной совокупности. На её основе строим выборочное среднее $\bar{x}$ и выборочное среднее квадратическое отклонение s. Параметры m и σ нормального распределения неизвестны.
Доказано, что случайная величина

$$T = \sqrt{n-1}\frac{\bar{x} - m}{s} \tag{1}$$

называемая статистикой Стьюдента, распределена по закону Стьюдента с $n-1$ степенями свободы. Пусть $f_T(x)$ — плотность вероятности этого распределения. Тогда

$$P\left(-x < \sqrt{n-1}\frac{\bar{x}-m}{s} < x\right) = P\left(-x < \sqrt{n-1}\frac{m-\bar{x}}{s} < x\right) = \\ = \int_{-x}^{x} f_T(t)dt = 2\int_{0}^{x} f_T(t)dt = 2\left(\int_{-\infty}^{x} f_T(t)dt - \frac{1}{2}\right) = 2F_T(x) - 1 \tag{2}$$

Здесь $F_T(x)$ — функция распределения Стьюдента с $n-1$ степенями свободы.
Полагаем $2F_T(x) - 1 = 1 - \alpha$, где α — выбранный уровень значимости. Тогда $F_T(x) = 1 - \alpha/2$. Пусть $t_{1-\alpha/2}(n1)$ — квантиль распределения Стьюдента с $n-1$ степенями свободы и порядка $1 - \alpha/2$. Из предыдущих равенств мы получаем

$$P\left(\bar{x} - \frac{sx}{\sqrt{n-1}} < m < \bar{x} + \frac{sx}{\sqrt{n-1}}\right) = 2F_T(x) - 1 = 1 - \alpha, \\ P\left(\bar{x} - \frac{st_{1-\alpha/2}(n1)}{\sqrt{n-1}} < m < \bar{x} + \frac{st_{1-\alpha/2}(n1)}{\sqrt{n-1}}\right) = 1 - \alpha, \tag{3}$$

что и даёт доверительный интервал для m с доверительной вероятностью $\gamma = 1 - \alpha$ [1, с. 457-458].

2.1.2 Доверительный интервал для среднего квадратического отклонения σ нормального распределения

Дана выборка $(x_1, x_2, ..., x_n)$ объёма n из нормальной генеральной совокупности. На её основе строим выборочную дисперсию s^2. Параметры m и σ нормального распределения неизвестны. Доказано, что случайная величина ns^2/σ^2 распределена по закону χ^2 с $n-1$ степенями свободы.
Задаёмся уровнем значимости α и находим квантили $\chi^2_{\alpha/2}(n-1)$ и $\chi^2_{1-\alpha/2}(n-1)$.
Это значит, что

$$\begin{gathered} P\left(\chi^2(n-1) < \chi^2_{\alpha/2}(n-1)\right) = \alpha/2, \\ P\left(\chi^2(n-1) < \chi^2_{1-\alpha/2}(n-1)\right) = 1-\alpha/2 \end{gathered} \tag{4}$$

Тогда

$$\begin{multline} P\left(\chi^2_{\alpha/2}(n-1) < \chi^2(n-1) < \chi^2_{1-\alpha/2}(n-1)\right) = \\ = P\left(\chi^2(n-1) < \chi^2_{1-\alpha/2}(n-1)\right) - P\left(\chi^2(n-1) < \chi^2_{\alpha/2}(n-1)\right) = \\ = 1-\alpha/2-\alpha/2 = 1-\alpha \end{multline} \tag{5}$$

Отсюда

$$\begin{multline} P\left(\chi^2_{\alpha/2}(n-1) < \frac{ns^2}{\sigma^2} < \chi^2_{1-\alpha/2}(n-1)\right) = P\left(\frac{1}{\chi^2_{1-\alpha/2}(n-1)} < \frac{\sigma^2}{ns^2} < \frac{1}{\chi^2_{\alpha/2}(n-1)}\right) = \\ = P\left(\frac{s\sqrt{n}}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}} < \sigma < \frac{s\sqrt{n}}{\sqrt{\chi^2_{\alpha/2}(n-1)}}\right) = 1-\alpha \end{multline} \tag{6}$$

Окончательно

$$P\left(\frac{s\sqrt{n}}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}} < \sigma < \frac{s\sqrt{n}}{\sqrt{\chi^2_{\alpha/2}(n-1)}}\right) = 1-\alpha, \tag{7}$$

что и даёт доверительный интервал для σ с доверительной вероятностью $\gamma = 1-\alpha$ [1, с. 458-459].

2.2 Доверительные интервалы для математического ожидания m и среднего квадратического отклонения σ произвольного распределения при большом объёме выборки. Асимптотический подход

При большом объёме выборки для построения доверительных интервалов может быть использован асимптотический метод на основе центральной предельной теоремы.

2.2.1 Доверительный интервал для математического ожидания m произвольной генеральной совокупности при большом объёме выборки

Выборочное среднее $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ при большом объёме выборки является суммой большого числа взаимно независимых одинаково распределённых случайных величин. Предполагаем, что исследуемое генеральное распределение имеет конечные математическое ожидание m и дисперсию σ^2. Тогда в силу центральной предельной теоремы центрированная и нормированная случайная величина $(\bar{x} - M\bar{x})/\sqrt{D\bar{x}} = \sqrt{n}(\bar{x} - m)/\sigma$ распределена приблизительно нормально с параметрами 0 и 1. Пусть

$$\Phi(x) = \frac{1}{2\pi}\int_{-\infty}^{x} e^{-t^2/2}dt \tag{8}$$

- функция Лапласа. Тогда

$$P\left(-x < \sqrt{n}\frac{\bar{x} - m}{\sigma} < x\right) = P\left(-x < \sqrt{n}\frac{m - \bar{x}}{\sigma} < x\right) \approx \\ \approx \Phi(x) - \Phi(-x) = \Phi(x) - [1 - \Phi(x)] = 2\Phi(x) - 1 \tag{9}$$

Отсюда

$$P\left(\bar{x} - \frac{\sigma x}{\sqrt{n}} < m < \bar{x} - \frac{\sigma x}{\sqrt{n}}\right) \approx 2\Phi(x) - 1 \tag{10}$$

Полагаем $2\Phi(x) - 1 = \gamma = 1 - \alpha$; тогда $\Phi(x) = 1 - \alpha/2$. Пусть $u_{1-\alpha/2}$ — квантиль нормального распределения N(0,1) порядка $1 - \alpha/2$. Заменяя в равенстве (10) σ на s, запишем его в виде

$$P\left(\bar{x} - \frac{s u_{1-\alpha/2}}{\sqrt{n}} < m < \bar{x} - \frac{s u_{1-\alpha/2}}{\sqrt{n}}\right) \approx \gamma, \tag{11}$$

что и даёт доверительный интервал для m с доверительной вероятностью $\gamma = 1 - \alpha$ [1, с. 460].

2.2.2 Доверительный интервал для среднего квадратического отклонения σ произвольной генеральной совокупности при большом объёме выборки

Выборочная дисперсия $s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$ при большом объёме выборки является суммой большого числа практически взаимно независимых случайных величин (имеется одна связь $\sum_{i=1}^{n} x_i = n\bar{x}$, которой при большом n можно пренебречь). Предполагаем, что исследуемая генеральная совокупность имеет конечные первые четыре момента. В силу центральной предельной теоремы центрированная и нормированная случайная величина $(s^2 - Ms^2)/\sqrt{Ds^2}$ при большом объёме выборки n распределена приблизительно нормально с параметрами 0 и 1. Пусть $\Phi(x)$ — функция Лапласа (8). Тогда

$$P\left(-x < \frac{s^2 - Ms^2}{\sqrt{Ds^2}} < x\right) \approx \Phi(x) - \Phi(-x) = \Phi(x) - [1 - \Phi(x)] = 2\Phi(x) - 1 \tag{12}$$

Положим $2\Phi(x) - 1 = \gamma = 1 - \alpha$. Тогда $\Phi(x) = 1 - \alpha/2$. Пусть $u_{1-\alpha/2}$ — корень этого уравнения — квантиль нормального распределения N(0,1) порядка $1-\alpha/2$. Известно, что $Ms^2 = \sigma^2 - \frac{\sigma^2}{n} \approx \sigma^2$ и $Ds^2 = \frac{\mu_4 - \mu_2^2}{n} + o(\frac{1}{n}) \approx \frac{\mu_4 - \mu_2^2}{n}$. Здесь μ_k — центральный момент k-го порядка генерального распределения; $\mu_2 = \sigma^2$; $\mu_4 = M[(x - Mx)^4]$; $o(\frac{1}{n})$— бесконечно малая высшего порядка, чем 1/n, при $n \to \infty$. Итак, $Ds^2 \approx \frac{\mu_4 - \mu_2^2}{n}$. Отсюда

$$Ds^2 \approx \frac{\sigma^4}{n}(\frac{\mu_4}{\sigma^4} - 1) = \frac{\sigma^4}{n}((\frac{\mu_4}{\sigma^4} - 3) + 2) = \frac{\sigma^4}{n}(E + 2) \approx \frac{\sigma^4}{n}(e + 2), \tag{13}$$

где E $= \frac{\mu_4}{\sigma^4} - 3$ — эксцесс генерального распределения, e $= \frac{m_4}{s^4} - 3$ — выборочный эксцесс; $m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4$ — четвёртый выборочный центральный момент. Далее,

$$\sqrt{Ds^2} \approx \frac{\sigma^2}{\sqrt{n}}\sqrt{e + 2} \tag{14}$$

Преобразуем неравенства, стоящие под знаком вероятности в формуле
$P\left(-x < \frac{s^2 - Ms^2}{\sqrt{Ds^2}} < x\right) = \gamma$:

$$\begin{aligned}
-\sigma^2 U < s^2 - \sigma^2 < \sigma^2 U;\\
\sigma^2(1 - U) < s^2 < \sigma^2(1 + U);\\
1/[\sigma^2(1 + U)] < 1/s^2 < 1/[\sigma^2(1 - U)];\\
s^2/(1 + U) < \sigma^2 < s^2/(1 - U);\\
s(1 + U)^{-1/2} < \sigma < s(1 - U)^{-1/2},
\end{aligned} \tag{15}$$

где $U = u_{1-\alpha/2}\sqrt{(e + 2)/n}$ или
$s(1 + u_{1-\alpha/2}\sqrt{(e + 2)/n})^{-1/2} < \sigma < s(1 - u_{1-\alpha/2}\sqrt{(e + 2)/n})^{-1/2}$.
Разлагая функции в биномиальный ряд и оставляя первые два члена, получим

$$s(1 - 0.5U) < \sigma < s(1 + 0.5U) \tag{16}$$

или

$$s(1 - 0.5u_{1-\alpha/2}\sqrt{(e + 2)/n}) < \sigma < s(1 + 0.5u_{1-\alpha/2}\sqrt{(e + 2)/n}) \tag{17}$$

Формулы (15) или (17) дают доверительный интервал для σ с доверительной вероятностью $\gamma = 1 - \alpha$ [1, с. 461-462].
Замечание. Вычисления по формуле (15) дают более надёжный результат, так как в ней меньше грубых приближений.

3 Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm. Использованы библиотеки scipy для генерации выборки и вычисления коэффициентов. Исходный код лабораторной работы приведён в приложении в виде ссылки на репозиторий GitHub.

4 Результаты

4.1 Доверительные интервалы для параметров нормального распределения

n = 20	m	σ
	-0.48 $< m <$ 0.55	0.84 $< \sigma <$ 1.55
n = 100	m	σ
	-0.12 $< m <$ 0.29	0.91 $< \sigma <$ 1.21

Таблица 1: Доверительные интервалы для параметров нормального распределения

4.2 Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

n = 20	m	σ
	-0.44 $< m <$ 0.51	0.97 $< \sigma <$ 1.23
n = 100	m	σ
	-0.12 $< m <$ 0.29	0.97 $< \sigma <$ 1.11

Таблица 2: Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

5 Обсуждение

- Генеральные характеристики ($m = 0$ и $\sigma = 1$) накрываются построенными доверительными интервалами.
- Также можно сделать вывод, что для большей выборки доверительные интервалы являются соответственно более точными, т.е. меньшими по длине.
- Доверительные интервалы для параметров нормального распределения более надёжны, так как основаны на точном, а не асимптотическом распределении.

6 Приложения

Код программы на GitHub, URL: `https://github.com/DariaKrup/Statistics`

Список литературы

[1] Вероятностные разделы математики. Учебник для бакалавров технических направлений.//Под ред. Максимова Ю.Д. — Спб.: «Иван Федоров», 2001. — 592 с., илл.

[2] Вентцель Е.С. Теория вероятностей: Учеб. для вузов. — 6-е изд. стер. — М.: Высш. шк., 1999.— 576 с.