Научно-технологический университет «Сириус» Научный центр информационных технологий и искусственного интеллекта

Отчёт

по заданию №4

Дисциплина: методы математичес	кой статистики для обработки данных
Выполнил	
Студент группы М01МР-24	A. C. Кондратьев
Преподаватель	Ю. В. Якубович

Санкт-Петербург

Задание

В файле записаны результаты измерений радиусов планет в формате < номер варианта>, < планета>, < наблюдение>.

- 1. Выделить данные назначенного варианта (вариант №5).
- 2. Для каждой планеты найти точечную оценку среднего и дисперсии наблюдений.
- 3. Для каждой планеты найти правосторонний (вида $[a, +\infty)$) и центральный доверительные интервалы для среднего наблюдений на уровнях доверия 95% и 99%.
- 4. Для каждой планеты найти левосторонний и центральный доверительный интервалы для дисперсий наблюдений на уровнях доверия 95% и 99%.
- 5*. Для каждой планеты найти кратчайший интервал, в который с вероятностью 0.95 попадет следующее измерение радиуса. Для этого необходимо применить теорему Фишера и подобрать константу c, чтобы случайная величина $\frac{c(X_{n+1}-\mu)}{S}$, где μ выборочное среднее, а S^2 выборочная дисперсия, построенные по имеющимся n наблюдениям, имела известное распределение, и воспользоваться квантилями этого распределения. Для имеющихся данных найти количество измерений, которые попали в этот интервал.
- 6. Какие из полученных результатов необходимо использовать, чтобы с вероятностью ошибки не более 1% найти нижнюю границу радиуса каждой из планет?

1 Выделение данных по назначенному варианту

Выделим из исходных данных те, которые соответствуют варианту 5. В выборке содержатся наблюдения для четырех планет. Количество наблюдений для Венеры – 19, для Меркурия – 18, для Земли – 21, для Марса – 27.

2 Расчет точечной оценки среднего и дисперсии наблюдений

Выборочное среднее для полученных данных рассчитывается по формуле

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1}$$

где n – количество наблюдений, X_i – отдельное наблюдение (радиус планеты).

Несмещенная выборочная дисперсия рассчитывается по формуле

$$S^{2} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{i} - \bar{X})^{2}$$
 (2)

Результаты расчетов выборочного среднего и выборочной дисперсии представлены в таблице 1.

Таблица 1 — Результаты расчетов выборочного среднего и выборочной дисперсии

Планета	$ar{X}$	S^2
Венера	6048.40	1573.03
Меркурий	2445.95	906.95
Земля	6362.42	763.76
Mapc	3379.71	1207.06

3 Расчет доверительных интервалов для среднего

Правосторонний доверительный интервал имеет следующие границы

$$\left[\bar{X} - t_{\alpha} \cdot \frac{S}{\sqrt{n}}, +\infty\right) \tag{3}$$

где t_{α} – квантиль распределения для выбранного уровня доверия, S – выборочное стандартное отклонение.

Центральный доверительный интервал имеет следующие границы

$$\left[\bar{X} - t_{\frac{1+\alpha}{2}} \cdot \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{1+\alpha}{2}} \cdot \frac{S}{\sqrt{n}}\right] \tag{4}$$

Результаты расчетов границ правостороннего и центрального доверительных интервалов среднего с уровнями доверия 0.95, 0.99 для каждой планеты представлены в таблице 2.

Таблица 2 — Результаты расчетов границ доверительных интервалов среднего с уровнями доверия 0.95, 0.99

Планета		Правосторонний	Центральный	
	Уровень	доверительный интервал	доверительный интервал	
	доверия	Правая граница	Правая	Левая
			граница	граница
Венера	0.95	6064.18	6029.28	6067.52
	0.99	6071.62	6022.21	6074.59
Меркурий	0.95	2458.30	2430.98	2460.93
	0.99	2464.17	2425.38	2466.53
Земля	0.95	6372.82	6349.84	6375.00
	0.99	6377.67	6345.26	6379.58
Марс	0.95	3391.12	3365.97	3393.46
	0.99	3396.29	3361.13	3398.29

4 Расчет доверительных интервалов для дисперсий наблюдений

Левосторонний доверительный интервал для дисперсии имеет следующие границы

$$\left[0, \frac{(n-1)S^2}{\chi_\alpha^2}\right] \tag{5}$$

где χ_{α}^2 – квантиль хи-квадрат распределения для уровня доверия α .

Центральный доверительный интервал для дисперсии имеет следующие границы

$$\[\frac{(n-1)S^2}{\chi_{\alpha/2}^2}, \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2} \]$$
 (6)

Результаты расчетов границ левостороннего и центрального доверительных интервалов дисперсий с уровнями доверия 0.95, 0.99 для каждой планеты представлены в таблице 3.

Таблица 3 — Результаты расчетов границ доверительных интервалов дисперсий с уровнями доверия 0.95, 0.99

Планета		Левосторонний	Центральный	
	Уровень	доверительный интервал	доверительный интервал	
	доверия	Левая граница	Правая	Левая
			граница	граница
Венера	0.95	3015.25	898.12	3440.10
	0.99	4036.35	762.04	4519.63
Меркурий	0.95	1777.97	510.68	2038.30
	0.99	2406.16	431.66	2706.25
Земля _	0.95	1407.76	447.04	1592.69
	0.99	1849.21	381.91	2054.82
Марс	0.95	2040.66	748.60	2266.97
	0.99	2572.82	649.90	2812.10

5 Поиск кратчайшего интервала, в который с вероятностью 0.95 попадет следующее измерение радиуса планеты

Кратчайший интервал для нового наблюдения имеет следующие границы

$$\left[\bar{X} - c \cdot S_p, \bar{X} + c \cdot S_p\right] \tag{7}$$

где c — квантиль распределения для уровня доверия 0.95, задающий ширину интервала, S_p — стандартная ошибка предсказания, которая вычисляется по формуле

$$S_p = S\sqrt{1 + \frac{1}{n}}\tag{8}$$

Результаты расчетов границ кратчайшего интервала для нового наблюдения представлены в таблице 4.

Таблица 4 - Границы кратчайшего интервала для нового наблюдения

Планета	Правая граница	Левая граница
Венера	5962.91	6133.89
Меркурий	2380.67	2511.23
Земля	6303.42	6421.43
Марс	3306.99	3452.44

6 Результаты, необходимые для поиска нижней границы радиуса каждой из планет с вероятностью ошибки не более 1%

Чтобы с вероятностью ошибки не более 1% найти нижнюю границу радиуса каждой из планет, необходимо использовать односторонний доверительный интервал для среднего радиуса. Поскольку необходимо найти именно нижнюю границу, доверительный интервал будет правосторонним, т.к. он позволяет определить, что с вероятностью 99% истинное значение будет больше или равно нижней границе, т.е. вероятность ошибки будет не более 1%.