

Доклад

Научные статистические расчеты. Программное обеспечение для
реализации статистических расчётов.

Смирнов-Мальцев Е. Д.

Содержание

| | |
|--|----|
| Цель работы | 4 |
| Теоретическое введение | 5 |
| Комплекс используемых программ | 5 |
| Подсчет основных статистических метрик | 6 |
| Создание случайной выборки | 6 |
| Проверка статистических гипотез | 8 |
| Выводы | 9 |
| Заключение | 10 |
| Список литературы | 11 |

Список иллюстраций

| | | |
|---|-----------------------------------|---|
| 1 | Гистограмма выборки X | 6 |
|---|-----------------------------------|---|

Цель работы

- Выделить задачи научной статистики,
- Выполнить подсчет основных статистических метрик.

Теоретическое введение

Статистика — совокупность числовых показателей, характеризующих те или иные явления и процессы. Статистика имеет дело, прежде всего с количественной стороной явлений и процессов. Статистические данные языком цифр характеризуют размеры и количественные соотношения (объемы, структуру, темпы развития и т.п.) явлений и проявляющиеся в них закономерности.

Важнейшими элементами статистической методологии являются: массовое наблюдение, группировки, применение обобщающих (сводных) характеристик. Статистическое исследование включает в себя: - разработку программы статистического наблюдения (определение объекта, единицы и формы наблюдения, разработку методик расчета запрашиваемых показателей и предполагаемые результаты обработки полученных данных); - сбор массовых данных о статистической совокупности (непосредственно статистическое наблюдение); - обработку данных (сводку, группировку); - анализ полученной информации.

Комплекс используемых программ

Для подсчета статистических метрик в данной работе был использован Octave с пакетом statistics. Эта программная система была выбрана, поскольку:

1. Имеет совместимый с Matlab язык.
2. Бесплатная.

Подсчет основных статистических метрик

Создание случайной выборки

В качестве статистической выборки X возьмем 1000 нормально распределенных случайных чисел. Построим гистограмму полученной выборки. Также создадим выборку Y с выбросом. Ей будет выборка X к которой добавлено число 100.

```
> X = randn(1,1000);  
> hist(X, 30, 'FaceColor', 'yellow', 'EdgeColor', 'red', 'LineWidth', 2);  
> Y = X;  
> Y(1,1001) = 100;
```

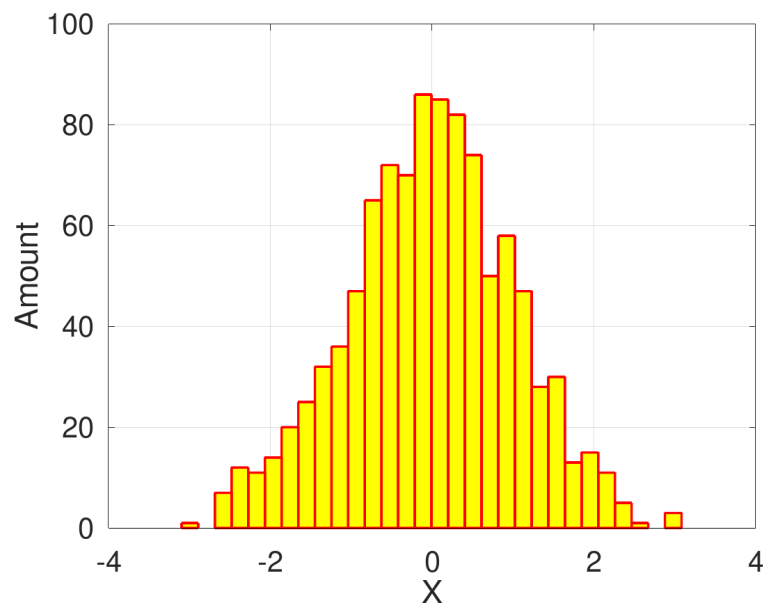


Рис. 1: Гистограмма выборки X

Среднее значение и медиану можно посчитать с помощью встроенных функций `mean` и `median`:

```
> mean(X)
ans = 0.0427
> median(X)
ans = 0.058198
> mean(Y)
ans = 0.14256
> median(Y)
ans = 0.059815
```

Как видно из эксперимента, медиана более устойчива к выбросам, чем среднее значение. Поэтому в случае если в выборке возможны выбросы, лучше использовать ее.

Стандартное отклонение можно посчитать с помощью встроенной функции `std(x,f)`. Если параметр `f` равен 0, то считается несмещенное отклонение. Если он равен 1, то смещенное.

```
> std(X,0)
ans = 1.0165
> std(X,1)
ans = 1.0160
> std(Y,0)
ans = 3.3187
> std(Y,1)
ans = 3.3170
```

Проверка статистических гипотез

Важной частью статистических расчетов является проверка статистических гипотез. В качестве примера, проверим является ли случайная величина Y нормально распределенной с найденным нами средним и стандартным отклонением равным 1.

```
> kolmogorov_smirnov_test(Y, "norm", mean(Y), 1)
pval: 0.00655609
```

pval — вероятность получить такую же или более экстремальную выборку, при условии, что гипотеза верна. Если pval меньше критического уровня, то мы можем отвергнуть гипотезу. Обычно критический уровень берут равным 0.05. В таком случае мы можем отвергнуть гипотезу. Однако в некоторых областях используют другой критический уровень. Например, в ядерной физике используется правило 5 сигм, по которому pval должен быть меньше $3 \cdot 10^{-7}$. В таком случае мы не можем отвергнуть нашу гипотезу.

Проверим является ли случайная величина Y нормально распределенной со средним значением равным медиане Y и стандартным отклонением 1.

```
> kolmogorov_smirnov_test(Y, "norm", median(Y), 1)
pval: 0.40823
```

Эту гипотезу мы отвергнуть не можем.

Выводы

Мы посчитали основные статистические метрики и провели проверку статистических гипотез.

Заключение

Для исследования модели Хищник-жертва в нашей работе будут использованы метод Эйлера и методы Рунге-Кутты, а программная реализация будет выполнена в системе математических вычислений Octave.

Список литературы