

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ им. ПЕТРА ВЕЛИКОГО

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ И МЕХАНИКИ

КАФЕДРА ПРИКЛАДНОЙ МАТЕМАТИКИ

ЛАБОРАТОРНАЯ РАБОТА №5  
ВЫБОРОЧНЫЕ КОЭФФИЦИЕНТЫ  
КОРРЕЛЯЦИИ И ЭЛЛИПСЫ  
РАСSEИВАНИЯ

Студент группы 3630102/70301

Камянский Д.В.

Преподаватель

Баженов А. Н.

САНКТ-ПЕТЕРБУРГ  
2020 г.

# Содержание

|                             |    |
|-----------------------------|----|
| 1. Список иллюстраций ..... | 3  |
| 2. Список таблиц .....      | 3  |
| 3. Постановка задачи .....  | 4  |
| 4. Теория.....              | 4  |
| 5. Реализация.....          | 4  |
| 6. Результаты .....         | 5  |
| 7. Выводы .....             | 9  |
| 8. Список литературы .....  | 10 |
| 9. Приложения .....         | 10 |

## 1 Список иллюстраций

|   |  |   |
|---|--|---|
| 1 | Графики двумерного нормального распределения и смеси для размера выборки $n = 20$ .....  | 7 |
| 2 | Графики двумерного нормального распределения и смеси для размера выборки $n = 60$ .....  | 8 |
| 3 | Графики двумерного нормального распределения и смеси для размера выборки $n = 100$ ..... | 9 |

## 2 Список таблиц

|   |   |   |
|---|---|---|
| 1 | Двумерное нормальное распределение, $n = 20$ .....  | 5 |
| 2 | Двумерное нормальное распределение, $n = 60$ .....  | 5 |
| 3 | Двумерное нормальное распределение, $n = 100$ ..... | 6 |
| 4 | Смесь нормальных распределений .....                | 6 |

### 3 Постановка задачи

Сгенерировать двумерные выборки размерами 20, 60, 100 для двумерного нормального распределения  $N(x, y, 0, 0, 1, 1, \rho)$ .

Коэффициент корреляции  $\rho$  взять равным 0, 0.5, 0.9.

Каждая выборка генерируется 1000 раз и для неё вычисляются: среднее значение, среднее значение квадрата и дисперсия коэффициентов корреляции Пирсона, Спирмена и квадрантного коэффициента корреляции. Повторить все вычисления для смеси нормальных распределений:

$$f(x, y) = 0.9N(x, y, 0, 0, 1, 1, 0.9) + 0.1N(x, y, 0, 0, 10, 10, -0.9) \quad (1)$$

Изобразить сгенерированные точки на плоскости и нарисовать эллипс равновероятности.

### 4 Теория

1. Двумерное нормально распределение:

$$N(x, y, \bar{x}, \bar{y}, \sigma_x, \sigma_y, \rho) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \\ \times \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\bar{x})^2}{\sigma_x^2} - 2\rho\frac{(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y} + \frac{(y-\bar{y})^2}{\sigma_y^2}\right]\right) \quad (2)$$

2. Коэффициент корреляции Пирсона:

$$r = \frac{\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2 \frac{1}{n} \sum (y_i - \bar{y})^2}} \quad (3)$$

3. Квадрантный коэффициент корреляции:

$$r_Q = \frac{(n_1 + n_3) - (n_2 + n_4)}{n} \quad (4)$$

где  $n_1, n_2, n_3, n_4$  – количества точек с координатами  $(x_i, y_i)$ , попавшими соответственно в I, II, III и IV квадранты декартовой системы с осями  $x' = x - medx, y' = y - medy$  и с центром в точке с координатами  $(medx, medy)$

4. Коэффициент корреляции Спирмана:

$$r_S = \frac{\frac{1}{n} \sum (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\frac{1}{n} \sum (u_i - \bar{u})^2 \frac{1}{n} \sum (v_i - \bar{v})^2}} \quad (5)$$

где  $u$  и  $v$  – ранги, соответствующие значениям переменной  $X$  и  $Y$  соответственно.

### 5 Реализация

Работы была выполнена на языке *Python3.8.2* Для генерации выборок использовался модуль *stats* библиотеки *scipy*. Для построения графиков использовалась библиотека *matplotlib*.

## 6 Результаты

Таблица 1: Двумерное нормальное распределение,  $n = 20$

|              |         |          |       |
|--------------|---------|----------|-------|
| $\rho = 0$   | Pearson | Spearman | Quad  |
| $E(z)$       | 0.009   | 0.001    | 0.004 |
| $E(z^2)$     | 0.05    | 0.05     | 0.05  |
| $D(z)$       | 0.05    | 0.05     | 0.05  |
| $\rho = 0.5$ | Pearson | Spearman | Quad  |
| $E(z)$       | 0.49    | 0.46     | 0.32  |
| $E(z^2)$     | 0.27    | 0.25     | 0.15  |
| $D(z)$       | 0.03    | 0.03     | 0.05  |
| $\rho = 0.9$ | Pearson | Spearman | Quad  |
| $E(z)$       | 0.893   | 0.865    | 0.69  |
| $E(z^2)$     | 0.801   | 0.754    | 0.5   |
| $D(z)$       | 0.003   | 0.05     | 0.03  |

Таблица 2: Двумерное нормальное распределение,  $n = 60$

|              |         |          |         |
|--------------|---------|----------|---------|
| $\rho = 0$   | Pearson | Spearman | Quad    |
| $E(z)$       | -0.003  | -0.004   | -0.0004 |
| $E(z^2)$     | 0.02    | 0.2      | 0.02    |
| $D(z)$       | 0.02    | 0.02     | 0.02    |
| $\rho = 0.5$ | Pearson | Spearman | Quad    |
| $E(z)$       | 0.497   | 0.47     | 0.33    |
| $E(z^2)$     | 0.257   | 0.24     | 0.13    |
| $D(z)$       | 0.009   | 0.01     | 0.02    |
| $\rho = 0.9$ | Pearson | Spearman | Quad    |
| $E(z)$       | 0.8984  | 0.883    | 0.706   |
| $E(z^2)$     | 0.8078  | 0.782    | 0.508   |
| $D(z)$       | 0.0007  | 0.001    | 0.009   |

Таблица 3: Двумерное нормальное распределение,  $n = 100$

|              |         |          |        |
|--------------|---------|----------|--------|
| $\rho = 0$   | Pearson | Spearman | Quad   |
| $E(z)$       | -0.004  | -0.004   | -0.001 |
| $E(z^2)$     | 0.01    | 0.01     | 0.01   |
| $D(z)$       | 0.01    | 0.01     | 0.01   |
| $\rho = 0.5$ | Pearson | Spearman | Quad   |
| $E(z)$       | 0.499   | 0.481    | 0.329  |
| $E(z^2)$     | 0.254   | 0.237    | 0.118  |
| $D(z)$       | 0.005   | 0.006    | 0.009  |
| $\rho = 0.9$ | Pearson | Spearman | Quad   |
| $E(z)$       | 0.8981  | 0.8841   | 0.708  |
| $E(z^2)$     | 0.8071  | 0.7823   | 0.506  |
| $D(z)$       | 0.0004  | 0.0007   | 0.004  |

Таблица 4: Смесь нормальных распределений

|           |         |          |       |
|-----------|---------|----------|-------|
| $n = 20$  | Pearson | Spearman | Quad  |
| $E(z)$    | -0.09   | -0.09    | -0.06 |
| $E(z^2)$  | 0.05    | 0.06     | 0.06  |
| $D(z)$    | 0.05    | 0.05     | 0.05  |
| $n = 60$  | Pearson | Spearman | Quad  |
| $E(z)$    | -0.09   | -0.004   | -0.06 |
| $E(z^2)$  | 0.03    | 0.02     | 0.02  |
| $D(z)$    | 0.02    | 0.02     | 0.02  |
| $n = 100$ | Pearson | Spearman | Quad  |
| $E(z)$    | -0.094  | -0.079   | -0.06 |
| $E(z^2)$  | 0.018   | 0.016    | 0.01  |
| $D(z)$    | 0.009   | 0.009    | 0.01  |

Рис. 1: Графики двумерного нормального распределения и смеси для размера выборки  $n = 20$

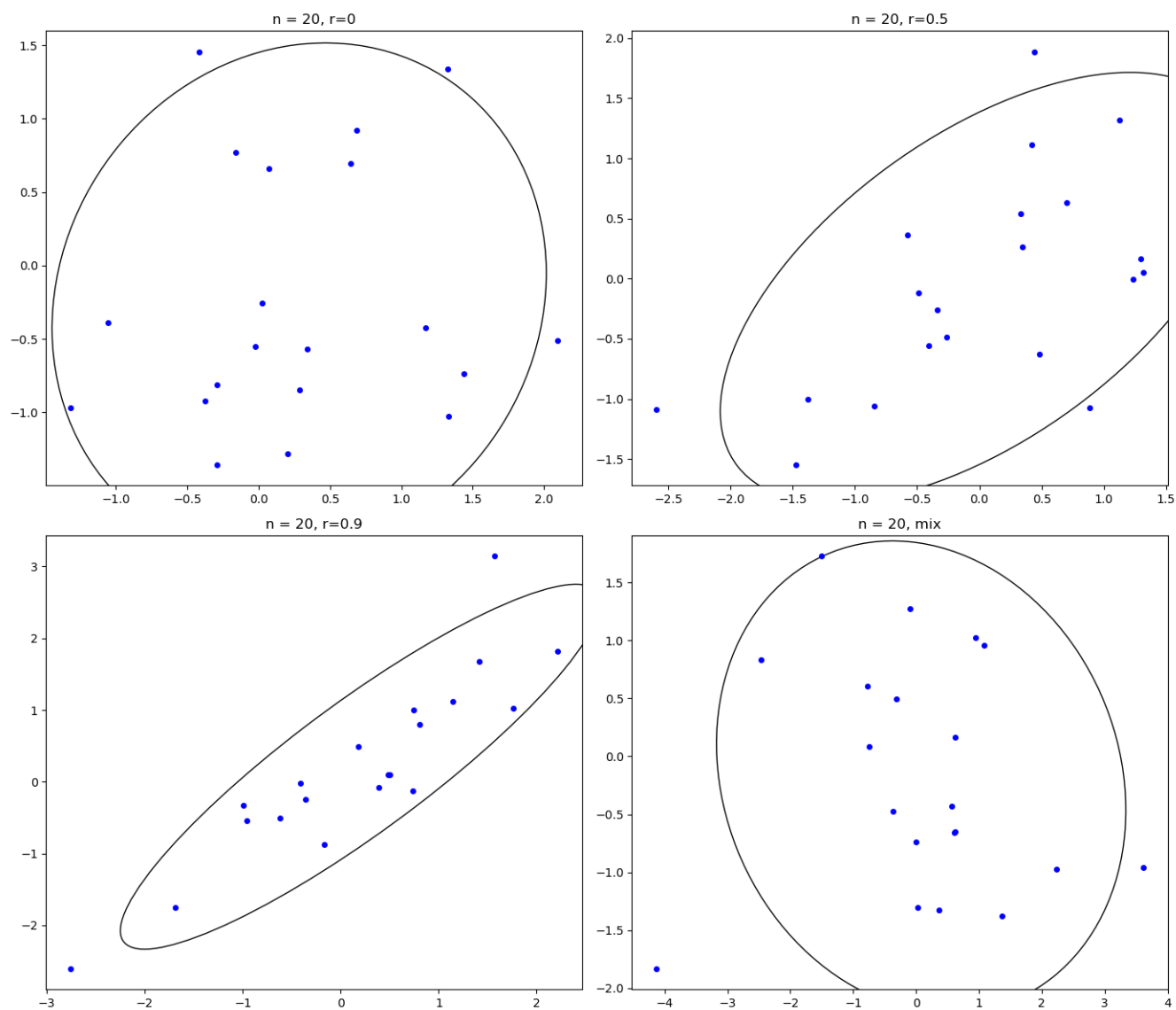


Рис. 2: Графики двумерного нормального распределения и смеси для размера выборки  $n = 60$

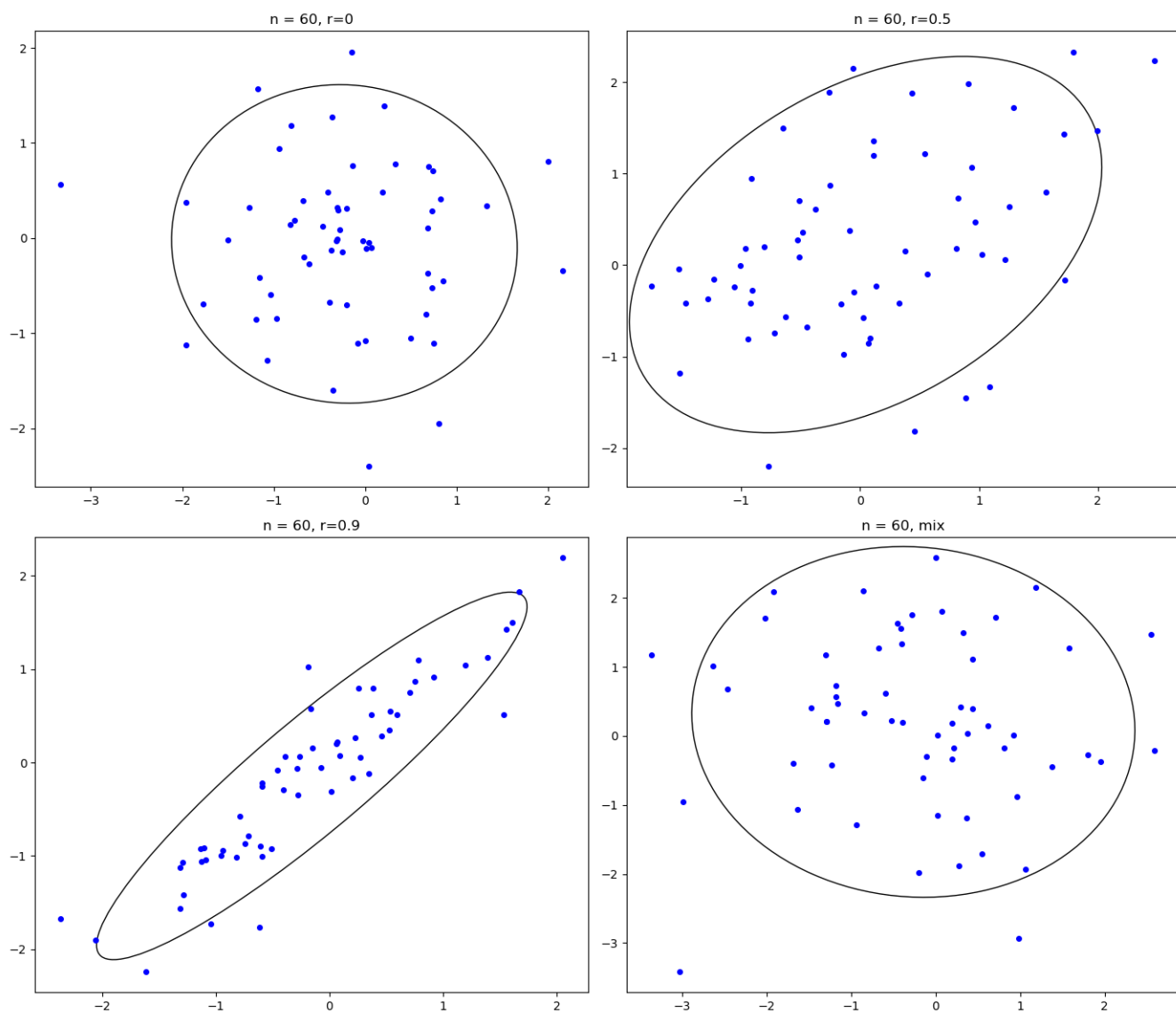
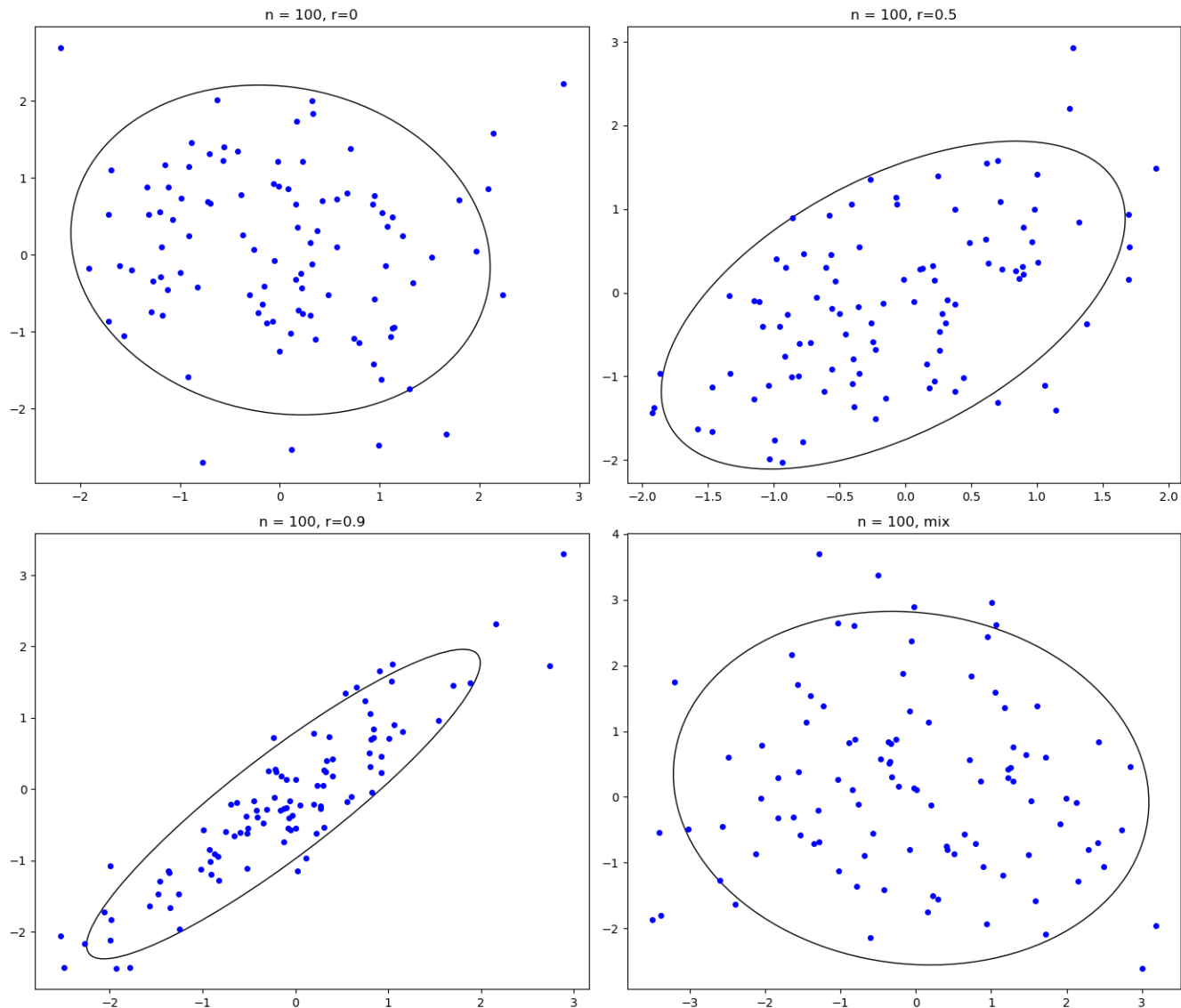




Рис. 3: Графики двумерного нормального распределения и смеси для размера выборки  $n = 100$



## 7 Выводы

По таблицам ??, ??, 4, ??, видно, что, при увеличении объёма выборки, подсчитанные коэффициенты корреляции стремятся к теоретическим.

Ближе всех к данному коэффициенту корреляции находится коэффициент Пирсона.

По графикам видно, что при уменьшении корреляции эллипс равновероятности стремится к окружности, а при увеличении растягивается.

## 8 Список литературы

- [1] Модуль numpy - <https://physics.susu.ru/vorontsov/language/numpy.html>
- [2] Модуль matplotlib - <https://matplotlib.org/users/index.html>
- [3] Модуль scipy - <https://docs.scipy.org/doc/scipy/reference/>
- [4] Двумерное нормальное распределение: [https://en.wikipedia.org/wiki/Multivariate\\_normal\\_distribution](https://en.wikipedia.org/wiki/Multivariate_normal_distribution)
- [5] Коэффициент корреляции Пирса: <http://statistica.ru/theory/koeffitsient-korreljatsii/>
- [6] Коэффициент корреляции Спирмана: [http://economic-definition.com/Exchange\\_Terminology/Koefficient\\_korreljicii\\_Correlation\\_coefficient\\_\\_eto.html](http://economic-definition.com/Exchange_Terminology/Koefficient_korreljicii_Correlation_coefficient__eto.html)

## 9 Приложения

Код отчёта: <https://github.com/MisterProper9000/MatStatLabs/blob/master/MatStatLab5/MatStatLab5.tex>

Код лабораторной: <https://github.com/MisterProper9000/MatStatLabs/blob/master/MatStatLab5/MatStatLab5.py>