

# СВОДКА И ГРУППИРОВКА

## Сводка

- *Сводка* – обобщение данных для выявления типичных черт и закономерностей
  - средний балл, общая посещаемость, процент отличников
  - При статистическом наблюдении собираются данные по каждой единице
  - При сводке получают обобщенные данные по совокупности в целом
- Результаты сводки представляют в форме статистических таблиц

## Определения

- *простая сводка*
  - подсчет итогов по совокупности в целом
- *группировка*
  - деление совокупности на группы по существенным признакам
- *сложная сводка*
  - группировка единиц
  - подсчет итогов по каждой группе

## Виды статистических группировок

- Цель группировки
  - *оценка распределения* по фактическим данным
  - для изучения общих закономерностей
- типологическая
  - типы явлений
- структурная
  - структура явления
- аналитическая
  - анализ связи между явлениями и признаками

## Типологическая группировка

- деление *разнородной* совокупности на классы, типы, однородные группы единиц
- задача: отследить зарождение, развитие и отмирание типов
  - Пример. Группировка предприятий
    - по отраслям промышленным
    - по видам продукции
    - по форме собственности

## Структурная группировка

- деление *однородной* совокупности
- задача: изучить структуру явления
  - Пример. Состав населения
    - по полу
    - по возрасту
    - по месту проживания
  - Пример. Группировка предприятий
    - по численности работников
    - по стоимости основных фондов

## Аналитическая группировка

- деление *однородной* совокупности по факторному признаку
  - признаки
    - факторные (причина)
    - результативные (следствие)
- задача: *анализ зависимости* результата от фактора
  - *изменение* фактора сопровождается *изменением* результата

## Домашнее задание

- простая и сложная группировка

## Группирующие признаки

- атрибутивные – качественные свойства
  - выражаются словами (КАКОЙ?)
    - пол: М/Ж
    - форма собственности: АО, ФГУП, ИП
      - » *атрибут* – свойство
      - » [L. *attributum* = *at-*, *ad-* (к-, при-) + *tribuere* (дань, налог)]  
= «данное, приписанное»
- вариационные – количественные
  - измеряются числами (СКОЛЬКО?)
    - вес, рост, зарплата

## Аналитическая группировка по вариационному признаку

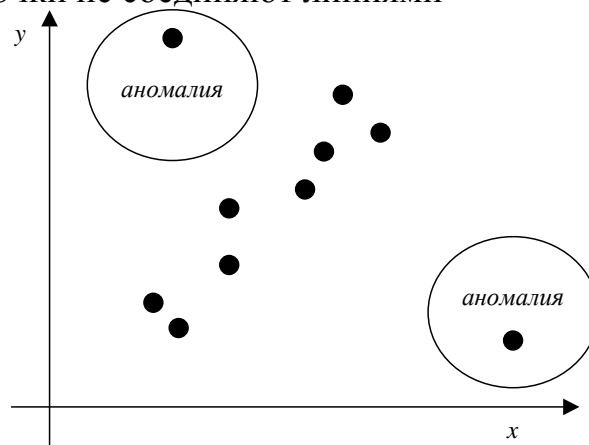
- Диапазон значений факторного признака делится на интервалы
  - число групп (интервалов)  $k$  зависит от числа единиц совокупности  $n$
  - формула Стерджесса  $k = 1 + 3,322 \cdot \lg n$ 
    - чем больше единиц в группе, тем ниже погрешность
    - нежелательно иметь пустые или малочисленные группы
  - величина интервала 
$$h = \frac{R}{k} = \frac{x_{\max} - x_{\min}}{k}$$
  - верхняя и нижняя границы интервалов

## Удаление аномальных наблюдений

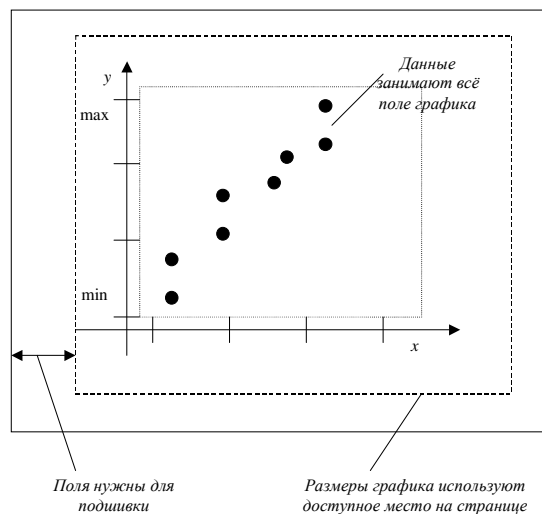
- Аномалия (выброс)
  - отклонение от статистической закономерности
  - выпадение из общей картины
    - причины
      - погрешности при сборе и обработке данных
        - » грубые ошибки
        - » искажение отчетности
      - неоднородность совокупности
        - » объекты другого типа
      - Аномалия – отклонение от нормы, нарушение закона, закономерности, неправильный, необычный
        - » [L. *anomalía*
        - » Гр.  $\alpha\nu\omicron\mu\omicron\tau\omicron\varsigma$  = *an-* (не) + *homos* (одинаковый, такой же)]
  - аномалии удаляют *перед* группировкой данных

## Диаграмма разброса

- диаграмма рассеяния
- точки не соединяют линиями



## Оформление графика



## Построение графика

- Поля и отступы страницы
- Место для графика
- Оси координат
- Обозначения осей и единицы измерения
- Масштаб по осям
  - min ... max
  - округляем до «стандартных» чисел
- Точки

## Заголовки: «Что? Где? Когда?»

- Заголовки документа, графика, таблицы

Для однозначного понимания *читателем*:

- *Что?*

- *Изучаемое явление, объект исследования*

- *Где?*

- *Место, предприятие*

- *Когда?*

- *Время, дата, год*

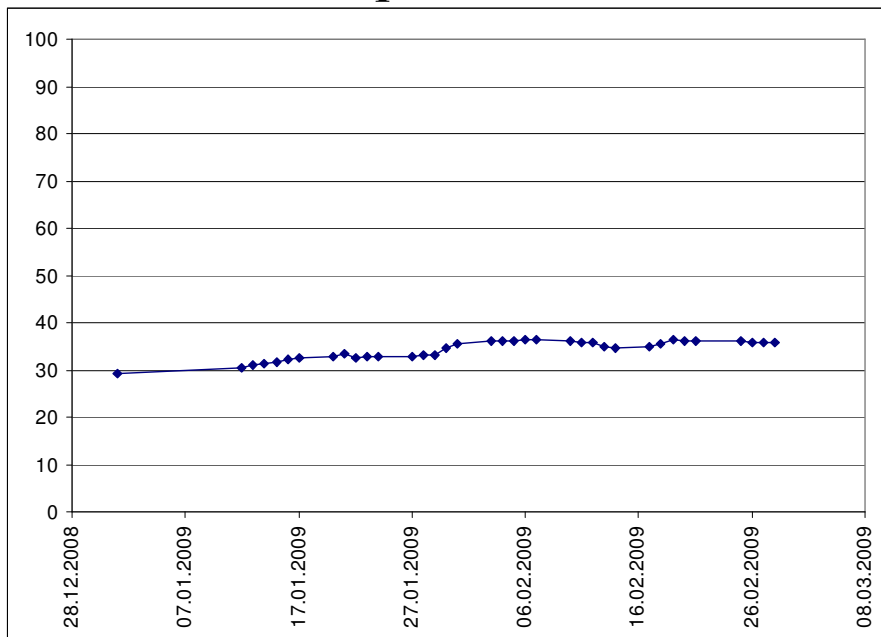
- *Дополнительные сведения*

- *условия сбора данных*
  - *особенности работы предприятия*

## Пример оформления графика



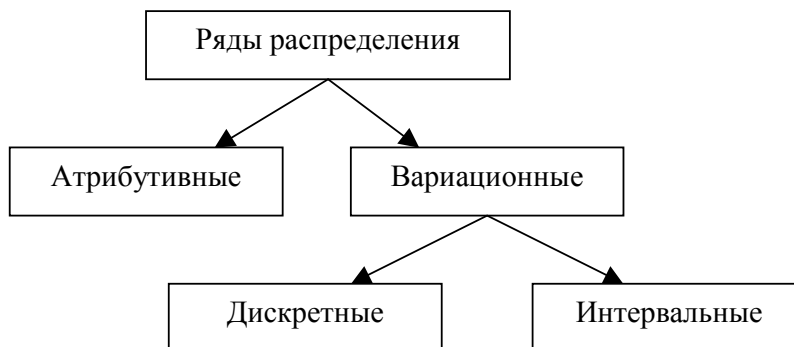
## Выбор масштаба



## Ряды распределения

- Ряды распределения позволяют оценить закон распределения по результатам наблюдения
  - результат сводки и группировки данных
  - оценка формы и параметров распределения

# Классификация



## Результат группировки данных

| Значения признака | Частота (размер группы) | Частость (относительная частота) | Накопленные (кумулятивные) частоты                                       | Условное среднее |
|-------------------|-------------------------|----------------------------------|--------------------------------------------------------------------------|------------------|
| $X_1$             | $n_1$                   | $n_1 \%$                         | $\begin{matrix} \longrightarrow K_1 \\ \longrightarrow K_2 \end{matrix}$ | $Y_1$            |
| $X_2$             | $n_2$                   | $n_2 \%$                         |                                                                          | $Y_2$            |
| ...               | ...                     | ...                              | ...                                                                      | ...              |
| $X_k$             | $n_k$                   | $n_k \%$                         | 100%                                                                     | $Y_k$            |
| Всего             | $n$                     | 100%                             | —                                                                        | —                |

## Частость

- относительная частота
- доля единиц совокупности
- оценка вероятности

$$n_i (\%) = \frac{n_i}{n} \cdot 100\% = \hat{p}_i$$

## Накопленные частоты

- кумулята = «накопление»  
» [*L. cumulus* (куча, масса) + *-ate* (обр. глагола от сущ.)]
- последнее значение = 1 или 100%

$$K_1 = n_1$$

$$K_2 = n_1 + n_2 = K_1 + n_2$$

$$K_3 = n_1 + n_2 + n_3 = K_2 + n_3$$

...

$$K_k = \sum_{i=1}^k n_i = K_{k-1} + n_k = 1$$

## Условное среднее

- Среднее значение результативного признака для единиц группы

$$\bar{y}_k = \hat{\mu}(y | x_k)$$

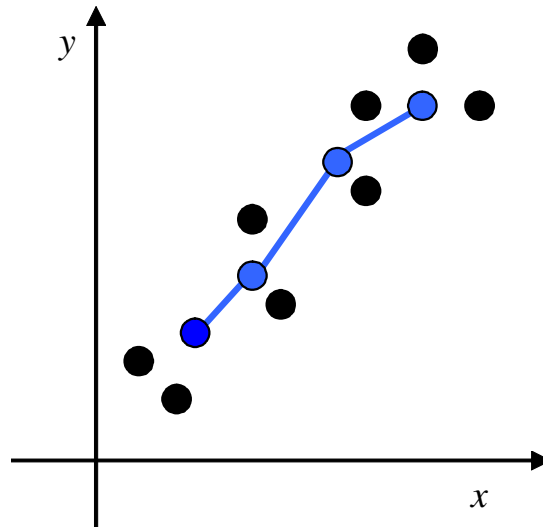
| – «при условии», «при»

$$\bar{y}_k = \frac{\sum_{i=1}^{n_k} y_i}{n_k} \quad \forall i: \quad x_i \in (инт.)$$

## Пример группировки

| $X_i$    | $n_i$ | $n_i (\%)$ | $K_i (\%)$ | $Y_i$ |
|----------|-------|------------|------------|-------|
| 10 .. 20 | 5     | 25%        | 25%        | 21    |
| 20 .. 30 | 10    | 50%        | 75%        | 24    |
| 30 .. 40 | 5     | 25%        | 100%       | 32    |
| Всего    | 20    | 100%       | —          | —     |

## Линия условного среднего



## Графическое изображение рядов распределения

- Оценка функции плотности вероятности
  - гистограмма
  - полигон
- Оценка функции распределения
  - кумулята
  - огива

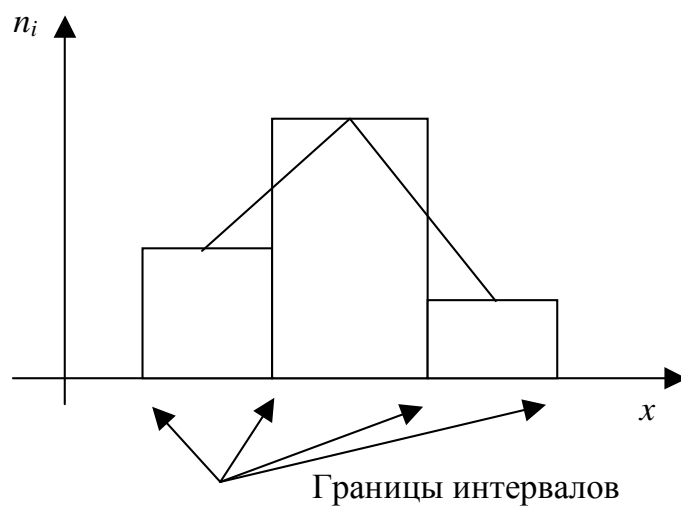
# Гистограмма

- столбиковая диаграмма частот
- столбик –диапазон (интервал) значений
- высота
  - частость
  - частота
    - плотность вероятности = частота в расчете на единицу ширины интервала.
      - » [Gr. *histos* (живая ткань, материал) + *gra-* (изображение)] = «рисование кусочками»
      - » гистология – наука о тканях живого организма
      - » *history* + *gram*

# Полигон

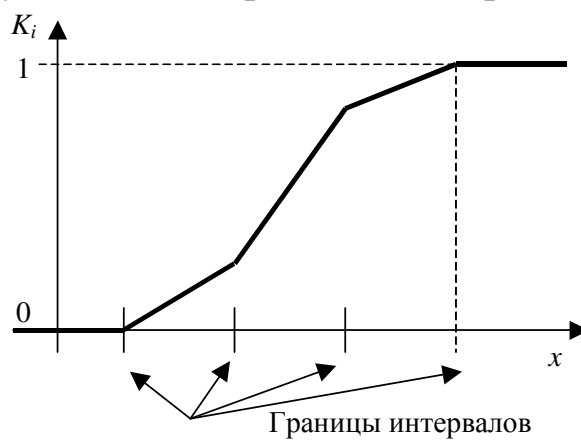
- график в виде ломаной линии
- варианты (середины интервалов) и частоты
  - Полигон можно построить по гистограмме, соединив середины верхних сторон прямоугольников
    - » [**Gr.** *poly* (много) + *gonia* (угол) = многоугольник, ломаная линия]

## Гистограмма и полигон



## Кумулята

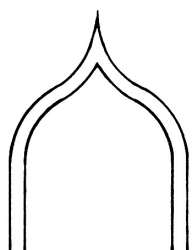
- Кумулята – интеграл от гистограммы



## Огива

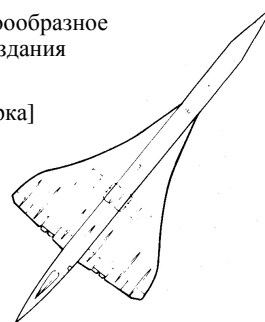
- *Огива* – изображение кумуляты, где оси меняются местами
  - зеркальное отображение относительно главной диагонали

- » [Е. *ogive* – стрельчатый свод, дугообразное перекрытие, соединяющее стены здания
- » Е. *ogee* – S-образная кривая
- » Fr. *augive* – диагональная арка]

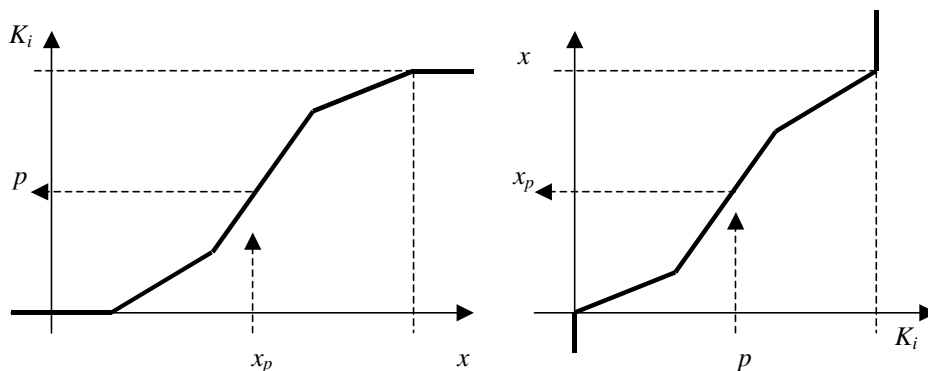


*ogee arch* (keel arch)

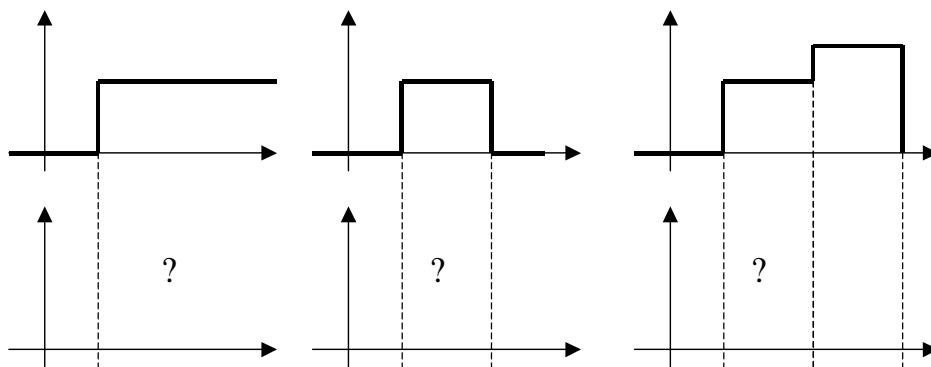
*ogee wing* (ogival wing)



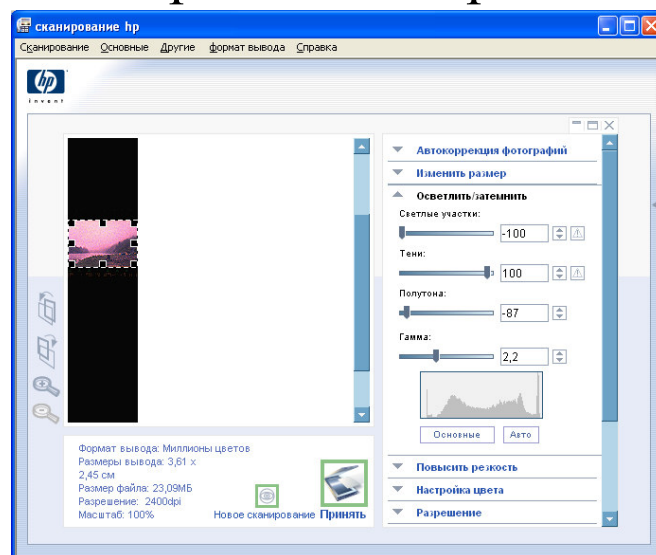
## Кумулята и огива



## Задача: Интеграл от графика



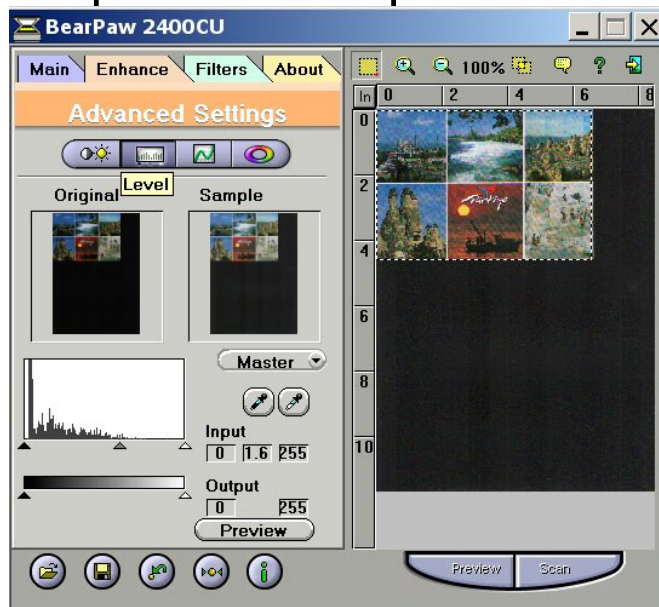
## Гистограмма в комп. графике: Настройка сканера HP



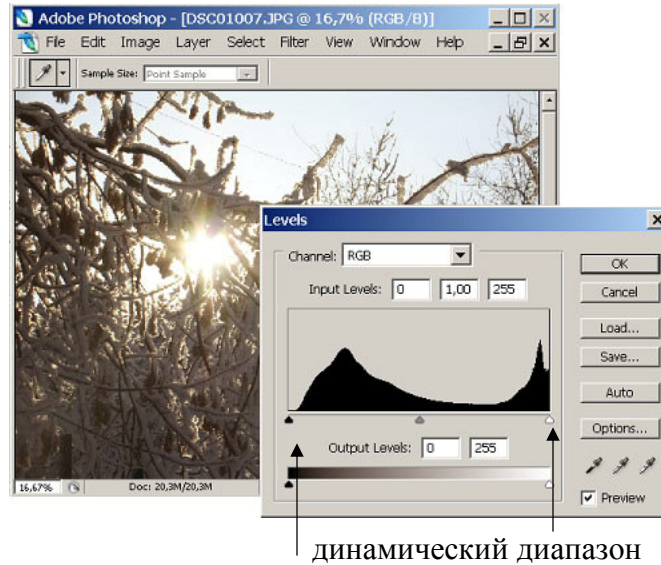
## Гистограмма: Состав изображения



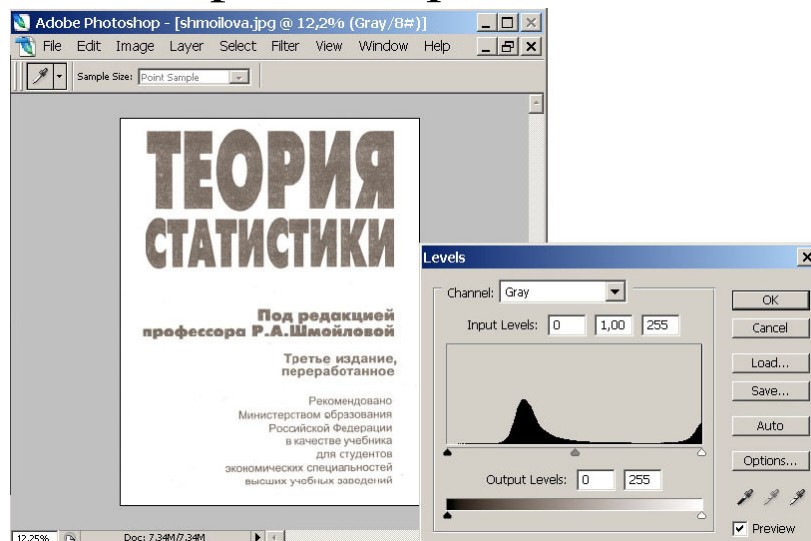
## Настройка сканера MUSTEK



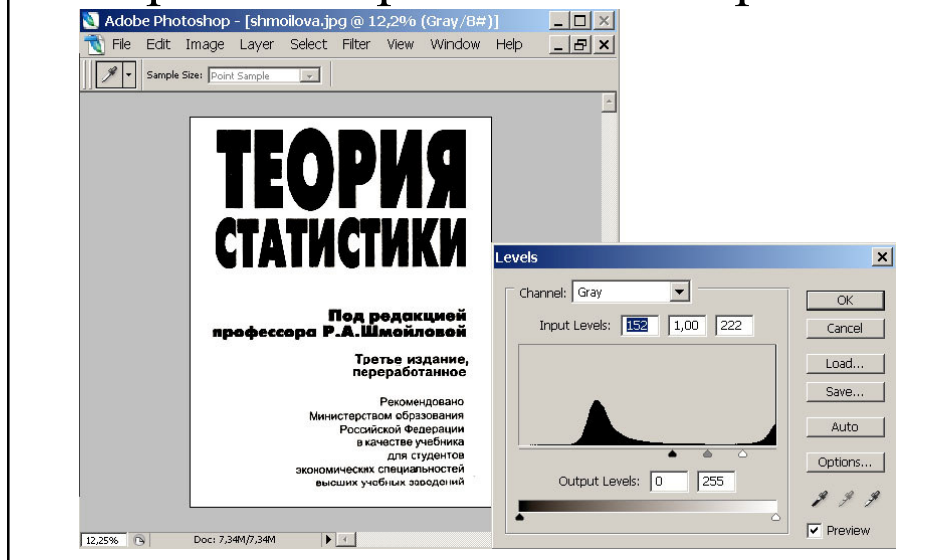
## Photoshop: Image – Adjustments – Levels



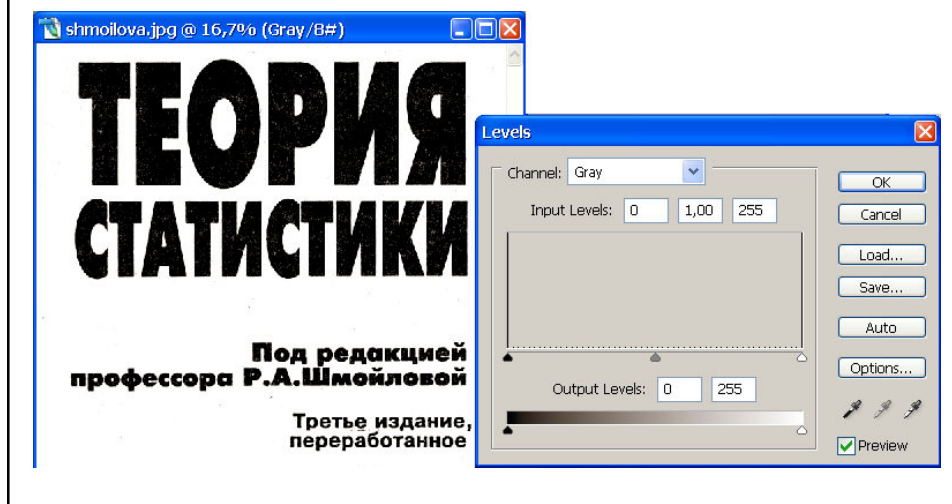
## Photoshop: Сканирование текста



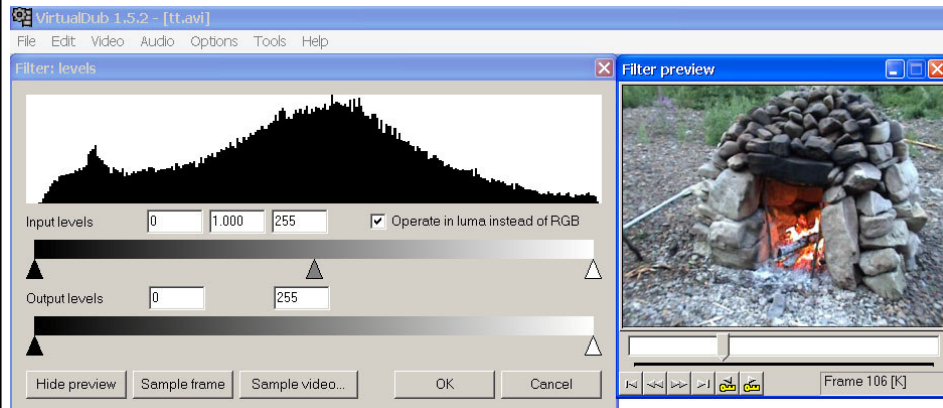
## Настройка штрихового изображения



## Гистограмма штрихового изображения



# Virtual Dub: Настройка видео



## Домашнее задание

- Ознакомиться с настройкой изображения с помощью гистограммы
  - динамический диапазон фотографии
  - контраст сканированного текста