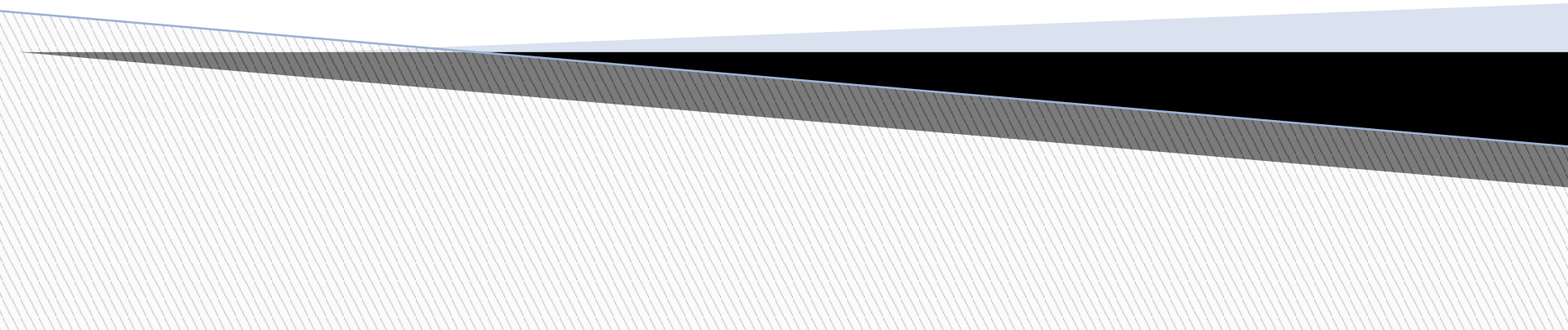


# Эконометрика

Семинар 2



# Законы распределения

1. Закон распределения — совокупность свойств распределения, описываемая плотностью вероятности и функцией распределения.
2. Наиболее важным законом распределения в математической статистике является закон нормального распределения, плотности вероятности которого записывается следующим образом:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right).$$

Здесь  $\mu$  — математическое ожидание,  $\sigma$  — среднеквадратичное отклонение.

3. Для статистических методов построения эмпирических зависимостей очень важно, чтобы результаты наблюдений подчинялись нормальному закону распределения.

# Свойства нормального распределения

1. Кривая нормального распределения симметрична относительно своего среднего значения.
2. Если  $x \rightarrow \pm\infty$ , то  $f(x) \rightarrow 0$ , то есть очень большие и очень маленькие значения  $x$  маловероятны.
3. Примерно  $2/3$  всех наблюдений лежит в площади, отсекаемой перпендикулярами к оси  $Ox$  ( $\mu \pm \sigma$ ). При большом объеме выборки примерно 90 % всех наблюдений лежит между  $-1,64\sigma$  и  $+1,64\sigma$ .
4. Для нормального распределения среднее, мода и медиана совпадают. Коэффициенты асимметрии и эксцесса равны нулю.

# Критерии проверки распределения на нормальность

1. Слабые («прикидочные») критерии:
  - Среднее абсолютное отклонение (САО-критерий).
  - Размах варьирования ( $R/S$ -критерий).
  - Коэффициенты асимметрии и эксцесса.
  - ...
2. Сильные («строгие») критерии:
  - Критерий  $\chi^2$ .
  - Критерий Колмогорова—Смирнова (К—С-критерий).

# R/S критерий

Размах варьирования выборки  $R$  определяется как

$$R = x_{max} - x_{min}.$$

Подсчитывают соотношение  $R/S$  и сопоставляют с критическими табличными верхними и нижними границами этого отношения.

Если  $R/S$  меньше нижней или больше верхней критической границы, то распределение не является нормальным, в противном случае — является.

# R/S критерий

Задача. Проверить подготовленную ранее выборку ВРП российских регионов на нормальность при помощи  $R/S$ -критерия.  $R = 926\,624,9$ ,  $S = 229\,052$ .

$$R/S = 4,045.$$

Выборка состояла из 83 наблюдений, то есть  $n = 83$ . Таблица критических значений  $R/S$ -критерия содержит значения для 80 и 85 наблюдений. Возьмем значения для объема выборки в 85 наблюдений. Проследим по таблице, начиная с какого уровня ошибки интервал между критическими значениями  $R/S$  статистики начнет покрывать расчетное значение 4,04. Легко видеть, что это произойдет при уровне ошибки в 1 % (0,01), следовательно, нам следует взять уровень на ступень выше — 5 %, следовательно, можно говорить, что данная выборка распределена нормально с вероятностью ошибки 5 %.

# Критерий асимметрии и эксцесса

Пусть  $k_a$  — коэффициент асимметрии,  $k_e$  — коэффициент эксцесса.

Вычислим несмещенные оценки для  $k_a$  и  $k_e$ :

$$K_a = \frac{\sqrt{n(n-1)}}{n-2} k_a,$$

$$K_e = \frac{n-1}{(n-2)(n-3)} \cdot [(n+1)k_e + 6].$$

Также вычислим среднеквадратические отклонения для показателей асимметрии и эксцесса:

$$S_{K_a} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}},$$

$$S_{K_e} = \sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}}.$$

# Критерий асимметрии и эксцесса

Наконец, следует проверить выполнение системы условий:

$$|K_a| \leq 3S_{K_a}, \quad |K_e| \leq 5S_{K_e}.$$

Если оба этих условия выполняются, то гипотеза нормальности исследуемого распределения может быть принята.



# Задача

Задача. Проверить подготовленную ранее выборку ВРП российских регионов на нормальность при помощи критерия асимметрии и эксцесса.  $k_a = 1,311$ .  $k_e = 1,014$ .

$$K_a = \frac{\sqrt{83(83-1)}}{83-2} 1,311 \approx 1,32.$$

$$K_e = \frac{83-1}{(83-2)(83-3)} [(83+1)k_e + 6] \approx 1,15.$$

$$S_{K_a} = \sqrt{\frac{6 \cdot 83 \cdot (83-1)}{(83-2)(83+1)(83+3)}} \approx 0,26.$$

$$S_{K_e} = \sqrt{\frac{24 \cdot 83 \cdot (83-1)^2}{(83-3)(83-2)(83+3)(83+5)}} \approx 0,52.$$

# Критерий $\chi^2$

1. Разобьем выборку на классы (подмножества). Длина интервалов разбиения определится согласно правилу Скотта:

$$\Delta = 3,5Sn^{-1/3}.$$

2. Определяем середины классов  $x$ .
3. Подсчитываем частоты  $B$  для всех классов.
4. Для каждого класса вычисляем  $Bx$  и  $Bx^2$ .
5. Находим  $\bar{x}$  и  $\bar{S}$ :

$$\bar{x} = \frac{\sum Bx}{n}, \quad \bar{S} = \sqrt{\frac{\sum Bx^2 - \frac{(\sum Bx)^2}{n}}{n-1}}.$$

# Критерий $\chi^2$

6. Вычисляем  $k = n\Delta/\bar{S}$ .

7. Для каждого класса рассчитываем

$$z = \left| \frac{x - \bar{x}}{\bar{S}} \right|.$$

8. Пусть

$$y = f(z) \simeq 0,4e^{-z^2/2} —$$

уравнение кривой стандартного нормального распределения. Для каждого класса рассчитываем  $E = f(z)k$ , затем —

$$\frac{(B - E)^2}{E}.$$

## Критерий $\chi^2$

9. Сумма последних выражений для всех классов даст значение критерия  $\chi^2$

$$\chi^2 = \sum_{m=1}^k \frac{(B - E)^2}{E}.$$

10. Проверяем выполнение гипотезы о нормальности распределения по выполнению неравенства

$$\chi^2 \leq \chi_{\text{табл.}}^2 = \chi_{\nu, \alpha}^2,$$

где  $\alpha = 1 - p$  — уровень значимости (обычно 10% или 5%),  $\nu = k - 2 - 1$ .

11. Если неравенство из п. 10 выполняется, считаем, что выборка распределена нормально.

# Критерий $\chi^2$

Задача. Определить интервалы разбиения исследуемой выборки на классы. Построить интервалы разбиения.

$$x_{min} = 18\,782,9;$$

$$x_{max} = 944\,407,8;$$

$$n = 78;$$

$$\bar{S} = 229\,052.$$

## Преобразование распределения выборки к нормальному

Если распределение имеет крутую левую ветвь гистограммы и пологую правую, выполняют преобразование выборки по формулам:

1.  $x' = \lg(x \pm a) \cdot 10^b$ ,
2.  $x' = 1/x$ ,
3.  $x' = 1/\sqrt{x}$ .

После каждого преобразования вычисляют критерий  $\chi^2$  и принимают то преобразование, которое дало наименьший  $\chi^2$ . Для распределений, смещенных вправо, исходные данные преобразуют по формуле  $x' = x^a$  (при  $a = 1,5; 2$ ) и после каждого преобразования вычисляют  $\chi^2$ , выбирают  $a$ , при котором  $\chi^2$  минимален.