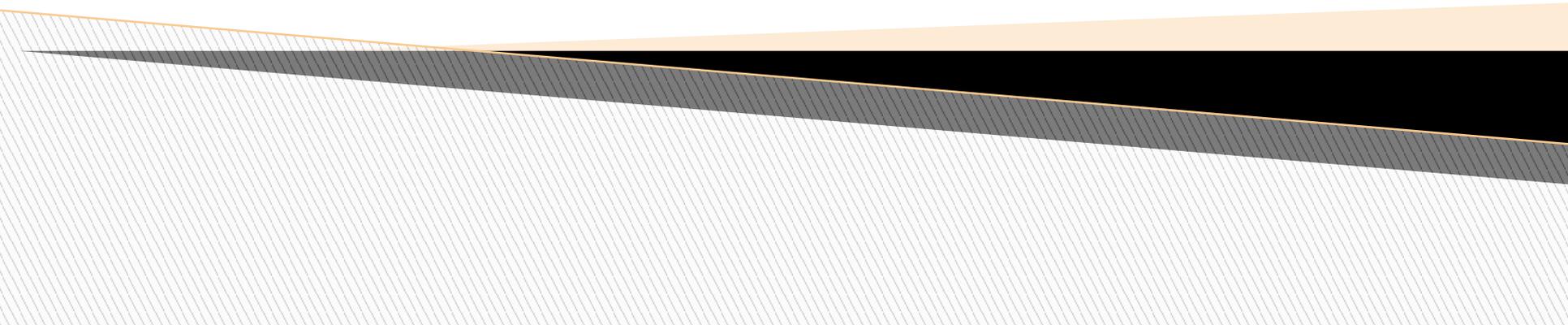
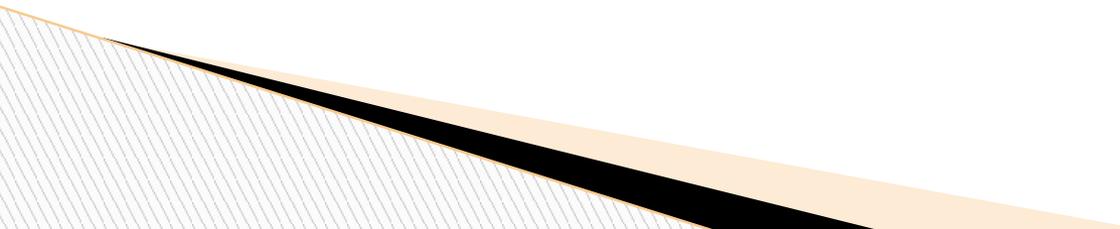


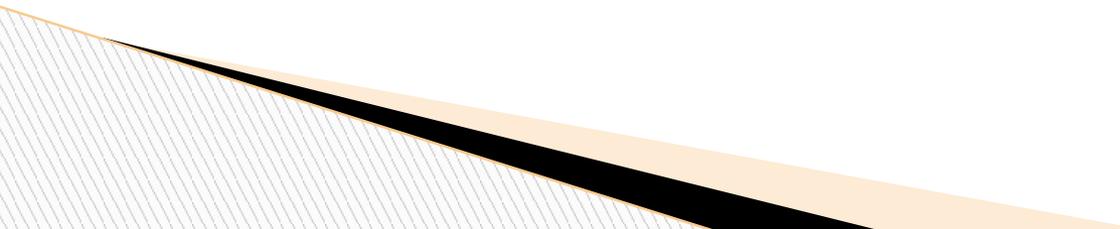
Эконометрика

Семинар 1



- ▣ Типы данных эконометрических исследований
 - ▣ Предварительная обработка данных
 - Простые описательные статистики
 - Проверка выборки на однородность
 - Отсев аномальных наблюдений
- 

Типы данных эконометрических исследований

- пространственные данные (англ. cross-sectional data) — наборы показателей экономических переменных, полученных в определенный момент времени,
 - временные ряды (англ. time series) — выборки наблюдений, в которых важны не только сами наблюдаемые значения случайных величин, но и порядок их расположения друг за другом,
 - пространственно-временные данные (англ. panel data) - прослеженные во времени пространственные выборки.
- 

Простые описательные статистики. Моменты

Пусть x_1, x_2, \dots, x_n – ограниченный ряд наблюдений случайной величины ξ .

Общая формула вычисления момента запишется так:

$$m_k = \frac{1}{n} \sum_{i=1}^n (x_i - c)^k.$$

Здесь k – порядок момента. Если $c = 0$, то момент называется начальным, если $c = \bar{x}$ – центральным.

Простые описательные статистики. Моменты

- ▣ Задание. Записать на доске формулы:
 - начального момента первого порядка,
 - центрального момента второго порядка,
 - центрального момента третьего порядка, отнесенного к кубу среднеквадратичного отклонения.

Простые описательные статистики. Моменты

- Важнейшими моментами эмпирического распределения, имеющими геометрическую интерпретацию, являются:

начальный момент первого порядка – выборочное среднее (\bar{x});

центральный момент второго порядка – выборочная дисперсия (S^2);

центральный момент третьего порядка, отнесенный к кубу среднеквадратичного отклонения – коэффициент асимметрии (g_1);

центральный момент четвертого порядка, отнесенный к четвертой степени среднеквадратичного отклонения за вычетом трех – коэффициент эксцесса (g_2).

Простые описательные статистики. Моменты

Оценки выборочных характеристик должны обладать свойствами несмещенности, состоятельности и эффективности. Все перечисленные выше выборочные характеристики, кроме выборочной дисперсии, обладают данными свойствами. Оценка выборочной дисперсии является смещенной, поэтому вводится оценка *исправленной выборочной дисперсии*

$$\overline{S^2} = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

которая обладает всеми тремя свойствами.

Аналогично случаю «обычной» дисперсии вводится исправленное среднеквадратичное отклонение

$$\overline{S} = \sqrt{\overline{S^2}}.$$

Простые описательные статистики. Моменты

Задание. Вычислить выборочное среднее, выборочную дисперсию и исправленную выборочную дисперсию для следующего ряда наблюдений.

Таблица: Объем добычи меди в Красноярском крае в 2004–2008 гг., тыс. т

	2004	2005	2006	2007	2008
Объем добычи	353,0	361,0	351,2	338,4	338,5

Источник: годовые отчеты ОАО «ГМК „Норильский никель“».

Ответы: $\bar{x} = 348,42$; $S^2 = 77,154$; $\overline{S^2} = 96,442$.

Проверка выборки на однородность

Однородность совокупности – это свойство устойчивости значения признака. Мерой однородности является, например, вариация – различие значений какого-либо признака у разных единиц совокупности за один и тот же промежуток времени. Коэффициент вариации вычисляется как

$$v = \frac{\bar{s}}{\bar{x}}$$

Дальнейший статистический анализ выборки (выборок) возможен, как правило, если $0,03 < v < 0,33$. Важно понимать, что эти границы условны и не могут использоваться как строгий критерий.

Проверка выборки на однородность

Одним из важнейших макроэкономических показателей развития государств является валовый внутренний продукт (ВВП). На региональном уровне вычисляется его аналог – валовый региональный продукт (ВРП), – который аналогичным образом рассматривается в качестве одного из ключевых индикаторов экономического развития территории.

Задание. По выборкам, основанным на наблюдениях над ВРП ряда российских регионов в 2008 г. (млн руб.), вычислить значения коэффициентов вариации и сделать выводы об однородности совокупностей.

Проверка выборки на однородность.

Пример №1

Белгородская область	319 071,4
Воронежская область	289 322,3
Республика Коми	295 614,3
Архангельская область	298 111,5
Республика Саха	304 986,8

Ответ: $v \approx 0,037$, то есть вариация значений признака незначительна, следовательно, данная выборка в значительной степени однородна. С экономической точки зрения это означает, что в выборку попали регионы с близким уровнем экономического развития. Обратите внимание, что несмотря на близкий уровень ВРП, Республика Саха и, например, Воронежская область существенно отличаются по площади и природно-климатическим условиям.

Проверка выборки на однородность.

Пример № 2

Еврейская автономная область	26 695,2
Амурская область	136 784,0
г. Санкт-Петербург	1 420 830,1
Красноярский край	740 233,1
г. Москва	8 441 206,2

Ответ: $v \approx 1,658$, то есть вариация значений признака крайне высока, следовательно, данная выборка не может считаться однородной и явно содержит уровни, которые являются «аномальными». С экономической точки зрения это означает, что в выборку попали регионы с кардинально различающимся уровнем экономического развития. Действительно, выборка была сконструирована так, что в нее попали самый «отсталый» с точки зрения экономического развития регион страны – Еврейская автономная область, – и самый высокоразвитый – г. Москва.

Отсев аномальных наблюдений

□ Малая выборка ($n \leq 25$)

- 1 Выделить элемент выборки, который, по предположению, является аномальным.
- 2 Вычислить для выбранного элемента статистику

$$\tau = \frac{|x_i - \bar{x}|}{\bar{s}}$$

- 3 Из таблицы распределения τ (например, из учебника Львовского) взять значения для данного объема выборки $p = 5\%$.
- 4 Сравнить вычисленное значение τ с табличными. Если

$$\tau \leq \tau_{1-p},$$

то наблюдение *не отсеивают*, в противном случае наблюдение отсеивают.

Отсев аномальных наблюдений

Большая выборка ($n > 25$). Известно, что критическое значение τ_p выражается через критическое значение распределения Стьюдента $t_{p, n-2}$:

$$\tau_{(p, n)} = \frac{t_{(p, n-2)} \sqrt{n-1}}{\sqrt{n-2 + [t_{(p, n-2)}]^2}}.$$

- 1** Выбирается элемент, имеющий максимальное отклонение от среднего.
- 2** По приведенной в алгоритме для малой выборки формуле вычисляется τ .
- 3** Из таблицы распределения Стьюдента находят процентные точки $t_{1\%, n-2}$, $t_{5\%, n-2}$.
- 4** Вычисляют $\tau_{(p, n)}$.

Отсев аномальных наблюдений

Большая выборка ($n > 25$). Далее возникают три альтернативы:

- 1** $\tau \leq \tau_{(5\%, n)}$: наблюдение отсеивать нельзя.
- 2** $\tau_{(5\%, n)} < \tau < \tau_{(0,1\%, n)}$: наблюдение можно отсеять при наличии других аналогичных соображений экспериментатора.
- 3** $\tau > \tau_{(0,1\%, n)}$: наблюдение отсеивается в любом случае.

Отсев аномальных наблюдений

Задача. Выборка объемом 83 наблюдений состоит из значений ВРП всех субъектов РФ в 2008 г.

Аномальным наблюдением будем считать наблюдение над ВРП г. Москвы (ВРП превышает 8 трлн руб. при среднем по стране 446,5 млрд руб.).

$\bar{x} = 446\,531,48$ млн руб., $\bar{s} = 1\,003\,611,904$.

τ -статистика для ВРП Москвы составляет 7,97.

Табличные значения t -статистики: $t_{5\%;83-2} = 1,664$;

$t_{1\%;83-2} = 2,373$.

Отсев аномальных наблюдений

Вычислим $\tau_{p,n}$ через t -статистику:

$$\tau_{5\%,83} = \frac{1,664 \cdot \sqrt{82}}{\sqrt{81 + (1,664)^2}} \approx 1,646;$$

$$\tau_{0,1\%,83} = \frac{3,195 \cdot \sqrt{82}}{\sqrt{81 + (3,195)^2}} \approx 3,029.$$

Очевидно, что $\tau \gg \tau_{0,1\%,83} > \tau_{5\%,83}$, следовательно, рассматриваемое наблюдение следует обязательно удалить из выборки.

Далее следует пересчитать выборочные характеристики и провести ту же процедуру для следующего «подозрительного» наблюдения. Таким образом будут отброшены еще 5 наблюдений.