Модуль 1 · Python и анализ данных

Временные ряды: сезонность и аномалии

Ph.D., старший преподаватель Мадина Турсумбаева

Кафедра метеорологии и гидрологии

Курс: Современные цифровые технологии в метеорологии

Цель лекции

Научить разделять временной ряд на тренд, сезонность и остаток, рассчитывать климатическую норму и аномалии в pandas и правильно их интерпретировать.

Температура в Алматы зимой и летом отличается на десятки градусов. Чтобы увидеть изменение климата, нужно вычесть сезонный ход.

Результаты обучения

После лекции студент будет:

Знать

Компоненты ряда: тренд, сезонность, остаток

Норма и аномалия; базовый период ВМО

Уметь

Считать норму по месяцам и аномалии

Выделять сезонные средние и скользящее среднее

Понимать

Аномалии зависят от базового периода

Сезонность не равна тренду

Компоненты временного ряда

Тренд

Медленное долгосрочное изменение: потепление, похолодание, многолетние колебания.

Сезонность

Повторяющийся годовой ход: зима и лето, периоды с периодом около 12 месяцев.

Остаток

То, что не объяснено трендом и сезонностью: погодные события, шум измерений.

```
Y = T + S + R        # аддитивная модель: для температуры в °C
```

- Для температуры в градусах обычно берут аддитивную модель: компоненты складываются.
- Мультипликативная модель ($Y = T \cdot S \cdot R$) уместна для величин, которые по природе положительны, например осадков.

Норма и аномалия

Норма (климатическая)

Среднее значение за базовый период. ВМО использует 30-летние периоды; текущий стандарт — 1991–2020 гг.

Аномалия

Отклонение от нормы того же месяца: аномалия = значение − норма. Знак показывает, теплее или холоднее обычного.

```
# аномалия за месяц m:
anomaly_m = T_m - norm[month(m)]
```

Смена базового периода сдвигает все аномалии на одну и ту же величину. Сравнивать можно только аномалии с одной нормой.

Годовой ход: условные данные



Как читать

Сезонный ход виден сразу: минимум зимой, максимум летом.

Это условные данные из лекции 1.

Вопрос

Какой месяц по этим данным ближе всего к норме года? Норму каждого месяца нужно считать отдельно.

Норма и аномалии в pandas

```
m = s.resample("MS").mean()                    # среднемесячные значения
# норма — по базовому периоду 1991–2020 гг.
base = m[(m.index.year >= 1991) & (m.index.year <= 2020)]
norm = base.groupby(base.index.month).mean()  # норма по месяцам
# аномалия: из каждого месяца вычитаем норму того же месяца
anom = m - norm.reindex(m.index.month).values
```

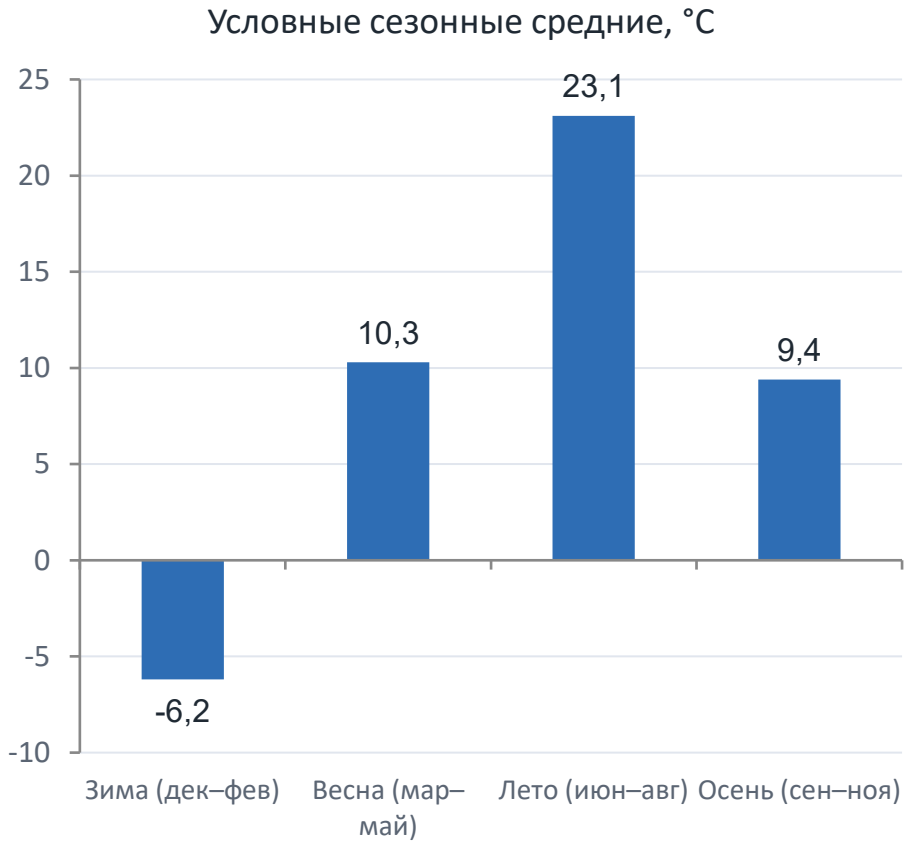
- base — только 1991–2020: норма не должна зависеть от всего ряда.
- reindex по номеру месяца сопоставляет каждой дате норму её месяца.

Сезоны: группировка по месяцам

```
# 0 — зима (дек-фев), 1 — весна, 2 — лето, 3 — осень
season = (m.index.month % 12) // 3
seasonal = m.groupby(season).mean()
# ВНИМАНИЕ: декабрь относится к зиме следующего года
```

Сезоны — не календарные месяцы

Сезоны удобно сравнивать между годами. Для Северного полушария обычно используют декабрь–февраль (зима), март–май (весна), июнь–август (лето), сентябрь–ноябрь (осень).



Значения — условные (из лекции 1), для иллюстрации.

Скользящее среднее и декомпозиция

```
# скользящее среднее за 12 месяцев: сглаживает сезонность
trend_ma = m.rolling(12, center=True).mean()

# декомпозиция (нужен ряд без пропусков)
from statsmodels.tsa.seasonal import seasonal_decompose
res = seasonal_decompose(m, model="additive", period=12)
res.trend, res.seasonal, res.resid
```

Что получаем

trend — медленная составляющая;

seasonal — повторяющийся годовой ход;

resid — остаток.

statsmodels устанавливается отдельно: conda install -c conda-forge statsmodels.

Интерпретация аномалий

Знак

Положительная аномалия — месяц теплее нормы; отрицательная — холоднее.

Величина

Оценивайте, на сколько градусов и насколько часто: одна тёплая зима не делает тренд.

Базовый период

Аномалии из разных источников сравнимы только при одной норме.

Сезонный ход — не тренд. Тренд ищут в аномалиях или в скользящем среднем, а не в сырых значениях.

Типичные ошибки и их исправление

Ошибка	Что происходит	Как исправить
Норма по всему ряду	Тренд попадает в норму	Фиксированный базовый период
Сравнение сырых значений по годам	Сезонность маскирует изменение	Сравнивать аномалии или сезонные средние
Декабрь — в том же году	Зима разорвана на два года	Относить декабрь к следующему зимнему сезону
Декомпозиция с пропусками	Ошибка или искажение	Заполнить или выбрать непрерывный отрезок
Разные нормы в одном графике	Аномалии не сопоставимы	Один базовый период для всех рядов

Заключение

1

Компоненты

тренд, сезонность, остаток

2

Норма

1991–2020, по месяцам

3

Аномалия

значение − норма месяца

4

Сезоны

группировка по season

Главный вывод

Временной ряд состоит из тренда, сезонности и остатка. Чтобы увидеть изменение климата, из ряда вычитают норму соответствующего месяца. Величина аномалии зависит от выбранного базового периода, а сезонность нельзя принимать за тренд.

Вопросы для самопроверки

1. Почему аномалия зависит от базового периода, а сырое значение — нет?
2. Как группировка по сезонам меняется, если декабрь относить к текущему году?
3. Что показывает скользящее среднее за 12 месяцев и чего оно не показывает?
4. Почему тренд нельзя оценивать по одному тёплому месяцу?

Семинар 5 и самостоятельная работа

Семинар 5 · Маусымдық өзгерістер

Сезонные изменения температуры по данным из семинара 4.

Нормы по месяцам, сезонные средние, аномалии.

СРО/СРОП 5 · 20 баллов

Самостоятельная работа по анализу сезонности и аномалий для выбранной станции.

Следующая лекция — радиозонды: структура данных и их обработка.

Литература

- Wilks, D. S. (2019). Statistical Methods in the Atmospheric Sciences (4th ed.). Elsevier.
- von Storch, H., & Zwiers, F. W. (1999). Statistical Analysis in Climate Research. Cambridge University Press.
- WMO (2017). WMO Climate Normals: сайт ВМО по климатическим нормам, период 1991–2020 гг. https://public.wmo.int
- statsmodels developers. Seasonal decomposition. https://www.statsmodels.org/
- pandas development team. Resampling and groupby. https://pandas.pydata.org/docs/