Лекция 8. Моделирование временных рядов

Моделирование колебаний

- Рассмотрим процесс построения аддитивной модели*:

$$Y=T+S+E\text{, где}$$

  - T – трендовая компонента
  - S – сезонная компонента
  - E – случайнвя компонента

---

* речь идет о моделях временного ряда с отсутствием циклической компоненты.

Шаги построения моделей:

- Выравнивание исходного ряда методом скользящей средней.
- Рассчет значений сезонной компоненты S.
- Устранение сезонной компоненты из исходных уровней ряда и получение выравненных данных (T+E) в аддитивной или (T*E) в мультипликативной.
- Аналитическое выравнивание уровней (T+E) или (T*E) и расчет значений T с использованием полученного уравнения тренда.
- Расчет полученных по модели значений (T+E) или (T*E).
- Расчет абсолютных и относительных ошибок(E).

Аддитивная модель

- Если временной ряд содержит сезонные колебания с определенной периодичностью и амплитуда этих колебаний приблизительно одинакова, значит, для моделирования подходит ***аддитивная модель***.
- Для её построения выполним необходимые расчеты и сведем их в таблицу.

t	y_t	Итого за 4 квартала	Скользящая средняя за 4 квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	2	3	4	5	6
1	6,0	-	-	-	-
2	4,4			-	-
3	5,0	24,4	6,1		
4	9,0	25,6	6,4	6,250	-1,250
5	7,2	26,0	6,5	6,450	2,550
6	4,8	27,0	6,75	6,625	0,575
7	6,0	28,0	7,0	6,875	-2,075
8	10,0	28,8	7,2	7,100	-1,100
9	8,0	29,6	7,4	7,300	2,700
10	5,6	30,0	7,5	7,450	0,550
11	6,4	31,0	7,75	7,625	-2,025
12	11,0	32,0	8,0	7,875	-1,475
13	9,0	33,0	8,25	8,125	2,875
14	6,6	33,6	8,4	8,325	0,675
15	7,0	33,4	8,35	8,375	-1,775
16	10,8	-	-	-	

назад

Шаг 1. Проведем выравнивание исходных уровней ряда методом скользящей средней:

- 1.1. Просуммируем уровни ряда последовательно за каждые четыре квартала со сдвигом на один момент времени ($\mathbf{y_1+y_2+y_3+y_4}$, затем $\mathbf{y_2+y_3+y_4+y_5}$, затем $\mathbf{y_3+y_4+y_5+y_6}$ и т.д.) и определим условные годовые объемы потребления. (см. **столбец 3**)
- 1.2. Разделим полученные суммы на 4 – находим *скользящие средние* (см. **столбец 4**)

- 1.3. Приведем эти значения в соответствие с фактическими моментами времени, для чего найдем средние значения из двух последовательных скользящих средних – *центрированные скользящие средние* (см. **столбец 5**)

Таблица

Шаг 2.

Рассчитаем *оценки сезонной компоненты* как разность между фактическими уровнями ряда (y_t) и центрированными скользящими средними ($y_3 - ц_1$, $y_4 - ц_2$,..., $y_{14} - ц_{12}$, где *ц* – значение столбца 5), получим **столбец 6**.

Таблица

Теперь на основе этих оценок рассчитаем *значения сезонной компоненты для S*. Для этого найдем *средние* за каждый квартал (по всем годам) оценки сезонной компоненты S_i (таблица 9).

Таблица 9

Показатель	Год	Номер квартала, *i*			
		I	II	III	IV
	1	-	-	-1,250	2,550
	2	0,575	-2,075	-1,100	2,700
	3	0,550	-2,025	-1,475	2,875
	4	0,675	-1,775	-	-
Итого за *i*—тый квартал (сумма ∑ за все годы)		1,800	-5,875	-3,825	8,125
Средняя оценка сезонной компоненты для *i*-го квартала, S_i (= ∑/3)		0,600	-1,958	-1,275	2,708
Скорректированная сезонная компонента, S_i		0,581	-1,977	-1,294	2,690

В моделях с сезонной компонентой предполагают, что сезонные воздействия за период взаимопогашаются, это означает, что в аддитивной модели сумма значений сезонной компоненты по всем кварталам должна быть равна нулю. Просуммируем $\overline{S}_i$:

$$0{,}6-1{,}958-1{,}275+2{,}708=0{,}075$$

Это значение больше, чем ноль, поэтому определим корректирующий коэффициент:

$$k= 0{,}075/4=0{,}01875$$

Скорректируем значения $\overline{S}_i$:

$S_i=\overline{S}_i - k$ *(при вычитании учитываем знак* $\overline{S}_i$*).*

Полученные значения занесены в последнюю строку таблицы 9.

Проверим их еще раз, просуммируем:
0,581 – 1,977 – 1,294 + 2,690=0. Теперь сумма равна 0.
Окончательно, получены **значения сезонной компоненты**:

I квартал: S1 = 0,581;
II квартал: S2 = -1,977;
III квартал: S3 = -1,294;
IV квартал: S4 = 2,690.

Далее необходимо выявить трендовую (тенденцию) и случайную компоненты. Для расчетов заведем новую таблицу 10. В первые два ее столбца внесем исходный временной ряд, в столбец 3 занесем полученные значения сезонной компоненты (они повторяются через каждые 4 квартала).

Шаг 3.

Исключим теперь влияние сезонной компоненты, вычитая ее значения из каждого уровня временного ряда. Получим T + E = Y – S (**столбец 4** таблицы 10).

t	y_t	S_i	$T+E=y_t+S_i$	T	$T+S$	$E = y_t - (T + S)$	E^2
1	2	3	4	5	6	7	8
1	6,0	0,581	5,419	5,902	6,483	-0,483	0,2333
2	4,4	-1,977	6,337	6,088	4,111	0,289	0,0835
3	5,0	-1,294	6,294	6,275	4,981	0,019	0,0004
4	9,0	2,690	6,310	6,461	9,151	-0,151	0,0228
5	7,2	0,581	6,619	6,648	7,229	-0,029	0,0008
6	4,8	-1,977	6,777	6,834	4,857	-0,057	0,0032
7	6,0	-1,294	7,294	7,020	5,727	0,273	0,0745
8	10,0	2,690	7,310	7,207	9,896	0,104	0,0108
9	8,0	0,581	7,419	7,393	7,974	0,026	0,0007
10	5,6	-1,977	7,577	7,580	5,603	-0.003	0,0009
11	6,4	-1,294	7,694	7,766	6,472	-0,072	0,0052
12	11,0	2,690	8,310	7,952	10,642	0,358	0,1282
13	9,0	0,581	8,419	8,139	8,720	0,28	0,0784
14	6,6	-1,977	8,577	8,352	6,348	0,252	0,0635
15	7,0	-1,294	8,294	8,519	7,218	-0,218	0,0475
16	10,8	2,690	8,110	8,698	11,388	-0,588	0,3457

Шаг 4.

Определим компоненту *T* модели. Для этого проведем аналитическое выравнивание ряда *(T+E)* модели с помощью линейного тренда: определяем уравнение парной линейной регрессии $y = a + bx$, в котором роль у играет $T + E$, а роль x – время t. Найдем коэффициенты уравнения, стандартную ошибку коэффициента регрессии b и коэффициент детерминации (например, используя программу «Регрессия» в Excel). Получим:

a= 5,715416
b= 0,186421

Стандартная ошибка коэффициента регрессии

Sb= 0,015188
$R2$ = 0,914971
n = 16

Число степеней свободы n-2 = 14.
В результате **получен линейный тренд** (прямая) вида:
$T = 5{,}715 + 0{,}186 * t$
Подставим имеющиеся значения t (t=1,…,16) в это уравнение, получим значения T для каждого момента времени, внесем их в таблицу (**столбец 5**).

Шаг 5.

Для вычисления ошибки (остатков) E найдем значения уравнений ряда $\hat{y}_t$, вычисленные по модели, т. е. посчитаем сумму $T + S$, добавляя к каждому значению тренда T соответствующее значение сезонной компоненты S_i по кварталам. Полученные значения внесем в **столбец 6** таблицы 10.

Шаг 6.Рассчитываем ошибку: E=Y-(T+S)

1. Для оценки качества модели используем анализ суммы квадратов ошибки E^2 (см. **столбец 7**) .
2. Подсчитаем значения $\sum \mathbf{E^2=1,10}$ и вычислим сумму квадратов отклонений уровня ряда от среднего значения: $\sum (\mathbf{y_t - \bar{y}_t})^2 = \mathbf{71,59}$.
3. Вычислим долю ошибки: 1,1/71,59=0,015365 или 1,536%. Оставшиеся - 98,46%-доля дисперсии уровней временного ряда, объясненная аддитивной моделью.

Вывод

- Полученная аддитивная модель Y=T+S+E, в которой тренд **T=5,715+0,186t** , сезонная компонента S составляет по кварталам: I квартал:$\mathbf{S_1=0,581}$;II квартал:$\mathbf{S_2=-1,977}$; III квартал $\mathbf{S_3=-1,294;}$ IV квартал: $\mathbf{S_4=2,690}$, объясняет около 98,5 % общей вариации уровней временного ряда потребления электроэнергии за последние 16 кварталов.