МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Тульский государственный университет»

Институт прикладной математики и компьютерных наук
Кафедра «Прикладная математика и информатика»

Утверждено на заседании кафедры
«Прикладная математика и информатика»
24 января 2023 г., протокол № 5

И.о. заведующего кафедрой

Н.В. Ларин

МЕТОДИЧЕСКИЕ УКАЗАНИЯ
по выполнению практических (семинарских) занятий
по дисциплине (модулю)
«Методы прогнозирования»

основной профессиональной образовательной программы
высшего образования – программы магистратуры

по направлению подготовки
01.04.02 Прикладная математика и информатика

с направленностью (профилем)
Перспективные методы искусственного интеллекта
в сетях передачи и обработки данных

Форма обучения: очная

Идентификационный номер образовательной программы: 010402-01-22

Тула 2023 год

Разработчик методических указаний

Кочетыгов А.А., профессор каф. ПМиИ, к.т.н., доцент

(ФИО, должность, ученая степень, ученое звание)

(подпись)

Целью освоения дисциплины «Методы прогнозирования» является получение навыков прогнозирования вероятных ситуаций методами экстраполяции, статистического моделирования временных рядов, методами экспертного прогнозирования

Практические занятия по этой дисциплине имеют целью закрепление знаний по всем основным разделам курса.

Предусматриваются примеры по следующим разделам курса.

Применение многофакторных моделей регрессии для прогнозирования

Статистические характеристики временных рядов. Предварительный анализ временных рядов. Тесты на стационарность временных рядов.

Сглаживание временных рядов и прогноз сглаживающими методами.

Прогнозирование нестационарных временных рядов.

Моделирование одномерных временных рядов. Автокорреляция уровней временных рядов и выявление его структуры.

Модели тренда.

Модели сезонности и цикличности

Построение моделей ARIMA и прогнозирование с их помощью временных рядов.

Экспертные методы прогнозирования

Предлагаются индивидуальные задания для самостоятельного решения.

Предлагаемые ниже задачи, примеры и задания предназначены для закрепления материала, изучаемого на лекциях.

На практических занятиях могут обсуждаться и рассматриваться также задачи и примеры из других приложений (области деятельности).

Пример 5.2. Имеются условные данные о стоимости акций некоторой компании за 16 кварталов (табл. 5.3).

Таблица 5.3

Стоимость акций некоторой компании за 16 кварталов

Квартал t	Стои-мость акций y_t	Расчётные значения 8 уровней ряда	Коэффициенты модели Хольта–Уинтерса			Модель Хольта–Уинтерса y_t^p
			a_t	b_t	s_t	
1	304	308,65	310,73	9,22	0,9723	297,32
2	320	317,25	320,87	9,50	0,9947	316,97
3	334	325,85	329,58	9,26	1,0157	336,68
4	347	334,45	338,54	9,17	1,0258	348,02
5	323	343,05	343,06	7,77	0,9538	338,08
6	342	351,65	348,74	7,14	0,9862	348,94
7	365	360,25	356,92	7,45	1,0199	361,47
8	375	368,85	364,73	7,56	1,0272	373,77
9	342		368,17	6,32	0,9389	355,09
10	365		373,18	5,93	0,9813	369,52
11	378		376,56	5,17	1,0103	386,65
12	399		383,74	5,77	1,0347	392,11
13	363		388,64	5,51	0,9360	365,71
14	388		394,52	5,62	0,9826	386,78
15	419		404,52	6,93	1,0256	404,26
16	418		409,21	6,26	1,0268	425,73

Спрогнозировать цены на эти акции на 4 квартала пятого года.

Решение. Используем мультипликативную тренд–сезонную модель Хольта–Уинтерса с линейным ростом.

$$y_{t+\tau}^p = (a_t + \tau \cdot b_t) \cdot s_{t+\tau-L}.$$

Здесь:

$y_{t+\tau}^p$ – расчётное (прогнозное) значение показателя для периода $t+\tau$; τ – период упреждения;

a_t, b_t, s_t – коэффициенты модели (они адаптируются, уточняются по мере перехода от членов ряда с номером t–1 к t);

$s_{t+\tau-L}$ – значение коэффициента сезонности того периода, для которого рассчитывается показатель;

L – период сезонности.

Уточнение (адаптация к новому значению параметра времени t) коэффициентов производится с помощью формул:

$$a_t = \frac{\alpha_1 \cdot y_t}{s_{t-L}} + (1-\alpha_1)(a_{t-1} + b_{t-1});$$

$$b_t = \alpha_2 (a_t - a_{t-1}) + (1-\alpha_2) b_{t-1};$$

$$s_t = \frac{\alpha_3 \cdot y_t}{a_t} + (1-\alpha_3) s_{t-L}.$$

Параметры сглаживания α_1, α_2, α_3 подбираются путём перебора так, чтобы расчётные данные наилучшим образом соответствовали фактическим.

Для расчёта a_1, b_1 необходимо оценить значения этих коэффициентов для предыдущего периода времени (для $t = 0$).

Для оценки этих величин a_0, b_0 применим линейную модель к первым 8 значениям исходного ряда y_t. Линейная модель имеет вид:

$$y_t^p = a_0 + b_0 t.$$

Метод наименьших квадратов даёт возможность определить коэффициенты этой модели по формулам:

$$b_0 = \frac{\sum_{t=1}^{n} (y_t - \overline{y})(t - \overline{t}\,)}{\sum_{t=1}^{n} (t - \overline{t}\,)^2};$$

$$a_0 = \overline{y} - b_0 \overline{t}\,;$$

$$\overline{y} = \frac{1}{n} \sum_{t=1}^{n} y_t; \quad \overline{t} = \frac{1}{n} \sum_{t=1}^{n} t.$$

Применяя линейную модель к первым 8 значениям исходного ряда y_t находим: $a_0 = 300,05; \quad b_0 = 8,60.$

С учётом полученных коэффициентов линейная модель имеет вид:

$$y_t^p = 300,05 + 8,6t.$$

Из этого уравнения находим расчётные значения y_t^p и сопоставляем их с фактическими. Такое сопоставление позволяет оценить приближённые значения коэффициентов сезонности четырёх кварталов s_{-3}, s_{-2}, s_{-1}, s_0 для года предшествующего первому году наблюдений.

Эти значения необходимы для расчёта коэффициентов сезонности первого года s_1, s_2, s_3, s_4 и для других параметров модели Хольта–Уинтерса.

Коэффициент сезонности есть отношение фактического значения показателя к значению, рассчитанному по линейной модели. Поэтому в качестве оценки коэффициента сезонности первого квартала может служить отношение фактических и расчётных значений показателя y_t первого квартала первого года, равное $y_1 \,/\, y_1^p$. Такое же отношение и для первого квартала второго года, т.е. за 5–й квартал $(t=5)$ $y_5 \,/\, y_5^p$. Для окончательной, более точной, оценки этого коэффициента сезонности можно использовать среднее арифметическое значение этих двух величин:

$$s_{-3} = (y_1 \,/\, y_1^p + y_5 \,/\, y_5^p) / 2 = (304 / 308{,}65 + 323 / 343{,}05) / 2 = 0{,}9633.$$

Аналогично находим оценки коэффициентов сезонности для второго, третьего и четвёртого кварталов:

$$s_{-2} = (y_2 \,/\, y_2^p + y_6 \,/\, y_6^p) / 2 = (320 / 317{,}25 + 342 / 351{,}65) / 2 = 0{,}9907.$$

$$s_{-1} = (y_3 \,/\, y_3^p + y_7 \,/\, y_7^p) / 2 = (324 / 325{,}85 + 365 / 360{,}25) / 2 = 1{,}0191.$$

$$s_0 = (y_4 \,/\, y_4^p + y_8 \,/\, y_8^p) / 2 = (347 / 334{,}45 + 375 / 368{,}85) / 2 = 1{,}0271.$$

Оценив значения $a_0,\ b_0,\ s_{-3},\ s_{-2},\ s_{-1},\ s_0$ можно перейти к построению адаптивной мультипликативной модели Хольта–Уинтерса.

Путём перебора возможных значений параметров сглаживания было установлено, что лучшими являются следующие их значения:

$$\alpha_1 = 0{,}3;\quad \alpha_2 = 0{,}3;\quad \alpha_3 = 0{,}6.$$

Рассчитаем значения $y_t^p,\ a_t,\ b_t,\ s_t$ для $t=1$

Полагая $t=0,\ \tau=1$ находим из исходного уравнения метода:

$$y_{0+1}^p = y_1^p = (a_0 + 1\cdot b_0)s_{0+1-4} = (a_0 + 1\cdot b_0)s_{-3} =$$
$$= (300{,}05 + 1\cdot 8{,}6)\cdot 0{,}9633 = 297{,}32.$$

Уточняем параметры адаптации (полагая $t=1$):

$$a_1 = \frac{\alpha_1 \cdot y_1}{s_{1-4}} + (1-\alpha_1)(a_{1-1} + b_{1-1}) = \frac{\alpha_1 \cdot y_1}{s_{-3}} + (1-\alpha_1)\cdot(a_0 + b_0) =$$
$$= \frac{0{,}3\cdot 304}{0{,}9633} + (1-0{,}3)\cdot(300{,}05 + 8{,}6) = 310{,}73;$$

$$b_1 = \alpha_2(a_1 - a_0) + (1-\alpha_2)b_0 = 0{,}3\cdot(310{,}73 - 300{,}05) + (1-0{,}3)8{,}6 = 9{,}22;$$

$$s_1 = \frac{\alpha_3 \cdot y_1}{a_1} + (1-\alpha_3)s_{-3} = \frac{0{,}6\cdot 304}{310{,}73} + (1-0{,}6)\cdot 0{,}9633 = 0{,}9723.$$

Рассчитаем значения $y_t^p,\ a_t,\ b_t,\ s_t$ для $t=2$ (полагая в правой части формул $t=1,\ \tau=1$):

$$y_2^p = (a_1 + 1\cdot b_1)s_{-2} = (310{,}73 + 1\cdot 9{,}22)\cdot 0{,}9907 = 316{,}97.$$

$$a_2 = \frac{\alpha_1 \cdot y_2}{s_{-2}} + (1-\alpha_1)(a_1 + b_1) = \frac{0,3 \cdot 320}{0,9907} + (1-0,3)\cdot(310,73 + 9,22) = 320,87;$$

$$b_2 = \alpha_2(a_2 - a_1) + (1-\alpha_2)b_1 = 0,3\cdot(320,87 - 310,73) + (1-0,3)\cdot 9,22 = 9,50;$$

$$s_2 = \frac{\alpha_3 \cdot y_2}{a_2} + (1-\alpha_3)s_{-2} = \frac{0,6\cdot 320}{320,87} + (1-0,6)\cdot 0,9907 = 0,9947.$$

Рассчитаем значения $y_t^p,\ a_t,\ b_t,\ s_t$ для $t = 3$ (полагая в правой части формул $t = 2,\ \tau = 1$):

$$y_3^p = (a_2 + 1\cdot b_2)s_{-1} = (320,87 + 1\cdot 9,5)\cdot 1,0191 = 336,68.$$

$$a_3 = \frac{\alpha_1 \cdot y_3}{s_{-1}} + (1-\alpha_1)(a_2 + b_2) = \frac{0,3\cdot 334}{1,0191} + (1-0,3)(320,87 + 9,5) = 329,58;$$

$$b_3 = \alpha_2(a_3 - a_2) + (1-\alpha_2)b_2 = 0,3\cdot(329,58 - 320,87) + (1-0,3)9,5 = 9,26;$$

$$s_3 = \frac{\alpha_3 \cdot y_3}{a_3} + (1-\alpha_3)s_{-1} = \frac{0,6\cdot 334}{329,58} + (1-0,6)\cdot 1,0191 = 1,0157.$$

Рассчитаем значения $y_t^p,\ a_t,\ b_t,\ s_t$ для $t = 4$ (полагая в правой части формул $t = 3,\ \tau = 1$):

$$y_4^p = (a_3 + 1\cdot b_3)s_0 = (329,58 + 1\cdot 9,26)\cdot 1,0271 = 348,02.$$

$$a_4 = \frac{\alpha_1 \cdot y_4}{s_0} + (1-\alpha_1)(a_3 + b_3) = \frac{0,3\cdot 347}{1,0271} + (1-0,3)(329,58 + 9,26) = 338,54;$$

$$b_4 = \alpha_2(a_4 - a_3) + (1-\alpha_2)b_3 = 0,3\cdot(338,54 - 329,58) + (1-0,3)9,26 = 9,17;$$

$$s_4 = \frac{\alpha_3 \cdot y_4}{a_4} + (1-\alpha_3)s_0 = \frac{0,6\cdot 347}{338,54} + (1-0,6)\cdot 1,0271 = 1,0258.$$

Рассчитаем значения $y_t^p,\ a_t,\ b_t,\ s_t$ для $t = 5$ (полагая в правой части формул $t = 4,\ \tau = 1$).

Здесь и в дальнейшем используются коэффициенты сезонности, уточнённые в предыдущем году.

$$y_5^p = (a_4 + 1\cdot b_4)s_1 = (338,54 + 1\cdot 9,17)\cdot 0,9723 = 338,08.$$

$$a_5 = \frac{\alpha_1 \cdot y_5}{s_1} + (1-\alpha_1)(a_4 + b_4) = \frac{0,3\cdot 323}{0,9723} + (1-0,3)(338,54 + 9,17) = 343,06;$$

$$b_5 = \alpha_2(a_5 - a_4) + (1-\alpha_2)b_4 = 0,3\cdot(343,06 - 338,54) + (1-0,3)9,17 = 7,77;$$

$$s_5 = \frac{\alpha_3 \cdot y_5}{a_5} + (1-\alpha_3)s_1 = \frac{0,6\cdot 323}{343,06} + (1-0,6)\cdot 0,9723 = 0,9538.$$

Продолжая аналогично для $t = 6, 7, 8, \ldots 16$, строят адаптивную мультипликативную модель Хольта–Уинтерса (табл. 5.4).

Максимальное значение t, для которого можно находить коэффициенты модели равно количеству имеющихся данных по показателю y_t. У нас максимальное значение $t = 16$.

Таблица 5.4

Стоимость акций некоторой компании за 16 кварталов

Квартал t	Стоимость акций y_t	Модель Хольта–Уинтерса y_t^p	Абсолютная погрешность ε_t	Относительная погрешность $\|\varepsilon_t\| / y_t$	Точки поворота
1	304	297,32	6,68	2,20	–
2	320	316,97	3,03	0,95	0
3	334	336,68	–2,68	0,80	1
4	347	348,02	–1,03	0,29	1
5	323	338,08	–15,08	4,67	1
6	342	348,94	–6,94	2,03	0
7	365	361,47	3,53	0,97	1
8	375	373,77	1,23	0,33	0
9	342	355,09	–13,09	3,83	1
10	365	369,52	–4,32	1,18	1
11	378	386,65	–8,65	2,29	1
12	399	392,11	6,89	1,73	1
13	363	365,71	–2,71	0,75	1
14	388	386,78	1,22	0,31	0
15	419	404,26	14,74	3,52	1
16	418	425,73	–7,73	1,85	–

Проверка качества модели. Для того чтобы модель была качественной уровни (значения) остатков $\varepsilon_t = y_t - y_t^p$ (разности между фактическими и расчётными значениями) должны удовлетворять определённым условиям (точности и адекватности).

Для проверки этих условий сведём необходимые данные в табл. 5.4.

1. Проверка точности модели. Будем считать, что условие точности выполнено, если относительная погрешность в среднем не превышает 5%.

Вычислим относительную погрешность как отношение абсолютного отклонения показателя к его фактическому значению:

$$\bar{\varepsilon}_{отн} = \frac{1}{n}\sum_{t=1}^{n}\frac{|\varepsilon_t|}{y_t} = \frac{27,7}{16}1,73\%.$$

Условие точности выполнено

2. Проверка адекватности модели

Для адекватности модели ряд остатков $\varepsilon_t = y_t - y_t^p$ должен обладать свойствами:

1) случайности,

2) независимости последовательности уровней,

3) нормальности распределения.

2.1. Проверку **случайности уровней** остатков ряда проводим на основе **критерия поворотных точек**. Для этого каждый уровень ряда остатков сравниваем с двумя соседними.

Если он больше (либо меньше) обоих соседних уровней, то точка считается поворотной и в соответствующей строчке табл. 5.4 ставится 1, в противном случае ставится 0.

В первой и последней строке ставится прочерк, т.к. у этих уровней нет двух соседних точек.

Общее число поворотных точек в примере равно $N_{пов} = 10$.

Рассчитаем значение критерия поворотных точек как целую часть выражения:

$$Z = int[2(n-2)/3 - 2\sqrt{(16n-29)/90}].$$

При $n = 16$ получаем:

$$Z = int[2(16-2)/3 - 2\sqrt{(16 \cdot 16 - 29)/90}] =$$

$$= int[28/3 - 2\sqrt{227/90}] = int[9,33 - 3,18] = 6.$$

Если количество поворотных точек $N_{пов}$ больше Z, то условие случайности выполнено.

В нашем случае

$$N_{пов} = 10; \; Z = 6.$$

Это означает, что условие случайности выполнено.

2.2. **Проверка отсутствия автокорреляции** (независимости уровней остатков).

Эту проверку проведём двумя методами:

1) по критерию Дарбина–Уотсона;

2) по первому коэффициенту автокорреляции.

1. Вычислим значение критерия Дарбина–Уотсона:

$$D_H = \frac{\sum_{t=2}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n} (\varepsilon_t)^2} = \frac{1765,08}{949,77} = 1,86.$$

Полученное значение критерия сравнивают с критическими точками $D_{кр1}$, $D_{кр2}$, взятыми из соответствующих таблиц.

В нашем случае

$$D_{кр1} = 1,08; \;\; D_{кр2} = 1,36.$$

Если $D_{кр2} < D < 2$, то уровни ряда остатков являются независимыми.

В нашем случае $D_{кр2} = 1{,}36 < D_H = 1{,}86 < 2.$

Следовательно, уровни ряда остатков являются независимыми.

2. Вычислим первый коэффициент автокорреляции остатков по формуле

$$r(1) = \frac{\sum_{t=2}^{n}(\varepsilon_t \cdot \varepsilon_{t-1})}{\sum_{t=1}^{n}(\varepsilon_t)^2} = \frac{15,05}{949,77} = 0,02.$$

Если модуль вычисленного первого коэффициента автокорреляции меньше критического значения $r_{кр}$ (которое можно взять из соответствующей таблицы), то уровни остатков ряда будут независимы.

У нас $r(1) = 0,02;\ r_{кр} = 0,32.$

Следовательно, уровни ряда остатков являются независимыми.

2.3. Проверим соответствие ряда остатков нормальному распределению по RS **– критерию**.

Рассчитаем значение RS – критерия по формуле:

$$RS = (\varepsilon_{max} - \varepsilon_{min}) / \sigma_\varepsilon.$$

Здесь ε_{max}, ε_{min} – максимальное и минимальное значения уровней ряда остатков, а σ_ε – среднее квадратическое отклонение ряда остатков.

Имеем:

$$\varepsilon_{max} = 14,74;\ \varepsilon_{min} = -15,08.$$

Среднее квадратическое отклонение ряда остатков вычислим по формуле

$$\sigma_\varepsilon = \sqrt{\frac{\sum_{t=2}^{n}(\varepsilon_t)^2}{n-1}} = \sqrt{\frac{949,77}{15}} = 7,96.$$

Рассчитаем значение RS – критерия по формуле:

$$RS = (\varepsilon_{max} - \varepsilon_{min}) / \sigma_\varepsilon = (14,74 - (-15,08)) / 7,96 = 3,75$$

Полученное значение критерия сравнивают с табличными значениями, которые зависят от количества точек и уровня значимости.

Для уровня значимости 5% и числа точек 16 полученное значение RS для нормального распределения должно находиться в интервале от 3,00 до 4,21.

Так как 3,00 < RS =3,75 < 4,21, то ряд остатков подчиняется нормальному распределению.

Все условия адекватности и точности выполнены. Можно говорить о возможности прогноза показателя на 4 квартала вперёд.

3. Расчёт прогнозных значений показателя

Рассчитаем значения y_t^p для $t = 17$ (полагая в правой части формул $t = 16,\ \tau = 1$):

$$y_{17}^{p}=(a_{16}+1\cdot b_{16})s_{16+1-4}=(a_{16}+1\cdot b_{16})s_{13}=$$
$$=(409,21+1\cdot 6,26)\cdot 0,9360=388,9.$$

Рассчитаем значения y_t^p для $t=18$ (полагая в правой части формул $t=16,\ \tau=2$):

$$y_{18}^{p}=(a_{16}+2\cdot b_{16})s_{16+2-4}=(a_{16}+2\cdot b_{16})s_{14}=$$
$$=(409,21+2\cdot 6,26)\cdot 0,9826=414,4.$$

Рассчитаем значения y_t^p для $t=19$ (полагая в правой части формул $t=16,\ \tau=3$):

$$y_{19}^{p}=(a_{16}+3\cdot b_{16})s_{16+3-4}=(a_{16}+3\cdot b_{16})s_{15}=$$
$$=(409,21+3\cdot 6,26)\cdot 1,0256=438,9.$$

Рассчитаем значения y_t^p для $t=20$ (полагая в правой части формул $t=16,\ \tau=4$):

$$y_{20}^{p}=(a_{16}+4\cdot b_{16})s_{16+4-4}=(a_{16}+4\cdot b_{16})s_{16}=$$
$$=(409,21+4\cdot 6,26)\cdot 1,0268=445,9.$$

Расчётные данные хорошо согласуются с фактическими, что говорит об удовлетворительном качестве прогноза.

Пример 5.3. Имеются данные об объёмах потребления электроэнергии жителями города за 18 кварталов (табл. 5.5). Различными методами спрогнозировать потребление электроэнергии жителями города на ближайшие полгода.

Решение. Исследуем структуру этого ряда.

Определим коэффициенты автокорреляции различных порядков.

Таблица 5.5

Потребление электроэнергии жителями города, млн. кВт·ч

t	y_t	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}
1	30				
2	22	30			
3	25	22	30		
4	45	25	22	30	
5	36	45	25	22	30
6	24	36	45	25	22
7	30	24	36	45	25
8	50	30	24	36	45
9	48	50	30	24	36
10	28	48	50	30	24
11	32	28	48	50	30
12	55	32	28	48	50
13	45	55	32	28	48
14	33	45	55	32	28
15	35	33	45	55	32
16	54	35	33	45	55
17	42	54	35	33	45
18	30	42	54	35	33

Значения АКФ этого ряда и коррелограмма приведены в табл. 5.6.

Коэффициент автокорреляции первого порядка $(r_{yy[1]} = 0{,}180)$ свидетельствует о слабой зависимости текущих уровней ряда от непосредственно им предшествующих уровней.

Структура ряда такова, что каждый следующий уровень y_t зависит от уровня y_{t-4} и y_{t-2} в гораздо большей степени, чем от уровня y_{t-1}.

Анализ значений АКФ позволяет сделать вывод о наличии в изучаемом временном ряде линейной тенденции и сезонных колебаний периодичностью в четыре квартала.

Данный вывод подтверждается и графическим анализом структуры ряда.

Таблица 5.6

Коррелограмма временного ряда потребления электроэнергии

Лаг $\boldsymbol{\tau}$	Коэффициенты автокорреляции уровней $\left\|r_{yy[\tau]}\right\|$	Коррелограмма
1	0,1802	**
2	0,6115	*******
3	0,1237	*
4	0,9190	************
5	0,0884	
6	0,7718	*********
7	0,0208	
8	0,8171	***********
9	0,0698	

1. Построим аддитивную модель ряда потребления электроэнергии жителями города за 4,5 года.

Данный ряд содержит сезонные колебания периодичностью 4.

Объёмы потребления электроэнергии в осенне–зимний период времени (I и IV кварталы) выше, чем весной и летом (II и III кварталы).

По графику этого ряда можно установить наличие приблизительно равной амплитуды колебаний.

Это свидетельствует о возможном существовании аддитивной модели временного ряда.

Рассчитаем её компоненты.

Шаг 1. Проведём выравнивание исходных уровней ряда методом скользящей средней. Для этого:

1) просуммируем уровни ряда последовательно за каждые четыре квартала со сдвигом на один момент времени и определим условные годовые объёмы потребления электроэнергии (гр.3 табл. 5.7);

2) разделив полученные суммы на 4, найдём скользящие средние (гр.4 табл. 5.7). Полученные таким образом выравненные значения уже не содержат сезонной компоненты;

3) приведём эти значения в соответствие с фактическими моментами времени, для чего найдём средние значения из двух последовательных скользящих средних, т.е. центрированные скользящие средние (гр.5 табл. 5.7).

Таблица 5.7

Расчёт оценок сезонной компоненты в аддитивной модели

№ квар–тала, t	Потребление электро–энергии, y_t	Итого по 4 квар-талам	Скользящая средняя по кварталам	Центрированная скользящая средняя	Оценка сезонной компоненты
1	**2**	**3**	**4**	**5**	**6**
1	30				
2	22	122	30,50		
3	25	128	32,00	31,250	–6,250
4	45	130	32,50	32,250	12,750
5	36	135	33,75	33,125	2,875
6	24	140	35,00	34,375	–10,375
7	30	152	38,00	36,500	–6,500
8	50	156	39,00	38,500	11,500
9	48	158	39,50	39,250	8,750
10	28	163	40,75	40,125	–12,125
11	32	160	40,00	40,375	–8,375
12	55	165	41,25	40,625	14,375
13	45	168	42,00	41,625	3,375
14	33	167	41,75	41,875	–8,875
15	35	164	41,00	41,375	–6,375
16	54	161	40,25	40,625	13,375
17	42				
18	30				

Шаг 2. Найдём оценки сезонной компоненты как разность между фактическими уровнями ряда и центрированными скользящими средними (гр.6 табл. 5.7).

Найдём (табл. 5.8) средние за каждый квартал (по всем годам) оценки сезонной компоненты $\overline{s_i}$.

В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются.

В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна нулю:

$$\sum_{i=1}^{L} s_i = 0.$$

Для данной модели имеем:

$$5{,}000 - 10{,}458 - 6{,}875 + 13{,}000 = 0{,}668.$$

Определим корректирующий коэффициент $k = \dfrac{0,668}{4} = 0,167$.

Рассчитаем скорректированные значения сезонной компоненты как разность между её средней оценкой и корректирующим коэффициентом *k*:

$$s_i = \overline{s_i} - \frac{\sum_{i=1}^{L} \overline{s_i}}{L}, \quad \text{где } i = \overline{1,4}.$$

Проверим условие равенства нулю суммы значений сезонной компоненты:

$$4,833 - 10,625 - 7,041 + 12,833 = 0.$$

Таким образом, получены следующие значения сезонной компоненты по кварталам года:

$$s_1 = 4,833;\ s_2 = -10,625;\ s_3 = -7,041;\ s_4 = 12,833.$$

Занесём полученные значения в табл. 5.9 для соответствующих кварталов каждого года (гр.3 табл. 5.9).

Таблица 5.8

Расчёт значений сезонной компоненты в аддитивной модели

Показатели	№ квартала, *i*			
	I	II	III	IV
Сезонная компонента 1–го года			–6,250	12,750
Сезонная компонента 2–го года	2,875	–10,375	–6,500	11,500
Сезонная компонента 3–го года	8,750	–12,125	–8,375	14,375
Сезонная компонента 4–го года	3,375	–8,875	–6,375	13,375
Итого за *i*–й квартал (за все годы)	15,000	–31,375	–25,500	52,000
Средняя оценка сезонной компоненты для *i*–го квартала, $\overline{s_i}$	5,000	–10,458	–6,875	13,000
Скорректированная сезонная компонента, s_i	4,833	–10,625	–5,041	12,833

Шаг 3. Элиминируем влияние сезонной компоненты, вычитая её значение из каждого уровня исходного ряда.

Получим величины $\overline{y}_t + \varepsilon_t = y_t - s_t$ (гр.4 табл. 5.9).

Эти значения рассчитываются за каждый момент времени и содержат только тенденцию и случайную компоненту.

Шаг 4. Определим компоненту $\overline{y}_t$ данной модели.

Для этого проведём аналитическое выравнивание ряда ($\overline{y}_t + \varepsilon_t$) с помощью линейного тренда.

Результаты аналитического выравнивания следующие:

$$\overline{y}_t = 29,995 + 0,759\,t.$$

Подставляя в это уравнение значения t = 1, ..., 18, найдём уровни $\overline{y}_t$ для каждого момента времени (гр.5 табл. 5.9).

Шаг 5. Найдём значения уровней ряда, полученные по аддитивной модели. Для этого прибавим к уровням тренда значения сезонной компоненты для соответствующих кварталов $\overline{y}_t + \varepsilon_t$.

Таблица 5.9

Расчёт выравненных значений тренда и ошибок в аддитивной модели

t	y_t	s_t	$\overline{y}_t + \varepsilon_t$	$\overline{y}_t$	$\overline{y}_t + s_t$	ε_t	ε_t^2
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	30	4,833	25,167	30,756	35,589	–5,589	31,237
2	22	10,625	32,625	31,515	20,890	1,110	1,232
3	25	–5,041	32,041	32,274	25,233	–0,233	0,054
4	45	12,833	32,167	33,033	45,866	–0,866	0,750
5	36	4,833	31,167	33,792	38,625	–2,625	6,891
6	24	–10,625	34,625	34,551	23,926	0,074	0,005
7	30	–5,041	35,041	35,310	28,269	1,731	2,996
8	50	12,833	35,167	36,069	48,902	1,098	1,206
9	48	4,833	43,167	36,828	41,661	6,339	40,183
10	28	–10,625	38,625	35,587	26,962	1,038	1,077
11	32	–5,041	39,041	38,346	31,305	0,695	0,483
12	55	12,833	42,167	39,105	51,938	3,062	9,376
13	45	4,833	40,167	39,864	44,697	0,303	0,092
14	33	–10,625	43,625	40,623	29,998	3,002	9,012
15	35	–5,041	42,041	41,382	34,341	0,659	0,434
16	54	12,833	41,167	42,141	54,974	–0,974	0,949
17	42	4,833	35,167	42,900	45,733	–5,733	32,867
18	30	–10,625	40,625	43,659	33,034	–3,034	9,205

Шаг 6. В соответствии с методикой построения аддитивной модели расчёт ошибки производится по формуле

$$\varepsilon_t = y_t - \left(\overline{y}_t + s_t\right).$$

Это абсолютная ошибка.

Численные значения абсолютных ошибок приведены в гр.7 табл. 5.9.

По аналогии с моделью регрессии для оценки качества построенной модели или для выбора наилучшей модели можно применять сумму квадратов полученных абсолютных ошибок.

Для данной аддитивной модели сумма квадратов абсолютных ошибок равна

$$S_{ocm}^2 = \sum_{t=1}^{n} \varepsilon_t^2 = 148,05.$$

Среднее значение уровней исходного ряда:

$$\bar{y} = \frac{1}{18}\sum_{t=1}^{18} y_t = \frac{668}{18} = 36,889.$$

Общая сумма квадратов отклонений уровней ряда от его среднего уровня:

$$S^2_{общ} = \sum_{t=1}^{n} (y_t - \bar{y})^2 = 1867,778.$$

Коэффициент детерминации равен

$$R^2 = 1 - \frac{S^2_{ост}}{S^2_{общ}} = 1 - \frac{148,05}{1867,778} = 0,9207.$$

Аддитивная модель объясняет 92,07% общей вариации уровней ряда потребления электроэнергии за последние 18 кварталов.

Шаг 7. Спрогнозируем потребление электроэнергии жителями города в течение второго полугодия последнего года по аддитивной модели ряда.

Прогнозное значение $y^{пр}_{t+\tau}$ уровня ряда в аддитивной модели есть сумма трендовой и сезонной составляющих.

Объём электроэнергии, потреблённой в течение второго полугодия пятого года, рассчитывается как сумма объёмов потребления электроэнергии в III и в IV кварталах пятого года, соответственно $y^{пр}_{19}$ и $y^{пр}_{20}$.

Для определения трендовой компоненты воспользуемся уравнением тренда

$$\bar{y}_t = 29,997 + 0,759\, t.$$

Получим

$$\bar{y}_{19} = 29,997 + 0,759 \cdot 19 = 44,418;$$
$$\bar{y}_{20} = 29,997 + 0,759 \cdot 20 = 45,177.$$

Значения сезонной компоненты равны

$s_3 = -7,041$ (III квартал); $s_4 = 12,833$ (IV квартал).

Таким образом,

$$y^{пр}_{19} = \bar{y}_{19} + s_3 = 44,418 - 7,041 = 37,377;$$
$$y^{пр}_{20} = \bar{y}_{20} + s_4 = 45,177 + 12,833 = 58,010.$$

Прогноз объёма потребления электроэнергии на второе полугодие последнего года составит

$$37,377 + 58,010 = 95,387 \text{ млн. кВт·ч.}$$

2. Построим мультипликативную модель ряда.

Поскольку амплитуда сезонных колебаний немного изменяется со временем, можно предположить существование мультипликативной модели. Определим её компоненты.

Шаг 1. Проведём выравнивание исходных уровней ряда методом скользящей средней (по аналогии с методикой выравнивания при построении аддитивной модели). Результаты расчётов оценок сезонной компоненты представлены в табл. 5.10.

Шаг 2. Найдём оценки сезонной компоненты как частное от деления фактических уровней ряда на центрированные скользящие средние (гр.6 табл. 5.10).

Используем эти оценки для расчета значений сезонной компоненты (табл. 5.11). Для этого найдём средние за каждый квартал оценки сезонной компоненты S_i.

Таблица 5.10

Расчёт оценок сезонной компоненты в мультипликативной модели

№ квартала, t	Потребление электро–энергии, y_t	Итого за четыре квартала	Скользящая средняя по кварталам	Центриро–ванная скользящая средняя	Оценка сезонной компоненты
1	**2**	**3**	**4**	**5**	**6**
1	30				
2	22				
3	25	122	30,50	31,250	0,800
4	45	128	32,00	32,250	1,395
5	36	130	32,50	33,125	1,087
6	24	135	33,75	34,375	0,698
7	30	140	35,00	36,500	0,822
8	50	152	38,00	38,500	1,299
9	48	156	39,00	39,250	1,223
10	28	158	39,50	40,125	0,698
11	32	163	40,75	40,375	0,793
12	55	160	40,00	40,325	1,354
13	45	165	41,25	41,625	1,081
14	33	168	42,00	41,875	0,788
15	35	167	41,75	41,375	0,846
16	54	164	41,00	40,625	1,329
17	42	161	40,25		
18	30				

Взаимопогашаемость сезонных воздействий в мультипликативной модели выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна числу периодов в цикле.

В нашем случае число периодов одного цикла (год) равно 4 (четыре квартала). Имеем:

$$1{,}130+0{,}728+0{,}815+1{,}344=4{,}017.$$

Определим корректирующий коэффициент

$$k=\frac{4}{4{,}017}=0{,}996.$$

Определим скорректированные значения сезонной компоненты, умножив её средние оценки на корректирующий коэффициент k:

$$s_i=\overline{s_i}\cdot k;\quad i=\overline{1,4}.$$

Таблица 5.11

Расчёт сезонной компоненты в мультипликативной модели

Показатели	№ квартала, i			
	I	II	III	IV
Сезонная компонента 1–го года			0,800	1,395
Сезонная компонента 2–го года	1,087	0,698	0,822	1,299
Сезонная компонента 3–го года	1,223	0,698	0,793	1,354
Сезонная компонента 4–го года	1,081	0,788	0,846	1,329
Итого за i–й квартал (за все годы)	3,391	2,184	3,260	5,377
Средняя оценка сезонной компоненты для i–го квартала, $\overline{s_i}$	1,130	0,728	0,815	1,344
Скорректированная сезонная компонента, s_i	1,125	0,725	0,812	1,338

Проверим условие равенства суммы значений сезонной компоненты s_i числу 4:

$$1{,}125+0{,}725+0{,}812+1{,}338=4.$$

Получим следующие значения сезонной компоненты s_i:

$$s_1=1{,}125;\quad s_2=0{,}725;\quad s_3=0{,}812;\quad s_4=1{,}338.$$

Занесём полученные значения в табл. 5.12 для соответствующих кварталов каждого года (гр.3).

Шаг 3. Разделим каждый уровень исходного ряда на соответствующие значения сезонной компоненты.

Получим величины $\overline{y}_t\varepsilon_t=\dfrac{y_t}{s_t}$ (гр.4 табл. 5.12), которые содержат только тенденцию и случайную компоненту.

Шаг 4. Определим параметры линейного тренда, используя уровни $\overline{y}_t\varepsilon_t$. Результаты аналитического выравнивания этого ряда дают уравнение тренда в виде:

$$\overline{y}_t=29{,}630+0{,}801\,t.$$

Подставляя в это уравнение значения t = 1, ..., 18, найдём уровни $\overline{y}_t$ для каждого момента времени (гр.5 табл. 5.12).

Шаг 5. Найдём уровни ряда по мультипликативной модели, умножив уровни $\overline{y}_t$ на значения сезонной компоненты для соответствующих кварталов (гр.6 табл. 5.12).

Шаг 6. Расчёт ошибки в мультипликативной модели производится по формуле $\varepsilon_t = \dfrac{y_t}{\overline{y}_t s_t}$ (гр.7 табл. 5.8).

Таблица 5.12

Расчёт тренда и ошибок в мультипликативной модели

t	y_t	s_t	$\overline{y}_t\varepsilon_t = \dfrac{y_t}{s_t}$	$\overline{y}_t$	$\overline{y}_t s_t$	$\varepsilon_t = \dfrac{y_t}{\overline{y}_t s_t}$	$\Delta\varepsilon_t = y_t - \overline{y}_t s_t$	$\left(\Delta\varepsilon_t\right)^2$
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	30	1,125	26,667	30,431	34,235	0,876	–4,235	15,934
2	22	0,725	30,345	31,232	22,643	0,972	–0,643	0,414
3	25	0,812	30,788	32,033	26,011	0,961	–1,011	1,022
4	45	1,338	33,632	32,834	43,932	1,024	1,068	1,141
5	36	1,125	32,000	33,635	35,839	0,951	–1,839	3,383
6	24	0,725	33,103	34,436	24,966	0,961	–0,966	0,933
7	30	0,812	36,946	35,237	28,612	1,048	1,388	1,925
8	50	1,338	35,369	36,038	48,219	1,037	1,781	3,173
9	48	1,125	42,667	36,839	41,444	1,158	6,556	42,983
10	28	0,725	38,621	35,640	25,289	1,026	0,711	0,506
11	32	0,812	39,409	38,441	31,214	1,025	0,786	0,618
12	55	1,338	41,106	39,242	52,506	1,048	2,494	6,221
13	45	1,125	40,000	40,043	45,048	0,999	–0,048	0,002
14	33	0,725	45,517	40,844	29,612	1,114	3,388	11,479
15	35	0,812	43,103	41,645	33,816	1,035	1,184	1,402
16	54	1,338	40,359	42,446	56,793	0,951	–2,793	5,799
17	42	1,125	35,333	43,247	48,653	0,863	–6,653	44,261
18	30	0,725	41,379	44,048	31,935	0,939	–1,935	3,743

Если ряд ошибок не содержит автокорреляции, его можно использовать вместо исходного для изучения его взаимосвязи с другими рядами.

Для того чтобы сравнить мультипликативную модель и другие модели временного ряда, можно по аналогии с аддитивной моделью использовать сумму квадратов абсолютных ошибок. Абсолютные ошибки в мультипликативной модели определяются как

$$\Delta\varepsilon_t = y_t - \overline{y}_t s_t .$$

В данной модели сумма квадратов абсолютных ошибок составляет 148,940. Общая сумма квадратов отклонений фактических уровней этого ряда от среднего значения равна

$$\sum_{t=1}^{n}(y_t-\bar{y})^2=1867,778.$$

Коэффициент детерминации модели равен

$$R^2=1-\frac{S^2_{ост}}{S^2_{общ}}=1-\frac{148,94}{1867,778}=0,9203.$$

Доля объяснённой дисперсии уровней ряда равна 0,9203 или 92,03%.

Выявление и устранение сезонного эффекта («десезонализация уровней ряда») используются в двух направлениях.

1. Воздействие сезонных колебаний следует устранять на этапе предварительной обработки исходных данных при изучении взаимосвязи нескольких рядов. Поэтому в российских и международных статистических сборниках часто публикуются данные, в которых устранено влияние сезонной компоненты (если это помесячная или поквартальная статистика), например, показатели объёмов производства в отдельных отраслях промышленности, уровня безработицы и т.д.

2. Сезонный эффект используется в прогнозировании уровней ряда в будущие моменты времени.

Шаг 7. Спрогнозируем потребление электроэнергии на ближайшие полгода по мультипликативной модели.

Прогнозное значение $y^{пр}_{t+\tau}$ уровня ряда в мультипликативной модели есть произведение трендовой и сезонной составляющих.

Для определения трендовой компоненты за каждый квартал воспользуемся уравнением тренда

$$\bar{y}_t=29,630+0,801\,t.$$

Получим

$$\bar{y}_{19}=29,630+0,801\cdot 19=44,849;$$

$$\bar{y}_{20}=29,630+0,801\cdot 20=45,650.$$

Значения сезонной компоненты равны

$s_3=0,812$ (III квартал); $s_4=1,338$ (IV квартал).

Таким образом,

$$y^{пр}_{19}=\bar{y}_{19}\cdot s_3=44,849\cdot 0,812=36,417;$$

$$y^{пр}_{20}=\bar{y}_{20}\cdot s_4=45,650\cdot 1,338=61,080.$$

Прогноз объёма потребления электроэнергии на второе полугодие последнего года составит

$36,417+61,080=97,497$ млн. кВт·ч.

3. Можно строить **модели регрессии с включением фактора времени и фиктивных переменных**. Количество фиктивных переменных в такой модели должно быть на единицу меньше числа моментов (периодов) времени внутри одного цикла колебаний. Например, при моделировании поквартальных данных модель должна включать 4 независимые переменные – фактор времени и три фиктивные переменные.

Каждая фиктивная переменная отражает сезонную (циклическую) компоненту ряда для какого–либо одного периода. Она равна единице для данного периода и нулю для всех остальных периодов.

Пусть имеется временной ряд, содержащий циклические колебания периодичностью L. Модель регрессии с фиктивными переменными для этого ряда будет иметь вид:

$$y_t = a + bt + c_1 x_1 + ... + c_j x_j + ... + c_{L-1} x_{L-1} + \varepsilon_t,$$

где $x_j = \begin{cases} 1 - \text{для каждого } j\text{-го квартала внутри каждого года;} \\ 0 - \text{ в других случаях.} \end{cases}$

Например, при моделировании сезонных колебаний на основе поквартальных данных за несколько лет число кварталов внутри одного года $L = 4$, а общий вид модели следующий:

$$y_t = a + bt + c_1 x_1 + c_2 x_2 + c_3 x_3 + \varepsilon_t,$$

где $x_1 = \begin{cases} 1 - \text{для первого квартала;} \\ 0 - \text{в других случаях;} \end{cases}$ $x_2 = \begin{cases} 1 - \text{для второго квартала;} \\ 0 - \text{в других случаях;} \end{cases}$

$$x_3 = \begin{cases} 1 - \text{для третьего квартала;} \\ 0 - \text{в других случаях.} \end{cases}$$

Уравнения трендов для кварталов будет иметь следующий вид:

для 1–го квартала: $y_t = a + bt + c_1 + \varepsilon_t$;

для 2–го квартала: $y_t = a + bt + c_2 + \varepsilon_t$;

для 3–го квартала: $y_t = a + bt + c_3 + \varepsilon_t$.

Фиктивные переменные позволяют дифференцировать величину свободного члена уравнения регрессии для каждого квартала. Она составит для I кв.: $(a + c_1)$; для II кв.: $(a + c_2)$; для III кв.: $(a + c_3)$; для IV кв.: a.

Параметр b в этой модели характеризует среднее абсолютное изменение уровней ряда под воздействием тенденции.

Рассмотренная модель есть аналог аддитивной модели ряда, поскольку фактический уровень ряда есть сумма трендовой, сезонной и случайной составляющих.

Основной недостаток модели с фиктивными переменными для описания сезонных и циклических колебаний – наличие большого количества переменных.

Применительно к нашей задаче о потреблении электроэнергии в модели будут четыре независимые переменные: t, x_1, x_2, x_3 и результативная переменная y_t. Составим матрицу исходных данных (табл. 5.13).

Обычным МНК оценим параметры уравнения регрессии

$$y_t = a + bt + c_1 x_1 + c_2 x_2 + c_3 x_3 + \varepsilon_t.$$

Результаты оценки уравнения регрессии приведены в табл. 5.14.

Уравнение регрессии имеет вид:

$$\bar{y}_t = 43,5 + 0,75t - 10,05x_1 - 23,6x_2 - 19,75x_3.$$

Коэффициент детерминациии этого уравнения $R^2 = 0,929$.

Проанализируем полученные результаты.

Влияние сезонной компоненты в каждом квартале статистически значимо (фактические значения T–критерия по модулю больше 2 для параметров при переменных x_1, x_2, x_3 и константы a).

Параметр a = 43,5 есть сумма начального уровня ряда и сезонной компоненты в IV квартале.

Сезонные колебания в I, II и III кварталах приводят к снижению этой величины, о чем свидетельствуют отрицательные оценки параметров при переменных x_1, x_2 и x_3.

Отметим, что эти параметры не равны значениям сезонной компоненты, поскольку они характеризуют не сезонные изменения уровней ряда, а их отклонения от уровней, учитывающих сезонные воздействия в IV квартале.

Положительная величина параметра b = 0,75 при переменной времени свидетельствует о наличии возрастающей тенденции в уровнях ряда.

Его абсолютное значение говорит о том, что средний за квартал абсолютный прирост объёма потребления электроэнергии составляет 0,75 млн. кВт·ч.

Так как фактическое значение T–критерия равно 5,14, можно утверждать, что существование в уровнях ряда тенденции установлено.

Общая сумма квадратов отклонений уровней ряда от общей средней составляет величину $S^2_{общ} = \sum_{t=1}^{n} \left(y_t - \bar{y} \right)^2 = 1867,778$.

С использованием коэффициента детерминации модели (R^2 = 0,929) определим остаточную сумму квадратов уровней ряда

$$S^2_{ост} = (1 - R^2) \cdot S^2_{общ} = (1 - 0,929) \cdot 1867,778 = 132,612.$$

Остаточная сумма квадратов абсолютных ошибок по аддитивной модели была рассчитана ранее (табл. 5.9) и составляет 148,05. Следовательно, модель регрессии с фиктивными переменными описывает динамику ряда потребления электроэнергии лучше, чем аддитивная модель.

Таблица 5.13

Данные для расчёта параметров уравнения регрессии
с фиктивными переменными по ряду потребления электроэнергии

t	x_1	x_2	x_3	y_t
1	1	0	0	30
2	0	1	0	22
3	0	0	1	25
4	0	0	0	45
5	1	0	0	36

6	0	1	0	24
7	0	0	1	30
8	0	0	0	50
9	1	0	0	48
10	0	1	0	28
11	0	0	1	32
12	0	0	0	55
13	1	0	0	45
14	0	1	0	33
15	0	0	1	35
16	0	0	0	54
17	1	0	0	42
18	0	1	0	30

Таблица 5.14

Уравнение регрессии с фиктивными переменными

Переменная	Коэффициент	Стандартная ошибка	T−критерий
Константа	43,50	2,17	20,09
t	0,75	0,15	5,14
x_1	−10,05	2,15	−4,67
x_2	−23,60	2,15	−11,00
x_3	−19,75	2,27	−8,71

Выполним прогнозирование по полученной модели

$$\overline{y}_t = 43,5 + 0,75t - 10,05x_1 - 23,6x_2 - 19,75x_3.$$

Потребление электроэнергии за третий квартал последнего года:

$$\overline{y}_{19} = 43,5 + 0,75 \cdot 19 - 10,05 \cdot 0 - 23,6 \cdot 0 - 19,75 \cdot 1 = 38.$$

Потребление электроэнергии за четвёртый квартал последнего года:

$$\overline{y}_{20} = 43,5 + 0,75 \cdot 20 - 10,05 \cdot 0 - 23,6 \cdot 0 - 19,75 \cdot 0 = 58,5.$$

Потребление электроэнергии за ближайшие полгода:

$$\overline{y}_{19+20} = 38 + 58,5 = 96,5.$$

Пример 5.4. Пусть имеются три конкурирующих изделия X_1, X_2, X_3.

С целью определения спроса на эти изделия произведён опрос 100 человек. Оказалось, что изделие X_1 покупает 50 человек, изделие X_2 – 20 человек, а X_3 – 30 человек.

Предположим, что поведение покупателей в каждый следующий месяц обусловлено только их поведением в предыдущий месяц (таким образом, исследуется простая цепь Маркова).

По истечении месяца оказалось, что из 50 человек, покупавших изделие X_1, 45 человек продолжают его покупать, 4 человека стали покупать изделие X_2 и 1 – изделие X_3. Из 20 человек, покупавших изделие X_2, 6 человек продолжают его покупать, 8 стали покупать изделие X_1, 6 – изделие X_3. Из 30 человек, покупавших изделие X_3, 6 человек продолжают его покупать, 21 человек стал покупать изделие X_1, 3 – изделие X_2.

Требуется определить, какое изделие будет пользоваться наибольшим спросом по истечении месяца? Через 2 месяца? Через год?

Решение. Если $P_k(t_0)$ – вероятности потребности в изделии X_k в момент t_0, то из условий задачи $P_0(t_0) = (0,5;\ \ 02;\ \ 0,3)$.

Переходные вероятности определяются из условия задачи (рис. 5.2):

$$P_{11} = \frac{45}{50} = 0{,}9; \quad P_{12} = \frac{4}{50} = 0{,}08; \quad P_{13} = \frac{1}{50} = 0{,}02;$$

$$P_{21} = \frac{8}{20} = 0{,}4; \quad P_{22} = \frac{6}{20} = 0{,}3; \quad P_{23} = \frac{6}{20} = 0{,}3;$$

$$P_{31} = \frac{21}{30} = 0{,}7; \quad P_{32} = \frac{3}{30} = 0{,}1; \quad P_{33} = \frac{6}{30} = 0{,}2.$$

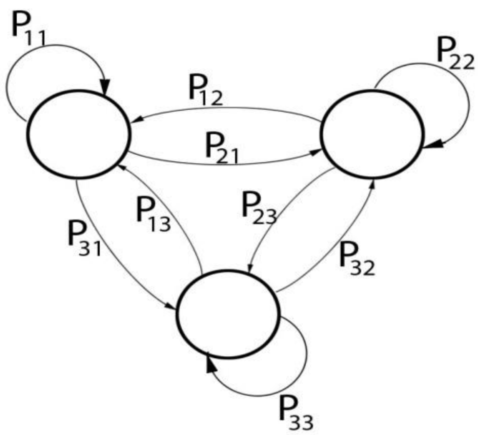

Рис. 5.2. Граф переходов и состояний системы

Искомые вероятности получаются умножением вектора вероятностей состояния цепи Маркова на переходную матрицу вероятностей:

$$P_j(k) = \sum_{i=1}^{n} P_i(k-1)P_{ij}^k,$$

где P_{ij}^k – вероятность перехода системы из i–го состояния в j–е на k–м шаге.

Получаем вероятности спроса изделий для первого месяца:

$$P(t_1) = P(t_0)\left\|P_{ij}\right\| = (0,5;\ 0,2;\ 0,3)\begin{vmatrix} 0,9 & 0,08 & 0,02 \\ 0,4 & 0,3 & 0,3 \\ 0,7 & 0,1 & 0,2 \end{vmatrix} = (0,74;\ 0,13;\ 0,13).$$

Через месяц наибольшим спросом будет пользоваться изделие X_1.

Если предположить, что поведение покупателей со временем не меняется, т.е. что цепь однородна по времени, то аналогично можно определить, какое изделие будет пользоваться наибольшим спросом по истечении двух, трёх т.д. месяцев. Вероятности спроса изделий через два месяца:

$$P(t_2) = P(t_1)\left\|P_{ij}\right\| = (0,74;\ 0,13;\ 0,13)\begin{vmatrix} 0,9 & 0,08 & 0,02 \\ 0,4 & 0,3 & 0,3 \\ 0,7 & 0,1 & 0,2 \end{vmatrix} = (0,809;\ 0,111;\ 0,080).$$

Определим, какое изделие будет пользоваться наибольшим спросом по истечении достаточно продолжительного периода (год).

Все элементы матрицы перехода положительны, т.е. условие эргодичности выполняется, следовательно, предельные вероятности P_1, P_2, P_3 потребления изделий X_1, X_2, X_3 существуют.

Система уравнений в данном случае имеет вид $P_j = \sum_{i=1}^{n} P_i P_{ij}$, т.е.

$$\begin{cases} P_1 = 0,9P_1 + 0,4P_2 + 0,7P_3; \\ P_2 = 0,08P_1 + 0,3P_2 + 0,1P_3; \\ P_3 = 0,02P_1 + 0,3P_2 + 0,2P_3. \end{cases}$$

Система линейно зависима. Заменяя третье уравнение системы уравнением $P_1 + P_2 + P_3 = 1$, получим систему

$$\begin{cases} P_1 = 0,9P_1 + 0,4P_2 + 0,7P_3; \\ P_2 = 0,08P_1 + 0,3P_2 + 0,1P_3; \\ P_1 + P_2 + P_3 = 1. \end{cases}$$

Решение системы:

$$P_1 = 0,84;\ P_2 = 0,10;\ P_3 = 0,06$$