**Компьютерный практикум**

**Выполнила студентка группы ОБ19-8**

**Арутюнян Элеонора**

**Условие:**

Считая, что в данной таблице представлен помесячный объем производства молока за несколько лет, построить аддитивную модель временного ряда, выяснив длину цикла и проведя выравнивание методом скользящей средней. Определить циклическую компоненту, построить линейный тренд, оценить качество полученной модели тренда. Если уравнение является статистически значимым, построить прогноз производства молока на следующие два месяца.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Месяц\год | 2001 | 2002 | 2003 | 2004 | 2005 |
| январь | 22,10 | 21,45 | 20,74 | 18,41 | 19,06 |
| февраль | 23,48 | 21,83 | 20,60 | 19,50 | 20,56 |
| март | 29,27 | 29,34 | 27,88 | 26,58 | 27,18 |
| апрель | 32,53 | 33,24 | 31,35 | 30,30 | 28,84 |
| май | 37,85 | 40,57 | 36,68 | 39,48 | 35,54 |
| июнь | 46,61 | 50,35 | 46,51 | 47,88 | 48,30 |
| июль | 46,66 | 49,88 | 46,96 | 45,95 | 47,49 |
| август | 39,52 | 43,33 | 39,83 | 41,51 | 42,37 |
| сентябрь | 29,00 | 32,13 | 30,45 | 33,46 | 32,61 |
| октябрь | 24,82 | 23,43 | 22,84 | 23,57 | 22,72 |
| ноябрь | 17,89 | 17,47 | 16,11 | 17,57 |  |
| декабрь | 18,87 | 18,28 | 16,36 | 18,46 |  |

Данные выбираются из приведенной таблицы в соответствии с номером варианта.

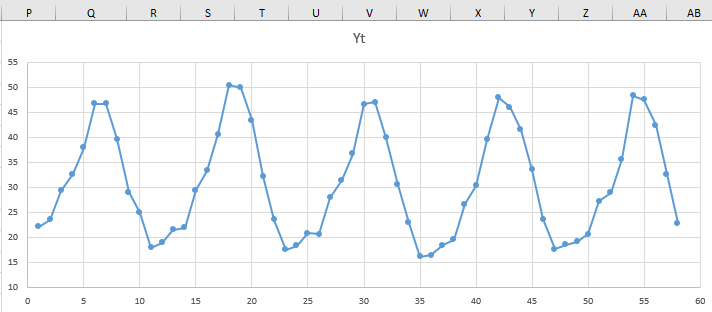
Вариант 1: с января 2001 года до октября 2005 года

**Решение:**

Аддитивная модель – модель, в которой временной ряд представлен как сумма его компонент.

Для начала переведем исходные данные в нужный нам вид. Для этого построим временной ряд длинною в 58 периодов (т. к. период с января 2001 по октябрь 2005 занимает 58 месяцев) и сопоставим их с производительностью молока за эти месяцы.

Начнем 1 этап. Для наглядности построим точечный график.



По графику видно, что цикл составляет 12 месяцев. Допустим это значение. Амплитуда колебания примерна одинакова, следовательно, можно построить аддитивную модель, где каждый уровень ряда представим в виде суммы компонент T, S, E.

Найдем Yt-1со сдвигом, так как при анализе временных рядов важное значение имеет автокорреляция – зависимость значений уровней временного ряда от предыдущих (сдвиг на 1, 2 и более) уровней того же ряда. Количественно автокорреляцию можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями ряда, сдвинутыми на несколько шагов вперед, т. е. между исходным рядом yt и этим же рядом, сдвинутым во времени на величину τ.

Тогда берем наш исходный ряд и копируем эти значения со сдвигом 12 раз (по кол-ву месяцев или по 1 циклу). Для того, чтобы найти коэффициенты автокорреляции (они показывают связь слабую, сильную) мы должны использовать пары значений, потому значения ниже последнего периода мы удаляем. Коэффициент автокорреляции = , где

За x рассматривается ряд ут, начиная со второго элемента, а в качестве переменной у = у + наш сдвиг. Или воспользуемся функцией ЭКСЕЛЬ [КОРРЕЛ

(где берем два массива: исходный ряд и последующие)].

Высокое значение коэффициента автокорреляции показывает, что ряд содержит линейную тенденцию (направленность процесса, при котором уровень изменения остается постоянным). Высокий коэффициент автокорреляции порядка τ свидетельствует о том, что ряд содержит циклические колебания с периодичностью τ.

Поставим порядок коэффициентов автокорреляции - 12. Далее по полученным рядам находим эти коэффициенты по функции категории статистически [КОРРЕЛ (первый массив = значения столбца B \*начиная со строки 3, т. к. нам необходимы только парные элементы, далее продолжаем этот ряд попутно меняя необходимые значения\*)]. После, по данному массиву строим коррелограмму (первая гистограмма), удаляем синий ряд, т. к. он означает значения 1, 2, 3 и т. д. График автокорреляционной функции называется коррелограммой и показывает, как изменение показателя yt отражается на его последующих значениях. Видим, что наибольший коэффициент корреляции принадлежит 12 порядку и равен 0,98. Близость найденного коэффициента к единице показывает очень тесную зависимость между объемами производства молока, текущего и непосредственно предшествующего годов.

Нужно проверить является ли данное значение коэффициента автокорреляции статистически значимым, тогда выдвинем гипотезу:

**Нулевая гипотеза:** Коэффициент корреляции 12-го порядка во всей генеральной совокупности равен 0.

Изображение выглядит как стол

Автоматически созданное описаниеИзображение выглядит как стол

Автоматически созданное описание

Проверяется это при помощи критерия Стьюдента (проверка гипотез, по ф-ле – наблюдаемое значение статистики Стьюдента для проверки гипотезы о статистической значимости значения коэффициента корреляции).

Наблюдаемым значением является наибольший коэффициент автокорреляции по модулю. Находим наблюдаемое значение по ф-ле: = \* = 34,511

Находим критическое значение статистики Стьюдента на уровне значимости 0,05 [СТЬЮДЕНТ.ОБР. 2Х] = 2,015,

Изображение выглядит как стол

Автоматически созданное описание

тогда делаем вывод, что t крит. <t набл., значит отвергаем нулевую гипотезу.

Далее проверим по P-уровню. [СТЬЮДЕНТ.РАСП.2Х] = 1,67762E-33, P-знач <0,05, следовательно, отвергаем нулевую гипотезу.

Значит, принимаем альтернативную гипотезу, что коэффициент автокорреляции 12-го порядка статистически значим, так как не принимает значение ноль при уровне значимости 0,05. А значит, мы можем распространить его на генеральную совокупность для того, чтобы найти длину цикла. Наш сезонный цикл имеет длину 12. А это значит, что сезонная компонента будет состоять из 12 компонентов. Мы сделали предварительно исследование модели. Амплитуда колебаний примерно постоянна. Тогда переходим к след. шагу.

**Построение аддитивной модели**.

Скопируем наши исходные данные на след. лист. Тут будем строить оценку сезонной компоненты (Колебания, которые имеют периодический характер и завершаются в течение одного года, называют сезонными). Для этого выровняем исходные данные методом скользящей средней (сглаживание временного ряда, основан на расчете и анализе так называемых скользящих (подвижных) средних. Скользящими средними называются средние арифметические значения показателя, исчисленные по новым m-членным укрупненным интервалам). Скользящая средняя – динамическая величина, которая последовательно рассчитывается путем передвижения на один интервал при заданной продолжительности периода. Рассчитывается с целью определения тенденции изменения случайной величины. Продолжительность периода = 12. Тогда, берем сначала 12 значений по предыдущему шагу. Находим скользящую среднюю по ф-ле [СРЕДНЕЕ] и за массив берем наш цикл = 12 месяцев. Продолжаем его до того, пока в массив не будут входить пустые ячейки.

Изображение выглядит как стол

Автоматически созданное описание

Далее, находим центрическую скользящую среднюю — это среднее значение двух смежных ячеек скользящей средней.

Продолжаем его также.

Изображение выглядит как текст, стена

Автоматически созданное описание

Затем, находим разности (это оценки сезонных компонент) = исходные данные – ЦСС. Для наглядности приведем полученные данные в таблицу.

Изображение выглядит как стол

Автоматически созданное описание

Видно, что значения за один месяц в разные года разнятся, но не сильно. Найдем среднее в каждый месяц.

Изображение выглядит как стол

Автоматически созданное описание

Сумма средних должна быть равна 0, но у нас она отрицательно и равна -0,17. Тогда эту разность мы разбросаем по этим 12 компонентам поровну. Для этого найдем среднее нашей суммы, затем в столбец сезонные компоненты запишем формулу [среднее за месяцы – среднее суммы]. Проверим, найдя сумму сез. комп., она равна 0. Все верно.

Сезонная компонента показывает влияние месяца на тренд. Тренд - изменение, определяющее общее направление развития, основную тенденцию временного ряда

**Строим тренд.**

Аддитивная модель Y = E+T+S, где

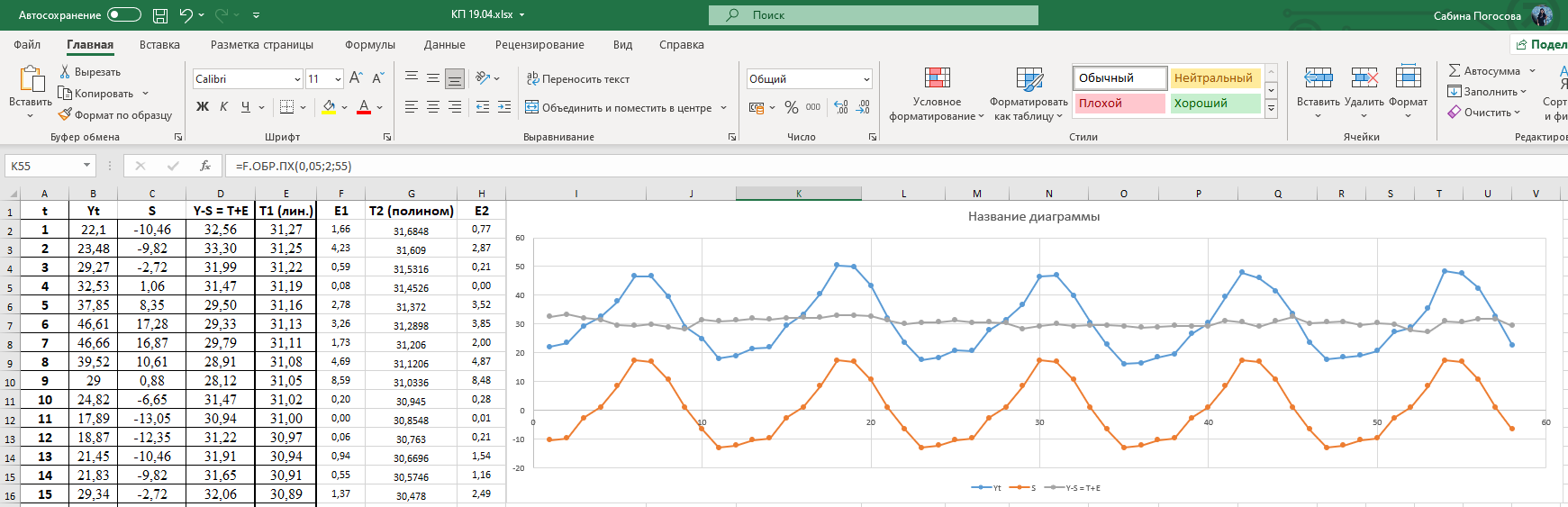
Y – фактическое значение

E - Ошибка

T – трендовое значение

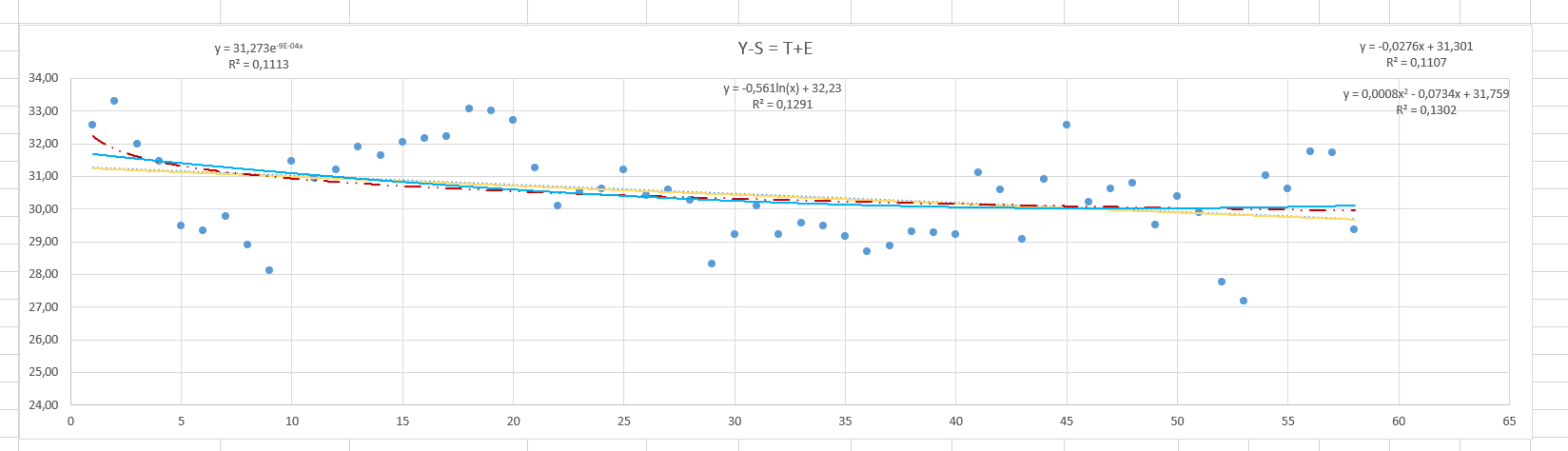
S – сезонная вариация

S мы уже нашли. Значит теперь эту компоненту из Y удалим, что (по ф-ле видно) равно T+E, таким образом выровняв данные в аддитивной модели. Теперь построим график.



Он получается очень приближенным. Зависимость от времени может принимать различные формы, поэтому для ее формализации используют различные виды функций.

Построим другой график с массивами t и Y-S=Y+E. На нем построим для наглядности линии тренда. Линейная – синяя, экспоненциальная – желтая, логарифмическая – красная, полином – голубой. Выберем тренд для построения – линейный. Критерием отбора наилучшей формы тренда является наибольшее значение коэффициента детерминации R2. (в данном случае для удобства вычислений выбираем линейную)



Проверяем является ли данное уравнение статистически значимым. Для этого используем анализ данных. Строим регрессию.

**Нулевая гипотеза** – уравнение является статистически незначимым, воздействие признака t на Y несущественно.

Значимость F равна 0,010715561. Данное значение меньше 0,05, следовательно, нулевую гипотезу о зависимости коэффициентов у и т отвергаем (уравнение является статистически значимым, воздействие признака t на Y существенно). Можем предположить, что объем производства молока зависит от месяца производства.

Для **построения прогноза** необходимо проверить, отличаются ли показатели уравнения от нуля.

**Нулевая гипотеза** – показатели a и b незначительно отличаются от нуля, т. е. являются статистически незначимыми.

Данную гипотезу можем проверить при помощи доверительных интервалов. По анализу данных доверительный интервал на уровне 0,05:

**a** = (-0,048618246; -0,00666906) - интервал НЕ включает в себя 0, значит параметр а может НЕ принимать нулевое значение. Тогда, нулевая гипотеза для параметра а не принимается

**b** = (30,58940864;32,01228392) – данный интервал не пробегает значение 0, и значительно от него отличается, следовательно для параметра b мы отвергаем нулевую гипотезу.

Проверим наш вывод через Р-значение каждого параметра, для этого сравним P-значение данных параметров из анализа данных с уровнем значимости 0,05.

a = 0,01071556 < 0,05 следовательно для а нулевая гипотеза отвергается

b = 9,1368E-62 < 0,05 следовательно для b нулевая гипотеза отвергается

Следовательно, дальнейшее построение прогноза по линейной модели возможно. Так как прогноз можно построить, если параметры статистически значимы.

Найдем трендовую компоненту (Т), подставив значение t в уравнение парной линейной регрессии t = -0,028x + 31,3

Найдем случайную компоненту (Е) = или как разность Y-S=E+T и T в квадрате, так как E является квадратом отклонения.

Строим прогноз. Прогноз на 2 месяца. Приписываем значение S за ноябрь и декабрь 2005г, которые равны соответственно -13,05 и -12,35.

Изображение выглядит как стол

Автоматически созданное описание

Далее найдем T для ноября и декабря 2005г., подставляя t в уравнение регрессии. Получим значения 29,67 и 29,64 соответственно. Складываем S и T для ноября и декабря, получим: 16,62 и 17,30.

Мы спрогнозировали объем производства молока за ноябрь и декабрь с помощью линейной модели.