

ЛАБОРАТОРНАЯ РАБОТА 8.

ВРЕМЕННЫЕ РЯДЫ.

ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ДАННЫХ

Цели работы:

Получить навыки по выявлению аномальных данных.

Ознакомиться с простыми приемами диагностики тренда временного ряда.

Изучить способы сглаживания временных рядов.

Общее задание на лабораторную работу:

Провести предварительный анализ экспериментальных данных.

Для выполнения данной работы необходимо выполнить 4 этапа:

1. Сгруппировать экспериментальные данные по 20-минутным интервалам.
2. Выявить и заменить аномальные данные в экспериментальных данных.
3. Определить наличие тренда у экспериментальных данных несколькими методами.
4. Сгладить полученный ряд различными методами сглаживания.

В качестве исходных экспериментальных данных используются данные из файла лабораторной работы № содержащие значения (в виде временного ряда) о количестве срабатывания регистратора.

Понятие временных рядов

Обычно для представления динамических процессов, протекающих в биологических, технических, экономических и др. системах, используют последовательность (ряд) значений некоторого показателя, упорядоченного по времени. Подобный ряд значений показателя отражает ход развития изучаемого процесса. Последовательность наблюдений одного показателя (признака), упорядоченная в зависимости от последовательно возрастающих или убывающих значений другого показателя, называется динамическим рядом, или рядом динамики. Если в качестве признака, в зависимости от которого происходит упорядочивание, берется время, то такой динамический ряд называется временным рядом.

Отдельные элементы рядов динамики, являющиеся значениями наблюдаемого показателя, называются уровнями ряда. Каждому уровню ряда соответствуют момент или интервал времени, к которым он относится. Временные ряды, в которых значения показателя относятся к определенным моментам времени, называются моментными (например, температура в печи при закалке стальных изделий, фиксируемая посекундно). Если уровни временного ряда образуются суммированием, усреднением или каким-либо другим

методом агрегирования за некоторый промежуток времени, то такие ряды называют интервальными временными рядами (например, ряд наблюдений за количеством обслуживаемых покупателей в магазине по часам, ряд значений усредненной температуры окружающего воздуха в течение суток). Под длиной временного ряда понимают время, прошедшее от начального момента наблюдений до конечного, или число уровней ряда.

Если во временном ряду проявляется длительная закономерность изменения уровней, то говорят, что имеет место тренд (тенденция) ряда. Тренд характеризует общее направление развития рассматриваемого процесса. Если математическая модель развития изучаемой системы описывается с учетом тренда ее основных показателей, то такая модель называется трендовой моделью.

Для выявления наличия тренда временных рядов, определения параметров тренда, а также для анализа трендовых моделей используется аппарат теории вероятностей и математической статистики. Однако следует помнить, что этот аппарат предназначен для обработки простых статистических совокупностей, и поэтому применение методов теории вероятности и математической статистики требует определенных поправок. Временные ряды отличаются от простых статистических совокупностей тем, что уровни временного ряда упорядочены во времени и их перестановка недопустима, тогда как элементы статистической совокупности не являются упорядоченными и перестановка этих элементов не изменяет значений оценок статистических показателей – среднего значения, дисперсии и т. д. Кроме того, уровни временного ряда зависят друг от друга, а элементы статистической совокупности являются независимыми.

Структурный состав временного ряда

При рассмотрении временного ряда показателя Y_t , состоящего из n уровней ($Y_1, Y_2, Y_3, \dots, Y_n$), в его составе, в общем случае, можно выделить три структурных элемента:

- тренд;
- колебания (циклические, сезонные и др.);
- остаточная компонента.

Основным структурным элементом временного ряда является тренд U_t , характеризующий наличие общего направления изменения наблюдаемого показателя в течение продолжительного времени. Кроме тренда во временных рядах могут наблюдаться близкие к повторяющимся колебания V_t относительно основной тенденции (тренда). Колебания с относительно небольшим периодом (например, день, месяц, год), обусловленные влиянием природно-климатических условий на рассматриваемый показатель, обычно называются сезонными колебаниями. Колебания с достаточно большим периодом (несколько лет) называются циклическими.

Сезонные колебания заметно проявляются в сельском хозяйстве, в до-

бывающих отраслях, а также в потреблении энергоносителей.

Под сезонностью понимают значимое различие в поведении рассматриваемой системы в течение регулярно повторяющихся временных периодов, обусловленных влиянием природных факторов или особенностей технологического процесса. Примером, иллюстрирующим сезонные колебания, являются графики расхода воды на полив в оросительной системе, в которых прослеживается влияние дневного и ночного времени. Большие циклические колебания обусловлены общими спадами и подъемами в рассматриваемом процессе.

Примером, иллюстрирующим циклические колебания, является временной ряд интенсивности солнечной активности, имеющий повторяющиеся циклы с периодом около 11 лет.

Тренд и колебания (различного периода) называются регулярными, или систематическими компонентами. Они характеризуют основное изменение показателя временного ряда «условно очищенное» от возможных случайных факторов. Если из временного ряда выделить и убрать систематические компоненты, то оставшаяся часть будет составлять *остаточную компоненту* e_t . Она является неотъемлемой (обязательной) частью любого временного ряда, так как такие процессы всегда сопровождаются небольшими изменениями, вызванными слабыми влияниями большого числа случайных (или неучтенных неслучайных) факторов. Если регулярная компонента временного ряда выделена правильно, что является целью при построении трендовой модели, то остаточная компонента будет обладать следующими свойствами:

- случайностью изменения своих значений;
- соответствием нормальному закону распределения;
- равенством нулю математического ожидания;
- независимостью значений уровней друг от друга, т. е. отсутствием существенной автокорреляции.

Проверка наличия этих четырех свойств у остаточной последовательности является основой проверки на адекватность трендовой модели.

В зависимости от вида связи между компонентами временного ряда может быть построена аддитивная модель, имеющая вид:

$$Y_t = U_t + V_t + e_t$$

или, если компоненты ряда умножаются, то получаем мультипликативную модель:

$$Y_t = U_t * V_t * e_t$$

Существует также и смешанная модель вида:

$$Y_t = U_t * V_t + e_t .$$

В большинстве случаев при анализе временных рядов наличием колебательной компоненты V_t пренебрегают, а если требуется ее учитывать, то для ее выделения применяют специальные методы, основанные на спектральном анализе.

Изучение соотношения между закономерностью и случайностью в формировании значений уровней ряда, оценка количественной меры их влияния есть **основная цель статистического анализа временных рядов**. Трендовая модель, построенная по значениям показателя в прошлом, используется для прогнозирования значений этого показателя в будущем, а учет случайности позволяет определить степень отклонения будущих значений от выявленного закономерного развития.

Этапы построения прогноза по временным рядам

Экстраполяционное прогнозирование процессов (прогнозирование «на будущее» при условии сохранения поведения «как в прошлом»), представленных одномерными временными рядами, сводится к выполнению следующих основных этапов:

- 1) предварительный анализ данных;
- 2) построение моделей: формирование набора аппроксимирующих функций (кривых роста) и численное оценивание параметров этих моделей;
- 3) проверка адекватности моделей и оценка их точности;
- 4) выбор лучшей модели.

В данной лабораторной работе рассматривается первый из этапов.

Задание

1. Подготовка экспериментальных данных. Проведите группировку временного ряда для интервала 20 минут (см. лабораторную работу 7).

Предварительный анализ данных временного ряда

К процедурам предварительного анализа относятся:

- выявление аномальных наблюдений;
- проверка наличия тренда;
- сглаживание временного ряда.

Выявление аномальных наблюдений

Аномальными уровнями считаются такие значения временного ряда, которые «сильно выбиваются» из общей тенденции. Их наличие существенно искажает основные статистические характеристики ряда, в том числе и соответствующую трендовую модель. Причинами аномальных значений часто являются ошибки технического характера (сбои при фиксации и передаче данных, ошибки при вычислениях).

Так как наличие аномальных наблюдений приводит к искажению результатов анализа данных, то необходимо убедиться в отсутствии аномалий. Для выявления аномальных значений ряда разработаны различные критерии, например, критерий Ирвина.

Согласно критерию Ирвина, аномальной считается точка Y_t , отстоящая от предыдущей точки Y_{t-1} на величину, большую среднеквадратичного отклонения. Для всех точек временного ряда рассчитывается критерий Ирвина:

$$\lambda_t = \frac{|Y_t - Y_{t-1}|}{\sigma},$$

где σ – оценка среднеквадратичного отклонения временного ряда,

$$\sigma = \sqrt{\frac{1}{n-1} \cdot \sum_{t=1}^n (Y_t - \bar{Y})^2};$$

$$\bar{Y} \text{ – среднее значение временного ряда, } \bar{Y} = \frac{1}{n} \cdot \sum_{t=1}^n Y_t$$

Точка ряда Y_t считается аномальной, если выполняется условие $\lambda_t > \lambda_{таб}$. Табличные значения $\lambda_{таб}$ уменьшаются с ростом длины ряда, их значения приведены в таблице для доверительных вероятностей p равных 0.95 и 0.99.

Критические значения параметра $\lambda_{таб}$

n		10	20	30	40	50	100	200	500	1000
$\lambda_{таб}$	$p = 0.95$	1.46	1.27	1.20	1.15	1.11	1.02	0.95	0.87	0.83
	$p = 0.99$	2.03	1.80	1.70	1.63	1.60	1.47	1.38	1.28	1.22

Задание

Выявление аномальных данных. После группировки проверьте полученный временной ряд на аномальные значения по критерию Ирвина. Обнаруженные аномальные значения замените путем интерполирования (усреднения) по соседним точкам

Рекомендации:

1. После однократной замены аномальных точек временной ряд все равно может содержать другие аномальные точки. Поэтому процедуру обнаружения и замены аномалий необходимо повторять циклически, до тех пор, пока аномальные точки не перестанут обнаруживаться.
2. Так как аномальные точки могут располагаться подряд (блоком),

а их замена осуществляется путем усреднения их соседей, то одновременно заменять аномальные точки во всем блоке нецелесообразно. За одну итерацию «проверки-замены» следует заменять только по одной точке в каждом блоке. Заменяемая точка выбирается по максимальному значению критерия Ирвина в пределах данного блока.

	H	I	J	K	L	M	N	O	P	Q	R	S
1	Группировка по 20 мин	Критерий Ирвина 1	Аномалии 1				Без аномалий 1	Критерий Ирвина 2	Аномалии 2			Без аномалий 2
2	6:00	5			кол точек	46	5					5
3	6:20	9	0,167295	0	крит. знач.	1,12	9	0,165885	0			9
4	6:40	9	0	0	СКО	23,91	9	0	0	СКО	24,11	9
5	7:00	34	1,045592	0	кол. аномалий	6	34	1,036784	0	кол. аномалий	4	34
6	7:20	43	0,376413	0			43	0,373242	0			43
7	7:40	37	0,250942	0			37	0,248828	0			37
8	8:00	38	0,041824	0			38	0,041471	0			38
9	8:20	40	0,083647	0			40	0,082943	0			40
10	8:40	32	0,33459	0			32	0,331771	0			32
11	9:00	44	0,501884	0			44	0,497656	0			44
12	9:20	61	0,711003	0			61	0,705013	0			61
13	9:40	47	0,585532	0			47	0,580599	0			47
14	10:00	49	0,083647	0			49	0,082943	0			49
15	10:20	54	0,209118	0			54	0,207357	0			54
16	10:40	63	0,376413	0			63	0,373242	0			63
17	11:00	51	0,501884	0			51	0,497656	0			51
18	11:20	69	0,752826	0			69	0,746484	0			69
19	11:40	114	1,882066	1			114	1,866211	1			75,5
20	12:00	61	2,216655	1			82	1,327083	1			82
21	12:20	50	0,460061	0			50	1,327083	1			50
22	12:40	43	0,292766	0			43	0,290299	0			43

Пример выявления и замены аномальных данных по критерию Ирвина

	H	I	J	K	L	M	N
1	Группировка по 20 мин	Критерий Ирвина 1	Аномалии 1				Без аномалий 1
2	6:00	5			кол точек	=СЧЁТ(I2:I47)	=I2
3	6:20	9	=ABS(I3-I2)/\$M\$4	=ЕСЛИ(J3>\$M\$3;1;0)	крит. знач.	1,12	=I3
4	6:40	9	=ABS(I4-I3)/\$M\$4	=ЕСЛИ(J4>\$M\$3;1;0)	СКО	=СТАНДОТКЛОН(I2:I47)	=I4
5	7:00	34	=ABS(I5-I4)/\$M\$4	=ЕСЛИ(J5>\$M\$3;1;0)	кол. аномалий	=СУММ(K3:K47)	=I5
6	7:20	43	=ABS(I6-I5)/\$M\$4	=ЕСЛИ(J6>\$M\$3;1;0)			=I6
7	7:40	37	=ABS(I7-I6)/\$M\$4	=ЕСЛИ(J7>\$M\$3;1;0)			=I7
8	8:00	38	=ABS(I8-I7)/\$M\$4	=ЕСЛИ(J8>\$M\$3;1;0)			=I8
9	8:20	40	=ABS(I9-I8)/\$M\$4	=ЕСЛИ(J9>\$M\$3;1;0)			=I9
10	8:40	32	=ABS(I10-I9)/\$M\$4	=ЕСЛИ(J10>\$M\$3;1;0)			=I10
11	9:00	44	=ABS(I11-I10)/\$M\$4	=ЕСЛИ(J11>\$M\$3;1;0)			=I11
12	9:20	61	=ABS(I12-I11)/\$M\$4	=ЕСЛИ(J12>\$M\$3;1;0)			=I12
13	9:40	47	=ABS(I13-I12)/\$M\$4	=ЕСЛИ(J13>\$M\$3;1;0)			=I13
14	10:00	49	=ABS(I14-I13)/\$M\$4	=ЕСЛИ(J14>\$M\$3;1;0)			=I14
15	10:20	54	=ABS(I15-I14)/\$M\$4	=ЕСЛИ(J15>\$M\$3;1;0)			=I15
16	10:40	63	=ABS(I16-I15)/\$M\$4	=ЕСЛИ(J16>\$M\$3;1;0)			=I16
17	11:00	51	=ABS(I17-I16)/\$M\$4	=ЕСЛИ(J17>\$M\$3;1;0)			=I17
18	11:20	69	=ABS(I18-I17)/\$M\$4	=ЕСЛИ(J18>\$M\$3;1;0)			=I18
19	11:40	114	=ABS(I19-I18)/\$M\$4	=ЕСЛИ(J19>\$M\$3;1;0)			=I19
20	12:00	61	=ABS(I20-I19)/\$M\$4	=ЕСЛИ(J20>\$M\$3;1;0)			=I19+I21)/2
21	12:20	50	=ABS(I21-I20)/\$M\$4	=ЕСЛИ(J21>\$M\$3;1;0)			=I21
22	12:40	43	=ABS(I22-I21)/\$M\$4	=ЕСЛИ(J22>\$M\$3;1;0)			=I22

Пример выявления и замены аномальных данных по критерию Ирвина (ре-

жим отображения формул)

Проверка наличия тренда

Для определения наличия тренда временного ряда используются различные методы. В данной лабораторной работе рассмотрен метод проверки разностей средних уровней [и метод Фостера-Стьюарта.

Метод проверки разностей средних уровней

Согласно этому методу временной ряд разбивается на две примерно равные по числу уровней части n_1 и n_2 ($n = n_1 + n_2$), каждая из которых рассматривается как некоторая самостоятельная выборочная совокупность, имеющая нормальное распределение.

Если временной ряд имеет тенденцию к тренду, то средние, вычисленные для каждой совокупности, должны существенно (значимо) различаться между собой. Если же расхождение несущественно (случайно), то временной ряд не имеет тенденции к тренду. Таким образом, проверка наличия тренда в исследуемом ряду сводится к проверке гипотезы о равенстве средних двух нормально распределенных совокупностей, проверяемой с помощью t -критерия Стьюдента. Заметим, что метод применим только для случая, когда обе части выборок имеют одинаковую дисперсию, поэтому рассматриваемый метод диагностики наличия тренда нужно предварить проверкой гипотезы о равенстве дисперсий, для чего используется критерий Фишера.

Для каждой из частей ряда вычисляются среднее значение и оценка дисперсии:

$$\bar{Y}_1 = \frac{\sum_{t=1}^{n_1} Y_t}{n_1}; \quad S_1^2 = \frac{\sum_{t=1}^{n_1} (Y_t - \bar{Y}_1)^2}{n_1 - 1};$$

$$\bar{Y}_2 = \frac{\sum_{t=1}^{n_2} Y_t}{n_2}; \quad S_2^2 = \frac{\sum_{t=1}^{n_2} (Y_t - \bar{Y}_2)^2}{n_2 - 1};$$

Далее, рассчитывается F -критерий Фишера:

Производится проверка равенства (однородности) дисперсий обеих частей ряда с помощью F-критерия Фишера, которая основана на сравнении расчетного значения этого критерия:

$$F = \begin{cases} S_1^2 / S_2^2, & \text{если } S_1^2 > S_2^2 \\ S_2^2 / S_1^2, & \text{если } S_2^2 > S_1^2 \end{cases}$$

с табличным (критическим) значением критерия Фишера $F_{\text{таб}}$ с заданным уровнем значимости (уровнем ошибки). В качестве уровня значимости (q) чаще всего берут значения 0,1 (10 %-я ошибка), 0,05 (5 %-я ошибка), 0,01 (1 %-я ошибка). Величина $P = 1 - q$ называется доверительной вероятностью.

Если полученное значение F меньше табличного значения $F_{\text{таб}}$ (со степенями свободы $n_i - 1$ и $n_k - 1$, где i – индекс части ряда с большей дисперсией, k – индекс части с меньшей дисперсией), то гипотеза об однородности дисперсий не отвергается (принимается) и переходят к следующему этапу расчета. Если $F \geq F_{\text{таб}}$, то гипотеза об однородности дисперсий отклоняется и метод не дает ответа на вопрос о наличии или отсутствии тренда.

Окончательная проверка гипотезы об отсутствии тренда (точнее, о равенстве средних значений двух выборок) производится с использованием t-критерия Стьюдента, вычисляемого по формуле:

$$t = \frac{|\bar{Y}_1 - \bar{Y}_2|}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

где σ – среднеквадратичное отклонение разности средних:

$$\sigma = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Если расчетное значение t меньше табличного значения $t_{\text{таб}}$, с заданным уровнем значимости, то гипотеза о равенстве средних значений не отвергается (принимается), т. е. тренда нет, в противном случае – тренд есть. Для определения табличного значения $t_{\text{таб}}$ число степеней свободы принимается равным $n_1 + n_2 - 2$.

Задания

Определение наличия тренда.

1. После удаления аномальных данных проверьте полученный временной ряд на наличие тренда методом проверки разностей средних. Сделайте вывод о наличии или отсутствии тренда временного ряда.

Рекомендации:

1. Для расчета $F_{\text{таб}}$ критерия Фишера можно использовать встроенную Excel-функцию
 $\text{=FРАСПОБР(вероятность; степени_свободы1; степени_свободы2)}$.

2. Для расчета $t_{\text{таб}}$ критерия Стьюдента можно использовать встроенную Excel-функцию
 $\text{=СТЮДРАСПОБР(вероятность; степени_свободы)}$.

	А	В	С	Д	Е	Ф	Г	Н	І	Ј
1	Время	Исходные данные без аномалий		Метод проверки разностей средних						
2	6:00	5	Объем 1		23		Объем 2		23	
3	6:20	9	от		6:00		от		13:40	
4	6:40	9	до		13:20		до		21:00	
5	7:00	26	Среднее1		45,2173913		Среднее2		26,021739	
6	7:20	43	Дисперсия1		403,0187747		Дисперсия2		415,26087	
7	7:40	37	Гипотеза об однородности дисперсии - критерий Фишера							
8	8:00	38	Критерий		1,030375991		=H6/E6			
9	8:20	40	F табл		2,047770309		$\text{=FРАСПОБР(0,05;H2-1;E2-1)}$			
10	8:40	32	Вывод		дисперсии одинаковы, проверять дальше					
11	9:00	44								
12	9:20	61	Гипотеза об отсутствии тренда - t-критерий Стьюдента							
13	9:40	47	СКО разности средних		20,227205		$\text{=КОРЕНЬ(((E2-1)*E6+(H2-1)*H6)/(E2+H2-2))}$			
14	10:00	49	t-критерий		3,218221383		$\text{=ABS(E5-H5)/(F13*КОРЕНЬ(1/E2+1/H2))}$			
15	10:20	54	t табличный		2,015367547		$\text{=СТЮДРАСПОБР(0,05;E2+H2-2)}$			
16	10:40	63	Вывод		тренд есть					
17	11:00	51								

Проверка наличия тренда методом разностей средних уровней

Задание

2. Проверьте еще раз временной ряд на наличие тренда методом проверки разностей средних с использованием инструментов из пакета «Анализ данных» – «Двухвыборочный F-тест для дисперсии» и «Двухвыборочный t-тест с одинаковыми дисперсиями» (см. рисунки).

Замечание. Режим работы инструмента «Двухвыборочный F-тест для дисперсий» служит для проверки гипотезы H_0 о равенстве дисперсий двух нормальных распределений ($D_1 = D_2$). При этом, в качестве альтернативной рассматривается гипотеза $H_1: D_1 < D_2$, если $S_1 < S_2$; или гипотеза $H_1: D_1 > D_2$, если $S_1 > S_2$. Из-за этого условие принятия гипотезы H_0 отличается:

– если дисперсия первой части больше дисперсии второй части ($S_1 > S_2$), то гипотеза H_0 не отвергается при условии $F < F_{\text{кр}}$;

– если дисперсия первой части меньше дисперсии второй части ($S_1 < S_2$), то гипотеза H_0 не отвергается при условии $F > F_{\text{кр}}$.

На рис. 2.4 дисперсия первой части (403,01) меньше дисперсии

второй части (420,33), поэтому условием не отвержения гипотезы H_0 является $F > F_{кр}$ ($0,9588 > 0,4883$).

	A	B	C	D	E	F	G	H	I	J
5	7:00	26		Среднее1	45,2173913		Среднее2	26,021739		
6	7:20	43		Дисперсия1	403,0187747		Дисперсия2	415,26087		
7	7:40	37		Гипотеза об однородности дисперсии - критерий Фишера						
8	8:00	38		Критерий	1,030375991	=H6/E6				
9	8:20	40		F табл	2,047770309	=FРАСПОБР(0,05;H2-1;E2-1)				
10	8:40	32		Вывод	дисперсии одинаковы, проверять дальше					
11	9:00	44								
12	9:20	61		Гипотеза об отсутствии тренда - t-критерий Стьюдента						
13	9:40	47		СКО разности средних		20,227205	=КОРЕНЬ(((E2-1)*E6+(H2-1)*H6)/(E2+H2-2))			
14	10:00	49		t-критерий		3,218221383	=ABS(E6-H6)/(F13*КОРЕНЬ(1/E2+1/H2))			
15	10:20	54		t табличный		2,015367547	=СТЫОДРАСПОБР(0,05;E2+H2-2)			
16	10:40	63		Вывод	тренд есть					
17	11:00	51								
18	11:20	69		Проверка посредством "Пакета анализа"						
19	11:40	75,5		Двухвыборочный F-тест для дисперсии						
20	12:00	82								
21	12:20	62,5		Переменная 1 Переменная 2						
22	12:40	43		Среднее	45,2173913	27,91304348				
23	13:00	55		Дисперсия	403,0187747	420,333004				
24	13:20	45		Наблюдения	23	23				
25	13:40	44		df	22	22				
26	14:00	58,5		F	0,958808304					
27	14:20	73		P(F<=f) однос	0,461152257					
28	14:40	60,5		F критическое	0,488336019		Вывод	проверять дальше		
29	15:00	48								
30	15:20	37								
31	15:40	27								
32	16:00	30								
33	16:20	25								
34	16:40	37								
35	17:00	24,5								
36	17:20	12								
37	17:40	17								
38	18:00	27								
39	18:20	25								
40	18:40	9								
41	19:00	16								
42	19:20	17								
43	19:40	4								

Двухвыборочный F-тест для дисперсии

Входные данные

Интервал переменной 1:

Интервал переменной 2:

☐ Метки

Альфа:

Параметры вывода

☒ Выходной интервал:

☐ Новый рабочий лист:

☐ Новая рабочая книга

OK

Отмена

Справка

Задание параметров для инструмента «Двухвыборочный F-тест для дисперсии»

Метод Фостера-Стьюарта

Этот метод дает более надежные результаты по сравнению с предыдущим. Метод позволяет установить наличие, как самого тренда, так и наличие тренда дисперсии: при отсутствии тренда дисперсии разброс уровней ряда постоянен, при наличии тренда дисперсии дисперсия увеличивается или уменьшается. Согласно методу, выполняется сравнение каждого уровня ряда с предыдущим и формируются две последовательности:

	C	D	E	F	G	H	I	J	K	L
12		Гипотеза об отсутствии тренда - t-критерий Стьюдента								
13		СКО разности средних		20,28979767	=КОРЕНЬ(((E2-1)*E6+(H2-1)*H6)/(E2+H2-2))					
14		t-критерий		2,892187469	=ABS(E5-H5)/(F13*КОРЕНЬ(1/E2+1/H2))					
15		t табличный		2,015367547	=СТЮДРАСПОБР(0,05;E2+H2-2)					
16		Вывод	тренд есть							
17										
18		Проверка посредством "Пакета анализа"								
19		Двухвыборочный F-тест для дисперсии								
20										
21		Переменная 1		Переменная 2						
22		Среднее	45,2173913	27,91304348						
23		Дисперсия	403,0187747	420,333004						
24		Наблюдения	23	23						
25		df	22	22						
26		F	0,958808304							
27		P(F<=f) однос	0,461152257							
28		F критическое	0,488336019							
29										
30										
31		Двухвыборочный t-тест с одинаковыми дисперсиями								
32										
33		Переменная 1		Переменная 2						
34		Среднее	45,2173913	27,91304348						
35		Дисперсия	403,0187747	420,333004						
36		Наблюдения	23	23						
37		Объединенная	411,6758893							
38		Гипотетическая	0							
39		df	44							
40		t-статистика	2,892187469							
41		P(T<=t) однос	0,002962989							
42		t критическое	1,680229977							
43		P(T<=t) двухс	0,005925978							
44		t критическое	2,015367547							
45		Вывод	тренд есть							

Двухвыборочный t-тест с одинаковыми дисперсиями

Входные данные

Интервал переменной 1: \$8\$2:\$8\$24

Интервал переменной 2: \$8\$25:\$8\$47

Гипотетическая средняя разность:

☐ Метки

Альфа: 0,05

Параметры вывода

☒ Выходной интервал: \$D\$31

☐ Новый рабочий лист:

☐ Новая рабочая книга

OK

Отмена

Справка

Двухвыборочный t-тест с одинаковыми дисперсиями

Входные данные

Интервал переменной 1: \$8\$2:\$8\$24

Интервал переменной 2: \$8\$25:\$8\$47

Гипотетическая средняя разность:

☐ Метки

Альфа: 0,05

Параметры вывода

☒ Выходной интервал: \$D\$31

☐ Новый рабочий лист:

☐ Новая рабочая книга

OK Отмена Справка

Задание параметров для инструмента «Двухвыборочный t-тест с одинаковыми дисперсиями»

$$k_t = \begin{cases} 1, & \text{если } Y_t \text{ больше всех предыдущих уровней} \\ 0, & \text{в противном случае;} \end{cases}$$

$$l_t = \begin{cases} 1, & \text{если } Y_t \text{ меньше всех предыдущих уровней} \\ 0, & \text{в противном случае;} \end{cases}$$

$$t = 2, 3, \dots, n.$$

Вычисляются величины s и d , характеризующие изменения среднего значения и дисперсии временного ряда:

$$s = \sum_{t=2}^n (k_t + l_t), \quad d = \sum_{t=2}^n (k_t - l_t).$$

Величина s характеризует изменение дисперсии временного ряда, она может принимать значение от 0 (когда все уровни ряда равны) до $n - 1$ (ряд

монотонно изменяется). Величина d характеризует изменение среднего значения временного ряда и изменяется от $-(n-1)$ (когда ряд монотонно убывает) до $+(n-1)$ (когда ряд монотонно возрастает). Величины s и d являются случайными с математическим ожиданием равным μ для значения s и равным 0 для значения d .

Далее проверяются гипотезы о случайности отклонения величин s и d от их математических ожиданий с помощью критерия Стьюдента:

$$t_s = \frac{|s - \mu|}{\sigma_1}, \quad \sigma_1 = \sqrt{2 \cdot \ln(n) - 3.4253},$$

$$t_d = \frac{|d - 0|}{\sigma_2}, \quad \sigma_2 = \sqrt{2 \cdot \ln(n) - 0.8456}, \quad \mu = \sigma_2^2,$$

где μ – оценка математического ожидания величины s для случайного временного ряда;

σ_1 – оценка среднеквадратического отклонения s для случайного временного ряда;

σ_2 – оценка среднеквадратического отклонения d для случайного временного ряда.

Формулы для σ_1 и σ_2 применимы при $n \geq 50$. Значения σ_1 и σ_2 при $n < 50$ приведены в таблице.

Значения σ_1 и σ_2 для критерия Фостера-Стьюарта [5]

n	10	15	20	25	30	35	40	45	50
σ_1	1,288	1,512	1,677	1,791	1,882	1,956	2,019	2,072	2,121
σ_2	1,964	2,153	2,279	2,373	2,447	2,509	2,561	2,606	2,645

Полученные значения t_s и t_d необходимо сравнить с табличными значениями критерия Стьюдента $t_{\text{таб}}$ с n степенями свободы.

Если $t_{\text{таб}}$ больше расчетного значения, то соответствующий тренд отсутствует: например, если $t_s > t_{\text{таб}}$, а $t_d < t_{\text{таб}}$, то тренд дисперсии есть, а тренда ряда нет.

Задание

3. Проверьте временной ряд без аномалий на наличие тренда методом Фостера-Стьюарта. Сделайте вывод о наличии или отсутствии тренда временного ряда

	L	M	N	O	P	Q	R	S	T	U
1	Время	Исходные данные без аномалий	Метод Фостера-Стьюарта							
2	6:00	5	k	I	k-I			s (дисперсия)	d (среднее)	
3	6:20	9	1	0	1			12	6	
4	6:40	9	0	0	0		СКО	2,057178357	2,609920074	
5	7:00	26	1	0	1		мат ожидание	6,811682793		
6	7:20	43	1	0	1					
7	7:40	37	0	0	0		Гипотеза о случайности отклонений s и d			
8	8:00	38	0	0	0		от их мат. ожиданий			
9	8:20	40	0	0	0		t-критерий	2,522055119	2,298920974	
10	8:40	32	0	0	0		t табл	2,014103359	2,014103359	
11	9:00	44	1	0	1		Вывод	есть тренд	есть тренд	
12	9:20	61	1	0	1					
13	9:40	47	0	0	0					
14	10:00	49	0	0	0					
15	10:20	54	0	0	0					
16	10:40	63	1	0	1					

Пример проверки наличия тренда методом Фостера-Стьюарта

Сглаживание временного ряда

Довольно часто уровни временного ряда колеблются, так что тенденция развития процесса (регулярная составляющая) скрыта случайными отклонениями. Сглаживание временного ряда позволяет отфильтровать мелкие случайные колебания и выявить основную тенденцию изменения исследуемой величины. При механическом сглаживании выравнивание отдельных уровней производится с использованием значений соседних уровней. Для сглаживания используются следующие методы:

Простая (среднеарифметическая) скользящая средняя:

$$\tilde{Y}_t = \frac{\sum_{i=t-p}^{t+p} Y_i}{2 \cdot p + 1}, \quad p < t < n - p.$$

Сглаженное значение Y_t является среднеарифметическим из $2 \cdot p + 1$ соседних точек. Наиболее часто используется сглаживание по 5 точкам:

$$\tilde{Y}_t = \frac{Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2}}{5}.$$

Взвешенная (средневзвешенная) скользящая средняя:

$$\tilde{Y}_t = \frac{\sum_{i=t-p}^{t+p} \alpha_i \cdot Y_i}{\sum_{i=t-p}^{t+p} \alpha_i}, \quad p < t < n - p.$$

В этом методе каждая из точек входит в общую сумму с весовым коэффициентом α_i . Для сглаживания по 5 точкам используют весовые коэффициенты $(-3, 12, 17, 12, -3)$. Для сглаживания по 7 точкам используются коэффициенты $(-2, 3, 6, 7, 6, 3, -2)$ или $(5, -30, 75, 131, 75, -30, 5)$.

Задания

4. Сглаживание временного ряда.

4.1. Выполните сглаживание временного ряда, полученного после удаления аномальных данных методом «среднеарифметическая по 5 точкам»

На одной диаграмме постройте графики исходного ряда и сглаженные ряды.

4.2. Выполните экспоненциальное сглаживание временного ряда с использованием инструмента «Экспоненциальное сглаживание» из пакета «Анализ данных» (см. рисунок).

Замечание. В случае если в окне «Экспоненциальное сглаживание» не задан «фактор затухания» (рисунок), то он, по умолчанию, принимается равным 0.3.

		Сглаживание						
Время	Исходные данные без аномалий	Средне-арифметическое по 5 точкам	Средне-взвешенное по 5 точкам	Средне-взвешенное по 7 точкам	Средне-хронологическое по 12 точкам	Экспоненциальное	Альфа для эксп. сглаживания	Экспоненциальное из "Пакета анализа"
6:00	5					5	0,3	
6:20	9					7,8		
6:40	9	18,4	12,257143			8,64		
7:00	26	24,8	26,514286	26		20,792		
7:20	43	30,6	38,457143	34,38095		36,3376		
7:40	37	36,8	40,085714	41		36,80128		
8:00	38	38	38,428571	38,71429	34,41667	37,640384		
8:20	40	38,2	36,485714	35	38,125	39,292115		
8:40	32	43	35,857143	40,80952	42,25	34,187635		
9:00	44	44,8	45,8	45,38095	45,54167	44,05639		
9:20	61							
9:40	47							
10:00	49							
10:20	54							
10:40	63							
11:00	51							
11:20	69							
11:40	75,5							
12:00	82							
12:20	62,5							
12:40	43							
13:00	55							
13:20	45							
13:40	44							

Экспоненциальное сглаживание

Входные данные

Входной интервал:

\$G\$79:\$G\$124

Фактор затухания:

☐ Метки

Параметры вывода

Выходной интервал:

\$N\$79

Новый рабочий лист:

Новая рабочая книга

☐ Вывод графика☐ Стандартные погрешности

OK

Отмена

Справка

Экспоненциальное сглаживание

Входные данные

Входной интервал:

Фактор затухания:

☐ Метки

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

☐ Вывод графика

☐ Стандартные погрешности

OK

Отмена

Справка

Пример задания параметров для инструмента «Экспоненциальное сглаживание»

Контрольные вопросы

1. Поясните, в чем состоят характерные отличия временных рядов от статистических выборок?
2. Что такое тренд, трендовая модель?
3. В чем особенность моментных временных рядов, интервальных временных рядов?
4. Какова общая структура временного ряда?
5. На чем основан критерий Ирвина по определению аномальных значений ряда?
6. Для каких целей может быть использован метод Фостера-Стюарта?
7. Когда метод сравнения разностей средних уровней не дает ответа на вопрос о наличии тренда?
8. Какой метод позволяет определить тренд дисперсии?
9. Влияние каких компонент временного ряда устраняется с помощью методов сглаживания?