

## Упражнение 4

### Приложение на статистическия модул в Excel



#### Цели:

- Да се представят възможностите за графично представяне на данните от минали периоди за откриване на тенденции;
- Да се изследват основните статистически функции в Excel;
- Да се разгледат възможностите за съставяне на прогнози за бъдещи периоди.



**Ключови думи:** графично представяне на данните, откриване на тенденции, времева сесия, статистическа обработка на информация, прогнози.



**Ресурси:** За реализация на настоящото упражнение е нужно наличие на РС с инсталиран програмен продукт Excel и надеждна и сравнително бърза връзка с Интернет.

#### Графично представяне на данните с цел откриване на тенденция.

Изследването на бизнес дейностите често изисква анализ на данни, които са събирани за определен минал период от време. Всяка такава поредица от данни се нарича *времева серия*. С помощта на Excel могат да бъдат извършени редица пресмятания, които да помогнат за анализа и прогнозирането на бъдещите стойности на времевите серии.

Анализът на всяка поредица от данни трябва да започне с графичното представяне на информацията. Има 2 основни графика, които всеки един статистически анализ трябва да съдържа:

- 1) Промяна на данните с течение на времето – помага за откриване на тенденция;
- 2) Съотношение на данните от текущия и предходния период – показва в каква степен се променят данните с изтичане на всеки един период.



#### Задача 1

- 1) Да се представят графично данните от продажбите в голям магазин за последните 5 години.
- 2) Да се изготви графика за

промените в продажбите спрямо всеки предходен период. 3) Да се направят изводи за наблюдаваната тенденция.

**Указания за изпълнение:**

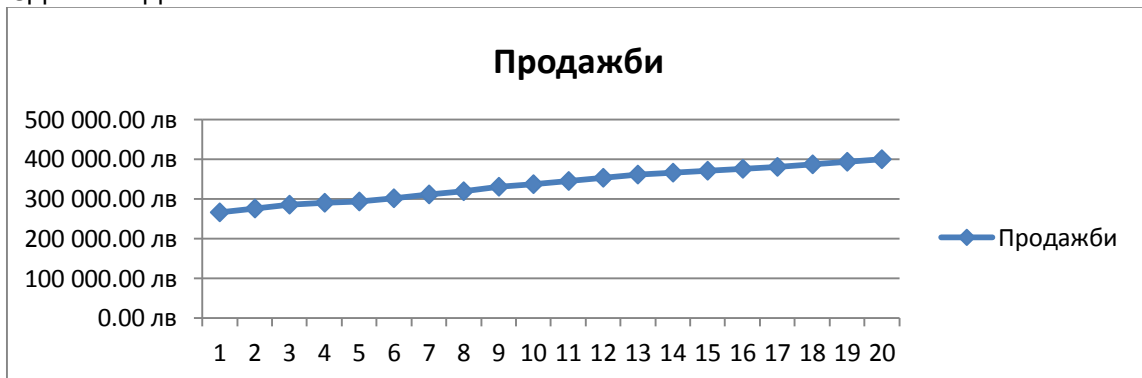
Данните за продажбите се намират във файла U4-analysis.xls (фиг. 1) В колона **Е** пресметнете промяната на продажбите спрямо всеки от предходните период, въвеждайки в клетка **Е3** формулата **=D3/D2\*100** и разпространявайки я до клетка **Е21**.

	A	B	C	D	E	F
	Година	Тримесечие	Пореден период	Продажби	Промяна текущ / предходен период	
1						
2	2001	I	1	266 200.00 лв		
3		II	2	276 000.00 лв	103.68	
4		III	3	284 200.00 лв	102.97	
5		IV	4	290 600.00 лв	102.25	
6	2002	I	5	293 400.00 лв	100.96	
7		II	6	301 600.00 лв	102.79	
8		III	7	310 400.00 лв	102.92	
9		IV	8	318 300.00 лв	102.55	
10	2003	I	9	329 600.00 лв	103.55	
11		II	10	337 200.00 лв	102.31	
12		III	11	345 200.00 лв	102.37	
13		IV	12	352 800.00 лв	102.20	
14	2004	I	13	361 100.00 лв	102.35	
15		II	14	365 000.00 лв	101.08	
16		III	15	370 100.00 лв	101.40	
17		IV	16	375 700.00 лв	101.51	
18	2005	I	17	380 400.00 лв	101.25	
19		II	18	386 600.00 лв	101.63	
20		III	19	392 700.00 лв	101.58	
21		IV	20	398 900.00 лв	101.58	
22						
23						
24						

Фигура 1 Получени данни в файла E4-analysis.xls.

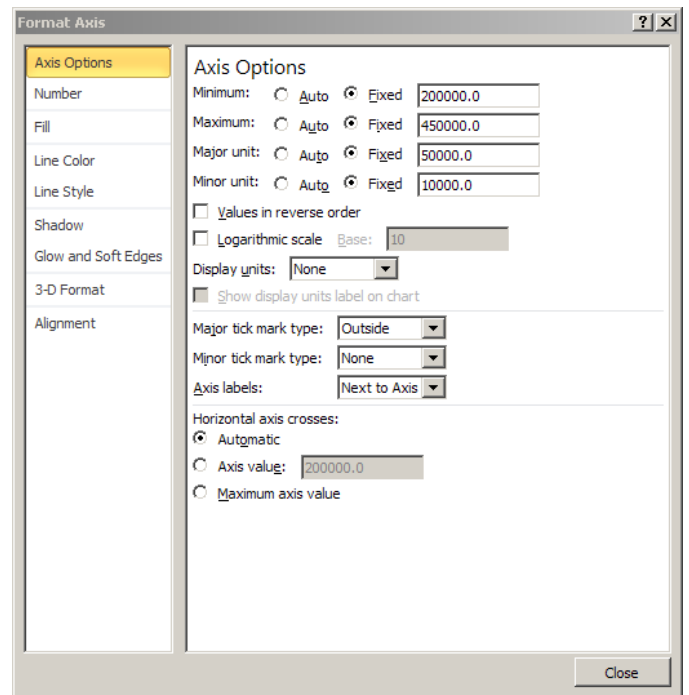
За да направите първата графика използвайте данните в колона **D** и графика от вида **Line (Line with markers displayed at each data value)**.

При правилно изпълнение на указанията графиката трябва да има следния вид:

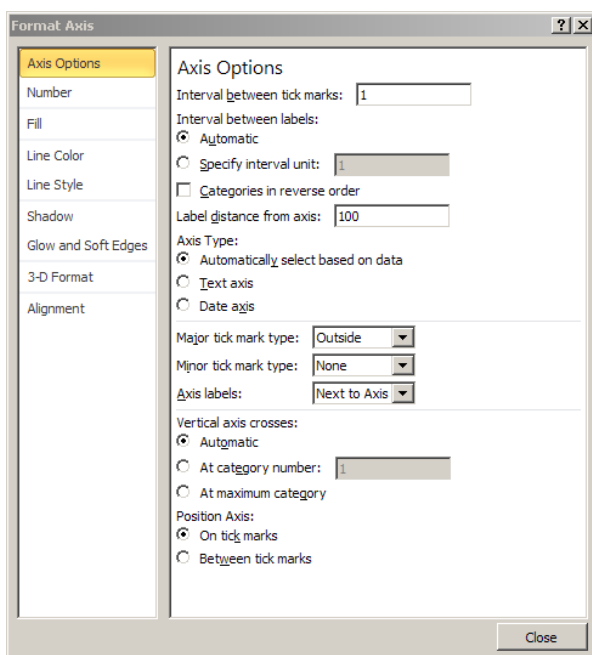


Фигура 2 Графика на колона D от файла E4-analysis.xls

Ако долните или горните стойности на графиката са твърде ниски или високи, щракнете два пъти върху числата (по ос у), представящи стойността на продажбите, задайте подходящи минимални и максимални стойности (фиг.3). При първоначалната графика, период 1 не започва в началото на оста. За да преместите периодите, така че точно да отговарят на съответните стойности, щракнете два пъти върху числата на Position Axis, представящи периодите и махнете отметката от опцията *between tick marks* както е показано на Фигура 4.



Фигура 3 Настройки на графиката по продажби (ос у).



Фигура 4 Настройки на графиката по периоди.

Направете подобна графика и за промяната в продажбите спрямо предходните периоди, използвайки данните в колона Е. Вижда се тенденцията към нарастване на продажбите, но в последните години темповете на растеж спадат (фиг.5).



Фигура 5 Графика на колона Е от файла E4-analysis.xls  
Запазете файла!

## Основни статистически функции в Excel. Описателни статистики.



### Задача 2 Средни, минимални и максимални стойности, стандартно отклонение

Използвайте основните статистически функции в Excel, за да опишете състоянието на данните за броя на обажданията в кол центъра на ОББ за последните 90 дни. Изгответе отчет съдържаш:

- 1) средния, максималния и минималния брой на обажданията;
- 2) средния брой на обажданията, изключвайки 5% от най-ниските и най-високите стойности, за да се елиминира вероятността от погрешно въведени данни;
- 3) степента, в която броя на обажданията се колебае около средния брой.

**Указания за изпълнение:**

Отворете секцията *calls* във файла *У4-analysis.xls*. За да намерите минималния и максималния брой на обажданията в разглеждания период използвайте функциите **MIN** и **MAX**. За да намерите средната стойност, използвайте функцията **AVERAGE**.

=MIN(*начална\_клетка:крайна\_клетка*)  
 =MAX(*начална\_клетка:крайна\_клетка*)  
 =AVERAGE(*начална\_клетка:крайна\_клетка*)



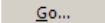
Понякога при въвеждането на данни се получават грешки, които могат да доведат до изкривена информация за изчислените по-горе показатели. В конкретния случай в клетки B19 и B50 са въведени стойности, които се различават значително от останалите. За да се преодолее този недостатък, може да се използва функцията **TRIMMEAN**, която изчиства зададен процент от минималните и максималните стойности.

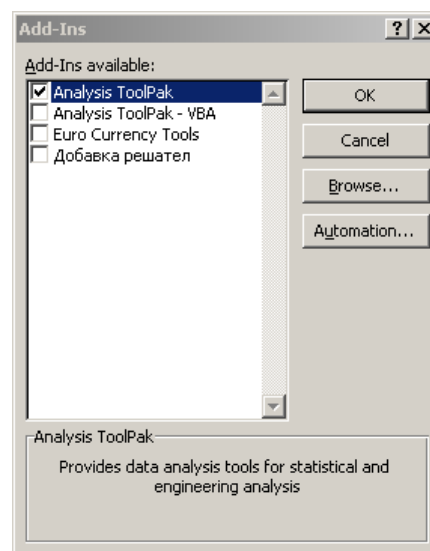
=TRIMMEAN(*начална\_клетка:крайна\_клетка;процент като десетична дроб*)

За да откриете степента, в която броя на обажданията се колебае около средната стойност трябва да намерите стандартното отклонение. Можете да използвате функцията **STDEV** ако имате извадка от данните или **STDEVP** ако разполагате с всички данни за определен период. В конкретния случай ще използвате втората функция, тъй като разполагате с всички данни за 90-дневния период.

=STDEVP(*начална\_клетка:крайна\_клетка*)

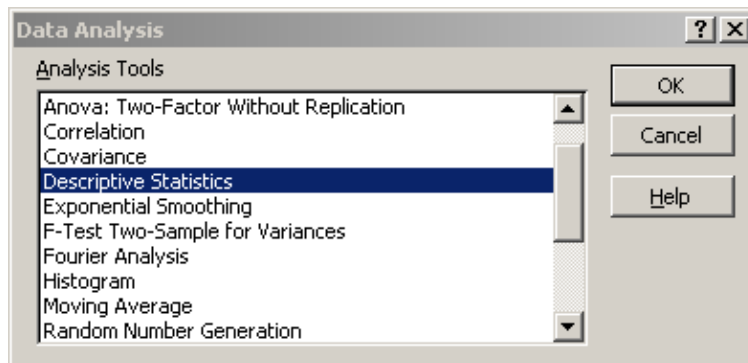
Всички тези резултати могат да бъдат получени и чрез използване на опцията **Descriptive Statistics** (Описателни статистики) от пакета Data Analysis

от раздел  **Developer**,  **Add-Ins**. Ако опцията **Add-Ins** не е активна може да я активирате по следния начин: от менюто **File** изберете **Options** след което посочете **Add-Ins**, от появилия се диалогов прозорец натиснете бутона  **Go...** В ново появилия се прозорец поставете отметка пред **Analysis ToolPak** (фиг. 6).



Фигура 6 Диалогов прозорец Add-Ins.

При изпълнение на задачата е нужно да се активира **Descriptive Statistics** от **Data Analysis** (фиг.7).



Фигура 7 Диалогов прозорец Data Analysis.

В полето **Input Range** (фиг. 8) въведете клетките, в които се намират изследваните данни. Ако в първият ред има наименование на колоната, сложете отметка на опцията **Labels in first row**. Задайте изчисленията да се показват в нов лист поставяйки знак

пред **New Worksheet Ply**. Маркирайте и опцията **Summary Statistics**, за да получите исканата информация. Като резултат ще се изведе следната информация:  
**Mean** – средна стойност. Същият резултат се получава и с функцията AVERAGE.

**Median** – показва стойността, която е в центъра на масива от данни.

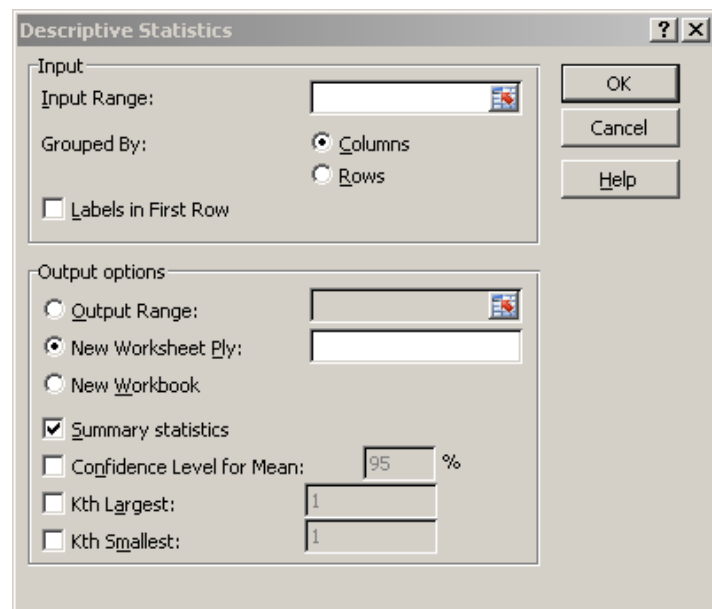
**Mode** – показва най-често срещаната стойност в масива от данни.

**Standard Deviation** – стандартно отклонение. Резултатът може да бъде получен и чрез функцията STDEV или STDEVP.

**Minimum; Maximum** – минималната и максималната стойност.

Алтернатива е използването на **MIN** и **MAX**.

**Sum; Count** – сума и брой на наблюденията.



Фигура 8 Прозорец на Descriptive Statistics.



### Задача 3 – самостоятелна работа

Използвайте данните за сумата на покупките от файла [U1-report.xls](#), използван в Упражнение 1, за да покажете описателните статистики за продукт номер 735. Флоат 4 мм. 3210/2250.

**Указания:** Сортирайте данните по номер и като входящи данни за статистическия анализ, използвайте само редовете, показващи съответния продукт.



#### Задача 4

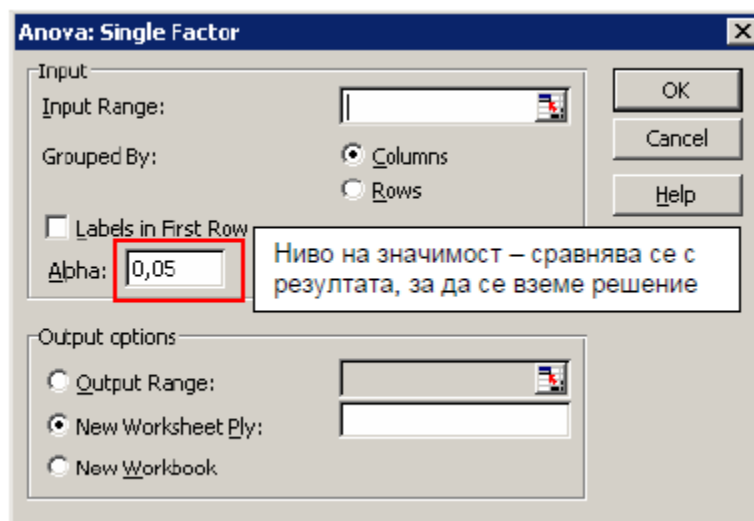
СофтПрес е специализирано издателство за компютърна литература. Наскоро компанията решава да направи проучване сред книжарниците в страната и да установи дали позицията на нейните книги на рафтовете с компютърна литература оказва влияние върху продажбите. По-конкретно,

издателството иска да провери какво е влиянието върху търсенето, ако неговите книги се поставят в началото, в средата или в края на рафтовете с компютърна литература. Събрана е информация от 25 книжарници:

Начало	Среда	Край
7	12	10
10	13	11
8	15	12
9	16	11
11	11	11
14	13	12
12	14	
13	15	
11		
10		
12		

**Указания:** За да се установи дали има или не значими различия в средните стойности на продажбите се използва статистическата функция дисперсионен анализ (Analysis Of Variance, ANOVA).

От менюто Data изберете Data Analysis и от появилия се прозорец посочете Anova: Single Factor. В полето Входящи данни (Input Range) маркирайте цялата информация за продажбите, включително наименованията на колоните. Сложете отметка пред Labels in First Row и потвърдете с ОК(фиг.9).



Фигура 9 Настройки на прозорец Anova: Single Factor.

Anova: Single Factor

## SUMMARY

Groups	Count	Sum	Average	Variance
Начало	11	117	10,63636364	4,454545455
Среда	8	109	13,625	2,839285714
Край	6	67	11,16666667	0,566666667

## ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	43,78621212	2	21,89310606	7,161653619	0,00402356	3,443356779
Within Groups	67,25378788	22	3,056990358			
Total	111,04	24				

Фигура 3.10 Получени данни от операцията Anova: Single Factor.

Стойностите, които стоят в колоната Average представляват средния брой продажби според местоположението на книгите. За да се определи дали между тези средни стойности има значителна статистическа разлика трябва да се погледне стойността в колоната *P-Value*. Ако тя е по-малка от нивото на значимост (0,05 в случая), може да се каже че има значителни разлики в продажбите в зависимост от местоположението на книгите и от резултатите може да се види, че най-добре е те да бъдат разположени в средата на рафтовете.

**САМОСТОЯТЕЛНА РАБОТА.** Фирма, която досега е работила само в региона на град Русе, решава да разшири дейността си на националния пазар. По тази причина, тя иска да разбере дали покупките на нейните продукти са еднакви във всички региони на България. Компанията събира

Таблица 1 Примерни данни за продажбите

Русе	Плевен	Варна	Бургас	Пловдив
10	11	9	14	14
11	13	9	17	16
9	14	16	17	16
13	17	17	18	18
14	11	12	16	17

примерни данни за обема на продажбите (в хиляди броя за седмица) в сравнително еднакви магазини в страната. Резултатите са показани в таблицата 1. Има ли значими разлики в продажбите в

различните региони на страната?

## Прогнозиране

Една от най-полезните характеристики на MS Excel е възможността на програмата за съставяне на прогнози. С помощта на функцията Data Analysis могат да се направят прогнози за бъдещи периоди посредством няколко основни метода – Подвижни средни стойности (Moving average), Изглаждане (Exponential Smoothing) и Регресия (Regression).



### Задача 5 Прогнозиране с подвижни средни стойности и изглаждане

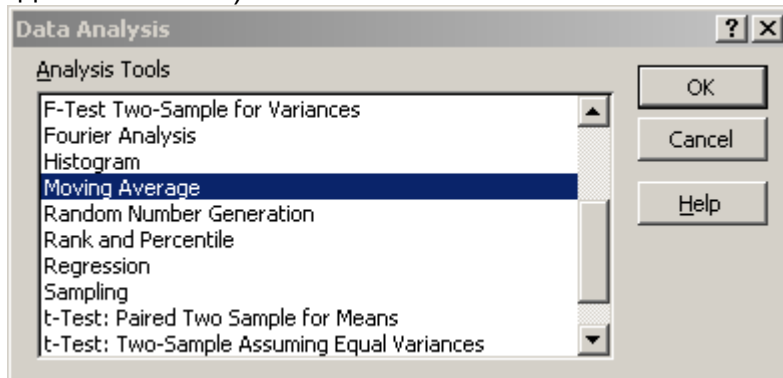
Използвайте данните от кол центъра на ОББ ([Y4-analysis.xls](#)), за да направите прогноза колко най-вероятно ще са обажданията за

следващия ден, така че най-добре да разпределите наличния си персонал.

**Указания:**

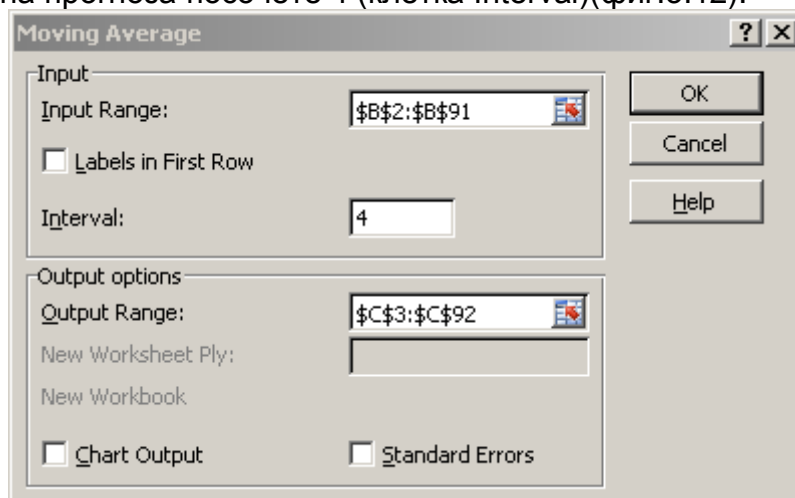
А) За прогноза с подвижни средни стойности се използват винаги данните от последните няколко периода (Може да се експериментира с различен брой периоди, за да се проследи коя прогноза ще даде по-добри резултати). В клетка С1 запишете **Прогноза 1**.

От менюто Data изберете Data Analysis и посочете Moving Average (фиг.11) (Подвижни средни стойности).



Фигура 3.11 Избор на Moving Average от Data Analysis.

Маркирайте съответните клетки на изходните данни и колоната в която да се появи прогнозата. Като брой на стойностите, които ще се използват за съставянето на прогноза посочете 4 (клетка Interval)(фиг.3.12).



Фигура 3.12 Примерни настройки в прозореца Moving Average за реализация на поставената задача.

Прогнозата за период 91 трябва да изглежда така:

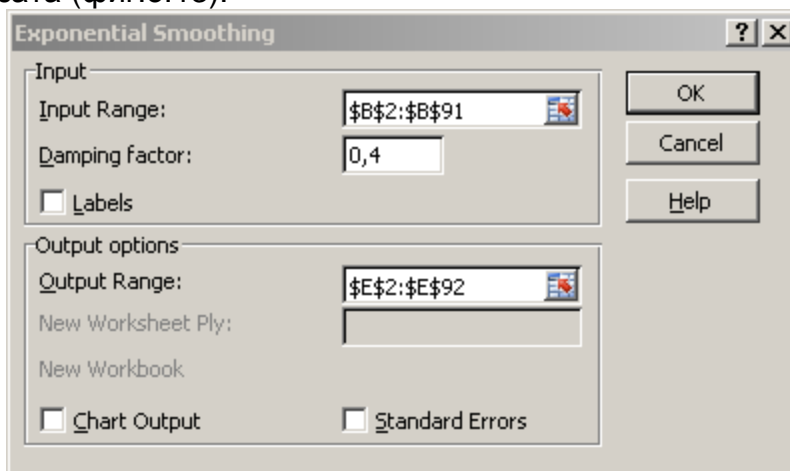
87	387	366.5
88	356	367.5
89	398	376.0
90	342	373.0
		370.8

Направете подобна прогноза в колона **D**, използвайки данните от 6 предходни периода.

Б) За прогноза с изглаждане се използват реалните данни за обажданията за предходния период и прогнозата за предходния период. Така не е необходимо да имате голямо количество информация, за да съставите



сравнително точна прогноза. В клетка E1 напишете **Прогноза 3**. От менюто Data изберете Data Analysis и посочете Exponential Smoothing. Маркирайте съответните клетки на изходните данни и колоната в която да се появи прогнозата (фиг.3.13).



Фигура 3.13 Примерни настройки на Exponential Smoothing.

В полето **Damping Factor** можете да записвате стойности между 0 и 1. Това е коефициента на изглаждане, който се използва за съставяне на прогнозата. Направете в колона F нова прогноза като използвате коефициент на изглаждане 0,7.



#### Задача 6 – самостоятелна работа

За да определите коя от съставените 4 прогнози е най-добра, трябва да пресметнете доколко всяка от тях се различава от реалните данни. Направете 4 нови колони с наименования **Разлика 1, Разлика 2, Разлика 3, Разлика 4**. Във всяка от тях трябва да пресметнете абсолютното отклонение на прогнозата от реалните данни. Например за **Разлика 1** в клетка G2 трябва да напишете следната формула **=ABS(B2-C2)**. Така ще получите абсолютната разлика между реалните данни за обажданията и прогнозните. В края на всяка колона с функцията AVERAGE трябва да пресметнете и средното отклонение. *Прогнозата с най-малко отклонение от действителните данни е най-добра.*

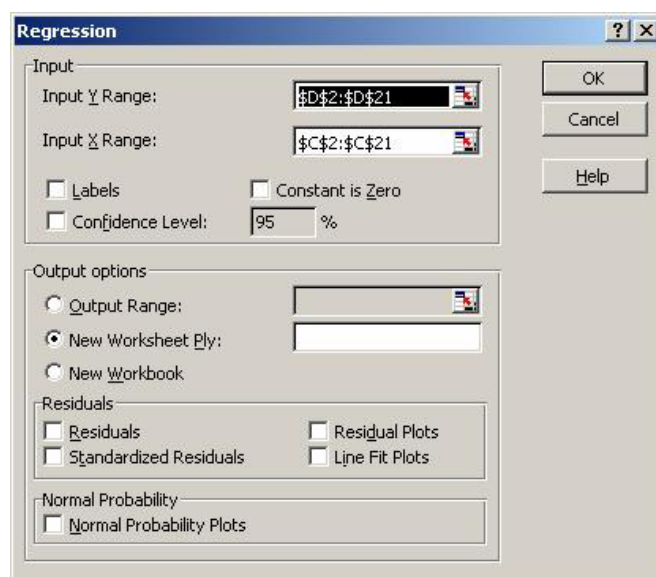


#### Задача 7 Прогнозиране с помощта на регресия

Регресията е много полезен инструмент за прогнозиране когато данните следват тенденция към увеличаване или намаляване. Използвайки данните от [У4-analysis.xls](#) за продажбите (sales) направете прогноза за продажбите през период 21 (първото тримесечие на 2006 година).

#### Указания:

От менюто Data изберете Data Analysis и посочете Regression. Въведете данните както е указано на фигурата 3.14



Фигура 3.14 Примерни настройки в прозореца Regression.

Посочете, получения резултат да излезе на нов лист в таблицата (New Worksheet Ply) и потвърдете с ОК.

Резултатът от тези действия трябва да е нов лист, изглеждащ приблизително по следния начин.

Важните данни от този анализ се намират в колоната с наименование Coefficients.

Това са коефициентите на уравнението за прогнозата, което в конкретния случай ще има вида:

Прогноза =  $262556 + 7070X$  ,

където X е номера на поредния период, за който се прави прогнозата. Например, за да намерите прогнозата за период 21, трябва да заместите X с 21.

#### SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,996337502
R Square	0,992688418
Adjusted R Square	0,992282219
Standard Error	3688,450782
Observations	20

#### ANOVA

	df	SS	MS
Regression	1	33247735955	33247735955
Residual	18	244884045,1	13604669,17
Total	19	33492620000	

	Coefficients	Standard Error	t Stat
Intercept	262556,3158	1713,401413	153,236897
X Variable 1	7070,827068	143,0319864	49,43528539



#### Задача 8 – самостоятелна работа

Съществува и графичен способ за намиране на регресионното уравнение. За тази цел можете да използвате графиката, която направихте в началото на упражнението. Щракнете с десен бутон върху линията, показваща продажбите и изберете Add Trendline. В показалия се прозорец щракнете Options и сложете отметка пред Display equation on chart. Така върху графиката ще се покаже уравнение с правилните коефициенти.