**ВСЕРОССИЙСКИЙ ЗАОЧНЫЙ ФИНАНСОВО-ЭКОНОМИЧЕСКИЙ ИНСТИТУТ**

**КАФЕДРА СТАТИСТИКИ**

**О Т Ч Е Т**

о результатах выполнения

компьютерной лабораторной работы

**Автоматизированный корреляционно-регрессионный анализ взаимосвязи статистических данных в среде MS Excel**

Вариант № 3

**Выполнил:** ст. III курса гр. ФК-2

**Проверил: доцент Лосева О.В**.

Пенза, 2008г.

# 1. Постановка задачи статистического исследования

Корреляционно-регрессионный анализ взаимосвязи признаков является составной частью проводимого статистического исследования деятельности 30-ти предприятий и частично использует результаты ЛР-1.

В ЛР-2 изучается взаимосвязь между факторным признаком *Среднегодовая стоимость основных производственных фондов* (признак **Х**) и результативным признаком *Выпуск продукции* (признак **Y**), значениями которых являются исходные данные ЛР-1 после исключения из них аномальных наблюдений.

|  |  |  |
| --- | --- | --- |
| Исходные данные | | |
| Номер предприятия | Среднегодовая стоимость основных производственных фондов, млн. руб. | Выпуск продукции, млн. руб. |
| 5 | 215,00 | 175,00 |
| 23 | 232,50 | 232,50 |
| 27 | 252,50 | 200,00 |
| 1 | 260,00 | 257,50 |
| 8 | 270,00 | 275,00 |
| 32 | 275,00 | 290,00 |
| 22 | 295,00 | 247,50 |
| 19 | 302,50 | 237,50 |
| 2 | 307,50 | 282,50 |
| 3 | 317,50 | 315,00 |
| 13 | 320,00 | 335,00 |
| 26 | 327,50 | 307,50 |
| 9 | 332,50 | 322,50 |
| 4 | 335,00 | 350,00 |
| 28 | 342,50 | 312,50 |
| 17 | 345,00 | 320,00 |
| 6 | 352,50 | 300,00 |
| 14 | 352,50 | 365,00 |
| 25 | 352,50 | 325,00 |
| 7 | 362,50 | 405,00 |
| 31 | 377,50 | 325,00 |
| 18 | 382,50 | 380,00 |
| 10 | 385,00 | 402,50 |
| 20 | 387,50 | 325,00 |
| 24 | 395,00 | 372,50 |
| 29 | 397,50 | 342,50 |
| 15 | 405,00 | 442,50 |
| 12 | 422,50 | 425,00 |
| 21 | 432,50 | 437,50 |
| 16 | 465,00 | 475,00 |

В процессе статистического исследования необходимо решить ряд задач.

1. Установить наличие ***статистической связи*** между факторным признаком **Х** и результативным признаком **Y** графическим методом.
2. Установить наличие ***корреляционной связи*** между признаками **Х** и **Y** методом аналитической группировки.
3. Оценить тесноту связи признаков **Х** и **Y** на основе эмпирического корреляционного отношения ***η***.
4. Построить однофакторную линейную регрессионную модель связи признаков **Х** и **Y**, используя инструмент **Регрессия** надстройки **Пакет анализа,** и оценить тесноту связи признаков **Х** и **Y** на основе линейного коэффициента корреляции ***r***.
5. Определить адекватность и практическую пригодность построенной линейной регрессионной модели, оценив:

а) значимость и доверительные интервалы коэффициентов ***а0, а1***;

б) индекс детерминации **R2** и его значимость;

в) точность регрессионной модели.

1. Дать экономическую интерпретацию:

а) коэффициента регрессии ***а1***;

б) коэффициента эластичности ***К*Э**;

в) остаточных величин ***εi***.

1. Найти наиболее адекватное нелинейное уравнение регрессии с помощью средств инструмента **Мастер диаграмм**.

**2. Выводы по результатам выполнения лабораторной работы[[1]](#footnote-2)**

**Задача 1**. Установление наличия ***статистической связи*** между факторным признаком **Х** и результативным признаком **Y** графическим методом.

Статистическая связь является разновидностью стохастической (случайной) связи, при которой с изменением факторного признака **X** ***закономерным*** образом изменяется какой–либо из обобщающих статистических показателей распределения результативного признака **Y**.

**Вывод:**

Точечный график связи признаков (диаграмма рассеяния, полученная в ЛР-1 после удаления аномальных наблюдений) позволяет сделать вывод, что имеет место статистическая связь. Предположительный вид связи – линейная прямая.

**Задача 2.** Установление наличия ***корреляционной связи*** между признаками **Х** и **Y** методом аналитической группировки.

Корреляционная связь – важнейший частный случай стохастической статистической связи, когда под воздействием вариации факторного признака **Х** ***закономерно*** изменяются от группы к группе ***средние групповые значения***  результативного признака **Y** (усредняются результативные значения , полученные под воздействием фактора ). Для выявления наличия корреляционной связи используется метод аналитической группировки.

**Вывод:**

Результаты выполнения аналитической группировки предприятий по факторному признаку *Среднегодовая стоимость основных производственных фондов* даны в табл. 2.2 Рабочего файла, которая показывает, что с увеличением значений факторного признака **Х** закономерно увеличиваются средние групповые значения результативного признака **.** Следовательно, между признаками **Х** и **Y** существует прямая связь.

**Задача 3**.Оценка тесноты связи признаков **Х** и **Y** на основе эмпирического корреляционного отношения.

Для анализа тесноты связи между факторным и результативным признаками рассчитывается показатель ***η*** – эмпирическое корреляционное отношение, задаваемое формулой

,

где  и  - соответственно межгрупповая и общая дисперсии результативного признака **Y** - *Выпуск продукции* (индекс ***х*** дисперсии  означает, что оценивается мера влияния признака **Х** на **Y)***.*

Для качественной оценки тесноты связи на основе показателя эмпирического корреляционного отношения служит шкала Чэддока:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Значение **η** | 0,1 – 0,3 | 0,3 – 0,5 | 0,5 – 0,7 | 0,7 – 0,9 | 0,9 – 0,99 |
| Сила связи | Слабая | Умеренная | Заметная | Тесная | Весьма тесная |

Результаты выполненных расчетов представлены в табл. 2.4 Рабочего файла.

**Вывод:**

Значение коэффициента ***η* =**0,90, что в соответствии с оценочной шкалой Чэддока говорит о весьма тесной степени связи изучаемых признаков.

**Задача 4**. Построение однофакторной линейной регрессионной модели связи изучаемых признаков с помощью инструмента **Регрессия** надстройки **Пакет анализа** и оценка тесноты связи на основе линейного коэффициента корреляции **r**.

4.1. Построение регрессионной модели заключается в нахождении аналитического выражения связи между факторным признаком **X** и результативным признаком **Y.**

Инструмент **Регрессия** на основе исходных данных (***xi , yi***), производит расчет параметров ***а0***и ***а1*** уравнения однофакторной линейной регрессии , а также вычисление ряда показателей, необходимых для проверки адекватности построенного уравнения исходным (фактическим) данным.

***Примечание****. В результате работы инструмента* ***Регрессия*** *получены четыре результативные таблицы (начиная с заданной ячейки* ***А75****).* ***Эти таблицы выводятся в Рабочий файл без нумерации, поэтому необходимо присвоить им номера табл.2.5 – табл.2.8 в соответствии с их порядком****.*

**Вывод:**

Рассчитанные в табл.2.7 (ячейки **В91** и **В92**) коэффициенты ***а0*** и ***а1***позволяют построить линейную регрессионную модель связи изучаемых признаков в виде уравнения -44,297+1,089x

4.2. В случае линейности функции связи для оценки ***тесноты связи*** признаков **X** и **Y,** устанавливаемой по построенной модели, используется ***линейный коэффициент корреляции r***.

Значение коэффициента корреляции **r** приводится в табл.2.5 в ячейке **В78** (термин "*Множественный* **R**").

**Вывод:**

Значение коэффициента корреляции ***r*** =0,91 , что в соответствии с оценочной шкалой Чэддока говорит о весьма тесной степени связи изучаемых признаков.

**Задача 5**. Анализ адекватности и практической пригодности построенной линейной регрессионной модели.

Анализ адекватности регрессионной модели преследует цель оценить, насколько построенная теоретическая модель взаимосвязи признаков отражает фактическую зависимость между этими признаками, и тем самым оценить практическую пригодность синтезированной модели связи.

Оценка соответствия построенной регрессионной модели исходным (фактическим) значениям признаков **X** и **Y** выполняется в 4 этапа:

1. оценка статистической значимости коэффициентов уравнения ***а0, а1*** и определение их доверительных интервалов для заданного уровня надежности;
2. определение практической пригодности построенной модели на основе оценок линейного коэффициента корреляции ***r*** и индекса детерминации **R2**;
3. проверка значимости уравнения регрессии в целом по **F**-критерию Фишера;
4. оценка погрешности регрессионной модели**.**
   1. ***Оценка статистической значимости коэффициентов уравнения а0, а1 и определение их доверительных интервалов***

Так как коэффициенты уравнения ***а0 , а1*** рассчитывались, исходя из значений признаков только для 30-ти пар (***xi , yi***), то полученные значения коэффициентов являются лишь ***приближенными оценками*** фактических параметров связи ***а0 , а1***. Поэтому необходимо:

* + 1. проверить значения коэффициентов на неслучайность (т.е. узнать, насколько они типичны для всей генеральной совокупности предприятий отрасли);
    2. определить (с заданной доверительной вероятностью **0,95** и **0,683**) пределы, в которых могут находиться значения ***а0, а1*** для генеральной совокупности предприятий.

Для анализа коэффициентов ***а0, а1*** линейного уравнения регрессии используется табл.2.7, в которой:

– значения коэффициентов ***а0, а1*** приведены в ячейках **В91** и **В92** соответственно;

– рассчитанный уровень значимости коэффициентов уравнения приведен в ячейках **Е91** и **Е92**;

– доверительные интервалы коэффициентов с уровнем надежности **Р=0,95** и **Р=0,683** указаны в диапазоне ячеек **F91:I92**.

*5.1.1. Определение значимости коэффициентов уравнения*

Уровень значимости – это величина ***α***=1–**Р**, где **Р** – заданный уровень надежности (доверительная вероятность).

Режим работы инструмента **Регрессия** использует по умолчанию уровень надежности **Р=0,95.** Для этого уровня надежности **у**ровень значимости равен ***α*** = 1 – 0,95 = **0,05.** Этот уровень значимости считается ***заданным***.

В инструменте **Регрессия** надстройки **Пакет анализа** для каждого из коэффициентов **а0** и ***а1*** вычисляется **у**ровень его значимости ***αр*,** который указан в результативной таблице (табл.**2.7** термин "Р-*значение*"). Если ***рассчитанный*** для коэффициентов ***а0, а1*** уровень значимости ***αр*,** меньше ***заданного*** уровня значимости ***α*= 0,05**, то этот коэффициент признается ***неслучайным*** (т.е. типичным для генеральной совокупности), в противном случае – ***случайным***.

***Примечание****. В случае, если признается случайным свободный член* ***а0****, то уравнение регрессии целесообразно построить заново без свободного члена* ***а0.*** *В этом случае в диалоговом окне* **Регрессия** *необходимо задать те же самые параметры за исключением лишь того, что следует активизировать флажок* ***Константа-ноль*** *(это означает, что модель будет строиться при условии* ***а0=0****). В лабораторной работе такой шаг не предусмотрен.*

*Если незначимым (случайным) является коэффициент регрессии* ***а1,*** *то взаимосвязь между признаками* ***X*** *и* ***Y*** *в принципе не может аппроксимироваться линейной моделью.*

**Вывод:**

Для свободного члена ***а0*** уравнения регрессии рассчитанный уровень значимости есть ***αр*** =0,1734.Так как он больше заданного уровня значимости ***α*=0,05**, то коэффициент ***а0*** признается случайным.

Для коэффициента регрессии ***а1*** рассчитанный уровень значимости есть ***αр*** =0,00.. Так как он меньше заданного уровня значимости ***α*=0,05**, то коэффициент ***а1*** признается типичным.

*5.1.2. Зависимость доверительных интервалов коэффициентов уравнения от заданного уровня надежности*

Доверительные интервалы коэффициентов ***а0, а1*** построенного уравнения регрессии при уровнях надежности **Р=0,95** и **Р**=**0,683** представлены в табл.2.7, на основе которой формируется табл.2.9.

Таблица 2.9

**Границы доверительных интервалов коэффициентов уравнения**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Коэффициенты | Границы доверительных интервалов | | | |
| Для уровня надежности **Р=0,95** | | Для уровня надежности **Р=0,683** | |
| нижняя | верхняя | нижняя | верхняя |
| ***а0*** | -109,25 | 20,66 | -76,60 | -11,98 |
| ***а1*** | 0,90 | 1,27 | 0,99 | 1,18 |

**Вывод:**

В генеральной совокупности предприятий значение коэффициента ***а0*** следует ожидать с надежностью **Р=0,95** в пределах **-109,25*а0*20,66,** значение коэффициента ***а1*** в пределах 0,90***а1*1,27.** Уменьшение уровня надежности ведет к сужению доверительных интервалов коэффициентов уравнения.

* 1. ***Определение практической пригодности построенной регрессионной модели.***

Практическую пригодность построенной модели можно охарактеризовать по величине линейного коэффициента корреляции ***r***:

* + близость  к единице свидетельствует о хорошей аппроксимации исходных (фактических) данных с помощью построенной линейной функции связи ;
  + близость  к нулю означает, что связь между фактическими данными **Х** и **Y** нельзя аппроксимировать как построенной, так и любой другой линейной моделью, и, следовательно, для моделирования связи следует использовать какую-либо подходящую нелинейную модель.

Пригодность построенной регрессионной модели для практического использования можно оценить и по величине ***индекса детерминации*** **R2**, показывающего, ***какая часть общей вариации признака* Y *объясняется в построенной модели вариацией фактора*** **X.**

В основе такой оценки лежит равенство **R = *r***(имеющее место для линейных моделей связи), а также шкала Чэддока, устанавливающая качественную характеристику тесноты связи в зависимости от величины ***r***.

Согласно шкале Чэддока ***высокая степень тесноты связи*** признаков достигается лишь при >0,7, т.е. при **** >0,7. Для индекса детерминации **R2** это означает выполнение неравенства **R2*>*0,5.**

При недостаточно тесной связи признаков **X**, **Y** (слабой, умеренной, заметной) имеет место неравенство 0,7, а следовательно, и неравенство **.**

С учетом вышесказанного, практическая пригодность построенной модели связи  оценивается по величине **R2** следующим образом:

* + неравенство **R2 >0,5** позволяет считать, что построенная модель пригодна для практического применения, т.к. в ней достигается высокая степень тесноты связи признаков **X** и **Y**, при которой более 50% вариации признака **Y** объясняется влиянием фактора **Х**;
  + неравенство  означает, что построенная модель связи практического значения не имеет ввиду недостаточной тесноты связи между признаками **X** и **Y**, при которойменее 50% вариации признака **Y** объясняется влиянием фактора **Х**, и, следовательно, фактор **Х** влияет на вариацию **Y** в значительно меньшей степени, чем другие (неучтенные в модели) факторы.

Значение индекса детерминации **R2** приводится в табл.2.5 в ячейке **В79** (термин "**R -** *квадрат*").

**Вывод:**

Значение линейного коэффициента корреляции ***r*** и значение индекса детерминации **R2** согласно табл. 2.5 равны: ***r*** =0,91, **R2** =0,83. Поскольку  и  то построенная линейная регрессионная модель связи пригодна для практического использования.

* 1. ***Общая оценка адекватности регрессионной модели по F-критерию Фишера***

Адекватность построенной регрессионной модели фактическим данным (***xi, yi***) устанавливается по критерию Р.Фишера, оценивающему статистическую значимость (неслучайность) индекса детерминации **R2**.

Рассчитанная для уравнения регрессии оценка значимости **R2** приведена в табл.2.6 в ячейке **F86** (термин "*Значимость* ***F***"). Если она меньше заданного уровня значимости ***α***=**0,05**, то величина **R2** признается неслучайной и, следовательно, построенное уравнение регрессии  может быть использовано как модель связи между признаками **Х** и **Y** для ***генеральной совокупности*** предприятий отрасли.

**Вывод:**

Рассчитанный уровень значимости ***αр*** индекса детерминации **R2** есть ***αр=***0,00… Так как он меньше заданного уровня значимости ***α***=0,05, то значение **R2** признается типичным и модель связи между признаками **Х** и **Y** -44,297+1,089x применима для генеральной совокупности предприятий отрасли в целом.

* 1. ***Оценка погрешности регрессионной модели***

Погрешность регрессионной модели можно оценить по величине стандартной ошибки ****** построенного линейного уравнения регрессии . Величина ошибки ****** оценивается как среднее квадратическое отклонение по совокупности отклонений  исходных (фактических) значений ***yi*** признака **Y** от его теоретических значений , рассчитанных по построенной модели.

Погрешность регрессионной модели выражается в процентах и рассчитывается как величина **.100.**

В адекватных моделях погрешность не должна превышать 12%-15%.

Значение ****** приводится в выходной таблице "*Регрессионная статистика*" (табл.2.5) в ячейке **В81** (термин "*Стандартная ошибка*"), значение  – в таблице описательных статистик (ЛР-1, ***Лист 1***, табл.3, столбец 2).

**Вывод:**

Погрешность линейной регрессионной модели составляет **.100**=29,9177/326,0833\*100=9,1748 %, что подтверждает адекватность построенной модели -44,297+1,089x

**Задача 6**. Дать экономическую интерпретацию:

1) коэффициента регрессии ***а1***;

3) остаточных величин ***i***.

2) коэффициента эластичности ***К*Э**;

***6.1. Экономическая интерпретация коэффициента регрессии а1***

В случае линейного уравнения регрессии ****=*a0+a1x*** величина коэффициента регрессии ***a1*** показывает, на сколько ***в среднем*** (в абсолютном выражении) изменяется значение результативного признака **Y** при изменении фактора **Х** на единицу его измерения. Знак при ***a1*** показывает направление этого изменения.

**Вывод:**

Коэффициент регрессии ***а1*** =1,089 показывает, что при увеличении факторного признака *Среднегодовая стоимость основных производственных фондов* на 1 млн руб. значение результативного признака *Выпуск продукции* увеличивается в среднем на 1,089 млн руб.

***6.2. Экономическая интерпретация коэффициента эластичности.***

С целью расширения возможностей экономического анализа явления используется коэффициент эластичности , который измеряется в процентах и показывает, ***на сколько процентов*** изменяется ***в среднем*** результативный признак при изменении факторного признака на 1%.

Средние значения  и  приведены в таблице описательных статистик (ЛР-1, ***Лист 1***, табл.3).

Расчет коэффициента эластичности:

=1,089\*(340/326,0833) =1,1355%

**Вывод:**

Значение коэффициента эластичности ***Кэ***=1,1355 показывает, что при увеличении факторного признака *Среднегодовая стоимость основных производственных фондов* на 1% значение результативного признака *Выпуск продукции* увеличивается в среднем на 1,1355%.

***6.3. Экономическая интерпретация остаточных величин εi***

Каждый их остатков  характеризует отклонение фактического значения ***yi*** от теоретического значения , рассчитанного по построенной регрессионной модели и определяющего, какого среднего значения  следует ожидать, когда фактор **Х** принимает значение ***xi***.

Анализируя остатки, можно сделать ряд практических выводов, касающихся выпуска продукции на рассматриваемых предприятиях отрасли.

Значения остатков ***i*** (таблица остатков из диапазона **А98:С128**) имеют как положительные, так и отрицательные отклонения от ожидаемого в среднем объема выпуска продукции  (которые в итоге уравновешиваются, т.е.).

Экономический интерес представляют ***наибольшие расхождения*** между фактическим объемом выпускаемой продукции ***yi*** и ожидаемым усредненным объемом .

**Вывод:**

Согласно таблице остатков максимальное превышение ожидаемого среднего объема выпускаемой продукции  имеют три предприятия - с номерами 6, 20, 27, а максимальные отрицательные отклонения - три предприятия с номерами 8, 26, 24 .Именно эти шесть предприятий подлежат дальнейшему экономическому анализу для выяснения причин наибольших отклонений объема выпускаемой ими продукции от ожидаемого среднего объема и выявления резервов роста производства.

**Задача 7.** Нахождение наиболее адекватного ***нелинейного*** уравнения регрессии с помощью средств инструмента **Мастер диаграмм.**

Уравнения регрессии и их графики построены для 3-х видов нелинейной зависимости между признаками и представлены на диаграмме 2.1 ***Рабочего файла***.

Уравнения регрессии и соответствующие им индексы детерминации **R2** приведены в табл.2.10 (при заполнении данной таблицы ***коэффициенты уравнений необходимо указывать не в компьютерном формате, а в общепринятой десятичной форме чисел***).

## Таблица 2.10

## Регрессионные модели связи

|  |  |  |
| --- | --- | --- |
| Вид уравнения | Уравнение регрессии | Индекс  детерминации **R2** |
| Полином 2-го порядка | 0,0006x2+0,6816x+21,986 | 0,8353 |
| Полином 3-го порядка | -0,4x3-0,0119x2+4,867x-421,85 | 0,8381 |
| Степенная функция | 0,4087x1,1452 | 0,8372 |

Выбор наиболее адекватного уравнения регрессии определяется максимальным значением индекса детерминации **R2**: ***чем ближе значение R2 к единице, тем более точно регрессионная модель соответствует фактическим данным***.

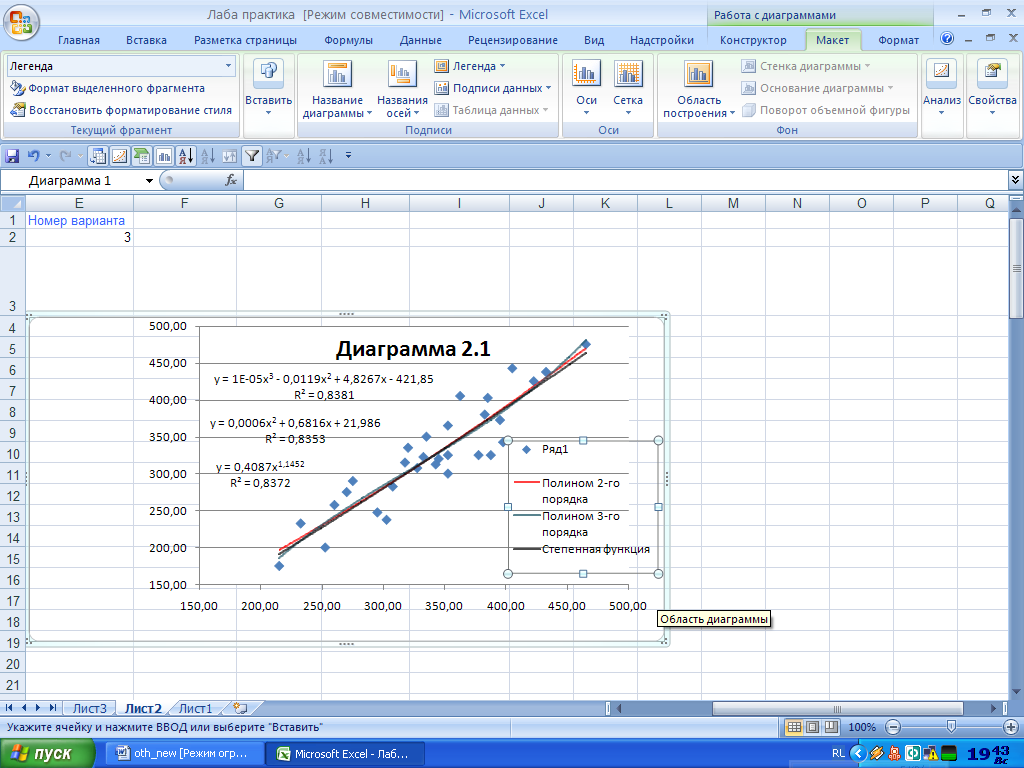
**Вывод:**

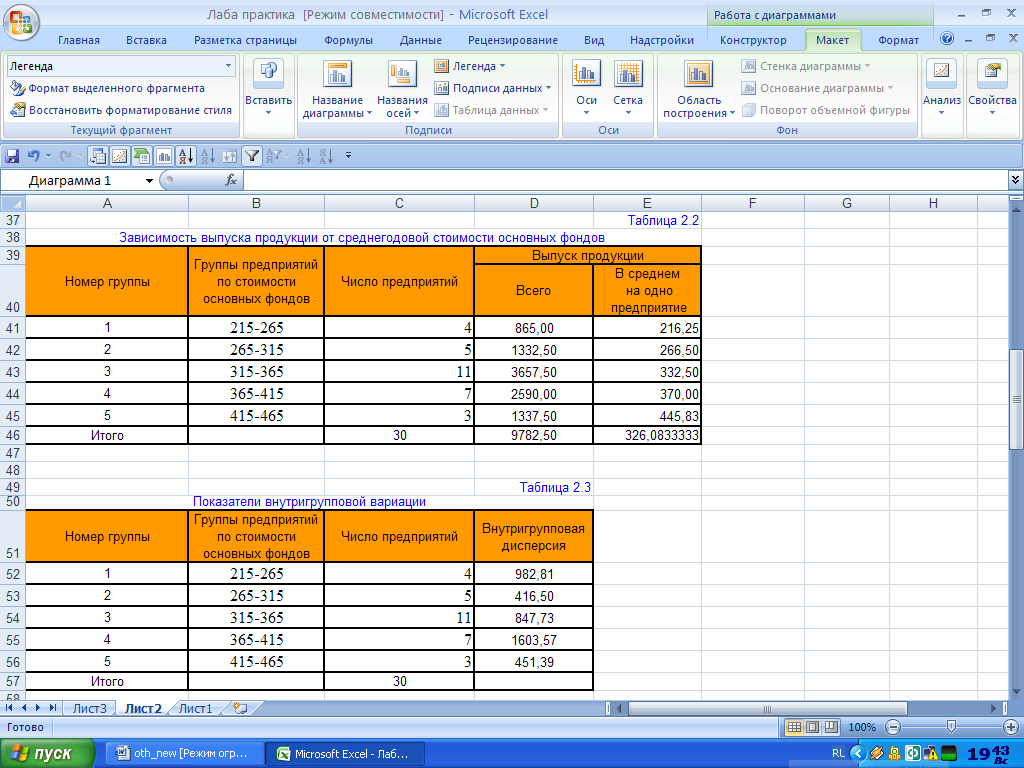
Максимальное значение индекса детерминации **R2** =0,8381.Следовательно, наиболее адекватное исходным данным нелинейное уравнение регрессии имеет вид 0,4087x1,1452.

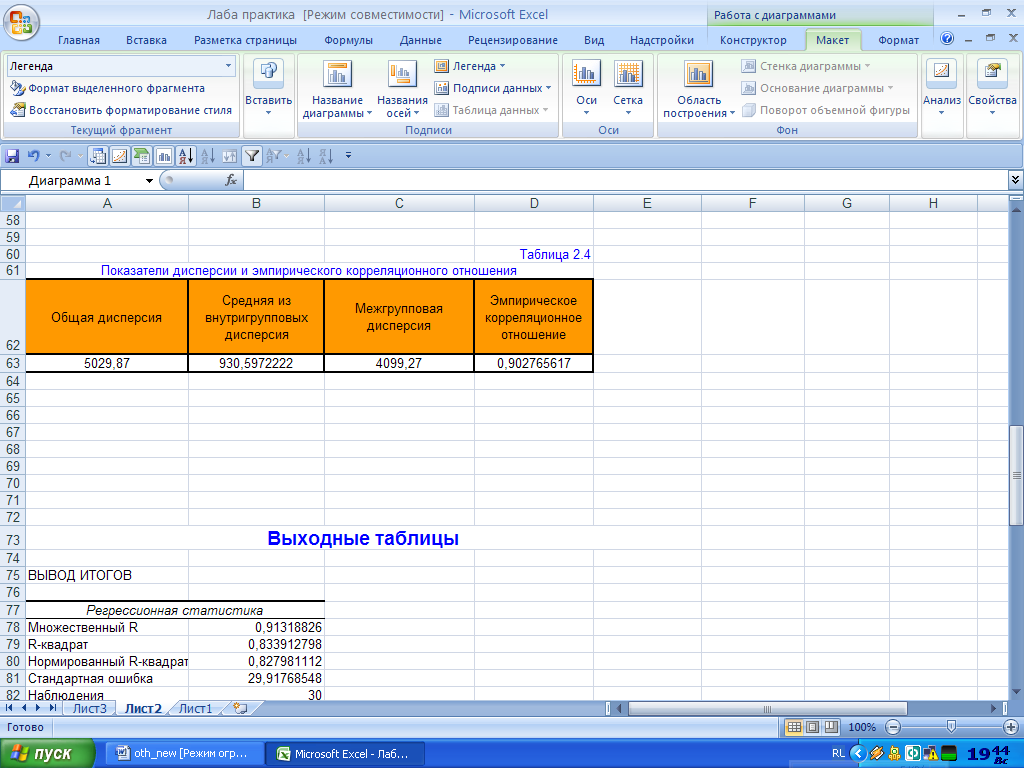
**ПРИЛОЖЕНИЕ**

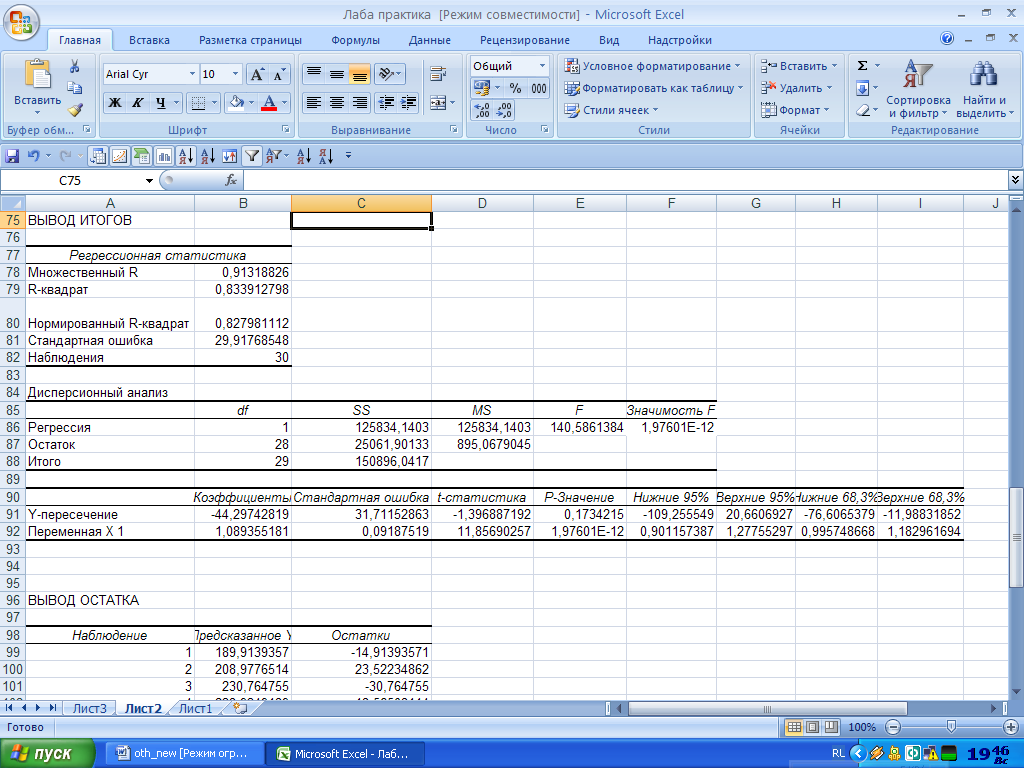
**Результативные таблицы и графики**

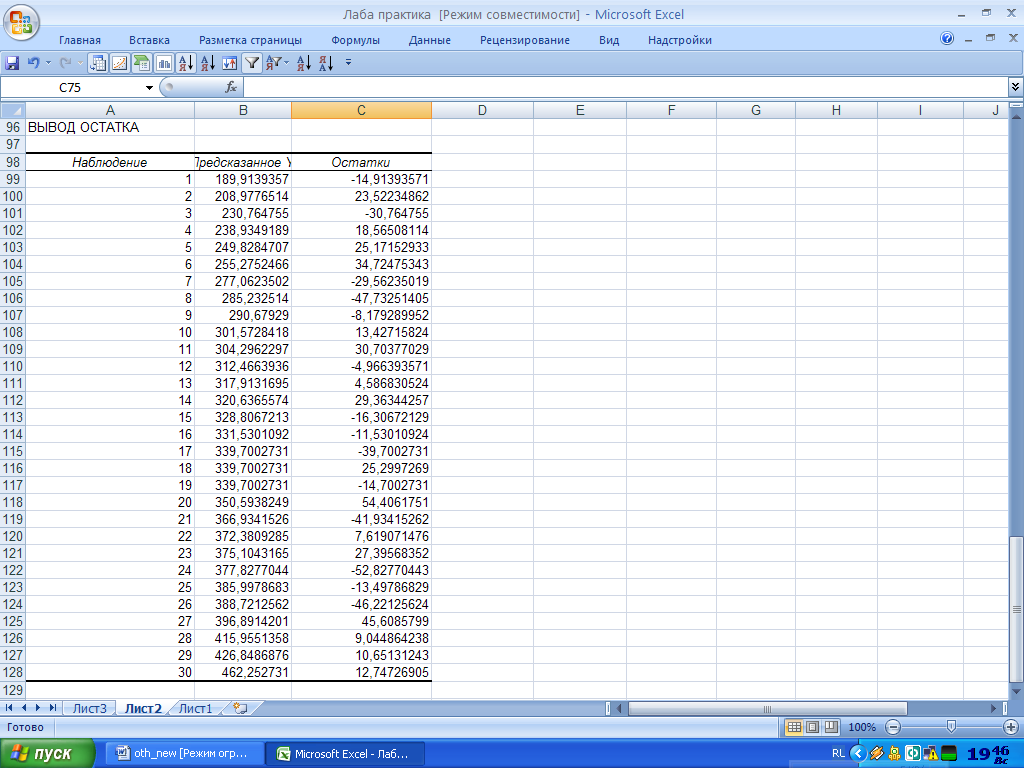
|  |  |  |
| --- | --- | --- |
| Номер предприятия | Среднегодовая стоимость основных производственных фондов, млн.руб. | Выпуск продукции, млн. руб. |
| 5 | 215,00 | 175,00 |
| 23 | 232,50 | 232,50 |
| 27 | 252,50 | 200,00 |
| 1 | 260,00 | 257,50 |
| 8 | 270,00 | 275,00 |
| 32 | 275,00 | 290,00 |
| 22 | 295,00 | 247,50 |
| 19 | 302,50 | 237,50 |
| 2 | 307,50 | 282,50 |
| 3 | 317,50 | 315,00 |
| 13 | 320,00 | 335,00 |
| 26 | 327,50 | 307,50 |
| 9 | 332,50 | 322,50 |
| 4 | 335,00 | 350,00 |
| 28 | 342,50 | 312,50 |
| 17 | 345,00 | 320,00 |
| 6 | 352,50 | 300,00 |
| 14 | 352,50 | 365,00 |
| 25 | 352,50 | 325,00 |
| 7 | 362,50 | 405,00 |
| 31 | 377,50 | 325,00 |
| 18 | 382,50 | 380,00 |
| 10 | 385,00 | 402,50 |
| 20 | 387,50 | 325,00 |
| 24 | 395,00 | 372,50 |
| 29 | 397,50 | 342,50 |
| 15 | 405,00 | 442,50 |
| 12 | 422,50 | 425,00 |
| 21 | 432,50 | 437,50 |
| 16 | 465,00 | 475,00 |



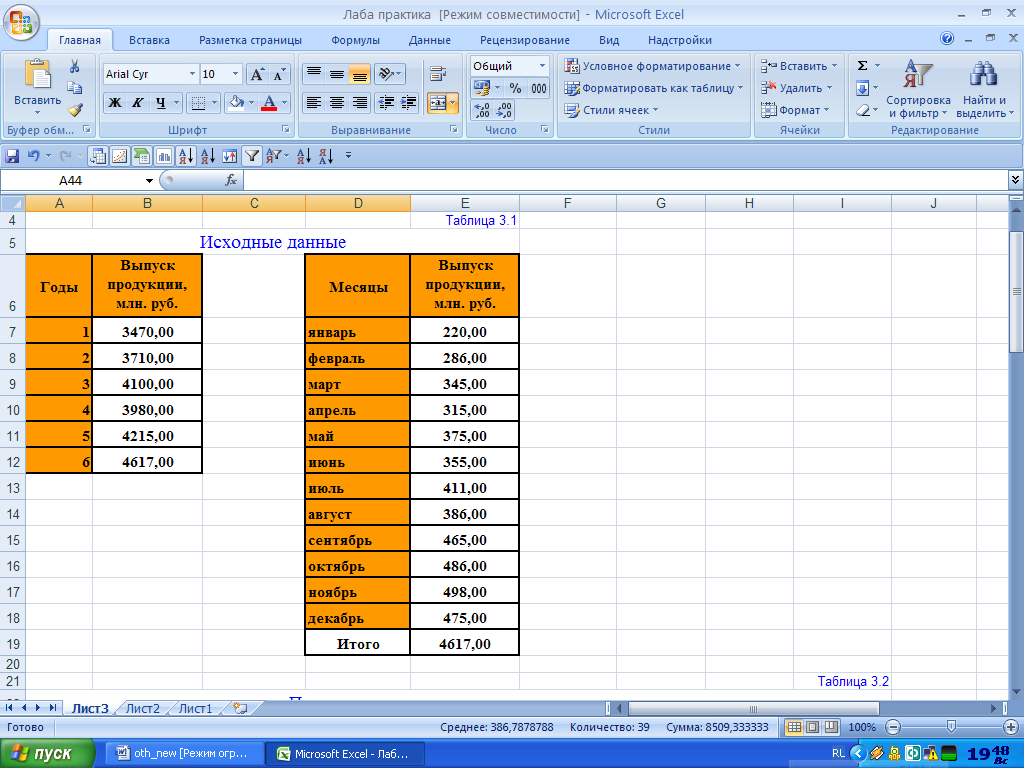


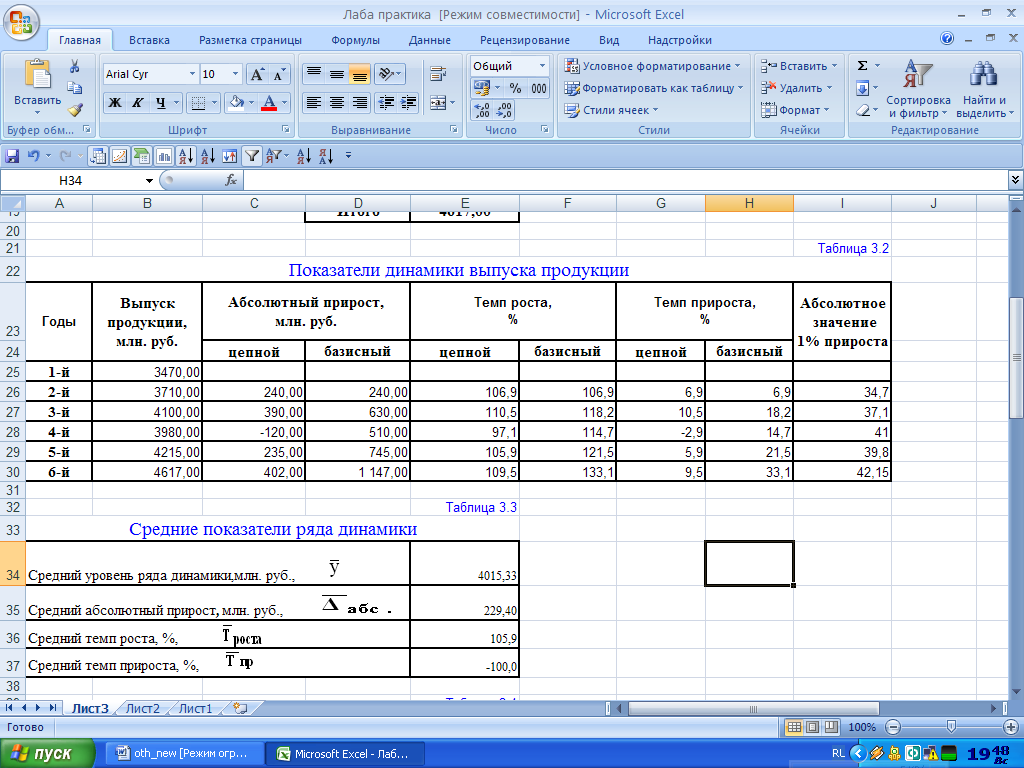


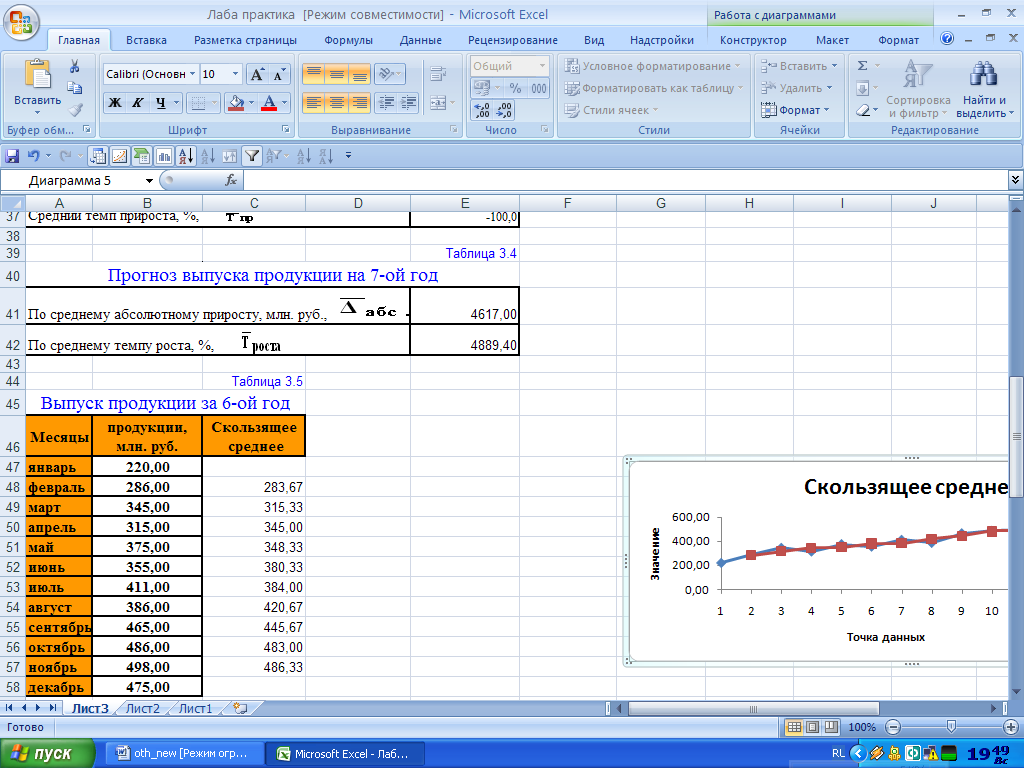


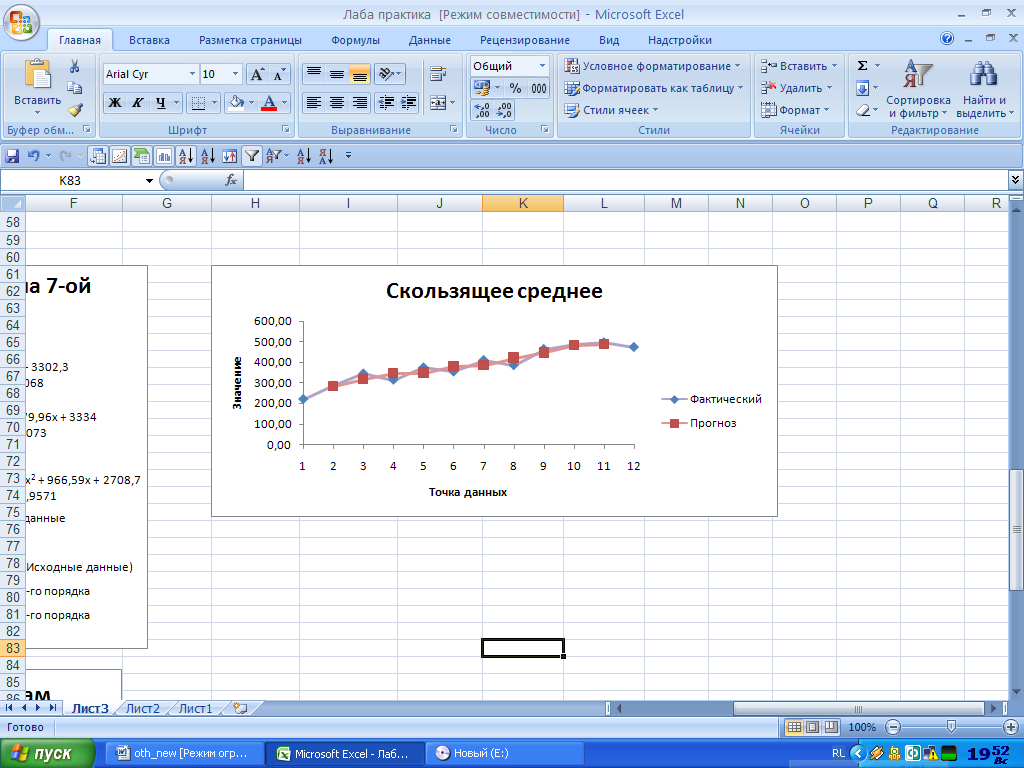


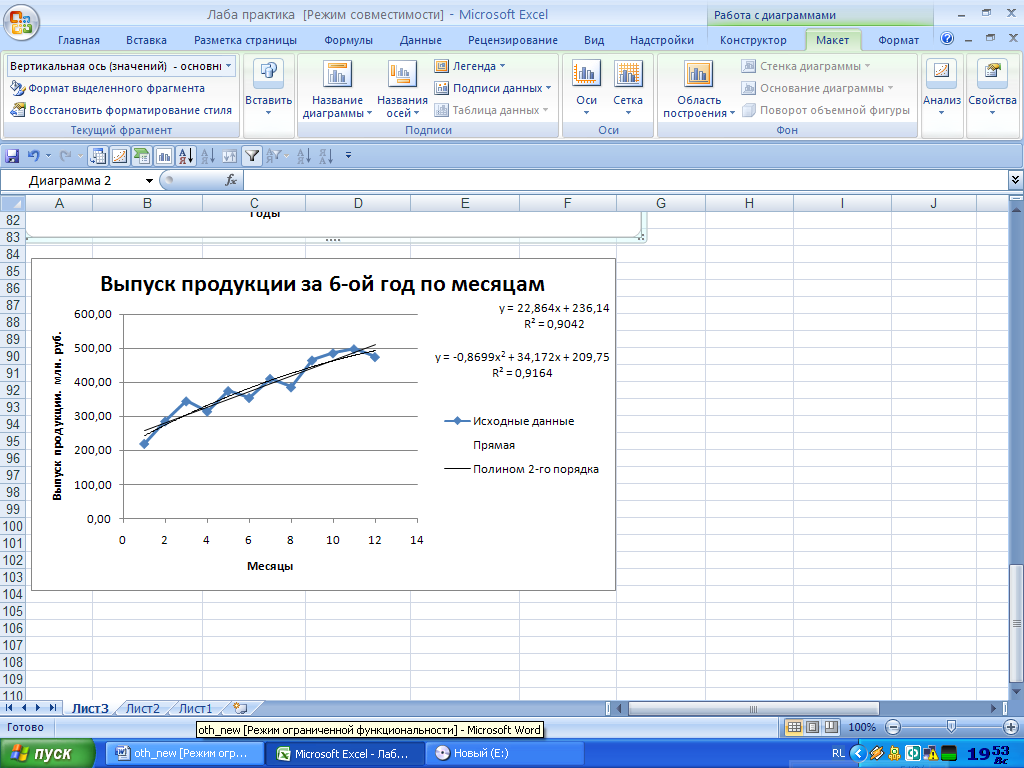
***Лабораторная работа №3***

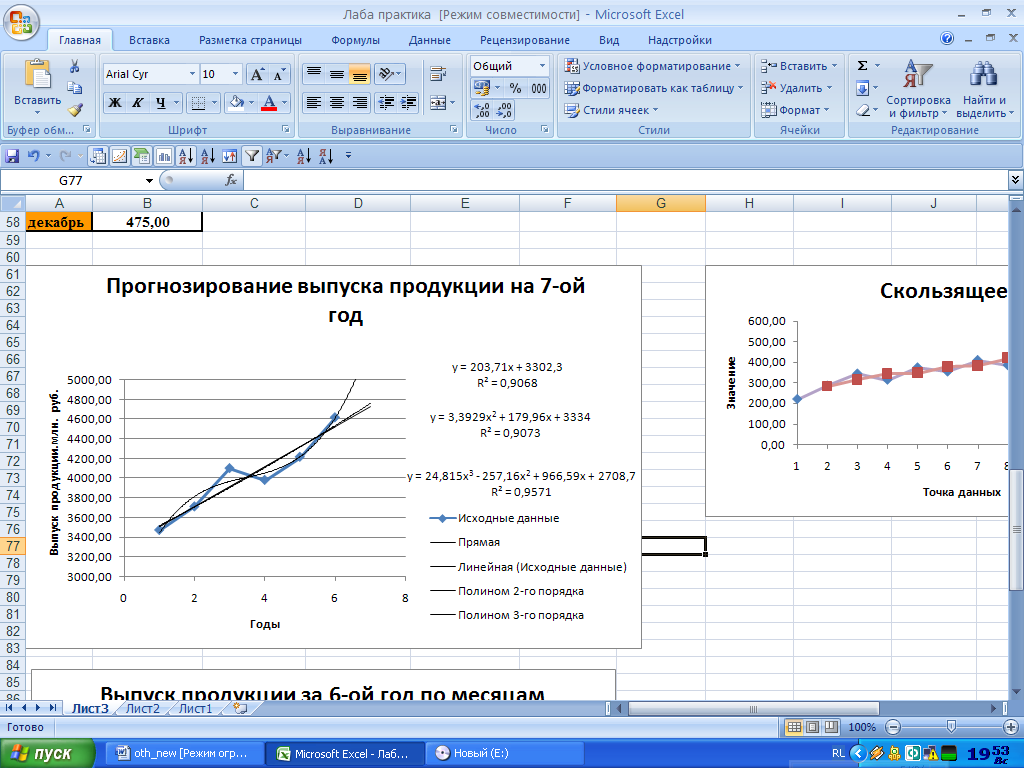












1. Все статистические показатели необходимо представить в таблицах с точностью до 4-х знаков после запятой. Таблицы и пробелы в формулировках выводов заполнять ***вручную***. В выводах при выборе альтернативного варианта ответа ***ненужный вариант*** ***вычеркивается***. [↑](#footnote-ref-2)