МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное образовательное учреждение высшего образования

Национальный исследовательский технологический университет «МИСиС»

Институт экономики и управления промышленными предприятиями им. В.А. Роменца

Кафедра промышленного менеджмента

**Курсовая работа**

По дисциплине «Статистика»

на тему:

“Статистическое исследование выборки”

Выполнил: Студент группы

БЭК-20-1: Михальцов Дмитрий Сергеевич

Научный руководитель: к.э.н., доцент

Богачев А. С.

Москва 2021

# **1 Описательная статистика**

В качестве исходных данных были взяты значения за период **с 1995 по 2021 года** по следующим макроэкономическим показателям **Эстония** (таблица 1):

1. Валовой внутренний продукт, текущие цены (млрд. $);
2. Общие инвестиции, % от ВВП;
3. Инфляция, % на конец периода;
4. Численность населения, млн. человек;
5. Уровень безработицы, % от общей рабочей силы.

Таблица 1 – Исходные данные

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Год | Валовой внутренний продукт, текущие цены (млрд. $) | Общие инвестиции, % от ВВП | Инфляция, % на конец периода | Численность населения млн, человек | Уровень безработицы, % от общей рабочей силы |
| 1995 | 3,909 | 26,408 | 38,163 | 9,645 | 1,437 |
| 1996 | 4,785 | 27,487 | 45,697 | 9,881 | 1,416 |
| 1997 | 5,16 | 30,726 | 49,952 | 9,65 | 1,4 |
| 1998 | 5,678 | 31,623 | 54,338 | 9,831 | 1,386 |
| 1999 | 5,772 | 25,904 | 56,022 | 12,201 | 1,39 |
| 2000 | 5,707 | 28,475 | 58,224 | 14,602 | 1,397 |
| 2001 | 6,269 | 29,58 | 61,5 | 13,009 | 1,388 |
| 2002 | 7,391 | 32,941 | 63,707 | 11,227 | 1,379 |
| 2003 | 9,891 | 35,198 | 64,59 | 10,342 | 1,371 |
| 2004 | 12,158 | 34,655 | 66,553 | 10,14 | 1,363 |
| 2005 | 14,126 | 33,356 | 69,293 | 8,031 | 1,355 |
| 2006 | 17,047 | 39,832 | 72,372 | 5,912 | 1,347 |
| 2007 | 22,476 | 40,008 | 77,253 | 4,592 | 1,341 |
| 2008 | 24,433 | 31,634 | 85,447 | 5,455 | 1,337 |
| 2009 | 19,693 | 21,01 | 85,617 | 13,549 | 1,335 |
| 2010 | 19,568 | 21,692 | 87,963 | 16,707 | 1,331 |
| 2011 | 23,21 | 25,679 | 92,432 | 12,325 | 1,327 |
| 2012 | 23,034 | 29,361 | 96,332 | 10,023 | 1,323 |
| 2013 | 25,116 | 27,198 | 99,459 | 8,628 | 1,318 |
| 2014 | 26,641 | 27,099 | 99,933 | 7,351 | 1,315 |
| 2015 | 22,893 | 25,001 | 100 | 6,185 | 1,315 |
| 2016 | 24,066 | 25,119 | 100,8 | 6,758 | 1,316 |
| 2017 | 26,915 | 26,382 | 104,48 | 5,763 | 1,317 |
| 2018 | 30,503 | 26,913 | 108,045 | 5,371 | 1,322 |
| 2019 | 31,049 | 26,083 | 110,496 | 4,448 | 1,327 |
| 2020 | 30,626 | 30,244 | 109,795 | 6,806 | 1,329 |
| 2021 | 36,287 | 30,524 | 114,722 | 6,181 | 1,33 |

**Первым делом определим наличие значимых и незначимых выбросов.**

Выбросы определяются через анализ значений – попадают ли они или нет в пределы так называемых «внутренних границ» и «внешних границ». Значение, лежащее вне «внутренних границ», классифицируется как «незначительный выброс», в то время как значение, находящееся за «внешними границами», классифицируется как «значительный выброс». Чтобы найти внутренние границы, необходимо умножить межквартильный диапазон на 1,5; результат нужно прибавить к Q3 и вычесть из Q1. Межквартильный диапазон рассчитывается как Q3-Q1.

Нахождение внешних границ делается таким же образом, как для внутренних границ, за исключением того, что межквартильный диапазон умножается на 3, а не на 1,5. Результат нужно прибавить к Q3 и вычесть из Q1. Два найденных числа являются внешними границами набора данных.

Таблица 2 – Определение внутренних и внешних границ для поиска выбросов

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Показатель | Валовой внутренний продукт, текущие цены (млрд. $) | Общие инвестиции, % от ВВП | Инфляция, % на конец периода | Численность населения, млн. человек | Уровень безработицы, % от общей рабочей силы |
| Q1 | 6,83 | 26,2325 | 62,6035 | 6,183 | 1,325 |
| Q3 | 24,7745 | 31,6285 | 99,9665 | 10,7845 | 1,3825 |
| Q3-Q1 | 17,9445 | 5,396 | 37,363 | 4,6015 | 0,0575 |
| 1,5∙(Q3-Q1) | 26,91675 | 8,094 | 56,0445 | 6,90225 | 0,08625 |
| 3∙(Q3-Q1) | 53,8335 | 16,188 | 112,089 | 13,8045 | 0,1725 |
| Внутренние границы | -20,08675 | 18,1385 | 6,559 | -0,71925 | 1,23875 |
| 51,69125 | 39,7225 | 156,011 | 17,68675 | 1,46875 |
| Внешние границы | -47,0035 | 10,0445 | -49,4855 | -7,6215 | 1,1525 |
| 78,608 | 47,8165 | 212,0555 | 24,589 | 1,555 |

Анализ показал, что в исходной выборке выбросов не наблюдается

**Определение типа распределения для показателя «Инфляция, потребительские цены на конец периода, процентное изменение».**

Самый простой графический способ проверки характера распределения данных – построение гистограммы. Если гистограмма имеет колоколообразный симметричный вид, можно сделать заключение о том, что анализируемая переменная имеет примерно нормальное распределение.

Построим гистограмму распределения частот.

По формуле Стерджесса рассчитаем количество интервалов:

, т.е. количество интервалов выбираем равное 6 (n – количество элементов в выборке).

Определим параметры для построения гистограмм.

Таблица 3 – Параметры для построения гистограмм

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Показатель | Валовой внутренний продукт, текущие цены (млрд. $) | Общие инвестиции, % от ВВП | Инфляция, % на конец периода | Численность населения млн, человек | Уровень безработицы, % от общей рабочей силы |
| Коэффициент вариации | 0,553254065 | 0,159637974 | 0,281012802 | 0,359348378 | 0,026057494 |
| Максимальное значение | 36,287 | 40,008 | 114,722 | 16,707 | 1,437 |
| Минимальное значение | 3,909 | 21,01 | 38,163 | 4,448 | 1,315 |
| Ширина интервала | 5,396333333 | 3,166333333 | 12,75983333 | 2,043166667 | 0,020333333 |
| Границы интервалов | 3,909 | 21,01 | 38,163 | 4,448 | 1,315 |
| 9,305333333 | 24,17633333 | 50,92283333 | 6,491166667 | 1,335333333 |
| 14,70166667 | 27,34266667 | 63,68266667 | 8,534333333 | 1,355666667 |
| 20,098 | 30,509 | 76,4425 | 10,5775 | 1,376 |
| 25,49433333 | 33,67533333 | 89,20233333 | 12,62066667 | 1,396333333 |
| 30,89066667 | 36,84166667 | 101,9621667 | 14,66383333 | 1,416666667 |
| 36,287 | 40,008 | 114,722 | 16,707 | 1,437 |

Рисунок 1 – Гистограммы показателей

Рисунок 2 – Q-Q графики показателей

Другим очень часто используемым графическим способом проверки характера распределения данных является построение графиков квантилей. На таких графиках изображаются квантили двух распределений – эмпирического и теоретически ожидаемого стандартного нормального распределения. При нормальном распределении проверяемой переменной точки на графике квантилей должны выстраиваться в прямую линию, исходящую под углом 45 градусов из левого нижнего угла графика. Графики квантилей особенно полезны при работе с небольшими по размеру совокупностями, для которых невозможно построить гистограммы, принимающие какую-либо выраженную форму.

Рисунок 3 – Блочные гистограммы показателей

При работе с выборками малого объема приходится ограничиться применением более простых и грубых критериев, используемых в качестве «прикидочных» при большом объеме выборочных данных. Они основаны на использовании коэффициента вариации, среднего абсолютного отклонения, размаха варьирования, показателей асимметрии и эксцесса. Более или менее уверенный вывод о подтверждении гипотезы нормальности на основе данных малой выборки можно сделать только при получении положительных результатах проверки несколькими критериями этой группы.

**Проверка с помощью критерия среднего абсолютного отклонения**

Таблица 4 – Подготовка данных для критерия среднего абсолютного отклонения (“Валовой внутренний продукт, текущие цены (млрд. $)” и “Общие инвестиции, % от ВВП”)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Год | Валовой внутренний продукт, текущие цены (млрд. $) | |  | | --- | |  | | Год | Общие инвестиции, % от ВВП | |  | | --- | |  | |
| 1995 | 3,909 | 14,03185185 | 1995 | 26,408 | 2,856148148 |
| 1996 | 4,785 | 13,15585185 | 1996 | 27,487 | 1,777148148 |
| 1997 | 5,16 | 12,78085185 | 1997 | 30,726 | 1,461851852 |
| 1998 | 5,678 | 12,26285185 | 1998 | 31,623 | 2,358851852 |
| 1999 | 5,772 | 12,16885185 | 1999 | 25,904 | 3,360148148 |
| 2000 | 5,707 | 12,23385185 | 2000 | 28,475 | 0,789148148 |
| 2001 | 6,269 | 11,67185185 | 2001 | 29,58 | 0,315851852 |
| 2002 | 7,391 | 10,54985185 | 2002 | 32,941 | 3,676851852 |
| 2003 | 9,891 | 8,049851852 | 2003 | 35,198 | 5,933851852 |
| 2004 | 12,158 | 5,782851852 | 2004 | 34,655 | 5,390851852 |
| 2005 | 14,126 | 3,814851852 | 2005 | 33,356 | 4,091851852 |
| 2006 | 17,047 | 0,893851852 | 2006 | 39,832 | 10,56785185 |
| 2007 | 22,476 | 4,535148148 | 2007 | 40,008 | 10,74385185 |
| 2008 | 24,433 | 6,492148148 | 2008 | 31,634 | 2,369851852 |
| 2009 | 19,693 | 1,752148148 | 2009 | 21,01 | 8,254148148 |
| 2010 | 19,568 | 1,627148148 | 2010 | 21,692 | 7,572148148 |
| 2011 | 23,21 | 5,269148148 | 2011 | 25,679 | 3,585148148 |
| 2012 | 23,034 | 5,093148148 | 2012 | 29,361 | 0,096851852 |
| 2013 | 25,116 | 7,175148148 | 2013 | 27,198 | 2,066148148 |
| 2014 | 26,641 | 8,700148148 | 2014 | 27,099 | 2,165148148 |
| 2015 | 22,893 | 4,952148148 | 2015 | 25,001 | 4,263148148 |
| 2016 | 24,066 | 6,125148148 | 2016 | 25,119 | 4,145148148 |
| 2017 | 26,915 | 8,974148148 | 2017 | 26,382 | 2,882148148 |
| 2018 | 30,503 | 12,56214815 | 2018 | 26,913 | 2,351148148 |
| 2019 | 31,049 | 13,10814815 | 2019 | 26,083 | 3,181148148 |
| 2020 | 30,626 | 12,68514815 | 2020 | 30,244 | 0,979851852 |
| 2021 | 36,287 | 18,34614815 | 2021 | 30,524 | 1,259851852 |

Таблица 5 – Подготовка данных для критерия среднего абсолютного отклонения (“Инфляция, % на конец периода” и “ Численность населения млн, человек”)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Год | Инфляция, % на конец периода | |  | | --- | |  | | Год | Численность населения млн, человек | |  | | --- | |  | |
| 1995 | 38,163 | 42,32533333 | 1995 | 9,645 | 0,585259259 |
| 1996 | 45,697 | 34,79133333 | 1996 | 9,881 | 0,821259259 |
| 1997 | 49,952 | 30,53633333 | 1997 | 9,65 | 0,590259259 |
| 1998 | 54,338 | 26,15033333 | 1998 | 9,831 | 0,771259259 |
| 1999 | 56,022 | 24,46633333 | 1999 | 12,201 | 3,141259259 |
| 2000 | 58,224 | 22,26433333 | 2000 | 14,602 | 5,542259259 |
| 2001 | 61,5 | 18,98833333 | 2001 | 13,009 | 3,949259259 |
| 2002 | 63,707 | 16,78133333 | 2002 | 11,227 | 2,167259259 |
| 2003 | 64,59 | 15,89833333 | 2003 | 10,342 | 1,282259259 |
| 2004 | 66,553 | 13,93533333 | 2004 | 10,14 | 1,080259259 |
| 2005 | 69,293 | 11,19533333 | 2005 | 8,031 | 1,028740741 |
| 2006 | 72,372 | 8,116333333 | 2006 | 5,912 | 3,147740741 |
| 2007 | 77,253 | 3,235333333 | 2007 | 4,592 | 4,467740741 |
| 2008 | 85,447 | 4,958666667 | 2008 | 5,455 | 3,604740741 |
| 2009 | 85,617 | 5,128666667 | 2009 | 13,549 | 4,489259259 |
| 2010 | 87,963 | 7,474666667 | 2010 | 16,707 | 7,647259259 |
| 2011 | 92,432 | 11,94366667 | 2011 | 12,325 | 3,265259259 |
| 2012 | 96,332 | 15,84366667 | 2012 | 10,023 | 0,963259259 |
| 2013 | 99,459 | 18,97066667 | 2013 | 8,628 | 0,431740741 |
| 2014 | 99,933 | 19,44466667 | 2014 | 7,351 | 1,708740741 |
| 2015 | 100 | 19,51166667 | 2015 | 6,185 | 2,874740741 |
| 2016 | 100,8 | 20,31166667 | 2016 | 6,758 | 2,301740741 |
| 2017 | 104,48 | 23,99166667 | 2017 | 5,763 | 3,296740741 |
| 2018 | 108,045 | 27,55666667 | 2018 | 5,371 | 3,688740741 |
| 2019 | 110,496 | 30,00766667 | 2019 | 4,448 | 4,611740741 |
| 2020 | 109,795 | 29,30666667 | 2020 | 6,806 | 2,253740741 |
| 2021 | 114,722 | 34,23366667 | 2021 | 6,181 | 2,878740741 |

Таблица 6 – Подготовка данных для критерия среднего абсолютного отклонения (“Уровень безработицы, % от общей рабочей силы”)

|  |  |  |  |
| --- | --- | --- | --- |
| Год | Уровень безработицы, % от общей рабочей силы | |  | | --- | |  | |
| 1995 | 1,437 | 0,084703704 |
| 1996 | 1,416 | 0,063703704 |
| 1997 | 1,4 | 0,047703704 |
| 1998 | 1,386 | 0,033703704 |
| 1999 | 1,39 | 0,037703704 |
| 2000 | 1,397 | 0,044703704 |
| 2001 | 1,388 | 0,035703704 |
| 2002 | 1,379 | 0,026703704 |
| 2003 | 1,371 | 0,018703704 |
| 2004 | 1,363 | 0,010703704 |
| 2005 | 1,355 | 0,002703704 |
| 2006 | 1,347 | 0,005296296 |
| 2007 | 1,341 | 0,011296296 |
| 2008 | 1,337 | 0,015296296 |
| 2009 | 1,335 | 0,017296296 |
| 2010 | 1,331 | 0,021296296 |
| 2011 | 1,327 | 0,025296296 |
| 2012 | 1,323 | 0,029296296 |
| 2013 | 1,318 | 0,034296296 |
| 2014 | 1,315 | 0,037296296 |
| 2015 | 1,315 | 0,037296296 |
| 2016 | 1,316 | 0,036296296 |
| 2017 | 1,317 | 0,035296296 |
| 2018 | 1,322 | 0,030296296 |
| 2019 | 1,327 | 0,025296296 |
| 2020 | 1,329 | 0,023296296 |
| 2021 | 1,33 | 0,022296296 |

Таблица 7 – Данные для критерия среднего абсолютного отклонения

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Показатель | ВВП | Инвестиции | Инфляция | Население | Безработица |
| |  | | --- | |  | | 234,7944444 | 98,4961481 | 537,36867 | 72,59125926 | 0,813481481 |
| |  | | --- | |  | | 9,030555556 | 3,78831339 | 20,668026 | 2,79197151 | 0,031287749 |
| Sn | 9,92584922 | 4,67166933 | 22,618252 | 3,255603141 | 0,035237453 |
| |  | | --- | |  | | 0,111917246 | 0,01302758 | 0,1158919 | 0,059705074 | 0,090027221 |
| |  | | --- | |  | | 0,078446454 | 0,07844645 | 0,0784465 | 0,078446454 | 0,078446454 |

.

Для теоретического нормального распределения отношение САО к стандартному отклонению равно . Для выборки, имеющей приближенно нормальный закон распределения, должно выполняться соотношение

1. ВВП: 0,112 > 0,078, значит показатель не проходит проверку на нормальность распределения по критерию САО
2. Общие инвестиции: 0,013 < 0,078 – показатель проходит проверку на нормальность распределения по критерию САО
3. Инфляция: 0,115 > 0,078 - значит показатель не проходит проверку на нормальность распределения по критерию САО
4. Численность населения: 0,059 < 0,078 - показатель проходит проверку на нормальность распределения по критерию САО
5. Уровень безработицы: 0,09 > 0,078 - значит показатель не проходит проверку на нормальность распределения по критерию САО

**Проверка по размаху варьирования**

Продемонстрируем расчет на примере показателя «Валовой внутренний продукт, текущие цены (млрд. $)».

Рассчитывается критериальное отношение , которое сопоставляется с критическими значениями верхней и нижней границ для различных уровней значимости. Если рассчитанное отношение лежит в пределах границ, гипотеза о нормальности распределения принимается, если же оно меньше нижней границы или больше верхней – гипотеза отвергается.

1. Валовой внутренний продукт, текущие цены (млрд. $)

Критериальное значение составило 3,26, что ниже допустимых границ при вероятности ошибки в 5 %. Следовательно гипотеза о нормальности распределения по размаху варьирования отвергается.

1. Общие инвестиции, % от ВВП

Критериальное значение составило 4,07, что входит в допустимые границы при вероятности ошибки в 5 %. Следовательно гипотеза о нормальности распределения по размаху варьирования принимается.

1. Инфляция средние потребительские цены (индекс)

Критериальное значение составило 3,38, что входит в допустимые границы при вероятности ошибки в 5 %. Следовательно гипотеза о нормальности распределения по размаху варьирования принимается.

1. Численность населения млн, человек

Критериальное значение составило 3,76, что входит в допустимые границы при вероятности ошибки в 5 %. Следовательно гипотеза о нормальности распределения по размаху варьирования принимается.

1. Уровень безработицы, % от общей рабочей силы

Критериальное значение составило 3,46, что входит в допустимые границы при вероятности ошибки в 5 %. Следовательно гипотеза о нормальности распределения по размаху варьирования принимается.

**Проверка с помощью показателей асимметрии и эксцесса**

Продемонстрируем расчет на примере показателя «Валовой внутренний продукт, текущие цены (млрд. $)».

Величины асимметрии и эксцесса для выборки могут быть рассчитаны с помощью следующих формул

Несмещенные оценки G1 и G2 для g1 и g2 соответственно рассчитываются как

а их среднеквадратические отклонения и

Гипотеза нормальности исследуемого распределения может быть принята, если совместно выполняются условия и . В противном случае гипотеза должна быть отвергнута.

Таблица 8 – Показатели для проверки с помощью показателей асимметрии и эксцесса

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Показатель | ВВП | Инвестиции | Инфляция | Население | Безработица |
| |  | | --- | |  | | -0,03736 | 1,22028815 | 1,1472881 | -1,137983 | -1,000006511 |
|  | -1,30074 | -1,3535876 | -1,615696 | -1,645303 | -1,999982648 |
| |  | | --- | |  | | -0,0396 | 1,29327517 | 1,2159089 | -1,206047 | -1,059818204 |
| |  | | --- | |  | | -1,31823 | -1,3823529 | -1,700377 | -1,736301 | -2,166645613 |
| |  | | --- | |  | | 0,447852 | 0,44785202 | 0,447852 | 0,447852 | 0,447852016 |
|  | 0,890049 | 0,89004916 | 0,8900492 | 0,8900492 | 0,890049156 |

В анализируемых рядах данных условие выполнено, следовательно гипотеза о нормальности распределения подтверждается для всех показателей.

и

и

и

и

и

Таблица 9 – Допустимость гипотезы о нормальности распределения

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Показатель | Валовой внутренний продукт, текущие цены (млрд. $) | Общие инвестиции, % от ВВП | Инфляция, % на конец периода | Численность населения млн, человек | Уровень безработицы, % от общей рабочей силы |
| Проверка по коэффициенту вариации | Недопустимо | Допустимо | Допустимо | Недопустимо | Допустимо |
| Критерий среднего абсолютного отклонения | Недопустимо | Допустимо | Недопустимо | Допустимо | Недопустимо |
| Проверка по размаху варьирования | Недопустимо | Допустимо | Допустимо | Допустимо | Допустимо |
| Проверка с помощью показателей асимметрии и эксцесса | Допустимо | Допустимо | Допустимо | Допустимо | Допустимо |
| Q-Q график | Допустимо | Допустимо | Допустимо | Допустимо | Допустимо |

**Таким образом можно сделать вывод о нормальности распределения показателей «Общие инвестиции, % от ВВП», «Инфляция, % на конец периода» и «Уровень безработицы, % от общей рабочей силы».**

Показатели отсеивались следующим образом. На первом этапе смотрелся показатель вариации. Если он превышал 0,33 – показатель автоматически исключался. Далее осуществлялась проверка с помощью 3 методов. В случае, если 2 из них показывали неудовлетворительный результат – показатель исключался. И третий этап – проверка по Q-Q графику. Наглядные выводы по гистограммам не учитывались, т.к. гистограммы способны правильно преподносить данные, только если количество элементов выборки превышает 50.

# **2 Изучение взаимосвязи данных**

При изучении взаимосвязи двумерных данных, следует придерживаться трех основных целей:

1. Описание и понимание взаимосвязи (При изучении сложной системы важно знать, какие факторы наиболее тесно взаимодействуют друг с другом, а какие вообще не оказывают влияния друг на друга);
2. Прогнозирование и предсказание нового наблюдения (Понимание некоторой взаимосвязи может позволить использовать информацию об одном из измерений для более качественного предсказания другого измерения);
3. Регулирование и управление процессом (Когда требуется вмешаться в какой-либо процесс (например, регулирование уровня производства, вводя некоторые технологические изменения или новый тип обслуживания), необходимо определить объем этого вмешательства. Если существует непосредственная взаимосвязь между вмешательством и результатом и имеется понимание данной взаимосвязи, то такое знание может помочь выполнить оптимальное регулирование).

Существуют два базовых инструмента, с помощью которых анализируют двумерные данные: корреляционный анализ, позволяющий оценить степень взаимосвязи между двумя факторами (если такая взаимосвязь вообще существует), и регрессионный анализ, показывающий, как можно предсказать или управлять одной из двух переменных с помощью другой.

Когда приходится иметь дело с двумерными данными, следует нарисовать **диаграмму рассеяния**, которая позволяет увидеть структуру. Так же как гистограмма отображает структуру одномерных данных (нормальное распределение, асимметрия, выбросы и т.д.), диаграмма рассеяния показывает все, что происходит с двумерными данными. Если данные содержат какие-то проблемы (например, выбросы или какие-либо неожиданные особенности), зачастую единственный способ их обнаружения состоит как раз в анализе соответствующей диаграммы рассеяния.

Диаграмма рассеяния представляет каждое наблюдение в пространстве двух измерений, соответствующих двум факторам. Диаграмма рассеяния позволяет определить тип взаимосвязи данных: линейный, нелинейный и отсутствие связи.

Возьмем за ключевой показатель ВВП и определим типы взаимосвязи (рис. 4).

Рисунок 4 – Диаграммы рассеяния ВВП с иными показателями

Линейная взаимосвязь играет такую же особую роль для двумерных данных, как и нормальное распределение для одномерных данных. Линейная взаимосвязь проявляется в двумерной совокупности данных, если точки на диаграмме рассеяния случайным образом концентрируются вокруг прямой линии.

Взаимосвязь в двумерной совокупности данных полностью отсутствует, если соответствующая диаграмма рассеяния точек носит совершенно случайный характер.

В двумерной совокупности данных присутствует нелинейная взаимосвязь, если точки на диаграмме рассеяния группируются вокруг некоторой кривой, а не вокруг прямой линии. Поскольку разновидностей такого рода кривых может быть чрезвычайно много, анализ такой взаимосвязи существенно сложнее. Применительно к некоторым задачам бывает полезно преобразовать одну или обе переменные таким образом, чтобы получить между ними линейную взаимосвязь.

**Таким образом мы можем сделать вывод, что связь между показателями ВВП и численностью населения – линейная, между ВВП и инфляцией – линейная, между ВВП и инвестициями –связь отсутствует и связь отсутствует между данными показателей ВВП и уровнем безработицы.**

**Коэффициент корреляции** ( r ), является числом в диапазоне от -1 до 1, характеризующим силу взаимосвязи в данных. Корреляция, равная 1, указывает на идеальную взаимосвязь в виде прямой линии, причем более высокие значения одной переменной соответствуют идеально предсказуемым более высоким значениям другой переменной. Корреляция -1 указывает на идеальную отрицательную взаимосвязь в виде прямой линии, причем одна переменная уменьшается с ростом другой.

Обычная интерпретация промежуточных корреляций в диапазоне от -1 до 1 заключается в том, что величина корреляции указывает на «силу» взаимосвязи. А знак (положительный или отрицательный) указывает направление (увеличение или уменьшение).

**При линейной связи необходимо находить коэффициент корреляции Пирсона, для остальных типов – коэффициент корреляции Кендалла или Спирмена.** Именно поэтому необходимо на первом этапе строить диаграмму рассеяния и лишь с опорой на нее находить корреляцию, т.к. зачастую, видя коэффициент корреляции более 0,7 по модулю, делается вывод о линейной связи (исходя из определения корреляции), однако в действительности линейная связь отсутствует или выбросы существенно искажают выводы.

В таблицах 10-13 представлены расчеты коэффициентов корреляции между ВВП и иными показателями.

Таблица 10 – Расчет коэффициента корреляции Пирсона для показателей ВВП и численность населения

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Год | ВВП, млрд. долл. | Население, млн. человек | |  | | --- | |  | | |  | | --- | |  | | |  | | --- | |  | | |  | | --- | |  | | |  | | --- | |  | |
| 1995 | 3,909 | 1,437 | -14,0319 | 0,084704 | -1,188549822 | 196,8929 | 0,007175 |
| 1996 | 4,785 | 1,416 | -13,1559 | 0,063704 | -0,838076488 | 173,0764 | 0,004058 |
| 1997 | 5,16 | 1,4 | -12,7809 | 0,047704 | -0,60969397 | 163,3502 | 0,002276 |
| 1998 | 5,678 | 1,386 | -12,2629 | 0,033704 | -0,413303525 | 150,3775 | 0,001136 |
| 1999 | 5,772 | 1,39 | -12,1689 | 0,037704 | -0,458810785 | 148,081 | 0,001422 |
| 2000 | 5,707 | 1,397 | -12,2339 | 0,044704 | -0,546898488 | 149,6671 | 0,001998 |
| 2001 | 6,269 | 1,388 | -11,6719 | 0,035704 | -0,41672834 | 136,2321 | 0,001275 |
| 2002 | 7,391 | 1,379 | -10,5499 | 0,026704 | -0,281720118 | 111,2994 | 0,000713 |
| 2003 | 9,891 | 1,371 | -8,04985 | 0,018704 | -0,150562044 | 64,80011 | 0,00035 |
| 2004 | 12,158 | 1,363 | -5,78285 | 0,010704 | -0,061897933 | 33,44138 | 0,000115 |
| 2005 | 14,126 | 1,355 | -3,81485 | 0,002704 | -0,010314229 | 14,55309 | 7,31E-06 |
| 2006 | 17,047 | 1,347 | -0,89385 | -0,0053 | 0,004734104 | 0,798971 | 2,81E-05 |
| 2007 | 22,476 | 1,341 | 4,535148 | -0,0113 | -0,051230377 | 20,56757 | 0,000128 |
| 2008 | 24,433 | 1,337 | 6,492148 | -0,0153 | -0,099305822 | 42,14799 | 0,000234 |
| 2009 | 19,693 | 1,335 | 1,752148 | -0,0173 | -0,030305674 | 3,070023 | 0,000299 |
| 2010 | 19,568 | 1,331 | 1,627148 | -0,0213 | -0,034652229 | 2,647611 | 0,000454 |
| 2011 | 23,21 | 1,327 | 5,269148 | -0,0253 | -0,133289933 | 27,76392 | 0,00064 |
| 2012 | 23,034 | 1,323 | 5,093148 | -0,0293 | -0,149210377 | 25,94016 | 0,000858 |
| 2013 | 25,116 | 1,318 | 7,175148 | -0,0343 | -0,246081007 | 51,48275 | 0,001176 |
| 2014 | 26,641 | 1,315 | 8,700148 | -0,0373 | -0,324483303 | 75,69258 | 0,001391 |
| 2015 | 22,893 | 1,315 | 4,952148 | -0,0373 | -0,184696785 | 24,52377 | 0,001391 |
| 2016 | 24,066 | 1,316 | 6,125148 | -0,0363 | -0,222320192 | 37,51744 | 0,001317 |
| 2017 | 26,915 | 1,317 | 8,974148 | -0,0353 | -0,316754192 | 80,53533 | 0,001246 |
| 2018 | 30,503 | 1,322 | 12,56215 | -0,0303 | -0,380586562 | 157,8076 | 0,000918 |
| 2019 | 31,049 | 1,327 | 13,10815 | -0,0253 | -0,331587599 | 171,8235 | 0,00064 |
| 2020 | 30,626 | 1,329 | 12,68515 | -0,0233 | -0,29551697 | 160,913 | 0,000543 |
| 2021 | 36,287 | 1,33 | 18,34615 | -0,0223 | -0,409051155 | 336,5812 | 0,000497 |
|  |  |  |  |  |  |  |  |
|  | ср. знач | ср.знач |  | Сумма | -8,180893815 | 2561,585 | 0,032284 |
|  | 17,94085185 | 1,352296296 |  |  |  |  |  |

Таблица 11 – Расчет коэффициента корреляции Пирсона для показателей ВВП и уровня инфляции

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Год | ВВП, млрд. долл. | Инфляция, % | |  | | --- | |  | | |  | | --- | |  | | |  | | --- | |  | | |  | | --- | |  | | |  | | --- | |  | |
| 1995 | 3,909 | 38,163 | -14,0319 | -42,3253 | 593,9028069 | 196,8929 | 1791,434 |
| 1996 | 4,785 | 45,697 | -13,1559 | -34,7913 | 457,7096271 | 173,0764 | 1210,437 |
| 1997 | 5,16 | 49,952 | -12,7809 | -30,5363 | 390,2803524 | 163,3502 | 932,4677 |
| 1998 | 5,678 | 54,338 | -12,2629 | -26,1503 | 320,6776635 | 150,3775 | 683,8399 |
| 1999 | 5,772 | 56,022 | -12,1689 | -24,4663 | 297,7271857 | 148,081 | 598,6015 |
| 2000 | 5,707 | 58,224 | -12,2339 | -22,2643 | 272,3785556 | 149,6671 | 495,7005 |
| 2001 | 6,269 | 61,5 | -11,6719 | -18,9883 | 221,6290136 | 136,2321 | 360,5568 |
| 2002 | 7,391 | 63,707 | -10,5499 | -16,7813 | 177,0405805 | 111,2994 | 281,6131 |
| 2003 | 9,891 | 64,59 | -8,04985 | -15,8983 | 127,979228 | 64,80011 | 252,757 |
| 2004 | 12,158 | 66,553 | -5,78285 | -13,9353 | 80,58596817 | 33,44138 | 194,1935 |
| 2005 | 14,126 | 69,293 | -3,81485 | -11,1953 | 42,7085381 | 14,55309 | 125,3355 |
| 2006 | 17,047 | 72,372 | -0,89385 | -8,11633 | 7,25479958 | 0,798971 | 65,87487 |
| 2007 | 22,476 | 77,253 | 4,535148 | -3,23533 | -14,67271598 | 20,56757 | 10,46738 |
| 2008 | 24,433 | 85,447 | 6,492148 | 4,958667 | 32,19239862 | 42,14799 | 24,58838 |
| 2009 | 19,693 | 85,617 | 1,752148 | 5,128667 | 8,986183802 | 3,070023 | 26,30322 |
| 2010 | 19,568 | 87,963 | 1,627148 | 7,474667 | 12,16239002 | 2,647611 | 55,87064 |
| 2011 | 23,21 | 92,432 | 5,269148 | 11,94367 | 62,9329491 | 27,76392 | 142,6512 |
| 2012 | 23,034 | 96,332 | 5,093148 | 15,84367 | 80,69414154 | 25,94016 | 251,0218 |
| 2013 | 25,116 | 99,459 | 7,175148 | 18,97067 | 136,1173438 | 51,48275 | 359,8862 |
| 2014 | 26,641 | 99,933 | 8,700148 | 19,44467 | 169,1714807 | 75,69258 | 378,0951 |
| 2015 | 22,893 | 100 | 4,952148 | 19,51167 | 96,62466395 | 24,52377 | 380,7051 |
| 2016 | 24,066 | 100,8 | 6,125148 | 20,31167 | 124,4119675 | 37,51744 | 412,5638 |
| 2017 | 26,915 | 104,48 | 8,974148 | 23,99167 | 215,304771 | 80,53533 | 575,6001 |
| 2018 | 30,503 | 108,045 | 12,56215 | 27,55667 | 346,1709291 | 157,8076 | 759,3699 |
| 2019 | 31,049 | 110,496 | 13,10815 | 30,00767 | 393,3449402 | 171,8235 | 900,4601 |
| 2020 | 30,626 | 109,795 | 12,68515 | 29,30667 | 371,7594084 | 160,913 | 858,8807 |
| 2021 | 36,287 | 114,722 | 18,34615 | 34,23367 | 628,0559203 | 336,5812 | 1171,944 |
|  |  |  |  |  |  |  |  |
|  | ср. знач | ср.знач |  | Сумма | 5653,131091 | 2561,585 | 13301,22 |
|  | 17,94085185 | 80,48833333 |  |  |  |  |  |

Таблица 12 – Расчет коэффициента корреляции Спирмена между показателями ВВП и суммарные инвестиции

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Год | ВВП, млрд. долл. | Суммарные инвестиции, % | Px | Py | |  | | --- | |  | | |  | | --- | |  | |
| 1995 | 3,909 | 26,408 | 1 | 9 | -8 | 64 |
| 1996 | 4,785 | 27,487 | 2 | 13 | -11 | 121 |
| 1997 | 5,16 | 30,726 | 3 | 19 | -16 | 256 |
| 1998 | 5,678 | 31,623 | 4 | 20 | -16 | 256 |
| 1999 | 5,772 | 25,904 | 6 | 6 | 0 | 0 |
| 2000 | 5,707 | 28,475 | 5 | 14 | -9 | 81 |
| 2001 | 6,269 | 29,58 | 7 | 16 | -9 | 81 |
| 2002 | 7,391 | 32,941 | 8 | 22 | -14 | 196 |
| 2003 | 9,891 | 35,198 | 9 | 25 | -16 | 256 |
| 2004 | 12,158 | 34,655 | 10 | 24 | -14 | 196 |
| 2005 | 14,126 | 33,356 | 11 | 23 | -12 | 144 |
| 2006 | 17,047 | 39,832 | 12 | 26 | -14 | 196 |
| 2007 | 22,476 | 40,008 | 15 | 27 | -12 | 144 |
| 2008 | 24,433 | 31,634 | 20 | 21 | -1 | 1 |
| 2009 | 19,693 | 21,01 | 14 | 1 | 13 | 169 |
| 2010 | 19,568 | 21,692 | 13 | 2 | 11 | 121 |
| 2011 | 23,21 | 25,679 | 18 | 5 | 13 | 169 |
| 2012 | 23,034 | 29,361 | 17 | 15 | 2 | 4 |
| 2013 | 25,116 | 27,198 | 21 | 12 | 9 | 81 |
| 2014 | 26,641 | 27,099 | 22 | 11 | 11 | 121 |
| 2015 | 22,893 | 25,001 | 16 | 3 | 13 | 169 |
| 2016 | 24,066 | 25,119 | 19 | 4 | 15 | 225 |
| 2017 | 26,915 | 26,382 | 23 | 8 | 15 | 225 |
| 2018 | 30,503 | 26,913 | 24 | 10 | 14 | 196 |
| 2019 | 31,049 | 26,083 | 26 | 7 | 19 | 361 |
| 2020 | 30,626 | 30,244 | 25 | 17 | 8 | 64 |
| 2021 | 36,287 | 30,524 | 27 | 18 | 9 | 81 |
|  |  |  |  |  |  |  |
|  |  |  |  |  | Сумма | 3978 |

Таблица 13 – Расчет коэффициента корреляции Спирмена между показателями ВВП и суммарные инвестиции

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Год | ВВП, млрд. долл. | Уровень безработицы, % | Px | Py | |  | | --- | |  | | |  | | --- | |  | |
| 1995 | 3,909 | 9,645 | 1 | 14 | -13 | 169 |
| 1996 | 4,785 | 9,881 | 2 | 17 | -15 | 225 |
| 1997 | 5,16 | 9,65 | 3 | 15 | -12 | 144 |
| 1998 | 5,678 | 9,831 | 4 | 16 | -12 | 144 |
| 1999 | 5,772 | 12,201 | 6 | 22 | -16 | 256 |
| 2000 | 5,707 | 14,602 | 5 | 26 | -21 | 441 |
| 2001 | 6,269 | 13,009 | 7 | 24 | -17 | 289 |
| 2002 | 7,391 | 11,227 | 8 | 21 | -13 | 169 |
| 2003 | 9,891 | 10,342 | 9 | 20 | -11 | 121 |
| 2004 | 12,158 | 10,14 | 10 | 19 | -9 | 81 |
| 2005 | 14,126 | 8,031 | 11 | 12 | -1 | 1 |
| 2006 | 17,047 | 5,912 | 12 | 6 | 6 | 36 |
| 2007 | 22,476 | 4,592 | 15 | 2 | 13 | 169 |
| 2008 | 24,433 | 5,455 | 20 | 4 | 16 | 256 |
| 2009 | 19,693 | 13,549 | 14 | 25 | -11 | 121 |
| 2010 | 19,568 | 16,707 | 13 | 27 | -14 | 196 |
| 2011 | 23,21 | 12,325 | 18 | 23 | -5 | 25 |
| 2012 | 23,034 | 10,023 | 17 | 18 | -1 | 1 |
| 2013 | 25,116 | 8,628 | 21 | 13 | 8 | 64 |
| 2014 | 26,641 | 7,351 | 22 | 11 | 11 | 121 |
| 2015 | 22,893 | 6,185 | 16 | 8 | 8 | 64 |
| 2016 | 24,066 | 6,758 | 19 | 9 | 10 | 100 |
| 2017 | 26,915 | 5,763 | 23 | 5 | 18 | 324 |
| 2018 | 30,503 | 5,371 | 24 | 3 | 21 | 441 |
| 2019 | 31,049 | 4,448 | 26 | 1 | 25 | 625 |
| 2020 | 30,626 | 6,806 | 25 | 10 | 15 | 225 |
| 2021 | 36,287 | 6,181 | 27 | 7 | 20 | 400 |
|  |  |  |  |  |  |  |
|  |  |  |  |  | Сумма | 5208 |

Расчет коэффициента корреляции Пирсона для показателей ВВП и численность населения

Расчет коэффициента корреляции Пирсона для показателей ВВП и уровень инфляции

Расчет коэффициента корреляции Спирмена между показателями ВВП и суммарные инвестиции

Расчет коэффициента корреляции Спирмена между показателями ВВП и уровень безработицы

# **3 Множественный регрессионный анализ**

Прогнозирование единственной переменной Y на основании двух или нескольких переменных X называется множественной регрессией. Прогнозирование единственной переменной Y на основании единственной переменной X называется простой регрессией. Пользуясь множественной регрессией, преследуются те же цели, что и в случае простой регрессии.

**Первым этапом построения множественной регрессионной модели является проверка на мультиколлинеарность.**

Одна из проблем множественной регрессии – это проблема мультиколлинеарности, возникающая тогда, когда некоторые из объясняющих переменных X оказываются слишком схожими. Решением данной проблемы является удаление из уравнения некоторых из переменных. (Например, если 2 из объясняющих X переменных имеют высокий уровень корреляции между собой, то в уравнении целесообразно оставить ту из них, которая сильнее коррелирует с показателем Y).

Это означает, что ни один из объясняющих показателей не должен сильно коррелировать с каким-либо другим. В частности, ни одна из предикторных переменных не должна быть линейной комбинацией других; иными словами, нельзя включать в качестве предикторов переменные А, В и А+В в одну модель. Модели, содержащие сильно скоррелированные предикторы, обычно нестабильны, то есть добавление или удаление одной переменной из модели может кардинально поменять коэффициенты и значимость остальных предикторов.

В таблице 11 представлена корреляционная матрица по исходным данным с учетом преобразований.

Анализ на мультиколлинеарность показал, что среди переменных Х есть 2 показателя, которые сильно коррелируют между собой, а именно ВВП и Инфляция. Так как инфляция коррелирует с основным показателем больше чем ВВП, оставляем ее и убираем ВВП.

Таблица 14 – Корреляционная матрица исходных данных

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *ВВП* | Сумм. инвестиции | Инфляция | Ур. безработицы | Население |
| ВВП | 1 |  |  |  |  |
| Сумм. инвестиции | -0,2027978 | 1 |  |  |  |
| Инфляция | 0,96725868 | -0,30855624 | 1 |  |  |
| Ур. безработицы | -0,5612626 | -0,35033548 | -0,4432560 | 1 |  |
| Население | -0,9252412 | 0,186904925 | -0,9496057 | 0,408265933 | 1 |

**Статистический вывод начинается с общей проверки, которую называют F-тестом.** Цель F-теста заключается в том, чтобы выяснить, объясняют ли Х-переменные значимую долю вариации Y. Если регрессия не является значимой, говорить больше не о чем. Если же регрессия оказывается значимой, можно продолжить анализ статистических выводов.

Нулевая гипотеза для F-теста утверждает, что в генеральной совокупности между Х-переменными и Y прогнозирующая связь отсутствует. Иначе говоря, Y является чисто случайной величиной и значения Х-переменных не оказывают на Y никакого влияния. Это имеет смысл быть, в том случае, если все коэффициенты регрессии в генеральной совокупности равны 0.

Альтернативная гипотеза утверждает, что в генеральной совокупности между Х-переменными и Y существует определенная прогнозирующая взаимосвязь. Таким образом, переменная Y уже не является чисто случайной величиной и должна зависеть, по крайней мере, от одной из Х-переменных.

Чтобы выполнить F-тест, необходимо обратить внимание на p-значение. Если оно больше, чем 0,05, то полученный результат не является значимым. Если же это значение меньше, чем 0,05, то полученный результат является значимым. Если p<0,01, тогда полученный результат является высоко значимым, и т.д.

По данным регрессионного анализа (рисунок 5) делаем вывод, что F-значимость <0,01 , следовательно можем сделать вывод, что среди предикторных переменных Х есть как минимум одна, объясняющая изменчивость показателя «Численность населения».

Т.к. F-тест является значимым, то известно, что одна или несколько Х-переменных могут быть полезны в прогнозировании Y и, следовательно, можно продолжать анализ с помощью t-тестов для отдельных коэффициентов регрессии с целью выяснить, какие именно из Х-переменных действительно полезны. Эти t-тесты определяют, оказывает ли значимое влияние на Y та или иная Х-переменная, если все другие Х-переменные остаются при этом неизменными. Следует помнить, что, приняв нулевую гипотезу, делается слабое заключение и, по сути, тем самым не доказана бесполезность Х-переменной, а просто не хватило убедительных доказательств наличия взаимосвязи. Таким образом, взаимосвязь может существовать, но вследствие действия фактора случайности или из-за небольшого размера выборки мы не в состоянии обнаружить ее с помощью тех данных, которые имеются в нашем распоряжении.

Достаточно простой, приблизительный способ определения значимости коэффициентов: значимыми будут те коэффициенты регрессии, для которых t-статистика по абсолютному значению равна или больше 2, поскольку для достаточно больших n и уровня значимости 5% значения из таблицы приблизительно равно 2.

По результатам анализа рисунка 5 можно сделать вывод, что показатель «Суммарные инвестиции, %» и «Уровень безработицы, %» следует исключить из модели множественной регрессии, т.к. t-статистика по модулю меньше 2, а р-значение больше 0,05.

После исключения показателя проведем регрессию еще раз (рисунок 6).

F-тест имеет показатель 7,58509E-13 , что говорит о значимости полученного результата. Все объясняющие показатели Х имеют t-статистику по модулю более 2. Следовательно все показатели могут быть использованы в регрессионной модели, которая будет иметь вид:

где – Население, млн. человек;

– Инфляция, % на конец текущего периода

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Регрессионная статистика* | |  |  |  |  |
| Множественный R | 0,943680489 |  |  |  |  |
| R-квадрат | 0,890532865 |  |  |  |  |
| Нормированный R-квадрат | 0,876254543 |  |  |  |  |
| Стандартная ошибка | 0,012395645 |  |  |  |  |
| Наблюдения | 27 |  |  |  |  |
|  |  |  |  |  |  |
| Дисперсионный анализ | |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Значимость F* |
| Регрессия | 3 | 0,028749633 | 0,009583211 | 62,36957449 | 3,35179E-11 |
| Остаток | 23 | 0,003533996 | 0,000153652 |  |  |
| Итого | 26 | 0,03228363 |  |  |  |
|  |  |  |  |  |  |
|  | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* |  |
| Y-пересечение | 1,525925178 | 0,034392419 | 44,36806828 | 8,66781E-24 |  |
| Переменная X 1 | -0,001123247 | 0,000659472 | -1,703251308 | 0,101998189 |  |
| Переменная X 2 | -0,001607516 | 0,000144272 | -11,14227025 | 9,56575E-11 |  |
| Переменная X 3 | -0,001255194 | 0,001029609 | -1,219096905 | 0,235163264 |  |

Рисунок 5 – Выводы регрессионного анализа

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Регрессионная статистика* | |  |  |  |  |
| Множественный R | 0,936105557 |  |  |  |  |
| R-квадрат | 0,876293615 |  |  |  |  |
| Нормированный R-квадрат | 0,871345359 |  |  |  |  |
| Стандартная ошибка | 0,012639132 |  |  |  |  |
| Наблюдения | 27 |  |  |  |  |
|  |  |  |  |  |  |
| Дисперсионный анализ |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Значимость F* |
| Регрессия | 1 | 0,028289939 | 0,028289939 | 177,0914274 | 7,58509E-13 |
| Остаток | 25 | 0,003993691 | 0,000159748 |  |  |
| Итого | 26 | 0,03228363 |  |  |  |
|  |  |  |  |  |  |
|  | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* |  |
| Y-пересечение | 1,46967876 | 0,009149963 | 160,6212724 | 3,33226E-39 |  |
| Переменная X 1 | 0,00145 | 0,00010959 | -13,3075703 | 7,58509E-13 |  |

Рисунок 6 – Выводы повторного регрессионного анализа

Поскольку все коэффициенты регрессии могут быть выражены в разных единицах измерения, непосредственное их сравнение весьма затруднительно: небольшой коэффициент может на самом деле оказаться более важным чем большой. Стандартизированные коэффициенты регрессии позволяют решить эту проблему за счет представления коэффициентов регрессии в терминах единого множества имеющих статистический смысл единиц измерения, что позволяет по крайней мере попытаться проводить сравнение.

Стандартизированный коэффициент регрессии, который вычисляется путем умножения коэффициента регрессии на и деления полученного произведения на представляет собой ожидаемое изменение Y (в стандартизированных единицах ), вызванное увеличением на одну соответствующую стандартизированную единицу, когда все другие X-переменные остаются неизменными. Абсолютные значения стандартизированных коэффициентов регрессии можно сравнить, получая при этом некоторое представление об относительной важности соответствующих переменных. Каждый стандартизированных коэффициент регрессии измеряется в единицах стандартных отклонений Y на одно стандартное отклонение .

Стандартизированный коэффициент показателя «Уровень безработицы, % от общей рабочей силы» составляет

# **4 Анализ временных рядов**

Главная **цель анализа временных рядов** заключается в прогнозировании будущего. Для описания временных рядов требуется модель. Модель (математическая модель или процесс) представляет собой систему уравнений, которая позволяет получить некий набор искусственных данных в форме временных рядов. Одним из наиболее важных методов при анализе временных рядов в сфере бизнеса является **анализ трендов / сезонности**, представляющий собой непосредственный, интуитивный подход к оцениванию базовых компонентов временных рядов: долгосрочный тренд (тенденция), сезонности, циклическая вариация и нерегулярный компонент. Прогнозы получаются путем наложения обычных сезонных моделей на долгосрочную тенденцию.

Долгосрочный **тренд (тенденция)** указывает действительно долгосрочное поведение временного ряда – как правило, в виде прямой линии или экспоненциальной кривой. Это бывает полезно в случае, если требуется увидеть картину в целом.

Точно повторяющийся **сезонный компонент** определяет влияние времени года. Каждый период времени в течение года характеризуется своим сезонным индексом, который свидетельствует о том, насколько выше или ниже соответствующий показатель в данный период времени по сравнению с другими периодами.

Среднесрочный **циклический компонент** состоит из последовательных повышений и понижений, которые не повторяются каждый год и поэтому исключаются из сезонного компонента. Поскольку эти повышения и понижения чередуются, их нельзя считать достаточно случайными и рассматривать как часть независимой случайной ошибки (нерегулярная компонента). Циклическую вариацию особенно трудно прогнозировать за пределами ближайшего будущего. Тем не менее она может быть очень важна, поскольку основные явления экономического цикла (такие как экономический спад) рассматриваются как часть циклической вариации в экономических показателях.

Краткосрочный **нерегулярный (случайный) компонент** представляет остаточную вариацию, которую невозможно объяснить. В нем проявляется действие тех однократных событий, которые происходят с течением времени случайно, а не систематически.

Эти четыре компонента временного ряда можно оценивать различными способами, одним из которых является **отношение к скользящему среднему**.

Скользящее среднее используется для устранения сезонных эффектов путем усреднения по всему году. Для уменьшения нерегулярного компонента и получения комбинации тренда и циклического компонента.

Деление исходного ряда на сглаженный ряд скользящего среднего дает отношение к скользящему среднему, которое включает как сезонные, так и нерегулярные значения. Выполняя группирование по времени года, а затем усреднение в полученных группах, находим сезонный индекс для каждого времени года. Выполняя деление каждого значения ряда на соответствующий сезонный индекс для соответствующего времени года, находим значения с сезонной поправкой.

Регрессия ряда с сезонной поправкой (Y) по времени (Х) служит для оценивания долгосрочного тренда в виде прямой линии как функции от времени. Этот тренд (тенденция) не отражает сезонных колебаний и дает возможность получить прогноз с сезонной поправкой.

Прогнозирование можно выполнять с помощью сезонности тренда. Получая из уравнения регрессии прогнозируемые значения (тренд) для будущих периодов времени и затем умножая их на соответствующий сезонный индекс, получаем прогнозы, которые отражают как долгосрочную тенденция, так и сезонное поведение.

Преимуществом метода отношения к скользящему среднему является легкость вычислений и интерпретации результатов. Основной недостаток заключается в том, что соответствующая модель не является полностью определенной.

Проведем перечисленные операции с показателем «Инфляция на конец периода, %» и представим результаты в таблицах 15-17.

Таблица 15 – Представление исходных данных с поправкой на циклические колебания

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Год | Инфляция, % изменение в ценах на конец периода | Скользящее среднее | Отношение к скользящему среднему | Сезонный индекс | Инфляция с поправкой на циклические колебания |
| 1995 | 38,163 |  |  | 0,997 | 38,2778335 |
| 1996 | 45,697 |  |  | 1,003 | 45,56031904 |
| 1997 | 49,952 | 49,269875 | 1,013844667 | 1,002 | 49,85229541 |
| 1998 | 54,338 | 53,068125 | 1,023929148 | 0,998 | 54,44689379 |
| 1999 | 56,022 | 56,0775 | 0,999010298 | 0,997 | 56,19057172 |
| 2000 | 58,224 | 58,692125 | 0,992024058 | 1,003 | 58,04985045 |
| 2001 | 61,5 | 60,93425 | 1,009284598 | 1,002 | 61,37724551 |
| 2002 | 63,707 | 63,046375 | 1,010478398 | 0,998 | 63,83466934 |
| 2003 | 64,59 | 65,061625 | 0,992751103 | 0,997 | 64,78435306 |
| 2004 | 66,553 | 67,118875 | 0,991569063 | 1,003 | 66,35393819 |
| 2005 | 69,293 | 69,784875 | 0,992951553 | 1,002 | 69,15469062 |
| 2006 | 72,372 | 73,7295 | 0,981588102 | 0,998 | 72,51703407 |
| 2007 | 77,253 | 78,13175 | 0,988752972 | 0,997 | 77,48545637 |
| 2008 | 85,447 | 82,121125 | 1,040499628 | 1,003 | 85,19142572 |
| 2009 | 85,617 | 85,967375 | 0,995924326 | 1,002 | 85,44610778 |
| 2010 | 87,963 | 89,225375 | 0,985851839 | 0,998 | 88,13927856 |
| 2011 | 92,432 | 92,31625 | 1,001253842 | 0,997 | 92,71013039 |
| 2012 | 96,332 | 95,54275 | 1,0082607 | 1,003 | 96,04386839 |
| 2013 | 99,459 | 97,985 | 1,015043119 | 1,002 | 99,26047904 |
| 2014 | 99,933 | 99,4895 | 1,004457757 | 0,998 | 100,1332665 |
| 2015 | 100 | 100,675625 | 0,993289091 | 0,997 | 100,3009027 |
| 2016 | 100,8 | 102,31725 | 0,985171122 | 1,003 | 100,4985045 |
| 2017 | 104,48 | 104,64325 | 0,998439938 | 1,002 | 104,2714571 |
| 2018 | 108,045 | 107,079625 | 1,009015487 | 0,998 | 108,261523 |
| 2019 | 110,496 | 109,48425 | 1,009241055 | 0,997 | 110,8284855 |
| 2020 | 109,795 |  |  | 1,003 | 109,4666002 |
| 2021 | 114,722 |  |  | 1,002 | 114,493014 |

Таблица 16 – Определение сезонного индекса

|  |  |  |  |
| --- | --- | --- | --- |
| Первые года цикла | Вторые года цикла | Третьи года цикла | Четвертые года цикла |
|  |  | 1,013844667 | 1,023929148 |
| 0,999010298 | 0,992024058 | 1,009284598 | 1,010478398 |
| 0,992751103 | 0,991569063 | 0,992951553 | 0,981588102 |
| 0,988752972 | 1,040499628 | 0,995924326 | 0,985851839 |
| 1,001253842 | 1,0082607 | 1,015043119 | 1,004457757 |
| 0,993289091 | 0,985171122 | 0,998439938 | 1,009015487 |
| 1,009241055 |  |  |  |
| Средние значения отношений к скользящему среднему | | | |
| 0,99738306 | 1,003504914 | 1,002328707 | 0,998278317 |

Таблица 17 – Прогнозирование значений долгосрочного тренда

|  |  |
| --- | --- |
| Временные периоды, Х | Тренд и прогноз с поправкой на циклические колебания, прогнозируемые значения Y, % |
| 1 | 43,7148 |
| 2 | 46,5436 |
| 3 | 49,3724 |
| 4 | 52,2012 |
| 5 | 55,03 |
| 6 | 57,8588 |
| 7 | 60,6876 |
| 8 | 63,5164 |
| 9 | 66,3452 |
| 10 | 69,174 |
| 11 | 72,0028 |
| 12 | 74,8316 |
| 13 | 77,6604 |
| 14 | 80,4892 |
| 15 | 83,318 |
| 16 | 86,1468 |
| 17 | 88,9756 |
| 18 | 91,8044 |
| 19 | 94,6332 |
| 20 | 97,462 |
| 21 | 100,2908 |
| 22 | 103,1196 |
| 23 | 105,9484 |
| 24 | 108,7772 |
| 25 | 111,606 |
| 26 | 114,4348 |
| 27 | 117,2636 |
| 28(2022 год) | 120,0924 |
| 29(2023 год) | 122,9212 |

Рисунок 7 – Оценка тренда показателя

Для того чтобы спрогнозировать значения на 2021 и 2022 года необходимо учесть сезонность в долгосрочном тренде, вернув ему ожидаемую сезонную вариацию. Для этого умножим значение тренда на значение сезонного индекса для того периода времени, который прогнозируется. Этот процесс является обратным по отношению к внесению поправки на сезонные колебания. Результирующий прогноз включает долгосрочный тренд и сезонную вариацию.

**Таким образом прогнозное значение на 2021 год составит 21582,96 ∙ 0,995 = 21466,64 млрд. $, а в 2022 году 22129,14 ∙ 1,000 = 22127,25 млрд. $.**

**Оценим устойчивость тенденции с помощью коэффициента корреляции рангов Спирмена.**

Таблица 18 – Расчет коэффициента корреляции Спирмена по показателю «Инфляция на конец периода, %»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Год | Временные периоды | Инфляция, % | Ранг по Px | Ранг по Py | di | di^2 |
| 1995 | 1 | 38,163 | 1 | 1 | 0 | 0 |
| 1996 | 2 | 45,697 | 2 | 2 | 0 | 0 |
| 1997 | 3 | 49,952 | 3 | 3 | 0 | 0 |
| 1998 | 4 | 54,338 | 4 | 4 | 0 | 0 |
| 1999 | 5 | 56,022 | 5 | 5 | 0 | 0 |
| 2000 | 6 | 58,224 | 6 | 6 | 0 | 0 |
| 2001 | 7 | 61,5 | 7 | 7 | 0 | 0 |
| 2002 | 8 | 63,707 | 8 | 8 | 0 | 0 |
| 2003 | 9 | 64,59 | 9 | 9 | 0 | 0 |
| 2004 | 10 | 66,553 | 10 | 10 | 0 | 0 |
| 2005 | 11 | 69,293 | 11 | 11 | 0 | 0 |
| 2006 | 12 | 72,372 | 12 | 12 | 0 | 0 |
| 2007 | 13 | 77,253 | 13 | 13 | 0 | 0 |
| 2008 | 14 | 85,447 | 14 | 14 | 0 | 0 |
| 2009 | 15 | 85,617 | 15 | 15 | 0 | 0 |
| 2010 | 16 | 87,963 | 16 | 16 | 0 | 0 |
| 2011 | 17 | 92,432 | 17 | 17 | 0 | 0 |
| 2012 | 18 | 96,332 | 18 | 18 | 0 | 0 |
| 2013 | 19 | 99,459 | 19 | 19 | 0 | 0 |
| 2014 | 20 | 99,933 | 20 | 20 | 0 | 0 |
| 2015 | 21 | 100 | 21 | 21 | 0 | 0 |
| 2016 | 22 | 100,8 | 22 | 22 | 0 | 0 |
| 2017 | 23 | 104,48 | 23 | 23 | 0 | 0 |
| 2018 | 24 | 108,045 | 24 | 24 | 0 | 0 |
| 2019 | 25 | 110,496 | 25 | 26 | 0 | 0 |
| 2020 | 26 | 109,795 | 26 | 25 | 0 | 0 |
| 2021 | 27 | 114,722 | 27 | 27 | 0 | 0 |

Коэффициент Спирмена определяется по формуле

где n – количество данных по каждому показателю.

Так как рассчитанный коэффициент равен 1, тенденция изменения показателя устойчива.

**Рассчитаем показатель колеблемости показателя.**

Для определения показателей колеблемости следует рассчитать теоретические значения уровней для каждого периода по уравнению тренда. Затем по отклонениям исходных показателей от теоретических определяются поворотные точки.

Рисунок 8 – Оценка тренда показателя Инфляция

Таблица 19 – Расчет коэффициента колеблемости показателя «Инфляция на конец периода, %»

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Теоретическое значение | Исходное - Теоретическое | Поворотные точки | Абсолютное отклоненние | Квадрат абсолютного |
|
| 42,7774 | -4,6144 | 0 | 4,6144 | 21,29268736 |
| 45,6878 | 0,0092 | 0 | 0,0092 | 8,464E-05 |
| 48,5982 | 1,3538 | 0 | 1,3538 | 1,83277444 |
| 51,5086 | 2,8294 | 1 | 2,8294 | 8,00550436 |
| 54,419 | 1,603 | 0 | 1,603 | 2,569609 |
| 57,3294 | 0,8946 | 1 | 0,8946 | 0,80030916 |
| 60,2398 | 1,2602 | 1 | 1,2602 | 1,58810404 |
| 63,1502 | 0,5568 | 0 | 0,5568 | 0,31002624 |
| 66,0606 | -1,4706 | 0 | 1,4706 | 2,16266436 |
| 68,971 | -2,418 | 0 | 2,418 | 5,846724 |
| 71,8814 | -2,5884 | 1 | 2,5884 | 6,69981456 |
| 74,7918 | -2,4198 | 0 | 2,4198 | 5,85543204 |
| 77,7022 | -0,4492 | 0 | 0,4492 | 0,20178064 |
| 80,6126 | 4,8344 | 1 | 4,8344 | 23,37142336 |
| 83,523 | 2,094 | 0 | 2,094 | 4,384836 |
| 86,4334 | 1,5296 | 1 | 1,5296 | 2,33967616 |
| 89,3438 | 3,0882 | 0 | 3,0882 | 9,53697924 |
| 92,2542 | 4,0778 | 0 | 4,0778 | 16,62845284 |
| 95,1646 | 4,2944 | 1 | 4,2944 | 18,44187136 |
| 98,075 | 1,858 | 0 | 1,858 | 3,452164 |
| 100,9854 | -0,9854 | 0 | 0,9854 | 0,97101316 |
| 103,8958 | -3,0958 | 1 | 3,0958 | 9,58397764 |
| 106,8062 | -2,3262 | 0 | 2,3262 | 5,41120644 |
| 109,7166 | -1,6716 | 1 | 1,6716 | 2,79424656 |
| 112,627 | -2,131 | 0 | 2,131 | 4,541161 |
| 115,5374 | -5,7424 | 1 | 5,7424 | 32,97515776 |
| 118,4478 | -3,7258 | 0 | 3,7258 | 13,88158564 |
|  | **ИТОГ** | **10** | **63,922** | **205,479266** |

В таблице 20 представлены данные для расчета необходимого количества поворотных точек для разных типов колеблемости.

Таблица 20 – Типы колеблемости

|  |  |
| --- | --- |
| Характеристика типа колеблемости | Расчет числа поворотных точек, соответствующих данному типу колеблемости |
| Пилообразная | n-2 = 27-2 = 25 |
| Циклическая | 2∙n/l = 2∙27/4 = 13,5 |
| Случайно распределенная | 2/3\*(n-2) = 2/3\*(27-2) = 16,6 |
| *(где n-число уровней ряда и l - длина цикла)* | |

Таким образом, можем сделать вывод, что показатель «Инфляция на конец периода, %» имеет циклический тип колеблемости.

Коэффициент колеблемости рассчитывается по формуле

где – среднее квадратическое отклонение;

– среднее значение исходных данных.

Среднее квадратическое отклонение составит

Коэффициент колеблемости составит

Наблюдается низкая степень колеблемости исследуемого показателя.