Долгосрочное домашнее задание №1

по теории дискретных случайных процессов

ККСО-01-20

Семин Владислав

# ОПИСАНИЕ ДАННЫХ

В данной работе будет рассмотрена динамика изменения урожайности пшеницы в центнерах с одного гектара убранной площади с 1990 по 2020 год (всего 31 данных).

|  |  |
| --- | --- |
| год | урожай |
| 1990 | 21,0 |
| 1991 | 17,3 |
| 1992 | 19,6 |
| 1993 | 18,2 |
| 1994 | 15,3 |
| 1995 | 13,9 |
| 1996 | 15,5 |
| 1997 | 18,4 |
| 1998 | 13,5 |
| 1999 | 15,7 |
| 2000 | 16,1 |
| 2001 | 20,6 |
| 2002 | 20,7 |
| 2003 | 17,1 |
| 2004 | 19,8 |
| 2005 | 19,3 |
| 2006 | 19,5 |
| 2007 | 21,0 |
| 2008 | 24,5 |
| 2009 | 23,2 |
| 2010 | 19,1 |
| 2011 | 22,6 |
| 2012 | 17,7 |
| 2013 | 22,3 |
| 2014 | 25,0 |
| 2015 | 23,9 |
| 2016 | 26,8 |
| 2017 | 31,2 |
| 2018 | 27,2 |
| 2019 | 27,0 |
| 2020 | 29,8 |

Построим диаграмму:

На данной диаграмме можно заметить, что с 1990 по 1998 динамика была убывающей, но после этого тенденция меняется на возрастающую динамику.

# АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИЯ

Автокорреляционная функция при лаге равном определяется, как корреляция между наборами данных и , где – общее количество данных.

По АКФ можно сделать предварительный вывод о виде тренда. Возможно 3 случая.

1. Значения велики по модулю и одного знака. Как правило, уменьшаются по модулю с ростом лага, но не до малых значений. Это указывает на тренд с чёткой динамикой (возрастающей или убывающей). Предсказания по такому тренду достаточно грубы, но их можно использовать для предсказания на далёкое время.
2. С ростом лага знак коэффициента автокорреляции не меняется, но модуль резко убывает к малым значениям (на очень малых может меняется знак). Это означает динамику со стабилизацией.
3. Значения резко растут, меняя знак. Это означает смену тенденции (наличие экстремума).

Рассчитаем значения автокорреляционной функции для лагов от 1 до 10.

Среднее от до : .

Среднее от до : .

Среднее выборочное от до : .

Среднее выборочное от до :

…

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  | 0,7775 | 0,7037 | 0,6914 | 0,5788 | 0,5057 | 0,4927 | 0,4912 | 0,3797 | 0,5654 | 0,3703 |

Мы наблюдаем случай 1 с большими значениями одного знака, которые медленно уменьшаются с ростом лага. Значит, скорее всего, тренд будет иметь чёткую динамику.

# ТРЕНДОВАЯ МОДЕЛЬ

Трендом называется регрессия наблюдаемых значений от времени.

## ЛИНЕЙНЫЙ ТРЕНД

где – -ый член выборки, – временной такт, – ошибка.

Рассчитаем и по методу наименьших квадратов.

где – наша выборка, – временные такты.

где – среднее выборочное выборки, – среднее выборочное временных тактов.

Следовательно линейный тренд имеет вид

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| 21,0 | 14,8026 | 6,1974 | 38,4075 |
| 17,3 | 15,1981 | 2,1019 | 4,4178 |
| 19,6 | 15,5937 | 4,0063 | 16,0507 |
| 18,2 | 15,9892 | 2,2108 | 4,8877 |
| 15,3 | 16,3847 | -1,0847 | 1,1766 |
| 13,9 | 16,7802 | -2,8802 | 8,2958 |
| 15,5 | 17,1758 | -1,6758 | 2,8082 |
| 18,4 | 17,5713 | 0,8287 | 0,6868 |
| 13,5 | 17,9668 | -4,4668 | 19,9524 |
| 15,7 | 18,3623 | -2,6623 | 7,0880 |
| 16,1 | 18,7579 | -2,6579 | 7,0642 |
| 20,6 | 19,1534 | 1,4466 | 2,0927 |
| 20,7 | 19,5489 | 1,1511 | 1,3250 |
| 17,1 | 19,9444 | -2,8444 | 8,0908 |
| 19,8 | 20,3310 | -0,54 | 0,2916 |
| 19,3 | 20,7355 | -1,4355 | 2,0606 |
| 19,5 | 21,1310 | -1,631 | 2,6602 |
| 21,0 | 21,5265 | -0,5265 | 0,2772 |
| 24,5 | 21,9221 | 2,5779 | 6,6458 |
| 23,2 | 22,3176 | 0,8824 | 0,7787 |
| 19,1 | 22,7131 | -3,6131 | 13,0545 |
| 22,6 | 23,1086 | -0,5086 | 0,2587 |
| 17,7 | 23,5042 | -5,8042 | 33,6882 |
| 22,3 | 23,8997 | -1,5997 | 2,5590 |
| 25,0 | 24,2952 | 0,7048 | 0,4967 |
| 23,9 | 24,6908 | -0,7907 | 0,6252 |
| 26,8 | 25,0863 | 1,7138 | 2,9369 |
| 31,2 | 25,4818 | 5,7182 | 32,6981 |
| 27,2 | 25,8773 | 1,3227 | 1,7495 |
| 27,0 | 26,2728 | 0,7272 | 0,5288 |
| 29,8 | 26,6683 | 3,1317 | 9,8073 |

## КВАДРАТИЧНЫЙ ТРЕНД

Рассчитаем коэффициенты по методу наименьших квадратов.

После решения этой системы получаем: .

Следовательно квадратичный тренд имеет вид

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| 21,0 | 18,0632 | 2,9368 | 8,6250 |
| 17,3 | 17,8066 | -0,5066 | 0,2566 |
| 19,6 | 17,5950 | 2,005 | 4,0202 |
| 18,2 | 17,4283 | 0,7717 | 0,5955 |
| 15,3 | 17,3067 | -2,0067 | 4,0267 |
| 13,9 | 17,2300 | -3,33 | 11,0887 |
| 15,5 | 17,1983 | -1,6983 | 2,8841 |
| 18,4 | 17,2115 | 1,1885 | 1,4125 |
| 13,5 | 17,2697 | -3,7697 | 14,2109 |
| 15,7 | 17,3729 | -1,6729 | 2,7987 |
| 16,1 | 17,5211 | -1,4211 | 2,0195 |
| 20,6 | 17,7143 | 2,8857 | 8,3275 |
| 20,7 | 17,9524 | 2,7476 | 7,5495 |
| 17,1 | 18,2355 | -1,1355 | 1,2893 |
| 19,8 | 18,5635 | 1,2365 | 1,5289 |
| 19,3 | 18,9366 | 0,3634 | 0,1321 |
| 19,5 | 19,3546 | 0,1454 | 0,0211 |
| 21,0 | 19,8176 | 1,1824 | 1,3982 |
| 24,5 | 20,3255 | 4,1745 | 17,4263 |
| 23,2 | 20,8784 | 2,3216 | 5,3896 |
| 19,1 | 21,4763 | -2,3763 | 5,6470 |
| 22,6 | 22,1192 | 0,4808 | 0,2311 |
| 17,7 | 22,8071 | -5,1071 | 26,0822 |
| 22,3 | 23,5399 | -1,2399 | 1,5373 |
| 25,0 | 24,3177 | 0,6823 | 0,4656 |
| 23,9 | 25,1405 | -1,2405 | 1,5387 |
| 26,8 | 26,0082 | 0,7918 | 0,6270 |
| 31,2 | 26,9209 | 4,2791 | 18,3106 |
| 27,2 | 27,8786 | -0,6786 | 0,4605 |
| 27,0 | 28,8813 | -1,8813 | 3,5391 |
| 29,8 | 29,9289 | -0,1289 | 0,0166 |

Так как квадратичный тренд даёт значительный прирост коэффициента детерминации , то для дальнейшего исследования выберем его.

## ИССЛЕДОВАНИЕ АВТОКОРРЕЛЯЦИИ ОСТАТКОВ С ПОМОЩЬЮ КРИТЕРИЯ ДАРБИНА-УОТСОНА

Рассчитаем коэффициент автокорреляции остатков с лагом 1.

Среднее от до :

Среднее от до :

Среднее выборочное от до :

Среднее выборочное от до :

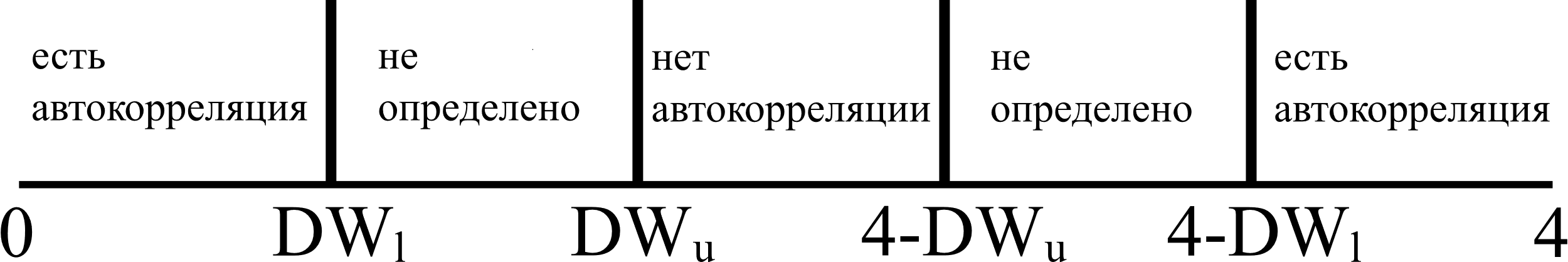
Коэффициент автокорреляции равен

Наблюдаемое значение статистики Дарбина-Уотсона равно

Так как , то .

Нижние и верхние критические значения статистики Дарбина-Уотсона ( и ) при уровне значимости для размера выборки и количества степеней свободы равны и соответственно.

Сравнив с и согласно схеме ниже, мы сможем определить наличие или отсутствие автокорреляции остатков.



В данном случае наблюдаемое значение статистики Дарбина-Уотсона больше , но меньше . Автокорреляции остатков нет.

В случае наличия автокорреляции либо выбирают другую трендовую модель, либо выполняют оценку остатков.

## ОЦЕНКА ОСТАТКОВ

В реальности при отсутствии автокорреляции остатков их оценка не производится. Но так как данная работа направлена на отработку основных методов работы с временными рядами, то мы её проведём.

Для этого нужно построить линейную регрессию вида .

Данные коэффициенты рассчитываем по методу намётших квадратов. То есть равен отношению ковариации между и и дисперсии , а равен разности среднего и произведения среднего на .

Ковариация между и равна .

Дисперсия равна 5,1147.

Среднее равно -0,0979.

Среднее равно 0,0043.

Трендовая модель после оценки остатков принимает вид

Рассчитаем ошибки оценки остатков.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| 2,9368 | -0,5066 | 0,3452 | 0,7255 |
| -0,5066 | 2,005 | -0,1751 | 4,7529 |
| 2,005 | 0,7717 | 0,2044 | 0,3218 |
| 0,7717 | -2,0067 | 0,018 | 4,0994 |
| -2,0067 | -3,33 | -0,4017 | 8,5746 |
| -3,33 | -1,6983 | -0,6017 | 1,2025 |
| -1,6983 | 1,1885 | -0,3551 | 2,3828 |
| 1,1885 | -3,7697 | 0,081 | 14,8283 |
| -3,7697 | -1,6729 | -0,6681 | 1,0097 |
| -1,6729 | -1,4211 | -0,3513 | 1,1445 |
| -1,4211 | 2,8857 | -0,3133 | 10,2336 |
| 2,8857 | 2,7476 | 0,3375 | 5,8089 |
| 2,7476 | -1,1355 | 0,3166 | 2,1085 |
| -1,1355 | 1,2365 | -0,2701 | 2,2698 |
| 1,2365 | 0,3634 | 0,0883 | 0,0757 |
| 0,3634 | 0,1454 | -0,0436 | 0,0357 |
| 0,1454 | 1,1824 | -0,0766 | 1,5851 |
| 1,1824 | 4,1745 | 0,0801 | 16,7639 |
| 4,1745 | 2,3216 | 0,5322 | 3,2019 |
| 2,3216 | -2,3763 | 0,2522 | 6,9093 |
| -2,3763 | 0,4808 | -0,4576 | 0,8805 |
| 0,4808 | -5,1071 | -0,0259 | 25,8183 |
| -5,1071 | -1,2399 | -0,8702 | 0,1367 |
| -1,2399 | 0,6823 | -0,2859 | 0,9374 |
| 0,6823 | -1,2405 | 0,0045 | 1,55 |
| -1,2405 | 0,7918 | -0,286 | 1,1616 |
| 0,7918 | 4,2791 | 0,0211 | 18,1306 |
| 4,2791 | -0,6786 | 0,548 | 1,5045 |
| -0,6786 | -1,8813 | -0,2011 | 2,823 |
| -1,8813 | -0,1289 | -0,3828 | 0,0645 |

# МОДЕЛЬ АВТОРЕГРЕССИИ

Модель авторегрессии порядка k представляет собой множественную регрессию вида

Трендовая модель делает достаточно грубые предсказания, но при этом предоставляет возможность делать их на относительно далёкое будущее. Модель авторегрессии позволяет сделать предсказание только на следующий временной такт, и после получения его истинного значения она перестраивается. Но при этом данные краткосрочные предсказания будут более точными.

Рассчитаем модель авторегрессии 3-го порядка.

,

Следовательно авторегрессия третьего порядка примет вид:

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 18,2 | 20,3161 | 4,4781 |
| 15,3 | 18,6831 | 11,4456 |
| 13,9 | 17,5246 | 13,1375 |
| 15,5 | 15,9454 | 0,1984 |
| 18,4 | 15,8334 | 6,5872 |
| 13,5 | 17,285 | 14,3261 |
| 15,7 | 15,2857 | 0,1717 |
| 16,1 | 16,8256 | 0,5264 |
| 20,6 | 15,8617 | 22,4516 |
| 20,7 | 19,1657 | 2,354 |
| 17,1 | 19,8904 | 7,7861 |
| 19,8 | 19,1667 | 0,4011 |
| 19,3 | 20,3178 | 1,0359 |
| 19,5 | 19,283 | 0,0471 |
| 21 | 20,1436 | 0,7335 |
| 24,5 | 20,8849 | 13,069 |
| 23,2 | 23,1483 | 0,0027 |
| 19,1 | 23,2711 | 17,3982 |
| 22,6 | 21,7899 | 0,6563 |
| 17,7 | 22,9243 | 27,293 |
| 22,3 | 19,2966 | 9,0202 |
| 25 | 22,4016 | 6,7517 |
| 23,9 | 23,058 | 0,709 |
| 26,8 | 24,1242 | 7,1598 |
| 31,2 | 26,471 | 22,3635 |
| 27,2 | 29,0365 | 3,3726 |
| 27 | 28,1274 | 1,2711 |
| 29,8 | 28,8385 | 0,9244 |

В данном примере получаются следующие средние квадратов ошибок при разных порядках модели авторегрессии:

|  |  |
| --- | --- |
|  | Среднее квадратов ошибок |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Как можно увидеть, авторегрессия третьего порядка даёт значительное уменьшение ошибок по сравнению с первым и вторым порядком. При этом переход на более высокие порядки особых преимуществ не имеет. Поэтому для дальнейшего рассмотрения выберем авторегрессию третьего порядка.

# СРАВНЕНИЕ МОДЕЛЕЙ

Для того, чтобы проверить точность предсказаний по данным моделям, попробуем вычислить значение для 2021 года и сравнить его с реальным значением, равным 27,2.

Для тренда без оценки остатков:

Для тренда с оценкой остатков:

Для авторегрессии:

|  |  |  |  |
| --- | --- | --- | --- |
| Модель | Формула | Среднее квадратов ошибок | Ошибка в предсказании на 2021 год |
| Тренд |  |  |  |
| Тренд с оценкой остатков |  |  |  |
| Авторегрессия |  |  |  |

**Вывод.** Оценка остатков лишь незначительно увеличила точность тренда, что было ожидаемо, так как автокорреляции ошибок у тренда и так изначально не было. Тренд в среднем даёт меньшую ошибку, чем авторегрессия. Но на короткой дистанции авторегрессия сделала более точное предсказание. Следовательно, тренд лучше использовать для долгосрочных предсказаний, а авторегрессию для краткосрочных.