Министерство науки и высшего образования

Российской Федерации

Федеральное Государственное

Автономное Образовательное Учреждение

Высшего Образования

Национальный ядерный университет «МИФИ»

Кафедра: «Финансовый мониторинг»

Отчет по Лабораторной работе №5:

Студент Монастырский М. О.

Группа С21-703

Проверила: Домашова Д. В.

Оглавление

[Введение 3](#_Toc164885994)

[Описание лабораторной работы 3](#_Toc164885995)

[Постановка задачи 4](#_Toc164885996)

[Порядок работы 4](#_Toc164885997)

[3.1 Подбор и построение модели множественной регрессии 4](#_Toc164885998)

[3.1.1 Линейная модель множественной регрессии 4](#_Toc164885999)

[Подбор нелинейной модели 10](#_Toc164886000)

[Сравнение моделей 14](#_Toc164886001)

[Вывод 16](#_Toc164886002)

# Введение

Одним из самых сложных этапов спецификации модели регрессии – параметризация, заключающаяся в выборе параметрического семейства функций , в рамках которого ищется неизвестная функция регрессии. Иногда подбор параметрического класса удается провести из соображений содержательного (экономического) характера, других соображений. Естественны попытки свести выбранную модель к линейной в целях упрощения оценки и исследование параметров модели.

В случае невозможности линеаризации модели оценка параметров модели может быть осуществлена методом наименьших квадратов, приводящего к решению нелинейной оптимизационной задачи МНК. В вычислительном плане, в настоящее время, это не представляет трудностей, но остаются проблемы с изучением статистических свойств оценок.

Большой интерес представляет построение линеаризуемой нелинейной зависимости в случаях, когда последняя неизвестна, методом Бокса-Кокса. При этом построенную нелинейную степенную зависимость можно рассматривать и как хорошую аппроксимацию достаточно широкого класса функций более общего вида. Освоение приемов подбора нелинейной регрессионной зависимости является целью предлагаемой работы.

# Описание лабораторной работы

Лабораторная работа включает следующие этапы:

- постановку задачи;

- ознакомление с порядком выполнения работы в пакете Statistica 6.0;

- выполнение расчетов для индивидуальных задач;

- подготовку письменного отчета;

- защиту лабораторной работы.

# Постановка задачи

По данным Приложения Б провести регрессионный анализ:

1 Из экономических или других соображений подобрать параметрический класс нелинейных зависимостей для модели регрессии.

2 Линеаризовать модель, оценить параметры и провести содержательный анализ.

3 Подобрать нелинейную модель, используя подход Бокса-Кокса и провести анализ модели.

# Порядок работы

# 3.1 Подбор и построение модели множественной регрессии

# 3.1.1 Линейная модель множественной регрессии

|  |  |
| --- | --- |
| **Признак** | **Значение** |
| **Y2** | Ожидаемая продолжительность жизни при рождении (лет) |
| **X3** | Потребление сахара и кондитерских изделий на душу населения в кг |
| **X4** | Потребление картофеля на душу населения в кг |
| **X5** | Потребление масло растительного и других жиров на душу населения в кг |
| **X6** | Потребление мяса и мясных продуктов на душу населения в кг |
| **X7** | Потребление овощей и бахчевых на душу населения в кг |
| **X8** | Потребление фруктов и ягод на душу населения в кг |
| **X13** | Валовой региональный продукт на душу населения (тыс рублей) |

выявим зависимость между результативным признаком и объясняющими переменными. Вначале построим линейную функцию регрессии:



Multiple Regression Results

Dependent: Y2 Multiple R = ,72347295 F = 12,08079

R?= ,52341311 df = 7,77

No. of cases: 85 adjusted R?= ,48008703 p = ,000000

Standard error of estimate: 1,708779411

Intercept: 65,870050916 Std.Error: 1,656768 t( 77) = 39,758 p = 0,0000

X3 b\*=,229 X4 b\*=-,02 X5 b\*=-,13

X6 b\*=-,17 X7 b\*=,620 X8 b\*=-,08

X13 b\*=,385

(significant b\* are highlighted in red)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N=85 | Regression Summary for Dependent Variable: Y2 (Sheet1 in данные лабы 5) R= ,72347295 R?= ,52341311 Adjusted R?= ,48008703 F(7,77)=12,081 p | | | | | |
| |  | | --- | | b\* | | |  | | --- | | Std.Err. of b\* | | |  | | --- | | b | | |  | | --- | | Std.Err. of b | | |  | | --- | | t(77) | | |  | | --- | | p-value | |
| |  | | --- | | Intercept | |  |  | 65,87005 | 1,656768 | 39,75817 | 0,000000 |
| |  | | --- | | X3 | | 0,229087 | 0,097355 | 0,10301 | 0,043777 | 2,35311 | 0,021171 |
| |  | | --- | | X4 | | -0,022009 | 0,088773 | -0,00427 | 0,017243 | -0,24793 | 0,804852 |
| |  | | --- | | X5 | | -0,127484 | 0,096749 | -0,16055 | 0,121846 | -1,31768 | 0,191517 |
| |  | | --- | | X6 | | -0,170314 | 0,090982 | -0,03024 | 0,016152 | -1,87196 | 0,065010 |
| |  | | --- | | X7 | | 0,619567 | 0,126195 | 0,07944 | 0,016181 | 4,90960 | 0,000005 |
| |  | | --- | | X8 | | -0,081283 | 0,130588 | -0,01366 | 0,021945 | -0,62244 | 0,535493 |
| |  | | --- | | X13 | | 0,384996 | 0,089305 | 0,00027 | 0,000062 | 4,31101 | 0,000048 |

Поскольку можно допустить, нормальный характер распределения регрессионных остатков

После исключения мультиколлинеарности методом включения переменных

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N=85 | Regression Summary for Dependent Variable: Y2 (Sheet1 in данные лабы 5) R= ,72154752 R?= ,52063082 Adjusted R?= ,49029100 F(5,79)=17,160 p | | | | | |
| |  | | --- | | b\* | | |  | | --- | | Std.Err. of b\* | | |  | | --- | | b | | |  | | --- | | Std.Err. of b | | |  | | --- | | t(79) | | |  | | --- | | p-value | |
| |  | | --- | | Intercept | |  |  | 65,82588 | 1,585595 | 41,51493 | 0,000000 |
| |  | | --- | | X7 | | 0,560059 | 0,088519 | 0,07181 | 0,011350 | 6,32697 | 0,000000 |
| |  | | --- | | X13 | | 0,374203 | 0,084774 | 0,00026 | 0,000059 | 4,41412 | 0,000032 |
| |  | | --- | | X6 | | -0,181604 | 0,087074 | -0,03224 | 0,015458 | -2,08562 | 0,040241 |
| |  | | --- | | X3 | | 0,220622 | 0,095530 | 0,09921 | 0,042957 | 2,30944 | 0,023532 |
| |  | | --- | | X5 | | -0,128926 | 0,092522 | -0,16237 | 0,116522 | -1,39346 | 0,167387 |

 на основании отчета о результатах регрессионного анализа, делаем вывод:

* модель регрессии значима (p-value=0,016<0,05);

|  |  |
| --- | --- |
| **Признак** | **Значение** |
| **Y2** | Ожидаемая продолжительность жизни при рождении (лет) |
| **X3** | Потребление сахара и кондитерских изделий на душу населения в кг |
| **X4** | Потребление картофеля на душу населения в кг |
| **X5** | Потребление масло растительного и других жиров на душу населения в кг |
| **X6** | Потребление мяса и мясных продуктов на душу населения в кг |
| **X7** | Потребление овощей и бахчевых на душу населения в кг |
| **X8** | Потребление фруктов и ягод на душу населения в кг |
| **X13** | Валовой региональный продукт на душу населения (тыс рублей) |

* Переменные, существенно влияющие на результативный признак – Ожидаемая продолжительность жизни (лет)

X6 Потребление мяса и мясных продуктов на душу населения в кг

X7 Потребление овощей и бахчевых на душу населения в кг

X3 Потребление сахара и кондитерских изделий на душу населения в кг

X13 Валовой региональный продукт на душу населения (тыс рублей)

* коэффициент детерминации составил 0,52.

Оценка уравнения регрессии выглядит следующим образом:

Ожидаемая продолжительность жизни снижается на 0,032 года при увеличении Потребления мяса и мясных продуктов на душу населения на 1 кг

Ожидаемая продолжительность жизни увеличивается на 0,072 года при увеличении Потребление овощей и бахчевых на душу населения на 1 кг

Ожидаемая продолжительность жизни увеличивается на 0,1 года при увеличении Потребления сахара и кондитерских изделий на душу населения в кг на 1 кг

Ожидаемая продолжительность жизни увеличивается на года при увеличении Валовой региональный продукт на душу населения (тыс рублей) на 1 тыс рублей

На основании графического анализа проверим гипотезу о наличии положительной автокорреляции (Но: автокорреляция отсутствует) и проверим ее с помощью критерия Дарбина – Уотсона



|  |  |  |
| --- | --- | --- |
|  | Durbin-Watson d (Sheet1 in данные лабы 5) and serial correlation of residuals | |
| |  | | --- | | Durbin- Watson d | | |  | | --- | | Serial Corr. | |
| |  | | --- | | Estimate | | 1,403521 | 0,284676 |

Для расчета критического значения критерия воспользуемся таблицей значений статистики Дарбина-Уотсана. Для n=87, k=3 получаем dн=1,55; dв=1,75 Значение статистики Дарбина-Уотсона для модели находится в интервале DW ∈ (0; dн ) = (0; 1,55), то гипотеза об отсутствии автокорреляции отклоняется, принимается гипотеза о наличии положительной автокореляции.

# Подбор нелинейной модели

Анализируя данные, можно предположить, что функцию можно искать в форме Кобба-Дугласа:

Модель нелинейной регрессии:

где - регрессионные остатки.

Линеаризуем модель логарифмированием:

или

где

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N=85 | Regression Summary for Dependent Variable: lnY (Sheet1 in данные лабы 5) R= ,67916755 R?= ,46126856 Adjusted R?= ,41229297 F(7,77)=9,4183 p | | | | | |
| |  | | --- | | b\* | | |  | | --- | | Std.Err. of b\* | | |  | | --- | | b | | |  | | --- | | Std.Err. of b | | |  | | --- | | t(77) | | |  | | --- | | p-value | |
| |  | | --- | | Intercept | |  |  | 3,944751 | 0,103786 | 38,00835 | 0,000000 |
| |  | | --- | | lnX3 | | 0,148159 | 0,106530 | 0,027483 | 0,019761 | 1,39077 | 0,168300 |
| |  | | --- | | lnX4 | | -0,089598 | 0,096278 | -0,013790 | 0,014818 | -0,93062 | 0,354960 |
| |  | | --- | | lnX5 | | -0,098910 | 0,106814 | -0,017236 | 0,018614 | -0,92601 | 0,357334 |
| |  | | --- | | lnX6 | | -0,189775 | 0,099578 | -0,043671 | 0,022915 | -1,90579 | 0,060410 |
| |  | | --- | | lnX7 | | 0,693749 | 0,134039 | 0,119335 | 0,023057 | 5,17572 | 0,000002 |
| |  | | --- | | lnX8 | | -0,099396 | 0,142477 | -0,016155 | 0,023157 | -0,69763 | 0,487512 |
| |  | | --- | | lnX13 | | 0,241563 | 0,102738 | 0,007074 | 0,003008 | 2,35125 | 0,021269 |

После устранения мультиколлинеарности:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N=85 | Regression Summary for Dependent Variable: lnY (Sheet1 in данные лабы 5.stw) R= ,67249808 R?= ,45225367 Adjusted R?= ,41758618 F(5,79)=13,045 p | | | | | |
| |  | | --- | | b\* | | |  | | --- | | Std.Err. of b\* | | |  | | --- | | b | | |  | | --- | | Std.Err. of b | | |  | | --- | | t(79) | | |  | | --- | | p-value | |
| |  | | --- | | Intercept | |  |  | 3,961825 | 0,101790 | 38,92137 | 0,000000 |
| |  | | --- | | lnX7 | | 0,626554 | 0,098916 | 0,107776 | 0,017015 | 6,33423 | 0,000000 |
| |  | | --- | | lnX13 | | 0,237426 | 0,093802 | 0,006952 | 0,002747 | 2,53114 | 0,013354 |
| |  | | --- | | lnX6 | | -0,215237 | 0,096519 | -0,049530 | 0,022211 | -2,23000 | 0,028586 |
| |  | | --- | | lnX4 | | -0,112191 | 0,092069 | -0,017267 | 0,014170 | -1,21856 | 0,226639 |
| |  | | --- | | lnX3 | | 0,100570 | 0,097202 | 0,018655 | 0,018030 | 1,03466 | 0,303986 |



то на основании отчета делаем выводы:

* модель регрессии незначима;
* коэффициент детерминации составил 0,45.

В результате получили следующее уравнение регрессии:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N=85 | Regression Summary for Dependent Variable: lnY (Sheet1 in данные лабы 5.stw) R= ,67249808 R?= ,45225367 Adjusted R?= ,41758618 F(5,79)=13,045 p | | | | | |
| |  | | --- | | b\* | | |  | | --- | | Std.Err. of b\* | | |  | | --- | | b | | |  | | --- | | Std.Err. of b | | |  | | --- | | t(79) | | |  | | --- | | p-value | |
| |  | | --- | | Intercept | |  |  | 3,961825 | 0,101790 | 38,92137 | 0,000000 |
| |  | | --- | | lnX7 | | 0,626554 | 0,098916 | 0,107776 | 0,017015 | 6,33423 | 0,000000 |
| |  | | --- | | lnX13 | | 0,237426 | 0,093802 | 0,006952 | 0,002747 | 2,53114 | 0,013354 |
| |  | | --- | | lnX6 | | -0,215237 | 0,096519 | -0,049530 | 0,022211 | -2,23000 | 0,028586 |
| |  | | --- | | lnX4 | | -0,112191 | 0,092069 | -0,017267 | 0,014170 | -1,21856 | 0,226639 |
| |  | | --- | | lnX3 | | 0,100570 | 0,097202 | 0,018655 | 0,018030 | 1,03466 | 0,303986 |

+

|  |  |  |
| --- | --- | --- |
|  | Durbin-Watson d (Sheet1 in данные лабы 5.stw) and serial correlation of residuals | |
| |  | | --- | | Durbin- Watson d | | |  | | --- | | Serial Corr. | |
| |  | | --- | | Estimate | | 1,443130 | 0,262631 |

Для расчета критического значения критерия воспользуемся таблицей значений статистики Дарбина-Уотсана. Для n=85, k=2 получаем dн=1,6; dв=1,7 Значение статистики Дарбина-Уотсона для модели находится в интервале DW ∈ (0; dн ) = (0; 1,6), то гипотеза об отсутствии автокорреляции отвергается. Принимается гипотеза о наличии положительной автокорелляции

Перейдем к уравнению регрессии с исходными показателями:

Из полученной модели следует:

|  |  |
| --- | --- |
| **Признак** | **Значение** |
| **Y2** | Ожидаемая продолжительность жизни при рождении (лет) |
| **X3** | Потребление сахара и кондитерских изделий на душу населения в кг |
| **X4** | Потребление картофеля на душу населения в кг |
| **X5** | Потребление масло растительного и других жиров на душу населения в кг |
| **X6** | Потребление мяса и мясных продуктов на душу населения в кг |
| **X7** | Потребление овощей и бахчевых на душу населения в кг |
| **X8** | Потребление фруктов и ягод на душу населения в кг |
| **X13** | Валовой региональный продукт на душу населения (тыс рублей) |

Ожидаемая продолжительность жизни при рождении (лет) увеличится на 0,027% с ростом Потребления сахара и кондитерских изделий на душу населения в кг на 1%

Ожидаемая продолжительность жизни при рождении (лет) уменьшится на 0,014% с ростом Потребления картофеля на душу населения в кг на 1%

Ожидаемая продолжительность жизни при рождении (лет) уменьшится на 0,017% с ростом Потребления масла растительного и других жиров на душу населения в кг на 1%

Ожидаемая продолжительность жизни при рождении (лет) уменьшится на 0,05% с ростом Потребления мяса и мясных продуктов на душу населения в кг на душу населения в кг на 1%

Ожидаемая продолжительность жизни при рождении (лет) увеличится на 0,12% с ростом Потребления овощей и бахчевых на душу населения в кг на душу населения в кг на 1%

Ожидаемая продолжительность жизни при рождении (лет) увеличится на 0,007% с ростом Валового регионального продукта на душу населения (тыс рублей) на 1%

Ожидаемая продолжительность жизни при рождении (лет) уменьшится на 0,016% с ростом Потребления фруктов и ягод на душу населения в кг на 1%

# Сравнение моделей

По имеющимся данным были построены две модели: линейная и нелинейная в форме Кобба-Дугласа. Для этих моделей сравним оцененные значения результативного признака.

|  |  |  |
| --- | --- | --- |
| **Y2** | **Yлин** | **Yэксп** |
| 73,03 | 73,10 | 71,15136 |
| 71,22 | 73,37 | 71,62783 |
| 70,8 | 72,74 | 71,52806 |
| 72,45 | 74,42 | 73,11574 |
| 70,57 | 73,32 | 71,69057 |
| 72,48 | 71,96 | 70,6022 |
| 69,9 | 72,60 | 70,77765 |
| 71,54 | 75,03 | 72,8197 |
| 72,12 | 73,97 | 72,82787 |
| 73,78 | 75,37 | 73,52254 |
| 70,73 | 72,00 | 70,07043 |
| 72,14 | 72,68 | 71,32343 |
| 70,35 | 71,63 | 69,90205 |
| 72,01 | 71,63 | 70,66804 |
| 69,94 | 73,40 | 72,1014 |
| 71,86 | 72,68 | 71,40965 |
| 71,55 | 72,90 | 71,2698 |
| 78,17 | 80,12 | 74,13281 |
| 69,03 | 72,84 | 70,55415 |
| 69,94 | 72,83 | 71,79966 |
| 70,74 | 71,78 | 70,29893 |
| 70,93 | 74,30 | 72,38299 |
| 71,56 | 74,12 | 72,46199 |
| 73,07 | 72,71 | 71,25276 |
| 73,33 | 72,49 | 72,05056 |
| 70,16 | 72,74 | 71,61605 |
| 70,45 | 71,69 | 70,23428 |
| 68,95 | 71,89 | 70,26922 |
| 75,77 | 76,51 | 74,17008 |
| 73,6 | 73,98 | 71,72048 |
| 73,49 | 72,19 | 71,36556 |
| 71,97 | 74,74 | 74,04661 |
| 72,92 | 74,60 | 73,62068 |
| 71,8 | 75,42 | 74,00691 |
| 73,24 | 74,29 | 73,04475 |
| 72 | 76,43 | 74,52346 |
| 74,57 | 75,08 | 73,01586 |
| 78,22 | 77,69 | 74,80691 |
| 78,34 | 74,46 | 72,14794 |
| 75,51 | 74,42 | 72,35164 |
| 75,32 | 75,67 | 73,56322 |
| 74,7 | 74,95 | 73,40341 |
| 74,61 | 73,85 | 71,34711 |
| 74,29 | 75,60 | 74,09267 |
| 72,98 | 73,18 | 71,21385 |
| 71,9 | 73,56 | 71,64721 |
| 73,16 | 73,82 | 71,31632 |
| 74,92 | 74,80 | 73,04858 |
| 72,13 | 74,38 | 73,09229 |
| 72,49 | 74,79 | 73,10609 |
| 70,9 | 73,53 | 72,85003 |
| 71,31 | 74,29 | 72,64577 |
| 71,49 | 74,68 | 73,167 |
| 71,24 | 73,20 | 71,96888 |
| 72,07 | 73,12 | 71,67855 |
| 72,14 | 72,71 | 72,77222 |
| 72,85 | 75,02 | 73,78857 |
| 71,34 | 72,95 | 71,8373 |
| 69,88 | 76,40 | 73,6152 |
| 71,31 | 74,65 | 73,43278 |
| 75,41 | 74,14 | 72,20855 |
| 74,82 | 76,91 | 74,59215 |
| 74,59 | 74,33 | 72,67972 |
| 72,16 | 73,30 | 72,84616 |
| 68,47 | 71,43 | 68,93029 |
| 67,11 | 69,23 | 66,12931 |
| 70,57 | 73,72 | 71,67464 |
| 69,96 | 71,83 | 70,14728 |
| 70,58 | 73,60 | 72,30156 |
| 69,31 | 72,44 | 70,98155 |
| 69,64 | 72,38 | 71,27208 |
| 71,49 | 73,15 | 71,82034 |
| 71,45 | 73,24 | 71,6868 |
| 72,33 | 71,52 | 70,83266 |
| 69,35 | 70,85 | 69,16798 |
| 72,67 | 71,55 | 69,93313 |
| 67,75 | 72,26 | 69,55091 |
| 68,77 | 70,82 | 69,62647 |
| 69,71 | 72,37 | 70,99851 |
| 69,96 | 72,02 | 71,70902 |
| 68,17 | 71,52 | 69,32468 |
| 68,45 | 70,91 | 69,47707 |
| 70,37 | 70,23 | 69,07258 |
| 67,7 | 72,33 | 69,52582 |
| 66,2 | 70,63 | 68,90304 |

Оцененные значения результативного признака для линейной и нелинейной моделей отличаются значительно. Сравним другие характеристики наших моделей.

|  |  |  |
| --- | --- | --- |
| № | Линейная модель | Нелинейная модель |
|  | 0,52 | 0,46 |
| F | 17,16 | 9,42 |
| p-value | <0, 000 | <0,0000 |

Коэффициент детерминации различается незначительно, и в обоих случаях модели являются значимыми.

# Вывод

В ходе работы на основе данных, представленных в приложении А, были построены две модели регрессии: Одна линейная, вторая – нелинейная в форме Кобба-Дугласа. Оценки уравнений имеют следующий вид:

1. Линейная Модель

Ожидаемая продолжительность жизни снижается на 0,032 года при увеличении Потребления мяса и мясных продуктов на душу населения на 1 кг

Ожидаемая продолжительность жизни увеличивается на 0,072 года при увеличении Потребление овощей и бахчевых на душу населения на 1 кг

Ожидаемая продолжительность жизни увеличивается на 0,1 года при увеличении Потребления сахара и кондитерских изделий на душу населения в кг на 1 кг

Ожидаемая продолжительность жизни увеличивается на года при увеличении Валовой региональный продукт на душу населения (тыс рублей) на 1 тыс рублей

1. Нелинейная модель

Ожидаемая продолжительность жизни при рождении (лет) увеличится на 0,027% с ростом Потребления сахара и кондитерских изделий на душу населения в кг на 1%

Ожидаемая продолжительность жизни при рождении (лет) уменьшится на 0,014% с ростом Потребления картофеля на душу населения в кг на 1%

Ожидаемая продолжительность жизни при рождении (лет) уменьшится на 0,017% с ростом Потребления масла растительного и других жиров на душу населения в кг на 1%

Ожидаемая продолжительность жизни при рождении (лет) уменьшится на 0,05% с ростом Потребления мяса и мясных продуктов на душу населения в кг на душу населения в кг на 1%

Ожидаемая продолжительность жизни при рождении (лет) увеличится на 0,12% с ростом Потребления овощей и бахчевых на душу населения в кг на душу населения в кг на 1%

Ожидаемая продолжительность жизни при рождении (лет) увеличится на 0,007% с ростом Валового регионального продукта на душу населения (тыс рублей) на 1%

Ожидаемая продолжительность жизни при рождении (лет) уменьшится на 0,016% с ростом Потребления фруктов и ягод на душу населения в кг на 1%