Ушаков Владимир Александрович, группа МПиКИ

Лабораторная работа № 1

**Регрессионный анализ**

**Цель работы:**

Целью лабораторной работы является проведение регрессионного анализа при помощи Excel и R, а также познакомиться с основными метриками и сделать выбор об исследуемой выборке

**Задание:**

1. Для заданного набора данных построить линейную модель множественной регрессии.

2. Оценить адекватность и значимость построенного уравнения регрессии.

3. Выделить значимые и незначимые факторы в модели.

4. Построить уравнение регрессии со статистически значимыми факторами. Дать экономическую интерпретацию параметров модели.

5. Оценить гетероскедастичность дисперсии остатков с помощью теста

Гольдфельда– Квандта.

6. Определить наличие автокорреляции остатков с помощью

a. теста Дарбина-Уотсона;

b. коэффициента автокорреляции

**Код программы (с выделением внесенных изменений):**

lr1 <- read.csv2("data.csv")

View(lr1)

model <- lm(data = lr1, Y ~ X1 + X2 + X3 + X4)

print("Результаты регрессионного анализа")

print(summary(model))

print("Результаты регрессионного анализа со

статистически значимыми коэффициентами")

model <- lm(data = LR1, Y ~ X1 + X2 + X3)

print(summary(model))

# install.packages("lmtest")

library("lmtest")

print(gqtest(model, order.by = ~X3, data = LR1, fraction = 8))

print(dwtest(model))

# install.packages("car")

library(car)

print(durbinWatsonTest(model))

**В результате выполнения задания:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Дисперсионный анализ |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Значимость F* |  |  |  |
| Регрессия | 4 | 14066,02 | 3516,506 | 86,07833 | 6,69E-12 |  |  |  |
| Остаток | 19 | 776,1956 | 40,8524 |  |  |  |  |  |
| Итого | 23 | 14842,22 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* | *Верхние 95,0%* |
| Y-пересечение | -203,552 | 82,87974 | -2,456 | 0,023847 | -377,022 | -30,0831 | -377,022 | -30,0831 |
| X1 | 0,665526 | 0,065781 | 10,11725 | 4,36E-09 | 0,527844 | 0,803208 | 0,527844 | 0,803208 |
| X2 | 1,239273 | 0,330975 | 3,744309 | 0,001374 | 0,546534 | 1,932012 | 0,546534 | 1,932012 |
| X3 | 6,980195 | 2,52993 | 2,759047 | 0,012487 | 1,684992 | 12,2754 | 1,684992 | 12,2754 |
| X4 | 1,090743 | 2,846426 | 0,383197 | 0,705827 | -4,8669 | 7,048381 | -4,8669 | 7,048381 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Дисперсионный анализ | | |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Значимость F* |  |  |  |
| Регрессия | 3 | 14060,03 | 4686,675 | 119,834 | 5,97E-13 |  |  |  |
| Остаток | 20 | 782,1944 | 39,10972 |  |  |  |  |  |
| Итого | 23 | 14842,22 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* | *Верхние 95,0%* |
| Y-пересечение | -178,454 | 49,68996 | -3,59135 | 0,001825 | -282,105 | -74,8025 | -282,105 | -74,8025 |
| X1 | 0,664395 | 0,064298 | 10,33304 | 1,82E-09 | 0,530271 | 0,798518 | 0,530271 | 0,798518 |
| X2 | 1,272114 | 0,312794 | 4,066934 | 0,000602 | 0,619636 | 1,924591 | 0,619636 | 1,924591 |
| X3 | 7,351938 | 2,286163 | 3,215841 | 0,004336 | 2,583085 | 12,12079 | 2,583085 | 12,12079 |

|  |  |  |
| --- | --- | --- |
|  | DW | 1,352191 |
|  |  |  |
| du | 1,66 |  |
| dl | 1,1 |  |
|  |  |  |
| dl < DW < du | | автокорреляция не определена |

|  |  |
| --- | --- |
| F | 32,50232 |
| Fтабл | 6,388233 |
| F > Fтабл => модель гетероскедастична | |

Residuals:

Min 1Q Median 3Q Max

-8.495 -3.019 -1.616 1.562 15.571

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -203.55241 82.87974 -2.456 0.02385 \*

X1 0.66553 0.06578 10.117 4.36e-09 \*\*\*

X2 1.23927 0.33098 3.744 0.00137 \*\*

X3 6.98020 2.52993 2.759 0.01249 \*

X4 1.09074 2.84643 0.383 0.70583

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.392 on 19 degrees of freedom

Multiple R-squared: 0.9477, Adjusted R-squared: 0.9367

F-statistic: 86.08 on 4 and 19 DF, p-value: 6.693e-12

**Анализ результатов:**

Регрессионная модель: Y=-203.552 + 0.665X1 + 1.239X2 + 6.98X3 + 1.09X4

Регрессионная модель без фактора X4: Y=-178.45 + 0.664X1 + 1.27X2 + 7.35X3

F=81.82/2.51=32.5

Fтабл=6.38

Так как F > Fтабл => модель гетероскедастична

DW=1.35 ,du=1.66, dl=1,1, dl < DW < du => автокорреляция не определена

**Вывод:**

В результате проведения регрессионного анализа были построены модели для 4 и 3 факторов, а также было выяснено, что модель гетероскедастична и автокорреляция остатков не определена