Федеральное агентство по образованию

Государственное образовательное учреждение

высшего профессионального образования

«Южно-Уральский государственный университет»

Кафедра "Прикладной математики"

ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №0

по дисциплине «Эконометрика»

Тема: «Построение регрессионной модели»

Выполнил:

Проверил: Литвинов А.О.

«\_\_\_» \_\_\_\_\_\_\_\_\_ 2016 г.

Челябинск - 2016

**library**("dplyr")

**library**("ggplot2")

**library**("GGally")

**library**("psych")

*#1. Посторойте поле корреляции и сформулируйте гипотезу о форме связи.*

*#2. Рассчитайте параметры уравнения линейной, показательной и гиперболической парной регрессии.*

*#3.Оцените тесноту связи с помощью показателей корреляции и детерминации*

*#4. Дайте с помощью общего коэффициента эластичности сравнительную оценку силы связи фактора с результатом.*

*#5. Оцените с помощью средней ошибки аппроксимации качество уравнений.*

#modelCoef[1] - intercept

#modelCoef[2] - coeff

X = **data.frame**(meanSalary = **c**(240,226,221,226,220,250,237,232,215,220,222,231,229), *#target*

minSalary = **c**(178,202,197,201,189,302,215,166,199,180,181,186,250)) *#train*

nd <- **data.frame**(minSalary = **sort**(X$minSalary))

plotNewData <- **function**(x, predictedY) {

**ggplot**(data = X, **aes**(x = X$minSalary, y = X$meanSalary)) +

**geom\_point**(**aes**(x = X$minSalary, y = X$meanSalary)) +

**geom\_line**(**aes**(x = x, y = predictedY ))

}

getStd <- **function**(predicted, y){

sum = 0

for (i in 1:13)

{

sum = sum + (predicted[i] - y[i]) \* (predicted[i] - y[i])

}

sum = **sqrt**(sum)

return(sum / 13)

}

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_linear\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

linearModel <- **lm**(meanSalary~minSalary, data = X)

linearCoef <- **coef**(linearModel)

#**summary**(linearModel)

x <- nd$minSalary

y <- linearCoef[1] + linearCoef[2] \* x

**plotNewData**(x, y)



linearStd = **getStd**(y, X$meanSalary)

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_polynomial\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

linearizedDataX <- **log**(X$minSalary)

linearizedDataY <- **log**(X$meanSalary)

polynomialModel <- **lm**(linearizedDataY~linearizedDataX)

polynomialCoef <- **coef**(polynomialModel)

#**summary**(polynomialModel)

x <- nd$minSalary

#y = b0 \* x ^ b1

y <- **exp**(polynomialCoef[1]) \* (x ^ polynomialCoef[2])

**plotNewData**(x, y)



polynomialStd = **getStd**(y, X$meanSalary)

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_exponential\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

linearizedDataX <- **exp**(X$minSalary \* 0.005)

linearizedDataY <- X$meanSalary

exponentialModel <- **lm**(linearizedDataY~linearizedDataX)

exponentialCoef <- **coef**(exponentialModel)

#**summary**(exponentialModel)

x <- nd$minSalary

#y = b0 + e ^ (x \* 0.005)

y <- exponentialCoef[1] + exponentialCoef[2] \* **exp**(x \* 0.005)

**plotNewData**(x, y)



exponentialStd = **getStd**(y, X$meanSalary)

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_inverse\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

linearizedDataX <- 1000 / X$minSalary

linearizedDataY <- X$meanSalary

inverseModel <- **lm**(linearizedDataY~linearizedDataX)

inverseCoef <- **coef**(inverseModel)

#summary(inverseModel)

x <- nd$minSalary

#y = b0 + b1 \* (1000 / x)

y <- inverseCoef[1] + inverseCoef[2] \* (1000 / x)

**plotNewData**(x, y)



inverseStd = **getStd**(y, X$meanSalary)

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_hyperbolic\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

linearizedDataX <- X$minSalary

linearizedDataY <- 1/X$meanSalary

hyperbolicModel <- **lm**(linearizedDataY~linearizedDataX)

hyperbolicCoef <- **coef**(hyperbolicModel)

#**summary**(hyperbolicModel)

x <- nd$minSalary

#y = 1 / (b0 + b1 \* x)

y <- 1 / (hyperbolicCoef[1] + hyperbolicCoef[2] \* x)

**plotNewData**(x, y)



hyperbolicStd = **getStd**(y, X$meanSalary)

#linearModel, polynomialModel, exponentialModel, inverseModel, hyperbolicModel

result = **data.frame**(*modelName* = c("linear", "polynomial", "exponential", "inverse", "hyperbolic"),

*correlation* = **cor**(X$minSalary, X$meanSalary),

*determincation* = **c**(**summary**(linearModel)$r.squared, **summary**(polynomialModel)$r.squared, **summary**(exponentialModel)$r.squared,**summary**(inverseModel)$r.squared, **summary**(hyperbolicModel)$r.squared),

*stdError* = **c**(linearStd, polynomialStd, exponentialStd, inverseStd, hyperbolicStd),

*F* = **c**(**summary**(linearModel)$fstatistic[1], **summary**(polynomialModel)$fstatistic[1], **summary**(exponentialModel)$fstatistic[1], **summary**(inverseModel)$fstatistic[1], **summary**(hyperbolicModel)$fstatistic[1]))

| **modelName** | | **correlation** | **R squred** | **stdError** | **F** |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| **1** | linear | 0.5618762 | 0.3157049 | 3.025802 | 5.074936 |
| **2** | polynomial | 0.5618762 | 0.2600144 | 2.978663 | 3.865154 |
| **3** | exponential | 0.5618762 | 0.3604340 | 3.052483 | 6.199163 |
| **4** | inverse | 0.5618762 | 0.2228003 | 2.957549 | 3.153377 |
| **5** | hyperbolic | 0.5618762 | 0.2920475 | 2.998704 | 4.537765 |

По результатам анализа лучшей моделью оказалась обратная y = 4.935e-03 + -2.699e-06 \* (1000 / x). Но не смотря на это мы можем использовать любую из моделей, так как различие между ними минимальное. При учете значений целевой переменной ***mean*** = 228.384, ***sd*** = 9.674, и при визуальном анализе, где мы можем увидеть, что есть по крайней мере один выброс, который может сильно влиять на всю модель и несогласующийся с остальными данными и априорными предположениями – минимальная зп – 250, средняя 229. Вывод – берем любую модель.