Университет ИТМО

Задание №2

**Современные инструменты анализа данных**

Выполнил: Балтабаев Дамир  
Группа: P3210

Преподаватель: Максимова Татьяна Геннадьевна

Санкт-Петербург  
2021

1. Парнас



**Trend Lines Model**

A linear trend model is computed for Стоимость given Площадь. The model may be significant at p <= 0,05.

|  |  |
| --- | --- |
| **Model formula:** | ( Площадь + intercept ) |
| **Number of modeled observations:** | 69 |
| **Number of filtered observations:** | 0 |
| **Model degrees of freedom:** | 2 |
| **Residual degrees of freedom (DF):** | 67 |
| **SSE (sum squared error):** | 1,71553e+09 |
| **MSE (mean squared error):** | 2,56049e+07 |
| **R-Squared:** | 0,675712 |
| **Standard error:** | 5060,13 |
| **p-value (significance):** | < 0,0001 |

**Individual trend lines:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Panes** | | **Line** | | **Coefficients** | | | | |
| **Row** | **Column** | **p-value** | **DF** | **Term** | **Value** | **StdErr** | **t-value** | **p-value** |
| Стоимость | Площадь | < 0,0001 | 67 | Площадь | 369,034 | 31,233 | 11,8155 | < 0,0001 |
|  | | | | intercept | 9894,49 | 1576,31 | 6,27698 | < 0,0001 |

y=369.034 \* S(площадь) +9894.49

Коэффициент детерминации: 0.675712

1. Петроградская



**Trend Lines Model**

A linear trend model is computed for Стоимость given Площадь. The model may be significant at p <= 0,05.

|  |  |
| --- | --- |
| **Model formula:** | ( Площадь + intercept ) |
| **Number of modeled observations:** | 65 |
| **Number of filtered observations:** | 0 |
| **Model degrees of freedom:** | 2 |
| **Residual degrees of freedom (DF):** | 63 |
| **SSE (sum squared error):** | 1,39739e+10 |
| **MSE (mean squared error):** | 2,21808e+08 |
| **R-Squared:** | 0,68656 |
| **Standard error:** | 14893,2 |
| **p-value (significance):** | < 0,0001 |

**Individual trend lines:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Panes** | | **Line** | | **Coefficients** | | | | |
| **Row** | **Column** | **p-value** | **DF** | **Term** | **Value** | **StdErr** | **t-value** | **p-value** |
| Стоимость | Площадь | < 0,0001 | 63 | Площадь | 857,048 | 72,958 | 11,7471 | < 0,0001 |
|  | | | | intercept | 6334,11 | 4630,06 | 1,36804 | 0,17616 |

y=857.048\* S(площадь) +6334.11

Коэффициент детерминации: 0.68656

1. Множественная регрессия стоимости от площади



**Trend Lines Model**

A linear trend model is computed for Стоимость given Площадь. The model may be significant at p <= 0,05. The factor Петр-1, Парнас-0 may be significant at p <= 0.05.

|  |  |
| --- | --- |
| **Model formula:** | Петр-1, Парнас-0\*( Площадь + intercept ) |
| **Number of modeled observations:** | 134 |
| **Number of filtered observations:** | 0 |
| **Model degrees of freedom:** | 4 |
| **Residual degrees of freedom (DF):** | 130 |
| **SSE (sum squared error):** | 1,56895e+10 |
| **MSE (mean squared error):** | 1,20688e+08 |
| **R-Squared:** | 0,799583 |
| **Standard error:** | 10985,8 |
| **p-value (significance):** | < 0,0001 |

**Analysis of Variance:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field** | **DF** | **SSE** | **MSE** | **F** | **p-value** |
| **Петр-1, Парнас-0** | 2 | 1.813895e+10 | 9,06948e+09 | 75,148 | < 0,0001 |

**Individual trend lines:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Panes** | | **Line** | | **Coefficients** | | | | |
| **Row** | **Column** | **p-value** | **DF** | **Term** | **Value** | **StdErr** | **t-value** | **p-value** |
| 0 | Площадь | < 0,0001 | 67 | Площадь | 369,034 | 31,233 | 11,8155 | < 0,0001 |
|  | | | | intercept | 9894,49 | 1576,31 | 6,27698 | < 0,0001 |
| 1 | Площадь | < 0,0001 | 63 | Площадь | 857,048 | 72,958 | 11,7471 | < 0,0001 |
|  | | | | intercept | 6334,11 | 4630,06 | 1,36804 | 0,17616 |
|  | | | |  |  |  |  |  |

**Вывод**: При отдельном рассмотрении регрессий мы можем заметить, что коэффициент детерминации Петроградского района выше, следовательно модель данного района более действительна. Если рассматривать модель множественной регрессии, коэффициент регрессии от двух аргументов имеет более высокий показатель, нежели отдельные модели, отсюда можно сделать вывод, что данная модель является наиболее точной.