**Тест № 2.6**

**Регрессионный анализ**

**Выполнил:** студент 1 курса магистратуры,

образовательная программа

«Распределенные вычислительные технологии»,

группа 21.М12-ПУ,

Романычев Леонид

**г. Санкт-Петербург, 2021**

Регрессионный анализ находит широкое применение в маркетинговых исследованиях, когда изучается взаимосвязь двух и более переменных. Рассмотрим изготовителя шариковых ручек Click, который заинтересован в исследовании эффективности маркетинговых усилий своей фирмы. Компания использует оптовых торговцев для реализации продукции Click и в дополнение к их усилиям прибегает к персональным продажам и коротким рекламным телероликам. Компания планирует использовать в качестве меры оценки эффективности ежегодный объем продаж по территориям. ( Г.А. Черчилль «Маркетинговые исследования»)

|  |  |  |  |
| --- | --- | --- | --- |
| Территория | Продажа  (тыс. долларов) | Телевизионная реклама (число показов в месяц) | Число торговых представителей |
| 005  019  033  039  061  082  091  101  115  118  133  149  162  164  178  187  189  205  222  237  242  251  260  266  279  298  306  332  347  358 | 270,8  280,0  270,5  411,0  431,2  315,3  569,1  574,0  421,1  317,0  405,6  229,5  343,6  645,6  522,4  327,5  420,0  335,2  452,4  421,8  249,6  505,3  370,7  275,5  620,6  449,5  283,1  368,0  545,1  575,0 | 6  8  7  10  12  8  11  16  13  6  10  5  9  17  19  10  12  6  13  14  3  16  8  5  18  18  4  8  12  13 | 4  6  2  5  8  4  7  9  4  2  6  5  4  8  6  3  6  2  5  5  4  7  5  3  6  5  2  6  7  6 |

Построены парные регрессии:

|  |  |
| --- | --- |
| **По х1** | |
| 22,6339616 | 167,701139 |
| 2,58178251 | 29,5686431 |
| 0,73296895 | 62,4645548 |
| 76,8567191 | 28 |
| 299881,13 | 109250,977 |

|  |  |
| --- | --- |
| **По х2** | |
| 46,3298429 | 172,128796 |
| 8,17202298 | 44,0582451 |
| 0,53442921 | 82,4794254 |
| 32,1412303 | 28 |
| 218652,149 | 190479,957 |

Диаграммы корреляционных полей с линиями регрессий:

Графики остатков по фактору:

Графики остатков по номеру наблюдения:

Проведена проверка значимости коэффициентов парных регрессий:

|  |  |  |
| --- | --- | --- |
|  | **значимость** | |
|  | **х1** | **х2** |
| **a** | 5,67158723 | 3,90684639 |
| **b** | 8,7667964 | 5,66932362 |
| **Tкрит** | 2,36845175 |

Таким образом, все коэффициенты являются значимыми.

Регрессии проверены по критерию Фишера:

|  |  |
| --- | --- |
| **F1** | 5,48031314 |
| **F2** | 6,96680962 |
| **Fкрит** | 4,19597182 |

Поскольку значение F-статистики превышает табличное значение, то линейная регрессия статистически значима по критерию Фишера в обоих случаях.

Проведена проверка остатков на нормальность по критерию Жарка-Бера:

|  |  |  |
| --- | --- | --- |
| **критерий Жарка-Бера** | | |
| **ХИ2крит** | **х1** | **х2** |
| 17,7083662 | 9,50738468 | 11,0969532 |

По результатам можно сказать, что остатки распределены нормально.

Далее с помощью функции ЛИНЕЙН построена множественная регрессия:

|  |  |  |
| --- | --- | --- |
| **b2** | **b1** | **a** |
| 19,3334967 | 17,4199396 | 124,839588 |

Все коэффициенты являются значимыми:

|  |  |  |  |
| --- | --- | --- | --- |
| **t** | 2,62125351 | 5,66165247 | 3,96739894 |
| **tкрит** | 0,03162972 |  |  |

Тест Фишера показал значимость множественной регрессии:

|  |  |
| --- | --- |
| **F** | 49,9214068 |
| **Fкр** | 4,21000847 |

Тест Жарка-Бера показал, что остатки распределены нормально:

|  |  |
| --- | --- |
| **ЖБ** | 14,1978562 |
| **хи2кр** | 16,927875 |

При сравнении парных регрессий и множественной видно, что, хотя парные регрессии являются состоятельными, множественная дает меньший разброс остатков и имеет больший коэффициент детерминации.

Далее был проведен прогноз показателя для значений факторов на 185% превышающих средние значения факторов с уровнем доверия 95%.

|  |  |
| --- | --- |
| **x1^** | 19,5483333 |
| **х2^** | 9,37333333 |
|  |  |
| **у\_х1** | 610,157365 |
| **у\_х2** | 606,393857 |
| **у\_мн** | 646,589684 |

Стандартные ошибки:

|  |  |
| --- | --- |
| **mx1** | 67,5987985 |
| **mx2** | 90,9299054 |
| **mмн** | 1701,36656 |

Доверительные интервалы:

|  |  |  |
| --- | --- | --- |
|  | **Доверительный интервал** | |
| **ух1** | 450,052873 | 770,2618578 |
| **ух2** | 391,030763 | 821,7569505 |
| **yмн** | 592,775933 | 700,4034336 |

Предсказанные значения попали в соответствующие доверительные интервалы.

**Выводы**

Парный регрессионный анализ показал, что все коэффициенты регрессий как по фактору 𝑥1, так и по фактору 𝑥2 значимы. Также принимается гипотеза о значимости в целом обеих парных регрессий.

На основе этого можно сделать вывод, что существует линейная связь как между продажами и рекламой, так и между продажами и числом торговых представителей. То есть можно предсказать уровень продаж, основываясь только на числе показов рекламы или числе торговых представителей.

В то же время, основываясь на виде диаграммы остатков, можно сделать вывод, что регрессии не являются адекватными для описания истинных зависимостей между продажами и факторами. Иначе говоря, предсказанный уровень продаж может не совпадать с действительностью. Следовательно, стоит рассмотреть множественную регрессию показателя на все факторы.

Все коэффициенты множественной линейной регрессии значимы. Также принимается гипотеза о значимости множественной регрессии в целом. На основе этого можно сделать вывод, что между продажами существует линейная связь с числом показов рекламы и числом торговых представителей. То есть зная число торговых представителей и зная число показов рекламы можно спрогнозировать уровень продаж.

Коэффициент регрессии, стоящий при x2, больше стоящего при х1. Это может говорить о том, что на уровень продаж большее влияние оказывает число торговых представителей, чем число показов рекламы. Основываясь на визуальном анализе диаграммы остатков, нельзя сказать, что регрессия адекватно моделирует связь между показателем и факторами. Другими словами, предсказанный уровень продаж может не совпадать с действительностью. Следовательно, стоит рассмотреть наличие нелинейной связи между показателем продаж и факторами.

Рассчитав прогноз показателя **y** с помощью всех трёх регрессий, можно заметить, что спрогнозированные точечные результаты довольно сильно отличаются. Если рассматривать интервальные прогнозы, то видно, что у прогноза для множественной линейной регрессии он имеет размах существенно меньший, чем размахи интервалов парных линейных регрессий. Из этого можно сделать вывод, что с помощью множественной линейной регрессии можно с большей точностью предсказать объёмы продаж по обоим факторам.

Сравнивая все три регрессионные модели, можно сказать, что все они статистически значимы и могут быть использованы для прогнозирования y. Но коэффициент детерминации множественной регрессии больше коэффициентов детерминации парных регрессии, то есть для множественной регрессии доля объясненной части общей вариации больше, а значит использование множественной линейной регрессии более эффективно.