**Тест № 1.16**

**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.

Компания Joseph Machine Company была озабочена улучшением процедуры подбора торгового персонала. Изучение информационных источников, касающихся эффективности работников торговли, показало, что главными причинами успеха часто оказываются опыт работы и интеллектуальные способности торгового представителя. По этой причине компания решила назначить тесты коэффициентов интеллекта (IQ) всех торговых представителей. В качестве критерия эффективности использовался индекс объема продаж за год по отношению к территориальной квоте. (Г. А. Черчилль «Маркетинговые исследования»)

|  |  |  |  |
| --- | --- | --- | --- |
| Торговый  представитель | Индекс эффективности | Тест (IQ) | Время работы  (в месяцах) |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30 | 120  107  102  94  96  105  99  115  77  100  116  122  87  109  95  92  91  102  120  94  98  101  97  101  97  98  112  115  111  97 | 128  99  92  80  97  113  86  84  114  113  91  80  88  82  91  83  84  113  84  98  114  111  108  85  91  91  80  102  113  91 | 79  49  82  63  99  64  73  86  60  56  118  104  50  111  78  62  75  83  116  103  54  90  76  72  75  80  88  85  107  110 |

**ЗАДАНИЕ:**

Постройте парные линейные регрессии — зависимости результативного признака *y* от факторов  и  взятых по отдельности. В каждом случае постройте поле корреляции, вычислите коэффициенты уравнения линейной регрессии, коэффициент детерминации, коэффициент корреляции, значение *F* статистики. Для вычислений воспользуйтесь встроенной функцией **ЛИНЕЙН.** Проверьте гипотезу гомоскедастичности для каждой из регрессий.

**Диаграмма корреляционного поля с линией регрессии.**

Для первого случая

Для второго случая

**Найденное уравнение линейной регрессии.**

Для у и х1: 𝑦 = −0,0279𝑥 + 105,02

Для у и х2: 𝑦 = 0,2785𝑥 + 79,609

|  |  |  |
| --- | --- | --- |
| **По х1** | | |
| -0,027879621 | | 105,015353 |
| 0,148753886 | | 14,4444543 |
| 0,001252952 | | 10,7646847 |
| 0,035126679 | | 28 |
| 4,070424687 | | 3244,59624 |
| **По х2** | | |
| 0,278481672 | 79,6092289 | |
| 0,085434177 | 7,16968718 | |
| 0,275081372 | 9,17103161 | |
| 10,62502484 | 28 | |
| 893,6476847 | 2355,01898 | |

**График остатков**

Для первого случая

Для второго случая

**Коэффициент корреляции и детерминации, средняя ошибка аппроксимации, результаты *F-*статистики и выводы по критерию Фишера. Выводы по критерию Жарка-Бера.**

Для двух регрессий посчитаны коэффициенты корреляции, детерминации, а также средняя ошибка аппроксимации:

|  |  |  |
| --- | --- | --- |
|  | **х1** | **х2** |
| Коэфф. Корреляции | -0,0353971 | 0,524482 |
| Коэфф. Детерминации | 0,00125295 | 0,27508137 |
| Ср. ошибка аппр. | 0,08319707 | 0,07058821 |

Коэффициент детерминации получился близким к полученному с помощью функции ЛИНЕЙН.

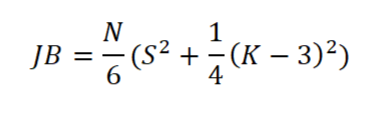
Затем посчитана статистика Фишера

|  |  |  |
| --- | --- | --- |
| F | 0,03512668 | 10,6250248 |
| **Fкрит** | **4,19597182** |  |

По формуле:



По полученному значению (при сравнении с критическим значением) видно, что первая регрессия не является значимой на достаточном уровне (0,05), а вторая **является значимой**.

Статистика Жарка-Бера посчитана по формуле:

Где

***N*** – Количество наблюдений;

***S*** – Коэффициент асимметрии;

***K*** – Коэффициент эксцесса;

|  |  |
| --- | --- |
| ЖБ1 | 12,6019625 |
| ЖБ2 | 13,6667214 |
| ХИ2КР | 17,7083662 |

При сравнении с критическим значением можно сделать вывод о том, что остатки регрессий распределены нормально. Таким образом, метод наименьших квадратов может быть использован.

Далее посчитана значимость коэффициентов регрессий и сравнена с критическим значением:

|  |  |  |
| --- | --- | --- |
| **Значимость** | **x1** | **x2** |
| **а** | 7,27028871 | 11,1035847 |
| **b** | -0,187421128 | 3,25960501 |
|  |  |  |
| Tкр | 0,03161926 |  |

**В первом случае:**

Поскольку значение статистики для **b** меньше критического значения, то оценка статистически не значима. А для параметра **а** – статистически значима.

**Во втором случае:**

Поскольку значение статистики для обоих параметров больше критического значения, то оценки считаем статистически значимыми.

Также посчитаны доверительные интервалы для коэффициентов обеих регрессий на уровне значимости 0,05:

|  |  |  |
| --- | --- | --- |
|  | **Доверительные интервалы** | |
|  | Левая гр. | Правая гр. |
| a1 | 104,5586299 | 105,472076 |
| a2 | 79,38252872 | 79,8359291 |
| b1 | -0,032583109 | -0,0231761 |
| b2 | 0,275780306 | 0,28118304 |

Далее рассмотрены: значение точечного прогноза, ошибка точечного прогноза и доверительный интервал прогноза для значений факторов на 50% превышающих средние выборочные для каждого из факторов

|  |  |  |
| --- | --- | --- |
| **Превышающие на 50%** | | |
|  | **х1** | **х2** |
| **х** | 144,3 | 122,4 |
| **m** | 13,0742481 | 9,9529704 |
| **y** | 100,992324 | 113,695386 |
|  |  |  |
| **Доверительные интервалы** | | |
|  | **Левая гр.** | **Правая гр.** |
| **по х1** | 100,578926 | 101,405722 |
| **по х2** | 113,38068 | 114,010091 |

В обоих случаях оценка попала в доверительный интервал.

**Гомоскедастичность**

Была проведена проверка гипотезы гомоскедастичности для наблюдений каждой из регрессий. Для проверки использовался метод Гольдфельда-Квандта. В каждом наборе наблюдений было исключено ровно m = 8 наблюдений, следует отметить, что должно выполняться этом (𝑛−𝑚)/2 >2. Получившиеся значения F-статистики:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| F1 | 0,359961 |  | F2 | 0,42254242 |
|  |  |  |  |  |
| Fкр | 3,1788931 |  |  |  |

По методу Гольдфельда-Квандта, так как 𝐹<𝐹крит, то гипотеза гомоскедастичности **принимается**.

В нашем случае принятие гипотезы о гомоскедастичности означает, что дисперсии наблюдений равны, следовательно, они не влияют на точность оценивания параметров регрессии. Таким образом, можно сделать вывод, что полученные оценки параметров регрессии адекватны.

**Выводы**

Случай 1:

Так как по критерию Фишера было выявлено, что уравнение регрессии статистически незначимо, то можно сделать вывод, что между переменными y и x1 нет линейной связи. Это подтверждается принятием гипотезы о статистической незначимости коэффициента β линейной регрессии. Анализ остатков, а именно критерий Жарка-Бера о нормальности распределения остатков, позволяет сделать вывод, что между остатки распределены по нормальному закону.

Случай 2:

Так как по критерию Фишера было выявлено, что уравнение регрессии статистически значимо, то можно сделать вывод, что между переменными y и x2 существует линейная связь. Другими словами, индекс эффективности торгового представителя линейно зависит от времени его работы. Это подтверждается отклонением гипотезы о статистической незначимости коэффициентов α и β линейной регрессии и отклонением гипотезы о статистической незначимости регрессии в целом. Критерий Жарка-Бера подтверждает нормальность распределения остатков.

В обоих случаях гипотеза гомоскедастичности принимается. Это означает, что дисперсии наблюдений можем считать равными, следовательно, они не влияют на точность оценивания параметров регрессии. Таким образом, можно сделать вывод, что полученные оценки параметров регрессии адекватны.

Итак, анализ показал, что присутствует линейная зависимость между **коэффициентом продуктивности** и **временем работы**, а **уровень IQ не влияет** линейно на продуктивность. Возможно, есть какая-то другая зависимость от уровня IQ.