Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет Информационных технологий и управления

Кафедра Интеллектуальных информационных технологий

**ОТЧЁТ**

по дисциплине «Статистические основы индуктивного вывода»

ИПР

Тема: Построение множественной линейной регрессионной модели

Выполнил: Мартысюк М. Ю.

гр. 221703

Проверил: Ефремов А. А.

Минск 2024

**Постановка задачи:** известны статистические данные наблюдений за некоторым

количеством однородных экономических объектов.

**Требуется:**

1. Осуществить выбор факторных признаков для построения регрессионной модели.

2. Построить линейное уравнение регрессии, описывающее зависимость между

факторами и результатом.

3. Оценить качество уравнения регрессии с экономической и математической точки

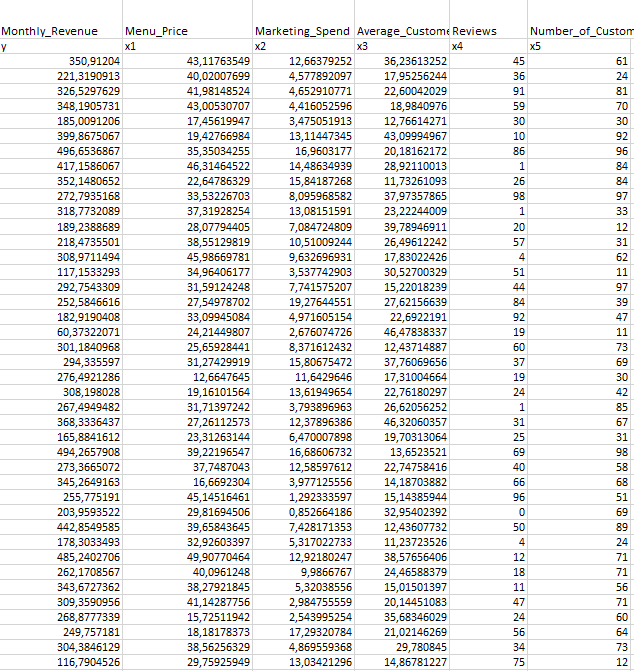
зрения.

4. Найти прогнозное значение результата.

**Описание работы:**

Известны статистические данные ресторанов. Постройте линейное уравнение регрессии, описывающее зависимость между факторами и результатом для предсказания месячной выручки, если прогнозный объем факторных признаков должен составить 115% от их среднего уровня.

**Шаг 1: Таблица с исходными данными:**



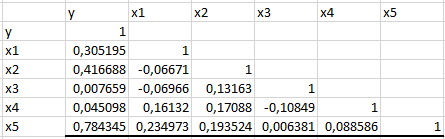
Зависимая переменная: Montly revenue

Факторами являются:

1. Menu\_Price
2. Marketing\_Spend
3. Average\_Customer
4. Reviews
5. Number\_of\_Customers

**Шаг 1. Осуществить выбор факторных признаков для построения регрессионной модели.** Для подбора факторных признаков используем инструмент КОРРЕЛЯЦИЯ. Вызываемкоманду ДАННЫЕ – АНАЛИЗ ДАННЫХ (рис. 3.2). В меню АНАЛИЗ ДАННЫХ выбираем инструментКОРРЕЛЯЦИЯ.

Результат выполнения команды:



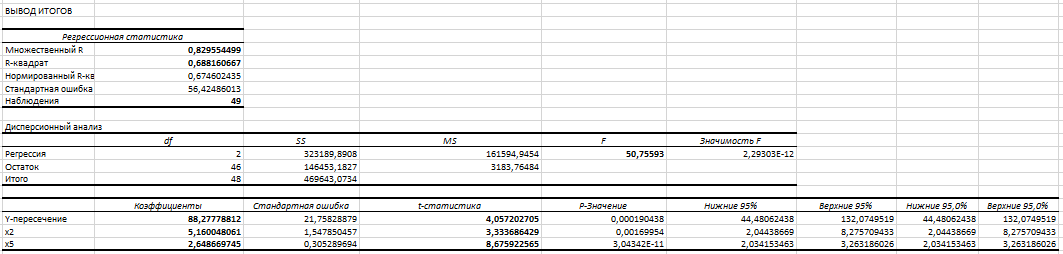
Корреляционная матрица

Анализ коэффициентов парной корреляции: зависимая переменная у имеет

тесную связь с переменной , и не слишком низкое влияние на нее оказывает переменная . Остальные факторы можно исключить из модели.

**Шаг 2. Построить линейное уравнение регрессии, описывающее зависимость между факторами и результатом.**

Регрессия:



Таким образом, мы можем записать полученную модель, описывающую зависимость между выручкой, затратами на рекламу и количеством клиентов,:

**Шаг 3. Оценить качество уравнения регрессии с экономической и математической точки зрения.**

α=0,05 – уровень значимости, n=49 – число наблюдений, число независимых переменных: m=2.

2,013

3,2

Сравним t-статистики с :

Константа (88): t = **4,06 >**  – значима

Коэффициент (5,16) > – значима

Коэффициент (2,65) > – значима

Сравним F-статистики:

F = 50,76; F > – уравнение значимо

**Математическая значимость**:

Все коэффициенты (включая константу) статистически значимы, так как их t-статистики превышают ​. Уравнение в целом значимо, так как F-статистика значительно превышает

**Экономическая значимость**:

означает, что 69% вариации зависимой переменной y объясняется независимыми переменными ​, что указывает на хорошее качество модели.

**Шаг 4. Рассчитайте эластичность по каждому фактору и ранжируйте факторы по их силе влияния на результативный показатель.**

Средние значения:

9,32

56,49

285,77

Таким образом при увеличении x на 1% y увеличится на E%.

Ранжирование факторов по силе влияние: чем выше коэффициент, тем больше влияние данного фактора на результативный показатель. Следовательно наибольше влияние оказывает фактор , затем .

Таким образом можно сделать общий вывод о качестве уравнения и возможности его использования для прогноза:

1) Все коэффициенты в модели статистически значимы (t-значения превышают критическое значение​), а также сама модель в целом статистически значима (F=50,76> =3.2F = 50.8 > = 3.2F=50.8> ​=3.2). Это говорит о том, что факторы​ оказывают значительное влияние на результативную переменную y.

2) Коэффициент детерминации показывает, что 69% изменчивости зависимой переменной y объясняется изменениями факторов и ​. Это высокий уровень объяснения, особенно для экономических моделей, где коэффициент выше 0.6 считается хорошим результатом.

3) Значение стандартной ошибки (если оно доступно) позволяет оценить разброс предсказаний модели. Если стандартная ошибка относительно невелика по сравнению с диапазоном значений зависимой переменной, это также подтверждает качество модели.

**Шаг 5. Найти прогнозное значение результата.**

Построенная в примере 3 модель имеет хорошее качество, является адекватной и ее

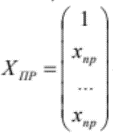
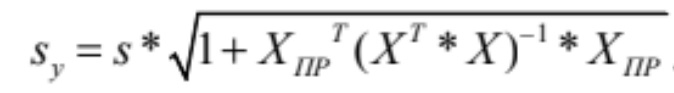
параметры соответствуют экономической теории, следовательно, она может быть использована

для прогноза. Рассчитаем прогноз объема выручки, если прогнозный объем

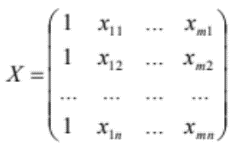
независимых факторов составит 115% от их среднего уровня. Для расчета средних значений

воспользуемся встроенной функцией СРЗНАЧ.

Прогнозный объем выручки составит . Рассчитаем ошибку прогноза. Для этого сформируем таблицу (рис. 3.7), в которой сформируем столбец Y62:Y110 из единиц и столбцы В45:С60 значений факторов , , а также вектор прогнозных значений AD60:AF60. Ошибка прогноза находится по формуле:

**

- - вектор-столбец прогнозных значений независимых факторов;



* матрица наблюдаемых значений m независимых переменных; s – стандартная ошибка.

Предельная ошибка прогноза:

Дополнительный интервал прогноза:

Произведем необходимые расчеты (см. рис. 3.7). Выделим ячейки AD63:AF65 и вызовем математическую функцию МУМНОЖ. В окошке Массив 1 введем формулу ТРАНСП(Y62:AA110), а в окошке Массив 2 – диапазон ячеек Y62:AA110. Далее нажмем сочетание клавиш Сtrl-Shift-ОК, которое всегда используется при работе с массивами. В результате будет рассчитан массив ячеек AD63:AF65. Выделим диапазон ячеек AD69:AF72 и вызовем математическую функцию МОБР. В качестве аргумента в окошке Массив введем адреса ячеек AD63:AF65. Снова воспользуемся сочетанием клавиш Сtrl-Shift-ОК. выделим ячейку AD75 и вызовем функцию МУМНОЖ, в которой в окошке Массив 1 введем диапазон AD61:AF61, а в окошке Массив 2 – формулу МУМНОЖ(AD51:AF71;ТРАНСП(AD61:AF61)). Не забудем про сочетание клавиш Сtrl-Shift-ОК. И, наконец, в ячейке AD78 найдем ошибку прогноза, воспользовавшись формулой=стандартная ошибка \*КОРЕНЬ(1+AD75). В ячейке AD81 рассчитаем предельную ошибку прогноза. В результате получим, что прогнозное значение фактора y находится в пределах 114,89, т.е. между значениями 430,3 и 200,52.

