

Задание для самостоятельной работы:

Целью лабораторной работы является изучение регрессионных моделей, построение регрессионной модели, проверка адекватности построенной регрессионной модели в RStudio.

- умение работать в RStudio;
- умение строить однофакторные регрессионные модели;
- умение строить многофакторные регрессионные модели;
- умение проверять адекватность построенной модели;
- умение проводить корректировку построенной модели.

Осваивается умение строить адекватные регрессионные линейные уравнения, проводить их мониторинг и давать качественную интерпретацию результатов моделирования.

Базовый уровень:

1. Построить однофакторную регрессионную модель, отражающую зависимость одного фактора от другого. Выборка должна включать не менее 70 наблюдений. В качестве метода нахождения оценок неизвестных параметров регрессионной модели использовать метод наименьших квадратов.

2. Проанализировать адекватность построенного однофакторного регрессионного уравнения: оценить качество подгонки; проверить различные гипотезы относительно параметров уравнения; проверить условия для получения состоятельных, несмещенных, эффективных оценок \tilde{a} и \tilde{b} ; провести содержательный анализ модели и корректировку модели (при необходимости), построить прогноз по модели.

3. Построить многофакторную регрессионную модель, отражающую зависимость одного фактора от нескольких других. Выборка должна включать не менее 70 наблюдений. В качестве метода нахождения оценок неизвестных параметров регрессионной модели использовать метод наименьших квадратов.

4. Проанализировать адекватность построенного многофакторного регрессионного уравнения: оценить качество подгонки; проверить различные гипотезы относительно параметров уравнения; проверить условия для получения состоятельных, несмещенных, эффективных оценок; исследовать наличие мультиколлинеарности, применить методы оценивания коэффициентов для ее устранения: гребневую регрессию и метод Lasso; провести содержательный анализ модели, построить прогноз по модели.

5. Результатом выполнения задания является отчет по лабораторной работе. К отчету предъявляются следующие требования:

- Четкое формулирование поставленной цели исследования (например: цель: определить какое влияние на уровень сахара в крови оказывают фактор «курение» / факторы посредством построения линейной регрессии)
- Формулирование задач, решение которых необходимо для достижения поставленной цели.
- Описание в виде пунктов, тех действий, которые требуются для решения поставленных задач. Все рисунки и таблицы последовательно нумеруются и описываются. Каждый пункт решения поставленных задач сопровождается анализом принятого решения. При проведении статистических тестов на определение статистической значимости модели в целом и отдельных ее параметров

обязательно выписывается нулевая и альтернативная гипотеза, указывается критическая область отказа от нулевой гипотезы в пользу альтернативной.

- В заключении выписывается уравнение регрессии и проводится его содержательный анализ и получившийся прогноз.

Повышенный уровень: Применить взвешенный метод наименьших квадратов для оценки параметров линейной регрессии.

Дополнительное задание (необязательное): Протестировать линейную регрессию на возможную точку излома структуры, используя тесты Чоу, Квандта-Эндрюса или Рамсея.

Данные можно найти в следующих источниках:

1. <https://www.kaggle.com/datasets>
2. <http://www.gks.ru/>
3. The R Datasets Package - <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.html>