**Моделирование статистических зависимостей**

**Домашнее задание 2**

Максимум- 10 баллов

Домашнее задание необходимо направить на почты [a.a.egorov07@gmail.com](mailto:a.a.egorov07@gmail.com) и [akhabirzanova@gmail.com](mailto:akhabirzanova@gmail.com). Допустимые форматы - R html или PDF файл с описанием результатов + скрипт файл с кодом, на основе которого были получены результаты. Задание 3 может быть выполнено «от руки», в данном случае скан должен быть интегрирован в PDF файл. **Работу необходимо сдать до 13.12.2019 23:59.**

**Задание 1**

По данным по Великобритании о потреблении цыплят (Y), среднедушевом доходе (X1), стоимости одного фунта цыплят (X2), стоимости одного фунта свинины (X3) и стоимости одного фунта говядины (X4) Необходимо построить, сравнить и проинтерпретировать уравнения регрессии вида:

1) (функция спроса)

2) (функция потребления)

3) (функция спроса-потребления)

4) (функция спроса с учетом цены на товары-заменители)

Исходные данный для исследования приведены в табл. 1

Таблица 1

Исходные данные

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **t** | **Y** | **X1** | **X2** | **X3** | **X4** |
| 1 | 31,2 | 492,9 | 37,3 | 54,7 | 77,4 |
| 2 | 33,3 | 528,6 | 38,1 | 63,7 | 80,2 |
| 3 | 35,6 | 560,3 | 39,3 | 69,8 | 80,4 |
| 4 | 36,4 | 624,6 | 37,8 | 65,9 | 83,9 |
| 5 | 36,7 | 666,4 | 38,4 | 64,5 | 85,5 |
| 6 | 38,4 | 717,8 | 40,1 | 70 | 93,7 |
| 7 | 40,4 | 768,2 | 38,6 | 73,2 | 106,1 |
| 8 | 40,3 | 843,3 | 39,8 | 67,8 | 104,8 |
| 9 | 41,8 | 911,6 | 39,7 | 79,1 | 114 |
| 10 | 40,4 | 931,1 | 52,1 | 95,4 | 124,1 |
| 11 | 40,7 | 1021,5 | 48,9 | 94,2 | 127,6 |
| 12 | 40,1 | 1165,9 | 58,3 | 123,5 | 142,9 |
| 13 | 42,7 | 1349,6 | 57,9 | 129,9 | 143,6 |
| 14 | 44,1 | 1449,4 | 56,5 | 117,6 | 139,2 |
| 15 | 46,7 | 1575,5 | 63,7 | 130,9 | 165,5 |
| 16 | 50,6 | 1759,1 | 61,6 | 129,8 | 203,3 |
| 17 | 50,1 | 1994,2 | 58,9 | 128 | 219,6 |
| 18 | 51,7 | 2258,1 | 66,4 | 141 | 221,6 |
| 19 | 52,9 | 2478,7 | 70,4 | 168,2 | 232,6 |

**Задание 2**

Используйте переменные и из задания 1. Постройте несколько полиномиальных регрессий используя полиномы разной степени. Выберите оптимальную степень полинома для моделирования данной зависимости с помощью ANOVA теста. Постройте график, на котором точками будут отмечены исходные данные, а линией – оцененная полиномиальная регрессия.

**Задание 3**

Задана случайная величина , распределение которой неизвестно. Из данной случайной величины была получена выборка . Найдите оценки в точках 5 и 10, используя:

а) Ядро Епанечникова

б) Гауссово ядро

**Задание 4**

Для выполнения этого задания используйте набор данных **“Smarket”** из библиотеки **“ISLR”** в R. Этот набор данных содержит информацию о доходности акций, входящих в индекс S&P 500 за период 1250 дней. Также данные содержат доходность, которая наблюдалась в 5 предыдущих дней, по отношению к текущему: *Lag1-Lag5*; объем проданных акций в предыдущий день: *volume;* а также бинарная переменная *direction*, которая отражает, рос или падал рынок в отдельно взятый день.

1) Постройте матрицу парных коэффициентов корреляции. Что можно сказать о взаимосвязи признаков?

2) Постройте модель бинарного выбора, которая бы предсказывала рост или падение рынка на основе данных о доходности в предыдущие 5 дней, а также на основе объема торгов в предыдущий день. Проинтерпретируйте результаты.

3) Посчитайте вероятность того, что рынок будет расти, для каждого из дней в выборке. Постройте предсказание на основе модели для каждого дня в выборке в терминах «рынок растет», «рынок падает». Посчитайте долю верных прогнозов.

**Задание 5**

Для выполнения этого задания используйте данные *hdi* (приложены к письму). Эти данные содержат информацию о значении индекса человеческого развития (<http://hdr.undp.org/en/content/human-development-index-hdi>), а также его компонентов. Используйте метод главных компонент для построения индекса человеческого развития на основе его компонентов. Сравните полученные результаты с фактическими значениями индекса.