

УРФУ

УРАЛЭНИН

КАФЕДРА

АВТОМАТИЗИРОВАННЫЕ ЭЛЕКТРИЧЕСКИЕ СИСТЕМЫ

**Методические указания
к выполнению
лабораторных работ
по дисциплине
«Математические основы оптимизации
развития электроэнергетических систем»**

Авторы:

А.С. ТАВЛИНЦЕВ

С.И. СЕМЕНЕНКО

Е.А. ЛЮХАНОВ



15 ноября 2021 г.

Оглавление

	Стр.
Введение	2
Лабораторная работа 1. Корреляционный анализ	4
1.1 Цель	4
1.2 Исходные данные	4
1.3 Задание	4
1.4 Отчёт	4
1.5 Рекомендация к работе	5
Лабораторная работа 2. Прогнозирование	6
2.1 Цель	6
2.2 Исходные данные	6
2.3 Задание	6
2.4 Отчёт	6
2.5 Рекомендация к работе	7
Лабораторная работа 3. Анализ модели	8
3.1 Цель	8
3.2 Исходные данные	8
3.3 Задание	8
3.4 Отчёт	8
3.5 Рекомендация к работе	9
3.5.1 F-тест	9
3.5.2 T-тест	10
3.5.3 Критерий Дарбина-Ватсона	10
3.5.4 Оценка доверительных интервалов	11
Литература	12
Список иллюстраций	13

Введение

Основной целью данных лабораторных работ является разработка программного кода для выполнения прогнозирования максимума мощности нагрузки по заданной выборке данных различных факторов.

Общие требования к программному коду:

1. Должен быть выполнен расчёт коэффициентов линейной корреляции.
2. Расчёт точечной оценки параметров модели.
3. Расчёт F-статистики (критерий Фишера), T-статистики (критерий Стьюдента) и критерия Дарбина-Ватсона.
4. Оценка доверительных интервалов параметров модели.
5. Оценка доверительных интервалов прогноза по модели.

В качестве результата работы программы должен быть краткий отчёт в текстовом или табличном виде. Пример текстового отчета, который выдаёт библиотека statsmodels, представлен на рисунке 1.

В рамках лабораторных работ для выполнения прогноза максимума электропотребления P рассматривать линейные модели следующего вида:

$$P = \sum_{i=1}^n a_i \cdot x_i,$$

где P — модельное значение максимума электропотребления; n — число параметров модели; a_i — параметр модели при i влияющем факторе; x_i — значение i влияющего фактора.

```

In [8]: mod = smf.ols(formula='Lottery ~ Literacy + Wealth + Region', data=df)

In [9]: res = mod.fit()

In [10]: print(res.summary())

```

OLS Regression Results

```

=====
Dep. Variable:          Lottery    R-squared:                0.338
Model:                  OLS        Adj. R-squared:           0.287
Method:                 Least Squares    F-statistic:            6.636
Date:                  Sat, 28 Nov 2020    Prob (F-statistic):     1.07e-05
Time:                  14:39:43          Log-Likelihood:         -375.30
No. Observations:      85              AIC:                   764.6
Df Residuals:          78              BIC:                   781.7
Df Model:              6
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	38.6517	9.456	4.087	0.000	19.826	57.478
Region[T.E]	-15.4278	9.727	-1.586	0.117	-34.793	3.938
Region[T.N]	-10.0170	9.260	-1.082	0.283	-28.453	8.419
Region[T.S]	-4.5483	7.279	-0.625	0.534	-19.039	9.943
Region[T.W]	-10.0913	7.196	-1.402	0.165	-24.418	4.235
Literacy	-0.1858	0.210	-0.886	0.378	-0.603	0.232
Wealth	0.4515	0.103	4.390	0.000	0.247	0.656

```

=====
Omnibus:                 3.049    Durbin-Watson:           1.785
Prob(Omnibus):           0.218    Jarque-Bera (JB):        2.694
Skew:                   -0.340    Prob(JB):               0.260
Kurtosis:                2.454    Cond. No.               371.
=====

```

Рис. 1 — Пример текстового отчета, генерируемого библиотекой Statsmodels (Python)

Лабораторная работа 1. Корреляционный анализ

1.1 Цель

Сформировать рекомендации по учёту влияющих факторов в модели на основе корреляционного анализа.

1.2 Исходные данные

Индивидуальные варианты заданий на курсовой проект.

1.3 Задание

Разработать программный код для выполнения корреляционного анализа и формирования рекомендации по результатам расчётов коэффициентов парной корреляции.

1.4 Отчёт

В отчёте представить:

1. Цель выполнения работы и решаемые задачи.
2. Краткое описание алгоритма.
3. Разработанный программный код.
4. Результаты расчётов с использованием заданных исходных данных.

Требования к оформлению отчёта соответствуют требованиям предъявляемым к курсовым и дипломным проектам. Структура разделов отчёта:

1. Задание и цель выполнения работы.
2. Краткое описание разработанной программы, с представлением необходимых математических формул.
3. Разработанный программный код.
4. Результаты расчётов с использованием заданных исходных данных.

1.5 Рекомендация к работе

Коэффициент линейной корреляции можно рассчитать по следующей формуле:

$$r_{x_1, x_2} = \frac{\sum_{i=1}^N \left((x_{1,i} - \mathbb{E}(x_1)) (x_{2,i} - \mathbb{E}(x_2)) \right)}{N \cdot \sigma_{x_1} \cdot \sigma_{x_2}},$$

где x_1 и x_2 — параметры между которыми выполняется оценка коэффициента линейной корреляции; $\mathbb{E}(x_j)$ — математическое ожидания параметра j , $j \in \{1, 2\}$; σ_{x_2} — среднеквадратическое отклонение j параметра; N — размер выборки.

В качестве практического условия значимости коэффициентов корреляции можно ориентироваться на практические критерии. Зависимость считать значимой, если $r_{x_1, x_2} > 0,8$

Лабораторная работа 2. Прогнозирование

2.1 Цель

Выполнить прогноз максимума электропотребления.

2.2 Исходные данные

Индивидуальные варианты заданий на курсовой проект.

2.3 Задание

Разработать программный код для

1. Формирования расчётной модели;
2. Выполнения прогноза максимума электропотребления на основе регрессионного анализа;
3. Расчёта ошибок моделирования по ретроспективным данным¹ значений максимума электропотребления.

2.4 Отчёт

В отчёте представить:

1. Цель выполнения работы и решаемые задачи.
2. Краткое описание алгоритма.
3. Разработанный программный код.
4. Результаты расчётов с использованием заданных исходных данных (в т.ч. в графическом виде).

Требования к оформлению отчёта соответствуют требованиям предъявляемым к курсовым и дипломным проектам. Структура разделов отчёта:

1. Задание и цель выполнения работы.
2. Краткое описание разработанной программы, с представлением необходимых математических формул.
3. Разработанный программный код.
4. Результаты расчётов с использованием заданных исходных данных.

¹Исходная выборка.

2.5 Рекомендация к работе

В качестве основного метода прогнозирования рекомендуется использовать метод наименьших квадратов [3].

Для расчётов рекомендуется использовать МНК в матричном виде:

$$\mathbf{A} = (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{B},$$

где \mathbf{A} — вектор коэффициентов линейной модели; \mathbf{X} — матрица ретроспективных значений влияющих факторов; \mathbf{B} — вектор значений максимума электропотребления в прошлом.

Ошибка моделирования оценивается следующим образом:

$$\varepsilon_i = P_{i,\text{факт}} - P_{i,\text{мод}},$$

где ε_i — ошибка моделирования i измерения максимума электропотребления, $i \in 1 \dots N$; N — глубина выборки; $P_{i,\text{факт}}$ — ретроспективное значение i максимума электропотребления; $P_{i,\text{мод.}}$ — модельное значение i максимума электропотребления.

Лабораторная работа 3. Анализ модели

3.1 Цель

Оценить статистическую значимость модели и её коэффициентов, а также рассчитать доверительные интервалы параметров модели и прогнозного значения максимума электропотребления.

3.2 Исходные данные

Индивидуальные варианты заданий на курсовой проект.

3.3 Задание

Разработать программный код для

1. Расчёта F-статистики (критерий Фишера), T-статистики (критерий Стьюдента) и критерия Дарбина-Ватсона;
2. Оценки доверительных интервалов параметров модели;
3. Оценки доверительных интервалов прогноза по модели.

3.4 Отчёт

В отчёте представить:

1. Цель выполнения работы и решаемые задачи.
2. Краткое описание алгоритма.
3. Разработанный программный код.
4. Результаты расчётов с использованием заданных исходных данных.

Требования к оформлению отчёта соответствуют требованиям предъявляемым к курсовым и дипломным проектам. Структура разделов отчёта:

1. Задание и цель выполнения работы.
2. Краткое описание разработанной программы, с представлением необходимых математических формул.
3. Разработанный программный код.
4. Результаты расчётов с использованием заданных исходных данных.

3.5 Рекомендация к работе

3.5.1 F-тест

Проверка статистической значимости выбранной модели прогнозирования основана использовании F-теста или критерия Фишера.

Несмещённая оценка дисперсии ошибки моделирования:

$$\sigma_\varepsilon^2 = D(\varepsilon) = \frac{\sum_{i=1}^N (\varepsilon_i)^2}{N - n},$$

где σ_x — среднее квадратичное отклонение случайной величины x ; $D(x)$ — дисперсия случайной величины x ; N — глубина выборки; n — число параметров модели, при $N > n$.

$$\sigma_P^2 = D(P) = \frac{\sum_{i=1}^N (P_i - \mathbb{E}(P))^2}{N - 1},$$

где P_i — фактическое (ретроспективное) значение максимума электропотребления в момент времени i , $i \in 1 \dots N$.

Тогда расчётное значение F-критерия можно оценить:

$$F_{\text{расч}} = \frac{D(P)}{D(\varepsilon)}$$

Для проверки значимости рассматриваемой модели выдвинем нулевую гипотезу о том, что модель не улавливает закономерностей случайного процесса и оценка дисперсии $D(P)$ (с числом степеней свободы $\lambda_P = N - 1$) исходной выборки соизмерима с дисперсией ошибки $D(\varepsilon)$ (с числом степеней свободы $\lambda_\varepsilon = N - n$).

Отношение дисперсий двух случайных величин подчиняется распределению Фишера [3]. Задавая близкий к 1 доверительный уровень $\beta \in \{0,95; 0,99\}$ можно определить по стандартному значению распределения Фишера $F_{\text{табл.}}$.

При значениях $F_{\text{расч}} > F_{\text{табл.}}(\beta, \lambda_P, \lambda_\varepsilon)$ дисперсия ошибки мала по сравнению с дисперсией показателя P и нулевая гипотеза отвергается, а следовательно, подтверждается гипотеза о виде модели.

В том случае, когда проверка по F-распределению удовлетворительна, модель принимается для дальнейшего исследования; если нет, то необходимо изменить или вид модели, или состав показателей для повышения адекватности модели. Плохая модель может быть также обусловлена наличием авторегрессионных связей. Под авторегрессионными связями понимается взаимозависимость ошибок соседних наблюдений.

В рамках данных лабораторных работ необходимо выдать расчётное значение $F_{\text{расч}}$ и вывод о значимости или не значимости рассматриваемой модели.

3.5.2 Т-тест

На данном этапе выполняется оценка статистической значимости коэффициентов модели. Если коэффициент модели не значим, то его можно исключить из модели.

Статистическая значимость коэффициентов проверяется на основе выдвижения нулевой гипотезы, отвергающей статистическую значимость коэффициентов модели.

Для проверки значимости можно воспользоваться величиной τ , определяемой как отношение случайной величины к её ошибке с $\nu = N - n$ степенями свободы.

Пусть

$$\tau_{i,\text{расч}} = \frac{a_i}{\sigma_{a_i}},$$

где σ_{a_i} ошибка коэффициента модели a_i , которая может быть оценена по ошибке моделирования σ_ε и коэффициенту обратной информационной матрицы:

$$\sigma_{a_i} = \sigma_\varepsilon \cdot \sqrt{(\mathbf{X}^T \cdot \mathbf{X})_{i,i}^{-1}},$$

где $(\mathbf{X}^T \cdot \mathbf{X})_{i,i}^{-1}$ — диагональный элемент обратной информационной матрицы¹, который соответствует i параметру модели.

Распределение случайной величины τ подчиняется распределению Стьюдента [3].

Выдвинем нулевую гипотезу, отвергающую статистическую значимость коэффициентов модели. Если нулевая гипотеза верна и коэффициенты модели незначимы, то их ошибка велика и статистически не-отличима от самого коэффициента, т. е. τ с высокой вероятностью мало (уровень значимости $\alpha = 0,05$ и менее, а уровень незначимости $\beta = 0,95$ и более, $\alpha + \beta = 1$). Уровень незначимости β задаёт вероятность с которой проверка нулевой гипотезы будет давать истинный результат.

Если теперь определить по стандартному распределению значение τ_{max} и сравнить его с расчетным значением, то можно утверждать, что при $\tau_{i,\text{расч}} < \tau_{max}$ нулевая гипотеза о незначимости коэффициентов подтверждается и данный коэффициент можно не учитывать, а если $\tau_{i,\text{расч}} > \tau_{max}$, то коэффициенты модели значимы.

3.5.3 Критерий Дарбина-Ватсона

В рамках данной работы проверку на отсутствия авторегрессии выполняется по критерию Дарбина-Ватсона:

$$D = 2 \cdot \left(1 - \frac{\sum_{i=2}^N (\varepsilon_i \cdot \varepsilon_{i-1})^2}{\sum_{i=1}^N (\varepsilon_i)^2} \right)$$

¹След матрицы.

Если значение критерия находится в интервале $1,5 < D < 2,5$, то авторегрессия отсутствует; в противном случае ошибки соседних наблюдений ε_i и ε_{i-1} нельзя считать независимыми и следует учесть наличие авторегрессии.

3.5.4 Оценка доверительных интервалов

Для корректного описания модели необходимо получить оценку доверительных интервалов коэффициентов модели и прогноза.

Если коэффициент модели значим, то его ошибка не превосходит величину:

$$|\pm \Delta a_i| \leq \sigma_\varepsilon \cdot \tau_{max}.$$

При этом максимальное значение доверительного интервала с вероятностью $\beta = 0,95 \dots 0,99$ и числом степеней свободы $\nu = N - n$ равно

$$\Delta a_i = \sigma_\varepsilon \cdot \tau_{max}.$$

Доверительный интервал прогноза на момент времени j можно оценить:

$$\sigma_j = \sigma_\varepsilon \cdot \sqrt{1 + \mathbf{X}_j \cdot (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}_j^T},$$

где \mathbf{X}_j — строка значений влияющих факторов в момент времени j .

$$\mathbf{X}_j = \begin{bmatrix} x_{j,1} & x_{j,2} & \dots & x_{j,n} \end{bmatrix}$$

Тогда оценку доверительного интервала прогнозного значения максимума электропотребления можно выполнить:

$$\Delta P_j = \sigma_j \cdot \tau_{max}, \text{ при } \beta = 0,95 \dots 0,99 \text{ и } \nu = N - n.$$

Литература

- [1] Мызин Модели оптимизации развития энергосистем / Д. А. Арзамасцев, А. В. Липес, А. Л. – М.:Высшая школа, 1987.
- [2] Модели развития электроэнергетических систем : учебное пособие / С. С. Ананичева, П. Е. Мезенцев, А. Л. Мызин. – Екатеринбург : УрФУ, 2014. – 148 с. ISBN 978-5-321-02313-6
- [3] Гмурман В. Е. Теория вероятностей и математическая статистика / В. Е. Гмурман. М.: Высшая школа, 1997. 479 с
- [4] <http://www.machinelearning.ru/>

Список иллюстраций

- | | | |
|---|--------------------------------------------------------------------------|---|
| 1 | Пример текстового отчета, генерируемого библиотекой Statsmodels (Python) | 3 |
|---|--------------------------------------------------------------------------|---|