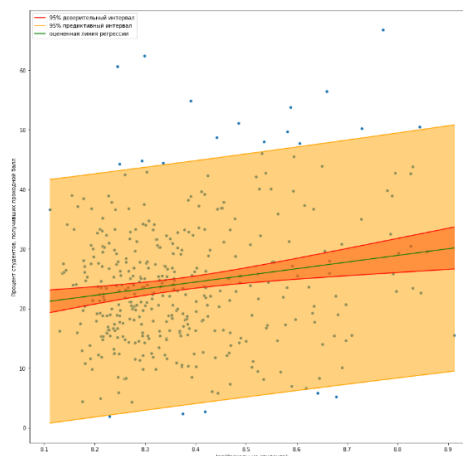


## Эконометрика. Домашнее задание 1.

Выполнил Авилов Илья, БЭК 181

А) Исследователь хочет оценить взаимосвязь между проходным баллом и расходом на студента. (Проходной балл интерпретировался как агрегированный средний балл на каких-то внутриуниверситетских экзаменах среди студентов данного ВУЗа.)

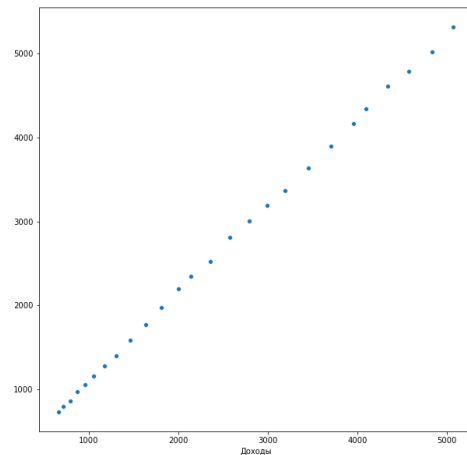
- Исследовательский вопрос: Приводит ли увеличение расходов на студента к увеличению процента студентов, получивших проходной балл?  
Гипотеза: Увеличение расходов на студента увеличивает процент студентов, получивших проходной балл.
- Функциональная спецификация модели  $\widehat{math10} = \hat{\alpha} + \hat{\beta} \times \ln \text{expend}$ . Где  $\widehat{math10}$  – процент студентов, получивших проходной балл, а  $\text{expend}$  – расходы на студента
- Была выбрана данная модель, т.к. логично предположить, что при увеличении расходов будет происходить нелинейное изменение процента студентов, получивших проходной балл. Так, при низких расходах на студента, их увеличение сильно скажется на проценте студентов, получивших проходной балл, в то время как при высоких расходах на студента, их увеличение не сильно повлияет на объясняемую переменную.
- Факторы, входящие в ошибку: психологическое состояние студентов (их мотивированность обучаться, настроение и т.д.), сложность экзаменов. Эти факторы могут влиять на количество студентов, получивших проходной балл.
- Оцененное уравнение регрессии:  $\widehat{math10} = -69.3411 + 11.1644 \times \ln \text{expend}$
- Интерпретация коэффициента  $\hat{\beta}$ : При увеличении расходов на студента на 1%, процент студентов, получивших проходной балл, увеличится на 0.1116
- Средние расходы на студента: 4376.578  
Ожидаемый средний процент студентов, получивших проходной балл при средних расходах на студента: 24.26  
Медианные расходы на студента: 4145  
Ожидаемый средний процент студентов, получивших проходной балл при медианных расходах на студента: 23.65
- Эластичность при среднем значении объясняющей переменной: 0.46  
Предельный эффект при среднем значении объясняющей переменной: 0.00255  
Эластичность при медианном значении объясняющей переменной: 0.472  
Предельный эффект при медианном значении объясняющей переменной: 0.00269
- График доверительного и предиктивного интервалов, диаграмма рассеяния и линия оцененной регрессии:



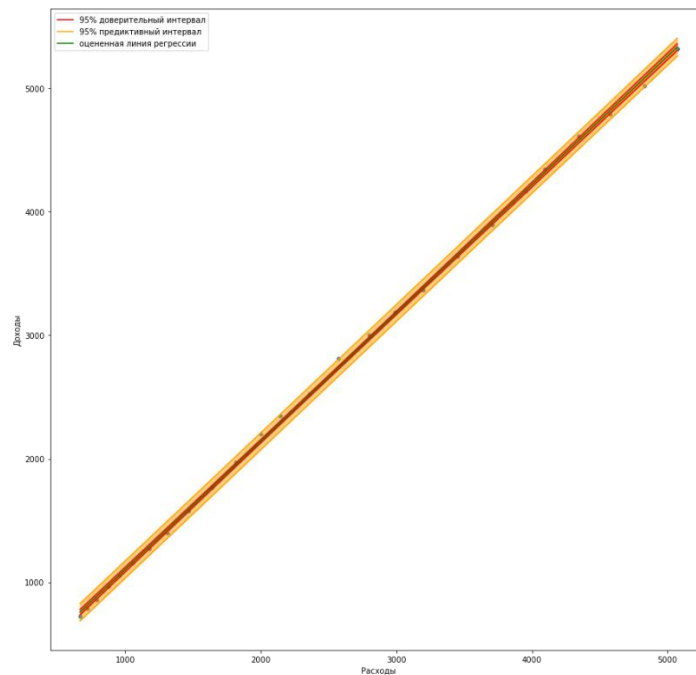
В) Исследователь хочет оценить предельную склонность к потреблению на основании наблюдений о сбережениях и доходе.

- Исследовательский вопрос: Склонны ли домохозяйства потреблять больше, чем зарабатывают?  
Гипотеза: Домохозяйства тратят меньше, чем зарабатывают (то есть предельная склонность к потреблению  $< 1$ ).
- Предельная склонность к потреблению:  $MPC = \frac{dC}{dY}$ . Оценим предельную склонность к потреблению с помощью регрессии вида

- Функциональная спецификация модели  $\hat{C} = \hat{\alpha} + \hat{\beta} \times Y$  В этой модели оцененный коэффициент  $\hat{\beta}$  будет предельной склонностью к потреблению.
- Была взята модель lin-lin, т.к. на графиках наблюдается строгая линейная зависимость. Логично было выбрать модель Lin-log, т.к. домохозяйства с очень большим доходом скорее всего не начинают тратить сильно больше, при увеличении дохода, а начинают сберегать, но видимо в данном датасете таких домохозяйств не представлено, поэтому lin-log модель не сможет адекватно описать такую зависимость.

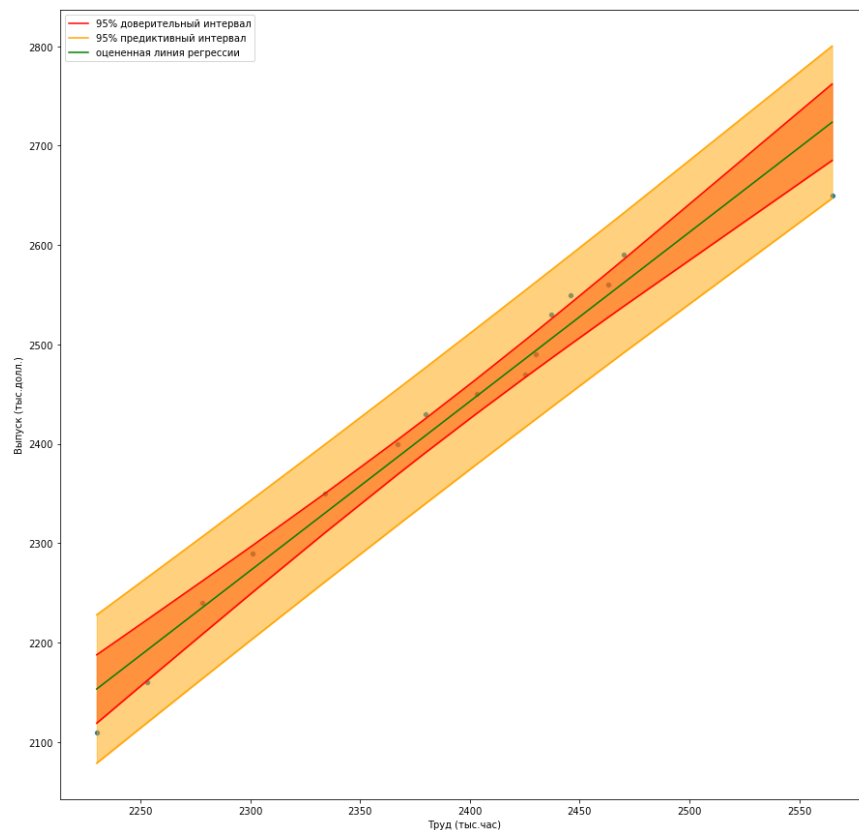


- Факторы, входящие в ошибку: большое разнообразие домохозяйств (при одном и том же уровне дохода одни домохозяйства могут тратить сильно больше или меньше других), ошибки измерения (не все доходы и расходы декларируются).
- Оцененное уравнение регрессии:  $\hat{C} = -62.4227 + 0.9623 \times Y$
- Интерпретация коэффициента  $\hat{\beta}$ : При увеличении дохода домохозяйства на 1, потребление вырастет на 0.9623. То есть предельная склонность к потреблению = 0.9623
- Средние доходы домохозяйств: 2645.119  
Ожидаемое среднее потребление при среднем доходе домохозяйства: 2483.030  
Медианные доходы домохозяйств: 2434.85  
Ожидаемое среднее потребление при медианном доходе домохозяйства: 2280.68
- Эластичность при среднем значении объясняющей переменной: 1.025  
Предельный эффект (предельная склонность к потреблению) при среднем значении объясняющей переменной: 0.962  
Эластичность при медианном значении объясняющей переменной: 1.027  
Предельный эффект (предельная склонность к потреблению) при медианном значении объясняющей переменной: 0.962
- График доверительного и предиктивного интервалов, диаграмма рассеяния и линия оцененной регрессии:



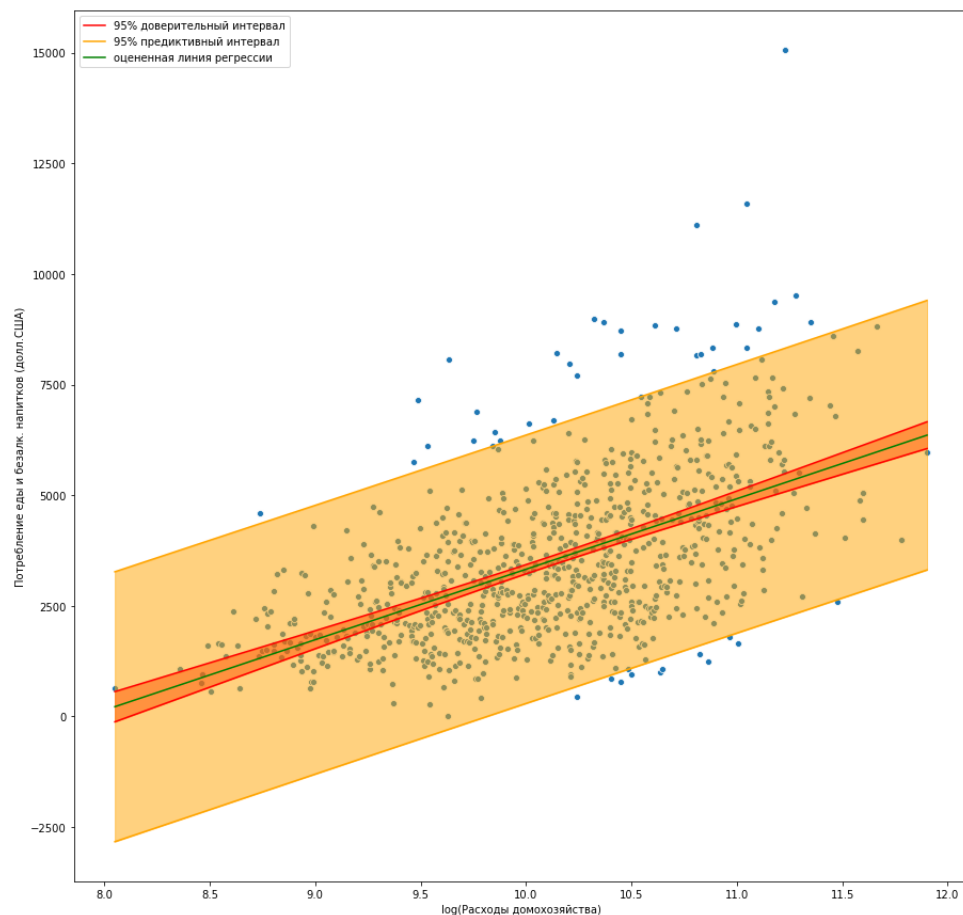
С) Исследователь оценивает зависимость выпуска от вложенных объемов труда и капитала, используя набор данных *prfunc.dta*.

- Исследовательский вопрос: Вызывает ли увеличение объемов вложенного труда и капитала увеличение выпуска? Гипотеза: Увеличение объемов труда и вложенного капитала приводит к увеличению объемов выпуска
- Функциональная спецификация модели  $\hat{Y} = \alpha + \beta \times L$ . Где  $Y$  – объем выпуска(тыс.дол), а  $L$  – объем труда(тыс.час).
- Логично было бы выбрать функцию Кобба-Дугласа, прологорифмировать ее и оценить необходимые коэффициенты с помощью линейной регрессии, однако между регрессорами (объемы труда и капитала) присутствует сильная корреляция (коэф. Корреляция Пирсона 0.99). Логарифмы этих переменных также сильно скорелированы (0.99). Если использовать оба этих регрессора, то возникнет проблема мультиколлинеарности, что приведет к большой дисперсии коэффициентов регрессии и затруднит их интерпретацию. Поэтому было решено оставить лишь 1 регрессор (объем труда). Также по графику видно, что между объемом труда и объемом выпуска присутствует сильная линейная связь, то есть фирмы из данного датасета не сталкиваются с убывающим предельным продуктом, поэтому объем труда не был прологорифмирован.
- Факторы, входящие в ошибку: различия в технологиях производства (при одинаковом количестве труда разные фирмы могут иметь разный выпуск), различие в менеджменте (способ управления ресурсами может влиять на объем производства)
- Оцененное уравнение регрессии:  $\hat{Y} = -1642.9401 + 1.7024 \times L$
- Интерпретация коэффициента  $\hat{\beta}$ : При увеличении объемов труда на 1 тыс.час. выпуск увеличивается на 1.7024 тысячи долларов
- Средние объемы труда: 2385.47  
Ожидаемый средний выпуск при среднем объеме труда: 2418  
Медианные объемы труда: 2403  
Ожидаемый средний выпуск при медианном объеме труда: 2447.85
- Эластичность при среднем значении объясняющей переменной: 1.679  
Предельный эффект при среднем значении объясняющей переменной: 1.702  
Эластичность при медианном значении объясняющей переменной: 1.671  
Предельный эффект при медианном значении объясняющей переменной: 1.702
- График доверительного и предиктивного интервалов, диаграмма рассеяния и линия оцененной регрессии:



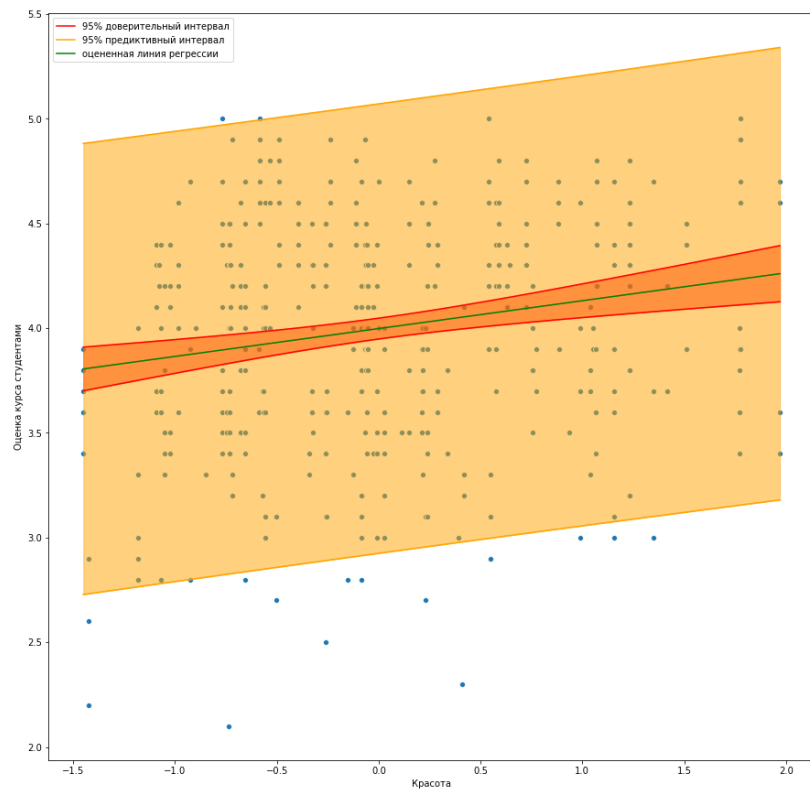
D) Исследователь оценивает зависимость потребления еды и безалкогольных напитков от доходов домохозяйств.

- Исследовательский вопрос: Увеличивается ли потребление еды и безалкогольных напитков с ростом доходов домохозяйств?  
Гипотеза: Рост доходов ведет к увеличению потребления еды и безалкогольных напитков.
- Функциональная спецификация модели  $\widehat{fdho} = \hat{\alpha} + \hat{\beta} \times \ln \text{expend}$ . Где fdho – потребление еды и безалкогольных напитков, а expend – общие расходы.
- Т.к. исследователь хочет оценить зависимость потребления еды и безалкогольных напитков от доходов, то было бы логично использовать в качестве регрессора доход, однако данных о доходах нет. Поэтому предполагая, что доход положительно влияет общее потребление, мы будем оценивать влияние общего потребления на потребление еды и безалкогольных напитков. Была выбрана lin-log модель, т.к. предполагается, что при очень больших общих расходах домохозяйства не начинают потреблять сильно больше еды и безалкогольных напитков при большем росте общих расходов.
- Факторы, входящие в ошибку: разные предпочтения у домохозяйств (при одном и том же доходе домохозяйства будут потреблять еду и безалкогольные напитки в разных объемах).
- Оцененное уравнение регрессии:  $\hat{Y} = -12600 + 1593.0363 \times \ln L$
- Интерпретация коэффициента  $\hat{\beta}$ : При увеличении общих расходов 1%, потребление еды и безалкогольных напитков увеличивается на 15.93 долл. США
- Средние расходы домохозяйств: 29932.447  
Ожидаемое среднее потребление еды и безалк. напитков при средних расходах домохозяйства: 3816.799  
Медианные расходы домохозяйств: 25582.8  
Ожидаемое среднее потребление еды и безалк. напитков при медианных расходах домохозяйства: 3566.65
- Эластичность при среднем значении объясняющей переменной: 0.417  
Предельный эффект при среднем значении объясняющей переменной: 0.053  
Эластичность при медианном значении объясняющей переменной: 0.447  
Предельный эффект при медианном значении объясняющей переменной: 0.062
- График доверительного и предиктивного интервалов, диаграмма рассеяния и линия оцененной регрессии:



Е) Исследователь оценивает зависимость рейтинга курсов среди студентов от внешней привлекательности. (Внешняя привлекательность интерпретировалась как внешняя привлекательность преподавателя, проводившего курс).

- Исследовательский вопрос: Получают ли курсы с привлекательными преподавателями более высокие оценки от студентов?  
Гипотеза – привлекательная внешность преподавателя положительно влияет на оценку курса студентами.
- Функциональная спецификация модели  $\widehat{courseval} = \alpha + \beta \times beauty$ .
- Была выбрана lin lin модель, т.к. предполагается, что при разных уровнях красоты ее увеличение будет приводить к одинаковому увеличению ожидаемой средней оценки курса.
- Факторы, входящие в ошибку: красота – вещь субъективная, поэтому при ее оценке будет возникать ошибка, т.к. нет четких критериев красоты. На разных курсах студенты скорее всего будут уделять разное внимание внешности преподавателя (на важном для карьеры студента курсе (например мат.анализ для математиков) студентов вряд ли будет волновать внешность преподавателя, а на, например, физкультуре, возможно будет волновать).
- Оцененное уравнение регрессии:  $\widehat{courseval} = 3.9983 + 0.1330 \times beauty$
- Интерпретация коэффициента  $\hat{\beta}$ : При увеличении оценки внешней привлекательности на 1 средняя ожидаемая оценка курса увеличится на 0.1330. При этом если внешняя привлекательность была оценена на 0, то средняя ожидаемая оценка курса = 3.9983
- Средняя оценка красоты:  $\sim 0$   
Ожидаемая средняя оценка курса студентами при средней оценке внешней привлекательности: 3.9983  
Медианная оценка красоты: -0.0680143  
Ожидаемая средняя оценка курса студентами при медианной оценке внешней привлекательности 3.9892
- Эластичность при среднем значении объясняющей переменной:  $\sim 0$   
Предельный эффект при среднем значении объясняющей переменной: 0.133  
Эластичность при медианном значении объясняющей переменной: -0.0025  
Предельный эффект при медианном значении объясняющей переменной: 0.133
- График доверительного и предиктивного интервалов, диаграмма рассеяния и линия оцененной регрессии:



## Код на Python:

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
from statsmodels.iolib.regression_pearson import wls_prediction_std
from scipy.stats import t

class hw_X():
    def __init__(self, X, y, model_type = 'lin lin'):
        self.model_type = model_type
        self.X_initial = X
        self.y_initial = y
        if model_type == 'lin lin':
            self.X = X
            self.y = y
        elif model_type == 'lin log':
            self.X = np.log(X)
            self.y = y
        elif model_type == 'log lin':
            self.X = np.log(X)
            self.y = np.log(X)
        elif model_type == 'log log':
            self.X = np.log(X)
            self.y = np.log(y)
        else:
            print("Введите model_type из списка ['lin lin', 'log log', 'log lin', 'lin log']")

        self.y = y[np.argsort(self.X)]
        self.X = np.sort(self.X)
        self.y_label = model_type.split(':')[0]
        self.x_label = model_type.split(':')[1]

    def fit(self):
        self.X = sm.add_constant(self.X)
        model = sm.OLS(self.y, self.X)
        self.results = model.fit()
        print(self.results.summary())

        self.alpha = self.results.params[0]
        self.betta = self.results.params[1]

    def calc_intervals(self):
        # y горизонталь
        y_var = self.results.cov_params()[0][0] + 2*self.X[:, 1]*self.results.cov_params()[1][0] \
            + self.X[:, 1]**2*self.results.cov_params()[1][1]
        # y вертикальная
        df = self.X.shape[0] - self.X.shape[1] # степени свободы
        t_stat_bounds = t.interval(0.95, df = df)
        self.y_lower_bound = self.results.predict() + t_stat_bounds[0]*np.sqrt(y_var)
        self.y_upper_bound = self.results.predict() + t_stat_bounds[1]*np.sqrt(y_var)
        # OLS residuals
        residuals = self.results.resid
        residuals_var = np.sum((residuals)**2)/(len(residuals)-2)
        print(residuals_var)
        # y вертикальная, горизонтальная
        self.y_lower_bound_predictiv = self.results.predict() + t_stat_bounds[0]*np.sqrt(residuals_var + y_var)
        self.y_upper_bound_predictiv = self.results.predict() + t_stat_bounds[1]*np.sqrt(residuals_var + y_var)

    def print_intervals(self, x_axis_name, y_axis_name):
        plt.figure(figsize = (15, 10))
        - sns.lmplot(x = self.X[:, 1], y = self.y_lower_bound, color = 'red')
        - sns.lmplot(x = self.X[:, 1], y = self.y_upper_bound, color = 'red', label = "95% доверительный интервал")
        - sns.lmplot(x = self.X[:, 1], y = self.y_lower_bound_predictiv, color = 'orange')
        - sns.lmplot(x = self.X[:, 1], y = self.y_upper_bound_predictiv, color = 'orange', label = "95% предиктивный интервал")

        - sns.lmplot(x = self.X[:, 1], y = self.results.predict(), color = 'green', label = "оценочная линия регрессии")
        - sns.scatterplot(x = self.X[:, 1], y = self.y)

        - plt.fill_between(self.X[:, 1], self.y_upper_bound, self.y_lower_bound, color = 'red', alpha = 0.5)
        - plt.fill_between(self.X[:, 1], self.y_upper_bound_predictiv, self.y_lower_bound_predictiv, color = 'orange',
            alpha = 0.5)

        if self.x_label == 'log':
            - plt.xlabel('log(' + x_axis_name + ')')
        else:
            - plt.xlabel(x_axis_name)
        if self.y_label == 'log':
            - plt.ylabel('log(' + y_axis_name + ')')
        else:
            - plt.ylabel(y_axis_name)

    def calc_elasticities(self, X):
        if self.model_type == 'lin lin':
            ME = self.betta
            E = self.betta * X/self.results.predict([1, X])
        elif self.model_type == 'log log':
```

```

        ME = self.betta * np.exp(self.results.predict([1, np.log(X)])) / X
        E = self.betta

    elif self.model_type == 'log lin':
        ME = self.betta * np.exp(self.results.predict([1, X]))
        E = self.betta * X

    elif self.model_type == 'lin log':
        ME = self.betta/X
        E = self.betta/self.results.predict([1, np.log(X)])
    return(ME, E)

```

Пункт А

```

data_a = pd.read_stata('HEAP93.dta')

# Регрессоры
X = data_a['expend'].values
# Объясняемая переменная
y = data_a['math10'].values
# Индикаторный класс
point_a = hu_1(X, y, 'lin log')
# Функция модели
point_a.fit()
# Числен объясняемый и предсказанный интервал
point_a.calc_intervals()
# График интервалов, линии регрессии и скатерплов
point_a.print_intervals('Расходы на студента', 'Процент студентов, получивших проходной балл')
# Числен эластичности

print('Среднее значение регрессора', np.mean(X))
print('Предсказанное значение при среднем регрессоре', point_a.results.predict([1, np.log(np.mean(X))]))
print('Предельный эффект и эластичность при среднем регрессоре', point_a.calc_elasticities(np.mean(X)))

print('Медианное значение регрессора', np.median(X))
print('Предсказанное значение при медианном регрессоре', point_a.results.predict([1, np.log(np.median(X))]))
print('Предельный эффект и эластичность при медианном регрессоре', point_a.calc_elasticities(np.median(X)))

=====
OLS Regression Results
=====
Dep. Variable: y R-squared: 0.838
Model: OLS Adj. R-squared: 0.827
Method: Least Squares F-statistic: 12.41
Date: Mon, 28 Sep 2009 Prob (F-statistic): 0.000475
Time: 13:58:51 Log-Likelihood: -1531.4
No. Observations: 488 AIC: 3067.
Df Residuals: 486 BIC: 3075.
Df Model: 1
Covariance Type: nonrobust
=====
coef std err t P>|t| [0.025 0.975]
-----
const -69.9411 26.538 -2.634 0.009 -121.495 -17.188
x1 11.5644 3.169 3.523 0.000 4.935 17.394
=====
Omnibus: 28.397 Durbin-Watson: 1.992
Prob(Omnibus): 0.000 Jarque-Bera (JB): 34.689
Skew: 0.551 Prob(JB): 3.05e-08
Kurtosis: 3.880 Cond. No. 440.
=====

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
107.1126978379386
Среднее значение регрессора 4376.57841137248
Предсказанное значение при среднем регрессоре [24.26144043]
Предельный эффект и эластичность при среднем регрессоре (0.0025589413147627273, array([0.46017032]))
Медианное значение регрессора 4165.0
Предсказанное значение при медианном регрессоре [23.65449425]
Предельный эффект и эластичность при медианном регрессоре (0.002693468732941457, array([0.47197774]))

```

png

png

Пункт В

```

data_a = pd.read_stata('sav_inc.dta')

# Регрессоры
X = data_a['income'].values
# Объясняемая переменная
y = (data_a['income']-data_a['savings']).values

plt.figure(figsize = (10, 10))
_ = sns.scatterplot(X, y)
_ = plt.xlabel('Потребление')
_ = plt.ylabel('Доходы')

png

png

# Индикаторный класс
point_a = hu_1(X, y, 'lin lin')
# Функция модели
point_a.fit()
# Числен объясняемый и предсказанный интервал
point_a.calc_intervals()
# График интервалов, линии регрессии и скатерплов
point_a.print_intervals('Расходы', 'Доходы')
# Числен эластичности

print('Среднее значение регрессора', np.mean(X))
print('Предсказанное значение при среднем регрессоре', point_a.results.predict([1, np.mean(X)]))
print('Предельный эффект и эластичность при среднем регрессоре', point_a.calc_elasticities(np.mean(X)))

print('Медианное значение регрессора', np.median(X))
print('Предсказанное значение при медианном регрессоре', point_a.results.predict([1, np.median(X)]))
print('Предельный эффект и эластичность при медианном регрессоре', point_a.calc_elasticities(np.median(X)))

```

```

=====
OLS Regression Results
=====
Dep. Variable: y R-squared: 1.000
Model: OLS Adj. R-squared: 1.000
Method: Least Squares F-statistic: 5.168e+04
Date: Mon, 28 Sep 2009 Prob (F-statistic): 1.05e-41
Time: 13:58:53 Log-Likelihood: -125.24
No. Observations: 26 AIC: 254.5
Df Residuals: 24 BIC: 257.8
Df Model: 1
Covariance Type: nonrobust
=====
coef std err t P>|t| [0.025 0.975]
-----
const -62.4227 12.761 -4.892 0.000 -88.768 -36.886
x1 0.9623 0.006 227.146 0.000 0.954 0.971
=====
Omnibus: 1.662 Durbin-Watson: 0.868
Prob(Omnibus): 0.436 Jarque-Bera (JB): 1.171
Skew: -0.515 Prob(JB): 0.557
Kurtosis: 2.863 Cond. No. 6.30e+03
=====

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 6.3e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
968.6799438824742
Среднее значение регрессора 2645.119
Предсказанное значение при среднем регрессоре [2483.03046981]
Предельный эффект и эластичность при среднем регрессоре (0.9623208864854715, array([1.02513972]))
Медианное значение регрессора 2434.85
Предсказанное значение при медианном регрессоре [2288.68442292]
Предельный эффект и эластичность при медианном регрессоре (0.9623208864854715, array([1.02737016]))

```

png

png

Пункт С

```

data_a = pd.read_stata('prhinc.dta')
data_a = data_a.drop('l', axis = 'columns')

```

```

data_a.corr()

```

q

l

k

q

```
1.00000
0.982740
0.965160
l
0.98274
1.000000
0.992491
k
0.96516
0.992491
1.000000
```

```
data_a['log(q)'] = np.log(data_a['q'])
data_a['log(l)'] = np.log(data_a['l'])
data_a['log(k)'] = np.log(data_a['k'])
```

```
data_a[['log(q)', 'log(l)', 'log(k)']].corr()
```

```
log(q)
log(l)
log(k)
1.000000
0.981894
0.982806
log(l)
0.981894
1.000000
0.992118
log(k)
0.982806
0.992118
1.000000
```

```
# Регрессор
X = data_a[1:].values
# Объясненная переменная
y = data_a[0].values
# Инициализируем класс
point_a = hp.M(X, y, 'lin lin')
# Обучаем модель
point_a.fit()
# Численность наблюдений и предельный интервал
point_a.calc_intervals()
# Группы интервалов, линия регрессии и scatterplot
point_a.print_intervals('Группа (тыс.час)', 'Выпуск (тыс.долл.)')
# Численность значимости
print('Среднее значение регрессора', np.mean(X))
print('Предказанное значение при среднем регрессоре', point_a.results.predict([1, np.mean(X)]))
print('Предельный эффект и эластичность при среднем регрессоре', point_a.calc_elasticities(np.mean(X)))
print('Медианное значение регрессора', np.median(X))
print('Предказанное значение при медианном регрессоре', point_a.results.predict([1, np.median(X)]))
print('Предельный эффект и эластичность при медианном регрессоре', point_a.calc_elasticities(np.median(X)))
```

```
C:\Users\Shhire\Anaconda3\lib\site-packages\scipy\stats\stats.py:1684: UserWarning: kurtosistest only valid for n>20 ... continuing anyway, n=15
"anyway, n=%i" % int(n))
```

```
=====
OLS Regression Results
=====
Dep. Variable: y R-squared: 0.966
Model: OLS Adj. R-squared: 0.963
Method: Least Squares F-statistic: 366.9
Date: Mon, 28 Sep 2020 Prob (F-statistic): 6.59e-11
Time: 13:58:54 Log-likelihood: -71.554
No. Observations: 15 AIC: 147.1
Df Residuals: 13 BIC: 148.5
Df Model: 1
Covariance Type: nonrobust
=====
coef std err t P>|t| [0.025 0.975]
-----
const -1642.9401 212.165 -7.744 0.000 -2181.294 -1184.586
x1 1.7024 0.009 19.254 0.000 1.510 1.894
=====
Omnibus: 6.981 Durbin-Watson: 1.179
Prob(Omnibus): 0.490 Jarque-Bera (JB): 4.111
Skew: -1.242 Prob(SB): 0.128
Kurtosis: 3.639 Cond. No. 6.48e+04
=====
```

```
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 6.4e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
939.9817350387928
Среднее значение регрессора 2385.4666666666667
Предказанное значение при среднем регрессоре [2448.]
Предельный эффект и эластичность при среднем регрессоре (1.7023671583118398, array([1.67946241]))
Медианное значение регрессора 2483.0
Предказанное значение при медианном регрессоре [2447.84817084]
Предельный эффект и эластичность при медианном регрессоре (1.7023671583118398, array([1.67117729]))
```

```
png
```

```
png
```

Пункт D

```
data_a = pd.read_stata('foodexp.dta')
```

```
# Регрессор
X = data_a['expend'].values
# Объясненная переменная
y = data_a['lnho'].values
```

```
sns.scatterplot(np.log(X), y)
```

<AxesSubplot>

```
png
```

```
png
```

```
# Инициализируем класс
point_a = hp.M(X, y, 'lin log')
# Обучаем модель
point_a.fit()
# Численность наблюдений и предельный интервал
point_a.calc_intervals()
# Группы интервалов, линия регрессии и scatterplot
point_a.print_intervals('Расходы домохозяйства', 'Потребление еды и безалк. напитков (долл.США)')
# Численность значимости
print('Среднее значение регрессора', np.mean(X))
print('Предказанное значение при среднем регрессоре', point_a.results.predict([1, np.log(np.mean(X))]))
print('Предельный эффект и эластичность при среднем регрессоре', point_a.calc_elasticities(np.mean(X)))
print('Медианное значение регрессора', np.median(X))
print('Предказанное значение при медианном регрессоре', point_a.results.predict([1, np.log(np.median(X))]))
print('Предельный эффект и эластичность при медианном регрессоре', point_a.calc_elasticities(np.median(X)))
```

```
=====
OLS Regression Results
=====
Dep. Variable: y R-squared: 0.910
Model: OLS Adj. R-squared: 0.910
Method: Least Squares F-statistic: 390.2
Date: Mon, 28 Sep 2020 Prob (F-statistic): 5.99e-72
Time: 14:34:08 Log-likelihood: -7022.4
No. Observations: 869 AIC: 1.52e+04
Df Residuals: 867 BIC: 1.524e+04
Df Model: 1
Covariance Type: nonrobust
=====
coef std err t P>|t| [0.025 0.975]
-----
const -3.26e+04 836.773 -15.420 0.000 -1.42e+04 -1.1e+04
x1 1593.0363 80.650 19.752 0.000 1434.743 1751.329
=====
Omnibus: 147.850 Durbin-Watson: 1.940
Prob(Omnibus): 0.000 Jarque-Bera (JB): 138.150
Skew: 0.000 Prob(SB): 1.69e-78
Kurtosis: 5.566 Cond. No. 159.
=====
```

```
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
2384241.7208349356
Среднее значение регрессора 29932.447
Предказанное значение при среднем регрессоре [3816.79900754]
Предельный эффект и эластичность при среднем регрессоре (0.05322105097888594, array([0.41737495]))
Медианное значение регрессора 25852.8
```



Предсказанное значение при медианном регрессоре [3566.65498352]  
Предельный эффект и эластичность при медианном регрессоре (0.06226981617338196, array([0.44664716]))

png

png

Python

```
data_a = pd.read_stata('beauty.dta')

data_a['course_eval'].isnull().sum()

0

# Регрессоры
X = data_a['beauty'].values
# Объясняемая переменная
y = data_a['course_eval'].values
# Метод оценивания
point_a = lm_2(X, y, 'lin lin')
# Оценка модели
point_a.fit()
# Численные (аналитический и предельный) интервалы
point_a.calc_intervals()
# График интервалов, линии регрессии и скалярных
point_a.print_intervals('график', 'Оценка курса студентам')
# Числовая эластичность

print('Среднее значение регрессора', np.mean(X))
print('Предсказанное значение при среднем регрессоре', point_a.results.predict([1, np.mean(X)]))
print('Предельный эффект и эластичность при среднем регрессоре', point_a.calc_elasticities(np.mean(X)))

print('Медианное значение регрессора', np.median(X))
print('Предсказанное значение при медианном регрессоре', point_a.results.predict([1, np.median(X)]))
print('Предельный эффект и эластичность при медианном регрессоре', point_a.calc_elasticities(np.median(X)))
```

OLS Regression Results

Dep. Variable:	y	R-squared:	0.836
Model:	OLS	Adj. R-squared:	0.834
Method:	Least Squares	F-statistic:	17.08
Date:	Mon, 28 Sep 2020	Prob (F-statistic):	4.25e-05
Time:	15:08:41	Log-Likelihood:	-375.32
No. Observations:	463	AIC:	754.6
DF Residuals:	461	BIC:	762.9
DF Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	3.9983	0.025	157.727	0.000	3.948	4.048
x1	0.1130	0.032	4.133	0.000	0.049	0.156

Omnibus: 15.299 Durbin-Watson: 1.300  
Prob(Omnibus): 0.000 Jarque-Bera (JB): 16.485  
Skew: -0.453 Prob(JB): 0.000274  
Kurtosis: 2.831 Cond. No. 1.27

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.  
0.2975179795498137  
Среднее значение регрессора 6.5912004e-08  
Предсказанное значение при среднем регрессоре [3.99827214]  
Предельный эффект и эластичность при среднем регрессоре (0.1330801449534438, array([2.19256887e-09]))  
Медианное значение регрессора -0.0080143  
Предсказанное значение при медианном регрессоре [3.98922613]  
Предельный эффект и эластичность при медианном регрессоре (0.133081449534438, array([-0.00226761]))

png

png