

<https://www.kaggle.com/fedesoriano/the-boston-houseprice-data>

Input features in order:

- 1) CRIM: per capita crime rate by town (continuous)
- 2) ZN: proportion of residential land zoned for lots over 25,000 sq.ft. (continuous)
- 3) INDUS: proportion of non-retail business acres per town (continuous)
- 4) CHAS: Charles River dummy variable (1 if tract bounds river; 0 otherwise) (binary)
- 5) NOX: nitric oxides concentration (parts per 10 million) [parts/10M] (continuous)
- 6) RM: average number of rooms per dwelling (discrete)
- 7) AGE: proportion of owner-occupied units built prior to 1940 (continuous)
- 8) DIS: weighted distances to five Boston employment centres (continuous)
- 9) RAD: index of accessibility to radial highways (discrete)
- 10) TAX: full-value property-tax rate per \$10,000 [\$ /10k] (continuous)
- 11) PTRATIO: pupil-teacher ratio by town (continuous)
- 12) B: The result of the equation  $B = 1000(B_k - 0.63)^2$  where  $B_k$  is the proportion of blacks by town (continuous)
- 13) LSTAT: % lower status of the population (continuous)

Output variable:

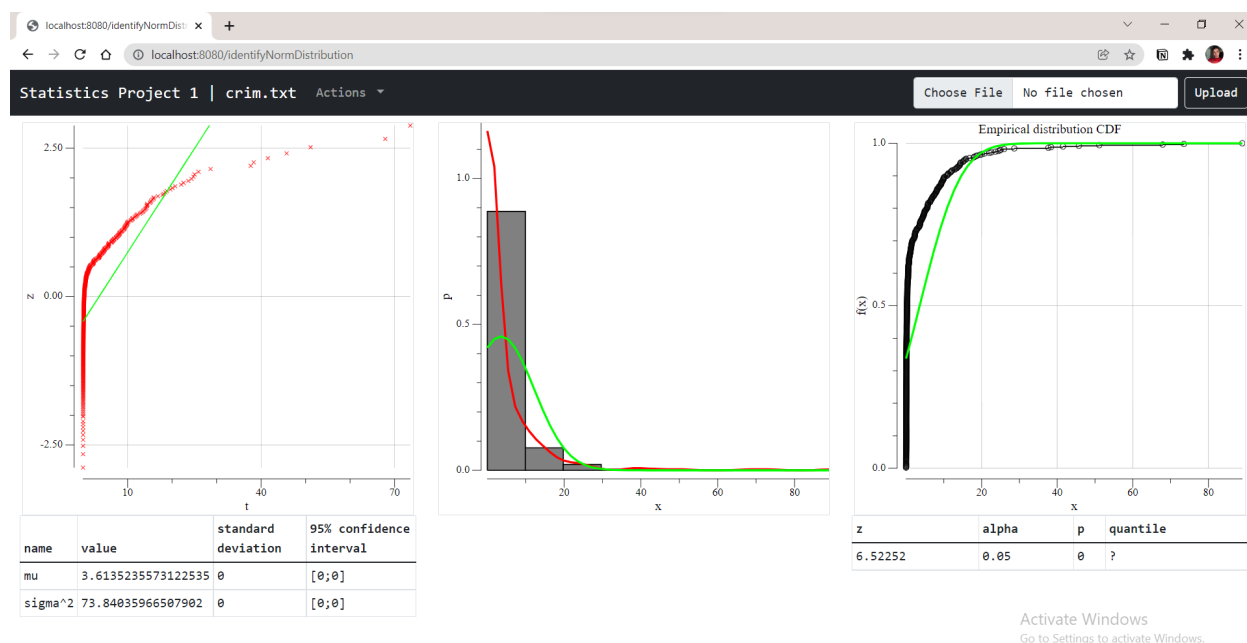
- 1) MEDV: Median value of owner-occupied homes in \$1000's [k\$]

## 1) CRIM: per capita crime rate by town (continuous)

name	value	standard deviation	95% confidence interval
mean	3.6135235573122535	0.382385	[2.862097;4.36495]
median	0.26042	0	[0.19802;0.3494]
standard deviation	8.60154510533249	0.270387	[8.070206;9.132884]
skewness	5.207652387859716	0.10825	[4.99493;5.420375]
kurtosis	36.75278626142282	0.214587	[36.331101;37.174471]
antikurtosis	0.15860475857033832	0	[0;0]
min	0.00632	0	[0;0]
max	88.9762	0	[0;0]

|u\_A|: 48.10763028416543>1.960392179315197

|u\_E|: 171.27241815131563>1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по гистограмме:

Гистограмма НЕ унимодальная  $\Rightarrow$  не идентифицируется

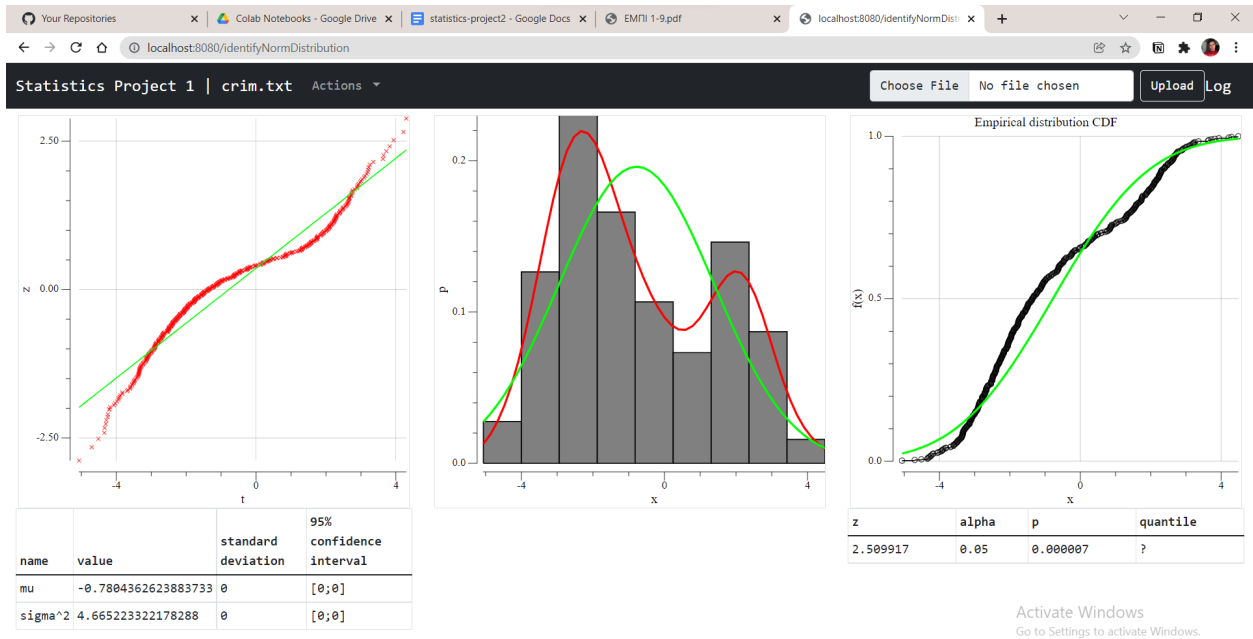
Идентификация нормального распределения по вероятностной бумаге:

В линию не складывается  $\Rightarrow$  не идентифицируется

Судя по гистограмме и по эмп. функции распределения, норм распределение НЕ вероятно.

$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

## Проведем логарифмирование



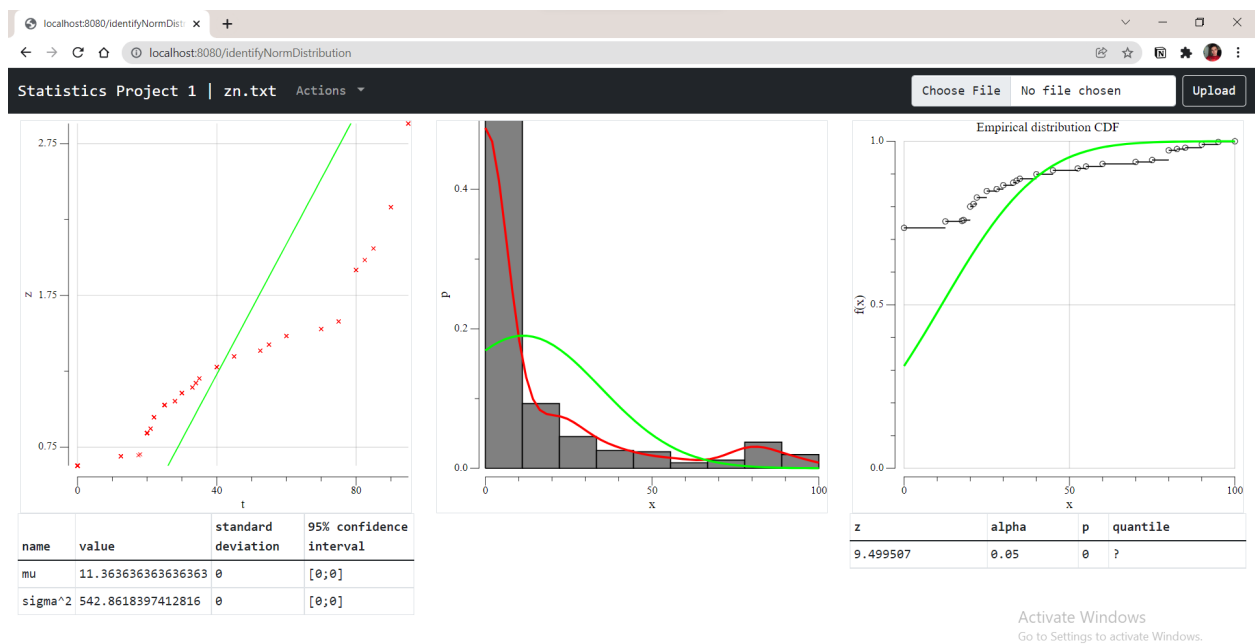
Данные неоднородны. Норм распределение не идентифицируется.

## 2) ZN: proportion of residential land zoned for lots over 25,000 sq.ft. (continuous)

name	value	standard deviation	95% confidence interval
mean	11.363636363636363	1.03681	[9.326198;13.401074]
median	0	0	[0;0]
standard deviation	23.322452994515153	0.733135	[21.881767;24.763139]
skewness	2.21906305714842	0.10825	[2.006341;2.431786]
kurtosis	3.979948773991354	0.214587	[3.558264;4.401634]
antikurtosis	0.37850697098038	0	[0;0]
min	0	0	[0;0]
max	100	0	[0;0]

|u\_A|: 20.499422230910586>1.960392179315197

|u\_E|: 18.547041462142307>1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по гистограмме:

Гистограмма НЕ унимодальная  $\Rightarrow$  не идентифицируется

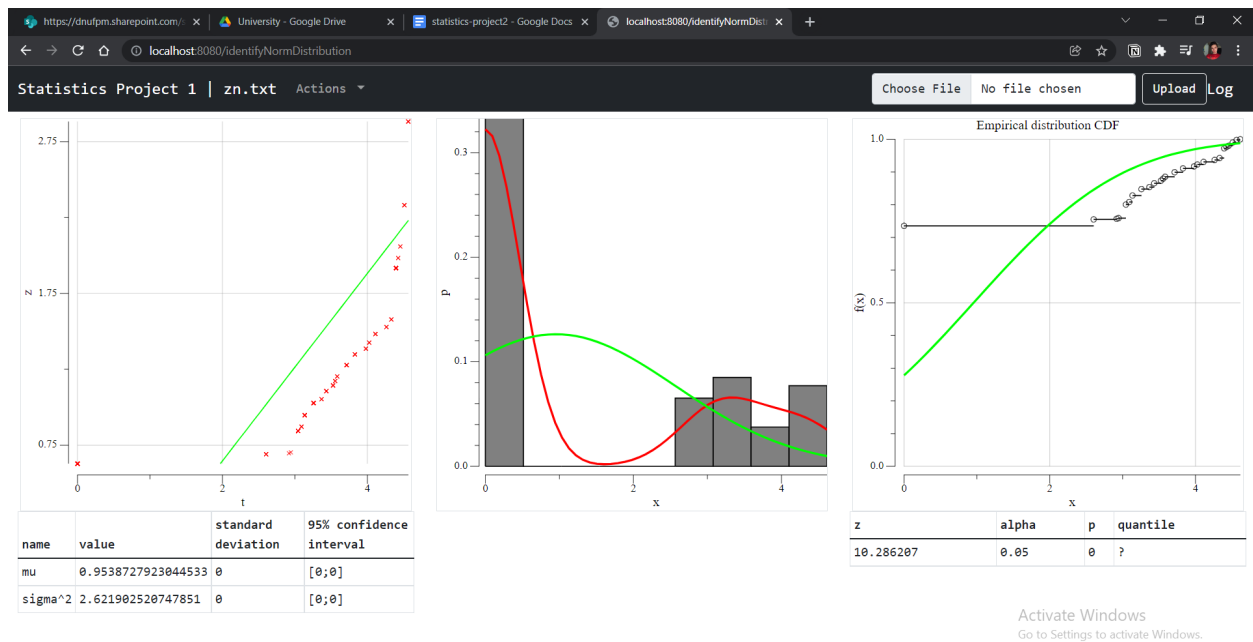
Идентификация нормального распределения по вероятностной бумаге:

В линию не складывается  $\Rightarrow$  не идентифицируется

Судя по гистограмме и по эм. функции распределения, норм распределение НЕ вероятно.

$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

## Проведем логарифмирование



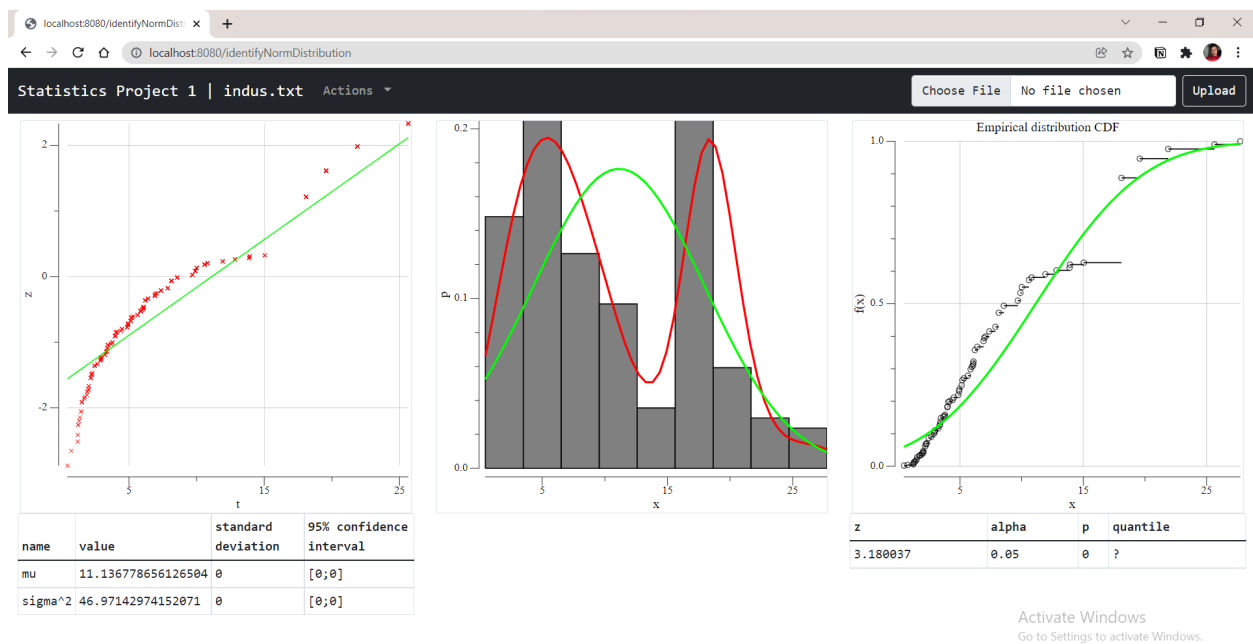
Данные неоднородны. Норм распределение не идентифицируется.

### 3) INDUS: proportion of non-retail business acres per town (continuous)

name	value	standard deviation	95% confidence interval
mean	11.136778656126477	0.30498	[10.537462;11.736096]
median	9.69	0	[8.14;10.01]
standard deviation	6.860352940897581	0.215653	[6.436572;7.284134]
skewness	0.2941462768441882	0.10825	[0.081424;0.506869]
kurtosis	-1.2332184715629375	0.214587	[-1.654904;-0.811533]
antikurtosis	0.7523303374698953	0	[0;0]
min	0.46	0	[0;0]
max	27.74	0	[0;0]

|u\_A|: 2.7172858866065255 > 1.960392179315197

|u\_E|: 5.746946863594796 > 1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по гистограмме:

Гистограмма НЕ унимодальная  $\Rightarrow$  не идентифицируется

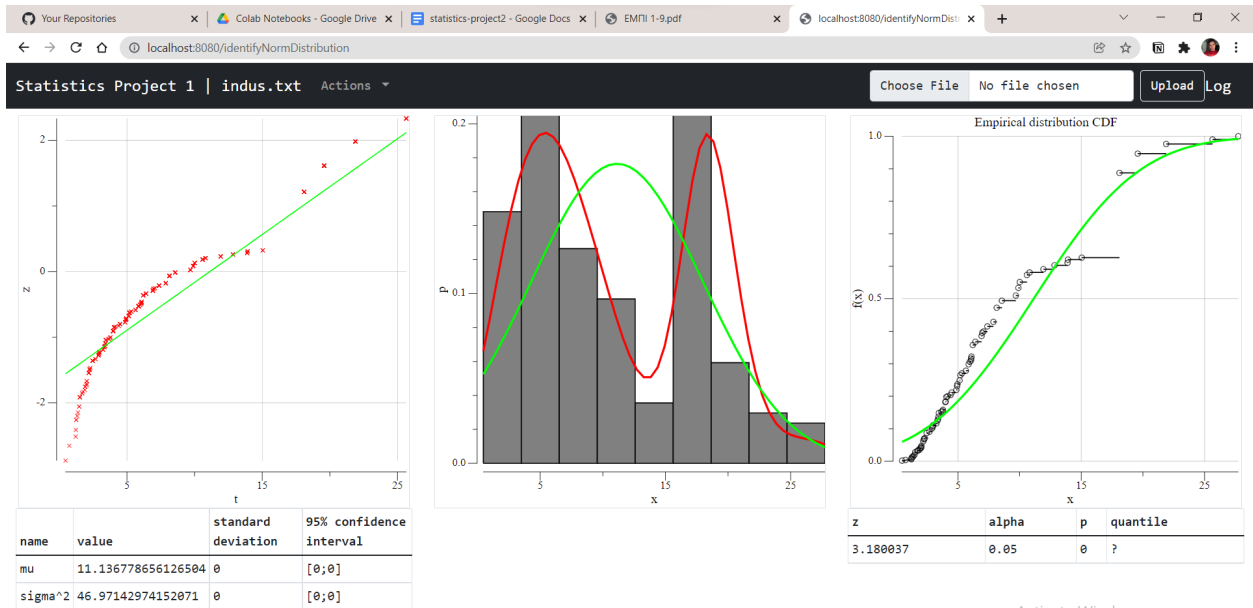
Идентификация нормального распределения по вероятностной бумаге:

В линию не складывается  $\Rightarrow$  не идентифицируется

Судя по гистограмме и по эмп. функции распределения, норм распределение НЕ вероятно.

$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

Проведем логарифмизацию



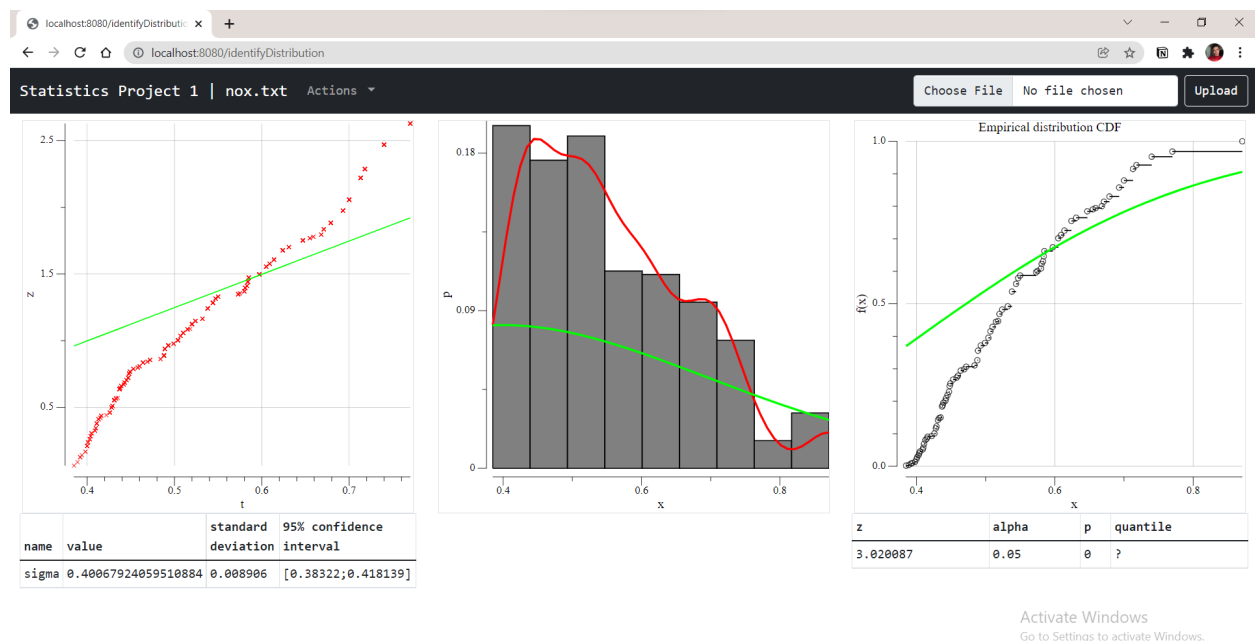
Данные неоднородны. Норм распределение не идентифицируется.

## 5) NOX: nitric oxides concentration (parts per 10 million) [parts/10M] (continuous)

name	value	standard deviation	95% confidence interval
mean	0.5546950592885365	0.005151	[0.544572;0.564818]
median	0.538	0	[0.52;0.544]
standard deviation	0.11587767566755586	0.003643	[0.10872;0.123036]
skewness	0.7271441597740629	0.10825	[0.514422;0.939867]
kurtosis	-0.07586421510233521	0.214587	[-0.497549;0.345821]
antikurtosis	0.5847917433315095	0	[0;0]
min	0.385	0	[0;0]
max	0.871	0	[0;0]

|u\_A|: 6.717265246668584 > 1.960392179315197

|u\_E|: 0.353536395290034 < 1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по гистограмме:

Гистограмма НЕ унимодальная  $\Rightarrow$  не идентифицируется

Идентификация нормального распределения по вероятностной бумаге:

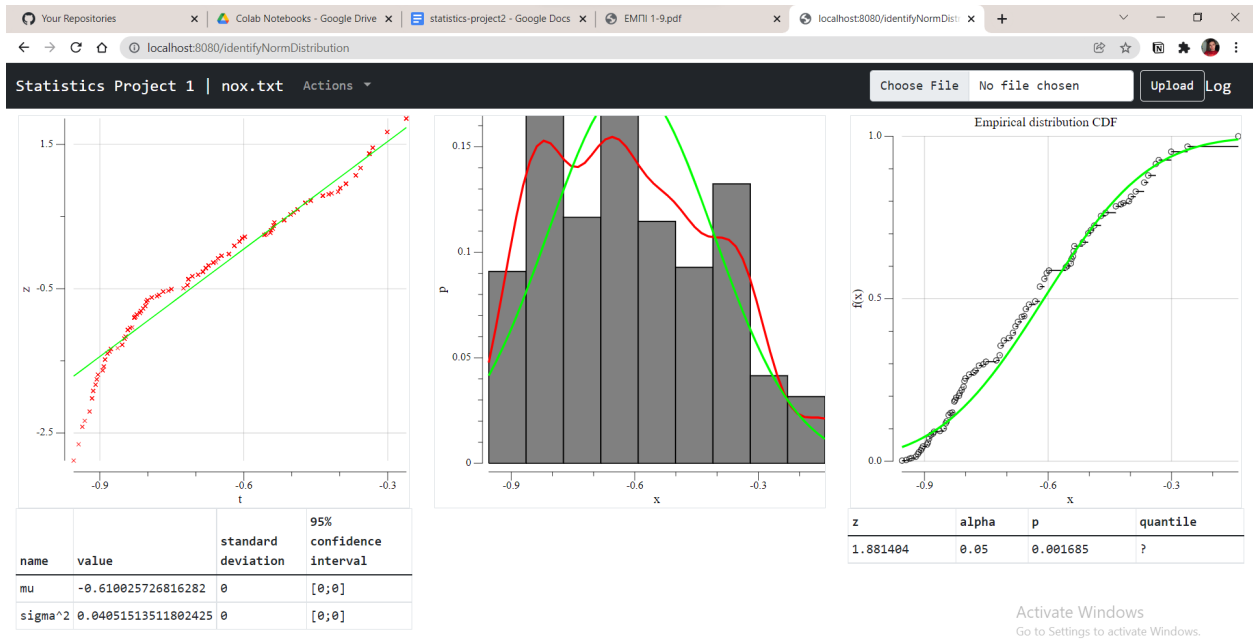
В линию не складывается  $\Rightarrow$  не идентифицируется



Судя по гистограмме и по эмп. функции распределения, норм распределение НЕ вероятно.

$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

## Проведем логарифмирование



Распределение стало более похоже на нормальное, хотя таковым все таки не является.

name	value	standard deviation	95% confidence interval
mean	0.43858690631137576	0.003243	[0.432214;0.44496]
median	0.43048287108345223	0	[0.41871;0.434376]
standard deviation	0.07294762699592305	0.002293	[0.068441;0.077454]
skewness	0.5826134297445467	0.10825	[0.369891;0.795336]
kurtosis	-0.3851957285385952	0.214587	[-0.806881;0.036489]
antikurtosis	0.618415558240484	0	[0;0]
min	0.32570013963930183	0	[0;0]
max	0.6264730472919525	0	[0;0]

$|u_A|: 5.382108748671584 > 1.960392179315197$

$|u_E|: 1.7950585683244167 < 1.960392179315197$

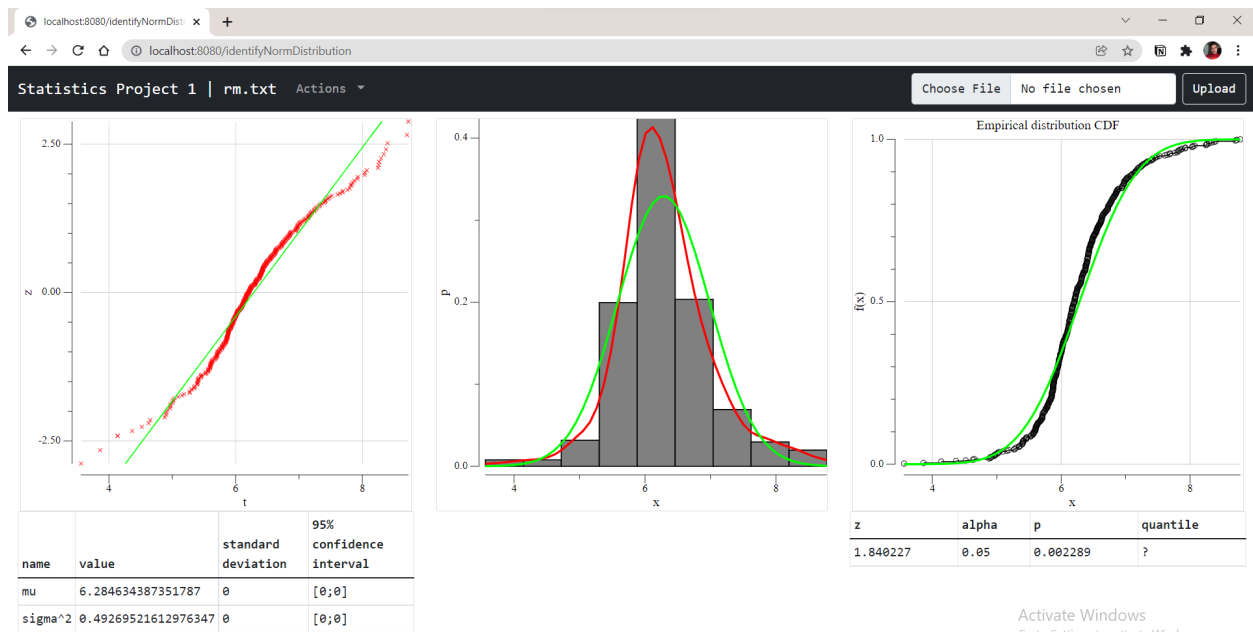
Экссесс характерен для нормального распределения, остальные параметры нет.

## 6) RM: average number of rooms per dwelling (continuous)

name	value	standard deviation	95% confidence interval
mean	6.2846343873517805	0.031235	[6.223254;6.346015]
median	6.209	0	[6.151;6.273]
standard deviation	0.7026171434153236	0.022087	[0.659215;0.74602]
skewness	0.40241466802495474	0.10825	[0.189692;0.615137]
kurtosis	1.8610269725310182	0.214587	[1.439342;2.282712]
antikurtosis	0.4535612928626693	0	[0;0]
min	3.561	0	[0;0]
max	8.78	0	[0;0]

|u\_A|: 3.717455511316511 > 1.960392179315197

|u\_E|: 8.672610222337738 > 1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по гистограмме:

Гистограмма унимодальная  $\Rightarrow$  идентифицируется

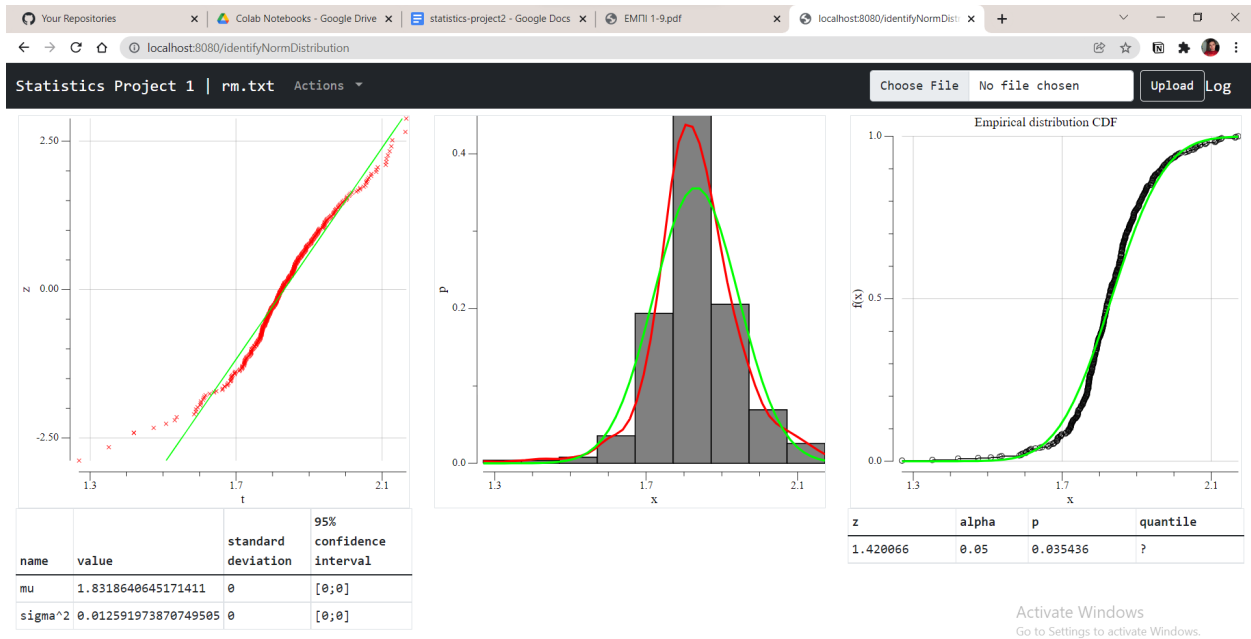
Идентификация нормального распределения по вероятностной бумаге:

В линию не складывается (но это было близко)  $\Rightarrow$  не идентифицируется

Судя по гистограмме и по эмп. функции распределения, норм распределение вероятно.

$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

Проведем логарифмирование



name	value	standard deviation	95% confidence interval
mean	1.9811392365329155	0.00429	[1.972709;1.989569]
median	1.9753302454224415	0	[1.967252;1.984169]
standard deviation	0.09649950672483172	0.003033	[0.090538;0.102461]
skewness	-0.18238673662151278	0.10825	[-0.395109;0.030336]
kurtosis	2.4331167462504073	0.214587	[2.011432;2.854802]
antikurtosis	0.429017967842524	0	[0;0]
min	1.5175418977295634	0	[0;0]
max	2.280339484046726	0	[0;0]

$|u_A|: 1.6848654711627715 \leq 1.960392179315197$   
 $|u_E|: 11.338617589714012 > 1.960392179315197$

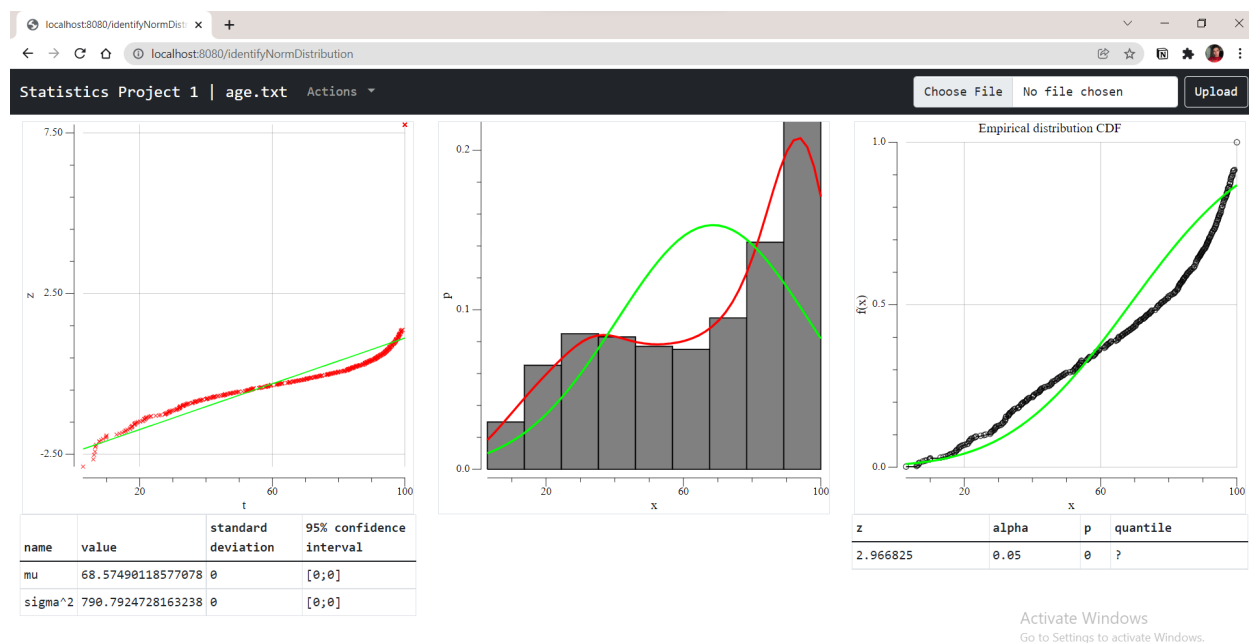
Смещение вправо улучшило ситуацию - распределение похоже на нормальное. Но вот эксцесс вскочил еще выше - больше похоже на Лапласа.

## 7) AGE: proportion of owner-occupied units built prior to 1940 (continuous)

name	value	standard deviation	95% confidence interval
mean	68.57490118577077	1.25137	[66.115831;71.033972]
median	77.7	0	[72.5;82.6]
standard deviation	28.148861406903645	0.884852	[26.410036;29.887687]
skewness	-0.5971855948016171	0.10825	[-0.809908;-0.384463]
kurtosis	-0.9700139266403975	0.214587	[-1.391699;-0.548329]
antikurtosis	0.7018648138893193	0	[0;0]
min	2.9	0	[0;0]
max	100	0	[0;0]

|u\_A|: 5.516724555717318 > 1.960392179315197

|u\_E|: 4.520381928989622 > 1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по гистограмме:

Гистограмма НЕ унимодальная  $\Rightarrow$  не идентифицируется

Идентификация нормального распределения по вероятностной бумаге:

В линию не складывается  $\Rightarrow$  не идентифицируется

Судя по гистограмме и по эмп. функции распределения, норм распределение НЕ вероятно.

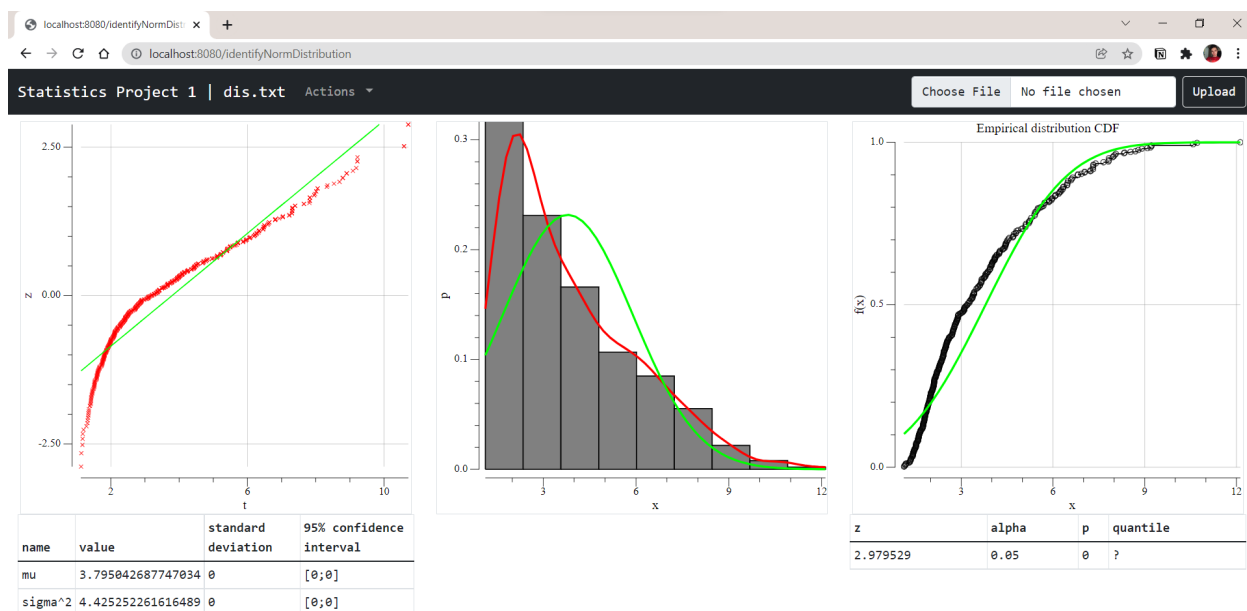
$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

## 8) DIS: weighted distances to five Boston employment centres (continuous)

name	value	standard deviation	95% confidence interval
mean	3.7950426877470314	0.09361	[3.611089;3.978996]
median	3.2157	0	[2.872;3.4952]
standard deviation	2.105710126627611	0.066192	[1.975635;2.235785]
skewness	1.008778756515231	0.10825	[0.796056;1.221501]
kurtosis	0.47129857161099853	0.214587	[0.049613;0.892984]
antikurtosis	0.536727707186047	0	[0;0]
min	1.1296	0	[0;0]
max	12.1265	0	[0;0]

|u\_A|: 9.318969824116873 > 1.960392179315197

|u\_E|: 2.1963082052313427 > 1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по гистограмме:

Гистограмма НЕ унимодальная  $\Rightarrow$  не идентифицируется

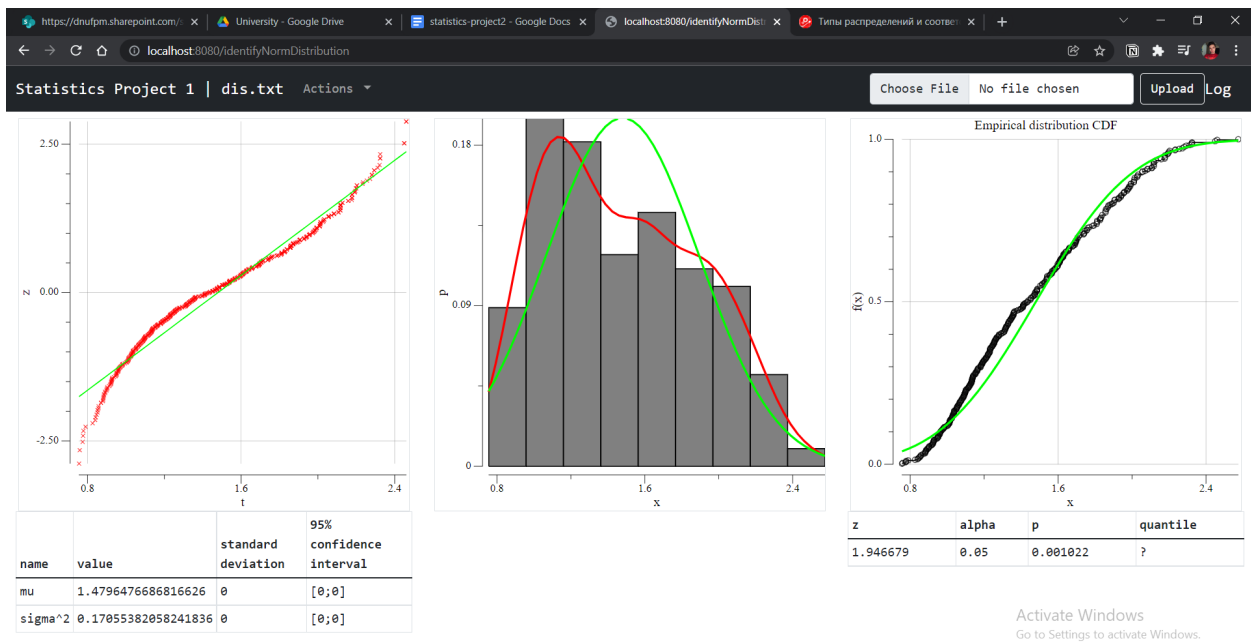
Идентификация нормального распределения по вероятностной бумаге:

В линию не складывается  $\Rightarrow$  не идентифицируется

Судя по гистограмме и по эм. функции распределения, норм распределение НЕ вероятно.

$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

Проведем логарифмирование



name	value	standard deviation	95% confidence interval
mean	1.479647668681662	0.018377	[1.443534;1.515761]
median	1.4388156512119903	0	[1.353771;1.50301]
standard deviation	0.41339031304592194	0.012995	[0.387854;0.438926]
skewness	0.3305777426343017	0.10825	[0.117855;0.5433]
kurtosis	-0.9080125069208811	0.214587	[-1.329698;-0.486327]
antikurtosis	0.691385804063767	0	[0;0]
min	0.7559341686587101	0	[0;0]
max	2.5746330876618586	0	[0;0]

|u\_A|: 3.053835133062914>1.960392179315197

|u\_E|: 4.2314478327107095>1.960392179315197

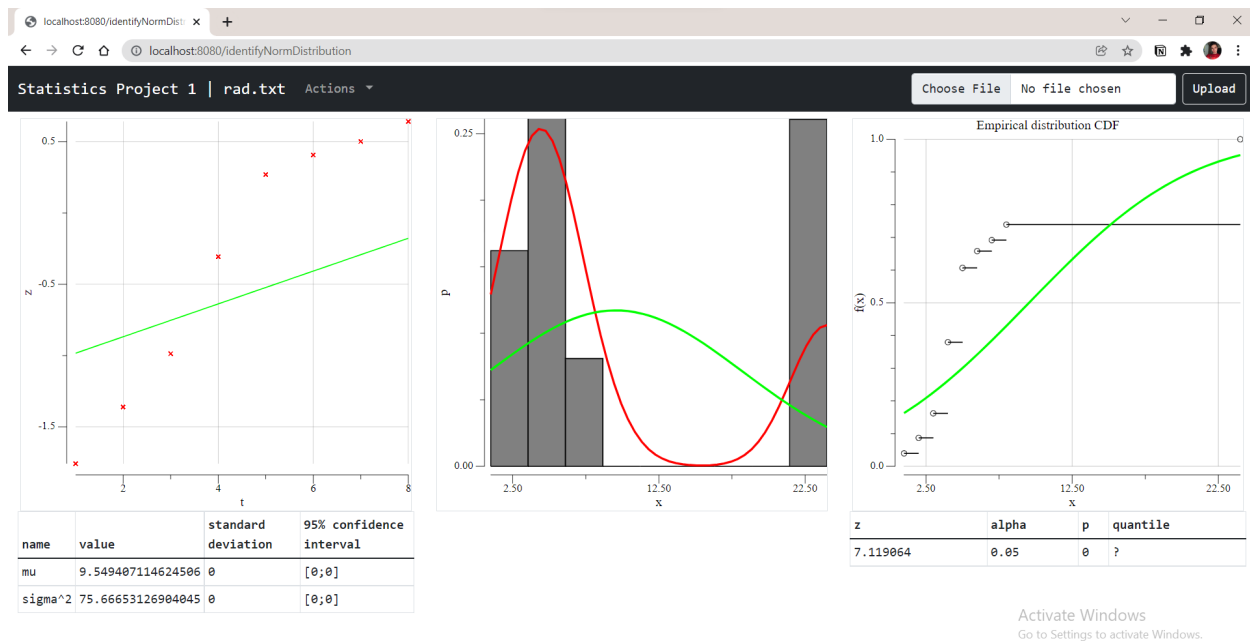
Асимметрия уменьшилась, хотя до нормального далековато. Экссесс вырос. По бумаге и эм. Функции не идентифицируется.

## 9) RAD: index of accessibility to radial highways (discrete)

name	value	standard deviation	95% confidence interval
mean	9.549407114624506	0.387085	[8.788745;10.310069]
median	5	0	[5;5]
standard deviation	8.707259384239377	0.27371	[8.16939;9.245129]
skewness	1.0018334924536938	0.10825	[0.789111;1.214556]
kurtosis	-0.8705205036905599	0.214587	[-1.292206;-0.448835]
antikurtosis	0.6852724442476346	0	[0;0]
min	1	0	[0;0]
max	24	0	[0;0]

|u\_A|: 9.254810358236>1.960392179315197

|u\_E|: 4.056730574298817>1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по вероятностной бумаге:

В линию не складывается  $\Rightarrow$  не идентифицируется

Судя по гистограмме и по эмп. функции распределения, норм распределение НЕ вероятно.



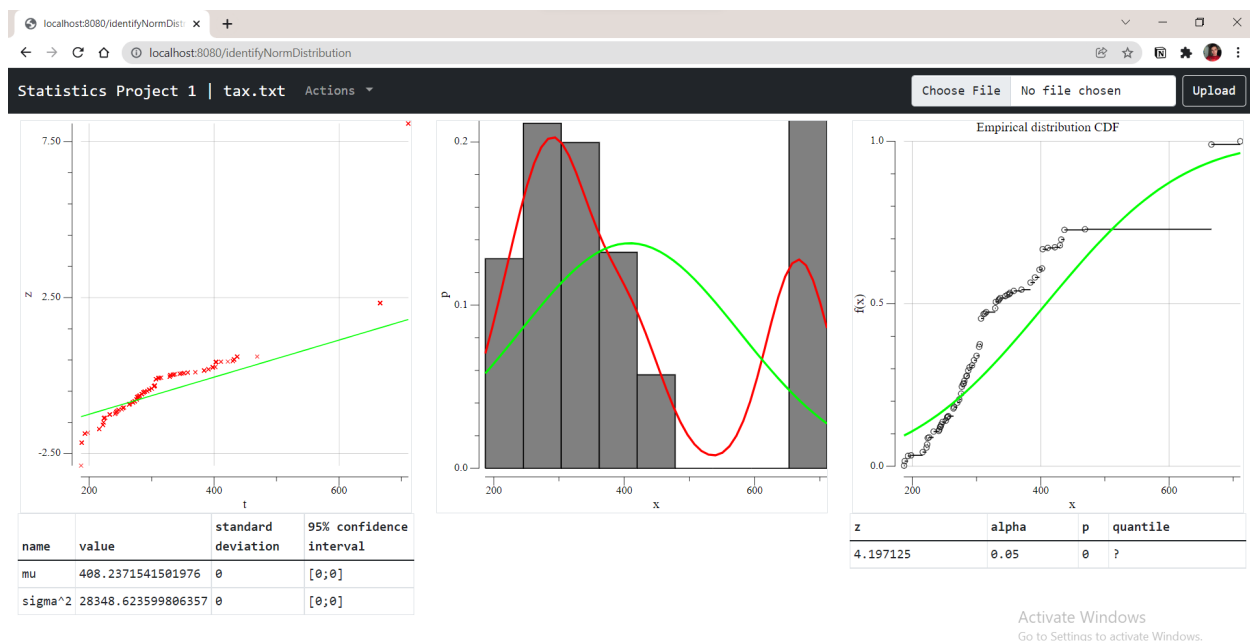
$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

## 10) TAX: full-value property-tax rate per \$10,000 [\$ /10k] (continuous)

name	value	standard deviation	95% confidence interval
mean	408.2371541501976	7.492389	[393.513835;422.960473]
median	330	0	[311;384]
standard deviation	168.53711605495914	5.297919	[158.126157;178.948075]
skewness	0.667968268723776	0.10825	[0.455246;0.880691]
kurtosis	-1.142984884835201	0.214587	[-1.56467;-0.7213]
antikurtosis	0.7338246238879247	0	[0;0]
min	187	0	[0;0]
max	711	0	[0;0]

|u\_A|: 6.170605893018203>1.960392179315197

|u\_E|: 5.32644746288548>1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по гистограмме:

Гистограмма НЕ унимодальная  $\Rightarrow$  не идентифицируется

Идентификация нормального распределения по вероятностной бумаге:

В линию не складывается  $\Rightarrow$  не идентифицируется

Судя по гистограмме и по эмп. функции распределения, норм распределение НЕ вероятно.

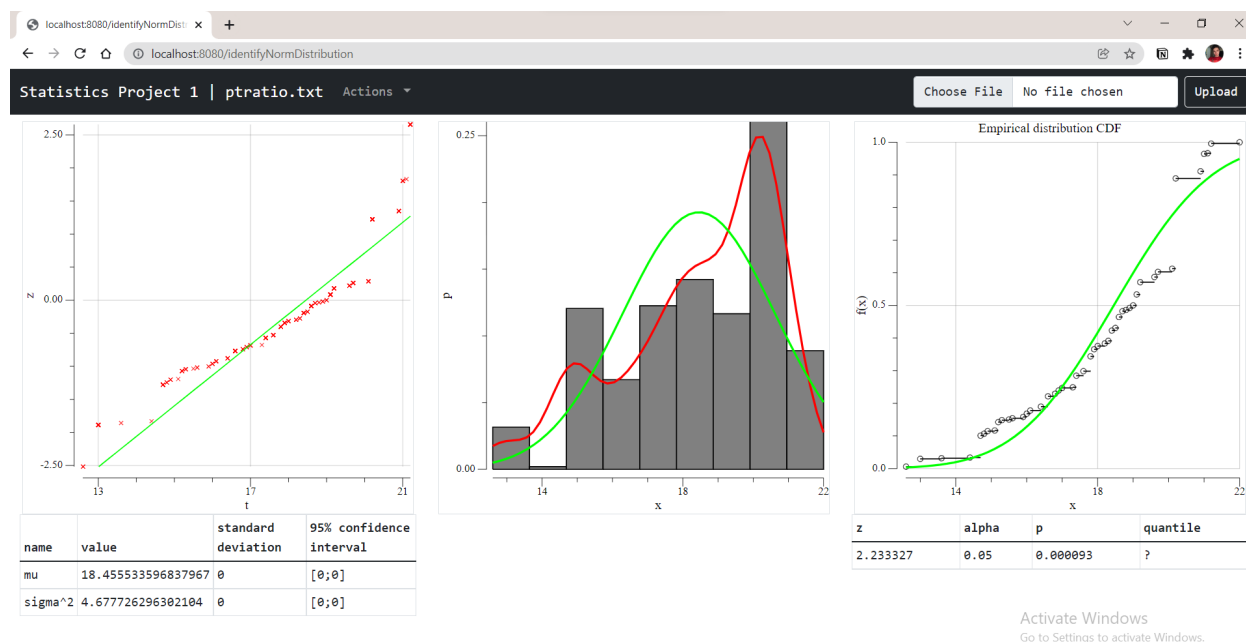
$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

## 11) PTRATIO: pupil-teacher ratio by town (continuous)

name	value	standard deviation	95% confidence interval
mean	18.455533596837927	0.096244	[18.266405;18.644662]
median	19.1	0	[18.6;19.2]
standard deviation	2.1649455237144384	0.068054	[2.031211;2.29868]
skewness	-0.7999445320367345	0.10825	[-1.012667;-0.587222]
kurtosis	-0.2941163777003575	0.214587	[-0.715801;0.127569]
antikurtosis	0.6079186162837703	0	[0;0]
min	12.6	0	[0;0]
max	22	0	[0;0]

|u\_A| : 7.389785824564069 > 1.960392179315197

|u\_E| : 1.3706178048198887 <= 1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по гистограмме:

Гистограмма НЕ унимодальная  $\Rightarrow$  не идентифицируется

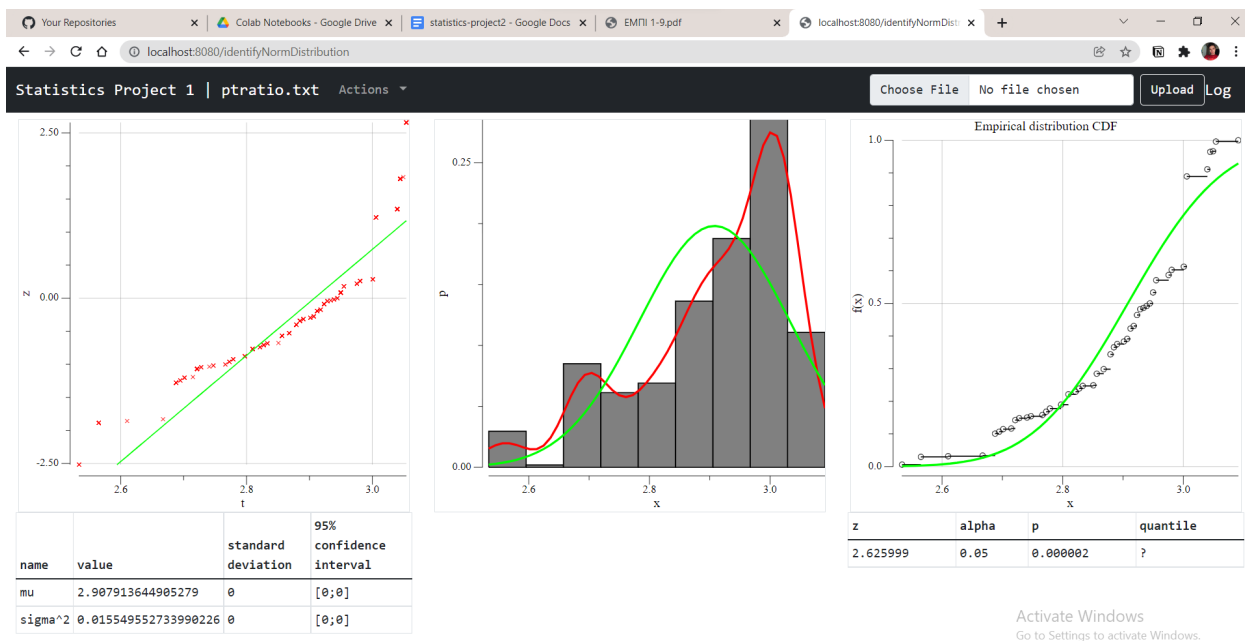
Идентификация нормального распределения по вероятностной бумаге:

В линию не складывается  $\Rightarrow$  не идентифицируется

Судя по гистограмме и по эмп. функции распределения, норм распределение НЕ вероятно.

$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

## Проведем логарифмирование



name	value	standard deviation	95% confidence interval
mean	2.9614624142511143	0.005242	[2.95116;2.971764]
median	3.0007198150650303	0	[2.97553;3.005683]
standard deviation	0.11792573709911214	0.003707	[0.110641;0.12521]
skewness	-1.0000064811450267	0.10825	[-1.212729;-0.787284]
kurtosis	0.19063003860345562	0.214587	[-0.231055;0.612315]
antikurtosis	0.559837228565339	0	[0;0]
min	2.6100697927420065	0	[0;0]
max	3.1354942159291497	0	[0;0]

$|u_A|: 9.237932660183947 > 1.960392179315197$

$|u_E|: 0.8883589791439261 \leq 1.960392179315197$

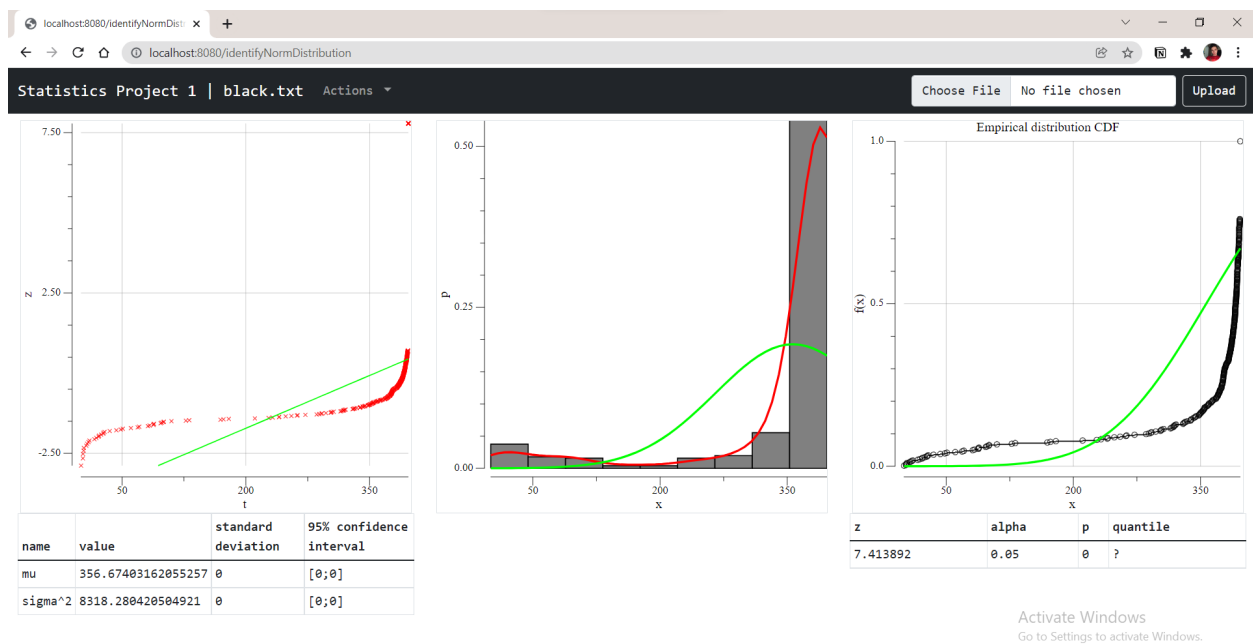
Данные еще более сместились вправо, что ожидаемо. А вот эксцесс стал ближе к нормальному.

12) B: The result of the equation  $B=1000(B_k - 0.63)^2$  where  $B_k$  is the proportion of blacks by town (continuous)

name	value	standard deviation	95% confidence interval
mean	356.674031620552	4.058552	[348.698557;364.649506]
median	391.475	0	[390.43;392.74]
standard deviation	91.29486438415753	2.869829	[85.655352;96.934376]
skewness	-2.8817983452627622	0.10825	[-3.094521;-2.669076]
kurtosis	7.143769291496026	0.214587	[6.722084;7.565454]
antikurtosis	0.313978795049101	0	[0;0]
min	0.32	0	[0;0]
max	396.9	0	[0;0]

|u\_A|: 26.62168651475577>1.960392179315197

|u\_E|: 33.290826784305644>1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по гистограмме:

Гистограмма НЕ унимодальная  $\Rightarrow$  не идентифицируется

Идентификация нормального распределения по вероятностной бумаге:

В линию не складывается  $\Rightarrow$  не идентифицируется

Судя по гистограмме и по эмп. функции распределения, норм распределение НЕ вероятно.

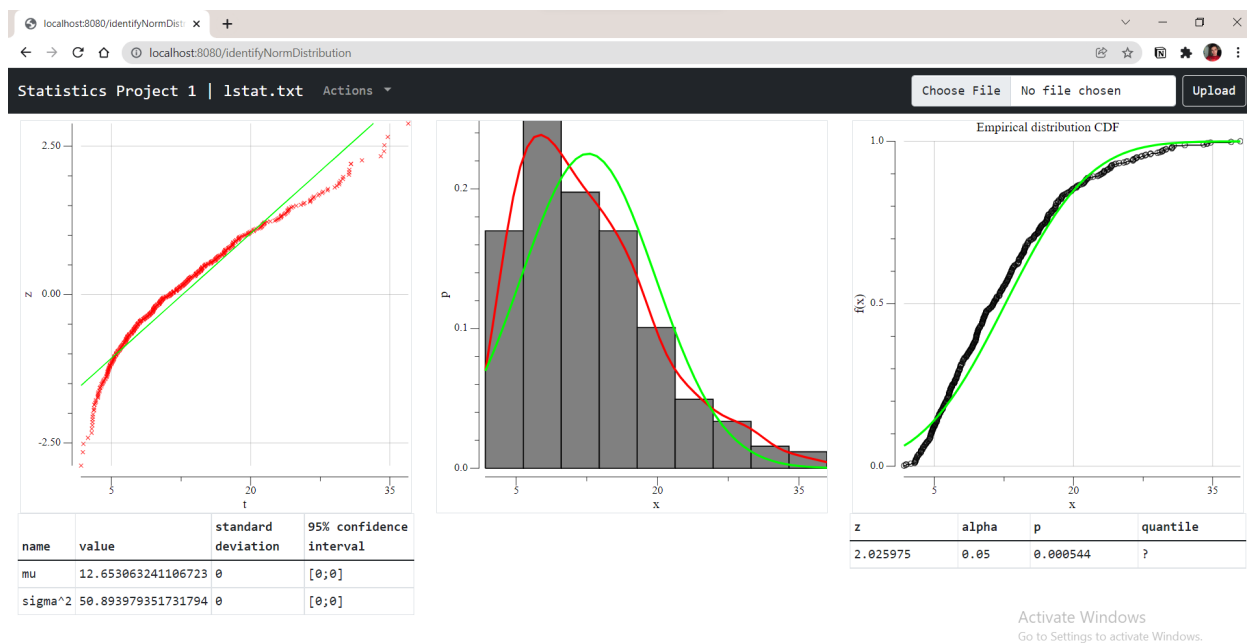
$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

### 13) LSTAT: % lower status of the population (continuous)

name	value	standard deviation	95% confidence interval
mean	12.653063241106734	0.317459	[12.029224;13.276903]
median	11.395	0	[10.3;12.34]
standard deviation	7.1410615113485685	0.224477	[6.69994;7.582183]
skewness	0.9037707431346086	0.10825	[0.691048;1.116493]
kurtosis	0.47654475572974286	0.214587	[0.05486;0.89823]
antikurtosis	0.5363225872121025	0	[0;0]
min	1.73	0	[0;0]
max	37.97	0	[0;0]

$|u_A|$ : 8.34891915476606 > 1.960392179315197

$|u_E|$ : 2.220756056169586 > 1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по гистограмме:

Гистограмма НЕ унимодальная  $\Rightarrow$  не идентифицируется

Идентификация нормального распределения по вероятностной бумаге:

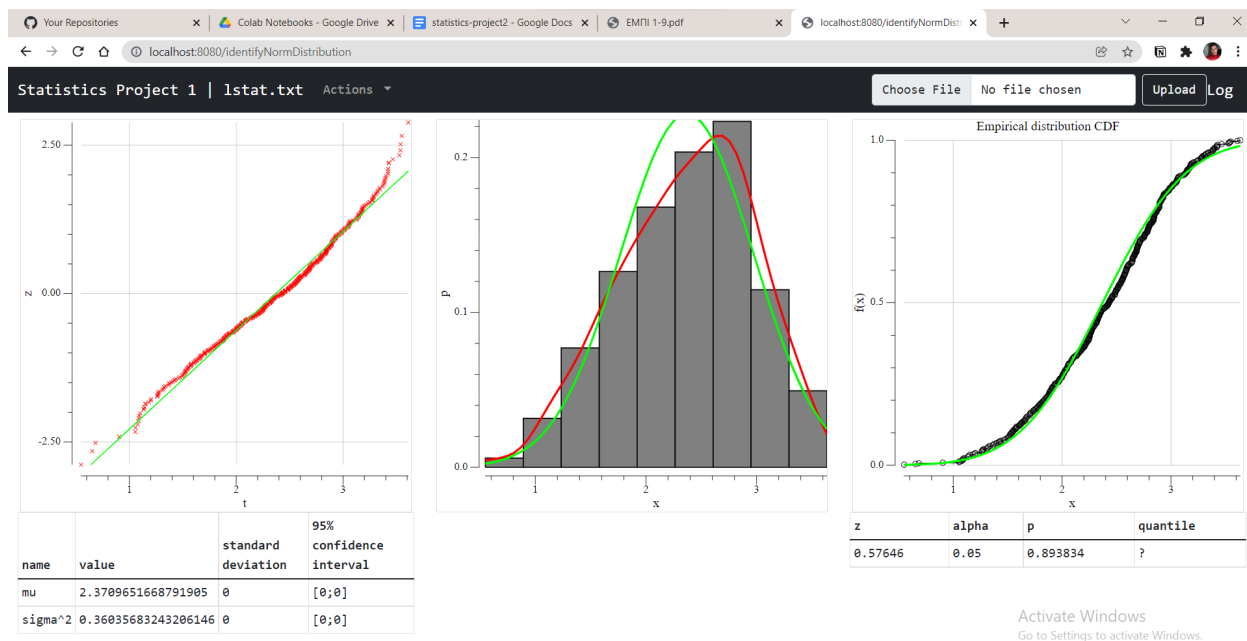
В линию не складывается  $\Rightarrow$  не идентифицируется



Судя по гистограмме и по эмп. функции распределения, норм распределение НЕ вероятно.

$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

## Логарифмизация



name	value	standard deviation	95% confidence interval
mean	2.3709651668791873	0.026713	[2.318472;2.423459]
median	2.4331737962860847	0	[2.332144;2.512846]
standard deviation	0.6008913465162111	0.018889	[0.563773;0.63801]
skewness	-0.3192822699479334	0.10825	[-0.532005;-0.10656]
kurtosis	-0.43905902932755536	0.214587	[-0.860744;-0.017374]
antikurtosis	0.6248851670715596	0	[0;0]
min	0.5481214085096876	0	[0;0]
max	3.636796374243711	0	[0;0]

$|u_A|: 2.9494889933037616 > 1.960392179315197$

$|u_E|: 2.0460680485340874 > 1.960392179315197$

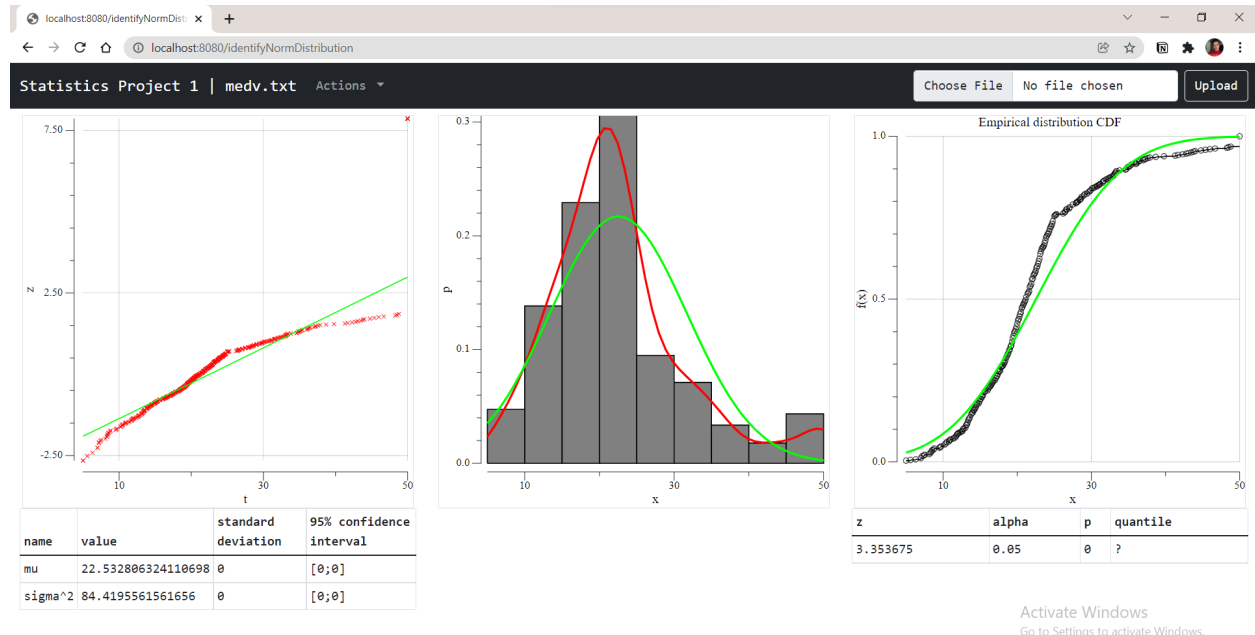
Асимметрия стала больше похожа на асимметрию нормального распределения, эксцесс почти не изменился. Но все же данные смещены слишком вправо. Бумага и эмп. функции весьма похожи на нормальное распределение.

# 1) MEDV: Median value of owner-occupied homes in \$1000's [k\$]

name	value	standard deviation	95% confidence interval
mean	22.532806324110673	0.408861	[21.729352;23.336261]
median	21.2	0	[20.5;21.9]
standard deviation	9.197104087379817	0.289108	[8.628976;9.765232]
skewness	1.1048108228646367	0.10825	[0.892088;1.317533]
kurtosis	1.4686287722747524	0.214587	[1.046944;1.890314]
antikurtosis	0.4730563333431139	0	[0;0]
min	5	0	[0;0]
max	50	0	[0;0]

|u\_A|: 10.20610183663977>1.960392179315197

|u\_E|: 6.843987266840675>1.960392179315197



Идентификация нормального распределения по коэффициентам асимметрии и эксцесса:

$u_A > u_{1-0.05/2} \Rightarrow$  не идентифицируется

Идентификация нормального распределения по гистограмме:

Гистограмма НЕ унимодальная  $\Rightarrow$  не идентифицируется

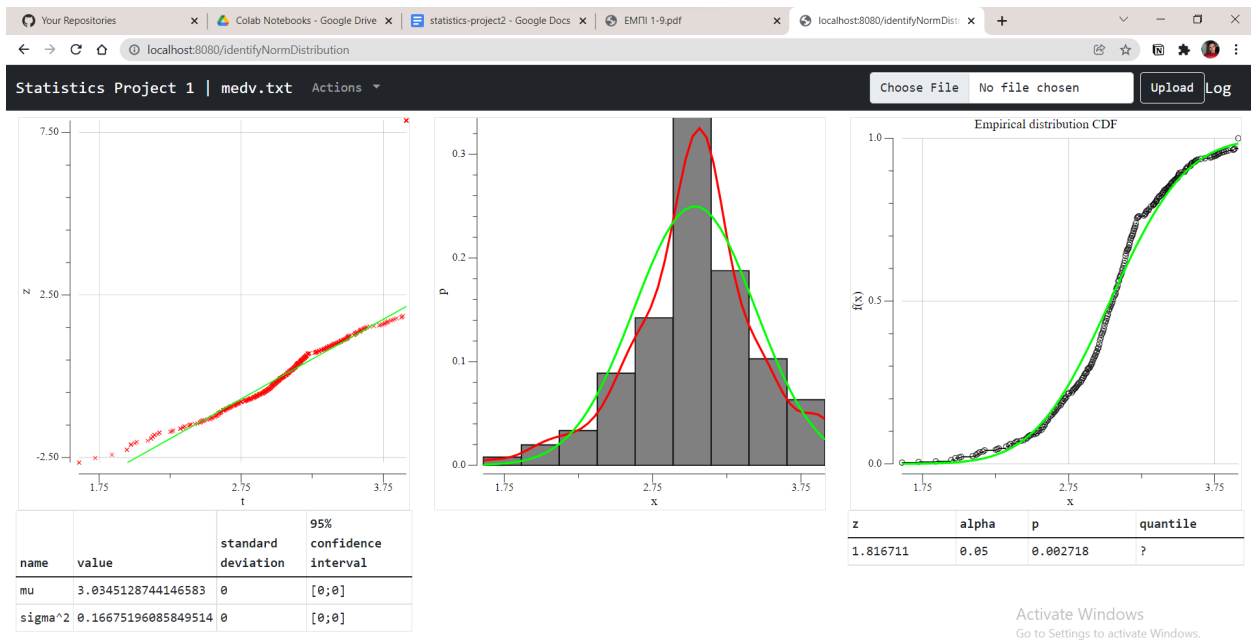
Идентификация нормального распределения по вероятностной бумаге:

В линию не складывается  $\Rightarrow$  не идентифицируется

Судя по гистограмме и по эм. функции распределения, норм распределение НЕ вероятно.

$p > \alpha \Rightarrow$  норм распределение НЕ вероятно

Проведем логарифмирование



name	value	standard deviation	95% confidence interval
mean	3.0854366214365796	0.017203	[3.051631;3.119242]
median	3.100092288878234	0	[3.068053;3.131137]
standard deviation	0.38696564180661636	0.012164	[0.363062;0.410869]
skewness	-0.24052778511556885	0.10825	[-0.45325;-0.027805]
kurtosis	0.6555768010148184	0.214587	[0.233892;1.077262]
antikurtosis	0.523024512953821	0	[0;0]
min	1.791759469228055	0	[0;0]
max	3.9318256327243257	0	[0;0]

|u\_A|: 2.221965080922886>1.960392179315197

|u\_E|: 3.0550669871679275>1.960392179315197

Асимметрия стала меньше, довольно близка к нормальному распределению. Эксцесс также стал ближе. По бумаге и эм. функции ситуация также значительно улучшилась.