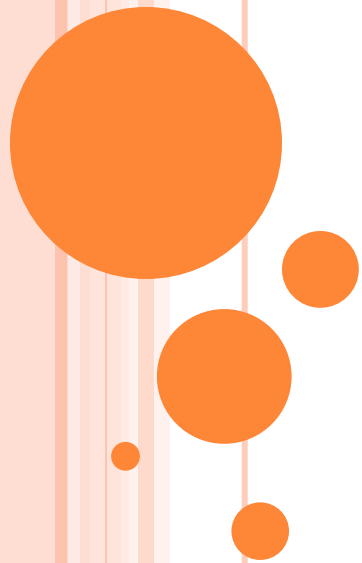


СТАТИСТИКА ЗАДАЧА 2

КОМАНДА №3

Надежда Бойко
Алексей Кулиш
Карина Чумак
Антон Лютов



ПОСТАНОВКА ЗАДАЧИ

В таблице приведены данные Госкомстата об основных показателях ведения охотничьего хозяйства с 1996 по 2017 гг.

Роки	Площа мисливських угідь на кінець року, тис.га	Витрати на ведення мисливського господарства (у фактичних цінах), тис.грн	Облікова кількість штатних працівників, зайнятих у мисливському господарстві, на кінець року, осіб	Чисельність тварин у мисливських господарствах, тис. голів			Кількість розплідників, вольєрів, ферм для штучного розведення мисливських тварин, одиниць
				копитні тварини	хутрові звірі	перната дичина	
1996	50583,5	12784,9	4558	219,4	2528,9	10888,8	17
1997	48701,6	15268,7	5048	205,3	2554	9975,4	28
1998	47721,8	16388,2	5262	192,3	2464,2	9828	26
1999	47903,9	20336,3	4906	182,6	2422,3	8953,9	29
2000	47806,9	20811,1	4965	177,8	2416,4	9417,5	44

НЕОБХОДИМО ПРОВЕСТИ ВИЗУАЛЬНИЙ І ДЕСКРИПТИВНИЙ АНАЛІЗ ДАННИХ:

- А) ВИСЛИТИ ВИБОРЧНІ ХАРАКТЕРИСТИКИ (СРЕДНЕЕ, МЕДИАНА, ДИСПЕРСИЯ, ИНТЕРКВАРТИЛЬНИЙ РАЗМАХ), НАЙТИ ВИБОРЧНУЮ КОРРЕЛЯЦИОННУЮ МАТРИЦУ КОЕФФИЦИЕНТОВ ПИРСОНА І СПИРМЕНА.
- В) ДЛЯ НАБОРОВ НАБЛЮДЕНИЙ ПОСТРОИТИ ГРАФИКИ ТИПА «УСАТЫХ КОРОБОЧЕК»
- С) ВИБРАТИ ЗАВИСИМУЮ ВЕЛИЧИНУ Y І НАБОР ФАКТОРОВ $X_1, X_2 \dots X_p$ ДЛЯ ДАЛЬНЕЙШЕГО РЕГРЕССИОННОГО АНАЛІЗА
- Д) ОЦЕНИТИ ПАРАМЕТРИ **ЛИНЕЙНОЇ** РЕГРЕССИИ ЗАВИСИМОСТИ ВЕЛИЧИН (Y) ОТ (X) І ПРОАНАЛІЗОВАТИ РЕЗУЛЬТАТИ (АДЕКВАТНОСТЬ, ИСПОЛЬЗОВАНИЕ ДЛЯ ПРОГНОЗА).



ВЫБОРОЧНЫЕ ХАРАКТЕРИСТИКИ

	Mean	Median	Variance	IQR
Area_mln_hectares	4.554428e+01	4.722615e+01	1.357150e+01	2.6702
Spending_mln_hrn	1.202456e+02	8.174905e+01	1.053927e+04	186.0390
Num_of_workers	5.984818e+03	5.973500e+03	6.174582e+05	1102.2500
Hoofed_animals	2.102091e+05	2.130500e+05	5.102508e+08	43200.0000
Hur_animals	2.212559e+06	2.284350e+06	7.424128e+10	314900.0000
Birds	9.811314e+06	9.792100e+06	2.826696e+11	686275.0000
Animal_nurseries	1.337727e+02	1.390000e+02	5.792085e+03	133.5000

МАТРИЦА КОЭФФИЦИЕНТОВ ПИРСОНА

```
#Корреляция Пирсона  
corr = df_stat.corr()  
corr
```

	Area_mln_hectares	Spending_mln_hrn	Num_of_workers	Hoofed_animals	Hur_animals	Birds	Animal_nurseries
Area_mln_hectares	1.000000	-0.853679	-0.510364	-0.498136	0.972034	0.048485	-0.783467
Spending_mln_hrn	-0.853679	1.000000	0.785662	0.763282	-0.919471	0.361212	0.918721
Num_of_workers	-0.510364	0.785662	1.000000	0.804859	-0.629380	0.435128	0.889425
Hoofed_animals	-0.498136	0.763282	0.804859	1.000000	-0.607175	0.708028	0.763938
Hur_animals	0.972034	-0.919471	-0.629380	-0.607175	1.000000	-0.085774	-0.872199
Birds	0.048485	0.361212	0.435128	0.708028	-0.085774	1.000000	0.266335
Animal_nurseries	-0.783467	0.918721	0.889425	0.763938	-0.872199	0.266335	1.000000

ВЫБОРОЧНЫЕ ХАРАКТЕРИСТИКИ

МАТРИЦА КОЭФФИЦИЕНТОВ СПИРМЕНА

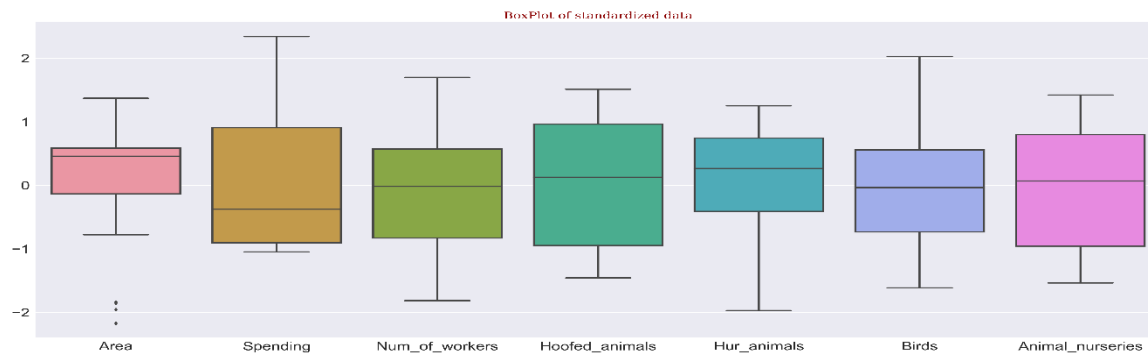
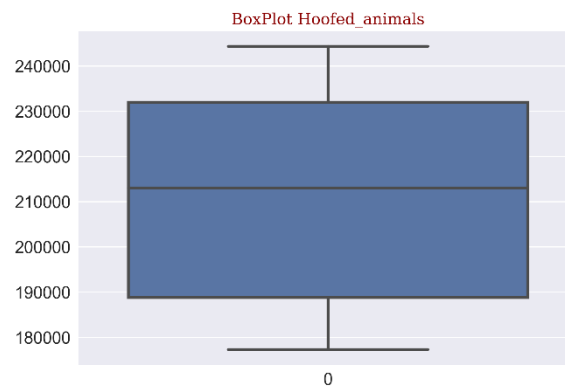
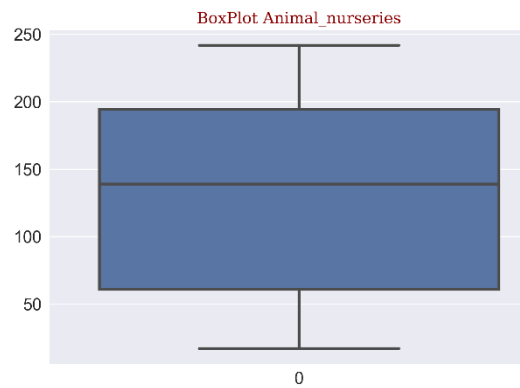
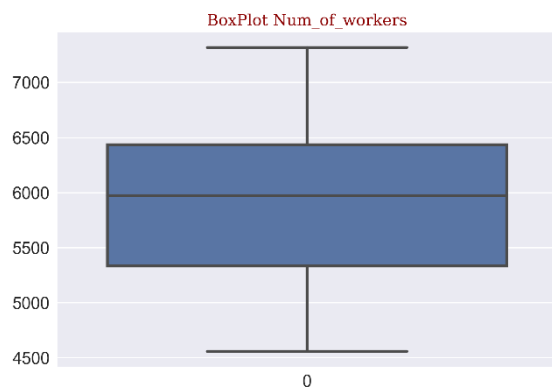
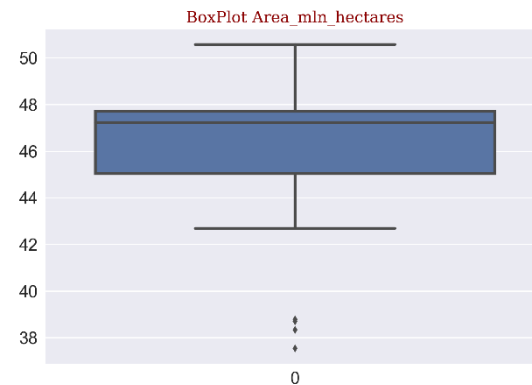
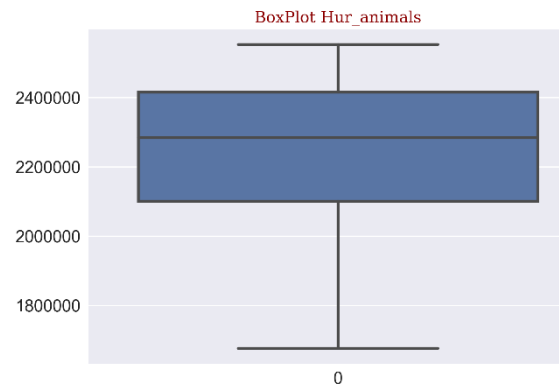
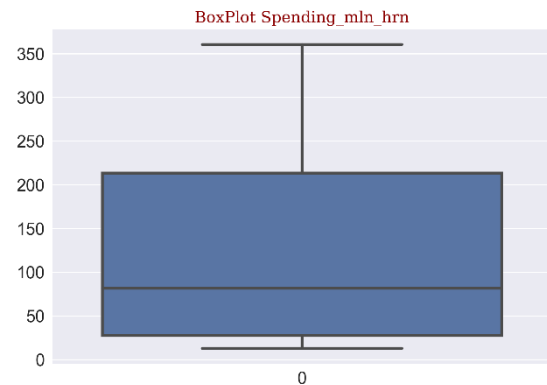
	Area mln hectares	Spending mln hrn	Num_of_workers	Hoofed_animals	Hur_animals	Birds	Animal_nurseries
Area mln hectares	1.000000	-0.964992	-0.838509	-0.712027	0.954828	-0.221909	-0.963862
Spending mln hrn	-0.964992	1.000000	0.869001	0.719932	-0.969509	0.300960	0.981931
Num_of_workers	-0.838509	0.869001	1.000000	0.836251	-0.809147	0.503106	0.858837
Hoofed_animals	-0.712027	0.719932	0.836251	1.000000	-0.692829	0.705251	0.722191
Hur_animals	0.954828	-0.969509	-0.809147	-0.692829	1.000000	-0.229814	-0.952569
Birds	-0.221909	0.300960	0.503106	0.705251	-0.229814	1.000000	0.262564
Animal_nurseries	-0.963862	0.981931	0.858837	0.722191	-0.952569	0.262564	1.000000



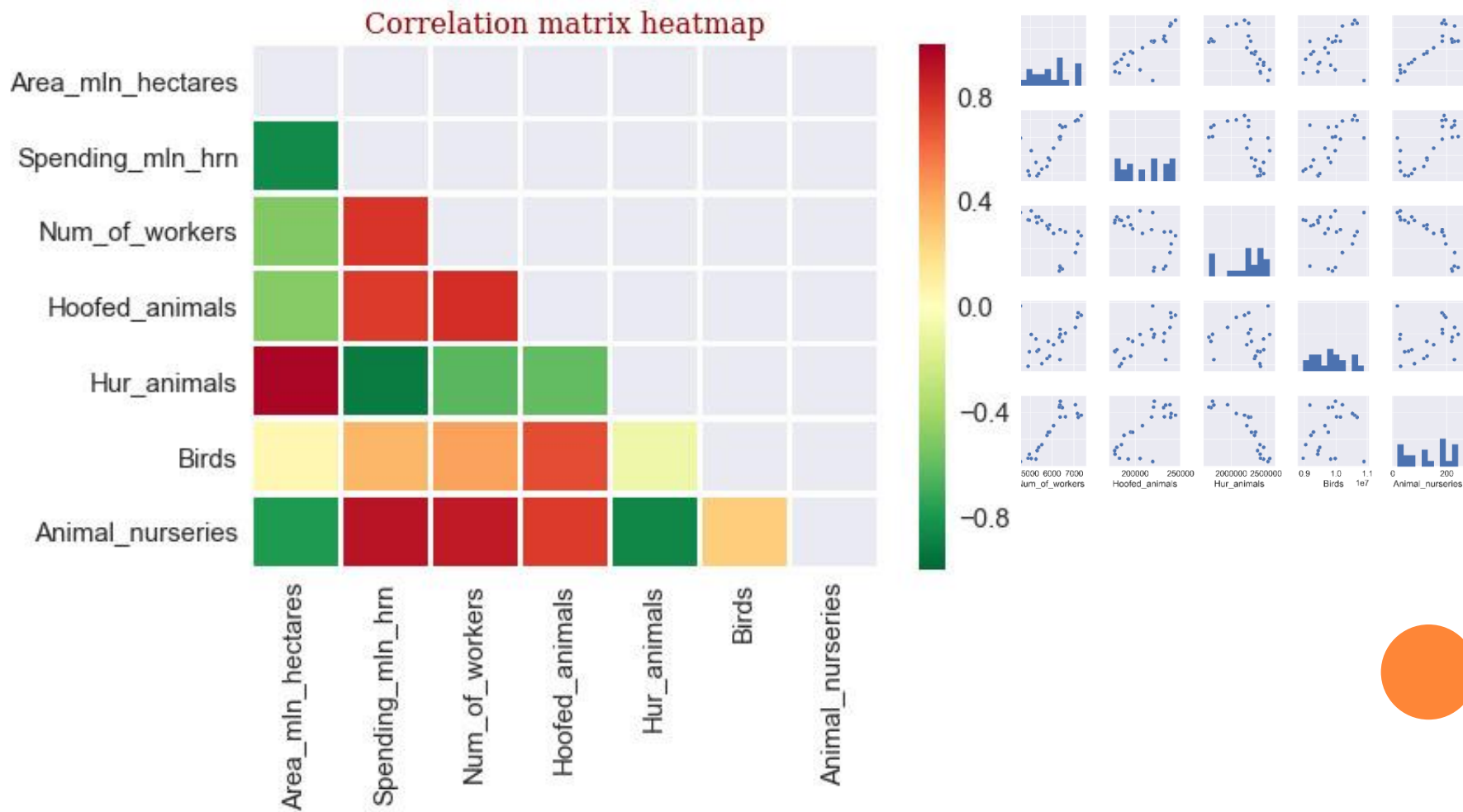
ГРАФИКИ ТИПА “ЯЩИК С УСАМИ”



ДИАГРАММЫ РАЗМАХА



КОРРЕЛЯЦИОННЫЕ ДИАГРАММЫ



РЕГРЕССИОННЫЙ АНАЛИЗ

- Площадь охотничьих угодий ('Area_mln_hectares')
- Расходы на ведение охотничьего хозяйства ('Spending_mln_hrn')
- Количество штатных работников, занятых в охотничьем хозяйстве ('Num_of_workers')
- Количество питомников, вольеров, ферм ('Animal_nurseries')

Набор
факторов
(регрессоры)

- Копытные ('Hoofed_animals')
- меховые ('Hur_animals')
- Дичь ('Birds')

Зависимые
величины



РЕГРЕССИОННЫЙ АНАЛИЗ

КОПЫТНЫЕ ЖИВОТНЫЕ (HOOFED_ANIMALS)

OLS model: Hoofed_animals ~ Num_of_workers

Dep. Variable:	Hoofed_animals	R-squared:	0.648
Model:	OLS	Adj. R-squared:	0.630
Method:	Least Squares	F-statistic:	36.79
Date:	Sat, 23 Jun 2018	Prob (F-statistic):	6.28e-06
Time:	23:04:44	Log-Likelihood:	-240.29
No. Observations:	22	AIC:	484.6
Df Residuals:	20	BIC:	486.8
Df Model:	1		
Covariance Type:	nonrobust		

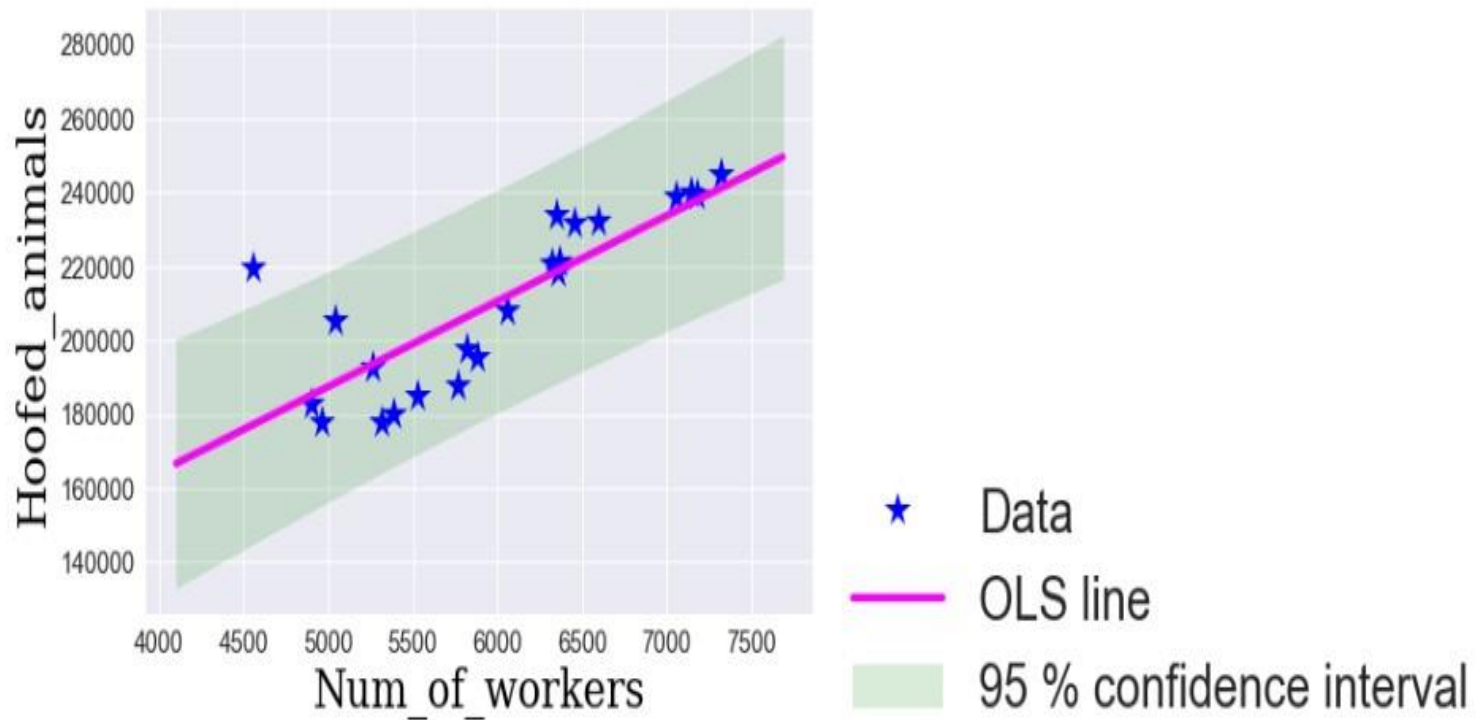
	coef	std err	t	P> t	[95.0% Conf. Int.]
Intercept	7.174e+04	2.3e+04	3.115	0.005	2.37e+04 1.2e+05
Num_of_workers	23.1370	3.815	6.065	0.000	15.180 31.095



РЕГРЕССИОННЫЙ АНАЛИЗ

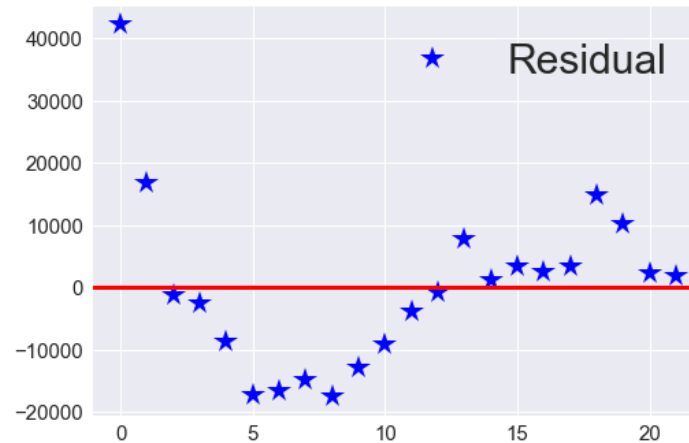
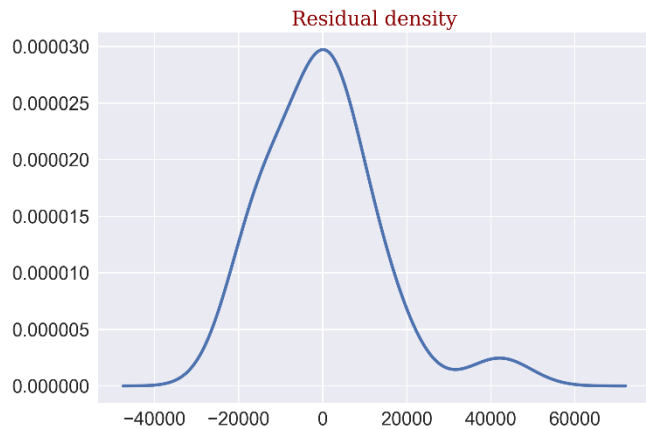
КОПЫТНЫЕ ЖИВОТНЫЕ (HOOFED_ANIMALS)

Prediction for the number of hoofed animals depending on the number of workers



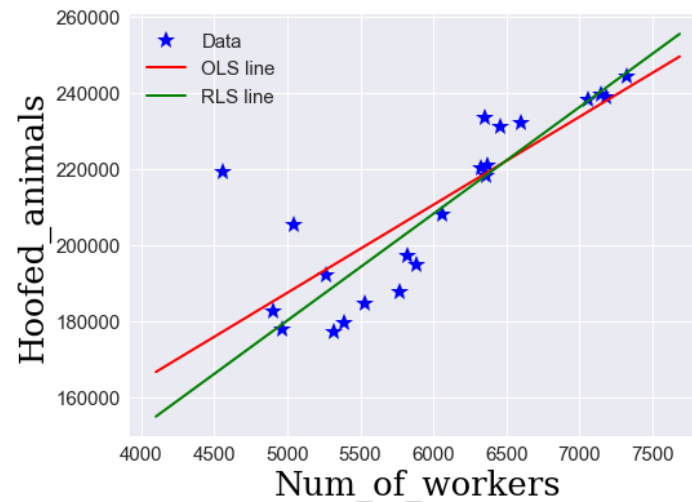
РЕГРЕССИОННЫЙ АНАЛИЗ

КОПЫТНЫЕ ЖИВОТНЫЕ (HOOFED_ANIMALS)



**ОТСУТСТВИЕ
ГОМОСКЕДАСТИЧНОСТИ
ОЦЕНКИ МОДЕЛИ
(ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ) НЕ
ДОСТОВЕРНЫ.**

Prediction for the number of hoofed animals depending on the number of workers



**WLS MODEL?
ROBUST LINEAR MODEL**



РЕГРЕССИОННЫЙ АНАЛИЗ 2

МЕХОВЫХ ЖИВОТНЫЕ (HUR_ANIMALS)

OLS model Hur_animals ~ Area_mln_hectares + Animal_nurseries

OLS Regression Results

Dep. Variable:	Hur_animals	R-squared:	0.977
Model:	OLS	Adj. R-squared:	0.974
Method:	Least Squares	F-statistic:	395.6
Date:	Sat, 23 Jun 2018	Prob (F-statistic):	3.28e-16
Time:	23:31:48	Log-Likelihood:	-265.27
No. Observations:	22	AIC:	536.5
Df Residuals:	19	BIC:	539.8
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[95.0% Conf. Int.]
Intercept	-1.684e+05	2.13e+05	-0.793	0.438	-6.13e+05 2.76e+05
Area_mln_hectares	5.529e+04	4181.297	13.224	0.000	4.65e+04 6.4e+04
Animal_nurseries	-1025.7405	202.399	-5.068	0.000	-1449.366 -602.115



РЕГРЕССИОННЫЙ АНАЛИЗ 2

МЕХОВЫХ ЖИВОТНЫЕ (HUR_ANIMALS)

OLS model Hur_animals ~ 0 + Area_mln_hectares + Animal_nurseries

OLS Regression Results

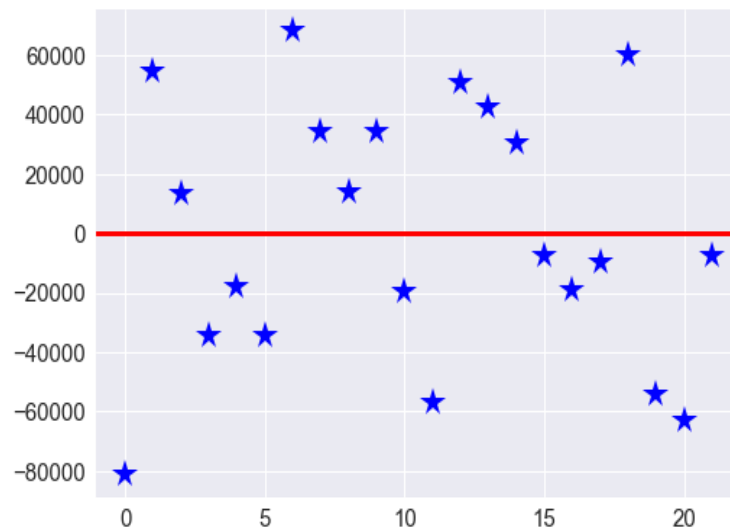
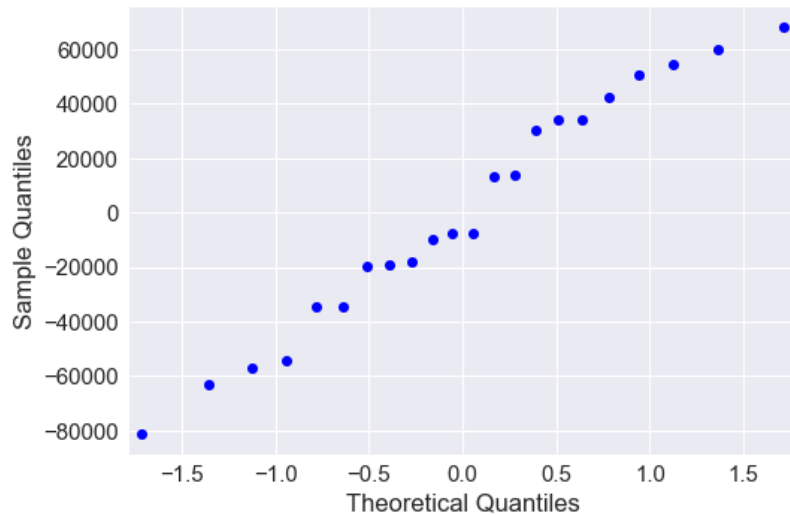
Dep. Variable:	Hur_animals	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	2.762e+04
Date:	Sat, 23 Jun 2018	Prob (F-statistic):	3.85e-35
Time:	23:38:26	Log-Likelihood:	-265.63
No. Observations:	22	AIC:	535.3
Df Residuals:	20	BIC:	537.4
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[95.0% Conf. Int.]
Area_mln_hectares	5.199e+04	377.289	137.803	0.000	5.12e+04 5.28e+04
Animal_nurseries	-1158.7853	112.013	-10.345	0.000	-1392.440 -925.131



РЕГРЕССИОННЫЙ АНАЛИЗ 2

МЕХОВЫХ ЖИВОТНЫЕ (HUR_ANIMALS)

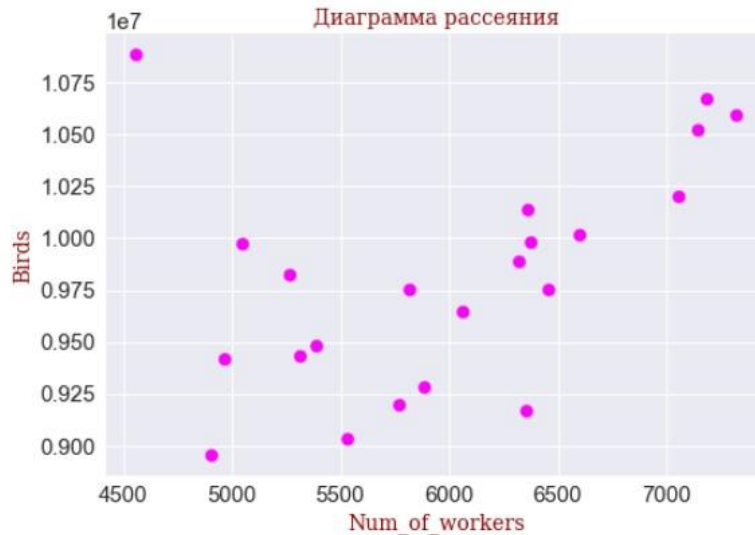


★ Residual



РЕГРЕССИОННЫЙ АНАЛИЗ 3

Дичь (BIRDS)



Dep. Variable:	Birds	R-squared:	0.189
Model:	OLS	Adj. R-squared:	0.149
Method:	Least Squares	F-statistic:	4.671
Date:	Sat, 23 Jun 2018	Prob (F-statistic):	0.0430
Time:	23:54:20	Log-Likelihood:	-318.95
No. Observations:	22	AIC:	641.9
Df Residuals:	20	BIC:	644.1
Df Model:	1		
Covariance Type:	nonrobust		

Для прогноза по дичи нужны другие данные которые описывают количество ПЕРНАТЫХ.

ЛИБО НЕОБХОДИМО БРАТЬ НЕЛИНЕЙНУЮ МОДЕЛЬ



Спасибо за внимание

