

▼ Отчёт по лабораторной работе по курсу SAS

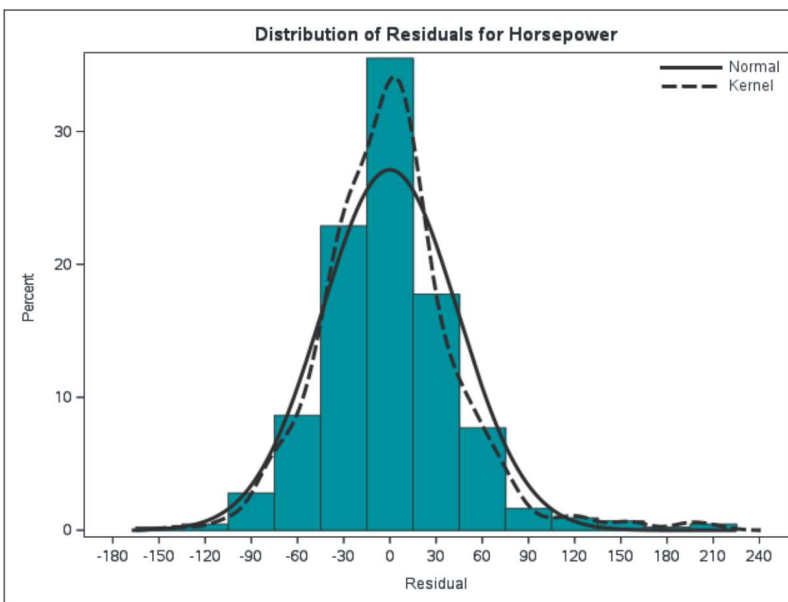
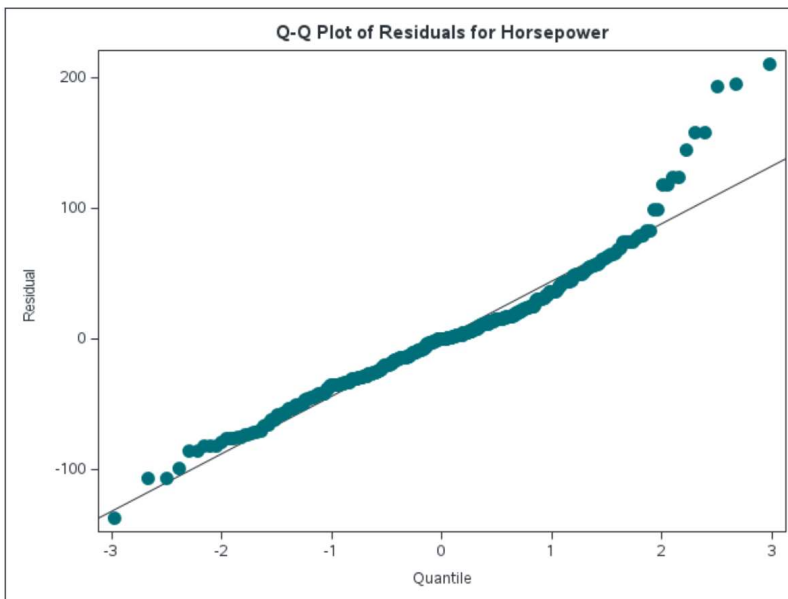
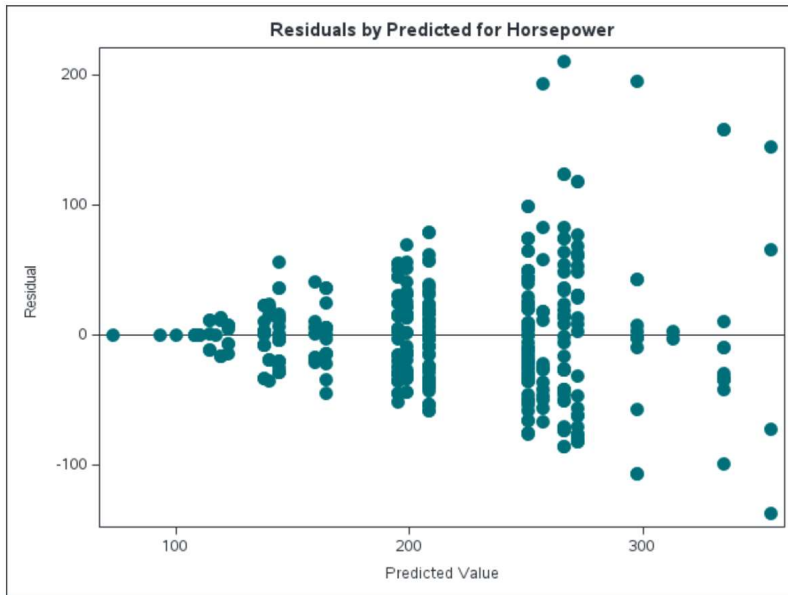
Выполнил: Сухарников Андрей, группа Б05-812

▼ Дисперсионный анализ

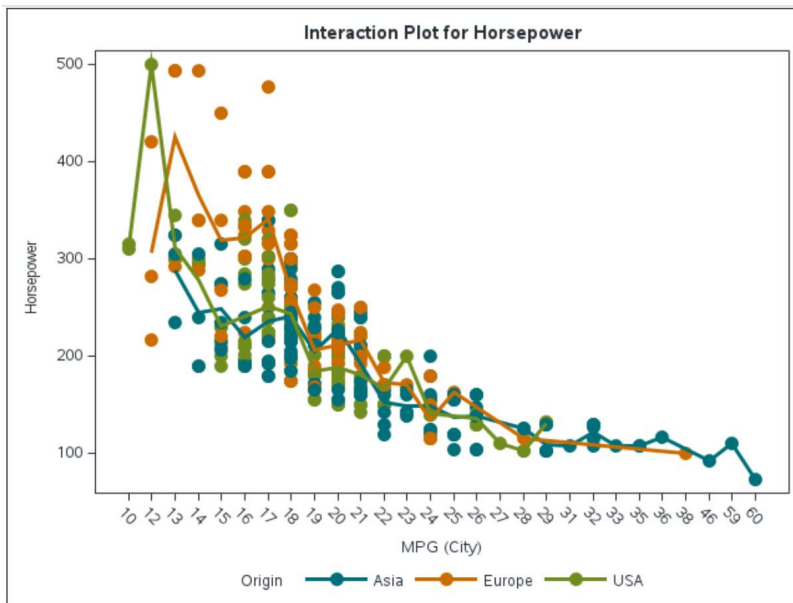
1) Для levene's test имеем p-value менее 0.001, поэтому гипотеза о равенстве групповых дисперсий отвергается на (стандартном) уровне значимости 0.05.

Levene's Test for Homogeneity of Horsepower Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
MPG_City	17	1.5093E9	88784854	5.11	<.0001
Error	398	6.9108E9	17363849		

2) По QQ-plot для остатков и диаграмме можно наблюдать, что остатки имеют нормальное распределение. По графику остатков от предсказанной переменной можно видеть неоднородность дисперсий (увеличение разброса слева направо), что является дополнительным подтверждением вывода из пункта 1.



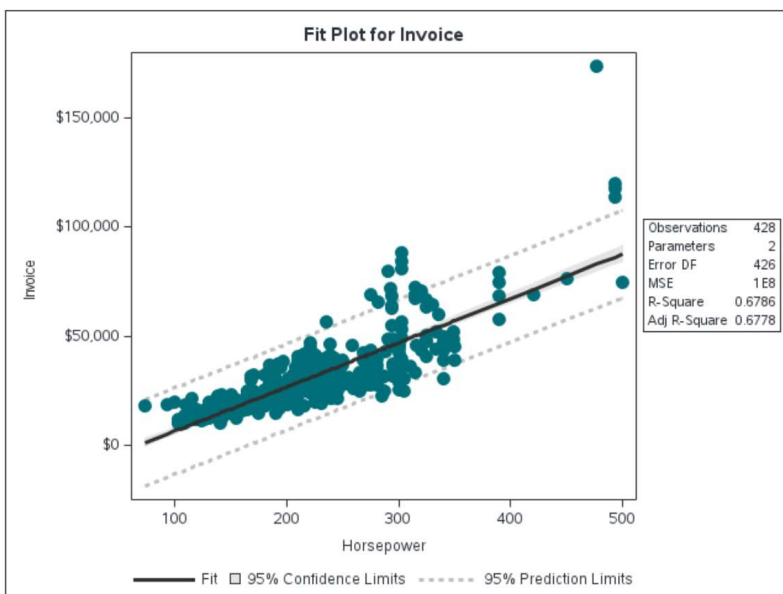
3) Переменные являются взаимодействующими: это можно наблюдать из соответствующего графика (кривые пересекаются).

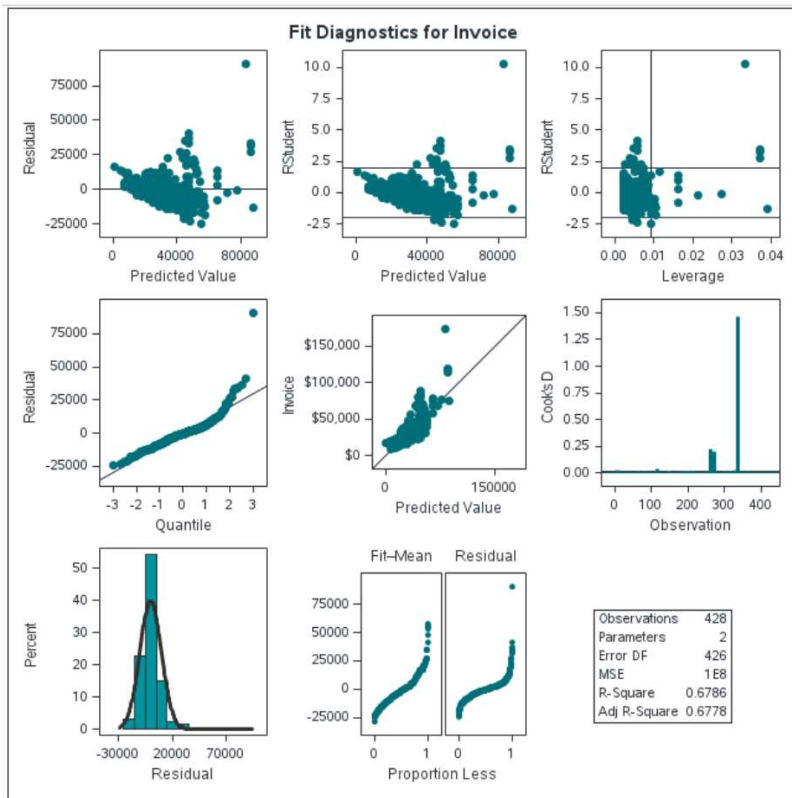


▼ Линейная регрессия

1) Для всех параметров модели p -value менее 0.05, поэтому гипотеза о том, что эти параметры являются нулевыми отвергается, и переменные можно считать зависимыми. Далее, исследуем корректность. По графику зависимости Invoice от HorsePower можно визуально оценить, что приближение линейной моделью вполне корректно (точки примерно ложатся на прямую). Остатки также можно считать нормальными из QQ-plot и диаграммы, однако увеличение разброса с ростом предсказанной переменной (на графике остатков от предсказанной переменной) может говорить о неоднородной дисперсии.

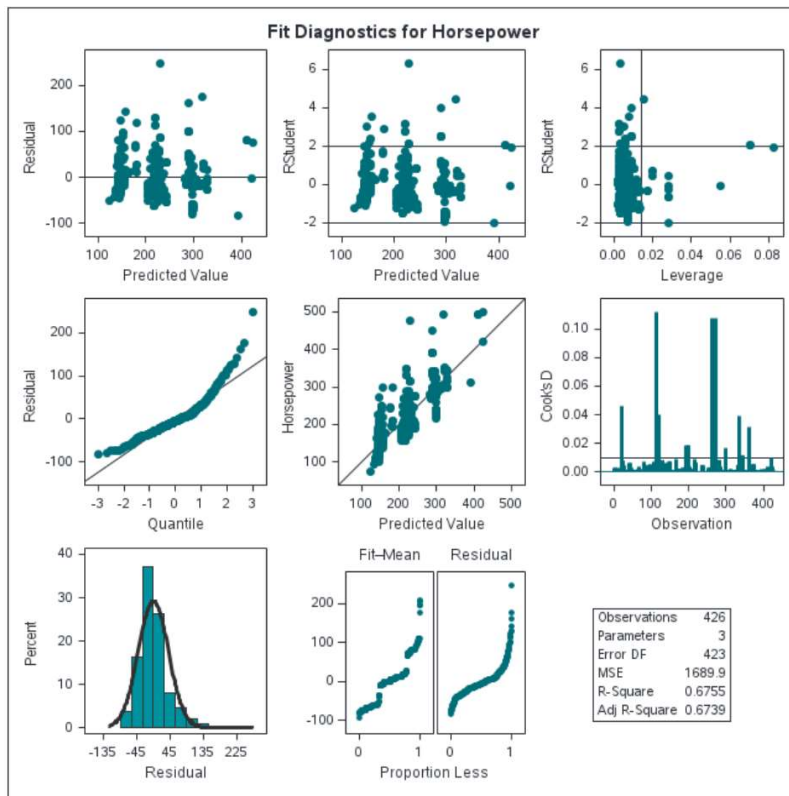
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-13660	1534.72473	-8.90	<.0001
Horsepower	1	202.30283	6.74613	29.99	<.0001





2) Для параметров модели p -value менее 0.05, поэтому гипотеза о том, что эти параметры являются нулевыми отвергается, и можно считать, что модель множественной регрессии существует. Однако, из QQ-plot для остатков можно увидеть, что они, вероятно, имеют ненормальное распределение. Это подтверждается в пункте 4. Поэтому применение линейной регрессии, скорее всего, будет некорректным;

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	9.76682	8.02354	1.22	0.2242
EngineSize	Engine Size (L)	1	21.35488	4.31305	4.95	<.0001
Cylinders		1	23.70296	3.05403	7.76	<.0001



3) Для каждого из критериев (RSquare, Adjusted RSquare, Mallows' Cp) наиболее оптимальным является добавление всех независимых переменных.

Summary of Forward Selection								
Step	Variable Entered	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	Cylinders		1	0.6567	0.6567	54.6802	810.90	<.0001
2	EngineSize	Engine Size (L)	2	0.0188	0.6755	30.5678	24.51	<.0001
3	Wheelbase	Wheelbase (IN)	3	0.0178	0.6933	7.8311	24.51	<.0001
4	Weight	Weight (LBS)	4	0.0035	0.6968	5.0000	4.83	0.0285

4) Все представленные критерии проверки нормальности дают p -value менее 0.001, значит гипотеза о нормальности отвергается;

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.906185	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.141318	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.594449	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	9.076174	Pr > A-Sq	<0.0050

5) Так как коэффициент корреляции между остатком и предсказанными значениями значительно отличен от нуля (≈ 0.1), то дисперсию ошибки можно считать непостоянной.

Spearman Correlation Coefficients, N = 426 Prob > r under H0: Rho=0		
	abserror	pred
abserror	1.00000	0.10547 0.0295
pred Predicted Value of Horsepower	0.10547 0.0295	1.00000

✓ 0 сек. выполнено в 21:51

×