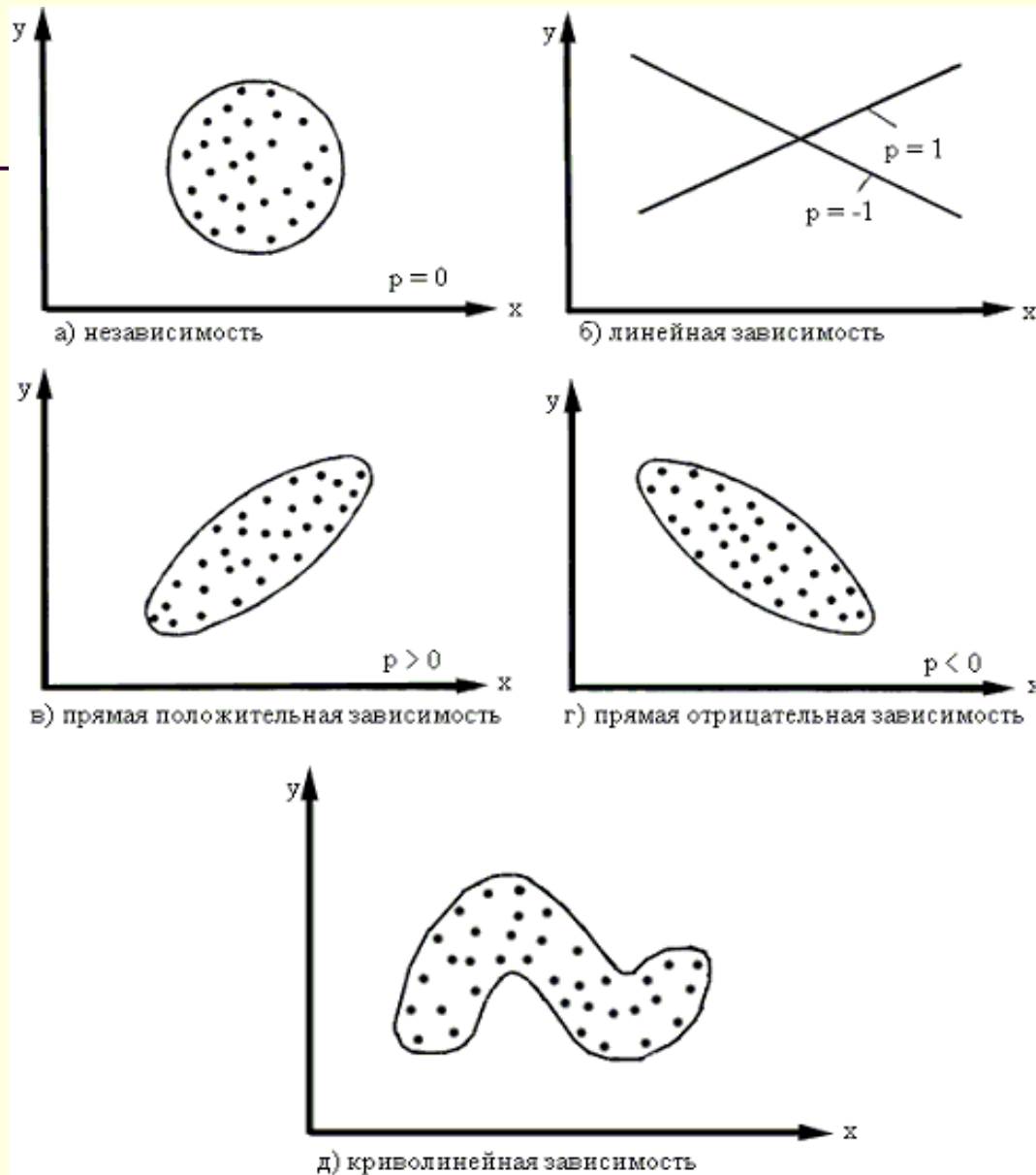


Корреляционный анализ данных. Графический метод.



Корреляционный и регрессионный анализ данных. Множественный регрессионный анализ.

Значения взаимных корреляций основных параметров ДВС

а) по результатам испытаний двигателя ВАЗ-21081

б)

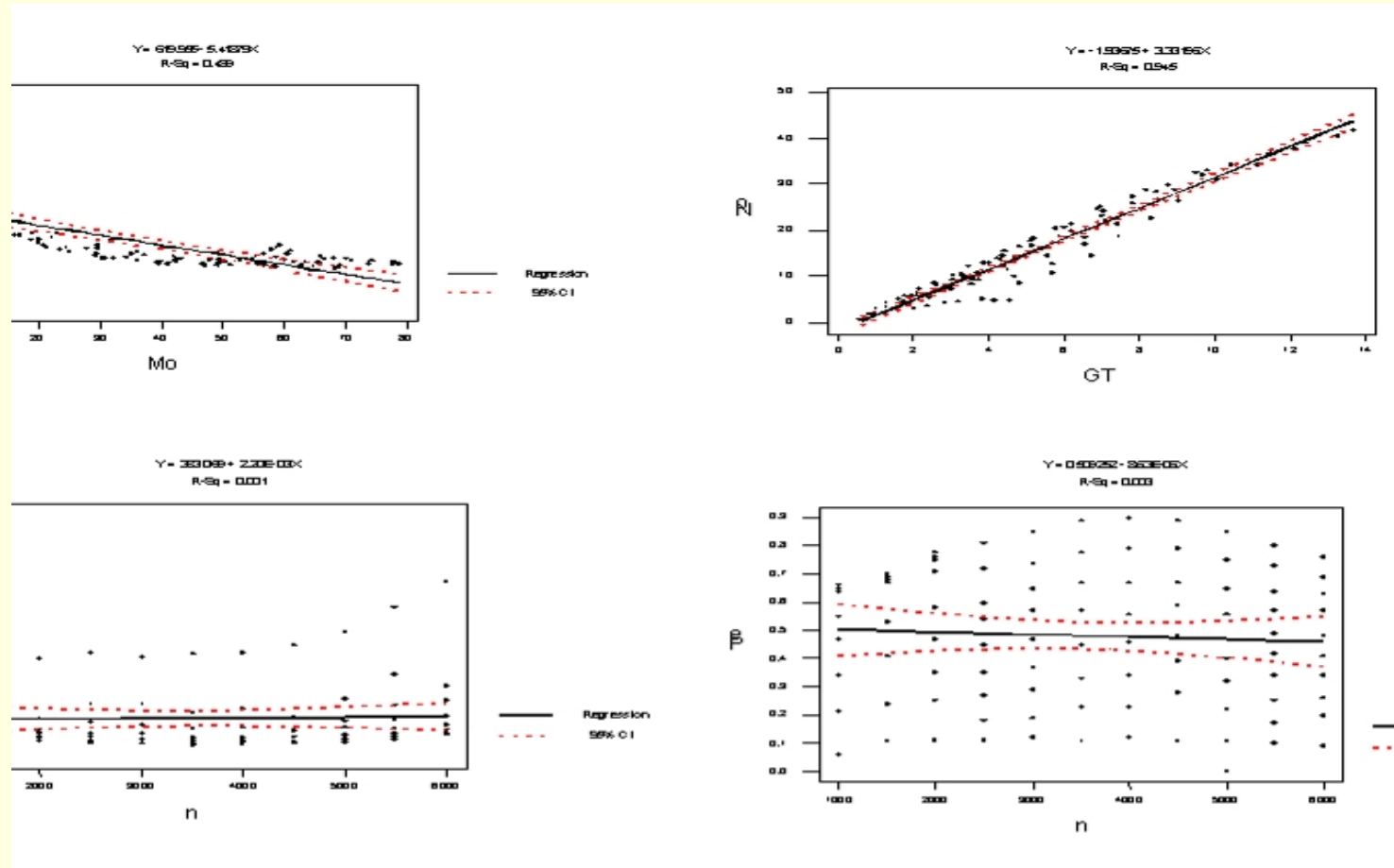
	$N_o, \text{мин}^{-1}$	$M_o, \text{Н} \cdot \text{м}$	$P_e, \text{МПа}$	$G_e, \text{кг}(\text{Квт} \cdot \text{час})$	$G_t, \text{кг/час}$
$N_o, \text{мин}^{-1}$	1.000				
$M_o, \text{Н} \cdot \text{м}$	0.695	1.000			
$P_e, \text{МПа}$	0.695	1.000	1.000		
$G_e, \text{кг}(\text{Квт} \cdot \text{час})$	-0.498	-0.699	-0.699	1.000	
$G_t, \text{кг/час}$	0.972	0.572	0.573	-0.359	1.000

с) по результатам испытаний двигателя МеМЗ-245

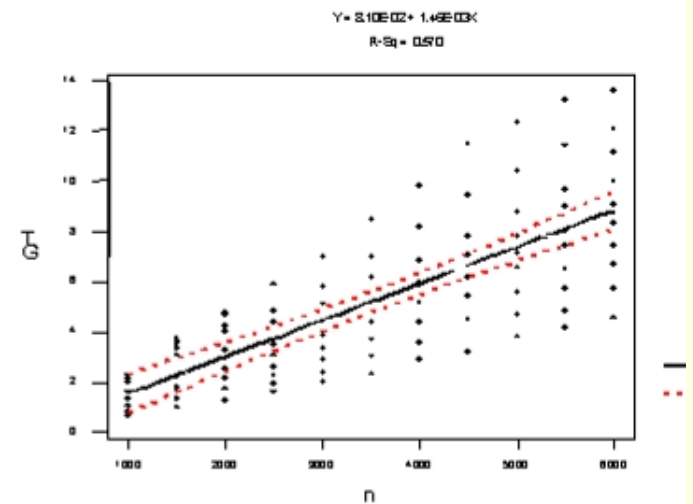
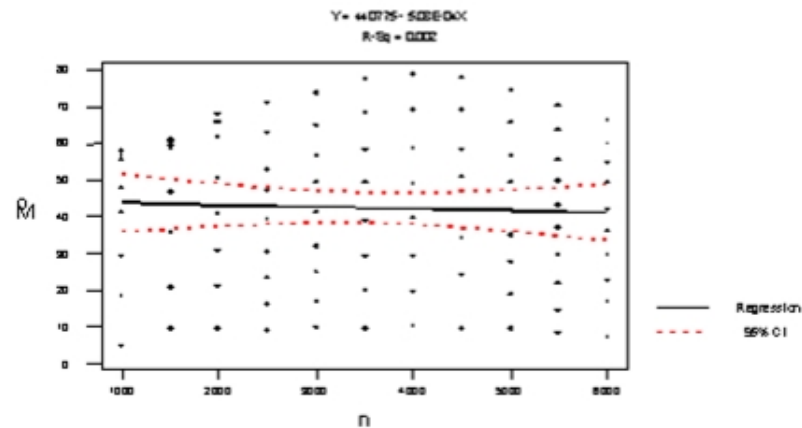
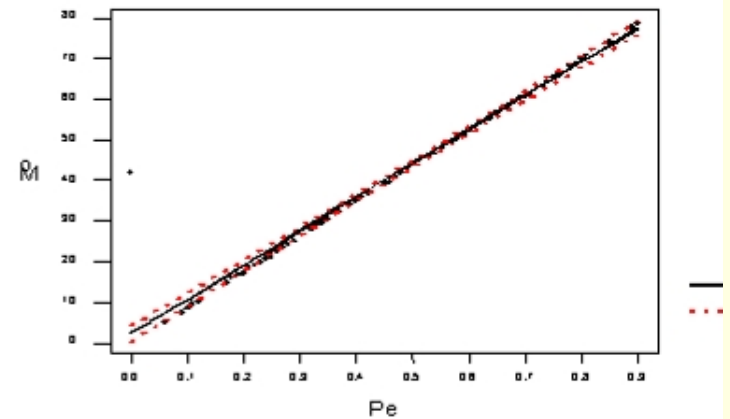
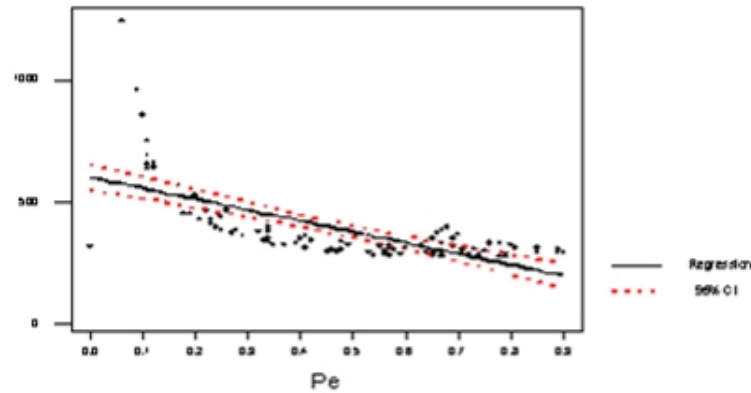
д)

	$M, \text{Н} \cdot \text{м}$	$N_e, \text{кВт}$	$G_t, \text{кг/час}$	$Q_e, \text{кг/час}$	$D_{h_0}, \text{мм рт. ст.}$	$Q, \text{м}^3/\text{час}$	$G_e, \text{кг}(\text{Квт} \cdot \text{час})$
$M, \text{Н} \cdot \text{м}$	1.000						
$N_e, \text{кВт}$	0.710	1.000					
$G_t, \text{кг/час}$	0.595	0.978	1.000				
$Q_e, \text{кг/час}$	-0.731	-0.445	-0.265	1.000			
$D_{h_0}, \text{мм рт. ст.}$	-0.986	-0.706	-0.610	0.680	1.000		
$Q, \text{м}^3/\text{час}$	-0.752	-0.154	-0.043	0.535	0.784	1.000	
$G_e, \text{кг}(\text{Квт} \cdot \text{час})$	0.145	0.682	0.744	-0.006	0.134	0.426	1.000

Корреляционный и регрессионный анализ данных. Множественный регрессионный анализ.



Корреляционный и регрессионный анализ данных. Множественный регрессионный анализ.



Шкала Чеддока

Количественная мера тесноты связи	Качественная характеристика силы связи
0,1 - 0,3	Слабая
0,3 - 0,5	Умеренная
0,5 - 0,7	Заметная
0,7 - 0,9	Высокая
0,9 - 0,99	Весьма высокая

Корреляционный и регрессионный анализ
данных.
Множественный регрессионный анализ.

Парный коэффициент корреляции:

Линейный коэффициент корреляции Пирсона:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n \sigma_x \sigma_y}$$

Регрессионный анализ.

- Впервые термин употреблен в работе Pearson (1908)
- Анализ связи между несколькими независимыми переменными (регрессорами или предикторами) и зависимой переменной

Регрессионный анализ.

- Впервые термин употреблен в работе Pearson (1908)
- Анализ связи между несколькими независимыми переменными (регрессорами или предикторами) и зависимой переменной

Условия применения регрессионного анализа.

- Использование метрических переменных
- Равенство условных дисперсий: $D(Y / X) = const$;
- Независимость ошибок от предикторов и нормального их распределения с нулевым средним и постоянной дисперсией;
- Попарное нормальное распределение всех признаков модели;
- Независимость предикторов между собой
- Достаточное количество наблюдений (обычно >15 , в зависимости от конкретного характера распределений наблюдений и сложности искомой зависимости)

Регрессионный анализ.

$$Y = a + b \cdot X; \text{ где:}$$

- Y – зависимая переменная,
- a - константа
- b - угловой коэффициент
- X – независимая переменная

Для многомерной регрессии:

$$Y = a + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_p \cdot X_p$$

Метод наименьших квадратов.

- Цель - минимизировать квадраты отклонений линии регрессии от наблюдаемых точек.
- По этим данным строим диаграмму рассеяния

Остаток

Отклонение отдельной точки от линии регрессии (от предсказанного значения) называется остатком.

Чем меньше разброс значений (дисперсия) остатков около линии регрессии по отношению к общему разбросу значений, тем лучше прогноз

Регрессионные коэффициенты.

Это независимые вклады каждой независимой переменной в предсказание зависимой переменной:

переменная X_1 коррелирует с переменной Y после учета влияния всех других независимых переменных (частная корреляция)

Пример.

Успеваемость = $1 + 0.02 \cdot IQ$, где:

$$a = 1$$

$$b = 0,02$$

IQ – независимая переменная

При $IQ=130$:

$$\text{Успеваемость} = 1 + 0.02 \cdot 130 = 3,6$$

Регрессионный анализ.

- Цель - минимизировать квадраты отклонений линии регрессии от наблюдаемых точек.
- По этим данным строим диаграмму рассеяния

Корреляционный и регрессионный анализ данных. Множественный регрессионный анализ.

Для аналитического описания связи между признаками могут быть использованы следующие виды уравнений:

- 1) $\bar{y} = a_0 + a_1x$ — прямая, линейная функция;
- 2) $\bar{y} = a_0 + a_1x^2 + a_2x$ — парабола;
- 3) $\bar{y} = a_0 + a_1 \frac{1}{x}$ — гипербола;
- 4) $\bar{y} = a_0x^{a_1}$ — степенная функция;
- 5) $\bar{y} = \exp(a_0 + a_1x)$ — экспонента и др.