



MONASH University

Information Technology

程序代写代做 CS编程辅导

FIT1006



Business Information Analysis

WeChat: cstutorcs

Assignment Project Exam Help

Lecture 8

Email: tutorcs@163.com

Linear Regression

QQ: 749389476

<https://tutorcs.com>

Topics covered:

程序代写代做 CS编程辅导



- Estimating the regression equation by eye.
- Fitting a regression model using Excel and SYSTAT.
- Measuring the goodness of fit.
- Modelling with the regression equation.

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

程序代写代做 CS编程辅导

Linear Regression

- Regression is the process of describing the (linear) relationship between two or more quantitative variables.
- Thus if we know the value of one variable, we can estimate the value of the related variable of interest.
- Origin: The 19th century scientist Francis Galton collected data on the heights of fathers and their sons. He found that tall fathers had slightly shorter sons and that short fathers had slightly taller sons. Thus in each case there was a regression (reversion) to the mean. Over time the details of the investigation have been forgotten but the name has stuck to this method of modelling.



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutores@163.com

QQ: 749389476

<https://tutores.com>

Motivating Question

程序代写代做 CS编程辅导

- In 1998, *Choice* tested 1500 toothbrushes



- A summary of price and functionality scores is on the right.

WebChat: cstutorcs

- What is the relationship between price and functionality?

Assignment Project Exam Help

Email: tutorcs@163.com

- How reliable is the model?

QQ: 749389476

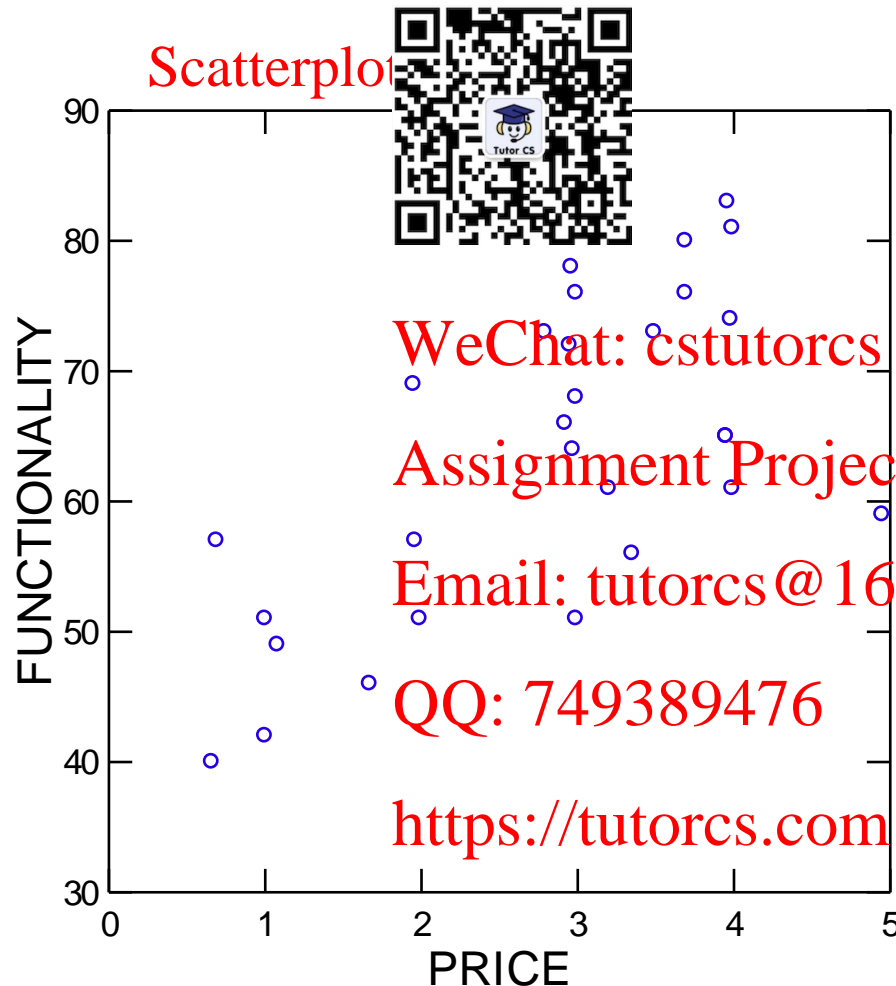
(Selvanathan 4th Ed p 679) <https://tutorcs.com>

- Answers later...

Price	Functionality
3.96	83
3.99	81
3.69	80
2.96	78
3.69	76
2.99	76
3.98	74
2.79	73
3.49	73
2.95	72
1.95	69
2.99	68
2.92	66
3.95	65
3.95	65
2.97	64
3.99	61
3.20	61
4.95	59
0.69	57
1.96	57
3.35	56
1.00	51
2.99	51
1.99	51
1.08	49
1.67	46
1.00	42
0.66	40

Motivating Question

程序代写代做 CS编程辅导



Price	Functionality
3.96	83
3.99	81
3.69	80
2.96	78
3.69	76
2.99	76
3.98	74
2.79	73
3.49	73
2.95	72
1.95	69
2.99	68
2.92	66
3.95	65
3.95	65
2.97	64
3.99	61
3.20	61
4.95	59
0.69	57
1.96	57
3.35	56
1.00	51
2.99	51
1.99	51
1.08	49
1.67	46
1.00	42
0.66	40

程序代写代做 CS编程辅导

The underlying assumption



When we calculate regression of y on x , we are assuming that the relationship between x and y is linear and thus we can say that $y = ax + b + e$, where e are random, normally distributed errors.

WeChat: tutores

We want to find the value of a and b .

Assignment Project Exam Help

(Note: the textbook uses slightly different notation)

Email: tutores@163.com

QQ: 749389476

<https://tutorcs.com>

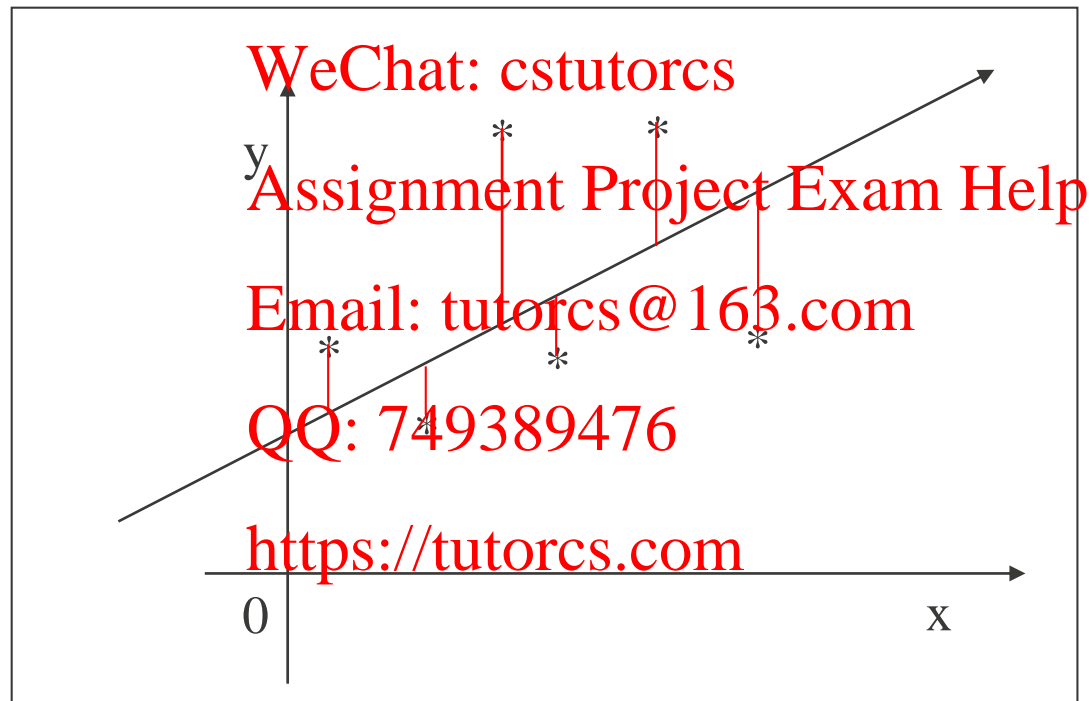


The basic idea

程序代写代做 CS编程辅导



We want to fit a line to the data that minimises the sum of the squared error differences, between the fitted model (line) and the data.

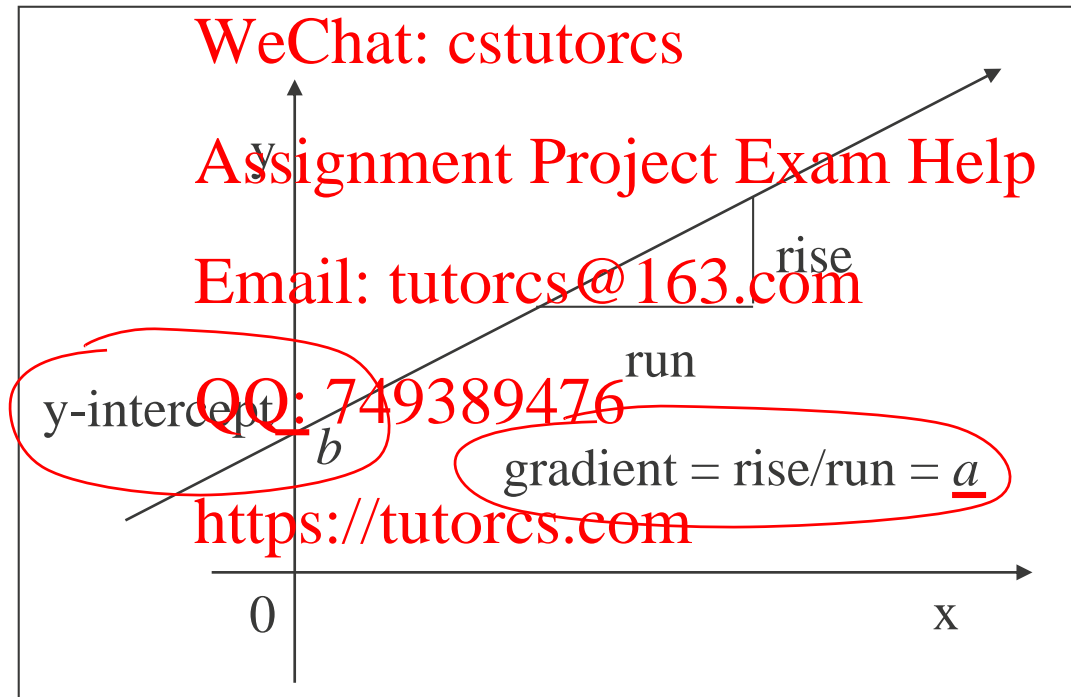


程序代写代做 CS编程辅导

The equation of a straight line



We can use the basic equation of a straight line as the model for our regression equation. A line with gradient 'a' and y-intercept 'b' has equation: $y = ax + b$.



程序代写代做 CS编程辅导

Least Squares Regression



- Ordinary Least Squares Regression minimises the sum of squared residuals in the data.
- The regression of y on x as $y = ax + b$ is:

WeChat: cstutorcs

$$a = \frac{s_{xy}}{s_x^2} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \quad \text{and } b = \bar{y} - a\bar{x}$$

Assignment Project Exam Help

Email: tutormcs@163.com

QQ: 749389476

<https://tutormcs.com>

$$\text{note: } \bar{y} = \frac{\sum y}{n} \text{ etc.}$$

<https://flux.qa> (Feed code: SJ6KGV)

Question 1: In which plot is 'b' greatest?

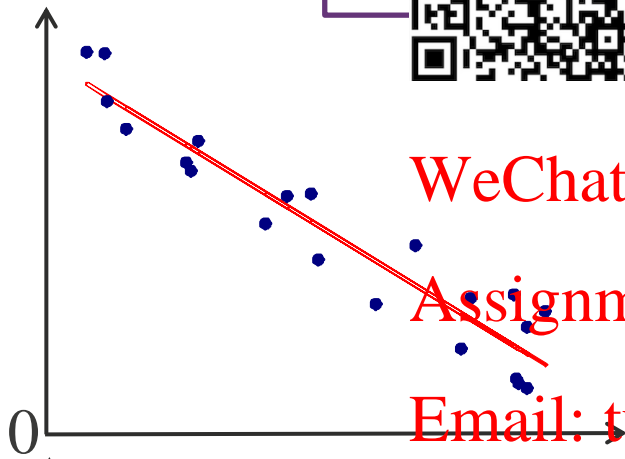
As



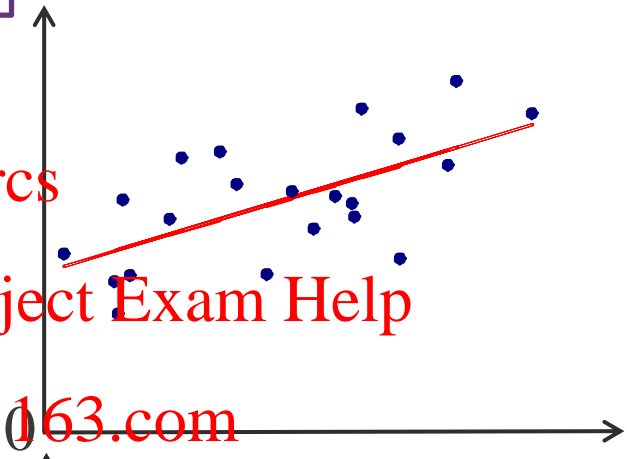
$ax + b$

✓

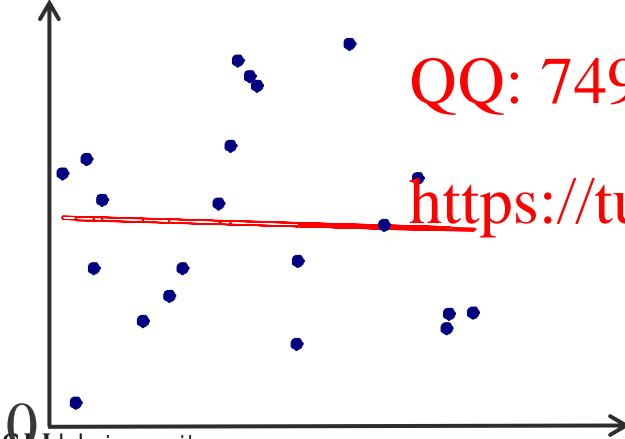
A.



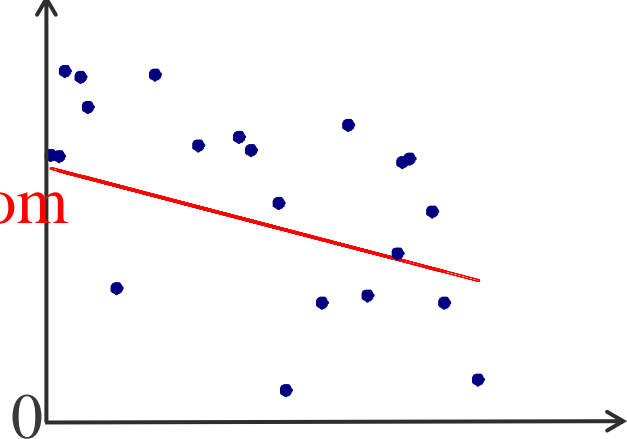
B.



C.



D.



WeChat: cstutorcs

Assignment Project Exam Help

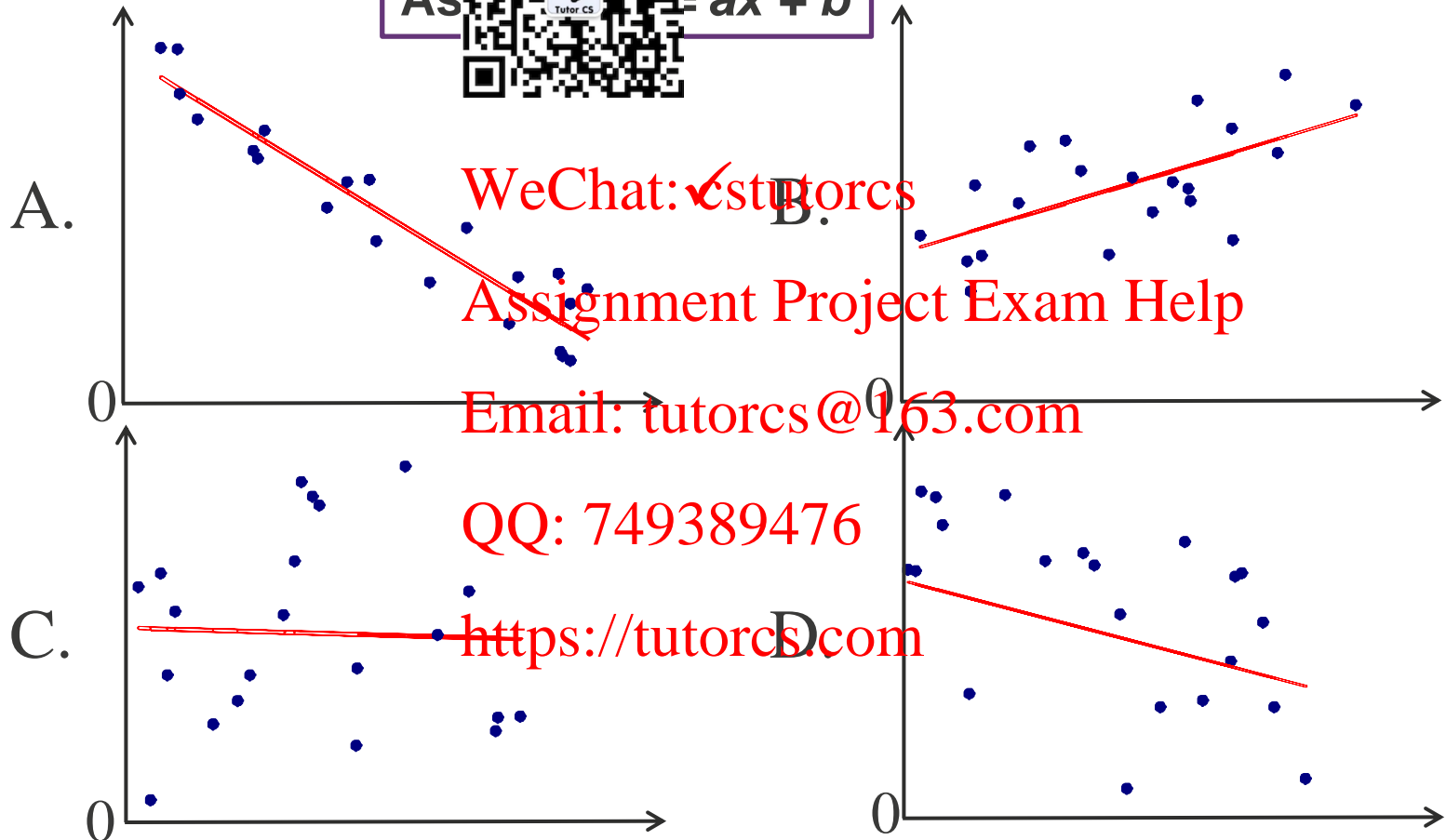
Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

<https://flux.qa> (Feed code: 8U0KGV)

Question 2: Which plot is 'a' greatest



WeChat: ✓stutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

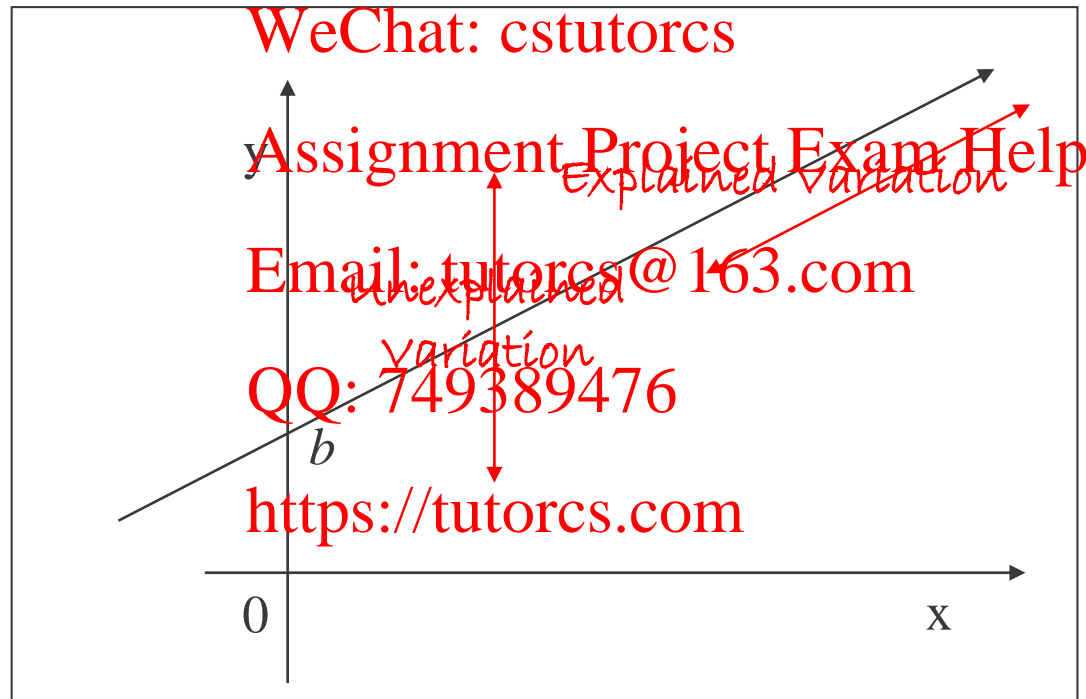
<https://tutorcs.com>

程序代写代做 CS编程辅导

How good is the fit?



- One measure of how well the regression model is the proportion of variation in y that is explained by the regression equation.



程序代写代做 CS编程辅导

Coefficient of Determination



- The coefficient of determination is the proportion of variation in y that is explained by variation in x through the regression equation.

WeChat: cstutorcs

- The coefficient of determination is r^2 – the square of Pearson's correlation coefficient r .

Assignment Project Exam Help

Email: tutorcs@163.com

- Often 100 r^2 is calculated and the result expressed as a percentage.

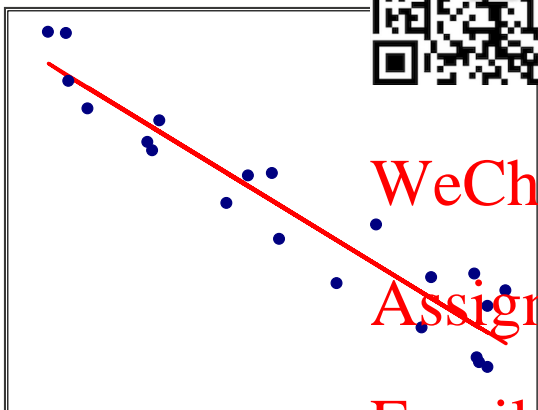
QQ: 749389476

<https://tutorcs.com>

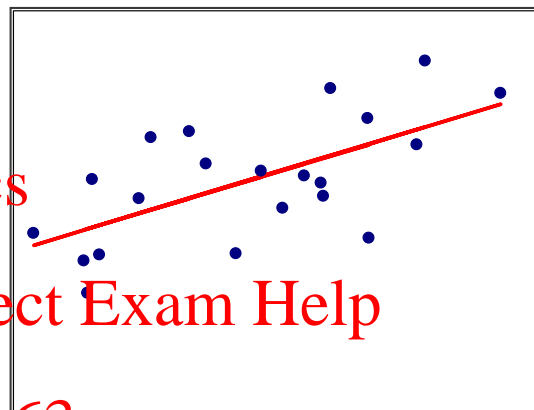
<https://flux.qa> (程序代写代做CS编程辅导)
(Feed code: SJ6KGV)

Question 3: In  plot has r^2 closest to 1?

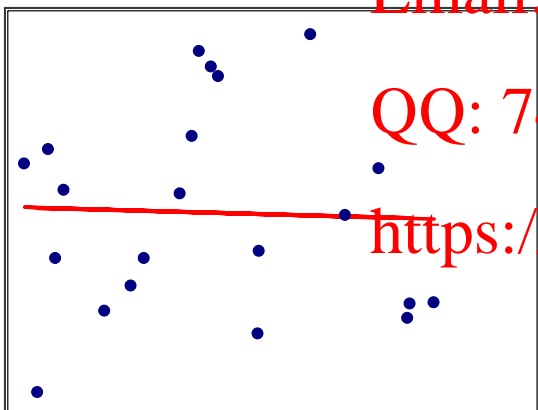
✓ A.



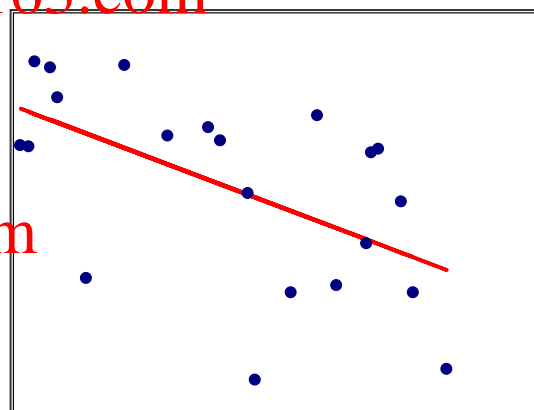
B.



C.



D.



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

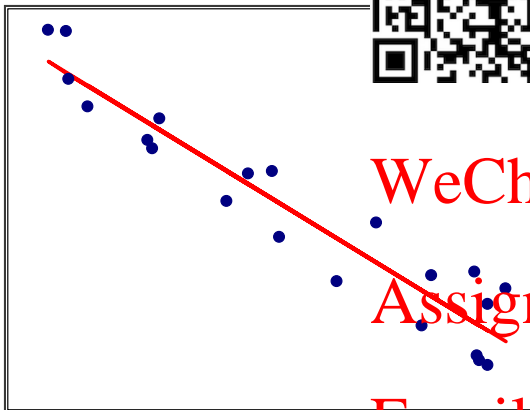
<https://tutorcs.com>

<https://flux.qa> (作代码做CS编程辅导V)

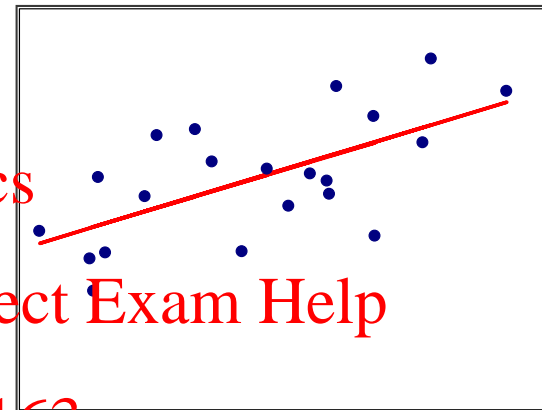
Question 4: In which plot has r^2 closest to 0?



A.



B.



WeChat: cstutorcs

Assignment Project Exam Help

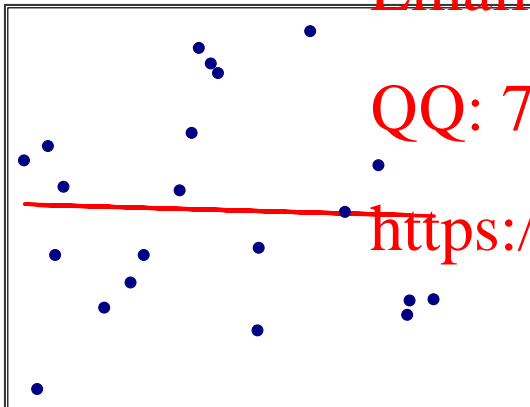
Email: tutorcs@163.com

QQ: 749389476

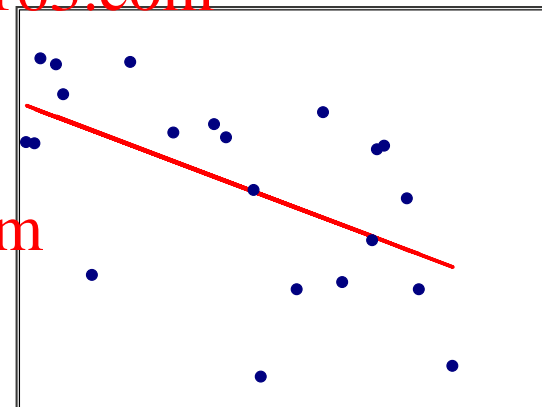
<https://tutorcs.com>



C.



D.



Regression in EXCEL

程序代写代做 CS编程辅导



- Regression is a b analysis function, or you can also calculate formulas manually with:

– $a = \text{SLOPE}(y \text{ values}, x \text{ values})$

– $b = \text{INTERCEPT}(y \text{ values}, x \text{ values})$

- Also,

– $r = \text{CORREL}(y \text{ values}, x \text{ values})$

– $r^2 = \text{CORREL}(y \text{ values}, x \text{ values})^2$

- Regression is also a Chart Tool if you first draw a scatter plot and then choose this option.

x	y
Price	Functionality
3.96	83
3.99	81
3.69	80
2.96	78
3.69	76
2.99	76
3.98	74
2.79	73
3.49	73
2.95	72
1.95	69
2.99	68
2.92	66
3.95	65
3.95	65
2.97	64
3.99	61
3.20	61
4.95	59
0.69	57
1.96	57
3.35	56
1.00	51
2.99	51
1.99	51
1.08	49
1.67	46
1.00	42
0.66	40

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

程序代写代做 CS编程辅导

Regression in SYSTAT



SYSTAT calculates regression and gives a diagnostic output for the fitted model.

- Select: Analyze > Regression > Linear > Least Squares

WeChat: cstutorcs

- The dependent variable is the one we're trying to predict.

Assignment Project Exam Help

- The independent variable is the one that is free to change.

Email: tutorcs@163.com

- The model and residuals can be saved to a data file.

QQ: 749389476

<https://tutorcs.com>

Regression by hand

程序代写代做 CS编程辅导

We stopped here ... will continue the rest next week

- Use the same formulas you calculated for the least squares correlation: Σx , Σy , Σx^2 , Σy^2 , Σxy , and n .

WeChat: cstutorcs

$$a = \frac{s_{xy}}{s_x^2} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} \quad \text{and} \quad b = \bar{y} - a\bar{x}$$

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

$\bar{y} = \frac{\Sigma y}{n}$

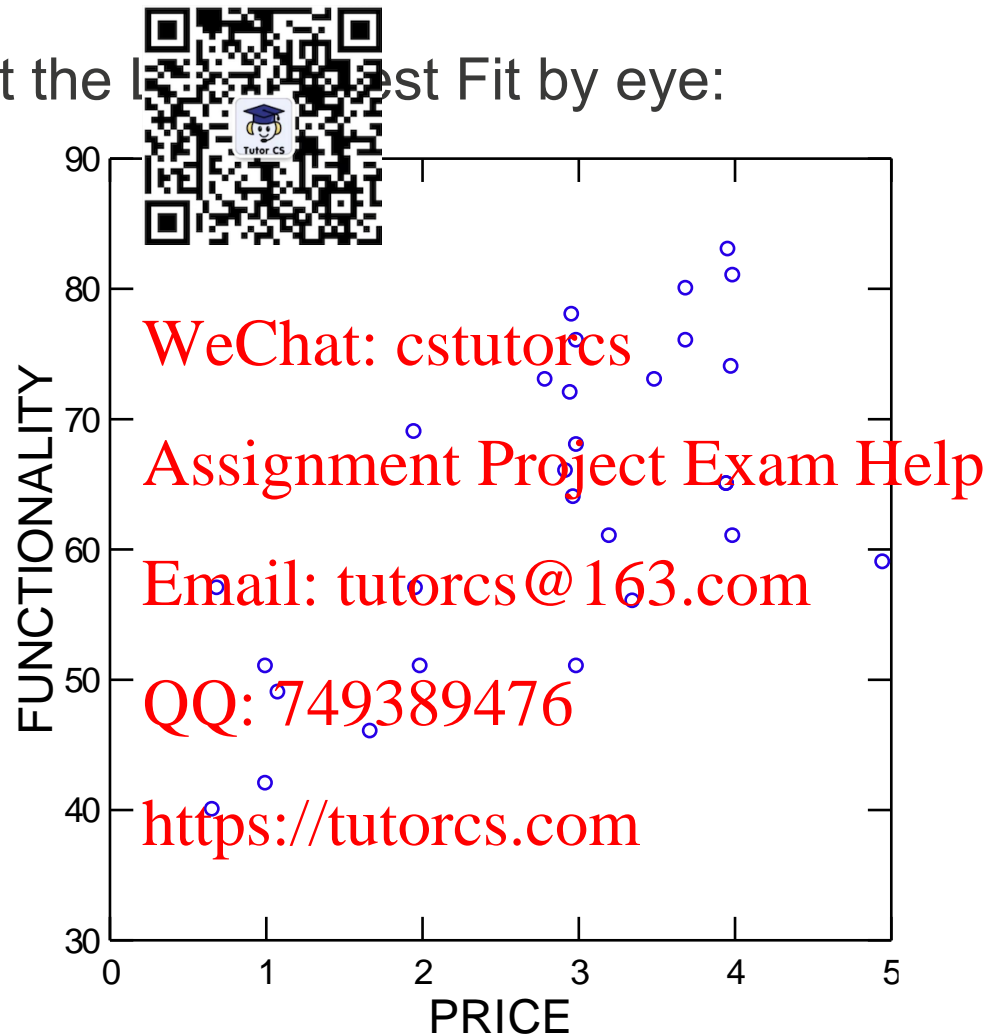
<https://tutorcs.com>

- Know how to calculate the regression equation using your calculator.

Motivating Question

程序代写代做 CS编程辅导

- Let's try and fit the Linear Least Fit by eye:



<https://flux.qa> (Feed code: SJ6KGV)

Question 5: For the toothbrush problem, which assumption is true:

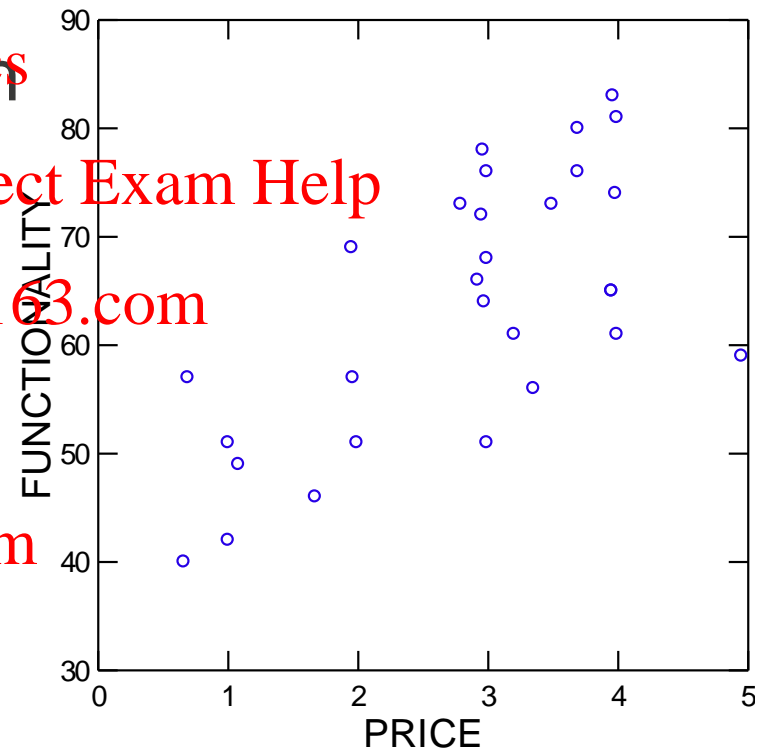


A. Price and function are both independent.

B. Function is independent.

C. Price is dependent.

D. Price is independent.



程序代写代做 CS编程辅导

SYSTAT Output (a) report

Dependent Variable

N

Multiple R

Squared Multiple R

Adjusted Squared Multiple

Standard Error of Estimate



QUALITY

Pearson's r

Coeff of determination: r^2

Regression Coefficients $B = (X'X)^{-1}X'Y$

WeChat: cstutorcs

Effect	Coefficient	Std. Error	Coefficient	Tolerance	t	p-Value
CONSTANT	44.025	4.567	0.000	.	9.640	0.000
PRICE	6.939	1.503	0.664	1.000	4.618	0.000

Assignment Project Exam Help

Email: tutorcs@163.com

Y-intercept: $b = 44.025$
QQ: 749389476

Gradient: $a = 6.939$
<https://tutorcs.com>

Least square regression line is: $Y = 6.639x + 44.025$

SYSTAT Output (a) report

程序代写代做 CS编程辅导

Analysis of Variance

Source	SS	df	Mean Squares	F-Ratio	p-Value
Regression	1,800.032	1	1,800.032	21.325	0.000
Residual	2,279.003	27	84.408		

WeChat: cstutorcs

Durbin-Watson D-Statistic | 0.946
First Order Autocorrelation | 0.482

Assignment Project Exam Help

Information Criteria

Email: tutorcs@163.com

AIC | 214.860
AIC (Corrected) | 215.820
Schwarz's BIC | 218.962

QQ: 749389476

<https://tutorcs.com>

程序代写代做 CS编程辅导

SYSTAT Output (a) report



- Tweaking the c changing data point 19 from (4.95, 59) to (4 results in a warning:

Analysis of Variance

WeChat: cstutorcs

Source	SS	df	Mean Squares	F-Ratio	p-Value
Regression	1,257.113	1	1,257.113	10.008	0.004
Residual	3,391.577	27	125.614		

Assignment Project Exam Help

Email: tutorcs@163.com

*** WARNING *** :

QQ: 749389476

Case 19 is an Outlier (Studentized Residual : -4.698)

<https://tutorcs.com>

- Other warnings are for 'leverage' and large residuals.

<https://flux.qa> (Feedback Code: 84660V)

Question 6

If regression equation is: Function = $44 + \text{Price} \times 7$,
a toothbrush has Price of \$3 would have a
Function of:



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

A. 44

B. 51

C. 54

D. 65

E. None of the above.

程序代写代做 CS编程辅导

The Challenger Disaster

- The Space Shuttle Challenger disaster occurred on January 28, 1986, when Space Shuttle Challenger broke apart 73 seconds into its flight, leading to the deaths of its seven crew members... (Text and images: Wikipedia);



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com


QQ: 749389476

<https://tutorcs.com>

程序代写代做 CS编程辅导

The Challenger Disaster

Data



Flight #	Date	Temp F	Temp C	# Failures
1	01-17-83	66	18.9	0
2	01-28-83	70	21.1	1
3	02-05-83	69	20.6	0
4	02-18-83	80	26.7	*
5	03-09-83	68	20.0	0
6	04-04-83	67	19.4	0
7	06-18-83	72	22.2	0
8	08-30-83	73	22.8	0
9	11-26-83	70	21.1	0
10	02-03-84	57	13.9	1
11	04-06-84	63	17.2	1
12	06-10-84	71	21.7	1
13	10-05-84	78	25.6	0
14	11-08-84	67	19.4	0
15	01-21-85	63	17.7	3
16	04-12-85	67	19.4	0
17	04-29-85	75	23.9	0
18	06-10-85	70	21.1	0
19	07-29-85	81	27.2	0
20	08-27-85	76	24.4	0
21	10-03-85	79	26.1	0
22	10-30-85	75	23.9	2
23	11-26-85	76	24.4	0
24	01-12-86	58	14.4	1

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Temperature on launch

31

-0.6

From: http://wps.aw.com/wps/media/objects/15/15719/projects/ch5_challenger/index.html

The Challenger Disaster

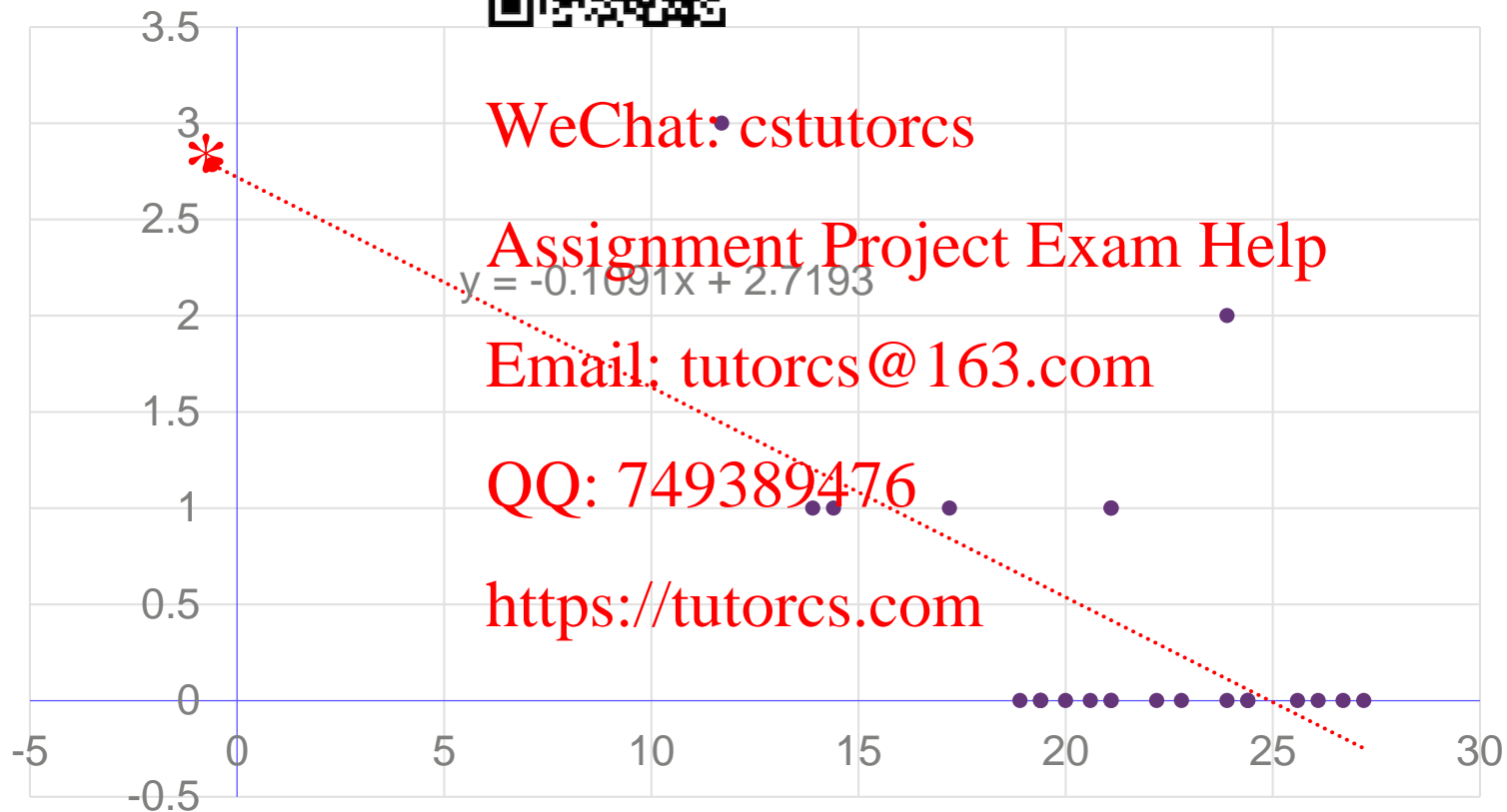
程序代写代做 CS编程辅导

Temperature at



*

failures



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Group activity

程序代写代做 CS编程辅导

$$Y_f = 0.43x_f + 138.56$$

$$Y_m = 0.425x_m + 150.47$$



- The Height v Weight of 102 elite male and 100 elite female athletes at the 2000 Sydney Olympic Games.
- See: FIT1006 Lecture 08 Worksheet.
- In groups work out the target height for a 100kg male and an 80kg female athlete.
- The most accurate prediction can be made for which gender? Why?
- Comment on the differences between genders of their weight profile.
- Data Source: <http://www.statsci.org/data/oz/ais.html>

WeChat: [estutorcs](#)

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

One More Thing

程序代写代做 CS编程辅导



- If we interchange x and y of our model we get a different regression equation, not just the inverse equation.

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

- Why?

程序代写代做 CS编程辅导

Necessary Skills



- Calculate the least squares regression by hand (using your calculator) for all data set.
- Interpret the basic SYSTAT output and comment on any data points that have a significant affect on the regression model.
- Draw a scatterplot and superimpose the line of best fit.
- Calculate Pearson's r , r^2 and Comment on the goodness of fit of the regression.

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutores@163.com

QQ: 749389476

<https://tutorcs.com>

程序代写代做 CS编程辅导

Reading/Questions



■ Reading:

- 7th Ed Section 15.4, 15.7, 16.1*, 16.2*.
- *Additional reading on multiple regression.

WeChat: cstutorcs

■ Questions:

- 7th Ed Questions 15.6, 15.7, 15.8, 15.10, 15.12, 15.14, 15.19, 15.21, 15.17, 15.63, 15.64.

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>