Due 11:59 PM CT 06/12/2022

Consider a well-known dataset on per capita income and per capita spending in public schools by state in the United States in 1979. (Available on blackboard). This dataset has been widely analyzed in various statistical. As in those previous analyses, we take per capita spending (Expenditure) as the dependent variable and per capita income as the predictor variable.

1. **Draw a scatter-plot to check the relationship between Income and Expenditure and interpret the relationship between Income and Expenditure.**

Chart, scatter chart

Description automatically generated

1. **Find and interpret the slope for the least squares regression line**

After running the following code;

lm(Expenditure ~ Income, data = spend)

we get;

Call:

lm(formula = Expenditure ~ Income, data = spend)

Coefficients:

(Intercept) Income

-151.26509 0.06894

So the slope for the least squares regression line is 0.06894

1. **Find and interpret y-intercept for the least squares regression line**

Off the results from part b, the y intercept for the least squares regression line is -151.265

1. **Find the least square regression equation and circle the results from your outputs.**

Y = 0.06894(X) – 151.265

Output

Call:

lm(formula = Expenditure ~ Income, data = spend)

Coefficients:

(Intercept) Income

-151.26509 0.06894

1. **Find proportion of the variation that can be explained by the least squares regression line (i.e., R2).**

Call:

lm(formula = Expenditure ~ Income, data = spend)

Residuals:

Min 1Q Median 3Q Max

-112.390 -42.146 -6.162 30.630 224.210

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -151.26509 64.12183 -2.359 0.0224 \*

Income 0.06894 0.00835 8.256 9.05e-11 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 61.41 on 48 degrees of freedom

Multiple R-squared: 0.5868, Adjusted R-squared: 0.5782

F-statistic: 68.16 on 1 and 48 DF, p-value: 9.055e-11

1. **Find the estimator of σ2 (i.e., s2) and interpret the value of this estimator.**

The value of sigma^2 = 3771.15, and this represents the variance

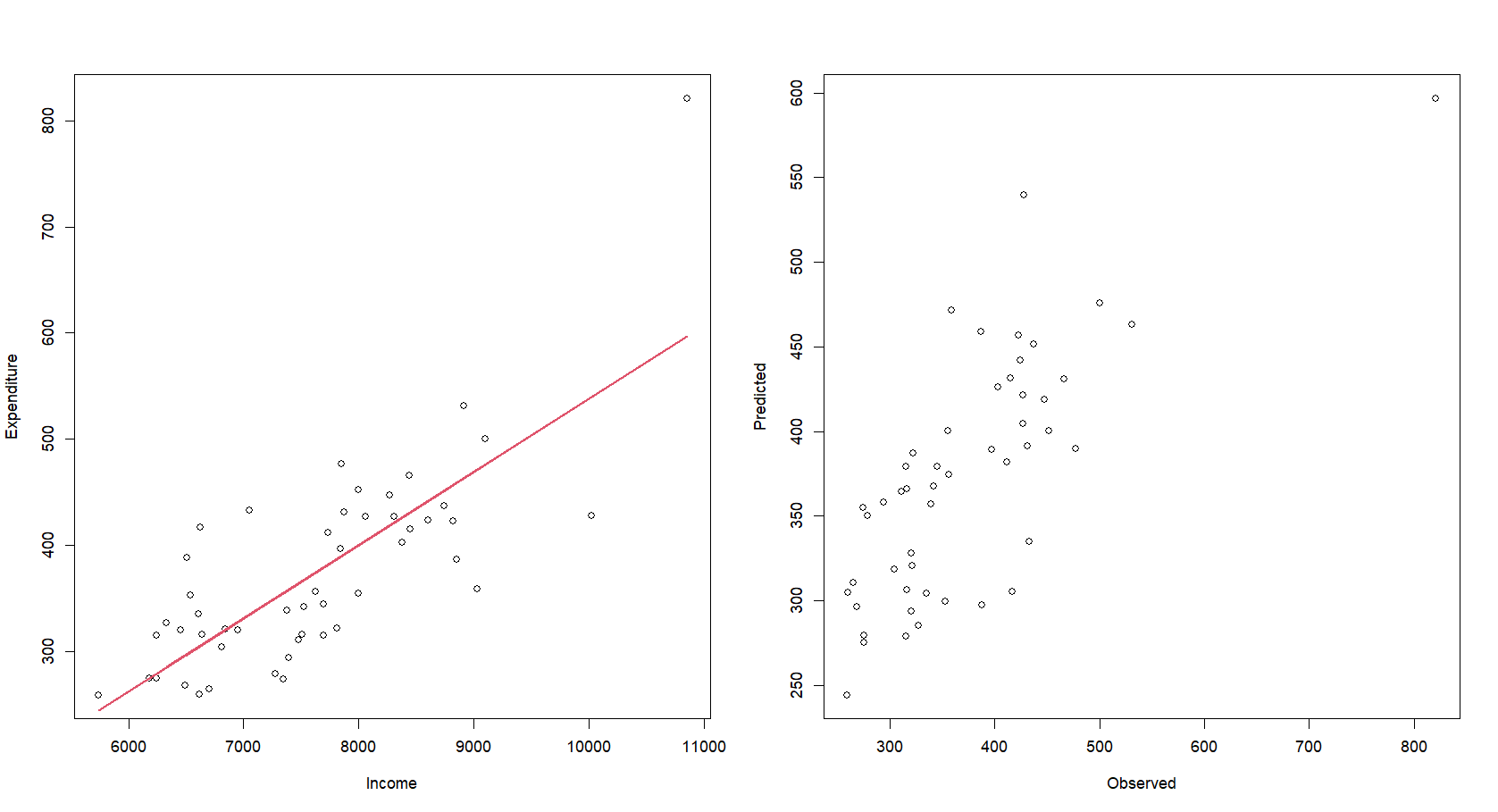
1. **Check if the data contain any outlier or influential points?**

Both the residual plot and the boxplot show one extreme outlier

Chart, box and whisker chart

Description automatically generated

1. **Fit a single linear model and conduct 10-fold CV to estimate the error. In addition, draw the scatter plot with the fitted line and the scatter plot between the observed and fitted values below.**



1. **Fit a quadratic model and conduct 10-fold CV to estimate the error and draw the scatter plot with the fitted line and the scatter plot between the observed and fitted values.**

Chart, scatter chart

Description automatically generated

I’m not sure why the fitted line is so squiggly, I’ve double checked the code and don’t see any problems

1. **Fit a mars model with optimal tuning parameters that you choose and conduct 10-fold CV to estimate the error and draw the scatter plot with the fitted line and the scatter plot between the observed and fitted values.**

Chart, scatter chart

Description automatically generated

1. **Compare the three fitted models in terms of RMSE and R2, and then make a recommendation based on your criteria.**

Graphical user interface, text, application

Description automatically generated