TotalIncome	TotalTax
12854	242
46361	8412
34921	5323
16198	2231
62964	13325
33567	5816
4349	168
18382	2753
23438	3202
30104	3798
31198	1479
58533	7926
21596	1102
29015	336
368	0
54033	7223
65627	13748
17424	2510
19948	3005
13359	0
35247	974
11540	1143
19669	3032
24804	107
27484	4051
26992	3282
24731	2239
24150	108
29360	4530
34815	4496
40755	8318
58657	12700
39440	5758
51313	6121
26110	3301
3243	206
25742	2202
44286	8484
1283	58
32964	3409
39229	4564
26426	2192
8962	743
11248	24
2912	24
36918	5024
41594	2419
44646	8023

45981	5070
14231	178
33055	4627
56289	10276
3387	0
28213	3539
45377	5813
41939	2474
43990	2614
37518	7877
14492	1822
37090	5795
20921	3280
46534	8528
53005	10251
7154	375
37604	2816

Here we have 65 observations. It summarizes the total income of 65 individus and their tax paiment. We want to see if there is a linear relationship between total income (*Totalincome*) and total tax paiments in order to predict the total tax paiment (*Totaltax*) of any individu who has a certain income. We use simple linear regression model to check if there is a linear relationship. The regression equation is expressed as the following:

$$Totaltax = \beta_0 + \beta_1 Totalinc + \epsilon$$

.

Given the following output, answer the questions coming after the output:

Summary of Fit

r^2	0.745
r_a	0.741
Root mean square	1776.81
$\overline{y}$	3930.24
observations	65

Parameter Estimates

	Etimates	Std Error	test statistics	P-value
Intercept	-1626.644	464.8847	-3.50	0.0009
Totalinc	0.184705	0.013606	13.58	0.0001

answer the following questions:

1. What are the estimation of the regression coefficient β_0 and β_1?
2. What is the regression line?
3. Are total Income and totaltax correlated? why
4. Can you use the outputs to perform the following tests: (1) $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$ and (2) $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ and answer the following questions: is $\beta_0 \neq 0$? is $\beta_1 \neq 0$ using the rejection point ($\alpha = 0.05$)
5. Can you use the outputs to perform the following tests: (1) $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$ and (2) $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ and answer the following questions: is $\beta_0 \neq 0$? is $\beta_1 \neq 0$ using the p-value?($\alpha = 0.05$)
6. Can you predict the TotalTax paiment of a person whose $Totalinc = \$115,000$?
7. Is the proposed regression model a good one and what is the coefficient that indicates this in the output.

Answer:

1. $$\widehat{\beta}_0 = -1626.644,\ \widehat{\beta}_1 = 0.184705$$
2. the regression line is given by:
$$Totaltax = -1626.644 + 0.1847 Totalinc$$
3. Yes TotalTax and Totalinc are strongly correlated because the correlation coefficient
$$r = \sqrt{r^2} = \sqrt{0.745} = 0.86313$$
so the totaltax and totlainc are positively correlated.
4. To perform (1) $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0,$ we need to calculate the test Statistics $test\_Statistics(\beta_0)$ and also $t_{(\alpha/2,n-1)}$. Using the t-table, $t_{(0.025,64)} = 2.00$ and using the output, we have
$$test\_Statistics(\beta_0) = -3.50$$
we reject H_0 if $test\_Statistics > t_{(\alpha/2,n-1)}$ *or* $test\_Statistics < -t_{(\alpha/2,n-1)}$. which is true in this case because $-3.50 < -2.00$.
We do the same thing for $\beta_1,$ in fact: To perform (2) $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0,$ we need to calculate the test Statistics $test\_Statistics(\beta_1)$ and also $t_{(\alpha/2,n-1)}$. Using the t-table, $t_{(0.025,64)} = 2.00$ and using the output, we have
$$test\_Statistics(\beta_1) = 13.58$$
we reject H_0 if $test\_Statistics > t_{(\alpha/2,n-1)}$ *or* $test\_Statistics < -t_{(\alpha/2,n-1)}$. which is true in this case because $13.58 > 2.00$
5. Using the P-value, it is straighforward:
for (1): $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0,$ the p-value=0.0009<0.05, so we reject the null hypothesis
for (2):$H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0,$ the p-value =0.0001<0.05, so we reject the null hypothesis
6.

$$\begin{aligned} Totaltax &= -1626.644 + 0.1847 Totalinc \\ &= -1626.644 + 0.1847(115,000) \\ &= \$19,614 \end{aligned}$$

7. The coefficient that determines if the model is a good model is the coefficient of determination r^2 which in this case equal to 0.745. It is very high so the proposed model of regression is a good model.