Statistics & Spreadsheets Advanced Answer

Stanley Austin Susanto

1. Correlation

Question

1. To understand the data in general, what variable(s) has a strong relationship to Adibi value of orders?

Answer

We will use correlation to see which variable(s) has a strong relationship to Adibi value of orders or amount paid on Adibi. The greater value of r (greater positive or greater negative), the stronger relationship.

In this case, we use **independent variables** such as **income, total number of order**, **Balena and Celinna quantity of order**, **amount spent on Balena and Celinna**, **promotion on Adibi, Balena, and Celinna**. **Dependent variable** we used here is **amount paid on Adibi**.

(Cont'd Answer 1)
It turns out that the **strongest relationship** to **Adibi value of order** or **amount spent on Adibi** is **amount spent on Celinna**. It means that when the amount spent on Celinna is increasing, then the amount spent on Adibi is increasing too.
However, the value of correlation of this relationship is close to 1 which means that **multicollinearity does appear**.

	Income	Total Number of Order	Balena Quantity of Order	Celinna Quantity of Order	Amount Spent on Balena	*Amount Spent on Celinna*	Promotion Adibi (%)	Promotion Balena (%)	Promotion Celinna (%)	Amount Spent on Adibi
Income	1									
Total Number of Order	-0,03648	1								
Balena Quantity of Order	0,050133	-0,04196	1							
Celinna Quantity of Order	-0,06523	-0,0166	-0,04321	1						
Amount Spent on Balena	0,028595	0,053938	0,006258	0,045752	1					
Amount Spent on Celinna	0,026359	0,068693	-0,00717	0,035907	0,982681	1				
Promotion Adibi (%)	0,012663	0,002841	0,046469	-0,01324	-0,10931	-0,10966	1			
Promotion Balena (%)	-0,0597	-0,02493	-0,01273	-0,05736	-0,03727	-0,05135	-0,02537718	1		
Promotion Celinna (%)	-0,01332	-0,08856	0,017175	0,014318	-0,01337	-0,02479	0,010270005	-0,0265482	1	
Amount Spent on Adibi	0,027238	0,068734	-0,00652	0,03701	0,982512	**0,999818**	-0,10959324	-0,05069863	-0,02542024	1

2. Linear Regression

Question

2. What variable has the biggest impact on the Adibi value of orders?

Answer

We use **dependent variable** of **amount paid on Adibi** (Adibi value of orders) while independent variables are **income**, **total number of order**, **Balena quantity of order**, **Celinna quantity of order**, **amount paid on Balena**, **amount paid on Celinna**, **promotion on Adibi**, **promotion on Balena**, and **promotion on Celinna**.

The **biggest impact** on the **Adibi value of orders** must have the **greatest positive or negative** value of **r squared** when we make a regression line using dependent line and various independent variables.

Using significance level of 5%, we will **eliminate variable(s)** that has **p-value larger than 5%** ($p \geq 0.05$).

(Cont'd Answer 2)
It turns out that only ***p*-value** of **Amount Spent on Celinna** has a value less than 0.05. Thus, we will count on this variable only.

ANOVA

	df	SS	MS	F	Significance F
Regression	9	684459,4331	76051,05	150509,2	0
Residual	487	246,0770322	0,505292		
Total	496	684705,5102			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%
Intercept	**-0,586217243**	0,257365754	-2,27776	**0,023173**	-1,0919	-0,08053	-1,0919	-0,08053
Income	**8,78973E-07**	7,7896E-07	1,128393	**0,25971**	-6,5E-07	2,41E-06	-6,5E-07	2,41E-06
Total Number of Order	**0,002551214**	0,022482892	0,113474	**0,909702**	-0,04162	0,046727	-0,04162	0,046727
Balena Quantity of Order	**0,009085091**	0,011512059	0,78918	**0,430391**	-0,01353	0,031705	-0,01353	0,031705
Celinna Quantity of Order	**0,016430552**	0,011243913	1,461284	**0,144582**	-0,00566	0,038523	-0,00566	0,038523
Amount Spent on Balena	**-0,000396758**	0,003596139	-0,11033	**0,912194**	-0,00746	0,006669	-0,00746	0,006669
Amount Spent on Celinna	**0,695281843**	0,003262595	213,107	**0**	0,688871	0,701692	0,688871	0,701692
Promotion Adibi (%)	**0,001783038**	0,039696329	0,044917	**0,964192**	-0,07621	0,07978	-0,07621	0,07978
Promotion Balena (%)	**0,006129147**	0,006833569	0,896917	**0,370206**	-0,0073	0,019556	-0,0073	0,019556
Promotion Celinna (%)	**-0,007993557**	0,011186706	-0,71456	**0,475224**	-0,02997	0,013987	-0,02997	0,013987

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,999820288
R Square	0,999640609
Adjusted R Square	0,999633967
Standard Error	0,710838693
Observations	497

(Cont'd Answer 2)
When we have only variable of **Amount Paid on Celinna**, we will have this output. It turns out that the value of *r* squared is 0.9996 which suggests that **the biggest impact of Adibi value of order** is **amount paid on Celinna!**

Using this output, we will have a regression line of
y = 0.695x - 0.300
where **x**, **y** be **the amount paid on Celinna and Adibi** in thousand US Dollars.
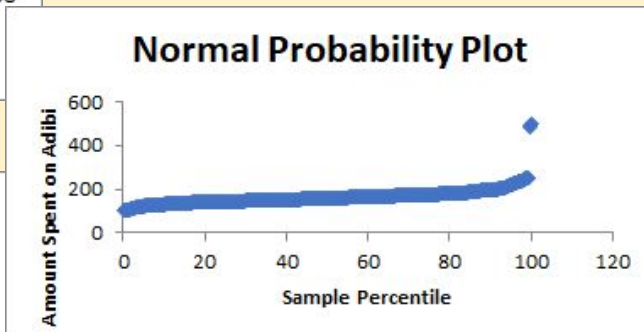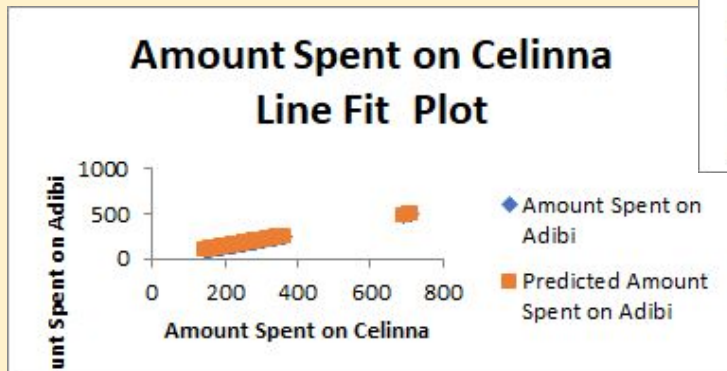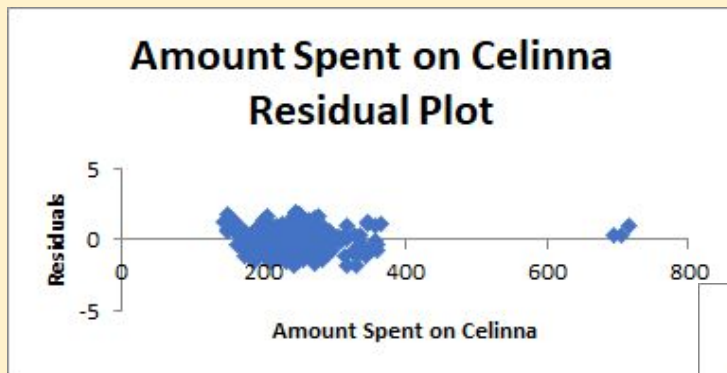
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,999818
R Square	0,999637
Adjusted R Square	**0,999636**
Standard Error	0,708598
Observations	497

ANOVA

	df	SS	MS	F	ignificance F
Regression	1	684457	684457	1363157	0
Residual	495	248,5453	0,502112		
Total	496	684705,5			

	Coefficients	andard Err	t Stat	P-value	Lower 95%	Upper 95%	ower 95,0%	pper 95,0%
Intercept	**-0,29933**	0,144148	-2,07655	0,038359	-0,58255	-0,01611	-0,58255	-0,01611
Amount Spent on Celinna	**0,694959**	0,000595	1167,543	0	0,69379	0,696129	0,69379	0,696129

Answer

(Cont'd Answer 2) Based on residual plot, we have seemingly scattered plot, which means the graphic is **homoscedasticity**.

Since *r* squared to 1, it is obvious that its line fit plot fits predicted amount spent on Adibi.

Its normal probability plot shows that **the residual** is **normally distributed**.

Amount Spent on Celinna Residual Plot

Normal Probability Plot

Amount Spent on Celinna Line Fit Plot

2. Linear Regression

Question

3. Mr. Chu needs to manage his expenses, mainly because he has spent a lot of money on Adibi. Hence, he asked Luxura team to predict his Adibi value of orders based on following characteristics:

Answer

We use a regression line of

$$y = 0.695x - 0.300$$

where **x, y** be **the amount paid on Celinna and Adibi in thousand US Dollars** based on the following data made by Mr. Chu.

Income	Total Number of Order	Balena Quantity of Order	Celinna Quantity of Order	Balena Values of Order	Celinna Values of Order	Promotion Adibi (%)	Promotion Balena (%)	Promotion Celinna (%)
1,820,312	17	3	2	49	67	2%	3%	1%

2. Linear Regression

Question

3. Mr. Chu needs to manage his expenses, mainly because he has spent a lot of money on Adibi. Hence, he asked Luxura team to predict his Adibi value of orders based on following characteristics:

Answer

(Cont'd Answer 3)
For Celinna values of order/**the amount paid** on **Celinna** of **67** thousand US Dollars ($x = 67$), we will have **the amount paid on Adibi**, that is

$$y = 0.695(67) - 0.300$$
$$\boldsymbol{y = 46.265}$$

or we can say that **Mr. Chu** has **paid 46.265 thousand US Dollars on Adibi**.

2. Linear Regression

Question

4. Based on point 1-3 what recommendation(s) can you provide? Please consider providing recommendations for each point.

Answer

We have recommendations for each point.
- Point 1: focusing on customers with high or very high income to buy Adibi and Celinna products by giving them suitable and recommended respective products.
- Point 2: focusing on making ads about Adibi and Celinna products in order to buy them.
- Point 3: Mr. Chu needs to buy Adibi and Celinna products in order to have increase on amount paid on Adibi. Giving him promotions and recommended products is good idea.