Statistics 2 coursework

Aim

For this report I will simply be studying to see if advertising affects humans, I have collected data to see if video games advertising on TV actually has an effect on people to purchase more video games.

The data I have collected is the amount of adverts shown on TV per month about video games and the amount of games sold per month, since the amount of games sold in the entire UK or London will be very hard to find out, I will be finding out the amount of games sold in Streath am (a fairly average town in London).

Finding a correlation will prove and justify, companies spending millions on advertising there products on TV adverts and show it has an affect. The final result may convince people that TV adverts actually do have an effect on people, showing TV adverts can influence people to buy products, this very effect in my opinion makes this exercise worth doing. Showing the data in statistical form will hopefully give a more clearer picture if a link is present.

Data Collection

I have collected 58 pairs of data from various internet search engines these include Google, AltaVista, Lycos and government statistic sites like www.statistics.gov.uk. I decided to use the internet as it is a great source of information especially for statistical data and it makes it easy to get large sets of data easily and quickly.

To ensure that the data was of good quality I did not just collect the data from one source, I gathered various statistics about the data from multiple sources to hopefully get a better spread of data. In total I managed to collect 58 sets of data, I think it is appropriate to include all the data since it is from various sources making it a good set of data.

Amount of video games sold per month in Streatham (X)	adverts on TV about Video games per month(Y)
1432	343
470	244
941	322
980	335
1750	453
1683	512
512	163
488	286
1006	193
437	183
683	261
611	280
556	129
2429	505

2400	368
660	160
1230	182
1634	257
379	136
401	97
1548	413
721	105
574	84
792	105
1065	136
1448	145
1410	169
2005	214
648	77
2081	500
971	109
944	111
2013	420
889	87
1559	260
2230	430
2515	460
2800	537
280	195
401	218
663	323
1916	256
1829	320
662	231
875	290
1131	297
1144	381
1446	273
2099	305
2562	402
2177	382
1742	357
973	265
791	223
1268	256
1384	223
889	240
992	283

Modelling Procedures

Advertising vs Sales figures
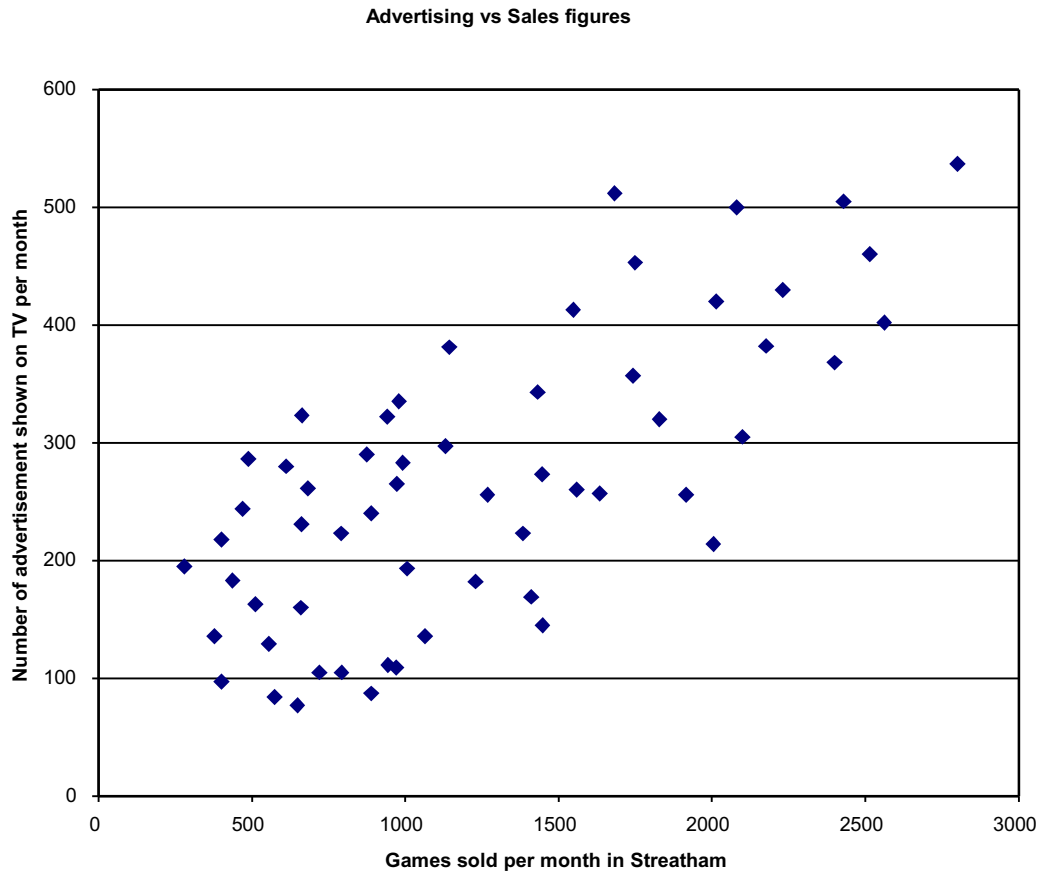
By applying the method mentioned above I was able to collect 58 pairs of data, which are both independent variables as well as random. From the scatter diagram above you can see there is roughly a positive correlation between the points, forming an elliptical shape. It i s essential to draw a line of best fit in order to formulate a rough prediction of the values on the y -axis by using values from the x-axis. Next in order to ensure that this line is accurate I will draw a regression line by adopting appropriate calculatio ns.

I will carry out more statistical analysis to see if there exists a relationship between the two variables. The mean of the two variables and a correlation coefficient will be worked out. To present a strong argument I will try to retest if the implication provided by the correlation coefficient is true by using a hypothesis test.

Amount of video games sold per month in Streatham (X)	Adverts on TV about Video games per month (Y)	X^2	Y^2	XY	
1432	343	2050624	117649	491176	
470	244	220900	59536	114680	
941	322	885481	103684	303002	
980	335	960400	112225	328300	
1750	453	3062500	205209	792750	
1683	512	2832489	262144	861696	
512	163	262144	26569	83456	
488	286	238144	81796	139568	
1006	193	1012036	37249	194158	
437	183	190969	33489	79971	
683	261	466489	68121	178263	
611	280	373321	78400	171080	
556	129	309136	16641	71724	
2429	505	5900041	255025	1226645	
2400	368	5760000	135424	883200	
660	160	435600	25600	105600	
1230	182	1512900	33124	223860	
1634	257	2669956	66049	419938	
379	136	143641	18496	51544	
401	97	160801	9409	38897	
1548	413	2396304	170569	639324	
721	105	519841	11025	75705	
574	84	329476	7056	48216	
792	105	627264	11025	83160	
1065	136	1134225	18496	144840	
1448	145	2096704	21025	209960	
1410	169	1988100	28561	238290	
2005	214	4020025	45796	429070	
648	77	419904	5929	49896	
2081	500	4330561	250000	1040500	
971	109	942841	11881	105839	
944	111	891136	12321	104784	
2013	420	4052169	176400	845460	
889	87	790321	7569	77343	
1559	260	2430481	67600	405340	
2230	430	4972900	184900	958900	
2515	460	6325225	211600	1156900	
2800	537	7840000	288369	1503600	
280	195	78400	38025	54600	
401	218	160801	47524	87418	
663	323	439569	104329	214149	
1916	256	3671056	65536	490496	
1829	320	3345241	102400	585280	
662	231	438244	53361	152922	
875	290	765625	84100	253750	
1131	297	1279161	88209	335907	
1144	381	1308736	145161	435864	

1446	273	2090916	74529	394758	
2099	305	4405801	93025	640195	
2562	402	6563844	161604	1029924	
2177	382	4739329	145924	831614	
1742	357	3034564	127449	621894	
973	265	946729	70225	257845	
791	223	625681	49729	176393	
1268	256	1607824	65536	324608	
1384	223	1915456	49729	308632	
889	240	790321	57600	213360	
992	283	984064	80089	280736	
Σx	Σy	Σx^2	Σy^2	Σxy	
72119	15491	114746411	4980045	22566980	
					Correlation
Mean of x	Mean of y	Sx	Sy	Covariance(S_{xy})	Coefficient(r)
1243.431	267.086	657.469	120.531	63808.651	0.805
		S_x^2			
		432265.659			

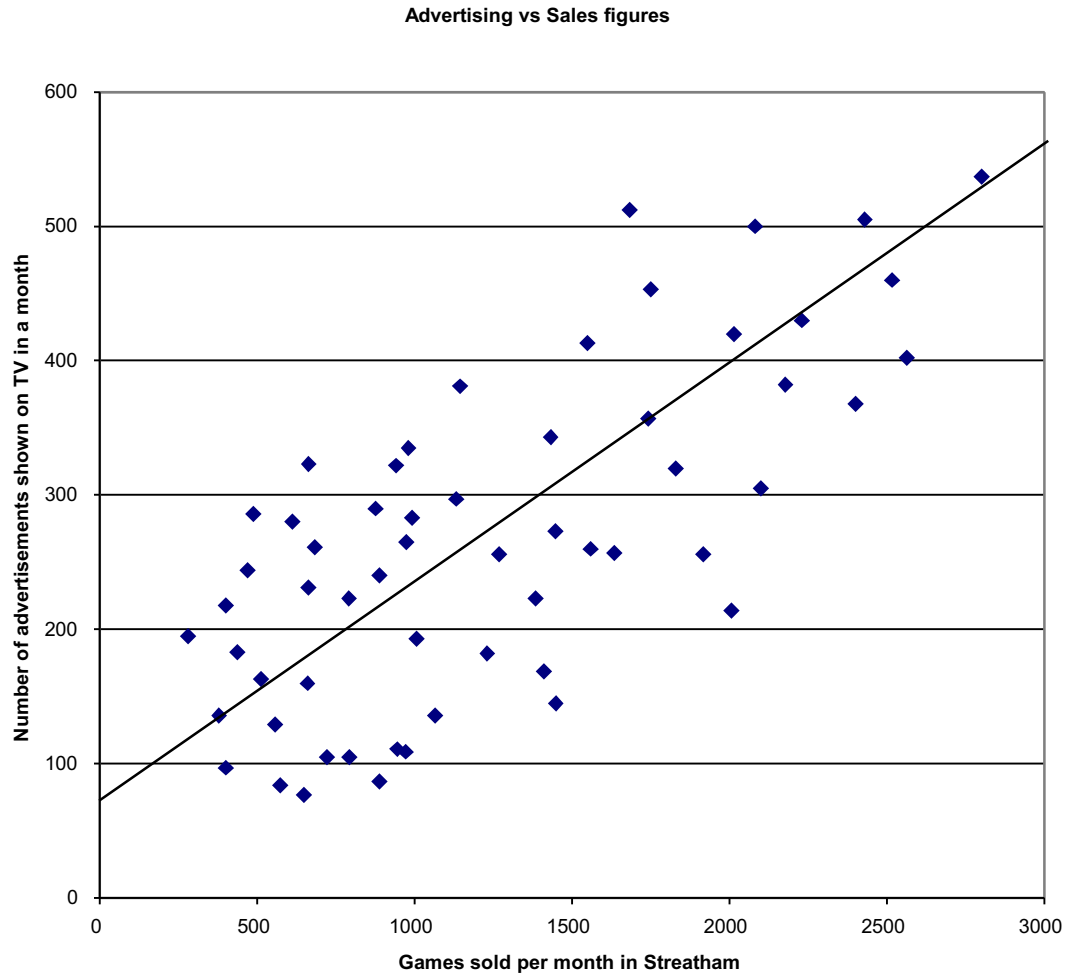## Analysis

Regression line

$$y - \bar{y} = \frac{S_{xy}}{S_x^2} (x - \bar{x})$$

$$y - 267.086 = \frac{63808.651}{432265.659} (x - 1243.431)$$

$$y - 267.086 = 0.148 (x - 1243.431)$$

$$y - 267.086 = 0.148x - 184.028$$

Y = 0.148X + 83.058

Advertising vs Sales figures

Hypothesis test

H_o : $\rho = 0$ There is no correlation between the two variables.

H_1 : $\rho > 0$ There is a positive correlation between the number of adverts about video games shown on TV and the amount of video games sold.

Significance level is **5%**.

The critical value for **n = 58** at a **5%** significance level from the product moment correlation coefficient (r) table is **0.2181**.

The correlation coefficient is calculated to be **0.805**.

$$0.805 > 0.2181$$
$$H_1 > H_o$$

Carrying out a hypothesis test using a 5% significant level reveals that there is a positive correlation between the numbers of TV adverts shown and video games sold.

Interpretation

By looking at the scatter diagrams and calculations it is safe to say that the variables being examined are roughly positively correlated. The calculations I carried out avoid any uncertainties about the relationship between the two variables by revealing that the two variables are directly proportional.

Finally I can conclude that you can see that the correlation coefficient, r, is pretty high and close to +1 so it is safe to say that there is a positive liner correlation between the amount of video game adverts shown and the amount of games sold. Again my hypothesis H_1 was proven to be correct which further strengthens the argument which was set in my aim showing that advertising does have and effect.

This exercise was worth doing because I find that indirectly influencing people or convincing people to do things which they may not normally want to, like buying certain products, is not really good thing, I hope by providing this statistical data that people can see how advertising influences them, and start to view things they see with a more head strong opinion and not get pushed around.

Accuracy and refinements

Restrictions and Possible sources of error

Due to the topic that was processed in this exercise being very large, I think the data I collected only represented a small portion of the whole thing. Another restriction was that I would have liked to obtain the data my self from individual high street stores, etc. However since there are too many these are not entirely possible, so I had to use other means of obtaining this data like the internet. Other possible errors could be to do with the data obtained and the date which it was recorded was not noted. So it is hard to say which time period it fits into, as the adverts and video game sales may be different at different seasons of the year.

Improvements

This exercise in my opinion can be carried out in much greater accuracy and detail if the time period it was recorded in was much larger. Another improvement could have been to contact the TV companies to get a timetable of what type of adverts where shown at what time, so a much more accurate overview can be obtained. Then last of all this exercise processed video game sales in general it would be interesting to have done work on an individual product, this in my opinion would have been much more accurate and convincing.