# Bivariate coursework is to discover whether there

The aim of this coursework is to discover whether there is a correlation between the heights of people and there shoe size. Is it true that if you are taller, your shoe size is bigger? In this investigation, I deem to find that out. This information would then be used to inform specialist clothes shows that tailor for tall men, which shoe sizes to stock. Should a shop that sells clothes for men over 6ft stock shoes which range from size 6 up to size 15? Or would it be more appropriate for them to just stock from sizes 10 -15? Data collectionThe population is adults who shop in Shirley, Birmingham. When I was working in a supermarket, I asked the first 50 people from 6pm onwards if they would fill in my table.

The sampling method I have used is definitely not random however; it is suitable for this investigation. The conclusion given at the end of this investigation will acknowledge that I have not used a random sampling method.(Table shown below)Height (cmShoe Size (Nearest Integer)All 50 people co-operated and filled in the table as requested. There are weaknesses with this form of data collection. However, asking one person, does not affect the second persons answer.

Also, I am relying on the customer’s results to be accurate. There was roughly an even number of men and women filling in my chart. Another point to mention is, the majority of the data was collected from adults; there were no children in my population and no elderly people, all by coincidence. The data will tell me if there is a correlation between height and shoe size.

As I mentioned before, the data is not 100% accurate because it isn’t random however, I am relying on the information collected to be reliable enough for the investigation. Height cmShoe Size (integer)136613871397141814281438144714571479149914981519152101 52915391539154101551015691571015810159111591016010161101619161 11116210162101631116311164101651116510166111671216710168111170 101711117111172111741117710177111791118412184121851218813Total 8019492Mean160. 389. 84(Sorted Data)For this investigation, I have decided to see whether there is a correlation between height and shoe size. To work out the correlation coefficient I will use Pearson’s Product Moment correlation coefficient. Calculating PMCC is an appropriate technique since both my X and Y variables are random, the data shown on the scatter graph is roughly elliptical and also there appears to be a linear correlation on the scatter graph.

If the value of ‘ r’ turns out to be close to 1, then I can deduce that there is a strong correlation. However, that is not the end of the story, after that, I will carry out a hypothesis test as a further test statistic. The working for Product Moment correlation coefficient can be seen below. x – x meany – y meanx – x mean sqy – y mean sq(x – x mean)(y – y mean)-24. 38-3.

84594. 384414. 745693. 6192-22. 38-2. 84500.

86448. 065663. 5592-21. 38-2. 84457. 10448.

065660. 7192-19. 38-1. 84375. 58443.

385635. 6592-18. 38-1. 84337. 82443.

385633. 8192-17. 38-1. 84302. 06443.

385631. 9792-16. 38-2. 84268. 30448.

065646. 5192-15. 38-2. 84236.

54448. 065643. 6792-13. 38-0. 84179. 02440.

705611. 2392-11. 38-0. 84129. 50440.

70569. 5592-11. 38-1. 84129. 50443.

385620. 9392-9. 38-0. 8487.

98440. 70567. 8792-8. 380.

1670. 22440. 0256-1. 3408-8.

38-0. 8470. 22440. 70567. 0392-7.

38-0. 8454. 46440. 70566. 1992-7.

38-0. 8454. 46440. 70566. 1992-6.

380. 1640. 70440. 0256-1. 0208-5. 380.

1628. 94440. 0256-0. 8608-4. 38-0.

8419. 18440. 70563. 6792-3. 380. 1611.

42440. 0256-0. 5408-2. 380.

165. 66440. 0256-0. 3808-1. 381. 161.

90441. 3456-1. 6008-1. 380.

161. 90440. 0256-0. 2208-0. 380.

160. 14440. 0256-0. 06080. 620. 160.

38440. 02560. 09920. 62-0. 840.

38440. 7056-0. 52080. 621. 160. 38441.

34560. 71921. 620. 162. 62440.

02560. 25921. 620. 162. 62440. 02560.

25922. 621. 166. 86441. 34563. 03922.

621. 166. 86441. 34563. 03923. 620.

1613. 10440. 02560. 57924. 621. 1621.

34441. 34565. 35924. 620. 1621. 34440.

02560. 73925. 621. 1631. 58441. 34566.

51926. 622. 1643. 82444. 665614. 29926.

620. 1643. 82440. 02561.

05927. 621. 1658. 06441. 34568.

83929. 620. 1692. 54440. 02561.

539210. 621. 16112. 78441.

345612. 319210. 621. 16112. 78441.

345612. 319211. 621. 16135. 02441. 345613.

479213. 621. 16185. 50441. 345615. 799216.

620. 16276. 22440. 02562.

659216. 621. 16276. 22441. 345619. 279218.

621. 16346. 70441. 345621.

599223. 622. 16557. 90444.

665651. 019223. 622. 16557.

90444. 665651. 019224. 622. 16606. 14444.

665653. 179227. 623. 16762. 86449.

985687. 27922. 27374E-137. 10543E-158233. 78112. 72862.

04n = 50 (50 pairs of data)Sx = 12. 83259911Sy = 1. 50146595Sxy = 17. 2408r = 0. 894802727There seems to be a positive correlation for my sample because r is very close to 1 (r = 0.

894802727). A hypothesis test will determine if there's enough evidence to suggest there is actually a positive correlation for my population. I will now carry out a 1 tailed hypothesis test. H0 : p = 0H1 : p > 0r = 1 r = -1'r' is often called the correlation coefficient, the quantity, r, provides a standardised measure of correlation. Its value always lies between -1 and 1, the value 1 means a complete positive correlation and -1 means a complete negative correlation as shown on the graphs above.

So far in this investigation, I have calculated the correlation coefficient for my data and it is very strong, 0. 89 (2dp) but I am now going to assume there is no correlation between the two variables. H0 : p = 0H1 : p > 0I am going to use the ‘ standard’ 5% significance level to carry out my hypothesis test. The critical value for n = 50 at the 5% significance level for a 1-tail test is found from the product moment correlation coefficient tables to be 0. 2353.

Since 0. 8948 > 0. 2353, the critical value, the null hypothesis is accepted. The evidence from this small sample suggests that if you are taller, your shoe size is bigger however it is important not to rely on this data too much. This conclusion is only fairly accurate for adults.

It is important to bear in mind that the conclusion could well be different for growing teenagers or younger children. The sampling technique could be improved by covering different ages of people. To get a more realistic perception of whether height affects shoe size then the sampling technique has to be much more random. An example of a more randomised test could be a proportional stratified sample, including the elderly, young, and middle aged people. Stratified sampling will usually lead to more accurate results about the entire population, and will also give useful information about the individual strata.