Scatter plot and linear regression worksheet

Still not sure about linear regressions and bivariant data? Well, don't be afraid! We explain everything you need to know for the linear regressions of ace. In this article, we will talk: Year 12 Advanced Mathematics: Bivariate Data AnalysisProbability and Statistics is widely used in mathematics, statistics, finance, science, artificial intelligence and many more areas of study. Surprisingly, the topic that sustains difficult probability problems is an area of mathematics called combinatorial (i.m. counting!) In fact, since the odds are the ratios of the number of ways in which an event can happen to the total number of possibilities, the step is to develop a theory of the count. Let's look at the following NESA curriculum points. Results of the NESA curriculumThe following results of the syllabus will be addressed in this Thematic Guide: Build a bivariant dispersion diagram to identify patterns in the data that suggest the presence of an associationUse bivariant dispersion diagrams (building them when necessary), to describe the patterns, characteristics and associations of bivariant data sets, justifying the conclusionsDewriting bivariant data sets in terms of form (linear/nonlinear) and in the linear case, also the direction (positive/negative) and the strength of the association (strong/moderate/weak)Identify the dependents and independents variables within the bivariant data sets if applicableCalculate and interpret Pearson's correlation coefficient (r) using technology to quantify the strength of a linear association of a model sample by adjusting an appropriate line of best fit to a scatter and use it to describe and quantify associationsInterpret and gradient of the adjusted lineUsing the right line of best fit , both by eye and by applying the equation of the adjusted line, make predictions by interpolation or extrapolationDistinguish between interpolation and extrapolation, recognizing the limitations of using the adjusted line to make predictions, and interpolating from data plotted to make predictions if necessary Solve problems that involve identifying, analyzing and describing associations between two numerical variablesAssumed KnowledgeStudents should only be familiar with arithmetic for the Students must already be familiar with the basics of probability. They must also understand basic algebraic techniques and expansion to understand the concepts explored in the following guide. Bivariar Data AnalysisBivariate Data are data collected in pairs $(x, y)$ being the results of some experiment or observation. Here, the variable is called a separate variable $x$ and the variable $y$ is called a dependent variable. for example, $x$ may be the amount of money invested in television advertising for a particular business and $y$ may be the corresponding benefit that the businesses obtained for this $x$. For example, a data point would mean $2500 was invested in advertising, and $7000 was made in profit. In the analysis of bivariate data, we aim to see if there is a relationship or association between the variable $x$ and $y$. This relationship will generally be only a trend, and not a deterministic relationship, because the variable $y$, as the benefit in the example above, may depend on other factors not captured by just $x$. Now, we will introduce some terminology and tools used to describe bivariate data. Scatter Scattering data Can be viewed using a scatter chart, where the independent variable is placed on the horizontal axis, and the dependent variable is placed on the vertical axis. LinearityThe data set is linear if the data has a more or less linear tendency. For example, the following charts indicate a linear data set: A data set is not linear if the data does not follow a line shape. Instead, you couldn't have any trends at all, or maybe a curved/more complicated trend. For example: Correlation Correlation measures the strength and direction of this linear relationship. We will see each of these below:Strength When there is a linear trend, the strength of the association can be three categories – strong correlation, average correlation and weak correlation. When there is no linear tendency, we say there is no correlation. The strength of the correlation has to do with the alignment of data points. Strong correlation: The following plots have a strong correlation. Average correlation: The following plots show an average correlation. Weak correlation: The following plots have a weak correlation. Without correlation: we can not have correlation, that is, data points show no signs of linearity. The correction of the direction can be positive or negative. Positive correlation: The following plots show positive correlation – as $x$ increases, $y$ increases. Check your linear regression skills! PearsonSo's correlation coefficient far away, we have only described correlation qualitatively. However, it is difficult to compare two plots that both could have a positive correlation of average strength. Pearson's correlation coefficient is a quantitative measure for correlation. It is denoted by the letter $r$ (so it is also called Pearson's $r$) and ranges between -1 and 1.La magnitude of Pearson's coefficient refers to the strength of the linear relationship. The sign of Pearson's coefficient describes the direction of the linear relationship. The following plots are noted with how to interpret Pearson's correlation. Best fitAs line we've seen, while our data on a scatter chart may not be exactly linear, there may be a linear trend. We can describe relationship qualitatively or using Pearson's Coefficient.Now, we will actually approach the data using a better fit line (also known as the regression line). A better fit line aims to better represent the data by hand with a straight line. When a better fit line is built by a computer or calculator, the drawn line will have as short vertical distance from data points as possible. There are several reasons to use a better fit line:Summarize data points in an approximate relationship between $x$ and $y$We can interpret the characteristics of this line (such as intercept and gradient)We can use the line to make predictions (via interpolation and extrapolation)Students should be able to draw an approximate line of better adjustment. For example, the best fit line is shown through the scattering plot below: Using the calculatorThe curriculum requires students to use technology to find Pearson's correlation coefficient $r$The equation of the best fit line $y = A + Bx$when data is provided. In exams, they can use their calculators to do so. Steps:Clear the calculator from any previous data entry (SHIFT $\rightarrow$ 9 $\rightarrow$ 3)Select the linear regression to the calculator from STAT mode (MODE $\rightarrow$ 2 $\rightarrow$ 2)Enter your data and press AC when you have just retrieved the correlation coefficient or the A and B parameters of your line, press SHIFT $\rightarrow$ 1 $\rightarrow$ 5 and , then select 1, 2 or 3 depending on what you're after. Interpolation and Extrapolation Once drawn, calculated or provided by the question, it can be used to predict the value of $y$ to a hypothetical value of $x$ . If we have drawn the best fit line by hand, we can approximate the value $y$ to the hypothesis value of $x$If a better-adjusted line equation is provided, we can replace the hypothesis value of $x$ in the equation to approximate the value of $y$. There are two types of prediction: Interpolation: Predict $y$ for a value of $x$ within the range of our dataExtrapolation: Predict $y$ by a value of $x$ outside the range of our dataAieu, the range of our data refers to the values between the first point and the last point as you go from left to right. Note: Extrapolation is prone to error because it assumes that the trend in existing data will continue for $x$ values outside of the available data. Unless the data analyst is confident that the linearity of the trend will continue rather than, for example, curve, extrapolation can produce meaningless data due to strong assumptions. Concept Check QuestionsThe CEO of a company wanted to see the relationship between the number of years for which each employee has been working and their annual salary. The are shown in the following dispersion: 1) Identify any atypical.2) Estimate the correlation coefficient.3) What is the equation of less square regression line?4) What is the meaning of the interceptor and gradient of this model?5) Matthew has been working at the company for 10 years. Predict your salary. Concept Check Solutions1) Point (4, 140) is an outlier, potentially an exceptional employee who is highly compensated for his work despite having less experience than others.2) 0.9 (The data is quite linear and positively correlated)3) $y = 5.25x + 58$4) The intercept approximates the average salary of an in-experience employee being around $58,000, while gradient suggests a $5250 increase in annual salary from a year's worth of experience.5) $110,000See your brands increase in a strong and positive trendLearn Maths Adv from home with Matrix + Online Course! We will provide you with video theory lessons, access to our complete math books sent to your door and question discussion tables to help guide your doubts. Learn more about Matrix + Online Maths Adv Course now.  Content.