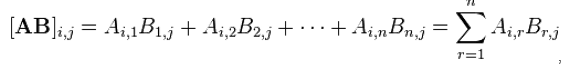
1. A is a 2 x 3 matrix and B is a 3 x 2 matrix. Specify a single summation formula to determine the matrix product AB and calculate the matrix product AB using the formula

Solution for first one, plus 2 practice ones ->



Try to solve for the other two samples

1. Cricket jumps in inches for a sample of 13 crickets were recorded.

7.92, 11.13, 9.76, 8.23, 14.51, 12.21, 12.43, 5.84, 12.11, 5.12, 4.96, 7.21, 8.32

i) Compute the range of the 13 cricket jumps. Give the units of measurement of the range.  *[2]*

ii) Compute the variance of the 13 cricket jumps. Give the units of measurement of the variance.  *[6]*

iii) Compute the standard deviation of the 13 cricket jumps. Give the units of measurement of the standard deviation. *[2]*

iv) Do you detect any outliers in the data? Explain with the aid of diagrams using both the box chart and z-score methods. Show all calculations. *[10]*

(20 Marks)

i) The range is the difference between the largest and smallest observations.

In this case it is 14.51 – 4.96 = 9.55 inches

ii) The variance is:

# hand-written formula for variance and standard deviation

jumps\_var = (sum(jumps^2) - sum(jumps)^2/length(jumps))/(length(jumps) - 1)

jumps\_var

9.451781

iii) The standard deviation is the square root of the variance

sqrt(jumps\_var)

3.074375

iv) A box plot, aka box and whisker plot shows if there are outliers.

Calculate mean = Sum Xi / Number X = 9.211538

Outliers are = > mean + 2\*sd

< mean – 2\*sd

On a whisker plot they 1.5 times the IQR.

To get IQR

Sort the data points:

4.96, 5.12, 5.84, 7.21, 7.92, 8.23, 8.32, 9.76, 11.13, 12.11, 12.21, 12.43, 14.51

Find the median:

8.32 – there should be 6 points lower and 6 points higher

Find 1Q – 25% quartile – median of lower 7 = 7.21

Find 3Q – 75% quartile – median of upper 7 = 12.11

Find IQR –> 3Q – 1Q -> = 12.11 – 7.21 = 4.9

1.5 times 4.9 = 7.35

There outliers are any number > 4.9 \* 1.5, and any value < 4.9 \* 1.5

So 8.32 + 7.35 is upper outlier, or median 8.32 – 7.35 is lower outlier.

There are no outliers.

Calculate z scores for each value and rejects those > 2, or less than – 2

Try yourself by making up your own dataset and run the formulae

**PROBABILITY**

1. A gymnasium is looking to offer two levels of service, one priced at €200/month and another at €75/month. The gymnasium offers free try out sessions to generate new customers, and finds that 10% of the attendees will sign up for the €200 service, 50% for the €50 service, and 40% will not sign up for anything.

Calculate and interpret the expected value of a gymnasium membership.

(5 Marks)

The expected value is a weighted mean as it takes the probability of expectations to determine the outcome.

1. Non-destructive evaluation (NDE) describes methods that quantitatively characterize materials, tissues, and structures by non-invasive means, such as X-ray computed tomography, ultrasonic, and acoustic emission. Recently (May 2005) NDE was used to detect defects in steel castings.

Assume that the probability that NDE detects a “hit” (i.e. predicts a defect in steel castings) is .97. Assume also that the probability that NDE detects a “hit” when in fact no defect exists is .005. Past experience has shown that a defect exists once in every 100 steel castings. If NDE detects a “hit” for a steel casting, what is the probability that an actual defect exists?

Hint: -

Define the events:

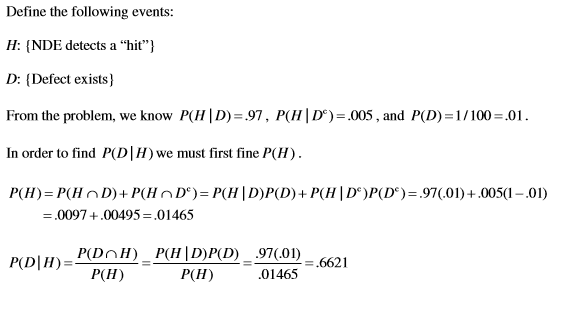
H: {NDE detects a “hit”}

D: {Defect exists}

: {Defect does not exist}

1. Determine P(H|D) and P(H| *[2]*
2. Determine P(H). *[4]*
3. Determine P(D|H) *[4]*

(10 Marks)

****

1. Failure rates for computer crashes can be caused by various problems like power surge, hardware controller, memory failure. Engineers have determined that the probabilities of power, hardware, and memory problems are .02, .04, and .03, respectively.

They have also determined that if the system experiences power surges, it is repairable 15% of the time. Similarly, if hardware controller the system is repairable 70% of the time, and if memory failure occurs, the system is repairable 90% of the time.

1. Formulate the problem mathematically using Bayes Rule. *[4]*
2. What is the probability that the current crash is due to a power surge? *[2]*
3. What is the probability that the current crash is due to a memory failure? *[2]*
4. What is the probability that the current crash is due to power failure? *[2]*

(10 Marks)

Try Yourself:

R: Repairable

PF: Power Failure

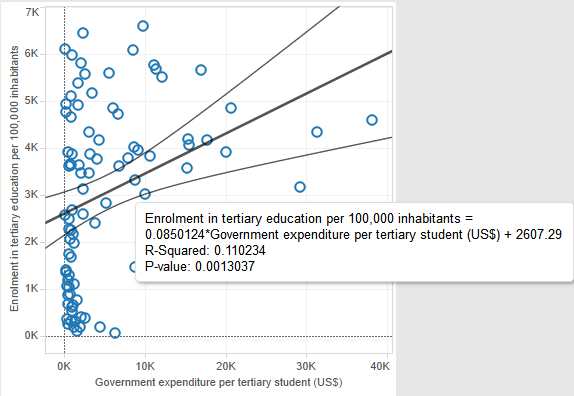
HF: Hardware Failure

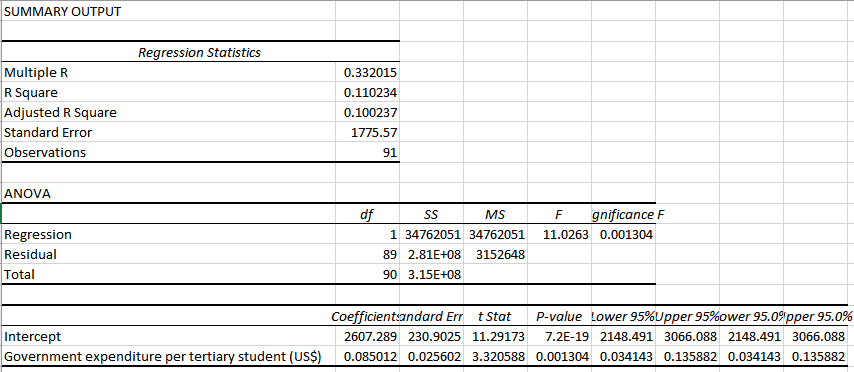
MF: Memory Failure

**REGRESSION ANALYSIS**

1. The World Bank annually collects available data across all countries on government expenditure per tertiary student as well as enrolment in tertiary education per 100,000 inhabitants. There is data available for 91 countries for the year 2010 on both indicators. The following Tableau scatter plot shows tertiary enrolment versus government spend per tertiary student in 2010 for the 91 countries for which data is available. A trend line has been fitted to the Tableau data. The output of an excel analysis are also included below.

What can you tell about the relationship between these two variables from the scatter plot, the fitted simple linear regression model and the correlation coefficient r?





(7 Marks)

As the correlation coefficient is low (r=0.33), this indicates that there is a weak correlation (linear relationship) between government expenditure per enrolled student and number of enrolled students. The Regression formula (y = 0.85012x + 2607.289) would not provide a good model to predict enrollment (independent variable) based on government spend (dependent variable).

1. An article published in the *American Educational Research Journal (Fall 1998)* used multiple regression to model the students’ perceptions of their ability in science classes. The sample consisted of 165 Grade 5-Grade 8 students in six performance-based science classrooms, all of which use hands-on activities as the main teaching tool. The dependent variable of interest, the student’s perception of his or her ability (y), was measured on a four point scale (where 1 = little or no ability and 4 = high ability). Two types of independent variables were included in the model, *control variables* and *performance behaviour variables.* The control variables are: prior science aptitude (measured on a 4-point scale), score on standardised science test, gender, and classroom (1, 2, 3, 4, 5, or 6). The performance behaviour variables (all measured on a numerical scale between 0 and 1) are: active-leading behaviour, passive assisting behaviour, and acting-manipulating behaviour.
2. Identify the independent variables as quantitative or qualitative. *[2]*
3. Individual *β*-tests on the independent variables all had *p*-values greater than .10 except for prior science attitude, gender, and active-leading behaviour. Which variables appear to contribute to the prediction of a student’s perception of his or her ability in science?  
    *[3]*
4. The estimated *β-*value for the active-leading behaviour variable is .88 with a standard error of .34. Use this information to construct a 95% confidence interval for this *β.* Interpret the interval. *[4]*
5. The following statistics for evaluating the overall predictive power of the model were reported: *R²* = .48, *F* = 12.84, *p* < .001. Interpret the results. *[3]*
6. Hypothesize the equation of the first-order main effects model for *E*(*y*) *[2]*
7. The researchers also considered a model that included all possible interactions between the control variables and the performance behaviour variables. Write the equation for this model for E(y). *[3]*
8. The researchers determined that the interaction terms in the model formulated in 6.above were not significant; therefore, they used the model from 5. above to make inferences. Explain the best way to conduct this test for interaction. Give the null hypothesis of the test.  
    *[3]*

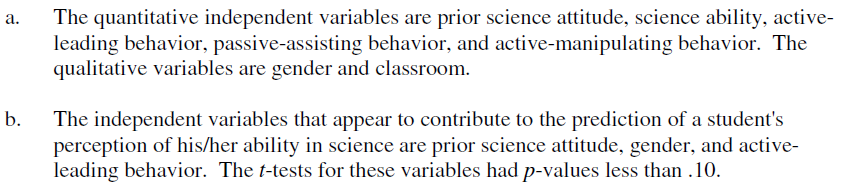
(20 Marks)

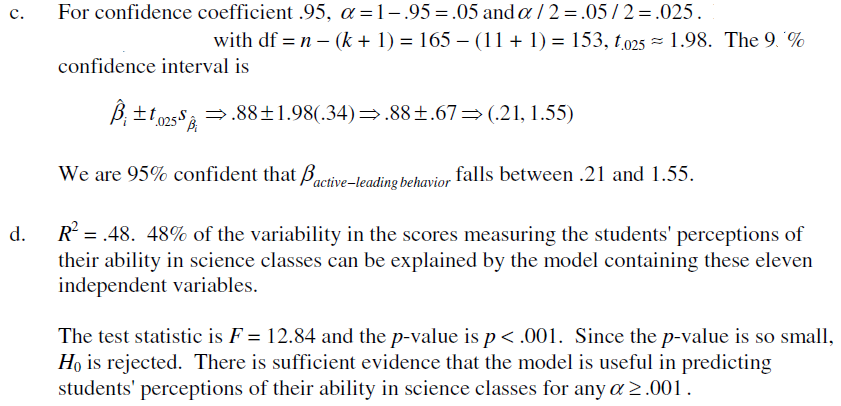
1. A popular website among college students is RateMyProfessors.com (RMP). Established over 10 years ago, RMP allows students to post quantitative ratings of their instructors. In *Practical Assessment, Research & Evaluation (May 2007)*, University of Maine researchers investigated whether instructor ratings posted on RMP are correlated with the formal in-class student evaluations of teaching (SET) that all universities are required to administer at the end of the semester. Data collected for *n* = 426 University of Maine instructors yielded a correlation between RMP and SET ratings of .68.
2. Give the equation of a linear model relating SET rating (y) to RMP rating (x). *[3]*
3. Give a practical interpretation of the value of r = .68. *[3]*
4. Is the estimated slope of the line, in 1. above positive or negative? Explain. *[3]*
5. A test of the null hypothesis is Hₒ: p = 0, yielded a p-value of .001. Interpret this result. *[3]*
6. Compute the coefficient of determination, r², for the analysis. Interpret the result. *[3]*

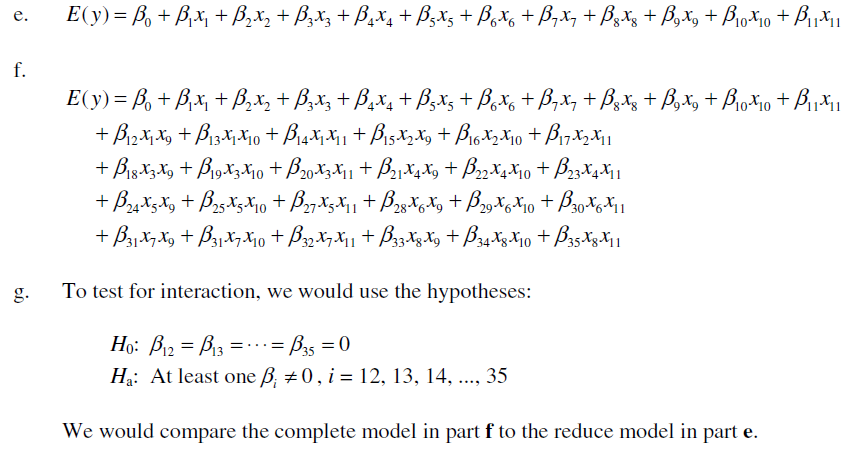
(15 Marks)

***Question 4 Solution***

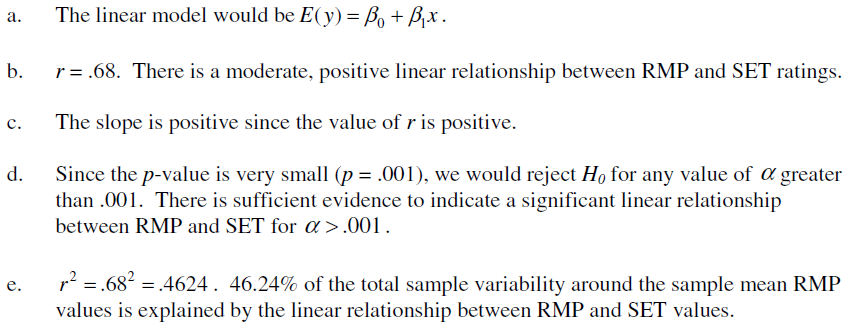
Part a)

****

****

****

Part b)

****

1. My wife’s spending each day over a 9 day period in the Christmas sales was recorded as follows:

|  |  |
| --- | --- |
| **Date** | **Spending** |
| 26-Dec-16 | 1858 |
| 27-Dec-16 | 1433 |
| 28-Dec-16 | 1836 |
| 29-Dec-16 | 1605 |
| 30-Dec-16 | 1282 |
| 31-Dec-16 | 1767 |
| 01-Jan-17 | 1443 |
| 02-Jan-17 | 1275 |
| 03-Jan-17 | 1716 |

*Data source: Receipts in Wife’s Handbag.*

Assume that you have been tasked with making a forecast of my wife’s spending for 4th January based on these figures.

1. Use a three-point simple moving average to estimate the spending for 4th January.

*[1]*

1. Use a three-point weighted moving average to estimate the closing price for 4th January (use weights of 3.0 for the most recent date, 2.0 for the next date, and 2.0 for the last date).

*[3]*

1. Specify and interpret the formula for calculating an estimate for the 4th January stock price using exponential smoothing with a smoothing constant (α) of 0.75.

*[4]*

(8 Marks)

Part c)

1. (B10\*B9\*B8)/3 = (1716+1275+1443)/3 = 1478
2. (3\*B10+2\*B9+2\*B8)/7 = (3\*1716 + 2\*1275 + 2\*1443)/7 = 1512
3. . The smoothing constant ( determines the level at which previous observations influence the forecast. Large weights result in faster changes in the fitted line. Small weights result in slower changes in the fitted line. Commonly used values between 0 and 1 with steps .2 or less. In this case implies larger weight applied to previous actual and faster change in forecast observations.