Math Studies

Project

@DAA

2014 – 2015

Rubrics for marking

Guidelines for writing

Help with calculations

Ideas for project topics

Useful websites

Math Studies Project Assessment Criteria & Guidelines

Criterion A: Introduction

In this context, the word "task" is defined as "what the student is going to do"; the word "plan" is defined as "how the student is going to do it". A statement of the task should appear at the beginning of each project. It is expected that each project has a clear title.

Achievement level	Descriptor
0	The project does not contain a clear statement of the task. *There is no evidence in the project of any statement of what the student is going to do or has done.*
1	The project contains a clear statement of the task. *For this level to be achieved, the task should be stated explicitly.*
2	The project contains a title, a clear statement of the task and a description of the plan. *The plan need not be highly detailed, but must describe how the task will be performed. If the project does not have a title, this achievement level cannot be awarded.*
3	The project contains a title, a clear statement of the task and a detailed plan that is followed. *The plan should specify what techniques are to be used at each stage and the purpose behind them, thus lending focus to the task.*

Guidelines for Criterion A: Introduction

- Have a **clear title** for your project (**on a title page**). Eg
    - "economics around the world"
    - "Does the engine size of a car have any effect on the performance of the car"
    - "Physical attributes of a person and intelligence; is there a connection?"

- Remember to include early in your introduction a **statement on why you chose this topic** and why it is interesting to you. Include an expectation of **what you expect to find and your reasons why**. It doesn't matter if in the end, you show your ideas were wrong. NOTE this IS NOT the null hypothesis but is just a general statement of what you think the data will show. (Your introduction shouldn't have the H_0 and H_1 in it).

- The **statement of the task should be clearly written in its own paragraph** in the introduction so it is easy to find. Consider making it **bold** or in *italics* so it stands out. Failure to do this means you can only get 0 in this criteria. The task statement should be a more specific question than the title. Eg
    - "Does the price of an economy airline ticket depend on the distance between cities or the number of airline companies which fly that route?"
    - "Are taller people more intelligent than shorter people?"

- You should clearly **state how you will collect the data**.
    - **Identify the variables** that you want to test and state how you will collect the data (doing experiments, carrying out surveys, searching for data from reliable sources (give citations for all internet sources used) etc).
    - Make sure your data is fair, relevant and appropriate for the task.
    - Make sure you collect enough data (about 100 "pieces")
    - Be aware of sources of error (biased survey, incomplete data on website)

- For a "detailed description of the plan" you need to **list what processes you will do** (that is, describe the statistical tests eg, "find mean, median and standard deviation"), **state why** you are doing them, **and make sure you do them**. Include how you will present your information (tables & graphs etc). **Go back at the end after your project is finished to edit this to ensure your plan is correct.** (this is where many people lose a mark)

- Remember that you will **need to do 2 variable analysis** (correlation, line of best fit or chi squared test) if you want to score highly. Therefore you will **need to collect data with 2 variables**.
    - Both variables need to be quantitative and continuous for regression. (this is what I recommend collecting)
    - Discrete data will work for a χ^2 test but not for correlation.

Overall your introduction should be about a page in length

Criterion B: Information/measurement

In this context, generated measurements include those that have been generated by computer, by observation, by prediction from a mathematical model or by experiment. Mathematical information includes geometrical figures and data that are collected empirically or assembled from outside sources. This list is not exhaustive and mathematical information does not solely imply data for statistical analysis. If a questionnaire or survey is used then a copy of this along with the raw data must be included.

Achievement level	Descriptor
0	The project does not contain any relevant information collected or relevant measurements generated. *No attempt has been made to collect any relevant information or to generate any relevant measurements.*
1	The project contains relevant information collected or relevant generated measurements. *This achievement level can be awarded even if a fundamental flaw exists in the instrument used to collect the information, for example, a faulty questionnaire or an interview conducted in an invalid way.*
2	The relevant information collected, or set of measurements generated, is organized in a form appropriate for analysis **or** is sufficient in both quality and quantity. *A satisfactory attempt has been made to structure the information/measurements ready for the process of analysis, **or** the information/measurement collection process has been thoroughly described and the quantity of information justified. The raw data must be included for this achievement level to be awarded.*
3	The relevant information collected, or set of measurements generated, is organized in a form appropriate for analysis **and** is sufficient in both quality and quantity. *The information/measurements have been properly structured ready for analysis **and** the information/measurement collection process has been thoroughly described and the quantity of information justified. If the information/measurements are too sparse or too simple, this achievement level cannot be awarded. If the information/measurements are from a **secondary** source, then there must be evidence of sampling if appropriate. All sampling processes should be completely described.*

Guidelines for Criterion B: Information/Measurement

- "*organized in a form appropriate for analysis*" – the **raw data needs to be in a table** which is easy to understand. Don't forget headings & titles. (Make sure your raw data is included in the main part of the project). For complex data, you may then need to group your results into another table, but this should follow the raw data.
- If you use **data from a secondary source** (eg a website), the URL must be put in a footnote and include a **printout or screen shot of the webpage in your appendix**. You need to **reorganize this data** in some way – you won't get credit for using someone else's table. A good way to do this is to sample the data provided on the site or to use information from 2 different places for your two variables. Otherwise, think of an alternative way to reorganize the data so it is not copied from the website (which is why the screenshot is so important – remember to do this when you collect the data, not later when the IA is due to be handed in).
- If you use **data from a secondary source** (eg a website), you **must explain how you sampled the data or justify why you didn't**. An example of this would be to sample the "best 100" or if you are using country data, restrict your data to Europe and Africa for example. Only sample if the amount of information is much more than 100, otherwise use it all and say that it wasn't sampled to make your results as representative as possible. Don't use "make it easier" or "alphabetical" as a reason. "Random" sampling would need to be done with a random number generator (Lots of online sources of these).
- The **data collected must be relevant** - so don't ask for any information that you won't use or isn't relevant for your task. This is especially true if you are using a survey – it can be tempting to ask extra questions "just in case". Deciding early on your research question helps you to ensure you only collect the information you need.
- The **data MUST be suitable for BOTH simple and complex calculations** – that is, you should be able to calculate a mean or median, and preferably be able to display it as a box-plot or a histogram as a visual summary. This means you **MUST collect some numerical data** – design some of your questions to get numbers as a response. It must also lend itself to correlation or χ^2.
- "*sufficient in both quality and quantity*" – **needs a minimum of 80 – 100 pieces of data** depending on what test you are doing. You must plan to collect this much data right from the start. If you are using data from the internet, you may need to choose a sample of all the data. **You need to explain and justify the method you used**. (eg a random sample – how did you make sure it was random)
- You also should collect "GPA", "number of hours", or whatever as **raw results** NOT as pre-grouped data.

  Eg: How many hours did you sleep on Saturday night?

  10 or more _____
  8 hours up to 10 _____
  6 hours up to 8 _____
  4 hours up to 6 _____
  Less than 4 hours _____

  compared to: Please state how many hours you slept to the nearest hour.

  This gives you much greater flexibility as you can change the grouping size if necessary for your chi square test.

- A copy of the survey with all the questions asked (including a scan of some replies) should be i the appendix.

Criterion C: Mathematical processes

When presenting diagrams, students are expected to use rulers where necessary and not merely sketch. A freehand sketch would not be considered a correct mathematical process. When technology is used, the student would be expected to show a clear understanding of the mathematical processes used. All graphs must contain all relevant information. The teacher is responsible for determining the accuracy of the mathematics used and must indicate any errors on the final project. If a project contains no simple mathematical processes, then the first two further processes are assessed as simple.

Achievement level	Descriptor
0	The project does not contain any mathematical processes. *For example, where the processes have been copied from a book, with no attempt being made to use any collected/generated information.* *Projects consisting of only historical accounts will achieve this level.*
1	**At least two** simple mathematical processes have been carried out. *Simple processes are considered to be those that a mathematical studies SL student could carry out easily, for example, percentages, areas of plane shapes, graphs, trigonometry, bar charts, pie charts, mean and standard deviation, substitution into formulae* ***and any*** *calculations and/or graphs using technology only.*
2	At least two simple mathematical processes have been carried out **correctly**. *A small number of isolated mistakes should not disqualify a student from achieving this level. If there is incorrect use of formulae, or consistent mistakes in using data, this level cannot be awarded.*
3	At least two simple mathematical processes have been carried out correctly. All processes used are **relevant**. *The simple mathematical processes must be relevant to the stated aim of the project.*
4	The simple relevant mathematical processes have been carried out correctly. In addition, **at least one relevant further process** has been carried out. *Examples of further processes are differential calculus, mathematical modelling, optimization, analysis of exponential functions, statistical tests and distributions, compound probability. For this level to be achieved, it is not required that the calculations of the further process be without error.* ***At least one further process must be calculated showing full working.***
5	The simple relevant mathematical processes have been carried out correctly. In addition, at least one relevant further process has been carried out. All processes, both simple and further, that have been carried out **are without error.** *If the measurements, information or data are limited in scope, then this achievement level cannot be awarded.*

Guidelines for Criterion C: Mathematical Processes

- Accuracy (or lack of it) is the biggest reason for losing marks. Check your answers with the calculator & consider getting a buddy so you can check each other's. **Every calculation must be correct!** Your teacher will NOT check your calculations for accuracy during the draft writing process – this is your responsibility.

- You **need at least 2 *relevant* simple processes**. These are most likely to be calculations of the mean (needed for regression), standard deviation (needed for regression) and some form of graphical representation (eg scatter plot, box-plot or histogram) so you can make some conclusions about your data before going on to further processes. If you are doing a chi squared test instead of regression, finding the quartiles and drawing a box plot will be relevant to help you group your data.

- The **maximum you can score is 3 if you only do simple processes**. Further processes (chi square, regression) will be considered *simple processes* in the following cases:

    - They are done with technology – if you use your calculator to do the calculation, it is now simple! **You need to do EVERY calculation by hand and show ALL the step**s in your project.
    - If you have no other simple calculations – if you haven't done any one-variable calculations, then your further process now counts as a simple one! You MUST do some calculations that can be considered simple first (and they must be relevant).

- **Graphs may be made with a computer or your calculator** (insert screen captures into your project) but make sure the axes scales and labels are added. I would recommend using a graphing programme over your calculator as the end result is much more polished unless using TI Nspire.
  You can use Excel or download free software from the internet. Each graph should be at least a page width – don't make them too small.
    - Macs have "Grapher" already installed and Windows laptops can download "Graph" (4.4 is the latest version) for free.
- Cumulative frequency curves should be **drawn by hand**. Scan them into your project. (not done very often as it is difficult to show relevancy)
- **Every graph and calculation MUST be commented on**. What does the graph or calculation tell you? You should be able to write at least 1 or 2 sentences about each thing you do (this helps to demonstrate relevancy). Do not draw two graphs from which you will draw the same conclusion. (No point in this!) **Each graph & calculation must be relevant to the task** and included in your introduction's plan.
- "*at least one relevant further process*" – you need to explain or justify why you used each process. What is the point? What will the process let you say? You must comment on what the result means. Further process means χ^2 test or correlation or distribution.

See pages 15 – 18 in this booklet for formulas and more guidance.

Criterion D: Interpretation of results

Use of the terms "interpretation" and "conclusion" refers very specifically to statements about what the mathematics used tells us after it has been used to process the original information or data. Discussion of limitations and validity of the processes is assessed elsewhere.

Achievement level	Descriptor
0	The project does not contain any interpretations or conclusions. *For the student to be awarded this level, there must be no evidence of interpretation or conclusions anywhere in the project, or a completely false interpretation is given without reference to any of the results obtained.*
1	The project contains at least one interpretation or conclusion. *Only minimal evidence of interpretations or conclusions is required for this level. This level can be achieved by recognizing the need to interpret the results and attempting to do so, but reaching only false or contradictory conclusions.*
2	The project contains interpretations and/or conclusions that are consistent with the mathematical processes used. *A "follow through" procedure should be used and, consequently, it is irrelevant here whether the processes are either correct or appropriate; the only requirement is consistency.*
3	The project contains a meaningful discussion of interpretations and conclusions that are consistent with the mathematical processes used. *To achieve this level, the student would be expected to produce a discussion of the results obtained and the conclusions drawn based on the level of understanding reasonably to be expected from a student of mathematical studies SL. This may lead to a discussion of underlying reasons for results obtained.* *If the project is a very simple one, with few opportunities for substantial interpretation, this achievement level cannot be awarded.*

Guidelines for Criterion D: Interpretation of results

- You **MUST comment on every process/calculation/graph** and talk about it shows. Demonstrate that you understand what the standard deviation tells you, or what a large IQR represents. Use your textbook and class notes to help you with this. This should be ongoing as you work and not just left until the conclusion. *Do not copy old projects!!!*

- For the chi squared test, you must get the right conclusion (look up your notes to be sure). Make sure you use the critical value table included here and not the p-value for accepting/rejecting the null hypothesis. What does 5% probability mean? Can you reject the H_0 at a lower probability? What would this mean?

- For correlation, what is the strength of the correlation? Do NOT make a regression line of the strength is weak. If you can make a regression line, use it to make a prediction.

- Do your conclusions agree with your initial ideas/expectations?

- Summarize your overall findings in the conclusion, as well as your feelings about it.

Criterion E: Validity

Validity addresses whether appropriate techniques were used to collect information, whether appropriate mathematics was used to deal with this information, and whether the mathematics used has any limitations in its applicability within the project. Any limitations or qualifications of the conclusions and interpretations should also be judged within this criterion. The considerations here are independent of whether the particular interpretations and conclusions reached are correct or adequate.

Achievement level	Descriptor
0	There is no awareness shown that validity plays a part in the project.
1	There is an indication, with reasons, if and where validity plays a part in the project. *There is discussion of the validity of the techniques used* **or** *recognition of any limitations that might apply. A simple statement such as "I should have used more information/measurements" is not sufficient to achieve this level. If the student considers that validity is not an issue, this must be fully justified.*

Guidelines for Criterion E: Validity

- Validity is one of the most poorly done aspects of the project. **Give it its own section or put it into the conclusion with a clear heading**. Questions you need to ask and answer honesty after you have finished are:

    - How random/representative was your data collection?
    - Is there enough data?
    - Was there any bias (or lack of honesty in replies) or could there have been?
    - How certain are you that your results /conclusions are true or accurate?
    - Look at your chi squared result – is it valid? (you MUST discuss this with a reason)

- Comment on the way data was collected, the sample size, possible bias or dishonesty, any problems with the survey – the way the questions were asked or how they were used – eg improving the structure of the questions.

- Make some good sensible suggestions about how the project could be extended - eg supplementary questions that could be asked or extending who the questions were asked to.

- If you do chi squared tests, you MUST discuss the expected values – are they all > 5? Are your results valid? Explain how you know they are. If the df = 1, did you use the Yates correction formula to calculate χ^2

- If you find the correlation coefficient, you must relate r to confidence in making conclusions based on correlation. (ie if r is close to 1 you can be more confident than if it is 0.6)

- **NEVER draw or calculate a regression line if the correlation is weak** and say so, make your understanding clear.

- look back at the way that you collected your data and see if there were any flaws in the process of carrying out the questionnaire or in taking measurements. Comment on these and suggest improvements.

- Look back at your calculations. Were any errors incurred? If so why and how could they be avoided in the future?

- Were any calculations or graphs in fact inappropriate or irrelevant with hindsight. Can you suggest what they could be replaced by if you were to repeat the process.

- If you have worked on lines of regression, will the equation you have found be relevant for all values of the variable x? Was a linear model suitable?

- What (if any) assumptions did you make?

Criterion F: Structure and communication

The term "structure" should be taken primarily as referring to the organization of the information, calculations and interpretations in such a way as to present the project as a logical sequence of thought and activities starting with the task and the plan, and finishing with the conclusions and limitations.

Communication is not enhanced by a large number of repetitive procedures. All graphs must be fully labelled and have an appropriate scale.

It is not expected that spelling, grammar and syntax are perfect, and these features are not judged in assigning a level for this criterion. Nevertheless, teachers are strongly encouraged to correct and assist students with the linguistic aspects of their work. Projects that are very poor linguistically are less likely to excel in the areas that are important in this criterion. Projects that do not reflect the significant time commitment required will not score highly on this assessment criterion.

Achievement level	Descriptor
0	No attempt has been made to structure the project. *It is not expected that many students will be awarded this level.*
1	Some attempt has been made to structure the project. *Partially complete and very simple projects would only achieve this level.*
2	The project has been structured in a logical manner so that it is easily followed. *There must be a logical development to the project. The project must reflect the appropriate commitment for this achievement level to be awarded.*
3	The project has been well structured in accordance with the stated plan **and** is communicated in a coherent manner. *To achieve this level, the project would be expected to read well, and contain footnotes and a bibliography, as appropriate. The project must be focused and contain only relevant discussions.*

Guidelines for Criterion F: Structure and Communication

- **Structure includes the order of your project** –title page, introduction – why you want to study this question, statement of task, data collection & raw data tables, simple calculations & displays of data, further calculations, and a conclusion that includes validity, bibliography, appendix

- **Use headings** to make your project easier to follow eg see above (**but not these criteria**). Make sure your pages are numbered and you have a contents page.

- **Tables should not be split over 2 pages** – make more columns, reduce font size – what ever it takes – keep each table on one page only.

- Don't forget a title page

- Try to **print in colour** – we send originals to the IB & if you have colour coded graphs etc, it looks clearer in colour than black & white.

- Structure also includes appendices (where you **MUST have a copy of your questionnaire/ screenshot of the website your data came from**

- Headings and Titles on tables and diagrams are important for communication.

- You MUST explain every calculation in words (in a sentence) before you do it and give **ONE fully worked example**. If you are doing many of the same type of calculation, you can reduce the amount of working you show after the first one.

- Remember to comment on the significance (meaning) of every process.

- You must include all your websites you used. Ask the librarians if you need help with citations.

End your project with a clear conclusion which

- summarizes your main results

- concludes whether or not your initial expectation was correct

- makes suggestions for further investigation (NOT "I would collect more data!")

- discuss validity of results and math processes.

- If you have used website addresses they should be noted in a **footnote** at the bottom of the page. Also quotes should be referenced in a footnote.
- If you have used books/journals these must be detailed in a **bibliography.** You need to record the author, title, date and place of publication and the name of the publisher.
- In the **appendix** you should include copies of questionnaires and printouts of webpages used.

Use footnotes when:

- You have a quote from a publication
- Use a formula from a math book or the formula booklet

Criterion G: Notation and terminology

This criterion refers to the use of correct terminology and mathematical notation. The use of calculator or spreadsheet notation is not acceptable.

Achievement level	Descriptor
0	The project does not contain correct mathematical notation or terminology. *It is not expected that many students will be awarded this level.*
1	The project contains some correct mathematical notation **or** terminology.
2	The project contains correct mathematical notation **and** terminology throughout. *Variables should be explicitly defined. An isolated slip in notation need not preclude a student from achieving this level. If it is a simple project requiring little or no notation and/or terminology, this achievement level cannot be awarded.*

- *Terminology*: make sure you use the correct words – chi square is a test for independence, NOT correlation or a relationship.
- Use IB symbols, not calculator ones eg Standard deviation of x is S_x not the σx that the calculator uses.
- You **must define any variables used**. This is important when finding the mean, standard deviation, etc and you use an "x" or "y". You MUST define what your "x" and "y" represent.

- Watch overall *notation*:
  - use the correct symbol for Chi squared: χ^2 (Use a font that shows it clearly – this is Times New Roman & it is of course a Greek letter).
  - Do NOT use ∧2 – use the superscript instead – ask if you don't know how!
  - Use xy not x*y unless you are explaining a formula in a spreadsheet, but these should start with =x*y and be explained in words clearly.
  - Download "Math Type" a free software add-on for a Mac or PC that lets you type fractions & square roots, formulae etc. Or use Microsoft equation editor. Do not type fractions with a / - type them using one of the above editors so it has a horizontal line.
  - If you are doing a box & whisker plot, there are free sites that let you do this online & you can save the image for your project.

Guidelines for Calculations (by hand)

You must **show full working with formulae** if you want to credit for your mathematical processes. Here are the formulas for the calculations you will typically be doing.

If you are calculating **r values or the regression line**, you should use a spreadsheet such as excel. You will need to include a full print-out of this in your appendix. Transfer the *sums from each column* into the main section of your project in a table with an explanation of where the numbers have come from. Then you can use these values in your calculations. Calculate the standard deviations in the "simple processes" section (after you have calculated the mean) and then in the "further processes" section you can start with calculating the covariance and from there calculate r. (Don't try to make one massive calculation of r in one step as it is much easier to make a mistake)

Use the following formulas (not random formulas you find on the internet – note citation below).

Pearson's correlation coefficient (r)

[1]

$$r = \frac{S_{xy}}{S_x S_y} \quad \text{where} \quad S_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n} \qquad S_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} \qquad S_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}}$$

S_{xy} is called the covariance of X and Y

S_x is the standard deviation of X and S_y is the standard deviation of Y

Set up your spreadsheet columns in the following order:

x, y, $(x-\bar{x})$, $y-\bar{y}$, $(x-\bar{x})^2$, $(y-\bar{y})^2$, $(x-\bar{x})(y-\bar{y})$.

You will need the sums of the forth and fifth columns for standard deviation and the sum of the last column for covariance.

After calculating the r value you need to say what type of correlation it shows. (your text has a table on p224). Only go on to calculate a regression line if the correlation is moderate or strong.

Regression equation (correlation) by hand

[2]

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

you will need to rearrange this after substitution to get it into gradient intercept form.

Here you obviously will have needed to calculate the mean of x and the mean of y already. You will need to show all working (as you usually only do this once in your project). See above for the formulas for covariance and standard deviation of x. The final equation for the line should be in the form of y = mx + c. Afterwards **check your result on your calculator**!

After finding the equation, use it! Make a prediction for what a certain value of y would be given a value of x. Make sure you interpolate and not extrapolate. Then you can comment on how likely your prediction is going to be true (ie another chance to show you understand the significance of r).

[1] [2]

For **chi squared calculations**, you first need to calculate the expected values and put these into an expected value table. Then, to calculate the chi squared value, you should set up a table with each aspect of the calculation as a column and work logically across the table so that the last column can be summed for your answer. Do not do this with excel.

χ^2 calculations.

$$\chi^2_{calc} = \sum \frac{(f_o - f_e)^2}{f_e} \qquad \begin{array}{rl} \text{where} & f_o \text{ is an observed frequency} \\ \text{and} & f_e \text{ is an expected frequency.} \end{array}$$ [3]

You must first calculate the expected values for each of the observed values (see your class notes). Typically you won't have many of these to calculate so you can show the working for all your expected values.

f_e *(A and B) = p(A) x p(B) x total frequency*

If any of your expected frequencies are less than 5, your χ^2 result is INVALID. You will get no credit if this is true, so you will need to reorganize your original table of observed data to correct this. See your teacher for help as soon as possible! This is why you need to collect lots of data in the first place!

Put all your expected values into a new table and finally set up a table for your χ^2 calculations using the following column headings. The sum of the final column is your χ^2 value. (see your lesson notes on chi square for projects)

$$f_o \qquad f_e \qquad f_o - f_e \qquad (f_o - f_e)^2 \qquad \frac{(f_o - f_e)^2}{f_e}$$

You also need to calculate the degrees of freedom by hand. (this should be done right after your table of observed values) *If your df = 1, see the Yates calculations below*.

Finally you need to compare the χ^2 value to the critical value. You will need to look up the table of critical values (page 18 – also note citation) to find the critical value for the degrees of freedom of your calculations. **Do not use the p-value** as this can only be calculated with a calculator.

[3] "Mathematics for the international student Mathematical Studies SL" pages 591

χ^2 calculations: Yates Continuity Correction

You **ONLY** need to do this for your project, never in an exam (so we don't teach it in class). If you have **1 df (degrees of freedom) you MUST use this correction**.

[4]

The following is Yates's corrected version of Pearson's chi-squared statistic:

$$\chi^2_{\text{Yates}} = \sum_{i=1}^{N} \frac{(|O_i - E_i| - 0.5)^2}{E_i}$$

where:

O_i = an observed frequency

E_i = an expected (theoretical) frequency, asserted by the null hypothesis

N = number of distinct events

This means you will have to set up your tables a little differently as you need to calculate the **absolute value of observed – expected** and then subtract 0.5 each time before you square the result, divide by the expected value and then finally sum your values. See table headings to use below.

f_o	f_e	$\lvert f_o - f_e \rvert$	$\lvert f_o - f_e \rvert - 0.5$	$(\lvert f_o - f_e \rvert - 0.5)^2$	$\frac{(\lvert f_o - f_e \rvert - 0.5)^2}{f_e}$

" The effect of Yates's correction is to prevent overestimation of statistical significance for small data." [5]

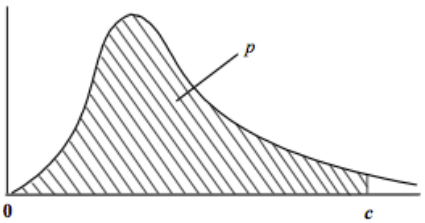
If you have a 2 x 2 contingency table then you need to do this correction regardless of how much data you have collected. (IB requirement).

Remember to **CHECK all your calculations** with a **calculator**! Incorrect calculations will lose you at least 1 point on the rubric. Your teachers WILL NOT do this for you – so don't ask!

[4, 5] from Wikipeadia.com http://en.wikipedia.org/wiki/Yates's_correction_for_continuity (29/09/2012)

6 # Critical values of the χ^2 distribution (topic 6.9)

$p = \mathrm{P}(X \leq c)$



p	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
$\nu = 1$	0.00004	0.0002	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

ν = number of degrees of freedom

[6] "IBO Diploma Programme Mathematical Studies SL Information Booklet, first examination 2006"

Examiner Feedback from May 2014 - lots of helpful notes

Criterion A:

Many candidates were awarded a level 2 out of a possible 3. This was mainly due to the fact that they did not give any reasons for the processes they were going to use.

Some candidates only scored 1 mark because their plan was not clear or their project had no title.

To award level 3 there should be no surprises when reading the project. For the plan to be considered detailed, the student should describe precisely all the mathematical processes to be used and the reasons for choosing each of these processes.

If any processes are used that are not discussed in the introduction then at most level 2 can be awarded.

If any process is explained in the introduction but not performed, then at most level 2 can be awarded.

Candidates with a clear statement of task and detailed plan discussing the processes to be used and the rationale behind their choices usually produced excellent projects.

Criterion B:

Most candidates collected data that was appropriate for their project but it was not always sufficient in quantity to perform the processes set out in their plan.

Few candidates described their sampling process clearly and so were not awarded full marks for this criterion.

The collection process must be thoroughly described and must be representative of the population. Saying that the data was randomly collected is insufficient. The sampling process must be explained. If sampling is not done then this must be justified.

If no real organization of the data is required then at most level 2 can be awarded for this criterion.

Raw data must be seen to consider level 2 for this criterion.

Calculations must be able to be checked.

Data that is too simple also limits the marks for other criteria such as the mathematical processes, interpretation and communication.

Criterion C:

Most of the changes in the new assessment criteria are in this criterion. Not all teachers and candidates paid attention to the changes and, as a result, did not score well.

Simple processes are considered relevant if they pertain to the statement of task and if these processes are used later in the development of further processes, as stated in their plan.

If there are no simple processes in the project, then two of the further processes will be considered to be simple processes and **not** further processes.

Repeated processes count as one process (e.g. producing two bar charts).

If the project includes only two processes and one is incorrect, then level 1 is the maximum which can be awarded.

If there is only one process used, simple or further, then the candidate is awarded level zero.

If the simple and further processes are not presented in order, the student will not be penalized in this criterion. However this may be penalized in criterion F.

To be awarded level 5 all further processes (and there only needs to be one) must be without error, and must be relevant.

Any process that is beyond the course needs to be fully explained to be considered a further process, for example the unsupported use of the *t*-test, whether performed wholly on the GDC or by substitution into the formula is deemed a simple process.

Although the processes are not limited to the chi-squared test and calculating the regression equation, the frequency with which they appear makes it worthwhile producing further guidance on how they should be marked.

Chi-squared test

A χ^2 test performed by hand is considered to be one further process.

For a completed χ^2 test candidates are expected to write down their hypotheses, degrees of freedom, show how to calculate at least one expected value and complete the table of expected values, work out the chi-squared test statistic using the formula and write down the conclusion (using either the critical value or the significance level).

If the observed values are not frequencies, then at most level 3 can be awarded for criterion C.

If any expected values are less than 5, then at most level 4 can be awarded for criterion C, and only if all the working is shown in full. If the working is not shown, then at most level 3 can be awarded.

If the degree of freedom is 1, then Yates continuity correction must be applied (and only when the degree of freedom is 1). If the correction factor is not applied and the test has been satisfactorily performed by hand then at most level 4 can be awarded.

Candidates should note that a χ^2 test does not prove anything. It supplies evidence or support only.

Correlation / regression

If the candidate draws a scatter diagram and it is clear from the diagram that there is no correlation then it is relevant to calculate the correlation coefficient, r, to verify that fact. However, it is not relevant to calculate the regression line.

If from the scatter diagram it seems that there is some correlation then it is relevant to calculate the correlation coefficient, r, and, if the correlation is strong enough, then it is relevant to find the regression line, provided it is used or its purpose explained.

If a scatter graph is not drawn, then the relevancy of a regression line will depend on the value of r.

If the value of r is written down from the GDC (or Excel) then this is a simple process.

If the summary statistics have been calculated from the GDC and then substituted into a formula to determine r this is also a simple process.

Calculation of the mean or standard deviation as part of calculating r is not considered a separate process. The exception to this is if the mean or standard deviation has been calculated independently as part of the stated plan.

Normal distribution

Sketching a normal distribution curve and calculating probabilities or percentages is a simple process.

Using z-scores is also a simple process.

If a χ^2 goodness of fit test is performed by hand, then this is a further process.

Criterion D:

The project flows better if the candidate writes partial interpretations/conclusions after each mathematical process.

Most candidates managed to give at least one interpretation that was consistent with their analysis. However, the wording in this criterion has now changed and, if there are any inconsistent conclusions/interpretations, then there must be at least two consistent conclusions/interpretations for the candidate to be awarded level 2.

Any irrelevant or unsupported conclusions (or personal beliefs) preclude the award of level 3.

Criterion E:

Many candidates now show more understanding of validity and are able to comment meaningfully on the mathematical processes used or recognize limitations and provide a discussion.

Recognizing and commenting on the need to use the Yates' continuity correction factor or combining groups in the χ^2 test is sufficient for this criterion.

Criterion F:

Overall the structure of the projects was good. However, this criterion covers more than the layout, it also deals with commitment. The project must demonstrate the required time commitment otherwise the maximum that can be awarded is level 1.

Some candidates included unsupported generalizations and this does not lead to a coherent project. Also, a large number of repetitive procedures preclude the award of level 3.

Graphs, tables or processes presented out of order also preclude the award of level 3.

If many pages of raw data or calculations via spreadsheet are presented, it is preferred that these be shown in an appendix; however this is not penalized.

If processes have been mentioned in the introduction and have not been performed or vice versa then the candidate is not penalized twice for the same error.

Criterion G:

Surprisingly few candidates scored full marks on this criterion. The most common level awarded was 1 due to incorrect notation and/or terminology or failure to define variables.

Candidates that use Excel or calculator screen dumps need to be aware that this notation is not acceptable. If there are examples of such notation this must be explained and corrected in the body of the text.

Candidates should avoid using their cameras to take pictures of calculator screens.

Isolated typographical errors are condoned, however if the candidate uses $x\text{^}2$ instead of x^2, for example, this is poor notation and the maximum that can be awarded is level 1.

Examples of notation:

Correct notation	**Incorrect notation**
x^2	$x\text{^}2$ **or** $x2$
$x\times 2$ **or** $2x$	x*2
1.2×10^{-3}	1.2 E-03
χ^2	X^2 **or** x^2
r^2:Coefficient of determination	r^2:Correlation coefficient
$\sqrt{\frac{2402}{16}}$ **or** $\sqrt{(2402/16)}$	$\sqrt{}$ 2402/16 **or** sqrt.

Hints for Collecting and Processing Times, Dates and other Measurements

- Use the metric system – if the original data is in inches and feet, then convert it into cm or metres. Show your working – this is a simple math process.
- Use accurate measuring devices and explain what you used eg the stopwatch on your iphone. Decide on the degree of accuracy you will record (eg to nearest second or tenth of a second) and explain your decision. Be consistent.
- Convert mixed units into one single unit eg
  A stopwatch time of 1.32.4 this is 1 minute and 32.4 seconds
  Either: convert the minutes to seconds (x by 60) and add to the seconds eg 92.4 s
  or convert seconds to a decimal of minutes (÷ by 60) eg 1.54 minutes
  Collect age in years + birthday month (especially with students)
  You can then convert age to a "continuous" scale with years and decimals.
  Ask for help with your calculations if you are doing this.

Where can data be collected from?

- take measurements using a whole range of measuring instruments, such as rulers, tape measures, compasses, protractors, scales and electronic devices
- collect data through surveys and questionnaires
- access timetables or loan-repayment schedules
- search the internet (ensure that sources are referenced and sampled if appropriate)
- carry out experiments.

Designing a questionnaire

There is skill involved in structuring a questionnaire to elicit the required information. The questions should be:

- clearly and courteously phrased
- appropriate in number
- capable of being answered by any person questioned
- capable of being answered easily and in a defined manner, for example, yes or no, a number, a place, a name
- capable of being answered truthfully and willingly
- regarded as being non-intrusive by those questioned.

The author of the questionnaire should always be identified, and the reason for seeking the requested information should be explained. It is always advisable to trial the questionnaire on a small group first so that the questions can be refined before it is distributed to a wider audience.

All data collected must be kept in a confidential and responsible manner and not divulged to any other person. Anonymity for each participant must be guaranteed. The data must not be used for any purpose other than that for which it was collected.

NOTE: Math studies students have often used HS homerooms for surveys. This is efficient and can ensure a large number of samples are obtained. You MUST arrange this through your math teacher and also have your questionnaire ready for printing on the deadline given. We try to send all surveys at the same time to minimize disruption to homerooms.

NOTE: Students in the past who have used survey-monkey questionnaires sent out to their friends have often been disappointed by an insufficient number of responses. This is probably the least efficient way for you to obtain the data you need.

Helpful suggestions (from a mixture of sources):

- **Word Count = 2000 words maximum** (not including tables. graphs and appendices). Include the word count on the title page.

- By far the best projects are those done on topics students are passionately interested in – be it sports, movies, books, music or a psychology or science experiment. Data from reliable sources on the Internet allows you to get started quicker and often provide extensive amounts of data; surveys take time to write, print, distribute and then collate and you need to ensure that enough valid responses are received.
- Everyone in the school should have their own UNIQUE topic – check the google doc that no one else has the same topic (first to choose it gets to keep it).
- The **IB guidelines for projects** (posted on edmodo) has excellent information on ideas for projects, suggestions for meeting each criteria and also has past projects (4 complete ones + parts of projects that illustrate how to (& not to) score well in each criteria. Please look over this document.*
  *note that the criteria has changed so you need to be a little cautious looking at past projects.

Possible Project Topics

Choose your topic and formulate this into a title. Here are some examples of topics and titles. Each student must select a different title.

- *Correlation between race and probability of breast cancer*
- *Can the number of hours spent watching Tv affect grades?*
- *The rate of obesity amongst adults*
- *Comparing memory and age*
- *Does gender affect taste ?(pepsi and coke)*
- *The effectiveness of subliminal learning*
- *How does the colour of light affect growth in pea plants*
- *Where is the best place for me to shoot in a basketball game.*
- *A comparison of M&Ms and smarties in relation to their weight and colour*
- *Does gender affect speeding rates*
- *Is there a relationship between the age of people and the age of their car?*
- *Is there a relationship between the GPA and the persons involvement in a sport?*
- *Does eating breakfast affect school grades*
- *Does age or gender affect reaction times*
- *Does gender affect a persons perception of time?*
- *Height weight and swimming performance.*
- *If you fly further does the airline ticket cost more?*
- *Colour of words affects accuracy of reading*
- *100m sprint times and 100m freestyle times in the Olympic Games*

The list is endless. Choose something that you are interested in.

Your textbook has a long list of possible projects as well – on page 465

Data/Information Websites

Some of this information may be out of date but the topic headings may give you ideas.

Astronomical and Time Information

http://www.skyandtelescope.com/

http://www.nmm.ac.uk/server/show/nav.2857

http://www.peterrussell.com/Odds/WorldClock.php

Computers

http://www.bandwidthplace.com/

http://www.numion.com/YourSpeed3/

http://www.snapshotspy.com/stats.htm#child

Design and Tiling

http://www.uwgb.edu/DutchS/symmetry/penrose.htm

http://www.quadibloc.com/math/pen01.htm

Educational Data – IB, SAT, and AP

http://www.ibo.org/facts/statbulletin/

http://professionals.collegeboard.com/gateway

http://www.ucas.ac.uk/about_us/stat_services/

http://eric.ed.gov/

Food, Nutrition and Diet

http://www.fao.org/waicent/portal/statistics_en.asp

http://fnic.nal.usda.gov

http://www.nal.usda.gov/fnic/foodcomp/Bulletins/faq.html

http://www.ars.usda.gov/Services/docs.htm?docid=8964

General National Statistics

http://www.statistics.gov.uk/

https://www.cia.gov/library/publications/the-world-factbook/

http://www.bized.co.uk/dataserv/datahome.ht

http://www.fedstats.gov/

Global Environment and Biological Population

http://www.metoffice.gov.uk/education/data/index.html

http://www.lib.umich.edu/govdocs/stenv.html

http://www.ulb.ac.be/ceese/meta/data.html

http://www.nws.noaa.gov/

Health and Safety

http://www.hse.gov.uk/statistics/

http://www.cdc.gov/nchs/

http://www.lib.gla.ac.uk/Depts/MOPS/Stats/medstats.shtml

Mathematics in Nature

http://www.mcs.surrey.ac.uk/Personal/R.Knott/Fibonacci/fibInArt.html

http://www.cut-the-knot.org/index.shtml

Physical Fitness and Body Mass Index

http://www.exrx.net/Testing/YouthNorms.html

http://www.cdc.gov/nchs/data/ad/ad347.pdf

http://www.halls.md/body-mass-index/bmirefs.htm

Psychological and Personality Testing

http://www.socialpsychology.org/cognition.htm#visual

http://www.humanmetrics.com/cgi-win/JTypes2.asp

http://similarminds.com/

Sports

http://it.stlawu.edu/~rlock/sports.html

http://www.soccerstats.com/

http://www.olympic.org/uk/passion/studies/index_uk.asp

Travel and Transport and Maps

http://www.dft.gov.uk/pgr/statistics/

http://ntl.bts.gov/tools.html

http://maps.google.co.uk/

More useful websites:

WorldMapper

http://www.worldmapper.org/data_sources.html

An extensive list of data sources for world indices of various kinds, and information about how to access that data.

Eg: " **B. World Bank**
The World Bank. World Development Indicators 2005
http://www.worldbank.org/data
To access data:

1 Under the Key Products heading, click on 'World Development Indicators'.
2 Then click on WDI 2005 Data
3 In the select a country section, highlight the countries in the left-hand box and click select. Click next.
4 Choose a variable of interest - the 'find' box may be of assistance. Click next.
5 Choose the year(s) of interest. Click next.
6 In 'data output options' choose 'save data as excel file'.
7 You can then save the file to your computer. " [8]

Gapminder

http://www.gapminder.org/data/

Amazing variety of data that you can download directly from the site. Playing some of the videos will give you an idea of the correlation strength – this data lends itself to both linear regression if you choose lots of countries (for example) and a health vs wealth indicator. Then you can group the data into categories (eg world regions) and do a chi square test on the data in each category. This use of both chi squared testing and a linear regression will make you eligible for maximum marks. The sheer amount of data also makes validity less of a problem as well (sample size will be large).

Math Type

A free download (for Macs or PCs – be careful which one you choose!) that will allow you to easily type math formulae. Click on the 30 day free trial – after it runs out you get the basic version still on your computer which is all that you ever need.

http://www.dessci.com/en/products/mathtype/

[8] from WorldMapper.com http://www.worldmapper.org/data_sources.html (29/09/2012)

Mathematical Studies SL Project Check List	Check ✓

INTRODUCTION:

Does your project have a front cover with the title of your project?	
Is your candidate name and number and the examination session on the front cover?	
Have you stated clearly what you are going to do?	
Have you explained how you are going to do it?	
Have you explained what mathematical processes you will use and why?	
Did you do everything that you said you would do?	

DATA:

Have you collected data or generated measurements or information?	
Is your raw data included in the project or in an appendix?	
Is your data relevant?	
Is your data sufficient in quantity?	
Do you have quality data?	
Is your data set up for use?	
Have you described the sampling process clearly?	

Math processes and calculations:

Have you performed at least two simple mathematical processes?	
Are these simple processes correct?	
Are the simple processes relevant?	
Have you performed a sophisticated mathematical process?	
Is this sophisticated process relevant?	
Is the sophisticated process correct?	

Analysis and conclusions:

Have you commented on your results?	
Are your comments consistent with your analysis?	
Have you commented thoroughly on everything that you have done?	
Have you commented on validity?	

Format, notation, presentation

Does the project contain only correct notation?	
Does the project contain only correct terminology?	
Is your project laid out in a logical manner?	
Do you have an appendix if one is needed?	
Do you have a bibliography?	

* from your textbook