TOPIC 4
PROBABILITY AND STATISTICS

SUBTOPIC 4.11
HYPOTHESIS TESTING

4.11 Teacher view

Hypothesis testing

Index

The big picture

How confident can we be in the predictions we make?

Suppose you want to determine whether there is a relationship between what sport someone plays and how much they enjoy school. You might be able to use a $1 - 10$ scale to have people rate how much they enjoy school, but the sport someone plays is not a number. So how could you predict whether there is a relationship between a sport and enjoying school? How confident could you be that any relationship you think exists is actually significant?

We used critical values to determine the significance of correlations with Pearson's r and Spearman's r_s, and we used technology to find p-values so we were not reliant on a table of critical values. But these relationships dealt only with ratio, interval and ordinal data. In

Overview (/study/ap 122-cid-754029/

this subtopic, we will explore tests that enable us to verify predictions made in contexts that involve nominal data - the χ^2 test for independence and the χ^2 goodness of fit test. Recall that nominal data are values that tell us how many items in a group fall into different categories. Since this type, nominal data, is the most general type of data, other types (ratio, interval and ordinal data) can be converted to nominal data. Therefore these tests have a broad range of applications in statistics.

We will also compare the means of different sets of data using the t-test.

Concept

The tests discussed in this subtopic help us test the validity of the claims we make about our data. We are able to recognise patterns within data, but do they occur by chance, or would we find the same pattern in another set of data? We can use a variety of statistical tests to ensure that the claims we make are valid.

Theory of Knowledge

As you have inevitably noticed, mathematics is built upon itself conceptually. Lower-level axioms provide the basis upon which more complex mathematics and subsequent axioms are built. This methodology gives maths its epistemic clout.

To what extent are other areas of knowledge built in a similar fashion? If you were to rank the areas of knowledge (AOK) in terms of validity of knowledge produced, would the most axiomatic AOKs be at the top of your list?

Knowledge Question: To what extent is new knowledge a result of prior knowledge?

4. Probability and statistics / 4.11 Hypothesis testing

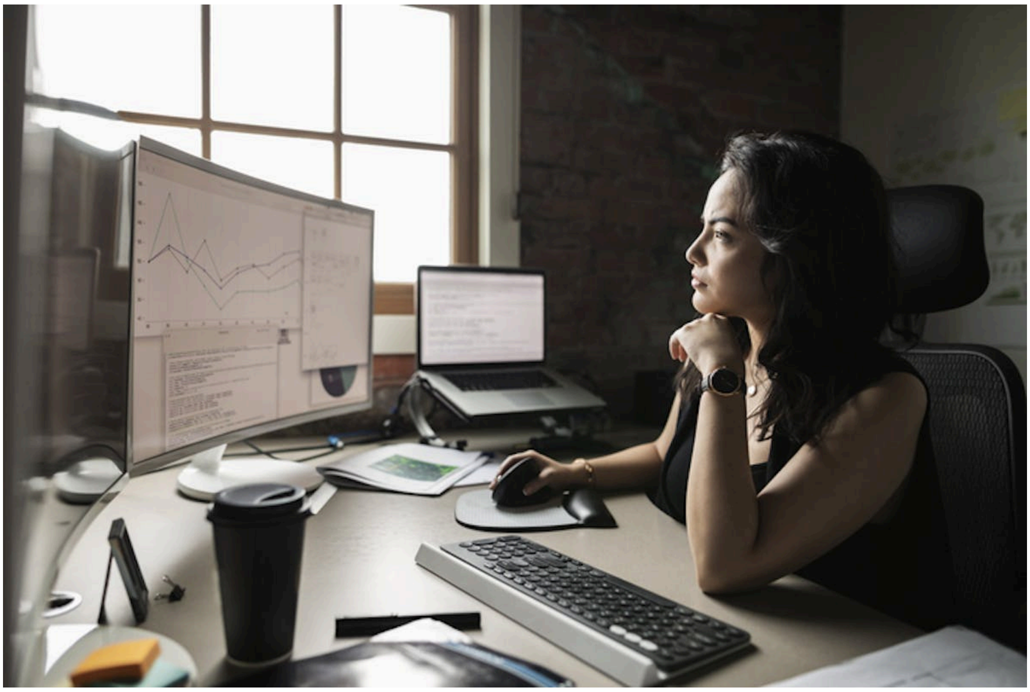
Hypothesis testing and the chi-squared distribution

Section Student... (0/0) Feedback Print (/study/app/m/sid-122-cid-754029/book/hypothesis-testing-and-the-chi-squared-id-27588/print/) Assign

Student view

Overview
(/study/ap
122-
cid-
754029/

Hypothesis testing



How do you verify apparent trends?

Credit: Hero Images Getty Images

Four stages of testing

When researchers in various fields recognise apparent trends in data, they form a hypothesis about the trend but must confirm or reject it somehow. To do this they use a hypothesis test , a formal procedure used to analyse data in order to determine if a hypothesis is likely to be true. The hypothesis test includes four stages.

1. State two hypotheses. These will vary depending on which test you are using. The name of these are the null hypothesis, H_0, and the alternative hypothesis, H_1.
2. Calculate the test statistic. This could be r or r_{s} from your previous studies, χ^2 (chi-squared), which you will use in this subtopic, or one of several other statistical values.
3. Find the probability that the test statistic occurred by chance. This is called the p-value, calculators tell you this probability.
4. State your conclusion. Compare the p-value with the level of confidence you want, called the alpha-level (α). If $p < \alpha$, then you reject H_0, assuming then that H_1 is most likely true. If $p > \alpha$, then you fail to reject H_0.

Student
view

Be aware

We use the phrase 'fail to reject H_0' rather than 'accept H_0' because we are simply saying that the data do not provide enough evidence to conclude H_0 is not true. It still may be false, but we cannot prove it with the data we have.

We will use these four stages when performing hypothesis tests in the sections to come.

The χ^2 distribution

Uses of the χ^2 test

Suppose you believe that there is an association between two variables or properties , such as:

- Most people who study physics also study higher level mathematics.
- People who own large properties tend to vote for conservative political parties .
- People who get the flu vaccination get the flu less than those who don't get the vaccination.

The χ^2 test for independence is used to suggest association, but in fact the test is performed by assuming that the variables are not associated at all. We then study the data set and calculate the probability of our data set occurring purely by chance if there is no association.

If there is no association, we expect the data to be evenly distributed over the categories, relative to the totals in each category.

If the data are heavily concentrated in one category or another, we might suggest that there is an association between those categories.

We can also use the χ^2 goodness of fit test to determine how well the data we observe fit a given distribution. If the data turn out to follow the distribution closely enough, we can make predictions based on the properties of that distribution.

Overview
(/study/ap
122-
cid-
754029/

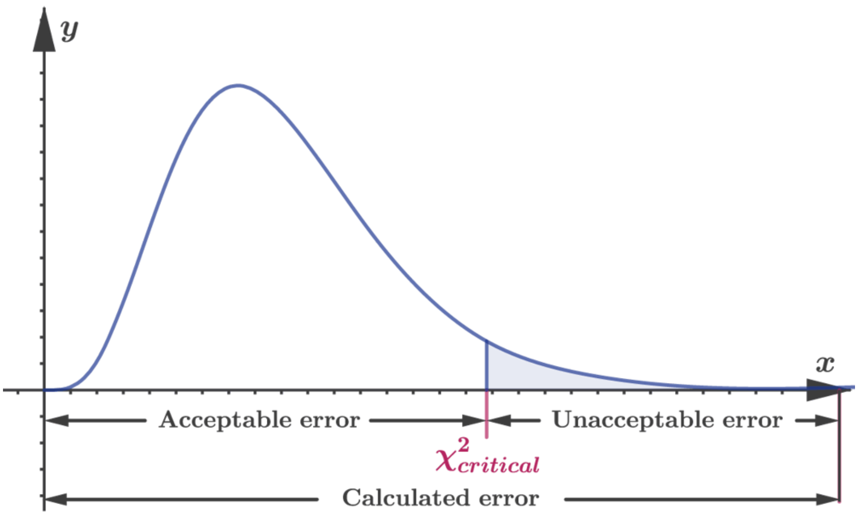
The χ^2 distribution curve

The χ^2 distribution curve is similar to the normal distribution curve. The area under the curve is 1 and the area shaded represents probability. The primary difference is that it is not symmetric.

The horizontal axis is a measure of error, which is calculated in a very particular way. We will see this in section 4.11.2 (/study/app/m/sid-122-cid-754029/book/calculating-the-chi-squared-statistic-id-27589/).

The error is a calculation based on the difference between the observed results (our data set) and the expected results (a hypothetical data set, calculated on the assumption that there is no association between the two variables under study). The error calculation is called χ^2_{calc} .

We are interested in the upper tail (the right tail). If our measure of error, χ^2_{calc}, makes it into the shaded zone, then we think that our data set is suspicious! The value where the red line is drawn in the diagram below is called the χ^2_{critical} value.



More information

The image shows a graph of the chi-square distribution. The Y-axis is labeled as 'y' and the X-axis is labeled as 'x' with indicators for 'Calculated error.' The graph illustrates a bell-shaped curve with a peak approximately midway along the x-axis. To the right of the curve, the area under the curve in the tail is shaded, which represents the 'Unacceptable error' or upper tail. This shaded region corresponds to 5% of the area under the curve. There is a vertical dashed line labeled as 'χ_critical^2,' indicating the critical value where unacceptable error begins. This line separates the 'Acceptable error' region on the left from the 'Unacceptable error' region on the right.

Student
view

[Generated by AI]

In the diagram above, 5% of the area has been shaded. The shaded part is the upper tail, or the right tail.

The horizontal axis is a measure of error, the difference between the observed outcomes of a trial and the mathematically calculated expected outcomes of a trial. If the error calculated is higher than the red line (the critical error, written as χ^2_{critical}, then we say: 'There is evidence at the 5% significance level that the variables are dependent.'

As we discovered before, the normal distribution extends infinitely to the left and right and has two parameters, the mean and standard deviation. The χ^2 distribution is different in these aspects as well.

For the χ^2 distribution, the minimum error is 0, that is, the curve starts at 0. This represents the observed results matching the expected results perfectly. More likely though, small differences between the observed and expected results are common – which is where the results of the bulk of trials end up – with a measure of error located in the acceptable error region of the curve. Very large differences between the observed and expected values are not very likely, and so we see the curve approaching the horizontal axis as the error increases.

The χ^2 curve has only one parameter: the amount of data being compared. For example, this data set, studied in the next section, records the distribution of students over the various Group 3 and 4 classes available in their school, and has $3 \times 3 = 9$ data points (see table below).

	Economics	**Geography**	**History**
Biology	25	46	15
Chemistry	15	44	15
Physics	10	10	20

Student view

The amount of error generated by a 3×3 table (9 cells) will be different from the amount of error generated by a 4×5 table (20 cells). This value is captured in the computation called degrees of freedom. The curve above has been drawn for a 3×3 table. On a 3×3 table,

Overview (/study/ap 122- cid- 754029/

there are $2 \times 2 = 4$ degrees of freedom. An explanation of what is meant by degrees of freedom can be found here (http://blog.minitab.com/blog/statistics-and-quality-data-analysis/what-are-degrees-of-freedom-in-statistics) .

2 section questions

4. Probability and statistics / 4.11 Hypothesis testing

Calculating the chi-squared statistic

Section Student... (0/0) Feedback Print (/study/app/m/sid-122-cid-754029/book/calculating-the-chi-squared-statistic-id-27589/print/) Assign

Observed and expected frequencies

A school with 200 IB Diploma Programme students offers the choice between Economics, Geography and History in Group 3. The school offers the choice between Biology, Chemistry and Physics in Group 4.

The data set showing student choices is shown in in the table below. A table like this comparing the number of elements in different categories of two different variables is known as a contingency table or a two-way frequency table. In this example the two variables are Group 3 class and Group 4 class, and each variable contains nominal data separated into three categories.

	Economics	Geography	History	Total
Biology	25	46	15	86
Chemistry	15	44	15	74
Physics	10	10	20	40
Total	50	100	50	200

Student view

Overview
(/study/ap
122-
cid-
754029/

Exam tip

The χ^2 statistic can be calculated using variables with any number of categories, meaning the contingency table can have as many rows and columns as needed. For this course there will be a maximum of four rows and four columns.

Expected frequencies from the formula

In the table above you see the observed number of students studying each subject combinations. To see whether these observations are consistent with the assumption that Group 3 and Group 4 subjects are chosen independently, you can find the expected frequencies and compare the two numbers. For example, you see from the table above that out of the 200 students, 86 study biology and 50 study economics. If these subject choices were independent, then the ratio of the number of students studying both biology and economics compared to the number of students studying economics, would be the same as the ratio of students studying biology compared to all students.

$$\frac{x}{50} = \frac{86}{200}$$

You can rearrange this to get the expected number of students studying both biology and economics.

$$x = \frac{50 \times 86}{200}$$

To get the expected frequencies, for each cell, we multiply the total for the column by the total for the row, and divide by the grand total .

	Economics	Geography	History	Total
Biology	$\frac{50 \times 86}{200}$			86
Chemistry				74
Physics				40
Total	50	100	50	200

Student view

Overview
(/study/ap
122-
cid-
754029/

Important

This formula is

$$\text{expected frequency} = \frac{\text{row total} \times \text{column total}}{\text{grand total}}.$$

We continue in like manner for each of the cells.

	Economics	Geography	History	Total
Biology	$\frac{50 \times 86}{200}$	$\frac{100 \times 86}{200}$	$\frac{50 \times 86}{200}$	86
Chemistry	$\frac{50 \times 74}{200}$	$\frac{100 \times 74}{200}$	$\frac{50 \times 74}{200}$	74
Physics	$\frac{50 \times 40}{200}$	$\frac{100 \times 40}{200}$	$\frac{50 \times 40}{200}$	40
Total	50	100	50	200

The resulting table is computed as below.

	Economics	Geography	History	Total
Biology	21.50	43	21.5	86
Chemistry	18.50	37	18.5	74
Physics	10.00	20	10 .0	40
Total	50.00	100	50 .0	200

This table agrees with the values computed by ratios shown in all the tables above.

Again, we need to answer the question: is the table of expected numbers of students in each cell relatively close to the original table? Or is the difference so great that the **expectation** of independence is **wrong** ?

Student view

Overview
(/study/ap
122-
cid-
754029/

Exam tip

From time to time , you will be required to compute expected frequencies using this formula in an exam paper.

Be aware

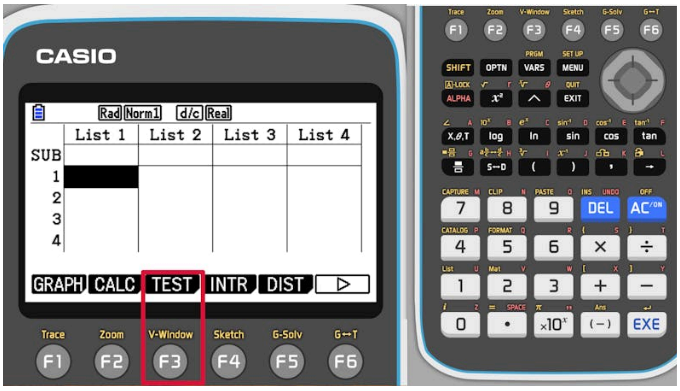
Each expected frequency must be at least 5 in order for the results of χ^2 tests to be reliable. This often requires a larger amount of data than you needed to test for correlation.

Expected frequencies using a calculator

Steps	Explanation
In these instructions you will see how to find the expected frequency matrix by calculator to confirm the manual calculation above. Open the statistics mode ...	

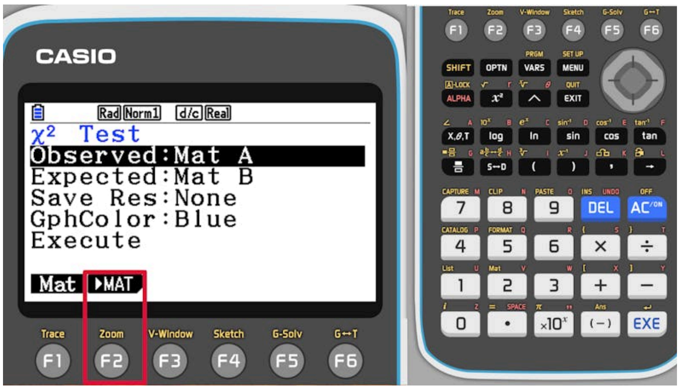
Student
view

Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
... press F3 to choose the test option ...	
... press F3 again for the χ^2 test ...	

Student
view

Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
... and finally press F2 (2WAY) for the χ^2 test for independence.	
You will need to tell the calculator the table of observed values. Press F2 to edit a matrix.	

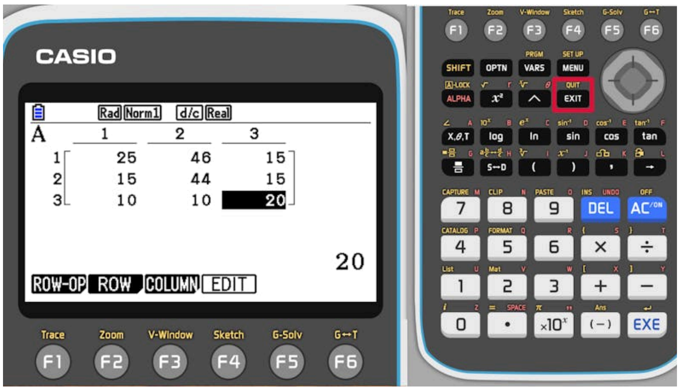
Student
view

Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
You can choose any matrix to store the observed frequencies. Press EXE to confirm your choice.	
Enter the dimensions of the data table ...	

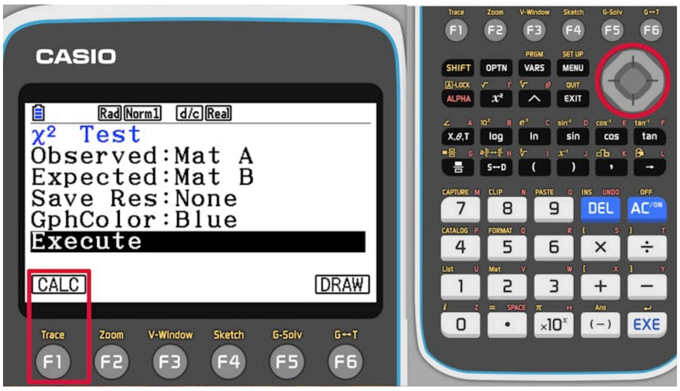
Student view

Overview
(/study/ap
122-
cid-
754029/(

Steps	Explanation
... and the data itself. Press EXIT when you are done.	
You can see that a 3×3 matrix is stored now in the memory. Press EXIT again to quit the data entering mode.	

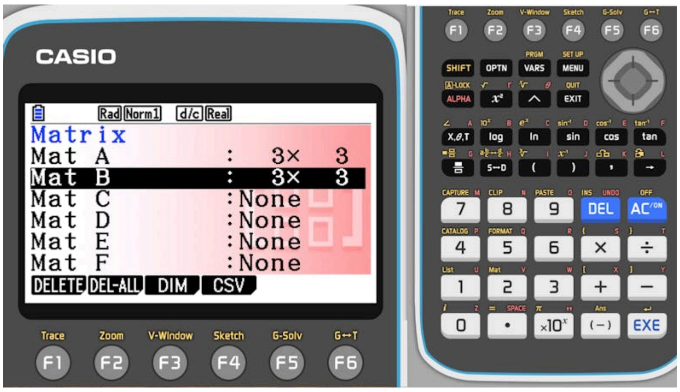
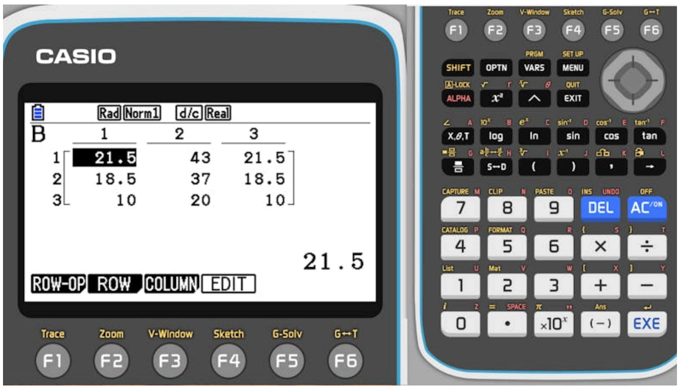
Student view

Overview
(/study/ap
122-
cid-
754029/

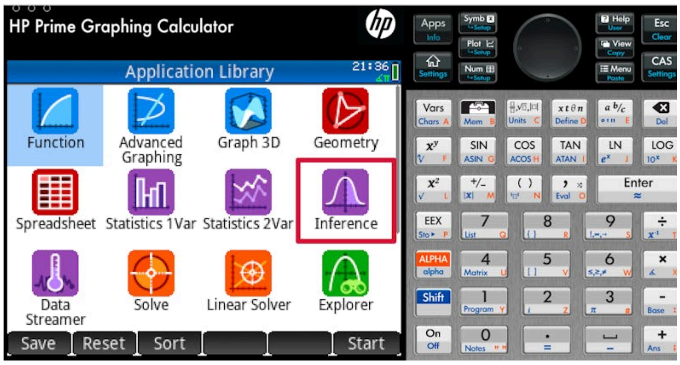
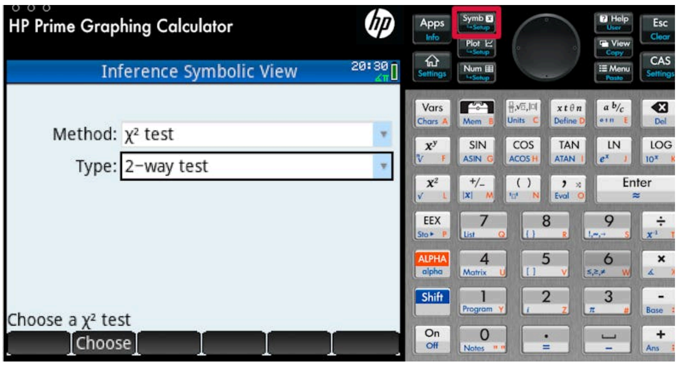
Steps	Explanation
On this screen you can see that the calculator expects the observed matrix in matrix A (change this if you stored it elsewhere) and will give the expected frequency matrix in matrix B. Move down to execute and press F1 to calculate the χ^2 statistics.	
On this screen you see some numbers. These will be important when you want to draw conclusions. You will see later in this section what these numbers mean. The goal now is to see the expected frequency matrix, so press F6.	

Student
view

Overview
(/study/ap
122-
cid-
754029/(

<table>
<tr><th>Steps</th><th>Explanation</th></tr>
<tr><td>Remember, the calculator stored the expected frequencies in matrix B, so choose that one.</td><td>
</td></tr>
<tr><td>These are the values to confirm the manual calculation.</td><td>
</td></tr>
</table>

Student
view

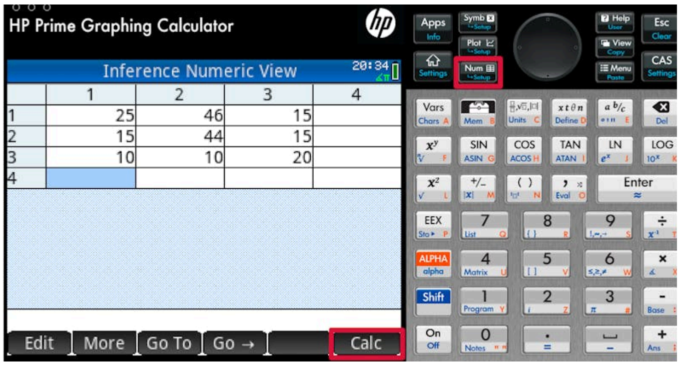
Steps	Explanation
In these instructions you will see how to find the expected frequency matrix by calculator to confirm the manual calculation above. Open the inference application ...	
--- and in symbolic view choose a 2-way χ^2 test to perform a χ^2 test for independence.	

<table>
<tr><th>Steps</th><th>Explanation</th></tr>
<tr><td>

In numeric view enter the data and when done, tap on Calc to calculate the χ^2 statistics.

The calculator uses the matrix called ObsMat to store this data. If you already have the data stored in another matrix in your calculator, on the home screen you can copy it with the store option to the ObsMat variable, and it will be displayed on this screen. No need to enter it twice.

</td><td>

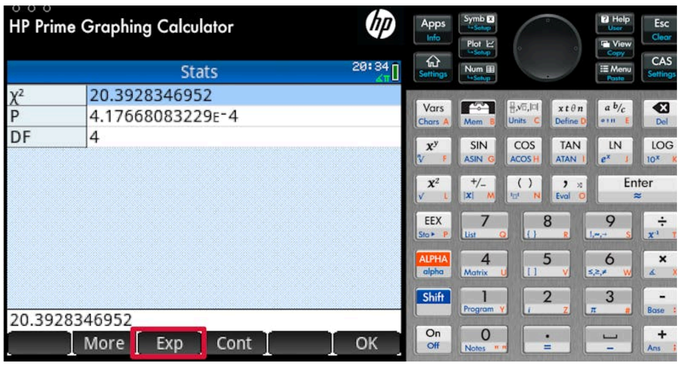


</td></tr>
<tr><td>

On this screen you see some numbers. These will be important when you want to draw conclusions. You will see later in this section what these numbers mean.

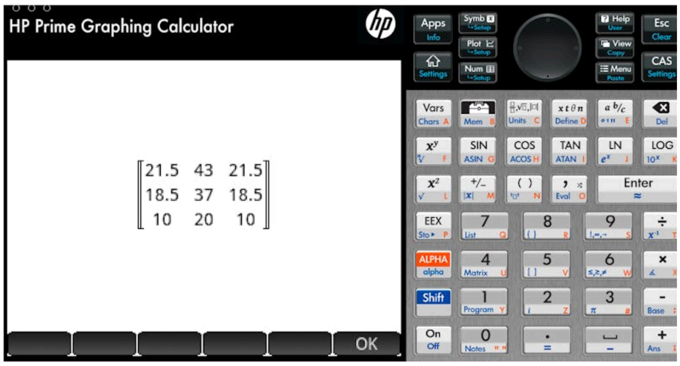
The goal now is to see the expected frequency matrix, so tap on Exp.

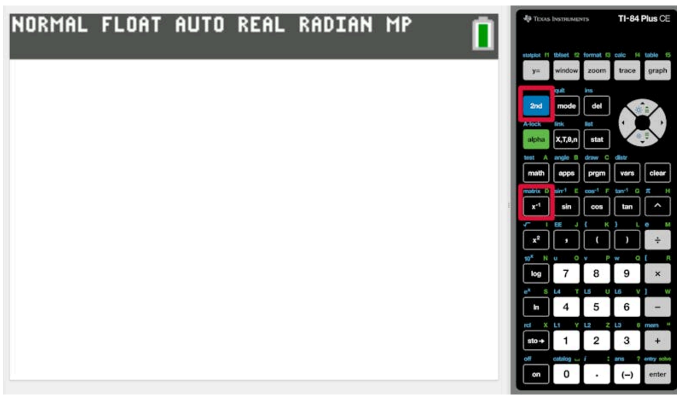
</td><td>



</td></tr>
</table>

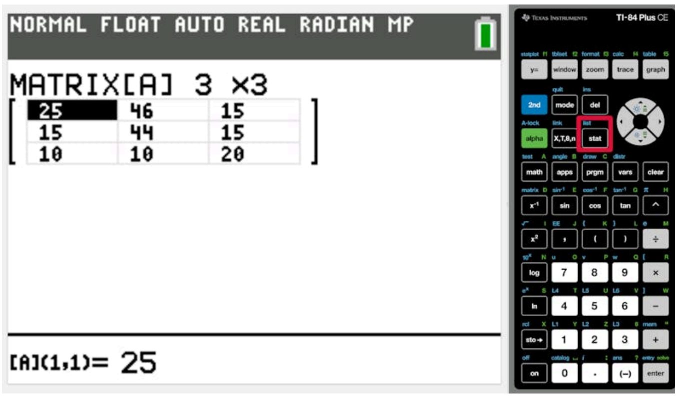
Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
These are the values to confirm the manual calculation. This matrix is stored in the ExpMat variable, you can view it any time on the home screen of any application.	

Steps	Explanation
In these instructions you will see how to find the expected frequency matrix by calculator to confirm the manual calculation above. First you need to enter the data in the calculator, so open the option to work with matrices ...	

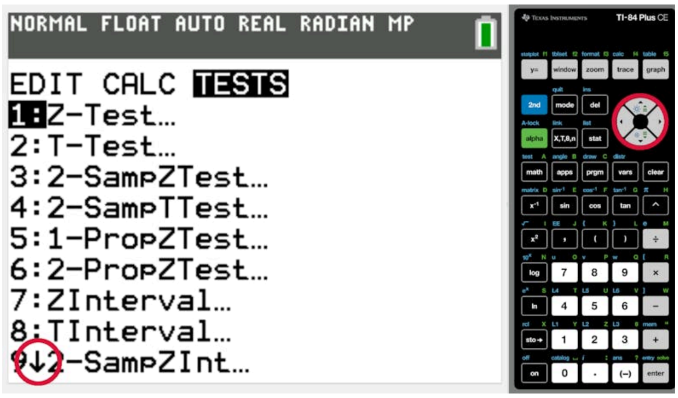
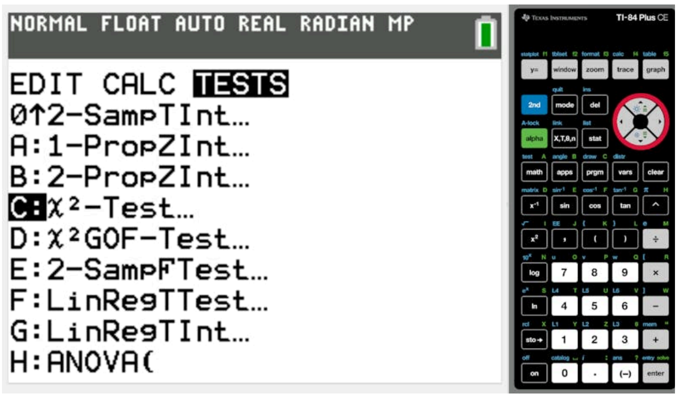
Student
view

Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
... and choose to edit a matrix. You can choose any matrix to store the observed frequencies.	
Enter the data and when done press stat to see the statistical calculation options.	

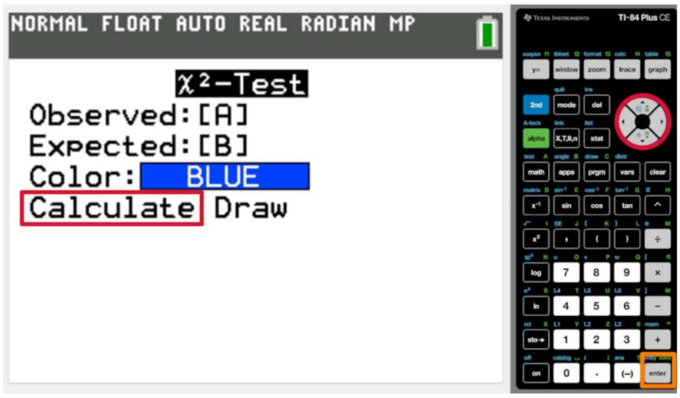
Student
view

Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
Select the test options and scroll down (the χ^2 tests are not among the first options) ...	
... and choose the χ^2 test (for independence).	

Student
view

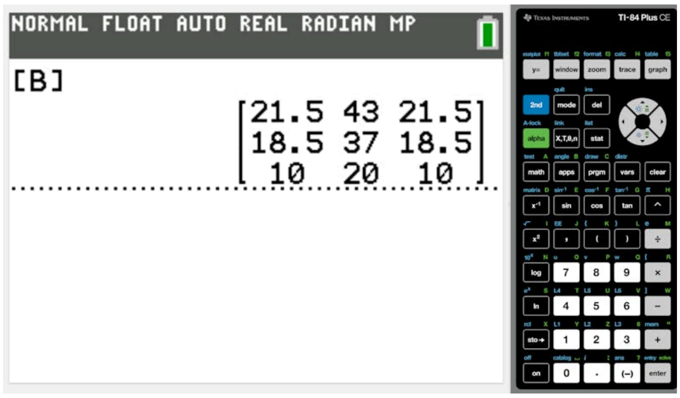
Overview
(/study/ap
122-
cid-
754029/

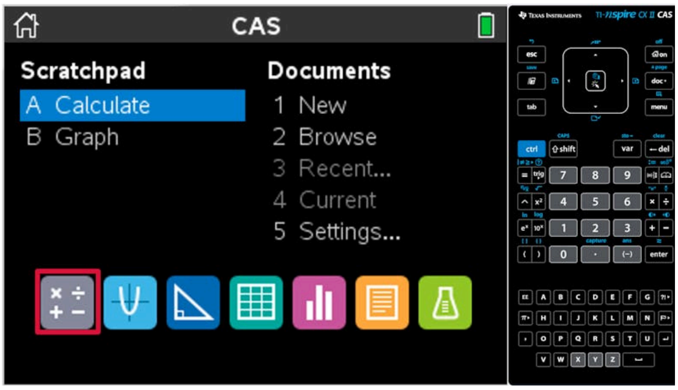
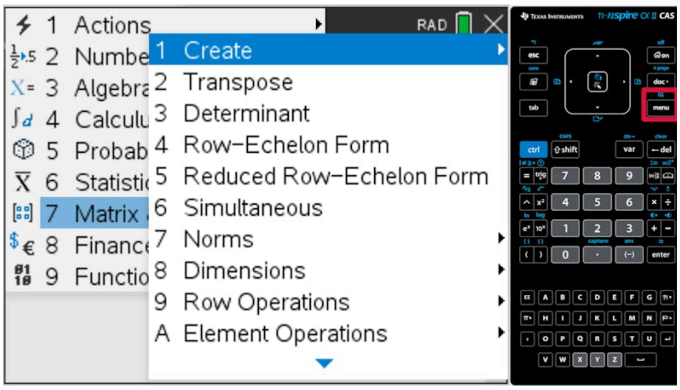
<table>
<tr><th>Steps</th><th>Explanation</th></tr>
<tr><td>
On this screen you can see that the calculator expects the observed matrix in matrix A (change this if you stored it elsewhere) and will give the expected frequency matrix in matrix B.
Move down to Calculate and press enter to calculate the χ^2 statistics.
</td><td>


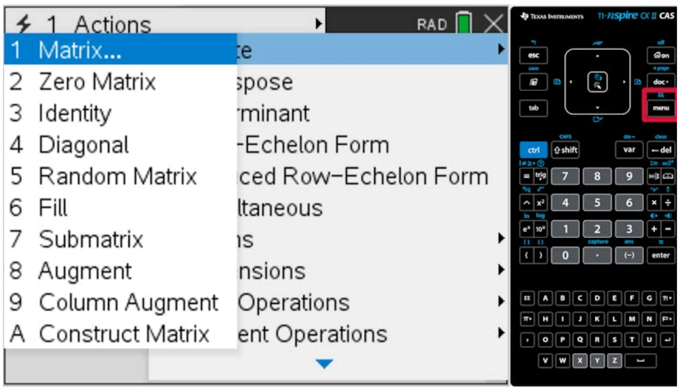
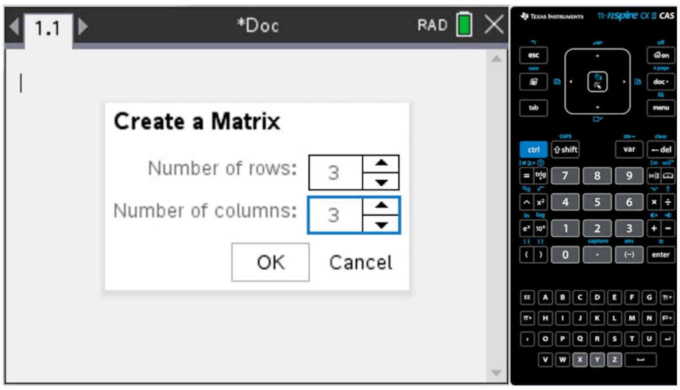
</td></tr>
<tr><td>
On this screen you see some numbers. These will be important when you want to draw conclusions. You will see later in this section what these numbers mean.
The goal now is to see the expected frequency matrix, so go back to the matrix options.
</td><td>


</td></tr>
</table>

Student
view

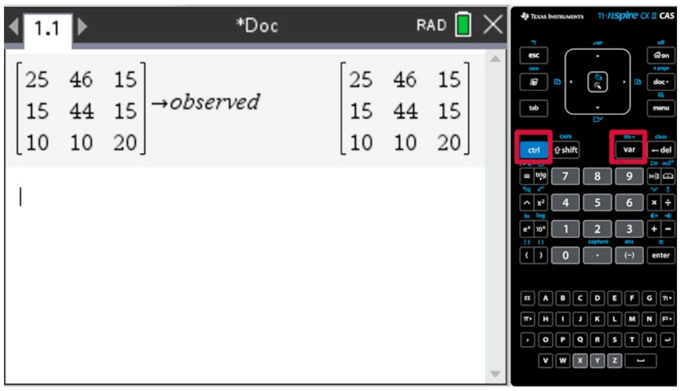
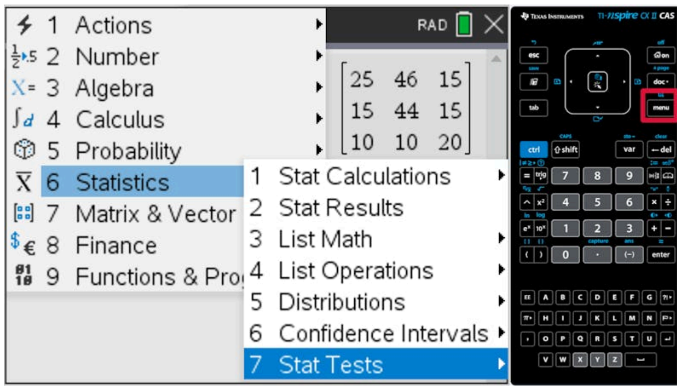
Steps	Explanation
Remember, the calculator stored the expected frequencies in matrix B, so choose that one.	
These are the values to confirm the manual calculation.	

Steps	Explanation
In these instructions you will see how to find the expected frequency matrix by calculator to confirm the manual calculation above. Open a calculator page.	
First you need to enter the data in the calculator, so open the option to work with matrices ...	

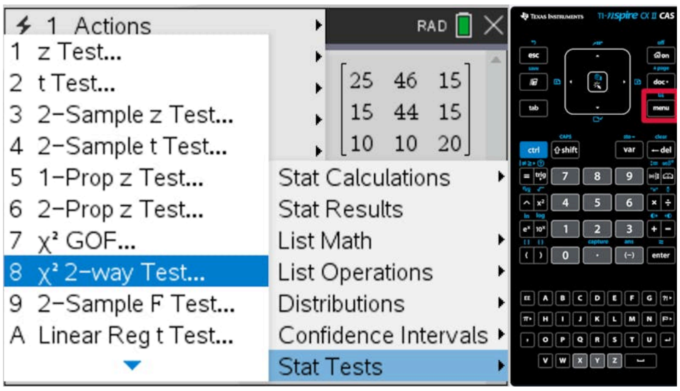
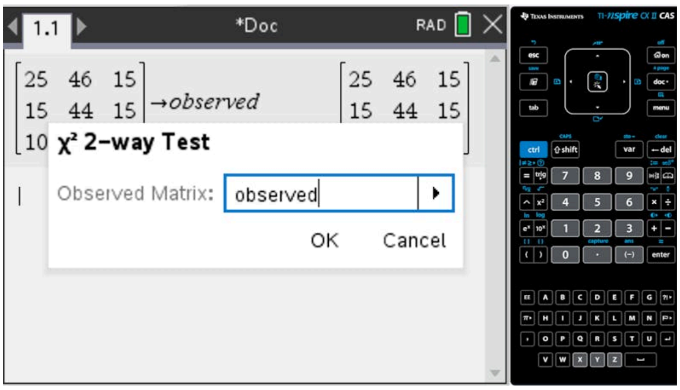
Steps	Explanation
... and choose to create a matrix.	
Enter the dimensions of the data table ...	

Overview
(/study/ap
122-
cid-
754029/

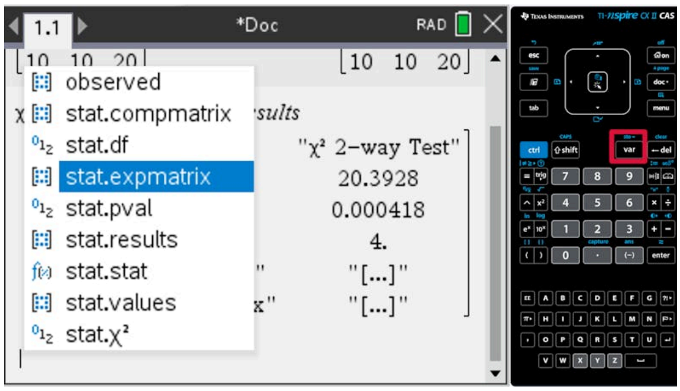
<table>
<tr><th>Steps</th><th>Explanation</th></tr>
<tr><td>... and enter the data itself.

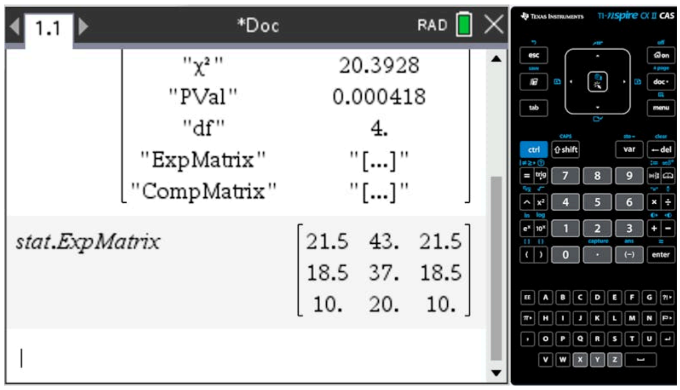
When you are done, store the matrix in the memory of the calculator. You can give any name (just make sure you remember it).</td><td>
</td></tr>
<tr><td>Once the data is stored in the calculator, open the menu, look for the statistical tests ...</td><td>
</td></tr>
</table>

Student view

Steps	Explanation
... and choose the χ^2 2-way test (for independence).	
Enter the name you gave for the data matrix and press OK to see the χ^2 statistics.	

Overview
(/study/ap
122-
cid-
754029/(

<table>
<tr><th>Steps</th><th>Explanation</th></tr>
<tr><td>
On this screen you see some numbers. These will be important when you want to draw conclusions. You will see later in this section what these numbers mean.
The goal now is to see the expected frequency matrix, so press vars and choose the variable where the expected frequency matrix is stored.
</td><td>


</td></tr>
<tr><td>
These are the values to confirm the manual calculation.
</td><td>


</td></tr>
</table>

We are now ready to calculate the error between the observed frequencies (the original data set) and the expected frequencies (those we have calculated in this section).

Student view

Overview (/study/app/m/sid-122-cid-754029/

Calculating the χ^2 statistic

Using the formula to find χ^2_{calc}

Important

The formula for calculating the test value χ^2_{calc} in the χ^2 test is

$$\chi^2_{\text{calc}} = \sum \frac{(\text{O} - \text{E})^2}{\text{E}}.$$

χ^2_{calc} can be calculated by using the formula and the lists on a calculator, by using a spreadsheet or by using the matrix on a calculator.

Before calculating χ^2_{calc} with the calculator, let's examine how to use the formula by calculating it for the data set regarding subject choices for Group 3 and Group 4 subjects. To do so, we can compile the data from the observed and expected frequencies already completed in the first and fourth tables, above, into the table below. Since we will need to examine each of the nine values in the table individually, we put them on separate rows in pairs.

Subject	Observed frequency	Expected frequency
Economics/Biology	25	21.5
Economics/Chemistry	15	18.5
Economics/Physics	10	10.0
Geography/Biology	46	43.0
Geography/Chemistry	44	37.0
Geography/Physics	10	20.0
History/Biology	15	21.5
History/Chemistry	15	18.5

Student view

Overview
(/study/ap
122-
cid-
754029/

Subject	Observed frequency	Expected frequency
History/Physics	20	10.0

Now for each of the nine rows, we find the difference $(\text{O} - \text{E})$, we square the difference $(\text{O} - \text{E})^2$, then we divide by the expected frequency $\frac{(\text{O} - \text{E})^2}{\text{E}}$. Finally, we add up the final column. The calculations are detailed below.

Subject	Observed frequency	Expected frequency	$(O - E)$	$(O - E)^2$	$\frac{(O - E)^2}{\mathbf{E}}$
Economics/Biology	25	21.5	3.5	12.25	0.5697674
Economics/Chemistry	15	18.5	−3.5	12.25	0.6621621
Economics/Physics	10	10.0	0	0	0
Geography/Biology	46	43.0	3.0	9.00	0.2093023
Geography/Chemistry	44	37.0	7.0	49.00	1.3243243
Geography/Physics	10	20.0	−10.0	100.00	5
History/Biology	15	21.5	−6.5	42.25	1.9651162
History/Chemistry	15	18.5	−3.5	12.25	0.6621621
History/Physics	20	10.0	10.0	100.00	10
Sum					20.392834

Student
view

Overview (/study/app/m/sid-122-cid-754029/

Using the calculator to find χ^2_{calc}

Note that the calculator method for computing the test value has already been demonstrated in detail earlier in this section. It appeared on the result screen, when the calculator was used to find the expected frequencies.

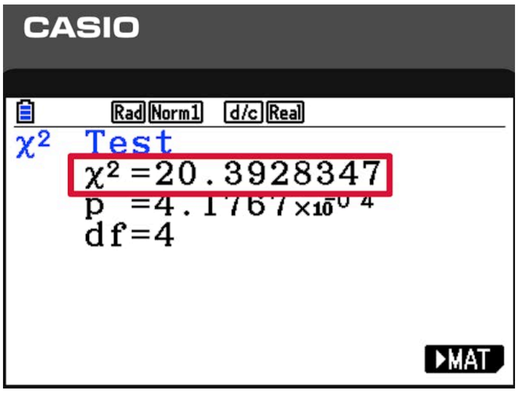
On this result screen the calculator reports three numbers.

- First, the χ^2 value is 20.4 (3 significant figures). This is the value we are looking for in this section.
- Second, the probability value (p -value) is $0.000\,418$ (3 significant figures). See section 4.11.3 (/study/app/m/sid-122-cid-754029/book/independence-and-goodness-of-fit-two-chi-squared-id-27591/) for more on the p -value.
- Finally, the number of degrees of freedom of this test is 4 . See section 4.11.3 (/study/app/m/sid-122-cid-754029/book/independence-and-goodness-of-fit-two-chi-squared-id-27591/) for more on degrees of freedom.

The result screens from the calculators

Student view

Steps	Explanation

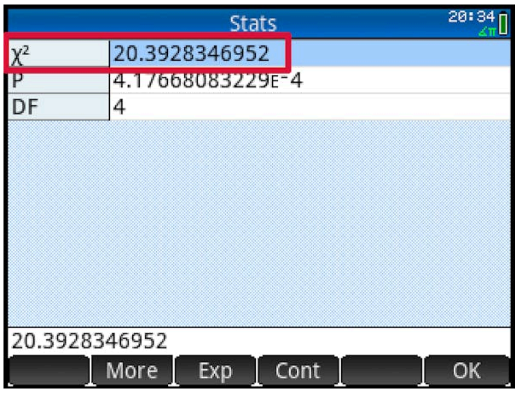


Casio fx-CG50

More information

The image is a screenshot from a Casio calculator display. It shows the results of a chi-squared test. There is a red box highlighting the main result: "χ² = 20.3928347." Below the highlighted text are additional results: "p = 4.1767 x 10^-4" and "df = 4." The top bar of the display includes menu options such as "Rad," "Norm1," "d/c," and "Real." There is also a "MAT" button at the bottom right of the display. The text is clearly legible, providing detailed statistical results from the calculator.

[Generated by AI]



HP Prime

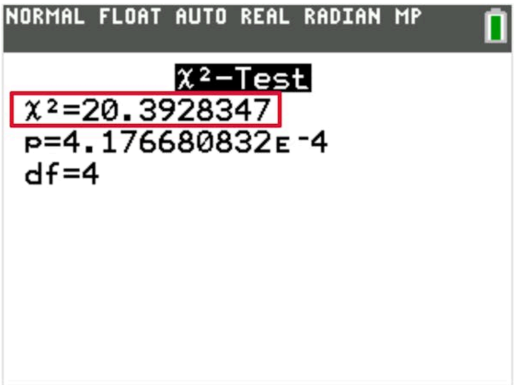
More information

The image shows a calculator screen with a "Stats" menu displaying statistical results. The top highlighted row shows "χ²" with a value of "20.3928346952." Below it, two more rows are visible, labeled "P" with a value "4.1768803229E-4" and "DF" with a "4." Below these rows, at the bottom of the screen, the same value "20.3928346952" is repeated. There are several buttons at the bottom labeled "More," "Exp," "Cont," and "OK," indicating more options for interaction with the statistics menu.

[Generated by AI]

Overview (/study/ap 122- cid- 754029/

Steps	Explanation



TI-84 plus CE

More information

A TI-84 calculator screen displays the results of a chi-squared test. The screen shows the following text:

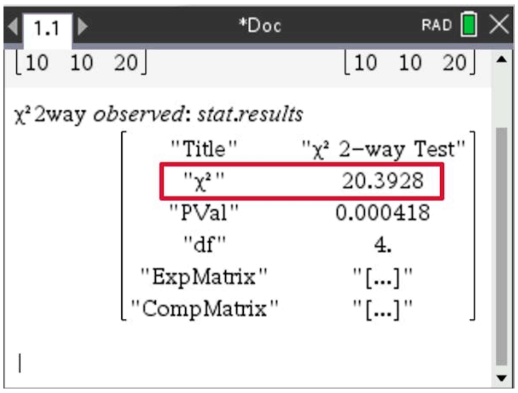
Top Bar: - NORMAL FLOAT AUTO REAL RADIAN MP

Main Display: - χ^2-Test (in bold within a black rectangle)

Results: - χ^2 = 20.3928347 - p = 4.176680832E-4 - df = 4

The heading indicates the type of test being performed, while the results give the chi-squared value (χ^2), the p-value (significance level), and the degrees of freedom (df).

[Generated by AI]



TI-nspire CX

More information

The image is a screenshot from a TI-nspire CX calculator displaying the results of a statistical test. The top text reads, 'x² 2way observed: stat.results'. Below this, there is a table with two columns.

The first column has the following labels: - "Title" - "x²" - "PVal" - "df" - "ExpMatrix" - "CompMatrix"

The second column, associated with a "x² 2-way Test," contains the respective values: - 20.3928 for "x²" - 0.000418 for "PVal" - 4 for "df"

The 'ExpMatrix' and 'CompMatrix' show placeholder values represented by '[...]'.

[Generated by AI]

3 section questions

4. Probability and statistics / 4.11 Hypothesis testing

Independence and goodness of fit – two chi-squared tests

Student view

Section Student... (0/0) Feedback Print (/study/app/m/sid-122-cid-754029/book/independence-and-goodness-of-fit- Assign

Overview (/study/app/m/sid-122-cid-754029/

The χ^2 test for independence

One application of the χ^2 statistic is to test whether two variables are independent . We will now use the four stages of the hypothesis test that we learned about in section 4.11.1 (/study/app/m/sid-122-cid-754029/book/hypothesis-testing-and-the-chi-squared-id-27588/) in order to test the data representing the choice of Group 3 and Group 4 subjects that we explored in section 4.11.2 (/study/app/m/sid-122-cid-754029/book/calculating-the-chi-squared-statistic-id-27589/).

We can test to see if the choice of Group 4 subject and the choice of Group 3 course for 200 IB Diploma Programme students are independent, or rather, if a preference for science goes alongside a shared preference for individuals and societies subject.

The data set showing student choices is shown in the table below.

	Economics	Geography	History	Total
Biology	25	46	15	86
Chemistry	15	44	15	74
Physics	10	10	20	40
Total	50	100	50	200

Do students who take a certain Group 4 course tend to have a preference for a certain Group 3 course?

We can test for independence between Group 3 and Group 4 courses using the χ^2 test. First, we need to state what our assumption is. *We put aside our belief that physics students take economics, and any other belief that we have.* We state that there is no association at all, that they are independent of each other.

Stage 1: The null and alternative hypotheses

The null hypothesis is given the notation H_0 , while the alternative hypothesis is given the notation H_1 . The null hypothesis is the statement that our two variables are independent. The alternative hypothesis is the statement that they are not independent. In our context:

Student view

Overview
(/study/ap
122-
cid-
754029/

- H_0 : The choice of Group 4 course and the choice of Group 3 course are independent.
- H_1 : The choice of Group 4 course and the choice of Group 3 course are not independent.

Remember that these hypotheses should be complementary, leaving no third option possible.

Stage 2: Calculate the test statistic

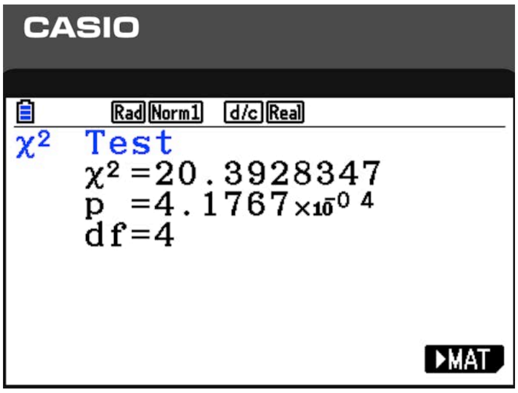
Use technology as described in section 4.11.2 (/study/app/m/sid-122-cid-754029/book/calculating-the-chi-squared-statistic-id-27589/) to calculate χ^2_{calc}.

The table above shows the results we found earlier.

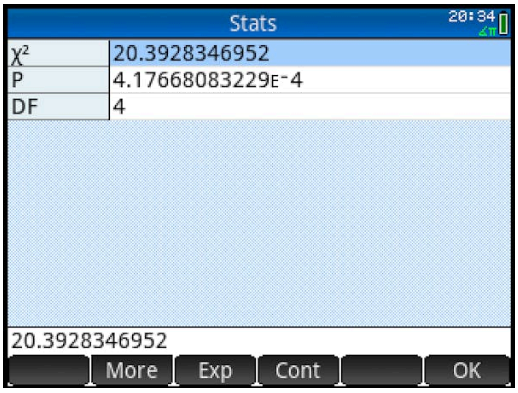
The result screens from the calculators

Student
view

Steps	Explanation



Casio fx-CG50



HP Prime

More information

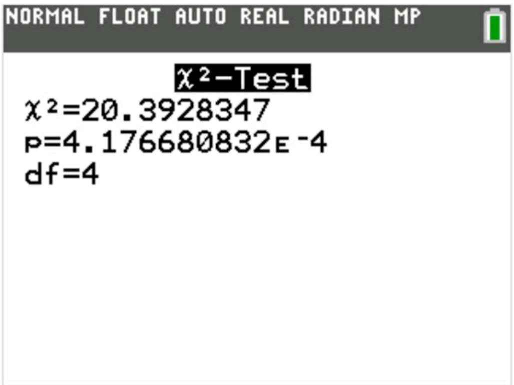
The image is a screenshot displaying a table with statistical data. It is titled "Stats" at the top. The table contains three rows with the following information:

- The first row has the label "χ^2" followed by the value "20.3928346952".
- The second row is labeled "P" with the value "4.17668083229e-4".
- The third row is labeled "DF" with the value "4".

Below the table, the value "20.3928346952" is repeated, and there are four buttons labeled "More," "Exp," "Cont," and "OK", indicating possible actions.

[Generated by AI]

Steps	Explanation



TI-84 plus CE

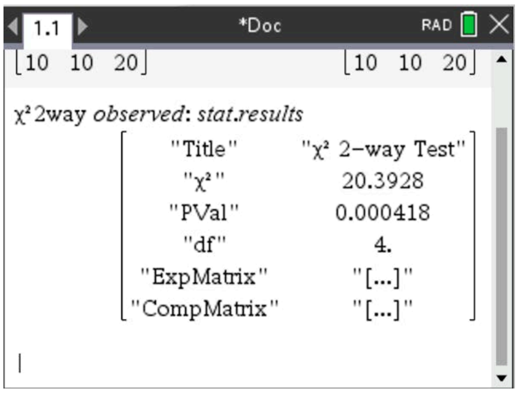
More information

The image shows the screen of a TI-84 Plus calculator. At the top of the screen, a menu displays options: "NORMAL FLOAT AUTO REAL RADIAN MP". Below, the main content includes the result of a chi-squared test, displayed as follows:

- "χ^2-Test" is shown highlighted.
- "χ^2 = 20.3928347"
- "p = 4.176680832E-4"
- "df = 4"

These lines indicate the chi-squared statistic value (χ^2), the p-value denoting significance, and the degrees of freedom (df). An icon resembling a battery is visible in the top right corner indicating the device's battery status.

[Generated by AI]



TI-nspire CX

More information

The image is a screenshot displaying statistical results on a TI-Nspire calculator. It shows the results of a chi-square test labeled as 'χ^2 2-way Test.' The text includes 'Title,' 'χ^2' with a value of 20.3928, 'PVal' with 0.000418, 'df' which is 4, and two additional elements labeled 'ExpMatrix' and 'CompMatrix' with unspecified values indicated by '[...]'.

[Generated by AI]

Stage 3: Find the p -value

Examining the output from the calculator, you can see that the p-value is given as $p = 0.000418$.

This means that, if the null hypothesis were true, you could look at the data for 10 000 groups of 200 students each and you would only expect to find four such groups with this strong an apparent correlation between their choice of subjects by chance.

Stage 4: State your conclusion

In order to come to a conclusion, we need to decide on the level of confidence, α, that we desire. We commonly use $\alpha = 0.05$.

Since $p < 0.05$, we reject H_0 and accept H_1: a student's choice of Group 4 subject is not independent of their choice of Group 3 subject.

Example 1

★★☆

The contingency table below shows the number of people in three age groups who like, dislike or have not seen a new movie.

	Like	Dislike	Haven't seen
10—16 years old	12	7	10
17—25 years old	19	5	8
26—35 years old	14	13	12

Use a χ^2 test for independence to determine at a 0.05 level of confidence whether someone's opinion of the movie is independent of their age.

Stage 1: State the hypotheses

H_0 : A person's opinion of the movie is independent of their age.

H_1 : A person's opinion of the movie is not independent of their age.

Stage 2: Find χ^2_{calc}

Using technology we get these results:

$$\chi^2 = 4.957584242$$
$$p = .2916773717$$
$$\mathrm{d}f = 4$$

Overview (/study/app/m/sid-122-cid-754029/)

We can see that $\chi^2_{\text{calc}} \approx 4.96$.

Stage 3: Find the p -value

Looking at the results again, we can see that $p \approx 0.292$.

Stage 4: State the conclusion

We are told to use $\alpha = 0.05$. Since $p > 0.05$, we fail to reject H_0. We conclude that it is likely that a person's opinion of the movie is independent of their age.

International Mindedness

The use of p -values and $\alpha = 0.05$ to determine statistical significance has been debated by researchers around the world for decades. In March 2019, *The American Statistician* , a statistics journal, devoted their entire issue to the topic (a summary can be found here (https://www.psychologytoday.com/us/blog/the-athletes-way/201903/rethinking-p-values-is-statistical-significance-useless)). Consider the different types of medical research done around the world. How could an arbitrary measure like the chosen α level affect strategies that researchers choose to pursue or discard? How could it impact whether government regulators approve or reject a new drug or treatment? With these implications in mind, how much of a difference is there really between $p = 0.051$ and $p = 0.049$?

Theory of Knowledge

This section on chi-squared tests can be examined from the TOK perspective to explore not only ethics (i.e. at what point unexpected results are deemed suspicious) but also reason. Much as reason allows us to rethink events and stories from the past and to analyse the world in front of us, reason also helps to dispel doubt for future predictions and ideas, and much of this is calculated.

Because it is obviously difficult to confirm the future, or to understand yet why something deemed unusual or significant might occur, a compelling exploration of reason can come from examining prior predictions of the future. One such audio file (from the American radio station National Public Radio, from 13 September 2013) is found here (http://www.npr.org/2013/09/13/219719902/what-predictions-from-1984-came-true) .

Student view

This audio file clearly raises real-life situation questions about confidence in predictability and about predictions made by Nicholas Negroponte at the onset of TED. In the audio from 1984, Negroponte challenges the idea of using a stylus to navigate a touchscreen, and, in 2007, Steve Jobs talks about these same advantages of touchscreen technology in the iPhone. This all seems so logical to us now. Why wouldn't we use our fingers to navigate a screen?

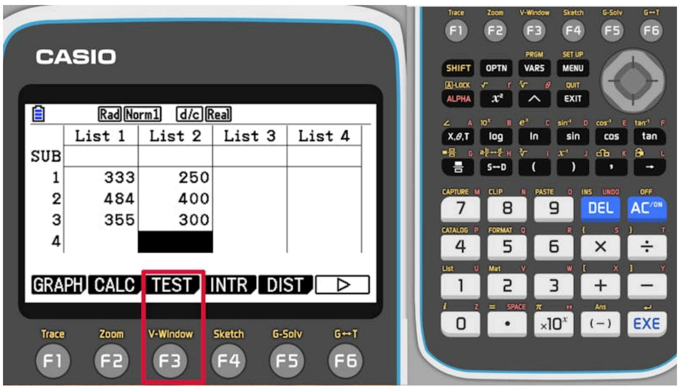
In other words, reason helps us feel confident in conclusions regarding the past in which we may have not participated, in events we have witness ed and in predictions we may never see. In IB courses, metacognition — thinking about your own thinking — is essential and is ongoing. Reason allows you to examine your thoughts, how you are thinking and, most interestingly, how you might be able to think differently despite the odds or potential statistical disconnections.

The χ^2 goodness of fit test

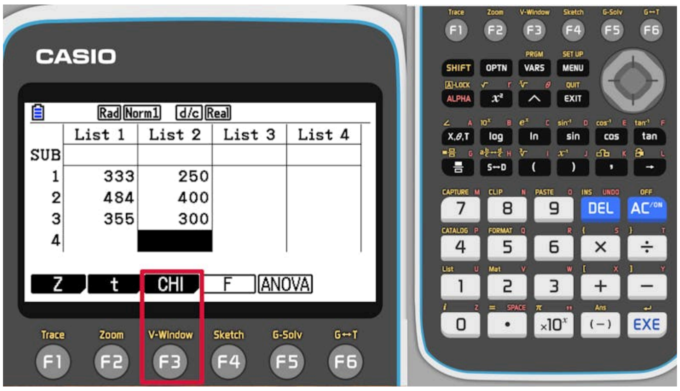
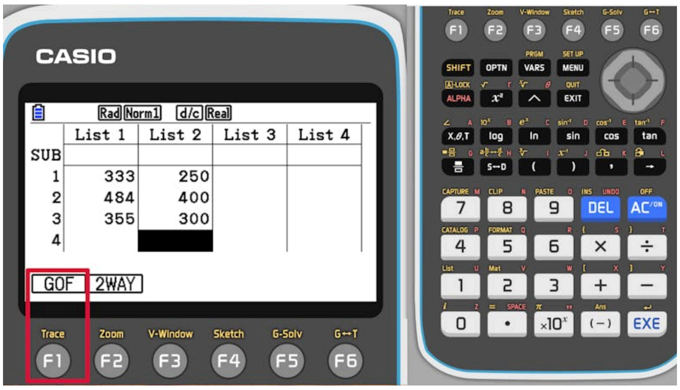
A second application of the χ^2 statistic is to test whether the observed distribution of a variable fits a given distribution. This fit does not have to be exact, but like the test for independence we can come to a conclusion with a certain degree of confidence by calculating the p-value. Note that the procedure we use to calculate the χ^2 statistic is the same, but the context of the problem and how we interpret the results changes. One other difference is in how you determine the degrees of freedom. On SL exams the degrees of freedom will always be $n-1$, where n represents the number of categories that the variable is divided into. For the purposes of this course, there will always be at least two degrees of freedom.

The following help shows you how to perform a χ^2 goodness of fit test on certain graphic display calculators.

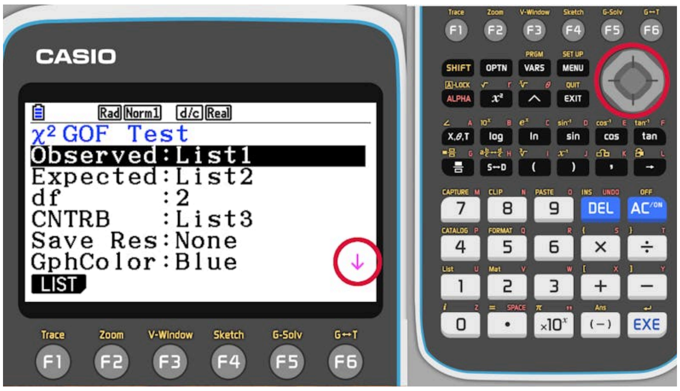
Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
To perform a χ^2 goodness of fit test, start with opening the statistics mode.	
Enter the data. In this example List1 contains the observed frequencies and List2 contains the expected frequencies. When done, press F3 to open the statistical test options ...	

Student
view

<table>
<tr><th>Steps</th><th>Explanation</th></tr>
<tr><td>... press F3 again to perform a χ^2 test ...</td><td>
</td></tr>
<tr><td>... and finally F1 for a χ^2 goodness of fit (GOF) test.</td><td>
</td></tr>
</table>

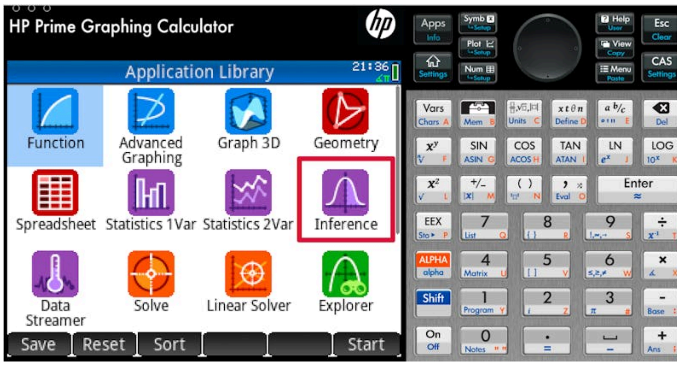
Overview (/study/ap 122- cid- 754029/

Steps	Explanation
Tell the calculator which list contains the observed and expected frequencies. Tell also the degrees of freedom (which in this example is $3 - 1 = 2$). Once done, scroll down ...	
... scroll down to the last line and press F1 to calculate the χ^2 statistics.	

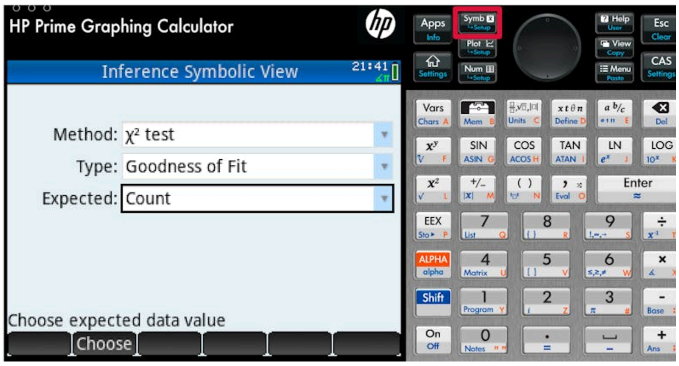
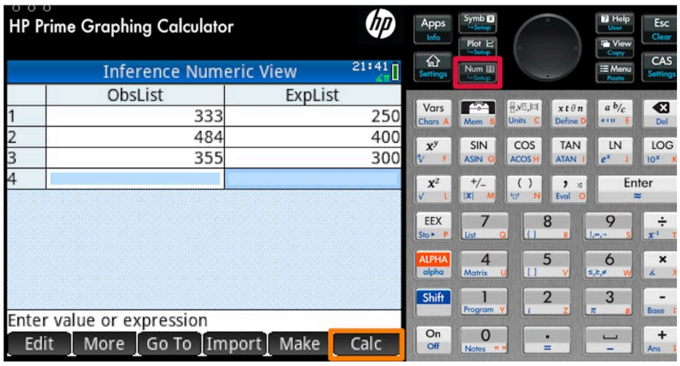
Student view

Overview
(/study/ap
122-
cid-
754029/

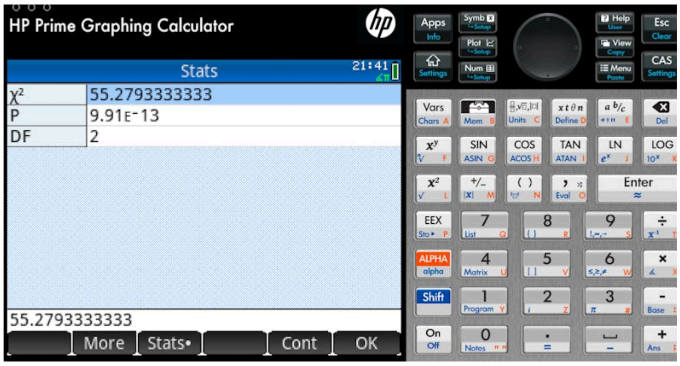
Steps	Explanation
The result screen tells you the χ^2-value and the p-value which you need to draw conclusion. Note, that in this case the calculator displays $p = 0$, which indicates that the p-value is closer to 0 than the accuracy of the calculator.	

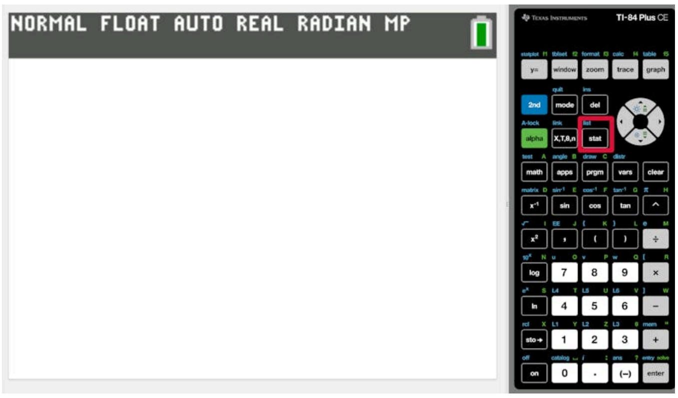
Steps	Explanation
To perform a χ^2 goodness of fit test, start with opening the inference application.	

Student view

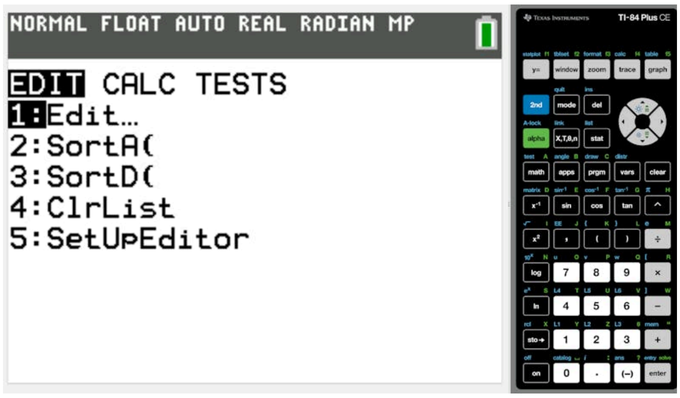
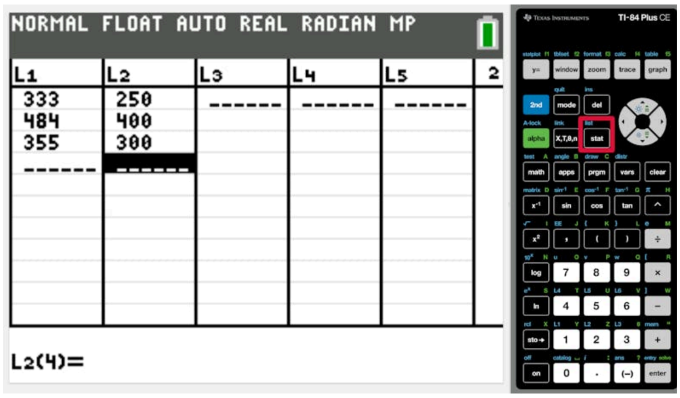
Steps	Explanation
In symbolic view, tell the calculator that you would like to perform a χ^2 goodness of fit test. Tell also, that you have a list of expected frequencies.	
In numeric view, enter the data. Once done, tap on calc to calculate the χ^2 statistics.	

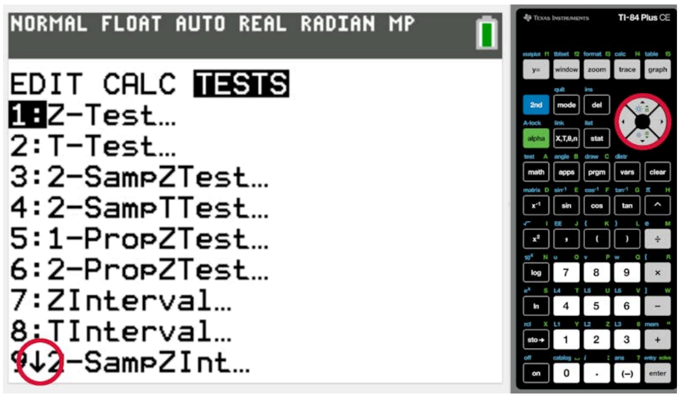
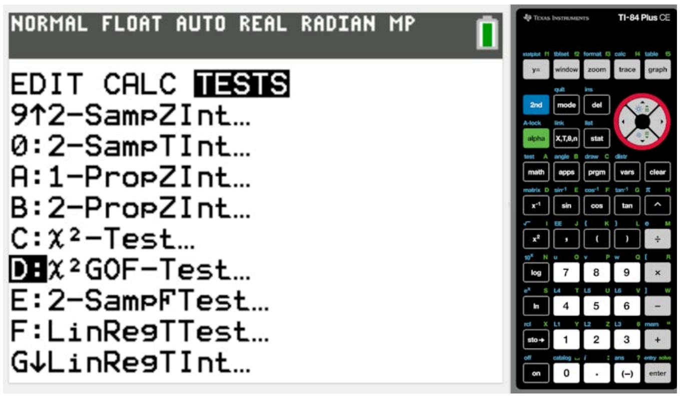
Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
The result screen tells you the χ^2-value and the p-value which you need to draw conclusion.	

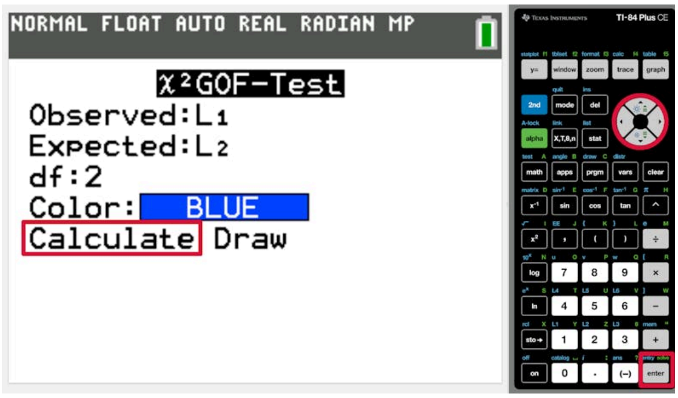
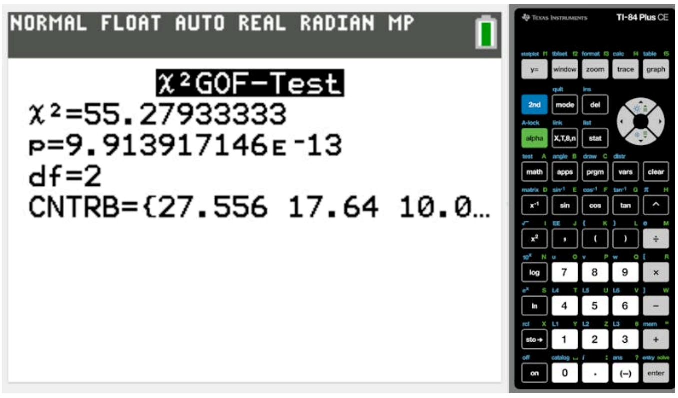
Steps	Explanation
To perform a χ^2 goodness of fit test, start with opening the statistics mode ...	

Student
view

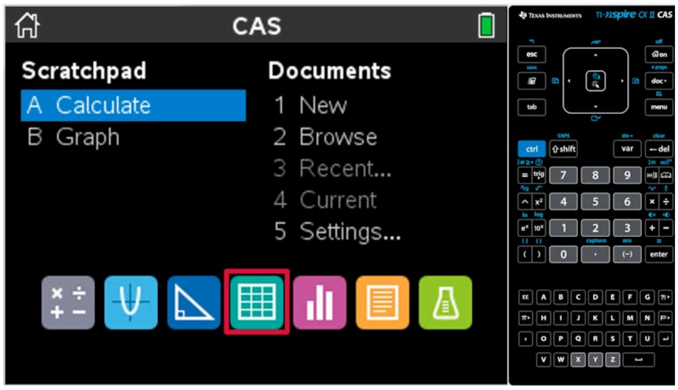
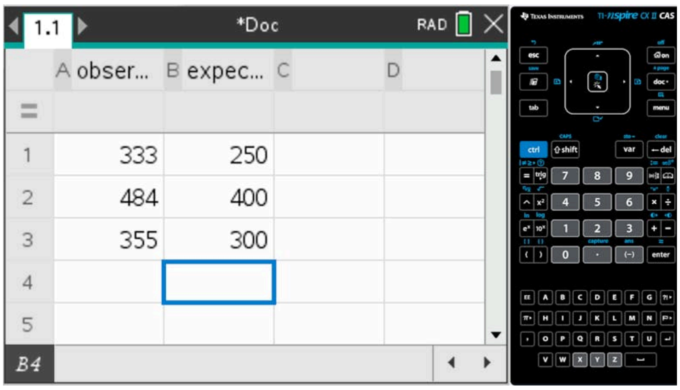
Steps	Explanation
... and choose to edit the lists.	
Enter the data. In this example L1 contains the observed frequencies and L2 contains the expected frequencies. When done, press stat to open the statistical test options ...	

Steps	Explanation
... move to the test options and scroll down ...	
... and choose the χ^2 goodness of fit (GOF) test.	

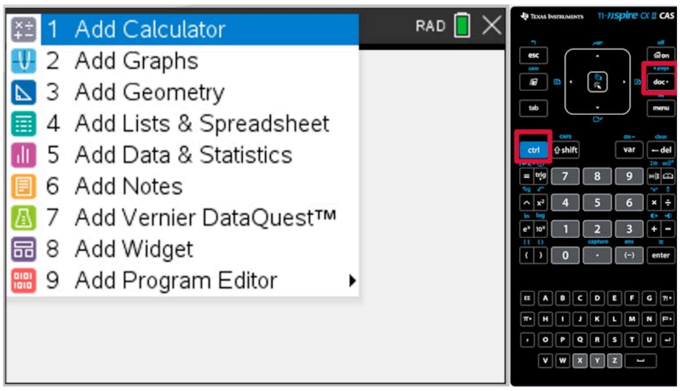
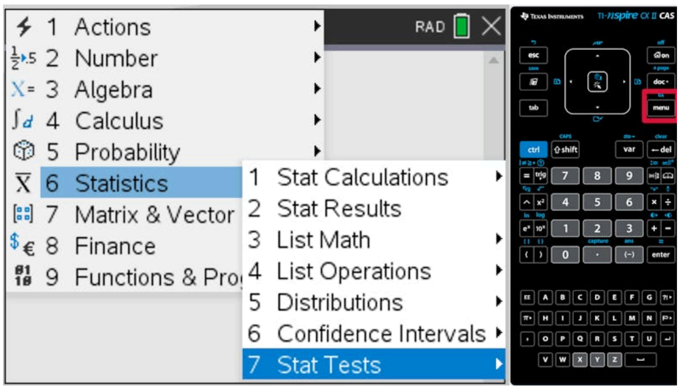
Student
view

Steps	Explanation
Tell the calculator which list contains the observed and expected frequencies. Tell also the degrees of freedom (which in this example is $3-1=2$). Once done, scroll down to calculate and press enter.	
The result screen tells you the χ^2-value and the p-value which you need to draw conclusion.	

Student view

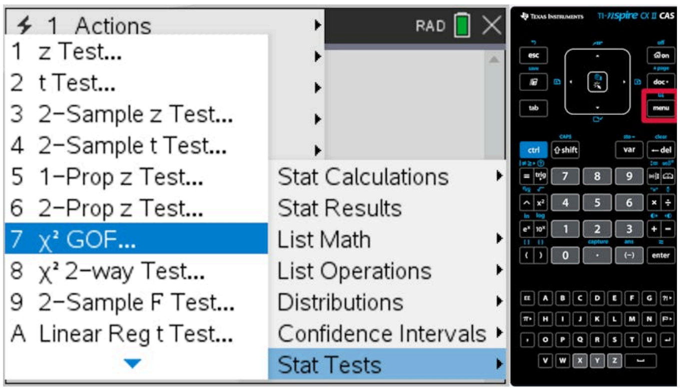
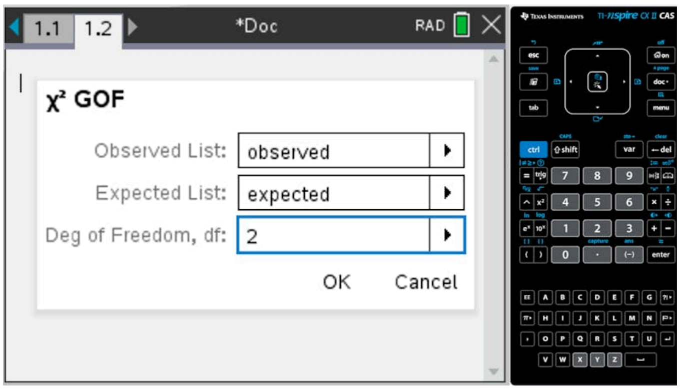
Steps	Explanation
To perform a χ^2 goodness of fit test, start with opening a spreadsheet page.	
Enter the data and give a name for the lists. You will need to remember these names, so use something meanngful.	

Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
Add a calculator page.	
On the calculator page use the menu to search for the statistical tests ...	

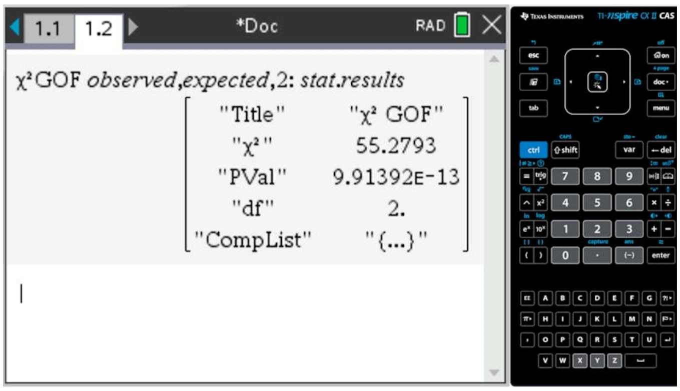
Student
view

Overview (/study/ap 122- cid- 754029/

Steps	Explanation
... and choose the χ^2 goodness of fit (GOF) test.	
Use the names you gave to your lists and specify the degrees of freedom (which in this example is $3 - 1 = 2$).	

Student view

Overview (/study/ap 122- cid- 754029/

Steps	Explanation
The result screen tells you the χ^2-value and the p-value which you need to draw conclusion.	

Example 2

★★☆

A company claims that in a certain bag of candies they produce 25% is red, 40% is green and 35% is yellow. You try to check this claim and buy a bag of candies. You find that there are 63 red, 72 green and 65 yellow candies in your bag.

Is your sample consistent with the distribution the company claims? Use a χ^2 goodness of fit test with a 0.05 level of significance.

To solve this problem, you can complete a hypothesis test.

Stage 1: The null and alternative hypotheses

First, assume that the company gave reliable information.

H_0 : The distribution of candies the company claims is correct.

Student view

Overview
(/study/ap
122-
cid-
754029/

Then state the opposite as the alternative hypothesis.

H_1: The distribution of candies the company claims is incorrect.

Stage 2: Calculate the test statistic

Let's collect the information in a table.

- In the first row, write the observations.
- In the second row, write the expected number of different colour candies based on the distribution the company claims. Since the distribution is given using percentages, you will need the total number of candies in your sample pack.

$$63 + 72 + 65 = 200$$

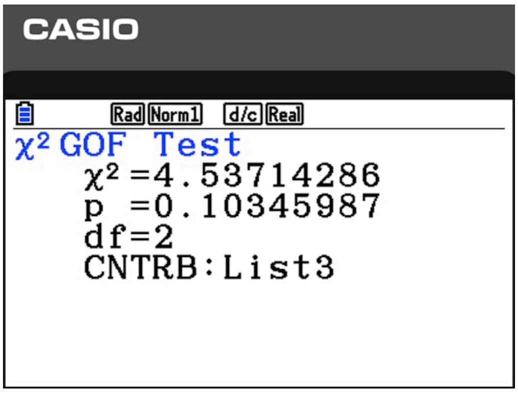
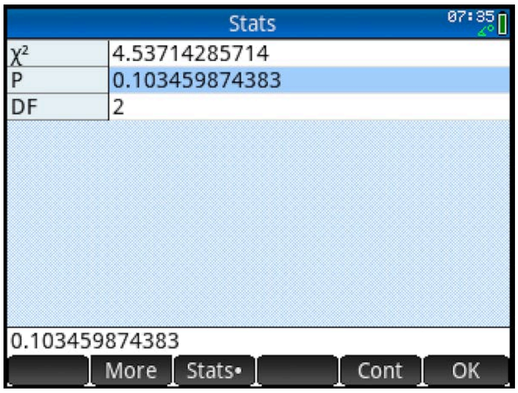
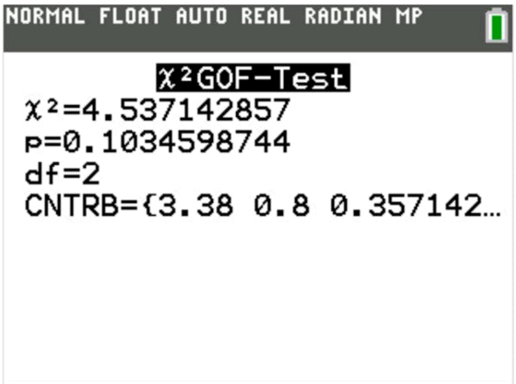
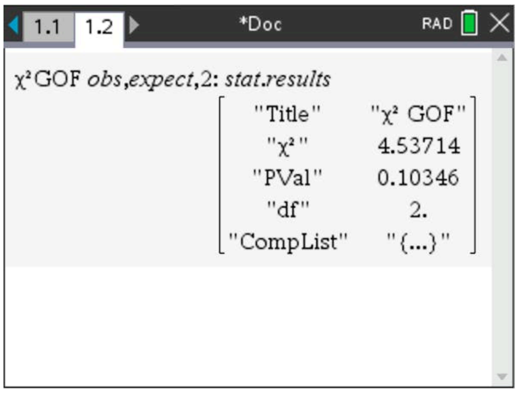
	Red	Green	Yellow
Observed	63	72	65
Expected	$0.25 \times 200 = 50$	$0.4 \times 200 = 80$	$0.35 \times 200 = 70$

Next, enter the data into the calculator and use the χ^2 GOF-Test function. Notice that since there are 3 type of candies, $n = 3$ and $df = 3 - 1 = 2$.

The result screens from the calculators

Student
view

Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
 Casio fx-CG50	 HP Prime
 TI-84 plus CE	 TI-nspire CX

Stage 3: Find the p -value

Based on the calculator output, $p \approx 0.103$.

Stage 4: State your conclusion

Since $p > 0.05$, we do not reject H_0. Even though this sample does not match the expected count, the difference can be because of chance. At the 0.05 level of significance, this sample is consistent with the distribution the company claims.

Student
view

Overview
(/study/app
122-
cid-
754029/

Activity

Try to predict the distribution of something that can be divided into at least three categories. For example, estimate what percentage of the 50 most popular songs comes from different genres. Then collect some data to determine the observed values. Conduct a χ^2 goodness of fit test to determine if your original prediction was valid or not.

Example 3

★★☆

Lotte likes to play board games and she often uses her favourite dice. To check if her dice is biased or not, she rolls it 100 times and records the result.

Result	1	2	3	4	5	6
Frequency	10	18	20	15	16	21

- Write an appropriate null and alternative hypothesis.
- Use the χ^2 goodness of fit test and the critical value 11.07 to state a conclusion.

Stage 1: The null and alternative hypotheses

First, assume that the dice is unbiased.

H_0 : All numbers come up with equal probability.

Then state the opposite as the alternative hypothesis.

H_1: The numbers do not come up with equal probability.

Stage 2: Calculate the test statistic

Let's collect the information in a table.

Student
view

- In the first row, write the observations.

Overview
(/study/ap
122-
cid-
754029/

- In the second row, write what Lotte expects to see if the dice is unbiased, so when the numbers come up with equal probability.

	1	2	3	4	5	6
Observed	10	18	20	15	16	21
Expected	$\frac{100}{6}$	$\frac{100}{6}$	$\frac{100}{6}$	$\frac{100}{6}$	$\frac{100}{6}$	$\frac{100}{6}$

Use your calculator to perform a χ^2 goodness of fit test. Since the table has 6 coloumns, the degree of freedom is $6-1=5$.

Stage 3: Find the χ^2 -value

On the outputs screen the calculator tells you the p-value and the χ^2-value. Since the significance level is not given in this question, the p-value is not useful in this case. You need to draw conclusion based on the χ^2-value.

$$\chi^2 = 4.76$$

Stage 4: State your conclusion

For the conclusion you need to compare this χ^2-value with the given critical value, 11.07. Since $4.76 < 11.07$, we do not reject H_0. Even though this sample does not match the expected count, the difference can be because of chance. Based on this sample, Lotte does not have enough evidence to doubt that her dice is fair.

Example 4

★★★

Joe wants to be a professional basketball player and practicing every day. On each day he finishes his practice with three free throws and records his results for a year. The table below contains the data.

Student
view

Overview
(/study/ap
122-
cid-
754029/

	Missed all three free throws	Missed two and made one free throws	Missed one and made two free throws	Made all three free throws
Frequencies	6	51	153	155

Joe thinks that on average he makes 70% of the free throws and that the success of each attempt is independent of the other attempts.

Is this data consistent with this assumption? Use an appropriate hypothesis test with a 0.05 level of significance.

Stage 1: The null and alternative hypotheses

The null hypothesis is that Joe's assumption is correct. Since Joe thinks that the attempts are independent, this means that the number of made free throws follow a binomial distribution.

H_0 : The number of made free throws follow a binomial distribution with $n = 3$ number of trials and $p = 0.7$ probability of success.

Then state the opposite as the alternative hypothesis.

H_1: The number of made free throws either does not follow a binomial distribution or the probability of success is not 0.7.

Stage 2: Calculate the test statistic

Let's collect the information in a table.

- In the first row, write the observations.
- In the second row, write the expected probabilities based on the binomial distribution assumption. You can find these probabilities using your graphic display calculator. If you need a reminder on how to do this, take a look at the help in Section 4.8.2.
- In the third row, multiply your numbers in the second row by 365 to find the expected counts.

Student
view

Overview
(/study/ap
122-
cid-
754029/

	Missed all three free throws	Missed two and made one free throws	Missed one and made two free throws	Made all three free throws
Observed count	6	51	153	155
Expected probability	0.027	0.189	0.441	0.343
Expected count	9.855	68.985	160.97	125.2

Use the first and third rows to perform a χ^2 goodness of fit test. Since the table has 4 coloumns, the degree of freedom is $4 - 1 = 3$.

Stage 3: Find the p -value

On the output screen the calculator tells you the p-value and the χ^2-value. Since the significance level is given in this question, you can draw conclusion based on the p-value.

$$p \approx 0.00336$$

Stage 4: State your conclusion

For the conclusion you need to compare this p-value with the given significance level, 0.05. Since $0.00336 < 0.05$, we reject H_0. It is unlikely that Joe got this distribution for the full year if his assumption is correct. It can happen that his attempts are not independent of each other (for example missing a free throw puts a pressure on him for the next attempts) or his percentage is not constant 70% (for example it is improving with practice during the year).

International Mindedness

In all our examples we investigated how probable it is if the χ^2 statistic is too high. This is what the p-value you get with your calculator expresses. We are usually interested in how likely it is to get a sample far away from the expected distribution. This is what you will be expected to explore on your IB exam.

Student
view

Overview
(/study/app/m/sid-122-cid-754029/

However, sometimes it can also be a valid question to ask how probable it is to get a data too close to the expectation.

An example of this is related to the data published in 1866 by Gregor Mendel who is famous for discovering the basic underlying principles of heredity through his work on garden pea plants (*Pisum sativum*). Statisticians, for example Sir Ronald Aymler Fisher in his paper published in 1936, investigated how close Mendel's published data is to the expectation. They found that getting this close match is very unlikely.

4 section questions

4. Probability and statistics / 4.11 Hypothesis testing

The t-test

Section Student... (0/0) Feedback Print (/study/app/m/sid-122-cid-754029/book/the-t-test-id-27592/print/) Assign

The t-test is a statistical method of comparing the means of two groups of data. There are a few different types of t-test, but we will be focusing only on the pooled two-sample t-test. This test is a comparison of the means of **two independent sets of data** that are samples selected from **a normally-distributed population** . It is called a pooled t-test, because it is also assumed that the **variance of the two populations is the same unknown value.**

Suppose you have two sets of data that are independent and come from normally distributed populations with equal variance. If the mean of the first set is μ_1 and the mean of the second set is μ_2, and you want to determine whether the means are the same or different, then the null hypothesis would be $\mathrm{H}_0\colon \mu_1 = \mu_2$. However, you might be interested in whether the mean μ_2 is greater or less than the mean μ_1. For this reason, the alternative hypotheses can vary depending on the context.

- For two-tailed tests the alternative hypothesis is $\mathrm{H}_1\colon \mu_1 \neq \mu_2$.
- For one-tailed tests the alternative hypothesis is $\mathrm{H}_1\colon \mu_1 < \mu_2$ or $\mathrm{H}_1\colon \mu_1 > \mu_2$.

The first possibility is known as a two-tailed test, because it is testing whether the mean of the first set is significantly different from the mean of the second set on either side. The latter two possibilities are known as one-tailed tests, because they are testing whether the mean of the first set is on one side or the other of the mean of the second set.

Student view

Overview
(/study/ap
122-
cid-
754029/

The idea of the t-test is similar to other hypothesis tests. Assuming that the conditions stated above and the null-hypothesis are true, you ask the calculator to to tell you how likely it is to get a data as extreme as you have in your sample. The calculator returns this probability as a p-value. The reason it is called a t-test is that the calculator uses a test statistics that follows a so-called t-distribution to determine this probability.

Example 1

One common example is when testing a new treatment for a medical condition. Suppose you want to determine if a new antibiotic cream will help heal a cut faster than the existing treatment will.

Does the new antibiotic cream help cuts heal faster?

Credit: Creative-Touch Getty Images

We give the first group of people the new cream and count how many days it takes for a cut to heal, giving us μ_1, and we give another group of people the old cream and count how many days it takes for a cut to heal (μ_2). Since we are dealing with two different groups of people, we can assume that the sets are independent. In addition, we want to know if the new cream works *faster* , so we would complete a one-tail test with the alternative hypothesis H_1: $\mu_1 < \mu_2$.

Now that we understand what kind of test to run, let's consider the data for the two groups and perform a hypothesis test to determine if the new cream decreased the healing time of a cut with a 0.05 level of significance.

Student
view

New cream: $3, 5, 4, 6, 6, 5, 3, 2, 3, 4, 5, 3, 4$

Old cream: $4, 6, 6, 7, 6, 4, 4, 4, 3, 6, 5, 4, 5$

Stage 1: State the null and alternative hypotheses

$\mathrm{H}_0: \mu_1 = \mu_2$

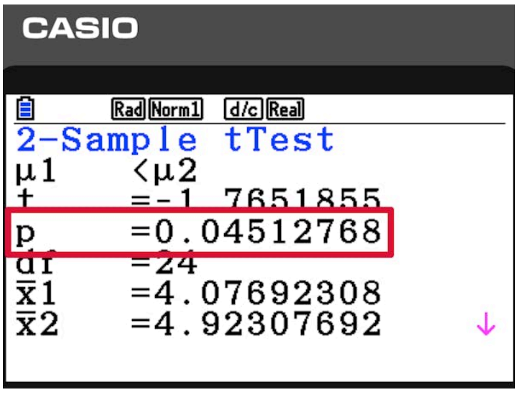
$\mathrm{H}_1: \mu_1 < \mu_2$

Stage 2: Calculate the p-value

Select the calculator you are using (below the solution box). This will show you how to run a two-sample t-test on your calculator.

The result screens from the calculators

Steps	Explanation

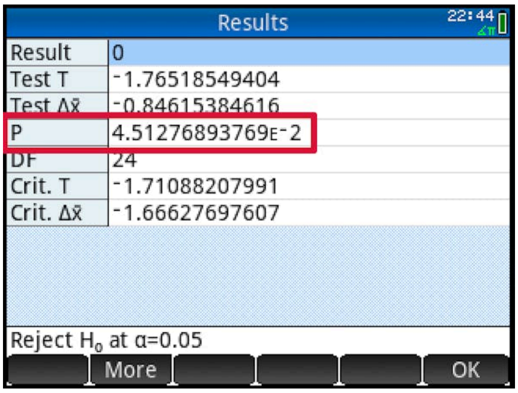


Casio fx-CG50

More information

The image shows a calculator screen displaying results from a 2-Sample t-Test. The text on the screen includes statistical test results with a focus on the p-value. Specifically, it shows the following: - μ1 < μ2 - p-value highlighted in red as: p = 0.04512768 - Degrees of freedom (df) = 24 - Means of the samples displayed as x̄1 = 4.07692308 and x̄2 = 4.92307692. The image seems to be from a Casio calculator, indicated by the branding on the top left.

[Generated by AI]



HP Prime

More information

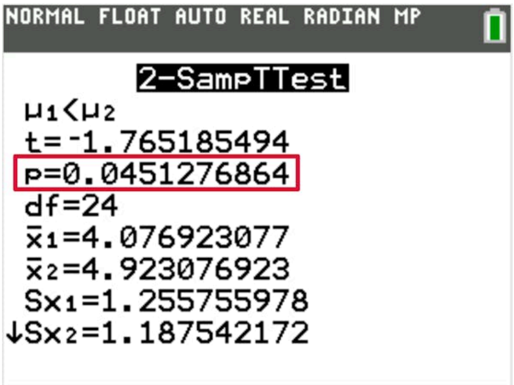
The image is a screenshot of a calculator screen showing statistical analysis results. The top of the screen has a header with the word "Results." Below the header, there are several rows with statistical data:

1. **Result**: On the first row, the label is "Result" followed by "0."
2. **Test T**: The second row shows "Test T" with a value of "-1.76518549404."
3. **Test Δx̄**: The third row states "Test Δx̄" with "-0.84615384616."
4. **P**: Highlighted in red, the fourth row reads "P" with a value of "4.51276893769e-2."
5. **DF**: On the fifth row, "DF" is followed by "24."
6. **Crit. T**: The sixth row lists "Crit. T" at "-1.71088207991."
7. **Crit. Δx̄**: The last row notes "Crit. Δx̄" at "-1.66627697607."

Below the data, the text "Reject H_0 at α=0.05" indicates a decision related to hypothesis testing. The bottom of the screen shows navigation buttons labeled "More" and "OK."

[Generated by AI]

Steps	Explanation

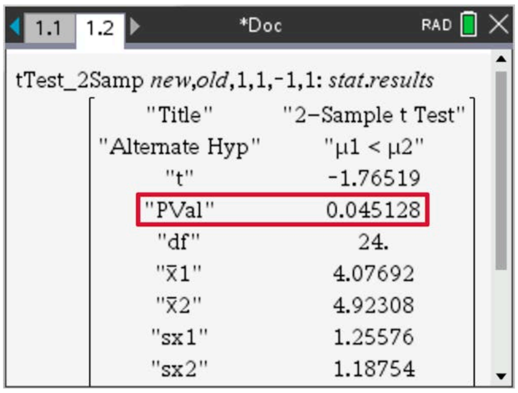


TI-84 plus CE

More information

The image shows a calculator display with a 2-SampTTest result. The display has several statistical values listed: (\mu_1 < \mu_2), (t = -1.765185494), (p = 0.0451276864) outlined in red, (df = 24). Additional statistics include (x_1 = 4.076923077), (x_2 = 4.923076923), (Sx_1 = 1.255755978), and (\downarrow Sx_2 = 1.187542172). These values relate to a two-sample t-test used in statistics to compare the means of two groups. The calculator display also shows symbols indicating the mode settings such as NORMAL, FLOAT, AUTO, REAL, RADIAN, and MP at the top, and a battery icon is present in the upper right corner.

[Generated by AI]



TI-nspire CX

More information

This image is a screenshot of a calculator display showing the results of a two-sample t-test. The results are contained within a table format. It lists parameters and their corresponding values as follows: - Title: "2-Sample t Test." - Alternate Hypothesis: "μ1 < μ2." - t-value: -1.76519. - P-Value (PVval): 0.045128 (highlighted in red). - Degrees of Freedom (df): 24. - Mean of sample 1 (x̄1): 4.07692. - Mean of sample 2 (x̄2): 4.92308. - Standard deviation of sample 1 (sx1): 1.25576. - Standard deviation of sample 2 (sx2): 1.18754.

This display is typical of statistical software outputs used in hypothesis testing. The red highlight emphasizes the p-value, which is typically a key value in determining statistical significance.

[Generated by AI]

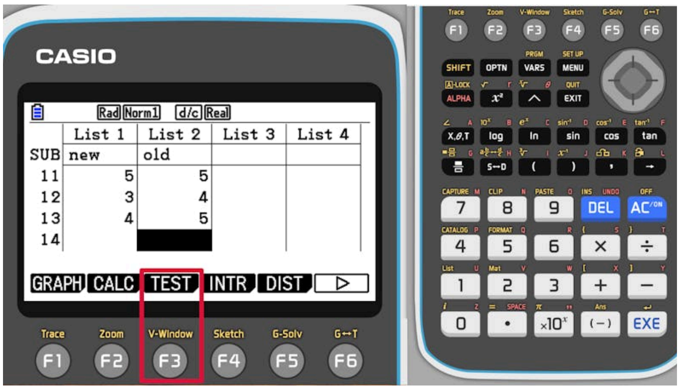
According to the calculator output, $p \approx 0.0451$.

Stage 3: State your conclusion

You were told to use a 0.05 level of significance, so $\alpha = 0.05$.

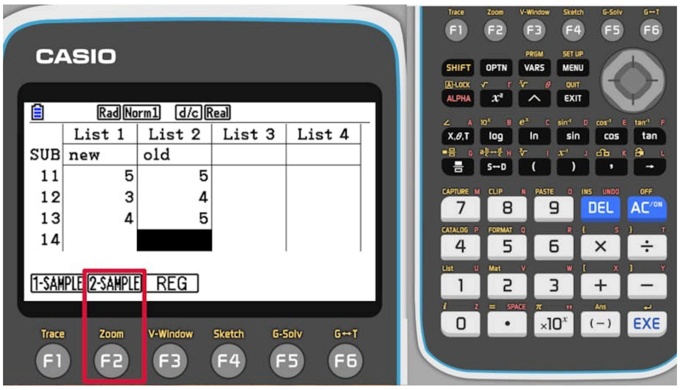
Since $p < 0.05$, we reject the null hypothesis and determine that μ_1 is significantly less than μ_2, meaning the new cream does significantly decrease the average healing time of a cut compared with the old cream.

Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
To perform a t-test, open the statistics option.	
Enter the data. In the context of this example, it is important to remember which list contains the data for the new and the old treatment. Once the data is stored, press F3 to access the statistical tests ...	

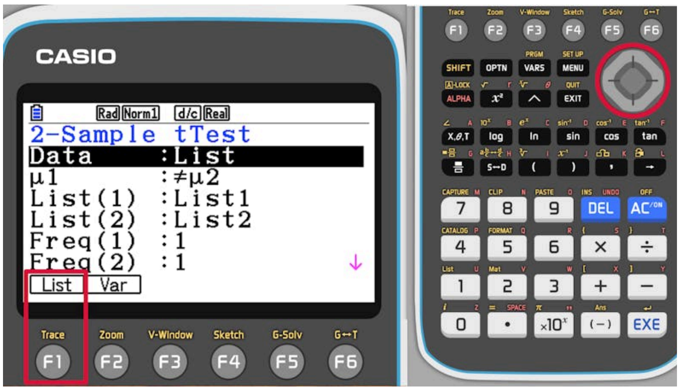
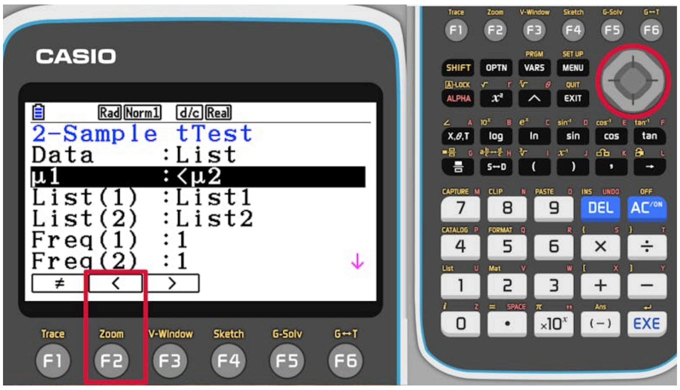
Student
view

Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
... press F2 for the t-test ...	
... and choose the 2-sample t-test.	

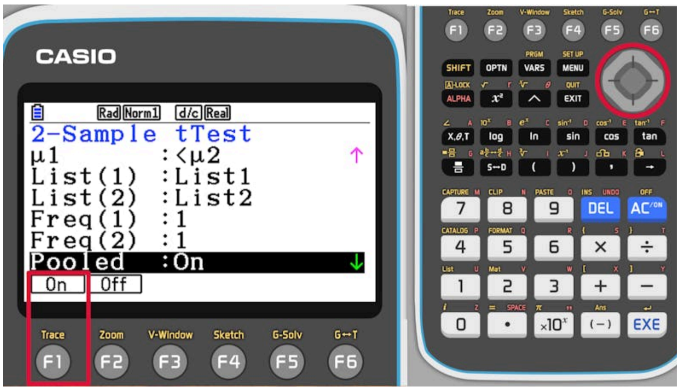
Student
view

Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
You will need to tell the calculator where the data is and some parameters of the test. Press F1 to choose the option that you have the data as a list. Then move down to the next line.	
In this example you do a one-tailed test. Since the new treatment data is stored in the first list and you want to see wheter it reduces healing time or not, press F2 to choose $\mu_1 < \mu_2$. Once done, move down to the next rows and tell the calculator where the data is. Leave the frequencies as 1.	

Student view

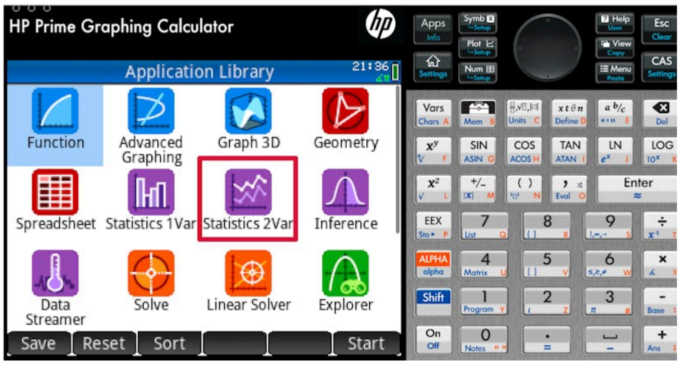
Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
Scroll further down and choose the option of a pooled test. The IB syllabus specifies, that in this course only pooled tests will be used.	
Once every detail is set, scroll further down and press F1 to run the t-test.	

Student
view

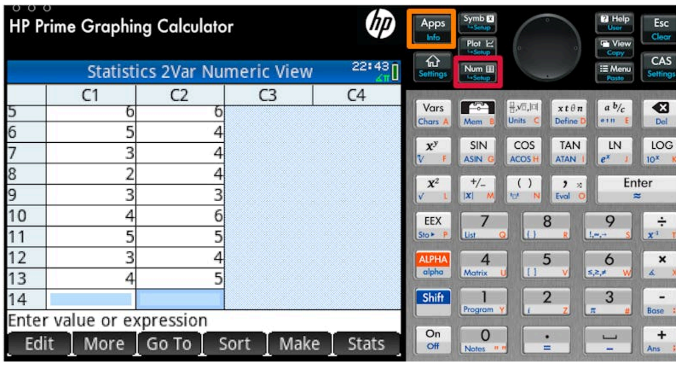
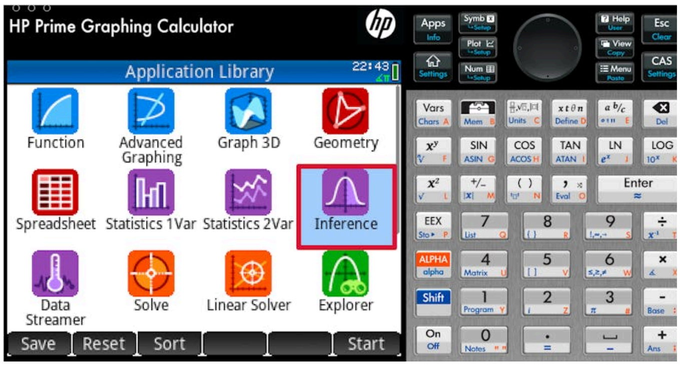
Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
The information you will need from this result screen is the p-value.	

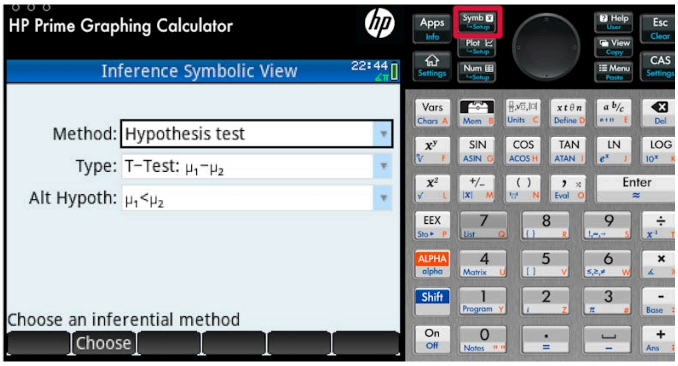
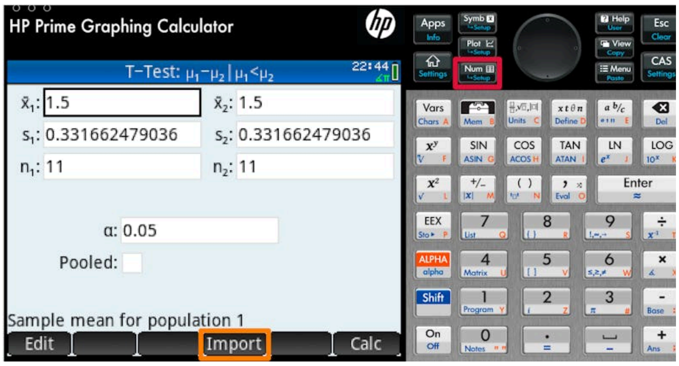
Steps	Explanation
To perform a t-test, you need to tell the calculator the data. You can do this for example using the 2-variable statistics application.	

Student view

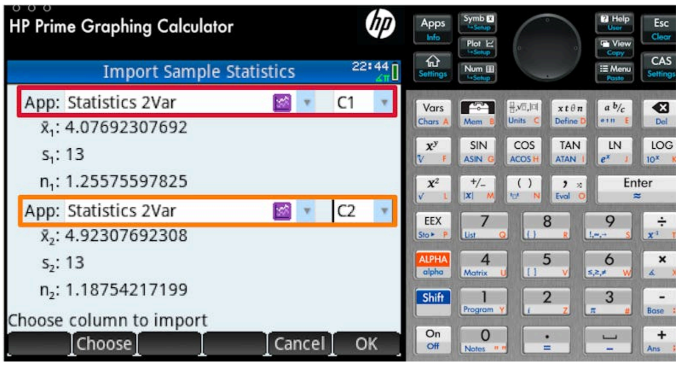
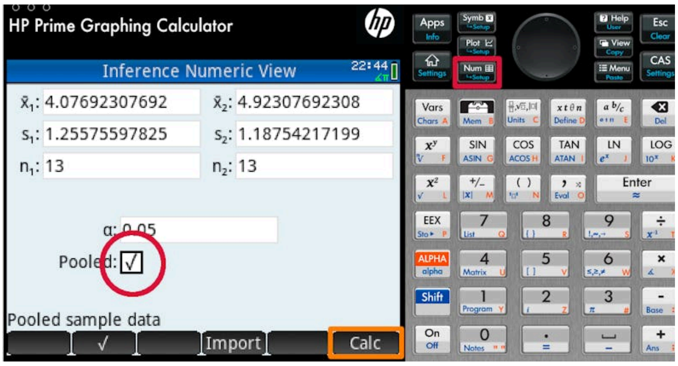
Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
Enter the data. In the context of this example, it is important to remember which list contains the data for the new and the old treatment. Once the data is stored, go back to the main screen to choose a different application.	
Choose the application for statistical inference.	

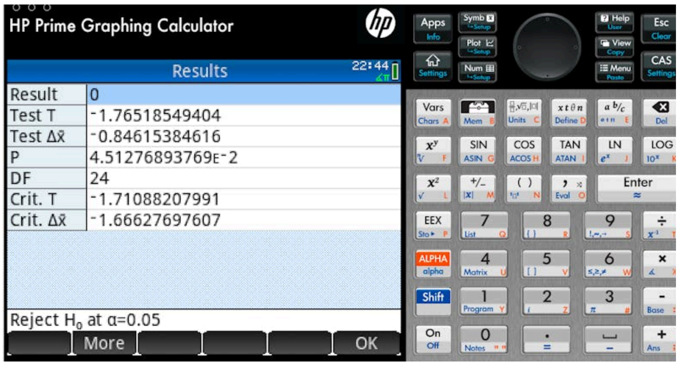
Student
view

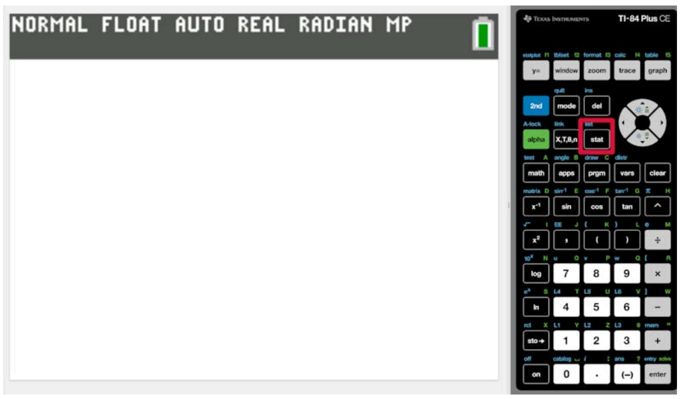
Steps	Explanation
In the symbolic view you can select what kind of test you would like to run. Select a to run a hypothesis test and select the 2-sample t-test (comparing the means μ_1 and μ_2). In this example you do a one-tailed test. Since the new treatment data is stored in the first list and you want to see wheter it reduces healing time or not, choose $\mu_1 < \mu_2$.	
Once you set the type of test you need, enter the numeric view. You will probably see a lot of random looking numbers on this screen. These are there from previous use of similar applications of the calculator. You will need to tell the calculator where the data is stored, so tap on Import.	

Overview
(/study/app/m/sid-122-cid-754029/

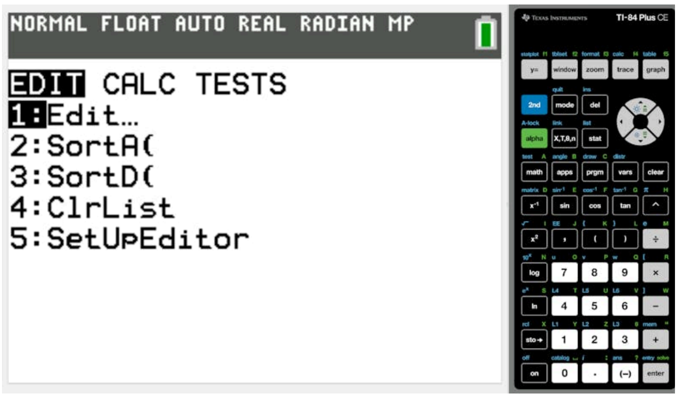
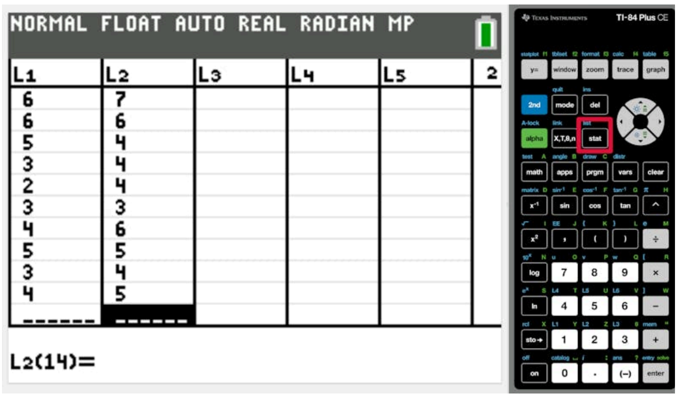
Steps	Explanation
At this step you need to remember, that you used the C1 and C2 columns of the 2-variable statistics application to store the data.	
Back in the numeric view you can see the mean and standard deviation of your data. It is important to tick the box for a pooled test. The IB syllabus specifies, that in this course only pooled tests will be used. Once done, tap on Calc to run the t-test. Note, that you can also set the significance level, but it is not needed for calculating the p-value.	

Student view

Steps	Explanation
The information you will need from this result screen is the p-value. Notice, that if you specify the significance level on the previous screen, the calculator also tells you the conclusion you should make about rejecting or not the null-hypothesis.	

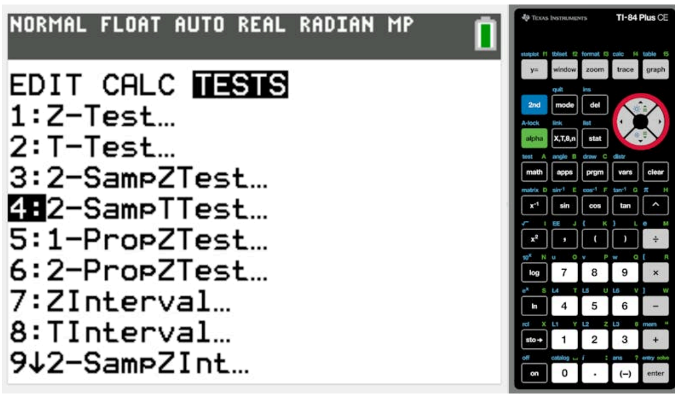
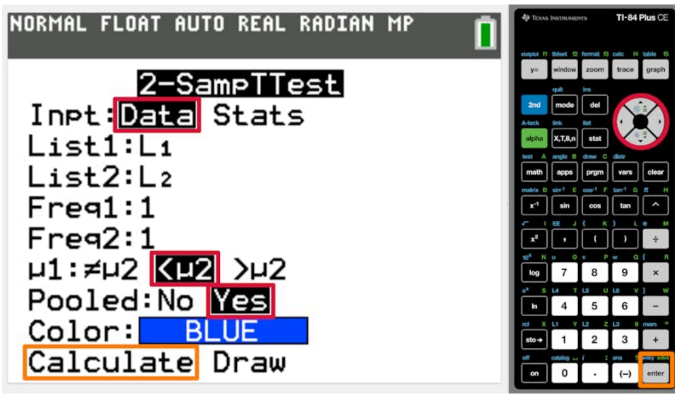
Steps	Explanation
To perform a t-test, open the statistics options.	

Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
You need to tell the calculator the data, so first choose to edit the lists.	
Enter the data. In the context of this example, it is important to remember which list contains the data for the new and the old treatment. Once the data is stored, press stat again ...	

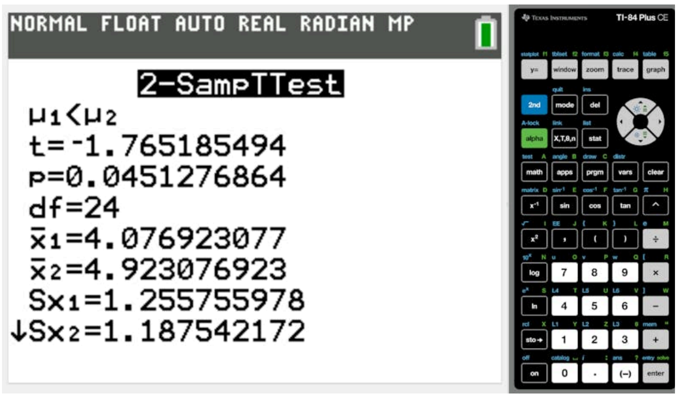
Student
view

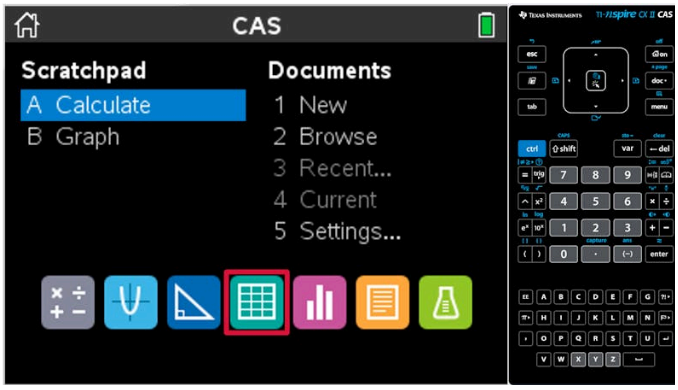
Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
... and choose to work with a 2 sample t-test.	
You need to set the parameters for the test. • Choose the option that you have the data as a list. • Tell the calculator where the data is. Leave the frequencies as 1. • In this example you do a one-tailed test. Since the new treatment data is stored in the first list and you want to see wheter it reduces healing time or not, choose $\mu_1 < \mu_2$. • Choose the option of a pooled test. The IB syllabus specifies, that in this course only pooled tests will be used. Once every detail is set, scroll further down to highlight calculate and press enter to run the t-test.	

Student
view

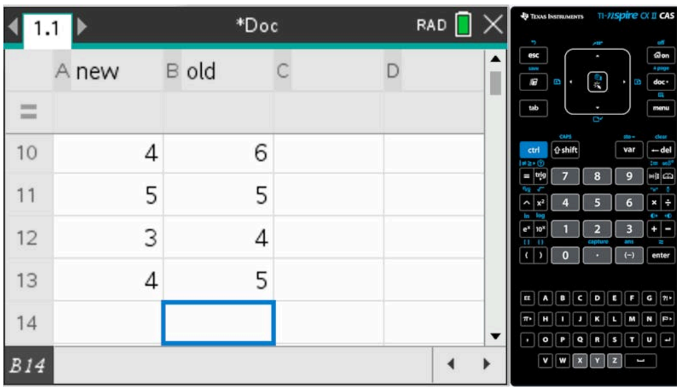
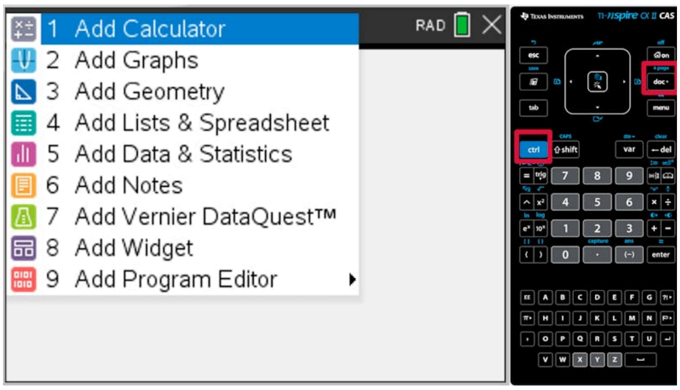
Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
The information you will need from this result screen is the p-value.	

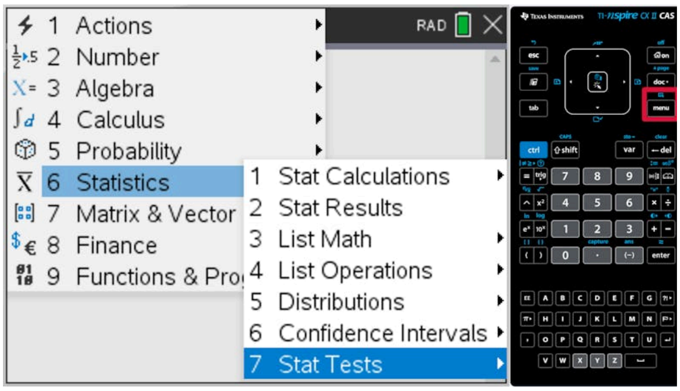
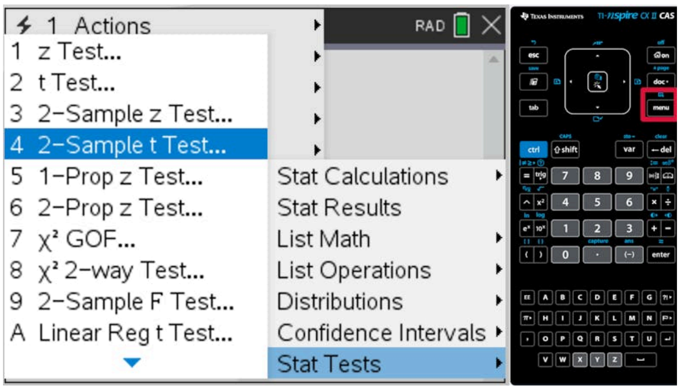
Steps	Explanation
To perform a t-test, you need to tell the calculator the data. You can do this for example using the spreadsheet.	

Student
view

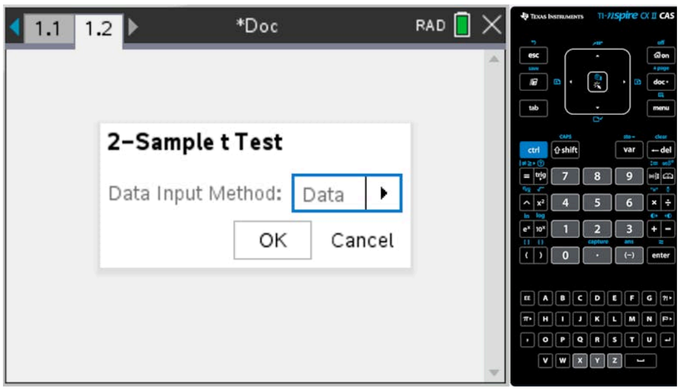
Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
Enter the data. In the context of this example, it is important to remember which list contains the data for the new and the old treatment, so give meaningful names to the lists.	
To run the t-test, add a calculator page.	

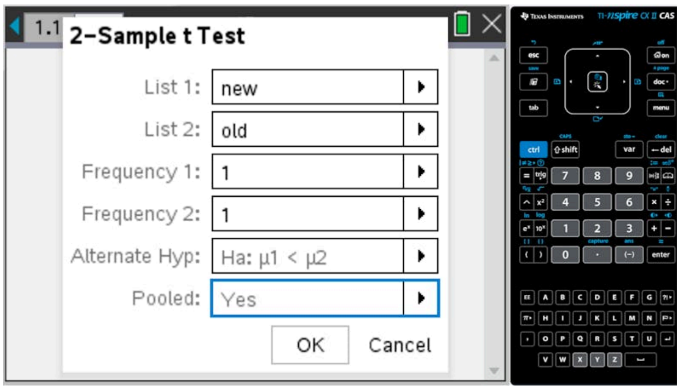
Student
view

Steps	Explanation
Open the menu and choose to work with statistical tests ...	
... and choose the 2-sample t-test.	

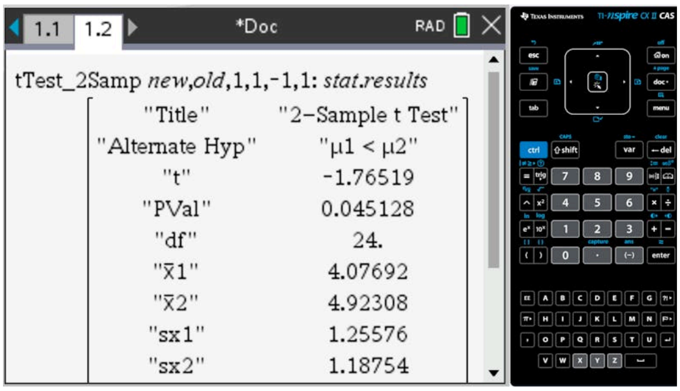
Student
view

<table>
<tr><th>Steps</th><th>Explanation</th></tr>
<tr><td>You have the data, so select that option.</td><td>
</td></tr>
<tr><td>You need to set the parameters for the test.

Tell the calculator where the data is. Use the names you gave to your lists. Leave the frequencies as 1.
In this example you do a one-tailed test. Since the new treatment data is stored in the first list and you want to see wheter it reduces healing time or not, choose $\mu_1 < \mu_2$.
Choose the option of a pooled test. The IB syllabus specifies, that in this course only pooled tests will be used.

Once every detail is set, press OK to run the t-test.</td><td>
</td></tr>
</table>

Overview
(/study/ap
122-
cid-
754029/

Steps	Explanation
The information you will need from this result screen is the p-value.	

Example 2

Two IB schools would like to compare the test results of their students. They randomly pick 10% of their second year students and check their test results on a test where the maximum mark was 60. The table below shows the results.

School A	48	40	53	48	45	57	54	46	47
School B	50	46	49	43	45	52	32		

Assume that the test results follow a normal distribution and the variance is the same in the two schools.

- State the appropriate null and alternative hypothesis for the two sample t-test.
- Find the p-value.
- Draw a conclusion at a 0.1 level of significance.

Student view

Stage 1: State the null and alternative hypotheses

Overview (/study/ap 122- cid- 754029/

The null-hypothesis is that the average test results are the same. Since neither of the schools claim a better performance, the scools can use a two-sided test. The alternative hypothesis is that the average test results are not the same.

H_0: $\mu_A = \mu_B$

H_1: $\mu_A \neq \mu_B$

Stage 2: Calculate the p-value

Use your calculator. Make sure that you are choosing the two-sided option and check that you are using a pooled test (on an IB exam you always use a pooled test). Find the p-value on the result screen.

$$p \approx 0.271$$

Stage 3: State your conclusion

You were told to use a 0.1 level of significance.

Since $0.271 > 0.1$, you do not reject the null hypothesis. Even though the average score for school A ($\overline{x}_A \approx 48.7$) is slightly greater than the average score for school B ($\overline{x}_B \approx 45.3$), the difference is not large enough, it can be explained by chance.

Example 3

A scientist would like to check how light affects the growth of wild garlic (*Allium ursinum*). He suspects that if the plant gets less light, it will grow shorter. To confirm this, he chooses two parts of the forest close to each other where the average light intensity is different. He collects eight plants from both areas and measures the length of the stem of each plant. The data (mesured in millimetres) is presented in the table below.

More light	278	159	280	341	305	333	332	303
Less light	263	196	224	131	124	235	247	131

Student view

Overview (/study/app/m/sid-122-cid-754029/

Assume that the height of the plants follow a normal distribution and the variance is the same in the two regions.

- State the appropriate null and alternative hypothesis for the two sample t-test.
- Find the p-value.
- Draw a conclusion at a 0.01 level of significance.

Stage 1: State the null and alternative hypotheses

The null-hypothesis is that the average height of the plants are the same in the two regions. Since the scientist would like to check the claim that plants grow shorter if they get less light, he can use a one-sided test. The alternative hypothesis is that the average height of the plants in the region with less light is less than the average height in the other region.

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 > \mu_2$

Stage 2: Calculate the p-value

Use your calculator. Make sure that you are choosing the appropriate one-sided option and check that you are using a pooled test (on an IB exam you always use a pooled test). Find the p-value on the result screen.

$$p \approx 0.00231$$

Stage 3: State your conclusion

You were told to use a 0.01 level of significance.

Since $0.00231 < 0.01$, you reject the null hypothesis in favour of the one-sided alternative hypothesis. It is unlikely to get these samples if light does not have a positive effect on the growth of wild garlic.

Student view

Overview
(/study/ap
122-
cid-
754029/

Example 4

In 1936 R.A.Fisher published a paper (https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-1809.1936.tb02137.x) investigating similarities and differences in the measurements of the flowers of 50 plants of the three species *Iris setosa* , *Iris versicolor* and *Iris virginica* . The following table shows the summary of the measurements of the petal length of two species.

	Iris versicolor	*Iris virginica*
Mean petal length	55.52	43.22
Sample standard deviation	5.52	5.36

Assume that the petal length of the flowers follow a normal distribution and the variance is the same for the two species.

Iris versicolor

Iris virginica

Student
view

Overview
(/study/ap
122-
cid-
754029/

The size of the *Iris versicolor* flowers seem to be bigger than the *Iris virginica* flowers. Based on the information about petal length, test this observation at a 0.01 level of significance.

- State the apprpriate null and alternative hypothesis.
- Find the p-value.
- Draw a conclusion.

Stage 1: State the null and alternative hypotheses

The null-hypothesis is that the average petal lengths are the same for the two species. Since you are asked to check the claim that the *Iris versicolor* flowers seem to be bigger than the *Iris virginica* flowers, you use a one-sided test. The alternative hypothesis is that the average petal length of *Iris versicolor* is greater than the the average petal length of *Iris virginica* .

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 > \mu_2$

Stage 2: Calculate the p-value

Use your calculator. Note that in this case the data itself is not given here (although you can find it in the original article by Fischer). Calculators have an option to give the summary statistics of the measurements given in the question above instead of the data itself. Make sure that you are choosing the appropriate one-sided option and check that you are using a pooled test (on an IB exam you always use a pooled test). Find the p-value on the result screen.

$$p \approx 9.21 \times 10^{-20}$$

Stage 3: State your conclusion

You were told to use a 0.01 level of significance.

Student view

Overview (/study/app/m/sid-122-cid-754029/...

Since $9.21 \times 10^{-20} < 0.01$, you reject the null hypothesis in favour of the one-sided alternative hypothesis. It is likely that the average *Iris versicolor* flower is indeed bigger than the average *Iris virginica* flower.

2 section questions

4. Probability and statistics / 4.11 Hypothesis testing

Checklist

Section Student... (0/0) Feedback Print (/study/app/m/sid-122-cid-754029/book/checklist-id-27593/print/) Assign

What you should know

By the end of this subtopic you should be able to:

- complete the four stages of a hypothesis test:
  1. State the null and alternative hypotheses
  2. Calculate the test statistic
  3. Find the p -value
  4. State your conclusion
- find the expected values from a contingency table using the formula or a calculator
- calculate the χ^2 statistic with the formula and with a calculator
- understand the difference between a test for independence and a goodness of fit test and in what context to use each one
- use p -values to determine significance
- compare the means of independent sets of data using either a one- or two-tailed
  t -test.

Student view

Investigation

2×2 contingency table

In any exam, the contingency table will be larger than a 2×2 table. However, often in an internal assessment or other project, we find ourselves with the smallest table one can use for a χ^2 contingency test, which is a 2×2 table.

In this case, the formula used to compute the χ^2 curve does not represent the correct probabilities for the regular χ^2_{calc} value. Therefore, the formula for the value χ^2_{calc} has been altered, just for the special case of a 2×2 table.

This adjustment is called the Yates continuity correction. Some statisticians no longer use this correction. However, for an IB internal assessment, it is expected to be used for a 2×2 contingency table.

The formula for calculating χ^2_{calc} for a 2×2 contingency table is:

$$\chi^2_{\text{calc}} = \sum \frac{(|\text{O} - \text{E}| - 0.5)^2}{\text{E}}$$

Explore the Yates correction further on your own. Why have some mathematicians stopped using the Yates correction?

Construct several 2×2 contingency tables and calculate χ^2_{calc} for each one with and without the Yates correction. Do you notice a difference in the value of χ^2_{calc}? The number of degrees of freedom for a 2×2 contingency table is $df = (2-1)(2-1) = 1$. Use a critical value table and compare your values to see whether the differences affect significance, and if so, at what level of significance.