Index

The big picture

When answering questions in mathematics, the emphasis is on making sure all of the calculations are correct. However, within statistics, it is very easy to come to incorrect or misleading conclusions even when you have done all of the calculations correctly.

Consider the following video about the perils of ‘lurking variables’.

Concept

The concepts of validity and reliability in statistics help mathematicians make reasonable interpretations of the results their data produce.

Valid data collection and analysis

Valid data collection methods

Making connections

In section 4.1.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/sampling-id-26063/) you learned about five different sampling techniques you could use to collect a random sample of data when collecting data for an entire population is not possible or impractical:

- simple
- convenience
- systematic
- quota
- stratified.

In this section, you will build on your knowledge from section 4.1.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/sampling-id-26063/) and explore how to design a survey or questionnaire to collect data effectively once you have determined which sampling technique you will use.

Example 1

★☆☆

Overview
(/study/ap
ai-
hl/sid-
132-
cid-
761618/ov

A coffee shop at a mall in Thailand

Credit: JohnnyGreig GettyImages

Xay is a student at an international school in Thailand. He would like to determine the average amount of money that an average Thai person spends each month on entertainment. He creates a questionnaire to help collect the information and decides to use a systematic sampling method of asking every tenth person that walks out of the Starbucks at a mall near his home. Describe some possible problems with Xay's chosen method.

Making connections

Recall from section 4.1.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/sampling-id-26063/) that a systematic sampling method involves randomly choosing a starting point within the data set and then systematically taking every n th value until you have collected your desired number of data points.

In this case, Xay has chosen to only question people who are walking out of a Starbucks. However, the group of people who visit a Starbucks might not be representative of Thai people – for example they might belong to one specific age group or earnings class. Also, by only asking people at a mall, Xay has chosen a group of people from a place where people are more likely to spend money. For these reasons, Xay has chosen a method that will most likely be very biased .

Student
view

Important

Overview
(/study/ap
ai-
hl/sid-
132-
cid-
761618/ov

The design of a data collection method shows bias when it would consistently give a value that either overestimates or underestimates the actual value for the entire population. The collection method is more likely to be unbiased if the sample group is representative of the entire population being studied.

Example 2

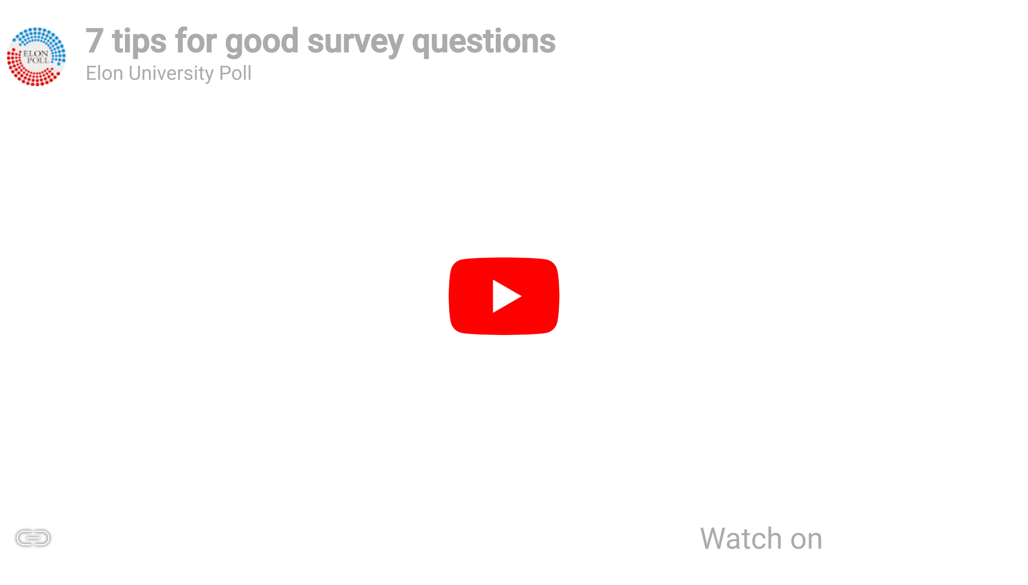
★☆☆

Describe the steps that Xay can take to remove the bias from his collection method.

Both of the issues mentioned in the answer to **Example 1** can be addressed by questioning people in more than one location. Because Starbucks might be favoured by a particular type of person, Xay could instead ask people coming out of various grocery stores, since everybody has to eat. To address the issue of only asking people at a mall, Xay could also include people at other places around town. In both cases he would need to consider whether all of his chosen places together are representative of the Thai population.

Creating a questionnaire or survey

When creating a questionnaire or survey, it is important to spend time designing the questions carefully. Watch this video about how to write good questions.



Here are some of the issues covered in the video:

Student
view

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

Personal questions: You should be very careful when writing questions that are personal in nature. As explained in the video, the reason for including the question should be explained. The question should be written in a way that encourages the participants to give accurate answers. Examples of personal information include examination results and class ranking, weight, medical history, or political preferences.

Precise questioning: It is important that your questions can be easily understood. If your question is vague or ambiguous, then the participants' answers will not be useful.

Structured vs. unstructured questions: A structured question is one that has specific answers that are provided by you to the participants. Just as with precise questions, you may find that having structured questions produce better results to analyse. However, when using this type of question, it is important that you have considered and provided all of the possible answers that participants may give. It may be helpful to include an answer option of 'other', but this can be easily overused within surveys.

Reliability and validity tests

Once you have decided on the design of your questionnaire or survey, you must consider whether it is reliable and valid.

Reliability

✓ **Important**

In statistics, reliability refers to how similar the data sets obtained from multiple trials are. A reliable method produces reliable, consistent data.

For this course, you will need to be familiar with two different tests that can be used to measure the reliability of a statistical data collection method.

Test–retest: You can distribute the same questionnaire or survey to a sample group of participants at two different times. You can then compare the two sets of data to see how well they match. If your method of collecting data set is reliable, the second set will be very similar.

Parallel forms: For this test method you will need to create two different surveys or questionnaires designed to measure the same variable. By giving both of the surveys or questionnaires to your sample group and comparing the results, you can determine the reliability of your data collection method.

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

Exam tip

There are statistical tests that can be carried out to quantify the reliability of a data collection method, but for this course you only need to be able to explain how each of the two tests above can be used qualitatively to determine the reliability of a data collection method.

Validity

Important

Validity describes whether the collected data set accurately measures the variable being investigated. Validity is also used to describe whether the distribution of the data sample matches the distribution of the population.

You will need to be familiar with two methods used to measure validity for this course.

Criterion-related validity: A statistical instrument is said to have criterion-related validity if it accurately predicts the outcome for another measure. An example is the SAT, a test used in schools in the USA. Over time it has gained a measure of criterion-related validity by effectively predicting a student's success in university-level courses.

Content validity: When considering the content validity of a statistical instrument, you should consider whether it is actually testing what you want it to test. For example, imagine that for your next history test your teacher decided to give you the test in Spanish instead of English, in which the course is taught. In this case, the test would be testing your knowledge of Spanish rather than of history and thus it would not have content validity.

4 section questions

4. Probability and statistics / 4.12 Collecting and analysing data

Categorising numerical data in a chi squared table

Student view
Section Student... (0/0) Feedback Print (/study/app/math-ai-hl/sid-132-cid-761618/book/categorising-numerical-data-in-a-chi-squared-table-id-27977/print/) Assign

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

Making connections

Recall from section 4.11.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/calculating-the-chi-squared-statistic-id-27881/) that one-way contingency tables are used for χ^2 goodness-of-fit test and two-way contingency tables are used for the χ^2 test for independence.

While χ^2 tests are generally used to analyse categorical data, they can also be used with numerical data by dividing the data into numerical intervals or categories. Let us explore how you can do this.

Cameron is conducting a study into whether a US graduate's income is dependent on the SAT score (the result of a standardised test) required for acceptance by their university. He has decided to use the χ^2 test for independence for this study and therefore, as both of his variables are numerical, needs to split the data into numerical intervals.

After collecting data for 100 universities in the USA, he finds that the SAT requirements ranged from 1210 to 1565. He also found that the average salaries for the graduates of the universities 10 years after graduation ranged from $39\,300$ to $91\,300$.

How should Cameron divide the ranges of both the SAT scores and salaries for use with the χ^2 test for independence?

Important

While there is no set procedure for determining numerical intervals, you should be able provide a justification for how you decide to divide the data into the intervals.

Cameron decides to divide the data as shown in the table.

Average salary 10 years after graduation	SAT acceptance requirement				
	$1200–1299$	$1300–1399$	$1400–1499$	$\geqslant 1500$	**Total**
$\$30\,000–39\,999$	0	0	1	0	1
$\$40\,000–49\,999$	2	5	9	0	16

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

Average salary 10 years after graduation	SAT acceptance requirement				
	1200–1299	1300–1399	1400–1499	$\geqslant$ 1500	**Total**
$50 000–59 999	3	15	19	3	40
$60 000–69 999	1	8	10	6	25
$70 000–79 999	0	3	3	7	13
$80 000–89 999	0	0	2	1	3
$\geqslant$ $90 000	0	0	0	2	2
Total	6	31	44	19	100

Example 1

★☆☆

Describe how Cameron can determine whether this division of the data is adequate for carrying out the χ^2 test for independence.

Cameron first creates the expected frequency table shown below.

Average salary 10 years after graduation	SAT acceptance requirement				
	1200–1299	1300–1399	1400–1499	$\geqslant$ 1500	**Total**
$30 000–39 999	**0.06**	**0.31**	**0.44**	**0.19**	1
$40 000–49 999	**0.96**	**4.96**	7.04	**3.04**	16
$50 000–59 999	**2.4**	12.4	17.6	7.6	40
$60 000–69 999	**1.5**	7.75	11	**4.75**	25
$70 000–79 999	**0.78**	**4.03**	5.72	**2.47**	13
$80 000–89 999	**0.18**	**0.93**	**1.32**	**0.57**	3

Student view

Overview
(/study/ap
ai-
hl/sid-
132-
cid-
761618/ov

Average salary 10 years after graduation	SAT acceptance requirement				
	1200–1299	1300–1399	1400–1499	$\geqslant 1500$	Total
$\geqslant \$90\,000$	**0.12**	**0.62**	**0.88**	**0.38**	2
Total	6	31	44	19	100

One important requirement for the χ^2 test for independence is that **all** of the expected frequencies must be greater than or equal to 5. This is known as the expected counts condition for χ^2 . As you can see from the values in bold in the table above, many cells do not meet this requirement.

Therefore, Cameron must determine new intervals for the data. The calculator instructions in section 4.11.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/calculating-the-chi-squared-statistic-id-27881/) show how to create expected frequency tables using a calculator.

Be aware

In addition to the expected counts condition that is mentioned in the above solution, remember that the values within the contingency tables for χ^2 must be frequencies (counts) and not proportions or actual data values.

In an attempt to correct the issue with the expected frequencies from the original numerical intervals, Cameron reorganises the tables as shown below.

Average salary 10 years after graduation	SAT acceptance requirement		
	1200–1400	> 1400	Total
$< \$50\,000$	7	10	17
$\$50\,000–70\,000$	27	38	65
$> \$70\,000$	3	15	18

Student view

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

Average salary 10 years after graduation	SAT acceptance requirement		
	1200–1400	> 1400	Total
Total	37	63	100

Example 2

★★☆

Cameron notices that there is still a cell in the table above that has a frequency less than 5. Determine whether he needs to reorganise the numerical intervals again.

Recall that the expected counts condition requires that all of the **expected frequencies** must be greater than or equal to 5. Therefore, to answer this question you must first create the expected frequency table.

Average salary 10 years after graduation	SAT acceptance requirement		
	1200–1400	> 1400	Total
$<$ \$50 000	6.29	10.71	17
\$50 000–70 000	24.05	40.95	65
$>$ \$70 000	6.66	11.34	18
Total	37	63	100

As all of the expected frequencies are ≥ 5 , these numerical intervals have now satisfied the expected counts condition necessary to continue carrying out the χ^2 test for independence.

Exam tip

Even though the above frequency table now satisfies the expected counts condition, that does not necessarily mean that the chosen variables will adequately answer the research question. Remember to use the techniques covered in section 4.12.1

Student view

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

(/study/app/math-ai-hl/sid-132-cid-761618/book/valid-data-collection-and-analysis-id-27560/) to analyse the selection of the variables.

International Mindedness

These examples are based upon data collected for universities within the United States. What importance do standardised tests have in the admission processes of universities in other countries around the world?

Activity

In this activity, you will practise categorising numerical data by investigating whether the height of a person is dependent on their nationality.

This can be completed independently or in pairs.

1. Go to the following Wikipedia page: List of average human height worldwide (https://en.wikipedia.org/wiki/List_of_average_human_height_worldwide)
2. Using the data found on the Wikipedia page, create a χ^2 two-way contingency table that can be used to investigate the possible dependency of height on nationality. To do this, you will need to decide on how to group the numerical data for heights into categories. You will also need to determine how to best group the countries.
3. Create the expected frequency table.
4. Return to Step 2 and regroup the data if your expected frequency table has any frequencies less than 5.
5. Continue the activity until you have successfully created **justifiable** categories for the data that yield an expected frequency table with all values greater than 5.

Yates continuity correction

Exam tip

On exams you will not be required to apply Yates continuity correction and you will not meet an example where it would be needed. The rest of this section can be skipped for exam preparation. However, it is expected that you use it in your exploration in case it is needed.

Student view

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

In the previous section, Cameron was able to successfully group his data into a table containing two columns and three rows. However, there may be times when your contingency tables will end up with only two columns and two rows.

Example 3

★☆☆

State how many degrees of freedom there are for a 2×2 contingency table.

Recall the degrees of freedom equation from **section 4.11.2** :

$$\begin{aligned}\text{degrees of freedom} &= (r-1)(c-1)\\ &= (2-1)(2-1)\\ &= 1\end{aligned}$$

Therefore, there is only one degree of freedom within a 2×2 contingency table.

In some cases you may be forced to reduce your contingency tables to a 2×2 size. At this size, the data is reduced to discrete binomial variables. However, as the χ^2 distribution is continuous, any approximation made using the discrete data will inevitably introduce some error. To reduce this error you will need to include the Yates continuity correction in your calculations. This is done by altering the χ^2 calculation as follows:

$$\chi^2 = \sum \frac{\left(|f_0 - f_e| - 0.5\right)^2}{f_e}$$

Example 4

★☆☆

Results of a questionnaire about eating breakfast on exam days is shown in the table below. Calculate the χ^2 value for this data.

	Yes	No
Boys	10	7
Girls	13	9

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

First create the observed contingency table.

	Yes	No	Total
Boys	10	7	17
Girls	13	9	22
Total	23	16	39

Then create the expected contingency table.

	Yes	No	Total
Boys	10.03	6.97	17
Girls	12.97	9.03	22
Total	23	16	39

Use the formula for χ^2 that includes the Yates continuity correction.

$$\begin{aligned}\chi^2 &= \sum \frac{(|f_o - f_e| - 0.5)^2}{f_e} \\ &= \frac{(|10 - 10.03| - 0.5)^2}{10.03} + \frac{(|7 - 6.97| - 0.5)^2}{6.97} + \frac{(|13 - 12.97| - 0.5)^2}{12.97} + \frac{(|9 - 9}{} \\ &= 0.0970\end{aligned}$$

Be aware

The answer for **Example 4** was obtained from the unrounded values for the expected values. The expected values were rounded only for the purpose of displaying the values and the full unrounded values were carried forward into the next calculation.

Remember that you should **never** round calculated values until the end of the given problem.

2 section questions

Student view

Estimating parameters

Degrees of freedom when estimating parameters

Making connections

In section 4.11.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/calculating-the-chi-squared-statistic-id-27881/) , when you studied the χ^2 goodness-of-fit test you learned that $n - 1$ is used to calculate the number of degrees of freedom.

The method you previously learned for calculating degrees of freedom does not work when estimated parameters are included in the situation.

Instead of only considering the number of categories, n, you must also consider the number of parameters, r, that you have estimated when creating your data tables. You can then calculate the number of degrees of freedom using the formula $\text{d.f.} = n - 1 - r$.

Example 1

★☆☆

Ramón is trying to decide whether the sample of data he has collected follows a normal distribution. He begins by calculating the mean and the standard deviation of his data. He then standardises the data and creates the following table.

	$z \leqslant -2$	$-2 < z \leqslant -1$	$-1 < z \leqslant 0$	$0 < z \leqslant 1$	$1 < z \leqslant 2$	$z > 2$
Observed	9	51	81	99	54	6
Expected	6.8	40.8	102.4	102.4	40.8	6.8

Find the number of degrees of freedom for his data.

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

For this situation, Ramón has divided his data into six different categories. Therefore, $n = 6$. However, in order to standardise his data, Ramón had to use the mean and standard deviation that he calculated from the same data. Since Ramón's data is a sample, the mean and standard deviation are both estimates for the population data. Therefore, $r = 2$.

Hence:

$$\begin{aligned} \text{d.o.f.} &= n - 1 - r \\ &= 6 - 1 - 2 \\ &= 3 \end{aligned}$$

The example above involved data that Ramon suspected as coming from a normal distribution. Similar tests can be used to check other distributions.

Important

- When estimating parameters for checking whether the distribution is normal, you need to subtract 2 from the usual degrees of freedom, since the mean and the standard deviation of the distribution are estimated.
- When estimating parameters for checking whether your data is from a binomial distribution, you need to subtract 1 from the usual degrees of freedom, since only the probability of success is estimated.
- When estimating parameters for checking whether your data is from a Poisson distribution, you need to subtract 1 from the usual degrees of freedom, since only the mean of the distribution is estimated.

The importance of degrees of freedom

As you learned in section 4.11.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/calculating-the-chi-squared-statistic-id-27881/), once you have calculated your χ^2 value you compare it with the χ^2 critical value to determine whether the difference between the observed and expected frequencies is large enough to reject the null hypothesis.

Let us explore where the critical value comes from and why the degrees of freedom are so important.

Student view

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

As you look at the interactive graph of the χ^2 distribution for different degrees of freedom shown below, consider the following questions:

- What happens to the curve as you change the number of degrees of freedom?
- What does the area of the shaded region represent?
- What does the x-coordinate of the point on the horizontal axis represent?

Interactive 1. Chi-square Distribution.

More information for interactive 1

This interactive allows users to explore the Chi-square distribution and understand the importance of degrees of freedom in statistical hypothesis testing. The Chi-square distribution is used to determine whether the difference between observed and expected frequencies is significant, and the degrees of freedom play a crucial role in shaping the distribution.

A distribution graph is displayed on the xy plane, where the x-axis ranges from 0 to 60 and the y-axis ranges from -0.02 to 0.24. A bell-shaped curve represents the chi-square distribution, with its shape determined by the degrees of freedom. Users can adjust the degrees of freedom using a horizontal slider at the top, ranging from 1 to 40. A movable red point on the x-axis divides the area of the curve, with the right side representing the shaded region and the left side representing the unshaded region.

Student view

As you change the degrees of freedom, the shape of the Chi-square curve will shift. With fewer degrees of freedom, the curve is highly skewed to the right, but as the degrees of freedom increase, the curve becomes more symmetric and approaches a normal distribution. This demonstrates how the degrees of freedom influence the critical values used in hypothesis testing.

As users move the red point, the interactive calculates and displays the area of the shaded region under the curve to the right of the point. For example, if the degree of freedom is 0, then the area of the shaded region is 0, and if the area is 5, then the area of the shaded region is 0.08.

This shaded area represents the p-value. This interactive tool helps users interpret the degrees of freedom affecting the Chi-square distribution and how the p-value is determined based on the Chi-square value. By experimenting with different degrees of freedom and Chi-square values.

Credit: GeoGebra (https://www.https://www.geogebra.org/m/xscjetub) Camille Fairbourn

Important

- On the applet above the area of the shaded region is the probability that the χ^2 statistics is above the x-coordinate of the moving point.
- The critical value for a certain level of significance is the bound for the region that has area the same as the significance level.

Exam tip

Your calculator knows the χ^2 distribution, so you can find critical values corresponding to certain probabilities. However, you will not need this on exams. If a question asks you to draw conclusion based on critical values, it will be given on the exam.

Performing the χ^2 test

In **Example 1** above you saw how to calculate the degree of freedom. Let's see how to use this degree of freedom to perform the test.

Example 2

★☆☆

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

Ramón carries out the χ^2 goodness-of-fit test on his data from the table given above, in **Example 1**. He is using a 5% significance level. The critical value for this test is 7.81.

Determine how he should interpret the results.

Method 1 (using the significance level)

Use you calculator to find the p-value of the test. Make sure you use 3 as the degree of freedom. This is what you found in **Example 1**.

$$p \approx 0.0066922822$$

This value is less than the 0.05 significance level. This means that the probability of getting a data this extreme is very small if the distribution is indeed normal.

Ramon should **reject** the assumption that the distribution is normal.

Method 2 (using the critical value)

Calculators also give you the χ^2 statistics for the data:

$$\chi^2 \approx 12.21162684$$

This value is greater than the given critical value (7.81). This means that the difference between the observed and expected (under the assumption that the distribution is normal) data is too large.

Ramon should **reject** the assumption that the distribution is normal.

Example 3

★★★

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/app/math-ai-hl/sid-132-cid-761618/ov

	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 the probability that he makes a free throw does not change over the year and that each of his attempts 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 and his accuracy does not change, 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 unknown p probability of success.

Then state the opposite as the alternative hypothesis.

H_1: The number of made free throws does not follow a binomial distribution.

Stage 2: Perform the test

To work out the expected number of times Joe makes $0, 1, 2$ and 3 free throws, you need to estimate the probability of success for the binomial distribution. You can use the formula that the mean of a binomial distribution is given by $\mu = np$.

First you can find the mean of the observed data.

$$\bar{x} = \frac{6 \times 0 + 51 \times 1 + 153 \times 2 + 155 \times 3}{6 + 51 + 153 + 155} = \frac{822}{365} \approx 2.2520548$$

If you divide this number by 3, the number of trials, you get the estimated percentage.

Student view

$$p = \frac{822}{3 \times 365} \approx 0.750685$$

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

Let's use this estimated accuracy to collect the information needed for the hypothesis test 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 (/study/app/math-ai-hl/sid-132-cid-761618/book/calculating-bionomial-probabilities-id-26116/).
- In the third row, multiply your numbers in the second row by 365 to find the expected counts.

	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	0.0155	0.140	0.421	0.423
Expected count	5.656	51.094	153.84	154.41

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

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.

$$p \approx 0.986$$

Stage 4: State your conclusion

For the conclusion you need to compare this p-value with the given significance level, 0.05. Since $0.986 > 0.05$, we do not reject H_0. This data does not give enough evidence against the suspected binomial distribution. The data is consistent with the assumption

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

that Joe's accuracy does not change during the year and his attempts are independent of each other.

4. Probability and statistics / 4.12 Collecting and analysing data

Checklist

Section Student... (0/0) Feedback Print (/study/app/math-ai-hl/sid-132-cid-761618/book/checklist-id-27979/print/) Assign

What you should know

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

- design valid questionnaires and surveys by
  - considering whether they are biased or unbiased
  - using structured and unstructured questions
  - asking precise questions
  - considering whether a question is personal and explaining its purpose if so
  - selecting the relevant variables to investigate
- categorise numerical data in a χ^2 table and justify the chosen intervals
- choose an appropriate number of degrees of freedom when estimating parameters in a χ^2 goodness-of-fit test
- define statistical reliability
- define statistical validity
- give examples of tests that are used to measure statistical reliability and statistical validity.

4. Probability and statistics / 4.12 Collecting and analysing data

Investigation

Student view

Section Student... (0/0) Feedback Print (/study/app/math-ai-hl/sid-132-cid-761618/book/investigation-id-27980/print/) Assign

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

Leaves on a tree

Credit: temmuzcan GettyImages

When discussing the normal distribution, many teachers often give real-life examples of manufactured products or things formed in nature. Now that you have explored the χ^2 goodness-of-fit test and degrees of freedom when estimating parameters, you can test this out for yourself.

Obtain a large amount of a product that you can measure. Some examples could include a bag of dry beans, a collection of leaves from the same plant or tree, or a large box of nails.

Make the appropriate measurements, such as mass or length, and then use the χ^2 goodness-of-fit test to investigate whether the distribution of your measurements follows the normal distribution.

Rate subtopic 4.12 Collecting and analysing data

Help us improve the content and user experience.

Student view