Determining the Probability of a Good Day for Sowing Cotton

**Teacher Notes & Solutions**

**ACMSP247**

Mathematics / Year 10 / Statistics and Probability / Chance

Use the language of ‘if ....then, ‘given’, ‘of’, ‘knowing that’ to investigate conditional statements and identify common mistakes in interpreting such language

**Elaborations**

* using two-way tables and Venn diagrams to understand conditional statements
* using arrays and tree diagrams to determine probabilities

**Aims**

* To consolidate the language and notation used to describe sets of data and reinforce connections between different representations of frequency and relative frequency (experimental probability) though their application to a real life context.
* To introduce the language of c[onditional probability](http://vocabulary.curriculum.edu.au/scot/13445" \t "_blank" \o "View information about this ScOT term (opens in a new window))
* To instruct students in the use of excel spreadsheet commands for conditionally formatting data.
* To encourage critical thinking regarding the analysis and application of statistical data.

**The Context**

The profitability of growing cotton depends on the prevailing economic conditions and the yield of the crop. Although soils and rainfall influence cotton yields, temperature is the dominant environmental factor.

Temperature conditions vary between locations. In the Trangie district, a farmer needs to be able to sow (i.e. plant) the crop on or between 15 September and 20 October for the yield to be profitable. This period is referred to as the planting season. There are two temperature conditions that need to be considered to determine whether a day in the planting season is suitable for sowing cotton.

Students are given the scenario of farmers in the Trangie district faced with the decision of whether or not they should invest in cotton production. Some of the farmers are more risk-averse than others. The risk-averse farmers will only plant cotton on a day when both temperature conditions are satisfied. The less risk-averse farmers will plant cotton provided that at least one of the two temperature conditions is satisfied. Students analyse historical temperature data at Trangie to determine the relative frequency these conditions having been satisfied in past years.

Before beginning this series of lessons, read the background information provided. You could give this information to your students and/or you could introduce the context using a video or a slide presentation such as the one at <https://prezi.com/tzrjuk06rqih/visual-story-of-cotton/> which shows how the cotton plant grows.

Cotton is sown in the spring and harvested in the autumn. Emphasise the importance of crop being grown when it will not be damaged by frosts, and that temperature determines the rate at which the cotton plant develops.

Find Trangie on the map of cotton growing districts. Ask them where this district is (north, south, east or west) *in relation to* other cotton growing districts.

Show the line graph “Relative yield of cotton at Trangie vs. Planting date” in the introduction.

From this graph, ask them to find the planting dates for which the cotton grower can expect to get a relative yield of at least 95%. This means that based on previous yields, the yield on average is expected to be within the top 5% of yields.

**The Mathematics Tasks**

There are three tasks. These are ordered sequentially, however each task has sufficient information within it to be able to stand alone.

Task 1 – Conditional formatting of Microsoft Excel spreadsheets

In this task, students learn to manipulate data in spreadsheets. The conditional formatting function in Excel is used to identify good days for sowing and spreadsheet formulae are used to count these days. This task also uses scatter graphs to determine linear trends in data **(ACMSP279)**.

Task 2 – Constructing and interpreting Venn diagrams and two-way tables

This task revises the use of Venn diagrams and the language and notation used to describe sets of data **(ACMSP204 and ACMSP205)**. It also relates data in Venn diagrams to data expressed in two-way tables **(ACMSP292)**.

Task 3- Calculating probability and conditional probability

This task investigates the concept of relative frequency (experimental probability) **(ACMSP226)** and the concept of dependence and independence **(ACMSP246)** using Venn diagrams, two-way tables and tree diagrams. During the investigative process, emphasis is given to understanding the meaning of events being dependent and the correct use of conditional language **(ACMSP247)**.

Detailed plans for the delivery of each task and solutions to the questions posed are given below.

Following the launch of each lesson, you will need to decide whether students work individually, in pairs or as small groups to answer the given questions, prior to discussing the solutions as a class.

**Task 1 – Conditional formatting of Microsoft Excel spreadsheets**

1) Check that all students understand the language on the webpage <http://www.csd.net.au/greenlight> “Have you got the greenlight for planting this season”

* *AEST* - Australian Eastern Standard Time
* *Forecast average temps* – predicted average daily temperatures (found by averaging the maximum and minimum temperature of each day)
* *The week following* – Note that this doesn’t include the day in question
* *On a rising plane* - trending upwards
* *Adjustments* - These are measures or actions taken to enable better sowing conditions eg. the planting rate (the amount of seeds sown per hectare) may need to be increased to allow for a potential loss of seedlings if the temperature drops

2) Check that all students understand what is meant by *conditionally formatting* a spreadsheet.

* *Conditionally* - subject to, or depending on one or more requirements

3) Show the line graph of “Average daily temperatures at Trangie, 15 September – 20 October”.

* Check that all students know the difference in meaning between *daily average temperatures* and *average daily temperature*.

The average daily temperature points in the graph are daily average temperatures that have been averaged over the days in the period from 15 September – 20 October (36 days) each year.

From the graph, ask them to find the maximum and minimum values for each line and the year in which these values occurred.

4) Show the spreadsheet *Trangie Planting Season Data*, or ask students to open it.

* Why do you think this pattern happens?

The pattern happens because the staff at Trangie Agricultural Research Station who recorded the measurements, were not employed on weekends.

In October there is missing data for 3 days in a row because of the October long weekend.

The diagonal pattern is formed because in consecutive years, the days of the week go forward by one (a year being 52 weeks and 1 day). When it is a leap year (52 weeks and 2 days) the day of the week goes forward by 2 days.

5) Ask students to open the spreadsheet *Testing for Good Sowing Days*.

Point out that there are 3 worksheets, one for each of 3 years.

Revise spreadsheet skills.

6) Give students the Task 1 sheet and ask them to follow the instructions (beginning on page 3) to conditionally format the data so they can see which days are good days for sowing cotton.

In the sheet, bolded numbers or letters are those to be typed into a box.

Questions to answer are marked with a dot.

* On how many days in the 1992 planting season, was the soil temperature measured? 25 days
* How many of these days have a soil temperature greater than 14°C? 9 days
* What fraction of the days when soil temperature was measured, have a soil temperature greater than 14°C? Write your answer as a decimal correct to 2 decimal places. 0.36
* By using data in this spreadsheet as “a forecast of average temperatures”, what assumption has been made?

It is assumed that the temperatures forecast were those that actually happened.

* How many of the days with soil temperature data, have mean temperatures for the following week on a rising plane? 15 days
* What fraction of the days with soil temperature data, have average temperatures for the following week on a rising plane? Write your answer as a decimal correct to 2 decimal places. 0.6
* From conditionally formatting your spreadsheet:
* how could you tell which days are **green light** days?

In their row, they have a red border for the cells in both Column B and Column D.

* how could you tell which days are **red light** days?

In their row, they have a no cells with a red border.

* how could you tell which days are **amber light** days?

In their row, they have only one cell with a red border.

* What are the two conditions for **red light** days?

A soil temperature of 14°C or less, and average temperatures for the following week *not* rising.

* Write this rule as an Excel formula to conditionally format these days.

(Note that the symbols <= written together mean ”less than or equal to”.)

=AND($B3<=14,$D3=”N”)

* What are the two conditions for **brown light** days?

A soil temperature greater than 14°C and avearge temperatures for the following week *not* rising.

* Write this rule as an Excel formula to conditonally format these days.

=AND($B3>14,$D3=”N”)

* What are the two conditions for **orange light** days?

A soil temperature of 14°C or less and average temperatures for the following week rising.

* Write this rule as an Excel formula to conditonally format these days.

AND($B3<=14,$D3=”Y”)

To count the number of **green light** days, go to Cell E51 and type the formula

**=COUNTIFS(B3:B38,">14",D3:D38,"=Y")**.

* Explain the meaning of the symbols in this formula.

There is more than one condition, so the formula is =COUNTIFS (not =COUNTIF) and the two conditions are separated by a comma.

Each condition is written in inverted commas after the location of the array of cells to which it is to be applied and a comma.

The formula says to count the number of days (rows) from row 3 to row 38 *if* they have a value of greater than 14 in Column B *and if* they have a Y in Column D.

* What formula would you type to get a count of the **red light** days into Cell E52?

=COUNTIFS(B3:B38,"<=14",D3:D38,"=N")

* What formula would you type to get a count of the **brown** **light** days into Cell E53?

=COUNTIFS(B3:B38,">14",D3:D38,"=N")

* What formula would you type to get a count of the **orange light** days into Cell E54?

=COUNTIFS(B3:B38,"<=14",D3:D38,"=Y")

* Why is it better to compare the proportion of days in each category rather than compare the number of days in each category?

Using proportions makes possible a comparison between different seasons, even when the number of days in each season for which there is data, differs.

* How many days are there in a planting season (15 September – 20 October)? 36 days
* What is the proportion of **green light** days in the 1992 season? of the days.
* In the coolest planting seasons, how many **green light** days are expected?

days (to the nearest day)

* What is the proportion of **amber light** days in the 1992 season? of the days.
* In the coolest planting seasons, how many **amber light** are expected?

days (to the nearest day)

* In the coolest planting seasons, on how many days during the coolest planting seasons are farmers with larger properties likely to sow cotton?

This is the number of green light days plus the number of amber light days, which together is

of the days

days (to the nearest day)

**Task 2 – Constructing and interpreting Venn diagrams and two-way tables**

1) Remind students of the context and what is meant by the various colours of day.

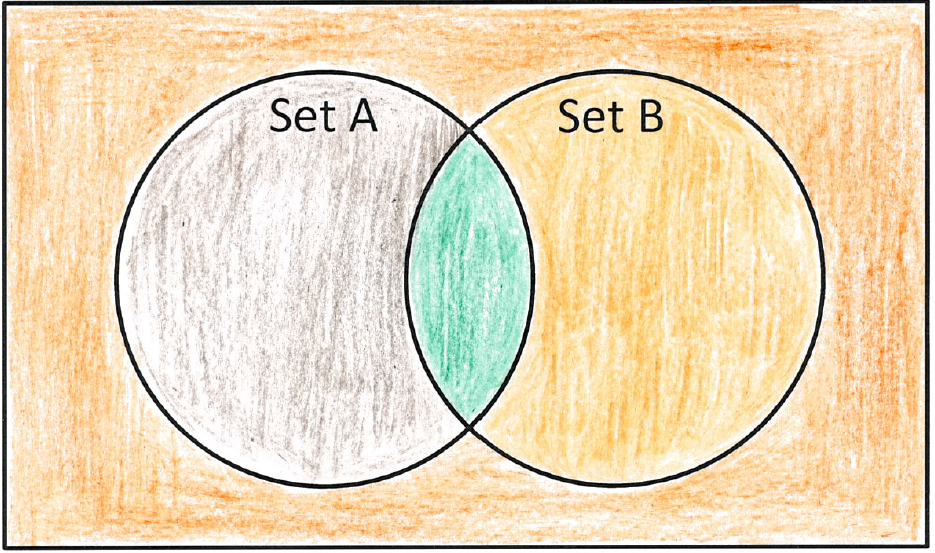
Emphasise the words in italics (eg. both, neither).

2) Check that all students understand what is meant by

* *The condition is satisfied* – the requirements have been fulfilled
* *A set* – a collection of distinct “things” (eg. numbers or objects) that have something in common

3) Give them the Task 2 sheet and ask them to follow the instructions and answer the questions.

Venn Diagrams



The two circles in the Venn diagram intersect (i.e. overlap) because some days satisfy Condition A *and* Condition B and are therefore in both Set A *and* Set B.

* What colour light was given to these days? Green

On some days, *neither* Condition A *nor* Condition B was satisfied.

* What colour light was given to these days? Red

On some days, Condition A *was* satisfied but Condition B *was not* satisfied.

* What colour light was given to these days? Brown

On some days, Condition A *was not* satisfied but Condition B *was* satisfied.

* What colour light was given to these days? Orange

Set Language and Notation

* What is the total number of days represented in the Venn diagram for the 1992 season? 25
* What colour days are in Set B? orange and green
* What does n(B) equal?

The Venn diagram below shows the data for the 2000 planting season.

* Shade it with a pencil to help you find the values of the following:
* Is the rule n(A ∪ B) = n(A) + n(B) - n(A ∩ B) true for the 2000 planting season?

Yes, because 24 = 23 + 14 – 13

Is this rule true for the 1996 planting season?

For the 1996 season: 20 = 9 + 15 – 4

Is this rule true for *any* number of elements in Set A and Set B? Explain why or why not.

Yes. The intersection is doubled when the two sets overlap so if its value is subtracted from the total of the two sets, it becomes the same as just each of the three regions added together.

Would this rule apply if there were *no* **green light** days (i.e. no overlap of the two sets)?

Yes. In this case n(A ∩ B) = 0

2006 planting season:

Set A

Set B

**12**

**0**

**11**

**0**

* How many days are in Set A? 12
* How many days are in Set B? 12

Two-way Tables

Colour the 4 empty cells of the table above as follows.

**Green light** days satisfy Condition A and Condition B, so colour the cell that is in the Set A column *and* in the Set B row, green.

|  |  |  |
| --- | --- | --- |
|  | **Set A** | **Set A’** |
| **Set B** |  |  |
| **Set B’** |  |  |

Colour the remaining 3 cells to indicate the sets in which there are **red light** days, **brown light** days and **orange light** days.

|  |  |  |  |
| --- | --- | --- | --- |
| 1992 | **Set A** | **Set A’** | **Marginal**  **Totals** |
| **Set B** | 4 | 11 | 15 |
| **Set B’** | 5 | 5 | 10 |
| **Marginal**  **Totals** | 9 | 16 | 25 |

* What does the total of Set B mean in this context?

There are 15 days when temperatures were on a rising plane.

* What does the total of Set A’ mean in this context?

There are 16 days when temperatures are not above 14°C.

* What is the sum of the two column totals? 25
* What is the sum of the two row totals? 25
* Why should the sum of the columns totals, be the same as the sum of the row totals?

Because every day has to be either in Set A or not in Set A, and every day has to be in Set B or not in Set B. Both these values are total of all the days (the universal set).

* Using the Venn diagrams for 2000 and 2006, make a two-way table for each of these seasons.

|  |  |  |  |
| --- | --- | --- | --- |
| 2000 | **Set A** | **Set A’** | **Marginal**  **Totals** |
| **Set B** | 13 | 1 | **14** |
| **Set B’** | 10 | 0 | **10** |
| **Marginal**  **Totals** | **23** | **1** | **24** |

|  |  |  |  |
| --- | --- | --- | --- |
| 2006 | **Set A** | **Set A’** | **Marginal**  **Totals** |
| **Set B** | 12 | 0 | **12** |
| **Set B’** | 11 | 0 | **11** |
| **Marginal**  **Totals** | **23** | **0** | **23** |

* Which of the three seasons (1992, 2000 or 2006) has the greatest proportion (i.e. fraction) of days in Set B?

1992:

2000:

2006:

1992 has the greatest proportion of days when the following week had temperatures on a rising plane.

* Did you expect all the seasons to have more than half of their days in Set B? Why or why not?

Yes. Because September and October are months of spring, temperatures overall are expected to be rising.

**Task 3 – Calculating probability and conditional probability**

1) Remind students of the context and what is meant by the various colours of day.

Emphasise the words in italics (eg. both, neither).

2) Check that all students understand the language of probability.

* *Event*
* *Outcome*
* *Frequency*
* *Sample space*
* *Relative frequency*
* *Experimental probability*

3) Introduce the words *mutually exclusive* using examples of mutually exclusive events and mutually exclusive outcomes.

eg. If throwing a dice is one event and drawing a card is another event, these *events* have no common elements (eg. you can’t draw an ace by throwing a die) .

eg. For the event of throwing a die once, the *outcomes* of throwing an odd number and the outcome of throwing an even number are mutually exclusive (because no number is both odd and even).

Ask students to find their own examples of mutually exclusive events and mutually exclusive outcomes.

4) Give students the sheet for Task 3. Ask them to read the first page and then answer the questions about probabilities on page 2.

Probabilities

* Are the two events (a day when soil temperature is greater than 14°C and a day when average temperatures for the following week are on a rising plane) mutually exclusive? Why or why not?

No. These events are not mutually exclusive because on the the same day there can be a soil temperature is greater than 14°C and average temperatures for the following week on a rising plane

* For Event A (whether soil temperature is greater than 14°C) what are the two possible outcomes?

The two outcomes are:

* A soil temperature greater than 14°C
* A soil temperature *not* greater than 14°C (i.e. less than or equal to 14°C)

Are these outcomes, mutually exclusive? Why or why not?

Yes, because a temperature cannot be both greater than 14°C and less than or equal to 14°C.

* How many days are in the sample space? 25
* Calculate the probabilities.

4) Ask them to read page 3 which is about conditional probabilities.

Work through this page with them using the table below and the Venn diagram for the 1992 season, so they understand the meaning of:

* *The sample space and a restricted sample space*
* *Marginal total*
* *Dependent and independent events*
* *Conditional probability*
* *Initial condition*

|  |  |  |  |
| --- | --- | --- | --- |
| 1992 | **Set A** | **Set A’** | **Marginal**  **Totals** |
| **Set B** | 4 | 11 | **15** |
| **Set B’** | 5 | 5 | **10** |
| **Marginal**  **Totals** | **9** | **16** | **25** |

Set A

Set B

**4**

**5**

**5**

**11**

5) Get students working in pairs to answer the questions on page 4.

Conditional Probabilities

* With a partner, decide whether the following sentences are true or false, i.e. whether they mean the same as *The probability that a day is in Set B, given that day is in Set A is* .

1) If a day is in Set B, then the probability that it is in Set A is True

2) If a day is in Set A, then the probability that it is in Set B is False

3) Given that a day is in Set B, the probability that it is in Set A is True

4) Knowing that a day is in Set B, the probability that it is also in Set A is False

5) If you already know that a day is in Set A, the probability that it is in Set B is True

* What is meant by ?

This is the probability that a day will have a rising plane of temperatures the following week, if you already know that the soil temperature that day is less than or equal to 14°C.

* What is meant by ?What is the value of in the 1992 season?

This is the probability that the soil temperature on a day is greater than 14°C , if you already know that there will a rising plane of temperatures over the following week.

* What is meant by ?What is the value of in the 1992 season?

This is the probability that the soil temperature on a day is greater than 14°C , if you already know that there will a rising plane of temperatures over the following week.

* With a partner, decide whether the pairs of events below are dependent or independent.

Read information at <http://cottonaustralia.com.au/australian-cotton/basics/cotton-facts>.

Then share your answers with others so you can convince them or be convinced by them.

(i) Whether a cotton crop is genetically modified, and the quantity of insecticide used on it.

On the webpage, if students click the link *Biotechnology and Cotton*, they will find that:

* The use of biotechnology in cotton has made a significant contribution in the dramatic reduction in insecticides applied to Australian cotton crops.
* Australian cotton growers have reduced their insecticide use by 89% over the last decade, with some crops not sprayed for insects at all.

Genetically modified cotton crops are crops where the cotton variety has been given a gene that reduces its susceptibility to insects, thereby reducing the amount of insecticide needed to control the insects that attack it.

Therefore the two events are dependent.

(ii) Whether an area is sown to cotton or left fallow, and the emission of greenhouse gases.

On the webpage, if students click the link *Climate Challenges and Cotton*, they will find that:

* Cotton growing has a better-than-neutral carbon footprint. Net on-farm emissions of greenhouse gases on cotton farms are negative because the cotton plants store more carbon than is released from production inputs used during growth.

“Left fallow” means that nothing is grown on the land.

When a crop grows, it uses the process of photosynthesis to take in carbon dioxide from the atmosphere and produce plant material. This is why an area sown to cotton takes in more carbon than is emitted by the processes of growing it (such as the energy used to plant and fertilise it).

Therefore the two events are dependent.

(iii) The weight of a bale of cotton, and what that bale of cotton is used to make.

On the webpage, if students click the link *Properties and Cotton Products*, they will find

* The cotton lint from one 227 kg bale can produce 215 pairs of denim jeans, 250 single bed sheets, 750 shirts, 1,200 t-shirts, 3000 nappies, 4,300 pairs of socks, 680,000 cotton balls, or 2,100 pairs of boxer shorts

227 kg is the statistical average weight of a bale of cotton in Australia. In other countries, the weight is different eg. In South Africa, cotton bales weigh about 200 kg. The weight of the bale has no effect on what the cotton is used to make.

Therefore the two events are independent.

(iv) Rain occurring on a day at Trangie, and rain occurring on the same day at Narromine.

If you find these two places on a map of NSW you will notice that they are close to each other. They are only 34 km apart by road, so if it is raining at one of these places it is likely to be raining at the other.

Therefore the two events are dependent.

6) Revise the topic of probability trees. Then ask students to answer the questions on Pages 5 & 6.

Probability Trees

* What probabilities should be written on the branches that split from the initial bottom branch?

Calculate their values. Then write them on these two branches.

The probabilities on the bottom two branches are *not* the same as the probabilities on the top two branches because Event A and Event B are *not* independent events.

Yes

No

**Event A**

Yes

No

Yes

No

**Event B**

**Outcome**

**Green light** day

**Brown light** day

**Orange light** day

**Red light** day

* The branches of the **green light** day outcome are and

Calculate the probability of a **green light** day using these two values.

* Using the tree diagram , write the calculations needed to find:
* the probability of a **brown light** day in the 1992 season
* the probability of an **orange light** day in the 1992 season.
* the probability of a **red light** day in the 1992 season.

The two-way table of frequencies in 2000 (a fairly typical planting season) is shown below.

* Draw a probability tree and calculate the probability of each of the four outcomes.

|  |  |  |  |
| --- | --- | --- | --- |
| 2000 | **Set A** | **Set A’** | **Marginal**  **Totals** |
| **Set B** | 13 | 1 | **14** |
| **Set B’** | 10 | 0 | **10** |
| **Marginal**  **Totals** | **23** | **1** | **24** |

Yes

No

**Event A**

Yes

No

Yes

No

**Event B**

**Outcome**

**Green light** day

**Brown light** day

**Orange light** day

**Red light** day

* For this season, could you have drawn a probability tree with fewer branches?

Yes. The last of the 4 branches isn’t needed because it has a zero probability.

Yes

No

**Event A**

Yes

No

Yes

**Event B**

**Outcome**

**Green light** day

**Brown light** day

**Orange light** day

The two-way table of frequencies in 2006 (the warmest of planting season) is shown below.

* Draw a probability tree and calculate the probability of each of the four outcomes.

|  |  |  |  |
| --- | --- | --- | --- |
| 2006 | **Set A** | **Set A’** | **Marginal**  **Totals** |
| **Set B** | 12 | 0 | **12** |
| **Set B’** | 11 | 0 | **11** |
| **Marginal**  **Totals** | **23** | **0** | **23** |

Yes

No

**Event A**

Yes

No

Yes

No

**Event B**

**Outcome**

**Green light** day

**Brown light** day

**Orange light** day

**Red light** day

0

0

* For this season, could you have drawn a probability tree with fewer branches?

Yes. Because Event A is certain i.e. P(A)= 1, only Event B needs to be considered.

Yes

No

**Event B**

**Outcome**

**Green light** day

**Brown light** day

* A farmer with a small property is interested in the probability of a **green light** day occurring.

Does this probability vary much between seasons?

1992:

2000:

2006:

There is a large variation due to the number of days with a temperature above 14°C increasing greatly from the coolest season (1992) to an average type of season (2000). For warmer than average seasons, little variation would be expected.

* A farmer with a large property is interested in the probability of a **green light** *or* a **brown light** *or* an **orange light** day occurring as he will sow on any of these days.

Calculate the probability of this occurring on a day in the 1992 season, the 2000 season and the 2006 season. Does this probability vary much between seasons?

This is the same as the complement of the probability of a red light day occurring.

1992:

2000:

2006:

This probability varies a lot from the coolest season to the average season. It is unlike to vary much once when seasons are warmer than average because in these seasons Condition A (soil temperature above 14°C) is always (or nearly always) satisfied.

* If the average soil temperature and the average air temperature over a planting season at Trangie increase in the future,
* would you expect the probability of **green light** days to increase or decrease? Why?
* would you expect the probability of **red** **light** days to increase or decrease? Why?

Soil temperatures increasing will increase the probability of green light days and brown light days. However there could still be orange light days because whether a temperature is on a rising plane or not over the next week, is *not* greatly affected by the temperature of the day. There can still be a very hot day followed by a week of decreasing temperatures.