**Diabetes**

Source: Pfannkuch, M., Seber, G., & Wild, C.J. (2002). Probability with less pain.

*Teaching Statistics,* 24(1) 24-30

A resource aimed at:

 Year 13 probability and probability distributions

connecting experimental (simulation - Tinkerplots) and theoretical approaches to probability

using a simulation model to ask ‘what if’ questions

using the Normal distribution, table of counts, conditional probability

Tinkerplots: <http://www.keycurriculum.com/products/tinkerplots>

The task could be adapted and used without Tinkerplots simulation

Teacher version

Diabetes is a major disease in New Zealand and is of concern because of its rising incidence and hence cost to the health system as well as personal costs.

Introduction:

Ask class about what they know of diabetes.

Show students either of these clips

<http://www.youtube.com/watch?v=4EEtubB74lM>

<http://www.youtube.com/watch?v=Qi6LYIhlFdw>

<http://www.youtube.com/watch?v=MGL6km1NBWE>

and/or read newspaper articles:

NZ's soaring diabetes rate puts us with world's worst

<http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=10734691>

Health professionals warn of impending diabetes crisis

<http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=10744635>

A standard test for diabetes is based on glucose levels in the blood after fasting for a prescribed period. For ‘healthy’ people, the mean fasting glucose level is 5.31 mmol/L and the standard deviation is 0.58 mmol/L. For untreated diabetes the mean is 11.74 and the standard deviation is 3.50. In both groups the levels appear approximately Normal.

Get students to graph these Normal distributions N(5.31, 0.58) and N(11.74, 3.50) on the same axis using information from graphic calculator. They can find the heights of the graph at different x values using their calculator – Dist Norm Ndp.

Distribution of blood glucose for undiagnosed diabetics

Distribution of blood glucose for healthy people

C

Discuss the appearance of these distributions: shape/ spread/ overlap.

*Healthy people- blood glucose effectively ranges from about 2.75 to 7.5 mmol/L*

*Graph is tall and skinny – so small variation in glucose level (small sd)*

*Mean glucose level is about 5.3 mmol/L. Approx 95% of people would have blood glucose levels between about 4 and 6.5 mmol/L*

*People with diabetes: graph is flat and wide. Blood glucose effectively ranges from about 2 to 21 mmol/L. There is a large variation in blood glucose levels for diabetics. (sd large). 95% of people would have blood glucose levels between about 4.5 and 18mmol/L. Mean glucose level is about 11.4 mmol/L*

*People with diabetes tend to have a higher and more varied glucose level than non-diabetics. The difference in the mean values is about 6 mmol/L*

Discuss where they might set the cut-off point, C, (in mmol/L) using this test. If

a person’s fasting glucose level is above *C* they ‘test positive’ and we say that they have diabetes, while if it is below *C* they ‘test negative’ and we say that they do not have diabetes.

What are the implications for people close to the cut-off point?

*Some people who do not have diabetes are classified as having diabetes- a false positive test result/ some people who have diabetes are classified as not having diabetes- a false negative test result.*

*If a person is diabetic but the test is negative, they will not receive the treatment that they need. Diabetes does have long term effects such as eye, kidney and heart damage. One would hope that eventually they would be correctly diagnosed before these effects became severe. Diabetes also has immediate effects such as trembling, sweating, thirst, moodiness, tiredness that a person could suffer without correct treatment. These symptoms would possibly aid diagnosis, but we are considering the case of a mass screening programme, where presumably people are not displaying strong symptoms.*

*If a person is not diabetic but the test is positive, they may receive treatment that they do not need, and such a diagnosis would make the person anxious about their health. Also the treatment and follow up screening is an unnecessary cost on the health system.*

Shade regions on the graph to indicate proportions of people who would be wrongly

diagnosed.



6.7% of people have diabetes but test is negative, based on C = 6.5

2% of people do not have diabetes but test is positive, based on C= 6.5

The *sensitivity* of such a test is the probability of the test giving a positive result when the person has the condition. (true positive)

The *specificity* of such a test is the probability of the test giving a negative result when the person does not have the condition. (true negative)

For example, if C = 6.5 mmol/L get students to calculate:

P(test is negative | person does not have diabetes) = 0.980 [N(5.31, 0.58); P(X < 6.5) = 0.980]

P(test is positive | person has diabetes) = 0.933

[N(11.74, 3.50); P(X > 6.5) = 0.933]

 and add this information to their graph

In 2012, 225 686 people had been diagnosed with diabetes. (<http://www.diabetes.org.nz/about_diabetes>)

The population estimate was 4 433 000. (Statistics NZ)

This gives a base rate (proportion of the population with diabetes) of about 5%.

Suppose there was a screening programme introduced where the entire population of New Zealand was tested for diabetes using this test and the cut-off point was taken as 6.5mmol/L

Set up a Tinkerplots simulation based on this information.

* Drag a sampler onto the page – this part is about the probability that something happens.

To represent that a person has or has not got diabetes (base rate = 5%) use a two-bar device Label one bar ‘diabetes’ and one ‘not’. Down arrow – show percent/proportion – change the ‘not’ bar to 95%/0.95.

* Drag two more bar devices into the sampler, one above the other. (Pink dots will appear in the sampler indicating where the new device can go – when you hover around the particular pink dot a black rectangle will appear and you put the new device in the rectangle). Use the ‘+’ button to make 2 bars in each. Label the bars ‘positive’ and ‘negative’

Change Attr1 to ‘disease’ andAttr2 to ‘test’

Change the heights of the bars to represent the probabilities.

Run to check that your model is working as you think it should.

* Highlight the column labelled ‘Join’ and drag a plot onto the page. Pull a circle in the plot to the right to separate.

Stack.

Count (N)

Change the repeat number to something sensible.

Students should experiment to find how many simulations is ‘sensible’. They should compare their results with others to see the consistency or variability of their results.

Run



*The table of counts for this simulation is [each student should get different, but similar, results from their simulation. These are here for a guide]:*

|  |  |  |
| --- | --- | --- |
|  |  | *Test result* |
| *Disease status* |  | *Positive* | *Negative* |
| *Has diabetes* | *4797* | *343* |
| *Does not have diabetes* | *2005* | *92855* |

*P(does not have diabetes | test positive) = 2005/6802 = 0.2948*

 *P(has diabetes | test negative) = 343/93198 = 0.0037*

*About 29% of those that test positive do not have diabetes – This is the false positive result.*

*Just under 4% of those that test negative do have diabetes. This is the false negative result.*

*2.3% of people would get an incorrect result (2348/100000).*

1. Use the model you have created in Tinkerplots to investigate the effect of different cut-off values.

Students should experiment with different cut-off values of their own choice; eg -

1. Base rate = 5%; cut-off = 6 mmol/L

P(test is negative | person does not have diabetes) = 0.883

P(test is positive | person has diabetes) = 0.949

|  |  |  |
| --- | --- | --- |
|  |  | Test result |
| Disease status |  | Positive | Negative |
| Has diabetes | 4721 | 259 |
| Does not have diabetes | 11136 | 83884 |

Use your table to estimate:

P(does not have diabetes | test positive) = 11136/15857 = 0.702

70% of those that test positive do not have diabetes.

P(has diabetes | test negative) = 259/84143 = 0.003

0.3% of those that test negative do have diabetes.

*11% of people would get an incorrect result (*11395/100000*).*

b). Base rate = 5%; cut-off = ……7… mmol/L

P(test is negative | person does not have diabetes) = 0.998

P(test is positive | person has diabetes) = 0.912

|  |  |  |
| --- | --- | --- |
|  |  | Test result |
| Disease status |  | Positive | Negative |
| Has diabetes | 4557 | 451 |
| Does not have diabetes | 215 | 94777 |

Use your table to estimate:

P(does not have diabetes | test positive) = 215/4772 = 0.045

4.5% of those that test positive do not have diabetes.

P(has diabetes | test negative) = 451/95228 = 0.005

0.5% of those that test negative do have diabetes

*0.7% of people would get an incorrect result (*666/100000*).*

c). Base rate = 5%; cut-off = ……… mmol/L

P(test is negative | person does not have diabetes) =

P(test is positive | person has diabetes) =

|  |  |  |
| --- | --- | --- |
|  |  | Test result |
| Disease status |  | Positive | Negative |
| Has diabetes |  |  |
| Does not have diabetes |  |  |

Use your table to estimate:

P(does not have diabetes | test positive) =

* 1. P(has diabetes | test negative)

Recommend a better cut-off value than 6.5 mmol/L. Explain why you would recommend this cut-off value.

Class completion to get the best cut off point in terms of minimising the percentage of incorrect diagnoses.

Class discussion as to their simulation results- is it better to have more false positives or more false negatives?

Teaching point: - the language of conditional probability

P(test is negative | person does not have diabetes)

P(test is positive | person has diabetes)

as opposed to

P(has diabetes | test is negative)

P(does not have diabetes | test is positive)

1. Investigate what happens if the base rate is different than 5%. For instance, in New Zealand the base rate for diabetes varies for different ethnic groups and ranges between 1 and 10%.

The 2002–2003 New Zealand Health Survey of people over fifteen years of age estimated that the self-identified prevalence of diagnosed diabetes within each ethnicity was as follows: Maori 8.0%, Pacific 10.1%, Asian 8.4% and European (including ‘Other’) 2.9%, so disparities in prevalence exist [5].Ethnic disparities have also existed in the management and complications of diabetes in New Zealand [6].

<http://www.diabetes.org.nz/news/nz_news/trends_in_the_management_of_risk_of_diabetes_complications>

1. Base rate = …10.1%…%; cut-off = …7…..mmol/L (cut off should be the best result that the class has found – 7 is used here as an example)

P(test is negative | person does not have diabetes) = 0.998

P(test is positive | person has diabetes) = 0.912

|  |  |  |
| --- | --- | --- |
|  |  | Test result |
| Disease status |  | Positive | Negative |
| Has diabetes | 9089 | 928 |
| Does not have diabetes | 173 | 89810 |

Use your table to estimate:

P(does not have diabetes | test positive) = 173/9262 = 0.019

1.9% of those that test positive do not have diabetes.

P(has diabetes | test negative) = 928/90738 = 0.010

1.0% of those that test negative do have diabetes

*1.1% of people would get an incorrect result (*1101/100000*).*

b). Base rate = 2.9…….. % cut-off = …7…… mmol/L

P(test is negative | person does not have diabetes) =

P(test is positive | person has diabetes) =

|  |  |  |
| --- | --- | --- |
|  |  | Test result |
| Disease status |  | Positive | Negative |
| Has diabetes | 2686 | 246 |
| Does not have diabetes | 197 | 96871 |

Use your table to estimate:

P(does not have diabetes | test positive) = 197/2883 = 0.068

6.8% of those that test positive do not have diabetes.

P(has diabetes | test negative) = 246/97117 = 0.003

0.3% of those that test negative do have diabetes

*0.4% of people would get an incorrect result (*443/100000*).*

Describe the effect of different base rates on the percentage of people who would get incorrect diagnoses.
*Based on a 7 mmol/L cut-off value-*

*Higher base rate gives greater proportion of people with incorrect diagnosis compared to lower base rate. (1.1%; 0.4%).*

*Higher base rate gives greater proportion of false negative results than lower base rate (1.9%, 1%)*

*Higher base rate gives smaller proportion of false positive results than lower base rate (1.9%, 6.8%)*

1. How do we estimate the base rate? (Extension- This may be appropriate only for students who are good algebraists.)

We want to estimate ***d*** = P(a person in New Zealand has diabetes)

To find a good estimate of the proportion of people in the population who test positive for diabetes, we collect data from a random sample of people. Suppose that 10.3% of such a sample test positive. Use the cut-off value, C =6.5 mmol/L.

Base rate = ***d*** cut-off =6.5mmol/L

P(test is negative | person does not have diabetes) = 0.98 [from task 1]

P(test is positive | person has diabetes) = 0.933

Fill in the table and then calculate the base rate, d.

|  |  |  |
| --- | --- | --- |
|  |  | Test result |
| Disease status |  | Positive | Negative | Total |
| Has diabetes | 0.933d | 0.067d | d |
| Does not have diabetes | 0.02(1 – d) | 0.98(1 – d) | 1 - d |
| Total | 0.103 | 0.897 | 1 |

Now solve an equation to calculate ***d.***

*0.933d + 0.02(1 – d) = 0.103*

 *d = 0.0909*

Conclusion:

If 10.3% of a random sample of New Zealanders tested positive for diabetes then about 9% of the population are estimated to have the disease.

References:

* + 1. <http://www.health.govt.nz/your-health/conditions-and-treatments/diseases-and-illnesses/diabetes>
		2. <http://www.diabetes.org.nz/about_diabetes>
		3. <http://www.diabetes.org.nz/news/nz_news/trends_in_the_management_of_risk_of_diabetes_complications>



This work is supported by:

The New Zealand Science, Mathematics and Technology Teacher Fellowship Scheme which is funded by the New Zealand Government and administered by the Royal Society of New Zealand

and



Department of Statistics

The University of Auckland