Getting *hands-on* with statistics

Communicating Statistics (9.6): STEM Showcase

Simon R. White

Education and Statistical Literacy Committee, Royal Statistical Society

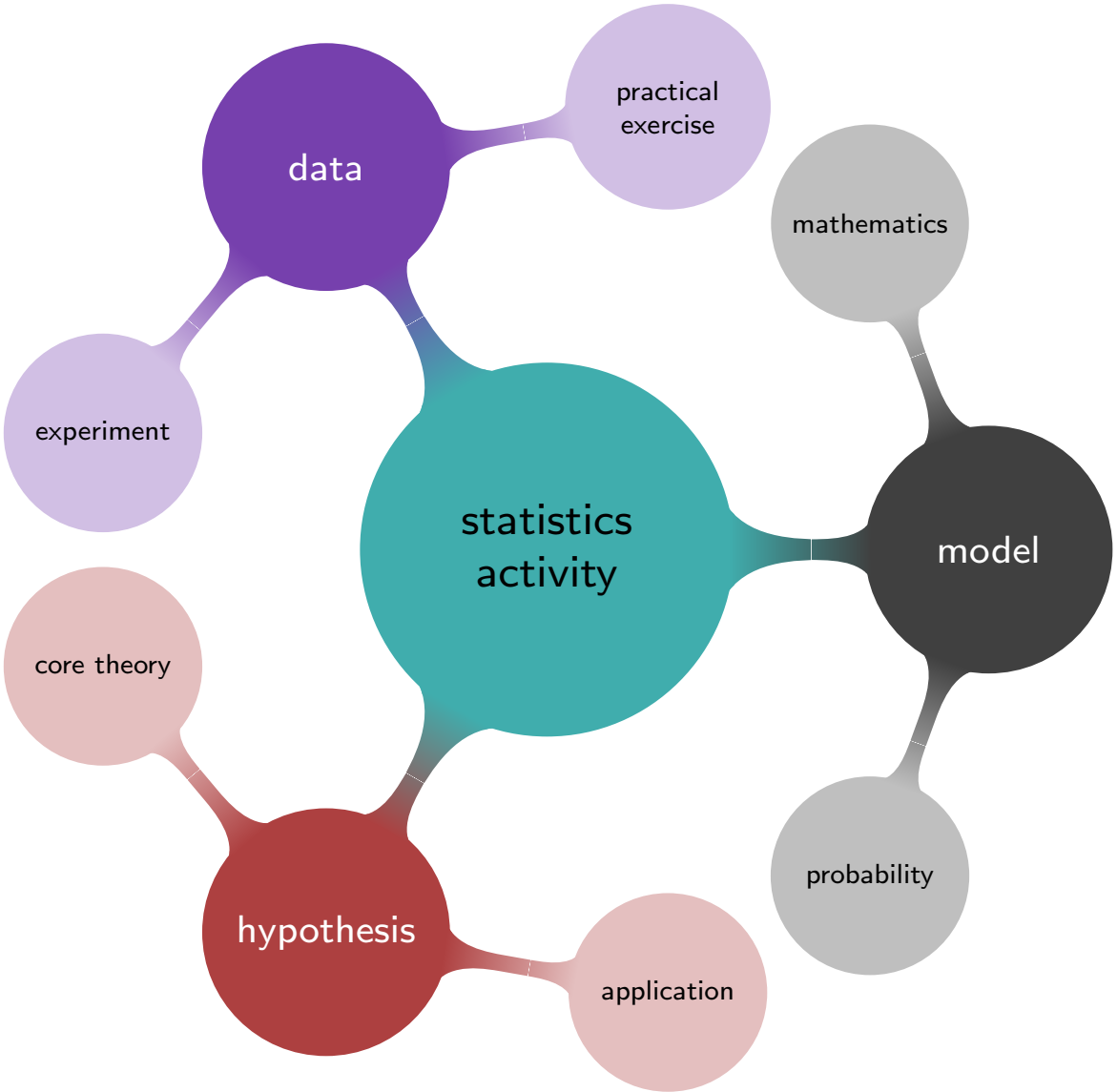
MRC Biostatistics Unit, University of Cambridge

Department of Psychiatry, University of Cambridge

5th September 2019

RSS2019 Conference

What makes a statistics activity?



Statistics and society

Wide range of contexts and topics to build an activity around

Department for Transport

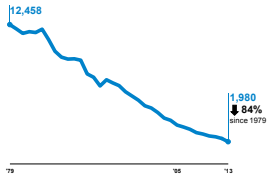
Facts on Child Casualties

June 2015

Overview

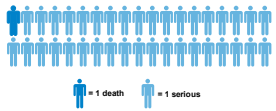
Children under the age of 16, are one of the most vulnerable road users, in particular child pedestrians which account for the majority of child casualties across all road user types. In contrast to older pedestrians, child pedestrians are not experienced and well educated about using the road. They therefore, are particularly vulnerable to injuries and accidents.

Killed or seriously injured, 1979-2013



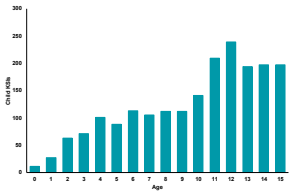
Average per week, 2013

An average of 1 death and 37 seriously injured children per week



Child KSIs by age, 2003-2013

The number of child KSI casualties increases with age



Child KSIs by time of day, 2003-2013

More children are killed or seriously injured on their way home from school than on their way to school....

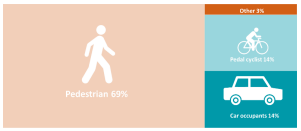
58% of child KSI casualties occur between 3 and 6:59pm

13% of child KSI casualties occur between 8 and 8:59am

28% of child KSI casualties occur during other times of the day

Child KSIs by road user type, 2013

Child pedestrians account for the majority of child KSI casualties



Contributory factors allocated

78% of child pedestrians involved in accidents failed to look properly.

38% of child pedestrians involved in accidents were careless, reckless or in a hurry.

Email: roadacc.stats@dft.gsi.gov.uk Contact: 020 7944 6595 Press enquiries: 020 7944 2813

1

BBC Sign in News Sport Weather iPlayer TV Radi

NEWS

Home UK World Business Politics Tech Science Health Education Entertain

Education School Report Global Education

Private schools extend lead in entrants to university

By Hannah Richardson
BBC News education reporter

4 August 2016 Education & Family Share

Universities Minister Jo Johnson says all students should be able to fulfil their promise

Private schools extended their lead over state schools in getting pupils into England's universities after higher fees were introduced, data shows.

Government figures show 85% of private school pupils went to higher education, compared with 62% of those from state schools by the age of 19 in 2013-14.

Statistics and society

Wide range of contexts and topics to build an activity around

Department for Transport

Facts on Child Casualties

June 2015

Overview

Children under the age of 16, are one of the most vulnerable road users, in particular child pedestrians which account for the majority of child casualties across all road user types. In contrast to older pedestrians, child pedestrians are not experienced and well educated about using the road. They therefore, are particularly vulnerable to injuries and accidents.

Killed or seriously injured, 1979-2013

Average per week, 2013

An average of 1 death and 37 seriously injured children per week

Mail Online

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Money

Latest Headlines | Health | Health Directory | Health Boards | Diets

How running is more risky than boxing or rugby: Sports that are perceived as dangerous actually have fewer injuries

- Research was conducted by UK health and wellbeing provider Benenden
- It found that football was sport most likely to result in a participant's injury
- Running came second, topping sports such as boxing, rugby and tennis

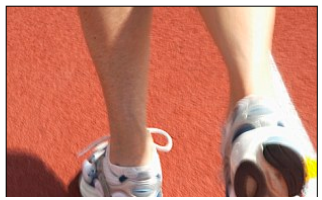
By DAILY MAIL REPORTER

PUBLISHED: 00:31, 5 August 2016 | UPDATED: 10:56, 5 August 2016

Share 154 shares 85 View comments

They're generally regarded as being among the most 'dangerous' sports which cause one injury after the next.

But people are more likely to suffer injuries while playing football or running than they are while boxing or horse riding.



BBC Sign In News Sport Weather iPlayer TV

NEWS

Home UK World Business Politics Tech Science Health Education

Education School Report Global Education

The Telegraph

Home Video News World Sport Business Money Comment Culture Travel Life

Politics Investigations Obits Education Science Earth Weather Health Royal

Current Account Starting from the date you open the account with us

Representative example After the fee-free period, if you use an arranged overdraft of £1,200 the amount we will charge you is 50p per day (variable)

HOME » NEWS » HEALTH » HEALTH NEWS

How teenage pregnancy collapsed after birth of social media

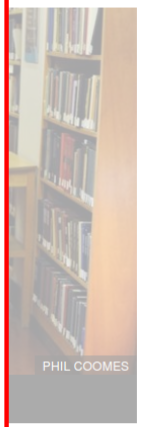
Have Facebook and Snapchat helped stop teenagers having sex? New figures show teenage pregnancies plunging to record low since social media explosion

3K 0 126 3K Email

Are teenagers simply spending more time at home? Photo: © Nick Moore / Alamy Stock Photo

Tony Jacklin's stunning win at Hazeltine

Sponsored by Standard



PHIL COOMES

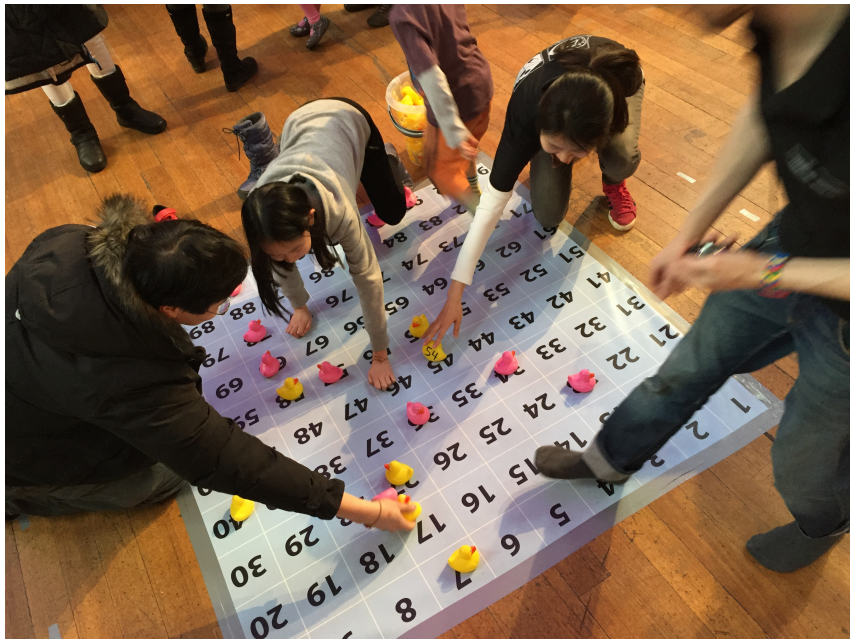
Statistics and STE(A)M

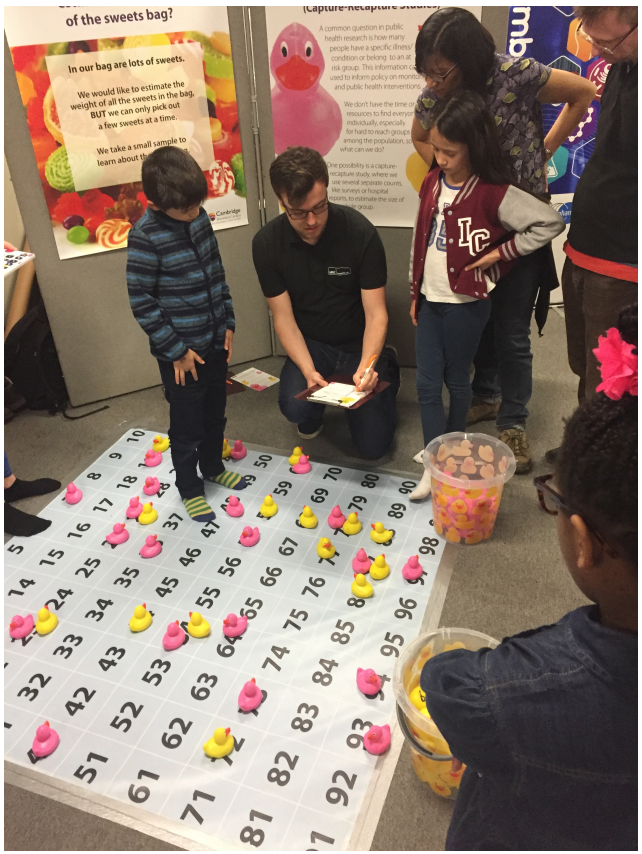
As an extension to many other subjects

Name a subject where we can't talk about statistics?

Counting Ducks

Capture Recapture



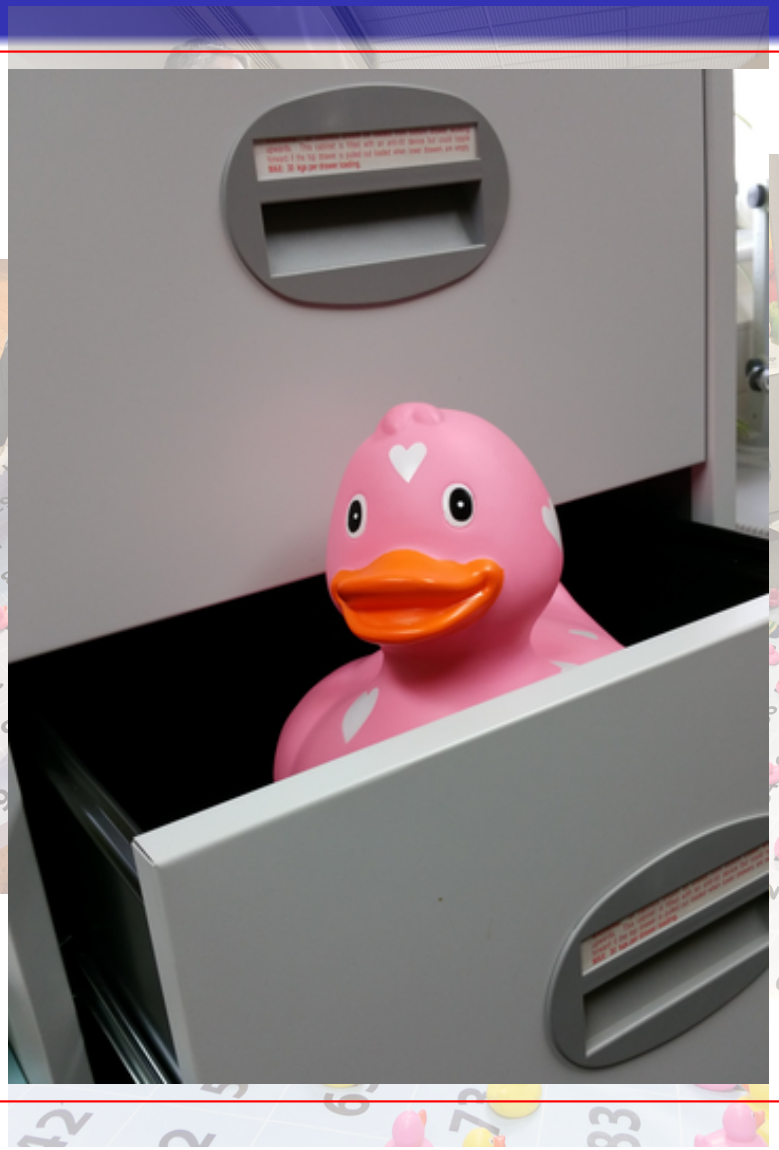


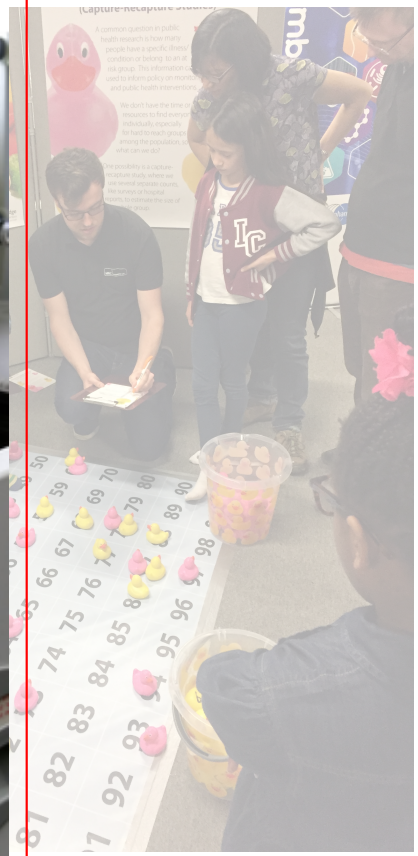
Counting Ducks

Capture Recapture

Counting Ducks

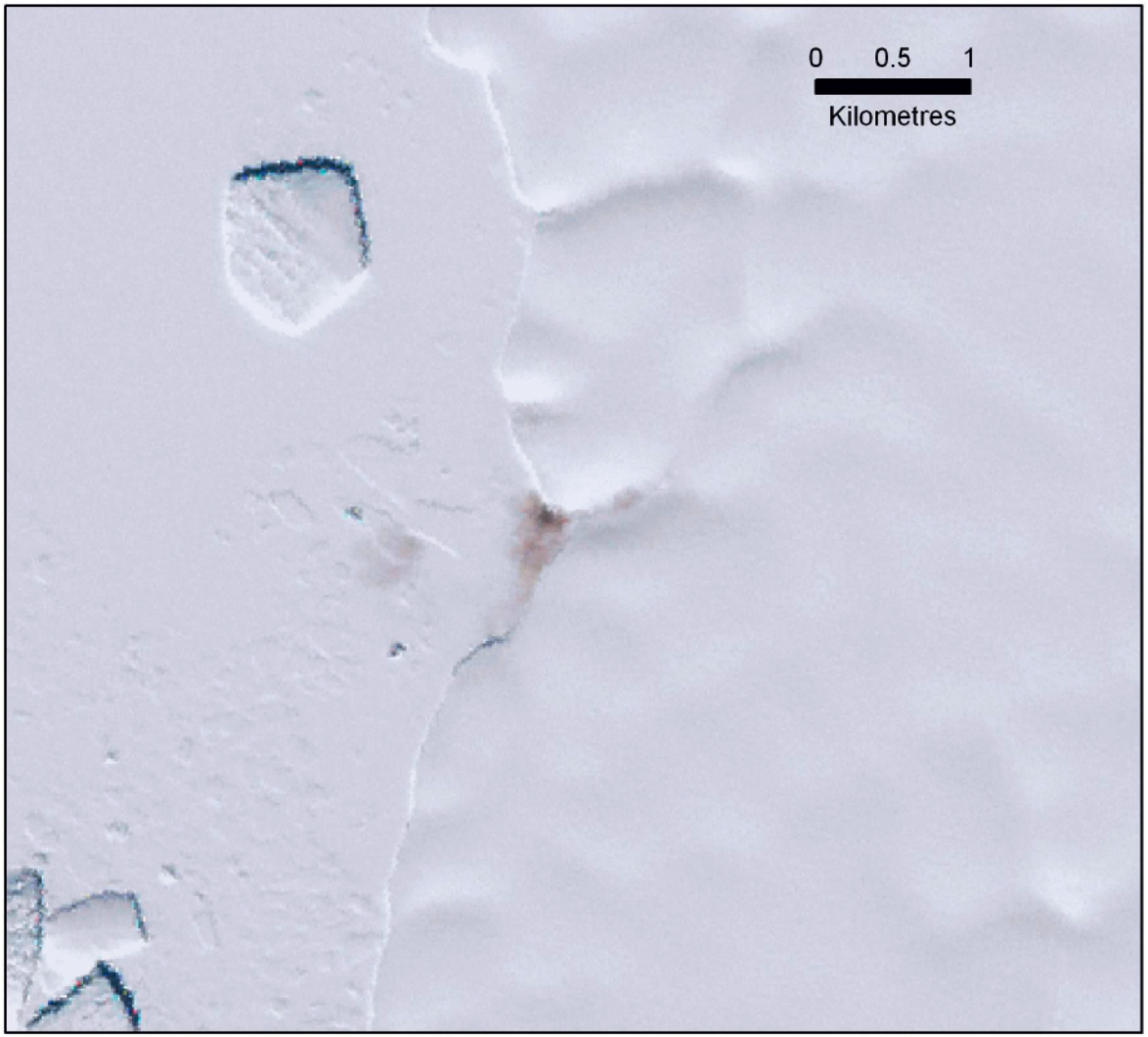
Capture Recapture





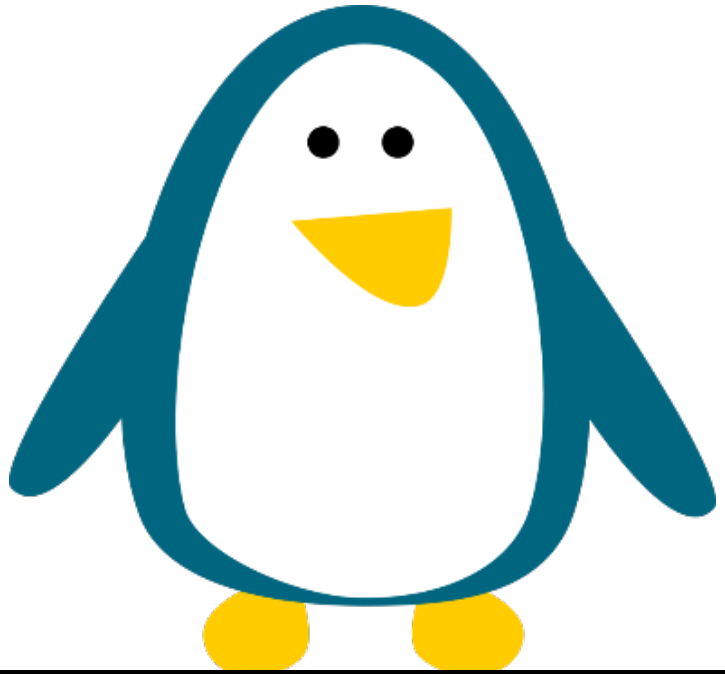
Counting penguin populations

View from above



A medium-resolution Landsat image of an emperor penguin colony in West Antarctica. At this resolution colonies appear as brown stains on the ice. Individuals or groups of birds cannot be differentiated from the surrounding guano, but these satellite images are an ideal tool for searching for undiscovered colony locations as they are freely available and cover large areas. Credit: British Antarctic Survey

Sample from closer to home

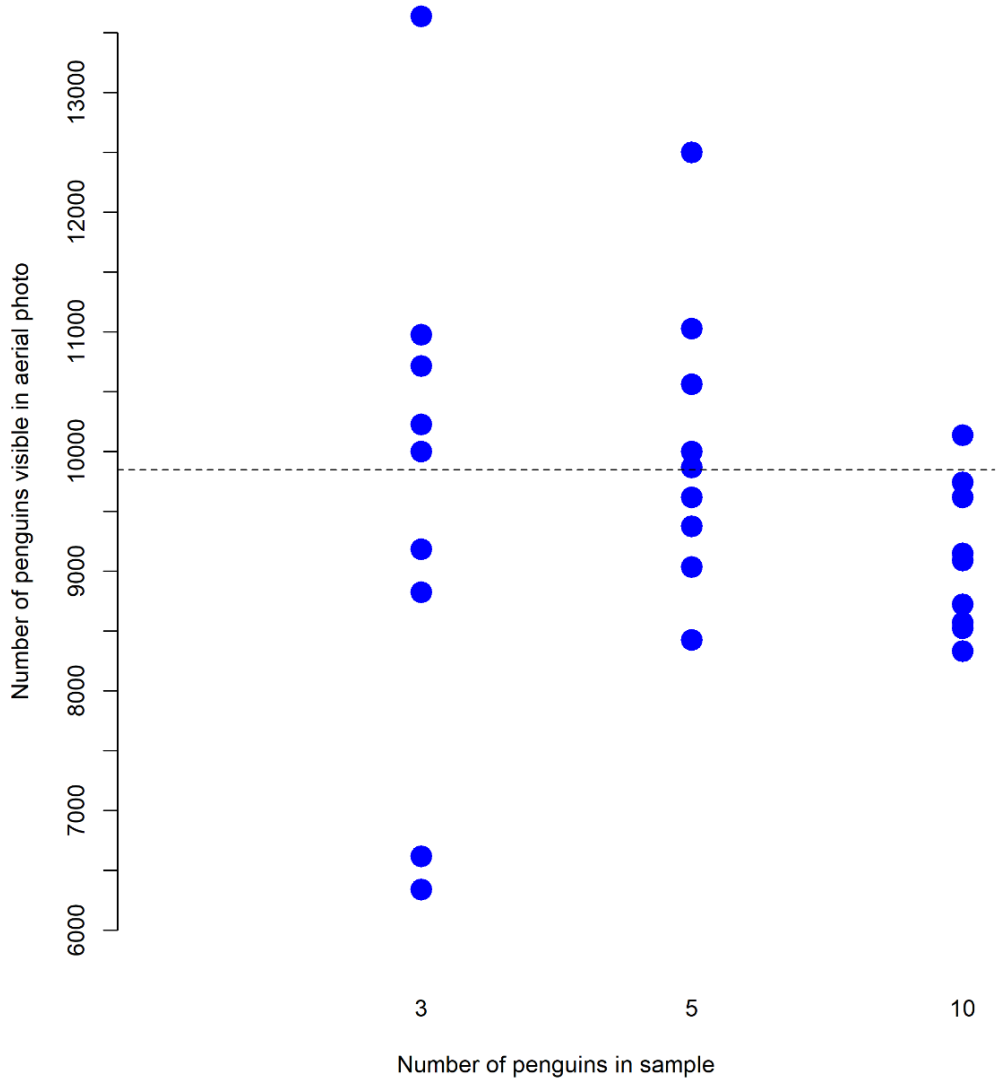


Estimation

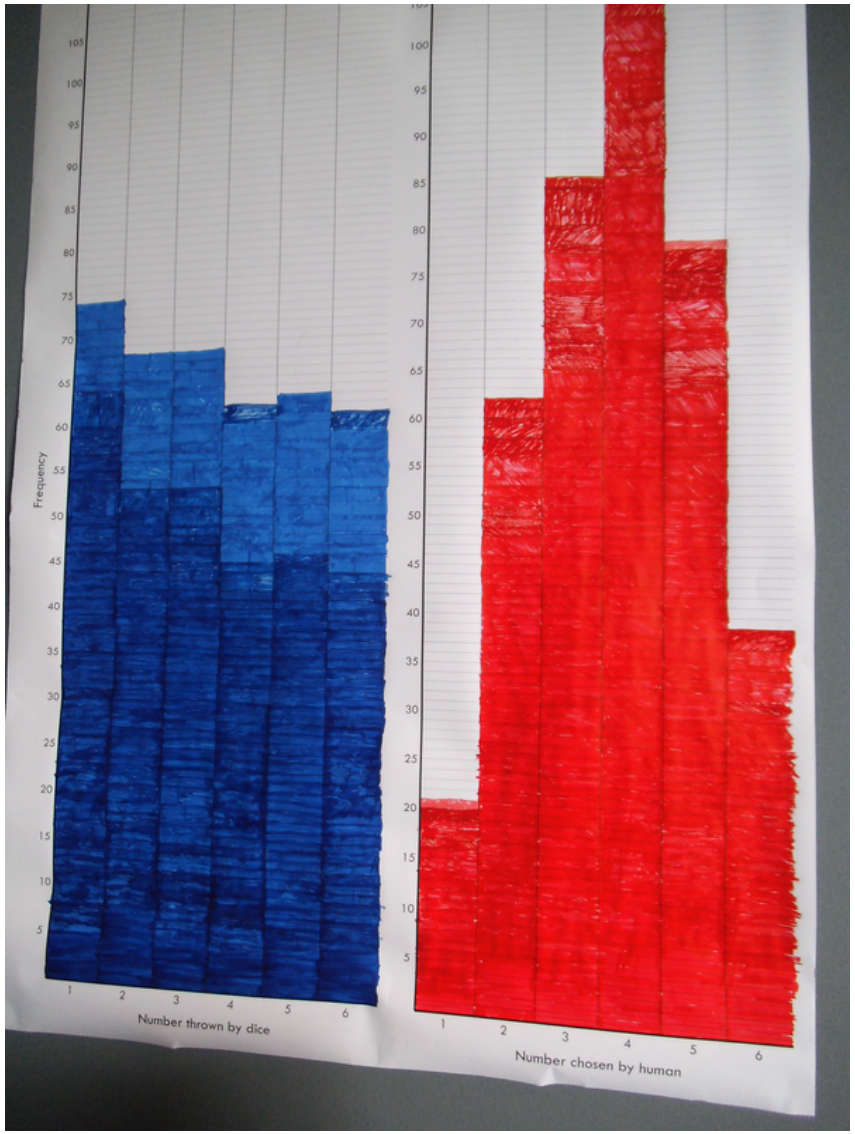
$$\text{Mean amount of guano} = \frac{\text{Penguin 1's guano} + \text{Penguin 2's guano} + \cdots + \text{Penguin n's guano}}{n}$$

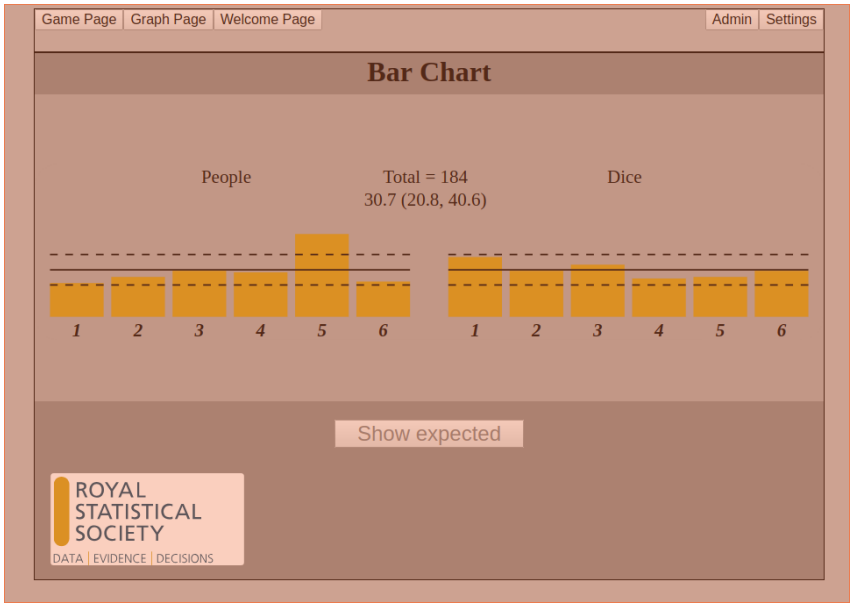
$$\text{Colony size} = \frac{\text{Surface area of guano of photo}}{\text{Mean amount of guano}}$$

Sample variability



How random are you?

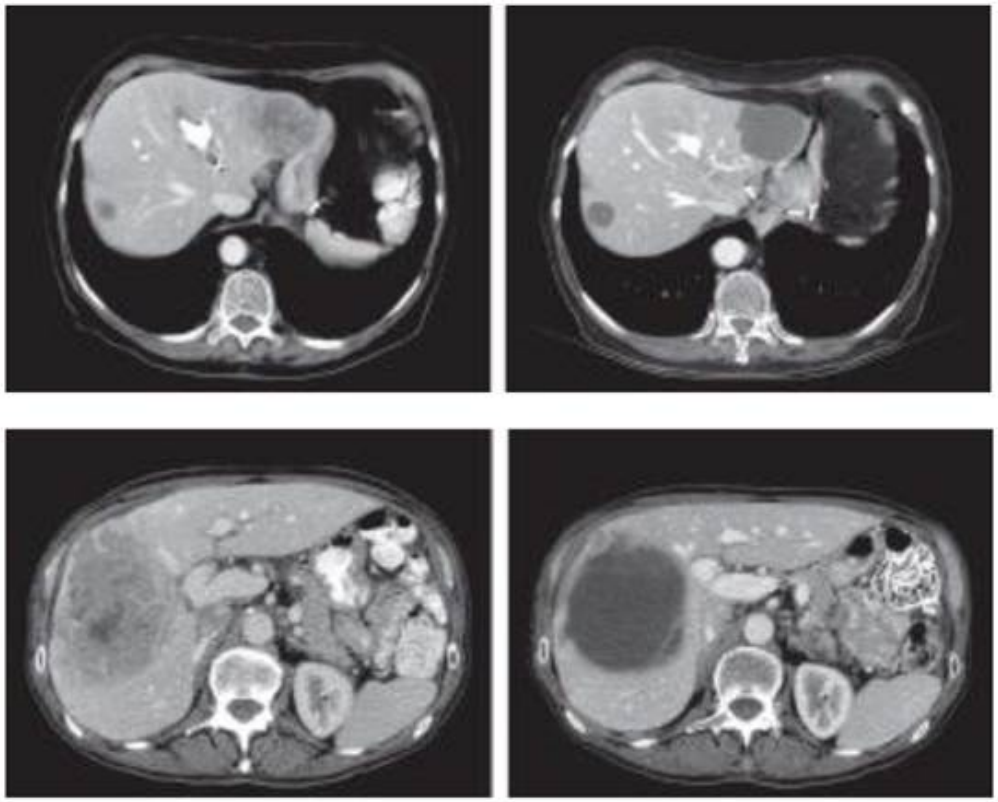


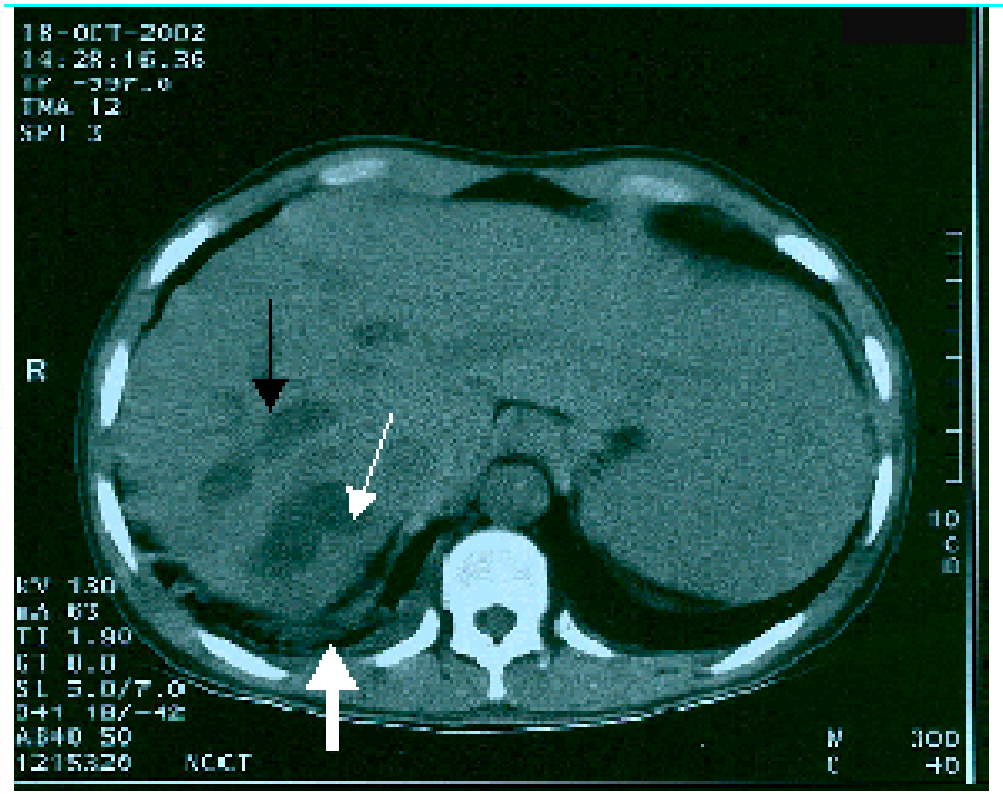


(Webapp)

Radiotherapy

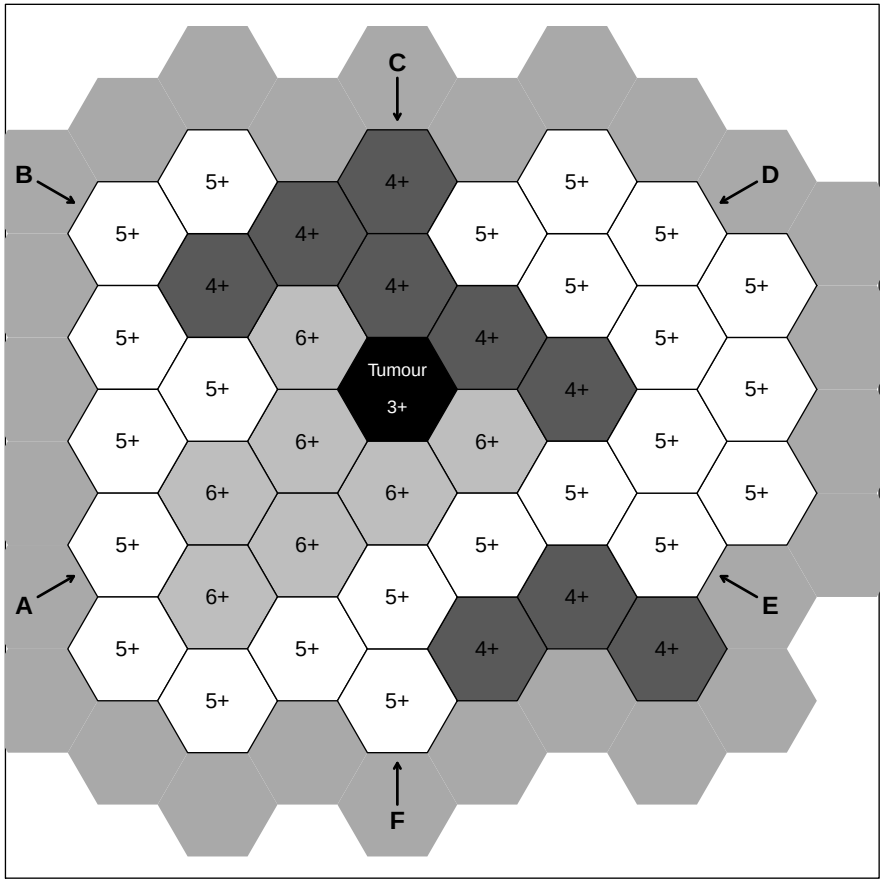
Medical treatment to shrink tumours





Radiotherapy

Simplified illustration



(Map)

What is a "good" outcome?

What dose maximises the probability of damaging only one healthy cell and the tumour absorbing three or more dice?

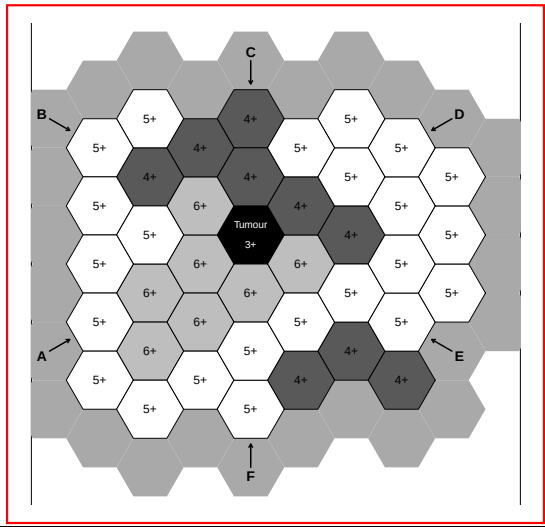
A cell is damaged if it has three or more dice.

Dose strength of 1 from starting point C

Probability tree

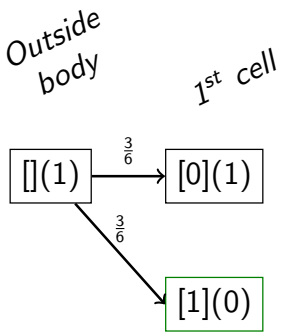
Outside body

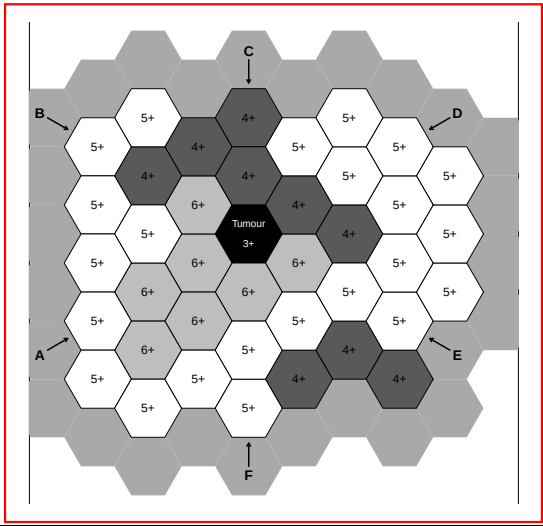
[](1)



Dose strength of 1 from starting point C

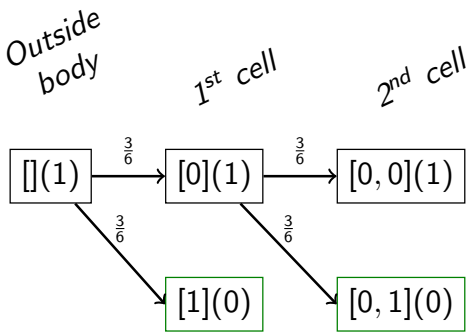
Probability tree

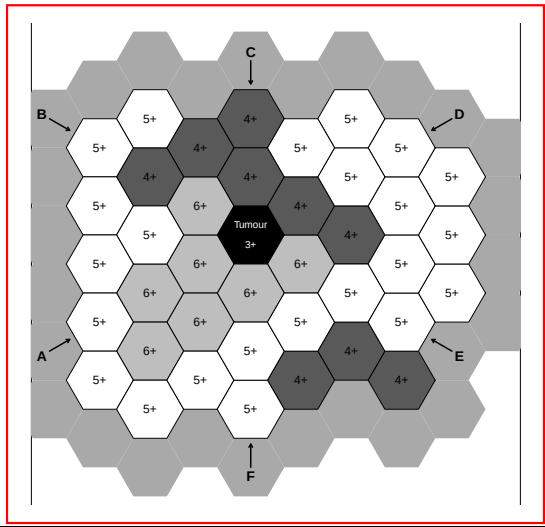




Dose strength of 1 from starting point C

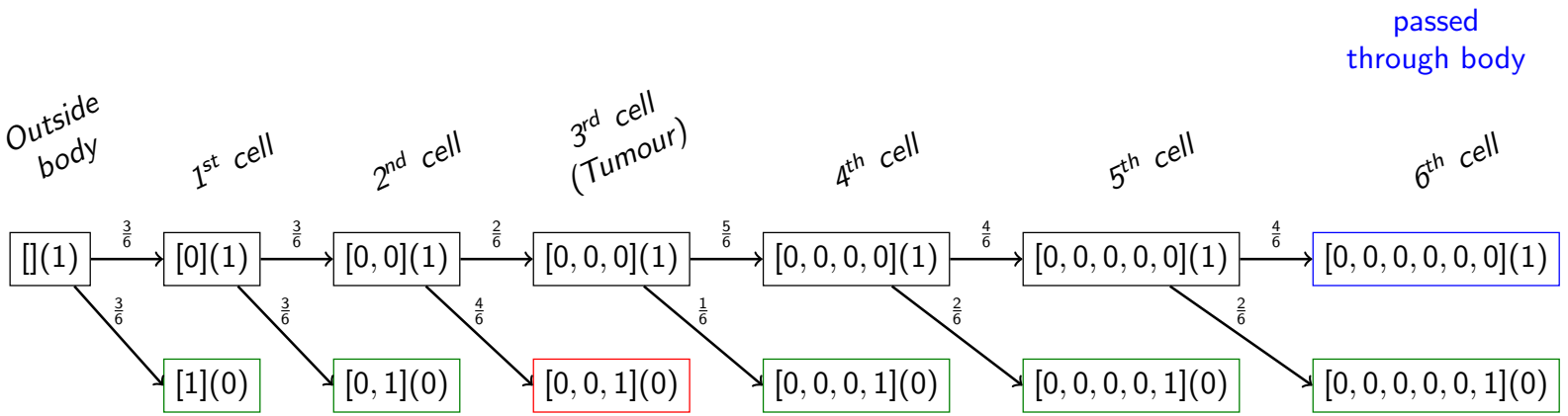
Probability tree

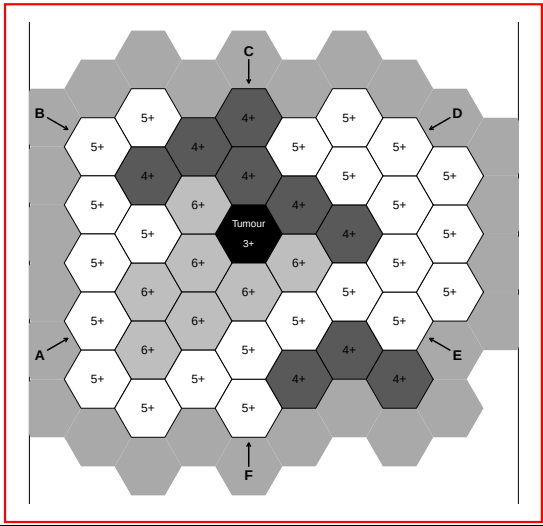




Dose strength of 1 from starting point C

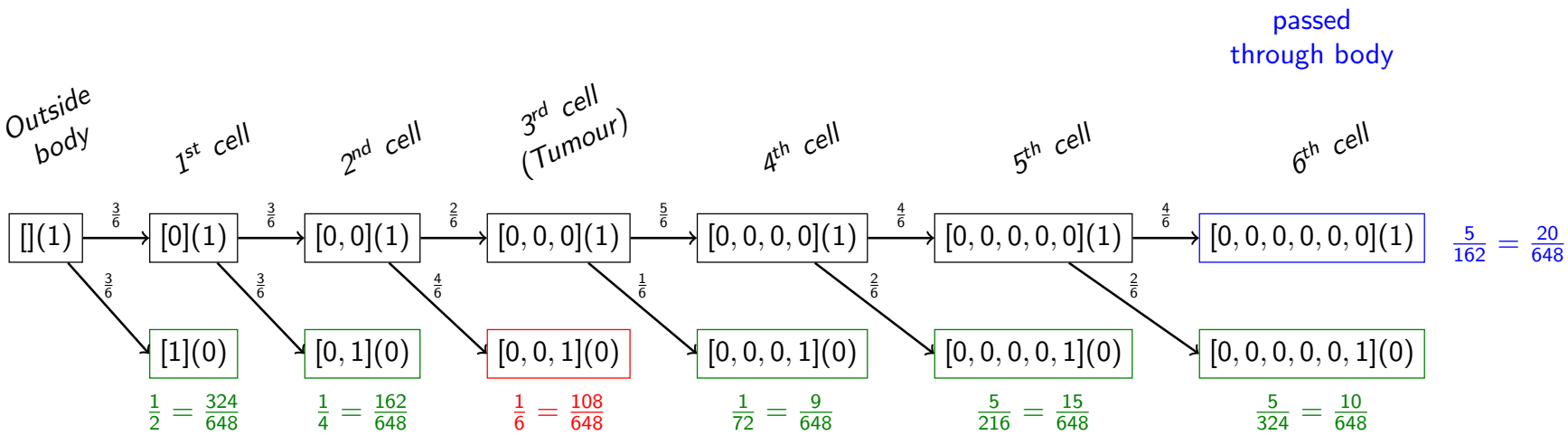
Probability tree

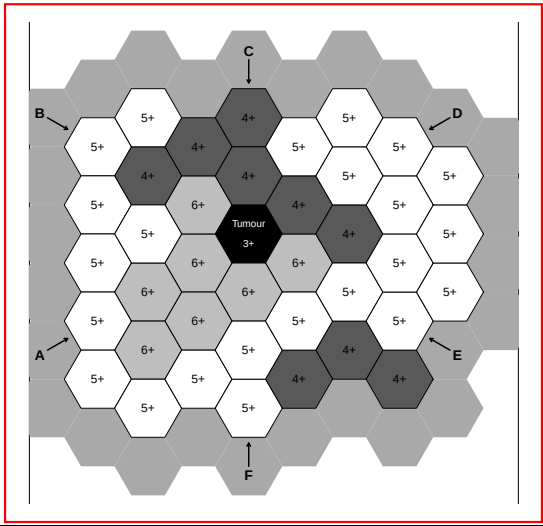




Dose strength of 1 from starting point C

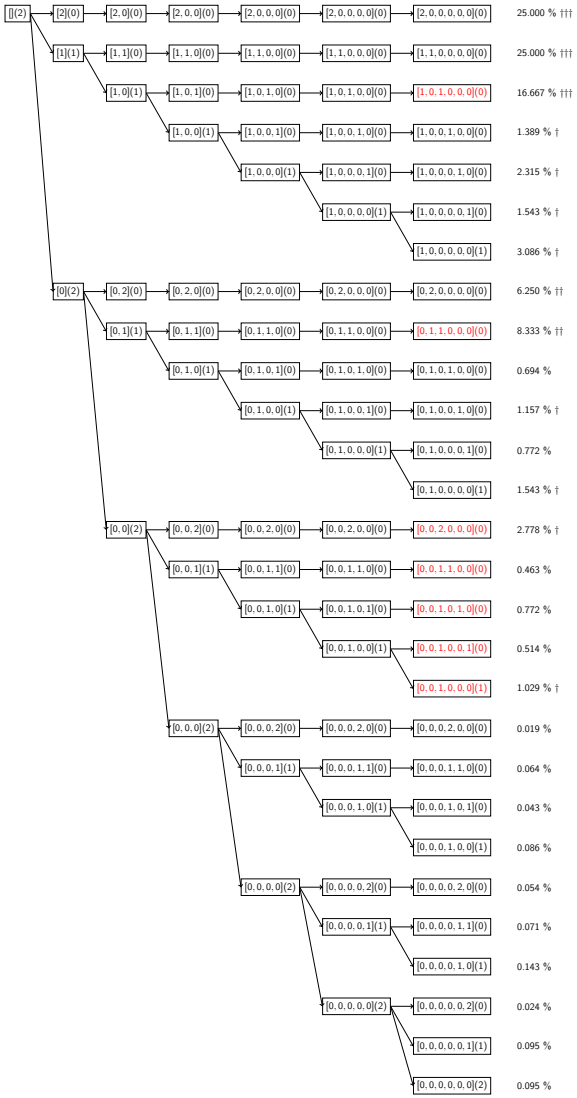
Probability tree





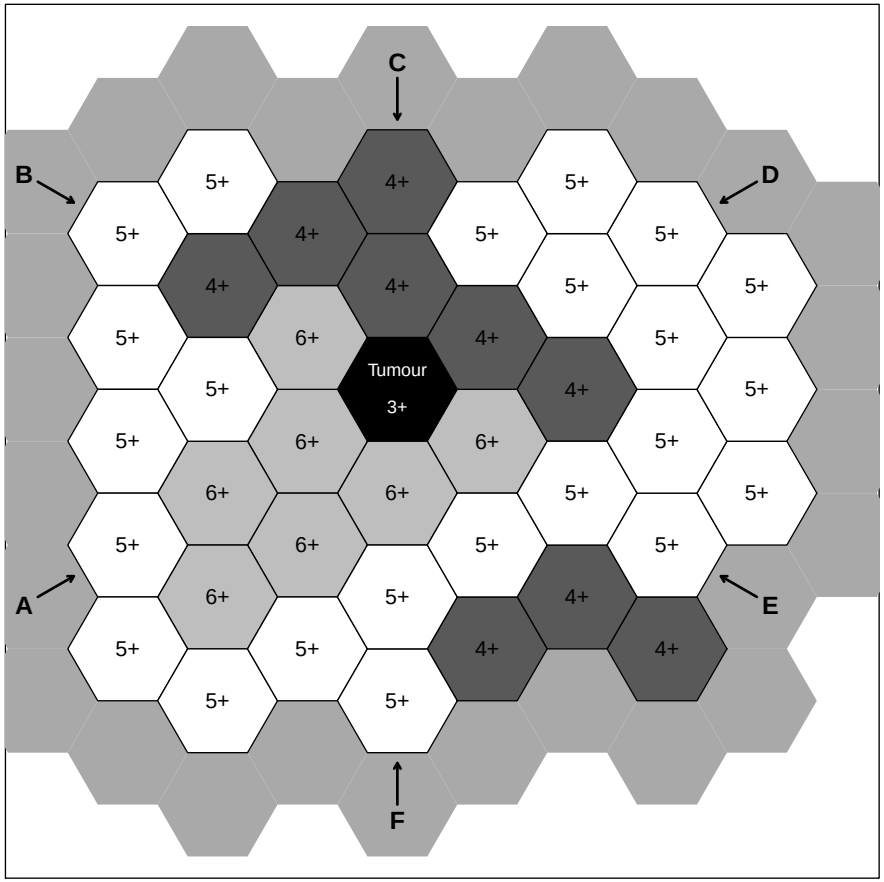
Dose strength of 2 from starting point C

Probability tree



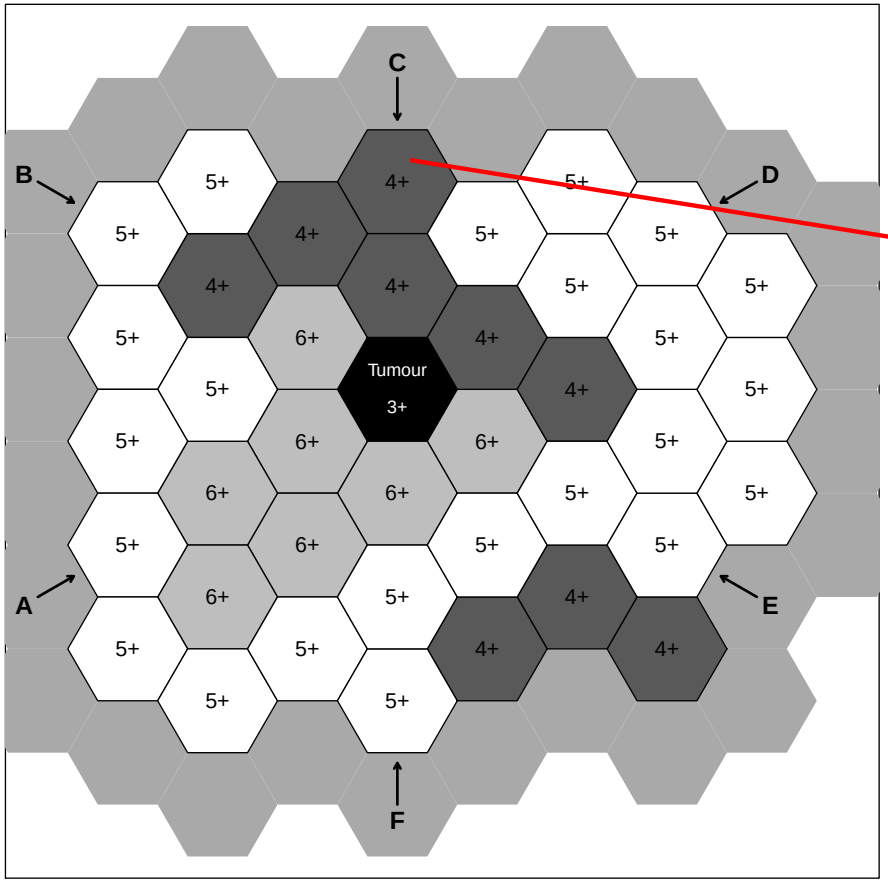
Simulating a dose

Five dice from C



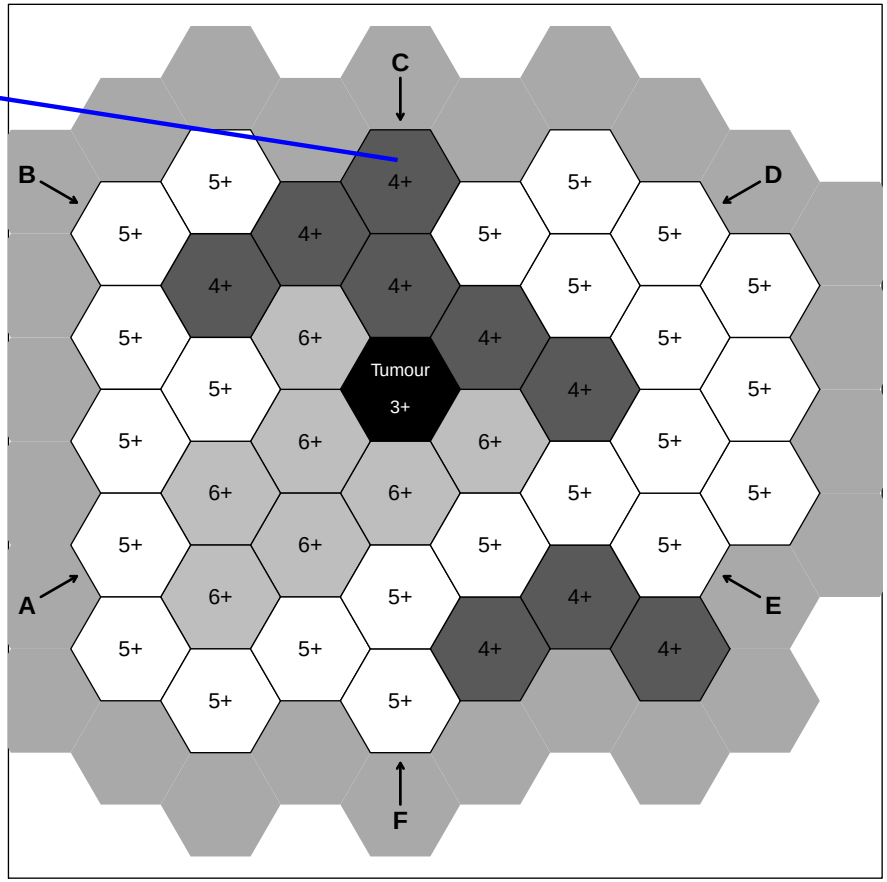
Simulating a dose

Five dice from C



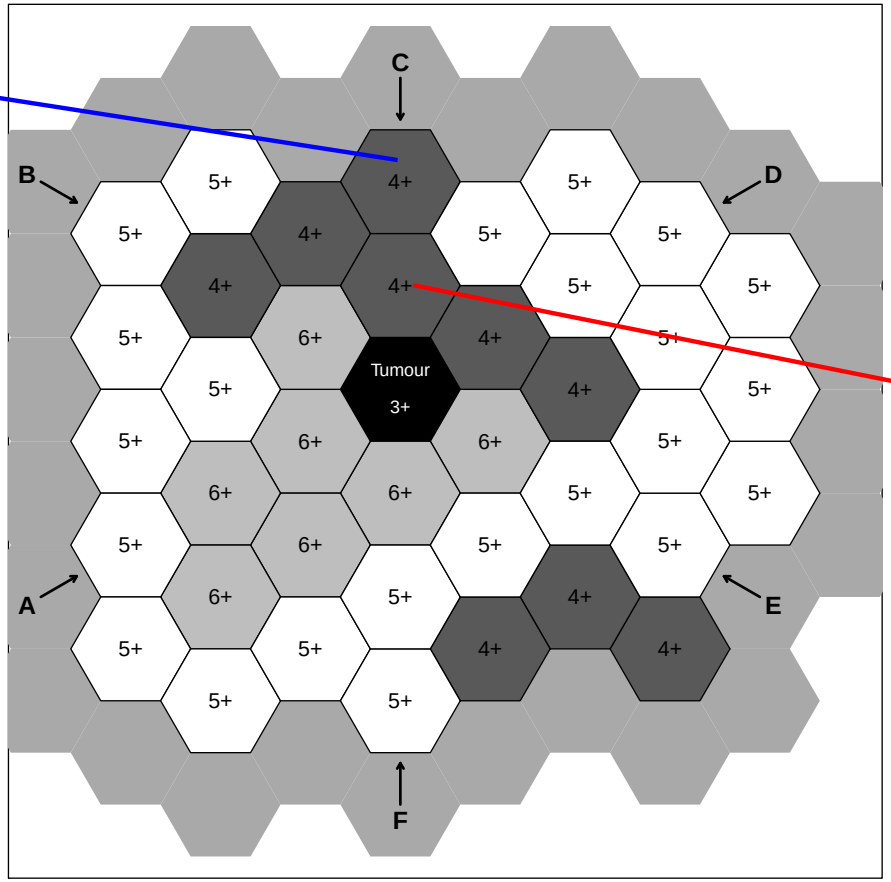
Simulating a dose

Five dice from C



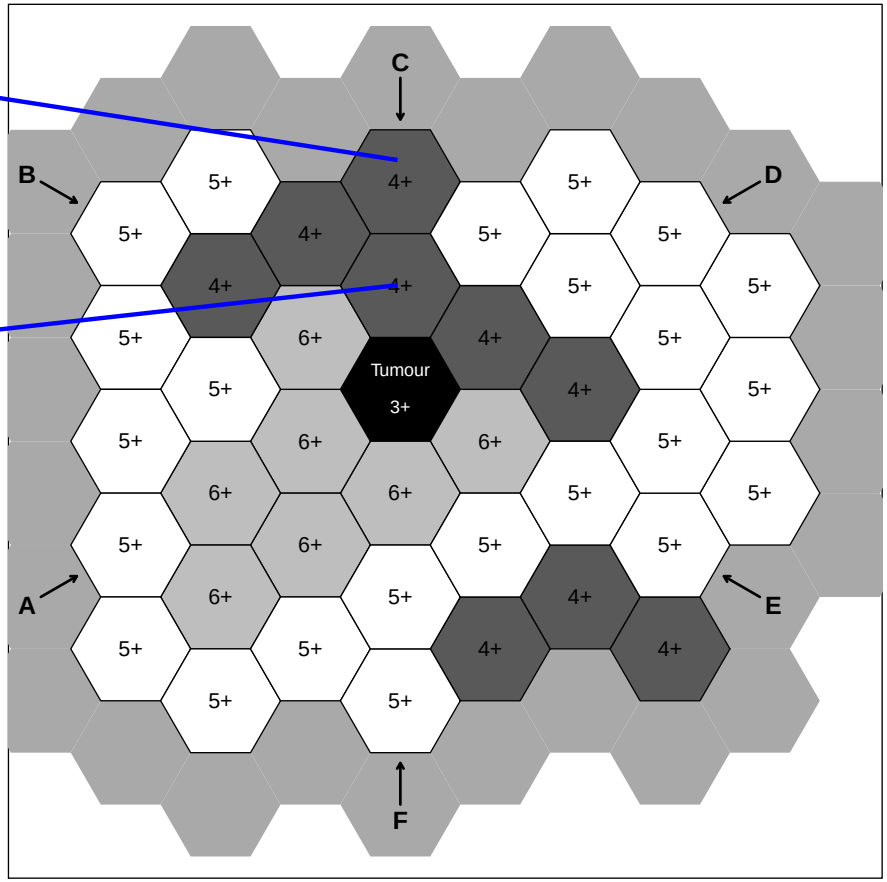
Simulating a dose

Five dice from C



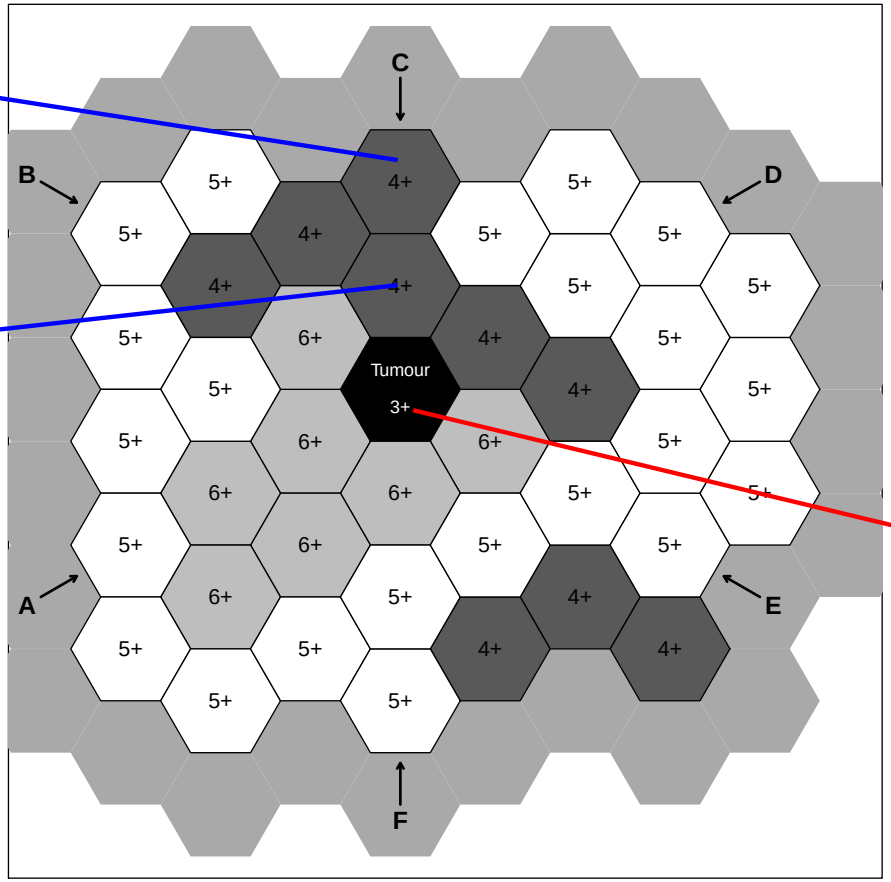
Simulating a dose

Five dice from C



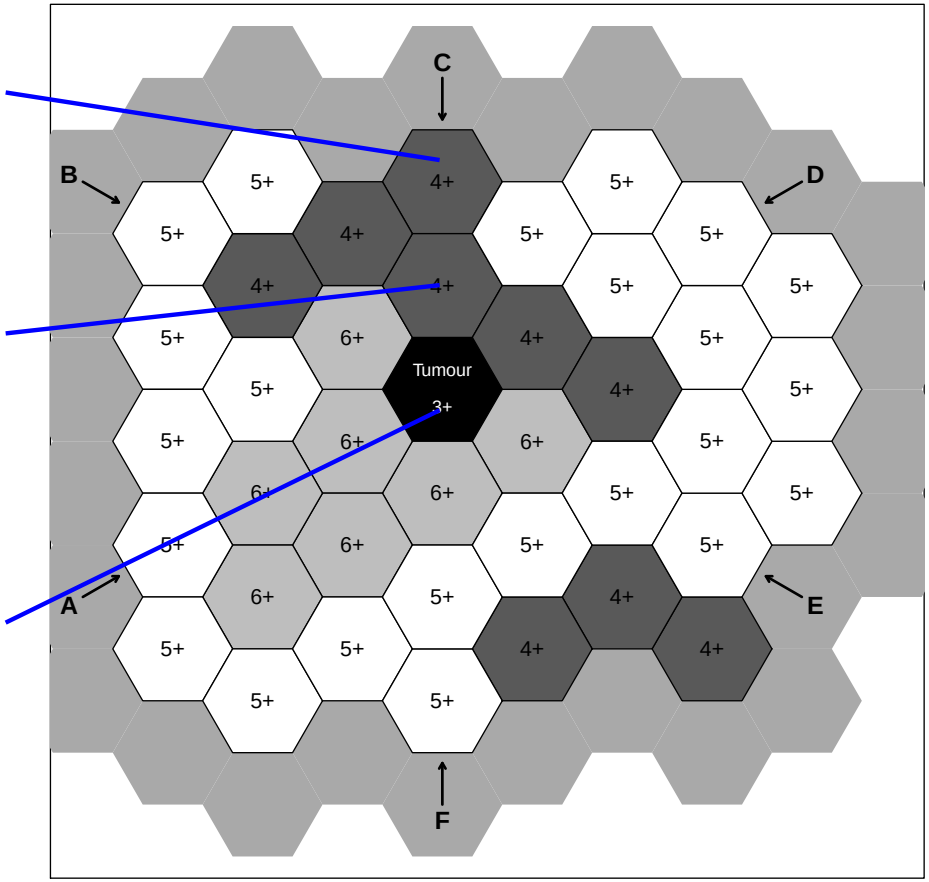
Simulating a dose

Five dice from C



Simulating a dose

Five dice from C



Recording simulations

Group: _____ Scenario: A B C D E F

Number of dice on tumour

Number of cells damaged	0	1	2	3	4	5	6	7	8	9	>9
0											
1											
2											
>2											

Note: a cell is damaged if it has three or more dice on it.

Note: a "good" outcome damages at most one healthy cell and has three or more dice on tumour.

"Good" = ☐

Total = ☐

(Sheet)

Recording simulations

Group: Alpha

Scenario:

A	B	C	D	E	F
-	-	5	-	-	-

Number of dice on tumour

Number of cells damaged	0	1	2	3	4	5	6	7	8	9	>9
0											
1		1									
2											
>2											

Note: a cell is damaged if it has three or more dice on it.

Note: a "good" outcome damages at most one healthy cell and has three or more dice on tumour.

"Good" =

Total =

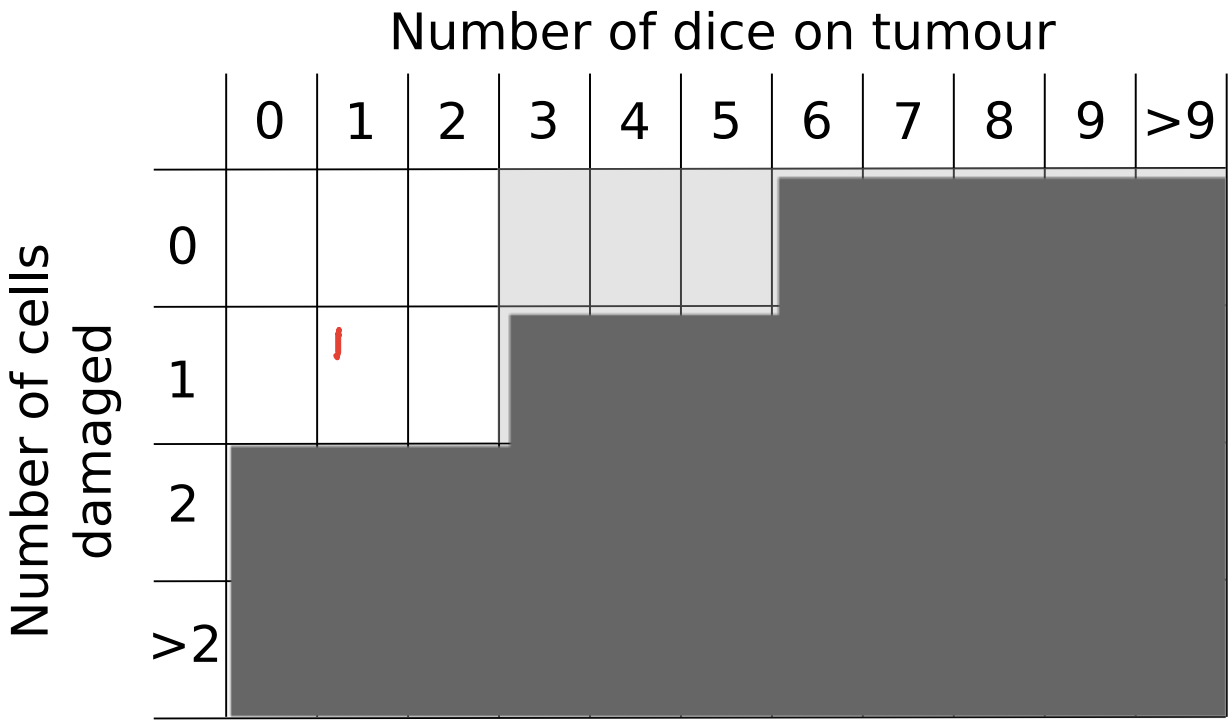
(Sheet)

Recording simulations

Group: Alpha

Scenario:

A	B	C	D	E	F
–	–	5	–	–	–



Note: a cell is damaged if it has three or more dice on it.

Note: a "good" outcome damages at most one healthy cell and has three or more dice on tumour.

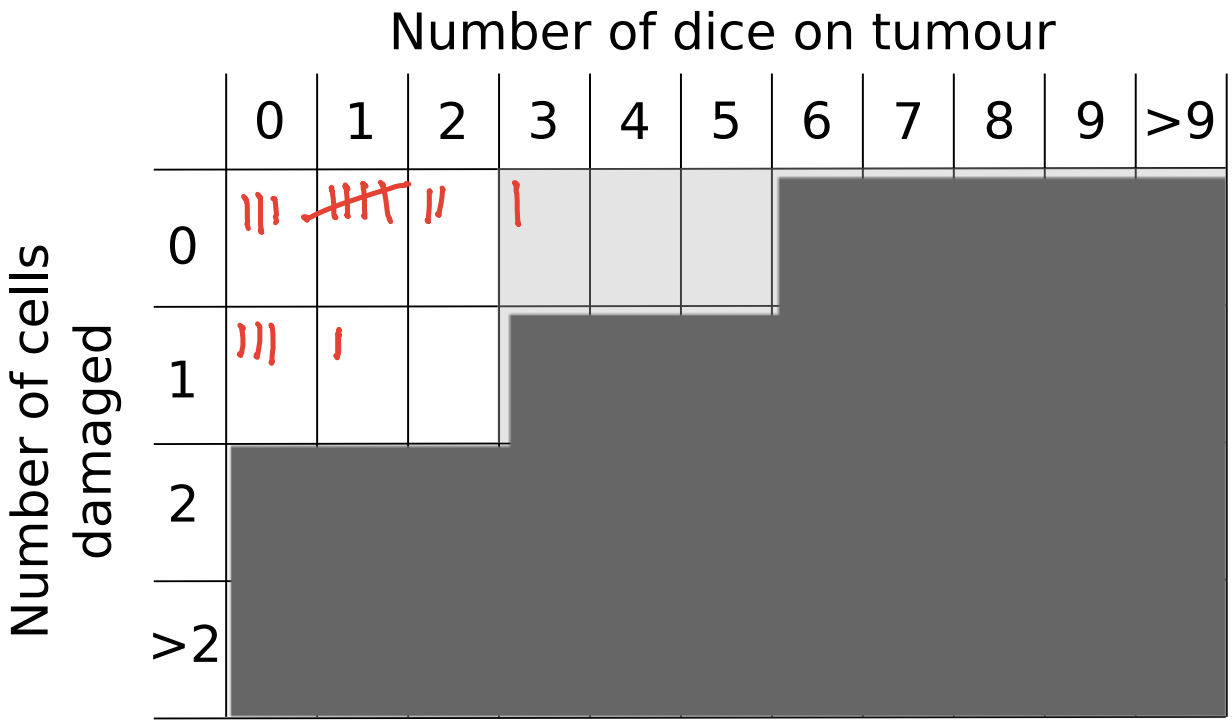
"Good" =

Total =

(Sheet)

Recording simulations

Group: Alpha Scenario: A B C D E F — — 5 — — —



Note: a cell is damaged if it has three or more dice on it.

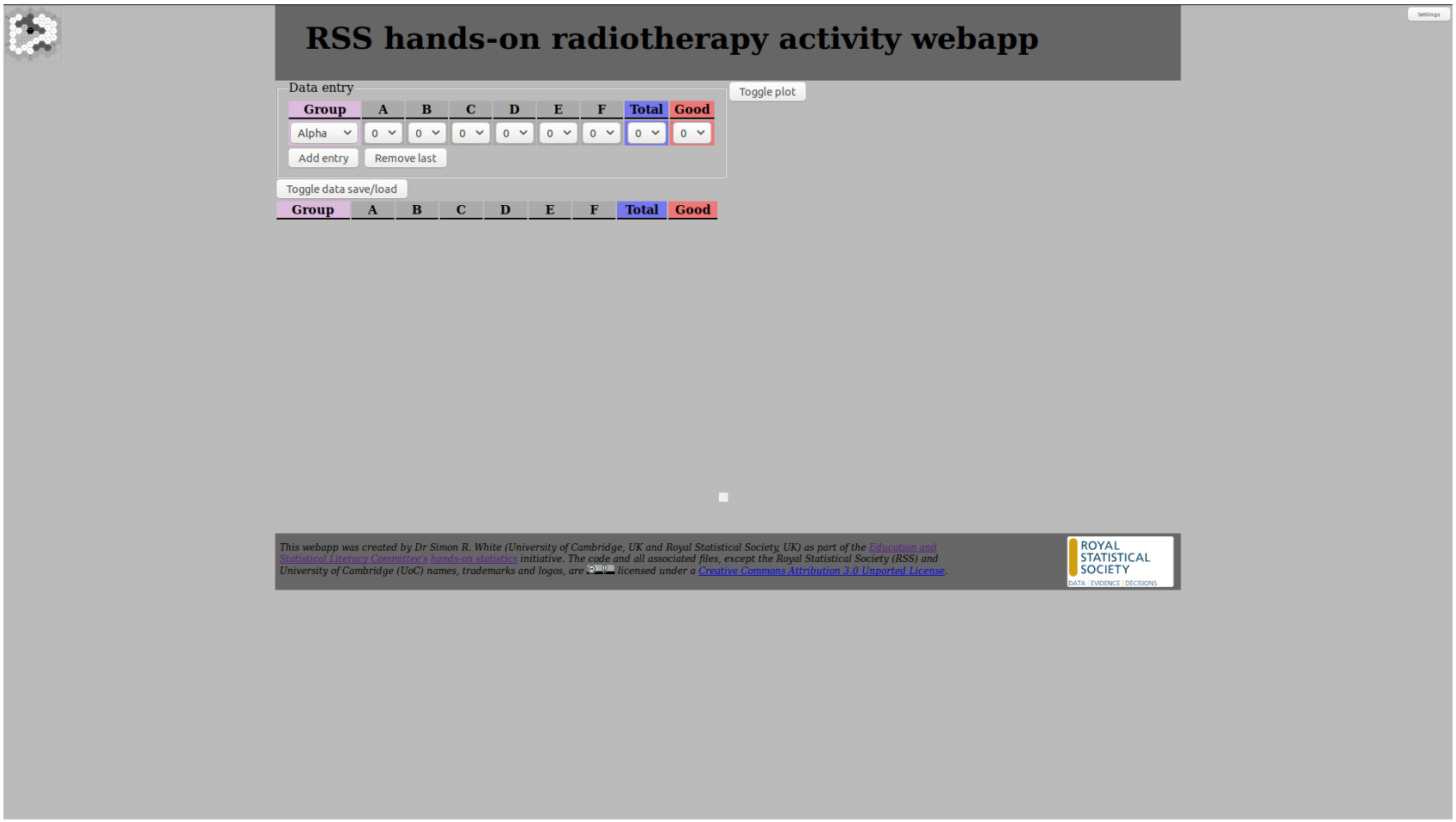
Note: a "good" outcome damages at most one healthy cell and has three or more dice on tumour.

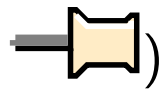
"Good" = 1

Total = 25

(Sheet)

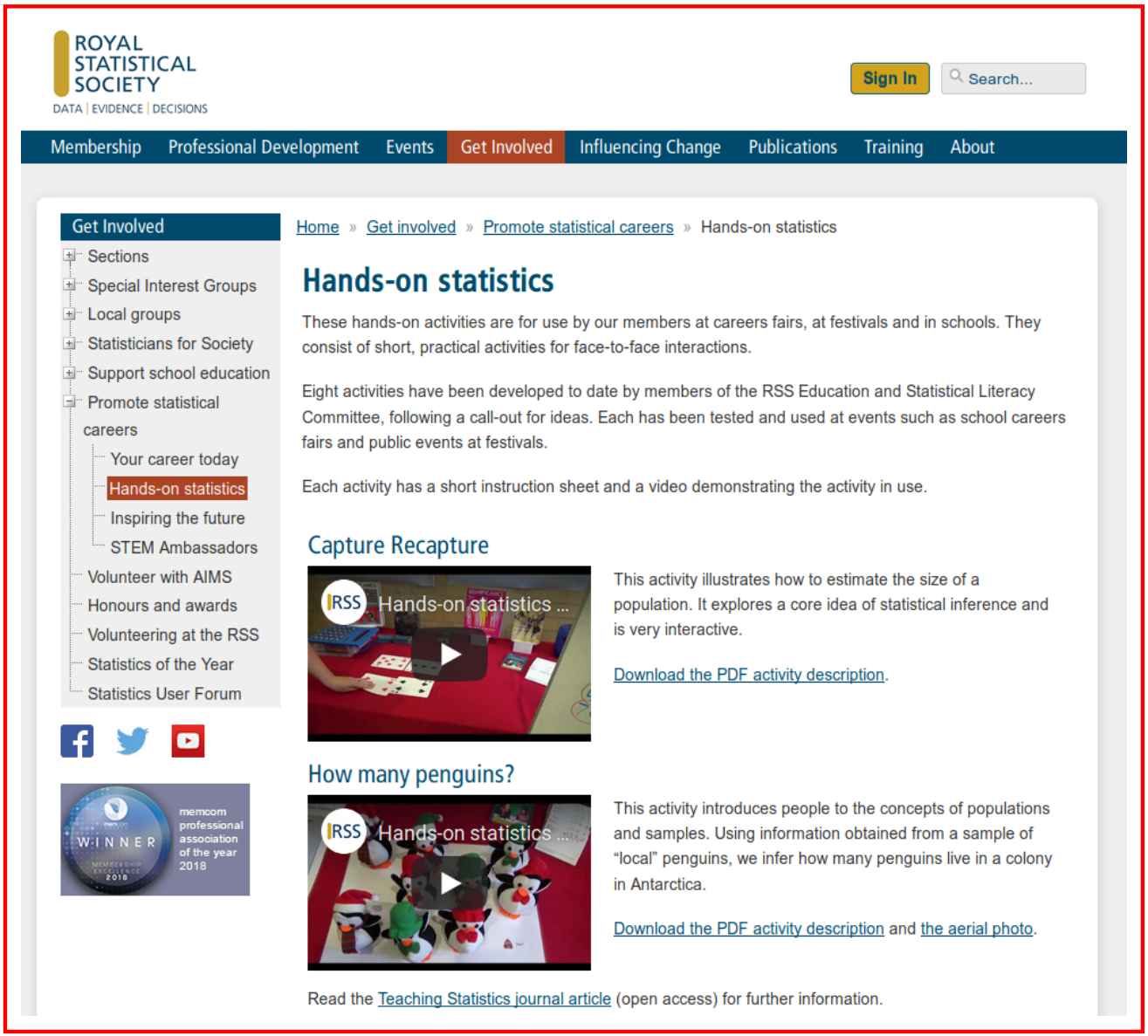
Plotting simulations



(Webapp )

Resources available from the Royal Statistical Society

Hands-on website



ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Sign In | Search...

Membership | Professional Development | Events | Get Involved | Influencing Change | Publications | Training | About

Get Involved
- Sections
- Special Interest Groups
- Local groups
- Statisticians for Society
- Support school education
- Promote statistical careers
  - Your career today
  - Hands-on statistics
  - Inspiring the future
  - STEM Ambassadors
- Volunteer with AIMS
- Honours and awards
- Volunteering at the RSS
- Statistics of the Year
- Statistics User Forum

memcom professional association of the year 2018 WINNER

Home » Get involved » Promote statistical careers » Hands-on statistics

Hands-on statistics

These hands-on activities are for use by our members at careers fairs, at festivals and in schools. They consist of short, practical activities for face-to-face interactions.

Eight activities have been developed to date by members of the RSS Education and Statistical Literacy Committee, following a call-out for ideas. Each has been tested and used at events such as school careers fairs and public events at festivals.

Each activity has a short instruction sheet and a video demonstrating the activity in use.

Capture Recapture

RSS Hands-on statistics ...

This activity illustrates how to estimate the size of a population. It explores a core idea of statistical inference and is very interactive.

Download the PDF activity description.

How many penguins?

RSS Hands-on statistics ...

This activity introduces people to the concepts of populations and samples. Using information obtained from a sample of "local" penguins, we infer how many penguins live in a colony in Antarctica.

Download the PDF activity description and the aerial photo.

Read the Teaching Statistics journal article (open access) for further information.

Resources available from the Royal Statistical Society

Hands-on worksheets

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Capture Recapture

Activity Summary:

This activity illustrates how to estimate the size of a population.

It explores a core idea of statistical inference and is very interactive.

Activity Learning Outcomes:	Suggested Resources:
• Understand the words "estimate", "sample" and "population". • Understand how the population size is estimated. • Understand capture-recapture studies.	• 50-60 uniquely identifiable objects (ideally numbered or labelled). • Opaque container for objects. • (optional) mat to place objects on. • (optional) second set of objects, see below. • Alternatively, a deck of playing cards (52 unique objects, no container needed).

How to run the activity:

Setup:

- Place objects in the container. If using a deck of cards, remove 1-5 cards beforehand.

Introduction:

- Explain that we want to know the total number of objects. We could simply count them all, but this is not viable for any large or difficult to reach **population,** such as fish in a river or birds. Define the technical term **population** as the whole group of interest. The objects are the **population** of interest.
- Get the participant to guess the size of the population. They have no real information to inform their guess. If using playing cards, be clear you have some playing cards (rather than a whole deck of playing cards) to avoid guesses of 52.
- Explain the participant is going to take a **(random) sample**, which is a random subset of the objects.

Activity (one set of objects in one container):

- Get the participant to take a **random sample** of objects from the container and record their labels (see example recording sheet). Take out a sample of 15-20 (out of the 50-60) objects.
- After this first sample, ask the participant again to estimate the size of the population. Although they may have a better idea, having drawn objects from the container, they really only know that the population has at least A objects (where A is the size of the first sample).
- Return the objects to the container and explain we will take a second sample. We call this a **capture-recapture study**. During the second sample, the **recapture**, record which objects are sampled. Some objects will not have been seen before, some will be seen for a second time. The second sample size (B) should again be between 15-20 objects (but does not need to equal the size of the first capture).
- For the final time, ask the participant what is the population size? They should deduce that the population must be at least $a+b+c$ in size, where c is the number of objects in common, a and b are the number of objects only seen in the first ($A=a+c$) and second ($B=b+c$) sample respectively.
- A statistical **estimate** is a numerical characteristic of interest obtained from the sample, in this case the size of the population. An estimate of the population size is $(AB)/c$.
- Finally, reveal the true population size and compare to the participant's estimate. Discuss that we do not expect the estimate to exactly equal the true population size.

Activity (two sets of distinguishable objects in two containers, plus mat):

- The only difference with the above description is that the capture and recapture can be conducted at the same time (note, the two containers must be identical – have the same objects labelled the same way, but visually different).
- A mat (see example photograph) including spaces for objects as they are drawn becomes a visual representation of the capture and recapture. The counts a, b, and c can be seen from which objects are on the mat.

Exploring the activity:

- Repeat the activity several times (keeping the sample sizes the same) to explore the variability of the estimate.
- Do we tend to over or under estimate the population size?
- What happens if the samples are too small? What if c=0?
- We assume capturing has no effect on being recaptured, is this reasonable? When might it be invalid? For example, animals if captured may be more difficult to recapture.

What's going on?

- Let the population size be N, so that there are d unseen objects, i.e. $N=a+b+c+d$.
- Assuming two random samples and that objects are sampled with the same probability (equal catchability assumption), then the proportion of the population in the first sample, A/N, that will be marked will equal the proportion of the second sample that is recaptured (marked), c/B.

$$A/N = (a+c)/N = c/(b+c) = c/B \quad \rightarrow \quad N = (a+c)(b+c)/c = AB/c = ab/c + (a + b + c)$$

Video demonstration:

A video demonstrating this activity is available on the RSS website at www.rss.org.uk/hands-on

Risk assessment:

Depending on the size of the objects and mat (if used), be mindful of slip hazards and choke hazards.

Additional information and taking it further:

Lincoln–Petersen method: https://en.wikipedia.org/wiki/Mark_and_recapture

Credits:

Dr Simon R. White (Medical Research Council Biostatistics Unit, University of Cambridge)

Version 4.0 Published 2018/September/05

Resources available from the Royal Statistical Society

Teaching Statistics articles

Original Article

May the odds be ever in your favour

Laura J. Bonnett
Department of Biostatistics, University of Liverpool, Liverpool, UK

Simon R. White
MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
e-mail: simon.white@mrc-bsu.cam.ac.uk

Summary Probability and chance are essential concepts, not just in statistics but in real life. We present an adaptable activity which investigates what we mean by bias, how we can identify bias, and how we can use it to our advantage!

Keywords: Teaching statistics; practical activity; two-dice experiment; probabilities; simulations.

INTRODUCTION

In the authors' experience, ask a secondary school student in the United Kingdom (age 11 to 18) what they understand by probabilities and chance and it is likely they will respond that probability relates to counters in a bag or socks in a drawer. We are keen to present the concepts in a fresh and inspiring way.

As working statisticians, our engagement with the next generation of possible statisticians usually takes place at science fairs such as The Big Bang and Cambridge Science Festival. These events require quick-hitting low-technology activities which immediately entice students to the stall, engage them and can then introduce and demonstrate a statistical concept in less than 5 min. Therefore, we have to move away from the student-perceived 'dull' counters in bags and socks in a drawer demonstrations whilst still teaching the concept of probability and chance.

Games are often thought of as a distraction – something to be done for fun. However, there is growing evidence to suggest that games can do much more, especially when it comes to classroom learning. A review of research into the effectiveness of games for educational purposes determined that mathematics was the subject area with the greatest percentage of results favouring games over conventional classroom instruction (Randel et al. 1992).

An example of an existing activity developed for the purpose of statistical outreach to illustrate probability involved the Hunger Games, an annual event in the fictional country of Panem (Caudle and Daniels 2015). The authors used statistical analysis and computer simulations to explore the possibility that the Gamemakers, those in charge of planning the Hunger Games, fixed the lottery. Additionally, a recent informal study introduced undergraduate students to two versions of a story involving probability (Passaretti and Vistocco 2017). The results of this study revealed the students had trouble detecting equal probabilities in a sampling scheme without replacement in which no information on earlier draws was available.

We therefore propose a novel activity to demonstrate chance via biased die and an outdoor-sized Ludo board game. The learning aim is to understand the concept of bias and determine when the method of estimation of probabilities is biased, and why that can be useful in real life applications. Our activity can be tailored to the available time, context and level of the target audience. In the following sections, we present a suggested template for delivering the activity as a single game play, and as a multiple game play version, together with additional extension activities.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2018 The Authors Teaching Statistics published by John Wiley & Sons Ltd on behalf of Teaching Statistics Trust, ••, ••, pp ••–••

Original Article

Biased sampling activity: an investigation to promote discussion

Simon R. White
MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
e-mail: simon.white@mrc-bsu.cam.ac.uk

Laura J. Bonnett
Department of Biostatistics, University of Liverpool, Liverpool, UK

Summary The statistical concept of sampling is often given little direct attention, typically reduced to the mantra "take a random sample". This low resource and adaptable activity demonstrates sampling and explores issues that arise due to biased sampling.

Keywords: Teaching; Teaching statistics; Sampling; Biased sampling; Practical activity.

INTRODUCTION

It is difficult to communicate statistical concepts in a meaningful way using only abstract definitions. Using interactive activities that highlight the key concept, consisting of a physical process that can be followed and manipulated, can greatly support learning. The focus of this article is to describe such an activity to demonstrate sampling.

Sampling is an interesting concept to communicate. With a reasoned line of questions, we can convey an intuitive understanding of the concept in genuine population contexts, namely, that 'taking a sample' is required to learn something about a large group without having to investigate everyone. By large group, we mean the population of interest, which naturally leads to a discussion about the term *population*; differentiating its common meaning – number of living things or items – from its traditional statistical usage covering all contexts and meaning a general situation described by a probability model.

There are many questions we can use to stimulate this discussion and give it a real-life context. For example, 'what pets do people own?', 'how many people have dementia?', 'how will people vote in the next election?' or 'in how many films does the lead actress have more lines than the lead actor?' In each case, we first have to consider what the population of interest is; for questions about people, we may consider a local region, a country or even all the people of the world, but that does not make sense for how people will vote in an election – with different political parties in different countries. When thinking about films, we have to think about what we mean by the word 'film': only films with a lead actress and lead actor – not all films have both, only fictional films or documentaries, major studio releases or fan-made short films, all films ever or only in recent years, what about different languages. This may seem pedantic, but it can be at the core of a discussion on sampling.

If we could find every member of the population, then we would find *the* answer to our question, there would be no uncertainty. However, i.e. typically not feasible (either it would take too long or cost too much money) so instead we only consider some members of the population – we take a sample. From a sample, we will obtain an estimate of a quantity of interest about the population.

The aim of this activity is to explore ideas around taking a sample, with a focus on the issue of biased sampling. The learning aim is to understand the need for a well-designed sampling scheme to ensure an unbiased sample from the population that will lead to an unbiased estimate.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors Teaching Statistics published by John Wiley & Sons Ltd on behalf of Teaching Statistics Trust. **41**, 1, pp 8–13

Thanks

@ljbcmshe

Thank you

@SimonWhite83

UNIVERSITY OF LIVERPOOL

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE