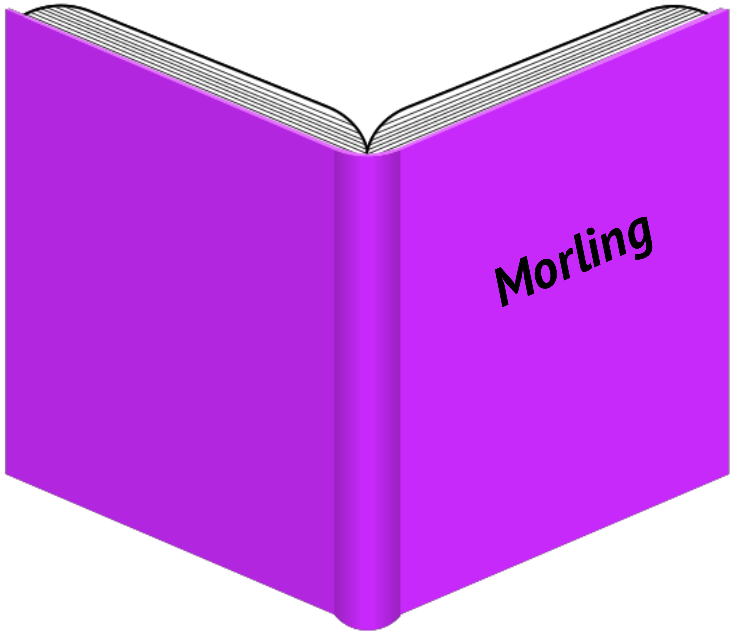
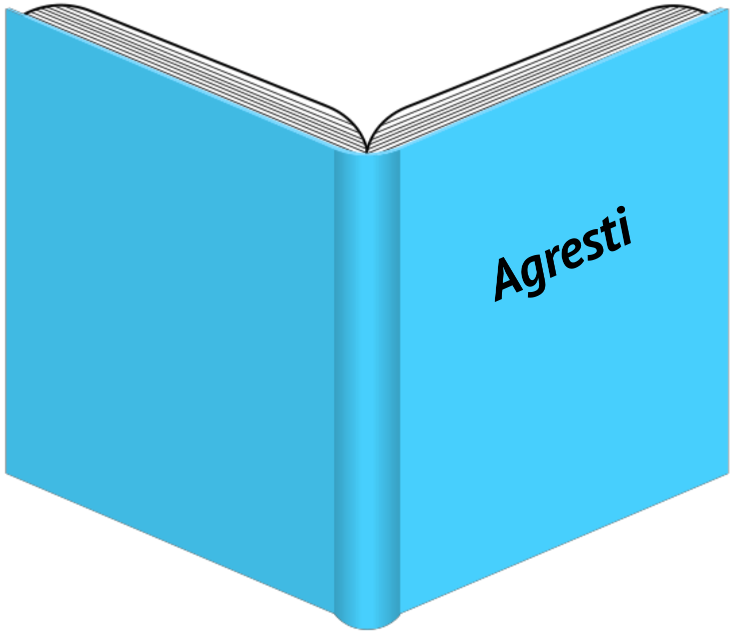


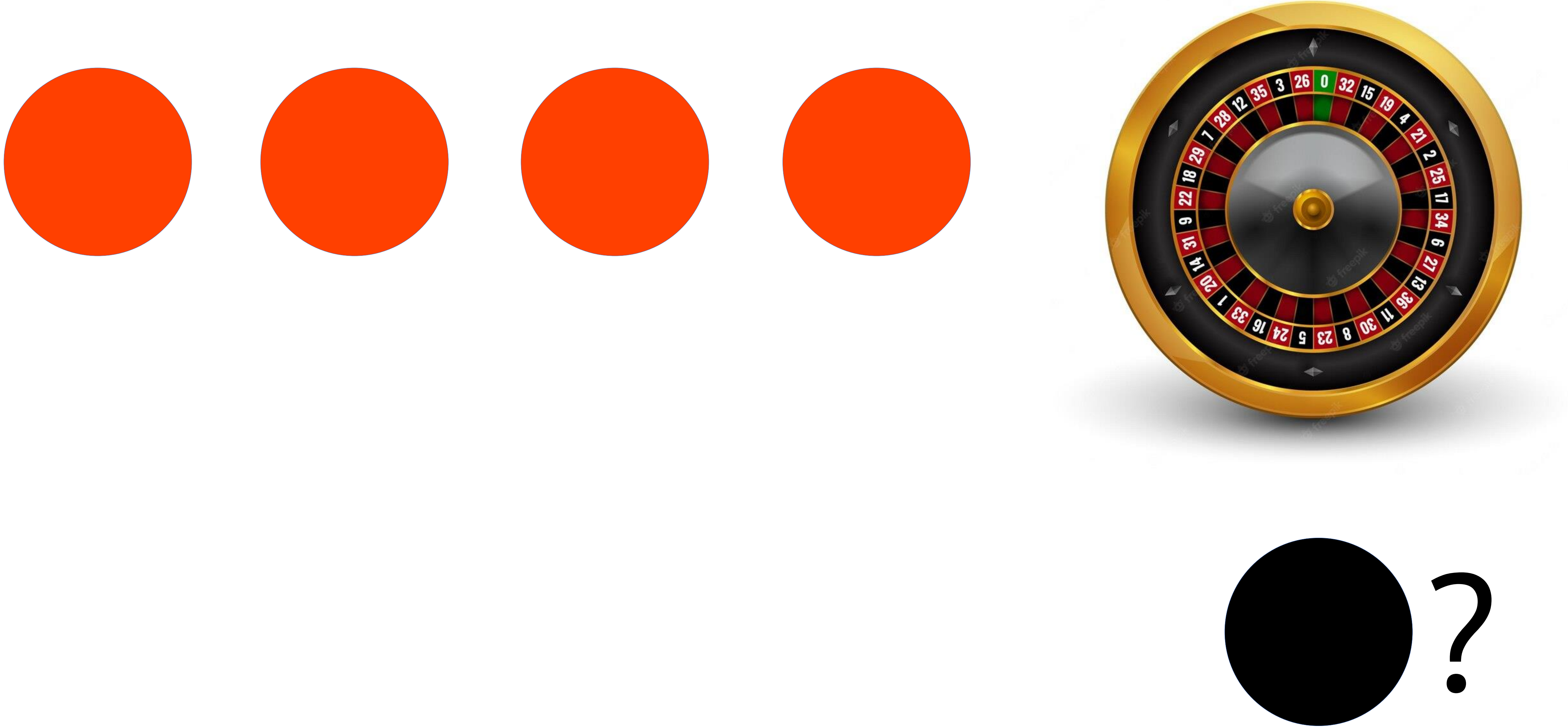


Research Methods and Statistics

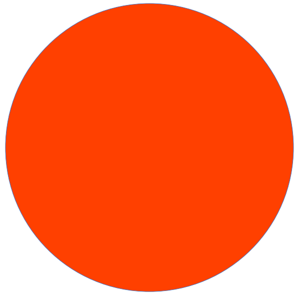
Lecture 6: Conditional Probability

Johnny van Doorn

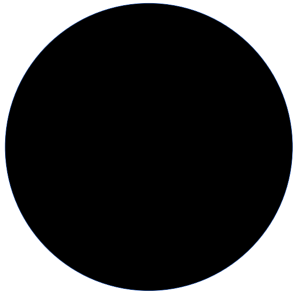
Pictures source: pixabay



Pictures source: freepik.com

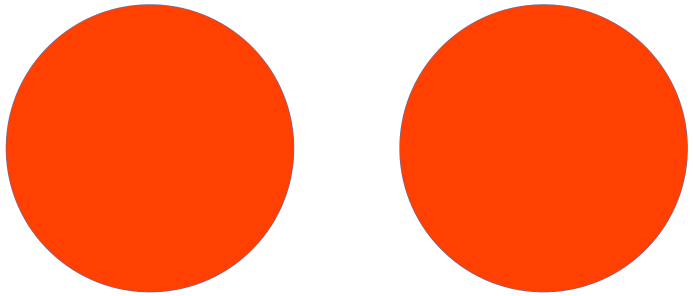


P(R, B)?

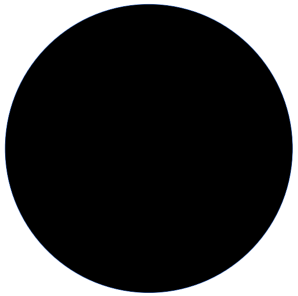

?

P(B | R)?

Pictures source: freepik.com

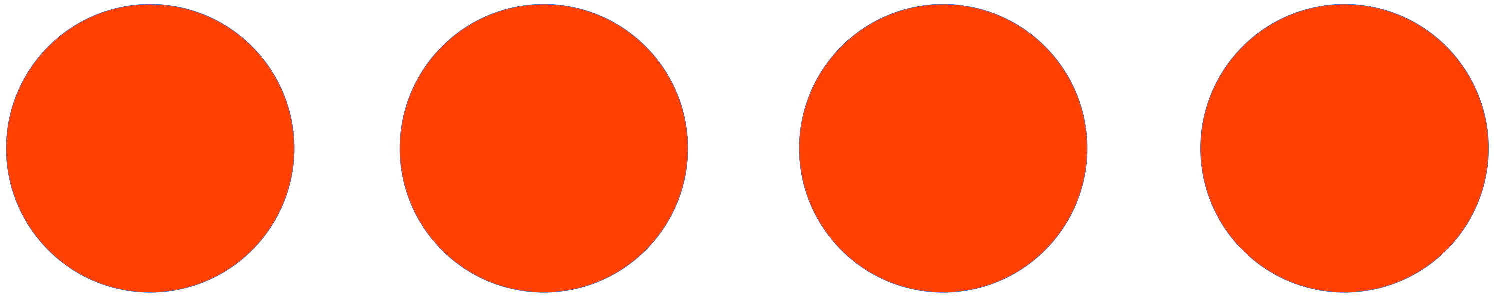


P(R, R, B)?

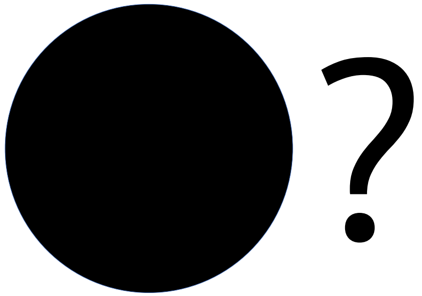
?

P(B | R, R)?

Pictures source: freepik.com

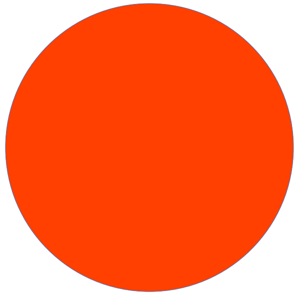


$P(R, R, R, R, B)$?

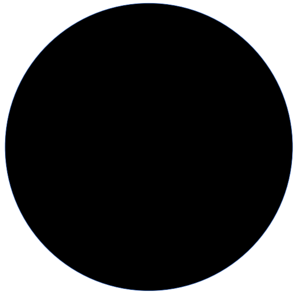


$P(B \mid R, R, R, R)$?

Pictures source: freepik.com



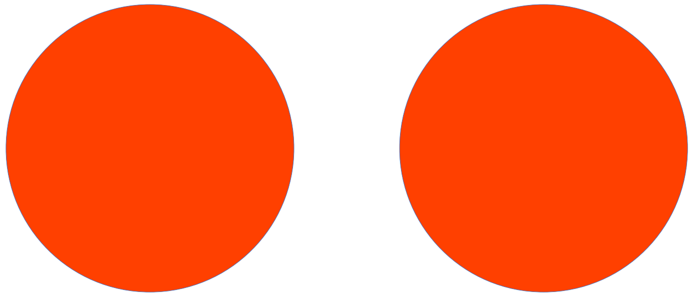
$$P(R, B) = 0.5 * 0.5$$



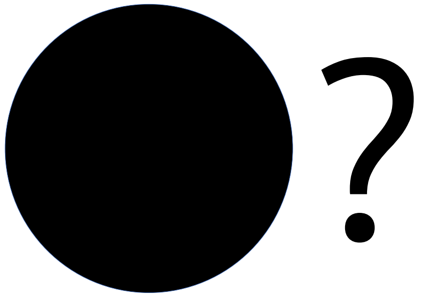
?

$$P(B \mid R) = 0.5$$

Pictures source: freepik.com

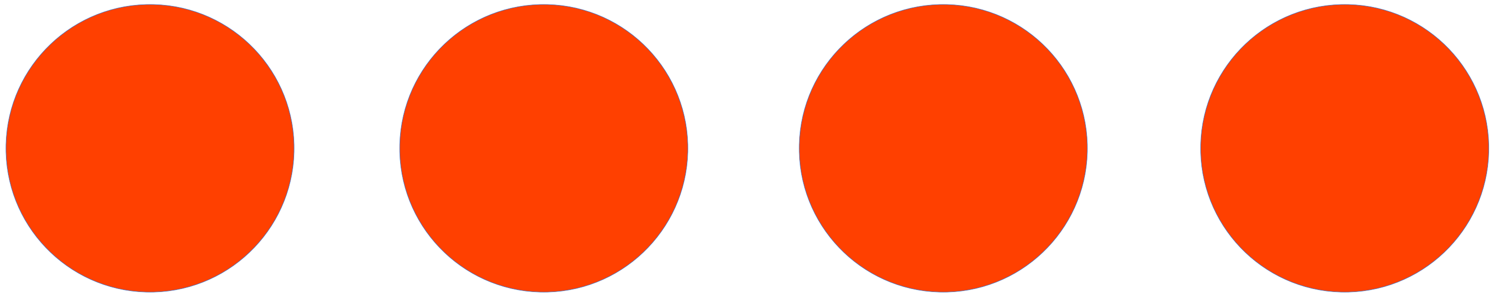


$$P(R, R, B) = 0.5^2 * 0.5$$

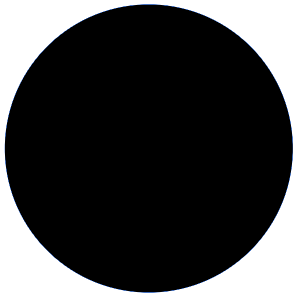


$$P(B \mid R, R) = 0.5$$

Pictures source: freepik.com



$$P(R, R, R, R, B) = 0.5^4 * 0.5$$

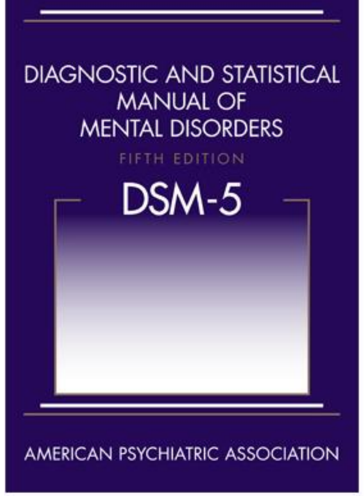


?

$$P(B \mid R, R, R, R) = 0.5$$

Pictures source: freepik.com

Conditional probabilities are often counterintuitive



- Probability of ADHD in children is 0.05 (prevalence/ base rate)
- Probability that the diagnostic test indicates that you have ADHD, when you really have ADHD is 0.90
- Probability that the diagnostic test indicates that you **don't** have ADHD, when indeed you don't have ADHD is 0.75
- **What is the probability that you have ADHD, when the test indicates that you have ADHD?**

P(D|Pos)=0.1593!

ADHD prevalence: 5.3% worldwide, 4.7% in NL (Nederlands Jeugdinstituut)

Source: https://en.wikipedia.org/wiki/Diagnostic_and_Statistical_Manual_of_Mental_Disorders

Base rate neglect

- Often, people ignore that P(D | Pos) and P(Pos | D) are different concepts (Tversky & Kahneman, 1982)

- P(D | Pos) and P(Pos | D) differ strongly if the prevalence (base rate) is low

→ Probability of death if aliens invade earth ≠ probability of alien invasion if you die

→ Base rate neglect

Nice video that explains this: https://www.youtube.com/watch?v=VeQXXzEJQrg

Diagnostic testing is hypothesis testing in miniature: P(Pos | M) is not P(M | Pos).
The same confusion appears with hypothesis testing, P(Data | Hypothesis), see lecture 13 (Gigerenzer & Hoffrage, 1995)

Source: https://en.wikipedia.org/wiki/Daniel_Kahneman

Today

1. **Conditional Probability**
    - Definition of conditional probability
    - Probability rules recall
2. Diagnostic tests
    - Three methods for calculating conditional probability
    - Probability rules extension
3. Independence revisited
    - Conditional probability is useful to determine independence
4. Closing
    - Next time
    - Example exam question

What is conditional probability?

- Chance of an event, when you know that another event has already occurred

"A given B"

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Examples of conditional probability statements

- Having ADHD **if** the rating scale score is above the cut-off
  - → Diagnostic tool quality!
- Probability of your pet wanting to cuddle **when** it wags its tail
- IQ score > 100 **for** blonde people

Simple scenario: three marbles

- The probability of each marble is 1/3
  - What is the probability of a red marble?
- Now we take two marbles in a row *(without replacement)*
  - What is the probability that the *second* marble is red?
  - First draw: P(red)=2/3
  - Second draw: P(red)=1/2 OR P(red)=1

Simple scenario: three marbles

- The probability of each marble is 1/3
  - What is the probability of a red marble?
- Now we take two marbles in a row *(without replacement)*
  - What is the probability that the *second* marble is red?
  - First draw: P(red)=2/3
  - Second draw: P(red)=1/2 OR P(red)=1
    - P(red)=0.5 if first draw is red

Simple scenario: three marbles

- The probability of each marble is 1/3
  - What is the probability of a red marble?

- Now we take two marbles in a row *(without replacement)*
  - What is the probability that the *second* marble is red?
  - First draw: P(red)=2/3
  - Second draw: P(red)=1/2 OR P(red)=1
    - P(red)=1 if first draw is green

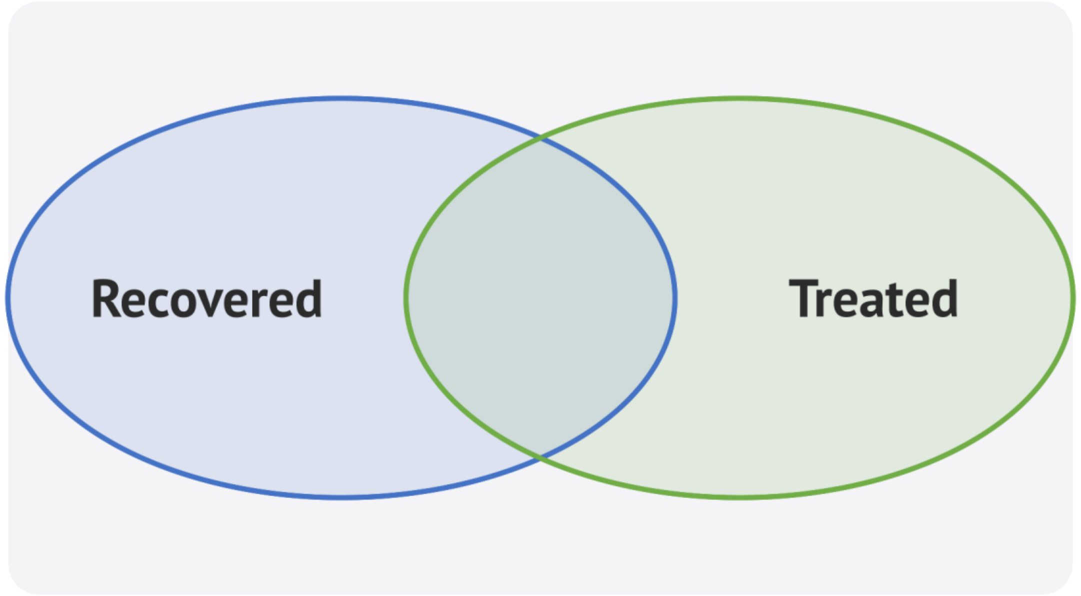
Simple scenario: three marbles

- The probability of each marble is 1/3
  - What is the probability of a red marble?

- Now we take two marbles in a row *(without replacement)*
  - What is the probability that the *second* marble is red?
  - First draw: P(red)=2/3
  - Second draw: P(red)=1/2 OR P(red)=1
    - P(red)=0.5 if first draw is red
    - P(red)=1 if first draw is green

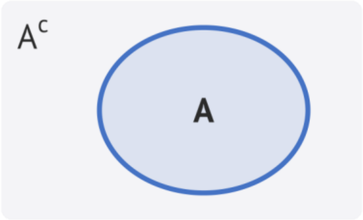
Another scenario: electroshock therapy

- Lecture 4: 80 patients, treated with electroshock therapy or not
- What is the chance that someone recovers **IF** they were treated?

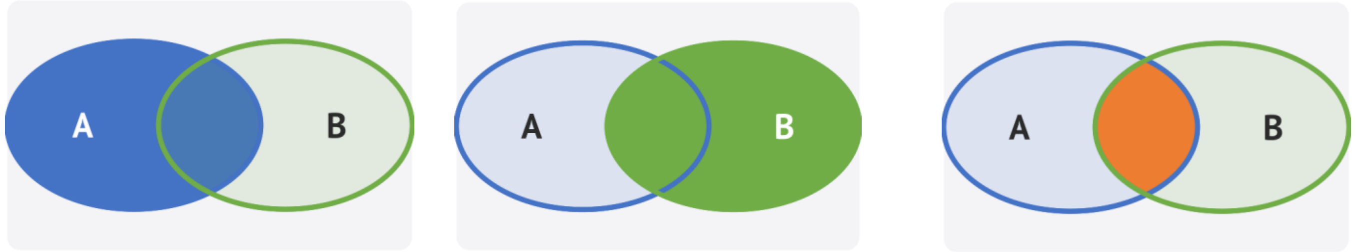


Probability rules

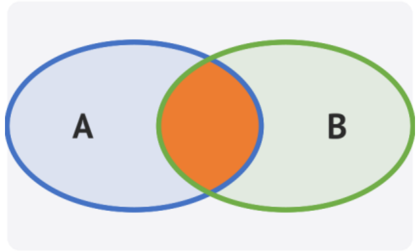
- Complement rule
  - $P(A^c) = 1 - P(A)$



- Addition rule
  - $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

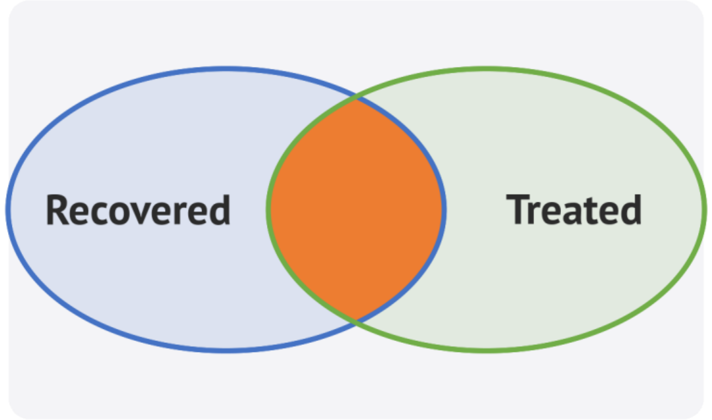


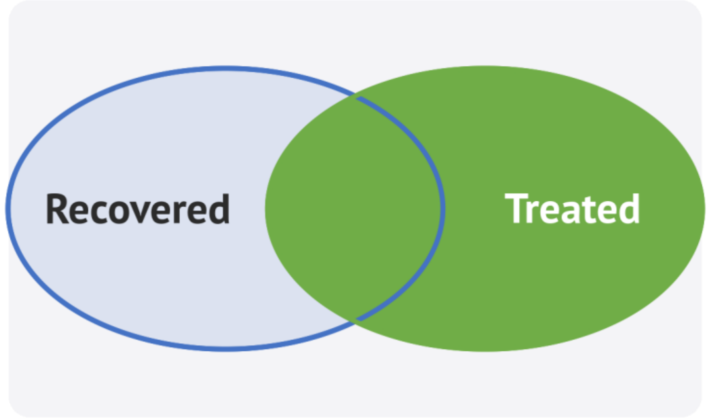
- Multiplication rule
  - $P(A \text{ and } B) = P(A) * P(B)$
  - **Assumes independence between A and B**



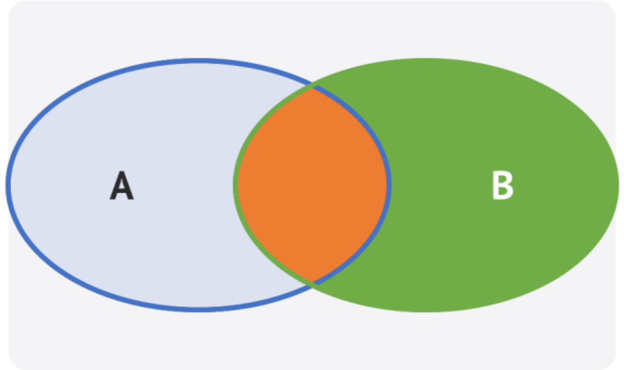
What is conditional probability?

- P(A | B) = P(A and B) / P(B)
- Event B has already occurred → we are only considering that case
- How many A are in B, relative to ALL B's?





Probability rules extended



- Conditional Probability
  - P(A|B) = P(A and B) / P(B)
- Multiplication rule
  - P(A and B) = P(A|B) * P(B)
  - Does **not** assume independence
- Bayes' theorem
  - P(A|B) = P(B|A) * P(A) / P(B)
  - "Flip P(B|A)"

Combination of Conditional probability rule and Multiplication rule

See also lecture 4, slides 40-44

How to compute conditional probability?

Contingency table

→ Conditional proportions: P(Recovered|Treated) = 27/40 = 0.68

Recovered?

Treatment?	No	Yes	Total
No	21	19	40
Yes	13	27	40
Total	34	46	80(= n)

Recovered?

Treatment?	No	Yes	Total
No	0.53	0.47	1
Yes	0.32	0.68	1

Conditional proportion: The proportion of the response variable, for *one level* of the explanatory variable

Today

1. Conditional Probability
   - Definition of conditional probability
   - Probability rules recall

2. **Diagnostic tests**
   - Three methods for calculating conditional probability
   - Probability rules extension

3. Independence revisited
   - Conditional probability is useful to determine independence

4. Closing
   - Next time
   - Example exam question

The Happy Herbologist

"[...] she cultivates crops of a magical plant called green codacle – a flowering plant that when consumed causes a witch or wizard to feel euphoric and relaxed."

Source:
Etz, A., & Vandekerckhove, J. (2017). Introduction to Bayesian Inference for Psychology. https://doi.org/10.31234/osf.io/q46q3

The Happy Herbologist

"However, it has turned out that one in a thousand codacle plants is afflicted with a mutation that changes its effects: Consuming those rare plants causes unpleasant side effects such as paranoia, anxiety, and spontaneous levitation."

Source:
Etz, A., & Vandekerckhove, J. (2017). Introduction to Bayesian Inference for Psychology. https://doi.org/10.31234/osf.io/q46q3
https://tumblr.com/search/sybil%20trelawney

The Happy Herbologist

Professor Sprout has developed a mutation-detecting spell. The new spell has a 99% chance to accurately detect an existing mutation, but also has a 2% chance to falsely indicate that a healthy plant is a mutant.

What is the probability that a codacle plant is a mutant, when the spell says that it is a **mutant**?

What is the probability that a codacle plant is a mutant, when the spell says that it is **healthy**?

Source:
Etz, A., & Vandekerckhove, J. (2017). Introduction to Bayesian Inference for Psychology. https://doi.org/10.31234/osf.io/q46q3
https://tumblr.com/search/sybil%20trelawney

Some terminology of diagnostic testing

Mutant	Pos	Neg	Prob
Yes (M)			1
No (M^c)			1

“The new spell has a 99% chance to accurately detect an existing mutation”

Some terminology of diagnostic testing

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99		1
No (M^c)			1

"The new spell has a 99% chance to accurately detect an existing mutation"

Mutant	Pos	Neg	Prob
Yes (M)	Sensitivity		1
No (M^c)			1

Some terminology of diagnostic testing

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)			1

Mutant	Pos	Neg	Prob
Yes (M)	Sensitivity	False negative	1
No (M^c)			1

When we know P(Pos|M), then we also know P(Neg|M), because: P(Pos|M) + P(Neg|M) = 1

"but also has a 2% chance to falsely indicate that a healthy plant is a mutant."

Some terminology of diagnostic testing

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02		1

Mutant	Pos	Neg	Prob
Yes (M)	Sensitivity	False negative	1
No (M^c)	False positive		1

"but also has a 2% chance to falsely indicate that a healthy plant is a mutant."

Some terminology of diagnostic testing

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

When we know P(Pos|M^c), then we also know P(Neg|M^c), because:
P(Pos|M^c) + P(Neg|M^c) = 1

Mutant	Pos	Neg	Prob
Yes (M)	Sensitivity	False negative	1
No (M^c)	False positive	Specificity	1

Some terminology of diagnostic testing

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

"it has turned out that one in a thousand codacle plants is afflicted with a mutation"

P(M) = 0.001

Mutant	Pos	Neg	Prob
Yes (M)	Sensitivity	False negative	1
No (M^c)	False positive	Specificity	1

P(M): Prevalence or base rate

Today

1. Conditional Probability
   - Definition of conditional probability
   - Probability rules recall

2. Diagnostic tests
   - **Three methods for calculating conditional probability**
   - Probability rules extension

3. Independence revisited
   - Conditional probability is useful to determine independence

4. Closing
   - Next time
   - Example exam question

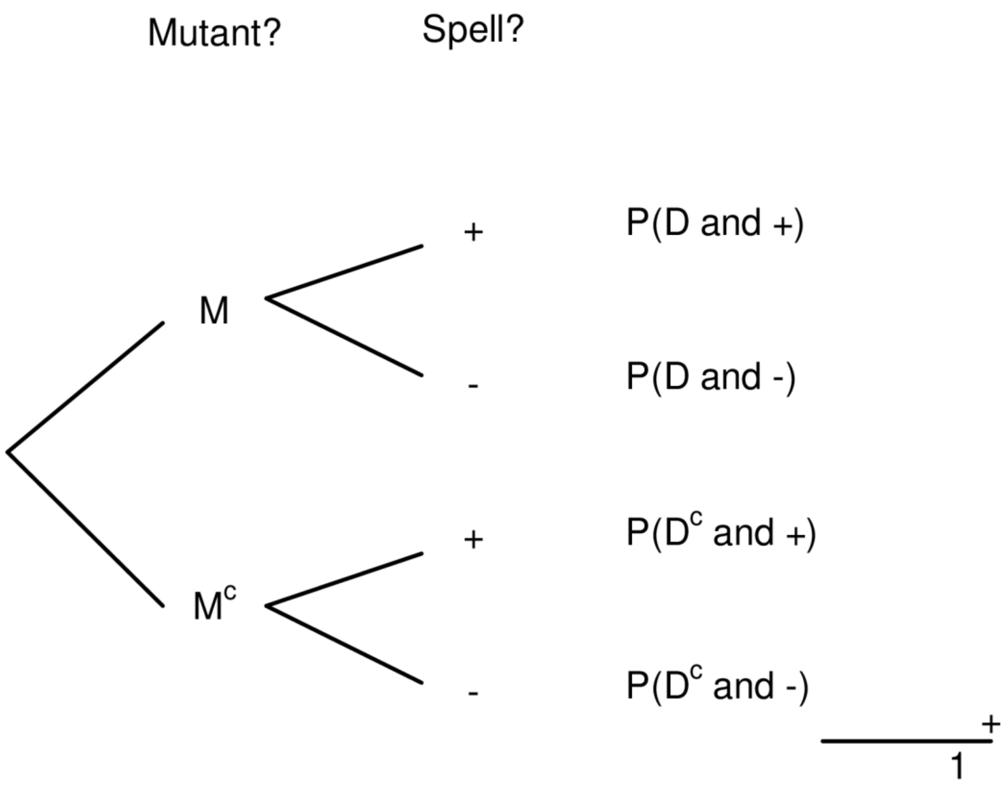
All 3 calculations:

- Start by identifying the different probabilities
- Write them in notation (e.g., P(A | B), where A is mutant and B is the diagnostic test result)
- Use any of the three methods to find the intersection probabilities (e.g., P(A and B))
- When you have the intersection probabilities, you can go to either conditional (e.g., conditional on A, or on B)

Calculation 1: Tree diagram

1. Who received a **positive** spell diagnosis?

2. How many of those are indeed mutants?



Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

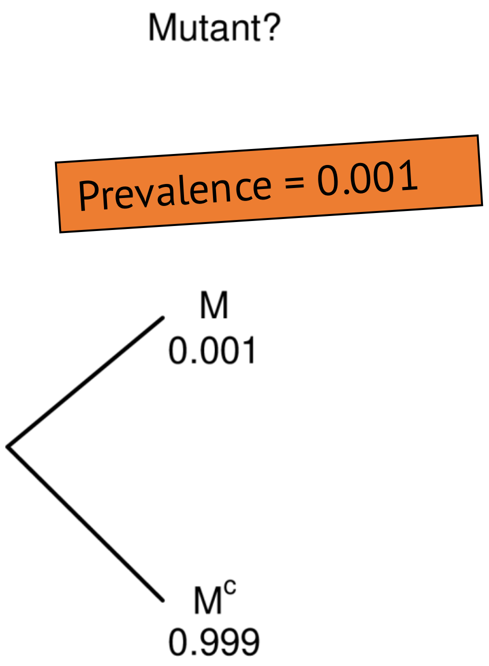
"P(M) = 0.001"

→ P(M^c) = 0.999

Calculation 1: Tree diagram

1. Who received a **positive** spell diagnosis?

2. How many of those are indeed mutants?



Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

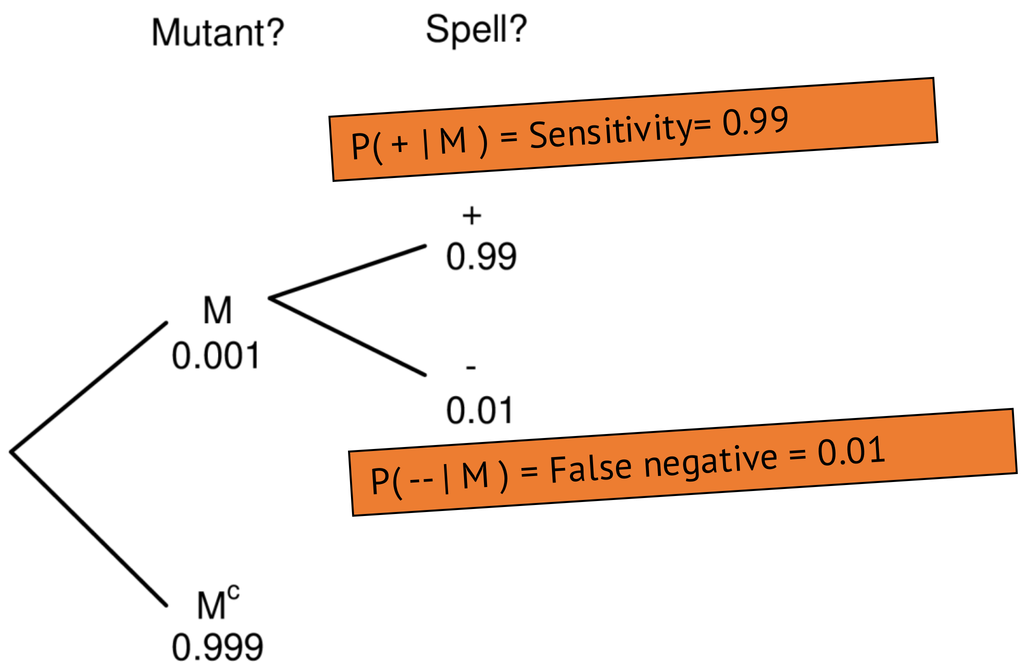
"P(M) = 0.001"

→ P(M^c) = 0.999

Calculation 1: Tree diagram

1. Who received a **positive** spell diagnosis?

2. How many of those are indeed mutants?



Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

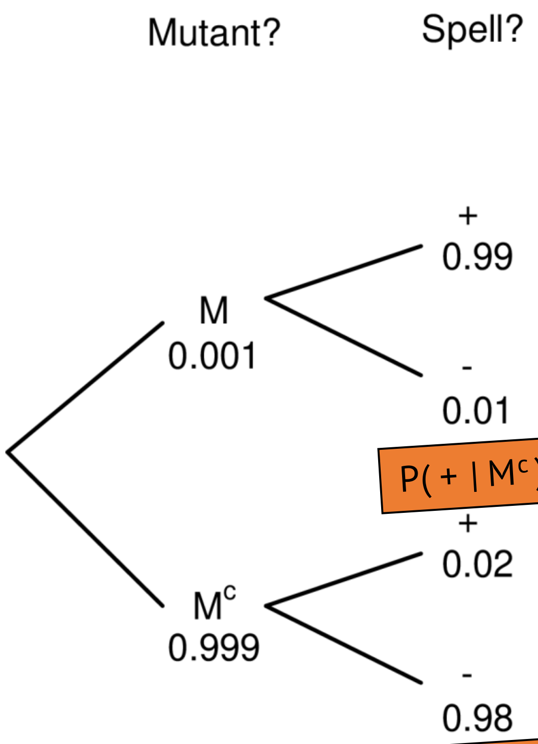
"P(M) = 0.001"

→ $P(M^c) = 0.999$

Calculation 1: Tree diagram

1. Who received a **positive** spell diagnosis?

2. How many of those are indeed mutants?



P(+ | M^c) = False positive = 0.02

P(-- | M^c) = Specificity = 0.98

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

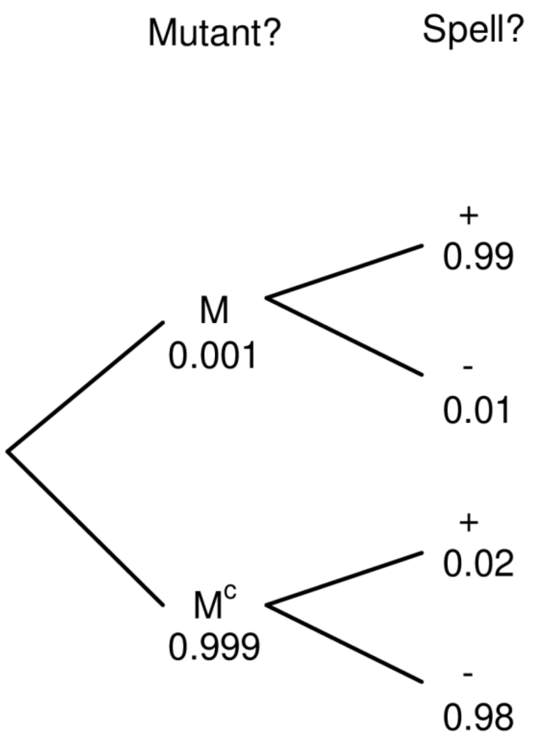
"P(M) = 0.001"

→ P(M^c) = 0.999

Calculation 1: Tree diagram

1. Who received a **positive** spell diagnosis?

2. How many of those are indeed mutants?



Now we can calculate the intersection probabilities:
P(M and +) = P(+ | M) * P(M) =
0.99 * 0.001 = 0.00099

P(M and +) = 0.00099

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

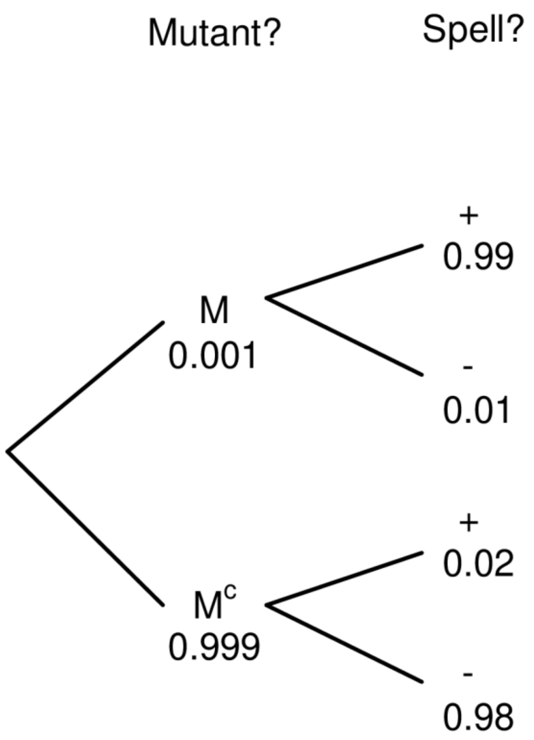
"P(M) = 0.001"

$\rightarrow$ P(M^c) = 0.999

Calculation 1: Tree diagram

1. Who received a **positive** spell diagnosis?

2. How many of those are indeed mutants?



Now we can calculate the intersection probabilities: $P(M \text{ and } -) = P(- \mid M) * P(M) = 0.01 * 0.001 = 0.00001$

$P(M \text{ and } +) = 0.00099$

$P(M \text{ and } -) = 0.00001$

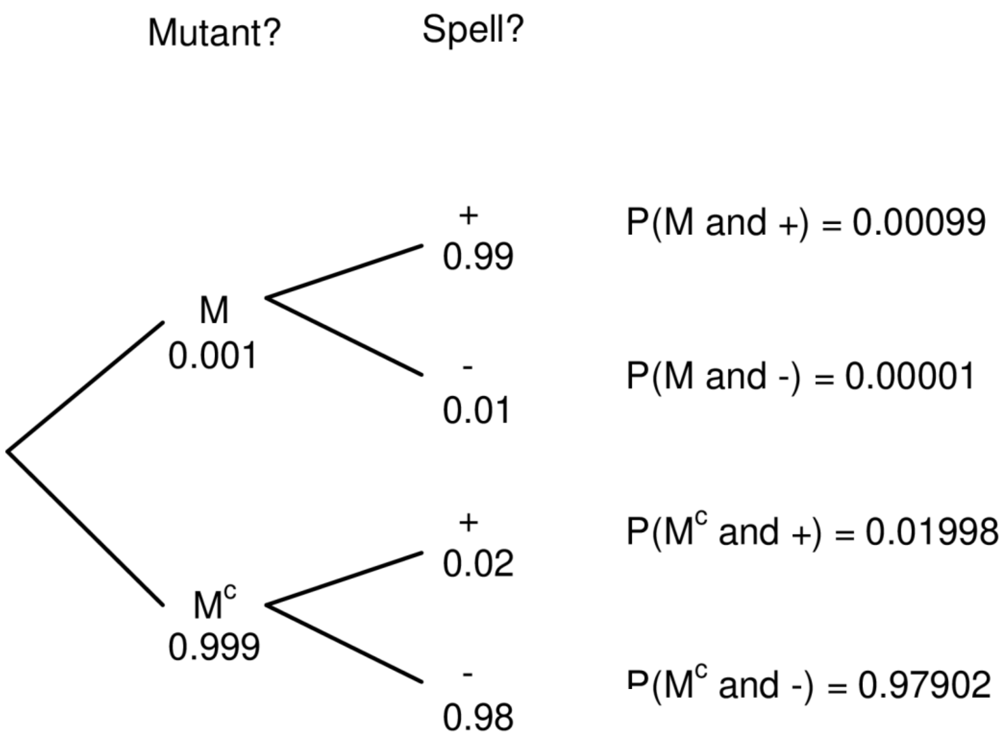
Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

"$P(M) = 0.001$"

$\rightarrow P(M^c) = 0.999$

Calculation 1: Tree diagram

1. Who received a **positive** spell diagnosis?
2. How many of those are indeed mutants?



Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

"P(M) = 0.001"

→ P(M^c) = 0.999

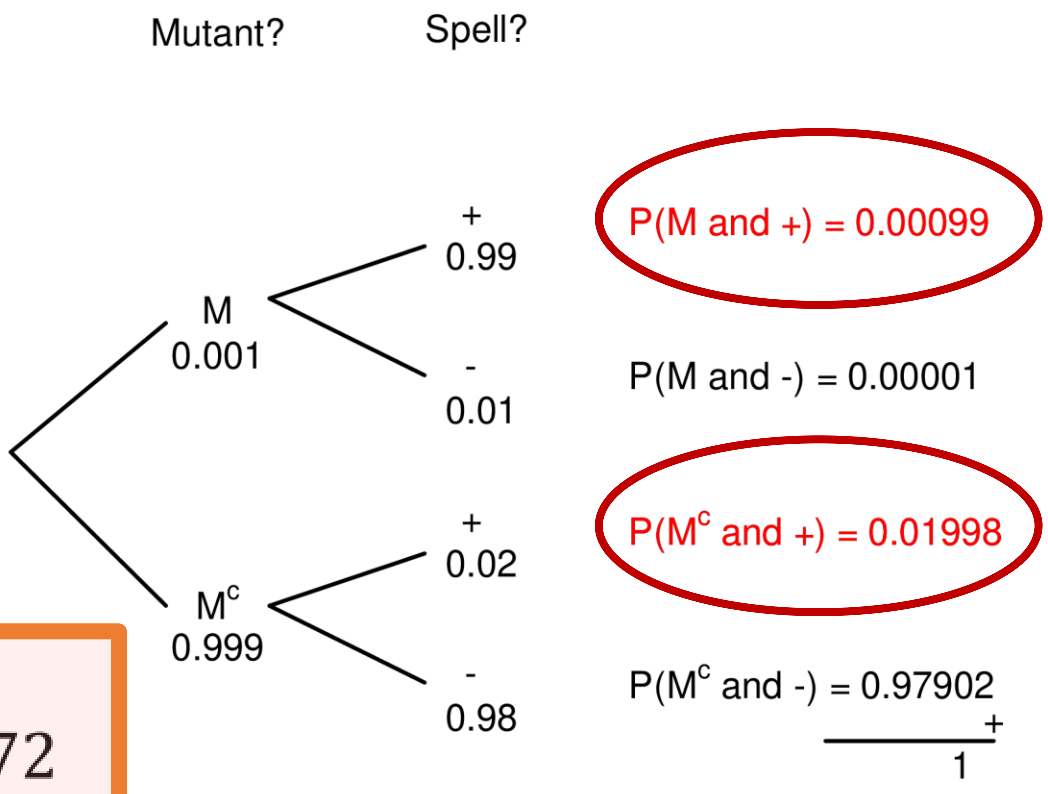
Calculation 1: Tree diagram

1. Who received a **positive** spell diagnosis?

   $\boxed{0.00099+0.01998}$

2. How many of those are indeed mutants?

$$P(M|Pos) = \frac{0.00099}{0.00099+0.01998} = 0.0472$$



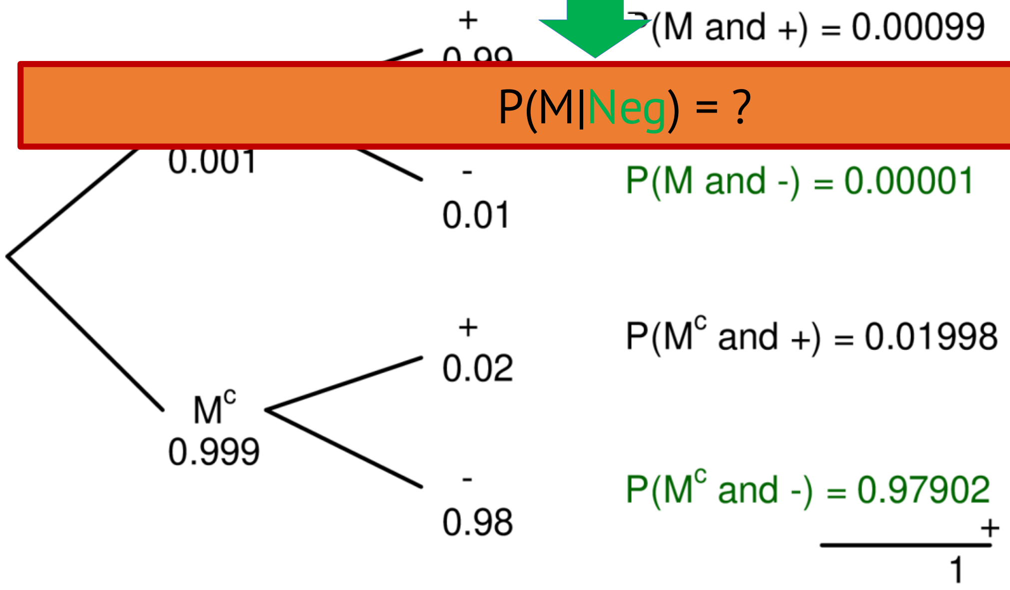
What is the chance of **M**utant if the spell has a **negative** diagnosis?

What is the probability that a codacle plant is a mutant, when the spell says that it is **healthy**?

P(M|Neg) = ?

1. Who received a **negative** spell diagnosis?

2. How many of those are indeed mutants?

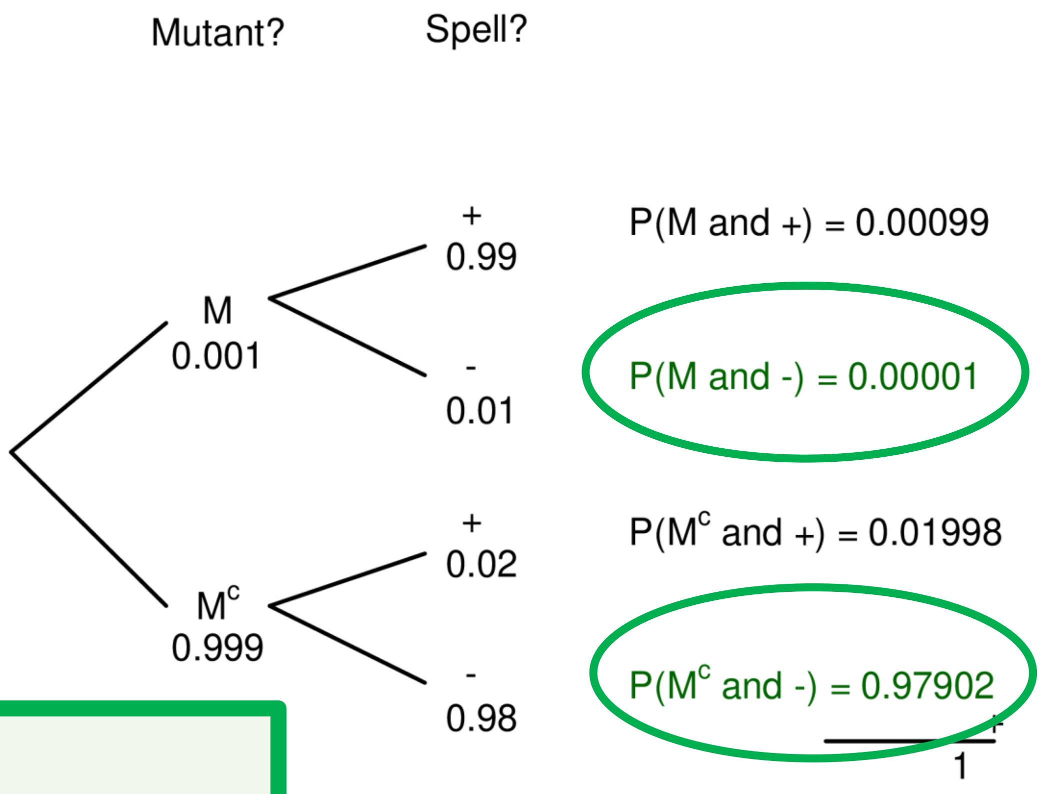


What is the chance of **M**utant if the spell has a **negative** diagnosis?

1. Who received a **negative** spell diagnosis?

$\boxed{0.00001+0.97902}$

2. How many of those are indeed mutants?



$$\boxed{P(M|Neg) = \frac{0.00001}{0.00001+0.97902} = 0.00001021419}$$

Calculation 2: Bayes' theorem

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

"P(M) = 0.001"

What is the probability that a codacle plant is a mutant, when the spell says that it is a **mutant**?

P(M|Pos) = ?

P(M| Pos) = P(Pos | M) * P(M) / P(Pos) = ?

Bayes' theorem

Calculation 2: Bayes' theorem

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

"P(M) = 0.001"

→ P(M^c) = 0.999

complement

P(Pos) = ?

P(Pos) = P(Pos and M) + P(Pos and M^c)

Addition rule for disjoint events

P(Pos and M) = P(Pos|M) * P(M)
P(Pos and M^c) = P(Pos|M^c) * P(M^c)

Multiplication rule

P(M| Pos) = P(Pos | M) * P(M) / P(Pos) = ?

Calculation 2: Bayes' theorem

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

"P(M) = 0.001"

→ P(M^c) = 0.999

P(Pos) = ?

P(Pos) = P(Pos and M) + P(Pos and M^c)

P(Pos and M) = P(Pos|M) * P(M) = 0.99 * 0.001 = 0.00099
P(Pos and M^c) = P(Pos|M^c) * P(M^c) = 0.02 * 0.999 = 0.01998

P(M| Pos) = P(Pos | M) * P(M) / P(Pos) = ?

Calculation 2: Bayes' theorem

Mutant	Pos	Neg	Prob
Yes (M)	$P(Pos\|M) = 0.99$	$P(Neg\|M) = 0.01$	1
No (M^c)	$P(Pos\|M^c) = 0.02$	$P(Neg\|M^c) = 0.98$	1

"$P(M) = 0.001$"

$\rightarrow P(M^c) = 0.999$

$P(Pos) = ?$

$P(Pos) = P(Pos \text{ and } M) + P(Pos \text{ and } M^c) = 0.00099 + 0.01998 = 0.02097$

$P(Pos \text{ and } M) = P(Pos|M) * P(M) = 0.99 * 0.001 = 0.00099$

$P(Pos \text{ and } M^c) = P(Pos|M^c) * P(M^c) = 0.02 * 0.999 = 0.01998$

$P(M|Pos) = P(Pos|M) * P(M) / P(Pos) = ?$

Calculation 2: Bayes' theorem

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

"P(M) = 0.001"

$\rightarrow$ P(M^c) = 0.999

P(Pos) = ?

P(Pos) = P(Pos and M) + P(Pos and M^c) = 0.00099 + 0.01998 = 0.02097

P(Pos and M) = P(Pos|M) * P(M) = 0.99 * 0.001 = 0.00099
P(Pos and M^c) = P(Pos|M^c) * P(M^c) = 0.02 * 0.999 = 0.01998

P(M| Pos) = P(Pos | M) * P(M) / P(Pos) = 0.99 * 0.001 / 0.02097 = 0.0472

Calculation 3: Create a frequency table using large numbers (say n = 100,000)

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

"P(M) = 0.001"

→ P(M^c) = 0.999

Multiplication rule:
P(A and B) = P(A|B) * P(B)

Mutant	Pos	Neg	Total
Yes (M)	P(Pos|M) * P(M) * n	P(Neg|M) * P(M) * n	
No (M^c)	P(Pos|M^c) * P(M^c) * n	P(Neg|M^c) * P(M^c) * n	
			n

Presented as frequencies instead of probabilities, students solved these problems far more often: 16% to 46% (Gigerenzer & Hoffrage, 1995)

Calculation 3: Create a frequency table using large numbers (say n = 100,000)

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

"P(M) = 0.001"

$\rightarrow$ P(M^c) = 0.999

Mutant	Pos	Neg	Total
Yes (M)	0.99 * 0.001 * n = 99	0.01 * 0.001 * n = 1	99 + 1 = 100
No (M^c)	0.02 * 0.999 * n = 1998	0.98 * 0.999 * n = 97,902	1998 + 97902 = 99900
Total			100,000

Presented as frequencies instead of probabilities, students solved these problems far more often: 16% to 46% (Gigerenzer & Hoffrage, 1995)

Calculation 3: Create a frequency table using large numbers (say n = 100,000)

Mutant	Pos	Neg	Prob
Yes (M)	P(Pos|M) = 0.99	P(Neg|M) = 0.01	1
No (M^c)	P(Pos|M^c) = 0.02	P(Neg|M^c) = 0.98	1

"P(M) = 0.001"

→ P(M^c) = 0.999

Mutant	Pos	Neg	Total
Yes (M)	0.99 * 0.001 * n = 99	0.01 * 0.001 * n = 1	99 + 1 = 100
No (M^c)	0.02 * 0.999 * n = 1998	0.98 * 0.999 * n = 97,902	1998 + 97902 = 99900
Total	99 + 1998 = 2097	1 + 97,902 = 97,903	100,000

What is the proportion of **Pos** that is really a **M**utant?

$$\frac{99}{99 + 1998} = 0.0472$$

All 3 calculations:

- Start by identifying the different probabilities
- Write them in notation (e.g., P(A | B), where A is mutant and B is spell)
- Use any of the three methods to find the intersection probabilities (e.g., P(A and B))
- When you have the intersection probabilities, you can go to either conditional (e.g., conditional on A, or on B)

→ practice: repeat calculation on previous slides, using a different prevalence (20%),

Solution: $P(M \mid Pos) = 0.9252$

Today

1. Conditional Probability
    - Definition of conditional probability
    - Probability rules recall

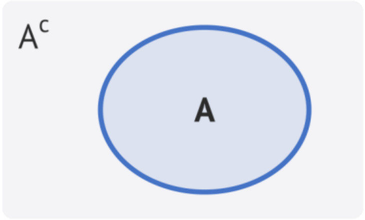
2. Diagnostic tests
    - Three methods for calculating conditional probability
    - **Probability rules extension**

3. Independence revisited
    - Conditional probability is useful to determine independence

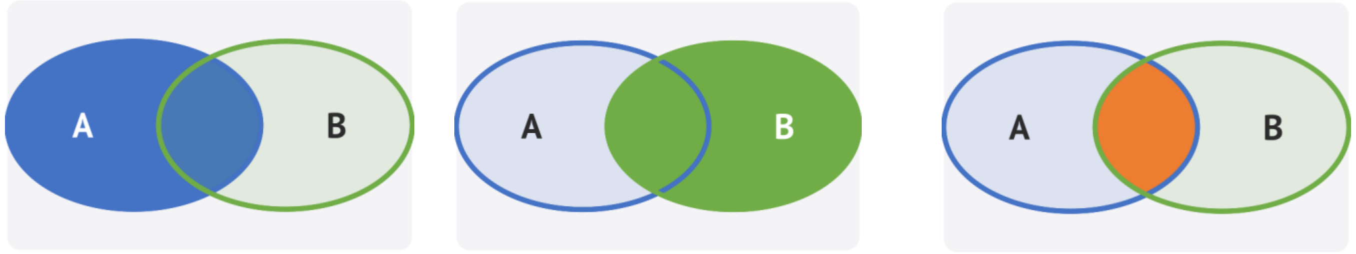
4. Closing
    - Next time
    - Example exam question

Probability rules

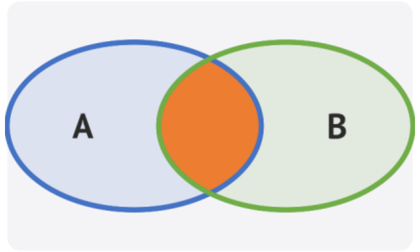
- Complement rule
  - $P(A^c) = 1 - P(A)$



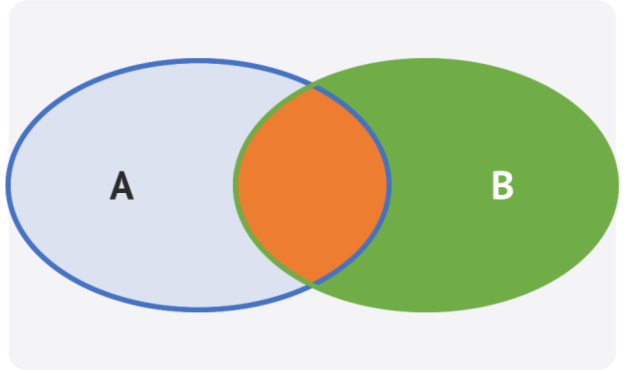
- Addition rule
  - P(A or B) = P(A) + P(B) – P (A and B)



- Multiplication rule
  - P(A and B) = P(A) * P(B)
  - **Assumes independence between A and B**



Probability rules extended



- Conditional Probability
  - P(A|B) = P(A and B) / P(B)
- Multiplication rule
  - P(A and B) = P(A|B) * P(B)
  - **Does not assume independence**
- Bayes' theorem
  - P(B|A) = P(A|B) * P(B) / P(A)
  - "Flip P(A|B)"

Today

1. Conditional Probability
   - Definition of conditional probability
   - Probability rules recall

2. Diagnostic tests
   - Three methods for calculating conditional probability
   - Probability rules extension

3. **Independence revisited**
   - Conditional probability is useful to determine independence

4. Closing
   - Next time
   - Example exam question

Dependence: When conditional proportions differ

Contingency table

Treatment? \ Recovered?	No	Yes	Total
No	21	19	40
Yes	13	27	40
Total	34	46	80(= n)

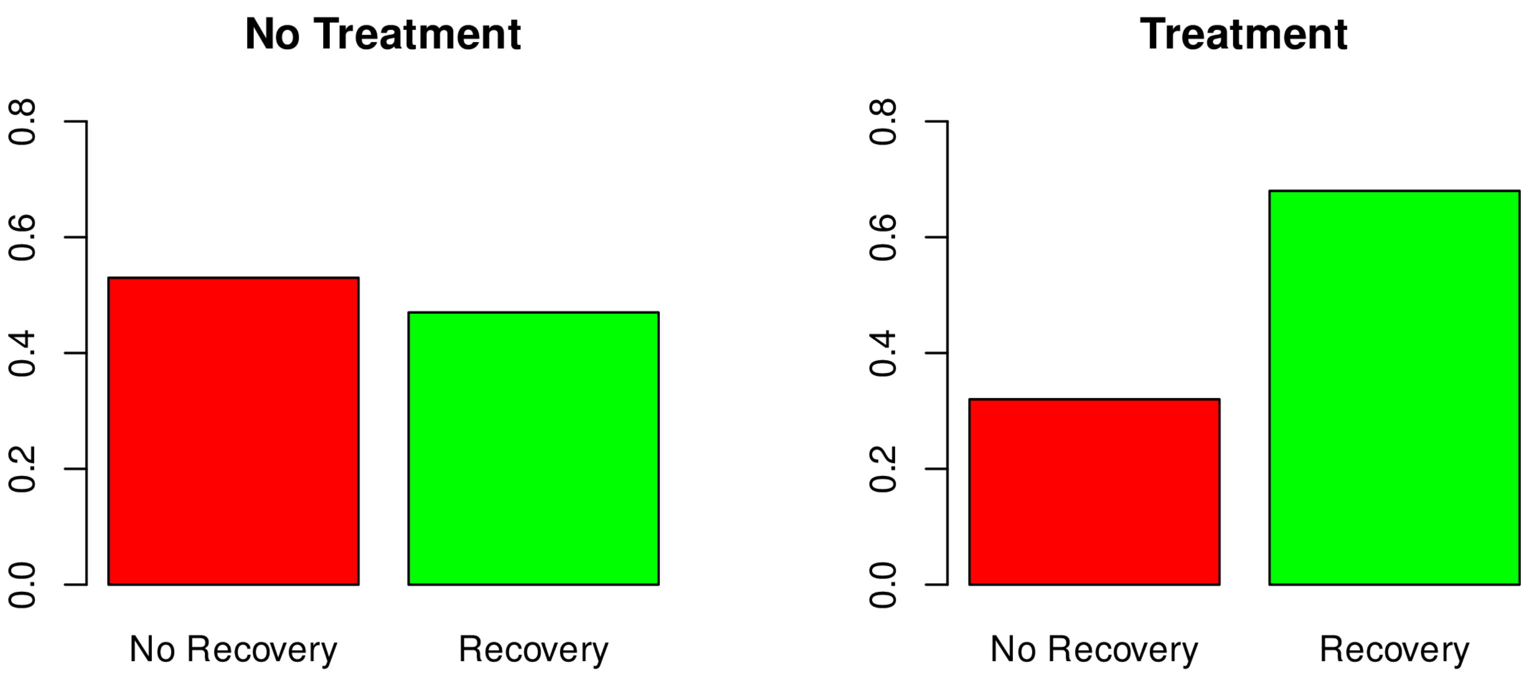
Conditional proportions

(e.g., 13 / 40 = 0.32)

Treatment? \ Recovered?	No	Yes	Total
No	0.53	0.47	1
Yes	0.32	0.68	1
Total			80(= n)

Conditional proportion: The proportion of the response variable, for *one level* of the explanatory variable

Dependence: When conditional proportions differ



- **NOTE**: Conditional probability difference only indicates a dependence, **NOT** causality

Simple scenario: three marbles

- The probability of each marble is 1/3
  - What is the probability of a red marble?

- Now we take two marbles in a row *(**without** replacement)*
  - What is the probability that the *second* marble is red?
  - First draw: P(red)=2/3
  - Second draw: P(red)=1/2 OR P(red)=1
    - P(red)=1/2 if first draw is red
    - P(red)=1 if first draw is green

Independence: Two events are independent if the occurrence of event B *does not change* the probability of event A occurring.

Simple scenario: three marbles

- The probability of each marble is 1/3
  - What is the probability of a red marble?
- Now we take two marbles in a row *(**with** replacement)*
  - What is the probability that the *second* marble is red?
  - First draw: P(red)=2/3
  - Second draw: P(red)=2/3
    - P(red)=2/3 if first draw is red
    - P(red)=2/3 if first draw is green

Independence: Two events are independent if the occurrence of event B *does not change* the probability of event A occurring.

You can check dependence with conditional probability

- P(A|B) = P(A and B) / P(B)

Conditional probability

- P(A and B) = P(A) * P(B) if A and B are independent

Multiplication rule

❼ P(A|B) = P(A and B) / P(B)
❼ P(A|B) = P(A) * P(B) / P(B)
❼ P(A|B) = P(A)

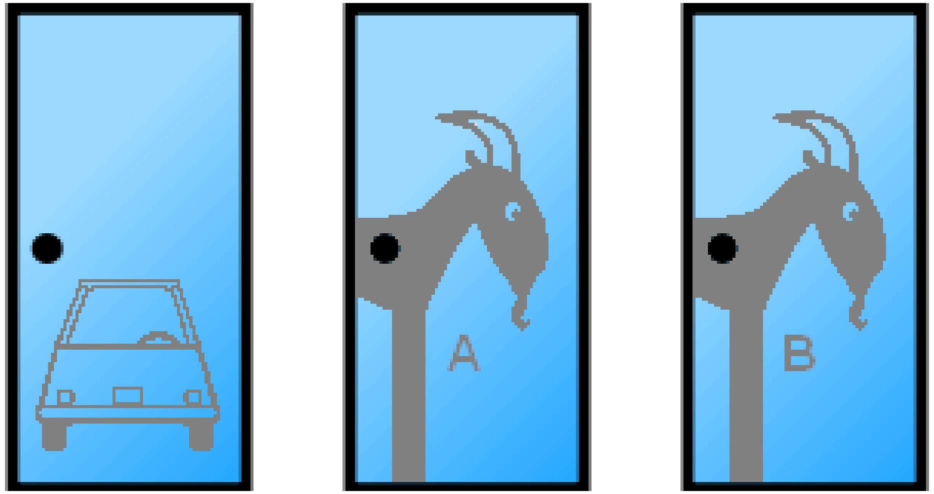
P(Recovery | Treatment) ≠ P(Recovery)
→ dependence!

Probability of someone sneezing right now IF I jump = Probability of someone sneezing right now

Monty Hall problem

Source: http://bestride.com/news/entertainment/throwback-thursday-the-cars-from-lets-make-a-deal

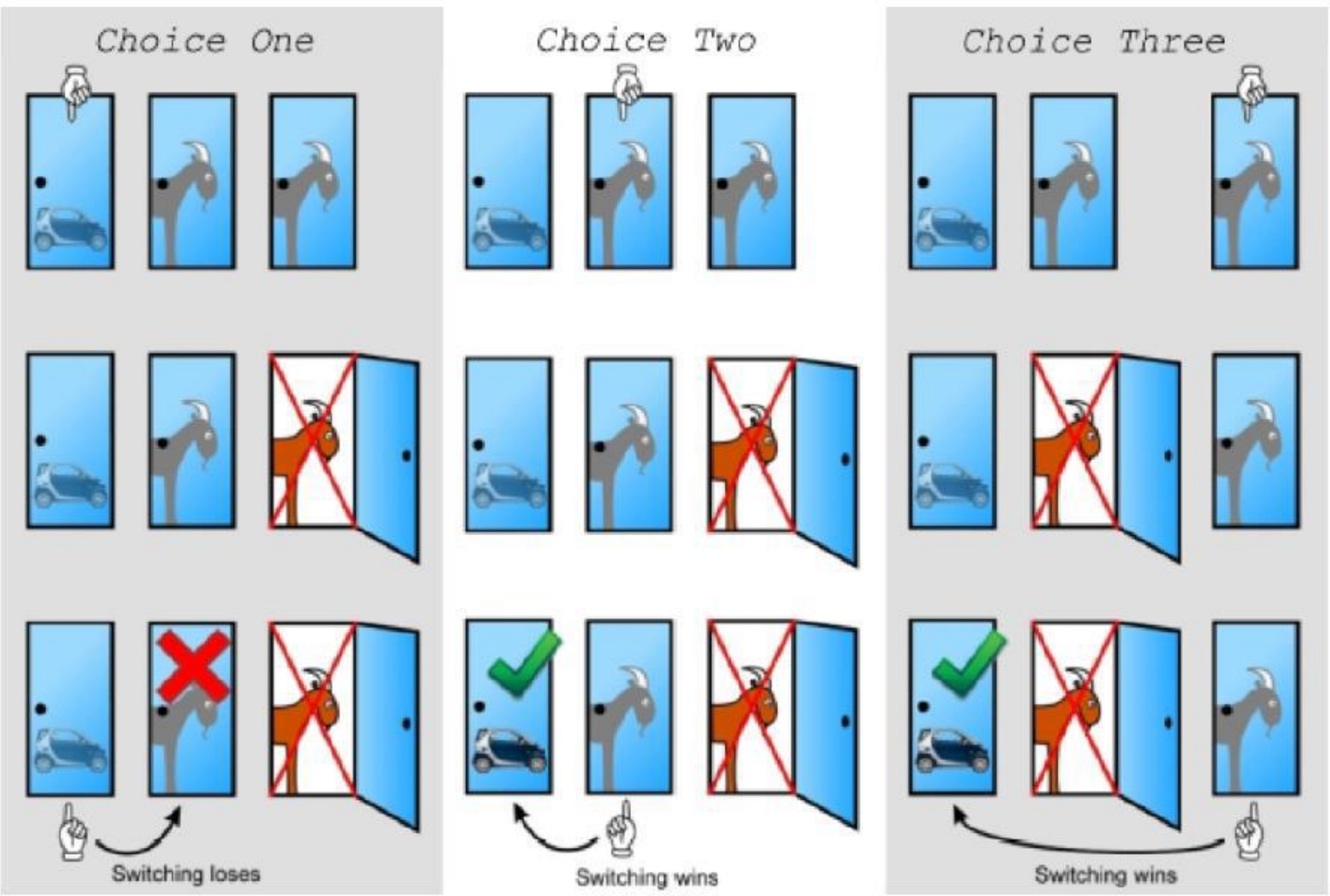
Monty Hall problem



- Choose a door! Behind one door is a car, the others hide a goat
- The show host Monty Hall opens one of the doors and reveals a goat
- Monty Hall asks the Question: "Do you want to switch doors?"
- What should you do?? (what is the probability of winning the car **IF** you switch or stay?)
  - P(winning) = 1/3
  - P(winning | switch) = ?

https://en.wikipedia.org/wiki/Monty_Hall_problem

Monty Hall solution



Source: http://knollubc.blogspot.com/2011/12/twenty-one-years-ago-reader-of-parade.html

Today

1. Conditional Probability
   - Definition of conditional probability
   - Probability rules recall

2. Diagnostic tests
   - Three methods for calculating conditional probability
   - Probability rules extension

3. Independence revisited
   - Conditional probability is useful to determine independence

4. **Closing**
   - Next time
   - Example exam question

Recap of Today

- We can look at the *conditional* probability of an event:
  - = the probability of that event happening, given that another event has happened
  - e.g., the probability of recovery, given that someone was treated
- We expanded the probability rules (conditional/multiplication/Bayes)

- It matters what you condition on: P(A|B) ≠ P(B|A)
  - the opening test caught 90% of real cases, yet only 16% of its positives were correct
  - a low base rate plus an imperfect test makes most positives false

A and B independent	A and B dependent
P(A|B) = P(A)	P(A|B) = P(A and B) / P(B)
P(A and B) = P(A) * P(B)	P(A and B) = P(A|B) * P(B)
	P(A and B) = P(B|A) * P(A)

Next time

Exam 1

Exams

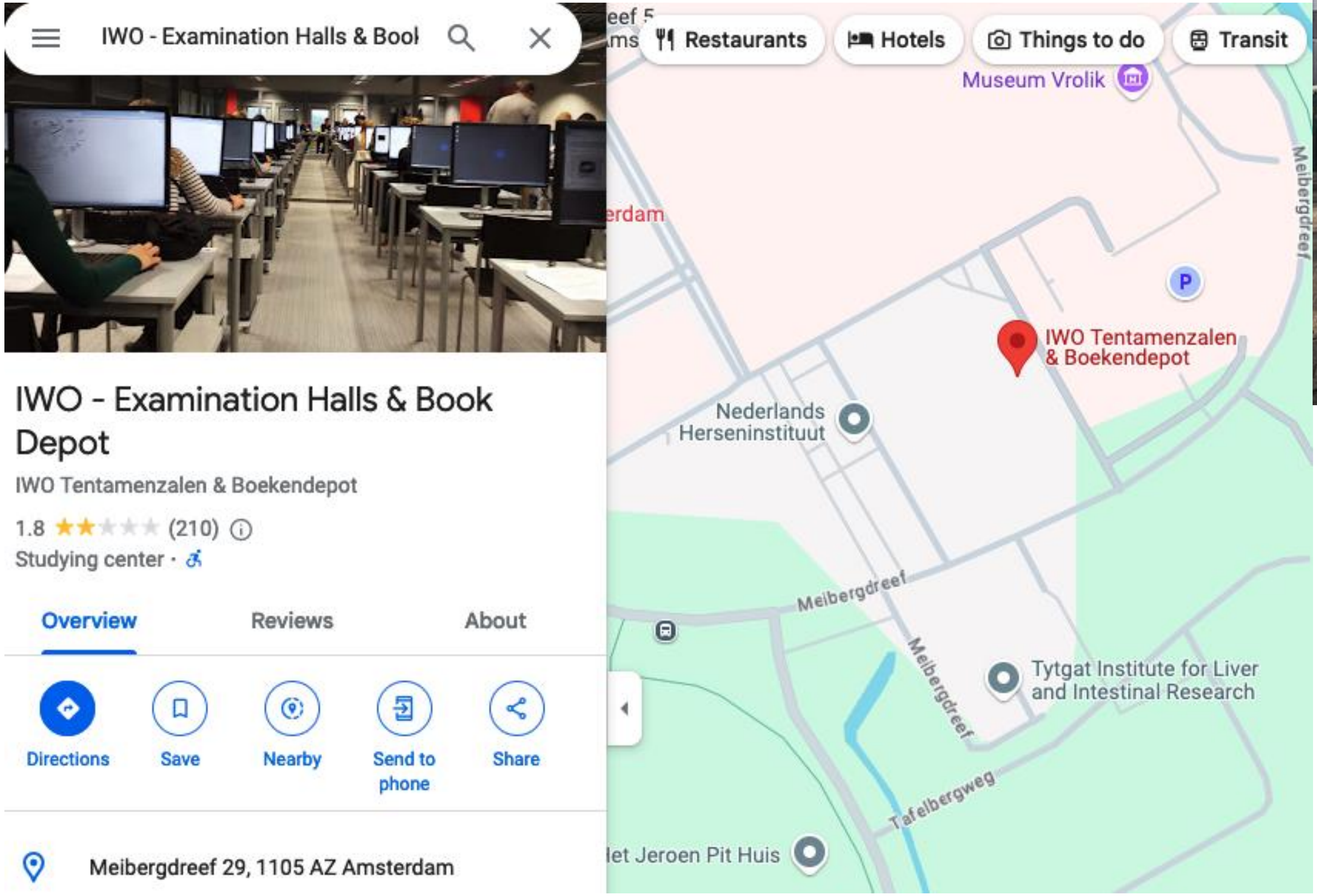
- There are four interim exams: See Canvas for the exam material
- Each interim exam:
  - 25 multiple choice questions (25 points)
  - Open question (5 points)
  - Total: 30 points
  - 90 minutes

  - Practice with the software
    - (e.g., Ans calculator)
  - Practice with the calculations (speed)
  - RM/S questions proportional to lectures

Exams

- ‘Introductory psychology and brain & cognition’ and ‘Research methods and statistics’ exams together
- You can choose order yourself
- Each 90 minutes
- Remember Uvanet id + password (write on piece of paper?)
- Bring ID card!
- Be on time
- Examination regulations on Canvas: https://canvas.uva.nl/courses/46982/pages/examinations-2

Exam Location



https://maps.app.goo.gl/huow75UZiqcLjdRQ8

Exam Tips!

- You can revisit complex questions later → go for the easier/quick ones first
- Rounding → use 4 decimals for inbetween steps
- Open question: note down your inbetween results (e.g., mean and sd when computing a correlation)



Pictures source:

Example exam question

A psychologist is asked to judge a person's math anxiety. She uses a test that in 90% of the cases correctly indicates someone's math anxiety.

Sensitivity

However, the specificity of the test is only 75%.

Specificity

What is the probability that someone with a positive test outcome has indeed math anxiety?

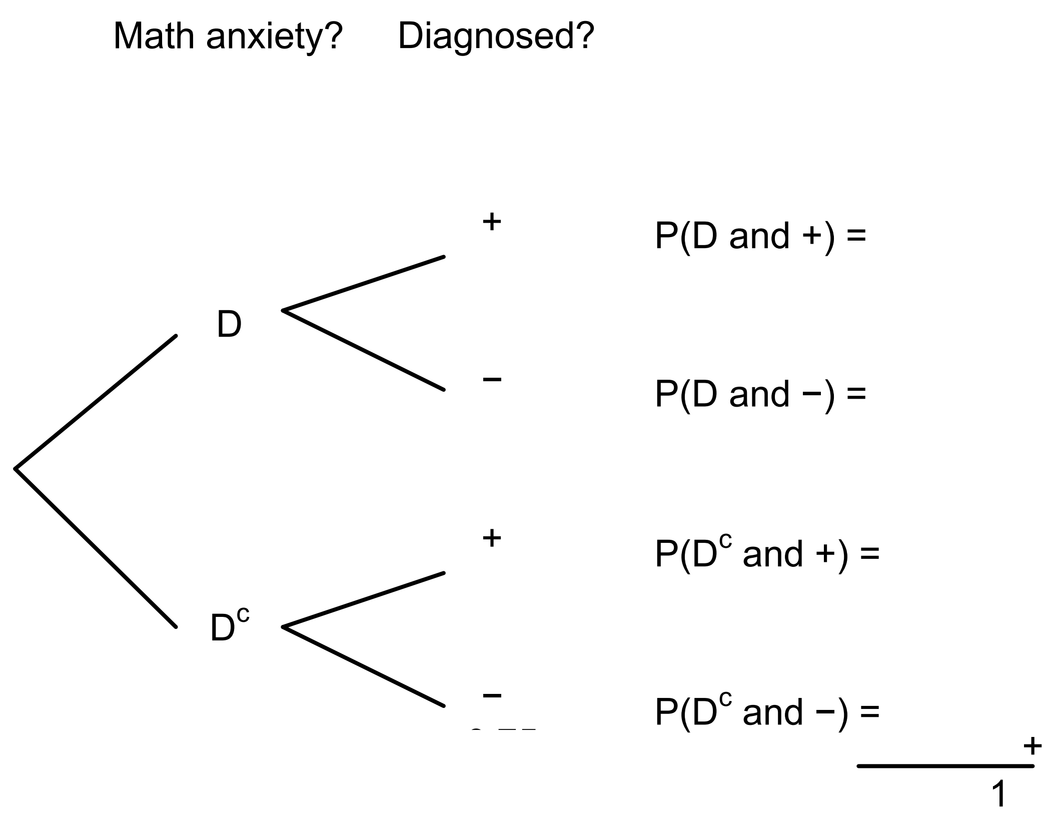
Assume in your answer that the prevalence of math anxiety is 1%.

Prevalence

Answer

1. Who received a **positive** diagnosis?

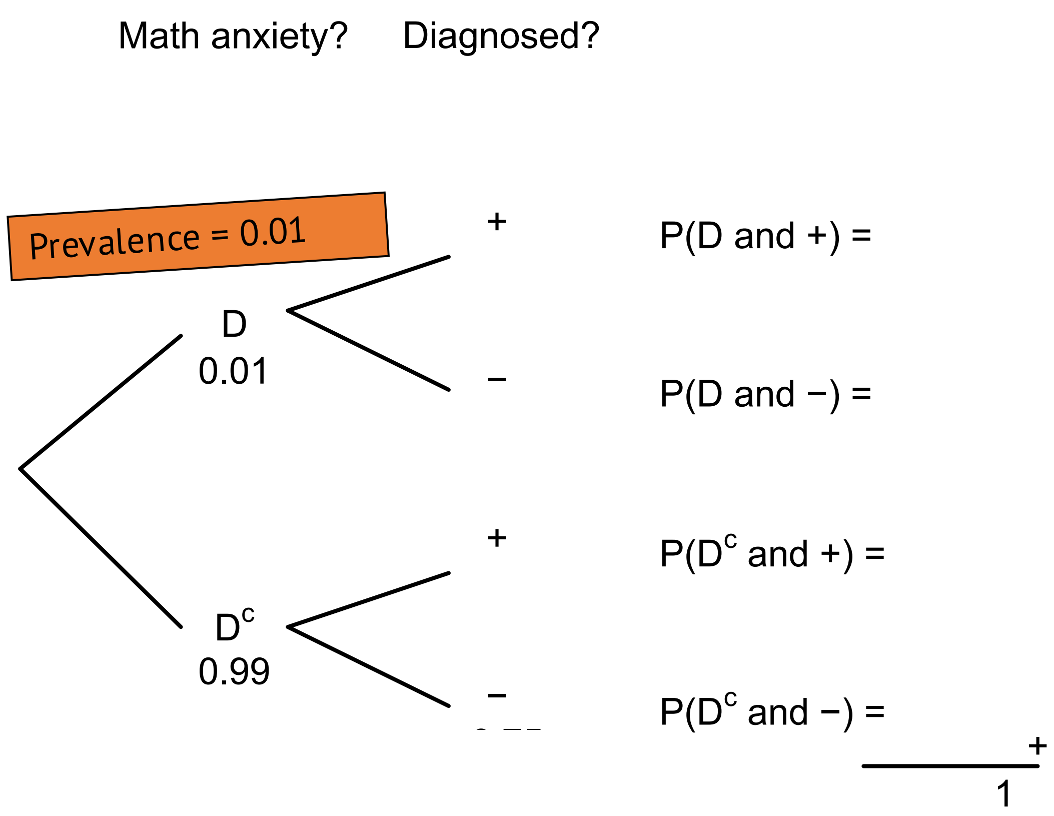
2. How many of those are indeed anxious?



Answer

1. Who received a **positive** diagnosis?

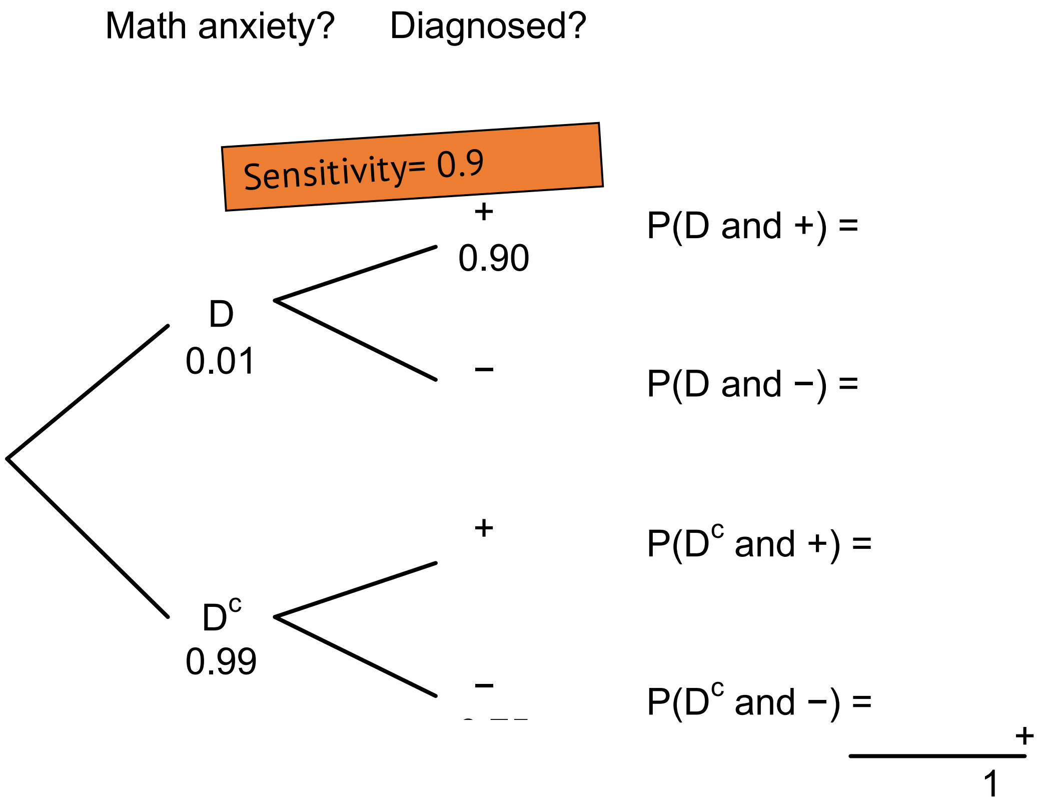
2. How many of those are indeed anxious?



Answer

1. Who received a **positive** diagnosis?

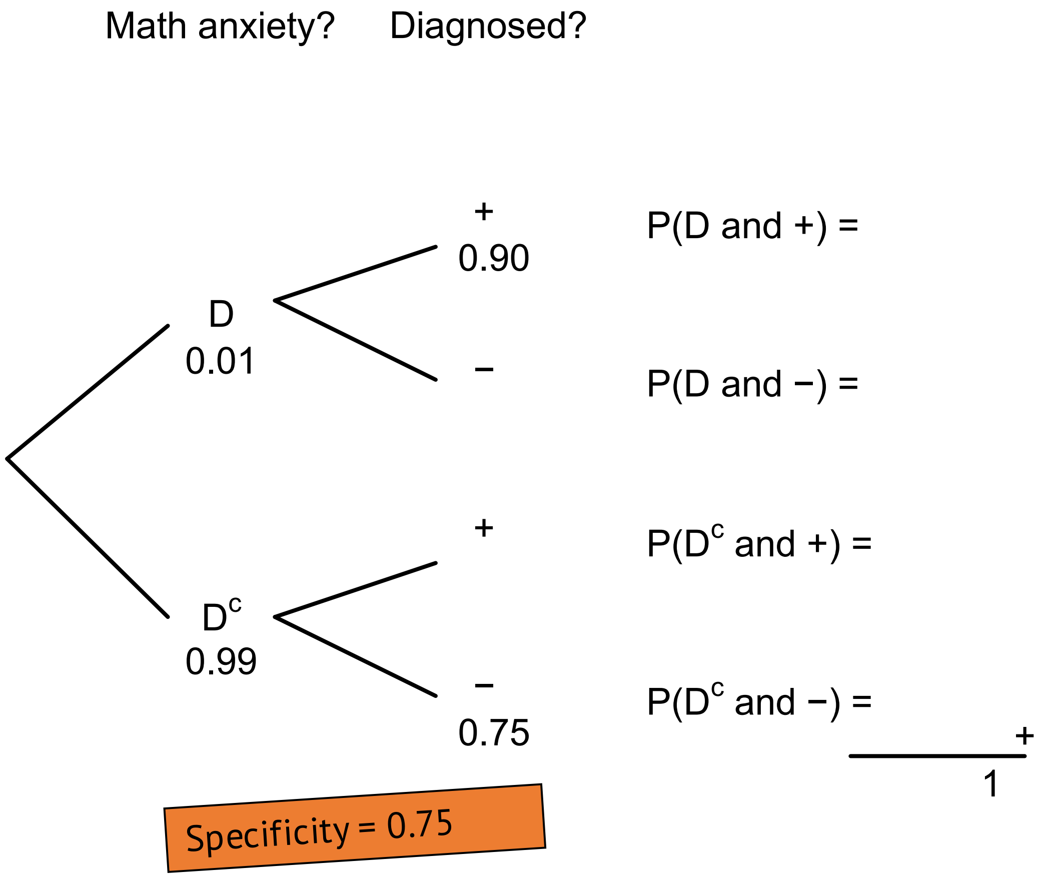
2. How many of those are indeed anxious?



Answer

1. Who received a **positive** diagnosis?

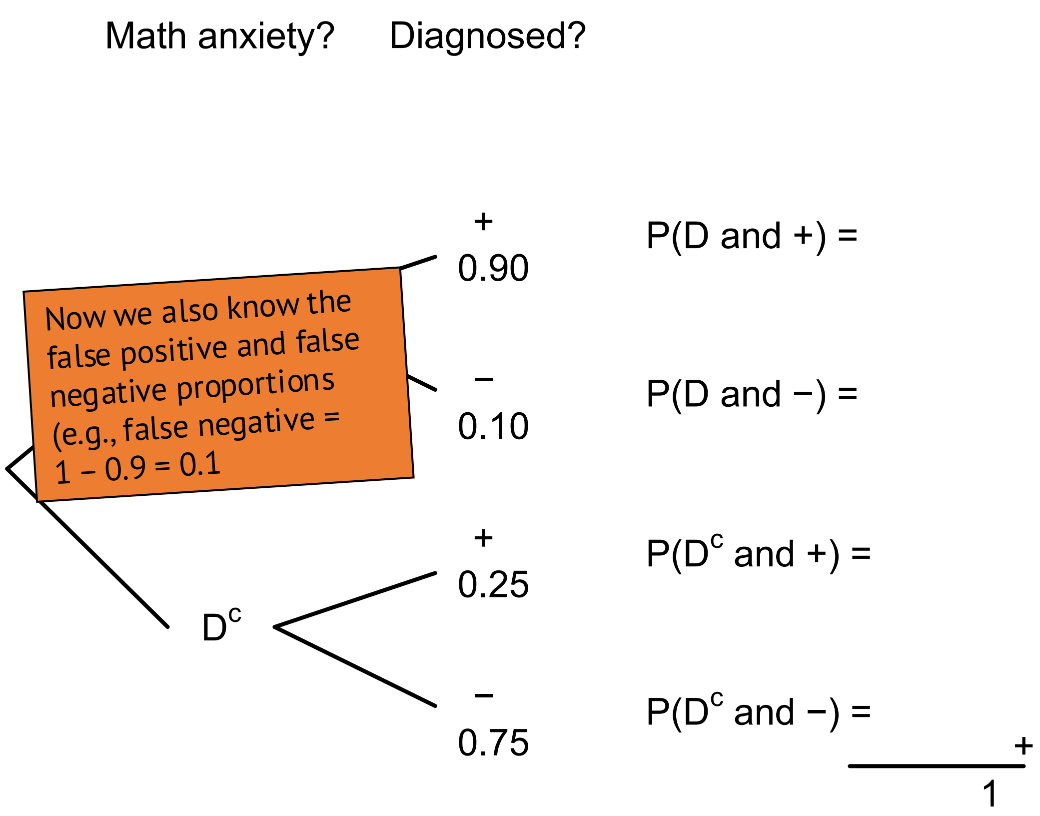
2. How many of those are indeed anxious?



Answer

1. Who received a **positive** diagnosis?

2. How many of those are indeed anxious?



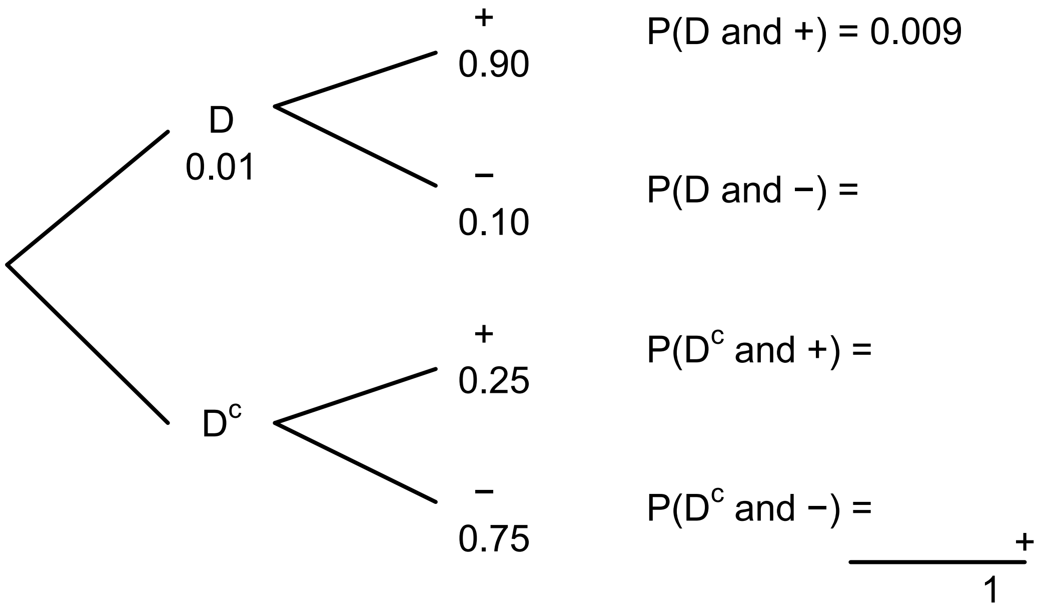
Answer

Math anxiety

Now we can calculate the intersection probabilities:
P(D and +) = P(+ | D) * P(D) =
0.9 * 0.01 = 0.009

1. Who received a **positive** diagnosis?

2. How many of those are indeed anxious?



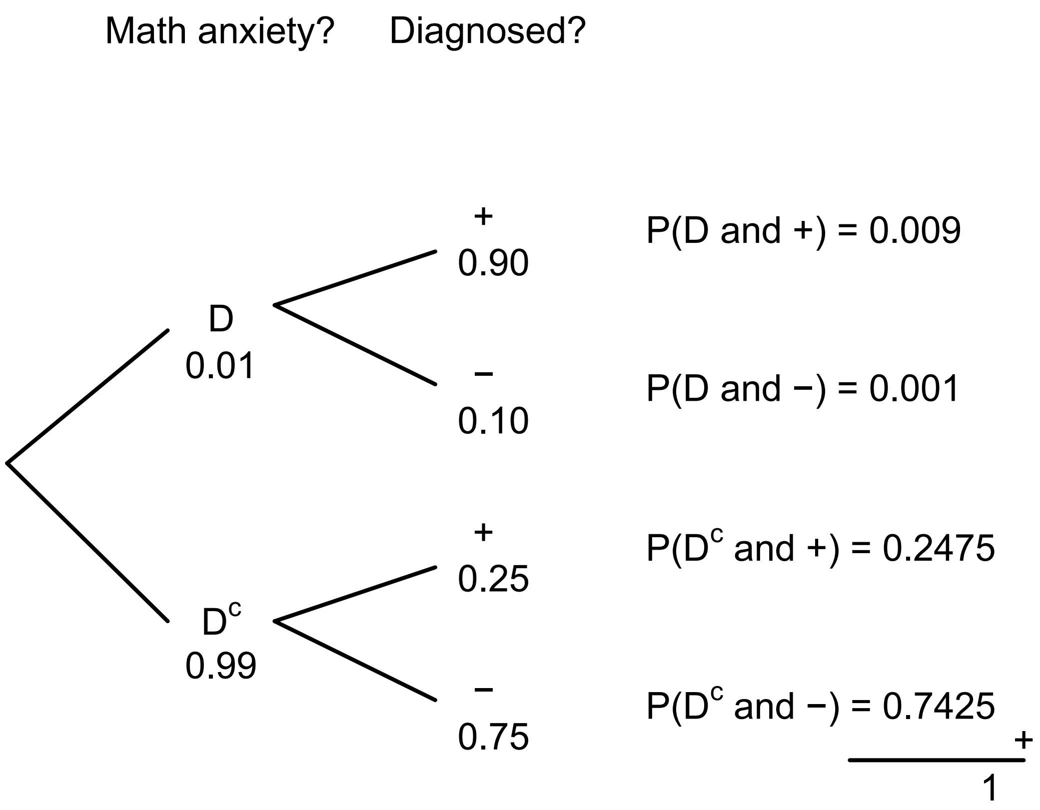
Answer

1. Who received a **positive** diagnosis?

$P(D \text{ and } +) + P(D^c \text{ and } +) =$
$0.009 + 0.2475 = 0.2565$

2. How many of those are indeed anxious?

$$\frac{0.009}{0.009+0.2475} \approx 0.035$$



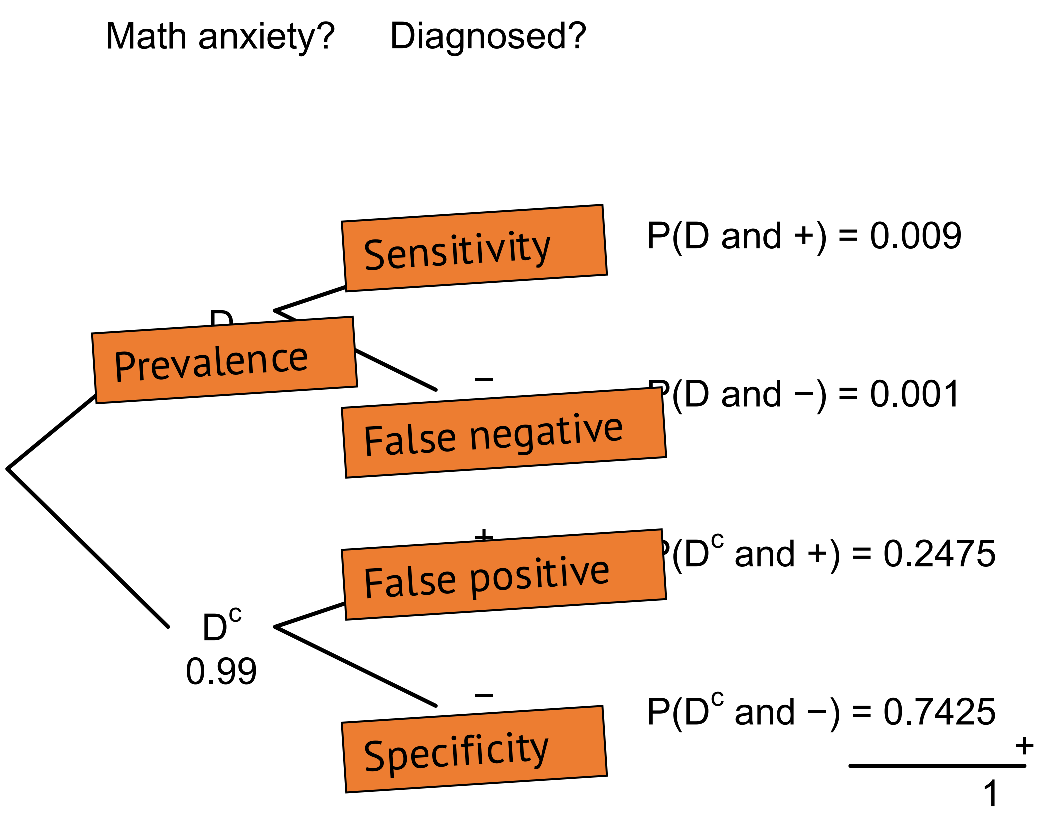
Answer

1. Who received a **positive** diagnosis?

$P(D \text{ and } +) + P(D^c \text{ and } +) =$
$0.009 + 0.2475 = 0.2565$

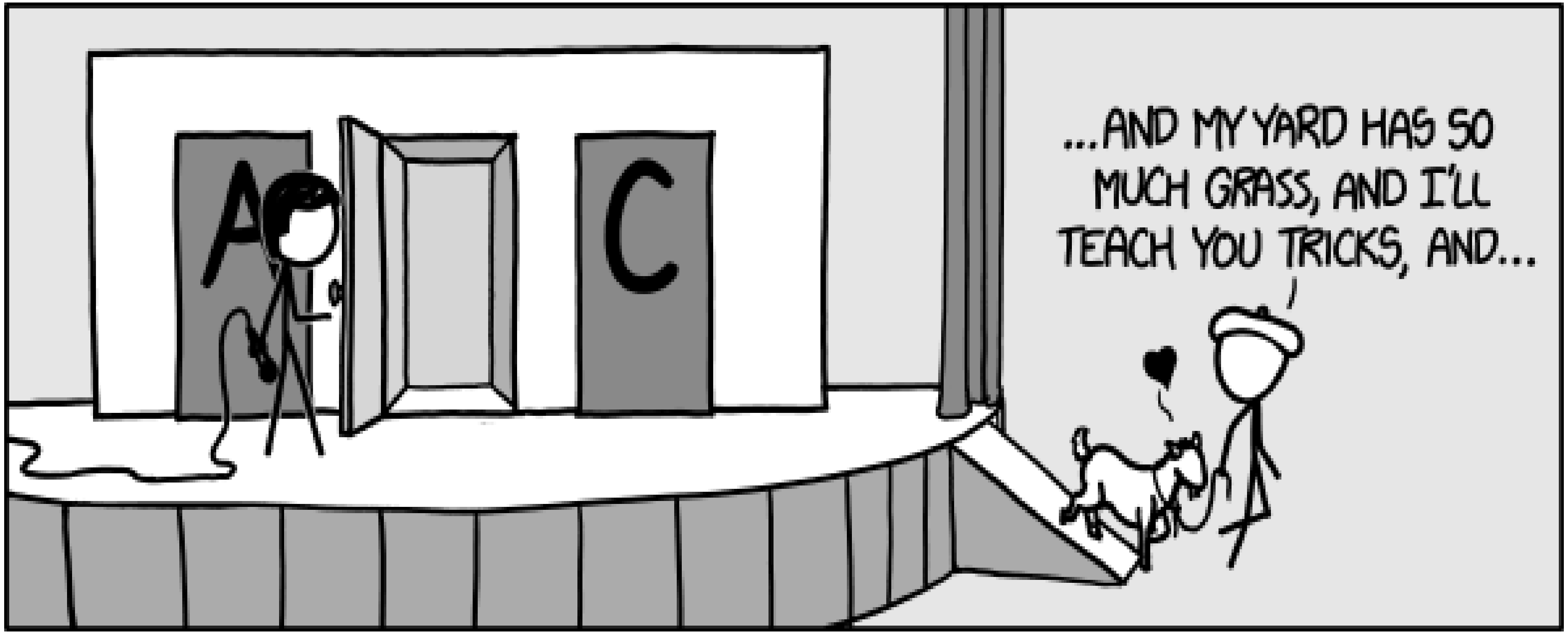
2. How many of those are indeed anxious?

$$\frac{0.009}{0.009+0.2475} \approx 0.035$$



Questions?

Thank you for your attention



https://www.xkcd.com/1282/

Bonus Video

Rosencratz & Guildenstern discuss tossing a coin:

https://www.youtube.com/watch?v=gOwLEVQGbrM

Conditional probabilities raise an interesting notion:

The probability of tossing a coin 1000 times and getting heads each time is equal to 0.5^{1000}, which gives a number with about 300 zeroes behind the decimal (so good luck with that!).

However, the probability of getting heads a 1000th time, when you already have gotten 999 heads, is simply 0.5. This is the probability of getting heads, conditional on observing 999 heads first.

Now think of playing roulette in a casino: it's tempting to think that, if the ball has landed 10 times on red, it will be more likely to land on black next. However, this is a cognitive heuristic, and is known as the gambler's fallacy (Wikipedia Link): we already observed the 10 reds, so we need to look at the *conditional* probability of the ball landing on black!

Highlighted exercises from the book

- 5.30 - Tax audit and income
- 5.32 - Cancer deaths
- 5.35 - Identifying spam

→ try yourself first, then check next slides for answers

5.30

a) P(Under 200k AND Not Audited) =
$\frac{141{,}686}{149{,}919} \approx 0.945$

b) P(Not Audited | Under 200k) = 141,686≈0.994
$\frac{141{,}686}{142{,}525} \approx 0.994$

c) P(Under 200k | Not Audited) =
$\frac{141{,}686}{148{,}985} \approx 0.951$

5.32

a) The first probability (25%) is not conditional, since it's about all deaths. The last three probabilities are conditional, since they are about the probability of a certain cause, IF the death is due to cancer

for example:

$P(A)$ = probability of death due to cancer (=25%)

$P(B)$ = probability of death due to tobacco

$P(B|A) = 30\%$

b)

$P(A \text{ and } B) = ?$

$P(A \text{ and } B) = P(B|A) * P(A) = 0.3 * 0.25 = 0.075$

→ Practice with the probability rules (see also formula sheet on Canvas)

5.35

→ Start with making a table of frequencies that are in the story

ASG

Spam		Yes	No	Total
	Yes	7005	835	7840
Spam	No	48		
	Total	7053		

P(ASG marks spam | spam) = 7005 / 7840 = 0.8935

P(spam | ASG marks spam) = 7005 / 7053 = 0.9932