

$$P(I'M\ NEAR\ |\ I\ PICKED\ UP) =$$
$$P(I\ PICKED\ UP\ |\ I'M\ NEAR) P(I'M\ NEAR)$$


STATISTICALLY SPEAKING, IF YOU PICK UP A
SEASHELL AND DON'T HOLD IT TO YOUR EAR,
YOU CAN PROBABLY HEAR THE OCEAN.

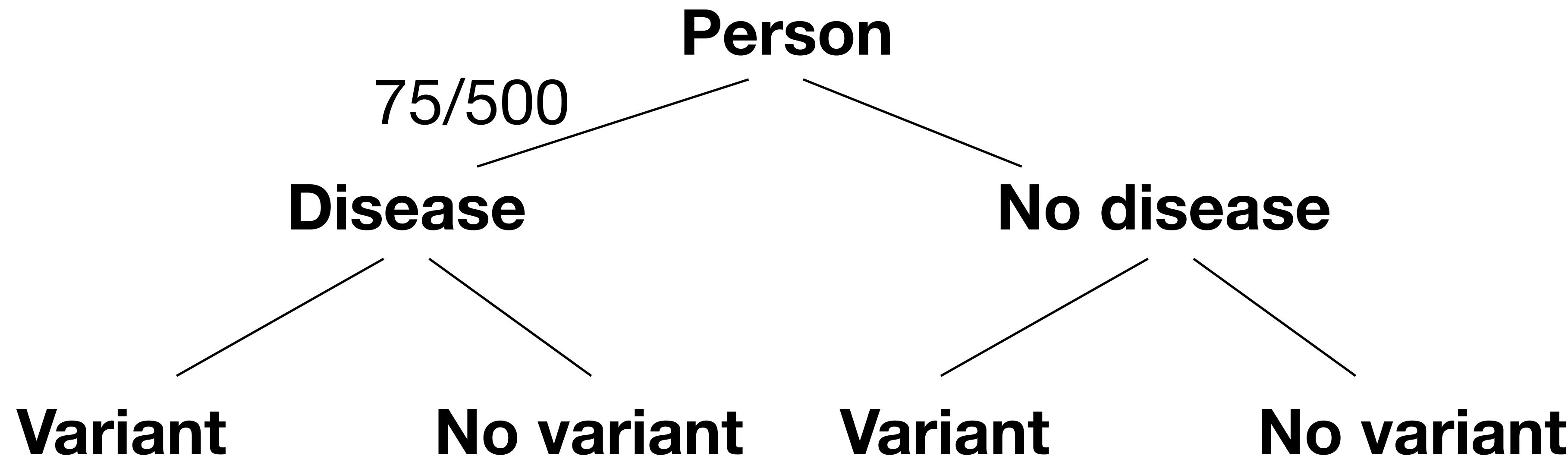
Lecture 03

9.28.21

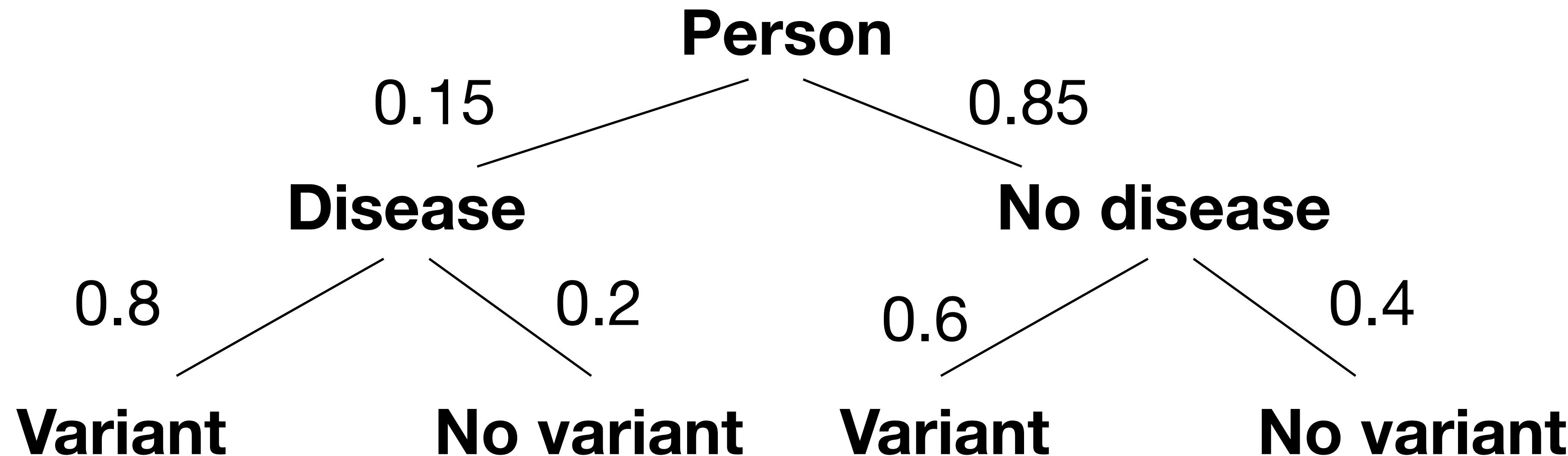
Refresher Quiz

Suppose you have a sample of 500 individuals. 75 of these individuals have a rare genetic disease and you identified a genetic variant that might be associated with this disease. 80% of individuals with the disease have the genetic variant and 60% of individuals without the disease have the genetic variant. What is the probability of randomly selecting an individual with the variant?

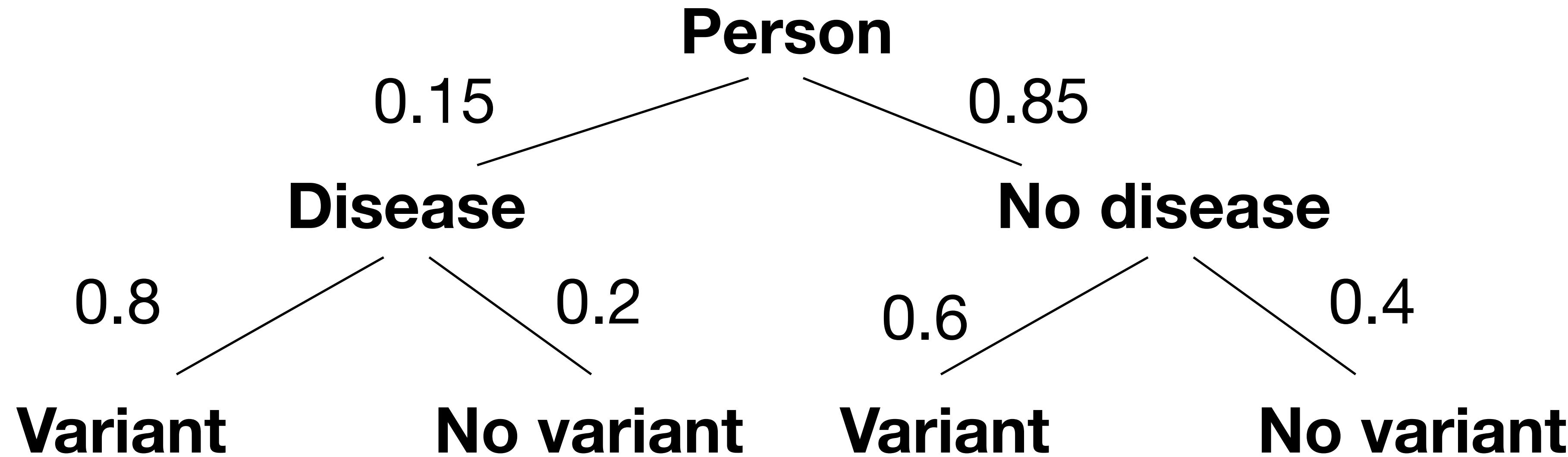
Suppose you have a sample of 500 individuals. 75 of these individuals have a rare genetic disease and you identified a genetic variant that might be associated with this disease. 80% of individuals with the disease have the genetic variant and 60% of individuals without the disease have the genetic variant. What is the probability of randomly selecting an individual with the variant?



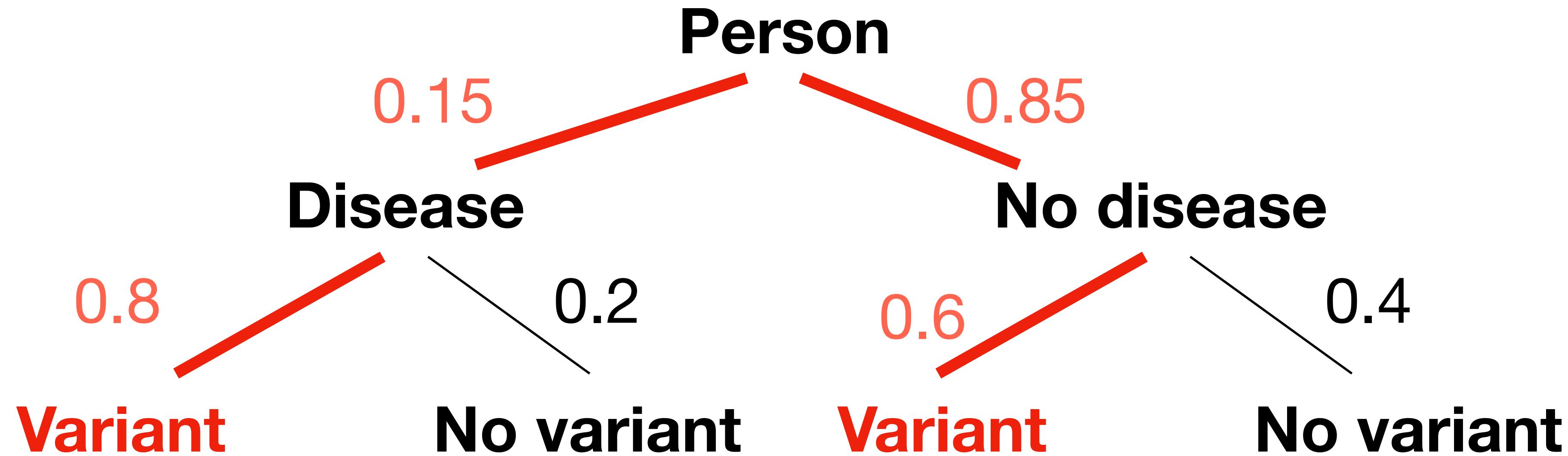
Suppose you have a sample of 500 individuals. 75 of these individuals have a rare genetic disease and you identified a genetic variant that might be associated with this disease. 80% of individuals with the disease and 60% of individuals without the disease have the genetic variant. What is the probability of randomly selecting an individual with the variant?



Suppose you have a sample of 500 individuals. 75 of these individuals have a rare genetic disease and you identified a genetic variant that might be associated with this disease. 80% of individuals with the disease and 60% of individuals without the disease have the genetic variant. **What is the probability of randomly selecting an individual with the variant?**

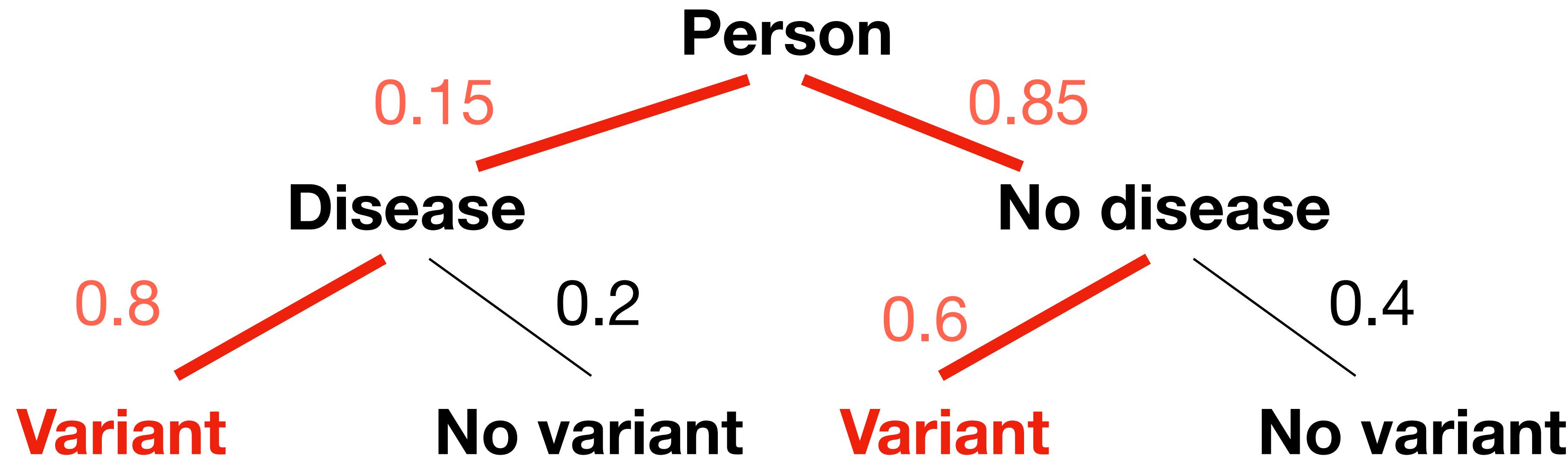


Suppose you have a sample of 500 individuals. 75 of these individuals have a rare genetic disease and you identified a genetic variant that might be associated with this disease. 80% of individuals with the disease and 60% of individuals without the disease have the genetic variant. **What is the probability of randomly selecting an individual with the variant?**



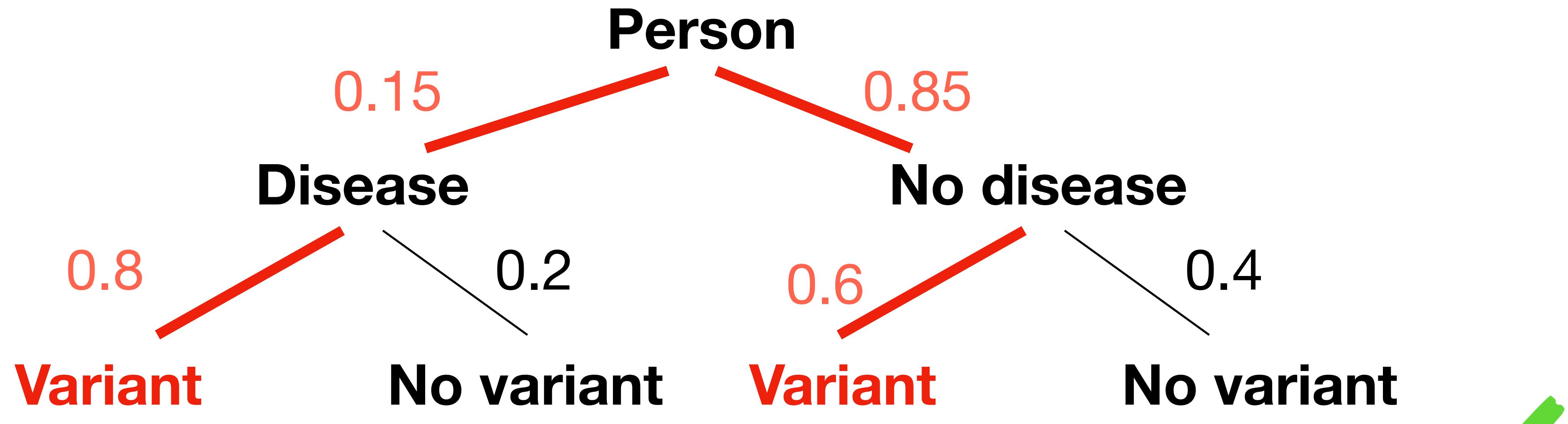
$$\Pr\{\text{variant}\} = \Pr\{\text{variant} \mid \text{disease}\} + \Pr\{\text{variant} \mid \text{no disease}\}$$

Suppose you have a sample of 500 individuals. 75 of these individuals have a rare genetic disease and you identified a genetic variant that might be associated with this disease. 80% of individuals with the disease and 60% of individuals without the disease have the genetic variant. **What is the probability of randomly selecting an individual with the variant?**



$$\Pr\{\text{variant}\} = (0.15)(0.8) + \Pr\{\text{variant} \mid \text{no disease}\}$$

Suppose you have a sample of 500 individuals. 75 of these individuals have a rare genetic disease and you identified a genetic variant that might be associated with this disease. 80% of individuals with the disease and 60% of individuals without the disease have the genetic variant. **What is the probability of randomly selecting an individual with the variant?**



$$\Pr\{\text{variant}\} = (0.15)(0.8) + (0.85)(0.6)$$

$$= 0.63$$



	Income			Total
	Low	Medium	High	
Stressed	526	274	216	1,016
Not stressed	1,954	1,680	1,899	5,533
Total	2,480	1,954	2,115	6,549

Given that someone in this study is from the highest income group, what is the probability that the person is stressed?

$$\Pr\{\text{Stressed} \mid \text{High income}\} = \frac{\frac{\# \text{ Stressed}}{(\text{High income})}}{\frac{\# \text{ Total}}{(\text{High income})}} = \frac{216}{2,115} = 10.21\%$$

$$\boxed{\Pr\{B \mid A\} = \Pr\{A \text{ and } B\} / \Pr\{A\}}$$

What is the probability of randomly selecting an individual who is male and loves chocolate?

Gender	Vanilla	Chocolate	Strawberry	Total
Male	47	15	29	91
Female	48	32	29	109
Total	95	47	58	200

Disjoint or non-disjoint?

What is the probability of randomly selecting an individual who is male and loves chocolate?

Gender	Vanilla	Chocolate	Strawberry	Total
Male	47	15	29	91
Female	48	32	29	109
Total	95	47	58	200

Disjoint or non-disjoint?

$$\Pr\{A \text{ and } B\} = \Pr\{A\} \times \Pr\{B | A\}$$

$$\Pr\{\text{Male} + \text{Chocolate}\} = \Pr\{\text{Male}\} * \Pr\{\text{Chocolate} | \text{Male}\}$$

$$(91/200)$$

$$(15/91)$$

$$= 0.075$$

What is the probability of randomly selecting an individual who is male and loves chocolate?

Gender	Vanilla	Chocolate	Strawberry	Total
Male	47	15	29	91
Female	48	32	29	109
Total	95	47	58	200

Independent or dependent?

$\Pr\{B | A\} = \Pr\{B\}$ (for independent events)

$\Pr\{\text{Chocolate} | \text{Male}\} ? \Pr\{\text{Male}\}$

$$(15/91) \neq (91/200)$$

What is the probability of randomly selecting an individual who is male and loves chocolate?

Gender	Vanilla	Chocolate	Strawberry	Total
Male	47	15	29	91
Female	48	32	29	109
Total	95	47	58	200

Independent or dependent?

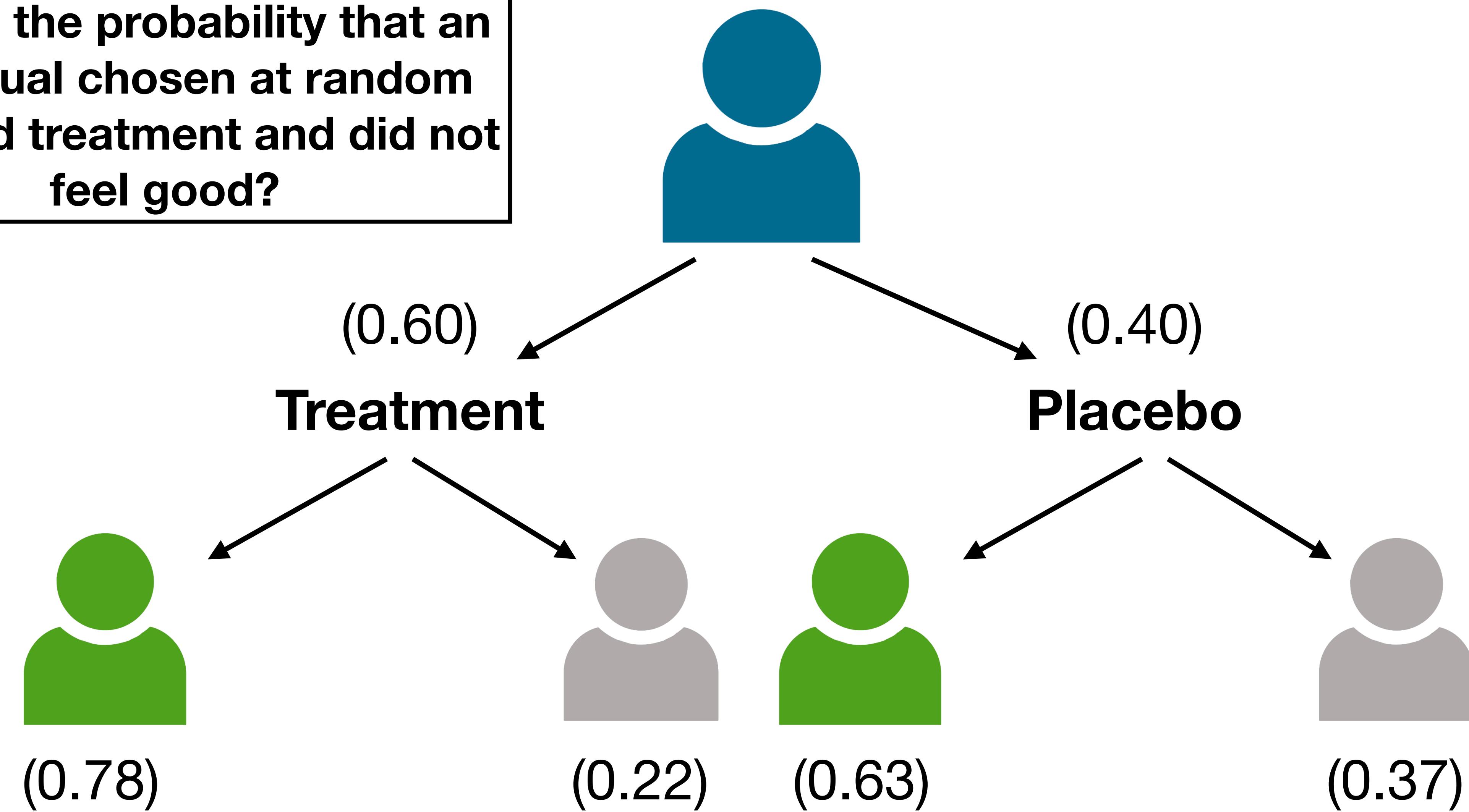
$\Pr\{B | A\} = \Pr\{B\}$ (for independent events)

$\Pr\{\text{Chocolate} | \text{Male}\} ? \Pr\{\text{Male}\}$

$$(15/91) \neq (91/200)$$

Conditional probability with Bayes' Theorem

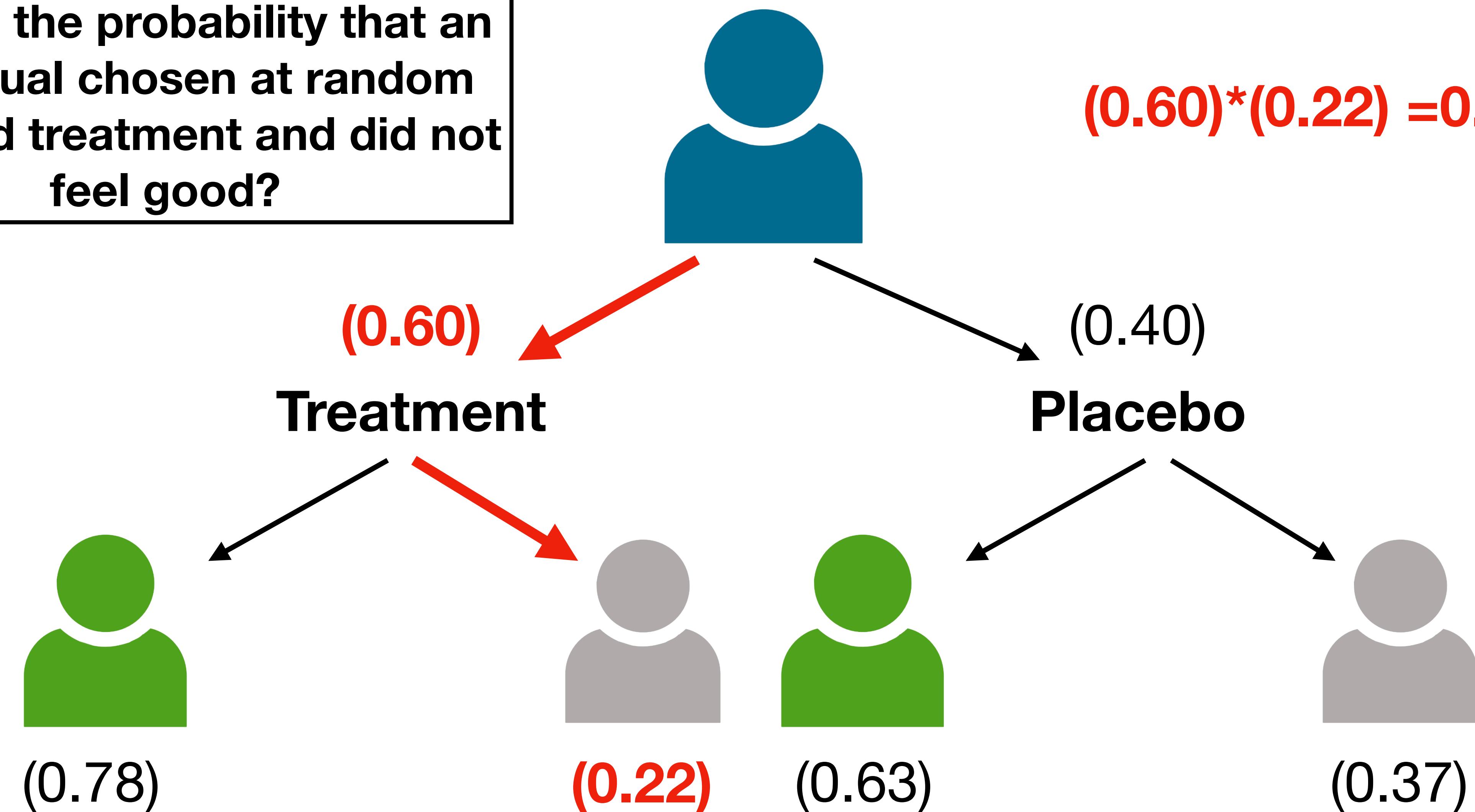
What is the probability that an individual chosen at random received treatment and did not feel good?



Conditional probability with Bayes' Theorem

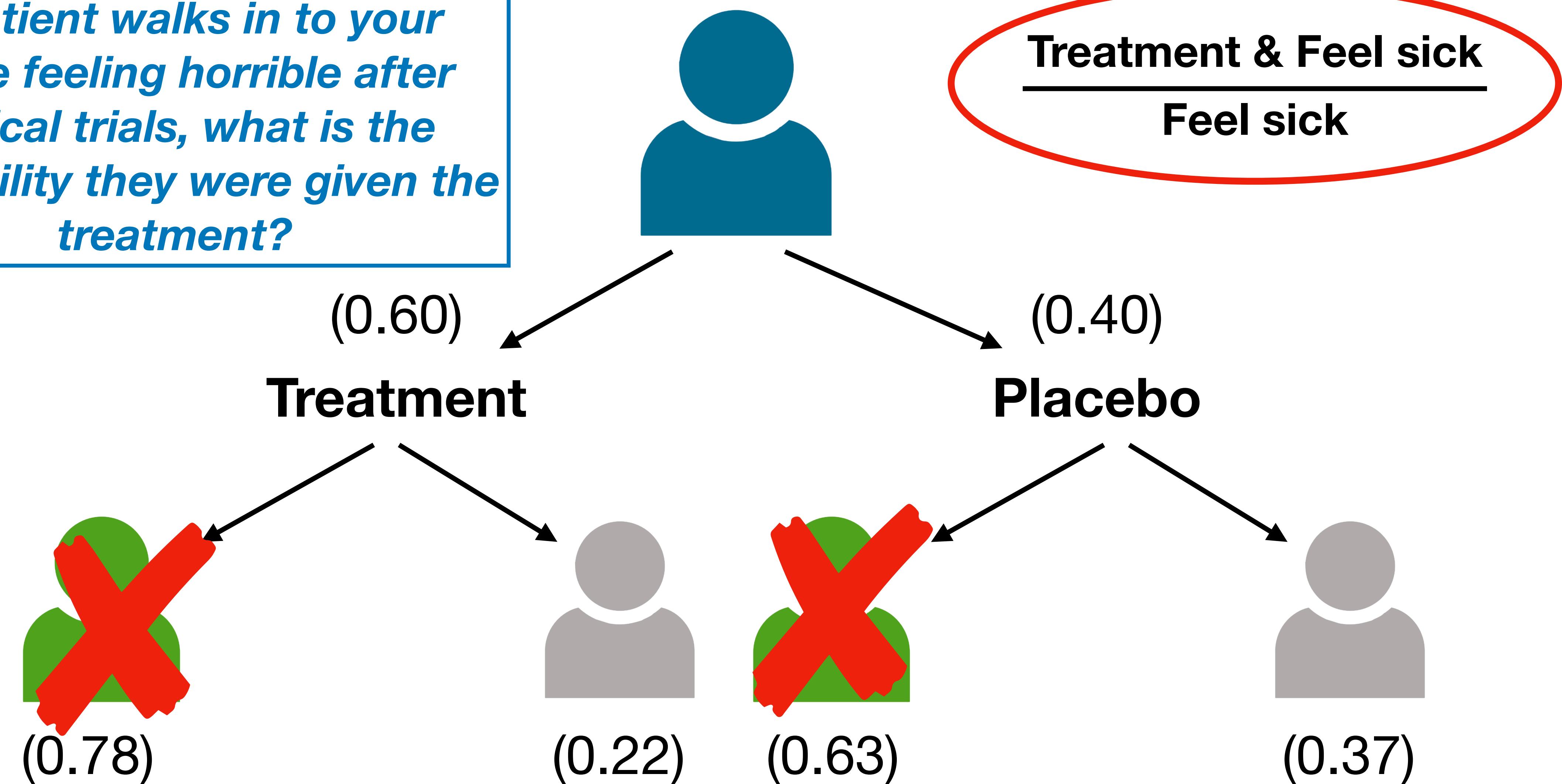
What is the probability that an individual chosen at random received treatment and did not feel good?

$$(0.60) * (0.22) = 0.132$$



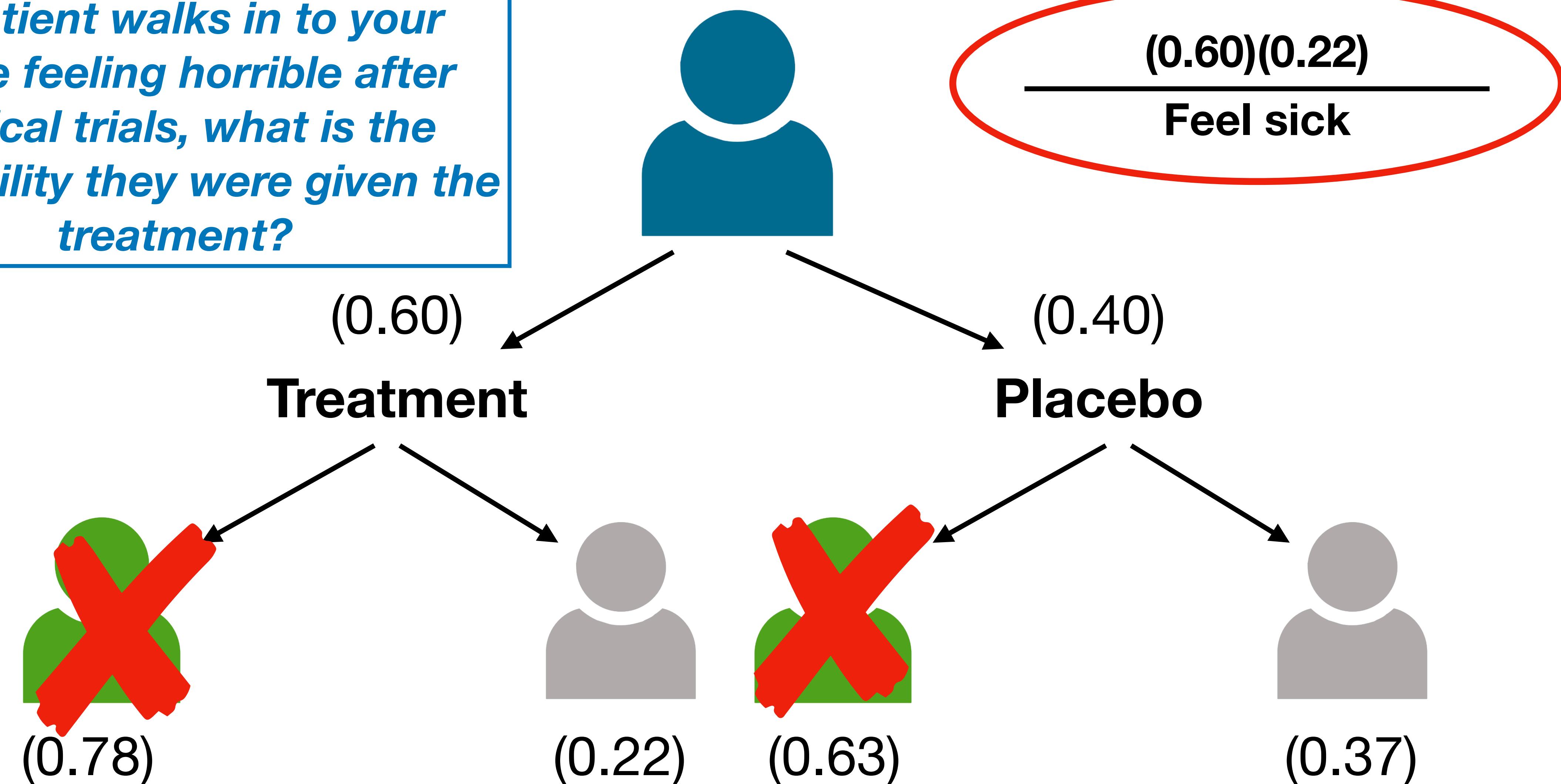
Conditional probability with Bayes' Theorem

A patient walks in to your office feeling horrible after clinical trials, what is the probability they were given the treatment?



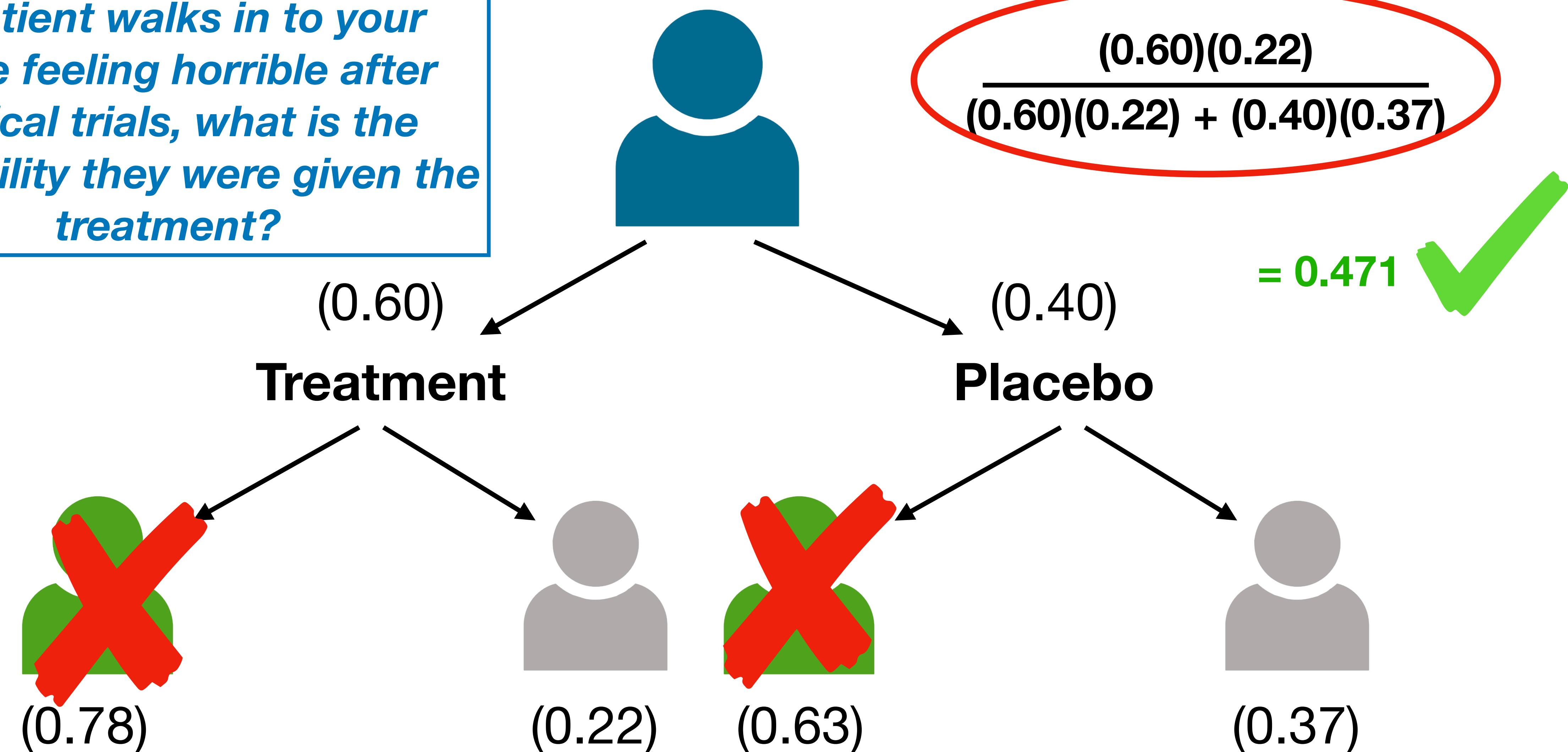
Conditional probability with Bayes' Theorem

A patient walks in to your office feeling horrible after clinical trials, what is the probability they were given the treatment?



Conditional probability with Bayes' Theorem

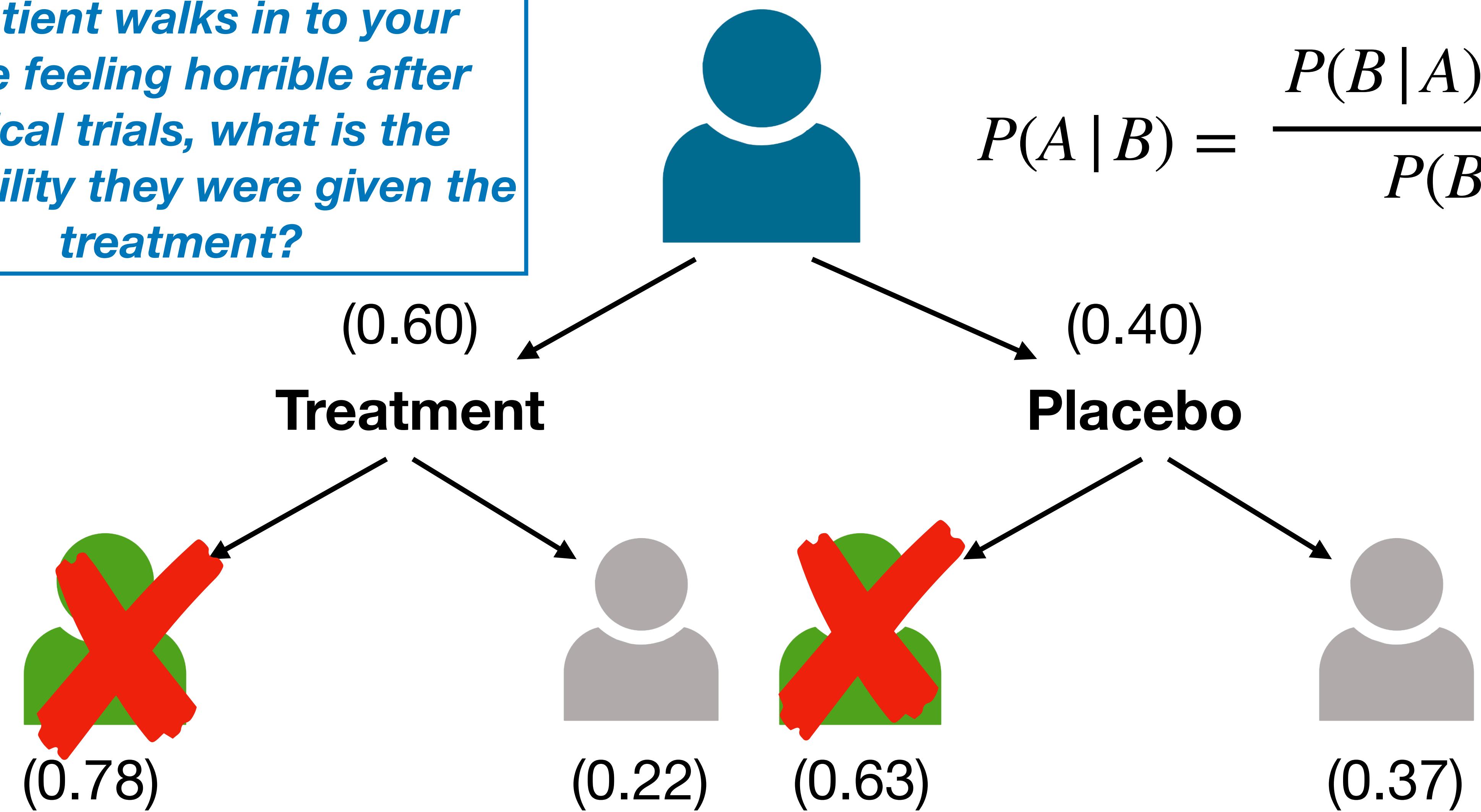
A patient walks in to your office feeling horrible after clinical trials, what is the probability they were given the treatment?



Conditional probability with Bayes' Theorem

A patient walks in to your office feeling horrible after clinical trials, what is the probability they were given the treatment?

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



Conditional probability with Bayes' Theorem

“The probability of A given B”

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

“Prior”

“Evidence”

“Likelihood”

“Posterior”

The diagram illustrates the components of Bayes' Theorem. It features a mathematical equation $P(A | B) = \frac{P(B | A)P(A)}{P(B)}$. The term $P(B | A)P(A)$ is highlighted with a teal oval and labeled "Likelihood". The term $P(A)$ is highlighted with a pink oval and labeled "Prior". The term $P(B)$ is highlighted with an orange oval and labeled "Evidence". Arrows point from the labels "Likelihood", "Prior", and "Evidence" to their respective highlighted terms in the numerator and denominator of the equation.

Conditional probability with Bayes' Theorem

**A patient walks in to your office feeling horrible after clinical trials,
what is the probability they were given the treatment?**

“The probability of A given B”

A: treatment;
B: feeling sick

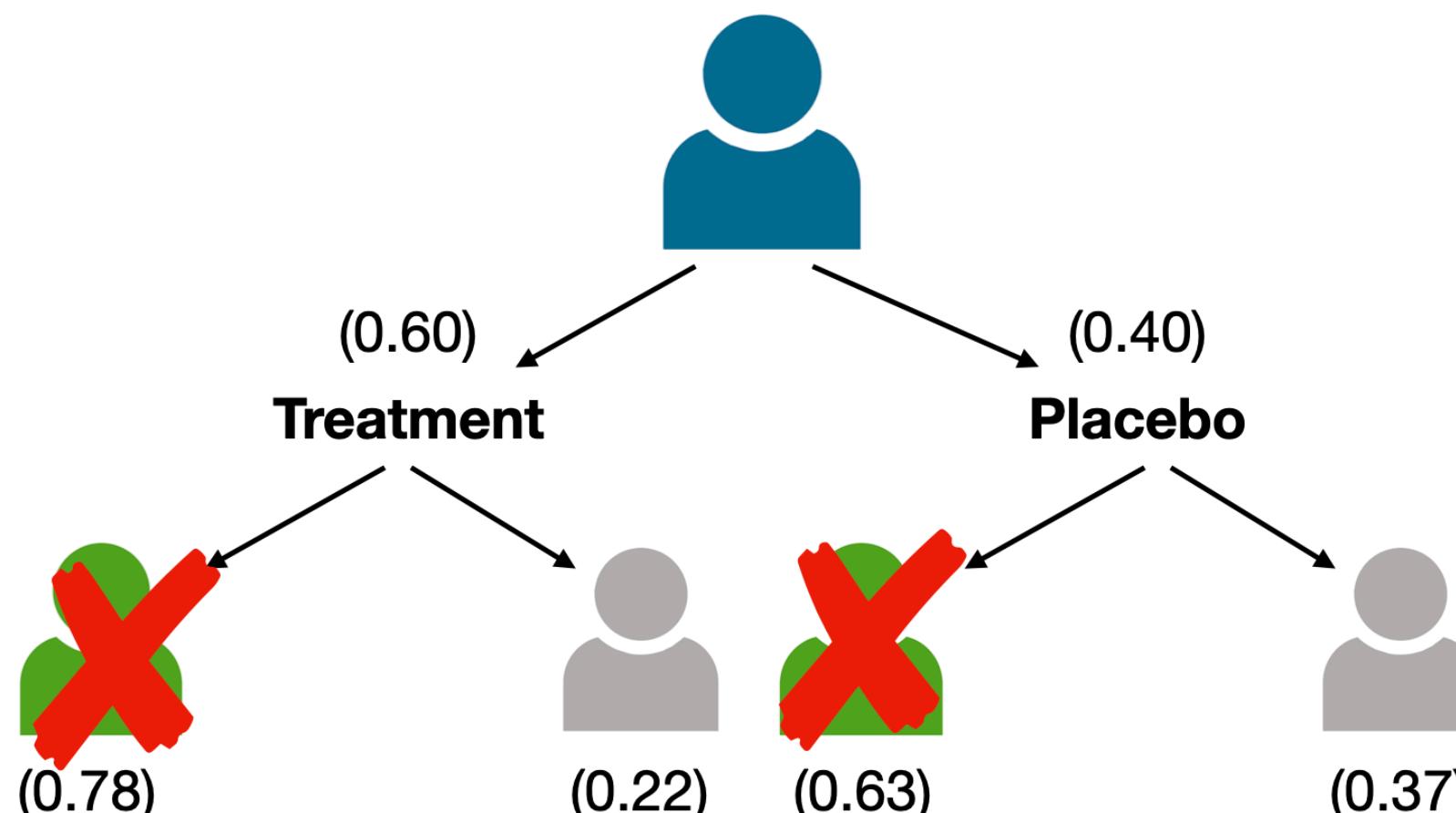
$$P(A | B) =$$

$$\frac{P(B | A)P(A)}{P(B)}$$

0.60
 $(0.60)(0.22) + (0.40)(0.37)$

$$\frac{(0.60)(0.22)}{(0.60)(0.22) + (0.40)(0.37)}$$

0.22



- Scenario: I always get itchy eyes and runny nose when I am around cats. Let's say I decide to get allergy tested:
 - The test is positive 80% of the time for people with allergies
 - The test is positive 10% of the time for people without allergies
- If 1% of the population has allergies and my test came back positive, what is the probability that I have allergies?

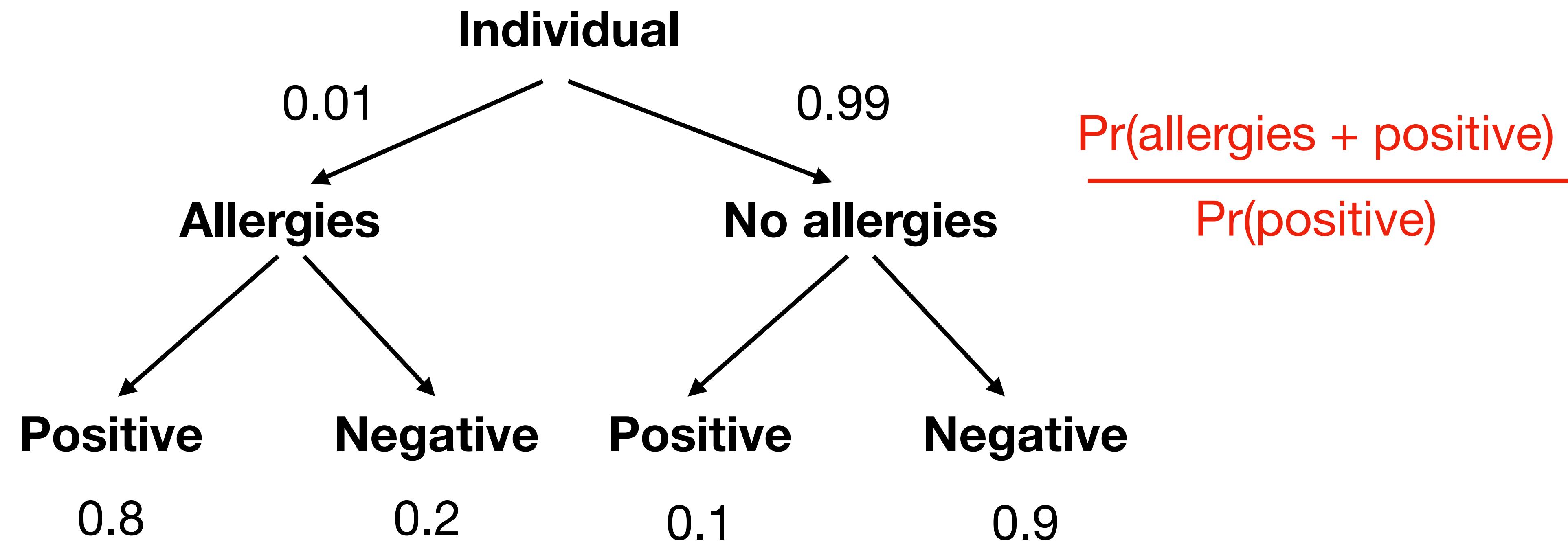
Contingency
Table

Probability
Tree

Bayes'
Theorem

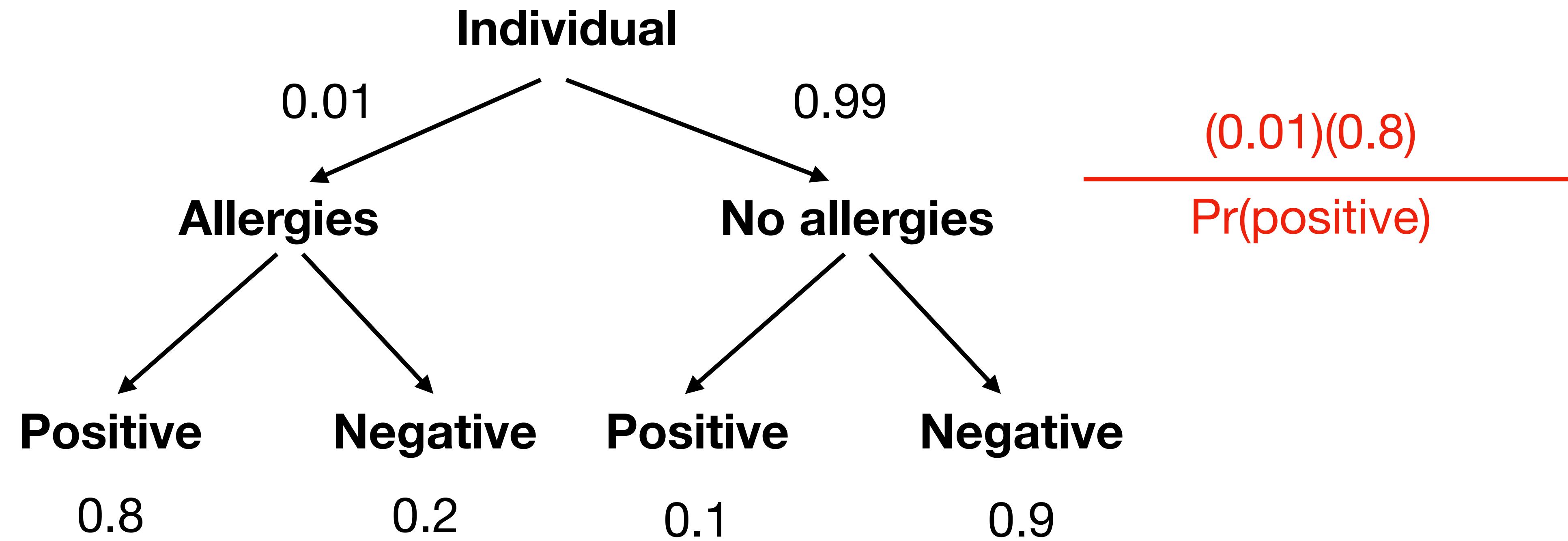
Probability Tree

- Scenario: I always get itchy eyes and runny nose when I am around cats. Let's say I decide to get allergy tested:
 - The test is positive 80% of the time for people with allergies
 - The test is positive 10% of the time for people without allergies
- If 1% of the population has allergies and my test came back positive, what is the probability that I have allergies?



Probability Tree

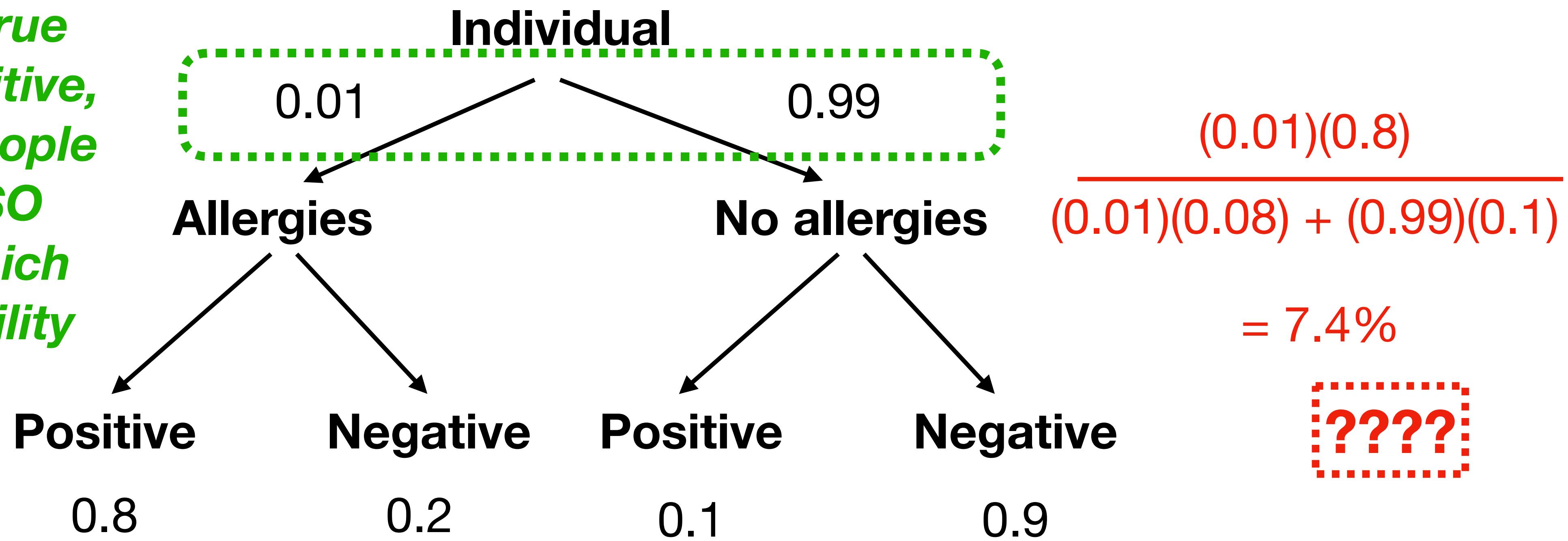
- Scenario: I always get itchy eyes and runny nose when I am around cats. Let's say I decide to get allergy tested:
 - The test is positive 80% of the time for people with allergies
 - The test is positive 10% of the time for people without allergies
- If 1% of the population has allergies and my test came back positive, what is the probability that I have allergies?



Probability Tree

- Scenario: I always get itchy eyes and runny nose when I am around cats. Let's say I decide to get allergy tested:
 - The test is positive 80% of the time for people with allergies
 - The test is positive 10% of the time for people without allergies
 - If 1% of the population has allergies and my test came back positive, what is the probability that I have allergies?

Even though the true positive > false positive, the initial rate of people with allergies is SO MUCH LOWER which affects the probability



- Scenario: I always get itchy eyes and runny nose when I am around cats. Let's say I decide to get allergy tested:
 - The test is positive 80% of the time for people with allergies
 - The test is positive 10% of the time for people without allergies
- If 1% of the population has allergies and my test came back positive, what is the probability that I have allergies?

Contingency
Table

Probability
Tree

Bayes'
Theorem

Bayes' Theorem

- Scenario: I always get itchy eyes and runny nose when I am around cats. Let's say I decide to get allergy tested:
 - The test is positive 80% of the time for people with allergies
 - The test is positive 10% of the time for people without allergies
 - If 1% of the population has allergies and my test came back positive, what is the probability that I have allergies?

A: allergies;

B: positive test

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

0.80 0.01
 $(0.01)(0.80) + (0.99)(0.1)$

- Scenario: I always get itchy eyes and runny nose when I am around cats. Let's say I decide to get allergy tested:
 - The test is positive 80% of the time for people with allergies
 - The test is positive 10% of the time for people without allergies
- If 1% of the population has allergies and my test came back positive, what is the probability that I have allergies?

**Contingency
Table**

**Probability
Tree**



Bayes' Theorem

Contingency
Table

- Scenario: I always get itchy eyes and runny nose when I am around cats. Let's say I decide to get allergy tested:
 - The test is positive 80% of the time for people with allergies
 - The test is positive 10% of the time for people without allergies
- If 1% of the population has allergies and my test came back positive, what is the probability that I have allergies?

	People with allergies	People without allergies	TOTAL
Positive test			
Negative test			
TOTAL			

Contingency
Table

- Scenario: I always get itchy eyes and runny nose when I am around cats. Let's say I decide to get allergy tested:
 - The test is positive 80% of the time for people with allergies
 - The test is positive 10% of the time for people without allergies
- If 1% of the population has allergies and my test came back positive, what is the probability that I have allergies?

	People with allergies	People without allergies	TOTAL
Positive test			
Negative test			
TOTAL	10	990	1000

Contingency
Table

- Scenario: I always get itchy eyes and runny nose when I am around cats. Let's say I decide to get allergy tested:
 - The test is positive 80% of the time for people with allergies
 - The test is positive 10% of the time for people without allergies
 - If 1% of the population has allergies and my test came back positive, what is the probability that I have allergies?

	People with allergies	People without allergies	TOTAL
Positive test	8		
Negative test	2		
TOTAL	10	990	1000

Contingency
Table

- Scenario: I always get itchy eyes and runny nose when I am around cats. Let's say I decide to get allergy tested:
 - The test is positive 80% of the time for people with allergies
 - The test is positive 10% of the time for people without allergies
- If 1% of the population has allergies and my test came back positive, what is the probability that I have allergies?

	People with allergies	People without allergies	TOTAL
Positive test	8	99	107
Negative test	2	891	893
TOTAL	10	990	1000

Contingency
Table

- Scenario: I always get itchy eyes and runny nose when I am around cats. Let's say I decide to get allergy tested:
 - The test is positive 80% of the time for people with allergies
 - The test is positive 10% of the time for people without allergies
- If 1% of the population has allergies and my test came back positive, what is the probability that I have allergies?

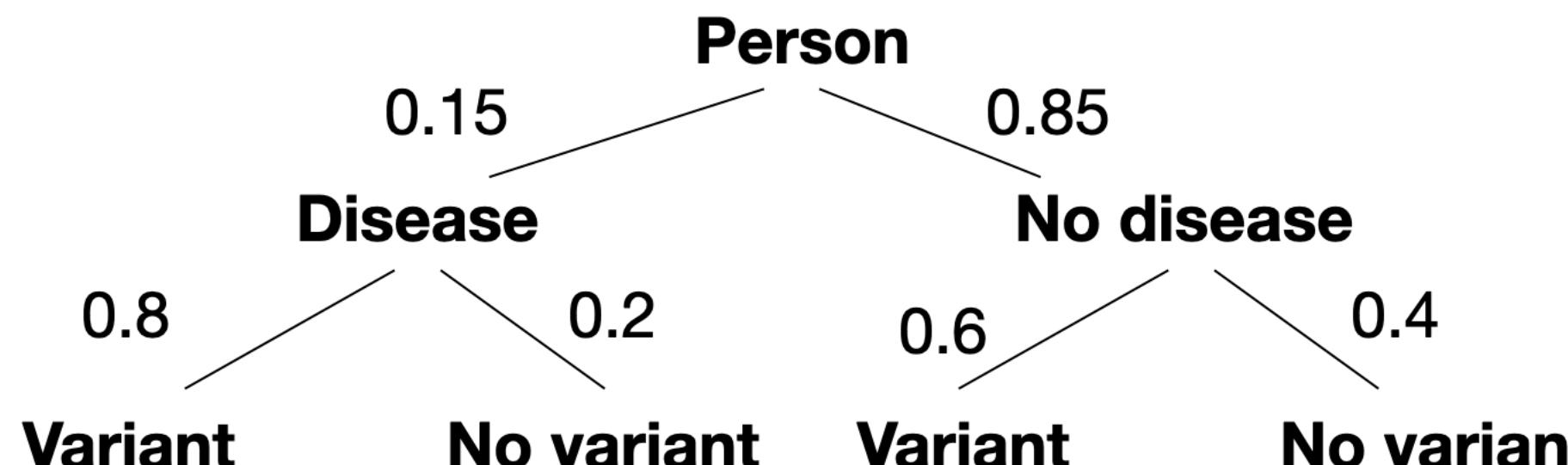
	People with allergies	People without allergies	TOTAL
Positive test	8	99	107
Negative test	2	891	893
TOTAL	10	990	1000


$$8/107 = 7.47\%$$

Suppose you have a sample of 500 individuals. 75 of these individuals have a rare genetic disease and you identified a genetic variant that might be associated with this disease. 80% of individuals with the disease and 60% of individuals without the disease have the genetic variant. **If an individual has the variant, what is the probability that they also have the disease?**

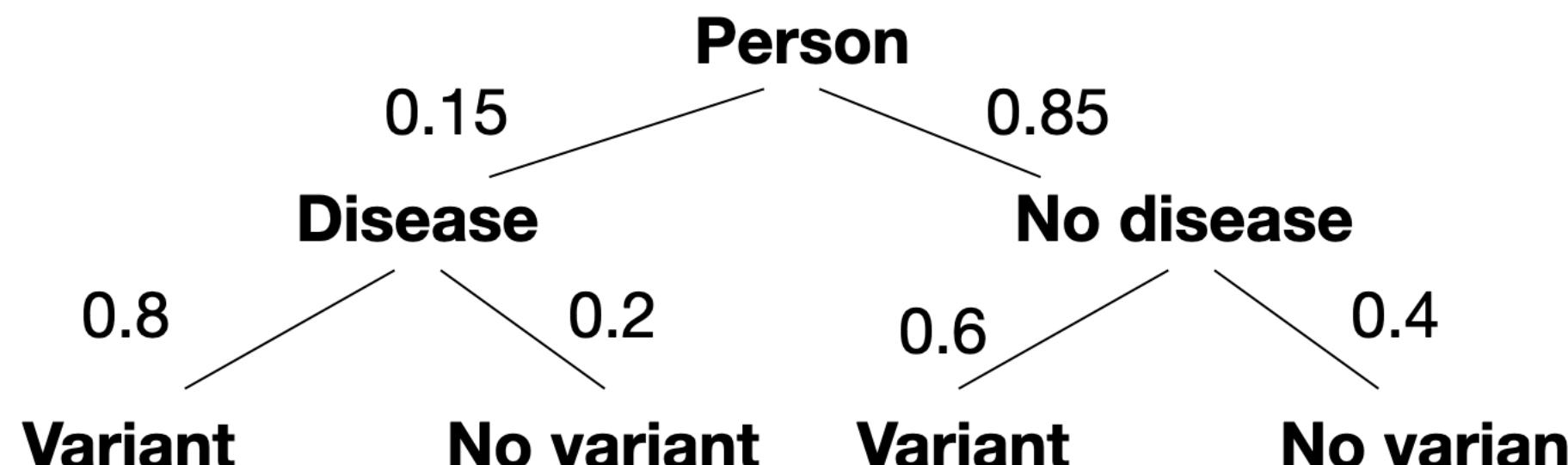
Suppose you have a sample of 500 individuals. 75 of these individuals have a rare genetic disease and you identified a genetic variant that might be associated with this disease. 80% of individuals with the disease and 60% of individuals without the disease have the genetic variant. **If an individual has the variant, what is the probability that they also have the disease?**

$$P(D | V) = \frac{P(V | D)P(D)}{P(V)}$$



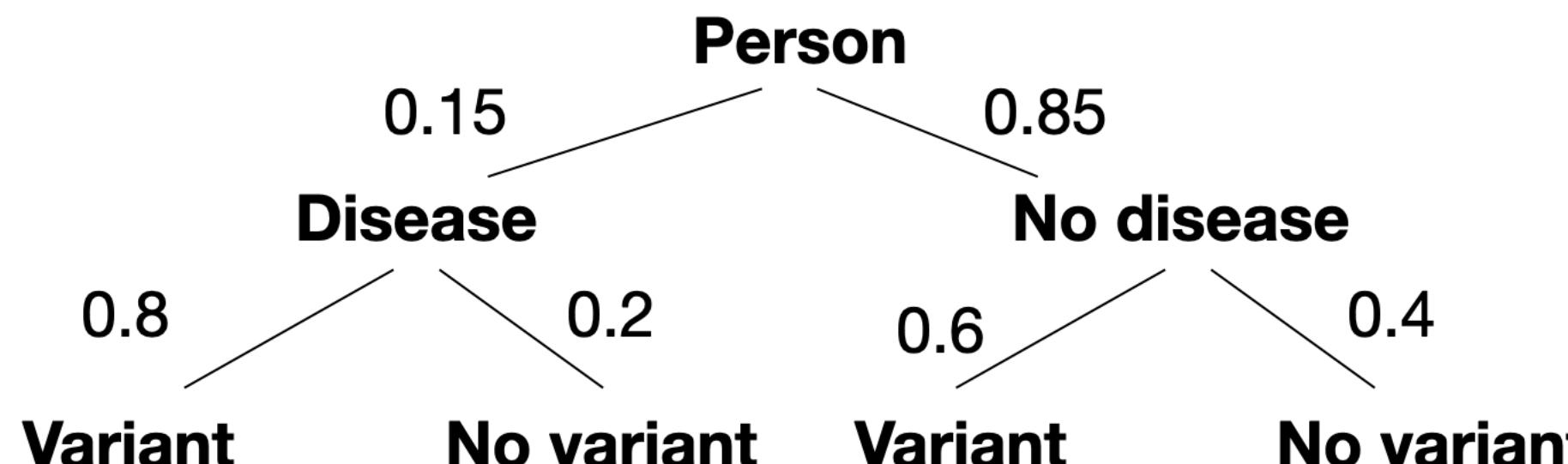
Suppose you have a sample of 500 individuals. 75 of these individuals have a rare genetic disease and you identified a genetic variant that might be associated with this disease. 80% of individuals with the disease and 60% of individuals without the disease have the genetic variant. **If an individual has the variant, what is the probability that they also have the disease?**

$$P(D | V) = \frac{P(V | D)(0.15)}{P(V)}$$



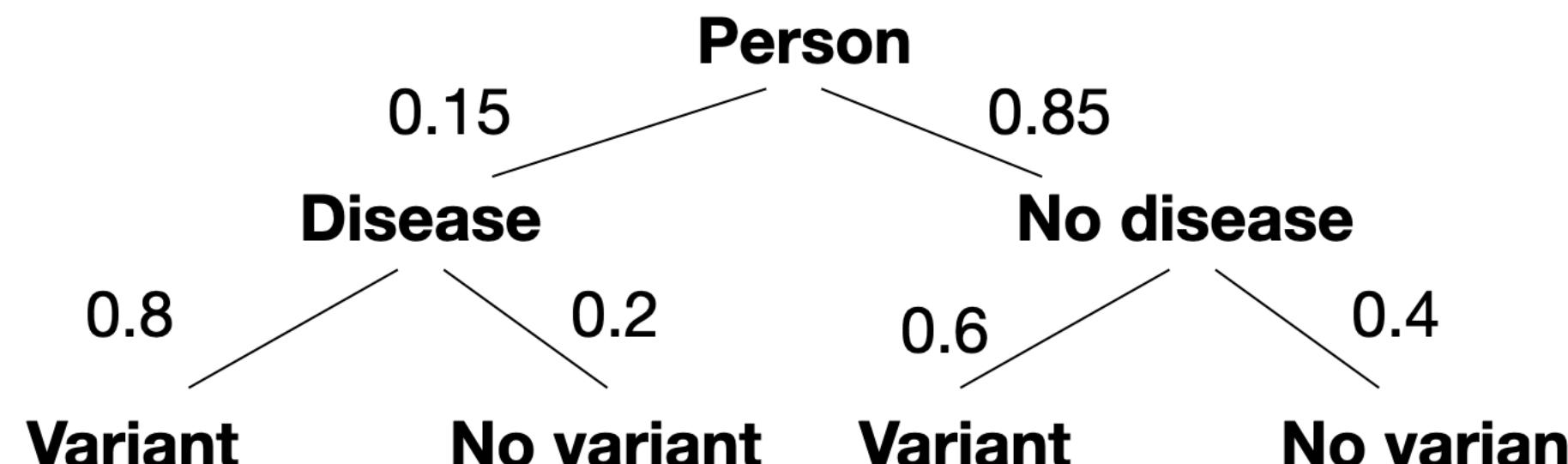
Suppose you have a sample of 500 individuals. 75 of these individuals have a rare genetic disease and you identified a genetic variant that might be associated with this disease. 80% of individuals with the disease and 60% of individuals without the disease have the genetic variant. **If an individual has the variant, what is the probability that they also have the disease?**

$$P(D | V) = \frac{(0.8)(0.15)}{P(V)}$$



Suppose you have a sample of 500 individuals. 75 of these individuals have a rare genetic disease and you identified a genetic variant that might be associated with this disease. 80% of individuals with the disease and 60% of individuals without the disease have the genetic variant. **If an individual has the variant, what is the probability that they also have the disease?**

$$P(D | V) = \frac{(0.8)(0.15)}{(0.8)(0.15) + (0.85)(0.6)} = 0.19$$



Conditional probability with Bayes' Theorem

- Tenant of **Bayes' Theorem** is that you can **modify the existing probably** of an event happening **based on observations**
- Helps address flaw of “**frequentists**” view of statistics that the **result of an experiment depends on the number of times** it is repeated
- Applications of Bayes' theorem:
 - Naive Bayes classifiers (machine learning)
 - Discriminant functions (similar to PCA/SVM grouping)
 - Bayesian parameter estimation (i.e. maximum likelihood estimate)

Bayesian statistical learning for big data biology

Christopher Yau & Kieran Campbell

Biophysical Reviews 11, 95–102 (2019) | Cite this article

5787 Accesses | **9** Citations | **15** Altmetric | [Metrics](#)

Abstract

Bayesian statistical learning provides a coherent probabilistic uncertainty in systems. This review describes the theoretical foundations of Bayesian statistics and outlines the computational frameworks for implementation in practice. We then describe the use of Bayesian learning in single- and multi-task learning with high-dimensional, large data sets.



Physics of Life Reviews

Volume 33, July 2020, Pages 88-108

Review

Morphogenesis as Bayesian inference: A variational approach to pattern formation and control in complex biological systems

Franz Kuchling^a, Karl Friston^b, Georgi Georgiev^c, Michael Levin^{a, d}



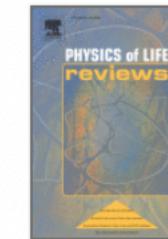
Virus Evolution, 2018, 4(1): vey001

doi: 10.1093/ve/vey016

Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10

Marc A. Suchard,^{1,2,3,*†} Philippe Lemey,^{4,‡} Guy Baele,^{4,§} Daniel
Alexei J. Drummond,^{6,7,*} and Andrew Rambaut^{8,*,**}

¹Department of Biomathematics, David Geffen School of Medicine, University of California Charles E. Young Dr., South, Los Angeles, CA, 90095 USA, ²Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, 650 Charles E. Young Dr., South, Los Angeles, CA, 90095 USA, ³Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, 650 Charles E. Young Dr., South, Los Angeles, CA, 90095 USA, ⁴Department of Microbiology, Immunology, and Molecular Genetics, David Geffen School of Medicine, University of California, Los Angeles, 650 Charles E. Young Dr., South, Los Angeles, CA, 90095 USA, ⁵Center for Bioinformatics and Computational Biology, University of Maryland, College Park, 125 Biomolecular Science Building, 1000 Maryland Hall for Life Sciences, University of Maryland, College Park, MD, 20742 USA.



Article | Open Access | Published: 15 July 2021

Bayesian analysis of home advantage in North American professional sports before and during COVID-19

Nico Higgs & Ian Stavness

Scientific Reports 11, Article number: 14521 (2021) | Cite this article

941 Accesses | 14 Altmetric | Metrics

Abstract

Home advantage in professional sports is a widely accepted phenomenon despite the lack of any controlled experiments at the professional level. The return to play of professional sports during the COVID-19 pandemic presents a unique opportunity to analyze the hypothesized effect of home advantage in neutral settings. While recent work has examined the effect of

Article | Open Access | Published: 20 April 2024

Genomic architecture and prediction of censored time-to-event phenotypes with a Bayesian genome-wide analysis

Sven E. Ojavee , Athanasios Kousathanas, Daniel Trejo Banos, Etienne J. Orliac, Marion Patxot, Kristi Läll, Reedik Mägi, Krista Fischer, Zoltan Kutalik & Matthew R. Robinson 

Nature Communications 12: Article number: 2337 (2021) | Cite this article

2887 Accesses | 69 Altmetric | Metrics

Abstract

While recent advancements in computation and modelling have improved the analysis of complex traits, our understanding of the genetic basis of the time at symptom onset remains limited. Here, we develop a Bayesian approach (BayesW) that provides probabilistic inference of the genetic architecture of age-at-onset phenotypes in a sampling scheme that

ensive simulation work the

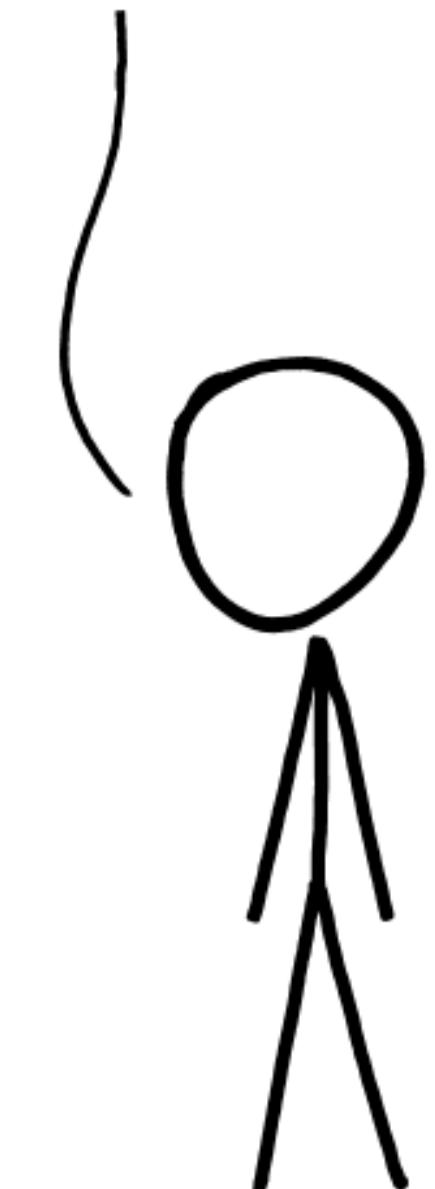
DID THE SUN JUST EXPLODE? (IT'S NIGHT, SO WE'RE NOT SURE.)

THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.

THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.

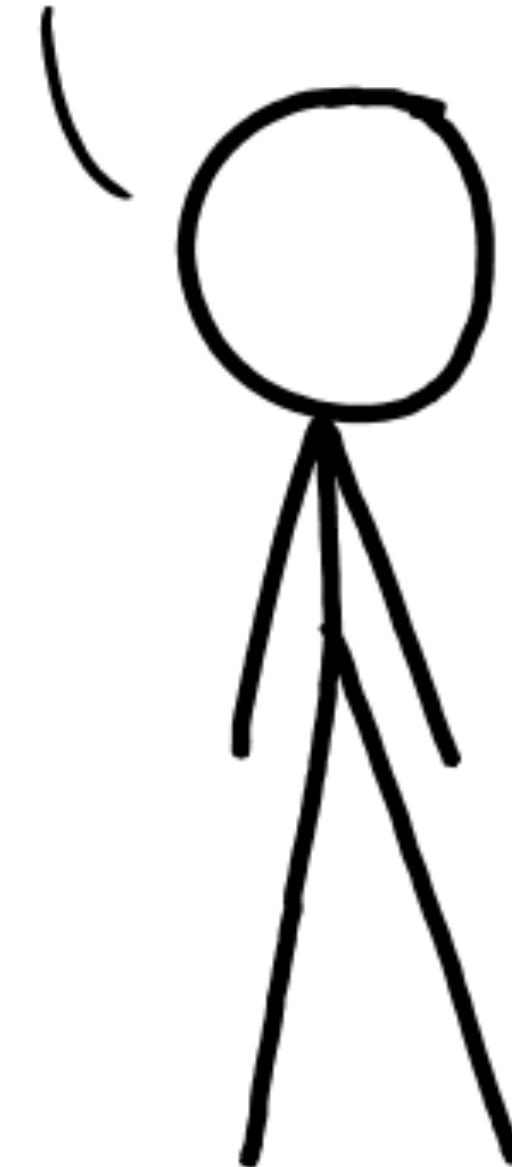
LET'S TRY.

DETECTOR! HAS THE SUN GONE NOVA?



FREQUENTIST STATISTICIAN:

THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS $\frac{1}{36} = 0.027$. SINCE $p < 0.05$, I CONCLUDE THAT THE SUN HAS EXPLODED.



BAYESIAN STATISTICIAN:

BET YOU \$50
IT HASN'T.



An introduction to random variables

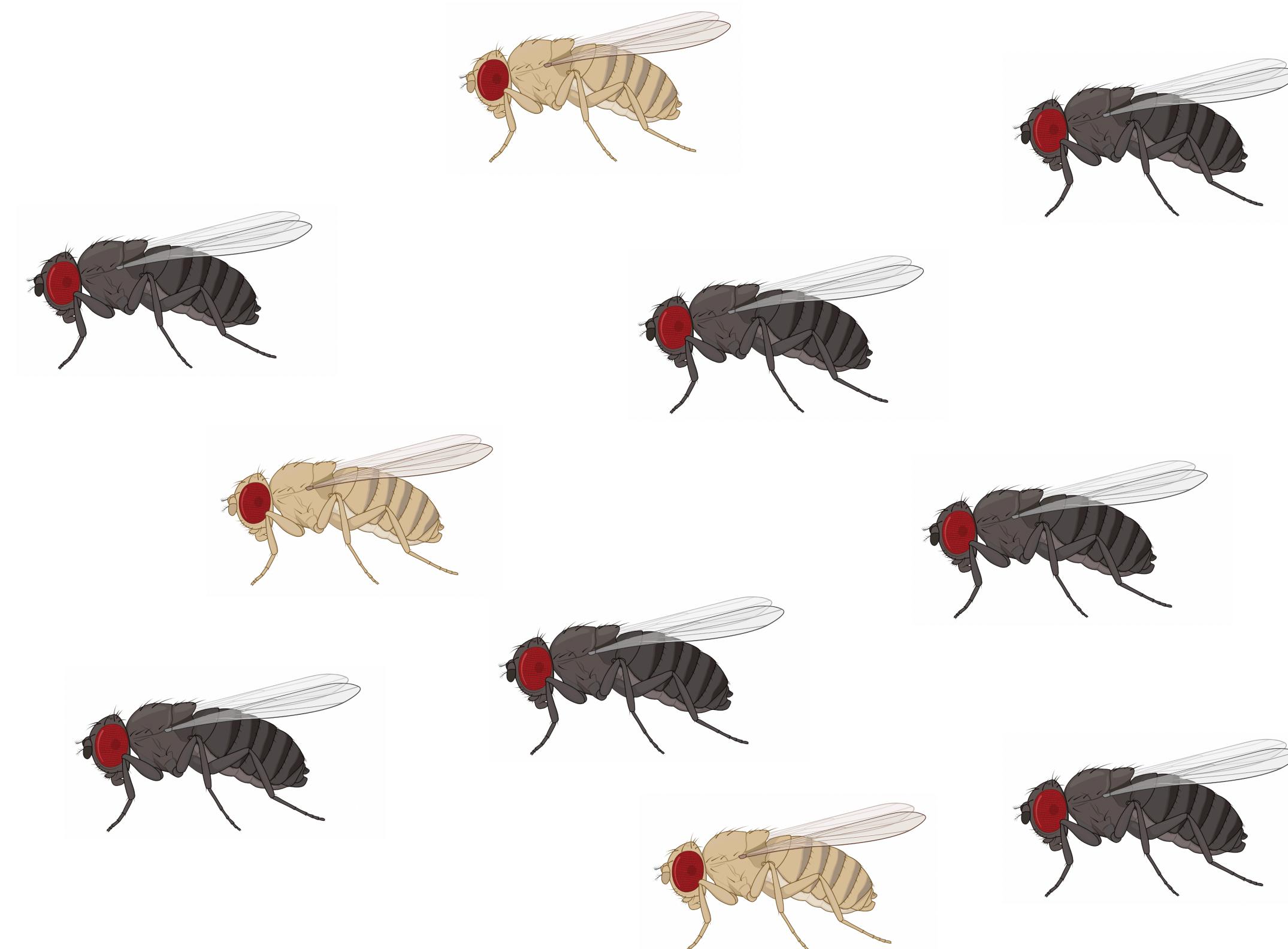
Random variable: variable that takes on a numerical value that depends on the outcome of a chance operation

$\Pr\{\text{Selecting a black fly}\} = ?$

$\Pr\{Y = \text{black}\} = ?$



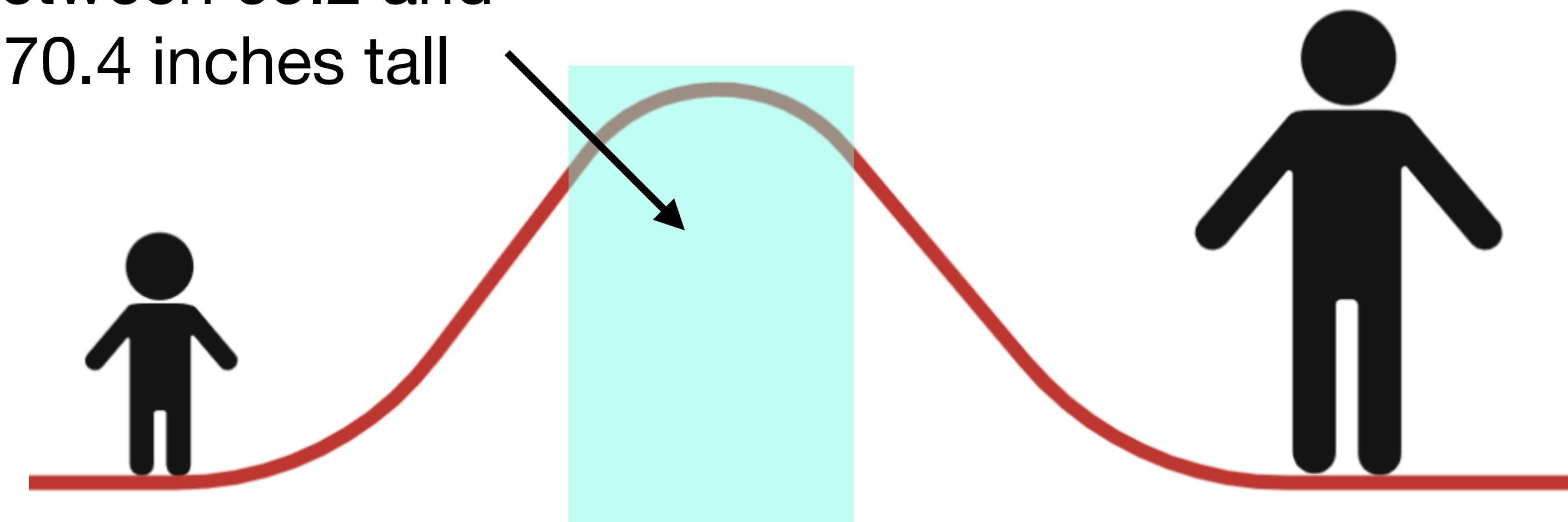
Random variable, Y
 $Y = \{\text{black, white}\}$



An introduction to random variables

Random variable, Y = height of individual chosen at random from population

46% of men are
between 65.2 and
70.4 inches tall



$$Pr[65.2 \leq Y \leq 70.4] = 0.46$$

Mean of a discrete random variable

$$\mu_Y = \sum y_i Pr(Y = y_i)$$

(Sum of all: possible value for Y * probability of that value for Y)

This is also called the **expected value, E(Y)**

Mean of a discrete random variable

$$E(Y) = \mu_Y = \sum y_i Pr(Y = y_i)$$

(Sum of all: possible value for Y * probability of that value for Y)

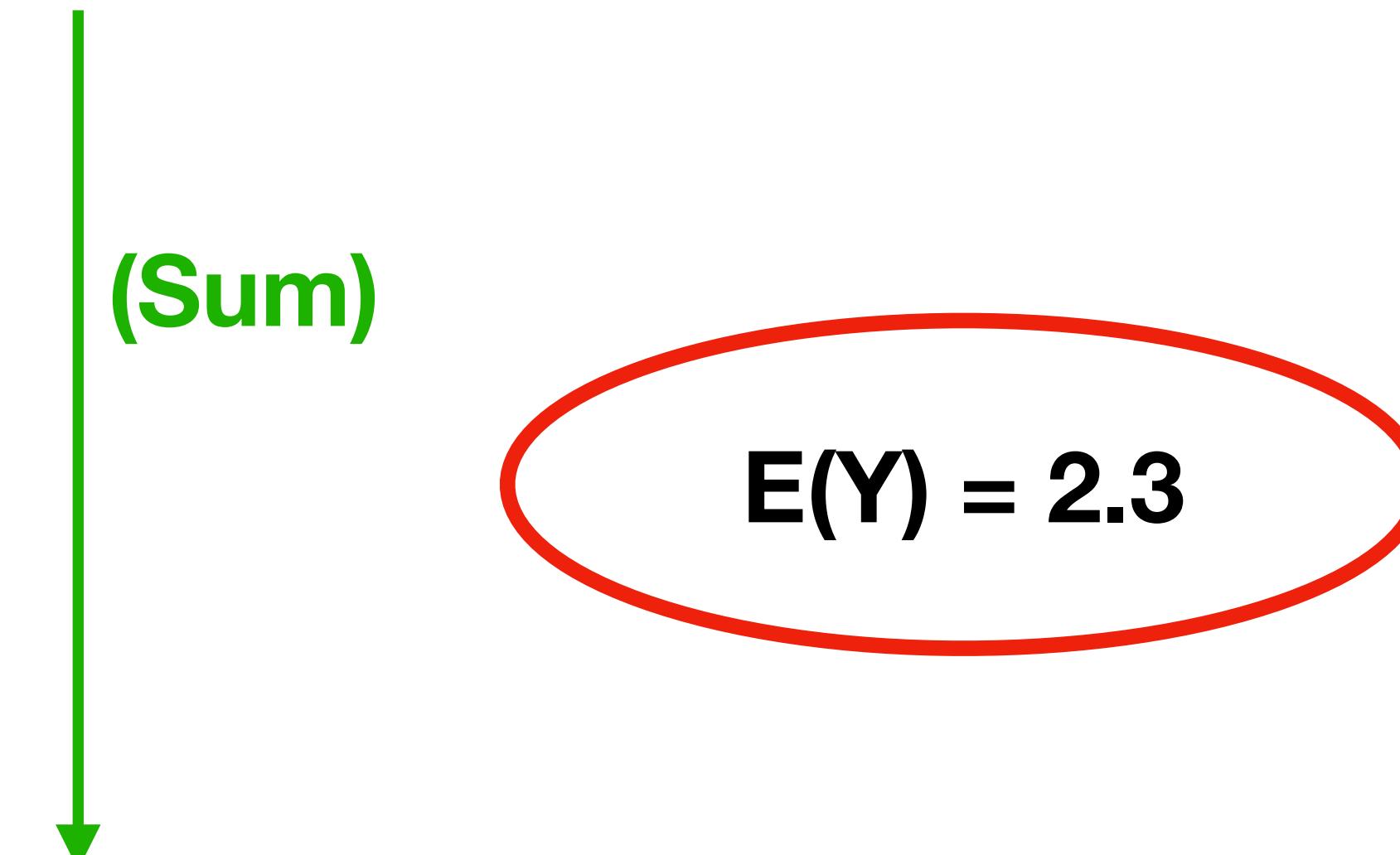
Number of markers	Percent of cells
1	40
2	28
3	19
5	9
8	4

Mean of a discrete random variable

$$E(Y) = \mu_Y = \sum y_i Pr(Y = y_i)$$

(Sum of all: possible value for Y * probability of that value for Y)

Number of markers	Percent of cells	(Y)(Pr(Y))
1	40	(1)(40/100)
2	28	(2)(28/100)
3	19	(3)(19/100)
5	9	(5)(9/100)
8	4	(8)(4/100)

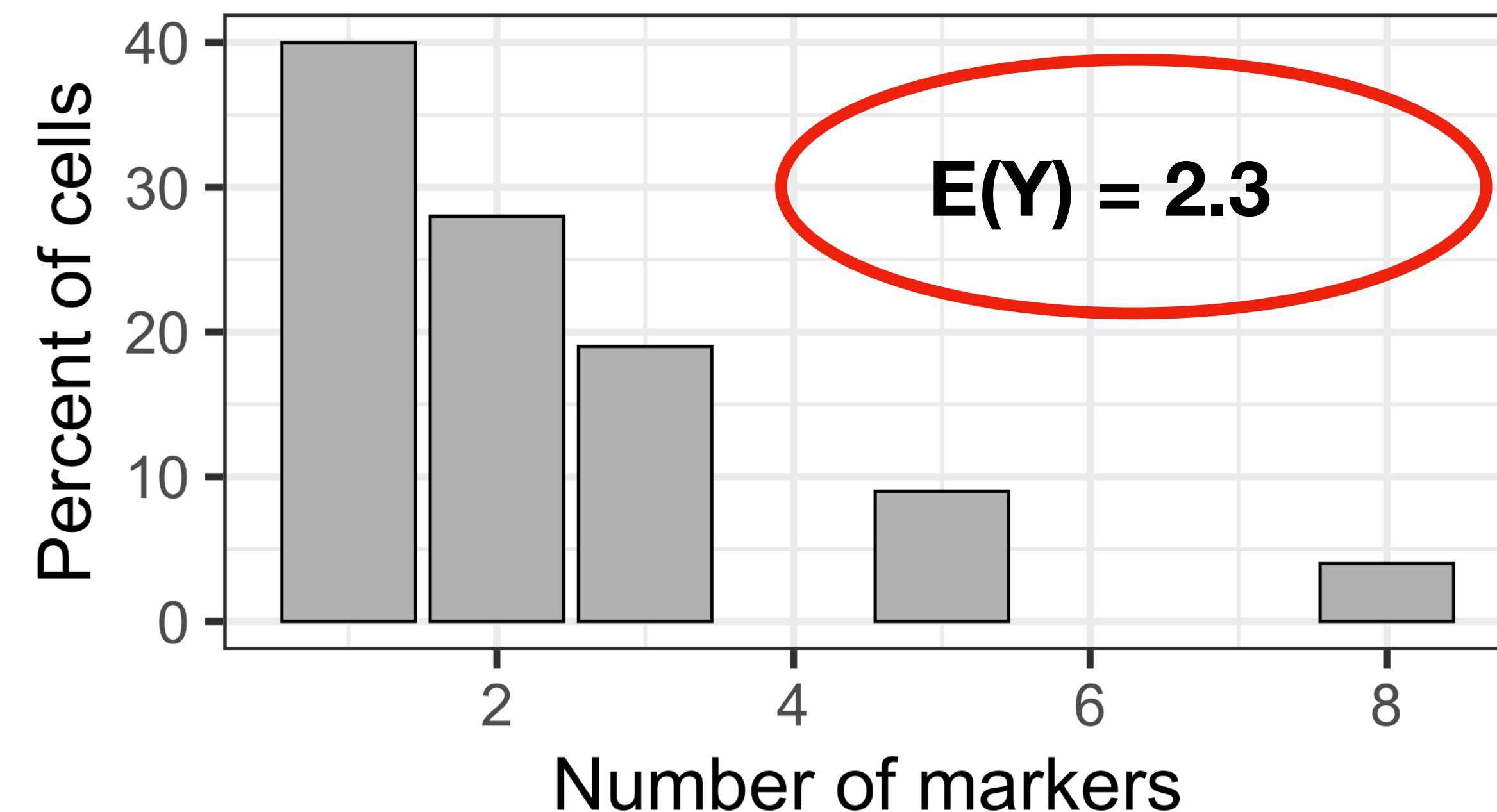


Mean of a discrete random variable

$$E(Y) = \mu_Y = \sum y_i Pr(Y = y_i)$$

(Sum of all: possible value for Y * probability of that value for Y)

Number of markers	Percent of cells
1	40
2	28
3	19
5	9
8	4

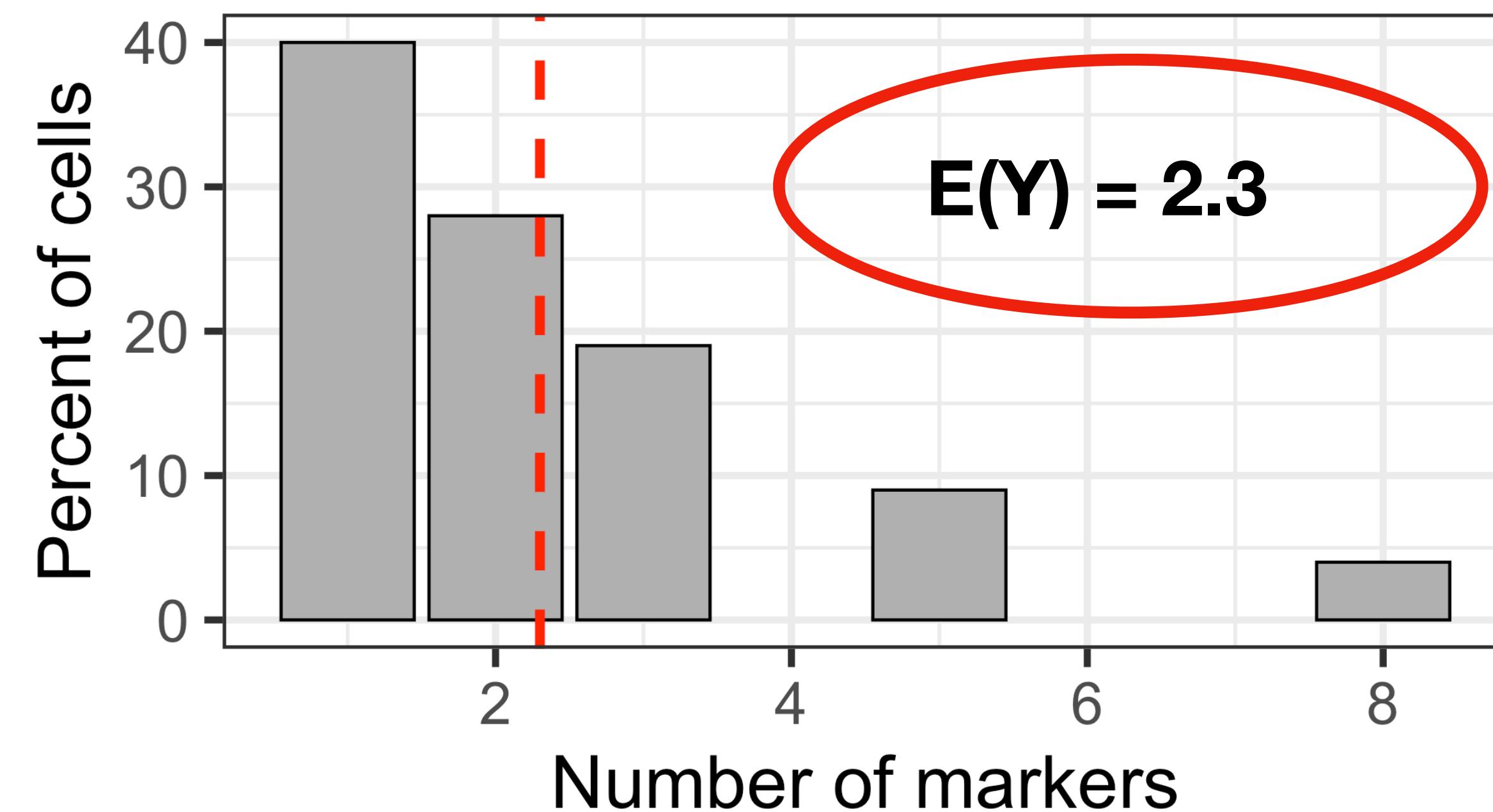


Mean of a discrete random variable

$$E(Y) = \mu_Y = \sum y_i Pr(Y = y_i)$$

(Sum of all: possible value for Y * probability of that value for Y)

Number of markers	Percent of cells
1	40
2	28
3	19
5	9
8	4



Mean of a discrete random variable

This is a theoretical mean value

i.e. if we were to create a sample with this distribution, the mean of that sample would be the expected value

Number of markers	Percent of cells	Total (sample)
1	40	400
2	28	280
3	19	190
5	9	90
8	4	49
		Total: 1000

Sample = 1, 1, 1, ..., 2, 2, ..., 3, ..., 5, ..., 8

mean(sample) = 2.3



E(Y) = 2.3

Variance of a discrete random variable

$$\text{(Variance)} \quad \sigma_Y^2 = \sum (y_i - \mu_Y)^2 Pr(Y = y_i)$$

(Sum of all: squared deviations for each value of Y * probability of that value for Y)

$$\text{(Standard deviation)} \quad \sigma = \sqrt(\sigma_Y^2)$$

Variance of a discrete random variable

(Variance) $\sigma_Y^2 = \sum (y_i - \mu_Y)^2 Pr(Y = y_i)$ $E(Y) = 2.3$

(Sum of all: squared deviations for each value of Y * probability of that value for Y)

Number of markers	Percent of cells	
1	40	$(1 - 2.3)^2(0.40)$
2	28	$(2 - 2.3)^2(0.28)$
3	19	$(3 - 2.3)^2(0.19)$
5	9	$(5 - 2.3)^2(0.09)$
8	4	$(8 - 2.3)^2(0.04)$

(Sum)

VAR(Y) = 2.75
SD(Y) = 1.65

Sample = 1, 1, 1, ..., 2, 2, ..., 3, ..., 5, ...

sd(sample) = 1.65



Summary: discrete random variables

$$E(Y) = \mu_Y = \sum y_i Pr(Y = y_i)$$

(Sum of all: possible value for Y * probability of that value for Y)

(Variance) $\sigma_Y^2 = \sum (y_i - \mu_Y)^2 Pr(Y = y_i)$

(Sum of all: squared deviations for each value of Y * probability of that value for Y)

(Similar definitions for continuous random variables, but involve integral calculus...)

Consider a population of *Drosophila* in which 30% of the individuals are black because of a mutation and 70% have the normal gray body color. Suppose three flies are chosen at random from the population, let Y denote the number of black flies out of the three.

Find $Pr[Y \geq 2]$

Y (No. Black)	Probability
0	0.343
1	0.441
2	0.189
3	0.027

$$0.189 + 0.027 = 0.216$$

Consider a population of *Drosophila* in which 30% of the individuals are black because of a mutation and 70% have the normal gray body color. Suppose three flies are chosen at random from the population, let Y denote the number of black flies out of the three.

Calculate the mean and standard deviation of the random variable

Y (No. Black)	Probability
0	0.343
1	0.441
2	0.189
3	0.027

$$\begin{aligned} & 0(0.343) \\ & + 1(0.441) \\ & + 2(0.189) \\ & + 3(0.027) \end{aligned}$$

$$E(Y) = 0.9$$

Consider a population of *Drosophila* in which 30% of the individuals are black because of a mutation and 70% have the normal gray body color. Suppose three flies are chosen at random from the population, let Y denote the number of black flies out of the three.

Calculate the mean and standard deviation of the random variable

Y (No. Black)	Probability
0	0.343
1	0.441
2	0.189
3	0.027

$$(0-0.9)(0)$$

$$\text{VAR}(Y) = 0.63$$

$$+ (1-0.9)*1(0.441)$$

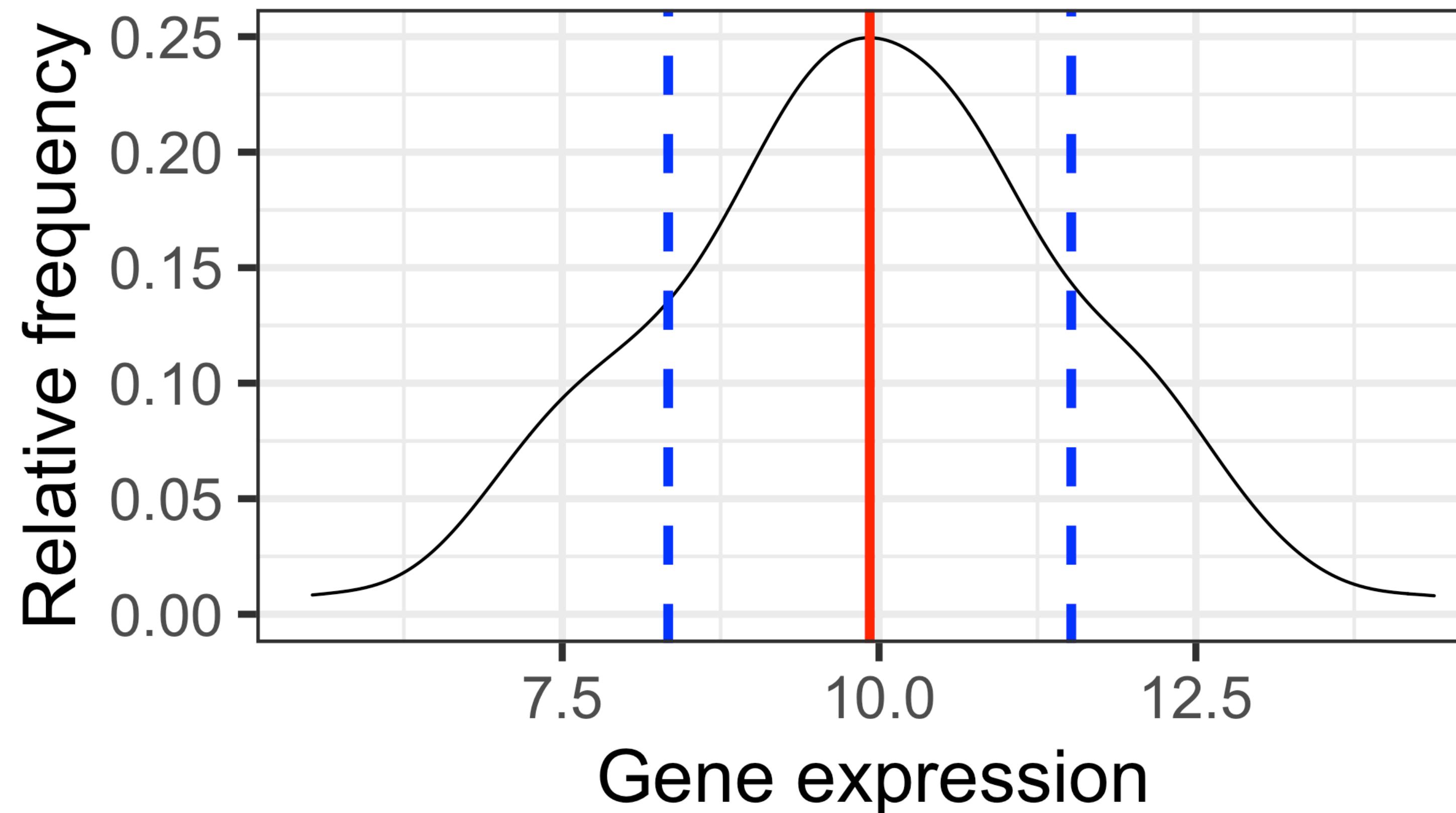
$$\text{SD}(Y) = 0.79$$

$$+ (2-0.9)*2(0.189)$$

$$+ (3-0.9)*3(0.027)$$

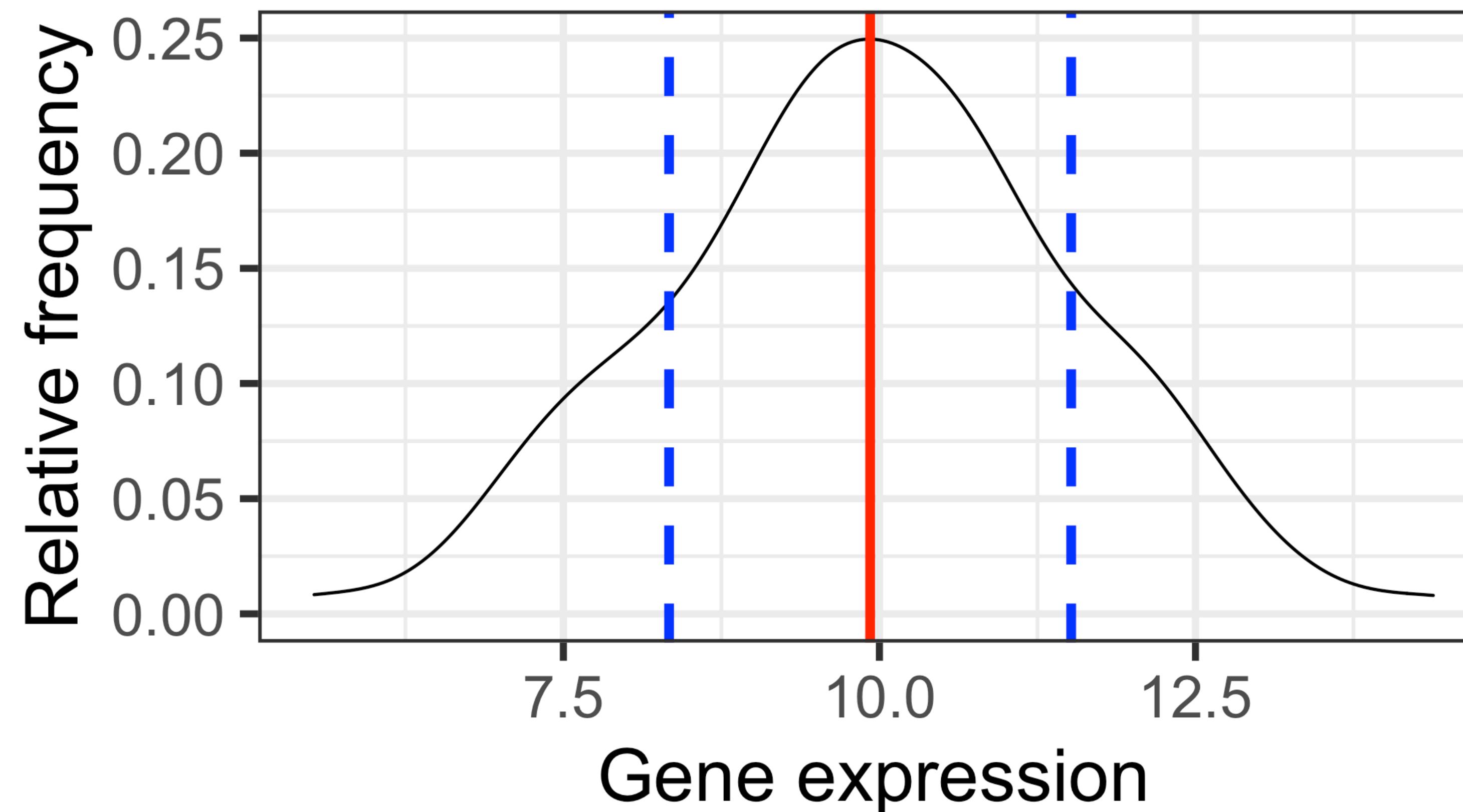
Adding and subtracting random variables

Let's say you performed single-cell RNA-seq within a specific population and found the distribution of your gene of interest to have a mean of 10 units with a standard deviation of 1.5 units:



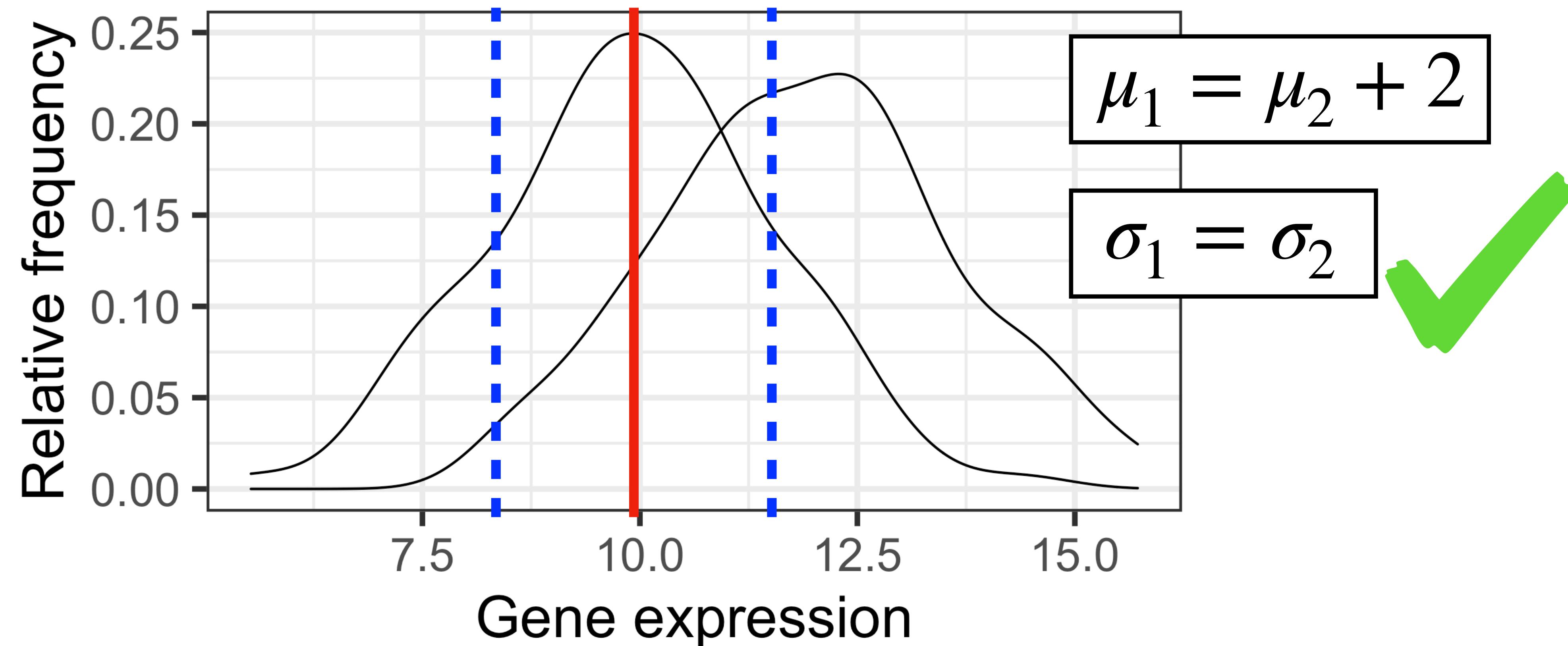
Adding and subtracting random variables

Let's say you performed single-cell RNA-seq within a specific population and found the distribution of your gene of interest to have a mean of 10 units with a standard deviation of 1.5 units. **If you know your second gene of interest is expressed, on average, two units more than the first, what should its distribution look like?**



Adding and subtracting random variables

Let's say you performed single-cell RNA-seq within a specific population and found the distribution of your gene of interest to have a mean of 10 units with a standard deviation of 1.5 units. If you know your second gene of interest is expressed, on average, two units more than the first, what should its distribution look like?



Adding and subtracting random variables

- (1) If we **add** two random variables that are *independent* of one another, then we **add their variances**
- (2) If we **subtract** two random variables that are *independent* of one another, then we **add their variances**
- (3) **standard deviation** = square root of **variance**

Adding and subtracting random variables



$$\mu_1 = 300; \sigma_1 = 22$$

$$\mu_2 = 368; \sigma_2 = 26$$

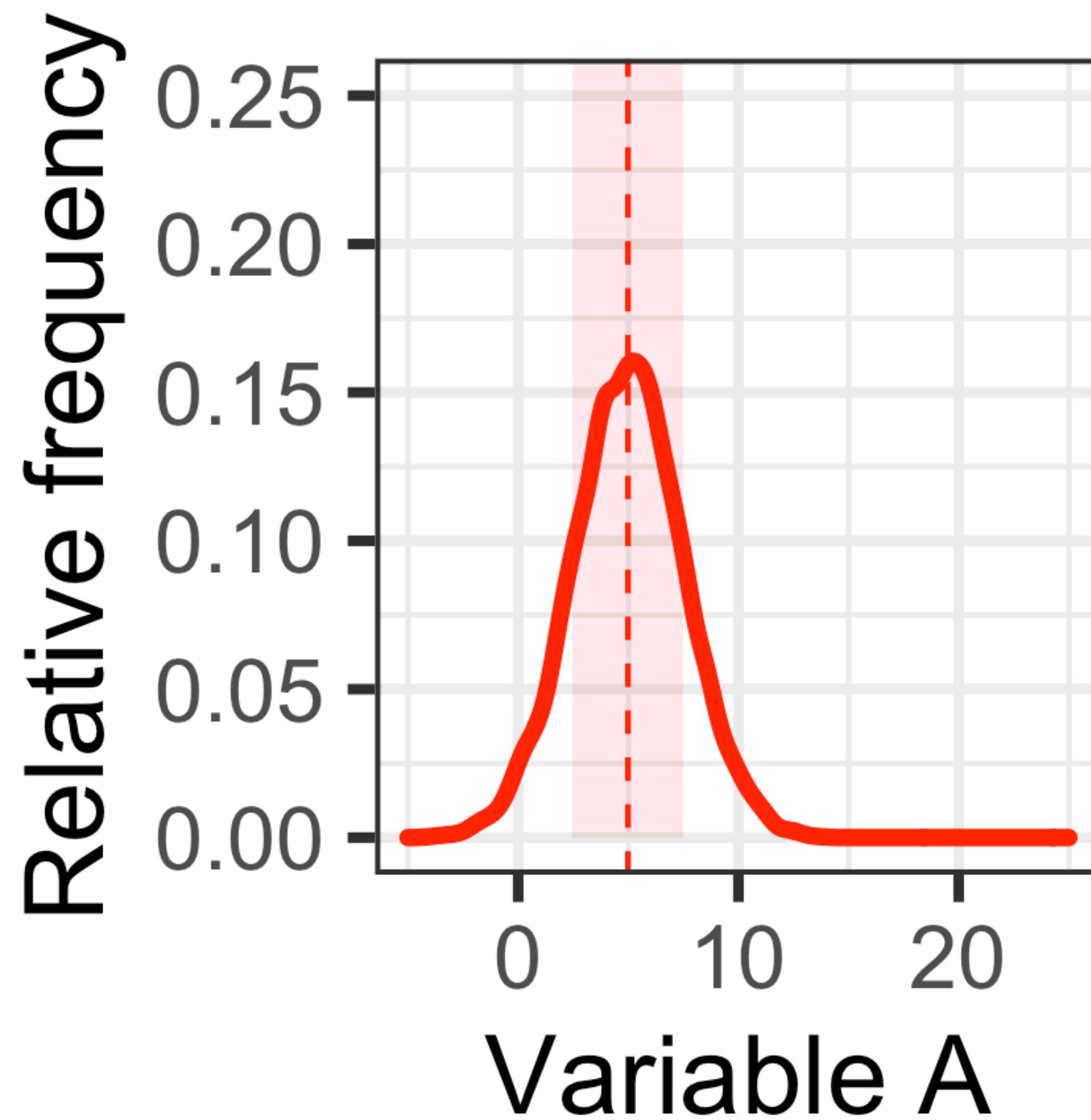
What is the overall difference in means between the populations (black and white)?

What is the variance in difference between these two populations?

$$\begin{aligned}\mu_{1-2} &= \mu_1 - \mu_2 \\ &= -68\end{aligned}$$

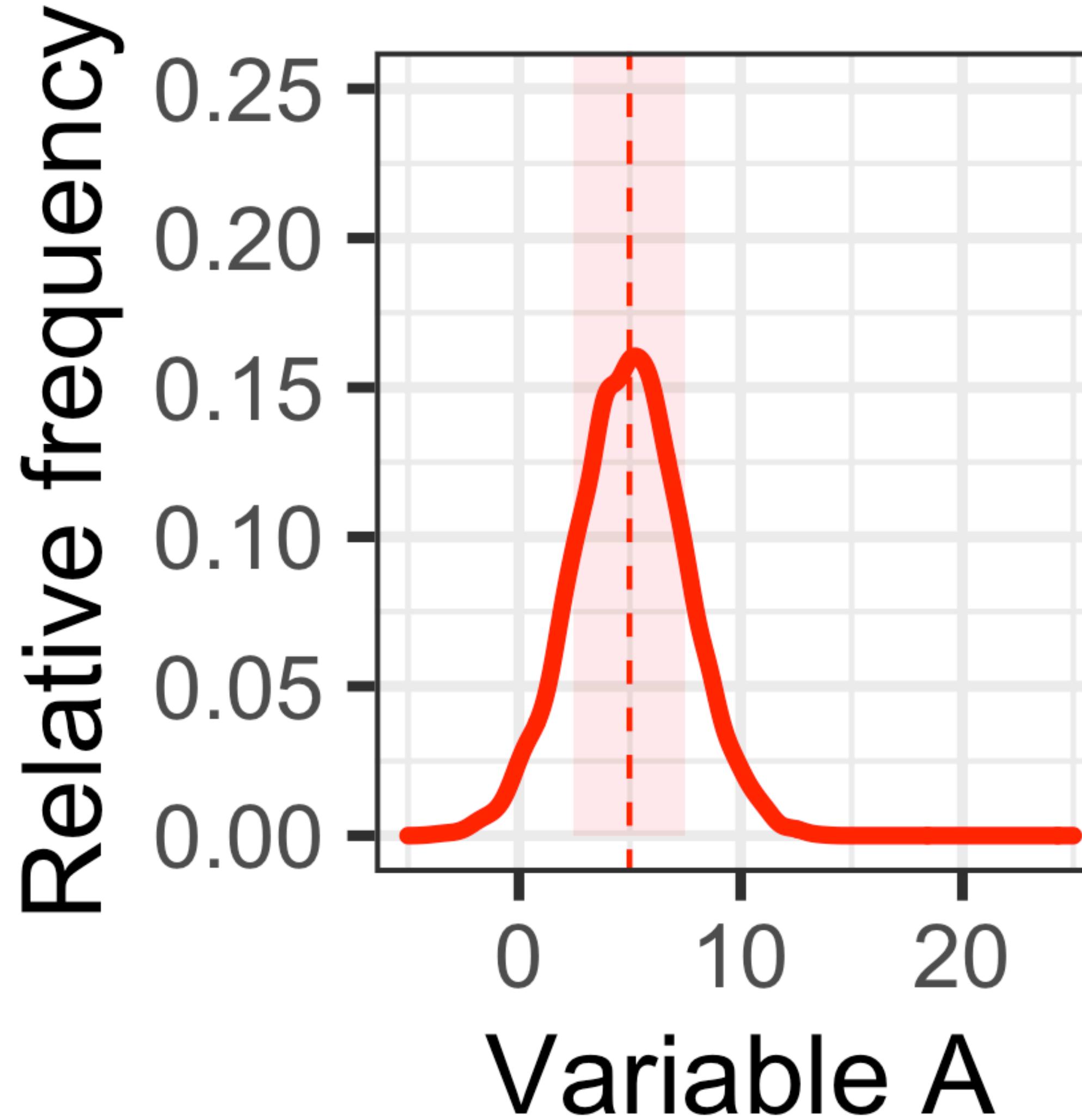
$$\begin{aligned}\sigma_{1-2}^2 &= \sigma_1^2 + \sigma_2^2 \\ \sigma_{1-2}^2 &= 22^2 + 26^2 \\ &= 1160 \\ \sigma &= 34.06\end{aligned}$$

Var A = {11, 5, 4, 3, 5}

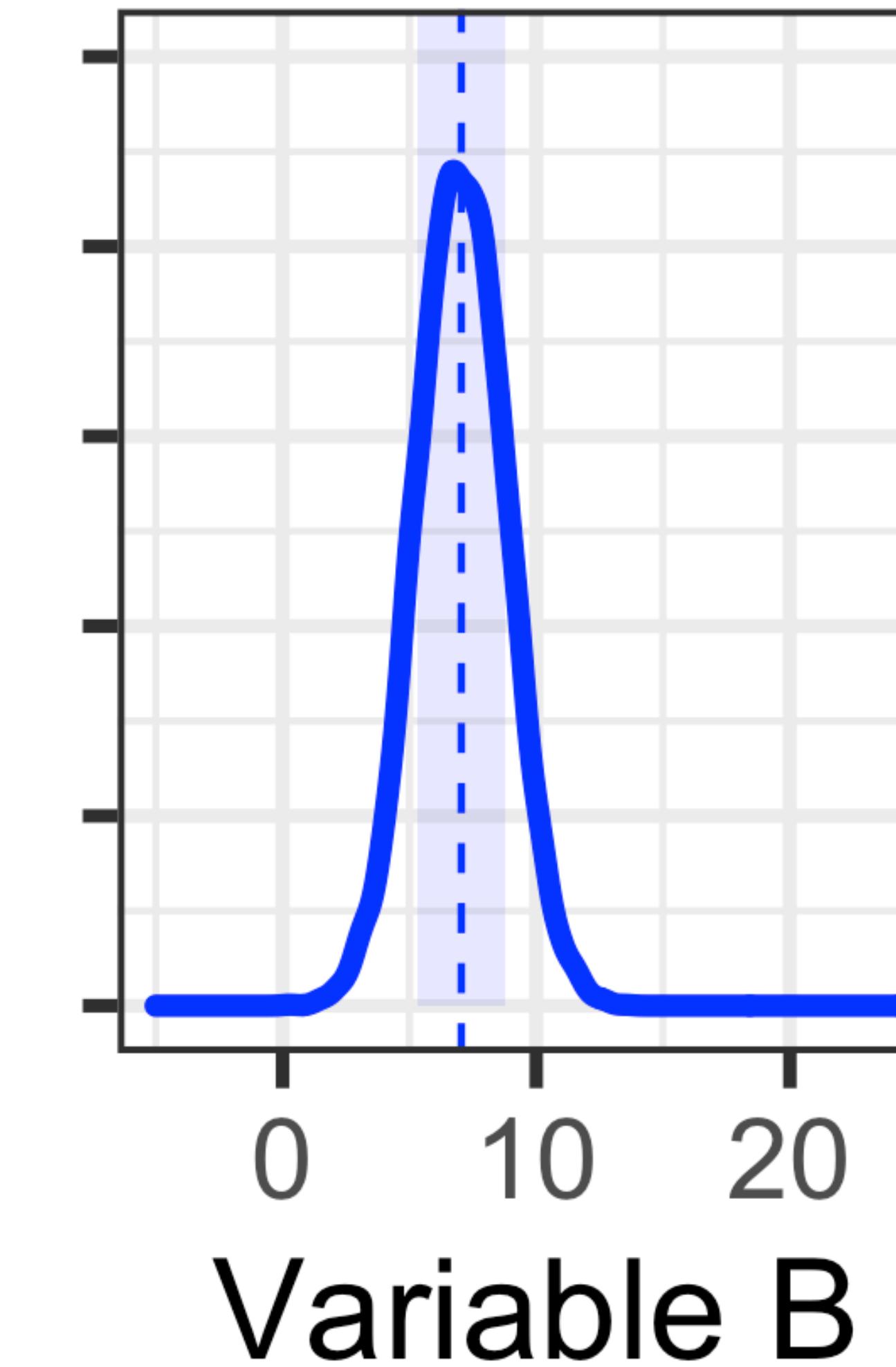


Relative frequency

Var A = {11, 5, 4, 3, 5}



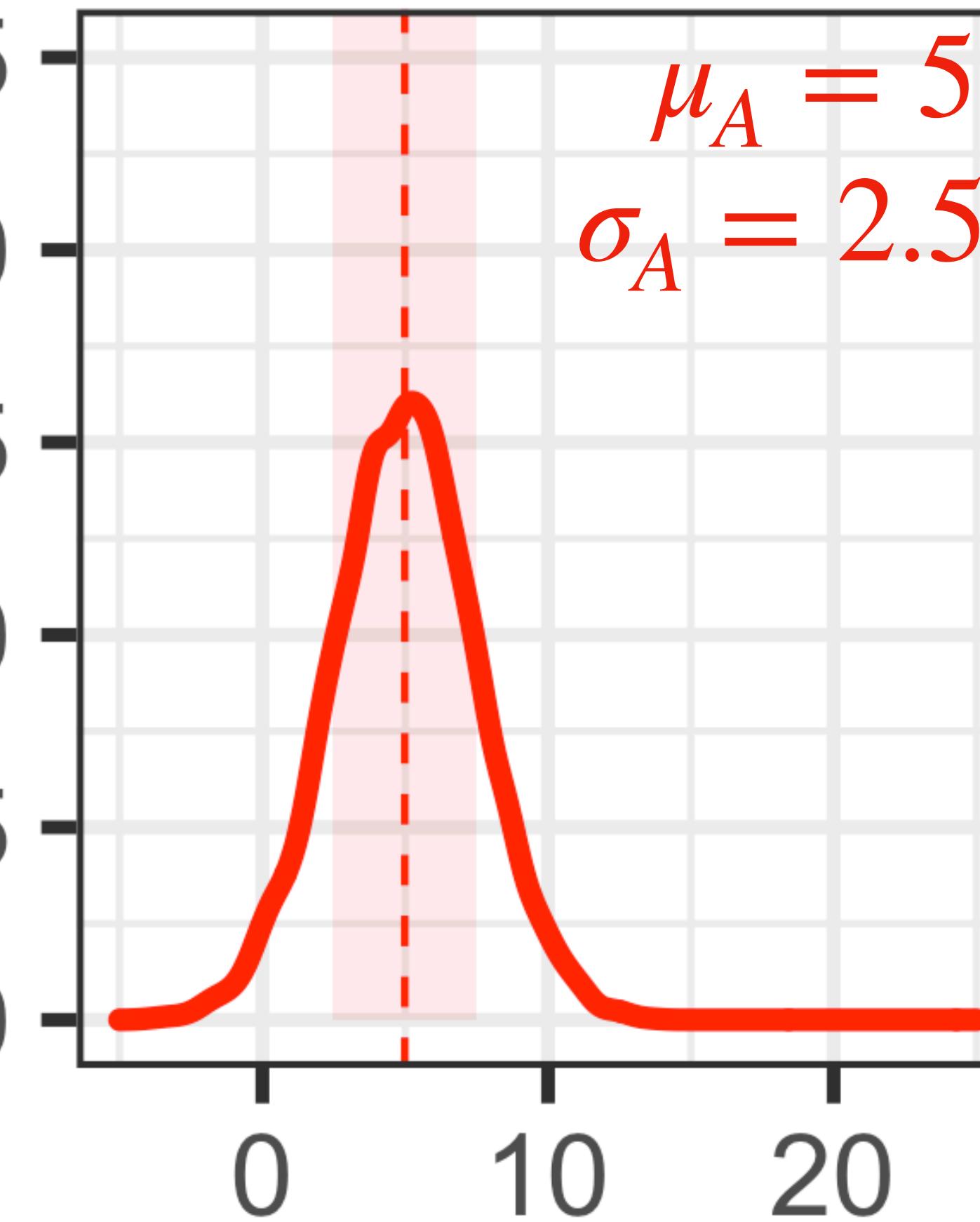
Var B = {6.5, 6.6, 6.8, 6.5, 6.1}



Relative frequency

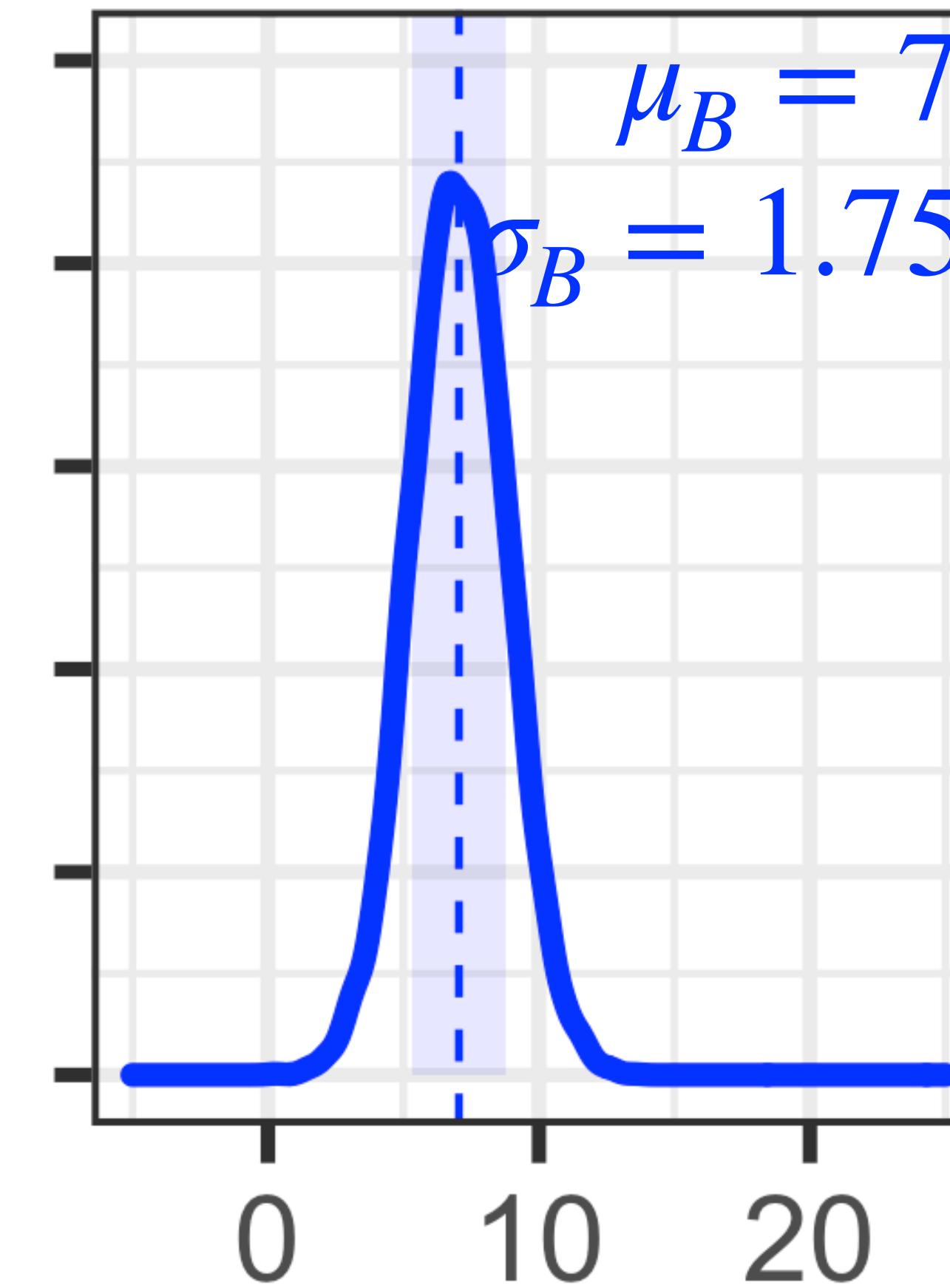
Var A = {11, 5, 4, 3, 5}

$$\begin{aligned}\mu_A &= 5 \\ \sigma_A &= 2.5\end{aligned}$$



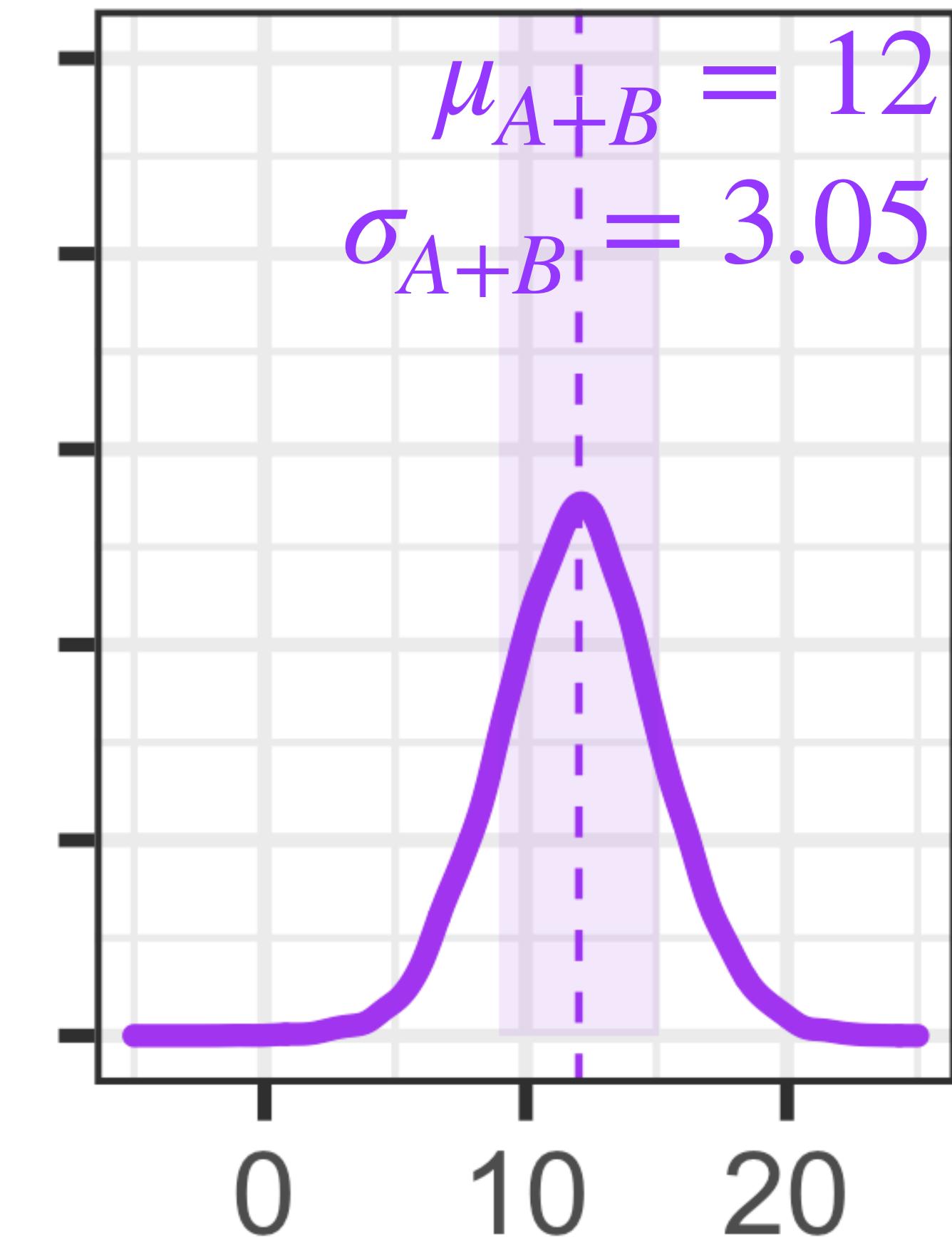
Var B = {6.5, 6.6, 6.8, 6.5, 6.1}

$$\begin{aligned}\mu_B &= 7 \\ \sigma_B &= 1.75\end{aligned}$$



Var (A+B) = {6.5+11, 6.6+5, 6.8+4, 6.5+3, 6.1+5}

$$\begin{aligned}\mu_{A+B} &= 12 \\ \sigma_{A+B} &= 3.05\end{aligned}$$

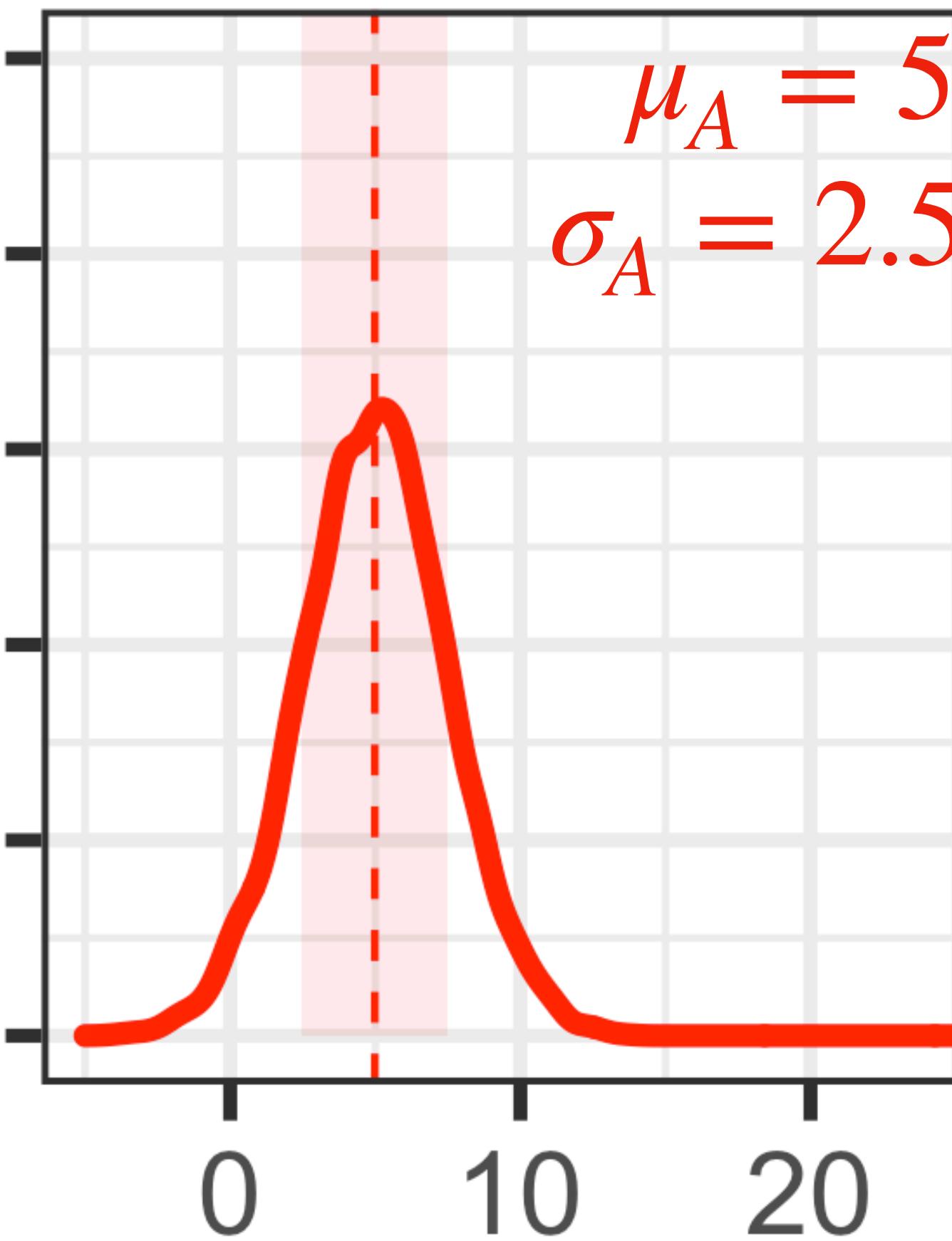


If you want to see the math: link [here](#)

Relative frequency

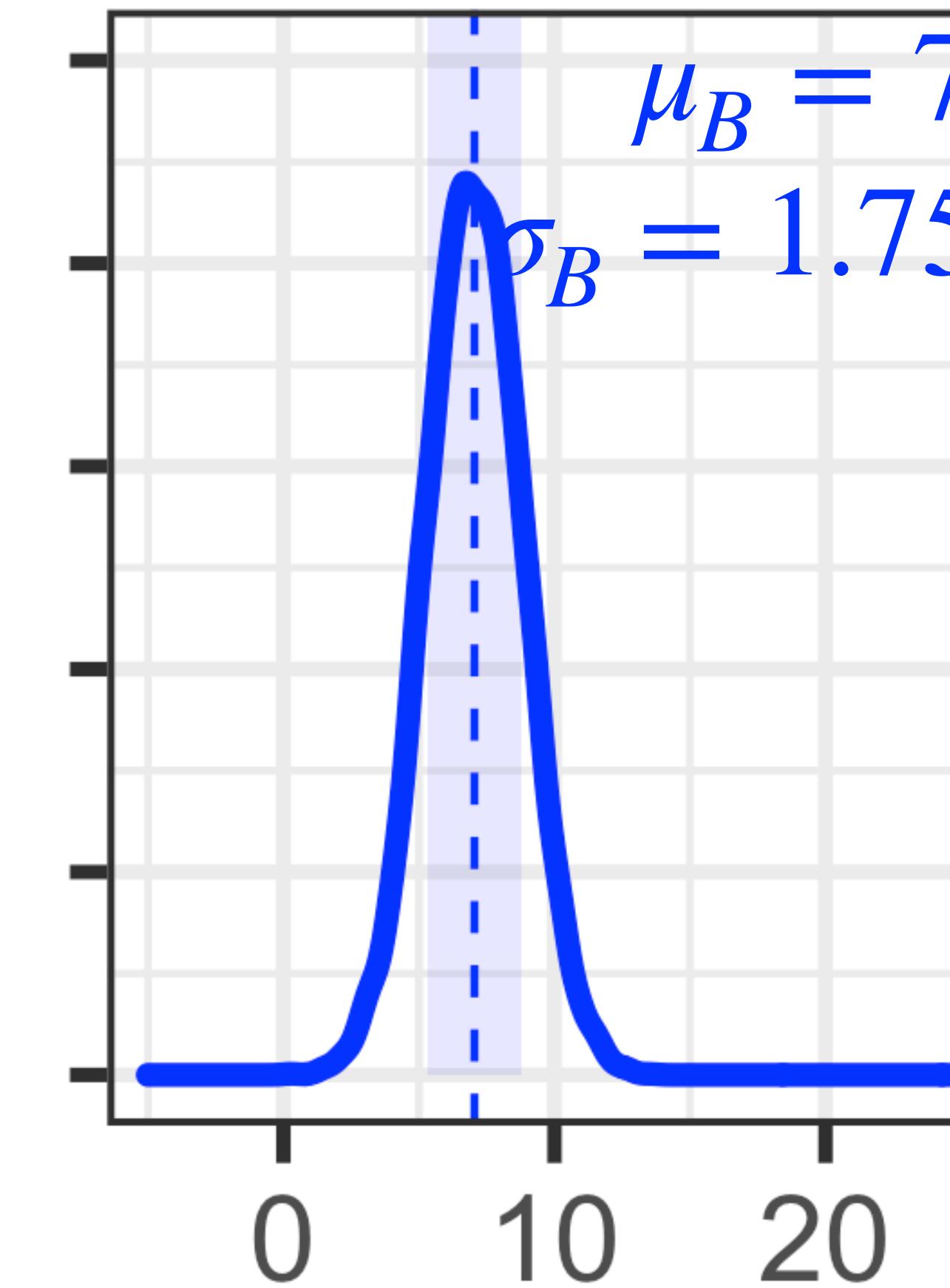
$$\text{Var A} = \{11, 5, 4, 3, 5\}$$

$$\begin{aligned}\mu_A &= 5 \\ \sigma_A &= 2.5\end{aligned}$$



$$\text{Var B} = \{6.5, 6.6, 6.8, 6.5, 6.1\}$$

$$\begin{aligned}\mu_B &= 7 \\ \sigma_B &= 1.75\end{aligned}$$



$$\text{Var (A+B)} = \{6.5-11, 6.6-5, 6.8-4, 6.5-3, 6.1-5\}$$

$$\begin{aligned}\mu_{A-B} &= 2 \\ \sigma_{A-B} &= 3.05\end{aligned}$$

