

Math 140

Introductory Statistics

Disjoint events

If they cannot occur at the same time. For all events

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

For disjoint events $P(A \text{ and } B) = 0$

$$P(A \text{ or } B) = P(A) + P(B)$$

An example

Suppose two dice are rolled.

A is the event of getting a sum of 12,
B is the event of getting two odd numbers.

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

An example

Suppose two dice are rolled.

A is the event of getting a sum of 12,
B is the event of getting two odd numbers.

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

The only way to get a sum of 12 from the dice is (6,6)
and both numbers are even.

So A and B are disjoint and
 $P(A \text{ and } B) = 0$.

Suppose you select a person at random from school.
Which of these pairs of events must be disjoint?

- a. the person has ridden a roller coaster; the person has ridden a Ferris wheel
- b. owns a classical music CD; owns a jazz CD
- c. is a senior; is a junior
- d. has brown hair; has brown eyes
- e. is left-handed; is right-handed
- f. has shoulder-length hair; is a male.

80% of the students carry a backpack B,
or a wallet, W.

Forty percent carry a backpack and 50% carry a wallet.

If a student is selected at random, find the probability
that the student carries both a backpack and a wallet.

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

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

$$0.8 = 0.4 + 0.5 - P(A \text{ and } B)$$

$$P(A \text{ and } B) = 0.1$$

10% of students carry both backpack and wallet

5.3 Conditional probability

The Titanic sank in 1912 without enough lifeboats for the passengers and crew.

Almost 1500 people died, most of them men.

Was that because men were less likely than women to survive?

Or did more men die simply because men outnumbered women by more than 3 to 1 on the Titanic?

The captain of the Titanic *“Women and Children first”*

Gender

		male	female	total
Survived?	yes	367	344	711
	no	1364	126	1490
	total	1731	470	2201

The captain of the Titanic

“Women and Children first”

Female survival percentage?

Male survival percentage?

Overall survival percentage?

The captain of the Titanic *“Women and Children first”*

Female survival percentage? 73%

Male survival percentage? 21%

Overall survival percentage? 32%

It seems that the chance of survival depends on the **condition** on whether or not you are a female

Conditional probability

Conditional Probability refers to the probability of a particular event where additional information is known.



Conditional probability

We write it the following way

$P(S | T)$ = probability of S given that T is known to have happened

$P(S | T)$ = the probability of S given T
(in short)

In other words: we are SURE T happened

Conditional probability on the Titanic

Let's find

$P(S | F)$ = the probability of Survival given you are a Female

		Gender		
		male	female	total
Survived?	yes	367	344	711
	no	1364	126	1490
	total	1731	470	2201

Conditional probability on the Titanic

We need to work ONLY where the condition of being female is true = use the sample space of 470 women

$$P(S | F) = 344/470 = 73\%$$

		Gender		
		male	female	total
Survived?	yes	367	344	711
	no	1364	126	1490
	total	1731	470	2201

Calculate

$P(F)$ = probability of being female

$P(S | F)$ = probability of survival given you are a female

$P(F | S)$ = probability of being female given you survived

$P(\text{not } F)$ = probability of not being a female

		Gender		
		male	female	total
Survived?	yes	367	344	711
	no	1364	126	1490
	total	1731	470	2201

Calculate

$P(F)$ = probability of being female

$P(S | F)$ = probability of survival given you are a female

$P(F | S)$ = probability of being female given you survived

$P(\text{not } F)$ = probability of not being a female

$$P(F) = 470/2201 = 0.213$$

$$P(S | F) = \text{earlier} = 344/470 = 0.732$$

$$P(F | S) = 344/711 = 0.483$$

$$P(\text{not } F) = 1731/2201 = 0.786$$

Calculate

$P(\text{not } F \mid S)$ = of not being female, given you survived

$P(S \mid \text{not } F)$ = probability of survival given you are not a female

$P(F \text{ and } S)$ = probability you were female and survived out of all passengers

Calculate

$P(\text{not F} | S)$ = of not being female, given you

$P(S | \text{not F})$ = probability of survival given you are not a female

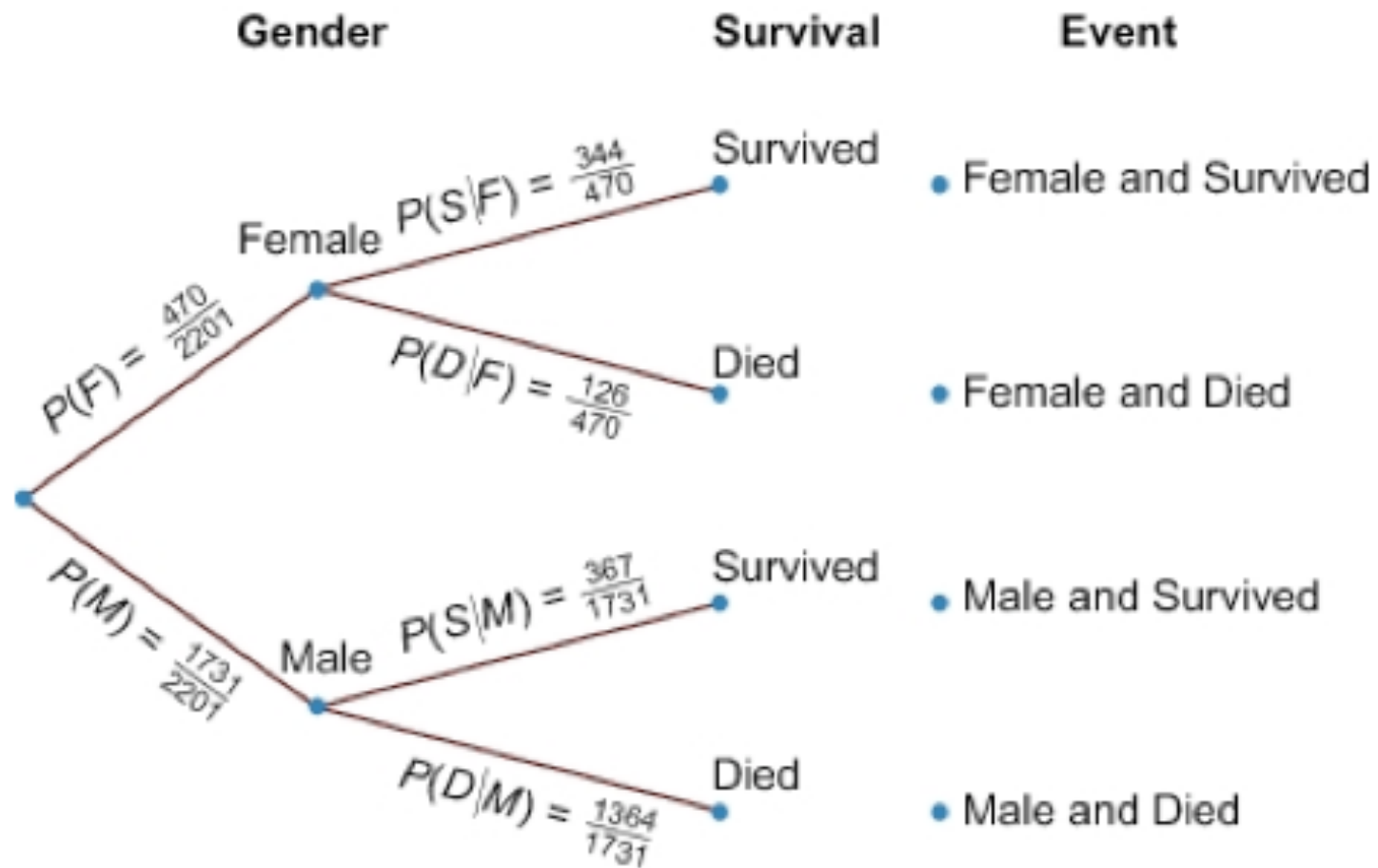
$P(\text{F and S})$ = probability you were female and survived out of all passengers

$$P(\text{not F} | S) = 367/711 = 0.516$$

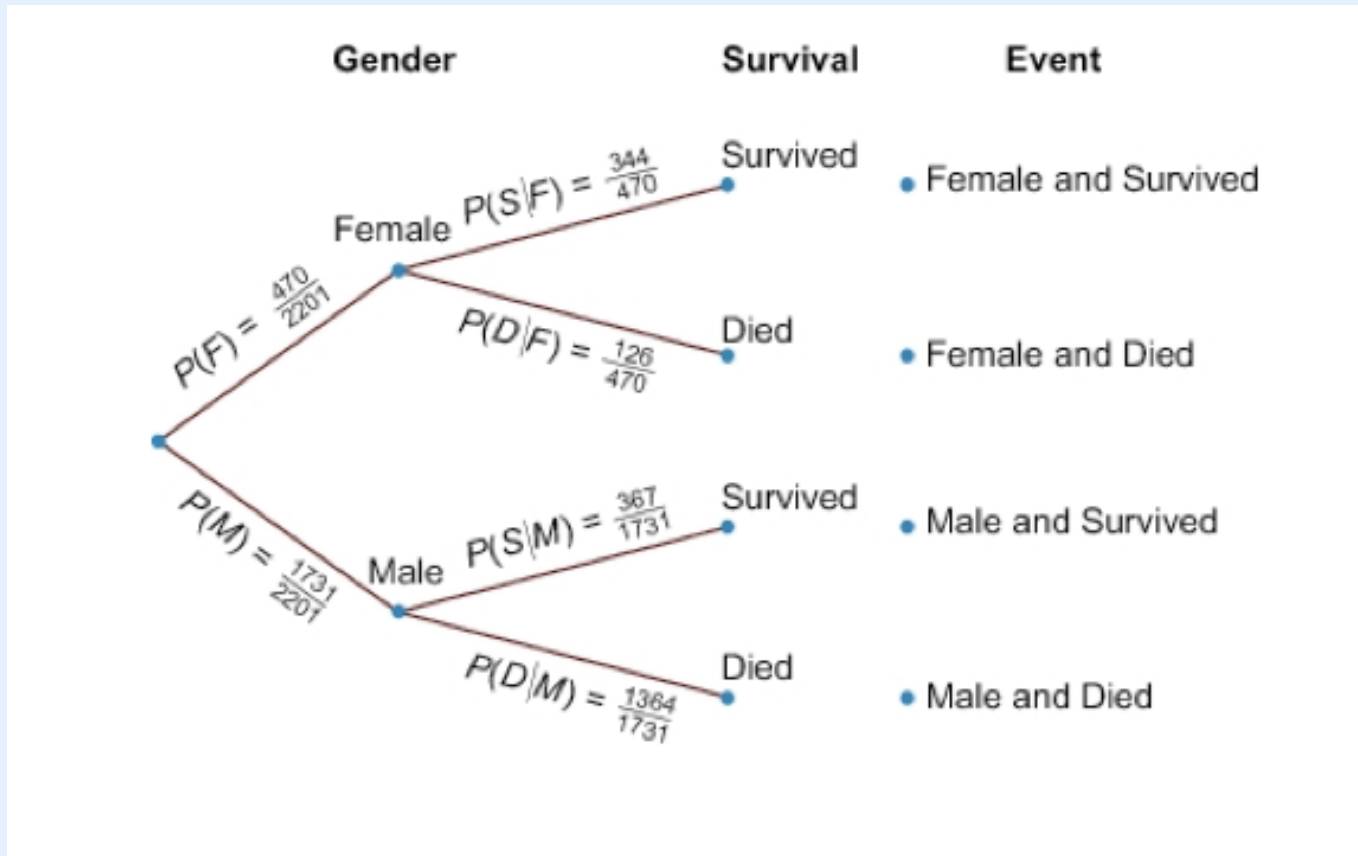
$$P(S | \text{not F}) = 367/1731 = 0.212$$

$$P(\text{F and S}) = 344/2201 = 0.153$$

The tree diagram

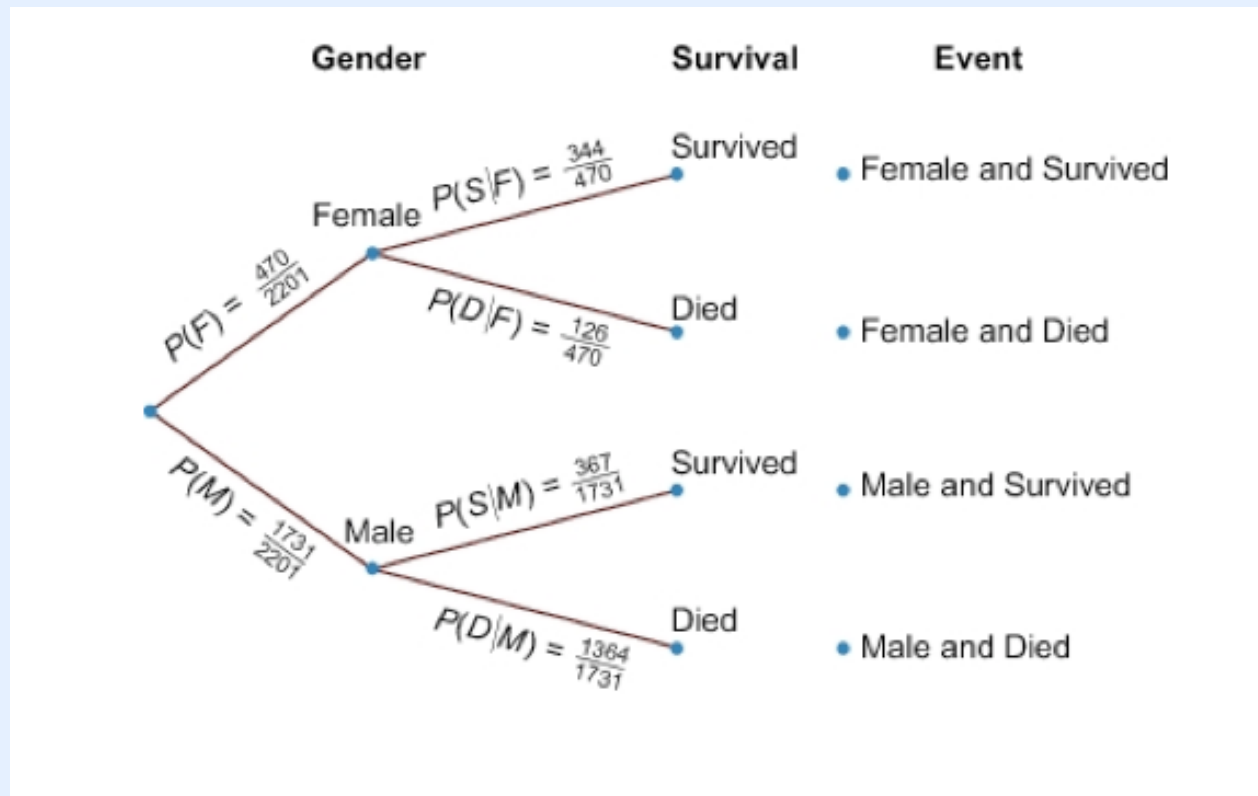


The tree diagram



$$P(F \text{ and } S) = \frac{344}{2201} = \frac{344}{470} * \frac{470}{2201}$$

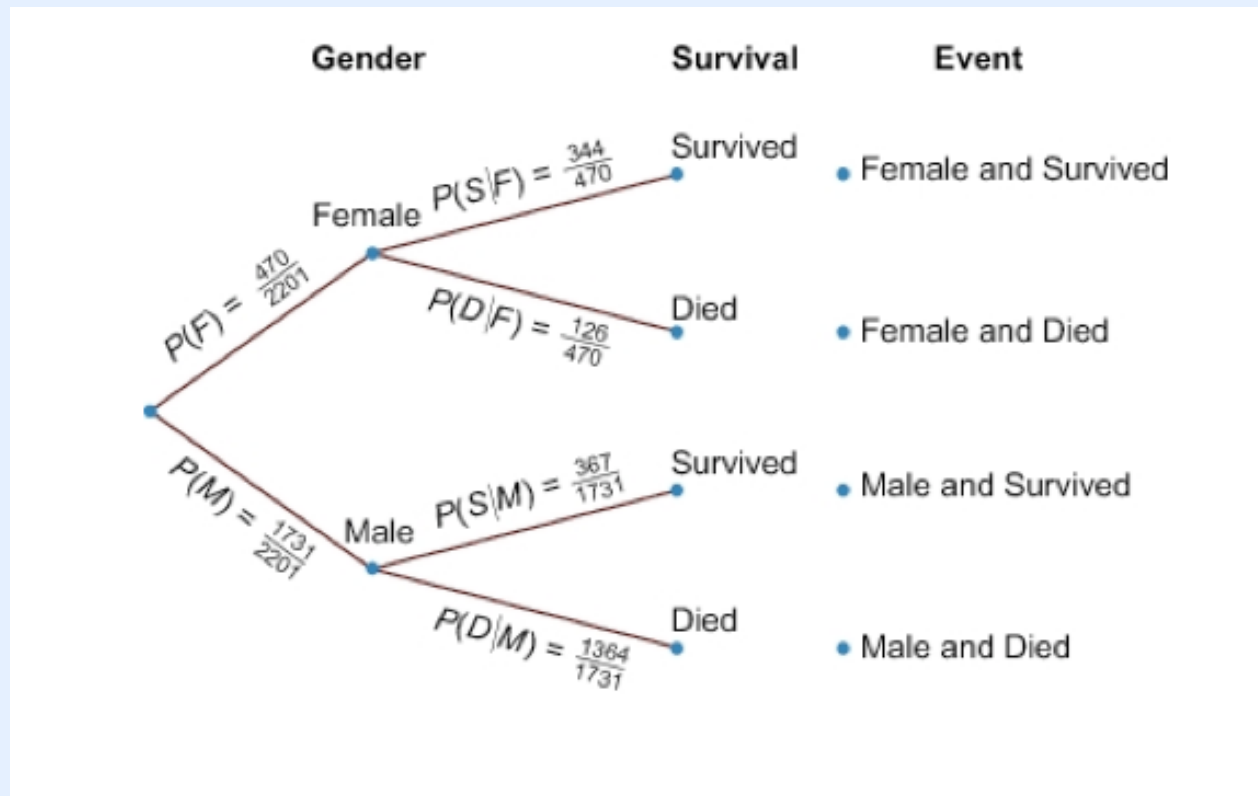
Try, just for fun



$$P(\text{F and S}) = 344/2201 = 344/470 * 470/2201$$

$$\text{But: } 344/470 = P(S | F)$$

$$\text{And: } 470/2201 = P(F)$$



So we can say that

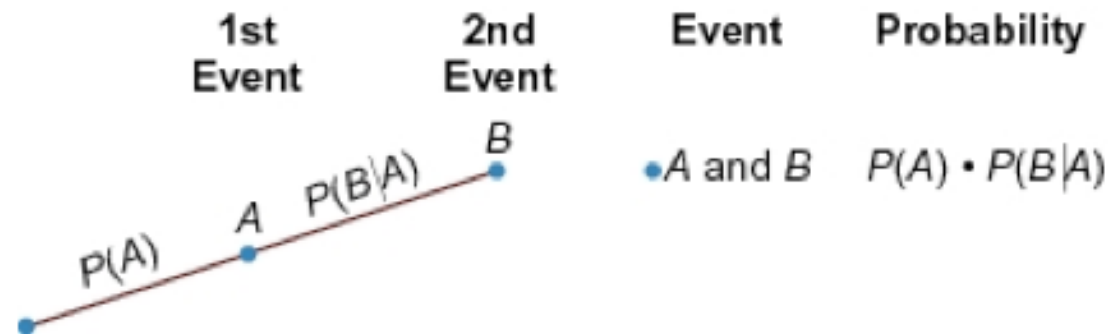
$$P(F \text{ and } S) = 344/2201 = 344/470 * 470/2201$$

$$P(F \text{ and } S) = P(S | F) * P(F) = P(F) * P(S | F)$$

This is true in general

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

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



Moving along the branch of the tree diagram

This is true in general

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

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

For the Titanic data, show that

$$P(\text{Male and Survive}) = P(\text{Survive}) * P(\text{Male} | \text{Survive})$$

$$P(\text{Male and Survive}) = P(\text{Male}) * P(\text{Survive} | \text{Male})$$

Conditional probability

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

Divide by $P(B)$ both sides

For any two events A and B such that $P(B) > 0$,

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

Example: Roll two dice.

Find the probability that you get a sum of 8 given that you rolled doubles.

Conditional probability

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

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

Example: Roll two dice.

Find the probability that you get a sum of 8 given that you rolled doubles.

$$P(8 | \text{doubles}) = P(8 \text{ and doubles}) / P(\text{doubles})$$

Conditional probability

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

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

Roll two dice.

Find the probability that you get a sum of 8 given that you rolled doubles.

$$\begin{aligned} P(8 | \text{doubles}) &= P(8 \text{ and doubles}) / P(\text{doubles}) \\ &= P(\text{two fours}) / P(\text{doubles}) \\ &\quad 1/36 \quad / \quad 6/36 \end{aligned}$$

Conditional probability

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

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

Roll two dice.

Find the probability that you get a sum of 8 given that you rolled doubles.

$$\begin{aligned} P(8 | \text{doubles}) &= P(8 \text{ and doubles}) / P(\text{doubles}) \\ &= 1/36 * 36/6 = 1/6 \end{aligned}$$

Screening tests

(ELISA - HIV or chest Xray - lung cancer)

Preliminary, less invasive but **less accurate**
Four possibilities

		Test Result		
		Positive	Negative	Total
Disease	Present	a	b	$a + b$
	Absent	c	d	$c + d$
	Total	$a + c$	$b + d$	$a + b + c + d$

Screening tests

		Test Result		
		Positive	Negative	Total
Disease	Present	a	b	$a + b$
	Absent	c	d	$c + d$
	Total	$a + c$	$b + d$	$a + b + c + d$

$P(\text{disease is present} \mid \text{test is positive}) =$

$P(\text{disease is absent} \mid \text{test is negative}) =$

Screening tests

		Test Result		
		Positive	Negative	Total
Disease	Present	a	b	$a + b$
	Absent	c	d	$c + d$
	Total	$a + c$	$b + d$	$a + b + c + d$

$$P(\text{disease is present} \mid \text{test is positive}) = \frac{a}{a+c}$$

$$P(\text{disease is absent} \mid \text{test is negative}) = \frac{d}{b+d}$$

Screening tests

		Test Result		
		Positive	Negative	Total
Disease	Present	a	b	$a + b$
	Absent	c	d	$c + d$
	Total	$a + c$	$b + d$	$a + b + c + d$

$P(\text{test is positive} \mid \text{disease present}) =$

$P(\text{test is negative} \mid \text{disease is absent}) =$

Screening tests

		Test Result		
		Positive	Negative	Total
Disease	Present	a	b	$a + b$
	Absent	c	d	$c + d$
	Total	$a + c$	$b + d$	$a + b + c + d$

$$P(\text{test is positive} \mid \text{disease present}) \\ = \frac{a}{a+b}$$

$$P(\text{test is negative} \mid \text{disease is absent}) = \\ \frac{d}{c+d}$$

Screening tests

		Test Result		
		Positive	Negative	Total
Disease	Present	a	b	$a + b$
	Absent	c	d	$c + d$
	Total	$a + c$	$b + d$	$a + b + c + d$

$P(\text{disease is absent} \mid \text{test is positive}) =$

$P(\text{disease is present} \mid \text{test is negative}) =$

Screening tests

		Test Result		
		Positive	Negative	Total
Disease	Present	a	b	$a + b$
	Absent	c	d	$c + d$
	Total	$a + c$	$b + d$	$a + b + c + d$

$$P(\text{disease is absent} \mid \text{test is positive}) = \frac{c}{a+c}$$

$$P(\text{disease is present} \mid \text{test is negative}) = \frac{b}{b+d}$$

Definitions

Positive predictive value (PPV)

$$P(\text{disease is present} \mid \text{test is positive}) = a / (a+c)$$

Negative predictive value (NPV)

$$P(\text{disease is absent} \mid \text{test is negative}) = d / (b+d)$$

Sensitivity

$$P(\text{test is positive} \mid \text{disease present}) = a / (a+b)$$

Specificity

$$P(\text{test is negative} \mid \text{disease is absent}) = d / (c+d)$$

Definitions

False positive rate

$$P(\text{disease is absent} \mid \text{test is positive}) = c / (a+c)$$

False negative rate

$$P(\text{disease is absent} \mid \text{test is negative}) = d / (b+d)$$

A rare disease - statistics

		Test Result		
		Positive	Negative	Total
Disease	Present	9	1	10
	Absent	50	9,940	9,990
	Total	59	9,941	10,000

Find

False Pos. Rate= $P(\text{no disease} \mid \text{test positive})$

False Neg. Rate= $P(\text{disease present} \mid \text{test negative})$

Sensitivity= $P(\text{test positive} \mid \text{disease present})$

Specificity= $P(\text{test negative} \mid \text{no disease})$

PPV = $P(\text{disease present} \mid \text{test positive})$

NPV= $P(\text{no disease} \mid \text{test negative})$

Hk

Page 255

P21, P22, P31, E29, E30, E33, E34, E35, E40,