Data Science Math Skills
Paul Bendich and Daniel Egger
Duke University

Sets: Medical Testing Example

Video companion

1 Example using set theory

VBS: "very bad syndrome"

X = set of people in a clinical trial

$S = \{x \in X : x \text{ has VBS}\}$
$H = \{x \in X : x \text{ does not have VBS}\}$

$X = S \cup H$ (you either have VBS or you don't)
$S \cap H = \emptyset$ (no one both has and doesn't have it)

Point of medical testing to figure out whether a person is in S or in H

2 Test

$P = \{x \in X : x \text{ tests positive for VBS}\}$
$N = \{x \in X : x \text{ tests negative for VBS}\}$

$P \cup N = X$ (you either test positive or negative)
$P \cap N = \emptyset$ (no one tests both positive and negative)

In a perfect world, S would equal P—the sick people would always test positive, and H would equal N—the healthy people would always test negative.

...but this is not always the case.

$S \cap P$	$H \cap N$	$S \cap N$	$H \cap P$
true positive	true negative	false negative	false positive

real life outcomes

3 Cardinality

$\frac{|S|}{|X|}$ = proportion of people in the study who do genuinely have VBS
$\frac{|H|}{|X|}$ = proportion of people in the study without VBS
$\frac{|S|}{|X|} + \frac{|H|}{|X|} = 1$

careful in choosing a representative sample, not a biased one, to represent the total population

$\frac{\lvert S \cap P \rvert}{\lvert S \rvert}$	true positive rate	would like to be close to 1
$\frac{\lvert H \cap P \rvert}{\lvert H \rvert}$	false positive rate	would like to be as small as possible
$\frac{\lvert S \cap N \rvert}{\lvert S \rvert}$	false negative rate	would like to be as small as possible
$\frac{\lvert H \cap N \rvert}{\lvert H \rvert}$	true negative rate	would like to be close to 1