

Math 247: Association and Categorical Variables (Section 5.3)

Conditional Probability: Conditional probability is the case where we narrow our focus to looking at probabilities (percentages) relative to a smaller group instead of looking at the entire group.

If you have the actual numbers of individuals in the A, B categories, then we can use this formula:

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

Example: Research question: How accurate are drug-screening tests used for workers, athletes, parolees, etc.?

Describe how we could do a controlled experiment for the accuracy of steroid drug tests for athletes. Assume we have a random sample of 200 adults to work with.

Is this ethical?

Suppose this was the result of your experiment:

	Took Steroid	Did Not Take Steroid	
Test Positive	95	3	
Test Negative	5	97	

What is the probability that a person tests positive given that he/she did not take steroids? (This result is called a “false positive”.) Write the probability statement using the proper notation.

What is the probability that a person tests negative given she/he did take the drug? (This result is called a “false negative”.) Write the probability statement using the proper notation.

Sequences of Events (Compound Events)

In 2012, Lance Armstrong was stripped of his 7 Tour de France cycling titles along with his Olympic bronze medal because of evidence that he had used performance-enhancing steroids. He had tested positive for steroids in 4 different tests in 1999. What is the probability Lance would have a false positive four times if, in fact, he hadn't used any steroids?

Write this probability problem in terms of "AND"

We need a rule now for dealing with AND statements (just as we have a rule for dealing with OR statements).

The General Multiplication Rule: $P(A \text{ and } B) = P(A) \cdot P(B|A)$

Multiplication Rule for Independent Events $P(A \text{ and } B) = P(A) \cdot P(B)$

For both of these rules, what mathematical operation do we use for "AND"? _____

Back to Lance: What is the probability of getting 4 false positives on steroid tests? Use $P(\text{FP}) = 1.5\%$

Example: If you roll a die twice, what is the theoretical probability that the die comes up 5 both times?

Write this probability problem in terms of "AND"

Example: If you draw two cards from a standard deck of cards, what is the theoretical probability both will be aces? Write each probability problem in terms of "AND"

(a) Situation 1: You put the first card back in the deck (*sample with replacement*)

(b) Situation 2: You don't put the first card back in the deck (*sample without replacement*)

Independent Events: Events A and B are independent if the fact that one event occurred doesn't affect the probability of the other event in any way.

Definition: A and B are independent when $P(A|B) = P(A)$ and $P(B|A) = P(B)$

If events are not independent, then we say they are "**associated**" (or "**dependent**").

Note: Association does not mean causation!

Use your knowledge of the world to determine whether these events are independent or are associated:

A person is a water polo player.
A person is a good swimmer.

A person is a water polo player.
A person has a white car.

A high school student is female.
A high school student dropped out (didn't graduate).

Given actual data, we can test for association by looking at the percentages and checking whether $P(A) \neq P(A|B)$ OR $P(B) \neq P(B|A)$

Example: Suppose there were 1000 boys and 1000 girls in a high school. If 120 of the boys dropped out and 80 of the girls dropped out, is there an association between gender and dropping out?

Now let's suppose there were 1200 boys and 800 girls. If 120 boys dropped out and 80 girls dropped out, is there an association between gender and dropping out?