

# Conditional Probability and Independence

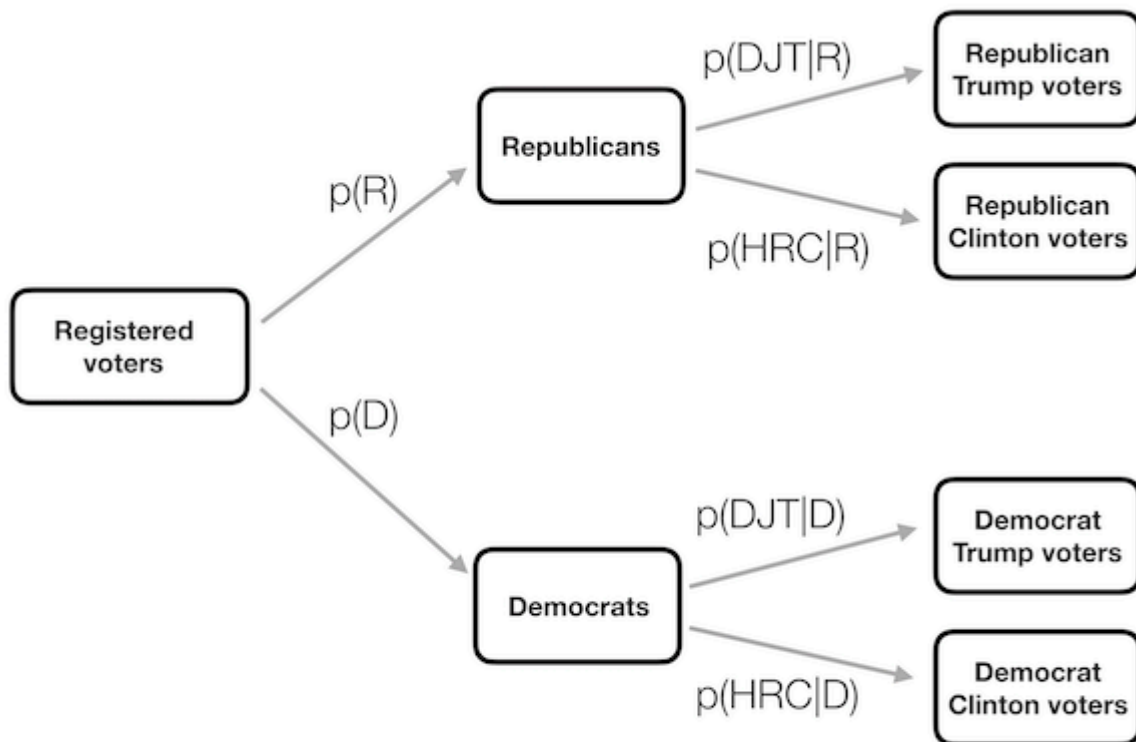
## What is Conditional Probability?

### Definition:

- Probability of A given B occurred
- Written as  $P(A | B)$
- Updates probability based on new information

### Formula:

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



So far we have limited ourselves to simple probabilities - that is, the probability of a single event or combination of events. However, we often wish to determine the probability of some event given that some other event has occurred, which are known as conditional probabilities.

## NHANES Example: Physical Activity

### Question:

What is  $P(\text{diabetes}|\text{inactive})$ ?

| total | inactive | diabetes | diabetes_given_inactive |
|-------|----------|----------|-------------------------|
| 5443  | 0.454    | 0.101    | 0.141                   |

### Joint Probabilities:

| Diabetes | PhysActive | n    | prob      |
|----------|------------|------|-----------|
| No       | No         | 2123 | 0.3900423 |
| No       | Yes        | 2770 | 0.5089105 |
| Yes      | No         | 349  | 0.0641191 |
| Yes      | Yes        | 201  | 0.0369282 |

We can compute conditional probabilities directly from data. Let's say that we are interested in the following question: What is the probability that someone has diabetes, given that they are not physically active? The NHANES dataset includes two variables that address the two parts of this question: Diabetes and PhysActive.

## Independence

### Statistical Independence:

$$P(A \mid B) = P(A)$$

### Key Points:

- B tells us nothing about A
- Different from everyday usage
- Must check with data

### Example: Jefferson State

- $P(\text{Jeffersonian}) = 0.014$
- $P(\text{Californian}) = 0.986$
- Not independent!
- Mutually exclusive

The term "independent" has a very specific meaning in statistics, which is somewhat different from the common usage of the term. Statistical independence between two variables means that knowing the value of one variable doesn't tell us anything about the value of the other.

For example, there is currently a move by a small group of California citizens to declare a new independent state called Jefferson. The new states might be politically independent, but they would not be statistically independent, because if we know that a person is Jeffersonian, then we can be sure they are not Californian!

## Mental Health and Physical Activity

**Question:** Are physical and mental health independent?

**Variables:**

- PhysActive: physically active?
- DaysMentHlthBad: bad mental health days
- Threshold: >7 days = bad mental health

| PhysActive | Bad Mental Health | Good Mental Health | Total |
|------------|-------------------|--------------------|-------|
| No         | 629               | 2510               | 3139  |
| Yes        | 471               | 3095               | 3566  |
| Total      | 1100              | 5605               | 6705  |

Let's look at another example, using the NHANES data: Are physical health and mental health independent of one another? To determine whether mental health and physical activity are independent, we would compare the simple probability of bad mental health to the conditional probability of bad mental health given that one is physically active.

**Bibliography**