ST 516: Foundations of Data Analytics

Statistical Independence

Independence of Events

The word **independence** has a special meaning in probability, statistics and data analysis.

If events A and B are independent, the probability of event A is the same whether or not event B occurs.

This is made more precise with conditional probability notation:

- two events, A and B, are independent if $\mathsf{P}(A|B) = \mathsf{P}(A)$
- equivalently, A and B are independent if $\mathsf{P}(B|A) = \mathsf{P}(B)$

Example

In one roll of a fair, six-sided die, let A be the event that a 2 appears, and let B be the event that an even number appears.

Are A and B independent events?

Example

In one roll of a fair, six-sided die, let A be the event that a 2 appears, and let B be the event that an even number appears.

Are A and B independent events?

- $\mathsf{P}(A) = \frac{1}{6}$

Example

In one roll of a fair, six-sided die, let A be the event that a 2 appears, and let B be the event that an even number appears.

Are A and B independent events?

- $\mathsf{P}(A) = \frac{1}{6}$
- $\mathsf{P}(A|B) = \frac{1}{3}$

Example

In one roll of a fair, six-sided die, let A be the event that a 2 appears, and let B be the event that an even number appears.

Are A and B independent events?

- $\mathsf{P}(A) = \frac{1}{6}$
- $\mathsf{P}(A|B) = \frac{1}{3}$
- Since these are not equal, A and B are not independent.

The Product Rule, Revisited

Recall that the product rule for joint probability is given by:

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

but if A and B are independent, we just saw that $\mathrm{P}(A|B) = \mathrm{P}(A)$.

This gives us a product rule for independent events:

For independent events, A and B, $\mathrm{P}(A \text{ and } B) = \mathrm{P}(A)\mathrm{P}(B)$.

This version of the product rule is very important in data analysis, and so it will be important for you to learn how to decide whether a sample of observations (or a collection of events) is independent.

Example

Let M represent the event that a man selected at random has a moustache and let B represent the event that a man selected at random has a beard. Suppose $\mathsf{P}(M) = \frac{1}{4}, \mathsf{P}(B) = \frac{1}{10}$, and

$\mathsf{P}(M \text{ and } B) = \dfrac{1}{25}$.

Are the events M and B independent?

- We know that if the two events are independent, then the product rule for independent events holds.
- In this case since $\mathsf{P}(M)\mathsf{P}(B) = \frac{1}{4}\frac{1}{10} = \frac{1}{40} \neq \frac{1}{25}$, the rule does not hold—the events re not independent.

Independence and Randomness

If a simple random sample is drawn from a population, then the observations in that sample are typically independent.

- We say "typically here," because there are situations when a simple random sample can result in some or all of the observations being non-independent.
    - This might happen when observations are located close together in space or time.
    - It can also happen when observations are clustered somehow, and certain observations belong to different clusters than other observations.
    - In these situations, it will be important for the data analyst to account for any non-independence in the sample.