MATH 3339
Statistics for the Sciences
Live Lecture Help

James West
jdwest@uh.edu

University of Houston

Session 3

Office Hours: see schedule in the "Office Hours" channel on Teams
Course webpage: www.casa.uh.edu

Email policy

When you email me you **MUST** include the following

- MATH 3339 Section 20024 and a description of your issue in the **Subject Line**
- Your name and ID# in the **Body**
- Complete sentences, punctuation, and paragraph breaks
- Email messages to the class will be sent to your Exchange account (user@cougarnet.uh.edu)

Updates

- Access code enforcement begins at midnight.

Using R and R-Studio

1. Download R from https://cran.r-project.org/
2. Download R-Studio from https://www.rstudio.com/

Outline

1. Recap
2. Examples
3. Student submitted questions

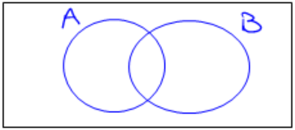
Conditional Probability

Let A and B be events with $P(B) > 0$. The **conditional probability** of A, given B is:

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

General rule for multiplication: For any two events E and F, $P(E \cap F) = P(E) \times P(F|E)$ or $P(E \cap F) = P(F) \times P(E|F)$.

Two Frequently Asked Questions



1. When do I add and when do I multiply?
   - Add when finding the chance of events A **or** B (or both) happening.

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

   - Multiply when finding the chance that both events A **and** B happen.

   $$P(A \text{ and } B) = P(A \cap B) = P(A) \times P(B, \text{ given } A) = P(A)P(B|A)$$

Two Frequently Asked Questions

2. What's the difference between disjoint (mutually exclusive) and independent?
   - Two events are disjoint if the occurrence of one prevents the other from happening.

     $$P(A \cap B) = 0$$

   - Two events are independent if the occurrence of one does not change the *probability* of the other.

     $$P(A|B) = P(A)$$

     or $P(A \cap B) = P(A) \cdot P(B)$

Bayes' Rule

- The probability of a person buying an iMac, given they are first-time buyers is an example of using **Bayes' rule**.
- Given a prior (initial) probability, then from sources we obtain additional information about the events.
- From these events we revise the probabilities and get a posterior probability.
- This is an application of the General Multiplication Rule.
- It might be easier to use the tree diagram to calculate this probability.

Let A and B be two events with $P(B) \neq 0$ then we have:

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

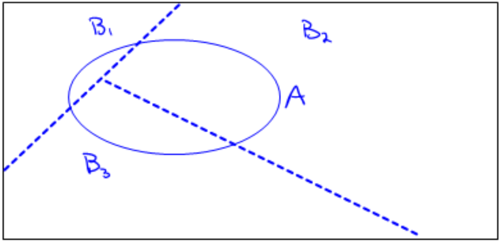
We use following facts:

1. $P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$
2. $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

Bayes' Rule

Let A and $B_1, B_2, \ldots, B_k$ be pairwise disjoint events such that each $P(B_i) > 0$ and $\Omega = B_1 \cup B_2 \cup \ldots \cup B_k$ and assume $P(A) > 0$. Then for each i,

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}$$



Mean

Mean

The mean is a measure of the center of data.

To calculate the mean for a sample with n values:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

For a population with N values:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Note: The mean is NOT robust against extreme values. The mean is pulled away from the center of the distribution toward the extreme value (“tails of graph").

mean(x)

Median

Median
The median is also a measure of the center of data.

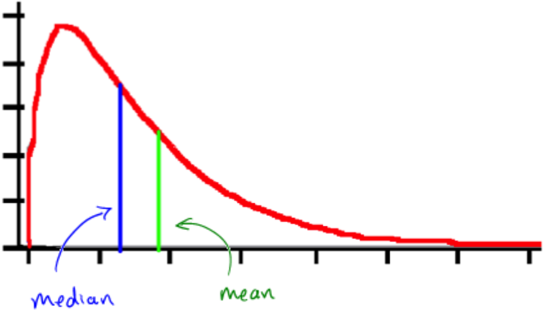
To calculate the median for a discrete data set, put the values in order from least to greatest and locate the center. If the list has an even number of elements, average the middle two values.

median(x)

Note: The median is resistant to extreme values (outliers) in data set.

Measures of Location

Of the 2 segments, where's the Mean with respect to the Median?



Remember the mean is pulled toward extreme values.

Percentiles (Quantiles)

In general, if you have n data measurements, x_1 represents the $\frac{100(1-0.5)}{n}^{th}$ percentile, x_2 represents the $\frac{100(2-0.5)}{n}^{th}$ percentile, and x_i represents the $\frac{100(i-0.5)}{n}^{th}$ percentile.

This is useful if you wish to calculate the percentile rank of a known measurement.

If you are looking for the measurement that has a desired percentile rank, the $100 \cdot P^{th}$ percentile, is the measurement with rank $nP + 0.5$.

The 25^{th} percentile is called the first quartile, Q_1. It represents the first $\frac{1}{4}$ of the data. Similarly, the 50^{th} and 75^{th} percentiles are the second and third quartiles, Q_2 and Q_3, respectively.

A party is held where everyone is offered and eats exactly one meal option. Of those in attendance 50% prefer the first meal (tacos), 30% prefer the second (pizza), and everyone else prefers the third (hot dog). Of the people who ate tacos, 1% got sick. Of the people who ate pizza, 2% got sick. 5% of the people who ate a hot dog got sick.

1. Draw a tree diagram for this problem.

A party is held where everyone is offered and eats exactly one meal option. Of those in attendance 50% prefer the first meal (tacos), 30% prefer the second (pizza), and everyone else prefers the third (hot dog). Of the people who ate tacos, 1% got sick. Of the people who ate pizza, 2% got sick. 5% of the people who ate a hot dog got sick.

2. If a guest is randomly selected, what is the probability that the guest ate pizza and did not get sick?

A party is held where everyone is offered and eats exactly one meal option. Of those in attendance 50% prefer the first meal (tacos), 30% prefer the second (pizza), and everyone else prefers the third (hot dog). Of the people who ate tacos, 1% got sick. Of the people who ate pizza, 2% got sick. 5% of the people who ate a hot dog got sick.

3. Given that a guest got sick, what is the probability that the guest ate hot dogs?

If two events A and B are both independent and mutually exclusive, which of the following must be true?

A, B are independent: $P(A \mid B) = P(A)$

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

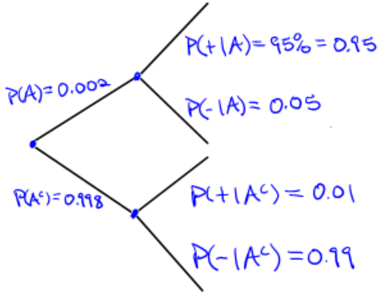
or $P(A \cap B) = P(A) \cdot P(B)$

A, B mutually exclusive: $P(A \cap B) = 0$

Thus, $0 = P(A) \cdot P(B)$

Example

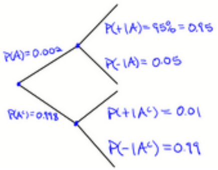
A rare disease exists in which only 1 in 500 are affected. A test for the disease exists but of course it is not infallible. A correct positive result (patient actually has the disease) occurs 95% of the time while a false positive result (patient does not have the disease) occurs 1% of the time. If a randomly selected individual is tested and the result is positive, what is the probability that the individual has the disease? $P(A) = \frac{1}{500} = 0.002$, $P(A^c) = \frac{499}{500} = 0.998$



$$P(A \cap +) = P(A) \cdot P(+|A) = (0.002)(0.95)$$
$$= 0.0019$$

$$P(A^c \cap +) = P(A^c) \cdot P(+|A^c)$$
$$= (0.998)(0.01)$$
$$= 0.00998$$

Continuing the last example, what is $P(A|-)$ (Round to four decimal places)?



We know $$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}$$

Here $$P(A|-) = \frac{P(-|A)\cdot P(A)}{P(-)}$$

$$P(-) = P\big((-\cap A)\cup(-\cap A^c)\big) = P(-\cap A) + P(-\cap A^c)$$
$$= P(A)\cdot P(-|A) + P(A^c)\cdot P(-|A^c)$$

$$P(A|-) = \frac{P(-|A)\cdot P(A)}{P(A)\cdot P(-|A) + P(A^c)\cdot P(-|A^c)}$$

The “mammals” data set is a built-in dataset in the “MASS” library of R. The “mammals” data set contains the result of a study of sleep in mammal species. First, load the “mammals” data set into your R workspace. In Rstudio, you can click on the “Packages” tab and then on the checkbox next to MASS. Without Rstudio, type the following command in R console:

```
data(mammals,package="MASS")
```

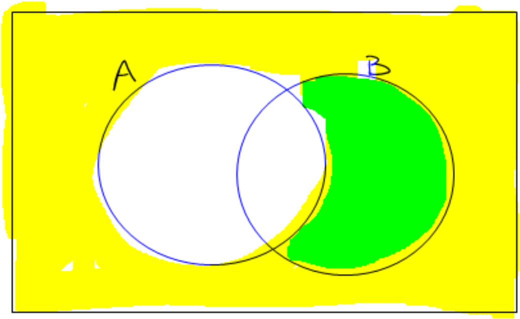
A random experiment is to choose one of the species listed in this data set. All outcomes are equally likely. You can obtain a list of the species in the event “brain > 500” with the command

```
subset(mammals,brain>500)
```

What is the probability of this event, i.e., what is the probability that you randomly select a species with a brain weight greater than 500g?

Hint: you can obtain a count of the species with brain weights greater than 500g, by

```
sum(mammals$brain>500)
```


A
B

Using R and R-Studio

1. Download R from https://cran.r-project.org/
2. Download R-Studio from https://www.rstudio.com/