

ENV 710: Lecture I

data description

roadmap

- modifications to syllabus
- logistics and questions
- pod work – descriptive statistics



syllabus

Diversity Statement

It is my intent that students from all diverse backgrounds, perspectives, and abilities be well served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength and benefit. It is my intent to present materials and activities that are respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups.

Data Acknowledgement

In this class we use a variety of different datasets selected because they have the appropriate numerical and design characteristics to illustrate data analysis and statistical concepts. In dealing with these data, we often overlook their context, and therefore, explicit and implicit biases as well as sensitive issues regarding gender, race, sexual orientation and diversity. The intent is not to ignore or minimize these issues of paramount importance, but rather to focus on the course objectives of understanding and using methods of data analysis.

syllabus

- modified the syllabus slightly (corrected some dates, etc.)
 - download new version, if necessary
- teaching philosophy again...

Pods

Name	Pod	Name	Pod	Name	Pod
Benaka, Isaac	1	Go, Li Jia	5	Palia, Sophia	9
Bernaus, Katrina	1	Gulino, Justin	5	Pang, MiaoJun	9
Bi, Yuntian	1	Haber, Jordan	5	Pike, Rachel	9
Bliska, Hanna	2	Harvey, Marla	6	Price, Noah	10
Brentjens, Emma	2	Hays, Brandon	6	Rowley, Caroline	10
Campos, Gabriel	2	Healey, Liam	6	Satagopan, Nanditha Ram	10
Carlson, Maria	3	Hyun, Jiwon	7	Seagle, Jenna	11
Davidson, Kelly	3	Jackson, Rachel	7	Sepe, Stevie	11
Diaz, Danae	3	Kuhlmann, Emily	7	Sun, Suri	11
Dye, Logan	4	Martinez, Laura	8	White, Libby	12
Franzetti, Tristan	4	Merritt, Melissa	8	Wong, Richard	12
Freedman, Jacob	4	Owens, Katie	8	Yang, Wookeun	12
				Zungailia, Isabel	12

- introduce yourselves
- report to the class the name of one of your pod mates and an interesting statistic about them
- as a pod, one fact learned or question from 1st video lecture or lab

I – data description

From memory, discuss the following terms/subjects (5 min.)

- what are the different types of data and examples of each?
- what do the terms population, sample, parameter and statistic mean?
- name three measures of central tendency and when they should be used
- name three measures of spread and when they should be used

- describe this curve



write down any questions that arise and any topics or terms that are not clear

I – data description

From memory, discuss the following terms/subjects (5 min.)

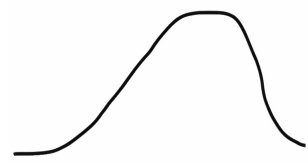
- what are the different types of data and examples of each?
- what do the terms population, sample, parameter and statistic mean?
- name three measures of central tendency and when they should be used
- name three measures of spread and when they should be used

numerical:
continuous, discrete
categorical: nominal,
ordinal

mean, median, mode

variance, range,
standard deviation

- describe this curve



I – data description

Type the following code into R:

```
set.seed(710)
nums <- rnorm(30, mean = 8, sd = 8)
```

- Using R, calculate the mean, median, variance, and standard deviation of nums.
- What does `set.seed()` do?
- Create a boxplot of nums

5 Rules of Probability

(So simple even your cat can understand them)

$$0 \leq P(A) \leq 1$$



1. The probability Felix will eat pizza for dinner tonight is somewhere between zero and 100%.

$$P(A) + P(B) = 1$$



2. Left alone in the yard, the probability Fido will dig up the flower garden = 88%; The probability he will find other "entertainment" = 12%
(Probability he'll get up to some kind of mischief = 88% + 12% = 100%).

$$P(\text{not } A) = 1 - P(A)$$



3. Felix is Asleep 75% of the time. So the probability he is not Asleep is
 $100\% - 75\% = 25\%$.

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



4. Felix is in a box, or he's not in a box (i.e. he can't be both). If he's in the box 40% of the time, then
the probability Felix is in a box OR he's not in a box = $40\% + 60\% = 100\%$

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



5. Fido brings you one of his favorite toys.



The probability it's a ball OR it's red = $50\% + 50\% (-25\%) = 75\%$

Probability Rules Cheat Sheet

complement rule

$$P(A) = 1 - P(A')$$

multiplication rules (joint probability)

dependent $P(A \cap B) = P(A) * P(B|A)$

independent $P(A \cap B) = P(A) * P(B)$

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

addition rules (union of events)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

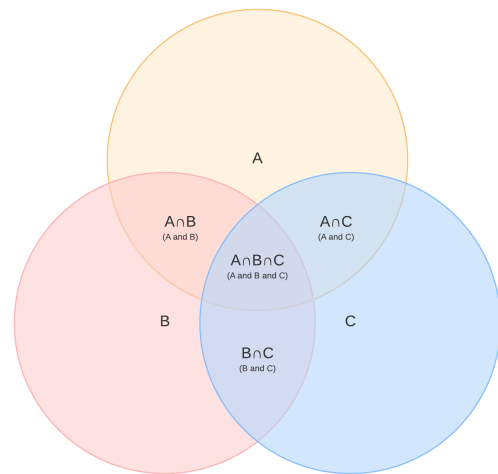
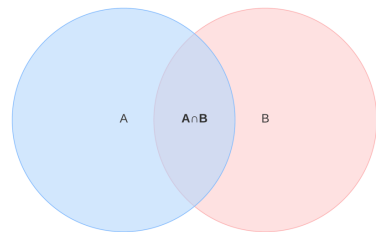
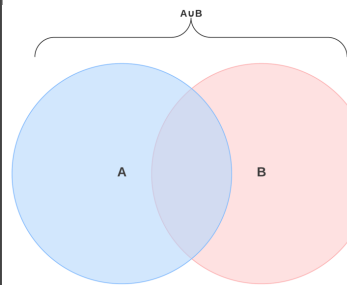
mutually exclusive $P(A \cup B) = P(A) + P(B)$

conditional probability

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

Bayes' Theorem

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

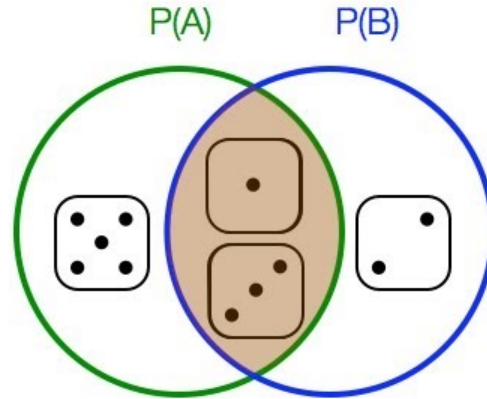


Conditional Probability

What is the Probability of
rolling a dice and it's
value is less than 4

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

knowing that the value is
an odd number





Questions?