

Hi!

DS501: Basic Statistics, Machine Learning, and Review

Prof. Randy Paffenroth
rcpaffenroth@wpi.edu

Worcester Polytechnic Institute



WPI

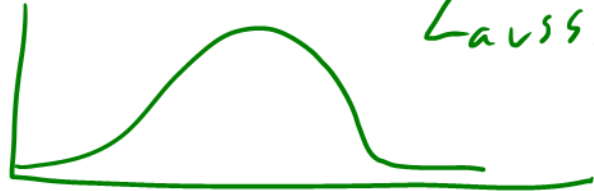
Announcements

- Case study 2 due next week! Wednesday 3/2.
- Midterm exam is **also** next week! Wednesday 3/2.
- The idea will be to have presentations the first half of the class and the midterm exam the second half.



https://upload.wikimedia.org/wikipedia/commons/f/f0/Fear_has_big_eyes.jpg

Descriptive statistics: Mean *pdf*

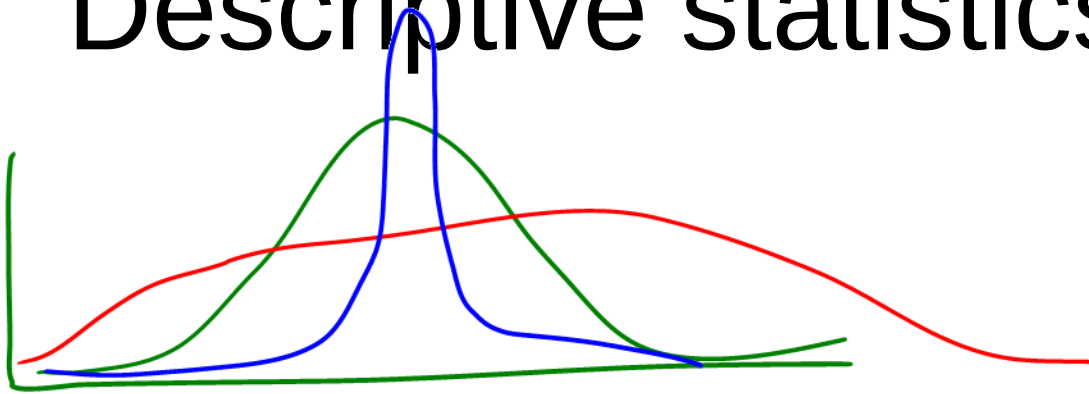
Random variable $X \Rightarrow$  *Gaussian*

mean $X = E[X] =$

$$\int_{-\infty}^{\infty} t \underbrace{pdf(t)}_N dt \quad \text{or} \quad \int_{-\infty}^{\infty} t \cdot \frac{1}{N} \cdot \sum_{i=1}^N \delta(t - x_i) dt$$

$$E[X] = \frac{1}{N} \sum_{i=1}^N X_i$$

Descriptive statistics: Variance



$$\text{Var}(x) = E[(X - E[X])^2]$$

$$= \frac{1}{n-1} \sum_{i=1}^n (x_i - E[X])^2$$

$$\text{Std}(x) = \sqrt{\text{Var}(x)}$$

Difference	$E[x]$	$E_s[x]$
	\uparrow	\uparrow
	True	sample mean

$$\bar{E}(x) - E_g(x) \approx \frac{\sqrt{\text{Var}(x)}}{\sqrt{N}}$$

error

$$\frac{1}{2}$$
$$\frac{1}{10}$$

Notg

4x

100X

Descriptive statistics: Covariance and Correlation

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Linearly predictable

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{std}(X) \text{std}(Y)}$$

↑
Correlation

Let's see it in python

