MATH:1260 Pokémath
The Mathematics of Pokémon Go®

Week 10 Wednesday, Spring 24

Popular curve:

Slugma-like curve

Plan for Today

- Module 2: Gotta Catch 'em All®
  - Statistics and Module 2 Wrap-up

Class Reminders

- Project 2 work day on Thursday.
  - Earn up to 3 points extra credit.
- Project 2 Stage 2: Data due tonight at Midnight.
- Project 2 Stage 3: Report due Wednesday April 3 at Midnight.

Statistics

Statistics is the study of data.

Two main types: Descriptive and Inferential.

Descriptive: Organize, Display, Describe a data set.

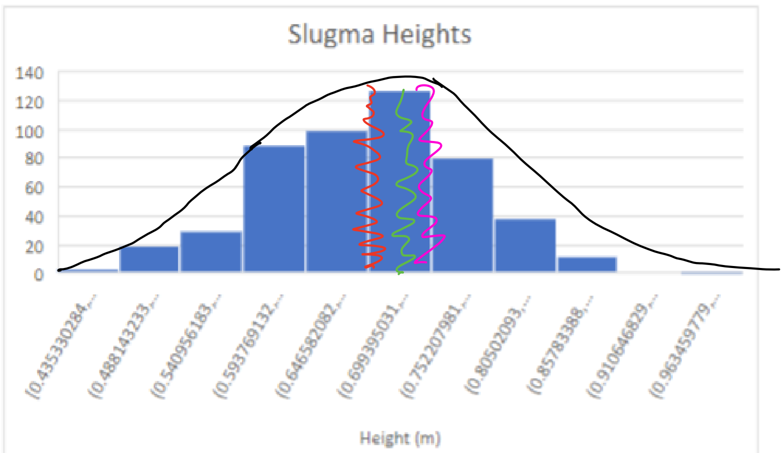
Inferential: Draw conclusions about an underlying population using a data set.

Recall data we recorded for Pikipek for GW.

We could compute the mean of the number of balls to catch. That would be descriptive.
We could use the data to estimate the underlying catch probability. That would be inferential.

Slugma

Here is a data set of the heights of many, many Slugma, collected by TAs over the years.



Mean for this data set of 500 Slugma is 0.699m.

middle data value

What do you notice about the shape of the curve? The median and mode?

most common data value

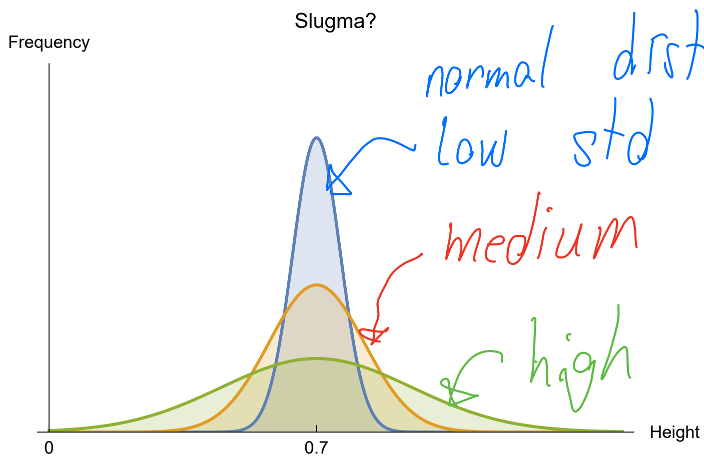
Normal Distribution

The height of each Pokemon is chosen from a Normal Distribution.

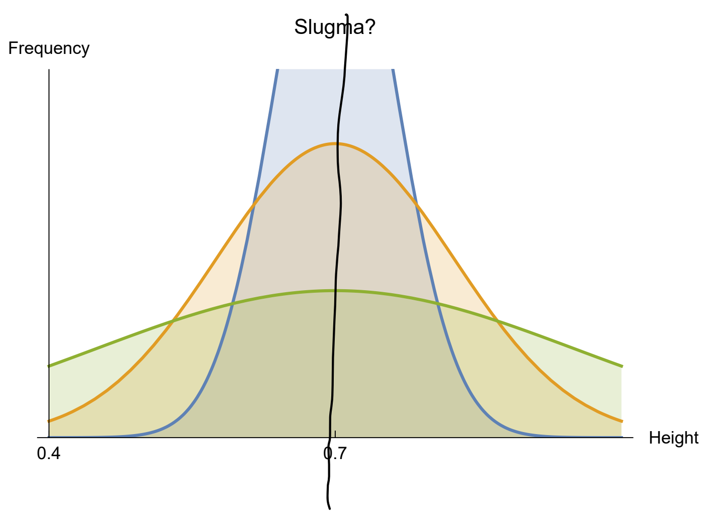
The means are given in this chart by species
https://pokemondb.net/pokedex/stats/height-weight

					height (m)			
	218	Slugma	FIRE	2'04"	0.7	77.2	35.0	71.4

all symmetric



normal distribution
low std deviation
medium std deviation
high std deviation



The normal curve is symmetric about the mean or average

A more manageable set of Slugma

Slugma	Height (m)
1	0.66
2	0.69
3	0.78
4	0.63
5	0.66

TopHat: What is the average height of these Slugma?

$$\frac{.66 + .69 + .78 + .63 + .66}{5} = .684$$

mean = .684

We can measure the “spread” of the data

We use “deviation from mean”. How did I calculate these? mean − height

Slugma	Height (m)	Deviation from mean
1	0.66	−.024
2	0.69	.006
3	0.78	.096
4	0.63	−.054
5	0.66	.684 − .66 = −.024

Negatives can be hard to work with, so we square the deviation

This makes the “distance from the mean” always positive

Slugma	Height (m)	Deviation from Mean	Square of Deviation
1	0.66	−.024	.000576
2	0.69	.006	.000036
3	0.78	.096	.009216
4	0.63	−.054	.002916
5	0.66	−.024	.000576

Variance

Population Variance is the average of the square of deviation.

Notation $\boxed{\sigma^2}$ sigma squared

$\sigma^2 = .002664$

Full Formulas

Population Variance

of data x_1, x_2, x_3 ... x_n with mean $\overline{x}$ is

$$\sigma^2 = \frac{(x_1-\overline{x})^2 + (x_2-\overline{x})^2 + (x_3-\overline{x})^2 + \ldots + (x_n-\overline{x})^2}{n}$$

Formula for Standard Deviation

Population Standard Deviation (STDEV in excel)
of data x_1, x_2, x_3 ... x_n with mean $\overline{x}$ is

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{(x_1-\overline{x})^2 + (x_2-\overline{x})^2 + (x_3-\overline{x})^2 + \ldots + (x_n-\overline{x})^2}{n}}$$

Notice how we square all the deviations from the mean, and then take the huge square root at the end. Do these cancel out? No!

$$\sqrt{2^2+3^2} = \sqrt{4+9} = \sqrt{13}$$

$$2+3 = 5 \neq \sqrt{13}$$

Formula for Correlation

Correlation (CORREL in excel)

of data $x_1, x_2, x_3 \ldots x_n$ with mean $\overline{x}$ and data $y_1, y_2, y_3 \ldots y_n$ with mean $\overline{y}$

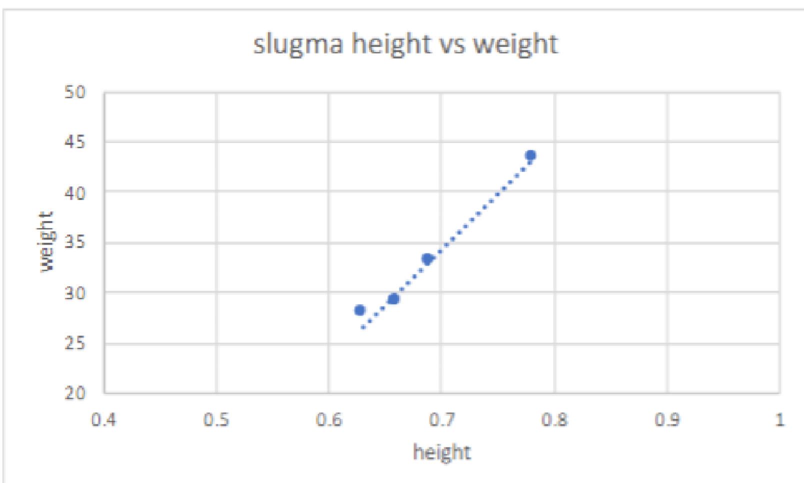
$$r = \frac{(x_1-\overline{x})(y_1-\overline{y}) + (x_2-\overline{x})(y_2-\overline{y}) + (x_3-\overline{x})(y_3-\overline{y}) + \ldots + (x_n-\overline{x})(y_n-\overline{y})}{n\,\sigma_x\,\sigma_y}$$

Formula for Correlation

Now let’s compute a correlation for height (h_1, h_2, ... h_5 with mean $\overline{h}$ and stdev s_h) and weight (w_1, w_2, ... w_5 with mean $\overline{w}$ and stdev s_w)

$$r = \frac{(h_1-\overline{h})(w_1-\overline{w}) + (h_2-\overline{h})(w_2-\overline{w}) + (h_3-\overline{h})(w_3-\overline{w}) + (h_4-\overline{h})(w_4-\overline{w}) + (h_5-\overline{h})(w_5-\overline{w})}{(n-1)\, s_h\, s_w}$$

Slugma	Height (m)	Weight (kg)
1	0.66	29.71
2	0.69	33.12
3	0.78	43.42
4	0.63	28.10
5	0.66	29.16



r = .991

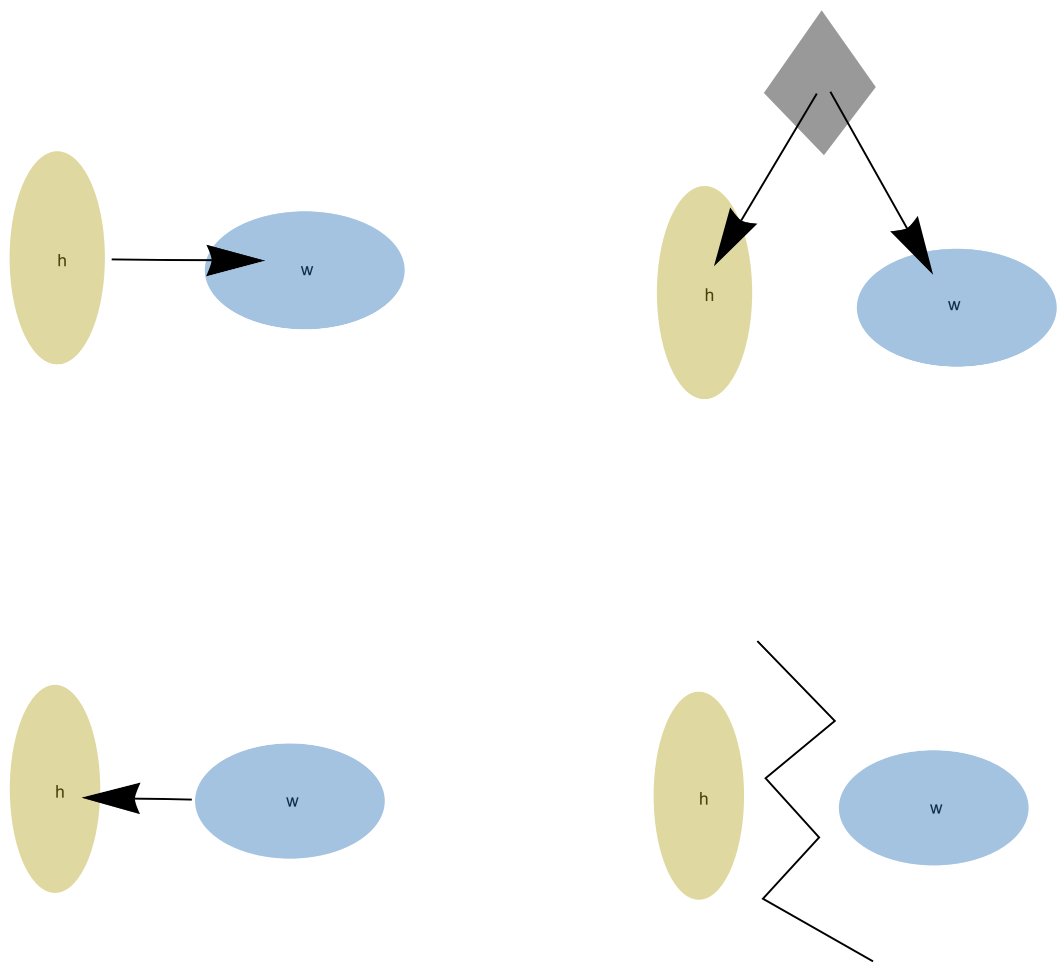
Correlation is NOT Causation

"Permitting statistical treatment and the hypnotic presence of numbers and decimal points to befog causal relationships is little better than superstition. And it is often more seriously misleading."

--Darrell Huff. *How to Lie with Statistics.*

Be careful!

Correlation can not distinguish between these cases...



Some funny examples of spurious correlations
http://tylervigen.com/spurious-correlations