if you have HW5 questions — I have time set aside at the beginning for these

Bias, estimators, bessel's correction, Bernoulli trials

Is the following problem best modeled with a binomial distribution or a poisson distribution?

↳ poisson → time, # of occurrences

I want to know what the probability is that I will observe four cars honking on my 7 minute walk to the T stop in the morning while I go to work.

↳ λ a prob. of hearing a car honk in a 7 min. period

Admin: ICAs

- Based on the feedback from your Canvas ICAs....
  - most of you either felt *shrug* or better about the Canvas ICAs than the gradescope ICAs
  - I'll be continuing to stop class 10 minutes early and have you all do the Canvas ICAs

Admin: Tests

- Test 1: if you haven't checked your grade **please do so now**. There was a segmenting issue that caused some tests to get ignored in gradescope—please send me an email ASAP if you have no grade for *either* quiz1_01 or quiz1_02

HW 5: clarifications

Section 1 - Late day count on Canvas - up to date

- Remember—we gave you an extra late day to account for the late release of HW 5—and so that we could have class together between the release and the due date.

  - I'm going over whatever clarifications you'd like for HW 5 now

  - You won't get credit for unused late days at the end of the semester

- Linearity of expectation/variance

  - holds for independent variables - both $E[X]$ and $Var(X)$

    $E[X+Y] = E[X] + E[Y]$ $\quad Var(X+Y) = Var(X) + Var(Y)$

  - for dependent variables, holds for expectation but not variance

Expected Value - from lec 13

What is the expected value of the total number of donuts that Felix eats in a given day?

$$E[X] = E[D_m + D_a] = 0 * P(X=0) + 1 * P(X=1) + 2 * P(X=2) + 3 * P(X=3) + 4 * P(X=4)$$

$$E[X] = E[D_m + D_a] = 0 * 0.5 + 1 * 0 + 2 * 0.5 + 3 * 0 + 4 * 0 = 1$$

What is the expected value of the total number of donuts that 3 Felixes eat in a given day?

$$E[3X] = E[X] + E[X] + E[X]$$

$= 3$

D_m

D_a

	donuts = 0	donuts = 1	donuts = 2
donuts = 0	0.5	0	0.05
donuts = 1	0	0.25	0
donuts = 2	0.2	0	0

Expected Value - from lec 13

What about the variance?

$$Var(X) = \sum_x P(X = x) * (x - E[X])^2$$

$Var(D_m + D_a) = \cancel{Var(D_m) + Var(D_a)}$

or, $Var(X) = E[X^2] - E[X]^2$

$Var(D_m + D_a) = E[X^2] - E[X]^2 = (0^2 * 0.5 + 1^2 * 0 + 2^2 * 0.5 + 3^2 * 0 + 4^2 * 0) - 1^2 = 2 - 1 = 1$

What about the variance of the total number of donuts that 3 Felixes eat in a given day?

$\underline{Var(3X) \neq Var(X) + Var(X) + Var(X)}$ (for dependent variables, this does not hold)

BUT! we are multiplying by a constant....

$Var(cX) = c^2 Var(X) = 9 * 1 = 9$ → means that the expected spread is much larger

↳ for all variables

Problem 2, part 6:

prob that the teams make the same # of goals

1) model an individual teams goals

2) $P(\text{team A} = \_\_\_, \text{team B} = \_\_\_) =$

$P(\text{die 1} = 6, \text{die 2} = 6) = P(\text{die 1} = 6) P(\text{die 2} = 2)$

How to account for different orderings

↳ look at distr., one of these does that for you

Add variance w/ constant ind. R.V. ex.

X and Y are coins w/ equal prob. distributions

$$E[X+Y] = E[X] + E[Y] = 0.5 + 0.5 = 1$$

$$E[3(X+Y)] = 3(E[X] + E[Y]) = 3$$

so $E[cX] = cE[X]$ where c is a constant

$$Var(X+Y) = E[(X+Y)^2] - E[X+Y]^2$$

$$Var(c(X+Y)) = E[(c(X+Y))^2] - E[c(X+Y)]^2$$

$$
\begin{aligned}
\mathrm{Var}(c(X+Y)) &= E\left[c^2(X+Y)^2\right] - E\left[c(X+Y)\right]^2 \\
&= c^2 E\left[(X+Y)^2\right] - \left(c E\left[X+Y\right]\right)^2 \\
&= c^2\left(E\left[(X+Y)^2\right] - E\left[X+Y\right]^2\right) \\
&= c^2 \,\mathrm{Var}(X+Y)
\end{aligned}
$$

for both independent and dependent variables
(we only need linearity of expectation)

Expectation/Variance/Standard Deviation

- **Expected value** is the average value that I expect to see after infinite random trials. written as $E[X]$.

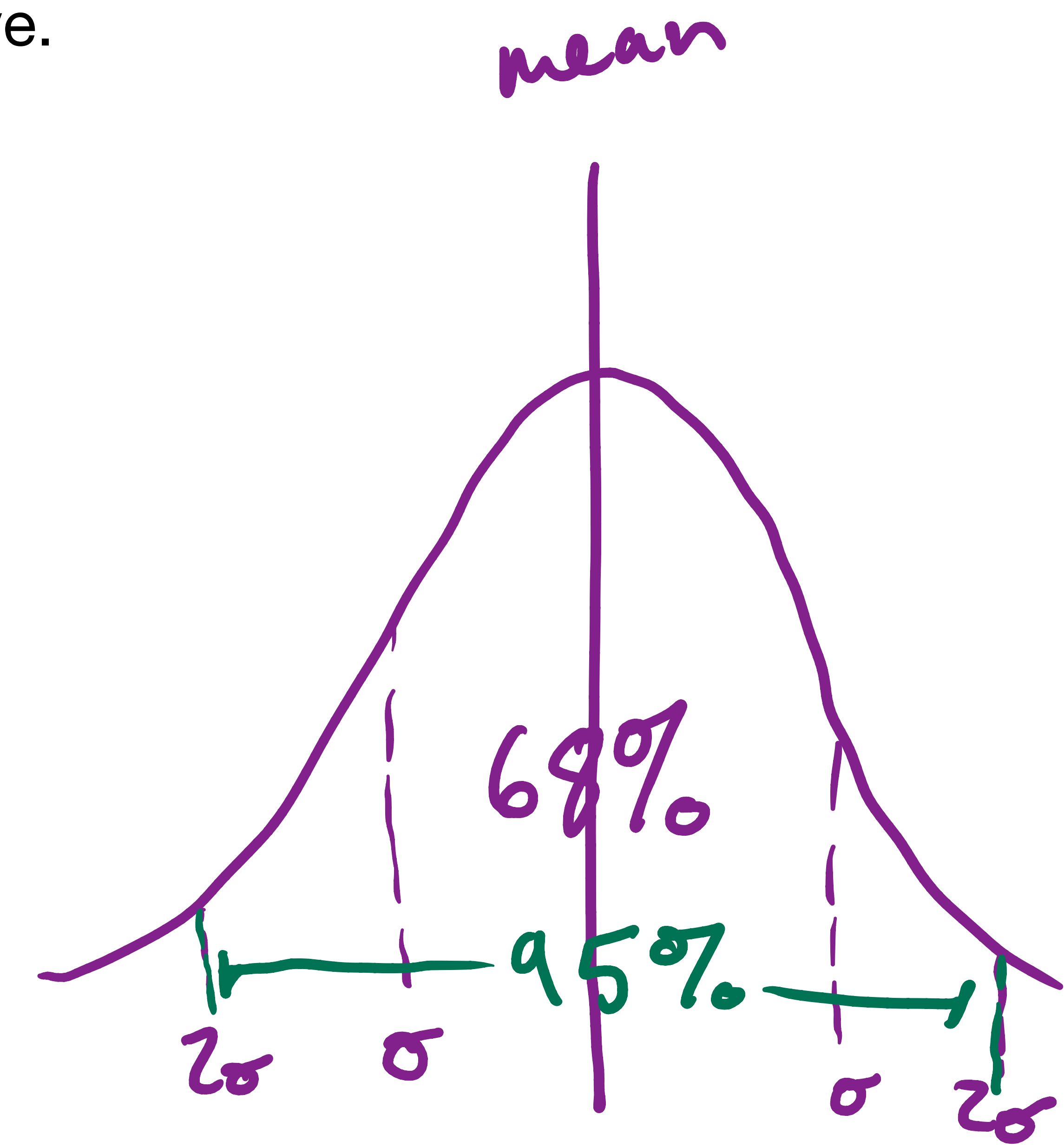
- **Variance** is a measure of how far actual values are from their expected value. (Their "variability"). written as $Var(X)$ or σ^2.

- **Standard deviation** is also a measure of the variation of a set of values. It has special properties when we are talking about normal distributions. written as σ.

$Var(X) = \sigma^2$ $\quad$ $\text{std dev}(X) = \sqrt{Var(X)}$

Normal Distributions

- A **normal** distribution is a bell-shaped curve.
- Right now, we're just using it to ground σ.



Binomial distributions - ICA Question 1

What does a binomial distribution model?

↳ two outcomes

↳ fixed probability

↳ independent trials

- model likelihood of some # of successes

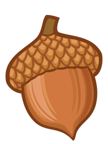
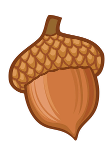
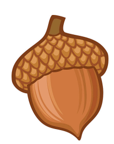
# of trials

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

# of successes

prob. of success

What should the value of p be if we want to model the chances of a squirrel finding a nut that it buried last fall on any given day if we observe:

$$P(S_9 = k) = \binom{n}{k} .6^k (.4)^{n-k}$$

$$P(S_q = 1) = \binom{1}{1} .6 (.4)^0$$
$$= .6$$

$$P(S_q = 1) = \binom{365}{1} .6^1 (.4)^{364}$$

Poisson distributions - ICA Question 2

$E[X] = Var(X) = \lambda$

What does a Poisson distribution model?

↳ over a fixed time period

↳ # of occurrences → successes

↳ events are independent

↳ two options

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

What should the value of λ be if we want to model the number of turkeys on the green line tracks on a given morning if we observe:

Mon - 3 turkeys

Tuesday - 5 turkeys

Wednesday - 0 turkeys

↳ between 8am + 12pm

$$\lambda = \frac{8}{3} = 2.667$$

λ time period needs to match your model

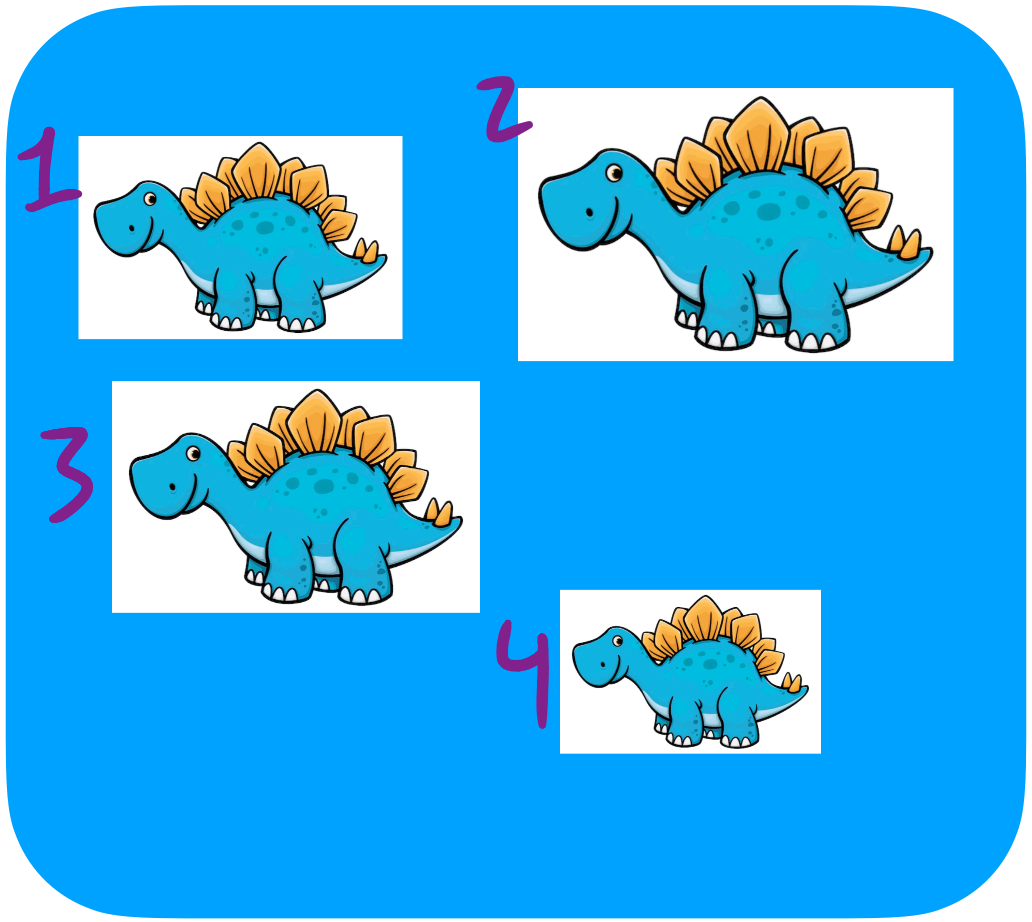
$P(X=2)$ for 8am - 9am, hourly lambda

$P(X=2)$ for 8am - 12Pm, 4-hour lambda

Observing data

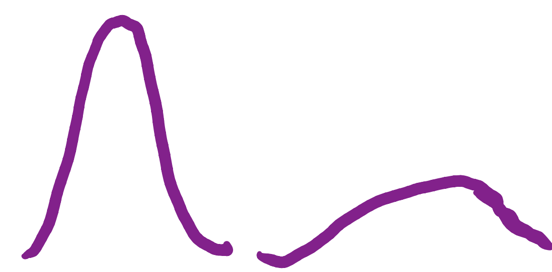
- Say that we're observing some data:

$$\bar{x} = \frac{(7.2+9+7.5+6)}{4}$$



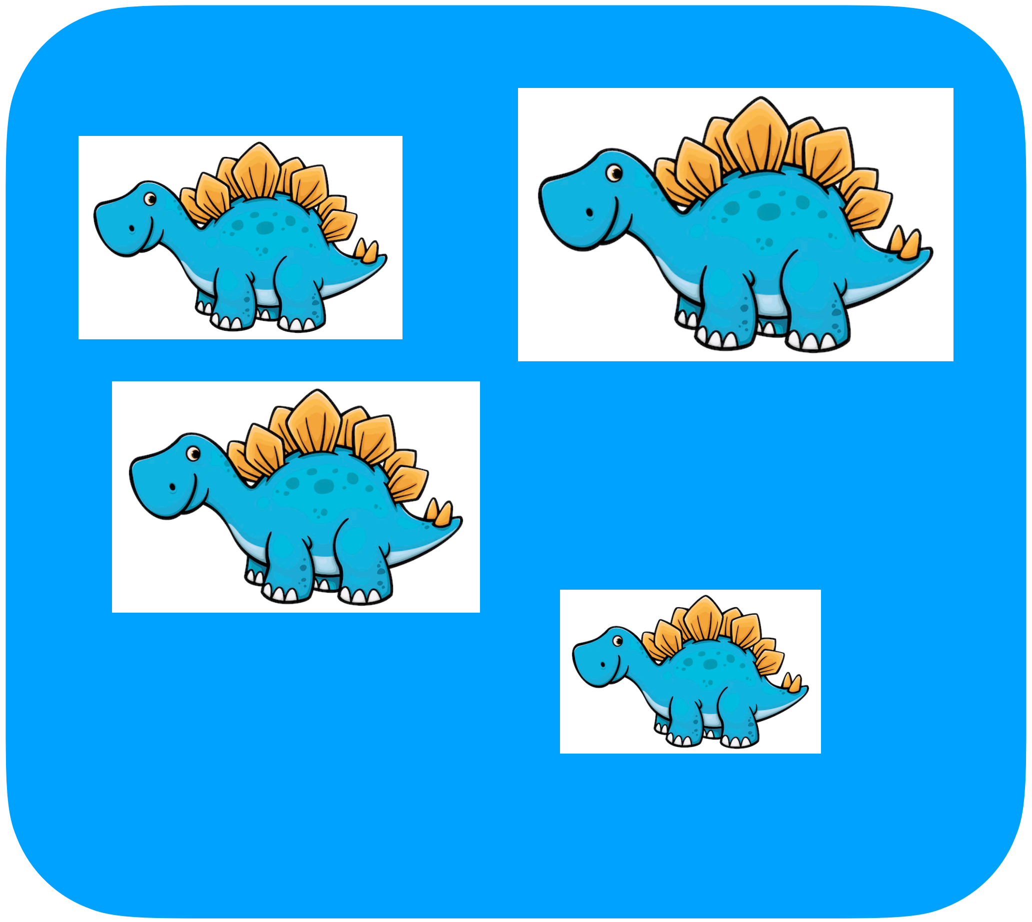
Observed Stego length
7.2
9
7.5
6

$$\bar{x} = 7.425$$



Observing data - estimators

- We'd like to know what the variance of the **population** is, but we only have our observed sample to work from.

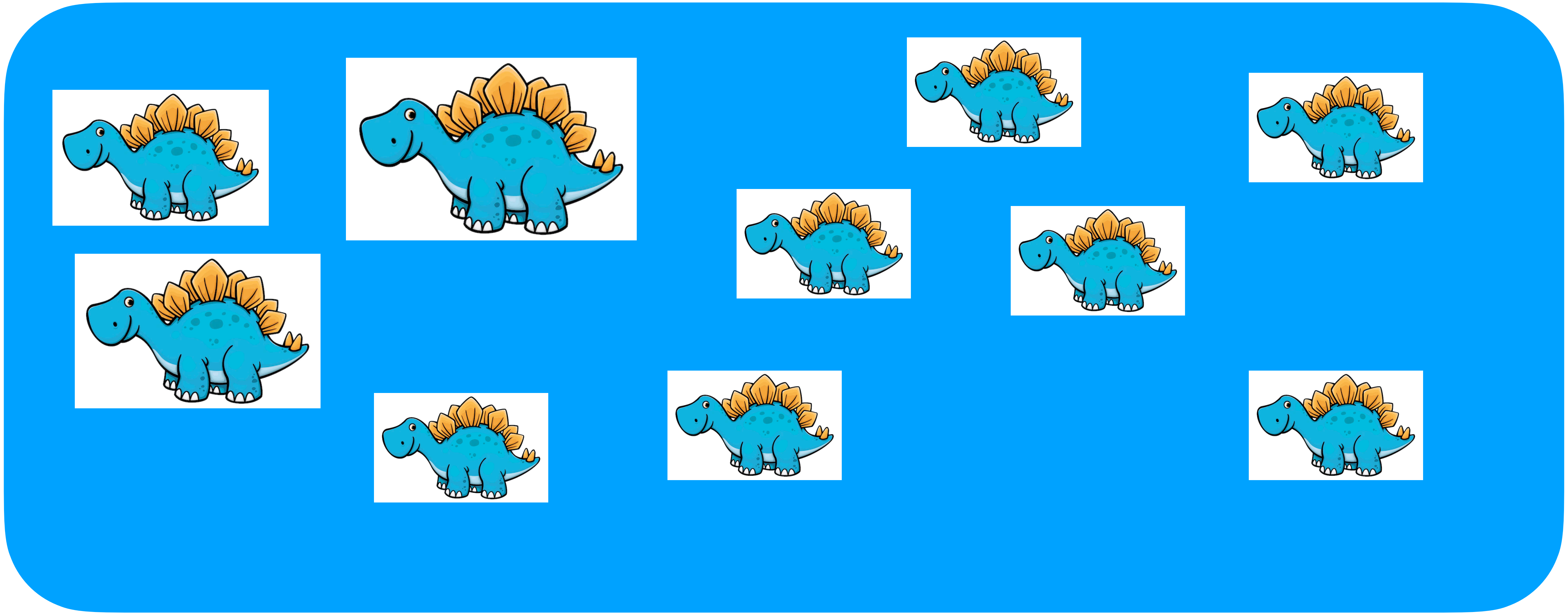


Observed Stego length
7.2
9
7.5
6

"how far from $\bar{X}$ are each of our observed X_i?"

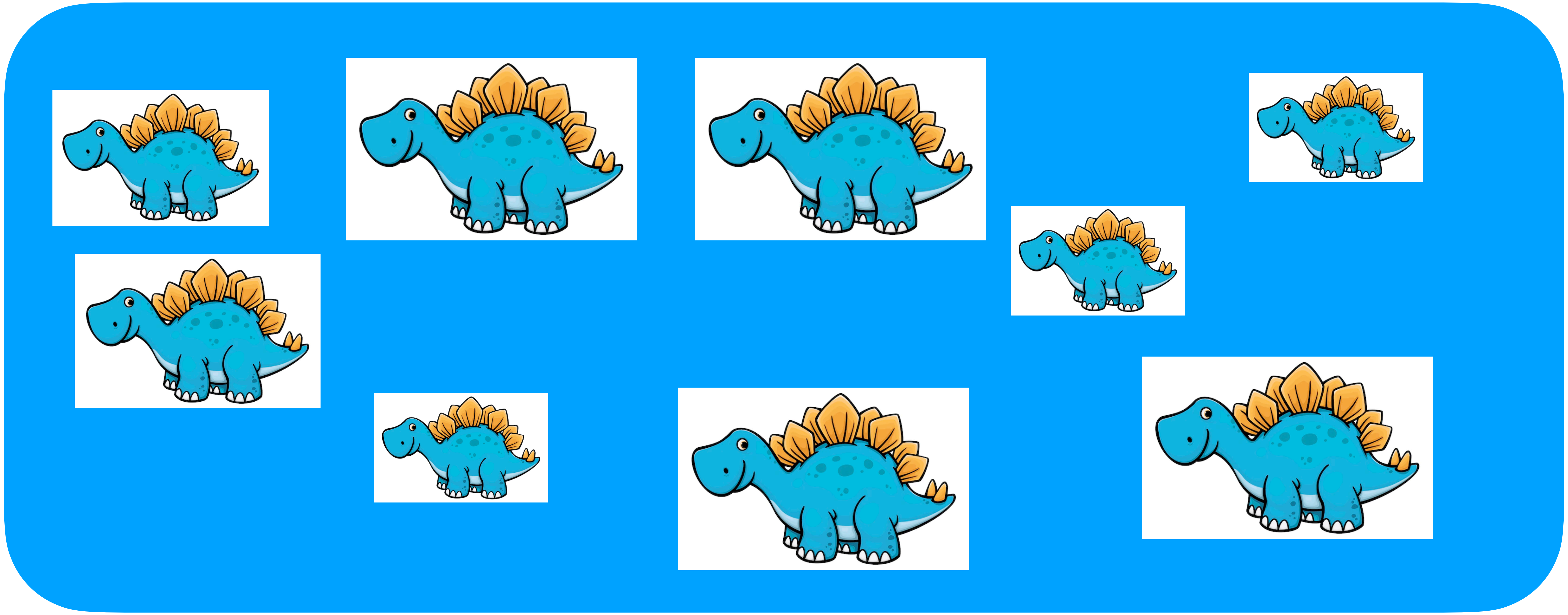
Observing data - estimators

- We'd like to know what the variance of the **population** is, but we only have our observed sample to work from.



Observing data - estimators

- We'd like to know what the variance of the **population** is, but we only have our observed sample to work from.



Observing data - estimators

- Working with what we have, we'll calculate an estimate of the **population** variance σ^2_{mean} by calculating "how far from the **sample** mean is each individual?"
- $$\sigma^2_{mean} = \frac{1}{N}\sum_i (x_i - \bar{x})^2$$

$$\frac{1}{4}\left((7.2-7.425)^2 + \ldots + (6-7.425)^2\right) = 1.14$$

Observed Stego length
7.2
9
7.5
6

Observing data - estimators

- But, is this estimate **biased**?
  - In statistics, **bias** is the difference in the value of an ~~expected~~ estimated parameter (like σ^2) and the real, ground truth value

    σ^2

- This was our estimate:

- $\sigma^2_{mean} = \dfrac{1}{N}\sum_i (x_i - \bar{x})^2$

- $\dfrac{1}{4}((7.2 - 7.425)^2 + (9 - 7.425)^2 + (7.5 - 7.425)^2 + (6 - 7.425)^2) = 1.14$

  $\sigma^2 \quad \sigma^2 \quad \sigma^2 \quad \cancel{\sigma^2}$

we've run out of degrees of freedom

Observing data - estimators

- This was our estimate:
- $\sigma^2_{mean} = \frac{1}{N}\sum_i (x_i - \bar{x})^2$
- $\frac{1}{4}((7.2 - 7.425)^2 + (9 - 7.425)^2 + (7.5 - 7.425)^2 + (6 - 7.425)^2) = 1.14$
- So the big question is $\sigma^2_{mean} = \sigma^2$?

No!

$$\underbrace{\sigma^2_{mean} = \frac{N-1}{N}\sigma^2}$$

Bessel's correction

- Our estimate was **biased** because $\sigma^2_{mean} \neq \sigma^2$
- Re-writing our estimate as:
  - $\sigma^2_{mean} = \dfrac{N-1}{N}\sigma^2$
- We can now **correct** for the bias by multiplying our calculable but biased estimator by $\dfrac{N}{N-1}$
- $\sigma^2_{bessel} = \dfrac{\cancel{N}}{\cancel{N-1}}\overbrace{\left(\dfrac{\cancel{N-1}}{\cancel{N}}\sigma^2\right)}^{\sigma^2_{mean}}$
- So the big question is $\sigma^2_{bessel} = \sigma^2$? Yes!

Bessel's correction

- Mathematically, $\sigma^2_{bessel} = \sigma^2$?
- Good news! We have an easier way to write down this correction:
- $\sigma^2_{bessel} = \left(\frac{1}{N-1}\right)\sum_i (x_i - \bar{x})^2$

σ^2_{mean} $\sigma^2_{\bar{x}}$

Bessel's Correction - ICA Question 3

What is the value of σ^2_{bessel} for our population of stegosauruses given that $\bar{x} = 7.425$?

7.2

9

7.5

6

$$\sigma^2_{bessel} = \frac{1}{N-1}\sum_i (x_i - \bar{x})^2$$

Do we expect this to be larger or smaller than σ^2_{mean}?

↳ larger

$\sigma^2_{bessel} = 1.5225$

Observing data - sampling a population

- Finally, we observe **12 more stegosauruses**

- How do we expect $\bar{x}$ and σ^2_{bessel} to act in relation to the underlying ground truth?

↳ should get closer to the ground truth

↳ $\bar{x}$, σ^2_{bessel}

↳ law of large #s

↳ but! 12 is not so many → 1000

more is better

Bessel's Correction - ICA Question 4

Say that the underlying ground truth $E[X]$ is revealed to you. Is your calculated variance with the true expected value going to be closer to σ^2_{bessel} or σ^2_{mean}? Do we have any guarantees?

→ true, unbiased

"biased" → mathematically

$\sigma^2_{mean} \neq \sigma^2$

A. σ^2_{bessel}

B. σ^2_{mean}

C. no guarantee

Break time!

back at 1:10

- We'll break for 13 minutes
  - stretch, get water, etc
  - Fill out the ICA quiz on canvas. Passcode: "water"
  - (you can ask me any other questions you have as well)

Bernoulli Trials

- What is a Bernoulli trial?
- !vocab alert! Bernoulli trials are also called **binomial** trials
- Independent, repeated trials of an experiment with exactly two outcomes.

  ↳ flipping a coin many times

- Wait, how is this different than a **binomial distribution**?

  ↳ "creates" the exp. data underlying a binomial distribution

Binomial distributions

- For a binomial distribution, we assume:
  - Each underlying Bernoulli trial is **independent**
  - Each event being summed is in fact a **Bernoulli trial**
  - Each Bernoulli trial is **identically distributed**

↳ p is the same across trials

Bernoulli experiments

- A **Bernoulli experiment** is modeled as $B(n, p)$ where n is the number of trials and p is the probability of success.

- $$B(n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

- This Bernoulli experiment produces a binomial dist. that we can evaluate to answer the same questions that we were answering last lecture

Bernoulli experiments

t1

t2

$\frac{6 \text{ happy}}{10 \text{ total}}$

- So, we want to estimate n and p so that we can evaluate this to answer questions like "what is the probability that two turtles are happy on one day?"

$$B(2, 0.6) = \binom{2}{k} 0.6^{k} (1-0.6)^{2-k}$$

$$B(2, 0.6) = P(X=1) = \binom{2}{1} 0.6^1 \, 0.4^1$$

↳ means what are the chances of 1 turtle out of 2 being happy

Admin

⚠ we are different ⚠

- On Thursday:
  - I'll be releasing your mini-project description
  - You'll have a combined HW 6 workshop + mini-project workshop
  - We do expect you to attend in person if you are able to
  - If you need to dial in, do expect to be showing us your work :)
  - You *will* get ICA credit for attending and working diligently

Schedule

Turn in ICA 14 on Canvas (make sure that this is submitted by 2pm!)

HW 5's final due date is on Tuesday

HW 6 will be released:

HW 6's official due date is March 20th. **Come to class on Thursday so you don't have to work over break.**

Mon	Tue	Wed	Thu	Fri	S	Sun
Lecture 14 - estimators, bias **HW 5 due @ 11:59pm**	**Felix OH Calendly**	**Felix OH Calendly**	**Felix OH Calendly** Lecture 15 - mini-project descriptions, HW 6 work day (yes, you will get ICA credit for this day)			
SPRING BREAK						**HW 6 due @ 11:59pm**
Lecture 16 - normal distributions	**Felix OH Calendly**	**Felix OH Calendly**	**Felix OH Calendly** Lecture 17 - hypothesis testing			

More recommended resources on these topics

- YouTube: 3Blue1Brown Binomial distributions | Probabilities of probabilities, part 1 (still useful!)
- YouTube: Ben Lambert Estimating the population variance from a sample - part one*
    - Warning: this uses some notation that we haven't covered in class
- Wikipedia Bessel's Correction: https://en.wikipedia.org/wiki/Bessel%27s_correction