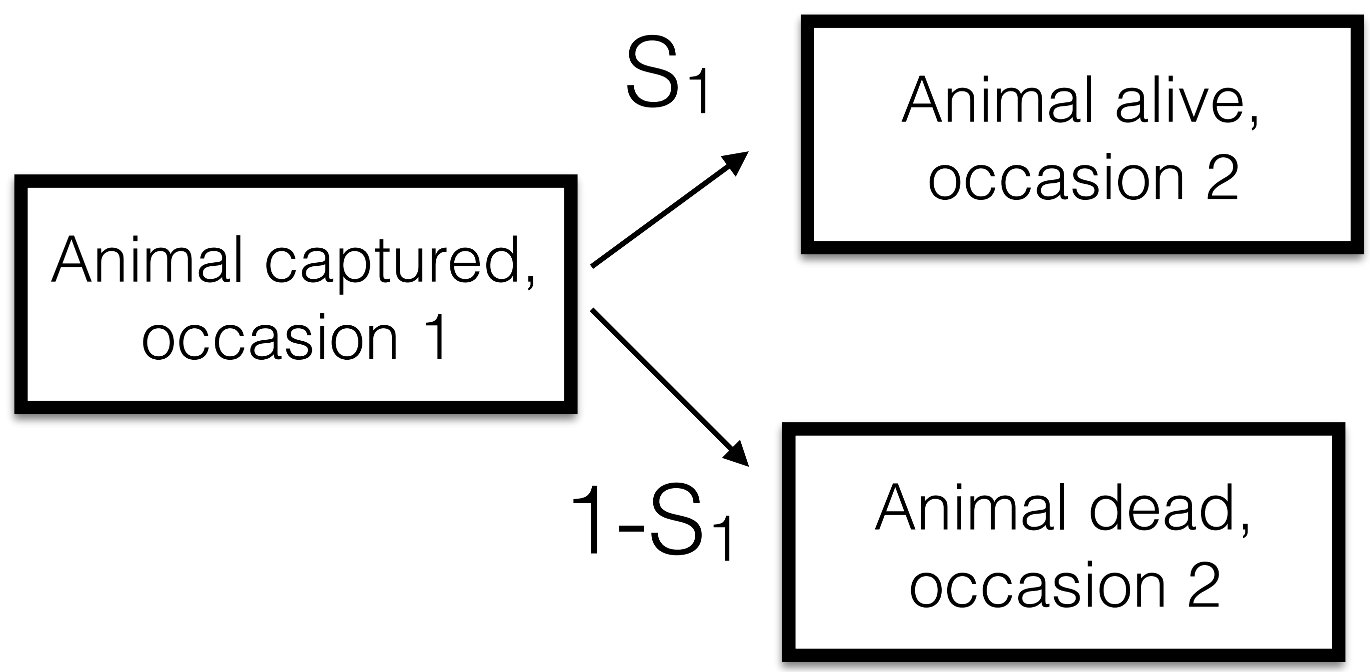
Applied population analysis, week 3

WLF 504, Fall 2019

Known fates data

- When we know the fates of all our animals…

- We can eliminate the detection probability (p), woo! The probabilities are so much simpler.



Known fates data

Each tagged animal either:

- Survives to end of study period (detected at each sampling occasion so fate is known for each occasion)
- Dies during study (death site found on first occasion after death so its fate is known)
- Survives up to a point where its fate is last known, after which it is censored (fate is kinda known)

The Kaplan-Meier (K-M) Estimator

$$\hat{S}_t = \prod_{i=1}^{t}\left(\frac{n_i - d_i}{n_i}\right)$$

The estimate for survival to time t

The Kaplan-Meier (K-M) Estimator

$$\hat{S}_t = \prod_{i=1}^{t} \left(\frac{n_i - d_i}{n_i} \right)$$

The product from 1 to t

The Kaplan-Meier (K-M) Estimator

$$\hat{S}_t = \prod_{i=1}^{t} \left(\frac{n_i - d_i}{n_i} \right)$$

Number alive at start of interval i

The Kaplan-Meier (K-M) Estimator

$$\hat{S}_t = \prod_{i=1}^{t} \left(\frac{n_i - d_i}{n_i} \right)$$

Number that died during interval i

Example: black ducks

	survived to occasion							
week	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
number alive at start	48	47	45	39	34	28	25	24
number dying	1	2	2	5	4	3	1	0
number alive at end	47	45	39	34	28	25	24	24
number censored	0	0	4	0	2	0	0	0

	survived to occasion							
week	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
number alive at start	48	47	45	39	34	28	25	24
number dying	1	2	2	5	4	3	1	0
number alive at end	47	45	39	34	28	25	24	24
number censored	0	0	4	0	2	0	0	0

Censoring reduces "at risk" pool, temporarily or permanently

$$\hat{S}_1 = (47/48) = 0.979$$

$$\hat{S}_2 = (45/47) = 0.957$$

$$\hat{S}_3 = (39/41) = 0.951$$ (*note*: only 41 because 4 were censored)

$$\hat{S}_4 = (34/39) = 0.872$$

$$\hat{S}_5 = (28/32) = 0.875$$ (*note*: only 32 because 2 were censored)

$$\hat{S}_6 = (25/28) = 0.893$$

$$\hat{S}_7 = (24/25) = 0.960$$

$$\hat{S}_8 = (24/24) = 1.000$$

$\hat{S}_1 = (47/48) = 0.979$

$\hat{S}_2 = (45/47) = 0.957$

$\hat{S}_3 = (39/41) = 0.951$ (*note*: only 41 because 4 were censored)

$\hat{S}_4 = (34/39) = 0.872$

$\hat{S}_5 = (28/32) = 0.875$ (*note*: only 32 because 2 were censored)

$\hat{S}_6 = (25/28) = 0.893$

$\hat{S}_7 = (24/25) = 0.960$

$\hat{S}_8 = (24/24) = 1.000$

- Full St model has 8 parameters, so 8 Si estimates
- Could fit simpler models, for example if S is constant
- Or if S is affected by a covariate
- Or a smooth function (S declines/increases w/time)

The binomial model

$$\hat{S}_t = \prod_{i=1}^{t}\left(\frac{n_i - d_i}{n_i}\right)$$

Kaplan-Meier estimator

$$\hat{S}_i = \frac{y_i}{n_i},$$

Binomial estimator

$$\hat{S}_t = \prod_{i=1}^{t}\left(\frac{n_i - d_i}{n_i}\right) = \prod_{i=1}^{t}\left(\frac{y_i}{n_i}\right).$$

St = KM estimate = a series of binomial estimates

The binomial model

$$\hat{S}_t = \prod_{i=1}^{t}\left(\frac{n_i - d_i}{n_i}\right) = \prod_{i=1}^{t}\left(\frac{y_i}{n_i}\right).$$

$$\mathcal{L}(\theta \mid n_i, y_i) = \prod_{i=1}^{t} S_i^{y_i}(1 - S_i)^{(n_i - y_i)}$$

θ is the survival model parameter for the t intervals

The binomial model

$$\mathcal{L}(\theta \mid n_i, y_i) = \prod_{i=1}^{t} S_i^{y_i}(1 - S_i)^{(n_i - y_i)}$$

$$\mathcal{L}(S_1 \mid n_1, n_2) = \binom{n_1}{n_2} S_1^{n_2}(1 - S_1)^{(n_1 - n_2)} = \binom{48}{47} S_1^{47}(1 - S_1)^{(48-47)}$$

	survived to occasion							
week	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
number alive at start	48	47	45	39	34	28	25	24
number dying	1	2	2	5	4	3	1	0
number alive at end	47	45	39	34	28	25	24	24
number censored	0	0	4	0	2	0	0	0

Known fates data

- In program mark/Rmark, we format known-fates data this way:
    - Follows a L/D (live/dead) format
    - 2 number to describe each interval:
    - First number is a 1 if animal is known to be alive at start of interval j
    - Second number is a 1 if animal is known to be dead at start of interval j

Known fates data

Type of model	Encounter history (CH)	Probability
Known Fates	10 10 10 10 10 10 11 00 10 11 00 00 11 00 00 00 10 00 00 10 00 00 10 11 10 00 00 00	$S_1S_2S_3S_4$ $S_1S_2(1-S_3)$ $S_1(1–S_2)$ $(1–S_1)$ S_1S_4 $S_3(1–S_4)$ S_1

Known fates data

Type of model	Encounter history (CH)	Probability
Known Fates	10 10 10 10	$S_1S_2S_3S_4$
	10 10 11 00	$S_1S_2(1-S_3)$
	10 11 00 00	$S_1(1–S_2)$
	11 00 00 00	$(1–S_1)$
	10 00 00 10	S_1S_4
	00 00 10 11	$S_3(1–S_4)$
	10 00 00 00	S_1

Tagged at start of 1st interval, survives the entire study (4 intervals)

Known fates data

Type of model	Encounter history (CH)	Probability
Known Fates	10 10 10 10 10 10 11 00 10 11 00 00 11 00 00 00 10 00 00 10 00 00 10 11 10 00 00 00	$S_1S_2S_3S_4$ $S_1S_2(1\text{-}S_3)$ $S_1(1–S_2)$ $(1–S_1)$ S_1S_4 $S_3(1–S_4)$ S_1

Tagged at start of 1st interval, dies during 3rd

Known fates data

Type of model	Encounter history (CH)	Probability
Known Fates	10 10 10 10 10 10 11 00 10 11 00 00 11 00 00 00 10 00 00 10 00 00 10 11 10 00 00 00	$S_1S_2S_3S_4$ $S_1S_2(1\text{-}S_3)$ $S_1(1–S_2)$ $(1–S_1)$ S_1S_4 $S_3(1–S_4)$ S_1

Tagged at start of 1st interval, survives for 1 intervals, dies during 2nd

Known fates data

Type of model	Encounter history (CH)	Probability
Known Fates	10 10 10 10 10 10 11 00 10 11 00 00 11 00 00 00 10 00 00 10 00 00 10 11 10 00 00 00	$S_1S_2S_3S_4$ $S_1S_2(1\text{-}S_3)$ $S_1(1–S_2)$ $(1–S_1)$ S_1S_4 $S_3(1–S_4)$ S_1

Tagged at start of 1st interval, dies during first interval

Known fates data

Type of model	Encounter history (CH)	Probability
Known Fates	10 10 10 10 10 10 11 00 10 11 00 00 11 00 00 00 10 00 00 10 00 00 10 11 10 00 00 00	$S_1S_2S_3S_4$ $S_1S_2(1\text{-}S_3)$ $S_1(1–S_2)$ $(1–S_1)$ S_1S_4 $S_3(1–S_4)$ S_1

Tagged at start of 1st interval, **censored** (not known) for 2 intervals, survival 4th interval

Known fates data

Type of model	Encounter history (CH)	Probability
Known Fates	10 10 10 10 10 10 11 00 10 11 00 00 11 00 00 00 10 00 00 10 00 00 10 11 10 00 00 00	$S_1S_2S_3S_4$ $S_1S_2(1\text{-}S_3)$ $S_1(1–S_2)$ $(1–S_1)$ S_1S_4 $S_3(1–S_4)$ S_1

Tagged at start of 3rd interval, dies in 4th interval

Known fates data

Type of model	Encounter history (CH)	Probability
Known Fates	10 10 10 10 10 10 11 00 10 11 00 00 11 00 00 00 10 00 00 10 00 00 10 11 10 00 00 00	$S_1S_2S_3S_4$ $S_1S_2(1\text{-}S_3)$ $S_1(1–S_2)$ $(1–S_1)$ S_1S_4 $S_3(1–S_4)$ S_1

Tagged occasion 1, unknown fate, **censored**

How to calculate variance of the product of survivals?

- We often estimate survival (daily, weekly) in intervals. What if we want to know the mean and the variance of some longer time period, say a month or a year?
- We can multiply the survival rates together:
- And we can determine the variance…
- But what do we mean by “variance?”

Sampling vs. process variance

- Sampling variance = variance due to estimation of our parameter from sample data. Affected by var. among individuals and sample size.
- This sampling var. is part of what our stats packages give us.
- Process variance = variation in parameter over time and space (long term varation in weather, disease, etc.)
- Further discussion of process variance:
  - White et al. (1982), Franklin et al. (2000), MARK book.

Method 1: The “delta’ method

- Works best for functions that are largely linear (or transformations to linear via link functions)
- Variance of complex functions may not be well approximated by delta function….
- Which leads us to bootstrapping (in a bit)

The “delta’ method

We use a method from calculus, the “Taylor series expansion”, which lets us derive a linear function that approximates a more complex function.

—> Therefor, the measure of variance we derive is an approximation, not an estimate… estimates come from data, approximations come from math shortcuts.

The “delta’ method

$$\mathrm{var}(G) = \mathrm{var}[f(X_1, X_2, \ldots, X_n)] = \sum_{i=1}^{n} \mathrm{var}(X_i) \left[\frac{\partial f}{\partial X_i} \right]^2$$

The approximate variance of a function G, where the variance is a function of one or more random variables (for example S-hat)…

The "delta' method

$$\mathrm{var}(G) = \mathrm{var}[f(X_1, X_2, \ldots, X_n)] = \sum_{i=1}^{n} \mathrm{var}(X_i) \left[\frac{\partial f}{\partial X_i} \right]^2$$

Is equal to the sum of…

The “delta’ method

$$\mathrm{var}(G) = \mathrm{var}[f(X_1, X_2, \ldots, X_n)] = \sum_{i=1}^{n} \mathrm{var}(X_i) \left[\frac{\partial f}{\partial X_i} \right]^2$$

Each random variable Xi’s variance, times…

The "delta' method

$$\mathrm{var}(G) = \mathrm{var}[f(X_1, X_2, \ldots, X_n)] = \sum_{i=1}^{n} \mathrm{var}(X_i) \left[\frac{\partial f}{\partial X_i} \right]^2$$

The derivative of the complex function with respect to the random variable Xi, squared.

The “delta’ method

$$\mathrm{var}(G) = \mathrm{var}[f(X_1, X_2, \ldots, X_n)] = \sum_{i=1}^{n} \mathrm{var}(X_i) \left[\frac{\partial f}{\partial X_i} \right]^2$$

For example:

if $Y = X^3$, and we want var(Y) w/respect to X

d(Y w/respect to X) = $3X^2$

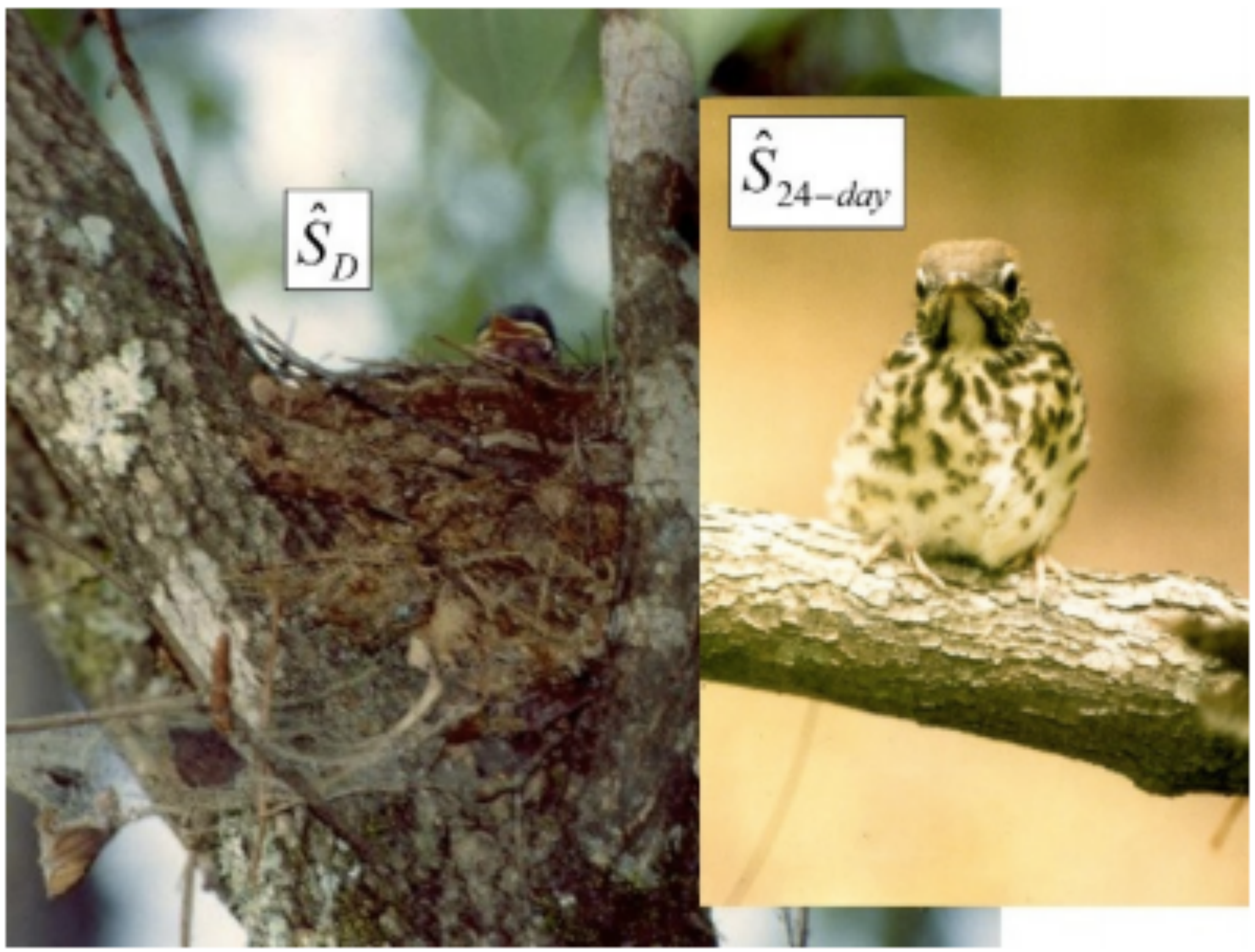
So var(Y) = var(X)*$(3X^2)^2$

Look up simple derivative rules when needed

Simple derivative rules

Function, generalized	Generalized derivative	Function, example	Derivative, example
c	0	2	0
x	1	S	1
cx	c	$2S$	2
x + c	x	$S + 2$	S
x^c	$cx^{(c-1)}$	S^2	$2S^1$
$\frac{x}{c}$	$\frac{1}{c}$	$\frac{S}{2}$	$\frac{1}{2}$
$\frac{c}{x}$	$-\frac{c}{x^2}$	$\frac{2}{S}$	$-\frac{2}{S^2}$

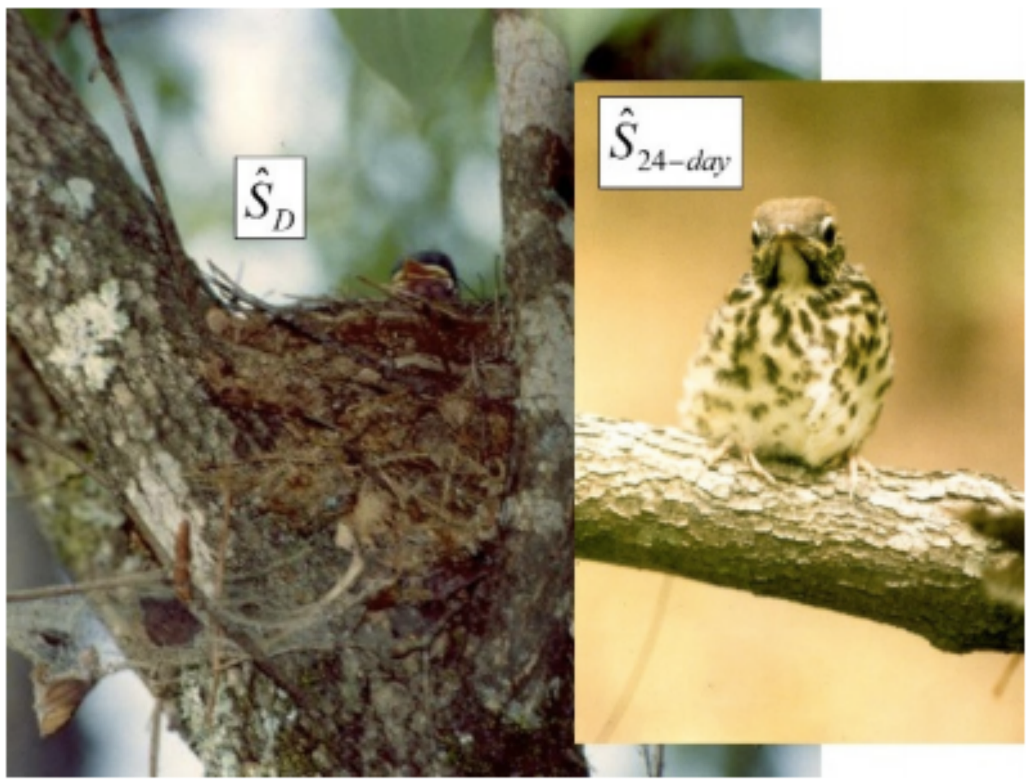
Worked delta method example… trust me it's not so bad!



Nest survival model:

- Successful nests found at higher prob. than unsuccessful nests
- Solution: estimate daily surv. rate (S_D)
- Use to calculate average S(t) until fledging
- $S(t=20) = S_D$^20

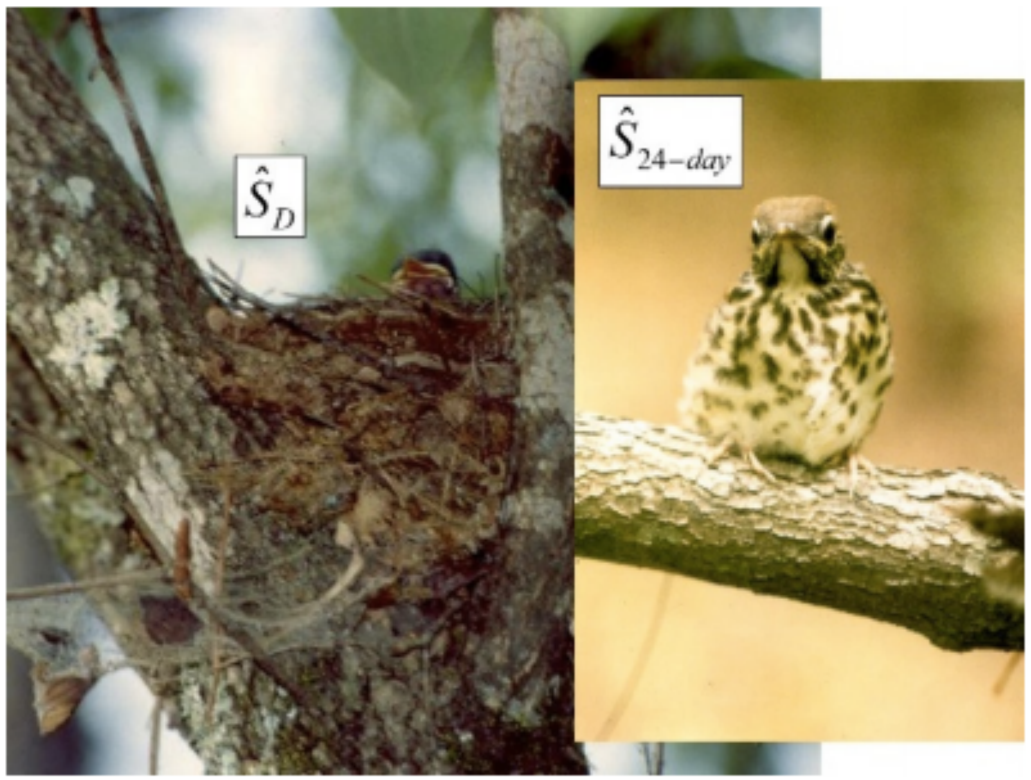
Worked delta method example… trust me it's not so bad!



$$\hat{S}_{24-day} = \hat{S}_D^{\ 24}$$

$$\hat{S}_{24-day} = 0.985^{24} = 0.696$$

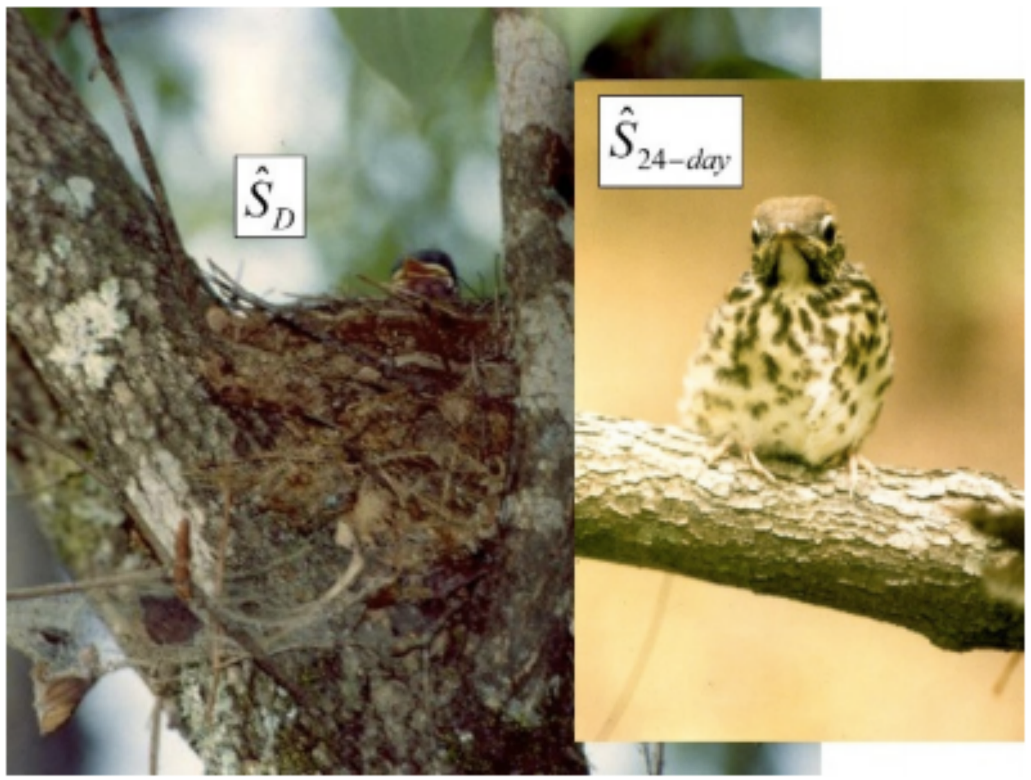
Worked delta method example… trust me it's not so bad!



$$\hat{\text{var}}(\hat{S}_{24-day}) = \hat{\text{var}}(\hat{S}_D)\left[\frac{\partial \hat{S}_{24-day}}{\partial \hat{S}_D}\right]^2$$

$$({S_D}^{24})' = 24{S_D}^{23}$$

Worked delta method example… trust me it's not so bad!



$$\hat{\text{var}}(\hat{S}_{24-day}) = \hat{\text{var}}(\hat{S}_D)(24\hat{S}_D^{\;23})^2$$

$$\hat{\text{var}}(\hat{S}_{24-day}) = 576\,\hat{\text{var}}(\hat{S}_D)\hat{S}_D^{\;46}$$

$$\hat{\text{var}}(\hat{S}_{24-day}) = 576 \cdot \left(0.023^2\right)0.985^{46} = 0.1520$$

$$S\hat{E}(\hat{S}_{24-day}) = \sqrt{0.1520} = 0.390$$

Method 2: Bootstrapping

- Re-sample our data (individuals) with replacement
- If we have groups, we must re-sample within groups so that number in each group remains the same
- Coding challenge this week: We will fit some known-fates models, then extend —> very simple bootstrapping functions “manually” and with the “boot” function in program R

Predictions using covariate models

- Extension x2 assignment to the coding challenge for this week, for those that get through bootstrapping challenge easily/quickly
- Using the observed range of a covariate, what Si do we predict? CI’s?
- Website URL w/link to example code for black duck data is in Coding Challenge handout.

Coding challenge

Further reading and code:

- http://www.phidot.org/software/mark/docs/book/pdf/chap16.pdf
- http://www.phidot.org/software/mark/docs/book/pdf/chap17.pdf
- https://rdrr.io/cran/RMark/man/mallard.html
- http://www.phidot.org/software/mark/docs/book/pdf/app_3.pdf
- https://sites.google.com/site/cmrsoftware/lecture-lab-schedule/9--known-fates-analysis