9/26/19

M156: Mortality Studies (on dragons) using Seriatim Data

Recall: X = rvr age at death

We want information on $Pr(X > z) = S_X(z)$

4 Cases

Case 1: We have complete individual data

↳ we know when each dragon dies

Example: X_i : 12 8 22 12 36

(Study on 5 dragons)

In this case, we approximate survival rate (probabilities) using the "empirical distribution". For this example, the empirical distribution is:

$\hat{X}$	Pr
8	1/5
12	2/5
22	1/5
36	1/5

made up value

Then we can approximate $\underbrace{Pr(X > 20)}_{S_X(20)} \approx \underbrace{Pr(\hat{X} > 20)}_{\hat{S}_5(20)} = \frac{2}{5}$

notation

$\hat{S}_5(20)$ ↳ # of dragons in study

$S_X(t) \approx \hat{S}_n(t)$

Remark: $\hat{S}_5(12) = \frac{2}{5} \approx Pr(X > 12) = S_X(12)$

Now consider a random sample of size 5.

Then $\underline{\underline{\hat{S}_5}}(12) = \frac{N}{5}$ where

$N =$ rvr the number of the 5 observations that are greater than 12.

The double underline distinguishes between the random variable (called the <u>estimator</u>) and the value (called the <u>estimate</u>)

<u>Remark</u>: $N \sim \text{Bin}(5, p)$ $\quad p = \Pr(X > 12)$

$$E[\underline{\underline{\hat{S}_5}}(12)] = E\left[\frac{N}{5}\right] = \frac{E[N]}{5} = \frac{5 \cdot p}{5} = p$$

$$Var(\underline{\underline{\hat{S}_5}}(12)) = Var\left(\frac{N}{5}\right) = \frac{Var(N)}{25} = \frac{5 \cdot p \cdot (1-p)}{25} = \frac{p(1-p)}{5}$$

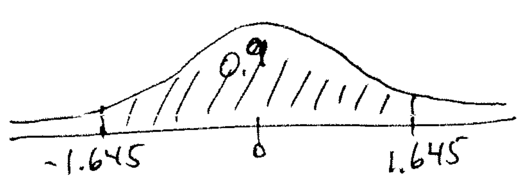
We estimate p by $\hat{p} = \frac{3}{5}$ (in the above example)

and we can estimate $Var(\underline{\underline{\hat{S}_5}}(12))$ by $\frac{\hat{p}(1-\hat{p})}{5} = \frac{3/5 \cdot 2/5}{5} = \frac{6}{125}$

I'll write $V_5 = \frac{6}{125}$

A "90%" confidence interval for a parameter is

estimate $\pm$ "1.645" $\cdot \sqrt{V_n}$



Example (LM.2 from L-TAM SOA Sample Questions)
(See Next Page)

Case 2: We have complete grouped data
In this case, we form the ogive by linearly interpolating between given $F_X(a)$-value where a is an endpoint of the given intervals

Example (~~LM.2~~ LM.5)
See next next page.

LM.2. In a study of 1,000 people with a particular illness, 200 died within one year of diagnosis. Calculate a 95% (linear) confidence interval for the one-year empirical survival function.

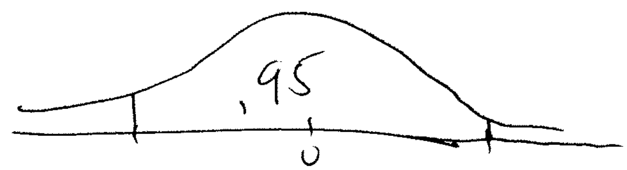
(A) (0.745, 0.855)

(B) (0.755, 0.845)

(C) (0.765, 0.835)

(D) (0.775, 0.825)

(E) (0.785, 0.815)



↳ 97.5 percentile = 1.96

Answer: $\hat{S}_{1000}(1) \pm 1.96 \cdot \sqrt{V_{1000}}$

$$\hat{S}_{1000}(1) = \frac{n}{1000} = \frac{800}{1000} = .8 = \hat{p}$$

$$V_{1000} = \frac{\hat{p}(1-\hat{p})}{1000} = \frac{.8(.2)}{1000}$$

Answer: $.8 \pm 1.96\sqrt{.00016}$

LM.5. In a study of workplace retention for a large employer, the following grouped data were collected from 100 new entrants.

Time to exit	Number of employees
0 – 5 years	28
5 – 10 years	19
10 – 20 years	15
20 – 30 years	30
Over 30 years	8

Calculate the probability that an employee exits within the first 12 years, using the ogive empirical distribution function.

(A) 0.49

(B) 0.50

(C) 0.51

(D) 0.52

(E) 0.53

We seek $Pr(X \leq 12) = F_X(12)$

Know $F_X(10) = \frac{28+19}{100} = 0.47$

and $F_X(20) = \frac{28+19+15}{100} = 0.62$

$\therefore \quad F_X(12) \approx \left(\frac{8}{10}\right)0.47 + \left(\frac{2}{10}\right)\cdot 0.62 =$