Chapter 23

The Accuracy of Averages

Confidence Intervals

Example 1. HANES women 18-24 have an average height of 64.3" with an SD of 2.6". Suppose we take a random sample of 100 of these women. What is the expected value of the average height of the women in the sample? It's SE?

$\boxed{\text{heights}} \rightarrow$ Draw 100 + consider AV of draws.

Box AV = 64.3", Box SD = 2.6"

EV for AV = Box AV = 64.3"

$$\text{SE for AV} = \frac{\text{Box SD} \times \sqrt{100}}{100} = \frac{(2.6)}{10}$$

$$= .26"$$

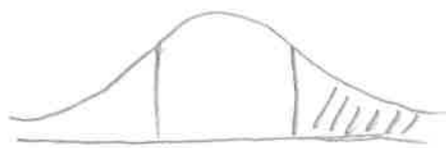
$$SE \text{ for } AV = \frac{\text{Box SD} \times \sqrt{100}}{100} = \frac{2.6 \times 10}{100} = .26$$

Example 2. HANES women 18-24 have an average height of 64.3" with an SD of 2.6". Suppose we take a random sample of 100 of these women.

a) What's the chance the sample average will be more than 64.5"?

AV of draws follows normal curve.

$$\frac{64.5 - 64.3}{.26} = .769$$



$A(.77) \approx 56\%$

$\frac{44\%}{2}$, 22%

b) What percentage of the women are taller than 64.5"?

heights follow normal curve

$$\frac{64.5 - 64.3}{2.6} = .07$$

$A(.07) \approx 6\%$

94% , 47%

- What is in the box?
- What is the average of the numbers in the box?

Women's heights: $SE \approx \frac{3.2 \times \sqrt{100}}{100} = .32$

68% C.I., $64.11 \pm .32$,

95% C.I., $64.11 \pm .64$

Family size: $SE \approx \frac{1.68 \times \sqrt{100}}{100} = .168$

68% C.I, $3.17 \pm .168$

95% C.I., $3.17 \pm 2(.168)$

$3.17 \pm .336$

The Bootstrap

When we do not know what is in the box, we estimate the SD of the box by the SD of the sample.

Confidence Intervals

A 95% confidence interval for the population average is given by

Sample average ± 2(SE for AV)

The confidence interval is valid if the number of draws is large enough.

Draw 300 + consider AV of draws

Example 3. A lake contains a large number of fish of a particular type. A simple random sample of 300 of these fish gives an average weight of 4.13 pounds with an SD of 2.1 pounds. Find a 95% confidence interval for the average weight of all the fish in the lake.

SE for AV $\approx \frac{(2.1)\sqrt{300}}{300} = .1212$

$4.13 \pm 2\,SE$

$4.13 \pm 2(.1212)$

$4.13 \pm .24$

$\boxed{\text{calories consumed}}$ → Draw 100 + consider AV of draws

Example 4. A nutrition student takes a simple random sample of 100 people from a large population and carefully monitors their caloric intake for 1 day. The average caloric intake for the sample is 2000 with an SD of 400.

a) Find a 95% confidence interval for the average caloric intake for the population.

SE for AV = $\frac{400 \times \sqrt{100}}{100} = 40$

95% C.I. : $2000 \pm 2(40)$

2000 ± 80

b) Is your confidence interval valid if the histogram for caloric intake is not normal?

yes (Normal approximation theorem)

12000 ages → Draw 500 + consider AV of draws

Example 5. A university has 12,000 students. A simple random sample of 500 students has average age 22.3 years with an SD of 4.1 years. Find a 90% confidence interval for the average age of all students at the university.

SE for AV $\approx \frac{(4.1)\sqrt{500}}{500} = .183$

$22.3 \pm (1.65)(.183)$

$22.3 \pm .3$

Reminder

Normal curve calculations, including confidence intervals, are valid if the number of draws is large enough.

How large is "large enough"? It depends on the box. If the box is a long way from normal (e.g. a box with lots of 0s and very few 1s) then the number of draws needs to be quite large.