Chapter 23: The Accuracy of Averages

Marius Ionescu

11/3/2011

Drawing at random from a box

Fact

When drawing at random from a box:

- *EV for average of draws = average of box.*
- *SE for average of draws=(SE for sum)/ (# of draws).*

Example

Example

Suppose that you have 10,000 dice and you roll 100 of them. What is the average of the draws? What is the error? a box with 1,2,3,4,5,6; $Avg_{box} = 3.5$; SD box= 1.7. $EV_{AVG} = 3.5$; $SE_{AVG} = \frac{\sqrt{100} \cdot 1.7}{100} = 0.17$

Normal approximation

Fact

Recall: When drawing at random from a box, the probability histogram for the average of the draws follows the normal curve, even it the contents of the box do not.

Example

Example (Rolling a die, con't)

Coming back to the die example, find the chance that the average of the draws is higher then 5. Average=3.5, SD =0.17; Z=8.82; the chance is basically 0

The sample average

Fact

With a simple random sample, the SD of the sample can be used to estimate the SD of the box. The estimate is good when the sample is large.

Example

Example

Suppose that a sample of 400 students at Colgate is taken. The total age of the sample persons is 8080. The SD of the sample is 0.8 years. Find a 95%-confidence interval for the average age of the students at Colgate.
$EV_{AVG} = \frac{8080}{400} = 20.2$; $SE_{AVG} = \frac{0.8}{\sqrt{400}} = 0.04$; $20.2 \pm 2 \cdot 0.04$

Which SE?

Fact (Sum up of SE's)

- sSE for sum $= \sqrt{\text{number of draws}} \times$ SD of box.
- SE for averages $= \frac{\text{SE for sum}}{\text{number of draws}}$
- SE for count = SE or sum, from a 0-1 box
- SE for percentage$= \frac{\text{SE for count}}{\text{number of draws}} \times 100\%$.
- The SE shows the likely size of the amount off. It is a give-or-take amount.

Examples

Example

A box of tickets has an average of 10,000, and an SD of 2000. Four hundred draws will be made at random with replacement from this box.

- Estimate the chance that the average of the draws will be in the range 8,000 to 12,000. $EV_{AVG} = 10,000$; $SE_{AVG} = \frac{\sqrt{400} \cdot 2000}{400} = 100$; Chance is basically 100%
- Estimate the chance that the average of the draws will be in the range 9,900 to 10100. 68%

Example

Example

A simple random sample of 400 firms was taken from the population of all manufacturing firms in the state. 16 in the sample had 250 or more employees.

- Estimate the percentage of manufacturing firms in the state with 250 or more employees. box with 16 of 1s and 384 of 0s; $Avg_{box} = \frac{16}{400} = 0.04$; $SD_{box} = \sqrt{\frac{16}{400} \cdot \frac{384}{400}} = 0.195$; $EV_{\%} = 4\%$; $SE_{\%} = \frac{\sqrt{400} \cdot 0.195}{400} \simeq 0.01 * 100\% = 1\%$
- Find a 68% confidence interval
- Find a 95% confidence interval

Example

Example

We make 1600 draws from a box. The average of the draws is 5.3 and the SD is 2. Find the 68% confidence interval for the average. $EV_{AVG} = 5.3$; $SE_{AVG} = \frac{\sqrt{1600} \cdot 2}{1600} = 0.005$

Example

100,000 tax forms are reported to have an average income of \$12,000 with an SD of \$6000. Additional study of 900 forms is proposed. What is the chance that income on these 900 forms will average between \$11,800 and \$12,200? $EV_{AVG} = 12,000$; $SE_{AVG} = \frac{\sqrt{900} \cdot 6000}{900} = 200$; 68%

Example

Example

740 Colgate students take 32 courses in 4 years. Suppose grades are given only with letter values (no +/-) and the numbers 0,1,2,3,4 are assigned to F,D,C,B,A. Let's test the idea that professors give out grades randomly.

- If grades are assigned randomly, how many students do we expect to have GPA of 3.0 of higher? $AVG_{box} = 2$; $SD_{box} = 1.41$; $EV_{AVG} = 2$; $SE_{AVG} = \frac{\sqrt{32} \cdot 1.41}{32} = 0.25$; Z=4; $P(X \geq 3) = \frac{100 - A(4)}{2} \simeq 0.003\%$; out of 740 students: $\simeq 0.02$ students; basically no honor students at Colgate