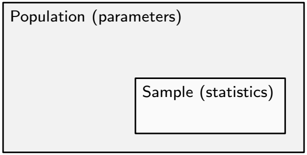
Math 207: Statistics

Chapter 20: Chance Error in Sampling



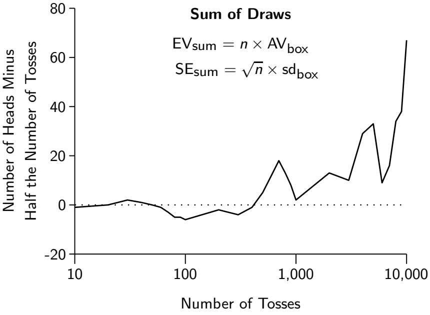
Dr. Ralph Wojtowicz

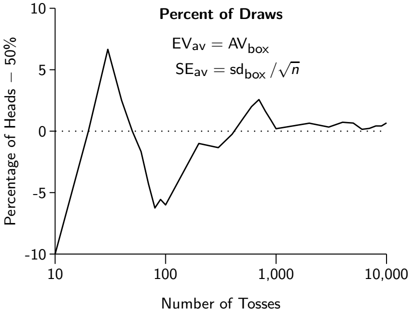
Mathematics Department

Shenandoah University

Expected Value, Standard Error, Central Limit Theorem

- Many statistics problems are modeled as samples from a box of numbered tickets.
- Solution procedure:
  - Formulate a box model.
  - Compute the average and SD of the contents of the box.
  - Determine if you are computing a sum or average (% in a 0/1 box is an average).
  - Use the appropriate formulas to compute the EV and SE for the sample.
  - Use the normal curve to compute chances of the sample being in a specified range.





Example (I)

- Running a political poll is like sampling from a box of number tickets.

Example (I)

- Running a political poll is like sampling from a box of number tickets.
- Take a random sample of size $n = 100$ from the box below.

$$\boxed{253{,}785\ \boxed{0}\text{s}\quad 433{,}211\ \boxed{1}\text{s}}$$

Example (I)

- Running a political poll is like sampling from a box of number tickets.
- Take a random sample of size $n = 100$ from the box below.

$$\boxed{253{,}785\ \boxed{0}\text{s}\quad 433{,}211\ \boxed{1}\text{s}}$$

- What is the chance that the percent of 1s is between 61% and 65%?

Example (I)

- Running a political poll is like sampling from a box of number tickets.
- Take a random sample of size $n = 100$ from the box below.

$$\boxed{253{,}785\ \boxed{0}\,\text{s}\quad 433{,}211\ \boxed{1}\,\text{s}}$$

- What is the chance that the percent of 1s is between 61% and 65%?
- $\text{AV}_{\text{box}} = \frac{433{,}211}{433{,}211+253{,}785} \approx 0.631$ and $\text{SD}_{\text{box}} = (1-0)\sqrt{0.631\,(1-0.631)} \approx 0.483$

Example (I)

- Running a political poll is like sampling from a box of number tickets.
- Take a random sample of size $n = 100$ from the box below.

$$\boxed{253{,}785\ \boxed{0}\text{s} \quad 433{,}211\ \boxed{1}\text{s}}$$

- What is the chance that the percent of 1s is between 61% and 65%?
- $\text{AV}_{\text{box}} = \frac{433{,}211}{433{,}211+253{,}785} \approx 0.631$ and $\text{SD}_{\text{box}} = (1-0)\sqrt{0.631\,(1-0.631)} \approx 0.483$
- $\text{EV}_{\%} = \text{AV}_{\text{box}} = 0.631 = 63.1\%$ and $\text{SE}_{\%} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{0.483}{\sqrt{100}} = 0.0483 = 4.83\%$.

Example (I)

- Running a political poll is like sampling from a box of number tickets.
- Take a random sample of size $n = 100$ from the box below.

$$\boxed{253{,}785\ \boxed{0}\text{s} \quad 433{,}211\ \boxed{1}\text{s}}$$

- What is the chance that the percent of 1s is between 61% and 65%?
- $\text{AV}_{\text{box}} = \frac{433{,}211}{433{,}211+253{,}785} \approx 0.631$ and $\text{SD}_{\text{box}} = (1-0)\sqrt{0.631\,(1-0.631)} \approx 0.483$
- $\text{EV}_{\%} = \text{AV}_{\text{box}} = 0.631 = 63.1\%$ and $\text{SE}_{\%} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{0.483}{\sqrt{100}} = 0.0483 = 4.83\%$.
- Compute the z scores for 61% and 65%:
  $z_1 = (61 - 63.1)/4.83 = -0.435$
  $z_2 = (65 - 63.1)/4.83 = 0.393$

Example (I)

- Running a political poll is like sampling from a box of number tickets.
- Take a random sample of size $n = 100$ from the box below.

$$\boxed{253{,}785 \ \boxed{0}\text{s} \quad 433{,}211 \ \boxed{1}\text{s}}$$

- What is the chance that the percent of 1s is between 61% and 65%?
- $\text{AV}_{\text{box}} = \frac{433{,}211}{433{,}211+253{,}785} \approx 0.631$ and $\text{SD}_{\text{box}} = (1-0)\sqrt{0.631\,(1-0.631)} \approx 0.483$
- $\text{EV}_{\%} = \text{AV}_{\text{box}} = 0.631 = 63.1\%$ and $\text{SE}_{\%} = \dfrac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{0.483}{\sqrt{100}} = 0.0483 = 4.83\%$.
- Compute the z scores for 61% and 65%:
  $z_1 = (61 - 63.1)/4.83 = -0.435$
  $z_2 = (65 - 63.1)/4.83 = 0.393$
- Area under the normal curve: $\texttt{pnorm}(0.393) - \texttt{pnorm}(-0.435) = 0.321 = 32.1\%$

Example (II)

- What if we increase the sample size to 1000?

Example (II)

- What if we increase the sample size to 1000?
- $\text{AV}_{\text{box}} \approx 0.631$ and $\text{SD}_{\text{box}} \approx 0.483$ don't change.

Example (II)

- What if we increase the sample size to 1000?
- $\text{AV}_{\text{box}} \approx 0.631$ and $\text{SD}_{\text{box}} \approx 0.483$ don't change.
- $\text{EV}_{\%} = \text{AV}_{\text{box}} = 0.631 = 63.1\%$ and $\text{SE}_{\%} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{0.483}{\sqrt{1000}} = 0.0153 = 1.53\%$.

Example (II)

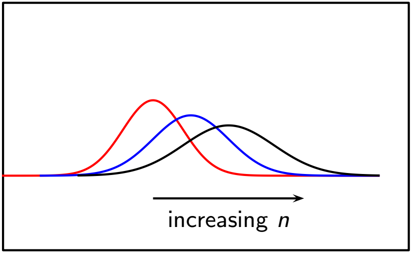
- What if we increase the sample size to 1000?
- $\text{AV}_{\text{box}} \approx 0.631$ and $\text{SD}_{\text{box}} \approx 0.483$ don't change.
- $\text{EV}_{\%} = \text{AV}_{\text{box}} = 0.631 = 63.1\%$ and $\text{SE}_{\%} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{0.483}{\sqrt{1000}} = 0.0153 = 1.53\%$.
- Compute the z scores for 61% and 65%:
  $z_1 = (61 - 63.1)/1.53 = -1.37$
  $z_2 = (65 - 63.1)/1.53 = 1.24$

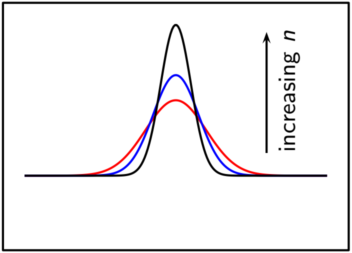
Example (II)

- What if we increase the sample size to 1000?
- $\text{AV}_{\text{box}} \approx 0.631$ and $\text{SD}_{\text{box}} \approx 0.483$ don't change.
- $\text{EV}_{\%} = \text{AV}_{\text{box}} = 0.631 = 63.1\%$ and $\text{SE}_{\%} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{0.483}{\sqrt{1000}} = 0.0153 = 1.53\%$.
- Compute the z scores for 61% and 65%:
  $z_1 = (61 - 63.1)/1.53 = -1.37$
  $z_2 = (65 - 63.1)/1.53 = 1.24$
- Area under the normal curve: $\texttt{pnorm}(1.24) - \texttt{pnorm}(-1.37) = 0.808 = 80.8\%$

Example (II)

- What if we increase the sample size to 1000?
- $\text{AV}_{\text{box}} \approx 0.631$ and $\text{SD}_{\text{box}} \approx 0.483$ don't change.
- $\text{EV}_{\%} = \text{AV}_{\text{box}} = 0.631 = 63.1\%$ and $\text{SE}_{\%} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{0.483}{\sqrt{1000}} = 0.0153 = 1.53\%$.
- Compute the z scores for 61% and 65%:
  $z_1 = (61 - 63.1)/1.53 = -1.37$
  $z_2 = (65 - 63.1)/1.53 = 1.24$
- Area under the normal curve: $\texttt{pnorm}(1.24) - \texttt{pnorm}(-1.37) = 0.808 = 80.8\%$





Example (III)

- Suppose we roll 100 dice.

Example (III)

- Suppose we roll 100 dice.
- What is the chance that the total will be greater than 375?

Example (III)

- Suppose we roll 100 dice.
- What is the chance that the total will be greater than 375?
- The sample is like the sum of 100 draws from:

$$\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}$$

Example (III)

- Suppose we roll 100 dice.
- What is the chance that the total will be greater than 375?
- The sample is like the sum of 100 draws from:

$$\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}$$

- $\text{AV}_{\text{box}} = \frac{1+2+\cdots+6}{6} = 3.5$ and $\text{SD}_{\text{box}} = \sqrt{\frac{(1-3.5)^2+(2-3.5)^2+\cdots+(6-3.5)^2}{6}} \approx 1.71$

Example (III)

- Suppose we roll 100 dice.
- What is the chance that the total will be greater than 375?
- The sample is like the sum of 100 draws from:

$$\boxed{1}\,\boxed{2}\,\boxed{3}\,\boxed{4}\,\boxed{5}\,\boxed{6}$$

- $\text{AV}_{\text{box}} = \frac{1+2+\cdots+6}{6} = 3.5$ and $\text{SD}_{\text{box}} = \sqrt{\frac{(1-3.5)^2+(2-3.5)^2+\cdots+(6-3.5)^2}{6}} \approx 1.71$
- $\text{EV}_{\text{sum}} = n \cdot \text{AV}_{\text{box}} = 350$ and $\text{SE}_{\text{sum}} = \sqrt{n} \cdot \text{SD}_{\text{box}} \approx 17.1.$

Example (III)

- Suppose we roll 100 dice.
- What is the chance that the total will be greater than 375?
- The sample is like the sum of 100 draws from:

$$\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}$$

- $\text{AV}_{\text{box}} = \frac{1+2+\cdots+6}{6} = 3.5$ and $\text{SD}_{\text{box}} = \sqrt{\frac{(1-3.5)^2+(2-3.5)^2+\cdots+(6-3.5)^2}{6}} \approx 1.71$
- $\text{EV}_{\text{sum}} = n \cdot \text{AV}_{\text{box}} = 350$ and $\text{SE}_{\text{sum}} = \sqrt{n} \cdot \text{SD}_{\text{box}} \approx 17.1$.
- Compute the z scores for 375:
  $z = (375 - 350)/17.1 = 1.46$

Example (III)

- Suppose we roll 100 dice.
- What is the chance that the total will be greater than 375?
- The sample is like the sum of 100 draws from:

$$\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}$$

- $\text{AV}_{\text{box}} = \frac{1+2+\cdots+6}{6} = 3.5$ and $\text{SD}_{\text{box}} = \sqrt{\frac{(1-3.5)^2+(2-3.5)^2+\cdots+(6-3.5)^2}{6}} \approx 1.71$
- $\text{EV}_{\text{sum}} = n \cdot \text{AV}_{\text{box}} = 350$ and $\text{SE}_{\text{sum}} = \sqrt{n} \cdot \text{SD}_{\text{box}} \approx 17.1$.
- Compute the z scores for 375:
  $z = (375 - 350)/17.1 = 1.46$
- Area under the normal curve: $1 - \texttt{pnorm}(1.46) \approx 0.0721 = 7.21\%$