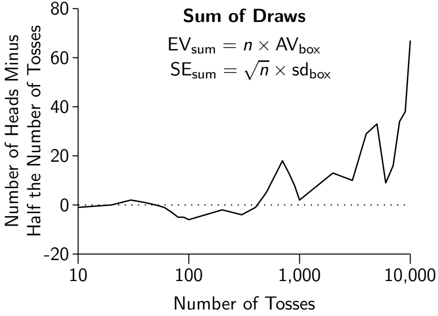
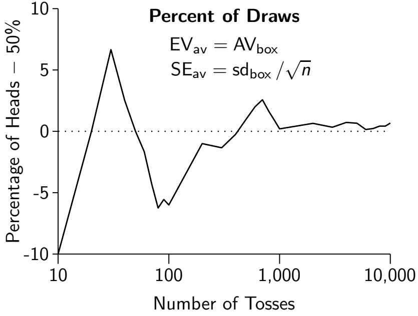
Math 207: Statistics

Chapter 18: The Normal Approximation for Probability Histograms





Dr. Ralph Wojtowicz

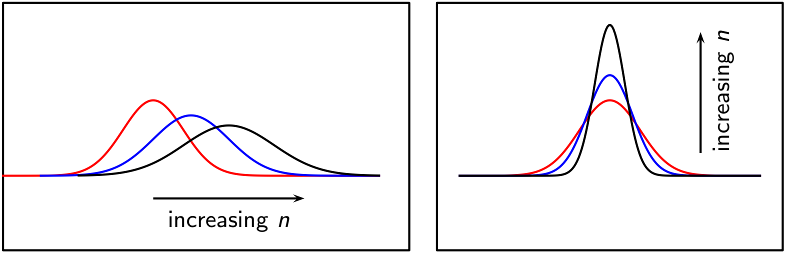
Mathematics Department

Shenandoah University

Using the Normal Curve

Central Limit Theorem: When drawing at random with replacement from a box, the probability histogram for the sum (and the average) will follow the normal curve, even if the contents of the box do not. The histogram must be put into standard units, and the number of draws must be reasonably large.

$$\begin{aligned} \mathrm{EV_{sum}} &= n \cdot \mathrm{AV_{box}} \\ \mathrm{SE_{sum}} &= \sqrt{n} \cdot \mathrm{SD_{box}} \end{aligned} \qquad \begin{aligned} \mathrm{EV_{av}} &= \mathrm{AV_{box}} \\ \mathrm{SE_{av}} &= \frac{\mathrm{SD_{box}}}{\sqrt{n}} \end{aligned}$$



Examples (I)

- A coin will be flipped 100 times. You are about 68% confident that the number of tails will fall in what range?
  The box is $\boxed{\boxed{0}\ \boxed{1}}$.
  $AV_{box} = 0.5$ and $SD_{box} = 0.5$.
  $EV_{sum} = 0.5 \cdot 100 = 50$ and $SE_{sum} = 0.5 \cdot \sqrt{100} = 5$
  Answer: 50 ± 5
- A coin will be flipped 10,000 times. You are about 95% confident that the number of tails will fall in what range?
  $EV_{sum} = 0.5 \cdot 10{,}000 = 5000$ and $SE_{sum} = 0.5 \cdot \sqrt{10000} = 50$
  Answer: $5000 \pm 2 \cdot 50 = 5000 \pm 100$
- A coin will be flipped 10,000 times. You are about 95% confident that the percent of tails will fall in what range?
  $EV_{av} = 0.5 = 50\%$ and $SE_{av} = 0.5/\sqrt{10000} = 0.005 = 0.5\%$
  Answer: $50\% \pm 2 \cdot 0.5\% = 50\% \pm 1\%$

Examples (II)

Examples (II)

- A die will be tossed 120 times. What is the chance that the number of 4s will be between 15 and 25?

  The box model is $\boxed{\boxed{0}\,\boxed{0}\,\boxed{0}\,\boxed{1}\,\boxed{0}\,\boxed{0}}$

  $\text{AV}_{\text{box}} = \frac{1}{6}$ and $\text{SD}_{\text{box}} = \frac{\sqrt{5}}{6} \approx 0.373$.

  $\text{EV}_{\text{sum}} = 120 \cdot \frac{1}{6} = 20$ and $\text{SE}_{\text{sum}} = \sqrt{120} \cdot \frac{\sqrt{5}}{6} \approx 4.08$

  $z_1 = \frac{15-20}{4.08} = -1.23$ and $z_2 = \frac{25-20}{4.08} = 1.23$

  $\text{Chance} = \text{pnorm}(1.23) - \text{pnorm}(-1.23) = 0.781 = 78.1\%$

Examples (II)

- A die will be tossed 120 times. What is the chance that the number of 4s will be between 15 and 25?

  The box model is $\boxed{\boxed{0}\,\boxed{0}\,\boxed{0}\,\boxed{1}\,\boxed{0}\,\boxed{0}}$

  $\text{AV}_{\text{box}} = \frac{1}{6}$ and $\text{SD}_{\text{box}} = \frac{\sqrt{5}}{6} \approx 0.373$.

  $\text{EV}_{\text{sum}} = 120 \cdot \frac{1}{6} = 20$ and $\text{SE}_{\text{sum}} = \sqrt{120} \cdot \frac{\sqrt{5}}{6} \approx 4.08$

  $z_1 = \frac{15-20}{4.08} = -1.23$ and $z_2 = \frac{25-20}{4.08} = 1.23$

  Chance $= \text{pnorm}(1.23) - \text{pnorm}(-1.23) = 0.781 = 78.1\%$

- A die will be tossed 12,000 times. What is the chance that the percentage of 4s will be between 16% and 17%?

  $\text{EV}_{\text{av}} = \frac{1}{6} = 0.166667 = 16.7\%$ and $\text{SE}_{\text{av}} = \frac{\sqrt{5}}{6}/\sqrt{12000} \approx 0.0034 = 0.34\%$

  $z_1 = \frac{16-16.7}{0.34} = -1.96$ and $z_2 = \frac{17-16.7}{0.34} = 0.98$

  Chance $= \text{pnorm}(0.98) - \text{pnorm}(-1.96) = 0.811 = 81.1\%$

SD Shortcut

- If there are only two different numbers on the tickets in a box then we can use a shortcut formula to compute the SD of the box.

$$\text{SD}_{\text{box}} = \left(\begin{array}{c}\text{big}\\ \text{number}\end{array} - \begin{array}{c}\text{small}\\ \text{number}\end{array}\right)\sqrt{\left(\begin{array}{c}\text{fraction of}\\ \text{tickets with}\\ \text{big number}\end{array}\right)\cdot\left(\begin{array}{c}\text{fraction of}\\ \text{tickets with}\\ \text{small number}\end{array}\right)}$$

- Examples:

	Box	SD
Flipping coins	$\boxed{\boxed{0}\ \boxed{1}}$	$(1-0)\sqrt{\frac{1}{2}\cdot\frac{1}{2}} = \frac{1}{2} = 0.5$
Counting 4s on die rolls	$\boxed{\boxed{0}\ \boxed{0}\ \boxed{0}\ \boxed{1}\ \boxed{0}\ \boxed{0}}$	$(1-0)\sqrt{\frac{1}{6}\cdot\frac{5}{6}} = \frac{\sqrt{5}}{6} \approx 0.373$
Sample from a box	$\boxed{\boxed{-1}\ \boxed{-1}\ \boxed{1}\ \boxed{1}\ \boxed{1}}$	$(1-(-1))\sqrt{\frac{3}{5}\cdot\frac{2}{5}} = \frac{2\sqrt{6}}{5} \approx 0.980$
Sample from a box	$\boxed{\boxed{1}\ \boxed{3}\ \boxed{3}\ \boxed{5}}$	We can't use the shortcut.