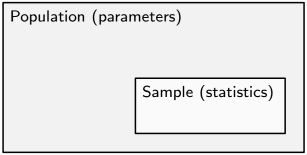
# Math 207: Statistics

## Chapter 21: The Accuracy of Percentages




Dr. Ralph Wojtowicz

Mathematics Department

Shenandoah University

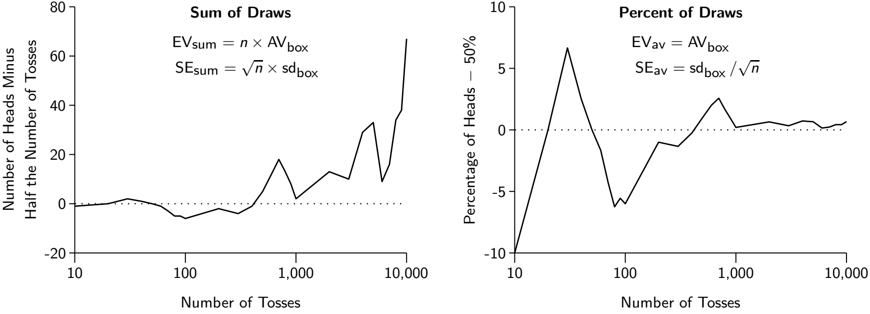
1 EV, SE and the Central Limit Theorem

2 Examples
- Example I
- Example I

3 Confidence Intervals
- Inferential Statistics
- Confidence Intervals
- Example

# 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: Box Contents Known

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

# Example: Box Contents Known

- 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: Box Contents Known

- 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: Box Contents Known

- 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: Box Contents Known

- 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: Box Contents Known

- 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: Box Contents Known

- 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%:
  $z_1 = (61 - 63.1)/4.83 = -0.435$
  $z_2 = (65 - 63.1)/4.83 = 0.393$
- Area under the normal curve: $\texttt{pnorm}(0.393) - \texttt{pnorm}(-0.435) = 0.321 = 32.1\%$

# Example: Box Contents Unknown

# Example: Box Contents Unknown

- Take a random sample of size $n = 100$ from the box below.

$$\boxed{\text{???}\ \boxed{0}\text{s}\quad \text{???}\ \boxed{1}\text{s}}$$

# Example: Box Contents Unknown

- Take a random sample of size $n = 100$ from the box below.

$$\boxed{???\ \boxed{0}\text{s}\quad ???\ \boxed{1}\text{s}}$$

- Suppose our sample has 63 $\boxed{1}$s.

# Example: Box Contents Unknown

- Take a random sample of size $n = 100$ from the box below.

$$\boxed{\text{???}\ \boxed{0}\text{s}\quad \text{???}\ \boxed{1}\text{s}}$$

- Suppose our sample has 63 $\boxed{1}$s.
- Estimate the fraction of $\boxed{1}$s in the box and put a plus or minus on the estimate.
- $AV_{\text{box}} =$ ??? Use 63% as our estimate.

# Example: Box Contents Unknown

- Take a random sample of size $n = 100$ from the box below.

$$\boxed{??? \ \boxed{0}\text{s} \quad ??? \ \boxed{1}\text{s}}$$

- Suppose our sample has 63 $\boxed{1}$s.
- Estimate the fraction of $\boxed{1}$s in the box and put a plus or minus on the estimate.
- $\text{AV}_{\text{box}} = ???$ Use 63% as our estimate.
- $\text{SD}_{\text{box}} = ???$. Use $(1-0)\sqrt{0.63\,(1-0.63)} \approx 0.483$ as our estimate.
- $\text{SE}_{\%} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{0.483}{\sqrt{100}} = 0.0483 = 4.83\%$.

# Example: Box Contents Unknown

- Take a random sample of size $n = 100$ from the box below.

$$\boxed{??? \boxed{0}\text{s} \quad ??? \boxed{1}\text{s}}$$

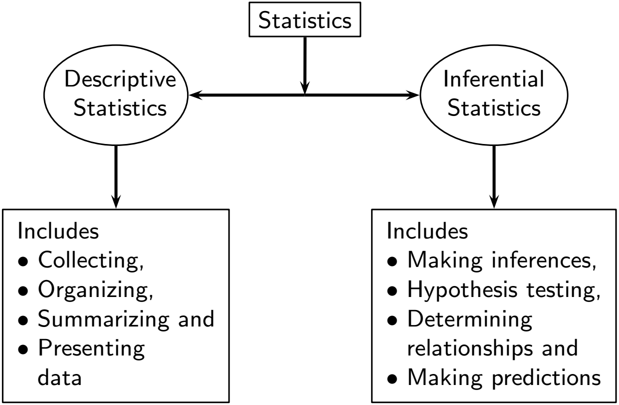
- Suppose our sample has 63 $\boxed{1}$s.
- Estimate the fraction of $\boxed{1}$s in the box and put a plus or minus on the estimate.
- $\text{AV}_{\text{box}} = ???$ Use 63% as our estimate.
- $\text{SD}_{\text{box}} = ???$. Use $(1-0)\sqrt{0.63\,(1-0.63)} \approx 0.483$ as our estimate.
- $\text{SE}_{\%} = \dfrac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{0.483}{\sqrt{100}} = 0.0483 = 4.83\%$.
- If we assume that the samples follow a normal curve, then

$$63\% \ \pm\ 2\cdot(4.83\%) = 63\% \ \pm 9.7\%$$

is 95% confidence interval on the fraction of $\boxed{1}$s.

# Inferential Statistics

- In real situations, we typically don't know the contents of the box!
- That is where inferential statistics is used.



# Confidence Intervals

- In real situations, we typically don't know the contents of the box!

# Confidence Intervals

- In real situations, we typically don't know the contents of the box!
- We can take a sample and compute a confidence interval on the average of the numbers in the box.

# Confidence Intervals

- In real situations, we typically don't know the contents of the box!
- We can take a sample and compute a confidence interval on the average of the numbers in the box.
- This gives an estimate of the average and puts a plus-or-minus on the estimate.

$$\mathrm{EV}_{\mathrm{av}} = \mathrm{AV}_{\mathrm{box}} \qquad \mathrm{SE}_{\mathrm{av}} = \frac{\mathrm{SD}_{\mathrm{box}}}{\sqrt{n}}$$

# Confidence Intervals

- In real situations, we typically don't know the contents of the box!
- We can take a sample and compute a confidence interval on the average of the numbers in the box.
- This gives an estimate of the average and puts a plus-or-minus on the estimate.

$$\text{EV}_{\text{av}} = \text{AV}_{\text{box}} \qquad \text{SE}_{\text{av}} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}}$$

- Assume the box contains only $\boxed{0}$s and $\boxed{1}$s.

# Confidence Intervals

- In real situations, we typically don't know the contents of the box!
- We can take a sample and compute a confidence interval on the average of the numbers in the box.
- This gives an estimate of the average and puts a plus-or-minus on the estimate.

$$\mathrm{EV}_{\mathrm{av}} = \mathrm{AV}_{\mathrm{box}} \qquad \mathrm{SE}_{\mathrm{av}} = \frac{\mathrm{SD}_{\mathrm{box}}}{\sqrt{n}}$$

- Assume the box contains only $\boxed{0}$s and $\boxed{1}$s.
- Take a random sample of size $n$.

# Confidence Intervals

- In real situations, we typically don't know the contents of the box!
- We can take a sample and compute a confidence interval on the average of the numbers in the box.
- This gives an estimate of the average and puts a plus-or-minus on the estimate.

$$\text{EV}_{\text{av}} = \text{AV}_{\text{box}} \qquad \text{SE}_{\text{av}} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}}$$

- Assume the box contains only $\boxed{0}$s and $\boxed{1}$s.
- Take a random sample of size $n$.
- The fraction $p$ of $\boxed{1}$s in the sample estimates the fraction of $\boxed{1}$s in the box.

# Confidence Intervals

- In real situations, we typically don't know the contents of the box!
- We can take a sample and compute a confidence interval on the average of the numbers in the box.
- This gives an estimate of the average and puts a plus-or-minus on the estimate.

$$\text{EV}_{\text{av}} = \text{AV}_{\text{box}} \qquad \text{SE}_{\text{av}} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}}$$

- Assume the box contains only $\boxed{0}$s and $\boxed{1}$s.
- Take a random sample of size $n$.
- The fraction $p$ of $\boxed{1}$s in the sample estimates the fraction of $\boxed{1}$s in the box.
- Use the shortcut method to estimate the SD: $\text{SD}_{\text{box}} \approx (1 - 0)\sqrt{p\,(1 - p)}$.

# Confidence Intervals

- In real situations, we typically don't know the contents of the box!
- We can take a sample and compute a confidence interval on the average of the numbers in the box.
- This gives an estimate of the average and puts a plus-or-minus on the estimate.

$$\mathrm{EV}_{\mathrm{av}} = \mathrm{AV}_{\mathrm{box}} \qquad \mathrm{SE}_{\mathrm{av}} = \frac{\mathrm{SD}_{\mathrm{box}}}{\sqrt{n}}$$

- Assume the box contains only $\boxed{0}$s and $\boxed{1}$s.
- Take a random sample of size $n$.
- The fraction $p$ of $\boxed{1}$s in the sample estimates the fraction of $\boxed{1}$s in the box.
- Use the shortcut method to estimate the SD: $\mathrm{SD}_{\mathrm{box}} \approx (1-0)\sqrt{p(1-p)}$.
- Then compute the SE.

# Confidence Intervals

- In real situations, we typically don't know the contents of the box!
- We can take a sample and compute a confidence interval on the average of the numbers in the box.
- This gives an estimate of the average and puts a plus-or-minus on the estimate.

$$\mathrm{EV}_{\mathrm{av}} = \mathrm{AV}_{\mathrm{box}} \qquad \mathrm{SE}_{\mathrm{av}} = \frac{\mathrm{SD}_{\mathrm{box}}}{\sqrt{n}}$$

- Assume the box contains only $\boxed{0}$s and $\boxed{1}$s.
- Take a random sample of size $n$.
- The fraction $p$ of $\boxed{1}$s in the sample estimates the fraction of $\boxed{1}$s in the box.
- Use the shortcut method to estimate the SD: $\mathrm{SD}_{\mathrm{box}} \approx (1-0)\sqrt{p\,(1-p)}$.
- Then compute the SE.
- The 95% confidence interval is:

$$\text{(observed fraction)} \pm 2 \cdot \text{(SE estimate)}$$

# Confidence Intervals

- In real situations, we typically don't know the contents of the box!
- We can take a sample and compute a confidence interval on the average of the numbers in the box.
- This gives an estimate of the average and puts a plus-or-minus on the estimate.

$$\mathrm{EV}_{\mathrm{av}} = \mathrm{AV}_{\mathrm{box}} \qquad \mathrm{SE}_{\mathrm{av}} = \frac{\mathrm{SD}_{\mathrm{box}}}{\sqrt{n}}$$

- Assume the box contains only $\boxed{0}$s and $\boxed{1}$s.
- Take a random sample of size $n$.
- The fraction $p$ of $\boxed{1}$s in the sample estimates the fraction of $\boxed{1}$s in the box.
- Use the shortcut method to estimate the SD: $\mathrm{SD}_{\mathrm{box}} \approx (1-0)\sqrt{p\,(1-p)}$.
- Then compute the SE.
- The 95% confidence interval is:

$$(\text{observed fraction}) \pm 2 \cdot (\text{SE estimate})$$

- Prior to taking the sample, we are 95% confident that this procedure will give an interval that contains the true average.

# Example 1 Page 378

- A university has 25,000 registered students.

# Example 1 Page 378

- A university has 25,000 registered students.
- Estimate the fraction who live at home by taking a random sample.

# Example 1 Page 378

- A university has 25,000 registered students.
- Estimate the fraction who live at home by taking a random sample.
- Box model:
  - 25,000 numbered tickets
  - a $\boxed{1}$ for every student who lives at home
  - a $\boxed{0}$ for others

# Example 1 Page 378

- A university has 25,000 registered students.
- Estimate the fraction who live at home by taking a random sample.
- Box model:
  - 25,000 numbered tickets
  - a $\boxed{1}$ for every student who lives at home
  - a $\boxed{0}$ for others
- Observed fraction of $\boxed{1}$ s: 317/400.

# Example 1 Page 378

- A university has 25,000 registered students.
- Estimate the fraction who live at home by taking a random sample.
- Box model:
  - 25,000 numbered tickets
  - a $\boxed{1}$ for every student who lives at home
  - a $\boxed{0}$ for others
- Observed fraction of $\boxed{1}$s: 317/400.
- Estimated fraction of $\boxed{1}$s in the box: $317/400 = 79.3\%$.

# Example 1 Page 378

- A university has 25,000 registered students.
- Estimate the fraction who live at home by taking a random sample.
- Box model:
  - 25,000 numbered tickets
  - a $\boxed{1}$ for every student who lives at home
  - a $\boxed{0}$ for others
- Observed fraction of $\boxed{1}$s: 317/400.
- Estimated fraction of $\boxed{1}$s in the box: $317/400 = 79.3\%$.
- Now compute a $\pm$ on this estimate:
  - $\text{SD}_{\text{box}} \approx \sqrt{0.793 \cdot (1 - 0.793)} = 0.41.$
  - $\text{SE}_{\text{av}} = 0.41/\sqrt{400} \approx 0.020 = 2.0\%$

# Example 1 Page 378

- A university has 25,000 registered students.
- Estimate the fraction who live at home by taking a random sample.
- Box model:
  - 25,000 numbered tickets
  - a $\boxed{1}$ for every student who lives at home
  - a $\boxed{0}$ for others
- Observed fraction of $\boxed{1}$s: 317/400.
- Estimated fraction of $\boxed{1}$s in the box: $317/400 = 79.3\%$.
- Now compute a $\pm$ on this estimate:
  - $\text{SD}_{\text{box}} \approx \sqrt{0.793 \cdot (1 - 0.793)} = 0.41.$
  - $\text{SE}_{\text{av}} = 0.41/\sqrt{400} \approx 0.020 = 2.0\%$
- 95% confidence interval:

$$79.3\% \ \pm \ 2 \cdot (2.0\%) = 79.3\% \ \pm \ 4\%$$