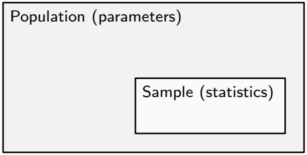
Math 207: Statistics

Chapter 23: The Accuracy of Averages



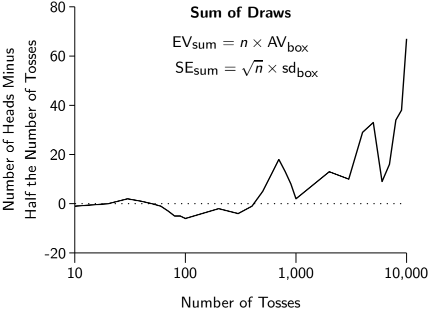
Dr. Ralph Wojtowicz

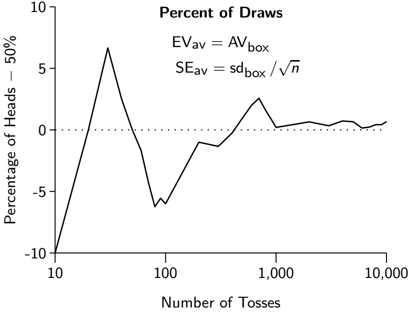
Mathematics Department

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: Box Contents Known

Example: Box Contents Known

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

$$\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}\ \boxed{7}}$$

Example: Box Contents Known

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

$$\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}\ \boxed{7}}$$

- $\text{AV}_{\text{box}} = \frac{1+2+\cdots+7}{7} = 4$ and $\text{SD}_{\text{box}} = \sqrt{\frac{(1-4)^2+(2-4)^2+\cdots+(7-4)^2}{7}} = \sqrt{\frac{28}{7}} = \sqrt{4} = 2.$

Example: Box Contents Known

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

$$\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}\ \boxed{7}}$$

- $\text{AV}_{\text{box}} = \frac{1+2+\cdots+7}{7} = 4$ and $\text{SD}_{\text{box}} = \sqrt{\frac{(1-4)^2+(2-4)^2+\cdots+(7-4)^2}{7}} = \sqrt{\frac{28}{7}} = \sqrt{4} = 2.$
- $\text{EV}_{\text{av}} = \text{AV}_{\text{box}} = 4$ and $\text{SE}_{\text{av}} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{4}{\sqrt{100}} = 0.4.$

Example: Box Contents Known

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

$$\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}\ \boxed{7}}$$

- $\text{AV}_{\text{box}} = \frac{1+2+\cdots+7}{7} = 4$ and $\text{SD}_{\text{box}} = \sqrt{\frac{(1-4)^2+(2-4)^2+\cdots+(7-4)^2}{7}} = \sqrt{\frac{28}{7}} = \sqrt{4} = 2.$
- $\text{EV}_{\text{av}} = \text{AV}_{\text{box}} = 4$ and $\text{SE}_{\text{av}} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{4}{\sqrt{100}} = 0.4.$
- We expect the average of the draws to be around 4 give or take 0.4 or so.

Example: Box Contents Known

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

$$\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}\ \boxed{7}}$$

- $\text{AV}_{\text{box}} = \frac{1+2+\cdots+7}{7} = 4$ and $\text{SD}_{\text{box}} = \sqrt{\frac{(1-4)^2+(2-4)^2+\cdots+(7-4)^2}{7}} = \sqrt{\frac{28}{7}} = \sqrt{4} = 2.$
- $\text{EV}_{\text{av}} = \text{AV}_{\text{box}} = 4$ and $\text{SE}_{\text{av}} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{4}{\sqrt{100}} = 0.4.$
- We expect the average of the draws to be around 4 give or take 0.4 or so.
- What is the chance that the average will be between 3.5 and 4.5?

Example: Box Contents Known

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

$$\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}\ \boxed{7}}$$

- $\text{AV}_{\text{box}} = \frac{1+2+\cdots+7}{7} = 4$ and $\text{SD}_{\text{box}} = \sqrt{\frac{(1-4)^2+(2-4)^2+\cdots+(7-4)^2}{7}} = \sqrt{\frac{28}{7}} = \sqrt{4} = 2.$
- $\text{EV}_{\text{av}} = \text{AV}_{\text{box}} = 4$ and $\text{SE}_{\text{av}} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{4}{\sqrt{100}} = 0.4.$
- We expect the average of the draws to be around 4 give or take 0.4 or so.
- What is the chance that the average will be between 3.5 and 4.5?
- Compute the z scores for 61% and 65%:
  $z_1 = (3.5 - 4)/0.4 = -1.25$
  $z_2 = (4.5 - 4)/0.4 = 1.25$

Example: Box Contents Known

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

$$\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}\ \boxed{7}}$$

- $\text{AV}_{\text{box}} = \frac{1+2+\cdots+7}{7} = 4$ and $\text{SD}_{\text{box}} = \sqrt{\frac{(1-4)^2+(2-4)^2+\cdots+(7-4)^2}{7}} = \sqrt{\frac{28}{7}} = \sqrt{4} = 2.$
- $\text{EV}_{\text{av}} = \text{AV}_{\text{box}} = 4$ and $\text{SE}_{\text{av}} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{4}{\sqrt{100}} = 0.4.$
- We expect the average of the draws to be around 4 give or take 0.4 or so.
- What is the chance that the average will be between 3.5 and 4.5?
- Compute the z scores for 61% and 65%:
  $z_1 = (3.5 - 4)/0.4 = -1.25$
  $z_2 = (4.5 - 4)/0.4 = 1.25$
- Area under the normal curve: $\texttt{pnorm}(1.25) - \texttt{pnorm}(-1.25) = 0.789 = 78.9\%$

Example: Box Contents Unknown

Example: Box Contents Unknown

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

$$\boxed{\boxed{?}\ \cdots\ \boxed{?}}$$

We don't know what numbers are in the box!

Example: Box Contents Unknown

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

$$\boxed{\boxed{?} \cdots \boxed{?}}$$

  We don't know what numbers are in the box!

- Suppose our sample of size 100 has an average of 10.2 and an SD of 0.8.

Example: Box Contents Unknown

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

$$\boxed{\boxed{?} \cdots \boxed{?}}$$

  We don't know what numbers are in the box!
- Suppose our sample of size 100 has an average of 10.2 and an SD of 0.8.
- Estimate the true average of the tickets to be 10.2 but also give a plus or minus on the estimate.

Example: Box Contents Unknown

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

$$\boxed{\boxed{?} \cdots \boxed{?}}$$

  We don't know what numbers are in the box!
- Suppose our sample of size 100 has an average of 10.2 and an SD of 0.8.
- Estimate the true average of the tickets to be 10.2 but also give a plus or minus on the estimate.
- Estimate $\mathrm{SD}_{\mathrm{box}} = 0.8$.
- $\mathrm{SE}_{\mathrm{av}} = \frac{\mathrm{SD}_{\mathrm{box}}}{\sqrt{n}} \approx \frac{0.8}{\sqrt{100}} = 0.08.$

Example: Box Contents Unknown

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

$$\boxed{\boxed{?} \cdots \boxed{?}}$$

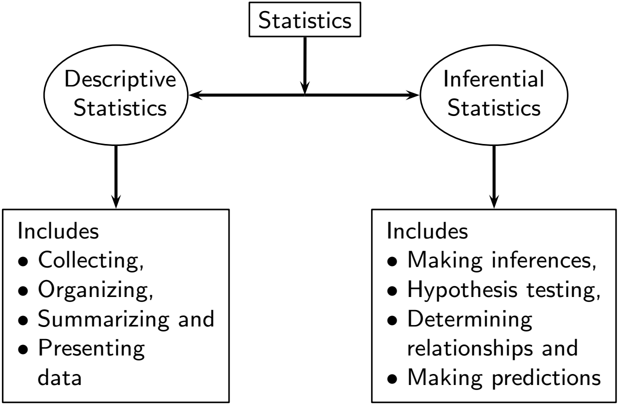
  We don't know what numbers are in the box!
- Suppose our sample of size 100 has an average of 10.2 and an SD of 0.8.
- Estimate the true average of the tickets to be 10.2 but also give a plus or minus on the estimate.
- Estimate $\text{SD}_{\text{box}} = 0.8$.
- $\text{SE}_{\text{av}} = \frac{\text{SD}_{\text{box}}}{\sqrt{n}} \approx \frac{0.8}{\sqrt{100}} = 0.08$.
- If we assume that the samples follow a normal curve, then

$$10.2 \pm 2 \cdot (0.08) = 10.2 \pm 0.16$$

  is 95% confidence interval on the average of the box.

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.

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

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.

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

- 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.

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

- Take a random sample of size n.
- Compute the average and SD of the sample.

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.

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

- Take a random sample of size n.
- Compute the average and SD of the sample.
- Use the SD of the sample to approximate the SD 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.
- This gives an estimate of the average and puts a plus-or-minus on the estimate.

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

- Take a random sample of size n.
- Compute the average and SD of the sample.
- Use the SD of the sample to approximate the SD of the box.
- Then compute the $\mathrm{SE}_{\mathrm{av}}$

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.

$$EV_{av} = AV_{box} \qquad SE_{av} = \frac{SD_{box}}{\sqrt{n}}$$

- Take a random sample of size n.
- Compute the average and SD of the sample.
- Use the SD of the sample to approximate the SD of the box.
- Then compute the SE_{av}
- The 95% confidence interval is:

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

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.

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

- Take a random sample of size n.
- Compute the average and SD of the sample.
- Use the SD of the sample to approximate the SD of the box.
- Then compute the SE_{av}
- The 95% confidence interval is:

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

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

Example Page 415

- A city has 25,000 residents.

Example Page 415

- A city has 25,000 residents.
- Estimate the average income of the residents by taking a random sample.

Example Page 415

- A city has 25,000 residents.
- Estimate the average income of the residents by taking a random sample.
- Box model:
  - 25,000 numbered tickets
  - Each ticket has the income of one resident.

Example Page 415

- A city has 25,000 residents.
- Estimate the average income of the residents by taking a random sample.
- Box model:
  - 25,000 numbered tickets
  - Each ticket has the income of one resident.
- Observed average: $\$62{,}396{,}714/1000 = \$62,400$.

Example Page 415

- A city has 25,000 residents.
- Estimate the average income of the residents by taking a random sample.
- Box model:
  - 25,000 numbered tickets
  - Each ticket has the income of one resident.
- Observed average: $\$62{,}396{,}714/1000 = \$62,400$.
- SD of the sample: $53,000 (given).
- Estimated average income: $62,400.

Example Page 415

- A city has 25,000 residents.
- Estimate the average income of the residents by taking a random sample.
- Box model:
  - 25,000 numbered tickets
  - Each ticket has the income of one resident.
- Observed average: $\$62{,}396{,}714/1000 = \$62,400$.
- SD of the sample: \$53,000 (given).
- Estimated average income: \$62,400.
- Now compute a $\pm$ on this estimate:
  - $\text{SD}_{\text{box}} \approx \$53{,}000$.
  - $\text{SE}_{\text{av}} = \$53{,}000/\sqrt{1000} \approx \$1{,}700$.

Example Page 415

- A city has 25,000 residents.
- Estimate the average income of the residents by taking a random sample.
- Box model:
  - 25,000 numbered tickets
  - Each ticket has the income of one resident.
- Observed average: $\$62{,}396{,}714/1000 = \$62,400$.
- SD of the sample: \$53,000 (given).
- Estimated average income: \$62,400.
- Now compute a $\pm$ on this estimate:
  - $\text{SD}_{\text{box}} \approx \$53{,}000$.
  - $\text{SE}_{\text{av}} = \$53{,}000/\sqrt{1000} \approx \$1{,}700$.
- 95% confidence interval:

$$\$53{,}000 \ \pm \ 2 \cdot (1{,}700) = \$53{,}000 \ \pm \$3{,}400$$