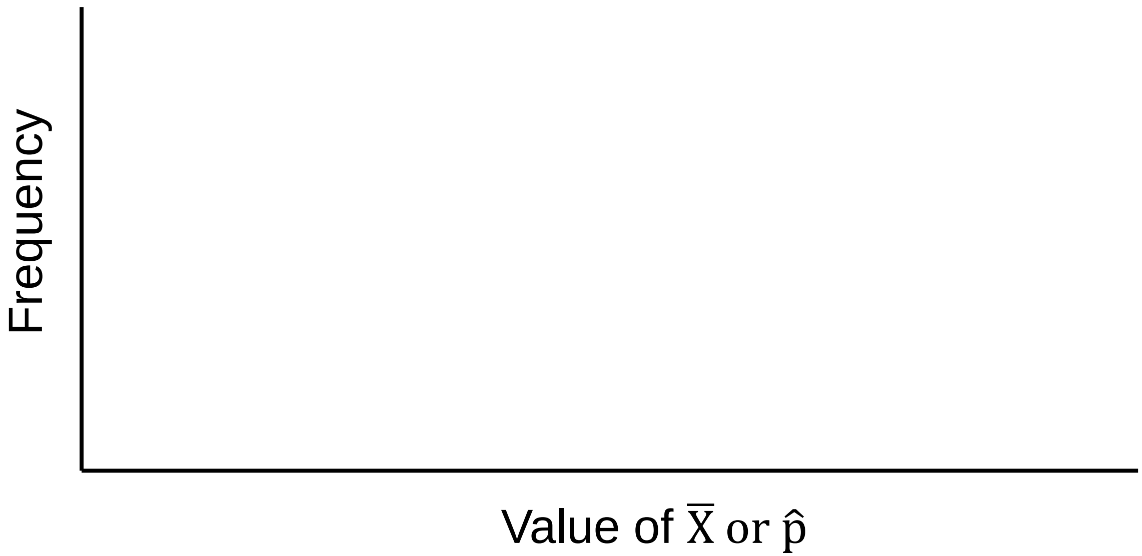
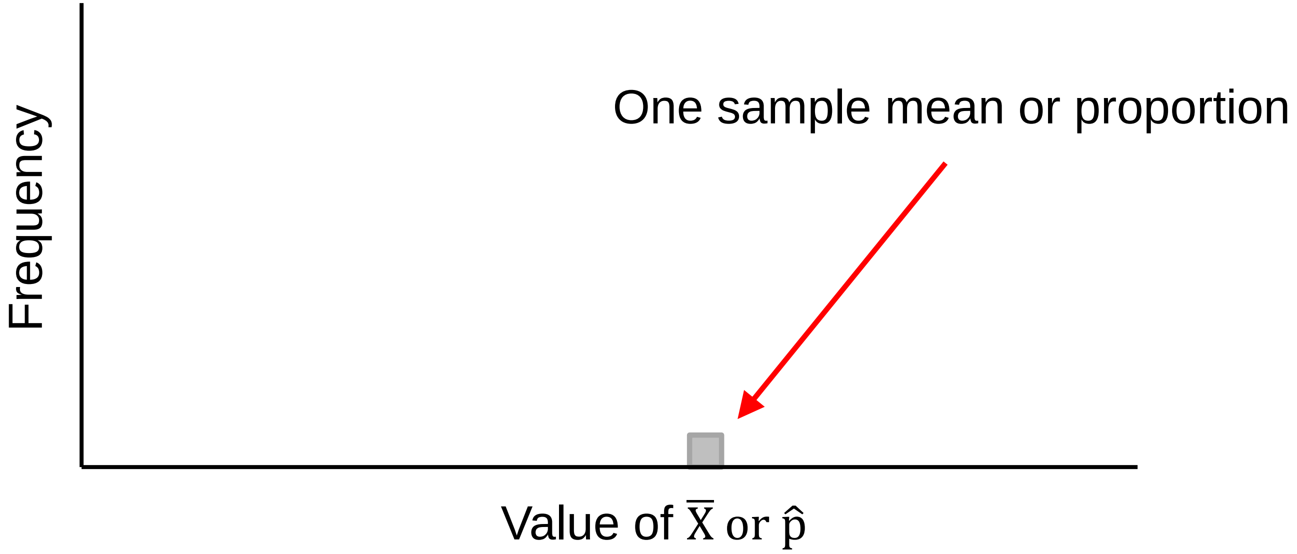
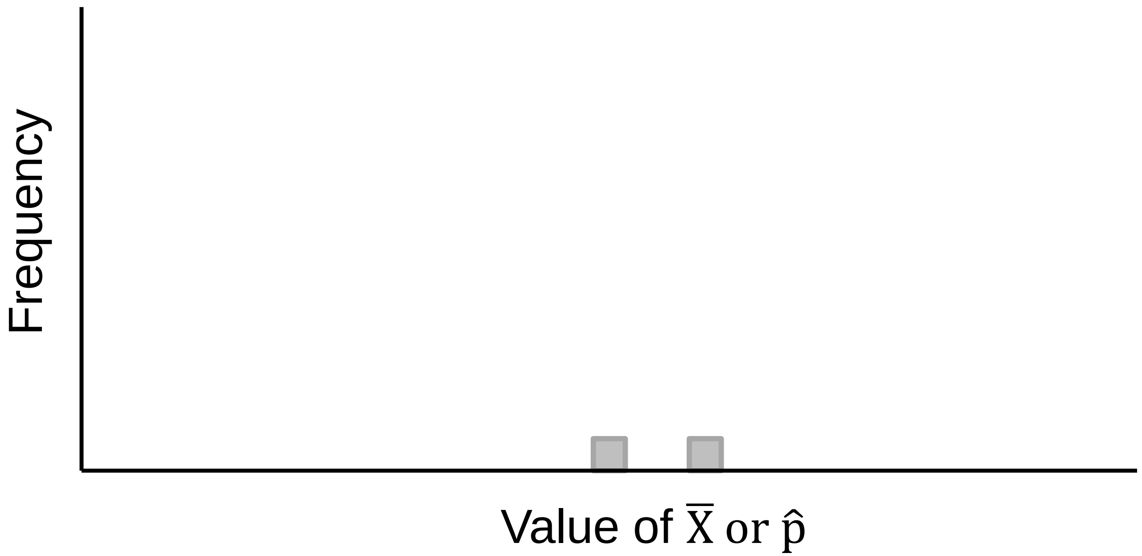
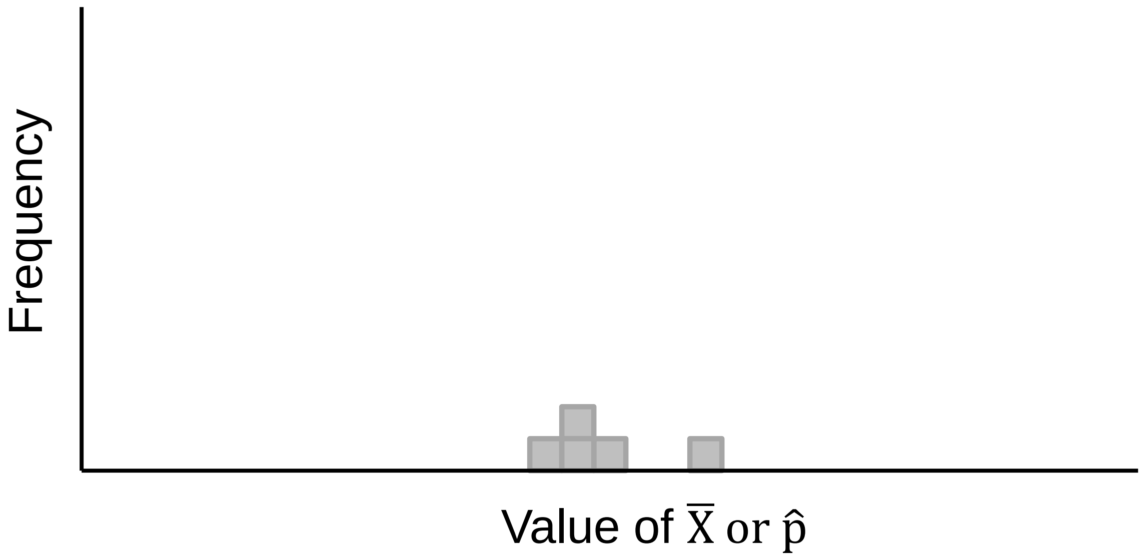
SOC 3811/5811: BASIC SOCIAL STATISTICS

Confidence Intervals

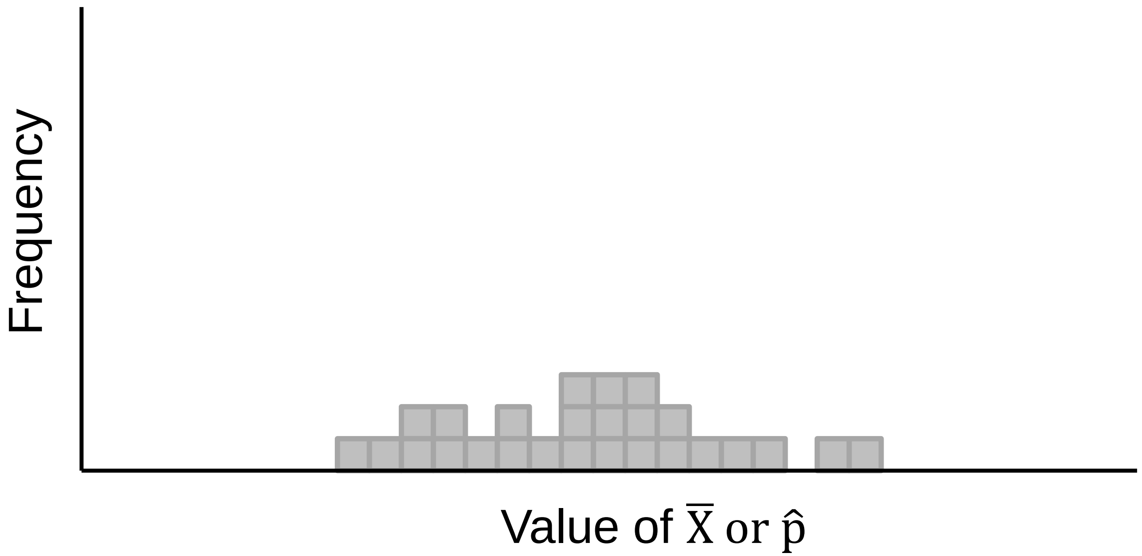

Frequency
Value of $\bar{X}$ or $\hat{p}$


One sample mean or proportion
Frequency
Value of $\overline{X}$ or $\hat{p}$

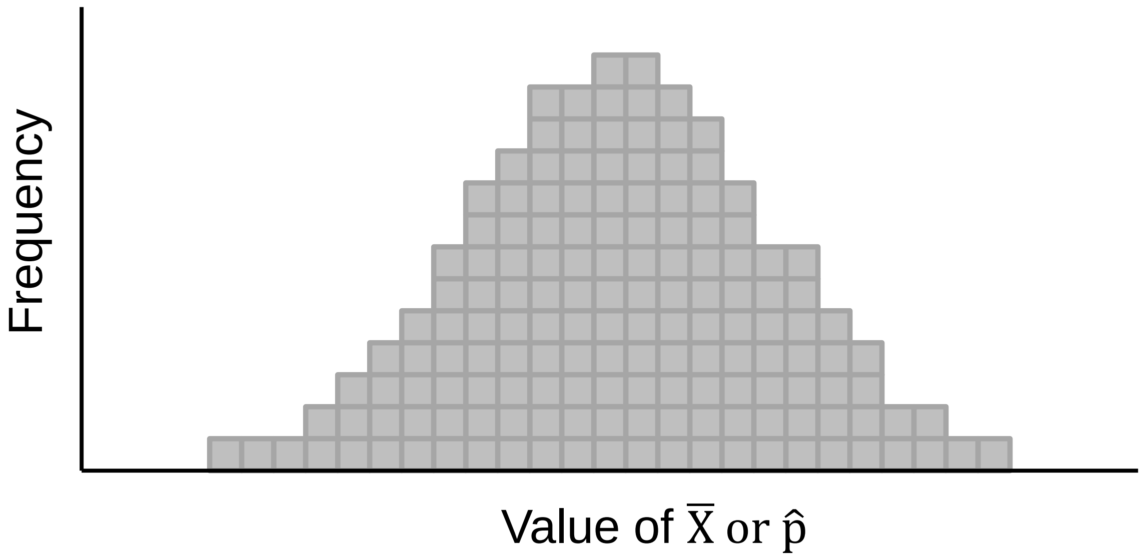

Frequency
Value of $\bar{X}$ or $\hat{p}$

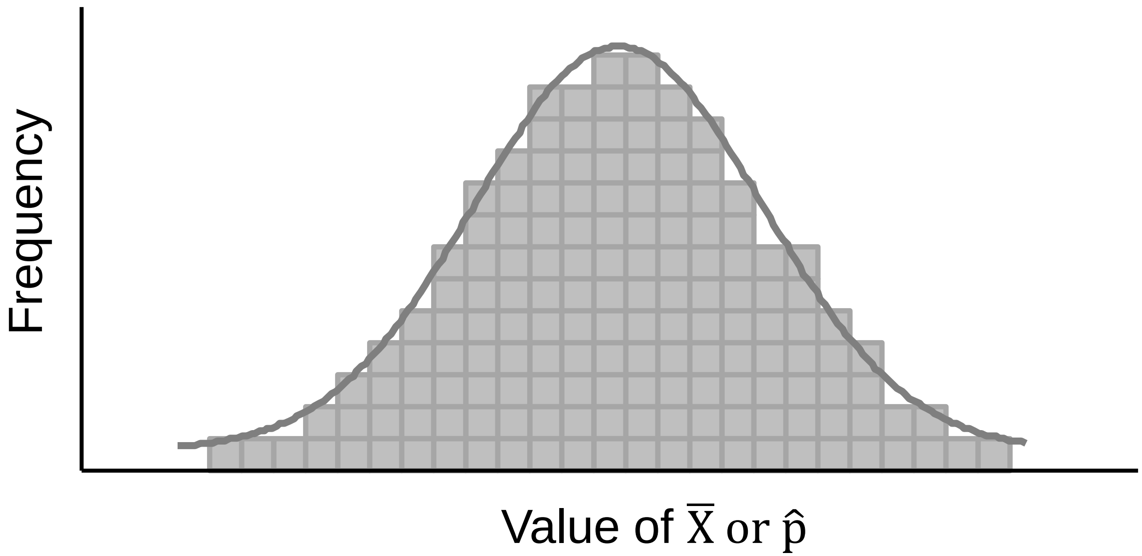


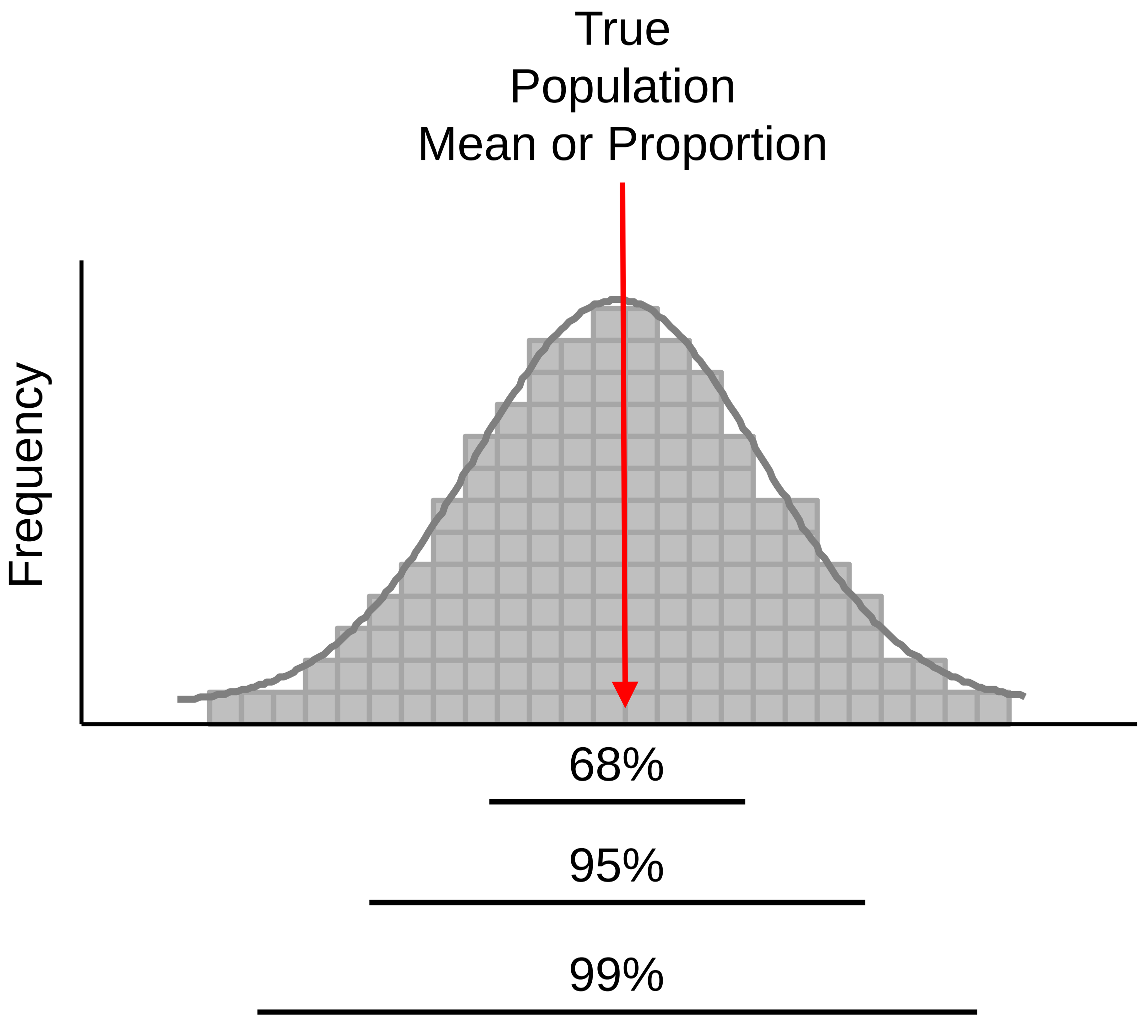
Frequency
Value of $\overline{X}$ or $\hat{p}$

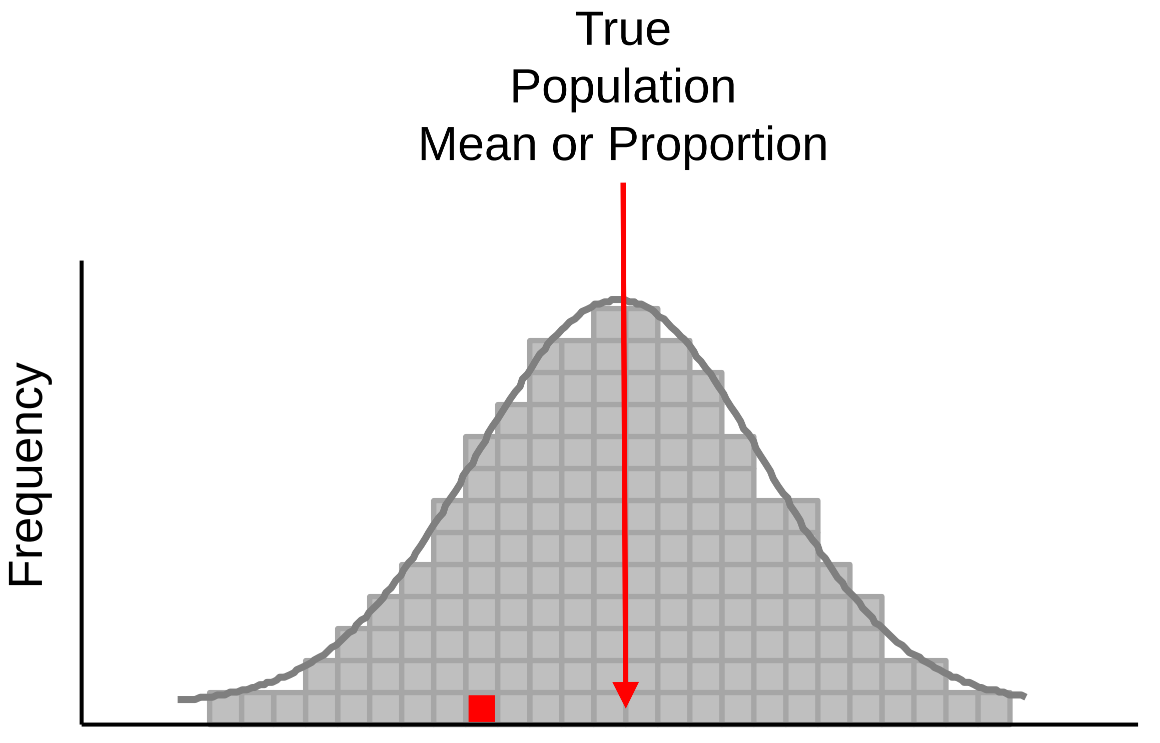


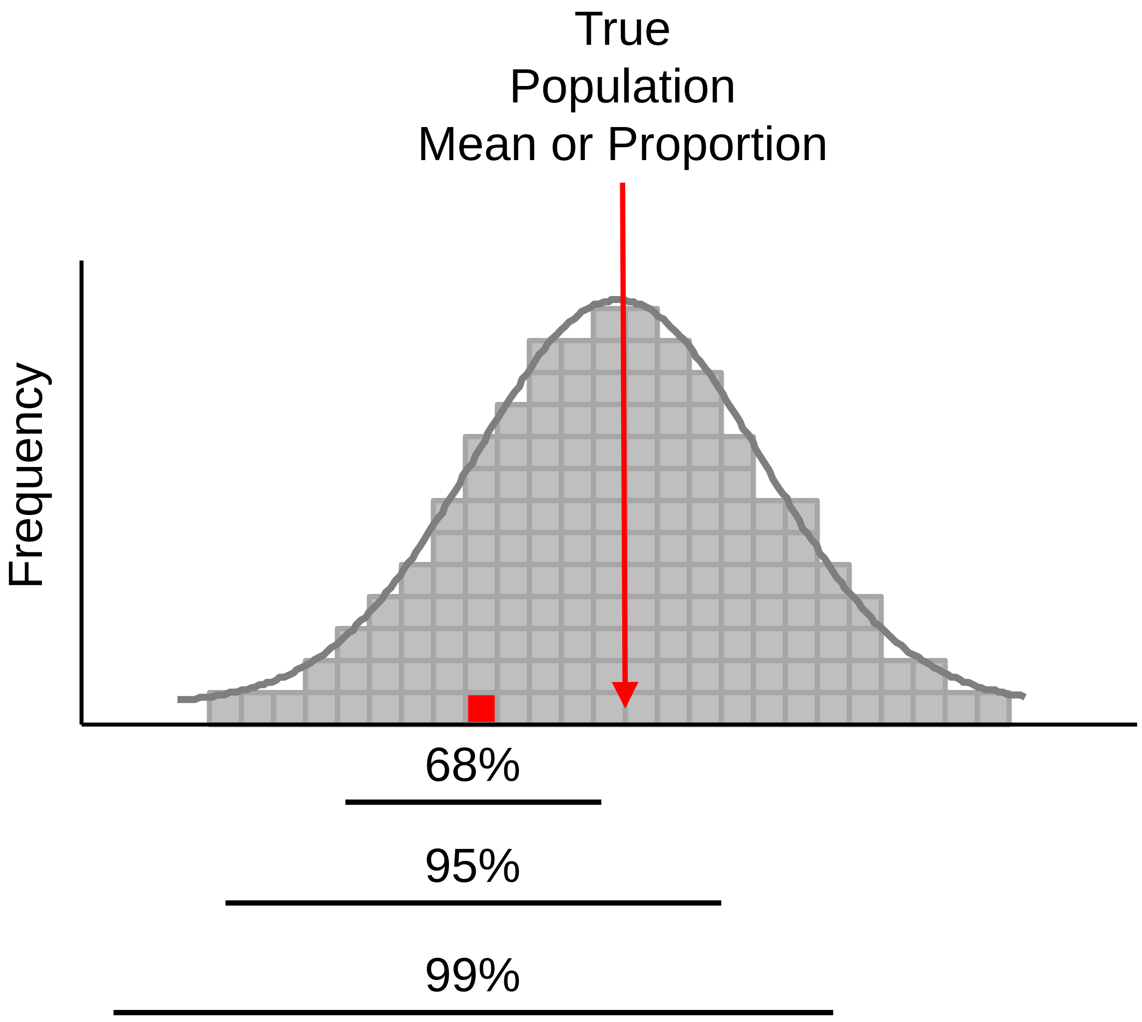
Frequency
Value of $\bar{X}$ or $\hat{p}$


Frequency
Value of $\overline{X}$ or $\hat{p}$


Frequency
Value of $\overline{X}$ or $\hat{p}$


True
Population
Mean or Proportion
Frequency
68%
95%
99%


True
Population
Mean or Proportion
Frequency


True
Population
Mean or Proportion
Frequency
68%
95%
99%

Confidence Intervals

Confidence Interval

A range of values that is "likely" to contain the population parameter (for example, a mean or proportion)

Just how "likely" it is that the confidence intervals contains the population proportion is called the *confidence level*

If our confidence level is C%, then we are saying that if we drew many random samples and computed many confidence intervals, then the true population parameter would be contained within the resulting confidence intervals C% of the time

Confidence Intervals

We will consider four different sorts of confidence intervals, all of which follow the same logic

Confidence Intervals for <u>Proportions</u>

Use $\hat{p}$ to infer p, the population proportion

Confidence Intervals for <u>Means</u>

Use $\bar{Y}$ to infer μ_Y, the population mean of Y

Confidence Intervals for <u>Differences in Proportions</u>

Use $\hat{p}_1 - \hat{p}_2$ to infer the difference between two population proportions, p_1 and p_2

Confidence Intervals for <u>Differences in Means</u>

Use $\bar{Y}_1 - \bar{Y}_2$ to infer the difference between two population means, μ_{Y1} and μ_{Y2}

Confidence Intervals

All confidence intervals can be written generally as:

$$\text{Sample Estimate} \pm \text{Multiplier} \times \text{Standard Error}$$

Or:

$$\overline{\mathrm{Y}} \pm (\mathrm{Z}_{\alpha/2})(\sigma_{\overline{\mathrm{Y}}})$$

Confidence Intervals

$$\text{Sample Estimate} \pm \text{Multiplier} \times \text{Standard Error}$$

For **proportions**:

$$se_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

For **means**:

$$se_{\bar{Y}} = \frac{s_Y}{\sqrt{n}}$$

For **differences in proportions**:

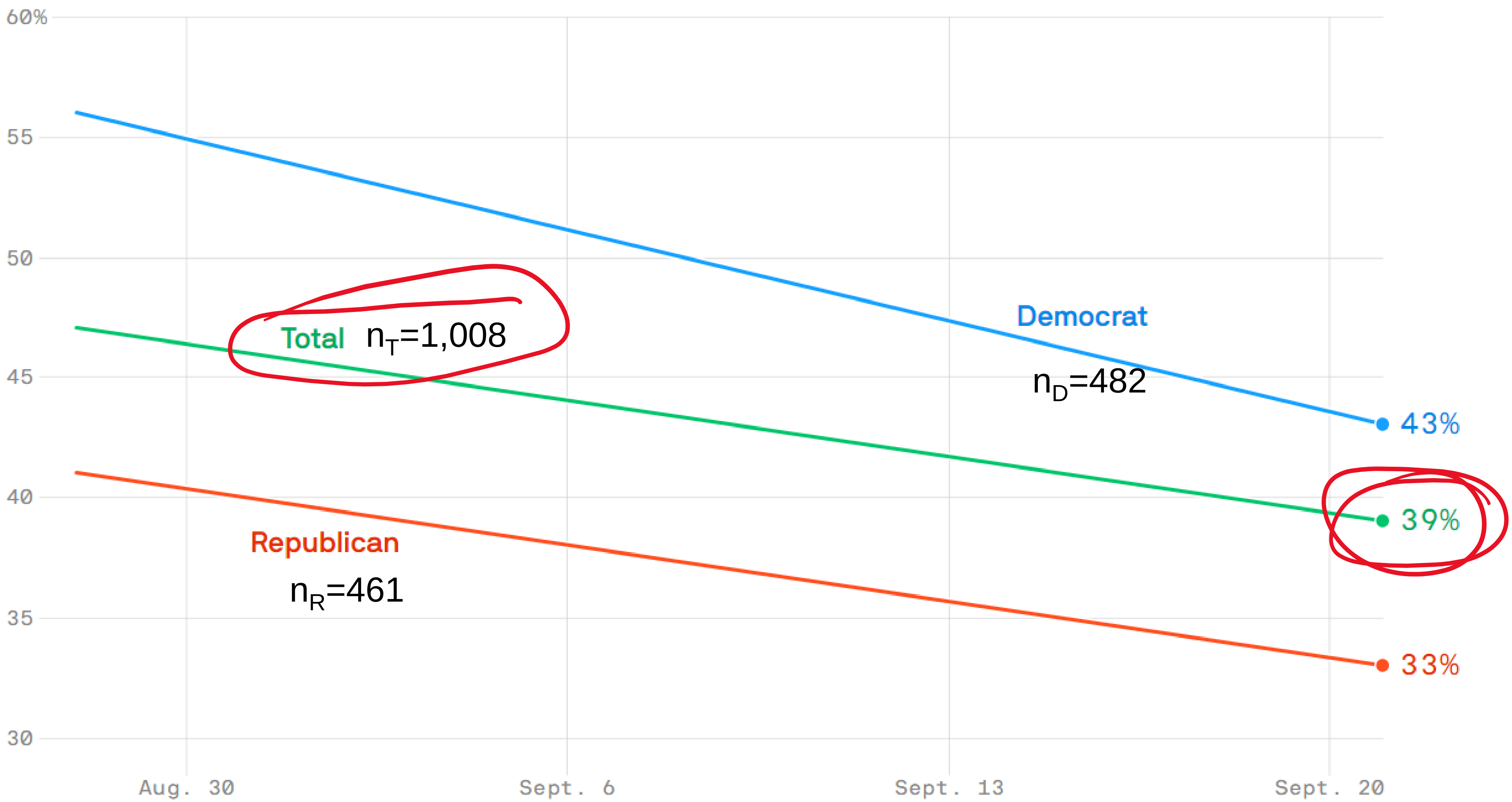
$$se_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

For **differences in means**:

$$se_{Y-X} = \sqrt{\frac{s_Y^2}{n_Y} + \frac{s_X^2}{n_X}}$$

Percent who say they are likely to get the first generation COVID-19 vaccine as soon as it is available

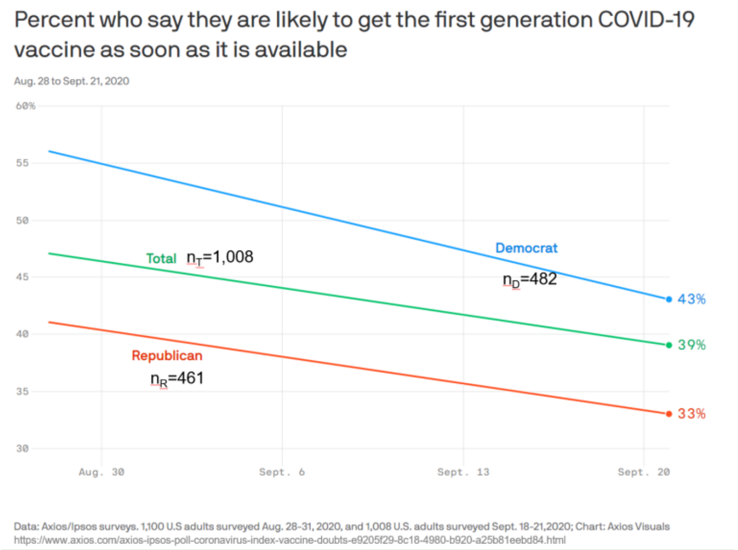
Aug. 28 to Sept. 21, 2020



Data: Axios/Ipsos surveys. 1,100 U.S adults surveyed Aug. 28-31, 2020, and 1,008 U.S. adults surveyed Sept. 18-21,2020; Chart: Axios Visuals

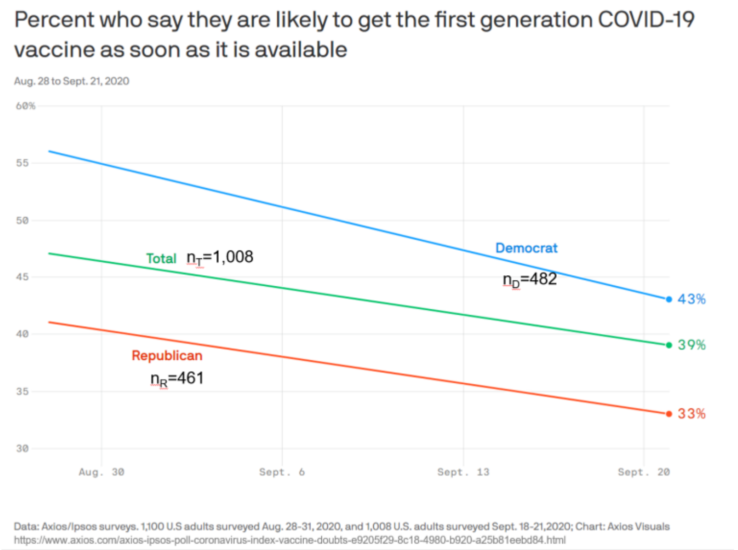
https://www.axios.com/axios-ipsos-poll-coronavirus-index-vaccine-doubts-e9205f29-8c18-4980-b920-a25b81eebd84.html

Construct an 80% confidence interval for the percentage of the total U.S. population that says they are likely to get the first generation COVID-19 vaccine as soon as it is available



Construct a 95% confidence interval for the percentage of the total U.S. population that says they are likely to get the first generation COVID-19 vaccine as soon as it is available

Construct an 95% confidence interval for the difference between Democrats and Republicans in the percentage of people Who say they are likely to get the first generation COVID-19 vaccine as soon as it is available



Construct an 80% confidence interval for the percentage of the total U.S. population that says they are likely to get the first generation COVID-19 vaccine as soon as it is available

$$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$.39 \pm 1.28\sqrt{\frac{.39(.61)}{1008}}$$

$$.39 \pm .02$$

$$.37 - .41$$

Construct a 95% confidence interval for the percentage of the total U.S. population that says they are likely to get the first generation COVID-19 vaccine as soon as it is available

$$.39 \pm z\sqrt{\frac{.39(.61)}{1008}}$$

$$.39 \pm 1.96\sqrt{\frac{.39(.61)}{1108}}$$

$$.39 \pm .03$$

$$.36 - .42$$

Construct an 80% confidence interval for the percentage of the total U.S. population that says they are likely to get the first generation COVID-19 vaccine as soon as it is available

$$\text{Sample Estimate} \pm \text{Multiplier} \times \text{Standard Error}$$

$$\hat{p} \pm 1.28\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .39 \pm 1.28\sqrt{\frac{0.39(0.61)}{1008}} = .39 \pm 0.02$$

Construct a 95% confidence interval for the percentage of the total U.S. population that says they are likely to get the first generation COVID-19 vaccine as soon as it is available

$$\text{Sample Estimate} \pm \text{Multiplier} \times \text{Standard Error}$$

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .39 \pm 1.96\sqrt{\frac{0.39(0.61)}{1008}} = .39 \pm 0.03$$

So, with 95% certainty, the population percentage is between 36% and 42%

Construct an 95% confidence interval for the difference between Democrats and Republicans in the percentage of people Who say they are likely to get the first generation COVID-19 vaccine as soon as it is available

$\hat{p}_D = .43 \quad n_D = 482$

$\hat{p}_R = .33 \quad n_R = 461$

$$\hat{p}_D - \hat{p}_R \pm z\sqrt{\frac{\hat{p}_D(1-\hat{p}_D)}{n_D} + \frac{\hat{p}_R(1-\hat{p}_R)}{n_R}}$$

$$.43 - .33 \pm 1.96\sqrt{\frac{.43(.57)}{482} + \frac{.33(.67)}{461}}$$

$$.10 \pm .062$$

$$.038 \text{ — } .162$$

$$3.8\% \text{ — } 16.2\%$$

Construct an 95% confidence interval for the difference between Democrats and Republicans in the percentage of people Who say they are likely to get the first generation COVID-19 vaccine as soon as it is available

$$\text{Sample Estimate} \pm \text{Multiplier} \times \text{Standard Error}$$

$$\widehat{p_D} - \widehat{p_R} \pm 1.96\sqrt{\frac{\widehat{p_D}(1-\widehat{p_D})}{\mathrm{n_D}} + \frac{\widehat{p_R}(1-\widehat{p_R})}{\mathrm{n}_R}}$$

$$0.43 - 0.33 \pm 1.96\sqrt{\frac{(0.43)(0.57)}{482} + \frac{(0.33)(0.67)}{461}}$$

$$0.10 \pm 0.062$$

So, with 95% certainty the difference between the two groups is between 3.8% and 16.2%

Construct a 95% confidence interval for the mean (average) length of the cobs of Gopher Corn (a variety of corn).

Your random sample of 100 cobs of the Gopher Corn variety averages 8.5 inches with a standard deviation of 0.35 inches.

$n = 100$

$\bar{x} = 8.5$

$s = 0.35$

$\bar{x} \pm t \frac{s}{\sqrt{n}}$

$8.5 \pm 1.984 \frac{.35}{\sqrt{100}}$

$8.5 \pm .07$

8.43" ~ 8.57"

Construct a 95% confidence interval for the difference between the mean (average) length of the cobs of two varieties of sweet corn (allowing them to grow the same number of days under the same conditions).

Call the two varieties Gopher Corn and Badger Corn. Your random sample of 100 cobs of the Gopher Corn variety averages 8.5 inches with a standard deviation of 0.35 inches; your random sample of 110 cobs of Badger Corn averages 7.5 inches with a standard deviation of 0.45.

Construct a 95% confidence interval for the mean (average) length of the cobs of Gopher Corn (a variety of corn).

Your random sample of 100 cobs of the Gopher Corn variety averages 8.5 inches with a standard deviation of 0.35 inches.

Construct a 95% confidence interval for the mean (average) length of the cobs of Gopher Corn (a variety of corn).

Your random sample of 100 cobs of the Gopher Corn variety averages 8.5 inches with a standard deviation of 0.35 inches.

$$\text{Sample Estimate} \pm \text{Multiplier} \times \text{Standard Error}$$

$$\bar{x} \pm 2.00 \frac{s_Y}{\sqrt{n}}$$

$$8.5 \pm 2.00 \frac{0.35}{\sqrt{100}}$$

8.5 ± 0.07

Thus, we are 95% confidence that the population mean length of Gopher Corn is between 8.43 and 8.57 inches

Construct a 95% confidence interval for the difference between the mean (average) length of the cobs of two varieties of sweet corn (allowing them to grow the same number of days under the same conditions).

Call the two varieties Gopher Corn and Badger Corn. Your random sample of 100 cobs of the Gopher Corn variety averages 8.5 inches with a standard deviation of 0.35 inches; your random sample of 110 cobs of Badger Corn averages 7.5 inches with a standard deviation of 0.45.

Construct a 95% confidence interval for the difference between the mean (average) length of the cobs of two varieties of sweet corn (allowing them to grow the same number of days under the same conditions).

Call the two varieties Gopher Corn and Badger Corn. Your random sample of 100 cobs of the Gopher Corn variety averages 8.5 inches with a standard deviation of 0.35 inches; your random sample of 110 cobs of Badger Corn averages 7.5 inches with a standard deviation of 0.45.

Sample Estimate $\pm$ Multiplier $\times$ Standard Error

$$\overline{X_G} - \overline{X_B} \pm 2.0\sqrt{\frac{s_G^2}{n_G} + \frac{s_B^2}{n_B}}$$

$Z_{\alpha/2}$

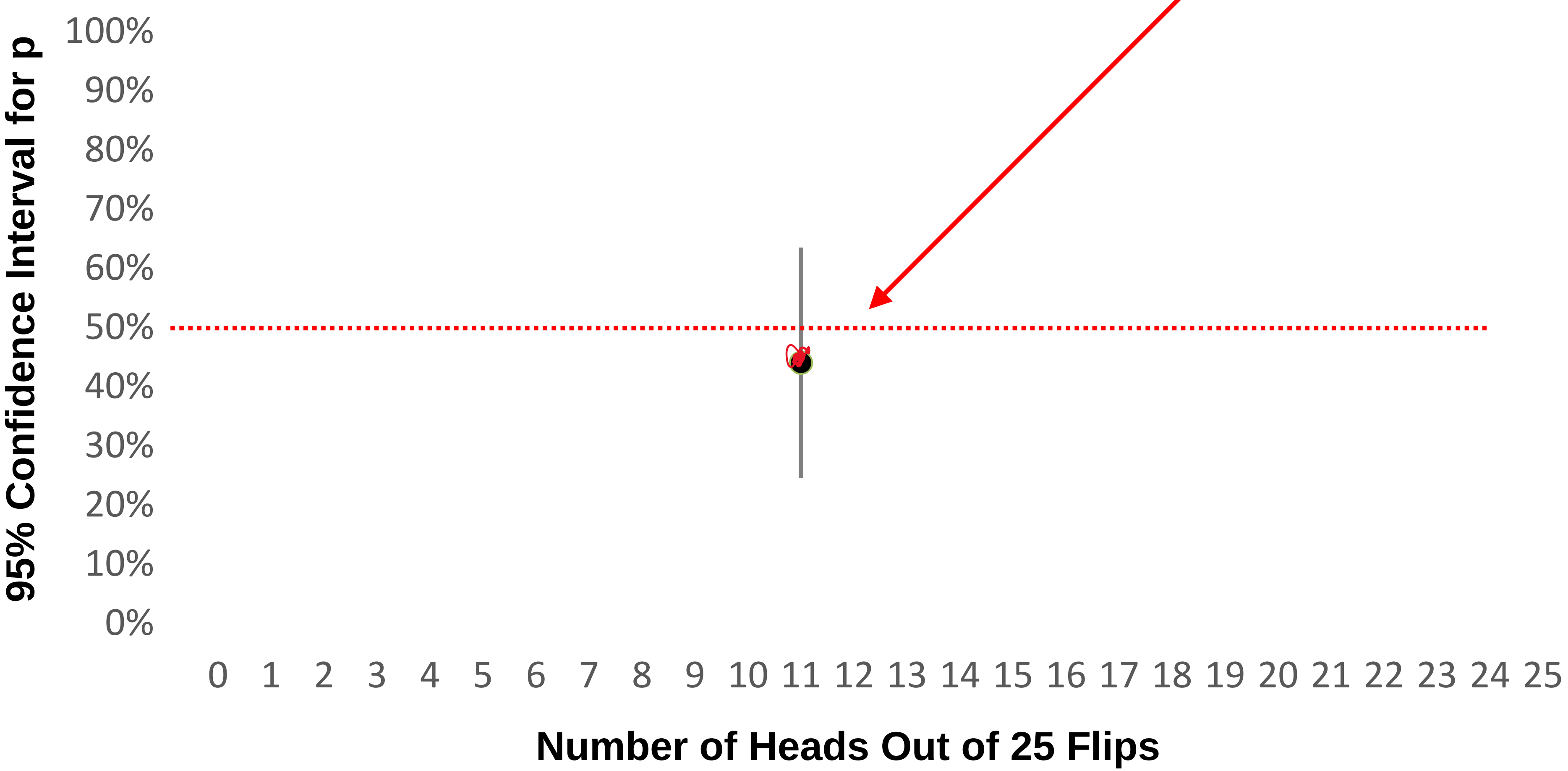
$$8.5 - 7.5 \pm 2.0\sqrt{\frac{0.35^2}{100} + \frac{0.45^2}{110}}$$

$$1.00 \pm 0.12$$

So, we are 95% confident that the population mean difference is such that Gopher Corn is between 0.88 and 1.12 inches longer than Badger Corn.

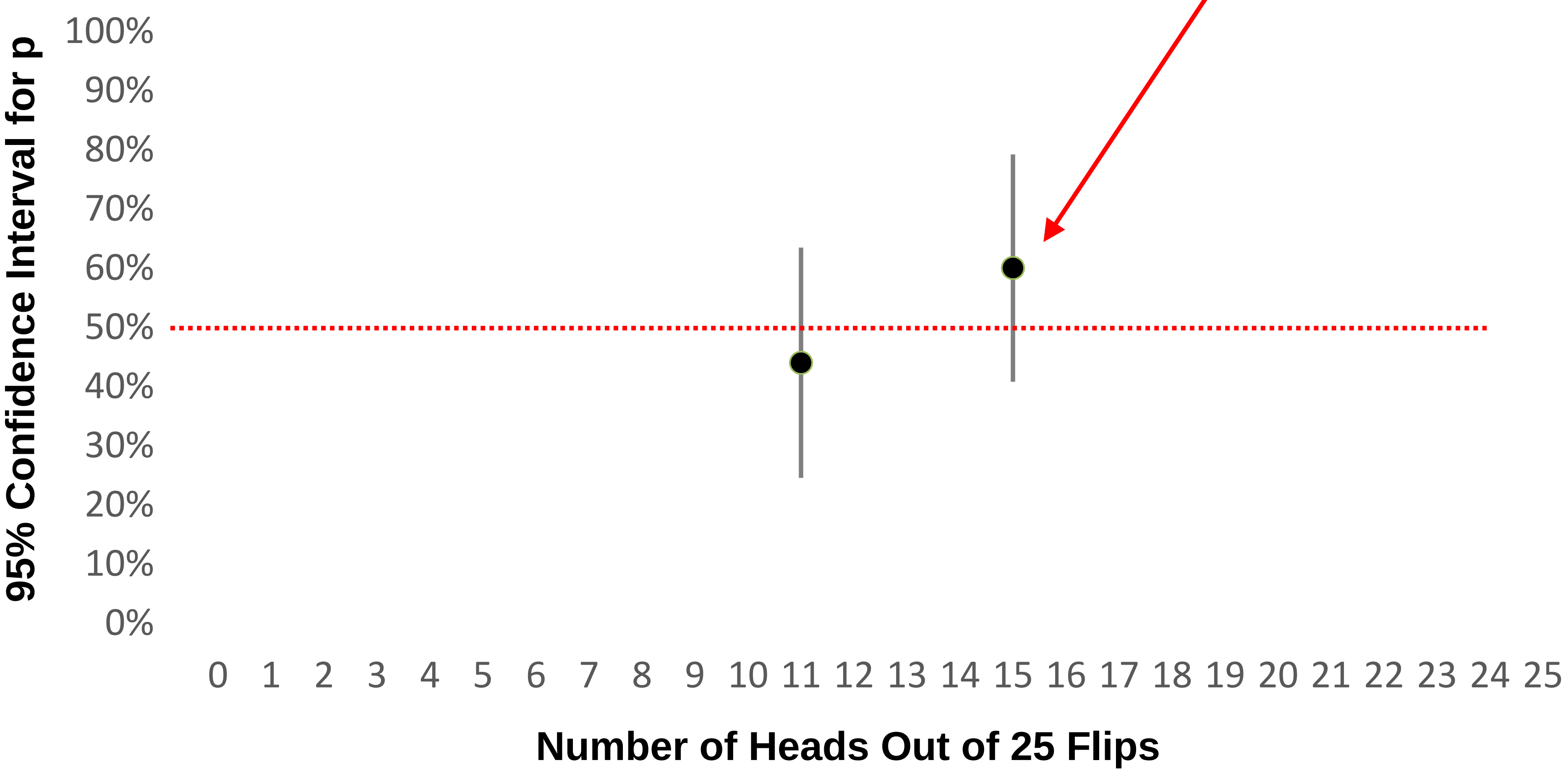
For 11 heads out of 25:

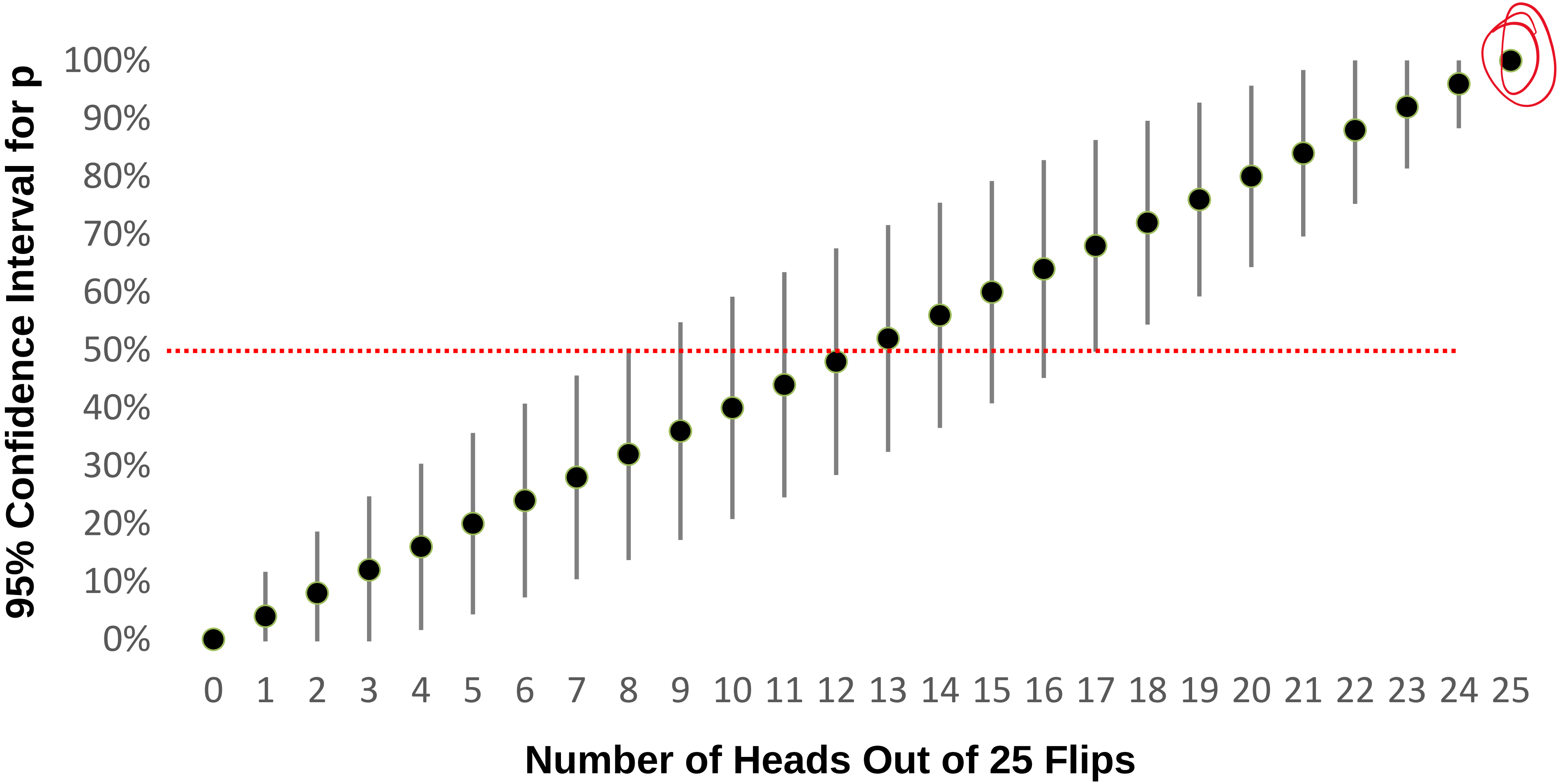
$$95\%\ \text{C.I.} = \hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.44 \pm 1.96\sqrt{\frac{0.44(0.56)}{25}} = 0.44 \pm 0.19, \text{so } 25\% \text{ to } 63\%$$

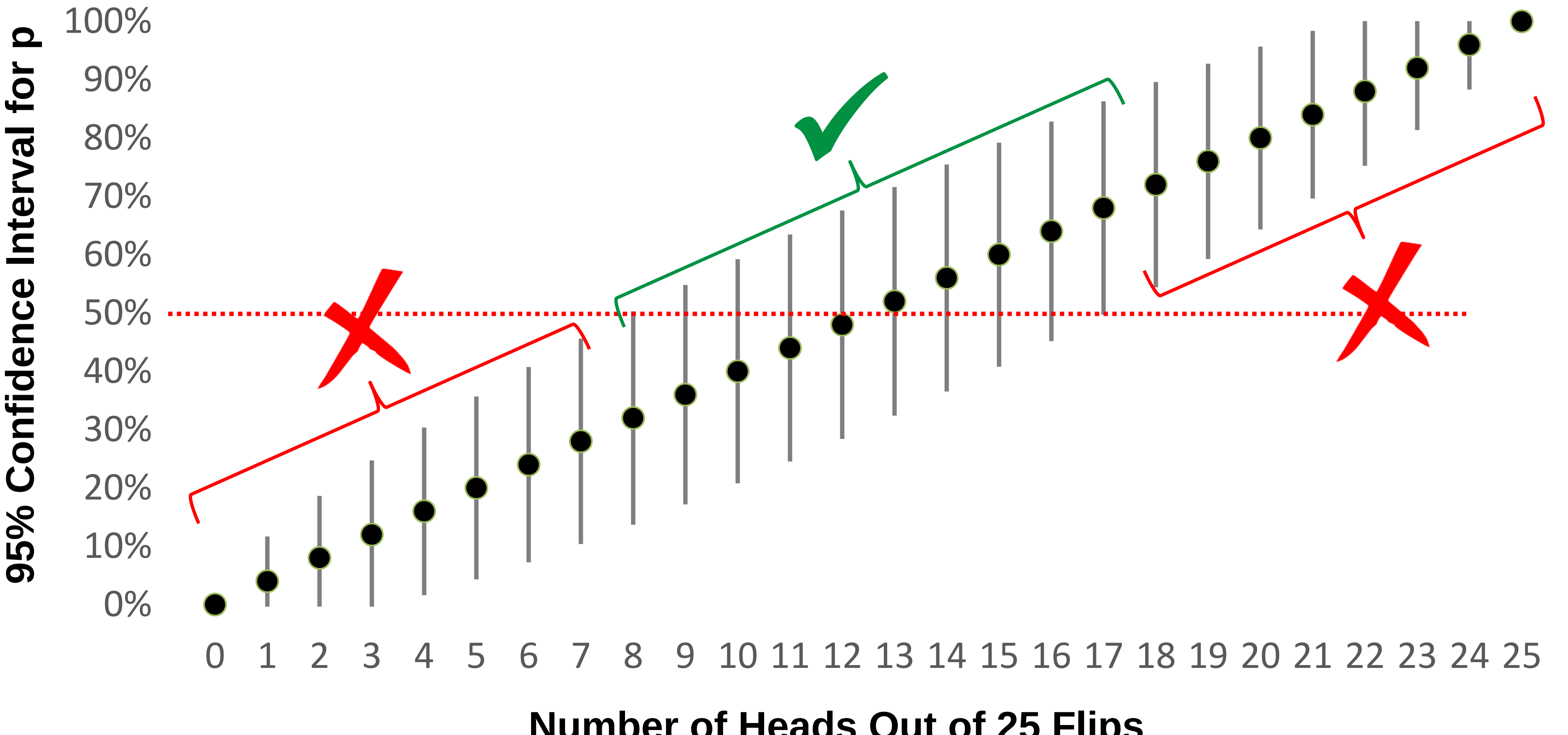


For 15 heads out of 25:

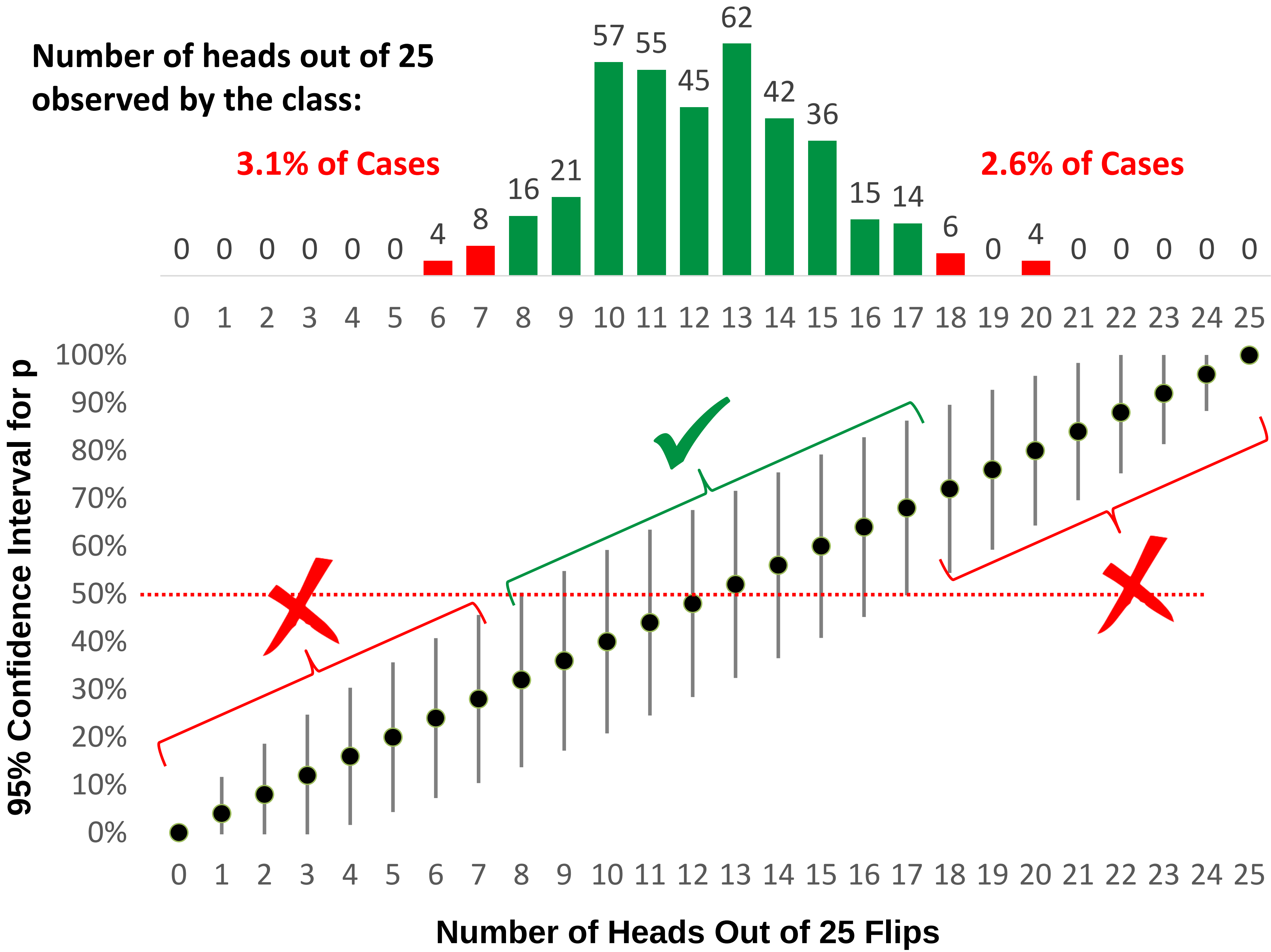
$$95\%\ \text{C.I.} = \hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.60 \pm 1.96\sqrt{\frac{0.6(0.4)}{25}} = 0.60 \pm 0.19, \text{so } 41\% \text{ to } 79\%$$




95% Confidence Interval for p
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Number of Heads Out of 25 Flips



95% Confidence Interval for p
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Number of Heads Out of 25 Flips


Number of heads out of 25 observed by the class:
3.1% of Cases
2.6% of Cases
0 0 0 0 0 0 4 8 16 21 57 55 45 62 42 36 15 14 6 0 4 0 0 0 0 0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
95% Confidence Interval for p
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Number of Heads Out of 25 Flips