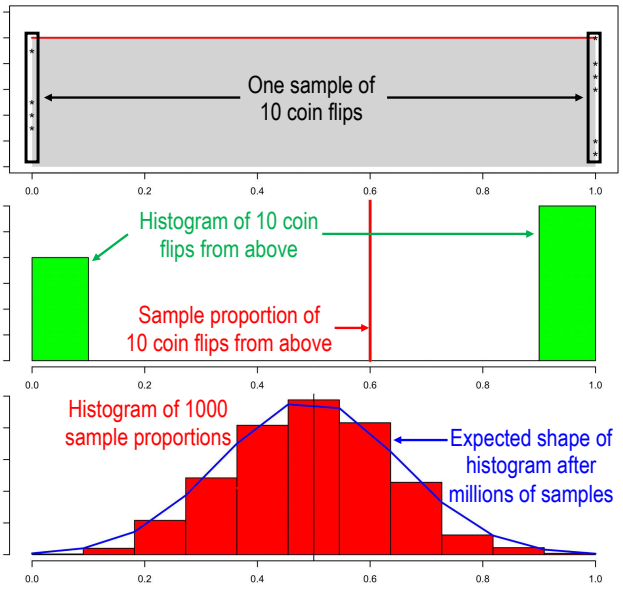
Inference for a Population Proportion

- Sampling Distribution of the Sample Proportion
- One-Sample Proportion Test

University of Pittsburgh

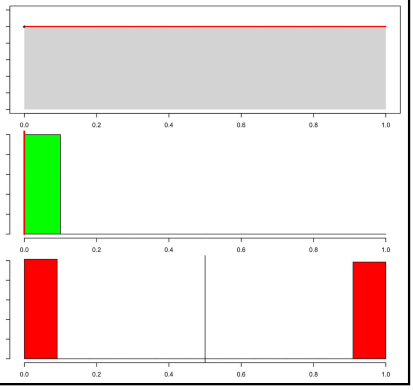
Simulation: Sampling Distribution

- **Scenario:** Flip a fair coin a given number of times, calculate the sample proportion, and plot this new sample proportion in a histogram. Repeat this 1000 times.
- **Goal:** Understand the _____ and _____ of this new distribution as we change the sample size.



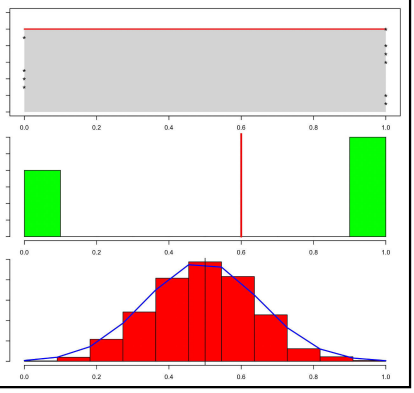
Simulation: Sampling Distributions

- **Scenario:** Flip a fair coin 1 time with "heads" as the success outcome. Calculate the sample proportion of heads.
- **Question:** What are the mean, standard deviation, and shape of the resulting histogram of sample proportions?
- **Answer:**
  - **Average Sample Proportion:** _____
  - **SD of Sample Proportions:** _____
  - **Shape of Sample Proportions:** _____
    - Only _____ sample proportions



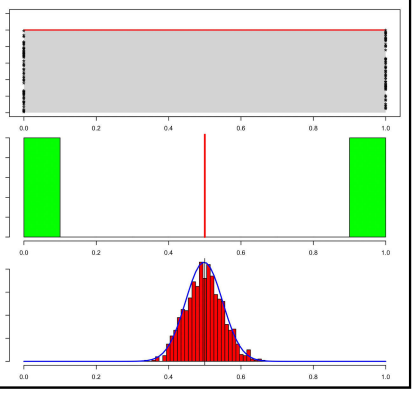
Simulation: Sampling Distributions

- **Scenario:** Flip a fair coin **10** times with "heads" as the success outcome. Calculate the sample proportion of heads.
- **Question:** What are the mean, standard deviation, and shape of the resulting histogram of sample proportions?
- **Answer:**
  - **Average Sample Proportion:** _____
  - **SD of Sample Proportions:** _____
  - **Shape of Sample Proportions:** _____
    - Like the die roll example, the sampling distribution is still not _____ to be considered _____ given how many samples were taken
    - Looks too much like a _____



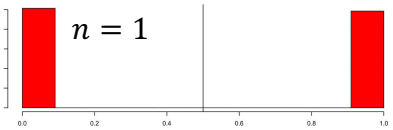
Simulation: Sampling Distributions

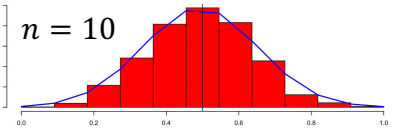
- **Scenario:** Flip a fair coin **100** times with "heads" as the success outcome. Calculate the sample proportion of heads.
- **Question:** What are the mean, standard deviation, and shape of the resulting histogram of sample proportions?
- **Answer:**
  - **Average Sample Proportion:** _____
  - **SD of Sample Proportions:** _____
  - **Shape of Sample Props.:** _____

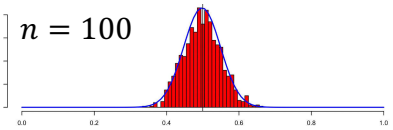


Simulation Discussion

- **Question:** As the sample size increases, what do you notice happening?



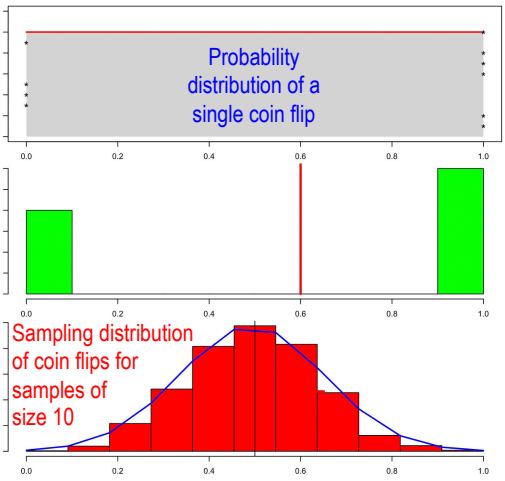




- **Answer:**
  - **Mean:** Average sample proportion _____ (probability of _____ on a fair coin)
    - With larger sample sizes, a flip of "heads" will usually be _____ with a flip of _____
  - **Standard Deviation:** Area that started in the _____
    - Spread of sample proportions became _____
  - **Shape:** Started _____ and became _____

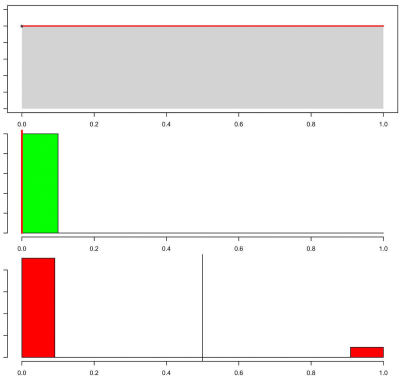
Sampling Distribution of a Sample Proportion

- **Sampling distribution:** For categorical data, the distribution of all possible sample proportions for a given population proportion p, and sample size n
  - Defined by three components:
    1. **Mean**
    2. **Standard error:** Standard deviation of the sampling distribution; measures how spread out the sample means tend to be
    3. **Shape**
  - Like the sampling distribution for a mean, it describes where we can expect a sample proportion to fall in relation to the population proportion.



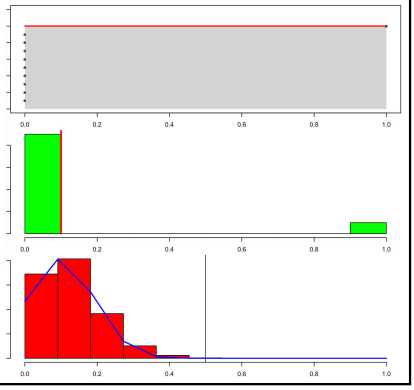
Motivation: Central Limit Theorem for Proportions

- **Scenario:** 10% of the world's population is left-handed. Randomly sample **1** person and calculate the sample proportion of left-handed people.
- **Question:** What are the mean, standard deviation, and shape of the resulting histogram of sample proportions?
- **Answer:**
  - **Average Sample Proportion:** _____
  - **SD of Sample Proportions:** _____
  - **Shape of Sample Proportions:** _____
    - Only two possible sample proportions, but sampling a _____-handed person is much more likely than sampling a _____-handed person



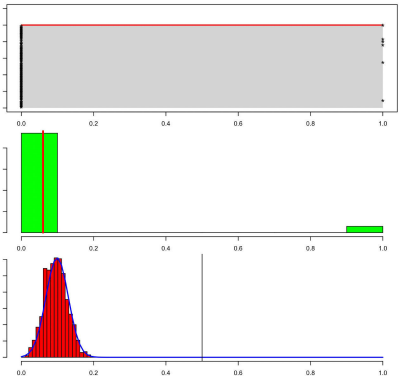
Motivation: Central Limit Theorem for Proportions

- **Scenario:** 10% of the world's population is left-handed. Randomly sample **10** people and calculate the sample proportion of left-handed people.
- **Question:** What are the mean, standard deviation, and shape of the resulting histogram of sample proportions?
- **Answer:**
  - **Average Sample Proportion:** _____
  - **SD of Sample Proportions:** _____
  - **Shape of Sample Proportions:** _____
    - Much more room for a sample proportion to _____ 0.10 than to be _____



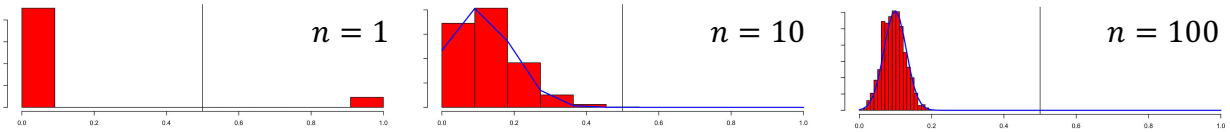
Motivation: Central Limit Theorem for Proportions

- **Scenario:** 10% of the world's population is left-handed. Randomly sample **100** people and calculate the sample proportion of left-handed people.
- **Question:** What are the mean, standard deviation, and shape of the resulting histogram of sample proportions?
- **Answer:**
  - **Average Sample Proportion:** _____
  - **SD of Sample Proportions:** _____
  - **Shape of Sample Props.:** _____



Sampling Distribution Changes

- **Question:** What do you notice about the changes in the sampling distributions as the sample size increased?



- **Answer:**
  - **Mean:** All were nearly _____ to the _____ proportion (_____)
  - **Standard Error:** Became _____
  - **Shape:** Became _____ in time

- ***Takeaway:*** *The sampling distribution of the sample proportion _____ starts out normal, but will _____ become normal; however, the necessary sample size to achieve normality gets _____ as p gets _____.*

Central Limit Theorem for Proportions

- **Central Limit Theorem:** The sampling distribution of the sample proportion approaches a normal distribution as the sample size increases regardless of the shape of the original population.

- ***Rules of Thumb:*** *Take a random sample of size n from some distribution X of categorical data where each trial has success probability p. Then the distribution of the sample proportion $\hat{p}$ is normal if **both** of the following are true:*
  1. $np \geq 10$ *(Expected number of successes is at least 10)*
  2. $n(1-p) \geq 10$ *(Expected number of failures is at least 10)*

Mean and Standard Error of Sample Proportion

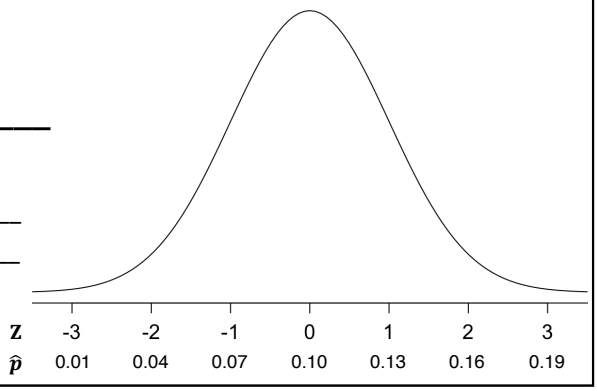
- Sample n observations from **any** population with categorical data that has success probability p. Then:
  1. **Mean:** $\mu_{\hat{p}} = p$
     - Mean of the sampling distribution of $\hat{p}$ equals the population proportion
  2. **Standard Error:** $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

- If the Central Limit Theorem holds, then the sample proportion is standardized as:

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

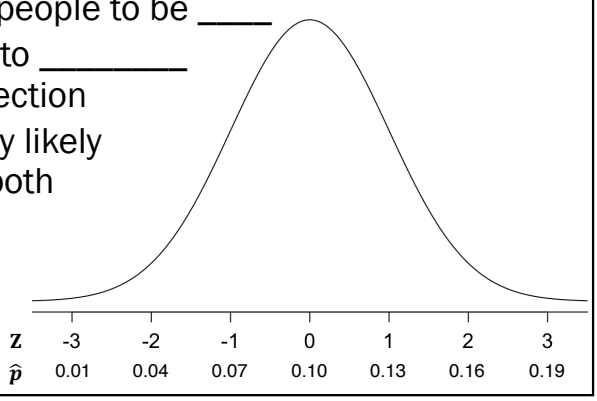
Example: Determining Sampling Distribution

- **Scenario:** 10% of the world's population is left-handed. Obtain the handedness of 100 randomly selected people.
- **Question:** What is the sampling distribution of the sample proportion?
- **Answer:**
  1. **Mean:** _____
  2. **Standard Error:** _____
  3. **Shape:** _____
     - **Exp. Successes:** _____ = ___ ≥ ___
     - **Exp. Failures:** _____ = ___ ≥ ___



Example: Understanding Sampling Distribution

- **Scenario:** 10% of the world's population is left-handed. Obtain the handedness of 100 randomly selected people.
- **Question:** What does the sampling distribution reveal?
- **Answer:**
  - Expect the proportion of left-handed people to be _____
  - Sample proportions of size 100 tend to _____ from _____ by about _____ in either direction
  - A sample proportion of _____ is equally likely as a sample proportion of _____, but both would be _____
    - ___ standard errors from the mean



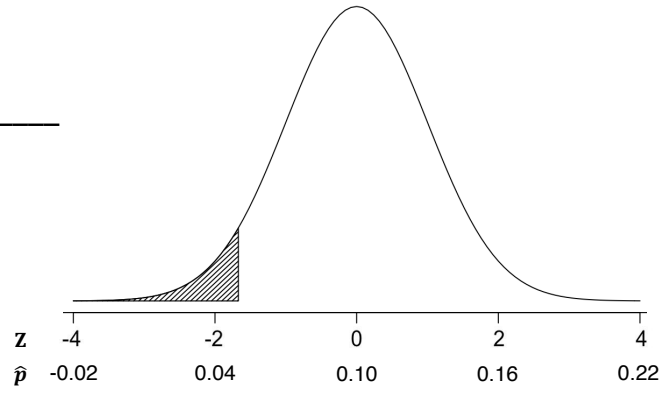
Example: Probability Calculation

- **Scenario:** 10% of the world's population is left-handed. Obtain the handedness of 100 randomly selected people.
- **Question:** What is the probability that less than 5% of a sample of 100 people is left-handed?
- **Answer:**

$P(\hat{p} < 0.05) =$ _____

$=$ _____

$=$ _____



Motivation: Inference for a Population Proportion

- **Scenario:** An April 2021 survey of 421 Pennsylvanian residents found that 244 of them were in favor of legalizing recreational marijuana.
- **Question:** Can we conclude that a majority of Pennsylvanians favor the legalization of recreational marijuana?
- **Observations:**
  - Working with _____ data – support is either _____
  - Performing inference on a _____

One-Sample Proportion Test

Step	Description
Used for	Performing inference on a single unknown population proportion where the variable has two possible outcomes
Conditions	Expected number of successes np and expected number of failures $n(1-p)$ must **both** be at least 10
Test Statistic	$Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ where p_0 stands for the hypothesized proportion
Confidence Interval	$\hat{p} \pm Z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

*Note: The hypothesis test uses the **hypothesized** proportion in the standard error, but the confidence interval uses the **sample** proportion.*

Example: One-Sample Proportion Test

- **Scenario:** An April 2021 survey of 421 Pennsylvanian residents found that 244 of them were in favor of legalizing recreational marijuana.
- **Question:** Can we conclude that a majority of Pennsylvanians favor the legalization of recreational marijuana?

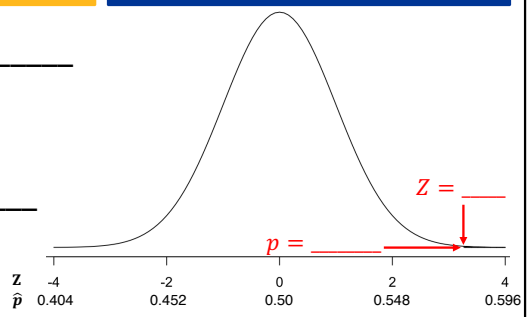
- **Hypotheses:** _____ vs. _____

- **Normality Condition:** _____
  - **Expected Successes:** _____ = _____ ≥ ___
  - **Expected Failures:** _____ = _____ ≥ ___

Example: One-Sample Proportion Test (Cont.)

- **Statistics:** $\hat{p}$ = _____ = _____, n = _____

- **Test Statistic:** Z = _____ = _____
  - **Interpretation:** If ____ of Pennsylvanians actually support the legalization of recreational marijuana, then _____ is _____ above the _____.



- **P-Value:** p = _____ = _____
  - **Interpretation:** If ____ of Pennsylvanians actually support the legalization of recreational marijuana, then the probability we would get a _____ proportion of _____ from a sample of _____ is _____.

One-Sample Proportion Test Using R

- **Function:**

Number of successes: `x = 244` | Sample size: `n = 421` | Hypothesized proportion: `p = 0.50`

```
prop.test(x = 244, n = 421, p = 0.50,
          alternative = "greater", correct = FALSE)
```

Form of alternative hypothesis: `alternative = "greater"`

Tells R to use normal distribution – must always be FALSE: `correct = FALSE`

```
> prop.test(x = 244, n = 421, p = 0.50, alternative = "greater", correct = FALSE)

        1-sample proportions test without continuity correction

data:  244 out of 421, null probability 0.5
X-squared = 10.663, df = 1, p-value = 0.0005466
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.5396158 1.0000000
sample estimates:
        p
0.5795724
```

Square of test statistic $3.27^2 \approx 10.663$

Ignore if alternative is one-sided and redo with `alternative = "two.sided"` to get correct C.I.

Sample proportion $\hat{p}$

Example: One-Sample Proportion Test (Cont.)

- 95% Confidence Interval: _____ $\pm$ _____ $\sqrt{\frac{\_\_\_\_\_}{\_\_\_\_}}$ ← Remember to use $\hat{p}$ in the C.I.

_____ $\pm$ _____

_____

- Conclusions:
  - _____and conclude that a _____ (significantly _____) of Pennsylvanians _____.
    - $p =$ _____ $<$ _____
  - We are 95% confident that the true percentage of Pennsylvanians who support legalizing recreational marijuana is _____ _____.

Sample Size Calculations

- To determine how large a sample is needed to attain a margin of error of m for a one-sample proportion confidence interval:

$$n = \left(\frac{Z\sqrt{\hat{p}(1-\hat{p})}}{m} \right)^2$$

  - Round up if n is a decimal

- **Problem:** Need __ to calculate __, but we need a _____ to find __
- **Solutions:**
  1. Use a value of $\hat{p}$ from a _____
  2. Use $\hat{p} =$ _____ as the most _____ estimate if no previous information exists

Example: Sample Size Calculations

- **Scenario:** We want to estimate the defect rate of a computer chip for the new machine. Failure rate for other machines is 2%.
- **Question:** How large a sample is needed to estimate the failure rate to within 1% using 95% confidence?

- **Answer:** $n =$ _____ $=$ _____ $\rightarrow$ Round up to _____
- **Question:** How would the sample size have changed if we had used the most conservative estimate of $\hat{p} = .50$?

- **Answer:** $n =$ _____ $=$ _____
- **Takeaway:** Sample sizes get smaller as $\hat{p}$ gets _____