HYPOTHESIS TESTING: APPLICATION OF CENTRAL LIMIT

CENTRAL LIMIT THEOREM

Central Limit Theorem: Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and standard deviation σ. Let $\bar{X}$ be the sample average of $X_1, X_2, \ldots, X_n$. Then the distribution of $\bar{X}$ is approximately normal with mean μ and standard deviation $\sigma/\sqrt{n}$.

- Mathematically, {Xi} is a sequence of IID with CDF F(x), mean E(Xi) = μ, and standard deviation std(Xi) = **σ**

$$\frac{1}{n}\sum_{i=1}^{n} x_i$$

- The average follows a normal distribution of mean μ and standard deviation **σ**/sqrt(n)

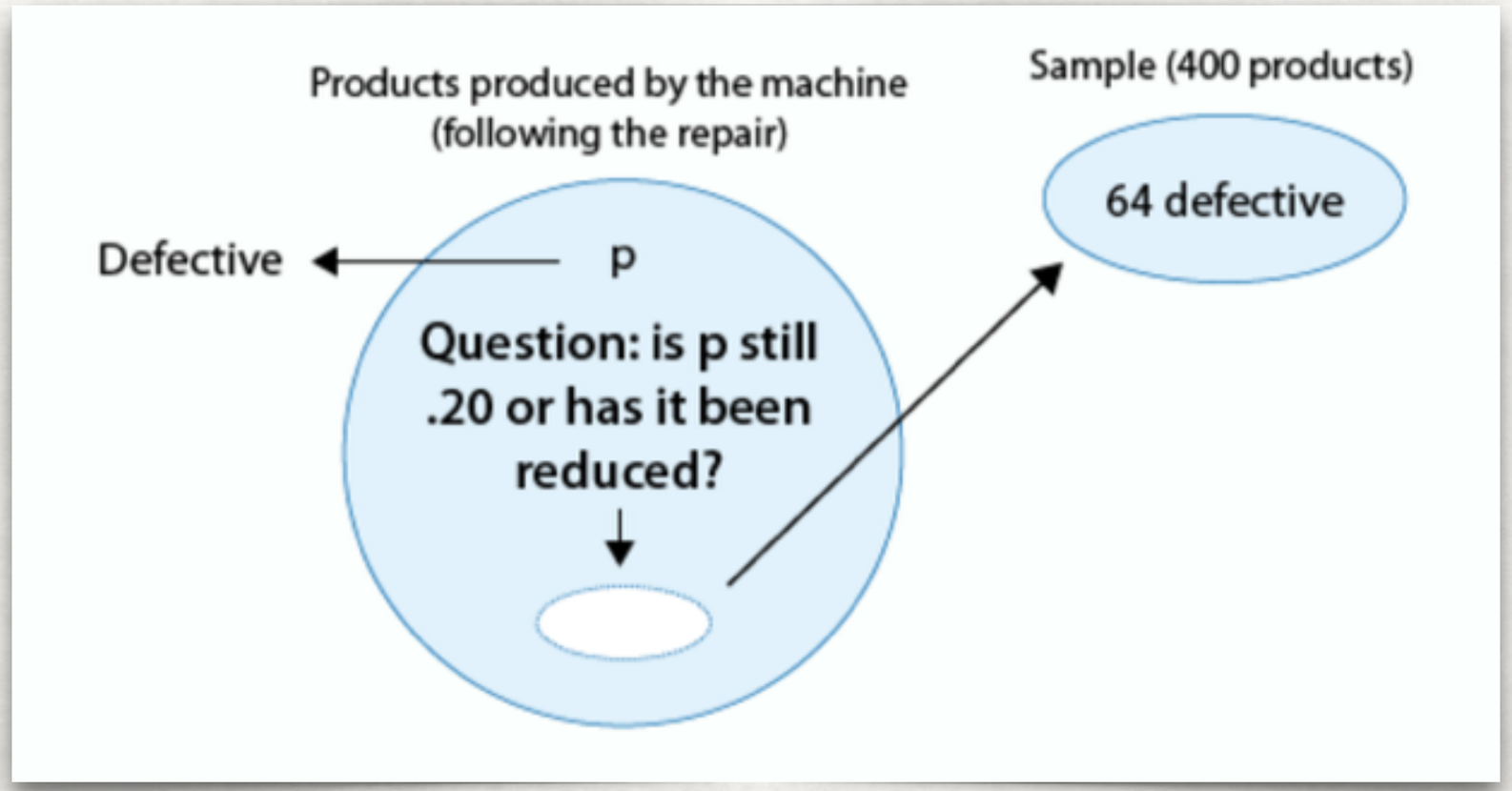
HYPOTHESIS TESTING

- Recall that there are basically 4 steps in the process of hypothesis testing:
- 1. State the null and alternative hypotheses.
- 2. Collect relevant data from a random sample and summarize them (using a test statistic).
- 3. Find the p-value, the probability of observing data like those observed assuming that Ho is true.
- 4. Based on the p-value and a predetermined α = significance level, decide whether we have enough evidence to reject Ho (and accept Ha), and draw our conclusions in context.
  If P-value is less than α, then we reject H0, accept Ha.

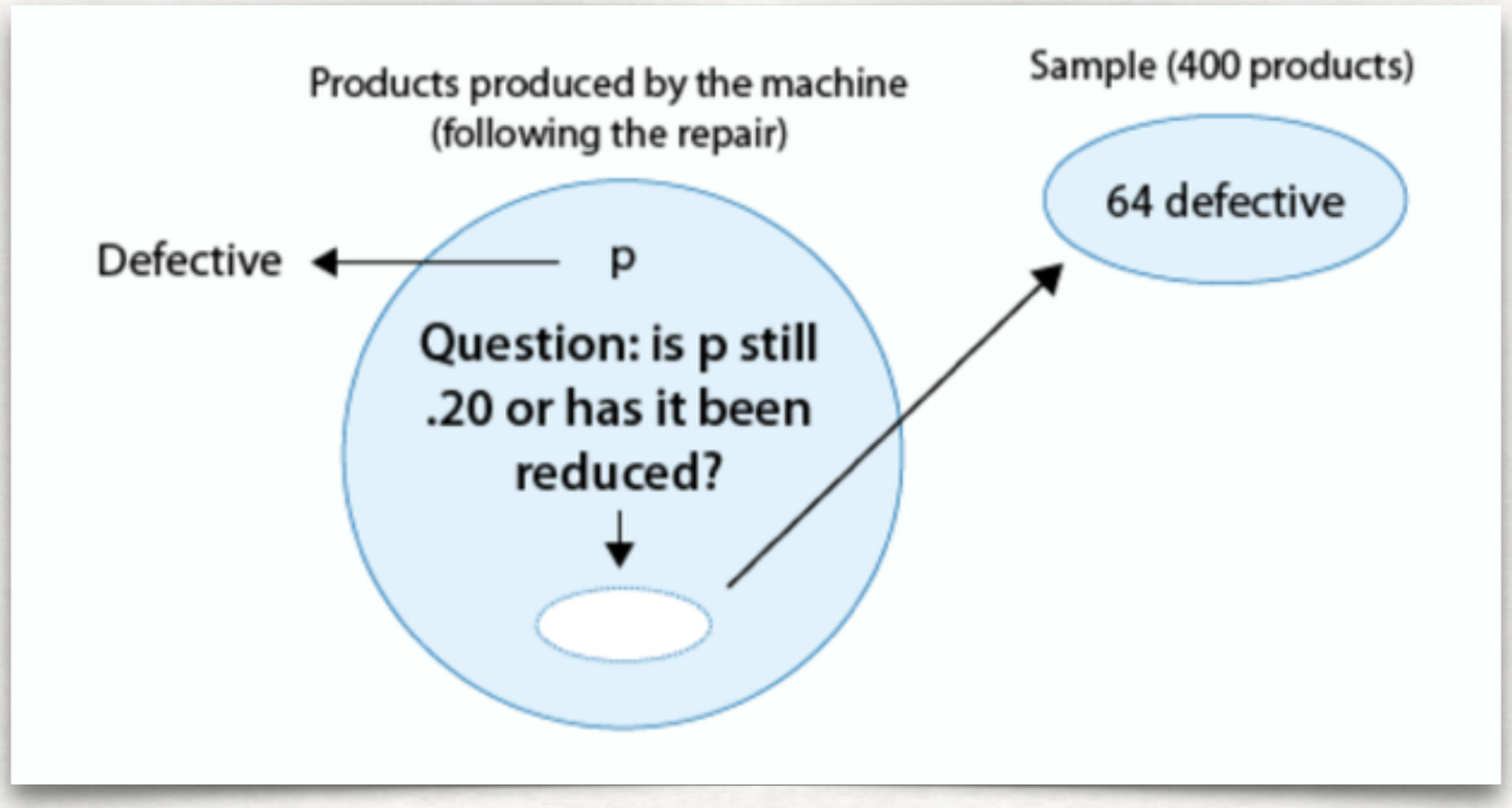
EXAMPLE

DETERMINE H0, HA

- A machine is known to produce 20% defective products, and is therefore sent for repair.
- After the machine is repaired, 400 products produced by the machine are chosen at random and 64 of them are found to be defective.
- Do the data provide enough evidence that the proportion of defective products produced by the machine (p) has been reduced as a result of the repair?

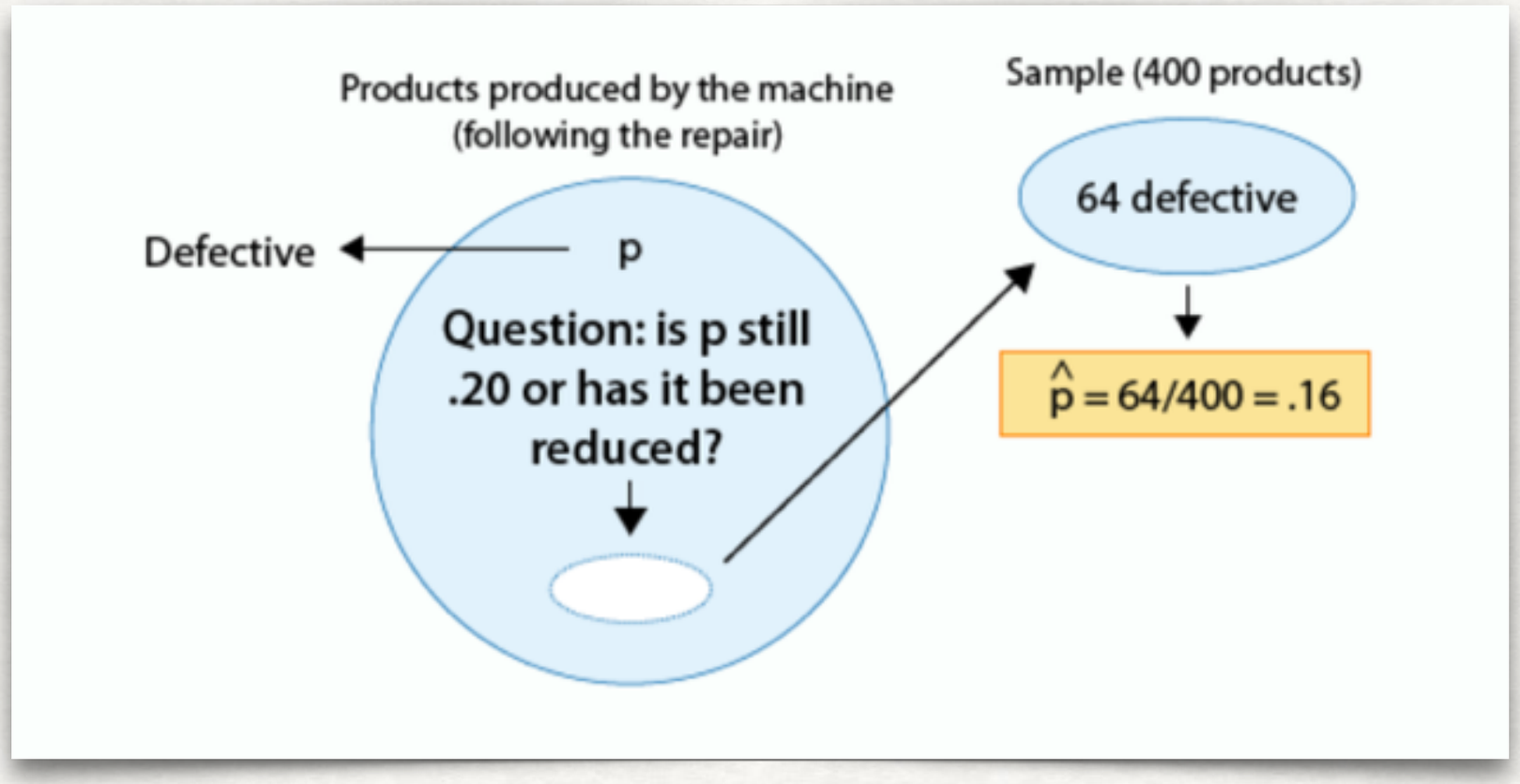


- 1. State the null and alternative hypotheses.



- Let p be the actual portion of defective product
- Ho: p = 0.20 (No change; the repair did not help).
- Ha: p < 0.20 (The repair was effective).

2. COLLECT RELEVANT DATA FROM A RANDOM SAMPLE AND SUMMARIZE THEM (USING A TEST STATISTIC)



- Data provides: 16% defective products
- The data are therefore 0.04 (or 4 percentage points) below the null hypothesis with respect to what they each tell us about p.

2. COLLECT RELEVANT DATA FROM A RANDOM SAMPLE AND SUMMARIZE THEM (USING A TEST STATISTIC)

- Under Null Hypothesis

* mean: p

  = 0.2

* standard deviation: $\sqrt{\frac{p(1-p)}{n}}$

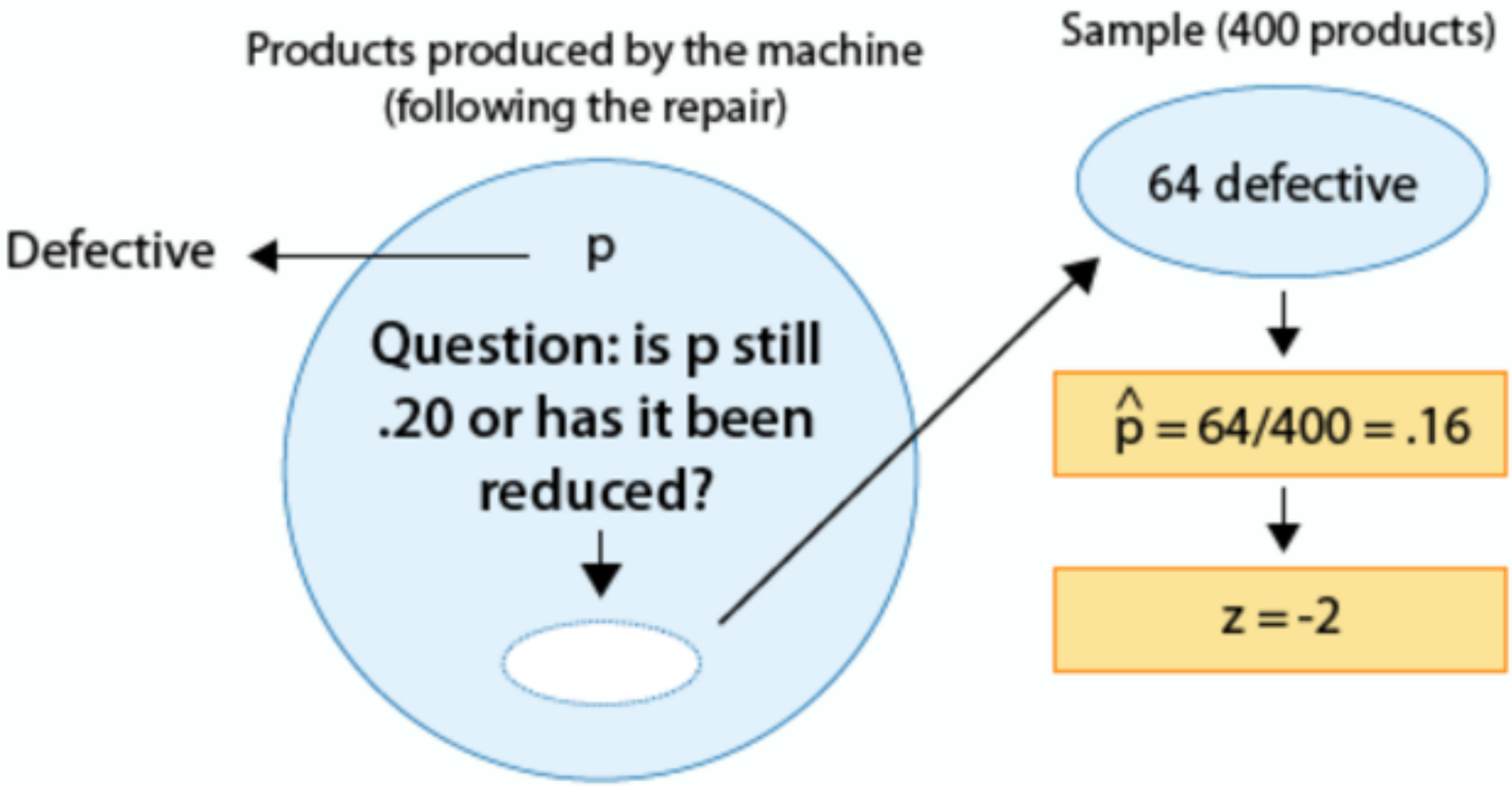
  = sqrt(0.2(1-0.2)/400) = 0.02

- Test statistics (or z-score):(0.16-0.2)/0.02 = -2

$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}}$$

- z represents # of standard deviations below or above the mean the value is.

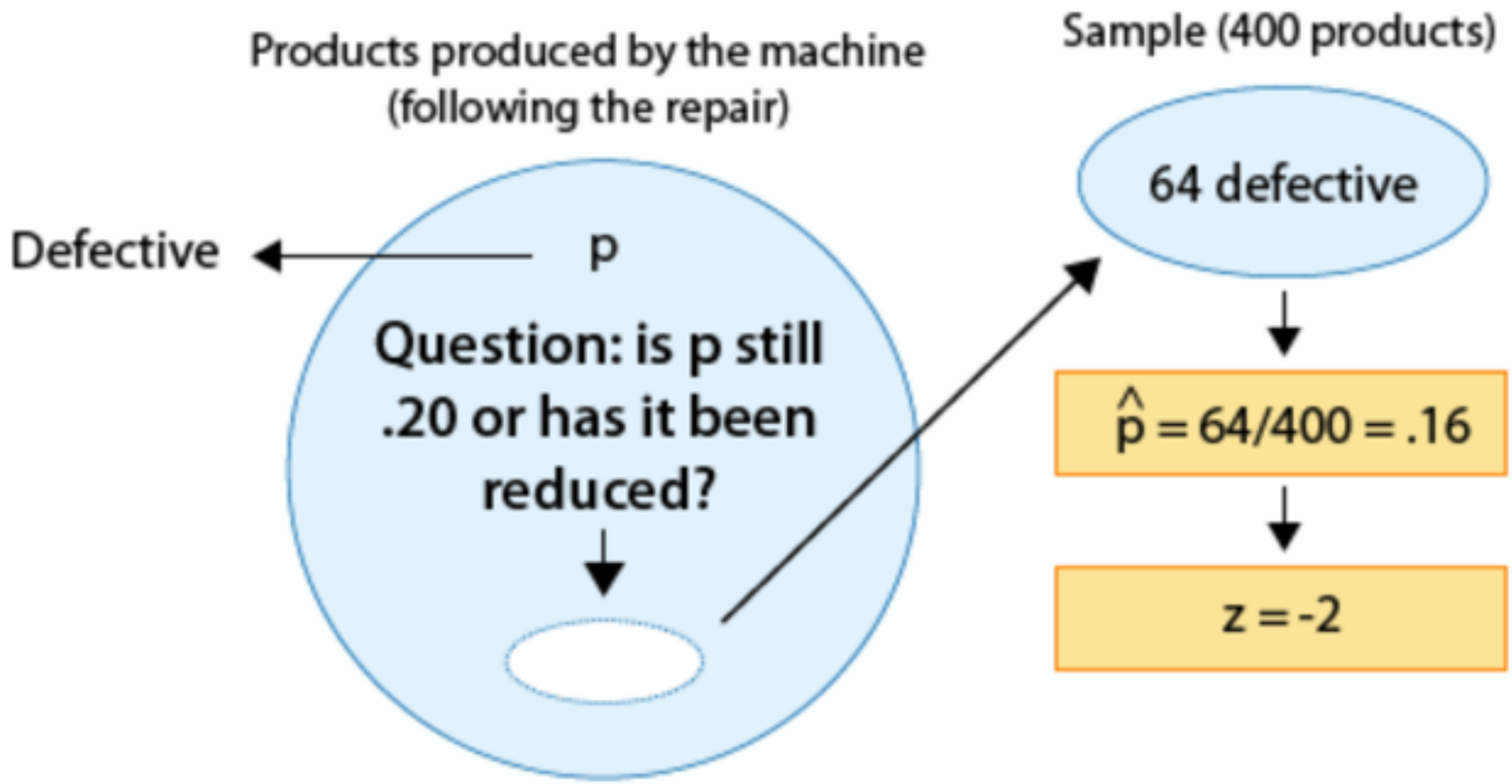
3. FIND THE P-VALUE, THE PROBABILITY OF OBSERVING DATA LIKE THOSE OBSERVED ASSUMING THAT HO IS TRUE.



Since the null hypothesis is H_o: p = 0.20, the standardized score of $\hat{p} = 0.16$ is: $z = \frac{0.16-0.20}{\sqrt{\frac{0.20(1-0.20)}{400}}} = -2.$

This z-score of -2 tells me that (assuming that Ho is true) the sample proportion is 2 standard deviations below the null value (0.20).

3. FIND THE P-VALUE, THE PROBABILITY OF OBSERVING DATA LIKE THOSE OBSERVED ASSUMING THAT HO IS TRUE.



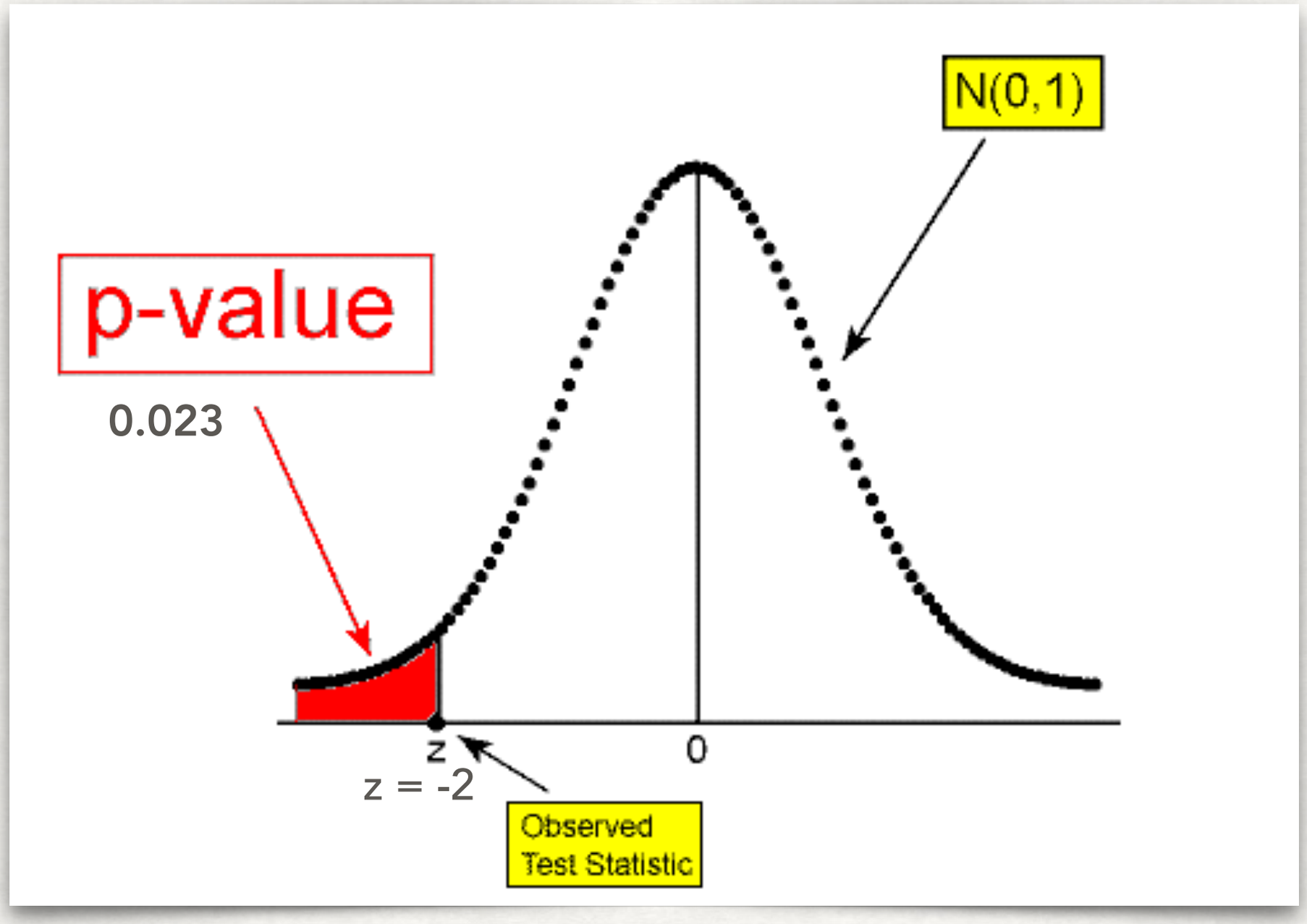
Since the null hypothesis is H_o: p = 0.20, the standardized score of $\hat{p} = 0.16$ is: $z = \frac{0.16-0.20}{\sqrt{\frac{0.20(1-0.20)}{400}}} = -2.$

Z~N(0,1)

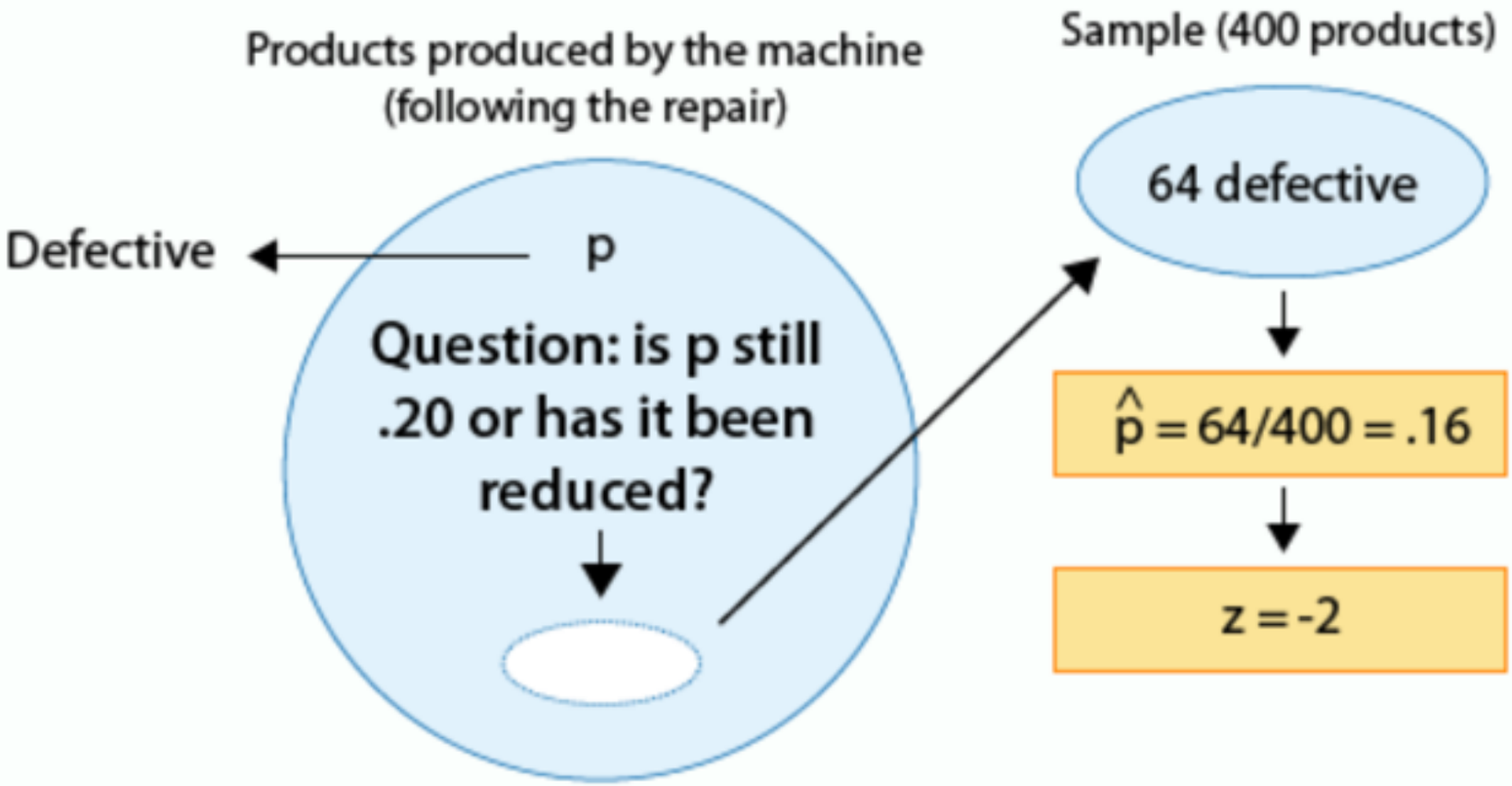
P-Value = P(Z<-2)

* Idea: assuming Ho is true, we observe a test statistic as small as -2 or smaller (i.e. we observe 16% defective or less defective, the "less" part is given by Ha)

3. FIND THE P-VALUE, THE PROBABILITY OF OBSERVING DATA LIKE THOSE OBSERVED ASSUMING THAT HO IS TRUE.



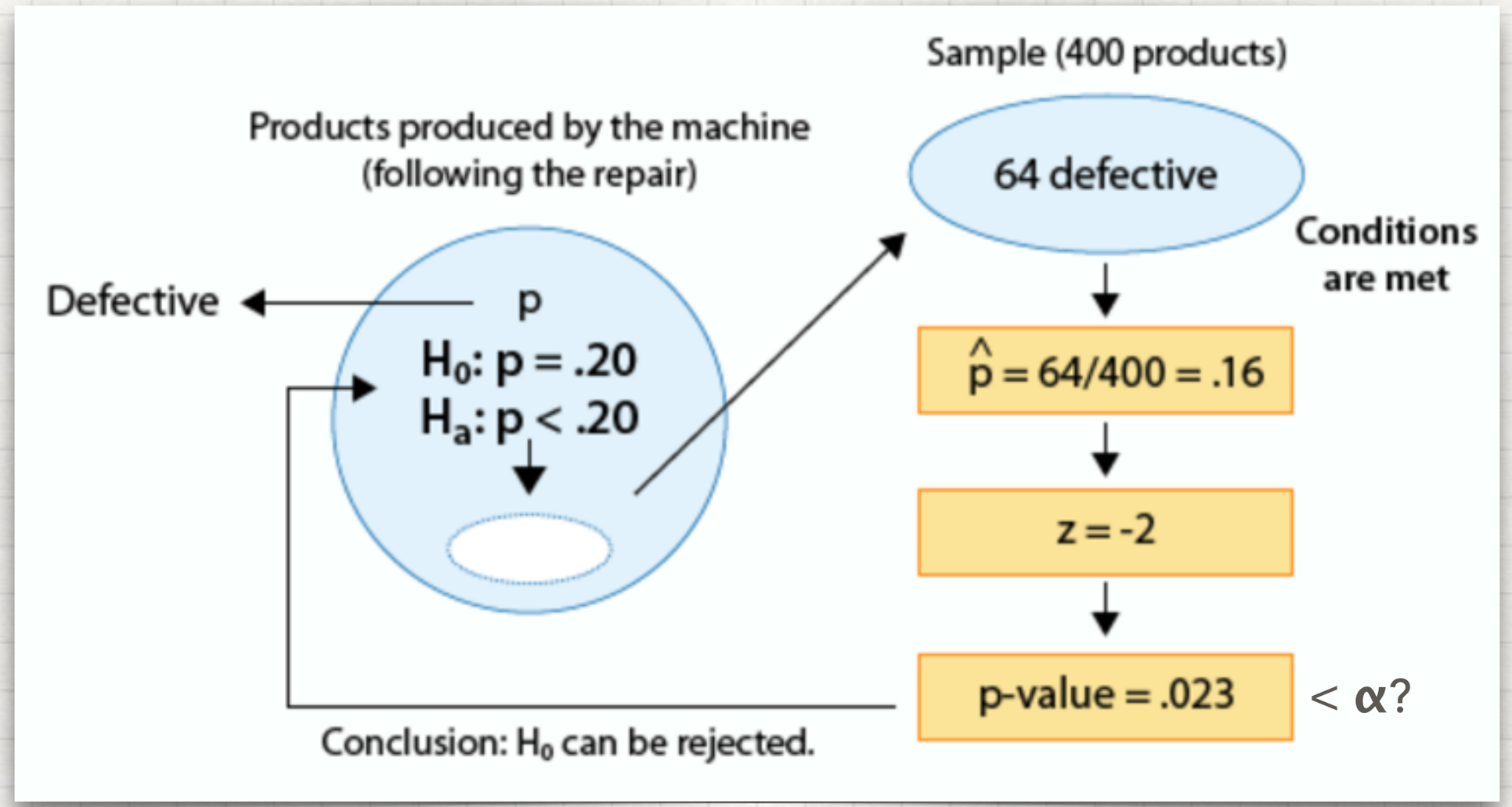
3. FIND THE P-VALUE, THE PROBABILITY OF OBSERVING DATA LIKE THOSE OBSERVED ASSUMING THAT HO IS TRUE.



Since the null hypothesis is H_o: p = 0.20, the standardized score of $\hat{p} = 0.16$ is: $z = \frac{0.16-0.20}{\sqrt{\frac{0.20(1-0.20)}{400}}} = -2.$

The value can be found using z-table P-Value = 0.023

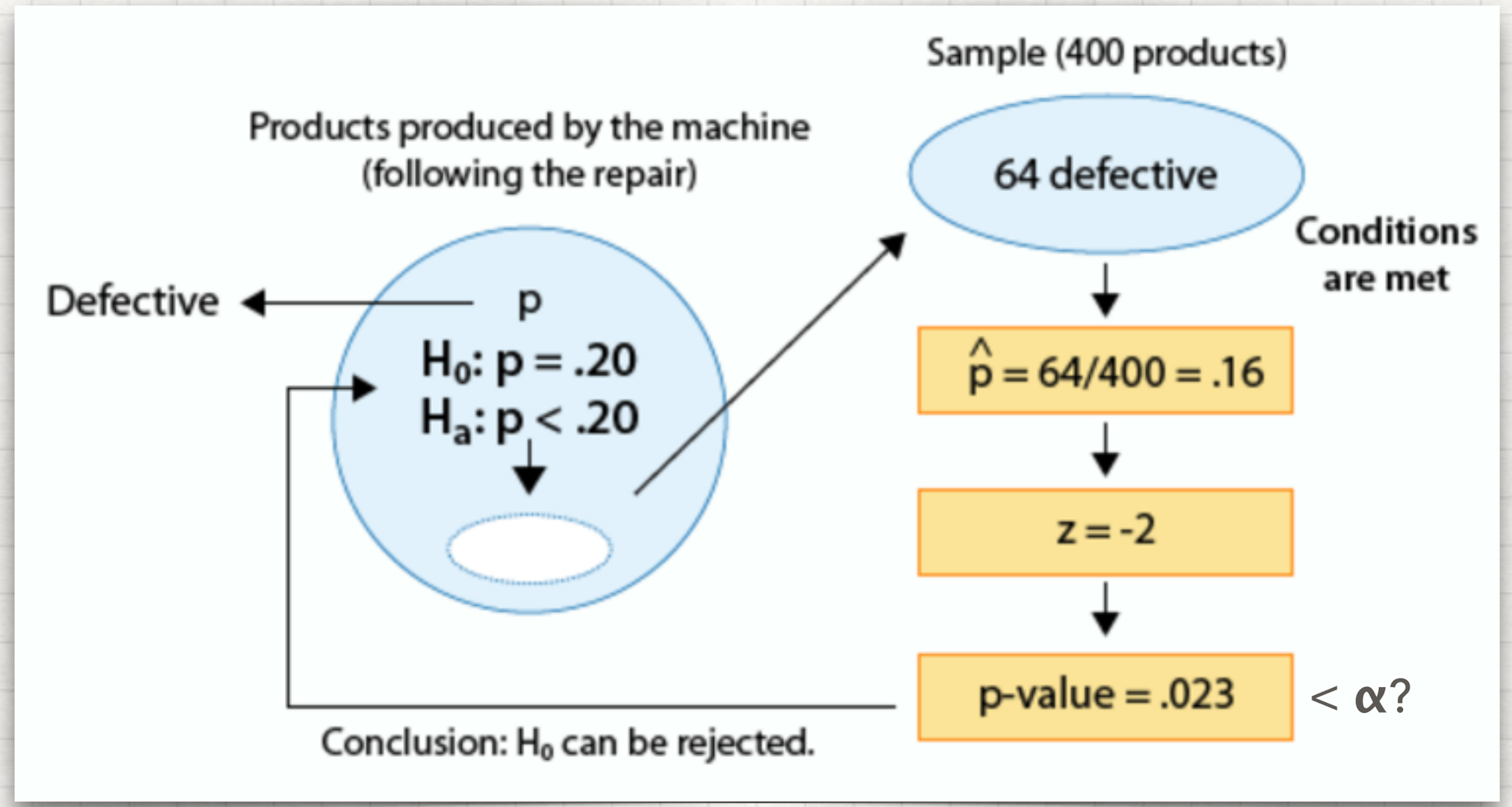
4. BASED ON THE P-VALUE, DECIDE WHETHER WE HAVE ENOUGH EVIDENCE TO REJECT HO (AND ACCEPT HA), AND DRAW OUR CONCLUSIONS IN CONTEXT.



Say, we want a significance level **α** = significance level 0.05 (or confidence level = 0.95)

Since 0.023 < 0.05), the data provide enough evidence to reject Ho and conclude that: as a result of the repair the proportion of defective products **has been reduced to below 0.20**.

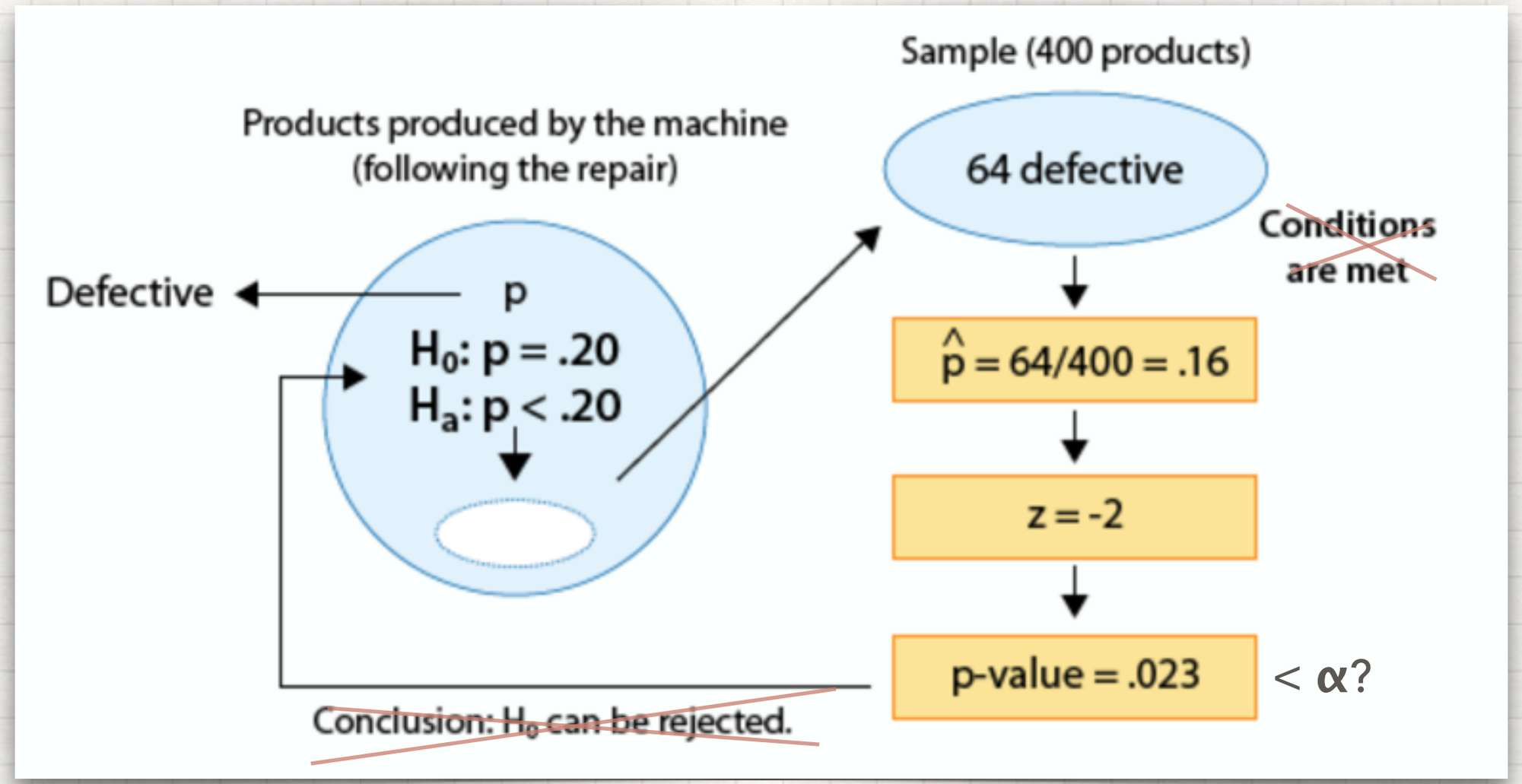
4. BASED ON THE P-VALUE, DECIDE WHETHER WE HAVE ENOUGH EVIDENCE TO REJECT HO (AND ACCEPT HA), AND DRAW OUR CONCLUSIONS IN CONTEXT.



Significance level α = significance level 0.05 and p-value 0.023 < 0.05

This means that we are more than 95% sure that the machine is fixed.

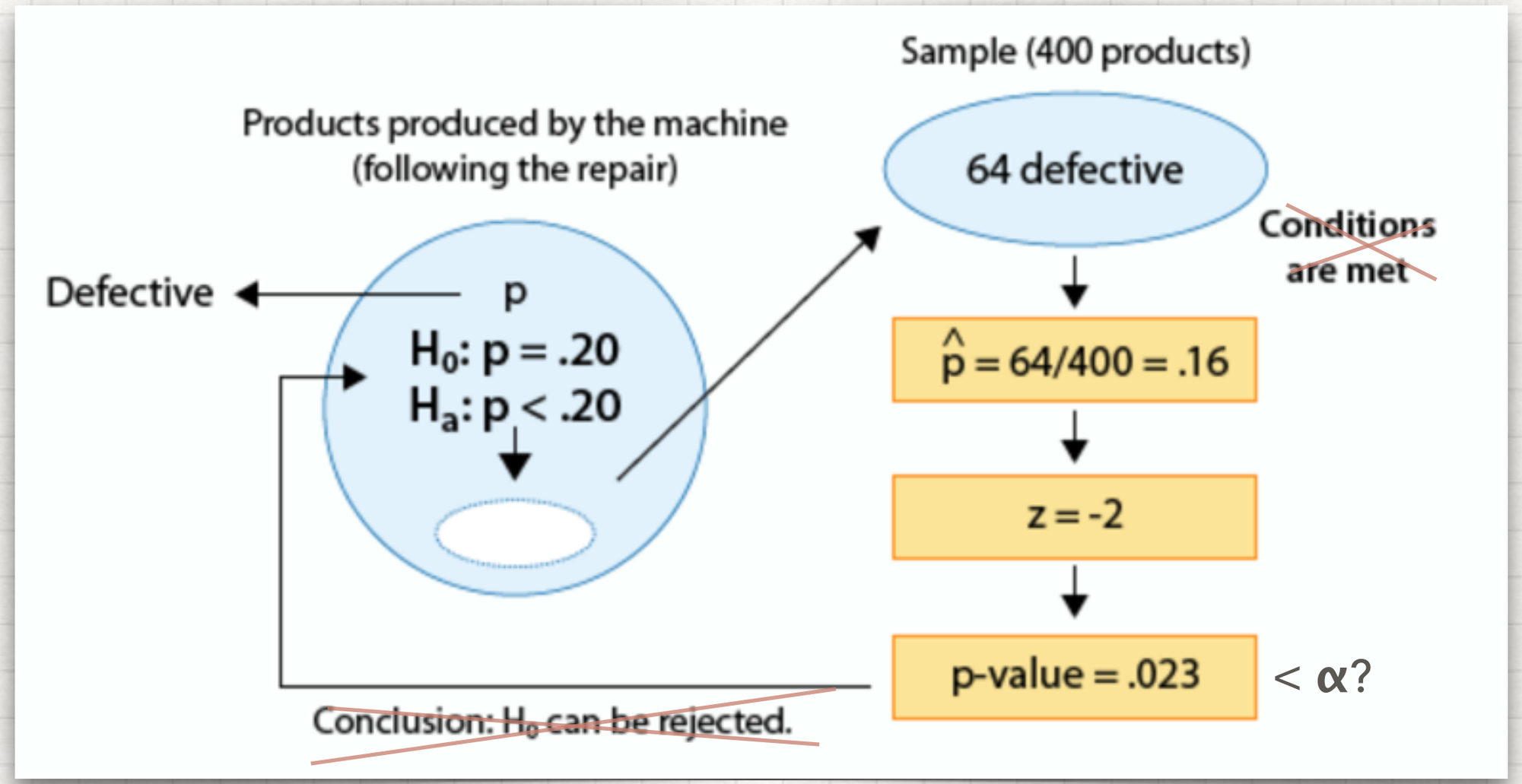
4. BASED ON THE P-VALUE, DECIDE WHETHER WE HAVE ENOUGH EVIDENCE TO REJECT HO (AND ACCEPT HA), AND DRAW OUR CONCLUSIONS IN CONTEXT.



But, if we are very strict and set our significance level **α** = significance level 0.01

Since 0.023 > 0.01, the data does not provide enough evidence to reject Ho. And we cannot draw any conclusion on whether or not the machine is fixed.

4. BASED ON THE P-VALUE, DECIDE WHETHER WE HAVE ENOUGH EVIDENCE TO REJECT HO (AND ACCEPT HA), AND DRAW OUR CONCLUSIONS IN CONTEXT.



Note, when p-value > α = significance level. The data does not provide enough evidence to reject Ho. And we cannot draw any conclusion on whether or not the machine is fixed.

This is like 'reasonable doubt' (you have to convict more than a reasonable doubt): a court does not have enough evidence to convict a suspect, but it does not mean suspect is not guilty.

MORE EXAMPLES

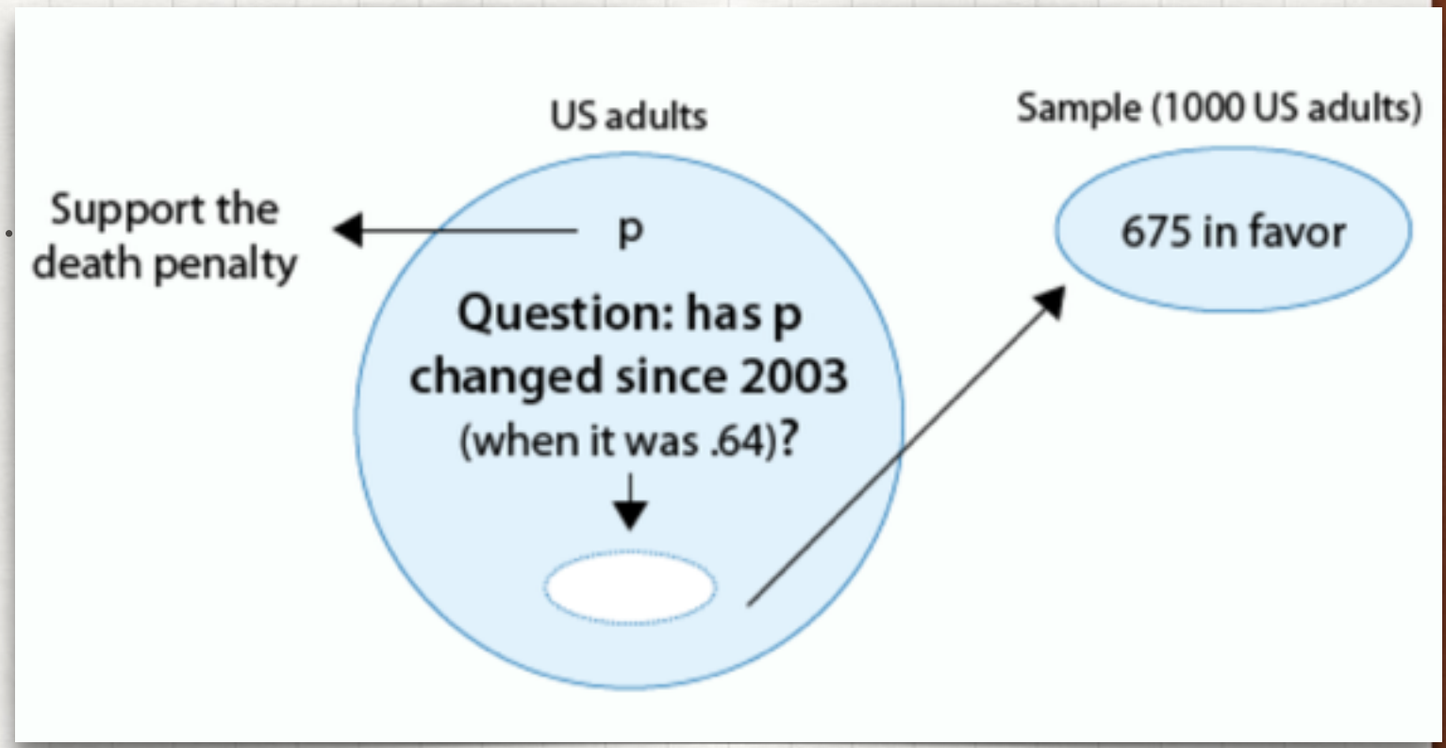
- Polls on certain topics are conducted routinely in order to monitor changes in the public's opinions over time. One such topic is the death penalty. In 2003 a poll estimated that 64% of U.S. adults support the death penalty for a person convicted of murder. In a more recent poll, 675 out of 1,000 U.S. adults chosen at random were in favor of the death penalty for convicted murderers. Do the results of this poll provide evidence that the proportion of U.S. adults who support the death penalty for convicted murderers (p) changed between 2003 and the later poll?

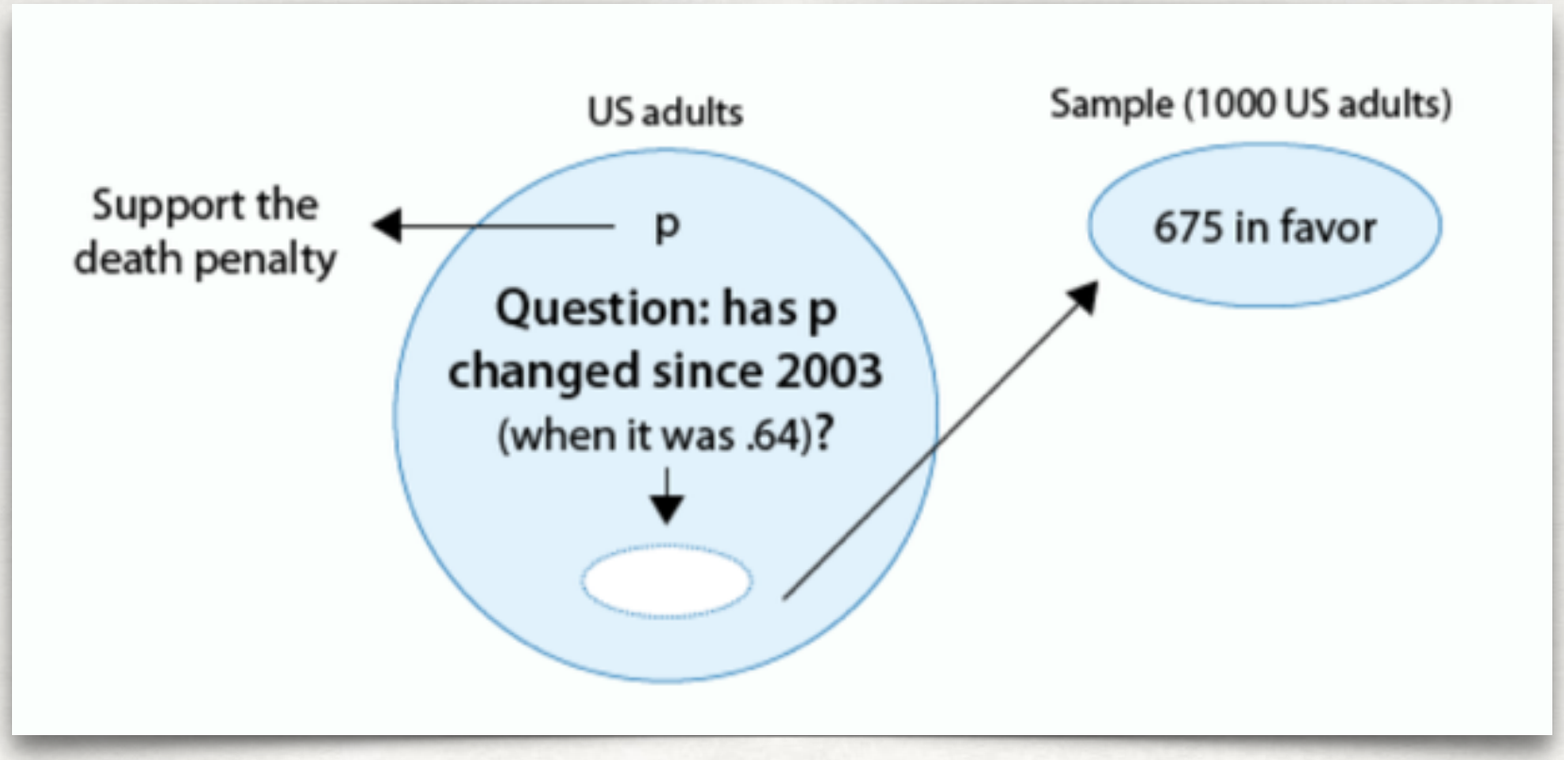
- Assume significancece level is 2.5%

SUMMARIZE DATA

HYPOTHESIS TESTING

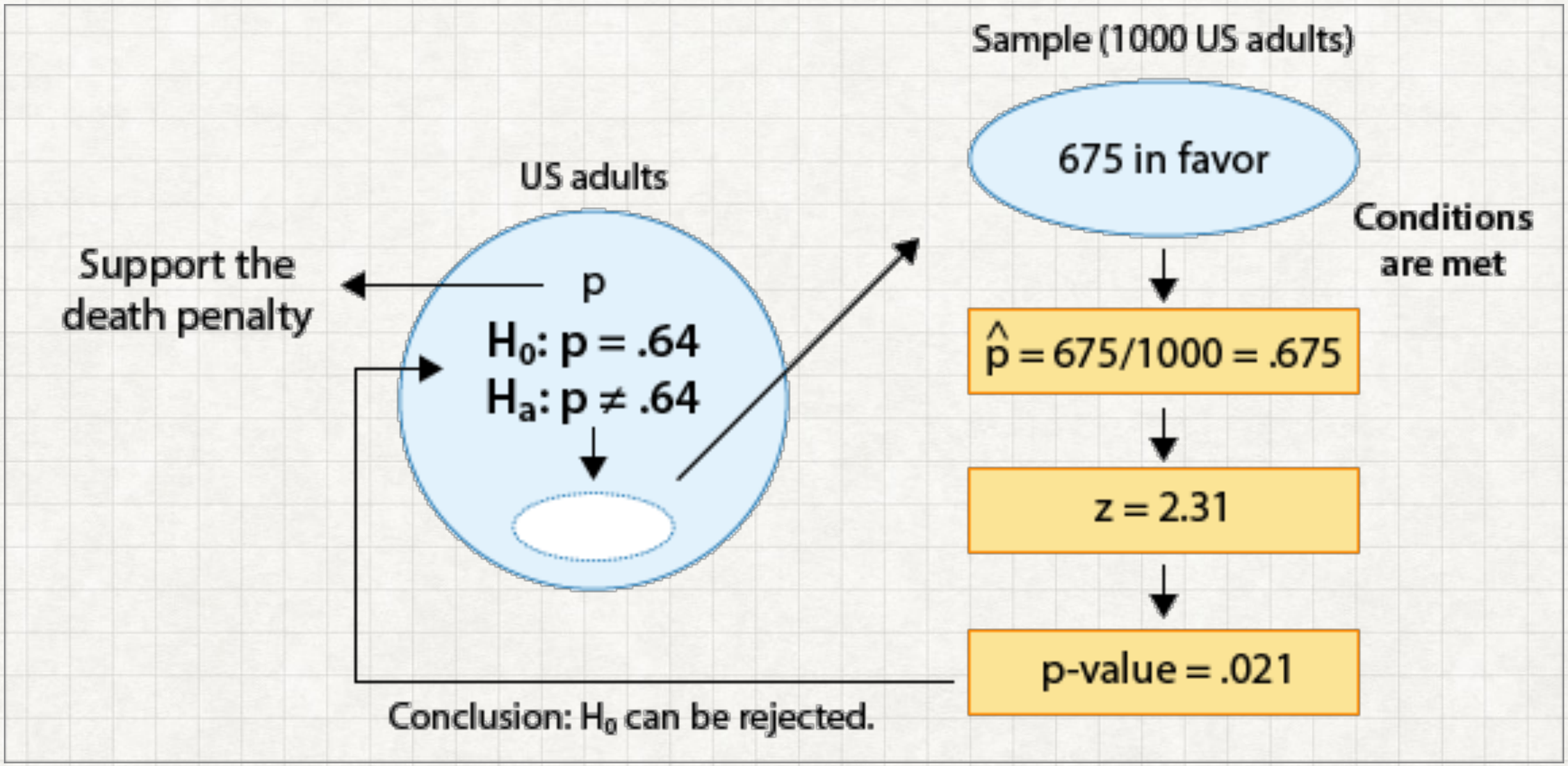
- Polls on certain topics are conducted routinely in order to monitor changes in the public's opinions over time.
- One such topic is the death penalty.
- In 2003 a poll estimated that 64% of U.S. adults support the death penalty for a person convicted of murder.
- In a more recent poll, 675 out of 1,000 U.S. adults chosen at random were in favor of the death penalty for convicted murderers.
- Do the results of this poll provide evidence that the proportion of U.S. adults who support the death penalty for convicted murderers (p) changed between 2003 and the later poll?





- Steps:
- 1. State the null and alternative hypotheses — Ho, Ha
- 2. Collect relevant data from a random sample and summarize them (using a test statistic):
  compute sample mean and standard deviation to obtain z-score
- 3. Use z-score to find the p-value
- 4. Based on the p-value and a predetermined α = significance level, decide whether we have enough evidence to reject Ho (and accept Ha), and draw our conclusions in context.
- For this question, set $\alpha = 0.025$

Since 0.021 < 0.025, the data provide enough evidence to reject Ho, and we conclude that the proportion of adults who support the death penalty for convicted murderers has changed since 2003. Here is the complete story of this example:



HYPOTHESIS TESTING FOR THE POPULATION PROPORTION P: SUMMARY

- **Step 1: State the null and alternative hypotheses:**
- Null Hypotheses is easy:
- Always assume nothing changes:

$$H_0 : p = p_0$$

- eg: machine is not repaired;
- eg: proportion of adults who support the death penalty for convicted murderers remains the same

where the choice of the appropriate alternative (out of the three) is usually quite clear from the context of the problem.

$$H_a : p \begin{Bmatrix} < \\ > \\ \neq \end{Bmatrix} p_0$$

eg: the defective rate is reduced (p < p0)
eg: the survival rate has increased (p>p0)
eg: proportion of adults who support the death penalty for convicted murderers has changed (P!=P0)

HYPOTHESIS TESTING FOR THE POPULATION PROPORTION P: SUMMARY

- **Step 2: Obtain data from a sample and calculate values**

  In reality, statisticians need to sample (do survey) from the population you are interested in and calculate $\hat{p}$
  eg: see if a die has heavy 1, you need to roll it n times. Then to calculate portion of ones.

- Random sample (or at least a sample that can be considered random in context) needs to satisfy:
  $n \cdot p0 \geq 10$, and $n \cdot (1 - p0) \geq 10$

- For the die example: p0 = 1/6
  $n \cdot 1/6 \geq 10$, $n \cdot (1 - 1/6) \geq 10$
  So you need at least 60 rolls!

Calculate the sample proportion , and summarize the data using the test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

Recall: This standardized test statistic represents how many standard deviations above or below p_0 our sample proportion $\hat{p}$ is.

HYPOTHESIS TESTING FOR THE POPULATION PROPORTION P: SUMMARY

- **Step 3: Find the p-value of the test either by using software or by using the test statistic as follows:**

* for $H_a : p < p_0 : P(Z \leq z)$

eg: the repair of the machine is effective, Ha is p<p0 , and we calculated P(Z<z)

* for $H_a : p > p_0 : P(Z \geq z)$

* for $H_a : p \neq p_0 : 2P(Z \geq |z|)$

Note: P(|Z| >= |z|) = 2P(Z>=|z|)

eg: the poll is different from before, Ha is p!=p0, and we calculated P(Z>|z|)

HYPOTHESIS TESTING FOR THE POPULATION PROPORTION P: SUMMARY

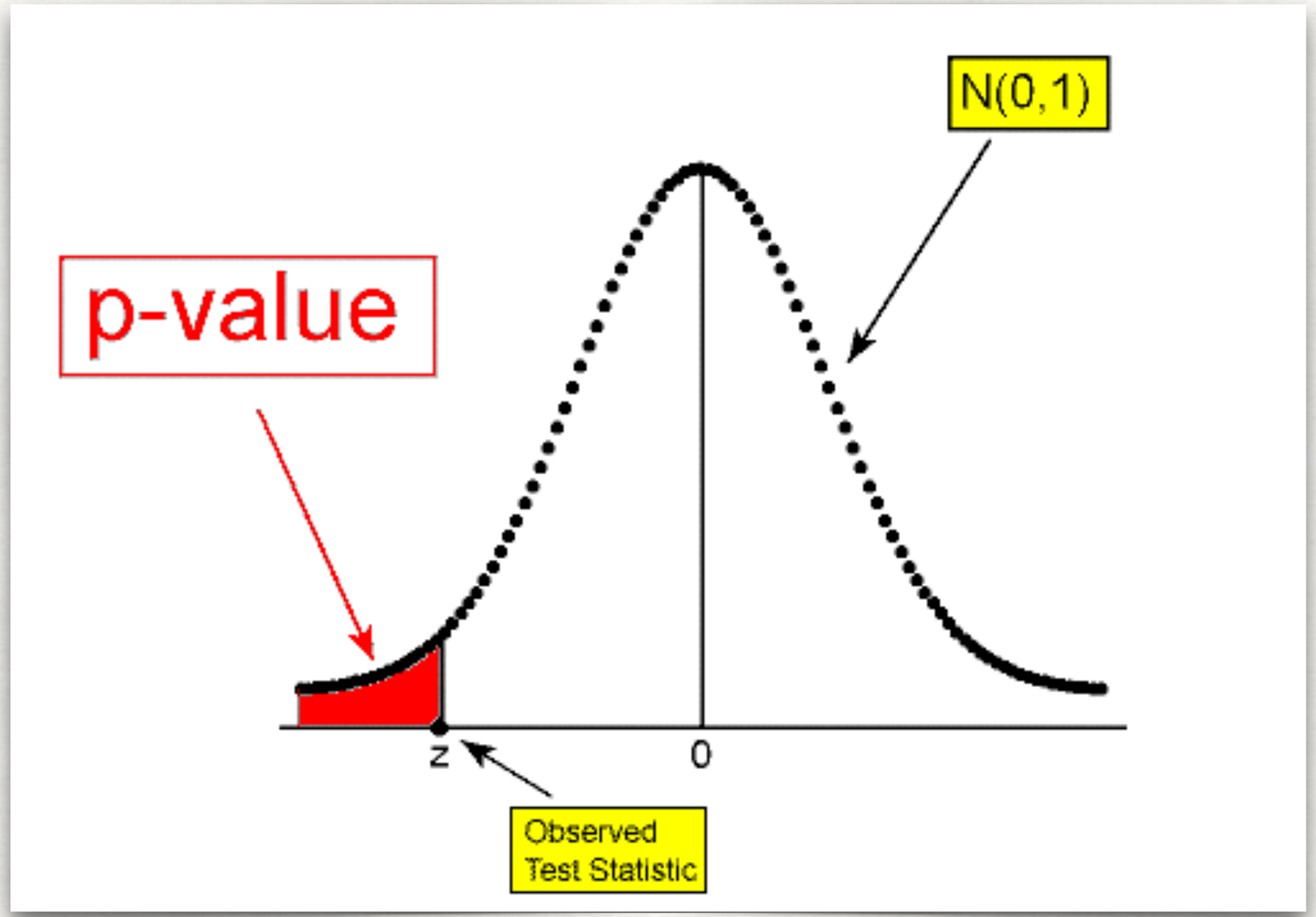
- Specifically, for the z-test for the population proportion:

- If the alternative hypothesis Ha is: P<P0(less than), then "extreme" means small, and the p-value is:
  The probability of observing a test statistic as small as that observed or smaller if the null hypothesis is true.

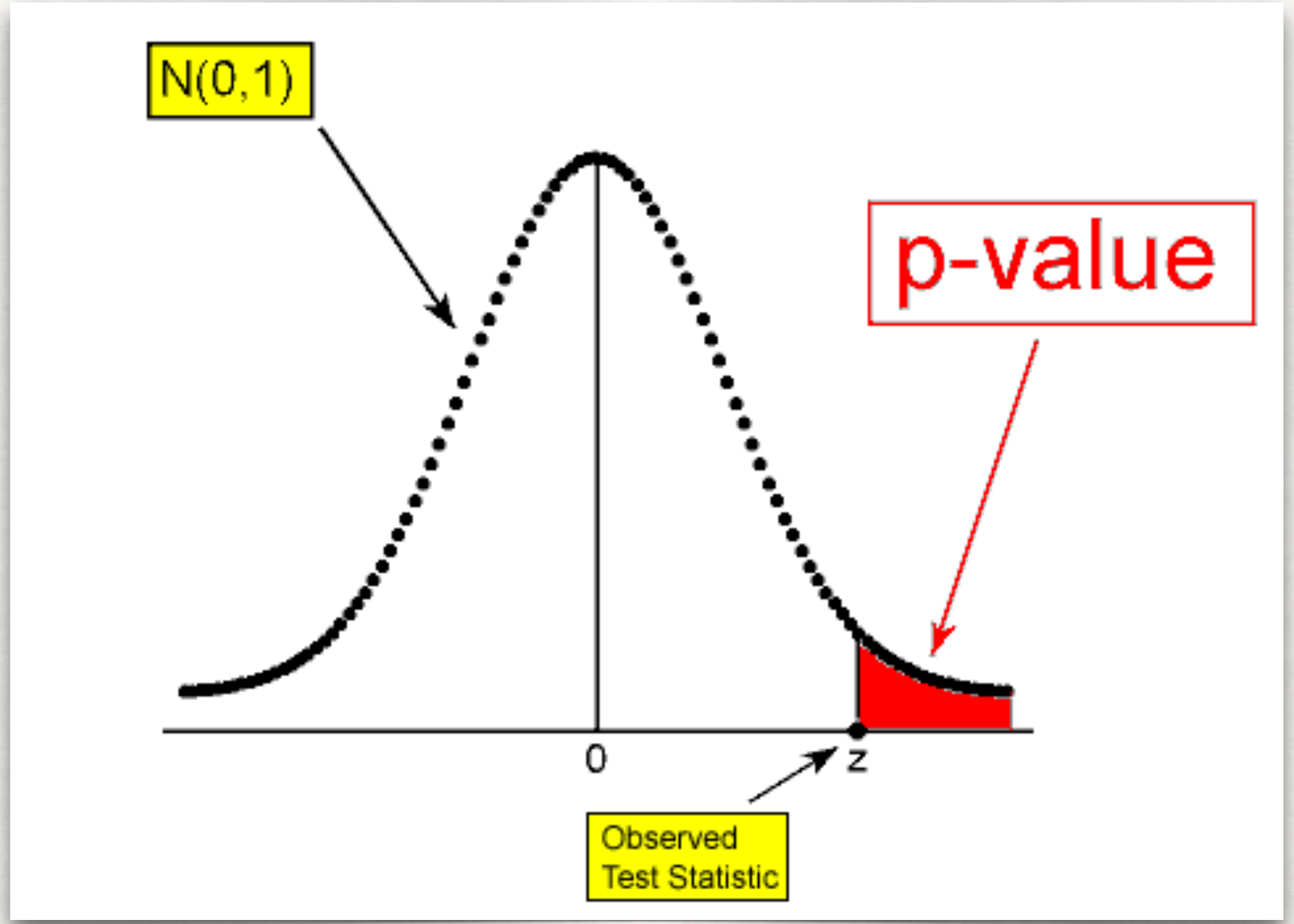
- If the alternative hypothesis is P>P0(greater than), then "extreme" means large, and the p-value is:
  The probability of observing a test statistic as large as that observed or larger if the null hypothesis is true.

- if the alternative is P!=P0(different from), then "extreme" means extreme in either direction either small or large (i.e., large in magnitude), and the p-value therefore is:
  The probability of observing a test statistic as large in magnitude as that observed or larger if the null hypothesis is true.

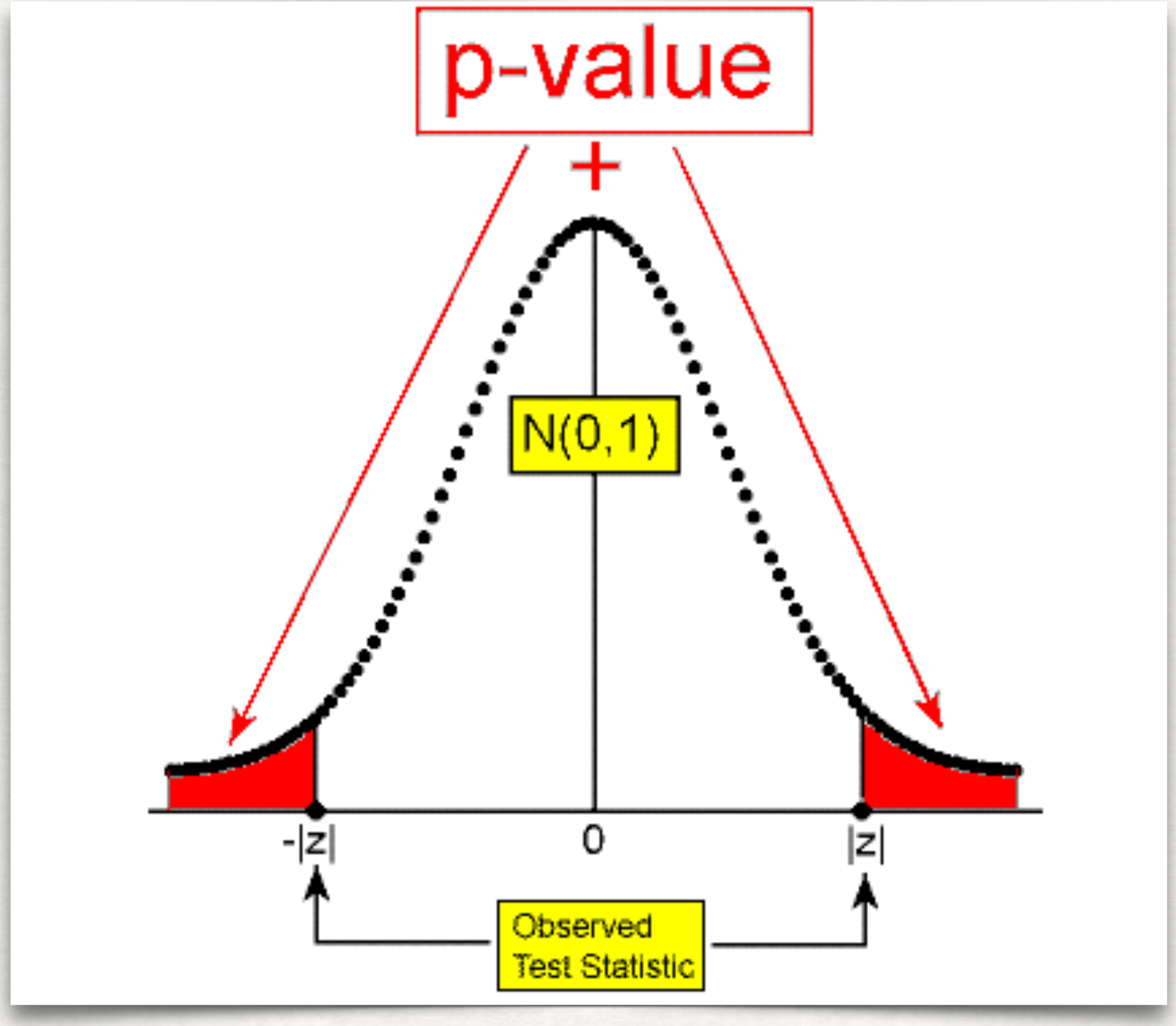
* for $H_a : p < p_0 : P(Z \leq z)$



* for $H_a : p > p_0 : P(Z \geq z)$



* for $H_a : p \neq p_0 : 2P(Z \geq |z|)$



HYPOTHESIS TESTING FOR THE POPULATION PROPORTION P: SUMMARY

- Step 4: Reach a conclusion first regarding the significance of the results, and then determine what it means in the context of the problem.

- Note that: significance level is usually set as $\alpha = 0.05$ (or confidence as 95%)

- If the p-value < α, the results are significant (in the sense that there is a significant difference between what was observed in the sample and what was claimed in Ho), and so we reject Ho.

- If the p-value is not small, we do not have enough statistical evidence to reject Ho, and so we continue to believe that Ho may be true. (Remember, in hypothesis testing we never "accept" Ho).

COIN FLIPPING EXAMPLE AGAIN

- We flip a coin 100 times, 60H 40T
  Question: Does this coin have heavy head?

- H0 =

- Ha =

- P-value =

- conclusion =

COIN FLIPPING EXAMPLE AGAIN

- We flip a coin 100 times, 60H 40T
  Question: Does this coin have heavy head?

- H0 = null hypothesis: observed fact is the same as general fact

- our case: H0 = the coin is a normal coin Po = 1/2

- Ha = alternative hypothesis: observed fact is different from the general population (>, <, or !=)

- our case: Ha = the coin has heavy head = P > Po (1/2)

COIN FLIPPING EXAMPLE AGAIN

- We flip a coin 100 times, 60H 40T
  Question: Does this coin have heavy head?

- Test statistics:

- mean = 60/100 = 0.6
  std = sqrt(0.5*0.5/100) = 0.05

- z-score = (0.6 - 0.5)/0.05 = 2

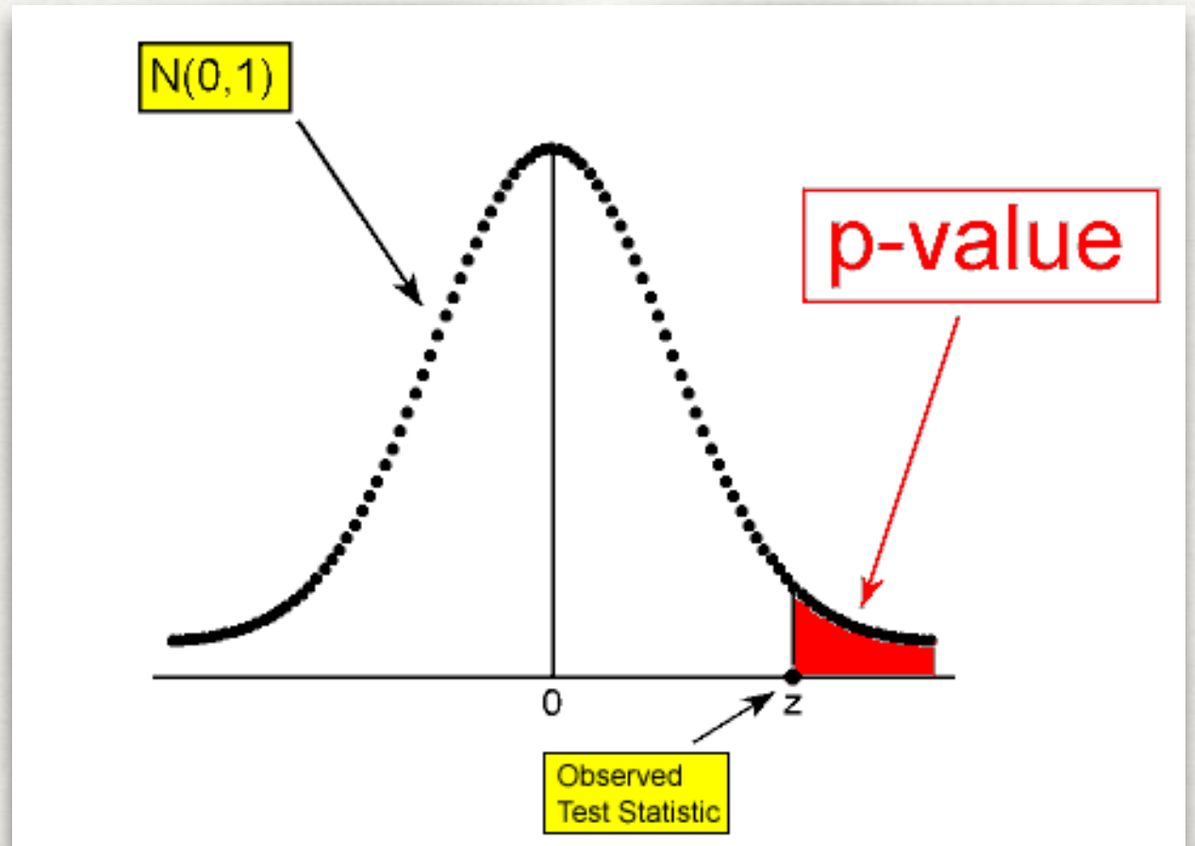
COIN FLIPPING EXAMPLE AGAIN

- We flip a coin 100 times, 60H 40T
  Question: Does this coin have heavy head?

- Since Ha = the coin has heavy head = P > Po,

- P-value = P(Z>z-score) = P(Z>2) = 0.0231

* for $H_a : p > p_0 : P(Z \geq z)$



COIN FLIPPING EXAMPLE AGAIN

- We flip a coin 100 times, 60H 40T
  Question: Does this coin have heavy head?

- p-value = 0.0231

- Let's use the general assumption that significance level **α** = 0.05

- Since P-value 0.0231 is less than **α** = 0.05, then we reject H0, accept Ha.

- We conclude that the coin is NOT fair, and it has a heavy head

COIN FLIPPING EXAMPLE AGAIN

- You observed some facts (eg. flip a coin 100 times, 60H 40T) — fair?
- H0 = null hypothesis: observed fact is the same as general fact
  our case: H0 = the coin is a normal coin P(H) = 1/2
- Ha = alternative hypothesis: observed fact is different from the general population
  our case: Ha = the coin has heavy head = Ph > 1/2
- P-value = probability that we observed the fact assuming H0
  our case: Assume the coin is fair, probability that we see 60H 40T
- **α** = significance level =
  If P-value is less than **α**, then we reject H0, accept Ha.
  our case: When assuming coin is fair, if the probability P of seeing 60H 40T is extremely low (say **α** = 5% and P< 5%). Then our assumption H0 is wrong and we accept Ha.

HYPOTHESIS TESTING FOR THE POPULATION PROPORTION P: CONFIDENCE INTERVALS

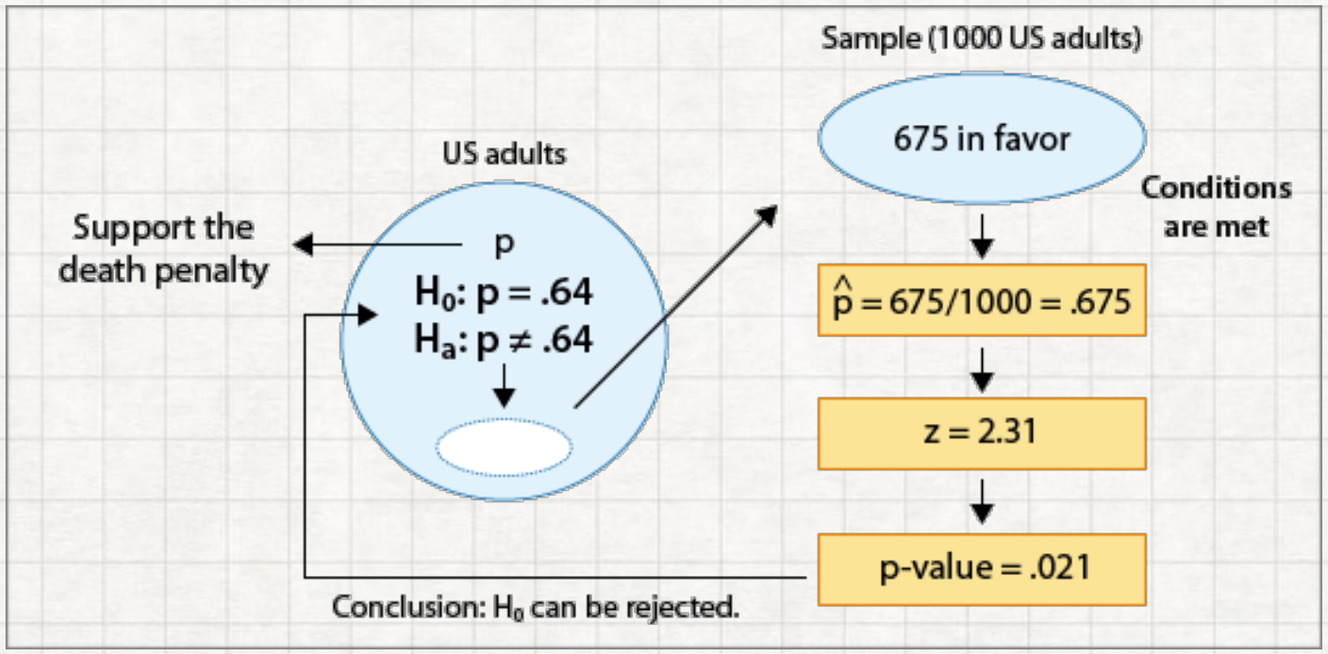
- Under significance level =0.05
  (Comment: Similarly, the results of a test using a significance level of 0.01 can be related to the 99% confidence interval.)

- Under significance level =0.05
  So, the confidence interval is [p-value - 2std, p-value+2std]

- If Po falls outside the confidence interval, reject Ho.

- If Po falls inside the confidence interval, do not reject Ho.

Recall the example
where we wanted to know whether the proportion of U.S. adults who support the death penalty for convicted murderers has changed since 2003, when it was 0.64.



We are testing:

$$H_0 : p = .64$$
$$Ha : p \neq .64$$

$$\hat{p} = 0.675$$

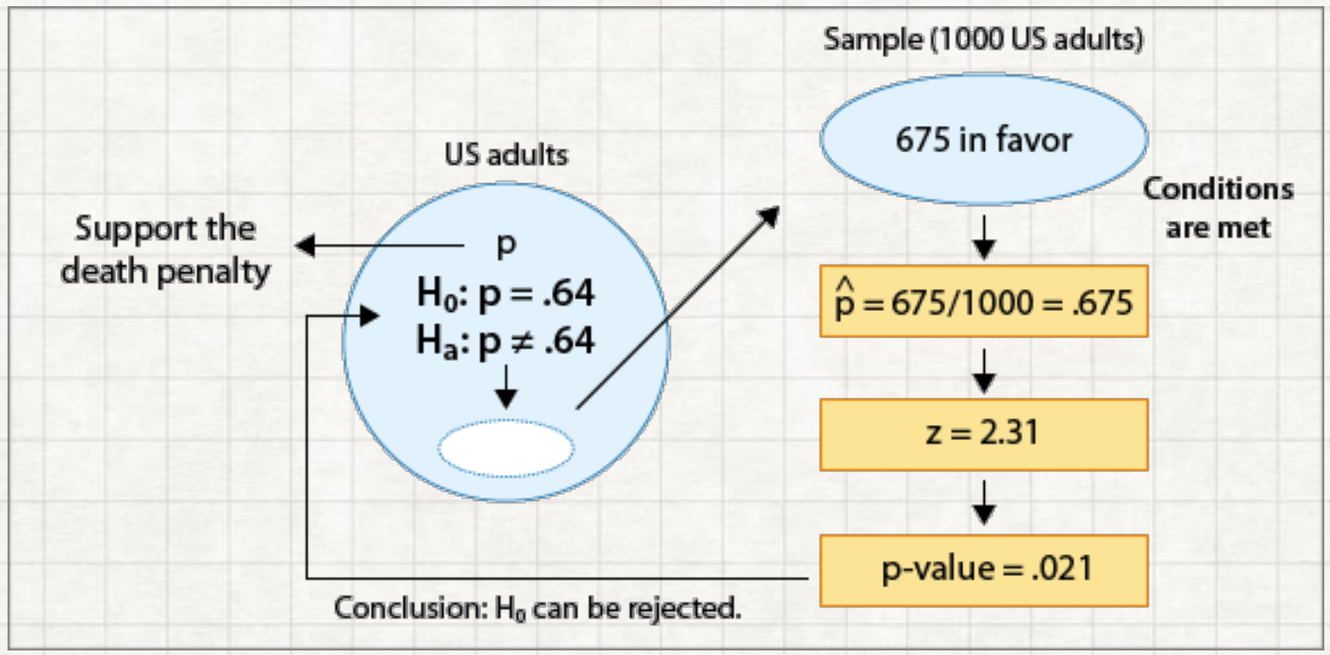
A 95% confidence interval for p, the proportion of all U.S. adults who support the death penalty, is:

$$0.675 \pm 2\sqrt{\frac{0.675(1-0.675)}{1000}} \approx 0.675 \pm 0.03 = (0.645,\ 0.705)$$

Since the 95% confidence interval for p does not include 0.64 as a plausible value for p.we can reject Ho and conclude (as we did before) that the proportion of U.S. adults who support the death penalty for convicted murderers has changed since 2003.

Recall the example

where we wanted to know whether the proportion of U.S. adults who support the death penalty for convicted murderers has changed since 2003, when it was 0.64.



We are testing:

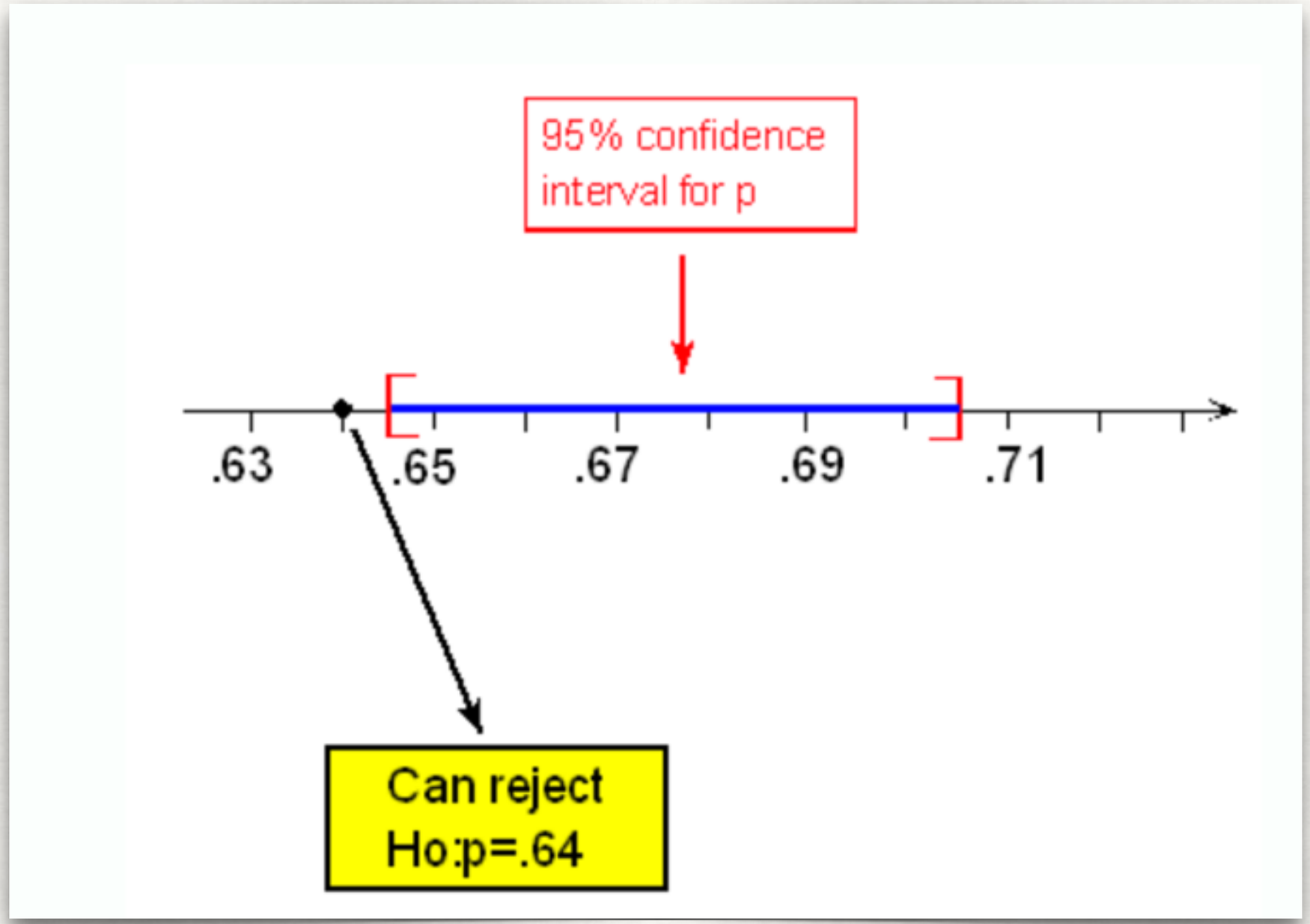
$$H_0 : p = .64$$
$$Ha : p \neq .64$$

$$\hat{p} = 0.675$$

$$0.675 \pm 2\sqrt{\frac{0.675(1-0.675)}{1000}} \approx 0.675 \pm 0.03 = (0.645,\ 0.705)$$

Since the 95% confidence interval for p does not include 0.64 as a plausible value for p.we can reject Ho and conclude (as we did before) that the proportion of U.S. adults who support the death penalty for convicted murderers has changed since 2003.

95% CONFIDENCE INTERVAL FOR P DOES NOT INCLUDE 0.64



MAXIMUM LIKELIHOOD ESTIMATION

MAXIMAL LIKELIHOOD ESTIMATION

- When we do data analysis in a parametric way, we start by characterizing our particular sample statistically;

  Then using a probability distribution (or mass function). This distribution has some parameters. Lets refer to these as $\boldsymbol{\lambda}$.

- If we assume that our data was generated by this distribution, then the notion of the true value of the parameter makes sense.

- Usually in life, there is no way of knowing if this was the true generating process, unless we have some physics or similar ideas behind the process.

- But lets stick with the myth that we can do this. Then let us call the true value of the parameters as $\boldsymbol{\lambda}*$.

- To know this true value, we'd typically need the entire large population eg Ho — null hypothesis

IT'S HARD TO DECIDE NULL HYPOTHESIS

- In the context of frequentist statistics, the assumption is that the parameters are fixed, and that there is this true value ($\boldsymbol{\lambda}*$), and that we can make some estimate of this from our sample ($\boldsymbol{\lambda}$)

- eg:

- Consider a hospital where 400 patients are admitted over a month for heart attacks

- A month later 72 of them have died and 328 of them have survived.

- We can ask, what's our estimate of the mortality (death) rate overall?

Under the frequentist paradigm (central limit)
we must first establish our reference population

i.e. Ho — hypothesis testing, we must have some general populations result!!

- What do we think our reference population is here?
- One possibility is we could think about heart attack patients in the region.
- Another is we could think about heart attack patients that are admitted to this hospital, but over a longer period of time.

FREQUENCY DIFFICULTY

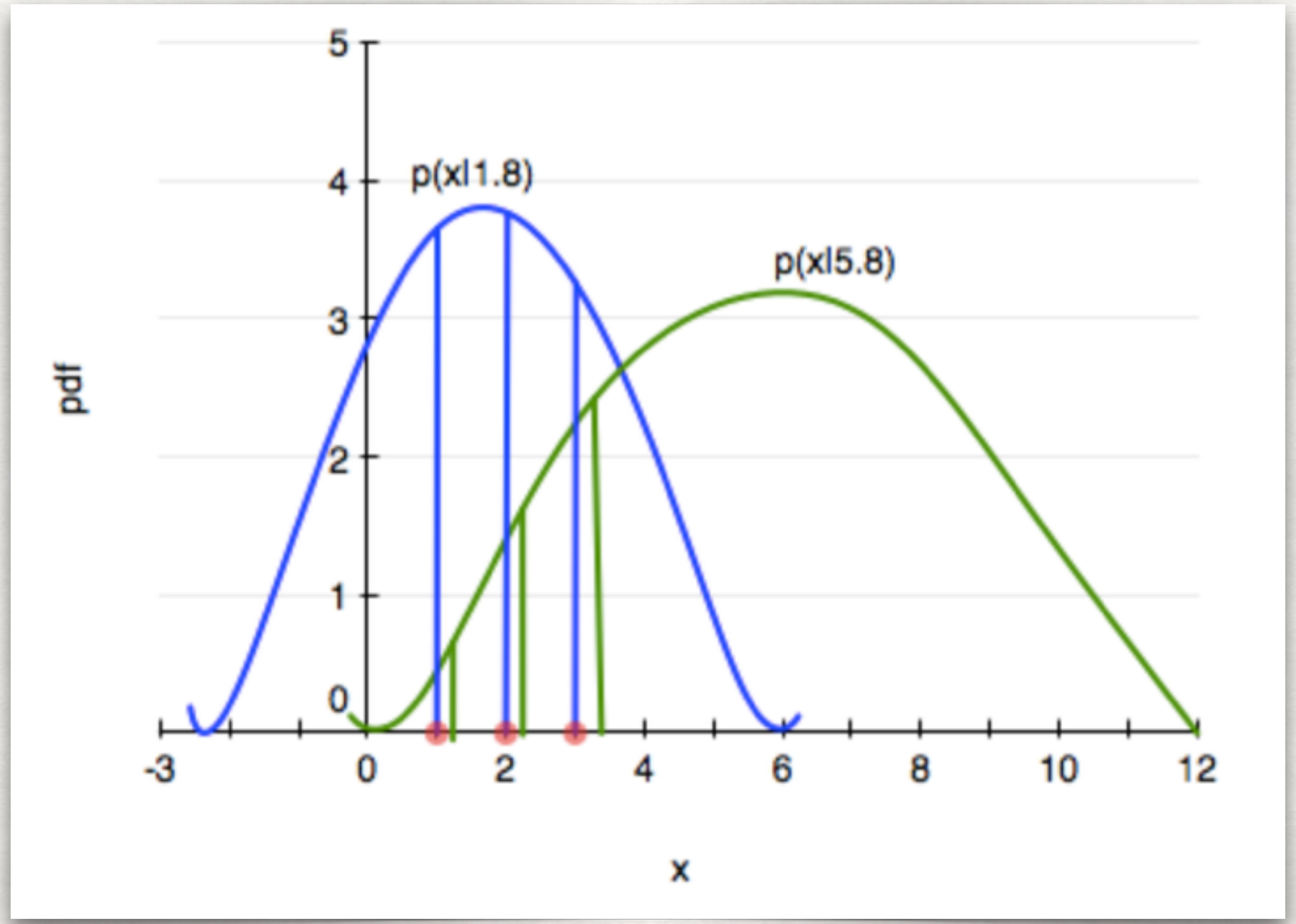
- In the context of frequentist statistics, the assumption is that the parameters are fixed, and that there is this true value ($\boldsymbol{\lambda}*$), and that we can make some estimate of this from our sample ($\boldsymbol{\lambda}$)

- Our question is: how do we estimate ($\boldsymbol{\lambda}$) ?
  And how do we compute this sampling distribution so that we can get a notion of the uncertainty that estimating from a sample rather than the population leaves us with?

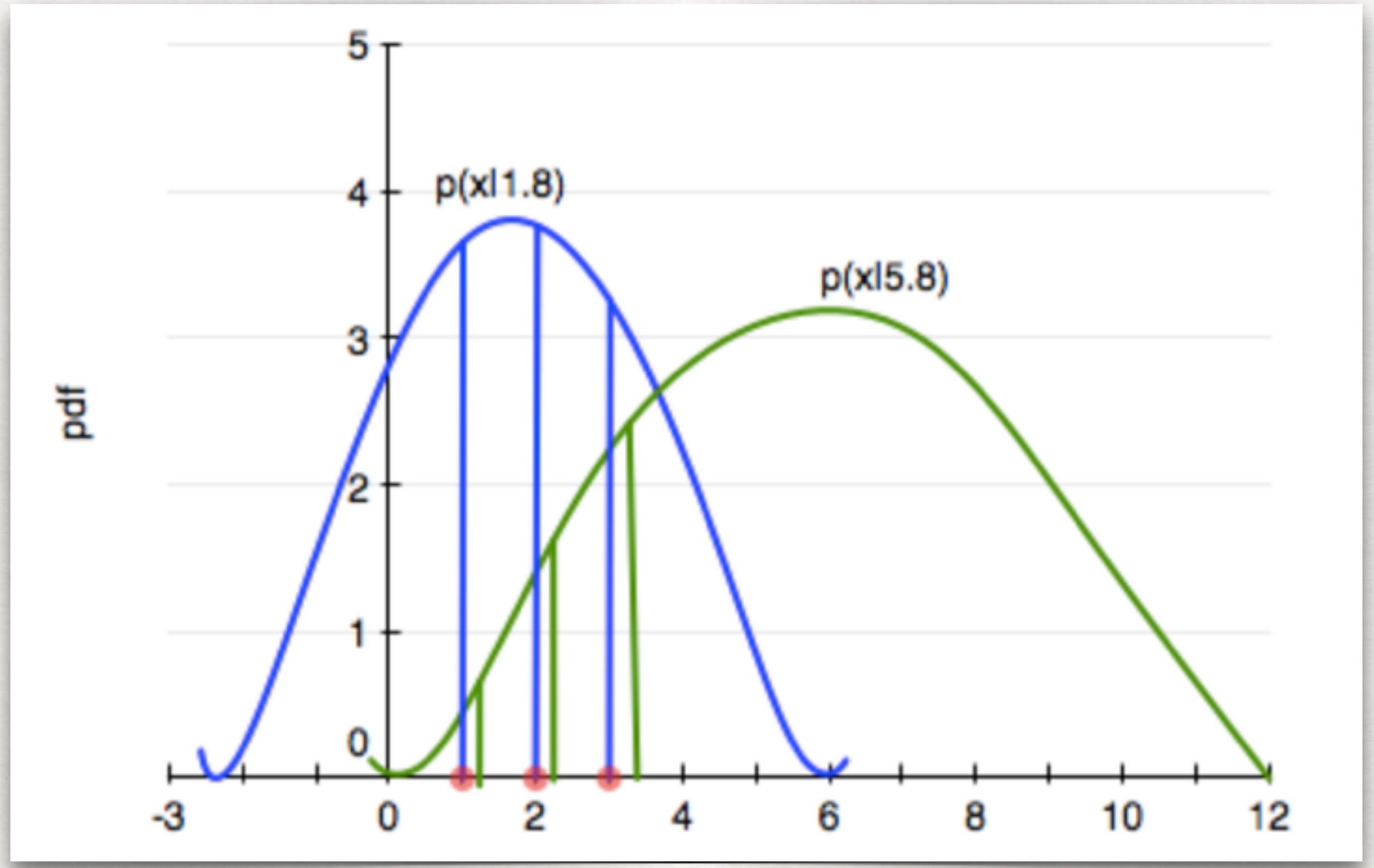
- The first question is tackled by the Maximum Likelihood estimate, or MLE. The second one is tackled by techniques like the bootstrap.

MLE

- The diagram below illustrates the idea behind the MLE.



MLE



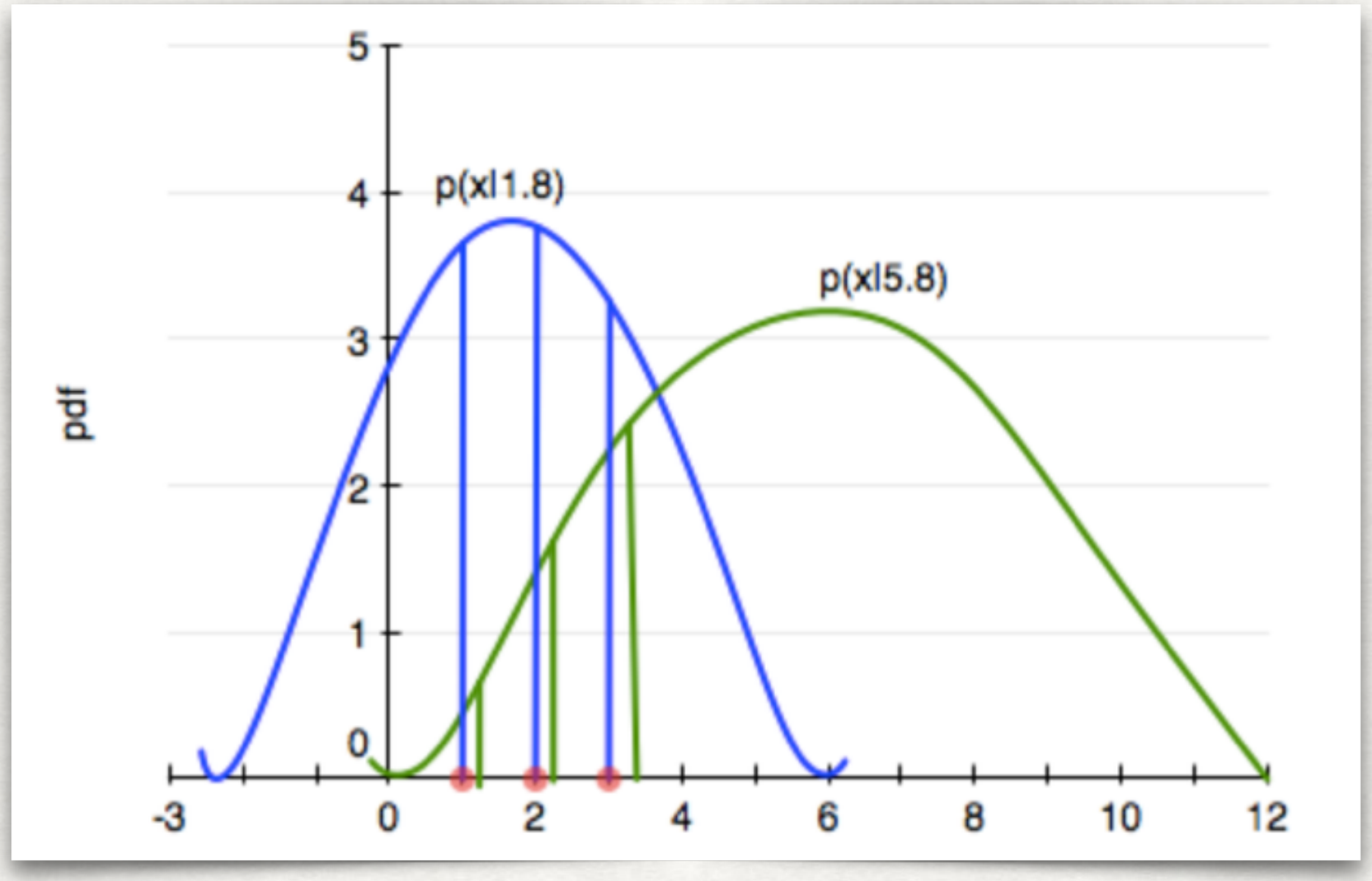
In our case the blue is more likely since the product of the height of the 3 vertical blue bars is higher than that of the 3 green bars.

x

- Consider two distributions in the same family, one with a parameter, lets call it θ, of value 1.8 (blue) and another of value 5.8. (green). Let's say we have 3 data points, at x=1,2,3.

MLE

- Indeed the question that MLE asks is: how can we move and scale the distribution, that is, change θ, until the product of the 3 bars is maximized!



MLE

- we can make some estimate of this from our sample (λ)
- eg:
- Consider a hospital where 400 patients are admitted over a month for heart attacks
- A month later 72 of them have died and 328 of them have survived.
- We can ask, what's our estimate of the mortality (death) rate overall?

MLE

- Conditional on the fixed value of λ, which distribution is the data more likely to have come from?
- change λ, until the product of the 3 bars is maximized!
- That is, the product

$$L(\lambda) = \prod_{i=1}^{n} P(x_i \mid \lambda)$$

MLE

$$L(\lambda) = \prod_{i=1}^{n} P(x_i \mid \lambda)$$

- Measure of how likely it is to observe values x1,…,xn given the parameters **λ**.
- Maximum likelihood fitting consists of choosing the appropriate "likelihood" function L=P(X|**λ**) to maximize for a given set of observations. How likely are the observations if the model is true?
- Often it is easier and numerically more stable to maximize the log likelihood

$$\ell(\lambda) = \sum_{i=1}^{n} ln(P(x_i \mid \lambda))$$

MLE

- The exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a homogeneous Poisson process.

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

- Given samples X = (x1,…,xn) follows exp(**λ**), what **λ** would give the best probability?