Suppose you are watching a radioactive source that emits particles at a rate described by the exponential density

$$f(t) = \lambda e^{-\lambda t} \ ,$$

where $\lambda = 1$, so that the probability $P(0,T)$ that a particle will appear in the next T seconds is $P([0,T]) = \int_0^T \lambda e^{-\lambda t}\, dt$. Find the probability that a particle (not necessarily the first) will appear

(c) between 3 and 4 seconds from now.

(d) after 4 seconds from now.

Need to assume "now" is at some value To

CENTRAL LIMIT, Z-TABLE, MLE

“

REFERENCES:

1) *Use z-table: https://github.com/letaoZ/MAT331/blob/master/week05/Use_of_z_table.pdf*

”

THE NORMAL DISTRIBUTION: AS MATHEMATICAL FUNCTION (PDF)

$$f(x) = \frac{1}{\boxed{\sigma}\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\boxed{\mu}}{\sigma}\right)^2}$$

This is a bell shaped curve with different centers and spreads depending on μ and σ

<u>Note constants:</u>

π=3.14159

e=2.71828

THE NORMAL DISTRIBUTION

- It's a probability function, so no matter what the values of μ and **σ**, must integrate to 1!

$$E(X)=\mu = \int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$Var(X)=\sigma^2 = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx) - \mu^2$$

$$\text{Standard Deviation}(X)=\sigma$$

THE NORMAL DISTRIBUTION

GAUSSIAN

- 68-95-99.7 Rule (more data later)
  in Math terms…

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = .68$$

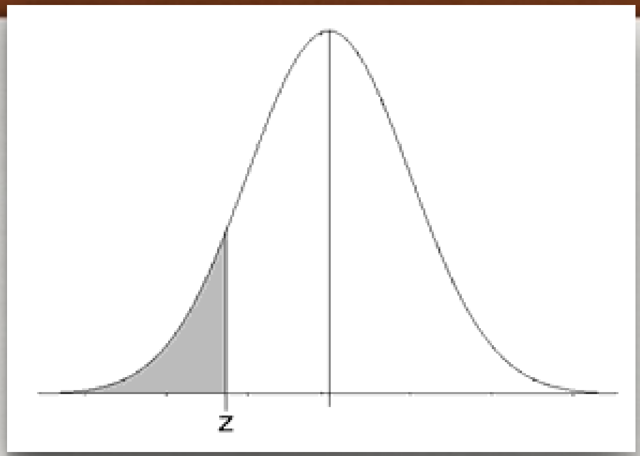
$$\int_{\mu-2\sigma}^{\mu+2\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = .95$$

$$\int_{\mu-3\sigma}^{\mu+3\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = .997$$

STANDARD NORMAL DISTRIBUTION

- $Z \sim N(0,1)$
- If a random variable Z follows a Gaussian distribution with mean 0 and variance 1, we say: X follows a **standard normal distribution**
- **Any Gaussian random variable can be written as**
- **$\mu + \sigma Z$, $Z \sim N(0,1)$**
- **(examples)**

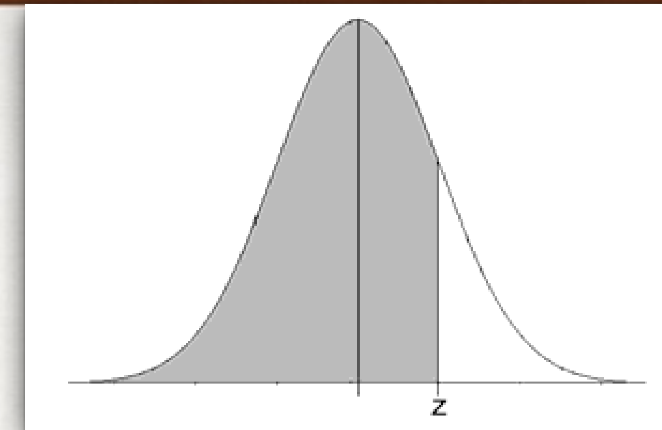
Standard Normal Cumulative Probability Table



Cumulative probabilities for NEGATIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985

Standard Normal Cumulative Probability Table



Cumulative probabilities for POSITIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890

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)

WHAT IS WEIRD?

GAUSSIAN

- **Normal curve:**

  No matter what μ and σ are,

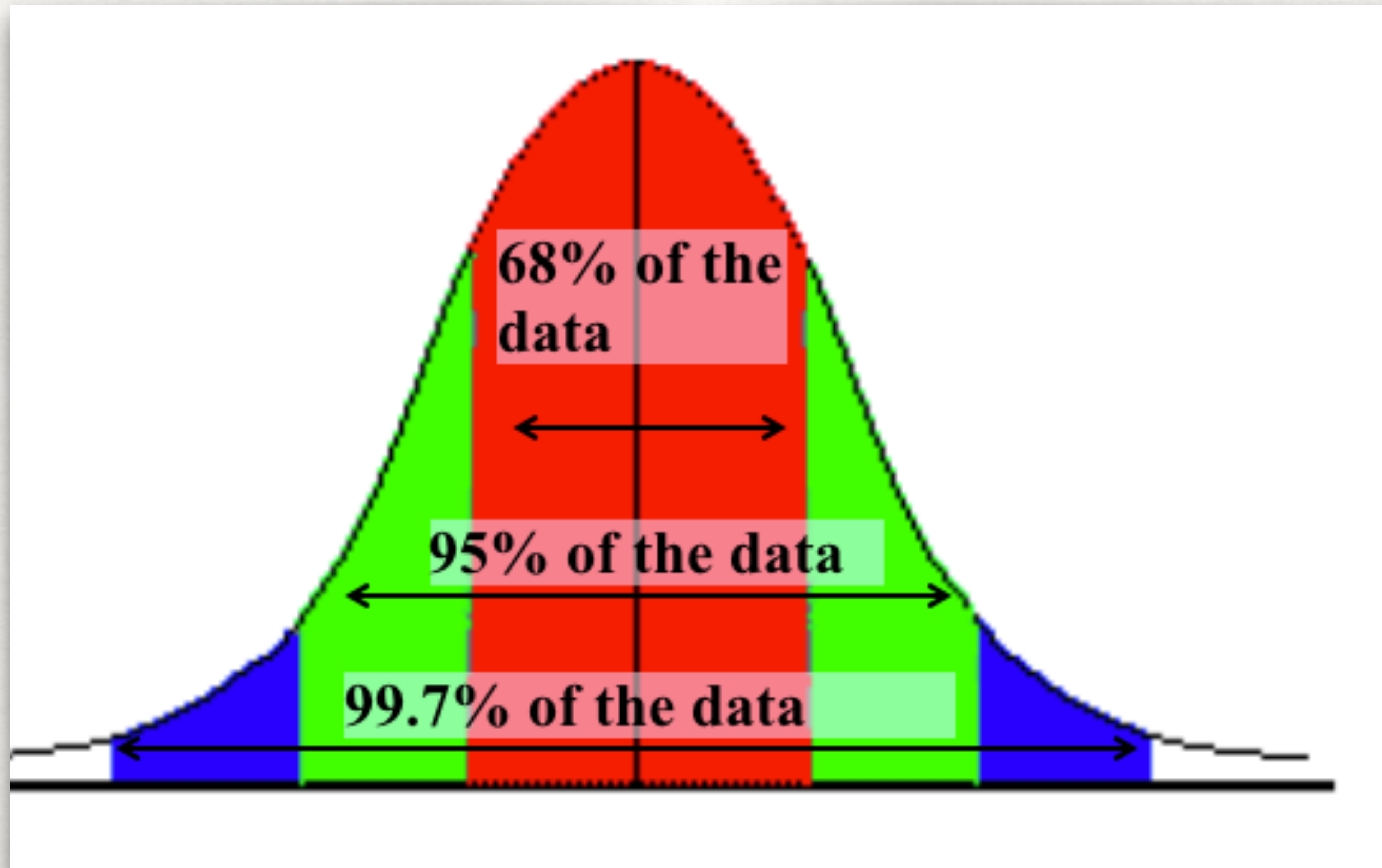
  - the area between $\mu-\sigma$ and $\mu+\sigma$ is about 68%;
  - the area between $\mu-2\sigma$ and $\mu+2\sigma$ is about 95%;
  - and the area between $\mu-3\sigma$ and $\mu+3\sigma$ is about 99.7%.

- Almost all values fall within 3 standard deviations.

- 68-95-99.7 Rule

THE NORMAL DISTRIBUTION

GAUSSIAN



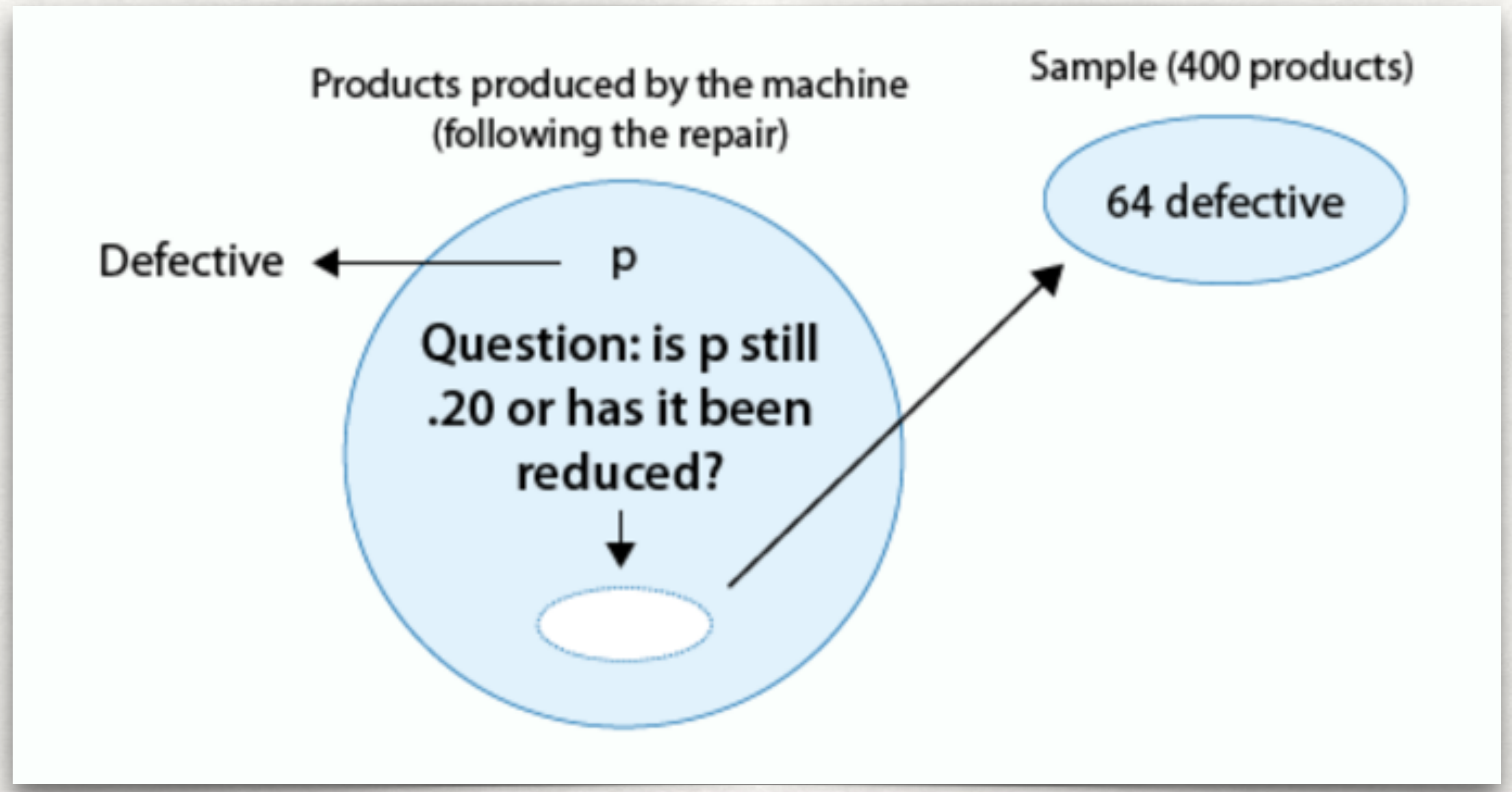
APPLICATION OF CENTRAL LIMIT: HYPOTHESIS TESTING

- You observed some facts (eg. flip a coin 100 times, 80H 20T)
- H0 = null hypothesis: observed fact is the same as general fact (coin is a fair coin: P(head) = 1/2)
- Ha = alternative hypothesis: observed fact is different from the general population
(coin has heavy head: P(head) > 1/2)
- P-value = probability that we observed the fact assuming H0
- **α** = confident level = 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?



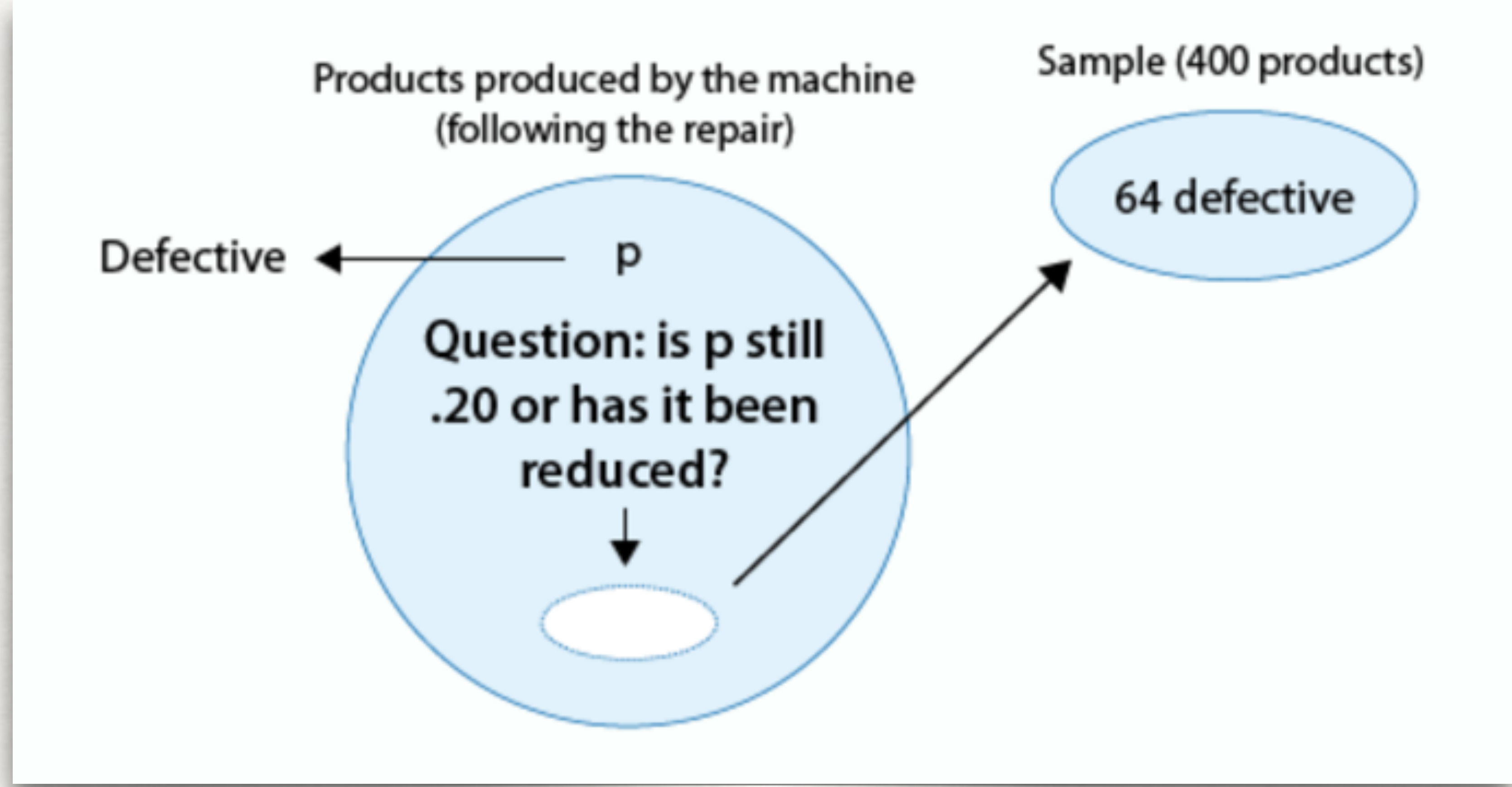
EXAMPLE

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?
- Ho — general population hypothesis
- Ha — alternative hypothesis (based on observation)

EXAMPLE

DETERMINE H0, HA



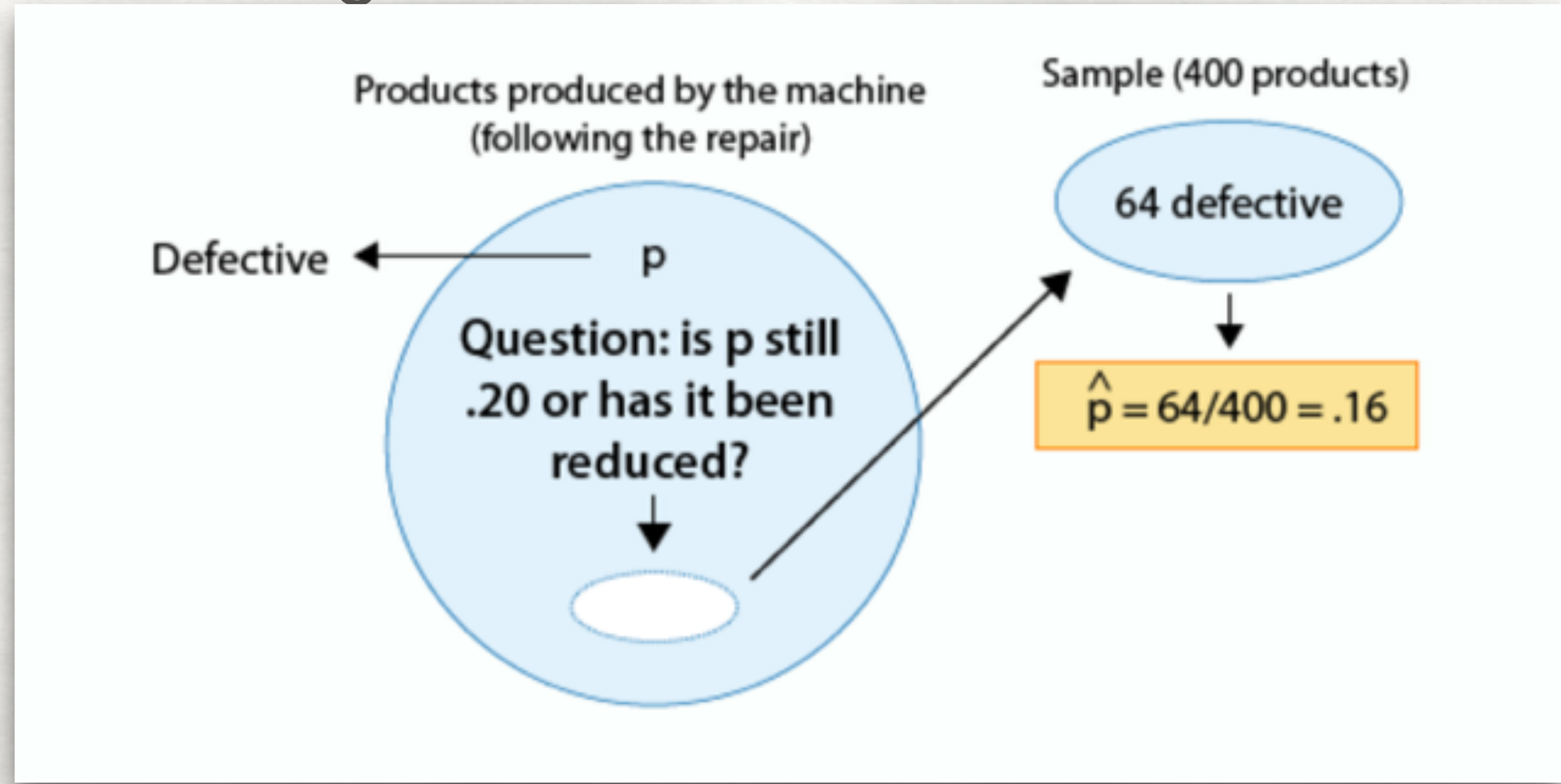
- Let p be the actual portion of defective product (there could be millions of products the machine made, and we only sampled out 400 of them)
- Ho: $p = 0.20$ (No change; the repair did not help).
- Ha: $p < 0.20$ (The repair was effective).

EXAMPLE

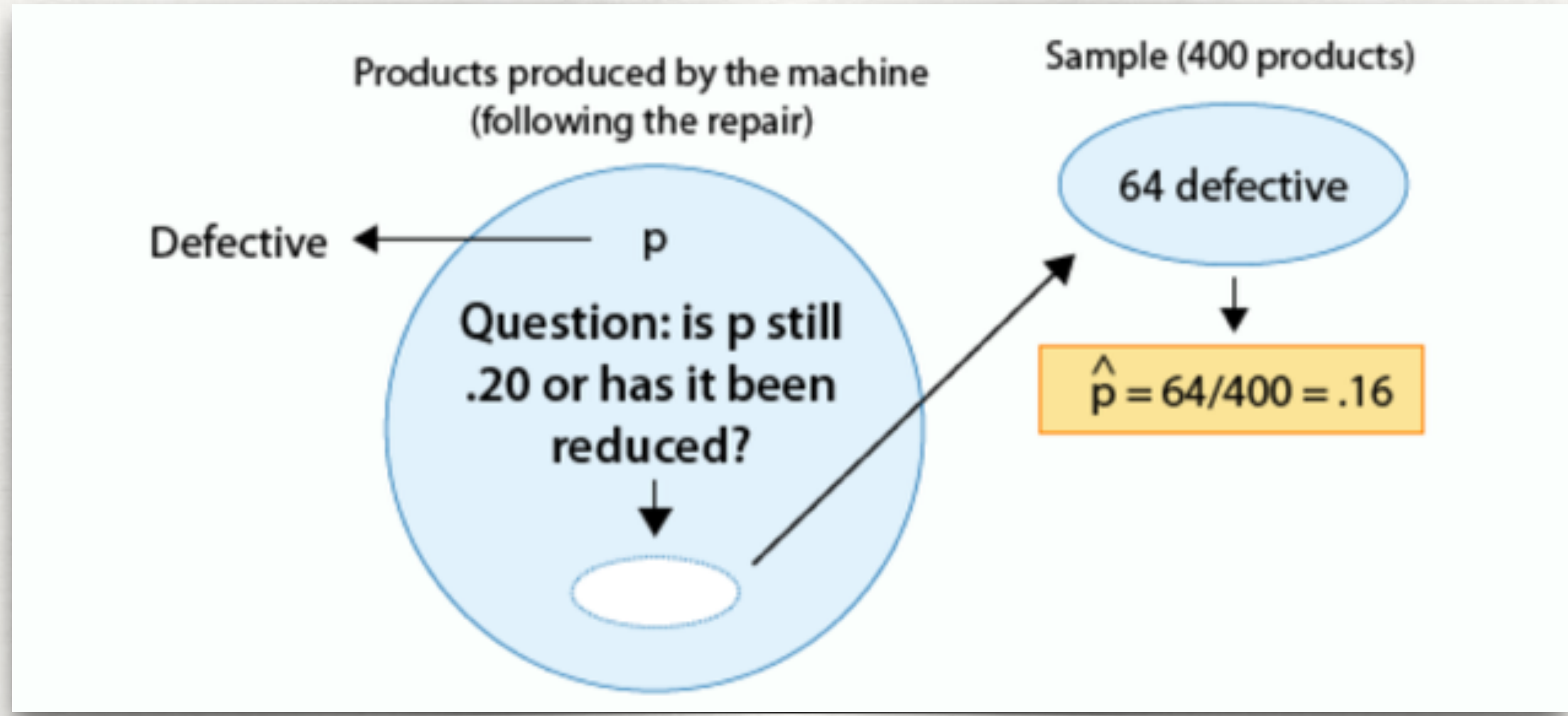
- Recall that there are basically 4 steps in the process of hypothesis testing:
- 1. State the null and alternative hypotheses.
- 2. **Collect relevant data** from a random sample and summarize them (using a test statistic). — calculating z
- 3. Find the p-value, the probability of observing data like those observed assuming that Ho is true. — calculation using N(0,1)
- 4. Based on the p-value, decide whether we have enough evidence to reject Ho (and accept Ha), and draw our conclusions in context.

TEST STATISTIC

- Has the proportion of defective products been reduced as a result of the repair?
- Ho: $p = 0.20$ (No change; the repair did not help).
- Ha: $p < 0.20$ (The repair was effective).
- Fact: portion changed to 0.16



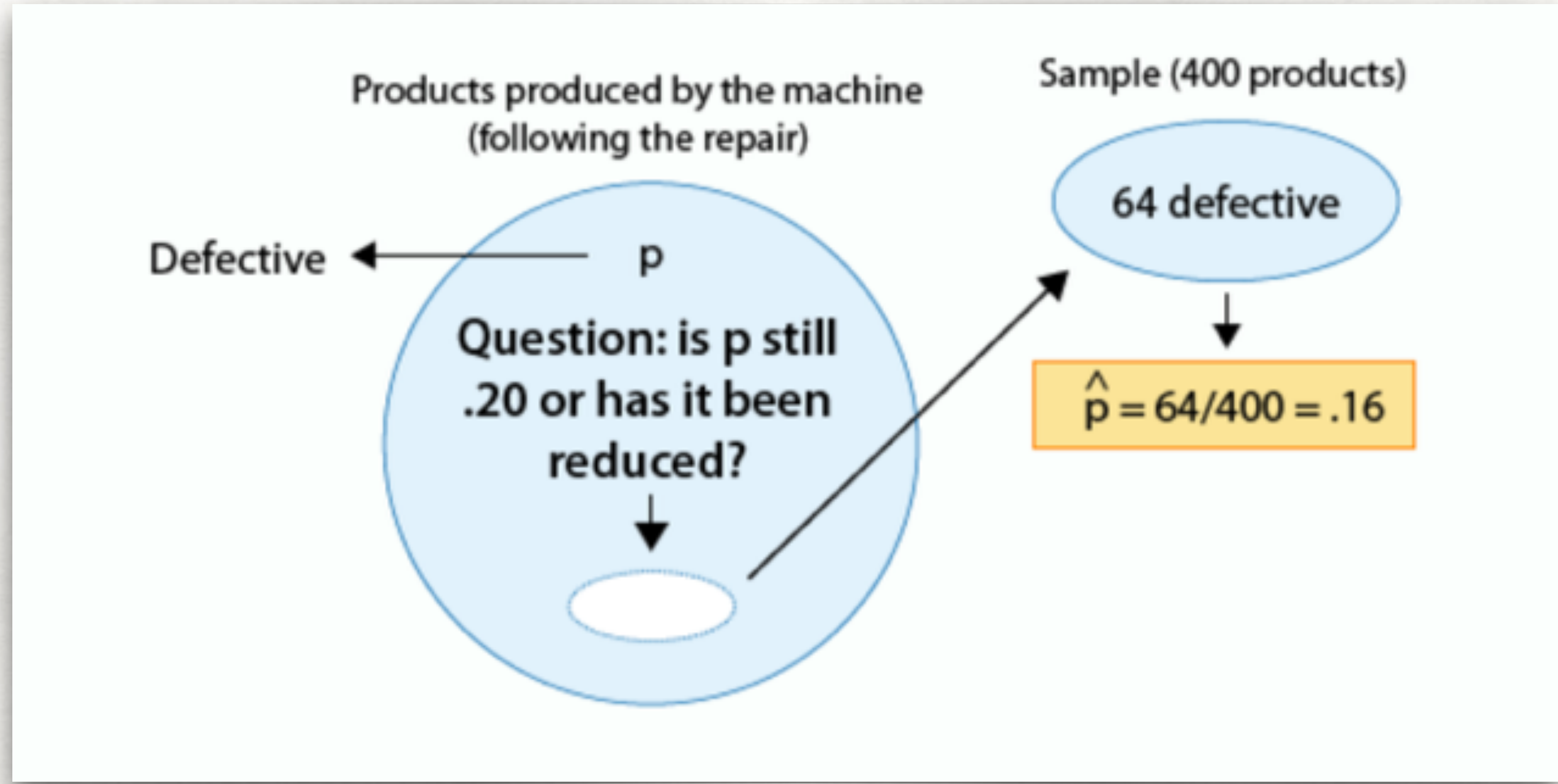
TEST STATISTIC



- The data estimate p to be 0.16
  The null hypothesis claims that p = 0.20
- The data are therefore 0.04 (or 4 percentage points) below the null hypothesis with respect to what they each tell us about p.

TEST STATISTIC

- Has the proportion of defective products been reduced as a result of the repair?



- test statistic: is a measure of how far the sample proportion 0.16 is from the null value 0.2, the value that the null hypothesis claims is the value of p.

- In other words, it is a measure of the "distance" between what the data tells us about p and what the null hypothesis claims p to be.

THIS IS WHERE CENTRAL LIMIT THEOREM COMES IN: TO GIVE A PROBABILISTIC MEASURE OF THE DIFFERENCE

”

HYPOTHESIS TESTING FOR THE POPULATION PROPORTION P: Z-SCORE

- For the reason illustrated in the examples at the end of the previous page, the test statistic cannot simply be the difference

- Central Limit says that (for a Bernoulli(p) distribution)

- 1. When we take a random sample of size n from a population (with mean p), the possible values of the sample proportion (when certain conditions are met) have approximately a normal distribution with:

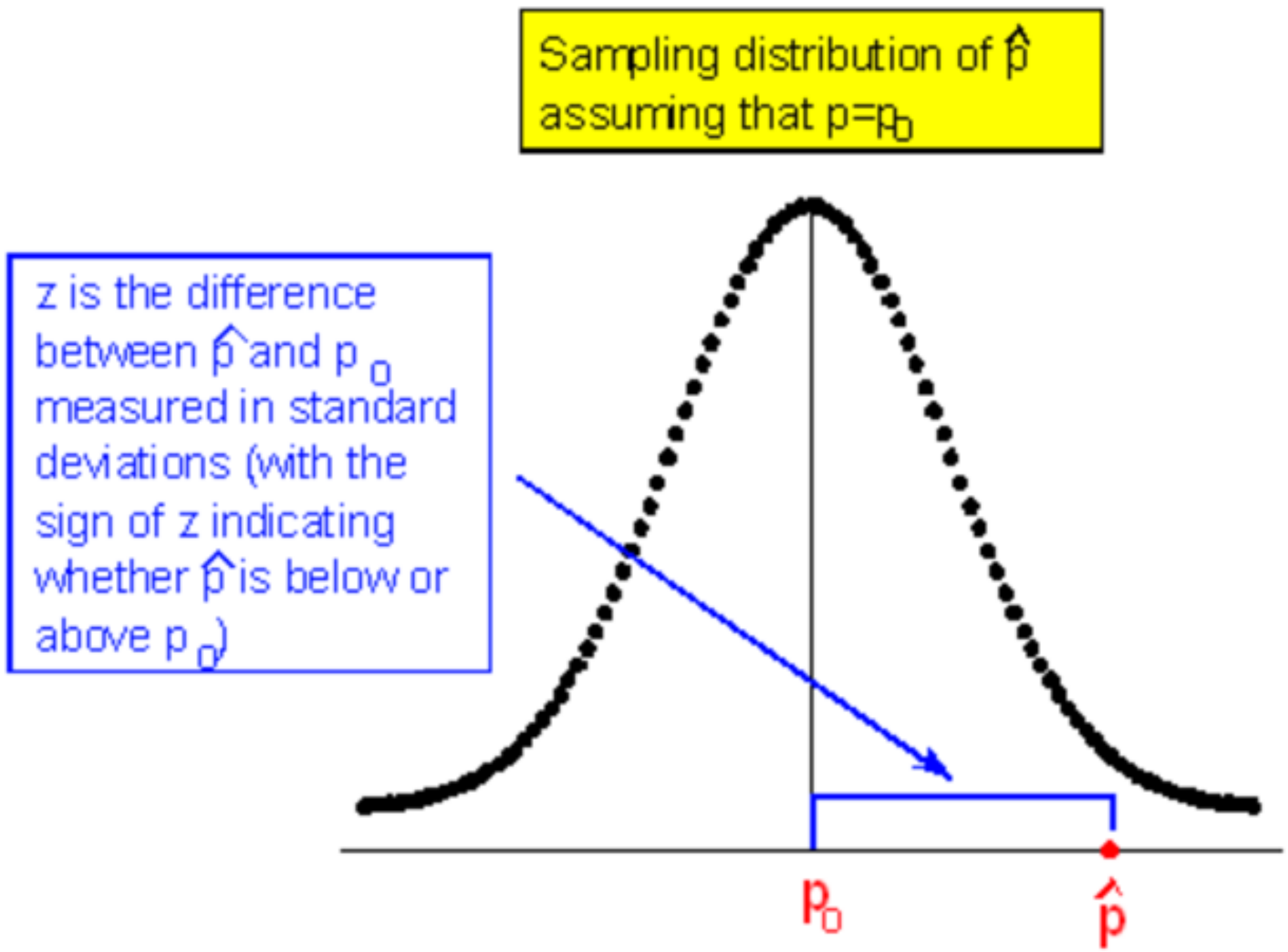
* mean: p

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

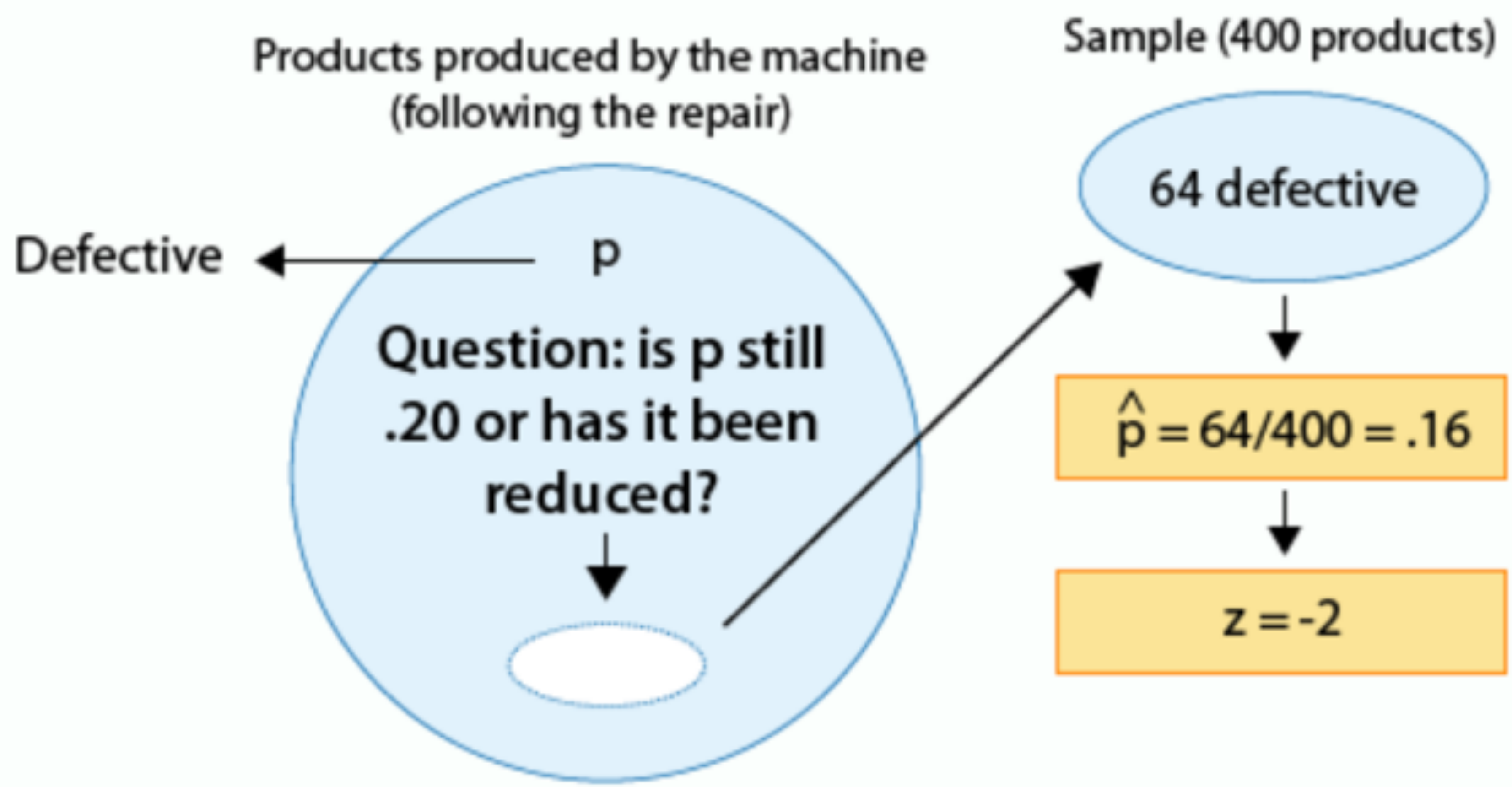
- 2. The z-score of a normal value (a value that comes from a normal distribution)

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

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



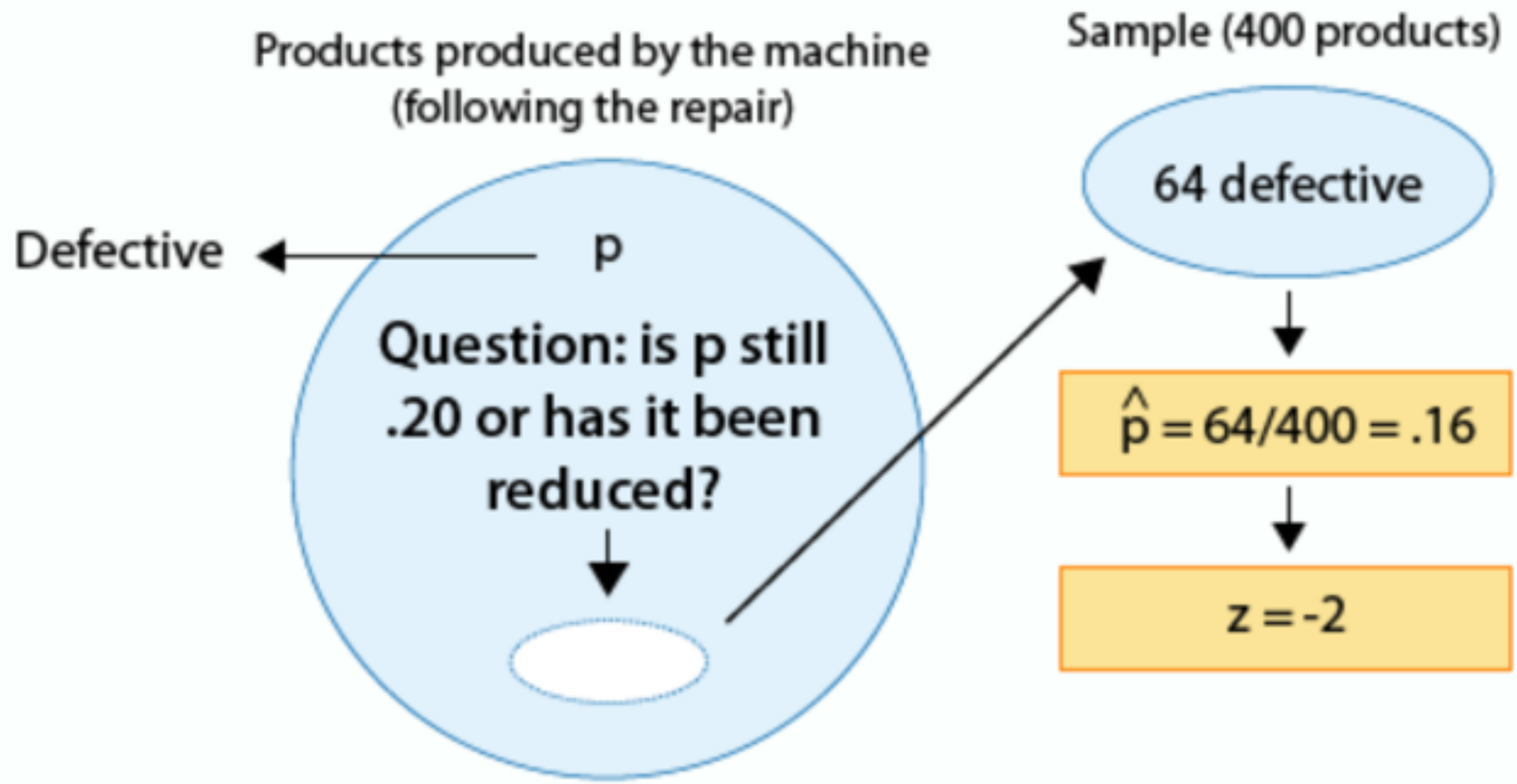
Sampling distribution of $\hat{p}$ assuming that $p=p_0$
z is the difference between $\hat{p}$ and p_0 measured in standard deviations (with the sign of z indicating whether $\hat{p}$ is below or above p_0)
p_0
$\hat{p}$



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

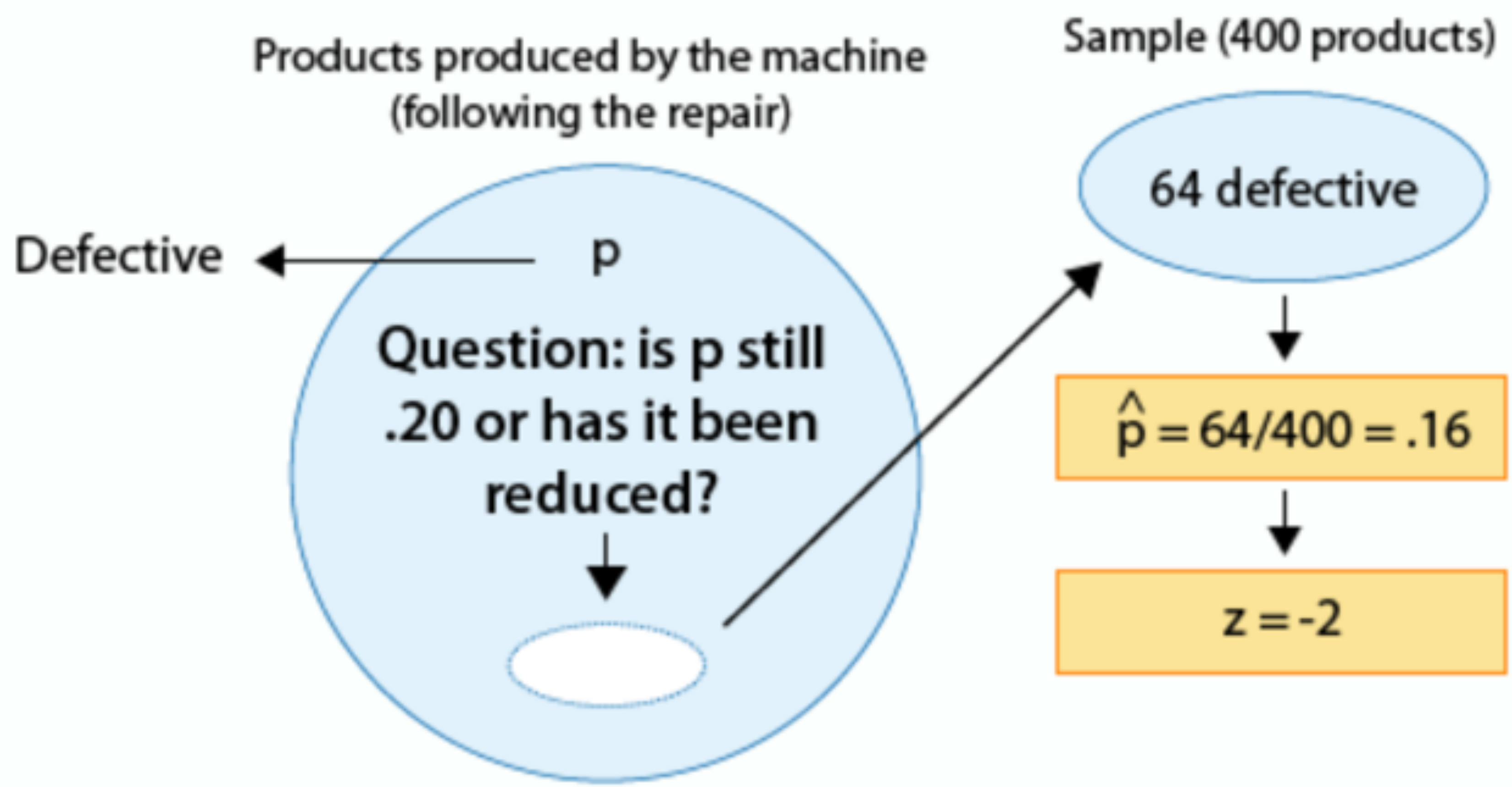
What does this tell me?

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).



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 probability of observing a test statistic as small as -2 or smaller, assuming that Ho is true.
* Intuitively, the p-value is the probability of observing data like those observed assuming that Ho is true.

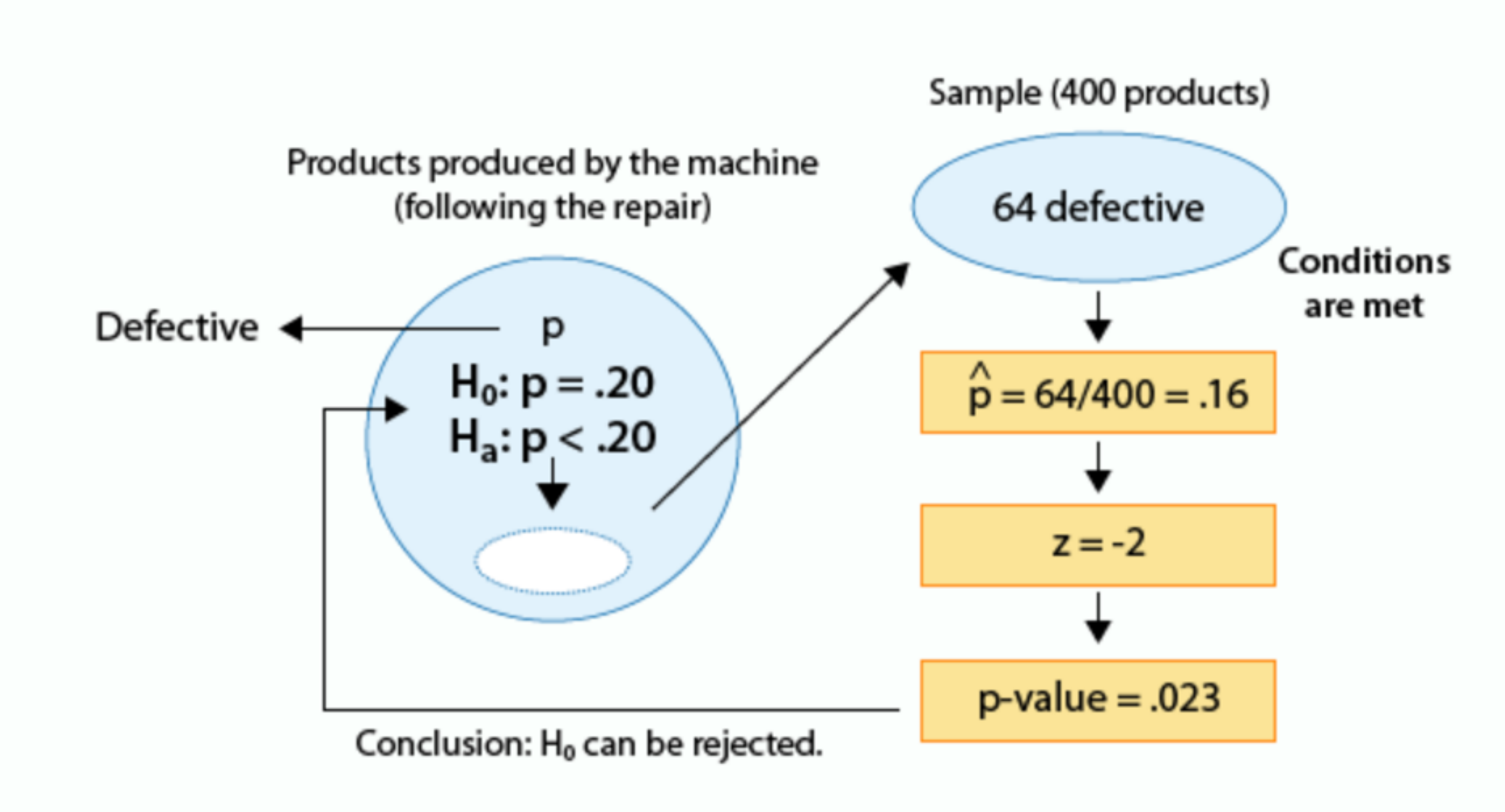


Since the null hypothesis is H_o: p = 0.20, the standardized score of $\hat{p} = 0.16$ is: $z = \frac{0.16-0.20}{\sqrt{\frac{0.20(1-0.20)}{400}}} = -2.$

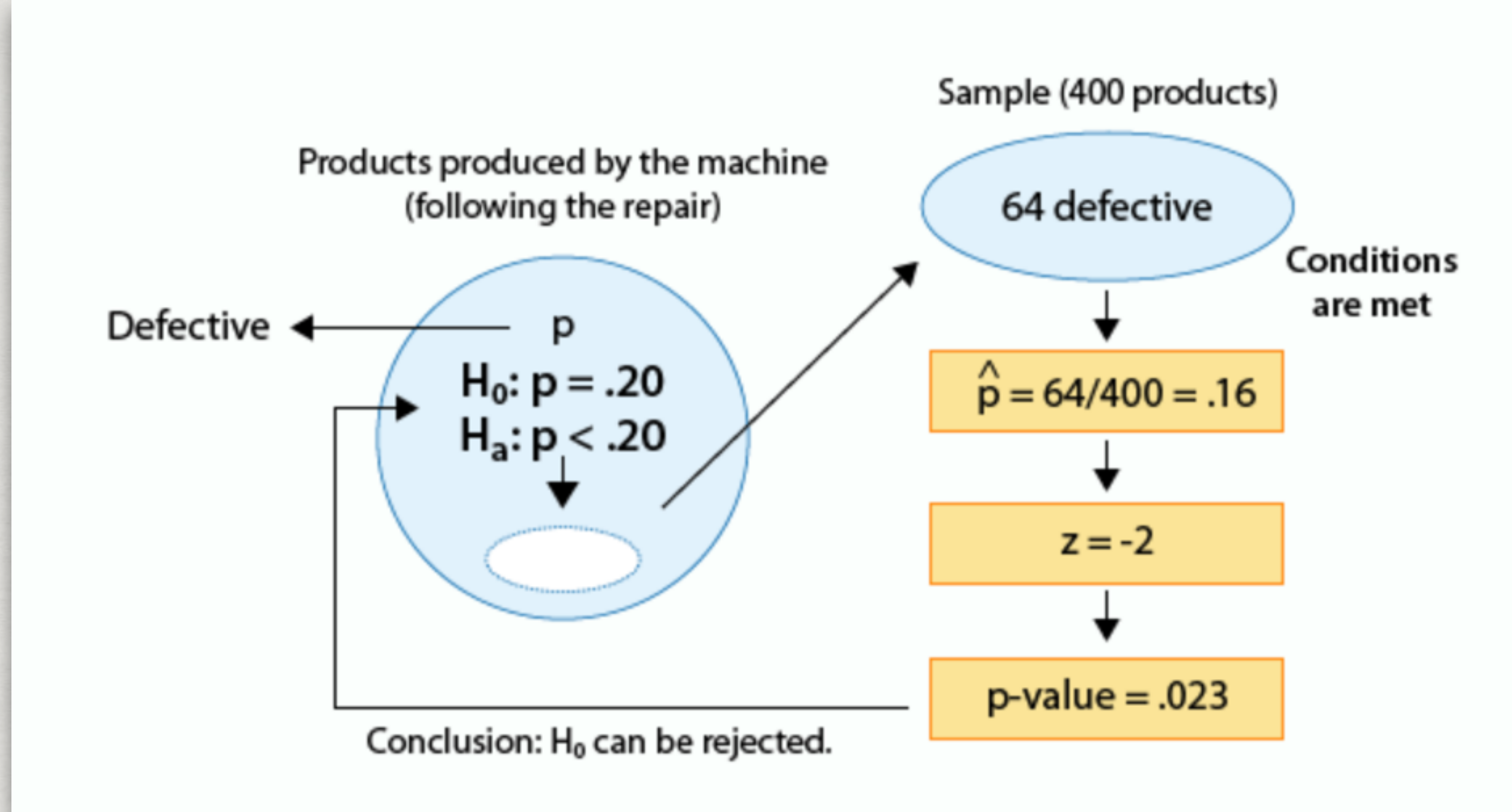
* The probability of observing a test statistic as small as -2 or smaller, assuming that Ho is true.

* Now, we need to find P(Z<-2)

* The value can be found using z-table



Since 0.023 is small (in particular, 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. The following figure is the complete story of this example, and includes all the steps we went through, starting from stating the hypotheses and ending with our conclusions:



The Idea is that: it is highly unlikely (with probability 0.023) that we observe what we see if we assume Ho. So either our theory is wrong or our assumption is wrong

because Central Limit is a Proven Theorem, so we reject Ho and accept Ha (with confident 1-0.023)
And then we will work with Ha.

USE 68-95-99.7 RULE

APPLICATION: P-VALUE AND CONFIDENCE INTERVAL

Single population mean (large n)

- Hypothesis test: first calculate s (sample standard deviation)

$$T_{n-1} = \frac{\text{observed mean} - \text{null mean}}{\frac{s}{\sqrt{n}}}$$

- Confidence Interval

$$\text{observed mean} \pm \mathrm{T}_{n-1,\alpha/2} * \left(\frac{s}{\sqrt{n}}\right)$$