Inferential Statistics

A decision, estimate, prediction, or generalization about a population, based on a sample.

Population: A collection of all possible individuals, objects, or measurements of interest.

Sample: A portion, or part, of the population of interest.

How well can sample represent population and conclusion on population drawn from sample?

Inferential Statistics

Probability:

A measure of the likelihood that a particular event will happen.

It expressed by a value between 0 and 1

Experiment:

Observation of some activity.

Act of taking some measurement.

Outcome:

A particular result of an experiment.

Event:

The collection of one or more outcomes of an experiment.

Definitions of Probability

Classical Probability:

Based on assumption that the outcomes of an experiment are *equally likely.*

$$\text{Probability of an event} = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}}$$

Relative Frequency Probability:

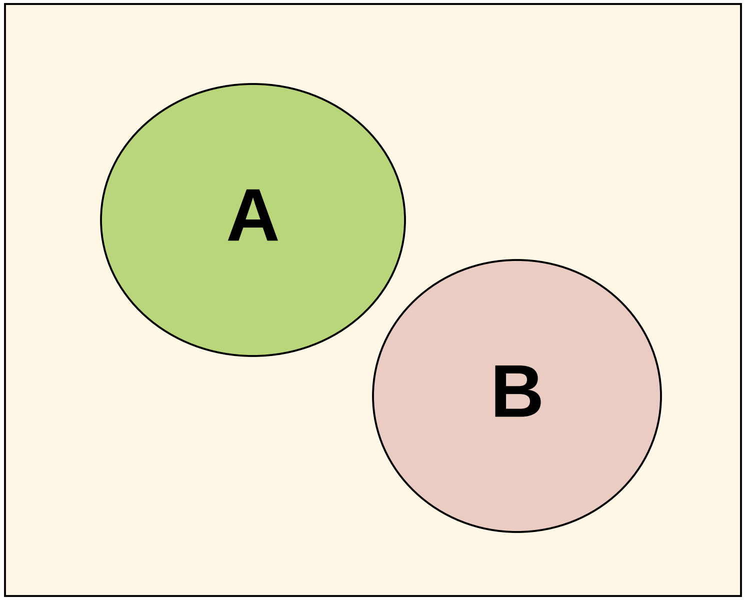
Based on relative frequencies.

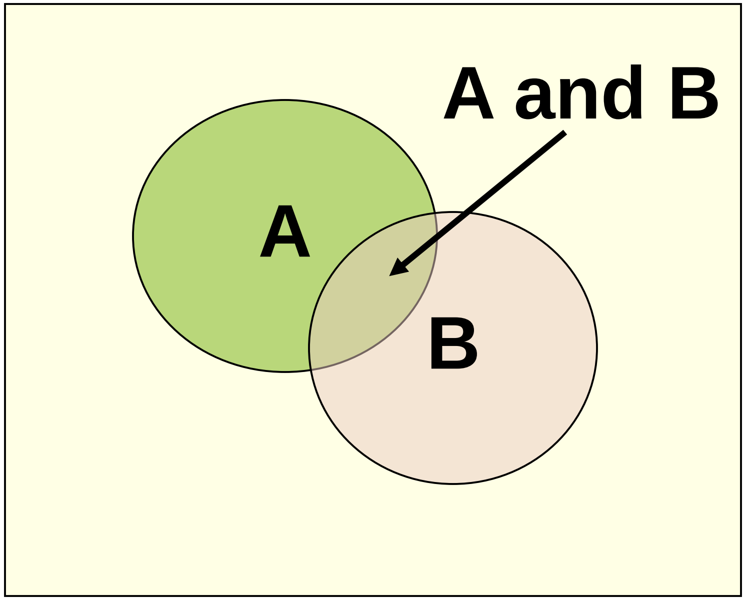
$$\text{Probability of event happening} = \frac{\text{Number of times event occurred in past}}{\text{Total number of observations}}$$

Rules of Addition

Mutually Exclusive:

Two events are mutually exclusive if one event happens, the other cannot happen.





Rules of Addition

Combine events that are mutually exclusive.

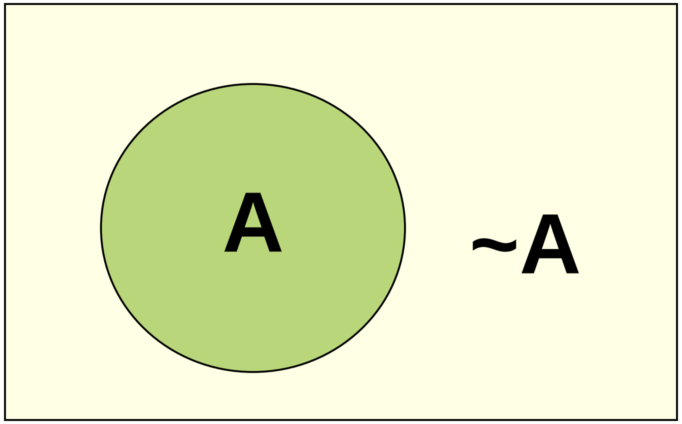
$$P(A \text{ or } B) = P(A) + P(B)$$

Combine events that are *not* mutually exclusive.

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Complement Rule:

$$P(A) = 1 - P(\sim A)$$



Rules of Multiplication

Independent:

Events are independent if the occurrence of one event has no effect on the occurrence of any other event.

Combine events that are *not* independent.

$$P(A \text{ and } B) = P(A) \times P(B|A)$$

Combine events that are independent.

$$P(A \text{ and } B) = P(A) \times P(B)$$

Conditional Probability:

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

Example 1

A financial analyst classifies his study firms by industry and whether the firm has high or low volume of trade.

As shown by the following Table, a total of 100 firms in his study group last year.

	Industry			
	PropCon	Financials	Services	Total
High Volume	17	11	24	52
Low Volume	22	14	12	48
Total	39	25	36	100

A. What is the probability of selecting a company to analyze and finding it is high volume? $\frac{52}{100}$

B. What is the probability of selecting a company to analyze and finding the company is in either financials or services sector? $\frac{25}{100}+\frac{36}{100}=\frac{61}{100}$

C. What is the probability of selecting a company that involved high volume or financial sector?

$$\frac{52}{100}+\frac{25}{100}-\frac{11}{100}=\frac{66}{100}$$

	Industry			
	PropCon	Financials	Services	Total
High Volume	17	11	24	52
Low Volume	22	14	12	48
Total	39	25	36	100

D. Consider only high volume, what is the probability that high volume company is in propcon sector? $\frac{17}{100} \div \frac{52}{100} = \frac{17}{52}$

E. Two companies are selected for analyze. What is the probability both are in financial sector? $\frac{25}{100} \times \frac{24}{99} = \frac{2}{33}$

Discrete Probability Distribution

Random Variable:

A numerical value determined by the outcome of an experiment.

Probability Distribution:

A listing of all the outcome of an experiment and the probability associated with each outcome.

1. Sum of all possible outcomes is 1.00.
2. Probability of outcome is between 0 and 1.
3. Outcomes are mutually exclusive.

Discrete Probability Distribution

Mean of discrete probability distribution:

$$E(X) = \mu = \sum_{i=1}^{n} \left[X_i \cdot P\left(X_i\right) \right]$$

Variance of discrete probability distribution:

$$E\left[(X_i - E(X))^2 \right] = \sigma^2 = \sum_{i=1}^{n} \left[(X_i - \mu)^2 P(X_i) \right]$$

Discrete Probability Distribution

Binomial Probability can be determined by:

$$P(X) = \frac{n!}{X!(n-X)!} p^x q^{n-X}$$

Mean of binomial probability distribution:

$$\mu = np$$

Variance of binomial probability distribution:

$$\sigma^2 = np(1-p)$$

Example 2

Tossing coin for 4 times, find the probability of turning head.

Event	0	1	2	3	4
Outcome	TTTT	HTTT, THTT, TTHT, TTTH	HHTT, THHT, TTHH, HTHT, THTH, HTTH	THHH, HTHH, HHTH, HHHT	HHHH
# Outcome	1	4	6	4	1
Probability	1/16	4/16	6/16	4/16	1/16

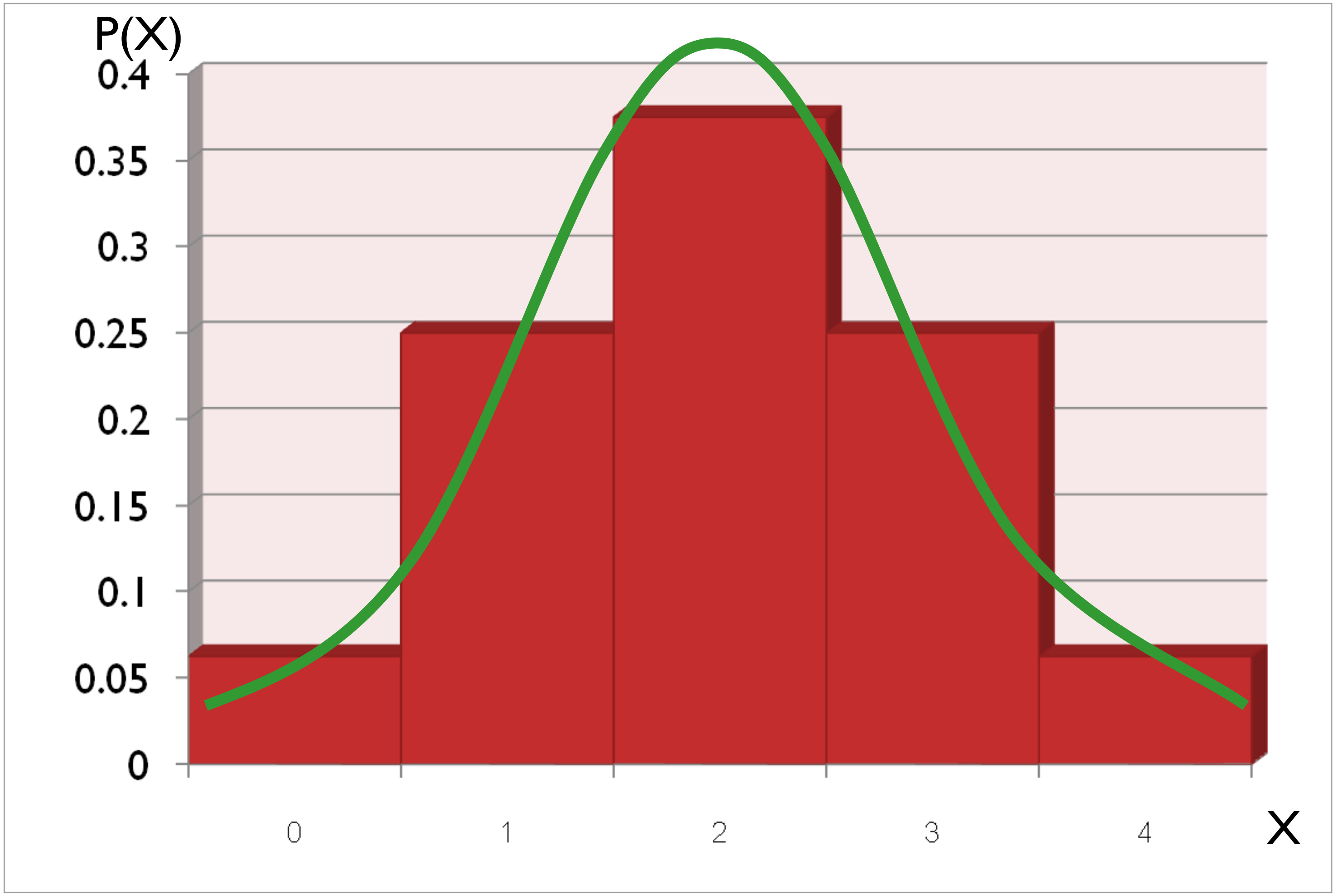
Example 2

Binomial Probability Distribution for $n=4, p=0.5$

Number of Outcome	Probability	
	Fraction	Decimal
0	1/16	0.0625
1	4/16	0.2500
2	6/16	0.3750
3	4/16	0.2500
4	1/16	0.0625
Total	16/16	1.0000

Example 2

Binomial Probability Distribution for $n=4, p=0.5$



Probability Distribution

Discrete Probability Distribution

- Binomial probability distribution.
- Poisson probability distribution.

Continuous Probability Distribution

- Normal probability distribution.
- Student *t* distribution.
- Chi-square distribution.
- *F* distribution.

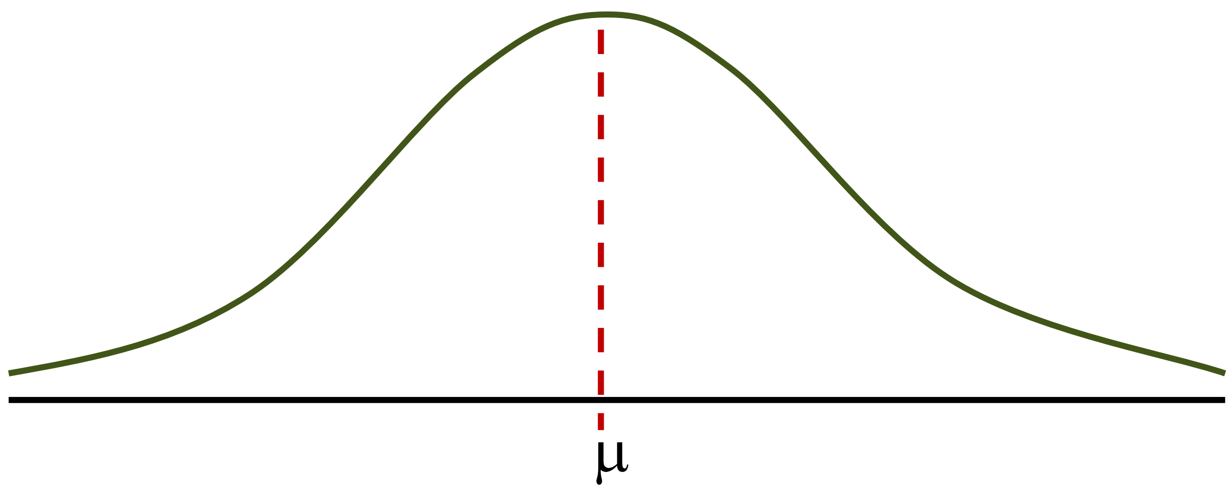
Normal Probability Distribution

A. Bell-shaped and mean and median are equal.

B. Symmetric.

C. Asymptotic (Curve approaches but never touches the X-axis).

D. Completely described by mean and standard deviation.



Standardized Normal Distribution

X can be transformed to be standardized normal distributed by:

$$Z = \frac{X - \mu}{\sigma}$$

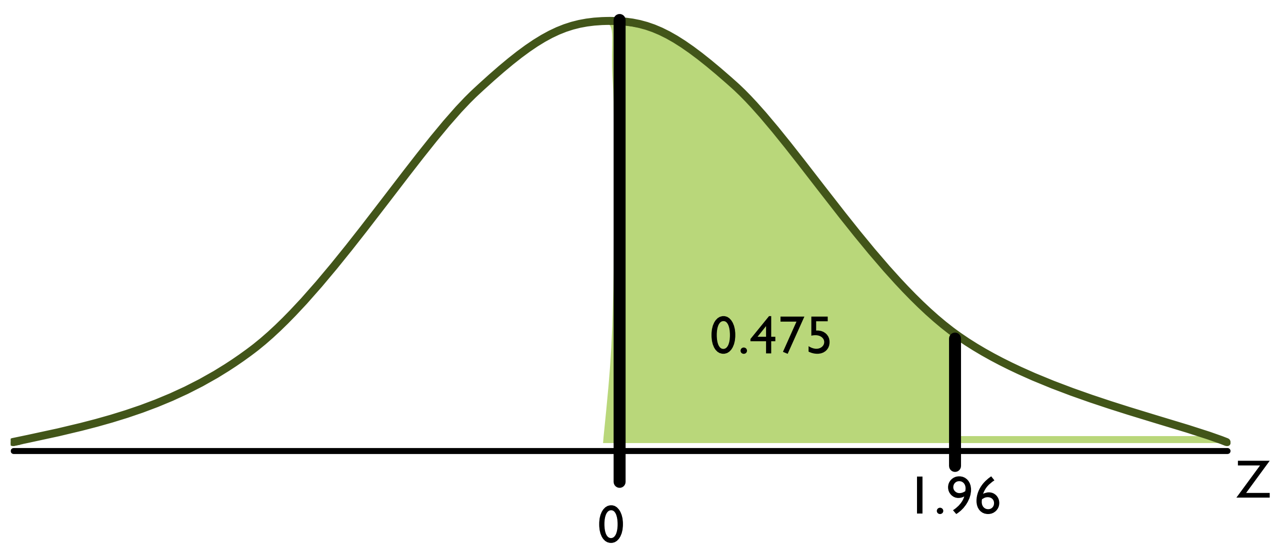
X	$Y = X - \mu$	$Z = \frac{X - \mu}{\sigma}$
1	1 – 2 = -1	$\frac{1-2}{0.8165} = -1.2247$
2	2 – 2 = 0	$\frac{2-2}{0.8165} = 0$
3	3 – 2 = 1	$\frac{3-2}{0.8165} = 1.2247$
$\mu_X = 2$	$\mu_Y = 0$	$\mu_Z = 0$
$\sigma_X = 0.82$	$\sigma_Y = 0.82$	$\sigma_Z = 1$

Standardized Normal Distribution

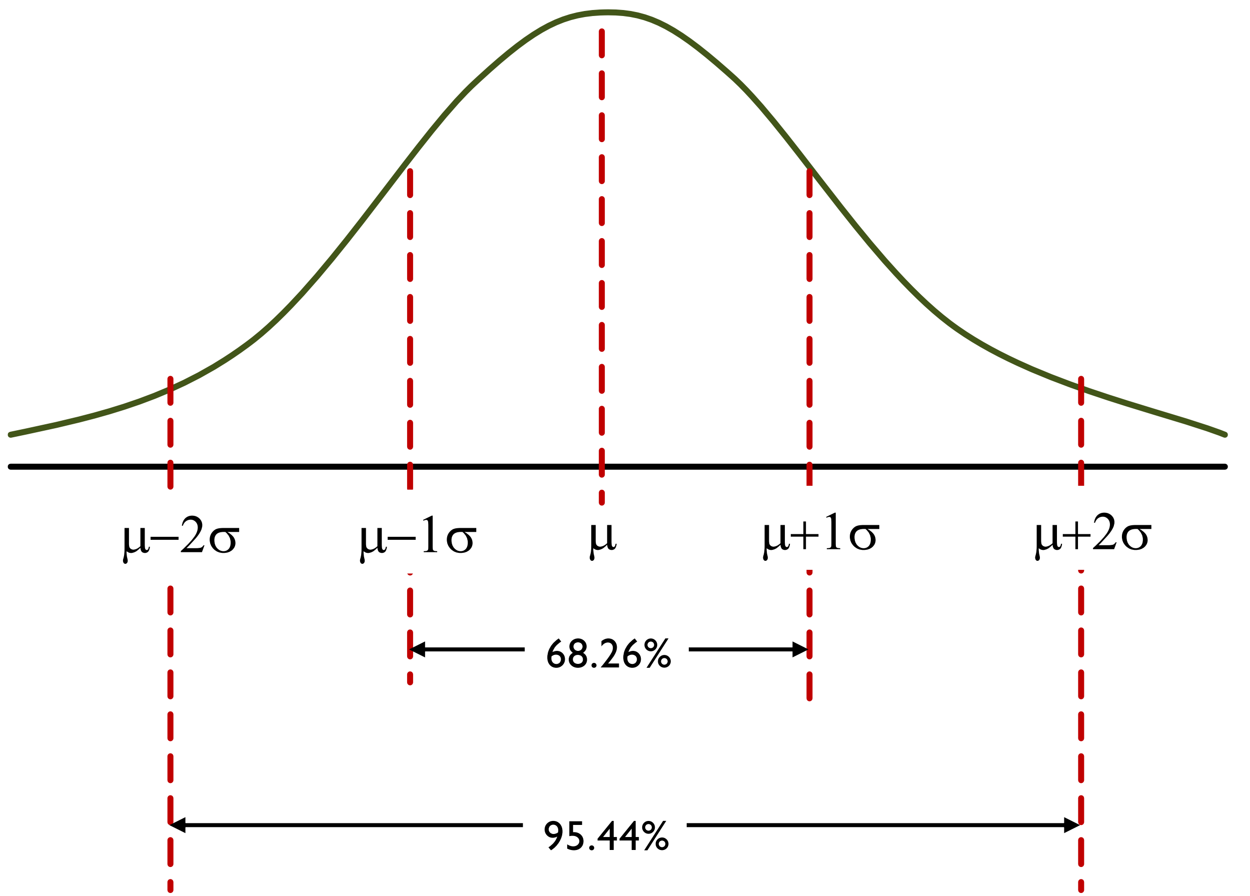
A. Mean equals to 0 and Standard deviation is 1.

B. Normal distribution can be converted to standard normal distribution by:

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$



Area Under Normal Curve



Probability Distribution Function

Normal Probability Density Function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2 / 2\sigma^2}$$

Cumulative Normal Probability Function

$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} \left(\frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2 / 2\sigma^2} \right) dx$$

Sampling Methods

Probability Sample

- Simple Random Sampling
- Systematic Sampling
- Stratified Sampling
- Cluster Sampling

Nonprobability Sample

- Convenience Sampling
- Judgement Sampling

Sampling Methods

Sampling error

The difference between the population parameter and the sample statistic.

Sampling Distribution for the means

A probability distribution of all possible sample means of a given sample size selected from a population and the probability of occurrence associated with each sample mean.

Sampling Methods

Central limit theorem

If the population is normal, then the sampling distribution of the sample means is also normal.

If the population is not normal, the sampling distribution of the sample means approaches normal as the size of the sample increases.

Sampling Distribution

Standard error of the mean

The variation in the sampling distribution of the sample mean.

Population standard error of the mean $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

Sample standard error of the mean $s_{\overline{X}} = \dfrac{s}{\sqrt{n}}$

Standardized sampling distribution of sample mean

$$z = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

Sampling Distribution

Point estimate: a single value used to estimate a population value.

Interval estimate: a range of values within which the population parameter is expected to occur.

The factors that make up a confidence interval for a mean are:

- Number of observations in the sample (*n*).
- Variability in the population—Sample standard deviation.
- Level of confidence.

Confidence interval for mean: $\bar{X} \pm Z \frac{s}{\sqrt{n}}$

Test of Hypotheses

Objective of hypothesis testing

To check the validity of statement about a population parameter.

Procedure used in hypothesis testing:

1. State the null hypothesis H_0 and the alternate hypothesis H_1.
2. Select the level of significance
3. Decide on the test statistic
4. State the decision rule
5. Take sample and make decision regarding H_0
6. The *p-value* is the probability that the test statistic in a hypothesis test is as extreme as the one obtained.

Test of Hypotheses

Type I error: when a true null hypothesis is rejected.

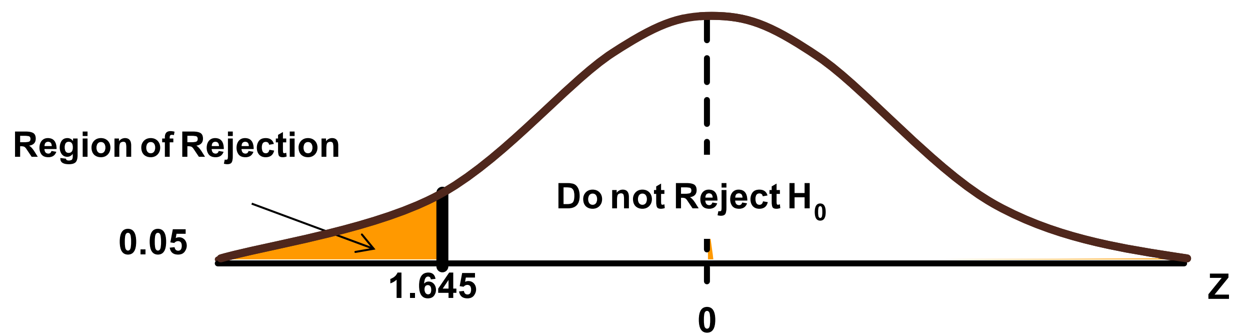
Type II error: when a false null hypothesis is not rejected.

	H_0 is true (No Tsunami)	H_0 is false (Tsunami)
Reject H_0 (Tsunami Warning)	Type I Error	OK
Fail to Reject H_0 (No Warning)	OK	Type II Error

Test of Hypotheses

One-tailed Test $H_0: \mu \geq ?$ or $\mu \leq ?$

There is one tail for region of rejecting null hypothesis – whether left or right.



Two-tailed Test $H_0: \mu = ?$

There are two tails for region of rejecting null hypothesis – both left and right.