Random Variable

- A variable that has a single numerical value, determined by chance, for each outcome of a procedure
  - The number of eggs a hen lays in a day
  - The amount of milk a cow produces in a day

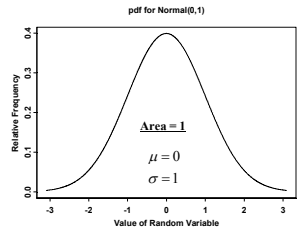
Probability Distribution

- A description that gives the probability for each value of the random variable
- Usually expressed as a graph, table, or formula

Requirements for a Probability Distribution

1. $\sum_{\text{all } x} P(x) = 1$
2. For every individual value of x

   $0 \le P(X \le x) \le 1$



Density Curve—graph of a continuous probability distribution

- Total area under the curve must equal 1
- Every point on the curve must have a vertical height that is 0 or greater

Statistical Inference

- Will be making statements about means or proportions
- Will take into account variability among samples
  - The value of a statistic depends on the particular values in the sample. Values generally vary from sample to sample
  - Variability of a statistic is called **sampling variability**

Central Limit Theorem

If

- The random variable has a distribution (any distribution) with mean μ and standard deviation σ
- Simple random samples all of the same size n are selected from the population

Then

- The distribution of sample means will approach a normal distribution
- The mean of all sample means is the population mean μ
- The standard deviation of all sample means is $\sigma/\sqrt{n}$

Can be extended to proportions

Confidence Intervals

Point Estimator —is a single value or point used to approximate a population

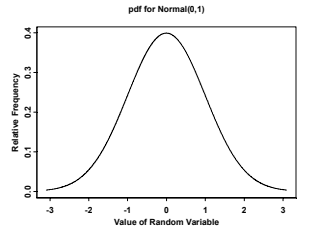
Confidence Interval (CI) —is a range of values used to estimate the true value of a population parameter

Confidence Level — represents the proportion of times that the confidence interval actually does contain the population parameter assuming that the estimation process is repeated a large number of times

Common choices for confidence levels are 90%, 95%, and 99%

Form of a Confidence Interval

point estimator ± (critical value × $SE_{\text{point estimator}}$)



Interpreting a CI

Correct:

We are 95% confident that the interval actually does contain the true value of the population parameter. This means that if we collected many different samples of size *n* and constructed the corresponding CIs, 95% of them would actually contain the value of the population parameter.

Interpreting a CI

Wrong:

There is a 95% chance (probability) that the true population parameter will fall between (lower bound, upper bound)—at any specific point in time a population has a fixed and constant value of the population parameter and a confidence interval constructed from a sample either includes the parameter value or does not.

Increasing Confidence

- Choose a higher confidence level –for a fixed sample size, this will effectively increase the width of the interval
- Increase the sample size –for a fixed confidence level, a larger sample will produce a narrower interval