Probability

- Develop an understanding of **probability** as the relative frequency of occurrence of an event over a very large number of observations or repetitions of the phenomenon.
- Understand the basic event relations and **probability laws**.
- Understand what is meant by **independent events**.

Relative Frequency and Probability

Relative frequency concept of probability:

If an experiment (observation/measurement) is repeated a large number of times and event **E** occurs 30% of the time, then .30 should be a very good approximation to the probability of event **E**.

Event:

A collection of possible outcomes of an experiment.

Outcome:

Each possible distinct result of a simple experiment.

Experiment	Outcome	Event
Coin toss	Head or Tail	Head
5 Coin Tosses	HHHHH or HTTHT …	At Least three heads
Select and weigh an individual	a Weight	weight < 50kg

Probability = relative frequency

If there are 35 white balls in a box of 100 balls, the probability that one ball selected at random from the box is ball is 0.35.

This could be determined by the following.

1. First number the balls from 1 to 100. Set the total to 0.
2. Use a random number table to select a number (hence a ball) at random between 1 and 100.
3. Determine the ball's color. Add a +1 to our total if the ball is white, 0 otherwise.
4. Repeat steps 2 and 3 at least 10,000 times, each time selecting from all 100 balls (simple random selection with replacement).
5. Divide the total observed by the total number of iterations (10,000) and that fraction would be the estimated probability of selecting a white ball at random.

Estimating Probabilities

Basic Experiment: Measure an individuals Height and Weight.

Weight and Height are random variables – their values vary from individual to individual. We use the symbols, W and H to represent these random variables.

Estimate the following Probabilities.

$P(W > 60\text{kg}) = 12/20 = 0.60$
$P(50 < W < 60) = 5/20 = 0.25$
$P(H < 1.8\text{m}) = 5/20 = 0.25$

$P(W > 60\text{kg and } H < 1.8\text{m}) = 2/20 = 0.1$

Can probability ever be greater than 1.0?
Can probability ever be less than 0.0?

Weight	Height
43.5	1.76
45.2	1.90
48.4	1.86
51.8	1.83
53.0	1.61
55.2	1.53
57.2	1.81
59.3	1.90
61.0	1.90
61.4	1.85
63.4	1.98
65.2	1.53
65.6	1.96
67.8	1.86
68.0	1.75
68.3	1.85
68.5	1.81
76.2	1.82
76.3	1.87
84.7	1.88

Probability Limits

Since probability is computed from relative frequencies, it follows that the probability of an event must be between zero and one.

$$0 \leq P(A) \leq 1$$

If all balls are white, the probability of selecting a white ball at random from the box is 1, that is, you will get a white ball 100% of the time.
Similarly, the probability of selecting a brown ball is 0, that is you will get a brown ball 0% of the time.

Mutually Exclusive Events

The occurrence of one of the events excludes the possibility of the occurrence of the other event.

Basic Experiment: 5 tosses of a fair coin.

Event A: At least three heads in 5 tosses of an unbiased coin.
Event B: At least three tails in 5 tosses of an unbiased coin.

Event A implies
3,4 or 5 heads which implies 2,1 or 0 tails.

Event B implies
3,4 or 5 tails which implies 2,1 or 0 heads.

Hence, if Event A occurs (e.g. we get 3 heads) then Event B cannot occur (e.g. we cannot get 3 tails as well).

Events A and B are mutually exclusive.

Mutually Exclusive Events (cont)

Basic Experiment: Measure an individuals weight and height.

Event A: An observed weight greater than 60 kg. (W>60)
Event B: An observed weight greater than 50 kg. (W>50)

A and B are not mutually exclusive. We could observe a weight of 61kg which would satisfy both events.

Event C: An observed weight less than 50 kg. (W<50)
Event D: An observed weight greater than 60 kg. (W>60)

C and D are mutually exclusive. If we observe a weight of less than 50kg we cannot simultaneously observe a weight of say 61kg.

If two events C and D are mutually exclusive, then the probability that either event occurs = P(C or D) = P(C) + P(D).

Mutually Exclusive Probabilities

P(W<50)=3/20 = 0.15,
P(W>60)=12/20 = 0.60,

P(C or D) = P(C) + P(D)

then P(W<50 or W>60) = 0.15 + 0.60 = .75
=(3+12)/20

Weight	Height
43.5	1.76
45.2	1.90
48.4	1.86
51.8	1.83
53.0	1.61
55.2	1.53
57.2	1.81
59.3	1.90
61.0	1.90
61.4	1.85
63.4	1.98
65.2	1.53
65.6	1.96
67.8	1.86
68.0	1.75
68.3	1.85
68.5	1.81
76.2	1.82
76.3	1.87
84.7	1.88

Complementary Events

Basic Experiment: Measure someone's weight

Event A: An observed weight of less than 60 kg. ($W<60$)
Event B: An observed weight of at least 60 kg. ($W\geq60$)

Events A and B are mutually exclusive. But, more than that, the two events are **complementary** in that the outcome of the experiment must fall in one or the other of the two events. There are no other options.

$P(W<60)=0.40$,
$P(W \geq 60)=0.60$,
$P(W< 60 \text{ or } W \geq 60) = 0.40 + 0.60 = 1.00$

Computing Probabilities

Basic Experiment: Toss a coin 5 times.

TTTTT	HTTTT
TTTTH	HTTTH
TTTHT	HTTHT
TTTHH	HTTHH
TTHTT	HTHTT
TTHTH	HTHTH
TTHHT	HTHHT
TTHHH	HTHHH
THTTT	HHTTT
THTTH	HHTTH
THTHT	HHTHT
THTHH	HHTHH
THHTT	HHHTT
THHTH	HHHTH
THHHT	HHHHT
THHHH	HHHHH

32 possible outcomes to 5 coin toss experiment.

Event A: At least 3 Heads

P(3-5H in 5 tosses) =16/32

Event B: At least 2 Tails

P(2-5T in 5 tosses) =26/32

Not Mutually Exclusive

Event C: 1 or 2 T

Event D: 1 or 2 H

Mutually Exclusive

Event C: 3 or 4 H

Event D: 3 or 4 T

P(C) =(5 +10)/32 = 15/32

P(D) =(5 +10)/32 = 15/32

P(C or D)= 30/32

Count them to make sure.

Some Probability Properties

If two events, A and B are mutually exclusive, then P(A) and P(B) must satisfy the following properties:

$$0 \le P(A) \le 1 \quad \text{and} \quad 0 \le P(B) \le 1$$

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

$$P(A) + P(\overline{A}) = 1 \quad \text{and} \quad P(B) + P(\overline{B}) = 1$$

The last line also holds for any events A and B, not necessarily mutually exclusive, where:

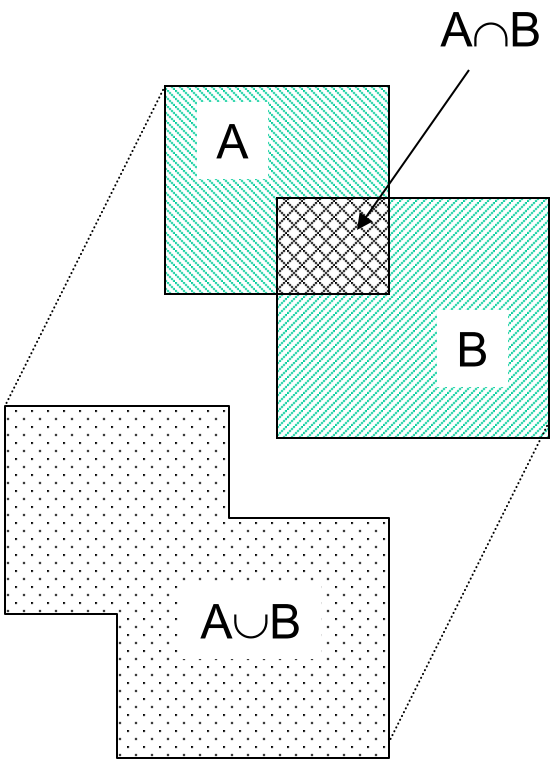
$\overline{A}$ Complement of A

$\overline{B}$ Complement of B

Union and Intersection of Events

The **INTERSECTION** of two events A and B is the set of all outcomes that are included in both A and B, and is denoted as $A \cap B$. (read as A **and** B)

The **UNION** of two events A and B is the set of all outcomes that are included in either A or B (or both) and is denoted as $A \cup B$). (read as A **or** B)



General Rule:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Important to remember

Add the probabilities then subtract the overlap so that we don't double count.

Union and Intersection Example

Basic Experiment: Measure an individuals Height and Weight.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

What is the probability of 60<W<70 or H>1.8)

P(60<W<70 $\cup$ H>1.8)
=P(60<W<70) + P(H>1.8) –P(60<W<70 $\cap$ H>1.8)
= 9/20 + 15/20 – 7/20 = (9+15-7)/20 = 17/20 = 0.85

$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$

What is the probability of W>60 and W<70)

P(W>60 $\cap$ W<70) = P(W>60) + P(W<70) –P(W>60 $\cup$ W<70)
= 12/20 + 17/20 – 20/20 = (12+17-20)/20 = 9/20 = 0.45

Weight	Height
43.5	1.76
45.2	1.90
48.4	1.86
51.8	1.83
53.0	1.61
55.2	1.53
57.2	1.81
59.3	1.90
61.0	1.90
61.4	1.85
63.4	1.98
65.2	1.53
65.6	1.96
67.8	1.86
68.0	1.75
68.3	1.85
68.5	1.81
76.2	1.82
76.3	1.87
84.7	1.88

Probability Algebra

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cup B) + P(A \cap B) = P(A) + P(B)$

$P(A \cap B) = P(A) + P(B) - P(A \cup B)$

Probabilities are like any algebraic symbols, we can add, subtract, multiply and divide them.

$1 = [P(A) + P(B) - P(A \cup B)] / P(A \cap B)$ {assuming $P(A \cap B) \neq 0$}

If event A and B are **complementary** they have no overlap,
hence $P(A \cap B) = 0$, and
$P(A \cup B) = P(A) + P(B)$, also,
since $A \cup B$ involves all possible events, $P(A \cup B)=1$,
thus $1=P(A) + P(B)$ for complementary event, or
$P(A) = 1-P(B)$

What if A and B are mutually exclusive?

Conditional Probability

Consider two events A and B with nonzero probabilities, P(A) and P(B).

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

The conditional probability of event A given event B.

Note: Event B must have non-zero probability.

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

The conditional probability of event B given event A.

Note: Event A must have non-zero probability.

When we compute a conditional probability we are essentially asking for the probability of an event under the constraint/knowledge that some second event has occurred.

Ex: P(W>50 given that H>1.8) - That is, what is the probability of finding someone with weight greater than 50kg among individuals who are greater than 1.8m tall?

Conditional Probability Example

Basic Experiment: Measure the height and weight of a sample.

Event A: Weight(X) is greater than 50 kg.
Event B: Height(Y) is greater than 1.8 m.

P(X>50 and Y>1.8) = 13/20
P(Y > 1.8) = 15/20

What is the probability of observing a weight greater than 50kg among individuals greater than 1.8m?

$$P(X > 50 \mid Y > 1.8) = \frac{P(X > 50 \cap Y > 1.8)}{P(Y > 1.8)}$$

=(13/20)/(15/20) = 13/15

Weight	Height
43.5	1.76
45.2	1.90
48.4	1.86
51.8	1.83
53.0	1.61
55.2	1.53
57.2	1.81
59.3	1.90
61.0	1.90
61.4	1.85
63.4	1.98
65.2	1.53
65.6	1.96
67.8	1.86
68.0	1.75
68.3	1.85
68.5	1.81
76.2	1.82
76.3	1.87
84.7	1.88

Intersection Probability using the Conditional Probability

The probability that two events occur together.

$$P(A \cap B) = P(A)P(B \mid A)$$

$$= P(B)P(A \mid B)$$

P(X>50 | Y>1.8) = 13/15
P(Y>1.8) = 15/20
P(X>50) = 17/20
P(X>50 and Y>1.8) = P(Y>18)P(X>50 | Y>1.8)
=(15/20)(13/15) = 13/20

Sometimes we know the conditional probability.

e.g. P(Lung Cancer and Smoking) =
P(Smoking)P(Lung Cancer | Smoking)

From general population surveys.

From a retrospective study of smokers.

Weight	Height
43.5	1.76
45.2	1.90
48.4	1.86
51.8	1.83
53.0	1.61
55.2	1.53
57.2	1.81
59.3	1.90
61.0	1.90
61.4	1.85
63.4	1.98
65.2	1.53
65.6	1.96
67.8	1.86
68.0	1.75
68.3	1.85
68.5	1.81
76.2	1.82
76.3	1.87
84.7	1.88

Nesting and Recruitment of Birds Example

Basic Experiment: Observe 1000 wading bird nests in the Everglades. Record number of nests in which eggs were laid, when eggs were laid and whether hatched chicks survived.

Event A = Nest is successful (chick survives).

Event B = Eggs laid in March.

P(A) = fraction of nests that are successful. (300/1000 = .3)

P(B) = fraction of nests with eggs laid in March. (100/1000 = .1)

$P(A \cap B)$ = fraction of nests that are successful and whose eggs were laid in March. (60/1000 = .06)

P(A|B) = Probability of a successful nest given that the eggs in the nest were laid in March

$$\boxed{P(A|B) = P(A \cap B)/P(B) = 0.06/\ 0.10 = 0.60\ (=60/100)}$$

P(B|A) = Probability eggs in the nest were laid in March given that the nest is successful.

$$\boxed{P(B|A) = P(A \cap B)/P(A) = 0.06/\ 0.30 = 0.20\ (=60/300)}$$

Independent Events

Events A and B are said to be *independent* if:

$$P(A \mid B) = P(A), \quad \text{or} \quad P(B \mid A) = P(B), \quad \text{or} \quad P(A \cap B) = P(A)P(B)$$

(Probability of one event does not depend on what happens with the other event.)

Event A: Weight(X) is greater than 50 kg.

Event B: Height(Y) is greater than 1.8 m.

Are A and B independent events?

P(A) =P(W>50)= 17/20 = 0.85

P(A|B) =P(W>50|H>1.8)= 13/15 = 0.8667

Close, but NO; they are dependent!

Weight	Height
43.5	1.76
45.2	1.90
48.4	1.86
51.8	1.83
53.0	1.61
55.2	1.53
57.2	1.81
59.3	1.90
61.0	1.90
61.4	1.85
63.4	1.98
65.2	1.53
65.6	1.96
67.8	1.86
68.0	1.75
68.3	1.85
68.5	1.81
76.2	1.82
76.3	1.87
84.7	1.88

This will become more important later when we develop test for independence.

Other Simple Experiments

- Toss a coin 5 times, count the number of heads.
  Outcome=number of heads.
- Throw six die.
  Outcome=sum of pips facing up.
- Throw two die.
  Outcome=sum of pips facing up.
- Toss a die until a 5 or 6 occurs.
  Outcome=number of tosses needed minus one.
- Toss a “frame” on spatial point patterns.
  Outcome=number of points in the frame.
- Draw a chip at random with replacement from bag.
  Outcome=value written on chip.

Replicate each experiment many times. Make a frequency chart of results. Discuss events and probabilities from distribution.

Probability Distribution

A **probability distribution** (function) is a list of the probabilities of the values (simple outcomes) of a random variable.

Table: Number of heads in two tosses of a coin

Can be envisioned as a table.

y *outcome*	P(y) *probability*
0	1/4
1	2/4
2	1/4

For some experiments, the probability of a simple outcome may be easily calculated mathematically using a specific **probability function.** If y is a simple outcome and p(y) is its probability.

$$0 \le p(y) \le 1$$

$$\sum_{\text{all } y} p(y) = 1$$

The Probability Density Function for Weight

Each of the weight values are unique, hence the probability we would assign to each observation, in the absence of any other information, would simply be 1/n.

Clearly this is not very informative. A *histogram* would provide a better estimate of the true underlying probabilities for the entire population, by binning the values.

A more sophisticated approach would involve *kernel density-based smoothing*, in order to estimate the underlying continuous **probability density function** curve. We will discuss this concept in the next lecture.

Weight	P(Y)
43.5	0.05
45.2	0.05
48.4	0.05
51.8	0.05
53.0	0.05
55.2	0.05
57.2	0.05
59.3	0.05
61.0	0.05
61.4	0.05
63.4	0.05
65.2	0.05
65.6	0.05
67.8	0.05
68.0	0.05
68.3	0.05
68.5	0.05
76.2	0.05
76.3	0.05
84.7	0.05

Distributions of Random Variables

In this Lecture we discuss the different types of random variables and illustrate the properties of typical probability distributions for these random variables.

What is a Random Variable?

A **variable** is any characteristic, observed or measured. A variable can be either **random** or **constant** in the population of interest.

Note this differs from common English usage where the word variable implies something that **varies** from individual to individual.

For a defined population, every **random variable** has an associated distribution that defines the **probability** of occurrence of each possible value of that variable (if there are a finitely countable number of unique values) or all possible sets of possible values (if the variable is defined on the real line).

Probability Distribution

A **probability distribution** (function) is a list of the probabilities of the values (simple outcomes) of a random variable.

Table: Number of heads in two tosses of a coin

y *outcome*	P(y) *probability*
0	1/4
1	2/4
2	1/4

For some experiments, the probability of a simple outcome can be easily calculated using a specific **probability function.** If y is a simple outcome and p(y) is its probability.

$$0 \leq p(y) \leq 1$$

$$\sum_{\text{all } y} p(y) = 1$$

Discrete Distributions

Relative frequency distributions for “counting” experiments.

- Bernoulli Distribution
- Binomial Distribution
- Negative Binomial
- Poisson Distribution
- Geometric Distribution
- Multinomial Distribution

Yes-No responses.

Sums of Bernoulli responses

Number of trials to k^{th} event

Points in given space

Number of trials until first *success*

Multiple possible outcomes for each trial

Binomial Distribution

- The experiment consists of **n identical trials** (simple experiments).
- Each trial results in one of **two outcomes** (success or failure)
- The probability of success on a single trial is equal to π and π remains the same from trial to trial.
- The trials are independent, that is, the outcome of one trial does not influence the outcome of any other trial.
- The random variable y is the number of successes observed during n trials.

$$P(y) = \frac{n!}{y!(n-y)!}\pi^{y}(1-\pi)^{n-y}$$

n!=1x2x3x…x n

$$\mu = n\pi$$

Mean

$$\sigma = \sqrt{n\pi(1-\pi)}$$

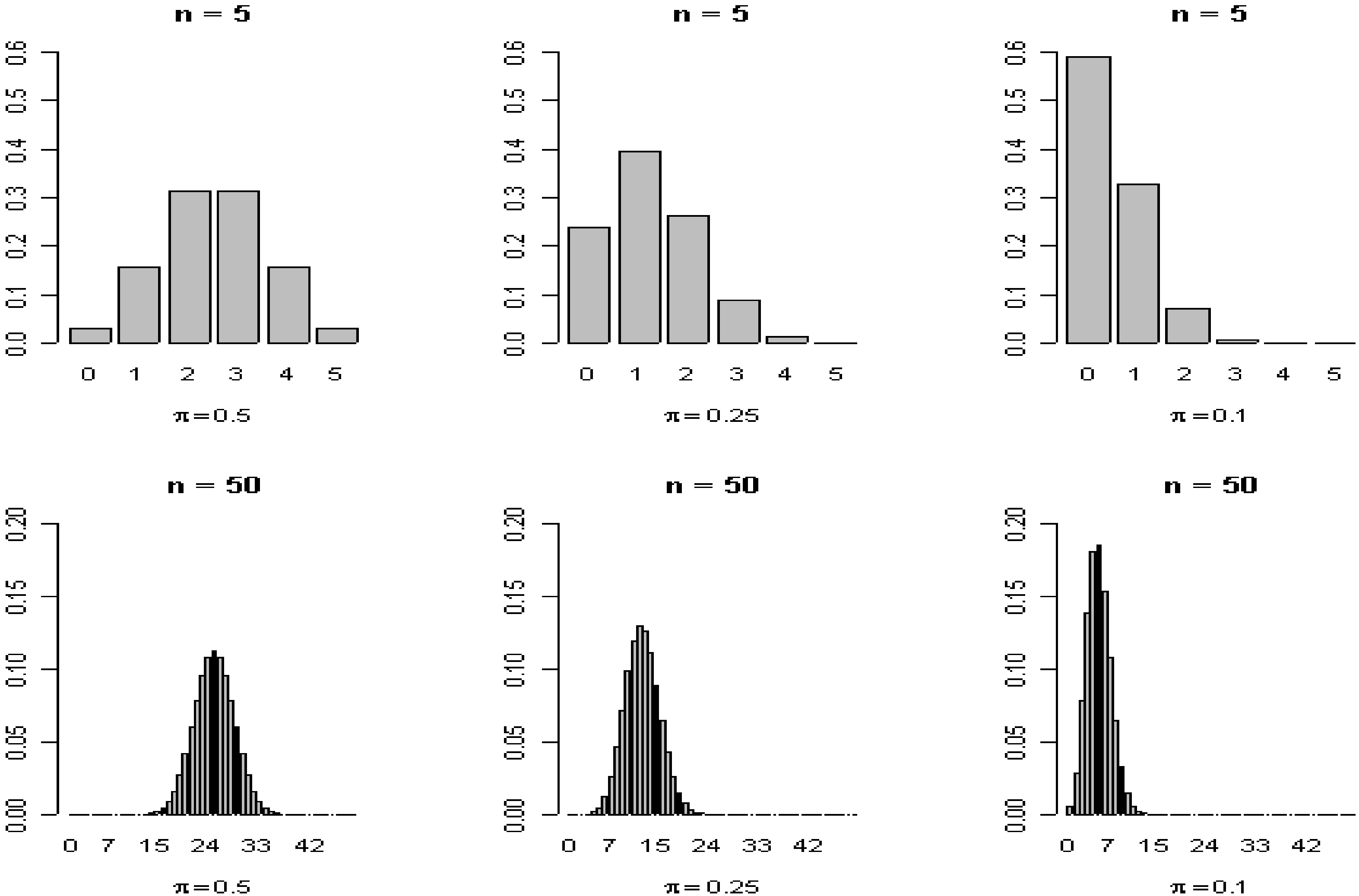
Standard deviation

Example n=5

$$P(y) = \frac{n!}{y!(n-y)!}\pi^{y}(1-\pi)^{n-y}$$

		n	π	π	π	π
		5	0.5	0.25	0.1	0.05
y	y!	n!/(y!)(n-y)!	P(y)	P(y)	P(y)	P(y)
0	1	1	0.03125	0.2373	0.59049	0.7737809
1	1	5	0.15625	0.3955	0.32805	0.2036266
2	2	10	0.31250	0.2637	0.07290	0.0214344
3	6	10	0.31250	0.0879	0.00810	0.0011281
4	24	5	0.15625	0.0146	0.00045	0.0000297
5	120	1	0.03125	0.0010	0.00001	0.0000003
		sum =	1	1	1	1

Binomial probability density function forms



As the n goes up, the distribution looks more symmetric and bell shaped.

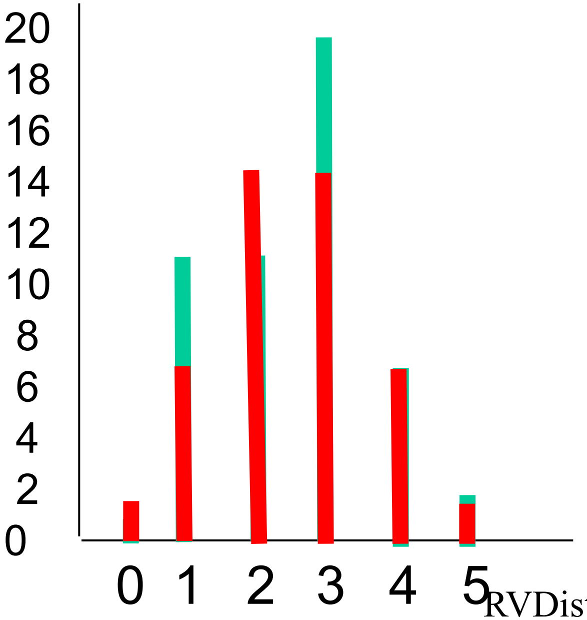
Binomial Distribution Example

Basic Experiment: 5 fair coins are tossed.
Event of interest: total number of heads.

Each coin is a trial with probability of a head coming up (a success) equal to 0.5. So the number of heads in the five coins is a binomial random variable with n=5 and π=.5.

The Experiment is repeated 50 times.

# of heads	*Observed*	*Theoretical*
0	1	1.56
1	11	7.81
2	11	15.63
3	19	15.63
4	6	7.81
5	2	1.56



Two Dice Experiment

Two dice are thrown and the number of “pips” showing are counted (random variable X). The simple experiment is repeated 50 times.

Outcome	Frequency	Relative Freq
2	2	0.04
3	4	0.08
4	4	0.08
5	3	0.06
6	4	0.08
7	7	0.14
8	9	0.18
9	3	0.06
10	5	0.10
11	7	0.14
12	2	0.04
total	50	1.00

Approximate probabilities for the random variable X:

$P(X \leq 8)=$ 33/50 =.66

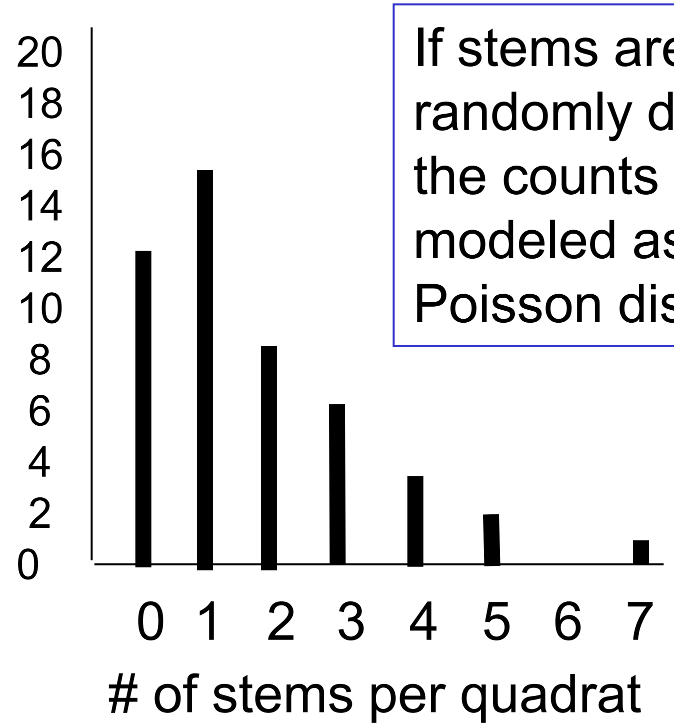
$P(X \geq 6)=$ 37/50 = .74

$P(4 \leq X \leq 10)=$ 34/50 = .68

Vegetation Sampling Data

A typical method for determining the density of vegetation is to use "quadrats", rectangular or circular frames in which the number of plant stems are counted. Suppose 50 "throws" of the frame are used and the distribution of counts reported.

Count (stems)	*Frequency (quadrats)*
0	12
1	15
2	9
3	6
4	4
5	3
6	0
7	1
total	50



If stems are randomly dispersed, the counts could be modeled as a Poisson distribution.

Poisson Distribution

A random variable is said to have a *Poisson Distribution* with rate parameter λ, if its probability function is given by:

e=2.718…

$$P(y) = \frac{\lambda^y}{y!} e^{-\lambda}, \qquad \text{for } y = 0,1,2,\ldots$$

$$\boxed{\mu = \lambda, \quad \sigma^2 = \lambda}$$

← Mean and variance for a Poisson

Ex: A certain type of tree has seedlings randomly dispersed in a large area, with a mean density of approx 5 per sq meter. If a 10 sq meter area is randomly sampled, what is the probability that no such seedlings are found?

$P(0) = 50^0(e^{-50})/0! =$ approx 10^{-22}

(Since this probability is so small, if no seedlings were actually found, we might question the validity of our model…)

Discrete Distributions Take Home Messages

- Primarily related to “counting” experiments.
- Probability only defined for “integer” values.
- Symmetric and non-symmetric distribution shapes.
- Best description is a frequency table.

Examples where discrete distributions are seen.

Wildlife - animal sampling, birds in a 2 km x 2 km area.
Botany - vegetation sampling, quadrats, flowers on stem.
Entomology - bugs on a leaf
Medicine - disease incidence, clinical trials
Engineering - quality control, number of failures in fixed time

Continuous Distributions

- **Normal Distribution**
- Log Normal Distribution
- Gamma Distribution
- **Chi Square Distribution**
- **F Distribution**
- **t Distribution**
- Weibull Distribution
- Extreme Value Distribution (Type I and II)

Foundations for much of statistical inference

Environmental variables

Time to failure, radioactivity

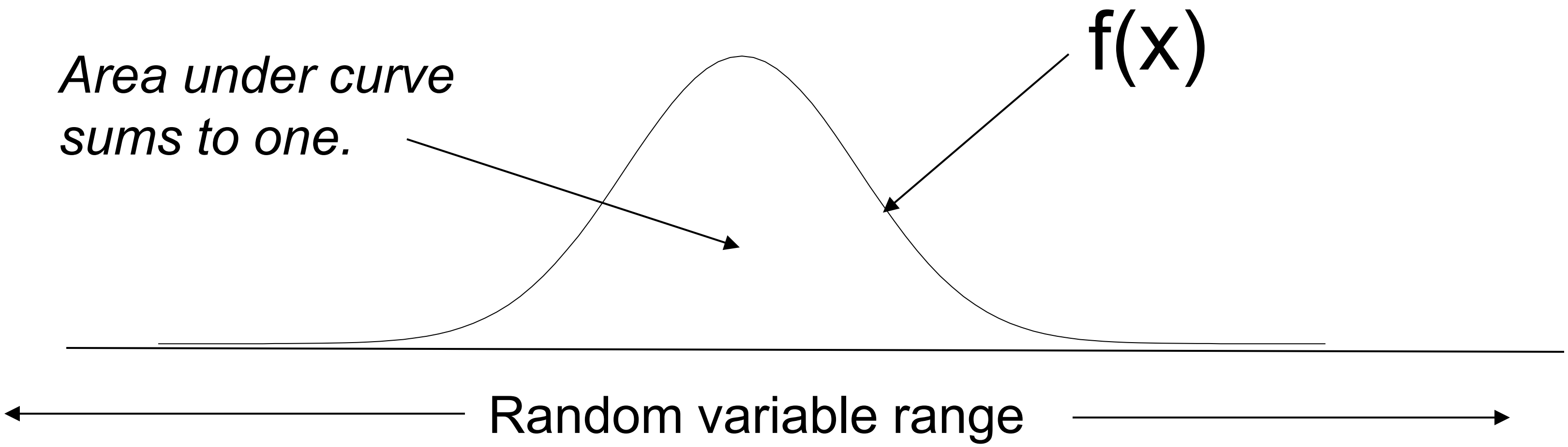
Basis for statistical tests.

Lifetime distributions

Continuous random variables are defined for continuous numbers on the real line. Probabilities have to be computed for all possible sets of numbers.

Probability Density Function

A function which integrates to 1 over its range and from which event probabilities can be determined.



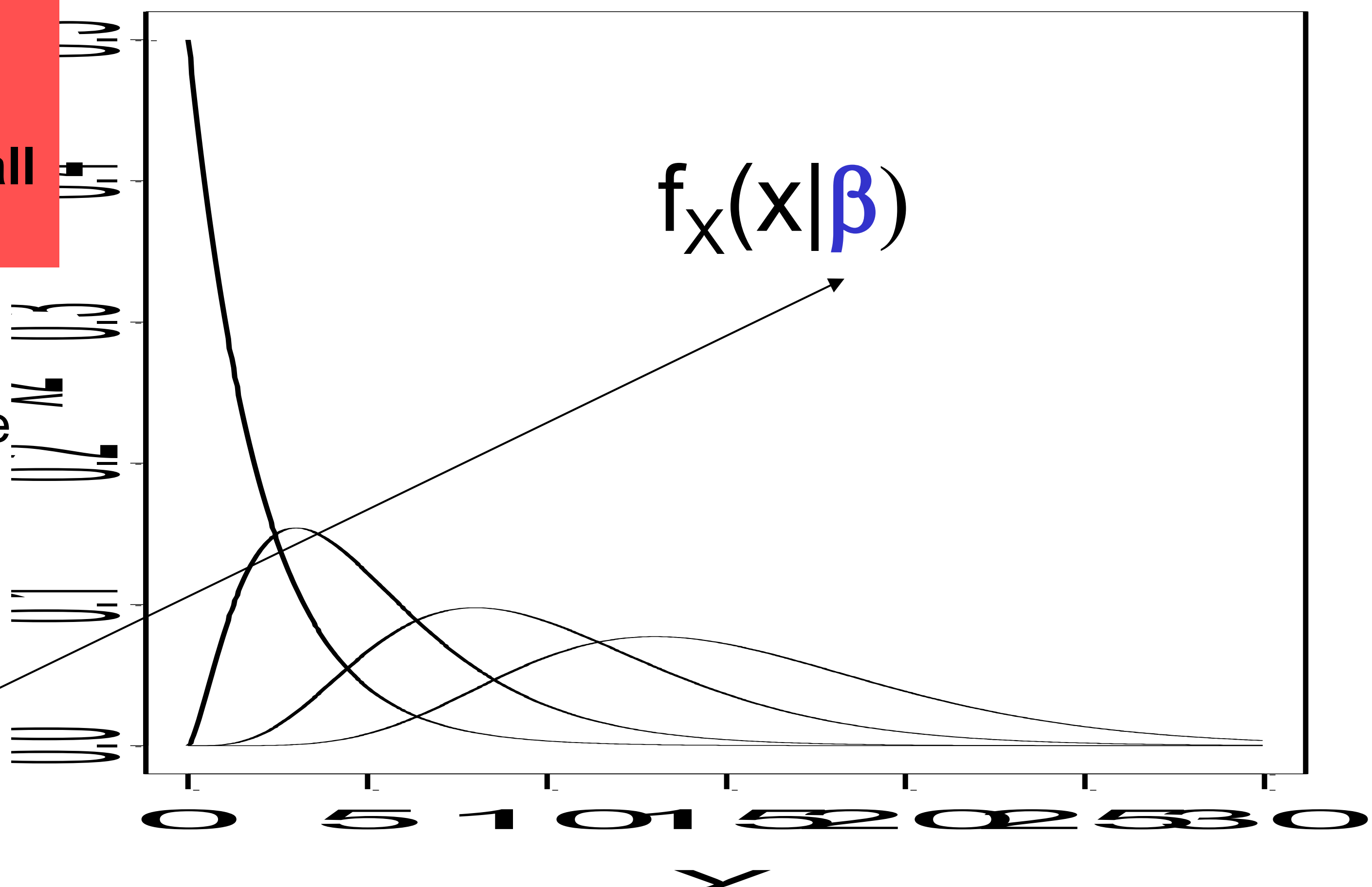
A theoretical shape - if we were able to sample the whole (infinite) population of possible values, this is what the associated histogram would look like.

A mathematical abstraction

Probability Density Function

The pdf does not have to be symmetric, nor be defined for all real numbers.

Chi Square density functions



The shape of the curve is determined by one or more **distribution parameters.**

Continuous Distribution Properties

Probability can be computed by integrating the density function.

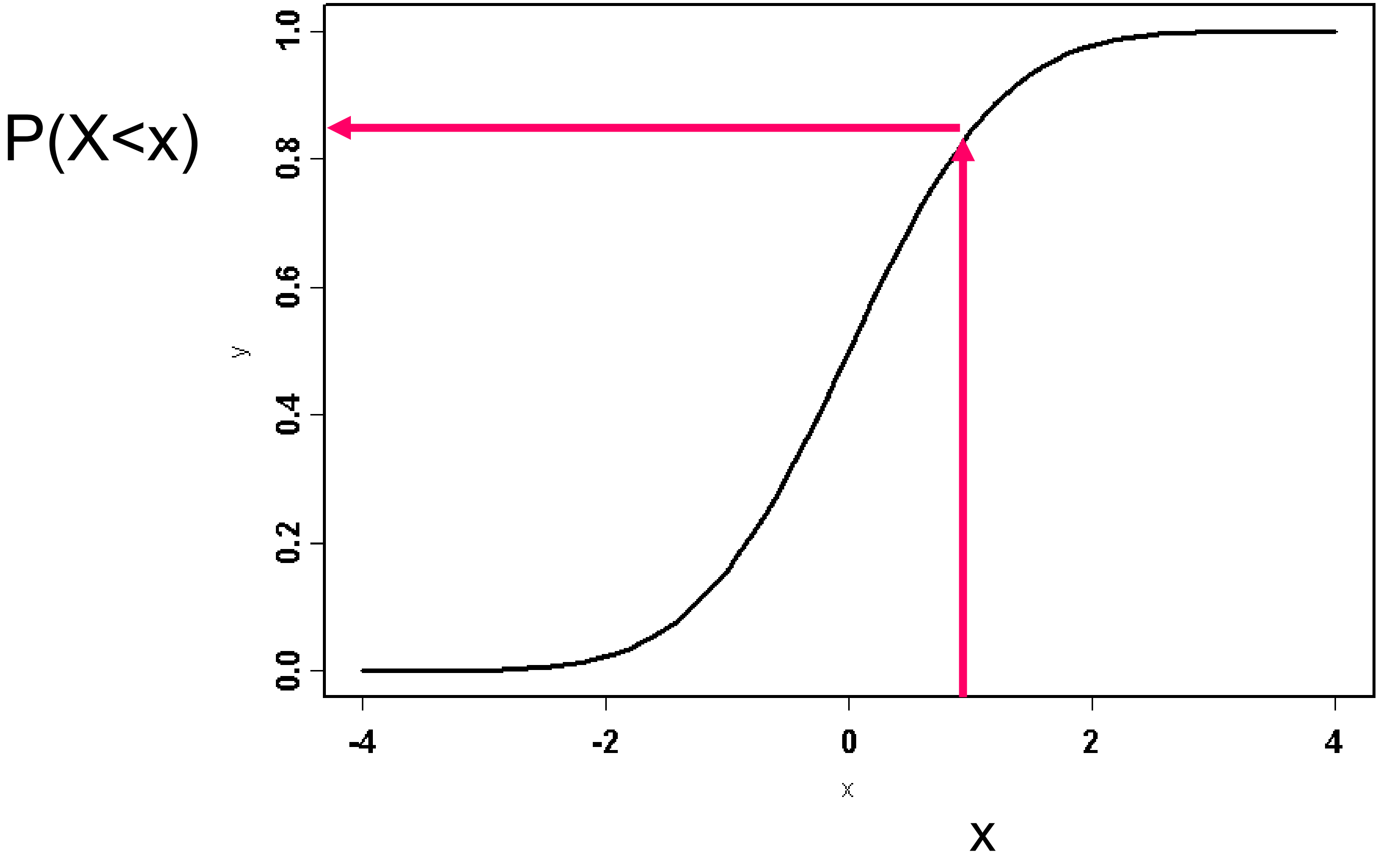
$$F(x_0) = P(X < x_0) = \int_{-\infty}^{x_o} f_X(x)dx$$

Continuous random variables only have positive probability for events which define intervals on the real line.

Any one point has **zero probability of occurrence**.

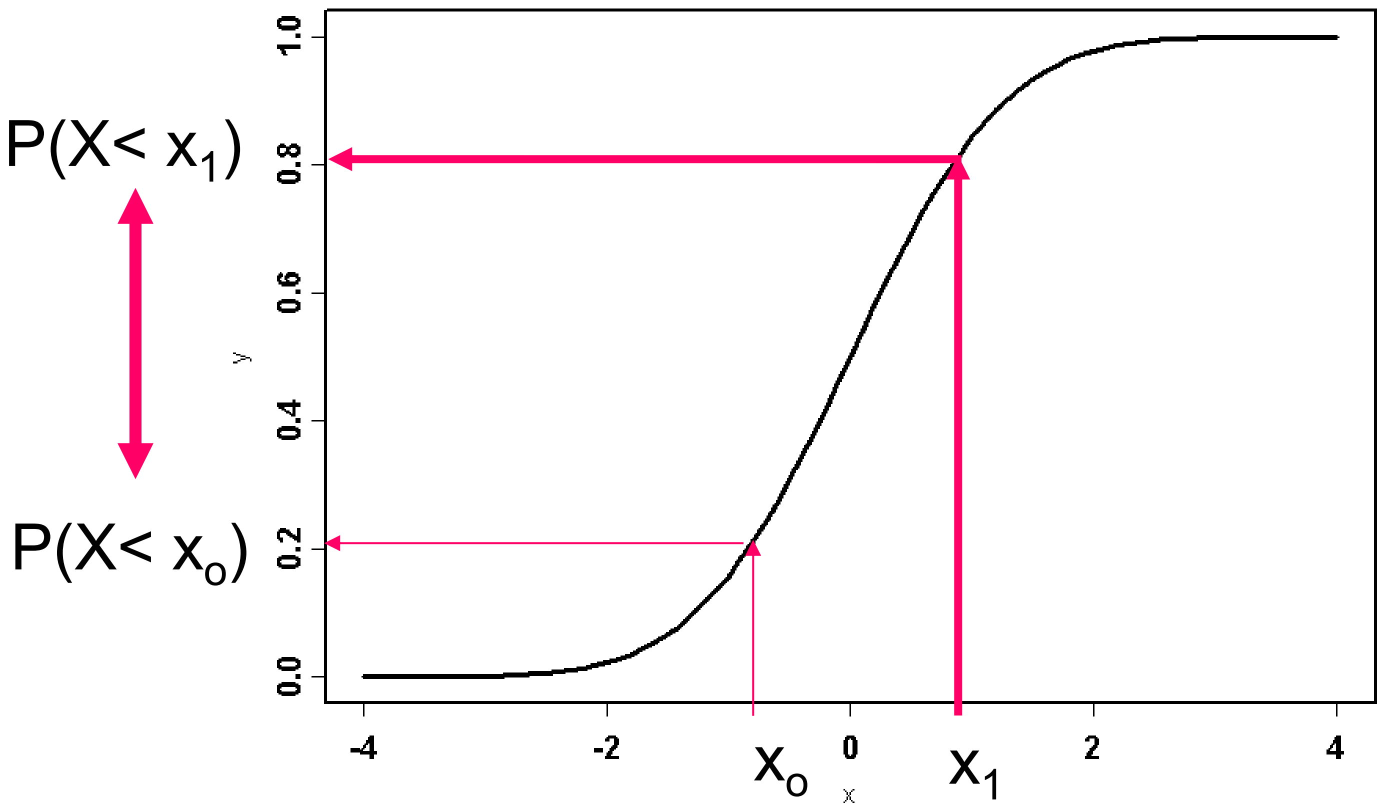
$$P(X = x_0) = \int_{x_0}^{x_0} f_X(x)dx = 0$$

Cumulative Distribution Function



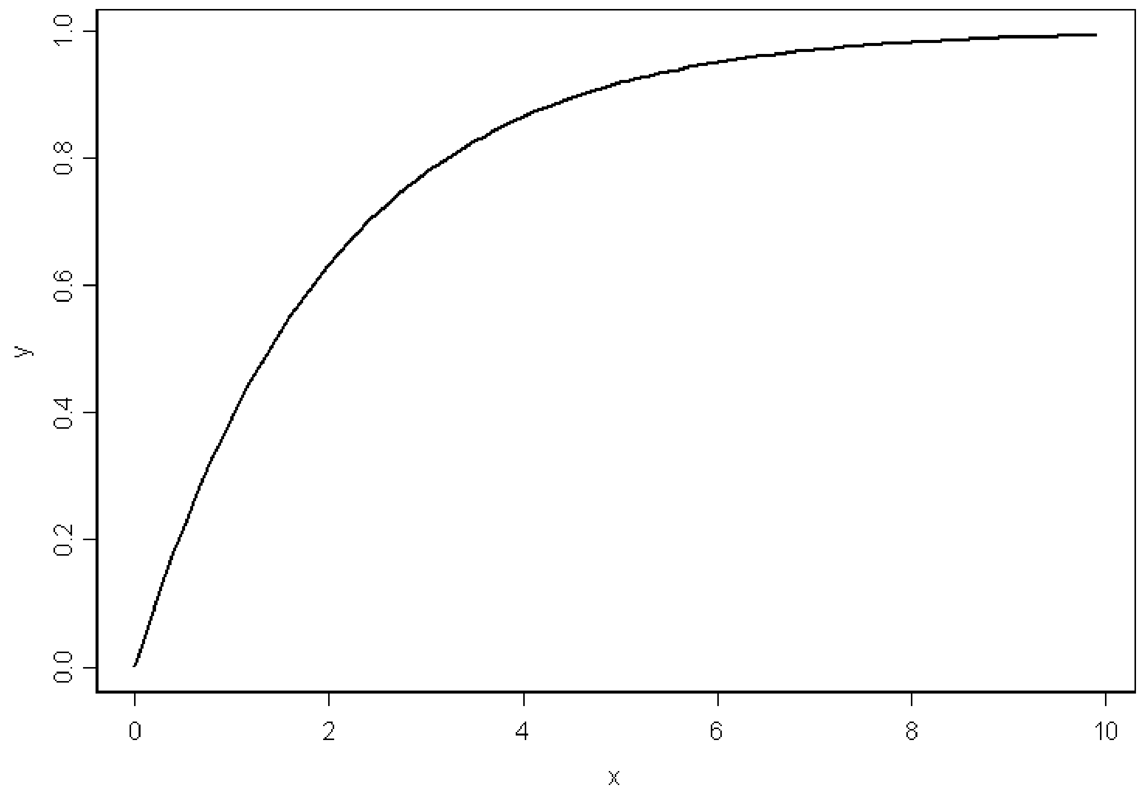
Using the Cumulative Distribution

$P(x_o < X < x_1) = P(X < x_1) - P(X < x_o) = .8 - .2 = .6$



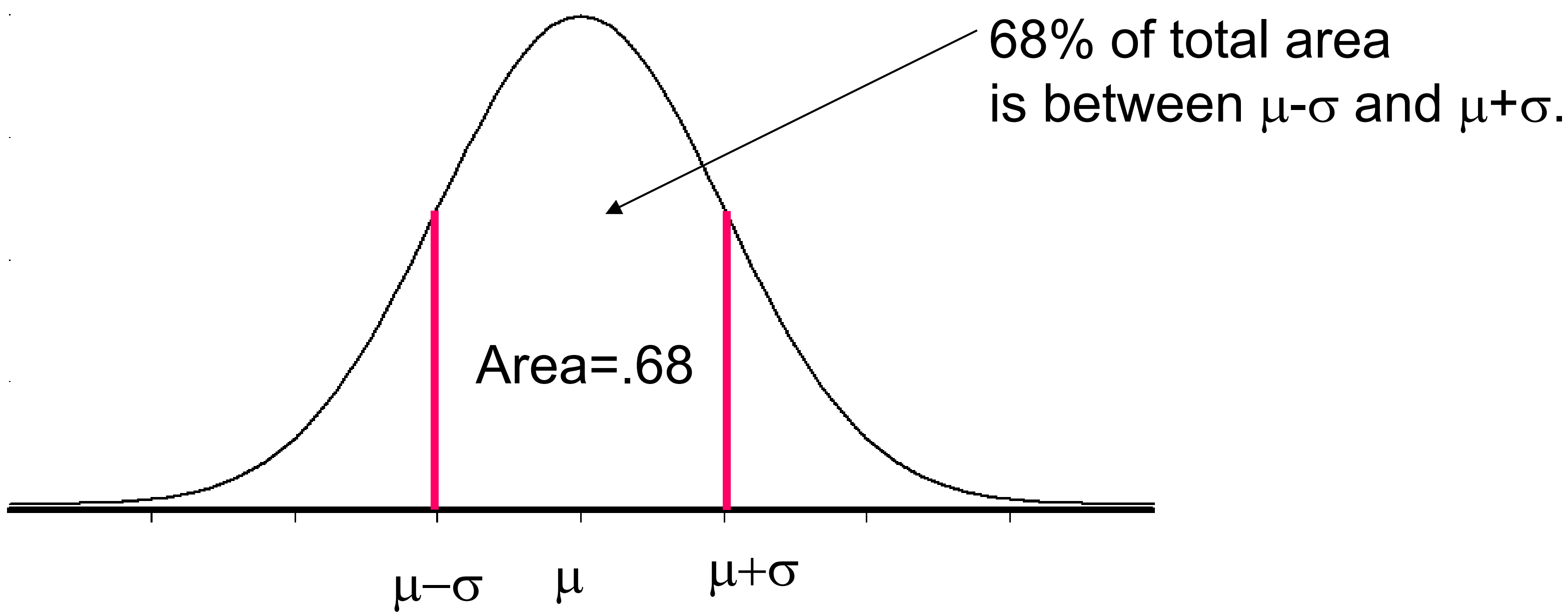
Chi Square Cumulative Distribution

Cumulative distribution does not have to be S shaped. In fact, only the normal and t-distributions have S shaped distributions.



Normal Distribution

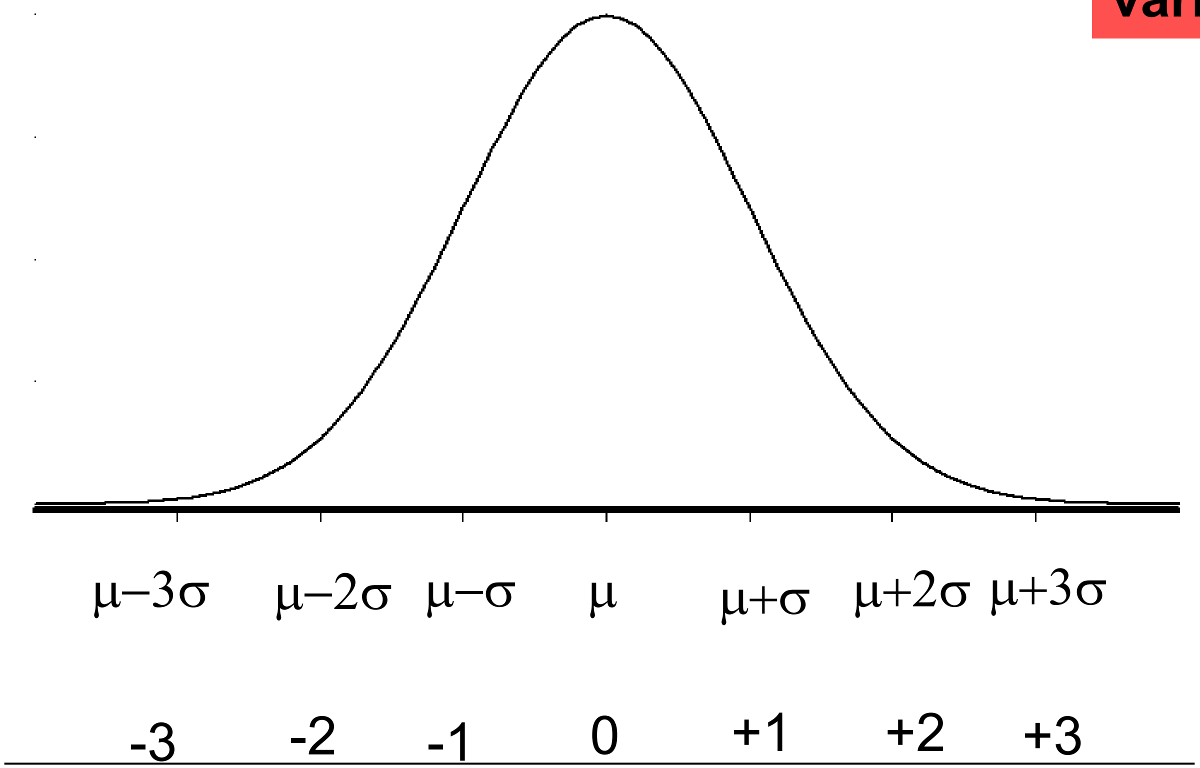
A symmetric distribution defined on the range -∞ to + ∞ whose shape is defined by two parameters, the **mean**, denoted μ, that centers the distribution, and the **standard deviation**, σ, that determines the spread of the distribution.



$$P(\mu - \sigma < X < \mu + \sigma) = .68$$

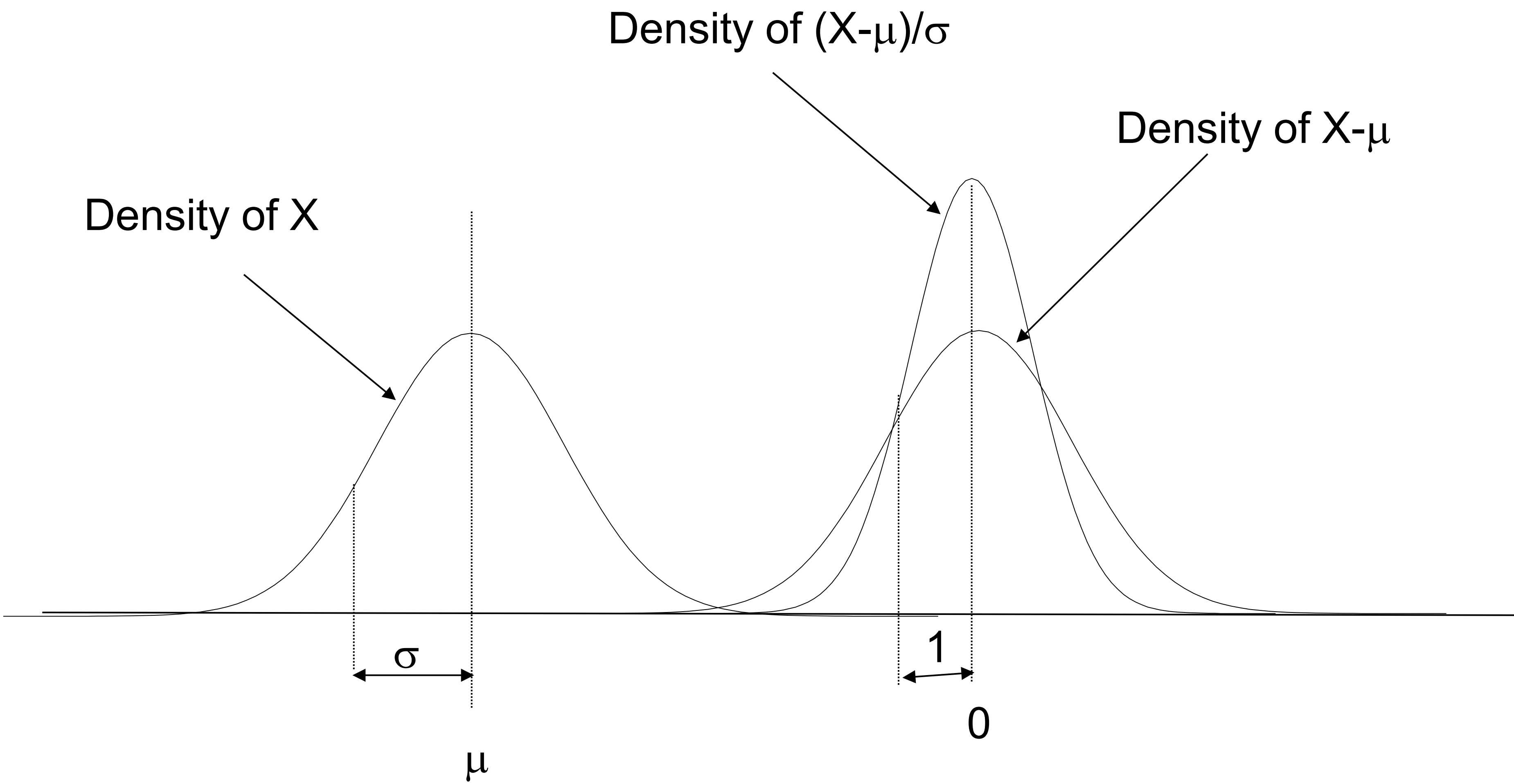
Standard Normal Distribution

All normal random variables can be related back to the **standard normal random variable**.



A Standard Normal random variable has mean 0 and standard deviation 1.

Illustration



Notation

Suppose **X** has a normal distribution with mean μ and standard deviation σ, denoted $\mathbf{X \sim N(\mu, \sigma)}$.

Then a new random variable defined as $\mathbf{Z=(X-\mu)/\sigma}$, has the standard normal distribution, denoted $\mathbf{Z \sim N(0,1)}$.

$\mathbf{Z=(X-\mu)/\sigma}$ — To **standardize** we subtract the mean and divide by the standard deviation.

$\mathbf{\sigma Z+\mu = X}$ — To create a random variable with specific mean and standard deviation, we start with a standard normal deviate, multiply it by the target standard deviation, and then add the target mean.

Why is this important? Because in this way, the probability of any event on a normal random variable with any given mean and standard deviation can be computed from tables of the standard normal distribution.

Relating Any Normal RV to a Standard Normal RV

$X \sim N(\mu, \sigma)$ $\qquad$ $Z \sim N(0,1)$

$$X = \sigma Z + \mu$$

$$\begin{aligned} P(X < x_0) &= P(\sigma Z + \mu < x_0) \\ & P(\sigma Z < x_0 - \mu) \\ & P\left(Z < \boxed{\frac{x_0 - \mu}{\sigma}}\right) \end{aligned}$$

This probability can be found in a table of standard normal probabilities (Table 1 in Ott and Longnecker)

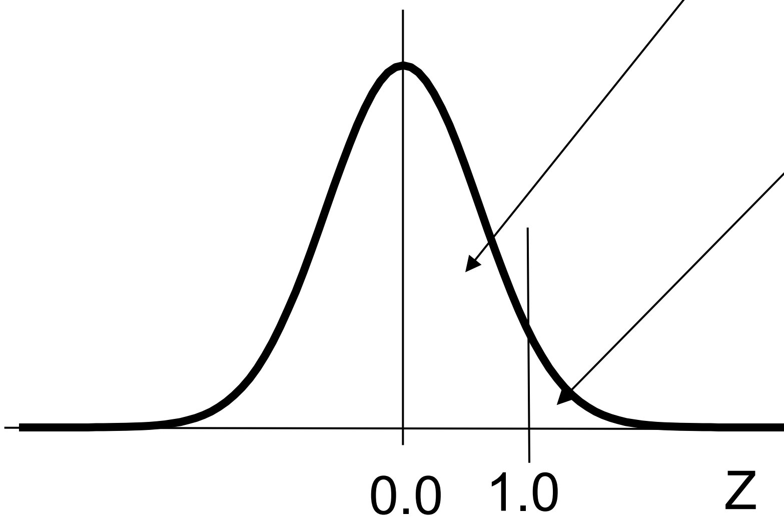
This value is just a number, usually between ±4

Other Useful Relationships

$P(Z < z_0) = 1 - P(Z > z_0)$ Probability of complementary events.

$P(Z < -z_0) = P(Z > +z_0)$ Symmetry of the normal distribution.

Normal Table



Ott & Longnecker, Table 1 page 676, gives areas left of z. This table from a previous edition gives areas right of z.

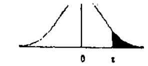


TABLE 2 Upper-tail Areas for the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.00	$\boxed{.5000}$	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.10	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.20	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.30	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.40	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.50	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.60	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.70	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.80	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.90	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.00	$\boxed{.1587}$	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.10	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.20	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.30	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.40	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.50	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.60	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.70	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.80	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.90	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.00	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.10	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.20	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.30	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.40	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.50	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.60	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.70	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.80	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.90	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.00	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

z	Area
3.500	.00023263
4.000	.00003167
4.500	.00000340
5.000	.00000029

Source: Computed by J. W. Stegeman using SAS.

Page A-3 OTT 1998
An Introduction To
STATISTICAL methods
and Data Analysis

Find P(2 < X < 4) when X ~ N(5,2).
The standarization equation for X is:

$$Z = (X-\mu)/\sigma = (X-5)/2$$

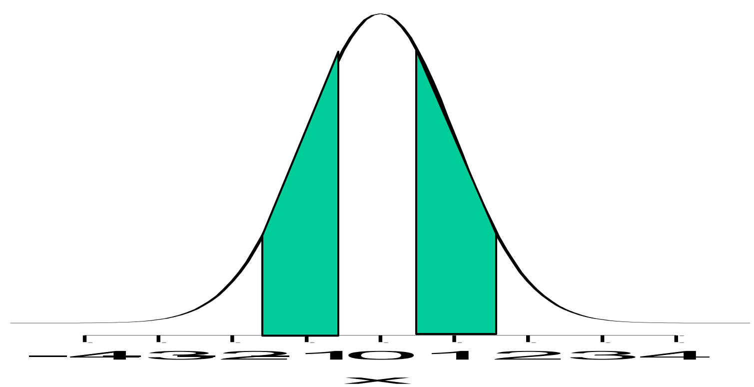
when X=2, Z= -3/2 = -1.5
when X=4, Z= -1/2 = -0.5

P(2<X<4) = P(X<4) - P(X<2)

P(X<2) = P(Z< -1.5)
= P(Z > 1.5) (by symmetry)

P(X<4) = P(Z < -0.5)
= P(Z > 0.5) (by symmetry)

P(2 < x < 4) = P(X<4)-P(X<2)
= P(Z>0.5) - P(Z > 1.5)
= 0.3085 - 0.0668 = 0.2417



Using a Normal Table

TABLE 2 Upper-tail Areas for the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.00	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.10	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.20	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.30	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.40	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.50	$\boxed{.3085}$	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.60	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.70	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.80	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.90	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.00	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.10	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.20	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.30	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.40	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.50	$\boxed{.0668}$	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.60	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.70	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.80	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.90	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.00	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.10	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.20	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.30	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.40	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.50	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.60	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.70	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.80	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.90	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.00	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

z	Area
3.500	.00023263
4.000	.00003167
4.500	.00000340
5.000	.00000029

Source: Computed by J. W. Stegeman using SAS.

Page A-3 OTT 1998
An Introduction To
STATISTICAL methods
and Data Analysis

Properties of the Normal Distribution

- Symmetric, bell-shaped density function.
- 68% of area under the curve between $\mu \pm \sigma$.
- 95% of area under the curve between $\mu \pm 2\sigma$.
- 99.7% of area under the curve between $\mu \pm 3\sigma$.

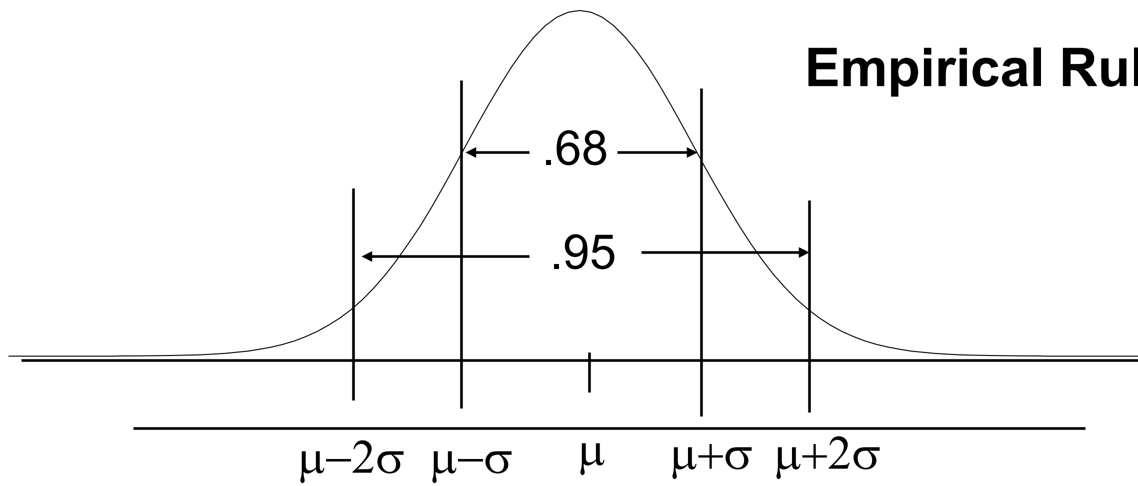
$$Y \sim N(\mu, \sigma)$$

$$Y - \mu \sim N(0, \sigma)$$

$$\frac{Y - \mu}{\sigma} \sim N(0,1)$$

Standard Normal Form

Empirical Rule



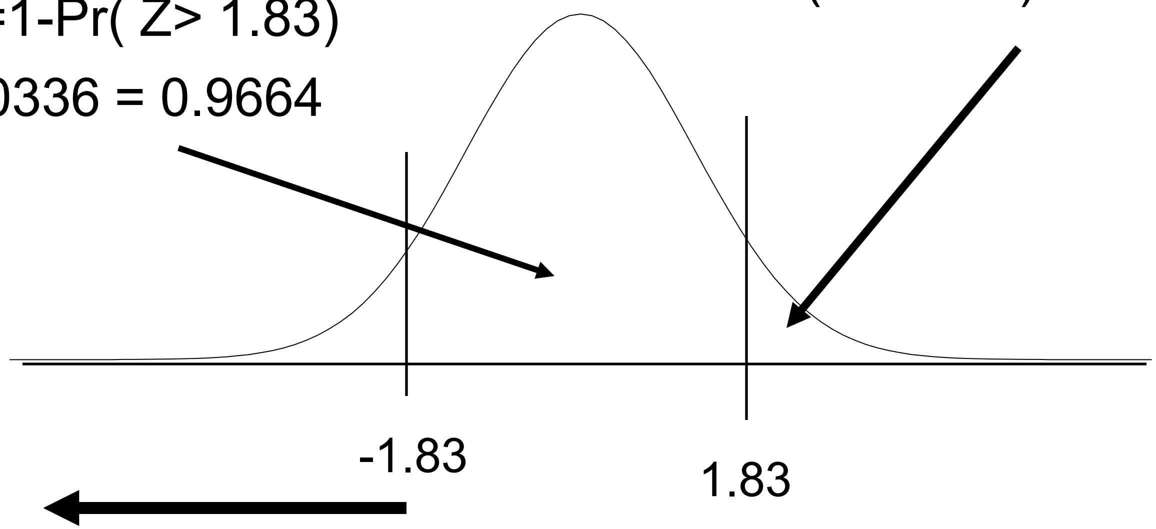
Probability Problems

Using symmetry and the fact that the area under the density curve is 1.

Pr(Z > 1.83) = 0.0336

Pr(Z < 1.83)=1-Pr(Z> 1.83)

=1-0.0336 = 0.9664



Pr(Z < -1.83) = Pr(Z> 1.83)

=0.0336

By Symmetry

Probability Problems

Cutting out the tails.

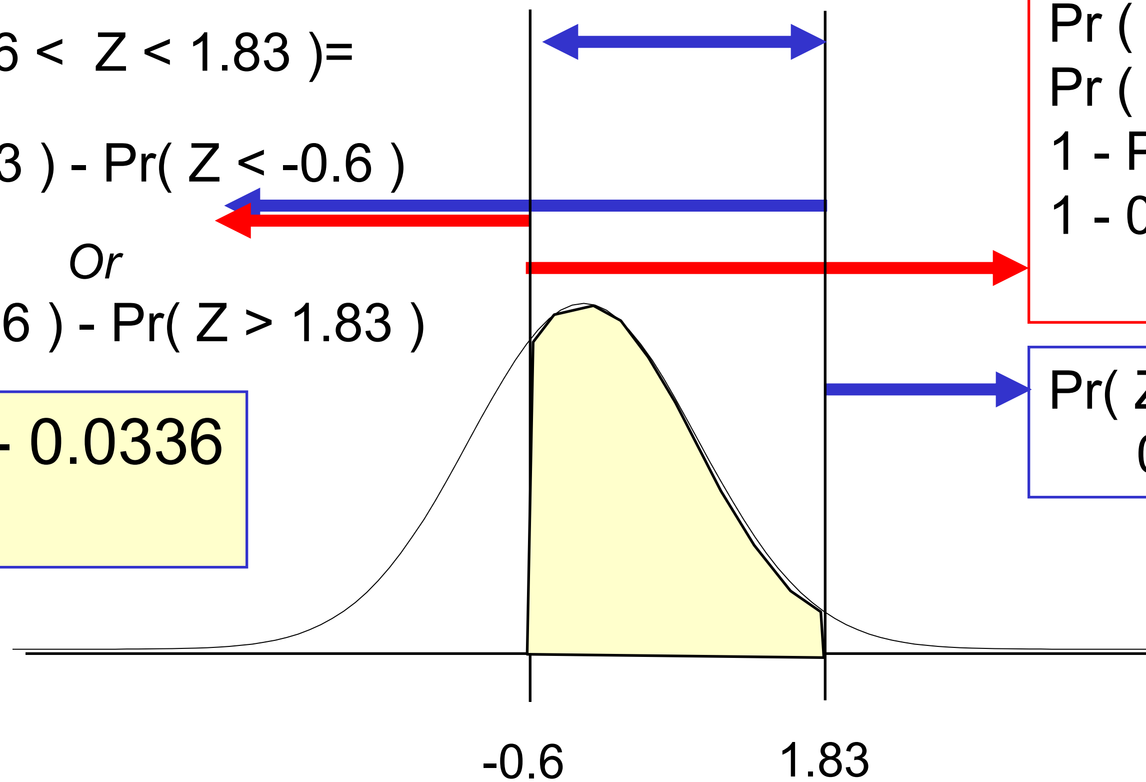
Pr(-0.6 < Z < 1.83)=

Pr(Z < 1.83) - Pr(Z < -0.6)

Or

Pr(Z > -0.6) - Pr(Z > 1.83)

= 0.7257 - 0.0336
= 0.6921

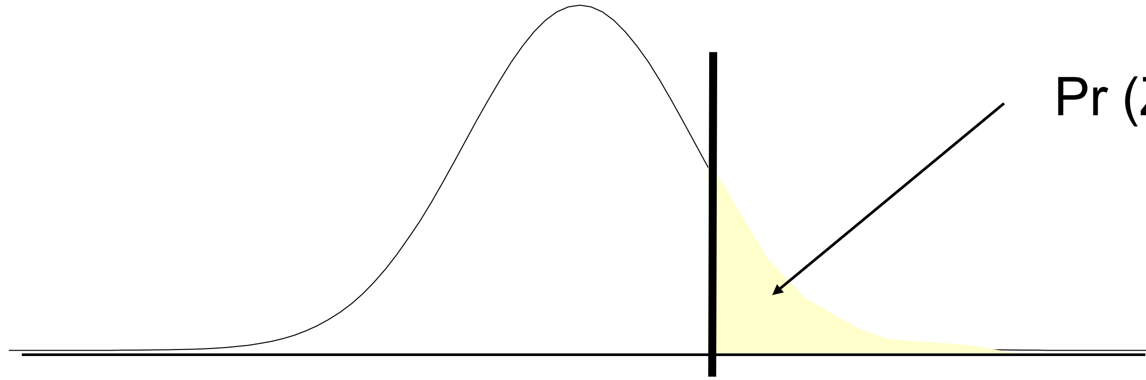


Pr (Z > -0.6) =
Pr (Z < +0.6) =
1 - Pr (Z > 0.6) =
1 - 0.2743 =
0.7257

Pr(Z > 1.83) =
0.0336

Given the Probability - What is Z_0 ?

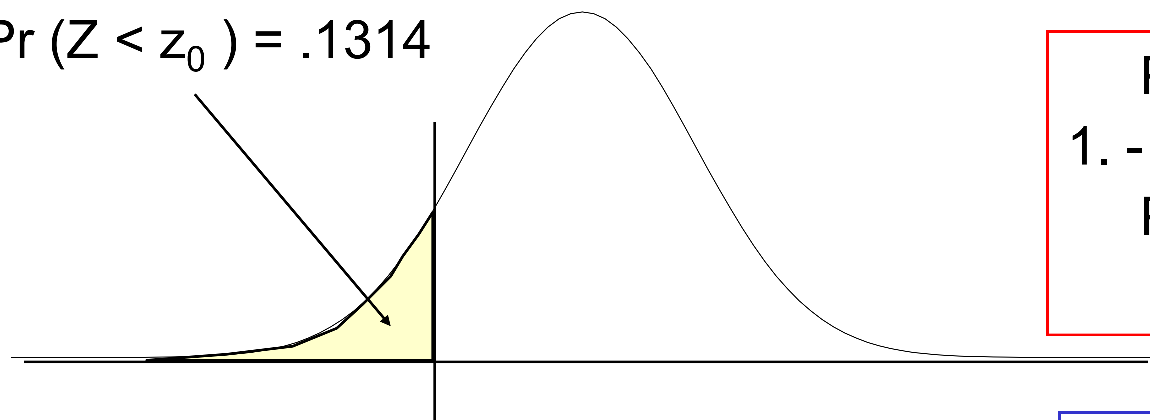
Working backwards.



$Pr(Z > z_0) = .1314$

$z_0 = 1.12$

$Pr(Z < z_0) = .1314$



$$Pr(Z < z_0) = 0.1314$$
$$1. - Pr(Z > z_0) = 0.1314$$
$$Pr(Z > z_0) = 1 - 0.1314$$
$$= 0.8686$$

z_0 must be negative, since $Pr(Z > 0.0) = 0.5$

$z_0 = -1.12$

Converting to Standard Normal Form

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

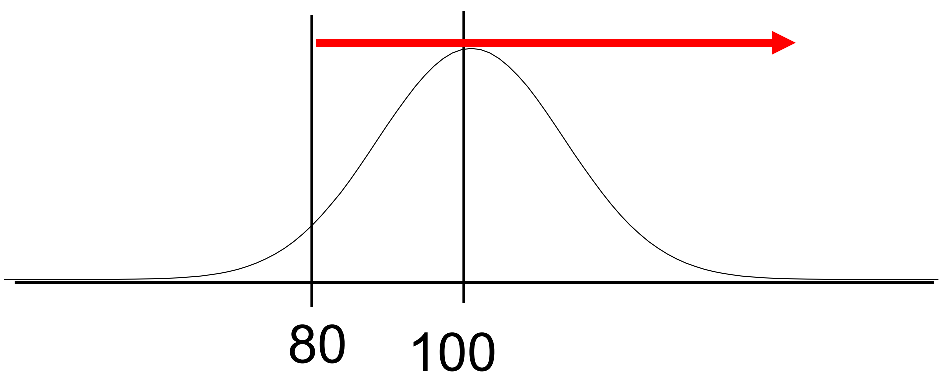
Suppose we have a random variable (say weight), denoted by W, that has a normal distribution with mean 100 and standard deviation 10.

$W \sim N(100, 10)$

$\Pr(W < 90) = \Pr(Z < (90 - 100) / 10) = \Pr(Z < -1.0)$

$= \Pr(Z > 1.0) = 0.1587$

$\Pr(W > 80) = \Pr(Z > (80 - 100) / 10) = \Pr(Z > -2.0)$

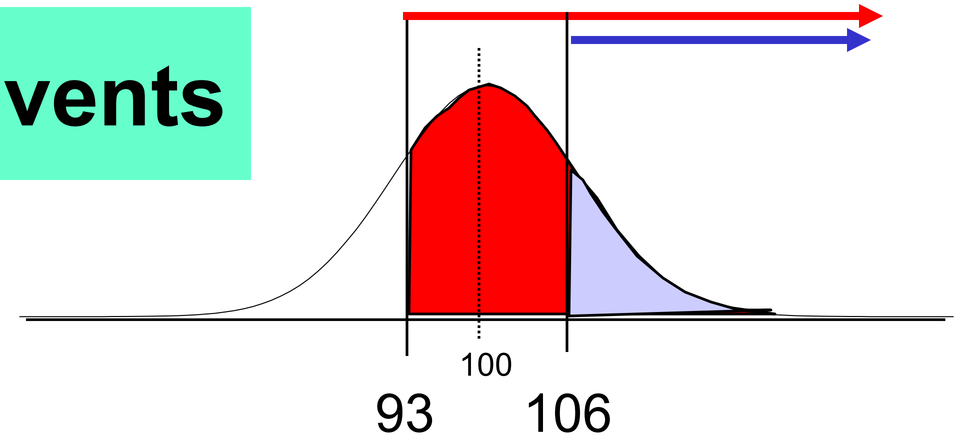


$= 1.0 - \Pr(Z < -2.0)$

$= 1.0 - \Pr(Z > 2.0)$

$= 1.0 - 0.0228 = 0.9772$

Decomposing Events



$W \sim N(100, 10)$

$\Pr(93 < W < 106) =$ $\Pr(W > 93)$ - $\Pr(W > 106)$

$= \Pr[Z > (93 - 100) / 10] - \Pr[Z > (106 - 100) / 10]$

$= \Pr[Z > -0.7] - \Pr[Z > +0.6]$

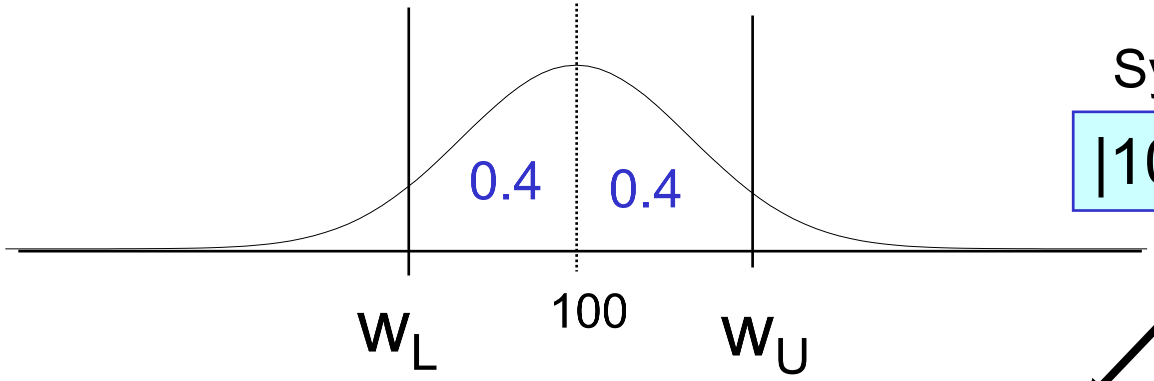
$= 1.0 - \Pr[Z > +0.7] - \Pr[Z > +0.6]$

$= 1.0 - 0.2420 - .2743 = 0.4837$

Finding Interval Endpoints

$W \sim N(100, 10)$

What are the two endpoints for a symmetric area centered on the mean that contains probability of 0.8 ?



Symmetry requirement

$$|100 - w_L| = |100 - w_U|$$

$$\mathbf{Pr(w_L < W < w_U) = 0.8}$$

$$Pr(100 < W < w_U) = Pr(w_L < W < 100) = 0.4$$

$$Pr(100 < W < w_U) = Pr(W > 100) - Pr(W > w_U) =$$
$$Pr(Z > 0) - Pr(Z > (w_U - 100)/10) = .5 - Pr(Z > (w_U - 100)/10) = 0.4$$

$$Pr(Z > (w_U - 100)/10) = 0.1 \Rightarrow (w_U - 100)/10 = z_{0.1} \approx 1.28$$

$$\boxed{w_U = 1.28 * 10 + 100 = 112.8}$$

$$\boxed{w_L = -1.28 * 10 + 100 = 87.2}$$

Probability Practice

Using Table 1 in Ott &Longnecker

Read probability in table using row (. 4) + column (.07) indicators.

$Pr(Z < .47) =$ .6808

$Pr(Z > .47) =$ $1-P(Z<.47)=$ $1-.6808=.3192$

$Pr(Z < -.47) =$ .3192

$Pr(Z > -.47) =$ $1.0 - Pr(Z < -.47)$ $= 1.0 - .3192 = .6808$

$Pr(.21 < Z < 1.56) =$ $Pr(Z < 1.56) - Pr(Z < .21)$

$.9406 - 0.5832 = .3574$

$Pr(-.21 < Z < 1.23) =$ $Pr(Z < 1.23) - Pr(Z < -.21)$

$.8907 - .4168$ $= .4739$

Finding Critical Values from the Table

Find probability in the Table, then read off row and column values.

$\Pr(Z > z_{.2912}) = 0.2912$ $\quad z_{.2912} = \boxed{0.55}$

$\Pr(Z > z_{.05}) = 0.05$ $\quad z_{.05} = \boxed{1.645}$

$\Pr(Z > z_{.025}) = 0.025$ $\quad z_{.025} = \boxed{1.96}$

$\Pr(Z > z_{.01}) = 0.01$ $\quad z_{.01} = \boxed{2.326}$

Sampling Distributions

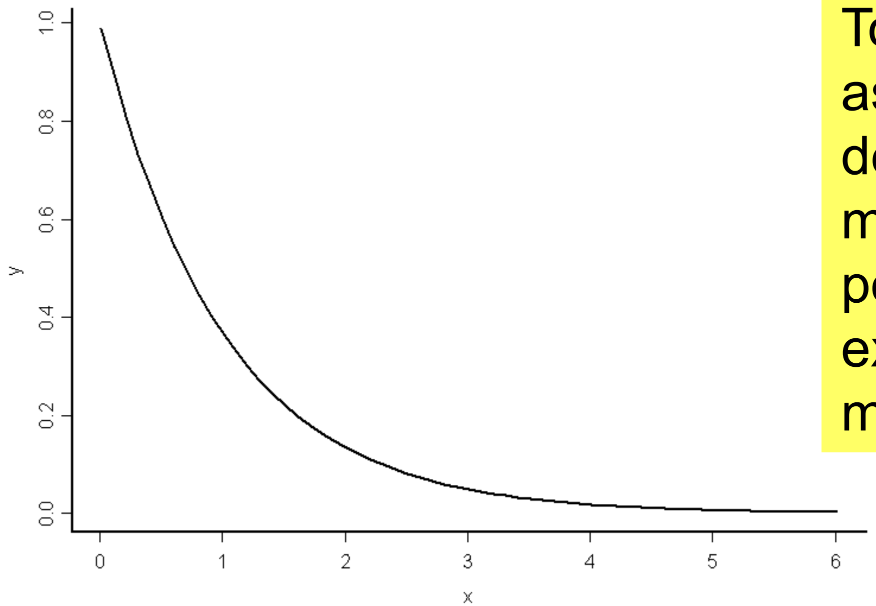
- Typically we select sample data from a population in order to compute some statistic of interest.
- If we were to take two random samples from the same population, it would be very unlikely that we would find that we have computed the exact same value of the statistics.
- Hence, the value of the statistic will vary from sample to sample. That is the statistic itself is a random variable.
- In this lecture we discuss how statistics (functions of data) have distributions of their own, and how those distributions can be determined in some cases by means of the Central Limit Theorem.

Sampling Distribution of the Mean

1. Because no one sample is exactly like the next, **the sample mean** will vary from sample to sample, and hence **is itself a random variable**.
2. Random variables have distributions, and since **the sample mean** is a random variable it **must have a distribution**.
3. Regardless of the distribution of the measurements (discrete or continuous), the **distribution of the sample mean can be approximated by a normal distribution** [Central Limit Theorem].
4. If the sample mean has a normal distribution, we can compute probabilities for specific events using the properties of the normal distribution.

A Resampling Experiment

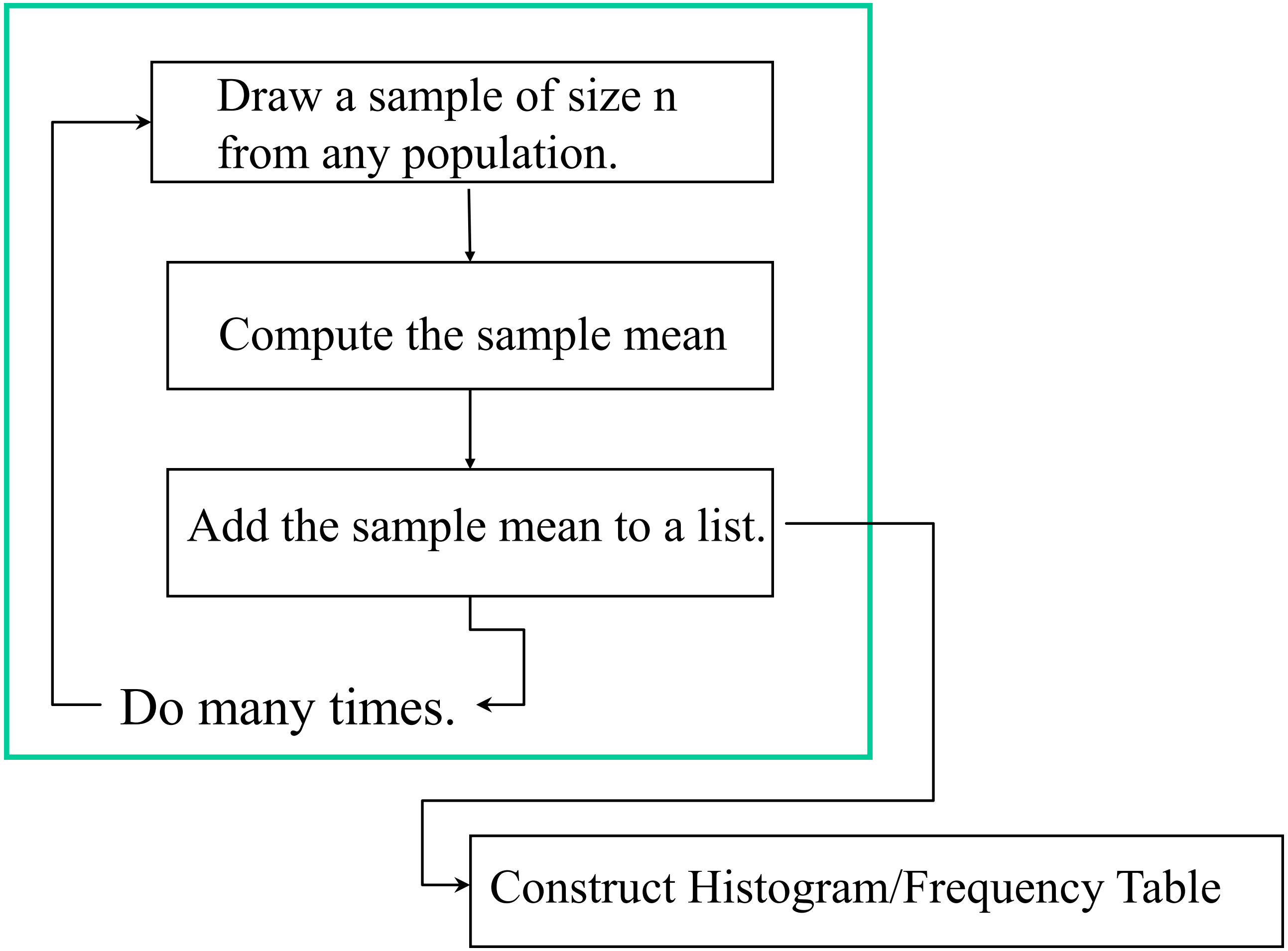
Suppose we are interested in finding the mean of a large population of individuals. A population too large to census. We decided to take a sample of 5 individuals, measure their responses and compute the sample mean. Now suppose we did this 1000 times (I.e. generate 1000 samples of size 5 and hence 1000 means). What would the distribution of these means look like?



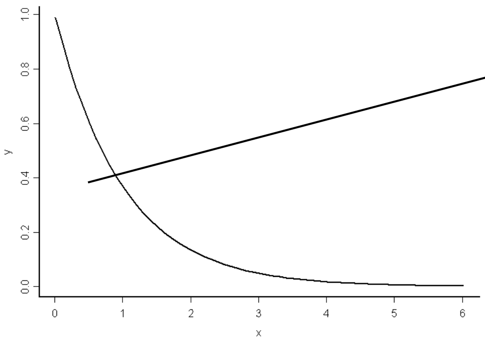
Area under curve is one.

To make things interesting, assume the probability density function of the measurements in the populations has an exponential shape, with mean 1.

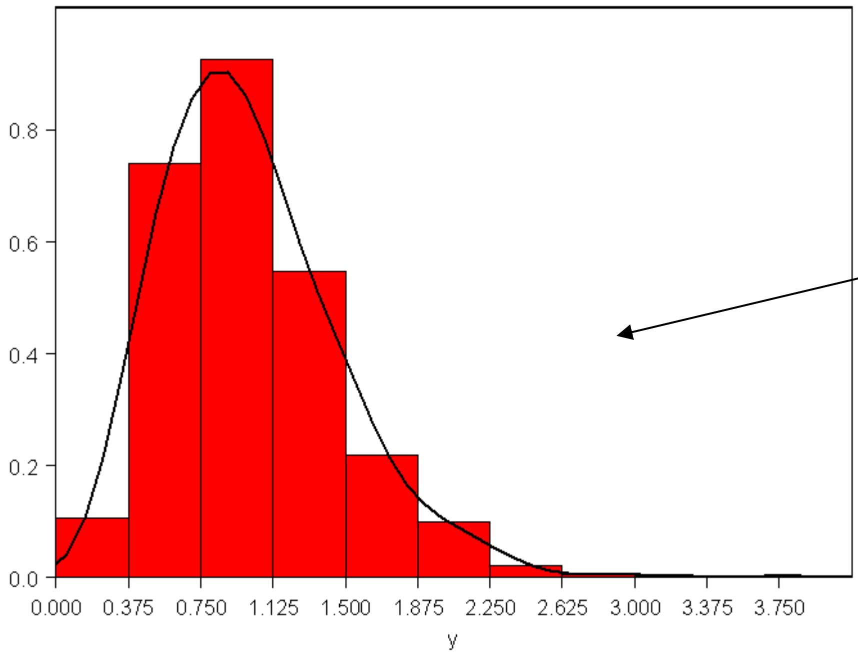
A sampling experiment:



Example Continued

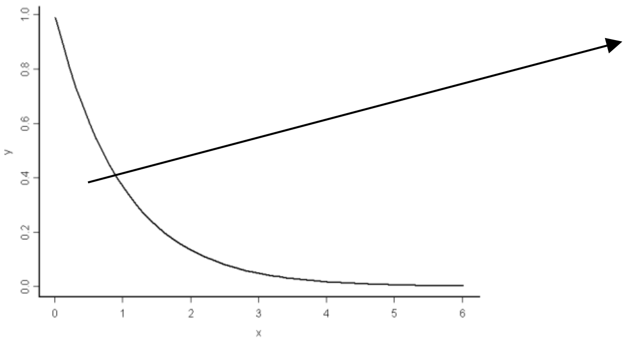


1. Draw a random sample of 5 individuals from population.
2. Compute sample mean.
3. Add mean to list.
4. If number of simulations less than 1000, return to 1 else go to 5.
5. Make histogram of 1000 means.

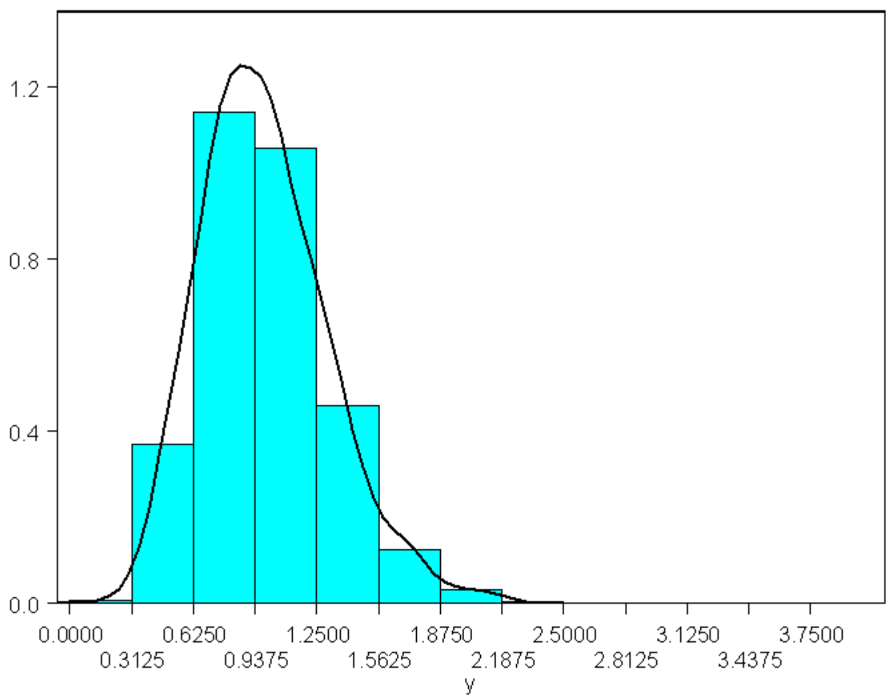


Sample	Mean
1	1.3498
2	0.6293
3	0.7390
4	0.7377
5	1.2206

Mean of 10 Observations

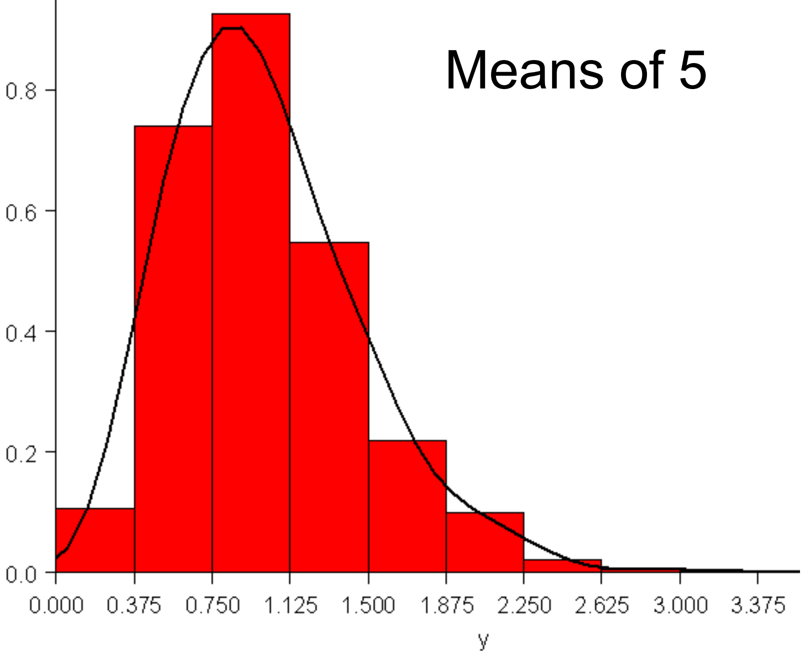


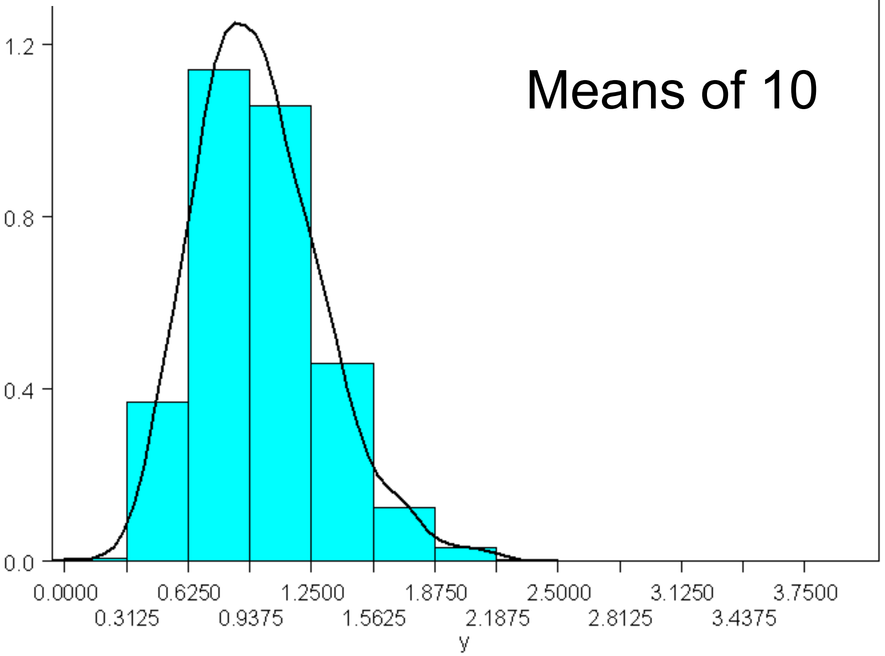
1. Draw a random sample of 10 individuals from population.
2. Compute sample mean.
3. Add mean to list.
4. If number of simulations less than 1000, return to 1 else go to 5.
5. Make histogram of 1000 means.



Comparison

Note: vertical scales are different.





- Both have mean of about 1.0.
- Spread of right plot is narrower than left plot. Somehow, when we look at the distribution of samples of size 10 we have less spread than if we look at samples of size 5. Is there a general rule here?

Central Limit Theorem of Statistics

If random samples, each with ***n*** measurements, are repeatedly drawn from the same population having true mean μ and standard deviation **σ**, then when ***n*** is large, the relative frequency histogram for the sample means (calculated from the repeated samples) will be approximately normal (bell-shaped) with mean μ and standard deviation **σ/√*n*,** that is,

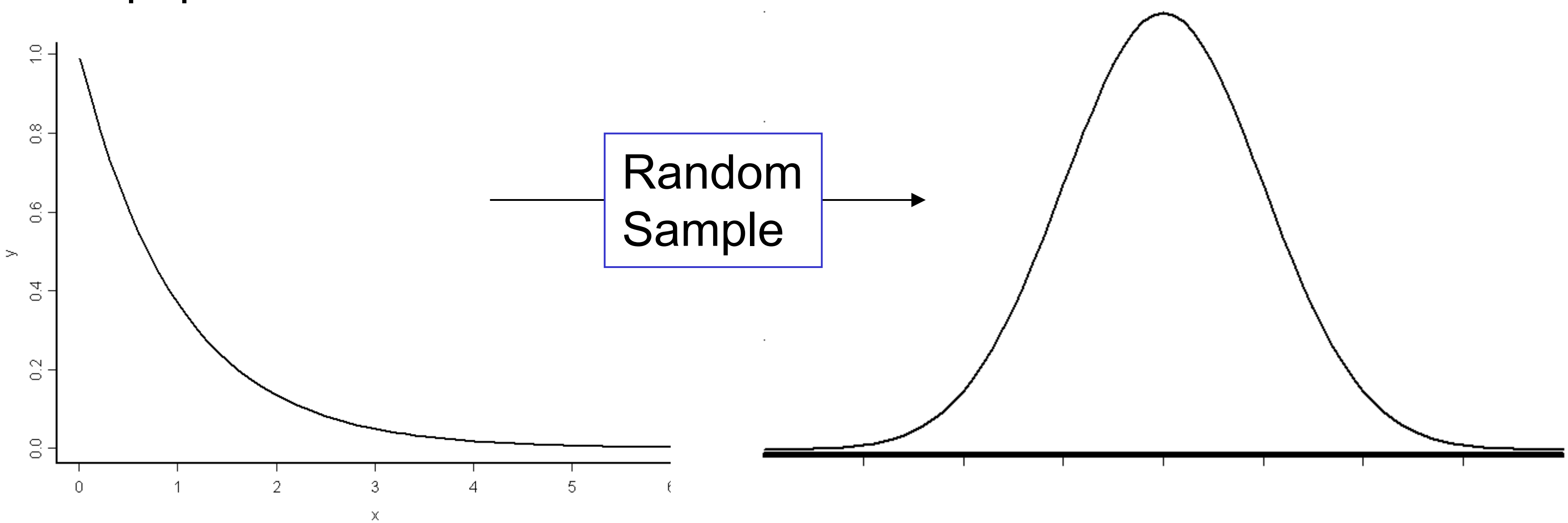
$$\bar{X}_n \sim \text{approx } \mathrm{N}(\mu, \sigma/\sqrt{n})$$

Note: In addition, the approximation becomes closer to true normal as ***n*** increases.

Population and Sampling Distribution

Distribution of measurements in population

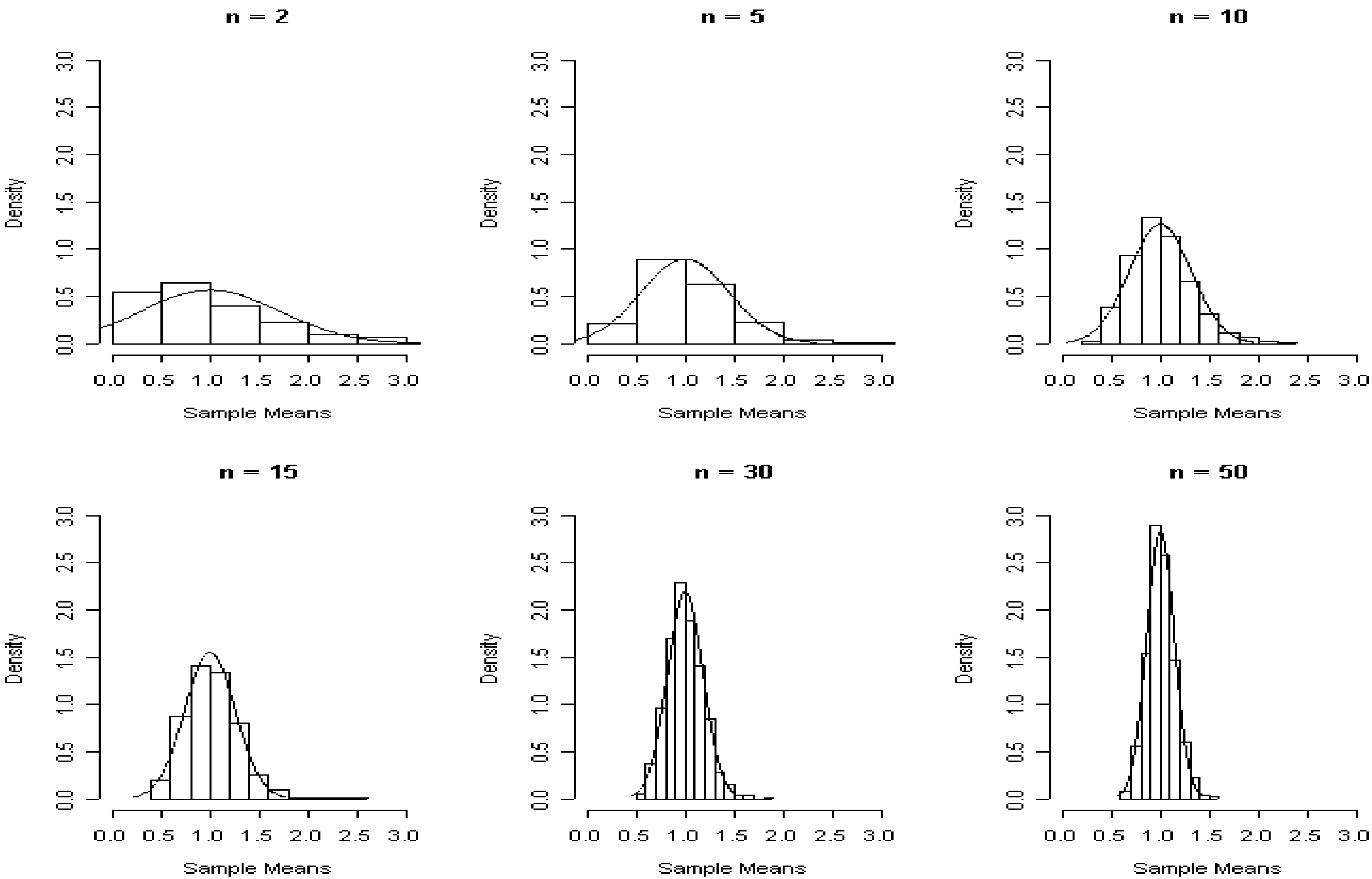
Distribution of means of random samples of size 10 from population.



Mean:	μ
Standard Deviation:	σ
Distribution:	**Anything**

Mean:	μ
Standard Deviation:	$\sigma/\sqrt{10}$
Distribution:	**Approx Normal**

Illustrating the CLT for the distribution of the sample mean when drawing samples from an exponential distribution of mean 1 (based on 1000 draws).



Standard Error of the Mean

The quantity σ is referred to as the **standard deviation**. It is a measure of spread in the population.

The quantity $\sigma/\sqrt{n}$ is referred to as the **standard error of the mean**. It is a measure of spread in the distribution of means of random samples of size n from a population of measurements having true standard deviation σ. I.e. it is just the standard deviation of $\bar{X}_n$.

Uses of the Central Limit Theorem

- The Central Limit Theorem is “central” to Statistics because it allows us to make inferences (decisions) about unknown population parameters, from sample estimates (statistics).
- We can estimate the true mean and standard deviation of a population using the sample mean and sample standard deviation.
- Using the sample mean, the sample standard deviation, and the central limit theorem, we can develop **hypothesis tests** to determine whether the TRUE population mean is equal to some specific value, AND/OR, construct **confidence intervals** for the true mean.

Population: mean μ, std. dev. σ

↓

Draw sample (size n)

↓

Estimate:

$$\mu \rightarrow \bar{x}, \quad \sigma \rightarrow s$$

↓

Draw Inferences:

Quantify uncertainty in estimates (confidence intervals and hypothesis tests).