



Chapter 7

Random Variables & Probability Distributions



Warm-UP

1) What, if anything, is the difference in these variables from the following examples?

A) $3x - 4 = 16$

B) $3x - 4 \leq 16$

C) $3x - 4y = 16$

2) Use the information to create normal distributions distributions based on the summary statistics:

Set A: $n = 24$; $\bar{x} = 69$, $s_x = 12.2$,

Set B: $n = 22$; $\bar{x} = 62.8$, $s_x = 15.7$



Warm-UP

Variables are symbols that represent some unknown quantity, but the unknown can take on different values depending upon the situation:

A) $3x - 4 = 16$; Here the unknown x , is simply a single value

B) $3x - 4 \leq 16$; Here the unknown x , actually varies, taking on a range of values

C) $3x - 4y = 16$; Here there are 2 unknowns, x and y , that have an infinite number of values, related to each other based upon a defined rule

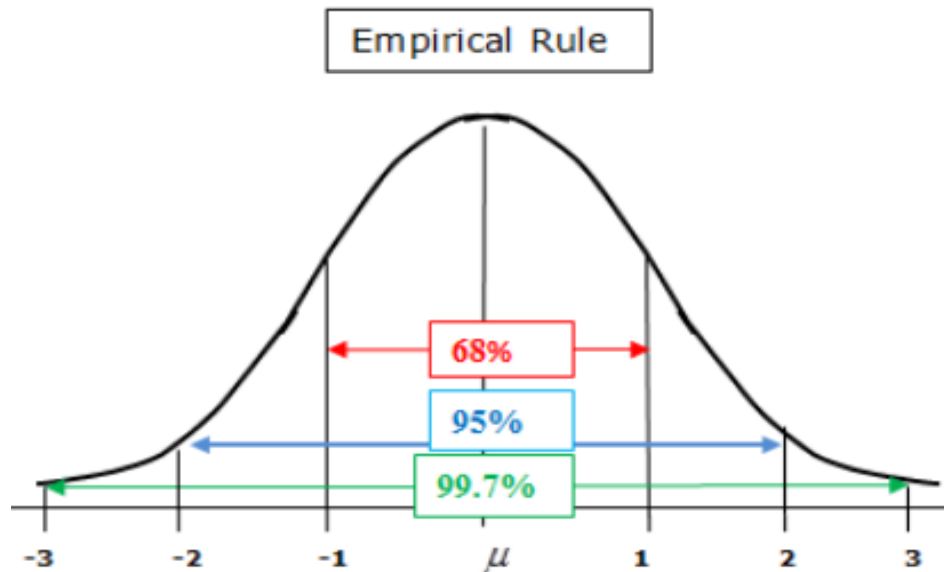


Warm-Up

Use the information to create normal distributions
distributions based on the summary statistics:

Set A: $n = 24$; $\bar{x} = 69$, $s_x = 12.2$,

Set B: $n = 22$; $\bar{x} = 62.8$, $s_x = 15.7$





7.1: Random Variables

A numerical variable whose value depends on the outcome of a chance experiment is called a **random variable**. A random variable associates a numerical value with each outcome of a chance experiment.

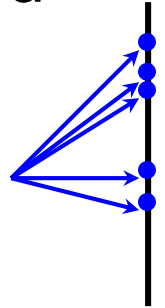




Discrete and Continuous Random Variables

A random variable is **discrete** if its set of possible values is a collection of isolated points on the number line.

Possible values of a
discrete random variable



A random variable is **continuous** if its set of possible values includes an entire interval on the number line.

Possible values of a
continuous random variable



We will use lowercase letters, such as x and y , to represent random variables.



Examples

1. Experiment: A fair die is rolled
Random Variable: The number on the up face
Type: **Discrete**
2. Experiment: A pair of fair dice are rolled
Random Variable: The sum of the up faces
Type: **Discrete**



Examples

3. Experiment: A coin is tossed until the 1st head turns up
Random Variable: The number of the toss that the 1st head turns up
Type: **Discrete**
4. Experiment: Choose and inspect a specified size for a manufactured part
Random Variable: The difference in length (mm) of the part compared to its prescribed optimum.
Type: **Continuous**



Examples

5. Experiment: Inspect the precision of Primary mirror (Hubble Telescope)

Random Variable: The number of defects on the surface of the mirror

Type: **Discrete (strictly a count)**

6. Experiment: Inspect the precision of Primary mirror (Hubble Telescope)

Random Variable: Percentage Variation in amount of curvature compared to optimum

Type: **Continuous (limit of measurement is strictly dependent on precision of tools)**



Examples

7. Experiment: Measure the voltage in a outlet in your room
Random Variable: The voltage
Type: Continuous
8. Experiment: Observe the amount of time it takes a bank teller to serve a customer
Random Variable: time in minutes
Type: Continuous



Examples

9. Experiment: Measure the time until the next customer arrives at a customer service window

Random Variable: The time

Type: Continuous

10. Experiment: Inspect a randomly chosen circuit board from a production line

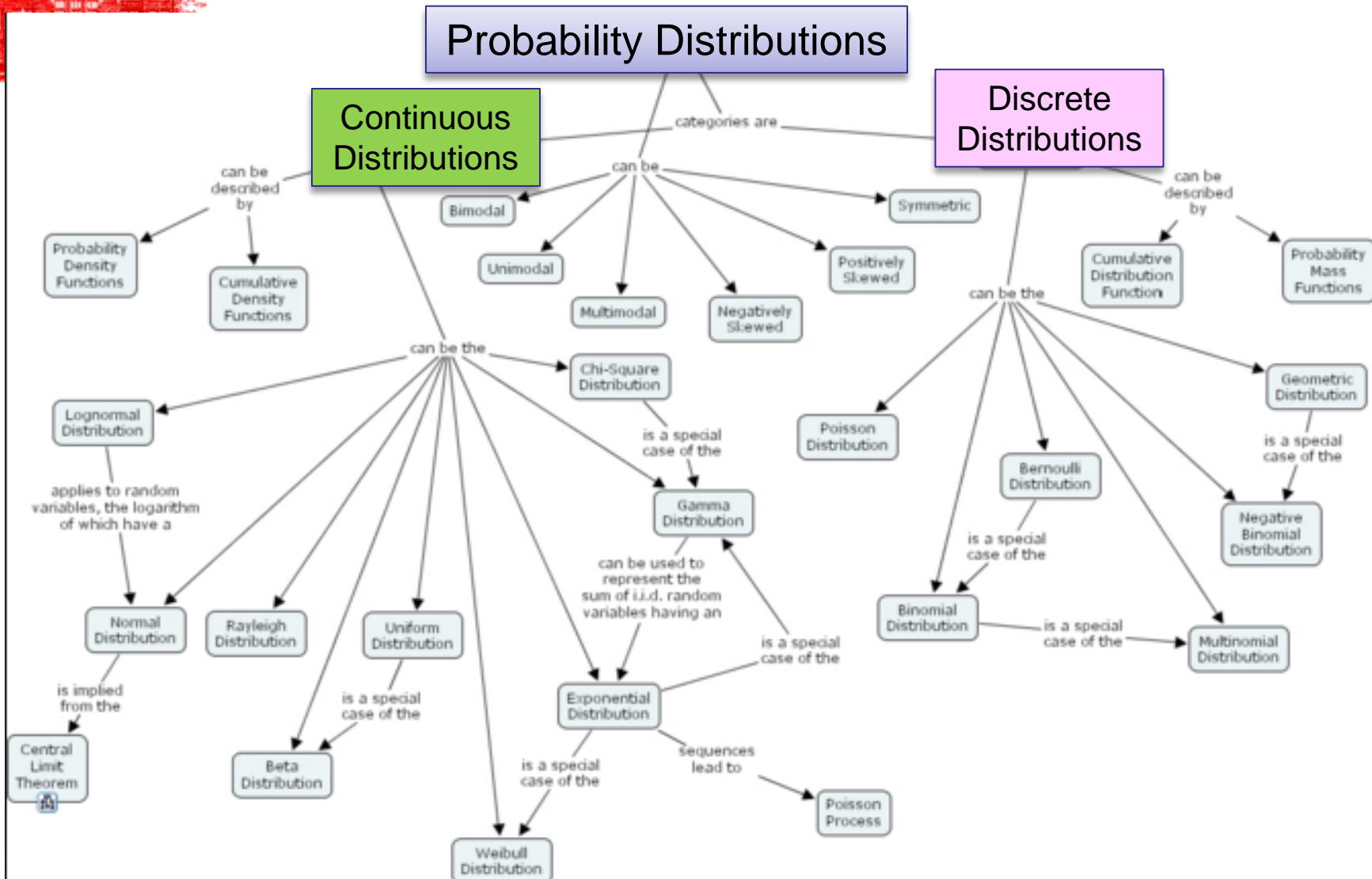
Random Variable:

1, if the circuit board is defective

0, if the circuit board is not defective

Type: Discrete

Schematic Typology





Notation for Random Variables

- For a probability $P(X \leq x)$, what do x and X mean here?

A chosen constant

- Consider X to be the random variable which represents the outcome of a single roll of a die, so that X can take on values of $\{1, 2, 3, 4, 5, 6\}$
- $P(X \leq 2)$ means what is the probability that the outcome will be 1 or 2.
- $P(X \leq 5)$ means what is the probability that the outcome will be 1, 2, 3, 4, or 5.



Notation: In general $P(X \leq x)$

...means the probability that the random variable X is less than or equal to the realization x .

Our textbook might show the following: Given two common dice, the random variable is the sum of the two dice $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

What is the probability that the sum is six or less?

$$P(X \leq 6) = p(x \leq 6)$$

What is the probability that the sum is 9?

$$P(X = 9) = p(x = 9) = p(9)$$



7.2: Probability Distributions for Discrete Random Variables

The **probability distribution of a discrete random variable x** gives the probability associated with each possible x value.

Each probability is the limiting relative frequency of occurrence of the corresponding x value when the experiment is repeatedly performed (LOLn).

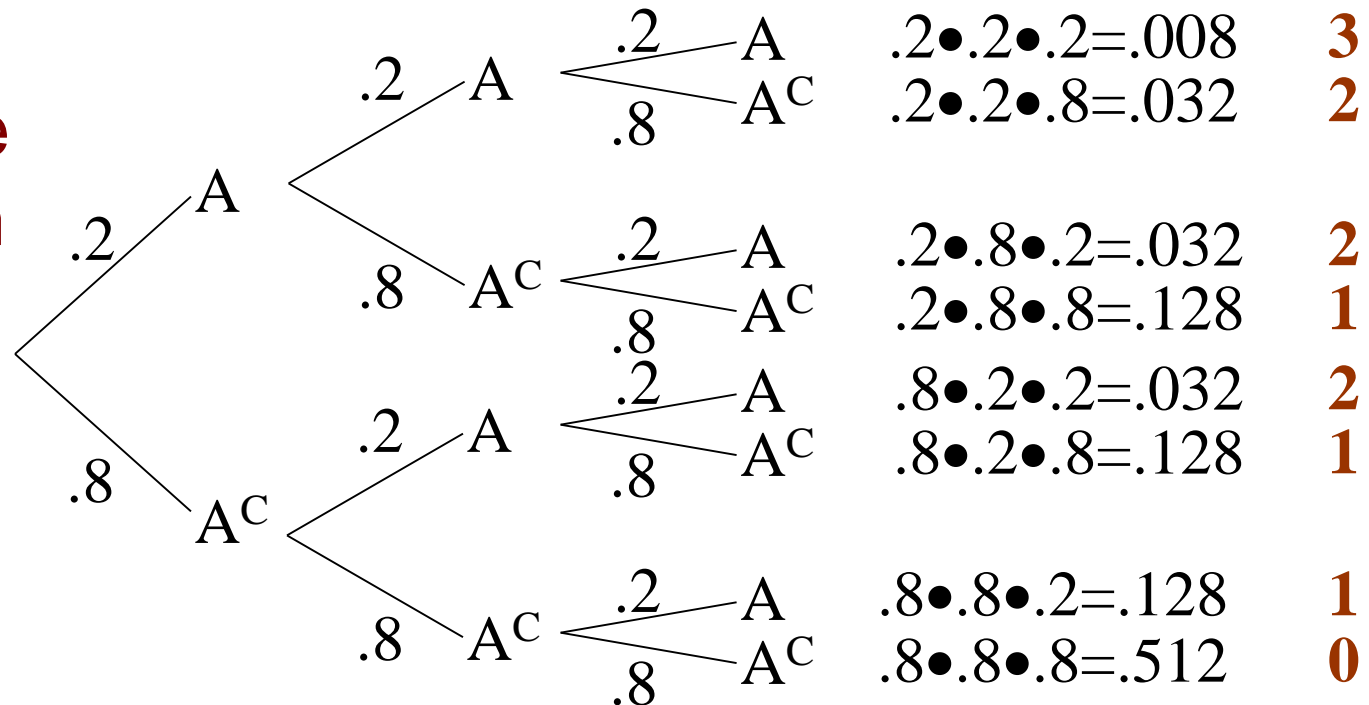
Roll	1	2	3	4	5	6
$p =$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$p =$	0.167	0.167	0.167	0.167	0.167	0.167



Example

Suppose that 20% of the apples sent to a sorting line are Grade A. If 3 of the apples sent to this plant are chosen randomly, determine the probability distribution of the number of Grade A apples in a sample of 3 apples.

Consider the tree diagram





The Results in Table Form

x	$p(x)$
0	$1(.8)^3$
1	$3(.8)^2(.2)^1$
2	$3(.8)^1(.2)^2$
3	$1(.2)^3$

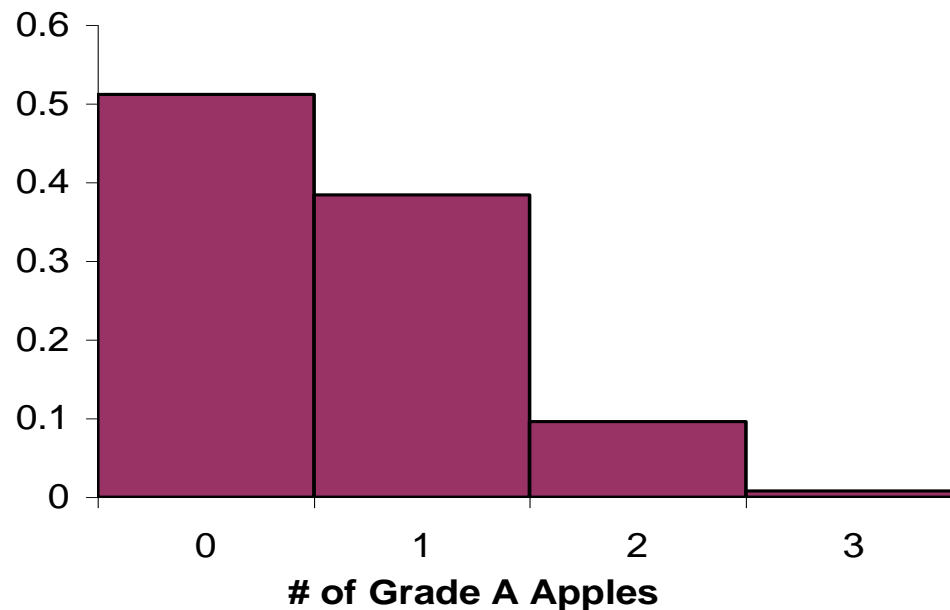
or

x	$p(x)$
0	0.512
1	0.384
2	0.096
3	0.008



Results in Graphical Form (Probability Histogram)

Probabilty Histogram



For a probability histogram, the area of a bar is the probability of obtaining that value associated with that bar.



Properties of Discrete Probability Distributions

The probabilities p_i must satisfy

1. $0 \leq p_i \leq 1$ for each i
2. $p_1 + p_2 + \dots + p_k = 1$

The probability $P(X \text{ in } A)$ of any event is found by summing the p_i for the outcomes x_i making up A .



Example

The number of items a given salesman sells per customer is a random variable. The table below is for a specific salesman (Wilbur) in a clothing store in the mall. The probability distribution of X is given below:

x	0	1	2	3	4	5	6
$p(x)$	0.20	0.35	0.15	0.12	0.10	0.05	0.03

Note: $0 \leq p(x) \leq 1$ for each x

$\sum p(x) = 1$ (the sum is over all values of x)



Example - continued

x	0	1	2	3	4	5	6
p(x)	0.20	0.35	0.15	0.12	0.10	0.05	0.03

The probability that he sells at least three items to a randomly selected customer is

$$P(X \geq 3) = 0.12 + 0.10 + 0.05 + 0.03 = \mathbf{0.30}$$

The probability that he sells *at most three items* to a randomly selected customer is

$$P(X \leq 3) = 0.20 + 0.35 + 0.15 + 0.12 = \mathbf{0.82}$$

The probability that he sells between (inclusive) 2 and 4 items to a randomly selected customer is

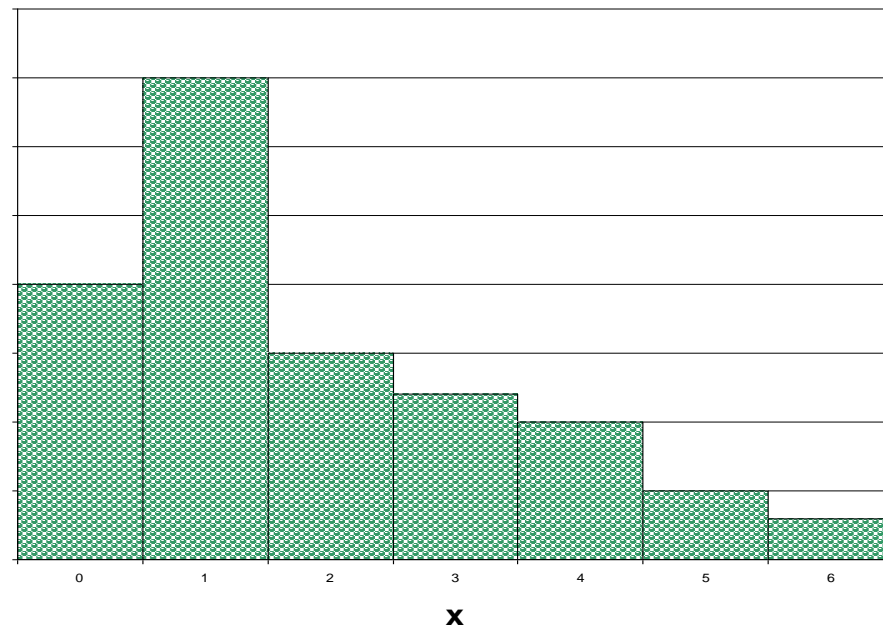
$$P(2 \leq X \leq 4) = 0.15 + 0.12 + 0.10 = \mathbf{0.37}$$



Probability Histogram

A probability histogram has its vertical scale adjusted in a manner that makes the area associated with each bar equal to the probability of the event that the random variable takes on the value describing the bar.

Probability Histogram





The **Chuck-a-LUCK** Game

Costs \$1 to play

Dice game of chance

Pick any number from 1 to 6. Roll the three dice.

If your chosen number comes up, you WIN!

If your number comes up 1 time = \$1

If your number comes up twice = \$2

If your number comes up thrice = \$3

Is this a fair game? What is the expected value?



The **Chuck-a-LUCK** Game

Costs \$1 to play Dice game of chance

Pick any number from 1 to 6. Roll the three dice

We must define the random variable ... X is?

X : the number of occurrences of my chosen number (1 to 6) on the dice

Now, create a probability table...

X	0	1	2	3
$P(X)$				



The **Chuck-a-LUCK** Game

What are the probabilities of the table?

X	0	1	2	3
P(X)	$\frac{125}{216}$	$\frac{25}{216}$	$\frac{5}{216}$	$\frac{1}{216}$

Is this a valid probability distribution ?

$$P(X) = P(x_1) + P(x_2) + P(x_3) + \cdots + P(x_n) = 1$$

$$P(X) = \frac{125}{216} + \frac{25}{216} + \frac{5}{216} + \frac{1}{216} \stackrel{?}{\Rightarrow} 1$$



The **Chuck-a-LUCK** Game

What are the probabilities of the table?

X	0	1	2	3
P(X)	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

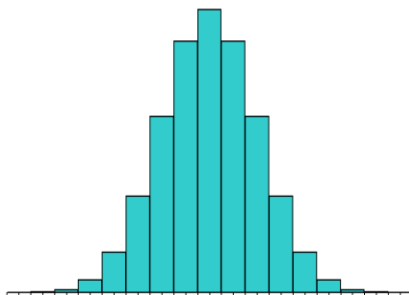
What is the Expected Value $E(X)$ or μ_x ?

$$E(X) \text{ or } \mu_x = \sum x \cdot p(x)$$
$$E(X) = 0 \cdot \frac{125}{216} + 1 \cdot \frac{75}{216} + 2 \cdot \frac{15}{216} + 3 \cdot \frac{1}{216} = 0.5$$

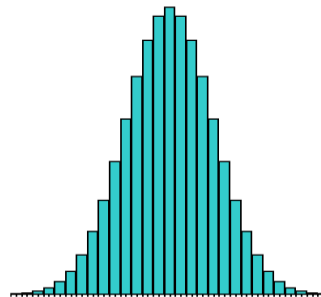


Continuous Probability Distributions

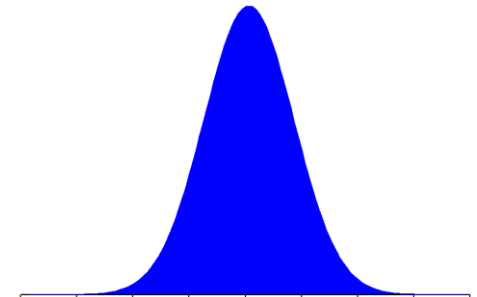
If one looks at the distribution of the actual amount of water (in ounces) in “one gallon” bottles of spring water they might see something such as



Amount measured
to nearest
hundredths of an
ounce.



Amount measured
to nearest ten
thousandths of an
ounce.



**Limiting curve as
the accuracy
increases**



The HOME stretch!





Warm-UP: Review

- 1) What is a z score?
- 2) Given a normal distribution, what is the proportion of observations that fall between z-scores of $-1.6 < x < 0.7$?
- 3) What is the z-score associated with the Standard normal probability of .892?
- 4) What is an Apgar score? (don't know...look it up!)



WARM-UP

1) What is a z score?

A z score gives the relationship between an observation (x_i) and the mean (\bar{x}) of a distribution in terms of some number of standard deviations. It is positive when it's greater than the mean, and negative when it's less than the mean.

2) Given a normal distribution, what is the proportion of observations that fall between z-scores of $-1.6 < x < 0.7$?

Approx. 0.703 or 70.3% of the observations



Warm-UP

3) What is the z-score associated with the Standard normal probability of .892?

z score = 1.237

4) What is an Apgar score?



Warm-Up

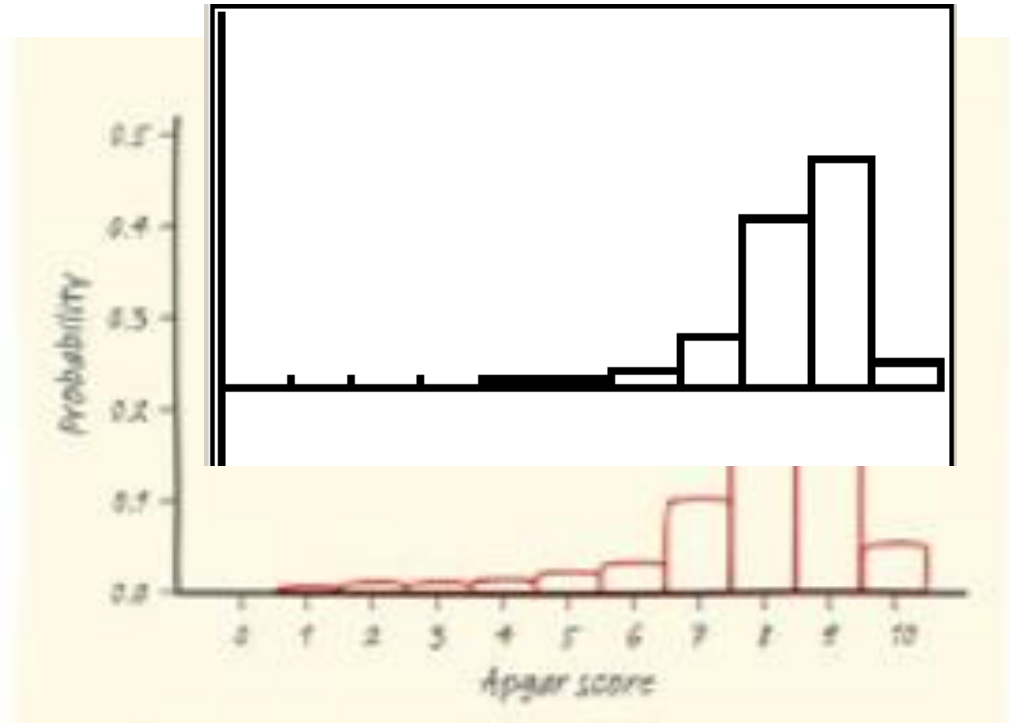
What is an Apgar score? What are the possible values for the random variable?

Virginia Apgar invented the **Apgar score** in 1952 as a method to quickly summarize the health of newborn children. Apgar was an anesthesiologist who developed the score in order to ascertain the effects of *obstetric anesthesia* on babies. The Apgar scale is determined by evaluating the newborn baby on five simple criteria on a scale from zero to two, then summing up the five values thus obtained. The resulting Apgar score ranges from zero to 10.



Histogram of Probability Distribution

L1	L2
0	.001
1	.006
2	.007
3	.008
4	.012
5	.02
6	.038
L2(1) = .001	



Can you make this histogram on your calculator?



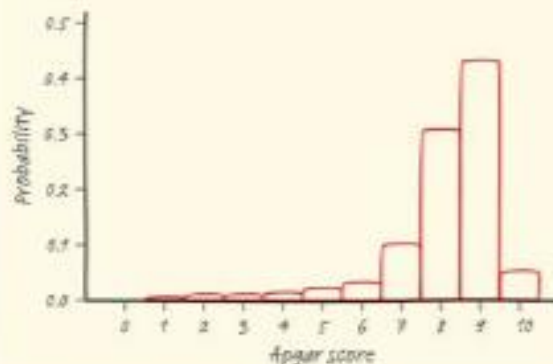
Apgar Scores

■ Example: Babies' Health at Birth (Apgar Scores)

- (a) Show that the probability distribution for X is legitimate.
- (b) Make a histogram of the probability distribution. Describe what you see.
- (c) Apgar scores of 7 or higher indicate a healthy baby. What is $P(X \geq 7)$?

Value:	0	1	2	3	4	5	6	7	8	9	10
Probability:	0.001	0.006	0.007	0.008	0.012	0.020	0.038	0.099	0.319	0.437	0.053

(a) All probabilities are between 0 and 1 and they add up to 1. This is a legitimate probability distribution.



(c) $P(X \geq 7) = .908$
We'd have a 91 % chance of randomly choosing a healthy baby.

Notice the difference between $>$ and \geq . With discrete variables, these are different but not with continuous variables.

(b) The left-skewed shape of the distribution suggests a randomly selected newborn will have an Apgar score at the high end of the scale. There is a small chance of getting a baby with a score of 5 or lower.



7.3: Probability Distribution for a Continuous Random Variable

A **probability distribution for a continuous random variable x** is specified by a mathematical function denoted by $f(x)$ which is called the **density function**. The graph of a density function is a smooth curve (the **density curve**).

The following requirements must be met:

1. $f(x) \geq 0$
2. The total area under the density curve is equal to 1.

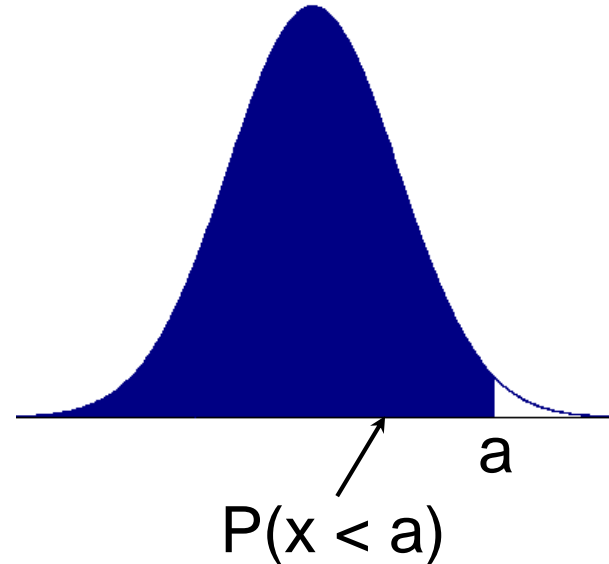
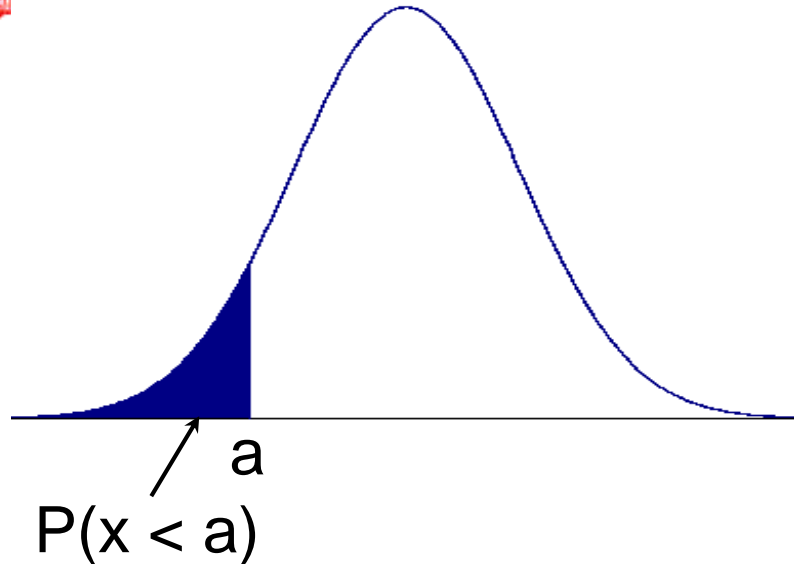
The probability that x falls in any particular interval is the area under the density curve that lies above the interval.



Probability Density Function

The probability that a discrete random variable X takes on a particular value x , that is, $P(X = x)$, is frequently denoted $f(x)$. The function $f(x)$ is typically called the **probability mass function**, although some authors also refer to it as the **probability function**, the **frequency function**, or **probability density function**. Most college level statistics courses will use the common terminology — the probability mass function — and its common abbreviation — the **p.m.f.**

Some Illustrations



Notice that for a continuous random variable x , $P(x = a) = 0$ for any specific value a because the “area above a point” under the curve is a line segment and hence has 0 area.

Specifically this means $P(x < a) = P(x \leq a)$.



Discrete Random Variables

Definition. The **probability mass function**, $P(X = x) = f(x)$, of a discrete random variable X is a function that satisfies the following properties:

(1) $P(X = x) = f(x) > 0$ if $x \in$ the support S

(2) $\sum_{x \in S} f(x) = 1$ or $\sum_{x \in S} f(x) = 1$

(3) $P(X \in A) = \sum_{x \in A} f(x)$



Continuous Random Variables

Definition. The **probability density function** ("p.d.f.") of a continuous random variable X with support S is an integrable function $f(x)$ satisfying the following:

(1) $f(x)$ is positive everywhere in the support S , that is, $f(x) > 0$, for all x in S

(2) The area under the curve $f(x)$ in the support S is 1, that is:

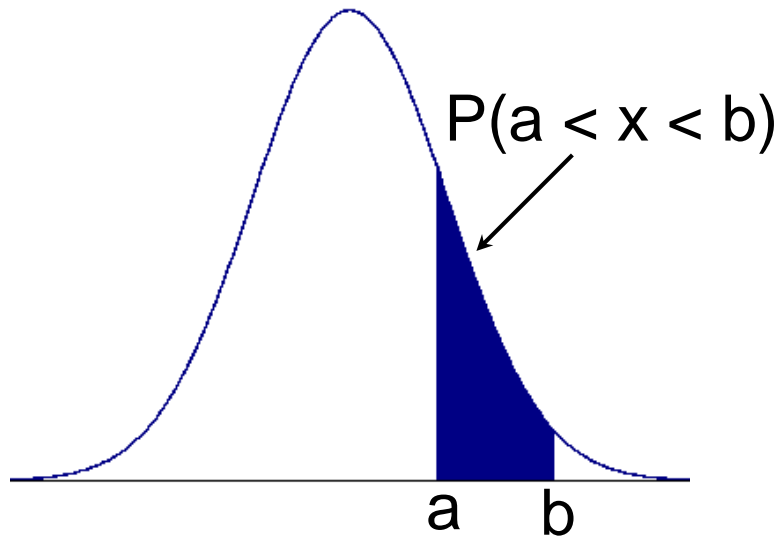
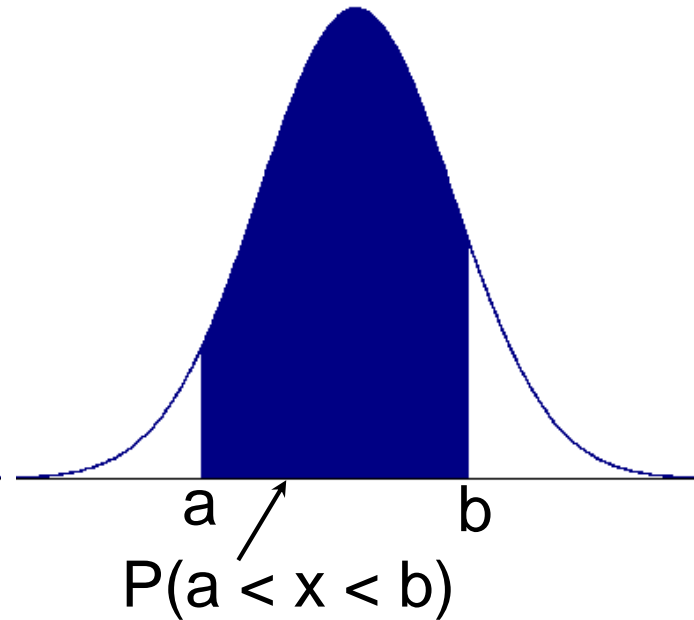
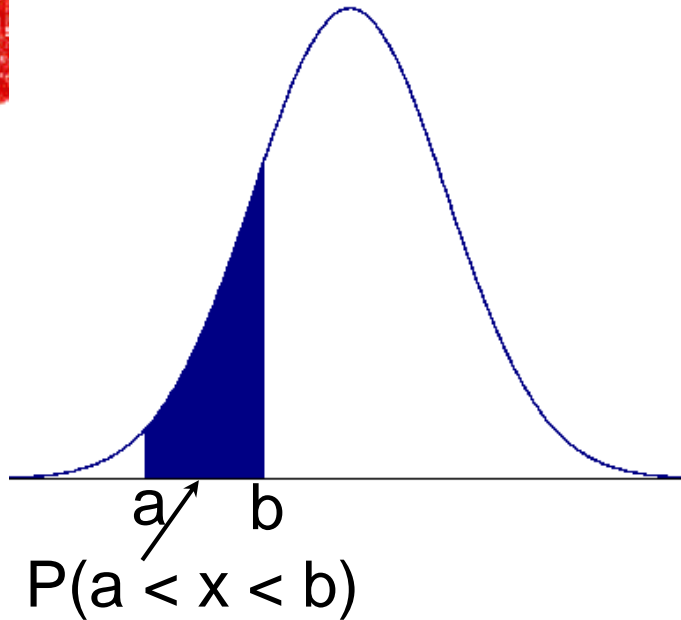
$$\int_S f(x) dx = 1$$

(3) If $f(x)$ is the p.d.f. of x , then the probability that x belongs to A , where A is some interval, is given by the integral of $f(x)$ over that interval, that is:

$$P(X \in A) = \int_A f(x) dx$$

****** As you can see, the definition for the p.d.f. of a continuous random variable differs from the definition for the p.m.f. of a discrete random variable by simply changing the summations that appeared in the discrete case to integrals in the continuous case.

Some Illustrations

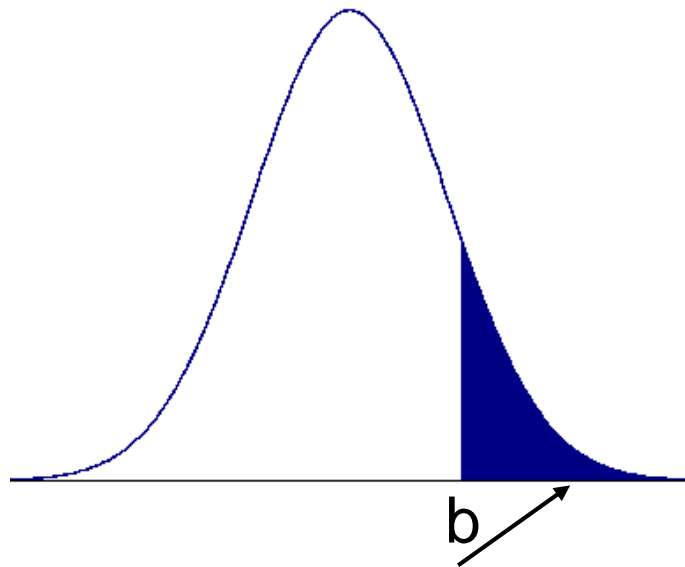


Note:

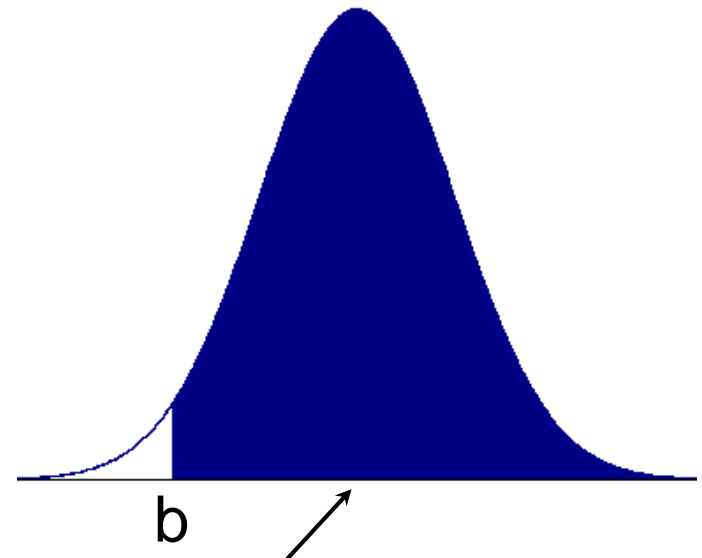
$$\begin{aligned} P(a < x < b) &= P(a \leq x < b) \\ &= P(a < x \leq b) \\ &= P(a \leq x \leq b) \end{aligned}$$



Some Illustrations



$$P(x > b) = P(b < x)$$



$$P(x > b) = P(b < x)$$



Example

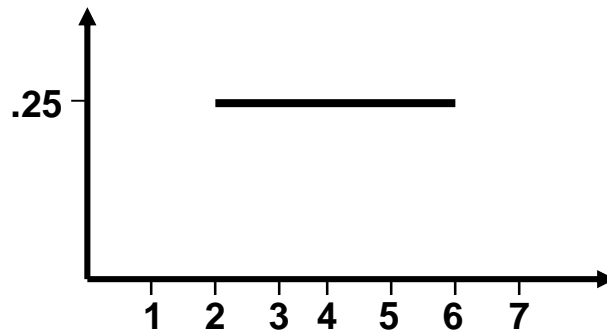
Define a continuous random variable x by

x = the weight of the crumbs in ounces left on the floor of a restaurant during a one hour period.

Suppose that x has a probability distribution with density function

$$f(x) = \begin{cases} .25 & 2 < x < 6 \\ 0 & \text{otherwise} \end{cases}$$

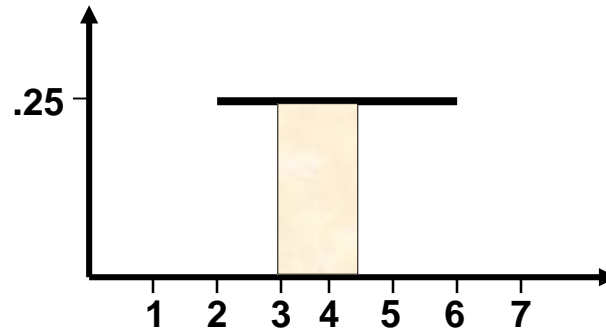
The graph looks like





Example

Find the probability that during a given 1 hour period between 3 and 4.5 ounces of crumbs are left on the restaurant floor.



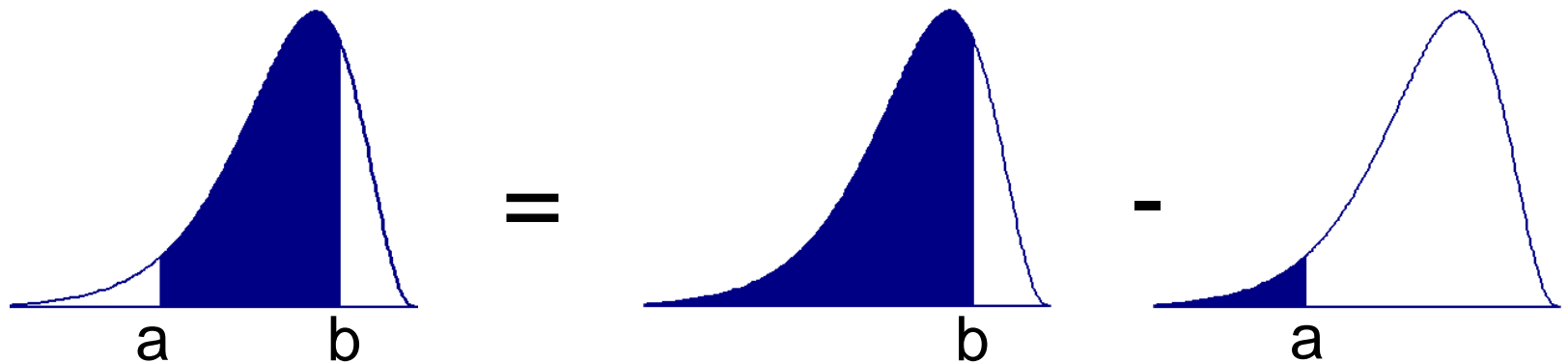
The probability is represented by the shaded area in the graph. Since that shaded area is a rectangle, $\text{area} = (\text{base})(\text{height}) = (1.5)(.25) = .375$



Method of Probability Calculation

The probability that a continuous random variable x lies between a lower limit a and an upper limit b is

$$\begin{aligned} P(a < x < b) &= (\text{cumulative area to the left of } b) - \\ &\quad (\text{cumulative area to the left of } a) \\ &= P(x < b) - P(x < a) \end{aligned}$$





7.4: Mean & Standard Deviation of Random Variable

The **mean value of a random variable x** , denoted by μ_x , describes where the probability distribution of x is centered.

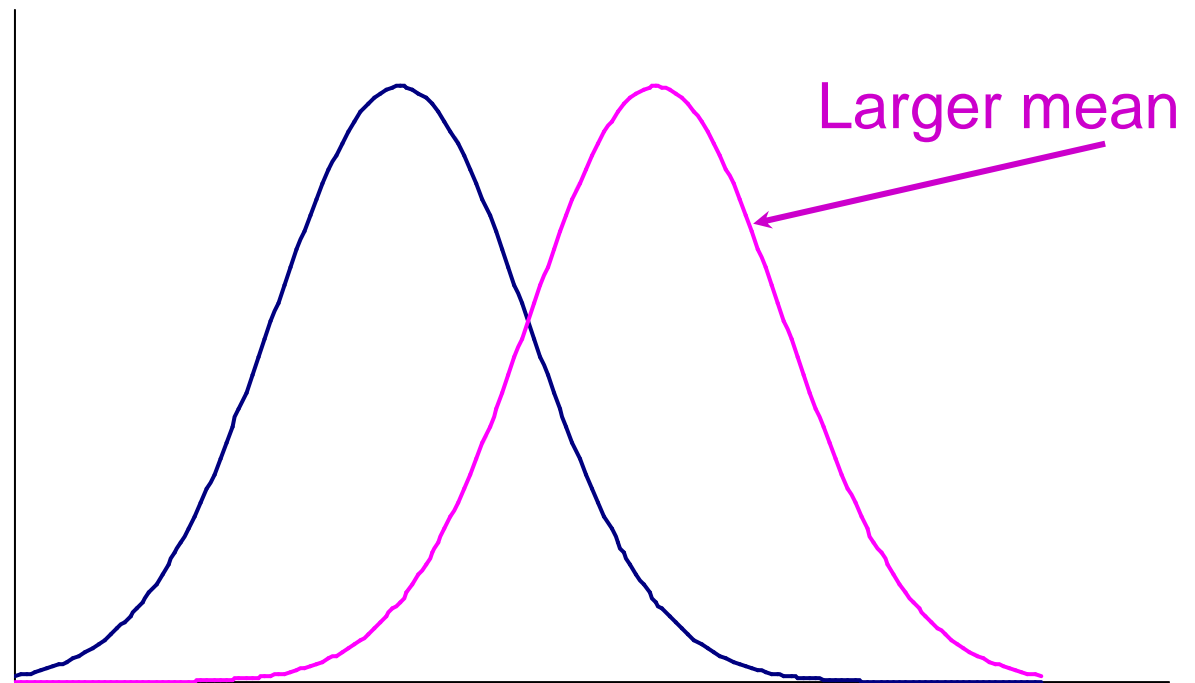
The **standard deviation of a random variable x** , denoted by σ_x , describes variability in the probability distribution.

1. When σ_x is small, observed values of x will tend to be close to the mean value and
2. when σ_x is large, there will be more variability in observed values.



Illustrations

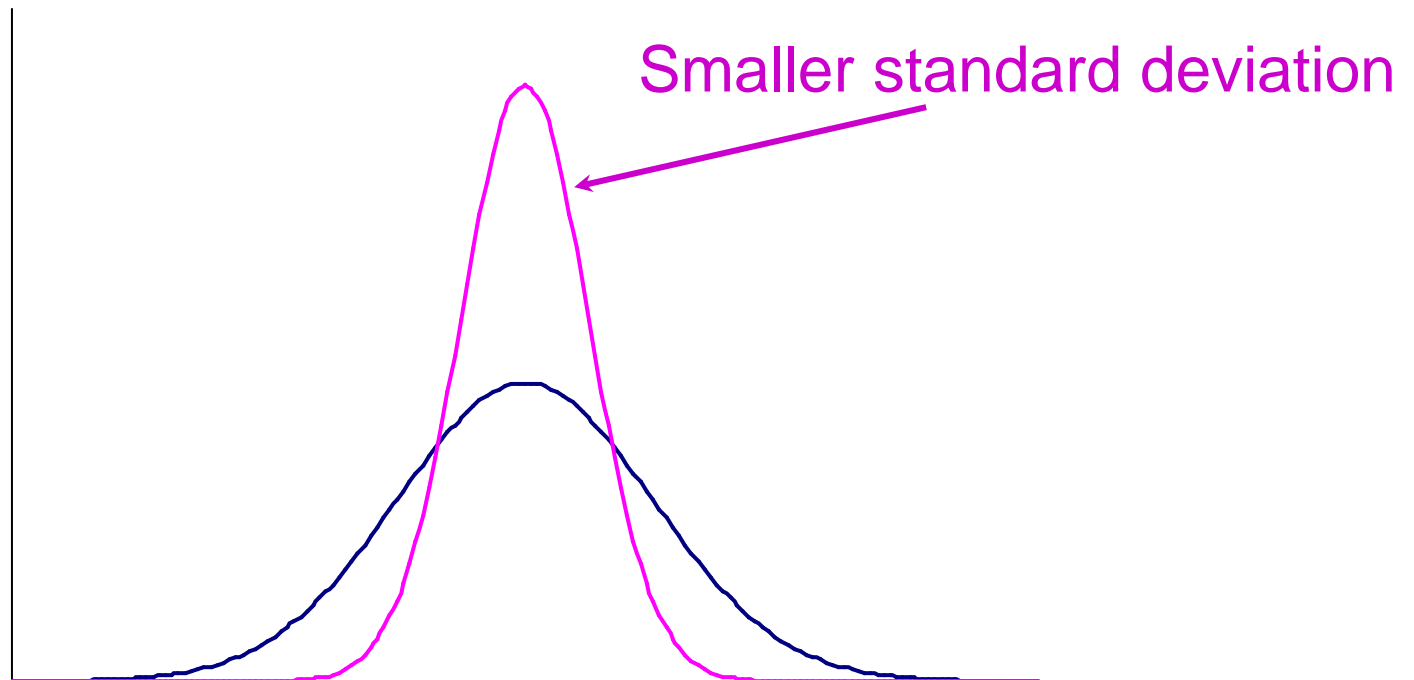
Two distributions with the same standard deviation with different means.





Illustrations

Two distributions with the same means and different standard deviation.





Mean of a Discrete Random Variable or the Expected Value

The **mean value of a discrete random variable x** , denoted by μ_x , is computed by first multiplying each possible x value by the probability of observing that value and then adding the resulting quantities. Symbolically,

$$\mu_x = \sum_{\text{all possible values of } x} x \cdot p(x)$$

alternatively, the notation $E(x)$ indicating Expected value



Variance and Standard Deviation of a Discrete Random Variable

The **Variance of a Discrete Random Variable x** , denoted by σ_x^2 is computed by first subtracting the mean from each possible x value to obtain the deviations, then squaring each deviation and multiplying the result by the probability of the corresponding x value, and then finally adding these quantities. Symbolically,

$$\sigma_x = \sqrt{\sigma_x^2}$$

The **standard deviation of x** , denoted by σ_x , is the square root of the variance.

$$\sigma_x^2 = \sum_{\text{all possible values of } x} (x - \mu)^2 \cdot p(x)$$



Example

A professor regularly gives multiple choice quizzes with 5 questions. Over time, he has found the distribution of the number of wrong answers on his quizzes is as follows

x	$P(x)$
0	0.25
1	0.35
2	0.20
3	0.15
4	0.04
5	0.01



Example

Multiply each x value by its probability and add the results to get μ_x .

x	$P(x)$	$x \cdot P(x)$
0	0.25	0.00
1	0.35	0.35
2	0.20	0.40
3	0.15	0.45
4	0.04	0.16
5	0.01	0.05
		<hr/> 1.41

$$\mu_x = 1.41$$



Quiz Distribution - continued

x	P(x)	x•P(x)	x - μ	$(x - \mu)^2$	$(x - \mu)^2 \cdot P(x)$
0	0.25	0.00	-1.41	1.9881	0.4970
1	0.35	0.35	-0.41	0.1681	0.0588
2	0.20	0.40	0.59	0.3481	0.0696
3	0.15	0.45	1.59	2.5281	0.3792
4	0.04	0.16	2.59	6.7081	0.2683
5	0.01	0.05	3.59	12.8881	0.1289
		<hr/> 1.41			<hr/> 1.4019

$$\text{Variance} = \sigma_x^2 = 1.4019$$

Standard deviation

$$= \sigma_x = \sqrt{1.4019} = 1.184$$



The Mean & Variance of a Linear Function

If x is a random variable with mean μ_x and variance σ_x^2 and a and b are numerical constants, the random variable y defined by $y = a + bx$ is called a **linear function of the random variable x** .

The mean of $y = a + bx$ is $\mu_y = \mu_{a+bx} = a + b\mu_x$

The variance of y is $\sigma_y^2 = \sigma_{a+bx}^2 = b^2 \sigma_x^2$

From which it follows that the standard deviation of y is $\sigma_y = \sigma_{a+bx} = |b| \sigma_x$



Example

Suppose x is the number of sales staff needed on a given day. If the cost of doing business on a day involves fixed costs of \$255 and the cost per sales person per day is \$110, find the mean cost (the mean of x or μ_x) of doing business on a given day where the distribution of x is given below.

x	$p(x)$
1	0.3
2	0.4
3	0.2
4	0.1

First consider: *What is the mean of the random variable?*



Example continued

We need to find the mean of $y = 255 + 110x$

x	p(x)	xp(x)
1	0.3	0.3
2	0.4	0.8
3	0.2	0.6
4	0.1	0.4
		2.1

What is the mean of the random variable?

$$\mu_x = 2.1$$

$$\begin{aligned}\mu_y &= \mu_{255+110x} = 255 + 110\mu_x \\ &= 255 + 110(2.1) = \$486\end{aligned}$$



Example continued

We need to find the variance and standard deviation of $y = 255 + 110x$

x	p(x)	$(x-\mu)^2 p(x)$
1	0.3	0.3630
2	0.4	0.0040
3	0.2	0.1620
4	0.1	0.3610
		0.8900

$$\sigma_x^2 = 0.89$$

$$\sigma_x = \sqrt{0.89} = 0.9434$$

$$\sigma_{255+110\mu_x}^2 = (110)^2 \sigma_x^2 = (110)^2 (0.89) = 10769$$

$$\sigma_{255+110\mu_x} = |110| \sigma_x = 110(0.9434) = 103.77$$



Means and Variances for Linear Combinations

If x_1, x_2, \dots, x_n are random variables and a_1, a_2, \dots, a_n are numerical constants, the random variable y defined as

$$y = a_1x_1 + a_2x_2 + \dots + a_nx_n$$

is a linear combination of the x_i 's.



Means and Variances for Linear Combinations

If x_1, x_2, \dots, x_n are random variables with means $\mu_1, \mu_2, \dots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ respectively, and $y = a_1x_1 + a_2x_2 + \dots + a_nx_n$ then

1. $\mu_y = a_1\mu_1 + a_2\mu_2 + \dots + a_n\mu_n$

(True for any random variables regardless of independence.)

2. If x_1, x_2, \dots, x_n are ***independent*** random variables then

$$\sigma_y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \dots + a_n^2\sigma_n^2$$

and

$$\sigma_y = \sqrt{a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \dots + a_n^2\sigma_n^2}$$



Example

A distributor of fruit baskets is going to put 4 apples, 6 oranges and 2 bunches of grapes in his small gift basket. The weights, in ounces, of these items are the random variables x_1 , x_2 and x_3 respectively with means and standard deviations as given in the following table.

	Apples	Oranges	Grapes
Mean μ	8	10	7
Standard deviation σ	0.9	1.1	2

Find the mean, variance and standard deviation of the random variable y = weight of fruit in a small gift basket.

Example continued



It is reasonable in this case to assume that the weights of the different types of fruit are **independent**.

	Apples	Oranges	Grapes
Mean μ	8	10	7
Standard deviation σ	0.9	1.1	2

$$a_1 = 4, a_2 = 6, a_3 = 2, \mu_1 = 8, \mu_2 = 10, \mu_3 = 7$$

$$\sigma_1 = 0.9, \sigma_2 = 1.1, \sigma_3 = 2$$

$$\mu_y = \mu_{a_1x_1+a_2x_2+a_3x_3} = a_1\mu_1 + a_2\mu_2 + a_3\mu_3$$

$$= 4(8) + 6(10) + 2(7) = 106$$

$$\sigma_y^2 = \sigma_{a_1x_1+a_2x_2+a_3x_3}^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + a_3^2\sigma_3^2$$

$$= 4^2(.9)^2 + 6^2(1.1)^2 + 2^2(2)^2 = 72.52$$

$$\sigma_y = \sqrt{72.52} = 8.5159$$



Another Example

Suppose “1 lb” boxes of Sugar Treats cereal have a weight distribution with a mean $\mu_T=1.050$ lbs and standard deviation $\sigma_T=.051$ lbs and “1 lb” boxes of Sour Balls cereal have a weight distribution with a mean $\mu_B=1.090$ lbs and standard deviation $\sigma_B=.087$ lbs. If a promotion is held where the customer is sold a shrink wrapped package containing “1 lb” boxes of both Sugar Treats and Sour Balls cereals, what is the mean and standard deviation for the distribution of promotional packages.



Another Example - continued

$$\mu_T = 1.05 \text{ lb and } \mu_B = 1.09 \text{ lb}$$

Combining these values we get

$$\mu_{T+B} = \mu_T + \mu_B = 1.05 + 1.09 = 2.14 \text{ lb}$$

$$\sigma_T^2 = .051^2 = 0.002601 \text{ and}$$

$$\sigma_B^2 = .087^2 = 0.007569$$

Combining these values we get

$$\sigma_{T+B}^2 = \sigma_T^2 + \sigma_B^2 = 0.01017 \text{ and}$$

$$\sigma_{T+B} = \sqrt{0.01017} = 0.1008 \text{ lb}$$



7.6: Normal Distributions

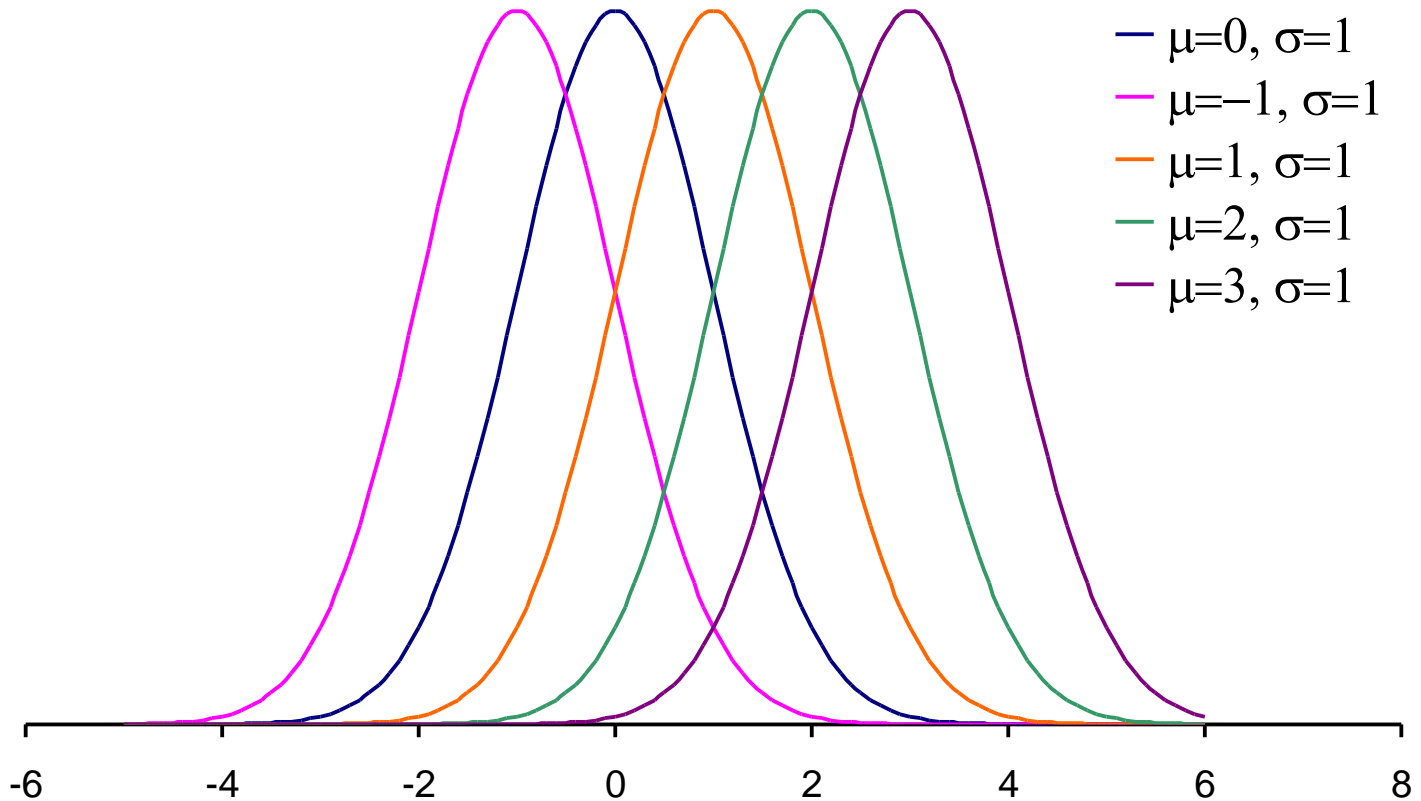
Two characteristic values (numbers)
completely determine a normal distribution

1. Mean - μ
2. Standard deviation - σ



Normal Distributions

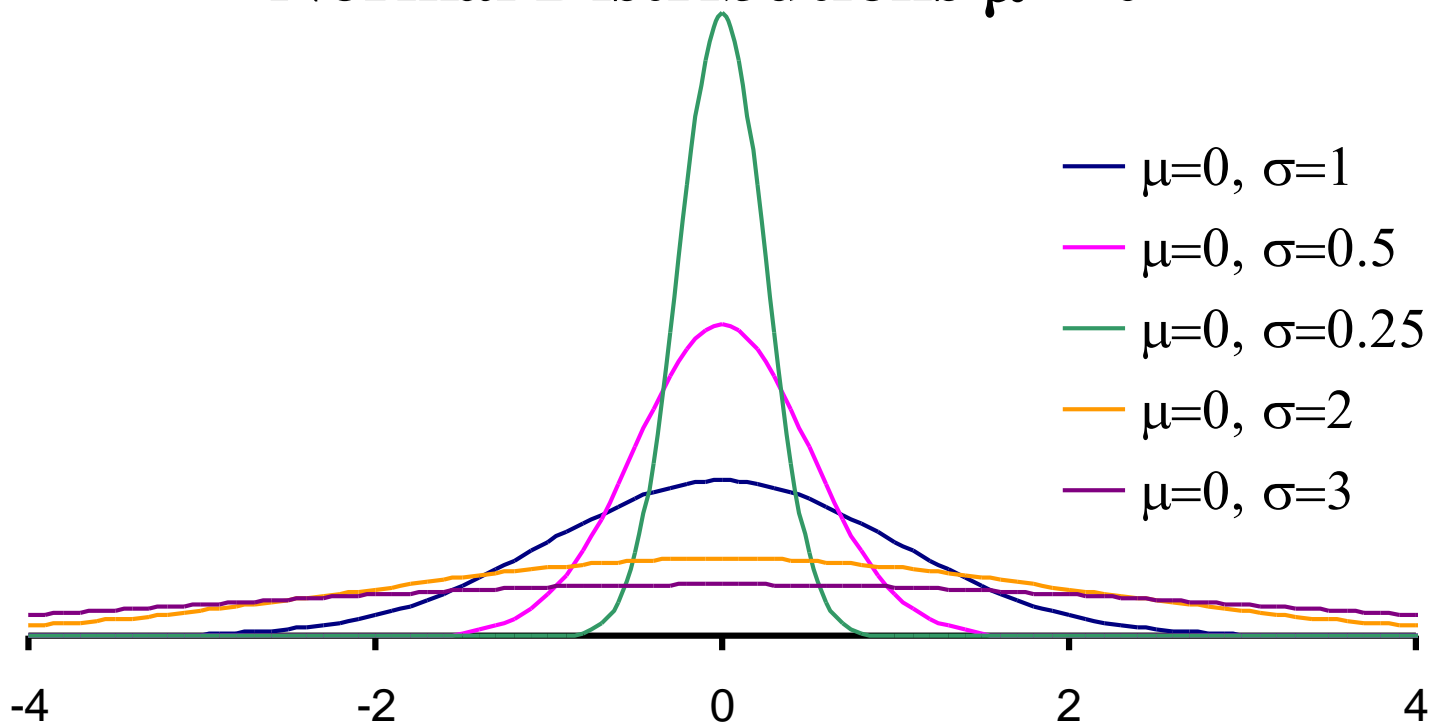
Normal Distributions $\sigma = 1$



Normal Distributions



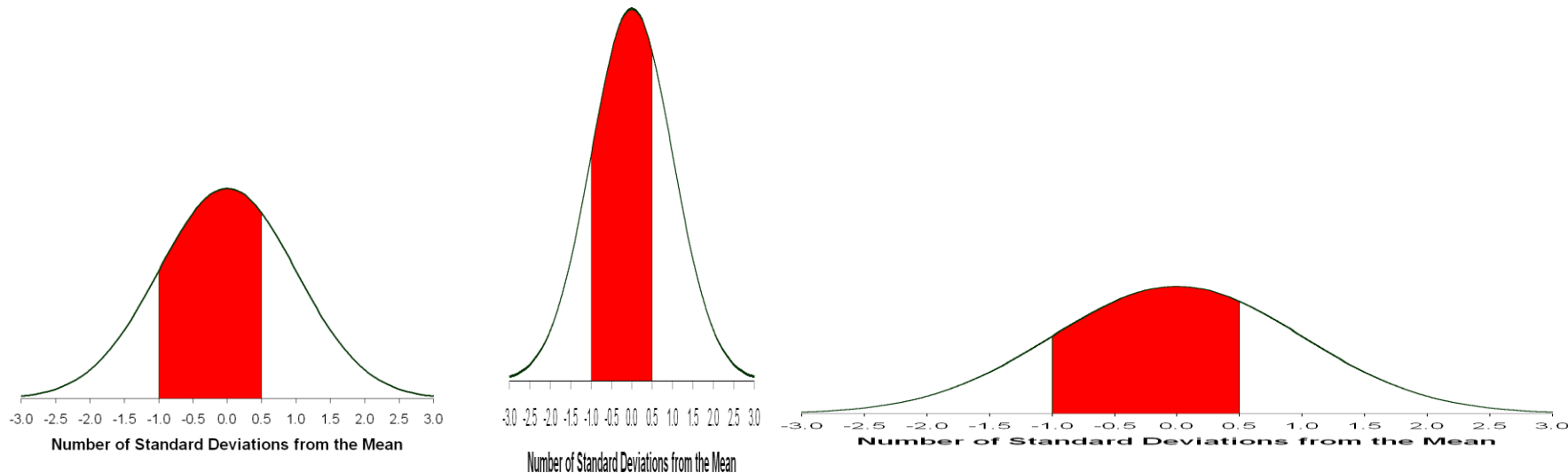
Normal Distributions $\mu = 0$





Major Principle

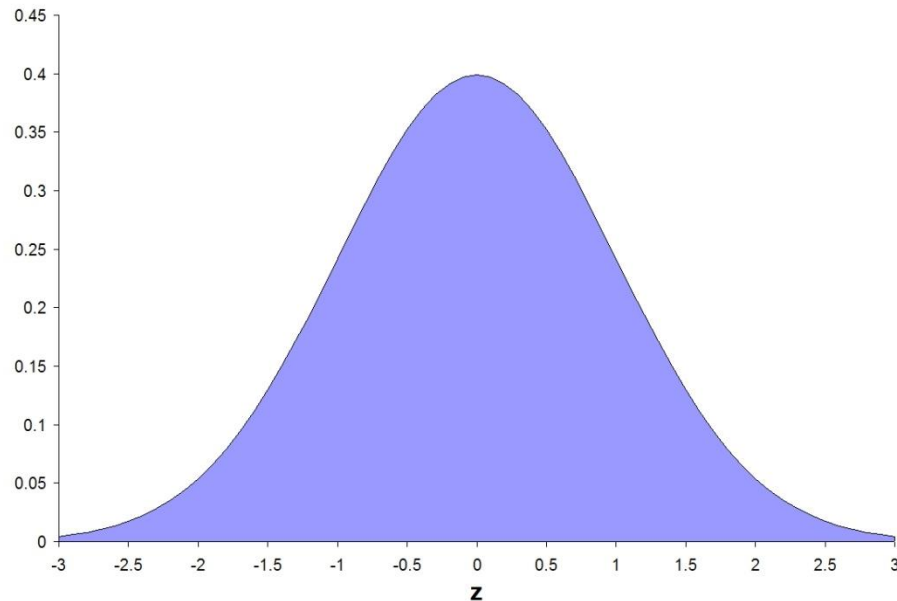
The proportion or percentage of a normally distributed population that is in an interval depends only on how many standard deviations the endpoints are from the mean.





Standard Normal Distribution

A normal distribution with mean 0 and standard deviation 1, is called the standard (or standardized) normal distribution.





Normal Tables

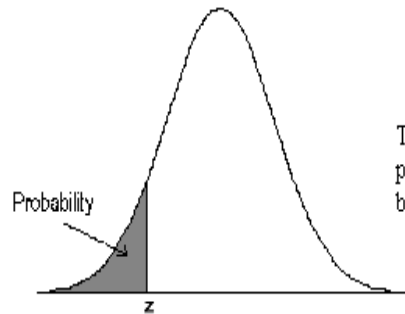


Table entry is
probability at or
below z .

z^*	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-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	0.0010	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	0.0019	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	0.0046	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0061	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
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641



Normal Tables



Table entry is
probability at or
below z .

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
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999



Using the Normal Tables

For any number z^* between -3.89 and 3.89 and rounded to two decimal places,

Appendix Table II gives

(Area under z curve to the left of z^*)

$$= P(z < z^*) = P(z \leq z^*)$$

where the letter z is used to represent a random variable whose distribution is the standard normal distribution



Using the Normal Tables

To find this probability, locate the following:

1. The row labeled with the sign of z^* and the digit to either side of the decimal point
2. The column identified with the second digit to the right of the decimal point in z^*

The number at the intersection of this row and column is the desired probability, $P(z < z^*)$.



Using the Normal Tables

Find $P(z < 0.46)$

Column
labeled
0.06

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

Row
labeled
0.4

$$P(z < 0.46) = 0.6772$$



Using the Normal Tables

Find $P(z < -2.74)$

$$P(z < -2.74) = 0.0031$$

Column
labeled
0.04

z^*	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-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	0.0010	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	0.0019	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	0.0046	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0061	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

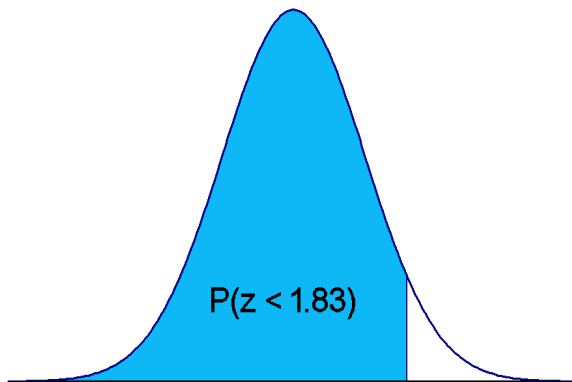
Row
labeled
-2.7



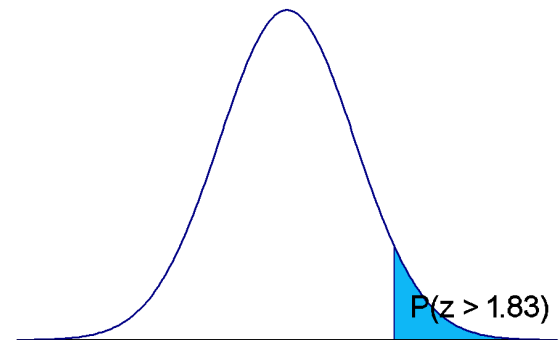
Sample Calculations Using the Standard Normal Distribution

Using the standard normal tables, find the proportion of observations (z values) from a standard normal distribution that satisfy each of the following:

(a) $P(z < 1.83)$
 $= 0.9664$



(b) $P(z > 1.83)$
 $= 1 - P(z < 1.83)$
 $= 1 - 0.9664$
 $= 0.0336$



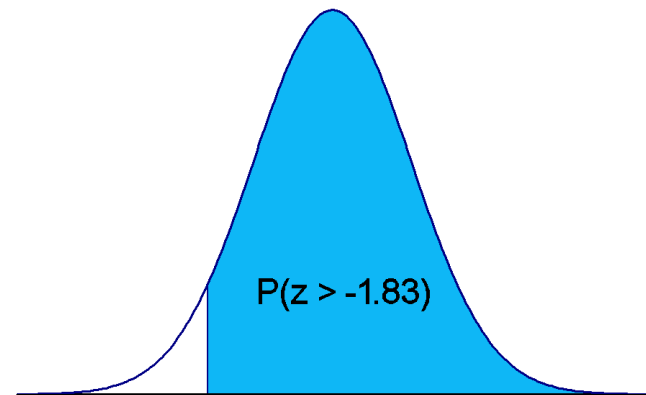
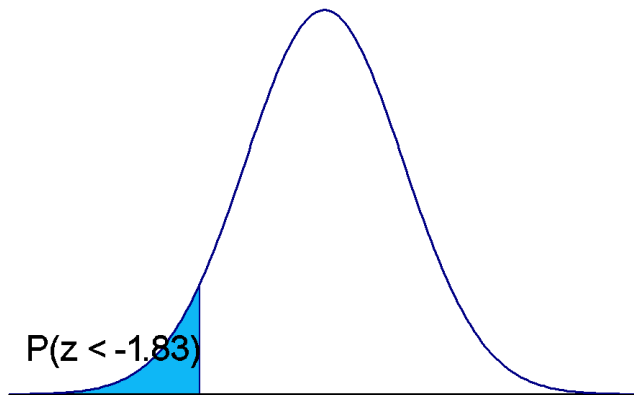


Sample Calculations Using the Standard Normal Distribution

Using the standard normal tables, find the proportion of observations (z values) from a standard normal distribution that satisfies each of the following:

c) $P(z < -1.83) = 0.0336$

(d) $P(z > -1.83)$
 $= 1 - P(z < -1.83)$
 $= 1 - 0.0336 = 0.9664$

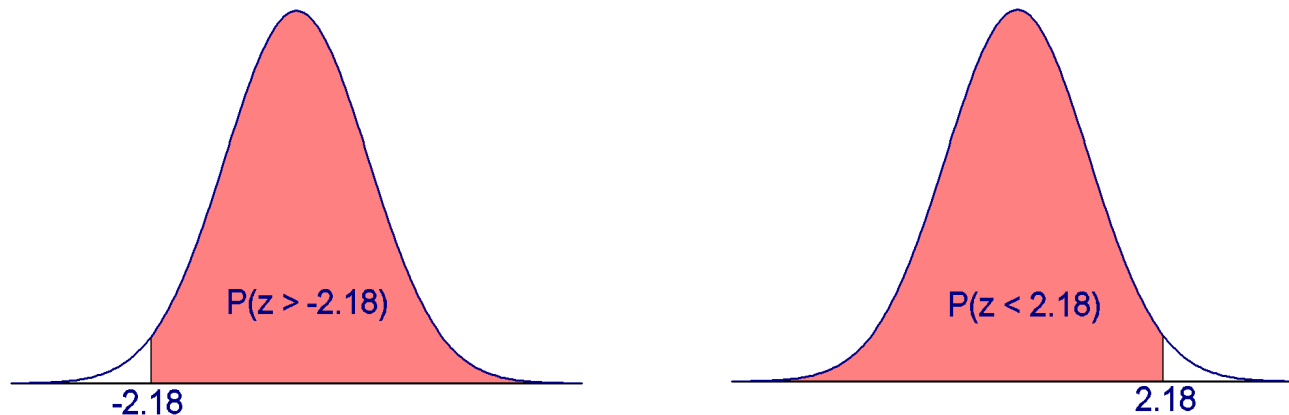




Symmetry Property

Notice from the preceding examples it becomes obvious that

$$P(z > z^*) = P(z < -z^*)$$

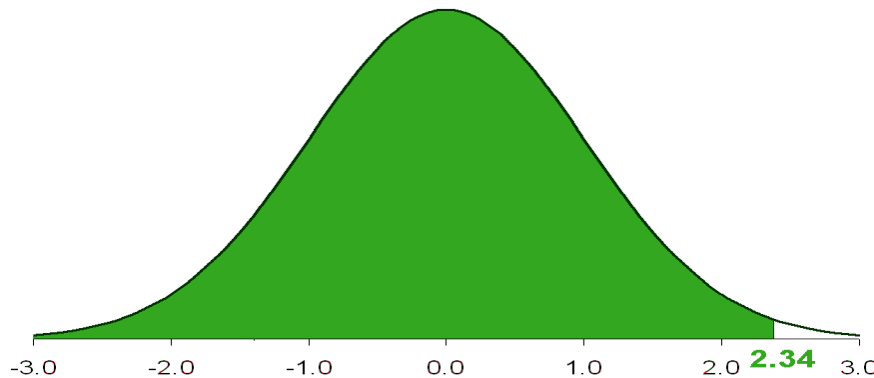


$$P(z > -2.18) = P(z < 2.18) = 0.9854$$



Sample Calculations Using the Standard Normal Distribution

Using the standard normal tables, find the proportion of observations (z values) from a standard normal distribution that satisfies $-1.37 < z < 2.34$, that is find $P(-1.37 < z < 2.34)$.



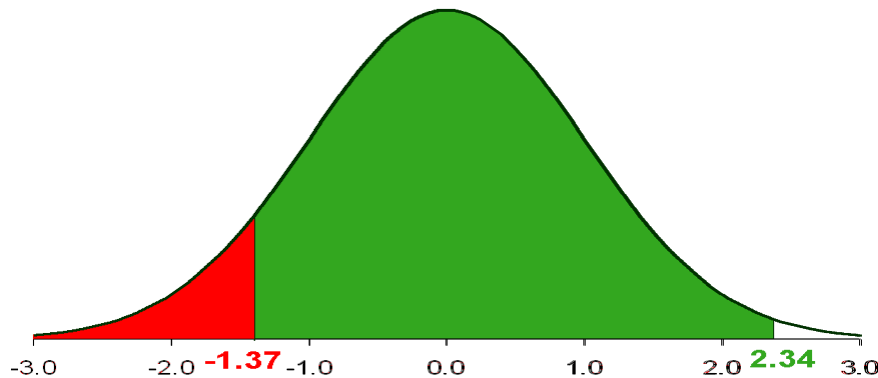
$$P(Z < 2.34) = 0.9904$$

$$P(-1.37 < z < 2.34) = 0.9904$$



Sample Calculations Using the Standard Normal Distribution

Using the standard normal tables, find the proportion of observations (z values) from a standard normal distribution that satisfies $-1.37 < z < 2.34$, that is find $P(-1.37 < z < 2.34)$.



$$P(Z < 2.34) = 0.9904$$

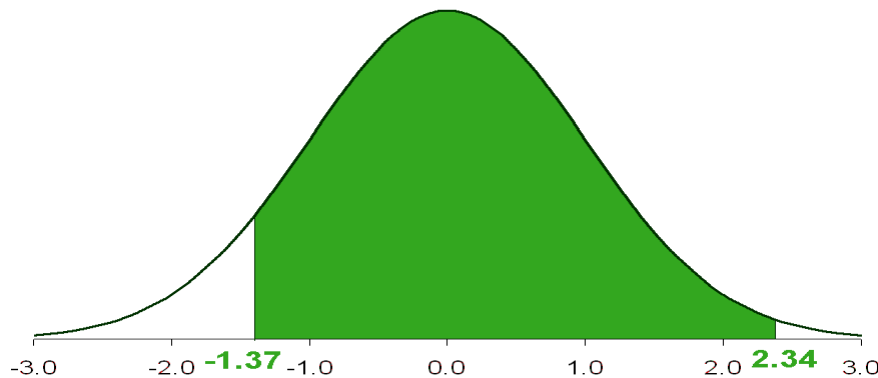
$$P(Z < -1.37) = 0.0853$$

$$P(-1.37 < z < 2.34) = 0.9904 - 0.0853$$



Sample Calculations Using the Standard Normal Distribution

Using the standard normal tables, find the proportion of observations (z values) from a standard normal distribution that satisfies $-1.37 < z < 2.34$, that is find $P(-1.37 < z < 2.34)$.



$$P(Z < 2.34) = 0.9904$$

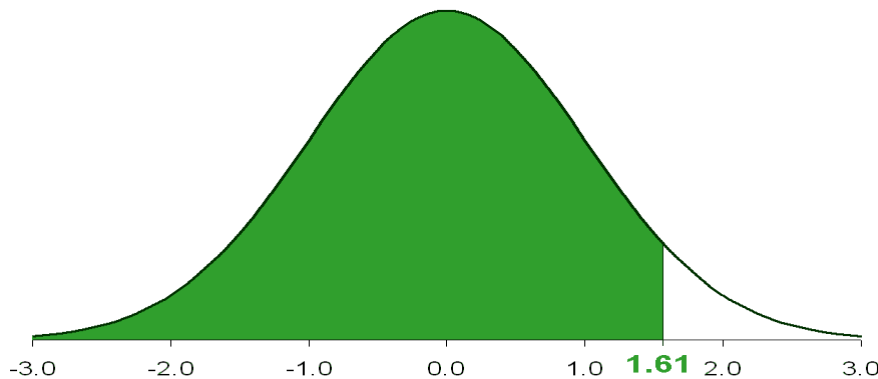
$$P(Z < -1.37) = 0.0853$$

$$P(-1.37 < z < 2.34) = 0.9904 - 0.0853 = 0.9051$$



Sample Calculations Using the Standard Normal Distribution

Using the standard normal tables, find the proportion of observations (z values) from a standard normal distribution that satisfies $0.54 < z < 1.61$, that is find $P(0.54 < z < 1.61)$.



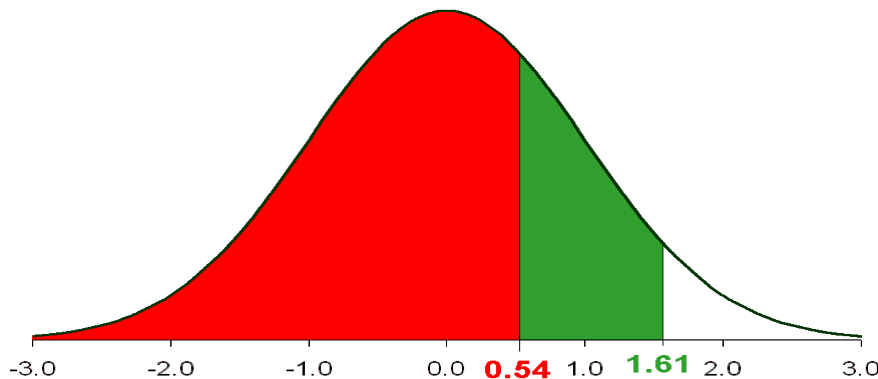
$$P(Z < 1.61) = 0.9463$$

$$P(0.54 < z < 1.61) = 0.9463$$



Sample Calculations Using the Standard Normal Distribution

Using the standard normal tables, find the proportion of observations (z values) from a standard normal distribution that satisfies $0.54 < z < 1.61$, that is find $P(0.54 < z < 1.61)$.



$$P(Z < 1.61) = 0.9463$$

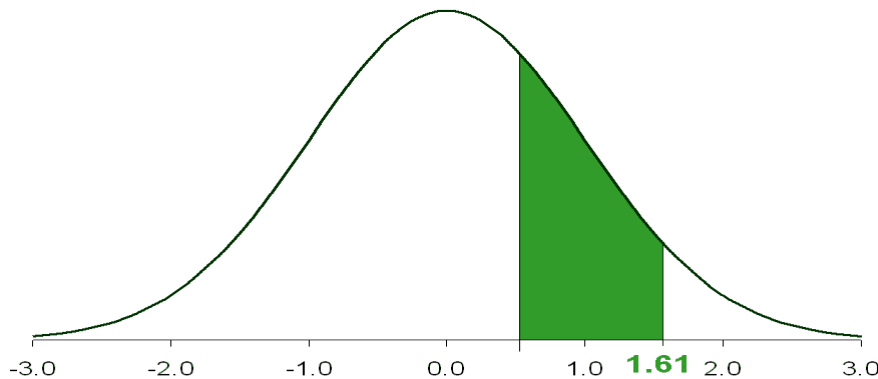
$$P(Z < .54) = 0.7054$$

$$P(0.54 < z < 1.61) = 0.9463 - 0.7054$$



Sample Calculations Using the Standard Normal Distribution

Using the standard normal tables, find the proportion of observations (z values) from a standard normal distribution that satisfies $0.54 < z < 1.61$, that is find $P(0.54 < z < 1.61)$.



$$P(Z < 1.61) = 0.9463$$

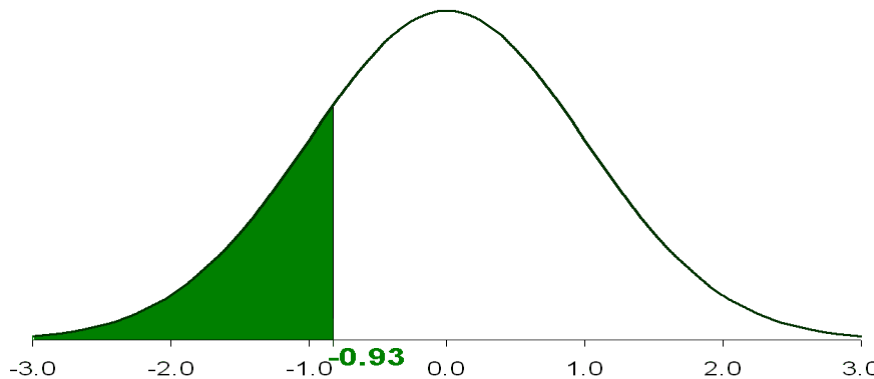
$$P(Z < .54) = 0.7054$$

$$P(0.54 < z < 1.61) = 0.9463 - 0.7054 = 0.2409$$



Sample Calculations Using the Standard Normal Distribution

Using the standard normal tables, find the proportion of observations (z values) from a standard normal distribution that satisfy $-1.42 < z < -0.93$, that is find $P(-1.42 < z < -0.93)$.



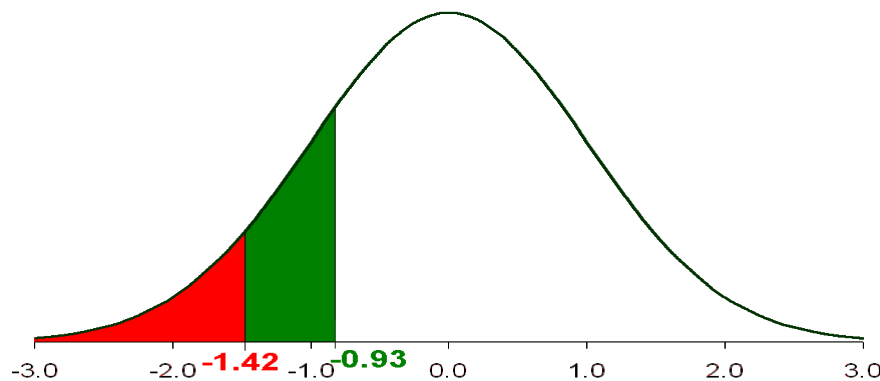
$$P(Z < -0.93) = 0.1762$$

$$P(-1.42 < z < -0.93) = 0.1762$$



Sample Calculations Using the Standard Normal Distribution

Using the standard normal tables, find the proportion of observations (z values) from a standard normal distribution that satisfy $-1.42 < z < -0.93$, that is find $P(-1.42 < z < -0.93)$.



$$P(Z < -0.93) = 0.1762$$

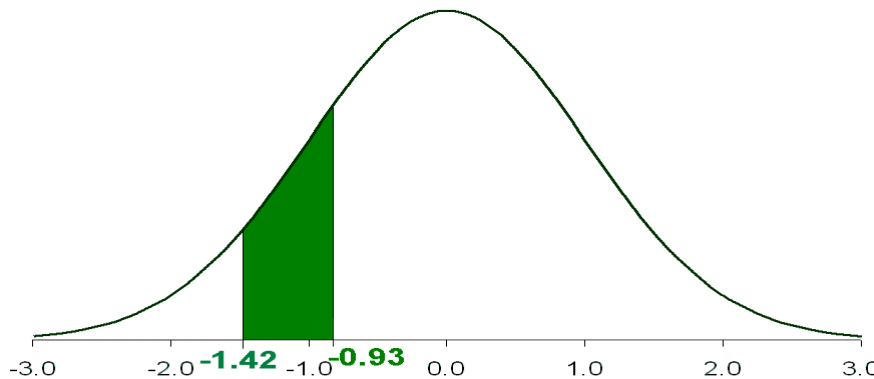
$$P(Z < -1.42) = 0.0778$$

$$P(-1.42 < z < -0.93) = 0.1762 - 0.0778$$



Sample Calculations Using the Standard Normal Distribution

Using the standard normal tables, find the proportion of observations (z values) from a standard normal distribution that satisfy $-1.42 < z < -0.93$, that is find $P(-1.42 < z < -0.93)$.



$$P(Z < -0.93) = 0.1762$$

$$P(Z < -1.42) = 0.0778$$

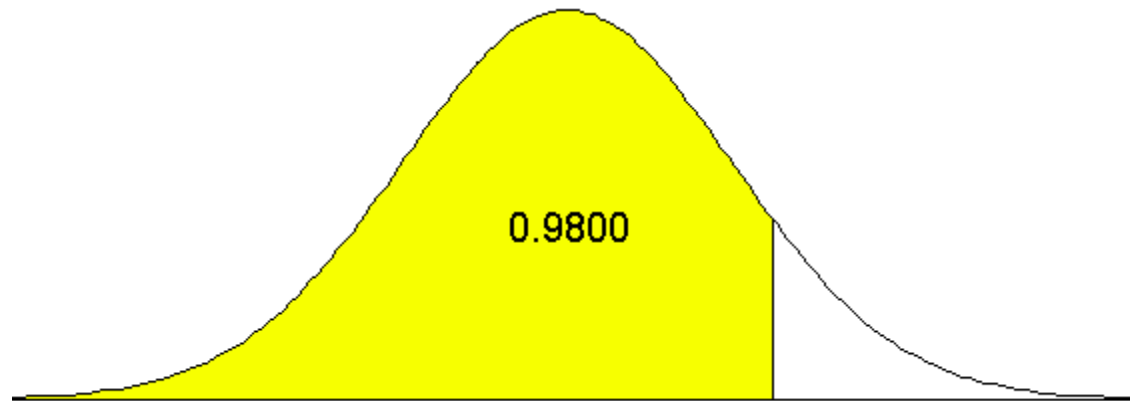
$$P(-1.42 < z < -0.93) = 0.1762 - 0.0778 = 0.0984$$



Example Calculation

Using the standard normal tables, in each of the following, find the z values that satisfy :

- (a) The point z with 98% of the observations falling below it.



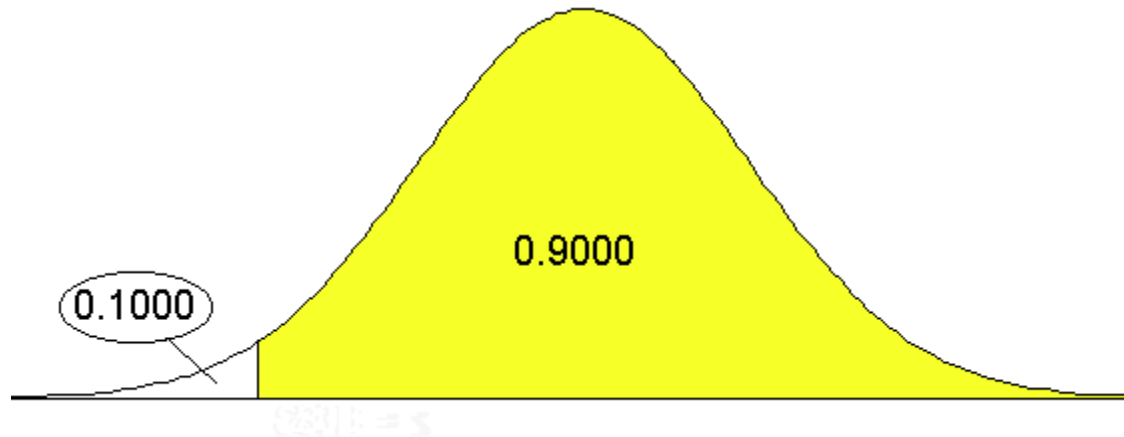
The closest entry in the table to 0.9800 is 0.9798 corresponding to a z value of 2.05



Example Calculation

Using the standard normal tables, in each of the following, find the z values that satisfy :

- (b) The point z with 90% of the observations falling above it.



The closest entry in the table to 0.1000 is 0.1003 corresponding to a z value of -1.28



Finding Normal Probabilities

To calculate probabilities for any normal distribution, standardize the relevant values and then use the table of z curve areas. More specifically, if x is a variable whose behavior is described by a normal distribution with mean m and standard deviation s , then

$$P(x < b) = p(z < b^*)$$

$$P(x > a) = P(a^* < z) = P(z > a^*)$$

where z is a variable whose distribution is standard normal and

$$a^* = \frac{a - \mu}{\sigma} \qquad b^* = \frac{b - \mu}{\sigma}$$



Standard Normal Distribution Revisited

If a variable X has a normal distribution with mean μ and standard deviation σ , then the standardized variable

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

has the normal distribution with mean 0 and standard deviation 1.

This is called the **standard normal distribution**.



Conversion to $N(0,1)$

The formula $z = \frac{x - \mu}{\sigma}$

gives the number of standard deviations that x is from the mean.

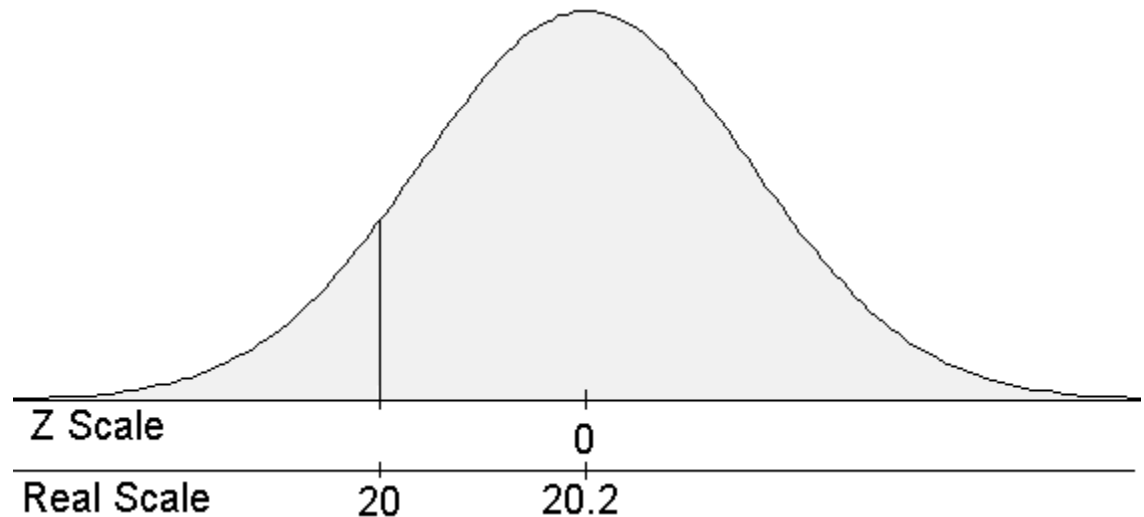
Where μ is the true population mean and σ is the true population standard deviation



Example 1

A Company produces “20 ounce” jars of a picante sauce. The true amounts of sauce in the jars of this brand sauce follow a normal distribution.

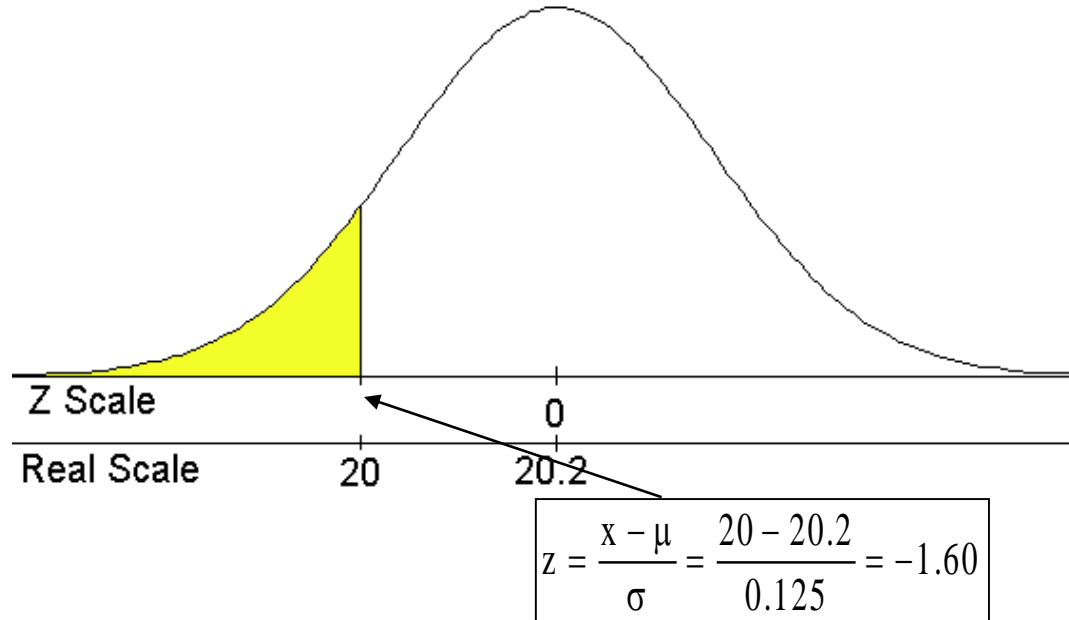
Suppose the companies “20 ounce” jars follow a $N(20.2, 0.125)$ distribution curve. (i.e., The contents of the jars are normally distributed with a true mean $\mu = 20.2$ ounces with a true standard deviation $\sigma = 0.125$ ounces.





Example 1

What proportion of the jars are under-filled (i.e., have less than 20 ounces of sauce)?



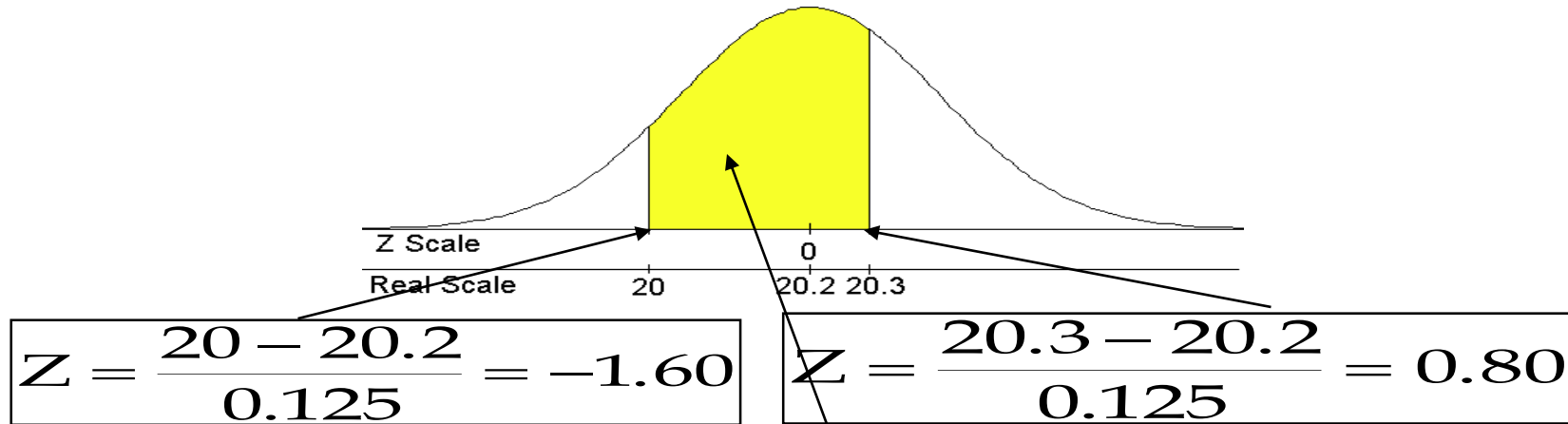
Looking up the z value -1.60 on the standard normal table we find the value 0.0548.

The proportion of the sauce jars that are under-filled is 0.0548 (i.e., 5.48% of the jars contain less than 20 ounces of sauce).



Example 1

What proportion of the sauce jars contain between 20 and 20.3 ounces of sauce.



Looking up the z values of -1.60 and 0.80 we find the areas (proportions) 0.0548 and 0.7881

The resulting difference

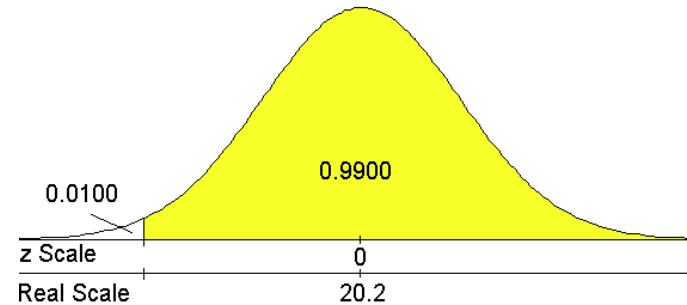
$$0.7881 - 0.0548 = 0.7333$$

is the proportion of the jars that contain between 20 and 20.3 ounces of sauce.

Example 1



99% of the jars of this brand of picante sauce will contain more than what amount of sauce?



When we try to look up 0.0100 in the body of the table we do not find this value. We do find the following

z	0.00	0.01	0.02	0.03	0.04	0.05
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071
-2.3	0.0107	0.0105	0.0102	0.0099	0.0096	0.0094
-2.2	0.0139	0.0135	0.0132	0.0129	0.0125	0.0122

The entry closest to 0.0100 is 0.0099 corresponding to the z value -2.33

Since the relationship between the real scale (x) and the z scale is $z = (x - \mu) / \sigma$ we solve for x getting $x = \mu + z\sigma$

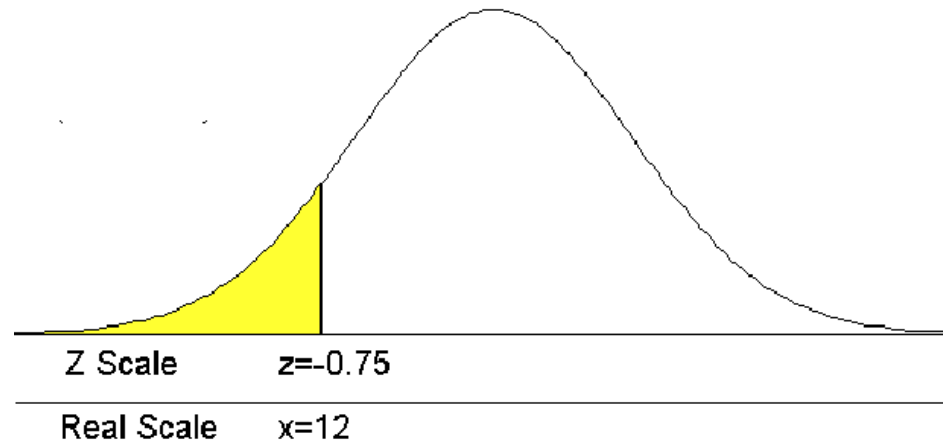
$$x = 20.2 + (-2.33)(0.125) = 19.90875 = 19.91$$



Example 2

The weight of the cereal in a box is a random variable with mean 12.15 ounces, and standard deviation 0.2 ounce.

What percentage of the boxes have contents that weigh under 12 ounces?



$$\begin{aligned} P(X < 12) &= P\left(Z < \frac{x - \mu}{\sigma}\right) = P\left(Z < \frac{12 - 12.15}{0.2}\right) \\ &= P(Z < -0.75) = 0.2266 \end{aligned}$$



Example 2

If the manufacturer claims there are 12 ounces in a box, does this percentage cause concern?

If so, what could be done to correct the situation?

- The machinery could be reset with a higher or larger mean.
- The machinery could be replaced with machinery that has a smaller standard deviation of fills.
- The label on the box could be changed.

Which would probably be cheaper immediately but might cost more in the long run?

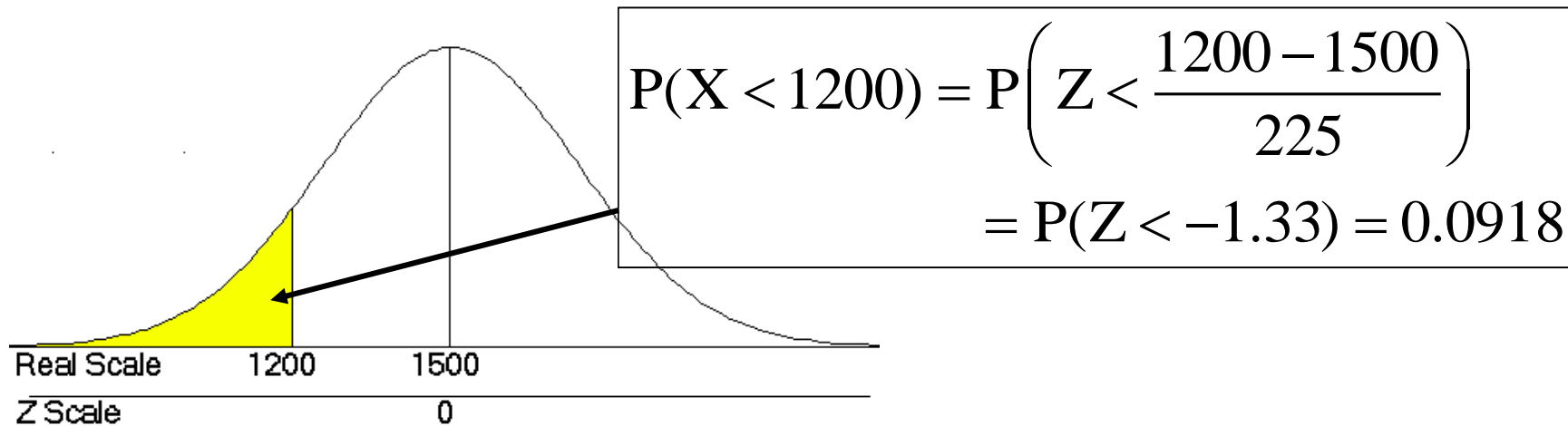
- In the long run it might be cheaper to get newer more precise equipment and not give away as much excess cereal.



Example 3

The time to first failure of a unit of a brand of ink jet printer is approximately normally distributed with a mean of 1,500 hours and a standard deviation of 225 hours.

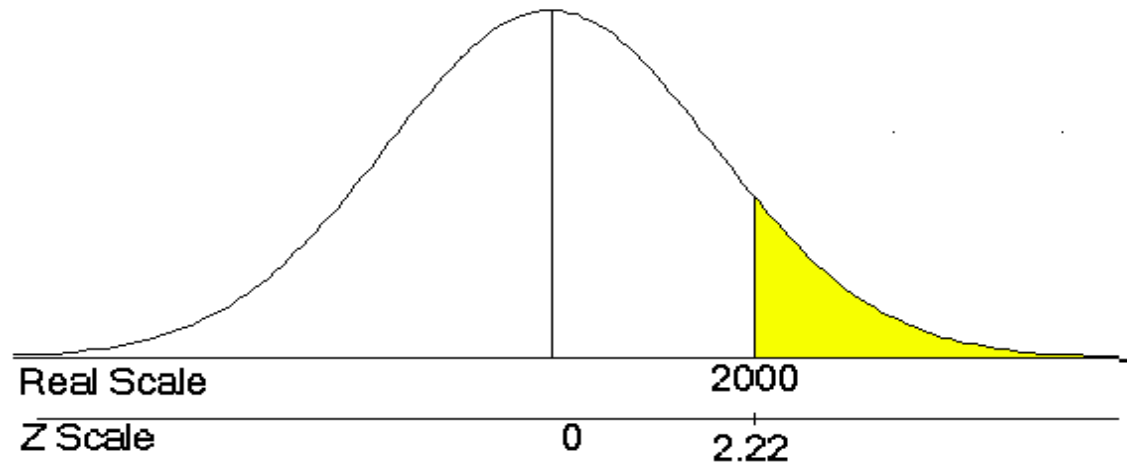
- a) What proportion of these printers will fail before 1,200 hours?





Example 3

- b) What proportion of these printers will not fail within the first 2,000 hours?

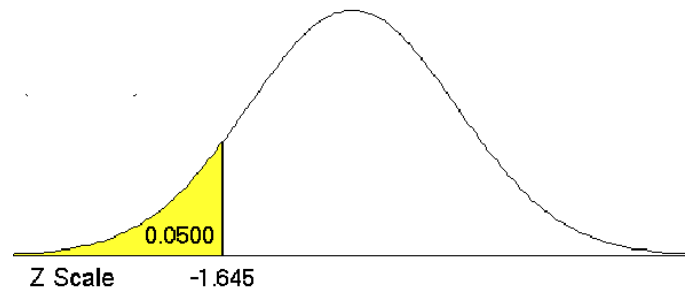


$$\begin{aligned} P(X > 2000) &= P\left(Z > \frac{2000 - 1500}{225}\right) \\ &= P(Z > 2.22) = 1 - 0.9868 \\ &= 0.0132 \end{aligned}$$



Example 3

- c) What should be the guarantee time for these printers if the manufacturer wants only 5% to fail within the guarantee period?



Notice that when you look for a proportion of 0.0500 in the body of the normal table you do not find it. However, you do find the value 0.0505 corresponding to a z score of -1.64 and value 0.0495 corresponding to the z score of -1.65. Since 0.0500 is exactly 1/2 way between -1.64 and -1.65, the z value we want is -1.645.



Example 3

The relationship connecting x , z , μ and σ is

$z = \frac{x - \mu}{\sigma}$. If you solve this equation for x you get $x = \mu + z \sigma$, so

$$x = \mu + z \sigma = 1500 + (-1.645)(225) = 1129.875$$

The guarantee period should be 1130 hours.



7.7: Checking for Normality

A **normal probability plot** is a scatter plot of the (normal score*, observation) pairs.

A substantially linear pattern in a normal probability plot suggests that population normality is plausible.

A systematic departure from a straight-line pattern (such as curvature in the plot) casts doubt on the legitimacy of assuming a normal population distribution. Can also use a box & whisker plot

*more on normal scores on the next slide.



Checking Normality

There are different techniques for determining the normal scores. Most statisticians use a statistical package such as Minitab, SPSS or SAS to create a normal probability plot. Graphing calcs have built in function to calculate. Even if Normal scores are different, overall look of the curve, nor correlation coefficient using normal scores will change significantly (book pg 409)

Bk	-1.539	-1.001	-.656	-.376	-.123	.123	.376	.656	1.001	1.539
TI	-1.644	-1.036	-.674	-.385	-.126	.126	.385	.674	1.036	1.644

Generate on calc: Seq(Invnorm(x),x,(1/(n+1)),((n+1)/(n+1)), (1/(n+1))) -> L#



Checking Normality

Calculate the regression between the normal score & observed score. If $r < \text{the critical } r$ for the corresponding n in the table below, considerable doubt is cast on the assumption of normality.

n	5	10	15	20	25	30	40	50	60	75
crit r	.832	.880	.911	.929	.941	.949	.960	.966	.971	.976

Table 7.2 in book, pg 412



Normal Probability Plot Example

Ten randomly selected shut-ins were each asked to list how many hours of television they watched per week. The results are

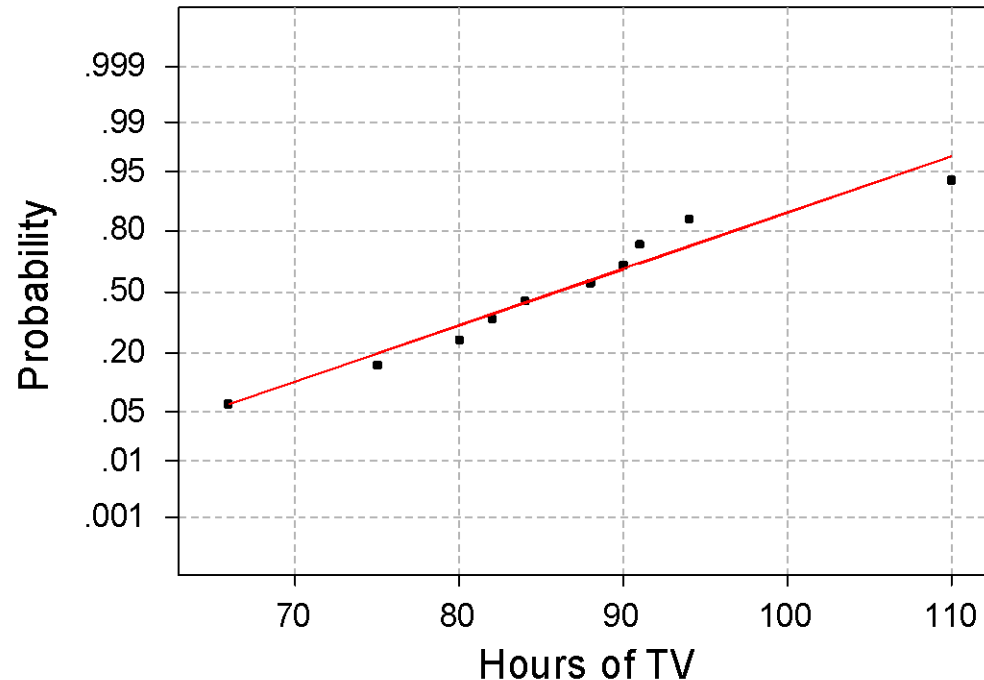
82	66	90	84	75
88	80	94	110	91

The normal probability plot is on the following slide.



Normal Probability Plot Example

Normal Probability Plot



Average: 86
StDev: 11.8415
N: 10

$r = 0.968$

Anderson-Darling Normality Test
A-Squared: 0.226
P-Value: 0.753

Notice that the points all fall nearly on a line so it is reasonable to assume that the population of hours of TV watched by shut-ins is normally distributed. The crit r with $n = 10$ is 0.88, so this verifies it seems normal.



Normal Probability Plot Example

A sample of times of 15 telephone solicitation calls (in seconds) was obtained and is given below.

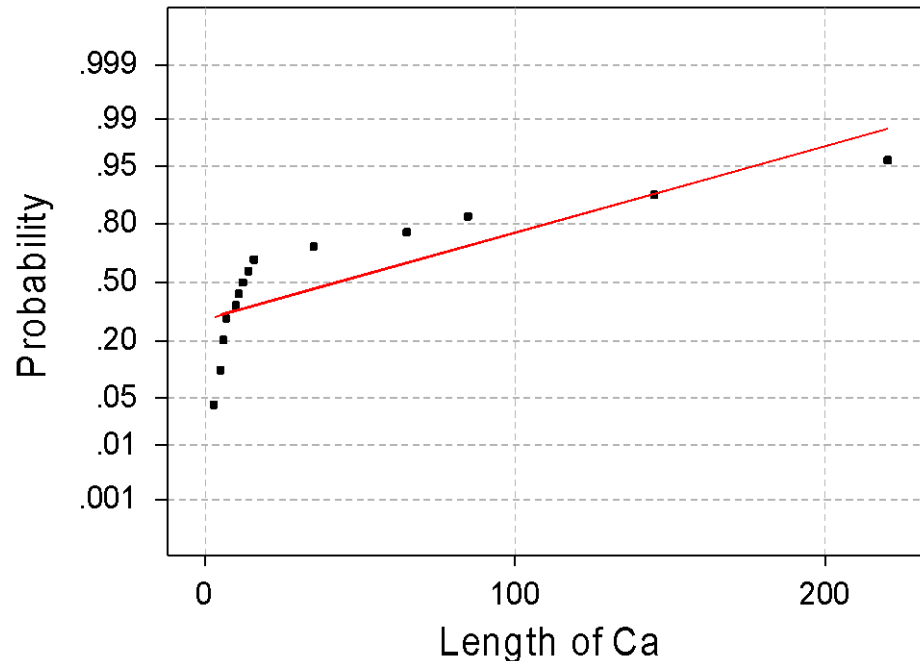
5	10	7	12	35
65	145	14	3	220
11	6	85	6	16

The normal probability plot follows on the next slide



Normal Probability Plot Example

Normal Probability Plot



Average: 42.6667
StDev: 63.1129
N: 15

$r = 0.814$

Anderson-Darling Normality Test
A-Squared: 2.087
P-Value: 0.000

Clearly the points do not fall nearly on a line. Specifically the pattern has a distinct nonlinear appearance. It is **NOT** reasonable to assume that the population of length of solicitation calls is normally distributed. The crit r with $n = 15$ is 0.911, so $r = .814$ also cast doubt on normality.

Checking Normality



Often, one needs to transform the original data set to obtain a more “Normal” data set for which parametric statistics can be performed.

Remember, most of the statistical calculations we are going to learn in this class (parametric statistics) have the assumption that the data are sampled from a population that is “Normal”.

If one must transform, use what was covered in unit 5.4



7.5: The Binomial Distribution

Properties of a Binomial Experiment

1. It consists of a fixed number of observations called trials.
2. Each trial can result in one of only two mutually exclusive outcomes labeled success (S) and failure (F).
3. Outcomes of different trials are independent.
4. The probability that a trial results in S is the same for each trial.

The **binomial random variable** x is defined as

x = number of successes observed when
experiment is performed

The probability distribution of x is called the **binomial probability distribution**.



The Binomial Distribution

Let

n = number of independent trials in a binomial experiment

π = constant probability that any particular trial results in a success.

Then

$P(x) = P(x \text{ successes among } n \text{ trials})$

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



Example 1

On the average, 1 out of 19 people will respond favorably to a certain telephone solicitation. If 25 people are called,

- a) What is the probability that exactly two will respond favorably to this sales pitch?

$$n = 25, \pi = \frac{1}{19}$$

$$p(2) = \frac{25!}{2!23!} \left(\frac{1}{19}\right)^2 \left(\frac{18}{19}\right)^{23} = 0.2396$$



Example 1 continued

On the average, 1 out of 19 people will respond favorably to a certain telephone sales pitch. If 25 people are called,

- b) What is the probability that at least two will respond favorably to this sales pitch?

$$\begin{aligned} P(x \geq 2) &= 1 - P(x < 2) = 1 - p(0) - p(1) \\ &= 1 - \frac{25!}{0!25!} \left(\frac{1}{19}\right)^0 \left(\frac{18}{19}\right)^{25} - \frac{25!}{1!24!} \left(\frac{1}{19}\right)^1 \left(\frac{18}{19}\right)^{24} \\ &= 1 - 0.2588 - 0.3595 = 0.3817 \end{aligned}$$



Example 2

The adult population of a large urban area is 60% African American. If a jury of 12 is randomly selected from the adults in this area, what is the probability that less than 3 are African American.

Clearly, $n = 12$ and $\pi = 0.6$, so

$$\begin{aligned} P(x < 3) &= P(x \leq 2) = p(0) + p(1) + p(2) \\ &= \frac{12!}{0!12!} (.6)^0 (.4)^{12} + \frac{12!}{1!11!} (.6)^1 (.4)^{11} + \frac{12!}{2!10!} (.6)^2 (.4)^{10} \\ &= 0.00002 + 0.00031 + 0.00249 = 0.00281 \end{aligned}$$



Another Version of the Binomial Formula

$$p(2) = \frac{25!}{2!23!} \left(\frac{1}{19}\right)^2 \left(\frac{18}{19}\right)^{23} = 0.2396$$

Can be computed as:

$$p(2) = (25 \text{ C } 2) (1/19)^2 (18/19)^{23} = 0.2396$$

(25 C 2) is the combinations function on your calculator found under MATH → PRB



Using the Binomial tables

On the average, 3 out of 10 MST students will take AP Stats. If 5 MST students are surveyed, what is the probability that exactly two signed up for AP stats?

n = 5

π

pg 864 in book

x	0.05	0.1	0.2	0.25	0.3	0.4	...
0	.774	.590	.328	.237	.168	.078	...
1	.203	.329	.409	.396	.360	.259	...
2	.022	.072	.205	.263	.309	.346	...
3	.001	.009	.051	.088	.132	.230	...
4	.000	.000	.007	.015	.029	.077	...
5	.000	.000	.000	.001	.002	.010	...



Mean & Standard Deviation of a Binomial Random Variable

The mean value and the standard deviation of a binomial random variable are, respectively,

$$\mu_X = n\pi$$

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



Example

A professor routinely gives quizzes containing 50 multiple choice questions with 4 possible answers, only one being correct.

Occasionally he just hands the students an answer sheet without giving them the questions and asks them to guess the correct answers.

Let x be a random variable defined by

x = number of correct answers on such an exam

Find the mean and standard deviation for x



Example - solution

The random variable is clearly binomial with $n = 50$ and $p = \frac{1}{4}$. The mean and standard deviation of x are

$$\mu_x = n\pi = 50\left(\frac{1}{4}\right) = 12.5$$

$$\sigma_x = \sqrt{50\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)} = \sqrt{9.375} = 3.06$$



The Geometric Distribution

Suppose an experiment consists of a sequence of trials with the following conditions:

1. The trials are independent.
2. Each trial can result in one of two possible outcomes, success and failure.
3. The probability of success is the same for all trials.

A **geometric random variable** is defined as
 x = number of trials until the first success is observed (including the success trial)

The probability distribution of x is called the **geometric probability distribution**.



The Geometric Distribution

If x is a geometric random variable with probability of success $= \pi$ for each trial, then

$$p(x) = (1 - \pi)^{x-1} \pi \quad x = 1, 2, 3, \dots$$



Example

Over a very long period of time, it has been noted that on Friday's 25% of the customers at the drive-in window at the bank make deposits.

What is the probability that it takes 4 customers at the drive-in window before the first one makes a deposit.



Example - solution

This problem is a geometric distribution problem with $\pi = 0.25$.

Let x = number of customers at the drive-in window before a customer makes a deposit.

The desired probability is

$$p(4) = (.75)^{4-1}(.25) = 0.105$$



Calculator Short Cuts

<code>normalpdf(x[, μ, σ])</code>	gives p @ x
<code>normalcdf(low, high[, μ, σ])</code>	gives p ($low \leq x \leq high$)
<code>invNorm(area [, μ, σ])</code>	gives x for a given cumulative area
<code>binompdf(numtrials, π[, x])</code>	gives p @ x
<code>binomcdf(numtrials, π[, x])</code>	gives $p \leq x$
<code>geometpdf(π, x)</code>	gives p 1 st occur @ x
<code>geometcdf(π, x)</code>	gives p 1 st occur $\leq x$

[] are optional inputs, without [, μ , σ] in normal, it computes x as a z score. Without [, x] the binomial function returns the p for each point or cumulative p for each point (i.e if numtrial is 25, it returns a p for each possible trial)

Under “APPS” turn on “CtlgHelp” to give input order