Lecture 5

The Normal Probability Distribution

Continuous Random Variables

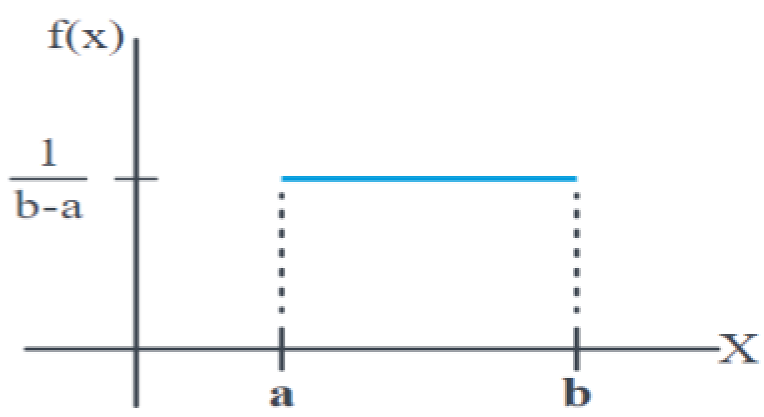
- Continuous random variables can assume the infinitely many values corresponding to points on a line interval.
- **Examples:**
  - Heights, weights
  - length of life of a particular product
  - experimental laboratory error

Continuous Uniform Distributions

The uniform distribution is a continuous probability distribution and is concerned with events that are equally likely to occur. A continuous random variable X has a **uniform distribution**, denoted **U(a,b)**, if its probability density function is:

$$f(x) = \frac{1}{b-a}$$

for two constants a and b, such that $a < x < b$. A graph of the p.d.f. looks like this:



Continuous Uniform Distributions

Cumulative distribution Function of a Uniform Random Variable X

The cumulative distribution function of a uniform random variable X is:

$$F(x) = P(X \leq x) = \frac{x - a}{b - a}$$

Properties:

Mean: $\frac{1}{2}(a + b)$

Variance: $\frac{1}{12}(b - a)^2$

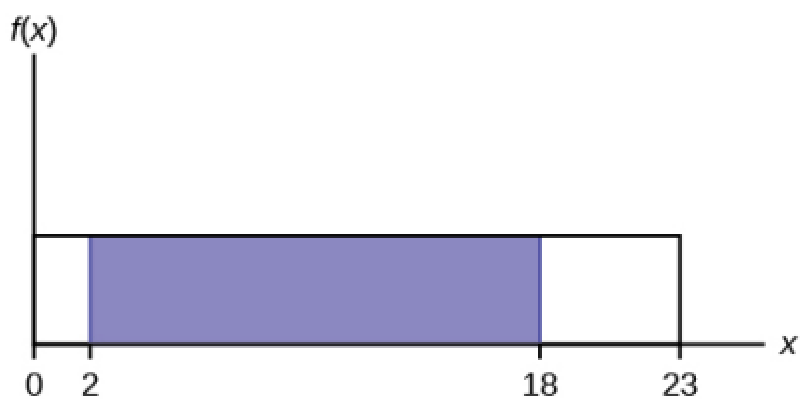
Continuous Uniform Distributions

Suppose the smiling times, in seconds, follow a uniform distribution between zero and 23 seconds, inclusive. This means that any smiling time from zero to and including 23 seconds is equally likely. Let X = length, in seconds, of an eight-week-old baby's smile.

a. What is the probability that a randomly chosen eight-week-old baby smiles between two and 18 seconds?

$$P(2 < x < 18) = P(X \leq 8) - P(X \leq 2) = \frac{18-0}{23-0} - \frac{2-0}{23-0} = 0.696$$

OR, $$P(2<x<18) = (\text{base})(\text{height}) = (18-2)\left(\frac{1}{23}\right) = \frac{16}{23} = 0.696$$



Continuous Uniform Distributions

Suppose the smiling times, in seconds, follow a uniform distribution between zero and 23 seconds, inclusive. This means that any smiling time from zero to and including 23 seconds is equally likely. Let X = length, in seconds, of an eight-week-old baby's smile.

b. Find the average and standard deviation for an eight-week-old baby's smiling time.

Mean: $\frac{1}{2}(a+b) = 11.5$

Variance: $\frac{1}{12}(b-a)^2 = 44.08333$

Standard deviation = 6.639528

Continuous Uniform Distributions

Suppose the smiling times, in seconds, follow a uniform distribution between zero and 23 seconds, inclusive. This means that any smiling time from zero to and including 23 seconds is equally likely. Let X = length, in seconds, of an eight-week-old baby's smile.

c. Find the 90th percentile for an eight-week-old baby's smiling time.

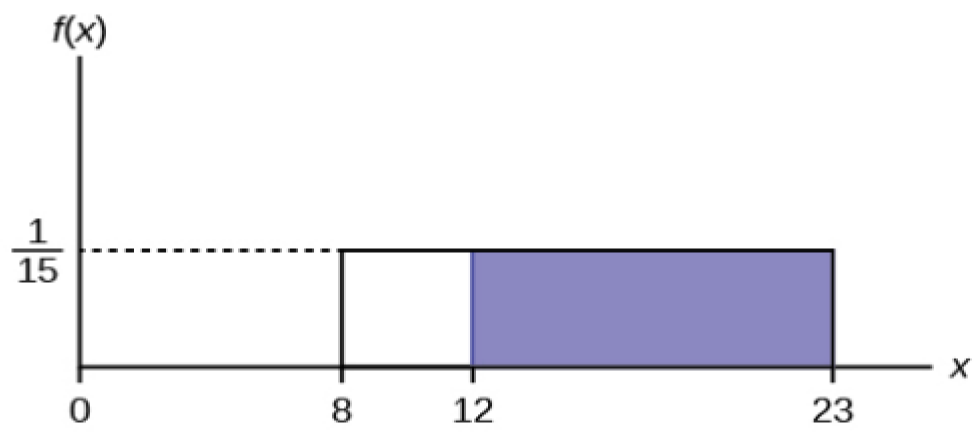
$$0.90 = \frac{x - 0}{23}$$

$$x = 20.7 \text{ seconds.}$$

d. Find the probability that a random eight-week-old baby smiles more than 12 seconds **knowing** that the baby smiles **more than eight seconds**.

Use the fact that this is a **conditional** and changes the sample space. The graph illustrates the new sample space. You already know the baby smiled more than eight seconds. Write a new $f(x)$:

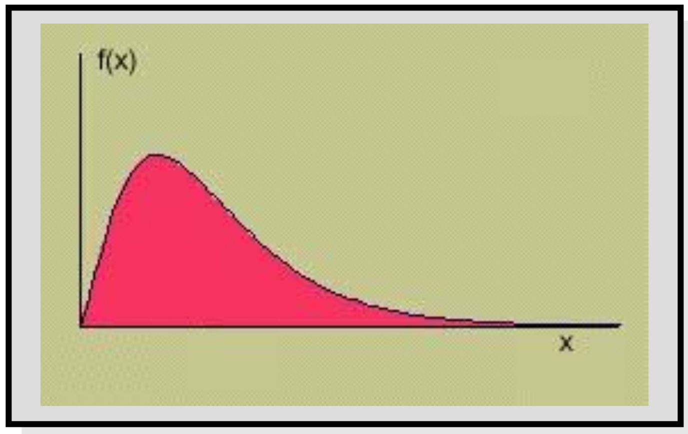
$$f(x) = \frac{1}{23-8} = \frac{1}{15}$$



$$P(x \geq 12 | x \geq 8) = (x) = (23-12)\frac{1}{15} = \frac{11}{15} = 0.73$$

Continuous Random Variables

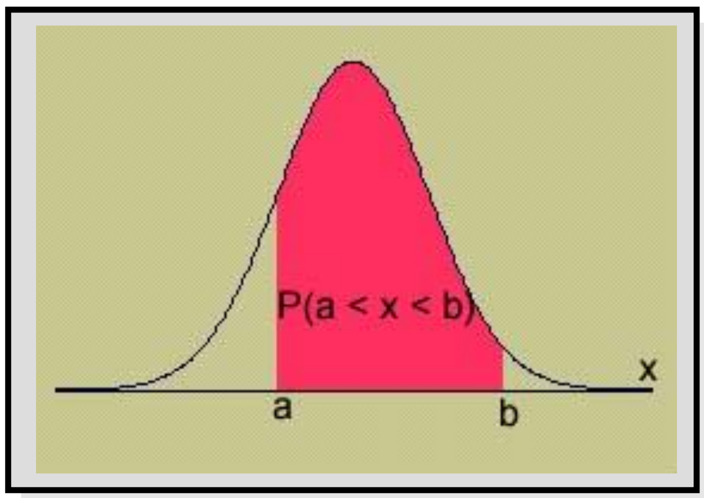
- A **smooth curve** describes the probability distribution of a continuous random variable.



- The depth or density of the probability, which varies with *x*, may be described by a mathematical formula $f(x)$, called the **probability distribution** or **probability density function** for the random variable *x*.

Properties of Continuous Probability Distributions

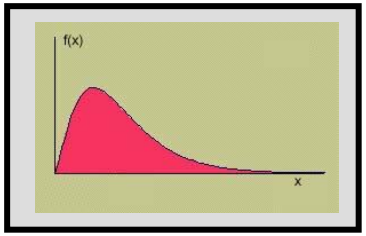
- The area under the curve is equal to **1.**
- $P(a \leq x \leq b)$ = **area under the curve** between a and b.



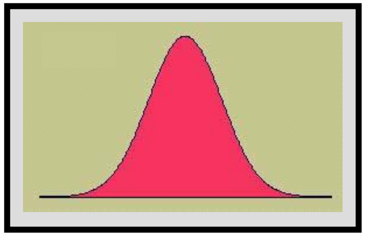
- There is no probability attached to any single value of *x*. That is, $P(x = a) = 0$.

Continuous Probability Distributions

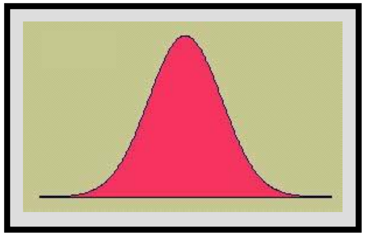
- There are many different types of continuous random variables



- We try to pick a model that
  - Fits the data well
  - Allows us to make the best possible inferences using the data.
- One important continuous random variable is the **normal random variable**.



The Normal Distribution



- The formula that generates the normal probability distribution is:

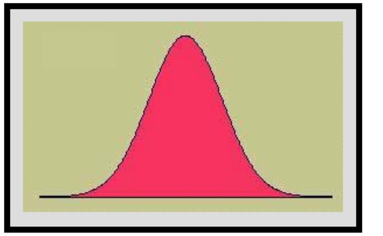
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } -\infty < x < \infty$$

$$e = 2.7183 \qquad \pi = 3.1416$$

μ and σ are the population mean and standard deviation.

- The shape and location of the normal curve changes as the mean and standard deviation change.

The Standard Normal Distributio



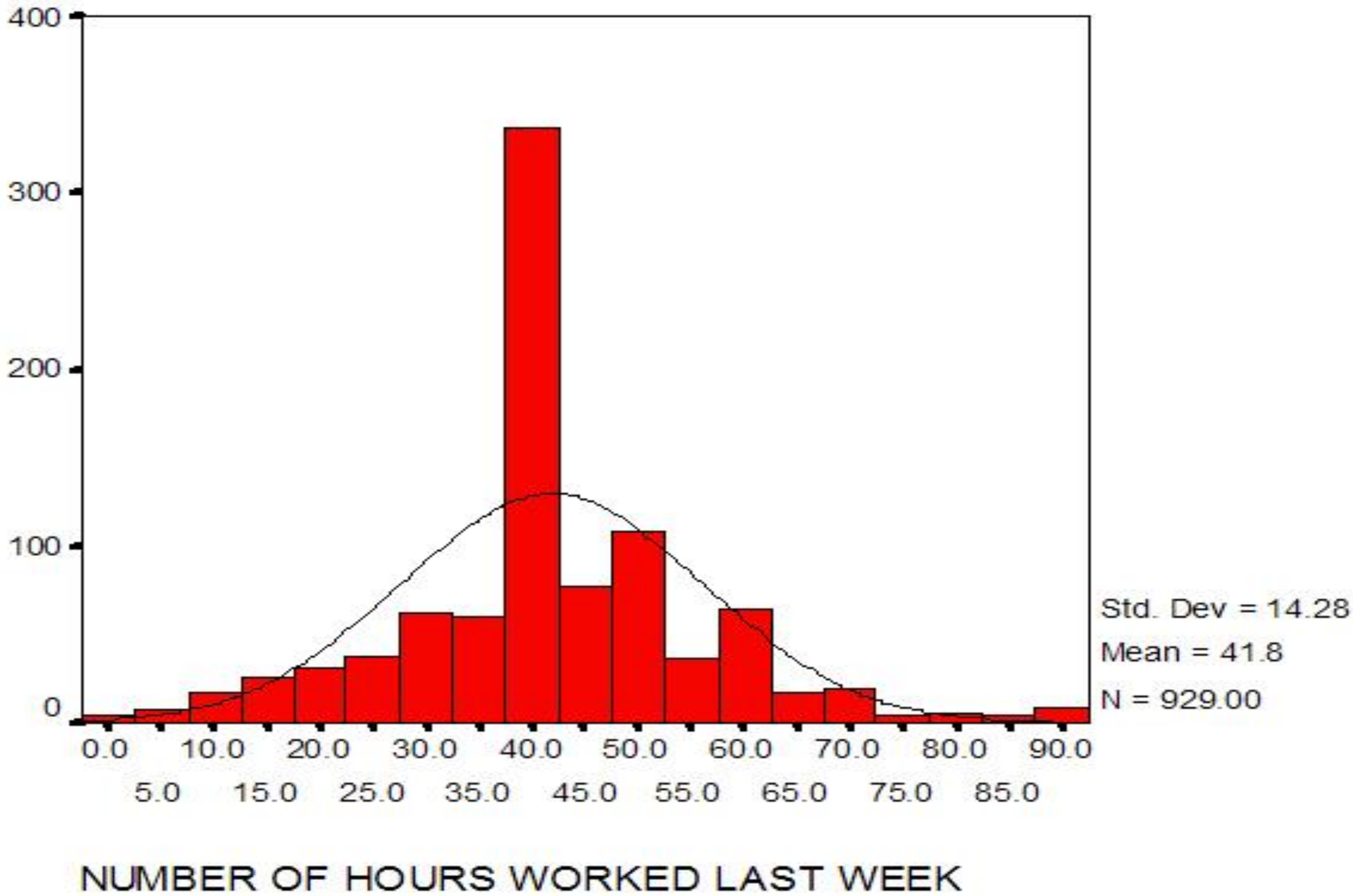
- To find P(a < x < b), we need to find the area under the appropriate normal curve.
- To simplify the tabulation of these areas, we **standardize** each value of x by expressing it as a z-score, the number of standard deviations σ it lies from the mean μ.

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

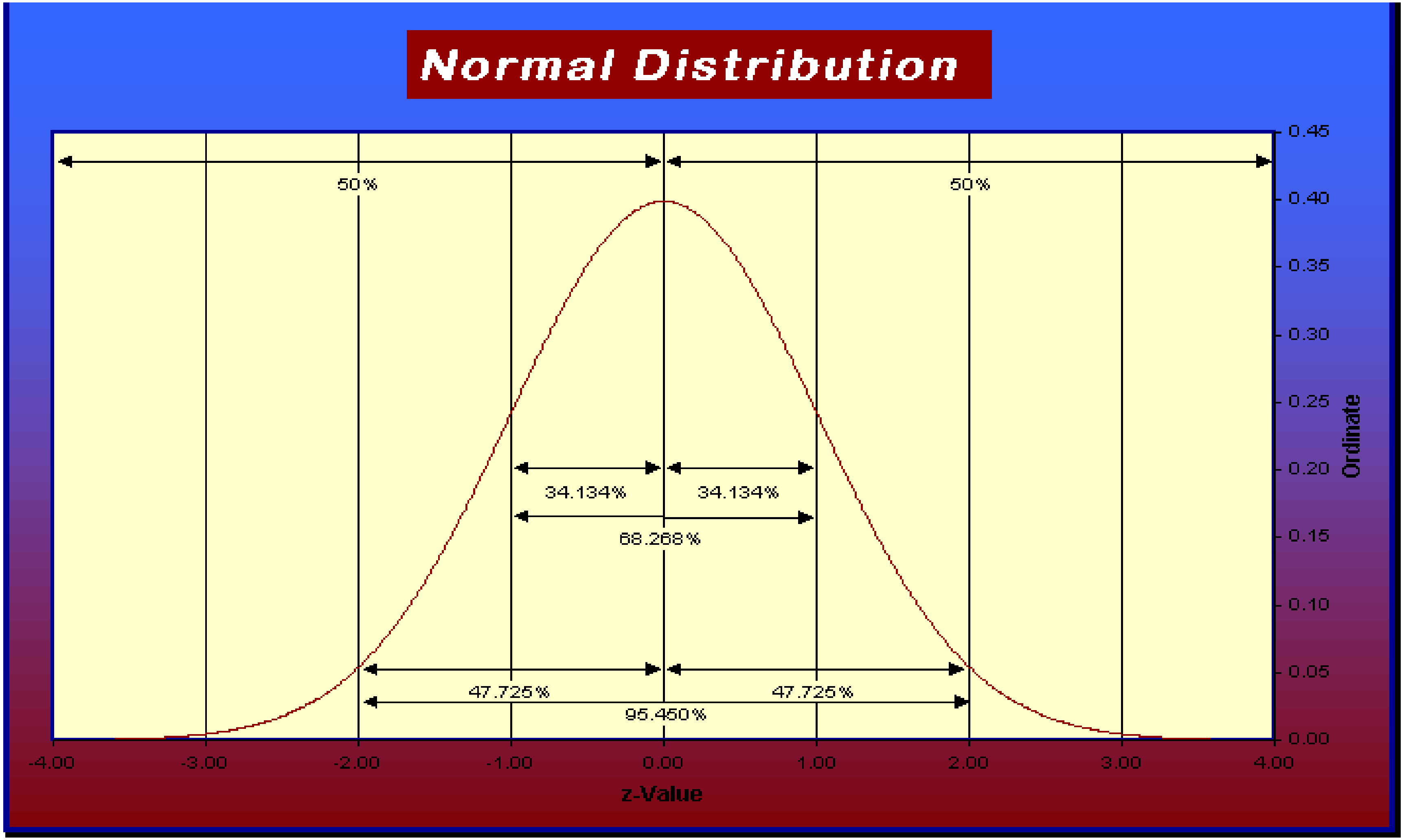
Carl Gauss

- The normal curve is often called the Gaussian distribution, after Carl Friedrich Gauss, who discovered many of its properties. Gauss, commonly viewed as one of the greatest mathematicians of all time (if not *the* greatest), is honoured by Germany on their 10 Deutschmark bill.
- From http://www.willamette.edu/~mjaneba/help/normalcurve.html

The Histogram and the Normal Curve



The Theoretical Normal Curve



Properties of the Normal Curve:

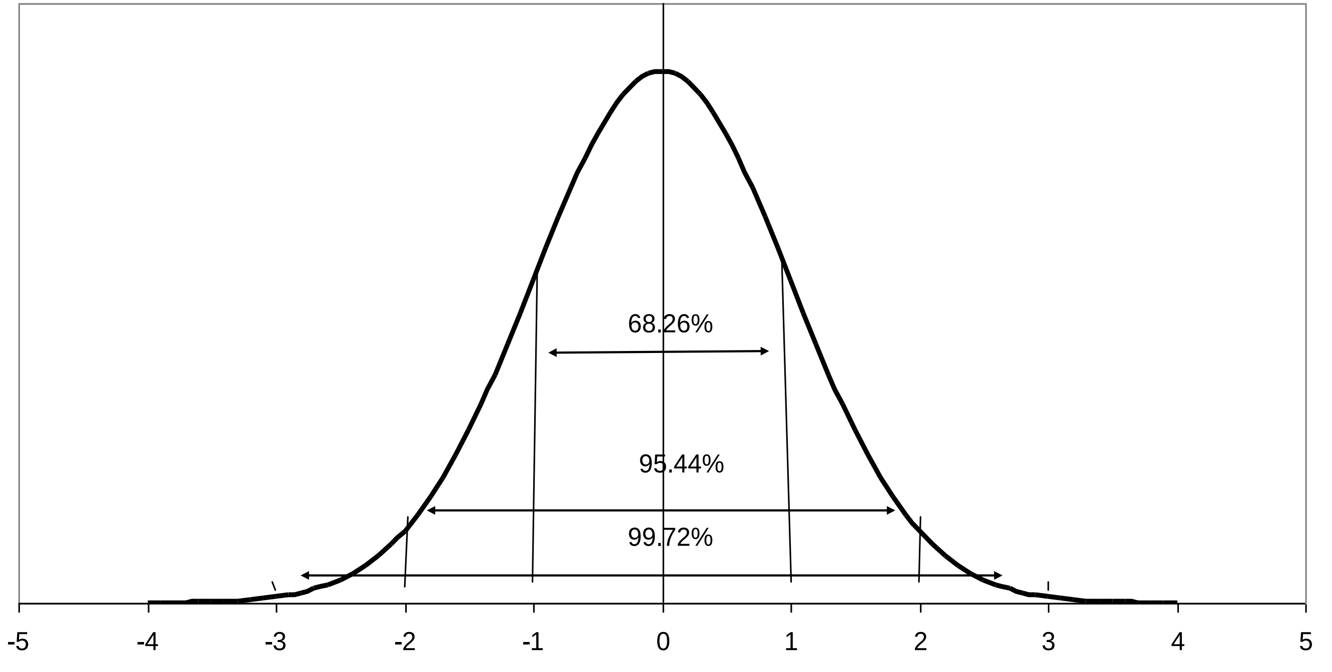
- Theoretical construction
- Also called Bell Curve or Gaussian Curve
- Perfectly symmetrical normal distribution
- The mean of a distribution is the midpoint of the curve
- The tails of the curve are infinite
- Mean of the curve = median = mode
- The “area under the curve” is measured in standard deviations from the mean

Properties (cont.)

- Has a mean = 0 and standard deviation = 1.
- General relationships: ±1 s = about 68.26%

  ±2 s = about 95.44%

  ±3 s = about 99.72%

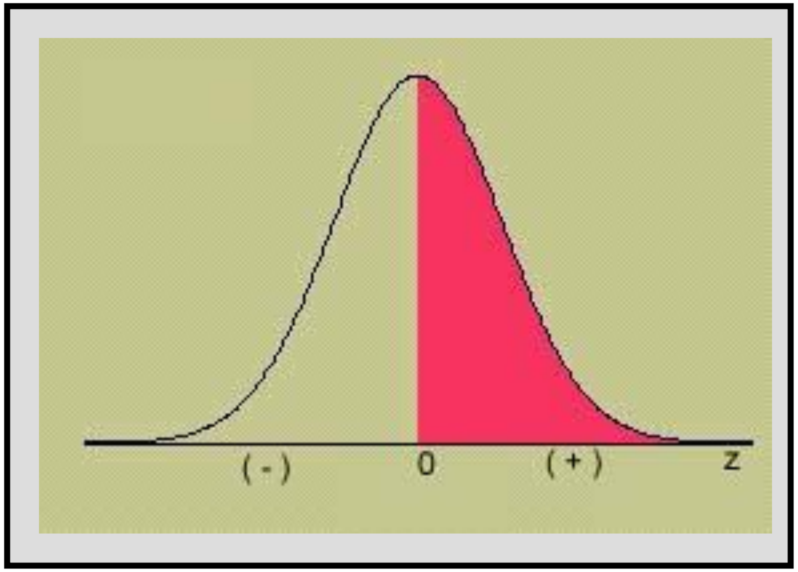


Z-Scores

- Are a way of determining the position of a single score under the normal curve.
- Measured in standard deviations relative to the mean of the curve.
- The Z-score can be used to determine an area under the curve known as a <u>probability</u>.

- Formula:

$$\boxed{z = \frac{x - \mu}{\sigma}}$$



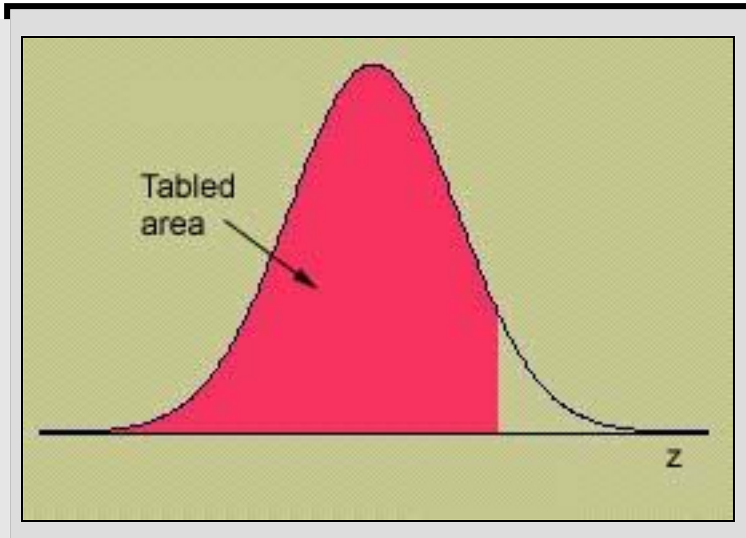
The Standard Normal (z) Distribution

- Mean = 0; Standard deviation = 1
- When $x = \mu$, $z = 0$
- Symmetric about $z = 0$
- Values of z to the left of center are negative
- Values of z to the right of center are positive
- Total area under the curve is 1.

Using Table 3

The four digit probability in a particular row and column of Table 3 gives the area under the *z* curve to the left that particular value of

z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8328	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278



Area for $z = 1.36$

Example

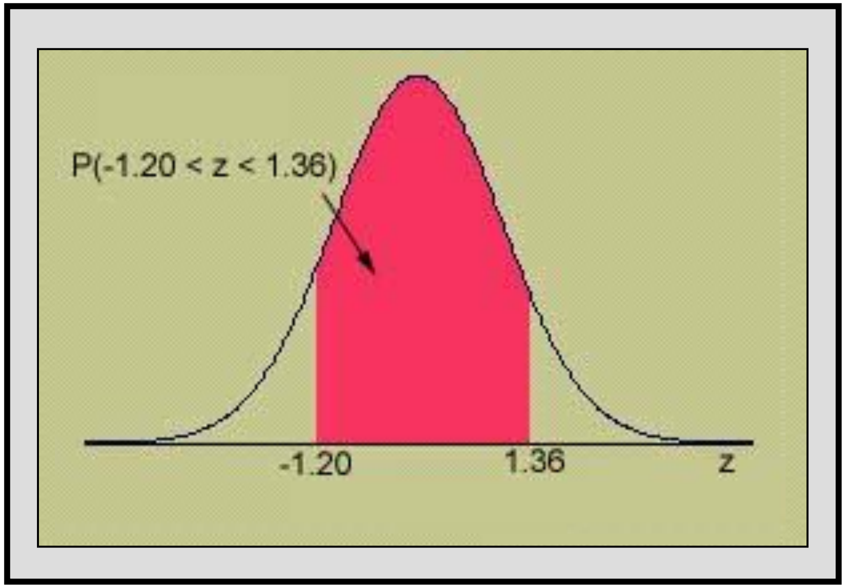
Use Table 3 to calculate these probabilities:

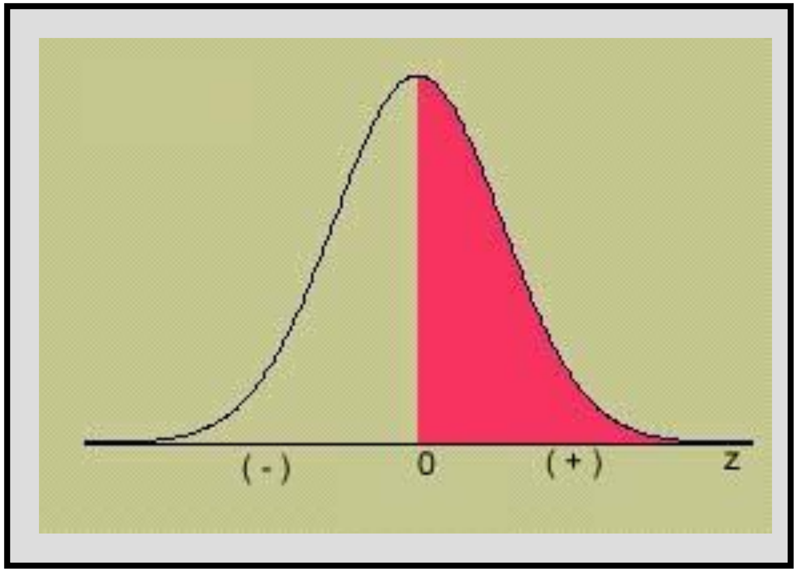
$P(z \le 1.36) = .9131$

$P(z > 1.36)$

$= 1 - .9131 = .0869$

$P(-1.20 \le z \le 1.36) =$
$.9131 - .1151 = .7980$





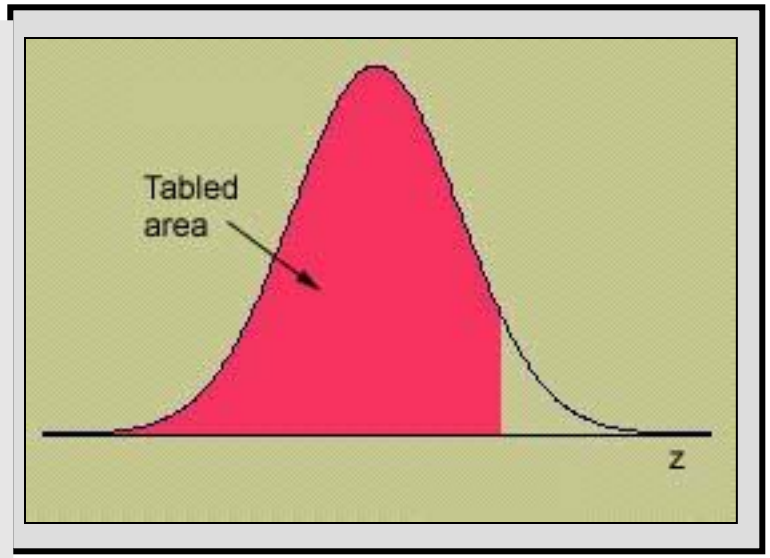
The Standard Normal (z) Distribution

- Mean = 0; Standard deviation = 1
- When $x = \mu$, $z = 0$
- Symmetric about $z = 0$
- Values of z to the left of center are negative
- Values of z to the right of center are positive
- Total area under the curve is 1.

Using Table 3

The four digit probability in a particular row and column of Table 3 gives the area under the *z* curve to the left that particular value of

z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8328	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278



Area for $z = 1.36$

Example

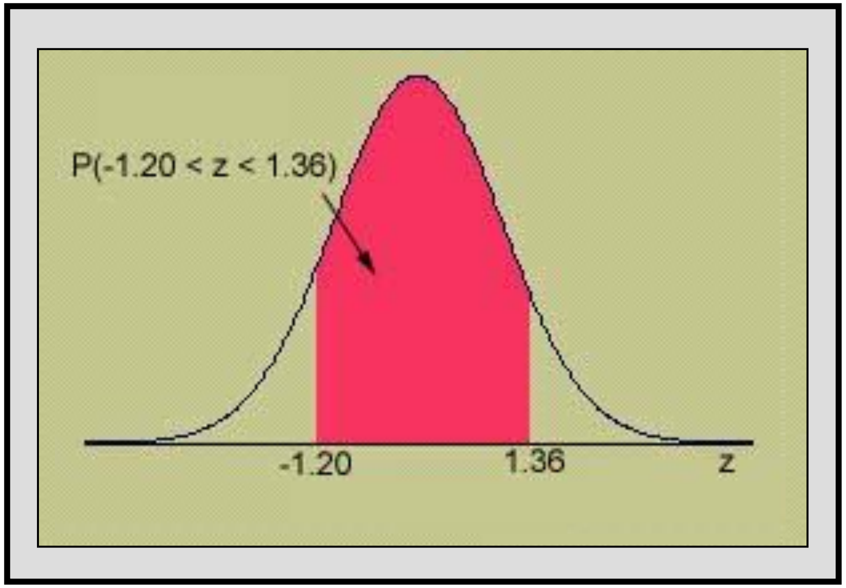
Use Table 3 to calculate these probabilities:

$P(z \leq 1.36) = .9131$

$P(z > 1.36)$

$= 1 - .9131 = .0869$

$P(-1.20 \leq z \leq 1.36) =$
$.9131 - .1151 = .7980$



Using the Normal Curve: Z Scores

- **Procedure:**

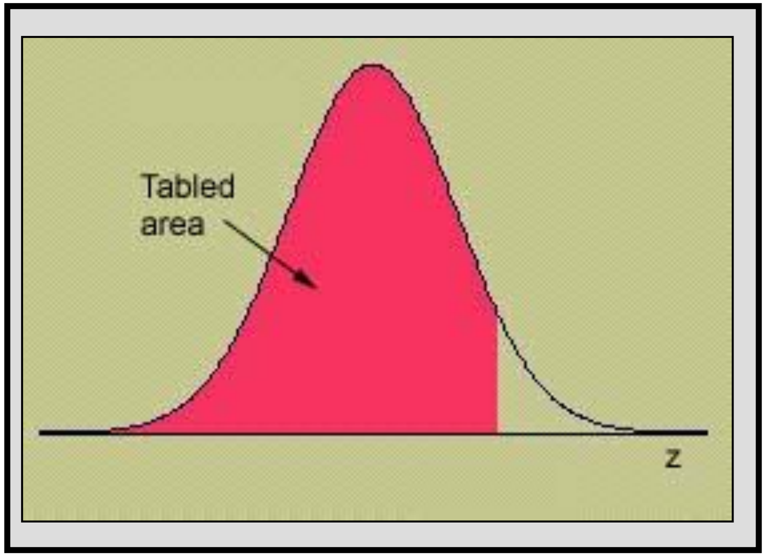
- To find areas, first compute Z scores.
- Substitute score of interest for X_i
- Use sample mean for $\bar{X}$ and sample standard deviation for S.
- The formula changes a “raw” score (X_i) to a standardized score (Z).

Finding Probabilities

- If a distribution has:
  - $\bar{X}$ = 13
  - s = 4

- What is the probability of randomly selecting a score of 19 or more?

P(19>=?) =
Z = 1.5
Probability =1-0.9332 = 0.0668
Percent 6.68%



In Class Example

- After an exam, you learn that the mean for the class is 60, with a standard deviation of 10. Suppose your exam score is 70. What is your Z-score?
- Where, relative to the mean, does your score lie?
- What is the probability associated with your score (use Z table Appendix I Table 3)?

To solve:

- Available information: $X_i = 70$

  $\bar{X} = 60$

  $S = 10$

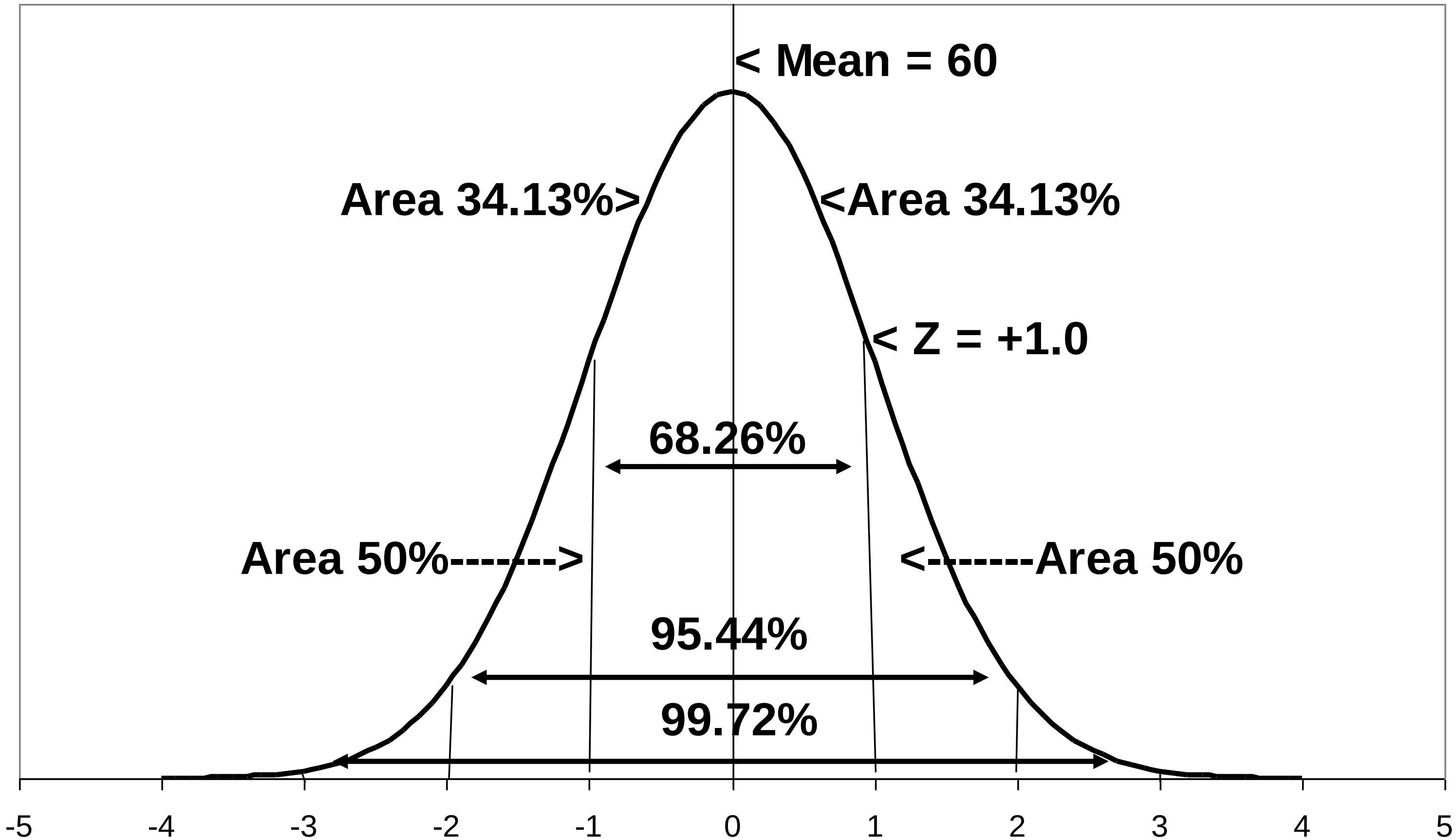
- Formula: $Z = (X_i - \bar{X}) / S$

  $= (70 - 60) / 10$

  $= +1.0$

Your Z-score of +1.0 is exactly 1 s.d. above the mean

You are at the **84.13** percentile.



What if your score is 72?

- Calculate your Z-score.
- What percentage of students have a score below your score? Above?
- What percentile are you at?

Answer:

- Z = 1.2

- The area beyond Z = .1151
  (11.51% of marks are above yours)

- Area between mean and Z = .8849
  (% of marks below = 88.49%)

- Your mark is at the 88^{th} percentile!

What if your mark is 55%?

- Calculate your Z-score.
- What percentage of students have a score below your score? Above?
- What percentile are you at?

Answer:

- Z = -.5

- Area between the mean and Z =1-0.3085= 0.6915
  (% of marks above = 69.15%)

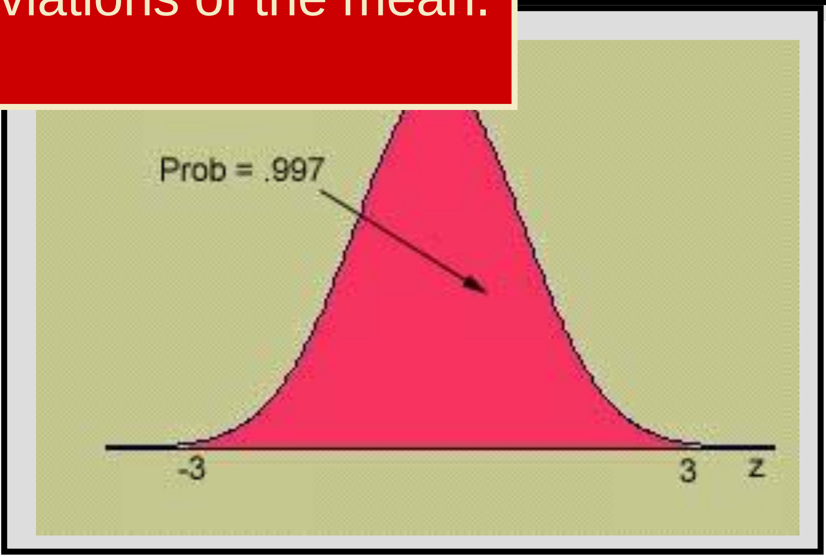
- The area beyond Z = .3085
  (30.85% of the marks are below yours)

- Your mark is only at the 31st percentile!

Using Table 3

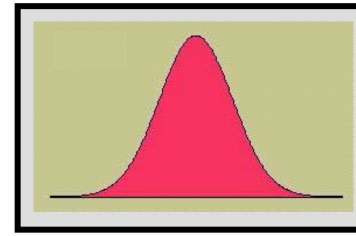
✓To find an area to the left of a *z*-value, find the area directly from the table.
✓To find an area to the right of a *z*-value, find the area in Table 3 and s
✓To find the are find the two areas in Table 3 other.

Remember the Empirical Rule: Approximately 95% of the measurements lie within 2 standard deviations of the mean.



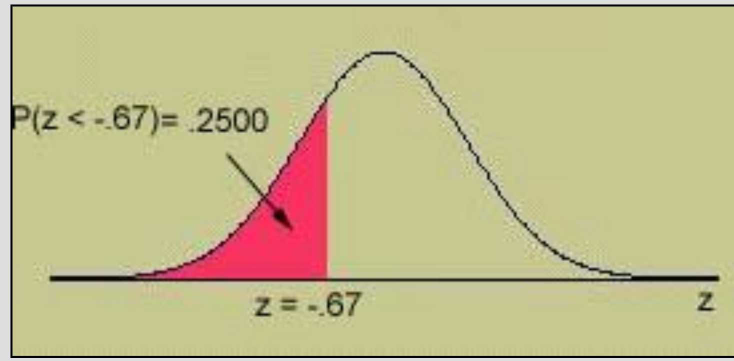
$P(-3 \le z \le 3)$
$= .9987 - .0013 = .9974$

Working Backwards



Find the value of *z* that has area .25 to its left.

1. Look for the four digit area closest to .2500 in Table 3.
2. What row and column does this value correspond to?
3. $z = -.67$

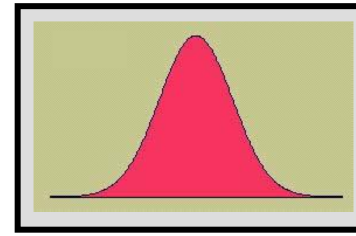


		.03	.04	.05	.06	.07	.08
0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635
0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894
0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177
[illegible]	.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483
[illegible]	.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810

4. What percentile does this value represent?

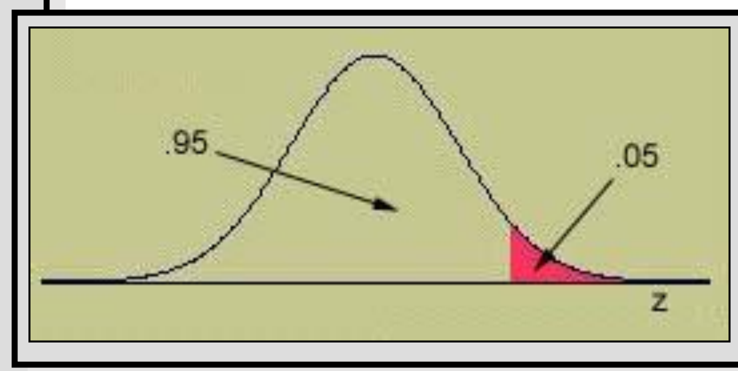
25th percentile,
or 1st quartile (Q_1)

Working Backwards



Find the value of z that has area .05 to its right.

1. The area to its left will be 1 - .05 = .95
2. Look for the four digit area closest to .9500 in Table 3.
3. Since the value .9500 is halfway between .9495 and .9505, we choose z halfway between 1.64 and 1.65.
4. $z = 1.645$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
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

Finding Probabilities for the General Normal Random Variable

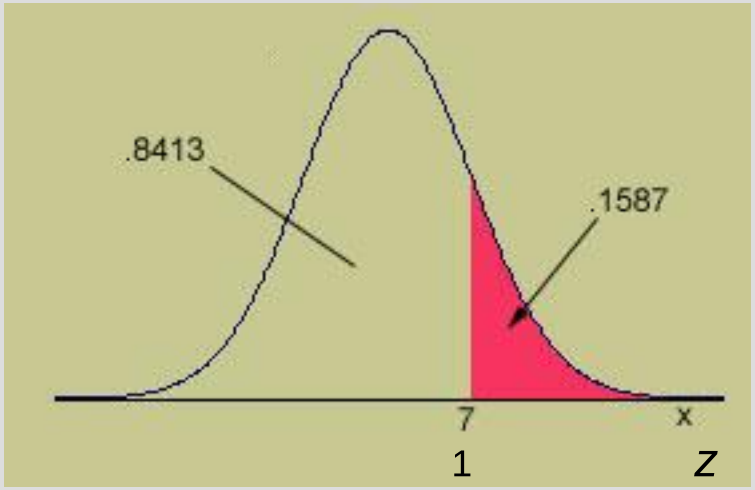
✓To find an area for a normal random variable *x* with mean μ and standard deviation σ, *standardize or rescale* the interval in terms of *z.*

✓Find the appropriate area using Table 3.

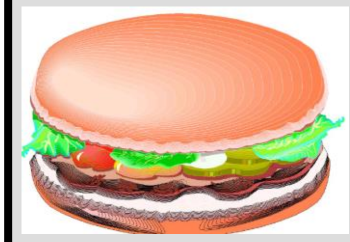
Example: *x* has a normal distribution with μ = 5 and σ = 2. Find P(*x* > 7).

$$P(x > 7) = P(z > \frac{7-5}{2})$$

$$= P(z > 1) = 1 - .8413 = .1587$$

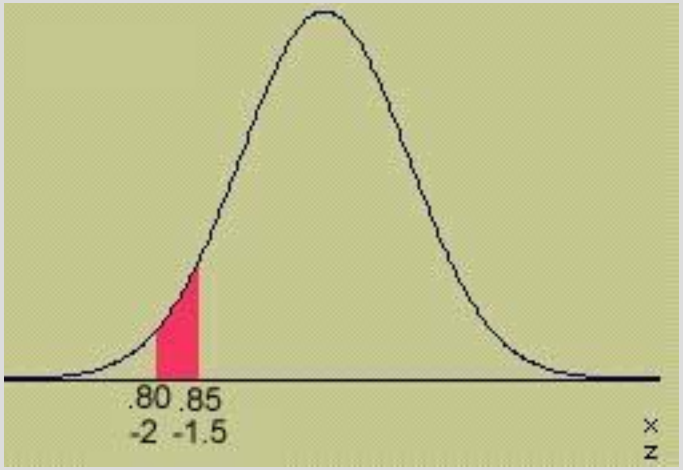


Example



The weights of packages of ground beef are normally distributed with mean 1 pound and standard deviation .10. What is the probability that a randomly selected package weighs between 0.80 and 0.85 pounds?

$$P(.80 < x < .85) =$$
$$P(-2 < z < -1.5) =$$
$$.0668 - .0228 = .0440$$



Example

John earned a score of 940 on a national achievement test. The mean test score was 850 with a standard deviation of 100. What proportion of students had a higher score than John? (Assume that test scores are normally distributed.)

Example

John earned a score of 940 on a national achievement test. The mean test score was 850 with a standard deviation of 100. What proportion of students had a higher score than John? (Assume that test scores are normally distributed.)

First, we transform John's test score into a z-score, using the z-score transformation equation.

$$z = (X - \mu) / \sigma = (940 - 850) / 100 = 0.90$$

- Then, using the standard normal distribution table, we find the cumulative probability associated with the z-score. In this case, we find $P(Z < 0.90) = 0.8159$.

- Therefore, the $P(Z > 0.90) = 1 - P(Z < 0.90) = 1 - 0.8159 = 0.1841$

Example

A normally distributed population of test scores has a mean of 80 and a standard deviation of 5.2.

a) Find the percentage of scores that lies below 73.
b) Find the percentage of scores that lies between 82 and 86.
c) Find the percentage of scores that lies above 73.

Example

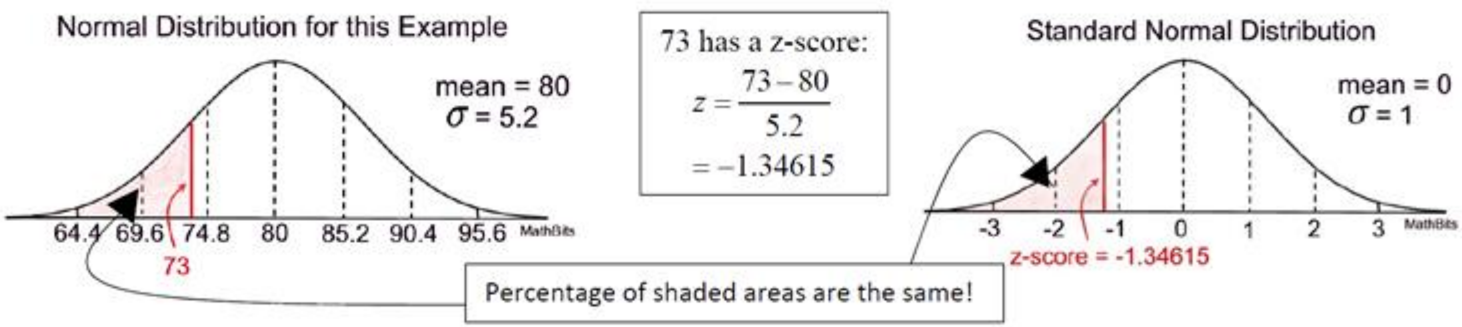
A normally distributed population of test scores has a mean of 80 and a standard deviation of 5.2.

a) Find the percentage of scores that lies below 73.

b) Find the percentage of scores that lies between 82 and 86.

c) Find the percentage of scores that lies above 73.

a) Find the percentage of scores that lies below 73.



Now, that the z-score is known, the Z-score Chart will show the percentage of the shaded area to the left of this z-score of **-1.34.**

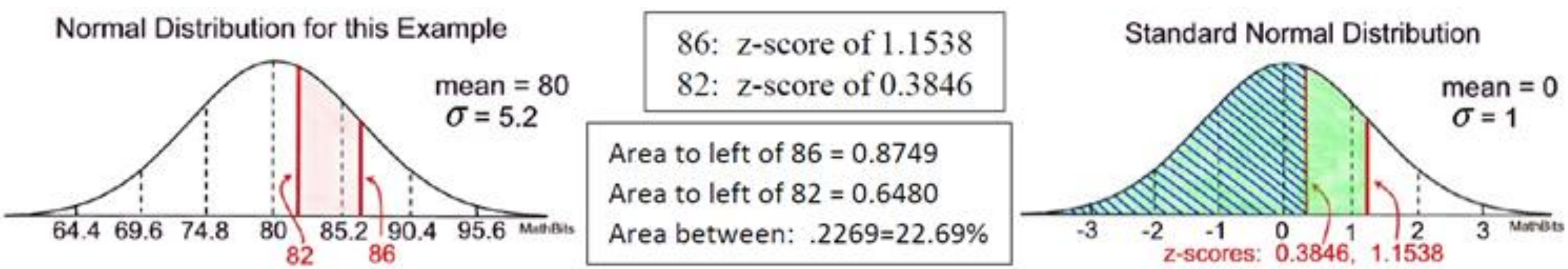
	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.0
-1.3	0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968

Percentage of scores below 73 is 9.01%.

Example

A normally distributed population of test scores has a mean of 80 and a standard deviation of 5.2.
a) Find the percentage of scores that lies below 73.
b) Find the percentage of scores that lies between 82 and 86.
c) Find the percentage of scores that lies above 73.

b) Find the percentage of scores that lies between 82 and 86.



Example

A normally distributed population of test scores has a mean of 80 and a standard deviation of 5.2.

a) Find the percentage of scores that lies below 73.
b) Find the percentage of scores that lies between 82 and 86.
c) Find the percentage of scores that lies above 73.

c) Find the percentage of scores that lies above 73.

To find the percentage of the area that lies "above" the z-score, take the total area under a normal curve (which is 1) and subtract the cumulative area to the left of the z-score.

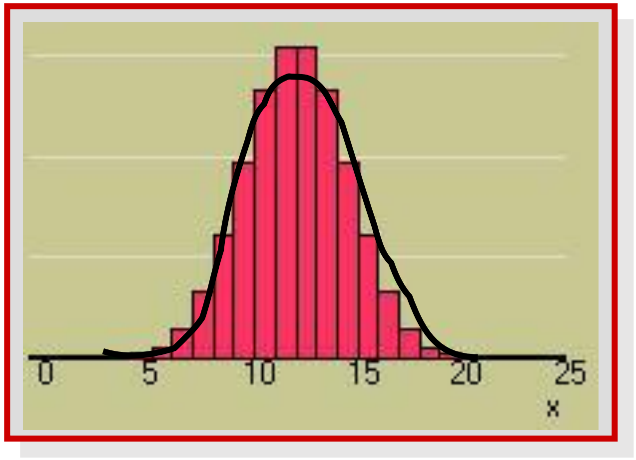
In part a, 73 had a z-score of -1.34615 with a cumulative area to the left of 0.0901 or 9.01%.

The area to the right of this z-score will be 1 - 0.0901 = 0.9099 or 90.99%.

The Normal Approximation to the Binomial

- We can calculate binomial probabilities using
  - The binomial formula
  - The cumulative binomial tables

- When n is large, and p is not too close to zero or one, areas under the normal curve with mean np and variance npq can be used to approximate binomial probabilities.



Approximating the Binomial

✓Make sure to include the entire rectangle for the values of *x* in the interval of interest. This is called the **continuity correction.**

✓Standardize the values of *x* using

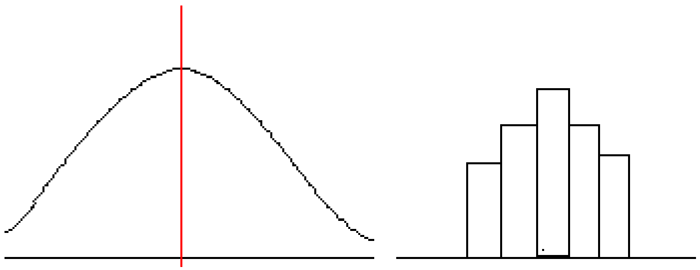
$$\boxed{z = \frac{x - np}{\sqrt{npq}}}$$

✓Make sure that *np* and *nq* are both greater than 5 to avoid inaccurate approximations!

Continuity Correction

The binomial are discrete random variables, whereas the normal distribution is continuous. We need to take this into account when we are using the normal distribution to approximate a binomial or Poisson using a **continuity correction**.

In the discrete distribution, each probability is represented by a rectangle (right hand diagram):



When working out probabilities, we want to include whole rectangles, which is what continuity correction is all about.

Example

let n=10 and p=12. What is the probability that exactly five people approve of the job the President is doing?

First, recognize in our case that the mean is:

$$\mu = np = 10\left(\frac{1}{2}\right) = 5$$

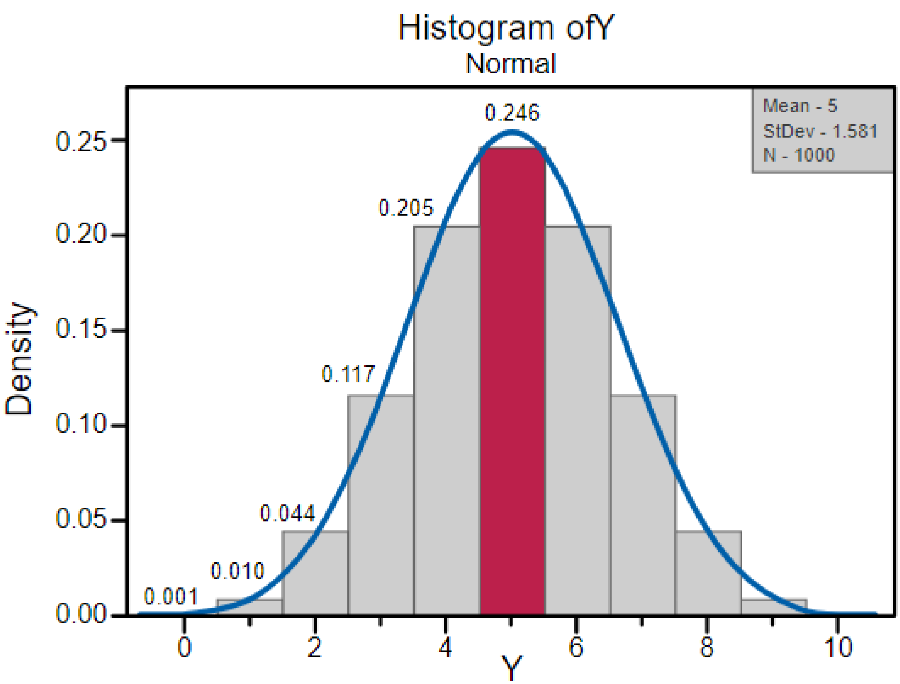
and the variance is:

$$\sigma^2 = np(1-p) = 10\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 2.5$$

Example

let n=10 and p=12. What is the probability that exactly five people approve of the job the President is doing?

Now, if we look at a graph of the binomial distribution with the rectangle corresponding to Y=5 shaded in red:



we should see that we would benefit from making correction for the fact that we are using a continuous distribution to approximate a discrete distribution. Specifically, it seems that the rectangle Y=5 really includes any Y greater than 4.5 but less than 5.5. That is:

$$P(Y=5)=P(4.5<Y<5.5)$$

Such an adjustment is called a "**continuity correction**." Once we've made the continuity correction, the calculation reduces to a normal probability calculation:

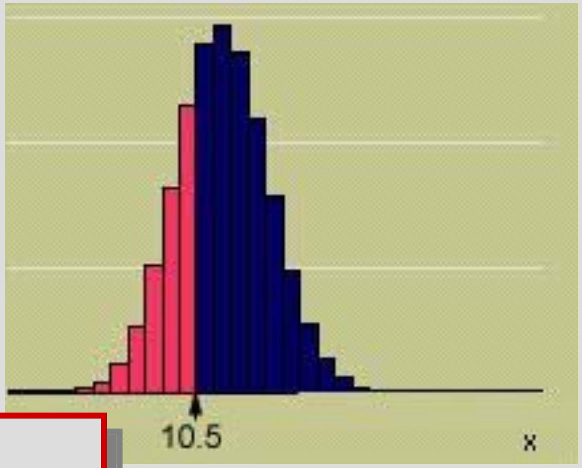
Example

Suppose x is a binomial random variable with $n = 30$ and $p = .4$. Using the normal approximation to find $P(x \leq 10)$.

$n = 30$ $\quad p = .4$ $\quad q = .6$

$np = 12$ $\quad nq = 18$

The normal approximation is ok!

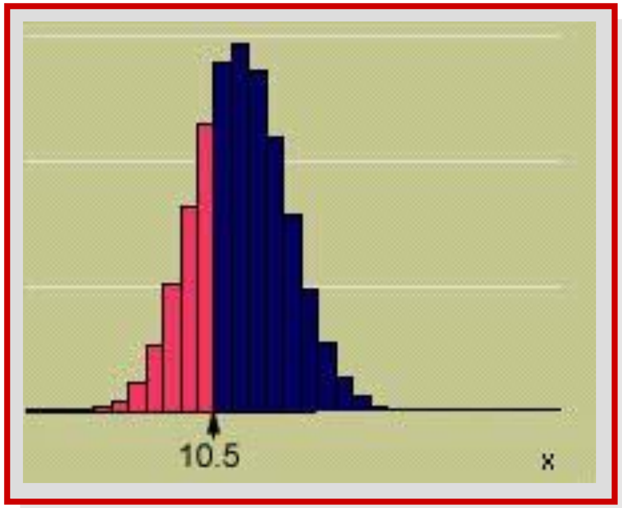


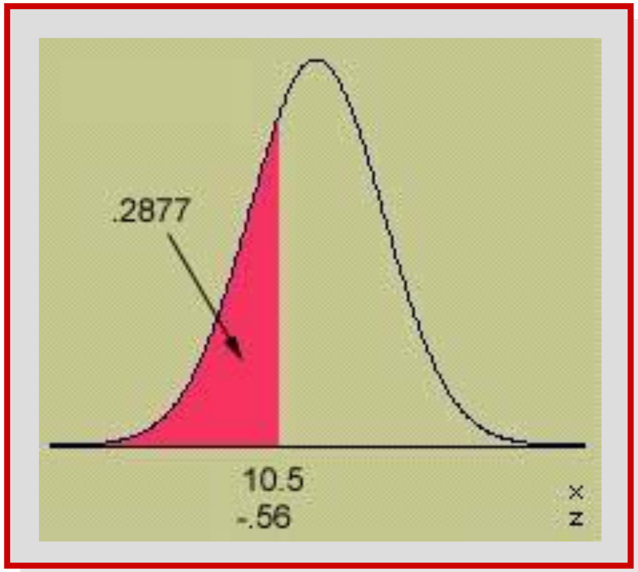
Calculate

$$\mu = np = 30(.4) = 12$$

$$\sigma = \sqrt{npq} = \sqrt{30(.4)(.6)} = 2.683$$

Example

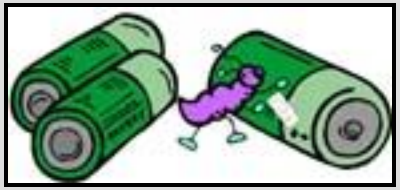




$$P(x \le 10) \approx P(z \le \frac{10.5-12}{2.683})$$

$$= P(z \le -.56) = .2877$$

Example



A production line produces AA batteries with a reliability rate of 95%. A sample of n = 200 batteries is selected. Find the probability that at least 195 of the batteries work.

Success = working battery $n = 200$

$p = .95$ $np = 190$ $nq = 10$

The normal approximation is ok!

$$P(x \geq 195) \approx P\left(z \geq \frac{194.5-190}{\sqrt{200(.95)(.05)}}\right)$$

$$= P(z \geq 1.46) = 1 - .9278 = .0722$$

Key Concepts

I. Continuous Probability Distributions

1. Continuous random variables
2. Probability distributions or probability density functions
   a. Curves are smooth.
   b. The area under the curve between *a* and *b* represents the probability that *x* falls between *a* and *b*.
   c. $P(x = a) = 0$ for continuous random variables.

II. The Normal Probability Distribution

1. Symmetric about its mean μ.
2. Shape determined by its standard deviation σ.

Key Concepts

III. The Standard Normal Distribution

1. The normal random variable *z* has mean 0 and standard deviation 1.
2. Any normal random variable *x* can be transformed to a standard normal random variable using

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

3. Convert necessary values of *x* to *z.*
4. Use Table 3 in Appendix I to compute standard normal probabilities.
5. Several important *z*-values have tail areas as follows:

Tail Area:	.005	.01	.025	.05	.10
z-Value:	2.58	2.33	1.96	1.645	1.28