STAT 139

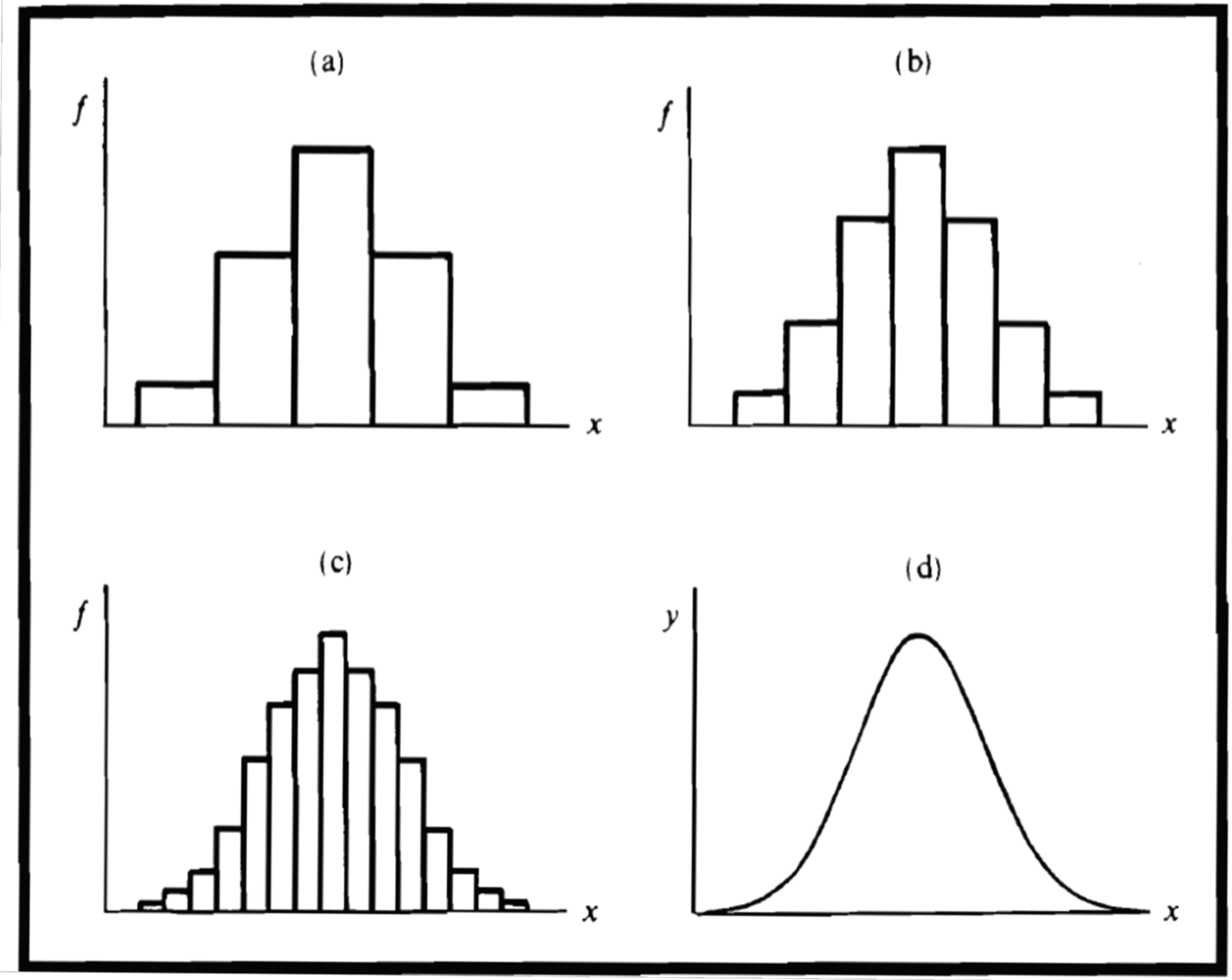
Lecture 9

The Normal Distribution

and

Continuous Probabilities

Normally distributed data



The normal distribution in a population

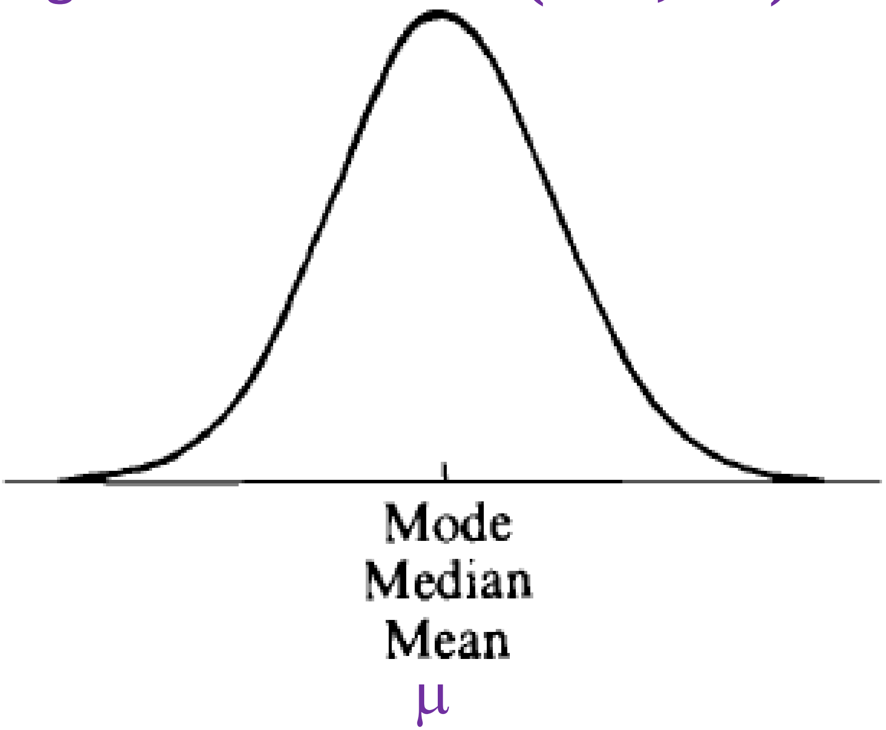
Usually described in shorthand as:

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

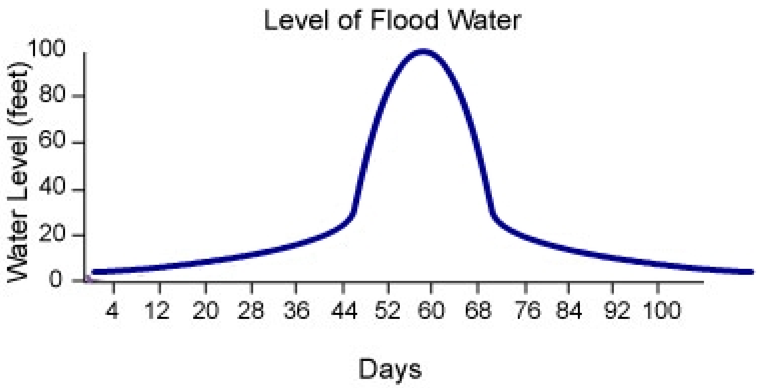
("myoo", "sigma")

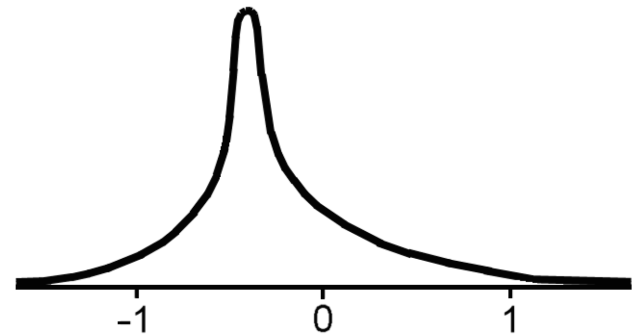
(Normal distribution with mean μ and standard deviation σ)

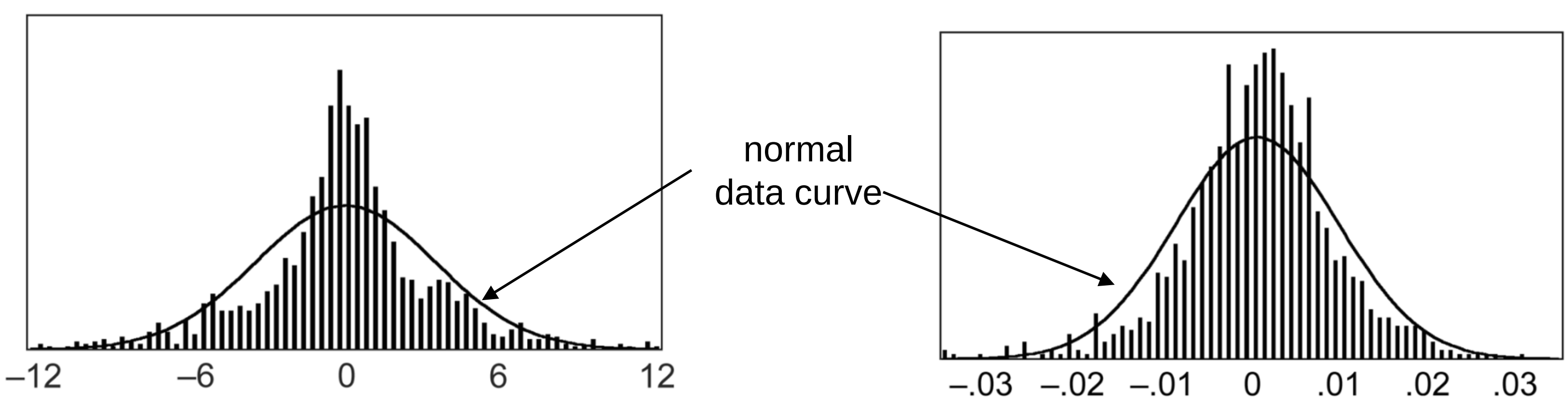
Example: Wing lengths in mm are **N (50.0, 4.5)**



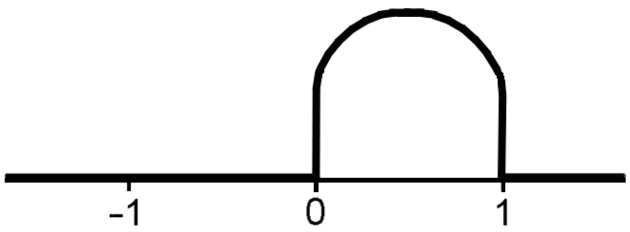
Leptokurtotic data

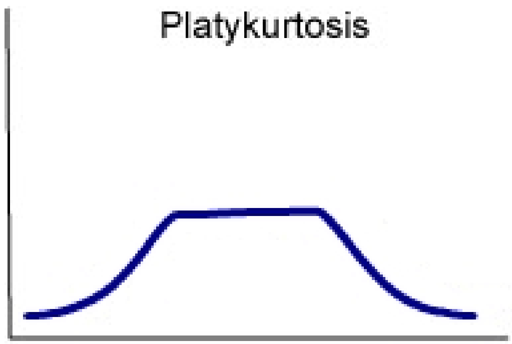


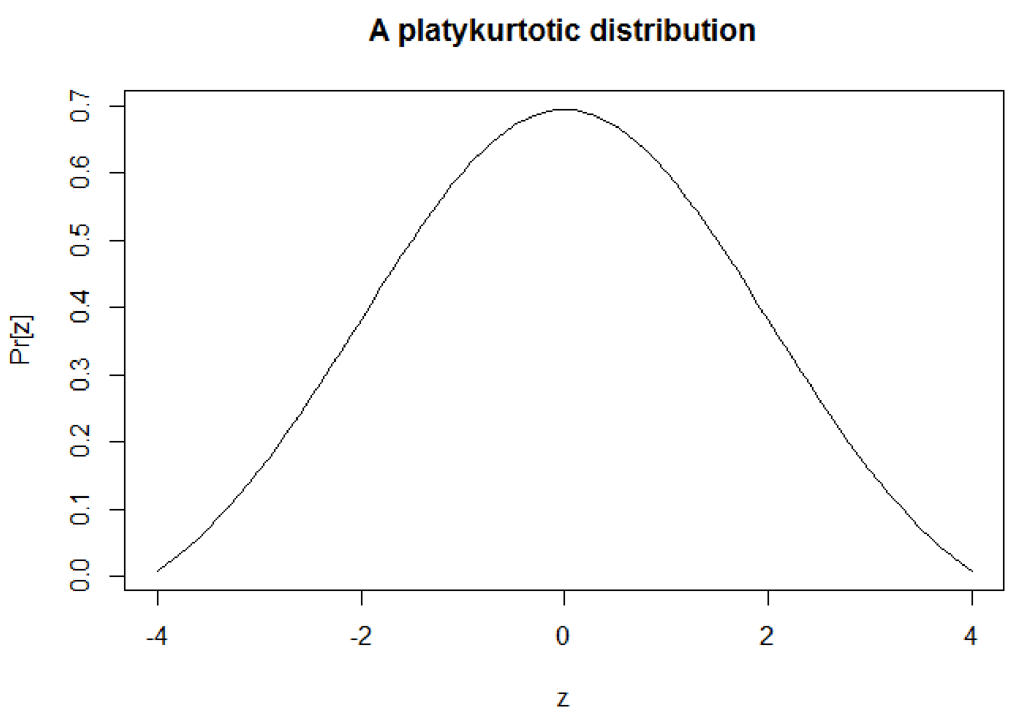




Platykurtotic data







Z scores (standardized data)

Observations converted to standard deviation scores (from a sample)

$\bar{x}$ = 50 mm, s = 4.5 mm;

$$z\text{-}score = \frac{x_i - \bar{x}}{s}$$

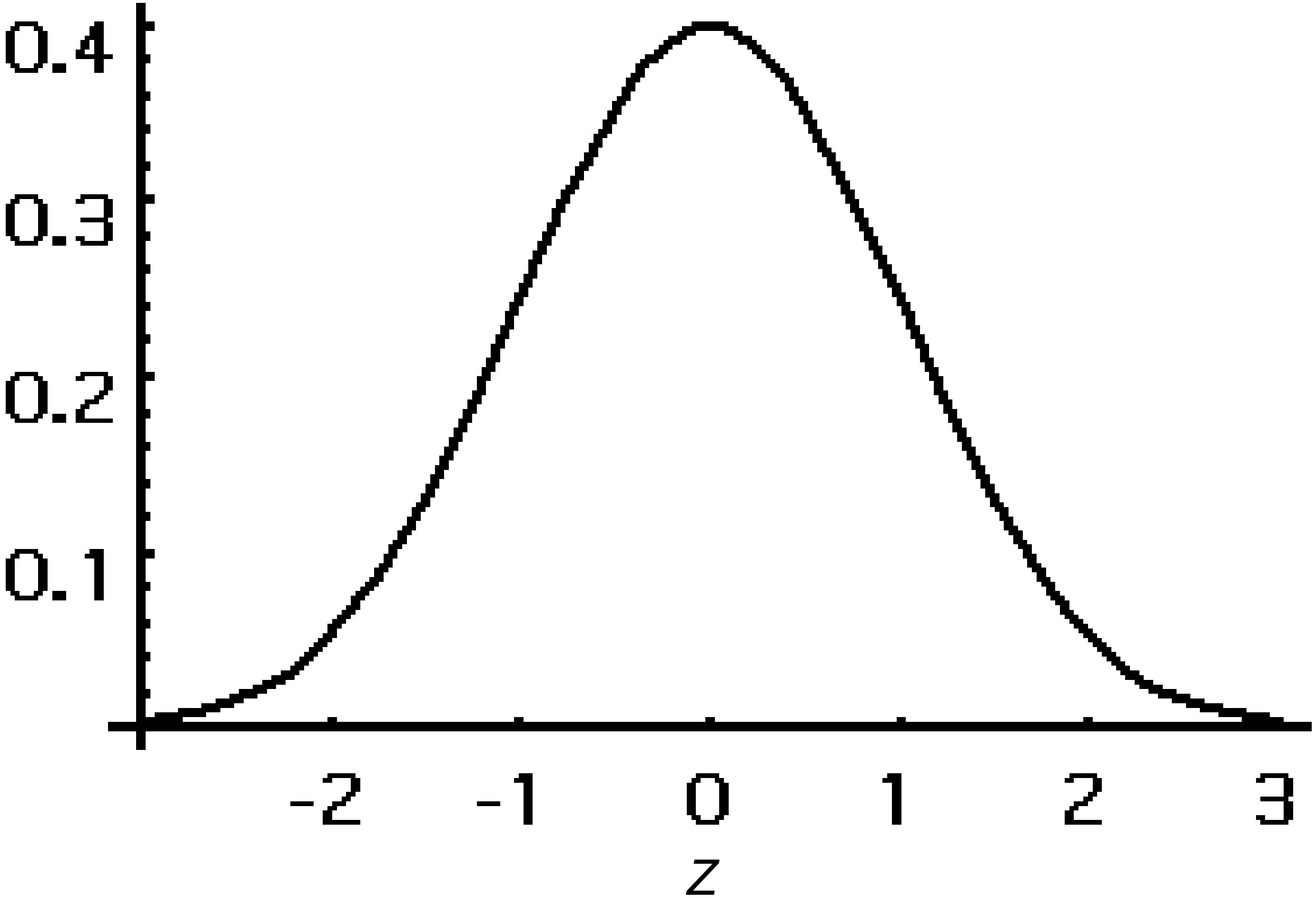
x_8 = 59 mm, $Z(x_8)$ = 2.0
x_9 = 41 mm, $Z(x_9)$ = -2.0
x_5 = 49 mm, $Z(x_5)$ = -0.22
x_6 = 50 mm, $Z(x_6)$ = 0.0

Z scores are unitless.

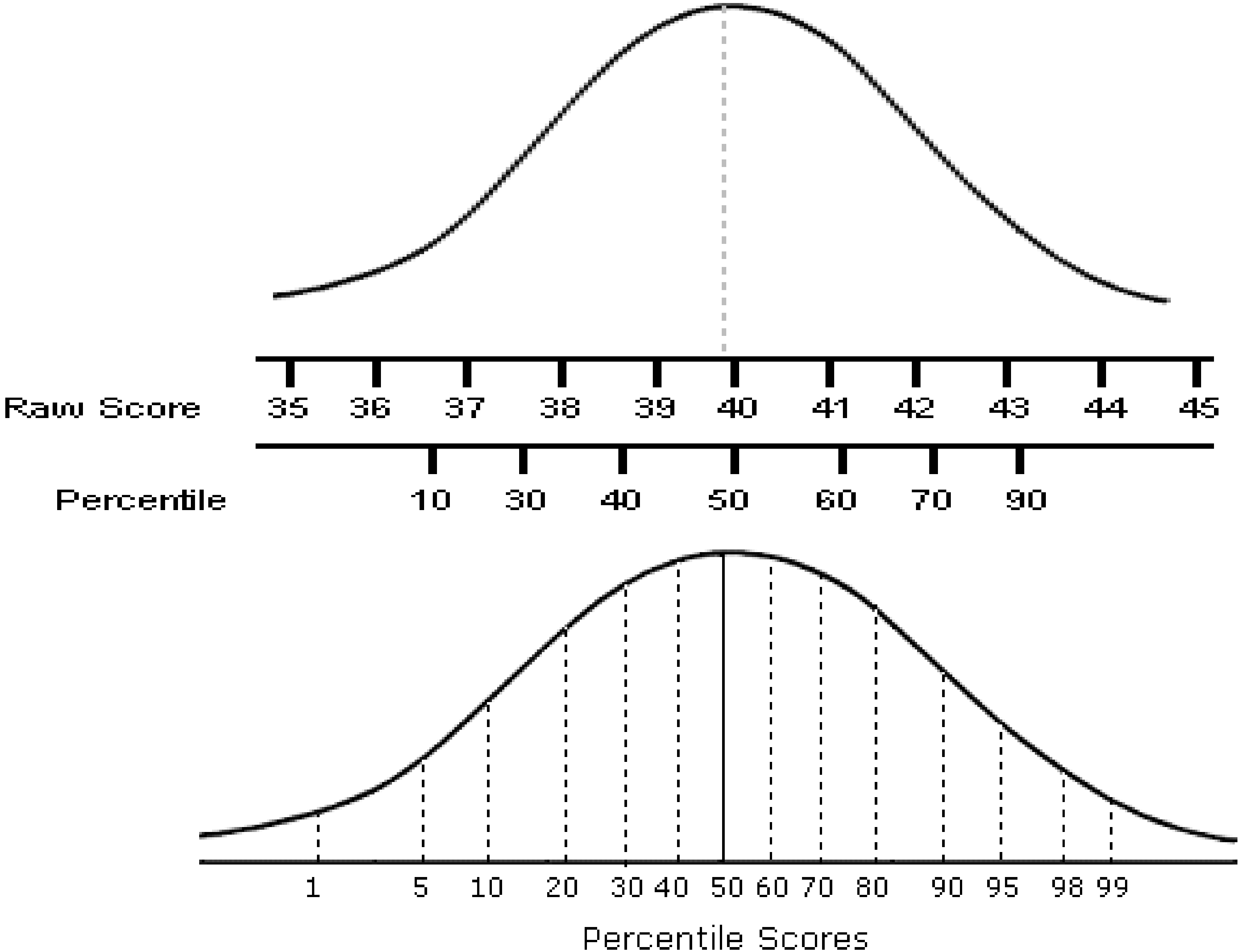
We can summarize an observation's relationship to the sample using ONE number!

The standard normal distribution

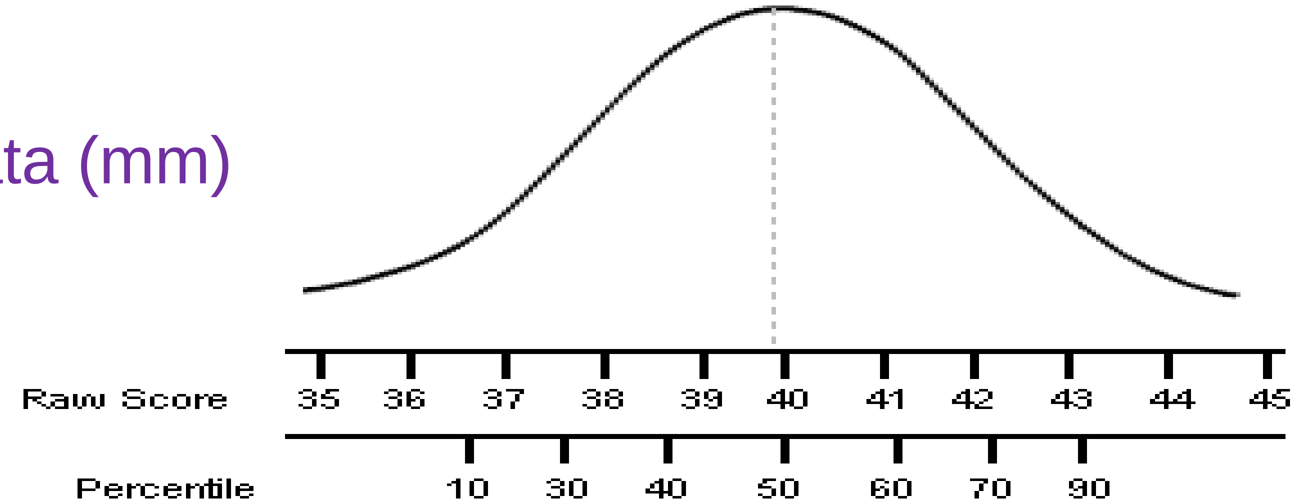
Once converted, *Z* scores are $N(0,1)$: **the standard normal distribution**



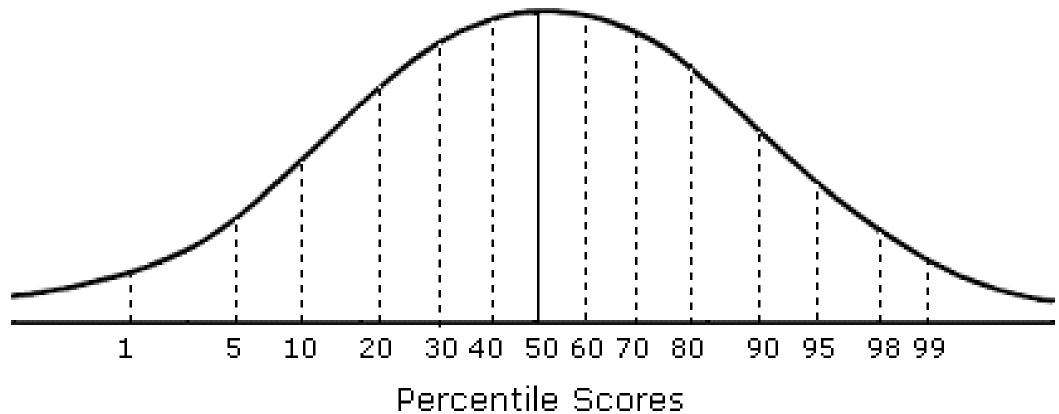
Percentiles are another scale



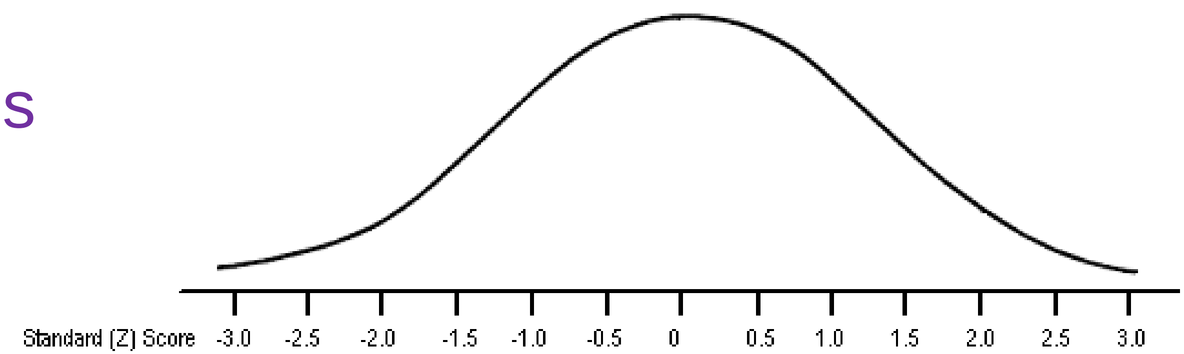
Raw data (mm)



Percentiles
(0-100)



Z scores



Continuous Random Variables

Situations that involve **measurements** often result in a **continuous random variable.**

A **continuous random variable *X*** takes on all values in an **interval** of numbers. The probability distribution of *X* is described by a **density curve.** The probability of any event is **the area under the density curve and related to *X*** (greater than *X* or less than *X*).

The probability model of a random variable *X* assigns a probability between 0 and 1 to each possible value of *X*.

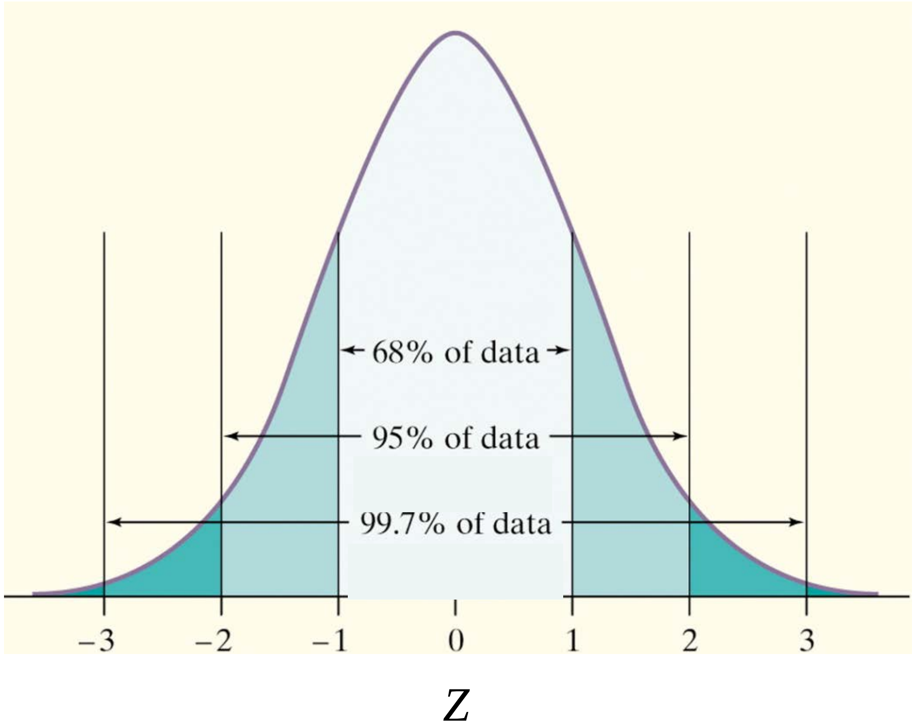
A continuous random variable *X* has *infinitely many* possible values.

So only *intervals* of values have useful probabilities.

The 68-95-99.7 Rule for Normal Data

In the Normal distribution with mean μ and standard deviation σ:

- Approximately **68%** of the observations fall within σ of μ.
- Approximately **95%** of the observations fall within 2σ of μ.
- Approximately **99.7%** of the observations fall within 3σ of μ.



Stature and the normal distribution

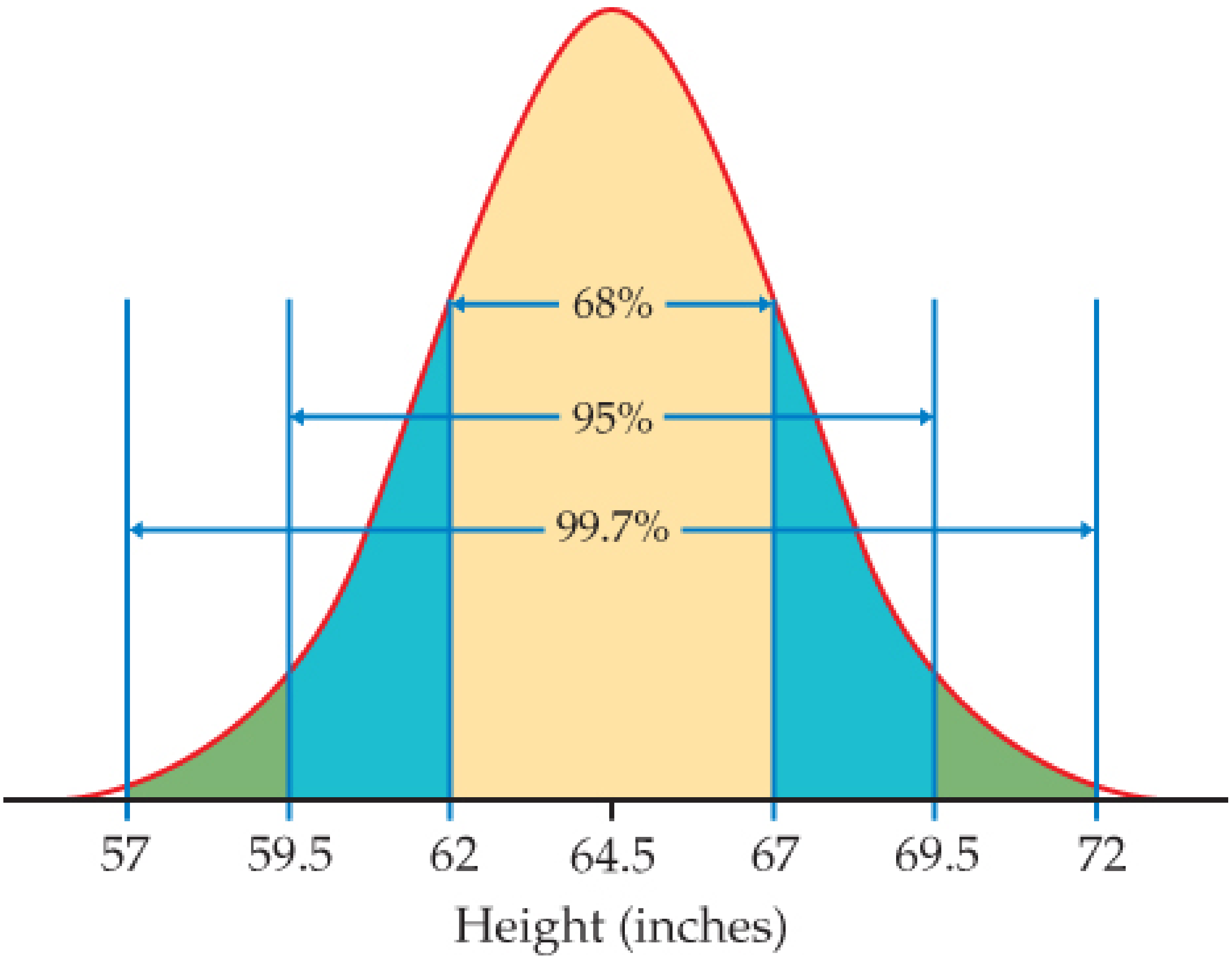
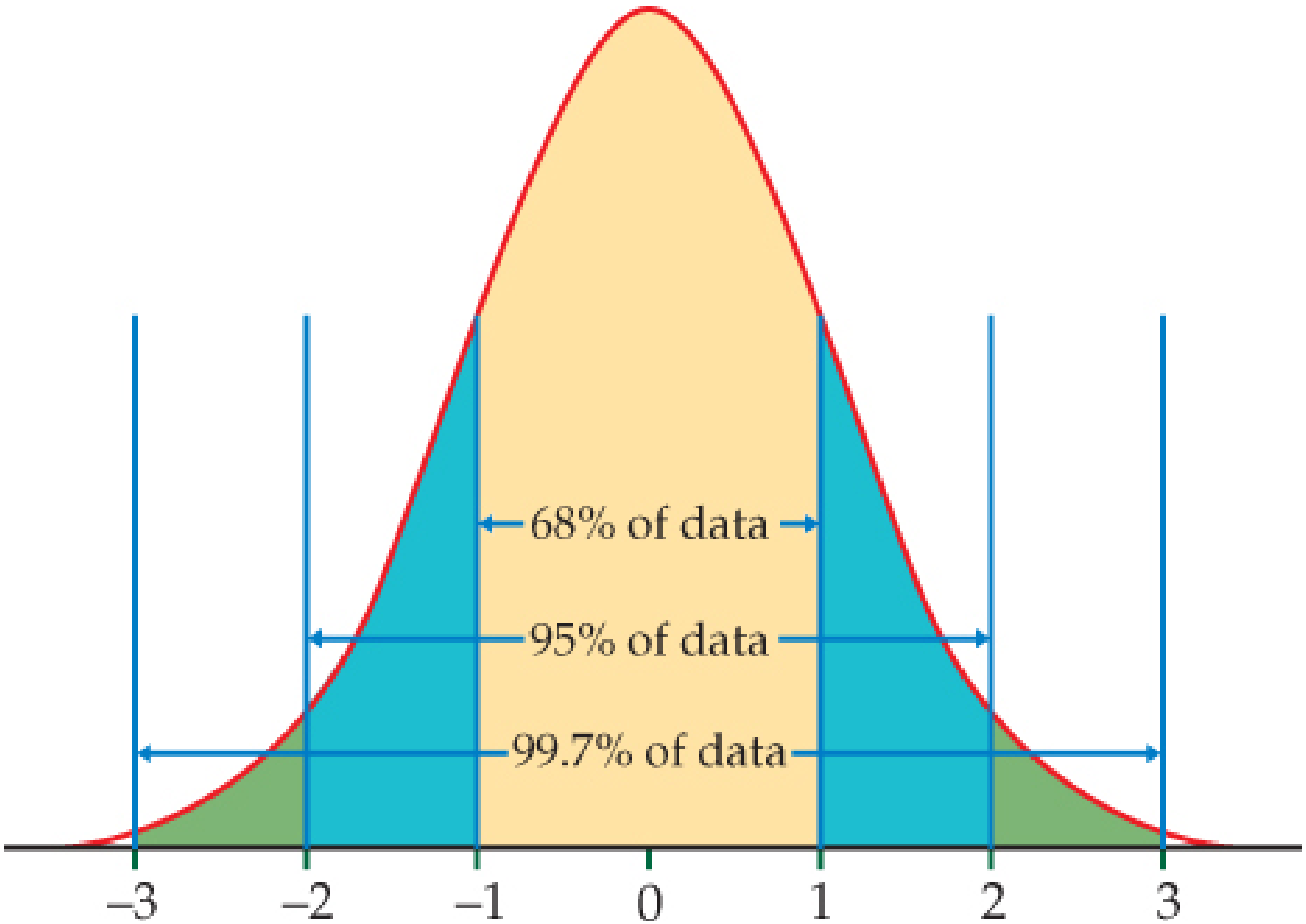


Figure 1.25, INTRODUCTION to the PRACTICE of STATISTICS, © 2014 W. H. Freeman

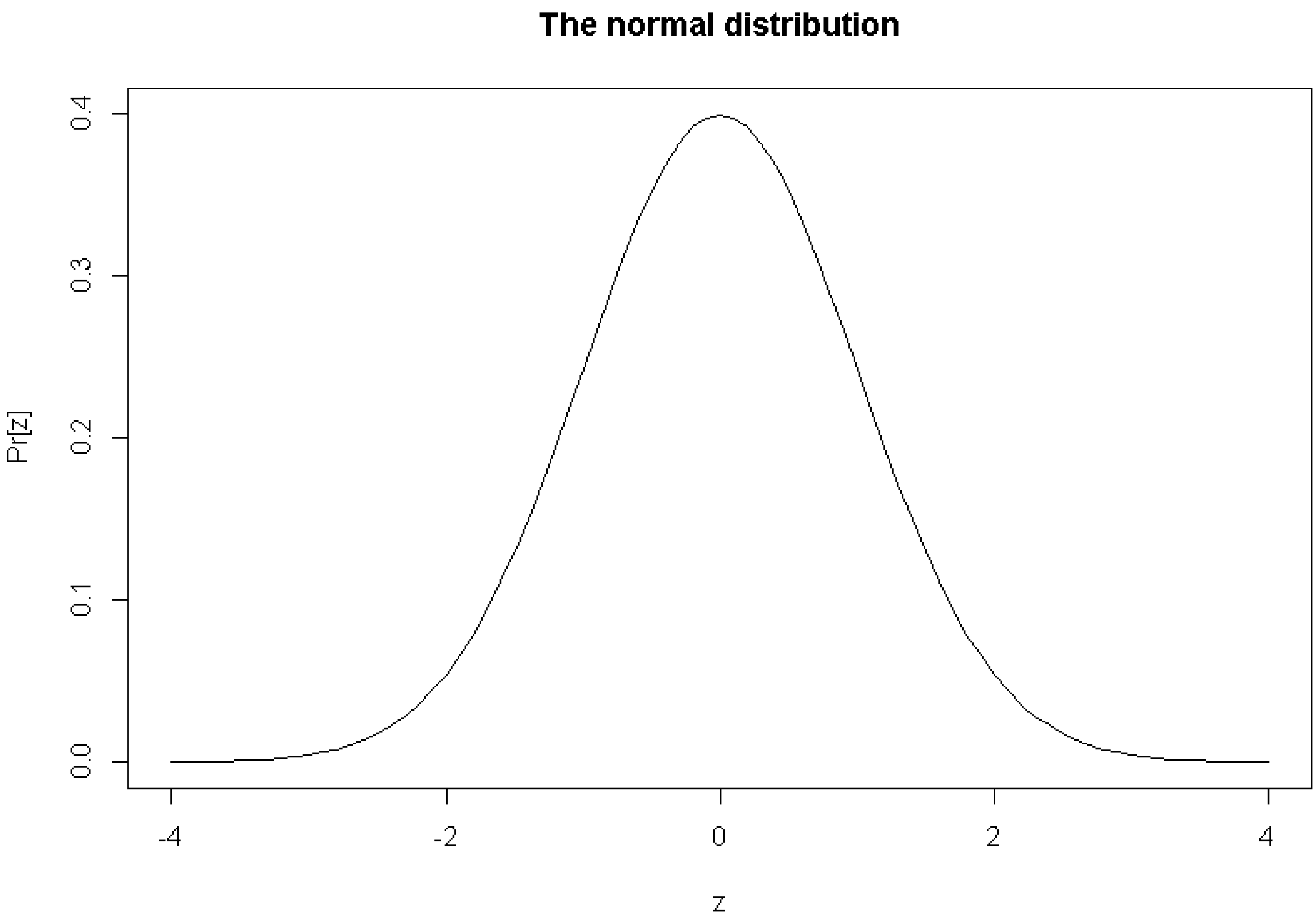
Stature and the normal distribution



Z

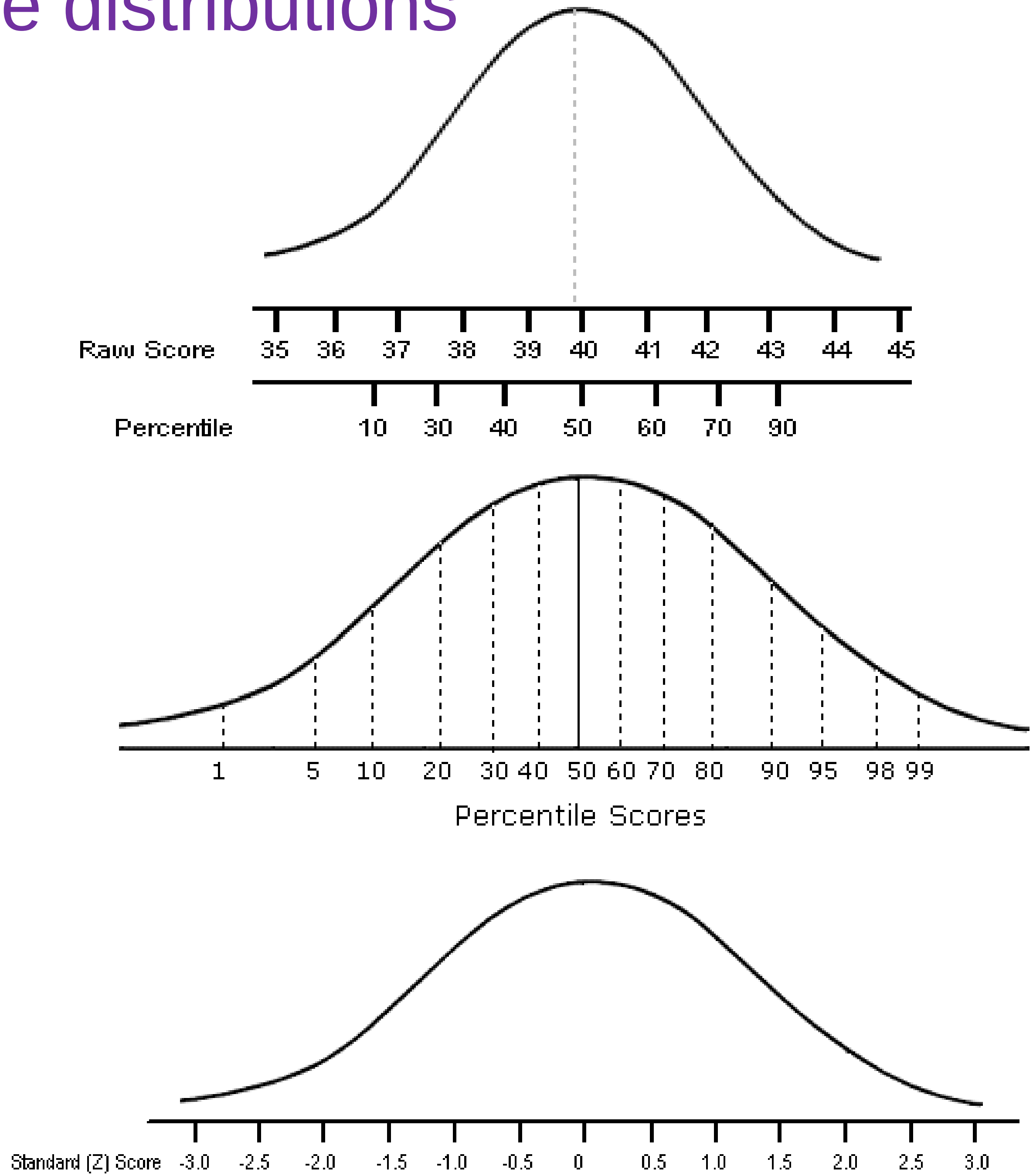
Figure 1.24, INTRODUCTION to the PRACTICE of STATISTICS, © 2014 W. H. Freeman

The Normal Distribution




```
x <- seq(-4,4,0.1)
# dnorm gives the normal probability density for a value
plot(x,dnorm(x), type="l",xlab="z",ylab="Pr[z]", main="The standard normal distribution")
```

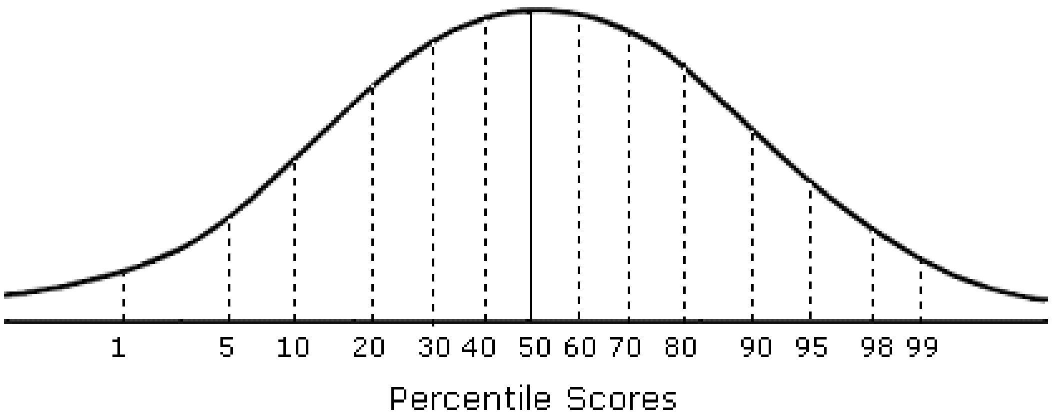
Sample distributions

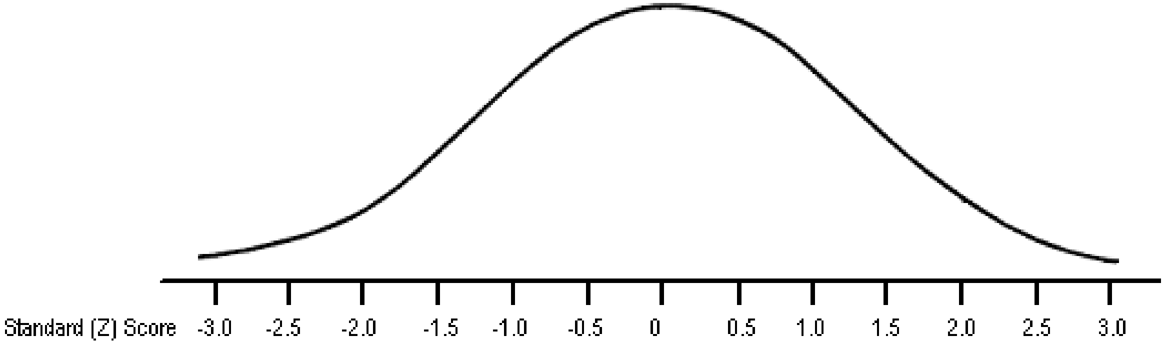


Sample distributions are probability densities

With normally distributed data, the mean is also the mode
The mode is the most common value, the most likely to be randomly picked
The mean is the expected random value

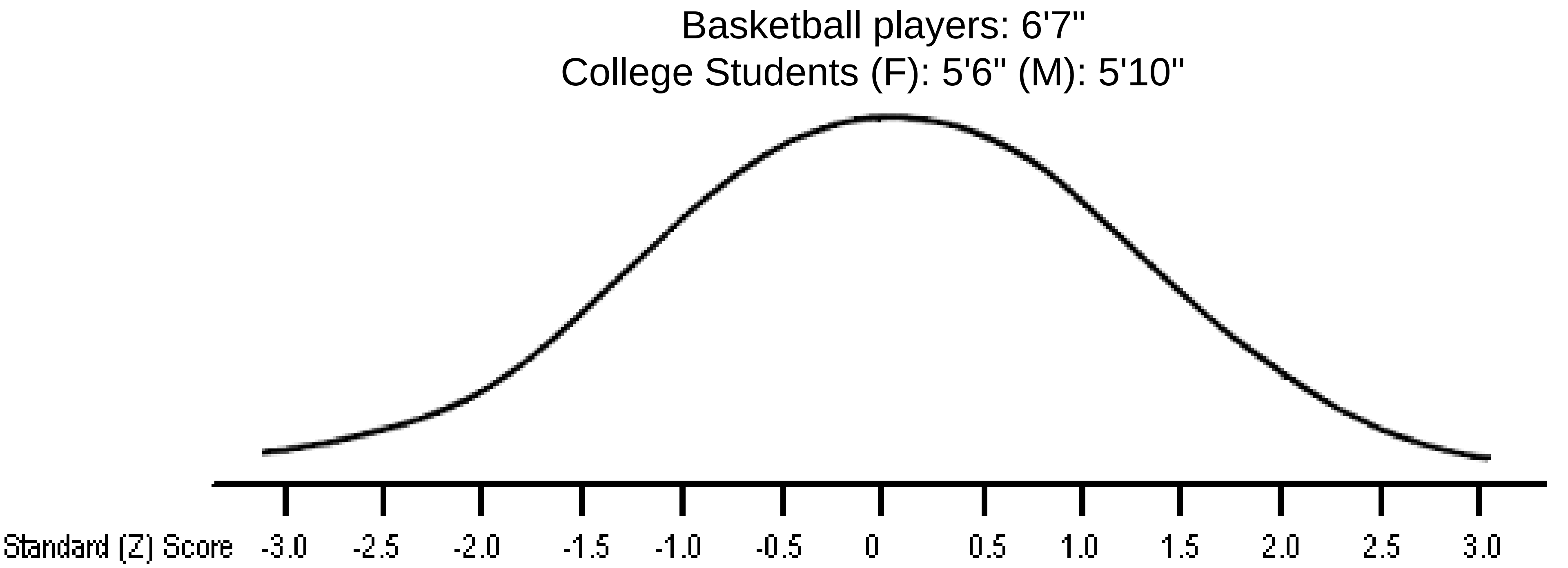
Basketball players: 6'7"
College Students (F): 5'6" (M): 5'10"





Sample distributions are probability densities

There are many individuals near the mean, they are more dense near it
How do we quantify "near the mean" ?
Using the standard deviation and Z score



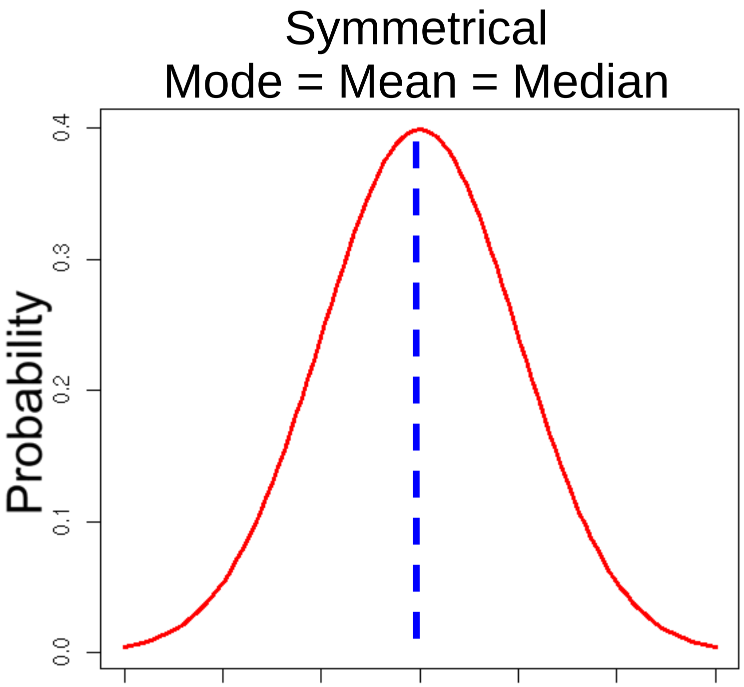
The normal distribution is a PDF (probability density function)

Area under curve = probability (total = 1)

The most probable value of a randomly picked individual is the mean

We know that there are many individuals NEAR the mean that could be picked

We also know that 50% of the sample will be less than the mean.



Normal Probability Models

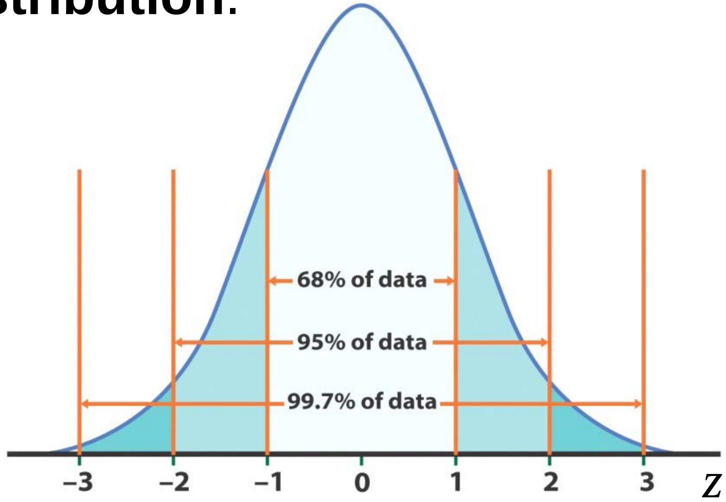
Often, the density curve used to assign probabilities to intervals of outcomes is the Normal curve.

<u>Normal distributions are probability models</u>:

Probabilities can be assigned to intervals of outcomes using the **Standard Normal probabilities**.

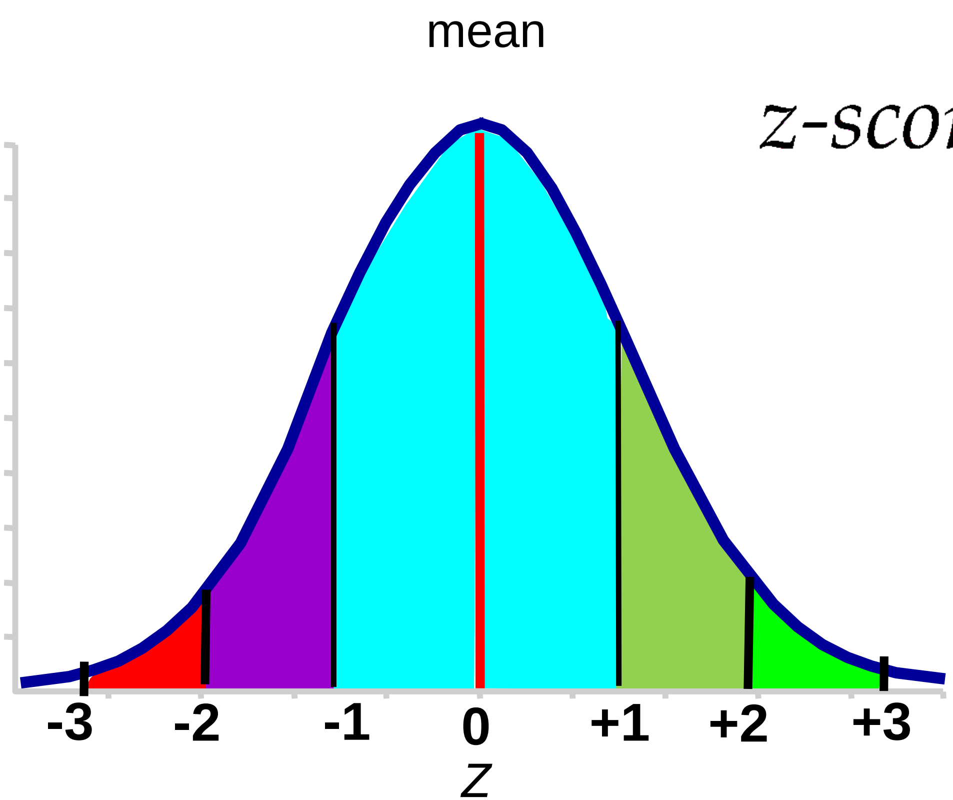
We standardize normal data by calculating *z*-scores so that any Normal curve $N(\mu,\sigma)$ can be transformed into the standard Normal curve $N(0,1)$ called the **standard normal distribution**.

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



The Standard Deviation and one-sided probabilities

$$z\text{-}score = \frac{x_i - \bar{x}}{s}$$

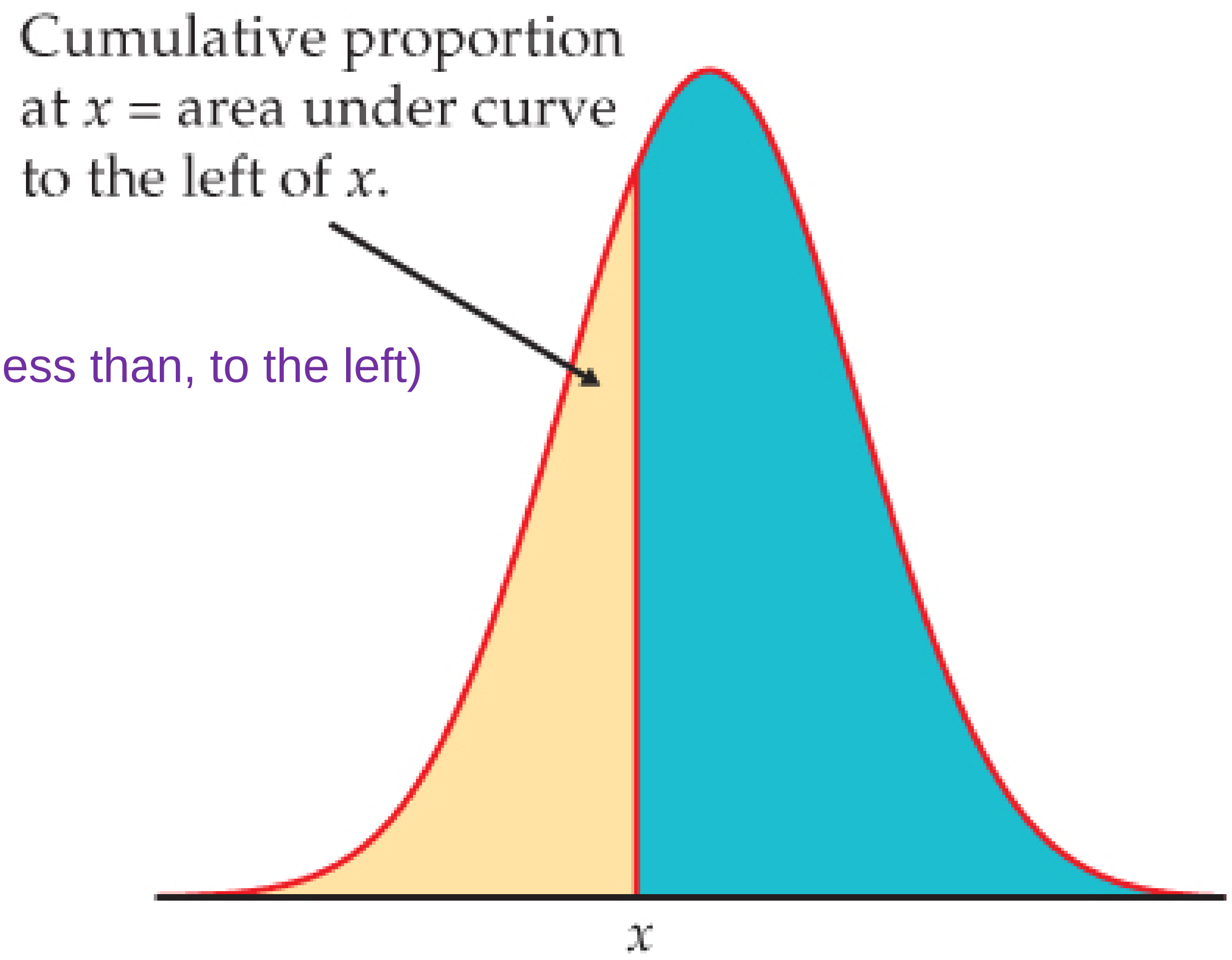


16 % | +/- 1 SD = 68 % | 16 %

2.5 % | +/- 2 SD = 95 % | 2.5 %

0.5 % | +/- 3 SD = 99 % | 0.5 %

Cumulative density function
Cumulative distribution function (cdf)
"proportion to the left"

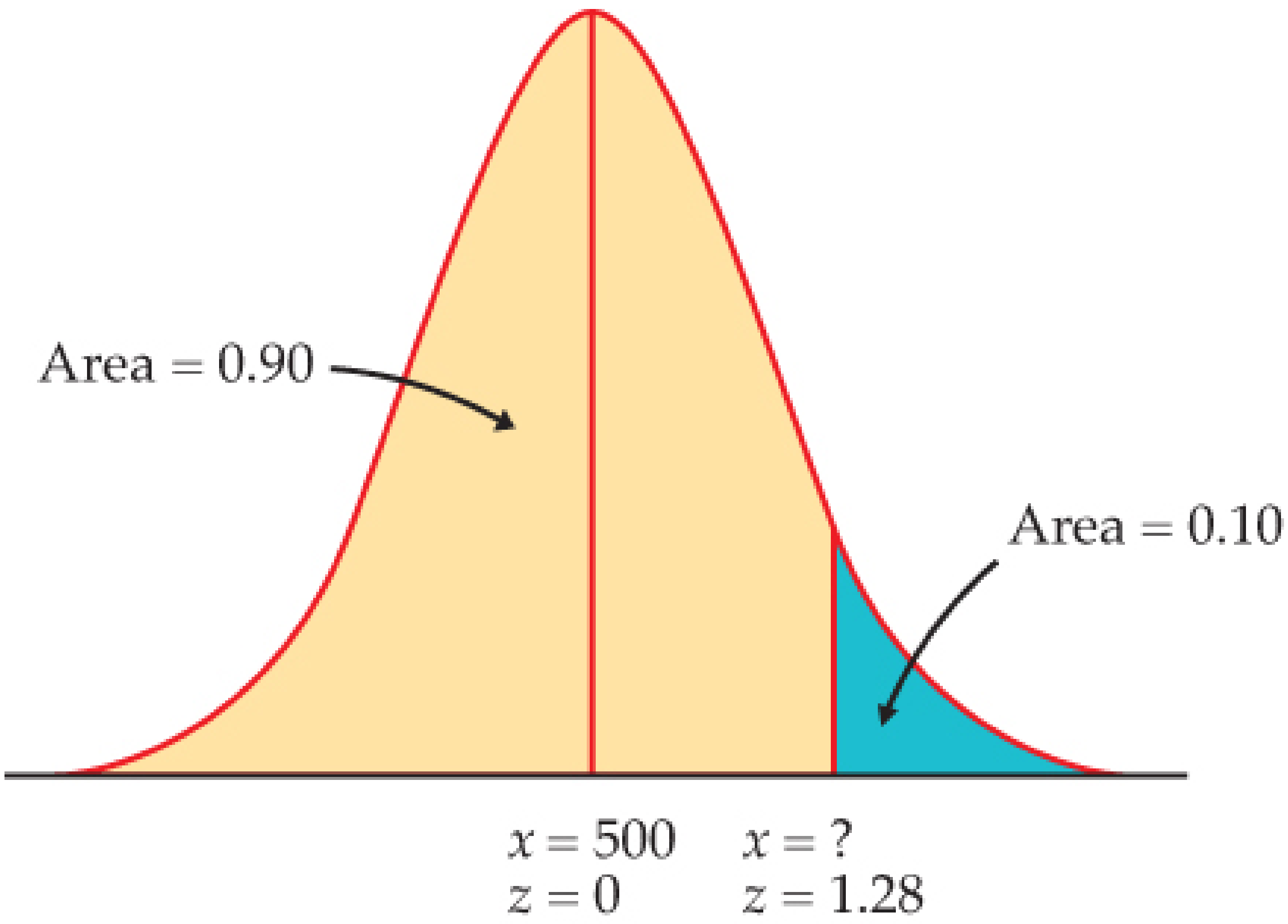


(less than, to the left)

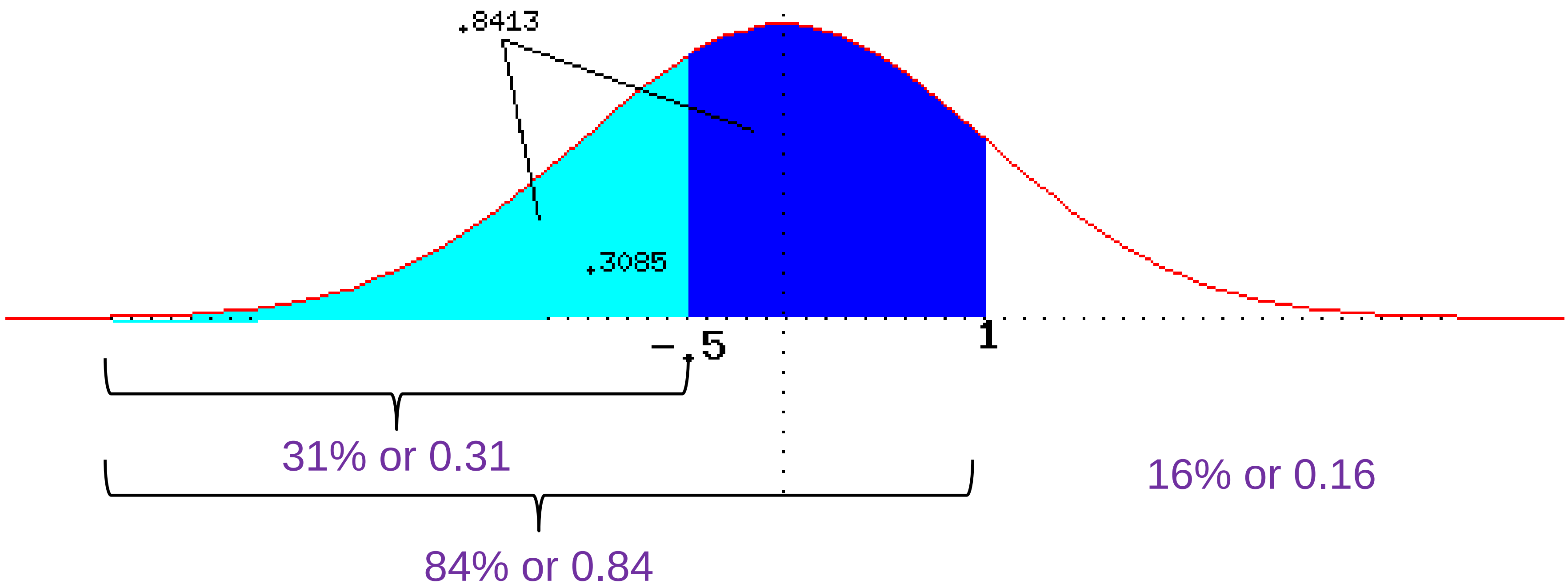
Figure 1.26, INTRODUCTION to the PRACTICE of STATISTICS, © 2014 W. H. Freeman

1 minus Cumulative distribution/density

1 - proportion to the left = **proportion to the right**



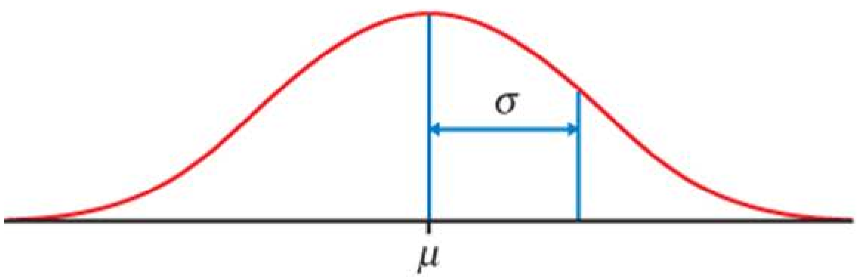
With *Z*, every proportion can be calculated

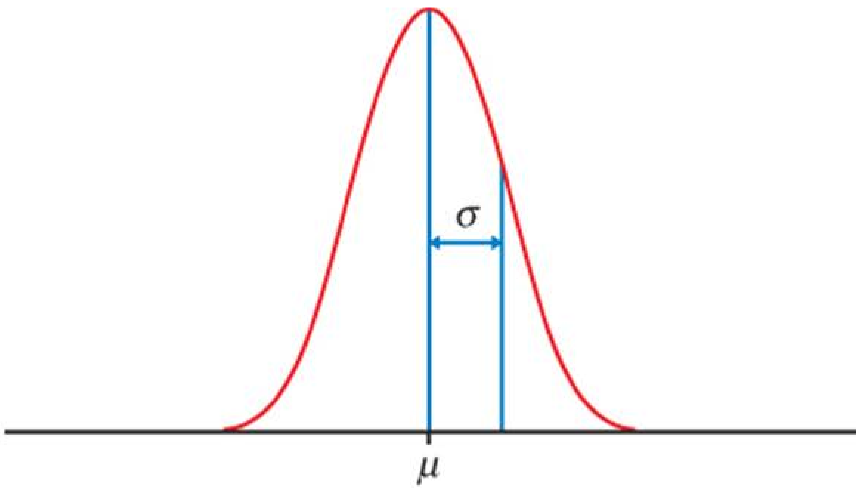



```
Total area = 1
Cumulative proportion less than Z =   1  = 0.84 or 84%
Cumulative proportion less than Z = -0.5 = 0.31 or 31%

Proportion in interval between -0.5 and 1:
0.84 - 0.31 = 0.53

(Proportion > Z = 1: 1 - 0.84 = 0.16 )
```

All proportions will be the same using *Z*

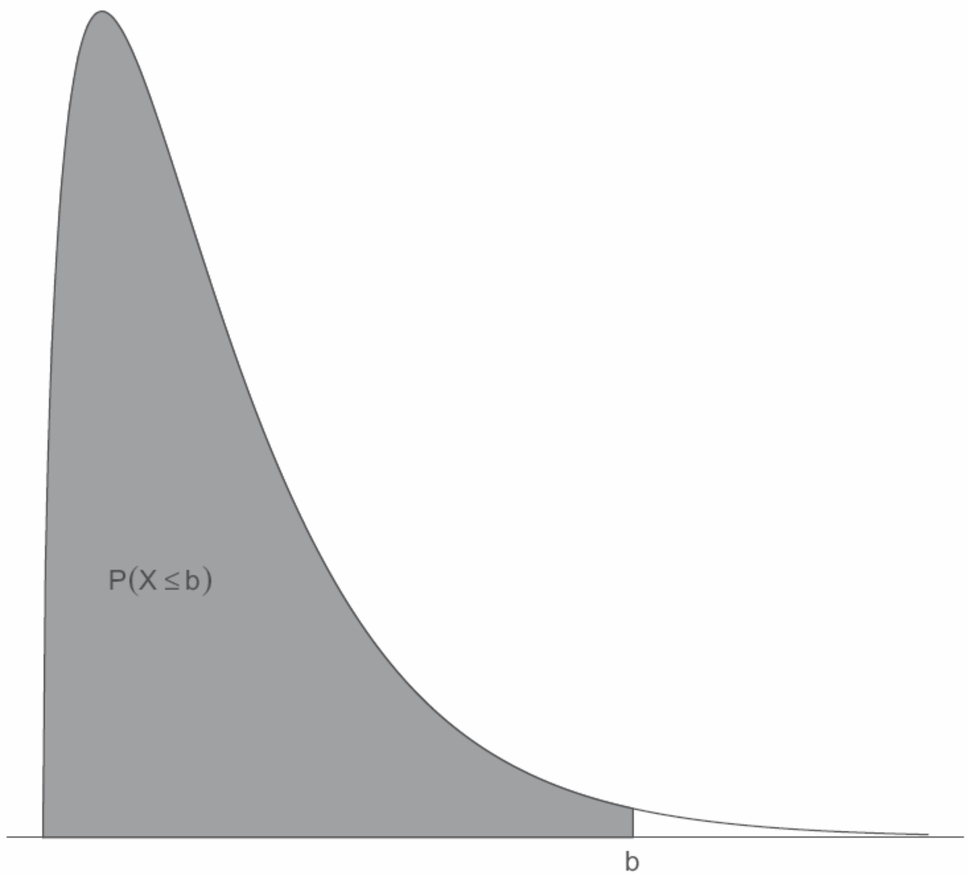
if the data are normally distributed
- meeting the equation requirement
- symmetrical, bell-shaped, unimodal





N(μ, σ) is a "recipe" for making a normal distribution

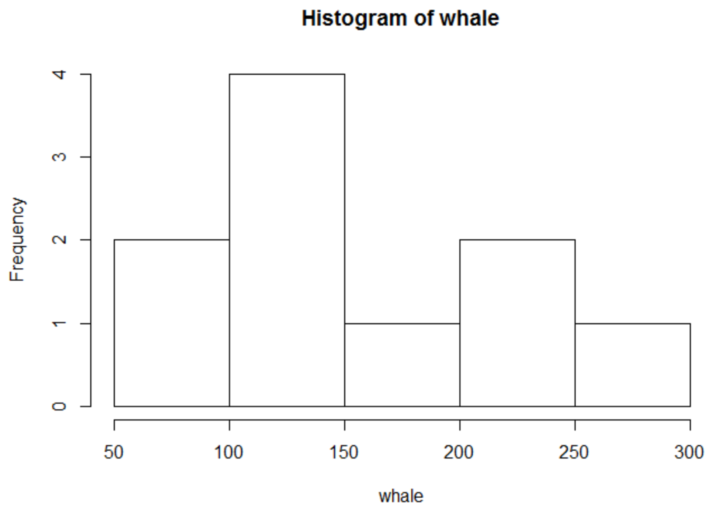
Another sample distribution



Only a few have values > b

Calculating Z scores in R

```
whale <- c(74, 122, 235, 111, 292, 111, 211, 133, 156, 79)
whale
 [1]  74 122 235 111 292 111 211 133 156  79

sd(whale)
[1] 71.50789
mean(whale)
[1] 152.4

# get difference from mean
whale-mean(whale)
 [1] -78.4 -30.4  82.6 -41.4 139.6 -41.4  58.6 -19.4   3.6 -73.4
```



How will we get the z score?

$$z\text{-}score = \frac{x_i - \bar{x}}{s}$$

Calculating *Z* scores in R

```
whale
 [1]  74 122 235 111 292 111 211 133 156  79
```

$$z\text{-}score = \frac{x_i - \bar{x}}{s}$$

```
#simply divide the differences by the SD
whale - mean(whale)/ sd(whale) # do these look correct?
 [1]  71.86877 119.86877 232.86877 108.86877 289.86877 108.86877 208.86877 130.86877
 [9] 153.86877  76.86877
```

Calculating Z scores in R

$$z\text{-}score = \frac{x_i - \bar{x}}{s}$$

```
whale
 [1]  74 122 235 111 292 111 211 133 156  79

#simply divide the differences by the SD
whale - mean(whale)/ sd(whale) # do these look correct?
 [1]  71.86877 119.86877 232.86877 108.86877 289.86877 108.86877 208.86877 130.86877
 [9] 153.86877  76.86877

(whale - mean(whale) ) / sd(whale)  # these numbers look MUCH better!
 [1] -1.0963826 -0.4251279  1.1551174 -0.5789571  1.9522322 -0.5789571  0.8194900
 [8] -0.2712987  0.0503441 -1.0264602

# how to check?  look at first one
#  get Z score here                    multiply by sd     add mean
(whale[1]-mean(whale))/sd(whale)  * sd(whale)     + mean(whale)
[1]  74
```

Calculating *Z* scores in R

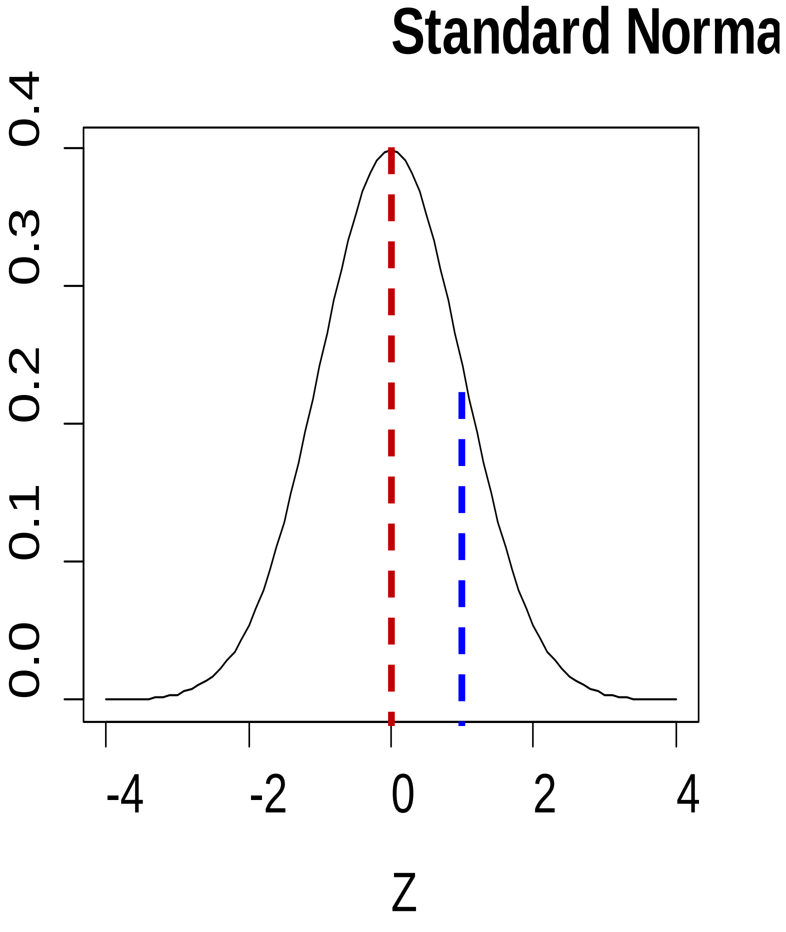
```
# another way to get standard normal Z scores: the scale() function
scale(whale)
            [,1]
 [1,] -1.0963826
 [2,] -0.4251279
 [3,]  1.1551174
 [4,] -0.5789571
 [5,]  1.9522322
 [6,] -0.5789571
 [7,]  0.8194900
 [8,] -0.2712987
 [9,]  0.0503441
[10,] -1.0264602
attr(,"scaled:center") (mean)
[1] 152.4
attr(,"scaled:scale") (sd)
[1] 71.50789
```

```
# get a single Z score
scale(whale)[1,1]
[1] -1.096383
```

```
# looks odd (a dataframe)
# convert to a vector using as.vector() function
as.vector(scale(whale))
 [1] -1.0963826 -0.4251279  1.1551174 -0.5789571  1.9522322 -0.5789571
 [7]  0.8194900 -0.2712987  0.0503441 -1.0264602
```

The Cumulative Distribution Function (CDF) OR The Cumulative Probability Function OR the Cumulative proportion

(proportion of sample less than a value)



Simplest case:
50% < mean

- can calculate
  for any Z

We need only specify:
μ and σ

Getting Probability Values in R:

pnorm (normal distribution)

The "p" functions returns the cumulative probability for a specific value

pnorm(value,mu,sigma)

```
pnorm(0) # if one value given, standard normal N(0,1);
 [1] 0.5

pnorm (1) # standard normal;
[1] 0.8413447

pnorm(1.65) # standard normal;
[1] 0.9505285

pnorm (2);
[1] 0.9772499

pnorm(59,50,4.5) # Prob < x8 in Z score example; mu = 50
# (59-50)/4.5 = 2
[1] 0.9772499
```

Getting Probability Values in R:

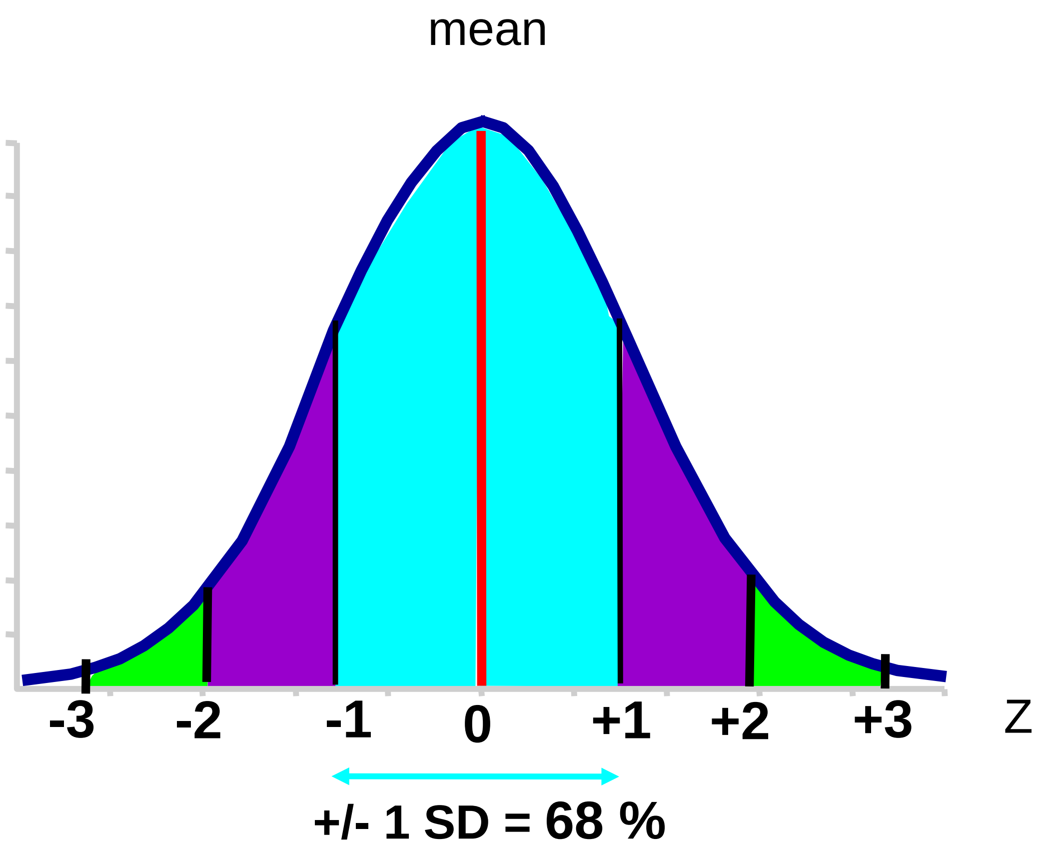
qnorm (normal distribution)

The "q" functions return the value for a quantile (0 to 1)

qnorm(quantile,mu,sigma)

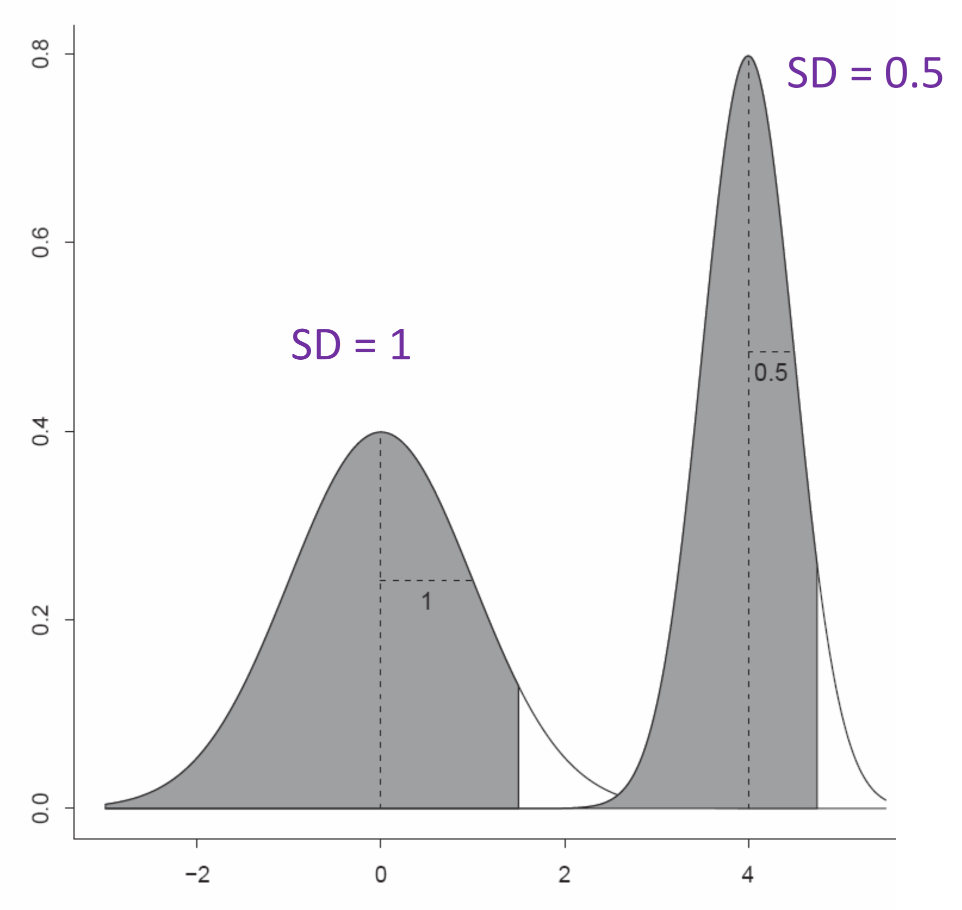
```
qnorm(0.25) # 25th percentile, 1st quartile, N(0,1);
[1] -0.6744898

qnorm(0.50) # 50th percentile, median;
[1] 0

qnorm (0.9772499) # #~97.5 ;
[1] 2.000001

qnorm (0.975) # EXACT 97.5 (equals 1.96);
[1] 1.959964

# Get value at which 97.5 % is lower;mu = 50, sigma = 4.5
qnorm(0.975,50,4.5)
[1] 58.81984
```

The Standard Normal Distribution




```
# probability/proportion/area between -1 and 1 (Z)
pnorm(1)-pnorm(-1)
[1] 0.6826895
diff(pnorm(c(-2,2)))  # between -2 and 2 using diff function
[1] 0.9544997
```

Standard normal distribution and probability




```
# cumulative prob density at Z = 1
pnorm(1, mean=0, sd=1)
[1] 0.8413
pnorm(4.5, mean=4, sd=0.5)   # same z-score as above
[1] 0.8413
```

Stature and the normal distribution

In a certain group of women, the mean stature is 64" and the SD is 3"

What proportion is **taller** than 70 "?

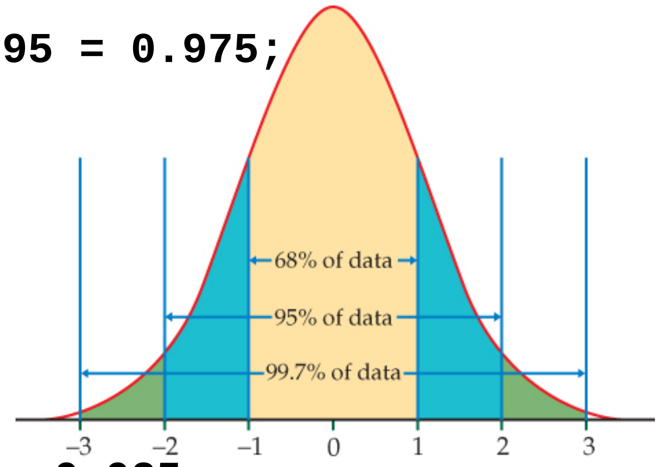
```
How to figure out?
Convert to a Z score.

(70-64)/3 #get the Z score.
[1] 2
+/- 2 Z includes 95%; cdf (Z = 2) = 0.025 + 0.95 = 0.975;
1 - 0.975 = 0.025 or 2.5% that are at least 70" tall
```

What proportion of women is **shorter** than 70"?

```
(70-64) / 3 = 2 = Z;
+/- 2 Z includes 95%; cdf (Z = 2) = 0.025 + 0.95 = 0.975;
 (97.5% are shorter than 70")
```

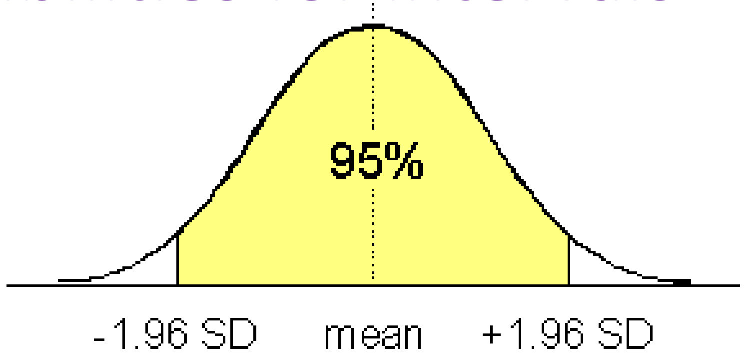
What proportion is shorter than 58"?

```
How to figure out?
Convert to a Z score.

(58-64) / 3 = -2;  +/- 2 includes 95%, so cdf = 0.025
```

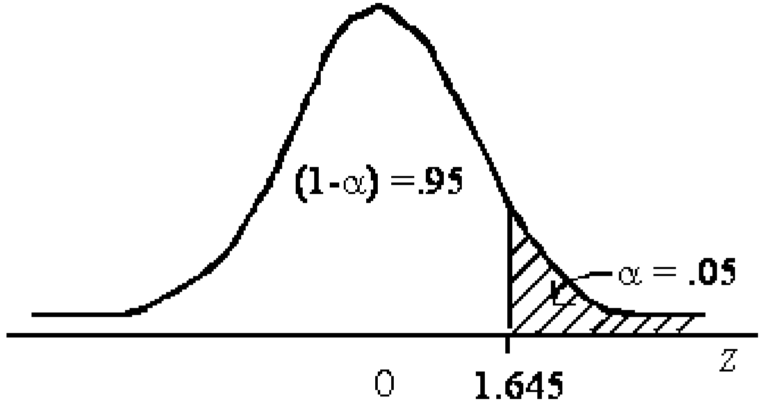


Continuous distributions: We can only get probabilities for intervals

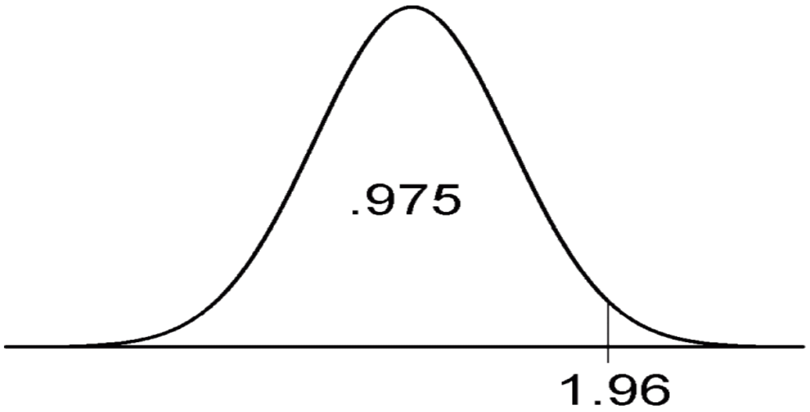
Intervals
- around mean
95% = mean +/-1.96 SD



-one-sided (upper)
95% = mean + 1.645 SD

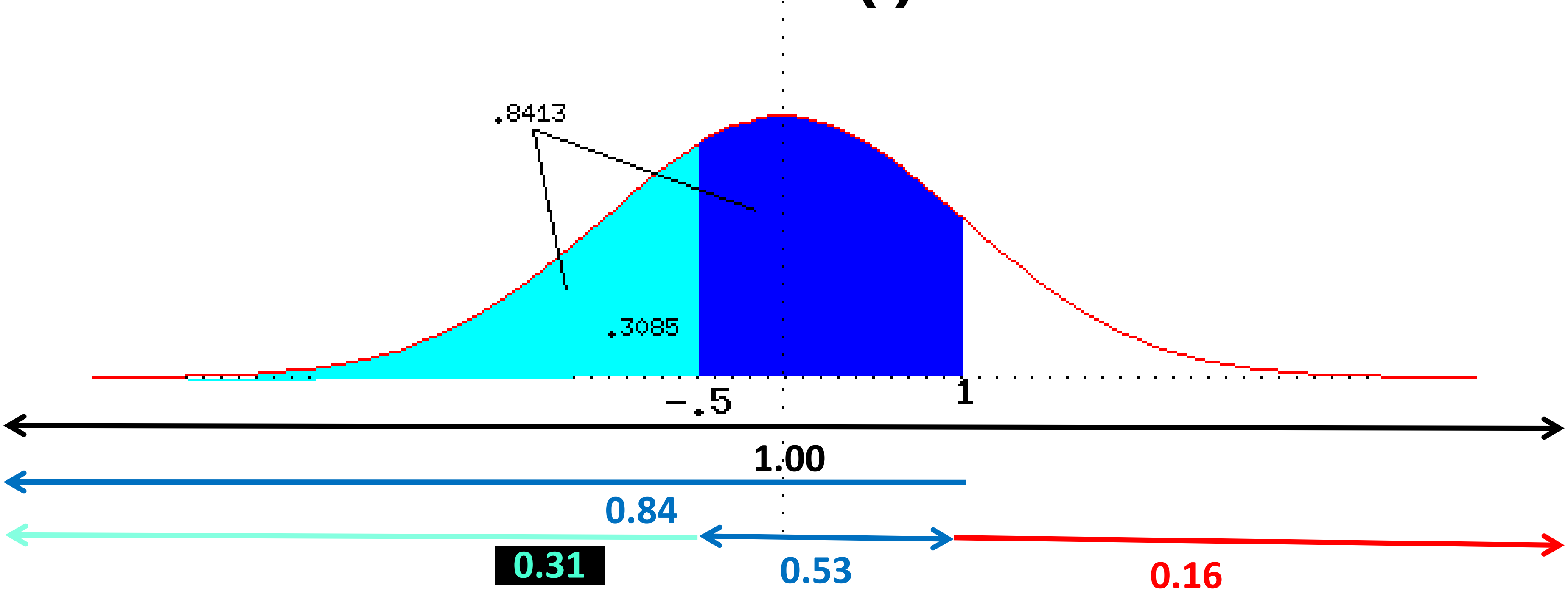


-one-sided (upper)
97.5 % = mean + 1.96 * SD



Getting normal probabilities in R:

`rnorm()`




```
# prob < Z of 1
pnorm(1)
[1] 0.8413447
pnorm(-0.5) # prob < Z of -0.5
[1] 0.3085375

# *** NOTE ****
#to get intervals, subtract one from the other
pnorm(1)-pnorm(-0.5) # prob between the two
[1] 0.5328072
```

Normal Probability Models

Often the density curve used to assign probabilities to intervals of outcomes is the Normal curve.

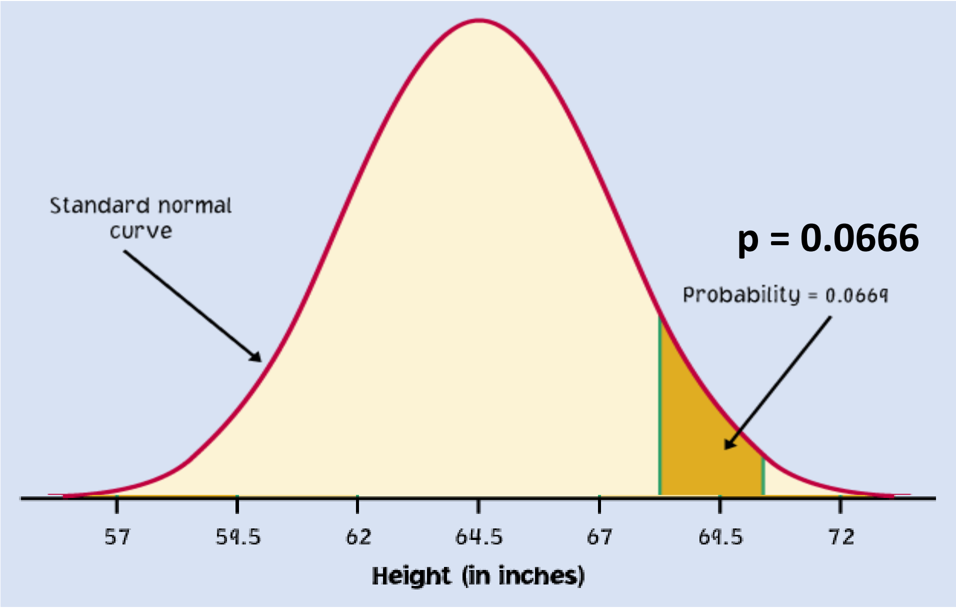
Women's heights are Normally distributed with mean 64.5 and standard deviation 2.5 in. If we pick one woman at random, what is the probability that her height will be between 68 and 70 inches $P(68 < X < 70)$? Because the woman is selected at random, X is a random variable.

$$z = \frac{(x - \mu)}{\sigma}$$

As before, we calculate the *z*-scores for 68 and 70.

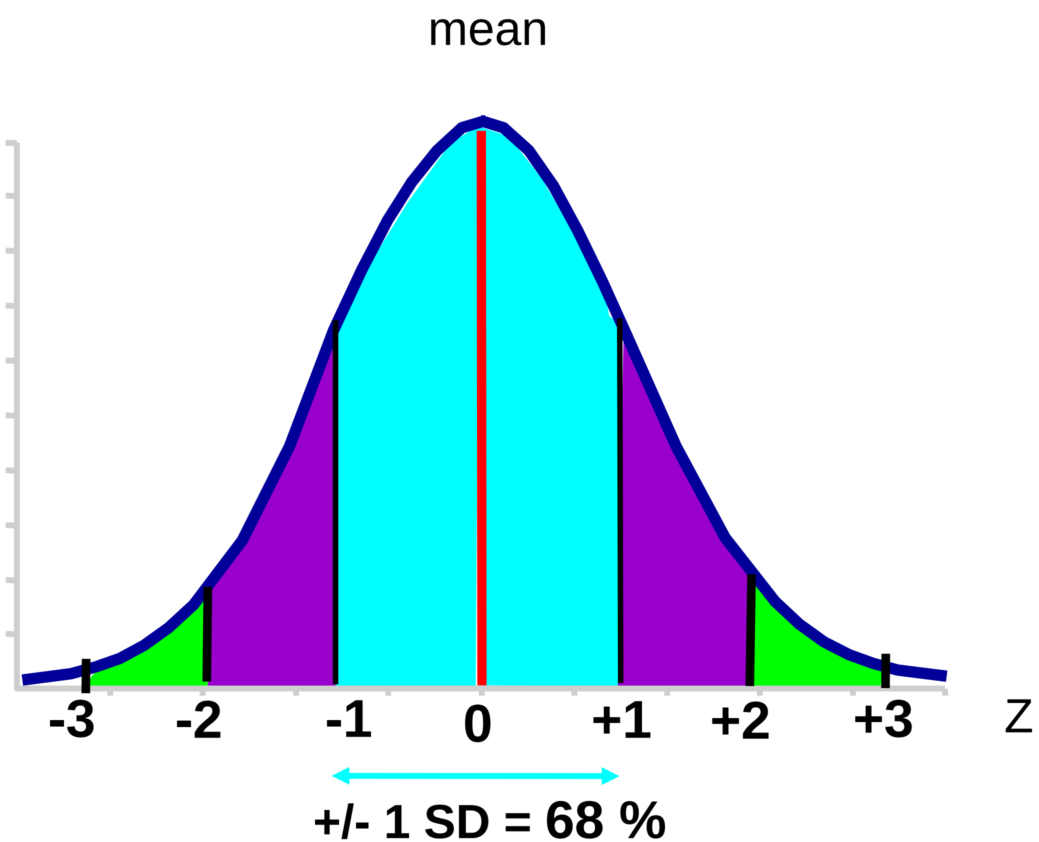
For x = 68", $$z = \frac{(68 - 64.5)}{2.5} = 1.4$$

For x = 70", $$z = \frac{(70 - 64.5)}{2.5} = 2.2$$




```
pnorm(2.2)- pnorm(1.4) # p between the two Z scores
[1] 0.06685321
```

The Standard Normal Distribution




```
# probability/proportion/area between -1 and 1 (Z)
pnorm(1)-pnorm(-1)
[1] 0.6826895
diff(pnorm(c(-2,2)))  # between -2 and 2 using diff function
[1] 0.9544997
```

Stature and the normal distribution

In a certain group, the average stature is 70.2 inches and the SD is 2.89 inches

What proportion of the population is shorter than 6 feet tall?

```
mu <- 70.2;
sigma <- 2.89;
pnorm(72, mean=mu, sd=sigma);  # cumulative;
[1] 0.7333
```

What proportion is taller than 6 feet?

```
# right tail, greater than, so subtract from 1;
1-pnorm(72, mean=mu, sd=sigma);
[1] 0.2666957
```

What proportion is taller than 2 meters?

```
# 1 meter = 39.37 inches; 1 inch = 0.0254 meters;
# 1 - again!;
1-pnorm((2*39.37), mean=mu, sd=sigma);
[1] 0.001563258
```

Homework 6

6.1 If Z is Normal(0,1), find the following:

1. $P(Z < 2.2)$.
2. $P(-1 < Z < 2)$.
3. $P(Z > 2.5)$.
4. b such that $P(-b < Z \; b) = 0.90$.

6.2 Suppose that the population of adult, male black bears has weights that are approximately distributed as Normal(350,75). What is the probability that a randomly observed male bear weighs more than 450 pounds?

Homework 6

6.3 A study found that foot lengths for Japanese women are normally distributed with mean 24.9 centimeters and standard deviation 1.05 centimeters.

For this population, find the probability that a randomly chosen foot is less than 26 centimeters long. What is the 95th percentile?

6.4 Assume that the average finger length for females is 3.20 inches, with a standard deviation of 0.35 inches, and that the distribution of lengths is normal. If a glove manufacturer makes a glove that fits fingers with lengths between 3.5 and 4 inches, what percent of the population will the glove fit?

Homework 6

6.5 Cereal is sold by weight not volume. This introduces variability in the volume due to settling. As such, the height to which a cereal box is filled is random. If the heights for a certain type of cereal and box have a Normal(12, 0.5) distribution in units of inches, what is the chance that a randomly chosen cereal box has cereal height of 10.7 inches or less?

DUE before class February 15