5 Continuous Probability Distribution

Normal Probability Distributions

- Key words:

Continuous random variable, normal distribution , standard normal distribution , T-distribution

Properties of continuous probability Distributions:

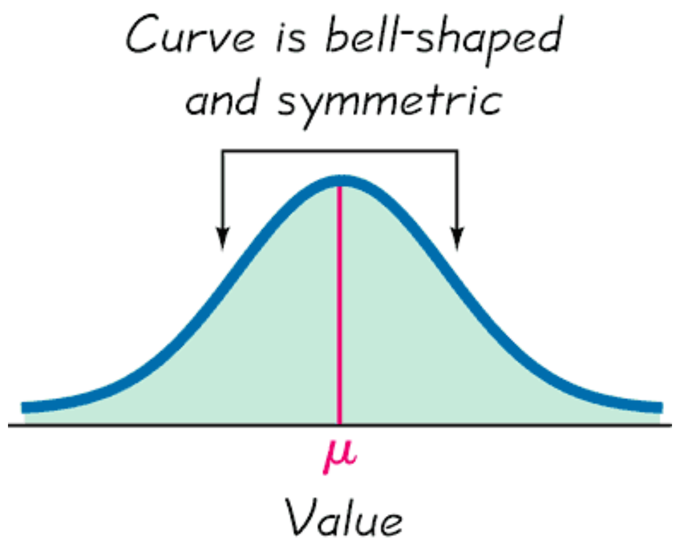
1- Area under the curve = 1.

2- $P(X = a) = 0$, where a is a constant.

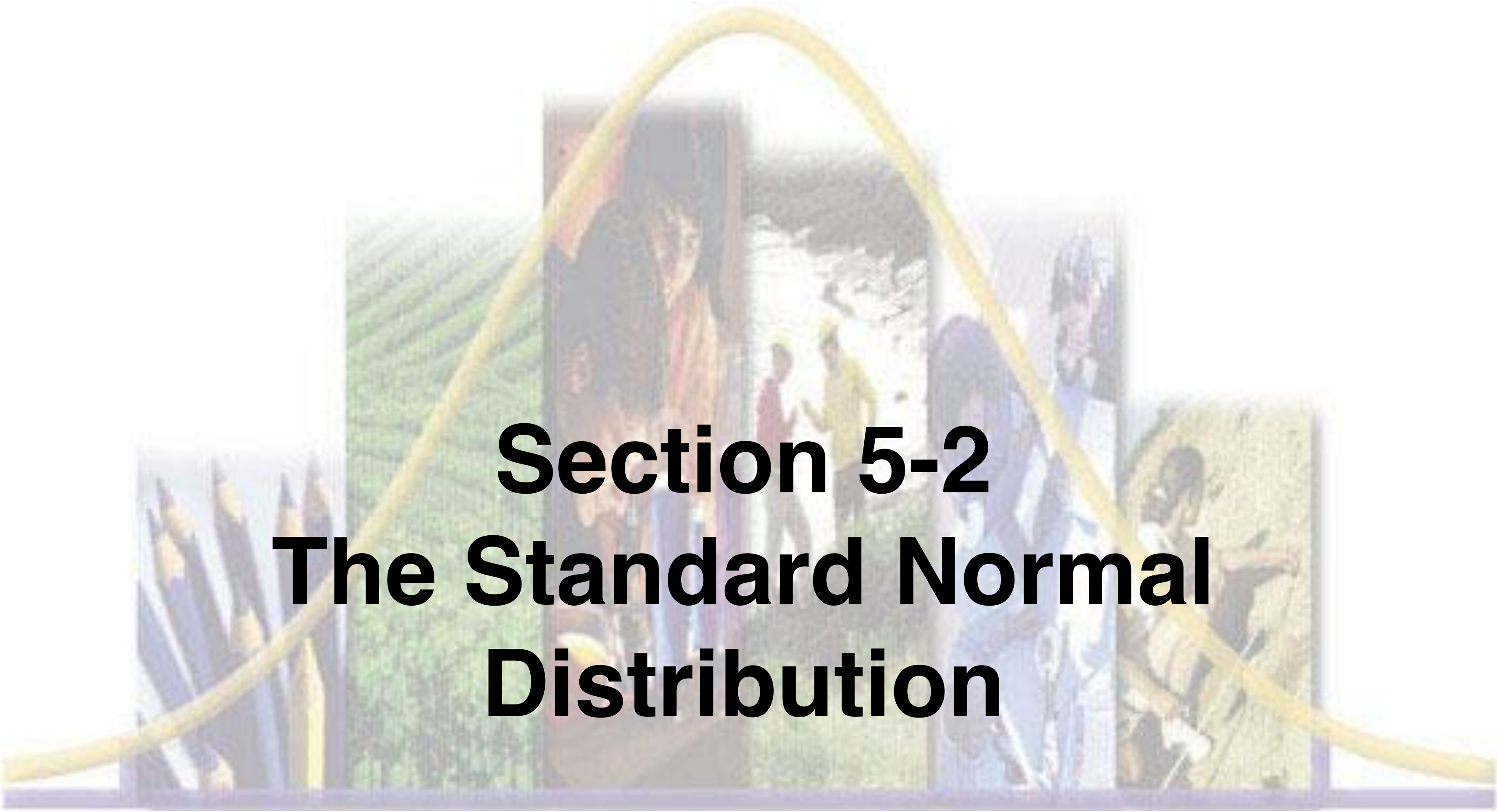
3- Area between two points a , b = $P(a<x<b)$.

Overview

- Continuous random variable
- Normal distribution P(X=x) = f(x)



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



Section 5-2
The Standard Normal Distribution

Definitions

- **Uniform Distribution** is a probability distribution in which the continuous random variable values are spread evenly over the range of possibilities; the graph results in a rectangular shape.

长方形

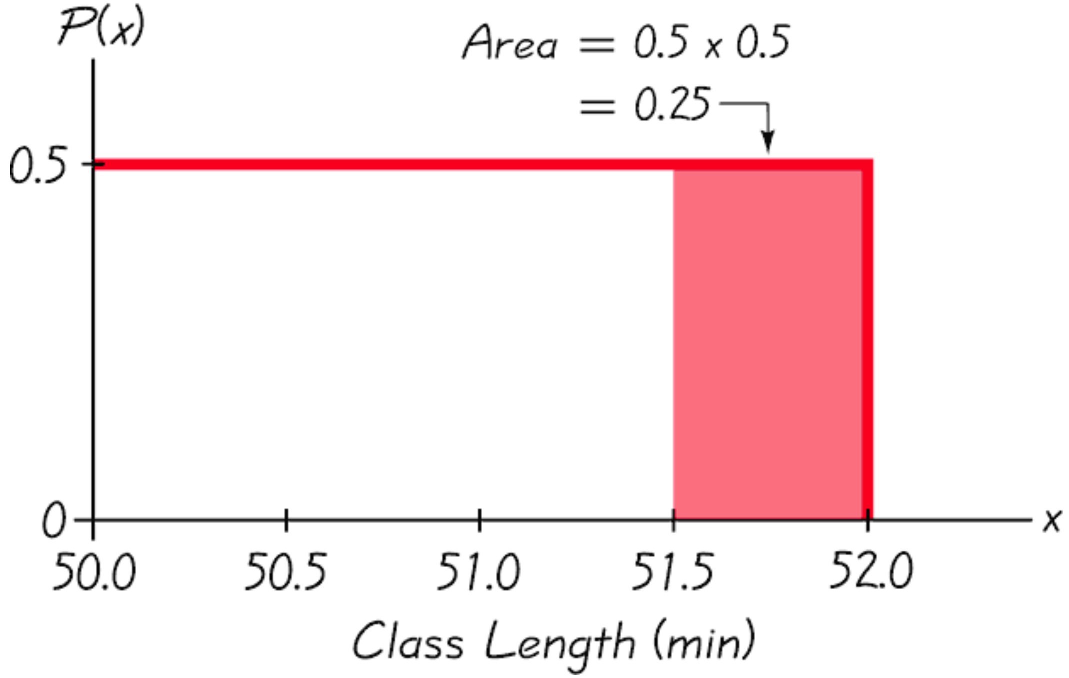
Definitions

- **Density Curve** (or probability density function is the graph of a continuous probability distribution.

1. The total area under the curve must equal 1.

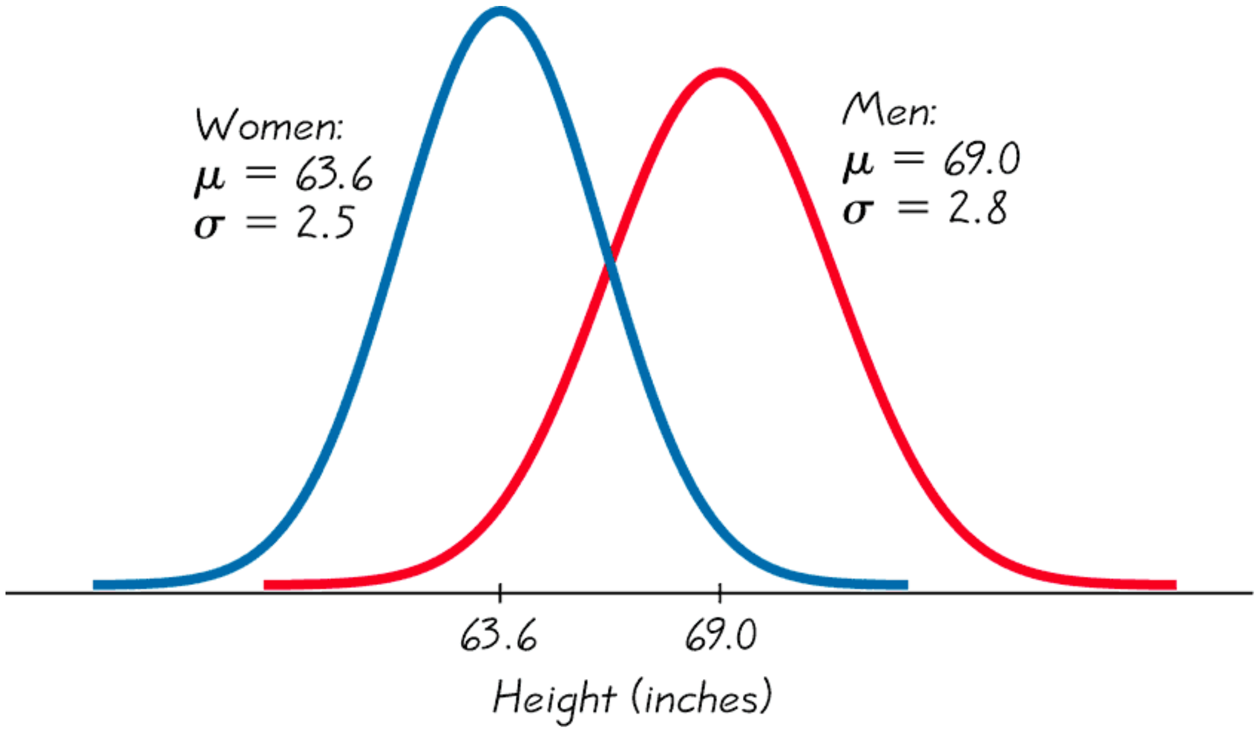
2. Every point on the curve must have a vertical height that is 0 or greater.

Using Area to Find Probability



Uniform distribution

Heights of Adult Men and Women



Law of large numbers

Let $\{X_1, ..., X_n\}$ be a random sample of size n, that is, a sequence of independent and identically distributed random variables drawn from distributions of expected values given by μ and finite variances given by σ^2. Suppose we are interested in the sample average

$$S_n := \frac{X_1 + \cdots + X_n}{n}$$

of these random variables. By the law of large numbers, the sample averages converge in probability and almost surely to the expected value μ as $n \to \infty$.

Central limit theorem

For large enough *n*, the distribution of S_n is close to the normal distribution with mean μ and variance σ^2/n.

$$S_n := \frac{X_1 + \cdots + X_n}{n}$$

Definition

Standard Normal Distribution:

a normal probability distribution that has a **mean of 0** and a **standard deviation of 1,** and the **total area under its density curve is equal to 1.**

Definition

Standard Normal Distribution:

a normal probability distribution that has a **mean of 0** and a **standard deviation of 1.**

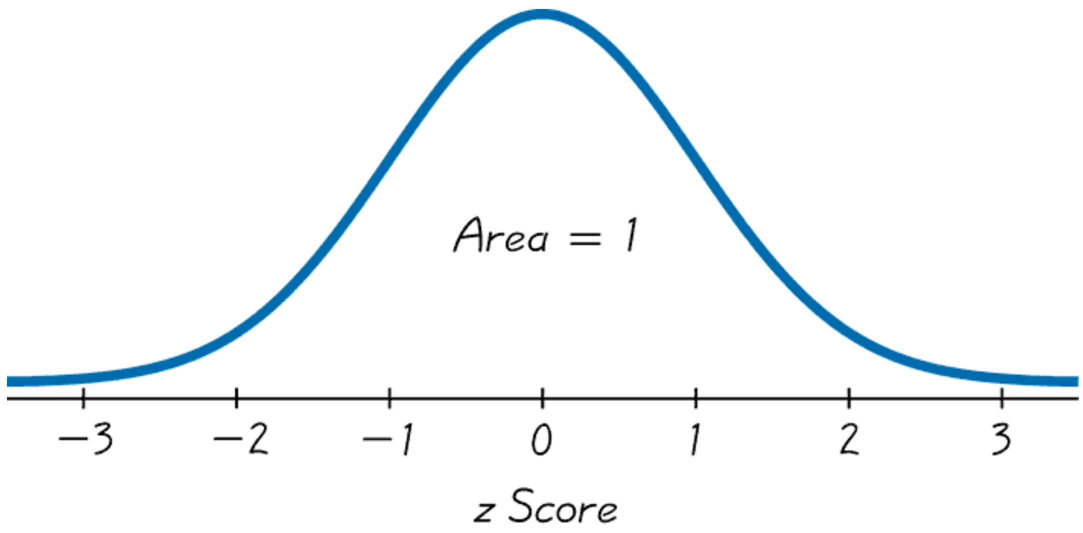


TABLE A-2	Standard Normal (z) Distribution: Cumulative Area from the LEFT									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.50 and lower	.0001									
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	* .0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	* .0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559

Notation

P($a < z < b$)

denotes the probability that the z score is between a and b

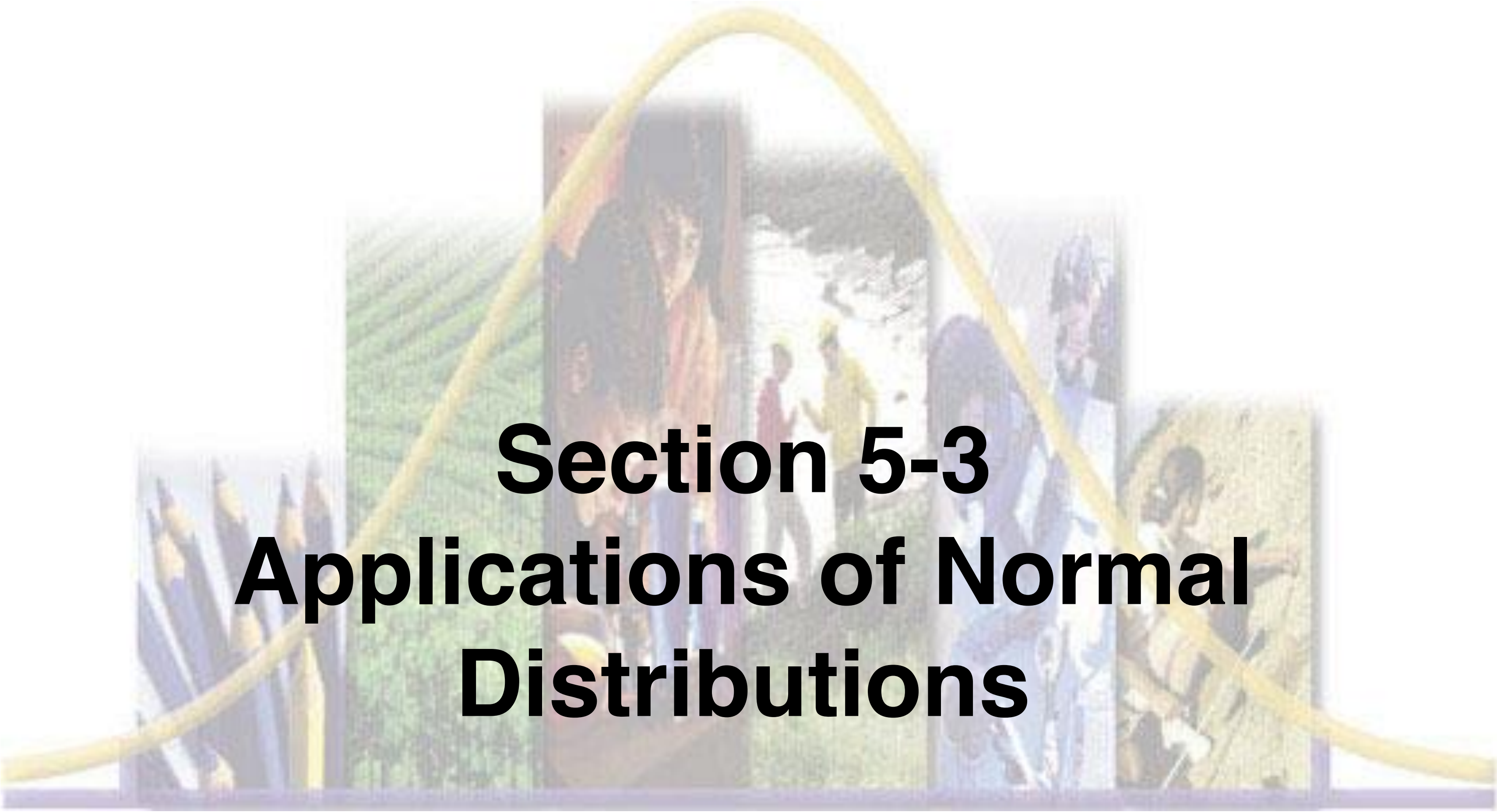
P($z > a$)

denotes the probability that the z score is greater than a

P($z < a$)

denotes the probability that the z score is less than a

Section 5-3
Applications of Normal Distributions

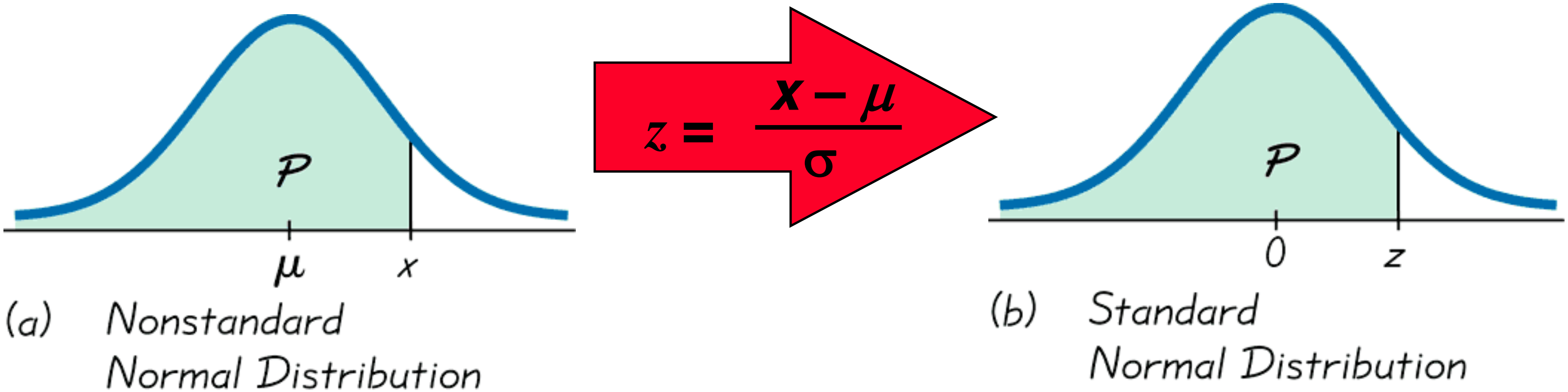


Nonstandard Normal Distributions

If $\mu \neq 0$ or $\sigma \neq 1$ (or both), we will convert values to standard scores using Formula, then procedures for working with all normal distributions are the same as those for the standard normal distribution.

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

Converting to Standard Normal Distribution



6 Sampling Distributions and Estimators

6-1 Sampling Distributions and Estimators

6-2 The Central Limit Theorem

6-3 Normal as Approximation to Binomial

6-4 Determining Normality

(Three important distributions)

Definition

Sampling Distribution of the mean is the probability distribution of sample means, with all samples having the same sample size n.

Definition

Sampling Variability:

The value of a statistic, such as the sample mean $\bar{x}$, depends on the particular values included in the sample.

Definition

The Sampling Distribution of the Proportion is the probability distribution of sample proportions, with all samples having the same sample size *n*.

Sampling Distributions

A population consists of the values 1, 2, and 5. We randomly select samples of size 2 with replacement. There are 9 possible samples.

a. For each sample, find the mean, median, range, variance, and standard deviation.

b. For each statistic, find the mean from part (a)

Table 5-2 Sampling Distributions of Different Statistics (for Samples of Size 2 Drawn with Replacement from the Population 1, 2, 5)

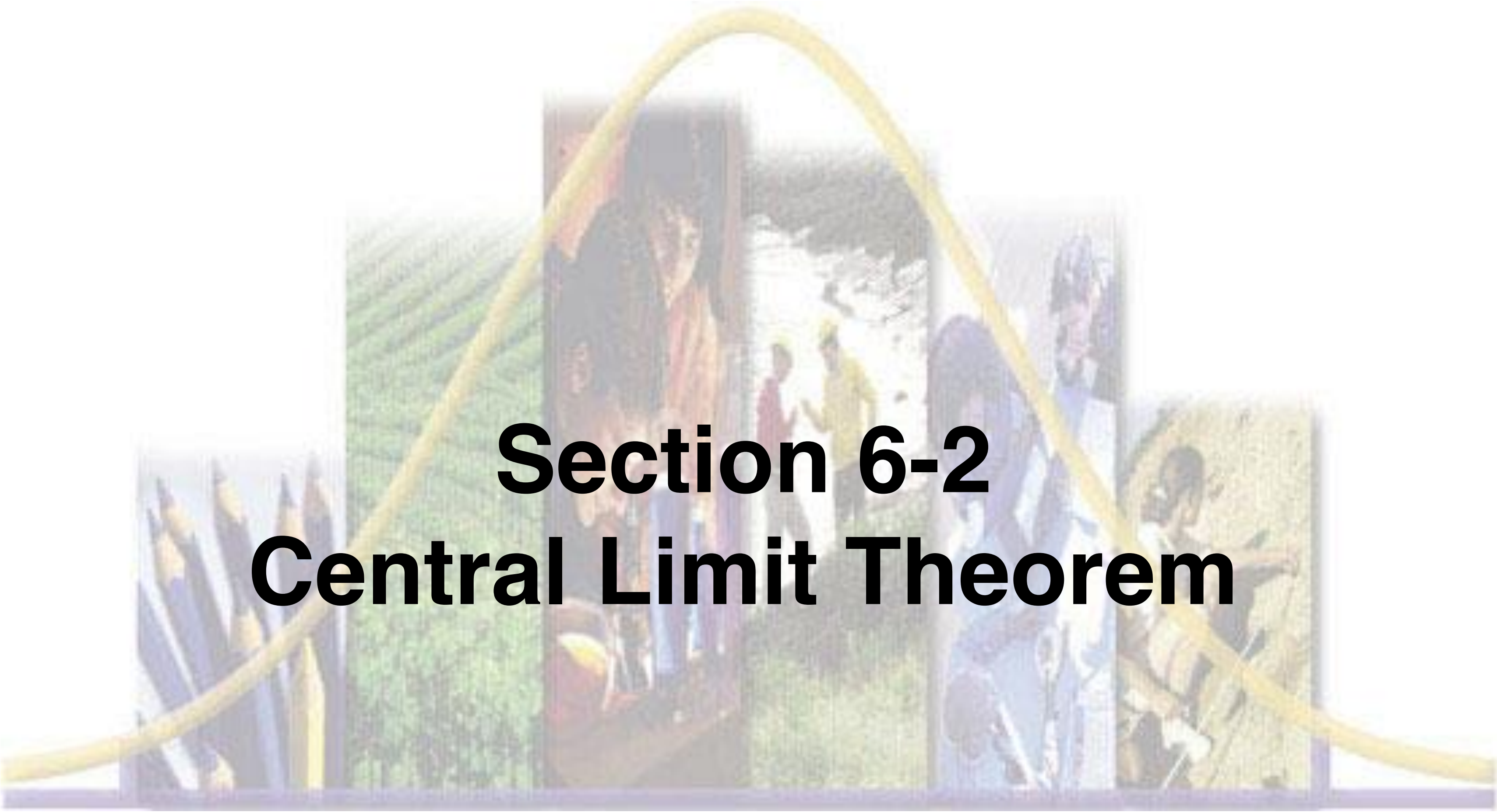
Sample	Mean $\bar{x}$	Median	Range	Variance s^2	Standard Deviation s	Proportion of Odd Numbers	Probability
1, 1	1.0	1.0	0	0.0	0.000	1	1/9
1, 2	1.5	1.5	1	0.5	0.707	0.5	1/9
1, 5	3.0	3.0	4	8.0	2.828	1	1/9
2, 1	1.5	1.5	1	0.5	0.707	0.5	1/9
2, 2	2.0	2.0	0	0.0	0.000	0	1/9
2, 5	3.5	3.5	3	4.5	2.121	0.5	1/9
5, 1	3.0	3.0	4	8.0	2.828	1	1/9
5, 2	3.5	3.5	3	4.5	2.121	0.5	1/9
5, 5	5.0	5.0	0	0.0	0.000	1	1/9
Mean of Statistic Values	2.7	2.7	1.8	2.9	1.3	0.667	
Population Parameter	2.7	2	4	2.9	1.7	0.667	
Does the sample statistic target the population parameter?	Yes	No	No	Yes	No	Yes	

Interpretation of Sampling Distributions

We can see that when using a sample statistic to estimate a population parameter, some statistics are good in the sense that they target the population parameter and are therefore likely to yield good results. Such statistics are called *unbiased estimators*.

Statistics that target population parameters: mean, variance, proportion

Statistics that do not target population parameters: median, range, standard deviation



Section 6-2
Central Limit Theorem

Central Limit Theorem

Given:

1. The random variable x has a distribution (which may or may not be normal) with mean μ and standard deviation σ.

2. Samples all of the same size n are randomly selected from the population of x values.

One sample i

$$\bar{x}_i = \frac{x_1 + x_2 + \cdots x_n}{n}$$

multi samples (distribution)

$$\bar{x}_1, \bar{x}_2, \bar{x}_3 \cdots$$

Central Limit Theorem

Conclusions:

1. The distribution of sample $\bar{x}$ will, as the sample size increases, approach a *normal* distribution.

2. The mean of the sample means will be the population mean μ.

3. The standard deviation of the sample means will approach $\sigma/\sqrt{n}$.

Practical Rules Commonly Used:

1. For samples of size n larger than 30, the distribution of the sample means can be approximated reasonably well by a normal distribution. The approximation gets better as the sample size n becomes larger.

2. If the original population is itself normally distributed, then the sample means will be normally distributed for any sample size n (not just the values of n larger than 30).

Notation

the mean of the sample means

$$\mu_{\bar{x}} = \mu$$

the standard deviation of sample mean

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

(often called <u>standard error of the mean</u>)

Distribution of 200 digits from Personal ID Numbers

(Last 4 digits from 50 students)

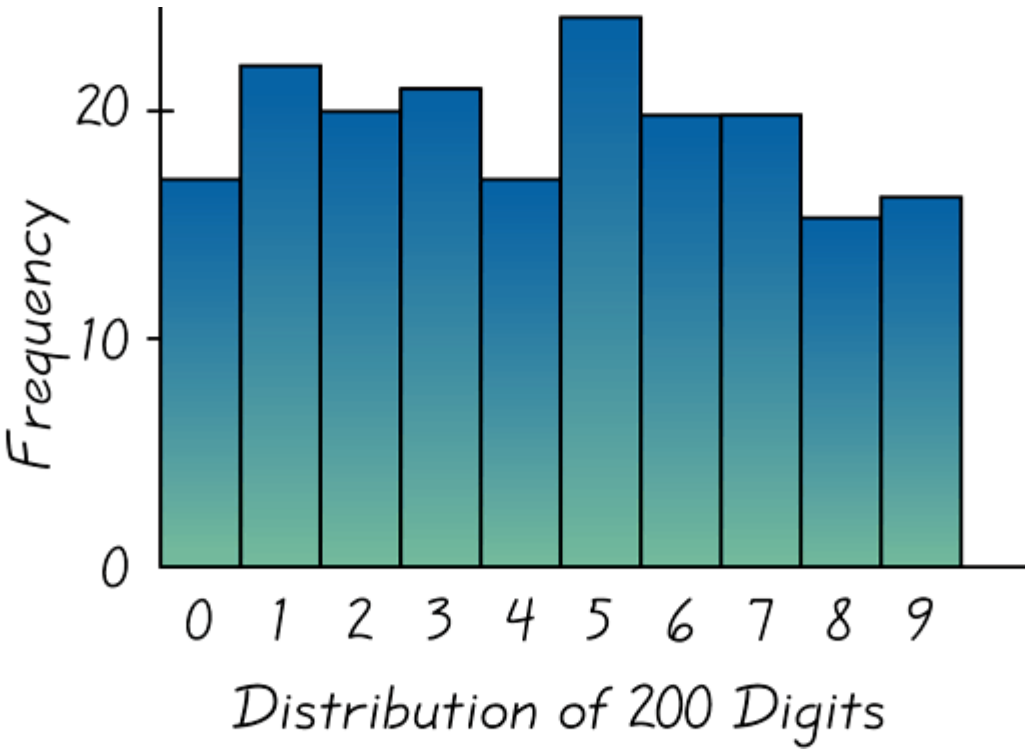
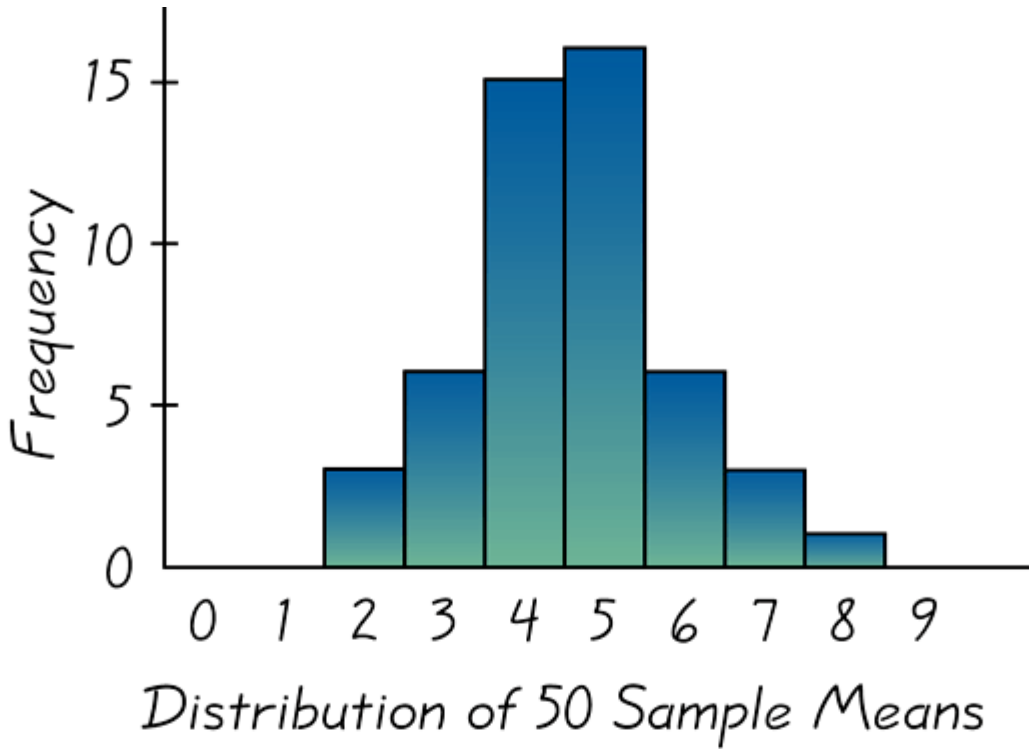


Table 5-6

Last 4 digits				$\bar{x}$
1	8	6	4	4.75
5	3	3	6	4.25
9	8	8	8	8.25
5	1	2	5	3.25
9	3	3	5	5.00
4	2	6	2	3.50
7	7	1	6	5.25
9	1	5	4	4.75
5	3	3	9	5.00
7	8	4	1	5.00
0	5	6	1	3.00
9	8	2	2	5.25
6	1	5	7	4.75
8	1	3	0	3.00
5	9	6	9	7.25
6	2	3	4	3.75
7	4	0	[illegible]	4.50
5	[illegible]	[illegible]	6	[illegible]
[illegible]	2	8	6	[illegible]
2	0	9	7	4.50
5	8	9	0	5.50
6	5	4	9	6.00
4	8	7	6	6.25
7	1	2	0	2.50
2	9	5	0	4.00
8	3	2	2	3.75
2	7	1	6	4.00
6	7	7	1	5.25
2	3	3	9	4.25
2	4	7	5	4.50
5	4	3	7	4.75
0	4	3	8	3.75
2	5	8	6	5.25
7	1	3	4	3.75
8	3	7	0	4.50
5	6	6	7	6.00

Distribution of 50 Sample Means for 50 Students



As the sample size increases, the sampling distribution of sample means approaches a normal distribution.

Example: **Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,**

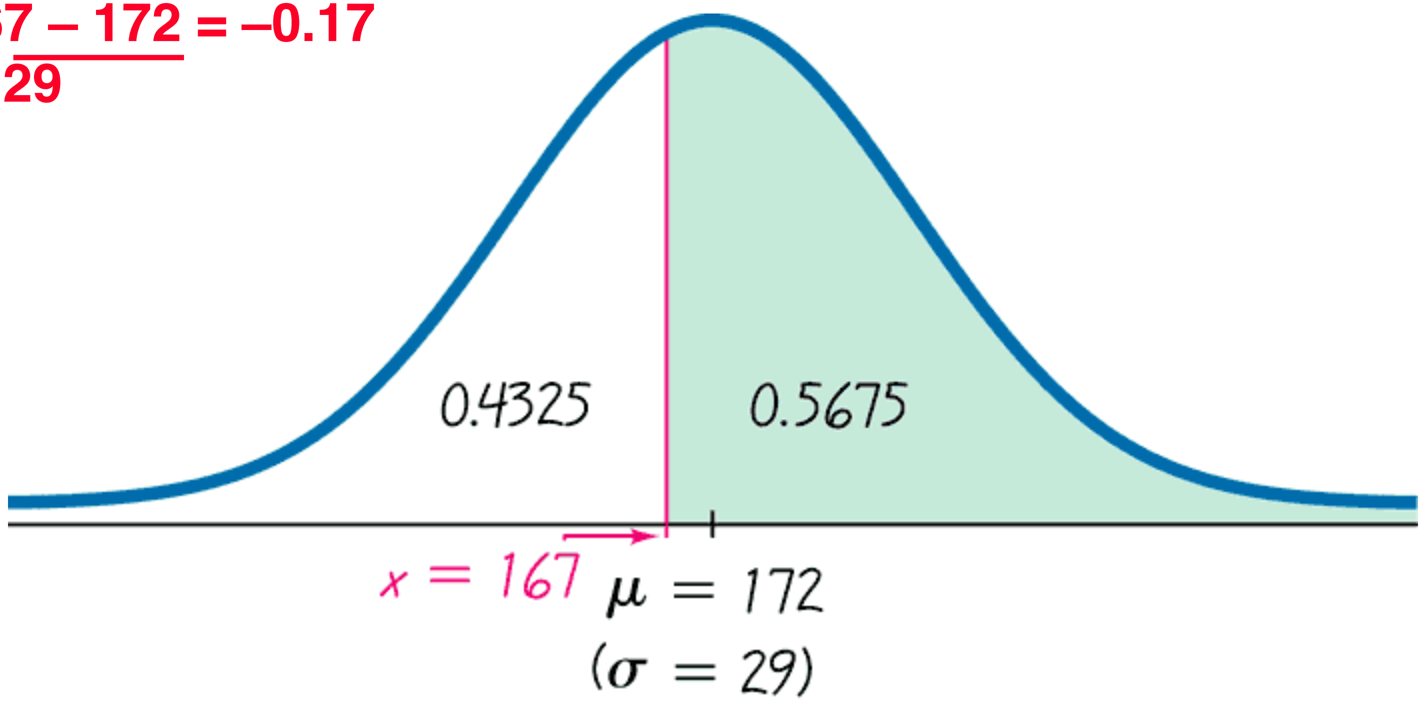
a) **if one man is randomly selected, find the probability that his weight is greater than 167 lb.**

b) **if 12 different men are randomly selected, find the probability that their mean weight is greater than 167 lb.**

Example: **Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,**

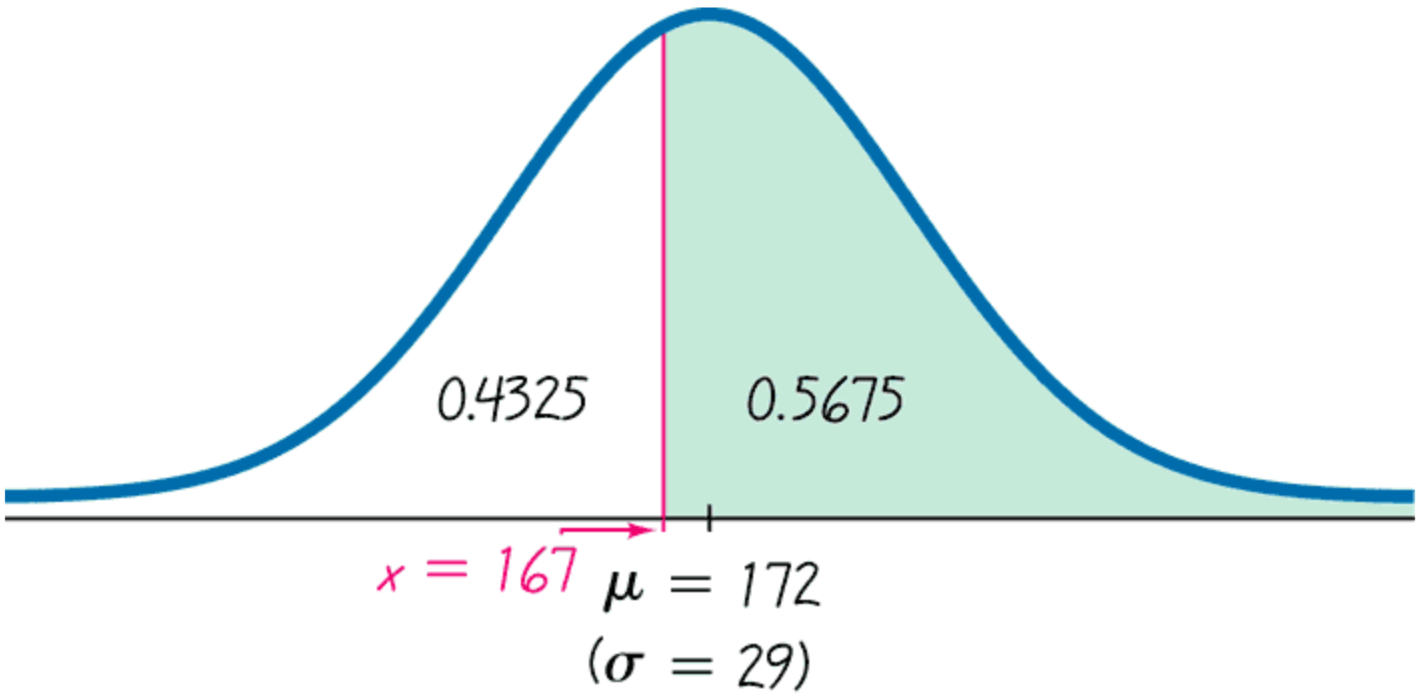
a) if one man is randomly selected, find the probability that his weight is greater than 167 lb.

$$z = \frac{167 - 172}{29} = -0.17$$



Example: **Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,**

a) if one man is randomly selected, the probability that his weight is greater than 167 lb. is 0.5675.

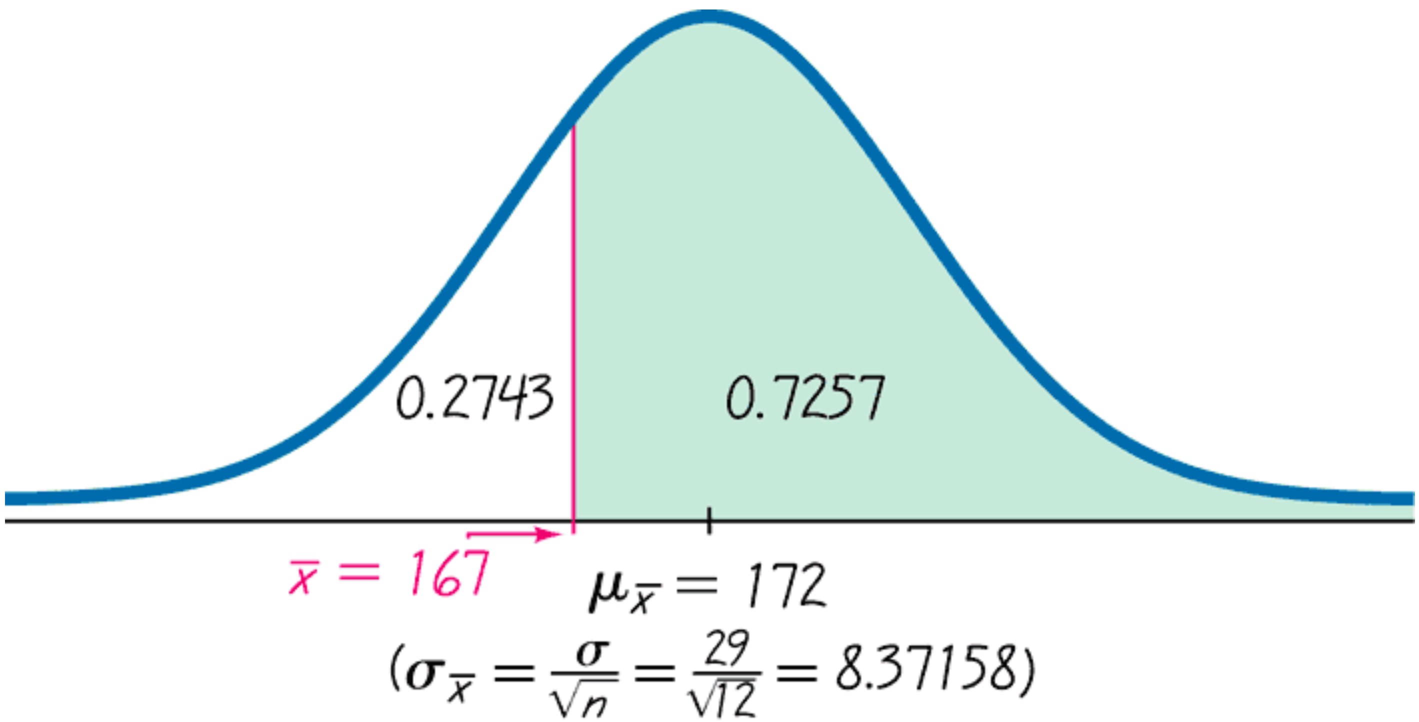


Example: **Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,**

b) if 12 different men are randomly selected, find the probability that their mean weight is greater than 167 lb.

Example: **Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,**

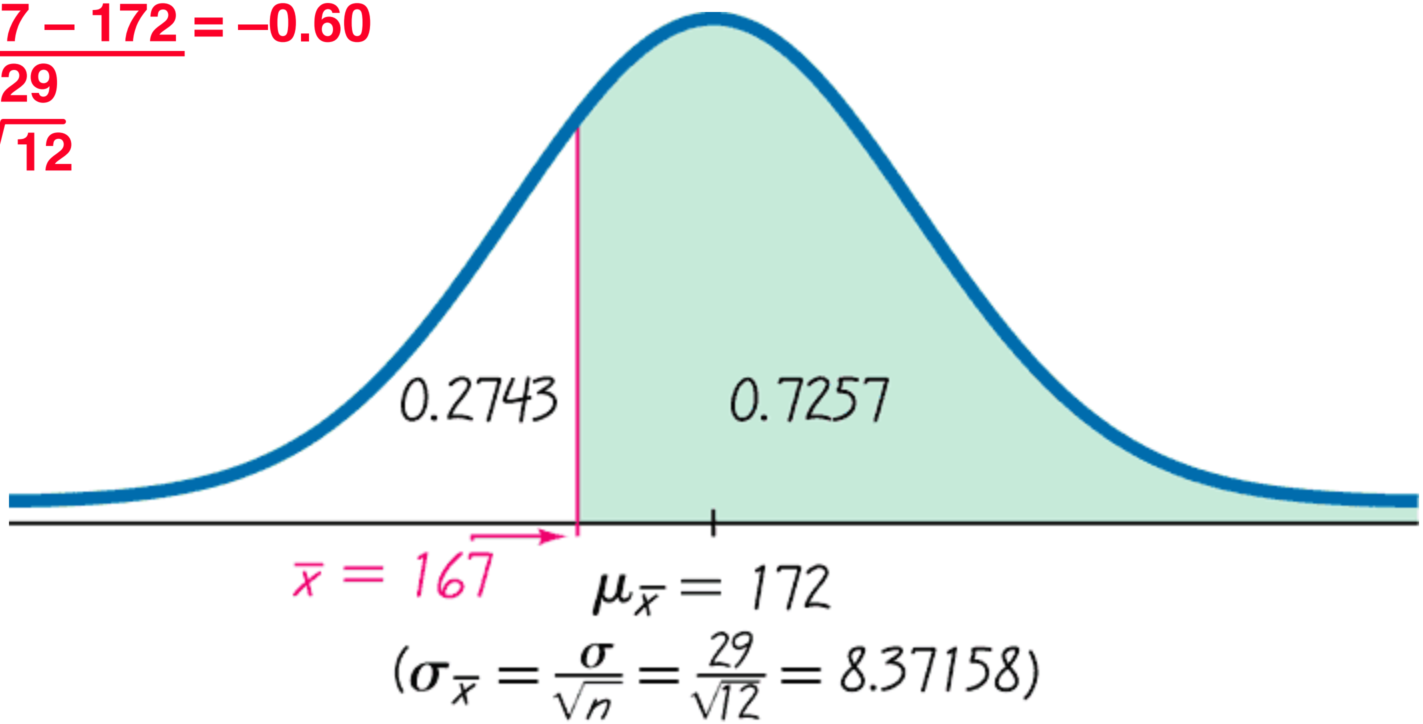
b) if 12 different men are randomly selected, find the probability that their mean weight is greater than 167 lb.



Example: **Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,**

b) if 12 different men are randomly selected, find the probability that their mean weight is greater than 167 lb.

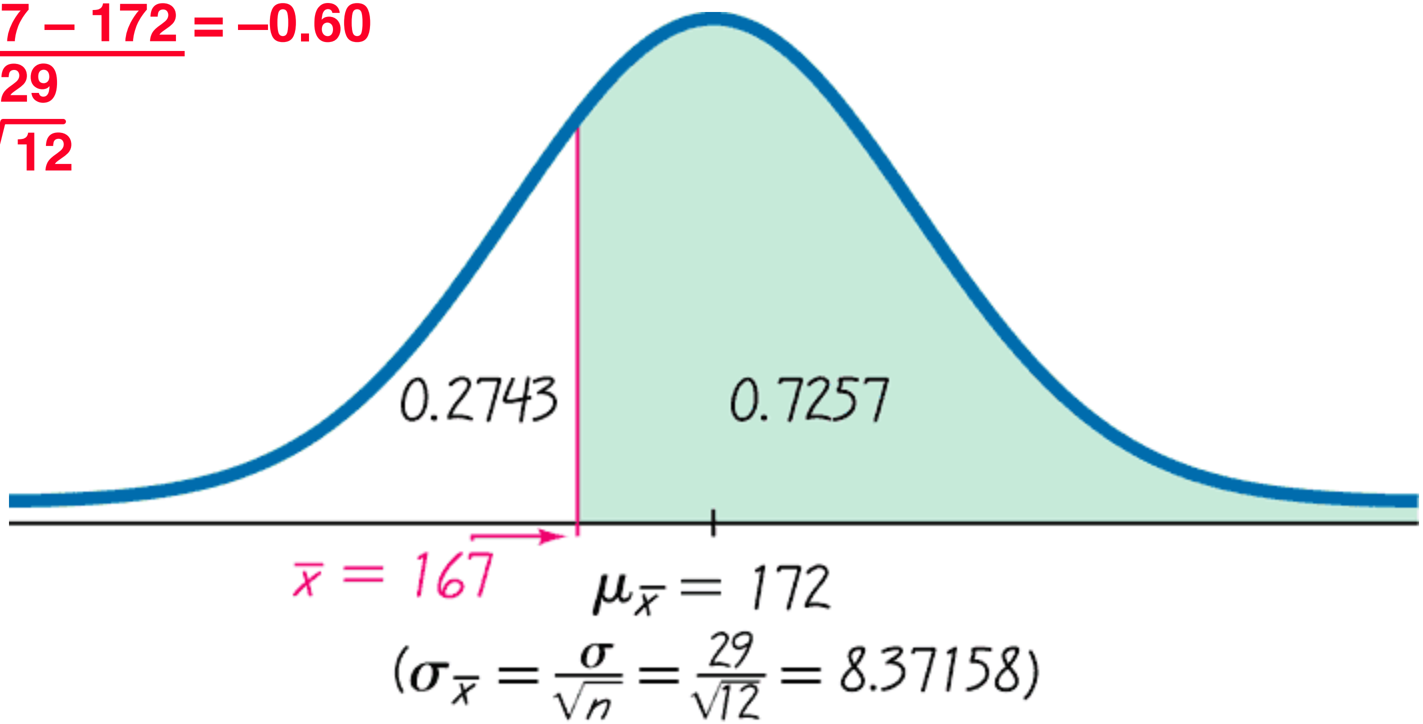
$$z = \frac{167 - 172}{\frac{29}{\sqrt{12}}} = -0.60$$



Example: **Given the population of men has normally distributed weights with a mean of 143 lb and a standard deviation of 29 lb,**

b.) if 12 different men are randomly selected, the probability that their mean weight is greater than 167 lb is 0.7257.

$$z = \frac{167 - 172}{\frac{29}{\sqrt{12}}} = -0.60$$



Example: **Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,**

a) if one man is randomly selected, find the probability that his weight is greater than 167 lb.

$$P(x > 167) = 0.5675$$

b) if 12 different men are randomly selected, their mean weight is greater than 167 lb.

$$P(\bar{x} > 167) = 0.7257$$

It is much easier for an individual to deviate from the mean than it is for a <u>group of 12</u> to deviate from the mean.

Sampling Without Replacement

If $n > 0.05$ N

$$\sigma_{\bar{x}} = \underbrace{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

finite population correction factor

Approximate a Binomial Distribution with a Normal Distribution if:

$$np \geq 5$$

$$nq \geq 5$$

then $\mu = np$ and $\sigma = \sqrt{npq}$

and the random variable has

a  distribution.

(normal)

Other important distributions

(small samples)

指数的，幂函数的

Exponential Distribution

- Another important continuous distribution is the ***exponential distribution*** which has this probability density function:

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0 \qquad\qquad \mu = \sigma = \frac{1}{\lambda}$$

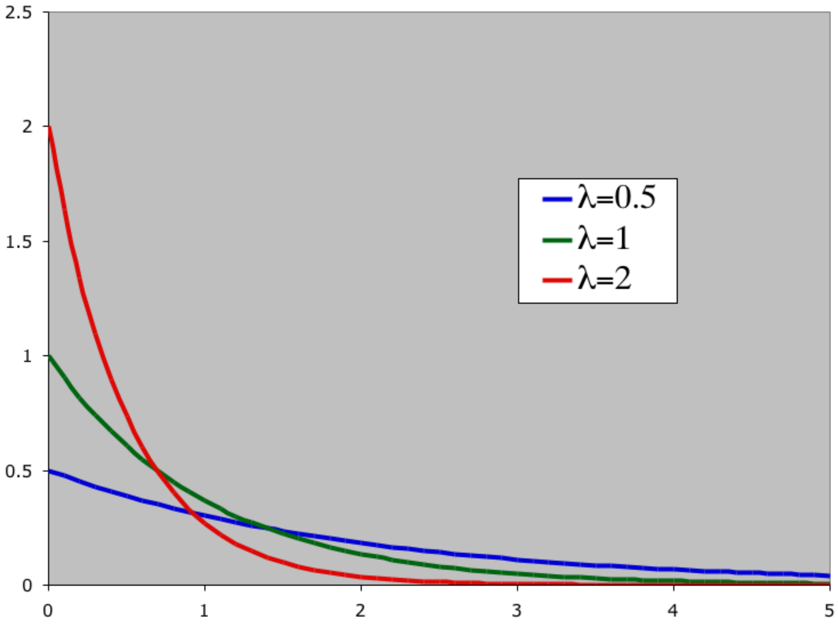
- Note that $x \geq 0$. Time (for example) is a non-negative quantity; the exponential distribution is often used for time related phenomena such as the length of time between phone calls or between parts arriving at an assembly station. Note also that the mean and standard deviation are equal to each other and to the inverse of the parameter of the distribution (lambda λ)

Exponential Distribution

- The exponential distribution depends upon the value of λ
- Smaller values of λ "flatten" the curve:

(E.g. exponential distributions for

λ = **.5**, **1**, **2**)



Student *t* Distribution

- Here the letter ***t*** is used to represent the random variable, hence the name. The density function for the Student ***t*** distribution is as follows…

$$f(t) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\left[1+\frac{t^2}{\nu}\right]^{-(\nu+1)/2}$$

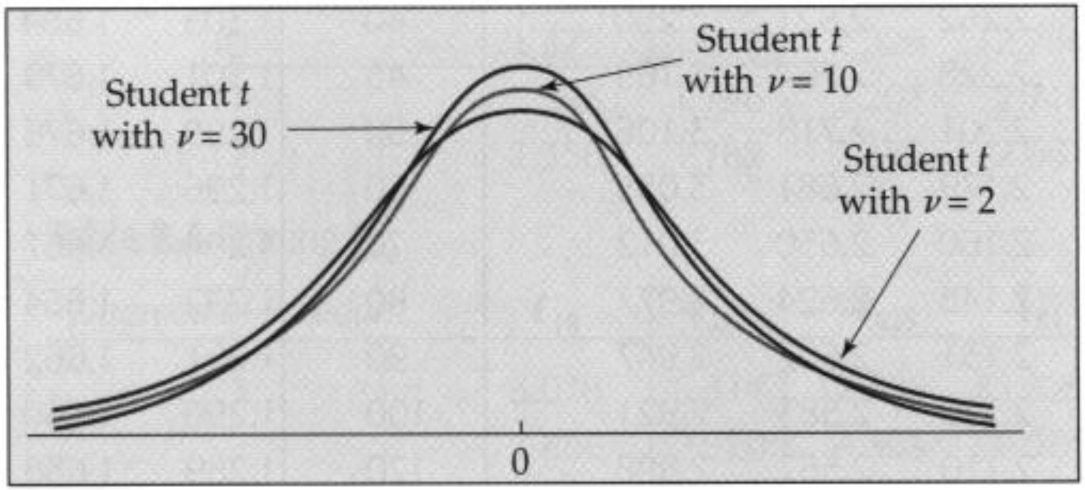
- ν (nu) is called ***the degrees of freedom (df)***, and
- Γ (Gamma function) is (k)=(k-1)(k-2)…(2)(1)

When population σ^2 is unknown and n is not so large, $t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$ follows df=n-1 t-distribution, where s is sample variance.

(small samples)

Student t Distribution

- In much the same way that μ and σ define the normal distribution [2 parameters], ν, the degrees of freedom, defines the Student [df]
- t Distribution:



- As the number of degrees of freedom increases, the t distribution approaches the standard normal distribution.

χ^2 Chi-Squared Distribution

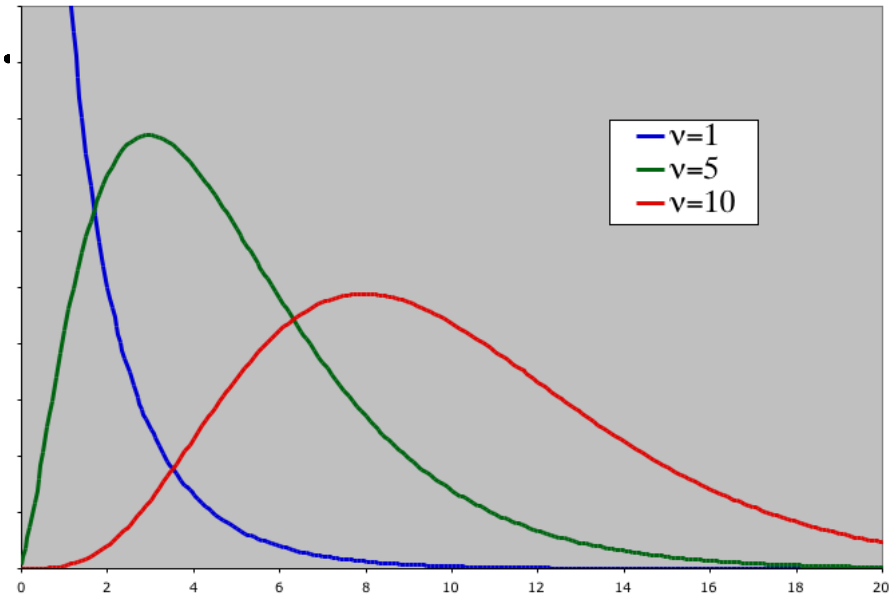
- The chi-squared density function is given by:

$$f(\chi^2) = \frac{1}{\Gamma(\nu/2)} \frac{1}{2^{\nu/2}} (\chi^2)^{(\nu/2)-1} e^{-\chi^2/2} \qquad ,\chi^2 > 0$$

- As before, the parameter ν is the number of degrees of freedom (df).

$$E(\chi^2) = \nu$$

$$V(\chi^2) = 2\nu$$



- Take n samples from normal distribution, then

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$$

or

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

Mean is unknown

follows Chi-square distribution with df = n-1.
Here s^2 is the sample variance with n samples.

F Distribution

- The ***F*** density function is given by:

$$f(F) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}\frac{F^{\frac{\nu_1-2}{2}}}{\left(1+\frac{\nu_1 F}{\nu_2}\right)^{\frac{\nu_1+\nu_2}{2}}}$$

- F > 0. Two parameters define this distribution,

- ν_1 is the “numerator” degrees of freedom (df_1) and
- ν_2 is the “denominator” degrees of freedom (df_2).

- Take n_1 and n_2 samples from two normal distributions, then the ratio of their normalized sample variances

$$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$$

follows F distribution with $df_1 = n_1-1$ and $df_2=n_2-1$.