The ______________ of the Sample Mean($V(\overline{X})$)

➢We also need to know the variance of the sampling distribution of ___for a given sample size n.

Notation: The variance of the values of $\overline{X}$ is denoted by either:

$$\boxed{V(\overline{X}) \quad or \quad \sigma^2_{\overline{X}}}$$

➢The _________ is the average of the squared deviations of the variable $\overline{X}$ about its mean $\mu_{\overline{X}}$.

$$\boxed{\sigma^2_{\overline{X}} = V(\overline{X}) = E(\overline{X} - \mu_{\overline{X}})^2 = \sum_{\overline{X}} (\overline{X} - \mu_{\overline{X}})^2 P(\overline{X})}$$

Continuation of the previous example:

$\overline{X}$	$P(\overline{X})$	$(\overline{X} - \mu_{\overline{X}})$	$(\overline{X} - \mu_{\overline{X}})^2$	$(\overline{X} - \mu_{\overline{X}})^2 P(\overline{X})$
4	$1/16$	**(4-7)=-3**		$9/16$
5	$2/16 = 1/8$	**(5-7)=-2**		$8/16$
6	$3/16$	**(6-7)=-1**		$3/16$
7	$4/16 = 1/4$	**(7-7)=0**		0
8	$3/16$	**(8-7)=1**		$3/16$
9	$2/16 = 1/8$	**(9-7)=2**		$8/16$
10	$1/16$	**(10-7)=3**		$9/16$

$$\sum\left(X_i - \overline{X}\right)^2 P(\overline{X}) = {}^{40}\!/_{16} = 2.5$$

Mean = ___
Variance = ____
Recall the population variance = ____

Notice that: $$\boxed{V(\overline{X}) = \sigma_{\overline{X}}^2 = {}^{\sigma^2}\!/_{n}}$$

i.e. 5/2= 2.5

➢***<u>This is not a coincidence either!!</u>***

<u>Recall:</u> The Var(X) = Var($X_i \ \forall$ i) $=\sigma^2$.
'n' is the sample size.
Sampling distribution is for all possible samples of size n.

Proof:

$$V(\overline{X}) = V\left(\frac{1}{n}\sum_i X_i\right)$$

$$= \left(\frac{1}{n}\right)^2 V\left(\sum X_i\right)$$

$$= \left(\frac{1}{n}\right)^2 \sum\left(V(X_i)\right)$$

and Since X's are independent under random sampling:

$$= \frac{1}{n^2}\left[\sum\left(\sigma^2 + \sigma^2 + \sigma^2 + \cdots + \sigma^2\right)\right]$$

$$= \frac{1}{n^2} n\sigma^2$$

$$= \frac{1}{n}\sigma^2.$$

Although we calculated the value of _____directly in this 4 element population of X_i's, in problems where there are many values of $\overline{X}$, direct calculation is impractical.

As long as we ______the variance of the population σ^2, we can calculate the $V(\overline{X})$.

This is because the variance of the random variable $\overline{X}$ is related to σ^2, the population variance, and to the sample size by the formula:

$$\boxed{V(\overline{X}) = \sigma^2/\_\_}$$

The variance of $\overline{X}$ is **always ____ than or equal** to the population variance.

The variance of the mean of a sample of n independent observations is $1/n$ times the ________ of the parent population (see footnote p. 256).

$$\boxed{V(\overline{X}) = \sigma^2_{\overline{X}} = \left(1/n\right)\sigma^2} \qquad \text{(Equation 7.6)}$$

When n=1, the samples contain only one observation and distribution of X and $\overline{X}$ are the _____.

- As n increases, $\sigma^2_{\overline{X}}$ becomes _______ because the sample means will tend to be closer to the value of the population mean $\mu_{\overline{X}}$.

When n = N (in a finite population) **all** sample means will _____ the population mean and the $V(\overline{X})$'s will equal ___.

With our example, the population variance (σ^2) is known (= 5) and n=2:

So the variance of $\overline{X}$ ($V(\overline{X})$) is:

$$\boxed{V(\overline{X}) = \sigma_{\overline{X}}^2 = \left(\tfrac{1}{n}\right)\sigma^2 = \left(\tfrac{1}{2}\right)(5) = 2.5}$$

<u>What Happens to</u> $V(\overline{X})$ **<u>as n _________?</u>**

➢Because each sample contains more information or more elements of the population as the sample size _________, the sample will be closer to the population, so expect ____ variability.

<u>Example</u>: Suppose X ~N(0, 100)

Randomly draw samples of size:
(i) 10
(ii) 100
(iii) 1000

from this population.

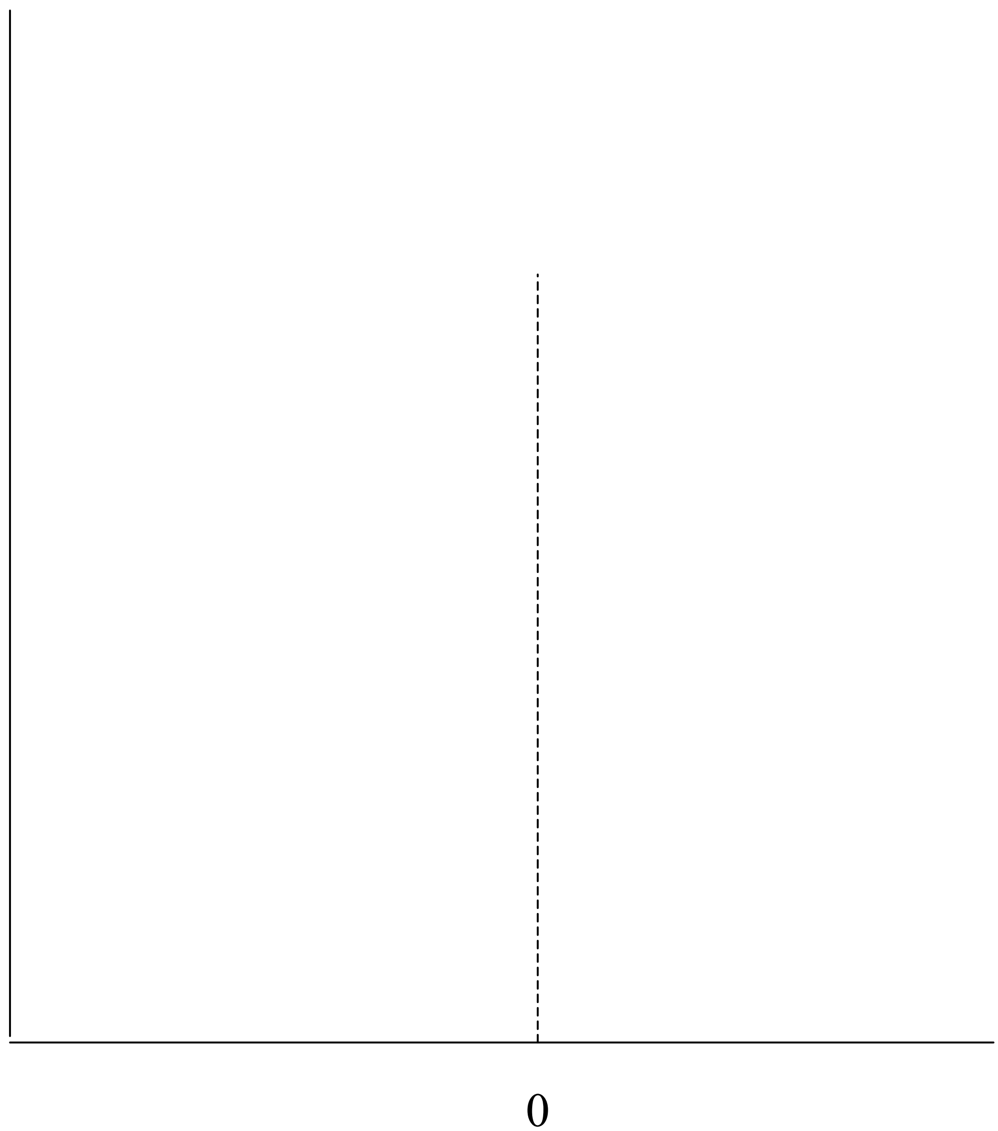
Calculate $\overline{X}_{10}$ for all possible samples of size 10.

Calculate $\overline{X}_{100}$ for all possible samples of size 100.

Calculate $\overline{X}_{1000}$ for all possible samples of size 1000.

Then we can show:

<u>Sampling Distributions for Xbar: Various Sample Sizes:</u>



For n=10: dispersion if Xbar is quite wide around the mean of 0.

For n=1000: less variation around the mean of zero.

- When n approaches ________, there is no dispersion and variance of Xbar =0.

Standard ________________ of the Mean

Notation: We usually denote the standard deviation of $\overline{X}$'s, $\sigma_{\overline{X}}$, the standard _____ of the mean.

- The error refers to ***sampling*** **_____**.

- $\sigma_{\overline{X}}$ is a measure of the standard expected _____ when the sample mean is used to obtain information or draw conclusions about the unknown population mean.

Standard Error of the mean:

$$\boxed{\sigma_{\overline{X}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}}$$ (Equation 7.7)

In our example: $$\boxed{\sigma_{\overline{X}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{5}{2}} = \frac{2.236}{\sqrt{2}} = 1.5811}$$

Notes: (i) $\mu_{\overline{X}}$ and $\sigma_{\overline{X}}$ are parameters of the population of sample averages for all conceivable samples of size n.

- These parameters are usually _______.

(ii) The population parameters (μ, σ^2) are also usually _______.

(iii) This means that we cannot use the relationships:

$$\boxed{\mu = \mu_{\overline{X}} \quad and \quad \sigma/\sqrt{n} = \sigma_{\overline{X}}}$$

to solve for values of one of these statistics.

But these relationships allow us to test hypotheses about the population parameters on the basis of sample results.

More on this later........

<u>Next:</u> We now have derived the mean and the variance of the sampling distribution, but have not said anything about the _____ of the sampling distribution of $\overline{X}$.

Recall that distributions with the same mean and variance can have very different ______.

➢**We must now specify an assumption about the entire distribution of $\overline{X}$'s:**

Section 7.5

Sampling distribution of $\overline{X}$, ______ Parent Population

➢It is typically not possible to specify the ***shape*** of the $\overline{X}$'s when the parent population is discrete and the sample ____ is small.

➢ However, the shape of a sample taken from a **normally distributed** parent population (X) can be specified.

In this case, the $\overline{X}$'s are distributed normally.

" The sampling distribution of $\overline{X}$'s drawn from a normal parent population is a ______ distribution."

Recall: The mean of the $\overline{X}$ s is $\mu_{\overline{X}} = \mu$ and the variance of $\overline{X}$ s is $\sigma^2_{\overline{X}} = \sigma^2/n$.

➢Hence the sampling distribution of $\overline{X}$ is:

$$\overline{X} \sim N(\mu_{\overline{X}}, \sigma^2_{\overline{X}}) = N(\mu, \sigma^2/n)$$

when ever the parent population is ______. X~N(μ,σ^2).

➢Meaning, regardless of the _____ of the parent population, the mean and variance of $\overline{X}$ equal: $\mu_{\overline{X}}$ and $\sigma^2_{\overline{X}} = \sigma^2/n$.

From the **last example**: X~N(0, 100).

Hence,

$$\bar{X}_{10} \sim N\left(0, {}^{100}\!/_{10}\right) = N(0, 10)$$

$$\bar{X}_{100} \sim N\left(0, {}^{100}\!/_{100}\right) = N(0, 1)$$

$$\bar{X}_{1000} \sim N\left(0, {}^{100}\!/_{1000}\right) = N(0, 0.1).$$

Remember: The normal distribution is a continuous distribution.
(I.e. infinite number of different samples could be drawn.)

(Error in text on page 259?)

Example: Suppose all the possible samples of size 10 are drawn from a ______ distribution that has a mean of 25 and a variance of 50. That is, X is normally distributed with a mean μ=25 and variance σ^2=50 : X~N(25,50).

- Since the population mean μ=25, the mean of $\bar{X}$ s equal $\mu_{\bar{X}}$ =25.
- Since the population variance σ^2=50, the variance of the $\bar{X}$'s equals $\sigma^2_{\bar{X}} = {}^{\sigma^2}\!/_{n} = {}^{50}\!/_{10} = 5.$
- Since X is ______, $\bar{X}$ is ________ distributed $\bar{X}$ ~N(25,5).

<u>What this means is</u>:

68.3% of the sample means will fall within ± one _______ error of the mean: $\sigma_{\overline{X}} = \sqrt{5} = 2.24.$

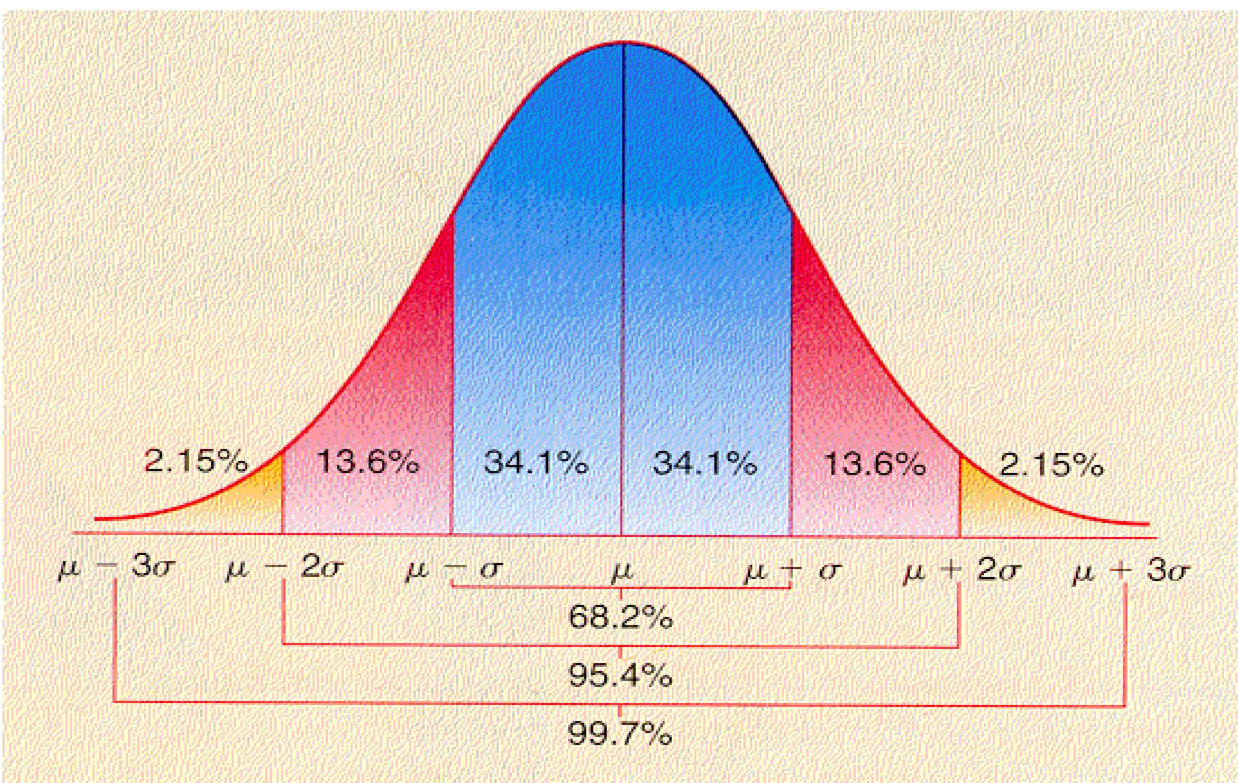
$$\boxed{\mu + 1\sigma_{\overline{X}} = 25 \pm (1)(2.24) = 22.76 \ \mathit{to} \ 27.24.}$$

_____% of the sample means will fall within ± two standard errors of the mean:

$$\boxed{\mu + 2\sigma_{\overline{X}} = 25 \pm (2)(2.24) = 25 \pm 4.48 \Rightarrow 20.52 \ \mathit{to} \ 29.48.}$$

99.7% of the sample means will fall within ± three standard errors of the ____:

$$\boxed{\mu + 3\sigma_{\overline{X}} = 25 \pm (3)(2.24) = 25 \pm 6.72 \Rightarrow 18.28 \ \mathit{to} \ 31.72.}$$



The Standardized Form of the Random Variable $\overline{X}$ and σ

In Economics 245, we saw from Chapter 6 that it is easier to work with the standard normal form of a variable than it is to leave it in its original units.

The same type of _______________ made on a random variable X, can be made on the random variable $\overline{X}$.

Recall, to _________ the random variable X to its standard normal form (Z), we subtract the mean from each value and divide by the standard deviation:

$$\boxed{Z = \frac{(X - \mu)}{\sigma}} \leftarrow Z \text{ has a mean} = 0 \ \& \ \text{variance} = 1.$$

Z~N(0,1).

The standardization of $\overline{X}$ is ___________ the same way:

$$\boxed{Z = \frac{(\overline{X} - \mu_{\overline{X}})}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}} \quad \text{(Equation 7.9)}$$

➢The random variable **Z** has a mean of zero and a variance of ___.

Thus: When sampling from a normal parent population, the distribution of $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ will be ______ with mean zero and variance equal to ___.
(See figure 7.4.)

Example: Suppose X is the height (in inches) of basketball players on all university teams in Canada during summer term. Suppose X~N(__,36). A random sample of nine players is drawn from this population.

- **What is the probability that the sample average team player height is less than 80 inches? (What is P (** $\overline{X} \leq 80$)?)

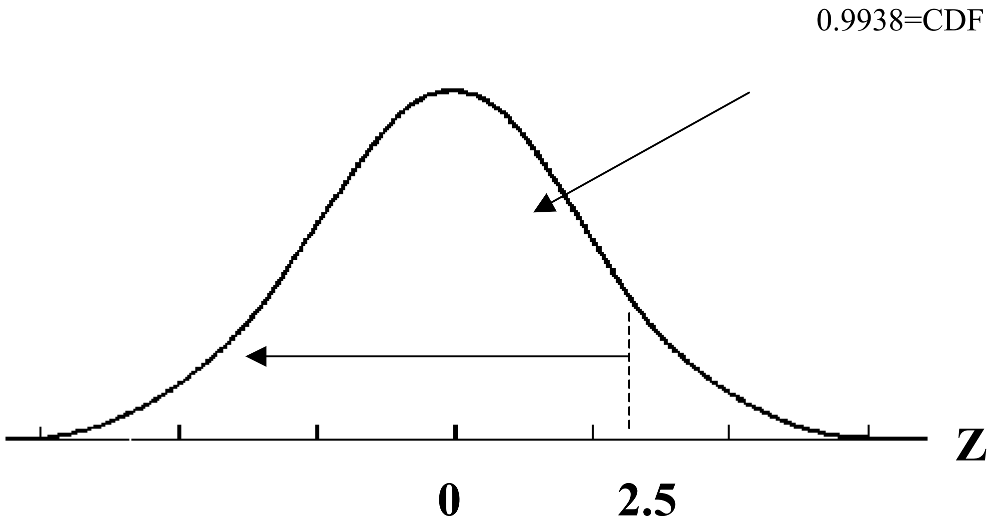
Solution: If X ~N(__,36), then $\overline{X}$ ~N(75,36/9=4).

- Standardize the variable $\overline{X}$:

$$\boxed{Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{80 - \_\_}{6/\sqrt{9}} = \frac{5}{6/3} = \frac{5}{2} = 2.5}$$

Looking at the Cumulative Standardized Normal Distribution Table F(Z), on page 891, the **P(Z ≤2.5) = 0.____**.

The probability that the average height of basketball players in our sample of size 9 is less than 80 inches is 99.38%.



<u>Example:</u> Let X be the amount of money customers owe on home mortgages at the Bank of Nova Scotia (in thousands of $).
Suppose X~N(150,____).
Draw a random sample of 25 from the population.

What is the probability that the average amount owing is greater that $200? $P(\overline{X} \geq 200)$ **?**

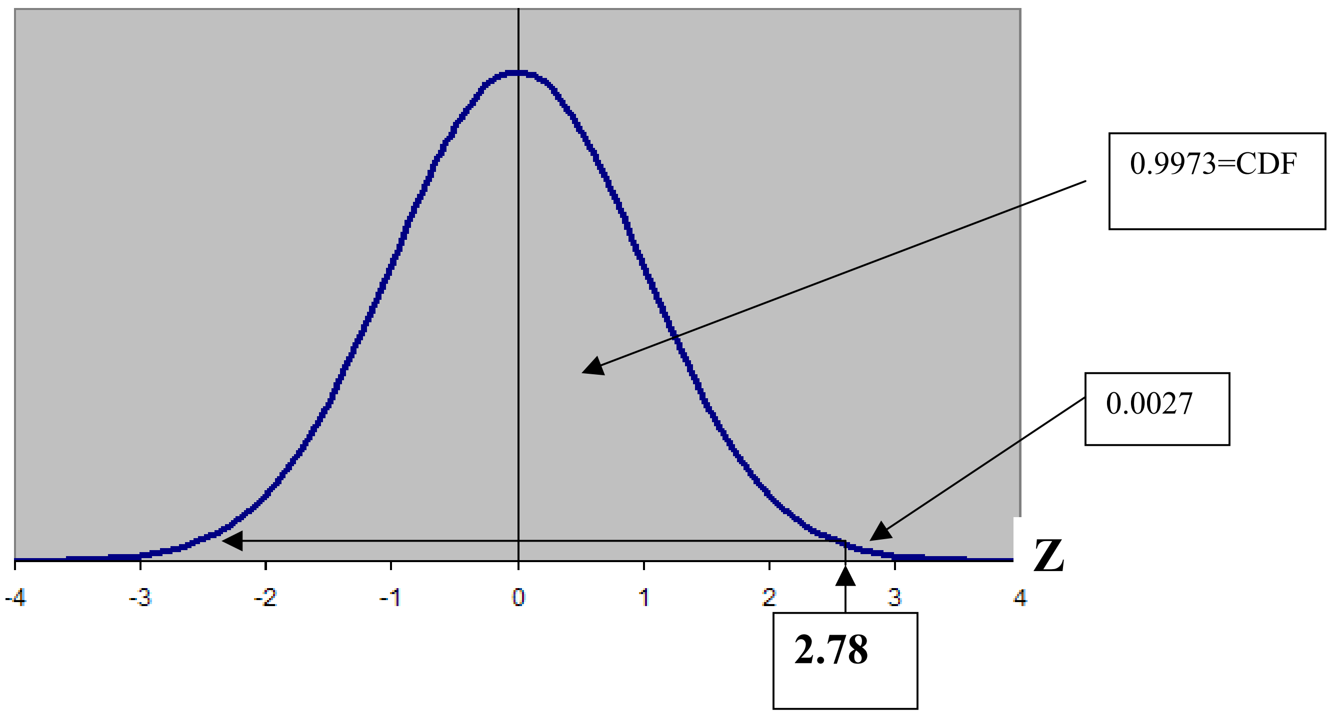
<u>Solution:</u> X~N(150,____), so

$\overline{X}$ ~N(150,____/25=324)=N(150, 324);

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{200 - 150}{90 / \sqrt{25}} = \frac{50}{90/5} = \frac{50}{18} = 2.78$$

P(Z ≥2.78) = (1-0.9973) = 0.____.

The probability that average amount owing on a mortgage is greater that $200 is .27%.



Section 7.6

The limitations from the last section is obvious:

" We cannot always assume that the parent population is ______."

What if the Population is ___-_____l?

Sampling Distribution of $\overline{X}$: Population Distribution Unknown and σ Known

- When the samples drawn are not from a normal population or when the population distribution is unknown, the ____ ***of the sample*** is extremely important.

- When the sample ____ is small, the shape of the distribution will depend mostly on the shape of the parent population.

- As the sample ____ increases, the shape of the sampling distribution of $\overline{X}$ will become more and more like a ______ distribution, regardless of the shape of the parent population.

Central Limit Theorem:

"Regardless of the distribution of the parent population, as long as it has a finite mean μ and variance σ^2, the distribution of the means of the random samples will approach a ______ distribution, with mean μ and variance σ^2/n, as the sample size n, goes to infinity."

(I) When the parent population is ______, the sampling distribution of $\overline{X}$ is **exactly** ______.

(II) When the parent population is not normal or unknown, the sampling distribution of $\overline{X}$ is **approximately** ______ as the sample size increases.

Example:
Let the sample be $(X_1, X_2, \ldots, X_n)$
Let $S=(X_1 + X_2 + X_3+\ldots+X_n)$

$E(S) = E(X_1) + E(X_2) + \ldots + E(X_n)$
$= \Sigma E(X_i) = n(E(X) =$ ___

$V(S) = V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n)$
$= \Sigma V(X_i) = nV(X_i) =$ ___.

Assuming independence.

So according to the CLT as $n \to \infty$

$$S \to N(n\mu, n\sigma^2)$$

Now, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{(X_1 + X_2 + \cdots + X_n)}{n} = \frac{S}{n}$.

The expected value of $\overline{X}$ is:

$$E(\overline{X}) = \frac{1}{n}E(S) = \frac{1}{n}n\mu = \mu$$

$$and\ the\ \text{var}\,iance\ of\ \ \overline{X}:$$

$$V(\overline{X}) = V\left(\frac{S}{n}\right) = \frac{1}{n^2}V(S) = \frac{1}{n^2}\sigma^2 n = \frac{\sigma^2}{n}.$$

So, according to the CLT: as n $\rightarrow \infty$,

$\overline{X}$ ~N(μ, σ^2/n) regardless of the form of the parent population distribution.

Notes on Page 262, Figure 7.5

(Distribution with discrete values of X.)
(Note: the CLT applies in discrete and continuous cases.)

➢The first row of diagrams in Figure 7.5 shows four different parent populations.

The next 3 rows show the sampling distribution of $\overline{X}$ for all possible repeated samples of size n=2, n=5, and n=30, drawn from the populations in the first row.

Column 1: Normal population

All sampling distributions are normal and have the same mean μ;

The variances decrease as $\sqrt{n}$ increases.

Column 2 : Uniform Population

At n=2, symmetrical

At n=5, normal looking distribution

Column 3: Bimodal Population

At n=2 the distribution is symmetrical.

At n=5, the distribution is bell-shaped.

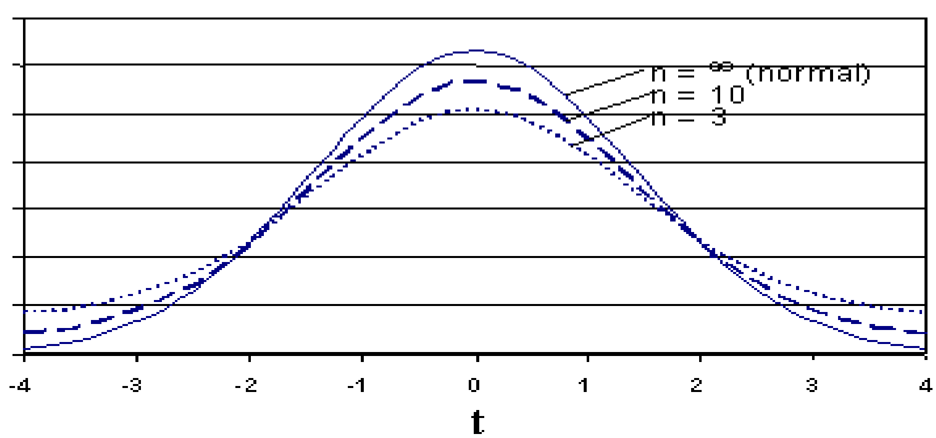
Column 4: Highly skewed exponential Population.

At n=2 and n=5, the distribution is still skewed.

At n=30, symmetrical bell-shaped distribution for

$$\overline{X} \rightarrow \_\_\_\_\_\_.$$

In General, if $n \geq$ __, the sampling distribution of $\overline{X}$ will be a good approximation.



Section 7.8
Sampling Distribution of $\overline{X}$, Normal Population, σ _______

Recall that if $X \sim N(\mu,\sigma^2)$, then $\overline{X} \sim N(\mu,\sigma^2/n)$;

Also recall that the standardized form of Z,

$$Z = \frac{(\overline{X} - \mu)}{\sigma / \sqrt{n}}$$

is important in the determination of probability of $\overline{X}$ taking some value, assuming that its population mean is μ.

We then use this probability distribution for problem solving and decision making.

But what happens if σ is _______?

➢In solving a problem where σ is _______, 's', the sample statistic for standard deviation of σ, can be applied to solve problems involving standardization.

It is legitimate because it can be shown that: $E(S^2)=\sigma^2$

and we can standardize creating a new ratio:

$$\boxed{t = \frac{\overline{X} - \mu}{s / \sqrt{n}}}.$$

Where "t-ratio" is **not** _________ distributed.

➢The resulting distribution no longer has a _______ equal to 1.

To determine the distribution of the ratio $\boxed{\dfrac{\bar{X} - \mu}{s / \sqrt{n}}}$ we follow these steps:

1) Collect all the possible samples of size n from a normal parent population.

2) Calculate $\bar{X}$ and s for each sample.

3) Subtract μ from each value of $\bar{X}$, and then divide this deviation by the appropriate value of $s / \sqrt{n}$.
This process will generate an infinite number of values of this random variable $\boxed{\dfrac{\bar{X} - \mu}{\_\_ / \_\_\ \_\_}}$.

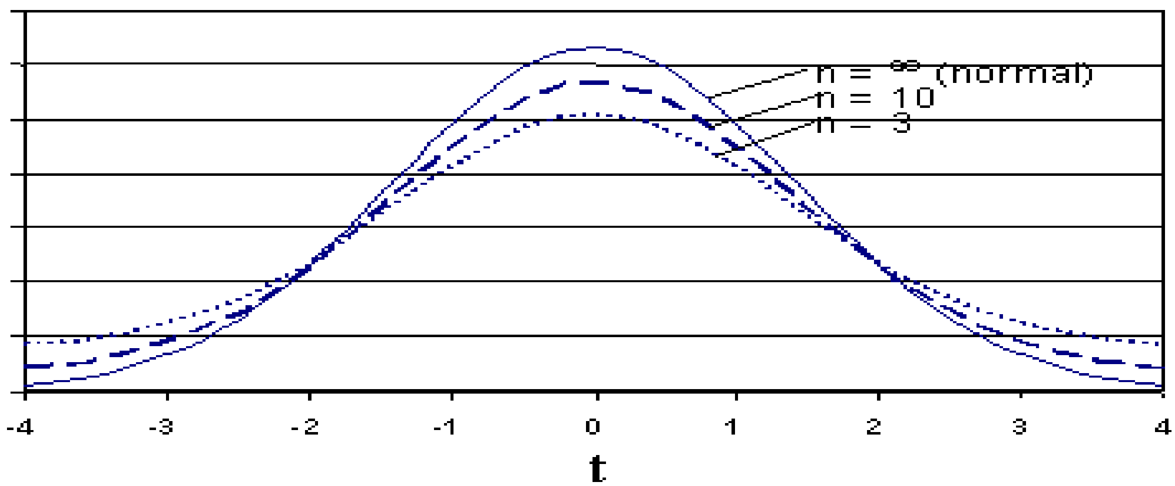
➢The **mean** of the t-distribution still equals __.

➢The **variance** no longer equals V(Z) = _. It is ______.

Because we use ‘s’ to standardize, the dispersion or the variation around the mean zero, will be wider.

➢ "s" introduces an element of uncertainty or ____because s^2 is a parameter estimate, not the **actual** population parameter.

Hence the more uncertainty there is, the more spread out the distribution.



Notes About The t- Distribution:

1) The t-distribution was developed by W.S. Gossett.

It consists of two random variables $\overline{X}$ and s. Hence, the variable "t" is a __________ random variable.

2) $[-\infty < t < \infty]$ __________ probability.

3) The t-distribution is ___________: E(t) =0= median = mode.

4) Variability of the t-distribution depends on the sample size (n), since n affects the ***reliability*** of the estimate of **'σ'** which **'s'** estimates.

➢When n is large, ‘s’ will be a **good** estimator of σ.
➢When n is small, ‘s’ may ___ be a good estimator.

The variability of the distribution depends on n:

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} \quad (7.11)$$

1) The t-distribution $\boxed{t = \dfrac{Z}{\sqrt{\chi^2 / \nu}} = \dfrac{(\overline{X} - \mu)\big/ \frac{\sigma}{\sqrt{n}}}{\sqrt{\chi^2 / \nu}}}$.

6) We characterize the t-distribution in terms of the sample size minus one, (n-1).

The (n-1) is referred to as the number of “**degrees of freedom**” (d.f.), which represent the number of ___________ pieces of information that are used to estimate the standard deviation of the parent population.

ν ← “nu” denotes degrees of freedom: ν=(n-1).

➢t-distribution is described by ν degrees of freedom.

(i) The mean of t-distribution =0; [E(t) = 0].

(ii) The variance for n≥3, is $\boxed{V(t) = \dfrac{\nu}{(\nu - 2)} = \dfrac{(n-1)}{(n-1-2)} = \dfrac{(n-1)}{(n-3)}}$.

7) For _____ sample sizes, the t-distribution is typically more ______ out than the normal distribution.

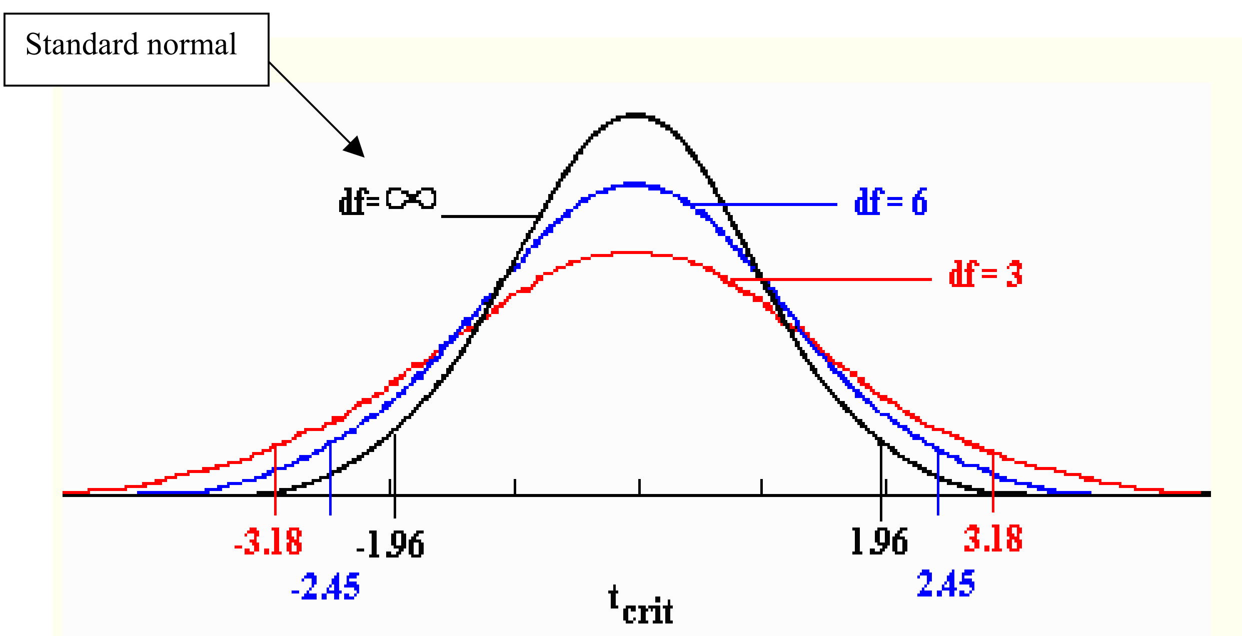
➢t-distribution typically has fatter tails than the Z for small degrees of freedom.

➢When the degrees of freedom are larger than 30, the t distribution resembles the ______ distribution.

➢In the limit, as n approaches infinity, the t and Z distributions are the same.

So, the t-tables usually have probability values for $\nu \leq 30$, since larger samples normally give a good approximation and are easier to use.

Although the distribution holds for any sample size, we usually use the t-distribution when we are using ______samples.



Probability Applications:

Probability questions involving a t-distributed random variable can be solved by forming the t-statistic: $\boxed{t = \frac{\bar{X} - \mu}{s/\sqrt{n}}}$,

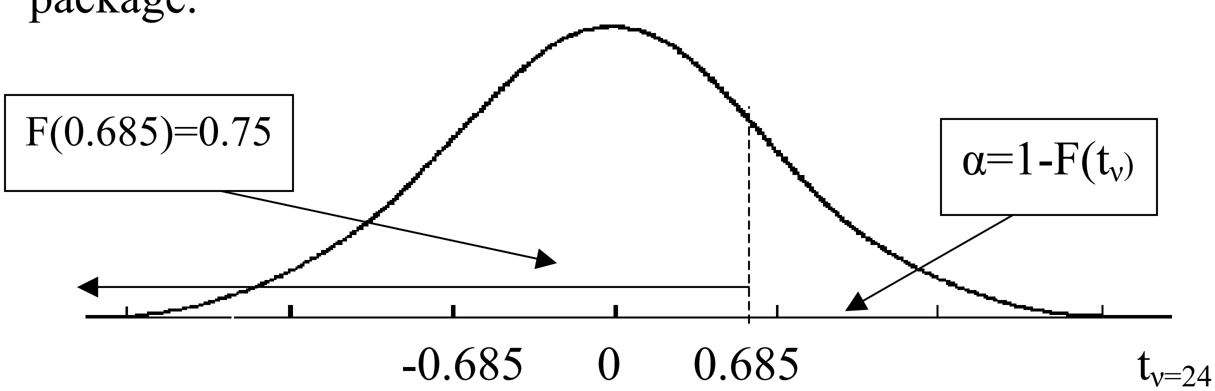
and determining the probability by using the Student t-table (H&M page 894) or a computer generated value (using the @ctdist(x,ν) command in EViews).

➢The Student t-table gives the values of "t" for _______ values of the cumulative probability F(t)=P($\boldsymbol{t}$<t) across the top of the table and for degree of freedom (ν) down the left margin.

➢Table VI gives probabilities for 7 selected t-values for each degree of freedom.

➢More extensive tables are available.

➢The easiest way to determine probabilities is to use a statistical package.



Recall, the t-distribution is the appropriate statistic for inference on a population mean whenever the parent population is normally distributed and σ is _______.

Example: A large restaurant reports its outstanding bills to suppliers are approximately normally distributed with a mean of \$1200. The standard deviation is unknown. A random sample of 10 accounts is taken. The mean of the sample $\overline{X}$ =___, with a standard deviation s=___. What is the probability that the sample mean will be \$980 or lower when μ=1200?

P($\overline{X} \leq 980$)?

To solve, standardize the values:

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} = \frac{980 - 1200}{210/\sqrt{10}} = \frac{-220}{66.4078} = -\_\_\_\_\_\_$$

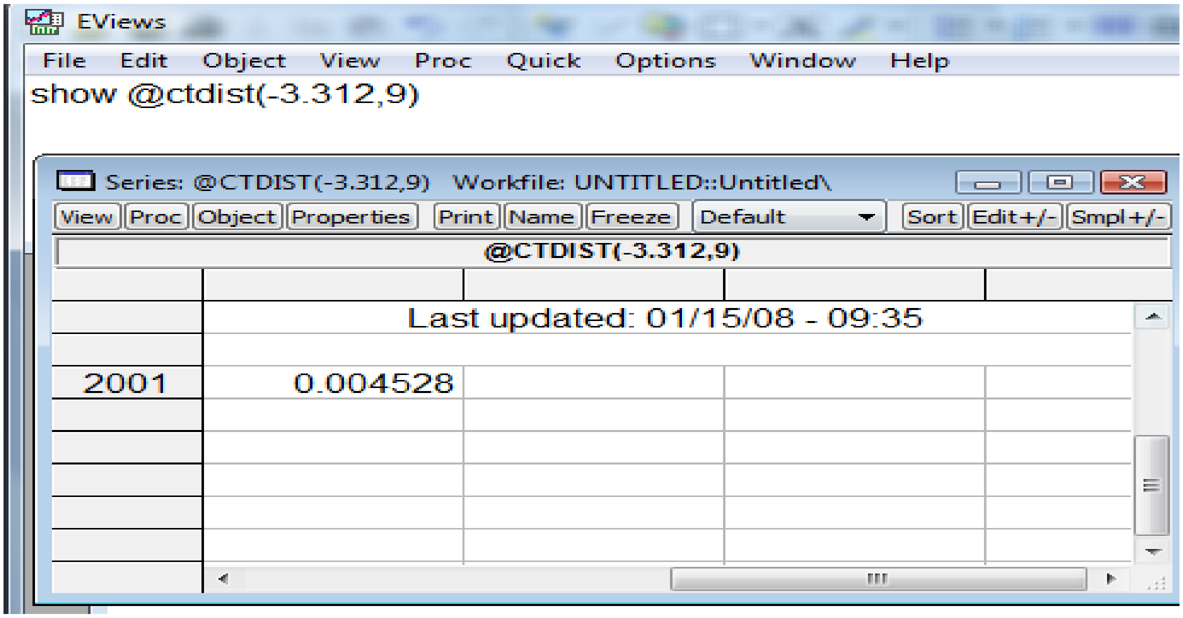
Using t=-3.312 does not appear in the row ν=9.
Use table to determine an upper and lower bound for P(t<-3.312):

F(-3.250) = 0.005
F(-4.781) = 0.0005

0.005 ≤ P(t≤-_.___) ≤ 0.0005.

A Sample mean as low or lower than 980 will occur approximately between 0.5% to 0.05% of the time with μ=\$1200. May be concerned with the accuracy of the sample.

Using EViews :



Draw another sample:

<u>Next sample</u>: What is the $P(\bar{X} \geq 1250)$?

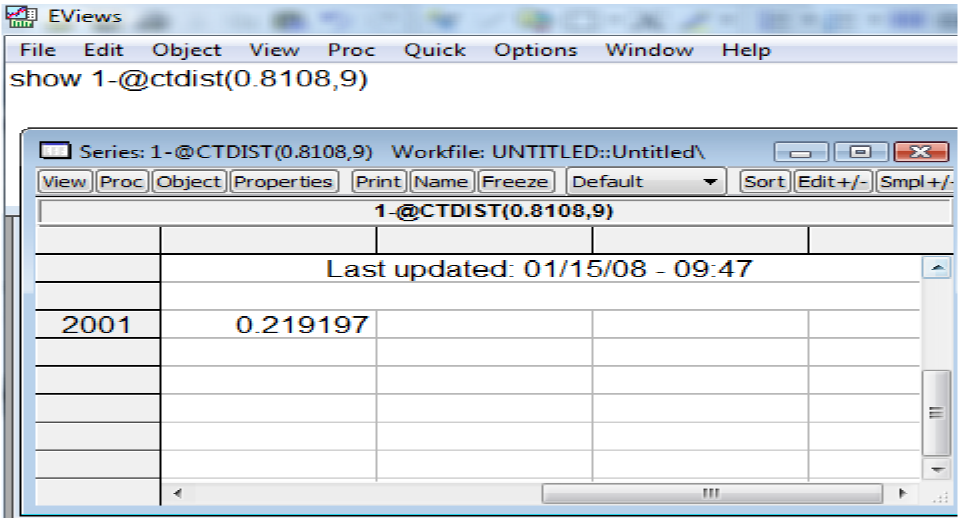
$\bar{X}$ =1250, s=195 and n=10

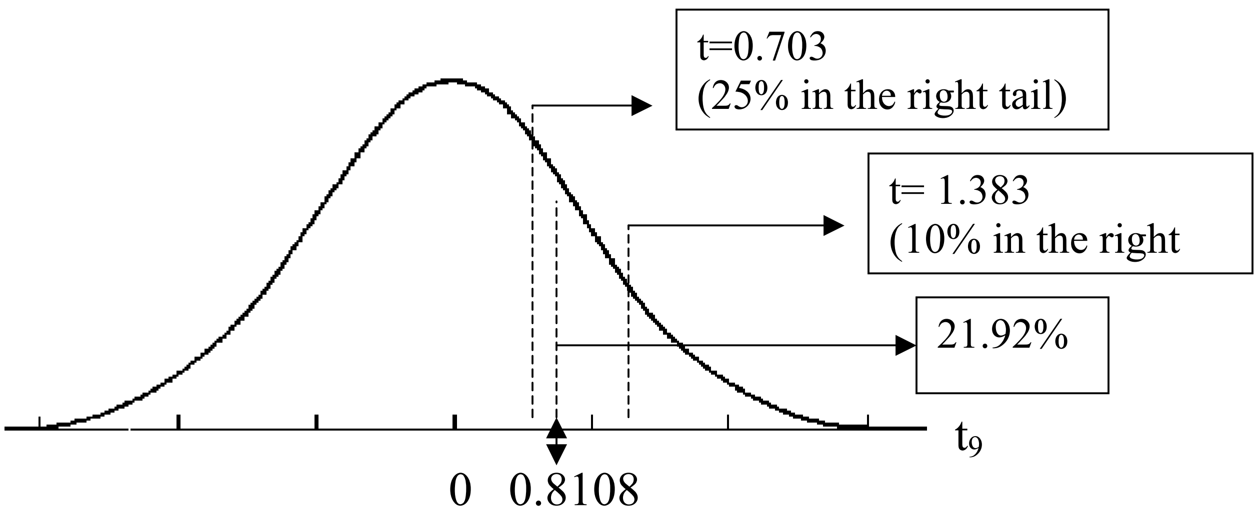
$$\boxed{t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{1250 - 1200}{195/\sqrt{10}} = \frac{50}{61.6644} = 0.8108.}$$

$P(\bar{X} \geq 1250)$: F(0.703)=0.75 $\rightarrow$ 1-F(0.703)=0.25
F(1.383)=0.90 $\rightarrow$ 1-F(1.383)=0.10

$$(0.25 \geq P(t \geq 0.8108) \geq 0.10)$$

Using EViews:

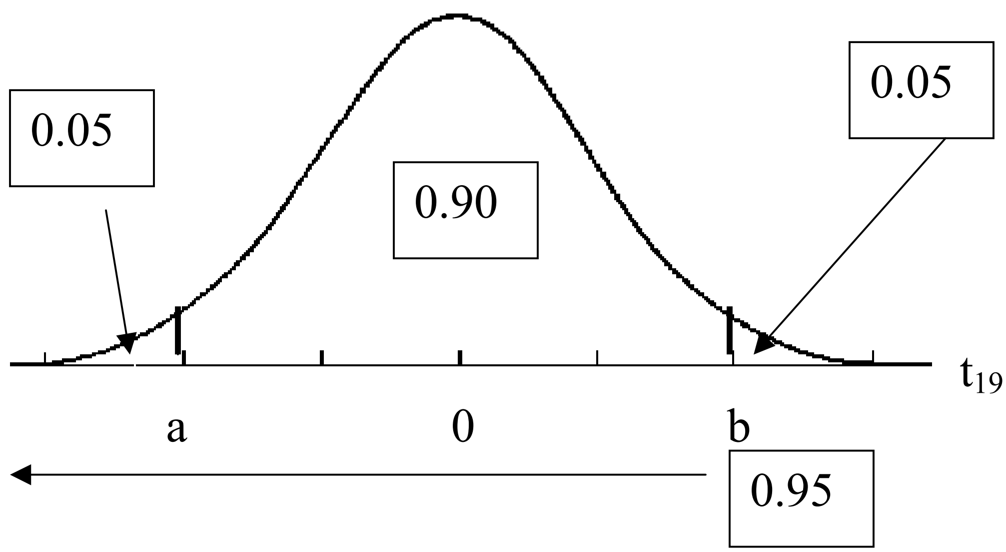




Example:
Determine an interval (a,b) such that $P(a \le t \le b) = 0.90$,
assuming n-1=__ degrees of freedom.

Put half of the excluded area in each tail of the distribution:

$$\left(\frac{1}{2}\right)(0.10) = .05$$



$$P(t \ge b) = 0.05 \quad \Rightarrow \quad F(\_\_\_\_\_\_) = 0.95 \quad \Rightarrow b = 1.729$$

Since the t-distribution is symmetrical, ‘a’ is the negative value of b:

- a = -1.729.

$$P(-\_\_\_\_\_\_ \le t \le 1.729) = 0.90$$

Use of the t-Distribution When the Population is Not ______

The discussion so far regarding the t-distribution assumes that samples drawn are from a normally distributed parent population.

- But often we cannot be sure or we cannot determine if the parent population is ______.

"So how important is this _________ assumption?"

- The ________ assumption can be relaxed without significantly changing the sampling distribution of the t-distribution.

- The distribution is said to be quite ***"robust"***, which implies the results still hold even if the assumptions about the parent population do not conform to the original assumption of _________.

- We must stress that the t-distribution is appropriate whenever 'x' is normal and σ is unknown, even though many t tables do not list values higher than $\nu=30$.

- Some texts suggest that the normal distribution be used to approximate the t-distribution when $\nu > 30$, since t and z-values will then be quite close.

Because of this procedure, the t-distribution is sometimes **erroneously** applied to only _____ samples. But, the t-distribution is ***always*** correct whenever **σ is ________** and x is normal.

Section 7.9
The Sampling Distribution of the Sample Variance s^2, Normal Population

We examined the sampling distribution of $\overline{X}$ to determine how good $\overline{X}$ is as an estimator of μ.

Now we need to examine the sampling distribution of s^2 to consider issues about σ^2.

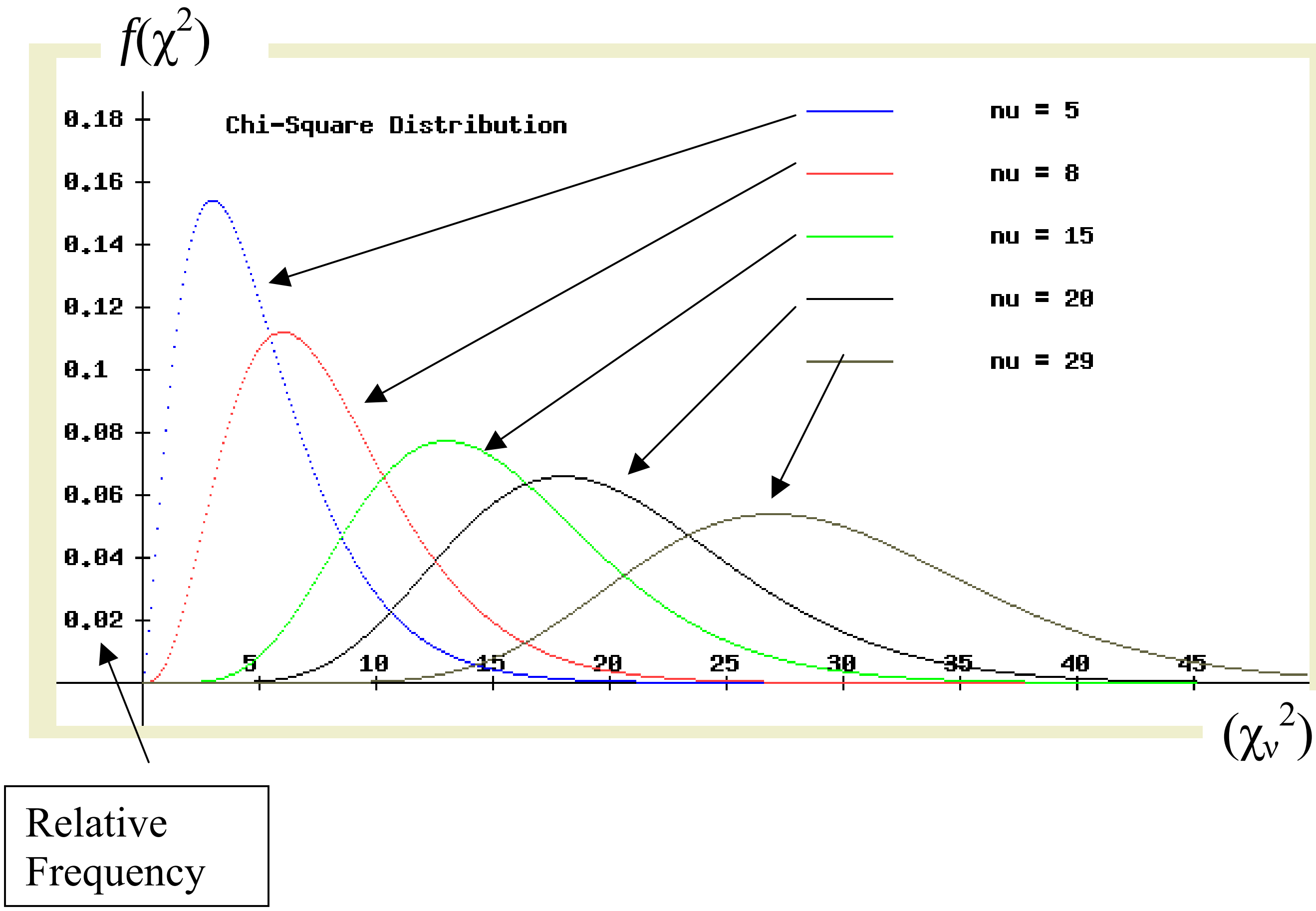
That is, need to explore the distribution that consists of all the possible values of s^2 calculated from samples of size n.

Characteristics of the sample variance:

1) s^2 must always be ________. Hence, the distribution of s^2 cannot be a normal distribution.

"s^2 " is a ________ distribution that is skewed to the _____ and looks like a smooth curve.

Sampling is from a normal population and it has one parameter, the degree of freedom.



The shape depends on the sample _____.

2) The usual application involving s^2, is analyzing whether s^2 will be larger or _______r than some observed value, given some assumed value of σ^2.

Example:

Given σ^2=0.020, what is the probability that a random sample of n=10 will result in a sample variance s^2 = 0.015?

P(s^2 ≥0.015) assuming (n-1) =9 and σ^2= 0.02?

- We cannot directly solve this type of problem.

We must transform it: "Multiply s^2 by (n-1) then divide the product by σ^2."

This new random variable is denoted "χ^2" $\rightarrow$ ___-______

- The Chi-squared distribution is part of a family of positively _______density functions, which depend on one parameter, n-1, which is its degree of freedom:

$$\boxed{\chi^2_{n-1} = \frac{v(s^2)}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2}} \quad (7.12)$$

If s^2 is the ________ of random samples of size n taken from a normal population having a variance of σ^2, then the variable

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

has the same distribution as a χ^2-variable with (n-1) d.f.

Solving a problem involving s^2 by 7.12 follows the same process as solving problems for $\overline{X}$.

<u>Example Continuation:</u>

$$P(s^2 \geq 0.015) = P\left[\frac{(n-1)s^2}{\sigma^2} \geq \frac{(9)0.015}{0.02}\right] = P(\chi^2_9 \geq 6.75)$$

Properties of χ^2 Distribution

1) The number of degrees of freedom in a χ^2 distribution determine its _____ f(χ^2).

- When the degrees of freedom is _____, the shape of the density function is highly skewed to the _____.

- As ν gets larger, the distribution becomes more ___________.
- As $\nu \rightarrow \infty$. The chi-square distribution becomes ______.

2) χ^2 is never less than ____. It has values between zero and positive infinity.

3) E(χ^2_υ) =_ _

4) V(χ^2_υ) =_ _ _

Table VII in Appendix C gives values of the cumulative χ^2 distribution for selected values of ν.

<u>Example:</u> Use the Chi-squared distribution to solve the following: Assume the sample variance equals $\$^2 16$, $s^2=16$,

the population variance $=\$^2 9$, $\sigma^2 = 9$,

and the sample is of size 11, n=11.

What is the probability that $P(s^2 \geq 16)$?

$$\upsilon s^2 / \sigma^2 = \frac{(n-1)s^2}{\sigma^2} = 10(16)/9 = \_\_\_\_\_\_$$

From Appendix C:

F(0.950) = _______
F(0.975) = _______

<u>Meaning</u>: $0.95 < P(\chi^2 \leq 17.78) < 0.975$

This implies $0.10 > P(\chi^2 \geq 17.78) > 0.05$