# 5) Sampling Theory.

**Def:** The probability distribution of a statistic is called Sampling Distribution.

So, the probability distribution of $\bar{X}$ is called the Sampling Distribution of the MEAN.

It is customary to refer to the standard deviation of the sampling distribution as the Standard Error of the statistic.

Therefore,

* the standard error of the mean is just the standard deviation of the sampling distribution of $\bar{X}$.

also,

* the standard error of the sample standard deviation for all possible samples of size $n$ selected from a specified population is the standard deviation of the statistic $S$.

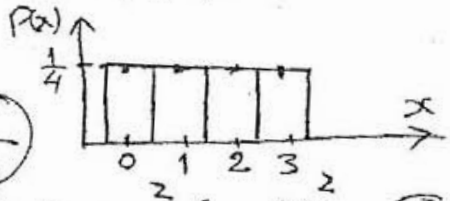
The sampling distribution of statistic will depend on,

* the size of the population,
* the size of the samples, and
* the method of choosing the samples.

## Sampling distribution of the MEAN :

(1) Consider a population consisting of the values 0, 1, 2, 3

$\therefore P(X=x) = \frac{1}{4}$ , $x = 0,1,2,3$ , zero, elsewhere




$\therefore \mu = E(X) = \sum_{x=0}^{3} x P(x) = \frac{0+1+2+3}{4} = \frac{3}{2}$

& $\sigma^2 = E(X-\mu)^2 = \sum_{x=0}^{3} (x-\mu)^2 P(x) = \frac{(0-\frac{3}{2})^2 + (1-\frac{3}{2})^2 + (2-\frac{3}{2})^2 + (3-\frac{3}{2})^2}{4} = \frac{5}{4}$

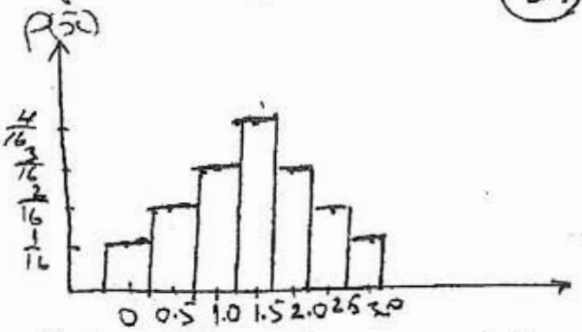
All possible samples of size 2, with replacement are:

| Sample | $\bar{x}$ | No. | Sample | $\bar{x}$ | No. | Sample | $\bar{x}$ | No. | Sample | $\bar{x}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0,0 | 0 | 5 | 1,0 | 0.5 | 9 | 2,0 | 1.0 | 13 | 3,0 | 1.5 |
| 0,1 | 0.5 | 6 | 1,1 | 1.0 | 10 | 2,1 | 1.5 | 14 | 3,1 | 2.0 |
| 0,2 | 1.0 | 7 | 1,2 | 1.5 | 11 | 2,2 | 2.0 | 15 | 3,2 | 2.5 |
| 0,3 | 1.5 | 8 | 1,3 | 2.0 | 12 | 2,3 | 2.5 | 16 | 3,3 | 3.0 |

Sampling distribution of X with replacement

| $\bar{x}$ | 0 | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 |
|---|---|---|---|---|---|---|---|
| f | 1 | 2 | 3 | 4 | 3 | 2 | 1 |
| $P(\bar{x})$ | $\frac{1}{16}$ | $\frac{2}{16}$ | $\frac{3}{16}$ | $\frac{4}{16}$ | $\frac{3}{16}$ | $\frac{2}{16}$ | $\frac{1}{16}$ |

Theoretical Sampling Distribution




Probability Histogram of $\bar{X}$ with replacement

$$\mu_{\bar{x}} = \sum \bar{x} P(\bar{x}) = \frac{3}{2} = \mu$$

$$\sigma^2_{\bar{x}} = \sum \left(\bar{x} - \frac{3}{2}\right)^2 P(\bar{x}) = \frac{5}{8} = \frac{\frac{5}{4}}{2} = \frac{\sigma^2}{n}$$

<u>Theorem①</u>: If all possible random samples of size n are drawn with replacement from a finite population of size N with mean $\mu$ and standard deviation $\sigma$, then for n sufficiently large the sampling distribution of the mean $\bar{x}$ will be approximately normally distributed with mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. Hence $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

<u>Note</u>: The above theorem is valid for any finite population when $n \geq 30$. If $n < 30$ the results will be valid only if the population being sampled is not too different from a normal population. If the population is known to be bell-shaped (Normal), the sampling distribution of $\bar{x}$ will be approximately a normal distribution, regardless of the size of the sample.

<u>ex②</u> Given the population 1,1,1,3,4,5,6,6,6,7. Find the probability that a random sample of size 36, selected with replacement, will yield a sample mean greater than 3.8 but less than 4.5 if the mean is measured to the nearest tenth.

<u>Sol.</u>

| $x$ | 1 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| $P(x)$ | 0.3 | 0.1 | 0.1 | 0.1 | 0.3 | 0.1 |

$\Rightarrow$ $\mu = E(X) = 4$

$\sigma^2 = E(X^2) - (E(X))^2 = 5$

$\mu_{\bar{x}} = \mu = 5$, $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{5}{36}$

$$\therefore P(3.85 < \bar{X} < 4.45) = P\left(\frac{3.85-4}{\sqrt{5}/6} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{4.45-4}{\sqrt{5}/6}\right) = P(-0.40 < Z < 1.21)$$

$$= P(Z < 1.21) - P(Z < -0.40)$$

$$= 0.8869 - [1 + 0.6554] = 0.5423$$

To verify the result of this example, we could write values of our population on tags and place them in a box from which we draw samples of size 36 with replacement. If we drew 100 samples of size 36 and computed the sample mean, we obtain what is known as an <u>Experimental Sampling Distribution</u> of $\bar{X}$. This means that approximately 54%, or 54 of our 100 sample means, fall within the interval from 3.85 to 4.45.

---

ex: 3

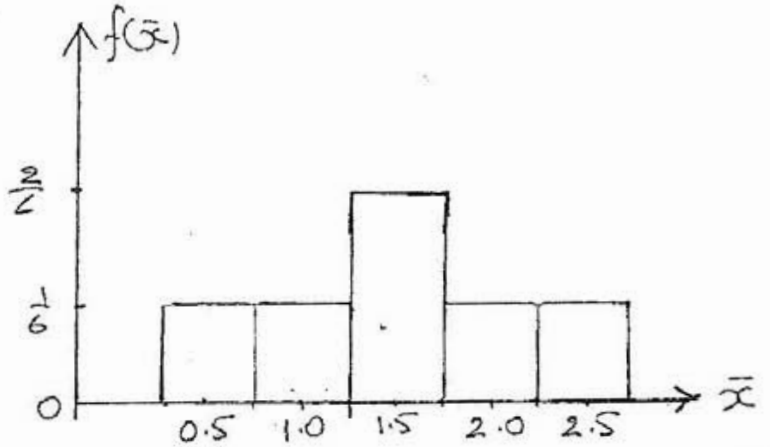
Suppose that we now draw all possible samples of size 2 from the uniform population of example 1 page 49, without replacement, and then for each compute $\bar{x}$.

| No. | Sample | $\bar{x}$ | No. | Sample | $\bar{x}$ |
|---|---|---|---|---|---|
| 1 | 0,1 | 0.5 | 7 | 1,0 | 0.5 |
| 2 | 0,2 | 1.0 | 8 | 2,0 | 1.0 |
| 3 | 0,3 | 1.5 | 9 | 3,0 | 1.5 |
| 4 | 1,2 | 1.5 | 10 | 2,1 | 1.5 |
| 5 | 1,3 | 2.0 | 11 | 3,1 | 2.0 |
| 6 | 2,3 | 2.5 | 12 | 3,2 | 2.5 |

Means of Random Samples without Replacement

| $\bar{x}$ | f | $f(\bar{x})$ |
|---|---|---|
| 0.5 | 2 | $\frac{1}{6}$ |
| 1.0 | 2 | $\frac{1}{6}$ |
| 1.5 | 4 | $\frac{1}{3}$ |
| 2.0 | 2 | $\frac{1}{6}$ |
| 2.5 | 2 | $\frac{1}{6}$ |

Sampling Distribution of $\bar{X}$ without Replacement




Probability Histrogram of $\bar{X}$ Without Replacement

Now, $\mu_{\bar{x}} = \sum \bar{x} f(\bar{x}) = \frac{3}{2} = \mu$

$$\sigma^2_{\bar{x}} = \sum (\bar{x} - \tfrac{3}{2})^2 f(\bar{x}) = \frac{5}{12} = \frac{\frac{5}{4}}{3} * \frac{2}{2} = \frac{\frac{5}{4}}{2} \quad \frac{2}{3} = \frac{\sigma^2}{n} \frac{(N-n)}{(N-1)}$$

Regardless of the size or form of the original population. When $n \geq 30$ and the population size is at least twice the sample size, we may apply the following theorem.

<u>Theorem ②</u>: If all possible random samples of size $n$ are drawn, without replacement, from a finite population of size $N$ with mean $\mu$ and standard deviation $\sigma$, then the sampling distribution of the sample mean $\bar{X}$ will be approximately normally distributed with a mean and standard deviation given by,

$$\mu_{\bar{x}} = \mu \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

<u>ex ④</u> Given the population 1,1,1,3,4,5,6,6,6,7 find the mean and standard deviation for the sampling distribution of means for samples of size 4 selected at random without replacement. Between what two values would you expect at least $\frac{3}{4}$ of the sample means fall?

<u>Sol</u> from ex ② we have $\mu = 4$, $\sigma^2 = 5$

$$\therefore \mu_{\bar{x}} = \mu = 4, \quad \sigma_{\bar{x}} = \frac{\sqrt{5}}{2} \sqrt{\frac{10-4}{10-1}} = 0.85$$

Now by chebyshev's theorem

$$\mu_{\bar{x}} \pm 2\sigma_{\bar{x}} = 4 \pm (2)(0.85) = 4 \pm 1.7 \Rightarrow (2.3, 5.7)$$

So, we expect at least $\frac{3}{4}$ of the sample means to fall between 2.3 and 5.7

The factor $\sqrt{\frac{N-n}{N-1}}$ in the formula for the standard deviation of $\bar{X}$ in the previous theorem is called <u>The Population Correction Factor</u>. For large $N$, relative to the sample size $n$, this correction factor will be close to 1, and $\sigma^2_{\bar{X}}$ will be approximately $\frac{\sigma^2}{n}$. Hence for a large or infinite populations, whether discrete or continuous, we use the following well-known theorem, called <u>The Central Limit Theorem</u>.

<u>Theorem ③</u>: If random samples of size $n$ are drawn from a large or infinite distribution with mean $\mu$ and variance $\sigma^2$, then the sampling distribution of the sample mean $\bar{X}$ is approximately normally distributed with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$. Hence $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \overset{r.v.}{\sim} N(0,1)$.

<u>Note</u> The normal approximation in theorem ③ will be good if $n \geq 30$ regardless of the shape of population. If $n < 30$, the approximation is good only if the population is not too different from a normal population. If the population is known to be normal, the sampling distribution of $\bar{X}$ will follow a normal distribution exactly, no matter how small the size of the samples.

<u>ex ⑤</u>: Refering to ex④ page (45), Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

<u>Sol</u> The sampling distribution of $\bar{X}$ will be approximately normal with $\mu_{\bar{X}} = \mu = 800$ and $\sigma^2_{\bar{X}} = \frac{1600}{16} = 100$ or $\sigma_{\bar{X}} = \frac{40}{\sqrt{16}} = 10$

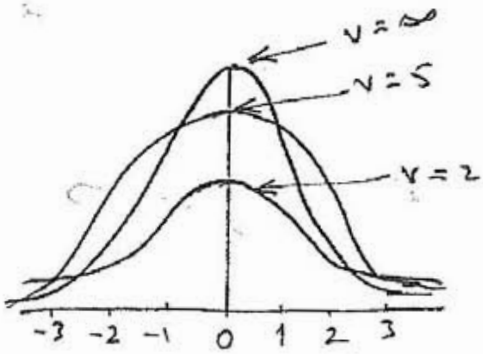
$$\therefore P(\bar{X} < 775) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{775-800}{40/4}\right) = P(Z < -2.5) = 0.0062$$

## t-Distribution:

**Theorem ④:** If $\bar{x}$ and $s^2$ are the mean and variance, respectively, of a random sample of size $n$ taken from a population that is normally distributed with mean $\mu$ and variance $\sigma^2$, then the $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ is a value of a random variable $T$ having t dist. with $v = n-1$ degress of freedom.

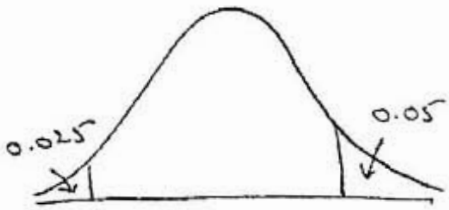
t distribution curves for $v = 2, 5$ & $\infty$

as $v \to \infty$ t distribution approches to standard normal distribution




**ex. ①** Find $P(-t_{0.025} < T < t_{0.05})$

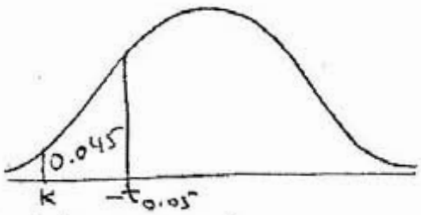
**sol**




$$P(-t_{0.025} < T < t_{0.05}) = 1 - (0.025 + 0.05)$$
$$= 0.925$$

**ex ②:** Find $k$ such that $P(k < T < -1.761) = 0.045$ for a random sample of size 15 selected from a normal dis

**sol** since $n = 15$, hence $v = n - 1 = 14$




$\because t_{0.05}(14) = 1.761$, $\therefore -t_{0.05}(14) = -1.761$

$\because$ k is to the left of $-t_{0.05}(14) = -1.761$, $\therefore$ let $k = -t_{\alpha}$

$\therefore \alpha = 0.05 - 0.045 = 0.005$

$\because t_{0.005}(14) = 2.997$, $\therefore k = -2.997$

$$\therefore P(-2.977 < T < -1.761) = 0.045$$

ex ③ A manufacturer of light bulbs claims that his bulbs will burn on the average 500 hours. To maintain this average, he tests 25 bulbs each month. If the computed $t$ value falls between $-t_{0.05}$ and $t_{0.05}$, he is satisfied with his claim. What conclusion should be draw from a sample that has a mean $\bar{x}=518$ hours and a standard deviation $s=40$ hours? Assume the distribution of burning times to be approximately normal.

Sol. $v = n-1 = 25-1 = 24 \Rightarrow t_{0.05}(24) = 1.711$

$\therefore$ He is satisfied if $-1.711 < t < 1.711$

Now $t = \dfrac{518-500}{40/\sqrt{25}} = 2.25$

since $t = 2.25 > 1.711 \Rightarrow$ The manufacturer may conclude that his bulbs are a better product than he thought, since $t = 2.25$ matches approximately $\alpha = 0.02$ $(2.064 < \alpha = 0.02 < 2.492)$

---

## Sampling Distribution of the differences of Means:

Theorem ⑤

If independent samples of size $n_1$ and $n_2$ are drawn from two large or infinite populations, discrete or continuous, with means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$, respectively, then the sampling distribution of the differeces of means, $\bar{X}_1 - \bar{X}_2$, is approximately normally distributed with mean and variance given by:

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2 \quad \text{and} \quad \sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Hence $$Z = \frac{(\bar{x}_1-\bar{x}_2) - (\bar{X}_1-\bar{X}_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}} \overset{r.v.}{\sim} N(0,1)$$

Note: If both $n_1$ and $n_2$ are greater than or equal to 30, the normal approximation for the distribution of $\bar{X}_1 - \bar{X}_2$ is very good regardless of the shapes of the two populations. However, even when $n_1$ and $n_2$ are less than 30, the normal approximation is reasonably good except when the populations are decidedly nonnormal.

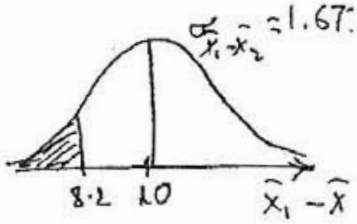
ex ①: A sample of size $n_1 = 5$ is drawn at random from a population is normally distributed with mean $\mu_1 = 50$ and variance $\sigma_1^2 = 9$, and the sample mean $x_1$ is recorded.
A second random sample of size $n_2 = 4$ is selected, independent of the first sample, from a different population that is also normally distributed, with mean $\mu_2 = 40$ and variance $\sigma_2^2 = 4$, and the sample mean $\bar{x}_2$ is recorded. What is the $P(\bar{X}_1 - \bar{X}_2 < 8.2)$

sol

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2 = 50 - 40 = 10$$

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{9}{5} + \frac{4}{4} = 2.8$$

$$\therefore \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{2.8} = 1.673$$




$$P(\bar{X}_1 - \bar{X}_2 < 8.2) = P\left(\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{1/2}} = \frac{8.2 - 10}{1.673}\right) = P(Z < -1.0$$

$$= 1 - P(Z < 1.08) = 1 - 0.8599 = 0.1401$$

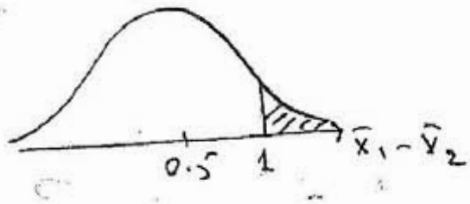
ex ②: The television picture tubes of manufacturer A have a mean lifetime of 6.5 years and a standard deviation of 0.9 years, while those of manufacturer B have a mean lifetime of 6.0 years and standard deviation of 0.8 years. What is the probability that a random sample of 36 tubes from manufacturer A will have a mean lifetime that is at least ONE year more than the mean lifetime of a sample of 49 tubes from manufacturer B?

Sol $\mu_{\bar{X}_1-\bar{X}_2} = 6.5 - 6.0 = 0.5$

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{0.81}{36}+\frac{0.64}{49}} = 0.189$$

| Pop. 1 | Pop. 2 |
|---|---|
| $\mu_1 = 6.5$ | $\mu_2 = 6.0$ |
| $\sigma_1 = 0.9$ | $\sigma_2 = 0.8$ |
| $n_1 = 36$ | $n_2 = 49$ |

$$P(\bar{X}_1-\bar{X}_2 \geq 1.0) = P(Z \geq 2.65)$$
$$= 1 - P(Z < 2.65)$$
$$= 1 - 0.9960 = 0.0040$$




Theorem 6: If the random variables $X$ and $Y$ are independent and normally distributed with means $\mu_X$ and $\mu_Y$ and variances $\sigma_X^2$ and $\sigma_Y^2$, respectively, then the distribution of the difference $X-Y$ is normally distributed with mean $\mu_{X-Y} = \mu_X - \mu_Y$ and variance $\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$

---

**EXERCISES**

1. Random samples of size 4 are drawn, with replacement, from the finite population 2, 4, and 6.
   (a) Assuming that the 81 possible samples are all equally likely to occur, construct the sampling distribution of $\bar{X}$.
   (b) Construct a probability histogram for the sampling distribution of $\bar{X}$.
   (c) Verify that $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}}^2 = \sigma^2/n$.
   (d) Between what two values would you expect the middle 68% of the sample means to fall?

2. If, in Exercise 1, a sample of size 54 is drawn with replacement, what is the probability that the sample mean will be greater than 4.1 but less than 4.4? Assume the means to be measured to the nearest tenth.

3. A finite population consists of the numbers 2, 2, 4, 6, and 6 written on 5 tags, each of a different color.
   (a) Assuming that the 25 possible samples of size 2 that can be selected at random, with replacement, are all equally likely to occur, construct the sampling distribution of $\bar{X}$.
   (b) Construct a probability histogram for the sampling distribution of $\bar{X}$.
   (c) Verify that $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}}^2 = \sigma_X^2/n$.

4. Random samples of size 2 are drawn, without replacement, from the finite population 1, 1, 1, 2, 2, 3, and 4.
   (a) Assuming that the 42 possible samples are all equally likely to occur, construct the sampling distribution of $\bar{X}$.
   (b) Verify that $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = (\sigma/\sqrt{n})\sqrt{(N-n)/(N-1)}$.
   (c) Between what two values would you expect at least $\frac{8}{9}$ of the sample means to fall?

5. A certain type of thread is manufactured with a mean tensile strength of 78.3 kilograms and a standard deviation of 5.6 kilograms. Assuming that the population is infinite, how is the standard error of the mean changed when the sample size is
   (a) increased from 64 to 196?
   (b) decreased from 784 to 49?

6. If the standard error of the mean for the sampling distribution of random samples of size 36 from a large or infinite population is 2, how large must the size of the sample become if the standard error is to be reduced to 1.2?

7. Find the value of the finite population correction factor when
   (a) $n = 2$ and $N = 5$;
   (b) $n = 10$ and $N = 1000$;
   (c) $n = 40$ and $N = 10{,}000$.

8. If all possible samples of size 16 are drawn from a normal population with mean equal to 50 and standard deviation equal to 5, what is the probability that a sample mean $\overline{X}$ will fall in the interval from $\mu_{\overline{X}} - 1.9\sigma_{\overline{X}}$ to $\mu_{\overline{X}} - 0.4\sigma_{\overline{X}}$? Assume that the sample means can be measured to any degree of accuracy.

9. A soft-drink machine is being regulated so that the amount of drink dispensed averages 240 milliliters with a standard deviation of 15 milliliters. Periodically, the machine is checked by taking a sample of 40 drinks and computing the average content. If the mean of the 40 drinks is a value within the interval $\mu_{\overline{X}} \pm 2\sigma_{\overline{X}}$, the machine is thought to be operating satisfactorily; otherwise, adjustments are made. In Section 8.1, the company official found the mean of 40 drinks to be $\overline{x} = 236$ milliliters and concluded that the machine needed no adjustment. Was this a reasonable decision?

10. The heights of 1000 students are approximately normally distributed with a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters. If 200 random samples of size 25 are drawn from this population and the means recorded to the nearest tenth of a centimeter, determine
    (a) the mean and standard error of the sampling distribution of $\overline{X}$;
    (b) the number of sample means that fall between 172.5 and 175.8 centimeters inclusive;
    (c) the number of sample means falling below 172.0 centimeters.

11. The random variable $X$, representing the number of cherries in a cherry puff, has the following probability distribution:

| $x$ | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| $P(X = x)$ | 0.2 | 0.4 | 0.3 | 0.1 |

    (a) Find the mean $\mu$ and the variance $\sigma^2$ of $X$.
    (b) Find the mean $\mu_{\overline{X}}$ and the variance $\sigma^2_{\overline{X}}$ of the mean $\overline{X}$ for random samples of 36 cherry puffs.

(c) Find the probability that the average number of cherries in 36 cherry puffs will be less than 5.5.

12. If a certain machine makes electrical resistors having a mean resistance of 40 ohms and a standard deviation of 2 ohms, what is the probability that a random sample of 36 of these resistors will have a combined resistance of more than 1458 ohms?

13. (a) Find $t_{0.025}$ when $\nu = 14$.
    (b) Find $-t_{0.01}$ when $\nu = 10$.
    (c) Find $t_{0.995}$ when $\nu = 7$.

14. (a) Find $P(T < 2.365)$ when $\nu = 7$.
    (b) Find $P(T > 1.318)$ when $\nu = 24$.
    (c) Find $P(-1.356 < T < 2.179)$ when $\nu = 12$.
    (d) Find $P(T > -2.567)$ when $\nu = 17$.

15. (a) Find $P(-t_{0.005} < T < t_{0.01})$.
    (b) Find $P(T > -t_{0.025})$.

16. Given a random sample of size 24 from a normal distribution, find $k$ such that
    (a) $P(-2.069 < T < k) = 0.965$.
    (b) $P(k < T < 2.807) = 0.095$.
    (c) $P(-k < T < k) = 0.90$.

17. A manufacturing firm claims that the batteries used in their electronic games will last an average of 30 hours. To maintain this average, 16 batteries are tested each month. If the computed $t$ value falls between $-t_{0.025}$ and $t_{0.025}$, the firm is satisfied with its claim. What conclusion should the firm draw from a sample that has a mean $\bar{x} = 27.5$ hours and a standard deviation $s = 5$ hours? Assume the distribution of battery lives to be approximately normal.

18. A normal population with unknown variance is believed to have a mean of 20. Is one likely to obtain a random sample of size 9 from this population that has a mean $\bar{x} = 24$ and a standard deviation of $s = 4.1$? If not, what conclusion would you draw?

19. A cigarette manufacturer claims that his cigarettes have an average nicotine content of 1.83 milligrams. If a random sample of 8 cigarettes of this type have nicotine contents of 2.0, 1.7, 2.1, 1.9, 2.2, 2.1, 2.0, and 1.6 milligrams, would you agree with the manufacturer's claim?

20. Let $\bar{X}_1$ represent the mean of a sample of size $n_1 = 2$, selected with replacement, from the finite population $-2$, 0, 2, and 4. Similarly, let $\bar{X}_2$ represent the mean of a sample of size $n_2 = 2$, selected with replacement, from the population $-1$ and 1.
    (a) Assuming that the 64 possible differences $\bar{x}_1 - \bar{x}_2$ are equally likely to occur, construct the sampling distribution of $\bar{X}_1 - \bar{X}_2$.