# CHAPTER 5
# PROBABILISTIC FEATURES OF THE DISTRIBUTIONS OF CERTAIN SAMPLE STATISTICS

**Key Words**

- Sampling Distributions
- Distribution of the Sample Mean
- Distribution of the difference between Two Sample Means
- Central Limit Theorem

# 5.1 INTRODUCTION

In real life calculating parameters of populations is prohibitive because populations are very large. Rather than investigating the whole population, we take a sample, calculate a statistic related to the parameter of interest, and make an inference.

The sampling distribution of the statistic is the tool that tells us how close is the statistic to the parameter.

# 5.2 SAMPLING DISTRIBUTIONS

## Definition

The sampling distribution of a statistic is the distribution of all possible values that can be assumed by that statistic, computed from samples of the same size randomly drawn from the same population.

# 5.3 DISTRIBUTION OF THE SAMPLE MEAN

- Suppose there is a population of size $N=4$.
- Random variable, *X*, is age of Individuals.
- Values of *X* : 18, 20, 22, 24 are measured in years.
- Everyone is one of these 4 ages in this population.

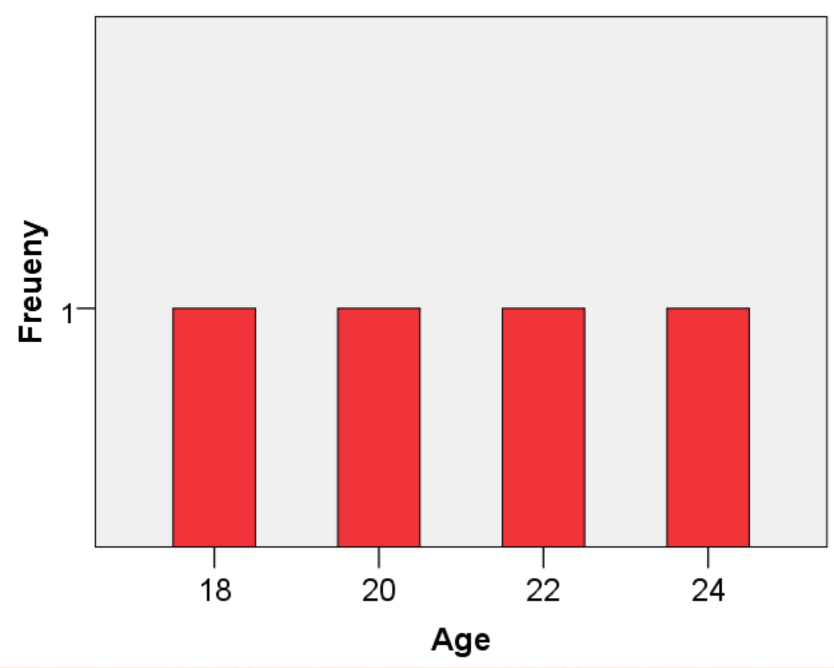
# POPULATION CHARACTERISTICS

## Summary Measure

$$\mu_X = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{18+20+22+24}{4} = 21$$

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}} = 2.236$$

## Population Distribution




**Uniform Distribution**

# All Possible Samples of Size n = 2

- **Draw a sample of size 2 with replacement.**
- **There are 16 samples.**

| 1st Obs | 2nd observation | | | |
|---|---|---|---|---|
| | 18 | 20 | 22 | 24 |
| 18 | (18,18) | (18, 20) | (18, 22) | (18, 24) |
| 20 | (20.,18) | (20, 20) | (20, 22) | (20, 24) |
| 22 | (22,18) | (22, 20) | (22, 22) | (22, 24) |
| 24 | (24,18) | (24, 20) | (24, 22) | (24, 24) |

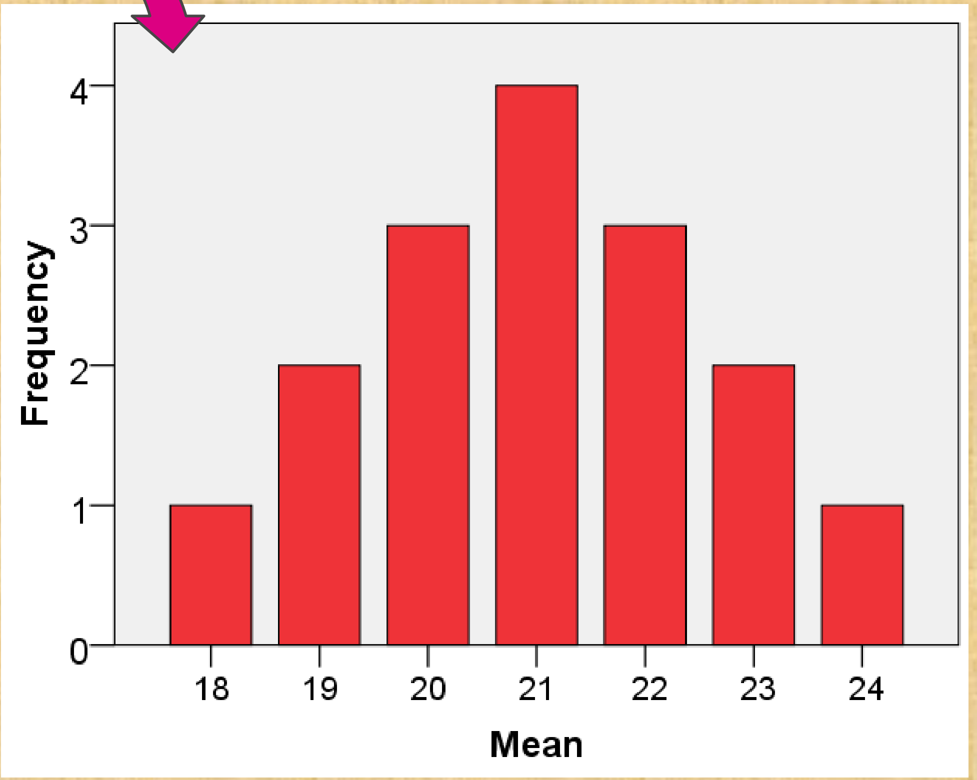
**16 Sample Means**

| 1st Obs | 2nd observation | | | |
|---|---|---|---|---|
| | 18 | 20 | 22 | 24 |
| 18 | 18 | 19 | 20 | 21 |
| 20 | 19 | 20 | 21 | 22 |
| 22 | 20 | 21 | 22 | 23 |
| 24 | 21 | 22 | 23 | 24 |

# Sampling Distribution of All Sample Means

## 16 Sample Means

| Mean | Frequency |
|---|---|
| 18 | 1 |
| 19 | 2 |
| 20 | 3 |
| 21 | 4 |
| 22 | 3 |
| 23 | 2 |
| 24 | 1 |



# Summary Measures for the Sampling Distribution

$$\mu_{\bar{X}} = \frac{\sum_{i=1}^{N} \bar{X}_i}{N} = \frac{18+19+19+\cdots+24}{16} = 21$$

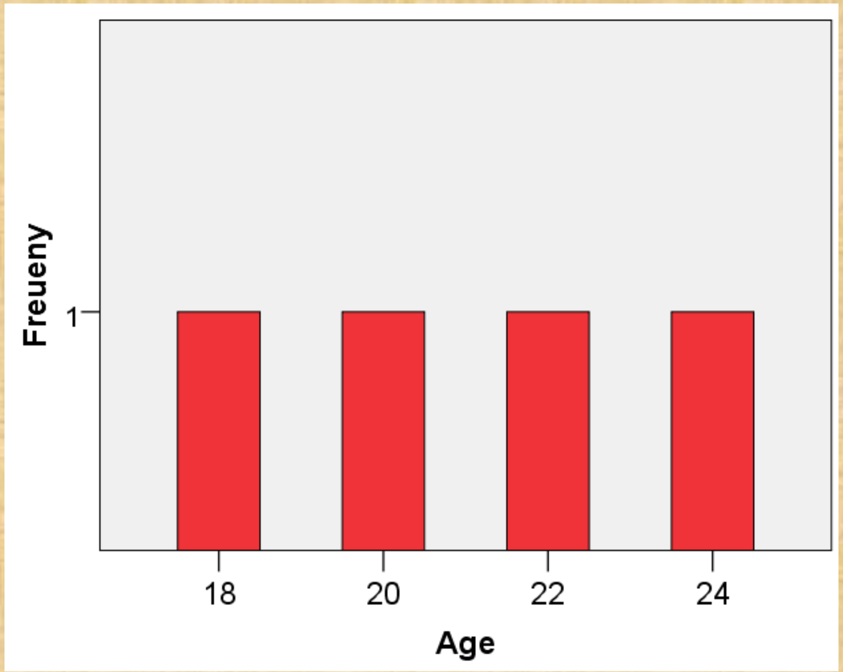
$$\sigma_{\bar{X}} = \sqrt{\frac{\sum_{i=1}^{N} \left(\bar{X}_i - \mu_{\bar{X}}\right)^2}{N}}$$

$$= \sqrt{\frac{(18-21)^2 + (19-21)^2 + \cdots + (24-21)^2}{16}} = 1.58$$

# Comparing the Population with its Sampling Distribution

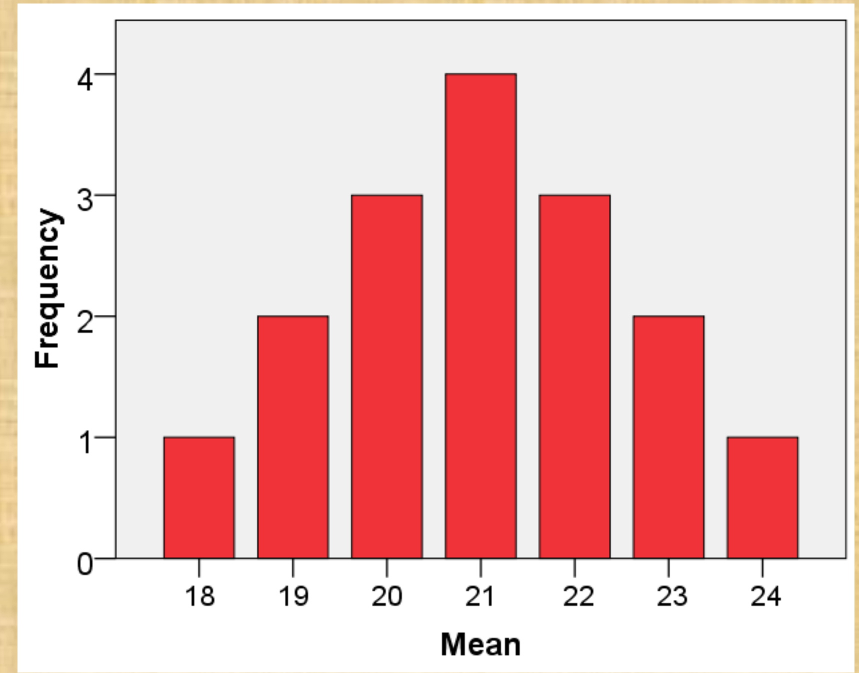
**Population**

$\mu = 21, \quad \sigma = 2.236$




**Sample Means Distribution, n = 2**

$\mu_{\bar{x}} = 21, \ \sigma_{\bar{X}} = 1.58$



# Results for other sample sizes

| Sample Size | Mean | Variance | St. Dev |
|---|---|---|---|
| 1 | 21 | 5 | 2.236 |
| 2 | 21 | 2.5 | 1.581139 |
| 3 | 21 | 1.66667 | 1.290994 |
| 4 | 21 | 1.25 | 1.118034 |
| 5 | | | |

$$\mu_{\bar{x}} = \mu_x, \qquad \sigma^2_{\bar{x}} = \frac{\sigma^2_x}{n}$$

# Sampling Distribution of the Sample Mean, $\bar{X}$ :

## Sampling From Normally distributed Population

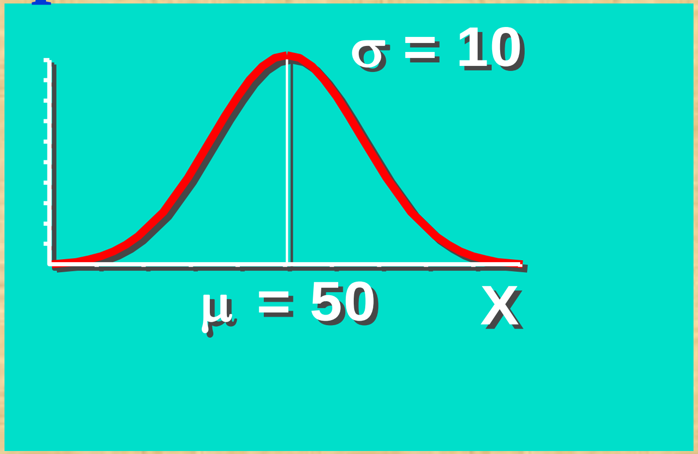
### Central Tendency

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

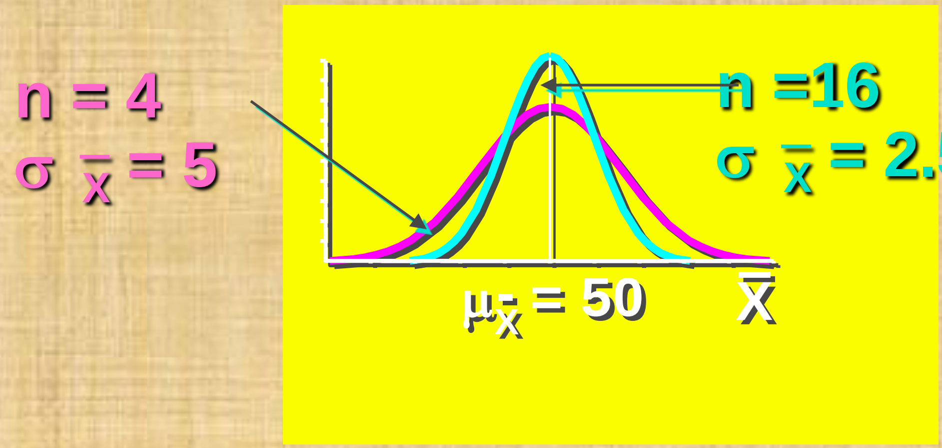
### Variation

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

### Population Distribution




### Sampling Distributions



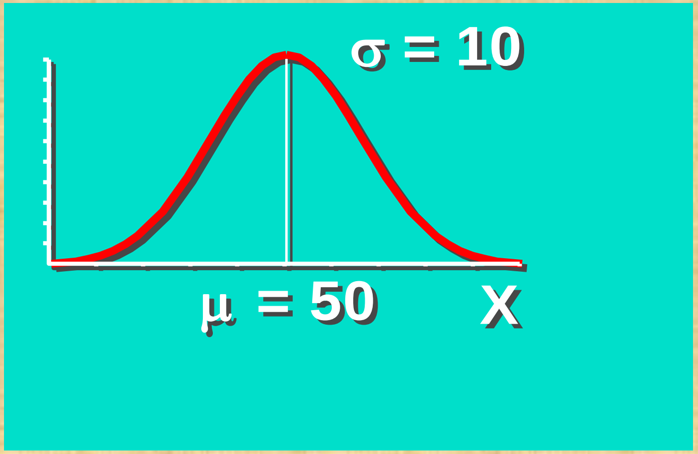
## Central Tendency

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

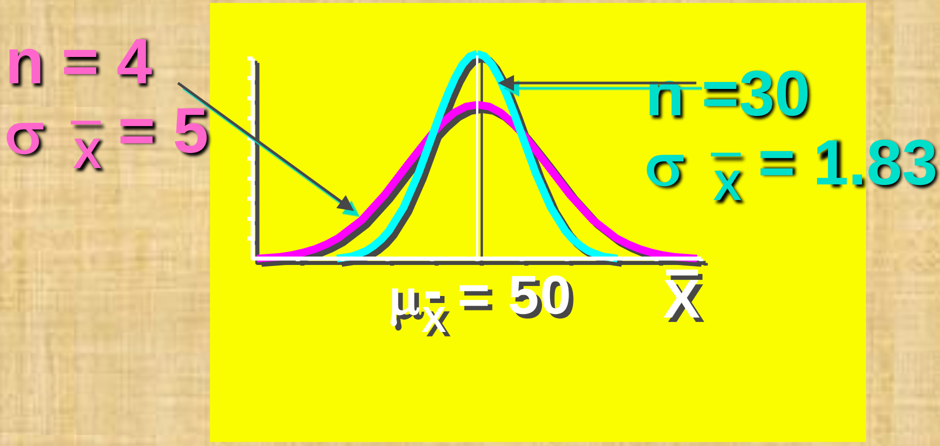
## Variation

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

## Population Distribution

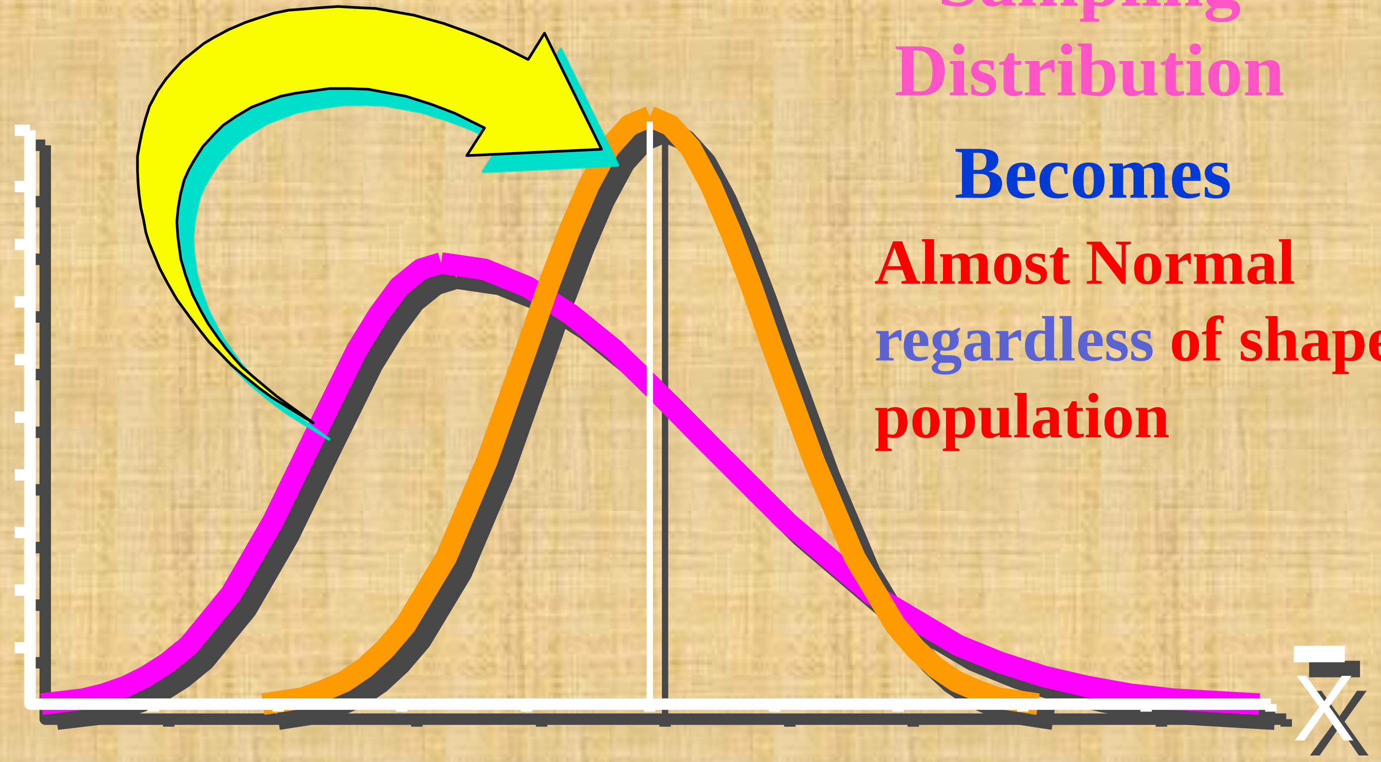



## Sampling Distributions



# CENTRAL LIMIT THEOREM

As Sample Size Gets Large Enough




Sampling Distribution

Becomes

Almost Normal regardless of shape of population

# CENTRAL LIMIT THEOREM

If a random sample is drawn from any population, with a mean $\mu$ and a finite variance $\sigma^2$, then the sampling distribution of the sample mean computed from samples of size *n* from this population will have mean $\mu$ and variance $\sigma^2/n$ and will be approximately normally distributed for a sufficiently large sample size.

# Central Limit Theorem

The larger the sample size, the more closely the sampling distribution of $\overline{X}$ will resemble a normal distribution. As the sample size increases the sampling distribution of the sample mean approaches the normal distribution with mean $\mu$ and variance $\sigma^2/n$.

$$\overline{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

Or

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1)$$

**Example: Exercise 5.3.3 Page 145**

If the uric acid values in normal adult males are approximately normally distributed with a mean and standard deviation of 5.7 and 1 mg percent, respectively.

The random variable X is uric acid value in normal adult male

a) Find the probability that a normal adult male picked at normal from this population will have a uric acid value greater than 6.

Solution:

$$P(X > 6) = P\left(\frac{X - \mu}{\sigma_x} > \frac{6 - 5.7}{1}\right)$$

$$= P(Z > 0.30) = 1 - P(Z < 0.30)$$

$$= 1 - 0.6179 \quad = 0.3821$$

b) Find the probability that a random sample of size 9 will yield a mean greater than 6

Solution

- The random variable here is the mean ($\overline{X}$) uric acid values.

$$P(\overline{X} > 6) = P\left(\frac{\overline{X} - \mu}{\sigma_{\overline{x}}} > \frac{6 - 5.7}{1/\sqrt{9}}\right)$$

$$= P(Z > 0.9) = P(Z < -0.9) = 0.1841$$

c) Find the probability that a random sample of size 9 will yield a mean between 5 and 6

Solution

$$P(5 < \bar{X} < 6) = P\left(\frac{5-5.7}{1/\sqrt{9}} < \frac{\bar{X}-\mu}{\sigma_{\bar{x}}} < \frac{6-5.7}{1/\sqrt{9}}\right)$$

$$= P(-2.1 < Z < 0.9) = P(Z < 0.9) - P(Z < -2.1)$$

$$= 0.8159 - 0.0179 = 0.7980$$

d) Find the probability that a random sample of size 9 will yield a mean less than 5.2

Solution

$$P(\overline{X} < 5.2) = P\left(\frac{\overline{X} - \mu}{\sigma_{\overline{x}}} < \frac{5.2 - 5.7}{1/\sqrt{9}}\right)$$

$$= P(Z < -1.5)$$

$$= 0.0668$$

# 5.4 Sampling Distribution of the Difference Between Two Means

- The difference between two means can become a parameter of interest when the comparison between two populations is studied.

- To make an inference about $\mu_1$ - $\mu_2$ we observe the distribution of $\bar{x}_1 - \bar{x}_2$

# Sampling From Normal Populations

Given two normally distributed populations with means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$, respectively, the sampling distribution of the difference $\bar{x}_1 - \bar{x}_2$ , between the means of independent samples of size $n_1$ and $n_2$ drawn from these populations is normally distributed with mean $\mu_1$ - $\mu_2$ and variance

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\mu_{\bar{X}_1-\bar{X}_2}=\mu_1-\mu_2$$

$$\sigma^2_{\bar{X}_1-\bar{X}_2}=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

$$Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\approx \mathbf{N(0,1)}$$

## Example: Exercise 5.4.5 Page 151

For a population of 17-year-old boys and 17-year-old girls, the mean and standard deviations, respectively of their sub scapular skin fold thickness values are as follows: boys, 9.7 and 6.0; girls, 15.6 and 9.5. Simple random samples of 40 boys and 35 girls are selected from the populations. What is the probability that the difference between sample means (girls – boys) will be grater that 10.

# Solution

$$P(\bar{X}_{girls} - \bar{X}_{boys} > 10) =$$

$$= P\left( \frac{(\bar{X}_{girls} - \bar{X}_{boys}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > \frac{10 - (15.6 - 9.7)}{\sqrt{\frac{9.5^2}{35} + \frac{6^2}{40}}} \right)$$

$$= P(Z > 2.20) = 0.0139$$