7.1 Sampling Distributions

Definition. Let $\bar{x}$ denote the sample mean. For different samples, the values of $\bar{x}$ are different, but there exists a probability distribution of those values, called the **sampling distribution of** $\bar{x}$.

7.2 Mean and Standard Deviation of $\bar{x}$

Definition. Consider a population with mean μ and standard deviation σ. Suppose a random sample of size n is drawn. The **mean of** $\bar{x}$ is

$$\mu_{\bar{x}} = E(\bar{x}) = \mu$$

and the **standard deviation** of $\bar{x}$ is

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

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

Note two properties:

- The expected value of $\overline{x}$ coincides with the true population mean μ. Hence, $\overline{x}$ is said to be an **unbiased estimator** of μ.
- The standard deviation of $\overline{x}$ decreases as the sample size n increases, as should be since if we sample the entire population, then there will be no error in estimation of μ.

7.3 Shape of Sampling Distribution of $\bar{x}$

The Central Limit Theorem (CLT)

If the sample size n is large enough ($n \geq 30$), then the sampling distribution of $\bar{x}$ is approximately **normal** with mean $\mu_{\bar{x}} = \mu$ and standard deviation

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

7.4 Applications of Sampling Distribution of $\bar{x}$

Example. In a very large group of gifted children, the average IQ is120 and the standard deviation is 15. If 100 children are randomly selected from this group, what is the probability that the average IQ of the children in the sample will be less than 117?

Solution. By the CLT, the average IQ in this sample $\overline{x}$ is approximately normal with mean $\mu = 120$ and standard deviation $\frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1.5.$ We compute

$$P(\bar{x} < 117) = P\left(\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < \frac{117 - 120}{1.5}\right) = P(z < -2) = 0.0228.$$

Example. The amount of time that a drive-through bank teller spends on a customer is a random variable with a mean of 3.2 minutes and a standard deviation of 1.6 minutes. If a random sample of 64 customers is observed, find the probability that their mean time at the teller's counter is at least 3 minutes.

Solution. We compute

$$P(\bar{x} \geq 3) = P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \geq \frac{3-3.2}{1.6/\sqrt{64}}\right) = P(z \geq -1)$$

$$= P(z \leq 1) = 0.8413.$$

Example. Suppose the number of DVDs owned by students has mean 74 and standard deviation 96. Find the approximate probability that the average number of DVDs owned when 100 students are asked is between 70 and 90.

Solution.

$$P(70 \le \bar{x} \le 90) = P\left(\frac{70-74}{96/\sqrt{100}} \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le \frac{90-74}{96/\sqrt{100}} \right)$$

$$= P(-0.42 \le z \le 1.67) = P(z \le 1.67) - P(z \le -0.42)$$

$$= 0.9525 - 0.3372 = 0.6193.$$

7.5 Population and Sample Proportions

Definition. Suppose N is the total size of a population, and X is the number of elements in the population that possess a specific characteristic. Then the **population proportion** is $p = \frac{X}{N}$

Definition. Suppose n is the size of a sample, and x is the number of elements in the sample that possess a specific characteristic. Then the **sample proportion** is

$$\hat{p} = \frac{x}{n}$$

(read "*p hat*")

7.5 Mean, Standard Deviation, and Shape of the Sampling Distribution of $\hat{p}$

By the Central Limit Theorem, if n is large ($n \geq 30$), then the **sampling distribution** of $\hat{p}$ is approximately normal with mean

$$\mu_{\hat{p}} = E(\hat{p}) = p$$

and standard deviation $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$

7.6 Applications of the Sampling Distribution of $\hat{p}$

Example. Suppose that 88% of the cases of car burglar alarms that go off are false alarms. Find the probability that in a sample of 100 alarms going off there will be at most 80% of false alarms.

Solution. The sample proportion of false alarms $\hat{p}$ is approximately normal with mean 0.88 and standard deviation $\sqrt{\frac{(0.88)(1-0.88)}{100}} = 0.0325$

Thus,

$$P(\hat{p} \le 0.8) = P\left(z \le \frac{0.8-0.88}{0.0325}\right) = P(z \le -2.46) = 0.0069$$

Example. Assuming that there are equal number of newborn boys and girls in a population, about what proportion of samples of size 200 contain at least 55% boys?

Solution. The sample proportion $\hat{p}$ of boys is approximately normal with mean 0.5 and standard deviation

$$\sqrt{\frac{(0.5)(1-0.5)}{200}} = 0.0354$$

Hence,

$$P(\hat{p} \geq 0.55) = P\left(z \geq \frac{0.55-0.5}{0.0354} \right)$$

$$= P(\mathrm{z} \geq 1.41) = P(z \leq -1.41) = 0.0793$$

Example. Suppose that in a large group of people 25% are obsessed with taking selfies. Approximately what percent of samples of size 50 contain fewer than 20% of people who are obsessed with taking selfies?

Solution. Denote by $\hat{p}$ the proportion of people in a sample of size 50 who are obsessed with taking selfies.

By the CLT, we know that $\hat{p}$ is approximately normally distributed with mean 0.25 and standard deviation $\sqrt{\dfrac{(0.25)(1-0.25)}{50}} = 0.061.$

Thus, $$P(\hat{p} < 0.2) = P\left(z < \frac{0.2-0.25}{0.061}\right)$$

$$= P(z < -0.82) = 0.206 \quad \text{or} \quad 20.6\%.$$

8.1 Estimation

Definition. To **estimate** a population parameter means to assign a value (called an **estimate** or **estimator**) based on the information collected from a sample.

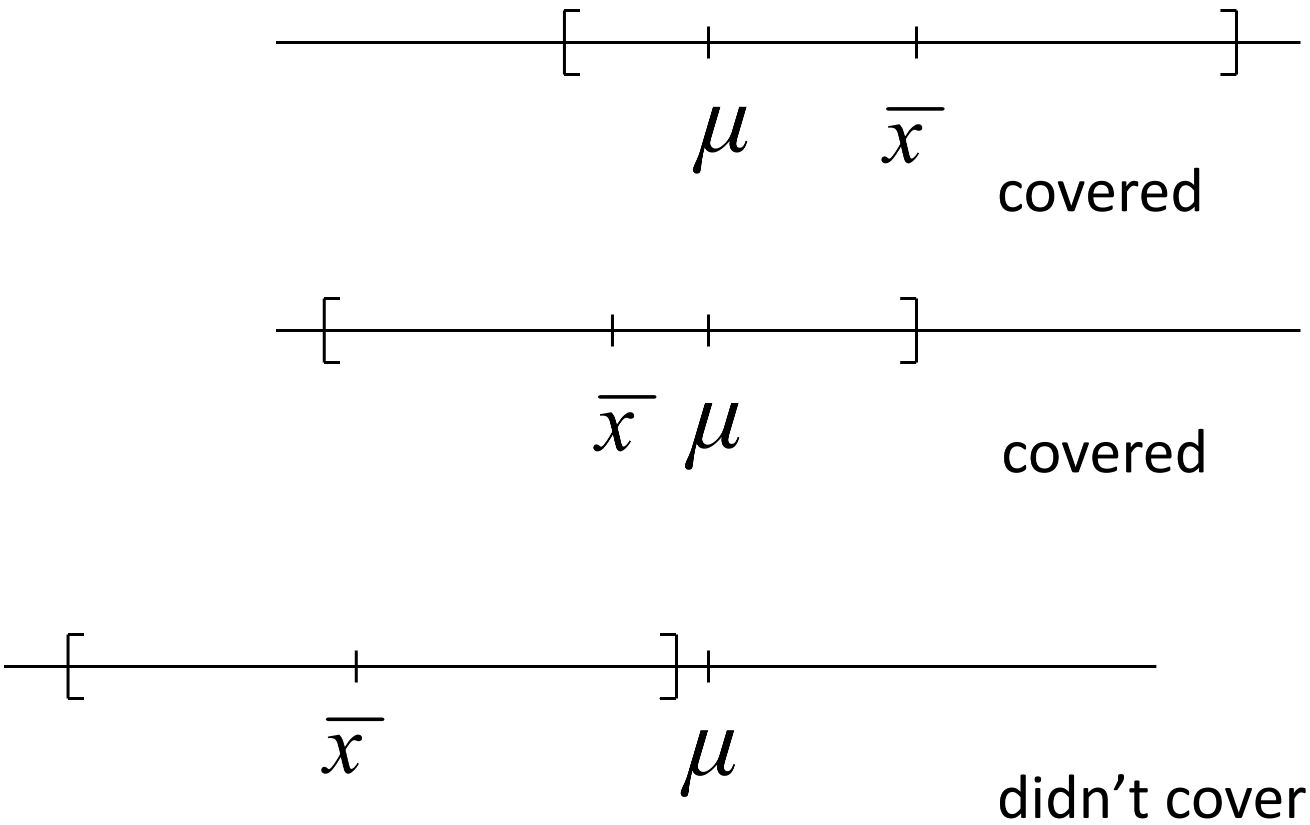
8.1 Point and Interval Estimators

Definition. When a population parameter is estimated by a value, this value is called a **point estimate** (or a **point estimator**).

Example. We estimate a population mean μ by the sample mean $\overline{x}$, so $\overline{x}$ is a point estimator of μ.

The drawbacks of point estimators are that their values are different for different samples, and surely none of them will be exactly equal to the population parameter. An alternative approach is to compute an interval estimator.

Definition. An **interval estimator** is an interval around the point estimator that is likely to contain the true population parameter.

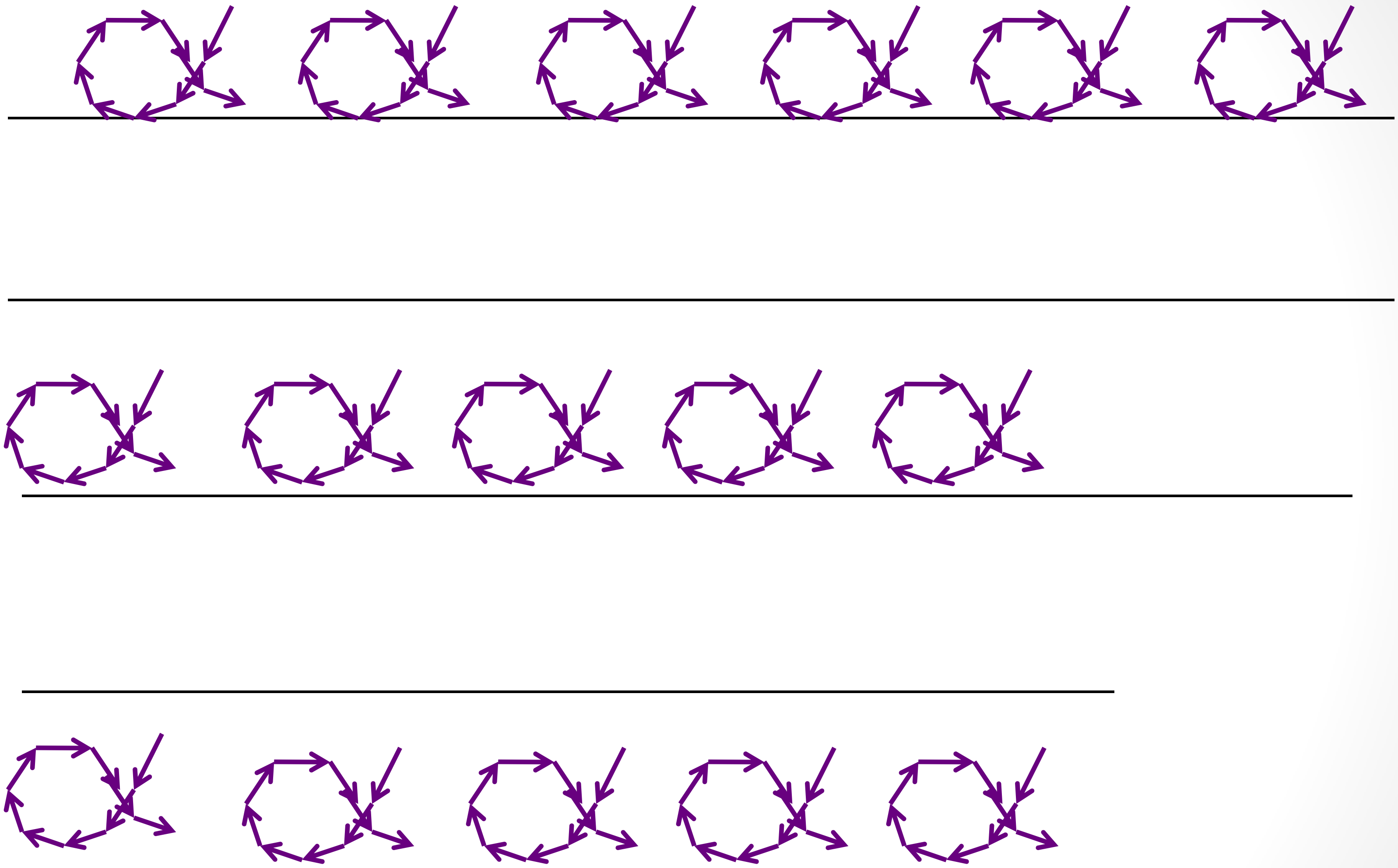


μ
x̄
covered
x̄
μ
covered
x̄
μ
didn't cover

Before we proceed, we have to practice writing the Greek letter

alpha

(pronounced /al-fuh/)

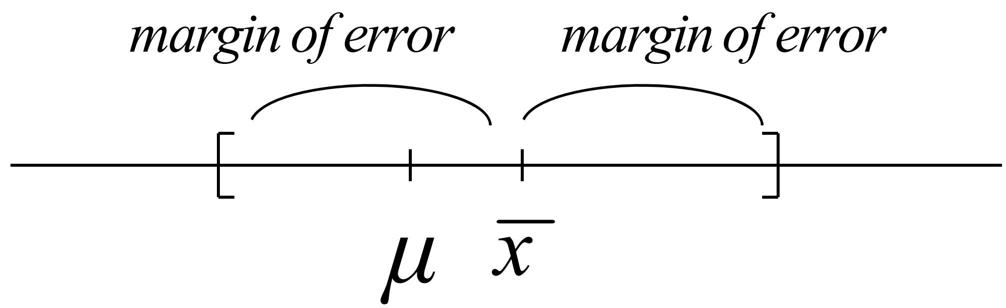


8.1 Confidence Interval

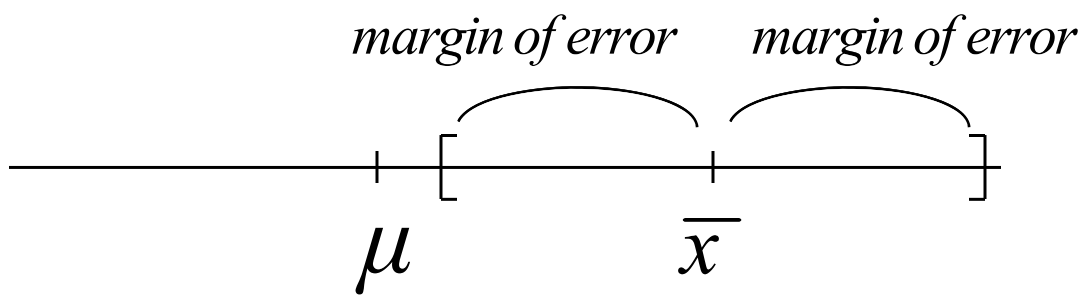
Definition. A $100 \cdot (1-\alpha)\%$ **confidence interval (CI)** for a population parameter has the form

point estimate +/- margin of error

This interval covers the true population parameter $100 \cdot (1-\alpha)\%$ of the time, and doesn't cover $100 \cdot \alpha\%$ of the time.



$$covers\ 100\cdot(1-\alpha)\%\ \ of\ \ the\ \ time$$



$$doesn't\ cover\ 100\cdot\alpha\%\ \ of\ \ the\ \ time$$

Note that:

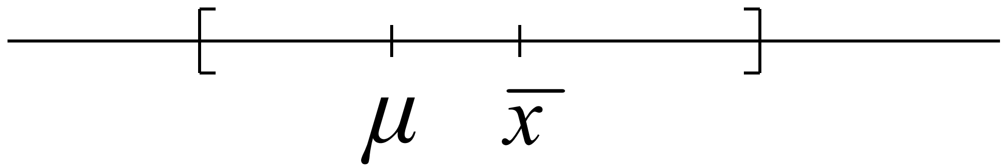
- A confidence interval is an interval estimator
- The center of a confidence interval is a point estimator
- The length of a confidence interval equals twice the margin of error

Definition. The quantity α is called the **confidence level**.

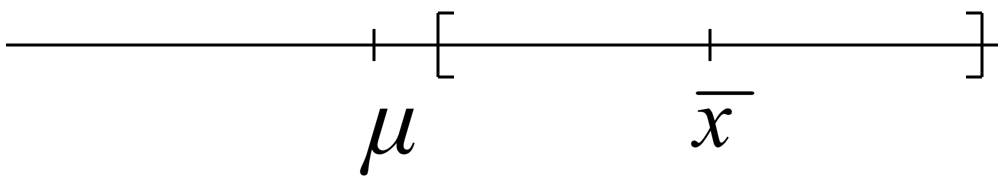
Typically three confidence levels are used: $\alpha = 0.1,\ 0.05,$ or 0.01, which correspond to 90%, 95%, or 99% confidence intervals.

α	$100 \cdot (1-\alpha)\%$
0.1	100·(1 - 0.1)% = 90%
0.05	100·(1 - 0.0.5)% = 95%
0.01	100·(1 - 0.01)% = 99%

95% Confidence Interval

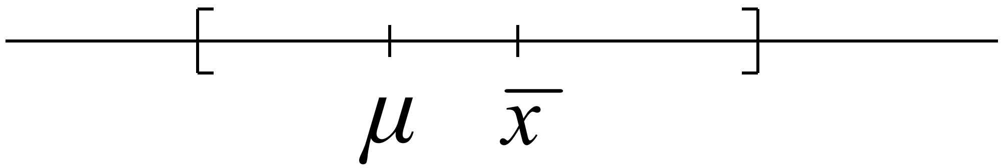


covers 95% *of the time*

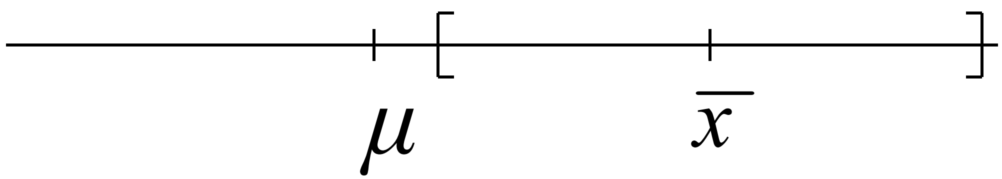


doesn't cover 5% *of the time*

99% Confidence Interval



covers 99% of the time



doesn't cover 1% of the time