```
In [1]:  ▶  1  from scipy import stats
            2  import numpy as np
            3  import pandas as pd
```

```
In [2]:  ▶  1  def ci_mu_known_sigma(data, sigma,confidence_level, side="double"):
            2      alpha = 1 - confidence_level
            3      z_dist = stats.norm(0,1)
            4      x_bar = np.mean(data)
            5      n = len(data)
            6      if side == "double":
            7          critical_value = np.round(z_dist.ppf(1 - (alpha/2)),2)
            8          interval = critical_value*sigma/np.sqrt(n)
            9          return (np.round(x_bar-interval,2), np.round(x_bar+interval,2)
           10      elif side == "low":
           11          critical_value = np.round(z_dist.ppf(1 - (alpha)),2)
           12          interval = critical_value*sigma/np.sqrt(n)
           13          return np.round(x_bar-interval,2)
           14      elif side == "up":
           15          critical_value = np.round(z_dist.ppf(1 - (alpha)),2)
           16          interval = critical_value*sigma/np.sqrt(n)
           17          return np.round(x_bar+interval,2)
```

1. A CI is desired for the true average stray-load loss μ (watts) for a certain type of induction motor when the line current is held at 10 amps for a speed of 1500 rpm. Assume that strayload loss is normally distributed with σ = 3.0.

a. Compute a 95% CI for μ when n = 25 and $\bar{x}$= 60.

```
In [3]:  ▶  1  sigma = 3
            2  n = 25
            3  x_bar = 60
            4  z_dist = stats.norm(0,1)
            5  critical_value = z_dist.ppf(.975) * sigma / np.sqrt(n)
            6  lower_band = x_bar - critical_value
            7  upper_band = x_bar + critical_value
            8  print(f"the 95% confidence interval is from {lower_band} to {upper_bar
```

```
the 95% confidence interval is from 58.82402160927597 to 61.1759783907240
3
```

b. Compute a 95% CI for μ when n = 100 and $\bar{x}$= 60.

In [4]:
```python
1  sigma = 3
2  n = 100
3  x_bar = 60
4  z_dist = stats.norm(0,1)
5  critical_value = z_dist.ppf(.975) * sigma / np.sqrt(n)
6  lower_band = x_bar - critical_value
7  upper_band = x_bar + critical_value
8  print(f"the 95% confidence interval is from {lower_band} to {upper_bar
```

the 95% confidence interval is from 59.41201080463799 to 60.5879891953620
1

c. Compute a 99% CI for μ when n = 100 and $\bar{x}$= 60.

In [5]:
```python
1  sigma = 3
2  n = 100
3  x_bar = 60
4  z_dist = stats.norm(0,1)
5  critical_value = z_dist.ppf(.995) * sigma / np.sqrt(n)
6  lower_band = x_bar - critical_value
7  upper_band = x_bar + critical_value
8  print(f"the 99% confidence interval is from {lower_band} to {upper_bar
```

the 99% confidence interval is from 59.22725120893533 to 60.7727487910646
7

d. Compute an 82% CI for μ when n = 100 and $\bar{x}$= 60.

In [6]:
```python
1  sigma = 3
2  n = 100
3  x_bar = 60
4  z_dist = stats.norm(0,1)
5  critical_value = z_dist.ppf(.91) * sigma / np.sqrt(n)
6  lower_band = x_bar - critical_value
7  upper_band = x_bar + critical_value
8  print(f"the 82% confidence interval is from {lower_band} to {upper_bar
```

the 82% confidence interval is from 59.59777348989294 to 60.4022265101070
6

e. What does this imply about the sample size and confidence level as pertaining to the length of the intervals? Justify your answer.

As the sample size increases, the standard error of the sample mean decreases. This means the distribution of the sample mean becomes more concentrated around the population mean.\\ A higher confidence level (e.g., 95% vs. 90%) means we want to be more sure that the population parameter lies within our interval. This results in a wider interval because we're accounting for more potential variability.

2. An engineer is doing a study in a manufacturing setting on the lengths of paperclips. The sample he collected is in the dataset paperclips.xlsx. He would like to create a 2-sided confidence interval for the true mean length. Assume that the distribution of paperclips is normal.

a. Estimate the mean, standard deviation, and standard error of the mean

In [8]: ▶
```
1  df = pd.read_excel("paperclips.xlsx")
2  x_bar = df['Length'].mean()
3  S = df['Length'].std()
4  n = len(df)
5  print(f"{x_bar = }")
6  print(f"{S = }")
7  print(f"{n = }")
```

```
x_bar = 5.367996000000001
S = 2.086490847308476
n = 50
```

Since the the distribution is normal we have

$$\mu = \bar{X}$$
$$\mu = 5.37$$
$$S = 2.09$$
$$n = 50$$
$$degree\ of\ freedom = 49$$
$$standard\ erorr\ of\ mean = \frac{S}{\sqrt{n}} = \frac{2.09}{\sqrt{50}} = 0.296$$

b. Find a two-sided 95% confidence interval for the mean and interpret

In [9]: ▶
```
1  stats.t.ppf(.975,49)
```

Out[9]: 2.009575234489209

$$\mu \in \bar{X} \pm t_{\alpha/2,\upsilon} * \frac{S}{\sqrt{n}}$$
$$t_{.025,49} = 2.01$$
$$\mu \in 5.37 \pm .296 * 2.01 = 5.37 \pm .595$$
$$\mu \in (4.775, 5.965)$$

c. Now the engineer decided that he is more concerned with the paperclips being too long and would like to know the average upper limit. Create the appropriate 95%

In [10]: ▶| 1   `stats.t.ppf(.95,49)`

Out[10]:   1.6765508919142629

$$\mu \leq \bar{X} + t_{\alpha,v} * \frac{S}{\sqrt{n}}$$
$$t_{.05,49} = 1.677$$
$$\mu \leq 5.37 + .296 * 1.677 = 5.37 + .496$$
$$\mu \leq 5.866$$

3. A Gallup poll was taken in which 1487 adults were surveyed and 43% of them said that they have a Facebook page.

a. Find the best point estimate of the proportion of all adults who have a Facebook page.

The best point estimate of the proportion of all adults who have a Facebook page is simply the sample proportion

$$\hat{P} = .43 = 43\%$$

b. Find the 90% confidence interval estimate of the population proportion p.

In [11]: ▶|
```
1 z_dist = stats.norm(0,1)
2 print(z_dist.ppf(.95))
3 print(np.sqrt(.43*(1-.43)/1487))
4 print(1.64*0.0128)
```

1.6448536269514722
0.012838555751569045
0.020992

$$\hat{P} \sim N(P, \sqrt{\frac{P(1-P)}{n}})$$
$$\alpha = .1$$
$$z_{\alpha/2} = z_{.05} = 1.64$$
$$\pi \sim \hat{p} \pm 1.64 * \sqrt{\frac{p*(1-p)}{n}} = 43 \pm 1.64 * 1.28 = 43 \pm 2.1$$

c. Based on the results, can we safely conclude that fewer than 50% of adults have Facebook pages?

The 90% confidence interval for the proportion of adults who have Facebook pages is (40.9, 45.1). Since the entire interval is below 0.50 (or 50%), we can conclude with 90% confidence that fewer than 50% of adults have Facebook pages.

So, based on the results, we can safely conclude that fewer than 50% of adults have Facebook pages at the 90% confidence level.

◄ ── ►

4. The sugar content of the syrup in canned peaches is normally distributed. A random sample of n=10 cans yields a sample standard deviation of s = 4.8 milligrams. Find a 95% two-side confidence interval for σ.

In [12]: ▶|
```
1  stats.chi2.ppf(.025,9)
```

Out[12]: 2.7003894999803584

$$(\sqrt{\frac{s^2 * (n-1)}{X^2_{v,\alpha/2}}}) \leq \sigma \leq (\sqrt{\frac{s^2 * (n-1)}{X^2_{v,1-\alpha/2}}})$$

$$\sqrt{\frac{207.36}{19.02}} \leq \sigma \leq \sqrt{\frac{207.36}{2.70}}$$

$$\sqrt{10.90} \leq \sigma \leq \sqrt{76.8}$$

$$3.30 \leq \sigma \leq 8.76$$