02_Normal_Distribution_Test

Continuous Data:

The data should be continuous, meaning it can take any value within a range.

Examples include height, weight, temperature, or time.

Interval or Ratio Scale:

The data should be on an interval or ratio scale.

Interval scales have meaningful differences between values but no true zero point (e.g., temperature in Celsius).

Ratio scales have both meaningful differences and a true zero point (e.g., weight).

import librares

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from scipy import stats
```

load the data

In [2]:
```
df=sns.load_dataset("titanic")
df.head()
```

Out[2]:

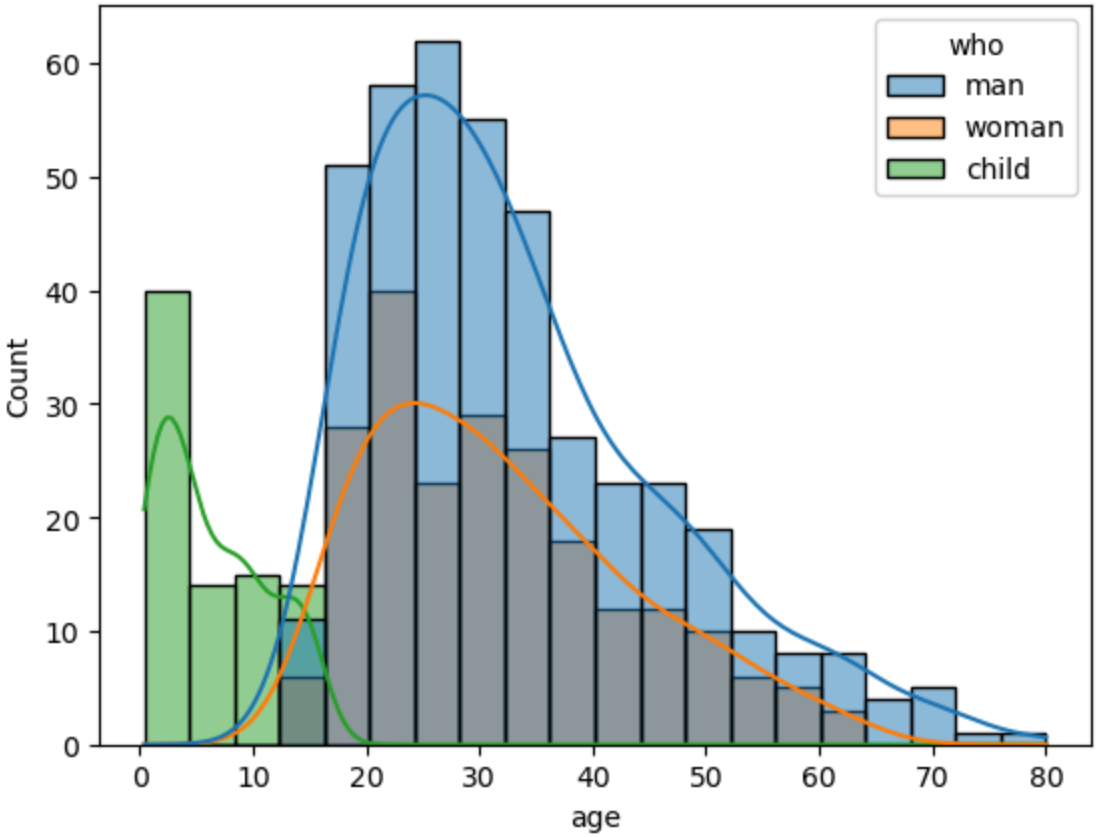
	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	Fals
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	Fals
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	Tru
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	Fals
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	Tru

Graphical representation

In [9]:
```
Graph=sns.histplot(data=df, x="age", hue="who", kde=True);
Graph
```

```
C:\Users\ustb\.anaconda\anwaar\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is dep
recated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\ustb\.anaconda\anwaar\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping with a length
-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` inste
ad of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
C:\Users\ustb\.anaconda\anwaar\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping with a length
-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` inste
ad of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
C:\Users\ustb\.anaconda\anwaar\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping with a length
-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` inste
ad of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
C:\Users\ustb\.anaconda\anwaar\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping with a length
-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` inste
ad of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
C:\Users\ustb\.anaconda\anwaar\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping with a length
-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` inste
ad of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
```

Out[9]:
```
<Axes: xlabel='age', ylabel='Count'>
```



Shapiro wilk test on normal distribution or Gaussion Distribution

Null Hypothesis: The data is Normaly Distributed

Alternative Hypothesis: the data is not Normaly Distributed

In [17]:
```
# Shapiro wilk test

statistics, p= stats.shapiro(df["age"])
#print the result using if condition
if p>0.05:
    print(f'p_value: {p}, sample looks Guassion/Normaly distributed (fail to reject H0)')
```

```
else:
    print(f'p_value: {p}, sample looks does not Guassion/Normaly distributed (reject H0)')
```

```
p_value: 1.0, sample looks Guassion/Normaly distributed (fail to reject H0)
```

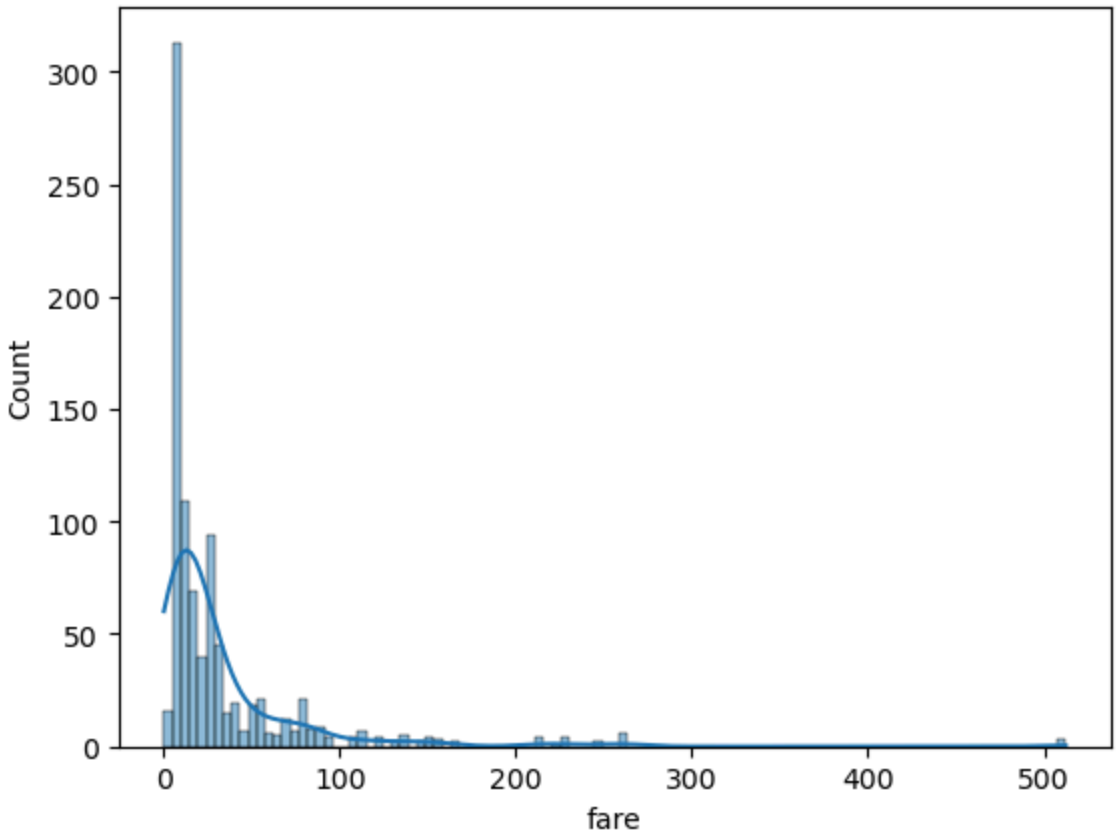
In [18]:
```
statistics, p= stats.shapiro(df["fare"])
#print the result using if condition
if p>0.05:
    print(f'p_value: {p}, sample looks Guassion/Normaly distributed (fail to reject H0)')
else:
    print(f'p_value: {p}, sample looks does not Guassion/Normaly distributed (reject H0)')
```

```
p_value: 1.0789998175301091e-43, sample looks does not Guassion/Normaly distributed (reject H0)
```

In [19]:
```
sns.histplot(data=df, x='fare', kde=True)
```

```
C:\Users\ustb\.anaconda\anwaar\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is dep
recated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

Out[19]:
```
<Axes: xlabel='fare', ylabel='Count'>
```



In [ ]: