ORCA 4500 : Foundations of Data Science (Homework - 3)

Importing Libraries

In [3]:

```
import pandas as pd
import numpy as np
import seaborn as sns
from scipy import stats
from scipy.stats import poisson
sns.set()
```

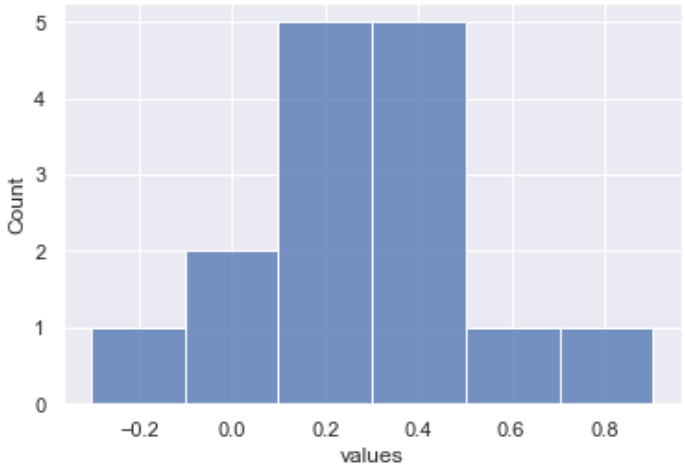
Problem 2

In [4]:

```
df = pd.read_csv('normal_samples.csv')
normal_dist_values = df['values']
sns.histplot(normal_dist_values)
```

Out[4]:

```
<AxesSubplot:xlabel='values', ylabel='Count'>
```



In [8]:

```
def CI_normal_dist(dist,alpha):
    n = len(normal_dist_values)
    mle_mue = np.mean(dist)
    mle_sigma = np.std(dist)
    t_alpha_2 = stats.t.ppf(1-alpha/2,n-1)
    lower_bound = np.mean(dist) - t_alpha_2*(np.std(dist,ddof=1))/np.sqrt(n)
    upper_bound = np.mean(dist) + t_alpha_2*(np.std(dist,ddof=1))/np.sqrt(n)
    return (lower_bound,upper_bound)

print("The t interval is : ",CI_normal_dist(normal_dist_values,0.10))
```

```
The t interval is :  (0.18001270248244733, 0.4236571314338406)
```

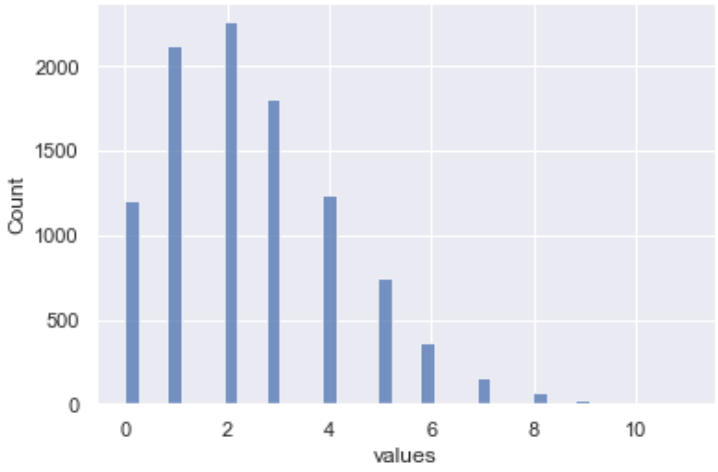
Problem 3

In [9]:

```
from scipy.optimize import minimize
df2 = pd.read_csv('accidents.csv')
dispersed_poisson_dist = df2['values']
sns.histplot(dispersed_poisson_dist)
```

Out[9]:

<AxesSubplot:xlabel='values', ylabel='Count'>



In [10]:

```
def dispersed_poisson_log_likelihood(X,lam1,lam2,p):
    likelihood = p*poisson.pmf(X,lam1) + (1-p)*poisson.pmf(X,lam2)
    return (-1)*np.sum(np.log(likelihood))

def bootstrap_CI(X,m,alpha):
    n = len(X)
    X_resample = np.random.choice(X, size=(m,n))
    mles = np.zeros((m,3))
    for i in range(0, m):
        f = lambda x: dispersed_poisson_log_likelihood(X_resample[i,:],lam1=x[0],lam2=x[
1],p=x[2])
        res = minimize(f, [1,2,0.5], bounds=((0, None), (0, None),(0,None)))
        mles[i,:] = res.x
    lower_bound = np.percentile(mles, alpha / 2,axis=0)
    upper_bound = np.percentile(mles, 100 - alpha/2,axis=0)
    return (lower_bound, upper_bound, mles)

lower,upper,mle = bootstrap_CI(dispersed_poisson_dist,m=100,alpha=5)
print("Maximum Likelihood Estimates for 100 samples :\n",mle.)
print("Bootstrap Confidence Interval for Lambda_1 :",(lower[0],upper[0]))
print("Bootstrap Confidence Interval for Lambda_2 :",(lower[1],upper[1]))
print("Bootstrap Confidence Interval for p :",(lower[2],upper[2]))
```

```
Maximum Likelihood Estimates for 100 samples :
 [[1.40735964 3.08056695 0.34327374]
 [1.00064816 2.87071809 0.22262764]
 [1.34519702 3.0959103  0.34689127]
 [1.4561242  3.15851917 0.40109322]
 [1.38219519 3.14134468 0.37486629]
 [1.39741564 3.17549457 0.3800703 ]
 [1.44155387 3.19035192 0.39934219]
 [1.21080451 3.05790928 0.31822843]
 [1.41065781 3.01758596 0.33697127]
 [1.18201818 3.02549958 0.29200149]
 [1.28029268 3.00983856 0.30889095]
 [1.40702557 3.08157248 0.36563579]
 [1.32296335 3.06679111 0.31883   ]
 [1.16701182 2.95094443 0.28546392]
 [1.20054618 2.98970518 0.29701583]
 [1.39368278 3.12333998 0.37431768]
 [1.36418738 3.10109037 0.36553093]
 [1.39369389 3.099132   0.35547072]
 [1.12046951 2.95863901 0.26087837]
 [1.24677    2.98833497 0.29205627]
 [1.43708913 3.18789569 0.3929575 ]
 [1.14677599 2.99189868 0.2728255 ]
 [1.32470592 3.0814262  0.3309701 ]
 [1.18989078 2.98594847 0.28286983]
 [1.37994631 3.07139464 0.35235836]
 [1.55081541 3.27874693 0.47672494]
```

```
[1.33332371 3.0384553  0.32639738]
[1.35392739 3.05843692 0.3353666 ]
[1.36983396 3.12999801 0.3666131 ]
[1.30706977 3.08485157 0.33893476]
[1.3152912  3.03894539 0.3225904 ]
[1.49253939 3.21905725 0.44340844]
[1.5163257  3.24076015 0.44719514]
[1.22324964 3.04196898 0.29964436]
[1.23190193 3.05332197 0.31015458]
[1.28236326 3.01957032 0.32636904]
[1.09938058 2.94637526 0.25017674]
[1.28435386 3.0748948  0.31237447]
[1.29539409 3.10511245 0.33065847]
[1.44499126 3.16506338 0.4060666 ]
[1.17913007 2.95935257 0.29010773]
[1.26954146 3.06071487 0.32332129]
[1.43797887 3.10062661 0.37820804]
[1.30161156 3.03234913 0.33011641]
[1.44925738 3.17914722 0.39617989]
[1.44169483 3.20332515 0.40883987]
[1.42162559 3.17822044 0.38450543]
[1.56396795 3.23829387 0.44871374]
[1.48862062 3.19354156 0.41552404]
[1.40491424 3.10192176 0.38074262]
[1.4147634  3.18353157 0.38294084]
[1.30989558 3.0949438  0.35542149]
[1.11354654 2.9762341  0.25218075]
[1.20594334 3.00047937 0.29716835]
[1.55381678 3.17257193 0.42988183]
[1.38222152 3.15469237 0.36857808]
[1.30693055 3.00604258 0.32395658]
[1.30930813 3.15872796 0.3696433 ]
[1.25804589 2.99353931 0.30281822]
[1.26137506 2.98880711 0.30288234]
[1.38745278 3.12899499 0.37391709]
[1.3082414  3.08849288 0.34045249]
[1.30598121 3.08261474 0.33800642]
[1.32470593 3.06781326 0.34152287]
[1.24550045 3.04736905 0.31548416]
[1.48919365 3.08249038 0.39759686]
[1.21986405 3.07118831 0.33078277]
[1.3645439  3.08533133 0.34898672]
[1.36995118 3.02552527 0.34037994]
[1.16143873 3.01393599 0.28863215]
[1.14073511 2.98002152 0.26174333]
[1.1464939  2.98763601 0.27751012]
[1.24213067 2.99357631 0.30550776]
[1.05099972 2.90368452 0.23633677]
[1.2391783  3.03613774 0.31259739]
[1.40812338 3.11233746 0.38031334]
[1.32188599 3.00826147 0.31532247]
[1.28113479 3.06004597 0.31274368]
[1.34156289 3.10492453 0.34588773]
[1.23790718 3.02962416 0.29944675]
[1.29379562 3.07160521 0.33355793]
[1.34796695 3.06698281 0.34134631]
[1.28624499 3.08126946 0.3428738 ]
[1.43754952 3.10110262 0.35484465]
[1.22774869 3.05212773 0.30389926]
[1.3107009  3.06174535 0.34153633]
[1.42621038 3.18910025 0.39866302]
[1.30643639 3.05134638 0.32141604]
[1.23890507 3.04432431 0.30110678]
[1.46770641 3.15674619 0.40007674]
[1.3634557  3.10428436 0.35442059]
[1.24226106 3.02954281 0.29505282]
[1.35246783 3.08400197 0.35858595]
[1.3114827  3.09829801 0.35510028]
[1.42488331 3.14230035 0.37556363]
[1.29435615 3.12709589 0.34816486]
[1.12722967 2.92120187 0.25106398]
[1.27394545 3.03470566 0.32015241]
```

```
 [1.47524752 3.11551171 0.40055257]
 [1.48313906 3.21150144 0.42456094]]
Bootstrap Confidence Interval for Lambda_1 : (1.1061094095045156, 1.5344327991627904)
Bootstrap Confidence Interval for Lambda_2 : (2.9331592295905446, 3.2291564749248645)
Bootstrap Confidence Interval for p : (0.2505981797144724, 0.4453964541600916)
```

In [ ]:

In [ ]: