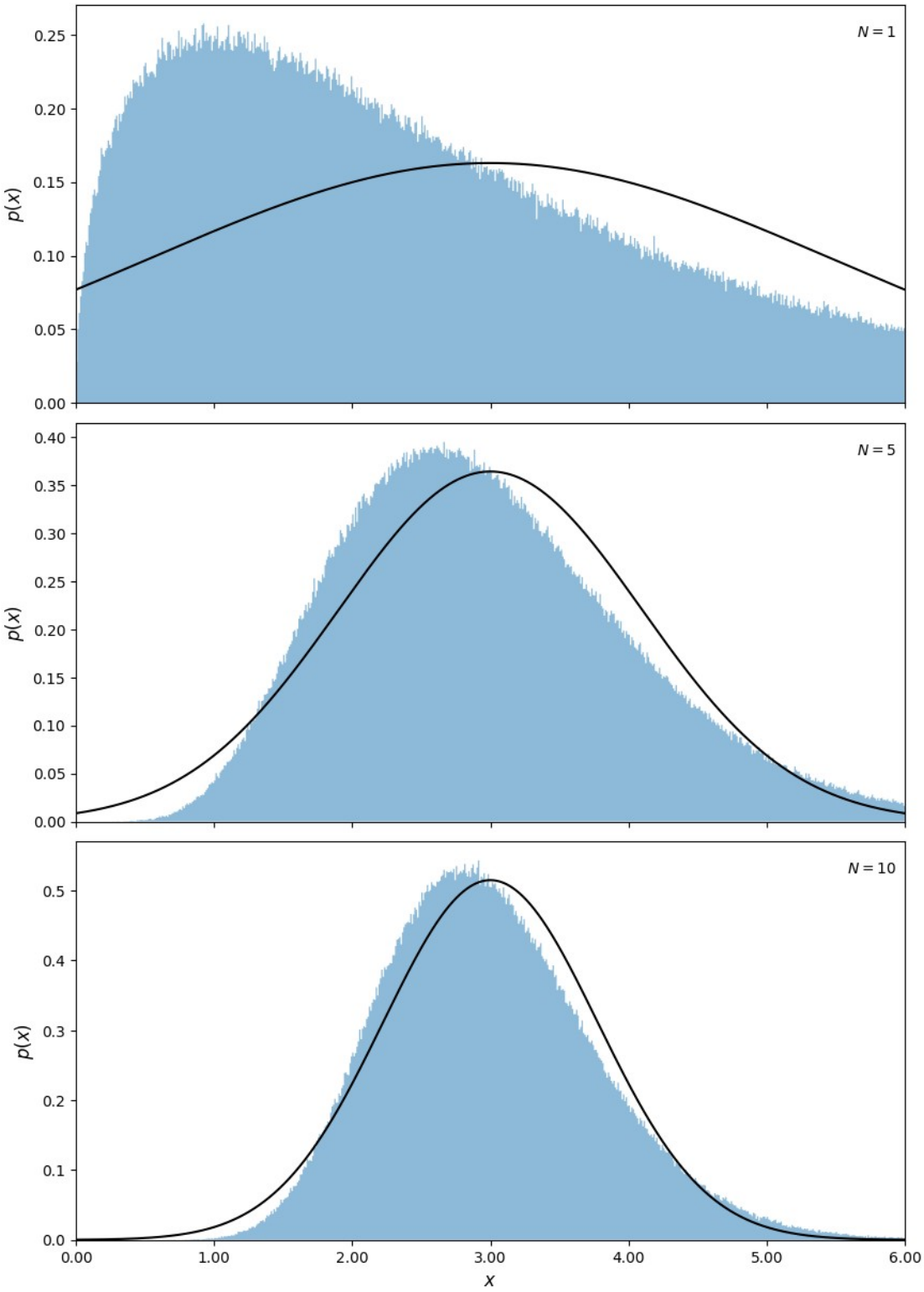
Assignment 2
EE20BTECH11047

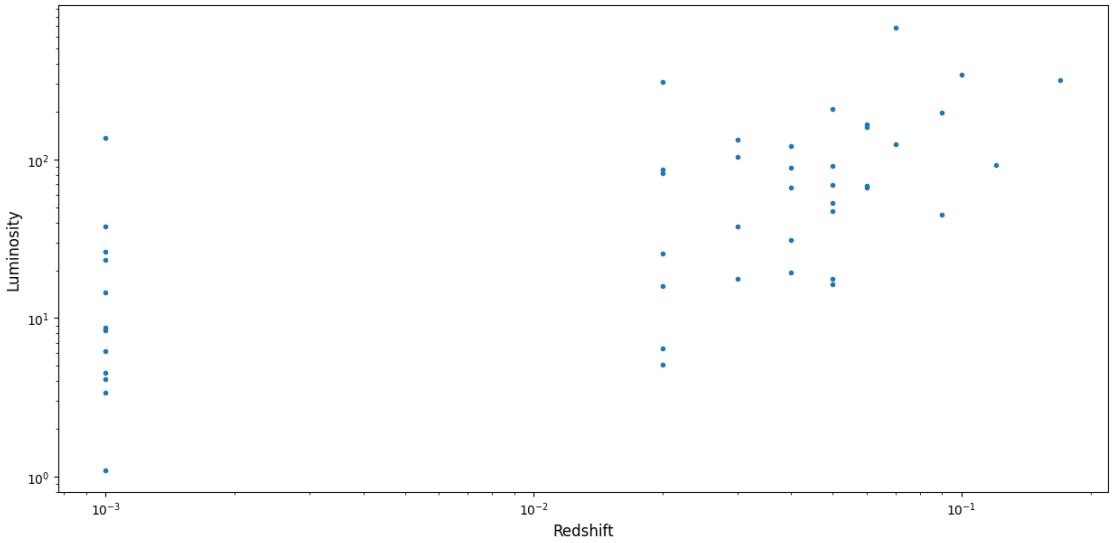
```
import numpy as np
import matplotlib.pyplot as plt
import scipy
from scipy import stats
from astroML import stats as asts
import pandas as pd

N = [1, 5, 10]
degree_of_freedom = 3
np.random.seed(42)
x = np.random.chisquare(degree_of_freedom,(max(N), int(1E6)))

fig = plt.figure(figsize=(10, 15))
fig.subplots_adjust(hspace=0.05)

for i in range(len(N)):
    # take the mean of the first N[i] samples
    x_i = x[:N[i], : ].mean(0)

    ax = fig.add_subplot(len(N),1,i+1)
    dist = stats.norm(degree_of_freedom, np.sqrt(2*degree_of_freedom /
N[i]))
    x_pdf = np.linspace(0, 10, 10000)
    ax.plot(x_pdf, dist.pdf(x_pdf), '-k')

    # histogram the data
    ax.hist(x_i, bins=np.linspace(0, 10, 1001),
            histtype='stepfilled', alpha=0.5, density=True)
    plt.xlim(0, 6)

    ax.text(0.99, 0.95, r"$N = %i$" % N[i],
            ha='right', va='top', transform=ax.transAxes)
    if i == len(N) - 1:
        ax.xaxis.set_major_formatter(plt.FormatStrFormatter('%.2f'))
        ax.set_xlabel(r'$x$', fontsize = 12)
    else:
        ax.xaxis.set_major_formatter(plt.NullFormatter())

    ax.set_ylabel('$p(x)$', fontsize = 12)


plt.show()
```



Question 2

```
dataset =pd.read_csv('http://www.iith.ac.in/~shantanud/test.dat' , sep
=' ')
```

```
X = dataset['#Lx']
Y = dataset['z']
plt.figure(figsize=(15,7))
plt.loglog(Y, X, '.')
plt.xlabel("Redshift", fontsize = 12)
plt.ylabel('Luminosity', fontsize = 12)
plt.show()
```




```
Spearman,p1 = stats.spearmanr(X,Y)
pearson,p2 = stats.pearsonr(X,Y)
kendall,p3 = stats.kendalltau(X,Y)
```

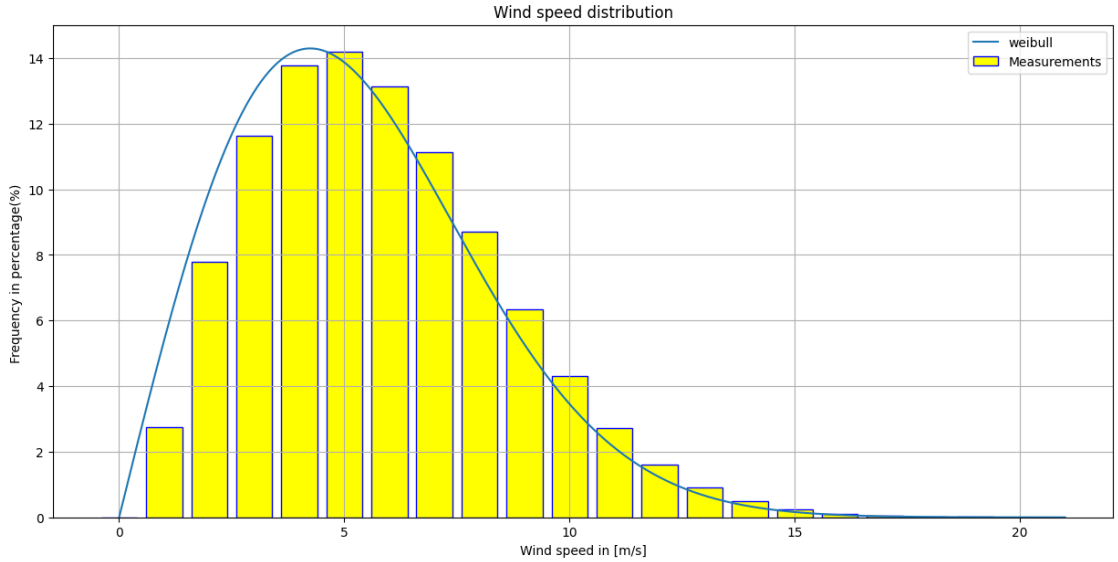
```
print(f'Spearman correlation coefficient is: {Spearman}, p-value is:
{p1}\n'
      f'Pearson correlation coefficient is : {pearson}, p-value: {p2}\
n'
      f'Kendall-tau correlation coefficient: {kendall}, p-value:
{p3}')
```

```
Spearman correlation coefficient is: 0.6596325957535455, p-value is:
6.166489759080966e-07
Pearson correlation coefficient is : 0.5144497852670242, p-value:
0.00025464716576124353
Kendall-tau correlation coefficient: 0.5029584682704178, p-value:
2.9696862274734036e-06
```

Question 3

```
percentage= [0, 2.75, 7.80, 11.64, 13.79, 14.20, 13.15, 11.14, 8.72,
6.34, 4.30, 2.73,
             1.62, 0.91, 0.48, 0.24, 0.11, 0.05, 0.02, 0.01, 0.00]
K,A = 2,6   #A is lamda
x = np.linspace(0, 21, 1000)
dist = scipy.stats.weibull_min(K, 0, A)
```

```
weibull = dist.pdf(x)*100

x1 = np.arange(0, 21)
plt.figure(figsize=(15,7))
plt.bar(x1, percentage ,color='yellow',edgecolor='blue', label =
"Measurements")
plt.plot(x, weibull ,label ="weibull" )
plt.xlabel('Wind speed in [m/s]')
plt.ylabel('Frequency in percentage(%) ')
plt.title('Wind speed distribution')
plt.legend()
plt.grid()
plt.show()
```



Question 4

```
mu , sigma, n = 0, 1, 1000
normal = stats.norm(mu, sigma)
dist1 = normal.rvs(n)
dist2 = normal.rvs(n)

corr_coe,p = stats.pearsonr(dist1,dist2)
t = corr_coe*np.sqrt((n - 2)/(1 - np.square(corr_coe)))
if t >= 0:
    p2= 2*(1 - stats.t.cdf(t, n - 2))
else:
    p2= 2*stats.t.cdf(t, n - 2)
print("Pearson Correlation Coefficient is : ",corr_coe)
print("P using Pearson  : ",p)
print("t value of students distribution is : ",t)
print("t value of students distribution is : ",p2)

Pearson Correlation Coefficient is :  0.03524337574216485
P using Pearson  :  0.26551719222730125
```

```
t value of students distribution is :  1.1140704525301182
t value of students distribution is :  0.2655171922273034
```

```
print("Here both the P values are almost same ")
```

```
Here both the P values are almost same
```