Week 2 Assignment

PART I: Probability Distributions

Question 1

1. It is given that the discrete random variable X satisfies

X ~ Binom (n, p)

Given further that P (X = 2) = P (X = 3), show that

E(X) = 3 − p

P (X=k) = $\binom{n}{k}$ p^k $(1-p)^{n-k}$

P(X=2) = $\binom{n}{2} p^2(1-p)^{n-2}$

P(X=3) = $\binom{n}{3} p^3(1-p)^{n-3}$

Since P (X = 2) = P (X = 3) then:

$\binom{n}{2} p^2(1-p)^{n-2} = \binom{n}{3} p^3(1-p)^{n-3}$

$= \frac{n!}{(n-2)!2!} p^2(1-p)^{n-2} = \frac{n!}{(n-3)!3!} p^3(1-p)^{n-3}$

$= \frac{n(n-1)}{2} p^2(1-p)^{n-2} = \frac{n(n-1)(n-2)}{2.3} p^3(1-p)^{n-3}$

Divide both side by $p^2(1-p)^{n-2}$

$= \frac{p^2(1-p)^{n-2}}{p^2(1-p)^{n-3}} = \frac{(n-2)}{3} \frac{p^3(1-p)^{n-3}}{p^2(1-p)^{n-3}}$

$= \frac{p^2(1-p)^{(n-2)-(n-3)}}{p^2} = \frac{(n-2)}{3} p$

$= \frac{p^2(1-p)^{(n-n-2+3)}}{p^2} = \frac{(n-2)}{3} p$

= (1-p) = $\frac{(n-2)}{3} p$

= 31n-p) = p(n-2)

= 3 – 3p = np – 2p

= 3 = np – 2p + 3p

= 3 = p(n-2+3)

= 3 = p(n+1)

p = 3/(n+1)

Note that E(X) = np, and p = 3/(n+1)

E(X) = n. $\frac{3}{n+1}$

To show that E(X) = 3−p, substitute p with $\frac{3}{n+1}$

= 3 - $\frac{3}{n+1}$

= $\frac{3(n+1)-3}{n+1}$

= $\frac{3n+3-3}{n+1}$ = $\frac{3n}{n+1}$ which is also equal to np i.e. E(X).

Therefore: E(X)

Question 2

2. A discrete random variable X has Poisson distribution where λ is a parameter.

Find an expression for the following in terms of λ;

(a) E(X)

(b) E(X2)

(a)

Since $\sum_{x=0}^{n} x\mathbb{p}(\mathrm{X} = x),\ \ where\ \ \mathbb{p}(\mathrm{X} = x) = \frac{e^{\lambda}\lambda^{x}}{x!}$

E(X) = $x\frac{e^{\lambda}\lambda^{x}}{x!}$

Rewrite $\frac{\lambda^{x}}{x!}$ as $\frac{\lambda\lambda^{x-1}}{x(x-1)!}$

E(X) = $x\frac{e^{\lambda}\lambda\lambda^{x-1}}{x(x-1)!}$

E(X) = $\frac{e^{\lambda}\lambda\lambda^{x-1}}{(x-1)!}$

If k=x−1, x=k+1

= $e^{\lambda}\lambda \sum_{x=0}^{n} \frac{\lambda^{k}}{k!}$

= $\sum_{x=0}^{n} \frac{\lambda^{k}}{k!}$ is the Taylor series expansion of e$^{\lambda}$

=E(X) = $e^{\lambda}\lambda e^{\lambda}$

= E(X) = $e^{2\lambda}\lambda$

Therefore, E(X) = $e^{2\lambda}\lambda$

(b)

Since $E(X^2)\sum_{x=0}^{n} x^2 \mathbb{p}(\mathrm{X} = x),\ \ where\ \mathbb{p}(\mathrm{X} = x) = \frac{e^{\lambda}\lambda^{x}}{x!}$

E(X^2) = $x^2\frac{e^{\lambda}\lambda^{x}}{x!}$

Rewrite x^2 as *(x(x−1)+x)*

E(X^2) = $e^{\lambda}(x(x-1)+x)\frac{\lambda^{x}}{x!}$

E(X^2) = $e^{\lambda}(x(x-1)\frac{\lambda^{x}}{x!} + x\frac{\lambda^{x}}{x!})$

=from (a), $e^{\lambda}.\, x\frac{\lambda^{x}}{x!}$ = $e^{2\lambda}\lambda$

=$e^{\lambda}(x(x-1)\frac{\lambda^{x}}{x!}$ = $e^{\lambda}\frac{\lambda^{2}\lambda^{x-2}}{(x-2)!}$

If k = x-2,

E(X^2) = $e^{\lambda}(\frac{\lambda^{2}\lambda^{k}}{k!})$ +

=$e^{\lambda}\lambda^{2}e^{\lambda}$ + $e^{2\lambda}\lambda$

Therefore, E(X^2) = $e^{\lambda}\lambda^{2}e^{\lambda}\ +\ e^{2\lambda}\lambda$

= $e^{2\lambda}\lambda^{2}+\ e^{2\lambda}\lambda$

=$e^{2\lambda}\lambda(\lambda+1)$

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv('/content/sleep.csv')
```

```
df.head()
```

	id	totSAS	sex	age	edlevel	fitrate	depress
0	83	10.0	female	42.0	secondary school	7.0	1.0
1	294	20.0	female	54.0	postgraduate degree	7.0	2.0
2	425	31.0	male	NaN	secondary school	5.0	10.0
3	64	34.0	female	41.0	postgraduate degree	7.0	3.0
4	536	25.0	female	39.0	postgraduate degree	5.0	0.0

Next steps: Generate code with df View recommended plots

1. Identify the data type of totSAS, sex, age and edlevel variable

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 271 entries, 0 to 270
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   id        271 non-null    int64
 1   totSAS    251 non-null    float64
 2   sex       271 non-null    object
 3   age       248 non-null    float64
 4   edlevel   269 non-null    object
 5   fitrate   266 non-null    float64
 6   depress   269 non-null    float64
dtypes: float64(4), int64(1), object(2)
memory usage: 14.9+ KB
```

```
print("The datatype of totSAS is:", df['totSAS'].dtype)
print("The datatype of sex is:", df['sex'].dtype)
print("The datatype of age is:", df['age'].dtype)
print("The datatype of edlevel is:", df['edlevel'].dtype)
```

```
The datatype of totSAS is: float64
The datatype of sex is: object
The datatype of age is: float64
The datatype of edlevel is: object
```

Answer

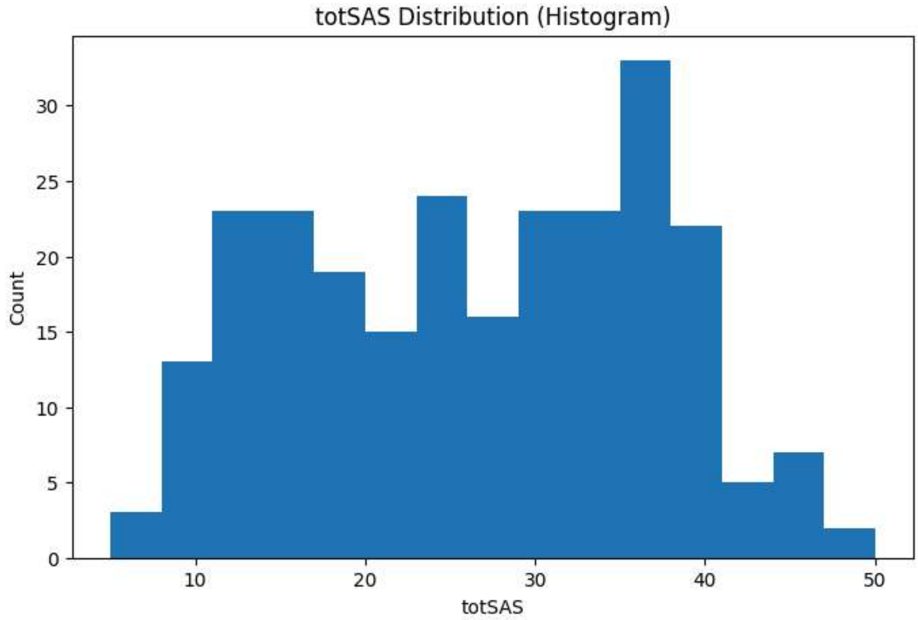
- totSAS has a datatype "float64"
- sex is of datatype "object"
- age is of datatype "float64" and
- edlevel is of datatype "object".

2. Which visualization method will you use for each variable

- 'totSAS': Histogram
- 'sex': Pie Chart, Bar Chart
- 'age': Histogram
- 'edlevel': Pie Chart, Bar Chart
- 'fitrate': Histogram
- 'depress': Histogram

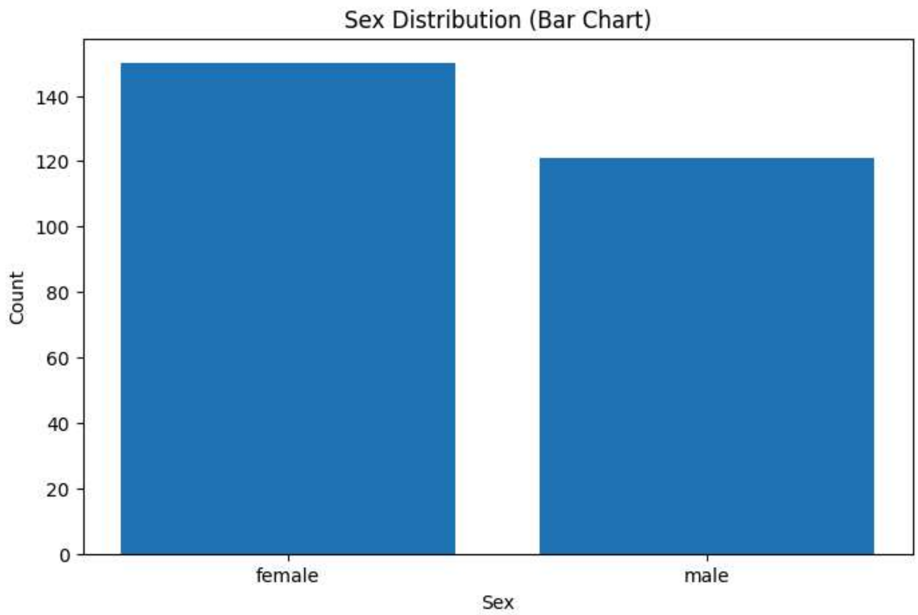
```
# histogram for totSAS
plt.figure(figsize=(8, 5))
plt.hist(df.totSAS, bins = 15)
plt.title('totSAS Distribution (Histogram)')
plt.xlabel('totSAS')
plt.ylabel('Count');

print('The totSAS Histogram is Negatively Skewed')
```

```
The totSAS Histogram is Negatively Skewed
```



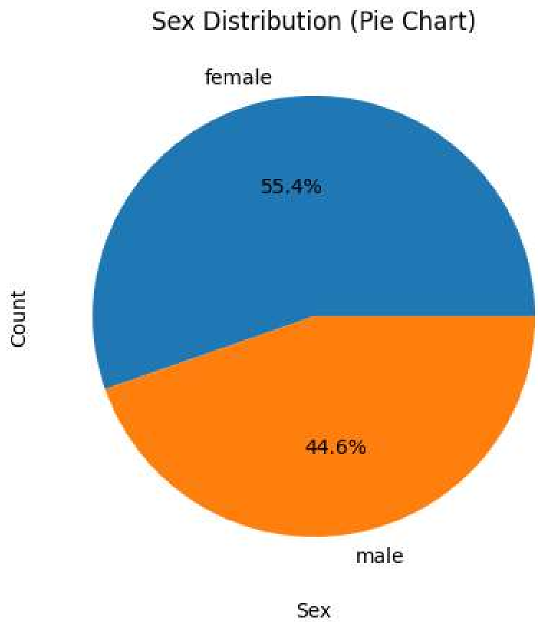
The totSAS Histogram is Negatively Skewed

```
# Pie Chart, Bar Chart for sex
plt.figure(figsize=(8, 5))
counts = df['sex'].value_counts()
plt.bar(counts.index, counts.values)
plt.title('Sex Distribution (Bar Chart)')
plt.xlabel('Sex')
plt.ylabel('Count');
```




```
#bar chart for Sex
plt.figure(figsize=(8, 5))
plt.pie(counts.values, labels=counts.index, autopct='%1.1f%%')
plt.title('Sex Distribution (Pie Chart)')
plt.xlabel('Sex')
plt.ylabel('Count')
```

```
Text(0, 0.5, 'Count')
```

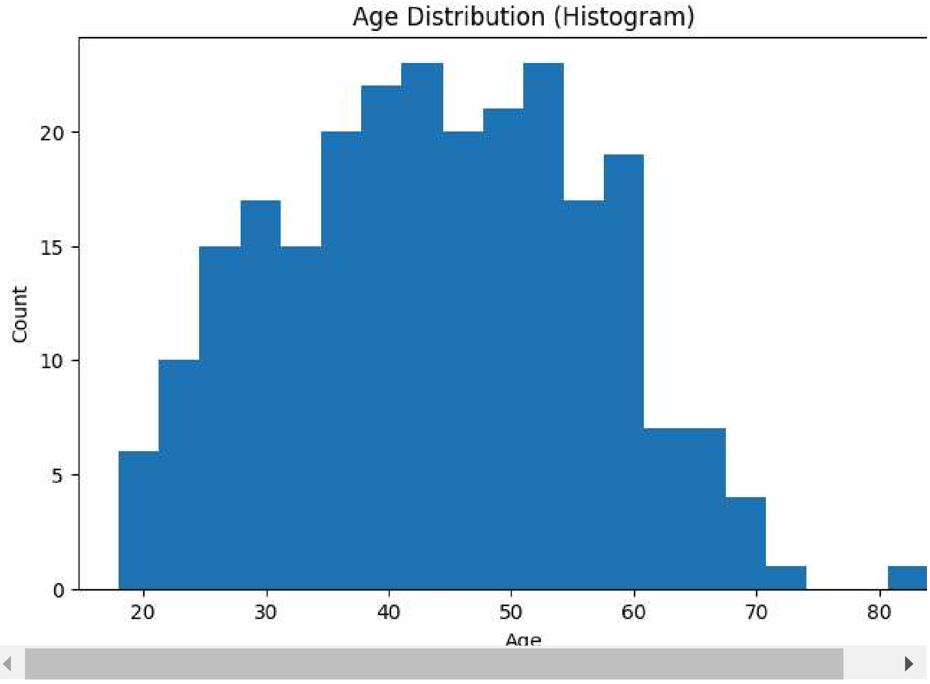



```
df.sex.value_counts()
```

```
sex
female    150
male      121
Name: count, dtype: int64
```

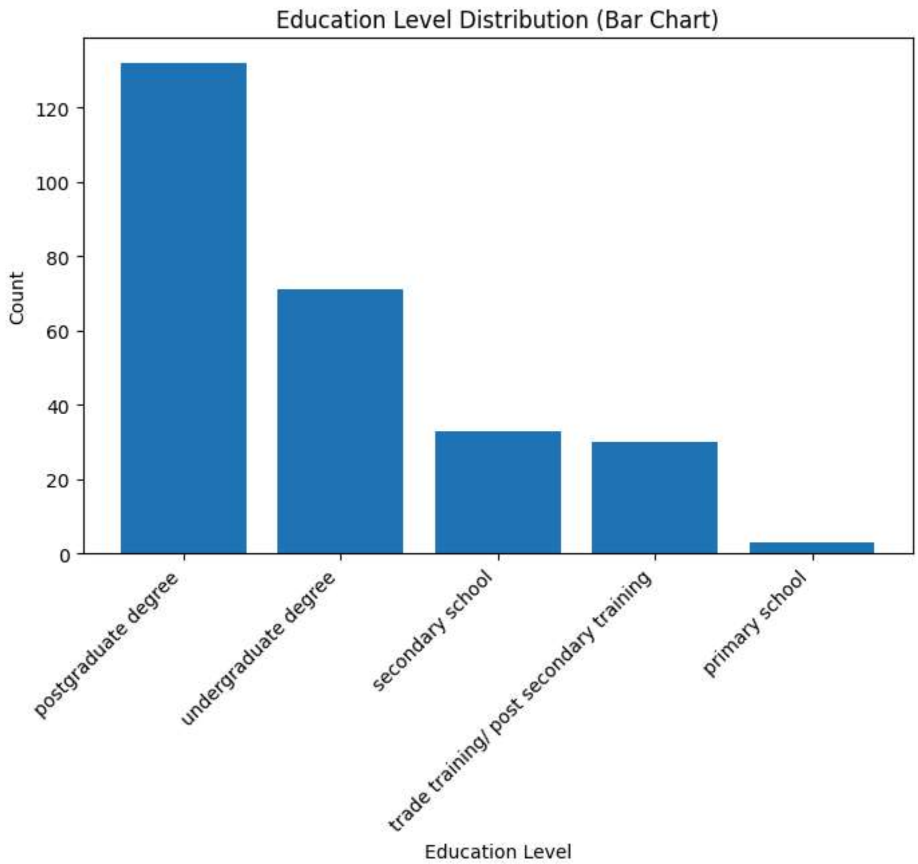
```
# histogram for age
plt.figure(figsize=(8, 5))
plt.hist(df.age, bins=20)
plt.title('Age Distribution (Histogram)')
plt.xlabel('Age')
plt.ylabel('Count');

print('The Age Histogram is Negatively Skewed')
```

```
The Age Histogram is Negatively Skewed
```

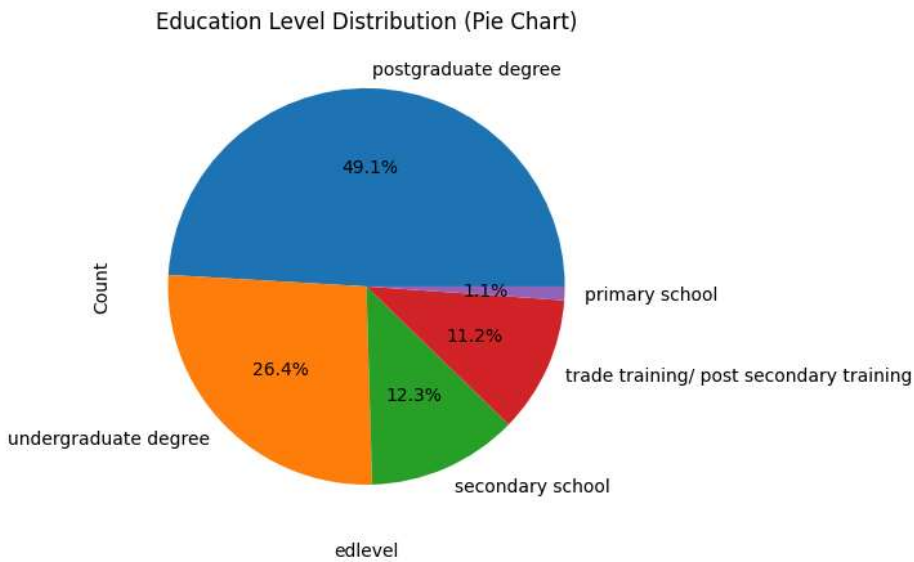


The Age Histogram is Negatively Skewed

```
edlevel_counts = df['edlevel'].value_counts()

# Plot bar chart for edlevel
plt.figure(figsize=(8, 5))
plt.bar(edlevel_counts.index, edlevel_counts.values)
plt.title('Education Level Distribution (Bar Chart)')
plt.xlabel('Education Level')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right');
```

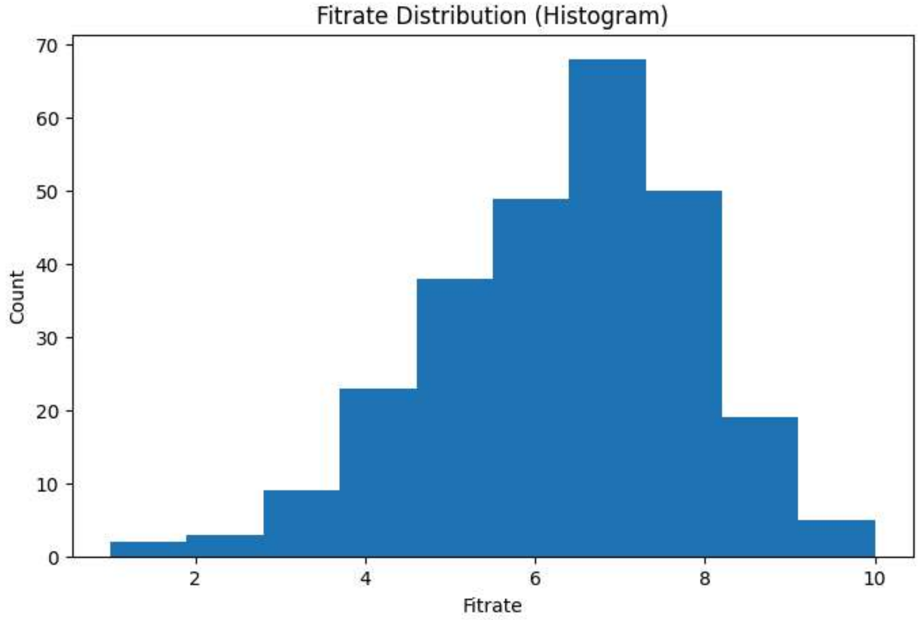



```
plt.figure(figsize=(8, 5))
plt.pie(edlevel_counts.values, labels=edlevel_counts.index, autopct='%1.1f%%')
plt.title('Education Level Distribution (Pie Chart)')
plt.xlabel('edlevel')
plt.ylabel('Count');
```




```
# histogram for fitrate
plt.figure(figsize=(8, 5))
plt.hist(df.fitrate)
plt.title('Fitrate Distribution (Histogram)')
plt.xlabel('Fitrate')
plt.ylabel('Count');

print('The Filtrate Histogram is Negatively Skewed')
```

```
The Filtrate Histogram is Negatively Skewed
```



The Filtrate Histogram is Negatively Skewed

```
# histogram for depress
```