San Francisco Bay University

MATH208 - Probability and Statistics
2023 Fall Homework #4

Due day: 12/7/2023

Kritika Regmi
ID:19702

1. Assuming that the lengths of American anchovies appease the normal distribution with the mean $\mu = 10.3cm$ and standard deviation $\sigma = 0.65cm$, please find the percentages of the lengths in the population of American anchovies.

   a. Less than *9cm*.

```
#Question no: 1a
from scipy.stats import norm
len_average = 10.3
std_dev = 0.65
z_score = (9 - len_average) / std_dev
percentage= norm.cdf(z_score) * 100
print(f"The percentage of anchovies having lengths less than 9 cm is: {percentage:.2f}%")
```

```
The percentage of anchovies having lengths less than 9 cm is: 2.28%
```

   b. Between *9.5cm* and *10.6cm.*

```
# Question no: 1b
zscore_9_5cm = (9.5 - len_average) / std_dev
zscore_10_6cm = (10.6 - len_average) / std_dev
percentage= (norm.cdf(zscore_10_6cm) - norm.cdf(zscore_9_5cm)) * 100
print(f"The percentage of anchovies having lengths between 9.5 cm and 10.6 cm {percentage:.2f}%")
```

```
The percentage of anchovies having lengths between 9.5 cm and 10.6 cm 56.86%
```

   c. What is the minimum length if a restaurant claimed that the lengths of the sold anchovies are in the top of 20%?

```
#Question no: 1c
zscore_top20 = norm.ppf(0.8)
min_len_top20 = len_average + zscore_top20 * std_dev
print(f"The minimum length for the top 20% {min_len_top20:.2f} cm")
```

```
The minimum length for the top 20% 10.85 cm
```

2. If the random variables X and Y are normal distributions with $\mu = 10$ & $\sigma = 3$ and $\mu = 15$ & $\sigma = 8$, namely, $X \sim N(10, 3)$ and $Y \sim N(15, 8)$, and they are independent, what is the probability distribution and statistical parameters of:

ANSWER:

(1) $X + Y$
$\mu x + \mu y$ =10+15
Mean = 25
Variance (σ^2_{x+y}) = $3^2 + 8^2$ = 9+ 64 = 73
So, Probability distribution = N(25, $\sqrt{73}$)

(2) $X - Y$
$\mu x - \mu y$ =10 - 15
Mean = -5
Variance (σ^2_{x+y}) = $3^2 + 8^2$ = 9+ 64 = 73
So, Probability distribution = N(-5, $\sqrt{73}$)

(3) $3X$
$\mu 3x = 3.\mu x$
Mean = 3*10 = 30
Variance($\sigma^2_{3x}) = (3.\sigma x)^2$ = $(3.3)^2$ = 81

So, Probability distribution = N(30,9)

(4) $4X + 5Y$
$\mu 4x + \mu 5y = 4.\mu x + 5.\mu y$
Mean = 4*10 + 5*15= 40 + 75 = 115
Variance($\sigma^2_{4x+5y}) = (4.\sigma x)^2 + (5.\sigma y)^2$ = $(4.3)^2$ + $(5.8)^2$= 344

So, Probability distribution = N(115, $\sqrt{344}$)

3. For the students in Engineering School, please write Python program to verify the mean $\mu = np$ and standard deviation $\sigma = \sqrt{npq}$ for p=*0.05* and selecting any n greater than *50* in binomial distribution.

ANSWER:

```python
#Question no:3
import numpy as np
from scipy.stats import binom
N = 100
P = 0.05
average= N * P
std_dev = np.sqrt(N * P* (1 - P))
binom_distribution = binom(N, P)
actual_avg = binom_distribution.mean()
actual_std = binom_distribution.std()
print(f"Mean: {average}")
print(f"Actual Mean: {actual_avg:.2f}")
print(f"Standard Deviation: {std_dev}")
print(f"Actual Standard Deviation: {actual_std:.2f}")
```

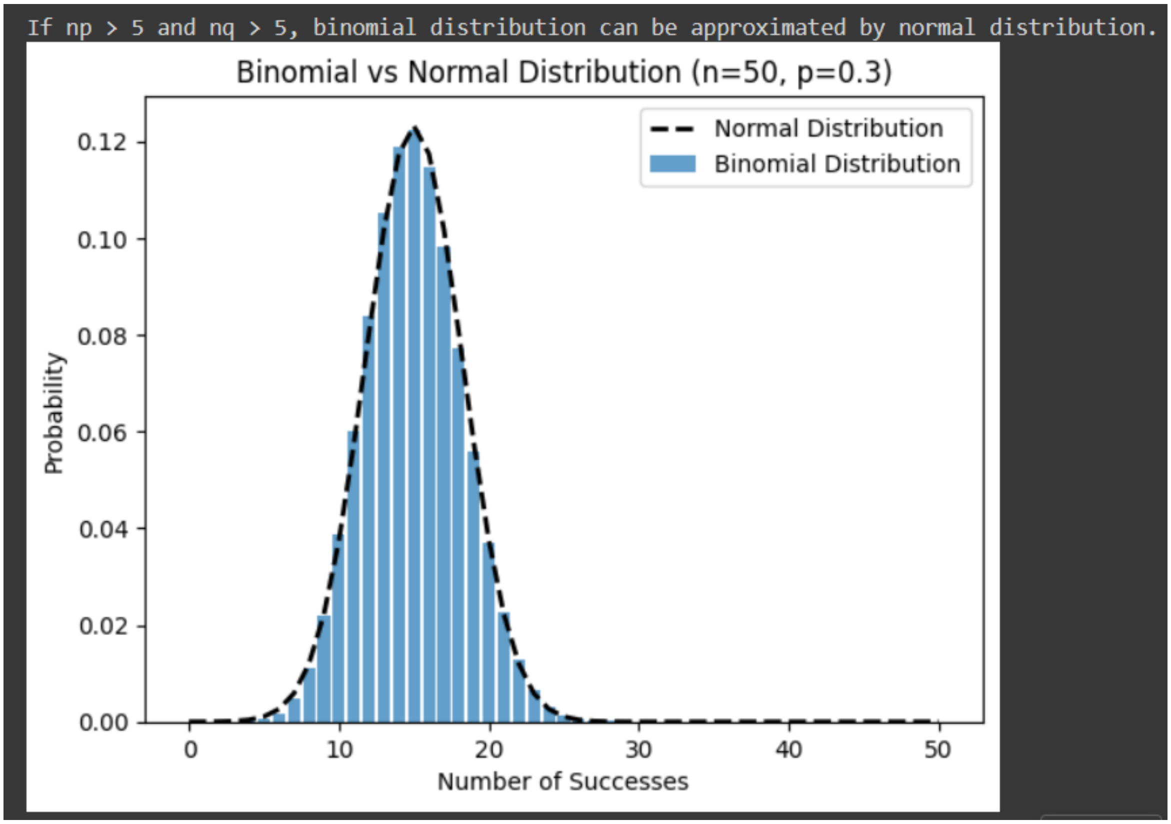
```
Mean: 5.0
Actual Mean: 5.00
Standard Deviation: 2.179449471770337
Actual Standard Deviation: 2.18
```

4. In general, if $np > 5$ and $nq > 5$ in binomial distribution, binomial probabilities can be approximated using the normal distribution. Please select any big enough n and p's values to verify in Python program or excel and plot the histogram.

ANSWER:

```python
#Question no:4
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import binom, norm
N = 50
P = 0.3
avg = N * P
std = np.sqrt(N * P * (1 - P))
if N * P > 5 and N * (1 - P) > 5:
    print("If np > 5 and nq > 5, binomial distribution can be approximated by normal distribution.")
else:
    print("If condition np > 5 and nq > 5 , binomial distribution cannot be approximated correctly by normal distribution.")
b_exmp = binom.rvs(N, P, size=1000)
x_vals = np.arange(0, N + 1)
b_probs = binom.pmf(x_vals, N, P)
n_probs = norm.pdf(x_vals, avg, std)
plt.bar(x_vals, b_probs, label="Binomial Distribution", alpha=0.7)
plt.plot(x_vals, n_probs, label="Normal Distribution", color='black', linestyle='dashed', linewidth=2)

plt.title(f'Binomial vs Normal Distribution (n={N}, p={P})')
plt.xlabel('Number of Successes')
plt.ylabel('Probability')
plt.legend()
plt.show()
```



5. In coin tossing experiments, please find the probability of the exact 6 heads from 12 tossing by ONLY using the normal distribution method.

ANSWER:

We know,

$\mu = np$

$\sigma = \sqrt{npq}$

Here,

n=12, p=6

Using normal distribution to find the probability, we get:

P(X =k) = $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

Substituting the values, we get:

$\mu = 12 * 0.5 = 6$

$\sigma = \sqrt{12 \times 0.5 \times 0 \cdot 5} = \sqrt{3}$

Now, using formula,

P(X= 6) = $\frac{1}{\sqrt{3}\sqrt{2\pi}} \exp\left(-\frac{(6-6)^2}{2\times 3}\right) = \frac{1}{\sqrt{6\pi}}$

Calculating, we get:
$\frac{1}{\sqrt{6\pi}}$ ~ 0.1508

Therefore, the probability of getting exactly 6 heads in 12 coins toss is 0.1508.

6. Given that the defective rate of a product of the batteries in a manufacturing company is 6%, 150 batteries are randomly selected from the population. Please find the probability of 12 or more defective ones in them by ONLY using the normal distribution method.

ANSWER:
Here, P =6%= 0.06, N = 150
So, Mean $(\mu) = NP = 150 * 0.06$ = 9

Calculating q, we get:
q = 1- p = 1- 0.06 = 0.94

Calculating standard deviation, we get:
Standard Deviation $(\sigma) = \sqrt{npq}$
$= \sqrt{150 \times 0.06 \times 0 \cdot 94}$
$= \sqrt{8.46}$ ~ 2.91

Calculating the z – score, we get:
Z = $\frac{X-\mu}{\sigma} = \frac{12-9}{2 \cdot 91}$ ~ 1.03

Using standard normal distribution:
P $(X \geq 12)$ = 1 – P (X <12) = 1 – 0.8485 ~ 0.1515 or 15.15%

The probability of having 12 or more defective batteries is 15.15%.

7. For the students in Engineering School, please write a Python program by calling functions in the following link to create *100* random numbers in T distribution with *df =10* (degree of freedom) and calculate the mean μ and standard deviation σ. After that, the *30* samples will be randomly selected from these random numbers in each sampling group. A total of 15 sampling groups should be created. Based on Central Limit Theorem (CLT), the mean value $\bar{\bar{x}}$ in total 15 sampling group is roughly the mean μ of *100* random numbers and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. Please verify it and plot the histogram, which should be normal distribution. For Business school students, complete the above process in Excel.

ANSWER:

```
#Question No:7
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import t

dfr=10
random_num=t.rvs(dfr, size=100)
mean1 = np.mean(random_num)
std_dev1 = np.std(random_num)
group=15
sample_each_group=30

avg=np.zeros(group)

for i in range(group):
  random_sample=np.random.choice(random_num, size=sample_each_group, replace=False)
  avg[i]=np.mean(random_sample)

average=np.mean(avg)
standard_deviation=np.std(avg, ddof=1)

print(f"First  Distribution: Mean  = {mean1:.2f}, Standard Deviation = {std_dev1:.2f}")
print(f"Sampling Groups: Mean = {average:.2f}, Standard Deviation  = {standard_deviation:.2f}")
plt.hist(avg, bins=15, edgecolor='black', alpha=0.7)
plt.title('Sampling Groups Mean Histogram')
plt.xlabel('Sample Mean')
plt.ylabel('Frequency')
plt.show()
```

```
First  Distribution: Mean  = 0.09, Standard Deviation = 1.02
Sampling Groups: Mean = 0.07, Standard Deviation  = 0.13
```