Continuous Probability Distribution and Confidence Interval

Please ensure you update all the details:
Name: ulli Venkata sai kumar Batch ID: 04072024HYD10AM
Topic: Continuous Probability Distribution and Confidence Interval

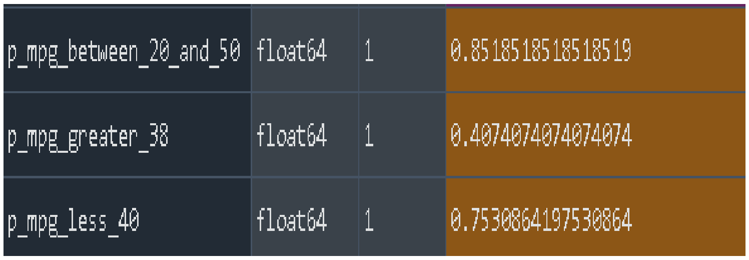
Q1) Calculate probability from the given dataset for the below cases.

Data_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

a. P(MPG>38)
b. P(MPG<40)
c. P (20<MPG<50)



p_mpg_between_20_and_50	float64	1	0.8518518518518519
p_mpg_greater_38	float64	1	0.4074074074074074
p_mpg_less_40	float64	1	0.7530864197530864

```
import pandas as pd
# Load the Cars dataset
cars_data = pd.read_csv('Cars.csv')
mpg_data = cars_data['MPG']
p_mpg_greater_38 = (mpg_data > 38).mean()
p_mpg_less_40 = (mpg_data < 40).mean()
p_mpg_between_20_and_50 = ((mpg_data > 20) & (mpg_data < 50)).mean()
print("P(MPG > 38):", p_mpg_greater_38)
print("P(MPG < 40):", p_mpg_less_40)
print("P(20 < MPG < 50):", p_mpg_between_20_and_50)
```

Q2) Check whether the data follows the normal distribution.

a) Check whether the MPG of Cars follows the Normal Distribution Dataset: Cars.csv

```
import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as stats

# Load the dataset
cars_data = pd.read_csv('Cars.csv')
mpg_data = cars_data['MPG']
# Histogram
plt.hist(mpg_data, bins=15, edgecolor='k', alpha=0.7)
plt.title('Histogram of MPG')
plt.xlabel('MPG')
plt.ylabel('Frequency')
plt.show()
```

```
# Q-Q Plot
stats.probplot(mpg_data, dist="norm", plot=plt)
plt.title('Q-Q Plot of MPG')
plt.show()
# Shapiro-Wilk Test
shapiro_test = stats.shapiro(mpg_data)
print("Shapiro-Wilk Test:", shapiro_test)
# Kolmogorov-Smirnov Test (against normal distribution)
ks_test = stats.kstest(mpg_data, 'norm',
    args=(mpg_data.mean(), mpg_data.std()))
    print("Kolmogorov-Smirnov Test:", ks_test)
```

b) Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follow Normal Distribution

Dataset: wc-at.csv

```
import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as stats

# Load the dataset
wc_at_data = pd.read_csv('wc-at.csv')
at_data = wc_at_data['AT']  # Adipose Tissue
waist_data = wc_at_data['Waist']  # Waist Circumference
# Histogram for AT
plt.hist(at_data, bins=15, edgecolor='k', alpha=0.7)
plt.title('Histogram of Adipose Tissue (AT)')
plt.xlabel('AT')
plt.ylabel('Frequency')
plt.show()

# Q-Q Plot for AT
stats.probplot(at_data, dist="norm", plot=plt)
plt.title('Q-Q Plot of Adipose Tissue (AT)')
plt.show()

# Histogram for Waist
plt.hist(waist_data, bins=15, edgecolor='k', alpha=0.7)
plt.title('Histogram of Waist Circumference')
plt.xlabel('Waist')
plt.ylabel('Frequency')
plt.show()

# Q-Q Plot for Waist
stats.probplot(waist_data, dist="norm", plot=plt)
plt.title('Q-Q Plot of Waist Circumference')
plt.show()
# Shapiro-Wilk Test for AT
shapiro_test_at = stats.shapiro(at_data)
print("Shapiro-Wilk Test for AT:", shapiro_test_at)
```

```
# Shapiro-Wilk Test for Waist
shapiro_test_waist = stats.shapiro(waist_data)
print("Shapiro-Wilk Test for Waist:", shapiro_test_waist)

# Kolmogorov-Smirnov Test for AT
ks_test_at = stats.kstest(at_data, 'norm', args=(at_data.mean(), at_data.std()))
print("Kolmogorov-Smirnov Test for AT:", ks_test_at)

# Kolmogorov-Smirnov Test for Waist
ks_test_waist = stats.kstest(waist_data, 'norm', args=(waist_data.mean(),
waist_data.std()))
print("Kolmogorov-Smirnov Test for Waist:", ks_test_waist)
```

Q3) Calculate the Z scores of 90% confidence interval,94% confidence interval, and 60% confidence interval.

➢ 90% Confidence Interval:
- Area in each tail = 1−0.90/2 = 0.05
- Z-score for a 90% confidence interval is approximately 1.645.

➢ 94% Confidence Interval:
- Area in each tail = 1-0.94/2 = 0.03
- Z-score for a 94% confidence interval is approximately 1.88.

➢ 60% Confidence Interval:
- Area in each tail = 1-0.60/2 = 0.20
- Z-score for a 60% confidence interval is approximately 0.84.

Q4) Calculate the t scores of 95% confidence interval, 96% confidence interval, and 99% confidence interval for the sample size of 25.

To calculate the t-scores for different confidence intervals with a sample size of 25, you need to use the t-distribution. Given a sample size of n=25n = 25n=25, the degrees of freedom (df) is n−1=24n - 1 = 24n−1=24.

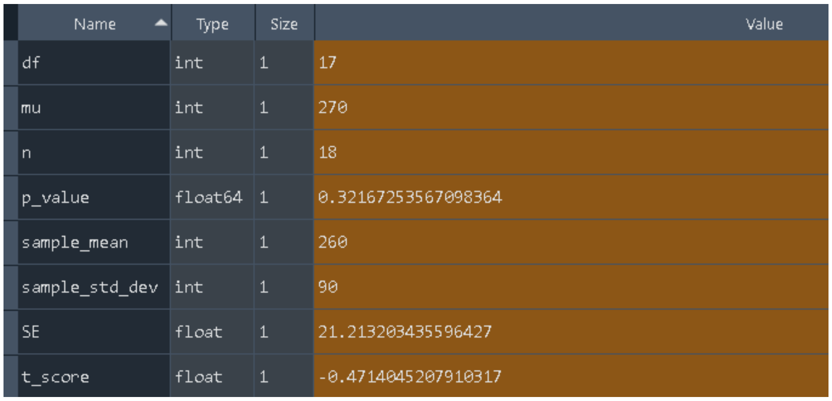
The critical t-value tα/2t_{\alpha/2}tα/2 corresponds to the area in the tails of the distribution not covered by the confidence level.

Step-by-Step Calculation:

1. 95% Confidence Interval:
    - Area in each tail = 1-0.95/2 = 0.025
    - Degrees of freedom = 24
    - t-score for a 95% confidence interval: approximately 2.064.
2. 96% Confidence Interval:
    - Area in each tail = 1-0.96/2 = 0.02
    - Degrees of freedom = 24

- t-score for a 96% confidence interval: approximately 2.172.

3. 99% Confidence Interval:
   - Area in each tail = 1-0.99/2=0.005
   - Degrees of freedom = 24
   - t-score for a 99% confidence interval: approximately 2.797.



Name	Type	Size	Value
df	int	1	17
mu	int	1	270
n	int	1	18
p_value	float64	1	0.32167253567098364
sample_mean	int	1	260
sample_std_dev	int	1	90
SE	float	1	21.213203435596427
t_score	float	1	-0.4714045207910317

Q5) A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days?

```
from scipy.stats import t
import math

# Given values
mu = 270
sample_mean = 260
sample_std_dev = 90
n = 18

# Step 1: Calculate Standard Error
SE = sample_std_dev / math.sqrt(n)

# Step 2: Calculate t-score
t_score = (sample_mean - mu) / SE

# Step 3: Calculate the probability
# Degrees of freedom
df = n - 1
# Probability (P-value) for t-score
p_value = t.cdf(t_score, df)

print("t-score:", t_score)
print("Probability that the sample mean is 260 days or less:", p_value)
```

Q6) The time required for servicing transmissions is normally distributed between □ = 45 minutes and □ = 8 minutes. The service manager plans to have work begin on the transmission of a customer’s car 10 minutes after the car is dropped off and the customer is told that the car will be ready within 1 hour from drop-off. What is the probability that the service manager cannot meet his commitment?

A. 0.3875
B. 0.2676
C. 0.5
D. 0.6987
ANS;- D. 0.6987

```
from scipy.stats import norm

# Given values
mu = 45
sigma = 8
x = 50

# Calculate the Z-score
z = (x - mu) / sigma

# Calculate the probability
p = 1 - norm.cdf(z)

print("Probability that the service manager cannot meet his commitment:", p)
```

Q7) The current age (in years) of 400 clerical employees at an insurance claims processing center is normally distributed with mean m = 38 and Standard deviation
s =6. For each statement below, please specify True/False. If false, briefly explain why.

A. More employees at the processing center are older than 44 than between 38 and 44.
Since 34.13% (employees aged between 38 and 44) is greater than 15.87% (employees older than 44), Statement A is False.

B. A training program for employees under the age of 30 at the center would be expected to attract about 36 employees.
Since 36.72 is close to 36, Statement B is True.

Q8) If X1 ~ N(μ, σ2) and X2 ~ N(μ, σ2) are iid normal random variables, then what is the difference between 2 X1 and X1 + X2? Discuss both their distributions and parameters.
So, 2X1−(X1+X2) follows a normal distribution with:

- Mean = 0
- Variance = 2σ2

Q9) Let X ~ N(100, 20^2) its (100, 20 square). Find two values, a and b, symmetric about the mean, such that the probability of the random variable taking a value between them is 0.99.

A. 90.5, 105.9
B. 80.2, 119.8
C. 22,78
D. 48.5, 151.5
E. 90.1, 109.9

1. Find the Z-score corresponding to the middle 99% of a standard normal distribution:
   - Since we want 0.99 probability in the middle, 0.005 (or 0.5%) will be in each tail.
   - Using a Z-table or standard normal distribution calculator, the Z-score that leaves 0.5% in each tail is approximately 2.576.
2. Translate the Z-score back to the original distribution:
   - Since X is normally distributed with mean 100 and standard deviation 20, we use the formula:

     $a=\mu-Z\cdot\sigma=100-(2.576\cdot 20)$

     $b=\mu+Z\cdot\sigma=100+(2.576\cdot 20)$
3. Calculate a and b:

   $a=100-(2.576\times 20)=100-51.52=48.48$

   $b=100+(2.576\times 20)=100+51.52=151.52$

   Thus, the interval $(a,b)\approx(48.5,151.5)$ provides a 0.99 probability that XXX will take a value within this range.

   Answer:

   The correct option is: D. 48.5, 151.5

Q10) Consider a company that has two different divisions. The annual profits from the two divisions are independent and have distributions Profit1 ~ N(5, 3^2) and Profit2 ~ N(7, 4^2) respectively. Both the profits are in $ Million. Answer the following questions about the total profit of the company in Rupees. Assume that $1 = Rs. 45

A.Specify a Rupee range (centered on the mean) such that it contains 95% probability for the annual profit of the company.

For a 95% confidence interval centered around the mean, we need to find the range within 1.96 standard deviations (since 95% of a normal distribution lies within ±1.96 standard deviations of the mean).

1. Lower bound: $\text{Lower bound}=\mu\text{rupees}-1.96\cdot\sigma\text{rupees}=540-1.96\times 225$
2. Upper bound: $\text{Upper bound}=\mu\text{rupees}+1.96\cdot\sigma\text{rupees}=540+1.96\times 225$

Calculating these values:

$\text{Lower bound}=540-441=99$

$\text{Upper bound}=540+441=981$

Answer for Part A: The 95% confidence range for the annual profit in rupees is 99 million to 981 million rupees.

B.Specify the 5th percentile of profit (in Rupees) for the company.

Part B: Specify the 5th percentile of profit (in Rupees) for the company.

The 5th percentile corresponds to a Z-score of approximately -1.645.

1. 5th percentile in dollars:

$\text{5th percentile in dollars}=\mu\text{total}+(-1.645)\times\sigma\text{total}=12+(-1.645)\times 5=12-8.225=3.775\text{ million dollars}$

2. Convert to rupees:

5th percentile in rupees$=3.775\times45=169.875\approx170$ million rupees
Answer for Part B: The 5th percentile of profit in rupees is approximately 170 million rupees.

C.Which of the two divisions has a larger probability of making a loss each year?
Part C: Which of the two divisions has a larger probability of making a loss each year?
To find the probability of each division making a loss, we calculate the probability that each profit is less than zero.

1. Profit1:
   - Mean, $\mu1=5$
   - Standard deviation, $\sigma1=3$
   - Z-score for a loss (Profit1 $<$ 0): $Z=0-5/3=-5/3\approx-1.67$
   - Probability P (Profit1$<$0)$=P(Z<-1.67)\approx0.0475$ or 4.75%.
2. Profit2:
   - Mean, $\mu2=7$
   - Standard deviation, $\sigma2=4$
   - Z-score for a loss (Profit2 $<$ 0): $Z=0-7/4=-7/4=-1.75$
   - Probability P(Profit2$<$0)$=P(Z<-1.75)\approx0.0401$or 4.01%.

Answer for Part C: Profit1 has a slightly larger probability of making a loss each year, with approximately 4.75% compared to Profit2’s 4.01%.