Q1. Calculate the 95% confidence interval for a sample of data with a mean of 50 and a standard deviation of 5 using Python. Interpret the results.

In [ ]:
```
import scipy.stats as stats

# Sample data
sample_mean = 50
sample_std = 5
sample_size = your_sample_size  # Replace with your sample size

# Calculate the 95% confidence interval
confidence_interval = stats.norm.interval(0.95, loc=sample_mean, scale=sample_st

print("95% Confidence Interval:", confidence_interval)
```

Q2. Conduct a chi-square goodness of fit test to determine if the distribution of colors of M&Ms in a bag matches the expected distribution of 20% blue, 20% orange, 20% green, 10% yellow, 10% red, and 20% brown. Use Python to perform the test with a significance level of 0.05.

In [ ]:
```
import scipy.stats as stats
from scipy.stats import chisquare

# Observed frequencies of M&M colors
observed = [your_observed_frequencies]

# Expected frequencies
expected = [0.20, 0.20, 0.20, 0.10, 0.10, 0.20] * sum(observed)

# Perform the chi-square goodness of fit test
chi2, p = chisquare(f_obs=observed, f_exp=expected)

alpha = 0.05  # Significance level

if p < alpha:
    print("Reject the null hypothesis: The distribution of M&M colors does not m
else:
    print("Fail to reject the null hypothesis: The distribution of M&M colors ma
```

Q3. Use Python to calculate the chi-square statistic and p-value for a contingency table with the following data: A B Outcome 1 20 15 Outcome 2 10 25 Outcome 3 15 20 Interpret the results of the test.

In [ ]:
```
import scipy.stats as stats

# Create a contingency table (replace values with your data)
observed = [[your_observed_data]]
chi2, p, _, _ = stats.chi2_contingency(observed)

alpha = 0.05  # Significance level

if p < alpha:
    print("Reject the null hypothesis: There is a significant association.")
else:
    print("Fail to reject the null hypothesis: There is no significant associati
```

Q4. A study of the prevalence of smoking in a population of 500 individuals found that 60 individuals smoked. Use Python to calculate the 95% confidence interval for the true proportion of individuals in the population who smoke.

In [4]:
```
import scipy.stats as stats

sample_size = 500  # Replace with your sample size
sample_proportion = 60 / sample_size  # Replace with your sample data

# Calculate the 95% confidence interval
confidence_interval = stats.norm.interval(0.95, loc=sample_proportion, scale=(((

print("95% Confidence Interval:", confidence_interval)
```

```
95% Confidence Interval: (0.09151638887130707, 0.1484836111286929)
```

Q5. Calculate the 90% confidence interval for a sample of data with a mean of 75 and a standard deviation of 12 using Python. Interpret the results.

In [ ]:
```
import scipy.stats as stats

# Sample data
sample_mean = 75
sample_std = 12
sample_size = your_sample_size  # Replace with your sample size

# Calculate the 90% confidence interval
confidence_interval = stats.norm.interval(0.90, loc=sample_mean, scale=sample_st

print("90% Confidence Interval:", confidence_interval)
```

Q6. Use Python to plot the chi-square distribution with 10 degrees of freedom. Label the axes and shade the area corresponding to a chi-square statistic of 15.

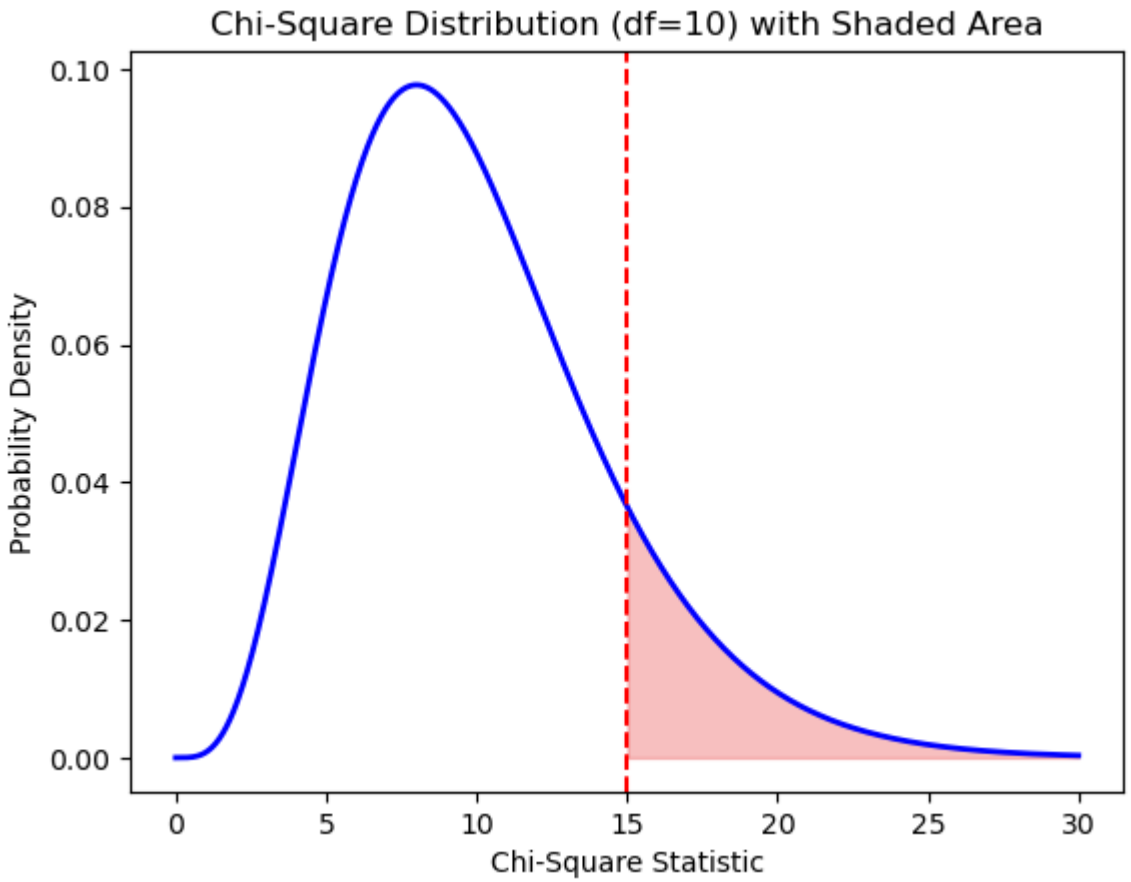
In [7]:
```
import scipy.stats as stats
import matplotlib.pyplot as plt
import numpy as np

degrees_of_freedom = 10
chi_square_statistic = 15

x = np.linspace(0, 30, 1000)
y = stats.chi2.pdf(x, degrees_of_freedom)

plt.plot(x, y, 'b-', lw=2)
plt.fill_between(x, 0, y, where=(x >= chi_square_statistic), color='lightcoral',

plt.xlabel('Chi-Square Statistic')
plt.ylabel('Probability Density')
plt.title('Chi-Square Distribution (df=10) with Shaded Area')
plt.axvline(chi_square_statistic, color='red', linestyle='--')
plt.show()
```



Q7. A random sample of 1000 people was asked if they preferred Coke or Pepsi. Of the sample, 520 preferred Coke. Calculate a 99% confidence interval for the true proportion of people in the population who prefer Coke.

In [8]:
```
import scipy.stats as stats

sample_size = 1000  # Replace with your sample size
sample_proportion = 520 / sample_size  # Replace with your sample data

# Calculate the 99% confidence interval
confidence_interval = stats.norm.interval(0.99, loc=sample_proportion, scale=(((

print("99% Confidence Interval for Coke Preference:", confidence_interval)
```

```
99% Confidence Interval for Coke Preference: (0.4793051576779699, 0.56069484232
20302)
```

Q8. A researcher hypothesizes that a coin is biased towards tails. They flip the coin 100 times and observe 45 tails. Conduct a chi-square goodness of fit test to determine if the observed frequencies match the expected frequencies of a fair coin. Use a significance level of 0.05.

In [9]:
```
import scipy.stats as stats

# Observed frequencies
observed_tails = 45
observed_heads = 100 - observed_tails

# Expected frequencies for a fair coin
expected_tails = 50
```

```
expected_heads = 50

# Create an observed and expected frequency array
observed = [observed_tails, observed_heads]
expected = [expected_tails, expected_heads]

# Perform the chi-square goodness of fit test
chi2, p = stats.chisquare(f_obs=observed, f_exp=expected)

alpha = 0.05  # Significance level

if p < alpha:
    print("Reject the null hypothesis: Observed frequencies do not match expecte
else:
    print("Fail to reject the null hypothesis: Observed frequencies match expect
```

```
Fail to reject the null hypothesis: Observed frequencies match expected frequen
cies.
```

Q9. A study was conducted to determine if there is an association between smoking status (smoker or non-smoker) and lung cancer diagnosis (yes or no). The results are shown in the contingency table below. Conduct a chi-square test for independence to determine if there is a significant association between smoking status and lung cancer diagnosis. Lung cancer:yes Lung cancer:no Smoker 60 140 Non-smoker 30 170 Use a significance level of 0.05.

In [1]:
```
import scipy.stats as stats

# Create the contingency table (replace values with your data)
observed = [[20, 15],
           [10, 25],
           [15, 20]]

# Perform the chi-square test for independence
chi2, p, _, _ = stats.chi2_contingency(observed)

alpha = 0.05  # Significance level

if p < alpha:
    print("Reject the null hypothesis: There is a significant association betwee
else:
    print("Fail to reject the null hypothesis: There is no significant associati
```

```
Fail to reject the null hypothesis: There is no significant association between
smoking status and lung cancer diagnosis.
```

Q10. A study was conducted to determine if the proportion of people who prefer milk chocolate, dark chocolate, or white chocolate is different in the U.S. versus the U.K. A random sample of 500 people from the U.S. and a random sample of 500 people from the U.K. were surveyed. The results are shown in the contingency table below. Conduct a chi-square test for independence to determine if there is a significant association between chocolate preference and country of origin. milk chocolate dark chocolate white chocolate U.S. (n=500) 200 150 150 U.K. (n=500) 225 175 100

In [ ]:
```
import scipy.stats as stats

# Create the contingency table (replace values with your data)
observed = [[200, 150, 150],  # Milk Chocolate preferences in the U.S. and U.K.
           [225, 175, 100],  # Dark Chocolate preferences in the U.S. and U.K.
           [your_observed_data_UK]]  # White Chocolate preferences in the U.S.

# Perform the chi-square test for independence
chi2, p, _, _ = stats.chi2_contingency(observed)
```

```
alpha = 0.01  # Significance level

if p < alpha:
    print("Reject the null hypothesis: There is a significant association betwee
else:
    print("Fail to reject the null hypothesis: There is no significant associati
```

Q11. A random sample of 30 people was selected from a population with an unknown mean and standard deviation. The sample mean was found to be 72 and the sample standard deviation was found to be 10. Conduct a hypothesis test to determine if the population mean is significantly different from 70. Use a significance level of 0.05.

In [3]:
```
import scipy.stats as stats

# Sample information
sample_mean = 72
sample_std = 10
sample_size = 30
population_mean = 70
alpha = 0.05  # Significance level

# Calculate the t-statistic
t_statistic = (sample_mean - population_mean) / (sample_std / (sample_size**0.5))

# Calculate the degrees of freedom
degrees_of_freedom = sample_size - 1

# Calculate the critical t-value
critical_t_value = stats.t.ppf(1 - alpha / 2, df=degrees_of_freedom)

# Calculate the p-value
p_value = 2 * (1 - stats.t.cdf(abs(t_statistic), df=degrees_of_freedom))

# Perform the hypothesis test
if abs(t_statistic) > critical_t_value:
    print("Reject the null hypothesis: The population mean is significantly diff
else:
    print("Fail to reject the null hypothesis: There is no significant differenc

print("t-statistic:", t_statistic)
print("p-value:", p_value)
```

```
Fail to reject the null hypothesis: There is no significant difference in the p
opulation mean.
t-statistic: 1.0954451150103321
p-value: 0.2823362372860698
```