In [2]:
```
import pandas as pd

# Data for Short Story 1
story1_controlled = [31.5, 31, 33.5, 21, 30.5, 28, 21, 32.5, 20, 21.5, 30, 26.
story1_experimental = [39, 39, 41.5, 42.5, 37, 40, 37, 35.5, 44, 40, 42, 37, 4

# Data for Short Story 2
story2_controlled = [32, 22, 22.5, 31, 29.5, 26, 32.5, 28.5, 22, 23, 25, 22.5,
story2_experimental = [37, 34.5, 39.5, 34, 33, 34.5, 33, 39, 40.5, 39, 39.5, 4

# Create DataFrames
df_story1 = pd.DataFrame({
    'Controlled Group': story1_controlled,
    'Experimental Group': story1_experimental
})

df_story2 = pd.DataFrame({
    'Controlled Group': story2_controlled,
    'Experimental Group': story2_experimental
})

# Print the DataFrames
print("DataFrame for Short Story 1:")
print(df_story1)

print("\nDataFrame for Short Story 2:")
print(df_story2)
```

```
DataFrame for Short Story 1:
    Controlled Group  Experimental Group
0               31.5                39.0
1               31.0                39.0
2               33.5                41.5
3               21.0                42.5
4               30.5                37.0
5               28.0                40.0
6               21.0                37.0
7               32.5                35.5
8               20.0                44.0
9               21.5                40.0
10              30.0                42.0
11              26.5                37.0
12              20.0                43.0
13              26.0                37.0
14              24.0                40.5

DataFrame for Short Story 2:
    Controlled Group  Experimental Group
0               32.0                37.0
1               22.0                34.5
2               22.5                39.5
3               31.0                34.0
4               29.5                33.0
5               26.0                34.5
6               32.5                33.0
7               28.5                39.0
8               22.0                40.5
9               23.0                39.0
10              25.0                39.5
11              22.5                45.0
12              25.0                42.0
13              23.5                40.0
14              26.0                40.5
```

In [5]:
```
controlled_group = (df_story1['Controlled Group'] + df_story2['Controlled Grou

# Calculate mean value for each index in Short Story 2
experimental_group = (df_story1['Experimental Group'] + df_story2['Experimenta

# Print the mean values
print("Mean value for each index in Short Story 1:")
print(controlled_group)

print("\nMean value for each index in Short Story 2:")
print(experiment_group)
```

```
Mean value for each index in Short Story 1:
0     31.75
1     26.50
2     28.00
3     26.00
4     30.00
5     27.00
6     26.75
7     30.50
8     21.00
9     22.25
10    27.50
11    24.50
12    22.50
13    24.75
14    25.00
Name: Controlled Group, dtype: float64

Mean value for each index in Short Story 2:
0     38.00
1     36.75
2     40.50
3     38.25
4     35.00
5     37.25
6     35.00
7     37.25
8     42.25
9     39.50
10    40.75
11    41.00
12    42.50
13    38.50
14    40.50
Name: Experimental Group, dtype: float64
```

In [6]:
```
from scipy import stats


# Perform Shapiro-Wilk test for normality
shapiro_stat_controlled, shapiro_p_controlled = stats.shapiro(controlled_group
shapiro_stat_experimental, shapiro_p_experimental = stats.shapiro(experimental

# Print the results
print("Controlled Group Shapiro-Wilk test:")
print("Statistic:", shapiro_stat_controlled, "p-value:", shapiro_p_controlled)
print("Experimental Group Shapiro-Wilk test:")
print("Statistic:", shapiro_stat_experimental, "p-value:", shapiro_p_experimen
```

```
Controlled Group Shapiro-Wilk test:
Statistic: 0.9765563607215881 p-value: 0.9405418634414673
Experimental Group Shapiro-Wilk test:
Statistic: 0.9520710706710815 p-value: 0.5576421022415161
```

For the controlled group, the Shapiro-Wilk test statistic is approximately 0.9766 and the p-value is approximately 0.9405. For the experimental group, the Shapiro-Wilk test statistic is approximately 0.9521 and the p-value is approximately 0.5576. For both groups, the p-values are greater than the typical significance level of 0.05. This suggests that we do not have sufficient evidence to reject the null hypothesis of normality. Therefore, we can assume that the data for both groups are approximately normally distributed.

Check Normal distribution in both group

In [13]:
```python
import seaborn as sns
import matplotlib.pyplot as plt

# Create density plots
plt.figure(figsize=(10, 5))

plt.subplot(1, 2, 1)
sns.kdeplot(controlled_group, color='blue', shade=True)
plt.title('Controlled Group')
plt.xlabel('Value')
plt.ylabel('Density')

plt.subplot(1, 2, 2)
sns.kdeplot(experimental_group, color='red', shade=True)
plt.title('Experimental Group')
plt.xlabel('Value')
plt.ylabel('Density')

plt.tight_layout()
plt.show()
```

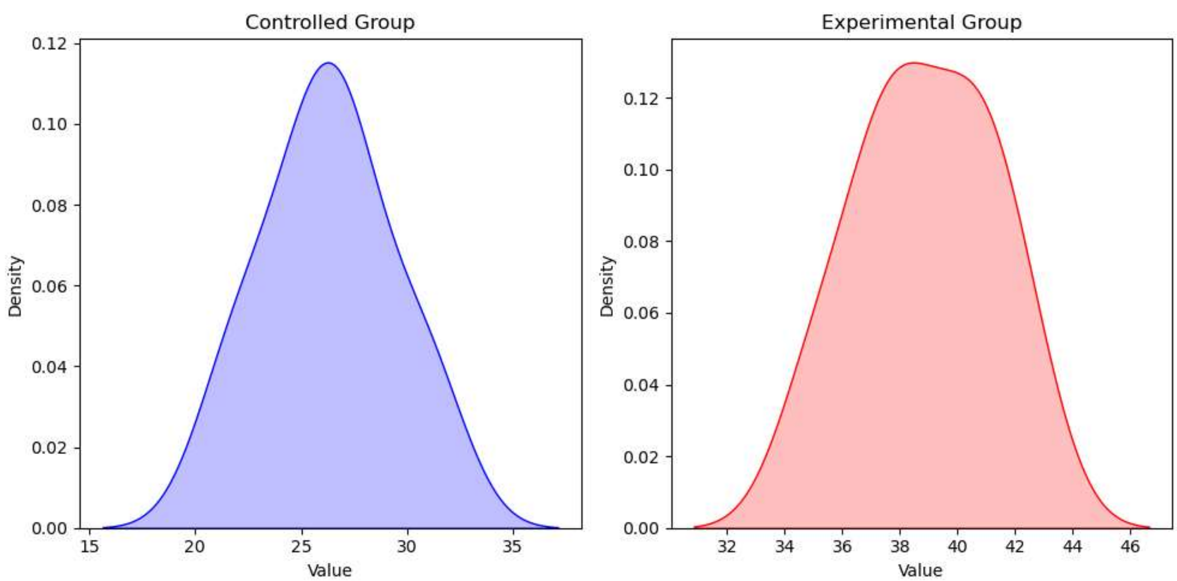
```
C:\Users\14198\AppData\Local\Temp\ipykernel_1976\1503984256.py:8: FutureWarni
ng:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(controlled_group, color='blue', shade=True)
C:\Users\14198\AppData\Local\Temp\ipykernel_1976\1503984256.py:14: FutureWarn
ing:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(experimental_group, color='red', shade=True)
```



In [14]:
```
from scipy import stats


# Perform t-test
t_statistic, p_value = stats.ttest_ind(controlled_group, experimental_group)

# Print the results
print("T-statistic:", t_statistic)
print("P-value:", p_value)

# Determine significance
alpha = 0.05
if p_value < alpha:
    print("The difference between the groups is statistically significant (rej
else:
    print("There is no statistically significant difference between the groups
```

```
T-statistic: -12.546591508815505
P-value: 5.171455465021041e-13
The difference between the groups is statistically significant (reject the nu
ll hypothesis)
```

Null Hypothesis:

There is no significant difference between the mean values of the variable in the controlled group and the experimental group. Alternative Hypothesis (:

There is a significant difference between the mean values of the variable in the controlled group and the experimental group.

Given that the p-value is much smaller than the typical significance level of 0.05, we reject the null hypothesis. This means that there is strong evidence to suggest that there is a statistically significant difference between the means of the controlled and experimental groups.

In other words, the data provides sufficient evidence to conclude that the difference in means is unlikely to have occurred by random chance alone. Therefore, we can infer that there is a genuine difference between the two groups.

In [ ]: