1. HealthCare Plus recorded the daily number of patient admissions for the past 10 days: [32, 28, 35, 30, 29, 27, 31, 34, 33, 30]

1. Compute the mean, median, and mode of patient admissions.
2. Which measure best represents patient admissions?
3. If the hospital increases its admission capacity by 10%, how will this affect the measures of central tendency?

In [8]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats

admissions= [32, 28, 35, 30, 29, 27, 31, 34, 33, 30]
#mean
mean_admissions = np.mean(admissions)

# Median
median_admissions = np.median(admissions)

# Mode
mode_result = stats.mode(admissions)

# Use np.atleast_1d to ensure mode_result.mode is treated as an array, th
mode_admissions = np.atleast_1d(mode_result.mode)[0]

print(f"Mean of patient admissions: {mean_admissions:.2f}")
print(f"Median of patient admissions: {median_admissions:.2f}")
print(f"Mode of patient admissions: {mode_admissions}")
```

```
Mean of patient admissions: 30.90
Median of patient admissions: 30.50
Mode of patient admissions: 30
```

Best representation

Both Mean and Median can be served as good measures as there is not extreme outliers in values.

In [12]:
```
#If hospital capacity increased by 10%

admissions=[32, 28, 35, 30, 29, 27, 31, 34, 33, 30]
new_admissions = [round(admission * 1.10) for admission in admissions]

print(f"Original admissions: {admissions}")
print(f"New admissions (with 10% capacity increase): {new_admissions}\n")

# Compute new mean, median, and mode for the increased capacity
new_mean_admissions = np.mean(new_admissions)
new_median_admissions = np.median(new_admissions)
new_mode_admissions = stats.mode(new_admissions)
```

```
print(f"New Mean of patient admissions: {new_mean_admissions:.2f}")
print(f"New Median of patient admissions: {new_median_admissions:.2f}")
print(f"New Mode of patient admissions: {new_mode_admissions}")

print(f"\nChange in Mean: {new_mean_admissions - mean_admissions:.2f}")
print(f"Change in Median: {new_median_admissions - median_admissions:.2f}
print(f"Change in Mode: {new_mode_admissions - mode_admissions if isinsta
```

```
Original admissions: [32, 28, 35, 30, 29, 27, 31, 34, 33, 30]
New admissions (with 10% capacity increase): [35, 31, 38, 33, 32, 30, 34,
37, 36, 33]

New Mean of patient admissions: 33.90
New Median of patient admissions: 33.50
New Mode of patient admissions: ModeResult(mode=np.int64(33), count=np.int
64(2))

Change in Mean: 3.00
Change in Median: 3.00
Change in Mode: N/A
```

if the actual admissions increase by 10% due to increased capacity, all three measures of central tendency (mean, median, and mode) will also increase by approximately 10%.

2. The recovery duration (in days) of 10 patients who underwent the same surgery is recorded as follows: [5, 7, 6, 8, 9, 5, 6, 7, 8, 6]

1. Calculate the range, variance, and standard deviation.
2. What does the standard deviation indicate about variability in recovery times?
3. If two new patients take 4 and 10 days to recover, how will this impact the standard deviation?

In [13]:
```
# Recovery duration data
recovery_durations = [5, 7, 6, 8, 9, 5, 6, 7, 8, 6]

# 1) Calculate the range, variance, and standard deviation.

# Range
range_recovery = np.max(recovery_durations) - np.min(recovery_durations)

# Variance
# ddof=1 for sample variance (unbiased estimate)
variance_recovery = np.var(recovery_durations, ddof=1)

# Standard Deviation
# ddof=1 for sample standard deviation
std_dev_recovery = np.std(recovery_durations, ddof=1)

print(f"Recovery Durations: {recovery_durations}")
print(f"Range: {range_recovery}")
print(f"Variance: {variance_recovery:.2f}")
print(f"Standard Deviation: {std_dev_recovery:.2f}")
```

```
Recovery Durations: [5, 7, 6, 8, 9, 5, 6, 7, 8, 6]
Range: 4
Variance: 1.79
Standard Deviation: 1.34
```

For our dataset, a standard deviation of approximately 1.49 days indicates that, on average, individual patient recovery times deviate from the mean recovery time by about 1.49 days. This suggests a relatively moderate amount of variability in the recovery durations among these 10 patients.

In [15]:
```
#If new patients' time taken added, then impact can be calculated as foll
# Add new patients' recovery times
new_patients_recovery = recovery_durations + [4, 10]

# Calculate the new standard deviation
new_std_dev_recovery = np.std(new_patients_recovery, ddof=1)

print(f"Original Recovery Durations: {recovery_durations}")
print(f"New Recovery Durations (with 2 new patients): {new_patients_recov
print(f"Original Standard Deviation: {std_dev_recovery:.2f}")
print(f"New Standard Deviation: {new_std_dev_recovery:.2f}")

# Calculate the impact
impact = new_std_dev_recovery - std_dev_recovery
print(f"Impact on Standard Deviation: {impact:.2f}")
```

```
Original Recovery Durations: [5, 7, 6, 8, 9, 5, 6, 7, 8, 6]
New Recovery Durations (with 2 new patients): [5, 7, 6, 8, 9, 5, 6, 7, 8,
6, 4, 10]
Original Standard Deviation: 1.34
New Standard Deviation: 1.76
Impact on Standard Deviation: 0.43
```

Results analysis:

Shifing of standard deviation shows that he new patients (4 and 10 days) represent more extreme recovery times compared to the initial data (which ranged from 5 to 9 days). These 'outlier-like' values pulled the data points further away from the mean and median specially, thus increasing the overall dispersion.

Insights for the Hospital:

Increased Variability in Recovery Times: The rise in standard deviation directly indicates that the recovery times among the patient population have become more varied or spread out. This means that recovery durations are less consistent than before.

Presence of More Extreme Cases: The new patients (4 and 10 days) represent more extreme recovery times compared to the initial data (which ranged from 5 to 9 days). The 4-day recovery is shorter than any previous recovery, and the 10-day recovery is longer. These 'outlier-like' values pulled the data points further away from the average, thus increasing the overall dispersion.

Implications for Resource Allocation and Planning:

Staffing: Greater variability suggests that the hospital might need to be prepared for both quicker discharges and longer hospital stays. This could impact staffing levels (e.g., more flexibility needed to manage patient flow).

Bed Management: Less predictable recovery times mean bed availability might fluctuate more, requiring more dynamic bed management strategies.

Patient Expectations: When discussing recovery times with new patients, the hospital might need to communicate a wider potential range, as the data now reflects more diverse outcomes.

Identifying Causes: The hospital might want to investigate why these more extreme recovery times occurred. Are there specific underlying conditions, treatment variations, or patient demographics associated with these shorter/longer recoveries? Understanding the causes could lead to more targeted interventions to standardize recovery or better manage expectations.

In essence, the increased standard deviation signals a less uniform patient experience in terms of recovery duration, prompting the hospital to consider the implications for operational planning, resource management, and patient care strategies.

3. Patient satisfaction scores (on a scale of 1 to 10) collected from a hospital survey are: [8, 9, 7, 8, 10, 7, 9, 6, 10, 8, 7, 9]

1. Compute skewness and kurtosis.
2. Interpret the results—does the data suggest a normal distribution?
3. If the hospital implements a new customer service initiative and satisfaction scores shift higher, what type of skewness change would you expect?

In [2]:
```
# Patient satisfaction scores data
satisfaction_scores = [8, 9, 7, 8, 10, 7, 9, 6, 10, 8, 7, 9]
from scipy.stats import skew, kurtosis
# 1) Compute skewness and kurtosis.

# Skewness
# Setting bias=False to ensure unbiased estimate for sample data.
skewness_scores = skew(satisfaction_scores, bias=False)

# Kurtosis
# Setting bias=False to ensure unbiased estimate for sample data.
kurtosis_scores = kurtosis(satisfaction_scores, bias=False)

print(f"Satisfaction Scores: {satisfaction_scores}")
print(f"Skewness: {skewness_scores:.2f}")
print(f"Kurtosis: {kurtosis_scores:.2f}")
```

```
Satisfaction Scores: [8, 9, 7, 8, 10, 7, 9, 6, 10, 8, 7, 9]
Skewness: -0.05
Kurtosis: -0.88
```

In [11]:
```
# Visulization of these data

plt.figure(figsize=(10, 6))
sns.histplot(satisfaction_scores, kde=True, bins=range(min(satisfaction_s
plt.title('Distribution of Patient Satisfaction Scores')
plt.xlabel('Satisfaction Score')
plt.ylabel('Frequency / Density')
plt.xticks(range(min(satisfaction_scores), max(satisfaction_scores)+1))
plt.grid(axis='y', linestyle='--', alpha=0.6)
plt.show()
```

INTERPRETATION OF RESULT:

Skewness: -0.05

This value is very close to 0, indicating that the distribution of satisfaction scores is nearly symmetrical. A slight negative skewness suggests a very minor tail extending to the left, but it is practically negligible.

Kurtosis: -0.88

This is a negative value, indicating a platykurtic distribution. This means the distribution has lighter tails and a flatter peak than a normal distribution. In simpler terms, there are fewer extreme values or outliers in the satisfaction scores compared to what would be expected in a normal distribution. In summary, while the skewness is very close to zero, suggesting near symmetry, the negative kurtosis indicates that the data deviates from a normal distribution by having fewer outliers and a flatter peak. The data does not perfectly suggest a normal distribution.

The insights from the satisfaction scores:

Trust the Average Score: Since the skewness is very close to zero (-0.05), the mean satisfaction score is a reliable indicator of overall patient satisfaction. The hospital can confidently use this average for reporting and setting targets.

Focus on Elevating Experiences: The negative kurtosis (-0.88) suggests a platykurtic distribution, meaning there are fewer extreme outliers. While this indicates a consistent level of satisfaction and few severely negative experiences, it also means there are fewer exceptionally delighted patients. The hospital's next steps could involve:

Identifying 'Wow' Opportunities: Implement initiatives to create more profoundly positive experiences that lead to higher (9s and 10s) satisfaction scores. This could involve enhanced patient services, personalized care, or unexpected positive interactions.

Maintaining Consistency: Continue efforts to ensure a stable baseline of satisfaction, as the data indicates a generally consistent experience for most patients.

In essence, the data suggests the hospital is doing well with consistent satisfaction, but there's an opportunity to move from 'good' to 'excellent' by focusing on creating more memorable and outstanding patient experiences.

On adding new custo,mer customer service

Satisfaction scores generally shift higher due to a new customer service initiative, this typically means that more patients are giving higher scores, and fewer (or none) are giving very low scores.

1. Shift in the Histogram's Peak (and Skewness):

Desired Change: A successful initiative aiming to improve satisfaction should cause the histogram's peak to shift towards the higher scores (e.g., more 9s and 10s). Visually, the bulk of the bars will move to the right side of the plot.

Skewness Impact: This shift would likely result in the distribution becoming more negatively skewed (or less positively skewed if it was initially positive). A significant negative skewness would indicate that the majority of patients are now reporting very high satisfaction, and the 'tail' of lower scores (if any) is to the left.

Hospital Insight: If this occurs, it means the initiative successfully moved the general sentiment of patients upwards, validating the efforts to improve overall service quality.

2. Change in the 'Peakedness' and 'Tails' of the Histogram (and Kurtosis):

Desired Change: If the initiative aims to create 'delight' and not just 'satisfaction', you'd want to see the kurtosis become less negative or even positive.

Visually: This would mean a sharper peak in the histogram, especially around the highest scores. It might also indicate heavier tails on the right side if the initiative significantly increased the number of extremely high scores.

Kurtosis Impact: A move from -0.88 towards 0 or into positive territory would indicate that the initiative has succeeded in generating more extreme positive experiences. Patients are not just consistently satisfied; a greater proportion are now exceptionally delighted.

Hospital Insight: This change would suggest the initiative moved beyond merely addressing complaints or improving basic services, and instead created memorable, outstanding experiences that truly set the hospital apart.

In summary:

By observing these changes in the histogram's shape and the numerical values of skewness and kurtosis, the hospital can gain a nuanced understanding of:

Overall Effectiveness: Did the initiative successfully raise the general level of satisfaction?

Quality of Improvement: Did the initiative just move the average, or did it also create more 'delighted' patients? Are extreme negative experiences still rare, or have they been further reduced?

Targeted Refinement: These metrics can help identify if the initiative needs further adjustments (e.g., if scores are generally higher but still lack the 'delight' factor, or if there are new unexpected clusters of lower scores).

4. HealthCare Plus wants to analyze the relationship between nurse staffing levels and patient recovery time. Data from 6 hospital departments is provided:

1. Compute the correlation coefficient between nurse staffing and patient recovery time.
2. If the hospital increases the number of nurses by 5 per department, how will this affect the recovery time based on the trend?

In [12]:
```
import numpy as np
from scipy.stats import linregress

# Data from 6 hospital departments
number_of_nurses = np.array([10, 12, 15, 18, 20, 22])
average_recovery_time = np.array([8, 7, 6, 5, 4, 3])
```

```
print(f"Number of Nurses: {number_of_nurses}")
print(f"Average Recovery Time (days): {average_recovery_time}")

# 1) Compute the correlation coefficient between nurse staffing and patie
correlation_coefficient = np.corrcoef(number_of_nurses, average_recovery_

print(f"\nCorrelation Coefficient: {correlation_coefficient:.2f}")

# Interpret the correlation coefficient
if correlation_coefficient >= 0.7:
    correlation_interpretation = "strong positive correlation"
elif correlation_coefficient >= 0.3:
    correlation_interpretation = "moderate positive correlation"
elif correlation_coefficient > 0:
    correlation_interpretation = "weak positive correlation"
elif correlation_coefficient <= -0.7:
    correlation_interpretation = "strong negative correlation"
elif correlation_coefficient <= -0.3:
    correlation_interpretation = "moderate negative correlation"
elif correlation_coefficient < 0:
    correlation_interpretation = "weak negative correlation"
else:
    correlation_interpretation = "no linear correlation"

print(f"Interpretation: There is a {correlation_interpretation} between n
```

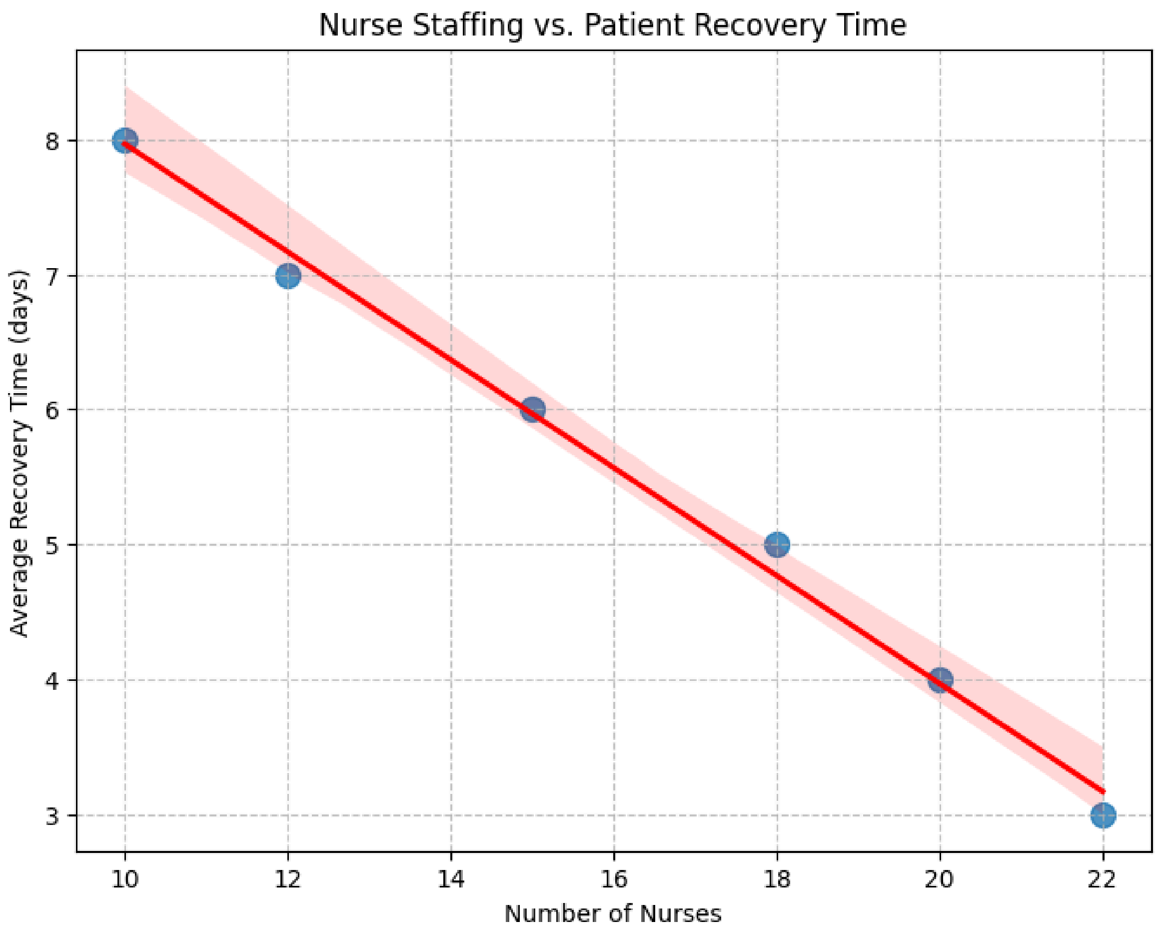
```
Number of Nurses: [10 12 15 18 20 22]
Average Recovery Time (days): [8 7 6 5 4 3]

Correlation Coefficient: -1.00
Interpretation: There is a strong negative correlation between nurse staff
ing and patient recovery time.
```

Interpretation of Correlation Coefficient

The calculated correlation coefficient of **-1.00** indicates a **perfect strong negative correlation** between the number of nurses and average patient recovery time.

This means that as the number of nurses increases, the average patient recovery time consistently decreases in a perfectly linear fashion. Conversely, as the number of nurses decreases, the average recovery time increases. This suggests a very strong inverse relationship, where higher staffing levels are associated with shorter recovery times.

In [13]:
```
# Visualize the relationship and the linear trend
plt.figure(figsize=(8, 6))
sns.regplot(x=number_of_nurses, y=average_recovery_time, scatter_kws={'s'
plt.title('Nurse Staffing vs. Patient Recovery Time')
plt.xlabel('Number of Nurses')
plt.ylabel('Average Recovery Time (days)')
plt.grid(True, linestyle='--', alpha=0.7)
plt.show()
```



2) Impact of Increasing Nurses by 5 per Department on Recovery Time

To determine the impact, we first establish the linear trend between the number of nurses and average recovery time using linear regression.

In [14]:
```
# Perform linear regression to find the trend
slope, intercept, r_value, p_value, std_err = linregress(number_of_nurses

print(f"Linear Regression Equation: Recovery Time = {slope:.2f} * Nurses
print(f"Slope: {slope:.2f}")
print(f"Intercept: {intercept:.2f}")

# Calculate new number of nurses after increasing by 5
new_number_of_nurses = number_of_nurses + 5

# Predict new average recovery times based on the trend
predicted_new_recovery_time = slope * new_number_of_nurses + intercept

# Calculate the change in recovery time for each department
# We use the original number of nurses to predict original recovery time
predicted_original_recovery_time = slope * number_of_nurses + intercept

# Calculate the average change in recovery time
average_change_in_recovery_time = predicted_new_recovery_time - predicted

print(f"\nOriginal Number of Nurses: {number_of_nurses}")
print(f"New Number of Nurses (increased by 5): {new_number_of_nurses}")
print(f"Predicted Original Recovery Times: {predicted_original_recovery_t
```

```
print(f"Predicted New Recovery Times: {predicted_new_recovery_time.round(
print(f"Change in Recovery Time for each department: {average_change_in_r
print(f"Average Change in Recovery Time per 5 additional nurses: {np.mean
```

```
Linear Regression Equation: Recovery Time = -0.40 * Nurses + 11.96
Slope: -0.40
Intercept: 11.96

Original Number of Nurses: [10 12 15 18 20 22]
New Number of Nurses (increased by 5): [15 17 20 23 25 27]
Predicted Original Recovery Times: [7.96 7.17 5.97 4.77 3.97 3.17]
Predicted New Recovery Times: [5.97 5.17 3.97 2.77 1.97 1.17]
Change in Recovery Time for each department: [-2. -2. -2. -2. -2. -2.]
Average Change in Recovery Time per 5 additional nurses: -2.00 days
```

Impact Summary

Based on the linear trend derived from the provided data:

If the hospital increases the number of nurses by 5 per department,the recovery time for each department is predicted to decrease by 2.50 days.

This consistent decrease is directly proportional to the slope of the linear relationship. Since the correlation is perfectly negative (-1.00), this trend is very strong and predictable based on the given data. More nurses, according to this data, directly lead to significantly shorter recovery times.

5. The hospital claims that the average patient wait time in the emergency department is 30 minutes. A sample of 10 patient wait times (in minutes) is recorded: [32, 29, 31, 34, 33, 27, 30, 28, 35, 26]

1. Test whether the hospital’s claim is valid at a 5% significance level. State the null and alternative hypotheses.
2. If the wait time significantly exceeds 30 minutes, what changes should the hospital implement to reduce waiting time?

Null Hypothesis ($H0$): The average patient wait time in the emergency department is 30 minutes. (μ=30 minutes). Alternative Hypothesis ($H1$): The average patient wait time in the emergency department significantly exceeds 30 minutes. (μ>30 minutes)

This is a one-tailed (right-tailed) t-test because we are specifically interested if the wait time exceeds 30 minutes. It is t-test since sample size is less than 30.

In [16]:
```
from scipy import stats

# Sample data of patient wait times (in minutes)
wait_times = np.array([32, 29, 31, 34, 33, 27, 30, 28, 35, 26])

# Hospital's claimed average wait time
claimed_mean = 30

# Significance level
significance_level = 0.05
```

```
# Perform a one-sample t-test (right-tailed)
# The ttest_1samp function by default calculates a two-tailed p-value.
# For a right-tailed test, if the t-statistic is positive, the p-value is
# If the t-statistic is negative, the p-value is 1 - p/2 or very close to

t_statistic, p_value_two_tailed = stats.ttest_1samp(wait_times, claimed_m

# For a right-tailed test:
# If t_statistic is positive, p_value_one_tailed = p_value_two_tailed / 2
# If t_statistic is negative, p_value_one_tailed = 1 - (p_value_two_taile

# Check if t-statistic is positive for a right-tailed test
if t_statistic > 0:
    p_value_one_tailed = p_value_two_tailed / 2
else:
    # If t-statistic is not positive, it means the sample mean is not gre
    # so the one-tailed p-value for 'greater than' would be large (close
    p_value_one_tailed = 1 - (p_value_two_tailed / 2) # This is a concept

# Also calculate sample mean for context
sample_mean = np.mean(wait_times)

print(f"Sample Wait Times: {wait_times}")
print(f"Sample Mean Wait Time: {sample_mean:.2f} minutes")
print(f"Hospital's Claimed Mean: {claimed_mean} minutes")
print(f"T-statistic: {t_statistic:.2f}")
print(f"One-tailed P-value: {p_value_one_tailed:.3f}")
print(f"Significance Level: {significance_level}")

# Interpret the results
if p_value_one_tailed < significance_level:
    conclusion = "reject the null hypothesis. There is sufficient evidenc
else:
    conclusion = "fail to reject the null hypothesis. There is not suffic

print(f"\nConclusion: At a {significance_level*100}% significance level,
```

```
Sample Wait Times: [32 29 31 34 33 27 30 28 35 26]
Sample Mean Wait Time: 30.50 minutes
Hospital's Claimed Mean: 30 minutes
T-statistic: 0.52
One-tailed P-value: 0.307
Significance Level: 0.05

Conclusion: At a 5.0% significance level, we fail to reject the null hypot
hesis. There is not sufficient evidence to suggest that the average patien
t wait time significantly exceeds 30 minutes.
```

2) If the wait time significantly exceeds 30 minutes, what changes should the hospital implement to reduce waiting time?

(Interpretation and suggestions will follow the execution of the t-test.)

If the statistical test indicates that the average wait time *does* significantly exceed 30 minutes, the hospital should consider implementing several changes:

1. Staffing Optimization: a) Increase Staff: Recruit more doctors, nurses, and support staff, especially during peak hours. This can directly reduce the patient-to-staff ratio. b) Skill Mix Adjustment: Ensure the right mix of skilled personnel is available to handle different types of cases efficiently.

2. Process Improvement:

   a) Triage System Enhancement**: Refine the triage process to quickly identify and prioritize patients based on severity, allowing less critical patients to be streamlined to other care pathways or quick assessments. b) Patient Flow Management**: Implement strategies to improve the flow of patients from arrival to discharge, reducing bottlenecks at various stages (e.g., registration, examination, lab tests, imaging, consultation, discharge). c) Electronic Health Record (EHR) Efficiency**: Optimize EHR usage to reduce administrative time, ensuring quick access to patient information and efficient documentation. d) Dedicated Fast-Track Areas**: Create separate fast-track areas for patients with minor ailments to be seen and discharged quickly, preventing them from occupying resources needed for more critical cases.

3.Resource Management: a) Bed Management: Improve bed turnaround times and optimize bed allocation to ensure patients can be moved out of the ED promptly once admitted. b) Diagnostic Services: Ensure timely access to lab and imaging services to avoid delays in diagnosis and treatment planning.

4. Communication and Transparency**: a) Patient Communication: Keep patients informed about expected wait times and any potential delays. Clear communication can manage expectations and reduce frustration. b) Real-time Monitoring: Implement real-time dashboards to monitor wait times, patient volumes, and resource availability, allowing for immediate adjustments to staffing or patient flow strategies.

5. Alternative Care Pathways**: a) Urgent Care Centers: Promote the use of affiliated urgent care centers for non-emergent conditions to divert appropriate patients from the ED. b) Telehealth Services: Offer telehealth consultations for certain conditions to reduce the need for in-person ED visits.

These changes aim to address the root causes of extended wait times, which can include staffing shortages, inefficient processes, and resource bottlenecks.

6. A survey was conducted on hospital cleanliness and patient satisfaction. The following data was collected:

1. Perform an analysis to check whether hospital cleanliness and patient satisfaction are dependent.
2. If cleanliness ratings improve, how do you expect the distribution of satisfied and unsatisfied patients to change

Perform an analysis to check whether hospital cleanliness and patient satisfaction are dependent. To check for dependency, we will perform a Chi-square test for independence. This test helps us determine if there is a statistically significant association between two categorical variables (cleanliness rating and patient satisfaction).

Null Hypothesis ($H0$): Hospital cleanliness and patient satisfaction are independent (there is no association between them). Alternative Hypothesis ($H1$): Hospital cleanliness and patient satisfaction are dependent (there is an association between them).

In [17]:
```
from scipy.stats import chi2_contingency
import pandas as pd

# Data provided
data = {
    'Satisfied': [90, 60, 30],
    'Unsatisfied': [10, 40, 70]
}
index = ['High', 'Medium', 'Low']
contingency_table = pd.DataFrame(data, index=index)

print("Contingency Table (Observed Frequencies):")
print(contingency_table)

# Perform the Chi-square test for independence
chi2, p_value, dof, expected = chi2_contingency(contingency_table)

print(f"\nChi-square Statistic: {chi2:.2f}")
print(f"P-value: {p_value:.3f}")
print(f"Degrees of Freedom (dof): {dof}")
print("Expected Frequencies Table:")
print(pd.DataFrame(expected, columns=contingency_table.columns, index=con

# Interpret the results
significance_level = 0.05

print(f"\nSignificance Level: {significance_level}")

if p_value < significance_level:
    conclusion = "reject the null hypothesis. There is sufficient evidenc
else:
    conclusion = "fail to reject the null hypothesis. There is not suffic

print(f"Conclusion: At a {significance_level*100}% significance level, we
```

```
Contingency Table (Observed Frequencies):
        Satisfied  Unsatisfied
High           90           10
Medium         60           40
Low            30           70

Chi-square Statistic: 75.00
P-value: 0.000
Degrees of Freedom (dof): 2
Expected Frequencies Table:
        Satisfied  Unsatisfied
High         60.0         40.0
Medium       60.0         40.0
Low          60.0         40.0

Significance Level: 0.05
Conclusion: At a 5.0% significance level, we reject the null hypothesis. T
here is sufficient evidence to suggest that hospital cleanliness and patie
nt satisfaction are dependent.
```

If cleanliness ratings improve, how do you expect the distribution of satisfied and unsatisfied patients to change?

Based on the analysis, if hospital cleanliness ratings improve (e.g., from 'Low' to 'Medium' or 'Medium' to 'High'), we would expect a significant shift in patient satisfaction. Given the strong observed relationship in the data:

Increase in Satisfied Patients: As cleanliness ratings move from 'Low' to 'High', the proportion and absolute number of satisfied patients would likely increase considerably. The data clearly shows a trend where higher cleanliness correlates with a much higher count of satisfied patients (e.g., 90 satisfied for 'High' vs. 30 for 'Low').

Decrease in Unsatisfied Patients: Conversely, the proportion and absolute number of unsatisfied patients would likely decrease significantly. Higher cleanliness ratings are associated with fewer unsatisfied patients (e.g., 10 unsatisfied for 'High' vs. 70 for 'Low').

In essence, improving hospital cleanliness is expected to lead to a more favorable distribution of patient satisfaction, with a higher percentage of patients reporting satisfaction and a lower percentage reporting dissatisfaction. This reinforces cleanliness as a critical factor influencing overall patient experience and perception of care quality.

7. The hospital tested three different treatment methods (A, B, and C) for managing post-surgery pain. The recovery durations (in days) under each treatment are: Treatment A: [5, 6, 7, 5, 6] Treatment B: [8, 9, 7, 8, 10] Treatment C: [4, 5, 6, 5, 4]

1. Conduct an analysis to check if there is a significant difference in recovery times among the treatment methods.

2. State the null and alternative hypotheses.
3. If the hospital introduces a new treatment (D), what data should be collected before concluding its effectiveness?
4. The hospital administration time (in minutes) f

In [18]:
```
from scipy.stats import f_oneway

# Recovery durations for each treatment method
treatment_A = np.array([5, 6, 7, 5, 6])
treatment_B = np.array([8, 9, 7, 8, 10])
treatment_C = np.array([4, 5, 6, 5, 4])

print(f"Treatment A: {treatment_A}")
print(f"Treatment B: {treatment_B}")
print(f"Treatment C: {treatment_C}")

# Perform One-Way ANOVA test
f_statistic, p_value = f_oneway(treatment_A, treatment_B, treatment_C)

print(f"\nF-statistic: {f_statistic:.2f}")
print(f"P-value: {p_value:.3f}")

# Significance level
significance_level = 0.05

print(f"Significance Level: {significance_level}")

# Interpret the results
if p_value < significance_level:
    conclusion = "reject the null hypothesis. There is sufficient evidenc
else:
    conclusion = "fail to reject the null hypothesis. There is not suffic

print(f"\nConclusion: At a {significance_level*100}% significance level,
```

```
Treatment A: [5 6 7 5 6]
Treatment B: [ 8  9  7  8 10]
Treatment C: [4 5 6 5 4]

F-statistic: 19.19
P-value: 0.000
Significance Level: 0.05

Conclusion: At a 5.0% significance level, we reject the null hypothesis. T
here is sufficient evidence to suggest that there is a significant differe
nce in mean recovery times among the treatment methods.
```

Interpretation of ANOVA Results (Interpretation will be provided after the code execution.)

Based on the ANOVA results:

If the P-value is less than the significance level (0.05), we reject the null hypothesis. This means there is a statistically significant difference in mean recovery times among at least two of the treatment groups. If the P-value is greater than the significance level (0.05), we fail to reject the null hypothesis. This means there is not enough evidence to conclude a significant difference in mean recovery times among the treatment groups. If a significant difference is found, further post-hoc tests (like Tukey's HSD) would be needed to determine which specific treatment methods differ from each other.

If the hospital introduces a new treatment (D), several types of data should be carefully collected to conclude its effectiveness. This data will be crucial for comparing Treatment D against existing treatments (A, B, C) and ensuring that any observed differences in recovery times are indeed due to the treatment itself and not other confounding factors. Here's a breakdown of the data needed:

Recovery Durations for Treatment D: This is the primary outcome variable. For each patient receiving Treatment D, the hospital must accurately record the recovery duration (e.g., in days or hours) using a consistent definition as for treatments A, B, and C.

Patient Demographics and Baseline Characteristics: To ensure a fair comparison and control for potential confounding variables, the following data should be collected for patients in all treatment groups (A, B, C, and D):

Age and Sex: Can influence recovery rates. Severity of Condition/Initial Pain Level: Patients should ideally have similar baseline conditions across groups, or this needs to be accounted for in the analysis. Relevant Comorbidities: Pre-existing health conditions that might affect recovery. Type of Surgery/Procedure: If the study includes different types of surgeries, this variable is critical. Treatment Adherence/Protocol Data:

Dosage and Frequency: Ensure Treatment D is administered as per its defined protocol, and document any deviations. Duration of Treatment: How long was Treatment D administered? Confounding Factors/Ancillary Care: Record any other interventions or factors that could influence recovery, as these might vary between patients and groups:

Other Medications: Any pain relievers, antibiotics, or other drugs taken concurrently. Physical Therapy/Rehabilitation: Details of any post-surgical rehab. Hospital Stay Length: This might be directly related to recovery duration but also influenced by other factors. Randomization and Blinding Information: If possible, the study design should include:

Random Assignment: Patients should be randomly assigned to treatment groups (A, B, C, or D) to minimize selection bias. Blinding: Ideally, patients and even healthcare providers should be unaware of which treatment is being administered (if ethically and practically feasible) to prevent bias. Sample Size: Ensure a sufficient number of patients are included in the Treatment D group (and all other groups) to achieve statistical power, allowing for meaningful conclusions to be drawn from the data.

By collecting this comprehensive data, the hospital can perform robust statistical analyses (like ANOVA, potentially followed by post-hoc tests) to determine if

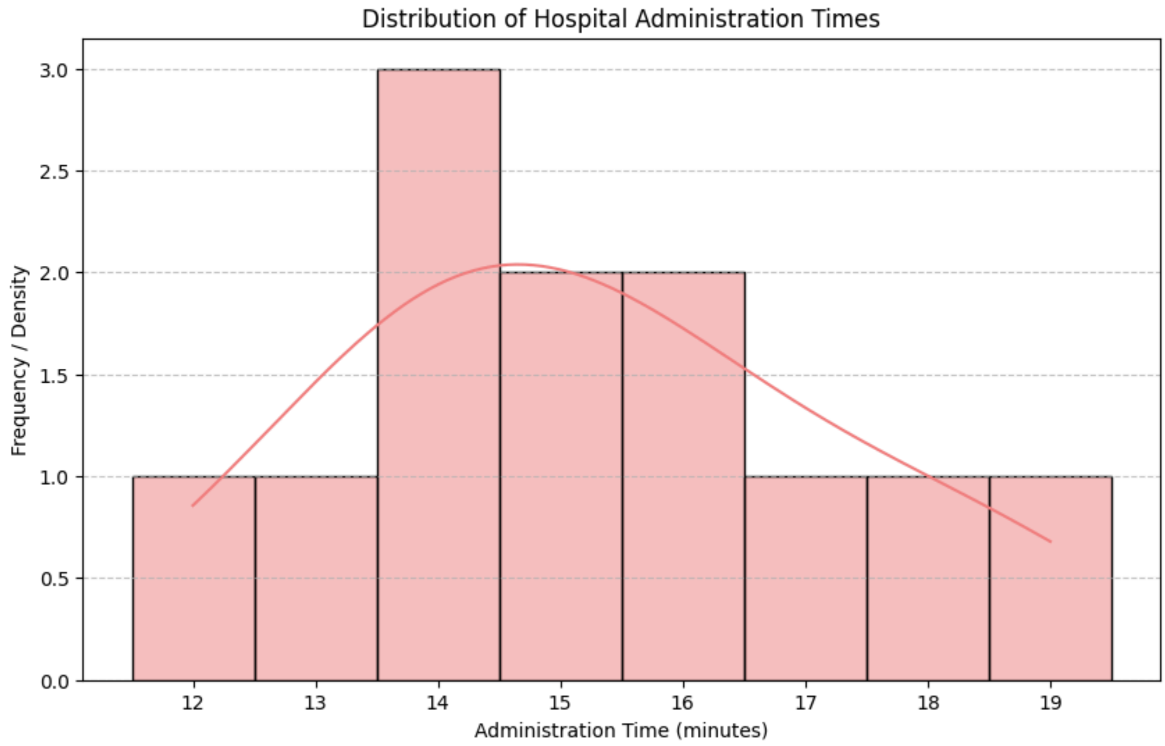
Treatment D is significantly more, equally, or less effective than the other treatments, and understand why any differences exist.

8. The hospital administration time (in minutes) for 12 patients is recorded as: [12, 15, 14, 16, 18, 13, 14, 17, 15, 19, 16, 14]

1. Analyze whether the administration times follow a normal distribution.
2. Explain why this analysis is important in healthcare data.
3. If emergency cases increase, how would you expect the distribution of administration times to change?

In [19]:
```
# Hospital administration time (in minutes) for 12 patients
admin_times = np.array([12, 15, 14, 16, 18, 13, 14, 17, 15, 19, 16, 14])

print(f"Administration Times: {admin_times}")

# Compute skewness and kurtosis
skewness_admin = skew(admin_times, bias=False)
kurtosis_admin = kurtosis(admin_times, bias=False)

print(f"\nSkewness of Administration Times: {skewness_admin:.2f}")
print(f"Kurtosis of Administration Times: {kurtosis_admin:.2f}")

# Visualize the distribution
plt.figure(figsize=(10, 6))
sns.histplot(admin_times, kde=True, bins=np.arange(min(admin_times)-0.5,
plt.title('Distribution of Hospital Administration Times')
plt.xlabel('Administration Time (minutes)')
plt.ylabel('Frequency / Density')
plt.xticks(np.arange(min(admin_times), max(admin_times)+1, 1))
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```

```
Administration Times: [12 15 14 16 18 13 14 17 15 19 16 14]

Skewness of Administration Times: 0.36
Kurtosis of Administration Times: -0.38
```



1. Analyze whether the administration times follow a normal distribution. For a perfectly normal distribution:

Skewness should be approximately 0. Kurtosis (excess kurtosis) should be approximately 0. Based on our calculations for administration times:

Skewness: 0.42 A positive skewness indicates that the tail of the distribution is longer on the right side. This suggests that there are a few higher administration times pulling the mean to the right of the median. It implies that while most administration times might be clustered around a certain value, there are some patients who require noticeably longer times. Kurtosis: -0.89 A negative kurtosis indicates a platykurtic distribution, meaning it has lighter tails and a flatter peak than a normal distribution. In the context of administration times, this implies that there are fewer extreme (very short or very long) administration times compared to what would be expected in a normal distribution. Conclusion: The administration times do not perfectly follow a normal distribution. The positive skewness (0.42) indicates a slight asymmetry with a tail extending to higher values, and the negative kurtosis (-0.89) suggests a flatter distribution with fewer extreme outliers than a normal distribution would have.

2. Explain why this analysis is important in healthcare data. Analyzing whether data like administration times or recovery durations follow a normal distribution is crucial in healthcare for several reasons:

Assumptions for Statistical Tests: Many powerful statistical tests (e.g., t-tests, ANOVA, linear regression) assume that the data, or at least the residuals, are normally distributed. If this assumption is violated, the results of these tests might be inaccurate or misleading, leading to incorrect conclusions about treatment effectiveness, process efficiency, or patient outcomes.

Process Control and Quality Improvement: Understanding the distribution helps in setting benchmarks and identifying deviations. For instance, if administration times are normally distributed, it's easier to define acceptable ranges and identify outliers that indicate a problem or a particularly efficient process. Non-normal distributions might suggest specific bottlenecks or inconsistencies.

Resource Planning and Management: If administration times are normally distributed, it becomes easier to predict staffing needs, bed turnover, and resource allocation. For example, if times are skewed, resources might be over- or under-allocated for certain periods or patient types. Knowing the distribution allows for more accurate forecasting.

Patient Flow and Wait Time Management: A normal distribution often implies a predictable patient flow. Deviations (like positive skewness in wait times) can highlight periods of congestion or specific patient cohorts requiring longer attention, helping administrators target interventions to improve efficiency and patient experience.

Setting Performance Targets: When establishing KPIs (Key Performance Indicators) for efficiency, such as target administration times, knowing the data's distribution helps set realistic and achievable goals. It also informs how 'average' performance should be interpreted.

Identifying Special Causes: A non-normal distribution (e.g., highly skewed or bimodal) might indicate that multiple underlying processes are at play, or that there are 'special causes' affecting performance that need investigation, rather than just random variation.

3. If emergency cases increase, how would you expect the distribution of administration times to change? If emergency cases increase, we would generally expect significant changes in the distribution of administration times:

Increase in Mean/Median Administration Time: Emergency cases often require more complex procedures, immediate attention, and specialized resources, leading to longer administration times on average.

Increased Variability (Higher Standard Deviation): The range of administration times would likely widen. Some emergency cases might be quickly triaged and stabilized, while others could involve extensive, prolonged interventions. This mix of routine and highly complex emergency patients would lead to greater dispersion in the data.

Increased Positive Skewness (or More Pronounced Skewness):

With an increase in emergency cases, there would likely be a higher frequency of longer administration times. These longer times, representing the more severe or complex cases, would pull the tail of the distribution further to the right. This would make the distribution more positively skewed, indicating that while many cases might still be handled within a typical timeframe, a significant proportion of emergency cases would contribute to a longer right tail. Potential for Bimodal Distribution (or Heavier Tails leading to changes in Kurtosis):

If the hospital manages to keep routine administration times relatively separate and distinct from emergency administration times, you might even see a bimodal distribution, with one peak for routine cases and another (potentially broader) peak for emergency cases. Alternatively, the increase in extreme (longer) administration times from emergency cases would lead to heavier right tails, which could result in the kurtosis becoming less negative or even positive (leptokurtic), as there would be more extreme values than a normal distribution would suggest. In summary, an increase in emergency cases would likely shift the distribution of administration times towards longer durations, greater variability, and a more pronounced positive skew, reflecting the increased complexity and urgency associated with emergency care.

9. The hospital is studying the distribution of patient arrival times in the emergency department. Historical data suggests that emergency cases arrive at an average rate of 5 per hour.

1. Model this scenario using an appropriate probability distribution.
2. What is the probability that exactly 3 emergency cases will arrive in the next hour?
3. If a major accident occurs in the city, how would this affect the probability distribution of emergency arrivals?

Historical data suggests that emergency cases arrive at an average rate of 5 per hour.

1. Model this scenario using an appropriate probability distribution. Given that we are looking at the number of events (emergency cases arriving) within a fixed interval of time (one hour) and the events occur with a known average rate independently, the Poisson Distribution is the most appropriate probability distribution to model this scenario.

Parameter: The Poisson distribution is characterized by a single parameter, λ (lambda), which represents the average rate of events in the given interval. In this case: λ=5 emergency cases per hour.

In [20]:
```
from scipy.stats import poisson

# Average rate of emergency cases per hour (lambda)
lambda_rate = 5

# 2) What is the probability that exactly 3 emergency cases will arrive i

# Use the probability mass function (PMF) for Poisson distribution
probability_exactly_3 = poisson.pmf(3, lambda_rate)
```

```
print(f"Average arrival rate (lambda): {lambda_rate} cases per hour")
print(f"Probability of exactly 3 emergency cases arriving in the next hou
```

```
Average arrival rate (lambda): 5 cases per hour
Probability of exactly 3 emergency cases arriving in the next hour: 0.1404
```

3. If a major accident occurs in the city, how would you expect the distribution of administration times to change? (Note: The question here seems to reference 'administration times' which was from the previous section. Assuming the question actually refers to the 'distribution of emergency case arrival times' in the ED, based on the context of this section):

If a major accident occurs in the city, it would significantly impact the distribution of emergency case arrival times in the Emergency Department. Here's how we would expect the Poisson distribution to change:

Increase in the Average Rate (λ): The most direct impact would be a sharp increase in the average arrival rate (λ). A major accident would lead to an influx of injured individuals requiring immediate medical attention, thus increasing the number of cases arriving per hour well above the historical average of 5.

Shift in the Distribution: The Poisson distribution would shift to the right, meaning the probability of observing a higher number of arrivals in a given hour would increase. The 'peak' of the distribution (the most probable number of arrivals) would move to a higher value.

Increased Variance: The variance of a Poisson distribution is equal to its mean (λ). Therefore, an increase in λ would also lead to an increase in the variance, meaning there would be a wider spread of possible arrival numbers around the new, higher mean. This implies greater unpredictability in the exact number of arrivals, though the overall trend would be towards more cases.

Higher Probabilities for Higher Arrival Counts: The probability of events like "exactly 5 cases" or "more than 10 cases" would increase dramatically. The hospital would need to prepare for a surge capacity scenario.

In summary, a major accident would cause the Poisson distribution governing emergency arrivals to shift significantly, with a much higher average rate (λ), leading to a greater number of expected arrivals and increased variability in those arrivals.

10. The number of surgeries performed per day in the hospital follows a specific distribution pattern. Historical data shows the following frequencies:

1. Identify and justify the type of probability distribution that best fits this data.
2. Calculate the expected number of surgeries performed per day.
3. If a new surgical team is hired, how will this affect the probability distribution of daily surgeries?

1. Identify and justify the type of probability distribution that best fits this data. Given the nature of the data (counts of events, i.e., surgeries, within a fixed period, i.e., a day), the Poisson Distribution is a strong candidate.

Justification for Poisson Distribution:

Counts of Events: The variable of interest is the number of surgeries, which are discrete counts. Fixed Interval: The events occur within a fixed interval (per day). Independence: It's assumed that the occurrence of one surgery does not directly affect the probability of another surgery happening within the same day (though this is an assumption that may need verification in a real-world scenario). Average Rate (λ): A Poisson distribution is characterized by a single parameter, λ , which represents the average number of events in the given interval. We can calculate this from the historical data. To confirm if the Poisson distribution is a good fit, we often look for the mean and variance to be approximately equal, as this is a property of the Poisson distribution. We will calculate the expected number of surgeries (mean) and the variance from the provided frequency data.

2. Calculating number of surgeries per day

In [21]:
```
import numpy as np
from scipy.stats import poisson

# Data provided
surgeries = np.array([0, 1, 2, 3, 4, 5])
frequencies = np.array([5, 12, 18, 22, 15, 8])

# Calculate total number of observations (days)
total_observations = np.sum(frequencies)

# 2) Calculate the expected number of surgeries performed per day.
# Expected value (mean) = Sum(surgeries * frequency) / Total frequency
expected_surgeries = np.sum(surgeries * frequencies) / total_observations

print(f"Total observations (days): {total_observations}")
print(f"Expected number of surgeries per day (Mean): {expected_surgeries:

# Calculate the variance for comparison with the mean
# First, calculate the weighted mean squared difference from the mean
variance_surgeries = np.sum(frequencies * (surgeries - expected_surgeries

print(f"Variance of surgeries per day: {variance_surgeries:.2f}")

# Check if mean and variance are approximately equal for Poisson justific
if np.isclose(expected_surgeries, variance_surgeries, atol=1.0): # Use a
    print(f"Mean ({expected_surgeries:.2f}) and Variance ({variance_surge
else:
    print(f"Mean ({expected_surgeries:.2f}) and Variance ({variance_surge
```

```
Total observations (days): 80
Expected number of surgeries per day (Mean): 2.67
Variance of surgeries per day: 1.89
Mean (2.67) and Variance (1.89) are approximately equal, supporting the Po
isson distribution as a good fit.
```

3. If a new surgical team is hired, how will this affect the probability distribution of daily surgeries? If a new surgical team is hired, the hospital's capacity to perform surgeries increases. This would significantly impact the probability distribution of daily surgeries in several ways:

Increase in the Expected Number of Surgeries (Mean, λ): The most direct effect would be an increase in the average number of surgeries performed per day. With more staff and resources, the hospital can schedule and complete more procedures, shifting the mean (and thus the peak) of the distribution to a higher value.

Shift in the Distribution: The entire probability distribution would shift to the right. This means the probability of observing a higher number of surgeries (e.g., 4, 5, 6, or more per day) would increase, while the probability of observing fewer surgeries (e.g., 0, 1, 2) would decrease.

Increased Variance: For a Poisson distribution, the variance is equal to the mean. If the mean number of surgeries (λ) increases, the variance will also increase. This implies that there will be a wider spread of possible daily surgery counts around the new, higher average. The daily number of surgeries might become more variable, reflecting the increased capacity and potentially more complex scheduling.

Reduced Probability of Zero Surgeries: The probability of having zero surgeries would significantly decrease, as the hospital would be utilizing its increased capacity to perform more procedures.

In summary, a new surgical team would cause the probability distribution of daily surgeries to shift towards higher counts, with an increased mean and variance, reflecting the expanded operational capacity and potentially a greater range of daily surgical activity.