To analyze the residuals and test the assumptions for a linear regression model, you should follow these steps:

### Step 1: Analyze Residuals

Residuals are the differences between actual and predicted values. Plotting and analyzing residuals helps verify if the assumptions of linear regression are met.

1. \*\*Residual Plot\*\*:

- Plot residuals against the predicted values to check for patterns. Ideally, residuals should be randomly scattered around zero, indicating that the model's assumptions are satisfied.

```python

residuals\_test = y\_test - y\_pred

residuals\_val = y\_val - y\_pred\_val

# Plotting residuals for the test set

plt.figure(figsize=(12, 6))

plt.scatter(y\_pred, residuals\_test, alpha=0.6)

plt.axhline(0, color='red', linestyle='--')

plt.xlabel('Predicted Values')

plt.ylabel('Residuals')

plt.title('Residuals vs Predicted Values (Test Set)')

plt.show()

```

2. \*\*Histogram of Residuals\*\*:

- Check if residuals are normally distributed by plotting a histogram. The distribution should resemble a normal curve.

```python

plt.hist(residuals\_test, bins=30, alpha=0.7)

plt.xlabel('Residuals')

plt.ylabel('Frequency')

plt.title('Histogram of Residuals (Test Set)')

plt.show()

```

### Step 2: Test Assumptions

To ensure that the basic assumptions of linear regression are met, you need to evaluate linearity, homoscedasticity (variance uniformity), and independence.

1. \*\*Linearity Assumption\*\*:

- Check if the relationship between the independent and dependent variables is linear. If the residuals plot shows a clear pattern (e.g., curvature), this indicates a potential violation of the linearity assumption.

- You can also plot the actual vs. predicted values to confirm linearity:

```python

plt.scatter(y\_test, y\_pred, alpha=0.6)

plt.plot([y\_test.min(), y\_test.max()], [y\_test.min(), y\_test.max()], 'r--')

plt.xlabel('Actual Values')

plt.ylabel('Predicted Values')

plt.title('Actual vs Predicted Values (Test Set)')

plt.show()

```

2. \*\*Homoscedasticity (Variance Uniformity)\*\*:

- Check if the variance of the residuals is constant across all levels of predicted values. This can be assessed using the residual plot:

- If residuals spread equally around zero across the plot, the variance is uniform (homoscedasticity). If residuals form a funnel shape (wider as values increase), heteroscedasticity is present.

- Use the residuals plot from Step 1 for this check.

3. \*\*Independence of Residuals\*\*:

- Check if residuals are independent of each other, especially for time-series data. This can be evaluated using the Durbin-Watson test.

- The Durbin-Watson statistic should be around 2 for independent residuals (values close to 0 or 4 indicate autocorrelation).

```python

from statsmodels.stats.stattools import durbin\_watson

dw\_test = durbin\_watson(residuals\_test)

print('Durbin-Watson statistic:', dw\_test)

```

### Step 3: Summary of Assumptions

- \*\*Linearity\*\*: Check the scatter plot of residuals vs. predicted values for patterns.

- \*\*Homoscedasticity\*\*: Confirm no clear patterns or funnels in the residuals plot.

- \*\*Independence\*\*: Ensure the Durbin-Watson statistic is close to 2.

### Interpreting Results:

- If the assumptions are violated (e.g., non-random residuals, heteroscedasticity), consider transforming variables, adding features, or using a different model (e.g., polynomial regression or robust regression).

- If residuals are normally distributed and spread randomly, the linear model is likely appropriate for the data.