In linear regression, hypothesis testing is used to determine whether certain relationships or parameters in the model are statistically significant. The primary hypothesis tests in linear regression involve testing the significance of individual coefficients (slope parameters) and the overall significance of the regression model. Here are the key hypothesis tests:

**1. Hypothesis Test for Individual Coefficients (Slope Parameters):**

* **Test**: This test assesses whether a specific independent variable (feature) has a statistically significant effect on the dependent variable (label).
* **Null Hypothesis (H0)**: The null hypothesis states that the coefficient of the independent variable is equal to zero, implying that the variable has no effect on the dependent variable.
* **Alternative Hypothesis (HA)**: The alternative hypothesis states that the coefficient of the independent variable is not equal to zero, indicating that the variable has a significant effect on the dependent variable.
* **How to Decide**: Compute the t-statistic for each coefficient and compare it to the critical value from a t-distribution or calculate the p-value associated with the coefficient. If the p-value is smaller than a chosen significance level (e.g., 0.05), you reject the null hypothesis.
* **Interpretation**: If the null hypothesis is rejected, it suggests that the corresponding independent variable has a statistically significant impact on the dependent variable.

**2. Hypothesis Test for the Overall Model (F-Test):**

* **Test**: This test assesses whether the entire regression model is statistically significant, i.e., whether at least one of the independent variables has a non-zero coefficient.
* **Null Hypothesis (H0)**: The null hypothesis states that all coefficients of the independent variables are equal to zero, indicating that none of the variables have a significant effect on the dependent variable.
* **Alternative Hypothesis (HA)**: The alternative hypothesis states that at least one coefficient of an independent variable is not equal to zero, implying that at least one variable has a significant effect on the dependent variable.
* **How to Decide**: Calculate the F-statistic for the overall model and compare it to the critical value from an F-distribution or calculate the associated p-value. If the p-value is smaller than a chosen significance level, you reject the null hypothesis.
* **Interpretation**: If the null hypothesis is rejected, it suggests that at least one independent variable has a statistically significant impact on the dependent variable, indicating that the overall model is significant.

import statsmodels.api as sm

import numpy as np

import pandas as pd

# Generate example data

np.random.seed(0)

X = np.random.rand(100, 2)

y = 2 \* X[:, 0] + 3 \* X[:, 1] + np.random.randn(100)

# Add a constant term for the intercept

X = sm.add\_constant(X)

# Fit the linear regression model

model = sm.OLS(y, X).fit()

# Hypothesis test for individual coefficients

# For example, testing the coefficient of the first independent variable

print(model.summary())

# Hypothesis test for the overall model (F-test)

# Null hypothesis: All coefficients are zero

f\_statistic = model.fvalue

p\_value = model.f\_pvalue

if p\_value < 0.05:

print("Overall model is statistically significant.")

else:

print("Overall model is not statistically significant.")

In this demo, we generate synthetic data and fit a linear regression model using the statsmodels library. The **model.summary()** provides detailed output, including the t-statistics and p-values for individual coefficients, as well as the F-statistic and associated p-value for the overall model. You can interpret the results based on the p-values and significance levels you choose.