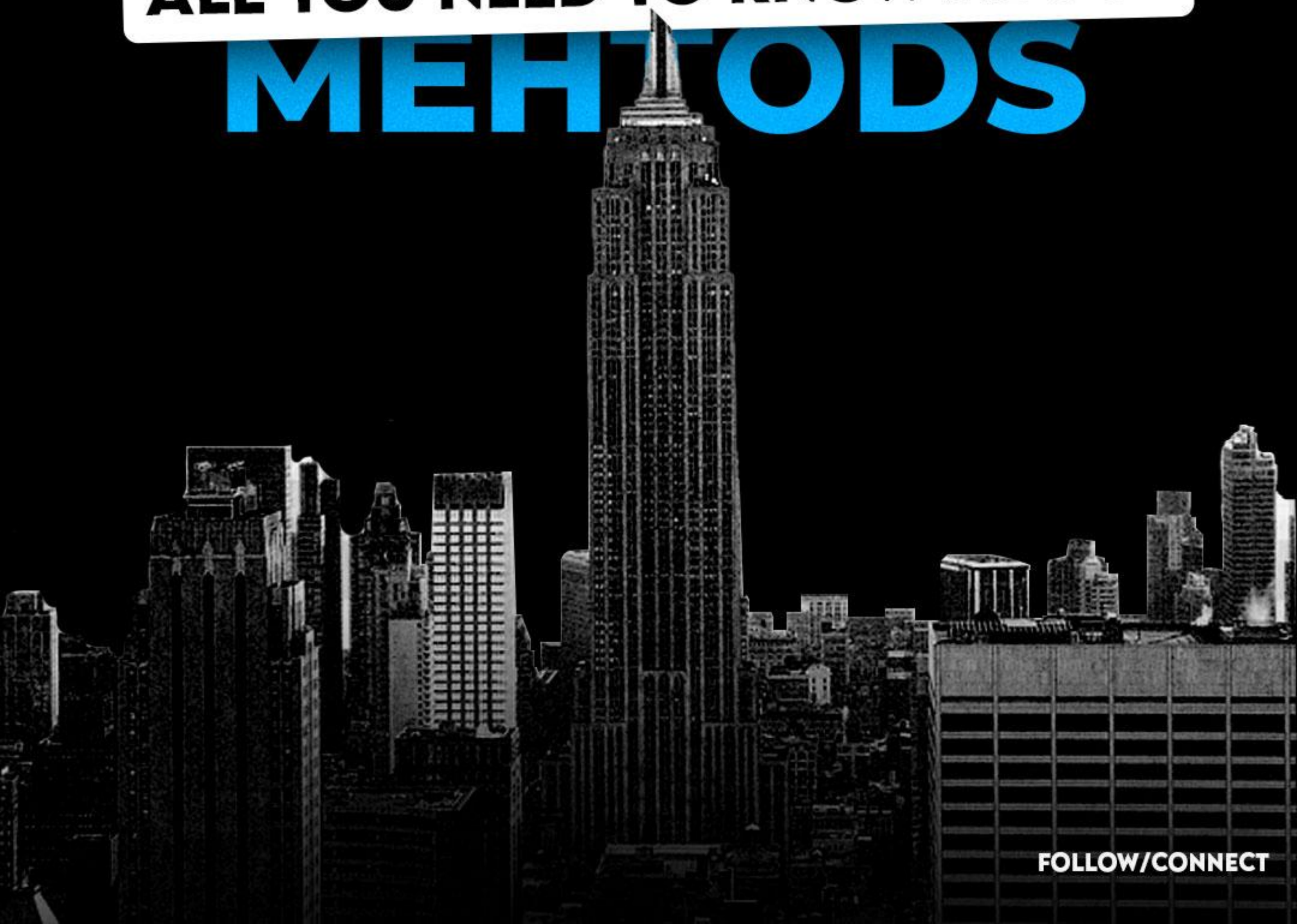
STATISTICAL METHODS

ALL YOU NEED TO KNOW ABOUT



1.

Post Hoc Tests

(Used after ANOVA to find exactly which groups differ)

python

```python
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import pandas as pd

df = pd.DataFrame({'score': [80, 85, 82, 90, 78, 84, 91, 89, 76],
                   'method': ['A', 'A', 'A', 'B', 'B', 'B', 'C', 'C', 'C']})

tukey = pairwise_tukeyhsd(df['score'], df['method'])
print(tukey)
```

Post hoc tests are crucial after a significant ANOVA result because they identify which specific groups differ from each other. While ANOVA tells you that at least one group mean is different, it doesn't show where the difference lies

2.

T-Tests

(Compare means between groups)

python

```python
from scipy.stats import ttest_1samp
import numpy as np

heights = np.array([172, 168, 174, 169, 171])
t_stat, p_val = ttest_1samp(heights, 170)
print("t-stat:", t_stat, "p-value:", p_val)
```

The t-test is important because it helps determine whether the difference between two group means is statistically significant or due to random chance.
It's widely used to compare sample data against a known value or between two groups

FOLLOW/CONNECT

3.

Correlation Tests

(Measure relationship strength and direction between variables)

python

```python
from scipy.stats import pearsonr

study_hours = [2, 4, 6, 8, 10]
scores = [55, 60, 65, 70, 80]

corr, p_val = pearsonr(study_hours, scores)
print("Correlation Coefficient:", corr, "p-value:", p_val)
```

Correlation tests are important because they measure the strength and direction of the relationship between two variables. They help identify patterns, such as whether an increase in one variable is associated with an increase or decrease in another

FOLLOW/CONNECT

4.

Column Relationship Tests

(To check dependency or association between columns)

python

```python
import pandas as pd
from scipy.stats import chi2_contingency

data = pd.DataFrame({
    'Product_A': [20, 15],
    'Product_B': [30, 35]
}, index=['Male', 'Female'])

chi2, p, dof, expected = chi2_contingency(data)
print("Chi-squared:", chi2, "p-value:", p)
```

Correlation tests are important because they measure the strength and direction of the relationship between two variables. They help identify patterns, such as whether an increase in one variable is associated with an increase or decrease in another

FOLLOW FOR

MORE

FOLLOW/CONNECT