

Statistical Analysis Quick Reference Guide

For Data Science with Python

1. Common Statistical Tests & Their Uses

Test Type	When to Use	Python Code	Key Assumptions
T-Test	Compare means of 2 groups	<code>stats.ttest_ind(group1, group2)</code>	<ul style="list-style-type: none">- Normal distribution- Equal variance- Independent samples
Chi-Square	Compare categorical data	<code>stats.chi2_contingency(table)</code>	<ul style="list-style-type: none">- Independent observations- Large enough sample size
ANOVA	Compare 3+ groups	<code>stats.f_oneway(group1, group2, group3)</code>	<ul style="list-style-type: none">- Normal distribution- Equal variance
Mann-Whitney U	Non-parametric comparison	<code>stats.mannwhitneyu(group1, group2)</code>	<ul style="list-style-type: none">- Independent observations- Similar shaped distributions

2. Testing Assumptions

Normality Check

```
Unset
# Visual check
sns.histplot(data)
stats.probplot(data, plot=plt)

# Statistical test
stats.normaltest(data) #  $p > 0.05$  suggests normality
```

Equal Variance Check

```
Unset
stats.levene(group1, group2) #  $p > 0.05$  suggests equal variance
```

3. Effect Size Measures

Measure	Use Case	Interpretation
Cohen's d	T-test	0.2 = small
		0.5 = medium
		0.8 = large
Cramer's V	Chi-square	0.1 = small
		0.3 = medium
		0.5 = large

4. Decision Flowchart

```
Unset
Start → Is data normal? (stats.normaltest)
↓
```

YES → Equal variance? (stats.levene)

↓

YES

↓

T-test

↓

NO

↓

Welch's t-test

NO → Mann-Whitney U test

5. Common Pitfalls to Avoid

- Using t-test without checking normality when the sample size is small
- Ignoring sample size requirements
- Drawing causation from correlation

6. Key Python Libraries

Unset

```
from scipy import stats          # Statistical tests
import numpy as np              # Numerical operations
import seaborn as sns           # Statistical visualization
import matplotlib.pyplot as plt # Plotting
```

7. Basic Statistical Measures

Unset

```
# Basic statistics
mean = np.mean(data)
median = np.median(data)
std = np.std(data)
quantiles = np.percentile(data, [25, 50, 75])

# Quick summary
df.describe()
```

8. Visualization Best Practices

Plot Type	Use Case	Python Code
Box Plot	Distribution comparison	<code>sns.boxplot(x='group', y='value', data=df)</code>
Histogram	Distribution shape	<code>sns.histplot(data=df['value'])</code>
Q-Q Plot	Normality check	<code>stats.probplot(data, plot=plt)</code>
Bar Plot	Category comparison	<code>sns.barplot(x='group', y='value', data=df)</code>

9. Reporting Results

Template:

Unset

```
Statistical test: [Test Name]
Test statistic: [Value]
P-value: [Value]
Effect size: [Value]
Interpretation: [Brief explanation]
```

10. Quick Checklist Before Analysis

- Data cleaned and formatted correctly
- Assumptions checked and documented
- Sample size adequate
- Appropriate test selected
- Effect size calculated
- Results properly interpreted

Note: This is a reference guide. Always consult documentation for detailed implementation.