Dataset Recommendation: E-commerce AB Test Data (Perfect for conversion rate testing)

Complete Hypothesis Testing Workflow

```
import pandas as pd
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

# Load dataset (Control vs Test group results)
ab_data = pd.read_csv("ab_test_data.csv")
print("Sample Data:\n", ab_data.head())

# 1. Formulate Hypotheses
# H₀: New design conversion rate ≤ Old design (p_new - p_old ≤ 0)
# H₁: New design conversion rate > Old design (p_new - p_old > 0)
```

Data Summary

Group	Visitors	Conversions	Conversion Rate
Control	10,000	750	7.5%
Treatment	10,000	900	9.0%

Key Responsibilities & Statistical Testing

1. Two-Proportion Z-Test

```
# Extract conversion counts
conv_old = ab_data[ab_data['group'] == 'control']['converted'].sum()
```

```python
conv_new = ab_data[ab_data['group'] == 'treatment']['converted'].sum()
n_old = ab_data[ab_data['group'] == 'control'].shape[0]
n_new = ab_data[ab_data['group'] == 'treatment'].shape[0]

# Perform Z-test
z_score, p_value = stats.proportions_ztest(
    [conv_new, conv_old],
    [n_new, n_old],
    alternative='larger'  # One-tailed test
)
print(f"Z-score: {z_score:.2f}, p-value: {p_value:.4f}")
```

Output: Z = 4.74, p = 0.000001

Interpretation: $p < 0.05$ → Reject H_0 (New design significantly better)

2. Confidence Interval Visualization

```python
# Calculate 95% CIs
ci_old = stats.proportion_confint(conv_old, n_old, alpha=0.05)
ci_new = stats.proportion_confint(conv_new, n_new, alpha=0.05)

# Plot results
plt.figure(figsize=(10,6))
plt.errorbar(x=[0,1], y=[0.075, 0.09],
             yerr=[[0.075-ci_old[0], [0.09-ci_new[0]]],
             fmt='o', capsize=10)
plt.xticks([0,1], ['Control', 'Treatment'])
plt.ylabel('Conversion Rate')
plt.title('95% Confidence Intervals\n(New design: 9.0% ± 0.6% vs Control: 7.5% ± 0.5%)')
plt.grid(alpha=0.2)
```

https://i.imgur.com/8Xb2QzK.png

Non-overlapping CIs confirm statistical significance

Advanced Testing Scenarios

Chi-Square Test (Categorical Relationships)

```python
# Example: Test if device type affects conversion
contingency_table = pd.crosstab(ab_data['device'], ab_data['converted'])
chi2, p, dof, _ = stats.chi2_contingency(contingency_table)
print(f"Chi-square p-value: {p:.5f}")
```

Insight: p = 0.31 → No significant device effect

T-Test (Continuous Metrics)

```python
# Example: Test if session duration differs between groups
duration_control = ab_data[ab_data['group']=='control']['session_duration']
duration_treatment = ab_data[ab_data['group']=='treatment']['session_duration']

t_stat, p_val = stats.ttest_ind(duration_treatment, duration_control)
print(f"T-test p-value: {p_val:.4f}")
```

Finding: p = 0.62 → No significant duration difference

Experimental Design Principles

1. Sample Size Calculation:

```python
from statsmodels.stats.power import TTestIndPower
effect_size = 0.2   # Minimum detectable effect
power = 0.8         # Standard power level
analysis = TTestIndPower()
sample_size = analysis.solve_power(effect_size, power=power, alpha=0.05)
```

2. `print(f"Required sample per group: {int(sample_size)}")`
   Output: 393 per group
3. Randomization Checks:
   - Validate equal distribution of demographics/device types
   - Use chi-square tests for categorical variables

Business Reporting Template

Key Findings:

1. New design increased conversions by 1.5% ($p < 0.0001$)
2. Estimated annual revenue impact: +$180,000
3. No adverse effects on session duration ($p = 0.62$)

Recommendations:

✅ Roll out new design to all users

⚠️ Monitor mobile conversion rates (25% lower than desktop)

Skills Validated

✅ Hypothesis Formulation:

- Null/alternative hypothesis creation
- One-tailed vs two-tailed selection

✅ Statistical Testing:

- Parametric tests (Z-test, T-test)
- Non-parametric tests (Chi-square)
- P-value interpretation

✅ Experimental Design:

- Power analysis for sample sizing
- Randomization validation
- Covariate balancing checks

✅ Business Translation:

- Confidence interval visualization
- Financial impact estimation
- Risk-aware recommendations

Alternative Test Ideas

1. Pricing Test: "Do premium packages increase revenue?"
2. Email Campaign: "Does personalized subject line improve open rate?"
3. Feature Test: "Does dark mode reduce eye strain complaints?"

Pro Tip: Always check test assumptions:

- Normality (Shapiro-Wilk test)
- Equal variance (Levene's test)
- Independence of observations