PRACTICAL COURSE ON

Python with Statistics

AFTERNOON SESSION HANDBOOK

Part 2: Hypothesis Testing in Action

⏰ Time: 15:15 - 16:45 (90 minutes)

Part 2: Hypothesis Testing in Action

What Is Hypothesis Testing?

In Part 1, we observed patterns: first-class passengers had higher survival rates, women survived more than men, etc. But here's the critical question:

Are these patterns REAL, or could they have occurred by random chance?

Hypothesis testing provides a mathematical framework to answer this question. It tells us whether observed differences are statistically significant or just noise.

The Logic of Hypothesis Testing

Every hypothesis test follows the same 5-step process:

1. State hypotheses: H_0 (null): No difference/effect. H_1 (alternative): There IS a difference/effect

2. Choose significance level: Usually $\alpha = 0.05$ (5% chance of being wrong)

3. Calculate test statistic: A number that measures how different the groups are

4. Find p-value: Probability of seeing this difference if H_0 were true

5. Make decision: If $p < 0.05$ → Reject H_0 (significant). If $p \geq 0.05$ → Fail to reject H_0 (not significant)

💡 **TIP:** p-value interpretation:
- $p < 0.05$: Strong evidence against H_0 → Reject H_0 → Difference is significant
- $p \geq 0.05$: Weak evidence against H_0 → Fail to reject H_0 → Difference not significant

Think of p-value as 'probability this happened by luck.' If $p = 0.02$, there's only a 2% chance the difference is random—so we conclude it's real.

Which Test Should I Use?

Different research questions require different statistical tests. Here's a decision guide:

Question Type	Data Condition	Test to Use
Comparing 2 groups (numeric data)	Normally distributed	Independent t-test
Comparing 2 groups (numeric data)	NOT normally distributed	Mann-Whitney U test
Comparing 2+ groups (numeric data)	Normally distributed	ANOVA (not covered today)
Comparing categorical variables	Any distribution	Chi-square test
Testing a proportion	Binary outcome	Binomial test

Today we'll cover:

- Independent t-test (parametric)
- Mann-Whitney U test (non-parametric alternative to t-test)
- Chi-square test (categorical relationships)
- Binomial test (proportion testing)

Test 1: Independent t-test

When to Use

Use an independent t-test when:

- Comparing TWO independent groups
- Outcome variable is numeric (continuous)
- Data is approximately normally distributed
- Example: Do males and females have different average ages?

Research Question: Age Difference by Gender

🎯 **RESEARCH QUESTION**
Is there a statistically significant difference in average age between male and female passengers on the Titanic?

Step 1: Visual Inspection

📋 **Visualize the comparison:**

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from scipy import stats

# Load data
df = sns.load_dataset('titanic')

# Remove missing ages
df_clean = df.dropna(subset=['age'])

# Side-by-side box plots
plt.figure(figsize=(10, 6))
sns.boxplot(data=df_clean, x='sex', y='age', palette=['lightblue', 'pink'])
plt.title('Age Distribution by Gender', fontsize=14, fontweight='bold')
plt.xlabel('Gender', fontsize=12)
plt.ylabel('Age (years)', fontsize=12)
plt.show()

# Overlapping histograms
plt.figure(figsize=(10, 6))
sns.histplot(data=df_clean[df_clean['sex']=='male'], x='age',
             color='blue', label='Male', kde=True, alpha=0.5, bins=30)
sns.histplot(data=df_clean[df_clean['sex']=='female'], x='age',
             color='pink', label='Female', kde=True, alpha=0.5, bins=30)
plt.title('Age Distribution by Gender (Overlapping)', fontsize=14,
fontweight='bold')
plt.xlabel('Age (years)', fontsize=12)
plt.ylabel('Frequency', fontsize=12)
plt.legend()
plt.show()
```

Visual observation: The distributions look similar, with males perhaps slightly older on average. But is this difference significant? Let's test it.

Step 2: Calculate Descriptive Statistics

📋 **Compare groups:**

```
# Separate by gender
male_ages = df_clean[df_clean['sex'] == 'male']['age']
female_ages = df_clean[df_clean['sex'] == 'female']['age']

# Calculate statistics
print(f"Male passengers:")
print(f"  Count: {len(male_ages)}")
print(f"  Mean age: {male_ages.mean():.2f} years")
print(f"  Std dev: {male_ages.std():.2f} years")

print(f"\nFemale passengers:")
print(f"  Count: {len(female_ages)}")
print(f"  Mean age: {female_ages.mean():.2f} years")
print(f"  Std dev: {female_ages.std():.2f} years")

print(f"\nDifference in means: {male_ages.mean() - female_ages.mean():.2f}
years")
```

Output:

```
Male passengers:
  Count: 453
  Mean age: 30.73 years
  Std dev: 14.68 years

Female passengers:
  Count: 261
  Mean age: 27.92 years
  Std dev: 14.11 years

Difference in means: 2.81 years
```

Step 3: State Hypotheses

H_0 (Null hypothesis): There is NO difference in average age between males and females (μ_male = μ_female)

H_1 (Alternative hypothesis): There IS a difference in average age between males and females (μ_male ≠ μ_female)

Significance level: $\alpha = 0.05$

Step 4: Run the t-test

📋 **Conduct the test:**

```
# Perform independent t-test
t_statistic, p_value = stats.ttest_ind(male_ages, female_ages)

print(f"t-statistic: {t_statistic:.4f}")
print(f"p-value: {p_value:.4f}")

# Interpret
if p_value < 0.05:
    print("\nConclusion: REJECT the null hypothesis")
    print("The difference in age between males and females IS statistically
significant (p < 0.05)")
else:
```

```
    print("\nConclusion: FAIL TO REJECT the null hypothesis")
    print("The difference in age between males and females is NOT
statistically significant (p ≥ 0.05)")
```

Output:

```
t-statistic: 2.4937
p-value: 0.0129

Conclusion: REJECT the null hypothesis
The difference in age between males and females IS statistically significant
(p < 0.05)
```

📊 **STATISTICAL CONCLUSION**

CONCLUSION:
Male passengers were significantly older than female passengers ($p = 0.013$). The 2.81-year difference is small but statistically significant.

In plain language: This age difference is unlikely to have occurred by random chance. Male passengers were genuinely older on average.

💡 **TIP:** How to report this in a paper:
"Male passengers ($M = 30.73$, $SD = 14.68$) were significantly older than female passengers ($M = 27.92$, $SD = 14.11$), $t(712) = 2.49$, $p = 0.013$."

✓ CHECKPOINT EXERCISE 4

Research Question: Did first-class passengers pay significantly more than third-class?

1. Filter for first-class and third-class passengers only
2. Create box plots comparing 'fare' between the two classes
3. Run an independent t-test
4. Interpret: Is the difference significant?

💡 *Hint:* first_class = df[df['pclass']==1]['fare'], third_class = df[df['pclass']==3]['fare']

Test 2: Mann-Whitney U Test (Non-Parametric)

When to Use

Use Mann-Whitney U test when:

- Comparing TWO independent groups
- Outcome variable is numeric OR ordinal (ranked)
- Data is NOT normally distributed (skewed, has outliers)
- Alternative to t-test when assumptions are violated
- Example: Comparing fare between survivors and non-survivors (fare is highly skewed)

🧭 **WHICH TEST TO USE?**

Why use Mann-Whitney instead of t-test?

The t-test assumes normal distribution. When data is heavily skewed (like fare), the t-test may give misleading results. Mann-Whitney U test works on RANKS instead of raw values, making it robust to outliers and skewness.

Research Question: Fare Difference by Survival

🎯 **RESEARCH QUESTION**

Did survivors pay significantly different fares than those who died? This could reveal whether ticket price (a proxy for class) affected survival.

Step 1: Check Distribution

📋 **Assess distribution:**

```
# Visualize fare distribution
plt.figure(figsize=(10, 6))
sns.histplot(data=df, x='fare', bins=50, kde=True)
plt.title('Distribution of Fare (Highly Right-Skewed)', fontsize=14,
fontweight='bold')
plt.xlabel('Fare (£)', fontsize=12)
plt.ylabel('Frequency', fontsize=12)
plt.show()

# Check skewness
print(f"Fare skewness: {df['fare'].skew():.2f}")
print("\nConclusion: Fare is highly right-skewed (skewness > 1).")
print("Therefore, use Mann-Whitney U test instead of t-test.")
```

Output:

```
Fare skewness: 4.79

Conclusion: Fare is highly right-skewed (skewness > 1).
Therefore, use Mann-Whitney U test instead of t-test.
```

Step 2: Visualize by Survival

📋 **Visualize comparison:**

```
# Box plot comparison
plt.figure(figsize=(10, 6))
```

```
sns.boxplot(data=df, x='alive', y='fare', palette=['red', 'green'])
plt.title('Fare Paid: Survivors vs Non-Survivors', fontsize=14,
fontweight='bold')
plt.xlabel('Survived', fontsize=12)
plt.ylabel('Fare (£)', fontsize=12)
plt.show()
```

Step 3: Calculate Descriptive Statistics

📋 **Compare medians:**

```
# Separate groups
survived = df[df['survived'] == 1]['fare']
died = df[df['survived'] == 0]['fare']

print(f"Survivors:")
print(f"  Count: {len(survived)}")
print(f"  Median fare: £{survived.median():.2f}")
print(f"  Mean fare: £{survived.mean():.2f}")

print(f"\nNon-survivors:")
print(f"  Count: {len(died)}")
print(f"  Median fare: £{died.median():.2f}")
print(f"  Mean fare: £{died.mean():.2f}")
```

Output:

```
Survivors:
  Count: 342
  Median fare: £26.00
  Mean fare: £48.40

Non-survivors:
  Count: 549
  Median fare: £10.50
  Mean fare: £22.12
```

Observation: Survivors paid nearly 2.5x higher median fare. This suggests wealthier passengers (higher-class cabins) had better survival chances.

Step 4: State Hypotheses

H_0 (Null hypothesis): There is NO difference in fare between survivors and non-survivors

H_1 (Alternative hypothesis): There IS a difference in fare between survivors and non-survivors

Step 5: Run Mann-Whitney U Test

📋 **Conduct the test:**

```
# Perform Mann-Whitney U test
u_statistic, p_value = stats.mannwhitneyu(survived, died, alternative='two-
sided')

print(f"U-statistic: {u_statistic:.2f}")
print(f"p-value: {p_value:.6f}")

# Interpret
if p_value < 0.05:
```

```
    print("\nConclusion: REJECT the null hypothesis")
    print("Survivors paid significantly different fares than non-survivors (p
< 0.05)")
else:
    print("\nConclusion: FAIL TO REJECT the null hypothesis")
    print("No significant difference in fares (p ≥ 0.05)")
```

Output:

```
U-statistic: 120849.00
p-value: 0.000000

Conclusion: REJECT the null hypothesis
Survivors paid significantly different fares than non-survivors (p < 0.05)
```

📊 **STATISTICAL CONCLUSION**

CONCLUSION:
Survivors paid significantly higher fares than non-survivors ($p < 0.001$). This makes sense: higher fares meant better cabins (first/second class), which were located on upper decks closer to lifeboats.

In plain language: Wealthier passengers had better survival chances, likely due to cabin location and priority access to lifeboats.

💡 **TIP:** When to use Mann-Whitney vs t-test:
- Data is roughly bell-shaped → Use t-test
- Data is skewed, has outliers, or ordinal → Use Mann-Whitney U
- Mann-Whitney is the 'safe choice' when in doubt

✓ CHECKPOINT EXERCISE 5

Research Question: Did age differ between survivors and non-survivors?

1. Separate age data by survival status
2. Create box plots comparing age
3. Check if age is normally distributed (optional: use histograms)
4. Run Mann-Whitney U test
5. Interpret the result

💡 *Hint:* survived_age = df[df['survived']==1]['age'].dropna(), died_age = df[df['survived']==0]['age'].dropna()

Test 3: Chi-Square Test of Independence

When to Use

Use Chi-square test when:

- Comparing TWO categorical variables
- Want to know if they are associated/related
- Example: Is survival related to passenger class?
- Example: Is smoking status related to disease occurrence?

Research Question: Survival and Passenger Class

🎯 **RESEARCH QUESTION**
Is there a statistically significant association between passenger class (First/Second/Third) and survival (Yes/No)?

Step 1: Create Contingency Table

📋 **Build contingency table:**

```
# Create contingency table (cross-tabulation)
contingency_table = pd.crosstab(df['class'], df['survived'])
contingency_table.columns = ['Died', 'Survived']

print("Contingency Table:")
print(contingency_table)

# Add row totals
contingency_table['Total'] = contingency_table.sum(axis=1)

# Add column totals
totals_row = pd.DataFrame(contingency_table.sum(axis=0)).T
totals_row.index = ['Total']
contingency_table = pd.concat([contingency_table, totals_row])

print("\nWith totals:")
print(contingency_table)
```

Output:

```
Contingency Table:
        Died  Survived
class
First     80       136
Second    97        87
Third    372       119

With totals:
        Died  Survived  Total
class
First     80       136    216
Second    97        87    184
Third    372       119    491
Total    549       342    891
```

Step 2: Calculate Survival Rates

📋 **Calculate percentages:**

```
# Calculate percentage survived per class
survival_rate = pd.crosstab(df['class'], df['survived'], normalize='index') *
100
survival_rate.columns = ['Died %', 'Survived %']

print("Survival rates by class:")
print(survival_rate)
```

Output:

```
Survival rates by class:
        Died %  Survived %
class
First   37.04       62.96
Second  52.72       47.28
Third   75.76       24.24
```

Observation: First class: 63% survived. Third class: Only 24% survived. Clear pattern—but is it statistically significant?

Step 3: Visualize the Association

📋 **Visualize association:**

```
# Stacked bar chart
survival_rate.plot(kind='bar', stacked=True, color=['red', 'green'],
                   figsize=(10, 6), edgecolor='black')
plt.title('Survival Rate by Passenger Class', fontsize=14, fontweight='bold')
plt.xlabel('Passenger Class', fontsize=12)
plt.ylabel('Percentage', fontsize=12)
plt.xticks(rotation=0)
plt.legend(title='Outcome')
plt.show()

# Heatmap of counts
plt.figure(figsize=(8, 6))
ct = pd.crosstab(df['class'], df['survived'])
ct.columns = ['Died', 'Survived']
sns.heatmap(ct, annot=True, fmt='d', cmap='YlOrRd', cbar_kws={'label':
'Count'})
plt.title('Survival by Class (Heatmap)', fontsize=14, fontweight='bold')
plt.ylabel('Class', fontsize=12)
plt.xlabel('Outcome', fontsize=12)
plt.show()
```

Step 4: State Hypotheses

H_0 (Null hypothesis): Survival and passenger class are INDEPENDENT (no association)

H_1 (Alternative hypothesis): Survival and passenger class are ASSOCIATED (dependent)

Step 5: Run Chi-Square Test

📋 **Conduct chi-square test:**

```
# Prepare contingency table (without totals)
ct = pd.crosstab(df['class'], df['survived'])

# Perform chi-square test
```

```
chi2, p_value, dof, expected = stats.chi2_contingency(ct)

print(f"Chi-square statistic: {chi2:.4f}")
print(f"p-value: {p_value:.6f}")
print(f"Degrees of freedom: {dof}")

print("\nExpected frequencies (if independent):")
print(expected)

# Interpret
if p_value < 0.05:
    print("\nConclusion: REJECT the null hypothesis")
    print("Passenger class and survival ARE significantly associated (p < 
0.05)")
else:
    print("\nConclusion: FAIL TO REJECT the null hypothesis")
    print("No significant association between class and survival (p ≥ 0.05)")
```

Output:

```
Chi-square statistic: 102.8890
p-value: 0.000000
Degrees of freedom: 2

Expected frequencies (if independent):
[[133.09090909  82.90909091]
 [113.37373737  70.62626263]
 [302.53535354 188.46464646]]

Conclusion: REJECT the null hypothesis
Passenger class and survival ARE significantly associated (p < 0.05)
```

📊 **STATISTICAL CONCLUSION**

CONCLUSION:

There is a statistically significant association between passenger class and survival ($\chi^2 = 102.89$, $p < 0.001$).

In plain language: Your social class dramatically affected your chances of survival. First-class passengers were 2.6 times more likely to survive than third-class passengers. This is not random—it reflects systematic differences in cabin location, lifeboat access, and rescue priority.

💡 **TIP:** How to report chi-square results:

"A chi-square test of independence revealed a significant association between passenger class and survival, $\chi^2(2) = 102.89$, $p < 0.001$. First-class passengers had a 63% survival rate compared to 24% for third-class passengers."

✓ CHECKPOINT EXERCISE 6

Research Question: Is survival associated with gender?

1. Create a contingency table for 'sex' and 'survived'
2. Calculate survival percentages by gender
3. Visualize with a stacked bar chart
4. Run chi-square test
5. Interpret: Is the association significant?

💡 *Hint:* pd.crosstab(df['sex'], df['survived'])

Test 4: Binomial Test

When to Use

Use Binomial test when:

- Testing a PROPORTION (percentage)
- Two possible outcomes (success/failure, yes/no)
- Want to compare observed proportion to an expected/theoretical proportion
- Example: Is survival rate different from 50%?
- Example: Did 'women and children first' policy hold true?

Research Question: 'Women and Children First' Policy

🎯 **RESEARCH QUESTION**

The maritime policy was 'women and children first' in lifeboats. If this policy were perfectly followed, 100% of women and children should survive. Let's test if the observed survival rate for women and children is significantly higher than the overall rate (38.38%).

Step 1: Calculate Observed Proportion

📋 **Calculate proportions:**

```
# Filter for women and children (age < 18)
women_children = df[(df['sex'] == 'female') | (df['age'] < 18)]

# Calculate survival stats
total_wc = len(women_children)
survived_wc = women_children['survived'].sum()
survival_rate_wc = (survived_wc / total_wc) * 100

print(f"Women and children aboard: {total_wc}")
print(f"Women and children survived: {survived_wc}")
print(f"Survival rate: {survival_rate_wc:.2f}%")

# Compare to overall rate
overall_rate = (df['survived'].sum() / len(df)) * 100
print(f"\nOverall survival rate: {overall_rate:.2f}%")
print(f"Difference: {survival_rate_wc - overall_rate:.2f} percentage points")
```

Output:

```
Women and children aboard: 417
Women and children survived: 286
Survival rate: 68.59%

Overall survival rate: 38.38%
Difference: 30.21 percentage points
```

Step 2: Visualize the Comparison

📋 **Visualize comparison:**

```
# Bar chart comparing survival rates
categories = ['Women & Children', 'Overall Population']
rates = [survival_rate_wc, overall_rate]

plt.figure(figsize=(8, 6))
```

```
bars = plt.bar(categories, rates, color=['pink', 'gray'], edgecolor='black',
linewidth=1.5)
plt.title('Survival Rate: Women & Children vs Overall', fontsize=14,
fontweight='bold')
plt.ylabel('Survival Rate (%)', fontsize=12)
plt.ylim(0, 100)

# Add percentage labels
for bar, rate in zip(bars, rates):
    height = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2., height + 2,
             f'{rate:.1f}%', ha='center', fontsize=12, fontweight='bold')

plt.axhline(y=50, color='red', linestyle='--', linewidth=2, label='50%
benchmark')
plt.legend()
plt.show()
```

Step 3: State Hypotheses

H_0 (Null hypothesis): Women & children survival rate = overall rate (38.38%)

H_1 (Alternative hypothesis): Women & children survival rate > overall rate (one-sided test)

Step 4: Run Binomial Test

📋 **Conduct binomial test:**

```
# Binomial test
# H0: probability of survival = overall rate (0.3838)
# H1: probability of survival > overall rate

from scipy.stats import binomtest

# Test if women & children survival rate > overall rate
result = binomtest(survived_wc, total_wc, overall_rate/100,
alternative='greater')

print(f"Test statistic (successes): {survived_wc}")
print(f"Total trials: {total_wc}")
print(f"Expected probability under H0: {overall_rate/100:.4f}")
print(f"p-value: {result.pvalue:.6f}")

# Interpret
if result.pvalue < 0.05:
    print("\nConclusion: REJECT the null hypothesis")
    print("Women and children had a significantly HIGHER survival rate than
the overall population (p < 0.05)")
else:
    print("\nConclusion: FAIL TO REJECT the null hypothesis")
    print("No significant difference in survival rates (p ≥ 0.05)")
```

Output:

```
Test statistic (successes): 286
Total trials: 417
Expected probability under H0: 0.3838
p-value: 0.000000

Conclusion: REJECT the null hypothesis
Women and children had a significantly HIGHER survival rate than the overall
population (p < 0.05)
```

📊 **STATISTICAL CONCLUSION**

CONCLUSION:
Women and children had a significantly higher survival rate (68.59%) compared to the overall rate (38.38%), p < 0.001.

In plain language: The 'women and children first' policy was genuinely implemented, though not perfectly (68.59% survived, not 100%). Women and children were prioritized for lifeboat access, resulting in dramatically better survival outcomes.

💡 **TIP:** One-sided vs Two-sided tests:
• Two-sided: Testing if proportion is DIFFERENT (higher OR lower)
• One-sided: Testing if proportion is specifically HIGHER (or specifically LOWER)

Here we used one-sided because we had a directional hypothesis: women & children should have HIGHER (not just different) survival rates.

✓ CHECKPOINT EXERCISE 7

Research Question: Did first-class passengers have better survival than chance (50%)?

1. Filter for first-class passengers only
2. Calculate survival count and total count
3. Run a binomial test comparing to 50% (random chance)
4. Interpret: Did first-class have significantly better than 50/50 odds?

💡 *Hint:* Use binomtest(successes, trials, 0.50, alternative='greater')