

Testes de Hipotese

Os testes estatísticos são ferramentas fundamentais na análise de dados, permitindo-nos tirar conclusões sobre a população a partir de amostras. Vamos explorar alguns dos principais testes, seus pressupostos e objetivos.

Testes de Normalidade

Esses testes verificam se os dados seguem uma distribuição normal.

Shapiro-Wilk Test

- **Função:** `shapiro(a)`
- **Hipótese Nula (H0):** A amostra vem de uma distribuição normal.
- **Uso:** É amplamente utilizado para pequenas amostras ($n < 50$).

```
from scipy.stats import shapiro

data = [1.2, 2.3, 3.1, 4.4, 5.0]
stat, p = shapiro(data)
print(f"Statistics={stat}, p-value={p}")
```

Testes de Homogeneidade de Variâncias

Teste de Levene

- **Função:** `levene(a, b)`
- **Hipótese Nula (H0):** As variâncias das amostras são iguais.
- **Uso:** Verifica se diferentes grupos têm variâncias semelhantes, um pressuposto importante para muitos testes paramétricos.

```
from scipy.stats import levene

data1 = [1.2, 2.3, 3.1, 4.4, 5.0]
data2 = [2.1, 2.2, 2.9, 3.8, 4.0]
stat, p = levene(data1, data2)
print(f"Statistics={stat}, p-value={p}")
```

Testes Paramétricos

Estes testes pressupõem que os dados seguem uma distribuição específica (geralmente normal) e são usados principalmente para comparar médias.

Teste t de Student

- **Função:** `ttest_ind(a, b, alternative="two-sided"|"greater"|"less")`
- **Distribuição:** t-Student
- **Uso:** Compara as médias de duas amostras independentes.

```
from scipy.stats import ttest_ind

data1 = [1.2, 2.3, 3.1, 4.4, 5.0]
data2 = [2.1, 2.2, 2.9, 3.8, 4.0]
stat, p = ttest_ind(data1, data2, alternative="two-sided")
print(f"Statistics={stat}, p-value={p}")
```

Teste z

- **Função:** `ztest(a, b, alternative="two-sided"|"greater"|"less")`
- **Distribuição:** Normal
- **Uso:** Compara as médias de duas populações conhecidas, útil para grandes amostras.

Teste t de Uma Amostra

- **Função:** `ttest_1samp(a, valor, alternative="two-sided"|"greater"|"less")`
- **Uso:** Compara a média da amostra com um valor conhecido.

Teste t Pareado

- **Função:** `ttest_rel(a, b, alternative="two-sided"|"greater"|"less")`
- **Uso:** Compara médias de duas amostras dependentes (ex.: medidas antes e depois do tratamento).

Teste t de Welch

- **Função:** `ttest_ind(a, b, equal_var=False)`
- **Uso:** Compara médias de duas amostras com variâncias ou tamanhos diferentes.

Testes Não Paramétricos

Esses testes não pressupõem nenhuma distribuição específica dos dados e são usados para comparar medianas e distribuições.

Teste de Wilcoxon

- **Função:** `wilcoxon(a, b, alternative="two-sided"|"greater"|"less")`
- **Uso:** Compara as medianas de duas amostras pareadas.

```
pythonCopiar código
from scipy.stats import wilcoxon

data1 = [1.2, 2.3, 3.1, 4.4, 5.0]
data2 = [1.1, 2.4, 3.0, 4.5, 5.1]
stat, p = wilcoxon(data1, data2, alternative="two-sided")
print(f"Statistics={stat}, p-value={p}")
```

Teste Mann-Whitney U

- **Função:** `mannwhitneyu(a, b, alternative="two-sided"|"greater"|"less")`
- **Uso:** Compara distribuições de duas amostras independentes, especialmente quando os tamanhos das amostras são diferentes.

Teste Qui-Quadrado (χ^2)

Esses testes são usados para variáveis categóricas, verificando a independência entre variáveis ou comparando distribuições observadas com distribuições esperadas.

Teste Qui-Quadrado de Independência

- **Função:** `chi2_contingency(contingency_table)`
- **Hipótese Nula (H0):** As variáveis são independentes.
- **Uso:** Verifica se há uma associação significativa entre duas variáveis categóricas.

Teste t de Student

Teste Wilcoxon

Teste Qui-Quadrado

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

data = {'Category1': [10, 20, 30], 'Category2': [15, 25, 35]}
df = pd.DataFrame(data)
stat, p, dof, expected = chi2_contingency(df)
print(f"Statistics={stat}, p-value={p}")
```

Conclusão

A escolha do teste estatístico apropriado depende dos pressupostos sobre os dados e do tipo de comparação que se deseja fazer. Testes paramétricos requerem suposições mais rígidas sobre a distribuição dos dados, enquanto testes não paramétricos são mais flexíveis, mas podem ser menos poderosos. Testes de normalidade e homogeneidade de variâncias ajudam a determinar se os dados atendem aos pressupostos dos testes paramétricos.

O parâmetro `alternative` em muitos testes estatísticos especifica a hipótese alternativa que está sendo testada. A hipótese alternativa é a afirmação que queremos testar contra a hipótese nula (H_0). Dependendo do contexto do teste, a hipótese alternativa pode ser formulada de diferentes maneiras. Em particular, `alternative` pode assumir um dos seguintes valores:

1. **"two-sided"** (ou "bilateral")
2. **"greater"** (ou "unilateral à direita")
3. **"less"** (ou "unilateral à esquerda")

Significado dos Valores de `alternative`

1. **"two-sided"**

- **Interpretação:** A hipótese alternativa é que a estatística de teste (por exemplo, a média) é diferente de um valor específico (ou que as duas populações são diferentes).
- **H_0 (Hipótese Nula):** Não há diferença (por exemplo, a média é igual ao valor especificado).
- **H_1 (Hipótese Alternativa):** Existe uma diferença (por exemplo, a média é diferente do valor especificado).

Exemplo

- **Testando se a média de uma amostra é diferente de um valor específico:**

```
pythonCopiar código
from scipy.stats import ttest_1samp
```

```
data = [1.2, 2.3, 3.1, 4.4, 5.0]
stat, p = ttest_1samp(data, 3, alternative="two-sided")
print(f"Statistics={stat}, p-value={p}")
```

- **H0:** A média é 3.
- **H1:** A média não é 3.

2. "greater"

- **Interpretação:** A hipótese alternativa é que a estatística de teste é maior que um valor específico.
- **H0 (Hipótese Nula):** A estatística de teste é menor ou igual a um valor específico.
- **H1 (Hipótese Alternativa):** A estatística de teste é maior que um valor específico.

Exemplo

- **Testando se a média de uma amostra é maior que um valor específico:**

```
pythonCopiar código
from scipy.stats import ttest_1samp

data = [1.2, 2.3, 3.1, 4.4, 5.0]
stat, p = ttest_1samp(data, 3, alternative="greater")
print(f"Statistics={stat}, p-value={p}")
```

- **H0:** A média é menor ou igual a 3.
- **H1:** A média é maior que 3.

3. "less"

- **Interpretação:** A hipótese alternativa é que a estatística de teste é menor que um valor específico.
- **H0 (Hipótese Nula):** A estatística de teste é maior ou igual a um valor específico.
- **H1 (Hipótese Alternativa):** A estatística de teste é menor que um valor específico.

Exemplo

- **Testando se a média de uma amostra é menor que um valor específico:**

```
pythonCopiar código
from scipy.stats import ttest_1samp

data = [1.2, 2.3, 3.1, 4.4, 5.0]
stat, p = ttest_1samp(data, 3, alternative="less")
print(f"Statistics={stat}, p-value={p}")
```

- **H0:** A média é maior ou igual a 3.
- **H1:** A média é menor que 3.

Aplicação em Diferentes Testes

Teste t de Student (Duas Amostras Independentes)

- **Função:** `ttest_ind(a, b, alternative="two-sided"|"greater"|"less")`
- **Comparação:** Compara as médias de duas amostras independentes.

Teste t Pareado

- **Função:** `ttest_rel(a, b, alternative="two-sided"|"greater"|"less")`
- **Comparação:** Compara as médias de duas amostras dependentes (ex.: antes e depois de um tratamento).

Teste Mann-Whitney U

- **Função:** `mannwhitneyu(a, b, alternative="two-sided"|"greater"|"less")`
- **Comparação:** Compara distribuições de duas amostras independentes, especialmente quando os tamanhos das amostras são diferentes.

Teste de Wilcoxon

- **Função:** `wilcoxon(a, b, alternative="two-sided"|"greater"|"less")`
- **Comparação:** Compara as medianas de duas amostras pareadas.

Conclusão

O parâmetro `alternative` é crucial para definir a natureza do teste estatístico. Ele especifica se estamos testando diferenças em ambas as direções ("two-sided") ou em uma direção específica ("greater" ou "less"). A escolha do valor adequado para `alternative` depende da hipótese de pesquisa e do contexto do teste.