



Aula #6: Design do Experimento II - Teoria

≡ Módulo	A/B Testing
# Aula	6

Aula #6: Design do Experimento II

Objetivo da Aula:

Entender os parâmetros do Teste A/B:

- ☐ Confident Level and Interval (**Nível e Intervalo de confiança**)
- ☐ Significant Level (**Nível de significância**)
- ☐ Effect Size (**Tamanho do Efeito**)
- ☐ Power Test (**Poder Estatístico**)
- ☐ Erros Tipo I e Tipo II

Conteúdo:

▼ **Nível de confiança**

Se todos os clientes visitassem a página atual (população), o **verdadeiro valor da conversão** seria de 19%.

O problema é que estou trabalhando em uma amostra de 500 clientes e o **valor da conversão sobre a amostra** é de 12%.

Existe uma **diferença entre o verdadeiro valor** da conversão (população) e o **valor estimado da conversão** sobre a amostra.

Portanto, existe uma **incerteza sobre o valor estimado** da conversão sobre a amostra. Logo, para ter mais segurança sobre o valor estimado, eu poderia pensar em um intervalo ao redor desse valor.

Eu posso criar um intervalo em torno da conversão estimada, que contenha a conversão verdadeira e desconhecida. Assim, se eu repetir o processo de amostragem 20x, estimar 20 valores de conversão, um para cada amostra e calcular um intervalo para cada valor estimado, **eu espero que o verdadeiro valor da conversão esteja em 19 intervalos**.

Ou seja, eu quero estar confiante de que 19 dos 20 intervalos contém o verdadeiro valor da conversão. Logo, **eu quero uma confiança de 95%**.

“Nível de confiança refere-se à taxa de sucesso do método no longo prazo, ou seja, com que frequência o intervalo vai capturar o parâmetro de interesse”

Melhor dizendo, **há 95% de chance que a escolha de uma amostra aleatória dessa população resulte em um intervalo de confiança que contém o verdadeiro valor que está sendo estimado**.

Porém, há 5% de chance do valor verdadeiro da conversão não estar dentro do intervalo.

▼ **Nível de significância (alpha)**

Eu gostaria de ter 95% de confiança de que, **95% dos intervalos construídos** em torno dos valores estimados para cada amostra **contenham o valor verdadeiro da conversão**.

Porém, há 5% de chance de que valor verdadeiro da conversão não esteja dentro do intervalo construído. Ou seja, **5% dos intervalos construídos em torno dos valores estimados para cada amostra não contenham o valor verdadeiro da conversão**.

Estou correndo um risco de 5% de tomar a decisão errada. A decisão errada em dizer que estou observando o valor verdadeiro da conversão, quando ela na verdade não está no intervalo.

Em outras palavras, estou correndo um risco de 5% de dizer que a amostra tem evidências suficientes para rejeitar a hipótese nula, quando a hipótese nula é verdadeira.

Logo, para um nível de confiança de 95%, eu estou assumindo no máximo 5% de risco ao rejeitar a hipótese nula, quando a hipótese nula é verdadeira. Para esse risco, damos o nome de **Nível de Significância**

“**Nível de Significância:** O risco máximo de rejeitar a hipótese nula, ou seja, observar evidências suficientes para afirmar que há um efeito, quando na verdade a hipótese nula é verdadeira, ou seja, não há efeito nenhum (foi uma ilusão).”

Se o risco de tomar a decisão errada estão no tamanho do intervalo, o que faz o intervalo aumentar ou diminuir: **Tamanho da amostra e variabilidade.**

▼ Effect Size

“Um forma padrão de **expressar a magnitude do resultado esperado** do seu estudo. Geralmente, baseado em estudos similares ou estudos anteriores”.

Quão grande é a diferença que esperamos que haja entre as taxas de conversão?

Obs: Diferenças pequenas exigem mais dados, diferenças maiores exigem menos dados.

▼ Formas de calcular o Effect Size

▼ Cohen's d

Cohen's d mede a diferença entre as médias de duas amostras

$$d = \frac{\mu_1 - \mu_2}{s}$$

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

u1, u2 : média da amostra do **grupo de tratamento** e amostra do **grupo controle**

n1, n2 : tamanho da amostra do **grupo de tratamento** e amostra do **grupo controle**

s1: desvio padrão da amostra do **grupo de tratamento** e amostra do **grupo controle**

▼ Código em Python

```
import numpy as np
import pingouin as pg

pg.compute_effsize(x, y, eftype='cohen')
```

▼ Glass' delta

Glass's delta é muito similar a métrica Cohen's d. A diferença se dá no uso somente do desvio-padrão da amostra do grupo controle, no cálculo

$$d = \frac{\mu_1 - \mu_2}{s_2}$$

u1, u2 : média da amostra do **grupo de tratamento** e amostra do **grupo controle**

s2: desvio padrão da amostra do **grupo controle**

▼ Código em Python

```
import numpy as np
import pingouin as pg

pg.compute_effsize(x, y, eftype='glass')
```

▼ Hedge's g

Cohens' d é um estimado enviesado para o tamanho do efeito a nível da população, principalmente para amostra pequenas, com $n < 20$. Para

esse contexto, a métricas de Hedge's corrige o tamanho da amostra por um fator baseado na função gamma.

$$g = d \times \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right)$$

u1, u2 : média da amostra do **grupo de tratamento** e amostra do **grupo controle**

s2: desvio padrão da amostra do **grupo controle**

▼ Código em Python

```
import numpy as np
import pingouin as pg

pg.compute_effsize(x, y, eftype='hedges')
```

▼ Cohen's h

Medida usada para comparar proporções de 2 amostras independentes.

$$h = 2(\arcsin(\sqrt{p1}) - \arcsin(\sqrt{p2}))$$

```
from statsmodels.stats.proportion import proportion_effectsize

cohen_h = proportion_effectsize(np.mean(x), np.mean(y))
print(f"Cohen's h: {cohen_h:.6f}")
```

▼ Poder Estatístico Power = (1 - beta)

O **poder estatístico** é a probabilidade em que um teste pode detectar corretamente um efeito real quando realmente existe um.

Estatistiquês: “O **poder estatístico** de um teste de hipótese é a probabilidade do teste corretamente rejeitar a hipóteses nula, quando a hipótese alternativa é verdadeira”.

Mas eu ainda posso cometer um erro. O erro de afirmar que não há nenhum efeito, quando realmente existir um. (Erro tipo II)

Power = 80% (O **poder estatístico** é a probabilidade em que um teste pode detectar corretamente um efeito real quando realmente existe um)

Beta = 20%

▼ Tamanho da Amostra

- Fórmula do Poder Estatístico

```
sample_n = sms.NormalIndPower().solve_power(  
    effect_size,  
    power=power,  
    alpha=significance_level )  
  
sample_n = ceil( sample_n )  
sample_n
```

▼ Tipos de Erros possível de acontecer:

Erro Tipo I: Erro ao dizer que há um efeito, quando realmente não existe nenhum.

(Erro ao rejeitar a hipótese nula, quando ela é realmente verdadeira)

Erro Tipo II: O erro ao afirmar que não há nenhum efeito, quando realmente existir um.

(Erro ao dizer que a hipótese nula é verdadeira, quando na verdade, a hipótese alternativa é a real)

▼ Na próxima aula ...

Aula #7: Coleta e Preparação dos dados