Data Science Academy

www.datascienceacademy.com.br

Business Analytics

Variância da Soma

Uma propriedade básica da variância é que a variância da soma de duas variáveis independentes aleatórias é a soma das variâncias.

$$Var(X+Y) = Var(X) + Var(Y)$$

$$Var(X-Y) = Var(X) + Var(Y)$$

Isso significa que a hipótese nula e alternativa terão a mesma variância que será a soma das variâncias para o grupo de controle e o grupo de teste.

$$Var(\hat{d}) = Var(\hat{p}_B - \hat{p}_A) = Var(\hat{p}_A) + Var(\hat{p}_B) = \frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}$$

O desvio padrão pode então ser calculado como:

$$\sigma = \sqrt{Var(\hat{d})} = \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$$

Se colocarmos esta equação em termos de desvio padrão para a distribuição de Bernoulli, temos:

$$\sigma = \sqrt{Var(\hat{d})} = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

e obtemos a aproximação de Satterthwaite para o erro padrão agrupado. Se calcularmos a probabilidade combinada e usarmos a probabilidade combinada para calcular o desvio padrão para ambos os grupos, obtemos:

$$\sigma = \sqrt{Var(\hat{d})} = \sqrt{\frac{s_p^2}{n_A} + \frac{s_p^2}{n_B}} = \sqrt{s_p^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)} = \sqrt{\hat{p}_p(1-\hat{p}_p)\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

Onde:

$$\hat{p}_p = \frac{p_A N_A + p_B N_B}{N_A + N_B}$$

Ambas as equações para o erro padrão combinado fornecerão resultados muito semelhantes.

Com isso, agora temos informações suficientes para construir as distribuições para a hipótese nula e a hipótese alternativa.