UFMG

DCC

DEPARTAMENTO DE
CIÊNCIA DA COMPUTAÇÃO

MÉTRICAS PARA *FAIRNESS* EM CIÊNCIA DE DADOS

GABRIEL LUCAS SILVA MACHADO

ORIENTADOR: MÁRIO SÉRGIO ALVIM

INTRODUÇÃO

- Discriminações algorítmicas são tratamentos feitos por algoritmos que desfavorecem alguém devido a uma característica “protegida” (raça, gênero, deficiência, etc);
  - Discriminação direta: a característica “protegida” afeta diretamente o resultado;
  - Discriminação indireta: resultados são afetados por características correlacionadas com atributos “protegidos”.
- Problema: os dados são discriminatórios!!!

MOTIVAÇÕES

- Discriminação é um problema mundial. Não queremos que o mesmo problema ocorra nos algoritmos;
- *Data mining* e *machine learning* já estão sendo usados em ferramentas que afetam diretamente a vida das pessoas.
  Exemplo: análise de CV, *credit scoring;*
- Justiça!!!

OBJETIVOS DO TRABALHO

- Estudo do estado da arte sobre o assunto (organização de conhecimento);
- Análise de métricas:
  - Como funcionam?
  - Como podem ser aplicadas ao problema?
  - Principais vantagens e desvantagens.
- Resultados esperados:
  - Artigo científico com uma análise das principais métricas que compõem o estado da arte.

CLASSIFICAÇÃO DE MÉTRICAS

Tipo de métrica	O que mede?	Tipo de discriminação
Testes estatísticos	Presença ou não de discriminação	Indireta
Medidas absolutas	Gravidade da discriminação	Indireta
Medidas condicionais	Gravidade da discriminação	Indireta
Medidas situacionais	Propagação da discriminação	Direta ou indireta

EXEMPLO: DIFERENÇA DE PROPORÇÕES

- Intuição: em um algoritmo justo, a classificação de indivíduos semelhantes pertencentes a grupos distintos deve ser parecida, logo compararemos as proporções dessas classificações;
- Hipótese nula: $P_0 = P_1$;
- Hipótese alternativa: $P_0 \neq P_1$.

CONTINUAÇÃO EXEMPLO: A ESTATÍSTICA Z

- Graus de significância: Tipicamente 0.01, 0.05 ou 0.07

  Teoricamente (0,1);

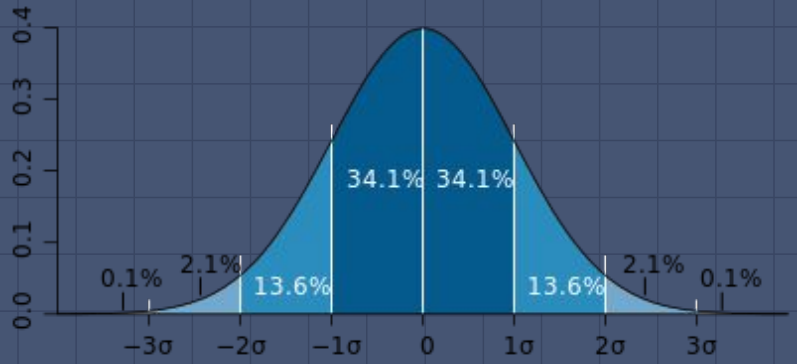
$$Z = \frac{Difference\ Between\ Proportions}{Standard\ Error}$$

- Os valores comuns que z assumirá caso a hipótese nula esteja correta seguem uma distribuição normal.

$$Z \sim N(\mu = 0, \sigma = 1)$$

CONTINUAÇÃO EXEMPLO: INTERPRETANDO RESULTADOS

- Após calcular a estatística z, calcular o p-valor;



- Comparar o p-valor com o grau de significância e decidir qual hipótese é a mais provável!

FEITO

- Análise dos testes estatísticos;
- Análise das medidas absolutas (a maior parte).

A FAZER

- Análise das medidas condicionais;
- Análise das medidas situacionais.