

Exercício – Tipos de variáveis

Data de entrega: 28/03/2024

Para as seguintes situações, classifique as variáveis em categóricas (nominais ou ordinais) ou numéricas (discretas ou contínuas). Quando necessário, justifique sua resposta.

1. Conjunto de dados experimentais de caracterização agronômica de cultivares de soja.

Pesquisadores maranhenses disponibilizaram recentemente um conjunto de dados contendo observações provenientes de um experimento planejado para coletar informações agronômicas de 40 cultivares de soja. O conjunto de dados, bem como outras informações relevantes, pode ser acessado no [Machine Learning Repository da UC Irvine](#). Foram coletadas medidas em duas temporadas de plantio subsequentes, e foram realizadas quatro repetições de cada medida. Cada observação foi fruto da média de 10 plantas durante a colheita. As variáveis disponíveis estão descritas no quadro abaixo.

Quadro 1. Descrição das variáveis disponíveis no conjunto de dados experimentais de caracterização agronômica de cultivares de soja.

Nome da variável	Tipo do dado	Descrição
Temporada	Inteiro	1 ou 2
Cultivar	Categórico	Nomes das cultivares
Repetição	Inteiro	1, 2, 3 ou 4
PH	Contínuo	Altura da planta (cm)
IFP	Contínuo	Altura da inserção da primeira vagem (cm)
NLP	Contínuo	Número de hastes (unidade)
NGL	Contínuo	Número de grãos por planta (unidade)
NS	Contínuo	Número de grãos por vagem (unidade)
MHG	Contínuo	Peso de 1000 sementes (g)
GY	Contínuo	Produtividade (kg/ha)

Fonte: adaptado de [Machine Learning Repository da UC Irvine](#) (Tradução livre)

2. Subconjunto de dados de uma pesquisa de saúde pública nos Estados Unidos entre 2013-2014

Subconjunto de dados da pesquisa nacional de saúde e nutrição (NHANES) uma população dos Estados Unidos, conduzida entre 2013-2014. Esse conjunto está disponível em [Machine Learning Repository da UC Irvine](#). Esse conjunto possui informações declaradas por pessoas entrevistadas, os entrevistados com mais de 65 anos foram rotulados como sêniores, e aqueles com menos de 65 anos foram rotulados como não sêniores. As variáveis disponíveis estão descritas no quadro abaixo.

Quadro 2. Descrição das variáveis disponíveis no conjunto de dados da pesquisa nacional de saúde e nutrição (NHANES) dos EUA, realizada entre 2013-2014

Nome da variável	Tipo do dado	Descrição
SEQN	Contínuo	Número de identificação do(a) entrevistado(a)
age_group	Categórico	Grupo de idade do(a) entrevistado(a) (sênior/não sênior)
RIDAGEYR	Contínuo	Idade do(a) entrevistado(a)
RIAGENDR	Contínuo	Gênero do(a) entrevistado(a) – 1 para masculino e 2 para feminino
PAQ605	Contínuo	Se o(a) entrevistado(a) participa de atividades esportivas ou recreativas de intensidade moderada ou intensa em uma semana típica - 1 para sim e 2 para não
BMXBMI	Contínuo	Índice de massa corporal
LBXGLU	Contínuo	Nível de glicose sanguínea após a alimentação
DIQ010	Contínuo	Se o(a) entrevistado(a) é diabético(a) - 1 para sim e 2 para não
LBXIN	Contínuo	Nível de insulina do entrevistado

Fonte: adaptado de em [Machine Learning Repository da UC Irvine](#) (Tradução livre)