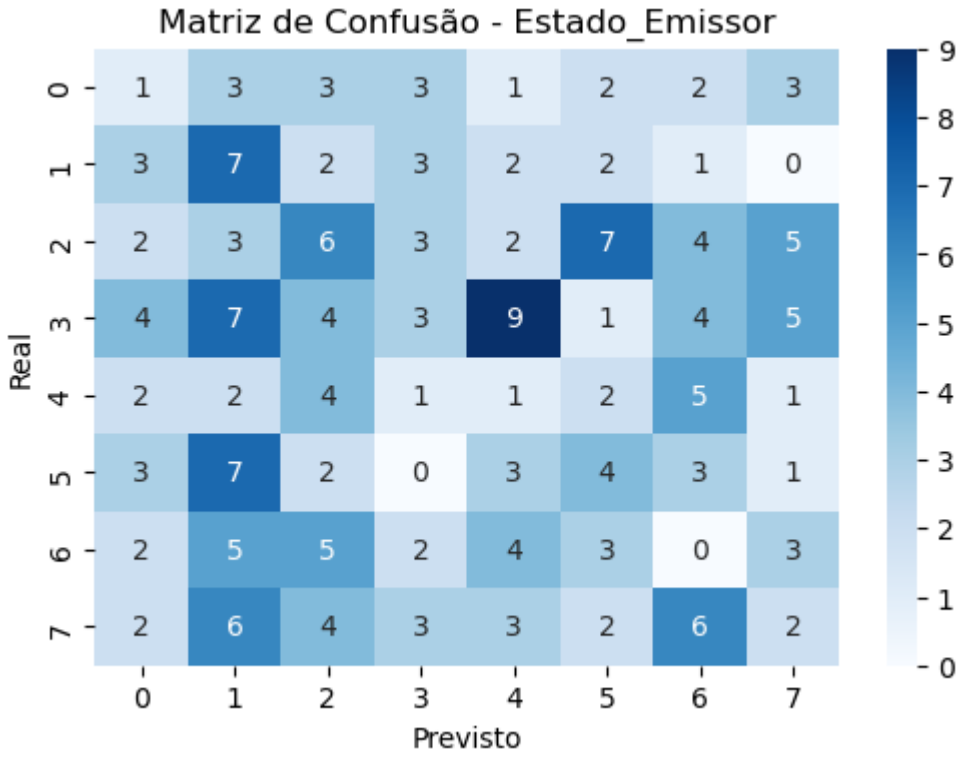
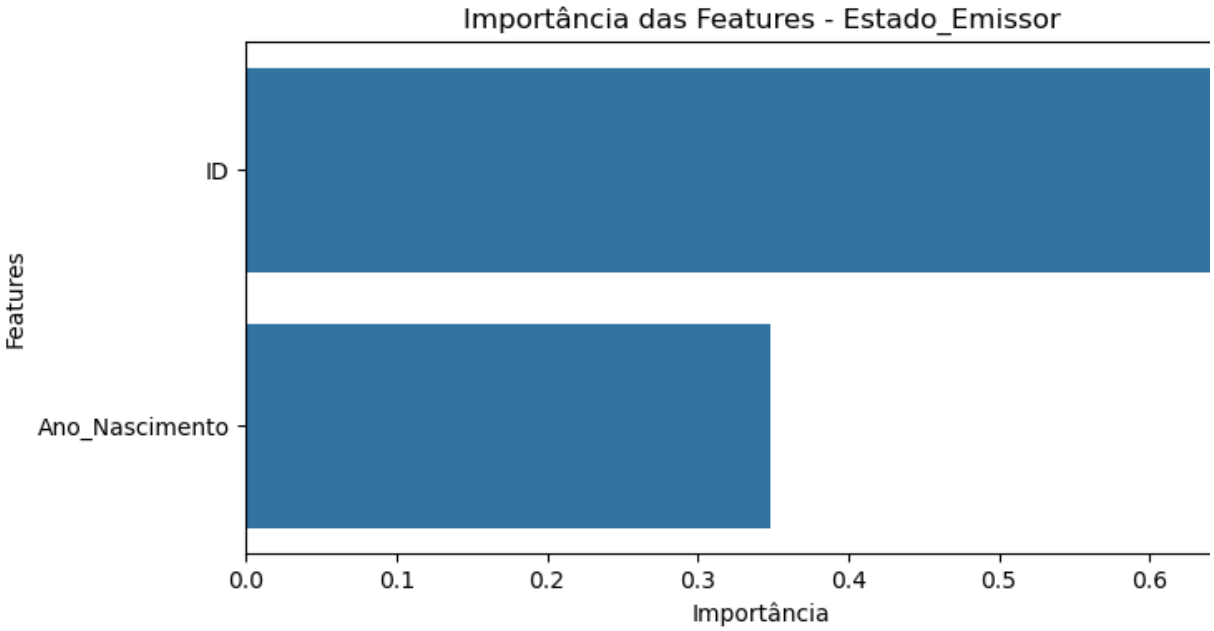
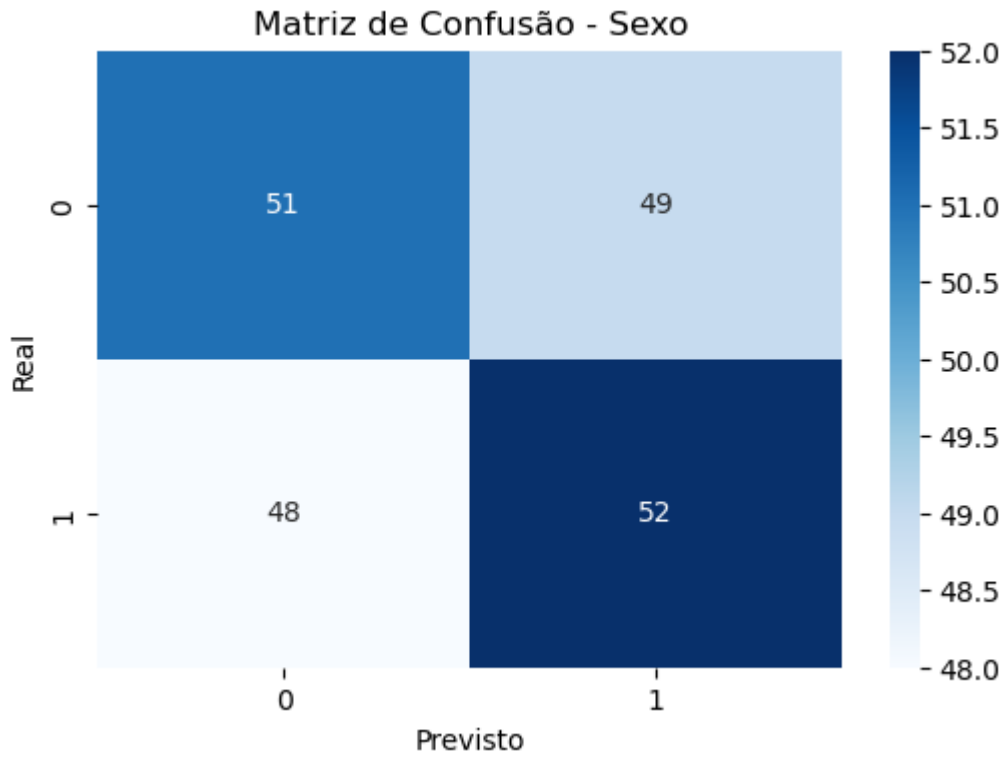
In [3]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, confusion_matrix

# Carregar os dados
df = pd.read_excel("dados_rg.xlsx")

# Converter datas para anos
df['Ano_Nascimento'] = pd.to_datetime(df['Data_Nascimento']).dt.year
df['Ano_Emissao'] = pd.to_datetime(df['Data_Emissão']).dt.year
df.drop(columns=['Data_Nascimento', 'Data_Emissão', 'Nome', 'RG'], inplac

# Codificar variáveis categóricas
label_encoders = {}
for col in ['Sexo', 'Estado_Emissor']:
    le = LabelEncoder()
    df[col] = le.fit_transform(df[col])
    label_encoders[col] = le

# Separar features e targets
X = df.drop(columns=['Estado_Emissor', 'Sexo', 'Ano_Emissao'])

def train_model(target_col):
    y = df[target_col]
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0
    model = RandomForestClassifier()
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    print(f'Acurácia na previsão de {target_col}: {accuracy:.2f}')

    # Matriz de Confusão
    cm = confusion_matrix(y_test, y_pred)
    plt.figure(figsize=(6, 4))
    sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
    plt.xlabel('Previsto')
    plt.ylabel('Real')
    plt.title(f'Matriz de Confusão - {target_col}')
    plt.show()

    # Importância das Features
    feature_importances = model.feature_importances_
    features = X.columns
    plt.figure(figsize=(8, 4))
    sns.barplot(x=feature_importances, y=features)
    plt.xlabel('Importância')
    plt.ylabel('Features')
    plt.title(f'Importância das Features - {target_col}')
    plt.show()

# Treinar modelos para cada variável
for target in ['Estado_Emissor', 'Sexo', 'Ano_Emissao']:
    train_model(target)
```

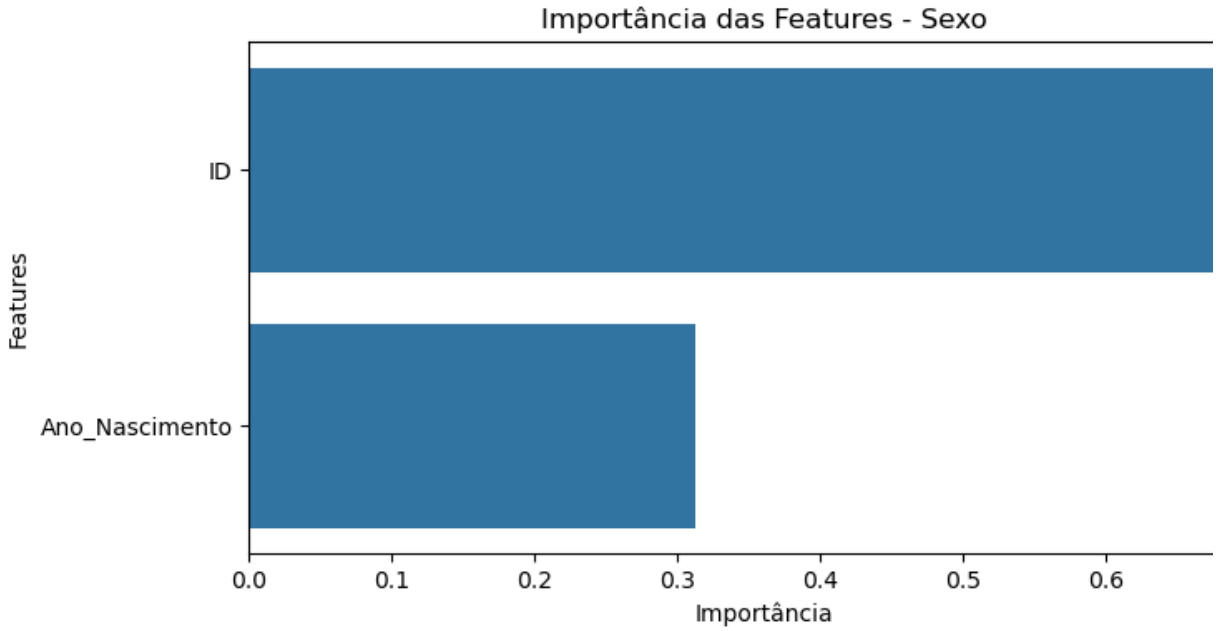
Acurácia na previsão de Estado_Emissor: 0.12



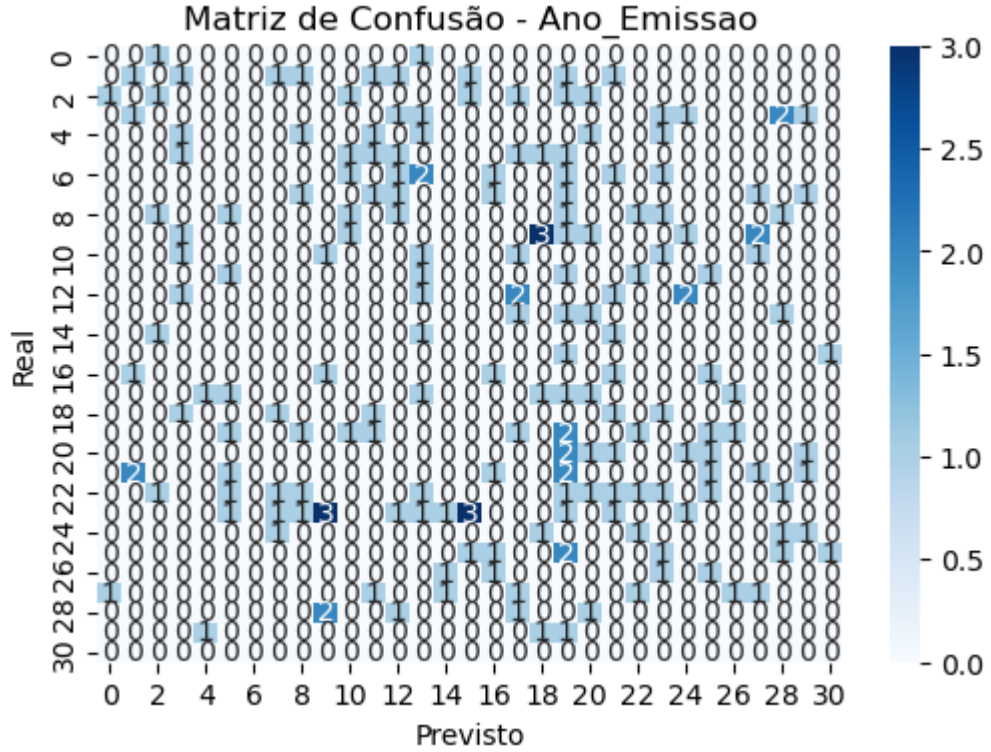


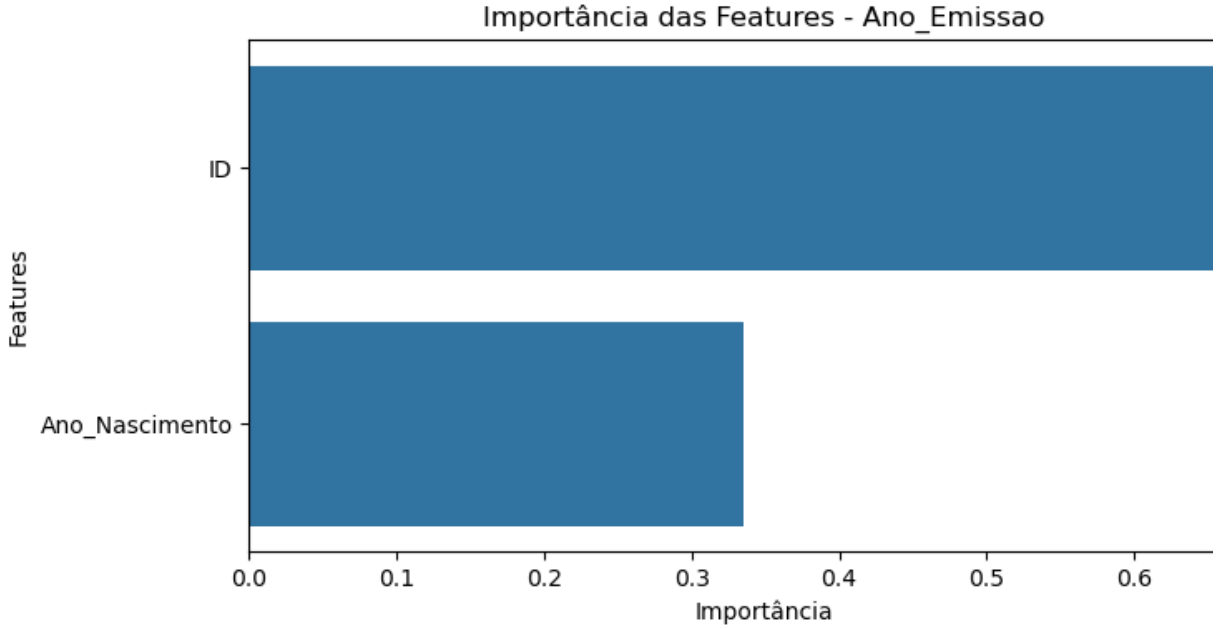
Acurácia na previsão de Sexo: 0.52





Acurácia na previsão de Ano_Emissao: 0.04


Matriz de Confusão - Ano_Emissao
Real
Previsto
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
3.0
2.5
2.0
1.5
1.0
0.5
0.0


Importância das Features - Ano_Emissao
Features
ID
Ano_Nascimento
0.0 0.1 0.2 0.3 0.4 0.5 0.6
Importância

In [ ]: