Relatório Estatístico do melhor grupo da FIAP

Integrantes:

Arthur Hieda Cunha - 551882

Henrique Marra Barbosa - 97672

Lucas Bueno Taets Gustavo - 552162

Com base nos dados adquiridos no Kaggle sobre diabetes, conseguimos analisar e chegamos na conclusão de que sobre a variável BMI (Seria o IMC, índice de massa corpórea):

1. É provável que seja menor que a média populacional
2. É muito provável que seja maior que a média populacional
3. É muito provável (Quase certo) que a variável esteja entre a média menos três vezes o desvio padrão e a média mais três vezes o evento padrão

Dentro da mesma variável é notável que:

1. Dentro de um nível de significância de 6% a amostra coletada é menor que a média populacional
2. Num nível de significância de 9%, mas agora com uma coleta amostra um pouco maior, a amostra se mostrou maior que a média populacional
3. Agora num nível de significância 2% e uma amostra menor é visto que não se diferiu da média populacional

Fizemos também nossas Análises Bivariadas e Multivariadas.

Nas Análises Bivariadas utilizamos as variáveis BMI e Age e descobrimos que:

1. Há uma correlação pequena entre as duas variáveis, portanto, não são tão dependentes
2. É um modelo significativo, entretanto, o R ajustado está distante de 1 (Esse R ajustado é o que mede a qualidade do ajuste, ou seja, o quanto uma variável explica a outra)

Conclusão: O modelo nos mostra que, por mais que as variáveis tenham correlação, a dependência entre elas é extremamente baixa o que nos prova que a ligar o IMC com a idade do paciente não diz muito sobre seu diagnóstico de diabetes, talvez a melhor decisão seria analisar outras vertentes

Dentro da Análise Multivariada foi utilizado as variáveis BMI, Age e, também o tipo da diabetes. No início tivemos ressalvas em escolher o tipo de diabetes como variável, porém testamos e vimos que é possível ver o tipo de diabetes que o paciente possui caso ele tenha a doença, dito isso descobrimos que:

1. Como visto na Análise Bivariada, há uma correlação entre as variáveis, contudo, não é uma grande correlação, o nível de dependência entre as variáveis é pequeno
2. Analisamos o modelo e vimos que o p-value nos diz que o modelo é significativo, entretanto, o R ajustado é pequeno

Conclusão: Na Análise Multivariada confirmamos o que foi visto na Análise Bivariada e realmente não é uma boa escolha analisar os pacientes a partir do IMC e da Idade.