

# ANÁLISE DOS NÍVEIS DE GLICOSE ENTRE ADULTOS E IDOSOS QUE TEM DIABETES

Murilo Rocha

## Definição do Problema

O Hospital INCAL está associado a uma grande empresa que disponibiliza serviços especializadas para a resolução de problemas na área da saúde utilizando técnicas e ferramentas de Análise de Dados e Inteligência Artificial. O INCAL acredita que pode haver uma diferença natural nos níveis de glicose entre adultos e idosos e pretende aumentar sua participação no mercado com a introdução de um tratamento diferenciado e especializado para cada tipo de paciente entre esses dois grupos. O objetivo deste projeto é analisar o conjunto de dados fornecido pelo hospital e extrair informações valiosas com foco principal em estudar se existe uma diferença significativa na glicemia entre pacientes adultos e idosos com diabetes.

Os dados fornecidos pelo hospital têm duas colunas principais:

- Coluna 1 - O nível de glicose no sangue em jejum de adultos (mg/dL)
- Coluna 2 - O nível de glicose no sangue em jejum de idosos (mg/dL)

## Análise Descritiva

Analizando a Tabela 1, percebe-se que as médias de glicemia entre adultos e idosos são muito próximas, embora os adultos tenham uma média um pouco mais alta que pode ser fruto do acaso, tudo parece indicar que não há diferença significativa. Assim como a média, o desvio padrão também não apresenta uma diferença muito significativa. Diferentemente da média e do desvio padrão, a diferença de variância entre os grupos de adultos e idosos parece ser significativa e merece ser estudada. Os valores da mediana estão muito próximos da média, este é um bom indicador de que a distribuição de ambas as variáveis segue uma distribuição normal.

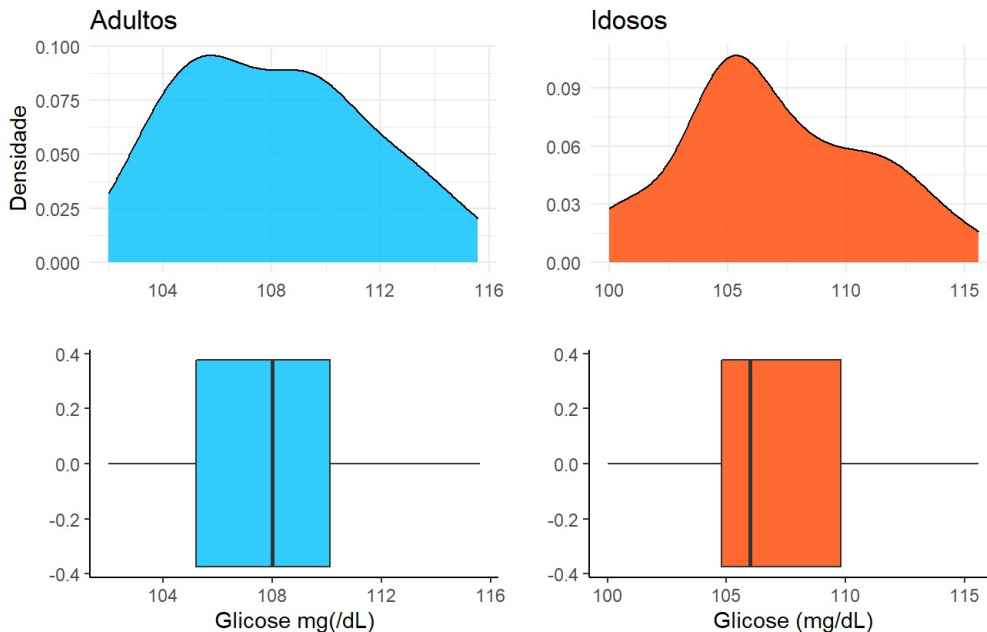
Um ponto importante que deve ser destacado é que embora a média dos dois grupos seja muito próxima, de acordo com as medidas de desvio padrão e variância, os níveis de glicemia dos idosos são mais dispersos, ou seja, estão mais distantes da média, abrindo uma margem para uma análise mais aprofundada.

Tabela 1: Estatísticas do nível de glicose entre adultos e idosos do hospital INCAL

Categoria	Média	Desvio Padrão	Variância	Mediana
Adulto	108.11	3.5	12.26	108
Idoso	107	3.96	15.67	106

Para que possamos obter informações mais precisas dos dados, aplicamos técnicas estatísticas que nos informam como os dados são distribuídos, utilizando valores de máximo, mínimo, mediana e intervalo interquartil (IQR). Analisando os valores máximos e mínimos dos dois grupos, verificamos que os idosos apresentam um maior intervalo de valores, onde a diferença entre o maior e o menor valor é de  $15,6 \frac{mg}{dL}$  e para os adultos essa diferença é de  $13,6 \frac{mg}{dL}$ . Portanto, se observarmos a Tabela 1 e a diferença entre os valores máximos e mínimos de ambos os grupos, podemos dizer que os valores de glicemia dos idosos abrangem uma faixa maior de valores em média em relação aos adultos.

# Distribuição de densidade dos níveis de glicose de pacientes com diabetes dados do hospital INCAL



O intervalo interquartil (IQR) de ambos as variáveis é equivalente aproximadamente a 5, ou seja, não há diferença entre os valores do primeiro e terceiro quartil. Os adultos tem como valores do primeiro e terceiro quartil, 105.2 e 110.1 respectivamente e os idosos possuem como primeiro e terceiro quartil 104.8 e 109.8 respectivamente.

Com base nas informações obtidas até o momento, podemos dizer que os adultos apresentam um nível médio de glicemia maior que os idosos, mas não o suficiente para ser significativo. Outra informação importante é que embora a média de glicemia dos adultos seja um pouco maior, os índices de glicemia dos idosos cobre um range maior dos valores e estão mais dispersos acima da mediana, como podemos ver no box plot acima e nas medidas de desvio padrão, variância, máximo e mínimo. Portanto, podemos concluir que embora os grupos apresentem algumas pequenas diferenças na distribuição e nas medidas descritivas, é bem provável que ambos os grupos não apresentam uma diferença significativa nos níveis de glicemia medidos em jejum.

## Teste de normalidade - Teste de Shapiro-Wilk

Para que possamos atingir o objetivo de verificar se existe de fato diferença significativa entre os níveis de glicose entre adultos e idosos, precisamos utilizar dois testes estatísticos, sendo o primeiro o teste de shapiro, que nos informa se os dados seguem uma distribuição normal e o teste t, que permite determinar se existe ou não diferença significativa entre as médias dos dois grupos. Para aplicarmos o teste t, primeiro temos que confirmar que os dados seguem uma distribuição normal e que ambas as variáveis são independentes.

O teste de Shapiro possui duas hipóteses, que chamamos de hipótese nula ( $H_0$ ) e hipótese alternativa ( $H_1$ ). A partir deste teste podemos aceitar ou rejeitar a hipótese nula com base no valor-p. As duas hipóteses são:

- $H_0$  - Se o valor-p for maior que 0,05 não rejeitamos a hipótese nula e podemos assumir a normalidade dos dados.
- $H_1$  - Se o valor-p for menor que 0,05 rejeitamos a hipótese nula e não podemos assumir a normalidade dos dados.

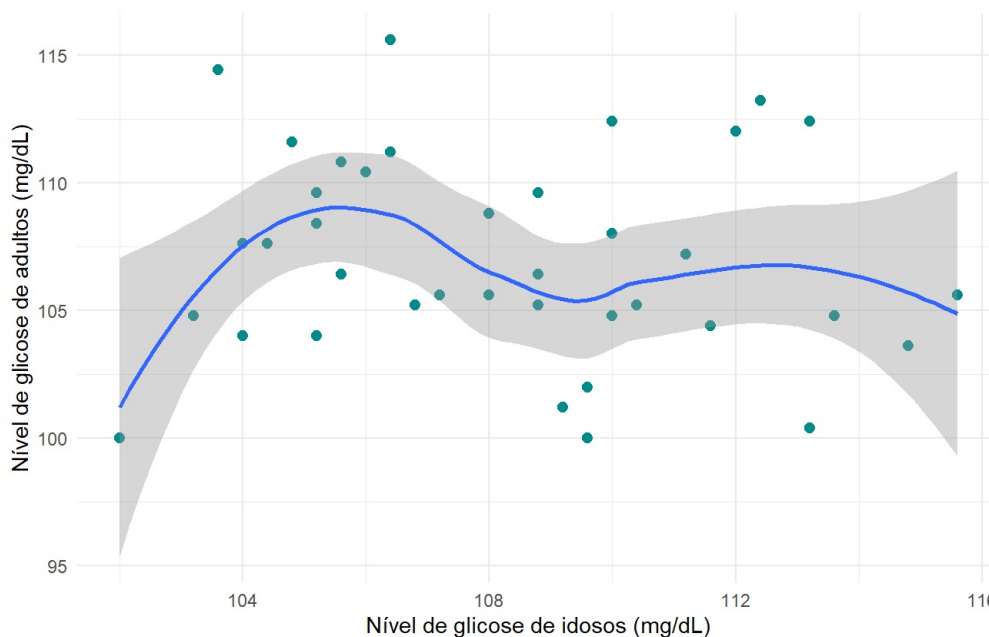
Tabela 2: Teste de Shapiro-Wilk

Catégorie	valor-p
Adulto	0.3788
Idoso	0.3065

Observando a tabela 2, vemos que para ambos os grupos o valor de p é maior que 0,05, portanto não temos evidências estatísticas para rejeitar  $H_0$ . Portanto, podemos assumir a normalidade dos dados.

Outro passo importante antes de aplicar o teste t é examinar se as variáveis são independentes. Para isso faremos um gráfico de dispersão e o coeficiente de correlação de Pearson, que nos permite quantificar a relação entre duas variáveis.

Relação entre os níveis glicêmicos entre adultos e idosos  
dados do hospital INCAL



Pode-se observar pelo gráfico acima que ambas as variáveis não apresentam nenhum tipo de relação ou tendência, mas sim um comportamento bastante aleatório, indicando que as duas variáveis são provavelmente independentes. Para melhor quantificar se existe ou não relação, usamos o coeficiente de correlação de Pearson, que nos dá -0,08 como resultado. Ou seja, as variáveis são independentes e podemos proceder ao teste t.

## Teste t de Student

O teste t é uma ferramenta importante para comparar médias, é através deste teste que podemos verificar se existe ou não uma diferença significativa entre os níveis clicêmicos entre adultos e idosos. Assim como o teste de Shapiro, também temos duas hipóteses, uma nula e uma alternativa.

Primeiramente vamos aplicar o teste para saber se existe diferença entre os grupos de adultos e idosos, esse tipo de teste é chamado de bicaudal. Se houver diferença então aplicamos o teste para descobrir se essa diferença é positiva ou negativa, esse chamado teste unicaudal.

- $H_0$  - Não há diferença significativa entre os níveis de glicose entre adultos e idosos.
- $H_1$  - Há diferença significativa entre os níveis de glicose entre adultos e idosos.

Tabela 3: Resultado do teste t de Student

t	df	valor-p	intervalo de confiança (95%)
1,32	76,85	0,18	(-0,5539748, 2,7739748)

Analisando o resultado do teste t na tabela 3, podemos concluir que falhamos em rejeitar a hipótese nula pois o valor-p é maior que o limite de 0,05. Significa dizer que há uma probabilidade alta de não haver diferença significativa entre os níveis de glicose medido em jejum entre adultos idosos.

No final de um teste estatístico, sempre surge uma pergunta, este teste é confiável? O tamanho da amostra usado para o processo de teste foi ideal? Para que possamos responder a essas perguntas e finalizar nossa análise, aplicamos outro teste estatístico, chamado teste de potência ou teste de força, que basicamente nos diz se podemos confiar no teste t que foi aplicado.

Aplicando o teste de força, obtemos um valor de 0,14. Este é um valor muito baixo, indicando que o teste t aplicado a este conjunto de dados não é forte o suficiente para rejeitar a hipótese de que não há diferença significativa entre os níveis de glicose de adultos e idosos. Basicamente, podemos dizer que o número de observações do conjunto de dados usado para o teste (40) não é suficiente estatisticamente. Aplicando o mesmo teste, mas para descobrir qual seria o número ideal de observações, temos 105 como resultado.

# Conclusão

Com as informações obtidas por meio de análise descritiva e testes de hipóteses realizados neste conjunto de dados, assumimos que não há diferença significativa nos níveis de glicemia entre adultos e idosos. Deve-se levar em consideração também que o tamanho da amostra é pequeno, abrindo assim uma maior margem de chance de erro, então seria melhor aumentar o número de observações, onde o número ideal seria 105.

## Resultado do teste de hipótese

1. O valor  $p$  do teste  $t$  é 0,18 e, portanto, não rejeitamos a hipótese nula. Portanto, podemos dizer que não há diferença significativa entre os níveis de glicose entre adultos e idosos medidos em jejum.
2. Deve haver um fator diferente da categoria (adulto e idoso) que afete a glicemia das pessoas.
3. O hospital INCAL não deve introduzir um tratamento diferenciado do diabetes baseado nos grupos de adultos e idosos.

## Intervalos de confiança

1. Intervalo de confiança de 95% para adultos [106,99, 109,22].
2. Intervalo de confiança de 95% para idosos [105,73, 108,26].

Assim, a respectiva médiapopulacional deve estar dentro dessa faixa para obter resultados consistentes.

## Força do teste e tamanho da amostra

1. A força do teste é 0,144, o que é baixo.
2. Tamanho da amostra: O teste deve ser realizado com um tamanho de amostra maior, pelo menos 105 registros.
3. É sempre melhor ter um tamanho de amostra maior para obter melhores resultados, mas também deve-se levar em consideração outros fatores como custo, praticabilidade e tempo.

## Recomendações

1. Estatisticamente, não há diferença significativa nos níveis de glicemia entre adultos e idosos.
2. O hospital INCAL deve continuar seu estudo e considerar outros fatores que afetam os níveis de glicose no sangue, como o tamanho da pessoa, o peso, raça, idade, se é fumante e se possui alguma comorbidade.
3. O INCAL também deve considerar a tecnologia predominante, os tipos de tratamentos de diabetes, e os produtos lançados por seus concorrentes.
4. A partir de agora, para aumentar sua participação no mercado, a empresa deve se concentrar em outras áreas como marketing, estratégia e seguir trabalhando em pesquisa e desenvolvimento.



© Feito por: **Murilo Rocha**

Instagram (<https://www.instagram.com/murilo.i.a/?hl=pt-br>) | Kaggle (<https://www.kaggle.com/upadorprofzs>)