

Vocês apresentam boa sumarização dos resultados, parabéns!

Laboratório 1

Camille Menezes e Michel Miler

26 de Março de 2023

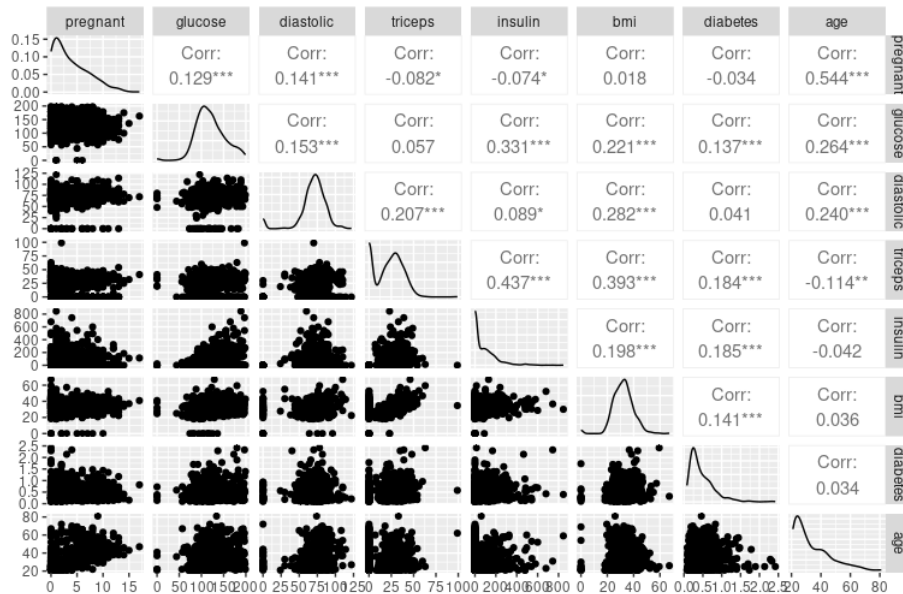
As figuras foram numeradas, mas em alguns trechos vocês escrevem com base nelas, mas não as citas. Sugiro rever isto nos próximos textos.

1 Sumarização dos resultados

As 768 mulheres da tribo Pina entre 21 e 81 anos, que residem próximo a Phoenix, apresentam, em média, quatro gestações. Sendo que, cerca de 31% delas apresentam o valor de glicose no sangue acima de 100 mg/dl, o que pode indicar pré-diabetes. Os níveis de pressão sanguínea diastólica de aproximadamente 75% delas é considerado normal, ou seja abaixo de 85 mmHg. Em relação aos níveis de insulina, indicador de diabetes, 75% das mulheres possuem níveis de insulina dentro dos padrões recomendados. O peso de 85% delas está acima considerado normal. Além disso, 35% delas obtiveram resultado positivo para o teste que avalia sinais de diabetes.

Senti falta de referências para as afirmações sobre limites recomendados e afins.

Figura 1: Gráficos de dispersão e correlação das variáveis no estudo acerca das mulheres de Pina da cidade de Phoenix no estado Arizona

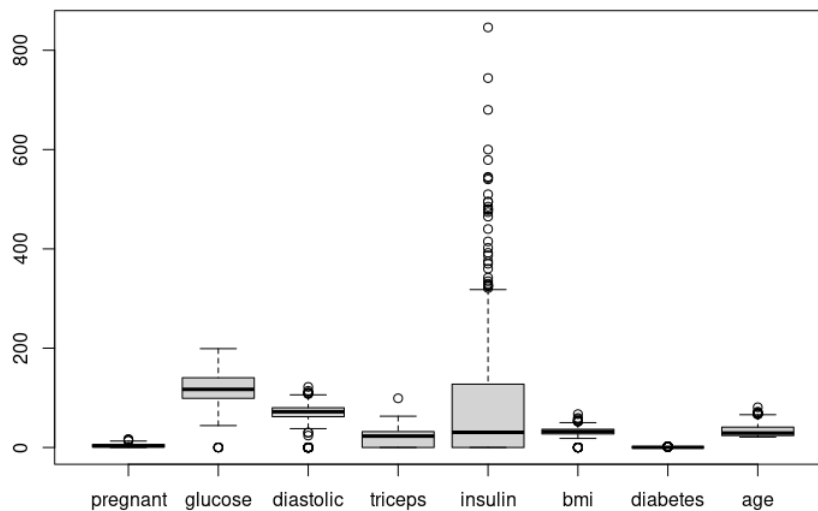


As variáveis, de modo geral, aparentam estar pouco correlacionadas, Figura 1. Apenas a idade que tem uma correlação moderada com a quantidade de gestações. Entretanto, pelo gráfico de dispersão Figura 4, a variável tríceps parece ter uma correlação moderada com o índice de massa corpórea, mas isso não aparece no valor da correlação de *pearson*. Isso pode explicado pela presença dos valores faltantes indicadas com o valor "zero".

Figura 2: Sumarização das variáveis acerca das mulheres de Pina da cidade de Phoenix no estado Arizona

pregnant	glucose	diastolic	triceps
Min. : 0.000	Min. : 0.0	Min. : 0.00	Min. : 0.00
1st Qu.: 1.000	1st Qu.: 99.0	1st Qu.: 62.00	1st Qu.: 0.00
Median : 3.000	Median :117.0	Median : 72.00	Median :23.00
Mean : 3.845	Mean :120.9	Mean : 69.11	Mean :20.54
3rd Qu.: 6.000	3rd Qu.:140.2	3rd Qu.: 80.00	3rd Qu.:32.00
Max. :17.000	Max. :199.0	Max. :122.00	Max. :99.00
insulin	bmi	diabetes	age
Min. : 0.0	Min. : 0.00	Min. :0.0780	Min. :21.00
1st Qu.: 0.0	1st Qu.:27.30	1st Qu.:0.2437	1st Qu.:24.00
Median : 30.5	Median :32.00	Median :0.3725	Median :29.00
Mean : 79.8	Mean :31.99	Mean :0.4719	Mean :33.24
3rd Qu.:127.2	3rd Qu.:36.60	3rd Qu.:0.6262	3rd Qu.:41.00
Max. :846.0	Max. :67.10	Max. :2.4200	Max. :81.00

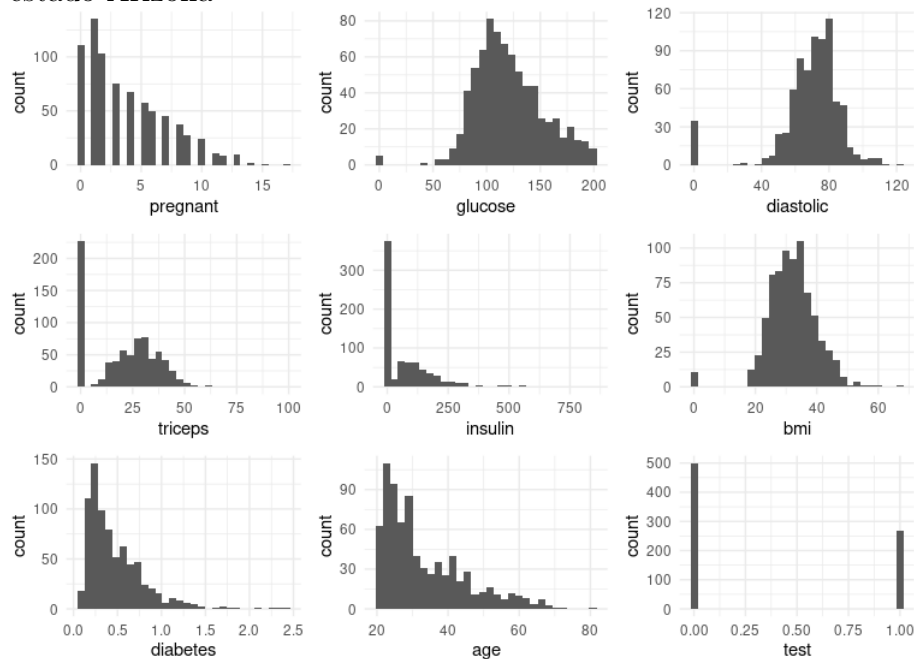
Figura 3: Boxplots das variáveis no estudo acerca das mulheres de Pina da cidade de Phoenix no estado Arizona



Com o boxplot do conjunto de variáveis, é possível notar que na variável insulina há um presença de diversos *outliers*. Estes valores parecem estar totalmente fora do que é observado nos seres humanos, contudo seria necessário confirmar com um especialista se existe alguma possibilidade desses valores serem observados na realidade, ou se estão incorretos.

O "excesso" do valor zero nas variáveis triceps e insulina atrapalharam a visualização mais adequada da distribuição dessas variáveis. As variáveis índice de massa corpórea (bmi), glicose (glucose) e diastólica (diastolic) apresentam uma distribuição de "sino", o que se assemelha a uma distribuição normal.

Figura 4: Histogramas das variáveis no estudo acerca das mulheres de Pina da cidade de Phoenix no estado Arizona



2 Mudanças após a retirada das observações não usuais

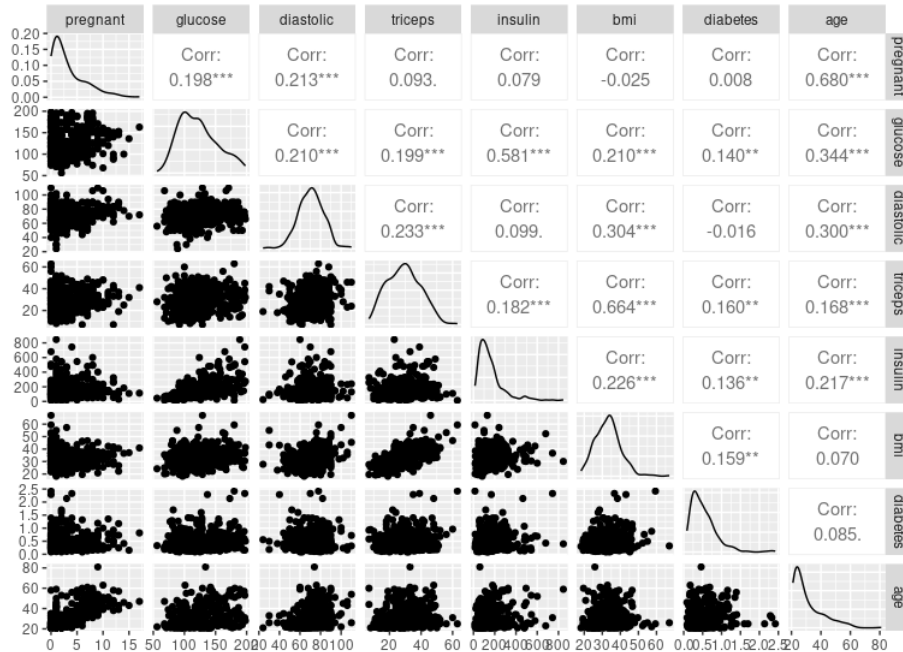
Após a retirada das observações não usuais, pode-se observar mudanças, sobretudo nos níveis de insulina, largura do tríceps e na pressão sanguínea diastólica. Agora, 61% das mulheres possuem níveis de glicose acima de 100 mg/dl, que indica pré-diabetes. A largura do tríceps que era em média 20,54 mm, apresenta em média 29,15 mm, Figuras 2 e 5. Em relação a pressão sanguínea diastólica, aproximadamente 87% das mulheres apresentam níveis normais.

Figura 5: Sumarização das variáveis acerca das mulheres de Pina da cidade de Phoenix no estado Arizona

pregnant	glucose	diastolic	triceps
Min. : 0.000	Min. : 56.0	Min. : 24.00	Min. : 7.00
1st Qu.: 1.000	1st Qu.: 99.0	1st Qu.: 62.00	1st Qu.:21.00
Median : 2.000	Median :119.0	Median : 70.00	Median :29.00
Mean : 3.301	Mean :122.6	Mean : 70.66	Mean :29.15
3rd Qu.: 5.000	3rd Qu.:143.0	3rd Qu.: 78.00	3rd Qu.:37.00
Max. :17.000	Max. :198.0	Max. :110.00	Max. :63.00
insulin	bmi	diabetes	age
Min. : 14.00	Min. :18.20	Min. :0.0850	Min. :21.00
1st Qu.: 76.75	1st Qu.:28.40	1st Qu.:0.2697	1st Qu.:23.00
Median :125.50	Median :33.20	Median :0.4495	Median :27.00
Mean :156.06	Mean :33.09	Mean :0.5230	Mean :30.86
3rd Qu.:190.00	3rd Qu.:37.10	3rd Qu.:0.6870	3rd Qu.:36.00
Max. :846.00	Max. :67.10	Max. :2.4200	Max. :81.00

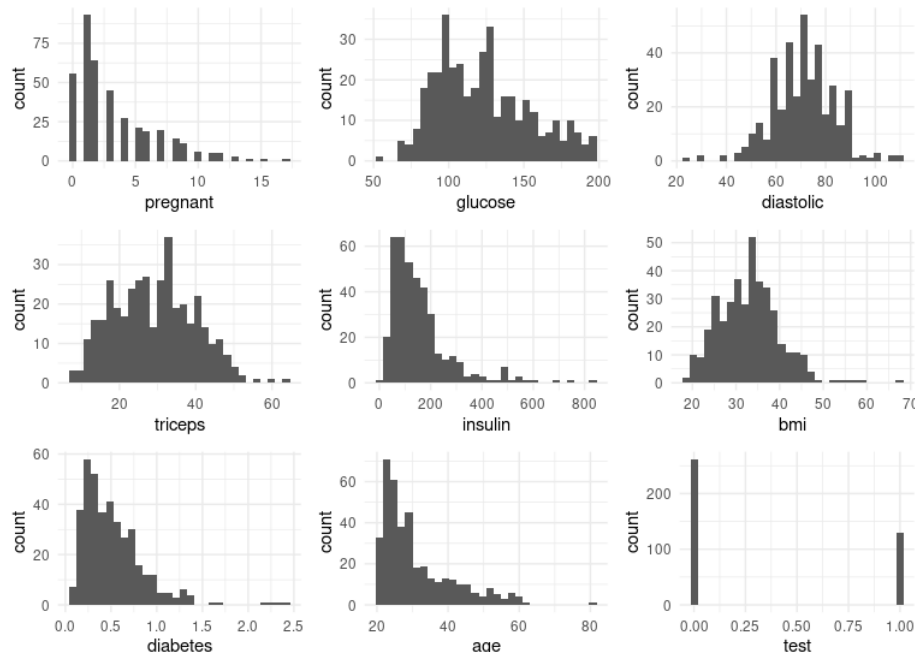
Agora podemos perceber que há uma maior correlação entre o nível de insulina e o nível de glicose no sangue. Do mesmo modo, pelo valor da correlação, conseguimos notar o que foi

Figura 6: Gráficos de dispersão e correlação das variáveis acerca das mulheres de Pina da cidade de Phoenix no estado Arizona



percebido anteriormente de que existia uma relação positiva entre o índice de massa corpórea (bmi) e o triceps.

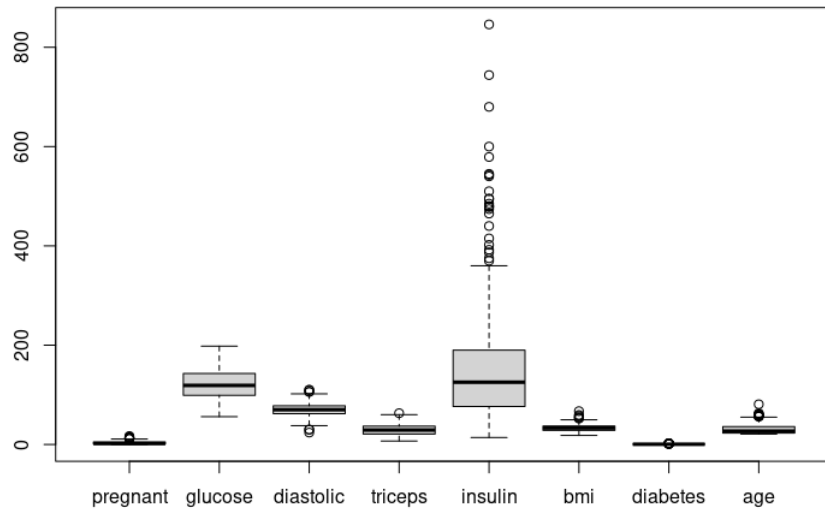
Figura 7: Histogramas das variáveis no estudo acerca das mulheres de Pina da cidade de Phoenix no estado Arizona



Agora é possível notar que variável triceps também tem uma distribuição "sino" e que a distribuição da variável insulina está sendo afetada por esses outliers citados anteriormente.

A variabilidade dentro dos quartis parece bem equivalente na variável insulina como é na variável glicose, se desconsiderarmos os outliers.

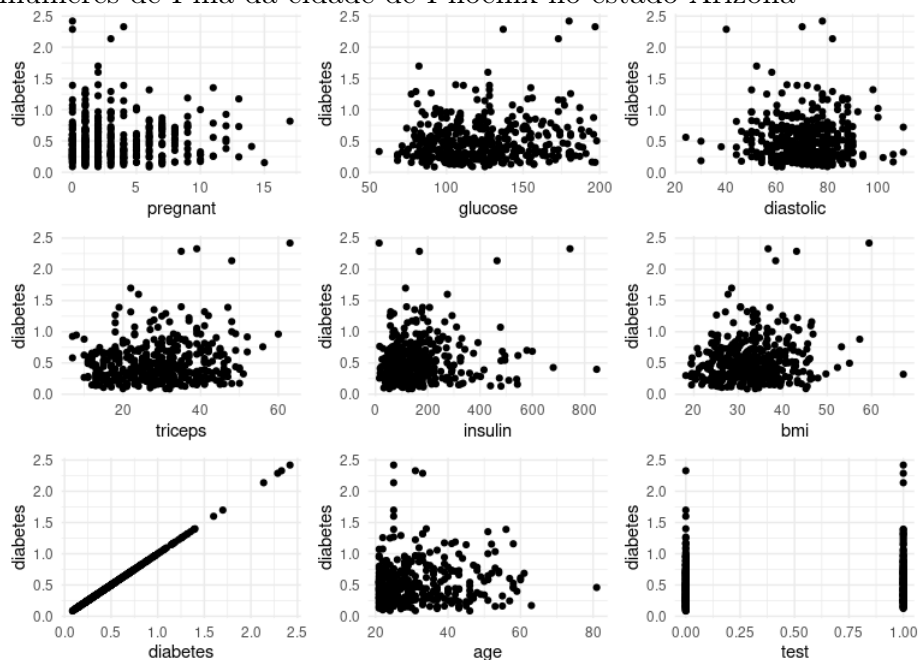
Figura 8: Boxplots das variáveis no estudo acerca das mulheres de Pina da cidade de Phoenix no estado Arizona



3 Relação da variável resposta (diabetes) com as demais variáveis

Não é possível notar relação entre a variável diabetes com as demais variáveis, o que é de se estranhar, haja vista que a variável diabetes é o nível de função diabética. Era esperado que esta variável tivesse uma relação com a glicose ou insulina, contudo isso não aconteceu de modo significamente notável.

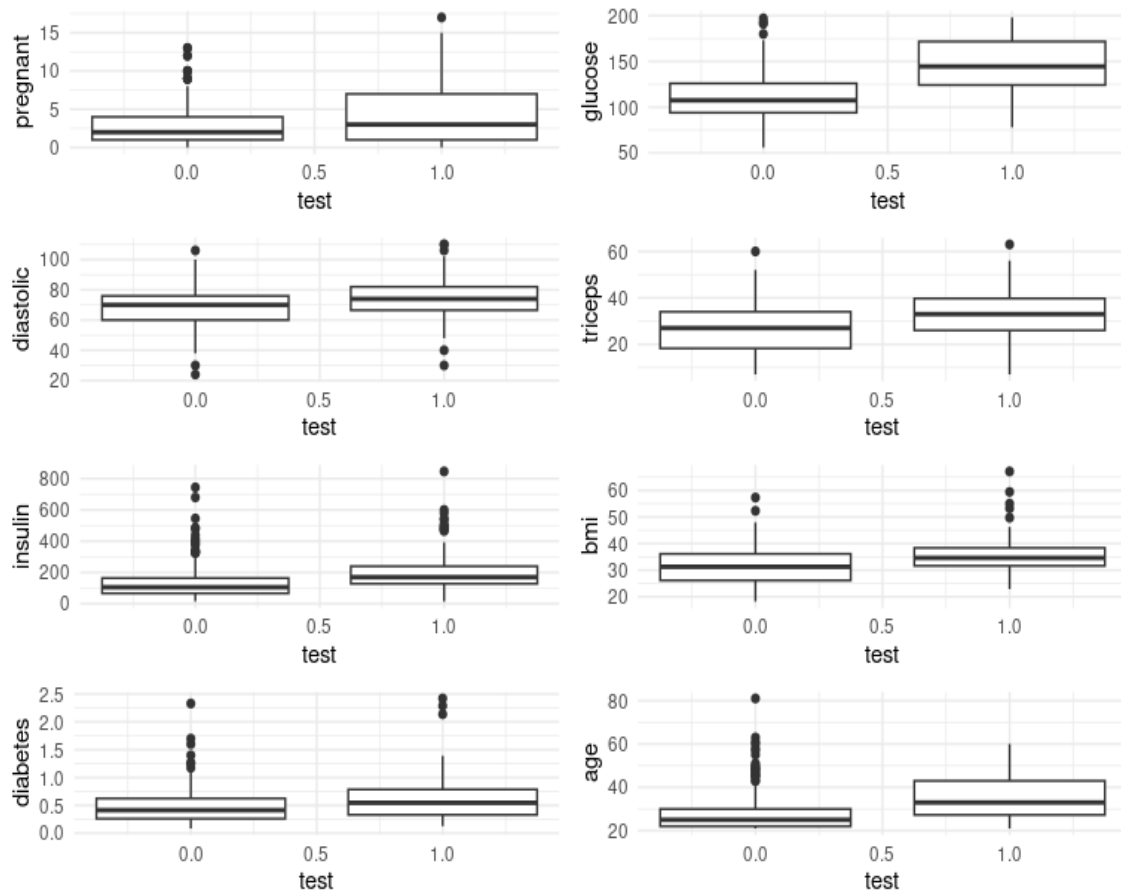
Figura 9: Gráficos de dispersão da variável diabetes com as demais variáveis presentes no estudo das mulheres de Pina da cidade de Phoenix no estado Arizona



4 Comparação do teste com as variáveis quantitativas

È possível perceber, através da Figura 10, que 75% das mulheres que apresentaram resultado **do teste positivo** possuem níveis mais altos de glicose que 75% das mulheres que negativaram no teste. Além disso, metade das mulheres que positivaram o teste possuem mais de 30 anos, enquanto 75% das mulheres com teste negativo possuem menos de 30 anos. As outras variáveis possuem diferenças menores entre os resultados do teste, sendo que a maior parte das mulheres com resultado positivo apresentam resultados mais altos em todas as variáveis.

Figura 10: Boxplots, em cada resultado do teste para avaliação de diabetes, das variáveis no estudo acerca das mulheres de Pina da cidade de Phoenix no estado Arizona



Vocês apresentaram um poder de síntese dos resultados muto bom.