

Universidade Federal do Rio de Janeiro



Universidade Federal
do Rio de Janeiro

Escola Politécnica

Estatística e Modelos Probabilísticos

Projeto Final - 2019.2

Aluno	Ian Patello de Freitas
DRE	116106085
Professor	Rosa M. M. Leão
Código	https://github.com/ianfreitas1/probest-2019/

Rio de Janeiro, 19 de Dezembro de 2019

Conteúdo

1	Introdução	1
2	Histograma e Função Distribuição Empírica	1
2.1	Idade	1
2.2	Peso	2
2.3	Carga Final	3
2.4	VO2 Máximo	4
2.5	Análise	5
3	Média, Variância e Boxplot	6
3.1	Idade	6
3.2	Peso	6
3.3	Carga Final	7
3.4	VO2 Máximo	8
4	Parametrizando distribuições	9
4.1	Idade	12
4.2	Peso	12
4.3	Carga Final	13
4.4	VO2 Máximo	14
5	QQPlot	14
5.1	Idade	15
5.2	Peso	15
5.3	Carga Final	16
5.4	VO2 Máximo	17
6	Teste de hipótese	17
6.1	Idade	18
6.2	Peso	18
6.3	Carga Final	19
6.4	VO2 Máximo	19
7	Análise de dependência entre as variáveis	20
8	Inferência Bayesiana	23
8.1	VO2 Máximo < 35	23
8.2	VO2 Máximo ≥ 35	24
8.3	$P[\text{VO2 Máximo} \geq 35 \mid \text{VO2 Máximo} < 35]$	24

1 Introdução

O objetivo deste trabalho é a análise de um conjunto de dados fornecidos pelo Professor Claudio Gil Soares de Araujo da CLINIMEX. Os dados possuem 4 variáveis, sendo elas idade, peso, carga final e VO2 máximo (variável objetivo que mede a capacidade aeróbica dos pacientes).

Assim, foi feito um estudo sobre as variáveis, suas distribuições e dependência entre elas. Para isso, utilizou-se Jupyter com Python 3.7 e as bibliotecas pandas, matplotlib, numpy, seaborn e scipy.

2 Histograma e Função Distribuição Empírica

Nesta etapa, foi calculado o histograma de cada variável, além de suas distribuições PDF e CDF. Para o cálculo do histograma, o tamanho do bin foi estimado a partir da regra de Sturges, definida como:

$$K = 1 + 3.322 * \log N \quad (1)$$

onde:

K = número de bins

N = número de observações

2.1 Idade

Abaixo, pode-se observar o histograma e as distribuições PDF e CDF em relação a variável idade. A partir da PDF, percebe-se que a variável possui uma distribuição semelhante à distribuição normal. Além disso, analisando o histograma nota-se que a maior faixa de valores encontra-se na faixa de 40-70 anos. Parece haver valores representativos em todas as faixas, portanto pode-se inferir que há uma grande variância mas com poucos ou nenhum outlier.

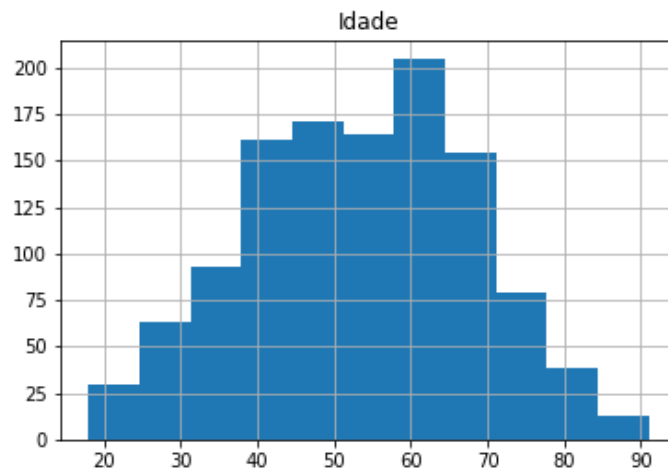
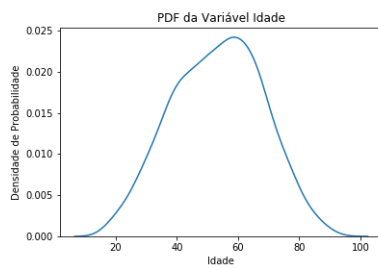
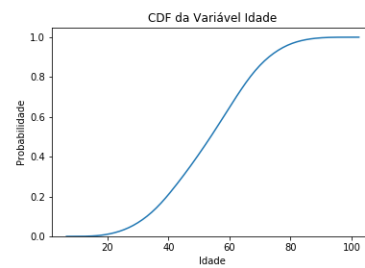


Figura 1: Histograma - Idade



(a) PDF - Idade



(b) CDF - Idade

Figura 2: Distribuições Empíricas - Idade

2.2 Peso

Abaixo, pode-se observar o histograma e as distribuições PDF e CDF em relação a variável peso. Diferentemente da idade, o peso apresenta-se mais concentrado em relação a faixa de 70-90kg, com outliers bem visíveis na faixa de 150+ kg.

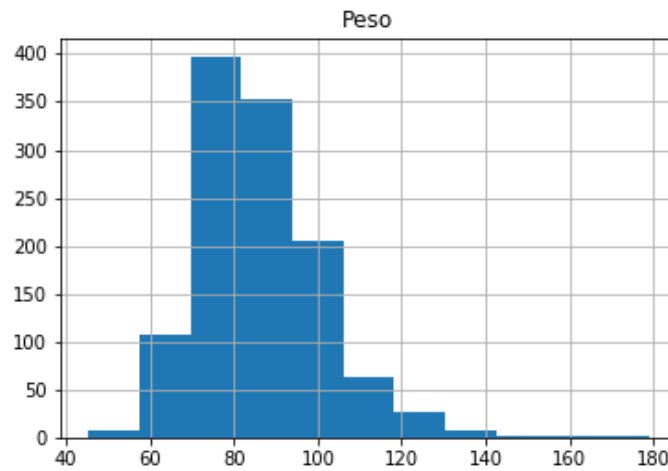
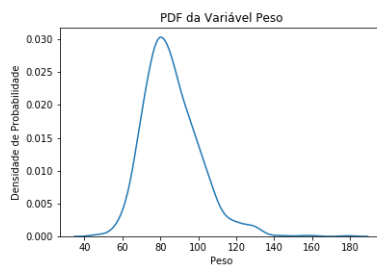
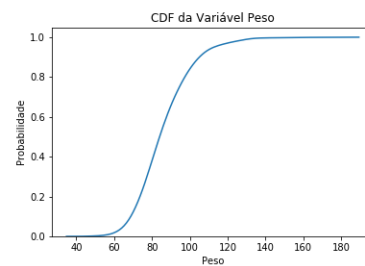


Figura 3: Histograma - Peso



(a) PDF - Peso



(b) CDF - Peso

Figura 4: Distribuições Empíricas - Peso

2.3 Carga Final

Abaixo, pode-se observar o histograma e as distribuições PDF e CDF em relação a variável carga final. Também pode-se notar aqui uma semelhança com a distribuição normal, com a maior parte dos valores de carga final entre 100-200W, cujo valor médio representa a carga esperada para um homem adulto. Os valores mais distantes, ou seja, os outliers, com uma carga final maior que 400W devem representar os atletas de alto nível.

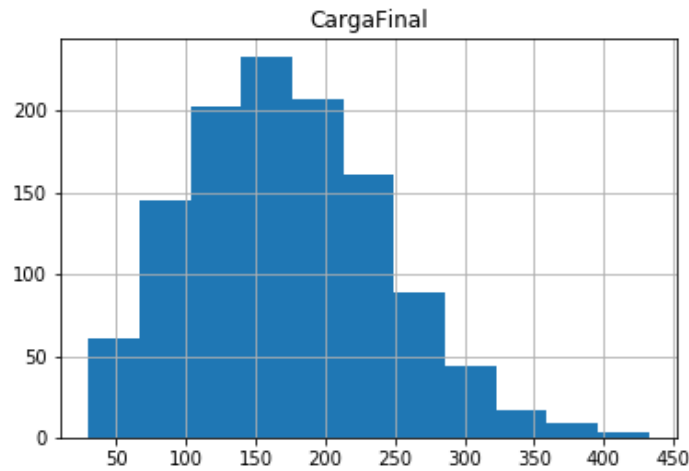
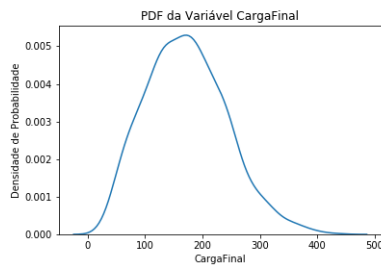
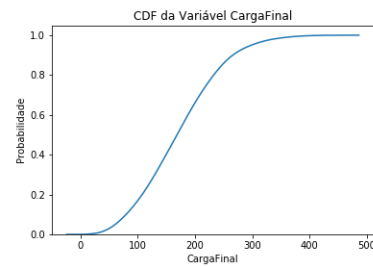


Figura 5: Histograma - Carga Final



(a) PDF - Carga Final



(b) CDF - Carga Final

Figura 6: Distribuições Empíricas - Carga Final

2.4 VO2 Máximo

Abaixo, pode-se observar o histograma e as distribuições PDF e CDF em relação a variável VO2 máximo. Novamente vemos uma distribuição parecida com a normal e, nesse caso, bem parecida com a distribuição da carga final, mostrando uma possível correlação entre as variáveis. Da mesma forma, os maiores valores representam outliers que possuem maior capacidade aeróbica, em comparação a média do homem adulto comum, em que espera-se um valor de 20-40 ml/(Kg.min).

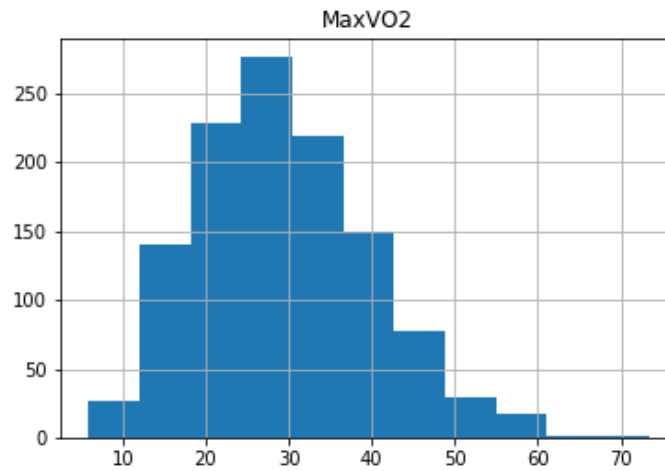
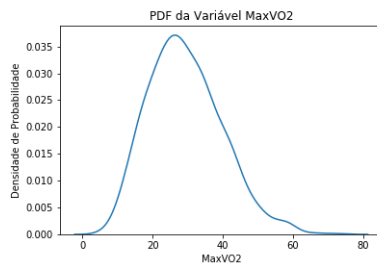
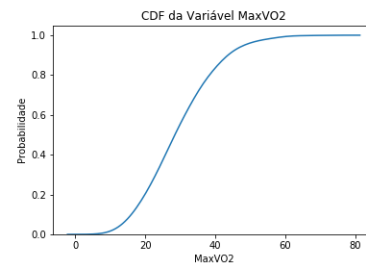


Figura 7: Histograma - VO2 Máximo



(a) PDF - VO2 Máximo



(b) CDF - VO2 Máximo

Figura 8: Distribuições Empíricas - VO2 Máximo

2.5 Análise

Com os histogramas e distribuições empíricas, percebe-se que as variáveis tem uma semelhança com a distribuição normal, variando a média e a variância de cada uma, obviamente. Mais nitidamente nesses gráficos, pôde-se observar os valores com maior e menor frequência, indicando a presença de possíveis outliers.

Contudo, não podemos tirar conclusões precipitadas apenas a partir dos gráficos, devendo ser realizada uma estatística mais profunda com resultados que nos dizem qual distribuição da literatura representa cada variável, como veremos em seguida.

3 Média, Variância e Boxplot

Nesta seção, apresenta-se a média, variância e boxplot de cada variável.

3.1 Idade

Média: 53.29

Variância: 217.45

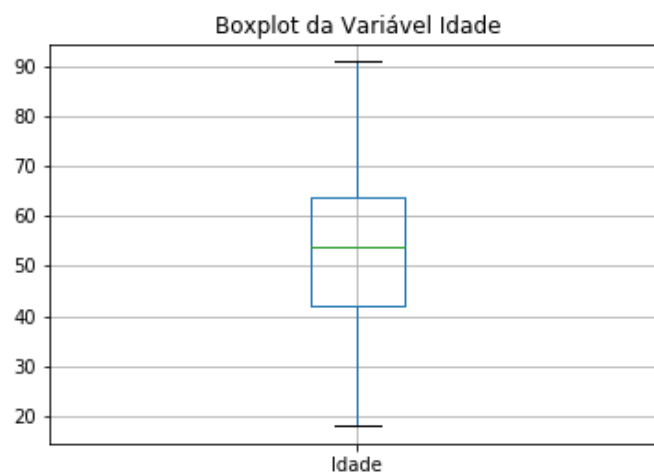


Figura 9: Boxplot - Idade

Como foi visto na seção anterior, percebe-se que há uma grande variância dos dados da variável idade, porém sem nenhum outlier, tendo em vista o boxplot. A maior parte dos valores parece estar entre 41-62 anos.

3.2 Peso

Média: 85.93

Variância: 219.01

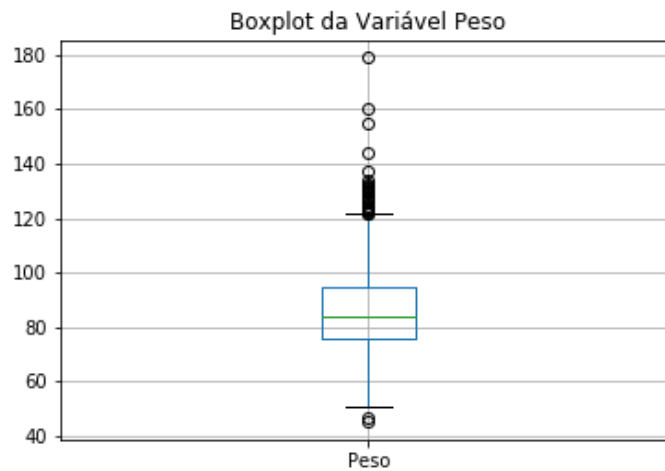


Figura 10: Boxplot - Peso

Já em relação a variável peso, pode-se notar novamente a concentração dos valores na faixa dos 80-100kg, com outliers significativos acima da faixa de 120kg, em especial uma amostra que possui 180kg. Também há um possível outlier na faixa menor que 50kg, mas acredita-se que não irá influenciar tanto na qualidade dos resultados.

3.3 Carga Final

Média: 172.27

Variância: 4913.05

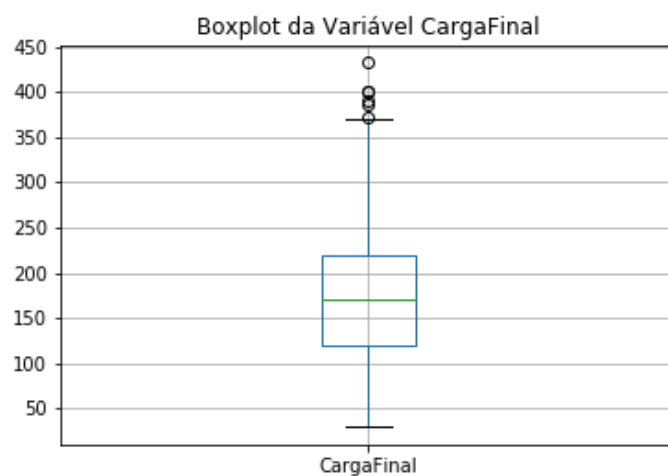


Figura 11: Boxplot - Carga Final

A carga final possui também alguns outliers, como analisamos na seção anterior, devido a possível presença de atletas de alto nível no conjunto de dados.

3.4 VO2 Máximo

Média: 29.39

Variância: 110.19

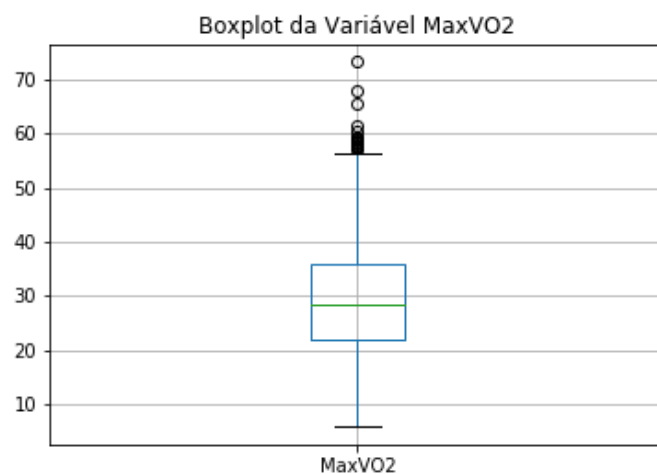


Figura 12: Boxplot - VO2 Máximo

De forma semelhante a variável carga final, os outliers representam os atletas que possuíram melhor rendimento.

4 Parametrizando distribuições

Neste item, utilizamos máxima verossimilhança para estimar os parâmetros de 4 distribuições da literatura: exponencial, normal, lognormal e weibull. Após isso, comparamos com a distribuição empírica de cada variável para verificar se alguma das distribuições da literatura pode representar essa variável.

- Exponencial

A PDF da distribuição exponencial é dada por:

$$\begin{aligned} f(x) &= \lambda \exp(-\lambda x) \quad \text{se } x \in \mathfrak{R} \\ f(x) &= 0 \quad \text{caso contrário} \end{aligned}$$

Queremos encontrar o valor do parâmetro λ que possui máxima verossimilhança em relação aos dados obtidos. Para isso, temos que a função likelihood $L(\lambda|x)$ é:

$$L(\lambda|x) = \lambda^n \exp(-\lambda \sum_{j=1}^n x_j) \quad (2)$$

A função log-likelihood:

$$l(\lambda|x) = n \ln \lambda - \lambda \sum_{j=1}^n x_j \quad (3)$$

Derivando e igualando a função acima a zero, temos que o parâmetro λ será dado por:

$$\lambda = n / \sum_{j=1}^n x_j \quad (4)$$

que representa o inverso da média amostral.

- Gaussiana

A PDF da distribuição gaussiana ou normal é dada por:

$$f(x) = (2\pi\sigma^2)^{-1/2} \exp(-1/2(x - \mu)/(\sigma^2)) \quad (5)$$

Queremos encontrar o valor dos parâmetro μ e σ^2 que possuem máxima verossimilhança em relação aos dados obtidos. Para isso, temos que a função likelihood $L(\mu, \sigma^2|x)$ é:

$$L(\mu, \sigma^2|x) = (2\pi\sigma^2)^{-n/2} \exp(-1/2\sigma^2(\sum_{j=1}^n (x_j - \mu)^2)) \quad (6)$$

A função log-likelihood:

$$l(\mu, \sigma^2|x) = -n/2 * \ln(2\pi) - n/2 * \ln(\sigma^2) - 1/2\sigma^2 \sum_{j=1}^n (x_j - \mu)^2 \quad (7)$$

Derivando e igualando a função acima a zero em relação a μ , temos que o parâmetro μ será dado por:

$$\mu = 1/n \sum_{j=1}^n x_j \quad (8)$$

que representa a média amostral.

O parâmetro σ^2 por sua vez será dado por:

$$\sigma^2 = 1/n \sum_{j=1}^n (x_j - \mu)^2 \quad (9)$$

que representa a variância amostral.

- Lognormal

A PDF da distribuição lognormal é dada por:

$$f(x) = 1/x * (2\pi\sigma^2)^{-1/2} \exp(-\ln(x - \mu)^2/(2\sigma^2)) \quad (10)$$

Queremos encontrar o valor dos parâmetro μ e σ^2 que possuem máxima verossimilhança em relação aos dados obtidos. Para isso, temos que a função likelihood $L(\mu, \sigma^2|x)$ é:

$$L(\mu, \sigma^2|x) = (2\pi\sigma^2)^{-n/2} \prod_{j=1}^n x_j^{-1} \exp(\sum_{j=1}^n (-\ln(x_j) - \mu)^2/2\sigma^2) \quad (11)$$

A função log-likelihood:

$$l(\mu, \sigma^2 | x) = -n/2 \ln(2\pi\sigma^2) - n\mu^2/2\sigma^2 - \sum_{j=1}^n \ln(x_j) - (\sum_{j=1}^n \ln(x_j)^2)/2\sigma^2 + (\sum_{j=1}^n \ln(x_j)\mu)/\sigma^2 \quad (12)$$

Derivando e igualando a função acima a zero em relação a μ , temos que o parâmetro μ será dado por:

$$\mu = (\sum_{j=1}^n \ln(x_j))/n \quad (13)$$

O parâmetro σ^2 por sua vez será dado por:

$$\sigma^2 = (\sum_{j=1}^n (\ln(x_j) - (\sum_{j=1}^n \ln(x_j))/n)^2)/n \quad (14)$$

- Weibull

A PDF da distribuição weibull é dada por:

$$f(x) = \beta/\eta (x/\eta)^{\beta-1} \exp(-(x/\eta)^\beta) \quad (15)$$

Queremos encontrar o valor dos parâmetro β e η que possuem máxima verossimilhança em relação aos dados obtidos. Para isso, temos que a função log-likelihood $l(\beta, \eta^2 | x)$ é:

$$l(\beta, \eta | x) = n \ln(\beta) - n\beta \ln(\eta) + (\beta - 1) \sum_{j=1}^n \ln(x_j) - \sum_{j=1}^n (x_j/\eta)^\beta \quad (16)$$

Derivando e igualando a função acima a zero em relação a η , temos que o parâmetro η será dado por:

$$\eta = [1/n \sum_{j=1}^n x_j^\beta]^{1/\beta} \quad (17)$$

O parâmetro β por sua vez será dado por:

$$\beta = n(1/n \sum_{j=1}^n \ln(x_j) + 1/\beta - (\sum_{j=1}^n (x_j^\beta \ln(x_j)))/\sum_{j=1}^n x_j^\beta) \quad (18)$$

4.1 Idade

Dessa forma, após parametrizar as distribuições, temos os gráficos com as distribuições da literatura e os das variáveis que queremos analisar. Abaixo segue o resultado para a variável idade.

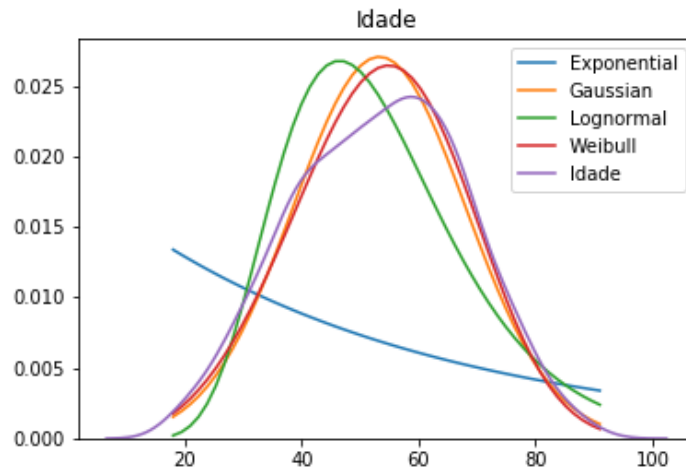


Figura 13: Distribuições - Idade

Analisando o gráfico, parece que as distribuições Weibull e Gaussiana foram as que mais se aproximaram da variável idade.

4.2 Peso

Para o peso, o resultado foi o seguinte:

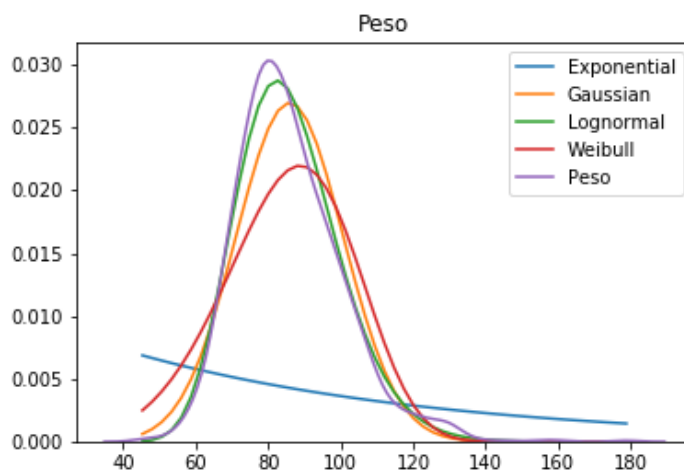


Figura 14: Distribuições - Peso

Nesse caso, a lognormal foi a que se apresentou com maior semelhança a variável peso.

4.3 Carga Final

Para a carga final, temos como resultado:

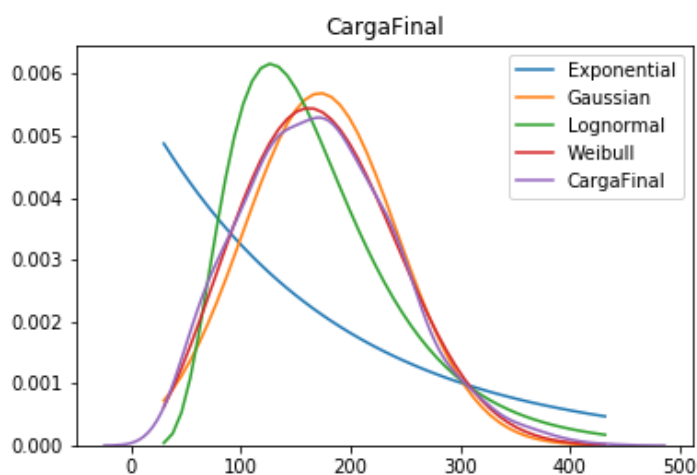


Figura 15: Distribuições - Carga Final

Nesse caso, a distribuição Weibull nitidamente é a que mais se assemelha

a carga final.

4.4 VO2 Máximo

Por fim, para o VO2 máximo temos o seguinte resultado:

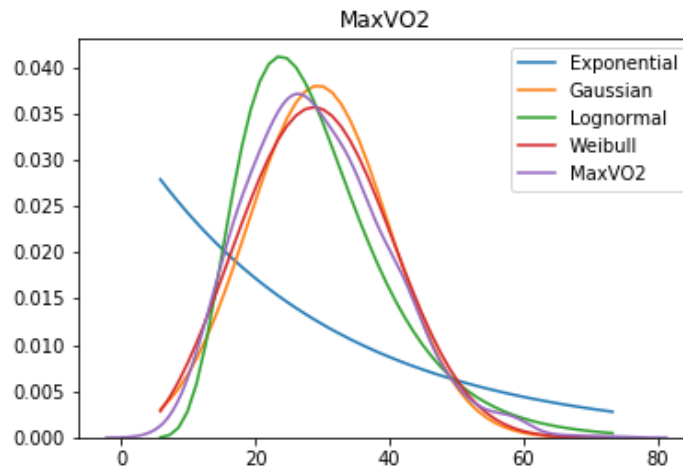


Figura 16: Distribuições - VO2 Máximo

Vemos que tanto a Gaussiana como a Weibull poderiam representar bem nossa variável VO2 Máximo, o que foi uma tendência para esse conjunto de dados. Além disso, em todas ficou bem clara a diferença da distribuição exponencial para a distribuição que queríamos encontrar.

Iremos realizar o QQPlot e o teste de hipótese para confirmar qual distribuição obteve o melhor resultado para representar nossa variável, como foi observado nos gráficos.

5 QQPlot

Nesta etapa do projeto, utilizamos o QQPlot como análise gráfica para verificar se as distribuições da literatura podem representar nossas variáveis. Nesse sentido, caso o resultado do gráfico seja aproximadamente uma linha reta, a distribuição representa bem a variável alvo.

5.1 Idade

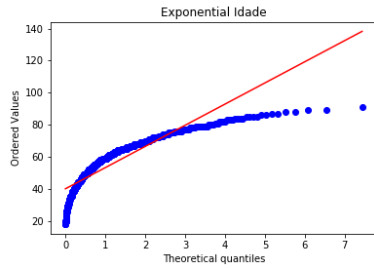


Figura 17: QQPlot - Idade - Exponencial

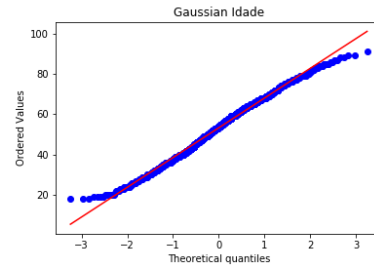


Figura 18: QQPlot - Idade - Gaussian

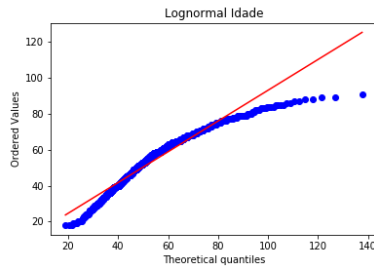


Figura 19: QQPlot - Idade - Lognormal

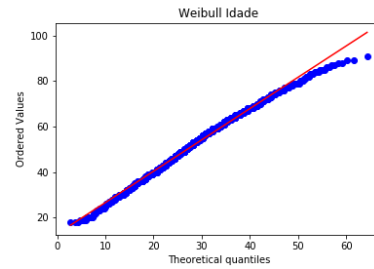


Figura 20: QQPlot - Idade - Weibull

Analisando os gráficos, pode-se perceber que a distribuição Normal e a Weibull foram as que melhor representaram a variável idade, com uma leve vantagem para a Weibull.

5.2 Peso

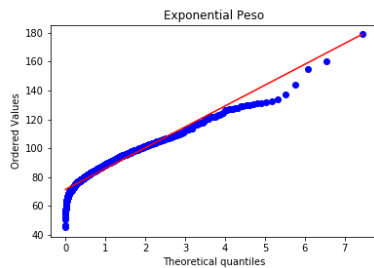


Figura 21: QQPlot - Peso - Exponencial

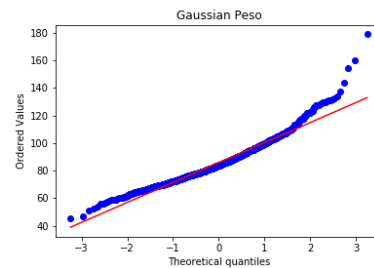


Figura 22: QQPlot - Peso - Gaussian

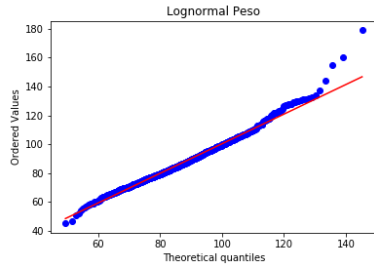


Figura 23: QQPlot - Peso - Log-normal

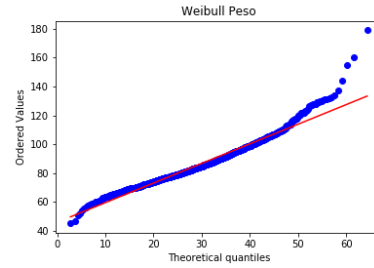


Figura 24: QQPlot - Peso - Weibull

Em relação a variável peso, as distribuições tiveram um resultado parecido, com exceção da exponencial. A lognormal parece ter obtido maior semelhança, seguida da Weibull.

5.3 Carga Final

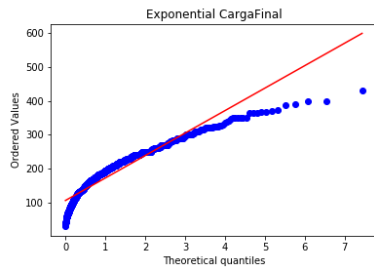


Figura 25: QQPlot - Carga Final - Exponencial

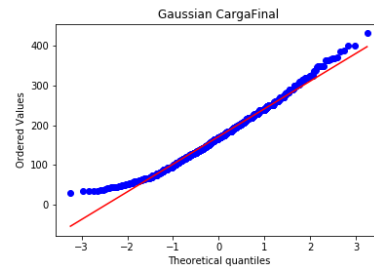


Figura 26: QQPlot - Carga Final - Gaussian

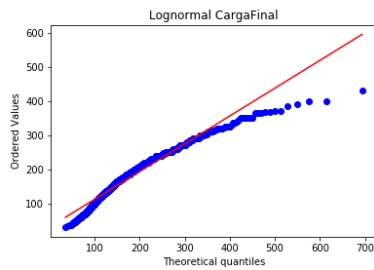


Figura 27: QQPlot - Carga Final - Lognormal

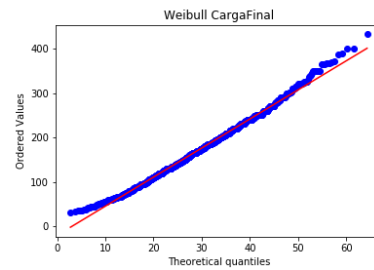


Figura 28: QQPlot - Carga Final - Weibull

Para a carga final, a Weibull se destacou em relação às demais, apresentando o melhor resultado. A Gaussiana também não obteve um mal resultado.

5.4 VO2 Máximo

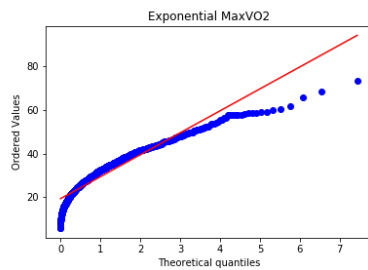


Figura 29: QQPlot - VO2 Máximo - Exponencial

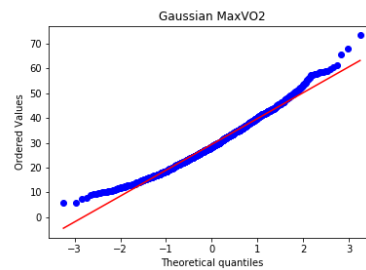


Figura 30: QQPlot - VO2 Máximo - Gaussian

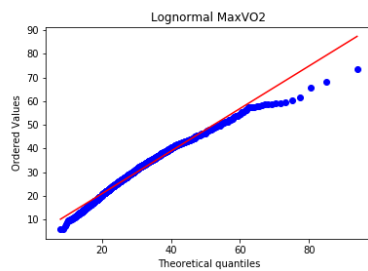


Figura 31: QQPlot - VO2 Máximo - Lognormal

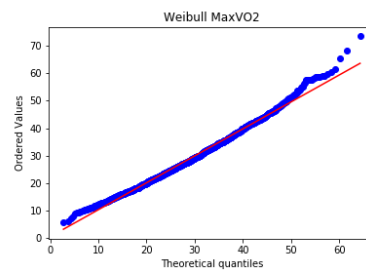


Figura 32: QQPlot - VO2 Máximo - Weibull

Por fim, como já havíamos verificado também a possível existência de correlação entre as variáveis carga final e VO2 máximo, nesse caso a distribuição Weibull também apresentou um melhor resultado em relação às demais.

Assim, pudemos ver que o gráfico com as PDFs pôde ser um bom indicativo de qual distribuição melhor representava a variável, como visto nos QQPlots acima.

6 Teste de hipótese

Nesta parte, utilizou-se o teste de Komolgorov-Sminorv para verificar a igualdade das distribuições. Além disso, o teste nos retorna o respectivo

p-valor. Adotando um $\alpha = 0.05$, iremos aceitar as hipóteses que possuem p-valor maior que α e rejeitá-la caso contrário.

A estatística é dada pela fórmula:

$$D_n = \sup_x |F_n(x) - F(x)| \quad (19)$$

Em que \sup_x é o supremo do conjunto de distâncias. Quando n tende ao infinito D_n tende a zero se a amostra vier da distribuição $F(x)$. Ou seja, queremos encontrar a distribuição que possui o menor valor de D_n .

6.1 Idade

- Exponencial

$$D_n = 0.373$$

$$\text{p-valor} = 1.109\text{e-}146$$

- Gaussiana

$$D_n = 0.044$$

$$\text{p-valor} = 0.02$$

- Lognormal

$$D_n = 0.0847$$

$$\text{p-valor} = 9.07\text{e-}08$$

- Weibull

$$D_n = 0.033$$

$$\text{p-valor} = 0.15$$

Portanto, a única hipótese aceita seria a da distribuição Weibull, que também apresentou o menor valor de D_n .

6.2 Peso

- Exponencial

$$D_n = 0.495$$

$$\text{p-valor} = 3.359\text{e-}266$$

- Gaussiana

$$D_n = 0.0666$$

$$\text{p-valor} = 5.758\text{e-}05$$

- Lognormal

$$D_n = 0.0323$$

$$\text{p-valor} = 0.17$$

- Weibull

$$D_n = 0.103$$

$$\text{p-valor} = 2.523\text{e-}11$$

Para a variável peso, a única distribuição aceita foi a Lognormal, com um p-valor de 0.17 e um D_n de 0.0323.

6.3 Carga Final

- Exponencial

$$D_n = 0.286$$

$$\text{p-valor} = 1.1721\text{e-}85$$

- Gaussiana

$$D_n = 0.0392$$

$$\text{p-valor} = 0.0527$$

- Lognormal

$$D_n = 0.0803$$

$$\text{p-valor} = 4.962\text{e-}07$$

- Weibull

$$D_n = 0.0245$$

$$\text{p-valor} = 0.479$$

Para a carga final, aceitamos as hipóteses das distribuições Gaussiana e Weibull, sendo a Weibull a que obteve melhor resultado no teste.

6.4 VO2 Máximo

- Exponencial

$$D_n = 0.335$$

$$\text{p-valor} = 1.059\text{e-}117$$

- Gaussiana

$$D_n = 0.0445$$

$$\text{p-valor} = 0.019$$

- Lognormal

$$D_n = 0.0405$$

$$\text{p-valor} = 0.0411$$

- Weibull

$$D_n = 0.0367$$

$$\text{p-valor} = 0.082$$

Por fim, somente a distribuição Weibull seria aceita no teste para a carga final, apresentando o melhor resultado.

Em geral, a Weibull pareceu ter uma maior correspondência com as variáveis, com exceção do peso que obteve um comportamento diferente das demais.

7 Análise de dependência entre as variáveis

Para a análise de dependência entre variáveis, temos o gráfico da matriz de correlação, junto com seus respectivos coeficientes de correlação e o scatter plot.

A matriz de correlação é a seguinte:

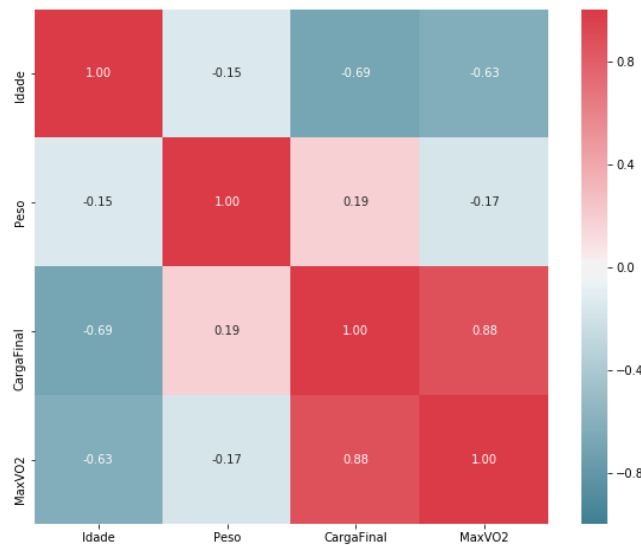


Figura 33: Matriz de Correlação

Logo de cara podemos ver a alta correlação entre as variáveis carga final e VO2 máximo. Além disso, há uma correlação negativa entre as variáveis VO2 máximo e Idade e VO2 máximo e Peso. Essa última, porém, não muito significativa.

Os valores encontrados para os coeficientes de correlação foram os seguintes:

- Idade - -0.630
- Peso - -0.174
- Carga Final - 0.878

E os respectivos scatter plots:

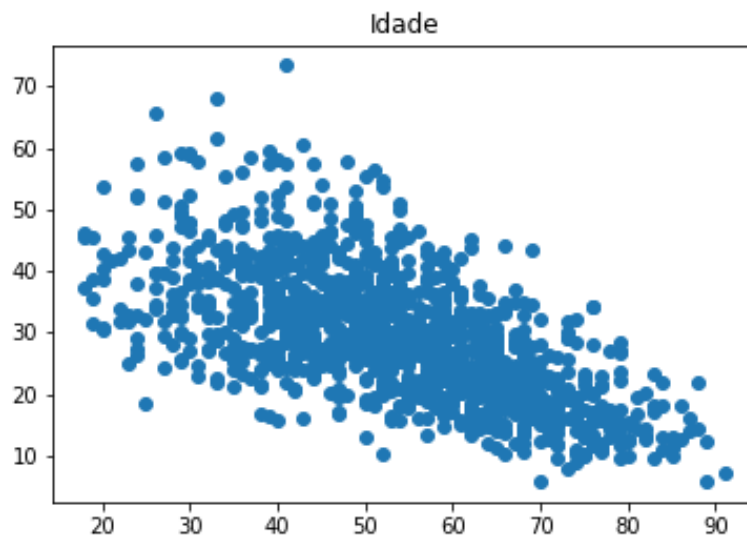


Figura 34: Scatter Plot - Idade x VO2 Máximo

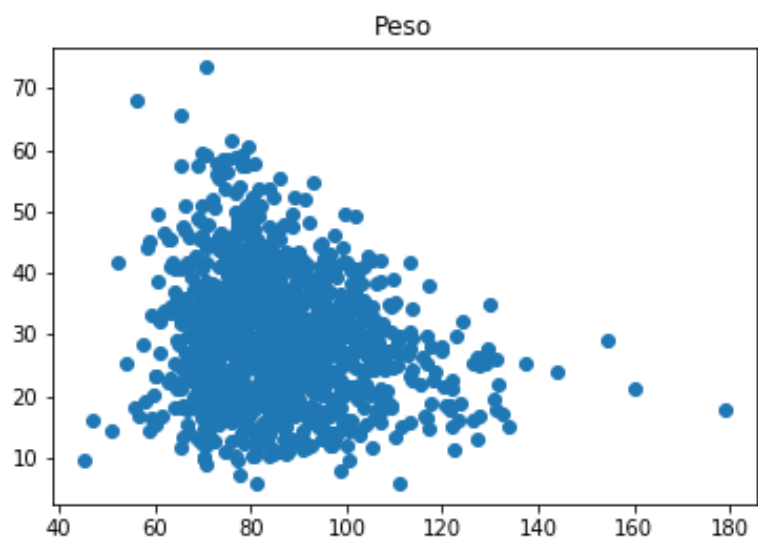


Figura 35: Scatter Plot - Peso x VO2 Máximo

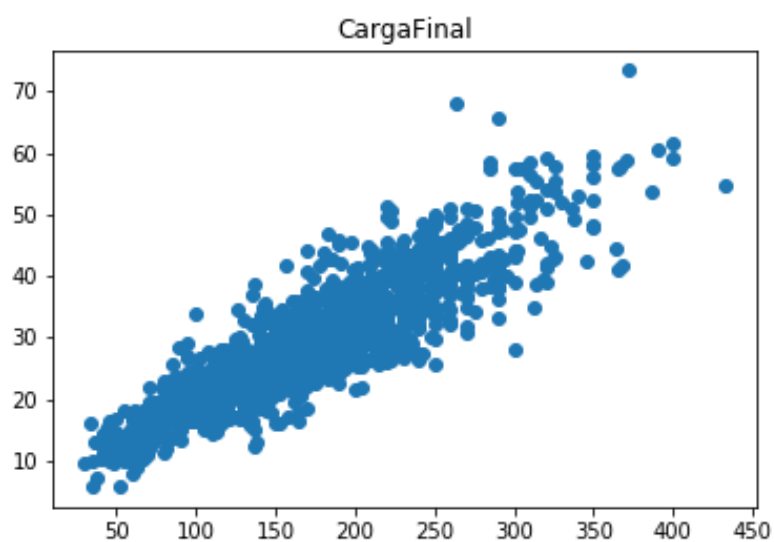


Figura 36: Scatter Plot - Carga Final x VO2 Máximo

Portanto, os scatter plots representam bem a correlação positiva e negativa entre as variáveis, como visto anteriormente, além da pouca correlação no caso do peso.

Já havíamos notado esse comportamento no decorrer do projeto, mas pudemos quantificar essa correlação agora. Faz sentido a carga máxima atingida pelo paciente ter uma alta correlação com a sua capacidade aeróbica, assim como sua idade influenciar negativamente na mesma.

8 Inferência Bayesiana

Para a tabela de inferência bayesiana, dividimos os valores de carga final em 8 faixas de valores, sendo o menor valor 30 e o maior valor 432, variando de 50.25 em 50.25. A carga final foi escolhida por ter o maior valor de correlação amostral com a variável VO2 máximo. Foram calculadas duas tabelas, uma para a que VO2 máximo < 35 e outra em que VO2 máximo ≥ 35 . Por fim, foi feita uma tabela em que utilizamos a posterior encontrado na tabela que VO2 máximo ≥ 35 para realizar uma previsão de melhora no condicionamento do paciente, ou seja, $P[\text{VO2 Máximo} \geq 35 \mid \text{VO2 Máximo} < 35]$.

Nessas tabelas, a prior representa a frequência dos valores de carga final naquela faixa considerada como hipótese e a likelihood a probabilidade de se obter um certo valor de VO2 máximo para aquela faixa.

8.1 VO2 Máximo < 35

Para o caso em que VO2 Máximo < 35 , observamos pela likelihood que temos 100% de probabilidade de encontrar esse valor na faixa de 30.0 - 130.5 de carga final e 0% na faixa de 331.5 - 432. Ou seja, não há nenhum paciente que possui carga final maior que 331.5 e possui um VO2 máximo menor que 35, o que faz sentido. Assim, calculamos a posterior e encontramos dado que VO2 Máximo < 35 , temos maior probabilidade de encontrar a hipótese da faixa de valores de carga final entre 130.5 e 180.75.

Hipótese	Prior	Likelihood	Numerador de Bayes	Posterior
(30.0, 80.25)	0.098	1.00	0.098	0.136
(80.25, 130.5)	0.211	1.00	0.211	0.292
(130.5, 180.75)	0.254	0.93	0.237	0.329
(180.75, 231.0)	0.241	0.62	0.148	0.206
(231.0, 281.25)	0.128	0.19	0.025	0.034
(281.25, 331.5)	0.049	0.05	0.003	0.004
(331.5, 381.75)	0.015	0.00	0.000	0.000
(381.75, 432.0)	0.003	0.00	0.000	0.000

Tabela 1: Inferência Bayesiana para VO2 Máximo < 35

8.2 VO2 Máximo ≥ 35

Para o caso em que VO2 Máximo ≥ 35 , observamos pela likelihood que temos 0% de probabilidade de encontrar esse valor na faixa de 30.0 - 130.5 de carga final e 100% na faixa de 331.5 - 432. Ou seja, não há nenhum paciente que possui carga final menor que 130.5 e possui um VO2 máximo maior que 35, o que faz sentido. Assim, calculamos a posterior e encontramos dado que VO2 Máximo ≥ 35 , temos maior probabilidade de encontrar a hipótese da faixa de valores de carga final entre 231.0 e 281.25.

Hipótese	Prior	Likelihood	Numerador de Bayes	Posterior
(30.0, 80.25)	0.098	0.00	0.00	0.00
(80.25, 130.5)	0.211	0.00	0.00	0.00
(130.5, 180.75)	0.254	0.07	0.017	0.06
(180.75, 231.0)	0.241	0.38	0.092	0.33
(231.0, 281.25)	0.128	0.81	0.103	0.37
(281.25, 331.5)	0.049	0.95	0.047	0.17
(331.5, 381.75)	0.015	1.00	0.015	0.05
(381.75, 432.0)	0.003	1.00	0.003	0.01

Tabela 2: Inferência Bayesiana para VO2 Máximo ≥ 35

8.3 $P[\text{VO2 Máximo} \geq 35 \mid \text{VO2 Máximo} < 35]$

Nessa tabela, foi feita a predição de encontrar um VO2 Máximo ≥ 35 , dado que um paciente possuía VO2 Máximo < 35 . Com os resultados, verificamos que é mais provável que um paciente aumente seu valor de VO2 máximo para acima da média, caso esteja na faixa de (180.75-230.0).

Hipótese	Prior	Likelihood1	Numerador de Bayes	Posterior1	Likelihood2	Prediction
(30.0, 80.25)	0.098	1.00	0.098	0.136	0.00	0.00
(80.25, 130.5)	0.211	1.00	0.211	0.292	0.00	0.00
(130.5, 180.75)	0.254	0.93	0.237	0.329	0.07	0.02
(180.75, 231.0)	0.241	0.62	0.148	0.206	0.38	0.08
(231.0, 281.25)	0.128	0.19	0.025	0.034	0.81	0.03
(281.25, 331.5)	0.049	0.05	0.003	0.004	0.95	0.00
(331.5, 381.75)	0.015	0.00	0.000	0.000	0.00	0.00
(381.75, 432.0)	0.003	0.00	0.000	0.000	0.00	0.00

Tabela 3: Inferência Bayesiana para $P[\text{VO2 Máximo} \geq 35 \mid \text{VO2 Máximo} < 35]$

9 Conclusão

Neste projeto, pôde-se fazer uma análise de um conjunto de dados médicos, sendo possível avaliar as distribuições das características dos pacientes, além da correlação entre essas variáveis. Assim, verificou-se quais fatores influenciam ou não na capacidade aeróbica dos pacientes, sendo de grande utilidade os resultados para a análise médica.