

## **Regressão por componentes principais para identificar os fatores de impacto na expectativa de vida no ranking das cidades com melhores índices de hábitos saudáveis**

Lucas Dutra Mendes<sup>1\*</sup>; Juliano Domingues da Silva<sup>2</sup>

<sup>1</sup>Optel Group. Bilingual Technical Support - Product Specialist. Rua James Clerk Maxwell, 280, Módulo 08 – Techno Park; 13069-380 Campinas, São Paulo, Brasil

<sup>2</sup>Universidade Estadual de Maringá. Professor Adjunto. Avenida Colombo, 5790 – Jardim Universitário; 87020-900 Maringá, Paraná, Brasil

\*autor correspondente: [lucas.dutra.mendes@gmail.com](mailto:lucas.dutra.mendes@gmail.com)

## **Regressão por componentes principais para identificar os fatores de impacto na expectativa de vida no ranking das cidades com melhores índices de hábitos saudáveis**

### **Resumo**

O presente trabalho buscou estudar as cidades que possuem os melhores índices de hábitos de vida saudáveis, com o objetivo de entender os fatores que impactam diretamente na expectativa de vida de uma população, sendo que a base de dados estudada, foi criada pelo time da Lenstore e contém quarenta e quatro cidades ao redor do mundo, ranqueadas com os melhores índices de estilo de vida saudável. A análise apontou que a sobrecarga do cotidiano e hábitos saudáveis possuem fator predominante na expectativa de vida da população. Dentre os modelos utilizados, estão; Regressão Linear, Análise por componentes principais [PCA] e Regressão por componentes principais [PCR]. Os principais resultados encontrados na pesquisa são que a sobrecarga do cotidiano, gerada pelo excesso de trabalho e poluição, afetam negativamente a expectativa de vida enquanto os hábitos saudáveis afetam positivamente a expectativa de vida. Essa pesquisa contribui para apontar quais fatores os agentes públicos devem priorizar para melhorar a qualidade de vida da população. **Palavras-chave:** PCA; PCR; Box-Cox; Regressão Linear; Expectativa de Vida

### **Introdução**

Nos dias atuais, muito se fala sobre a qualidade de vida presente no mundo em que vivemos, fazendo com que este tema possa ser assimilado diferentemente por cada um de nós. Ter uma qualidade de vida ou estilo de vida saudável pode ser referenciado como uma árvore, contendo diversos ramos e, cada um destes possuindo a sua importância dentro deste tema, podendo ser no âmbito social, biológico, econômico, político, humano, bem-estar, entre outros (Almeida et al., 2012). A perspectiva de vida segundo Minayo et al. (2000) pode ser compreendida de certa forma por condições elementares, como acesso as necessidades básicas, como água potável, habitação, saúde, lazer e alimentação. Segundo Fonseca et al. (2019), doenças crônicas ligadas a hábitos não saudáveis, dentre elas câncer, diabetes e hipertensão, que somente no Brasil correspondem a 72% das causas de morte e observasse a classe baixa como a mais afetada. Percebemos, portanto a importância de conhecermos e estudarmos tais comportamentos, especialmente para compreender os comportamentos significantes que definem a qualidade de vida e expectativa de vida de uma população.

A base de dados a ser estudada neste trabalho (ver <https://www.lenstore.co.uk/research/healthy-lifestyle-report/>), possui as seguintes observações de quarenta e quatro cidades, listadas como as melhores em estilo de vida saudáveis, sendo: horas de sol que cada cidade recebe por ano, custo de uma garrafa de água de 300ml, índice de obesidade do país que cada cidade faz parte, expectativa de vida do país, poluição que cada cidade produz, média anual de horas trabalhadas no país, índice de felicidade por país e atividades ao ar livre nas cidades, quantidade de restaurantes que fazem entrega de refeições e custo mensal de academia de ginástica.

Contudo, observasse a importância de estudar e entender os fatores que expliquem a expectativa de vida de uma população, sobretudo em cidades com os melhores índices de hábitos saudáveis, desta maneira será possível identificar os fatores preponderantes, que por sua vez podem ser implementados através de políticas públicas em cidades fora do ranking estudado.

Destacando a importância da tecnologia, aliada a captação de dados, matemática e capacidade de análise e processamento destes, para que os padrões possam ser estudados, possibilitando que análises possam ser realizadas, auxiliando nas tomadas de decisões. Norteando este trabalho para as seguintes perguntas: Quais os fatores importantes e comuns entre as melhores cidades ranqueadas e quais variáveis impactam diretamente na expectativa de vida de uma população? Para tanto, o objetivo dessa pesquisa é analisar o impacto dos indicadores de qualidade de vida das melhores cidades para se viver constante no “healthy lifestyle cities report 2021” na qualidade de vida da população.

## **Material e Métodos**

A metodologia deste trabalho consiste na análise qualitativa e quantitativa da base de dados “healthy lifestyle cities report 2021” criados pelo time da Lenstore onde quarenta e quatro cidades ao redor do mundo foram analisadas e ranqueadas com os melhores índices de estilo de vida saudável. Preparação dos dados através de data wrangling, que é a arte de importar os dados para o R e os deixar prontos para serem visualizados e modelados (Wickham e Grolemund, 2017).

Seguindo com a metodologia deste trabalho, também foi utilizado a Regressão Linear Múltipla como objetivo de entender as variáveis que explicam a expectativa de vida, teste de Shapiro-Francia para analisar a aderência dos resíduos a normalidade, teste variance inflation value [VIF], para verificar a presença de multicolinearidade, teste de Breusch-Pagan para identificar a presença de heterocedasticidade, normalização da variável Y por Box-Cox para adequar os resíduos a normalidade e tratar o fenômeno da heterocedasticidade no modelo de regressão, Análise por Componentes Principais [PCA], com o objetivo de tratar o fenômeno da multicolinearidade, agrupando as variáveis e capturando observações não latentes na base estudada, e, por fim, Regressão por Componentes Principais [PCR] com a premissa de obter os fatores preponderantes obtidos na PCA e que expliquem a expectativa de vida.

A Tabela 1 ilustra as variáveis a serem estudadas na base de dados, com uma contextualização de cada variável e a forma como cada uma delas é mensurada.

Tabela 1. Conceito das variáveis utilizadas na pesquisa.

Variável	Contextualização	Medida	Fonte
Horas de sol	Quantidade de exposição solar está ligada com a saúde física, mental e na prevenção de doenças.	Média de horas de exposição solar da cidade em um ano.	<a href="https://www.pucrs.br/blog/efeitos-e-beneficios-da-exposicao-luz-solar-para-imunidade/">https://www.pucrs.br/blog/efeitos-e-beneficios-da-exposicao-luz-solar-para-imunidade/</a>
Custo garrafa de água	O acesso a água sendo uma necessidade básica, impacta diretamente na saúde de uma população. Quanto menor o acesso, maiores os problemas de saúde.	£ / 300 ml	Não possui fonte descrita. <a href="https://www.numbeo.com/cost-of-living/">https://www.numbeo.com/cost-of-living/</a>
Índice de obesidade	Quanto maiores os níveis de obesidade, menores os hábitos saudáveis e diminuição da expectativa de vida	Porcentagem por país.	<a href="https://ourworldindata.org/obesity">https://ourworldindata.org/obesity</a>
Expectativa de vida	Expectativa de vida da população de cada país.	Anos de vida por país.	<a href="https://shorturl.ae/BbMVp">https://shorturl.ae/BbMVp</a>
Índice de poluição	Níveis de poluição altos impactam negativamente na vida de uma comunidade.	Índice de poluição por cidade, quanto maior, pior.	<a href="https://www.numbeo.com/pollution/rankings.jsp">https://www.numbeo.com/pollution/rankings.jsp</a>
Horas anuais de trabalho	Quanto mais horas trabalhadas, menos tempo para lazer e atividades físicas uma pessoa possui, interferindo em sua qualidade de vida.	Horas ano por cidade.	<a href="https://data.oecd.org/emp/hours-worked.htm">https://data.oecd.org/emp/hours-worked.htm</a>
Índice de felicidade	Quanto maior o índice, mais contente a população se encontra no país que vive.	Índice quanto maior melhor, por país	<a href="https://worldhappiness.report/ed/2021/">https://worldhappiness.report/ed/2021/</a>
Atividades ao ar livre por cidade.	Atividades ao ar livre proporcionadas pelas cidades observadas.	Número de atividades proporcionadas por cidade.	<a href="https://www.lawnstarter.com/blog/studies/best-cities-for-spring/">https://www.lawnstarter.com/blog/studies/best-cities-for-spring/</a>
Comida por delivery	Restaurantes disponíveis para entrega de comida, mostrando a facilidade no acesso a alimentação.	Número de restaurantes por cidade	Não possui fonte descrita.
Academia Preço	Acessibilidade a atividades físicas em locais fechados.	Mensal £ por cidade	Não possui fonte descrita.

Fonte: Dados originais obtidos na pesquisa.

## Resultados e Discussão

### Data Wrangling

Como as variáveis observadas estão em inglês e em um formato não muito amigável de leitura e entendimento, a organização dos dados é o primeiro passo a ser seguido. Com

auxílio da linguagem R, os nomes das variáveis foram renomeados para o português e foram utilizados nomes mais curtos para facilitar a compreensão e análise. Um ponto a se observar são as variáveis “Água” e “Academia”, onde ambas possuem o símbolo “£” antes de cada observação. Este símbolo foi removido e estas variáveis foram convertidas para o dólar americano. Já a variável obesidade possui o símbolo % antes de cada observação e este também foi removido. Seguindo em frente com a manipulação dos dados, o passo seguinte é o preenchimento dos valores faltantes, também conhecido como missing values. Primeiro valor a ser preenchido é na cidade Fukuoka para a variável poluição. Este dado não pode ser encontrado nas values estão concentrados na variável horas trabalhadas por ano, para as cidades: Johannesburg referências da base de dados e, portanto, foi utilizado o dado da cidade de Osaka, que está aproximadamente 600 quilômetros de distância de Fukuoka e possui o índice de poluição de 53,24. A variável horas de sol possui um dado faltante para a cidade de Geneva, no qual foi obtido nas referências da base de dados o valor de 1887 horas por ano de sol. Os demais missing, 2189 horas - São Paulo, 1706 horas - Shanghai e Beijing, 2168 horas - Hong Kong, 2148 horas - Mumbai, 2122 horas - Taipei, 2085 horas - Jakarta, 2018 horas - Buenos Aires, 1606 horas – Bangkok, 2090 horas. Todos os dados foram encontrados nas referências da base com exceção as horas trabalhadas anuais na cidade do Cairo, onde foi utilizada o dado da cidade de Johannesburgo, sendo 2189 horas anuais. A Figura 1 abaixo ilustra a base de dados após aplicação do Data Wrangling.

Posicao	Cidades	Sol	Agua	Obesidade	Vida	Poluicao	Trabalhadas	Felicidade	Atividades	Restaurantes	Academia
1	1 Amsterdam	1858	2.50	20.4	81.2	30.93	1434	7.44	422	1048	45.37
2	2 Sydney	2636	1.92	29.0	82.1	26.86	1712	7.22	406	1103	54.16
3	3 Vienna	1884	2.52	20.1	81.0	17.33	1501	7.29	132	1008	33.46
4	4 Stockholm	1821	2.24	20.6	81.8	19.63	1452	7.35	129	598	48.50
5	5 Copenhagen	1630	2.85	19.7	79.8	21.24	1380	7.64	154	523	42.29
6	6 Helsinki	1662	2.08	22.2	80.4	13.08	1540	7.80	113	309	45.80
7	7 Fukuoka	2769	1.01	4.3	83.2	53.24	1644	5.87	35	539	72.63
8	8 Berlin	1626	2.02	22.3	80.6	39.41	1386	7.07	254	1729	33.94
9	9 Barcelona	2591	1.55	23.8	82.2	65.19	1686	6.40	585	2344	49.14
10	10 Vancouver	1938	1.40	29.4	81.7	24.26	1670	7.23	218	788	40.35
11	11 Melbourne	2363	2.04	29.0	82.1	25.90	1712	7.22	243	813	47.96
12	12 Beijing	2671	0.34	6.2	75.4	85.43	2168	5.12	223	261	50.21

Figura 1. Base de dados após alterações  
Fonte: Dados originais obtidos na pesquisa

## Regressão Linear Múltipla

A análise inicia-se com o mapa de correlação de Pearson entre as variáveis preditoras, onde a variável expectativa de vida, a ser estudada e, portanto, removida do mapa de correlações. A Figura 2, ilustra o mapa de correlações entre as variáveis preditoras onde

podemos observar fortes correlações em modulo acima de 0.5 entre múltiplas variáveis, sendo está uma forte característica da presença de multicolinearidade.

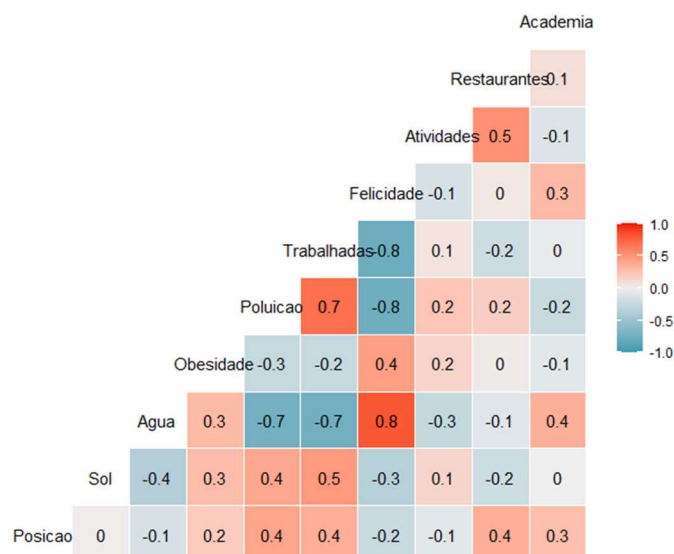


Figura 2. Matriz de correlações

Fonte: Resultados originais obtidos na pesquisa

Gerando o modelo de regressão linear múltipla com a variável expectativa de vida como Y (dependente ou resposta) e as demais como variáveis X (independes ou preditoras), foi obtido seguinte output, ilustrado na Figura 3.

```

Residuals:
    Min       1Q   Median       3Q      Max
-10.223  -1.919   1.146   1.905   5.271

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  77.6525192  12.7285857   6.101 6.38e-07 ***
sol          -0.0005499   0.0012888  -0.427  0.67230
Agua         -1.4237850   1.1542117  -1.234  0.22583
Obesidade    -0.0889106   0.0720917  -1.233  0.22592
Poluicao       0.0510994   0.0435259   1.174  0.24855
Trabalhadas  -0.0129272   0.0047077  -2.746  0.00957 **
Felicidade    3.1876377   1.1782964   2.705  0.01059 *
Atividades    0.0057991   0.0050368   1.151  0.25763
Restaurantes -0.0007419   0.0005729  -1.295  0.20409
Academia      0.1028749   0.0365539   2.814  0.00807 **
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.267 on 34 degrees of freedom
Multiple R-squared:  0.7001,    Adjusted R-squared:  0.6207
F-statistic:  8.82 on 9 and 34 DF,  p-value: 1.073e-06
    
```

Figura 3. Summary Modelo Regressão Linear Múltipla.

Fonte: Resultados originais obtidos na pesquisa

O summary do modelo exibe que o p-valor da distribuição do teste F é estatisticamente diferente de 0 a 95% de confiança, sendo, portanto, estatisticamente significativo para prever o fenômeno estudado. Dentre as variáveis estudadas, apenas o intercepto, horas trabalhadas, Índice de Felicidade e Academia se mostraram estatisticamente significantes.

A Figura 4 apresenta o teste de verificação da aderência dos resíduos a normalidade de Shapiro-Francia, onde para confirmar a hipótese nula  $H_0$ , e, portanto, a aderência dos dados a normalidade, o p-value deve ser superior a 0,05, no entanto, o p-value é inferior a este valor, indicando que a distribuição dos dados não segue distribuição normal.

#### Shapiro-Francia normality test

```
data: OLS_vida$residuals
W = 0.9182, p-value = 0.005673
```

Figura 4. Shapiro-Francia Teste de Normalidade.

Fonte: Resultados originais obtidos na pesquisa

### Regressão Não Linear Múltipla

Para tratar os resíduos não aderentes a normalidade, a variável Y foi normalizada através da transformação de BOX-COX e o valor de lambda gerado para o modelo BOX-COX é de 11.72205. O summary deste modelo, apresentado na Figura 5, mostra uma insignificante melhora no  $R^2$  ajustado, apenas duas variáveis estatisticamente significantes e o teste de verificação da aderência dos resíduos a normalidade de Shapiro-Francia, Figura 6, indica que os dados são aderentes a normalidade, após normalização da variável Y.

```
Residuals:
    Min       1Q   Median       3Q      Max
-7.773e+20 -3.287e+20 -3.284e+19  3.227e+20  8.519e+20
```

#### Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.221e+21	1.816e+21	1.223	0.2299
sol	2.373e+17	1.839e+17	1.290	0.2056
Agua	4.276e+19	1.647e+20	0.260	0.7967
obesidade	-2.172e+19	1.029e+19	-2.111	0.0422 *
Poluicao	-3.769e+18	6.211e+18	-0.607	0.5479
Trabalhadas	-1.555e+18	6.717e+17	-2.315	0.0268 *
Felicidade	2.362e+20	1.681e+20	1.405	0.1691
Atividades	4.871e+17	7.187e+17	0.678	0.5025
Restaurantes	1.295e+16	8.175e+16	0.158	0.8750
Academia	1.055e+19	5.216e+18	2.022	0.0511 .

---

signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```
Residual standard error: 4.661e+20 on 34 degrees of freedom
Multiple R-squared: 0.7073, Adjusted R-squared: 0.6298
F-statistic: 9.129 on 9 and 34 DF, p-value: 7.348e-07
```

Figura 5. Summary Modelo BOX-COX.

Fonte: Resultados originais obtidos na pesquisa

### shapiro-Francia normality test

```
data: OLS_BC_vida$residuals
W = 0.98388, p-value = 0.6962
```

Figura 6. Shapiro-Francia teste de normalidade.  
Fonte: Resultados originais obtidos na pesquisa

### Multicolinearidade

O teste VIF (Variance Inflation Factor), que pode ser definido por  $VIF = \frac{1}{Tolerance}$ , onde Tolerancia =  $1 - R^2$ , utilizado para diagnosticar a presença de multicolinearidade no modelo, sendo que o output deste teste é o mesmo para o modelo linear e modelo BOX-COX, pois não houve transformação nas variáveis X. De acordo com (Fávero e Belfiore, 2017), muitos autores afirmam que VIF acima de dez indicam problemas de multicolinearidade, no entanto, um VIF igual ou maior que quatro já representa um valor elevado para este teste. Ao verificar a Figura 7, percebe-se VIF elevado, próximos ou superiores à quatro para quatro variáveis preditoras, indicando, portanto, a presença de uma forte correlação elas.

	Variables	VIF
1	sol	2.124419
2	Agua	4.685874
3	obesidade	2.177049
4	Poluicao	3.562757
5	Trabalhadas	5.340954
6	Felicidade	5.496669
7	Atividades	1.653832
8	Restaurantes	2.551331
9	Academia	2.049322

Figura 7. VIF Modelo Linear & BOX-COX.  
Fonte: Resultados originais obtidos na pesquisa

### Heterocedasticidade

A Figura 8 mostra o output do teste Breusch-Pagan, utilizado para verificar a presença de heterocedasticidade nos modelos estudados até o momento. Este teste avalia a correlação entre as variáveis preditoras em relação aos termos de erro dos modelos. A hipótese nula,  $H_0$ , para p-value inferiores a 0,05 indica a presença de heterocedasticidade, uma vez que a hipótese alternativa  $H_1$ , para valores acima de 0,05 indica que o modelo não é heterocedastico. Ao comparar ambos os outputs, podemos perceber que o modelo linear



possui presença forte de heterocedasticidade, pois o valor da probabilidade  $> \chi^2$  é inferior a 0,05. Por sua vez, o modelo não linear, normalizando a Y por BOX-COX indicou melhora neste teste, anulando, portanto, a presença de heterocedasticidade.

Test Summary			Test Summary		
DF	=	1	DF	=	1
chi2	=	16.6139	chi2	=	0.05204538
Prob > chi2	=	4.581403e-05	Prob > chi2	=	0.8195416

Figura 8. Breusch-Pagan Modelo Linear e Não Linear.

Fonte: Resultados originais obtidos na pesquisa

### Análise Fatorial por Componentes Principais [PCA]

Esta é uma técnica multivariada exploratória, com o objetivo de tratar a multicolinearidade presente na base e obter novas observações que captem o comportamento conjunto das variáveis originais (Fávero e Belfiore, 2017). Após criar a matriz de correlações e aplicar o teste de esfericidade de Bartlett, que retornou o p-valor de 1.283726e-88, é possível concluir que a matriz é fatorável, validando a aplicação desta técnica.

A Figura 9 exibe o summary do modelo PCA gerado, onde foi adotando o método das raízes latentes, e, mantendo os componentes principais com valores de desvio padrão superiores a um, para obtenção dos fatores. Importante salientar que 86,23% da variância total da base foi capturada com estes quatro componentes.

Importance of components:									
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Standard deviation	1.9307	1.2747	1.1674	1.0249	0.66653	0.56561	0.47993	0.37915	0.30990
Proportion of Variance	0.4142	0.1805	0.1514	0.1167	0.04936	0.03555	0.02559	0.01597	0.01067
Cumulative Proportion	0.4142	0.5947	0.7461	0.8629	0.91222	0.94776	0.97336	0.98933	1.00000

Figura 9. Summary Modelo PCA.

Fonte: Resultados originais obtidos na pesquisa

Abaixo as Figuras 10 e 11 apresentam os pesos que cada componente principal captura em cada variável.

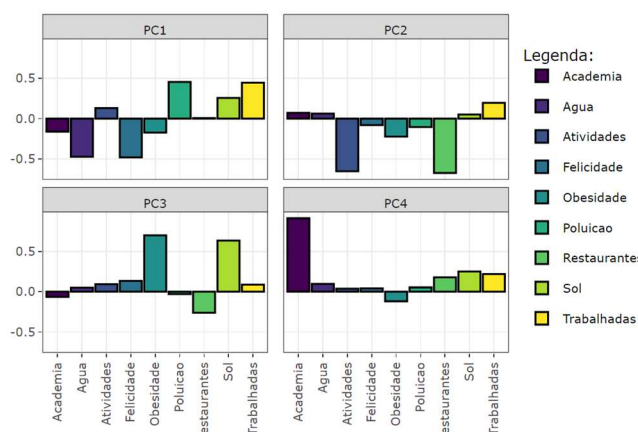


Figura 10. Pesos capturados pelas componentes principais.

Fonte: Resultados originais obtidos na pesquisa

	PC1	PC2	PC3	PC4
Sol	0.25632378	0.05173649	0.63367667	0.24974038
Agua	-0.47373063	0.06177191	0.04994505	0.09687390
Obesidade	-0.17383443	-0.22367118	0.69886180	-0.11965068
Poluicao	0.45472310	-0.10514722	-0.02923054	0.05284818
Trabalhadas	0.44639406	0.19619218	0.08742518	0.21799024
Felicidade	-0.48013600	-0.08016724	0.13317232	0.04025332
Atividades	0.13039188	-0.65382201	0.09315924	0.03697810
Restaurantes	0.00657942	-0.67453001	-0.26154649	0.17797806
Academia	-0.16137640	0.07092758	-0.06506842	0.91046819

Figura 11. Pesos capturados pelas componentes principais.

Fonte: Resultados originais obtidos na pesquisa

Ao analisar a primeira componente principal, é possível observar que esta captura negativamente as variáveis Água, referente ao custo de uma garrafa de água de 300 ml e Felicidade, referente ao índice de felicidade da população estudada, captando positivamente grande porção das variáveis Trabalhadas, sendo esta a média anual de horas trabalhadas e Poluição, referente ao índice de poluição que cada cidade apresenta, e Sol referente a média de horas anuais de luz solar em cada cidade. Levando em consideração estas observações, concluiu-se que quanto maior o número de horas trabalhadas, aumentasse a poluição, a provável exposição ao sol, índice de felicidade diminui e o acesso a água fica facilitado devido ao custo, porém esta observação pode indicar uma dificuldade ao ingerir água durante o dia, devido as longas jornadas. Portanto, ao obter os fatores principais, este será nomeado de "Sobrecarga\_Cotidiano".

A segunda componente principal, por sua vez, captura negativamente e de maneira muito forte as variáveis, Atividades, referente as atividades ao ar livre disponíveis nas cidades e Restaurantes, que se refere ao número de opções de delivery. Como os principais pesos desta componente são negativos, a multiplicação desta por -1 é adequada para que os quatro componentes capturados estejam vetorialmente no mesmo sentido, ilustrado na figura 12. Portanto, esta componente principal pode ser nomeada como "Lazer", representando o lazer oferecido por cada cidade aos seus habitantes.

A terceira componente principal, por sua vez, capta vigorosamente e de maneira positiva a variável Obesidade e Sol. Por este motivo, esta componente principal foi nomeada de "Sedentarismo".

A análise da quarta, e última componente principal, observa-se que somente a variabilidade da variável Academia é capturada de maneira significativa, sendo que as variáveis Restaurantes, Sol e Trabalhadas são positivas e discretamente capturadas. Portanto, essa componente principal será denominada de "Hbitos\_Saudaveis".

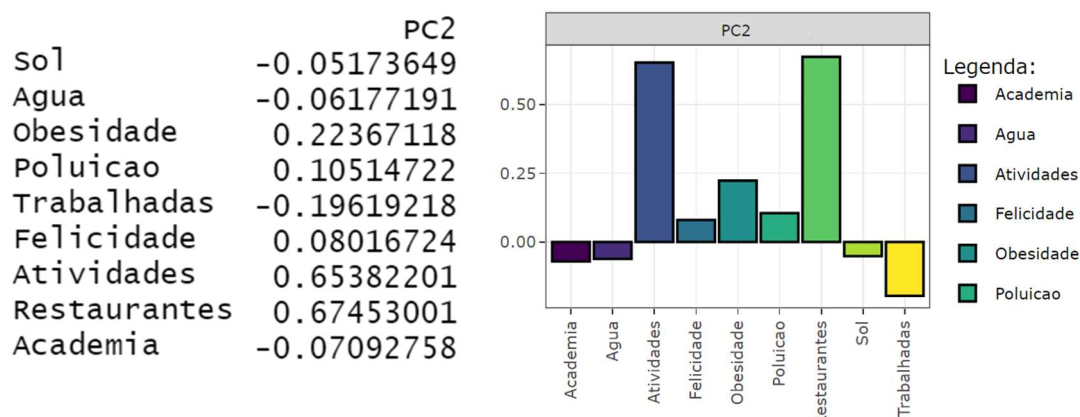


Figura 12. Resultado da multiplicação da Componente Principal 2 por -1.  
Fonte: Resultados originais obtidos na pesquisa

### Regressão por Componentes Principais [PCR] e Normalização Y Box-Cox

O modelo PCR foi elaborado com as novas variáveis obtidas por meio da técnica PCA, incluindo neste momento a variável Vida a ser predita, no entanto, normalizada por Box-Cox. O summary deste modelo mostra que a variável Sobrecarga\_cotidiano como a mais significativa para explicar a expectativa de vida, seguida por Habitoss\_Saudaveis. As demais variáveis não se mostraram estatisticamente significantes a 95% de confiança para explicar a expectativa de vida. Abaixo a Figura 13 ilustrando o summary do modelo.

```

Residuals:
    Min       1Q   Median       3Q      Max
-9.031e+20 -3.811e+20 -5.710e+19  3.056e+20  1.169e+21

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.600e+21  7.487e+19  21.365  < 2e-16 ***
Sobrecarga_Cotidiano -5.606e+20  7.573e+19  -7.402  6.06e-09 ***
Lazer         8.154e+19  7.573e+19   1.077   0.2882
Sedentarismo  -9.345e+19  7.573e+19  -1.234   0.2246
Habitoss_Saudaveis  1.833e+20  7.573e+19   2.420   0.0203 *
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.966e+20 on 39 degrees of freedom
Multiple R-squared:  0.6189,    Adjusted R-squared:  0.5798
F-statistic: 15.83 on 4 and 39 DF,  p-value: 8.857e-08
    
```

Figura 13. Summary Modelo PCR + Box-Cox Y.  
Fonte: Resultados originais obtidos na pesquisa

O teste de Shapiro-Francia, Figura 14, indica a aderência dos termos dos termos de erro a normalidade, sendo que o teste de Breusch-Pagan, Figura 15, apresenta que não há heterocedasticidade presente no modelo e o teste VIF, Figura 16, indica que não há multicolinearidade, sendo este fenômeno contornado pela aplicação do modelo PCA, tornando as novas observações ortogonais entre si.

```
shapiro-francia normality test

data:  OLS_BC_PCR$residuals
w = 0.96847, p-value = 0.2284
```

Figura 14. Shapiro-Francia teste de normalidade.

Fonte: Resultados originais obtidos na pesquisa

```
Test Summary
-----
DF          =    1
chi2        = 0.6339087
Prob > chi2 = 0.4259252
```

Figura 15. Breusch-Pagan teste.

Fonte: Resultados originais obtidos na pesquisa

```
variables  VIF
1 Sobrecarga_Cotidiano  1
2 Lazer                 1
3 Sedentarismo          1
4 Habitros_Saudaveis    1
```

Figura 16. Teste VIF.

Fonte: Resultados originais obtidos na pesquisa

Os resultados dessa pesquisa mostram que os principais fatores que explicam o índice de qualidade de vida das cidades listadas no “healthy lifestyle cities report 2021” são sobrecarga do cotidiano e hábitos saudáveis, enquanto lazer e sedentarismo não apresentam um poder de explicação significativo. De acordo com Almeida et al. (2012), a qualidade de vida pode ter fatores explicados por aspectos do âmbito social, biológico, econômico, político, humano, bem-estar, entre outros. Como o recorte da presente pesquisa busca avaliar apenas cidades que já possuem um status de qualidade de vida maior, notou-se que aspectos biológicos e social como sedentarismo e lazer não são fatores problemáticos para explicar a qualidade de vida na população. Contudo, os resultados dessa pesquisa não fornecem evidência de que esses fatores não são importantes para outras cidades, consideradas com um status de qualidade de vida mais variado. Portanto, futuros estudos podem explorar uma análise para avaliar se os fatores explicativos da expectativa de vida podem diferir entre cidades com status de qualidade de vida diferentes.

Ademais, Minayo et al. (2000) e Fonseca et al. (2019) argumentam que condições elementares, como acesso as necessidades básicas, como, água potável e hábitos de vida saudáveis, são preponderantes para aumentar a expectativa de vida da população. Os resultados dessa pesquisa corroboram com essa perspectiva, ao mostrar os efeitos importantes da sobrecarga do cotidiano (composta por altos níveis de poluição e excesso de trabalho e baixos níveis de felicidade) para prejudicar o índice da expectativa de vida, enquanto também ressalta o efeito positivo dos hábitos saudáveis, como a dedicação de horas do dia para frequentar academias.

## Conclusão

Essa pesquisa buscou estudar e compreender os fatores que explicam e determinam a expectativa de vida, sobretudo norteando políticas públicas em cidades não listadas, permitindo a aplicação de medidas de curto e longo prazo com o objetivo de elevar a qualidade e expectativa de vida de sua população.

A análise apontou através do modelo PCR, que a observação Sobrecarga\_cotidiano, se demonstrou negativa e estatisticamente significativa para explicar a expectativa de vida, ou seja, quanto maior as horas trabalhadas, mais poluição é gerada, o acesso a água e o índice de felicidade diminuem drasticamente, e, conseqüentemente reduzindo a expectativa de vida.

Por sua vez, a observação Habitats\_Saudaveis, retirada do modelo PCA e que se mostrou positiva e estatisticamente significativa para explicar a longevidade de uma população através do modelo PCR, explanou a importância da atividade física na expectativa de vida, sobretudo em ambientes fechados como ginásios e academias.

Portanto, o investimento e incentivo em horários de trabalho mais flexíveis ou até mesmo reduzidos, diminuição do tempo de deslocamento entre casa/trabalho, diminuição nos níveis de poluição gerados, facilitar o acesso a água que também pode estar ligado ao saneamento básico, aliando isso a prática de atividades físicas em academias e ginásios, podem ser políticas públicas a serem implementadas em regiões com o objetivo de aumentar a qualidade de vida e conseqüentemente a expectativa de vida da população.

## Referências

Almeida, M. A. B., Gutierrez, G. L., Marques, R. *Qualidade de Vida*. Each USP. São Paulo.

Fávero, L. P., & Belfiore, P. 2017. Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®. Elsevier Brasil.

Fonseca, C. D., Ting, M. L. B., & Sarti, F. M.. Evolução de indicadores de estilo de vida e potencial influência de políticas públicas na saúde da população brasileira: 2006-2017. São Paulo, 2019.

Lenstore. Healthy lifestyle cities report 2021. Disponível em: <https://www.lenstore.co.uk/-research/healthy-lifestyle-report/>

Minayo, M. C. D. S., Hartz, Z. M. D. A., Buss, P. M. Qualidade de vida e saúde: um debate necessário. *Ciência & Saúde Coletiva*, 5, 7-18, 2000.

Wickham, H., & Grolemund, G. (2017). R for data science: Import. Tidy, transform, visualize, and model data, 1.