AVALIAÇÃO CRÍTICA DE PROCEDIMENTOS DE ANÁLISE ESTATÍSTICA EM MODELOS DE PREVISÃO LINEAR DO ESTADO GERAL DE RODOVIAS

Wendy Fernandes Lavigne Quintanilha
Carla Marília Cavalcante Alecrim
Renan Santos Maia
Gledson Silva Mesquita Júnior
Programa de Pós-Graduação em Engenharia de Transportes (PETRAN)
Universidade Federal do Ceará (UFC)

RESUMO
Aspectos socioeconômicos e climáticos condicionam rodovias de cada região a enfrentarem problemas de características e escalas diferentes. Dada a baixa qualidade das rodovias cearenses, é fundamental a criação de ferramentas que possam orientar as autoridades para, por exemplo, direcionar os recursos para a gerência dessas rodovias. Nesse contexto, a utilização de ferramentas estatísticas de correlação e regressão podem auxiliar no desenvolvimento de modelos que ajudem a prever a ocorrência de fenômenos associados à deterioração da malha rodoviária. Neste estudo, foram estabelecidas análises para a avaliação da influência de variáveis relacionadas ao clima e à economia local na proporção de rodovias com pavimentos deteriorados. Foram criados modelos de regressão linear simples e múltipla, com variáveis definidas a partir do método *stepwise*. Observou-se uma melhor adequação às premissas relativas aos modelos de regressão no caso da regressão simples, na qual se constata uma predominância da precipitação na explicação do fenômeno.

ABSTRACT
Socioeconomic and climatic aspects condition force highways of each region to face problems of different characteristics and scales. Given the low quality of the highways of Ceará, it is fundamental to create tools that can better help authorities, for example, with allocation of resources for the management of these roads. In this context, the use of statistical tools of correlation and regression can help in the development of models that help to predict the occurrence of phenomena associated to the deterioration of the road network. In this study, analyzes were established to evaluate the influence of variables related to the climate and the local economy in the proportion of deteriorated roads. Simple and multiple linear regression models were created with variables defined from the stepwise method. It was observed a better adaptation to the assumptions regarding the development of regression models in the case of simple linear regression, in which a precipitation climatic factor predominates in the explanation of the phenomenon of the deterioration of the road network.

1. CONSIDERAÇÕES INICIAIS

A condição da malha rodoviária nacional é uma das faces mais perceptíveis da problemática tratada na engenharia rodoviária. De fato, existe uma dificuldade em administrar e direcionar os gastos com infraestrutura para o reparo e construção de novas rodovias, tanto no contexto nacional quanto nos contextos locais. As diferentes características socioeconômicas e climáticas de cada região e de cada estado condicionam as rodovias a revelar problemas de características e escalas diferentes. A necessidade de investimentos, embora seja uma resposta imediata, enfrenta outras questões como a disponibilidade de recursos e a ocorrência de outros fatores que podem prejudicar ainda mais a qualidade das infraestruturas.

Uma importante necessidade da gerência de pavimentos é o estabelecimento de ferramentas que viabilizem: (i) determinar a condição da rede pavimentada e o seu nível de serviço e (ii) estimar a qualidade dessa rede ao longo de sua vida útil (Shahin, 2005). Para este fim existem diversos meios, que avaliam os pavimentos a partir de diferentes perspectivas. É importante, nesse contexto, que sejam definidas quais características se deseja avaliar para, a partir disso, definir os parâmetros mais adequados. O *Pavement Condition Index* (PCI) é um dos mais recorrentes utilizado na gerência de pavimentos e se enquadra bem nos estudos de rodovias e aeródromos. Shahin (2005) ressalta que é decisivo para a gerência de pavimentos que sejam

definidos modelos para a previsão da qualidade das rodovias. O autor apresenta um meio para a concepção de modelos que utilizam a vida de serviço como variável para se estimar a vida útil das estruturas. A robustez de dados é uma característica recorrente para o estabelecimento de modelos estatísticos de desempenho para a gerência rodoviária (Yshiba, 2003; Benevides, 2006; Soncim e Fernandes Júnior, 2015)

Dadas tais complexidades, é importante que sejam elaborados modelos que sejam eficientes em explicar o fenômeno da deterioração da malha viária e auxilie os profissionais da área na definição de uma estratégia adequada, deixando-se de lado análises estatísticas simplórias, que podem conduzir a conclusões erradas ou incompletas a respeito de fenômenos estudados, especialmente em um contexto em que os parâmetros são atualizados apenas anualmente e que apresentam uma grande dependência dos ciclos econômicos, resultando em banco de dados limitado. Por exemplo, dados relacionados a décadas ou ciclos econômicos anteriores ao estudado no presente podem não ter validade frente à realidade observada, enviesando conclusões que doravante podem significar investimentos pouco inteligentes e o desperdício dos impostos pagos pelos contribuintes.

Dentro desse contexto, é frequente a preocupação de agências de infraestrutura de transportes nacionais em identificar quais os principais problemas de cada região. A Confederação Nacional do Transporte (CNT) elabora anualmente um panorama geral das rodovias brasileiras, realizado em larga escala, utilizado como um norte para o conhecimento da condição do pavimento. Apesar das limitações metodológicas, este estudo anual expõe os principais defeitos observados pelas equipes de inspeção e elabora relatórios estaduais que visam a apresentar dentro do contexto nacional quais estados apresentam problemas mais críticos. Em 2017, por exemplo, o Estado do Ceará apresentava um percentual de rodovias em estado regular, ruim ou péssimo de 60,6% (CNT, 2017). Já em 2018, esse percentual foi estimado em 72,1%, sendo considerados defeitos de geometria, sinalização e a qualidade do pavimento.

O objetivo deste trabalho é expor uma modelagem estatística mais confiável para a elaboração de modelos de previsão da qualidade de rodovias em estado ruim ou péssimo para o caso de limitações de bancos de dados, dando destaque para parâmetros que podem conduzir a conclusões erradas ou incompletas.

2. DEFINIÇÃO DAS VARIÁVEIS

O passo inicial para a concepção de um modelo eficaz em representar malhas rodoviárias requer a consideração das principais características da região, dentro do contexto viário. Por exemplo, no estado do Ceará, é usado majoritariamente o modo rodoviário para o escoamento de cargas e pessoas, indicando forte solicitação da rede pavimentada, que conta com 8.681 km de extensão (CNT, 2018). Os investimentos em reparo e construção de novas rodovias neste estado ocorrem principalmente por meio de investimentos públicos, que incorporam características muito peculiares de escassez de recursos. Além disso, existem particularidades quanto ao clima, típico do nordeste brasileiro, com considerável variabilidade na quadra chuvosa. Acredita-se que a ocorrência de precipitações implica diretamente na qualidade do pavimento, especialmente no caso de sistemas deficientes de drenagem. Dado o referido conhecimento prévio, devem ser buscadas para a proposição do modelo variáveis que guardem relação com o nível de tráfego e a realidade climática do estado.

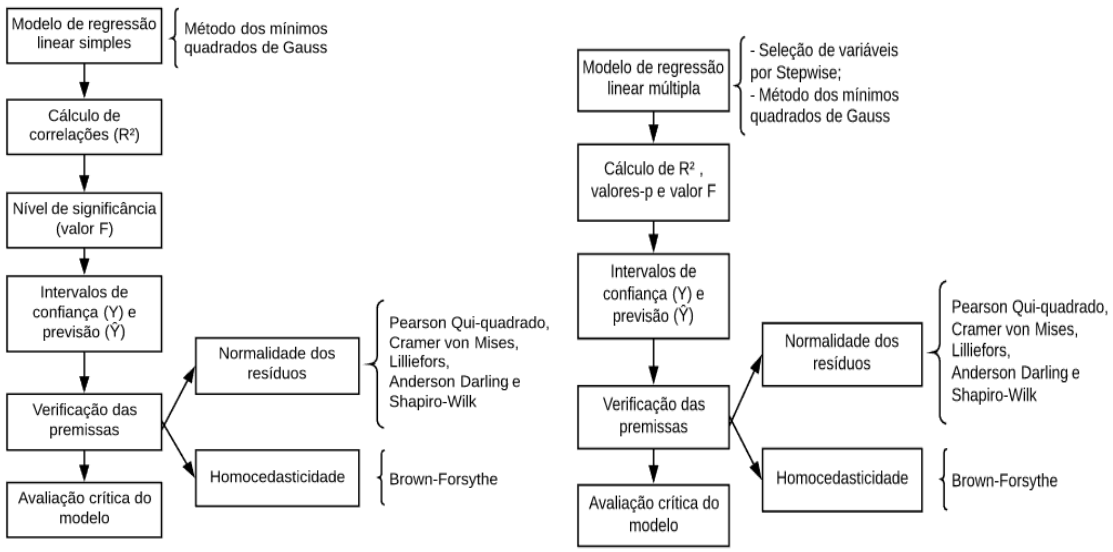
Para modelar o comportamento da variável dependente "percentual de rodovias cearenses com pavimento em estado ruim", foram selecionadas algumas variáveis explicativas, no período de 2005 a 2017 (com um hiato no ano de 2008, totalizando 12 observações), as quais se acredita terem relação com o fenômeno em estudo. Considerou-se que os dados obtidos representariam todo o território do estado do Ceará, embora reconheça-se que isto é uma aproximação, visto que alguns fenômenos por trás das variáveis (como a safra de feijão, por exemplo), podem ser referentes a locais específicos do estado e, portanto, apresentar limitações na explicação da variável dependente. As variáveis escolhidas, unidades e fontes são mostradas na Tabela 1. Neste ponto, deve-se questionar a possibilidade de que essas variáveis tenham correlação entre si, o que não se deseja em modelagem estatística, na qual se deseja evitar a sobreposição de efeitos das variáveis dependentes. Por exemplo, pode-se esperar que elevada correlação entre PIB/precipitação/produção agrícola; não sendo indicado, caso se confirme essa hipótese, que mais de uma dessas variáveis seja adicionada ao modelo. A variável "Safra de Feijão" foi escolhida porque além de ser um dos principais produtos agrícolas da economia do estado do Ceará, também é um grão sensível ao clima.

Tabela 1: Variáveis selecionadas

Variável	Tipo	Unidade	Fonte
Rodovias em condições "ruins e péssimas"	Dependente	%	CNT
Precipitação	Explicativa	mm	FUNCEME
PIB	Explicativa	bilhões de reais	IPECE
Veículos de carga	Explicativa	unidades	DENATRAN
Safra de feijão	Explicativa	toneladas	IBGE

3. MÉTODO DE ANÁLISE ESTATÍSTICA

Os métodos utilizados para a elaboração de modelos de regressão linear simples e múltipla são mostrados nos fluxogramas contidos na Figura 1. Ao fim, a decisão a respeito de qual desses produtos deve ser adotado parte da aderência de cada modelo à realidade que se deseja representar, das suas significância, complexidade e facilidade de obtenção e projeção de dados para as variáveis explicativas.



(a) Regressão Linear Simples (b) Regressão Linear Múltipla

Figura 1: Método de elaboração dos modelos de regressão

Realizou-se a determinação de apenas uma variável independente, dentre as previamente elencadas, a ser pareada com y. Do gráfico de dispersão dos pontos, calcularam-se os coeficientes angular e linear b_1 e b_0, respectivamente, com base no princípio dos mínimos quadrados de Gauss. Tal procedimento resultou em um coeficiente de determinação R^2, que indica o ajustamento dos valores observados ao modelo de regressão linear, muitas vezes encarado, erroneamente, como suficiente para definição da qualidade dos modelos de previsão.

O valor de R^2 é uma relação matemática que envolve os parâmetros SQE (Soma dos Quadrados dos Erros) e SQT (Soma dos Quadrados Totais). O SQE mantém relação diretamente proporcional com a variância dos valores de y para cada valor de x, o que é associado à variação em y não explicada pelo modelo. O SQT, por sua vez, é o somatório dos quadrados das diferenças entre os y e a média amostral de y ($\bar{Y}$). O SQT é a soma do SQE e do SQR (soma dos quadrados do modelo de regressão).

Os valores-p para os testes de hipótese dos valores de β_0 e β_1, com grau de confiança de 95%, e hipótese nula de que eles são iguais a 0, por sua vez, também foram obtidos nessa etapa. Um valor-p maior que o nível de significância adotado (5%) para o teste de hipótese do β_0 indica que a hipótese nula de que o intercepto é 0 não deve ser rejeitada. Caso esse valor seja menor que 5% a hipótese nula deve ser rejeitada e conclui-se que há indícios de que a variável independente x escolhida explica bem a variável dependente y segundo a relação linear estimada. Por último, nesta etapa, foi realizado o teste F, estatística calculada pela ponderação dos erros sobre os graus de liberdade, para se verificar o nível de significância do modelo de regressão linear. Seu valor deve ser comparado ao seu valor crítico.

Obtida a equação da reta para o modelo, pode-se encontrar os valores esperados de Y ($\hat{Y}$), substituindo os diferentes valores de x na equação da reta obtida. Calcula-se então o intervalo de confiança (IC) da equação, que determina o intervalo provável onde estará o valor médio populacional de Y ($\bar{Y}$) para cada valor de x, utilizando o desvio padrão (S) de Y. Também é calculado um intervalo de previsão (IP) para um valor de Y futuro para definir um intervalo de valores plausíveis para um Y futuro, utilizando o desvio padrão de X.

A técnica de regressão linear, tanto a simples quanto a múltipla, exige que algumas premissas sejam atendidas: é necessário que a variável dependente Y possua uma distribuição normal para cada valor da variável explicativa X (premissa de normalidade) e que essas distribuições tenham variâncias idênticas (homocedasticidade). Essas premissas podem ser checadas a partir dos resíduos do modelo por meio de testes formais. Na premissa de normalidade dos resíduos, o primeiro passo deve ser a plotagem de um gráfico entre os resíduos padronizados e o escore z esperado para uma distribuição normal, obtidos a partir dos percentis calculados para a amostra de resíduos; caso esses resíduos sigam uma distribuição normal, espera-se que esse gráfico tenha o formato de uma reta. Feita a aferição visual da normalidade dos veículos, procede-se à realização de testes formais de normalidade: Pearson Qui-quadrado, Cramer von Mises, Lilliefors (Kolmogorov-Smirnov), Anderson Darling e Shapiro-Wilk.

A premissa de homocedasticidade dos resíduos deve ser verificada em seguida, a partir do método de Brown-Forsythe; segundo esse método, divide-se a amostra de resíduos em dois grupos e realiza-se um teste de igualdade das médias (teste t) dos desvios absolutos dos resíduos para as medianas em cada grupo. Aqui, é ideal que não haja indícios para rejeitar a hipótese

nula de que essas médias são iguais, de forma a verificar a homocedasticidade dos resíduos.

Para selecionar as variáveis que irão compor o modelo final de regressão múltipla, procedeu-se à utilização do método conhecido como *stepwise*. O método *stepwise* funciona de forma semelhante ao método *forward*, segundo o qual as variáveis vão sendo adicionadas ao modelo e é verificada a sua significância; a primeira variável adicionada é aquela, dentre as variáveis disponíveis, que possuir o maior R^2 com a variável explicada; estima-se o modelo de regressão e verifica-se a significância da variável adicionada, permanecendo a variável no modelo se a mesma se mostrar significante; as demais variáveis são adicionadas ordenadamente, da maior para a menor redução no SQE do modelo, e têm sua significância avaliada.A principal diferença entre o *forward* e o *stepwise* é que, no primeiro, uma vez que a variável entra no modelo ela permanece no modelo; já no método *stepwise*, uma variável que entrou no modelo pode sair posteriormente, já que a cada variável selecionada para entrar no modelo, a significância das variáveis já adicionadas e, consequentemente, a permanência das mesmas no modelo, é reavaliada.

4. RESULTADOS

4.1. Análise de correlação

A matriz de colinearidade mostrada na Tabela 2 apresenta o apanhado dos resultados para o valor de r (coeficiente de correlação e estimador do parâmetro ρ), a partir dos quais procedeu-se as análises subsequentes (verde: correlação positiva; vermelho: correlação negativa). Analisando a Tabela 2, é possível selecionar qual(is) variável(is) apresenta(m) maior correlação com a variável independente, o que é priorizado quando da composição dos modelos de regressão, seja o linear ou múltiplo.

Tabela 2: Análise de correlação entre as variáveis

	% ruim/péssimo	Precipitação	PIB	Veículos de carga	Safra de feijão
% ruim/péssimo	1				
Precipitação (mm)	0,87	1			
PIB (bilhões)	-0,46	-0,33	1		
Veículos de carga	-0,51	-0,31	0,95	1	
Safra de feijão (t)	0,75	0,75	-0,41	-0,49	1

4.2. Regressão linear simples

A variável que apresentou melhor correlação com a variável dependente foi a precipitação de chuvas (r = 0,87). A partir disso, foi realizado um teste de correlação de Pearson (α = 5%), de onde se constatou que o valor de r obtido é superior ao valor crítico para n = 12 ($r_{crítico}$ = 0,576), rejeitando-se a hipótese nula de que não há correlação. Assim, procedeu-se à regressão linear simples com esta variável. A Figura 2 apresenta a dispersão dos dados de precipitação (abscissas) *versus* percentual de rodovias com condições ruim e péssima (ordenadas) no estado do Ceará.

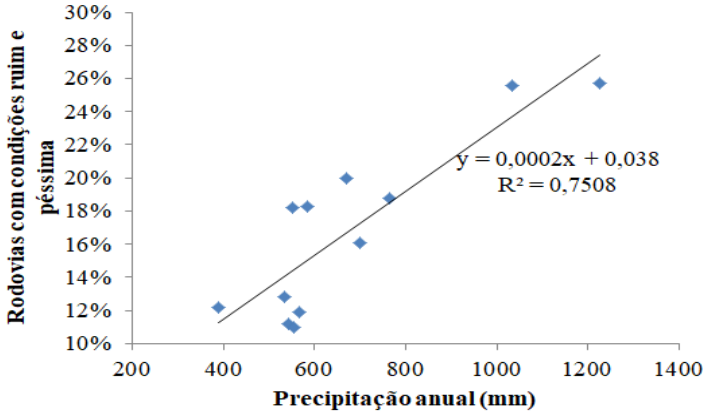


Figura 2: Dados de precipitação *versus* percentual de rodovias com condições ruim e péssima

A Figura 2 mostra, de maneira geral, que não há *outlier* evidente, embora tenha sido conduzido uma análise formal da presença de *outliers* e tenha se constatado que, de fato, não há, como será mostrado adiante. Esta figura também apresenta a linha de tendência amostral, assim como o seu R^2, que representa o percentual da variável dependente que é explicado por meio da relação linear com a variável independente, indicando uma alta relação de dependência entre as partes (0,75) para esse fenômeno. Vale-se ressaltar que o valor de R^2 ajustado também foi elevado, de 0,73. A Tabela 3 apresenta o resultado do modelo de regressão linear simples. Realizou-se, enfim, a estatística da regressão linear simples com os resultados para os coeficientes b_0 e b_1, da qual há indícios, a partir da análise do valor-p, para se rejeitar a hipótese nula de que os coeficiente b_1 é igual a zero, o que não foi observado para o coeficiente b_0. O modelo de regressão apresentou também um valor de F de significância igual a 0,00026, indicando que o modelo é significativo (>0,05).

Tabela 3: Resumo do modelo de regressão linear simples

Estatísticas de regressão	
R múltiplo	0,87
R-Quadrado	0,75
R-quadrado ajustado	0,73
Erro padrão	0,03
Observações	12
Parâmetros do modelo	
Coeficiente b0	3,80E-02
Coeficiente b1	1,93E-04
Valor-p b0	1,60E-01
Valor-p b1	2,66E-04

O modelo resultante é, portanto, $\hat{y}$ = 0,039 + 0.00019*x_1. A partir da análise dos dados da amostra, obtiveram-se os valores de desvio padrão (S), média e, consequentemente, $S_{\hat{y}}$. Com isso, obteve-se a estimação do valor esperado para Ȳ (valor médio), dado um X ($\mu_{y/x}$), com um intervalo de confiança de 95%. O intervalo de previsão para Ŷ (valor esperado), para um nível de confiança de 95%, também foi estimado. A Figura 3 mostra as os intervalos de confiança e de previsão encontrados. Analisando-a, tem-se que o intervalo de previsão mantém-se praticamente constante, enquanto que há aumento no intervalo de confiança para o valor esperado de Ȳ à medida que $x_1$cresce. Essa tendência de constância no intervalo de previsão à medida que se aumenta o valor da variável independente deve-se à normalidade dos resíduos

em relação à reta de regressão, que será abordada a seguir. Para se verificar as premissas de homocedasticidade e normalidade do modelo, realizou-se a plotagem dos resíduos, como mostrado na Figura 4.

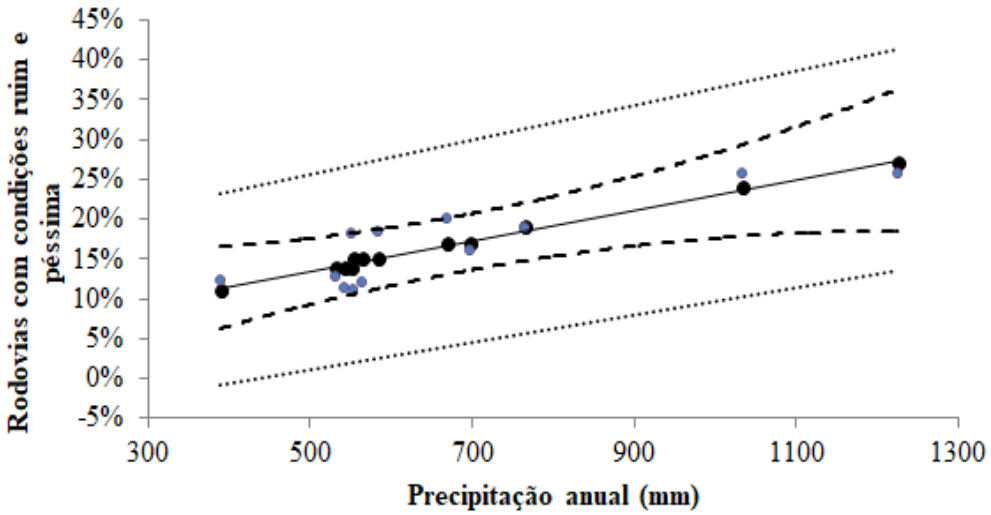


Figura 3: Intervalos de confiança para $\bar{Y}$ e de previsão de $\hat{Y}$

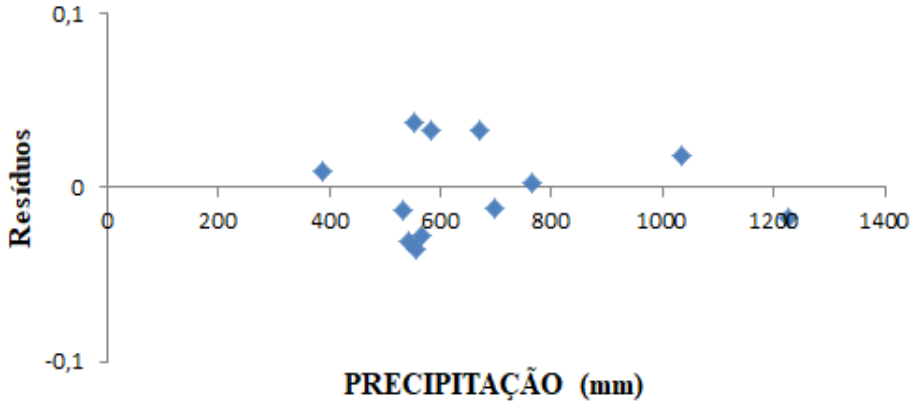


Figura 4: Distribuição dos resíduos do modelo de regressão linear simples

Para verificar a premissa de normalidade, realizou-se um teste de aderência Qui-quadrado de Pearson, que retornou um valor-p de 0,26, o que indica que não há evidências para se rejeitar a hipótese nula de que a distribuição dos resíduos pode ser bem representada pela distribuição normal. Realizou-se, ainda, o teste de normalidade de Cramer-von Mises (0,49), Lilliefors (0,45), Anderson-Darling (0,40) e Shapiro-Wilk (0,29) para corroborar a análise, com todos eles apontando para a mesma conclusão. A Figura 5 fornece graficamente a inferência de normalidade, indicando também que não há *outliers* na amostra, pois nenhum dos pontos está distante da curva relativa à plotagem de probabilidade normal.

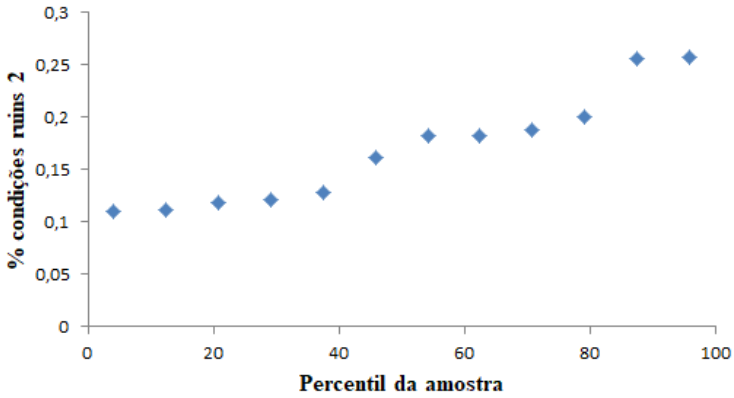


Figura 5: Plotagem de probabilidade normal

Com objetivo de se testar a homocedasticidade dos resíduos, utilizou-se a análise da variância proposta no teste de Brown-Forsythe, obtendo s^2= 0,00024, t_{bf}= 0,101 e t = 2,13. Com esse valor de t_{bf}, deixa-se de rejeitar a hipótese nula de que as médias dos dois grupos são iguais. Logo, conclui-se que a premissa de homocedasticidade dos resíduos é razoável. O modelo de regressão linear proposto apresentou elevado coeficiente de correlação entre a variável explicada e a explicativa; elevado nível de significância; e valores de coeficientes coerentes com as hipóteses inicialmente levantadas.

4.3. Regressão linear múltipla

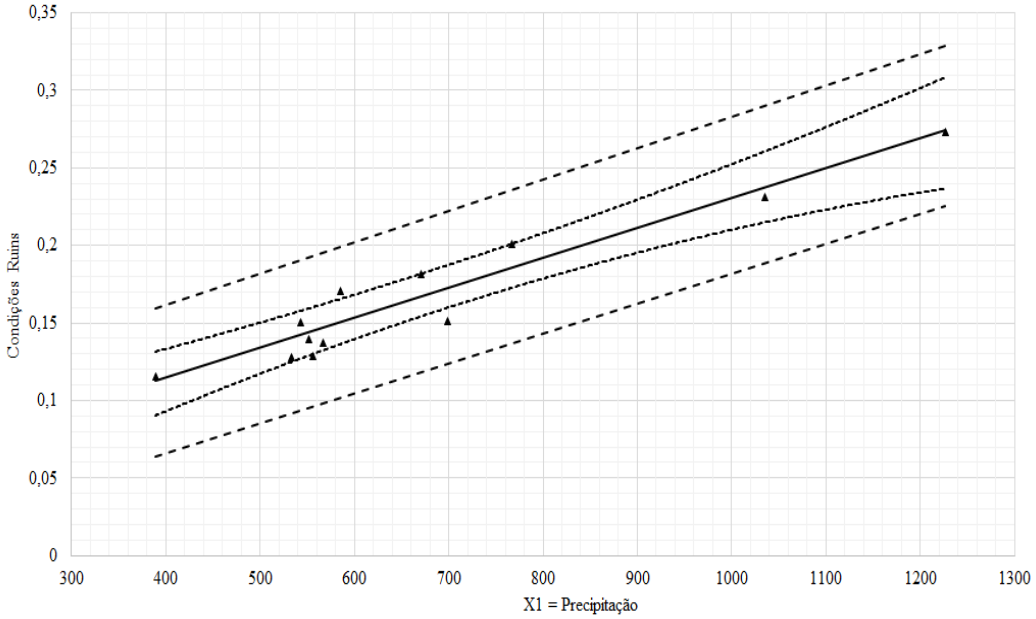
A determinação das variáveis para o modelo de regressão múltipla foi realizada por meio do método *Stepwise*. Como foi percebido que algumas variáveis dependentes apresentam correlação entre si, algumas dessas foram excluídas na composição do modelo de regressão linear múltipla (PIB e Safra do Feijão). Obteve-se um modelo baseado nas variáveis: (y) = variável dependente (proporção de rodovias com pavimentação em condições ruins/péssimas); (x_1) = variável independente - precipitação e (x_2) = variável independente - número de veículos de carga. Este modelo apresentou os melhores resultados de R² ajustado (0,77) e os respectivos p-valores: 0,04 para interseção, 0,00 para a variável x_1 e 0,11 para a variável x_2 (veículos de carga). Assim, acredita-se que a predominância desse modelo está na interseção, cuja hipótese nula de igualdade a zero é rejeitada, bem como na variável x_1, cujo p-valor encontrado foi menor que 0,05. Propôs-se, então, o seguinte modelo: y = 0,098 + 1,74E-04*x_1 - 4,30E-07* x_2. A Tabela 4 apresenta valores para a avaliação da qualidade do modelo proposto a partir da inferência de parâmetros.

Tabela 4: Resumo dos resultados para modelo de regressão linear múltipla

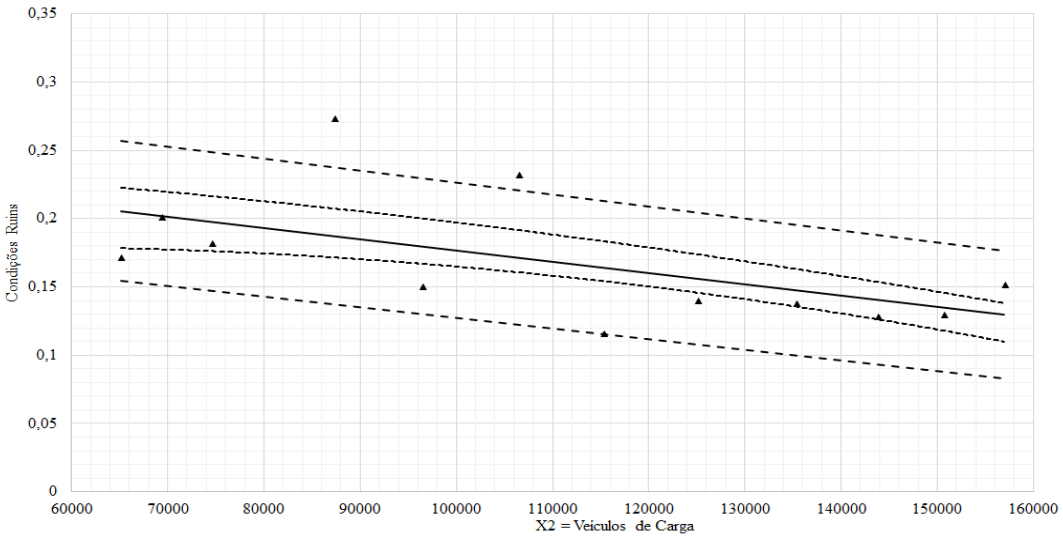
Estatística de regressão	
R múltiplo	0,90
R-Quadrado	0,81
R-quadrado ajustado	0,77
Erro padrão	0,025
Observações	12
F	19,75

A partir dos valores obtidos (número de observações, erro padrão, valores médios de x, além de Ŷ), foram construídos intervalos de confiança e previsão, apresentados na Figura 6, de forma similar ao que se propôs para a análise de regressão linear simples. No segundo caso, relativo à

variável x_2, foram observadas retas mais horizontais, o que seriam indícios mais fortes de que a variável em questão estaria pouco contribuindo para a explicação do fenômeno.



(a) IC e IP para a variável x_1.



(b) IC e IP para a variável x_2.

Figura 6: Intervalos de confiança e previsão para y a partir das variáveis x_1 (a) e x_2 (b)

Da análise matricial, obteve-se a matriz de correlações entre as variáveis utilizadas no modelo proposto. Como os valores obtidos fora da diagonal da matriz simétrica foram inferiores a 1, rejeita-se a existência de multicolinearidade. A matriz de correlações é exposta na Tabela 5.

Tabela 5: Matriz de correlação entre as variáveis escolhidas para o modelo composto

	Y	x_1	x_2
Y	**1,00**	0,00	0,00
x_1	0,00	**1,00**	0,31
x_2	0,00	0,31	**1,00**

Testes de normalidade não rejeitaram a premissa de normalidade dos resíduos obtidos quando se comparam os resultados Ŷ e os valores reais. Foram obtidos p-valores superiores a 0,70 para os testes de Shapiro-Wilk (0,98), Cramer-von-Mises (0,81), Lilliefors (0,74) e Anderson-Darling (0,86), deixando-se de rejeitar a hipótese de normalidade dos resíduos. Tal resultado é corroborado pelo gráfico de quantis normais, expostos na Figura 7.

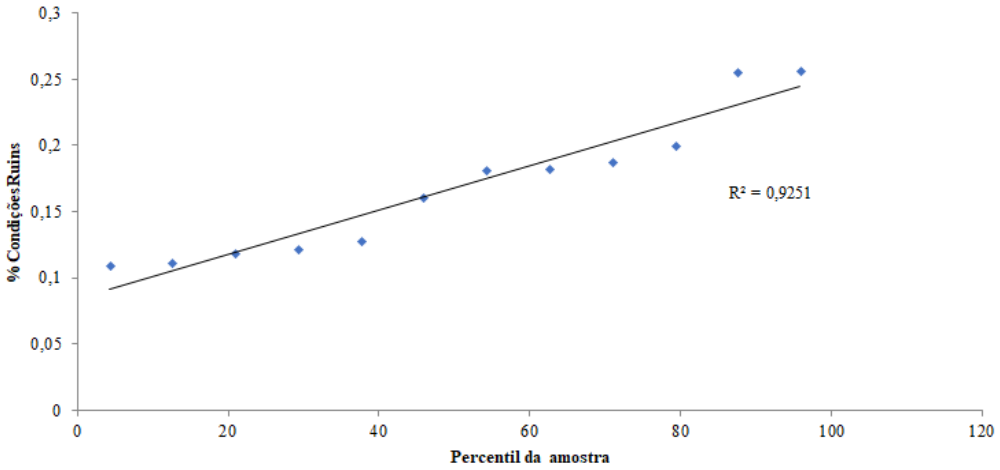


Figura 7: Quantis normais para os resíduos do modelo de regressão linear múltipla

Com objetivo de se testar a homocedasticidade dos resíduos, utilizou-se a análise da variância proposta no teste de Brown-Forsythe, e obteve-se um $t_{crítico}$ de 2,13, superior à estatística de teste, cujo valor encontrado foi de -0,02. Assim, deixa-se de rejeitar a hipótese nula de que as médias dos dois grupos de resíduos são iguais. Conclui-se que a premissa de homocedasticidade dos resíduos é razoável.

Este modelo apresentou um índice de correlação com os resultados observados considerado bom, acima de 0,7, entretanto algumas premissas necessárias à afirmação de que o modelo criado é razoável não foram cumpridas, especialmente quanto ao p-valor da segunda variável independente adicionada ao modelo. Tentou-se, inclusive, a criação de modelos com interação entre as variáveis, porém seguiu-se à mesma situação, na qual os p-valores, em geral, foram superiores a 0,05 (Triola, 2008; Devore, 2013) a pelo menos uma das variáveis independentes adicionadas no modelo, levando a crer que o modelo de regressão linear simples proposto, para o caso das variáveis selecionadas, é mais adequado para a explicação do fenômeno em estudo.

4.4. Validação

A validação dos modelos foi feita a partir dos dados atualizados de 2018. Os resultados de previsão (Ŷ) e o obtido pela Pesquisa CNT 2018 (Y) das condições do pavimento estão dispostos na Tabela 6.

Tabela 6: Validação dos modelos com dados de 2018

	X_0	X_1	X_2	Ŷ	Y	RESÍDUO	ERRO
MODELO 1	0,038	1,93E-04	-	0,192	0,178	0,014	8,12%
MODELO 2	0,098	1,74E-04	-4,30E-07	0,167	0,178	-0,011	-6,40%

Poder-se-ia chegar à conclusão de que, pelos resultados de R^2 e do erro, o modelo 2 seria mais

apropriado para a explicação do fenômeno. Essa conclusão seria equivocada, exatamente pela questão de que a significância de uma das variáveis não foi comprovada pelo valor-p, por mais que o R^2 tenha sido melhor. Isso poderia significar, por exemplo, que uma variação do tamanho da frota estaria enviesando as conclusões quanto à quantidade de rodovia com pavimento ruim/péssimo, porém isso não necessariamente seria correto. Além disso, a explicação do fenômeno perderia sentido lógico em função do sinal do coeficiente de x_2, negativo, o que indicaria que maiores números de veículos de carga poderiam contribuir para uma melhoria na qualidade do pavimento.

Por fim, considera-se o resultado de validação do modelo de regressão simples satisfatório, apresentando um erro de apenas 8,12%. Frisa-se, no entanto, que este modelo formulado visa apenas oferecer parâmetros de estimativa das condições gerais do pavimento nas rodovias do estado e que, apesar de dados da CNT não serem costumeiramente utilizados em gerenciamento de rodovias, sugere-se o mesmo rigor estatístico para criação de modelos utilizando outros bancos de dados.

5. CONCLUSÕES

No modelo de regressão linear simples, foi avaliada a relação entre a variável dependente "percentual de rodovias com pavimento em condições ruim e péssima" e a variável independente "precipitação". Nesta análise, obteve-se tanto uma alta correlação entre as variáveis, quanto a rejeição das hipóteses nulas de que b_1 é igual a 0. Essas informações reforçam a significância do modelo e a coerência dos coeficientes com as hipóteses levantadas. Durante a checagem das premissas do modelo, ele passou nos testes que verificam as premissas de normalidade e de homocedasticidade, corroborando, assim, para a confiabilidade das análises e inferências estatísticas realizadas a partir dele.

A análise obteve conclusões semelhantes para o modelo de regressão linear múltipla. Altos coeficientes de correlação foram obtidos, porém as premissas de p-valores inferiores a 0,05 não foram seguidas integralmente, apesar de que a normalidade dos resíduos e a homocedasticidade foram confirmadas a partir de testes formais. Considera-se, também, que a variável independente "veículos de carga" apresenta um comportamento diferente do esperado, já que, intuitivamente, acredita-se que, quanto maior a solicitação representada por essa variável, maior será o desgaste na via. Entretanto, no modelo, ela é representada em uma tendência oposta, sendo menor o índice de qualidade inferior quanto maior a solicitação de veículos de carga. Essa observação reforça a ideia de que a concepção de modelos a partir de valores de correlação pode ser incoerente com os fenômenos estudados, sendo fundamental a análise de diversas outras premissas, muitas vezes ignoradas pelos analistas.

Com relação ao valor do R^2 ajustado, indicador do percentual da variável Y que está sendo explicado pelas variáveis explicativas incluídas no modelo, percebe-se um pequeno aumento do seu valor no modelo de regressão múltipla, quando comparado àquele obtido na regressão linear simples, de 0,72 para 0,77. Além disso, o p-valor para todos os parâmetros, exceto b_2, obtiveram valor inferior a 0,05, mostrando-se significantes. Apesar disso, notou-se, por meio da estimativa dos valores esperados de Y, que a variável x_2 "veículos de carga" não possui contribuição significativa para explicar a variável. Além disso, a variável x_1 envolvida no modelo de regressão linear simples, quando em comparação com x_2, se mostra mais difícil de ser prevista e simulada. Isso se dá em função de que a variável x_1 é majoritariamente controlada por eventos de caráter mais imprevisíveis, como alterações climáticas. Diante dessas razões,

pode-se concluir que o modelo de regressão linear simples se apresentou como o mais indicado para explicar a variável em estudo.

A influência da precipitação na piora da qualidade do pavimento também expõe uma situação crítica das condições das rodovias do estado. É sabido que a infiltração da água no pavimento acelera o processo de degradação do mesmo, o que é intensificado em locais onde o trecho já está degradado, ao apresentar trincas e/ou não possuir sistema de drenagem adequado. Assim, verifica-se uma grave deficiência na manutenção dessas vias, cujo efeito é exponencial na piora das propriedades das mesmas. Existem limitações relacionadas ao modelo proposto no que diz respeito às variáveis, pois os dados utilizados são advindos de fontes que frequentemente não fornecem detalhes sobre as formas de amostragem/coleta ou apenas fornecem informações acerca do total geral daquela variável. Assim, o modelo fica dependente de que os dados obtidos estejam correspondendo a todo o espaço físico do estado do Ceará. Recomenda-se que em uma análise mais aprofundada acerca do alcance dos dados, antes de que se possa utilizar o modelo como critério para decisões técnicas.

Agradecimentos
Os autores agradecem às agências CAPES, CNPq e Funcap pelas bolsas concedidas.

REFERÊNCIAS BIBLIOGRÁFICAS

Benevides, S. A. S. (2006) *Modelos de Desempenho de Pavimentos Asfálticos para um Sistema de Gestão de Rodovias Estaduais do Ceará*. Tese de Doutorado. COPPE/UFRJ, Rio de Janeiro-RJ.

CNT (2005 a 2018) *Pesquisa CNT de Rodovias*. Brasília-DF.

DENATRAN (2019) Estatísticas - Frota de Veículos. Disponível em <https://infraestrutura.gov.br/component/content/article/115-portal-denatran/8552-estat%C3%ADsticas-frota-de-ve%C3%ADculos-denatran.html>. Acesso em Julho/2018.

Devore, J. L. (2013) *Probabilidade e Estatística para Engenharia e Ciências*. 1ª Edição - 4ª Reimpressão, Cengage Learning, São Paulo-SP.

FUNCEME (2019) Calendário das chuvas no Estado do Ceará. Disponível em: <http://www.funceme.br/app/calendario/produto/ceara/media/anual>. Acesso em Julho/2018.

IPECE (2005 a 2018) PIB Trimestral. Disponível em <https://www.ipece.ce.gov.br/pib-tabelas-especiais/>. Acesso em Julho/2018.

Shahin, M. Y. (2005) *Pavement Management for Airport, Roads and Parking Lots*. Springer. 2ª Edição. New York, NY. Estados Unidos.

Soncim, S. P; Fernandes Júnior, J. L. (2015) Modelo de previsão do índice de condição dos pavimentos flexíveis. *Journal of Transport Literature*, Vol.9(3), Pp. 25-29. DOI: http://dx.doi.org/10.1590/2238-1031.jtl.v9n3a5

Triola, M. F. (2008*) Introdução à Estatística*. 10ª Edição, LTC, Rio de Janeiro-RJ.

Yshiba, J.K. (2003) *Modelos de desempenho de pavimentos: estudo de rodovias do Estado do Paraná*. Tese de Doutorado. Escola de Engenharia de São Carlos/USP, São Carlos-SP.

Wendy Fernandes Lavigne Quintanilha (wendy@det.ufc.br)
Carla Marília Cavalcante Alecrim (cmariliac.civil@gmail.com)
Renan Santos Maia (renanmaia@det.ufc.br)
Gledson Silva Mesquita Júnior (gledson@det.ufc.br)
Programa de Pós-Graduação em Engenharia de Transportes (PETRAN), Departamento de Engenharia de Transportes (DET), Centro de Tecnologia, Universidade Federal do Ceará (UFC)
Campus do Pici, s/n – Bloco 703 – CEP: 60440-554 – Fortaleza, Ceará, Brasil