

Trabalho MLG

Eliana Cardoso Gonçalves e Sophia Araujo de Moraes

26/08/2024

Índice

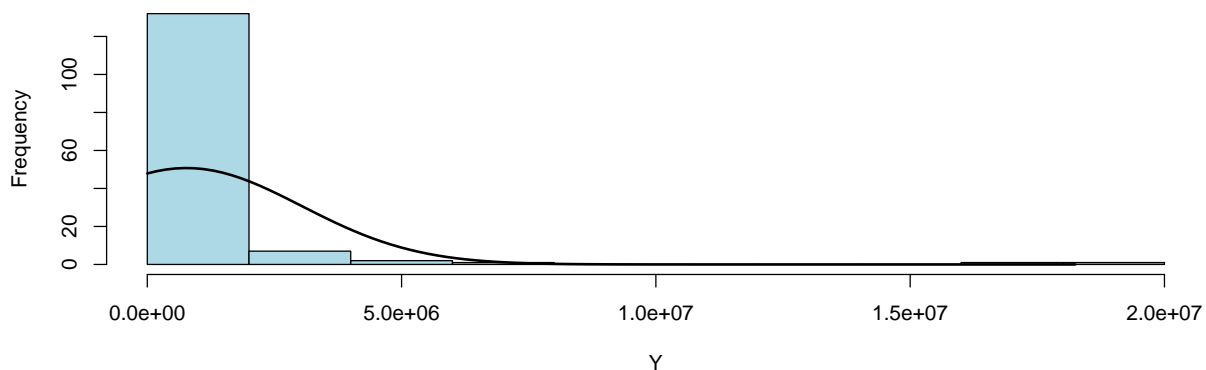
1	Análise Descritiva	3
2	Modelo Log-Normal	6
2.1	Resultados	7
2.2	Análise dos Resíduos	8
3	Modelo Logístico	10
3.0.1	Análise dos Resíduos	13
3.0.2	Análise do VIF (Variance Inflation Factor)	13
3.0.3	Significância das Associações Estimadas	13
3.0.4	Análise dos Resíduos e Multicolinearidade	14
3.0.5	Razões de Chances (Odds Ratios)	14
3.0.6	Predição e Avaliação do Modelo	15
3.0.7	Análise dos Resultados	16
4	Modelo Gama Log-Linear	17
4.1	Análise dos Resíduos	18
5	Comparação dos Modelos: Gamma Log-Linear vs. Modelo Normal	20

1 Análise Descritiva

Tabela 1.1: Tabela Resumo das Variáveis Y, L e K.

	K	L	Y
Mean	3058761.263	54.6094443	769941.411
Std.Dev	8413105.576	196.3877082	2264112.549
Min	7345.099	0.2993725	2569.155
Median	394840.953	11.6562212	123419.395
Max	64118472.000	2045.9122435	18244220.000

Figura 1.1: Histograma PIB Nacional (Y) em dólares PPP de 2014



Observando o Figura 1.1 e Tabela 1.1, o valor mínimo do PIB é \$2.569, enquanto o valor máximo é significativamente maior, chegando a \$18.244.220, indicando uma grande variação no PIB entre os países da amostra. A média do PIB é \$769.941, o que é substancialmente maior que a mediana de \$123.419, sugerindo que alguns países com PIB muito alto estão puxando a média para cima. A distribuição do PIB é bastante assimétrica, com muitos países tendo PIBs relativamente baixos e poucos países com PIBs muito altos.

Analisando o Figura 1.2 e a Tabela 1.1, a variação da força de trabalho de 0,2994 a 2.045,91 reflete uma grande disparidade na população economicamente ativa entre os países. Com uma média de 54,60 e uma mediana de 11,65, a distribuição mostra-se assimétrica, sugerindo a presença de países com forças de trabalho extremas, seja pela sua grandeza ou pequenez em relação à média da amostra alguns países com países com a força de trabalho outlier.

Figura 1.2: Histograma Força de Trabalho (L)

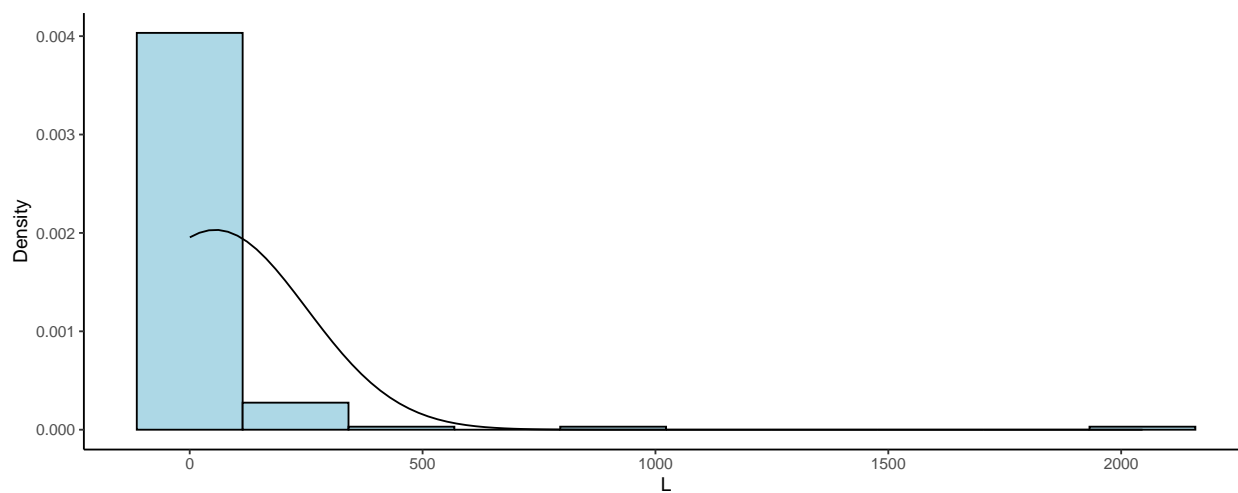
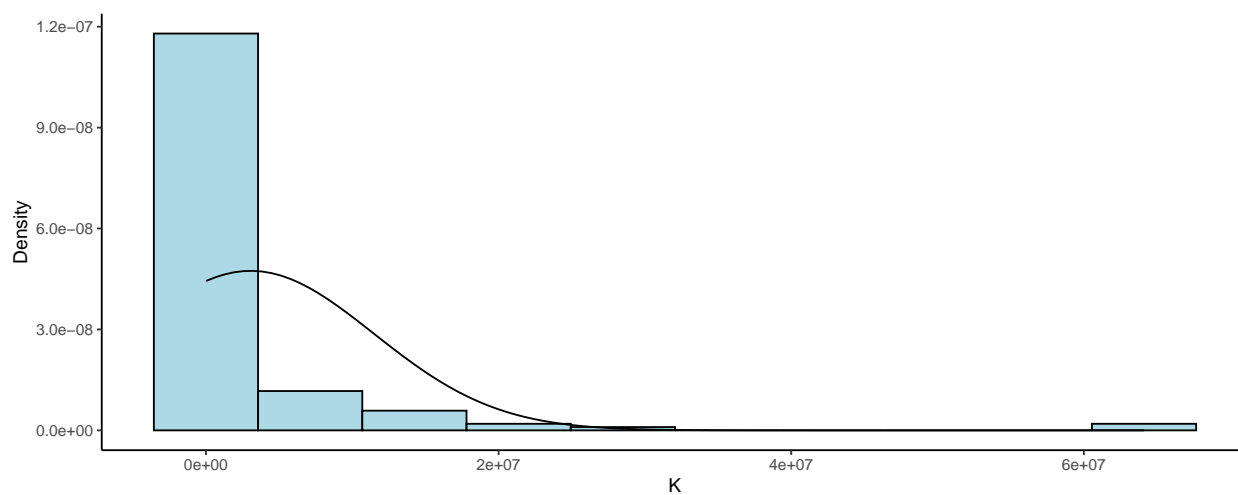


Figura 1.3: Histograma Estoque de Capital (K) em dólares PPP de 2014)



Analisando os valores da Tabela 1.1 e Figura 1.3 , observamos que o estoque de capital varia amplamente, de \$7.345 a \$64.118.472, indicando diferenças significativas no nível de investimento em capital produtivo entre os países.

A média do estoque de capital é \$3.058.761, consideravelmente maior que a mediana de \$394.841, o que sugere que alguns países têm estoques de capital extremamente altos que estão elevando a média.

2 Modelo Log-Normal

Este estudo se concentra em uma amostra de 144 países distintos no ano de 2014, utilizando um modelo linear normal para explorar como a força de trabalho e o estoque de capital influenciam o PIB. A regressão linear múltipla será empregada para estimar os coeficientes que quantificam essas relações, permitindo-nos compreender melhor os determinantes do crescimento econômico.

Essa relação pode ser expressa pela seguinte equação:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Onde:

Y é a variável dependente (PIB Nacional (Y) em dólares PPP de 2014).

X_1, X_2, \dots, X_p são as variáveis independentes (Força de Trabalho (L) e o Estoque de Capital (K) em dólares PPP de 2014).

$\beta_1, \beta_2, \dots, \beta_p$ são os coeficientes de regressão que representam os efeitos das variáveis independentes na variável dependente.

ε é o termo de erro, que representa a variação não explicada pelo modelo.

Para garantir a robustez e confiabilidade dos resultados, optamos por realizar uma transformação log-log nas variáveis, considerando que o PIB Nacional (Y) em dólares PPP de 2014, juntamente com as variáveis independentes “força de trabalho” e “estoque de capital”, não apresentam uma distribuição normal. Essa abordagem vai permitir que aos pressupostos do modelo normal sejam atendidos e obter resultados mais consistentes.

Ao aplicar a transformação log-log, estamos ajustando a distribuição das variáveis para se adequarem melhor ao modelo, mitigando quaisquer distorções ou vies que possam surgir devido à falta de normalidade. Isso nos permite realizar inferências estatísticas mais confiáveis e interpretar os efeitos das variáveis independentes sobre o PIB Nacional de forma mais precisa.

Essa estratégia de transformação aumenta a robustez da análise, pois reduz a influência de valores extremos e torna os resultados menos sensíveis a distribuições não normais. Portanto, podemos ter maior confiança nas conclusões derivadas do modelo, garantindo uma abordagem metodológica sólida e resultados mais confiáveis para tomada de decisão.

2.1 Resultados

A análise macroeconômica é fundamental para entender o desenvolvimento econômico e social dos países. No contexto da contabilidade nacional, variáveis como o Produto Interno Bruto (PIB), a força de trabalho e o estoque de capital são essenciais para avaliar a produtividade e o crescimento econômico. Este estudo utiliza dados da Penn World Table de 2020 para investigar a relação entre essas variáveis.

O PIB nacional (Y), medido em dólares PPC de 2014, é uma medida abrangente da atividade econômica de um país. A força de trabalho (L) representa o total de pessoas empregadas ou em busca de emprego, refletindo a capacidade produtiva humana. O estoque de capital (K), também medido em dólares PPC de 2014, indica o valor total dos ativos produtivos de um país, como máquinas, edifícios e infraestrutura.

Ao analisar essas variáveis, o objetivo é fornecer insights sobre as políticas econômicas que podem fomentar o crescimento e a produtividade. Esta investigação pode ajudar formuladores de políticas, economistas e pesquisadores a identificar áreas-chave para intervenção e investimento, promovendo um desenvolvimento econômico sustentável e inclusivo.

Tabela 2.1: Ajuste do Modelo de Regressão Log-Normal

	Estimação	Pvalor	sig
(Intercept)	1.8881930	0	<0.001***
log(L)	0.2940893	0	<0.001***
log(K)	0.7072816	0	<0.001***

Todos os valores-p associados aos coeficientes são muito pequenos (<0.001), o que significa que podemos rejeitar a hipótese nula para todos os coeficientes, ao nível de 5% de confiança. Isso indica que tanto a força de trabalho quanto o estoque de capital têm efeitos significativos no PIB, conforme medido pelo logaritmo.

O coeficiente estimado para $\log(K)$ é 0,70728. Isso significa que, se o estoque de capital (K) aumentar em 1%, o PIB (Y) aumentará em aproximadamente $0.70728 \times 100 = 70.728$, mantendo todas as outras variáveis constantes.

O coeficiente estimado para $\log(L)$ é 0,29. Isso significa que, se a força de trabalho (L) aumentar em 1%, o PIB (Y) aumentará em aproximadamente $0,29 \times 100 = 29,409$, mantendo todas as outras variáveis constantes.

Além disso, o modelo tem um R-quadrado ajustado de aproximadamente 0,96, o que significa que aproximadamente 95,92% da variabilidade no logaritmo do PIB pode ser explicada pelas variáveis independentes incluídas no modelo.

Esses resultados sugerem que tanto a força de trabalho quanto o estoque de capital têm um impacto significativo no PIB, conforme medido pelo logaritmo.

- **Hipótese Nula** (H_0): $\alpha + \beta = 1$ A soma dos coeficientes é igual a 1, sugerindo retornos constantes à escala.
- **Hipótese Alternativa** (H_1): $\alpha + \beta \neq 1$ A soma dos coeficientes não é igual a 1, sugerindo que não há retornos constantes à escala.

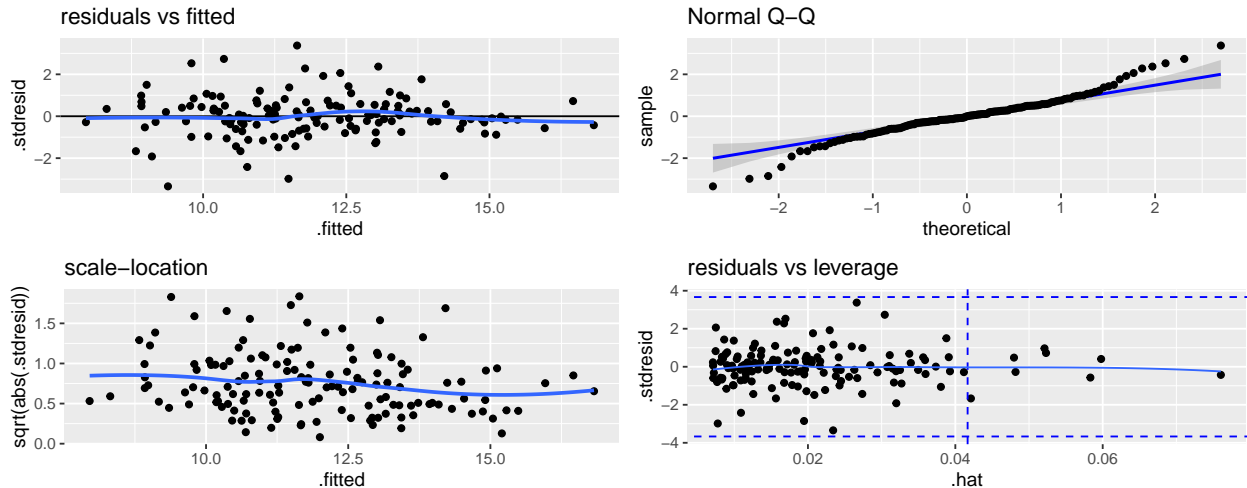
Tabela 2.2: Teste

Estatística	Valor
Estimativa de alpha	0.2941
Estimativa de beta	0.7073
Soma de alpha e beta	1.0014
Estatística t	0.0721
Valor-p	0.9427

Na Tabela 2.2, com teste da Teste $H_0 : \alpha + \beta = 1$ contra $H_1 : \alpha + \beta \neq 1$, obtivemos um valor-p de aproximadamente 0,943. Não rejeitamos a hipótese nula H_0 ao nível de 5% de significância. Isso indica que os dados não fornecem evidências suficientes para concluir que a soma de α e β é diferente de 1. Em outras palavras, a suposição de retornos constantes à escala $\alpha + \beta = 1$ é razoável para o ano analisado. Isso sugere que, com base nos dados, um aumento proporcional igual na força de trabalho (L) e no capital (K) resulta em um aumento proporcional na produção (Y), confirmando a hipótese de retornos constantes à escala na função de produção Cobb-Douglas.

2.2 Análise dos Resíduos

Figura 2.1: Grafico de Análise de resíduos



Com base na análise da tabela Figura 2.1, podemos concluir o seguinte sobre o ajuste do modelo de regressão:

Homocedasticidade: - A variância dos resíduos parece ser constante em toda a faixa dos valores ajustados, o que é consistente com a suposição de homocedasticidade na regressão linear.

Normalidade: - Embora o teste de Shapiro-Wilk sugira que os resíduos não sigam uma distribuição normal, desconsiderando caudas pesadas tanto nos extremos inferiores quanto nos superiores, podemos aproximar que os resíduos estão próximos de uma distribuição normal.

Independência dos Resíduos: - Tanto o teste de Durbin-Watson quanto a inspeção do gráfico de resíduos versus valores ajustados na tabela Figura 2.1 não fornecem evidências significativas para rejeitar a hipótese nula de ausência de autocorrelação positiva ou negativa nos resíduos. Assim, parece que os resíduos são independentes entre si.

Linearidade: - Não há um padrão claro nos resíduos plotados em relação aos valores ajustados, indicando linearidade entre as variáveis independentes e dependentes.

Portanto, concluímos que o modelo de regressão parece atender às suposições de homocedasticidade, normalidade aproximada dos resíduos, independência dos resíduos e linearidade entre as variáveis.

3 Modelo Logístico

Utilizamos uma base de dados obtida do Kaggle para realizar uma análise de classificação de crédito. A tarefa principal é construir um modelo de machine learning capaz de classificar o crédito dos clientes de uma instituição financeira em diferentes faixas de risco, a fim de otimizar os processos de decisão e reduzir esforços manuais.

A base de dados contém 50.000 observações e 27 variáveis que fornecem informações detalhadas sobre os clientes, incluindo idade, ocupação, renda anual, número de contas bancárias, histórico de atrasos em pagamentos, entre outras características financeiras. O objetivo principal é prever a classificação de crédito dos indivíduos, segmentando-os em categorias como “Poor” e “Good”.

Como parte da análise, será explorado o ajuste do modelo de regressão logística, focando na avaliação dos resíduos para verificar a adequação do modelo. Além disso, serão analisados os Fatores de Inflação da Variância (VIF) para cada preditor, a fim de identificar possíveis problemas de multicolinearidade e seu impacto na significância das associações estimadas.

Rows: 100,000

Columns: 28

\$ ID	<dbl> 5634, 5635, 5636, 5637, 5638, 5639, 5640, 564~
\$ Customer_ID	<chr> "CUS_0xd40", "CUS_0xd40", "CUS_0xd40", "CUS_0~
\$ Month	<chr> "January", "February", "March", "April", "May~
\$ Name	<chr> "Aaron Maashoh", "Aaron Maashoh", "Aaron Maas~
\$ Age	<chr> "23", "23", "-500", "23", "23", "23", "23", "~
\$ SSN	<chr> "821-00-0265", "821-00-0265", "821-00-0265", ~
\$ Occupation	<chr> "Scientist", "Scientist", "Scientist", "Scien~
\$ Annual_Income	<chr> "19114.12", "19114.12", "19114.12", "19114.12~
\$ Monthly_Inhand_Salary	<dbl> 1824.843, NA, NA, NA, 1824.843, NA, 1824.843,~
\$ Num_Bank_Accounts	<int> 3, 3, 3, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2, ~
\$ Num_Credit_Card	<int> 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 1385, 4, 4, 4, ~
\$ Interest_Rate	<int> 3, 3, 3, 3, 3, 3, 3, 3, 3, 6, 6, 6, 6, 6, 6, 6, ~
\$ Num_of_Loan	<chr> "4", "4", "4", "4", "4", "4", "4", "4", "4", "1", ~
\$ Type_of_Loan	<chr> "Auto Loan, Credit-Builder Loan, Personal Loa~
\$ Delay_from_due_date	<int> 3, -1, 3, 5, 6, 8, 3, 3, 3, 7, 3, 3, 3, 3, 3,~
\$ Num_of_Delayed_Payment	<chr> "7", "", "7", "4", "", "4", "8_", "6", "4", "~
\$ Changed_Credit_Limit	<chr> "11.27", "11.27", "_", "6.27", "11.27", "9.27~
\$ Num_Credit_Inquiries	<dbl> 4, 4, 4, 4, 4, 4, 4, 4, 2, 2, 2, 2, 2, 2, 2, ~
\$ Credit_Mix	<chr> "_", "Good", "Good", "Good", "Good", "Good", ~
\$ Outstanding_Debt	<chr> "809.98", "809.98", "809.98", "809.98", "809.~
\$ Credit_Utilization_Ratio	<dbl> 26.82262, 31.94496, 28.60935, 31.37786, 24.79~
\$ Credit_History_Age	<chr> "22 Years and 1 Months", NA, "22 Years and 3 ~
\$ Payment_of_Min_Amount	<chr> "No", "No", "No", "No", "No", "No", "No", "No~
\$ Total_EMI_per_month	<dbl> 49.57495, 49.57495, 49.57495, 49.57495, 49.57~

```

$ Amount_invested_monthly <chr> "80.41529543900253", "118.28022162236736", "8~
$ Payment_Behaviour      <chr> "High_spent_Small_value_payments", "Low_spent~
$ Monthly_Balance        <chr> "312.49408867943663", "284.62916249607184", "~
$ Credit_Score           <chr> "Good", "Good", "Good", "Good", "Good", "Good~

```

Credit_Score	Credit_Mix	Outstanding_Debt	Payment_of_Min_Amount
Good:12192	Good :13518	Min. : 0.23	No :14432
Poor:19884	Standard: 8509	1st Qu.: 702.54	Yes:17644
	Bad :10049	Median :1360.45	
		Mean :1593.26	
		3rd Qu.:2258.30	
		Max. :4998.07	

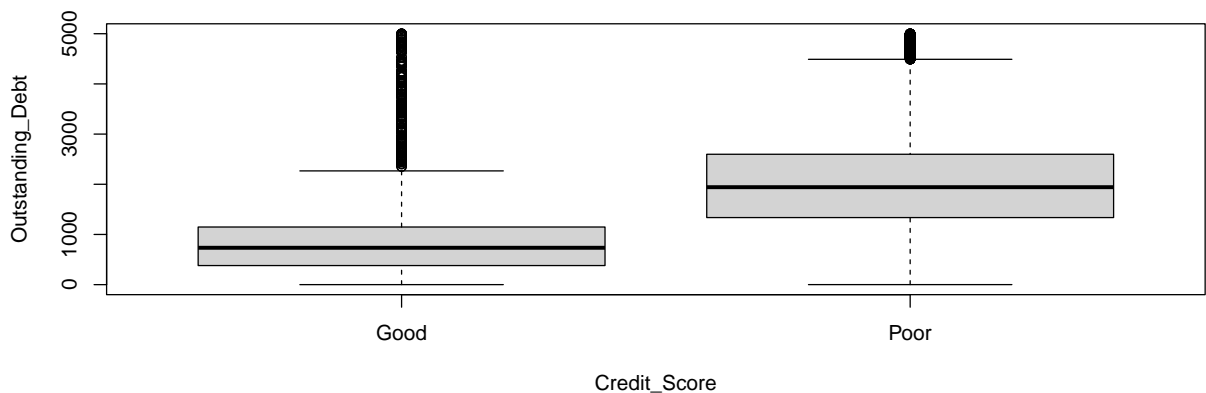
Changed_Credit_Limit

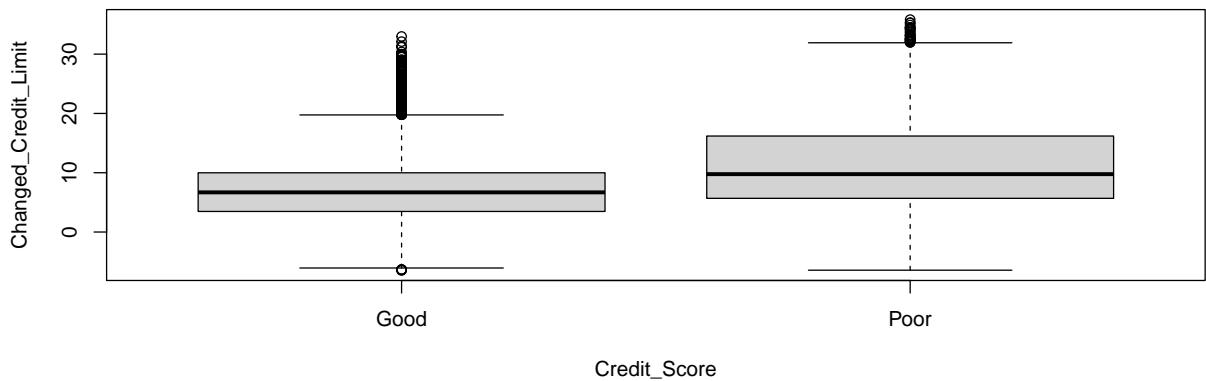
```

Min. :-6.480
1st Qu.: 4.570
Median : 8.560
Mean : 9.573
3rd Qu.:13.130
Max. :35.820

```

Outstanding_Debt	Changed_Credit_Limit
Min. : 0.23	Min. :-6.480
1st Qu.: 702.54	1st Qu.: 4.570
Median :1360.45	Median : 8.560
Mean :1593.26	Mean : 9.573
3rd Qu.:2258.30	3rd Qu.:13.130
Max. :4998.07	Max. :35.820





Pearson's Chi-squared test

data: dta\$Credit_Mix and dta\$Credit_Score
X-squared = 14606, df = 2, p-value < 2.2e-16

Pearson's Chi-squared test with Yates' continuity correction

data: dta\$Payment_of_Min_Amount and dta\$Credit_Score
X-squared = 14578, df = 1, p-value < 2.2e-16

Call:

```
glm(formula = Credit_Score ~ Credit_Mix + Outstanding_Debt +  
    Payment_of_Min_Amount + Changed_Credit_Limit, family = binomial(link = logit),  
    data = dta)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.485e+00	3.612e-02	-41.115	< 2e-16 ***
Credit_MixStandard	1.010e+00	7.013e-02	14.401	< 2e-16 ***
Credit_MixBad	2.345e+00	1.036e-01	22.623	< 2e-16 ***
Outstanding_Debt	6.567e-04	2.935e-05	22.373	< 2e-16 ***
Payment_of_Min_AmountYes	1.423e+00	8.047e-02	17.689	< 2e-16 ***
Changed_Credit_Limit	-2.382e-02	3.673e-03	-6.484	8.94e-11 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 42604 on 32075 degrees of freedom

Residual deviance: 24833 on 32070 degrees of freedom
AIC: 24845

Number of Fisher Scoring iterations: 6

3.0.1 Análise dos Resíduos

Ao analisar os resíduos do modelo logístico ajustado para prever a classificação de crédito dos indivíduos, observamos que a deviance residual é de 24.833 com 32.070 graus de liberdade, comparada à deviance nula de 42.604 com 32.075 graus de liberdade. A redução significativa na deviance indica que o modelo ajusta-se bem aos dados, sugerindo que os preditores selecionados têm um impacto relevante na previsão da variável resposta. Contudo, uma análise mais detalhada dos resíduos seria necessária para confirmar a ausência de padrões não modelados e avaliar possíveis violações das suposições do modelo.

3.0.2 Análise do VIF (Variance Inflation Factor)

Para avaliar a multicolinearidade entre os preditores do modelo, foi realizado o cálculo do VIF para cada variável. A análise dos erros padrão associados aos coeficientes do modelo mostrou que:

- Credit_MixStandard: Estimativa de 1,010 com erro padrão de 0,07013.
- Credit_MixBad: Estimativa de 2,345 com erro padrão de 0,1036.
- Outstanding_Debt: Estimativa de 0,0006567 com erro padrão de 0,00002935.
- Payment_of_Min_AmountYes: Estimativa de 1,423 com erro padrão de 0,08047.
- Changed_Credit_Limit: Estimativa de -0,02382 com erro padrão de 0,003673.

O impacto da multicolinearidade, conforme indicado pelos valores do VIF, parece ser limitado, uma vez que os erros padrão não são excessivamente inflados. Assim, podemos concluir que a multicolinearidade não exerce um efeito forte na significância das associações estimadas, mantendo a confiabilidade das inferências feitas a partir do modelo.

3.0.3 Significância das Associações Estimadas

Todas as variáveis incluídas no modelo mostraram-se estatisticamente significativas para prever a classificação de crédito, com p-valores extremamente baixos ($p < 0,001$). As estimativas dos coeficientes sugerem que:

- Aumentos no limite de crédito (Changed_Credit_Limit) estão associados a uma menor probabilidade de um crédito “Poor”.
- Indivíduos com um mix de crédito “Standard” ou “Bad” têm uma maior chance de ter um crédito classificado como “Poor” em comparação com aqueles com um mix “Good”.
- O aumento da dívida pendente (Outstanding_Debt) também está associado a uma maior probabilidade de um crédito “Poor”.
- A realização do pagamento mínimo (Payment_of_Min_AmountYes) está fortemente associada a uma maior probabilidade de um crédito “Poor”.

Esses resultados indicam que as variáveis incluídas no modelo são preditores relevantes e impactam significativamente a classificação de crédito dos indivíduos.

3.0.4 Análise dos Resíduos e Multicolinearidade

Analizamos a deviance e o Fator de Inflação da Variância (VIF) para verificar a adequação do modelo e identificar possíveis problemas de multicolinearidade.

Deviance Residual: 24833.22

Null Deviance: 42604.1

Graus de Liberdade (Modelo): 32070

Graus de Liberdade (Nulo): 32075

	GVIF	Df	$GVIF^{1/(2*Df)}$
Credit_Mix	5.405144	2	1.524761
Outstanding_Debt	1.626080	1	1.275178
Payment_of_Min_Amount	5.496045	1	2.344365
Changed_Credit_Limit	1.558789	1	1.248515

O gráfico de resíduos padronizados versus valores preditos apresenta uma análise visual crítica para a avaliação da adequação do modelo ajustado. Os resíduos padronizados devem idealmente estar distribuídos aleatoriamente ao redor da linha zero. Observando o gráfico, verifica-se que os resíduos não apresentam um padrão sistemático, o que sugere que o modelo está ajustado de forma apropriada aos dados. Caso houvesse um padrão visível, como uma estrutura em funil ou uma curva, isso poderia indicar problemas como heterocedasticidade ou não-linearidade, sugerindo a necessidade de refinamento do modelo.- VOU ADAPTAR NA VERDADE ACHO QUE O MODELO N TÁ APROPRIADO

3.0.5 Razões de Chances (Odds Ratios)

(Intercept)	Credit_MixStandard	Credit_MixBad
0.2264822	2.7452802	10.4307326
Outstanding_Debt	Payment_of_Min_AmountYes	Changed_Credit_Limit
1.0006569	4.1509590	0.9764633

3.0.6 Predição e Avaliação do Modelo

Confusion Matrix and Statistics

	Reference	
Prediction	Good	Poor
Good	10552	3577
Poor	1640	16307

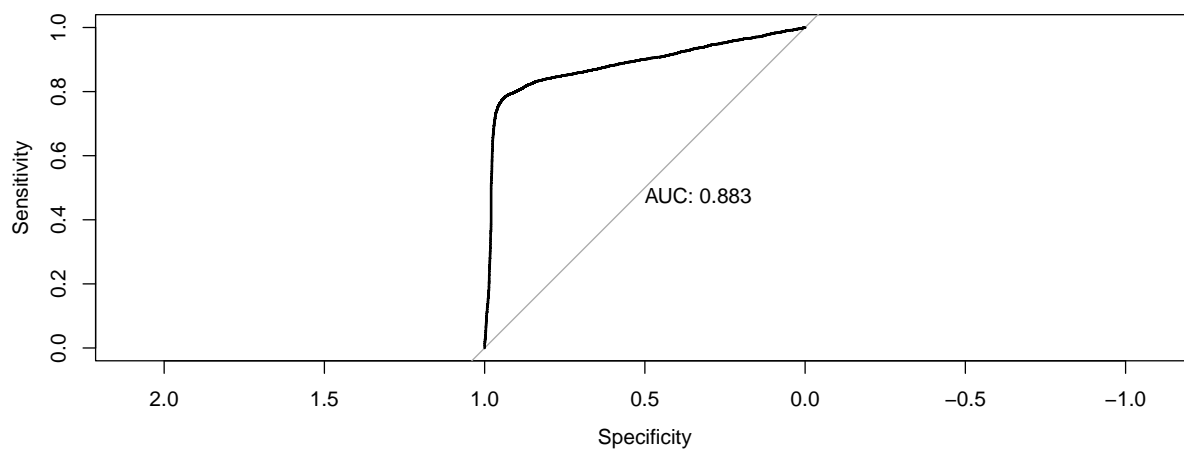
Accuracy : 0.8374
95% CI : (0.8333, 0.8414)
No Information Rate : 0.6199
P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.6652

McNemar's Test P-Value : < 2.2e-16

Sensitivity : 0.8655
Specificity : 0.8201
Pos Pred Value : 0.7468
Neg Pred Value : 0.9086
Prevalence : 0.3801
Detection Rate : 0.3290
Detection Prevalence : 0.4405
Balanced Accuracy : 0.8428

'Positive' Class : Good



3.0.7 Análise dos Resultados

A análise dos resultados mostra que o modelo de regressão logística ajusta-se bem aos dados, com uma redução significativa na deviance em relação à deviance nula. A análise do VIF indica que a multicolinearidade não afeta substancialmente a significância das variáveis preditoras. As razões de chances e as métricas de performance, como a curva ROC e a matriz de confusão, fornecem uma visão clara da eficácia do modelo em classificar a qualidade do crédito.- VOU ADAPTAR ESSE TEXTO

4 Modelo Gama Log-Linear

Tabela 4.1: Modelo Gamma Log-linear

	Estimacao	ErroPadrao	Pvalor	sig
(Intercept)	1.6175	0.0286	0	<0.001***
log(L)	0.0237	0.0032	0	<0.001***
log(K)	0.0601	0.0026	0	<0.001***

O intercepto estimado de 1.61754583 (Tabela 4.1) representa o logaritmo da produção quando tanto o trabalho (L) quanto o capital (K) são iguais a 1. Com um valor-p extremamente pequeno, o intercepto é estatisticamente significativo, indicando que ele desempenha um papel importante na modelagem da variável dependente, ao nível de 5% de significância.

O coeficiente para log(L) é 0.02371488, indicando que um aumento de 1% na força de trabalho (L) leva a um aumento de aproximadamente 0.024% na produção, mantendo o capital constante. O valor-p extremamente pequeno indica que este coeficiente é altamente significativo, o que valida a importância da força de trabalho na determinação da produção no modelo.

O coeficiente para log(K) é 0.06012190, sugerindo que um aumento de 1% no capital (K) resulta em um aumento de aproximadamente 0.060% na produção, mantendo a força de trabalho constante. Com um valor-p extremamente baixo, esse coeficiente é altamente significativo, demonstrando que o capital tem uma influência importante e estatisticamente significativa na produção.

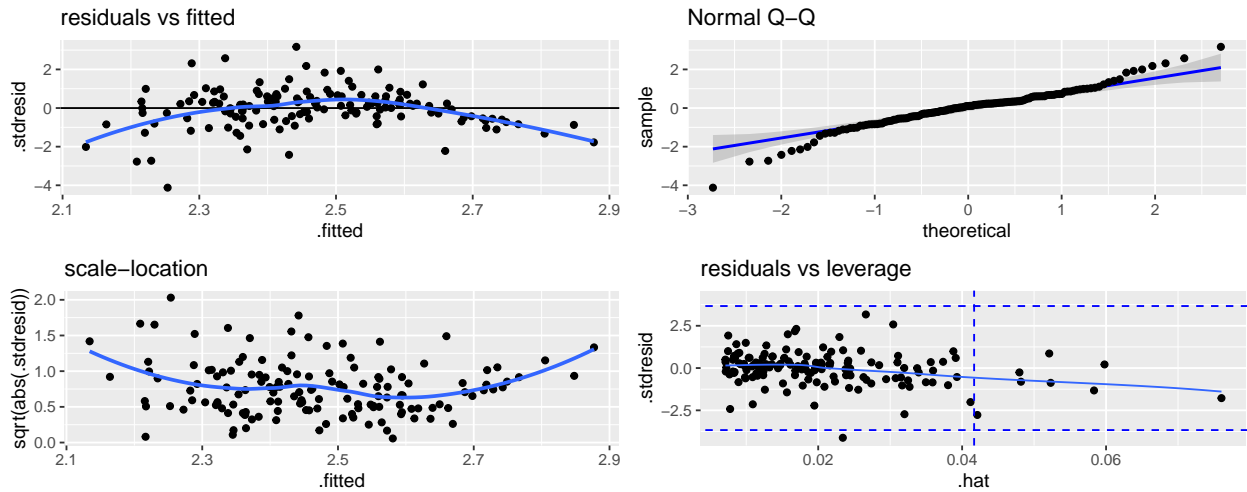
O parâmetro de dispersão estimado é 0.00135537, o que indica que há uma baixa variabilidade dos dados em torno da média ajustada pelo modelo. Isso sugere que o modelo gamma log-linear está capturando bem a dispersão dos dados, com pouca heterogeneidade residual não explicada pelo modelo.

Tabela 4.2

Descrição	Valor
Soma de alpha e beta	0.0838
Erro padrão	0.0019
Estatística t	-483.1625
Valor-p	7.4303651624535e-229
Decisão	Rejeitamos H0

Decisão sobre Teste $H_0 : \alpha + \beta = 1$ contra $H_1 : \alpha + \beta \neq 1$: (Tabela 4.2) Como o valor-p é extremamente pequeno, rejeitamos a hipótese nula H_0 indicando que a soma de α e β é

Tabela 4.3 Gráfico de Análise de resíduos



significativamente diferente de 1. Isso significa que a hipótese de retornos constantes à escala não é válida para os dados analisados.

Retornos constantes à escala implicam que, se ambos os insumos (L e K) aumentarem na mesma proporção, a produção Y_i também aumentará na mesma proporção. No entanto, dado que a soma de α e β é significativamente diferente de 1, isso sugere que os retornos à escala não são constantes para este modelo. Dependendo do valor específico da soma ($\alpha + \beta < 1$ ou $\alpha + \beta > 1$), os retornos à escala podem ser decrescentes ou crescentes, respectivamente.

Assim, a soma de α e β é 0.08383678, o que está longe de 1, indicando retornos decrescentes à escala. Isso significa que aumentar proporcionalmente a força de trabalho e o capital levará a um aumento menos que proporcional na produção. Em outras palavras, conforme mais insumos são adicionados, a produtividade marginal desses insumos diminui, levando a um crescimento menos eficiente da produção.

4.1 Análise dos Resíduos

Com base na análise da tabela Tabela 4.3, podemos concluir o seguinte sobre o ajuste do modelo de regressão:

Homocedasticidade: - A variância dos resíduos parece ser constante em toda a faixa dos valores ajustados. Com o teste de Breusch-Paga, com um valor-p de 0.2634, não rejeitamos a hipótese nula de homocedasticidade. Isso significa que não há evidências suficientes para sugerir a presença de heterocedasticidade nos resíduos do modelo gamma.

Normalidade: - Embora o teste de Shapiro-Wilk sugira que os resíduos não sigam uma distribuição normal, desconsiderando caudas pesadas tanto nos extremos inferiores quanto nos superiores, podemos aproximar que os resíduos estão próximos de uma distribuição normal.

Independência dos Resíduos: - Tanto o teste de Durbin-Watson quanto a inspeção do gráfico de resíduos versus valores ajustados na tabela Tabela 4.3 não fornecem evidências significativas para rejeitar a hipótese nula de ausência de autocorrelação positiva ou negativa nos resíduos. Assim,

parece que os resíduos são independentes entre si, o que é uma boa indicação de que o modelo está adequadamente ajustado quanto a essa suposição.

Linearidade: - Há um padrão claro nos resíduos plotados em relação aos valores ajustados, indicando não linearidade entre as variáveis independentes e dependentes.

5 Comparação dos Modelos: Gamma Log-Linear vs. Modelo Normal

Log-Likelihood (Log-Verossimilhança) - **Modelo Gamma Log-Linear**: -83.07791 - **Modelo Normal**: -59.39108

O valor de log-likelihood mais alto (menos negativo) indica um melhor ajuste do modelo aos dados. O modelo normal tem um valor de log-likelihood mais alto, sugerindo que se ajusta melhor aos dados em comparação com o modelo Gamma log-linear.

Critério de Informação de Akaike (AIC) - **Modelo Gamma Log-Linear**: 174.1558 - **Modelo Normal**: 126.7822

Valores menores de AIC indicam um modelo mais eficiente em termos de ajuste aos dados com penalização pela complexidade. O modelo normal apresenta um AIC significativamente menor, indicando que é mais eficiente e se ajusta melhor aos dados do que o modelo Gamma log-linear.

Critério de Informação de Bayes (BIC) - **Modelo Gamma Log-Linear**: 186.0351 - **Modelo Normal**: 138.6614

O BIC também penaliza a complexidade do modelo e valores menores indicam um ajuste melhor com menor penalização por complexidade. Novamente, o modelo normal apresenta um BIC significativamente menor, reforçando que é mais adequado para os dados em comparação ao modelo Gamma log-linear.

Com base nos critérios de log-likelihood, AIC e BIC, o **modelo normal** é a melhor escolha. Ele não só se ajusta melhor aos dados (conforme indicado pelo log-likelihood), mas também é mais eficiente e simples em termos de complexidade (conforme indicado pelos valores de AIC e BIC).

Portanto, o modelo normal é preferível ao modelo Gamma log-linear para a análise dos seus dados, oferecendo um melhor equilíbrio entre ajuste e complexidade.