

Trabalho de Dados Categorizados 1/2024

Profa. Maria Tereza Leão Costa

Bruno Gondim Toledo (15/0167636)

Rafael de Acypreste (20/0060023)

15/07/2024

Sumário

1	Introdução	2
1.1	Análise exploratória	2
2	Metodologia	4
2.1	Análise de Regressão Logística	4
2.1.1	Métricas para seleção do modelo logístico	6
3	Resultados	8
3.1	Modelo apenas com intercepto	8
3.2	Modelo acrescido do Nível de Fosfatase Ácida como variável preditora	9
3.2.1	Comparação dos modelos	11
3.3	Modelo Completo	11
3.3.1	Comparação dos modelos	14
3.4	Demais modelos possíveis	14
3.5	Análise a partir do modelo escolhido	14
3.6	Aplicação do modelo escolhido ao conjunto de teste	16
4	Conclusão	16
5	Referências	17
6	Apêndice A - códigos em R	17
7	Apêndice B - códigos em SAS	26
7.1	Códigos	26
7.2	Saídas SAS	32

1 Introdução

Ao se constatar que um paciente desenvolveu câncer, é fundamental, para se decidir qual tratamento utilizar, saber se o câncer já se espalhou para os linfonodos próximos. Um estudo foi realizado com o objetivo de medir a capacidade de predição para o envolvimento nodal em câncer de próstata de várias variáveis pré-operatórias cuja coleta é menos invasiva que uma cirurgia.

- Numa primeira etapa, desejava-se avaliar especificamente o efeito do nível de fosfatase ácida na predição para envolvimento nodal; e
- Na segunda etapa do estudo se considerou além desta variável as outras variáveis pré-operatórias.

Com isso, o presente estudo objetiva analisar os dados coletados e verificar a capacidade de predição para o envolvimento nodal em câncer de próstata.

1.1 Análise exploratória

Os dados avaliados consistem no acompanhamento clínico de 102 pacientes com câncer de próstata. Um resumo das variáveis pode ser visto na Tabela 1.

Tabela 1: Resumo dos dados.

Variável	N = 102 ¹
Resultado da Radiografia	
Negativo	72 (71%)
Positivo	30 (29%)
Estágio do Tumor	
- grave	47 (46%)
+ grave	55 (54%)
Nível da Fosfatase Ácida	62 (50, 79)
Envolvimento Nodal (x100)	
Não	60 (59%)
Sim	42 (41%)

¹n (%); Median (IQR)

Sobre as variáveis, dispomos no conjunto da amostra 4 colunas de variáveis, sendo elas:

- X1 (Resultado da radiografia): Variável categórica binária, em que o valor 0 indica resultado negativo e o valor 1 indica resultado positivo.

- X2 (Estágio do tumor): Variável categórica binária, em que o valor 0 indica que o tumor é menos grave, enquanto 1 indica tumor mais grave.
- X3 (Nível de fosfatase ácida): Variável quantitativa contínua indicando o nível desta enzima, aqui multiplicada por 100.
- X4 (Envolvimento nodal): Variável categórica binária, em que o valor 0 significa não haver envolvimento nodal, enquanto o valor 1 indica haver envolvimento nodal.

Existe ainda uma coluna de identificação do paciente, com um valor único por linha, que não será considerada no escopo desta análise.

A variável resposta de interesse é de Envolvimento Nodal. Nesse sentido, a Tabela 2 apresenta a distribuição das variáveis coletadas por pessoas que apresentaram ou não o envolvimento.

Tabela 2: Distribuição das variáveis por envolvimento nodal.

Envolvimento nodal	Não, N = 60^I	Sim, N = 42^I
Resultado da Radiografia		
Negativo	52 (87%)	20 (48%)
Positivo	8 (13%)	22 (52%)
Estágio do Tumor		
- grave	39 (65%)	8 (19%)
+ grave	21 (35%)	34 (81%)
Nível da Fosfatase Ácida	53 (50, 63)	74 (67, 84)

^I_n (%); Median (IQR)

O que se pode perceber é que as variáveis relativas ao resultado da tomografia, do estágio do tumor e do nível da Fosfatase ácida são mais elevados entre os pacientes que apresentaram envolvimento nodal. Entretanto, é preciso usar métodos estatísticos adequados para verificar se essas diferenças são significativas.

A relação entre o Envolvimento Nodal e o nível da Fosfatase Ácida pode ser vista na Figura 1. Por ela, também é reforçada a ideia de que o nível da Fosfatase Ácida é mais elevado entre os pacientes que apresentaram envolvimento nodal.

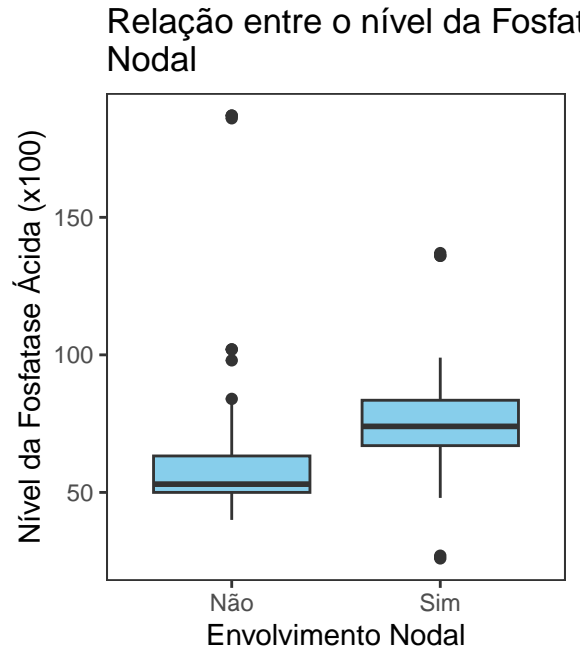


Figura 1: Relação entre o nível da Fosfatase Ácida e o Envolvimento Nodal.

2 Metodologia

2.1 Análise de Regressão Logística

A regressão logística é uma das técnicas disponíveis na família dos modelos lineares generalizados (MLG). Segundo Casella (2022), um MLG descreve uma relação entre a média de uma variável resposta Y e uma variável independente x . No caso do modelo de regressão logística, teremos três componentes, sendo eles: o aleatório, o sistemático e a função de ligação, em que:

- As variáveis resposta Y_1, \dots, Y_n são o componente aleatório. Elas são consideradas variáveis aleatórias independentes, cada uma com uma distribuição de uma família exponencial específica. Estes vetores não precisam ser identicamente distribuídos, mas cada um deles deve ter uma distribuição da mesma família.
- O componente sistemático é o modelo. Esta é a função da variável preditora x_i , linear **nos parâmetros**, que é relacionada à **média** de Y_i .
- Por fim, a função de ligação $g(\mu_i)$ estabelece ligação entre os dois componentes, afirmando que $g(\mu_i) = \beta_0 + \beta_i x_i$, em que $\mu_i = \mathbb{E}(Y_i)$

De forma geral, a função de ligação canônica para a regressão logística será a função *logito*, descrita como

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_{i(p \times 1)} X_{i(p \times 1)}, \quad (1)$$

em que $\beta_{i(p \times 1)} = [\beta_{i0}, \beta_{i1}, \dots, \beta_{ip-1}]^T$, e $\mathbf{X}_{i(p \times 1)} = [1, X_{i1}, X_{i2}, \dots, X_{ip-1}]^T$, visto que para este modelo o vetor de respostas Y_1, \dots, Y_n são independentes e $Y_i \sim \text{Bernoulli}(\pi_i), \forall i$.

Para a resposta às duas perguntas do problema, pode-se utilizar o modelo de regressão logística. Como a variável resposta, Envolvimento Nodal, é uma variável binária, pode-se codificá-la para que assuma o valor 1 se o paciente apresentou envolvimento nodal e 0 caso contrário, tal que a resposta esperada será igual a probabilidade $p = \mathbb{E}(Y)$ de um paciente apresentar envolvimento nodal. Segundo Morettin (2023) assim como no caso de modelos de regressão linear, o objetivo da análise é modelar a resposta esperada, que neste caso é uma probabilidade, como função da variável explicativa.

Numa primeira etapa, para se avaliar apenas a capacidade do nível de fosfatase ácida em prever o envolvimento nodal, pode-se ajustar o modelo de regressão logística com a variável nível de fosfatase ácida como preditora. O modelo utilizado é dado pela equação (2).

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_{3i} \quad (2)$$

em que π_i é a probabilidade de um paciente i apresentar envolvimento nodal, X_{3i} é o nível de fosfatase ácida do paciente i e β_0 e β_1 são os coeficientes do modelo. A função de ligação logito é dada por $\log\left(\frac{\pi_i}{1-\pi_i}\right)$, que tem como resposta o modelo linear visto na equação.

De maneira específica, o coeficiente e^{β_1} indica o quanto a razão de chances de um paciente apresentar envolvimento nodal aumenta para cada unidade de aumento no nível de fosfatase ácida.

Por fim, o modelo completo inclui todas as variáveis disponíveis de Estágio do Tumor, Resultado da Radiografia e Nível da Fosfatase Ácida. Nesse caso, o modelo é dado pela equação (3).

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} \quad (3)$$

em que X_{1i} , X_{2i} e X_{3i} são as variáveis Nível da Fosfatase Ácida, Resultado da Radiografia e Estágio do Tumor, respectivamente.

Na regressão logística, como não há mais uma conexão direta (por conta da função de ligação) entre Y_i e $\beta_{i(p \times 1)}$ como existe no caso da regressão linear, a opção de ajustar o modelo pelo método de mínimos quadrados não é mais possível. Portanto, o ajuste do modelo dar-se-á pelo método de máxima verossimilhança, que pode ser descrito da seguinte forma:

Se considerarmos o modelo geral $Y_i \sim \text{Bernoulli}(\pi_i)$, onde $\pi(x) = F(\alpha + \beta x)$, para seguir a notação de Casella (2022), em que α representa o intercepto do modelo, que na matéria

convencionamos chamar de β_0 , tal que $\alpha = \beta_0$. Portanto, se considerarmos que $F_i = F(\alpha + \beta x_i)$, então a função de verossimilhança é:

$$L(\alpha, \beta|y) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} = \prod_{i=1}^n F_i^{y_i} (1 - F_i)^{1-y_i}, \quad (4)$$

Com Log verossimilhança $\ell(\alpha, \beta|y)$ dada por:

$$\ell(\alpha, \beta|y) = \sum_{i=1}^n \left\{ \log(1 - F_i) + y_i \log\left(\frac{F_i}{1 - F_i}\right) \right\}. \quad (5)$$

Visto que o logarítmo é uma função monótona, a maximização desta verossimilhança pode ser obtida maximizando (5). Para isso, os estimadores de máxima verossimilhança de α e β correspondem à solução das equações de estimação $\sum_{i=1}^n \left\{ y_i - \frac{\exp(\hat{\alpha} + \hat{\beta} x_i)}{1 - \exp(\hat{\alpha} + \hat{\beta} x_i)} \right\} = 0$ e $\sum_{i=1}^n x_i \left\{ y_i - \frac{\exp(\hat{\alpha} + \hat{\beta} x_i)}{1 - \exp(\hat{\alpha} + \hat{\beta} x_i)} \right\} = 0$. Conforme Morettin (2023), esse sistema de equações não tem solução explícita, logo deve-se recorrer a métodos iterativos como o método de Newton-Raphson para uma solução numérica, que é o que os algoritmos implementados em ferramentas como SAS e R aplicam para estimar os modelos.

2.1.1 Métricas para seleção do modelo logístico

Para testar o ajuste aos dados, podemos aplicar técnicas como o teste da razão de verossimilhanças, o teste de Wald ou o teste score, que são descritos a seguir

2.1.1.1 Teste da razão de verossimilhança

Seja $\ell(\hat{\beta}^{(0)})$ o valor do logarítmo da função de verossimilhança do modelo sob H_0 , e $\ell(\hat{\beta}^{(1)})$ o valor do logarítmo da função de verossimilhança do modelo sob H_1 , então a estatística de teste é dada por $G = -2(\ell(\hat{\beta}^{(0)}) - \ell(\hat{\beta}^{(1)}))$, que segue uma distribuição χ^2 com p graus de liberdade, em que p é o número de parâmetros adicionados ao modelo da hipótese alternativa.

2.1.1.2 Teste de Wald

Seja o vetor de parâmetros do modelo ajustado $\hat{\beta}$ assintoticamente normal, tal que $\hat{\beta} \sim N_p(\beta, K^{-1})$, onde $K = \sigma^{-2} X^T X$ é a matriz de informação (Cordeiro, 2024, p. 79). Sob a hipótese nula $H_0) \beta_k = 0$, a estatística do teste é dada por $z^* = \frac{\beta_k}{s_{\{\beta_k\}}} \sim N(0, 1)$ no caso de um parâmetro, que é a situação onde este teste será aplicado.

2.1.1.3 Teste de escore

O teste de escore de Rao é definida por Cordeiro (2024) a partir da função escore como $S_R = U_1^T(\tilde{\beta})\widetilde{Cov}(\tilde{\beta}_1)U_1(\tilde{\beta})$, sendo $\widetilde{Cov}(\tilde{\beta}_1)$ a matriz $Cov(\tilde{\beta}_1)$ avaliada nos vetores de parâmetros dos modelos sob H_0 e H_1 , e $U_1(\tilde{\beta})$ o vetor de derivadas parciais da função de verossimilhança em relação aos parâmetros do modelo sob H_1 . Essa estatística tem distribuição qui-quadrado, com número de graus de liberdade igual a quantidade de parâmetros adicionais da hipótese alternativa.

2.1.1.4 Deviance

Uma das métricas possíveis para avaliação e seleção de modelos é a análise da *deviance* do modelo. Segundo Cordeiro et al. (2024), a análise de desvio (se referindo à *deviance*) é uma generalização da análise de variância (ANOVA) para os modelos lineares generalizados, visando obter, a partir de uma sequência de modelos encaixados, cada modelo incluindo mais termos do que os anteriores, os efeitos de variáveis explanatórias, os fatores e suas interações. Usa-se o desvio como uma medida de discrepância do modelo e forma-se uma tabela de diferença de desvios, montando assim o que pode ser chamado de tabela ANODEV (“Analysis of Deviance”), em analogia à tabela ANOVA. Buscaremos o modelo com menor valor de *deviance* entre os modelos que se ajustam aos dados.

A comparação entre modelos também pode ser feita utilizando o critério de informação de Akaike (AIC). Giolo (2017) define este critério como sendo: $AIC = -2(\log \text{verossimilhança} - \text{número de parâmetros do modelo})$, em que buscaremos o modelo que minimiza este valor. Podemos utilizar ainda o Critério Bayesiano de Schwarz $SBC_p = -2\ell(b) + p\ln(n)$, ou o critério Log da verossimilhança: $-2\ell(b)$.

2.1.1.5 Curva ROC

A curva ROC (Receiver Operating Characteristic, na sigla em inglês) é uma curva que permite analisar a sensibilidade e especificidade de um modelo para todas as probabilidades de corte (π_0) para definir se uma observação se adequa ou não à variável resposta. Neste estudo, a curva é apresentada para o modelo completo na Figura 6.

No eixo vertical, desenha-se a sensibilidade do modelo, que é a razão de verdadeiros positivos (TP), isto é $P(\hat{y} = 1|y = 1)$. No eixo horizontal, é representado o complementar da especificidade ($1 - P(\hat{y} = 0|y = 0)$). A curva de bissecção representa o modelo aleatório, que não tem poder preditivo. Quanto mais próximo da curva de bissecção, pior é o modelo. A área sob a curva (AUC) é uma medida de qualidade do modelo. Quanto mais próxima a área está de 1, maior a capacidade preditiva do modelo.

2.1.1.6 Teste de Hosmer-Lemeshow

O teste de Hosmer-Lemeshow é uma forma de avaliar a adequação do modelo de regressão logística aos dados. A tarefa principal consiste em dividir proporções da característica de interesse com as que foram previstas pelo modelo (Artes e Barroso, 2023). De maneira simplificada, deve-se ordenar as amostras pelos dados previstos pelo modelo, dividindo-a em g — 10 grupos costumam ser a saída mais adotada — grupos distintos, em ordem crescente de probabilidade. Em seguida, identifica-se a quantidade de observações com a característica em cada um dos grupos O_k , $k = 1, 2, \dots, g$, e calcula-se a média das probabilidades previstas para cada grupo $\hat{\pi}_k$.

A partir disso, determina-se a estatística do teste de Hosmer-Lemeshow, que é dada por:

$$HL = \sum_{k=1}^g \frac{(O_k - n_k \hat{\pi}_k)^2}{n_k \hat{\pi}_k (1 - \hat{\pi}_k)} \quad (6)$$

em que n_k é o número de observações no grupo k . Sob a hipótese nula de que o modelo se ajusta bem aos dados, a estatística de teste tem distribuição qui-quadrado com $g - 2$ graus de liberdade. Com isso, rejeita-se a hipótese de ajuste se $HL > \chi^2_{1-\alpha, g-2}$, em que $\chi^2_{1-\alpha, g-2}$ é o quantil de ordem $1 - \alpha$ da distribuição qui-quadrado com $g - 2$ graus de liberdade.

Com estas ferramentas, estamos preparados para analisar e selecionar um modelo útil que se ajuste bem aos dados.

3 Resultados

3.1 Modelo apenas com intercepto

Como estratégia inicial, pode-se avaliar o modelo com uma média única pra todos os níveis de Fosfatase Ácida. Nesse caso, o modelo é dado pela equação (7).

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 \quad (7)$$

O modelo com esse ajuste pode ser avaliado na Figura 2.

Tabela 3: Estatísticas do modelo de regressão logística com intercepto.

[H]	Deviance	AIC	BIC	Log Likelihood
	138.2	140.2	142.8	-69.1

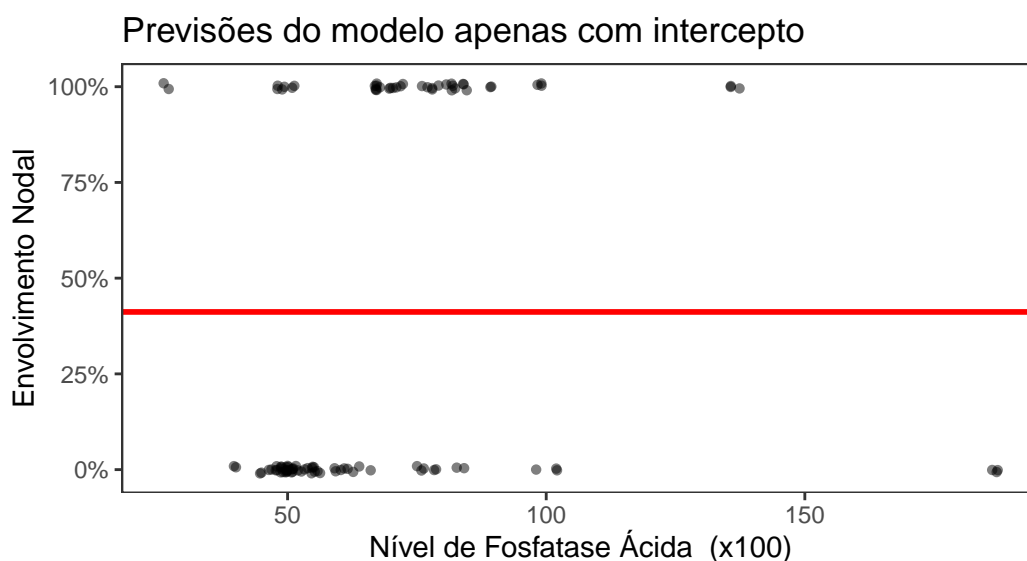


Figura 2: Modelo de regressão logística com intercepto. Os pontos foram ligeiramente agitados para facilitar a visualização.

As estatísticas de avaliação do modelo podem ser vistas na Tabela 3.

Como o modelo inclui apenas o intercepto, a estimativa pontual é dada por 0.412, que é a proporção de pacientes com envolvimento nodal. O intervalo de confiança de 95% para essa proporção é $[0.316; 0.507]$.

3.2 Modelo acrescido do Nível de Fosfatase Ácida como variável preditora

O modelo para o nível de Fosfatase Ácida como variável preditora é dado pela equação (2). As probabilidades previstas estão apresentadas na Figura 3.

Tabela 4: Estatísticas do modelo de regressão logística com o nível de Fosfatase Ácida como variável preditora.

[H]

Variável	Estimativa	Erro Padrao	Valor Z	Pr(> z)
Intercepto	-1.326	0.565	-2.347	0.019
Nível de fosfatase ácida	0.014	0.008	1.828	0.068

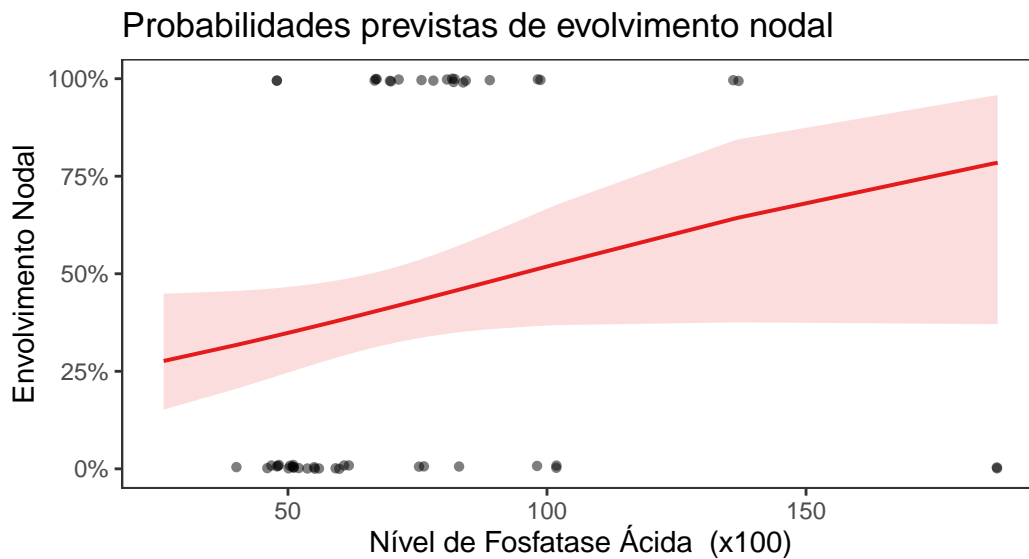


Figura 3: Modelo de regressão logística com o nível de Fosfatase Ácida como variável preditora. Os pontos foram ligeiramente agitados para facilitar a visualização.

As estatísticas do modelo quando é acrescido o nível de Fosfatase Ácida como variável preditora estão apresentadas na Tabela 4.

Como o modelo apresenta uma variável não categórica como variável explicativa, a deviance residual não apresenta distribuição aproximadamente qui-quadrado. Nesse caso, as maneiras tradicionais de análise dos resíduos (como o gráfico de resíduos de Pearson ou dos resíduos padronizados) não são apropriadas (Angresti, 2019, p. 131). Portanto, uma aproximação para avaliar a qualidade do ajuste é o teste de Hosmer-Lemeshow .

O teste de Hosmer-Lemeshow pode ser utilizado para avaliar a qualidade do ajuste. Para o modelo da apresentado na Tabela 4, a estatística χ^2 do teste de Hosmer-Lemeshow é de 26.48 com um p-valor de 8.7×10^{-4} . O teste sugere que o modelo ajustado não é adequado para os dados.

As estatísticas do modelo que inclui a variável “Nível de Fosfatase Ácida” como preditora estão apresentadas na Tabela 5.

Tabela 6: Estatísticas de comparação entre os modelos de regressão logística com a variável “Nível de Fosfatase Ácida” e apenas com intercepto.

[H]	Teste	Estatística	GL	P-valor
	Razao de Verossimilhanca	3.759	1	0.053
	Wald	3.340	1	0.068
	Score	3.759	1	0.053

Tabela 5: Estatísticas de qualidade do modelo de regressão logística com o nível de Fosfatase Ácida como variável preditora.

[H]	Deviance	AIC	BIC	Log Likelihood
	134.45	138.45	143.7	-67.225

Por fim, pela Tabela 4, o valor e_1^β fornece o incremento (ou redução) na razão de chances ao se aumentar uma unidade no Nível de Fosfatase Ácida é zero. O intervalo de confiança de 95% para essa proporção é $[0.9998; 1.031]$. Como o intervalo contém o 1 quando avaliado em 4 casas decimais, não há evidências de que o nível de Fosfatase Ácida influencia o envolvimento nodal ao nível de confiança desejado.

3.2.1 Comparação dos modelos

Uma forma de avaliar o modelo com a variável explicativa “Nível de Fosfatase Ácida” é comparar com o modelo apenas com intercepto. A Tabela 6 apresenta as estatísticas de comparação entre os modelos.

Portanto, percebe-se que nenhum dos 3 testes rejeitaram a hipótese nula a 5%. Isto é, o modelo não é significativamente melhor que o modelo que contém apenas o intercepto.

3.3 Modelo Completo

As demais informações disponíveis podem auxiliar um melhor ajuste do modelo. Nesse caso, o modelo completo é ajustado. As estatísticas do modelo completo estão apresentadas na Tabela 7.

Novamente, como há uma variável quantitativa como variável explicativa, o teste de Hosmer-Lemeshow é sugerido pela literatura. Para o modelo completo a estatística do teste é de 22.05 com um p-valor de 0.00482. O teste sugere que o modelo completo ajustado não é adequado para os dados.

Tabela 7: Estatísticas do modelo de regressão logística completo.

[H]

Variável	Estimativa	Erro Padrao	Valor Z	Pr(> z)
Intercepto	-5.1039	1.1454	-4.4559	0.0000
Resultado da Radiografia	2.7426	0.7360	3.7264	0.0002
Estágio do Tumor	3.2084	0.7543	4.2536	0.0000
Nível de Fosfatase Ácida	0.0278	0.0090	3.0762	0.0021

Pode-se inspecionar as previsões do modelo completo na Figura 4.

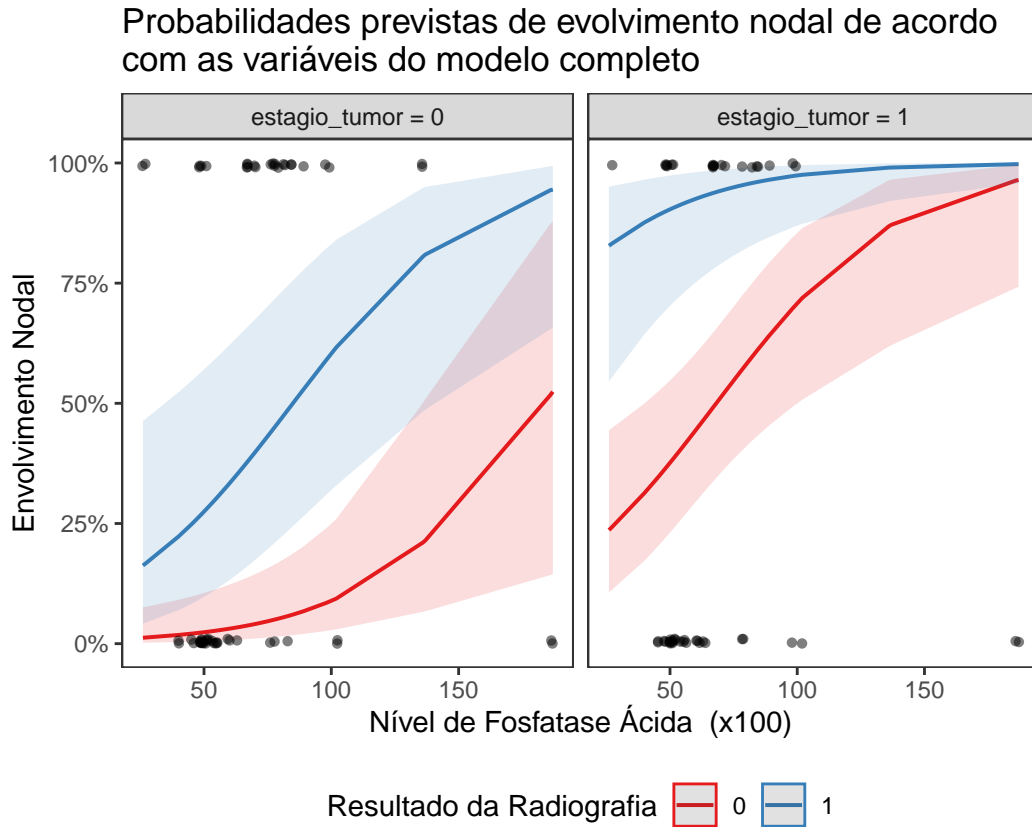


Figura 4: Modelo de regressão logística completo. Os pontos foram ligeiramente agitados para facilitar a visualização.

As estatísticas de qualidade do modelo completo estão apresentadas na Tabela 8.

Também é possível avaliar os resíduos por meio do gráfico de resíduos na Figura 5. Neste tipo de gráfico (“*binned*”, em inglês), os resíduos são agrupados em intervalos e a média

Tabela 8: Estatísticas de qualidade do modelo de regressão logística completo.

[H]	Deviance	AIC	BIC	Log Likelihood
	86.1937	94.1937	104.6936	-43.0969

dos resíduos é plotada contra os valores ajustados, uma vez que os resíduos diretos têm pouca interpretabilidade, uma vez que a variável resposta pode assumir apenas os valores 0 e 1. Nesse sentido, a o polígono azul indica duas vezes o desvio padrão dos resíduos (bandas positiva e negativa). Com isso, espera-se que 95% dos resíduos estejam dentro de tal polígono.

Gráfico residual armazenado

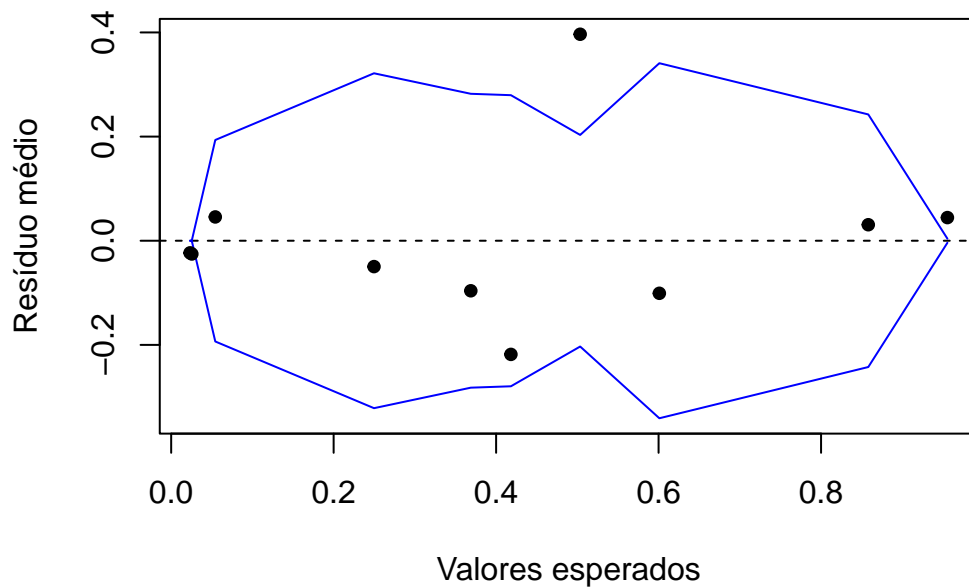


Figura 5: Gráfico de resíduos do modelo de regressão logística completo.

Portanto, pode-se visualizar as estimativas pontuais e seus respectivos intervalos de confiança para a razão de chances conforme a Tabela 9.

De acordo com os dados apresentados, a variável de Nível de Fosfatase Ácida se torna ligeiramente significativa para o modelo de regressão logística. A partir do modelo, o aumento em uma unidade no nível de fosfatase ácida aumenta a razão de chances de envolvimento nodal entre (1.011; 1.048) vezes.

Ademais, o resultado positivo na radiografia aumenta a razão de chances de envolvimento nodal entre (4.246; 82.785) vezes. Por fim, o estágio do tumor também aumenta a razão de

Tabela 9: Estimativas pontuais e intervalos de confiança para a razão de chances do modelo completo.

[H]	Variável	Estimativa Pontual	2.5 %	97.5 %	P-valor
	Resultado da Radiografia	15.5278	4.2459	82.7852	0.0002
	Estágio do Tumor	24.7386	6.6540	137.7293	0.0000
	Nível de Fosfatase Ácida	1.0282	1.0106	1.0480	0.0021

Tabela 10: Estatísticas de comparação entre os modelos de regressão logística completo e com intercepto.

[H]	Teste	Estatística	GL	P-valor
	Razao de Verossimilhanca	52.01516	3	0
	Score	52.01516	3	0

chances de envolvimento nodal entre (6.654; 137.729) vezes.

3.3.1 Comparação dos modelos

A Tabela 10 apresenta as estatísticas de comparação entre o modelo completo e o modelo com intercepto.

3.4 Demais modelos possíveis

Também é possível avaliar as demais combinações de modelos, cujas estatísticas de qualidade de ajuste estão na Tabela 11.

O método *stepwise* pode ser utilizado para selecionar o melhor modelo. Nesse caso, considerando tal método e os dados apresentados de qualidade de ajustes dos modelos, a escolha indicada foi exatamente a do modelo completo.

3.5 Análise a partir do modelo escolhido

Para avaliar a qualidade do modelo, pode-se utilizar a curva ROC. A Figura 6 apresenta a curva ROC do modelo escolhido.

Tabela 11: Estatísticas de qualidade dos modelos de regressão logística com as variáveis preditoras possíveis.

[H]

	Modelo	Deviance	AIC	BIC	Log Likelihood
1	$X4 \sim 1$	138.21	140.21	142.83	-69.10
3	$X4 \sim X1 + X2 + X3$	86.19	94.19	104.69	-43.10
4	$X4 \sim X1 + X2$	96.08	102.08	109.96	-48.04
5	$X4 \sim X1 + X3$	116.54	122.54	130.41	-58.27
6	$X4 \sim X2 + X3$	106.64	112.64	120.52	-53.32
8	$X4 \sim X1$	119.88	123.88	129.13	-59.94
7	$X4 \sim X2$	116.03	120.03	125.28	-58.01
2	$X4 \sim X3$	134.45	138.45	143.70	-67.22

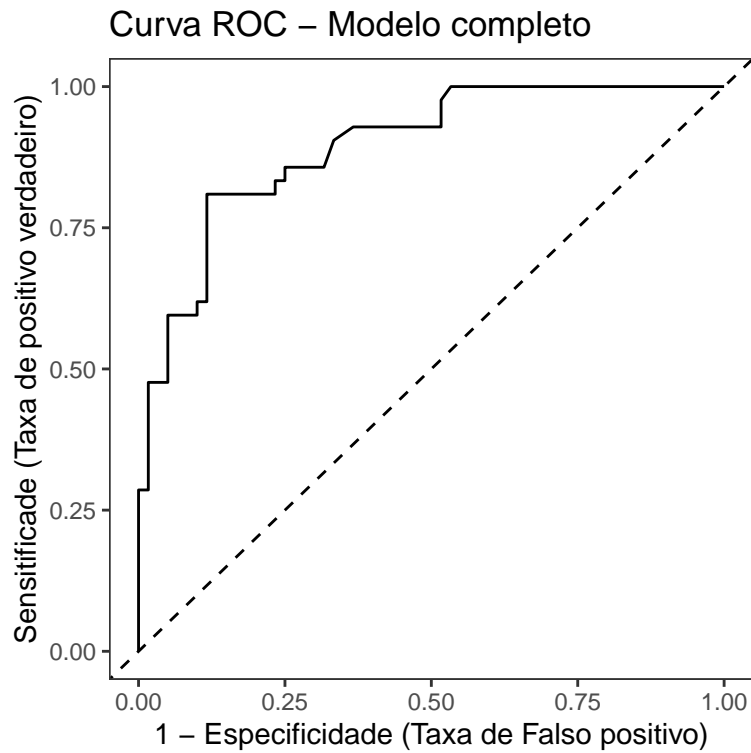


Figura 6: Curva ROC do modelo de regressão logística escolhido.

Tabela 12: Matriz de confusão do modelo de regressão logística aplicado ao conjunto de teste.

[H]	Observado\Predito	
	0	1
0	54	10
1	10	28

3.6 Aplicação do modelo escolhido ao conjunto de teste

O modelo escolhido é aplicado ao conjunto de teste. A Tabela 12 apresenta a matriz de confusão do modelo aplicado ao conjunto de teste. O valor de corte definido foi de 0.5.

Para esse caso específico, o modelo teve 80.39% de acertos, o que confere um caráter interessante para previsão dos resultados.

4 Conclusão

Como os dados levantados pela pesquisa conseguem fornecer informações importantes para o diagnóstico do espalhamento do câncer de próstata para os linfonodos, pode-se sugerir a coleta dessas informações para novos pacientes de modo a auxiliar no diagnóstico de maneira mais direta, rápida e menos invasiva. O modelo escolhido adota a estrutura de uma regressão logística. Com isso, o modelo estimado apresentou uma taxa de acertos de 80.39%, o que confere um caráter interessante para previsão dos resultados e bons indicativos diagnósticos iniciais.

De maneira específica, o modelo de regressão logística escolhido foi o modelo completo, que apresentou uma boa qualidade de ajuste, como as métricas apresentadas no trabalho indicaram. A curva ROC também sugere que o modelo é capaz de distinguir entre os grupos de interesse. Além disso, de maneira individual, as três variáveis preditoras apresentaram um bom poder de auxílio na previsão das razões de chance.

No que diz respeito aos dados coletados, a variável “Nível de Fosfatase Ácida” foi a que apresentou o menor poder, sendo ligeiramente significativa. Ainda assim, como sua escala de variação é considerável, à medida que o nível se eleva, pode-se considerar um aumento na chance de apresentar envolvimento nodal. Já o resultado das variáveis de resultado positivo da radiografia e de estágio grave do tumor apresentam indicações elevadas sobre o aumento da razão de chances de um paciente apresentar envolvimento nodal.

Por fim, para estudos futuros, seria interessante aumentar o tamanho da amostra coletada. Isso poderia fornecer mais informações para a análise e, possivelmente, melhorar a qualidade e precisão do modelo. Além disso, a inclusão de novas variáveis preditoras reduziria a chance de efeitos estimados serem influenciados por variáveis omitidas.

5 Referências

- Agresti, Alan. “An introduction to categorical data analysis.”. Third Edition John Wiley & Sons, Inc., 2019.
- ARTES, Rinaldo e BARROSO, Lucia Pereira. Métodos multivariados de análise estatística. . São Paulo: Blucher. . Acesso em: 14 jul. 2024. , 2023.
- Cordeiro, Gauss M.; Demétrio, Clarice G. B.; Moral, Rafael A. Modelos lineares generalizados e aplicações. Blucher. São paulo, 2024.
- George Casella, Roger L Berger. Inferência estatística. [tradução Solange Aparecida Visconde]. - São Paulo: Cengage Learning, 2022.
- Giolo, Suely Ruiz. Introdução à análise de dados categóricos com aplicações. Blucher. São paulo, 2017.
- Morettin, Pedro Alberto. Singer, Julio da Motta. Estatística e ciência de dados. 1ª ed. LTC. Rio de Janeiro, 2023.

6 Apêndice A - códigos em R

A seguir, os principais códigos utilizados neste estudo, além de alguma análises exploratórias adicionais.

```

# Pacotes ----
if (!require("pacman")) install.packages("pacman")
pacman::p_load(
  readxl,
  tidyverse,
  cowplot,
  mdscore,
  AICcmodavg,
  questionr,
  mlpack,
  ResourceSelection,
  lmtree,
  car,
  stats,
  knitr,
  pROC,
  ROCit,
  labelled,
  compareGroups,
  arm,
  performance
)

# Dados de análise e treino dos modelos ----
df <- read_excel("arquivos/Amostra_g06_Bruno_Rafael.xlsx")
colnames(df) <- c("ID", # Identificação do paciente
  "resultado_radiografia", # 0 = negativo | 1 - positivo
  "estagio_tumor", # 0 - menos grave | 1 - mais grave
  "nivel_fosfatase_acida", # x100
  "envolvimento_nodal" # 0 - não | 1 - sim
)

var_label(df) <- list(
  resultado_radiografia = "Resultado da Radiografia",
  estagio_tumor = "Estágio do Tumor",
  nivel_fosfatase_acida = "Nível da Fosfatase Ácida",
  envolvimento_nodal = "Envolvimento Nodal (x100)"
)

# Dados de validação do modelo (teste) ----
teste <- read_excel("arquivos/Amostra_VALIDACAO.xlsx")
colnames(teste) <- c("ID", # Identificação do paciente
  "resultado_radiografia", # 0 = negativo | 1 - positivo
  "estagio_tumor", # 0 - menos grave | 1 - mais grave
  "nivel_fosfatase_acida", # x100
  "envolvimento_nodal" # 0 - não | 1 - sim
)
#####
# Parte 1) - Análise exploratória ----

dados = df %>%
  mutate(resultado_radiografia = case_when(
    resultado_radiografia == "0" ~ "Negativo",
    resultado_radiografia == "1" ~ "Positivo"
  ), estagio_tumor = case_when(
    estagio_tumor == "0" ~ "- grave",
    estagio_tumor == "1" ~ "+ grave"
  ), envolvimento_nodal = case_when(
    envolvimento_nodal == "0" ~ "Não",
    envolvimento_nodal == "1" ~ "Sim"
  ), nivel_fosfatase_acida = nivel_fosfatase_acida)

dados = dados[,2:5]
dados$envolvimento_nodal = factor(dados$envolvimento_nodal)
dados$estagio_tumor = factor(dados$estagio_tumor)
dados$resultado_radiografia = factor(dados$resultado_radiografia)

compareGroups(envolvimento_nodal ~ . ,

```

```

        data=dados,
        method = c(nivel_fosfatase_acida=NA), alpha= 0.05 # Testando a
normalidade para decidir o teste
    )

tabela = compareGroups(envolvimento_nodal ~ . ,
        data=dados,
        method = c(nivel_fosfatase_acida=NA), alpha= 0.05 # Testando a
normalidade para decidir o teste
    )

createTable(tabela, show.ratio=TRUE)

descrTable(dados)

plot(tabela)
plot(tabela, bivar=TRUE)

# rm(dados,tabela)
#####
# Parte 2) Regressão logística: nivel_fosfatase_acida EXPLICANDO envolvimento_nodal ---
-

# Somente intercepto
fit0 <- glm(envolvimento_nodal ~ 1,
            family=binomial(link=logit),
            data=df)

plot(
  jitter(envolvimento_nodal, 0.01) ~ nivel_fosfatase_acida,
  xlab = "nivel_fosfatase_acida",
  ylab = "envolvimento_nodal",
  data = df,
  pch = 16
)

curve(
  predict(fit0, data.frame(nivel_fosfatase_acida = x), type = "resp"),
  add = T,
  col = "blue",
  lwd = 2
)

get_confint <- function(fit, variavel, digitos){
  if(!is.null(dim(confint(fit)))){
    round(exp(confint(fit)[variavel, ]), digitos)
  } else {
    round(confint(fit), digitos) |>
      str_c( collapse = "; ")
  }
}

fit0_pred <- predict(fit0, type="response", se.fit=TRUE)
fit0_fit <- fit0_pred$fit[1]
fit0_se <- fit0_pred$se.fit[1]

# Intervalo de confiança para a probabilidade
Lb_fit0 <- fit0_fit - qnorm(0.975) * fit0_se
Ub_fit0 <- fit0_fit + qnorm(0.975) * fit0_se

```

```

medidas0 <- as.data.frame(cbind(fit0$deviance, fit0$aic, BIC(fit0),
                                logLik(fit0)[1]))
colnames(medidas0) <- c("Deviance", "AIC", "BIC", "Log Likelihood")
medidas0
# fit1: somente nivel_fosfatase_acida

fit1 <- glm(envolvimento_nodal ~ nivel_fosfatase_acida,
            family=binomial(link=logit),
            data=df)

plot(jitter(envolvimento_nodal, 0.01) ~ nivel_fosfatase_acida,
     xlab="nivel_fosfatase_acida",
     ylab="envolvimento_nodal",
     data=df, pch=16)

curve(predict(fit1, data.frame(nivel_fosfatase_acida=x), type="resp"),
       add=T, col="blue", lwd=2)

summary(fit1); confint(fit1)

hoslem_fit1 <- ResourceSelection::hoslem.test(fit1$y, fitted(fit1), g = 10)

# DescTools::HosmerLemeshowTest(fitted(fit1), fit1$y)
# H_0) Valores observados e valores esperados são iguais para diferentes níveis de
# nivel_fosfatase_acida
# H_1) c.c.
# O teste de H-L rejeita a hipótese nula, portanto indica que o modelo não é adequado.

# Analisando a influência de cada valor no resultado da regressão logística
stats::influence.measures(fit1)

# teste verossimilhança
t1 <- lr.test(fit0, fit1)
# O teste de razão de verossimilhança não rejeita (por pouco) a hipótese nula, portanto
# o modelo fit0 é preferível

# teste de wald

thetahat <- fit1$coefficients
vcov1 <- vcov(fit1)
LL <- rbind(c(0, 1))

WaldTest = function(L, thetahat, Vn, h=0) {
  WaldTest = numeric(3)
  names(WaldTest) = c("W", "df", "p-value")
  r = dim(L)[1]
  W = t(L%*%thetahat-h) %*% solve(L%*%Vn%*%t(L)) %*%
    (L%*%thetahat-h)
  W = as.numeric(W)
  pval = 1-pchisq(W, r)
  WaldTest[1] = W; WaldTest[2] = r; WaldTest[3] = pval
  WaldTest}

# teste score
score1 <- anova(fit0, fit1, test="Rao")
resultados <- cbind(t1$LR, 1, t1$pvalue)
resultados2 <- matrix(WaldTest(LL, thetahat, vcov1), ncol=3)
resultados3 <- cbind(anova(fit0, fit1, test="Rao")[2, 4],
                    anova(fit0, fit1, test="Rao")[2, 3],
                    anova(fit0, fit1, test="Rao")[2, 6])

testes1 <- rbind(resultados, resultados2, resultados3)
rownames(testes1) <- c("Razao de Verossimilhança", "Wald", "Score")
colnames(testes1) <- c("Estatística", "GL", "P-valor")

```

```

testes1
# Vemos que nenhum dos 3 testes rejeitaram a hipótese nula a 5%. Isto é, o modelo não é
significativamente melhor que o modelo que contém apenas o intercepto.

coef1 <- summary(fit1)$coefficients

colnames(coef1) <- c("Estimativa", "Erro Padrao", "Valor Z", "Pr(>|z|)")
rownames(coef1) <- c("Intercepto", "Nível de \n fosfatase ácida")

coef1

odds.ratio(fit1)[2,]
# A Odds-ratio mostra que para cada 1 (x100) unidades do nível de fosfatase ácida,
aumenta em 1,41% a chance de não haver envolvimento nodal.
# Entretanto, o intervalo contém o valor 1, o que indica que é pouco significativo este
resultado, e pode não dizer nada na realidade.

# Medidas de qualidade de ajuste

medidas1 <- as.data.frame(cbind(fit1$deviance, fit1$aic, BIC(fit1),
                                logLik(fit1)[1]))
colnames(medidas1) <- c("Deviance", "AIC", "BIC", "Log Likelihood")
medidas1

# Intervalo de confiança para a odds ratio
fit1_ci <- get_confint(fit1 = fit1,
                      variavel = "nivel_fosfatase_acida",
                      digitos = 5)

# Plotando com I.C. 95%
temp.data <- data.frame(df$nivel_fosfatase_acida)
colnames(temp.data) = "nivel_fosfatase_acida"
predicted.data <- as.data.frame(predict(fit1, newdata = temp.data,
                                         type="link", se=TRUE))

new.data <- cbind(temp.data, predicted.data)

new.data$yci <- fit1$family$linkinv(new.data$fit - qnorm(0.975) * new.data$se)
new.data$yys <- fit1$family$linkinv(new.data$fit + qnorm(0.975) * new.data$se)
new.data$fit <- fit1$family$linkinv(new.data$fit)

p <- ggplot(df, aes(x=nivel_fosfatase_acida, y=envolvimento_nodal))
p + geom_point(na.rm = T) +
  geom_ribbon(data=new.data, aes(y=fit, ymin=yci, ymax=yys),
            fill="lightblue", alpha=0.3) +
  geom_line(data=new.data, aes(y=fit)) +
  labs(x="Nível fosfatase ácida (x100)", y="Probabilidade de envolvimento nodal")

#####
# Parte 3) Adicionando outras variáveis no modelo ----

# Modelo saturado:
fit2 <- glm(envolvimento_nodal ~ resultado_radiografia +
            + estagio_tumor +
            nivel_fosfatase_acida,
            family=binomial(link=logit),
            data=df)

summary(fit2); confint(fit2)
# Com o modelo saturado, todas as variáveis se tornam significativas a 5%, inclusive o
nível de fosfatase ácida.

hoslem_fit2 <- ResourceSelection::hoslem.test(fit2$y, fit2$fitted.values)
# H_0) Valores observados e valores esperados são iguais para diferentes níveis de
nivel_fosfatase_acida e outras variáveis.
# H_1) c.c.
# O teste de H-L rejeita a hipótese nula a 5%, portanto indica que o modelo não é
adequado.

```

```

# Outra opção de teste
# DescTools::HosmerLemeshowTest(fit = fitted(fit2),
#                               obs = df$envolvimento_nodal)

# Analisando a influência de cada valor no resultado da regressão logística (análise de
resíduos)
stats::influence.measures(fit2)

# teste verossimilhança
t2 <- lr.test(fit0,fit2)
# O teste de razão de verossimilhança rejeita a hipótese nula, portanto o modelo
saturado é preferível

# teste score
score2<-anova(fit0,fit2, test="Rao")
resultados <- cbind(t2$LR,3,t2$pvalue)
resultados3 <- cbind(anova(fit0,fit2, test="Rao") [2,4],
                    anova(fit0,fit2, test="Rao") [2,3],
                    anova(fit0,fit2, test="Rao") [2,6])

testes2 <- rbind(resultados,resultados3)
rownames(testes2) <- c("Razao de Verossimilhanca","Score")
colnames(testes2) <- c("Estatistica","GL","P-valor")

testes2
# Vemos que ambos os testes rejeitaram a hipótese nula a 5%.
# Isto é, o modelo é significativamente melhor que o modelo que contém apenas o
intercepto.

coef2 <- summary(fit2)$coefficients

colnames(coef2) <- c("Estimativa","Erro Padrao","Valor Z","Pr(>|z|)")
rownames(coef2) <- c("Intercepto",
                    "Resultado da Radiografia",
                    "Estágio do Tumor",
                    "Nível de Fosfatase Ácida"
                    )

coef2

odds.ratio(fit2)[2,]
# A odds ratio mostra que o resultado positivo na radiografia aumenta de 4,246 a 82,785
a chance de haver envolvimento nodal.

odds.ratio(fit2)[3,]
# A odds ratio mostra que o fato do tumor ser mais grave aumenta de 6,654 a 137,73 a
chance de haver envolvimento nodal.

odds.ratio(fit2)[4,]
# A Odds-ratio mostra que para cada 1 (x100) unidades do nível de fosfatase ácida,
aumenta em 2,82% a chance de não haver envolvimento nodal.
# Diferente do modelo simples, o intervalo não contém o valor 1, indicano que é
positiva a relação dessas variáveis a >95% de significância

fit2_odds <-
odds.ratio(fit2) |>
  as_tibble() |>
  slice(-1) |>
  mutate(Variável = c("Resultado da Radiografia",
                      "Estágio do Tumor",
                      "Nível de Fosfatase Ácida")) |>
  dplyr::select(Variável, everything()) |>
  rename("Estimativa Pontual" = OR, "P-valor" = p)

```

```
# Essas duas variáveis são extremamente significativas, e contribuem fortemente para a explicação de haver ou não envolvimento nodal.
```

```
# Medidas de qualidade de ajuste
medidas2 <- as.data.frame(cbind(fit2$deviance, fit2$AIC, BIC(fit2),
                                logLik(fit2)[1]))
colnames(medidas2) <- c("Deviance", "AIC", "BIC", "Log Likelihood")
medidas2
```

```
#####
```

```
# Outros modelos possíveis:
```

```
fit3 <- glm(envolvimento_nodal ~ resultado_radiografia + estagio_tumor,
            family=binomial(link=logit),
            data=df)
```

```
medidas3 <- as.data.frame(cbind(fit3$deviance, fit3$AIC, BIC(fit3),
                                logLik(fit3)[1]))
colnames(medidas3) <- c("Deviance", "AIC", "BIC", "Log Likelihood")
```

```
fit4 <- glm(envolvimento_nodal ~ resultado_radiografia + nivel_fosfatase_acida,
            family=binomial(link=logit),
            data=df)
```

```
medidas4 <- as.data.frame(cbind(fit4$deviance, fit4$AIC, BIC(fit4),
                                logLik(fit4)[1]))
colnames(medidas4) <- c("Deviance", "AIC", "BIC", "Log Likelihood")
```

```
fit5 <- glm(envolvimento_nodal ~ nivel_fosfatase_acida + estagio_tumor,
            family=binomial(link=logit),
            data=df)
```

```
medidas5 <- as.data.frame(cbind(fit5$deviance, fit5$AIC, BIC(fit5),
                                logLik(fit5)[1]))
colnames(medidas5) <- c("Deviance", "AIC", "BIC", "Log Likelihood")
```

```
fit6 <- glm(envolvimento_nodal ~ estagio_tumor,
            family=binomial(link=logit),
            data=df)
```

```
medidas6 <- as.data.frame(cbind(fit6$deviance, fit6$AIC, BIC(fit6),
                                logLik(fit6)[1]))
colnames(medidas6) <- c("Deviance", "AIC", "BIC", "Log Likelihood")
```

```
fit7 <- glm(envolvimento_nodal ~ resultado_radiografia,
            family=binomial(link=logit),
            data=df)
```

```
medidas7 <- as.data.frame(cbind(fit7$deviance, fit7$AIC, BIC(fit7),
                                logLik(fit7)[1]))
colnames(medidas7) <- c("Deviance", "AIC", "BIC", "Log Likelihood")
```

```
#####
```

```
# Comparando os modelos
```

```
fit = glm(
  envolvimento_nodal ~ nivel_fosfatase_acida + resultado_radiografia + estagio_tumor,
  family = binomial(link = logit),
  data = df
)
```

```
car::Anova(fit) # A iteração não é significativa.
stepAIC(fit, direction = c("both")) # Voltamos ao modelo saturado.
rm(fit)
```

```
stepwise <- stepAIC(fit2, direction = c("both"))
# O método stepwise indica ficar com o modelo saturado.
```

```

Modelo = c("X4 ~ 1", "X4 ~ X3", "X4 ~ X1 + X2 + X3", "X4 ~ X1 + X2", "X4 ~ X1 + X3", "X4 ~
X2 + X3", "X4 ~ X2", "X4 ~ X1")
medidas =
rbind(medidas0,medidas1,medidas2,medidas3,medidas4,medidas5,medidas6,medidas7)
medidas$Modelo = Modelo
medidas = medidas[c(1,3,4:6,8,7,2),c(5,1:4)]

kable(medidas)
# Pela tabela, devemos optar pelo modelo saturado, pois:
# 1) tem menor deviance 2) tem menor AIC 3) tem maior BIC 4) tem maior log-
verossimilhança.

# rm(fit0,fit1,fit3,fit4,fit5,fit6,fit7,
#   medidas0,medidas1,medidas3,medidas4,medidas5,medidas6,medidas7,
#   resultados,resultados2,resultados3,p,predicted.data,vcov1,Modelo,thetahat,
#   WaldTest,LL,new.data,score1,t1,testes1,temp.data,coef1,t2)
#####
# Métricas do modelo escolhido ----
kable(medidas2)

kable(coef2)
# Curva ROC do modelo fit2
ROC <- roc(response = df$envolvimento_nodal, predictor = predict(fit2, type =
"response"))

roc_data <- data.frame(
  Spec_comp = 1 - ROC$specificities,
  Sensit = ROC$sensitivities
)

plot(ROC, main = "Curva ROC - Modelo completo")

# Alternativo:
ROC2 <- rocit(score=predict(fit2, type = "response"),class=df$envolvimento_nodal)
plot(ROC2)
ksplot(ROC2)

performance_hosmer(fit2, n_bins = 10)

# Aplicando o modelo aos dados de teste:
teste$predict <- predict(fit2, teste,
                        type = "response")

# Usando um threshold = 0.5...
teste = teste %>%
  mutate(env_pred = ifelse(predict > .5,1,0))

# Matriz de confusão:
table_confusao <- table(teste$envolvimento_nodal, teste$env_pred)
kable(table_confusao)

acertos <- sum(diag(table_confusao)) / sum(table_confusao)

matriz_conf_alternativa <- table(teste$envolvimento_nodal, teste$env_pred) |>
  as.data.frame() |>
  pivot_wider(values_from = Freq,
              names_from = Var2) |>
  rename("Observado\\Predito" = Var1)

# Para os dados de validação, o modelo teve 54+28/102 = 80,4% de acerto, o que é
bastante razoável.

# Análise de resíduos
plot(residuals(fit2, "pearson"))

```



```

# Outra visualização: Binned residual plot
binnedplot(fitted(fit2),
            residuals(fit2, type = "response"),
            nclass = NULL,
            xlab = "Valores esperados",
            ylab = "Resíduo médio",
            main = "Gráfico residual armazenado",
            cex.pts = 0.8,
            col.pts = 1,
            col.int = "gray")

#Deveríamos observar 95% das observações sobre o intervalo produzido pelo ASE (linhas
cinzas)

y <- rbinom(102, 1, 0.5)
f1 <- fitted(fit2)
plot( residuals(fit2, "pearson"), (y-f1)/sqrt(f1*(1-f1)))
abline(0,1)

# Worm plot (está se popularizando para análise de MLGs)
p_load(gamlss)
wp(resid=resid(fit2), xvar=df$envolvimento_nodal)

# Pelo worm plot, o ajuste do modelo parece muito ruim.
# Mas isso aparenta ter mais a ver com a técnica (regressão logística) do que quanto ao
modelo
# específico ajustado.

```

7 Apêndice B - códigos em SAS

Também foram realizadas análises em SAS. A seguir, os principais códigos utilizados neste estudo.

7.1 Códigos

```
OPTIONS LS=80 PS=60 NODATE;
/* Importando dados da amostra */
libname trabalho '/home/u36587463/dados';
FILENAME REFFILE '/home/u36587463/dados/Amostra_g06_Bruno_Rafael.xlsx';

PROC IMPORT DATAFILE=REFFILE
    DBMS=XLSX
    OUT=trabalho.df REPLACE;
    GETNAMES=YES;
RUN;

/* Decodificando os dados*/
data trabalho.df;
    set trabalho.df;
/*    X3 = X3/100;*/
    Label ID='Nº do prontuário'
           X1='Resultado da radiografia'
           X2='Estágio do tumor'
           X3='Nível de fosfatase ácida'
           X4='Envolvimento nodal';
run;

/* Carregando os dados de validação do modelo */
FILENAME valid '/home/u36587463/dados/Amostra_VALIDACAO.xlsx';
PROC IMPORT DATAFILE=valid
    DBMS=XLSX
    OUT=trabalho.valid REPLACE;
    GETNAMES=YES;
RUN;

data trabalho.valid;
    set trabalho.valid;
/*    X3 = X3/100;*/
```

```

    Label ID='Nº do prontuário'
        X1='Resultado da radiografia'
        X2='Estágio do tumor'
        X3='Nível de fosfatase ácida'
        X4='Envolvimento nodal';
run;

/* Conferindo dados */
PROC CONTENTS
    DATA=trabalho.df;
RUN;

/* Criando rótulos para os valores das variáveis a que se referem */
proc format;
    value x1f 0='negativo'
              1='positivo';
    value x2f 0='menos grave'
              1='mais grave';
    value x4f 1='Sim'
              0='Não';
run;

/* Análise exploratória */
proc freq data=trabalho.df;
    tables X3 X4;
    format X4 x4f.;
run;

proc univariate data=trabalho.df plot;
    class X4;
    var X3;
    format X4 x4f.;
run;

proc means data=trabalho.df;
    class X4;
    var X3;
    format X4 x4f.;
run;

proc gplot data=trabalho.df;
    plot X4*X3;

```

```

run;

proc sgplot data=trabalho.df;
vbox X3/category=X4;
format X4 x4f.;
run;

proc tabulate data=trabalho.df;
class X4;
var X3;
table X4, X3*(mean std);
format X4 x4f.;
run;

/* Regressão binária logística: X3 (Nível de fosfatase ácida) explica X4 (Envolvimento nodal) */
proc logistic data=trabalho.df;
model X4 (event='1')= X3 /covb lackfit;
run;

/* Regressão binária logística: X1 (Resultado da radiografia),
X2 (Estágio do tumor),
X3 (Nível de fosfatase ácida)
EXPLICAM
X4 (Envolvimento nodal) */

proc logistic data=trabalho.df;
model X4 (event='Sim')= X1 X2 X3 /covb lackfit;
format X1 x1f. X2 x2f. X4 x4f.;
run;

/* Outros modelos possíveis */

proc logistic data=trabalho.df;
model X4 (event='Sim')= X1 /covb;
format X1 x1f. X4 x4f.;
run;

proc logistic data=trabalho.df;
model X4 (event='Sim')= X2 /covb;
format X2 x2f. X4 x4f.;
run;

```

```

proc logistic data=trabalho.df;
    model X4 (event='Sim')= X1 X3 /covb lackfit;
    format X1 x1f. X4 x4f.;
run;

proc logistic data=trabalho.df;
    model X4 (event='Sim')= X2 X3 /covb lackfit;
    format X2 x2f. X4 x4f.;
run;

proc logistic data=trabalho.df;
    model X4 (event='Sim')= X1 X2 /covb lackfit;
    format X1 x1f. X2 x2f. X4 x4f.;
run;

/* Olhando interações */
proc logistic data=trabalho.df;
    model X4 (event='Sim')= X1 X2 X3 X1*X2 X2*X3 X1*X3 X1*X2*X3 /covb lackfit;
    format X1 x1f. X2 x2f. X4 x4f.;
run;

proc logistic data=trabalho.df;
    model X4 (event='Sim')= X1 X2 X3 X1*X2 X2*X3 X1*X3 /covb lackfit;
    format X1 x1f. X2 x2f. X4 x4f.;
run;

proc logistic data=trabalho.df;
    model X4 (event='Sim')= X1 X2 X3 X1*X2 X2*X3 /covb lackfit;
    format X1 x1f. X2 x2f. X4 x4f.;
run;

proc logistic data=trabalho.df;
    model X4 (event='Sim')= X1 X2 X3 X1*X2 X1*X3/covb lackfit;
    format X1 x1f. X2 x2f. X4 x4f.;
run;

proc logistic data=trabalho.df;
    model X4 (event='Sim')= X1 X2 X3 X2*X3 X1*X3/covb lackfit;
    format X1 x1f. X2 x2f. X4 x4f.;
run;

proc logistic data=trabalho.df;

```

```

    model X4 (event='Sim')= X1 X2 X3 X1*X2 X2*X3 X1*X3/covb lackfit;
    format X1 x1f. X2 x2f. X4 x4f.;
run;

/* As interações nunca aparecem significativas quando a variável está presente no modelo.*/
/* Portanto, não utilizaremos. */
/* De todos os modelos testados, os que melhor se ajustam aparentam ser: */

/* Regressão binária logística: X1 (Resultado da radiografia),
                                X2 (Estágio do tumor),
                                X3 (Nível de fosfatase ácida)
                                EXPLICAM
                                X4 (Envolvimento nodal) */

proc logistic data=trabalho.df;
    model X4 (event='Sim')= X1 X2 X3 /covb lackfit;
    format X1 x1f. X2 x2f. X4 x4f.;
run;

/* Regressão binária logística: X1 (Resultado da radiografia),
                                X2 (Estágio do tumor),
                                EXPLICAM
                                X4 (Envolvimento nodal) */

proc logistic data=trabalho.df;
    model X4 (event='Sim')= X1 X2 /covb lackfit;
    format X1 x1f. X2 x2f. X4 x4f.;
run;

/* Utilizando abordagem Stepwise para escolha de modelo */
proc logistic data=trabalho.df;
    model X4 (event='Sim')= X1 X2 X3 X1*X2 X2*X3 X1*X3 X1*X2*X3
        / selection=stepwise
        slentry=0.3
        slstay=0.35
        details
        lackfit;
    format X1 x1f. X2 x2f. X4 x4f.;
run;

/* O método Stepwise sugere pelo modelo contendo todas as variáveis,
além de interação entre X1 e X3 */

```

```

/* Esta interação não é muito significativa, e o modelo sem interação
contendo todas as três funcionou muito bem como visto anteriormente */
/* Portando, ficaremos com o modelo mais parsimonioso, isto é:
X4 ~ X1 X2 X3 */

proc logistic data=trabalho.df outest=betas covout;
    model X4 (event='Sim')= X1 X2 X3 /covb lackfit;
    output out=pred p=phat lower=lcl upper=ucl
        predprob=(individual crossvalidate);
    ods output Association=Association;
    format X1 x1f. X2 x2f. X4 x4f.;
run;

/* Parâmetros do modelo */
proc print data=betas;
run;

/* Previsões do modelo sob os dados, com intervalo de confiança 95% */
proc print data=pred;
run;

/* Curva ROC do modelo */
proc logistic data=trabalho.df plots(only)=roc;
model X4 (event='Sim')= X1 X2 X3;
ods output roccurve=ROCdata;
format X1 x1f. X2 x2f. X4 x4f.;
run;

/* Testando o modelo nos dados de validação */
/* Salvando o modelo */
proc logistic data=trabalho.df outmodel=fit;
    model X4 (event='Sim')= X1 X2 X3 /covb lackfit;
    format X1 x1f. X2 x2f. X4 x4f.;
run;

/* Aplicando modelo sob dados de validação */
proc logistic inmodel=fit;
    score data=trabalho.valid out=validacao_resultado;
run;

/* Verificando resultado */
proc print data=validacao_resultado;

```

```
run;

/* "Matriz de confusão" */
proc freq data=validacao_resultado;
  tables X4*I_X4 / nopercnt norow nocol ;
  format X4 x4f.;
run;
```

7.2 Saídas SAS

The CONTENTS Procedure

Data Set Name	TRABALHO.DF	Observations	102
Member Type	DATA	Variables	5
Engine	V9	Indexes	0
Created	13/07/2024 11:18:51	Observation Length	40
Last Modified	13/07/2024 11:18:51	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	3265
Obs in First Data Page	102
Number of Data Set Repairs	0
Filename	/home/u36587463/dados/df.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	23630208228
Access Permission	rw-r--r--
Owner Name	u36587463
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
1	ID	Num	8	BEST.	Nº do prontuário
2	X1	Num	8	BEST.	Resultado da radiografia
3	X2	Num	8	BEST.	Estágio do tumor
4	X3	Num	8	BEST.	Nível de fosfatase ácida

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
5	X4	Num	8	BEST.	Envolvimento nodal

The FREQ Procedure

Nível de fosfatase ácida				
X3	Frequency	Percent	Cumulative Frequency	Cumulative Percent
26	1	0.98	1	0.98
27	1	0.98	2	1.96
40	2	1.96	4	3.92
45	2	1.96	6	5.88
46	1	0.98	7	6.86
47	1	0.98	8	7.84
48	5	4.90	13	12.75
49	7	6.86	20	19.61
50	8	7.84	28	27.45
51	7	6.86	35	34.31
52	2	1.96	37	36.27
53	2	1.96	39	38.24
54	1	0.98	40	39.22
55	5	4.90	45	44.12
56	2	1.96	47	46.08
59	2	1.96	49	48.04
60	1	0.98	50	49.02
61	1	0.98	51	50.00
62	1	0.98	52	50.98
63	1	0.98	53	51.96
64	1	0.98	54	52.94
66	1	0.98	55	53.92
67	6	5.88	61	59.80
68	1	0.98	62	60.78
70	3	2.94	65	63.73
71	1	0.98	66	64.71
72	2	1.96	68	66.67

Nível de fosfatase ácida				
X3	Frequency	Percent	Cumulative Frequency	Cumulative Percent
75	1	0.98	69	67.65
76	3	2.94	72	70.59
77	1	0.98	73	71.57
78	3	2.94	76	74.51
79	2	1.96	78	76.47
81	1	0.98	79	77.45
82	4	3.92	83	81.37
83	1	0.98	84	82.35
84	3	2.94	87	85.29
85	1	0.98	88	86.27
89	2	1.96	90	88.24
98	2	1.96	92	90.20
99	2	1.96	94	92.16
102	2	1.96	96	94.12
136	2	1.96	98	96.08
137	1	0.98	99	97.06
186	1	0.98	100	98.04
187	2	1.96	102	100.00

Envolvimento nodal				
X4	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Não	60	58.82	60	58.82
Sim	42	41.18	102	100.00

The UNIVARIATE Procedure
Variable: X3 (Nível de fosfatase ácida)
X4 = Não

Moments			
N	60	Sum Weights	60
Mean	64.416667	Sum Observations	3865
Std Deviation	31.5709004	Variance	996.721751

Moments			
Skewness	3.07158322	Kurtosis	9.64960429
Uncorrected SS	307777	Corrected SS	58806.5833
Coeff Variation	49.0104534	Std Error Mean	4.07578571

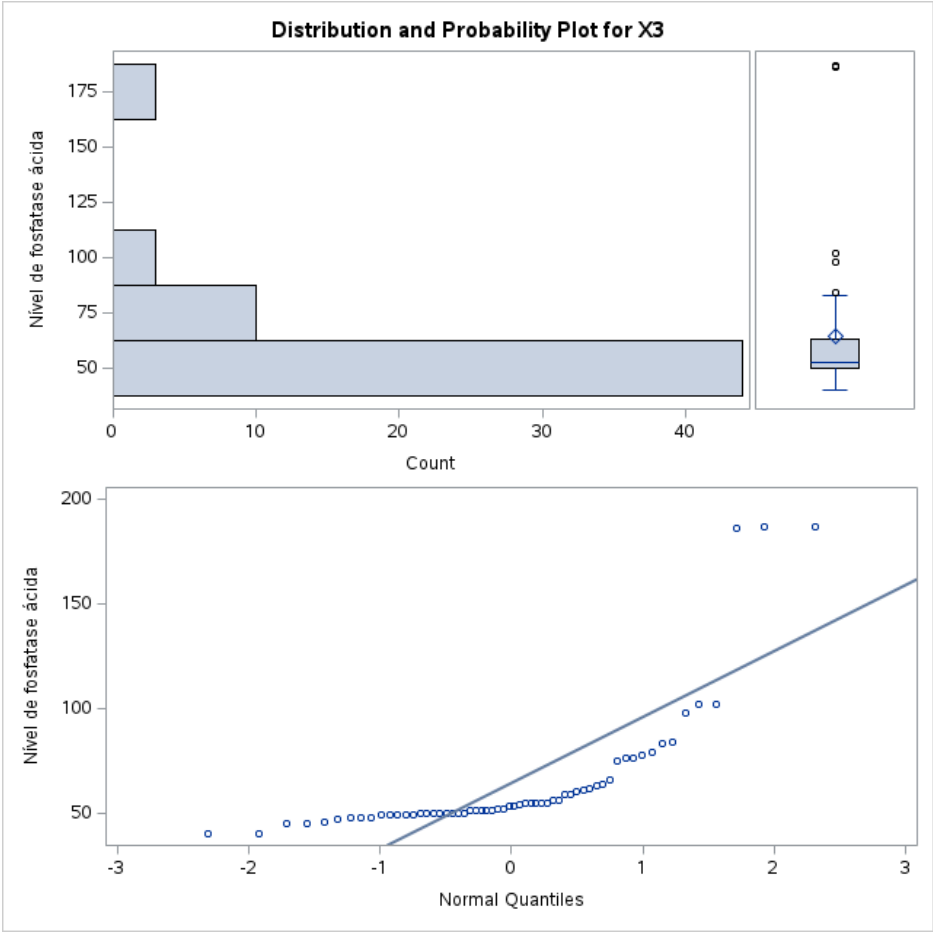
Basic Statistical Measures			
Location		Variability	
Mean	64.41667	Std Deviation	31.57090
Median	53.00000	Variance	996.72175
Mode	50.00000	Range	147.00000
		Interquartile Range	13.50000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	15.80472	Pr > t	<.0001
Sign	M	30	Pr >= M	<.0001
Signed Rank	S	915	Pr >= S	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	187.0
99%	187.0
95%	144.0
90%	91.0
75% Q3	63.5
50% Median	53.0
25% Q1	50.0
10%	47.5
5%	45.0
1%	40.0
0% Min	40.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
40	79	102	28
40	20	102	86

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
45	102	186	54
45	42	187	18
46	6	187	74



The UNIVARIATE Procedure
Variable: X3 (Nivel de fosfatase ácida)
X4 = Sim

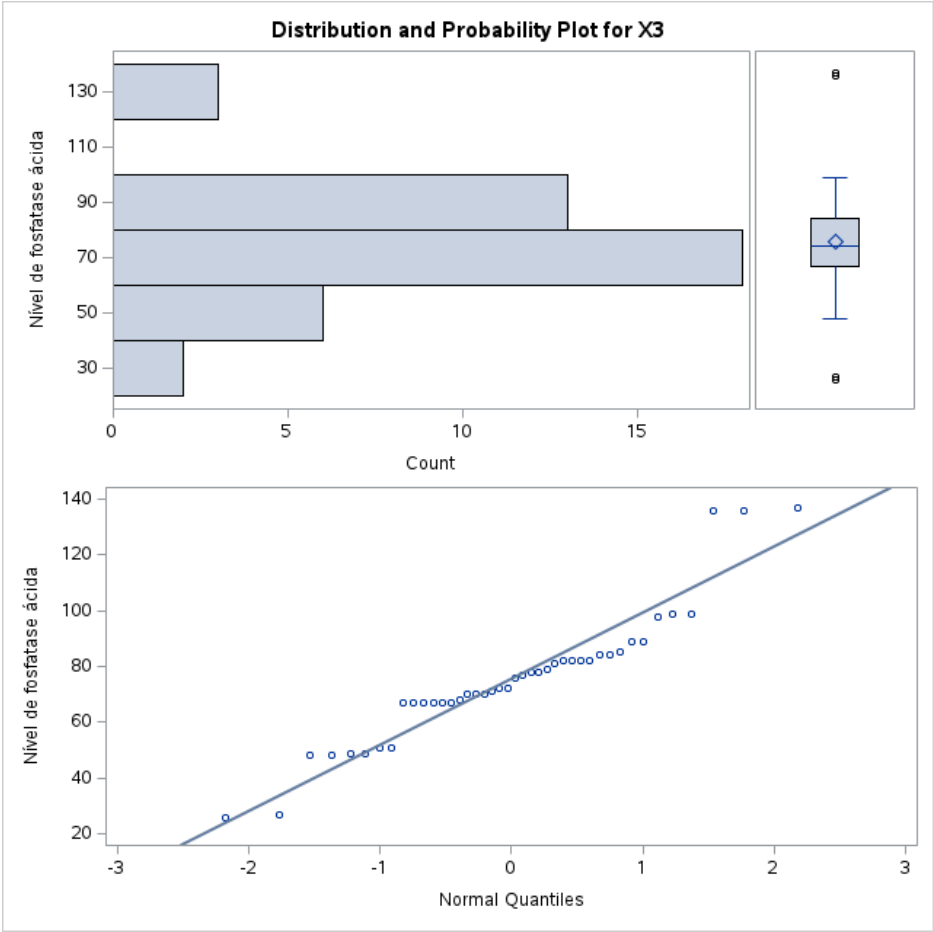
Moments			
N	42	Sum Weights	42
Mean	75.6428571	Sum Observations	3177
Std Deviation	23.7279514	Variance	563.015679
Skewness	0.69572446	Kurtosis	1.7801564
Uncorrected SS	263401	Corrected SS	23083.6429
Coeff Variation	31.3683966	Std Error Mean	3.66130239

Basic Statistical Measures			
Location		Variability	
Mean	75.64286	Std Deviation	23.72795
Median	74.00000	Variance	563.01568
Mode	67.00000	Range	111.00000
		Interquartile Range	17.00000

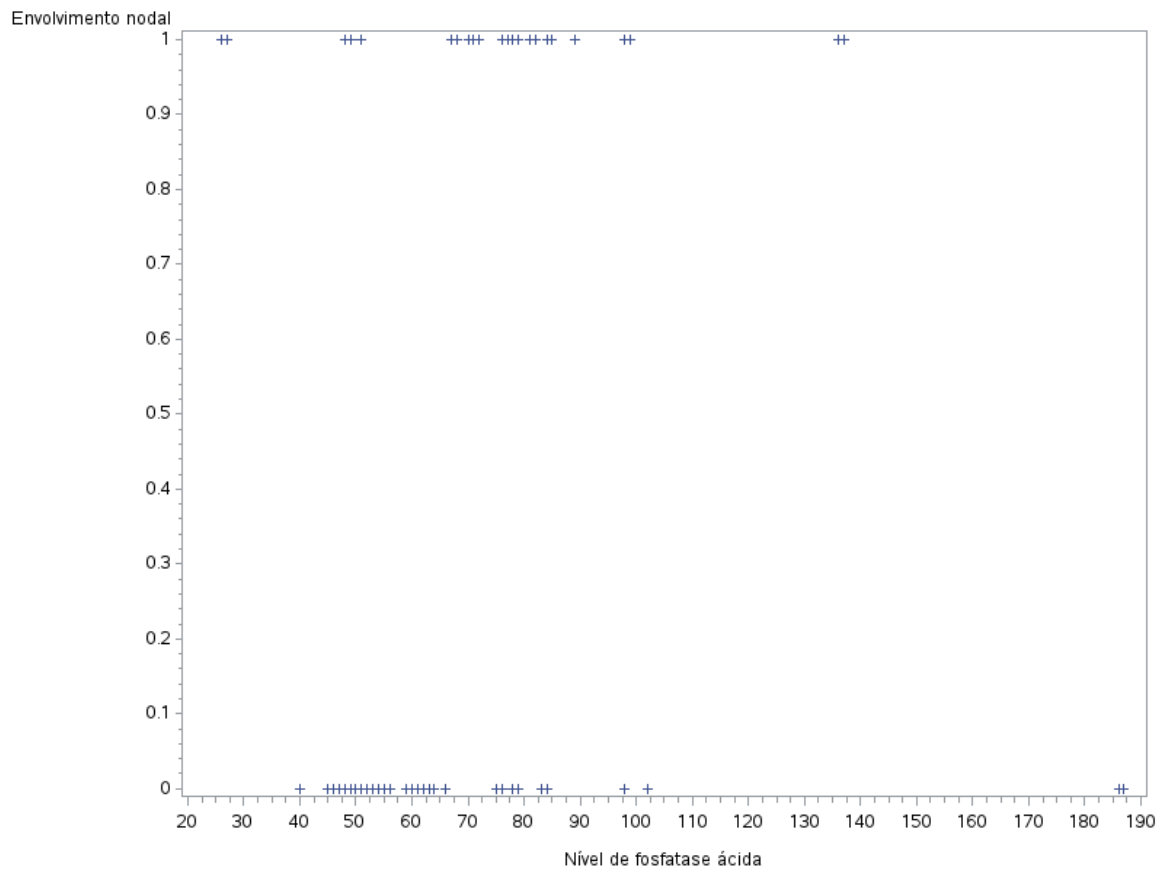
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	20.6601	Pr > t 	<.0001
Sign	M	21	Pr >= M 	<.0001
Signed Rank	S	451.5	Pr >= S 	<.0001

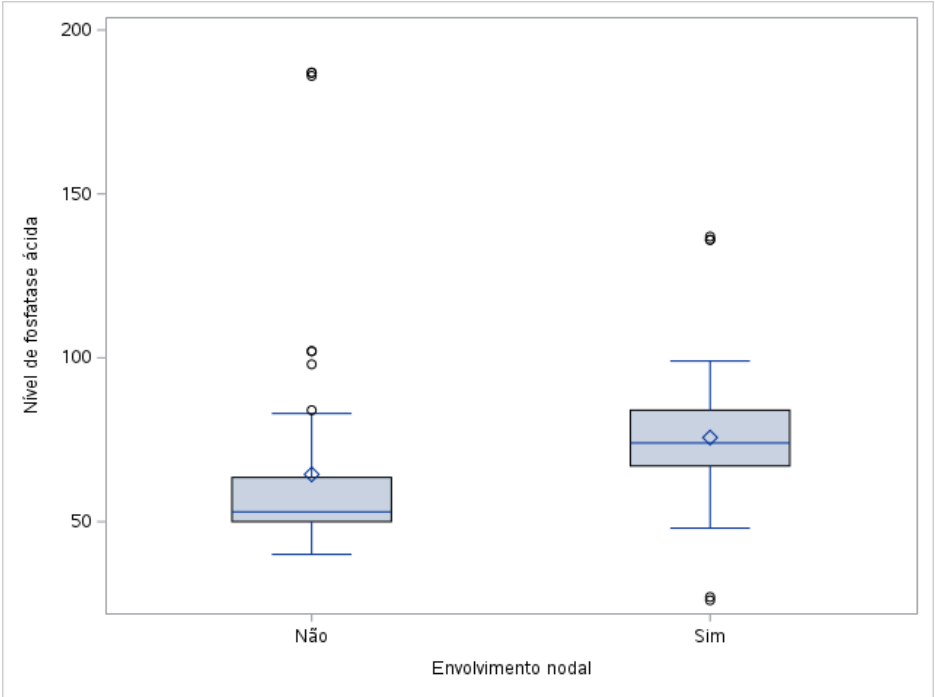
Quantiles (Definition 5)	
Level	Quantile
100% Max	137
99%	137
95%	136
90%	99
75% Q3	84
50% Median	74
25% Q1	67
10%	49
5%	48
1%	26
0% Min	26

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
26	96	99	17
27	71	99	73
48	82	136	19
48	60	136	75
49	84	137	55



Analysis Variable : X3 Nivel de fosfatase ácida						
Envolvimento nodal	N Obs	N	Mean	Std Dev	Minimum	Maximum
Não	60	60	64.4166667	31.5709004	40.0000000	187.0000000
Sim	42	42	75.6428571	23.7279514	26.0000000	137.0000000





	Nível de fosfatase ácida	
	Mean	Std
Envolvimento nodal		
Não	64.42	31.57
Sim	75.64	23.73

The LOGISTIC Procedure

Model Information

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	0	60
2	1	42

Probability modeled is X4='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	138.450
SC	142.834	143.700
-2 Log L	138.209	134.450

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	3.7590	1	0.0525
Score	3.7362	1	0.0532
Wald	3.3382	1	0.0677

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.3258	0.5649	5.5081	0.0189

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
X3	1	0.0140	0.00766	3.3382	0.0677

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X3	1.014	0.999	1.029

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	71.3	Somers' D	0.440
Percent Discordant	27.3	Gamma	0.446
Percent Tied	1.3	Tau-a	0.215
Pairs	2520	c	0.720

Estimated Covariance Matrix		
Parameter	Intercept	X3
Intercept	0.31912	-0.00403
X3	-0.00403	0.000059

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = 1		X4 = 0	
		Observed	Expected	Observed	Expected
1	13	4	4.24	9	8.76
2	7	2	2.42	5	4.58
3	8	0	2.79	8	5.21
4	11	2	3.89	9	7.11
5	10	0	3.67	10	6.33
6	12	6	4.76	6	7.24
7	11	8	4.65	3	6.35
8	11	9	4.93	2	6.07
9	11	8	5.33	3	5.67
10	8	3	5.33	5	2.67

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
28.0931	8	0.0005

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	94.194
SC	142.834	104.694
-2 Log L	138.209	86.194

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	52.0152	3	<.0001
Score	42.1766	3	<.0001

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Wald	21.9797	3	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.1039	1.1455	19.8514	<.0001
X1	1	2.7426	0.7361	13.8836	0.0002
X2	1	3.2084	0.7543	18.0894	<.0001
X3	1	0.0278	0.00904	9.4624	0.0021

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	15.528	3.669	65.712
X2	24.739	5.640	108.512
X3	1.028	1.010	1.047

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	89.3	Somers' D	0.788
Percent Discordant	10.5	Gamma	0.789
Percent Tied	0.2	Tau-a	0.385
Pairs	2520	c	0.894

Estimated Covariance Matrix				
Parameter	Intercept	X1	X2	X3
Intercept	1.312246	-0.49305	-0.71224	-0.0087
X1	-0.49305	0.54179	0.302397	0.002074
X2	-0.71224	0.302397	0.56904	0.003353
X3	-0.0087	0.002074	0.003353	0.000082

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	10	0	0.23	10	9.77
2	10	0	0.25	10	9.75

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
3	10	1	0.54	9	9.46
4	10	2	2.50	8	7.50
5	11	3	4.06	8	6.94
6	10	2	4.18	8	5.82
7	10	9	5.03	1	4.97
8	10	5	6.01	5	3.99
9	11	10	9.61	1	1.39
10	10	10	9.58	0	0.42

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
10.7117	8	0.2186

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	123.876
SC	142.834	129.126
-2 Log L	138.209	119.876

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	18.3327	1	<.0001
Score	18.1441	1	<.0001
Wald	16.1442	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.9555	0.2631	13.1878	0.0003
X1	1	1.9671	0.4896	16.1442	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	7.150	2.739	18.665

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	45.4	Somers' D	0.390
Percent Discordant	6.3	Gamma	0.755
Percent Tied	48.3	Tau-a	0.191
Pairs	2520	c	0.695

Estimated Covariance Matrix		
Parameter	Intercept	X1
Intercept	0.069231	-0.06923
X1	-0.06923	0.239685

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	120.029
SC	142.834	125.279
-2 Log L	138.209	116.029

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	22.1797	1	<.0001
Score	20.9969	1	<.0001
Wald	18.7463	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.5840	0.3881	16.6576	<.0001
X2	1	2.0659	0.4771	18.7463	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X2	7.892	3.098	20.107

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	52.6	Somers' D	0.460
Percent Discordant	6.7	Gamma	0.775
Percent Tied	40.7	Tau-a	0.225
Pairs	2520	c	0.730

Estimated Covariance Matrix		
Parameter	Intercept	X2
Intercept	0.150633	-0.15063
X2	-0.15063	0.227664

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	122.538
SC	142.834	130.413
-2 Log L	138.209	116.538

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	21.6704	2	<.0001
Score	21.0485	2	<.0001
Wald	18.1975	2	0.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.9102	0.6065	9.9193	0.0016
X1	1	1.9770	0.4973	15.8040	<.0001
X3	1	0.0137	0.00771	3.1433	0.0762

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	7.221	2.724	19.137
X3	1.014	0.999	1.029

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	79.7	Somers' D	0.600

Association of Predicted Probabilities and Observed Responses			
Percent Discordant	19.7	Gamma	0.603
Percent Tied	0.6	Tau-a	0.294
Pairs	2520	c	0.800

Estimated Covariance Matrix			
Parameter	Intercept	X1	X3
Intercept	0.36787	-0.08769	-0.00419
X1	-0.08769	0.247301	0.00022
X3	-0.00419	0.00022	0.000059

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	14	3	3.09	11	10.91
2	13	0	2.96	13	10.04
3	10	0	2.37	10	7.63
4	10	5	2.62	5	7.38
5	10	5	2.91	5	7.09
6	10	7	3.33	3	6.67
7	10	2	5.92	8	4.08
8	10	5	7.03	5	2.97
9	15	15	11.76	0	3.24

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
30.4936	7	<.0001

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	

Model Information		
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	112.643
SC	142.834	120.518
-2 Log L	138.209	106.643

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	31.5662	2	<.0001
Score	28.2044	2	<.0001
Wald	21.9174	2	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.6320	0.8710	17.3870	<.0001
X2	1	2.5493	0.5638	20.4437	<.0001
X3	1	0.0244	0.00840	8.4481	0.0037

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X2	12.798	4.239	38.641
X3	1.025	1.008	1.042

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	80.5	Somers' D	0.613
Percent Discordant	19.1	Gamma	0.616
Percent Tied	0.4	Tau-a	0.300
Pairs	2520	c	0.807

Estimated Covariance Matrix			
Parameter	Intercept	X2	X3
Intercept	0.758689	-0.35674	-0.00635
X2	-0.35674	0.317885	0.001878
X3	-0.00635	0.001878	0.000071

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	9	0	0.70	9	8.30
2	10	0	0.84	10	9.16
3	10	0	0.96	10	9.04
4	10	3	1.62	7	8.38
5	9	7	3.47	2	5.53
6	11	4	5.81	7	5.19
7	10	2	5.71	8	4.29
8	10	9	6.42	1	3.58
9	10	8	6.90	2	3.10
10	13	9	9.58	4	3.42

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
20.4259	8	0.0088

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	102.083
SC	142.834	109.957
-2 Log L	138.209	96.083

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	42.1263	2	<.0001
Score	36.2932	2	<.0001
Wald	22.3503	2	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.5945	0.5644	21.1290	<.0001
X1	1	2.4502	0.6292	15.1673	<.0001
X2	1	2.4937	0.5992	17.3172	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	11.591	3.377	39.780
X2	12.106	3.740	39.180

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	72.5	Somers' D	0.643
Percent Discordant	8.2	Gamma	0.797
Percent Tied	19.4	Tau-a	0.315
Pairs	2520	c	0.821

Estimated Covariance Matrix			
Parameter	Intercept	X1	X2
Intercept	0.31859	-0.22512	-0.28836
X1	-0.22512	0.395833	0.171955
X2	-0.28836	0.171955	0.359087

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	35	3	2.43	32	32.57
2	12	5	5.57	7	6.43
3	37	17	17.57	20	19.43
4	18	17	16.43	1	1.57

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
0.5109	2	0.7746

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Quasi-complete separation of data points detected.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	89.516
SC	142.834	110.516
-2 Log L	138.209	73.516

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	64.6928	7	<.0001

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Score	49.0345	7	<.0001
Wald	24.6557	7	0.0009

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.4464	1.2494	7.6084	0.0058
X1	1	-142.0	195.8	0.5261	0.4682
X2	1	0.0858	2.0686	0.0017	0.9669
X3	1	0.0131	0.0117	1.2525	0.2631
X1*X2	1	147.5	195.9	0.5670	0.4515
X2*X3	1	0.0376	0.0282	1.7813	0.1820
X1*X3	1	2.2272	3.0232	0.5427	0.4613
X1*X2*X3	1	-2.2664	3.0237	0.5618	0.4535

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	92.8	Somers' D	0.858
Percent Discordant	7.0	Gamma	0.859
Percent Tied	0.2	Tau-a	0.420
Pairs	2520	c	0.929

Estimated Covariance Matrix								
Parameter	Intercept	X1	X2	X3	X1X2	X2X3	X1X3	X1X2X3
Intercept	1.561097	-1.5611	-1.5611	-0.01267	1.561097	0.01267	0.01267	-0.01267
X1	-1.5611	38345.33	1.561097	0.01267	-38345.3	-0.01267	-591.688	591.688
X2	-1.5611	1.561097	4.278952	0.01267	-4.27895	-0.05395	-0.01267	0.053953
X3	-0.01267	0.01267	0.01267	0.000136	-0.01267	-0.00014	-0.00014	0.000136
X1X2	1.561097	-38345.3	-4.27895	-0.01267	38359.17	0.053953	591.688	-591.886
X2X3	0.01267	-0.01267	-0.05395	-0.00014	0.053953	0.000793	0.000136	-0.00079
X1X3	0.01267	-591.688	-0.01267	-0.00014	591.688	0.000136	9.139795	-9.13979
X1X2X3	-0.01267	591.688	0.053953	0.000136	-591.886	-0.00079	-9.13979	9.142885

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	9	0	0.13	9	8.87
2	12	0	0.69	12	11.31
3	10	0	0.65	10	9.35
4	10	3	1.17	7	8.83
5	9	3	2.53	6	6.47
6	10	0	3.44	10	6.56
7	10	8	5.09	2	4.91
8	10	7	7.24	3	2.76
9	11	10	10.36	1	0.64
10	11	11	10.72	0	0.28

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
14.0905	8	0.0794

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	97.258
SC	142.834	115.633
-2 Log L	138.209	83.258

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	54.9509	6	<.0001
Score	44.3960	6	<.0001
Wald	20.5241	6	0.0022

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.7704	1.2777	8.7082	0.0032
X1	1	-0.5836	2.4451	0.0570	0.8114
X2	1	1.3750	1.8929	0.5276	0.4676
X3	1	0.0164	0.0113	2.1220	0.1452
X1*X2	1	1.7736	1.8220	0.9476	0.3303
X2*X3	1	0.0189	0.0245	0.5918	0.4417
X1*X3	1	0.0390	0.0330	1.4012	0.2365

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	90.5	Somers' D	0.812
Percent Discordant	9.3	Gamma	0.814
Percent Tied	0.2	Tau-a	0.397
Pairs	2520	c	0.906

Estimated Covariance Matrix							
Parameter	Intercept	X1	X2	X3	X1X2	X2X3	X1X3
Intercept	1.632442	-1.43882	-1.55281	-0.01255	1.018684	0.011286	0.009503

Estimated Covariance Matrix							
Parameter	Intercept	X1	X2	X3	X1X2	X2X3	X1X3
X1	-1.43882	5.978506	0.831029	0.010585	-2.77197	-0.00095	-0.07253
X2	-1.55281	0.831029	3.583188	0.011741	-1.59215	-0.04207	-0.00039
X3	-0.01255	0.010585	0.011741	0.000127	-0.00633	-0.00011	-0.0001
X1X2	1.018684	-2.77197	-1.59215	-0.00633	3.319547	0.013569	0.024461
X2X3	0.011286	-0.00095	-0.04207	-0.00011	0.013569	0.000602	-0.00005
X1X3	0.009503	-0.07253	-0.00039	-0.0001	0.024461	-0.00005	0.001087

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	10	0	0.49	10	9.51
2	10	0	0.52	10	9.48
3	10	1	0.79	9	9.21
4	10	2	1.87	8	8.13
5	10	3	3.30	7	6.70
6	10	2	3.62	8	6.38
7	12	8	5.78	4	6.22
8	11	8	7.11	3	3.89
9	10	9	9.59	1	0.41
10	9	9	8.93	0	0.07

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
5.2155	8	0.7343

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	

Model Information		
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	97.464
SC	142.834	113.213
-2 Log L	138.209	85.464

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	52.7453	5	<.0001
Score	43.1149	5	<.0001
Wald	22.5598	5	0.0004

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.5214	1.2670	12.7358	0.0004
X1	1	2.3721	0.9823	5.8314	0.0157
X2	1	1.6998	1.8964	0.8035	0.3701
X3	1	0.0236	0.00986	5.7474	0.0165
X1*X2	1	0.7404	1.5044	0.2422	0.6226

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
X2*X3	1	0.0184	0.0239	0.5962	0.4401

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	89.7	Somers' D	0.796
Percent Discordant	10.1	Gamma	0.798
Percent Tied	0.2	Tau-a	0.390
Pairs	2520	c	0.898

Estimated Covariance Matrix						
Parameter	Intercept	X1	X2	X3	X1X2	X2X3
Intercept	1.605188	-0.80749	-1.60519	-0.01061	0.807488	0.010613
X1	-0.80749	0.964968	0.807488	0.003298	-0.96497	-0.0033
X2	-1.60519	0.807488	3.596308	0.010613	-1.25104	-0.04033
X3	-0.01061	0.003298	0.010613	0.000097	-0.0033	-0.0001
X1X2	0.807488	-0.96497	-1.25104	-0.0033	2.263107	0.008445
X2X3	0.010613	-0.0033	-0.04033	-0.0001	0.008445	0.000569

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	10	0	0.34	10	9.66
2	10	0	0.36	10	9.64
3	10	1	0.68	9	9.32
4	12	3	2.97	9	9.03
5	10	2	3.28	8	6.72
6	11	2	4.55	9	6.45
7	10	9	5.12	1	4.88
8	10	9	6.88	1	3.12
9	10	7	9.03	3	0.97
10	9	9	8.79	0	0.21

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
17.1315	8	0.0288

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	95.872
SC	142.834	111.621
-2 Log L	138.209	83.872

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	54.3372	5	<.0001
Score	44.1637	5	<.0001

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Wald	19.9062	5	0.0013

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.1533	1.2410	11.2012	0.0008
X1	1	-0.5924	2.5671	0.0533	0.8175
X2	1	2.7128	0.8385	10.4660	0.0012
X3	1	0.0202	0.0102	3.9606	0.0466
X1*X2	1	1.3727	1.6964	0.6548	0.4184
X1*X3	1	0.0415	0.0353	1.3815	0.2398

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	89.8	Somers' D	0.799
Percent Discordant	10.0	Gamma	0.800
Percent Tied	0.2	Tau-a	0.391
Pairs	2520	c	0.899

Estimated Covariance Matrix						
Parameter	Intercept	X1	X2	X3	X1X2	X1X3
Intercept	1.54003	-1.54003	-0.85959	-0.01075	0.859588	0.010752
X1	-1.54003	6.589816	0.859588	0.010752	-2.70637	-0.08185
X2	-0.85959	0.859588	0.70316	0.004232	-0.70316	-0.00423
X3	-0.01075	0.010752	0.004232	0.000103	-0.00423	-0.0001
X1X2	0.859588	-2.70637	-0.70316	-0.00423	2.877732	0.023836
X1X3	0.010752	-0.08185	-0.00423	-0.0001	0.023836	0.001246

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	10	0	0.41	10	9.59
2	10	0	0.43	10	9.57
3	10	1	0.74	9	9.26
4	11	2	2.26	9	8.74
5	12	5	4.61	7	7.39

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
6	10	0	4.19	10	5.81
7	10	9	4.84	1	5.16
8	10	7	6.04	3	3.96
9	10	9	9.58	1	0.42
10	9	9	8.89	0	0.11

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
16.5581	8	0.0351

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	96.324
SC	142.834	112.074
-2 Log L	138.209	84.324

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	53.8847	5	<.0001
Score	43.7892	5	<.0001
Wald	21.4871	5	0.0007

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.5073	1.2342	13.3377	0.0003
X1	1	0.9274	1.7984	0.2659	0.6061
X2	1	2.3734	1.6753	2.0069	0.1566
X3	1	0.0208	0.0110	3.5899	0.0581
X2*X3	1	0.0115	0.0227	0.2565	0.6125
X1*X3	1	0.0279	0.0276	1.0208	0.3123

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	91.0	Somers' D	0.822
Percent Discordant	8.8	Gamma	0.823
Percent Tied	0.2	Tau-a	0.402
Pairs	2520	c	0.911

Estimated Covariance Matrix						
Parameter	Intercept	X1	X2	X3	X2X3	X1X3
Intercept	1.52315	-0.73226	-1.20908	-0.01168	0.007682	0.003472
X1	-0.73226	3.234199	-0.3415	0.006077	0.009386	-0.045
X2	-1.20908	-0.3415	2.806768	0.009364	-0.03414	0.010041
X3	-0.01168	0.006077	0.009364	0.00012	-0.00009	-0.00006
X2X3	0.007682	0.009386	-0.03414	-0.00009	0.000517	-0.00015
X1X3	0.003472	-0.045	0.010041	-0.00006	-0.00015	0.000761

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	10	0	0.30	10	9.70
2	10	0	0.32	10	9.68
3	10	1	0.55	9	9.45
4	10	2	2.28	8	7.72
5	9	3	3.23	6	5.77
6	10	1	3.91	9	6.09
7	10	6	4.81	4	5.19
8	12	11	7.03	1	4.97
9	11	8	9.81	3	1.19
10	10	10	9.76	0	0.24

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
13.9799	8	0.0823

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	97.258
SC	142.834	115.633
-2 Log L	138.209	83.258

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	54.9509	6	<.0001
Score	44.3960	6	<.0001
Wald	20.5241	6	0.0022

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.7704	1.2777	8.7082	0.0032
X1	1	-0.5836	2.4451	0.0570	0.8114
X2	1	1.3750	1.8929	0.5276	0.4676
X3	1	0.0164	0.0113	2.1220	0.1452
X1*X2	1	1.7736	1.8220	0.9476	0.3303
X2*X3	1	0.0189	0.0245	0.5918	0.4417
X1*X3	1	0.0390	0.0330	1.4012	0.2365

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	90.5	Somers' D	0.812
Percent Discordant	9.3	Gamma	0.814
Percent Tied	0.2	Tau-a	0.397
Pairs	2520	c	0.906

Estimated Covariance Matrix							
Parameter	Intercept	X1	X2	X3	X1X2	X2X3	X1X3
Intercept	1.632442	-1.43882	-1.55281	-0.01255	1.018684	0.011286	0.009503

Estimated Covariance Matrix							
Parameter	Intercept	X1	X2	X3	X1X2	X2X3	X1X3
X1	-1.43882	5.978506	0.831029	0.010585	-2.77197	-0.00095	-0.07253
X2	-1.55281	0.831029	3.583188	0.011741	-1.59215	-0.04207	-0.00039
X3	-0.01255	0.010585	0.011741	0.000127	-0.00633	-0.00011	-0.0001
X1X2	1.018684	-2.77197	-1.59215	-0.00633	3.319547	0.013569	0.024461
X2X3	0.011286	-0.00095	-0.04207	-0.00011	0.013569	0.000602	-0.00005
X1X3	0.009503	-0.07253	-0.00039	-0.0001	0.024461	-0.00005	0.001087

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	10	0	0.49	10	9.51
2	10	0	0.52	10	9.48
3	10	1	0.79	9	9.21
4	10	2	1.87	8	8.13
5	10	3	3.30	7	6.70
6	10	2	3.62	8	6.38
7	12	8	5.78	4	6.22
8	11	8	7.11	3	3.89
9	10	9	9.59	1	0.41
10	9	9	8.93	0	0.07

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
5.2155	8	0.7343

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	

Model Information		
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	94.194
SC	142.834	104.694
-2 Log L	138.209	86.194

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	52.0152	3	<.0001
Score	42.1766	3	<.0001
Wald	21.9797	3	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.1039	1.1455	19.8514	<.0001
X1	1	2.7426	0.7361	13.8836	0.0002
X2	1	3.2084	0.7543	18.0894	<.0001
X3	1	0.0278	0.00904	9.4624	0.0021

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	15.528	3.669	65.712
X2	24.739	5.640	108.512
X3	1.028	1.010	1.047

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	89.3	Somers' D	0.788
Percent Discordant	10.5	Gamma	0.789
Percent Tied	0.2	Tau-a	0.385
Pairs	2520	c	0.894

Estimated Covariance Matrix				
Parameter	Intercept	X1	X2	X3
Intercept	1.312246	-0.49305	-0.71224	-0.0087
X1	-0.49305	0.54179	0.302397	0.002074
X2	-0.71224	0.302397	0.56904	0.003353
X3	-0.0087	0.002074	0.003353	0.000082

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	10	0	0.23	10	9.77
2	10	0	0.25	10	9.75
3	10	1	0.54	9	9.46
4	10	2	2.50	8	7.50
5	11	3	4.06	8	6.94
6	10	2	4.18	8	5.82
7	10	9	5.03	1	4.97
8	10	5	6.01	5	3.99
9	11	10	9.61	1	1.39
10	10	10	9.58	0	0.42

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
10.7117	8	0.2186

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	102.083
SC	142.834	109.957
-2 Log L	138.209	96.083

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	42.1263	2	<.0001
Score	36.2932	2	<.0001

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Wald	22.3503	2	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.5945	0.5644	21.1290	<.0001
X1	1	2.4502	0.6292	15.1673	<.0001
X2	1	2.4937	0.5992	17.3172	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	11.591	3.377	39.780
X2	12.106	3.740	39.180

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	72.5	Somers' D	0.643
Percent Discordant	8.2	Gamma	0.797
Percent Tied	19.4	Tau-a	0.315
Pairs	2520	c	0.821

Estimated Covariance Matrix			
Parameter	Intercept	X1	X2
Intercept	0.31859	-0.22512	-0.28836
X1	-0.22512	0.395833	0.171955
X2	-0.28836	0.171955	0.359087

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	35	3	2.43	32	32.57
2	12	5	5.57	7	6.43
3	37	17	17.57	20	19.43
4	18	17	16.43	1	1.57

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
0.5109	2	0.7746

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Stepwise Selection Procedure

Step 0. Intercept entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

-2 Log L	=	138.209
----------	---	---------

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.3567	0.2012	3.1430	0.0763

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
49.0345	7	<.0001

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
X1	1	18.1441	<.0001
X2	1	20.9969	<.0001
X3	1	3.7362	0.0532

Step 1. Effect X2 entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	120.029
SC	142.834	125.279
-2 Log L	138.209	116.029

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	22.1797	1	<.0001
Score	20.9969	1	<.0001
Wald	18.7463	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.5840	0.3881	16.6576	<.0001
X2	1	2.0659	0.4771	18.7463	<.0001

Odds Ratio Estimates		
Effect	Point Estimate	95% Wald Confidence Limits

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X2	7.892	3.098	20.107

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	52.6	Somers' D	0.460
Percent Discordant	6.7	Gamma	0.775
Percent Tied	40.7	Tau-a	0.225
Pairs	2520	c	0.730

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
36.8610	6	<.0001

Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
X2	1	18.7463	<.0001

Note: No effects for the model in Step 1 are removed.

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
X1	1	18.9236	<.0001
X3	1	10.9756	0.0009

Step 2. Effect X1 entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	102.083
SC	142.834	109.957

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
-2 Log L	138.209	96.083

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	42.1263	2	<.0001
Score	36.2932	2	<.0001
Wald	22.3503	2	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.5945	0.5644	21.1290	<.0001
X1	1	2.4502	0.6292	15.1673	<.0001
X2	1	2.4937	0.5992	17.3172	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	11.591	3.377	39.780
X2	12.106	3.740	39.180

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	72.5	Somers' D	0.643
Percent Discordant	8.2	Gamma	0.797
Percent Tied	19.4	Tau-a	0.315
Pairs	2520	c	0.821

Residual Chi-Square Test			
Chi-Square	DF	Pr > ChiSq	
13.7274	5	0.0174	

Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
X1	1	15.1673	<.0001

Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
X2	1	17.3172	<.0001

Note: No effects for the model in Step 2 are removed.

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
X3	1	11.9368	0.0006
X1*X2	1	0.5109	0.4748

Step 3. Effect X3 entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	94.194
SC	142.834	104.694
-2 Log L	138.209	86.194

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	52.0152	3	<.0001
Score	42.1766	3	<.0001
Wald	21.9797	3	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.1039	1.1455	19.8514	<.0001
X1	1	2.7426	0.7361	13.8836	0.0002
X2	1	3.2084	0.7543	18.0894	<.0001
X3	1	0.0278	0.00904	9.4624	0.0021

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	15.528	3.669	65.712
X2	24.739	5.640	108.512
X3	1.028	1.010	1.047

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	89.3	Somers' D	0.788
Percent Discordant	10.5	Gamma	0.789
Percent Tied	0.2	Tau-a	0.385
Pairs	2520	c	0.894

Residual Chi-Square Test			
Chi-Square	DF	Pr > ChiSq	
4.7892	4	0.3096	

Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
X1	1	13.8836	0.0002
X2	1	18.0894	<.0001
X3	1	9.4624	0.0021

Note: No effects for the model in Step 3 are removed.

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
X1*X2	1	0.1118	0.7381
X2*X3	1	0.4760	0.4903
X1*X3	1	1.3408	0.2469

Step 4. Effect X1*X3 entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	94.584
SC	142.834	107.709
-2 Log L	138.209	84.584

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	53.6251	4	<.0001
Score	43.4489	4	<.0001
Wald	20.7087	4	0.0004

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.6889	1.1937	15.4286	<.0001
X1	1	0.7043	1.8542	0.1443	0.7041
X2	1	3.1410	0.7520	17.4459	<.0001
X3	1	0.0228	0.0101	5.1272	0.0236
X1*X3	1	0.0315	0.0289	1.1932	0.2747

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X2	23.126	5.297	100.972

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	90.7	Somers' D	0.816
Percent Discordant	9.1	Gamma	0.818
Percent Tied	0.2	Tau-a	0.399
Pairs	2520	c	0.908

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
4.3308	3	0.2279

Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
X2	1	17.4459	<.0001
X1*X3	1	1.1932	0.2747

Note: No effects for the model in Step 4 are removed.

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
X1*X2	1	0.6841	0.4082
X2*X3	1	0.2576	0.6118

Note: No (additional) effects met the 0.3 significance level for entry into the model.

Summary of Stepwise Selection								
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
	Entered	Removed						
1	X2		1	1	20.9969		<.0001	Estágio do tumor
2	X1		1	2	18.9236		<.0001	Resultado da radiografia
3	X3		1	3	11.9368		0.0006	Nível de fosfatase ácida
4	X1*X3		1	4	1.3408		0.2469	

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	10	0	0.27	10	9.73
2	10	0	0.30	10	9.70
3	10	1	0.55	9	9.45
4	10	2	2.28	8	7.72
5	10	3	3.90	7	6.10
6	10	2	4.14	8	5.86
7	12	8	5.86	4	6.14
8	11	8	6.63	3	4.37
9	10	9	9.31	1	0.69
10	9	9	8.77	0	0.23

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
5.8827	8	0.6604

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	94.194
SC	142.834	104.694
-2 Log L	138.209	86.194

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	52.0152	3	<.0001
Score	42.1766	3	<.0001

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Wald	21.9797	3	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.1039	1.1455	19.8514	<.0001
X1	1	2.7426	0.7361	13.8836	0.0002
X2	1	3.2084	0.7543	18.0894	<.0001
X3	1	0.0278	0.00904	9.4624	0.0021

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	15.528	3.669	65.712
X2	24.739	5.640	108.512
X3	1.028	1.010	1.047

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	89.3	Somers' D	0.788
Percent Discordant	10.5	Gamma	0.789
Percent Tied	0.2	Tau-a	0.385
Pairs	2520	c	0.894

Estimated Covariance Matrix				
Parameter	Intercept	X1	X2	X3
Intercept	1.312246	-0.49305	-0.71224	-0.0087
X1	-0.49305	0.54179	0.302397	0.002074
X2	-0.71224	0.302397	0.56904	0.003353
X3	-0.0087	0.002074	0.003353	0.000082

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	10	0	0.23	10	9.77
2	10	0	0.25	10	9.75

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
3	10	1	0.54	9	9.46
4	10	2	2.50	8	7.50
5	11	3	4.06	8	6.94
6	10	2	4.18	8	5.82
7	10	9	5.03	1	4.97
8	10	5	6.01	5	3.99
9	11	10	9.61	1	1.39
10	10	10	9.58	0	0.42

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
10.7117	8	0.2186

Obs	_LINK_	_TYPE_	_STATUS_	_NAME_	Intercept	X1	X2	X3	_LNLIKE_	_ESTTYPE_
1	LOGIT	PARMS	0 Converged	X4	-5.10392	2.742629	3.208363981	0.027809	-43.0969	MLE
2	LOGIT	COV	0 Converged	Intercept	1.31225	-0.49305	-0.71223906	-0.008701	-43.0969	MLE
3	LOGIT	COV	0 Converged	X1	-0.49305	0.54179	0.302397316	0.002074	-43.0969	MLE
4	LOGIT	COV	0 Converged	X2	-0.71224	0.302397	0.569040191	0.003353	-43.0969	MLE
5	LOGIT	COV	0 Converged	X3	-0.00870	0.002074	0.003353139	0.000082	-43.0969	MLE

Obs	ID	X1	X2	X3	X4	_FROM_	_INTO_	IP_Não	IP_Sim	XP_Não	XP_Sim	_LEVEL_	phat	lcl	ucl
1	2	negativo	menos grave	56	Não	Não	Não	0.97199	0.02801	0.97152	0.02848	Sim	0.02801	0.00632	0.11547
2	3	negativo	menos grave	50	Não	Não	Não	0.97619	0.02381	0.97582	0.02418	Sim	0.02381	0.00502	0.10549
3	4	negativo	menos grave	52	Não	Não	Não	0.97486	0.02514	0.97446	0.02554	Sim	0.02514	0.00542	0.10868
4	5	negativo	menos grave	50	Não	Não	Não	0.97619	0.02381	0.97582	0.02418	Sim	0.02381	0.00502	0.10549
5	6	negativo	menos grave	49	Não	Não	Não	0.97683	0.02317	0.97647	0.02353	Sim	0.02317	0.00483	0.10394
6	7	positivo	menos grave	46	Não	Não	Não	0.74689	0.25311	0.72378	0.27622	Sim	0.25311	0.08557	0.55103
7	8	positivo	menos grave	62	Não	Não	Não	0.65411	0.34589	0.62280	0.37720	Sim	0.34589	0.14021	0.63164
8	10	positivo	menos grave	55	Não	Não	Não	0.69674	0.30326	0.66914	0.33086	Sim	0.30326	0.11408	0.59534
9	14	positivo	menos grave	67	Sim	Sim	Não	0.62201	0.37799	0.67544	0.32456	Sim	0.37799	0.16082	0.65835

Obs	ID	X1	X2	X3	X4	_FROM_	_INTO_	IP_Não	IP_Sim	XP_Não	XP_Sim	_LEVEL_	phat	lcl	ucl
10	15	negativo	menos grave	47	Não	Não	Não	0.97805	0.02195	0.97773	0.02227	Sim	0.02195	0.00447	0.10094
11	16	negativo	menos grave	49	Não	Não	Não	0.97683	0.02317	0.97647	0.02353	Sim	0.02317	0.00483	0.10394
12	17	negativo	menos grave	50	Não	Não	Não	0.97619	0.02381	0.97582	0.02418	Sim	0.02381	0.00502	0.10549
13	18	negativo	menos grave	78	Não	Não	Não	0.94954	0.05046	0.94841	0.05159	Sim	0.05046	0.01403	0.16561
14	19	negativo	menos grave	83	Não	Não	Não	0.94245	0.05755	0.94105	0.05895	Sim	0.05755	0.01659	0.18104
15	20	negativo	menos grave	98	Não	Não	Não	0.91518	0.08482	0.91249	0.08751	Sim	0.08482	0.02646	0.24012
16	21	negativo	menos grave	52	Não	Não	Não	0.97486	0.02514	0.97446	0.02554	Sim	0.02514	0.00542	0.10868
17	23	negativo	menos grave	99	Sim	Sim	Não	0.91300	0.08700	0.93812	0.06188	Sim	0.08700	0.02724	0.24485
18	24	negativo	menos grave	187	Não	Não	Sim	0.47593	0.52407	0.32775	0.67225	Sim	0.52407	0.14438	0.87783
19	25	positivo	menos grave	136	Sim	Sim	Sim	0.19455	0.80545	0.21461	0.78539	Sim	0.80545	0.48305	0.94830
20	27	negativo	mais grave	40	Não	Não	Não	0.68637	0.31363	0.67510	0.32490	Sim	0.31363	0.17272	0.50002
21	28	negativo	mais grave	50	Não	Não	Não	0.62366	0.37634	0.61209	0.38791	Sim	0.37634	0.23118	0.54771
22	29	negativo	mais grave	50	Não	Não	Não	0.62366	0.37634	0.61209	0.38791	Sim	0.37634	0.23118	0.54771
23	31	negativo	mais grave	55	Não	Não	Não	0.59051	0.40949	0.57867	0.42133	Sim	0.40949	0.26254	0.57461
24	32	negativo	mais grave	59	Não	Não	Não	0.56337	0.43663	0.55116	0.44884	Sim	0.43663	0.28792	0.59769
25	34	positivo	mais grave	51	Sim	Sim	Sim	0.09404	0.90596	0.09827	0.90173	Sim	0.90596	0.70789	0.97456
26	35	negativo	mais grave	49	Sim	Sim	Não	0.63016	0.36984	0.64946	0.35054	Sim	0.36984	0.22502	0.54259
27	36	negativo	mais grave	48	Não	Não	Não	0.63662	0.36338	0.62513	0.37487	Sim	0.36338	0.21892	0.53755
28	38	negativo	mais grave	102	Não	Não	Sim	0.28071	0.71929	0.24910	0.75090	Sim	0.71929	0.50748	0.86436
29	39	negativo	mais grave	76	Não	Não	Sim	0.44573	0.55427	0.42946	0.57054	Sim	0.55427	0.38950	0.70791
30	42	positivo	mais grave	84	Sim	Sim	Sim	0.03981	0.96019	0.04083	0.95917	Sim	0.96019	0.83346	0.99147
31	46	positivo	mais grave	78	Sim	Sim	Sim	0.04670	0.95330	0.04801	0.95199	Sim	0.95330	0.81658	0.98943
32	47	negativo	mais grave	70	Sim	Sim	Sim	0.48724	0.51276	0.50082	0.49918	Sim	0.51276	0.35558	0.66746
33	48	negativo	mais grave	67	Sim	Sim	Não	0.50809	0.49191	0.52204	0.47796	Sim	0.49191	0.33771	0.64767
34	49	negativo	mais grave	82	Sim	Sim	Sim	0.40498	0.59502	0.41806	0.58194	Sim	0.59502	0.42072	0.74826
35	50	negativo	mais grave	67	Sim	Sim	Não	0.50809	0.49191	0.52204	0.47796	Sim	0.49191	0.33771	0.64767
36	51	positivo	mais grave	72	Sim	Sim	Sim	0.05472	0.94528	0.05640	0.94360	Sim	0.94528	0.79743	0.98698
37	52	positivo	mais grave	89	Sim	Sim	Sim	0.03482	0.96518	0.03565	0.96435	Sim	0.96518	0.84602	0.99290
38	54	negativo	menos grave	49	Não	Não	Não	0.97683	0.02317	0.97647	0.02353	Sim	0.02317	0.00483	0.10394
39	55	negativo	menos grave	55	Não	Não	Não	0.97273	0.02727	0.97229	0.02771	Sim	0.02727	0.00609	0.11372
40	57	negativo	menos grave	53	Não	Não	Não	0.97417	0.02583	0.97376	0.02624	Sim	0.02583	0.00564	0.11032
41	59	negativo	menos grave	48	Não	Não	Não	0.97745	0.02255	0.97711	0.02289	Sim	0.02255	0.00464	0.10243
42	60	positivo	menos grave	45	Não	Não	Não	0.75211	0.24789	0.72948	0.27052	Sim	0.24789	0.08276	0.54629
43	61	positivo	menos grave	63	Não	Não	Não	0.64779	0.35221	0.61594	0.38406	Sim	0.35221	0.14421	0.63694
44	63	positivo	menos grave	54	Não	Não	Não	0.70258	0.29742	0.67550	0.32450	Sim	0.29742	0.11063	0.59028
45	64	negativo	menos grave	61	Não	Não	Não	0.96794	0.03206	0.96737	0.03263	Sim	0.03206	0.00763	0.12480

Obs	ID	X1	X2	X3	X4	_FROM_	_INTO_	IP_Não	IP_Sim	XP_Não	XP_Sim	_LEVEL_	phat	lcl	ucl
46	67	positivo	menos grave	68	Sim	Sim	Não	0.61545	0.38455	0.66843	0.33157	Sim	0.38455	0.16512	0.66375
47	69	negativo	menos grave	50	Não	Não	Não	0.97619	0.02381	0.97582	0.02418	Sim	0.02381	0.00502	0.10549
48	70	negativo	menos grave	51	Não	Não	Não	0.97553	0.02447	0.97515	0.02485	Sim	0.02447	0.00522	0.10707
49	71	negativo	menos grave	79	Não	Não	Não	0.94819	0.05181	0.94702	0.05298	Sim	0.05181	0.01451	0.16855
50	72	negativo	menos grave	84	Não	Não	Não	0.94092	0.05908	0.93947	0.06053	Sim	0.05908	0.01714	0.18436
51	74	negativo	menos grave	51	Não	Não	Não	0.97553	0.02447	0.97515	0.02485	Sim	0.02447	0.00522	0.10707
52	75	negativo	menos grave	76	Não	Não	Não	0.95214	0.04786	0.95110	0.04890	Sim	0.04786	0.01310	0.15994
53	76	negativo	menos grave	98	Sim	Sim	Não	0.91518	0.08482	0.93982	0.06018	Sim	0.08482	0.02646	0.24012
54	77	negativo	menos grave	186	Não	Não	Sim	0.48287	0.51713	0.33822	0.66178	Sim	0.51713	0.14262	0.87333
55	78	positivo	menos grave	137	Sim	Sim	Sim	0.19023	0.80977	0.20982	0.79018	Sim	0.80977	0.48709	0.95020
56	79	negativo	mais grave	81	Sim	Sim	Sim	0.41169	0.58831	0.42478	0.57522	Sim	0.58831	0.41571	0.74161
57	81	negativo	mais grave	51	Não	Não	Não	0.61711	0.38289	0.60550	0.39450	Sim	0.38289	0.23738	0.55292
58	84	negativo	mais grave	56	Não	Não	Não	0.58377	0.41623	0.57185	0.42815	Sim	0.41623	0.26888	0.58025
59	85	negativo	mais grave	60	Não	Não	Não	0.55651	0.44349	0.54419	0.45581	Sim	0.44349	0.29425	0.60367
60	88	negativo	mais grave	48	Sim	Sim	Não	0.63662	0.36338	0.65638	0.34362	Sim	0.36338	0.21892	0.53755
61	89	negativo	mais grave	49	Não	Não	Não	0.63016	0.36984	0.61864	0.38136	Sim	0.36984	0.22502	0.54259
62	90	positivo	mais grave	64	Não	Não	Sim	0.06743	0.93257	0.04145	0.95855	Sim	0.93257	0.76782	0.98300
63	95	positivo	mais grave	85	Sim	Sim	Sim	0.03876	0.96124	0.03974	0.96026	Sim	0.96124	0.83608	0.99178
64	97	positivo	mais grave	77	Sim	Sim	Sim	0.04795	0.95205	0.04932	0.95068	Sim	0.95205	0.81355	0.98905
65	98	negativo	mais grave	71	Sim	Sim	Sim	0.48029	0.51971	0.49378	0.50622	Sim	0.51971	0.36141	0.67414
66	99	positivo	mais grave	79	Sim	Sim	Sim	0.04548	0.95452	0.04673	0.95327	Sim	0.95452	0.81954	0.98980
67	100	negativo	mais grave	70	Sim	Sim	Sim	0.48724	0.51276	0.50082	0.49918	Sim	0.51276	0.35558	0.66746
68	101	negativo	mais grave	67	Sim	Sim	Não	0.50809	0.49191	0.52204	0.47796	Sim	0.49191	0.33771	0.64767
69	102	negativo	mais grave	82	Sim	Sim	Sim	0.40498	0.59502	0.41806	0.58194	Sim	0.59502	0.42072	0.74826
70	103	negativo	mais grave	67	Sim	Sim	Não	0.50809	0.49191	0.52204	0.47796	Sim	0.49191	0.33771	0.64767
71	106	positivo	mais grave	27	Sim	Sim	Sim	0.16827	0.83173	0.18132	0.81868	Sim	0.83173	0.55279	0.95184
72	108	negativo	menos grave	75	Não	Não	Não	0.95339	0.04661	0.95239	0.04761	Sim	0.04661	0.01265	0.15720
73	109	negativo	menos grave	99	Sim	Sim	Não	0.91300	0.08700	0.93812	0.06188	Sim	0.08700	0.02724	0.24485
74	110	negativo	menos grave	187	Não	Não	Sim	0.47593	0.52407	0.32775	0.67225	Sim	0.52407	0.14438	0.87783
75	111	positivo	menos grave	136	Sim	Sim	Sim	0.19455	0.80545	0.21461	0.78539	Sim	0.80545	0.48305	0.94830
76	112	negativo	mais grave	82	Sim	Sim	Sim	0.40498	0.59502	0.41806	0.58194	Sim	0.59502	0.42072	0.74826
77	114	negativo	mais grave	50	Não	Não	Não	0.62366	0.37634	0.61209	0.38791	Sim	0.37634	0.23118	0.54771
78	115	negativo	mais grave	50	Não	Não	Não	0.62366	0.37634	0.61209	0.38791	Sim	0.37634	0.23118	0.54771
79	116	negativo	mais grave	40	Não	Não	Não	0.68637	0.31363	0.67510	0.32490	Sim	0.31363	0.17272	0.50002
80	117	negativo	mais grave	55	Não	Não	Não	0.59051	0.40949	0.57867	0.42133	Sim	0.40949	0.26254	0.57461
81	118	negativo	mais grave	59	Não	Não	Não	0.56337	0.43663	0.55116	0.44884	Sim	0.43663	0.28792	0.59769

Obs	ID	X1	X2	X3	X4	_FROM_	_INTO_	IP_Não	IP_Sim	XP_Não	XP_Sim	_LEVEL_	phat	lcl	ucl
82	119	positivo	mais grave	48	Sim	Sim	Sim	0.10139	0.89861	0.10624	0.89376	Sim	0.89861	0.69173	0.97223
83	120	positivo	mais grave	51	Sim	Sim	Sim	0.09404	0.90596	0.09827	0.90173	Sim	0.90596	0.70789	0.97456
84	121	negativo	mais grave	49	Sim	Sim	Não	0.63016	0.36984	0.64946	0.35054	Sim	0.36984	0.22502	0.54259
85	122	negativo	mais grave	48	Não	Não	Não	0.63662	0.36338	0.62513	0.37487	Sim	0.36338	0.21892	0.53755
86	124	negativo	mais grave	102	Não	Não	Sim	0.28071	0.71929	0.24910	0.75090	Sim	0.71929	0.50748	0.86436
87	127	negativo	mais grave	66	Não	Não	Não	0.51504	0.48496	0.50173	0.49827	Sim	0.48496	0.33163	0.64118
88	128	positivo	mais grave	84	Sim	Sim	Sim	0.03981	0.96019	0.04083	0.95917	Sim	0.96019	0.83346	0.99147
89	130	positivo	mais grave	76	Sim	Sim	Sim	0.04924	0.95076	0.05066	0.94934	Sim	0.95076	0.81046	0.98866
90	131	negativo	mais grave	70	Sim	Sim	Sim	0.48724	0.51276	0.50082	0.49918	Sim	0.51276	0.35558	0.66746
91	132	positivo	mais grave	78	Sim	Sim	Sim	0.04670	0.95330	0.04801	0.95199	Sim	0.95330	0.81658	0.98943
92	135	negativo	mais grave	82	Sim	Sim	Sim	0.40498	0.59502	0.41806	0.58194	Sim	0.59502	0.42072	0.74826
93	136	negativo	mais grave	67	Sim	Sim	Não	0.50809	0.49191	0.52204	0.47796	Sim	0.49191	0.33771	0.64767
94	137	positivo	mais grave	72	Sim	Sim	Sim	0.05472	0.94528	0.05640	0.94360	Sim	0.94528	0.79743	0.98698
95	138	positivo	mais grave	89	Sim	Sim	Sim	0.03482	0.96518	0.03565	0.96435	Sim	0.96518	0.84602	0.99290
96	139	positivo	mais grave	26	Sim	Sim	Sim	0.17220	0.82780	0.18589	0.81411	Sim	0.82780	0.54514	0.95069
97	140	negativo	menos grave	49	Não	Não	Não	0.97683	0.02317	0.97647	0.02353	Sim	0.02317	0.00483	0.10394
98	141	negativo	menos grave	55	Não	Não	Não	0.97273	0.02727	0.97229	0.02771	Sim	0.02727	0.00609	0.11372
99	142	negativo	menos grave	51	Não	Não	Não	0.97553	0.02447	0.97515	0.02485	Sim	0.02447	0.00522	0.10707
100	143	negativo	menos grave	53	Não	Não	Não	0.97417	0.02583	0.97376	0.02624	Sim	0.02583	0.00564	0.11032
101	144	negativo	menos grave	51	Não	Não	Não	0.97553	0.02447	0.97515	0.02485	Sim	0.02447	0.00522	0.10707
102	146	positivo	menos grave	45	Não	Não	Não	0.75211	0.24789	0.72948	0.27052	Sim	0.24789	0.08276	0.54629

The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

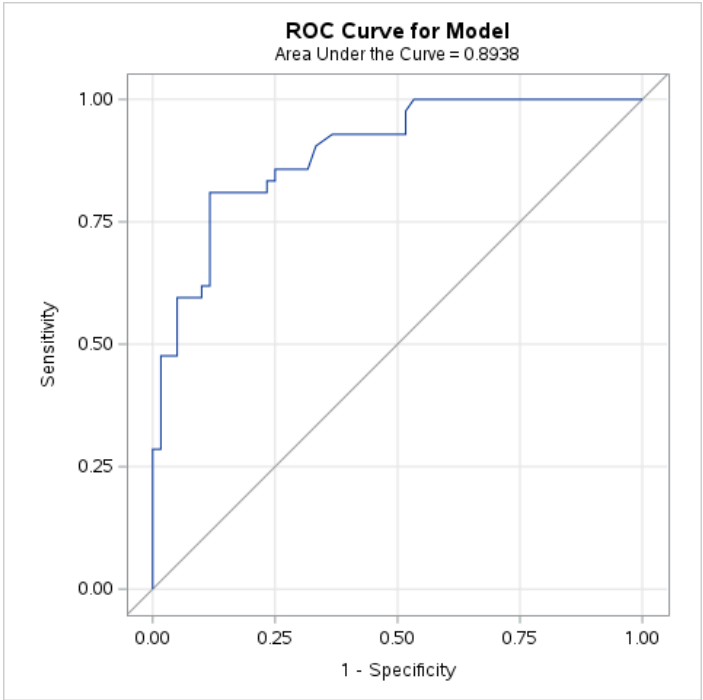
Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	94.194
SC	142.834	104.694
-2 Log L	138.209	86.194

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	52.0152	3	<.0001
Score	42.1766	3	<.0001
Wald	21.9797	3	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.1039	1.1455	19.8514	<.0001
X1	1	2.7426	0.7361	13.8836	0.0002
X2	1	3.2084	0.7543	18.0894	<.0001
X3	1	0.0278	0.00904	9.4624	0.0021

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	15.528	3.669	65.712
X2	24.739	5.640	108.512
X3	1.028	1.010	1.047

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	89.3	Somers' D	0.788
Percent Discordant	10.5	Gamma	0.789
Percent Tied	0.2	Tau-a	0.385
Pairs	2520	c	0.894



The LOGISTIC Procedure

Model Information		
Data Set	TRABALHO.DF	
Response Variable	X4	Envolvimento nodal

Model Information		
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	102
Number of Observations Used	102

Response Profile		
Ordered Value	X4	Total Frequency
1	Não	60
2	Sim	42

Probability modeled is X4='Sim'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.209	94.194
SC	142.834	104.694
-2 Log L	138.209	86.194

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	52.0152	3	<.0001
Score	42.1766	3	<.0001
Wald	21.9797	3	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.1039	1.1455	19.8514	<.0001
X1	1	2.7426	0.7361	13.8836	0.0002
X2	1	3.2084	0.7543	18.0894	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
X3	1	0.0278	0.00904	9.4624	0.0021

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	15.528	3.669	65.712
X2	24.739	5.640	108.512
X3	1.028	1.010	1.047

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	89.3	Somers' D	0.788
Percent Discordant	10.5	Gamma	0.789
Percent Tied	0.2	Tau-a	0.385
Pairs	2520	c	0.894

Estimated Covariance Matrix				
Parameter	Intercept	X1	X2	X3
Intercept	1.312246	-0.49305	-0.71224	-0.0087
X1	-0.49305	0.54179	0.302397	0.002074
X2	-0.71224	0.302397	0.56904	0.003353
X3	-0.0087	0.002074	0.003353	0.000082

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
1	10	0	0.23	10	9.77
2	10	0	0.25	10	9.75
3	10	1	0.54	9	9.46
4	10	2	2.50	8	7.50
5	11	3	4.06	8	6.94
6	10	2	4.18	8	5.82
7	10	9	5.03	1	4.97
8	10	5	6.01	5	3.99
9	11	10	9.61	1	1.39

Partition for the Hosmer and Lemeshow Test					
Group	Total	X4 = Sim		X4 = Não	
		Observed	Expected	Observed	Expected
10	10	10	9.58	0	0.42

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
10.7117	8	0.2186

Obs	ID	X1	X2	X3	X4	F_X4	L_X4	P_Não	P_Sim
1	2	0	0	56	0		Não	0.97199	0.02801
2	3	0	0	50	0		Não	0.97619	0.02381
3	5	0	0	50	0		Não	0.97619	0.02381
4	7	1	0	46	0		Não	0.74689	0.25311
5	8	1	0	62	0		Não	0.65411	0.34589
6	9	0	0	56	1		Não	0.97199	0.02801
7	10	1	0	55	0		Não	0.69674	0.30326
8	11	0	0	62	0		Não	0.96707	0.03293
9	12	0	0	71	0		Não	0.95809	0.04191
10	13	0	0	65	0		Não	0.96430	0.03570
11	14	1	0	67	1		Não	0.62201	0.37799
12	15	0	0	47	0		Não	0.97805	0.02195
13	17	0	0	50	0		Não	0.97619	0.02381
14	20	0	0	98	0		Não	0.91518	0.08482
15	21	0	0	52	0		Não	0.97486	0.02514
16	22	0	0	75	0		Não	0.95339	0.04661
17	23	0	0	99	1		Não	0.91300	0.08700
18	24	0	0	187	0		Sim	0.47593	0.52407
19	25	1	0	136	1		Sim	0.19455	0.80545
20	26	0	1	82	1		Sim	0.40498	0.59502
21	27	0	1	40	0		Não	0.68637	0.31363
22	31	0	1	55	0		Não	0.59051	0.40949
23	32	0	1	59	0		Não	0.56337	0.43663
24	33	1	1	48	1		Sim	0.10139	0.89861
25	34	1	1	51	1		Sim	0.09404	0.90596

Obs	ID	X1	X2	X3	X4	F_X4	I_X4	P_Não	P_Sim
26	36	0	1	48	0		Não	0.63662	0.36338
27	40	0	1	95	0		Sim	0.32163	0.67837
28	41	0	1	66	0		Não	0.51504	0.48496
29	42	1	1	84	1		Sim	0.03981	0.96019
30	44	1	1	76	1		Sim	0.04924	0.95076
31	45	0	1	70	1		Sim	0.48724	0.51276
32	46	1	1	78	1		Sim	0.04670	0.95330
33	47	0	1	70	1		Sim	0.48724	0.51276
34	49	0	1	82	1		Sim	0.40498	0.59502
35	50	0	1	67	1		Não	0.50809	0.49191
36	52	1	1	89	1		Sim	0.03482	0.96518
37	54	0	0	49	0		Não	0.97683	0.02317
38	56	0	0	51	0		Não	0.97553	0.02447
39	57	0	0	53	0		Não	0.97417	0.02583
40	58	0	0	51	0		Não	0.97553	0.02447
41	59	0	0	48	0		Não	0.97745	0.02255
42	60	1	0	45	0		Não	0.75211	0.24789
43	62	0	0	57	1		Não	0.97122	0.02878
44	63	1	0	54	0		Não	0.70258	0.29742
45	64	0	0	61	0		Não	0.96794	0.03206
46	65	0	0	70	0		Não	0.95920	0.04080
47	66	0	0	66	0		Não	0.96333	0.03667
48	68	0	0	48	0		Não	0.97745	0.02255
49	69	0	0	50	0		Não	0.97619	0.02381
50	71	0	0	79	0		Não	0.94819	0.05181
51	72	0	0	84	0		Não	0.94092	0.05908
52	73	0	0	97	0		Não	0.91732	0.08268
53	74	0	0	51	0		Não	0.97553	0.02447
54	75	0	0	76	0		Não	0.95214	0.04786
55	77	0	0	186	0		Sim	0.48287	0.51713
56	78	1	0	137	1		Sim	0.19023	0.80977
57	80	0	1	41	0		Não	0.68035	0.31965
58	81	0	1	51	0		Não	0.61711	0.38289
59	84	0	1	56	0		Não	0.58377	0.41623
60	85	0	1	60	0		Não	0.55651	0.44349
61	86	1	1	49	1		Sim	0.09888	0.90112

Obs	ID	X1	X2	X3	X4	F_X4	I_X4	P_Não	P_Sim
62	87	1	1	51	1		Sim	0.09404	0.90596
63	88	0	1	48	1		Não	0.63662	0.36338
64	89	0	1	49	0		Não	0.63016	0.36984
65	90	1	1	64	0		Sim	0.06743	0.93257
66	91	0	1	101	0		Sim	0.28636	0.71364
67	92	0	1	77	0		Sim	0.43888	0.56112
68	93	0	1	96	0		Sim	0.31559	0.68441
69	94	0	1	67	0		Não	0.50809	0.49191
70	95	1	1	85	1		Sim	0.03876	0.96124
71	97	1	1	77	1		Sim	0.04795	0.95205
72	100	0	1	70	1		Sim	0.48724	0.51276
73	101	0	1	67	1		Não	0.50809	0.49191
74	104	1	1	72	1		Sim	0.05472	0.94528
75	107	0	0	52	0		Não	0.97486	0.02514
76	108	0	0	75	0		Não	0.95339	0.04661
77	109	0	0	99	1		Não	0.91300	0.08700
78	110	0	0	187	0		Sim	0.47593	0.52407
79	111	1	0	136	1		Sim	0.19455	0.80545
80	114	0	1	50	0		Não	0.62366	0.37634
81	115	0	1	50	0		Não	0.62366	0.37634
82	116	0	1	40	0		Não	0.68637	0.31363
83	117	0	1	55	0		Não	0.59051	0.40949
84	118	0	1	59	0		Não	0.56337	0.43663
85	119	1	1	48	1		Sim	0.10139	0.89861
86	122	0	1	48	0		Não	0.63662	0.36338
87	124	0	1	102	0		Sim	0.28071	0.71929
88	125	0	1	76	0		Sim	0.44573	0.55427
89	128	1	1	84	1		Sim	0.03981	0.96019
90	129	1	1	81	1		Sim	0.04312	0.95688
91	130	1	1	76	1		Sim	0.04924	0.95076
92	131	0	1	70	1		Sim	0.48724	0.51276
93	132	1	1	78	1		Sim	0.04670	0.95330
94	133	0	1	70	1		Sim	0.48724	0.51276
95	134	0	1	67	1		Não	0.50809	0.49191
96	136	0	1	67	1		Não	0.50809	0.49191
97	138	1	1	89	1		Sim	0.03482	0.96518

Obs	ID	X1	X2	X3	X4	F_X4	I_X4	P_Não	P_Sim
98	139	1	1	26	1		Sim	0.17220	0.82780
99	140	0	0	49	0		Não	0.97683	0.02317
100	143	0	0	53	0		Não	0.97417	0.02583
101	144	0	0	51	0		Não	0.97553	0.02447
102	146	1	0	45	0		Não	0.75211	0.24789

The FREQ Procedure

Frequency	Table of X4 by I_X4			
	X4(Envolvimento nodal)	I_X4(Into: X4)		
		Não	Sim	Total
	Não	54	10	64
	Sim	10	28	38
	Total	64	38	102