# Análise de Dados Amostrais Complexos

*Djalma Pessoa e Pedro Nascimento Silva*

*2016-12-30*

# Contents

# Prefácio

# Chapter 1

# Introdução

# Chapter 2

# Referencial para Inferência

# Chapter 3

# Estimação Baseada no Plano Amostral

# Chapter 4

# Efeitos do Plano Amostral

# Chapter 5

# Ajuste de Modelos Paramétricos

# Chapter 6

# Modelos de Regressão

## 6.1 Modelo de Regressão Linear Normal

O problema considerado nesta seção é o de estimar os parâmetros num modelo de regressão linear normal especificado para um subconjunto das variáveis da pesquisa. O procedimento de máxima pseudo-verossimilhança, descrito na Seção **??**, é aplicado. Os resultados são derivados considerando pesos ordinários dados pelo inverso das probabilidades de inclusão das unidades na amostra. Resultados mais gerais considerando outros tipos de pesos (tais como os derivados de estimadores de razão ou regressão, por exemplo) estão discutidos em Nascimento Silva(1996, cap. 6).

### 6.1.1 Especificação do Modelo

Vamos supor que os dados da $i$-ésima unidade da população pesquisada incluam um vetor $\mathbf{z}_i = (z_{i1}, \ldots, z_{iP})^{'}$ de dimensão $P \times 1$ com os valores de variáveis $\mathbf{z}$, que são `preditoras` ou explanatórias num modelo de regressão $M$. Este modelo tem o objetivo de predizer ou explicar os valores de uma variável da pesquisa $y$, queé considerada como variável `resposta`. Denotemos por $Y_i$ e $\mathbf{Z}_i$ a variável e o vetor aleat'{o}rios que geram $y_i$ e $\mathbf{z}_i$, para $i \in U$. Sem perda de generalidade, suponhamos também que a primeira componente do vetor $\mathbf{z}_i$ de variáveis preditorasé sempre igual a 1, de modo a incluir sempre um termo de intercepto nos modelos de regressão linear considerados (tal hipótese não é essencial, mas será adotada no restante deste capítulo). Suponhamos agora que $\left(Y_i, \mathbf{Z}_i^{'}\right)^{'}$, $i \in U$, são vetores aleatórios independentes e identicamente distribuídos tais que

$$f\left(y_i | \mathbf{z}_i; \beta, \sigma_e\right) = \left(2\pi\sigma_e\right)^{-1/2} \exp\left[-\left(y_i - \mathbf{z}_i^{'}\beta\right)^2 / 2\sigma_e\right] \tag{6.1}$$

onde $\beta = (\beta_1, \ldots, \beta_P)^{'}$ e $\sigma_e > 0$ são parâmetros desconhecidos do modelo.

Observe que (6.1) constitui-se numa especificação (parcial) de um modelo marginal para um conjunto de variáveis da pesquisa, e não faz nenhuma referência direta à forma como elas se relacionam com variáveis auxiliares $\mathbf{x}$ que eventualmente possam estar disponíveis. A atenção é focalizada na estimação de $\beta$ e $\sigma_e$ e sua interpretação com respeito ao modelo agregado (6.1).

Modelos como (6.1) já foram considerados por vários autores, por exemplo Holt, Smith e Winter(1980), Nathan e Holt (1980), Skinner(1989b, p.81), Chambers(1986, 1995). Eles são simples, mesmo assim frequentemente usados pelos analistas de dados, pelo menos como uma primeira aproximação. Além disto, eles satisfazem todas as condições padrões de regularidade. Assim eles são adequados a uma aplicação de procedimentos de máxima pseudo-verossimilhança descritos na Seção @ref{modpar3}.

As funções escores para $\beta$ e $\sigma_e$ correspondentes ao modelo (6.1) podem ser facilmente obtidas como

$$\begin{aligned} \partial \log\left[f\left(y_i|\,\mathbf{z}_i;\beta,\sigma_e\right)\right]/\partial\beta &= \mathbf{z}_i\left(y_i-\mathbf{z}_i'\beta\right)/\sigma_e \\ &\propto \mathbf{z}_i\left(y_i-\mathbf{z}_i'\beta\right)=\mathbf{u}_i\left(\beta\right) \end{aligned} \tag{6.2}$$

e

$$\begin{aligned} \partial \log\left[f\left(y_i|\,\mathbf{z}_i;\beta,\sigma_e\right)\right]/\partial\sigma_e &= \left[\left(y_i-\mathbf{z}_i'\beta\right)^2-\sigma_e\right]/2\sigma_e^2 \\ &\propto \left(y_i-\mathbf{z}_i'\beta\right)^2-\sigma_e=u_i\left(\sigma_e\right)\ . \end{aligned}$$

### 6.1.2 Pseudo-parâmetros do Modelo

Se todos os elementos da população tivessem sido pesquisados, os EMVs de $\beta$ e $\sigma_e$ do censo, denotados por \$B \$ e $S_e$ respectivamente, poderiam ser facilmente obtidos como soluções das equações de verossimilhança do censo dadas por

$$\sum_{i\in U}\mathbf{u}_i\left(\mathbf{B}\right)=\sum_{i\in U}\mathbf{z}_i\left(y_i-\mathbf{z}_i'\beta\right)=\mathbf{z}_U^{'}\mathbf{y}_U-\left(\mathbf{z}_U^{'}\mathbf{z}_U\right)\mathbf{B}=\mathbf{0} \tag{6.3}$$

e

$$\sum_{i\in U}u_i\left(S_e\right)=\sum_{i\in U}\left[\left(y_i-\mathbf{z}_i'\mathbf{B}\right)^2-S_e\right]=\left(\mathbf{y}_U-\mathbf{z}_U'\mathbf{B}\right)^{'}\left(\mathbf{y}_U-\mathbf{z}\mathbf{z}_U'\mathbf{B}\right)-NS_e=0 \tag{6.4}$$

onde $\mathbf{z}_U=\left(\mathbf{z}_1,\ldots,\mathbf{z}_N\right)^{'}$ e $\mathbf{y}_U=\left(y_1,\ldots,y_N\right)^{'}$.

Se $\mathbf{z}_U^{'}\mathbf{z}_U$ for não-singular, as soluções para estas equações são facilmente obtidas como

$$\mathbf{B}=\left(\mathbf{z}_U^{'}\mathbf{z}_U\right)^{-1}\mathbf{z}_U^{'}\mathbf{y}_U \tag{6.5}$$

e

$$S_e=N^{-1}\sum_{i\in U}\left(y_i-\mathbf{z}_i'\mathbf{B}\right)^2=N^{-1}\left(\mathbf{y}_U-\mathbf{z}_U'\mathbf{B}\right)^{'}\left(\mathbf{y}_U-\mathbf{z}_U'\mathbf{B}\right)\ . \tag{6.6}$$

Com uma parametrização que isole o termo correspondente ao intercepto (primeira coluna do vetor $\mathbf{z}_i$) do modelo de regressão (6.1), pode ser facilmente mostrado (Nascimento Silva, 1996, p. 142) que os EMV de $\beta_2$ (igual a $\beta$ excluído o primeiro componente), $\beta_1$ e $\sigma_e$ são dados respectivamente por

$$\mathbf{B}_2=\mathbf{S}_{\mathbf{z}}^{-1}\mathbf{S}_{\mathbf{z}y}\ , \tag{6.7}$$

$$B_1=\bar{Y}-\bar{\mathbf{Z}}^{'}\mathbf{B}_2\ , \tag{6.8}$$

e

$$S_e=N^{-1}\sum_{i\in U}\left(y_i-B_1-\mathbf{z}_i^{'}\mathbf{B}_2\right)^2=N^{-1}\sum_{i\in U}e_i^2\ , \tag{6.9}$$

onde $\bar{Y} = N^{-1} \sum_{i \in U} y_i$, $\bar{\mathbf{Z}} = N^{-1} \sum_{i \in U} \mathbf{z}_i$ , $\mathbf{S_z} = N^{-1} \sum_{i \in U} \left(\mathbf{z}_i - \bar{\mathbf{Z}}\right) \left(\mathbf{z}_i - \bar{\mathbf{Z}}\right)^{'}$, $\mathbf{S_{zy}} = N^{-1} \sum_{i \in U} \left(\mathbf{z}_i - \bar{\mathbf{Z}}\right) \left(y_i - \bar{Y}\right)$ e $e_i = y_i - B_1 - \mathbf{z}_i^{'}\mathbf{B}_2 = \left(y_i - \bar{Y}\right) - \left(\mathbf{z}_i - \bar{\mathbf{Z}}\right)^{'} \mathbf{B}_2$ , sendo neste trecho os vetores de variáveis preditoras tomados sem o termo constante referente ao intercepto.

Os EMVs do censo dados em (6.1) a (6.9) coincidem com os estimadores de mínimos quadrados ordinários, sob as hipóteses mais fracas do modelo dadas por (6.10) a seguir (ver Nathan e Holt, 1980), onde se dispensou a hipótese de normalidade dos erros, isto é

$$\begin{aligned} E_M\left(Y_i | \, \mathbf{z}_i = \mathbf{z}_i\right) &= \beta_1 + \mathbf{z}_i^{'}\beta_2 \\ V_M\left(Y_i | \, \mathbf{z}_i = \mathbf{z}_i\right) &= \sigma_e \\ COV_M\left(Y_i, Y_j | \, \mathbf{z}_i = \mathbf{z}_i, \mathbf{z}_j = \mathbf{z}_j\right) &= 0 \quad \forall i \neq j \in U. \end{aligned} \tag{6.10}$$

### 6.1.3 Estimadores de MPV dos Parâmetros do Modelo

Quando apenas uma amostra de unidades da população é observada, são usados pesos $w_i$ para obter estimadores de máxima pseudo-verossimilhança de $\beta$ e $\sigma_e$, ou alternativamente de $\mathbf{B}$ e $S_e$, se as quantidades descritivas populacionais correspondentes forem escolhidas para alvo da inferência. Se os pesos $w_i$ satisfizerem às condições de regularidade discutidas na Seção @ref{modpar3}, será imediato obter as equações de pseudo-verossimilhança correspondentes ao modelo (6.1) como

$$\begin{aligned} \sum_{i \in s} w_i \mathbf{u}_i \left(\hat{\mathbf{B}}_w\right) &= \sum_{i \in s} w_i \mathbf{z}_i \left(y_i - \mathbf{z}_i' \hat{\mathbf{B}}_w\right) \\ &= \mathbf{z}_s^{'}\mathbf{W}_s\mathbf{y}_s - \left(\mathbf{z}_s^{'}\mathbf{W}_s\mathbf{y}_s\right)\hat{\mathbf{B}}_w = \mathbf{0} \end{aligned} \tag{6.11}$$

e

$$\begin{aligned} \sum_{i \in s} w_i u_i\left(s_e^w\right) &= \sum_{i \in s} w_i \left[\left(y_i - \mathbf{z}_i' \hat{\mathbf{B}}_w\right)^2 - s_e^w\right] \\ &= \left(\mathbf{y}_s - \mathbf{z}_s \hat{\mathbf{B}}_w\right)^{'} \mathbf{W}_s \left(\mathbf{y}_s - \mathbf{z}_s \hat{\mathbf{B}}_w\right) - \left(\mathbf{1}_s^{'}\mathbf{W}_s\mathbf{1}_s\right) s_e^w = 0 \end{aligned} \tag{6.12}$$

onde $\mathbf{z}_s$ e $\mathbf{y}_s$ são os análogos amostrais de $\mathbf{z}_U$ e $\mathbf{y}_U$, respectivamente, \$W _{s}=diag[$(w_{i_1}, \ldots, w_{i_n})$] \$é*umamatrizdiagonaln*$\times$ $n$ com os pesos dos elementos da amostra na diagonal principal, e $\hat{\mathbf{B}}_w$ e $s_e^w$ são estimadores MPV de $\beta$ e $\sigma_e$ respectivamente.

Supondo que $\mathbf{z}_s^{'}\mathbf{W}_s\mathbf{z}_s$ é não-singular e resolvendo (6.11) e (6.12) em $\hat{\mathbf{B}}_w$ e $s_e^w$ obtemos as seguintes expressões para os estimadores MPV dos parâmetros do modelo:

$$\widehat{\mathbf{B}}_w = \left(\mathbf{z}_s^{'}\mathbf{W}_s\mathbf{z}_s\right)^{-1} \mathbf{z}_s^{'}\mathbf{W}_s\mathbf{y}_s \tag{6.13}$$

e

$$\begin{aligned} s_e^w &= \left(\mathbf{1}_s^{'}\mathbf{W}_s\mathbf{1}_s\right)^{-1} \left(\mathbf{y}_s - \mathbf{z}_s\widehat{\mathbf{B}}_w\right)^{'} \mathbf{W}_s \left(\mathbf{y}_s - \mathbf{z}_s\widehat{\mathbf{B}}_w\right) \\ &= \left(\mathbf{1}_s^{'}\mathbf{W}_s\mathbf{1}_s\right)^{-1} \mathbf{y}_s^{'} \left[\mathbf{W}_s - \mathbf{W}_s\mathbf{z}_s\left(\mathbf{z}_s^{'}\mathbf{W}_s\mathbf{z}_s\right)^{-1}\mathbf{z}_s^{'}\mathbf{W}_s\right]\mathbf{y}_s \end{aligned} \tag{6.14}$$

sendo a segunda expressão para $s_e^w$ obtida mediante substituição do valor de $\widehat{\mathbf{B}}_w$ em (6.13) na primeira linha de (6.14).

Observe que a hipótese de não-singularidade de $\mathbf{z}_s^{'}\mathbf{W}_s\mathbf{z}_s$ não seria satisfeita se $ w_{i}=0$ para algum $i \in s$. Para evitar que se percam de vista as questões principais com relação à estimação dos parâmetros do modelo, admitiremos de agora em diante que $\mathbf{z}_s^{'}\mathbf{W}_s\mathbf{z}_s$ é não-singular.

Estimadores pontuais dos parâmetros do modelo podem ser derivados a partir de (6.13) e (6.14) para vários esquemas de ponderação de interesse pela simples substituição da matriz apropriada de ponderação $\mathbf{W}_s$. Se todos os elementos da pesquisa têm o mesmo peso (como no caso de planos amostrais autoponderados), ou seja, $w_i = \bar{w}$ e $\mathbf{W}_s = \bar{w}\mathbf{I}_n$, os estimadores pontuais não dependem do valor $\bar{w}$ dos pesos. Neste caso, eles ficam reduzidos às expressões correspondentes dos estimadores de mínimos quadrados ordinários (que são também estimadores de máxima verossimilhança sob normalidade) dos parâmetros do modelo, dados por:

$$\widehat{\mathbf{B}} = \left(\mathbf{z}_s^{'}\mathbf{z}_s\right)^{-1}\mathbf{z}_s^{'}\mathbf{y}_s \tag{6.15}$$

e

$$s_e = n^{-1}\left(\mathbf{y}_s - \mathbf{z}_s\widehat{\mathbf{B}}\right)^{'}\left(\mathbf{y}_s - \mathbf{z}_s\widehat{\mathbf{B}}\right) \ . \tag{6.16}$$

Substituindo $\mathbf{W}_s$ em (6.13) e (6.14) por $diag\left(\pi_i : i \in s\right) = \mathbf{\Pi}_s^{-1}$, onde os $\pi_i$ em geral não são todos iguais, obtemos estimadores, chamados de mínimos quadrados $\pi-$ponderados, dados por:

$$\widehat{\mathbf{B}}_\pi = \left(\mathbf{z}_s^{'}\mathbf{\Pi}_s^{-1}\mathbf{z}_s\right)^{-1}\mathbf{z}_s^{'}\mathbf{\Pi}_s^{-1}\mathbf{y}_s \tag{6.17}$$

e

$$s_e^\pi = \left(\mathbf{1}_s^{'}\mathbf{\Pi}_s^{-1}\mathbf{1}_s\right)^{-1}\left(\mathbf{y}_s - \mathbf{z}_s\widehat{\mathbf{B}}_\pi\right)^{'}\mathbf{\Pi}_s^{-1}\left(\mathbf{y}_s - \mathbf{z}_s\widehat{\mathbf{B}}_\pi\right) \ . \tag{6.18}$$

### 6.1.4 Estimação da Variância de Estimadores de MPV

O exercício de ajustar um modelo não estará completo sem a avalia ção da precisão e significância das estimativas dos parâmetros. Para isto é necessária a estimação das variâncias correspondentes. Nesta seção concentramos nossa atenção na estimação das variâncias dos estimadores de MPV dos coeficientes de regressão $\beta$. As expressões a seguir são obtidas por aplicação direta dos resultados gerais fornecidos na Seção **??**, observando-se que os escores correspondentes a $\beta$ no `ajuste do censo` do modelo (6.1) são dados por $\mathbf{u}_i\left(\mathbf{B}\right) = \mathbf{z}_i\left(y_i - \mathbf{z}_i'\mathbf{B}\right) = \mathbf{z}_i e_i$ , onde $e_i = \left(y_i - \bar{Y}\right) - \left(\mathbf{z}_i - \bar{\mathbf{Z}}\right)^{'}\mathbf{B}$ para $i \in U$, com o Jacobiano correspondente dado por

$$\begin{aligned} J\left(\mathbf{B}\right) &= \sum\nolimits_{i\in U}\partial\mathbf{z}_i\left(y_i - \mathbf{z}_i'\beta\right)/\partial\beta\Big|_{\beta=\mathbf{B}} \\ &= \partial\left(\mathbf{z}_U'\mathbf{y}_U - \mathbf{z}_U'\mathbf{z}_U\beta\right)/\partial\beta\big|_{\beta=\mathbf{B}} = -\mathbf{z}_U'\mathbf{z}_U \ . \end{aligned} \tag{6.19}$$

Substituindo em (6.7) e (6.8) os valores dos escores, do jacobiano e dos estimadores $\pi$ − *ponderados correspondentes, obtemos as seguintes expressões para a variância assintótica de aleatorização do estimador de MPV* e seu estimador consistente, dadas por

$$V_p\left(\widehat{\mathbf{B}}_\pi\right) = \left(\mathbf{z}_U'\mathbf{z}_U\right)^{-1}V_p\left(\sum_{i\in s}\pi_i^{-1}\mathbf{z}_i e_i\right)\left(\mathbf{z}_U'\mathbf{z}_U\right)^{-1} \tag{6.20}$$

e

$$\hat{V}_p\left(\widehat{\mathbf{B}}_\pi\right) = \left(\mathbf{z}_s'\mathbf{\Pi}_s^{-1}\mathbf{z}_s\right)^{-1} \hat{V}_p\left(\sum_{i\in s}\pi_i^{-1}\mathbf{z}_i e_i\right)\left(\mathbf{z}_s'\mathbf{\Pi}_s^{-1}\mathbf{z}_s\right)^{-1} \ , \tag{6.21}$$

onde

$$V_p\left(\sum_{i\in s}\pi_i^{-1}\mathbf{z}_i e_i\right) = \sum_{i\in U}\sum_{j\in U}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}e_i\mathbf{z}_i\mathbf{z}_j' e_j \ , \tag{6.22}$$

$$\hat{V}_p\left(\sum_{i\in s}\pi_i^{-1}\mathbf{z}_i\hat{e}_i\right) = \sum_{i\in s}\sum_{j\in s}\left(\pi_i^{-1}\pi_j^{-1}-\pi_{ij}^{-1}\right)\hat{e}_i\mathbf{z}_i\mathbf{z}_j'\hat{e}_j \ , \tag{6.23}$$

e $\hat{e}_i = y_i - \mathbf{z}_i'\widehat{\mathbf{B}}_\pi$ para $i \in s$.

Isto completa a especificação de um procedimento de máxima pseudo-verossimilhança para ajustar modelos normais de regressão como (6.1). Este procedimento é bastante flexível e aplicável numa ampla gama de planos amostrais.

## 6.2 Modelo de Regressão Logística

No modelo de regressão logística, a variável resposta $y$ é binária, isto é, assume os valores 0 e 1. Considerando um vetor $\mathbf{z}$ de variáveis explanatórias tal como o empregado no modelo de regressão linear discutido na Seção 6.1, o modelo de superpopulação é dado por

$$f(y_i|\mathbf{z}_i,\beta) = \left[p\left(\mathbf{z}_i'\beta\right)\right]^{y_i}\left[1-p\left(\mathbf{z}_i'\beta\right)\right]^{1-y_i} , \tag{6.24}$$

onde,

$$p\left(\mathbf{z}_i'\beta\right) = P\left(Y_i = 1|\,\mathbf{Z}_i = \mathbf{z}_i\right) = \exp\left(\mathbf{z}_i'\beta\right)/\left[1+\exp\left(\mathbf{z}_i'\beta\right)\right] \ .$$

A função escore de $\beta$ é

$$\mathbf{u}_i\left(\beta\right) = \partial\log(y_i|\mathbf{z}_i,\beta)/\partial\beta = \left[y_i - p\left(\mathbf{z}_i'\beta\right)\right]\mathbf{z}_i \tag{6.25}$$

e portanto a equação de verossimilhança do censo correspondente é dada por

$$\sum\nolimits_{i\in U}\mathbf{u}_i\left(\beta\right) = \sum\nolimits_{i\in U}\left[y_i - p\left(\mathbf{z}_i'\beta\right)\right]\mathbf{z}_i = \mathbf{0} \ . \tag{6.26}$$

O estimador de MPV do vetor de coeficientes $\beta$ no modelo (6.24) é a solução da equação

$$\sum\nolimits_{i\in s}w_i\mathbf{u}_i\left(\beta\right) = \sum\nolimits_{i\in s}w_i\left[y_i - p\left(\mathbf{z}_i'\beta\right)\right]\mathbf{z}_i = \mathbf{0}, \tag{6.27}$$

onde $w_i$ é o peso da $i$-ésima observação amostral.

A matriz de covariância do estimador de MPV de $\beta$ pode ser obtida conforme indicado na Seção **??**, bastando substituir os valores dos escores $\mathbf{u}_i\left(\beta\right) = \left[y_i - p\left(\mathbf{z}_i'\beta\right)\right]\mathbf{z}_i$ e do jacobiano correspondentes. Para maiores detalhes, o leitor interessado pode consultar Binder(1983), que aborda o problema da estimação da matriz

Table 6.1: Descrição das variáveis explicativas

| **Fatores** | **Níveis** | **Descrição dos níveis** |
|---|---|---|
| Sexo (sx) | sx(1) | Homens |
| | sx(2) | Mulheres |
| Anos de estudo (ae) | ae(1) | Até 4 |
| | ae(2) | De 5 a 8 |
| | ae(3) | 9 ou mais |
| Horas trabalhadas (ht) | ht(1) | Menos de 40 |
| | ht(2) | De 40 a 48 |
| | ht(3) | Mais de 48 |
| Idade em anos completos (id) | id(1) | Até 17 |
| | id(2) | De 18 a 25 |
| | id(3) | De 26 a 49 |
| | id(4) | 50 ou mais |
| Rendimento médio mensal (re) | re(1) | Menos de 1 |
| | re(2) | De 1 a 5 |
| | re(3) | Mais de 5 |

de covariância dos estimadores de MPV na família de modelos lineares generalizados, da qual o modelo de regressão logística é caso particular.

Vale observar que, tal como no caso da modelagem clássica, a obtenção dos estimadores de MPV dos parâmetros no modelo de regressão logística depende da solução por métodos numéricos de um sistema de equações. Portanto é importante dispor de um pacote computacional adequado para efetuar os cálculos. Hoje em dia já estão disponíveis vários pacotes com essa funcionalidade, conforme se discute no Capítulo 10.

Análise do perfil sócio-econômico das pessoas ocupadas no setor informal da economia na área urbana do Rio de Janeiro

Utilizando dados do Suplemento Trabalho da Pesquisa Nacional por Amos-tra de Domicílios (**PNAD**) de 90, Leote(1996) analisou o perfil sócio-econômico das pessoas ocupadas no setor informal da economia na área urbana do Rio de Janeiro.

Os dados utilizados são relativos a pessoas que:

- moravam em domicílios urbanos do estado do Rio de Janeiro;
- trabalhavam em atividades mercantis (não foram incluídos trabalhadores domésticos);
- na semana da pesquisa estavam trabalhando ou não estavam trabalhando por estarem de férias, licença, etc., mas tinham trabalho;
- desenvolviam atividades não agrícolas.

As pessoas que trabalhavam em locais com até cinco pessoas ocupadas foram classificadas no setor informal, independente da posição de ocupação delas, enquanto as que trabalhavam em locais com mais de cinco pessoas ocupadas foram classificadas no setor formal. O trabalho refere-se ao trabalho principal. Para a variável renda considerou-se a soma dos rendimentos de todos os trabalhos.

Foi considerada uma amostra de 6.507 pessoas (após a exclusão de 9 registros considerados atípicos), classificadas de acordo com as variáveis descritas na Tabela 6.1, todas tratadas como fatores na análise. A variável ht foi considerada como a soma de horas trabalhadas em todos os trabalhos, por semana. A variável re compreende a renda média mensal de todos os trabalhos, em salários mínimos.

| Fatores | Níveis D | escrição dos níveis |
|---|---|---|
| Sexo(sx) | sx(1) sx(2) | Homens Mulheres |
| Anos de estudo(ht) | ae(1) ae(3) | Até 4 ae(2) De 5 a 8 9 ou mais |
| Horas trabalhadas(ht) | ht(1) ht(2) ht(3) | Menos de 40 De 40 a 48 Mais de 48 |
| Idade e anos completos(id) | id(1) id(3) id(4) | Até 17 id(2) De 18 a 25 De 26 a 49 50 ou mais |
| Rendimento médio mensal(re) | re(1) re(2) re(3) | Menos de 1 De 1 a 5 Mais de 5 |

Os fatores considerados foram tomados como explicativos e a variável resposta foi o indicador de pertinência ao setor informal da economia. Foi ajus-tado um modelo logís-tico (Agresti, 1990) para explicar a probabilidade de uma pessoa pertencer ao setor informal da economia.

Para a seleção do modelo foi usada a função `glm` do **S-Plus**, aplicada aos dados tabelados. O modelo final selecionado foi escolhido passo a passo, incluindo em cada passo as interações que produziam maior decréscimo do desvio residual, considerando a perda de graus de liberdade. O modelo selecionado foi

$$\begin{aligned} \log\left(\frac{p_{ijklm}}{1-p_{ijklm}}\right) &= \mu + \beta_i^{sx} + \beta_j^{ae} + \beta_k^{ht} + \beta_l^{id} + \beta_m^{re} \\ &\quad + \beta_{ij}^{sx.id} + \beta_{ik}^{sx.ht} + \beta_{jk}^{ae.ht} + \beta_{kl}^{ht.id} + \beta_{km}^{ht.re}, \end{aligned} \tag{6.28}$$

onde $p_{ijklm}$ é a probabilidade de pertencer ao setor informal correspondente à combinação de níveis das variáveis explicativas, sendo i=1, 2 o nível de sx; j=1, 2, 3 o nível de ae; k=1, 2, 3 o nível de ht; l=1, 2, 3, 4 o nível de id e m=1, 2, 3 o nível de re.

Os efeitos foram adicionados sequencialmente na ordem da Tabela 6.1. Depois de introduzidos os efeitos principais, as interações de dois fatores foram introduzidas na ordem definida pela `função step do S-Plus`.

O *p*valor do teste de nulidade das interações não incluídas no modelo é 0,0515, aceitando-se a hipótese de nulidade destes efeitos ao nível $\alpha = 0,05$. O modelo obtido difere do selecionado em Leote(1996) só pela inclusão de mais um efeito, referente à interação ae:ht.

Uma descrição detalhada do plano amostral da PNAD 90 foi apresentada no Exemplo 6.2. Como se pode observar dessa descrição, o plano amostral da PNAD apresenta todos os aspectos de um plano amostral complexo, incluindo estratificação (geográfica), seleção de unidades primárias (municípios, ou setores nos municípios auto-representativos) ou secundárias (setores nos municípios não auto-representativos) com probabilidades desiguais, conglomeração (de domicílios em setores, e de pessoas nos domicílios) e seleção sistemática sem reposição de unidades. Nesse caso, fica difícil admitir a priori com confiança as hipóteses usuais de modelagem das observações amostrais como IID. Por esse motivo foram considerados métodos alternativos de modelagem e ajuste.

Apresentamos a seguir as estimativas dos efeitos principais e interações do modelo selecionado e seus respectivos desvios padrões, calculadas pela **PROC LOGISTIC** do pacote **SUDAAN**. Para facilitar a comparação incluímos na Tabela 6.3 os valores correspondentes estimados pelo **S-Plus**.

As estimativas calculadas pelo pacote **SUDAAN** são feitas pelo `Método de Máxima Pseudo-Verossimilhança`, resolvendo a equação (6.27). As estimativas dos desvios padrões são obtidas das variâncias calculadas pelo método de linearização descrito na Seção **??**, equação (**??**), considerando os escores tal como apresentados na equação (6.25). Para esses cálculos, os estimadores de variância considerados levaram em conta os pesos das observações, mas utilizaram uma aproximação que consiste em considerar que as unidades primárias de

Table 6.3: Estimativas dos efeitos e dos respectivos desvios padrões obtidas pelo **SUDAAN** e pelo **S-Plus**

| Variáveis independentes e efeitos | Ajuste no **SUDAAN** | | Ajuste no **S-Plus** | |
|---|---|---|---|---|
| | Estimativa do efeito | Desvio Padrão | Estimativa do efeito | Desvio Padrão |
| Intercepto | $-0,515$ | $0,260$ | $-0,514$ | $0,269$ |
| sx | $0,148$ | $0,222$ | $0,156$ | $0,228$ |
| ae1 | $0,745$ | $0,165$ | $0,740$ | $0,165$ |
| ae2 | $0,496$ | $0,156$ | $0,497$ | $0,159$ |
| ht1 | $-0,377$ | $0,317$ | $-0,386$ | $0,312$ |
| ht2 | $-0,697$ | $0,275$ | $-0,698$ | $0,268$ |
| id1 | $-0,239$ | $0,540$ | $-0,243$ | $0,492$ |
| id2 | $-0,729$ | $0,302$ | $-0,724$ | $0,314$ |
| id3 | $0,227$ | $0,231$ | $0,227$ | $0,234$ |
| re1 | $0,286$ | $0,277$ | $0,293$ | $0,245$ |
| re2 | $0,065$ | $0,144$ | $0,062$ | $0,145$ |
| ht1.re1 | $1,529$ | $0,356$ | $1,531$ | $0,332$ |
| ht2.re1 | $0,338$ | $0,320$ | $0,336$ | $0,284$ |
| ht1.re2 | $0,490$ | $0,233$ | $0,498$ | $0,221$ |
| ht2.re2 | $-0,115$ | $0,183$ | $-0,112$ | $0,178$ |
| ht1.id1 | $-1,420$ | $0,605$ | $-1,408$ | $0,515$ |
| ht2.id1 | $-0,413$ | $0,506$ | $-0,397$ | $0,465$ |
| ht1.id2 | $-0,124$ | $0,354$ | $-0,129$ | $0,351$ |
| ht2.id2 | $-0,109$ | $0,279$ | $-0,106$ | $0,286$ |
| ht1.id3 | $-0,220$ | $0,248$ | $-0,216$ | $0,253$ |
| ht2.id3 | $-0,537$ | $0,205$ | $-0,533$ | $0,201$ |
| sx.id1 | $0,878$ | $0,348$ | $0,870$ | $0,335$ |
| sx.id2 | $0,300$ | $0,231$ | $0,294$ | $0,226$ |
| sx.id3 | $-0,259$ | $0,190$ | $-0,263$ | $0,186$ |
| sx.ht1 | $-0,736$ | $0,206$ | $-0,737$ | $0,211$ |
| sx.ht2 | $-0,089$ | $0,185$ | $-0,093$ | $0,182$ |
| ae1.ht1 | $0,792$ | $0,240$ | $0,792$ | $0,239$ |
| ae2.ht1 | $0,739$ | $0,227$ | $0,735$ | $0,226$ |
| ae1.ht2 | $0,026$ | $0,197$ | $0,029$ | $0,196$ |
| ae2.ht2 | $0,089$ | $0,183$ | $0,087$ | $0,189$ |

amostragem foram selecionadas com reposição, especificando a opção WR do pacote SUDAAN. Veja (Shah et al., 1992), p. 4) e (Wolter, 1985), eq. 7.7.2.

Na Tabela 6.4 são apresentadas as probabilidades de significância dos tes-tes de nulidade dos efeitos do modelo. Todos os efeitos incluídos no modelo são significativos, nos níveis usuais de significância. A **PROC LOGISTIC** do pacote **SUDAAN** não inclui testes para os efeitos principais, por não ser possível separar tais efeitos das interações. A coluna de $p$ valores da Tabela 6.4, obtida pela PROC LOGISTIC do pacote SUDAAN, utiliza a estatística de Wald baseada no plano amostral com correção. Mais detalhes são encontrados em Shah et al.(1993).

Os testes da Tabela 6.4 indicam a significância de todas as interações de 2 fatores que entraram no modelo selecionado. O teste de qualidade global de ajuste, na primeira linha da Tabela 6.4, indica a necessidade de serem introduzidas novas interações.

Table 6.4: Testes de hipóteses de nulidade dos efeitos do modelo

| Contraste | Graus de liberdade | Graus de liberdade ajustados | Estatística F ajustada | *p*valor da estatística F ajustada |
|---|---|---|---|---|
| Modelo Global | 30 | 26,132 | 37,510 | 0,000 |
| Bondade do ajuste | 29 | 25,692 | 28,179 | 0,000 |
| ht:re | 4 | 3,946 | 6,040 | 0,000 |
| ht:id | 6 | 5,764 | 4,110 | 0,001 |
| sx:id | 3 | 2,969 | 7,168 | 0,000 |
| sx:ht | 2 | 1,993 | 9,166 | 0,000 |
| ae:ht | 4 | 3,959 | 4,814 | 0,001 |

Table 6.5: Estimativas das razões de vantagens, variando-se os níveis de ae para níveis fixos de ht

| ht | Mudança de nível de ae | **S-Plus** | **SUDAAN** |
|---|---|---|---|
| 1 | 1 para 2 | 0,741 | 0,739 |
| 1 | 2 para 3 | 0,291 | 0,291 |
| 2 | 1 para 2 | 0,831 | 0,830 |
| 2 | 2 para 3 | 0,558 | 0,557 |
| 3 | 1 para 2 | 0,785 | 0,780 |
| 3 | 2 para 3 | 0,608 | 0,608 |

Para comparação, apresentamos na Tabela 6.5 algumas estimativas de razões de vantagens, relevantes na análise, calculadas tanto pela função glm do **S-Plus** como pela **PROC LOGISTIC** do pacote **SUDAAN** e, na Tabela 6.6**,** os correspondentes intervalos de confiança de 95%. Na construção destes intervalos foi necessário utilizar estimativas pontuais dos efeitos bem como a matriz de covariância estimada dos estimadores dos efeitos do modelo. Deste modo, estes intervalos sumarizam, ao mesmo tempo, discrepâncias existentes tanto nas estimativas pontuais dos efeitos como nas variâncias e covariâncias das estimativas.

Além dos ajustes aqui comparados, foram feitos (embora não apresentados) os seguintes ajustes com a utilização do **S-Plus**: 1) dados individuais (resposta 0-1) considerando os pesos; 2) dados da tabela estimada considerando os pesos e 3) dados individuais com pesos normalizados. Em todas estas análises, como esperado, as estimativas pontuais dos efeitos coincidiram com as obtidas pela **PROC LOGISTIC** do pacote **SUDAAN**. Pode-se notar que, neste exemplo, há estreita concordância entre as estimativas pontuais obtidas

Table 6.6: Intervalos de confiança de 95% para razões de vantagens, variando-se os níveis de ae para níveis fixos de ht

| ht | Mudança de nível de ae | **S-Plus** | **SUDAAN** |
|---|---|---|---|
| 1 | 1 para 2 | (0,530; 1,036). | (0,516; 1,059) |
| 1 | 2 para 3 | (0,213; 0,399) | (0,212; 0,398) |
| 2 | 1 para 2 | (0,697; 0,991) | (0,693; 0,994) |
| 2 | 2 para 3 | (0,457; 0,680) | (0,452; 0,687) |
| 3 | 1 para 2 | (0,586; 1,050) | (0,577; 1,053) |
| 3 | 2 para 3 | (0,445; 0,831) | (0,448; 0,827) |

Table 6.7: Distribuição de freqüências dos pesos da amostra da PNAD-90 - Parte Urbana do Rio de Janeiro

| Valor do peso | Freqüência |
|---|---|
| 674 | 127 |
| 675 | 784 |
| 711 | 3288 |
| 712 | 712 |

pelos dois pacotes.

A concordância das estimativas dos coeficientes pode ser explicada, em parte, pela pequena variabilidade dos pesos das unidades, tal como se pode verificar na Tabela 6.7, que apresenta a distribuição de freq"{u}ências dos pesos.

Como foi visto na Tabela 6.3, o impacto do plano amostral nas estimativas de precisão é um pouco maior. As maiores diferenças entre os dois métodos ocorrem na estimação dos desvios das estimativas dos parâmetros do primeiro nível de idade (até 17 anos) e da interação deste com horas trabalhadas (tanto no nível de menos de 40 horas semanais como no nível de 40 a 48 horas semanais trabalhadas). Esta diferenciação maior no caso dos desvios padrões já era esperada. Quando não levamos em conta os pesos nem o plano amostral na estimação dos parâmetros, podemos até chegar em uma estimativa pontual dos coeficientes bem próxima de quando levamos ambos em conta, mas as estimativas dos desvios padrões são mais sensíveis a esta diferença entre as análises. A tendência revelada é de subestimação dos desvios padrões pelo **S-Plus** ao ignorar o plano amostral e a variação dos pesos.

Neste exemplo, foi utilizada a função glm do **S-Plus** na seleção do modelo. Feita a seleção, o mesmo modelo foi ajustado através da **PROC LOGISTIC** do **SUDAAN**. O propósito foi imitar uma situação onde o modelo já tivesse sido selecionado e ajustado por usuário secundário dos dados, sem considerar os pesos e o plano amostral, tal como é usual. Outra possibilidade seria repetir o processo de seleção do modelo usando-se a **PROC LOGISTIC** do **SUDAAN**. Isto poderia ser feito passo a passo, incluindo efeitos e interações que melhorassem mais a qualidade de ajuste, tal como foi feito automaticamente pela função **step do S-Plus**. Este procedimento possibilitaria comparar a seleção de modelos quando são considerados os pesos e o plano amostral na análise.

Diferentemente dos pacotes mais usados de análise estatística, tais como SAS, S-Plus, BMDP, etc., o SUDAAN não possui, atualmente, ferramentas usuais de diagnóstico de ajuste de modelos, como gráficos de resíduos padronizados, etc., tornando mais difícil seu uso na etapa de seleção de modelos. Considerando-se a maior dificuldade de seleção de modelos através do **SUDAAN**, preferiu-se usá-lo aqui apenas para ajustar um modelo já selecionado.

## 6.3 Teste de Hipóteses

Nas secões 6.1 e 6.2 discutimos formas de introduzir pesos e plano amostral em procedimentos de estimação pontual e de variâncias ao ajustar modelos com dados de pesquisas amostrais complexas. Neste contexto, procedimentos estatísticos de teste de hipóteses devem, também, sofrer adaptações. Nesta seção, esse problema será abordado de forma sucinta, para modelos de regressão.

De modo geral, testes de hipóteses em regressão surgem inicialmente na seleção de modelos e também para fornecer evidência favorável ou contrária a indagações levantadas pelo pesquisador.

Denotemos por $\beta = (\beta_1, \ldots, \beta_P)'$ o vetor de parâmetros num modelo de regressão. Como é sabido, para testar a hipótese $H_0 : \beta_j = 0$, para algum $j \in \{1, \ldots, P\}$, usamos um teste $t$, e para para testar a hipótese $H_0 : (\beta_{j_1}, \ldots, \beta_{j_R})' = \mathbf{0}_R$, onde $(j_1, \ldots, j_R) \subset (1, \ldots, P)$ e $\mathbf{0}_R$ é o vetor zero $R$-dimensional, usamos um teste

**F**. Tais testes $t$ e **F**, sob as hipóteses do modelo clássico de regressão com erros normais, são testes da Razão de Máxima Verossimilhança.

é pois natural tentar adaptar testes de Razão de Máxima Verossimilhança para pesquisas amostrais complexas, tal como foi feito na derivação de estimadores de MPV a partir de estimadores de Máxima Verossimilhança. A principal dificuldade é que no contexto de pesquisas complexas, devido aos pesos distintos das observações e ao plano amostral utilizado, a função de verossimilhança usual não representa a distribuição conjunta das observações. Apesar desta dificuldade ter sido contornada na derivação de estimadores de MPV, a adaptação fica bem mais difícil no caso de testes da Razão de Máxima Verossimilhança.

Por essa causa, é mais fácil basear os testes na estatística Wald, que mede a distância entre uma estimativa pontual e o valor hipotético do parâmetro numa métrica definida pela matriz de covariância do estimador. Pesos e plano amostral podem ser incorporados facilmente nessa estatística, bastando para isto utilizar estimativas apropriadas (consistentes sob aleatorização) dos parâmetros e da matriz de covariância, tais como as que são geradas pelo método de MPV. é essa abordagem que vamos adotar aqui.

Considere o problema de testar a hipótese linear geral

$$H_0 : \mathbf{C}\beta = \mathbf{c}, \tag{6.29}$$

onde $\mathbf{C}$ é uma matriz de dimensão $R \times P$ de posto pleno $R = P - Q$ e $\mathbf{c}$ é um vetor $R \times 1$.

Um caso particular de interesse é testar a hipótese aninhada $H_0 : \beta_2 = \mathbf{0}_R$, onde $\beta' = (\beta_1', \beta_2')$ , com $\beta_1$ de dimensão $Q \times 1$ e $\beta_2$ de dimensão $R \times 1$, \$C=[

$\mathbf{0}_{R\times Q}$ $\mathbf{I}_R$

]\$$ec = \mathbf{0}_R$ , sendo $\mathbf{0}_{R\times Q}$ matriz de zeros de dimensão $R \times Q$ e $\mathbf{I}_R$ a matriz identidade de ordem $R$.

A estatística de Wald clássica para testar a hipótese nula (6.29) é definida por

$$X_W^2 = \left(\mathbf{C}\widehat{\beta} - \mathbf{c}\right)' \left(\mathbf{C}\widehat{\mathbf{V}}\left(\hat{\beta}\right)\mathbf{C}'\right)^{-1} \left(\mathbf{C}\widehat{\beta} - \mathbf{c}\right), \tag{6.30}$$

onde os estimadores $\widehat{\beta}$ e $\widehat{\mathbf{V}}\left(\hat{\beta}\right)$ são obtidos pela teoria de mínimos quadrados ordinários. Sob $H_0$, a distribuição assintótica da estatística $X_W^2$ é $\chi^2(R)$.

Quando os dados são obtidos através de pesquisas amostrais complexas, a estatística $X_W^2$ deixa de ter distribuição assintótica $\chi^2(R)$, e usar esta '{u}ltima como distribuição de referência implica na obtenção de testes com níveis de significância incorretos. Esse problema é solucionado substituindo-se na expressão de $X_W^2$, $\hat{\beta}$ pela estimativa MPV $\widehat{\mathbf{B}}_\pi$ de $\beta$ dada em (6.17), e $\widehat{\mathbf{V}}\left(\hat{\beta}\right)$pela estimativa da matriz de covariância do estimador de MPV $\hat{V}_p\left(\widehat{\mathbf{B}}_\pi\right)$ dada em (6.21). Tais estimativas consideram os pesos diferentes das observações e o plano amostral efetivamente utilizado. A normalidade assintótica do estimador de MPV de $\beta$ e a consistência do estimador da matriz de covariância correspondente (Binder, 1983) implicam que

$$X_W^2 \sim \chi^2(R), \text{ sob } H_0.$$

Esta aproximação não leva em conta o erro amostral na estimação de $\mathbf{V}\left(\hat{\beta}\right)$. Uma alternativa é usar a aproximação

$$X_W^2/R \sim \mathbf{F}(R, \upsilon),$$

onde $\upsilon = m - H$ é o número de UPAs da amostra menos o n'{u}mero de estratos considerados no plano amostral para seleção das UPAs, que fornece uma medida de graus de liberdade apropriada para amostras complexas quando o método do conglomerado primário é empregado para estimar variâncias.

Com a finalidade de melhorar a aproximação da distribuição da estatística de teste, podem ser utilizados ajustes e correções da estatística $X^2_W$, que são apresentados com mais detalhes nos Capítulos 7 e 8 para o caso da análise de dados categóricos.

A especificação de um procedimento para testar hipóteses sobre os parâmetros de um modelo de regressão completa a abordagem para ajuste de modelos desse tipo partindo de dados amostrais complexos. Entretanto, uma das partes importantes da teoria clássica para modelagem é a que trata do diagnóstico dos modelos ajustados, muitas vezes empregando recursos gráficos. Nessa parte a abordagem baseada em MPV e em estatísticas de Wald deixa a desejar, pois não é possível adaptar de maneira simples as técnicas clássicas de diagnóstico. Por exemplo, é difícil considerar pesos ao plotar os resíduos do ajuste dum modelo via MPV. Essa é questão que ainda merece maior investigação e por enquanto é uma desvantagem da abordagem aqui preconizada.

## 6.4 Laboratório de R

Usar exemplo da amolim ou conseguir exemplo melhor? Reproduzir usando a survey os resultados do Exemplo 6.1???

```
library(survey)
```

```
## Loading required package: grid
```

```
## Loading required package: Matrix
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survey'
```

```
## The following object is masked from 'package:graphics':
##
##     dotchart
```

```
source("~\\GitHub\\adac\\data\\pnadrj90.R")
names(pnadrj90)
```

```
##  [1] "stra"     "psu"      "pesopes"  "informal" "sx"       "id"
##  [7] "ae"       "ht"       "re"       "um"
```

Preparação dos dados: Variáveis explicativas são fatores. Ver tipo de variável:

```
unlist(lapply(pnadrj90, mode))
```

```
##      stra       psu   pesopes  informal        sx        id        ae
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##        ht        re        um
## "numeric" "numeric" "numeric"
```

Transformar variáveis para fatores e mudar o nível básico do fator (último)

```
pnadrj90$sx<-as.factor(pnadrj90$sx)
pnadrj90$sx<-relevel(pnadrj90$sx,ref="2")
pnadrj90$id<-as.factor(pnadrj90$id)
pnadrj90$id<-relevel(pnadrj90$id,ref="4")
pnadrj90$ae<-as.factor(pnadrj90$ae)
pnadrj90$ae<-relevel(pnadrj90$ae,ref="3")
pnadrj90$ht<-as.factor(pnadrj90$ht)
pnadrj90$ht<-relevel(pnadrj90$ht,ref="3")
```

```
pnadrj90$re<-as.factor(pnadrj90$re)
pnadrj90$re<-relevel(pnadrj90$re,ref="3")
##transformar variável de resposta para 0,1:
pnadrj90$informal<-ifelse(pnadrj90$informal==1,1,0)
```

Cria objeto de desenho

```
pnad.des<-svydesign(id=~psu,strata=~stra,weights=~pesopes,data=pnadrj90,nest=TRUE)
```

Ajusta modelo de regressão logística na Tabela 6.2 Comparar resultado com o da página 106 de Pessoa e Silva (1998)

```
inf.logit<-svyglm(informal~sx+ae+ht+id+re+sx*id+sx*ht+ae*ht+ht*id+ht*re,design=pnad.des,family=binomial
```

```
## Warning: non-integer #successes in a binomial glm!
```

```
library(xtable)
print(xtable(inf.logit, digits= c(0,3,3,2,3)),type="html")
```

Estimate

Std. Error

t value

Pr(>|t|)

(Intercept)

-0.515

0.260

-1.98

0.048

sx1

0.148

0.222

0.67

0.506

ae1

0.745

0.165

4.53

0.000

ae2

0.496

0.156

3.18

0.002

ht1

-0.377

0.317

-1.19

0.236

ht2

-0.697

0.275

-2.53

0.012

id1

-0.239

0.540

-0.44

0.659

id2

-0.729

0.302

-2.41

0.016

id3

0.227

0.231

0.98

0.327

re1

0.286

0.277

1.03

0.302

re2

0.065

0.144

0.45

0.652

sx1:id1

0.878

0.348

2.52

0.012

sx1:id2

0.300

0.231

1.30

0.195

sx1:id3

-0.259

0.190

-1.36

0.173

sx1:ht1

-0.736

0.206

-3.57

0.000

sx1:ht2

-0.089

0.185

-0.48

0.631

ae1:ht1

0.792

0.240

3.29

0.001

ae2:ht1

0.739

0.227

3.26

0.001

ae1:ht2

0.026

0.197

0.13

0.895

ae2:ht2

0.089

0.183

0.49

0.626

ht1:id1

-1.420

0.605

-2.35

0.019

ht2:id1

-0.413

0.506

-0.82

0.414

ht1:id2

-0.124

0.355

-0.35

0.726

ht2:id2

-0.109

0.279

-0.39

0.696

ht1:id3

-0.220

0.248

-0.89

0.375

ht2:id3

-0.537

0.205

-2.62

0.009

ht1:re1

1.529

0.356

4.29

0.000

ht2:re1

0.338

0.320

1.06

0.292

ht1:re2

0.490

0.233

2.10

0.036

ht2:re2

-0.115

0.183

-0.63

0.530

Teste de Wald para a hipótese $H_0 : ht : re = 0$

```
regTermTest(inf.logit,"ht:re")
```

```
## Wald test for ht:re
##  in svyglm(formula = informal ~ sx + ae + ht + id + re + sx * id +
##     sx * ht + ae * ht + ht * id + ht * re, design = pnad.des,
##     family = binomial)
## F =  6.742662  on  4  and  616  df: p= 2.58e-05
```

# Chapter 7

# Testes de Qualidade de Ajuste

# Chapter 8

# Testes em Tabelas de Duas Entradas

# Chapter 9

# Agregação vs. Desagregação

# Chapter 10

# Pacotes para Analisar Dados Amostrais

# Chapter 11

# Placeholder

# Bibliography

Shah, B. V., Barnwell, B. G., Hunt, P. N., and LaVange, L. M. (1992). *SUDAAN User's Manual - Professional Software for SUrvey ANalysis for multistage sample designs - realease 6.0.* Research Triangle Park, NC:Research Triangle Institute.

Wolter, K. M. (1985). *Introduction to Variance Estimation.* Springer-Verlag, Nova Iorque.