

Aula 9:

'Que factores explicam a insatisfação com as condições de trabalho?'

Estimação de modelos de Regressão Logística

Docente: Daniela Craveiro
dcraveiro@iseg.ulisboa.pt

E se a nossa variável dependente não é uma variável contínua?

Existe um conjunto de opções ...

Sendo que a escolha do modelo de regressão mais adequado dependerá da natureza da variável dependente e (em alguns casos) da natureza das variáveis dependentes

Variável Dependente	Modelo	Variáveis Independentes
Contínua	Regressão Linear Multipla (OLS)	Todos
Nominal (2 Categorias)	Regressão Logística Regressão Probit	Todos
Nominal (2+ Categorias)	Regressão Multinomial	Todos
Nominal (2 Categorias)	Regressão Loglinear	Nominais
Ordinal	Ordinal Logit Ordered Probit	Todos
Contagens (contém apenas número inteiros, número finito de valores)	Poisson Regression	Contínuas
Tempo até um evento (Análise de Sobrevivência)	Cox Regression	Todos

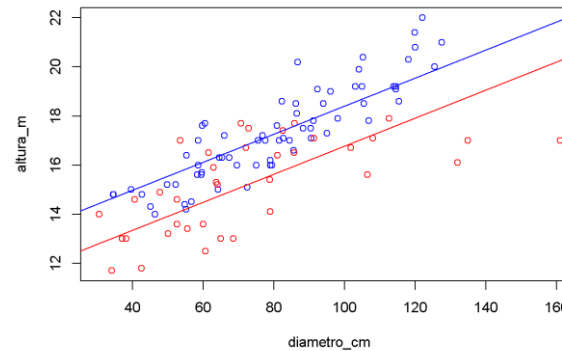
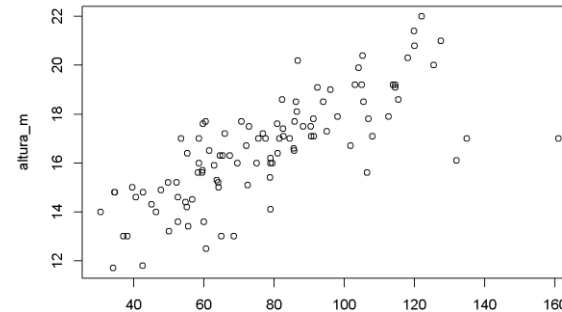
Nesta aula vamos focar-nos num dos modelos de regressão mais comum:

O Modelo de Regressão Logística

Porque é não podemos usar o modelo de regressão linear quando temos uma variável dependente binomial?

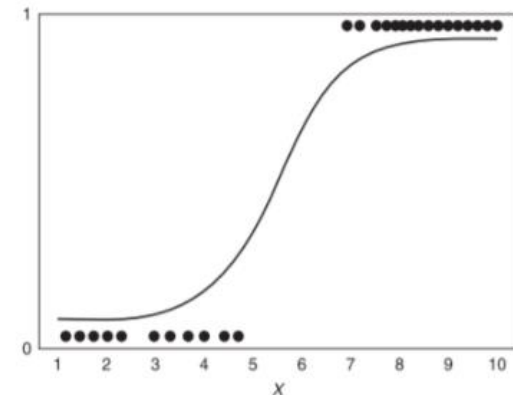
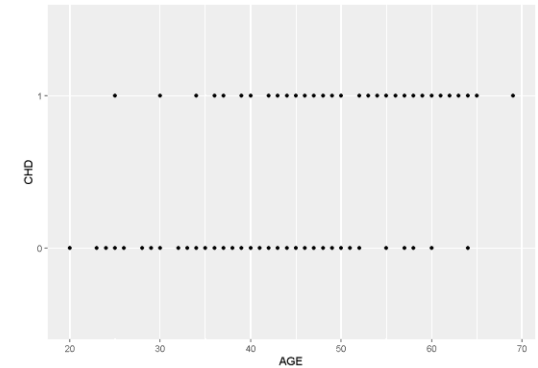
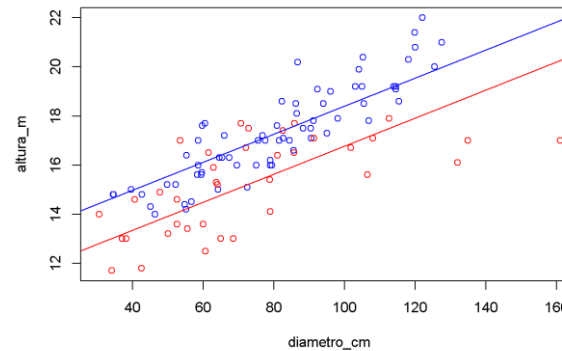
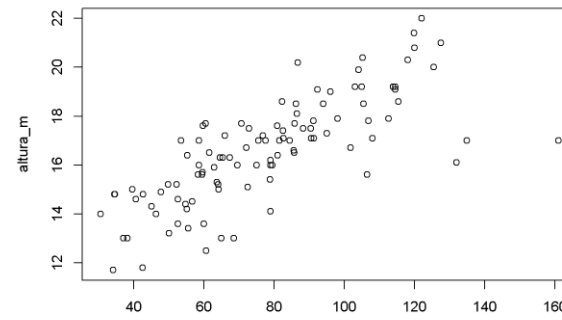
Porque é não podemos usar o modelo de regressão linear quando temos uma variável dependente binomial?

- Numa variável dependente contínua os valores distribuem-se à volta de uma linha (a linha de regressão).



Porque é não podemos usar o modelo de regressão linear quando temos uma variável dependente binomial?

- Numa variável dependente contínua os valores distribuem-se à volta de uma linha (a linha de regressão).
- Ora não é isso que acontece com uma variável dependente binária...
- A regressão logística utiliza a curva logística para assim representar a relação entre a variável dependente e as independentes.
- Esta curva assume que a partir de um dado valor da IV, o valor da DV não se altera



Modelo de regressão linear (simples)

$$Y = \beta_0 + \beta_1 X + E$$

Y	Variável Dependente
β_0	Constante (i.e., valor de Y quando X = 0)
$\beta_1 X$	Coefficiente Beta da variável X (mede o efeito de uma alteração unitária de X sobre o valor médio da variável Y, quando todas as outras variáveis estão fixas)
E	Erro aleatório ou estocástico (reflete a influência de outros factores no comportamento da variável Y que não podem ser explicadas linearmente pelo comportamento da variável X)

Modelo de regressão logística

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 X_1 + \dots + b_n X_n$$

$\ln\left(\frac{p}{1-p}\right)$	Logaritmo natural das chances (odds) de p acontecer
p	Probabilidade de um evento acontecer
$1-p$	Probabilidade de um evento não acontecer
b_0	Constante
$b_1 X_1$	Coefficiente Beta da variável X (mede o efeito de uma alteração unitária de X sobre o logaritmo das odds da variável dependente)

Implementação do Modelo de Regressão Logística

- 1. Criar a variável dependente*
- 2. Inspeccionar a variável dependente*
- 3. Explorar as relações entre variáveis*
- 4. Definir o modelo de regressão logística*
- 5. Interpretar os resultados do modelo de regressão logística*
- 6. Validar pressupostos do modelo de regressão logística*_{próxima aula!}*

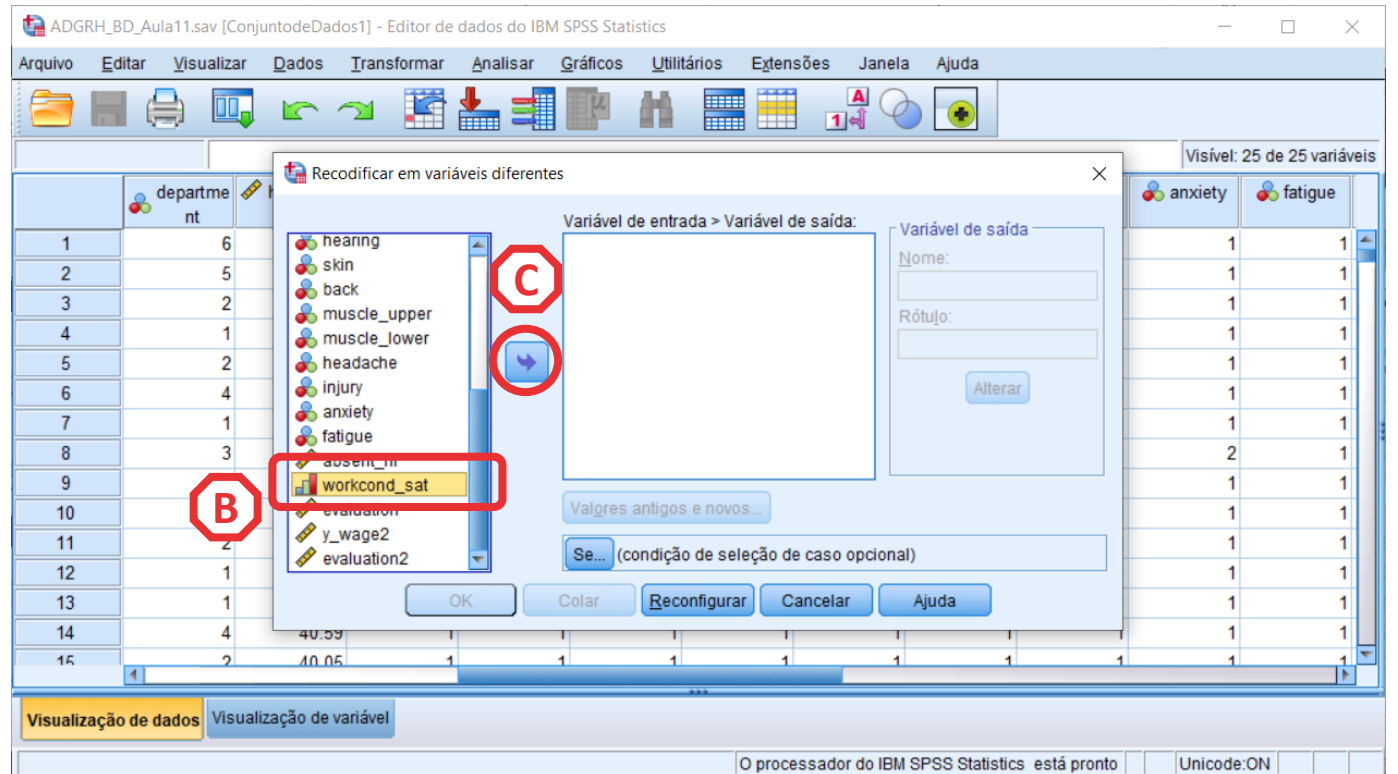
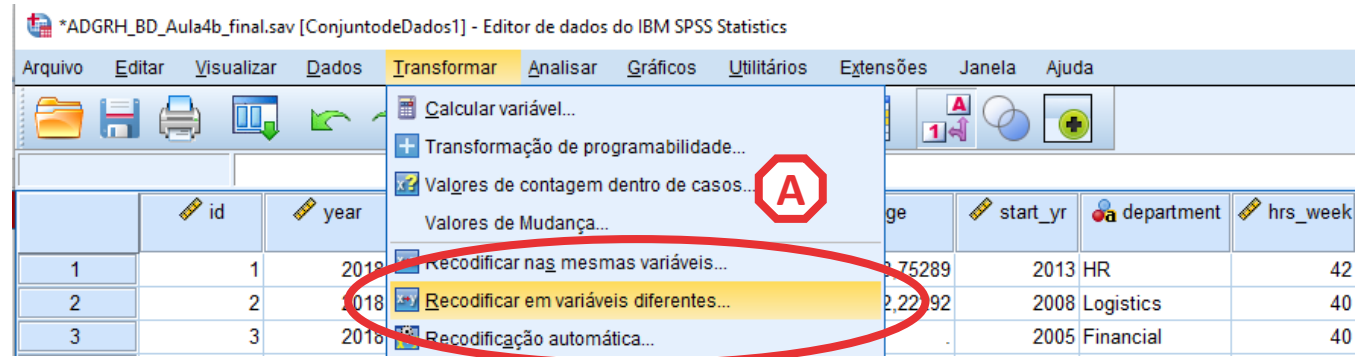
- **Objectivo:**
 - **Recodificar a variável que mede a satisfação com as condições de trabalho ('workcond_sat'), numa variável dummy em que 1 identifica pessoas que responderam 'Very satisfied' (1) ou 'Satisfied' (2).**

Criar a variável dependente

- Selecionar 'Transformar'/'Recodificar em variáveis diferentes' **A**

- Selecionar a variável 'workcond_sat'... **B**

... e colocar na caixa da 'variável de entrada'



Criar a variável dependente

- Seleccionar 'Transformar'/'Recodificar em variáveis diferentes' **A**

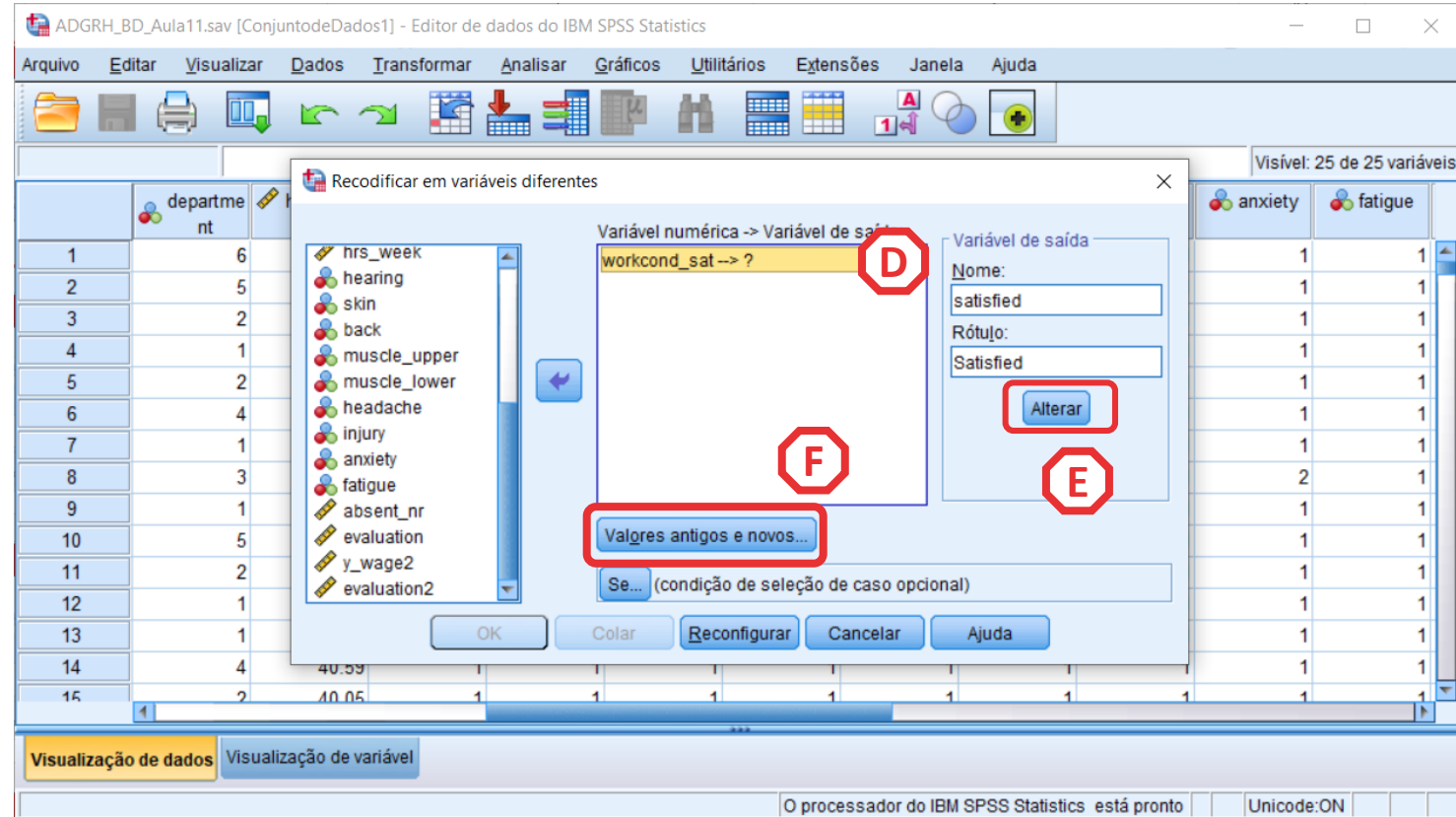
- Seleccionar a variável 'workcond_sat'... **B**

... e colocar na caixa da 'variável de entrada'

- Definir o nome da nova variável ('satisfied') e o rótulo da variável ('Satisfied') **D**

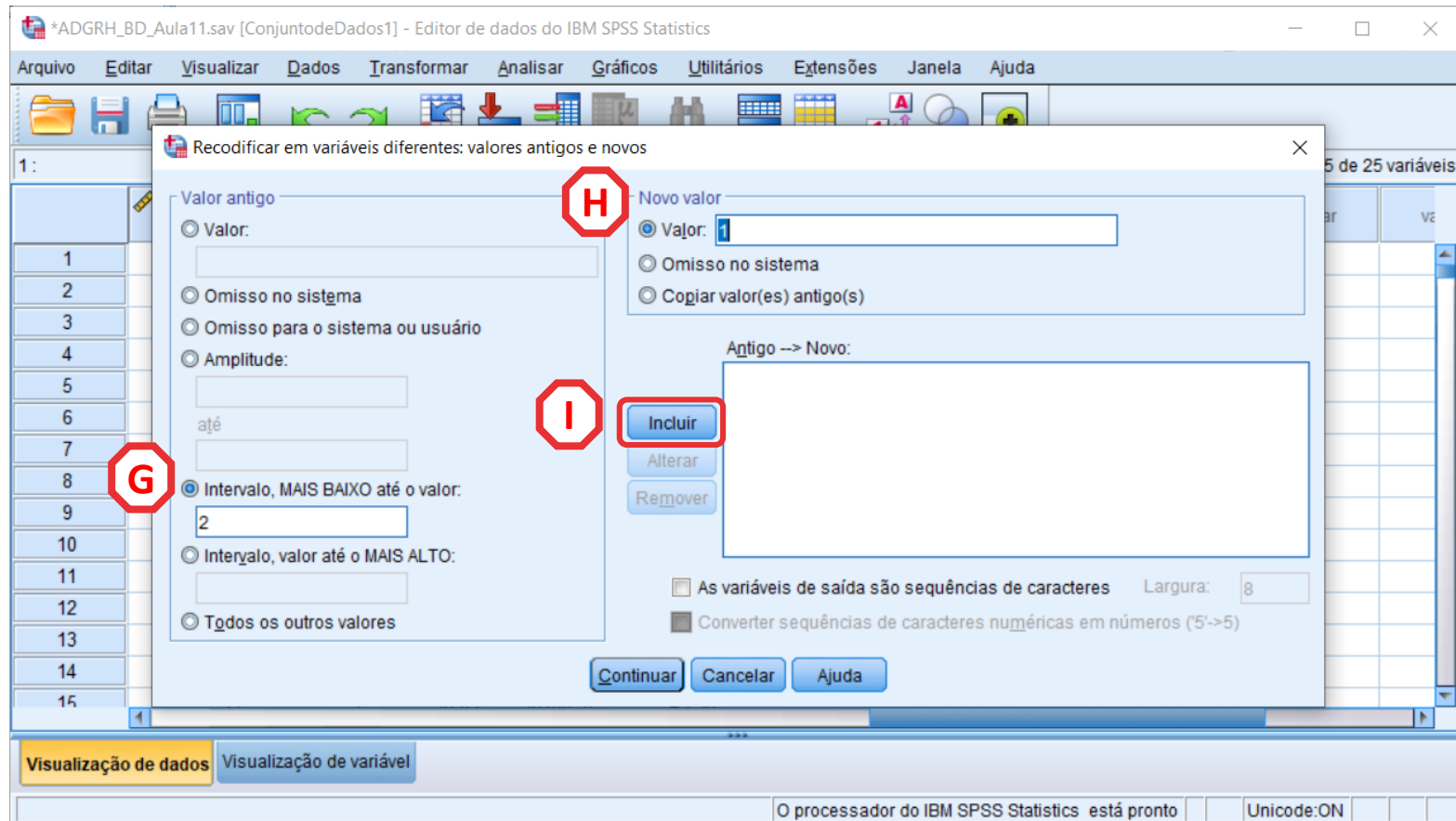
- Seleccionar o botão 'Alterar' **E**

- Seleccionar 'Valores antigo e novo' **F**



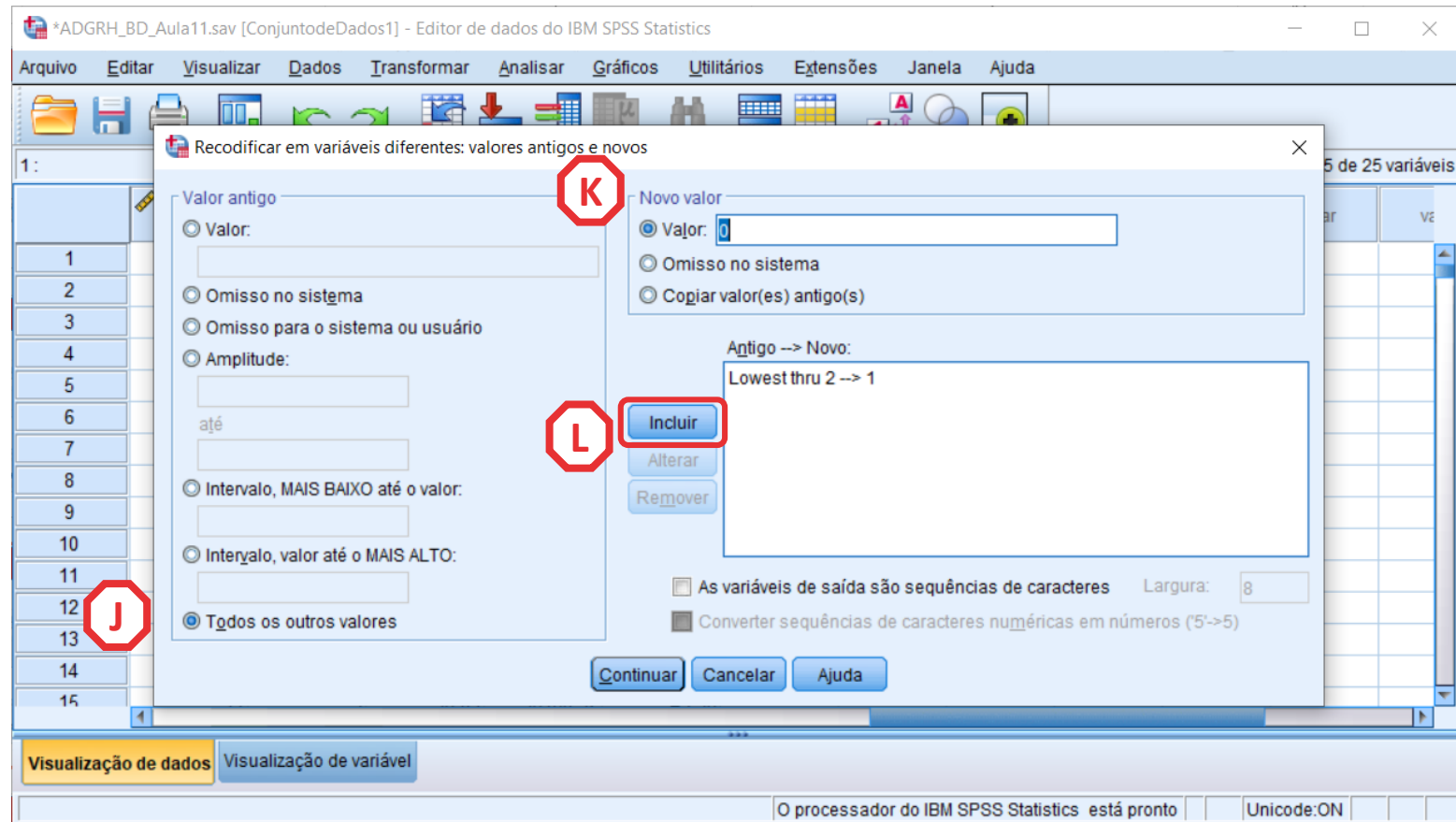
Criar a variável dependente

- Vamos definir que ao valor '1' e '2' na variável 'workcond_sat' **G**
- Corresponde o valor '1', na nova variável ('satisfied') **H**
- Selecionar 'Incluir' **I**



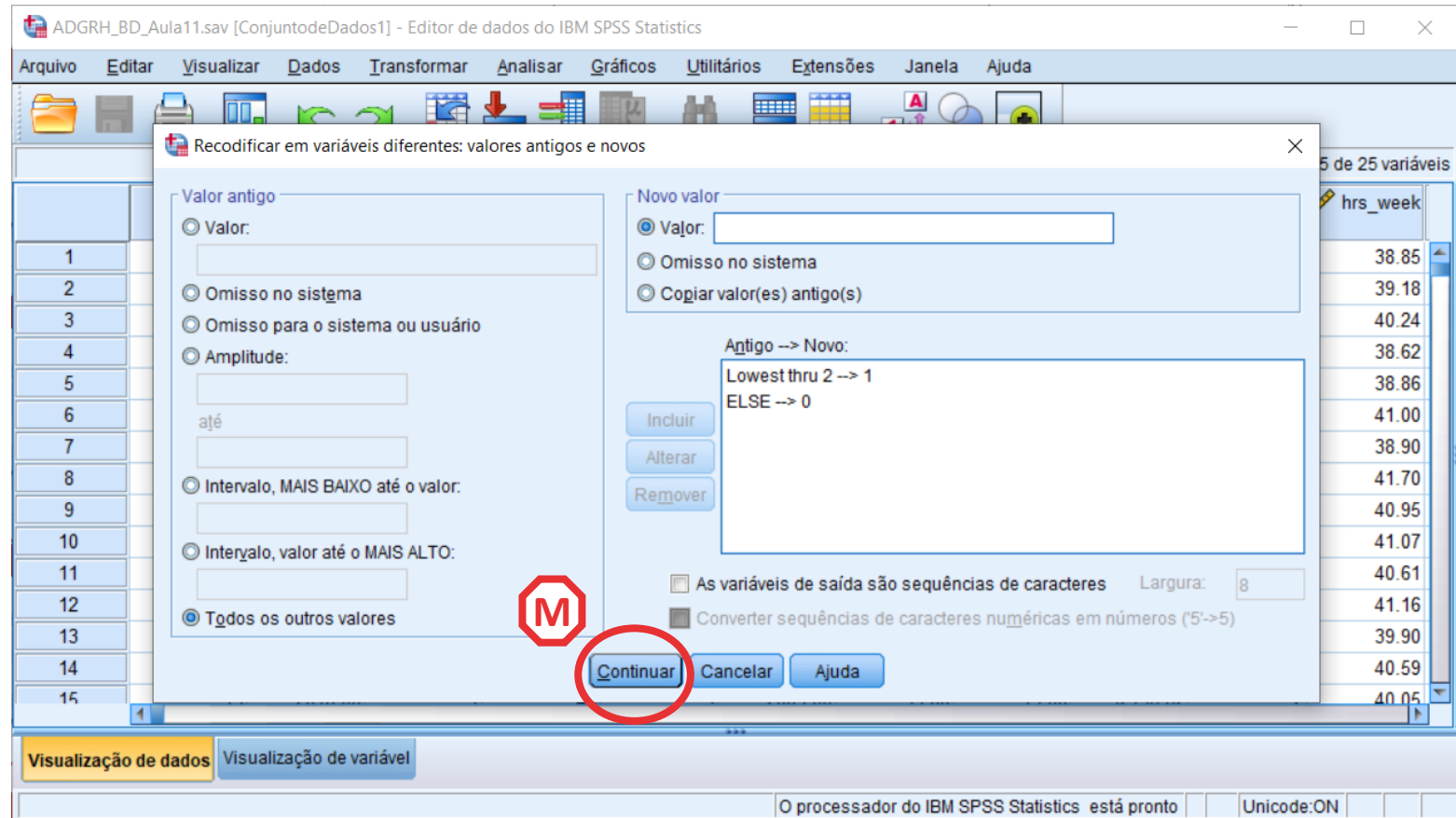
Criar a variável dependente

- Vamos definir que ao valor '1' e '2' na variável 'workcond_sat' **G**
- Corresponde o valor '1', na nova variável ('satisfied') **H**
- Selecionar 'Incluir' **I**
- Vamos definir que 'Todos os outros valores' na variável 'workcond_sat' **J**
- Corresponde o valor '0', na nova variável ('satisfied') **K**
- Selecionar 'Incluir' **L**



Criar a variável dependente

- Vamos definir que ao valor '1' e '2' na variável 'workcond_sat'
- Corresponde o valor '1', na nova variável ('satisfied')
- Selecionar 'Incluir'
- Vamos definir que 'Todos os outros valores' na variável 'workcond_sat'
- Corresponde o valor '0', na nova variável ('satisfied')
- Selecionar 'Incluir'
- Selecionar 'Continuar' / 'OK'



Criar a variável dependente

- Vamos verificar se a recodificação foi bem sucedida

Tabulação cruzada workcond_sat * satisfeito

Contagem

		satisfeito		Total
		,00	1,00	
workcond_sat	1 Very satisfied	0	719	719
	2 Satisfied	0	1633	1633
	3 Not Very Satisfied	1859	0	1859
	4 Not At All Satisfied	789	0	789
Total		2648	2352	5000

The screenshot shows the IBM SPSS Statistics interface. The main window displays a data view with the following columns: absent_nr, workcond_sat, evaluation, y_wage2, evaluation 2, and satisfied. The 'satisfied' column contains values of .00 and 1.00. The interface includes a menu bar (Arquivo, Editar, Visualizar, Dados, Transformar, Analisar, Gráficos, Utilitários, Extensões, Janela, Ajuda) and a toolbar with various icons. The status bar at the bottom indicates 'O processador do IBM SPSS Statistics está pronto' and 'Unicode:ON'.

Implementação do Modelo de Regressão Logística

2. Inspeccionar a variável dependente

- **Objectivo:**
 - **Perceber como se distribuem as observações na variável dependente ('satisfied').**

- **Neste caso, através de um Gráfico de Barras...**

Gráfico de Barras

- Seleccionar 'Gráficos' / 'Construtor de Gráfico'

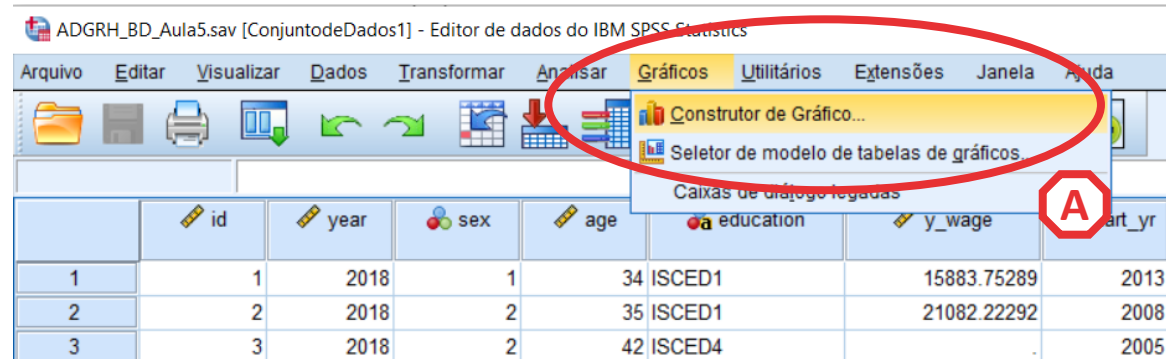


Gráfico de Barras

- Selecione 'Barras'
- Selecionar (com duplo-clique) o Gráfico de Barras (simples)

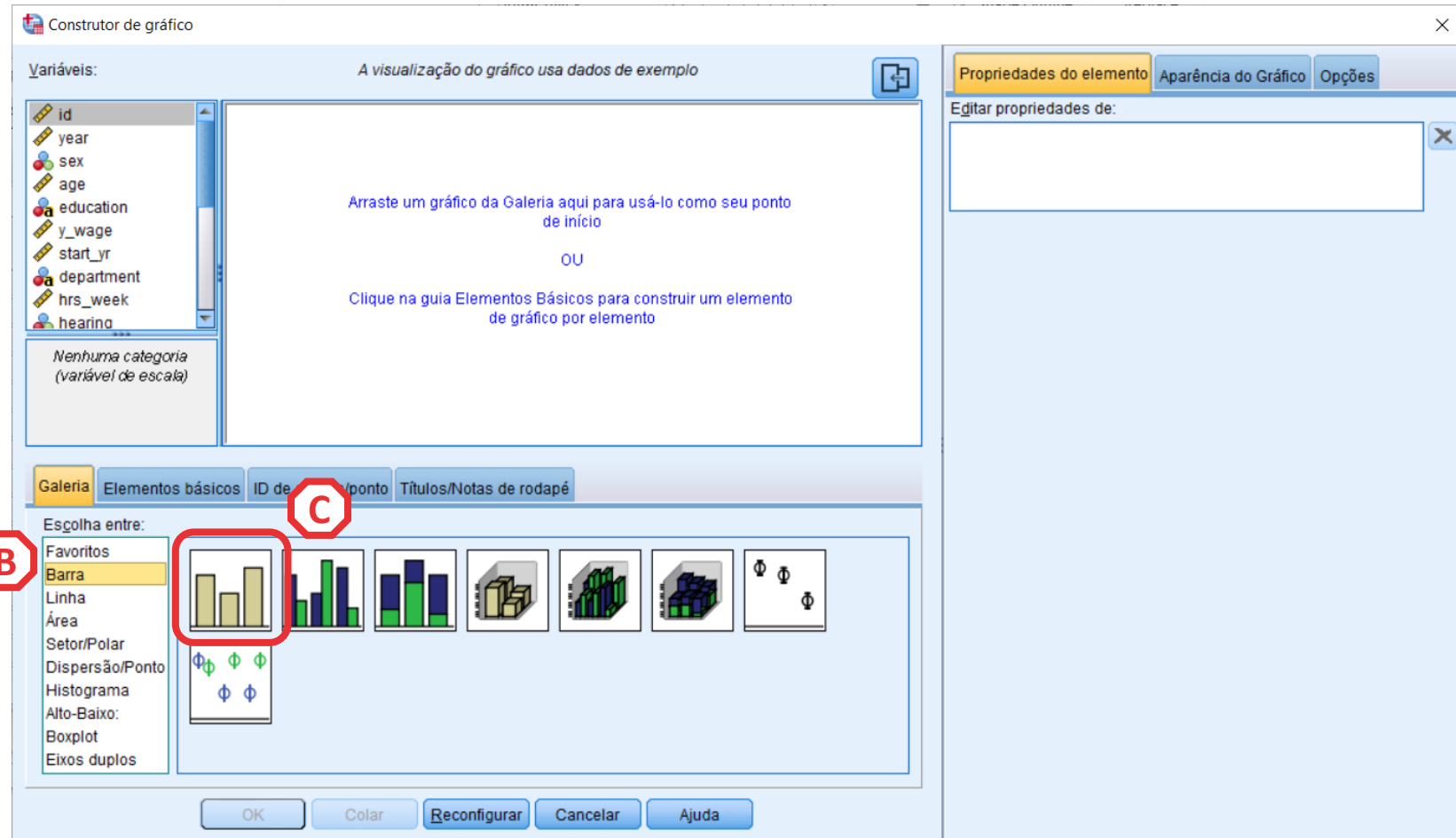


Gráfico de Barras

- Selecione 'Barras'
- Selecionar (com duplo-clique) o Gráfico de Barras (simples)
- Selecionar a variável 'satisfied'
- Colocar a variável 'satisfied' no 'Eixo X'

C

D

E

E

F

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- headache
- injury
- anxiety
- fatigue
- absent_nr
- workcond_sat
- evaluation
- wage2
- evaluation2
- Satisfied [satisfied]

Barra Simples Contagem

Eixo Y?

Filtro?

Eixo X?

Galeria Elementos básicos ID de grupos/pontos Eixos/Notas de rodapé

Escolha entre:

- Favoritos
- Barra
- Linha
- Área
- Setor/Polar
- Dispersão/Ponto
- Histograma
- Alto-Baixo:
- Boxplot
- Eixos duplos

Propriedades do elemento Aparência do Gráfico Opções

Editar propriedades de:

Barra1

X-Eixo1 (Barra1)

Y-Eixo1 (Barra1)

Título 1

Estadísticas

Variável:

Estatística:

Contagem

Configurar parâmetros...

Exibir barra de erros

Representação de Barras de Erros

Intervalos de confiança

Nível (%): 95

Erro padrão

Multiplicador: 2

Desvio padrão

Multiplicador: 2

Estilo de barra:

Barra

OK Colar Reconfigurar Cancelar Ajuda

Gráfico de Barras

- Selecione 'Barras' **C**
- Selecionar (com duplo-clique) o Gráfico de Barras (simples) **D**
- Selecionar a variável 'satisfied' **E**
- Colocar a variável 'satisfied' no 'Eixo X' **F**
- Selecionar 'Porcentagem' **G**
- Selecionar 'OK' **H**

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- id
- year
- sex
- age
- education
- start_yr
- start_age
- experience
- y_wage
- department

Nenhuma categoria (variável de escala)

Barra Simples Contagem de Satisfied

Contagem

Filtro?

Categoria 1 Categoria 2 [Mais...]

Satisfied

Galeria Elementos básicos ID de grupos/ponto Títulos/Notas de rodapé

Escolha entre:

- Favoritos
- Barra
- Linha
- Área
- Setor/Polar
- Dispersão/Ponto
- Histograma
- Alto-Baixo:
- Boxplot
- Eixos duplos

OK Colar Reconfigurar Cancelar Ajuda

Propriedades do elemento Aparência do Gráfico Opções

Editar propriedades de:

Barra1

X-Eixo1 (Barra1)

Y-Eixo1 (Barra1)

Título

Estadísticas

Variável:

Estatística:

- Contagem
- Contagem cumulativa
- Porcentagem (?)
- Porcentagem acumulativa
- Valor
- Média
- Mediana
- Mediana do grupo

Multiplicador: 2

Desvio padrão

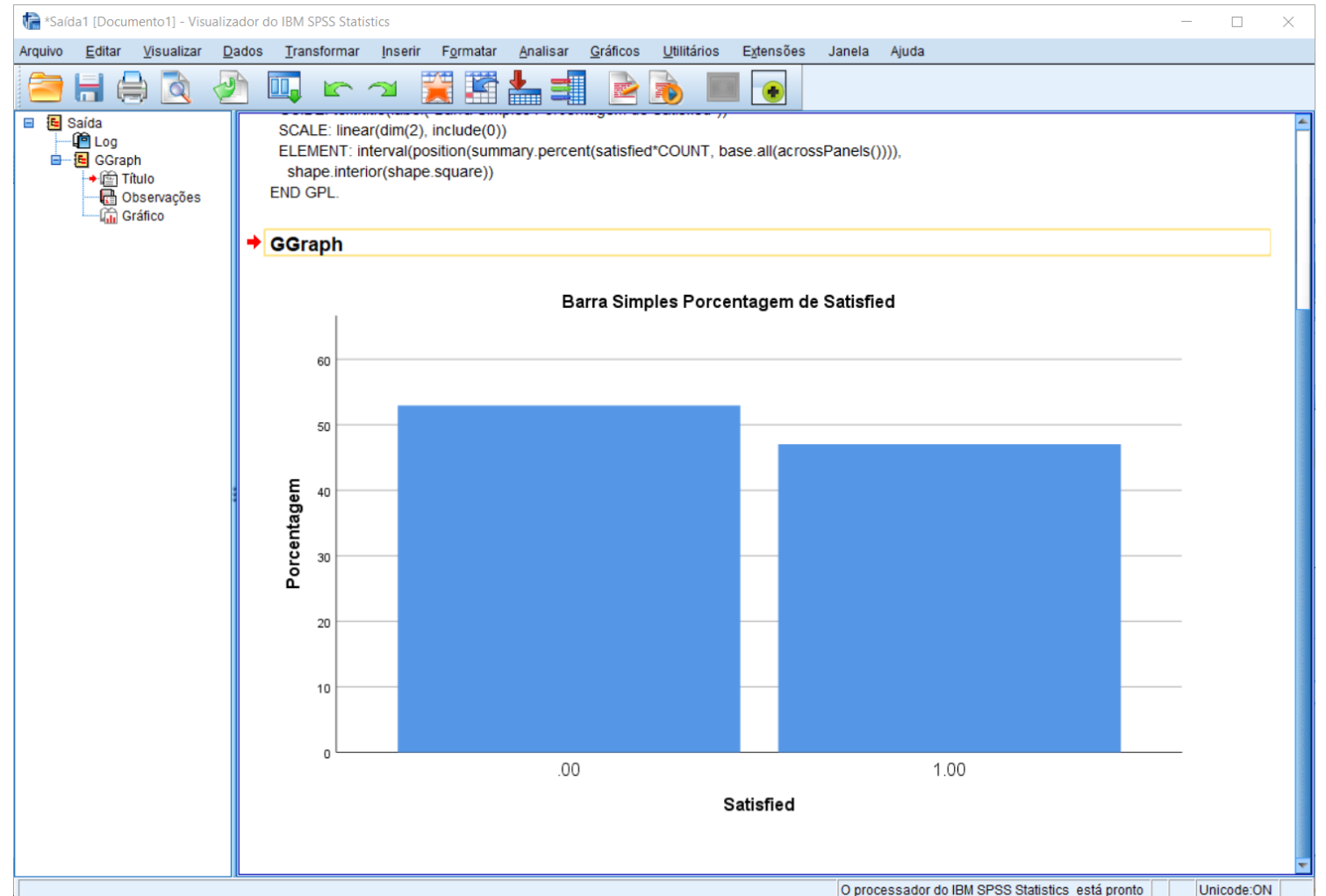
Multiplicador: 2

Estilo de barra:

Barra

Gráfico de Barras

- O gráfico é publicado no 'Visualizador de Resultados'
- Menos de metade dos trabalhadores está satisfeito com as condições de trabalho
- Gráfico dever ser complementado com tabela de frequências (ver Aula 5)



Implementação do Modelo de Regressão Logística

3. Explorar as relações entre variáveis

- **Objectivo:**

- **Perceber como se relaciona a variável dependente ('satisfied') com a variável independente 'sex' que queremos incluir no nosso modelo de regressão**

... neste caso, através de um Gráfico de Barras Sobrepostas

Gráfico de Barras Sobrepostas

- Selecionar 'Gráficos' / 'Construtor de Gráfico'

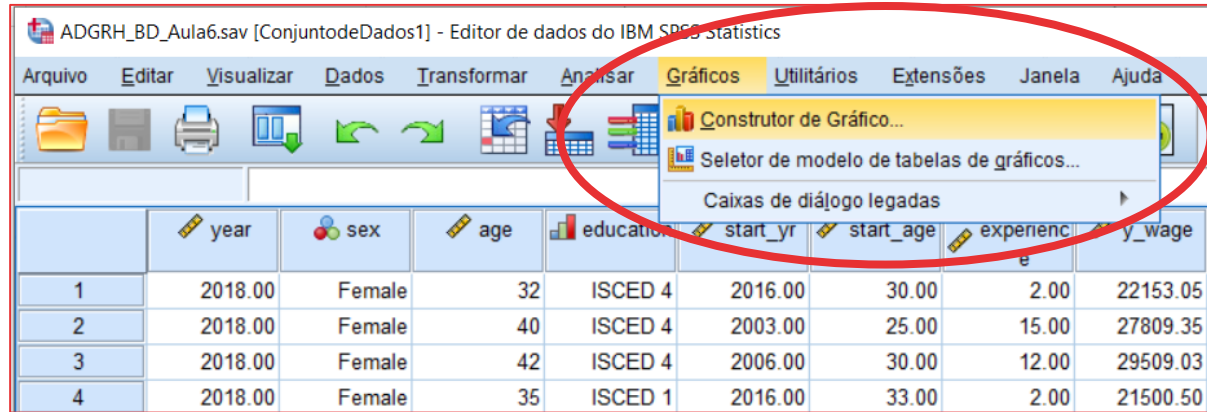


Gráfico de Barras Sobrepostas

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Barra'
- Duplo-Clique na opção 'Gráfico de Barras Sobrepostas'

A

B

C

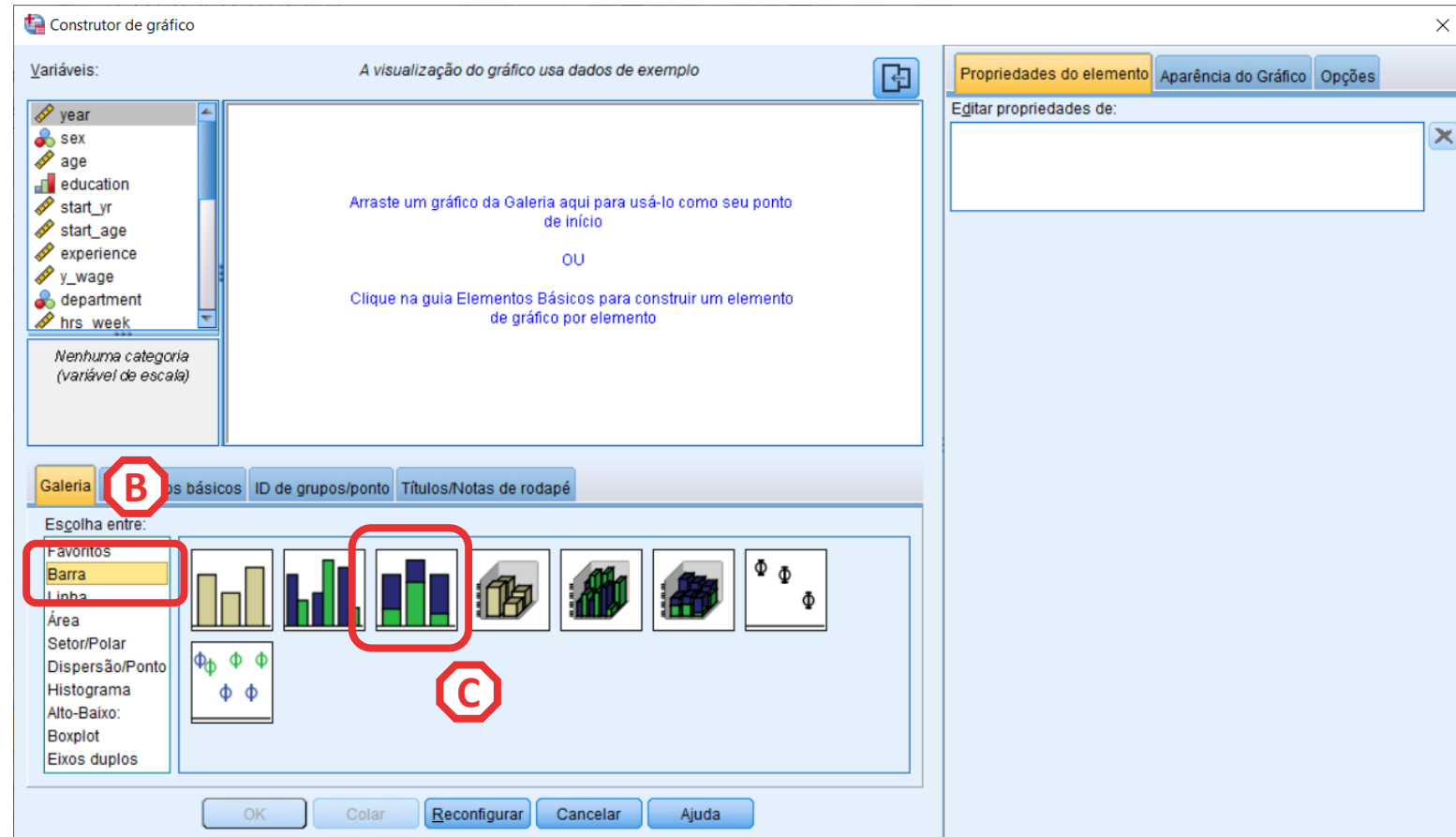


Gráfico de Barras Sobrepostas

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Barra'
- Duplo-Clique na opção 'Gráfico de Barras Sobrepostas'
- Selecionar a variável 'satisfied'...
- ... e colocar na caixa do 'Cluster em X'

Exercício: Colocar a variável 'sex' no 'Eixo X'

The screenshot shows the 'Construtor de gráfico' window with the following elements:

- Variáveis:** A list of variables including 'headache', 'injury', 'anxiety', 'fatigue', 'absent_nr', 'workcond_sat', 'evaluation', 'y_wage2', 'evaluation2', and 'Satisfied [satisfied]'. The 'Satisfied [satisfied]' variable is highlighted with a red box labeled 'C'.
- Galeria:** A gallery of chart types including 'Barra', 'Linha', 'Área', 'Setor/Polar', 'Dispersão/Ponto', 'Histograma', 'Alto-Baixo', 'Boxplot', and 'Eixos duplos'. The 'Barra' option is highlighted with a red box labeled 'B'.
- Chart Area:** A stacked bar chart titled 'Barra Empilhada Contagem' is displayed. The chart has three bars, each with a red bottom segment and a blue top segment. The X-axis is labeled 'Eixo X?'. A red box labeled 'D' highlights the 'Empilhar: definir oor' button, and a red box labeled 'E' highlights the 'Eixo X?' label.
- Propriedades do elemento:** A panel on the right showing properties for 'Barra1'. The 'Estatística' (Statistic) is set to 'Contagem' (Count). The 'Estilo de barra' (Bar style) is set to 'Barra'.

Gráfico de Barras Sobrepostas

- Seleccionar 'Barra 1'
- Definir 'Percentagens' como parâmetro de comparação



Gráfico de Barras Sobrepostas

- Seleccionar 'Barra 1'
- Definir 'Porcentagem' como parâmetro de comparação
- Seleccionar 'Configurar Parâmetros'



Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- id
- year
- sex
- age
- education
- start_yr
- start_age
- experience
- y_wage
- department

Female

Male

Barra Empilhada Porcentagem de sex por Satisfied

Empilhar: definir cor

Satisfied

Porcentagem

Female

Male

sex

Filtro?

Galeria

Elementos básicos

ID de grupos/ponto

Títulos/Notas de rodapé

Escolha entre:

Favoritos

Barra

Linha

Área

Setor/Polar

Dispersão/Ponto

Histograma

Alto-Baixo:

Boxplot

Eixos duplos

Configurar parâmetros...

Propriedades do elemento

Aparência do Gráfico

Opções

Editar propriedades de:

Barra1

X-Eixo1 (Barra1)

Y-Eixo1 (Barra1)

GroupColor (Barra1)

Estatísticas

Variável:

Estatística:

Porcentagem ()

Exibir barra de erros

Representação de Barras de Erros

Intervalos de confiança

Nível (%): 95

Erro padrão

Multiplicador: 2

Desvio padrão

Multiplicador: 2

Estilo de barra:

Barra

OK

Colar

Reconfigurar

Cancelar

Ajuda

Gráfico de Barras Sobrepostas

- Seleccionar 'Barra 1'
- Definir 'Porcentagem' como parâmetro de comparação
- Seleccionar 'Configurar Parâmetros'
- Seleccionar o denominador 'Total para cada categoria do eixo X'
- Seleccionar 'Continuar' / 'OK'

F

G

H

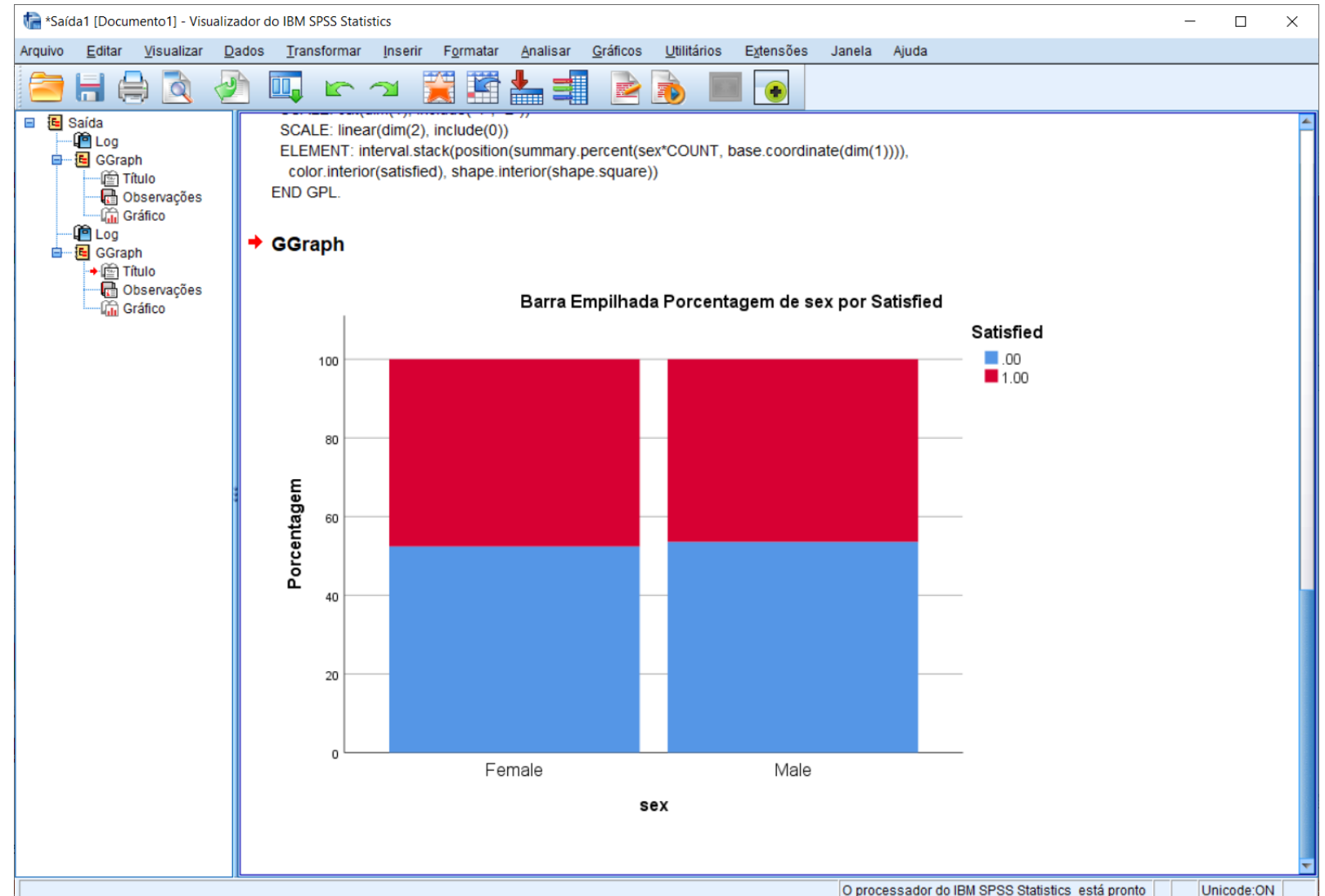
I

J

The screenshot shows the 'Construtor de gráfico' (Chart Builder) interface. The main window displays a stacked bar chart titled 'Barra Empilhada Porcentagem de sex por Satisfied'. The chart has two bars, one for 'Female' and one for 'Male', each divided into three segments representing different categories. The Y-axis is labeled 'Porcentagem'. A dialog box titled 'Propriedades do elemento: definir parâmetros' is open, showing a dropdown menu for 'Denominador para calcular a porcentagem:' with the following options: 'Total Geral', 'Total para cada categoria do eixo X', and 'Total para cada categoria de variável de legenda (mesma cor de preenchimento)'. The 'Total para cada categoria do eixo X' option is selected. The 'OK' button at the bottom of the dialog is circled in red. Other interface elements include a variable list on the left, a gallery of chart types at the bottom, and a properties panel on the right.

Gráfico de Barras Sobrepostas

- O resultado é publicado no 'Visualizador de Resultados'
- Não se notam diferenças significativas na satisfação com as condições de trabalho em função do sexo...
- Gráfico deve ser complementado com tabela de dupla entrada para estas variáveis (ver Aula 6)



- **Objectivo:**

- **Comparar a idade média dos trabalhadores que estão satisfeitos com as condições de trabalho com a idade média dos trabalhadores que não estão satisfeitos.**

... neste caso, através de um Gráfico de 'Barra de Erro Simples'

Gráfico de Barra de Erro Simples

- Selecionar 'Gráficos' / 'Construtor de Gráfico'

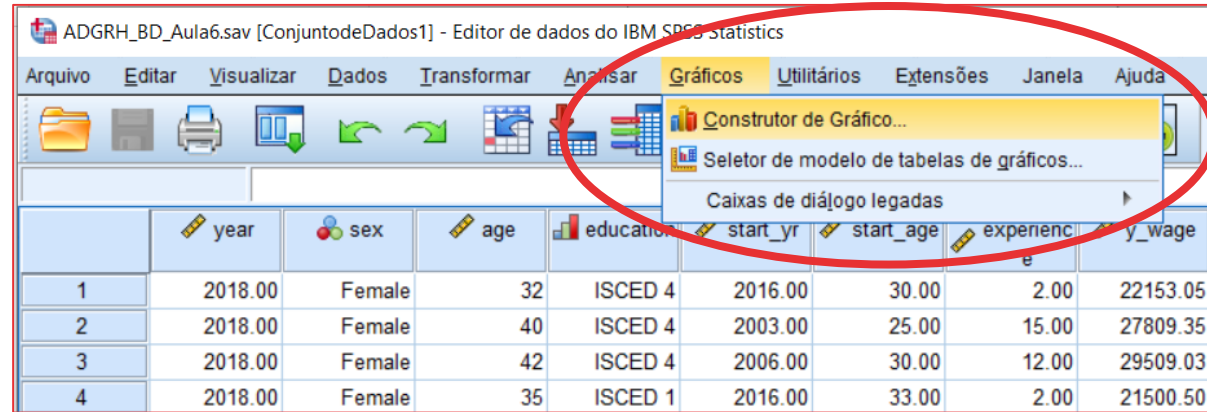


Gráfico de Barra de Erro Simples

- Seleccionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Barra'
- Duplo-Clique na opção 'Gráfico Barra de Erro Simples'

A

B

C

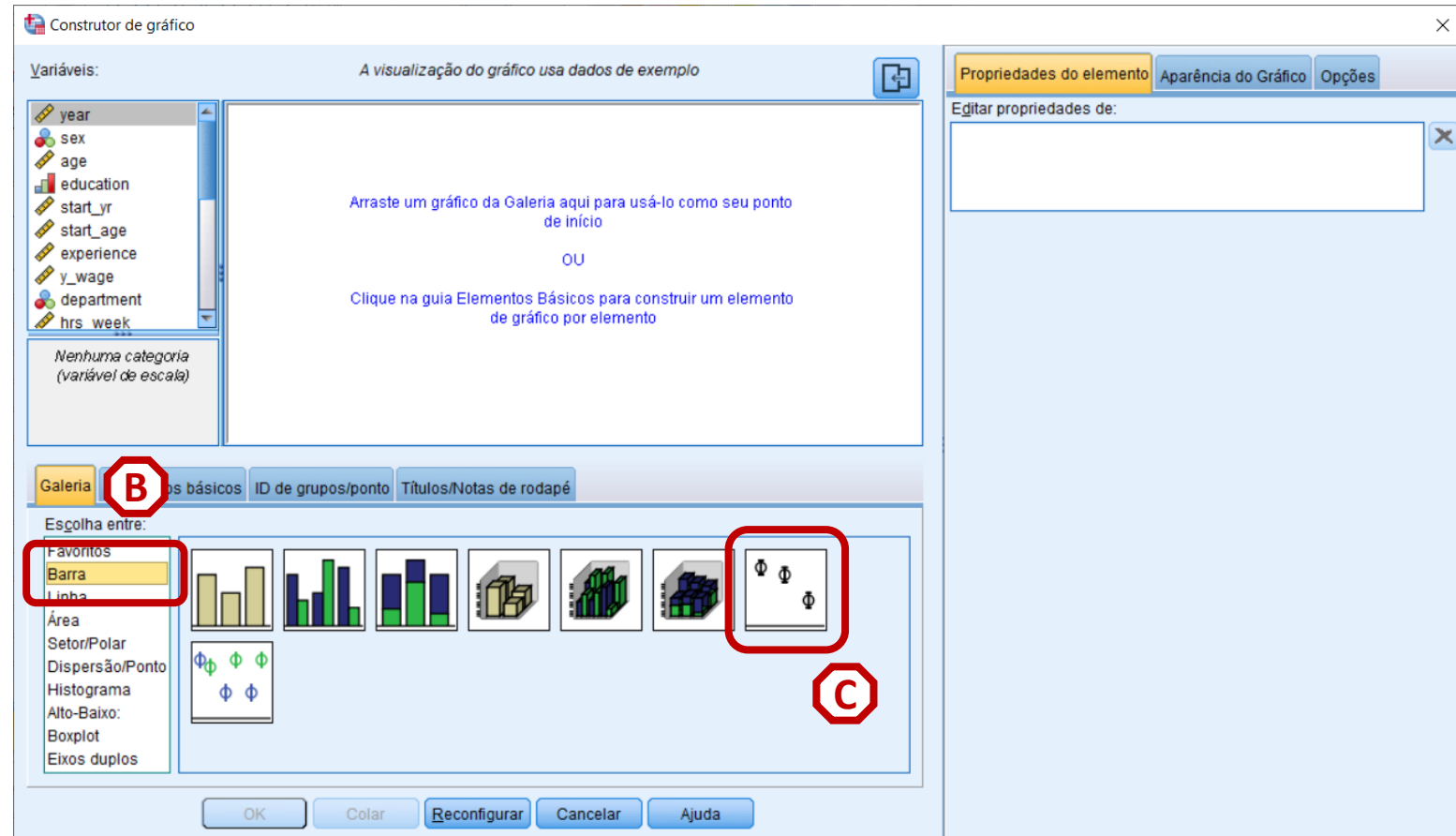


Gráfico de Barra de Erro Simples

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
 - Selecione 'Barra'
 - Duplo-Clique na opção 'Gráfico Barra de Erro Simples'
 - Selecionar a variável 'satisfied'...
- ... e colocar na caixa do 'Eixo X'

Exercício: Colocar a variável 'age' no 'Eixo Y'

A Variáveis:

- headache
- injury
- anxiety
- fatigue
- absent_nr
- workcond_sat
- evaluation
- y_wage2
- evaluation2
- Satisfied [satisfied]**
- Categoria 1
- Categoria 2

B Escolha entre:

- Favoritos
- Barra**
- Linha
- Área
- Setor/Polar
- Dispersão/Ponto
- Histograma
- Alto-Baixo:
- Boxplot
- Eixos duplos

C Gráfico Barra de Erro Simples

D Eixo X?

E Eixo X?

Propriedades do elemento

Estatísticas

Variável:

Estatística: Média

Exibir barra de erros

Representação de Barras de Erros

Intervalos de confiança

Nível (%): 95

Erro padrão

Multiplicador: 2

Desvio padrão

Multiplicador: 2

Linhas de Ajuste Lineares

Total

Subgrupos

Gráfico de Barra de Erro Simples

- Escolher a 'Média' como estatística a ser representada no gráfico [ou 'Valor' compara as soluções!]



Gráfico de Barra de Erro Simples

- Escolher a 'Média' como estatística a ser representada no gráfico
- Desligar a opção 'Exibir Barra de Erros'
- Clicar 'OK'



Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- id
- year
- sex
- age
- education
- start_yr
- start_age
- experience
- y_wage
- department

Nenhuma categoria (variável de escala)

Barra de Erro Simples Média de age por Satisfied

Propriedades do elemento

Aparência do Gráfico

Opções

Editar propriedades de:

Ponto1

X-Eixo1 (Ponto1)

Y-Eixo1 (Ponto1)

Título 1

Estatísticas

Variável: age

Estatística: Média

Exibir barra de erros

Representação de Barras de Erros

- Intervalos de confiança
- Nível (%): 95
- Erro padrão
- Multiplicador: 2
- Desvio padrão
- Multiplicador: 2

Empilhar valores idênticos

Exibir linhas de projeção verticais entre pontos

Linhas de Ajuste Lineares

- Total
- Subgrupos

Escolha entre:

Favoritos

Barra

Linha

Área

Setor/Polar

Dispersão/Ponto

Histograma

Alto-Baixo:

Boxplot

Eixos duplo

OK

Colar

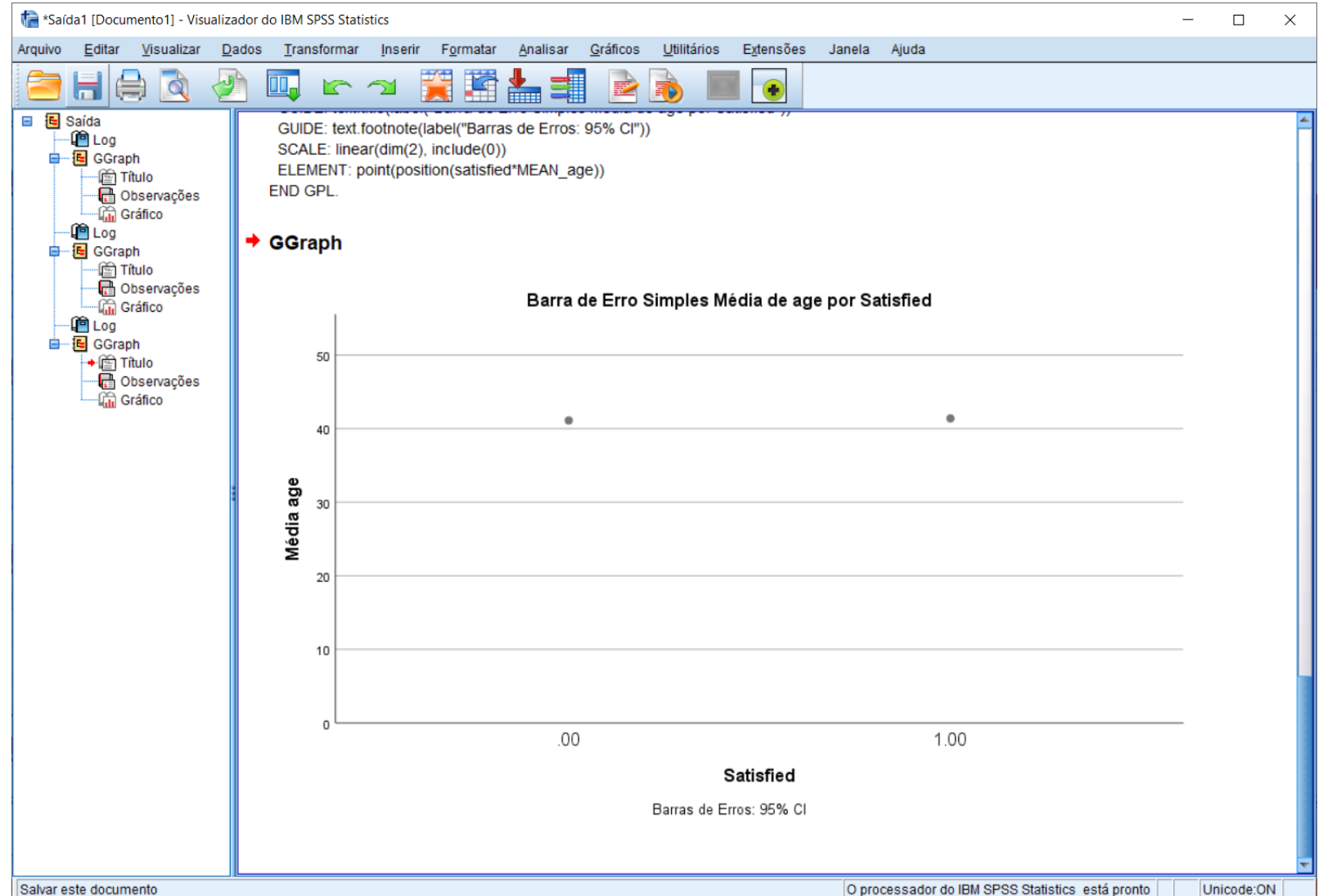
Reconfigurar

Cancelar

Ajuda

Gráfico de Barra de Erro Simples

- O resultado é publicado no ‘Visualizador de Resultados’
- Não se notam diferenças significativas em termos da idade média relativamente à sua satisfação com as condições de trabalho ...
- Gráfico deve ser complementado com tabela de comparação de médias (ver Aula 6)



- **Objectivo:**
 - **Perceber como se relaciona a variável dependente ('satisfied') com a variável independente 'education' que queremos incluir no nosso modelo de regressão**

... neste caso, através de um Gráfico de Barras Agrupadas

Gráfico de Barras Agrupadas

- Selecionar 'Gráficos' / 'Construtor de Gráfico'

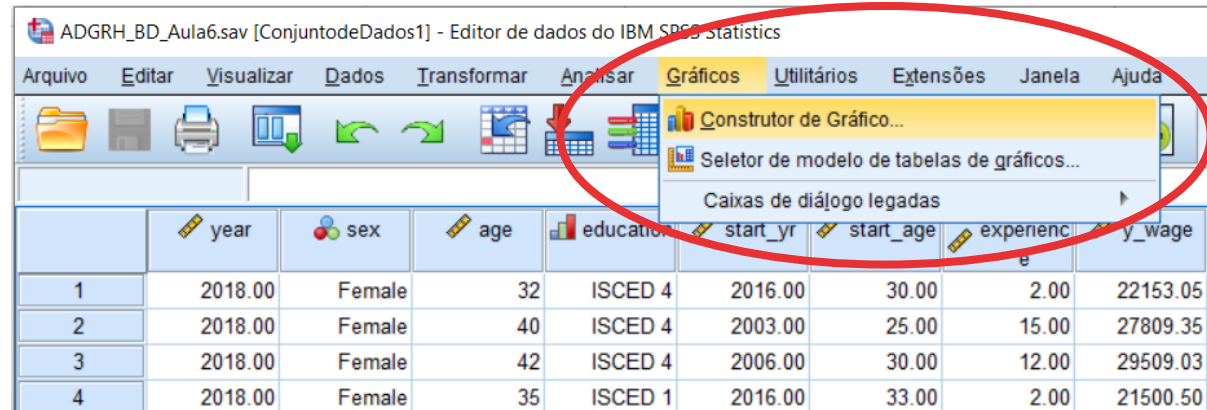


Gráfico de Barras Agrupadas

- Seleccionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Barra'
- Duplo-Clique na opção 'Gráfico de Barras Agrupadas'

A

B

C

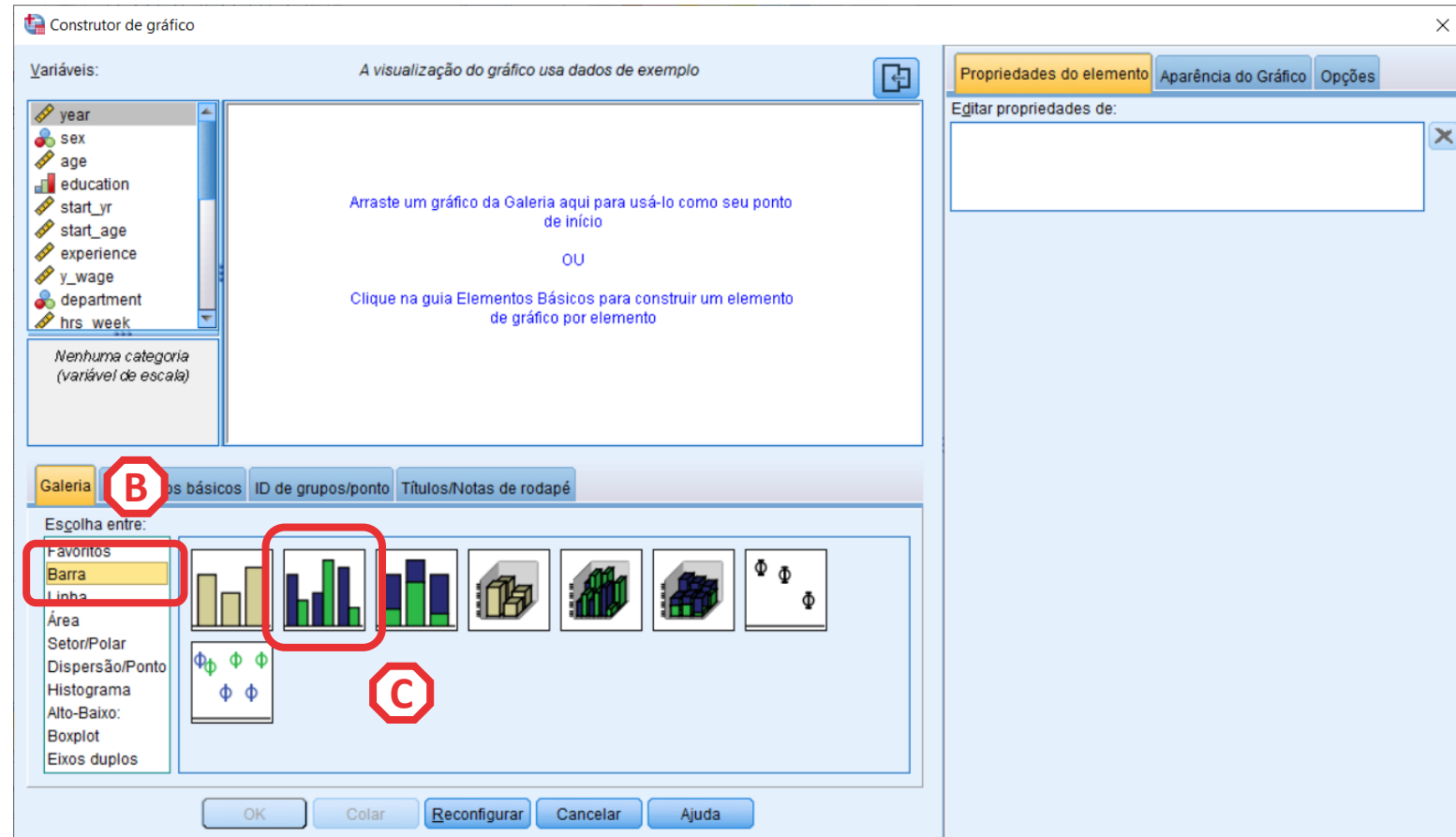


Gráfico de Barras Agrupadas

- Selecionar ‘Gráficos’ / ‘Construtor de Gráfico’
- Selecione ‘Barra’
- Duplo-Clique na opção ‘Gráfico de Barras Agrupadas’
- Selecionar a variável ‘satisfied’...

... e colocar na caixa do ‘Cluster em X’

A

B

C

D

E

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- headache
- injury
- anxiety
- fatigue
- absent_nr
- workcond_sat
- evaluation
- y_wage2
- evaluation2
- Satisfied [satisfied]

Barra em Cluster Contagem

Cluster em X: definir cor...

Eixo Y?

Eixo X?

Filtro?

Galeria

Elementos básicos ID de grupos/ponto Títulos/Notas de rodapé

Escolha entre:

- Favoritos
- Barra
- Linha
- Área
- Setor/Polar
- Dispersão/Ponto
- Histograma
- Alto-Baixo:
- Boxplot
- Eixos duplos

OK Colar Reconfigurar Cancelar Ajuda

Propriedades do elemento Aparência do Gráfico Opções

Editar propriedades de:

Barra1

X-Eixo1 (Barra1)

Y-Eixo1 (Barra1)

GroupColor (Barra1)

Estatísticas

Variável:

Estatística:

Contagem

Configurar parâmetros...

Exibir barra de erros

Representação de Barras de Erros

- Intervalos de confiança
- Nível (%): 95
- Erro padrão
- Multiplicador: 2
- Desvio padrão
- Multiplicador: 2

Estilo de barra:

Barra

Exercício: Colocar a variável
‘education’ no ‘Eixo X’

Gráfico de Barras Agrupadas

- Seleccionar 'Barra 1'
- Definir 'Porcentagem' como parâmetro de comparação



Construtor de gráfico

Variáveis: A visualização do gráfico usa dados de exemplo

Barra em Cluster Porcentagem de education por Satisfied

Cluster em X: definir cor Satisfied

Porcentagem

education

ISCED 1 ISCED 2 ISCED 3 ISCED 4 ISCED 5

Propriedades do elemento Aparência do Gráfico Opções

Editar propriedades de: Barra1

X-Eixo1 (Barra1)

Y-Eixo1 (Barra1)

Estatísticas

Variável:

Estatística:

Porcentagem ()

Contagem

Contagem cumulativa

Porcentagem ()

Porcentagem acumulativa

Valor

Média

Mediana

Mediana do grupo

Desvio padrão

Multiplicador: 2

Estilo de barra: Barra

OK Colar Reconfigurar Cancelar Ajuda

Gráfico de Barras Agrupadas

- Seleccionar 'Barra 1'
- Definir 'Porcentagem' como parâmetro de comparação
- Seleccionar 'Configurar parâmetros'



Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- headache
- injury
- anxiety
- fatigue
- absent_nr
- workcond_sat
- evaluation
- y_wage2
- evaluation2
- Satisfied [satisfied]

Barra em Cluster Porcentagem de education por Satisfied

Cluster em X: definir cor

Satisfied

Porcentagem

education

ISCED 1 ISCED 2 ISCED 3 ISCED 4 ISCED 5

Configurar parâmetros...

Propriedades do elemento

Aparência do Gráfico

Opções

Editar propriedades de:

Barra1

X-Eixo1 (Barra1)

Y-Eixo1 (Barra1)

GroupColor (Barra1)

Estadísticas

Variável:

Estadística:

Porcentagem ()

Exibir barra de erros

Representação de Barras de Erros

Intervalos de confiança

Nível (%): 95

Erro padrão

Multiplicador: 2

Desvio padrão

Multiplicador: 2

Estilo de barra:

Barra

OK Colar Reconfigurar Cancelar Ajuda

Gráfico de Barras Agrupadas

- Seleccionar 'Barra 1'
- Definir 'Porcentagem' como parâmetro de comparação
- Seleccionar 'Configurar parâmetros'
- Seleccionar o denominador 'Total para cada categoria do eixo X'
- Seleccionar 'Continuar' / 'OK'

F

G

H

I

J

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- start_yr
- start_age
- experience
- y_wage
- department
- hrs_week
- hearing
- skin
- back
- muscle_upper

Cluster em X: definir cor

anxiety

Filtro?

Propriedades do elemento: definir parâmetros

Denominador para calcular a porcentagem:

- Total Geral
- Total para cada categoria do eixo X
- Total para cada categoria de variável de legenda (mesma cor de preenchimento)

OK Colar Reconfigurar Cancelar Ajuda

Propriedades do elemento Aparência do Gráfico Opções

Editar propriedades de:

Barra1

X-Eixo1 (Barra1)

Y-Eixo1 (Barra1)

GroupColor (Barra1)

Estadísticas

Variável:

Estatística:

Porcentagem ()

Configurar parâmetros...

de erros

de Erros

95

ador: 2

padrão

Multiplicador: 2

Estilo de barra:

Barra

Escolha entre:

Favoritos

Barra

Linha

Área

Setor/Polar

Dispersão/Ponto

Histograma

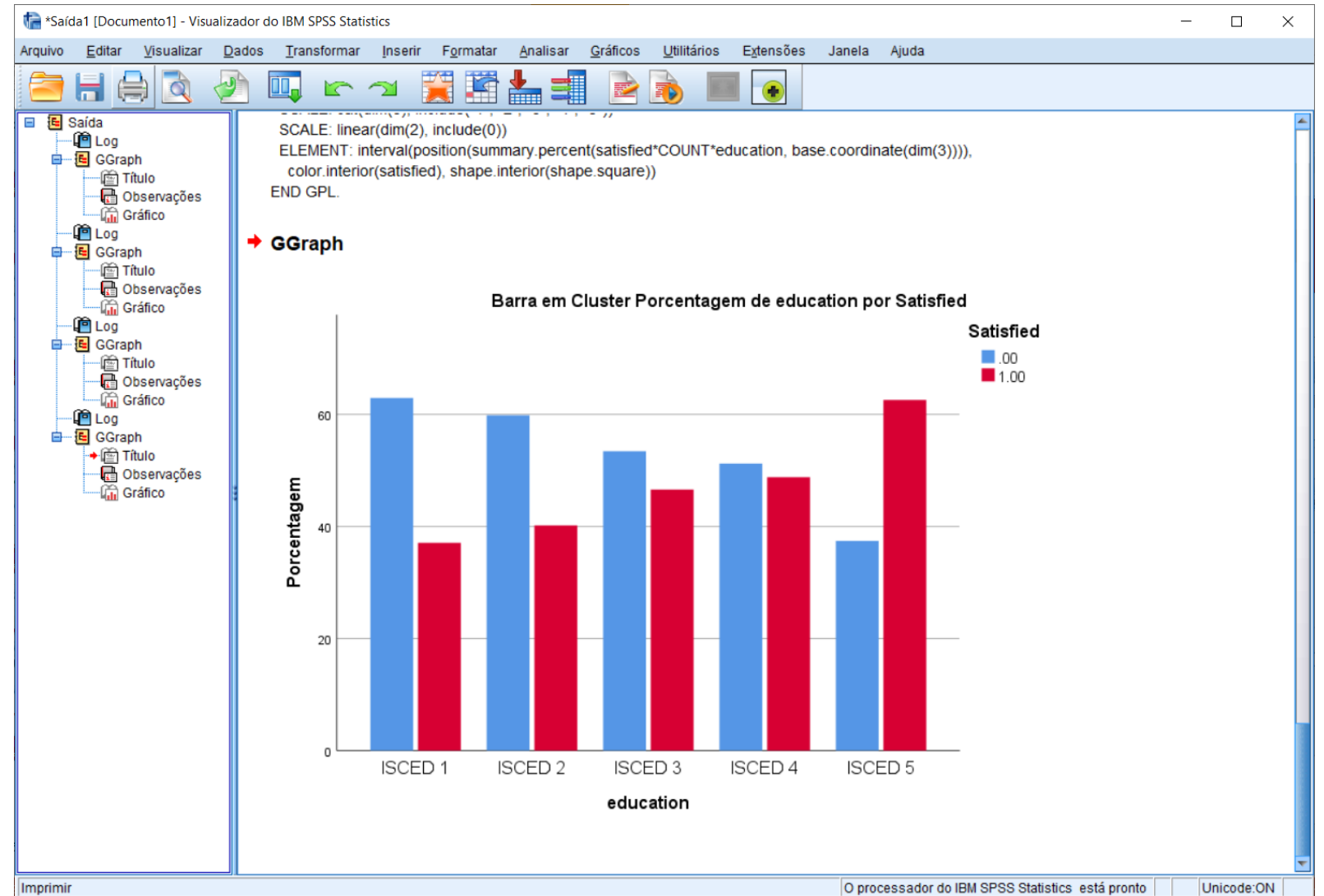
Alto-Baixo:

Boxplot

Eixos duplos

Gráfico de Barras Agrupadas

- O resultado é publicado no 'Visualizador de Resultados'
- Podemos perceber que a percentagem de trabalhadores satisfeitos com as condições do trabalho aumenta em função da educação
- Deve ser complementado com uma tabela de dupla entrada (ver Aula 6)



Implementação do Modelo de Regressão Logística

4. Definir o modelo de regressão logística

- **Objectivo:**
 - **Estimar um modelo de regressão logística que explique em que medida a probabilidade de se sentir satisfeito com as condições de trabalho varia em função do sexo, idade e educação.**

Regressão Logística

- Selecionar 'Analisar' / 'Regressão' / 'Logística binária'

The screenshot shows the IBM SPSS Statistics interface. The 'Analisar' menu is open, and the 'Regressão' option is selected, which has opened a sub-menu where 'Logística binária...' is highlighted. A red circle highlights the 'Analisar' menu and the 'Regressão' sub-menu. A red octagon with the letter 'A' is placed in the top right corner of the screenshot. Another red octagon with the letter 'A' is placed to the left of the data table.

	absent_nr	workcond_sat	evaluation
1	9	3	49.08
2	10	2	52.58
3	15	2	53.80
4	10	2	50.42
5	11	4	48.43
6	10	2	54.34
7	6	3	49.65
8	10	4	46.84
9	8	2	50.76
10	17	3	53.33
11	14	2	50.59

Regressão Logística

- Selecionar 'Analisar' / 'Regressão' / 'Logística binária'
- Selecionar a variável 'satisfied'
- Colocar na caixa 'Dependente'

Exercício: Colocar as variáveis 'sex', 'age' e 'education' na caixa 'Covariáveis'



*ADGRH_BD_Aula11.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

12: y_wage2 29668.46659292

absent_nr workcond_sat

	absent_nr	workcond_sat
1	9	2
2	10	2
3	15	2
4	10	2
5	11	4
6	10	2
7	6	3
8	10	4
9	8	3
10	17	2
11	14	2
12	7	2
13	13	2
14	7	2
15	11	3

Regressão logística

Dependente:

Covariáveis:

Método: Inserir

Variável de seleção:

OK Color Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

Regressão Logística

- Selecionar 'Analisar' / 'Regressão' / 'Logística binária'

- Selecionar a variável 'satisfied'

- Colocar na caixa 'Dependente'

Exercício: Colocar as variáveis 'sex', 'age' e 'education' na caixa 'Covariáveis'

- Clicar no botão 'Categórico'

A

B

C

D

The screenshot shows the IBM SPSS Statistics interface with the 'Regressão logística' dialog box open. The dependent variable is 'satisfied' and the independent variables are 'sex', 'age', and 'education'. The 'Categorical' button is highlighted. The background shows a data table with columns 'id', 'year', 'wage', 'department', and 'hrs_week'.

id	year	wage	department	hrs_week
1	2018.00	153.05	6	38.85
2	2018.00	309.35	5	39.18
3	2018.00	509.03	2	40.24
4	2018.00	500.50	1	38.62
5	2018.00	786.81	2	38.86
6	2018.00	153.80	4	41.00
7	2018.00	367.09	1	38.90
8	2018.00	399.42	3	41.70
9	2018.00	382.25	1	40.95
10	2018.00	331.50	5	41.07
11	2018.00	396.46	2	40.61
12	2018.00	315.75	1	41.16
13	2018.00	361.65	1	39.90
14	2018.00	177.37	4	40.59
15	2018.00	136.62	2	40.05

Regressão Logística

- Selecionar a variável 'sex'
- Colocar na caixa 'Covariáveis categóricas'



The screenshot displays the IBM SPSS Statistics interface. The main window shows a data table with columns: id, year, wage, department, and hrs_week. The 'Regressão logística' dialog box is open, titled 'Regressão logística: definir variáveis de categoria'. It has two main sections: 'Covariáveis' and 'Covariáveis categóricas'. The 'sex' variable is selected in the 'Covariáveis' list and is being moved to the 'Covariáveis categóricas' box. The 'Alterar Contraste' section is visible at the bottom of the dialog box, with 'Indicador' selected in the 'Contraste' dropdown and 'Último' selected in the 'Categoria de referência' radio buttons. The 'Continuar' button is highlighted.

id	year	wage	department	hrs_week
1	2018.0	2153.05	6	38.85
2	2018.0	7809.35	5	39.18
3	2018.0	9509.03	2	40.24
4	2018.0	1500.50	1	38.62
5	2018.0	6786.81	2	38.86
6	2018.0	1153.80	4	41.00
7	2018.0	0867.09	1	38.90
8	2018.0	4899.42	3	41.70
9	2018.0	0882.25	1	40.95
10	2018.0	9631.50	5	41.07
11	2018.0	2096.46	2	40.61
12	2018.0	8315.75	1	41.16
13	2018.0	1361.65	1	39.90
14	2018.0	30477.37	4	40.59
15	2018.0	32136.62	2	40.05

Regressão Logística

- Selecionar a variável 'sex'
- Colocar na caixa 'Covariáveis categóricas'
- Por defeito: define a categoria em primeiro lugar como categoria de referência
- Carregar no botão 'Alterar'
- Repetir esta operação para a variável 'education'
- Selecionar 'Continuar'



*ADGRH_BD_Aula11.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

12 : y_wage2 29668.46659292

Regressão logística

Regressão logística: definir variáveis de categoria

Covariáveis:

- age
- education

Covariáveis categóricas:

sex(Indicador)

Alterar Contraste

Contraste: Indicador

Categoria de referência: Último Primeiro

Alterar

Continuar Cancelar Ajuda

OK Colar Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

Regressão Logística

- Clicar no botão 'Opções'



The screenshot shows the IBM SPSS Statistics interface with the 'Regressão logística' dialog box open. The 'Dependente:' field contains 'Satisfied [satisfied]'. The 'Método:' dropdown is set to 'Inserir'. The 'Opções...' button is highlighted with a red circle. A red octagonal icon with a white letter J is overlaid on the 'Opções...' button. The background shows a data table with columns 'absent_nr' and 'workcond_sat'.

	absent_nr	workcond_sat
1	9	2
2	10	2
3	15	2
4	10	2
5	11	4
6	10	2
7	6	3
8	10	4
9	8	2
10	17	3
11	14	2
12	7	2
13	13	2
14	7	2
15	11	3

Regressão Logística

- Clicar no botão 'Opções'
- Selecionar a opção 'Qualidade do ajuste de Hosmer-Lemeshow'
- Selecionar a opção 'CI para exp(B)'
- Selecionar 'Continuar' / 'OK'



*ADGRH_BD_Aula11.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

12: y_wage2 296

Regressão logística: opções

Estadísticas e gráficos

- Gráficos de classificação
- Qualidade do ajuste de Hosmer-Lemeshow
- Listagem por caso de resíduos
- Valgres discrepantes no lado de fora 2 std. dev.
- Todos os casos
- Correlações de estimativas
- Histórico de iteração
- CI para exp(B): 95 %

Exibir

- Em cada passo
- Na última passo

Probabilidade para Stepwise

Entrada: 0.05 Remoção: 0.10 Corte de classificação: 0.5

Máximo de iterações: 20

Preservar memória para análises complexas ou grandes conjuntos de dados

Incluir constante no modelo

Continuar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode: ON

Implementação do Modelo de Regressão Logística

5. Interpretar os resultados do modelo de regressão logística

Regressão Logística

- O resultado é publicado no 'Visualizador de Resultados'
- Temos três grupos de resultados
 - Tabelas com informações sobre as variáveis incluídas no modelo
 - Tabelas com os resultados de um modelo de regressão sem variáveis independentes (Bloco 0)
 - Tabelas com os resultados do modelo de regressão que estimamos (Bloco 1)

LOGISTIC REGRESSION VARIABLES satisfied
/METHOD=ENTER sex age education
/CONTRAST (sex)=Indicator(1)
/CONTRAST (education)=indicator(1)
/PRINT=GOODFIT CI(95)
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Regressão logística

Resumo de processamento do caso

Casos não ponderados ^a		N	Porcentagem
Casos selecionados	Incluído na análise	5000	100.0
	Casos omissos	0	.0
Total		5000	100.0
Casos não selecionados		0	.0
Total		5000	100.0

a. Se a ponderação estiver em vigor, veja a tabela de classificação para o número total de casos.

Codificação de variável dependente

Valor original	Valor interno
.00	0
1.00	1

Codificações de variáveis categóricas

		Frequência	Codificação de parâmetro			
			(1)	(2)	(3)	(4)
education	ISCED 1	979	.000	.000	.000	.000
	ISCED 2	1023	1.000	.000	.000	.000
	ISCED 3	1007	.000	1.000	.000	.000

Regressão Logística

- O resultado é publicado no 'Visualizador de Resultados'
- Estas tabelas dizem-nos:
 - O número de observações que foi integrado no modelo
 - Como está operacionalizada a variável dependente



*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Inserir Formatar Analisar Gráficos Utilitários Extensões Janela Ajuda

LOGISTIC REGRESSION VARIABLES satisfied
/METHOD=ENTER sex age education
/CONTRAST (sex)=Indicator(1)
/CONTRAST (education)=indicator(1)
/PRINT=GOODFIT CI(95)
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Regressão logística

Resumo de processamento do caso

Casos não ponderados ^a		N	Porcentagem
Casos selecionados	Incluído na análise	5000	100.0
	Casos omissos	0	.0
Total		5000	100.0
Casos não selecionados		0	.0
Total		5000	100.0

a. Se a ponderação estiver em vigor, veja a tabela de classificação para o número total de casos.

Codificação de variável dependente

Valor original	Valor interno
.00	0
1.00	1

Codificações de variáveis categóricas

	Frequência	Codificação de parâmetro				
		(1)	(2)	(3)	(4)	
education	ISCED 1	979	.000	.000	.000	.000
	ISCED 2	1023	1.000	.000	.000	.000
	ISCED 3	1007	.000	1.000	.000	.000



Regressão Logística

- O resultado é publicado no ‘Visualizador de Resultados’
 - Estas tabelas dizem-nos:
 - O número de observações que foi integrado no modelo
 - Como está codificada a variável dependente
 - Como estão codificadas a variáveis independentes *
- * Podemos assim confirmar as categorias de referência – “tudo zeros”

A

B

C

Visualizador do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Inserir Formatar Analisar Gráficos Utilitários Extensões Janela Ajuda

Saída

- Log
- Regressão logística
 - Título
 - Observações
 - Resumo de processamento do caso
 - Codificação de variável dependente
 - Codificações de variáveis categóricas**
 - Bloco 0: Bloco Inicial
 - Título
 - Tabela de Classificação
 - Variáveis na equação
 - Variáveis não presentes na equação
 - Bloco 1: Método = Enter
 - Título
 - Testes de Omnibus do Modelo de
 - Resumo do modelo
 - Teste de Hosmer e Lemeshow
 - Tabela de contingência para teste
 - Tabela de Classificação
 - Variáveis na equação

a. Se a ponderação estiver em vigor, veja a tabela de classificação para o número total de casos.

Codificação de variável dependente

Valor original	Valor interno
.00	0
1.00	1

Codificações de variáveis categóricas

		Frequência	Codificação de parâmetro			
			(1)	(2)	(3)	(4)
education	ISCED 1	979	.000	.000	.000	.000
	ISCED 2	1023	1.000	.000	.000	.000
	ISCED 3	1007	.000	1.000	.000	.000
	ISCED 4	994	.000	.000	1.000	.000
	ISCED 5	997	.000	.000	.000	1.000
sex	Female	2514	.000			
	Male	2486	1.000			

Bloco 0: Bloco Inicial

Tabela de Classificação^{a,b}

Observado		Previsto		Porcentagem correta	
		Satisfied	1.00		
Passo 0	Satisfied	.00	2648	0	100.0
	1.00	2352	0	0	.0
Porcentagem global					53.0

a. A constante está incluída no modelo

O processador do IBM SPSS Statistics está pronto | Unicode:ON | H: 2,47, W: 5,22 in

Regressão Logística

- O resultado é publicado no 'Visualizador de Resultados'
- Esta tabela apresenta o resultado de um teste de Chi-Quadrado que determina se, comparado com um modelo sem variáveis independentes (só a constante), este modelo é estatisticamente significativo.
- Interpretação:
 - 'Sig.' < 0.01 / 0.05 / 0.1 o modelo é estatisticamente significativo
- Neste caso, o nosso modelo é estatisticamente significativo a um nível de confiança de 99%

*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Inserir Formatar Analisar Gráficos Utilitários Extensões Janela Ajuda

Score df Sig.

Passo 0	Variáveis	sex(1)	.667	1	.414
	age	3.450	1	.063	
	education	156.398	4	.000	
	education(1)	24.324	1	.000	
	education(2)	.110	1	.740	
	education(3)	1.530	1	.216	
	education(4)	120.837	1	.000	
	Estadísticas globais	157.196	6	.000	

Bloco 1: Método = Enter

Testes de Omnibus do Modelo de Coeficientes

Passo 1	Passo	Qui-quadrado	df	Sig.
	Bloco	158.358	6	.000
	Modelo	158.358	6	.000

Resumo do modelo

Passo	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	6755.581 ^a	.031	.042

a. Estimação finalizada no número de iteração 3 porque as estimativas de parâmetro mudaram foram alteradas para menos de .001.

Teste de Hosmer e Lemeshow

Passo	Qui-quadrado	df	Sig.
1	2.324	8	.969

O processador do IBM SPSS Statistics está pronto | Unicode:ON | H: 1,33, W: 3,81 in

Regressão Logística

- O resultado é publicado no 'Visualizador de Resultados'
- Esta tabela apresenta os resultados de um conjunto de testes, nomeadamente o teste R^2 Nagelkerke, que é um pseudo- R^2 .
- Não pode ser interpretado de forma literal como o R^2
- Interpretação:

O model explica aproximadamente 40% da variação na variável dependente



Estatísticas globais

	Escore	df	Sig.
Passo 0			
sex(1)	.667	1	.414
age	3.450	1	.063
education	156.398	4	.000
education(1)	24.324	1	.000
education(2)	.110	1	.740
education(3)	1.530	1	.216
education(4)	120.837	1	.000
Estatísticas globais	157.196	6	.000

Bloco 1: Método = Enter

Testes de Omnibus do Modelo de Coeficientes

	Qui-quadrado	df	Sig.
Passo 1			
Passo	158.358	6	.000
Bloco	158.358	6	.000
Modelo	158.358	6	.000

Resumo do modelo

Passo	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	6755.581 ^a	.031	.042

a. Estimação finalizada no número de iteração 3 porque as estimativas de parâmetro mudaram foram alteradas para menos de .001.

Teste de Hosmer e Lemeshow

Passo	Qui-quadrado	df	Sig.
1	2.324	8	.969

Regressão Logística

- O resultado é publicado no 'Visualizador de Resultados'
- Esta tabela dois indicadores da capacidade preditiva do modelo.
 - **Especificidade:** A proporção de casos negativos correctamente identificados (verdadeiros negativos)
 - **Sensibilidade:** A proporção de casos positivos correctamente identificados (verdadeiros positivos)
 - **Precisão:** A proporção de casos correctamente identificados

A

B

C

*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Inserir Formatar Analisar Gráficos Utilitários Extensões Janela Ajuda

Log
Regressão logística
Título
Observações
Resumo de processamento do caso
Codificação de variável dependente
Codificações de variáveis categóricas
Bloco 0: Bloco Inicial
Título
Tabela de Classificação
Variáveis na equação
Variáveis não presentes na equação
Bloco 1: Método = Enter
Título
Testes de Omnibus do Modelo de Coeficientes
Resumo do modelo
Teste de Hosmer e Lemeshow
Tabela de contingência para teste de Hosmer
Tabela de Classificação
Variáveis na equação

5	268	271.102	233	229.898	501
6	260	261.169	233	231.831	493
7	262	258.926	238	241.074	500
8	254	254.133	247	246.867	501
9	202	192.149	298	307.851	500
10	180	191.288	338	326.712	518

Tabela de Classificação^a

Observado		Previsto		Porcentagem correta
		Satisfied	1.00	
Passo 1	Satisfied	.00	2259	85.3
	1.00	1708	644	27.4
Porcentagem global				58.1

a. O valor de recorte é .500

Variáveis na equação

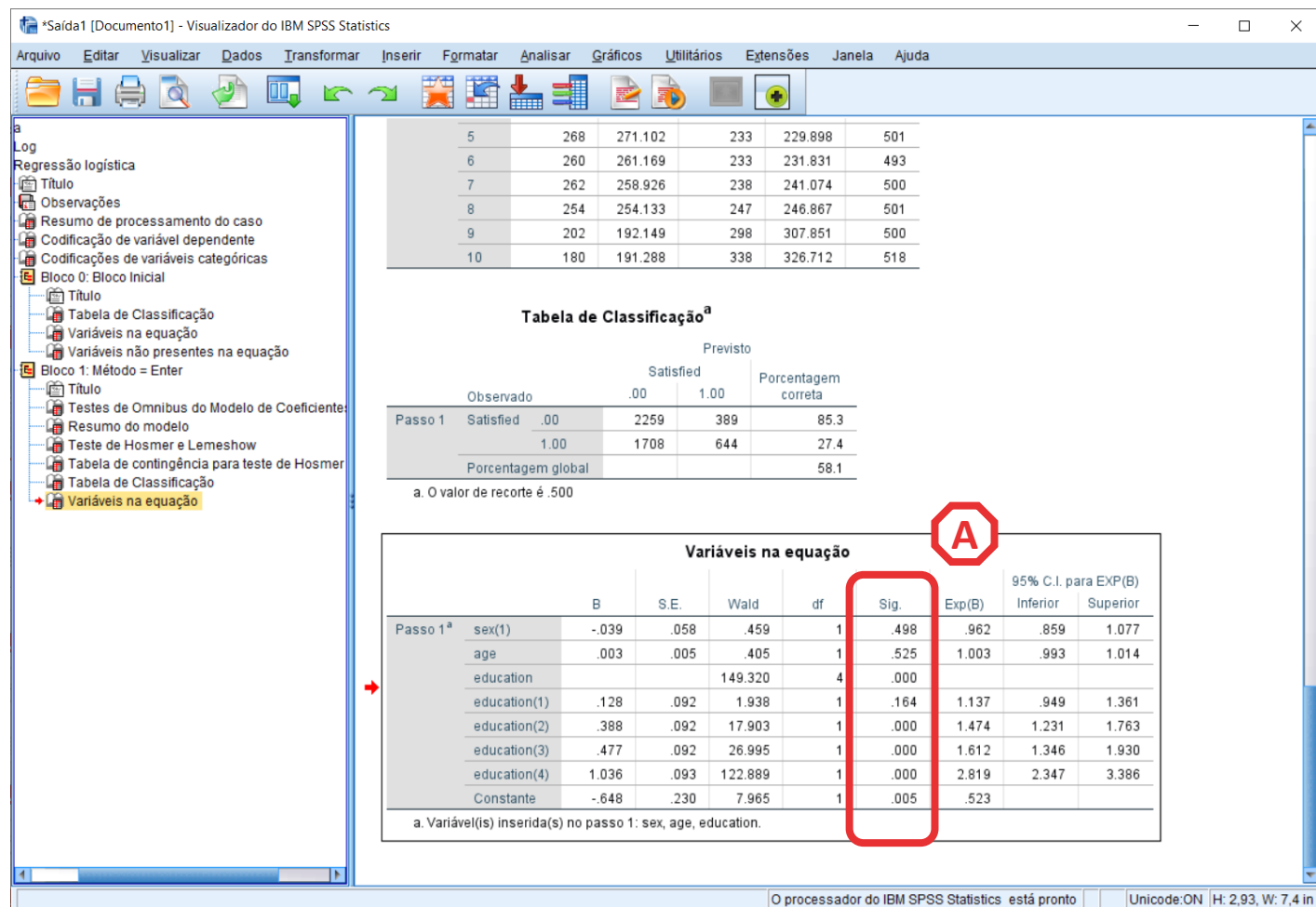
Passo 1 ^a		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. para EXP(B)	
								Inferior	Superior
	sex(1)	-.039	.058	.459	1	.498	.962	.859	1.077
	age	.003	.005	.405	1	.525	1.003	.993	1.014
	education			149.320	4	.000			
	education(1)	.128	.092	1.938	1	.164	1.137	.949	1.361
	education(2)	.388	.092	17.903	1	.000	1.474	1.231	1.763
	education(3)	.477	.092	26.995	1	.000	1.612	1.346	1.930
	education(4)	1.036	.093	122.889	1	.000	2.819	2.347	3.386
	Constante	-.648	.230	7.965	1	.005	.523		

a. Variável(is) inserida(s) no passo 1: sex, age, education.

O processador do IBM SPSS Statistics está pronto | Unicode:ON | H: 2,02, W: 4,58 in

Regressão Logística

- O resultado é publicado no ‘Visualizador de Resultados’
- Esta tabela apresenta a informação sobre o efeito das variáveis independentes sobre a variável dependente.
- A primeira coisa a olhar é se a relação das IV’s com a DV é estatisticamente significativa, i.e.
 - ‘Sig.’ < 0.01 / 0.05 / 0.1
- Neste caso, observa-se que as variáveis ‘sex’ e ‘age’ não são significativas.



*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Inserir Formatar Analisar Gráficos Utilitários Extensões Janela Ajuda

Log
Regressão logística
Título
Observações
Resumo de processamento do caso
Codificação de variável dependente
Codificações de variáveis categóricas
Bloco 0: Bloco Inicial
Título
Tabela de Classificação
Variáveis na equação
Variáveis não presentes na equação
Bloco 1: Método = Enter
Título
Testes de Omnibus do Modelo de Coeficientes
Resumo do modelo
Teste de Hosmer e Lemeshow
Tabela de contingência para teste de Hosmer
Variáveis na equação

5	268	271.102	233	229.898	501
6	260	261.169	233	231.831	493
7	262	258.926	238	241.074	500
8	254	254.133	247	246.867	501
9	202	192.149	298	307.851	500
10	180	191.288	338	326.712	518

Tabela de Classificação^a

	Observado	Previsto		Porcentagem correta	
		Satisfied	1.00		
Passo 1	Satisfied	.00	2259	389	85.3
	1.00	1708	644	27.4	
Porcentagem global					58.1

a. O valor de corte é .500

Variáveis na equação

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. para EXP(B)		
							Inferior	Superior	
Passo 1 ^a	sex(1)	-.039	.058	.459	1	.498	.962	.859	1.077
	age	.003	.005	.405	1	.525	1.003	.993	1.014
	education			149.320	4	.000			
	education(1)	.128	.092	1.938	1	.164	1.137	.949	1.361
	education(2)	.388	.092	17.903	1	.000	1.474	1.231	1.763
	education(3)	.477	.092	26.995	1	.000	1.612	1.346	1.930
	education(4)	1.036	.093	122.889	1	.000	2.819	2.347	3.386
	Constante	-.648	.230	7.965	1	.005	.523		

a. Variável(is) inserida(s) no passo 1: sex, age, education.

O processador do IBM SPSS Statistics está pronto | Unicode:ON | H: 2,93, W: 7,4 in

Regressão Logística

- O resultado é publicado no 'Visualizador de Resultados'
- De seguida, queremos ir ver os coeficientes do modelo:
 - Ser homem diminui o logaritmo da chance (log odds) de estar satisfeito com as condições de trabalho em 0.39...
 - Por cada ano adicional de vida, o logaritmo da chance (log odds) de estar satisfeito com as condições de trabalho aumenta em 0.03.
 - Comparado com ter apenas o ensino básico (ISCED1), ter uma licenciatura aumenta o logaritmo da chance (log odds) de estar satisfeito com as condições de trabalho em 1.036.

The screenshot shows the IBM SPSS Statistics interface. The left pane displays a tree view of the analysis results, with 'Variáveis na equação' selected. The main window displays the following tables:

Table 1: Observed vs. Predicted

Observado	Previsto	Satisfied		Porcentagem correta
		.00	1.00	
Satisfied .00		2259	389	85.3
1.00		1708	644	27.4
Porcentagem global				58.1


a. O valor de recorte é .500

Table 2: Variáveis na equação

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. para EXP(B)	
							Inferior	Superior
Passo 1 ^a sex(1)	-.039	.058	.459	1	.498	.962	.859	1.077
age	.003	.005	.405	1	.525	1.003	.993	1.014
education			149.320	4	.000			
education(1)	.128	.092	1.938	1	.164	1.137	.949	1.361
education(2)	.388	.092	17.903	1	.000	1.474	1.231	1.763
education(3)	.477	.092	26.995	1	.000	1.612	1.346	1.930
education(4)	1.036	.093	122.889	1	.000	2.819	2.347	3.386
Constante	-.648	.230	7.965	1	.005	.523		

a. Variável(is) inserida(s) no passo 1: sex, age, education.

Regressão Logística

- ISTO NÃO É NADA INTUITIVO!!
- Por isso é que nos interessa olhar para os odd-ratios (razão de chances) 
- Interpretação:
 - Um odds ratio maior que 1 significa que um aumento em x leva a um aumento nas chances de $y = 1$.
 - Um odds ratio menor que 1 significa que um aumento em x leva a uma diminuição nas chances de $y = 1$.

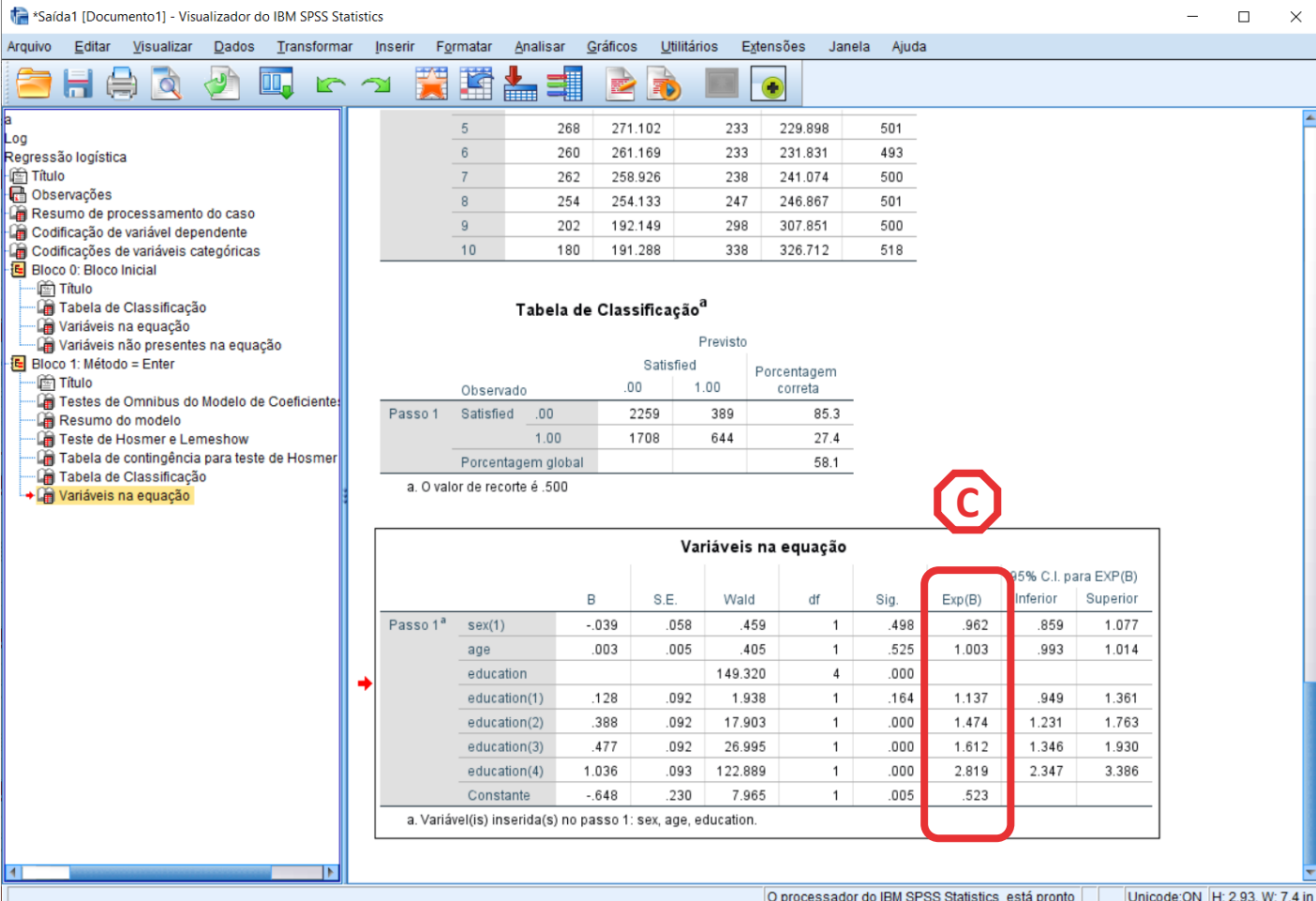


Tabela de Classificação^a

		Previsto		Porcentagem correta
		Satisfied	1.00	
Passo 1	Satisfied	.00	2259	85.3
	1.00	1708	644	27.4
Porcentagem global				58.1

a. O valor de corte é .500

Variáveis na equação

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. para EXP(B)	
								Inferior	Superior
Passo 1 ^a	sex(1)	-.039	.058	.459	1	.498	.962	.859	1.077
	age	.003	.005	.405	1	.525	1.003	.993	1.014
	education			149.320	4	.000			
	education(1)	.128	.092	1.938	1	.164	1.137	.949	1.361
	education(2)	.388	.092	17.903	1	.000	1.474	1.231	1.763
	education(3)	.477	.092	26.995	1	.000	1.612	1.346	1.930
	education(4)	1.036	.093	122.889	1	.000	2.819	2.347	3.386
	Constante	-.648	.230	7.965	1	.005	.523		

a. Variável(is) inserida(s) no passo 1: sex, age, education.

Regressão Logística

- Melhor ainda, podemos calcular a mudança percentual nas odds por cada unidade adicional na IV:

$$\Delta \% \text{ das Odds} = 100 * (\text{Odds Ratio} - 1)$$

- Exemplo:

- As chances de estar satisfeito/a com as condições de trabalho diminuem em - 3.8% [=100 * (0.962 - 1)] - entre os homens, em comparação com mulheres. [ainda que $p > 0.05$]
- Cada aumento de um ano na idade leva a um aumento de 1,3% - [= 100 *(1,013- 1)] das chances de estar satisfeito/a com as condições de trabalho. [ainda que $p > 0.05$]

Tabela de Classificação^a

		Previsto		Porcentagem correta	
		Satisfied	1.00		
Passo 1	Satisfied	.00	2259	389	85.3
	1.00	1708	644		27.4
Porcentagem global					58.1

a. O valor de recorte é .500

Variáveis na equação

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. para EXP(B)	
								Inferior	Superior
Passo 1 ^a	sex(1)	-.039	.058	.459	1	.498	.962	.859	1.077
	age	.003	.005	.405	1	.525	1.003	.993	1.014
	education			149.320	4	.000			
	education(1)	.128	.092	1.938	1	.164	1.137	.949	1.361
	education(2)	.388	.092	17.903	1	.000	1.474	1.231	1.763
	education(3)	.477	.092	26.995	1	.000	1.612	1.346	1.930
	education(4)	1.036	.093	122.889	1	.000	2.819	2.347	3.386
	Constante	-.648	.230	7.965	1	.005	.523		

a. Variável(is) inserida(s) no passo 1: sex, age, education.

Regressão Logística

- Melhor ainda, podemos calcular a mudança percentual nas odds por cada unidade adicional na IV:

$$\Delta \% \text{ das Odds} = 100 * (\text{Odds Ratio} - 1)$$

- Exemplo:

- O nível de educação aumenta as probabilidade dos colaboradores se sentirem satisfeitos. Em comparação com o nível mais baixo (ISCED1), o nível de educação de nível ISCED3 aumenta as chances relativas de satisfação em 47 %, o nível de educação de nível ISCED4 em 61 %, e o nível de educação de nível ISCED5 em 182%. Estima-se que percentagem de satisfação entre os mais qualificados é quase 3 vezes do que a entre os menos qualificados.

The screenshot displays the IBM SPSS Statistics interface with the following components:

- Log**: A table showing the log-likelihood values for each step of the model.
- Tabela de Classificação^a**: A confusion matrix showing the relationship between observed and predicted satisfaction levels.
- Variáveis na equação**: A table of coefficients and statistics for the variables in the model. The 'Exp(B)' column, representing the Odds Ratio, is highlighted with a red box.

Passo	Log	Log	Log	Log	Log
5	268	271.102	233	229.898	501
6	260	261.169	233	231.831	493
7	262	258.926	238	241.074	500
8	254	254.133	247	246.867	501
9	202	192.149	298	307.851	500
10	180	191.288	338	326.712	518

Observado	Previsto	Satisfied		Porcentagem correta
		.00	1.00	
Passo 1 Satisfied	.00	2259	389	85.3
	1.00	1708	644	27.4
Porcentagem global				58.1

a. O valor de recorte é .500

Passo 1 ^a	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. para EXP(B)	
								Inferior	Superior
	sex(1)	-.039	.058	.459	1	.498	.962	.859	1.077
	age	.003	.005	.405	1	.525	1.003	.993	1.014
	education			149.320	4	.000			
	education(1)	.128	.092	1.938	1	.164	1.137	.949	1.361
	education(2)	.388	.092	17.903	1	.000	1.474	1.231	1.763
	education(3)	.477	.092	26.995	1	.000	1.612	1.346	1.930
	education(4)	1.036	.093	122.889	1	.000	2.819	2.347	3.386
	Constante	-.648	.230	7.965	1	.005	.523		

a. Variável(is) inserida(s) no passo 1: sex, age, education.

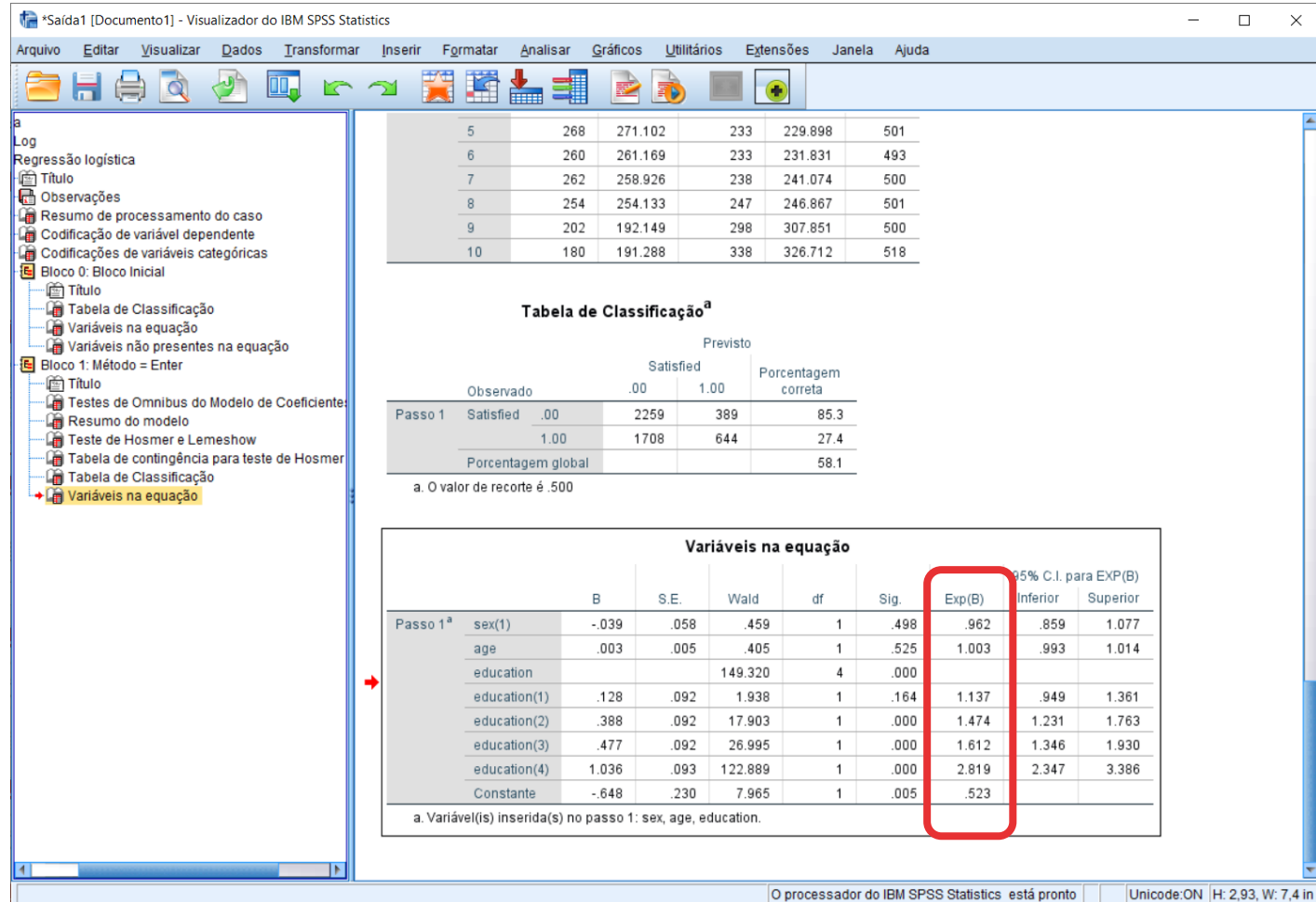
Reportar os resultados de um modelo de regressão linear

Procurou-se compreender o efeito das variáveis género, idade e educação nas probabilidades dos trabalhadores se sentirem satisfeitos com as condições de trabalho, com recurso a um modelo de regressão logística. O modelo estatisticamente significativo e portanto útil para estimar a variável dependente ($\chi^2_{(7)} = 159,422, p < 0,05$). Ainda assim, o modelo apresenta um R^2 Nagelkerke de 0,042, indicando que o modelo tem uma capacidade limitada de explicar a variabilidade na variável dependente. Comparando as previsões do modelo em relação aos resultados da amostra, o modelo apresenta uma precisão de 57,8%, sendo o modelo mais específico do que que sensível. Assumindo um nível de confiança de 95%, concluímos que o género e a idade não influenciam as chances dos trabalhadores se sentirem satisfeitos ($p > 0,05$). Em contrapartida o nível de escolaridade mostra-se relevante ($p < 0,05$). O nível de educação aumenta as probabilidades dos colaboradores se sentirem satisfeitos. Em comparação com o nível mais baixo (ISCED1), o nível de educação de nível ISCED 3 aumenta as chances relativas de satisfação em 47 %, o nível de educação de nível ISCED 4 em 61 %, e o nível de educação de nível ISCED5 em 182%. De acordo com o modelo, estima-se que a percentagem de satisfação entre os mais qualificados (ISCED 5) é quase 3 vezes do que a entre os menos qualificados (ISCED 1) (OR=2,82; IC95% [2,35-3,39]).

Regressão Logística

- Exercício
- Usar os slides para **gerar, interpretar e reportar** um novo modelo de regressão com
- VD: satisfação
- VI: age, sex, departement

Será que as % de estar satisfeito dependem do departamento?



*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Inserir Formatar Analisar Gráficos Utilitários Extensões Janela Ajuda

Log
Regressão logística
Título
Observações
Resumo de processamento do caso
Codificação de variável dependente
Codificações de variáveis categóricas
Bloco 0: Bloco Inicial
Título
Tabela de Classificação
Variáveis na equação
Variáveis não presentes na equação
Bloco 1: Método = Enter
Título
Testes de Omnibus do Modelo de Coeficientes
Resumo do modelo
Teste de Hosmer e Lemeshow
Tabela de contingência para teste de Hosmer
Tabela de Classificação
Variáveis na equação

	5	6	7	8	9	10
	268	260	262	254	202	180
	271.102	261.169	258.926	254.133	192.149	191.288
	233	233	238	247	298	338
	229.898	231.831	241.074	246.867	307.851	326.712
	501	493	500	501	500	518

Tabela de Classificação^a

	Observado	Previsto		Porcentagem correta	
		Satisfied	1.00		
Passo 1	Satisfied	.00	2259	389	85.3
	1.00	1708	644	27.4	
Porcentagem global					58.1

a. O valor de recorte é .500

Variáveis na equação

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. para EXP(B)	
								Inferior	Superior
Passo 1 ^a	sex(1)	-.039	.058	.459	1	.498	.962	.859	1.077
	age	.003	.005	.405	1	.525	1.003	.993	1.014
	education			149.320	4	.000			
	education(1)	.128	.092	1.938	1	.164	1.137	.949	1.361
	education(2)	.388	.092	17.903	1	.000	1.474	1.231	1.763
	education(3)	.477	.092	26.995	1	.000	1.612	1.346	1.930
	education(4)	1.036	.093	122.889	1	.000	2.819	2.347	3.386
	Constante	-.648	.230	7.965	1	.005	.523		

a. Variável(is) inserida(s) no passo 1: sex, age, education.

O processador do IBM SPSS Statistics está pronto | Unicode:ON | H: 2,93, W: 7,4 in