

Aula 10:

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

Validação de Modelos de Regressão Logística

Docente: Daniela Craveiro
dcraveiro@iseg.ulisboa.pt

- **Na Aula Anterior**
 - **Aprendemos a implementar e interpretar o resultado de um modelo de regressão logística**
- **Objetivos da Aula**
 - **Parte Teórica**
 - **Perceber qual a necessidade de fazermos diagnósticos aos pressupostos do nosso modelo de regressão**
 - **Saber quais são os pressupostos do modelo de regressão logística**
 - **Saber como, com a ajuda de gráficos e testes estatísticos, podemos conferir se os pressupostos do modelo estão a ser cumpridos**
 - **Parte Prática**
 - **Saber implementar os estudo dos pressupostos do modelo no SPSS**

Pressupostos do modelo de regressão logística

	PRESSUPOSTOS	DEFINIÇÃO	FORMA DE VALIDAÇÃO
I	Ausência de células vazias	Quando cruzamos a variável dependente com as variáveis independentes, não existem células vazias (i.e., sem observações).	<ul style="list-style-type: none">• Tabelas de Dupla Entrada

Pressupostos do modelo de regressão logística

	PRESSUPOSTOS	DEFINIÇÃO	FORMA DE VALIDAÇÃO
I	Ausência de células vazias	Quando cruzamos a variável dependente com as variáveis independentes, não existem células vazias (i.e., sem observações).	<ul style="list-style-type: none">• Tabelas de Dupla Entrada
II	Independência dos Erros	As observações são independentes entre si.	<ul style="list-style-type: none">• Só relevante se os dados são dependentes no tempo ou no espaço.

Pressupostos do modelo de regressão logística

	PRESSUPOSTOS	DEFINIÇÃO	FORMA DE VALIDAÇÃO
I	Ausência de células vazias	Quando cruzamos a variável dependente com as variáveis independentes, não existem células vazias (i.e., sem observações).	<ul style="list-style-type: none"> • Tabelas de Dupla Entrada
II	Independência dos Erros	As observações são independentes entre si.	<ul style="list-style-type: none"> • Só relevante se os dados são dependentes no tempo ou no espaço.
III	Ausência de Observações Influentes	Não existem observações que tenham uma influência anormal nos resultados do modelo.	<ul style="list-style-type: none"> • Cook's Distance

Pressupostos do modelo de regressão logística

	PRESSUPOSTOS	DEFINIÇÃO	FORMA DE VALIDAÇÃO
I	Ausência de células vazias	Quando cruzamos a variável dependente com as variáveis independentes, não existem células vazias (i.e., sem observações).	<ul style="list-style-type: none"> • Tabelas de Dupla Entrada
II	Independência dos Erros	As observações são independentes entre si.	<ul style="list-style-type: none"> • Só relevante se os dados são dependentes no tempo ou no espaço.
III	Ausência de Observações Influentes	Não existem observações que tenham uma influência anormal nos resultados do modelo.	<ul style="list-style-type: none"> • Cook's Distance
IV	Ausência de 'Separação Completa'	Não é possível a prever perfeitamente todos valores da variável dependente partir dos valores de uma (ou mais) variáveis independentes (<i>separação completa</i>).	<ul style="list-style-type: none"> • Gráfico de Dispersão

Pressupostos do modelo de regressão logística

	PRESSUPOSTOS	DEFINIÇÃO	FORMA DE VALIDAÇÃO
I	Ausência de células vazias	Quando cruzamos a variável dependente com as variáveis independentes, não existem células vazias (i.e., sem observações).	<ul style="list-style-type: none"> Tabelas de Dupla Entrada
II	Independência dos Erros	As observações são independentes entre si.	<ul style="list-style-type: none"> Só relevante se os dados são dependentes no tempo ou no espaço.
III	Ausência de Observações Influentes	Não existem observações que tenham uma influência anormal nos resultados do modelo.	<ul style="list-style-type: none"> Cook's Distance
IV	Ausência de 'Separação Completa'	Não é possível a prever perfeitamente todos valores da variável dependente partir dos valores de uma (ou mais) variáveis independentes (<i>separação completa</i>).	<ul style="list-style-type: none"> Gráfico de Dispersão
V	Ausência de multicolinearidade perfeita	As variáveis independentes não estão perfeitamente correlacionadas entre si.	<ul style="list-style-type: none"> Análise da Matriz de Correlações

Pressupostos do modelo de regressão logística

	PRESSUPOSTOS	DEFINIÇÃO	FORMA DE VALIDAÇÃO
I	Ausência de células vazias	Quando cruzamos a variável dependente com as variáveis independentes, não existem células vazias (i.e., sem observações).	<ul style="list-style-type: none"> Tabelas de Dupla Entrada
II	Independência dos Erros	As observações são independentes entre si.	<ul style="list-style-type: none"> Só relevante se os dados são dependentes no tempo ou no espaço.
III	Ausência de Observações Influentes	Não existem observações que tenham uma influência anormal nos resultados do modelo.	<ul style="list-style-type: none"> Cook's Distance
IV	Ausência de 'Separação Completa'	Não é possível a prever perfeitamente todos valores da variável dependente partir dos valores de uma (ou mais) variáveis independentes (<i>separação completa</i>).	<ul style="list-style-type: none"> Gráfico de Dispersão
V	Ausência de multicolinearidade perfeita	As variáveis independentes não estão perfeitamente correlacionadas entre si.	<ul style="list-style-type: none"> Análise da Matriz de Correlações
VI	Linearidade	O efeito das variáveis (independentes) de razão nos logaritmos da variável dependente é linear e aditivo.	<ul style="list-style-type: none"> Gráfico de Dispersão Teste do efeito da interação VI*ln(VI)

Validação do Modelo de Regressão Logística

- 1. Estimar o modelo de regressão com os diagnósticos*
- 2. Avaliação do Pressuposto I: Ausência de células vazias*
- 3. Avaliação do Pressuposto III: Ausência de Observações Influentes*
- 4. Avaliação do Pressuposto IV: Ausência de 'Separação Completa'*
- 5. Avaliação do Pressuposto V: Ausência de Multicolinearidade Perfeita*
- 6. Avaliação do Pressuposto VI: Linearidade*

Validação do Modelo de Regressão Logística

1. Estimar o modelo de regressão com os diagnósticos

- **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...**
 - ... incluindo a produção de diagnósticos para a avaliação da ausência de observações influentes.**

Regressão Logística

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



*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 29668.4665929268 8

	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

Relatórios

Estadística Descritiva

Estadísticas Bayesianas

Tabelas

Comparar Médias

Modelo Linear Geral

Modelos lineares generalizados

Modelos mistos

Correlacionar

Regressão

Log linear

Redes neurais

Classificar

Redução de dimensão

Escala

Modelagem Linear Automática...

Linear...

Curva de estimação...

Quadrados Mínimos Parciais...

Logística binária...



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'



The screenshot shows the IBM SPSS Statistics interface with the 'Regressão logística' dialog box open. The 'Dependente:' field is empty, and the 'Covariáveis:' field is empty. The 'Satisfied [satisfied]' variable is selected in the list on the left. The 'Método:' is set to 'Inserir'. The 'Variável de seleção:' field is empty. The 'OK' button is highlighted. Red octagons with letters A, B, and C are overlaid on the image to indicate specific steps: A points to the 'Analisar' menu, B points to the 'Satisfied [satisfied]' variable, and C points to the 'Dependente:' field.

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'



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

id	year	y_wage	department	hrs_week
1	2018.00	22153.05	6	38.85
2	2018.00	27809.35	5	39.18
3	2018.00	29509.03	2	40.24
4	2018.00	21500.50	1	38.62
5	2018.00	16786.81	2	38.86
6	2018.00	31153.80	4	41.00
7	2018.00	20867.09	1	38.90
8	2018.00	14899.42	3	41.70
9	2018.00	20882.25	1	40.95
10	2018.00	29631.50	5	41.07
11	2018.00	22096.46	2	40.61
12	2018.00	28315.75	1	41.16
13	2018.00	31361.65	1	39.90
14	2018.00	30477.37	4	40.59
15	2018.00	32136.62	2	40.05

Regressão Logística

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



*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:

- sex
- age
- education

Covariáveis categóricas:

Alterar Contraste

Contraste: Indicador Alterar

Categoria de referência: Último Primeiro

Continuar Cancelar Ajuda

OK Colar Reconfigurar Cancelar Ajuda

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

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

Regressão Logística

- Selecionar a variável 'sex'
- Colocar na caixa 'Covariáveis categóricas'
- Definir a categoria 'male' (valor 1 *) como categoria de referência na variável 'sex'

* neste caso, escolhemos o valor 1 que é o mais baixo (i.e., o primeiro)

- Carregar no botão 'Alterar'



*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

- Selecionar a variável 'sex'
- Colocar na caixa 'Covariáveis categóricas'
- Definir a categoria 'male' (valor 1 *) como categoria de referência na variável 'sex'

* neste caso, escolhemos o valor 1 que é o mais baixo (i.e., o primeiro)

- Carregar no botão 'Alterar'

Exercício: 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 Alterar

Categoria de referência: Último Primeiro

Continuar Cancelar Ajuda

OK Cancelar Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

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

Regressão Logística

- Clicar no botão 'Opções'



*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	2
10	17	3
11	14	2
12	7	2
13	13	2
14	7	2
15	11	3

Regressão logística

Dependente: Satisfied [satisfied]

Bloco 1 de 1

Anterior Próximo

Bloco 1 de 1

sex(Cat)
age
department(Cat)

Método: Inserir

Variável de seleção:

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'
- Selecionar a opção 'Qualidade do ajuste de Hosmer-Lemeshow'
- Selecionar a opção 'CI para exp(B)'
- Selecionar 'Continuar'

J

K

K

L

L

M

M

*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

id	departme	nt	hrs_week
1	6		38.85
2	5		39.18
3	2		40.24
4	1		38.62
5	2		38.86
6	4		41.00
7	1		38.90
8	3		41.70
9	1		40.95
10	5		41.07
11	2		40.61
12	1		41.16
13	1		39.90
14	4		40.59
15	2		40.05

Regressão Logística

- Clicar no botão 'Salvar'



*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	2
10	17	3
11	14	2
12	7	2
13	13	2
14	7	2
15	11	3

Regressão logística

Dependente: Satisfied [satisfied]

Bloco 1 de 1

Anterior Próximo

Bloco 1 de 1

sex(Cat)
age
department(Cat)

Método: Inserir

Variável de seleção:

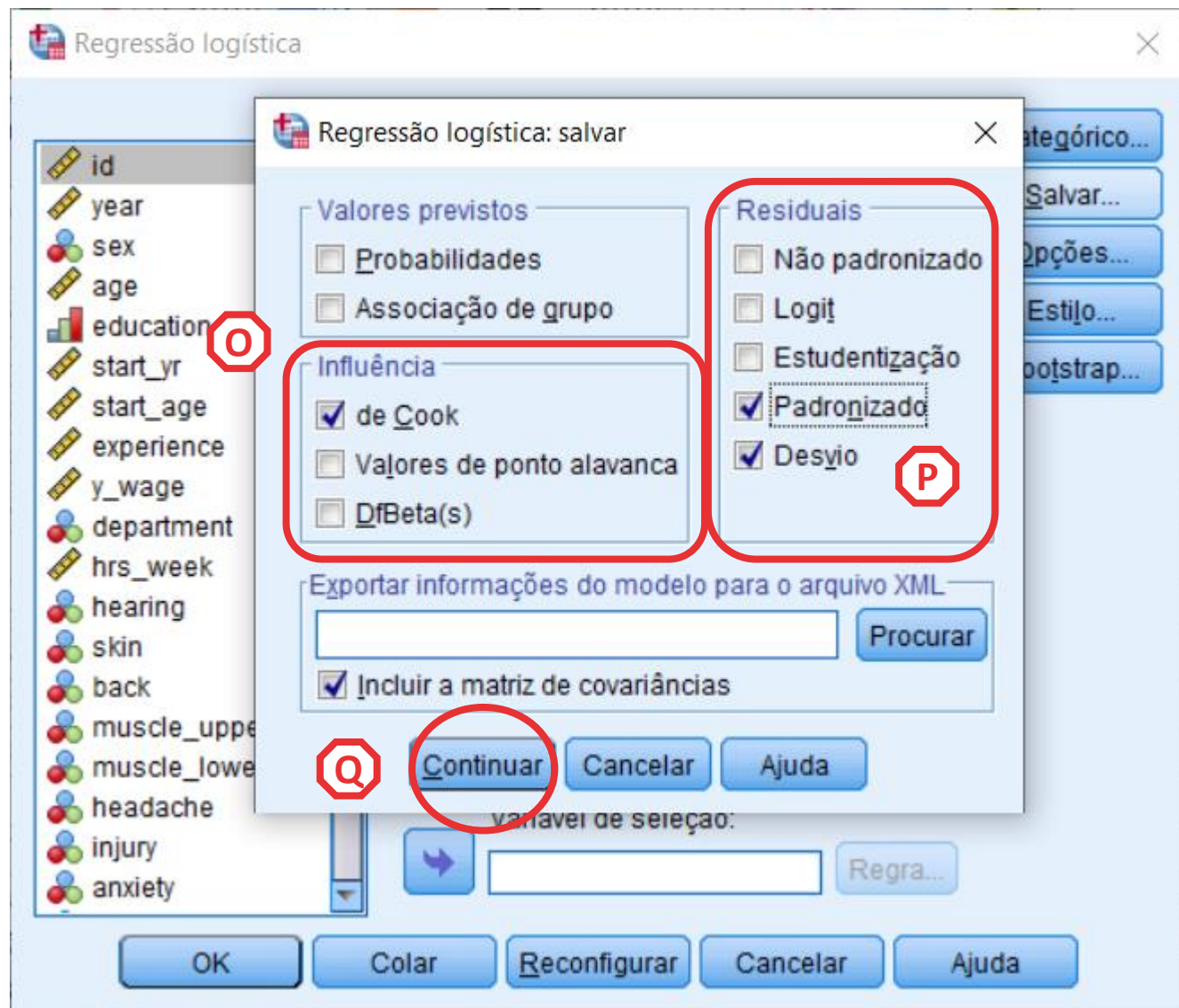
OK Colar Reconfigurar Cancelar Ajuda

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:ON

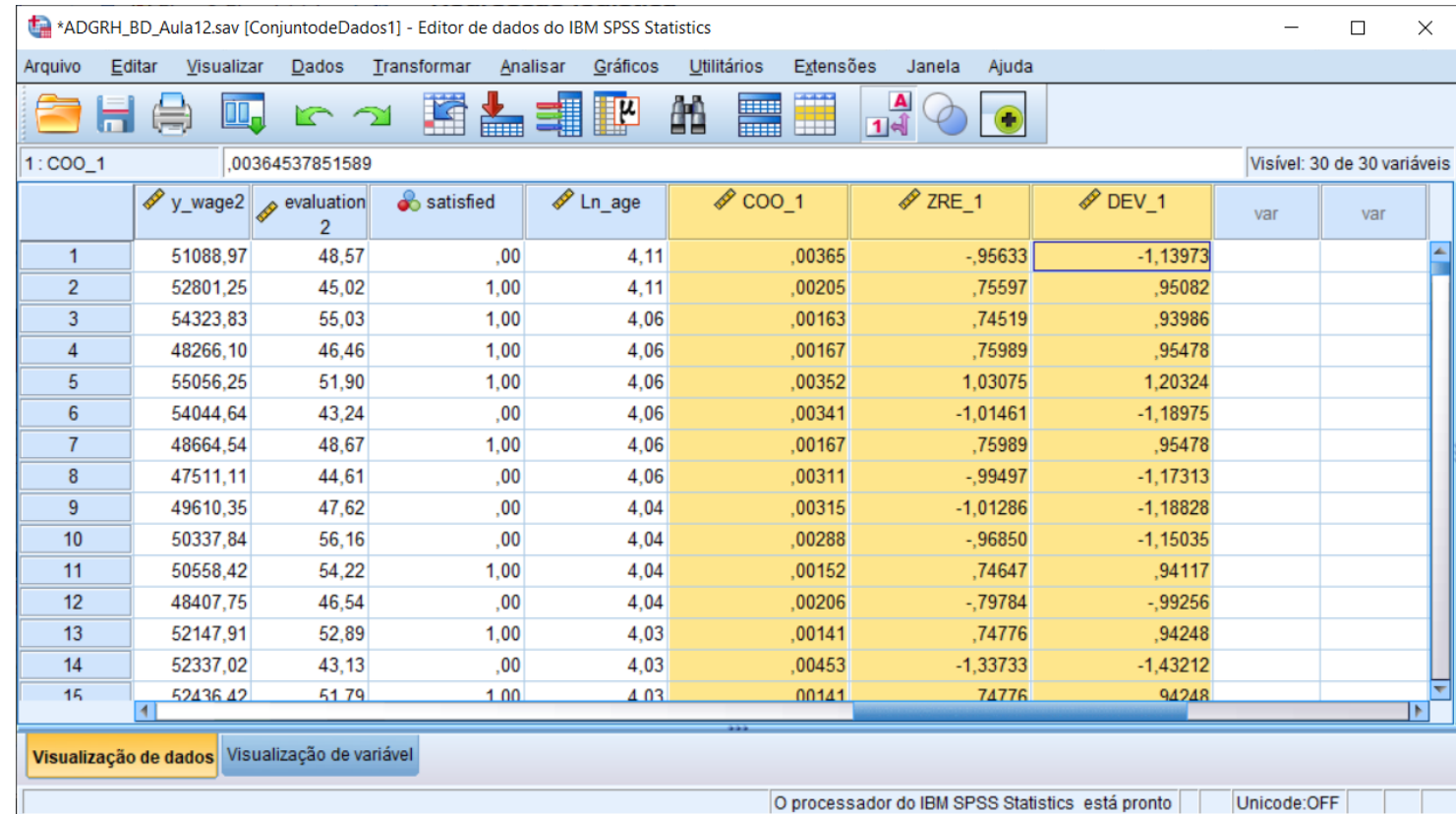
Regressão Logística

- Clicar no botão 'Salvar'
- Selecionar a opção 'de Cook'.
- Selecionar a opção 'Padronizado' e 'Desvio'.
- Clicar 'Continuar' / 'OK'



Regressão Logística

- Quando instruímos o SPSS para produzir os diagnósticos, é criado um conjunto variáveis
 - Uma variável que mede a distancia de Cook associada a cada observação (COO_1).
 - Uma variável com os 'Resíduos Padronizados' da variável dependente (ZRE_1) para cada observação.
 - Uma variável com os 'Resíduos de Desvio' (DEV_1) para cada observação.



*ADGRH_BD_Aula12.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: COO_1 ,00364537851589 Visível: 30 de 30 variáveis

	y_wage2	evaluation 2	satisfied	Ln_age	COO_1	ZRE_1	DEV_1	var	var
1	51088,97	48,57	,00	4,11	,00365	-,95633	-1,13973		
2	52801,25	45,02	1,00	4,11	,00205	,75597	,95082		
3	54323,83	55,03	1,00	4,06	,00163	,74519	,93986		
4	48266,10	46,46	1,00	4,06	,00167	,75989	,95478		
5	55056,25	51,90	1,00	4,06	,00352	1,03075	1,20324		
6	54044,64	43,24	,00	4,06	,00341	-1,01461	-1,18975		
7	48664,54	48,67	1,00	4,06	,00167	,75989	,95478		
8	47511,11	44,61	,00	4,06	,00311	-,99497	-1,17313		
9	49610,35	47,62	,00	4,04	,00315	-1,01286	-1,18828		
10	50337,84	56,16	,00	4,04	,00288	-,96850	-1,15035		
11	50558,42	54,22	1,00	4,04	,00152	,74647	,94117		
12	48407,75	46,54	,00	4,04	,00206	-,79784	-,99256		
13	52147,91	52,89	1,00	4,03	,00141	,74776	,94248		
14	52337,02	43,13	,00	4,03	,00453	-1,33733	-1,43212		
15	52436,42	51,79	1,00	4,03	,00141	,74776	,94248		

Visualização de dados Visualização de variável

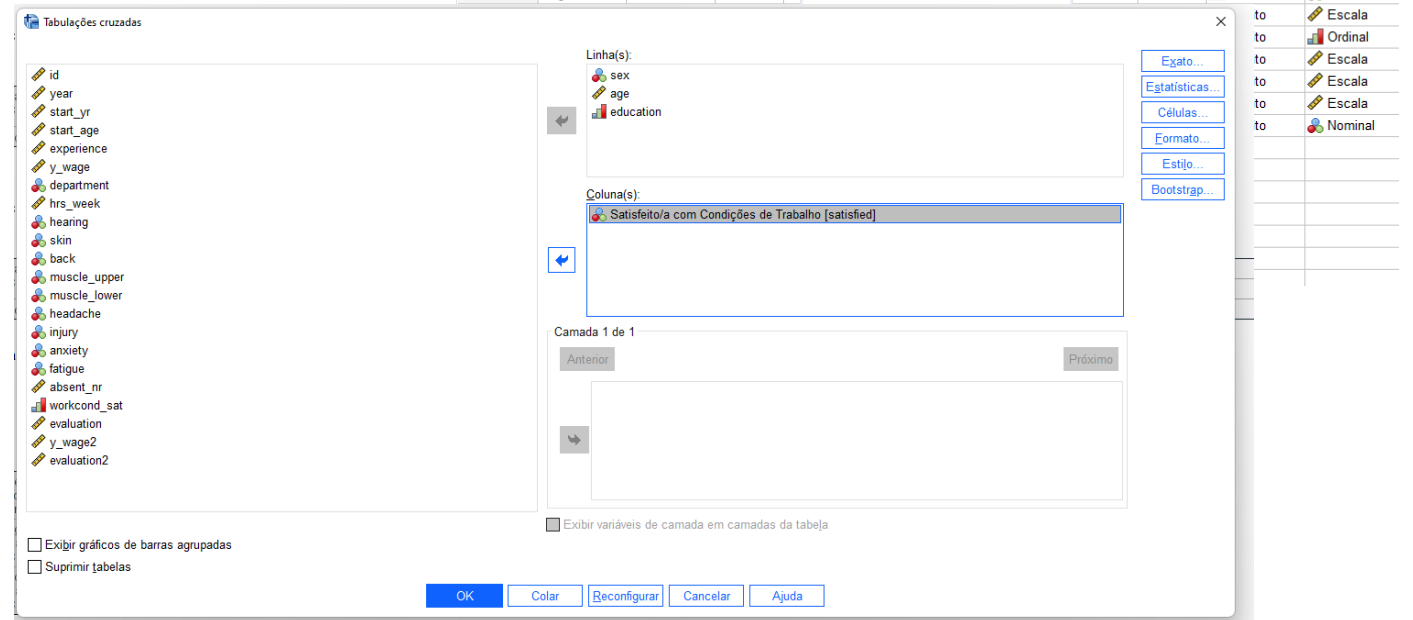
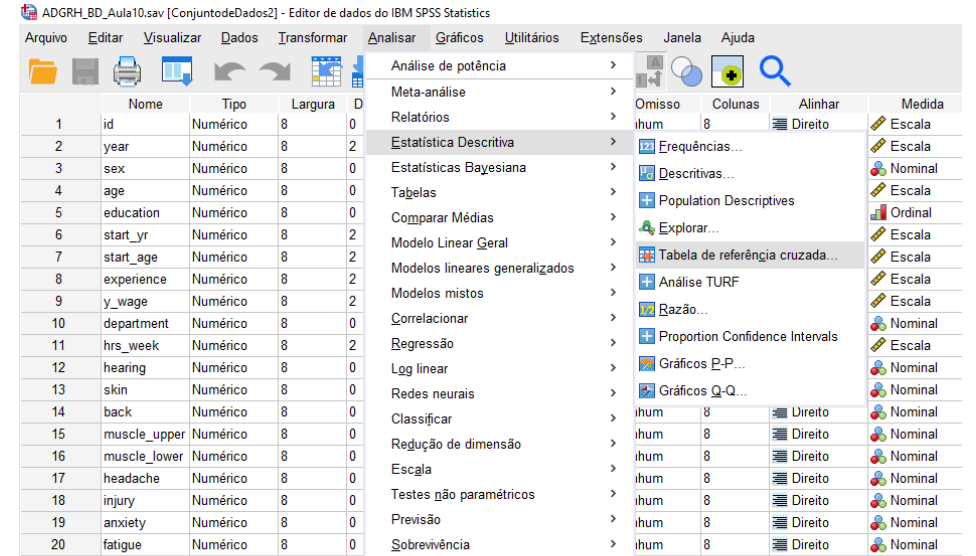
O processador do IBM SPSS Statistics está pronto Unicode:OFF

Validação do Modelo de Regressão Logística

2. Ausência de células vazias

Pressuposto III: *Ausência de células vazias*

- Tabelas de referência cruzada
- Linhas: VI's
- Colunas: VD
*ou vice-versa
- OK



Pressuposto II:

Ausência de células vazias

Tabulação cruzada education * satisfied Satisfeito/a com Condições de Trabalho

Contagem

		satisfied Satisfeito/a com Condições de Trabalho		Total
		,00	1,00	
education	1 ISCED 1	616	363	979
	2 ISCED 2	612	411	1023
	3 ISCED 3	538	469	1007
	4 ISCED 4	509	485	994
	5 ISCED 5	373	624	997
Total		2648	2352	5000

Tabulação cruzada sex * satisfied Satisfeito/a com Condições de Trabalho

Contagem

		satisfied Satisfeito/a com Condições de Trabalho		Total
		,00	1,00	
sex	1 Female	1317	1197	2514
	2 Male	1331	1155	2486
Total		2648	2352	5000

Tabulação cruzada age * satisfied Satisfeito/a com Condições de Trabalho

Contagem

age		satisfied Satisfeito/a com Condições de Trabalho		Total
		,00	1,00	
21	1	0	1	
22	0	1	1	
23	1	0	1	
24	1	0	1	
25	3	1	4	
26	4	3	7	
27	9	3	12	
28	5	6	11	
29	18	12	30	
30	20	15	35	
31	30	21	51	
32	52	44	96	
33	63	65	128	
34	71	69	140	
35	116	89	205	
36	126	93	219	
37	140	113	253	
38	174	162	336	
39	170	163	333	
40	208	166	374	
41	172	185	357	
42	207	165	372	
43	178	156	334	

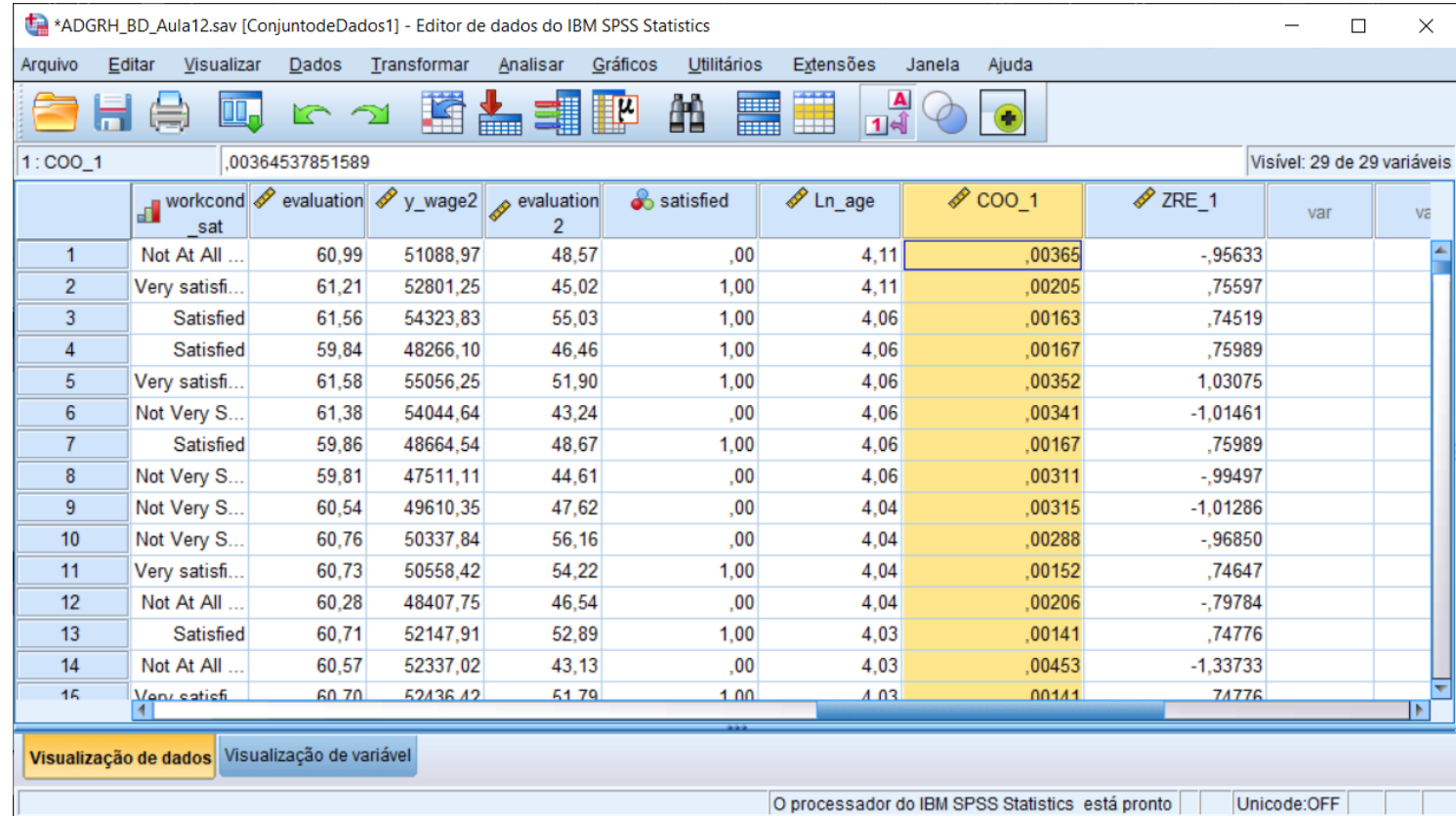
Validação do Modelo de Regressão Logística

3. Ausência de Observações Influentes

- **Objectivo:**
 - **Testar o pressuposto que os resultados do modelo que estima o impacto do sexo, da idade e do grau de educação sobre a satisfação dos trabalhadores não são enviesados pela presença de observações influentes.**

Pressuposto III: Observações Influentes

- Para testarmos a presença de observações influentes vamos usar a variável com os 'Distância de Cook' (COO_1) que acabamos de criar.



*ADGRH_BD_Aula12.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1: COO_1 ,00364537851589 Visível: 29 de 29 variáveis

	workcond_sat	evaluation	y_wage2	evaluation_2	satisfied	Ln_age	COO_1	ZRE_1	var	va
1	Not At All ...	60,99	51088,97	48,57	,00	4,11	,00365	-.95633		
2	Very satisfi...	61,21	52801,25	45,02	1,00	4,11	,00205	,75597		
3	Satisfied	61,56	54323,83	55,03	1,00	4,06	,00163	,74519		
4	Satisfied	59,84	48266,10	46,46	1,00	4,06	,00167	,75989		
5	Very satisfi...	61,58	55056,25	51,90	1,00	4,06	,00352	1,03075		
6	Not Very S...	61,38	54044,64	43,24	,00	4,06	,00341	-1,01461		
7	Satisfied	59,86	48664,54	48,67	1,00	4,06	,00167	,75989		
8	Not Very S...	59,81	47511,11	44,61	,00	4,06	,00311	-.99497		
9	Not Very S...	60,54	49610,35	47,62	,00	4,04	,00315	-1,01286		
10	Not Very S...	60,76	50337,84	56,16	,00	4,04	,00288	-.96850		
11	Very satisfi...	60,73	50558,42	54,22	1,00	4,04	,00152	,74647		
12	Not At All ...	60,28	48407,75	46,54	,00	4,04	,00206	-.79784		
13	Satisfied	60,71	52147,91	52,89	1,00	4,03	,00141	,74776		
14	Not At All ...	60,57	52337,02	43,13	,00	4,03	,00453	-1,33733		
15	Very satisfi...	60,70	52136,42	51,79	1,00	4,03	,00141	,74776		

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:OFF

Pressuposto II: Observações Influentes

- Selecionar 'Gráficos' / 'Construtor de Gráfico'

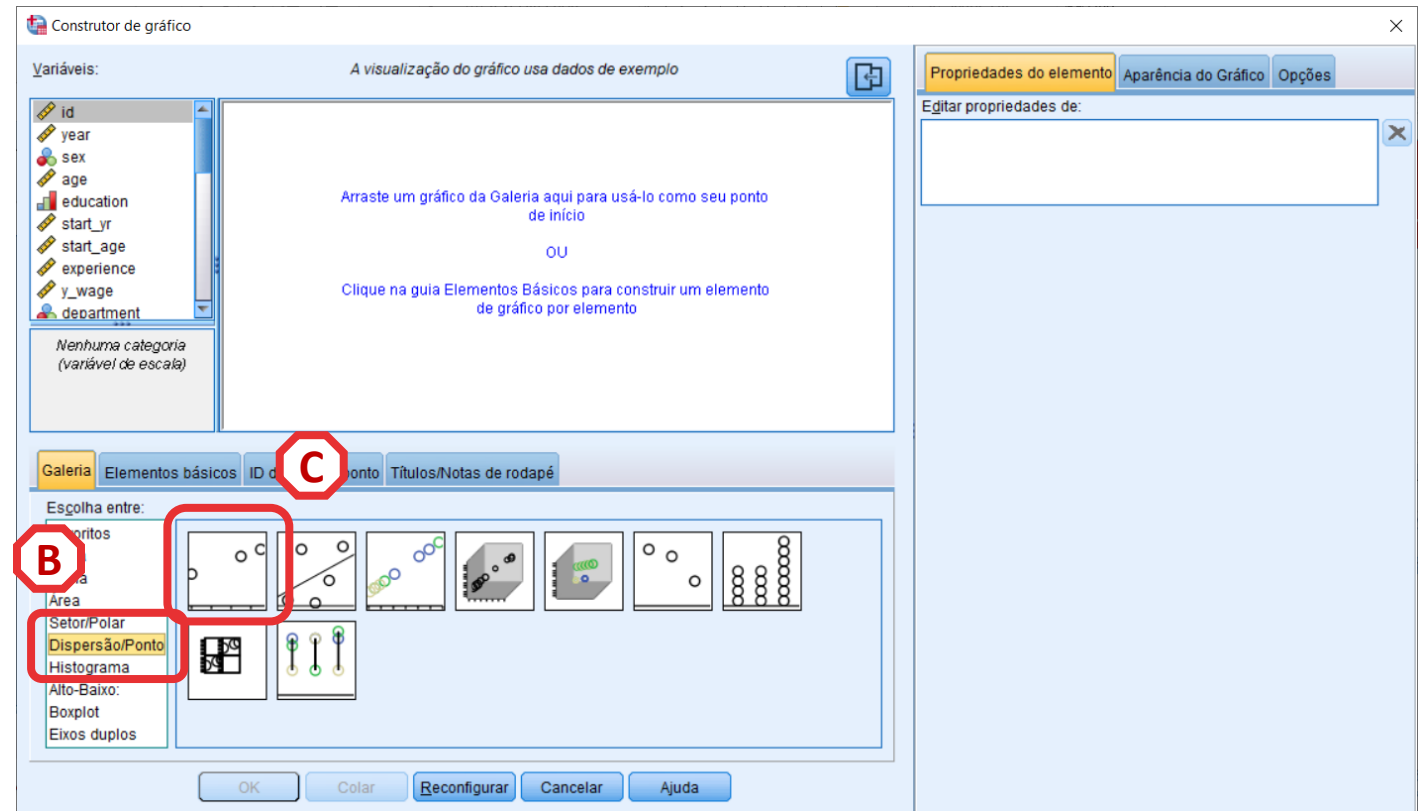
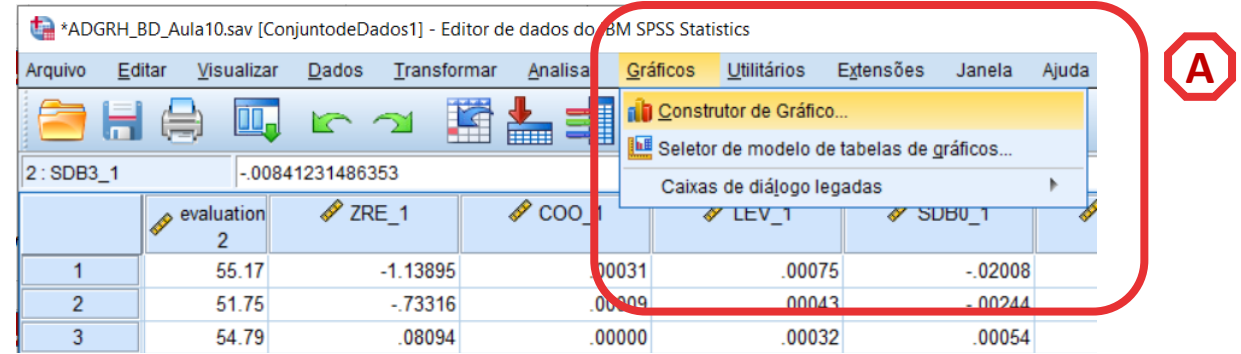
A

- Selecionar 'DispersãoPontos'

B

- Selecionar 'Dispersão (Simples)'

C



Pressuposto II: Observações Influentes

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'DispersãoPontos'
- Selecionar 'Dispersão (Simples)'
- Selecionar Variável 'id'
- Colocar no eixo 'x'

A

B

C

D

E

Pressuposto II: Observações Influentes

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'DispersãoPontos'
- Selecionar 'Dispersão (Simples)'
- Selecionar Variável 'id'
- Colocar no eixo 'x'
- Selecionar Variável 'Análogo às estatísticas de influência de Cook'
- Colocar no eixo 'y'

A

B

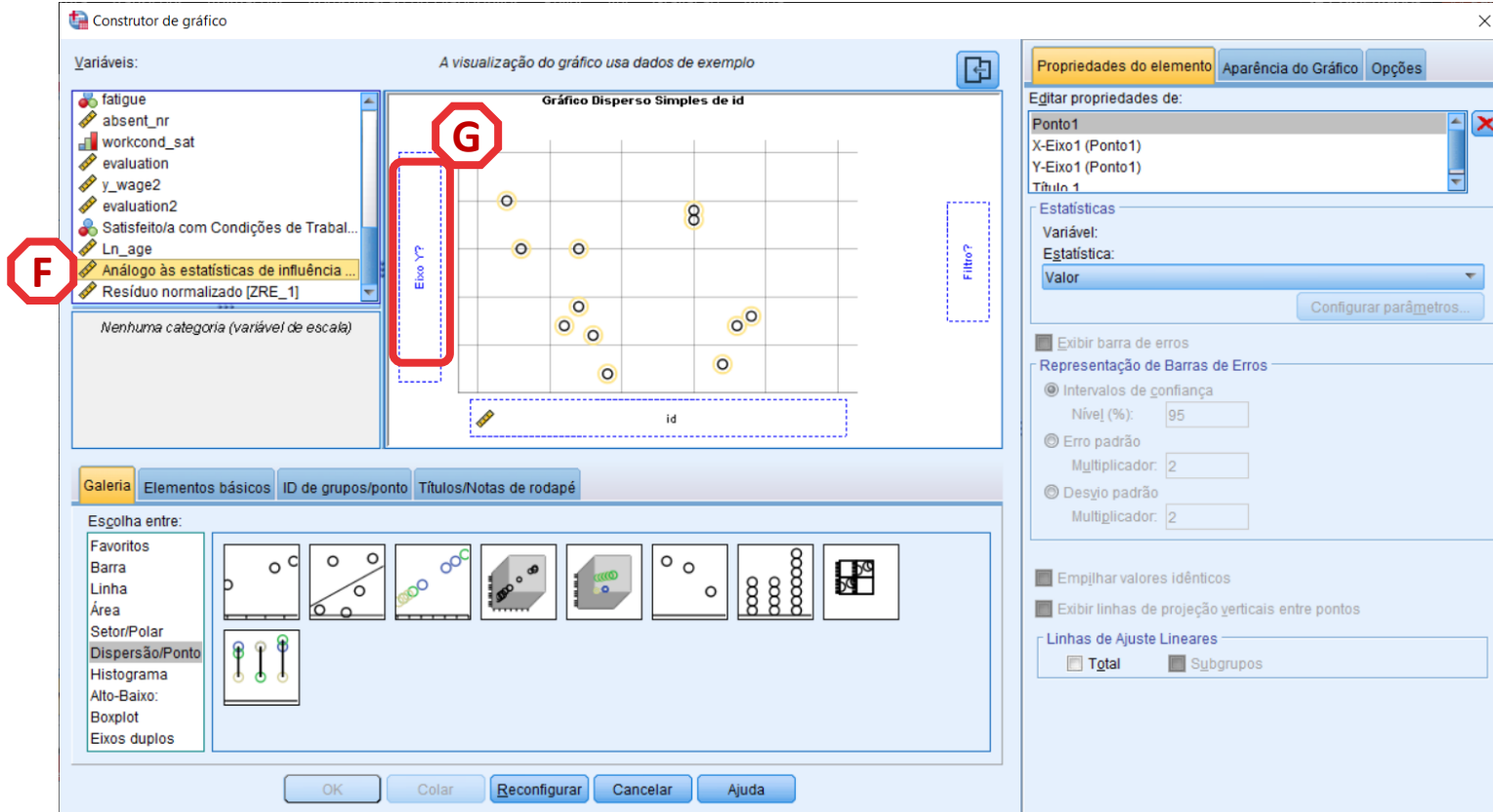
C

D

E

F

G



Pressuposto II: Observações Influentes

- Selecionar 'ID de grupos/ponto'
- Selecionar 'Rótulo da ID do Ponto'
- Selecionar Variável 'id'
- Colocar na caixa 'Variável do rótulo do ponto'
- Selecionar '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)

Gráfico Disperso Simples de Cook's Distance por id

Cook's Distance

id

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

Os itens marcados adicionam à tela áreas de lançamento para as quais as variáveis podem ser atribuídas.

- Variável de agrupamento em X
- Variável de agrupamento em Z
- Variável de agrupamento/empilhamento
- Variável do painel de linhas
- Variável do painel das colunas
- Rótulo da ID do ponto

OK Colar Reconfigurar Cancelar Ajuda

Propriedades do elemento Aparência do Gráfico Opções

Editar propriedades de: Ponto1

X-Eixo1 (Ponto1)

Y-Eixo1 (Ponto1)

Título 1

Estadísticas

Variável: Cook's Distance

Estadística: Valor

Configurar parâmetros...

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

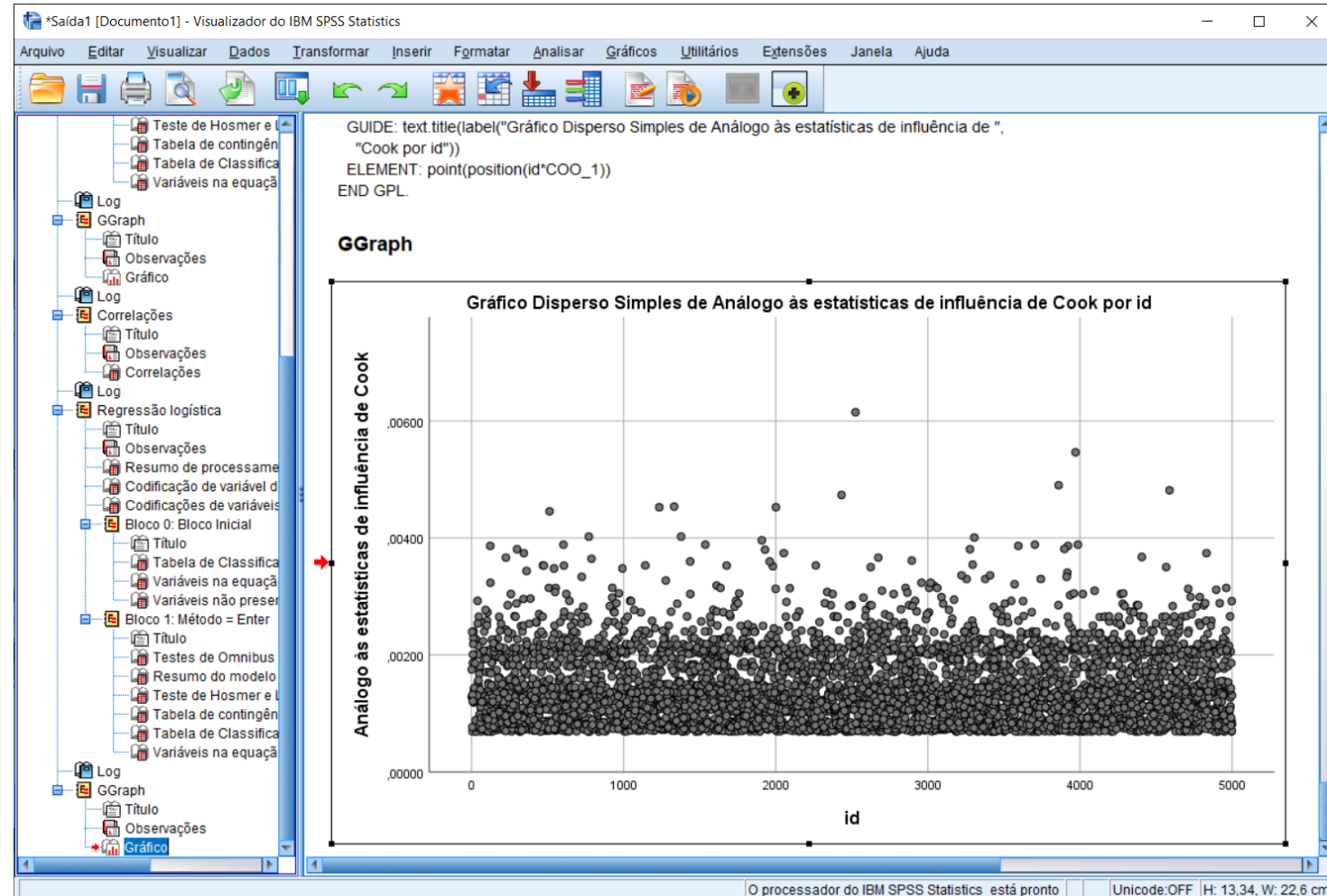
Pressuposto II: Observações Influentes

- Reparem que o gráfico permite identificar o ID dos outliers

- Interpretação

$CD > 4/n \rightarrow$ Caso Influente

- Neste caso, o valor de corte é $0.008 (4 / 5000)$
- Neste caso não há observações acima do valor de corte.
- Cumpre-se o pressuposto da ausência de observações influentes



Validação do Modelo de Regressão Logística

4. Ausência de 'Separação Completa'

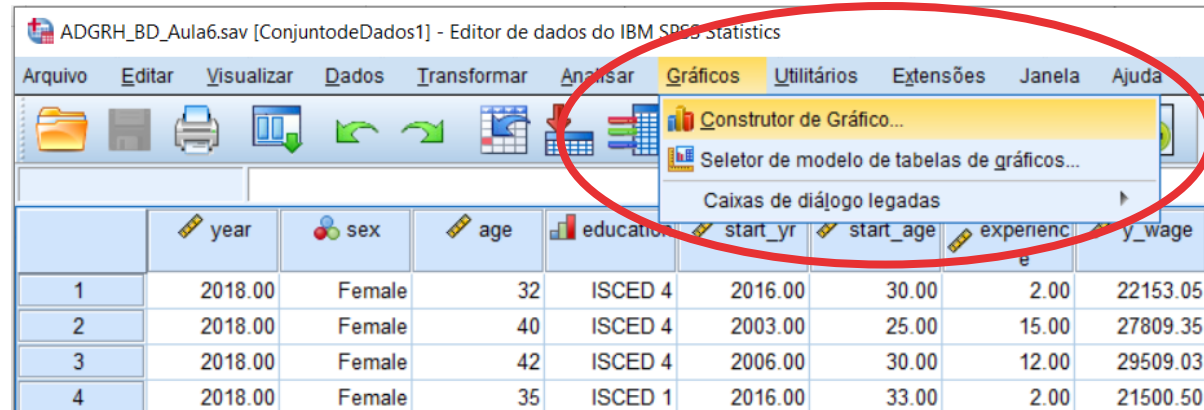
- **Objectivo:**

- **Testar se conseguimos prever todos valores da variável dependente partir dos valores de uma das variáveis independentes - que neste caso será a educação.**

Para esse efeito vamos representar graficamente a relação entre estas duas variáveis através de um Gráfico de Dispersão.

Pressuposto IV: Separação Perfeita

- Selecionar 'Gráficos' / 'Construtor de Gráfico'



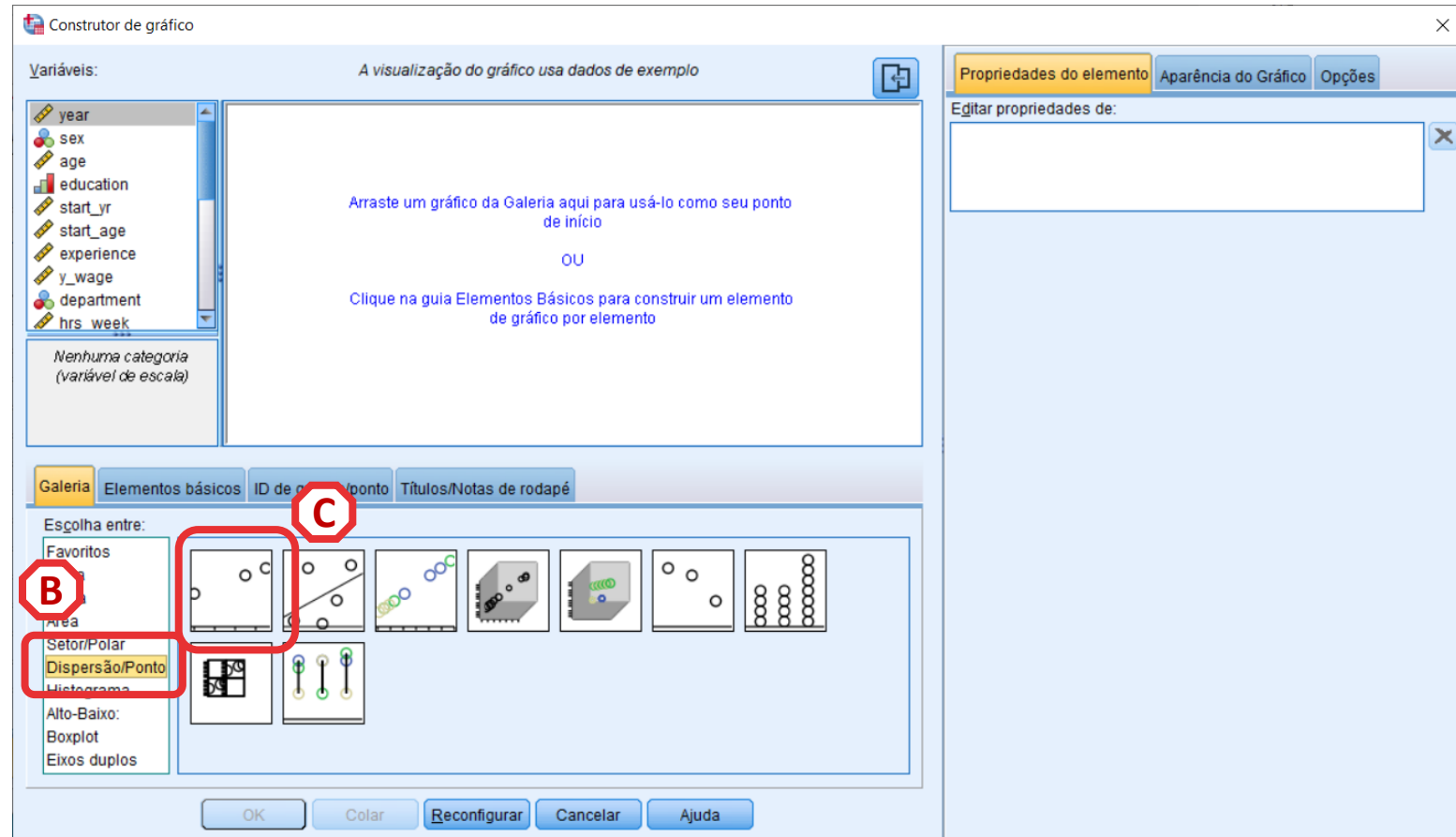
Pressuposto IV: Separação Perfeita

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Dispersão/Ponto'
- Duplo-Clique na opção 'Dispersão Simples'

A

B

C



Pressuposto IV: Separação Perfeita

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Dispersão/Ponto'
- Duplo-Clique na opção 'Dispersão Simples'
- Selecionar a variável 'age' ...
... e colocar na caixa do 'Eixo X'

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

Pressuposto IV: Separação Perfeita

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecione 'Dispersão/Ponto'
- Duplo-Clique na opção 'Dispersão Simples'
- Selecionar a variável 'age' ...
- ... e colocar na caixa do 'Eixo X'

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

- Selecionar 'OK'...

A

B

C

D

E

F

F

Construtor de gráfico

A visualização do gráfico usa dados de exemplo

Variáveis:

- injury
- anxiety
- fatigue
- absent_nr
- workcond_sat
- evaluation
- y_wage2
- evaluation2
- Satisfeito/a com ...
- Ln_age

Categoria 1

Categoria 2

Gráfico Disperso Simples de Satisfeito/a com Condições de Trabalho por age

age

Satisfeito/a com Condições de Trabalho

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:

Ponto1

X-Eixo1 (Ponto1)

Y-Eixo1 (Ponto1)

Título 1

Estatísticas

Variável: Satisfeito/a com Condições de Trabalho

Estatística:

Valor

Configurar parâmetros...

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

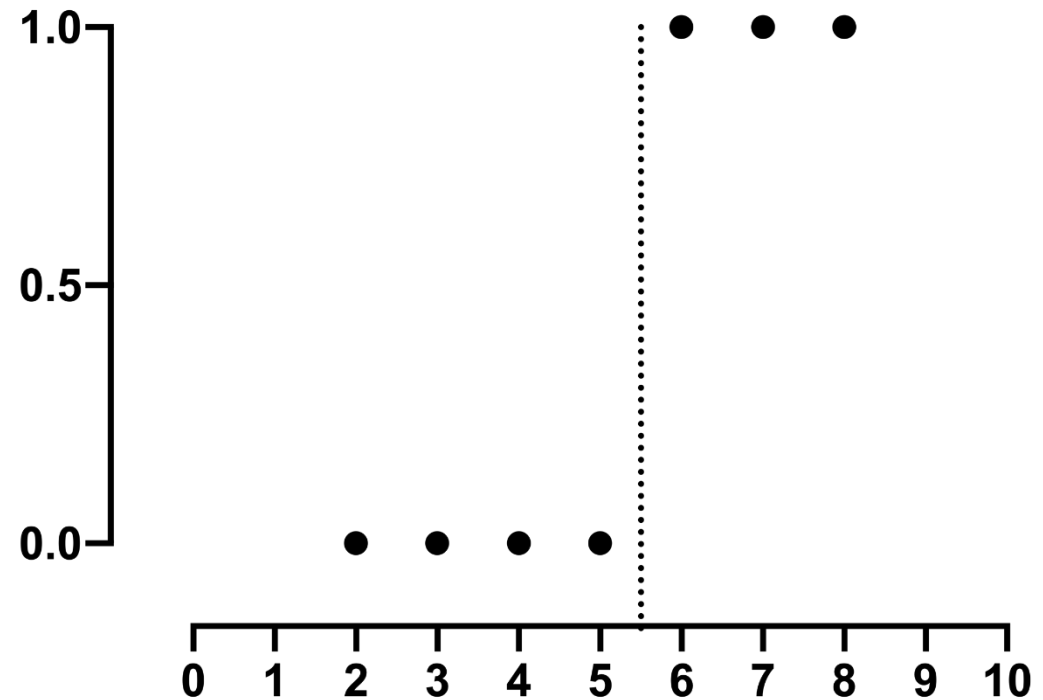
Exibir linhas de projeção verticais entre pontos

Linhas de Ajuste Lineares

- Total
- Subgrupos

Pressuposto IV: Separação Perfeita

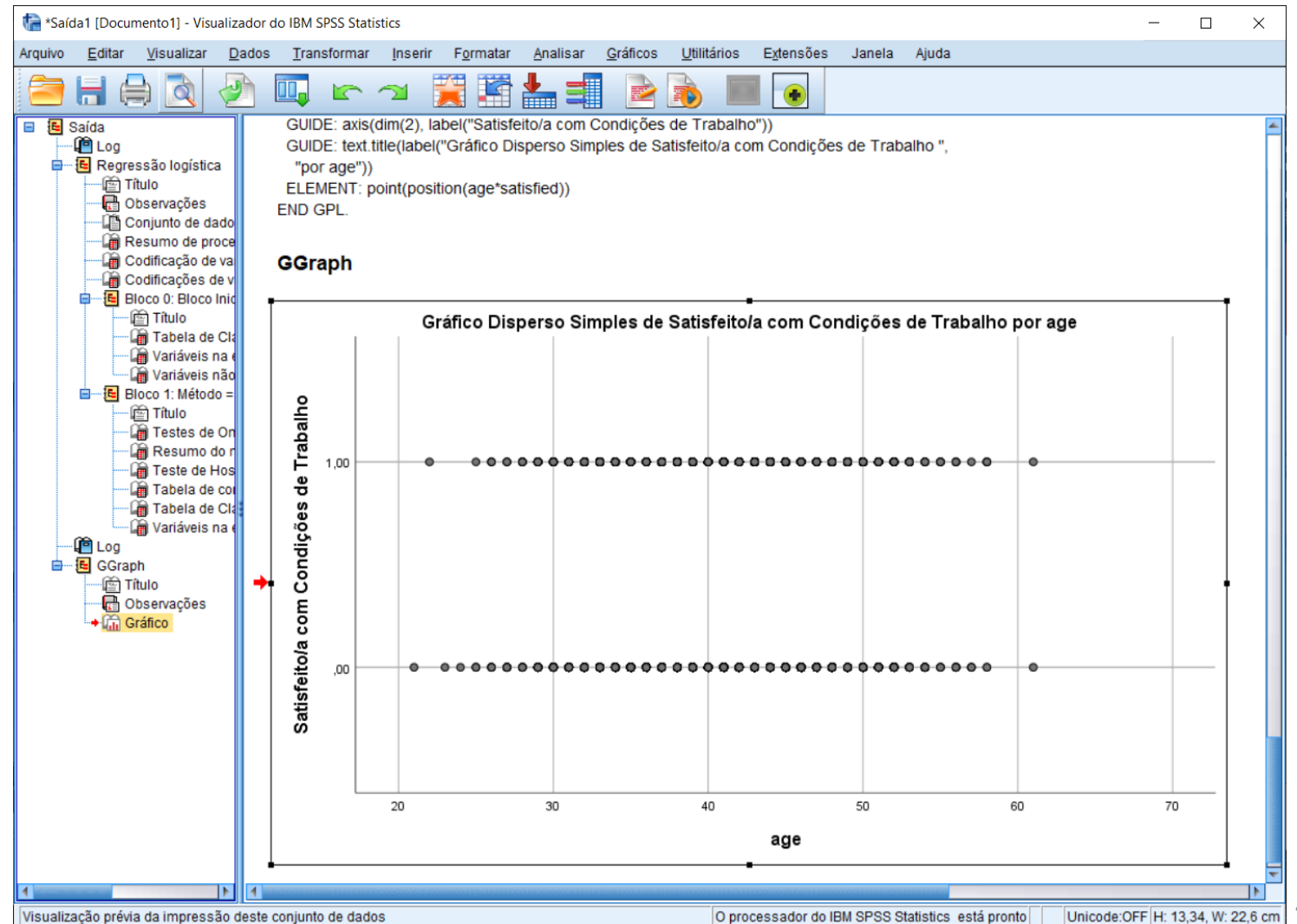
- O resultado é publicado no 'Visualizador de Resultados'
- Regista-se uma situação de separação perfeita quando podemos identificar um ponto na variável do Eixo X (identificado pela linha vertical) que separa de forma perfeita a distribuição das observações na variável dependente (Eixo Y).



Fonte: https://www.graphpad.com/guides/prism/latest/curve-fitting/reg_mult_logistic_dependence_and_separation.htm

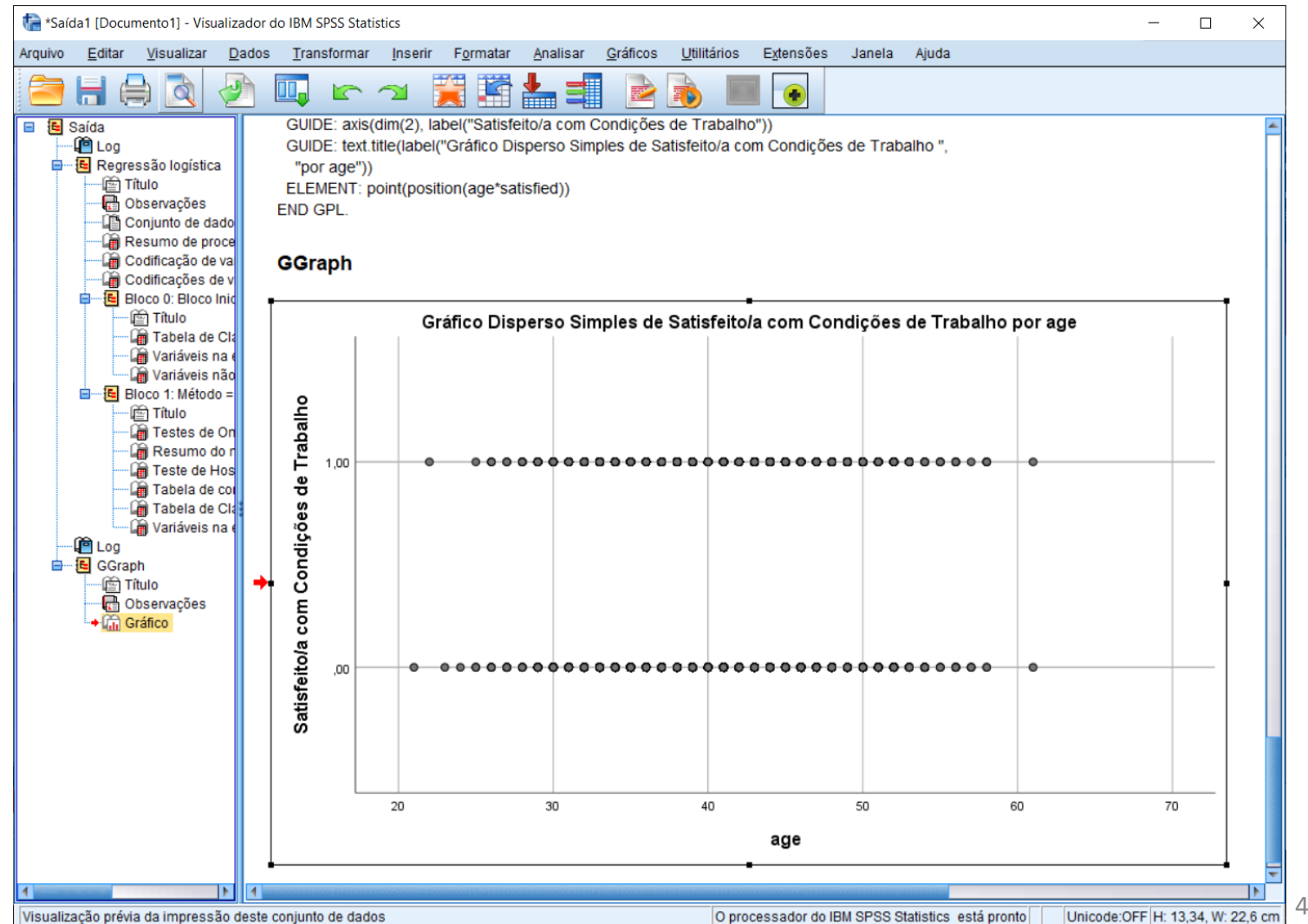
Pressuposto IV: Separação Perfeita

- O resultado é publicado no 'Visualizador de Resultados'
- Neste caso, confirma-se o pressuposto da inexistência de separação perfeita.



Pressuposto IV: Separação Perfeita

- E com as variáveis dicotômicas/categóricas?
- Exploramos as associações entre a VD e as VI



Validação do Modelo de Regressão Logística

5. Ausência de Multicolinearidade Perfeita

- **Objectivo:**

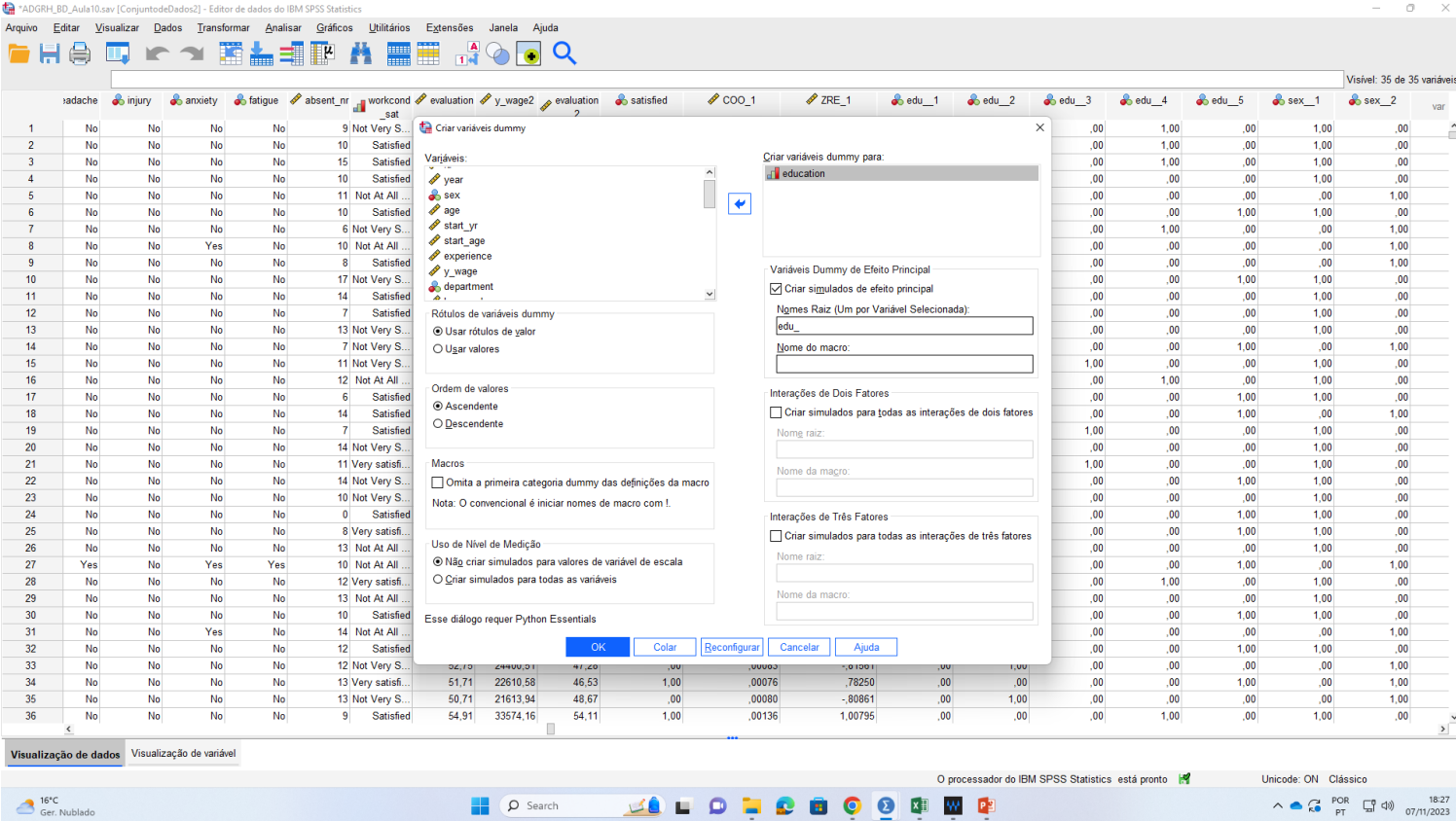
- Testar se existe uma situação de multicolinearidade perfeita entre as variáveis independentes do nosso modelo.

Podemos usar os comando da regressão linear para gerar as estatísticas adequadas (VIF), outra opção é estudar a associação entre as variáveis independentes: género, idade, educação.

Mas atenção, no nosso modelo, tratamos as variáveis género e educação como categóricas – e portanto temos de as tratar como tal.

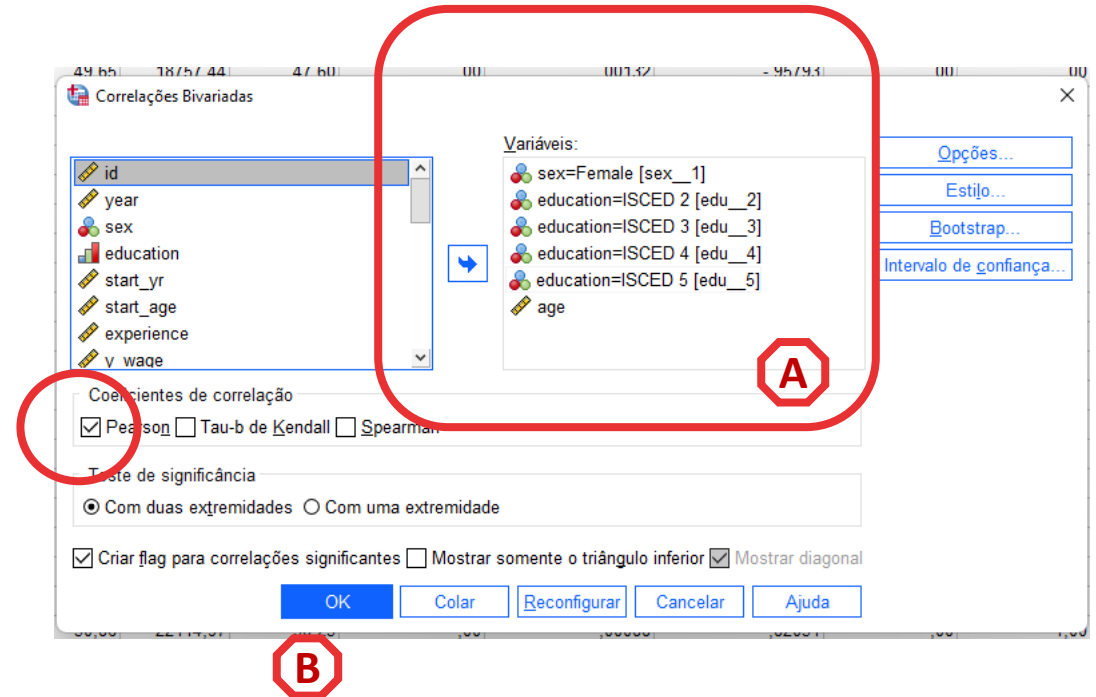
Pressuposto V: Multicolinearidade Perfeita

- Criar variáveis dummy para o género “sex_”
- Criar variáveis dummy para a educação “edu_”



Pressuposto V: Multicolinearidade Perfeita

- Selecionar 'Analisar' / 'Correlacionar' / 'Bivariada'
- Selecionar as variáveis independentes modelo... **A**
- e colocá-las na caixa 'Variáveis'
- Clicar 'OK' **B**



- **NOTA: as dummies são tratadas como variáveis contínuas. Por esse motivo o teste de associação R de Pearson é considerado adequado.**

Pressuposto V: Multicolinearidade Perfeita

- O resultado é publicado no 'Visualizador de Resultados'
- A existência de Coeficientes de Pearson com valores
- ≥ 0.8 ou
- ≤ -0.8)
- sugere a possibilidade de multiconlineariedade perfeita.

		Correlações					
		sex__1 sex=Female	edu__2 education=IS CED 2	edu__3 education=IS CED 3	edu__4 education=IS CED 4	edu__5 education=IS CED 5	age
sex__1 sex=Female	Correlação de Pearson	1	,016	,006	-,025	,020	-,048**
	Sig. (2 extremidades)		,244	,689	,079	,163	<,001
	N	5000	5000	5000	5000	5000	5000
edu__2 education=ISCED 2	Correlação de Pearson	,016	1	-,255**	-,253**	-,253**	-,037**
	Sig. (2 extremidades)	,244		<,001	<,001	<,001	,008
	N	5000	5000	5000	5000	5000	5000
edu__3 education=ISCED 3	Correlação de Pearson	,006	-,255**	1	-,250**	-,251**	,009
	Sig. (2 extremidades)	,689	<,001		<,001	<,001	,547
	N	5000	5000	5000	5000	5000	5000
edu__4 education=ISCED 4	Correlação de Pearson	-,025	-,253**	-,250**	1	-,249**	,024
	Sig. (2 extremidades)	,079	<,001	<,001		<,001	,087
	N	5000	5000	5000	5000	5000	5000
edu__5 education=ISCED 5	Correlação de Pearson	,020	-,253**	-,251**	-,249**	1	,078**
	Sig. (2 extremidades)	,163	<,001	<,001	<,001		<,001
	N	5000	5000	5000	5000	5000	5000
age	Correlação de Pearson	-,048**	-,037**	,009	,024	,078**	1
	Sig. (2 extremidades)	<,001	,008	,547	,087	<,001	
	N	5000	5000	5000	5000	5000	5000

** A correlação é significativa no nível 0,01 (2 extremidades).

Pressuposto V: Multicolinearidade Perfeita

- O resultado é publicado no ‘Visualizador de Resultados’
- A existência de Coeficientes de Pearson com valores ≥ 0.8 ou ≤ -0.8)
- sugere a possibilidade de multicolinearidade perfeita.
- Mas este não parece ser o caso aqui...
- pelo que se assume que se cumpre o pressuposto da ausência de multicolinearidade.

Correlações

		sex__1 sex=Female	edu__2 education=IS CED 2	edu__3 education=IS CED 3	edu__4 education=IS CED 4	edu__5 education=IS CED 5	age
sex__1 sex=Female	Correlação de Pearson	1	,016	,006	-,025	,020	-,048**
	Sig. (2 extremidades)		,244	,689	,079	,163	<,001
	N	5000	5000	5000	5000	5000	5000
edu__2 education=ISCED 2	Correlação de Pearson	,016	1	-,255**	-,253**	-,253**	-,037**
	Sig. (2 extremidades)	,244		<,001	<,001	<,001	,008
	N	5000	5000	5000	5000	5000	5000
edu__3 education=ISCED 3	Correlação de Pearson	,006	-,255**	1	-,250**	-,251**	,009
	Sig. (2 extremidades)	,689	<,001		<,001	<,001	,547
	N	5000	5000	5000	5000	5000	5000
edu__4 education=ISCED 4	Correlação de Pearson	-,025	-,253**	-,250**	1	-,249**	,024
	Sig. (2 extremidades)	,079	<,001	<,001		<,001	,087
	N	5000	5000	5000	5000	5000	5000
edu__5 education=ISCED 5	Correlação de Pearson	,020	-,253**	-,251**	-,249**	1	,078**
	Sig. (2 extremidades)	,163	<,001	<,001	<,001		<,001
	N	5000	5000	5000	5000	5000	5000
age	Correlação de Pearson	-,048**	-,037**	,009	,024	,078**	1
	Sig. (2 extremidades)	<,001	,008	,547	,087	<,001	
	N	5000	5000	5000	5000	5000	5000

** A correlação é significativa no nível 0,01 (2 extremidades).

Pressuposto V: Multicolinearidade Perfeita

- Outra alternativa é usar os comandos da regressão linear para obter a estatística VIF para o conjunto das variáveis
- $VIF < 3$, podemos aceitar o pressuposto da ausência de multicolinearidade elevada

Coefficientes^a

Modelo		Coeficientes não padronizados		Coeficientes padronizados	t	Sig.	Estatísticas de colinearidade	
		B	Erro Erro	Beta			Tolerância	VIF
1	(Constante)	,586	,058		10,095	<,001		
	edu__1 education=ISCED 1	-,253	,022	-,201	-11,399	<,001	,621	1,610
	edu__2 education=ISCED 2	-,223	,022	-,180	-10,169	<,001	,617	1,620
	edu__3 education=ISCED 3	-,159	,022	-,128	-7,250	<,001	,622	1,609
	edu__4 education=ISCED 4	-,137	,022	-,110	-6,214	<,001	,624	1,603
	age	,001	,001	,009	,636	,525	,987	1,013
	sex__1 sex=Female	,009	,014	,009	,677	,498	,996	1,004

a. Variável Dependente: satisfied Satisfeito/a com Condições de Trabalho

Validação do Modelo de Regressão Logística

5. Avaliação do Pressuposto VI: Linearidade

- **Objectivo:**
 - **Testar se há uma relação linear entre a variável que mede que se os trabalhadores estão satisfeitos ('satisfied') e o logaritmo das variáveis (independentes) de razão no nosso modelo de regressão - que neste caso é apenas a idade (age).**

- **Objectivo:**

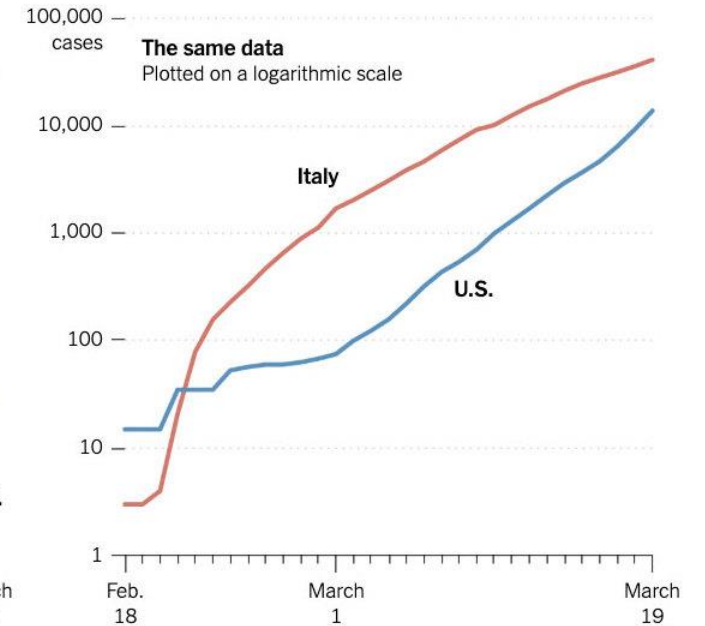
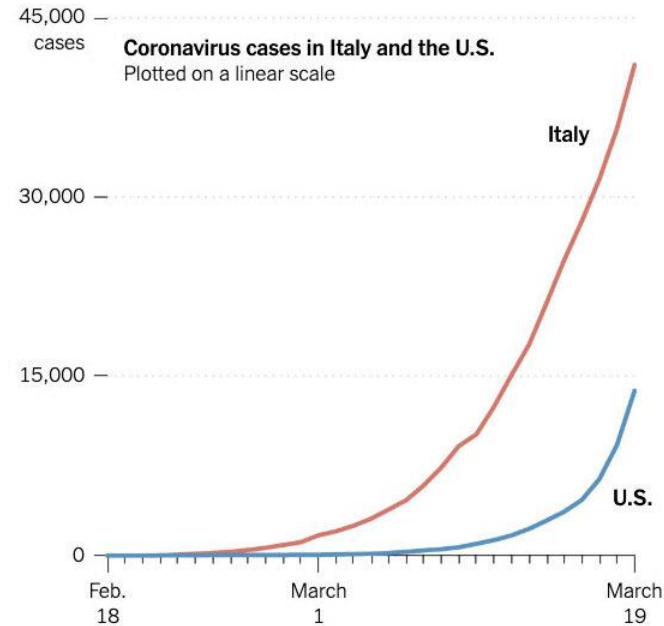
- **Testar se há uma relação linear entre a variável que mede que se os trabalhadores estão satisfeitos ('satisfied') e o logaritmo das variáveis (independentes) de razão no nosso modelo de regressão - que neste caso é apenas a idade (age).**

Para podermos fazer isto, temos de:

- a) Criar uma nova variável ('Ln_age') que transformamos a distribuição da variável de razão em causa ('age') numa escala logarítmica.**
- b) Adicionar ao nosso modelo um 'termo de interação' entre a variável idade ('age') e o seu logaritmo natural (Ln_age).**

Transformação Logarítmica

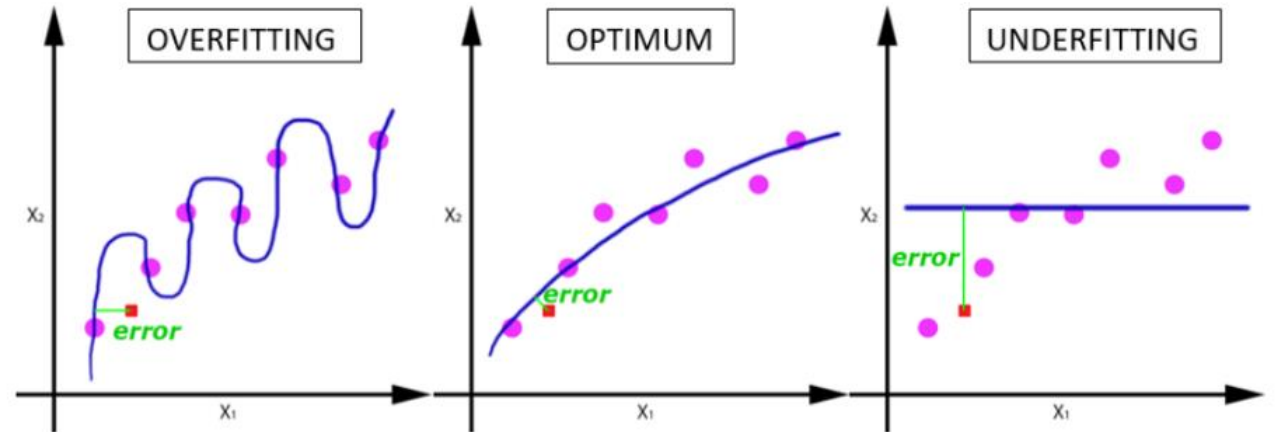
- A transformação logarítmica é um meio conveniente de transformar uma variável altamente enviesada numa distribuição mais normalizada.



Fonte: <https://static01.nyt.com/images/2020/03/20/science/virus-log-chart-1584728689795/virus-log-chart-1584728689795-videoSixteenByNineJumbo1600.jpg>

Transformação Logarítmica

- A transformação logarítmica é um meio conveniente de transformar uma variável altamente enviesada numa distribuição mais normalizada.
- Variáveis muito enviesadas podem gerar modelos sobreajustados (*overfitted*) que se ajustam muito bem ao conjunto de dados (i.e., um R^2 alto), mas se mostra ineficazes para prever novos resultados.
- Usar o logaritmo de uma ou mais variáveis ajuda a melhorar o ajuste do modelo.



Fonte: https://cdn-images-1.medium.com/max/1024/0*tBRkeyTgwJ5xpr2

Pressuposto VI: Linearidade

- Selecionar 'Transformar' / 'Calcular variável'



ADGRH_BD_Aula12.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janelas

Calcular variável...
+ Transformação de programabilidade...
x? Valores de contagem dentro de casos...
Valores de Mudança...
Recodificar nas mesmas variáveis...
Recodificar em variáveis diferentes...
Recodificação automática...
+ Criar variáveis dummy
Categorização visual...
Categorização ideal...

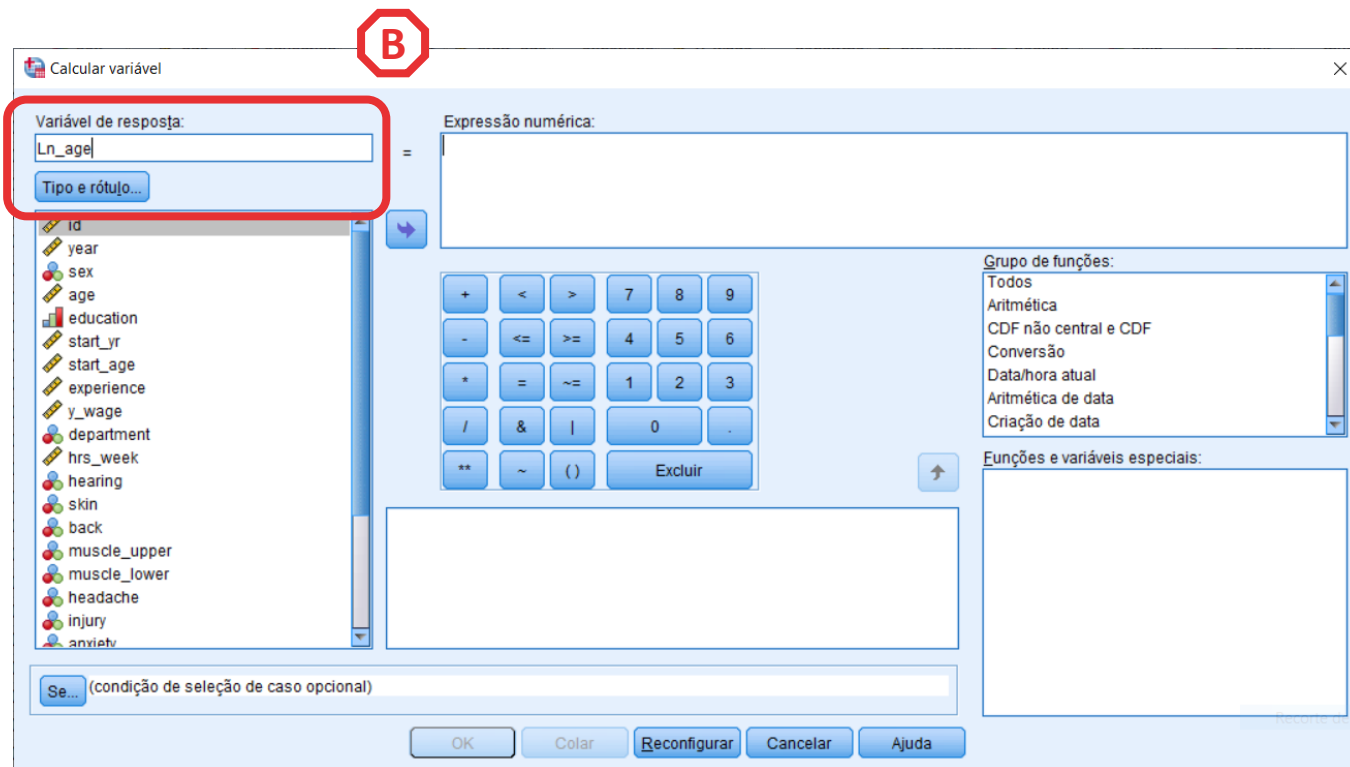
	id	year	start_age
1	1	2018,00	30,00
2	2	2018,00	25,00
3	3	2018,00	30,00
4	4	2018,00	33,00
5	5	2018,00	21,00
6	6	2018,00	32,00
7	7	2018,00	32,00

Pressuposto VI: Linearidade

- Selecionar 'Transformar' / 'Calcular variável'
- Definir o nome da variável que mede o logaritmo natural * da variável 'age'.

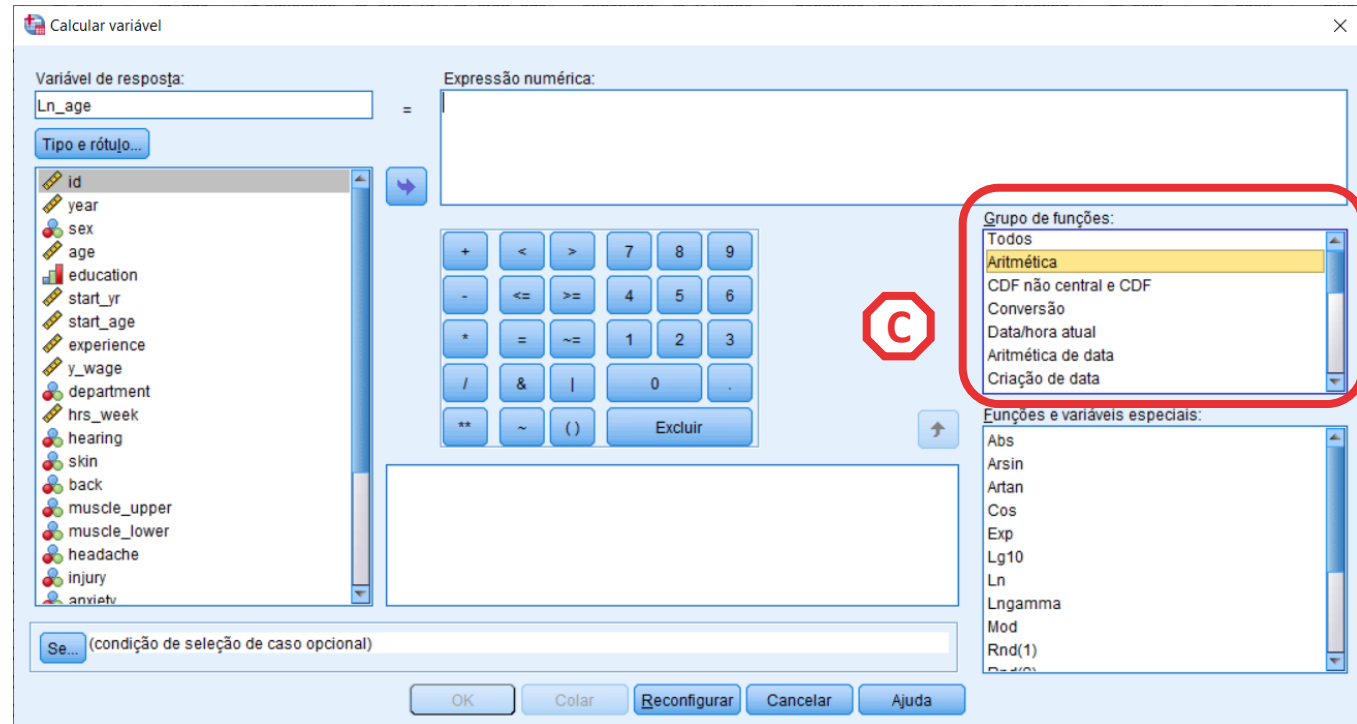
A

B



Pressuposto VI: Linearidade

- Selecionar 'Transformar' / 'Calcular variável'
- Definir o nome da variável que mede o logaritmo natural * da variável 'age'.
- Na caixa 'Grupo de funções', selecionar a opção 'Aritmética'.



Pressuposto VI: Linearidade

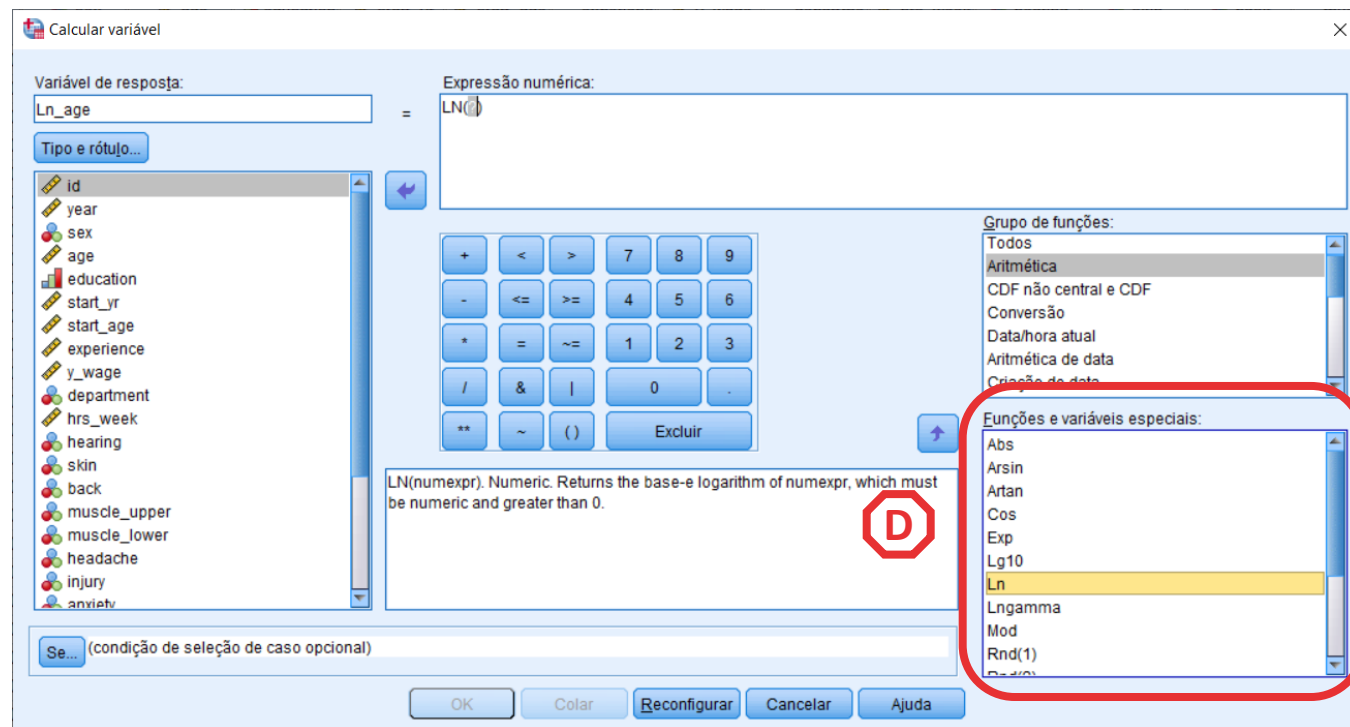
- Selecionar 'Transformar' / 'Calcular variável'
- Definir o nome da variável que mede o logaritmo natural * da variável 'age'.
- Na caixa 'Grupo de funções', selecionar a opção 'Aritmética'.
- Duplo-Clique na opção Ln (Logaritmo Natural).

A

B

C

D



Pressuposto VI: Linearidade

- Seleccionar 'Transformar' / 'Calcular variável'
- Definir o nome da variável que mede o logaritmo natural * da variável 'age'.
- Na caixa 'Grupo de funções', seleccionar a opção 'Aritmética'.
- Duplo-Clique na opção Ln (Logaritmo Natural).
- Seleccionar a variável que vai ser transformada ('age')...
- E colocá-la na dentro da função

A

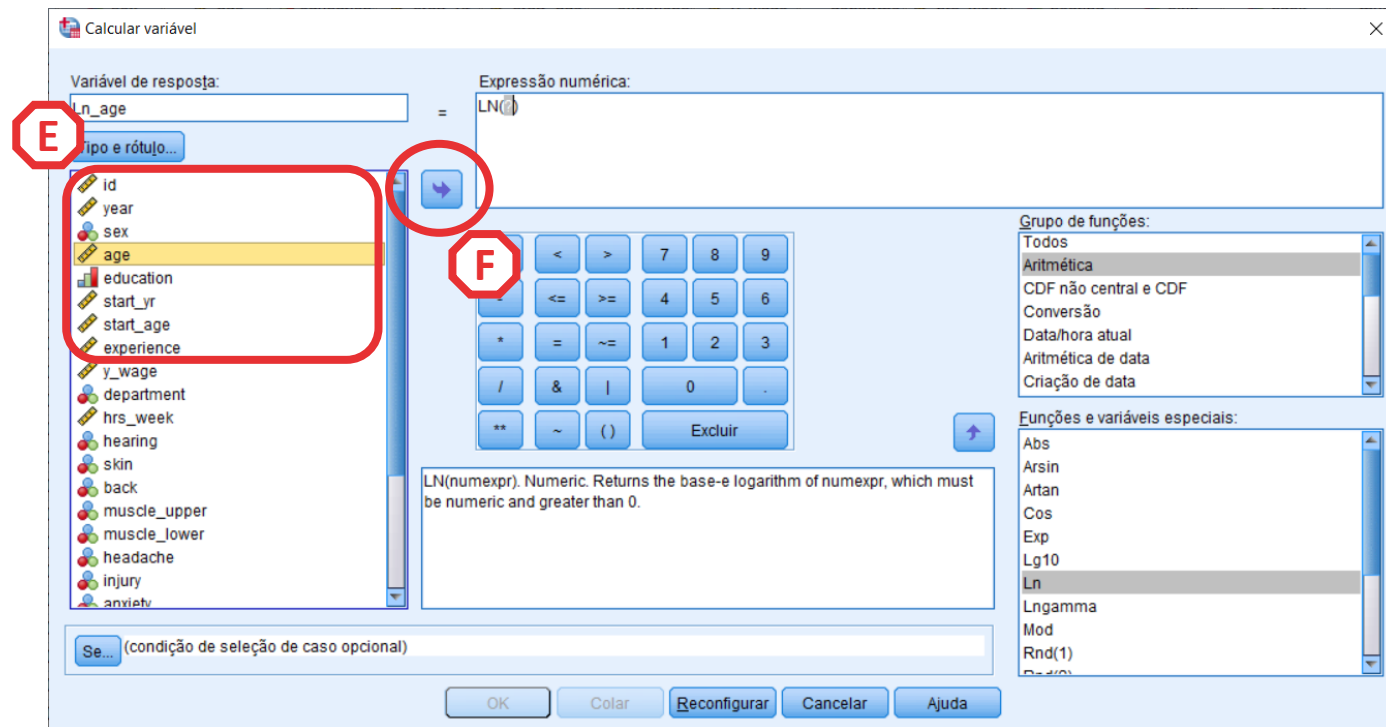
B

C

D

E

F



Pressuposto VI: Linearidade

- Selecionar 'Transformar' / 'Calcular variável'
- Definir o nome da variável que mede o logaritmo natural * da variável 'age'.
- Na caixa 'Grupo de funções', selecionar a opção 'Aritmética'.
- Duplo-Clique na opção Ln (Logaritmo Natural).
- Selecionar a variável que vai ser transformada ('age')...
- e colocá-la na dentro da função.
- Clicar em OK.

A

B

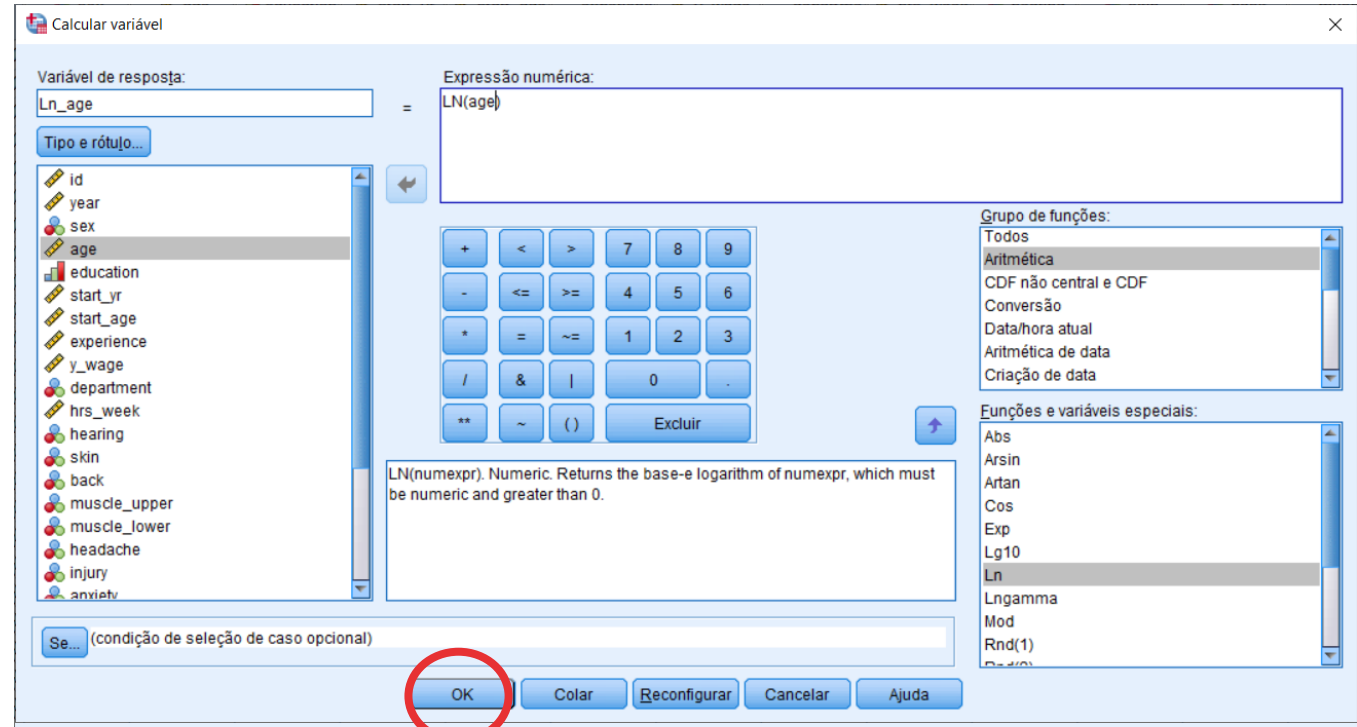
C

D

E

F

G



Pressuposto VI: Linearidade

- Selecionar 'Transformar' / 'Calcular variável'
- Definir o nome da variável que mede o logaritmo natural * da variável 'age'.
- Na caixa 'Grupo de funções', selecionar a opção 'Aritmética'.
- Duplo-Clique na opção Ln (Logaritmo Natural).
- Selecionar a variável que vai ser transformada ('age')...
- e colocá-la na dentro da função.
- Clicar em OK.

A

B

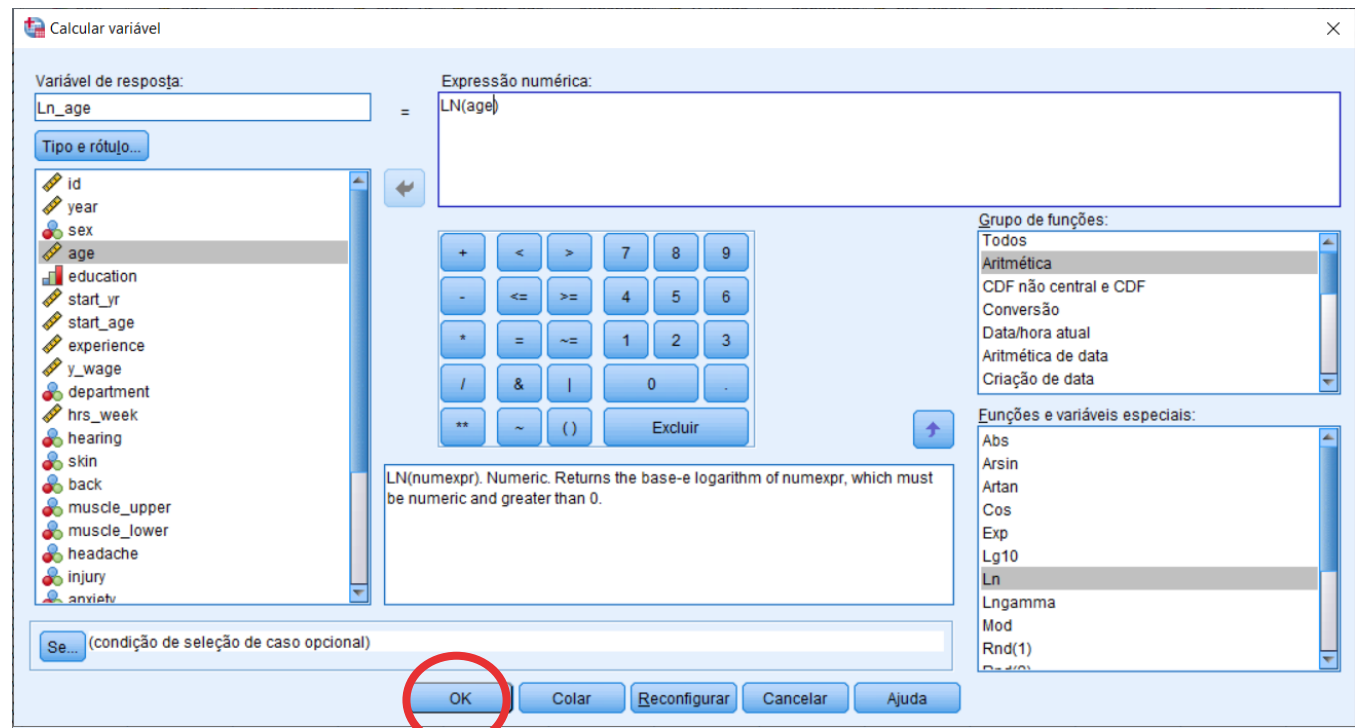
C

D

E

F

G



Pressuposto VI: Linearidade

- Vamos verificar se a recodificação foi bem sucedida

*ADGRH_BD_Aula12.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics

Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda

1 : Ln_age 3,46573590279973 Visível: 27 de 27 variáveis

	workcond_sat	evaluation	y_wage2	evaluation 2	satisfied	Ln_age	var	var	var	var	var
1	Not Very S...	49,08	17529,60	55,17	,00	3,47					
2	Satisfied	52,58	24843,10	51,75	1,00	3,69					
3	Satisfied	53,80	29874,72	54,79	1,00	3,74					
4	Satisfied	50,42	20994,46	58,30	1,00	3,56					
5	Not At All ...	48,43	10044,01	49,89	,00	3,58					
6	Satisfied	54,34	32321,53	52,74	1,00	3,76					
7	Not Very S...	49,65	18757,44	47,60	,00	3,58					
8	Not At All ...	46,84	9106,36	52,65	,00	3,43					
9	Satisfied	50,76	20432,69	48,48	1,00	3,66					
10	Not Very S...	53,33	29327,90	52,40	,00	3,71					
11	Satisfied	50,59	18223,72	47,21	1,00	3,61					
12	Satisfied	54,64	29668,47	48,39	1,00	3,81					
13	Not Very S...	56,64	35658,63	50,19	,00	3,91					
14	Not Very S...	55,36	35733,01	40,86	,00	3,87					
15	Not Very S...	56,09	36160,52	45,36	,00	3,87					

Visualização de dados Visualização de variável

O processador do IBM SPSS Statistics está pronto Unicode:OFF

Pressuposto VI: Linearidade

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



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

Arquivo Editar Visualizar Dados Transformar **Analisar** Gráficos Utilitários Extensões Janelas Ajuda

12 : y_wage2 29668.4665929268 8

	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

Relatórios
Estatística Descritiva
Estatísticas Bayesianas
Tabelas
Comparar Médias
Modelo Linear Geral
Modelos lineares generalizados
Modelos mistos
Correlacional
Regressão
Log linear
Redes neurais
Classificar
Redução de dimensão
Escala

Modelagem Linear Automática...
Linear...
Curva de estimação...
Quadrados Mínimos Parciais...
Logística binária...

Pressuposto VI: Linearidade

- Selecionar 'Analisar' / 'Regressão' / 'Logística binária' **A**
- Selecionar a variável 'satisfied' **B**
- Colocar na caixa 'Dependente' **C**

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

The screenshot shows the IBM SPSS Statistics interface with the 'Regressão logística' dialog box open. The 'Dependente:' field is empty, and the 'Covariáveis:' field is also empty. The 'Método:' is set to 'Inserir'. The 'Satisfied [satisfied]' variable is selected in the list on the left. Red circles and letters A, B, and C highlight the 'Analisar' menu, the 'Satisfied' variable, and the 'Dependente:' field respectively.

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

Pressuposto VI: Linearidade

- 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'



The screenshot shows the IBM SPSS Statistics interface with the 'Regressão logística' dialog box open. The 'Dependente:' field is set to 'Satisfied [satisfied]'. The 'Bloco 1 de 1' list contains 'sex', 'age', and 'education'. The 'Método:' dropdown is set to 'Inserir'. The 'Categórico...' button is highlighted with a red box and a red octagon icon with the letter D. The background shows a data table with columns 'id', 'year', 'y_wage', 'department', and 'hrs_week'.

id	year	y_wage	department	hrs_week
1	2018.00	22153.05	6	38.85
2	2018.00	27809.35	5	39.18
3	2018.00	29509.03	2	40.24
4	2018.00	21500.50	1	38.62
5	2018.00	16786.81	2	38.86
6	2018.00	31153.80	4	41.00
7	2018.00	20867.09	1	38.90
8	2018.00	14899.42	3	41.70
9	2018.00	20882.25	1	40.95
10	2018.00	29631.50	5	41.07
11	2018.00	22096.46	2	40.61
12	2018.00	28315.75	1	41.16
13	2018.00	31361.65	1	39.90
14	2018.00	30477.37	4	40.59
15	2018.00	32136.62	2	40.05

Pressuposto VI: Linearidade

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

The screenshot shows the IBM SPSS Statistics interface with a data file named 'ADGRH_BD_Aula11.sav'. A dialog box titled 'Regressão logística: definir variáveis de categoria' is open. In the 'Covariáveis' list on the left, the variable 'sex' is selected and highlighted with a red circle and the letter 'E'. A red arrow points from 'sex' to the 'Covariáveis categóricas' box on the right, which is also highlighted with a red circle and the letter 'F'. The dialog box also shows options for 'Alterar Contraste' (Contrast) and 'Categoria de referência' (Reference Category).

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

Pressuposto VI: Linearidade

- Seleccionar a variável 'sex' E
- Colocar na caixa 'Covariáveis categóricas'
- Definir a categoria 'male' (valor 1 *) como categoria de referência na variável 'sex'

* neste caso, escolhemos o valor 1 que é o mais baixo (i.e., o primeiro)

- Carregar no botão 'Alterar' H

The screenshot shows the IBM SPSS Statistics interface with a data file open. A dialog box titled 'Regressão logística: definir variáveis de categoria' is open. In the 'Covariáveis categóricas' list, 'sex(Indicador)' is selected. The 'Alterar Contraste' section shows 'Contraste' set to 'Indicador' and 'Categoria de referência' set to 'Primeiro'. The 'Alterar' button is highlighted with a red box and a red 'H' icon. Other buttons like 'Continuar', 'Cancelar', and 'Ajuda' are also visible. The background shows a data table with columns like 'id', 'year', 'wage', 'department', and 'hrs_week'.

Pressuposto VI: Linearidade

- Seleccionar a variável 'sex' E
- Colocar na caixa 'Covariáveis categóricas'
- Definir a categoria 'male' (valor 1 *) como categoria de referência na variável 'sex' F

* neste caso, escolhemos o valor 1 que é o mais baixo (i.e., o primeiro) G

- Carregar no botão 'Alterar' H

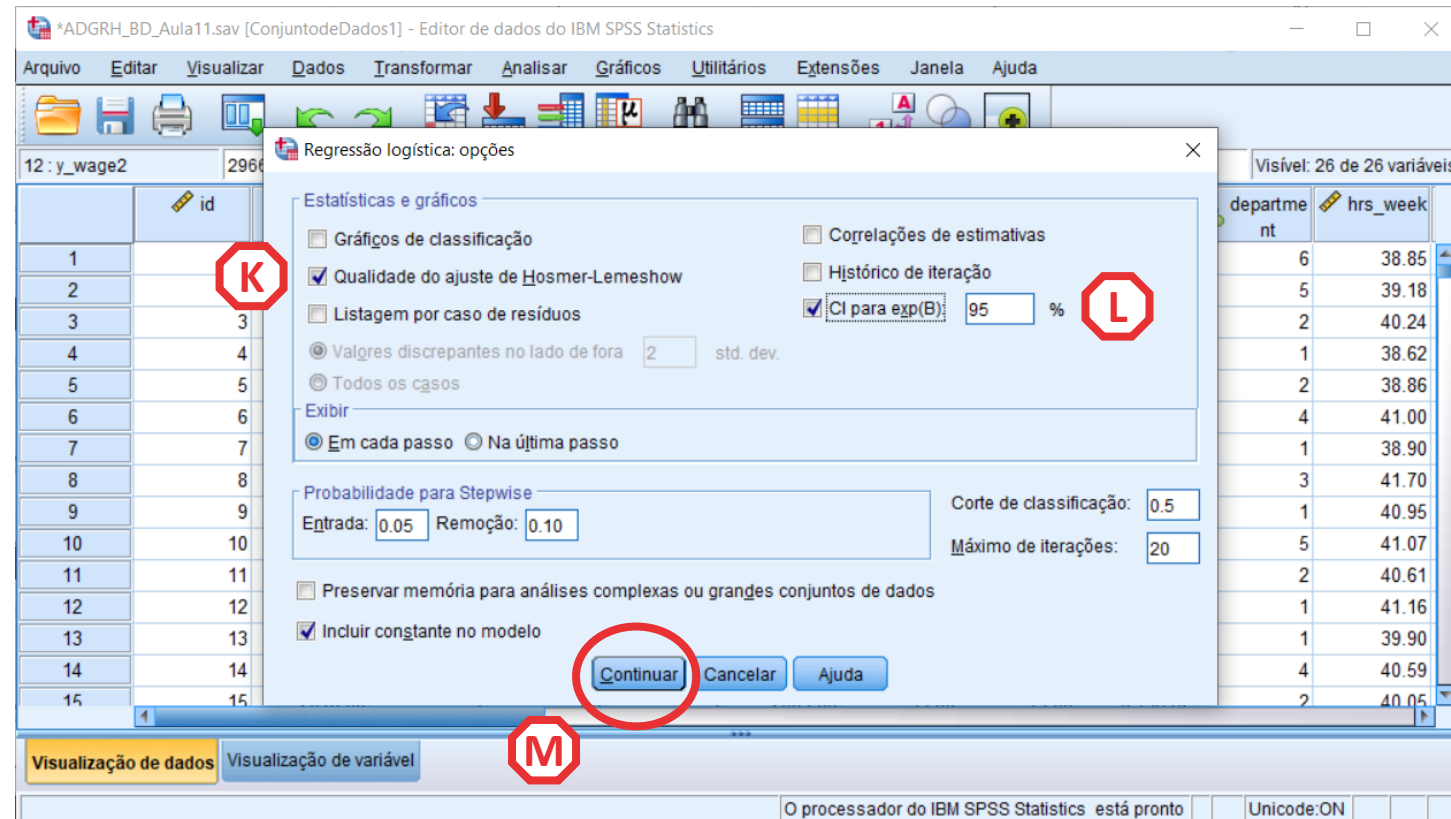
Exercício: Repetir esta operação para a variável 'education'

- Seleccionar 'Continuar' I

The screenshot displays the IBM SPSS Statistics interface. The main window shows a data viewer with columns: id, year, wage, department, and hrs_week. A dialog box titled 'Regressão logística: definir variáveis de categoria' is open. It has two main sections: 'Covariáveis:' and 'Covariáveis categóricas:'. The 'Covariáveis:' section contains 'age' and 'education'. The 'Covariáveis categóricas:' section contains 'sex(Indicador)'. Below these sections is the 'Alterar Contraste' section, which includes a 'Contraste:' dropdown set to 'Indicador' and an 'Alterar' button. The 'Categoria de referência:' section has two radio buttons: 'Último' and 'Primeiro', with 'Primeiro' selected. At the bottom of the dialog box, there are buttons for 'Continuar', 'Cancelar', and 'Ajuda'. The 'Continuar' button is circled in red. The background data viewer shows a table with 15 rows of data.

Pressuposto VI: Linearidade

- Clicar no botão 'Opções' J
- Selecionar a opção 'Qualidade do ajuste de Hosmer-Lemeshow' K
- Selecionar a opção 'CI para exp(B)' L
- Selecionar 'Continuar' M



Pressuposto VI: Linearidade

- Só nos falta criar o termo de interação entre a variável 'age' e variável 'ln_age'.

Regressão logística

Dependente: Satisfeito/a com Condições de ...

Bloco 1 de 1

Anterior Próximo

Covariáveis:

- sex(Cat)
- age
- education(Cat)

Método: Inserir

Variável de seleção:

OK Colar Reconfigurar Cancelar Ajuda

Categorico... Salvar... Opções... Estilo... Bootstrap...

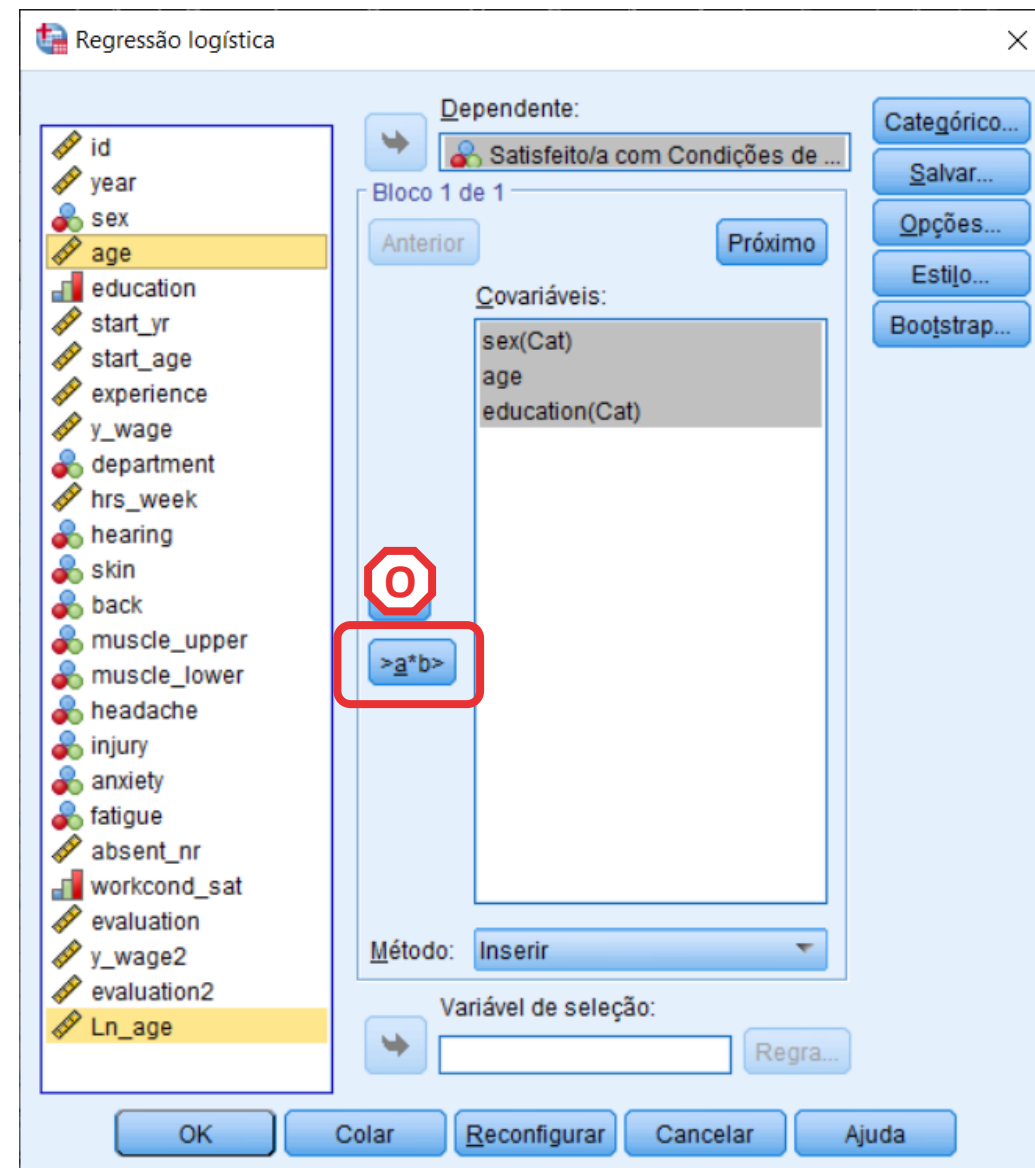
Pressuposto VI: Linearidade

- Só nos falta criar o termo de interação entre a variável 'age' e variável 'ln_age'.
- Enquanto clicamos na tecla 'CTRL', selecionamos as duas variáveis que vão formar o termo de interação.



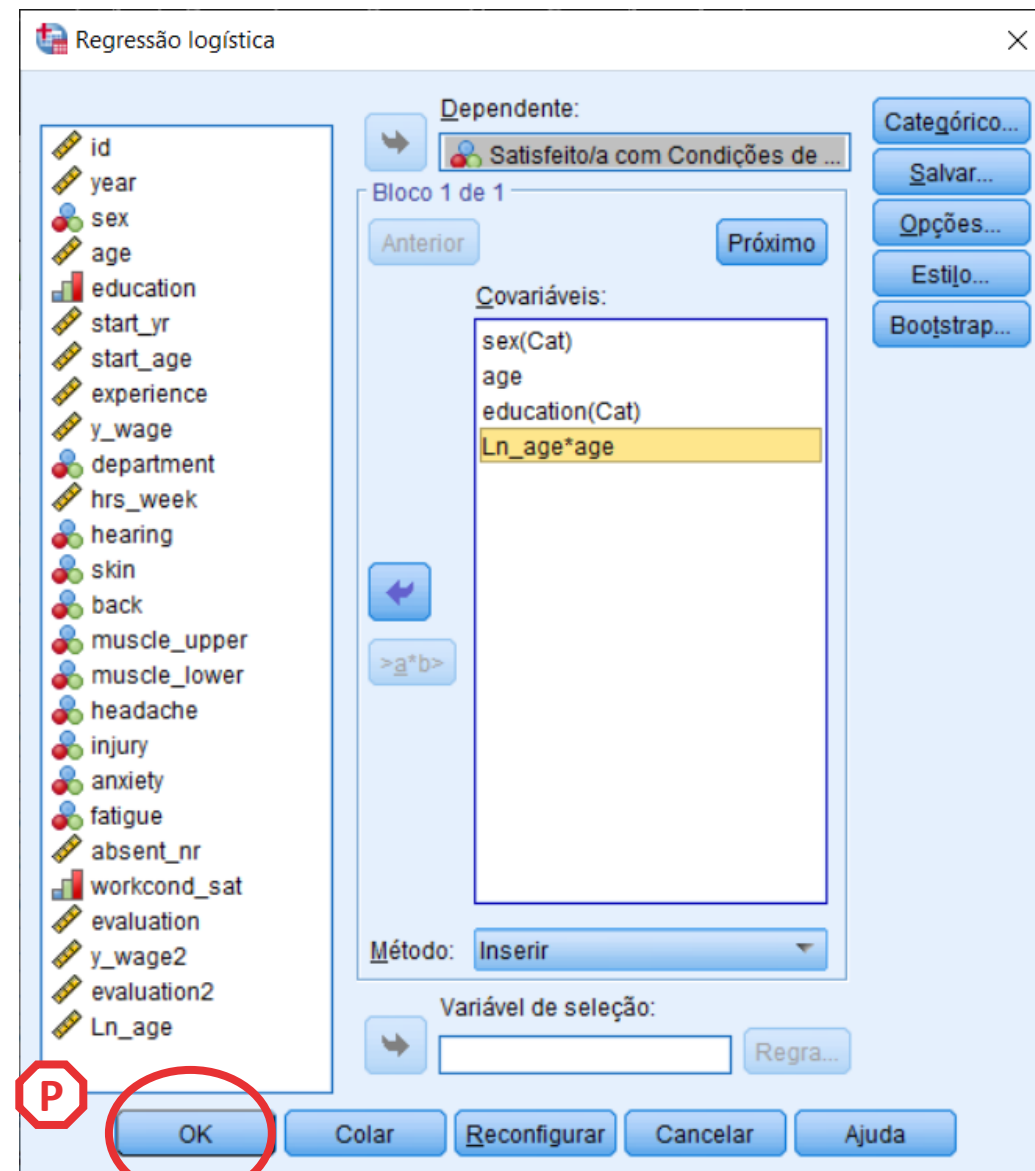
Pressuposto VI: Linearidade

- Só nos falta criar o termo de interação entre a variável 'age' e variável 'ln_age'.
- Enquanto clicamos na tecla 'CTRL',
- De seguida clicamos no botão que gera automaticamente termos de interação.



Pressuposto VI: Linearidade

- Só nos falta criar o termo de interação entre a variável 'age' e variável 'ln_age'.
- Enquanto clicamos na tecla 'CTRL', selecionamos as duas variáveis que vão formar o termo de interação.
- De seguida clicamos no botão que gera automaticamente termos de interação.
- Selecionar 'OK'



Pressuposto VI: Linearidade

- Se o 'termo de interação' não for estatisticamente significativo, i.e. se o valor de Sig. for superior a 0.01/0.05/0.1, então assume-se que a relação entre a variável dependente e o logaritmo da variável independente é linear.

*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

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

Saída

- Log
- Regressão logística
 - Título
 - Observações
 - Conjunto de dados ativo
 - Resumo de processament
 - Codificação de variável dep
 - Codificações de variáveis d
 - Bloco 0: Bloco Inicial
 - Título
 - Tabela de Classificaç
 - Variáveis na equação
 - Variáveis não presente
 - Bloco 1: Método = Enter
 - Título
 - Testes de Omnibus de
 - Resumo do modelo
 - Teste de Hosmer e Le
 - Tabela de contingênci
 - Tabela de Classificaç
 - Variáveis na equação

6	269	264,149	230	234,851	499
7	260	259,219	241	241,781	501
8	259	259,872	254	253,128	513
9	213	194,543	300	318,457	513
10	160	178,457	324	305,543	484

Tabela de Classificação^a

Observado	Previsto		Porcentagem correta
	Satisfeito/a com Condições de Trabalho	Satisfeito/a com Condições de Trabalho	
Etapa 1 Satisfeito/a com Condições de Trabalho	,00	2275	85,9
	1,00	1728	26,5
Porcentagem global			58,0

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
Etapa 1 ^a sex(1)	-,040	,058	,474	1	,491	,961	,858	1,076
age	,169	,273	,384	1	,535	1,185	,693	2,025
education			149,465	4	,000			
education(1)	,127	,092	1,896	1	,168	1,135	,948	1,359
education(2)	,387	,092	17,855	1	,000	1,473	1,231	1,763
education(3)	,477	,092	26,963	1	,000	1,611	1,346	1,929
education(4)	1,037	,094	122,887	1	,000	2,819	2,347	3,387
Ln_age by age	-,035	,058	,369	1	,544	,965	,862	1,082
Constante	-2,077	2,364	,772	1	,380	,125		

a. Variável(is) inserida(s) na etapa 1: sex, age, education, Ln_age * age .

O processador do IBM SPSS Statistics está pronto | Unicode:OFF | H: 8,02, W: 19 cm

Pressuposto VI: Linearidade

- Se o 'termo de interação' não for estatisticamente significativo, i.e. se o valor de Sig. for superior a 0.01/0.05/0.1, então assume-se que a relação entre a variável dependente e o logaritmo da variável independente é linear.
- Neste caso, o 'termo de interação' não é estatisticamente significativo.
- Cumpre-se o pressuposto da existência de uma relação linear entre a variável dependente e o logaritmo da variável de razão.
- O modelo poderá ser estimado apenas com as variáveis originais.

*Saída1 [Documento1] - Visualizador do IBM SPSS Statistics

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

Saída

- Log
- Regressão logística
 - Título
 - Observações
 - Conjunto de dados ativo
 - Resumo de processamento
 - Codificação de variável dependente
 - Codificações de variáveis independentes
 - Bloco 0: Bloco Inicial
 - Título
 - Tabela de Classificação
 - Variáveis na equação
 - Variáveis não presentes
 - Bloco 1: Método = Enter
 - Título
 - Testes de Omnibus de
 - Resumo do modelo
 - Teste de Hosmer e Lemeshow
 - Tabela de contingência
 - Tabela de Classificação
 - Variáveis na equação

	6	7	8	9	10
	269	260	259	213	160
	264,149	259,219	259,872	194,543	178,457
	230	241	254	300	324
	234,851	241,781	253,128	318,457	305,543
	499	501	513	513	484

Tabela de Classificação^a

Observado	Previsto		Porcentagem correta
	Satisfeito/a com Condições de Trabalho	1,00	
Etapa 1 Satisfeito/a com Condições de Trabalho	,00	2275	85,9
	1,00	1728	26,5
Porcentagem global			58,0

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
Etapa 1 ^a sex(1)	-,040	,058	,474	1	,491	,961	,858	1,076
age	,169	,273	,384	1	,535	1,185	,693	2,025
education			149,465	4	,000			
education(1)	,127	,092	1,896	1	,168	1,135	,948	1,359
education(2)	,387	,092	17,855	1	,000	1,473	1,231	1,763
education(3)	,477	,092	26,963	1	,000	1,611	1,346	1,929
education(4)	1,037	,094	122,887	1	,000	2,819	2,347	3,387
Ln_age by age	-,035	,058	,369	1	,544	,965	,862	1,082
Constante	-2,077	2,364	,772	1	,380	,125		

a. Variável(is) inserida(s) na etapa 1: sex, age, education, Ln_age * age .

O processador do IBM SPSS Statistics está pronto | Unicode:OFF | H: 8,02, W: 19 cm

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 mostrou-se estatisticamente significativo e portanto útil para estimar a variável dependente ($\chi^2_{(7)} = 159,422, p < 0,05$). Alguns estudos foram realizados para averiguar a adequabilidade do modelo aos dados. Observou-se a distribuição das variáveis independentes em função da variável dependente, não encontrando omissões de dados. Nenhuma das variáveis independentes se mostrou perfeitamente associada com as probabilidades de satisfação, e os estudos de associação entre as variáveis não sugerem níveis de multicolinearidade elevados, o que é corroborado pela estatística de VIF ($VIF < 3$). Por fim, avaliou-se a presença de observações influentes com base nas distâncias de Cook. Admitindo um critério de $0,008 (4 / N)$, admite-se a inexistência de valores extremos influentes na estimação do modelo, na medida em que as distâncias calculadas para todos os casos são inferiores ao critério.

Ainda há tempo?

**Analisar
o seguinte modelo
+ adequação aos
pressupostos**

VD: positive_eval (tem de ser criada!)

VI: sex, age, education