

Licenciaturas em

Gestão e Finanças e Contabilidade

**Estatística 2**

**Modelo de regressão linear com SPSS**

**Folhas de apoio às aulas**

*2023/2024*

|  |
| --- |
| UM EXEMPLO DE REGRESSÃO LINEAR MÚLTIPLA |

Vamos utilizar a base de dados **EXEMPLO\_RLM.sav**, com o intuito de explicar os salários de um conjunto de trabalhadores de um dado sector de actividade (**Salario**) como função do número de anos de estudo (**Anosestudo**), do número de anos de formação específica (**AnosFespecifica**), do número de anos de experiência profissional (**AnosExperiencia**).

Algumas estatísticas descritivas obtidas:

**Descriptive Statistics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | N | Minimum | Maximum | Mean | Std. Deviation |
| Salário Salário | 70 | 139 | 1355 | 667,17 | 223,567 |
| Anosestudo Anos de estudo | 70 | 2 | 18 | 9,30 | 3,200 |
| AnosFespecifica Anos de formação específica | 70 | 0 | 5 | 1,70 | 1,312 |
| AnosExperiencia Anos de experiência profissional | 70 | 4 | 24 | 11,77 | 4,709 |
| Valid N (listwise) | 70 |  |  |  |  |

Grau de correlação linear entre as variáveis quantitativas em estudo:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Correlations** | | | | | |
|  | | Salário | Anos de estudo | Anos de formação específica | Anos de experiência profissional |
| Salário | Pearson Correlation | 1 | ,946\*\* | ,324\*\* | ,191 |
| Sig. (2-tailed) |  | ,000 | ,006 | ,114 |
| N | 70 | 70 | 70 | 70 |
| Anos de estudo | Pearson Correlation | ,946\*\* | 1 | ,277\* | ,183 |
| Sig. (2-tailed) | ,000 |  | ,020 | ,130 |
| N | 70 | 70 | 70 | 70 |
| Anos de formação específica | Pearson Correlation | ,324\*\* | ,277\* | 1 | ,228 |
| Sig. (2-tailed) | ,006 | ,020 |  | ,058 |
| N | 70 | 70 | 70 | 70 |
| Anos de experiência profissional | Pearson Correlation | ,191 | ,183 | ,228 | 1 |
| Sig. (2-tailed) | ,114 | ,130 | ,058 |  |
| N | 70 | 70 | 70 | 70 |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). | | | | | |
| \*. Correlation is significant at the 0.05 level (2-tailed). | | | | | |

**MODELO DE RGERESSÃO LINEAR MÚLTIPLA**

🡺 Diagramas de dispersão para as variáveis duas a duas

***Graphs***

***Scatter/Dot***

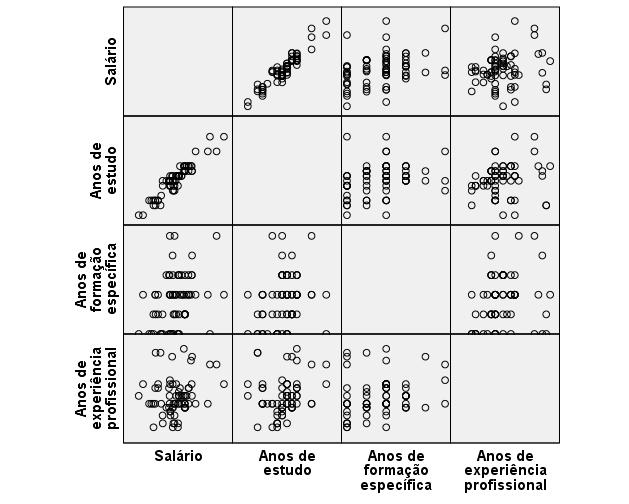
***Matrix Scatter***

***Define***

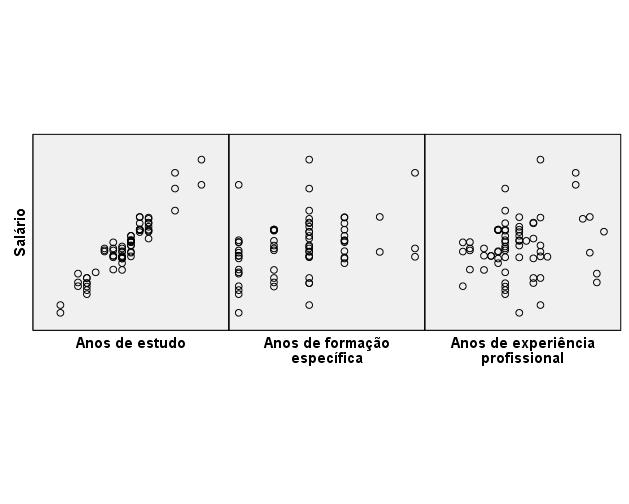
***Matrix variables: Salario, Anosestudo, AnosFespecifica,***

***AnosExperiencia***

**OUTPUTS \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**



Ou, com alguma sintaxe,

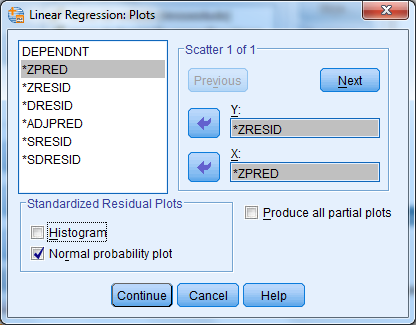
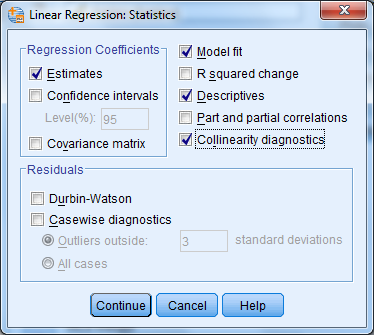
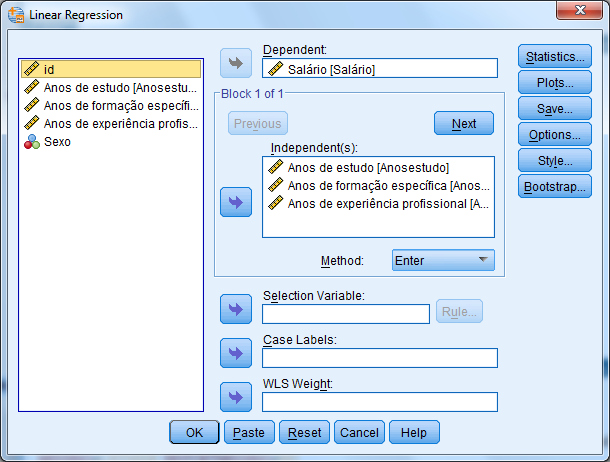


🡺 Estimação do Modelo de Regressão Linear Múltipla (forçando a entrada de todas as variáveis quantitativas – método Enter)

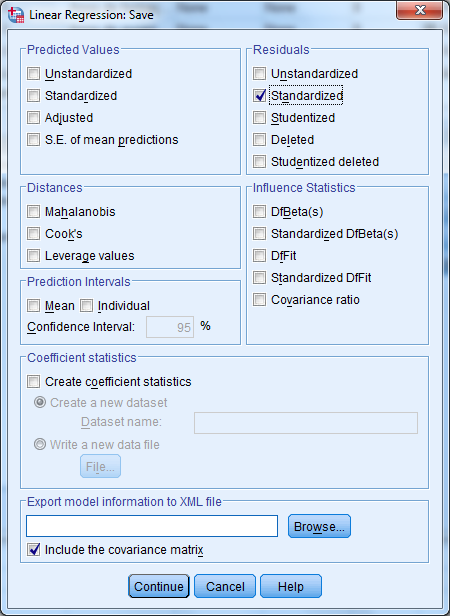
***Analize***

***Regression***

***Linear***



Os resíduos (estandartizados) podem ainda ser guardados na base de dados para posterior análise:



**OUTPUTS \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

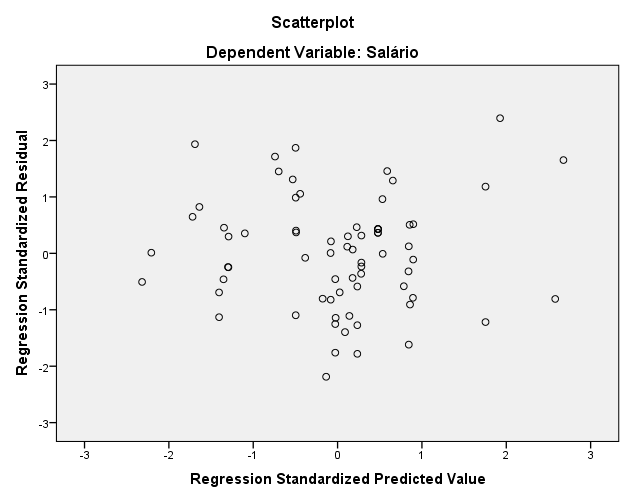
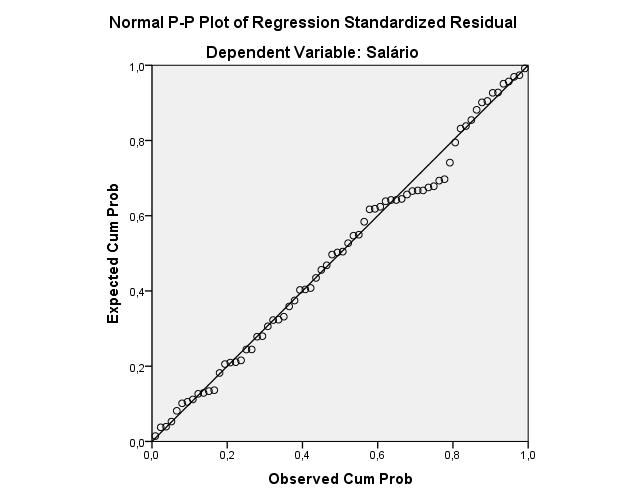
|  |  |  |  |
| --- | --- | --- | --- |
| **Variables Entered/Removeda** | | | |
| Model | Variables Entered | Variables Removed | Method |
| 1 | Anos de experiência profissional, Anos de estudo, Anos de formação específicab | . | Enter |
| a. Dependent Variable: Salário | | | |
| b. All requested variables entered. | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | ,948a | ,898 | ,894 | 72,909 |
| a. Predictors: (Constant), Anos de experiência profissional, Anos de estudo, Anos de formação específica | | | | |
| b. Dependent Variable: Salário | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 3097922,903 | 3 | 1032640,968 | 194,261 | ,000b |
| Residual | 350839,040 | 66 | 5315,743 |  |  |
| Total | 3448761,943 | 69 |  |  |  |
| a. Dependent Variable: Salário | | | | | | |
| b. Predictors: (Constant), Anos de experiência profissional, Anos de estudo, Anos de formação específica | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
| B | Std. Error | Beta | Tolerance | VIF |
| 1 | (Constant) | 42,724 | 32,202 |  | 1,327 | ,189 |  |  |
| Anos de estudo | 64,708 | 2,878 | ,926 | 22,484 | ,000 | ,908 | 1,101 |
| Anos de formação específica | 11,116 | 7,091 | ,065 | 1,568 | ,122 | ,891 | 1,123 |
| Anos de experiência profissional | ,320 | 1,930 | ,007 | ,166 | ,869 | ,933 | 1,072 |
| a. Dependent Variable: Salário | | | | | | | | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | | | | | |
| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions | | | |
| (Constant) | Anos de estudo | Anos de formação específica | Anos de experiência profissional |
| 1 | 1 | 3,587 | 1,000 | ,01 | ,01 | ,02 | ,01 |
| 2 | ,266 | 3,675 | ,03 | ,02 | ,96 | ,04 |
| 3 | ,100 | 5,976 | ,02 | ,36 | ,00 | ,74 |
| 4 | ,047 | 8,698 | ,95 | ,61 | ,01 | ,21 |
| a. Dependent Variable: Salário | | | | | | | |

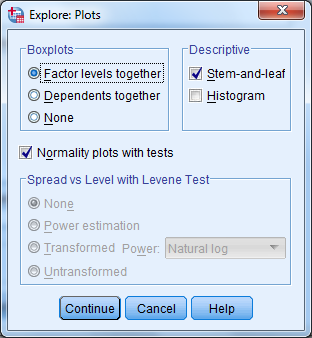
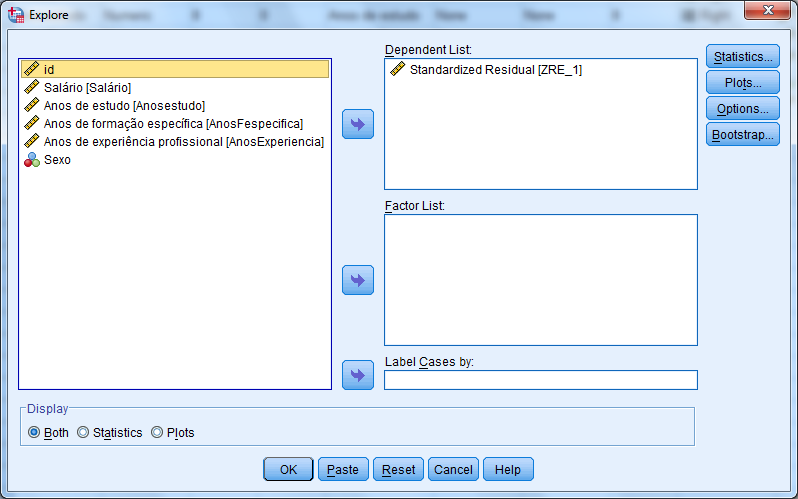


🡺 Análise exploratória dos resíduos

***Analize***

***Descriptive Statistics***

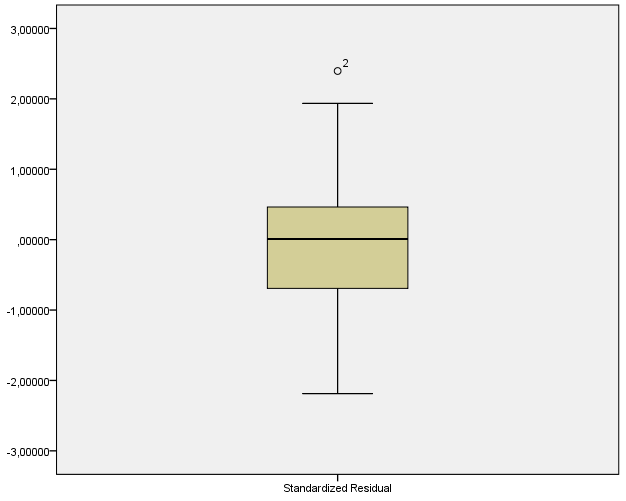
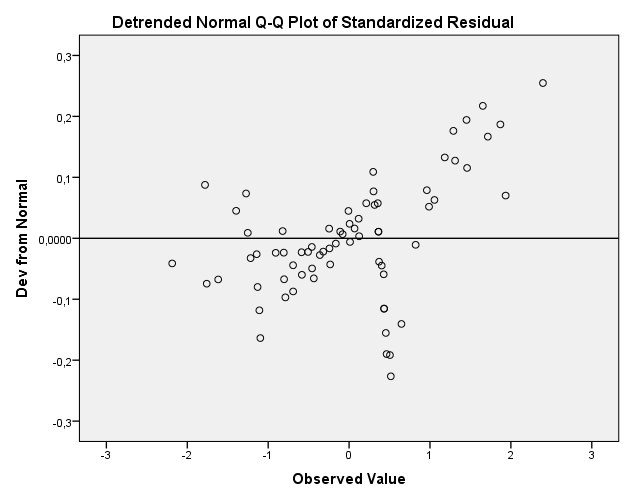
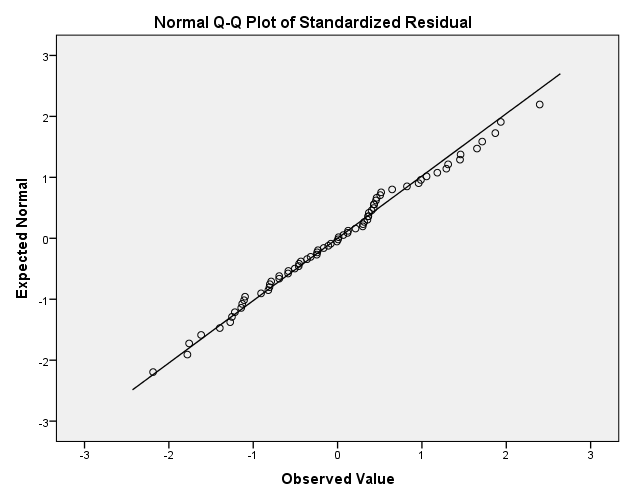
***Explore***



**OUTPUTS \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Descriptives** | | | | | | |
|  | | | Statistic | | Std. Error | |
| Standardized Residual | Mean | | | ,0000000 | | ,11689566 | |
| 95% Confidence Interval for Mean | Lower Bound | | -,2332005 | |  | |
| Upper Bound | | ,2332005 | |  | |
| 5% Trimmed Mean | | | -,0082780 | |  | |
| Median | | | ,0086494 | |  | |
| Variance | | | ,957 | |  | |
| Std. Deviation | | | ,97801929 | |  | |
| Minimum | | | -2,18698 | |  | |
| Maximum | | | 2,39583 | |  | |
| Range | | | 4,58282 | |  | |
| Interquartile Range | | | 1,18957 | |  | |
| Skewness | | | ,156 | | ,287 | |
| Kurtosis | | | -,265 | | ,566 | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tests of Normality** | | | | | | |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| Standardized Residual | ,085 | 70 | ,200\* | ,990 | 70 | ,857 |
| \*. This is a lower bound of the true significance. | | | | | | |
| a. Lilliefors Significance Correction | | | | | | |



🡺 Interpretação dos resultados

## Modelo: com



No caso



## Verificação dos pressupostos:

### 1. Linearidade da relação entre cada uma das variáveis *X* e *Y*

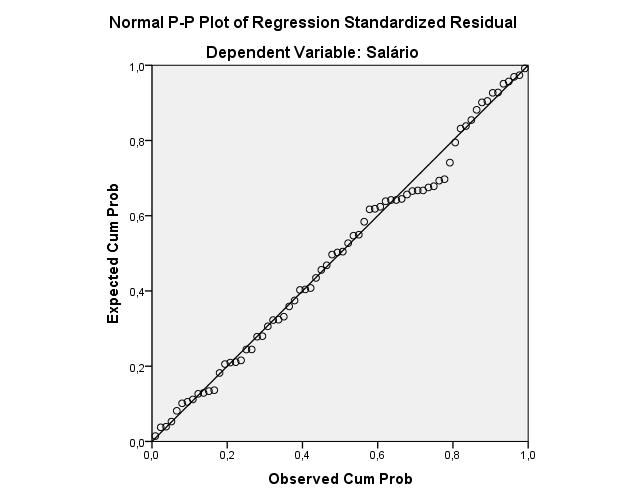
(verificar com *scatterplots* (*Xi, Yi*); análise das correlações)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Correlations** | | | | | |
|  | | Salário | Anos de estudo | Anos de formação específica | Anos de experiência profissional |
| Salário | Pearson Correlation | 1 | ,946\*\* | ,324\*\* | ,191 |
| Sig. (2-tailed) |  | ,000 | ,006 | ,114 |
| N | 70 | 70 | 70 | 70 |
| Anos de estudo | Pearson Correlation | ,946\*\* | 1 | ,277\* | ,183 |
| Sig. (2-tailed) | ,000 |  | ,020 | ,130 |
| N | 70 | 70 | 70 | 70 |
| Anos de formação específica | Pearson Correlation | ,324\*\* | ,277\* | 1 | ,228 |
| Sig. (2-tailed) | ,006 | ,020 |  | ,058 |
| N | 70 | 70 | 70 | 70 |
| Anos de experiência profissional | Pearson Correlation | ,191 | ,183 | ,228 | 1 |
| Sig. (2-tailed) | ,114 | ,130 | ,058 |  |
| N | 70 | 70 | 70 | 70 |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). | | | | | |
| \*. Correlation is significant at the 0.05 level (2-tailed). | | | | | |

### 2. Normalidade: os seguem distribuição aproximadamente normal



(verificar com um *Normal P-P plot* ou com um teste de Kolmogorov-Smirnov à normalidade dos resíduos)



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tests of Normality** | | | | | | |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| Standardized Residual | ,085 | 70 | ,200\* | ,990 | 70 | ,857 |
| \*. This is a lower bound of the true significance. | | | | | | |
| a. Lilliefors Significance Correction | | | | | | |

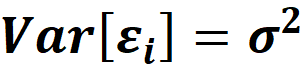
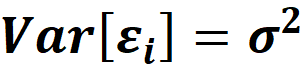
### 3. O valor esperado dos resíduos é nulo:



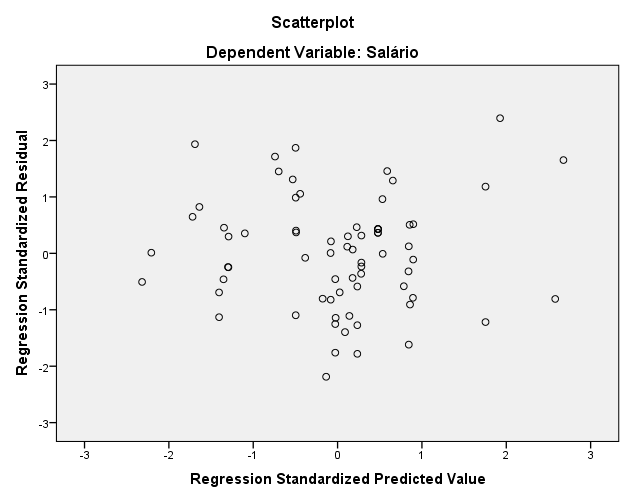
(este pressuposto **não é passível de verificação** porque os resíduos **são estimados de tal forma que a soma das estimativas é sempre nula**, o que se pode verificar na média da variável *Residual*, aquando do cálculo das *Residual Statistics*)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Residuals Statisticsa** | | | | | |
|  | Minimum | Maximum | Mean | Std. Deviation | N |
| Predicted Value | 175,98 | 1234,50 | 667,17 | 211,890 | 70 |
| Residual | -159,451 | 174,678 | **,000** | 71,307 | 70 |
| Std. Predicted Value | -2,318 | 2,677 | ,000 | 1,000 | 70 |
| Std. Residual | -2,187 | 2,396 | ,000 | ,978 | 70 |
| a. Dependent Variable: Salário | | | | | |

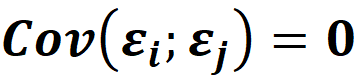
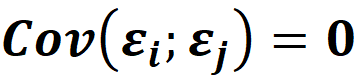
### 4. Homocedasticidade: a variância dos é constante, i.e.,



(verificar com um *scatterplot*  e ver se aumenta, ou não, a dispersão dos )



### 5. Inexistência de autocorrelação: os erros são independentes, i.e.,



(a sua validação apenas é relevante se os dados forem cronológicos)

### 6. Inexistência de multicolinearidade: as variáveis independentes não estão fortemente correlacionadas

(ver a matriz de correlações; ver os valores de *Tolerance (ok se >0,1)* e *VIF (ok se <10)*, analisar *Condition Indexes (ok se <30)*)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
| B | Std. Error | Beta | Tolerance | VIF |
| 1 | (Constant) | 42,724 | 32,202 |  | 1,327 | ,189 |  |  |
| Anos de estudo | 64,708 | 2,878 | ,926 | 22,484 | ,000 | ,908 | 1,101 |
| Anos de formação específica | 11,116 | 7,091 | ,065 | 1,568 | ,122 | ,891 | 1,123 |
| Anos de experiência profissional | ,320 | 1,930 | ,007 | ,166 | ,869 | ,933 | 1,072 |
| a. Dependent Variable: Salário | | | | | | | | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | | | | | |
| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions | | | |
| (Constant) | Anos de estudo | Anos de formação específica | Anos de experiência profissional |
| 1 | 1 | 3,587 | 1,000 | ,01 | ,01 | ,02 | ,01 |
| 2 | ,266 | 3,675 | ,03 | ,02 | ,96 | ,04 |
| 3 | ,100 | 5,976 | ,02 | ,36 | ,00 | ,74 |
| 4 | ,047 | 8,698 | ,95 | ,61 | ,01 | ,21 |
| a. Dependent Variable: Salário | | | | | | | |

## Teste F à validade do modelo (quadro de ANOVA):

H0 : O MRLM **não é** adequado vs H1 : O MRLM **é** adequado

Ou

H0: vs H1:



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 3097922,903 | 3 | 1032640,968 | 194,261 | ,000b |
| Residual | 350839,040 | 66 | 5315,743 |  |  |
| Total | 3448761,943 | 69 |  |  |  |
| a. Dependent Variable: Salário | | | | | | |
| b. Predictors: (Constant), Anos de experiência profissional, Anos de estudo, Anos de formação específica | | | | | | |

Neste caso p-value = 0.000 < α = 0.05, levando à rejeição da H0. Assim, o modelo é adequado.

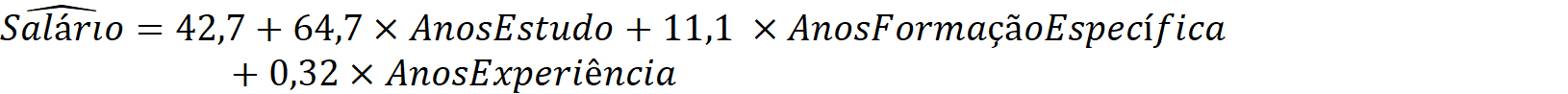
## Percentagem de variância explicada pelo modelo

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | ,948a | ,898 | ,894 | 72,909 |
| a. Predictors: (Constant), Anos de experiência profissional, Anos de estudo, Anos de formação específica | | | | |
| b. Dependent Variable: Salário | | | | |

89,8% da variância total é explicada pela regressão (comentário)

## Estimação dos parâmetros do modelo:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
| B | Std. Error | Beta | Tolerance | VIF |
| 1 | (Constant) | 42,724 | 32,202 |  | 1,327 | ,189 |  |  |
| Anos de estudo | 64,708 | 2,878 | ,926 | 22,484 | ,000 | ,908 | 1,101 |
| Anos de formação específica | 11,116 | 7,091 | ,065 | 1,568 | ,122 | ,891 | 1,123 |
| Anos de experiência profissional | ,320 | 1,930 | ,007 | ,166 | ,869 | ,933 | 1,072 |
| a. Dependent Variable: Salário | | | | | | | | |



Sendo a estimativa da dispersão dos erros (identificada por Std. Error of Estimate).



Mas os Anos de Formação Específica e os Anos de Experiência Profissional não revelam coeficientes significativamente diferentes de 0 (ver p-values dos testes t). Seria de ponderar refazer o modelo SEM estas variáveis.

## Interpretação dos coeficientes

|  |  |  |
| --- | --- | --- |
| **Coeficiente** | **B** |  |
| (Constant) | 42,724 | O salário médio estimado, sem a influência das variáveis independentes consideradas é de 42,7 u.m. |
| Anos de estudo | 64,708 | Por cada ano de estudo adicional espera-se um **aumento** de 64,7 u.m. no salário, mantendo as restantes variáveis constantes |
| Anos de formação específica | 11,116 | Por cada ano de formação específica adicional espera-se um aumento de 11,1 u.m. no salário, mantendo tudo o resto constante |
| Anos de experiência profissional | ,320 | Por cada ano adicional de experiência profissional espera-se um aumento de 0,3 u.m. no salário, mantendo tudo o resto constante |

Os coeficientes estandartizados (Beta) são utilizado essencialmente para identicar quais as variáveis mais influentes para o modelo, em termos relativos. A variável explicativa mais importante é Anos de estudo. A interpretação destes coeficientes faz-se em termos de desvios-padrão. Por exemplo, se , significa que, por cada aumento de um desvio-padrão nos anos de estudo, espera-se um aumento de 0,926 desvios-padrão no salário.

