

UNIVERSIDADE DE PASSO FUNDO FACULDADE DE CIÊNCIAS ECONÔMICAS, ADMINISTRATIVAS E CONTÁBEIS CENTRO DE PESQUISA E EXTENSÃO DA FEAC (www.upf.br/cepeac)

Texto para discussão

Texto para discussão Nº 01/2019

Regressão Linear Múltipla Como simplificar por meio do Excel e SPSS?

André da Silva Pereira

Thayane Woellner Sviercoski Manosso

Emanuele Canali Fossatti

Sandra Mara Berti

A presente apostila foi elaborada com o objetivo de orientar os leitores em relação ao uso do Excel e do SPSS durante a Regressão Linear Múltipla, a fim de clarear e simplificar os passos deste processo em ambas as ferramentas.

Regressão Linear Múltipla

Como simplificar por meio do Excel e SPSS?

Thayane Woellner Sviercoski Manosso Emanuele Canali Fossatti Sandra Mara Berti



SUMÁRIO)
---------	---

INTRODUÇÃO	. 3
INVESTIGĂÇÃO DESCRITIVA	. 4
INVESTIGAÇÃO CORRELACIONAL	. 5
INVESTIGAÇÃO EXPERIMENTAL	. 6
A Regressão Linear Múltipla (RLM)	7
1. Definição das variáveis	8
2. Desenho do gráfico de dispersão	8
3. Montagem da equação da RLM	9
4. Rodar a RLM	10
5. Substituir os dados na equação da RLM	10
6. Interpretação dos resultados	14
* Variáveis Dummy	14
Exemplo - Exercício Prático	16
Utilizando o EXCEL	18
Passo 1: Definição das Variáveis	18
Passo 2: Desenho do gráfico de dispersão	19
Passo 3: Montagem da equação da RLM	22
Passo 4: Rodar a RLM	22
Passo 5: Substituir os dados na equação da RLM	27
Passo 6: Interpretação dos resultados	29
* Variáveis Dummy	29
Escolhendo o melhor modelo de regressão	38
Utilizando o SPSS	44
Passo 1: Definição das Variáveis (Figura 49)	46
Passo 2: Desenho do gráfico de dispersão	46
Passo 3: Montagem da equação da RLM	50
Passo 4: Rodar a RLM	50
Passo 5: Substituir os dados na equação da RLM	55
Passo 6: Interpretação dos resultados	56
*Variáveis Dummy	56
Escolhendo o melhor modelo de regressão	62
REFERÊNCIAS	66

INTRODUÇÃO

Estatística é uma palavra que por si só, assusta. Isso ocorre em razão de algumas pessoas simplesmente não gostarem de números, o que não é raro, especialmente em áreas como as Ciências Sociais.

O que acontece, na verdade, é que tantos símbolos e palavras estranhas nos confundem, não permitindo que analisemos os números de forma lógica e simples. Além disso, não há muito material disponível que descomplique a estatística. Pensando nisso, desenvolvemos esta apostila, a qual tem por objetivo explicar de forma prática e bastante ilustrativa como realizar a análise de dados por meio da **Regressão Linear Múltipla** em dois softwares: o EXCEL, velho conhecido e amigo de todos que têm conhecimento básico em informática; e o SPSS, específico para cálculos estatísticos, muito comum àqueles que realizam análises estatísticas com mais frequência. Para isso, apresentaremos um exemplo prático, o qual será utilizado ao longo de toda a apostila.

Mas primeiro é necessário entender **O QUÊ** é, **PORQUÊ** precisamos usar e **COMO** chegamos até a Regressão Linear Múltipla, não é mesmo? Afinal, não esperamos que nossos leitores possuam algum conhecimento prévio de estatística para entender este material.

ESTATÍSTICA E CIÊNCIAS SOCIAIS

Sempre que buscamos entender ou prever um fenômeno, precisamos de dados que nos auxiliem, os quais podem ser coletados e analisados de diversas maneiras. A estatística faz parte da análise de dados **quantitativa**, um método que utiliza a linguagem matemática para descrever as causas de um fenômeno e as relações entre suas variáveis, por exemplo (FONSECA, 2002).

Para que possamos generalizar os resultados que obtemos por meio das análises estatísticas, precisamos que nossa amostra contemple uma representação viável da nossa população. A isso, damos o nome de **inferência estatística**, a qual nos permite tomar decisões generalizadas para toda a população, baseadas em uma amostragem.

As investigações com base em uma amostra, segundo Almeida e Freire (2000), podem ser de três tipos: **descritiva, correlacional ou experimental**. A investigação **descritiva** trata das variáveis separadamente, enquanto as investigações **correlacional** e **experimental** tratam da associação entre uma ou mais variáveis. A seguir, falaremos sobre cada uma delas.

INVESTIGAÇÃO DESCRITIVA

As técnicas que nos proporcionam analisar e interpretar as informações básicas dos dados coletados fazem parte de um conjunto chamado de **Investigação ou Estatística Descritiva.** A descrição dos dados é fornecida pelas **medidas de posição,** também chamadas de tendência central, as quais apresentam a frequência dos dados, conhecidas como **média, mediana** e **moda**; e **medidas de dispersão**, que nos dizem o quão dispersos ou distantes um do outro estão os valores de um conjunto de dados e são chamadas de **variância** e **desvio padrão**. Para entendermos melhor e de forma mais fácil, não utilizaremos as fórmulas matemáticas para explicar cada um desses termos, já que os softwares costumam nos fornecer esses valores quando solicitado, e o mais importante é saber interpretá-los.

- <u>Média:</u> é o valor médio dos dados e representa onde os dados se concentram. Para obtê-la, somam-se todos os dados de uma variável e divide-se pelo número de dados. Assim, teremos uma média para cada variável.
- <u>Mediana:</u> é o valor do meio de um conjunto de dados ordenado, o qual nem sempre é igual a média. Quando o número de dados é ímpar, a mediana é representada pelo número central do conjunto de dados, ou seja, exatamente 50% dos dados estão à sua esquerda (menores) e 50% à sua direita (maiores). Quando o número de dados é par, a mediana é calculada pela média dos dois valores do meio do conjunto.

* Quando existem valores excepcionalmente extremos no conjunto de dados (**outliers**), a mediana pode dar uma ideia melhor de um valor típico do que a média, visto que não é tão distorcida por esses valores.

<u>Moda:</u> é o valor que aparece com maior frequência em um conjunto de dados. Uma variável pode possuir uma única moda (unimodal), duas modas (bimodal), mais de duas modas (multimodal) ou nenhuma moda (amodal).

* A moda é útil quando os valores não são numéricos, e por isso a média e a mediana não podem ser definidas.

- <u>Variância</u>: é o valor que mostra o quão distante os valore reais do conjunto de dados está da média desse conjunto. Quanto menor a variância, mais próximos os valores estão da média; quanto maior a variância, mais distantes os valores estão da média.
- <u>Desvio padrão</u>: corresponde ao erro equivalente caso substituíssemos todos os valores reais de um conjunto de dados pela média desse conjunto.

*O desvio padrão é calculado pela raiz quadrada da variância. É mais fácil interpretar os dados utilizando-se o desvio padrão, porque a variância é um valor ao quadrado $(^2)$ e, portanto, não pode ser diretamente comparado aos valores reais do conjunto de dados.

INVESTIGAÇÃO CORRELACIONAL

Esse tipo de investigação consiste em descobrir se as variáveis que estão sendo estudadas possuem correlação (dependência) entre si. O índice de correlação é dado pelo coeficiente de correlação de Pearson (r), que pode variar de -1 à 1, conforme mostra a Figura 1.



Figura 1. Índice de correlação de Pearson (*r*)

Valores negativos indicam correlação negativa entre variáveis (-1 < r < 0), ou seja, quanto maior o valor de uma das variáveis, menor será o valor da segunda variável. Valores positivos indicam correlação positiva entre variáveis (0 < r < 1), indicando que quanto maior for o valor de uma das variáveis, maior o valor da outra variável. Quando a correlação for zero (r = 0), não há nenhuma correlação entre as variáveis. Assim, dizemos que a correlação entre as variáveis é nula. Índices de correlação entre |0,6| ou |0,7| e |1| indicam correlação FORTE entre as variáveis, e valores entre |0,01| e |0,59| ou |0,69| indicam correlação FRACA entre as variáveis (HAIR et al., 2009).

Normalmente, o resultado de uma correlação é apresentado em forma de uma matriz de correlação, o que possibilita que correlacionemos mais de duas variáveis de uma única vez. A correlação entre uma variável e ela mesma sempre será 1,0, de forma que, na matriz de correlação, a diagonal sempre será uma sequência de 1,0, conforme mostra a Figura 2.

Variáveis	1	2	3	4	5	6	7	8	9	10
1	1,000	0,231								
2	0,231	1,000								
3	0,089	0,231	1,000							
4	0,489	0,27	-0,066	1,000						
5	0,412	0,313	0,036	0,564	1,000					
6	0,228	-0,075	0,066	0,168	0,242	1 ,000				
7	0,326	0,008	0,206	0,336	0,377	0,671	1,000			
8	0,189	0,244	0,409	0,061	0,148	0,146	0,125	1,000		
9	-0,07	0,256	0,483	-0,077	0,045	-0,024	-0,072	0,404	1,000	
10	0,312	0,519	0,211	0,249	0,293	0,154	-0,033	0,397	0,281	1,000

Figura 2. Matriz de Correlação

Note que, neste caso, não há valores em cima da linha diagonal de 1,0. Isso acontece porque o índice de correlação entre a Variável 2 e a Variável 1 é o mesmo que já foi calculado para a Variável 1 x Variável 2, na primeira linha. Algumas vezes, dependendo do software utilizado ou por escolha do pesquisador, todas as casas estarão preenchidas, de forma que os valores se repetem quando as variáveis correlacionadas são as mesmas.

INVESTIGAÇÃO EXPERIMENTAL

Por fim, a investigação experimental procura relações causais e predições entre as variáveis, com o intuito de controlar o fenômeno que se deseja estudar. Inúmeros tipos de procedimentos podem ser utilizados nessa etapa de investigação. Um dos mais difundidos e que possui amplo poder de explicação de previsões, é a **Regressão**.

Quando realizamos a pesquisa teórica sobre o fenômeno que desejamos estudar, encontramos relações anteriores já estudadas e comprovadas por outros autores. A partir disso, temos uma base sólida para acreditar quais variáveis são explicadas e quais são as que explicam determinados fenômenos. À explicada, damos o nome de **variável dependente**, porque o valor assumido por esta depende da variação de outras variáveis. A(s) variável(is) explicativa(s) é(são) chamada(s) de **variável(is) independente(s)**, porque seu(s) valor(es) não se altera(m) quando o valor de outras variáveis muda.

A diferença principal entre a análise de regressão e a correlação, é que na segunda apenas sabemos que há uma associação entre as variáveis, mas não sabemos qual delas é a variável que depende da outra.

A Regressão é sempre Linear, porque supõe-se, previamente, que há correlação entre as variáveis que estão sendo analisadas, e isso, graficamente, é representado por uma "linha". Quando queremos analisar a relação de dependência entre duas variáveis, em que uma assume o papel de dependente e outra de independente, rodamos uma Regressão Linear Simples (RLS). Mas quando precisamos analisar a relação entre mais de duas variáveis, em que duas ou mais variáveis assumem o papel de independentes, precisamos calcular uma Regressão Linear Múltipla.

A Regressão Linear Múltipla (RLM)

Ressaltamos que a primeira coisa a se fazer, em qualquer estudo científico, é a pesquisa teórica. É a partir desta que poderemos desenvolver nossa ferramenta de coleta de dados, quantos casos serão necessários, quais dados (variáveis) precisamos coletar e qual a relação de causa e efeito entre essas variáveis. Mas esse não é o nosso foco aqui. Supõe-se que, nessa etapa da análise de dados, tudo isso já tenha sido feito. A partir daí, para entendermos a lógica da RLM, seguiremos seis passos para análise de dados:

- 1. Definição das variáveis;
- 2. Desenho do gráfico de dispersão;
- 3. Montagem da equação da RLM;
- 4. Rodar a RLM;
- 5. Substituir os dados na equação da RLM;

6. Interpretação dos resultados.

Esses passos serão utilizados ao longo de toda a apostila, para exemplificar o cálculo da RLM tanto no Excel, quanto no SPSS. Mas primeiro, cada um deles será explicado individualmente.

1. Definição das variáveis

A escolha das variáveis que serão coletadas parte, em primeiro lugar, da teoria. Após apreender o que outros autores da mesma área do conhecimento estão discutindo sobre o seu tema e definir quais variáveis serão analisadas na sua RLM, é preciso escolher qual delas será a variável dependente e quais serão as independentes.

Conforme Field (2009), para construir um modelo complexo com várias variáveis independentes, muito cuidado deve ser tomado ao selecionarmos tais variáveis, porque os valores dos **coeficientes de regressão** (\mathbb{R}^2) dependem delas.

Desta forma, as variáveis independentes incluídas e a forma com que elas são inseridas na RLM podem ter um grande impacto. Num mundo ideal, as variáveis independentes deveriam ser selecionadas baseadas em pesquisas anteriores. Não se deve de forma alguma selecionar centenas de variáveis independentes ao acaso, juntá-las todos em uma análise de regressão e torcer pelo melhor.

2. Desenho do gráfico de dispersão

Após a definição das variáveis, como forma de confirmar o pressuposto de que há correlação entre a variável dependente e cada uma das variáveis independentes, podemos gerar os gráficos de dispersão para cada uma das relações. Por exemplo, se foram escolhidas duas variáveis independentes, A e B, dois gráficos serão gerados, um para a relação entre a variável dependente e A e outro para a relação entre a variável dependente e B.

A interpretação se torna mais fácil quando se gera a linha de tendência sobre os pontos dispersos no gráfico, embora não seja necessário ainda se preocupar com a equação da reta, apenas com a lógica da imagem apresentada. A Figura 3 ilustra a lógica da interpretação de um gráfico de dispersão.



Há correlação entre a variável dependente e a variável independente. A linha de tendência pode aparecer mais ou menos inclinada, o que indica uma correlação forte ou fraca, respectivamente. O resultado apontado sugere que é importante manter a variável independente no modelo.

Figura 3. Interpretação de gráficos de dispersão

Não há correlação entre a variável dependente e a variável independente. Provavelmente o r da equação da reta será próximo de zero. O resultado apontado sugere que talvez seja necessário excluir a variável independente do modelo.

3. Montagem da equação da RLM

Confirmadas as correlações entre as variáveis estudadas, montaremos a equação que descreve a relação de dependência entre essas variáveis. A equação da RLM é a mesma equação da reta (Y = a + bx), com a diferença de que há múltiplas variáveis "b" que influenciam na inclinação da reta. A Figura 4 mostra como montar a equação e abaixo de cada termo estão seus respectivos significados.



Figura 4. Equação de RLM

Onde:

- Υ = valor previsto da variável dependente que será obtido por meio do modelo estimado;
- β₀ = representa a constante ou coeficiente linear; quando todos os χ forem iguais à 0, é o valor de β₀ que corresponde à Υ. A constante também nos mostra, no gráfico, qual o valor de Υ, quando χ for igual à 0, ou seja, qual o ponto em que a reta inicia no eixo Y do gráfico.
- β_n = é o coeficiente de cada variável independente ou coeficientes angulares. Esse valor indica quanto a variável dependente (Υ) vai variar com a variação de uma unidade de χ, quando todas as outras variáveis forem constantes. No gráfico, representam a inclinação da reta.
- $\chi_n = \acute{e}$ a **descrição** (nome) de cada variável independente;
- u = termo de erro ou resíduo, o qual equivale à diferença entre o valor real de Υ e o valor previsto de Υ. Quando menor o erro, melhor. Se o erro for muito alto, significa que outras variáveis, além das variáveis χ que foram incluídas na equação, afetam Υ;
- i = representa cada uma das variáveis da amostra (i = 1, 2, 3...*n*, em que *n* é o tamanho da amostra).

4. Rodar a RLM

Depois de montar a equação da RLM fica mais fácil identificar as relações entre as variáveis. O desenvolvimento do cálculo da RLM depende do software que se está utilizando. Por isso, esse passo será melhor explicado no exemplo posterior, que será aplicado tanto no Excel, quanto no SPSS.

De qualquer forma, podemos adiantar que os valores de Υ e dos χ sempre serão os nomes das variáveis, constituídos por palavras ou abreviações. Já os β sempre serão valores, e correspondem aos coeficientes das variáveis à que estão associados. O erro (u) também é sempre em formato numérico, e seu valor varia dependendo dos valores das outras variáveis.

5. Substituir os dados na equação da RLM

Após calcular os valores β de cada uma das variáveis, substituiremos esses valores na equação inicial. Mas antes, precisamos saber se as variáveis independentes realmente têm poder preditivo sob a variável dependente.

Para isso, primeiro, vamos falar da **significância**. O nível de significância sempre equivale à 1 menos (-) o intervalo de confiança que se está utilizando, ou seja, se quero ter 90% de confiança nos resultados, meu nível de confiança equivale à 1 - 0.9 = 0.1; se quero ter 95% de confiança nos resultados, meu nível de confiança equivale à 1 - 0.95 = 0.05; e se quero ter 99% de confiança nos resultados, meu nível de confiança equivale à 1 - 0.95 = 0.05; e se quero ter 99% de confiança nos resultados, meu nível de confiança equivale à 1 - 0.99 = 0.01. O intervalo de confiança mais utilizado é o de 95%, o que equivale à um nível de significância de 0.05. Mas o que isso significa? Vamos entender por meio da Figura 5.



Figura 5. Intervalo de confiança

Quando utilizamos o intervalo de confiança de 95%, consideramos que os resultados validos então dentro da área vermelha do gráfico. As "caudas" da curva, que juntas correspondem à 5% dos dados (ou 0,05), abrangem os valores que não fazem parte do intervalo de confiança que desejamos. Por isso, quando dizemos que um valor "é significativo à 95% de confiança" ou que o teste t (valor-P) é menor que 0,05, queremos dizer que ele está na parte vermelha do gráfico, portanto, é válido no modelo. Assim, as variáveis da RLM que apresentarem valores -P ou a significância do teste t menores que 0,05, são mantidas, e as que não apresentarem, são excluídas.

Mas afinal, o que são teste t, valor-P, teste z e teste f?

Quando coletamos dados em amostras grandes (mais do que 30, 50 ou 100 casos, dependendo do autor), a média da amostra tende a se comportar da mesma forma que a média de toda a população, visto que ela é representativa e pode ser generalizada. Nesses casos, levamos em conta as significâncias do **TESTE Z**, porque a distribuição é normal, em torno da média, ou seja, valores mais próximos da média tem mais probabilidade de aparecem do que valores mais distantes da média, como mostra a Figura 6.



Figura 6. Teste Z

O **Teste T** tem a mesma função do teste z, porém, quando as amostras são pequenas ou quando não se conhece o desvio padrão da população, uma vez que nesse caso não se pode extrapolar o desvio padrão da amostra para a população. O valor-P equivale ao nível de significância do teste t. Alguns softwares fornecem os valores de t e os respectivos valores-P e outros os valores de t e os níveis de significância de cada um, o que quer dizer a mesma coisa.

A Figura 7 ilustra a diferença entre Z e T. Como exemplo, desenhamos as curvas de duas amostras pequenas. Note que quanto menor a amostra, mais distante da média os dados podem estar, e por isso não se pode generalizar a média e o desvio padrão da amostra para a população.



Figura 7. Diferença entre teste z e teste t

Mas então, o que fazer se a amostra for pequena?

Quando não podemos verificar a normalidade dos dados em virtude de nossa amostra ser pequena demais, precisamos testar a normalidade de nossos dados, por meio dos **testes não-paramétricos de hipóteses**. Esse não é o nosso foco aqui, mas precisamos saber que eles existem para que, se necessário, possamos utilizá-los e saibamos pelo que procurar na literatura relacionada a estatística. Retiramos uma explicação básica sobre alguns testes do livro de Bruni (2012), o qual explica detalhadamente cada um deles.

a) **Teste de Kolmogorov-Smirnov:** analisa se os dados da amostra foram extraídos de uma população com uma distribuição peculiar de frequências, como a distribuição normal;

b) **Teste do qui-quadrado:** empregado na análise de frequências, quando uma característica da amostra é analisada;

c) **Teste do qui-quadrado para independência ou associação:** também empregado na análise de frequências, porém quando duas características da amostra são analisadas;

d) **Teste dos sinais:** empregado no estudo de dados emparelhados, quando um mesmo elemento é submetido a duas medidas;

e) **Teste de Wilcoxon:** também analisa dados emparelhados, permitindo, porém, uma consideração das magnitudes encontradas;

 f) Teste de Mann-Whitney: analisa se dois grupos originam-se de populações com médias diferentes;

g) **Teste da mediana:** analisa se dois grupos originam-se de populações com medianas diferentes;

h) Teste de Kruskal-Wallis: analisa se K (K > 2) grupos originam-se de populações com médias diferentes.

O **TESTE F**, por fim, testa a equação como um todo, e não fornece os valores de significância de cada variável. É esse valor que nos diz se nossa equação, como um todo, explica a nossa variável dependente.

Resumindo, os valores β que se mantém na equação da RLM são aqueles que possuem valores de significância menores que 0,05, quando se adota o intervalo de confiança de 95%.

6. Interpretação dos resultados

Finalmente, e talvez o passo mais importante, chega o momento da interpretação de todos esses resultados. Os números possuem diversas informações, mas muitas vezes seus significados estão obscuros. Assim, precisamos traduzi-los em forma de palavras para que outras pessoas vejam o que nós estamos vendo.

Para isso, desenvolveremos um exemplo, seguindo todo o passo a passo que foi apresentado, e interpretaremos os resultados gerados pelos softwares (Excel e SPSS).

* Variáveis Dummy

Mas, e quando as variáveis não são numéricas, mas sim são qualitativas?

Bom, nesse caso, precisaremos transformá-las em variáveis quantitativas, para que possamos realizar as análises estatísticas. Vamos entender a nomenclatura que utilizamos para cada tipo de variável através da Figura 8.

Variávais	Quantita	tiva	Qua	alitativa	
vallavels	Discreta Contínua	Nominal	Ordinal		
	São resultantes de	Podem assumir	sumir		
Descrição	contagens	qualquer valor	Não permitem	Permitem	
Descrição	representadas como	dentro de um	comparações	comparações	
	números inteiros	intervalo			
Evemplos	Nº de filhos	Peso e altura	Nome	Escolaridade e	
Exemptos	in de fillios		Nome	Likert	

Figura 8. Diferença entre variáveis quantitativas e qualitativas

Quando transformamos uma variável qualitativa em uma variável quantitativa (numérica), à ela damos o nome de variável dummy. Uma variável qualitativa com n

categorias gera *n*-1 variáveis *dummies*. Para atribuir valores numéricos às variáveis qualitativas sempre responderemos perguntas com "Sim" e "Não". Traduzindo: se a variável qualitativa tem 2 categorias, faremos uma pergunta, o que vai gerar uma *dummy*; se ela possuir 3 categorias, faremos 2 perguntas, gerando duas *dummies*. Uma forma de facilitar esse tipo de transformação é montar colunas diferentes para responder cada pergunta. Excepcionalmente, neste caso, daremos um exemplo a parte para facilitar.

Imaginemos que gostaríamos de analisar o quanto o grau de escolaridade influencia na renda mensal de uma amostra. Para isso, precisamos transformar o grau de escolaridade em uma variável quantitativa. As opções que oferecemos aos respondentes foram: ensino fundamental, ensino médio, ensino superior e pós-graduação. Como temos 4 categorias de escolaridade que se apresentam de forma qualitativa, precisaremos fazer 3 perguntas, criando assim 3 *dummies*. Quando a resposta à nossa pergunta for "Não", atribuiremos o número 0 ao χ , e quando a resposta for "Sim", atribuiremos o número 1 ao χ . Escolhemos uma variável (ensino fundamental), a qual não dará origem à nenhuma pergunta e, depois disso, montamos um quadro, conforme apresentado na Figura 9.

1. Tem ensino	2. Tem ensino	3. Tem pós-	
médio?	superior?	graduação?	Categoria
0	0	0	Ensino fundamental (E.F.)
1	0	0	Ensino médio (E.M.)
1	1	0	Ensino superior (E.S.)
1	1	1	Pós-Graduação (P.G.)

Figura 9. Elaboração de variáveis dummy

Portanto, o grau de escolaridade da amostra gera 3 variáveis *dummy* na equação da RLM. Vamos montar a equação para entendê-la melhor:

$$RENDA = \beta_0 + \beta_1 E.M. + \beta_2 E.S. + \beta_3 P.G.$$

Não precisamos de uma *dummy* para o ensino fundamental porque quando todos o χ forem iguais à 0, como mostra o quadro, o ensino fundamental equivale ao valor da constante. Vamos compreender melhor pela substituição da equação para cada categoria:

- i. Ensino fundamental: RENDA = $\beta_0 + \beta_1.0 + \beta_2.0 + \beta_3.0$ (a renda média será o valor da constante, pois todos os outros β são multiplicados por 0);
- ii. Ensino Médio: RENDA = $\beta_0 + \beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_3 \cdot 0$ (a renda média será o valor da constante + o valor de β_1);

- iii. Ensino Superior: RENDA = $\beta_0 + \beta_1 \cdot 1 + \beta_2 \cdot 1 + \beta_3 \cdot 0$ (a renda média será o valor da constante + o valor de β_1 + o valor de β_2);
- iv. **Pós Graduação:** RENDA = $\beta_0 + \beta_1 \cdot 1 + \beta_2 \cdot 1 + \beta_3 \cdot 1$ (a renda média será o valor da constante + o valor de β_1 + o valor de β_2 + + o valor de β_3).

Assim, se o grau de escolaridade for um preditor positivo da renda, ou seja, quanto maior o grau de escolaridade, maior a renda, todos os valores de β serão positivos, e à medida que se somam, a renda aumenta. Os exemplos a seguir terão seções específicas para explicar melhor essa transformação de variáveis qualitativas em *dummies* e como interpretá-las.

Exemplo - Exercício Prático

Com o objetivo de ilustrar a RLM, será apresentado um exemplo prático, no qual a regressão será realizada por meio do Excel e do SPSS. No exemplo apresentado na Tabela 1, observa-se um conjunto de dados de uma amostra formada por 36 filmes exibidos nos cinemas. Inicialmente, apresentam-se quatro variáveis: **código** (representando o título do filme), **faturamento** com o filme em milhões, **gasto** com o filme em milhões e **duração** do filme em minutos. Neste exemplo, busca-se identificar a relação existente entre as variáveis independentes (gasto e duração) e a variável dependente (faturamento). Logo depois, seguimos os passos, como apresentado na Introdução desta apostila.

Tabel	la 1. Amostra	de dados – filmes exibi	dos no cinema						
N°	CÓDIGO	FATURAMENTO (em milhões R\$)	GASTO (em milhões R\$)	DURAÇÃO (em minutos)	N°	CÓDIGO	FATURAMENTO (em milhões R\$)	GASTO (em milhões R\$)	DURAÇÃO (em minutos)
1	A1	81,843	18,500	137	19	I3	197,171	39,000	127
2	A2	194,125	140,000	144	20	I4	260,000	12,000	124
3	A3	147,540	50,000	138	21	M1	250,147	90,000	98
4	B1	75,600	72,000	177	22	M2	20,100	45,000	117
5	C1	12,006	0,300	105	23	P1	107,930	8,000	154
6	C2	100,853	90,000	153	24	R1	242,374	20,000	115
7	D1	67,155	104,000	112	25	S 1	178,091	70,000	170
8	D2	140,424	75,000	120	26	S2	96,067	25,000	197
9	E1	68,750	55,000	129	27	S3	103,001	15,000	111
10	F1	329,691	55,000	142	28	S4	48,068	110,000	121
11	G1	217,631	22,000	128	29	T1	36,900	6,400	108
12	G2	198,571	3,900	222	30	T2	65,000	62,000	114
13	G3	138,339	10,000	126	31	Т3	63,540	90,000	126
14	G4	181,280	6,000	110	32	T4	48,265	50,000	128
15	H1	47,000	0,300	93	33	Т5	56,876	35,000	132
16	H2	19,819	70,000	95	34	Т6	600,743	200,000	195
17	I1	72,219	17,000	100	35	Τ7	146,261	100,000	144
18	I2	306,124	75,000	142	36	V1	47,474	90,000	102

Utilizando o EXCEL

Inicialmente, iremos realizar cada um dos passos apresentados anteriormente utilizando a ferramenta Excel. Para isso, os dados apresentados na Tabela 1 devem ser transcritos para o Excel, conforme a Figura 10.

Arq	uivo Página Inicial Inserir Layou	ut da Página Fórmu	las Dados Rev	isāo Exibir
Área	$\begin{array}{c c} & & \\ & &$	$\begin{array}{c} \mathbf{A}^{\star} \ \mathbf{A}^{\star} \end{array} \equiv \equiv \equiv \\ \mathbf{A}^{\star} \ \mathbf{A}^{\star} \end{array} \equiv \equiv \equiv \parallel$	 ♥ - E[*] Quebrar 	Texto Automati e Centralizar 🔹
114	4 ▼ : × ✓ f _x	W-=11		
1	A	В	С	D
1	TÍTULO	FATURAMENTO	GASTO MILHÕES	DURAÇÃO
2	A1	81,843	18,5	137
3	A2	194,125	140	144
4	A3	147,54	50	138
5	B1	75,6	72	177
6	C1	12,006	0,3	105
7	C2	100,853	90	153
8	D1	67,155	104	112
9	D2	140,424	75	120
10	E1	68,75	55	129
11	F1	329,691	55	142
12	G1	217,631	22	128
13	G2	198,571	3,9	222
14	G3	138,339	10	126
15	G4	181,28	6	110
16	H1	47	0,3	93
17	H2	19,819	70	95
18	11	72,219	17	100
19	12	306,124	75	142
20	13	197,171	39	127
21	14	260	12	124

Figura 10. Dados no Excel

Passo 1: Definição das Variáveis

No exemplo citado, busca-se saber a influência que as variáveis gasto e duração têm sob o faturamento dos filmes exibidos no cinema. Assim, a variável dependente é o faturamento, enquanto as variáveis independentes são o gasto e a duração.

A fim de facilitar a regressão, sugere-se que as variáveis independentes estejam à direita da variável dependente. O código (título do filme) não é considerado uma variável, pois está sendo apresentado apenas com a intenção de ilustrar quais filmes foram avaliados nesta amostra. A Figura 11 apresenta a definição das variáveis no Excel.

Arq	uivo	Págir	ia Inicial	Inserir	Layout	da Págin	a Fórmu	ılas i	Dados	Revisã	o Exibir
0	Colar	¥ ≌ • ∛	Calibr	i • ≤ • ⊞	11 - / - 🙆 -	A A		%⁄- €≣ ₹≣	e Que	brar Te clar e C	do Automati entralizar 👻
Área	de Tra	nsf 🛱		Fonte		15			Alinhame	nto	
114	1	*	1 : 1	x	fx			VARI	ÁVEL INF	FPEN	DENTE
	а.				KG			VENN			1
1				A			В		C V		D
1			TÍT	TULO		FATU	RAMENTO	GAST	O MILHÕ	DES D	URAÇÃO
2	A1				-	8	1,843		18,5		137
3	A2		VARIA	AVEL DEPEN	DENTE	1	94,125	2	140		144
4	A3					1	.47,54		50		138
5	B1						75,6	5	72		177
6	C1					1	2,006		0,3		105
7	C2					1	00,853	2	90		153
8	D1					6	7,155		104		112
9	D2				-	14	40,424	×	75		120
10	E1					1	68,75	29	55		129
11	F1					3	29,691		55		142
12	G1					2	17,631		22		128
13	G2					1	98,571	1	3,9	2	222
14	G3					1	38,339	6) (5)	10		126
15	G4					1	81,28		6		110
16	H1						47		0,3		93
17	H2					1	9,819		70	2	95
18	11					7	2,219	0 	17	Č.	100
19	12					3	06,124		75		142
20	13					1	97,171	-	39		127
21	14				1		260	2 7	12		124
	4 F]	RLM 2	2 RLM 3	MATR	IZ DE CO	ORRELAÇÃO	M	DDELO 3	M	DELO 4

Figura 11. Definição das variáveis no Excel

Passo 2: Desenho do gráfico de dispersão

O segundo passo consiste na criação do gráfico de dispersão. Ressalta-se a importância de gerar o gráfico de dispersão individualmente, relacionando a variável dependente com cada uma das variáveis independentes, visto que, por meio deste gráfico, busca-se verificar se existe relação entre duas variáveis, assim como qual a intensidade desta relação. Assim, serão gerados dois gráficos de dispersão: um que relaciona faturamento (dependente) e gasto (independente) e outro, faturamento (dependente) e duração (independente). Os passos para gerar o primeiro gráfico serão os mesmos utilizados para o segundo gráfico, o que muda é apenas a variável independente.

Para gerar o gráfico de dispersão no Excel, deve-se clicar em "inserir" (1), selecionar as colunas que serão relacionadas (2) e gerar gráfico de dispersão (3), conforme apresentado na Figura 12 e na Figura 13.

Are	quivo	Página Inicial	Inserir	Lay	yout da Página	Fórmulas Da	dos Re	visão	Exibir	Q O que você deseja fazer					
Ta Din	abela âmica	Tabelas Dinâmicas Recomendadas Tabelas	Tabela Ir	nager	ns Imagens Online 🖦 *	🗎 Loja 🎝 Meus Suplem Suplemento	entos *	Gráf Recome	icos ndados Ø	Gráfico Binàmico + Gráfico + Binàmico + Grófico + Gráfico + Binamico + Gráfico + Binamico + Gráfico + Binamico + Gráfico + Binamico + Binamico + Gráfico + Binamico + Binamico + Gráfico + Binamico + Gráfico + Binamico + Gráfico + Binamico + Binamico + Gráfico + Binamico + Binamico + Binamico + Binamico + Gráfico + Binamico + Bina					
19			2	fx						Inserir Gráfico de Dispersão (X, Y) ou de Bolha					
		A		2	В	с	D		E	Use este tipo de gráfico para mostrar o relacionamento entre conjuntos de valores.					
1		TÍTULO	0)		FATURAMENTO	GASTO MILHÕ	ES DURA	ÇÃO		Clique na seta para ver os diferentes tipos de					
2	A1				81,843	18,5	13	7		gráficos de dispersão e de bolhas disponíveis e					
3	A2				194,125	140	14	4		pause o ponteiro nos icones para ter uma					
4	A3				147,54	50	13	8		visualização em seu documento.					
5	B1				75,6	72	17	7							
6	C1				12,006	0,3	10	5							
7	C2				100,853	90	15	3							
8	D1				67,155	104	11	2							
9	D2				140,424	75	12	0							
10	E1				68,75	55	12	9							
11	F1				329,691	55	14	2							
12	G1				217,631	22	12	8							
13	G2				198,571	3,9	22	2							
14	G3				138,339	10	12	5							
15	G4				181,28	6	11	0							
16	H1				47	0,3	93	1							
17	H2				19,819	70	95	i							
18	11				72,219	17	10	D							
19	12				306,124	75	14	2							
20	13				197,171	39	12	7							
21	14				260	12	12	4							
22	M1				250,147	90	98)							
23	M2				20,1	45	11	7							

Figura 12. Montagem do gráfico de dispersão de Faturamento X Gasto

Ar	quivo Página Inicial Inserir La	ayout da Página	Fórmulas Dados	s Revisão Exib	bir 🛛 Q O que você deseja fazer	
Ti Dir	abela Tabelas Dinâmicas Tabela nâmica Recomendadas Tabelas	ens Imagens Online 💁 🎽 Ilustrações	 Loja Meus Suplement Suplementos 	os - Gráficos Recomendado	s di v la v la v Gráfi Dispersão Minigráficos	()
G	iráfico 1 🔻 🗄 🗙 🏑 🎼				P. 9 18- 9 14 -1	
- 1	A	В	С	D E	J K	ě.
1	ΤΪΤυιο	FATURAMENTO	GASTO MILHÕES	DURAÇÃO	Dispersão	
2	A1	81,843	18,5	137	Use este tino de gráfico para:	
3	A2	194,125	140	144	Bolhas • Comparar pelo menos dois	
4	A3	147,54	50	138	conjuntos de valores ou pares	
5	B1	75,6	<u>72 </u>	177		
6	<u>C1</u>	12,006	0,3		entre conjuntos de valores	
7	C2	100,853	90	250	🐼 <u>Mai</u>	
8	DI	67,155	104		Os dados representarem	
9	D2	140,424	75	200	medidas separadas.	
10	El	68,75	55	200		
11	F1	329,691	55	150		
12	G1	217,631	22	150	•	
13	G2	198,571	3,9	100		
14	G3	138,339	10	100 00 0		
15	G4	181,28	0.2	F0		
10	HI	4/	0,3	50	•	
1/	H2	72 210	17		· · · · · · · · · · · · · · · · · · ·	
10	12	72,215		0 100	200 300 400 500 600 700	
20	12	197 171	20 0	127		
20	13	250	12	12/		
21	M1	250 147	90	99		
22	M2	201	45	117		
23	12 0 (. 1 1	~ 1 E				-

Figura 13. Gráfico de dispersão de Faturamento X Gasto

Para gerar a **linha de tendência** deve-se selecionar o gráfico, clicar no sinal "+" e assinalar o item "linha de tendência". Com esta linha será possível observar qual a tendência de comportamento entre as variáveis analisadas, conforme a Figura 14.



Figura 14. Linha de tendência da relação entre Faturamento X Gasto

Arquivo Página Inicial Inserir	Layout da Página	Fórmulas Dados	Revisão	Exib	ir Des	ign Form	atar 🛛	O que voi	:é deseja fazer					Entrar	A Comparti
Adicionar Elemento Layout Gráfico * Rápido * Layout de Gráfico	the line		Estilo	s de Gráfi			dis.	ĪЩ	in .	 Alter Linha/0 	nar Sele Coluna D Dados	cionar Alt ados de	erar Tipo Gráfico Tipo	Mover Gráfico Local	
Gráfico 2 * 1 × ✓ fa	2														
A .	В	c	D	F	F	G	н	1 I	i a i	ĸ	E.	м	N	0	р
11 F1	329 691	55	142		100	••				and the second s			7.50		
12 61	217,631	22	128					and a start							
13 G2	198,571	3.9	222		50										
14 G3	138,339	10	126				•								
15 G4	181.28	6	110					•							
16 H1	47	0.3	93		0	100	200	300	400 5	500 60	0 70	10			
17 H2	19,819	70	95												
18 11	72,219	17	100	(<u> </u>			0	- 11 -				ELEME	NTOS DO GI	RÁFICO
19 12	306,124	75	142					DURA	ÇAO) 🔽 Ei	xos	
20 13	197.171	39	127		250							1		tulos dos Eix	os
21 14	260	12	124										I TI	tulo do Gráfi	co
22 M1	250,147	90	98		200							7		ótulos de Da	dos
23 M2	20,1	45	117										B	arras de Erros	
24 P1	107,93	8	154		150								I Li	nhas de Grac	le
25 R1	242,374	20	115	(¢ .	5 m.						¢.	Le	egenda	
26 51	178,091	70	170		100		• •						🔽 Li	nha de Tend	ência
27 52	96,067	25	197			•									
28 53	103,001	15	111		50										
29 54	48,068	110	121		1.000										
30 T1	36,9	6,4	108		0										Título do (
31 T2	65	62	114		0	100	200	300	400 5	00 60	0 70	D			
32 T3	63,54	90	126	(6	1.2460		0		1963 (Shiji	a 1989	0			
33 T4	48,265	50	128												

Agora, faremos o mesmo com a variável duração, conforme Figura 15:

Figura 15. Linha de tendência da relação entre Faturamento X Duração

Observa-se que, por meio do gráfico de dispersão, é possível verificar como os dois conjuntos de dados comparáveis concordam entre si. Quanto mais os conjuntos de dados concordarem, mais os pontos dispersos tendem a se concentrar ao redor (próximo) da linha.

Passo 3: Montagem da equação da RLM

A Figura 16 apresenta a montagem da equação da RLM, conforme o exemplo prático apresentado.



Figura 16. Equação da RLM - Exemplo Prático

Passo 4: Rodar a RLM

O **quarto passo** consiste em gerar a RLM no Excel. Para isso, é necessário utilizar a ferramenta "Análise de Dados", conforme apresentado na Figura 17.

A B C D E F G H J K L P18 I Truco Gasto Millhöes 0.3 105 12 12 12 12 12 12 13 14 <t< th=""><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th>8</th><th>a fazer</th><th>ie você dese</th><th>Ωoq</th><th>Exibir</th><th>Revisão</th><th>Dados</th><th>Fórmulas</th><th>yout da Página</th><th>Inserir Lay</th><th>Página Inicia</th><th>ļuīvo</th><th>Ārq</th></t<>								8	a fazer	ie você dese	Ωoq	Exibir	Revisão	Dados	Fórmulas	yout da Página	Inserir Lay	Página Inicia	ļuīvo	Ārq
A B C DDL P Final Prenamenta de Dados Prenamenta de	Análise de Dados Solver 2	este de Planilha de pôtese + Previsão Previsão Estrutura de Tópicos 12 Anális			Teste de Hipótese	₽ 00 • ()	sto para	par plicar nçado	The Live Re	ificar Filtro	1 Ž↓ ZA Z↓ Class	onexões 1 ropriedades ditar Links	s Atualizar Tudo + G	Mostrar Consulta Da Tabela Fontes Recentes	Nova Consulta +	er Dados ernos *	Obte			
P18 I A B C D E F G H I J K L 1 TÍTULO FATURAMENTO GASTO MILHÕES DURAÇÃO Image: Construction of the structure of	Andrise	A	COS 18	a de topic	Estrutur	40	Previsa	P	Dados	riamentas di	1	riiular	Classificarie		oes	Conex	r e transformar	Obte		
A B C D E F G H I J K L 1 TTULO FATURAMENTO GASTO MILHÕES DURAÇÃO Image: Construction of the co																	$\times \checkmark f_x$	* E	18	P1
1 TÍTULO FATURAMENTO GASTO MILHÕES DURAÇÃO 2 A1 81,843 18,5 137 3 A2 194,125 140 144 4 A3 147,54 50 138 5 B1 75,6 72 177 6 C1 12,006 0,3 105 7 C2 100,853 90 153 8 D1 67,155 104 112 9 D2 140,424 75 120 11 FI 329,691 55 142 12 G1 217,631 22 128 13 G2 198,571 3,9 222 14 G3 138,339 10 126 15 G4 181,28 6 110 17 H2 19,819 70 95 18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171	M N	м	1	e	к	J	Ê.	1	н	G	F	Е	D	5	С	В		А		1
2 A1 81,843 18,5 137 3 A2 134,125 140 144 4 A3 147,54 50 138 5 B1 75,6 72 177 6 C1 12,006 0,3 105 7 C2 100,853 90 153 9 D2 140,424 75 120 10 E1 68,75 55 129 11 F1 329,691 55 142 12 G1 217,631 22 128 13 G2 198,571 3,9 222 14 G3 138,339 10 126 15 G4 1181,28 6 110 17 H2 19,819 70 95 18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171 39 127													URAÇÃO	VILHÕES D	GASTO M	FATURAMENTO	LO	τίτι		1
3 A2 194,125 140 144 4 A3 147,54 50 138 5 B1 75,6 72 177 6 C1 12,006 0,3 105 7 C2 100,853 90 153 8 D1 67,155 104 112 9 D2 140,424 75 120 10 E1 68,75 55 129 11 F1 329,691 55 142 12 G1 217,631 22 128 13 G2 198,571 3,9 222 14 G3 138,339 10 126 15 G4 181,28 6 110 17 H2 19,819 70 95 18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171 39 127													137	,5	18,	81,843			A1	2
4 A3 147,54 50 138 5 B1 75,6 72 177 6 C1 12,006 0,3 105 7 C2 100,853 90 153 8 D1 67,155 104 112 9 D2 140,424 75 120 10 E1 68,75 55 129 11 F1 329,691 55 142 13 G2 138,339 10 126 14 G3 138,339 10 126 15 G4 181,28 6 110 17 H2 19,819 70 95 18 I1 72,219 17 100 20 300 400 500 600 700 19 I2 306,124 75 142 20 300 400 500 600 700 19 I2 306,124 75 142 20 300 400 500 600 700 <td></td> <td></td> <td></td> <td></td> <td></td> <td>UÕEC</td> <td>5.411</td> <td>CACTO</td> <td>-</td> <td></td> <td></td> <td></td> <td>144</td> <td>10</td> <td>140</td> <td>194,125</td> <td></td> <td></td> <td>A2</td> <td>3</td>						UÕEC	5.411	CACTO	-				144	10	140	194,125			A2	3
5 B1 75,6 72 177 6 C1 12,006 0,3 105 7 C2 100,853 90 153 8 D1 67,155 104 112 9 D2 140,424 75 120 10 E1 68,75 55 129 11 F1 329,691 55 142 12 G1 217,631 22 128 13 G2 138,571 3,9 222 16 H1 47 0,3 93 16 H1 47 0,3 93 17 H2 19,819 70 95 18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171 39 127						LUCES	IVIIL	JASIU	C				138	0	50	147,54			A3	4
6 C1 12,006 0,3 105 7 C2 100,853 90 153 8 D1 67,155 104 112 9 D2 140,424 75 120 10 E1 68,75 55 129 11 F1 329,691 55 142 12 G1 217,611 22 128 13 G2 198,571 3,9 222 14 G3 138,339 10 126 15 G4 1181,28 6 110 17 H2 19,819 70 95 18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171 39 127											250		177	2	72	75,6			B1	5
7 C2 100,853 90 153 8 D1 67,155 104 112 9 D2 140,424 75 120 10 E1 68,75 55 129 11 F1 329,691 55 142 12 G1 217,631 22 128 13 G2 198,571 3,9 222 14 G3 138,339 10 126 15 G4 181,28 6 110 17 H2 19,819 70 95 18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171 39 127													105	,3	0,3	12,006			C1	6
8 D1 67,155 104 112 9 D2 140,424 75 120 10 E1 68,75 55 129 11 F1 329,691 55 142 12 G1 217,631 22 128 13 G2 138,339 10 126 16 H1 47 0,3 93 16 H1 47 0,3 93 18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171 39 127				•							200		153	0	90	100,853			C2	7
9 D2 140,424 75 120 10 E1 68,75 55 129 11 F1 329,691 55 142 12 G1 217,631 22 128 13 G2 198,571 3,9 222 14 G3 138,339 10 126 16 H1 47 0,3 93 17 H2 19,819 70 95 18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171 39 127													112)4	104	67,155			D1	8
10 E1 68,75 55 129 11 F1 329,691 55 142 12 G1 217,631 22 128 13 G2 198,571 3,9 222 14 G3 138,339 10 126 15 G4 181,28 6 110 17 H2 19,819 70 95 18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171 39 127											150		120	5	75	140,424			D2	9
11 F1 329,691 55 142 12 G1 217,631 22 128 13 G2 198,571 3,9 222 14 G3 138,339 10 126 15 G4 181,28 6 110 16 H1 47 0,3 93 17 H2 19,819 70 95 18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171 39 127									1				129	5	55	68,75			E1	10
12 G1 217,631 22 128 13 G2 138,571 3,9 222 14 G3 138,339 10 126 15 G4 131,28 6 110 16 H1 47 0,3 93 17 H2 19,819 70 95 18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171 39 127					and the second second					• •	100		142	5	55	329,691			F1	11
13 G2 198,571 3,9 222 14 G3 138,339 10 126 15 G4 181,28 6 110 16 H1 47 0,3 93 17 H2 19,819 70 95 18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171 39 127									•				128	2	22	217,631			G1	12
14 G3 138,339 10 126 15 G4 181,28 6 110 16 H1 47 0,3 93 17 H2 19,819 70 95 18 11 72,219 17 100 19 I2 306,124 75 142 20 I3 197,171 39 127							•			8	50		222	,9	3,9	198,571			G2	13
15 64 181,28 6 110 16 H1 47 0,3 93 17 H2 19,819 70 95 18 I1 72,219 17 100 19 I2 306,124 75 142 20 I3 197,171 39 127									1.				126	0	10	138,339			G3	14
16 H1 47 0,3 93 0 100 200 300 400 500 600 700 17 H2 19,819 70 95<								•		~ \$.	0 . •		110	i	6	181,28			G4	15
17 H2 19,819 70 95 18 I1 72,219 17 100 19 I2 306,124 75 142 20 I3 197,171 39 127		00	70	600	500	400		300	200	100	0		93	,3	0,3	47			H1	16
18 11 72,219 17 100 19 12 306,124 75 142 20 13 197,171 39 127													95	0	70	19,819			H2	17
19 12 306,124 75 142 DURAÇAO 20 13 197,171 39 127 250						i.o.		DUD					100	7	17	72,219			11	18
20 13 197,171 39 127 250						40	KAÇA	DUR					142	5	75	306,124			12	19
			250					127	9	39	197,171			13	20					
21 14 260 12 124									•				124	2	12	260			14	21
22 M1 250,147 90 98 200											:00		98	0	90	250,147			M1	22
23 M2 20,1 45 117										•			117	5	45	20,1			M2	23

Figura 17. Rodando a RLM

Caso essa ferramenta não esteja habilitada em seu Excel, é possível habilitá-la seguindo os seguintes passos: Arquivo > Opções > Suplementos > Suplementos do Excel > Ir > Ferramentas de Análise > Selecionar Análise de Dados > OK. Para gerar a RLM é necessário seguir os seguintes passos: Dados > Análise de Dados > Regressão > OK, conforme Figura 18.

Arquivo	Página Inicial Inserir Li	ayout da Página	Fórmulas Dado	s Revisão	Exibir	Qod	que vocé	deseja faz	cer	ii							
Obter Dados Externos *	Nova Consulta + Consult	Atualizar	onexões 1 ropriedades ditar Links 2↓	lassificar Filtr	 K Limp Reap Avar 	par blicar nçado	Texto pa Coluna			Teste de Hipóteses +	Planilha de Previsão	Agrup Desag	ar * rupar * tal	다. (1) (1) (1) (1) (1) (1) (1) (1) (1) (1)	Anális Solver	e de Dados 2]
	Obter e Transformar	Conex	ões	Classificar	e Filtrar	F	erramen	tas de Dad	los	Previ	são	Estrutura	de Tópicos	5	An	álise	
P18	▼ : × √ fx																
1	A	В	с	D	E	F	0	i	н	1	J	ĸ	1	8 H	м	N	
1	TÍTULO	FATURAMENTO	GASTO MILHÕES	DURAÇÃO					_								_
2 A1		81,843	18,5	137													
3 A2		194,125	140	144							LUÕEC						
4 A3		147,54	50	138					Ľ	JASTU IVII	LHUES						
5 B1		75,6	72	177	2	50			-	Análise de da	dos					? ×	\$0.
6 C1		12,006	0,3	105						<u>Ferramentas</u>	de análise					OK	
7 C2		100,853	90	153	2	00				Teste-F: dua	s amostras p	oara variânci	as		^	On	
8 D1		67,155	104	112						Histograma	ourier					Cancelar	
9 D2		140,424	75	120	1	50				Média móve					1	Aiuda	
10 E1		68,75	55	129				1		Ordem e per	centil	itorio					-
11 F1		329,691	55	142	1	.00	••		3	Regressão							
12 G1		217,631	22	128				• •		Teste-T: dua:	s amostras e	m par para r	nédias				
13 G2		198,571	3,9	222		50	. 8	and Section of		Teste-T: dua:	s amostras p	resumindo	variâncias	equivalen	ti Y		
14 G3		138,339	10	126				•									_
15 G4		181,28	6	110		0 . 9	. " 3			•							
16 H1		47	0,3	93		0	100	20	0	300	400	500	600	700			
17 H2		19,819	70	95													
18 11		72,219	17	100						DUDAG	20						
19 12		306,124	75	142						DURAÇ	AO						
20 13		197,171	39	127	25	50											
21 14		260	12	124													
22 M1		250,147	90	98	20	00											
23 M2		20,1	45	117			•										
D .	10 D 1 1 I		1														

Figura 18. Rodando a RLM no Excel

No "intervalo Y de entrada" deve-se selecionar toda a coluna da variável dependente; no "intervalo X de entrada" deve-se selecionar todas as colunas das variáveis independentes. É importante selecionar o campo "rótulos" para que os nomes de cada coluna estejam visíveis posteriormente. Do mesmo modo, é importante selecionar o campo "nível de confiança", visto que o pesquisador pode alterar o intervalo de confiança, se desejar, conforme Figura 19. O Excel, automaticamente, gera o intervalo padrão de confiança de 95% e assim, se o pesquisador desejar, pode inserir um novo intervalo para comparação. Quanto menores os níveis de confiança, mais estreitos serão os intervalos para conter um determinado parâmetro. Por outro lado, quanto maiores forem os níveis de confiança, maior amplitude terão os intervalos para conter este parâmetro.

An	uwo Página Inicial Inserir La	yout da Página	Fórmulas Dado	s Revisão	o Exibir Q	O que você d	eseja fazer	
Obte Exte	r Dados rrmos +	s D C Atualizar Tudo - D E	onexões opriedades litar Links	lassificar F	iltro	Texto para Colunas	Image: Second	Análise de Dado
	Obter e iransformar	Conex	Des	Classific	ar e Fiitrar	Perramenta	s de Dados Previsão Estrutura de iópicos	Analise
P1	\bullet							
	А	В	с	D	E F	G	Regressão	? × N
1	τίτυιο	FATURAMENTO	GASTO MILHÕES	DURAÇÃO			-	
2	A1	81,843	18,5	137			Intenzio V de entrada. EDE1-EDE37	ОК
3	A2	194,125	140	144	VARIAVEL DE	PENDENTE	SD\$1:SD\$57	Cancelar
4	A3	147,54	50	138	VARIÁVEL INDE	EPENDENTE	Intervalo 🛛 de entrada: SC\$1:SD\$37	
5	B1	75,6	72	177	250	-		Ajuda
6	C1	12,006	0,3	105			Nivel de confianca 95 %	
7	C2	100,853	90	153	200			
8	D1	67,155	104	112			Opções de saída	
9	D2	140,424	75	120	150		🔿 įntervalo de saída: 🛛 🚯	
10	E1	68,75	55	129			Nova planilha:	
11	F1	329,691	55	142	100	**	🔿 Nova pasta de trabalho	
12	G1	217,631	22	128			Resíduos	
13	G2	198,571	3,9	222	50		🗌 Resí <u>d</u> uos 📃 P <u>l</u> otar resíduos	
14	G3	138,339	10	126			Resíduos padronizados Plotar ajuste de linha	
15	G4	181,28	6	110	0		Probabilidade normal	
16	H1	47	0,3	93	0	100	Plotagem de probabilidade normal	
17	H2	19,819	70	95				
18	11	72,219	17	100		1	DURAÇÃO	
19	12	306,124	75	142			DURAÇAU	
20	13	197,171	39	127	250			
21	14	260	12	124			•	
22	M1	250,147	90	98	200			
23	M2	20,1	45	117		•	•	

Figura 19. Selecionando as variáveis da regressão

Após, o Excel irá gerar uma nova aba na planilha que está sendo utilizada, conforme a Figura 20.

Are	quivo	Página Inicia	l Inse	erir Laye	out da Página	Fórmi	ulas Dados	Revisão	Exibir Q	O que você deseja faz				
Obt Ext	er Dados ternos *	Nova Consulta +	Mostra	ar Consultas vela Recentes	Atualizar Tudo +	Conexõe Propriec Editar Li	es ⊉↓ dades nks Z↓ Cla	ssificar Filtro	K Limpar Reaplicar Avançado	Texto para Colunas	Teste de Pla Hipóteses - P	nilha de revisão	rupar • +]] sagrupar • -]] ototal	Análise de D.
120				ormar	Con	exoes		Classificar e r	littat	renamentas de Dad	DS Previsad	Estrutu	ra de lopicos 1a	Andrise
N	18		× ×	Jx										
- 24		Α			В		с	D	E	F	G	н	I.	0
1	RESUM	O DOS RESU	LTADOS											
2														
3		Es	tatística	de regress	ão									
4	R múlti	plo			0,5284	42429								
5	R-Quad	Irado			0,2792	51401								
6	R-quad	rado ajusta	do		0,2355	69668								
7	Erro pa	drão			101,13	32854								
8	Observ	ações		1.		36								
9														
10	ANOVA	4		1										
11					gl		SQ	MQ	F	F de significação				
12	Regres	são				2	130771,6866	65385,84329	6,392864477	0,004502283				
13	Resídu	D				33	337522,0664	10227,94141						
14	Total			1		35	468293,753							
15	1													
16				(Coeficientes		Erro padrão	Stat t	valor-P	95% inferiores	95% superiores	Inferior 95,0%	Superior 95,0%	6
17	Interse	ção			-88,606	67471	77,71696302	-1,140120139	0,262443891	-246,7230248	69,50967543	-246,7230248	69,5096754	3
18	GASTO	MILHÕES			0,8417	89202	0,390925457	2,153324086	0,038696133	0,04644538	1,637133025	0,04644538	1,63713302	5
19	DURAÇ	ÃO		1	1,3732	09971	0,585938149	2,34360909	0,025272059	0,181109844	2,565310098	0,181109844	2,56531009	8
20														
21														
22														
23	RESULT	ADOS DE RE	SÍDUOS											
	4. (F	Plani	lha2	FILMES	+									

Figura 20. Planilha de resultados da RLM

Abaixo, cada um dos dados apresentados na Figura 20 serão apresentados separadamente, por meio da Figura 21, Figura 22 e Figura 23.

• ESTATÍSTICA DE REGRESSÃO:

Estatística de re	gressão
R múltiplo	0,528442429
R-Quadrado	0,279251401
R-quadrado ajustado	0,235569668
Erro padrão	101,1332854
Observações	36

Figura 21. Estatística de regressão

O **R múltiplo** é utilizado quando existem vários previsores. /ele representa a correlação entre os valores de Y observados e previstos pelo modelo de regressão múltipla. Valores grandes de R múltiplo (mais próximos de 1) representam alta correlação entre os valores previstos e observados da variável dependente.

O \mathbb{R}^2 explica se a relação entre as variáveis é forte ou fraca. Quanto mais perto de 1 for o resultado, mais forte será a relação. Ressalta-se que o \mathbb{R}^2 relaciona todas as variáveis com a variável dependente. Se o \mathbb{R}^2 for igual a 1, o que dificilmente ocorrerá, não haverá resíduos para cada uma das observações da amostra em estudo e a variabilidade da variável Y estará totalmente explicada pelo vetor de variáveis X consideradas no modelo de regressão.

Quando há o intuito de comparar o coeficiente de ajuste (R^2) entre dois modelos ou entre um mesmo modelo com tamanhos de amostras diferentes, faz-se necessário o uso do R^2 **ajustado**, o qual é uma medida do R^2 da regressão estimada pelo método de mínimos quadrados ordinários ajustada pelo número de graus de liberdade, uma vez que a estimativa amostral de R^2 tende a superestimar o parâmetro populacional.

O **termo de erro ou resíduo** equivale à diferença entre o valor real de Y e o valor previsto de Y, visto que por meio de "u" é possível capturar o efeito das demais variáveis não incluídas no modelo de regressão utilizado.

O número de **observações** equivale ao número de casos analisados. Neste caso, foram 36 filmes.

ANOVA					
	gl	SQ	MQ	F	F de significação
Regressão	2	130771,6866	65385,84329	6,392864477	0,004502283
Resíduo	33	337522,0664	10227,94141		
Total	35	468293,753			

• ANOVA

Figura 22. Resultados ANOVA

Os **graus de liberdade (gl ou df)** representam a quantidade de informação, fornecida pelos dados, que você pode "gastar" para estimar os valores de parâmetros populacionais desconhecidos, e calcular a variabilidade dessas estimativas. Esse valor é determinado pelo número de observações em sua amostra e o número de parâmetros em seu modelo. Aumentar seu tamanho amostral fornece mais informações sobre a população e, desta forma, aumenta os graus de liberdade em seus dados. Adicionar parâmetros ao seu modelo (aumentando o número de termos em uma equação de regressão, por exemplo) "gasta" informações dos seus dados, e reduz os graus de liberdade disponíveis para estimar a variabilidade das estimativas de parâmetro. Na RLM deve-se estimar uma parâmetro para cada termo que você escolha incluir no modelo, e cada um consome um grau de liberdade. Portanto, incluir termos em excesso em um modelo de RLM reduz os graus de liberdade disponíveis para estimar a variabilidade dos parâmetros, e pode torná-lo menos confiável.

O **SQ** é a soma dos quadrados dos desvios totais, que representa a dispersão da variação aleatória de y em relação a sua média y total. O resultado final da SQ é obtido por meio da SQ da Regressão, que é a variação dos valores de y em torno de sua média (explicada pela regressão) + a SQ dos Resíduos, que é diferença entre os valores de y determinados e y' estimados (variação residual não explicada pela regressão).

O **Quadrado Médio (MQ)** é obtido pela divisão da Soma de Quadrados por seus respectivos graus de liberdade.

O **F**/ **F** de significação relaciona-se ao teste F, o qual avalia se o modelo proposto é útil para explicar a variável dependente, ou seja, busca identificar se pelo menos uma das variáveis independentes está relacionada à variável dependente. Assim, o F de significação deve ser < 0,005 para que o modelo seja considerado útil, visto que avalia a significância estatística geral do modelo estimado.

• **RESULTADO DA REGRESSÃO**

	Coeficientes	Erro padrão	Stat t	valor-P	95% inferiores	95% superiores	Inferior 95,0%	Superior 95,0%
Interseção	-88,60667471	77,71696302	-1,140120139	0,262443891	-246,7230248	69,50967543	-246,7230248	69,50967543
GASTO MILHÕES	0,841789202	0,390925457	2,153324086	0,038696133	0,04644538	1,637133025	0,04644538	1,637133025
DURAÇÃO	1,373209971	0,585938149	2,34360909	0,025272059	0,181109844	2,565310098	0,181109844	2,565310098
1								

Figura 23. Resultado da regressão

Os **coeficientes** são os números pelos quais as variáveis da equação serão multiplicadas. Esse valor representa o quando a variável dependente irá variar quando a respectiva variável independente variar 1 unidade.

O erro padrão é o mesmo desvio padrão de uma estimativa. O erro padrão do coeficiente mede o grau de precisão com que o modelo estima o valor desconhecido do coeficiente. Quanto menor o erro padrão, mais precisa é a estimativa.

Dividir o coeficiente pelo erro padrão calcula o valor-t, ou start t.

O valor-P representa o teste de significância individual. Este dado fornece a significância estatística de cada parâmetro a ser considerado no modelo de regressão. Assim, busca saber quais variáveis estão relacionadas com a variável dependente. Para que exista significância, o valor-P deve ser < 0.05 (à 95%).

Os **95% inferiores e superiores** equivalem-se ao nível de confiança da regressão. As duas primeiras colunas são geradas automaticamente pelo Excel e as duas últimas são geradas de acordo com a escolha do pesquisador, caso queira comparar o modelo com outro grau de confiança que não seja 95%. Como, neste caso, não foi escolhido um nível de confiança diferente de 95%, o Excel replica as duas primeiras colunas.

Passo 5: Substituir os dados na equação da RLM

Conforme apresentado anteriormente, este passo consiste em substituir os valores encontrados por meio da RLM na equação original, conforme Figura 24 e Figura 25. Abaixo apresenta-se a substituição de valores na equação da RLM.

$$Y = \beta_0 + \beta_1 . X_1 + \beta_2 . X_2 + u$$

FAT = - 88,6 + 0,84. GAS + 1,37 . DUR + 101,13

Arc	quivo	Página Inicial	Inserir	Layou	ıt da Pá	gina Fórm	ulas Dados	Revisão	Exibir Q () que você deseja faz	35		
Obte	er Dados ternos *	Nova Consulta + Co Obter e	Mostrar (Da Tabeli Fontes Ri Transform	Consultas a ecentes nar	Atuali: Tudo	Conexõ Proprie Conexões	dades inks Ž↓ Ž	ssificar Filtro	K Limpar Reaplicar Avançado iltrar	Texto para Colunas 🗟 *	Teste de Pla Hipóteses ~ P Previsão	inilha de revisão o Estruturo	upar • 🗐 🔚 An agrupar • 🗐 ?, So total a de Tópicos 💈
M	12	• = 2	$< \vee$	fx									
		A			в		с	D	Е	F	G	Н	1
1	RESUM	DOS RESULT	ADOS										
2													
3		Esta	tística d	e regressâ	0								
4	R múlti	plo				0,528442429							
5	R-Quad	rado				0,279251401							
6	R-quad	rado ajustado				0,235569668							
7	Erro pa	drão			u	101,1332854							
8	Observ	ações				36							
9													
10	ANOVA						142-11						
11					gl		SQ	MQ	F	F de significação			
12	Regress	ião				2	130771,6866	65385,84329	6,392864477	0,004502283			
13	Residuc)				33	337522,0664	10227,94141					
14	Total					35	468293,753						
15	<u> </u>			7. MP		1.000 C		And Anno 2000	a second and a second sec				
16				Co	peficier	ites	Erro padrão	Stat t	valor-P	95% inferiores	95% superiores	Inferior 95,0%	Superior 95,0%
17	Interse	ção			B0 -	88,60667471	77,71696302	-1,140120139	0,262443891	-246,7230248	69,50967543	-246,7230248	69,50967543
18	GASTO	MILHOES			B1	0,841789202	0,390925457	2,153324086	0,038696133	0,04644538	1,637133025	0,04644538	1,637133025
19	DURAÇ	40			DZ	1,373209971	0,585938149	2,34360909	0,025272059	0,181109844	2,565310098	0,181109844	2,565310098
20													
21													

Figura 24. Substituição de valores na RLM

É importante também analisar a significância de cada variável independente em relação à variável dependente por meio do "valor-P", conforme segue:

	Coeficientes	Erro padrão	Stat t	valor-P	95% inferiores	95% superiores	Inferior 95,0%	Superior 95,0%
Interseção	-88,60667471	77,71696302	-1,140120139	0,262443891	-246,7230248	69,50967543	-246,7230248	69,50967543
GASTO MILHÕES	0,841789202	0,390925457	2,153324086	0,038696133	0,04644538	1,637133025	0,04644538	1,637133025
DURAÇÃO	1,373209971	0,585938149	2,34360909	0,025272059	0,181109844	2,565310098	0,181109844	2,565310098

Figura 25. Valor-p (teste t)

O exemplo abaixo, exposto na Figura 26, busca clarear a significância de cada uma das variáveis.



Figura 26. Significância das variáveis

Passo 6: Interpretação dos resultados

Analisando os resultados, observa-se que ao relacionar as variáveis independentes com a variável dependente no modelo proposto, a relação entre elas pode ser considerada fraca, uma vez que o resultado do \mathbb{R}^2 foi 0,27. Além disso, por meio do valor do erro padrão (101,1), é possível que existam variáveis independentes que não estão sendo consideradas no modelo proposto. Ou seja, a diferença entre o valor real de Y e o valor previsto de Y pode ser considerada significativamente alta. Por outro lado, o teste F aponta que o modelo proposto é útil para explicar a variável dependente, visto que o F de significação foi de 0,004, mantendo-se abaixo de 0,05. Ainda, por meio do "valor – P" observa-se que as variáveis gasto e duração são significativamente relacionadas com a variável dependente, visto que ambas, individualmente, apresentaram valor < 0,05, em um intervalo de 95% de confiança. Por fim, constata-se que a cada aumento da variável gasto, o valor de Y aumentará 0,84. Do mesmo modo, para cada aumento na duração do filme, o valor de Y aumentará 1,37.

Conforme citado anteriormente, observa-se que outras variáveis podem não estar sendo consideradas neste modelo, visto que o valor de "u" é alto. Em outras palavras, é possível que **outras variáveis além do gasto e do tempo de duração do filme influenciem o faturamento de um filme**.

* Variáveis Dummy

A partir da interpretação dos resultados apresentada, acrescentamos aqui uma variável independente relativa ao período de lançamento dos filmes, separando os filmes lançados antes de 1990 dos lançados após 1990. Assim, antes de realizar novamente a RLM, é necessário transformar as variáveis qualitativas em variáveis *dummy*, atribuindo a elas valores numéricos. Neste exemplo, faz-se a pergunta: o filme foi lançado após de 1990? A resposta SIM equivale a 1, enquanto a resposta NÃO equivale à 0. Deste modo, neste modelo, os filmes lançados antes de 1990 estão identificados pelo número 0, enquanto os filmes lançados após 1990 estão identificados com o número 1, conforme Figura 27.

Arquivo	Página Inicial Inserir La	yout da Página	Fórmulas Da	dos Revisão	Exibir Q	O que voci	ê deseja faze	r						Entra
Obter Dado Externos *	Nova Consulta - Consulta	is Atualizar Tudo +	onexões 2↓ ropriedades ditar Links Å↓	Classificar Fi	tro	Texto pa Coluna	III = 1 III = 1 III III = 1 III = 1 III III = 1 III = 1 III III = 1 III III = 1 III IIII = 1 IIII = 1 III III IIII = 1 III IIII I	Teste de Hipótese	e Planilha de s * Previsão	🕮 Agru 👰 Desag 🔠 Subto	par + * grupar + otal	Ana ?, Solv	álise de Dados ver	
	Obter e Transformar	Conex	ões	Classifica	ir e Filtrar	Ferramer	ntas de Dado	is P	revisão	Estrutura	de Tópicos	rg .	Análise	
P5	▼ : × ✓ fx													
4	А	В	с	D	E	F	G	н	1	J	к	L	M	N
1	TÍTULO	FATURAMENTO	GASTO MILHÕ	ES DURAÇÃO	LANÇAMENTO									
2 A1		81,843	18,5	137	0				6		u uõre			
3 A2		194,125	140	144	1				G	ASTU M	ILHUES			
4 A3		147,54	50	138	1		250							
5 B1		75,6	72	177	1									
6 C1		12,006	0,3	105	1		200						•	
7 C2		100,853	90	153	1									
8 D1		67,155	104	112	1		150							
9 D2		140,424	75	120	1				1.00					
10 E1		68,75	55	129	1		100	••				Constant of the owner owner owner		
11 F1		329,691	55	142	1									
12 G1		217,631	22	128	1		50	. 8	8	•				
13 G2		198,571	3,9	222	0									
14 G3		138,339	10	126	1		0							
15 G4		181,28	6	110	0		0	100	200	300	400	500	600	700
16 H1		47	0,3	93	0									
17 H2		19,819	70	95	1					DUDA	-ño			
18 11		72,219	17	100	1					DURA	ÇAU			
19 12		306,124	75	142	1		250							
20 13		197,171	39	127	0									
21 14		260	12	124	0		200						•	
22 M1		250,147	90	98	1			•						
23 M2		20,1	45	117	1		150							

Figura 27. Inclusão da variável lançamento

Após substituir os anos por 0 e 1, é necessário seguir os mesmos passos apresentados anteriormente. O primeiro deles, como já sabemos quais são as variáveis independentes e dependentes, é desenhar o gráfico de dispersão, conforme Figura 28.





Após, é necessário gerar a **equação básica** de RLM. Observa-se que, com a inclusão de uma nova variável (lançamento), acrescentou-se também na equação mais uma variável.

 $Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + u$

O próximo passo consiste em realizar a **RLM** no Excel. O "intervalo Y de entrada" permanece o mesmo, porém no "intervalo X de entrada" é necessário selecionar todas as variáveis independentes, inclusive a *dummy* (lançamento), conforme a Figura 29.

Arguna	Página Inicial Inserir L	ayout da Página 🛛 f	órmulas 🛛	Dados Re	visão	Exibir Q	O que você d	eseja fazer En
Doter Dados Externos *	Dora Consulta Nova Consulta - Co Fontes Recentes Obter e Transformar	as Atualizar Tudo - Conexõ	nexões g opriedades tar Links es	2↓ <u>ZA</u> (↓ Classificar Clas	Filtro sificar e F	₩ Limpar Reaplicar Avançado iltrar	Texto para Colunas Ferramenta	Image: Constraint of the phane in the p
P8	▼ : × ✓ fx							
	A	В	С	D	i li	E	F	Regressão ? × N
1	TÍTULO	FATURAMENTO	GASTO MILH	IÕES DURA	ção la	NÇAMENTO		Tabuda
2 A1		81,843	18,5	137	7	0		Integrale V de entrada: CPC1.CPC37 OK
3 A2		194,125	140	144	1	1		Cancelar
4 A3		147,54	50	138	3	1	-	Intervalo X de entrada: SCS1:SES37
5 B1		75,6	72	177	7	1		Ajuda
6 C1		12,006	0,3	105	5	1	-	Vivel de confiance 95 %
7 C2		100,853	90	153	3	1		
8 D1		67,155	104	112	2	1		Opções de saída
9 D2		140,424	75	120)	1		🔿 Intervalo de saída:
10 E1		68,75	55	129	9	1		Noya planilha:
11 F1		329,691	55	142	2	1		O Nova pasta de trabalho
12 G1		217,631	22	128	3	1		Resíduos
13 G2		198,571	3,9	222	2	0		Residuos Plotar residuos
14 G3		138,339	10	126	5	1		Resíduos padronizados Plotar ajuste de linha
15 G4		181,28	6	110)	0		Probabilidade normal 20 700
16 H1		47	0,3	93		0		Plotagem de probabilidade normal
17 H2		19,819	70	95	Q	1		
18 11		72,219	17	100)	1		
19 12		306,124	75	142	2	1	1	50
20 13		197,171	39	127	7	0		
21 14		260	12	124		0	4	00
22 M1		250,147	90	98		1		• •
23 M2		20,1	45	117	7	1	1	50

Figura 29. RLM com variável dummy

Os resultados da RLM são apresentados na Figura 30.

	Α		В	С	D	E	F	G	н	I.	
1	RESUMO DOS RESULTADOS										Γ
2											
3	Estatística o	de regressão									
4	R múltiplo		0,606485796								
5	R-Quadrado		0,36782502								
6	R-quadrado ajustado		0,308558616								
7	Erro padrão		96,18400234								
8	Observações		36								
9											
10	ANOVA										
11		ç	gl	SQ	MQ	F	F de significação				
12	Regressão		3	172250,1592	57416,71973	6,206298902	0,001910885				
13	Resíduo		32	296043,5938	9251,362306						
14	Total		35	468293,753							
15											
16		Coefic	cientes	Erro padrão	Stat t	valor-P	95% inferiores	95% superiores	Inferior 95,0%	Superior 95,0%	
17	Interseção		-29,44842078	79,0177091	-0,372681278	0,711842811	-190,4022272	131,5053856	-190,4022272	131,5053856	
18	GASTO MILHÕES		1,285654149	0,426817829	3,012184734	0,005035474	0,416254681	2,155053616	0,416254681	2,155053616	
19	DURAÇÃO		1,297604692	0,55840613	2,32376513	0,026647909	0,160168628	2,435040757	0,160168628	2,435040757	
20	LANÇAMENTO		-93,83568537	44,3158922	-2,117427422	0,042088364	-184,1042038	-3,567166901	-184,1042038	-3,567166901	
21											

Figura 30. Resultados da regressão com variável dummy

Com base nos resultados apontados, é necessário novamente substituir a equação de

RLM:



Novamente os resultados serão interpretados. Observa-se que, ao relacionar as variáveis independentes com a variável dependente, a relação entre elas continua a ser considerada fraca, uma vez que o resultado do **R**² foi 0,36. Porém, observa-se que com a inclusão de mais uma variável independente a força da relação entre as variáveis aumentou. Os dados novamente apontam que ainda é possível existir outras variáveis independentes que não estão sendo consideradas no modelo, visto que o valor do **erro padrão** ainda é considerado alto (96,1). Por meio do **Teste F** observa-se que o modelo proposto é útil para explicar a variável dependente, visto que o **F de significação foi de 0,001**, mantendo-se abaixo de 0,05. Por meio do "valor – P" observa-se que as variáveis **gasto**, **duração e lançamento** são individualmente significativamente relacionadas com a variável dependente, visto que todas apresentaram valor < 0,05, em um intervalo de 95% de confiança. Por fim, ressalta-se que a cada aumento da variável gasto, **o valor de Y aumentará 1,29** e para cada aumento do ano de lançamento **o valor de Y diminuirá 93,83**. Neste caso o valor diminuirá devido ao resultado negativo apontado na regressão.

Conforme citado anteriormente, **observa-se ainda que outras variáveis podem não estar sendo consideradas neste modelo**. Assim, será acrescentada a variável independente "**faixa etária**", a qual pode ser: livre, maior que 14 anos e maior que 16 anos. Nas figuras seguintes, cada uma das abas da planilha foi renomeada de acordo com a regressão que está sendo realizada, a fim de nortear o leitor desta apostila. Assim, RLM 1 corresponde à primeira RLM, sem variáveis *dummy;* RLM 2 corresponde à RLM realizada com a variável *dummy* de lançamento e RLM 3 corresponde à RLM que está sendo realizada neste momento.

Conforme apresentado anteriormente, novamente deparamo-nos com uma variável qualitativa, visto que a faixa etária, neste caso, não corresponde à um valor quantitativo. Seguindo o exemplo anterior, poderíamos atribuir os seguintes valores para cada categoria (Tabela 2):

ruotiu =: cuteBornuo E uninity	
Categoria da Variável	Valor atribuído
Livre	0
Maior que 14	1
Maior que 16	2

Tabela 2 Categorias Dummy

Porém, é válido ressaltar que, diferente do exemplo anterior onde havia apenas duas categorias (lançados antes de 1990 ou lançados após 1990), esta variável *dummy* apresenta

três categorias. Assim, segundo Belfiore (2015), quando houver mais de duas categorias em uma variável *dummy*, é possível seguir dois caminhos:

- Se a variável independente for constante, pode-se utilizar os valores 1, 2, 3, 4, 5, e assim por diante. Porém, observa-se que geralmente a variável independente não é constante. Portanto:
- Criam-se duas variáveis *dummy*. Por exemplo, opta-se por fazer duas perguntas: o filme é para maiores de 14 anos? e; O filme é para maiores de 16 anos? A resposta SIM equivale ao número 1 e a resposta NÃO equivale ao número 0. Deste modo, quando ambas as perguntas receberem a resposta não, o filme terá uma faixa etária livre (Tabela 3).

Tabela 3. Categorias dummy com	mais de duas variáveis	
Categoria da Variável	É para maiores de 14?	É para maiores de 16?
Livre	0	0
Maior que 14	1	0
Maior que 16	0	1

Assim, seguindo o procedimento sugerido, as novas colunas inseridas na amostra serão "Faixa etária + 14" e "Faixa Etária + 16", conforme Figura 31.

Are	quivo Página Inicial Inserir L	ayout da Página	Fórmulas Dado	s Revisão	Exibir Q	O que você deseja fazer		
		→ A → A → B → A	• • • • • • •	Quebrar Text Mesclar e Ce	o Automaticament ntralizar 🔹	e Geral *	Formatação Condicional ≁	Formatar como Tabela * Célula *
Árei	a de Transf 🕞 Fonte	r _a	Alin	hamento		rs Número r	â	Estilo
14	• : × ~ f _x							
1	А	В	с	D	E	F	G	н
1	τίτυιο	FATURAMENTO	GASTO MILHÕES	DURAÇÃO	LANÇAMENTO	FAIXA ETÁRIA GERAL	É FAIXA + 14	É FAIXA ETÁRIA + 16
2	A1	81,843	18,5	137	0	t i	1 1	0
3	A2	194,125	140	144	1		1	0
4	A3	147,54	50	138	1		0 0	0
5	B1	75,6	72	177	1		1 1	0
б	C1	12,006	0,3	105	1		0 0	0
7	C2	100,853	90	153	1		1 1	0
8	D1	67,155	104	112	1		1 1	0
9	D2	140,424	75	120	1		2 0	1
10	E1	68,75	55	129	1		0 0	0
11	F1	329,691	55	142	1	V 1	L 1	0
12	G1	217,631	22	128	1	(0 0	0
13	G2	198,571	3,9	222	0		0 0	0
14	G3	138,339	10	126	1		1 1	0
15	G4	181,28	6	110	0		0 0	0
16	H1	47	0,3	93	0		2 0	1
17	H2	19,819	70	95	1		0 0	0
18	11	72,219	17	100	1		2 0	1
19	12	306,124	75	142	1		L 1	0
20	13	197,171	39	127	0		0 0	0
21	14	260	12	124	0		2 0	1
22	M1	250,147	90	98	1	¥¥	0 0	0
23	M2	20,1	45	117	1	- 1	1	0

Figura 31. Inclusão das colunas dummy

Novamente os passos apresentados anteriormente serão retomados. O primeiro deles consiste no gráfico de dispersão. Neste caso, o gráfico não será gerado, pois não podemos

inferir que há uma relação linear entre mais de duas variáveis. Assim, conforme citado anteriormente, é necessário **gerar a equação básica de RLM**. Observa-se que foram acrescentados o $\beta_{4.}X_{4}$ e o $\beta_{5.}X_{5}$ na equação, os quais correspondem respectivamente às variáveis independentes maior de 14 anos e maior de 16 anos, relacionadas à **faixa etária** do filme.

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_4 \cdot X_4 + \beta_5 \cdot X_5 + u$$

O próximo passo consiste em realizar a RLM no Excel, conforme apresentado anteriormente. O "intervalo Y de entrada" permanece o mesmo, porém no "intervalo X de entrada" é necessário selecionar todas as variáveis independentes, inclusiva as *dummy* (lançamento e faixa etária). A Figura 32 demonstra como gerar a RLM e a Figura 33 apresenta os resultados gerados por meio da respectiva regressão.

	Página Inicial Inserir	Layout da Página	Formulas Dade	Revisão	Exitur 🖓		seja fazer		Entrar A
Obter Dado Externos *	Nova Nova Comutta - Da Tabela Obter e Transform	onsultas Atualizas Tudo - Gines ar Cones	nneeden 21 opriedsdes Gae Linko Bes	(112) Interation Pro Operation	The Limpson The Prophese The Averagede Confliction	Testo para Columni Fernamentar	H ng Tente da H ng Hipóteses H Dados Prés	III Agnuer III Agn	
HB	(*) 1 (8) (8)	f.						Regressão	7 ×
- 16	A	8	c	D	E	F	G	Entrada	N N
1	TÍTULO	FATURAMENTO	GASTO MILHÕES	DURAÇÃO	LANÇAMENTO	É FAIXA + 14	É FAIXA ETÁRIA + 16	intervalo y de entrada: \$851-58537	
2 A1		81,843	18,5	137	0	1	0	Intenalo X de entrada: 6/51-5/517	Cancelar 15
3 A2		194,125	140	144	1	1	0	ACCURATE ACC	Aines
4 A3		147,54	50	138	1	0	0	Rótulos Constante é gero	Choine
5 81		75,6	72	177	1	1	0	P Nivel de confiança 95 %	
6 C1		12,006	0,3	105	1	0	0	Oprior de raida	
7 C2		100,853	90	153	1	1	0	Optimite de antes	
8 D1		67,155	104	112	1	1	0		
9 D2		140,424	75	120	1	0	1	Noya planifia:	-
10 E1		68,75	55	129	1	0	0	() Nova pesta de grabalho	
11 F1		329,691	55	142	1	1	0	Residuos	
12 G1		217,631	22	128	1	0	0	Residuos nadronizados Rotar ajuste de linha	
13 G2		198,571	3,9	222	0	0	0	Characteristic for the state	
14 G3		138,339	10	126	1	1	0	Probabilidade normal	1
15 G4		181,28	6	110	0	0	0	D giotagen de probabilidade homai	
16 H1		47	0,3	93	0	0	1		
17 H2		19,819	70	95	1	0	0	D	URAÇÃO
18 11		72,219	17	100	1	0	1	250	
19 12		306,124	75	142	1	1	0		
20 13		197,171	39	127	0	0	0	200	
21 14		260	12	124	0	0	1		
22 M1		250,147	90	98	1	0	0	150	
23 M2		20,1	45	117	1	1	0	and a second	

Figura 32. Gerando a RLM com mais de uma variável dummy

ter Dados dernos *	Nova Consulta - Dafe Obter e T	lostrar Consultas a Tabela ontes Recentes unsformar	Atualizar Tudo - Conest	tes daties inks	Classificar e F	To Resplicer To Resplicer To Avançado Iltrar	Testo para Colunas 🕷 • 🖗 Ferramentas de Dad	Teste de Pla Hipóteses - P Previsia	nilha de revisão s Estrutu	rspar • · · · · · · · · · · · · · · · · · ·	Análise 2. Solver Ani	de Dados	
(10		× 14											
a)	A		1	с	D	E	E.	G	н	1	1	ĸ	
RESUM	O DOS RESULTA	DOS								- 1161			
6			9										
	Estati	stica de regressà	io										
R mult	iplo		0,647498876	<u>.</u>									
R-Quar	drado		0,419254794										
R-quad	frado ajustado		0,322463926										
Erro pa	oEtbe	_	95,21193063										
Observ	rações	_	36										
ANOV	A												1
			gi	5Q	MQ	F	F de significação						-
Regres	são		5	196334,4009	39266,88019	4,331553214	0,004375223						
Residu	0		30	271959,352	9065,311735								
t Total		_	35	468293,753									
8		0	oeficientes	Erro padrão	Stat t	valor-P	95% inferiores	95% superiores	Inferior 95.0%	Superior 95.0%			
7 Interse	ecão		-30,60527802	78,33611574	-0.390691799	0.698785775	-190.5889695	129.3784135	-190.5889695	129,378413	5		
GASTO	MILHÕES		1,395256629	0,427981393	3,260087123	0,002772564	0,521202018	2,269311241	0,521202018	2,26931124	1		
DURAC	ÃO		1,248519028	0,561403332	2,223925217	0,033827946	0,101980465	2,39505759	0,101980465	2,3950575	9		
LANÇA	MENTO		-93,60739999	44,05605239	-2,124734172	0,041960894	-183,5818623	-3,632937652	-183,5818623	-3,63293765	2		
EFAIX	A + 14		-22,81584558	36,98913113	-0,61682567	0,54200298	-98,35772927	52,72603812	-98,35772927	52,7260381	2		
EFAIX	A ETÁRIA + 16		47,05523717	42,82781283	1,098707453	0,280635417	-40,41082534	134,5212997	-40,41082534	134,521299	7		

Figura 33. Resultado da RLM com mais de uma variável dummy

Com base nos resultados apontados, é necessário novamente substituir a equação:



Novamente os resultados serão interpretados. Ao relacionar as variáveis independentes com a variável dependente, observa-se que a relação entre elas pode ser considerada fraca, uma vez que o resultado do \mathbb{R}^2 foi 0,41. Porém, observa-se que com a inclusão de mais uma variável, a força da relação entre as variáveis aumentou. Ainda é possível que existam outras variáveis independentes que não estão sendo consideradas no modelo proposto, visto que o valor do erro padrão ainda é considerado alto (95,2). Porém, observa-se que com a inclusão de outras variáveis o valor do erro padrão diminuiu. Por meio do Teste F observa-se que o modelo é útil para explicar a variável dependente, visto que o F de significação foi de 0,004, mantendo-se ainda abaixo de 0,05. Por meio do "valor – P" observa-se que apenas as variáveis gasto, duração e lançamento são significativamente relacionadas com a variável dependente, visto que estas apresentaram valor < 0,05 em um intervalo de 95% de confiança. As variáveis relacionadas à faixa etária não apresentaram alta significância. Por fim, ressaltase que a cada aumento da variável gasto, o valor de Y aumentará 1,39; para cada aumento

da duração do filme, **o valor de Y aumentará 1,24**; para cada aumento do ano de lançamento, **o valor de Y diminuirá 93,60**; para cada faixa etária maior de 14 anos, **o valor de Y diminuirá 22,8** e para cada faixa etária maior de 16 anos, **o valor de Y diminuirá 47,05**. Para saber a relação com a variável **faixa etária livre**, é necessário substituir β 4 e β 5 por 0.

Com base no modelo proposto, ainda há um **próximo passo** a ser seguido, o qual consiste em verificar a existência ou não de **multicolinearidade** entre as variáveis independentes, ou seja, verificar se existe alta relação entre as variáveis independentes. Embora as variáveis independentes tenham relação com a variável dependente, quando existe alta correlação entre as variáveis independentes o R^2 tende a diminuir e assim é necessário retirá-las da equação. Para verificar a multicolinearidade do modelo proposto é necessário realizar a **Matriz de Correlação**, disponível em Dados > Análise de Dados > Correlação, conforme apresentado na Figura 34.

Arquivo	Página Inicial Inserir La	yout da Página	Fórmulas Dad	los Revisão	Exibir 🖓		eja fazer							Entrar	A co
Obter Dados Externos *	Nova Consulta - Consulta Da Tabela Do Tabela Do Tabela Consulta - Consulta - Consulta Obter e Transformar	s Conex	onexões ropriedades fitar Links ões	Classificar Filtr	 K Limpar Reaplicar Avançado Filtrar 	Texto para Colunas	e Dados Previ	Planilha de Previsão isão Est	Agrupar Desagru Subtota rutura de	r • + ipar • - I Tópicos	An 2., Sol	álise de Dad ver Análise	35		
H4	▼ : × √ f _x														
. d	A	В	с	D	E	F	G	н	I.	j	К	L	11	M	N
1	TÍTULO	FATURAMENTO	GASTO MILHÕE	S DURAÇÃO I	ANÇAMENTO	É FAIXA + 14	É FAIXA ETÁRIA + 16	Análise de da	dos					?	×
2 A1		81,843	18,5	137	0	1	0			(a)				-	10.1
3 A2		194,125	140	144	1	1	0	Ferramentas	de analis	e				OK	
4 A3		147,54	50	138	1	0	0	Anova: fato Anova: fato	r unico r duplo co	om repetic	io		^	Cancela	ar
5 B1		75,6	72	177	1	1	0	Anous: fato	duplo se	em repetiç	io		- 11	concerto	-
6 C1		12,006	0,3	105	1	0	0	Covariancia					-	Ajuda	
7 C2		100,853	90	153	1	1	0	Estatística d	lescritiva						
8 D1		67,155	104	112	1	1	0	Ajuste expo Teste-F: dua	nencial is amostra	as para var	lâncias				
9 D2		140,424	75	120	1	0	1	Análise de F	ourier						
10 E1		68,75	55	129	1	0	0	Histograma							
11 F1		329,691	55	142	1	1	0	5	0	. 5					
12 G1		217,631	22	128	1	0	0					•			
13 G2		198,571	3,9	222	0	0	0								
14 G3		138,339	10	126	1	1	0		0	100	é.	200	300		400
15 G4		181,28	6	110	0	0	0								
16 H1		47	0,3	93	0	0	1								
17 H2		19,819	70	95	1	0	0						DU	RAÇÃO)
18 11		72,219	17	100	1	0	1	250							
19 12		306,124	75	142	1	1	0					1			
20 13		197,171	39	127	0	0	0	200				1			
21 14		260	12	124	0	0	1								
22 M1		250,147	90	98	1	0	0	15/							
23 M2		20,1	45	117	1	1	0	15					•	•	

Figura 34. Gerando a matriz de correlação

O "intervalo Y de entrada" corresponde à variável dependente, enquanto o "intervalo X de entrada" corresponde à todas as variáveis independentes, conforme Figura 35.

Arquive	Página Inicial Inserir La	yout da Página	Fórmulas Da	idos Revisão	Exibir Q	O que você deseja fazer)			Entrar	Ac
Obter Dados Externos *	Nova Consulta - Co Fontes Recentes Obter e Transformar	s Conex	onexões ∄↓ ropriedades fitar Links Å↓	Classificar Filt	 K Limpar Reaplicar Avançado Filtrar 	Texto para Colunas S - B Ferramentas de Dados	Teste de Plan Hipóteses - Pr Previsão	Agrupar - + milha de evisão Estrutura de Tópicos	Análise de Dados		
114											
114	· · · · · · · ·										
1	А	В	С	D	E	F	G	Correlação		7 X	N
1	TÍTULO	FATURAMENTO	GASTO MILHÕ	DURAÇÃO	LANÇAMENTO	É FAIXA + 14 É FAIXA E	ETÁRIA + 16	conclução		1 14	c
2 A1		81,843	18,5	137	0	1	0	Entrada	(D44.65437 (52)	OK	3
3 A2		194,125	140	144	1	1	0	intervalo de gituada.	3031:30337	Cancelar	1
4 A3		147,54	50	138	1	0	0	Agrupado por: (<u>C</u> olunas		
5 B1		75,6	72	177	1	1	0		Linhas	Ajuda	
6 C1		12,006	0,3	105	1	0	0	Kotulos na primeira linna	_		
7 C2		100,853	90	153	1	1	0	Opções de saída			
8 D1		67,155	104	112	1	1	0	🔿 Intervalo de saída:			
9 D2		140,424	75	120	1	0	1	Noga planilha:			
10 E1		68,75	55	129	1	0	0	O Nova pasta de trabalho			
11 F1		329,691	55	142	1	1	0				
12 G1		217,631	22	128	1	0	0				
13 G2		198,571	3,9	222	0	0	0	0	• • • •		
14 G3		138,339	10	126	1	1	0	0 100	200	300 4	400
15 G4		181,28	6	110	0	0	0				
16 H1		47	0,3	93	0	0	1				
17 H2		19,819	70	95	1	0	0			DURAÇÃO	
18 11		72,219	17	100	1	0	1	250			
19 12		306,124	75	142	1	1	0				
20 13		197,171	39	127	0	0	0	200			
21 14		260	12	124	0	0	1	•			
22 M1		250,147	90	98	1	0	0	150 🐢			
23 M2		20,1	45	117	1	1	0	• • •			

Figura 35. Intervalos de entrada para matriz de correlação

A partir da geração da correlação, o Excel irá gerar uma nova planilha, conforme a Figura 36, que apresentará a Matriz de Correlação.

Arc	quivo	Página Inicia	ıl İnserir	Layo	ut da Página	Fórm	nulas (Dados	Rev	isão	Exibir Q () que você	deseja fazer			
Obt Ext	er Dados ternos *	Nova Consulta - Obte	Mostrar Co Da Tabela o Fontes Reco er e Transforma	nsultas entes	Atualizar Tudo + Cor	Conexô Proprie Editar L iexões	inks	↓ <mark>7</mark> ↓ Clas	ssificar Class	Filtro ficar e Fil	K Limpar Reaplicar Avançado Itrar	Texto para Colunas Ferrament	Image: marked black Image: marked black Image: marked black Image: marked black Image: marked black	Teste de Hipótese	Planilha de s - Previsão revisão	현물 Agrupa 현물 Desagn 문매 Subtota Estrutura d
K	11	•	× ✓	f_X												
A		А	В		с		D	1	E		F	1	G	н	1	J
1			FATURAN	IENTO	GASTO MIL	HÕES	DURAÇÂ	ío l	ANÇAN	MENTO	É FAIXA + 14	É FAIXA	ETÁRIA + 10	5		
2	FATUR	MENTO		1					940							
3	GASTO	MILHÕES	0,399	111968		1										
4	DURAÇ	ÃO	0,421	876249	0,2082	52623		1								
5	LANÇA	MENTO	-0,082	815832	0,4897	06903	0,047453	957		1						
6	ÉFAIXA	+ 14	-0,076	023408	0,206	54976	0,109352	054	0,152	286226		1				
7	É FAIXA	ETÁRIA + 1	6 0,189	579548	-0,1165	53577	0,084306	498	-0,035	714286	-0,42640143	3		1		
8																
9																
10																
11	1															
12																
13																
14																
15																
16																
17								_						1		1
18																
19								_						1		
20																
21								-						-		
22																
23		1 million	221 1 (2004) 000-00.V	1 1000 1 10				~ [0.100000	8 1	~				-	
	<. (E)	RLM	1 RLM 2	RLM	3 MATR	Z DE C	ORRELAÇ	AO	FILME	S	(+)				4	
Pro	nto															

Figura 36. Matriz de correlação

Caso exista uma correlação entre as variáveis independentes maior do que 0,6 é necessário excluí-las do modelo. No exemplo que está sendo utilizado, houve um valor

próximo à 0,6 entre lançamento e gasto, sugerindo uma possível multicolinearidade. Assim, iremos testar alguns modelos para análise, verificando qual deles apresenta o maior R^2 ajustado, já que este é utilizado quando há o intuito de comparar o coeficiente de ajuste (R^2) entre dois modelos ou entre um mesmo modelo com tamanhos de amostras diferentes.

Escolhendo o melhor modelo de regressão...

Modelo 1 - Considerando todas as variáveis

O Modelo 1 considera a relação entre a variável dependente e todas as variáveis independentes apresentadas. Observa-se que este modelo já foi realizado no item anterior (RLM3) e apresentou R^2 ajustado de 0,32. Ou seja, as cinco variáveis independentes explicam juntas 32% da variável dependente. A Figura 37 apresenta os resultados desta RLM, a qual considera todas as variáveis.

uivo Página Inicial Inse	rir Layout da Página Fórm	iulas Dados	Revisão	Exibir ⊉ () que você deseja faz	er		
Colar V I S V	$ \begin{array}{c c} \bullet & & \bullet \\ \hline \bullet & \bullet & \bullet \\ \hline \bullet & \bullet & \bullet \\ \hline \end{array} \bullet \begin{array}{c} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \\ \hline \bullet & \bullet & \bullet \\ \hline \bullet & \bullet & \bullet \\ \hline \end{array} \bullet \begin{array}{c} \bullet & \bullet \\ \bullet & \bullet \\ \hline \bullet & \bullet \\ \hline \end{array} \bullet \begin{array}{c} \bullet & \bullet \\ \bullet & \bullet \\ \hline \bullet & \bullet \\ \hline \end{array} \bullet \begin{array}{c} \bullet & \bullet \\ \bullet \\ \bullet \\ \hline \end{array} \bullet \begin{array}{c} \bullet & \bullet \\ \bullet \\ \bullet \\ \hline \end{array} \bullet \begin{array}{c} \bullet & \bullet \\ \bullet \\ \bullet \\ \bullet \\ \hline \end{array} \bullet \begin{array}{c} \bullet & \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \hline \end{array} \bullet \begin{array}{c} \bullet & \bullet \\	≫ - ₽(⊡ ⊡ □	Quebrar Texto Au Mesclar e Central	itomaticamente izar +	Geral ♀ % ∞∞ 5	→ Formataçã Condiciona	io Formatar como I * Tabela *	Estilos de Célula *
i de Iranst Isi Fo	inte la	Alinn	amento		s Numero	19	Estilo	
0 • I × V	f _x							
А	В	с	D	Е	F	G	н	1
RESUMO DOS RESULTADOS	******			· · · · · ·				
Estatística	de regressão							
R múltiplo	0,647498876							
R-Quadrado	0,419254794	-						
R-quadrado ajustado	0,322463926							
Erro padrão	95,21193063	0						
Observações	36							
ANOVA								
	gl	SQ	MQ	F	F de significação			
Regressão	5	196334,4009	39266,88019	4,331553214	0,004375223			
Resíduo	30	271959,352	9065,311735					
Total	35	468293,753						
	Confidential		01-11	and an D	0594 1- 5- 1	050/	1-5-1-05 0%	0
Intercoção	20 60527802	79 22611574	0 200601700	0 600705775	100 500605	120 2704125	100 5990605	120 270/125
GASTO MILHÕES	-30,00327802	0.427981393	3 260087123	0.002772564	-150,5665055	2 2692112/1	0 521202018	2 269311241
DURAÇÃO	1 248519029	0.561403333	2 223925217	0.033827946	0 101980465	2 39505759	0 101980465	2 39505759
	-93 60739999	44 05605239	-2 124734172	0.041960894	-183 5818623	-3 632937652	-183 5818673	-3 632937652
Ε ΕΔΙΧΔ + 14	-22.81584558	36,98913113	-0.61682567	0.54200298	-98 35772927	52,72603812	-98 35772927	52,72603812
É FAIXA ETÁRIA + 16	47.05523717	42.82781283	1.098707453	0.280635417	-40.41082534	134.5212997	-40.41082534	134,5212997
		.,	,		,		,	

Figura 37 Resultados da RLM do modelo 1

Modelo 2 – Considerando todas as variáveis exceto faixa etária

O Modelo 2 considera a relação entre a variável dependente e todas as variáveis independentes apresentadas, exceto a faixa etária, visto que esta não apresentou alta significância anteriormente (valor-P). Observa-se que este modelo já foi realizado (RLM2) e

apresentou **R² ajustado de 0,3**. Ou seja, as variáveis independentes **explicam juntas 30%** da variável dependente. A Figura 38 apresenta os resultados desta RLM, a qual considera todas as variáveis, exceto a faixa etária.

Arquivo	Página Inicial	Inserir I	ayout da Página Fó	rmulas Dado	s Revisão	Exibir 📿 🤇	O que você deseja faz	er			
ß	Calibri	+ 11	• A* A* ==	»· •	Quebrar Texto A	utomaticamente	Geral	•			
Colar		• 🖽 •	🏠 • 🛕 • 🔳 🗏 🗄		Mesclar e Centra	lizar =	🖓 - % 000 🏌	8 40 Formataçã	io Formatar.com	o Estilos de Inseri	r Excluir Fo
Área de Tr	ansf 🖙	Fonte	5	Alir	hamento		Número	ra .	Estilo	Celuiu	Células
M20	* : ×	€									
WILD			1								
1	A		В	C	D	E	F	G	н	1	J
1 RESU	JMO DOS RESULTAI	DOS									
2	F-1-1			-							
3	ESTATIS	tica ae regr	essao								
4 R mu	untipio		0,00048575	12							
6 R-0	uadrado aiustado		0.30855861	6							
7 Erro	padrão		96,1840023	4							
8 Obs	ervações		3	6							
9											
10 ANC	AVA										
11			gl	SQ	MQ	F	F de significação				
12 Regr	ressão			3 172250,159	2 57416,71973	6,206298902	0,001910885				
13 Resi	duo		3	2 296043,593	8 9251,362306						
14 Tota	il.		3	468293,75	3						
15											
16			Coeficientes	Erro padrão	Stat t	valor-P	95% inferiores	95% superiores	Inferior 95,0%	Superior 95,0%	
17 Inte	rseção		-29,4484207	8 79,017709	1 -0,372681278	0,711842811	-190,4022272	131,5053856	-190,4022272	131,5053856	
18 GAS	TO MILHÕES		1,28565414	9 0,42681782	9 3,012184734	0,005035474	0,416254681	2,155053616	0,416254681	2,155053616	
19 DUR	AÇÃO		1,29760469	0,5584061	3 2,32376513	0,026647909	0,160168628	2,435040757	0,160168628	2,435040757	
20 LAN	ÇAMENTO		-93,8356853	7 44,315892	2 -2,117427422	0,042088364	-184,1042038	-3,567166901	-184,1042038	-3,567166901	
21											
22											
23	DIALI	PLM 2		CORRELACÃO	FUNATE	0		: [4]			

Figura 38. Resultados da RLM do modelo 2

Modelo 3 – Considerando todas as variáveis, exceto gasto

O Modelo 3 considera a relação entre a variável dependente e todas as variáveis independentes apresentadas, exceto gasto. Este modelo será aplicado, uma vez que se observou certa multicolinearidade entre as variáveis gasto e lançamento. Assim iremos retirá-las individualmente de cada modelo, a fim de analisar as possíveis diferenças no R² ajustado. A Figura 39 apresenta a RLM sendo gerada e a Figura 40 apresenta os resultados desta RLM, a qual considera todas as variáveis, exceto gasto.

Arquive	Página Inicial Inserir Lay	rout da Página	Fórmulas Dad	os Revisão	Exibir 🛛	O que você deseja faze	f			Entrar 🞗 C
Obter Dados Externos *	Nova Consulta - Co Fontes Recentes	Atualizar Tudo	onexões Ž↓ opriedades litar Links	Z A Z A Classificar Filt	ro Kangado	Texto para Colunas	Teste de P Hipóteses	lanilha de Previsão	 +∃ Análise de Dados ar∃ 2, Solver 	
	Obter e Transformar	Conex	5es	Classificar	e Filtrar	Ferramentas de Dado	s Previsi	Regressão		? ×
H4	\bullet : \times \checkmark f_x							Entrada		ОК
	А	В	с	D	E	F	G	Intervalo <u>Y</u> de entrada:	SBS1:SBS37	Cancelar
1	TÍTULO	FATURAMENTO	GASTO MILHÕE	s duração	LANÇAMENTO	É FAIXA + 14 É FAIX	A ETÁRIA + 16	Intervalo X de entrada:	SDS1:SGS37	
2 A1		81,843	18,5	137	0	1	0	Rótulos	Constante é zero	Ajuda
3 A2		194,125	140	144	1	1	0	Nível de confiança	95 %	
4 A3		147,54	50	138	1	0	0			
5 B1		75,6	72	177	1	1	0	Opções de saída	202	
6 C1		12,006	0,3	105	1	0	0	O Intervalo de saída:		
7 C2		100,853	90	153	1	1	0	Noga planilha:		-
8 D1		67,155	104	112	1	1	0	🔿 Nova pasta de <u>t</u> rab	alho	
9 D2		140,424	75	120	1	0	1	Resíduos		
10 E1		68,75	55	129	1	0	0	Residuos	Plotar residuos	
11 F1		329,691	55	142	1	1	0	C Residgos padroniza	auos rigtar ajuste de limita	-
12 G1		217,631	22	128	1	0	0	Probabilidade normal		
13 G2		198,571	3,9	222	0	0	0	El Plotagem de proba	bilidade normal	_
14 G3		138,339	10	126	1	1	0			
15 G4		181,28	6	110	0	0	0			
16 H1		47	0,3	93	0	0	1			DUDAÇÃO
17 H2		19,819	70	95	1	0	0			DUKAÇAU
18 11		72,219	17	100	1	0	1	250		
19 12		306,124	75	142	1	1	0		•	4
20 13		197,171	39	127	0	0	0	200		
21 14		260	12	124	0	0	1		• •	
22 M1		250,147	90	98	1	0	0	150		
23 M2		20,1	45	117	1	1	0	¥		

Figura 39. Gerando a RLM do modelo 3

Arquiv	/o Página Inicial Inseri	r Layout da Página Fórm	ulas Dados	Revisão) Exibir	♀ O que você desej	a fazer		
Obter D Extern	Dados Nova Nova Consulta + Co Fontes R Obter e Transform	Consultas a ecentes mar	dades inks	ssificar Fi Classifica	Itro Kapic Marka Share Marka S	ado Ferramentas de	Ino PC T T T Dados	este de Pla póteses * Pr Previsão	nilha de evisão Estrutura
L11	• : × v	fx							
- 21	^	P	c	D	F	F	6	ц	
1 85		U	2	U	-		U		
2	SOMO DOS RESOLINDOS								
3	Estatística d	e rearessão							
4 R	múltiplo	0.462074531							
5 R-	Quadrado	0,213512873							
6 R-	quadrado ajustado	0,112030663							
7 Err	ro padrão	108,9994094	N.						
8 0	bservações	36							
9		()							
10 AN	NOVA								
11		gl	SQ	MQ	F	F de significação			
12 Re	egressão	4	99986,7444	24996,69	2,10394386	0,104145831			
13 Re	esíduo	31	368307,0086	11880,87					
14 To	otal	35	468293,753						
15									
16		Coeficientes	Erro padrão	Stat t	valor-P	95% inferiores	% superior	ferior 95,0	perior 95,0%
17 Int	terseção	-61,52512499	89,02011973	-0,69114	0,49462269	-243,0828562	120,0326	-243,083	120,0326
18 DL	JRAÇÃO	1,633556172	0,628315795	2,599897	0,014156522	0,352097659	2,915015	0,352098	2,915015
19 LA	NÇAMENTO	-24,70108875	44,25229631	-0,55819	0,58072564	-114,9542421	65,55206	-114,954	65,55206
20 É F	FAIXA + 14	-12,59723202	42,19315205	-0,29856	0,767267187	-98,65073297	73,45627	-98,6507	73,45627
21 É F	FAIXA ETÁRIA + 16	35,08513508	48,84912056	0,718235	0,477991152	-64,54330314	134,7136	-64,5433	134,7136
22									
23				MODELO					22
		2 KLM 3 MATRIZ DE CO	JKKELAÇAO	WODELO	5 FILMES	(+)		: 4	

Figura 40. Resultados da RLM do modelo 3

Observa-se que neste modelo o R² ajustado foi de 0,11. Ou seja, as variáveis independentes explicam juntas 11% da variável dependente.

Modelo 4 – Considerando todas as variáveis, exceto lançamento

O Modelo 4 considera a relação entre a variável dependente e todas as variáveis independentes apresentadas, exceto lançamento. Este modelo será aplicado, uma vez que se observou certa multicolinearidade entre as variáveis gasto e lançamento. Assim iremos retirá-las individualmente de cada modelo, a fim de analisar as possíveis diferenças no R² ajustado. Para realizar a regressão, é necessário modificar o local das colunas para que todas as variáveis independentes permaneçam lado a lado, conforme Figura 41. A Figura 42 demonstra a RLM do Modelo 4 sendo gerada e a Figura 43 apresenta os resultados desta RLM, considerando todas as variáveis, exceto lançamento.

Arc	uīvo	Página Inicial	Inserir La	yout da Página	Fórmulas	Dados	Revisã	0	Exibi	r Q) que voc	ê deseja fa:	zer		
Obte Ext	er Dados ernos *	Nova Consulta - Co Obter e	Mostrar Consulta Da Tabela Fontes Recentes Transformar	Atualizar Tudo - Conexi	onexões opriedades litar Links ões	∆↓ [] Z↓ Cla	Z A A Z ssificar F Classific	iltro ar e F	Te Li Te Ri Te A	mpar eaplicar vançado	Texto pa Coluna Ferramer	ara as = + 1	re Rest Hipóte	de Planilha de ses Previsão Previsão	Agrupar Besagru Bubtotal Estrutura de
E1		>	< 🗸 f _x						Ĩ.			_		<u> </u>	
		A		В	с		D		E		:	G		н	1
1		TÍTULO)	FATURAMENTO	GASTO MIL	HÕES	DURAÇÃO			É FAIX	A + 14 É	FAIXA ET	ÁRIA + 16	LANCAMENTO	
2	A1			81,843	18,5		137				1		0	0	
3	A2			194,125	140		144	1	l.	0	1		0	1	250
4	A3			147,54	50		138	1			0		C	1	
5	B1			75,6	72		177			- e	1		0	1	200
6	C1			12,006	0,3		105			-	0		0	1	200
7	C2			100,853	90		153			2	1		0	1	150
8	D1			67,155	104		112		M		1		C	1	150
9	D2			140,424	75		120		M	2	0		1	1	100
10	E1			68,75	55		129		V		0		0	1	100
11	F1			329,691	55		142			2	1		0	1	
12	G1			217,631	22		128		Λ		0		0	1	50
13	G2			198,571	3,9		222		Λ	2	0		0	C	
14	G3			138,339	10		126		\square		1		C	1	0
15	G4			181,28	6		110		\square	-	0		0	C	6
16	H1			47	0,3		93				0		1	0	
17	H2			19,819	70		95				0		0	1	250
18	11			72,219	17		100				0		1	1	250
19	12			306,124	75		142			2	1		0	1	200
20	13			197,171	39		127				0		C	C	200
21	14			260	12		124			2	0		1	0	
22	M1			250,147	90		98				0		C	1	150
23	M2			20,1	45		117			-	1		0	1	

Figura 41. Reorganização das colunas para RLM do modelo 4

Anguise	Página Inicial Inserir Laj	yout da Página	Fórmulas Dado	s Revisão	Exibir 🛛	O que você deseja fazer	•		Entrar 🤱 Con
Obter Dados Externos *	Nova Consulta - Consulta	is Q Co Atualizar Tudo - Q Ed	onexões opriedades litar Links	assificar Filtr	 Limpar Reaplicar Avançado 	Texto para Colunas	Teste de Pl Hipóteses - T	Análise de Dados Construita de Previsão Carevisão	1
	Obter e Transformar	Conexi	ões 🛛	Classificar e	e Filtrar	Ferramentas de Dados	Previsã	Regressão	? ×
C1	\bullet : $\times \checkmark f_x$							Entrada	OK
	A	B	с	D	E	F	G	Intervalo Y de entrada: SB\$1:SB\$37	Cancelar
1	τίτυιο	FATURAMENTO	GASTO MILHÕES	DURAÇÃO É	FAIXA + 14 É F	AIXA ETÁRIA + 16	NÇAMENTO	Intervalo X de entrada: SCS1:SFS37	
2 A1		81,843	18,5	137	1	0	0	Rótulos Constante é zero	Ajuda
3 A2		194,125	140	144	1	0	1	Nível de confianca 95 %	
4 A3		147,54	50	138	0	0	1		
5 B1		75,6	72	177	1	0	1	Opções de saída	
6 C1		12,006	0,3	105	0	0	1	🔿 Intervalo de saída:	
7 C2		100,853	90	153	1	0	1	Noya planilha:	
8 D1		67,155	104	112	1	0	1	🔿 Nova pasta de <u>t</u> rabalho	
9 D2		140,424	75	120	0	1	1	Resíduos Linhas de Grac	le PrincipaisEixoHorizo
10 E1		68,75	55	129	0	0	1	Resiguos Plotar residuos	
11 F1		329,691	55	142	1	0	1	Resíduos padronizados Plotar ajuste de linha	
12 G1		217,631	22	128	0	0	1	Probabilidade normal	
13 G2		198,571	3,9	222	0	0	0	Plotagem de probabilidade normal	
14 G3		138,339	10	126	1	0	1		
15 G4		181,28	6	110	0	0	0		
16 H1		47	0,3	93	0	1	0	DU	RACÃO
17 H2		19,819	70	95	0	0	1	350	
18 11		72,219	17	100	0	1	1	250	
19 12		306,124	75	142	1	0	1		
20 13		197,171	39	127	0	0	0	200	
21 14		260	12	124	0	1	0	• • •	
22 M1		250,147	90	98	0	0	1	150	•
23 M2		20,1	45	117	1	0	1		

Figura 42.Gerando a RLM do modelo 4

Arquiv	o Página Inicial Inser	ir Layout da Página Fórm	ulas Dados	Revisão	Exibir Q () que você deseja fazi	:F		
Obter D Externe	ados S5 * Consulta + Co Fontes F Obter e Transfor	Consultas la Atualizar Mar Conexões	dades inks Ž↓ Ž	ssificar Filtro	K Limpar Reaplicar Avançado Itrar	Texto para Colunas S + C	Teste de Pla Hipóteses + P Previsão	nilha de revisão	唱 Agrupar マ 唱 Desagrupar マ 目 Subtotal strutura de Tópicos
к10		fx							
	A	В	с	D	E	F	G	н	1 1
1 RE	SUMO DOS RESULTADOS								
2	Mile State and and the	an ann an tha ann an th							
3	Estatística c	le regressão							
4 Rr	nultiplo	0,576075004							
5 R-0	Quadrado	0,331862411							
7 Err	o nadrão	100 4641903							
8 Ob	servações	36							
9									
10 AN	IOVA								
11		gl	SQ	MQ	F	F de significação			
12 Re	gressão	4	155409,0937	38852,27343	3,849407251	0,011853958			
13 Re	síduo	31	312884,6593	10093,05352					
14 To	tal	35	468293,753						
15									-
16		Coeficientes	Erro padrão	Stat t	valor-P	95% inferiores	95% superiores	ferior 95,0	perior 95,0%
17 Int	erseção ~	-88,29461776	77,53374636	-1,138789519	0,263515042	-246,425736	69,83650049	-246,426	69,8365
18 GA	STO MILHOES	0,958988281	0,396225355	2,420310229	0,021559588	0,15088134	1,767095221	0,150881	1,767095
19 DU	RAÇAO	1,341395603	0,590574154	2,271341531	0,030219884	0,136911675	2,545879532	0,136912	2,54588
20 EF	AIXA + 14	-29,62041401	38,8830323	-0,761782512	0,451946827	-108,9228812	49,6820532	-108,923	49,68205
21 21	AIXA ETAKIA + 16	41,05032259	45,09185989	0,9103/1022	0,30965149	-50,91513198	133,0157772	-50,9151	133,0158
22									
1	RLM 1 RLM	2 RLM 3 MATRIZ DE CO	ORRELAÇÃO	MODELO 3	MODELO 4	FILMES (F) : [4]		54
Pronto						-			

Figura 43. Resultados da RLM do modelo 4

Observa-se que neste modelo o R² ajustado foi de 0,24. Ou seja, as variáveis independentes explicam juntas 24% da variável dependente.

Modelo 5 – Considerando apenas a variável duração

O **Modelo 5** considera a relação entre a variável dependente e a variável independente de duração, somente. Assim, podemos descobrir, por meio de uma RLS, o quanto essa variável independente, sozinha, explica o faturamento dos filmes. A Figura 44 demonstra a RLS do Modelo 5 sendo gerada e a Figura 45 apresenta os resultados desta RLS, considerando apenas a variável duração.

	Página Inicial Inserir	Layout da Página	Formulas Dad	os Revisão	Exibir 💱	O que você deseja fac	1et			Entrar
Dibter Dado Externos *	Mostriai Consu Nova Consulta - Co Fontes Recents	itas Do E p Atualizar Tudo - D to	onesilles 21 opriedades 51 6 Star Linka	Classificar Fit	tro	Tests para	NG Teste de P Hipóteses	Agrup Sanifha de Strevalio	ar - 13 HAnálise de Dados rupor - 13 2. Solver tal	
	Obter e Transformar	Conex	ões -	Classifica	r e Filtrar	Ferramentas de Da	dos Previs	Regressão		? X
D1	• 1 × 4 4	al						Entrada		OK
di i	A	В	c	D	E	F	G	Intervalo <u>Y</u> de entr	ida: \$8\$1:\$8\$37	Cancelar
3	τίτυιο	FATURAMENTO	GASTO MILHÕE	S DURAÇÃO	É FAIXA + 14 É	FAIXA ETÁRIA + 16	LANCAMENTO	Intervalo X de entr	ada: \$D\$1:\$D\$37	
Al		81,843	18.5	137	1	0	0	Di Dátular	Coortente é suro	Ajuda
A2		194,125	140	144	1	0	1	Nicel de confia	nra 05 %	
A3		147,54	50	138	0	0	1	E Bier de contra	100 IS	
81		75,6	72	177	1	0	1	Opções de saída		
C1		12,006	0,3	105	0	0	1	O Intervalo de sai	da:	
C2		100,853	90	153	1	0	1	Noya planiiha;		
D1		67,155	104	112	1	0	1	O Nova pasta de 1	rabalho	
02		140,424	75	120	0	1	1	Residuos		
E1		68,75	55	129	0	0	1	Resíduos	Plotar residuos	
F1		329,691	55	142	1	0	1	Residuos padro	nizados Pl <u>o</u> tar ajuste de linha	
G1		217,631	22	128	0	0	1	Probabilidade non	nal	
GZ		198,571	3,9	222	0	0	0	Plotagem de pr	obabilidade normal	
G3		138,339	10	126	1	0	1			
G4		181,28	6	110	0	0	0	177		
H1		47	0,3	93	0	1	0		D	IRACÃO
H2		19,819	70	95	0	0	1			
11		72,219	17	100	0	1	1	250		
12		306,124	75	142	1	0	1	1225	•	
13		197,171	39	127	0	0	0	200		
14		260	12	124	0	1	0		• •	
M1		250,147	90	98	0	0	1	150	Co Suntana	
1 M2		20,1	45	117	1	0	1			

Figura 44. Gerando a RLM do modelo 5

Arquivo	Página Inicial	Inserir	Layout da Página Fórm	nulas Dados	Revisão	Exibir Q (O que vocé deseja faze	:r		
Obter Da Externo	dos s * Consulta * Co	Mostrar Cons Da Tabela Fontes Recent	ultas Atualizar Tudo * Conexi E Proprie Editar L	inks 2↓ Cla	ssificar Filtro	Climpar Reaplicar	Texto para Colunas	Teste de Pla Hipóteses • P	nilha de revisão	upar • • Ar agrupar • • ?, So total
	Obter e	Iransformar	Conexoes		Classificar e F	itrar	Ferramentas de Dadi	os Previsad	Estrutu	ra de lopicos 🖓
17	* : >	< - fs	¢							
-	А		в	с	D	Е	F	G	н	1
1 RES	UMO DOS RESULT	ADOS				8015		300		
2										
3	Esta	tística de rea	aressão							
4 R m	últiplo		0,421876249							
5 R-0	uadrado		0,17797957							
6 R-q	uadrado ajustado		0,153802498							
7 Erro	padrão		106,4047534	3						
8 Obs	servações		36							
9										
10 AN	OVA									
11			gl	SQ	MQ	F	F de significação			
12 Reg	ressão		1	83346,72061	83346,72061	7,361502395	0,010382881			
13 Res	íduo		34	384947,0324	11321,97154					
14 Tot	al		35	468293,753						
15										
16			Coeficientes	Erro padrão	Stat t	valor-P	95% inferiores	95% superiores	Inferior 95,0%	Superior 95,0%
17 Inte	erseção		-78,15759208	81,60833468	-0,957715807	0,344967075	-244,0056821	87,69049798	-244,0056821	87,69049798
18 DU	RAÇÃO		1,635965373	0,602963299	2,713208874	0,010382881	0,41059652	2,861334227	0,41059652	2,861334227
19										
20										
21										
22 RES	ULTADOS DE RESÍ	DUOS								
23	1 100 million	L Santan I	the second s	ter I teoperateuro						
4	• RLM 2	RLM 3	MATRIZ DE CORRELAÇA	O MODELC	3 MODELO	MODE	O 5 FILMES	(+) : [•]		

Figura 45. Resultados da RLM do modelo 5

Observa-se que neste modelo o \mathbb{R}^2 ajustado foi de 0,15. Ou seja, a variável de duração explica sozinha 15% da variável dependente. Esse é um \mathbb{R}^2 alto, quando comparado aos outros modelos gerados. No Modelo 1, em que todas as variáveis foram inseridas, o valor do \mathbb{R}^2 foi de 0,32, ou seja, a duração é responsável por explicar metade do \mathbb{R}^2 , sozinha, e portanto, é uma variável muito importante.

Outros modelos também poderiam ser testados, e a escolha dependerá do objetivo de cada pesquisador.

Utilizando o SPSS

O SPSS é um software criado pela IBM para análises estatísticas nas Ciências Sociais. Não é nosso objetivo central ensinar a usá-lo, mas algumas dicas são necessárias para darmos continuidade às nossas análises.

Se você já tem uma base de dados digitada em outro programa, ou mesmo no próprio SPSS, pode importá-la clicando em:

Arquivo > Abrir > Dados. Na caixa de diálogo, em "Arquivos do tipo", selecione a extensão em que o arquivo original está (SPSS, Excel, SAS ou texto) > Abrir, conforme mostram as Figuras 46 e 47.

SPSS Statistics	Arquivo Editar Visualizar Dados Tra	ansformar	Analisar Ma	rketing diret	o Gráficos
	Novo	•	IntodeDados1] - I	Editor de dad	os do IBM SPSS
	Abrir	×	🛅 Dados		
	Abrir Base de Dados Ler dados do texto Ler Dados do Cognos	*	🐨 Syntaxe 😨 Saída 資 Script		14 9 9
var	🖾 Fechar	¥W.	var	var	var v
1	🔚 Salvar	ЖS			
2	Salvar como				
3	🖏 Salvar todos os dados				
4	Exportar	•			
5	Marcar arquivo como somente leitura				
7	Renomear conjunto de dados Exibir informações do arquivo de dados	•			
9	Armazenar dados em cache O Parar o processador R Alternar servidor	₩.			
.0	Repositório	•			
1	Nicualização prévia da improceão				
2	A Imprimir	9#P			
3		8.0.1			1
4	Diálogo de Boas-vindas				
5	Dados usados recentemente				
6	Arquivos usados recentemente				
7					
8		-			

Figura 46. Abrir ou importar arquivo no SPSS

SPSS Statistics	Arquivo	Editar Vis	sualizar	Dados	Transformar	Analisar	Marketing direto	Gráficos	Utilitários	Janela	Ajuda
				Sen	título2 [Conju	ntodeDados	1] - Editor de dados	do IBM SPS	S Statistics		
			1	P	M 🕷	¥	4	₿ 🌏 ¶	ARG .		
var	var	var	var	V	if yar	Var	var	var	var	var	var
1						4	Abrir dados		12		
•			and a state of the				And a second				
4		Procui	rar em:					i.			
5		Apr	plications	tmp							
6		📃 📄 bin		Users							
7		cor	es	usr							
8		etc		Volume	5						
9		hor	ne								
10	_	Lib	rary								
11		Net	twork								
12		priv	vate								
13		sbi	n taun								
14		Sys	tern							70.	
15		Nome	do arquivo	D:						A	brir
16		Armula	as da tina			5.00.0 0				6	olar
17		Arquiv	os do tipo	SPS	S Statistics (*.s	av)				C	orar
18		Codifie	ação:	SPSS	Statistics (".sa	npactado (*.	zsav)			Car	ncelar
19				Porta	átil (*.por)		1999 P			A	iuda
20			Minimizar	larg Exce	l (".xls, ".xlsx,	*.xlsm)					
21				Lotu Sylk	s (*.w*) (*.slk)						

Figura 47. Selecionando a extensão do arquivo

No nosso caso, como já realizamos a mesma análise no Excel, apenas importaremos os dados. Mas você também pode digitá-los direto no SPSS, caso desejar (Figura 48).

rquivo	Editar	Visualiza	r <u>D</u> ados	Transt	formar	Analisar	Marketing (
			, <u>r</u>	24		╆ ╡	
5:							
	Ť	ÍTULO	FATURAME	NTO	GAST	MILHOES	DURAÇÃO
1	A1			81,84		18,50	137,0
2	A2		1	94,13		140,00	144,0
3	A3		1	47,54		50,00	138,0
4	B1			75,60		72,00	177,0
5	C1			12,01		,30	105,0
6	C2		1	00,85		90,00	153,0
7	D1			67,16		104,00	112,0
8	D2		1	40,42		75,00	120,0
9	E1			68,75		55,00	129,0
10	F1		3	29,69		55,00	142,0
11	G1		2	17,63		22,00	128,0
12	G2		1	98,57		3,90	222,0
13	G3		1	38,34		10,00	126,0
14	G4		1	81,28		6,00	110,0
15	H1		33	47,00		,30	93,0
16	H2			19,82		70,00	95,0
17	11			72,22		17,00	100,0
18	12		3	06,12		75,00	142,0
19	13		1	97,17		39,00	127,0
20	14		2	60,00		12,00	124,0
21	M1		2	50,15		90,00	98,0
22	M2			20,10		45,00	117,0
23	P1		1	07,93		8,00	154,0
	4						No. of Concession, Name
Viewaliz	acão do c	Indon Vis	veh očastileu	ariável			

Figura 48. Digitando os dados no SPSS

Arquivo	Editar	Visualizar	Dados <u>T</u> rar	sformar	Analisar	Marketing di
	H () 🛄			*	
15 :						
	Ť	ÍTULO	FATURAMENTO	GAST	OMILHOES	DURAÇAO
1	A1		81,8	4	18,50	7,00
2	A2	_	194.1	3	140.00	200
3	A3		VARIÁVEL		VARIÁ	VEIS
4	B1	D	FPENDENTE		INDEPEN	DENTES
5 C1						
6	C2		100,8	5	90,00	153,00
7	D1		67,1	6	104.00	112,00
8	D2		140,4	2	75,00	120,00
9	E1		68,7	5	55,00	129,00
10	F1		329,6	9	55,00	142,00
11	G1		217,6	3	22,00	128,00
12	G2		198,5	7	3,90	222,00
13	G3		138,3	4	10,00	126,00
14	G4		181,2	8	6,00	110,00
15	H1		47,0	0	,30	93,00
16	H2		19,8	2	70.00	95,00
17	11		72,2	2	17,00	100,00
18	12		306,1	2	75,00	142,00
19	13		197,1	7	39,00	127,00
20	14		260,0	0	12,00	124,00
21	M1		250,1	5	90,00	98,00
22	M2		20,1	0	45,00	117,00
23	P1		107,9	3	8,00	154,00
	4					and a state of
Visualiz	ação de d	tados Visu	alização da variáv	el		

Passo 1: Definição das Variáveis (Figura 49)

Figura 49. Definição das variáveis no SPSS

Passo 2: Desenho do gráfico de dispersão

Para gerar o gráfico de dispersão, é necessário seguir o passo a passo descrito a seguir (Figura 50): Gráficos > Construtor de Gráfico > Ok.

	-		(12)(12)		-	-	
			· F 🎥 🗐	N.	Constr	rutor de gráfico.	
					Seleto	r de modelo de	tabelas de grafic
	TÍTULO	FATURAMENTO	GASTOMILHOES	DURAÇÃO	LAN official	arar subgrupos	
1	A1	81,84	18,50	137,00	Gratico	os de variavei d	le Regressao
2	A2	194,13	140,00	144,00	Caixas	s de diálogo leg	jadas
3	A3	147,54	50,00	138,00	1,00	,00	00, 0
4	B1	75,60	72,00	177,00	1,00	1,00	00, 0
5	C1	12,01	,30	105,00	1,00	00,	00, 0
6	C2	100,85	90,00	153,00	1,00	1,00	00, 0
7	D1	67,16	104,00	112,00	1,00	1,00	00, 0
8	D2	140,42	75,00	120,00	1,00	,00	1,00
9	E1	68,75	55,00	129,00	1,00	,00	00, 0
10	F1	329,69	55,00	142,00	1,00	1,00	00, 0
11	G1	217,63	22,00	128,00	1,00),00	00, 0
12	G2	198,57	3,90	222,00	,00),00	00, 0
13	G3	138,34	10,00	126,00	1,00	1,00	00, 0
14	G4	181,28	6,00	110,00	,00	,00	00,
15	H1	47,00	.30	93,00	,00	,00	1,00
16	H2	19,82	70,00	95,00	1,00),00	00, 0
17	11	72,22	17,00	100,00	1,00	00,	1,00
18	12	306,12	75,00	142,00	1,00	1,00	00,
19	13	197,17	39,00	127,00	,00	,00	00, 0
20	14	260,00	12,00	124,00	,00	,00	1,00
21	M1	250,15	90,00	98,00	1,00	00,00	00, 0
22	M2	20,10	45,00	117,00	1,00	1,00	00, C
23	P1	107,93	8,00	154,00	1,00	,00	1,00

Figura 50. Construtor de gráfico no SPSS

Para escolher o tipo de gráfico é necessário seguir o passo a passo descrito:

• Galeria > Dispersão/Ponto > Primeira opção (Gráfico Disperso Simples).

Depois, precisamos definir os eixos do gráfico. Para isso é necessário arrastar com o mouse a variável *faturamento* para o eixo Y e a variável *gastoemmilhões* para o eixo X > Ok (Figura 51). O gráfico será gerado na tela de saída de dados (Figura 52).



Figura 51. Gerando gráfico no SPSS



Figura 52. Gráfico de Dispersão Faturamento X GastoMilhões

Observa-se que não é possível interpretar a relação entre as duas variáveis apenas com os pontos dispersos no gráfico. Por isso, gera-se a linha de tendência. Para gerar a linha de tendência é necessário clicar duas vezes sob o gráfico e clicar no ícone "adicionar linha de ajuste no total" (Figura 53).



Figura 53. Gerando linha de tendência

É necessário selecionar o tipo de linha que se quer visualizar. Neste caso, o método de ajuste é linear, visto que busca-se descobrir qual relação entre a variável dependente e independente. É necessário desmarcar a opção "anexar rótulo à linha" e clicar em "aplicar", conforme a Figura 54.

ropriedades				×
Tamanho do gráfico	Linhas Linha d	le ajuste Var	riáveis	
Exibir a Método o	aumentos 📄 : de ajuste O Média de Y Elinear Loess & de pontos a Kernel: Epan	Suprimir intero	ceptação	CO
-Intervalo ◎ <u>N</u> ent ◎ Médi ◎ Indiv %: 95	s de confiança — num a dual			
Anexar	rótulo à linha		olicar) Cance	lar Ajuda

Figura 54. Linha de ajuste

Para gerar o gráfico com a variável duração, é necessário criar um novo gráfico a partir dos passos citados, porém deve-se arrastar com o mouse a variável *duração* para o eixo X (Figura 55).



Figura 55. Gráfico de Dispersão Faturamento X Duração

Passo 3: Montagem da equação da RLM

A Figura 56 apresenta a montagem da equação da RLM, conforme o exemplo prático apresentado. A equação é a mesma que montamos no Excel.



Figura 56. Equação da RLM

Passo 4: Rodar a RLM

Visto o gráfico e a equação, podemos iniciar a Regressão Múltipla no SPSS. Na tela principal é necessário selecionar Analisar > Regressão > Linear, conforme Figura 57.

Anglaser Anglaser Marketing direto Gráficos Utilitários Janela Auda 	trab.	ALHO.sav [Conjunto	_de_dados1] - IBM SPS	S Statistics	rs Editor de dados			2011 10	14 V224 14			
Relatórios Relatórios 6: Estatísticas descritivas 1 A1 8: Comparar médias 0 MAIS14 2 A2 1 A1 8: 1.00 2 A2 1 A1 8: 00 2 A2 1 A1 8: 00 4: B1 7: D1 6: C2 100 F1 329; E1 6: C2 100 F1 329; Estafisticar 9: E1 6: C2 100 F1 329; Estafisticar 8: D2 11 G1 21: A2; 13 G3 13 G3 14 G4 6: Previsão 16 H2 17 Malitiplas respostas <	Arquivo	<u>E</u> ditar <u>V</u> isualiz	ar <u>D</u> ados <u>T</u> rans	formar	Analisar Marketing direto	Gráficos	Utilit	ários Ja	anela Ajuda	5		
6: TiTULO FATURAMENTO GASTO Tabelas MAIS14 MAIS16 var var var 1 A1 81.84 Comparar médias Modelo linear geral 1.00 .00					Relatórios Estatísticas descritivas	1		4	\$ ₩	14 0	•	ABG
TÍTULO FATURAMENTO GASTO Comparar médias MAIS14 MAIS16 var var var 1 A1 81.84 Modelo linear geral 1.00 .00<	6:				Ta <u>b</u> elas)						
1 A1 81.84 Modelo linear geral 1.00 .00 2 A2 194,13 Modelos lineares generalizados 1.00 .00 3 A3 147,54 Modelos mistos .00 .00 4 B1 75,60 Correlacionar 1.00 .00 6 C2 100.85 Lgg linear Image: Correlacionar Ima		TÍTULO	FATURAMENTO	GASTO	Comparar médias	,	M.	AIS14	MAIS16	var	var	var
2 A2 194,13 Modelos lineares generalizados 1,00 ,00 3 A3 147,54 Modelos mistos ,00 ,00 ,00 4 B1 75,60 Correlacionar 1,00 ,00 ,00 ,00 5 C1 12,01 Regressão Lgg linear Lgg linear Lgg linear Lgg linear Lgg linear Ecassificar Ecositica Ecosificar Ecositica Ecositica E	1	A1	81,84		Modelo linear geral		8	1,00	,00			
3 A3 147,54 Modelos mistos 0.0 .00 .00 4 B1 75,60 Correlacionar 1.00 .00 .00 5 C1 12,01 Regressão Image: Modelagem Linear Automática Image: Modelagem Linear Automática 6 C2 100,85 Log linear Image: Modelagem Linear Automática Image: Modelagem Linear Automática 7 D1 67,16 Regressão Image: Modelagem Linear Automática Image: Modelagem Linear Automática 9 E1 68,75 Redução de dimensão Image: Modelagem Linear Automática Image: Modelagem Linear Automática 11 G1 217,63 Regução de dimensão Image: Modelagem Linear Image: Modelagem Linear 12 G2 198,67 Testes não paramétricos Image: Modelagem Linear Image: Modelagem Linear Image: Modelagem Linear 15 H1 47,00 Múltiplas respostas Image: Modelagem Linear Image: Modelagem Linear Image: Modelagem Linear 16 H2 19,82 Análise de valor ausente Image: Modelagem Linear Image: Modelagem Linear	2	A2	194,13		Modelos lineares general	izados I		1,00	,00			
4 B1 75,60 Correlacionar 1,00 ,00 5 C1 12,01 Regressão Modelagem Linear Automática 6 C2 100,85 Lg linear Lg linear 7 D1 67,16 Redes neurais Classificar 9 E1 68,75 Redução de dimensão Classificar 9 E1 68,75 Redução de dimensão Classificar 10 F1 329,69 Escala Correiado parciais Classificar 11 G1 217,63 Testes não paramétricos Correiado provisão Correiado provisão 13 G3 138,34 Sobrevivência Mútiplas respostas Mútiplas respostas 16 H2 19,82 Análise de valor ausente Mínimos quadrados de 2 estágios 18 12 306,12 Amostras complexas Escala ideal (CATREG) Escala ideal (CATREG) 19 13 197,17 Simulação 00 00 00 20 14 260,01 Controle de guaidade 00 00 00 00	3	A3	147,54		Modelos mistos			,00	,00			
5 C1 12,01 Regressão Image: Control de gualidade 6 C2 100,85 Log linear Image: Control de gualidade Image: Control de gualidad	4	B1	75,60		Correlacionar			1,00	,00			
6 C2 100,85 Lgg linear Lgg line	5	C1	12,01		Regressão	1	T	Modelar	nem Linear Aut	omática		
7 D1 67,16 Redes neurais 8 D2 140,42 Classificar Classificar 9 E1 68,75 Redução de dimensão Mínimos quadrados parciais 10 F1 329,69 Escala Logística binária Logística binária 11 G1 217,63 Testes não paramétricos Logística multinomial Logística multinomial 12 G2 198,67 Previsão Múltiplas respostas Múltiplas respostas 14 G4 181,28 Sobrevivência Múltiplas respostas Mínimos quadrados de 2 estágios 16 H2 19,82 Análise de valor ausente Mínimos quadrados de 2 estágios 18 12 306,12 Amostras complexas Mínimos quadrados de 2 estágios 19 13 197,17 Simulação Controle de gualidade ,00 1,00 20 14 260,00 Controle de gualidade ,00 1,00 20 14 260,00 Controle de gualidade ,00 1,00 20 1400 00	6	C2	100,85		Log linear	,		R Linear	gennenreernen	ornanoa	_	
8 D2 140,42 9 E1 68,75 10 F1 329,69 11 G1 217,63 12 G2 198,57 13 G3 138,34 14 G4 181,28 15 H1 47,00 Múttiplas respostas Mútiplas respostas 16 H2 19,82 17 I1 72,22 Imputação mútipla Moitras complexas 19 I3 197,17 20 14 260,00 20 44 260,00 20 14 260,00 20 14 260,00 20 14 260,00 20 14 260,00 20 14 260,00 20 14 260,00 20 14 260,00 20 14 260,00 20 14 260,00 20 14 260,00 20 14 260,00	7	D1	67,16		Redes neurais							
9 E1 68,75 Redução de dimensão 10 F1 329,69 Escala Image: Secala Image: Secala <t< td=""><td>8</td><td>D2</td><td>140,42</td><td></td><td>Classificar</td><td>,</td><td></td><td><u>C</u>urva d</td><td>e estimaçao</td><td></td><td></td><td></td></t<>	8	D2	140,42		Classificar	,		<u>C</u> urva d	e estimaçao			
10 F1 329,69 Escala Image: Control of Con	9	E1	68,75		Redução de dimensão	1		🔠 Mínimos	s quadrados p	arciais		
11 G1 217,63 Testes não paramétricos Image: Control of Control	10	F1	329,69		Escala			🛃 Logístic	a binária		_	
12 G2 198,67 Previsão Imorphalanteuros 13 G3 138,34 Previsão Imorphalanteuros 14 G4 181,28 Sobrevivência Imorphalanteuros 15 H1 47,00 Múltiplas respostas Imorphalanteuros 16 H2 19,82 Imputação múltipla Imputação múltipla 17 I1 72,22 Imputação múltipla Imputação múltipla 18 I2 306,12 Amostras complexas Escala ideal (CATREG) 19 I3 197,17 Simulação Imputação Imputação 20 I4 260,00 Controle de gualidade Imputação Imputação 21 M1 250,15 Curva ROC Imputação Imputação	11	G1	217,63		Testes não paramétricos	a i		🛃 Logístic	a <u>m</u> ultinomial.	÷		
13 G3 138,34 1100300 1100300 1100300 1100300 1100300 1100300 1100300 1100300 1100300 1100300 1100300 1100300 1100300 1100300 1100300 1100300 1100300 1100300 11003000 11003000 11003000 <	12	G2	198,57		Previsão			Ordinal.				
14 64 181,28 Sourceventua Image: Constraint of the second	13	G3	138,34		Sobravivância		. U	Probito.				
15 H1 47,00 initializas resposads <	14	G4	181,28					Não line			_	
16 H2 19,82 Analise de valor ausente M Estimação ponderada 17 I1 72,22 Imputação múltipla Minimos quadrados de 2 estágios 18 I2 306,12 Amostras complexas Escala ideal (CATREG) 19 I3 197,17 Simulação 20 I4 260,00 Controle de gualidade 11 250,15 Curva ROC	15	H1	47,00		Multiplas respostas	,			~			
17 I1 72,22 Imputação múltipla Imputação múltip	16	H2	19,82		Analise de valor ausente			Estimac	ao ponderada		_	
18 12 306,12 Amostras complexas Escala ideal (CATREG) 19 13 197,17 It Simulação .00 .00 20 14 260,00 Controle de gualidade .00 1,00 21 M1 250,15 Curva ROC .00 .00	17	1	72,22		Imputação múltipla	,		Mínimos	s quadrados d	e <u>2</u> estágios		_
19 13 197,17 ∰ Simulação ,00 ,00 20 I4 260,00 Controle de gualidade ▶ ,00 1,00 21 M1 250,15 ☑ Curva ROC ,00 ,00	18	12	306,12		Amostras complexas		8	Escala i	ideal (CATREG)		
20 I4 260,00 Controle de gualidade .00 1,00 21 M1 250,15 Image: Curva ROC .00 .00 .00	19	13	197,17		🕎 Simulação		F	,00	,00			
21 M1 250,15 Curva ROC ,00 ,00	20	14	260,00		Controle de gualidade	1		,00	1,00			
00 110 00 10	21	M1	250,15		Curva ROC			,00	,00			
ZZ MZ 20,10 43,00 117,00 1,00 1,00 ,00	22	M2	20,10		45,00 117,00	1,00		1,00	,00			
23 P1 107,93 8,00 154,00 1,00 1,00	23	P1	107,93	_	8,00 154,00	1,00	k	,00	1,00			
4		4		0								
Visualização de dados Visualização da variável	Visualiz	zação de dados Vi	sualização da variável									
Linear IBM SPSS SI	Linear											IBM SPSS Sta

Figura 57. Regressão no SPSS

Na tela que abrirá em seguida, é necessário clicar em FATURAMENTO e na seta azul ao lado da caixa Dependente (1). Em seguida, clicar em GASTOMILHOES e na seta azul da caixa Independente (2). Após, faça o mesmo com a variável DURAÇÃO. Depois, selecione Estatísticas (3). A Figura 58 ilustra este passo a passo.

ta Regressão linear		3 ×
GASTOMILHOES CARÇÃO CURAÇÃO CURAÇÃO CURAÇÃO MAIS10 MAIS14 MAIS16	Dependente: FATURAMENTO Bioco 1 de 1 Anterior Próximo Independente(s): GASTOMILHOES DURAÇÃO Método: Inserir	E <u>s</u> tatísticas Gráficos S <u>a</u> lvar Opções Esti <u>j</u> o <u>B</u> ootstrap
	Variáv <u>e</u> l de seleção: Regra Rótulos de <u>c</u> aso: Ponderação WLS:	

Figura 58. Entrada da variável dependente e independente

Assim, uma nova aba abrirá na qual você pode selecionar os *flags* que desejar. Marcamos aqui todos os *flags* (Figura 59). Nem todos eles são essenciais na análise de RLM e, portanto, não serão todos interpretados nos passos seguintes.

🍓 Regressão linear: Estatísticas		х
Coeficientes de regressão ✓ Estimativas ✓ Intervalos de confiança Nível (%): 95 ✓ Matriz de covariância	 Ajuste do modelo Alteração quadrada de R Descritivos Correlações parciais e de parte Diagnósticos de colinearidade 	
 Residuais ✓ Durbin-Watson ✓ Diagnóstico por caso ● Valores discrepantes no ● Todos os casos 	lado de fora: 3 desvios padrão	,
Continuar	Cancelar Ajuda	

Figura 59. Opções de "Estatísticas"

Os resultados da RLM são obtidos na caixa de saída dos resultados, como nas figuras abaixo. A Figura 61 de estatísticas descritivas nos mostra a média e desvio padrão de cada variável do conjunto de dados, bem como o número de casos que foram avaliados (36 filmes). Por exemplo, sabemos que a média de faturamento destes filmes foi de 137,9716 milhões. Essa tabela não é necessária para interpretar o modelo de regressão, mas é útil como resumo dos dados.

E	Estatísticas descritivas								
	Média	Desvio Padrão	Ν						
FATURAMENTO	137,9716	115,67118	36						
GASTOMILHOES	53,6500	44,70890	36						
DURAÇÃO	132,1111	29,82882	36						

Figura 60. Estatísticas descritivas

Na matriz de correlações, pode-se observar a correlação de *Pearson* entre cada par de variáveis. O item 1 da Figura 62 representa as correlações entre o faturamento e as demais variáveis. Nota-se que sempre a correlação entre uma variável e ela mesma será 1,000. O item 2 diz respeito à significância das correlações entre as variáveis, ou seja, a correlação entre faturamento e gasto em milhões foi de 0,008, representando uma correlação significativa a 95% de confiança, já que 0,008 < 0,05.

Correlações									
		FAT	URAMEN TO	GASTOMILHO ES	DURAÇÃO				
Correlação de Pearson	FATURAMENTO	- 94	1,000	,399	,422				
	GASTOMILHOES	1	,399	1,000	,208				
	DURAÇÃO		,422	,208	1,000				
Sig. (1 extremidade)	FATURAMENTO		*	,008	,005				
	GASTOMILHOES	2	,008	5 %	,111				
	DURAÇÃO		,005	,111	(84				
N	FATURAMENTO		36	36	36				
	GASTOMILHOES		36	36	36				
	DURAÇÃO		36	36	36				

Figura 61. Correlações

A matriz de correlações é extremamente útil para fornecer uma ideia aproximada do relacionamento entre as variáveis independentes e a variável dependente, bem como para o

primeiro exame da multicolinearidade. Se não existir multicolinearidade nos dados, não deve existir valores de correlação substanciais (R>0,90) entre os previsores (FIELD, 2009). Com base na matriz de correlação, o SPSS indica quais variáveis devem ser inseridas no modelo (Figura 63). Neste caso, as duas variáveis independentes em teste possuem relação com a variável dependente.

	Variáveis Inseridas/Removidas ^a									
Modelo	Variáveis inseridas	Variáveis removidas	Método							
1	DURAÇÃO, GASTOMILHO ES ^b		Inserir							
a. Variáv b. Todas	el Dependente: FA as variáveis solic	ATURAMENTO itadas inseridas.								

Figura 62. Variáveis Inseridas/Removidas

O resumo do modelo descreve se o modelo é eficaz. Para interpretar o resumo do modelo, observa-se os dados fornecidos no R, R², R² ajustado, no Erro Padrão da Estimativa e no *Durbin-Watson*, conforme Figura 64. Na figura 64 podemos observar, na coluna denominada R, o valor do coeficiente de correlação múltipla entre os previsores e a saída. As próximas colunas fornecem o valor de R^2 , o R² ajustado e o Erro Padrão de Estimativa que são foram melhor explicados na página 25. A alteração de R² nos diz se essa mudança de valor é significativa. Neste exemplo, a alteração não se mostrou significativa, pois é maior do que 0,05.

A estatística de *Durbin-Watson*, é encontrada na última coluna da tabela. Essa estatística nos informa se a hipótese de independência dos erros é satisfeita. Como uma regra conservadora, o autor sugere que valores menores do que 1 ou maiores do que 3 devem, definitivamente, ser motivos de preocupação (FIELD, 2009). Quanto mais próximo de 2 o valor estiver, melhor.

				Re	sumo do modelo	b				
0		Estatísticas de mu			as de muda	nça				
Modelo R	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa	Alteração de R quadrado	Alteração F	df1	d12	Sig. Alteração F	Durbin- Watson
1	.528 ^a	.279	.236	101,13329	279	6.393	2	33	.005	1,661

Figura 63. Resumo do modelo

Na Figura 65 podemos verificar os valores das análises de variância (ANOVA). Ressalta-se que os significados de cada coluna podem ser revistos nas páginas 25 e 26 desta apostila.

	ANOVAª									
Mod	elo	Soma dos Quadrados	df	Quadrado Médio	Z	Sig.				
1	Regressão	130771,687	2	65385,843	6,393	,005 ^b				
	Resíduo	337522,066	33	10227,941						
	Total	468293,753	35							
a. V	a. Variável Dependente: FATURAMENTO									
b. P	reditores: (Consta	ante), DURAÇÃO, (GASTOMILH	OES						

Figura 64. Análise de variância

A Figura 66 nos apresenta o resultado final da RLM. A constante (b0), como explicado anteriormente, representa o valor da variável dependente caso todas as variáveis independentes fossem 0. Neste caso, B0 é igual a -88,607. A significância de T (valor-p) indica se as variáveis são significativas.

Coeficientes ^a													
Coeficientes não padronizados		Coeficientes padronizados			95,0% Intervalo de Confiança para B		Correlações			Estatísticas de colinearidade			
Mode	lo	в	Erro Padrão	Beta	t	Sig.	Limite inferior	Limite superior	Ordem zero	Parcial	Parte	Tolerância	VIF
1	(Constante)	-88,607	77,717		-1,140	,262	-246,723	69,510					
	GASTOMILHOES	,842	,391	,325	2,153	,039	,046	1,637	,399	,351	,318	,957	1,045
	DURAÇÃO	1,373	.586	,354	2,344	,025	.181	2,565	.422	.378	.346	.957	1.045

Figura 65. Coeficientes

Passo 5: Substituir os dados na equação da RLM

Conforme já apresentado no Excel, este passo consiste em substituir os valores encontrados por meio da RLM na equação original, conforme Figura 64 e Figura 66. Abaixo apresenta-se a substituição de valores na equação da RLM.



É importante também analisar a significância de cada variável independente em relação à variável dependente por meio do "Sig.", conforme exemplo abaixo, exposto na Figura 68. Este dado busca clarear a significância de cada uma das variáveis.



Figura 67. Significância das variáveis

Passo 6: Interpretação dos resultados

Conforme já apresentado na página 29 desta apostila, a interpretação dos resultados é mesma fornecida no modelo de Excel, visto que os mesmos dados foram utilizados, apenas com ferramentas diferentes.

*Variáveis Dummy

Conforme citado anteriormente, observa-se que outras variáveis podem não estar sendo consideradas neste modelo, visto que o valor do erro padrão da estimativa é alto (Figura 64).

Assim, acrescentaremos a variável LANÇAMENTO, conforme foi realizado no modelo do Excel (Figura 69).

quivo <u>E</u>	Editar <u>V</u> isualiz	ar <u>D</u> ados <u>T</u> rans	formar <u>A</u> nalisar	<u>Marketing dire</u>	to <u>G</u> ráficos
a 1:		, r			
1:			40		
	TÍTULO	FATURAMENTO	GASTOMILHOES	DURAÇÃO	LANÇAMENTO
1	A1	81,84	18,50	137,00	,00
2	A2	194,13	140,00	144,00	1,00
3	A3	147,54	50,00	138,00	1,00
4	B1	75,60	72,00	177,00	1,00
5	C1	12,01	,30	105,00	1,00
6	C2	100,85	90,00	153,00	1,00
7	D1	67,16	104,00	112,00	1,0
8	D2	140,42	75,00	120,00	1,0
9	E1	68,75	55,00	129,00	1,0
10	F1	329,69	55,00	142,00	1,0
11	G1	217,63	22,00	128,00	1,0
12	G2	198,57	3,90	222,00	,0
13	G3	138,34	10,00	126,00	1,0
14	G4	181,28	6,00	110,00	,0
15	H1	47,00	.30	93,00	,0
16	H2	19,82	70,00	95,00	1,0
17	11	72,22	17,00	100,00	1,0
18	12	306,12	75,00	142,00	1,0
19	13	197,17	39,00	127,00	,0
20	14	260,00	12,00	124,00	,0
21	M1	250,15	90,00	98,00	1,0
22	M2	20,10	45,00	117,00	1,0
23	P1	107,93	8,00	154,00	1,0
	4			destroated.	

Figura 68. Inclusão da variável lançamento

Antes de realizar novamente a RLM, é necessário transformar as variáveis qualitativas em variáveis *dummy*, atribuindo a elas valores numéricos, conforme apresentado anteriormente. Após, é necessário gerar a **equação básica** de RLM.

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + u$$

O próximo passo consiste em realizar a **RLM** no SPSS. A variável dependente permanece a mesma, porém é necessário selecionar todas as variáveis independentes, inclusive a *dummy* (lançamento). A Figura 70 ilustra este passo a passo.

🕼 Regressão linear		×
CASTOMILHOES CASTOMILHOES CURAÇÃO CURAÇÃO CANÇAMENTO CANÇAMENTO CANCAME	Dependente:	Estatísticas Gráficos S <u>a</u> lvar Opções Esti <u>l</u> o <u>B</u> ootstrap
	Método: Inserir Variáv <u>e</u> l de seleção: Regra Rótulos de <u>c</u> aso: Ponderação WLS:	

Figura 69. Inclusão da variável lançamento

O resumo do modelo, apresentado na Figura 71, indica que o R² ajustado aumentou com relação ao modelo anterior, sem a variável *dummy lançamento*. A Figura 72 mostra os resultados que serão substituídos na equação.

		Resumo do modelo ⁶										
						Estatístic	cas de muda	nça				
Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa	Alteração de R quadrado	Alteração F	df1	df2	Sig. Alteração F	Durbin- Watson		
1	,606ª	,368	,309	96,18400	,368	6,206	3	32	,002	1,761		
a. Preditores: (Constante), DURAÇÃO, LANÇAMENTO, GASTOMILHOES b. Variável Dependente: FATURAMENTO												

Figura 70. Resumo do modelo

Coeficientes não padronizados		Coeficientes padronizados			95,0% Intervalo de Confiança para B		Correlações			Estatísticas de colinearidade			
Mode	10	в	Erro Padrão	Beta	t	Sig.	Limite inferior	Limite superior	Ordem zero	Parcial	Parte	Tolerância	VIF
1	(Constante)	-29,448	79,018		-,373	,712	-190,402	131,505					
	LANÇAMENTO	-93,836	44,316	-,342	-2,117	.042	-184,104	-3,567	-,083	-,351	-,298	.757	1,3
	GASTOMILHOES	1,286	,427	,497	3,012	,005	,416	2,155	,399	,470	,423	,726	1.3
	DURAÇÃO	1,298	,558	.335	2,324	,027	,160	2,435	.422	,380	,327	.953	1.

Figura 71. Coeficientes

Com base nos resultados apontados, é necessário novamente substituir a equação de RLM (Figura 73):



Figura 72. Equação do modelo 1

A interpretação dos resultados é idêntica a do Excel. Conforme citado anteriormente, **observa-se ainda que outras variáveis podem não estar sendo consideradas neste modelo**. Assim, será acrescentada a variável independente **"faixa etária"**, a qual pode ser: livre, maior que 14 anos e maior que 16 anos. Observa-se que esta nova variável, também qualitativa, apresenta três categorias. Portanto, é necessário transformá-la em valores numéricos, conforme apresentado anteriormente. Assim, seguindo o mesmo procedimento sugerido no Excel, gera-se um novo modelo de regressão. O passo a passo para realização desta RLM é apresentado na Figuras 74.

🕼 Regressão linear		×
CASTOMILHOES CASTOMILHOES DURAÇÃO CANÇAMENTO MAIS14 MAIS16	Dependente: Image: Proximo Bloco 1 de 1 Anterior Independente(s): Image: Proximo Independente(s): Image: Proximo Independente(s): Image: Proximo Image: Pro	E <u>s</u> tatísticas Gráficos S <u>a</u> lvar Opções Estilo <u>B</u> ootstrap
ОК	Ponderação WLS: Colar Redefinir Cancelar Ajuda	l

Figura 73. Inserção das variáveis independentes dummy

Observa-se que a variável dependente permanece a mesma, porém na variável independente é necessário selecionar todas as variáveis independentes, inclusive as *dummy* (lançamento e faixa etária). As saídas geradas podem ser vistas nas Figuras 75 e 76.

				Re	esumo do modelo	b _p				
						Estatísti	cas de muda	ança		
Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa	Alteração de R quadrado	Alteração F	df1	df2	Sig. Alteração F	Durbin- Watson
1	,647 ^a	,419	,322	95,21193	,419	4,332	5	30	,004	1,677
a. Predito b. Variávo	ores: (Const el Depender	ante), MAIS16, L nte: FATURAMEI	ANÇAMENTO, DI NTO	JRAÇÃO, MAIS14,	GASTOMILHOES					

Figura 74. Resumo do modelo

						Coefici	ientes ^a						
		Coeficientes nã	o padronizados	Coeficientes padronizados			95,0% Intervalo para	de Confiança a B	с	orrelações		Estatísticas de c	olinearidade
Modelo		в	Erro Padrão	Beta	t	Sig.	Limite inferior	Limite superior	Ordem zero	Parcial	Parte	Tolerância	VIF
1	(Constante)	-30,605	78,336		-,391	,699	-190,589	129,378					
	LANÇAMENTO	-93,607	44,056	-,341	-2,125	,042	-183,582	-3,633	-,083	-,362	-,296	,751	1,332
	GASTOMILHOES	1,395	,428	,539	3,260	,003	,521	2,269	,399	,511	,454	,707	1,414
	DURAÇÃO	1,249	,561	,322	2,224	.034	,102	2,395	,422	,376	,309	,924	1,083
	MAIS14	-22,816	36,989	-,098	-,617	,542	-98,358	52,726	-,076	-,112	-,086	,774	1,291
	MAIS16	47,055	42,828	,172	1,099	.281	-40,411	134,521	,190	,197	,153	.794	1,259

Figura 75. Coeficientes

Substituindo os valores encontrados, teremos a equação apresentada na Figura 77.



Figura 76. Equação

A interpretação dos resultados segue o mesmo padrão do Excel. Assim, com base no modelo proposto, ainda há um **próximo passo** a ser seguido, o qual consiste em verificar a existência ou não de **multicolinearidade** entre as variáveis independentes. Para verificar a multicolinearidade do modelo proposto é necessário realizar a **Matriz de Correlação**, na saída gerada pelo SPSS é a tabela com o nome correlações, a qual será apresentada na Figura 78.

		Ce	orrelações				
		FATURAMEN TO	LANÇAMENT O	GASTOMILHO ES	DURAÇÃO	MAIS14	MAIS16
Correlação de Pearson	FATURAMENTO	1,000	-,083	,399	,422	-,076	,190
	LANÇAMENTO	-,083	1,000	,490	,047	,152	-,036
	GASTOMILHOES	,399	,490	1,000	,208	,207	-,117
	DURAÇÃO	,422	,047	,208	1,000	,109	,084
	MAIS14	-,076	,152	,207	,109	1,000	-,426
	MAIS16	,190	-,036	-,117	,084	-,426	1,000
Sig. (1 extremidade)	FATURAMENTO	a -	,316	,008	,005	,330	,134
	LANÇAMENTO	,316		,001	,392	,188	,418
	GASTOMILHOES	,008	,001		,111	,113	,249
	DURAÇÃO	,005	,392	,111		,263	,312
	MAIS14	,330	,188	,113	,263		,005
	MAIS16	,134	,418	,249	,312	,005	3
Ν	FATURAMENTO	36	36	36	36	36	36
	LANÇAMENTO	36	36	36	36	36	36
	GASTOMILHOES	36	36	36	36	36	36
	DURAÇÃO	36	36	36	36	36	36
	MAIS14	36	36	36	36	36	36
	MAIS16	36	36	36	36	36	36

Figura 77. Correlações

Escolhendo o melhor modelo de regressão...

Modelo 1 – Considerando todas as variáveis

O Modelo 1 considera a relação entre a variável dependente e todas as variáveis independentes apresentadas. Observa-se que este modelo já foi realizado anteriormente e apresentou R² ajustado de 0,32. Ou seja, as cinco variáveis independentes explicam juntas 32% da variável dependente. A Figura 78 apresenta os resultados desta RLM, a qual considera todas as variáveis.

						Estatístic	as de muda	inça		
Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa	Alteração de R quadrado	Alteração F	df1	df2	Sig. Alteração F	Durbin- Watson
1	,647 ^a	,419	,322	95,21193	,419	4,332	5	30	,004	1,677

Figura 78. Resumo do Modelo 1

Modelo 2 – Considerando todas as variáveis exceto faixa etária

O Modelo 2 considera a relação entre a variável dependente e todas as variáveis independentes apresentadas, exceto a faixa etária, visto que esta não apresentou alta significância anteriormente (valor-P). Observa-se que este modelo já foi realizado e apresentou R² ajustado de 0,3. Ou seja, as variáveis independentes explicam juntas 30% da variável dependente. A Figura 79 apresenta os resultados desta RLM, a qual considera todas as variáveis, exceto a faixa etária.

						Estatístic	as de muda	inça		
Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa	Alteração de R quadrado	Alteração F	df1	df2	Sig. Alteração F	Durbin- Watson
1	,606 ^a	,368	,309	96,18400	,368	6,206	3	32	,002	1,761

Figura 79. Resumo do Modelo 2

Modelo 3 - Considerando todas as variáveis, exceto gasto

O Modelo 3 considera a relação entre a variável dependente e todas as variáveis independentes apresentadas, exceto gasto. Este modelo será aplicado, uma vez que se observou certa multicolinearidade entre as variáveis gasto e lançamento. A Figura 80 apresenta a RLM sendo gerada e a Figura 81 apresenta os resultados desta RLM, a qual considera todas as variáveis, exceto gasto.

Observa-se que neste modelo o R² ajustado foi de 0,11. Ou seja, as variáveis independentes explicam juntas 11% da variável dependente.

🕼 Regressão linear		×
CK	Dependente: Image: FATURAMENTO Bloco 1 de 1 Anterior Próximo Independente(s): Image: Duração Image: Duração Image: Duração	Estatísticas Gráficos Salvar Opções Estilo Bootstrap

Figura 80. Modelo 3

				R	esumo do modelo	р ^ь				
			8			Estatístic	as de muda	inça		
Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa	Alteração de R quadrado	Alteração F	df1	df2	Sig. Alteração F	Durbin- Watson
1	,462 ^a	,214	,112	108,99941	,214	2,104	4	31	,104	1,892
a. Preditor	,462 ^a res: (Const	,214 ante), MAIS16, L	,112 ANÇAMENTO, DU	108,99941 JRAÇÃO, MAIS14	,214	2,104	4	31	,104	1,

Figura 81. Resumo do Modelo 3

Modelo 4 – Considerando todas as variáveis, exceto lançamento

O Modelo 4 considera a relação entre a variável dependente e todas as variáveis independentes apresentadas, exceto lançamento. Este modelo será aplicado, uma vez que se observou certa multicolinearidade entre as variáveis gasto e lançamento. A Figura 82 demonstra a RLM do Modelo 4 sendo gerada e a Figura 83 apresenta os resultados desta RLM, considerando todas as variáveis, exceto lançamento. Observa-se que neste modelo o R²

ajustado foi de 0,24. Ou seja, as variáveis independentes **explicam juntas 24%** da variável dependente.

ta Regressão linear	×
Independente: Independente(s): In	Estatísticas Gráficos Salvar Opções Estilo Bootstrap
OK Colar Redefinir Cancelar Ajud	3

Figura 82. Modelo 4

						Estatístic	as de muda	inça		
Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa	Alteração de R quadrado	Alteração F	df1	df2	Sig. Alteração F	Durbin- Watson
1	,576 ^a	,332	,246	100,46419	.332	3,849	4	31	.012	1,57

Figura 83. Resumo do Modelo 4

Modelo 5 - Considerando apenas a variável duração

O **Modelo 5** considera a relação entre a variável dependente e a variável independente de duração, somente. A Figura 84 demonstra a RLM do Modelo 5 sendo gerada e a Figura 85 apresenta os resultados desta RLM, considerando apenas a variável duração. Observa-se que neste modelo o **R² ajustado foi de 0,15**. Ou seja, a variável de duração **explica sozinha 15%** da variável dependente.

🕼 Regressão linear		×
CASTOMILHOES CASTOMILHOES CURAÇÃO CURAÇÃO CANÇAMENTO MAIS14 MAIS16 MAIS16	Dependente: FATURAMENTO Bloco 1 de 1 Anterior Próximo Independente(s): DURAÇÃO	E <u>s</u> tatísticas Gráficos S <u>a</u> lvar <u>O</u> pções Esti <u>l</u> o <u>B</u> ootstrap
	Método: Inserir Variáv <u>e</u> l de seleção: Regra. Rótulos de <u>c</u> aso: Ponderação WLS:	
ОК	Colar <u>R</u> edefinir Cancelar Ajuda	

Figura 84. Modelo 5

				Re	esumo do modelo	b				
Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa	Alteração de R quadrado	Estatistic Alteração F	df1	df2	Sig. Alteração F	Durbin- Watson
1	,422 ^a	.178	.154	106,40475	.178	7.362	1	34	.010	1,865

Figura 85. Resumo do modelo 5

REFERÊNCIAS

ALMEIDA, L. S.; FREIRE, T. Metodologia da investigação em psicologia e educação. 2000.

BELFIORE, P. Estatística aplicada a administração, contabilidade e economia com Excel e SPSS. Rio de Janeiro: Elsevier, 2015.

BRUNI, A. L. SPSS-Guia prático para pesquisadores. São Paulo: Atlas, 2012.

FIELD, A. Descobrindo a estatística usando o SPSS. 2 ed. Porto Alegre: Bookman, 2009.

FONSECA, J. J. S. Metodologia da Pesquisa Científica. 2002.

HAIR, J. F.; BLACK, B.; BABIN, B.; ANDERSON, R. E.; TATHAM, R. L. Análise multivariada de dados. 6. ed. Porto Alegre: Bookman, 2009.