

LISTA DE EXERCÍCIOS 2 – INE 7001 – PROF. MARCELO MENEZES REIS
ANÁLISE BIDIMENSIONAL – GABARITO

- 1)
 a) Calculando os percentuais em relação aos totais de cada COLUNA obtemos: 18,57% de favoráveis entre os Estudantes, 8,57% entre os Servidores e 38,89% entre os Professores. Há diferença entre as categorias, porque a diferença mínima é de 10% e pode ser tão alta quanto 30%.
 b) Calculando os percentuais em relação ao total GERAL (300 entrevistados) obtemos: 8,67% de favoráveis entre os Estudantes, 2% entre os Servidores e 11,67% entre os Professores.

Com base nos percentuais não recomendaria a realização do convênio. Somando esses três percentuais (isso é possível porque foram calculados em relação ao MESMO total), obtêm-se 22,33%, ou seja, o convênio não deve ser realizado porque menos de um quarto da comunidade universitária é favorável à sua realização.

Há diferenças entre os percentuais calculados em relação ao total de cada categoria e aqueles calculados em relação ao total geral. Isso ocorre porque as diferenças entre os percentuais são consideráveis, de 6% (entre os Servidores) a 26% (entre os Professores).

c) Com base na resposta do item a parece haver associação entre a opinião e a categoria do entrevistado: os Servidores apresentam o menor percentual de favoráveis, enquanto os professores apresentam o maior. Calculando o coeficiente de contingência modificado é possível avaliar a força da associação.

Calculando as frequências esperadas para cada célula da tabela:

Est. – Favorável = 31,27 Est. – Desfavorável = 97,53 Est. – Indiferente = 11,20
 Ser. – Favorável = 15,63 Ser. – Desfavorável = 48,77 Ser. – Indiferente = 5,60
 Pro. – Favorável = 20,10 Pro. – Desfavorável = 62,70 Pro. – Indiferente = 7,20

O valor de χ^2 será igual a 31,1357. Substituindo na fórmula do coeficiente de contingência modificado (lembrando que N vale 300, e k vale 3):

$$C^* = \sqrt{\frac{\chi^2}{\chi^2 + N}} \times \sqrt{\frac{k}{k-1}} = \sqrt{\frac{31,1357}{31,1357 + 300}} \times \sqrt{\frac{3}{3-1}} = 0,375554$$

Com este valor de C* podemos classificar a associação como moderada, pois o coeficiente varia está entre 0,3 e 0,7 (valores abaixo de 0,3 são considerados fracos, e acima de 0,7 fortes, lembrando que C* varia de 0 a 1).

- 2) a) Para comparar os destinos precisamos calcular os percentuais em relação aos totais das LINHAS. Os resultados estão nas tabelas abaixo:

Destino	Opinião sobre a companhia aérea				Total
	Ruim	Regular	Boa	Ótima	
Porto Seguro	11,29	57,26	20,97	10,48	100,00
Manaus	24,27	32,04	34,95	8,74	100,00
Pantanal	6,93	33,77	54,98	4,33	100,00
Orlando	12,12	19,19	54,55	14,14	100,00
Total	12,03	36,09	43,63	8,26	100,00

Destino	Opinião sobre o translado no local				Total
	Ruim	Regular	Bom	Ótimo	
Porto Seguro	45,97	29,84	13,71	10,48	100,00
Manaus	35,92	27,18	21,36	15,53	100,00
Pantanal	6,06	21,65	44,16	28,14	100,00
Orlando	8,08	19,19	52,53	20,20	100,00
Total	20,83	24,06	34,65	20,47	100,00

Destino	Opinião sobre a hospedagem				Total
	Ruim	Regular	Boa	Ótima	
Porto Seguro	14,52	54,03	21,77	9,68	100,00
Manaus	19,42	33,98	21,36	25,24	100,00
Pantanal	14,72	19,91	36,80	28,57	100,00
Orlando	7,07	5,05	35,35	52,53	100,00
Total	14,18	27,47	30,34	28,01	100,00

Para tomar a decisão, devem-se somar os percentuais dos valores Ruim e Regular dentre as características:

Porto Seguro teve 68,55% dos clientes considerando a companhia aérea Ruim ou Regular. Também no traslado no local Porto Seguro teve o maior percentual de Ruim ou Regular, 75,81%. E, novamente, Porto Seguro recebeu o maior percentual de classificação Ruim ou Regular, 68,55%, no que se refere à hospedagem. O destino que deve ser retirado do portfólio da agência deve ser Porto Seguro, pois apresentou os maiores percentuais de classificação Ruim ou Regular em todas as características levantadas junto aos clientes.

b) Pelos percentuais calculados parece haver relação entre a opinião dos clientes e o destino da viagem para cada característica: Porto Seguro foi considerado Ruim ou Regular em todas. Para medir a força da associação, vamos calcular o coeficiente de contingência modificado para cada característica.

Para a companhia aérea, calculando as frequências esperadas:

P. Seg. – Ruim = 14,92 P. Seg. – Regular = 44,75 P. Seg. – Boa = 54,10 P. Seg. – Ótima = 10,24

Manaus – Ruim = 12,39 Manaus – Regular = 37,17 Manaus – Boa = 44,94 Manaus – Ótima = 8,51

Pantanal – Ruim = 27,79 Pantanal – Regular = 83,36 Pantanal – Boa = 100,78 Pantanal – Ótima = 19,08

Orlando – Ruim = 11,91 Orlando – Regular = 35,73 Orlando – Boa = 43,19 Orlando – Ótima = 8,18

O valor de χ^2 será igual a 77,0752. Substituindo na fórmula do coeficiente de contingência modificado (lembrando que N vale 557, e k vale 4):

$$C^* = \sqrt{\frac{\chi^2}{\chi^2 + N}} \times \sqrt{\frac{k}{k-1}} = \sqrt{\frac{77,0752}{77,0752 + 557}} \times \sqrt{\frac{4}{4-1}} = 0,402584$$

Com este valor de C^* (0,402) podemos classificar a associação entre destino e opinião sobre a companhia aérea como moderada, pois o coeficiente está entre 0,3 e 0,7 (valores abaixo de 0,3 são considerados fracos, e acima de 0,7 fortes, lembrando que C^* varia de 0 a 1).

Para o traslado local, calculando as frequências esperadas:

P. Seg. – Ruim = 25,82 P. Seg. – Regular = 29,83 P. Seg. – Boa = 42,97 P. Seg. – Ótima = 25,38

Manaus – Ruim = 21,45 Manaus – Regular = 24,78 Manaus – Boa = 35,69 Manaus – Ótima = 21,08

Pantanal – Ruim = 48,11 Pantanal – Regular = 55,57 Pantanal – Boa = 80,04 Pantanal – Ótima = 47,28

Orlando – Ruim = 20,62 Orlando – Regular = 23,82 Orlando – Boa = 34,30 Orlando – Ótima = 20,26

O valor de χ^2 será igual a 134,4919. Substituindo na fórmula do coeficiente de contingência modificado (lembrando que N vale 557, e k vale 4):

$$C^* = \sqrt{\frac{\chi^2}{\chi^2 + N}} \times \sqrt{\frac{k}{k-1}} = \sqrt{\frac{134,4919}{134,4919 + 557}} \times \sqrt{\frac{4}{4-1}} = 0,509242$$

Com este valor de C^* (0,509) podemos classificar a associação entre destino e opinião sobre o traslado local como moderada, pois o coeficiente está entre 0,3 e 0,7, bem no meio do intervalo (valores abaixo de 0,3 são considerados fracos, e acima de 0,7 fortes, lembrando que C^* varia de 0 a 1).

Para a hospedagem, calculando as frequências esperadas:

P. Seg. – Ruim = 17,59 P. Seg. – Regular = 34,06 P. Seg. – Boa = 37,62 P. Seg. – Ótima = 34,73

Manaus – Ruim = 14,61 Manaus – Regular = 28,29 Manaus – Boa = 31,25 Manaus – Ótima = 28,85

Pantanal – Ruim = 32,76 Pantanal – Regular = 63,45 Pantanal – Boa = 70,09 Pantanal – Ótima = 64,70

Orlando – Ruim = 14,04 Orlando – Regular = 27,19 Orlando – Boa = 30,04 Orlando – Ótima = 27,73

O valor de χ^2 será igual a 108,0968. Substituindo na fórmula do coeficiente de contingência modificado (lembrando que N vale 557, e k vale 4):

$$C^* = \sqrt{\frac{\chi^2}{\chi^2 + N}} \times \sqrt{\frac{k}{k-1}} = \sqrt{\frac{108,0968}{108,0968 + 557}} \times \sqrt{\frac{4}{4-1}} = 0,465515$$

Com este valor de C^* (0,465) podemos classificar a associação entre destino e opinião sobre a hospedagem como moderada, pois o coeficiente está entre 0,3 e 0,7, quase no meio do intervalo (valores abaixo de 0,3 são considerados fracos, e acima de 0,7 fortes, lembrando que C^* varia de 0 a 1).

3) Se há correlação entre as variáveis o diagrama de dispersão deve apresentar algum padrão. A correlação é tão mais forte quanto mais próximos os pontos estiverem entre si. A correlação será linear se for razoável ajustar uma reta aos dados (que passe por "entre" eles).

a) Há correlação entre as variáveis porque é possível identificar um padrão no diagrama da dispersão. Parece ser uma correlação forte porque os pontos estão bastante próximos entre si. O modelo linear parece ser apropriado para descrever o relacionamento entre as variáveis porque é possível ajustar uma reta aos dados. Deve ser usada uma reta crescente, porque à medida que aumentam os valores de X aumentam os valores de Y.

b) Há correlação entre as variáveis porque é possível identificar um padrão no diagrama da dispersão. Não parece ser uma correlação muito forte porque os pontos não estão bastante próximos entre si. O modelo linear parece ser apropriado para descrever o relacionamento entre as variáveis porque é possível ajustar uma reta aos dados. Deve ser usada uma reta crescente, porque à medida que aumentam os valores de X aumentam os valores de Y.

c) Há correlação entre as variáveis porque é possível identificar um padrão no diagrama da dispersão. Parece ser uma correlação forte porque os pontos estão bastante próximos entre si. O modelo linear parece ser apropriado para descrever o relacionamento entre as variáveis porque é possível ajustar uma reta aos dados. Deve ser usada uma reta decrescente, porque à medida que aumentam os valores de X diminuem os valores de Y.

d) Há correlação entre as variáveis porque é possível identificar um padrão no diagrama da dispersão. Não parece ser uma correlação muito forte porque os pontos não estão bastante próximos entre si. O modelo linear parece ser apropriado para descrever o relacionamento entre as variáveis porque é possível ajustar uma reta aos dados. Deve ser usada uma reta decrescente, porque à medida que aumentam os valores de X diminuem os valores de Y.

e) Há correlação entre as variáveis porque é possível identificar um padrão no diagrama da dispersão. Parece ser uma correlação forte porque os pontos estão bastante próximos entre si. O modelo linear não parece ser adequado para descrever o relacionamento entre as variáveis, porque parece ser inadequado ajustar uma reta aos dados. Poderia ser usado uma parábola, ou um modelo exponencial para descrever o relacionamento entre as variáveis.

f) Há correlação entre as variáveis porque é possível identificar um padrão no diagrama da dispersão. Parece ser uma correlação forte porque os pontos estão bastante próximos entre si. O modelo linear não é adequado para descrever o relacionamento entre as variáveis, porque parece ser inadequado ajustar uma reta aos dados. Poderia ser usada uma parábola para descrever o relacionamento entre as variáveis.

g) Aparentemente não há correlação entre as variáveis porque não é possível identificar um padrão no diagrama de dispersão. Desta forma não é possível recomendar qualquer modelo para descrever o relacionamento entre as variáveis.

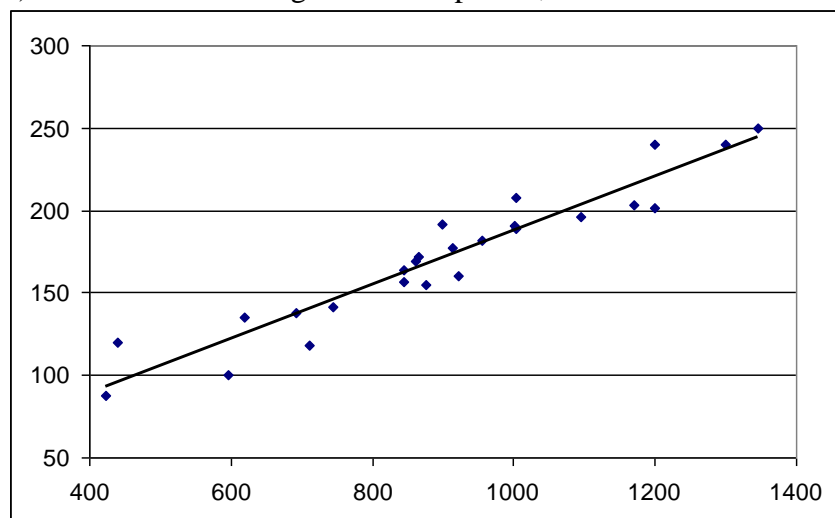
h) Há correlação entre as variáveis porque é possível identificar um padrão no diagrama da dispersão. Parece ser uma correlação forte porque os pontos estão bastante próximos entre si. O modelo linear não é adequado para descrever o relacionamento entre as variáveis, porque parece ser inadequado ajustar uma reta aos dados. Poderia ser usado uma parábola ou um modelo exponencial para descrever o relacionamento entre as variáveis.

4) Basta comparar o valor do r ($|r| > 0,7$ significa correlação linear APARENTEMENTE forte) com o diagrama de dispersão. Se for possível ajustar uma reta aos dados, o modelo linear é recomendado.

- a) O valor do coeficiente de correlação linear corresponde à disposição dos pontos no diagrama: $r = 0,9747$, $|r|$ é maior do que 0,7 indicando forte correlação linear, e é possível ajustar uma reta aos dados, os pontos estão muito próximos, e à medida que os valores de X aumentam, aumentam os de Y. Portanto, o modelo linear (utilizando uma reta crescente) é apropriado para descrever o relacionamento entre as variáveis.
- b) O valor do coeficiente de correlação linear corresponde à disposição dos pontos no diagrama: $r = 0,85339$, $|r|$ é maior do que 0,7 indicando forte correlação linear, é possível ajustar uma reta aos dados, os pontos estão próximos, e à medida que os valores de X aumentam, aumentam os de Y. Portanto, o modelo linear (utilizando uma reta crescente) é apropriado para descrever o relacionamento entre as variáveis.
- c) O valor do coeficiente de correlação linear corresponde à disposição dos pontos no diagrama: $r = -0,984$, $|r| > 0,7$ indicando forte correlação linear, é possível ajustar uma reta aos dados, os pontos estão muito próximos, e à medida que os valores de X aumentam, diminuem os de Y. Portanto, o modelo linear (utilizando uma reta decrescente) é apropriado para descrever o relacionamento entre as variáveis.
- d) O valor do coeficiente de correlação linear corresponde à disposição dos pontos no diagrama: $r = -0,619667$, porque apesar de ser possível ajustar uma reta aos dados os pontos não estão muito próximos, e há dois valores discrepantes, o que contribuiu para $|r|$ ser inferior a 0,7. Talvez o modelo linear não seja o mais apropriado para descrever o relacionamento entre as variáveis, a menos que os dois pontos discrepantes sejam eliminados.
- e) O valor do coeficiente de correlação linear não corresponde à disposição dos pontos no diagrama: $r = 0,851789$, pois parece não ser apropriado ajustar uma reta aos dados, não obstante $|r|$ ser maior do que 0,7, talvez uma parábola seja uma melhor opção.
- f) O valor do coeficiente de correlação linear não corresponde à disposição dos pontos no diagrama: $r = -0,826334$, pois parece não ser apropriado ajustar uma reta aos dados, não obstante $|r|$ ser maior do que 0,7, talvez uma parábola ou equação exponencial seja uma melhor opção.
- g) O valor do coeficiente de correlação linear corresponde à disposição dos pontos no diagrama: r (0,019939) aproximadamente igual a zero, embora haja um padrão nos dados, não é possível ajustar uma reta a eles, tornando o modelo linear inapropriado para descrever o relacionamento entre as variáveis.
- h) O valor do coeficiente de correlação linear corresponde à disposição dos pontos no diagrama: r (0,053803) aproximadamente igual a zero, aparentemente não há padrão nos dados, e não é possível ajustar uma reta a eles, tornando o modelo linear inapropriado para descrever o relacionamento entre as variáveis.

5) a) Número de clientes é a variável independente, e vendas é a variável dependente. Porque o número de clientes pode influenciar o valor das vendas.

b) Construindo um diagrama de dispersão, sendo X o número de clientes e Y as vendas:



Com base no diagrama de dispersão é recomendável a adoção de um modelo linear para descrever o relacionamento entre as variáveis, porque é possível ajustar uma reta aos dados.

c) $r = 0,95$ $r^2 = 0,90$. Com base nesses resultados é recomendável a adoção de um modelo linear para descrever o relacionamento entre as variáveis, porque o modelo linear consegue explicar 90% da variação Y a partir da variação de X.

A conclusão é coerente com a resposta do item b, pois pelo diagrama de dispersão é possível observar que a correlação linear entre as variáveis é forte.

d) b (coeficiente angular) = 0,164897 a (coeficiente linear) = 22,84791.

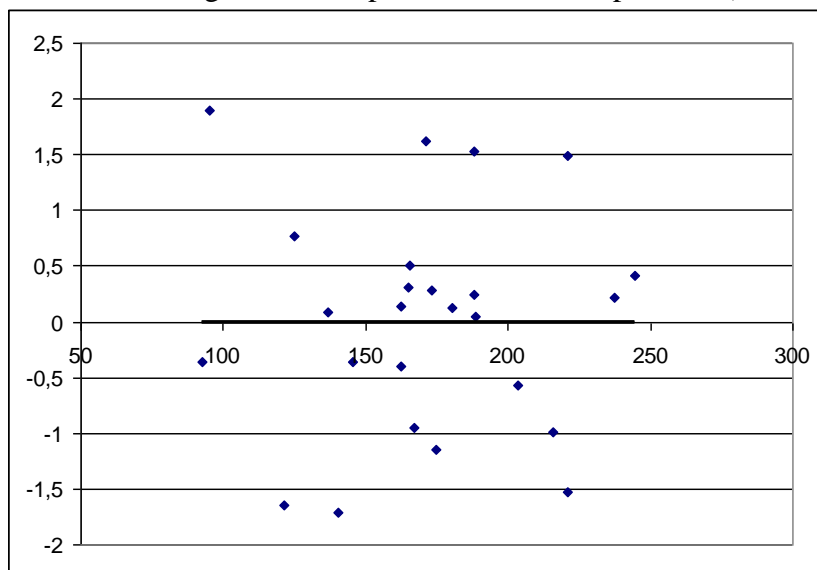
Equação da reta: $Y = 0,164897 X + 22,84791$. O coeficiente angular positivo significa que a reta é crescente.

e) Com base na equação da reta é possível calcular os valores preditos de Y: substitui-se cada valor do número de clientes (X) na equação e obtém-se Y. Posteriormente calcula-se a diferença (resíduo) entre o Y realmente observado e o Y predito. Calcula-se o desvio padrão dos resíduos. Para obter os resíduos padronizados basta dividir cada valor de resíduo pelo desvio padrão dos resíduos. Os resultados estão na tabela a seguir:

Y predito	Resíduo	Resíduo Padronizado
93	-5	-0,3547924
171	21	1,62566369
203	-7	-0,5716193
188	3	0,23836531
121	-21	-1,6424037
221	19	1,4868962
165	4	0,30938708
237	3	0,21489678
162	-5	-0,4000407
95	25	1,8974149
175	-15	-1,1480633
125	10	0,76490055
167	-12	-0,9486385
146	-5	-0,3622649
245	5	0,41388678
165	7	0,50264402
216	-13	-0,985646
137	1	0,08047789
180	2	0,12923428
173	4	0,27777916
162	2	0,13993211
188	1	0,04592738
188	-20	-1,52428788
221	-20	-1,5215238
140	-22	-1,7167015

O desvio padrão dos resíduos vale 12,9636155.

Fazendo um diagrama de dispersão dos valores preditos (eixo X) pelos resíduos padronizados (eixo Y):



Análise do diagrama de dispersão dos resíduos padronizados:

- o número de resíduos positivos é próximo do número de resíduos negativos.
- a distância dos resíduos positivos e dos negativos a zero é praticamente igual.
- aparentemente não há padrões nos resíduos.

Por esses motivos a adoção de um modelo linear para descrever o relacionamento entre as variáveis parece ser apropriado.

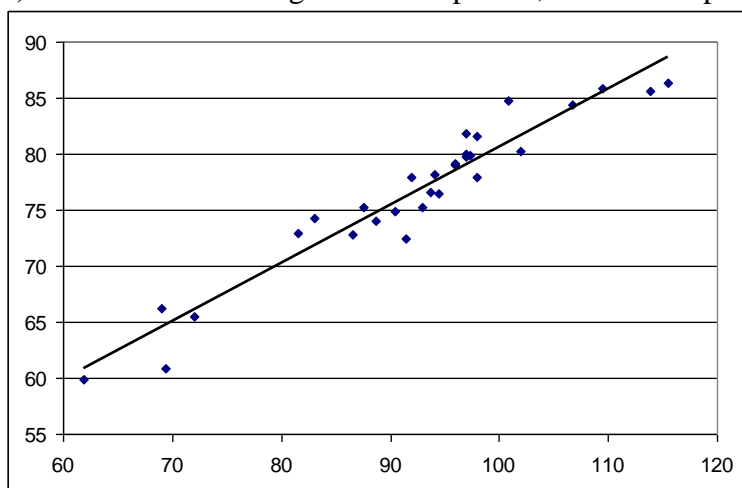
f) Para fazer a previsão basta substituir o número de clientes (X) na equação da reta obtida no item d):

$$Y = 0,164897 \times (900) + 22,84791 = 171,2553152$$

De acordo com o modelo linear, o valor das vendas para 900 clientes seria de 171,255 mil reais. Como este valor é inferior ao custo operacional (190 mil reais) a loja não deve ser aberta em Joinville.

6) a) Preço de avaliação é a variável independente, e preço de venda é a variável dependente. Porque o preço de avaliação pode influenciar o preço de venda.

b) Construindo um diagrama de dispersão, sendo X os preços de avaliação e Y os preços de venda:



Com base no diagrama de dispersão é recomendável a adoção de um modelo linear para descrever o relacionamento entre as variáveis, porque é possível ajustar uma reta aos dados.

c) $r = 0,96$ $r^2 = 0,93$. Com base nesses resultados é recomendável a adoção de um modelo linear para descrever o relacionamento entre as variáveis, porque o modelo linear consegue explicar 93% da variação Y a partir da variação de X.

A conclusão é coerente com a resposta do item b, pois pelo diagrama de dispersão é possível observar que a correlação linear entre as variáveis é forte.

d) b (coeficiente angular) = 0,51950302 a (coeficiente linear) = 28,65137485.

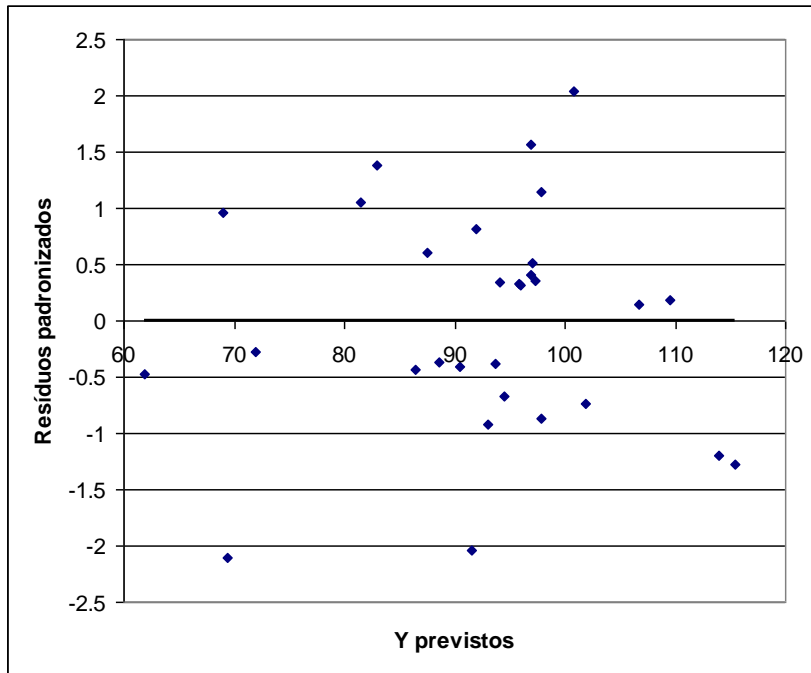
Equação da reta: $Y = 0,51950302 X + 28,65137485$. O coeficiente angular positivo significar que a reta é crescente.

e) Com base na equação da reta é possível calcular os valores preditos de Y: substitui-se cada valor do preço de avaliação (X) na equação e obtém-se Y. Posteriormente calcula-se a diferença (resíduo) entre o Y realmente observado e o Y predito. Calcula-se o desvio padrão dos resíduos. Para obter os resíduos padronizados basta dividir cada resíduo pelo desvio padrão dos resíduos. Resultados na tabela a seguir:

Y previsto	Resíduo	Resíduos padronizados
77,53660905	0,63339095	0,343534465
81,58873261	-1,348732607	-0,731516821
74,70531759	-0,67531759	-0,366274363
88,65397368	-2,343973681	-1,271309203
74,10788912	1,112110883	0,603179469
66,0555923	-0,515592304	-0,279643601
76,1859012	-3,755901197	-2,037101268
87,82276885	-2,212768849	-1,200147179
64,67371427	-3,873714271	-2,100999956
78,99121751	2,888782494	1,566799064
78,52366479	0,586335212	0,318012679
60,8086118	-0,878611801	-0,476535755
76,96515573	-1,695155728	-0,91940754
85,53695556	0,34304444	0,186058213
77,35478299	-0,714782993	-0,387679351
84,0823471	0,277652896	0,150591572
70,990871	1,949129004	1,057155915
77,74441026	-1,244410258	-0,674935144
64,49708324	1,782916756	0,967006796
78,99121751	0,748782494	0,40611978
73,5883861	-0,808386097	-0,438447194
79,51072053	-1,610720526	-0,873612125
71,77012553	2,539874474	1,377560601
79,19901871	0,650981286	0,353074997
81,01727928	3,762720715	2,040799993
79,51072053	2,099279474	1,13859355
75,66639818	-0,746398177	-0,40482659
79,04316781	0,936832192	0,508112952
76,44565271	1,514347293	0,821341837
78,47171449	0,598285514	0,324494207

O desvio padrão dos resíduos vale 1,843747906

Fazendo um diagrama de dispersão dos valores preditos (eixo X) pelos resíduos padronizados (eixo Y):



Análise do diagrama de dispersão dos resíduos padronizados:

- o número de resíduos positivos é próximo do número de resíduos negativos.

- a distância dos resíduos positivos e dos negativos a zero é semelhante.

- aparentemente não há padrões nos resíduos.

Por esses motivos a adoção de um modelo linear para descrever o relacionamento entre as variáveis parece ser apropriado.

f) Para fazer a previsão basta substituir o preço de avaliação (X) na equação da reta obtida no item d:

$$Y = 0,51950302 \times (100) + 28,65137485 = 80,60167687$$

De acordo com o modelo linear, o preço de venda de uma casa avaliada em 100 mil reais seria de 80,601 mil reais. Como o modelo linear é considerado adequado para descrever o relacionamento entre as variáveis (pelo diagrama de dispersão, pelo coeficiente de correlação linear e pela análise de resíduos) o valor obtido é confiável.

7) Três aspectos precisam ser observados ao fazer a análise de resíduos: número de resíduos positivos e negativos, grandeza dos resíduos positivos e negativos, e existência de padrões NÃO aleatórios no diagrama de dispersão.

a) Análise dos resíduos:

- o número de resíduos positivos e negativos é bastante próximo;

- a distância dos resíduos positivos a zero é maior do que a da maioria dos negativos;

- há claramente um padrão no diagrama de dispersão, os resíduos padronizados assumem maiores valores (positivos) quando os valores de Y predito aproximam-se dos extremos, e menores para valores intermediários (formato de parábola).

Devido às distâncias diferentes dos resíduos positivos e negativos a zero e devido à presença de padrão no diagrama, o modelo de regressão que gerou os resíduos NÃO parece ser adequado para descrever o relacionamento entre as variáveis.

b) Análise dos resíduos:

- o número de resíduos positivos e negativos é bastante próximo;

- a distância dos resíduos negativos a zero é maior do que a da maioria dos positivos;

- há claramente um padrão no diagrama de dispersão, à medida que os valores de Y predito aumentam, os resíduos padronizados passam de negativos e próximos a zero para positivos e depois tornam-se negativos novamente, distanciando-se mais acentuadamente do zero (formato de parábola invertida).

Devido às distâncias diferentes dos resíduos positivos e negativos a zero e devido à presença de padrão no diagrama, o modelo de regressão que gerou os resíduos NÃO parece ser adequado para descrever o relacionamento entre as variáveis.

c) Análise dos resíduos:

- o número de resíduos positivos e negativos é bastante próximo;

- a distância dos resíduos positivos e negativos a zero é bastante semelhante;

- não há padrões no diagrama (os pontos distribuem-se como uma "nuvem").

Como os números de resíduos positivos e negativos são semelhantes, as distâncias dos resíduos negativos e positivos a zero são semelhantes e não há padrão no diagrama, o modelo de regressão que gerou os resíduos parece ser adequado para descrever o relacionamento entre as variáveis.

d) Análise dos resíduos:

- o número de resíduos positivos e negativos é bastante próximo;
- a distância dos resíduos positivos e negativos a zero é bastante semelhante;
- não há padrões no diagrama (os pontos distribuem-se como uma "nuvem").

Como os números de resíduos positivos e negativos são semelhantes, as distâncias dos resíduos negativos e positivos a zero são semelhantes e não há padrão no diagrama, o modelo de regressão que gerou os resíduos parece ser adequado para descrever o relacionamento entre as variáveis.

e) Análise de resíduos:

- o número de resíduos positivos e negativos é bastante próximo;
- a distância dos resíduos negativos a zero é maior do que a da maioria dos positivos, especialmente para valores mais altos de Y predito;
- há claramente um padrão no diagrama de dispersão, à medida que os valores de Y predito aumentam, os resíduos padronizados vão se tornando negativos, e afastam-se de zero de forma acentuada.

Devido às distâncias diferentes dos resíduos positivos e negativos a zero e devido à presença de padrão no diagrama, o modelo de regressão que gerou os resíduos NÃO parece ser adequado para descrever o relacionamento entre as variáveis.

f) Análise dos resíduos:

- o número de resíduos positivos e negativos é bastante próximo;
- a distância dos resíduos positivos e negativos a zero é bastante semelhante;
- não há padrões no diagrama (os pontos distribuem-se como uma "nuvem").

Como os números de resíduos positivos e negativos são semelhantes, as distâncias dos resíduos negativos e positivos a zero são semelhantes e não há padrão no diagrama, o modelo de regressão que gerou os resíduos parece ser adequado para descrever o relacionamento entre as variáveis.

g) Análise de resíduos:

- o número de resíduos positivos e negativos é bastante próximo;
- as distâncias dos resíduos positivos e negativos a zero são semelhantes;
- há um padrão claro nos dados, quando os valores de Y predito ultrapassam 200 os resíduos padronizados (tanto positivos quanto negativos) aumentam muito, indicando que o modelo não proporciona bom ajuste para todo o conjunto de dados.

Devido à presença de padrão no diagrama, o modelo de regressão que gerou os resíduos NÃO parece ser adequado para descrever o relacionamento entre as variáveis.

h) Análise de resíduos:

- o número de resíduos positivos e negativos é bastante próximo;
- a distância dos resíduos negativos a zero é maior do que a da maioria dos positivos;
- há claramente um padrão no diagrama de dispersão, os resíduos padronizados assumem maiores valores (negativos) quando os valores de Y predito aproximam-se dos extremos, e menores para valores intermediários (formato de parábola invertida).

Devido às distâncias diferentes dos resíduos positivos e negativos a zero e devido à presença de padrão no diagrama, o modelo de regressão que gerou os resíduos NÃO parece ser adequado para descrever o relacionamento entre as variáveis.