

UNIDADE 4

1) Uma rede de lojas de vendas por atacado quer avaliar o desempenho de suas filiais, e de quebra verificar a viabilidade de implantar uma nova loja em Joinville, SC. Produziu a tabela a seguir, relacionando o número de clientes com as vendas em milhares de reais em um determinado mês. Com base nela responda as questões apresentadas.

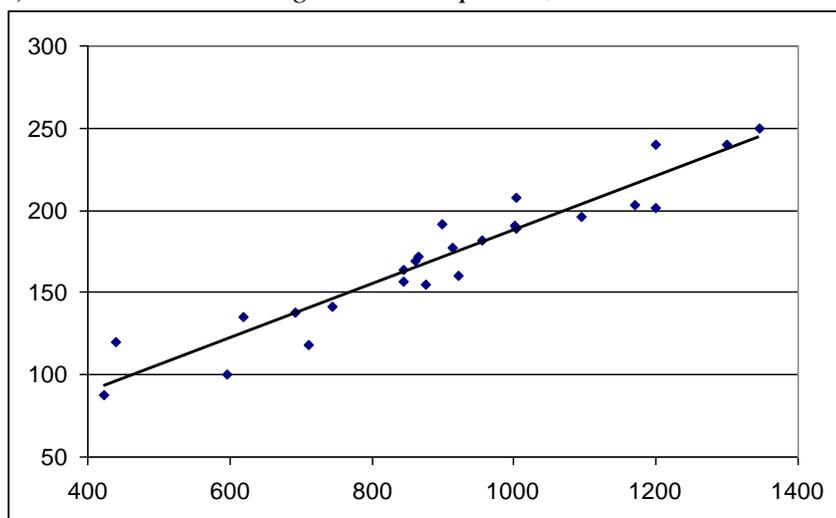
Filial	Número de clientes	Vendas (R\$ mil)
1	423	88
2	898	192
3	1095	196
4	1001	191
5	597	100
6	1200	240
7	862	169
8	1300	240
9	845	157
10	440	120
11	922	160
12	620	135
13	876	155
14	745	141
15	1345	250
16	865	172
17	1170	203
18	692	138
19	955	182
20	913	177
21	845	164
22	1004	189
23	1003	208
24	1200	201
25	712	118

- Qual é a variável independente? Qual é a variável dependente? JUSTIFIQUE sua resposta.
- Construa um diagrama de dispersão. Com base no diagrama você sugere a adoção de um modelo linear (reta) para o relacionamento entre as variáveis? JUSTIFIQUE sua resposta.
- Calcule os coeficientes de correlação linear de Pearson e de determinação com os dados da tabela ao lado. Com base nos resultados você sugere a adoção de um modelo linear para o relacionamento entre as variáveis? JUSTIFIQUE sua resposta. Compare com a resposta do item b.
- Calcule os coeficientes angular e linear da melhor reta que pode ser ajustada aos dados. Interprete o significado do coeficiente angular.
- Faça a análise de resíduos: calcule os valores preditos de Y pelo modelo, calcule os resíduos (diferença entre valores reais e preditos de Y), calcule os resíduos padronizados (dividindo cada resíduo pelo desvio padrão de todos os resíduos), e construa o diagrama de dispersão dos resíduos padronizados pelos valores preditos de Y. Com base neste diagrama de dispersão você sugere a adoção de um modelo linear para o relacionamento entre as variáveis? JUSTIFIQUE sua resposta/ Compare com as respostas dos itens b e c.
- Uma pesquisa de mercado detectou que uma nova loja em Joinville teria cerca de 900 clientes em potencial. O custo operacional mensal de uma loja capaz de atender tal número está por volta de 190 mil reais. Usando o modelo desenvolvido no item d estime o valor de vendas para 900 clientes e decida se a loja deve ser aberta ou não.

Adaptado de LEVINE, D.M., BERENSON, M.L., STEPHAN, D., Estatística: Teoria e Aplicações usando Microsoft® Excel em Português. Rio de Janeiro: LTC, 2000.

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.

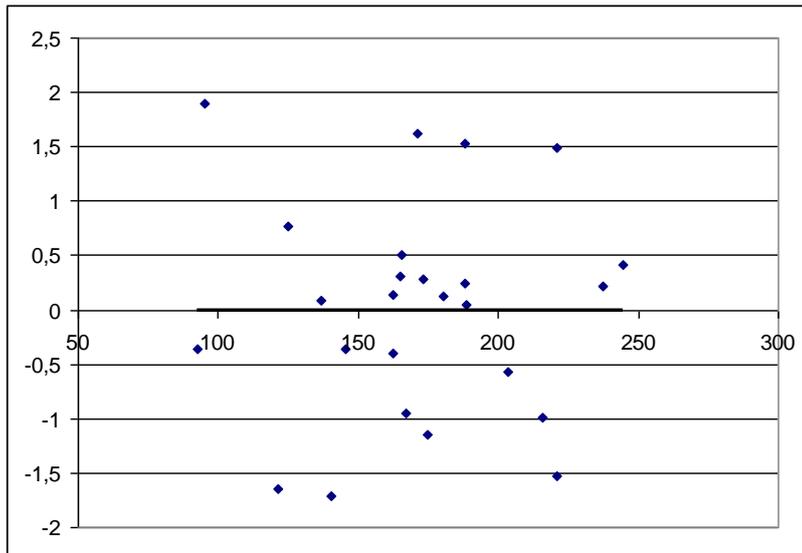
Equação da reta: $Y = 0,164897 X + 22,84791$. O coeficiente angular positivo significa que a reta é crescente.

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.

2) Como corretor incansável você decidiu estudar um pouco mais os valores de avaliação e de venda dos últimos imóveis negociados pela sua imobiliária. Há interesse em construir um modelo relacionando preço de venda e valor de avaliação do imóvel, com o objetivo de fazer algumas previsões. Você somente negocia com casas, e obteve os valores de avaliação e de venda de 30 unidades, descritas na tabela a seguir. Com base nela responda as questões apresentadas.

a) Qual é a variável independente? Qual é a variável dependente? JUSTIFIQUE sua resposta.

Casa	Preço de avaliação (R\$ 1000)	Preço de venda (R\$ 1000)
1	94,1	78,17
2	101,9	80,24
3	88,65	74,03
4	115,5	86,31
5	87,5	75,22
6	72	65,54
7	91,5	72,43
8	113,9	85,61
9	69,34	60,8
10	96,9	81,88
11	96	79,11
12	61,9	59,93
13	93	75,27
14	109,5	85,88
15	93,75	76,64
16	106,7	84,36
17	81,5	72,94
18	94,5	76,5
19	69	66,28
20	96,9	79,74
21	86,5	72,78
22	97,9	77,9
23	83	74,31
24	97,3	79,85
25	100,8	84,78
26	97,9	81,61
27	90,5	74,92
28	97	79,98

29	92	77,96
30	95,9	79,07

b) Construa um diagrama de dispersão. Com base no diagrama você sugere a adoção de um modelo linear (reta) para o relacionamento entre as variáveis? JUSTIFIQUE sua resposta.

c) Calcule os coeficientes de correlação linear de Pearson e de determinação com os dados da tabela ao lado. Com base nos resultados você sugere a adoção de um modelo linear para o relacionamento entre as variáveis? JUSTIFIQUE sua resposta. Compare com a resposta do item b.

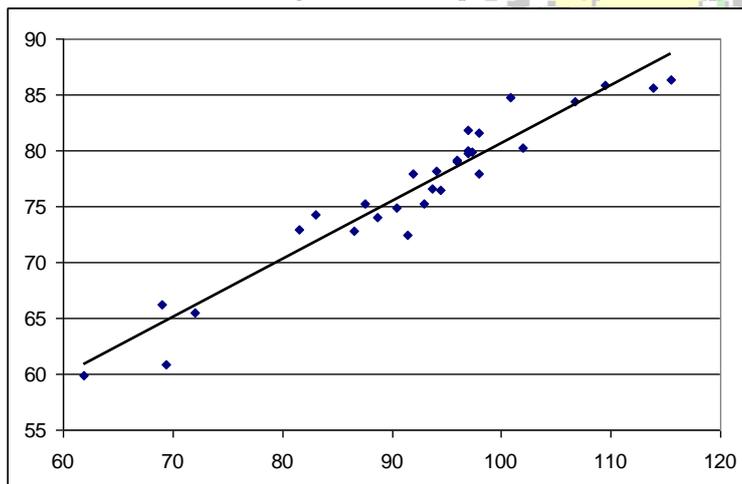
d) Calcule os coeficientes angular e linear da melhor reta que pode ser ajustada aos dados. Interprete o valor do coeficiente angular.

e) Faça a análise de resíduos: calcule os valores preditos de Y pelo modelo, calcule os resíduos (diferença entre valores reais e preditos de Y), calcule os resíduos padronizados (dividindo cada resíduo pelo desvio padrão de todos os resíduos), e construa o diagrama de dispersão dos resíduos padronizados pelos valores preditos de Y. Com base neste diagrama de dispersão você sugere a adoção de um modelo linear para o relacionamento entre as variáveis? JUSTIFIQUE sua resposta/ Compare com as respostas dos itens b e c.

Adaptado de LEVINE, D.M., BERENSON, M.L., STEPHAN, D., Estatística: Teoria e Aplicações usando Microsoft® Excel em Português. Rio de Janeiro: LTC, 2000.

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 significa que a reta é crescente.

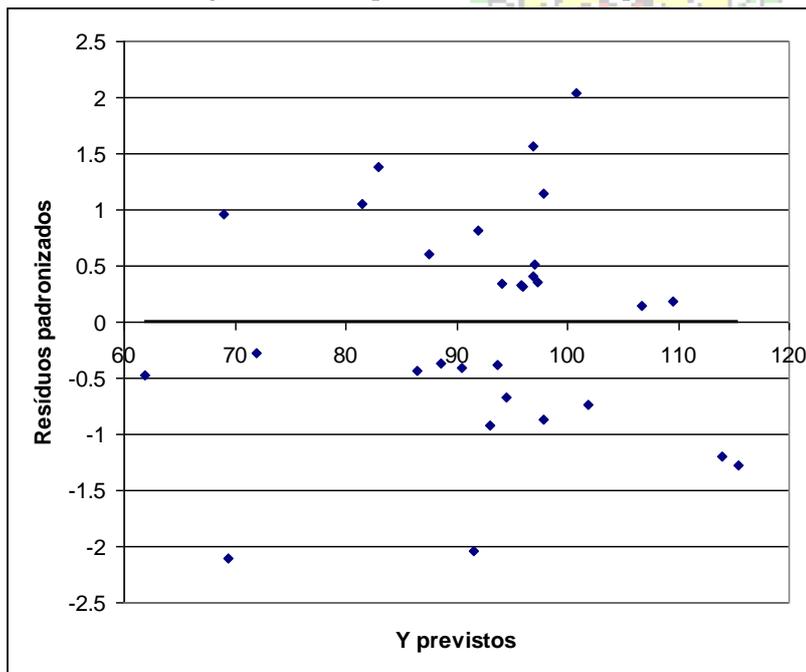
e) Com base na equação da reta é possível calcular os valores preditos de Y: substitui-se cada valor do preço de avaliação (X) na equação e obtém-se Y. Posteriormente calcula-se a diferença (resíduo) entre o Y realmente observado e o Y predito. Calcula-se o desvio padrão dos resíduos. Para obter os resíduos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

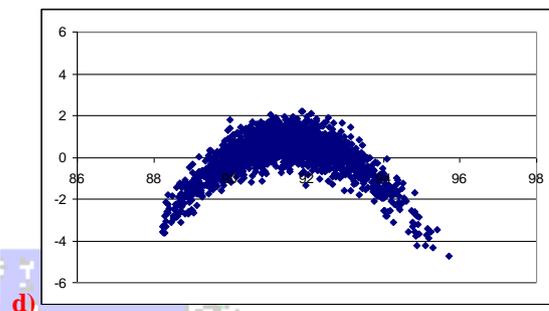
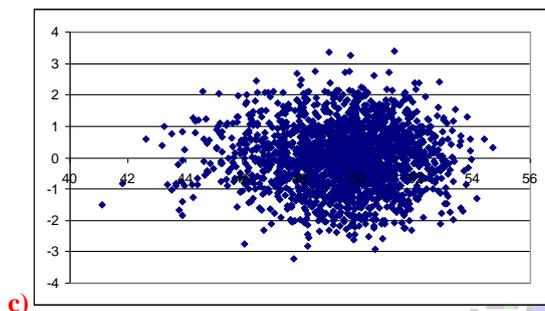
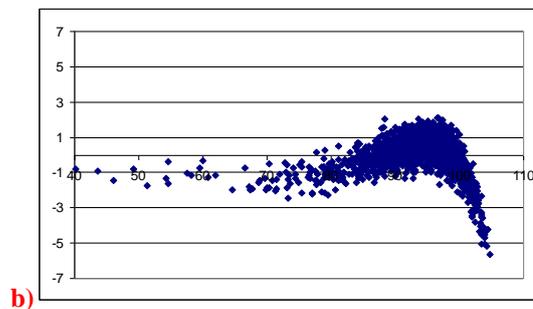
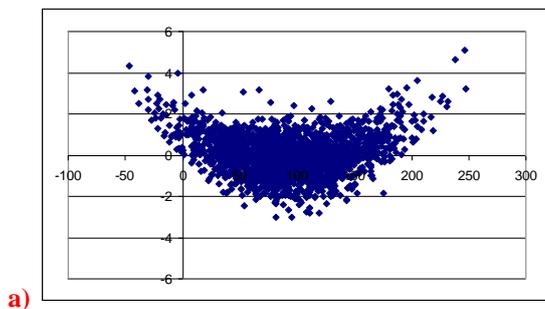
Por esses motivos a adoção de um modelo linear para descrever o relacionamento entre as variáveis parece ser apropriado.

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.

3) Os diagramas de dispersão mostram no eixo horizontal os valores preditos pelo modelo de regressão e no eixo vertical os resíduos padronizados. Faça a análise dos resíduos e emita uma opinião sobre a adequação do modelo obtido para o relacionamento entre as variáveis. JUSTIFIQUE a sua resposta.



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").

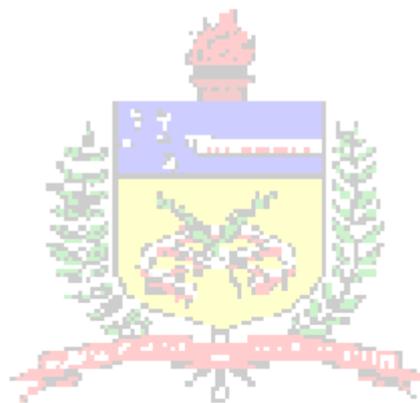
Como os números de resíduos positivos e negativos são semelhantes, as distâncias dos resíduos negativos e positivos a zero são semelhantes e não há padrão no diagrama, o modelo de regressão que gerou os resíduos parece ser adequado para descrever o relacionamento entre as variáveis.

d) Análise 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.



**Marcelo
Menezes
Reis**