

Instituto Federal de Educação, Ciência e Tecnologia do Amazonas

Disciplina: Probabilidade & Estatística

Docente: Silvestre Monteiro

Discentes:

1 - Carlos Alexandre

2 - Carlos Souza

3 - Rafael Kanda

4 - Saile Costa

Data: 29/06/2022

Apresentação - Diagrama de Dispersão

O conteúdo do diagrama de dispersão

O método do diagrama de dispersão é uma representação simples que é popularmente usada em comércio e estatística para encontrar a correlação entre duas variáveis. Essas duas variáveis são plotadas ao longo dos eixos X e Y em um gráfico bidimensional e o padrão representa a associação entre essas variáveis dadas. O estudo de tal representação gráfica envolvendo duas variáveis e usando tal diagrama é conhecido como análise de diagrama de dispersão.

Os alunos devem ser muito particulares ao traçar esses gráficos. Diagramas de dispersão em estatística e comércio são uma ferramenta vital que requer precisão, pois sua análise depende de tais representações.

Interpretação do Diagrama de Dispersão

Duas variáveis envolvidas em um estudo são representadas nos eixos X e Y. Essas variáveis podem ser tomadas como variáveis independentes, embora isso torne a segunda variável dependente desta anterior. Correspondentemente, todos esses pontos são plotados no gráfico e sua totalidade é chamada de diagrama de dispersão.

Depois de traçar todos esses pontos em um gráfico, os perfis gerados desses gráficos de dispersão são usados para desenhar uma extrapolação. Consequentemente, os alunos também podem calcular a correlação do coeficiente desses dados dados usando sua representação plotada. Notavelmente, a correlação do diagrama de dispersão é uma medida quantitativa de variáveis aleatórias e sua associação entre si.

Tipos de diagramas de dispersão

Ao compreender seus vários tipos, é importante descrever o diagrama de dispersão com exemplos para uma melhor compreensão dos alunos. Notavelmente, embora possa haver muitas representações, cada uma das quais sugere diferentes tipos de correlação, as mais comuns e vitais são explicadas abaixo.

Temos 7 tipos de Diagramas de Dispersão (Resumo Breve)

1. **Correlação Positiva Perfeita:** Um diagrama de dispersão é conhecido por ter uma correlação positiva perfeita se todos os pontos plotados estiverem em uma linha reta quando representados em um gráfico.
2. **Correlação Negativa Perfeita:** Entre os exemplos de diagramas de dispersão, uma correlação negativa perfeita é recíproca do tipo anterior. Aqui, cada ponto traçado está em uma linha reta, sem exceção também.
3. **Alto Grau de Correlação Positiva:** Se um diagrama de dispersão representa um alto grau de correlação positiva, então todos os seus pontos plotados estão aproximadamente ao longo de uma linha reta, mesmo que não criem uma linha claramente.
4. **Alto grau de correlação negativa:** Assim como as 2 correlações perfeitas, altos graus de correlação positiva e negativa são recíprocos entre si. Representando o significado e os valores do diagrama de dispersão, no caso de um alto grau de correlação negativa, cada ponto plotado forma uma banda que cai do canto superior esquerdo ao canto inferior direito.
5. **Baixo Grau de Correlação Positiva:** Aqui, temos um caso de um baixo grau de correlação, seja ela positiva ou negativa, esses pontos traçados ficam dispersos.
6. **Baixo Grau de Correlação Negativa:** Baixos graus de correlação negativa são representados em um gráfico com pontos de dispersão. No entanto, apesar de dispersos, esses pontos têm uma tendência geral de cair do canto superior esquerdo de um gráfico para o canto inferior direito.
7. **Sem Correlação:** Enquanto a definição do diagrama de dispersão procura encontrar a correlação entre as variáveis, os alunos devem notar que pode haver representações incoerentes e dispersas. Esta também é uma análise válida, pois mostra que as 2 variáveis dadas não estão correlacionadas. Nesses casos, esses pontos plotados são espalhados aleatoriamente em um gráfico.

Quais são as aplicações do diagrama de dispersão?

1° - Apresentação da relação entre quaisquer duas variáveis:

Um diagrama de dispersão representa essencialmente a relação distinta entre quaisquer duas variáveis. Essa relação pode ser linear ou não linear, positiva ou negativa, forte ou fraca. Tal representação fornece uma observação das duas variáveis que auxilia na análise dos dados através dos pontos de dados ou pontos mostrados no diagrama.

2° - Reconhecendo a relação correlacional:

Existe uma correlação entre as variáveis independentes presentes no eixo horizontal e as variáveis dependentes presentes no eixo vertical. Essa correlação pode ser identificada e estudada por observadores para derivar um valor provável das variáveis dependentes no eixo vertical.

3° - Observando padrões de dados:

Dados os vários pontos de dados ou pontos, no diagrama de dispersão o observador pode identificar os padrões comuns e incomuns de dados, seus respectivos valores e os pontos **outliers** (pontos distanciados anormais) presentes quando há lacunas de dados.

Usando a Linguagem R para demonstrar o Diagrama de Dispersão

Aqui já será outra ideia ... Salvando os trechos de códigos desenvolvidos

```
1 # Scatter Plot in R
2
3 # Considere o modelo 'Y = 2 + 3X^2 + ε', sendo 'Y' a variável dependente,
4 # 'X' a variável independente e 'ε' um termo de erro,
5 # tal que 'X~U(0, 1)' e 'ε~N(0, 0.25)'.
6
7 set.seed(12)
8
9 n <- 100
10 x <- runif(n)
11 eps <- rnorm(n, 0, 0.25)
12
13 y <- 2 + 3 * x^2 + eps
14
15 # Para plotar as observações, você pode digitar:
16 plot(x, y, pch = 19, col = "gray52")
17
```



```
1 # Gráfico de dispersão em R com cores diferentes #
2
3 # Se você tem uma variável que categoriza os pontos de dados em
4 # alguns grupos, você pode defini-la como parâmetro do argumento
5 # 'col' para plotar os pontos de dados 'com cores diferentes',
6 # dependendo do seu grupo, ou até mesmo definir símbolos
7 # diferentes por grupo. # nolint
8
9 group <- as.factor(ifelse(x < 0.5, "Group 1", "Group 2"))
10
11 plot(x, y, pch = as.numeric(group), col = group)

```



```
1 # Gráfico de dispersão com linha de regressão #
2
3 plot(x, y, pch = 19, col = "gray52")
4
5 # Modelo subjacente
6 lines(seq(0, 1, 0.05), 2 + 3 * seq(0, 1, 0.05)^2, col = "2", lwd = 3, lty = 2)
7
8 # Ajuste linear
9 abline(lm(y ~ x), col = "orange", lwd = 3)
10
11 # Ajuste suave
12 lines(lowess(x, y), col = "blue", lwd = 3)
13
14 # Legenda
15 legend("topleft",
16       legend = c("Theoretical", "Linear", "Smooth"),
17       lwd = 3, lty = c(2, 1, 1), col = c("red", "orange", "blue")
18 )

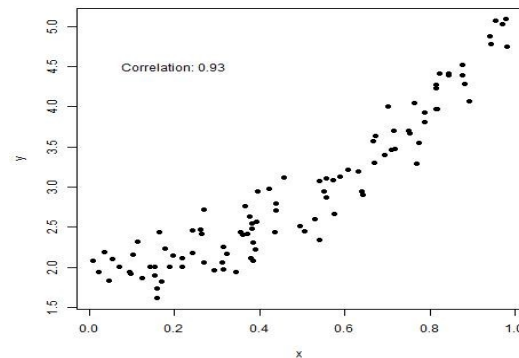
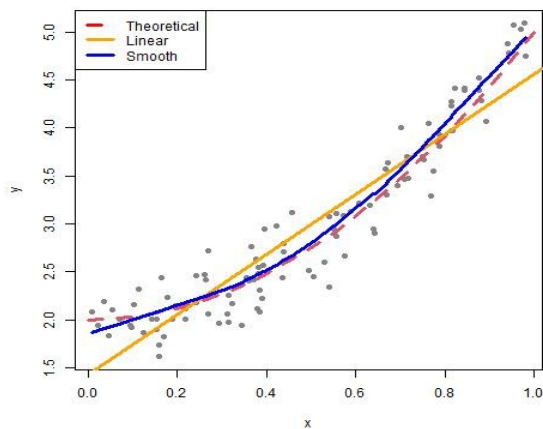
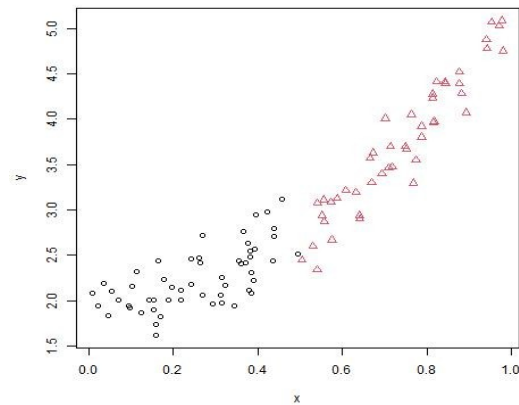
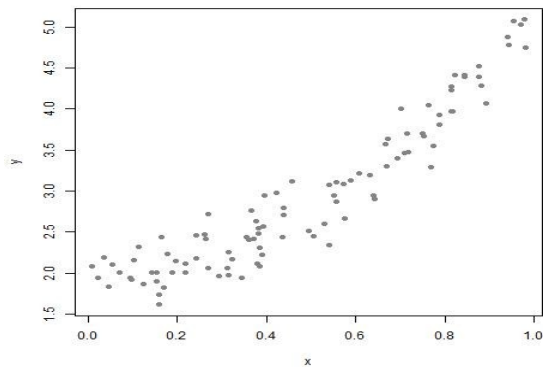
```



```
1 # Além disso, você pode adicionar a "correlação de Pearson"
2 # entre as variáveis que você pode calcular com a função
3 # 'cor'. # Então, você pode colocar a saída em algumas
4 # coordenadas do gráfico com a função 'text'. # nolint
5
6 # Calcule correlação
7 correlacao <- cor(x, y)
8
9 # Crie o gráfico e adicione o valor calculado
10 plot(x, y, pch = 19)
11
12 text(paste("Correlation:", round(correlacao, 2)), x = 0.2, y = 4.5)

```

Resultados dos códigos



Exemplo 01

Imagine que você queira analisar a altura, o peso e a idade das crianças que frequentam a escola em sua vizinhança. Vamos supor que os dados abaixo são o que você coletou. Uma possibilidade de pesquisa é a correlação entre 'altura' e 'idade'.

Dados coletados

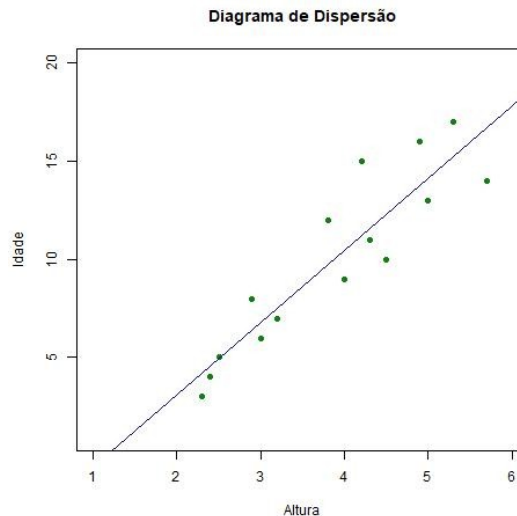
Age	Height	Weight in kg
3	2.3	6
4	2.4	9
5	2.5	10
6	3	8
7	3.2	7
8	2.9	9
9	4	12
10	4.5	15
11	4.3	18
12	3.8	22
13	5	19
14	5.7	20

Código

```
1 # Altura X Idade
2
3 jpeg("resExemp01.jpeg")
4
5 # altura em U.M americana
6 x <- c(
7   2.3, 2.4, 2.5, 3, 3.2, # nolint
8   2.9, 4, 4.5, 4.3, 3.8, # nolint
9   5, 5.7, 4.2, 4.9, 5.3
10 )
11
12 y <- c(
13   3, 4, 5, 6, 7, # nolint
14   8, 9, 10, 11, 12, # nolint
15   13, 14, 15, 16, 17
16 )
17
18 plot(x, y,
19   pch = 19, main = "Diagrama de Dispersão",
20   xlim = c(1, 6), ylim = c(1, 20),
21   xlab = "Altura", ylab = "Idade",
22   col = "#158115"
23 )
24
25 abline(lm(y ~ x), col = "#090979", lwd = 1)
26
27 dev.off()
```



Resultado



Exemplo 02

Imagine que você queira investigar se há uma relação entre o número de alunos e as notas obtidas. Vamos supor que os dados abaixo são o que você coletou.

Dados coletados

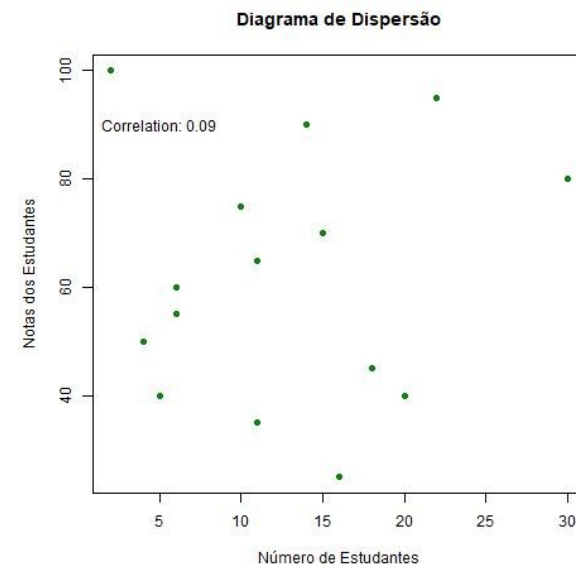
No of students	Marks Obtained	Percentage Of Students
5	40	2.5
6	60	3
25	70	12.5
11	65	5.5
30	80	15
4	50	2
6	55	3
10	75	5
14	90	7
18	45	9
20	40	10
22	95	11
2	100	1
11	35	5.5
16	25	8

Código

```
1 # Num. Estudante X Notas Estudantes
2
3 jpeg("resExemp02.jpeg")
4
5 x <- c(
6   5, 6, 15, 11, 30,
7   4, 6, 10, 14, 18,
8   20, 22, 2, 11, 16
9 )
10
11 y <- c(
12   40, 60, 70, 65, 80,
13   50, 55, 75, 90, 45,
14   40, 95, 100, 35, 25
15 )
16
17 # Calcule correlação
18 correlacao <- cor(x, y)
19
20 plot(x, y,
21   pch = 19, main = "Diagrama de Dispersão",
22   xlab = "Número de Estudantes",
23   ylab = "Notas dos Estudantes",
24   col = "#158115"
25 )
26
27 text(paste("Correlation:", round(correlacao, 2)), x = 5, y = 90)
28
29 dev.off()
```



Resultado



Referências Bibliográficas

[Scatter Diagram - Introduction, Interpretation and Types \(vedantu.com\)](https://vedantu.com/scatter-plot-introduction-types)

[SCATTER PLOT in R programming !\[\]\(c507f772dba2b921f86777f01218e570_img.jpg\) WITH EXAMPLES\] \(r-coder.com\)](https://r-coder.com/scatter-plot-in-r-programming-with-examples/)

<https://ppcexpo.com/blog/scatter-plot-examples>