

Análises do Desempenho de Campanhas de Marketing Digital (Independência do Qui-Quadrado e Regressão Linear Simples)

João Capelas¹, Miguel Velez²

¹ 2ºL_EI-SW-05, 201901812

² 2ºL_EI-SW-05, 201900678

Resumo: Este trabalho consiste na “recolha” e análise de dados referentes ao desempenho de campanhas de marketing digital através de dois testes diferentes, um para dados qualitativos e outro para dados quantitativos.

Palavras-chave: Qui-Quadrado, Análise, Dados, Regressão, Resíduos

1. Introdução

Este relatório foi feito sobre o trabalho de grupo da Unidade Curricular de Métodos Estatísticos.

O trabalho tem como objetivo utilizar o Teste de Independência do Qui-Quadrado e a Regressão Linear Simples para analisar possíveis associações entre determinadas variáveis referentes a dados sobre o desempenho de campanhas de marketing digital.

2. Análise das variáveis

Este capítulo irá abordar os dois tipos de análises realizadas para a concretização deste trabalho, análises para variáveis qualitativas e análises para variáveis quantitativas.

2.1. Análises das variáveis qualitativas

Para estudar a associação entre variáveis qualitativas, utiliza-se o teste de independência do Qui-Quadrado. Para chegar a uma conclusão através deste teste é necessário definir hipóteses e níveis de confiança e significância. Cada um dos três testes foi realizado com 95% de confiança e 5% de significância, por serem os valores mais usuais.

Para quantificar as associações existentes, utilizaram-se as referências de medidas presentes nas Figuras 1 e 2 abaixo.

	Associação		
	fraca	moderada	elevada
coeficiente de contingência	[0.10, 0.30[[0.30, 0.50[≥ 0.50
V de Crámer ($k = 2^{(*)}$)	[0.10, 0.30[[0.30, 0.50[≥ 0.50
V de Crámer ($k = 3^{(*)}$)	[0.07, 0.20[[0.20, 0.35[≥ 0.35
V de Crámer ($k = 4^{(*)}$)	[0.06, 0.17[[0.17, 0.29[≥ 0.29

(*) k representa o número mínimo de categorias nas linhas ou nas colunas

Figura 1 - Medidas dos níveis de associação

Medidas de Associação \rightarrow as 2 das variáveis são qualitativas ordinais

- **coeficiente τ_b de Kendall:** assume valores entre -1 e 1, mas os valores -1 e 1 só são atingidos em tabelas em que o número de linhas é igual ao número de colunas. Valores próximos de -1 ou de 1 indicam forte associação. Valores próximos de zero indicam fraca associação.

Sinal do coeficiente:

- ▶ sinal positivo indica que o "aumento" de uma das variáveis é acompanhado pelo "aumento" da outra variável;
- ▶ sinal negativo indica que o "aumento" de uma das variáveis é acompanhado pela "diminuição" da outra variável.

Figura 2 - Medidas do coeficiente de Kendall

2.1.1. Associação: Plataforma – Anúncio

Este teste foi realizado com as seguintes hipóteses:

H_0 : As variáveis plataforma e anuncio não estão relacionadas.

Vs.

H_1 : As variáveis plataforma e anuncio estão relacionadas.

Após definir as hipóteses e os níveis de confiança/significância foi realizada a tabela de contingência. Ao consultar a base de dados criada foi obtida a seguinte tabela:

Tabela 1 - Tabela de contingência da associação: Plataforma – Anúncio

Plataformas	Anúncios		
	Banner	Texto	Vídeo
Facebook	24	17	30
Google	45	27	21
Instagram	23	33	22

Com a realização do teste de independência do Qui-Quadrado com esta tabela de contingência (Tabela 1), obtiveram-se os seguintes resultados: Graus de liberdade = 4 e P-Value = 0.0071.

Dado que o P-Value é inferior ao nível de significância tem-se que rejeitar H_0 , o que indica que, com 5% de significância e com base nas variáveis e bases de dados estudadas, há evidência estatística de que as variáveis estão relacionadas.

Como existe uma associação é possível quantificá-la utilizando as três medidas de associação:

- Coeficiente de Contingência = 0.2344 → através da Figura 1 vê-se que corresponde a uma relação fraca;
- Coeficiente V de Crámer = 0.1705 → através da Figura 1 vê-se que corresponde a uma relação fraca ($K=3$);
- Coeficiente T_b de Kendall = -0.0268 → através da Figura 2 vê-se que corresponde a uma relação fraca, e como é negativa sabe-se também que quando as plataformas aumentam, os anúncios diminuem (ou vice-versa).

2.1.2. Associação: Plataforma – Mercado

Este teste foi realizado com as seguintes hipóteses:

H_0 : As variáveis plataforma e mercado não estão relacionadas.

Vs.

H_1 : As variáveis plataforma e mercado estão relacionadas.

Após definir as hipóteses e os níveis de confiança/significância foi realizada a tabela de contingência. Ao consultar a base de dados criada foi obtida a seguinte tabela:

Tabela 2 - Tabela de contingência da associação: Plataforma – Mercado

Plataformas	Anúncios		
	Alimentação	Moda	Tecnologia
Facebook	28	18	25
Google	40	32	21
Instagram	33	22	23

Com a realização do teste de independência do Qui-Quadrado com esta tabela de contingência (Tabela 2), obtiveram-se os seguintes resultados: Graus de liberdade = 4 e P-Value = 0.459.

Dado que o P-Value é superior ao nível de significância não se rejeita H_0 , o que indica que, com 5% de significância e com base nas variáveis e bases de dados estudadas, há evidência estatística de que as variáveis não estão relacionadas.

Como não existe uma associação não é possível quantificá-la.

2.1.3. Associação: Anúncio - Mercado

Este teste foi realizado com as seguintes hipóteses:

H_0 : As variáveis anuncio e mercado não estão relacionadas.

Vs.

H_1 : As variáveis anuncio e mercado estão relacionadas.

Após definir as hipóteses e os níveis de confiança/significância foi realizada a tabela de contingência. Ao consultar a base de dados criada foi obtida a seguinte tabela:

Tabela 3 - Tabela de contingência da associação: Anuncio – Mercado

Plataformas	Anúncios		
	Alimentação	Moda	Tecnologia
Facebook	37	32	23
Google	29	17	31
Instagram	35	23	15

Com a realização do teste de independência do Qui-Quadrado com esta tabela de contingência (Tabela 3), obtiveram-se os seguintes resultados: Graus de liberdade = 4 e P-Value = 0.7977.

Dado que o P-Value é superior ao nível de significância não se rejeita H_0 , o que indica que, com 5% de significância e com base nas variáveis e bases de dados estudadas, há evidência estatística de que as variáveis não estão relacionadas.

Como não existe uma associação não é possível quantificá-la.

2.2. Análises das variáveis quantitativas

Para estudar a associação entre variáveis quantitativas, utiliza-se a regressão linear simples. Para quantificar as associações existentes, utilizaram-se as referências presentes nas Figuras 3, 4 e 5.

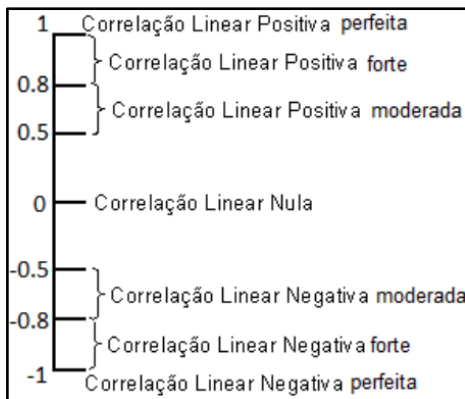
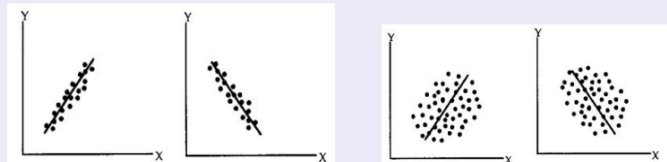


Figura 3 - Correlação Linear

Intensidade

- A correlação é tanto mais forte quanto menor for a dispersão dos pontos em torno da linha reta, isto é, quanto mais concentrados os pontos estiverem em torno dessa reta.



- A correlação diz-se **perfeita** se todos os pontos coincidirem com a reta.

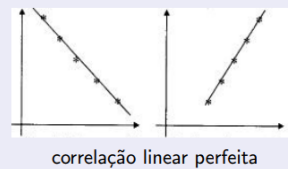
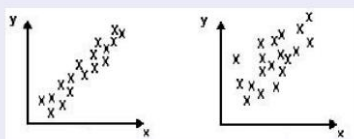


Figura 4 - Diagrama de Dispersão – Intensidade

Sinal

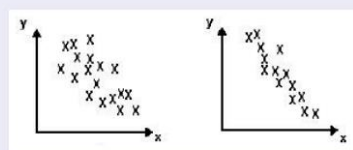
Correlação linear positiva

→ A maiores valores de uma variável tendem a corresponder maiores valores da outra variável.



Correlação linear negativa

→ A maiores valores de uma variável tendem a corresponder menores valores da outra.



- Correlação linear nula** → Não existe associação linear entre as variáveis.



Figura 5 - Diagrama de Dispersão – Sinal

2.2.1. Associação: Cliques – Investimento

Para este teste foram definidas as seguintes variáveis: cliques como variável independente e investimento como variável dependente.

Ao realizar a análise podemos concluir que o coeficiente de correlação linear entre as duas variáveis é $r_{xy} = -0.8384$, o que mostra que existe uma correlação linear negativa forte (tal como está indicado na Figura 3), mas muito perto de poder ser uma correlação linear negativa moderada.

Para o modelo de Regressão Linear Simples iremos utilizar as variáveis definidas e a seguinte equação da reta: $y = a + bx$. Ao analisar a Figura 6 podemos verificar que existe uma correlação linear negativa (referenciada na Figura 5) e que aparenta ser forte (referenciada na Figura 4), ou seja, a relação entre cliques e investimento é negativa forte e a equação da reta é:

$$\hat{y} = 15.5105 - 0.01106x. \quad (1)$$

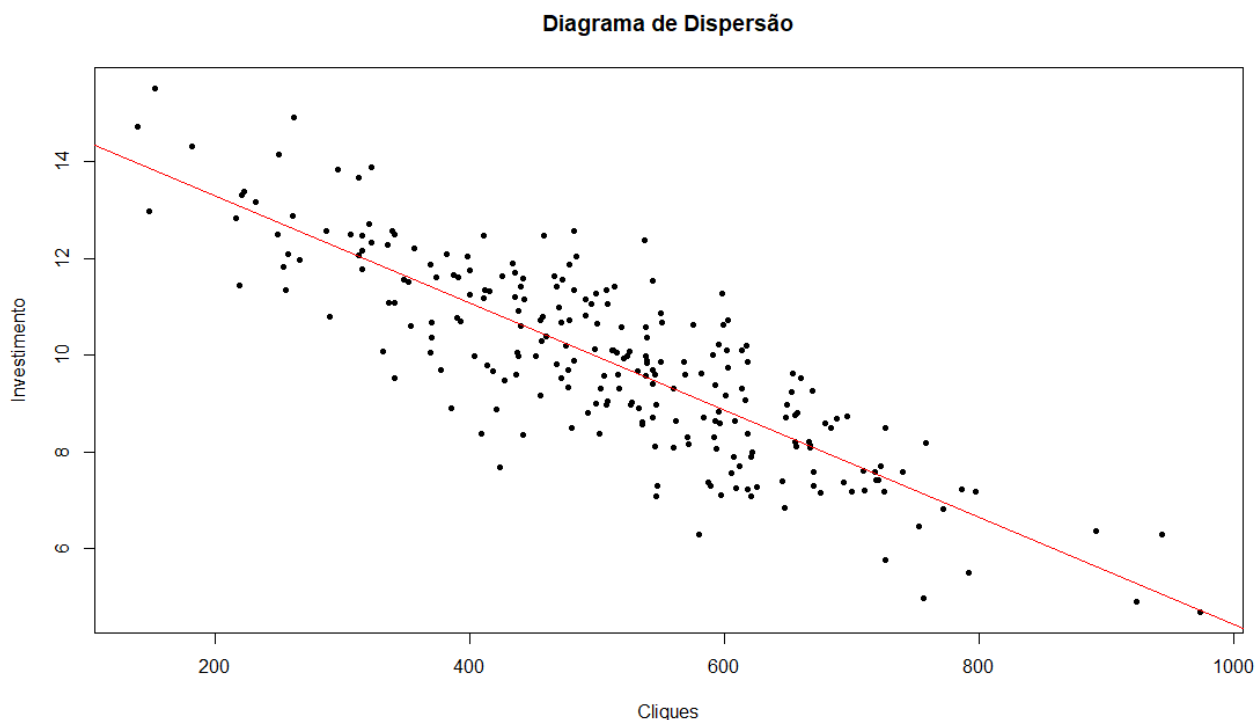


Figura 6 - Diagrama de dispersão da associação cliques - Investimento

Em relação aos resíduos, como podemos verificar na Figura 7 abaixo, os pontos distribuem-se de forma aleatória em torno da reta que corresponde ao resíduo zero (ou seja, não parece existir um padrão associado), o que indica que o modelo ajustado é bom.

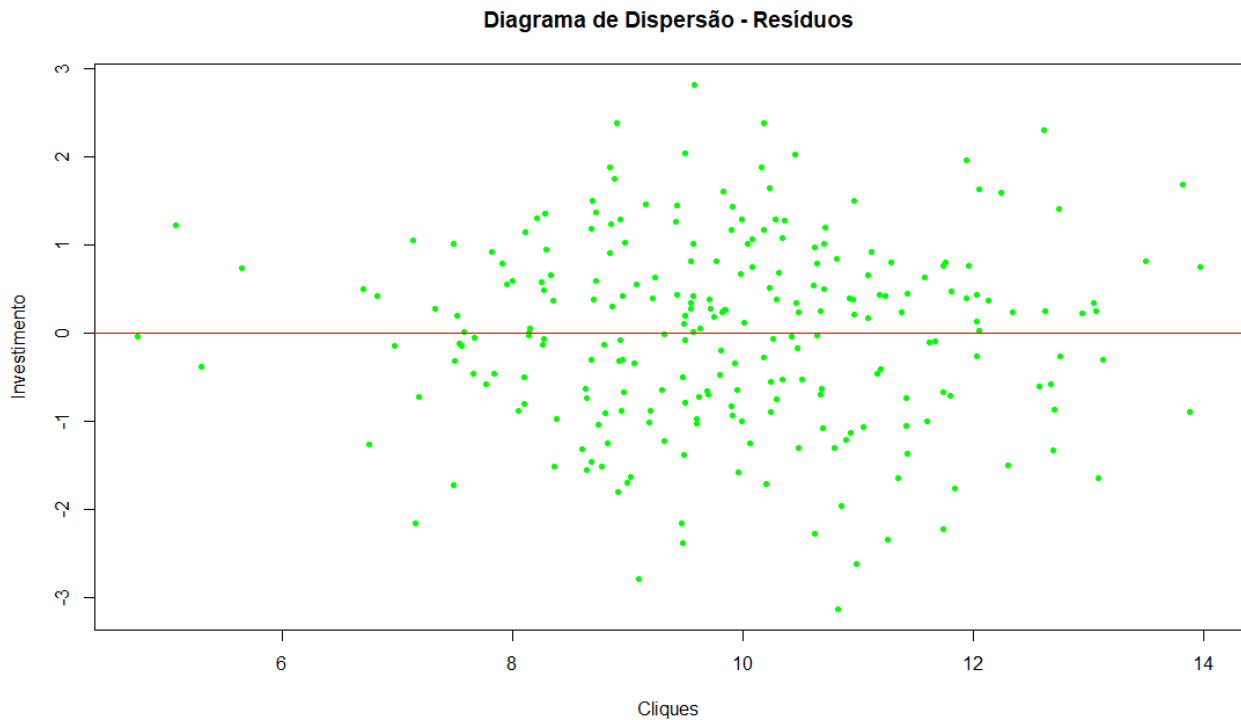


Figura 7 - Diagrama de dispersão dos resíduos da associação cliques – Investimento

2.2.2. Associação: Conversões – Investimento

Para este teste foram definidas as seguintes variáveis: conversões como variável independente e investimento como variável dependente.

Ao realizar a análise podemos concluir que o coeficiente de correlação linear entre as duas variáveis é $r_{xy} = 0.8946$, o que mostra que existe uma correlação linear positiva forte (tal como está indicado na Figura 3).

Para o modelo de Regressão Linear Simples iremos utilizar as variáveis definidas e a seguinte equação da reta: $y = a + bx$. Ao analisar a Figura 8 podemos verificar que existe uma correlação linear positiva (referenciada na Figura 5) e que aparenta ser forte (referenciada na Figura 4), ou seja, a relação entre conversões e investimento é positiva forte e a equação da reta é:

$$\hat{y} = 5.6672 - 0.08608x. \quad (2)$$

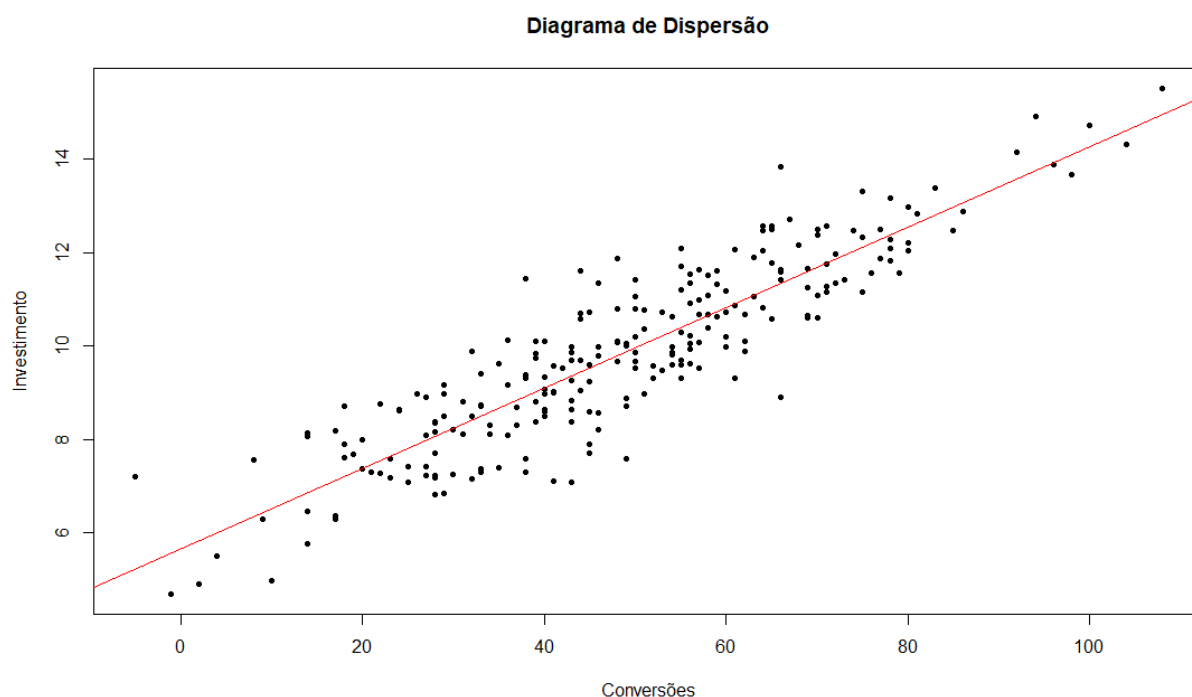


Figura 8 - Diagrama de dispersão da associação Conversões - Investimento

Em relação aos resíduos, como podemos verificar na Figura 9 abaixo, os pontos distribuem-se de forma aleatória em torno da reta que corresponde ao resíduo zero (ou seja, não parece existir um padrão associado), o que indica que o modelo ajustado é bom.

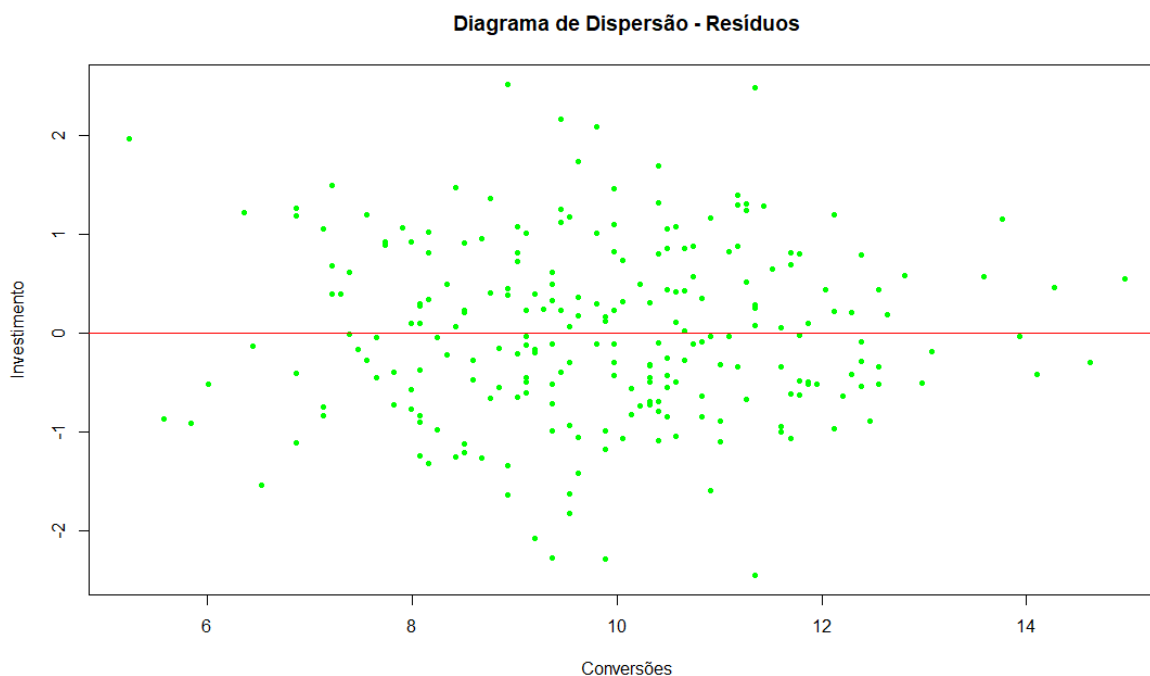


Figura 9 - Diagrama de dispersão dos resíduos da associação Conversões - Investimento

2.2.3. Associação: Cliques – Conversões

Para este teste foram definidas as seguintes variáveis: cliques como variável independente e conversões como variável dependente.

Ao realizar a análise podemos concluir que o coeficiente de correlação linear entre as duas variáveis é $r_{xy} = -0.759$, o que mostra que existe uma correlação linear negativa moderada (tal como está indicado na Figura 3).

Para o modelo de Regressão Linear Simples iremos utilizar as variáveis definidas e a seguinte equação da reta: $y = a + bx$. Ao analisar a Figura 10 podemos verificar que existe uma correlação linear negativa (referenciada na Figura 5) e que aparenta ser moderada (referenciada na Figura 4), ou seja, a relação entre cliques e conversões é negativa moderada e a equação da reta é:

$$\hat{y} = 101.8976 - 0.1041x. \quad (3)$$

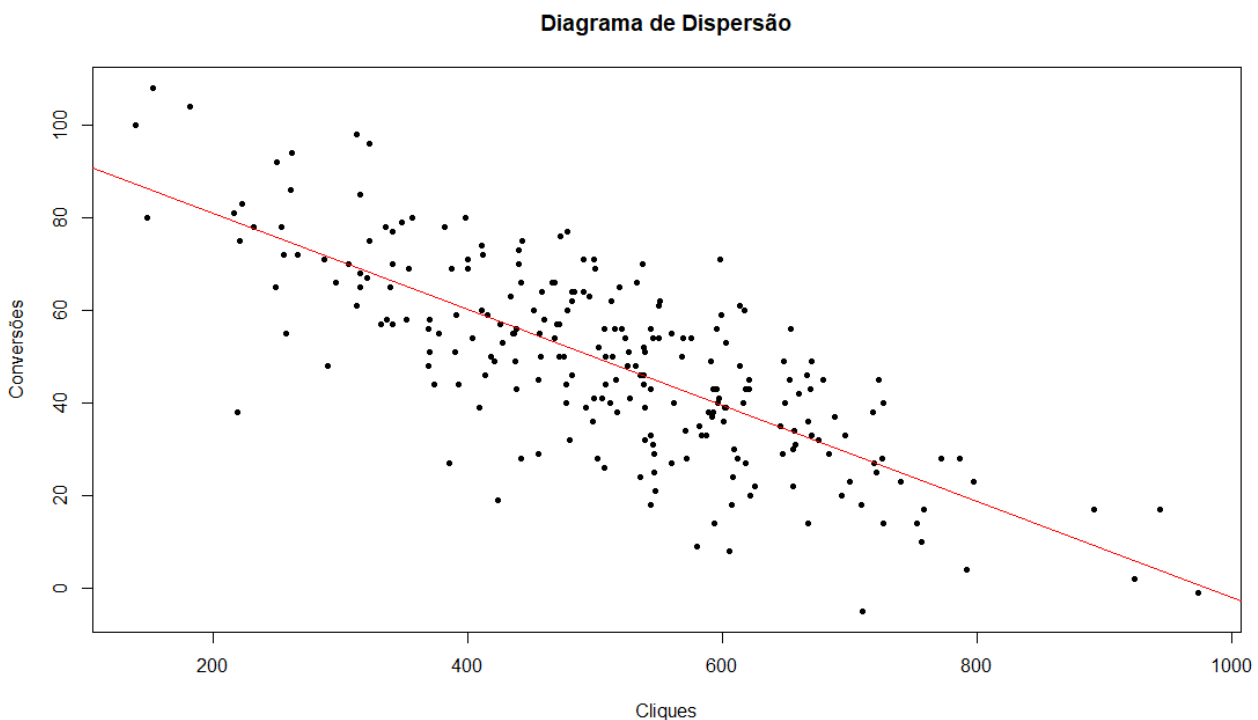


Figura 10 - Diagrama de dispersão da associação Cliques - Conversões

Em relação aos resíduos, como podemos verificar na Figura 11 abaixo, os pontos distribuem-se de forma aleatória em torno da reta que corresponde ao resíduo zero (ou seja, não parece existir um padrão associado), o que indica que o modelo ajustado é bom.

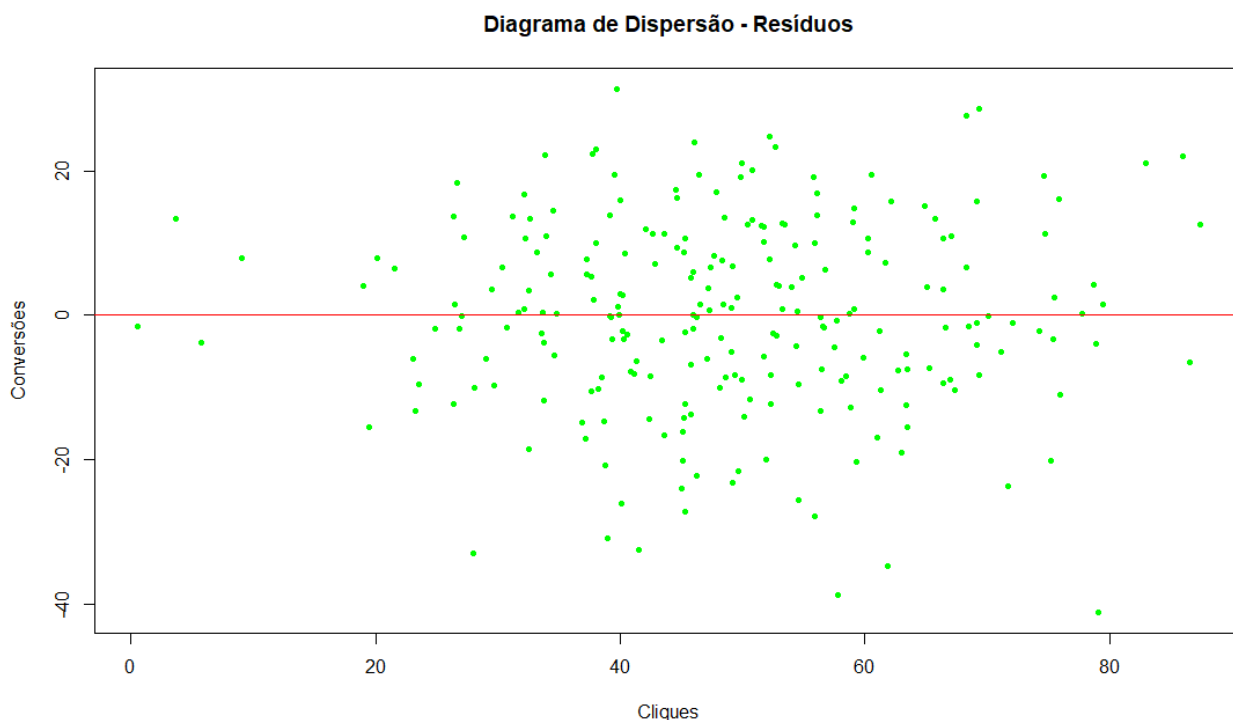


Figura 11 - Diagrama de dispersão dos resíduos da associação Cliques – Conversões

3. Conclusões

Com a realização deste trabalho foi possível aprofundar o conhecimento sobre os testes utilizados (Qui-Quadrado e Regressão Linear Simples) e observar que, para a base de dados utilizada e para as variáveis qualitativas, apenas se conseguiu identificar uma associação existente. Essa associação indica-nos que os anúncios realizados nas campanhas aparentam estar dependentes de plataformas específicas, e também se verificou que à medida que as plataformas aumentam os anúncios diminuem (ou vice-versa).

Para as variáveis quantitativas observou-se que nos três casos analisados existem associações, uma delas sendo positiva (Conversões - Investimento) e as outras duas negativas (Cliques - Investimento e Cliques - Conversões), e em todas elas os resíduos indicaram que os modelos seguidos estudados foram bons, pois nos três casos o seu diagrama não continha qualquer padrão.

Referências Bibliográficas

Slides disponibilizados pelas professoras, na plataforma Moodle – Documento “Capítulo 6 – Testes de Hipóteses Não Paramétricos – parte 2 – Teste de Independência”.

Slides disponibilizados pelas professoras, na plataforma Moodle – Documento “Capítulo 7 – Regressão Linear Simples”.