

# Medida de Associação entre Variáveis

Um Guia Prático para Entender Insights

Versão 1.0

Alexandre Soares



# Medida de Associação entre Variáveis

Um Guia Prático para Entender Insights

por

Alexandre Soares

# Notas Sobre o Estudo

Todos os programas fonte e dados utilizados para a elaboração deste estudo, bem como o PDF deste material, estão disponíveis para download no repositório GitHub, no endereço: <https://github.com/Alxsoa/Artigos>.

# Conteúdo

<b>1</b>	<b>Objetivo Deste Estudo</b>	<b>1</b>
<b>2</b>	<b>Uma Visão Geral do Método</b>	<b>2</b>
<b>3</b>	<b>Sobre os Dados</b>	<b>3</b>
3.1	Dados Utilizados . . . . .	3
3.2	Características Principais das Variáveis . . . . .	4
3.3	Entendendo a Cardinalidade . . . . .	5
3.4	Sobre Dados Ausentes . . . . .	6
3.5	Sobre Balanceamento de Dados . . . . .	7
<b>4</b>	<b>Diferenças Entre Pearson e Cramer V</b>	<b>8</b>
4.1	Principais Diferenças Entre os Métodos . . . . .	8
<b>5</b>	<b>Resultados e Conclusões do Uso do Método Cramer V</b>	<b>9</b>
5.1	Resultados Obtidos pelo Cramer V . . . . .	9
5.2	Associações Fortes . . . . .	11
5.3	Associações Moderadas . . . . .	12
5.4	Associações Fracas . . . . .	13
5.5	Conclusão das Associações Obtidas . . . . .	14
<b>6</b>	<b>Procedimento de Cálculo do Método Cramer V</b>	<b>15</b>
6.1	Contexto do Problema . . . . .	15
6.2	Equação do Qui-Quadrado . . . . .	16
6.3	Equação do Cramer V . . . . .	16
6.4	Obtendo o Valor do Qui-Quadrado . . . . .	16
6.5	Obtendo o Valor do Cramer V . . . . .	17
6.6	Conclusão dos Cálculos Obtidos . . . . .	17
<b>7</b>	<b>Conclusão</b>	<b>18</b>

# Lista de Figuras

3.1	Estrutura das Componentes do Dataset . . . . .	4
3.2	Relatório da Faltância de Dados . . . . .	6
3.3	Relatório Sobre o Balanceamento dos Dados . . . . .	7
5.1	Resultado de Todas as Associações . . . . .	9
5.2	Associações Consideradas Fortes . . . . .	11
5.3	Associações Consideradas Moderadas . . . . .	12
5.4	Associações Consideradas Fracas . . . . .	13

# Lista de Tabelas

3.1	Detalhamento dos Campos da Base de Dados . . . . .	3
3.2	Detalhamento da Cardinalidade . . . . .	5
5.1	Todos os Resultados . . . . .	10
6.1	Frequência Observada . . . . .	15
6.2	Frequências Esperadas . . . . .	16

# 1

## Objetivo Deste Estudo

As correlações são uma ferramenta essencial na ciência de dados, ajudando-nos a compreender as relações entre variáveis. Frequentemente, usamos correlações fortes e as direções dessas relações para eliminar variáveis indesejadas de conjuntos de dados, reduzindo assim a dimensionalidade e a complexidade dos problemas.

Tradicionalmente, o coeficiente de Pearson é utilizado para avaliar a relação entre duas variáveis. No entanto, esse método tem limitações, especialmente quando aplicado a dados nominais ou quando o tamanho da amostra varia significativamente.

Como alternativa, apresentaremos o método Cramer V, derivado do teste qui-quadrado, que normaliza a força da associação de acordo com o tamanho da amostra e o número de categorias. Isso resulta em uma medida mais confiável e compreensível.

Diferente de outras métricas, o Cramer V não assume linearidade e pode ser aplicado a variáveis nominais, tornando-se particularmente útil em análises onde as variáveis são categóricas.

Este estudo explora as limitações e vantagens do Cramer V em comparação com o coeficiente de Pearson, apresentando exemplos e aplicações práticas.

# 2

## Uma Visão Geral do Método

O método Cramer V é uma perspectiva estatística usada para calcular a força de associação entre duas variáveis categóricas, oferecendo um indicativo da magnitude da relação, independente do tamanho da amostra.

Esse método encontra um valor entre 0 e 1, que representa o grau de associação entre as variáveis. Um valor zero indica ausência de associação, enquanto o valor 1 aponta para uma associação perfeita. A relação entre as variáveis se fortalece à medida que o valor de Cramer V aumenta.

Diferentemente de outras métricas, o Cramer V não assume linearidade e pode ser aplicado com variáveis nominais, tornando-se particularmente útil em análises onde as variáveis são categóricas.

Uma característica significativa do método é que, sendo uma medida padronizada de associação, o número de observações não influencia o seu valor. Isso facilita a comparação do Cramer V entre diferentes conjuntos de dados, mesmo quando as quantidades de dados variam.

Este método é especialmente interessante ao trabalhar com grandes volumes de dados, permitindo examinar as relações entre múltiplas variáveis categóricas de forma padronizada, o que facilita a interpretação e comparação da força das relações.



# 3

## Sobre os Dados

### 3.1. Dados Utilizados

Os dados utilizados neste estudo é um subconjunto dos dados de clientes de uma empresa de telecomunicações fornecida pela IBM, este subconjunto, está no github como apontado na Seção **Notas Sobre o Estudo** a estrutura de dados é apresentado na Tabela-3.1.

**Tabela 3.1:** Detalhamento dos Campos da Base de Dados

Variável	Descrição
customerID	Código de identificação do cliente
gender	Gênero do Cliente
SeniorCitizen	Cliente aposentado
Partner	Cliente casado
Dependents	Cliente tem dependentes em seu plano
tenure	Quantos meses a pessoa é cliente
PhoneService	Possui serviço telefônico
MultipleLines	Possui múltiplas linhas telefônicas
InternetService	Qual provedor de serviço de internet
OnlineSecurity	Possui serviço de segurança online
OnlineBackup	Possui serviço de backup online
DeviceProtection	Cliente possui proteção de sistema
TechSupport	Possui serviço de suporte técnico
StreamingTV	Possui serviço de streamer
Continua na Próxima Página	

**Tabela 3.1 – Continuação da Página Anterior**

Variável	Descrição
StreamingMovies	Possui serviço de filmes
Contract	Tipo do Contrato
PaperlessBilling	Cliente utiliza faturamento sem papel
PaymentMethod	Método de pagamento
MonthlyCharges	Pagamento mensal atual
TotalCharges	Valor pago durante todo o tempo
Churn	Cliente abandonou o serviço

## 3.2. Características Principais das Variáveis

A Tabela-3.1 fornecida é uma ferramenta útil para a análise do conjunto de dados, fornecendo uma visão clara das características principais de cada variável.

Componente	Tipo	Categorico	Valor Categorico	Minimo	Maximo
0	gender	object	Sim	[Female, Male]	Female Male
1	SeniorCitizen	int64	Sim	[0, 1]	0 1
2	Partner	object	Sim	[Yes, No]	No Yes
3	Dependents	object	Sim	[Yes, No]	No Yes
4	tenure	int64	Não	[N/A]	0 72
5	PhoneService	object	Sim	[Yes, No]	No Yes
6	MultipleLines	object	Sim	[Yes, No, No phone service]	No Yes
7	InternetService	object	Sim	[Fiber optic, DSL, No]	DSL No
8	OnlineSecurity	object	Sim	[Yes, No, No internet service]	No Yes
9	OnlineBackup	object	Sim	[Yes, No, No internet service]	No Yes
10	DeviceProtection	object	Sim	[Yes, No, No internet service]	No Yes
11	TechSupport	object	Sim	[Yes, No, No internet service]	No Yes
12	StreamingTV	object	Sim	[Yes, No, No internet service]	No Yes
13	StreamingMovies	object	Sim	[Yes, No, No internet service]	No Yes
14	Contract	object	Sim	[Two year, One year, Month-to-month]	Month-to-month Two year
15	PaperlessBilling	object	Sim	[Yes, No]	No Yes
16	PaymentMethod	object	Sim	[Bank transfer (automatic), Electronic check, Mailed check, Credit card (automatic)]	Bank transfer (automatic) Mailed check
17	MonthlyCharges	float64	Não	[N/A]	18.25 118.75
18	TotalCharges	float64	Não	[N/A]	18.8 8684.8
19	Churn	object	Sim	[No, Yes]	No Yes

**Figura 3.1: Estrutura das Componentes do Dataset**

### 3.3. Entendendo a Cardinalidade

A cardinalidade em ciência de dados refere-se à quantidade de elementos distintos ou únicos presentes em um conjunto de dados, ou seja, ao número de valores únicos que uma variável (ou característica) pode assumir. A Tabela-3.2 mostra os resultados desta avaliação.

A cardinalidade de uma variável é crucial para a escolha do modelo e o processamento dos dados. Em modelos que exigem a transformação de dados categóricos, variáveis com alta cardinalidade, que apresentam muitos valores únicos, podem representar um desafio significativo. Técnicas específicas são necessárias para modelar essas variáveis de modo a garantir que o modelo possa lidar eficazmente com a diversidade dos dados.

Para integrar corretamente variáveis de alta cardinalidade no modelo, técnicas avançadas de codificação, como hashing ou embeddings, podem ser necessárias. Essas técnicas ajudam a reduzir a complexidade e a dimensionalidade dos dados, facilitando a criação de modelos mais eficientes e precisos.

Variáveis de baixa cardinalidade possuem poucos valores únicos em relação ao número total de observações. Por exemplo, em um grande conjunto de dados, uma variável "Gênero" com apenas os valores "Masculino" e "Feminino" teria baixa cardinalidade. Em contraste, uma variável como "ID do Cliente", que geralmente tem um valor único para cada cliente, é um exemplo de alta cardinalidade.

Identificar a cardinalidade das variáveis permite que os cientistas de dados reconheçam a diversidade e a distribuição dos dados, o que é essencial para uma interpretação correta e para tomar decisões informadas durante o processo de análise. Isso assegura que todas as variações relevantes sejam consideradas, levando a insights mais profundos e robustos.

Tabela 3.2: Detalhamento da Cardinalidade

Componente	Variabilidade	Avaliação
customerID	7043	Cardinalidade Presente
gender	2	Cardinalidade Presente
SeniorCitizen	2	Cardinalidade Presente
Partner	2	Cardinalidade Presente
Dependents	2	Cardinalidade Presente
tenure	73	Cardinalidade Presente
PhoneService	2	Cardinalidade Presente
MultipleLines	3	Cardinalidade Presente
InternetService	3	Cardinalidade Presente
OnlineSecurity	3	Cardinalidade Presente
OnlineBackup	3	Cardinalidade Presente
DeviceProtection	3	Cardinalidade Presente
TechSupport	3	Cardinalidade Presente
Continua na Próxima Página		

Tabela 3.2 – Continuação da Página Anterior

Componente	Variabilidade	Avaliação
StreamingTV	3	Cardinalidade Presente
StreamingMovies	3	Cardinalidade Presente
Contract	3	Cardinalidade Presente
PaperlessBilling	2	Cardinalidade Presente
PaymentMethod	4	Cardinalidade Presente
MonthlyCharges	1585	Cardinalidade Presente
TotalCharges	6530	Cardinalidade Presente
Churn	2	Cardinalidade Presente

3.4. Sobre Dados Ausentes

Como resultado do estudo preliminar sobre a faltância de dados, foi detectado que a componente TotalCharges possuía 16% de dados ausentes, para contornar este cenário foi aplicado uma imputação de dados multivariado utilizando o regressor Random Forest, após este procesimento os dados foram preenchidos, como demonstra a Figura-3.2.

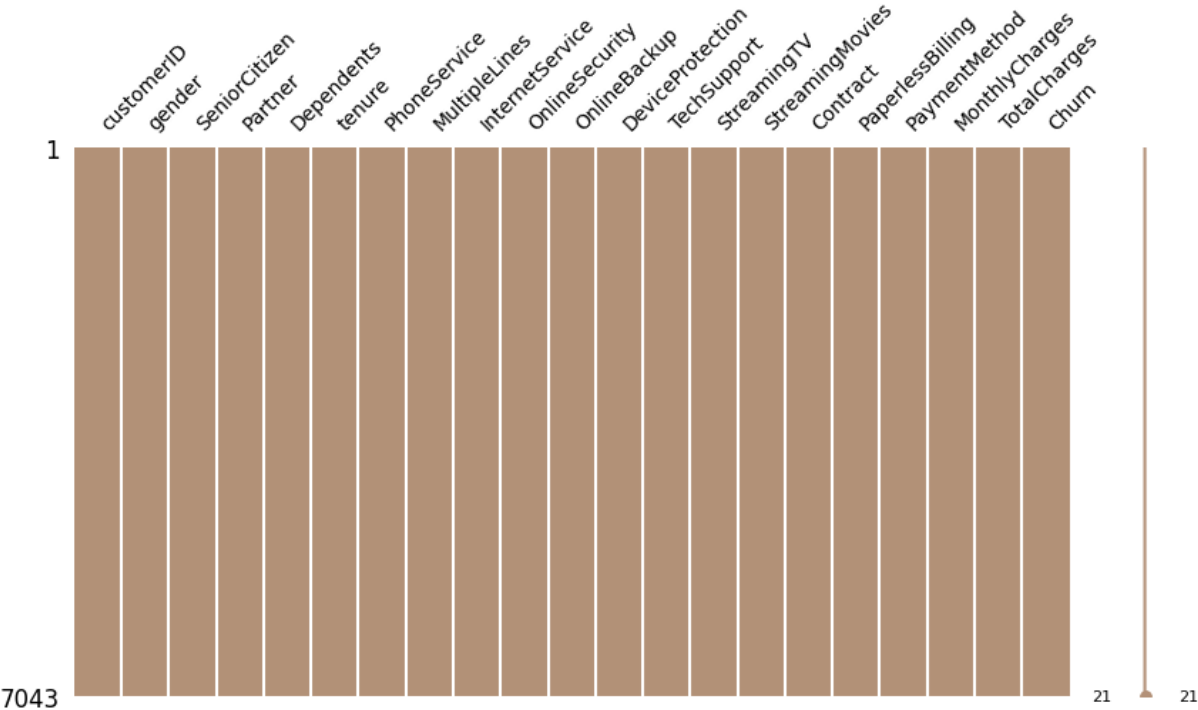
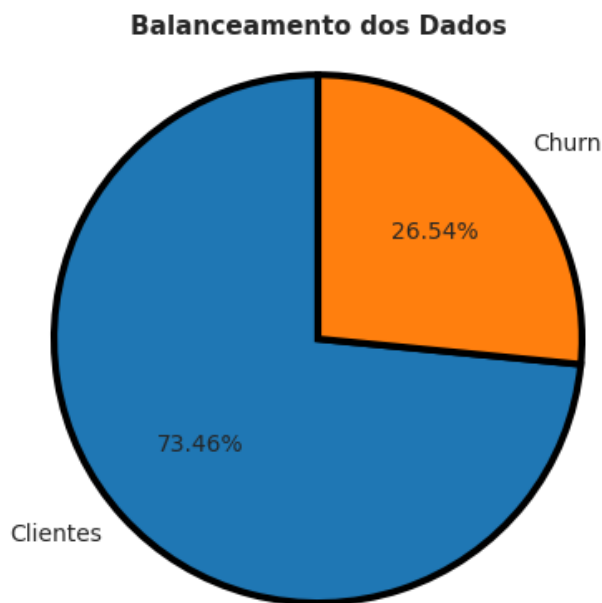


Figura 3.2: Relatório da Faltância de Dados

### 3.5. Sobre Balanceamento de Dados

O desbalanceamento observado na Figura-3.3 entre os clientes ativos e os que cancelaram os serviços é considerado moderado. Como as taxas de churn são geralmente menores do que as taxas de retenção, é comum que muitos datasets relacionados ao churn apresentem uma maior proporção de clientes ativos em comparação com aqueles que deixaram a empresa.

Para garantir que o modelo a ser desenvolvido seja eficaz na identificação de clientes em risco de churn e mantenha alta precisão para a classe majoritária, é crucial implementar técnicas específicas para lidar com esse desbalanceamento durante a fase de modelagem. Estas técnicas ajudam a otimizar a performance do modelo em ambas as classes, proporcionando uma avaliação mais equilibrada e precisa.



**Figura 3.3:** Relatório Sobre o Balanceamento dos Dados

# 4

## Diferenças Entre Pearson e Cramer V

### 4.1. Principais Diferenças Entre os Métodos

A correlação de Pearson é frequentemente utilizada para medir a relação linear entre duas variáveis contínuas, com valores que variam de -1 a 1. Onde o valor 1 indica uma correlação perfeita positiva, enquanto -1 indica uma correlação perfeita negativa, e 0 significa que não há correlação.

Esta medida é particularmente útil quando as variáveis seguem uma distribuição normal e a relação entre elas é linear, facilitando a interpretação e a aplicação em modelos estatísticos.

Em contraste, o Cramer V é uma medida de associação usada para avaliar a força da relação entre duas variáveis categóricas. Derivado do teste qui-quadrado, o Cramer V é aplicável a dados de qualquer tamanho, e seus valores, que variam de 0 a 1, este intervalo indica o grau de associação, sendo 0 uma ausência total de associação.

O Cramer V se destaca ao lidar com variáveis categóricas de várias categorias, oferecendo uma métrica precisa da intensidade da associação. Isso representa uma vantagem significativa sobre outros métodos.

A principal distinção entre a correlação de Pearson e o Cramer V reside no tipo de dados analisados e na maneira como as relações são quantificadas. Enquanto Pearson é restrito a relações lineares e variáveis contínuas, o Cramer V adapta-se a variáveis categóricas de qualquer dimensão.

Esta diferenciação é crucial para a análise de dados, pois orienta a escolha da métrica mais adequada conforme o tipo de variáveis envolvidas. A aplicação correta dessas medidas garante análises e interpretações precisas, revelando a complexidade e a diversidade dos dados reais.

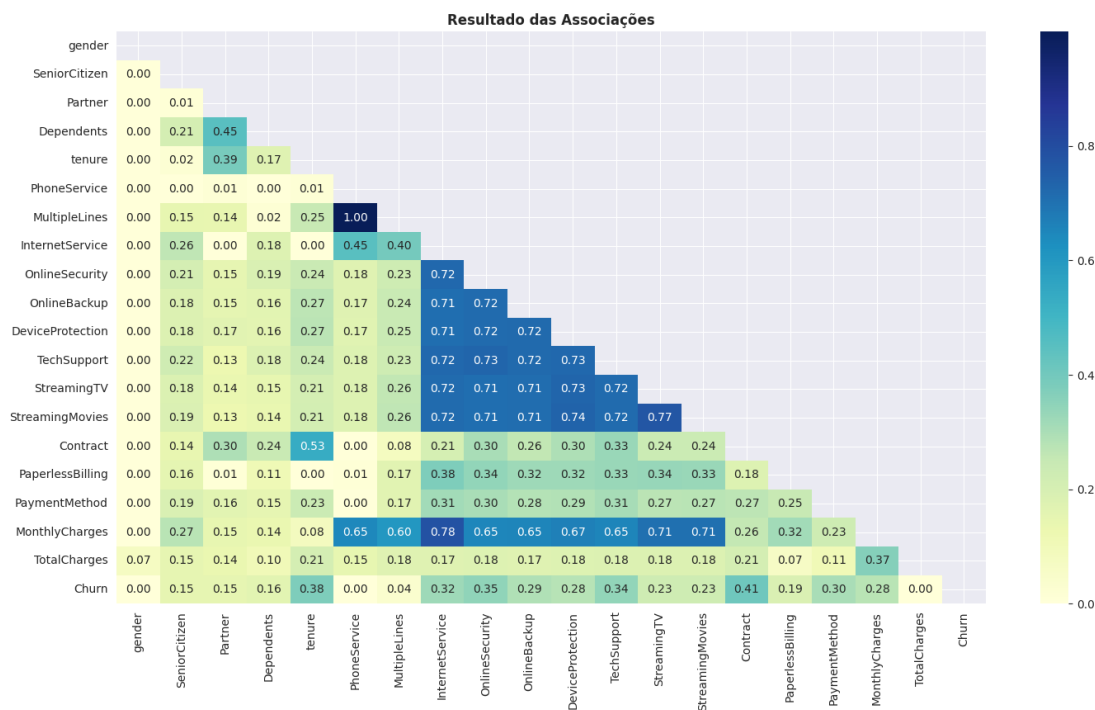


# 5

## Resultados e Conclusões do Uso do Método Cramer V

### 5.1. Resultados Obtidos pelo Cramer V

A Figura-5.1, apresenta os resultados das associações para todas as variáveis do dataset. Apesar da presença de uma escala orientativa, a identificação intuitiva das associações como fortes, moderadas ou fracas ainda é dificultada.



**Figura 5.1:** Resultado de Todas as Associações

Para facilitar a visualização desses resultados, é necessária uma separação clara dessas associações. Essa classificação requer o uso de critérios para definir o tamanho do efeito, que são detalhados na Tabela-5.1.

Interpretação do Tamanho do Efeito	
Tamanho do Efeito (TE)	Interpretação
$TE \leq 0.2$	O resultado é fraco. Embora o resultado seja estatisticamente significativo, os campos são apenas fracamente associados.
$0,2 < TE \leq 0,6$	O resultado é moderado. Os campos são moderadamente associados.
$TE > 0.6$	O resultado é forte. Os campos são fortemente associados..

**Tabela 5.1:** Todos os Resultados

## 5.2. Associações Fortes

As variáveis com associações fortes (valores próximos de 1) devem ser prioritárias, pois indicam uma relação significativa entre as variáveis categóricas.

A Figura-5.2 apresenta o resultado das associações fortes, destacam-se diversas combinações envolvendo o serviço de internet, como InternetService e MultipleLines com uma associação perfeita (1.00).

Outras combinações importantes incluem InternetService e OnlineSecurity (0.72), InternetService e OnlineBackup (0.71), InternetService e DeviceProtection (0.71), InternetService e TechSupport (0.72), InternetService e StreamingTV (0.72) e InternetService e StreamingMovies (0.72). Essas associações indicam que os tipos de serviço de internet estão fortemente correlacionados com diversos outros serviços adicionais oferecidos.

Além disso, observa-se uma forte associação entre MultipleLines e StreamingMovies (0.77), sugerindo que usuários com múltiplas linhas de telefone tendem a usar serviços de streaming de filmes.

Também é relevante a associação entre MonthlyCharges e InternetService (0.78), indicando que os custos mensais estão fortemente relacionados ao tipo de serviço de internet contratado.

Essas associações devem ser exploradas com maior profundidade, pois podem revelar padrões importantes de comportamento do consumidor e oportunidades de otimização dos serviços oferecidos.

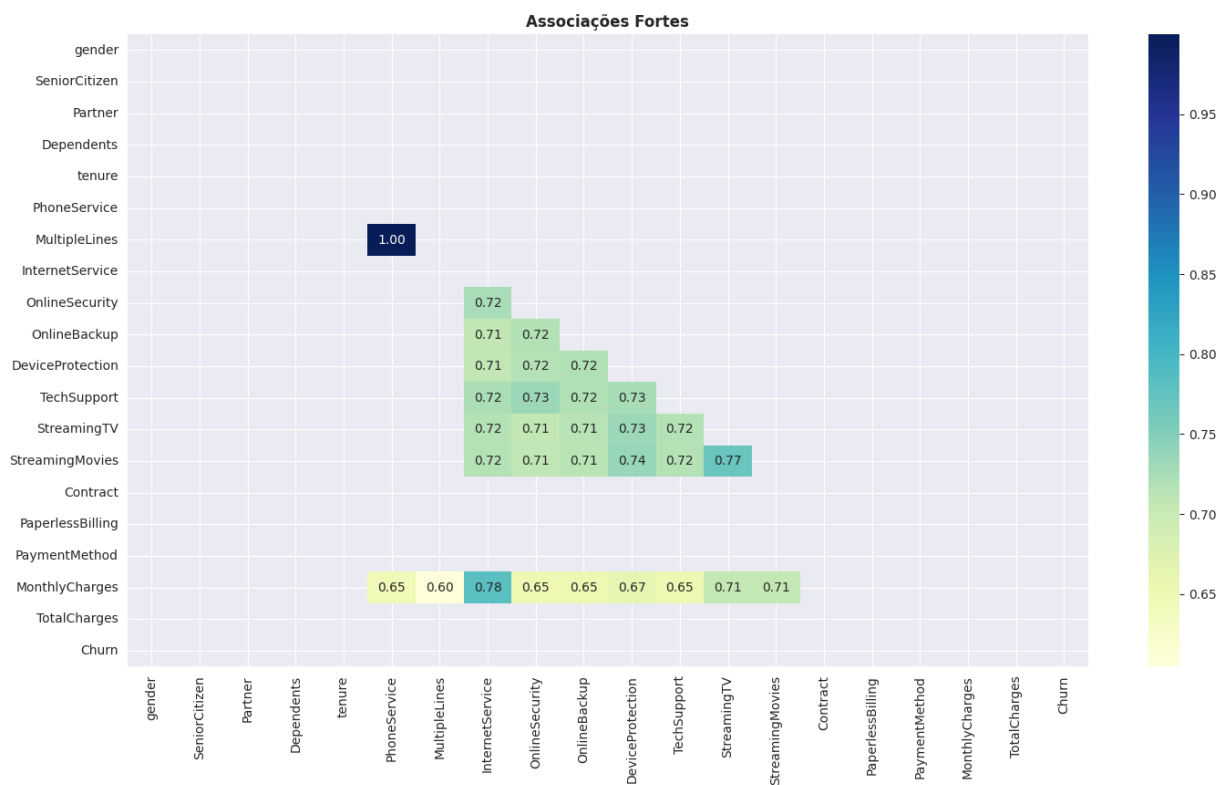


Figura 5.2: Associações Consideradas Fortes

### 5.3. Associações Moderadas

Variáveis com associações moderadas (valores entre 0.3 e 0.5) também são relevantes, embora a relação entre elas seja menos intensa. Essas associações indicam que, embora não sejam tão fortemente correlacionadas, ainda há uma ligação significativa que merece atenção.

A Figura-5.3 apresenta o resultado das associações moderadas, destacam-se algumas combinações importantes, como Contract e tenure (0.53), sugerindo que o tipo de contrato está moderadamente associado ao tempo de permanência do cliente. Essa associação pode ser explorada para entender melhor a fidelidade dos clientes e os fatores que influenciam a duração do contrato.

Outra associação moderada relevante é entre Churn e tenure (0.38), indicando que a taxa de cancelamento está relacionada ao tempo de permanência do cliente, o que pode fornecer insights valiosos sobre os padrões de cancelamento.

Além disso, a relação entre PaymentMethod e PaperlessBilling (0.38) sugere que a escolha do método de pagamento está associada à preferência por faturamento sem papel. Por fim, a associação entre Contract e PaymentMethod (0.37) também destaca uma ligação moderada que pode ser investigada para entender como diferentes tipos de contrato influenciam a escolha do método de pagamento. Essas associações moderadas podem ajudar a identificar áreas de melhoria nos serviços e estratégias para aumentar a satisfação e retenção dos clientes.

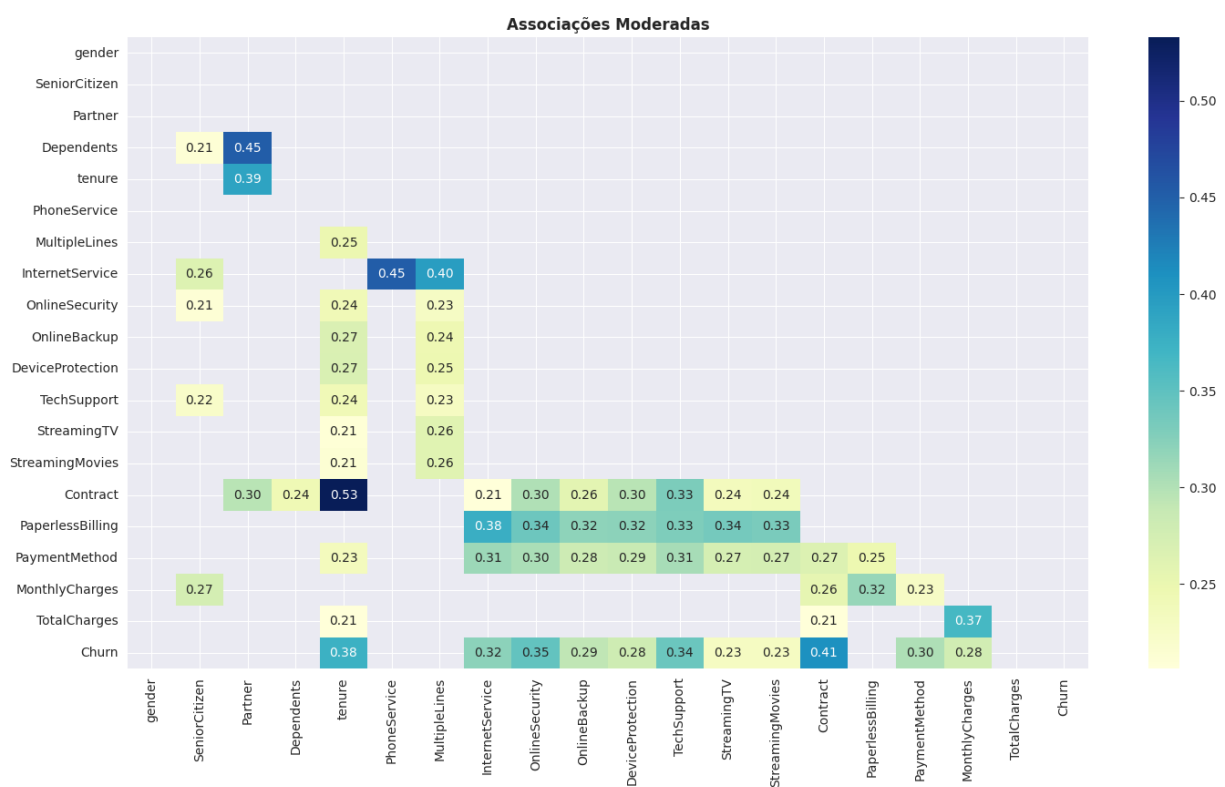


Figura 5.3: Associações Consideradas Moderadas

## 5.4. Associações Fracas

Embora as associações fracas (valores abaixo de 0.3) indiquem uma relação menos significativa, elas podem ser consideradas importantes quando complementam outras variáveis relevantes. Mesmo com uma correlação mais baixa, essas associações fracas podem fornecer insights adicionais que ajudam a construir um panorama mais completo da análise.

A Figura-5.4 apresenta o resultado das associações fracas, por exemplo, a relação entre Partner e Dependents (0.17) indica que ter um parceiro está levemente associado à presença de dependentes, o que pode ser útil na segmentação de clientes para ofertas personalizadas.

Além disso, a associação entre OnlineSecurity e OnlineBackup (0.18) e entre OnlineBackup e DeviceProtection (0.17) sugere que, embora fracas, essas relações podem indicar um padrão de comportamento entre os clientes que optam por serviços adicionais de segurança e proteção de dados.

Essas correlações podem ser exploradas em conjunto com associações mais fortes para desenvolver estratégias de marketing cruzado e identificar oportunidades de venda de pacotes de serviços. Portanto, mesmo associações fracas devem ser consideradas no contexto da análise global para maximizar os insights e a efetividade das estratégias de negócio.

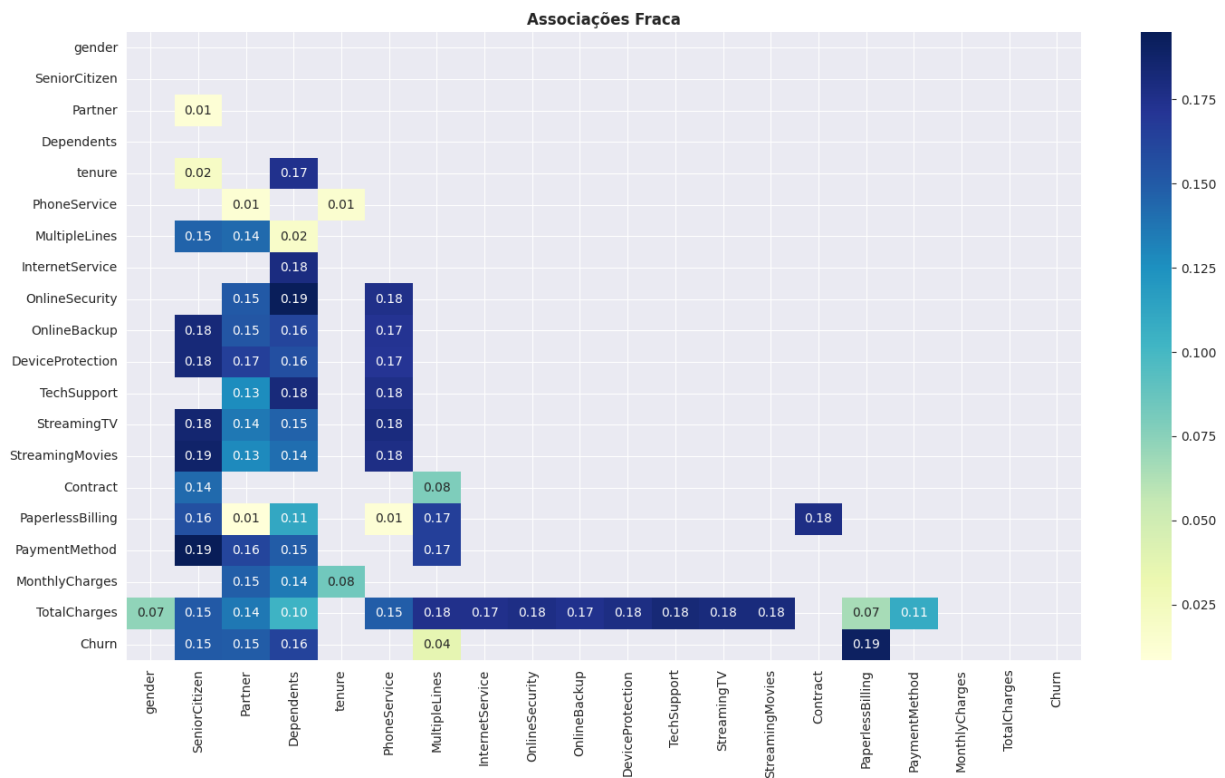


Figura 5.4: Associações Consideradas Fracas

## 5.5. Conclusão das Associações Obtidas

Com base nas Figuras-5.2, 5.3 e 5.4 podemos observar as variáveis com associações fortes e moderadas que devem ser priorizadas em no processo de análise, pois mostram relações mais significativas e potencialmente mais impactantes.

Essas variáveis indicam áreas onde há uma forte ou moderada interdependência entre os dados, o que pode revelar padrões importantes e tendências significativas.

Ao concentrar-se nessas variáveis, você pode identificar fatores cruciais que influenciam o comportamento do cliente e as operações de negócio, permitindo a formulação de estratégias mais eficazes e direcionadas.

Assim, focar nas variáveis relacionadas ao `InternetService`, `MultipleLines`, `tenure`, `PaymentMethod`, `Contract` e `MonthlyCharges` pode proporcionar insights valiosos e influenciar diretamente os resultados da sua análise de dados.

Essas variáveis destacam-se por suas correlações mais pronunciadas, sugerindo que aspectos como o tipo de serviço de internet, a presença de múltiplas linhas telefônicas, a duração do contrato, o método de pagamento e os encargos mensais são determinantes chave na análise do comportamento do cliente e na tomada de decisões estratégicas.

Explorar essas associações permitirá uma compreensão mais profunda dos fatores que afetam a retenção de clientes, a satisfação e o potencial de receita, resultando em uma análise de dados mais robusta e acionável.



# 6

## Procedimento de Cálculo do Método Cramer V

### 6.1. Contexto do Problema

Considere uma tabela de contingência que registra a frequência de duas variáveis categóricas: "Gênero"(Masculino, Feminino) e "Preferência por Produto"(A, B).

A tabela é apresentada da seguinte forma: 20 homens preferem o Produto A, enquanto 30 mulheres preferem o mesmo produto. Por outro lado, 40 homens preferem o Produto B, e 10 mulheres preferem esse produto.

Assim, a tabela de contingência (Tabela-6.1) é organizada com os seguintes valores: Produto A - Masculino (20), Produto A - Feminino (30), Produto B - Masculino (40), Produto B - Feminino (10). O total de observações é de 100, com 60 homens e 40 mulheres, e 50 preferências para cada produto.

Associadamente, podemos elaborar a organização de frequências esperadas (Tabela-6.2) para todas as preferências e gêneros.

Preferência por Produto	Masculino	Feminino	Total
A	20	30	50
B	40	10	50
Total	60	40	100

**Tabela 6.1:** Frequência Observada

Preferência por Produto	Masculino	Feminino	Total
A	30	20	50
B	30	20	50
Total	60	40	100

**Tabela 6.2:** Frequências Esperadas

## 6.2. Equação do Qui-Quadrado

O cálculo do valor do Qui-Quadrado ( $\chi^2$ ) é dado pela equação (6.1), onde  $O_i$  é a frequência observada e  $E_i$  é a frequência esperada.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (6.1)$$

## 6.3. Equação do Cramer V

O cálculo do Cramer V ( $V$ ) é dado pela equação (Equação-6.2), onde ( $V$ ) é o valor do qui-quadrado calculado,  $N$  é o total de observações,  $k$  é o menor número entre (número de colunas) e (número de linhas).

$$V = \sqrt{\frac{\chi^2}{N \cdot (k - 1)}} \quad (6.2)$$

## 6.4. Obtendo o Valor do Qui-Quadrado

Aplicando a Equação-6.1, no problema proposto, obtemos o valor ( $\chi^2$ ) segundo a Equação-6.3.

$$\chi^2 = \frac{(20 - 30)^2}{30} + \frac{(30 - 20)^2}{20} + \frac{(40 - 30)^2}{30} + \frac{(10 - 20)^2}{20} \quad (6.3)$$

$$\chi^2 = \frac{100}{30} + \frac{100}{20} + \frac{100}{30} + \frac{100}{20} \approx 3.33 + 5 + 3.33 + 5 = 16.66$$

## 6.5. Obtendo o Valor do Cramer V

Aplicando a Equação-6.2, no problema proposto, e considerando o número total de observações ( $n = 120$ ) e o número de categorias ( $k_{\min} = \min(2 - 1, 3 - 1) = 1$ ), obtemos o valor ( $V$ ) segundo a Equação-6.4.

$$V = \sqrt{\frac{16.66}{100 \cdot 1}} = \sqrt{0.1666} \approx 0.408 \quad (6.4)$$

## 6.6. Conclusão dos Cálculos Obtidos

O valor do Cramer V obtido pela Equação-6.4 é aproximadamente 0.408, indicando uma associação moderada, conforme mostrado na Tabela-5.1, entre as variáveis "Gênero" e "Preferência por Produto".

Este exemplo demonstra o cálculo do V de Cramer, permitindo avaliar a força da associação entre variáveis categóricas em uma tabela de contingência.

# 7

## Conclusão

O método Cramer V é uma ferramenta poderosa para avaliar a força da associação entre variáveis categóricas em uma tabela de contingência. Ao analisar os dados de forma precisa, este método permite identificar as relações mais significativas, ajudando a direcionar a atenção para áreas que podem impactar diretamente os resultados de um estudo.

A aplicação prática do método foi demonstrada em diversos exemplos, onde foram calculadas as associações entre variáveis como "Gênero" e "Preferência por Produto". Esses cálculos ilustram como o Cramer V pode ser utilizado para extrair insights valiosos a partir de dados categóricos.

Os resultados mostraram que variáveis com associações fortes e moderadas devem ser priorizadas em análises, pois indicam relações significativas e potencialmente impactantes. Por exemplo, variáveis como "InternetService" e "MultipleLines" mostraram associações perfeitas, enquanto outras combinações como "Contract" e "tenure" apresentaram associações moderadas.

Esses achados sugerem que aspectos como o tipo de serviço de internet e a duração do contrato são determinantes cruciais no comportamento do cliente e nas operações do negócio. Além disso, mesmo associações fracas podem complementar a análise ao fornecer contextos adicionais, como visto nas relações entre "Partner" e "Dependents".

Logo o método Cramer V é essencial para compreender a complexidade das interações entre variáveis categóricas. Sua aplicação permite uma análise mais detalhada e precisa, que pode orientar estratégias de marketing, otimização de serviços e melhoria na satisfação do cliente.

Ao analisar as associações mais fortes e moderadas, o cientista de dados pode identificar padrões de comportamento significativos e tomar decisões informadas que beneficiam tanto os negócios quanto os consumidores. A análise contínua e detalhada usando o Cramer V assegura que todas as nuances das relações entre variáveis sejam consideradas, resultando em uma compreensão mais completa e eficaz dos dados.