

O Uso do SPSS (Statistical Package for the Social Sciences) na Ciência Política: uma breve introdução

Mauro Meirelles¹

Resumo

Este texto trata de alguns conceitos básicos de estatística, alguns métodos de análise e sua utilidade para pesquisadores das mais diversas áreas. Em especial, concentramos nosso argumento no modo como o SPSS pode ser utilizado com vistas à realização de um grande número de inferências a partir de certo conjunto de dados quantitativos. Sobretudo, buscamos através destes conceitos básicos e do manuseio do referido software apresentar algumas rotinas que visam a instrumentalizar, minimamente, aqueles que pela seara da análise quantitativa desejam se aventurar. Em especial, nos ocupamos do uso do Teste de Correlação de Pearson, do Teste Qui-Quadrado e a Análise de Regressão Simples e Múltipla.

Palavras-chave: SPSS. Teste de Correlação de Pearson. Teste Qui-Quadrado. Análise de Regressão Simples e Múltipla.

The Use of SPSS (Statistical Package for the Social Sciences) in Political Science: a brief introduction

Abstract

This paper deals with some basic concepts of statistics a few analysis methods and their usefulness for researchers from various fields. In particular we focus our argument on how SPSS can be used with a view to carrying out a large number of inferences from certain set of quantitative data. Above all, we seek through these basic concepts and handling of said software routines equip, minimally, those who wish to venture quantitative analysis. In particular, we are concerned the use of the Pearson Correlation Test, Chi-Square Test and Analysis of Simple and Multiple Regression.

Keywords: SPSS, Pearson Correlation Test. Chi-Square Test. Analysis of Simple and Multiple Regression.

Introdução

A estatística é um campo de saber que gera certa resistência no interior das Ciências Sociais. Contudo, é muito usada entre os sociólogos e os cientistas políticos para realização de análises e inferências quando se utilizam bases de dados que envolvem um grande número de casos, como quando, por exemplo, das análises que envolvem dados censitários, pesquisas de opinião e/ou preferência eleitoral. Neste sentido, enquanto ferramenta para análise da realidade social, a estatística pode ser de extrema valia para todos aqueles que se

¹ Doutor em Antropologia Social e Professor da Unilasalle/Canoas. Outrossim, a confecção deste texto, também contou com o apoio da FAPERGS, através do Edital ARD 2012.

ocupam com o estudo da sociedade em suas múltiplas dimensões.

Ainda nesta direção, e dado o fato de haver certa dicotomia entre as metodologias quantitativas e qualitativas, importa lembrar que, estas, são complementares e que, juntas, auxiliam o pesquisador na busca de respostas a questões por ele levantadas. Deste modo, não nos deteremos nessa discussão, muito menos em elencar os prós e os contras de cada uma destas abordagens metodológicas em vista do fato que, neste texto, parte-se do pressuposto de que cada uma delas tem sua especificidade. E, no caso das abordagens quantitativas, estamos nos referindo à possibilidade que, estas, oferecem ao cientista político – e, também, ao sociólogo – de realizar certo conjunto de inferências a partir de uma base de dados qualquer, testando hipóteses, aceitando-as ou refutando-as.

Assim, se por um lado as metodologias qualitativas fornecem ao pesquisador ferramentas que permitem o aprofundamento de determinadas questões e uma investigação mais acurada de determinados padrões e tendências existentes no conjunto de uma dada população, por outro, as metodologias quantitativas e, em especial, a estatística, permitem ao pesquisador que dela faça uso – por meio da manipulação de uma quantidade infindável de dados – verificar tendências, perceber padrões de comportamento e ação, prever resultados, analisar prevalências e riscos, definir estratégias de ação e elaborar modelos de análise. Tudo isto, a um custo relativamente baixo, de forma rápida e com o uso de aplicações informáticas específicas para as Ciências Sociais, as quais facilitam a elaboração de rotinas e a manipulação de grandes quantidades de dados.

No mercado, existem muitas aplicações informáticas que cumprem essa função e o SPSS é apenas uma delas. Isto posto, no presente texto nos ocuparemos exclusivamente desta aplicação em específico, de suas potencialidades e do modo como o SPSS pode ser utilizado por aqueles que desejam trabalhar com uma abordagem quantitativa com vistas ao estabelecimento de padrões e tendências de comportamento amostral relacionadas a uma população específica.

Desta maneira, para fins didáticos, o presente texto está dividido em duas partes. A primeira em que apresentamos alguns conceitos básicos e sua utilidade para a pesquisa social e a segunda em que algumas rotinas e testes estatísticos são apresentados e aplicados a partir de uma base de dados fictícia cujo intuito é colocar em evidência a análise do teste em questão.

1. Conceitos básicos

Na estatística tem-se que a menor unidade de análise é aquilo que chamamos de dado. Este dado é a menor unidade de contabilização ou contagem de uma dada característica ou atributo. A variação que é observada em um atributo específico que compõe um certo conjunto de dados em relação a uma determinada característica é aquilo que chamamos de variável. Uma variável contém assim um conjunto de dados relativos a um atributo qualquer. A forma como esse atributo pode ser mensurado² pode variar, sendo tanto de natureza qualitativa³ como quantitativa⁴.

A população é todo um conjunto de indivíduos que possui determinadas características comuns, como por exemplo ser brasileiro. Quando fazemos um estudo que se ocupa de toda a população, chamamos este de censo. Quando somente uma parte da população, considerada representativa do todo é extraída desta, tem-se uma amostra. Uma amostra é um conjunto de indivíduos que, em função de determinadas características ou atributos que possuem, são tidos como representativos do todo do qual foram extraídos, ou seja, a população. Deste modo, uma amostra é um construto mental do pesquisador que, a partir de diferentes técnicas de amostragem – que podem ser probabilísticas ou não-probabilísticas, que podem relacionar-se a um recorte intencional ou aleatório –, extrai da população, um conjunto de indivíduos detentores de determinadas características.

Tais características ou atributos podem ser mensuradas de quatro formas distintas⁵. De forma discreta, quando apenas contabilizam um valor bruto, como

² Mensurar refere-se ao ato de determinar a medida, de calculá-la. É um sinônimo de medir. Deste modo, quando falamos de algo que é mensurável, estamos nos referindo a tudo aquilo que pode ser medido. Toda determinação de medida tem como base uma referência de medição padrão que podem ser metros, anos, dias, reais. Muitas vezes, uma variável não pode ser medida através uma variável discreta de modo que, nestes casos, nos utilizamos de categorias ou escalas para isso, como é o caso do uso que fazemos, muitas vezes, da escala e Likert.

³ São aquelas cujas características não possuem valores quantitativos, mas, ao contrário, são definidas por várias categorias e representam uma classificação artificial feita dos indivíduos. São tidas como nominais ou ordinais. São nominais aquelas que não exigem ordenação entre as categorias como é o caso, por exemplo, do sexo, da cor dos olhos etc. São ordinais todas aquelas que exigem uma ordenação entre as categorias, como é caso, por exemplo, da escolaridade do indivíduo, do mês de realização da observação etc.

⁴ São aquelas que têm sua variação medida em uma escala quantitativa e que apresentam valores numéricos que fazem sentido. Tais variáveis podem ser tidas como contínuas ou discretas. As variáveis discretas são aquelas que apenas podem assumir um número finito ou infinito contável de valor de modo que, nesse caso, somente fazem sentido valores inteiros e, geralmente, são resultado de contagens como é o caso, por exemplo, do número de filhos de uma família, do número de cigarros fumados por dia etc. Já as variáveis contínuas são aquelas que assumem valores existentes em uma escala contínua (uma reta real), onde os valores fracionais fazem sentido. Em geral, estas últimas são medidas através de algum instrumento, como por exemplo, o peso de uma criança (balança), a altura de um adolescente (régua), o tempo dedicado ao trabalho (relógio) etc.

⁵ Como colocado em evidência nas notas 2 e 3, a saber: de forma discreta, contínua, nominal e/ou ordinal.

é o caso da idade mensurada em anos. Através de categorias que reúnem indivíduos com certa característica, como é o caso da idade em faixas etárias, uma vez que, não sabemos a idade do indivíduo em si, e sim, a faixa em que este se situa de modo que, neste caso, não temos acesso à medida discreta (idade em anos), pois a idade é aqui medida através de categorias agrupadoras, ou seja, faixas de idade.

Esse último tipo de variável (a categórica) serve ainda para mensurar outras características da população ou amostra em questão que, por serem menos objetivas e não possuem um valor discreto, podem ser mensuradas através de categorias nominais, como é o caso, por exemplo, da preferência eleitoral do entrevistado (ou seja, do partido político que esse possui maior afinidade), ou ainda, se ele se considera, por exemplo, como um eleitor que situa seu voto num espectro ideológico que vai da extrema esquerda à extrema direita, passando por posições intermediárias – que o situam como simplesmente de esquerda ou mais próximo ao centro – ou ao centro. Ou ainda, como um eleitor de centro-direita e/ou extrema direita.

Neste caso, o atributo principal não reside na sua preferência eleitoral (ou no partido como o qual ele simpatiza), e sim no lugar em que ele se situa num gradil ideológico que vai da extrema-esquerda à extrema-direita. Tal característica é, neste caso, difícil e quase impossível de ser mensurada através de números discretos que correlacionam valores e categorias sendo, para isso, na maioria dos casos, utilizada para construção dessas categorias uma Escala de Likert de dez pontos. A Escala de Likert é um tipo de escala de resposta psicométrica que é comumente usada em questionários e pesquisas de opinião e têm por objetivo mensurar o grau de adesão a uma resposta e/ou a afinidade do entrevistado com determinada afirmação ligada a um determinado atributo que se quer mensurar.

É necessário contudo que diferenciemos o que é uma Escala Likert e um item de Likert. A Escala de Likert é a soma das respostas dadas a cada item Likert, uma escala é composta por vários itens Likert. Sua soma é o que mede e/ou mensura a presença e/ou ausência, assim como o grau e intensidade, que um determinado atributo possui para um dado indivíduo, um conjunto destes (amostra) ou para a população como um todo. Um item Likert corresponde, portanto, a apenas uma afirmação à qual o entrevistado responde através de um critério que pode ser objetivo ou subjetivo. Deste modo, em geral, o que se mede a partir de um item Likert é o grau de concordância ou não concordância a uma

afirmação⁶ que quando colocada em relação com outros itens Likert nos fornece um medida confiável (ligada a um modelo teórico dado) em relação a determinado atributo.

Assim dado que essa predileção pela direita ou esquerda não pode ser mensurada simplesmente a partir de números discretos, mas em categorias que englobam determinadas características e escalas, este tipo de variável também é chamada de categórica nominal. Quando há uma ordem entre as categorias, ela também será categórica, contudo, será nomeada como categórica ordinal.

Ainda no que se refere às variáveis tem-se que estas podem ser de três tipos distintos quando analisadas em relação a outra variável, ou seja, quando colocadas em relação como no caso das análises bivariada e multivariada. As análises bivariadas, como o próprio nome já diz, envolvem apenas duas variáveis, e neste caso, as variáveis envolvidas podem ser classificadas na relação entre elas como independentes e dependentes. Variável independente é aquela que apresenta variação no atributo, mas que, esta variação, independe de outra variável. Variável dependente é aquela que tem sua variação determinada por uma outra variável. Vejamos o caso por exemplo das variáveis sexo e renda.

Neste exemplo, quando colocadas em relação, tem-se que o sexo do respondente (masculino/feminino) pode influir na renda por este percebida no mercado de trabalho, contudo, a variação da renda (se de R\$ 1 ou R\$ 100.000), não influi e não pode determinar o sexo do respondente. De modo que, neste caso, podemos dizer que sexo é a variável independente, e renda, a variável dependente. Contudo, aproveitemos essas duas variáveis e coloquemos nessa relação – agora, multivariada, pois envolve mais de duas variáveis – a variável escolaridade.

Novamente, as mesmas perguntas devem ser feitas: A renda pode determinar o sexo do indivíduo? A escolaridade pode definir o sexo do indivíduo e sua renda? E logo veremos que, tanto a escolaridade como a renda não podem definir o sexo do indivíduo. Também nos será possível concluir que parte da variação da renda observada é explicada pela variável sexo e que, esta variação, sofre influência do grau de escolaridade que os indivíduos possuem. E, deste modo, tem-se então que, estas duas últimas são tidas como variáveis dependentes.

⁶ Usualmente são usados cinco níveis de respostas, apesar de alguns pesquisadores preferirem usar sete ou nove níveis. Contudo, eu pessoalmente prefiro as escalas como maior gradil de variação na medida em que, estas, captam tendências mais nuançadas que as primeiras.

Contudo, neste caso, a escolaridade, além de variável dependente, é também, uma variável interveniente, na medida em que, além da renda variar em função do sexo do entrevistado, esta varia, também, de acordo os diferentes níveis de escolaridade existentes no interior de cada um dos estratos da variável sexo. E neste caso, testes específicos tais como o de *Correlação de Pearson*, o *Qui-Quadrado* e a *Análise de Regressão* podem, como veremos mais adiante, respectivamente, nos dizer se há uma correlação positiva ou negativa entre elas, se elas estão associadas, e também, em que medida a variação de uma delas explica a variação de outra.

Quando se tratam de variáveis discretas – como é o caso da idade do respondente, da renda mensal que esse percebe, por exemplo, onde a variação é contabilizada ou mensurada por um número ou unidade de medida discreta, as medidas de tendência central tais como a média, a moda, o desvio padrão e os valores máximos e mínimos, nos permitem avaliar o modo como se dá essa variação e quão homogênea/heterogênea é nossa população e/ou amostra em função da distribuição amostral das respostas dadas pelos entrevistados.

No quadro abaixo apresentamos uma síntese das principais tendências de medida central e o que elas podem nos dizer em relação a nossa amostra.

Quadro 1 : Medidas (tendência central e dispersão), conceito e aplicabilidade

Medida	Conceitualização	Aplicabilidade
Média (Medida de Tendência Central)	É o valor que aponta para onde mais se concentram os dados de uma dada distribuição de casos. As mais usadas em estatística são as médias “aritmética simples” e “ponderada” (com pesos diferenciados).	Indica o valor médio observado em uma dada série de dados.
Moda (Medida de Tendência Central)	É o valor que mais aparece em uma distribuição. Em alguns casos, uma dada amostra ou distribuição de dados pode apresentar mais de um valor modal.	Indica qual (ou quais) são os valores mais comuns em um dado conjunto de dados. Serve para detectar onde se situam o maior número de respostas, independentemente do valor médio observado para as respostas.
Mediana (Medida de Tendência Central)	É o valor que se encontra mais ao meio em uma distribuição de dados qualquer, esta, ordenada em ordem crescente. Este valor separa a metade inferior da amostra, população ou distribuição de probabilidade, da metade superior.	Indica possíveis desvios em relação a média. Serve para detectar possíveis desvios gerados por valores anômalos (ou fora do comum) em uma dada série de dados.
Desvio Padrão (Medida de Dispersão)	É o valor que indica a dispersão dos valores observados em relação à média (ou valor esperado).	Um valor baixo de desvio padrão indica que os dados tendem a estar próximos da média, denotando maior homogeneidade da amostra. Um valor alto de desvio padrão indica que os dados tendem a estar espalhados por uma gama de valores bastante distante da média observada,

		denotando maior heterogeneidade da amostra.
Máximo (Medida de Dispersão)	É o maior valor observado em uma dada série de dados.	Indica o maior valor observado para uma dada série de dados. Junto com a mediana, serve para indicar a existência de possíveis valores anômalos (ou fora do comum) em uma dada série de dados.
Mínimo (Medida de Dispersão)	É o menor valor observado em uma dada série de dados.	Indica o menor valor observado para uma dada série de dados. Junto com a mediana, serve para indicar a existência de possíveis valores anômalos (ou fora do comum) em uma dada série de dados.

Fonte: BARBETTA (2003); AGUIRRE; NIÑO; SIMONETTI (2005).

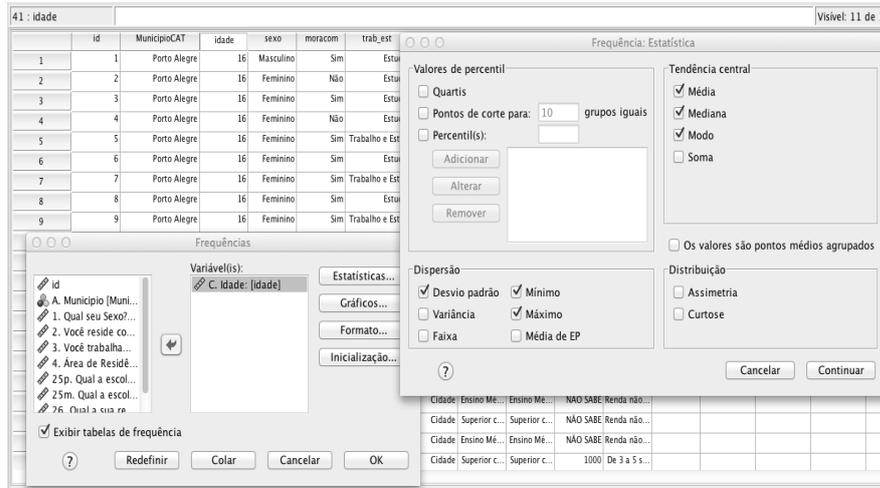
Da mesma forma, para exemplificar o exposto, daqui para diante nos utilizaremos de alguns dados de uma base de dados fictícia adaptada para os fins propostos neste texto, esta, com 430 casos⁷. O software utilizado para realização das rotinas é o SPSS 20 e as saídas apresentadas correspondem aos outputs do mesmo. Isto posto, quando for solicitada alguma rotina para realização de algum teste ou inferência seu caminho estará referendado dentro parênteses com base nos menus da referida versão em português.

Isto posto, para iniciarmos, com uma breve demonstração do exposto, com a base de dados aberta – você pode abrir qualquer base que tenha disponível e que esteja no formato .sav⁸, ou mesmo importar uma outra que tenha sido montada no Microsoft Excel se utilizando das ferramentas disponíveis no *Menu Arquivo (Arquivo / Abrir banco de dados / Nova consulta)* e seguir as instruções apresentadas em tela –, solicite uma frequência (*Analisar / Estatísticas Descritivas / Frequências*) da variável idade, a única em nossa base que apresenta valores discretos e que, por esta razão, nos permite a realização da avaliação das medidas de tendência central (no seu caso, você pode utilizar outras que tenha acesso em sua base de dados e que apresentem valores discretos). Ao fazer isto entre na opção *Estatísticas* e marque as medidas de tendência central: *média*, *mediana* e *moda*, assim como aquelas relativas a dispersão da amostra, a saber: *desvio padrão*, *máximo* e *mínimo*, como mostra a Figura 1.

⁷ Os dados apresentados são reais e correspondem aos dados coletados em pesquisa realizada pelo próprio autor no ano de 2005, apenas renomeados para fins deste texto. Os valores de renda utilizados para categorização da renda correspondem ao salário mínimo da época, a saber: R\$ 260,00.

⁸ Especificamente a terminação .sav refere-se aos arquivos gerados através do SPSS.

Figura 1: Tela de seleção das medidas de tendência central dispersão relativas as frequências da variável idade



Fonte: SPSS 20.

Em seguida clique em continuar e depois em OK e, uma janela com uma saída semelhante a apresentada na Figura 2, lhe será apresentada.

Figura 2: Saída do SPSS 20 com as medidas de tendência central e dispersão solicitadas na Figura 1

Statistics

C. Idade:

N	Valid	387
	Missing	43
Mean		17,78
Median		16,00
Mode		16
Std. Deviation		5,191
Minimum		14
Maximum		52

Fonte: SPSS 20. Base de Dados Fictícia. N= 430.

ANÁLISE DA ESTATÍSTICA: Assim, a partir dessa simples tabela podemos constatar que a idade média dos jovens entrevistados está bem próxima aos 18 anos (mean = 17,78). Contudo, como nos mostra a mediana (median = 16) tem-se que há um pequeno desvio nestes valores para baixo, algo que nos é confirmado, também, pelo valor modal (mode = 16). No que se refere à dispersão tem-se que esta amostra é bastante heterogênea em vista do desvio padrão observado (Std.

Deviation = 5,19), algo que se confirma ao agregar-se à análise os valores mínimos (Minimum = 14) e máximos (Maximum = 52) observados para o conjunto de dados relativos à idade.

Contudo, o uso do SPSS não se restringe somente à análise de dados discretos, produzidos a partir de uma simples análise de frequência, mas também, é extremamente útil quando lidamos com certa quantidade de dados que estão organizados em diferentes variáveis categóricas. E deste modo, permite ao usuário realizar uma série *Tabelas Cruzadas (Analisar / Estatísticas Descritivas / Tabelas de referência cruzada)* ou “cruzamentos” a partir dos quais diferentes hipóteses de pesquisa podem ser testadas através da realização do *Teste de Correlação de Pearson* (Ver Figura 3), do *Teste Qui-Quadrado* (Figura 5) e da *Análise de Regressão (Analisar / Regressão / Linear)*, como veremos no item seguinte.

2. Modelos de análise de dados quantitativos e qualitativos

Quando passamos para um outro conjunto de dados, chamados de categóricos e passamos a estabelecer entre estes uma série de relações hipotéticas, o SPSS tem uma vasta gama de testes a nos oferecer. Contudo, dado o espaço limitado dessa exposição me ocuparei apenas dos três mais comuns e mais utilizados na pesquisa social. São eles: o *Teste de Correlação de Pearson* (p), o *Teste Qui-Quadrado* (X^2) e a *Análise de Regressão* (R^2). Grosso modo, podemos dizer que existe uma ordem lógica entre os testes e que, a medida que aprofundamos o estudo de uma relação entre variáveis, vamos qualitativamente complexificando nossa análise.

2.1. Das finalidades e propósitos do Teste de Correlação de Pearson, do Qui-Quadrado e da Análise de Regressão

Quando levantamos a hipótese de que existe uma relação entre duas variáveis, X e Y, por exemplo, a primeira questão que nos vêm a mente é: como podemos saber isso? Nesse sentido, para responder essa questão, mais do que ficarmos nos debatendo em discussões e especulações filosóficas infundáveis, podemos – e devemos – utilizar o teste de *Correlação de Pearson*, uma vez que sua finalidade reside justamente em testar a existência ou não dessa relação e se, esta, é positiva ou negativa. Algo que, tanto Émile Durkheim (1967) com relação ao suicídio quanto Maurice Duverger (1977, 1980) em relação aos sistemas

políticos partidários, já se ocuparam com muita propriedade, como nos lembram Figueiredo Filho e Silva Júnior (2009), buscando através da análise de diferentes variáveis estabelecer correlações e padrões de normalidade comuns.

A partir de sua análise acerca da dinâmica de diferentes grupos sociais em relação ao seu nível de integração social, Durkheim (1967) constatou, entre outras recorrências, que: 1) a ocorrência de suicídio é maior entre os solteiros, viúvos e divorciados do que entre os casados; 2) as taxas de suicídio são maiores entre pessoas que não tem filhos e, também, entre protestantes do que entre católicos e judeus. Suas conclusões se assentam, sobretudo, no fato de que como ele mesmo demonstrou em seu livro “Le Suicide. Étude de Sociologie” as comunidades católicas tendem a ser um pouco mais integradas que as protestantes, com laços familiares mais próximos de modo que, de forma similar aquilo que ele observa em sua primeira constatação de Durkheim, os católicos mais que os protestantes e judeus, possuem um maior nível de integração social e, portanto, possuem, assim como as pessoas casadas e/ou com filhos, laços de solidariedade mais fortes que lhes fornecem mais motivos para viver.

Já Duverger (1977, 1980), também, correlacionando o modelo organizacional dos partidos, o tipo de partido, o modo de escolha dos dirigentes, a sua taxa de renovação interna, assim como o resultado obtido por estes, conseguiu a partir da identificação dessas tendências e padrões de comportamento estabelecer as bases para um dos modelos teóricos mais caros a Ciência Política, no que se refere ao entendimento da dinâmica dos diferentes sistemas eleitorais e partidários existentes, a chamada “Lei de Duverger”. Tal modelagem de análise, a partir da observação e do estabelecimento de correlações entre tipos de estrutura partidária, tipo de sistema eleitoral e resultados obtidos chegou, entre outras, a três grandes formulações: 1) o sistema majoritário de um só turno tende, na maioria dos casos, ao dualismo dos partidos, com a alternância de grandes partidos independentes no poder; 2) o sistema majoritário de dois turnos tende, na maioria dos casos, a fomentar a existência de um sistema de partidos múltiplos, flexíveis, dependentes e relativamente estáveis; e, 3) a representação proporcional tende, na maioria dos casos, a um sistema de partidos múltiplos, rígidos, independentes e estáveis.

Diante disto, além dos testes destinados à avaliação das medidas de tendência central tratadas no item anterior temos o *Teste de Correlação de Pearson* (p). Esse teste tem por finalidade apenas testar uma hipótese teórica que, de algum modo, intui que haja uma relação entre duas variáveis quaisquer,

contudo, no caso destas possuírem um relação, seja positiva, seja negativa, isto é, de estarem correlacionadas, este teste não informa nem mensura em que medida essas duas variáveis estão associadas.

Para mensurar o grau de associação entre duas variáveis, depois de testada e confirmada nossa hipótese teórica de que havia uma relação entre duas variáveis quaisquer, utilizamos o *Teste Qui-Quadrado* (X^2) para medir o grau dessa associação, como veremos mais adiante em item específico. Contudo, o fato de sabermos que duas variáveis estão correlacionadas, ou seja, que variam juntas, não nos informa em que medida a variação de X explica, por exemplo, a variação de Y, e vice-versa. Neste sentido, para sabermos o poder explicativo de uma variável X e o quanto a variação estimada numa variável Y é explicada pela variação de X, utilizamos a *Análise de Regressão* (R^2).

A análise de Regressão nada mais é, portanto, do que um modelo matemático que estima a relação que se dá entre duas variáveis (X e Y) e a variância que existe em torno dos valores observados dessas duas variáveis (AGUIRRE; NIÑO; SIMONETTI, 2005). Grosso modo, podemos dizer, segundo Barbetta (2003, p.288) que o modelo de regressão linear, em sua formulação mais simples, têm como fim relacionar *uma variável Y, chamada de variável de resposta ou dependente, com uma variável X, denominada de variável explicativa ou independente.*

2.2. Modelos de análise e correlação entre dados quantitativos: o Teste de Correlação de Pearson

Quando trabalhamos com mais de uma variável é normal que, para além de uma análise meramente descritiva, como muito bem nos lembram Figueiredo Filho e Silva Júnior (2009), que, também, estejamos testando uma relação entre variáveis e que, deste modo, estejamos a testar uma hipótese de pesquisa. E que, em função de nossa hipótese queiramos saber se elas estão positivamente correlacionadas e/ou negativamente correlacionadas.

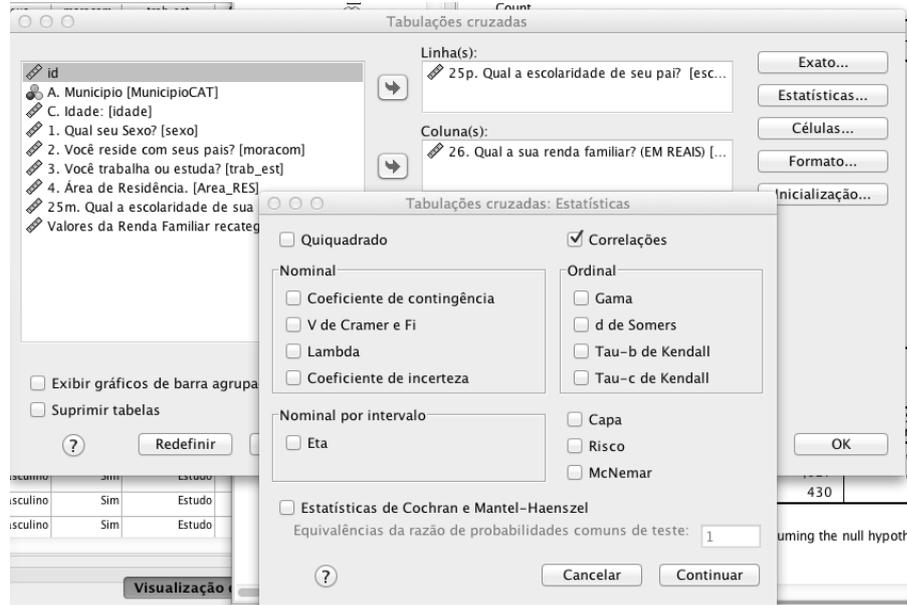
Neste sentido, dizemos que duas variáveis X (Grau de Escolaridade do Pai) e Y (Renda Familiar) estão positivamente correlacionadas (correlação positiva) quando sua variação se dá num mesmo sentido, ou seja, que a medida que aumentam os anos de escolaridade aumenta, também, a renda familiar percebida de modo que a valores pequenos de X correspondem valores pequenos de Y e que, a valores grandes de X correspondem valores grandes de Y. Da mesma forma, dizemos que duas variáveis estão correlacionadas negativamente

(correlação negativa) quando sua variação se dá em sentido oposto, ou seja, que a valores pequenos de X correspondem valores grandes de Y e que, a valores grandes de X correspondem valores pequenos de Y. Vejamos, agora, um exemplo usando o SPSS.

No SPSS vá em *Tabelas de referência cruzada* (**Analisar / Estatísticas Descritivas / Tabelas de referência cruzada**) e na janela que se abrirá a sua frente clique em *Estatísticas* e marque a opção **correlações** como mostra a Figura 3. Em seguida clique em continuar e depois em OK e você terá uma saída igual a apresentada na Figura 4, a partir da qual, você poderá aceitar ou rejeitar sua hipótese de natureza teórica, a qual, intui que a medida em que aumentam os anos de “Escaridade do Pai”, há uma tendência de que o valor da “Renda Familiar” percebida seja maior. Neste caso, como vimos no item anterior tem-se que a variável “Grau de Escaridade do Pai” é tida como a independente e que, esta, pode estar influenciando a variação da “Renda Familiar” percebida pelos nossos depoentes. De modo que temos duas hipóteses a serem testadas.

A primeira hipótese é de que há uma relação entre a variação dos valores observados para as variáveis “Grau de Escaridade do Pai” e “Renda familiar”. A segunda é que esta correlação é positiva e que as duas variáveis variam juntas, ou seja, na medida em que aumentam (ou diminuem) os anos de escolaridade aumenta (ou diminui) o valor da renda percebida pela família. Ainda nesta direção, estaremos considerando como variável independente a variável “Grau de Escaridade do Pai” e, portanto, esta se situará no eixo X (ou na Linha), enquanto a outra, “Renda Familiar” será considerada como dependente e situar-se-á no eixo Y (ou na Coluna), como também é mostrado na Figura 3. Prestemos atenção nas duas figuras que são apresentadas a seguir e, com base no apresentado na Figura 4, testemos nossas hipóteses e rejeitemos (ou não) a hipótese alternativa (nossa hipótese) de que “existe uma correlação positiva” entre as duas variáveis supracitadas.

Figura 3: Tela de seleção das estatísticas e do Teste de Correlação de Pearson para o cruzamento das variáveis “Escaridade do Pai” e “Renda Familiar”



Fonte: SPSS 20.

Figura 4: Saída do SPSS 20, com o Teste de Correlação de Pearson e os valores para ele observados em relação às variáveis “Escaridade do Pai” e “Renda Familiar”

Symmetric Measures					
		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval	Pearson's R	-,124	,022	-2,585	,010 ^c
Ordinal by Ordinal	Spearman Correlation	-,027	,046	-,563	,573 ^c
N of Valid Cases		430			

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.
c. Based on normal approximation.

Fonte: SPSS 20. Base de Dados Fictícia. N= 430.

Com base no apresentado na Figura 4, podemos concluir que existe uma fraca correlação negativa entre as duas variáveis uma vez que o valor observado no *Teste de Correlação de Pearson* ($p = -0,124$) está bastante próximo de 0. E que, em função disto, as duas variáveis tendem a variar de modo independente de modo que, devemos rejeitar nossas duas hipóteses de pesquisa (hipótese alternativa) e aceitar a hipótese nula no caso da primeira, qual seja, de que não há correlação entre as duas variáveis e que, em consequência disto, elas tendem a variar de modo independentemente de modo que, outras explicações ou hipóteses devem ser testadas, tais como a “Ocupação do Pai”, “O Município de residência do depoente e de sua família”, “Se este reside em área rural ou urbana”, entre

outras.

Havendo uma relação entre duas variáveis, identificadas a partir do *Teste de Correlação de Pearson*, podemos, em seguida nos utilizarmos de outros testes estatísticos com vistas a saber o grau de associação existente entre as duas variáveis (*Teste Qui-Quadrado*) e em que medida a variação de uma delas, considerada como independente, explica a variação de outras, consideradas como dependentes (*Análise de Regressão*). E, é com estes dois testes que nos ocupamos nos itens 2.3 e 2.4.

2.3. Modelos de análise e associação entre dados categóricos: o Teste Qui-Quadrado

O *Teste Qui-Quadrado* é um dos mais antigos e mais usados em estatística e permite testar a significância (ou grau de associação) existente entre duas variáveis qualitativas ou categóricas (BARBETTA, 2003). Contudo, para sua utilização é necessário que em cada “casela” da tabela de dados cruzados tenhamos mais de 5 casos, caso contrário, a estatística do teste só será válida para aquelas “caselas” onde esta exigência é cumprida⁹ de modo que, o próprio SPSS dirá que o valor de X^2 é válido somente para “X”% dos dados. Outra observação importante reside no grau de liberdade¹⁰ e no erro amostral¹¹ com o qual se trabalha com vistas à realização do teste de modo que, em geral, adota-se uma margem de erro de 5% com o qual cruzamos com o Grau de Liberdade (df) que nos é dado para a tabela de referência cruzada e para o teste com o SPSS para encontrarmos o valor de X^2 para um df = “X” e um intervalo de confiança de 5%, a partir do qual, assim como no caso do Teste de Correlação de Pearson, aceitamos ou refutamos nossa hipótese alternativa.

A hipótese alternativa, neste caso, assim como no Teste de Correlação de Pearson, é aquela que pressupõe haver associação entre as variáveis, de modo

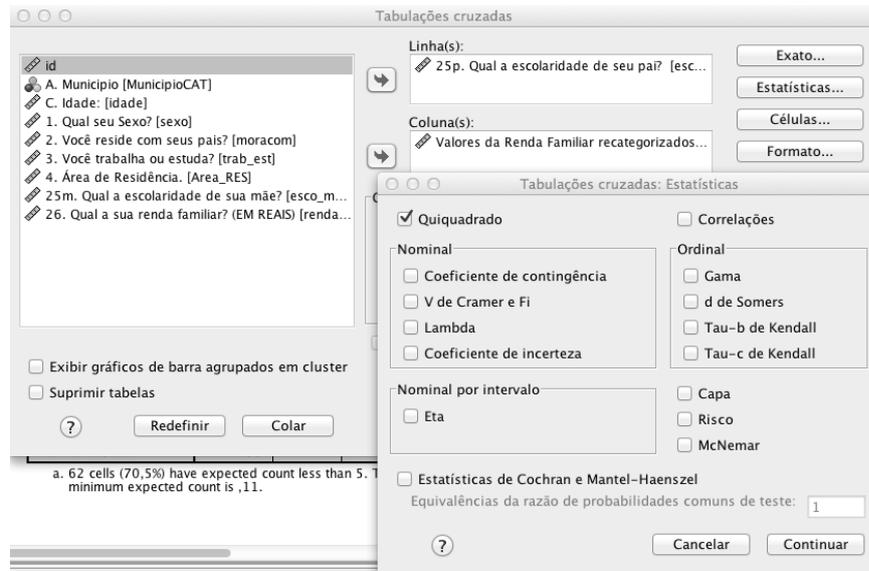
⁹ Para mais detalhes ver Louzã Neto et al. (2007). Ainda nesse sentido, como indica o referido autor, se o número de casos for inferior a 5 em cada casela deve-se usar o *Teste de Fischer*.

¹⁰ O qual corresponde ao número de variáveis independentes que estão sendo somadas. Na prática o Grau de Liberdade refere-se, em estatística, ao número de determinações independentes (dimensão da amostra) menos o número de parâmetros estatísticos a serem avaliados na população e consiste em um estimador do número de categorias independentes num teste particular ou experiência estatística. Para encontrar o Grau de Liberdade de uma distribuição deve-se usar a fórmula $n-1$, onde n é o número de elementos na amostra (que também pode ser representado por $k-1$, onde, k é o número de grupos) quando, no caso deste último se realizam operações com grupos e não com sujeitos individuais.

¹¹ Ou variabilidade amostral é a diferença entre a estimativa da amostra e o parâmetro da população e decorre, grosso modo, da própria noção de amostra na medida em que, quando se recolhe uma amostra, alguma coisa se perde da população de onde foi retirada, embora todos os cuidados tenham sido tomados para garantir a sua fidedignidade, uma amostra pode não ser representativa da população. Isto posto, tem-se que O erro amostral é um erro aleatório na medida em que as estimativas comportam-se aleatoriamente em torno do verdadeiro valor do parâmetro.

que, a hipótese nula é, portanto, a negação desta, qual seja, de que não há associação entre as variáveis testadas. Isto posto, e considerando o banco até agora utilizado no SPSS, gere uma Tabela de referência cruzada (Analisar / Estatísticas Descritivas / Tabelas de referência cruzada) e não esqueça de clicar em “estatísticas” e marcar a opção “Qui-Quadrado”, como mostra a Figura 5.

Figura 5: Tela de seleção das estatísticas e do Teste Qui-Quadrado para o cruzamento das variáveis “Escaridade do Pai” e “Renda Familiar (em faixas de renda)”



Fonte: SPSS 20.

Cabe notar que no caso do Teste Qui-Quadrado, mostrado na Figura 6, que estamos nos utilizando de uma variável categórica (Faixas de Renda), e não de uma variável discreta (Renda em Reais), como no caso do Teste de Correlação de Pearson, como mostrado na Figura 3. Ou seja, se o Teste de Correlação de Pearson testa a existência de uma relação entre duas variáveis discretas, o Teste Qui-Quadrado o faz, também, através do teste de associação – depois de constatada a existência de correlação entre elas – para variáveis categóricas, dada as especificidades do mesmo e a necessidade de se ter pelo menos cinco casos em cada casela como vimos anteriormente (ver nota de rodapé 7).

Figura 6: Saída do SPSS 20, com o teste de associação Qui-Quadrado para o cruzamento das variáveis “Escolaridade do Pai” e “Renda Familiar (em faixas de renda)”

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	144,517 ^a	70	,000
Likelihood Ratio	131,271	70	,000
Linear-by-Linear Association	14,490	1	,000
N of Valid Cases	430		

a. 62 cells (70,5%) have expected count less than 5. The minimum expected count is ,11.

Fonte: SPSS 20. Base de Dados Fictícia. N= 430.

ANÁLISE DA ESTATÍSTICA DO TESTE: Quando olhamos para os dados apresentados na Figura 6, relativos a distribuição normal dos casos, da tabela cruzada “Escolaridade do Pai” e “Renda Familiar (em faixas de renda)” podemos observar que nele consta um grau de liberdade (df=70) e que o valor do qui-quadrado (χ^2) é igual a 144,517. Sendo que neste há uma nota (a) que aparece abaixo da saída, informando que em 70,5% das células existe uma contagem de casos inferior a 5 de modo que, o teste só é válido para 29,5% da amostra, não aplicando-se, assim, a essas células. Ainda nesta direção, importa destacar que, caso o valor de χ^2 seja superior ao observado na “Tabela de distribuição dos valores de χ^2 ” (Anexo 1), a chance de incorrerem em erro ao assumirmos como verdadeira a hipótese alternativa, qual seja, que a variação do grau de “Escolaridade do Pai” está associada a “Renda Familiar (em Faixas de Renda)” percebida e que a medida que cresce o grau de “Escolaridade do Pai”, cresce a “Renda Familiar” é de 0% (Sig=0,000). Nos resta então, comparar o valor apresentado pelo SPSS com o valor da “Tabela de distribuição dos valores de χ^2 ” de modo que: Se $\chi^2 < \text{Valor da Tabela para GL} = 70$, \square rejeito a hipótese alternativa e aceito a hipótese nula, qual seja, que não existe associação entre as variáveis. Ou, ainda, Se $\chi^2 > \text{Valor da Tabela para GL} = 70$, \square aceito a hipótese alternativa e rejeito a hipótese nula. Neste sentido, como o valor esperado para χ^2 na Tabela (Anexo 1) é igual a 51,7393 tem-se que o valor apresentado para o teste qui-quadrado é maior que o observado na “Tabela de distribuição dos valores de χ^2 ” (144,517 > 51,7393) de modo que devemos aceitar a hipótese alternativa e rejeitar a hipótese nula, pois como mostrou o teste χ^2 , as duas variáveis estão associadas e variam junto.

Ainda no que tange as múltiplas possibilidades que o SPSS pode oferecer à pesquisa social, é importante destacar, antes de encerrar esse texto, as potencialidades que a Análise de Regressão pode oferecer ao pesquisador na medida em que, esta, permite a ele descortinar possibilidades não previstas em modelos de análise existentes através da exploração do poder explicativo de uma variável sobre a outra, testando assim, a interveniência de um dado conjunto de variáveis sobre uma em específica, como veremos a seguir.

2.4. Modelos de análise e associação entre dados quantitativos e categóricos: a Análise de Regressão

O modelo de regressão, ou simplesmente análise de regressão, em sua formulação mais simples têm por objetivo, através do estabelecimento de um modelo estatístico-matemático, relacionar uma variável, chamada variável resposta ou dependente (comumente denominada de Y) a uma variável explicativa ou independente (comumente denominada de X) (BARBETTA, 2003; HAIR et al., 2005), como vimos no início deste texto. Assim sendo, como no caso da análise de correlação, a regressão, também, parte de um conjunto de observações pareadas ligadas a duas ou mais variáveis.

Quando nos ocupamos somente com a relação e o poder explicativo de uma variável X sobre Y, ou seja, com duas variáveis somente, dizemos que estamos lidando com um modelo de análise de regressão linear simples. Quando buscamos analisar o poder explicativo ou produzir um modelo de análise que leve em consideração um conjunto de variáveis independentes e/ou intervenientes qualquer (denominadas de X, W, Z) sobre outra dependente (Y) dizemos que estamos lidando com um modelo de regressão múltiplo¹².

Para este fim e com vistas a demonstrar o modo como o referido teste é realizado e como devemos proceder na análise das saídas que nos são oferecidas pelo SPSS, no caso deste item em específico nos utilizaremos de outro banco de dados, onde é apresentado o número de católicos, evangélicos e o número total de cristãos para os estados de São Paulo, Rio de Janeiro e Rio Grande do Sul. Outrossim, para fins didáticos separamos a exposição em dois itens distintos, um

¹² A Análise de Regressão estuda, grosso modo, a relação existente entre uma variável chamada de independente e outras variáveis denominadas de dependentes. Essa relação é representada por um modelo matemático que associa a variação observada em de uma variável independente com outras tidas como dependentes. Quando tal modelo se ocupa somente da relação que se estabelece entre duas variáveis, uma independente e uma dependente estamos nos referindo a um modelo de Regressão Linear Simples. Todavia, se nos utilizarmos de mais de uma variável independente, dizemos que estamos trabalhando com um modelo de Regressão Linear Múltiplo.

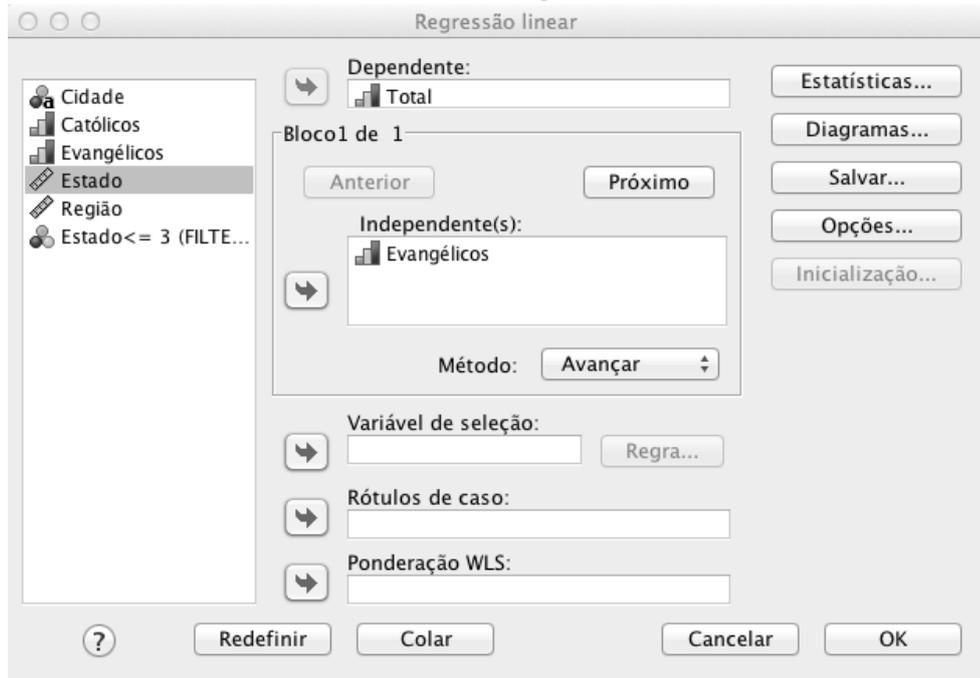
onde nos ocupamos do modelo de regressão linear simples e, outro em que nosso foco é o modelo de regressão linear múltipla.

2.4.1. Regressão Linear Simples

Nesse caso, para fins de nossa análise consideremos a seguinte hipótese, qual seja, de que: A presença de um maior número de evangélicos num estado do que no outro, independente do catolicismo, explica em grande parte, a maior variação observada no número de cristãos¹³. Para isto, dado que tal relação é somente entre duas variáveis, nos utilizaremos aqui somente da realização de um *Teste de Regressão Linear Simples*, no menu *Analisar*, vá em ***Regressão / Linear*** e uma janela igual a apresentada da Figura 7 se apresentará a você, onde você deverá colocar a variável “Total de Cristãos” em nosso caso considerada dependente e, a outra, “Número de evangélicos” considerada como independente e que, de certa forma parece influir de forma diferenciada nos dois estados em questão (ver nota de rodapé 11). Neste sentido, importa destacar que tanto o estado do Rio de Janeiro quanto o de São Paulo são aqueles que possuem, em números absolutos, o maior número de evangélicos e que são eles, em grande parte, o berço de grande parte das novas religiões evangélicas recentes. Dado isto, passemos aos dados.

¹³ Quando comparados os dados referentes aos estados do Rio de Janeiro e Rio Grande do Sul, neste caso em específico.

Figura 7: Saída do SPSS 20, com o teste de Regressão Linear para o cruzamento das variáveis “Total de Cristãos” e “Número de evangélicos”



Fonte: SPSS 20. Base de Dados Cristãos do Brasil. N= 109.

Assim, teremos acesso a duas tabelas com as estatísticas do teste para os estados do Rio de Janeiro (RJ) e do Rio Grande do Sul (RS). Uma relativa ao modelo sumarizado do teste para o RJ (Figura 8) e, outra, para o RS (Figura 10) onde é apresentado o coeficiente R^2 (R Square) que nos mostra em quanto % a variável predictor, considerada como independente, qual seja, o “Número de evangélicos” influencia no “Total de Cristãos” existentes em cada um desses Estados, esta, tida como variável resposta. E, outra relativa ao Teste F de análise de variância do modelo e a probabilidade que eu tenho de estar errado (Sig.) ao aceitar a hipótese alternativa, como mostram as figuras 9 e 11 relativas ao caso do RJ e do RS.

Figura 8: Saída do SPSS 20, com o teste de modelo sumarizado e a estatística do teste para o cruzamento das variáveis “Total de Cristãos” e “Número de evangélicos”, no RJ

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
	Estado = RJ (Selected)			
1	,866 ^a	,749	,735	3,94246

a. Predictors: (Constant), Evangélicos

Fonte: SPSS 20. Base de Dados Cristãos do Brasil. N= 109.

Figura 9: Saída do SPSS 20, ANOVA e teste F de análise de variância do modelo para o cruzamento das variáveis “Total de Cristãos” e “Número de evangélicos”, no RJ

ANOVA^{a,b}

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	790,211	1	790,211	50,840	,000 ^c
	Residual	264,231	17	15,543		
	Total	1054,442	18			

- a. Dependent Variable: Total
 b. Selecting only cases for which Estado = RJ
 c. Predictors: (Constant), Evangélicos

Fonte: SPSS 20. Base de Dados Cristãos do Brasil. N= 109.

Figura 10: Saída do SPSS 20, com o teste de modelo sumarizado e a estatística do teste para o cruzamento das variáveis “Total de Cristãos” e “Número de evangélicos”, no RS.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,490 ^a	,240	,233	6,35363

- a. Predictors: (Constant), Evangélicos

Fonte: SPSS 20. Base de Dados Cristãos do Brasil. N= 109.

Figura 11: Saída do SPSS 20, ANOVA e teste F de análise de variância do modelo para o cruzamento das variáveis “Total de Cristãos” e “Número de evangélicos”, no RS

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1363,509	1	1363,509	33,776	,000 ^b
	Residual	4319,438	107	40,369		
	Total	5682,947	108			

- a. Dependent Variable: Total
 b. Predictors: (Constant), Evangélicos

Fonte: SPSS 20. Base de Dados Cristãos do Brasil. N= 109.

ANÁLISE DA ESTATÍSTICA DO TESTE: Ao olharmos para as Figuras 8 e 9 veremos que o coeficiente de R² é igual a 0,749 e que o resultado do teste F de análise de variância e a chance de eu estar errado ao aceitar a hipótese alternativa é igual a 0,000. Deste modo, podemos dizer que, no caso do Rio de Janeiro, o “Número de evangélicos” explica em 74,9% a variação no número “Total de Cristãos” existentes em cada uma das regiões das diferentes regiões monitoradas (R² = 0,749) e que a chance de eu estar errado ao afirmar que o número de evangélicos existente explica em 74,9% o “Total de Cristãos” é nula (sig=0,000).

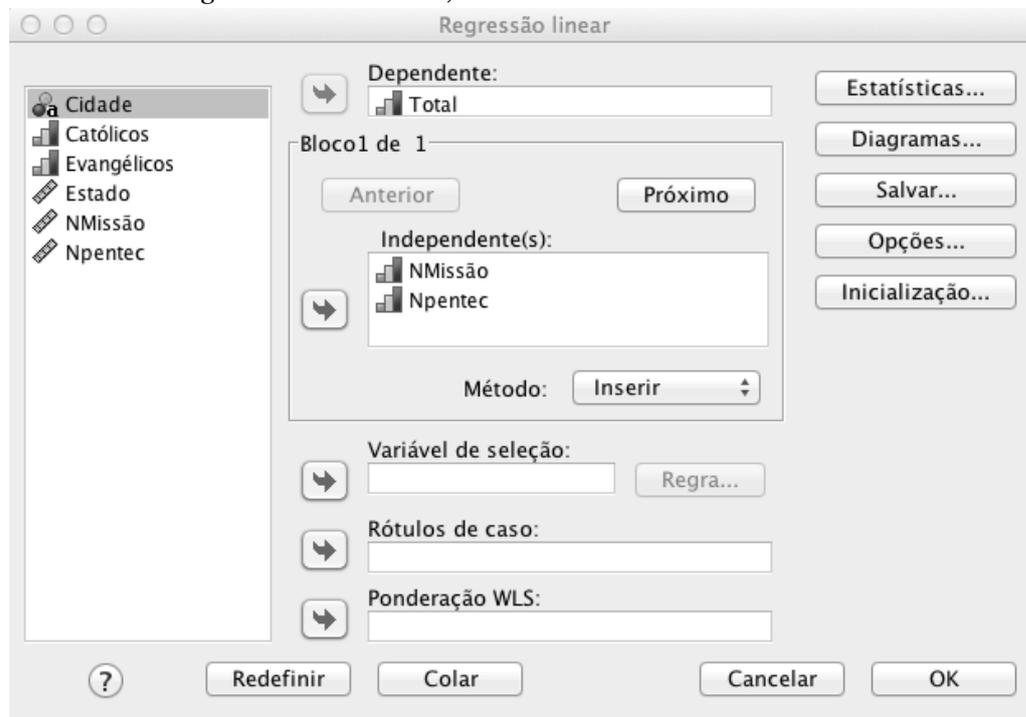
No caso do Rio Grande do Sul, quando olhamos para as saídas apresentadas nas Figuras 10 e 11, e realizamos análise semelhante, olhando para os coeficientes de R^2 ($R^2 = 0,240$) e do teste F ($Sig = 0,000$) podemos dizer que, no caso do RS, a variação no “Número de Evangélicos” nas diferentes regiões do Estado explica em apenas 24% a variação no número “Total de Cristãos”. Algo que, de acordo com outros estudos, têm assento na forte presença das religiões evangélicas históricas no RS, a qual no RJ é bem menor. E, também, ao fato de que foi no RJ que grande parte do movimento de renovação e reavivamento evangélico se fez mais presente através de diferentes grupos pentecostais e neopentecostais de modo que, neste, o número de conversões e adeptos das religiões evangélicas da segunda e terceira onda foi bem maior do que no RS, o qual, se mostra mais resistente a essa tendência do que os estados do centro do país.

2.4.2. Regressão múltipla

Diferentemente da análise de regressão linear simples que se ocupa somente do poder explicativo de uma variável em relação a outra, na análise de regressão múltipla se ocupa de um conjunto de variáveis tidas como predictoras (independentes) e no quanto a variação destas explica a variação e o comportamento da variável resposta (dependente). Neste sentido, a mesma metodologia de análise feita para a *Análise de Regressão Linear Simples* deve ser aplicada para a *Análise de Regressão Múltipla* com a diferença de que, neste modelo estatístico-matemático mais de uma variável será considerada como predictoras (independente) do comportamento da variável resposta (dependente).

Para ilustrar o exposto, neste caso nos ocuparemos de dados relativos somente ao RS, onde, dentre os evangélicos, separamos aqueles relativos às igrejas de missão e aos pentecostais com vistas a buscar-se perceber onde reside a maior influência no RS, se por parte das igrejas históricas ou daquelas denominadas como pentecostais. Dito isto, vamos novamente em ***Analisar / Regressão / Linear*** para solicitar uma nova modelagem para nossa Análise de Regressão, desta vez, considerando não mais o “Número de evangélicos do estado do RS” como variável predictoras, mas sim, as duas variáveis criadas, uma correspondente somente ao “Número de Evangélicos de Missão” em cada uma das regiões do RS e, outra, ao “Número de Evangélicos Pentecostais” nestas mesmas regiões do Estado, como mostra a Figura 12.

Figura 12: Saída do SPSS 20, com o teste de Análise de Regressão Linear Múltipla para o cruzamento das variáveis “Total de Cristãos”, “Número de Evangélicos de Missão” e “Número de Evangélicos Pentecostais”, no RS



Fonte: SPSS 20. Base de Dados Cristãos do Brasil. N= 35.

Feito isto você terá acesso a duas tabelas iguais às que apresentamos a seguir com as estatísticas do teste para o estado do Rio Grande do Sul (RS). Uma relativa ao modelo sumarizado do teste para o estado do RS (Figura 13) onde é apresentado o coeficiente R^2 (R Square) que nos mostra em quanto % as variáveis predictoras, consideradas como independentes, a saber, o “Número de Evangélicos de Missão” e o “Número de Evangélicos Pentecostais”, existentes nas diversas regiões do estado influencia no “Total de Cristãos” do RS, esta última, tida como variável resposta ou dependente da anterior. Da mesma forma, também me é apresentada a chance que eu tenho de estar errado ao aceitar a hipótese alternativa, qual seja, que estas duas variáveis influenciam no “Total de Cristãos” existentes no estado do RS, como podemos observar na Figura 14.

Não esqueça que, assim como no *Teste de Correlação de Pearson*, no *Teste Qui-Quadrado* e na *Análise de Regressão Linear Simples*, quando aceitamos a hipótese alternativa negamos, automaticamente, a hipótese nula, qual seja, de que as duas variáveis tidas como predictoras (ou independentes), neste caso, em nada explicam a variação existente na variável resposta (ou dependente).

Figura 13: Saída do SPSS 20, com o teste de modelo sumarizado e a estatística do teste para o cruzamento das variáveis “Total de Cristãos”, “Número de Evangélicos de Missão” e “Número de Evangélicos Pentecostais”, no RS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,334 ^a	,111	,056	8,84617

a. Predictors: (Constant), Npentec, NMissão

Fonte: SPSS 20. Base de Dados Cristãos do Brasil. N= 35.

Figura 14: Saída do SPSS 20, ANOVA e teste F de análise de variância do modelo para o cruzamento das variáveis “Total de Cristãos”, “Número de Evangélicos de Missão” e “Número de Evangélicos Pentecostais”, no RS.

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	314,014	2	157,007	2,006	,151 ^b
	Residual	2504,149	32	78,255		
	Total	2818,163	34			

a. Dependent Variable: Total
b. Predictors: (Constant), Npentec, NMissão

Fonte: SPSS 20. Base de Dados Cristãos do Brasil. N= 35.

ANÁLISE DA ESTATÍSTICA DO TESTE: Com base nas Figuras 13 e 14 podemos concluir que quase a metade, ou seja, 11,1%, da variação da variável “Total de Cristãos” no RS se deve a variação no “Número de Evangélicos de Missão” e ao “Número de Evangélicos Pentecostais” no RS e que, ao aceitarmos a hipótese alternativa, temos 84,9% de chance de estarmos certos em nossa análise ($R^2=0,111$; $Sig=0,151$). Contudo, ainda não sabemos qual das duas variáveis predictoras exerce maior influência e possui maior peso na definição da variável resposta. Para isso é necessário que realizemos uma nova Análise de Regressão Linear Simples com vistas a se buscar saber se, a maior parte desta variação observada na variável resposta se deve aos evangélicos pentecostais ou aos evangélicos de missão. Assim, ao realizarmos o teste novamente constatamos que, como mostram as Figuras 15, 16, 17 e 18, essa variação nas respostas de deve muito mais a presença dos evangélicos históricos do que aos pentecostais, como consta na literatura sócio-antropológica sobre o tema, uma vez que, a probabilidade de estarmos errados ao afirmamos isso é menor quando assumimos como predictor a variável “Número de Evangélicos de Missão”

($R^2=0,079$; $Sig=0,103$) do que a variável “Número de Evangélicos Pentecostais” ($R^2=0,052$; $Sig=0,186$). Ou seja, que 7,9 da variação da variável resposta responde a primeira variável predictor e 5,2 a segunda. Da mesma forma, segundo o F de análise de variância do modelo para o cruzamento das variáveis “Total de Cristãos” e “Número de Evangélicos de Missão” no RS tem-se que, nossa chance de errar cai praticamente pela metade quando assumimos que é “Número de Evangélicos de Missão” que exerce maior influência sobre a variável dependente, como mostram as Figuras 15 e 16, quando comparadas as Figuras 17 e 18.

Figura 15: Saída do SPSS 20, com o teste de modelo sumarizado e a estatística do teste para o cruzamento das variáveis “Total de Cristãos” e “Número de Evangélicos de Missão” no RS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,280 ^a	,079	,051	8,87068

a. Predictors: (Constant), NMissão

Fonte: SPSS 20. Base de Dados Cristãos do Brasil. N= 35.

Figura 16: Saída do SPSS 20, ANOVA e teste F de análise de variância do modelo para o cruzamento das variáveis “Total de Cristãos” e “Número de Evangélicos de Missão” no RS

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	221,425	1	221,425	2,814	,103 ^b
	Residual	2596,739	33	78,689		
	Total	2818,163	34			

a. Dependent Variable: Total
b. Predictors: (Constant), NMissão

Fonte: SPSS 20. Base de Dados Cristãos do Brasil. N= 35.

Figura 17: Saída do SPSS 20, com o teste de modelo sumarizado e a estatística do teste para o cruzamento das variáveis “Total de Cristãos” e “Número de Evangélicos Pentecostais” no RS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,229 ^a	,052	,024	8,99566

a. Predictors: (Constant), Npentec

Fonte: SPSS 20. Base de Dados Cristãos do Brasil. N= 35.

Figura 18: Saída do SPSS 20, ANOVA e teste F de análise de variância do modelo para o cruzamento das variáveis “Total de Cristãos” e “Número de Evangélicos Pentecostais” no RS

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	147,738	1	147,738	1,826	,186 ^b
	Residual	2670,425	33	80,922		
	Total	2818,163	34			

a. Dependent Variable: Total

b. Predictors: (Constant), Npentec

Fonte: SPSS 20. Base de Dados Cristãos do Brasil. N= 35.

Considerações finais

O que buscamos mostrar nesse texto é como uma aplicação como o SPSS, muitas vezes subutilizada no campo das ciências sociais, pode ser servir a pesquisadores de diferentes áreas na exploração de diferentes hipóteses e problemas de pesquisa. Em especial centramos nosso foco nos testes mais usados – no caso o *Teste de Correlação de Pearson*, o *Teste Qui-Quadrado* e a *Análise de Regressão* – e no modo como um mesmo dado pode ser explorado de diferentes formas. Lembramos, contudo, que SPSS apenas realiza os testes que você pede e que, pensar o instrumento de coleta de dados e as questões que pretende explorar no momento da montagem do instrumento com vistas a sua análise utilizando o SPSS é tão importante quanto o conhecimento dos testes e o modo como se utiliza deles em sua análise.

Assim, se pretendes usar o SPSS em suas pesquisas daqui para diante, é interessante que, antes de montar a pesquisa em si, sempre realize um piloto com poucos casos testando o instrumento e utilizando o próprio programa para ver se

ele está adequado a aquilo que você deseja explorar e se, este, consegue dar conta da análise e testagem de suas hipóteses, sejam elas teóricas ou empíricas.

Sugestões de leitura - Além dos livros referendados no item referências, aqueles que se interessarem podem consultar as seguintes obras, indicamos aqui algumas obras. Especial destaque deve ser dado ao livro Introdução a Estatística de Mario F. Triola, publicado pela LTC e a leitura dos capítulos 2, 7 e 9. Outros dois livros que sugerimos são: a) Estatística aplicada as Ciências Humanas, de Jack Levin, publicado pela editora Harbra, em especial, os capítulos 4, 10 e 11; e, b) A Estatística e suas aplicações, de Rivadavia S. Pereira, em especial os capítulos 7, 8 e 21, este, publicado pela Grafosul. Especificamente, sobre a análise de regressão em específico, sugerimos a leitura do livro “Análise de Modelos de Regressão Linear com aplicações” organizado por Reinaldo Charnet, Clarice Azevedo De Luna Freire, Eugênia Reginato Charnet e Heloísa Bovino, publicado pela Editora da Unicamp em 1999.

Referências

AGUIRRE, César N.; NIÑO, M. Fernanda; SIMONETTI, Eduardo. **Estadística aplicada en las Ciencias Sociales y Humanas**. Posadas: Universidade Nacional de Misiones-Editorial Universitaria, 2005.

BARBETTA, Pedro Alberto. **Estatística aplicada às Ciências Sociais**. 5 ed. rev. Florianópolis: UFSC, 2003.

DURKHEIM, Émile. **Le suicide. Étude de sociologie**. Paris: Les Presses universitaires de France, 1967.

DUVERGER, Maurice. Sociologia dos partidos políticos. In: GURVITCH, G. (Org.). **Tratado de Sociologia**, v. 2. São Paulo: Martins Fontes, 1977.

DUVERGER, Maurice. **Os Partidos políticos**. Brasília: UnB, 1980.

FIGUEIREDO FILHO, Dalson Britto; SILVA JÚNIOR, José Alexandre da. Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r). **Revista Política Hoje**, Recife, v. 18, n. 1, 2009.

HAIR, Joseph et al. **Análise Multivariada de dados**. Porto Alegre: Bookman, 2005.

LOUZÃ NETO, Mário Rodrigues et al. **Psiquiatria básica**. Porto Alegre: Artmed, 2007.

Texto enviado em: 18/04/2014
Aceito em: 03/06/2014

Anexo 1 – Tabela de distribuição dos valores de χ^2

Tabela que fornece valores c tais que
 $P(\chi_n > c) = p$, onde n é o número de graus de liberdade.
 Em destaque na tabela a coluna mais utilizada, ou seja, o intervalo de confiança de 5%.

GL	0,99	0,95	0,9	0,5	0,1	0,05	0,025	0,01	0,005
1	0,0002	0,0039	0,0158	0,4549	2,7055	3,8415	5,0239	6,6349	7,8794
2	0,0201	0,1026	0,2107	1,3863	4,6052	5,9915	7,3778	9,2104	10,5965
3	0,1148	0,3518	0,5844	2,366	6,2514	7,8147	9,3484	11,3449	12,8381
4	0,2971	0,7107	1,0636	3,3567	7,7794	9,4877	11,1433	13,2767	14,8602
5	0,5543	1,1455	1,6103	4,3515	9,2363	11,0705	12,8325	15,0863	16,7496
6	0,8721	1,6354	2,2041	5,3481	10,6446	12,5916	14,4494	16,8119	18,5475
7	1,239	2,1673	2,8331	6,3458	12,017	14,0671	16,0128	18,4753	20,2777
8	1,6465	2,7326	3,4895	7,3441	13,3616	15,5073	17,5345	20,0902	21,9549
9	2,0879	3,3251	4,1682	8,3428	14,6837	16,919	19,0228	21,666	23,5893
10	2,5582	3,9403	4,8652	9,3418	15,9872	18,307	20,4832	23,2093	25,1881
11	3,0535	4,5748	5,5778	10,341	17,275	19,6752	21,92	24,725	26,7569
12	3,5706	5,226	6,3038	11,3403	18,5493	21,0261	23,3367	26,217	28,2997
13	4,1069	5,8919	7,0415	12,3398	19,8119	22,362	24,7356	27,6882	29,8193
14	4,6604	6,5706	7,7895	13,3393	21,0641	23,6848	26,1189	29,1412	31,3194
15	5,2294	7,2609	8,5468	14,3389	22,3071	24,9958	27,4884	30,578	32,8015
16	5,8122	7,9616	9,3122	15,3385	23,5418	26,2962	28,8453	31,9999	34,2671
17	6,4077	8,6718	10,0852	16,3382	24,769	27,5871	30,191	33,4087	35,7184
18	7,0149	9,3904	10,8649	17,3379	25,9894	28,8693	31,5264	34,8052	37,1564
19	7,6327	10,117	11,6509	18,3376	27,2036	30,1435	32,8523	36,1908	38,5821
20	8,2604	10,8508	12,4426	19,3374	28,412	31,4104	34,1696	37,5663	39,9969
21	8,8972	11,5913	13,2396	20,3372	29,6151	32,6706	35,4789	38,9322	41,4009
22	9,5425	12,338	14,0415	21,337	30,8133	33,9245	36,7807	40,2894	42,7957
23	10,1957	13,0905	14,848	22,3369	32,0069	35,1725	38,0756	41,6383	44,1814
24	10,8563	13,8484	15,6587	23,3367	33,1962	36,415	39,3641	42,9798	45,5584
25	11,524	14,6114	16,4734	24,3366	34,3816	37,6525	40,6465	44,314	46,928
26	12,1982	15,3792	17,2919	25,3365	35,5632	38,8851	41,9231	45,6416	48,2898
27	12,8785	16,1514	18,1139	26,3363	36,7412	40,1133	43,1945	46,9628	49,645
28	13,5647	16,9279	18,9392	27,3362	37,9159	41,3372	44,4608	48,2782	50,9936
29	14,2564	17,7084	19,7677	28,3361	39,0875	42,5569	45,7223	49,5878	52,3355
30	14,9535	18,4927	20,5992	29,336	40,256	43,773	46,9792	50,8922	53,6719
35	18,5089	22,465	24,7966	34,3356	46,0588	49,8018	53,2033	57,342	60,2746
40	22,1642	26,5093	29,0505	39,3353	51,805	55,7585	59,3417	63,6908	66,766
45	25,9012	30,6123	33,3504	44,3351	57,5053	61,6562	65,4101	69,9569	73,166
50	29,7067	34,7642	37,6886	49,3349	63,1671	67,5048	71,4202	76,1538	79,4898
60	37,4848	43,188	46,4589	59,3347	74,397	79,082	83,2977	88,3794	91,9518
70	45,4417	51,7393	55,3289	69,3345	85,527	90,5313	95,0231	100,4251	104,2148
80	53,54	60,3915	64,2778	79,3343	96,5782	101,8795	106,6285	112,3288	116,3209
90	61,754	69,126	73,2911	89,3342	107,565	113,1452	118,1359	124,1162	128,2987
100	70,065	77,9294	82,3581	99,3341	118,498	124,3421	129,5613	135,8069	140,1697

Fonte: Tabela extraída e adaptada de:

<http://www.est.ufba.br/mato27/Tabela%20Qui%20Quadrado.pdf>