MÉTODOS E TÉCNICAS DE PESQUISA PARA ECONOMIA CRIATIVA E DA CULTURA

MARCELO MILAN
GUSTAVO MÖLLER
DÉBORA WOBETO
(Orgs.)

ITAÚ CULTURAL

Presidente
Alfredo Setubal

Diretor
Eduardo Saron

NÚCLEO OBSERVATÓRIO

Gerência
Jader Rosa

Coordenação
Luciana Modé

Produção
Ediana Borges
Rafael Gama Figueiredo

NÚCLEO DE COMUNICAÇÃO E RELACIONAMENTO

Gerência
Ana de Fátima Sousa

Coordenação de conteúdo
Carlos Costa

Direção de arte
Yoshiharu Ararkaki

Produção editorial
Luciana Araripe

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor
Carlos André Bulhões Mendes

Vice-Reitora
Patricia Pranke

FACULDADE DE CIÊNCIAS ECONÔMICAS

Diretora
Maria de Lurdes Furno da Silva

Vice-Diretor
André Moreira Cunha

NÚCLEO DE ESTUDOS EM ECONOMIA CRIATIVA E DA CULTURA

Coordenação
Marcelo Milan

Gerência
Gustavo Möller

Coordenação de Ensino e Pesquisa
Débora Wobeto

Projeto gráfico e editoração
Carolina Nobre

Dados Internacionais de Catalogação na Publicação (CIP)

M593 Métodos e técnicas de pesquisa para economia criativa e da cultura / Organizadores Marcelo Milan, Gustavo Möller, Débora Wobeto. – Porto Alegre : UFRGS/FCE; Itaú Cultural, 2022.
recurso digital

Modo de acesso: internet.

ISBN: 978-65-5973-163-3 (recurso eletrônico)

1. Metodologia da pesquisa. 2. Economia criativa. 3. Políticas públicas. 4. Financiamento público. 5. Bens e serviços culturais. I. Milan, Marcelo, organizador. II. Möller, Gustavo, organizador. III. Wobeto, Débora, organizadora. IV. Universidade Federal do Rio Grande do Sul. Faculdade de Ciências Econômicas. Núcleo de Estudos em Economia Criativa e da Cultura V. Itaú Cultural. Núcleo Observatório. VI. Título.

CDU 316.7

Elaborada pelo Núcleo de Publicações da Faculdade de Ciências Econômicas – UFRGS
publicafce@ufrgs.br

MÉTODOS E TÉCNICAS DE PESQUISA PARA ECONOMIA CRIATIVA E DA CULTURA

MARCELO MILAN
GUSTAVO MÖLLER
DÉBORA WOBETO
(ORGS)

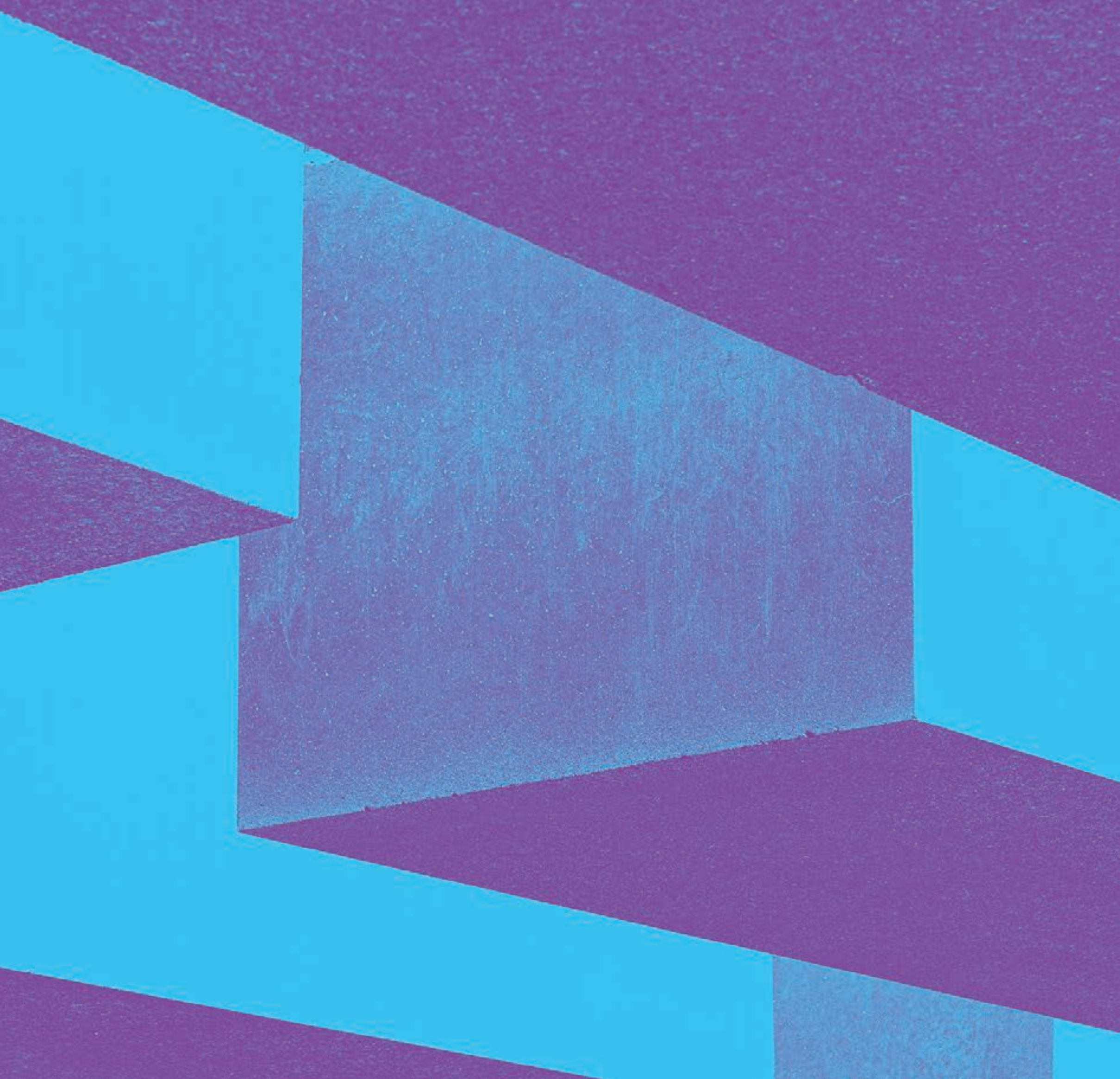
ISBN: 978-65-5973-163-3

9 786559 731633

capítulo 3

FUNDAMENTOS DE ESTATÍSTICA E ECONOMETRIA

SÉRGIO MONTEIRO



1. INTRODUÇÃO

God tirelessly plays dice under laws which he has himself prescribed.
(Albert Einstein, em carta para Paul Epstein em 1945)

Uma das poucas certezas que temos é a de que o mundo caracteriza-se pela incerteza. No campo da economia, mesmo que o conhecimento sobre o comportamento passado de variáveis como crescimento econômico, inflação, desemprego, taxa de juros, taxa de câmbio e taxa de juros possa ser útil para antecipar o futuro, quase que por definição ele é incerto. Na medida em que se reduz o grau de incerteza relacionado ao comportamento de variáveis de interesse, aumenta-se a possibilidade de um manejo adequado dos instrumentos de política pública que podem levar a maiores níveis de bem-estar social em todas as áreas, em especial no campo da economia da cultura.

A estatística oferece a possibilidade de redução da incerteza na medida em que busca identificar padrões e regularidades que ajudem a ordenar o aparente "caos" que se observa na realidade. Na comparação com outras ciências, o uso da estatística na economia guarda uma especificidade quanto à possibilidade de condução de experimentos aleatórios, que são a forma de se identificar esses padrões e regularidades. Um experimento aleatório consiste na repetição de um procedimento em que há a possibilidade de diferentes resultados. A partir da observação dos resultados obtidos no experimento, pode-se identificar tais padrões e regularidades que permitam fazer inferências e previsões sobre a realidade. No estudo dos fenômenos econômicos utilizam-se modelos e dados históricos para compreender o funcionamento do mundo real. A estratégia para lidar com esses fenômenos, já que a repetição não é possível na história, é considerar que os dados são o resultado de algum experimento aleatório realizado pela natureza.

A econometria é o ramo da economia que se propõe a aplicar o instrumental matemático e estatístico de análise aos fenômenos econômicos. É por meio da econometria que se pode testar, de forma empírica ou experimental, as teorias econômicas. O pesquisador utiliza a formulação matemática da teoria para verificá-la por meio de métodos estatísticos. Utilizando-se a econometria, é possível estabelecer relação entre variáveis, medir o efeito de mudanças em uma variável sobre outras e fazer previsões a partir de um conjunto de dados.

Esse capítulo foi elaborado de forma a cobrir o conteúdo básico de estatística e de econometria. A segunda seção trata da estatística descritiva, que fornece um meio de sintetizar as informações contidas em um conjunto de dados. Pode-se dizer que é uma forma de organizar a informação que está dispersa nos dados em estado bruto. A terceira seção refere-se à inferência estatística. Por meio dela é possível tirar conclusões sobre as características de um conjunto amplo de dados a partir de um subconjunto – uma amostra – desses dados. Na quarta seção será apresentado o método econométrico de especificação de modelos, incluindo a estimação da relação entre variáveis e a formulação de previsões.

2. ESTATÍSTICA DESCRITIVA

2.1 TIPOS DE VARIÁVEIS

Suponha que um pesquisador deseja obter informações sobre os funcionários de um banco. Usando informações do Departamento de Pessoal, ele elaborou a tabela a seguir com sete variáveis.

VARIÁVEL	MEDIDA/CLASSIFICAÇÃO
Estado civil	Casado/não casado
Escolaridade	Fundamental/médio/superior
Número de filhos	Número inteiro
Salário	Reais
Idade	Número de anos
Gênero	Feminino/Masculino
Local de nascimento	Capital/Interior

Nesse conjunto de dados, podemos identificar dois tipos de variáveis. As **variáveis quantitativas** são aquelas que permitem alguma forma de mensuração ou de contagem. Na tabela, elas estão representadas pelas variáveis número de filhos, salário e idade. As **variáveis qualitativas** referem-se a um atributo. São desse tipo as variáveis estado civil, escolaridade, gênero e procedência. Observe-se que a variável escolaridade poderia ser representada pelo número de anos de estudo. Nesse caso, seria uma variável quantitativa.

As variáveis quantitativas podem ser **discretas** ou **contínuas.** As discretas formam um conjunto finito ou enumerável de valores. Na tabela são discretas as variáveis número de filhos e idade. As contínuas são aquelas em que os valores situam-se em um intervalo do conjunto dos números Reais e resultam de uma mensuração, como o salário.

As variáveis qualitativas podem ser **ordinais**, quando podem ser ordenadas a partir de algum atributo, como é o caso da escolaridade, ou **nominais**, quando não há ordenação possível. As variáveis estado civil, gênero e local de nascimento se enquadram nessa categoria. Em algumas situações, pode-se atribuir valores numéricos a variáveis qualitativas. Um caso particularmente interessante é o das **variáveis dicotômicas**, para as quais supomos duas situações possíveis, ocorre ou não ocorre, atribuindo-se o valor 0 ou 1 à variável. Por exemplo, a variável estado civil pode ser representada pelos valores 0 para casados e 1 para não-casados. Também poderia ser representada pelos valores 0 para não casados e 1 para casados.

2.2 DISTRIBUIÇÃO DE FREQUÊNCIA

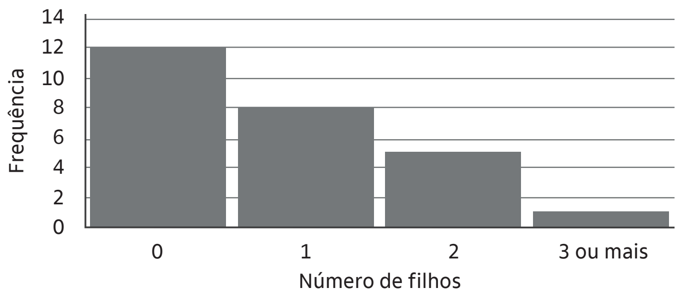
Ao coletar as informações sobre os funcionários da empresa, o pesquisador deseja conhecer o comportamento das variáveis, analisando suas propriedades. Uma forma particularmente útil de descrever as características de um conjunto de dados é a distribuição de frequência. Ela é construída a partir de uma proporção dos dados com relação a determinadas características ou intervalos de valores. Por exemplo, no caso de uma variável discreta, como o número de filhos das mulheres da empresa, suponha que estes sejam os dados:

NÚMERO DE FILHOS	FREQUÊNCIA ABSOLUTA	FREQUÊNCIA RELATIVA (PROPORÇÃO)
0	12	0,4616
1	8	0,3076
2	5	0,1924
3 ou mais	1	0,0384
Total	26	1

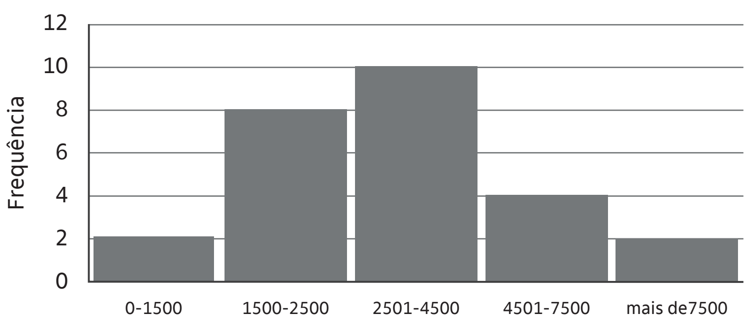
No caso de variáveis contínuas, como é o caso do salário das mulheres da empresa, a proporção é construída a partir de intervalos, conforme tabela a seguir.

SALÁRIO EM R$	FREQUÊNCIA ABSOLUTA	FREQUÊNCIA RELATIVA (PROPORÇÃO)
0 – 1.500	2	0.0769
1.501 – 2.500	8	0.3076
2.501 – 4.500	10	0.3847
4.501 – 7.500	4	0.1539
Mais de 7.500	2	0.0769
Total	26	1

A representação gráfica de uma distribuição de frequência é feita por um histograma. Nele se faz a relação da característica com a frequência. No caso da variável discreta número de filhos, apresentada anteriormente, temos:



E no caso da variável contínua salário das mulheres da empresa, o histograma pode ser visualizado a seguir:



2.3 MEDIDAS DE TENDÊNCIA CENTRAL (POSIÇÃO)

Uma distribuição de frequência oferece uma espécie de resumo das informações contidas em um conjunto de dados. É uma representação que permite uma melhor visualização das características desse conjunto. As medidas de tendência central, também chamadas de medidas de posição, são outra forma de se identificar algumas dessas características. Essas medidas mostram qual é a tendência de agrupamento dos dados em torno de um determinado valor e são representativas da tendência central do conjunto de dados. São empregadas três medidas de tendência central: média, mediana e moda.

2.3.1 Média Aritmética

A média aritmética é simplesmente a razão entre o somatório dos valores da variável e o número de observações. É definida como:

$$\bar{X} = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Por exemplo, se os preços em reais dos ingressos de cinco salas de cinema em uma cidade são (20, 22, 25, 30 e 35), a média será (20+22+25+30+35)/5 = 26,40.

A média só pode ser calculada para variáveis quantitativas. Ela é uma medida de tendência central mais estável, no sentido de que não é tão afetada por um valor específico, mas pode ser afetada por valores extremos.

2.3.2 Mediana

Considerando-se um conjunto de valores ordenados, a mediana é o valor que se encontra no centro. Quando o total de elementos desse conjunto é ímpar, a mediana é o valor central, ou seja, é o valor que divide o conjunto de elementos da série em dois subconjuntos com o mesmo número de elementos. Quando o total de elementos ordenados é par, a mediana é a média aritmética dos dois valores centrais.

Usando o mesmo exemplo dos preços de ingresso de cinema, a série de valores (20, 22, **25**, 30 e 35) tem como mediana o valor 25. Se for inaugurado mais um cinema na cidade cobrando o preço de 10 pelo ingresso, a sequência seria (10, 20, **22, 25**, 30 e 35), e a mediana será (22+25)/2 = 23,50. A mediana é uma medida de tendência central recomendável quando há valores extremos que podem afetar a média. Ela também só pode ser calculada para variáveis quantitativas.

2.3.3 Moda

A moda é o valor que ocorre com maior frequência em um conjunto de valores. Por exemplo, na série (1, 2, 3, 4, 4, 4, 5, 5, 6 e 7), a moda é 4, pois é o valor que aparece mais vezes (três) na série. Quando não há valores repetidos, diz-se que a série é amodal. Quando houver dois ou mais valores com o mesmo número de repetições na série, ela será bimodal, trimodal, e assim sucessivamente.

A moda é a única medida de tendência central que se aplica para variáveis nominais. Ela é recomendável quando se deseja uma medida de posição que identifique o valor típico do conjunto de dados.

2.4 Medidas de dispersão (variabilidade)

As medidas de tendência central não contêm informação sobre quão dispersas em relação ao centro estão os elementos de um conjunto de dados. Ao usar-se apenas uma medida de posição como representativa, deixa-se de considerar parte da informação desse conjunto. Suponha, por exemplo, que os valores a seguir sejam as notas de quatro grupos de alunos:

A: 3, 4, 5, 6, 7
B: 1, 3, 5, 7, 9
C: 5, 5, 5, 5, 5
D: 3, 5, 5, 7

É possível observar que são quatro conjuntos de dados bem diferentes, mas todos possuem média, mediana e moda (nos casos C e D) iguais a 5. Ou seja, essa não é uma informação representativa útil dos conjuntos de dados. Combinadas com as medidas de posição, as medidas de dispersão permitem uma melhor caracterização desses conjuntos. Uma medida simples de dispersão é a **amplitude total**, definida como a diferença entre o maior e o menor valor do conjunto de dados. Entretanto, ela não é uma boa medida porque só leva em conta os valores extremos e não informa quão dispersos estão os dados no interior da série.

As medidas de dispersão mais utilizadas são a **variância** e o **desvio-padrão**, justamente porque incorporam a informação de todas as observações da série de dados. A variância é definida como a média aritmética do quadrado do desvio de cada observação em relação à média do conjunto de dados. Sua fórmula é dada por[1]:

$$s^2 = \frac{\sum_i^n (X_i - \bar{X})^2}{n}$$

Note-se que os desvios em relação à média são calculados para cada observação da série e estão elevados ao quadrado. Ocorre que o somatório dos desvios em relação à média é igual a zero, porque os valores positivos dos desvios acima da média se anulam com os valores negativos dos desvios abaixo da média. Elevando-se os desvios ao quadrado esse problema desaparece, porque todos os valores passam a ser positivos. Entretanto, como a unidade de medida da variável é afetada ao elevar-se os valores ao quadrado, o mais comum é utilizar-se o desvio-padrão como medida de dispersão. Ele é simplesmente a raiz quadrada da variância:

$$s = \sqrt{\frac{\sum_i^n (X_i - \bar{X})^2}{n}}$$

---

[1] Será visto a seguir que o denominador dessa expressão sofrerá uma modificação relacionada ao uso de dados amostrais para estimação.

2.5 Correlação

As medidas de tendência central e de dispersão vistas anteriormente referem-se às características de uma única variável. Quando se considera mais de uma variável, pode haver relação entre elas. Espera-se, por exemplo, que exista relação entre o peso e a altura de um grupo de indivíduos; entre o nível de poluição e a ocorrência de doenças; e entre o nível de renda e o acesso a eventos culturais. É possível medir a relação entre variáveis por meio do coeficiente de correlação linear. Ele é definido em um intervalo de -1 a 1, e indica o grau de relação entre as variáveis. A correlação entre duas variáveis *X* e *Y* pode ser calculada pela fórmula:

$$r = \frac{\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i^n (X_i - \bar{X})^2}\sqrt{\sum_i^n (Y_i - \bar{Y})^2}}$$

Valores próximos de zero indicam que não há relação entre as variáveis. Valores próximos de 1 sugerem que há uma relação positiva muito forte entre as variáveis. Esse pode ser o caso, por exemplo, do peso e da altura de um conjunto de indivíduos. Valores próximos de -1 indicam uma forte relação negativa entre as variáveis. Por exemplo, quanto maior o número de faltas de um estudante, menor tende a ser a sua nota.

3. INFERÊNCIA ESTATÍSTICA

No capítulo 2 foram apresentadas algumas formas de se organizar, resumir e apresentar um conjunto de dados. Nesse capítulo vamos estudar como fazer afirmações e previsões sobre o comportamento das variáveis e sobre a relação entre elas a partir de um subconjunto desses dados.

3.1 POPULAÇÃO E AMOSTRA

Com base em uma parte dos dados, pode-se fazer afirmações sobre o conjunto de dados. Há dois conceitos fundamentais quanto se fala de inferência estatística. Chama-se de **população** o conjunto de todos os elementos sob análise, e chama-se de **amostra** um subconjunto dessa população. Por exemplo, se estamos interessados em saber características da idade dos 1.000 empregados de uma empresa, podemos selecionar uma parte desses empregados, digamos, 50 indivíduos, e a partir deles obter informações que podem ser estendidas ao total de empregados. A idade dos 1.000 empregados constitui a população em análise e a idade dos 50 selecionados constitui a amostra.

Em alguns casos, não conhecemos a população, e ainda assim é possível identificarmos suas características a partir de uma amostra. Suponha, por exemplo, que se deseja estudar qual é o tempo de duração de um determinado equipamento, medido em horas úteis. A população seria constituída por todos os equipamentos que já foram fabricados e pelos que serão fabricados, os quais, obviamente, ainda não existem. Uma amostra de 200 equipamentos fabricados pode ser testada aferindo-se o tempo de uso decorrido até que deixem de funcionar. Assume-se que as características observadas na amostra com relação ao tempo de duração dos equipamentos testados assemelham-se às características da população.

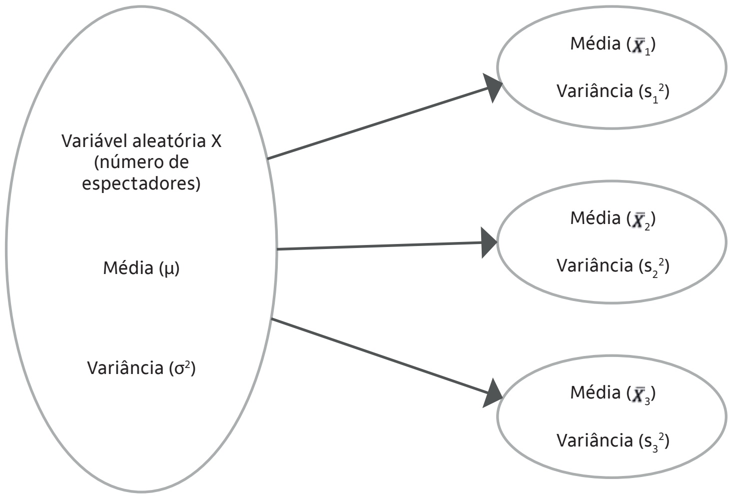
As características de uma população podem ser representadas pelo que chamamos de **variáveis aleatórias**, que são variáveis que assumem valores de acordo com uma distribuição de probabilidade[2]. Essa distribuição fornece informações sobre a população, tais como média, variância, valor máximo, valor mínimo, entre outras. O problema é que, em geral, não se conhece a distribuição da população, por isso recorremos à amostra para obter informações sobre as variáveis aleatórias. Por exemplo, suponha que a nossa variável aleatória em estudo seja o número de espectadores em um teatro. Podemos admitir que essa variável comporta-se de acordo com uma distribuição de frequência com uma certa média e uma certa variância que não são conhecidas. Podemos coletar informações sobre o número de espectadores em um determinado período de tempo e inferir características dessa variável a partir da distribuição da amostra. Para isso precisamos fazer a distinção entre estatísticas e parâmetros.

3.2 ESTATÍSTICAS E PARÂMETROS

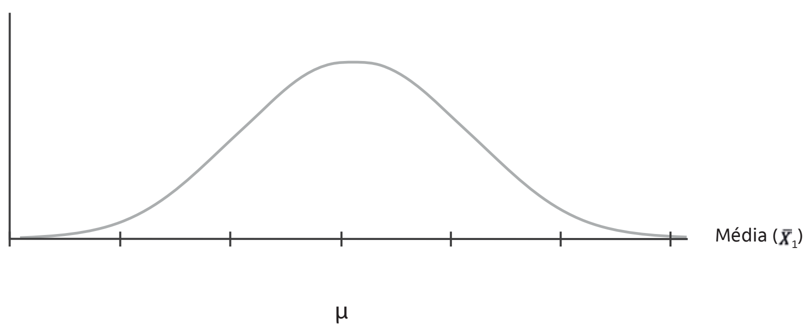
Uma vez extraída uma amostra[3], o que se deseja é, a partir dela, identificar-se características da população. A **estatística** é uma característica da amostra e o **parâmetro** é uma característica da população. Por exemplo, no caso mencionado anteriormente, no qual se deseja estudar o número de espectadores em um teatro, suponha que estamos interessados em uma característica específica dessa variável, como a sua média. A média da população seria o parâmetro e a média da amostra seria a estatística. Normalmente iremos trabalhar com amostras, e as estatísticas mais relevantes são exatamente a média e a variância, definidas na seção anterior. Costuma-se usar uma letra do alfabeto grego para representar os parâmetros e uma letra do alfabeto romano para representar uma estatística. Assim, a média da população é μ e a média amostral é . A variância da população é σ^2 e a da amostra é *s2*. Usando um diagrama de Venn, podemos representar esquematicamente a diferença entre estatística e parâmetro:

[2] Na seção 3.3 trataremos do conceito de distribuição de probabilidade.

[3] Há regras para seleção de uma amostra que constituem um campo próprio de estudo na Estatística, a chamada **Amostragem**. Essas regras não serão tratadas nesse capítulo.



Podemos associar às amostras uma distribuição de frequência para a média, como mostrado na representação a seguir:



3.3 DISTRIBUIÇÕES DE PROBABILIDADE

Nessa seção discutiremos o conceito de distribuição de probabilidade, que pode ser pensada como uma espécie de distribuição de frequência. Duas definições são importantes para chegarmos a esse conceito. A primeira é a definição de

espaço amostral, que é o conjunto de todos os resultados possíveis de um experimento. A segunda é a de **evento**, definido como um subconjunto de um espaço amostral. Vejamos um exemplo. Suponha que o nosso experimento é o de lançar uma moeda não viciada duas vezes. O conjunto de resultados possíveis – o espaço amostral – pode ser expresso por:

1. cara, cara (cara no primeiro e no segundo lançamento);
2. cara, coroa (cara no primeiro e coroa no segundo lançamento);
3. coroa, cara (coroa no primeiro e cara no segundo lançamento);
4. coroa, coroa (coroa no primeiro e no segundo lançamento).

Nesse espaço amostral podemos definir uma variável aleatória *X* tal que ela é igual ao número de caras obtida nesse experimento. Essa variável, portanto, pode assumir os valores 0 (na situação 4), 1 (nas situações 2 e 3) e 2 (na situação 1). Teríamos então a correspondência entre os elementos do espaço amostral e a variável aleatória *X*:

ESPAÇO AMOSTRAL	X (NÚMERO DE CARAS)
cara, cara	2
cara, coroa	1
coroa, cara	1
coroa, coroa	0

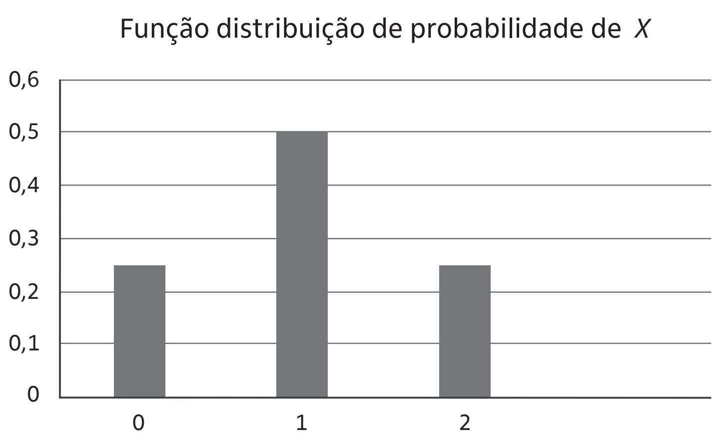
O que faz dessa variável *X* uma variável aleatória é o fato de podermos associar uma probabilidade a cada um dos possíveis resultados. Como temos quatro elementos no espaço amostral, cada um ocorrerá com uma probabilidade de ¼. Isso significa que cada um dos resultados tem uma chance de 25% de ocorrer. Teríamos então:

ESPAÇO AMOSTRAL	X (NÚMERO DE CARAS)	PROBABILIDADE
cara, cara	2	¼
cara, coroa	1	¼
coroa, cara	1	¼
coroa, coroa	0	¼

Definimos uma distribuição de probabilidade a partir da correspondência dos valores que a variável *X* pode assumir (0, 1 e 2) e a probabilidade de ocorrência desses valores. Nesse exemplo, teríamos:

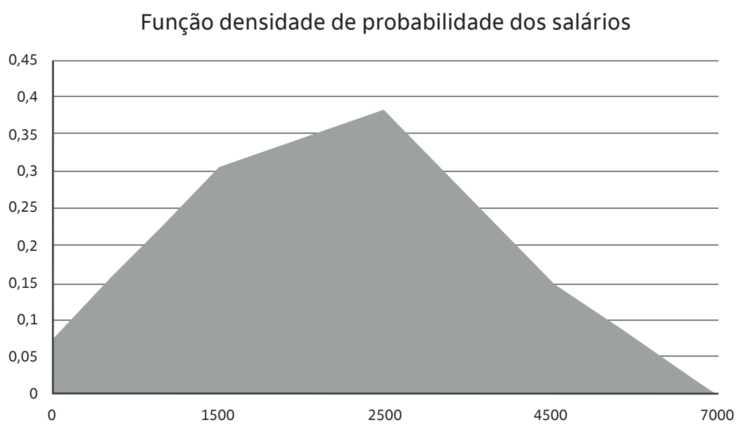
X	f(X)
0	¼
1	½
2	¼
	Total = 1

Essa distribuição nos diz que o resultado de nenhuma cara ocorrerá com uma chance de 25%, o de uma cara com uma chance de 50% e o de duas caras com chance de 25%. A representação gráfica dessa distribuição de probabilidade é mostrada a seguir:



A associação com a distribuição de frequência pode ser pensada facilmente a partir dos exemplos que vimos no capítulo anterior. Na distribuição de frequência associamos os valores à frequência absoluta. No caso da distribuição de probabilidade, eles são associados à frequência relativa. Quando lidamos com variáveis discretas, como a variável número de filhos das mulheres da empresa, cada elemento do espaço amostral é associado a um valor numérico e cada valor pode ser associado a uma probabilidade de ocorrência.

Em situações em que o espaço amostral não é composto por variáveis discretas, como é o caso dos salários das mulheres da empresa, não é possível associar uma probabilidade a um valor específico. Por definição, uma variável aleatória contínua pode assumir infinitos valores. Quando a distribuição de probabilidade é de uma variável contínua, chamamos essa função de densidade de probabilidade. Em uma função desse tipo, a probabilidade só pode ser calculada para um intervalo de valores. Usando como exemplo a distribuição de frequência dos salários, a função densidade de probabilidade seria:



Apresentada dessa forma, não seria possível calcular a probabilidade de que o salário seja de 1.500,00, por exemplo. Podemos calcular a probabilidade de que o salário se encontre em um determinado intervalo a partir da área abaixo da curva. A título de ilustração, definindo o salário como uma variável aleatória X, a probabilidade de que o salário de uma funcionária se encontre entre os valores 1.500 e 1.800 é dada por:

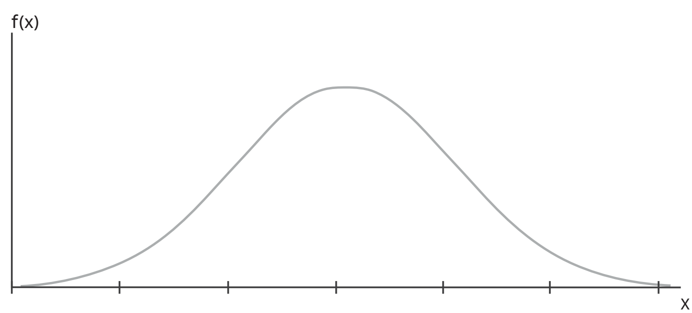
$$P\,(1.500 < X < 1.800) = \int_{1500}^{1800} f(x)dx$$

3.4 DISTRIBUIÇÃO NORMAL

Algumas distribuições contínuas são particularmente importantes e são mais comuns no mundo real. A **distribuição normal**, também conhecida como distribuição gaussiana, é a mais utilizada em estatística porque é capaz de descrever a maioria dos dados. Dizemos que a variável aleatória *X* tem distribuição normal com média μ e variância σ^2 se a sua função densidade de probabilidade é dada por:

$$P\,(1.500 < X < 1.800) = \int_{1500}^{1800} f(x)dx$$

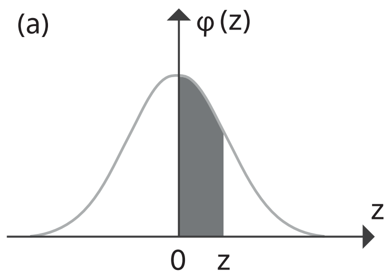
A representação da variável aleatória com essa distribuição é $X \sim N(\mu, \sigma^2)$. A distribuição normal tem o formato de sino e é simétrica em torno da média, que é seu ponto de máximo. A área total abaixo da curva é 1 e a área abaixo da média ou acima da média é 0,5. Ela está representada no gráfico a seguir:



Quando uma variável aleatória *X* tem distribuição normal com média μ e variância σ^2, podemos definir uma variável aleatória *z*, chamada de variável aleatória normal padrão ou reduzida, tal que:

$$z = \frac{X - \mu}{\sigma}$$

A variável aleatória *z* tem média zero e desvio padrão igual a 1, ou seja, $z \sim N(0,1)$. Essa variável permite que se possa fazer, a partir de uma tabela, o cálculo das probabilidades de qualquer intervalo de valores. Considere a área mostrada na figura a seguir. Vejamos como obter as probabilidades a partir da tabela de distribuição normal.



A área representada no gráfico corresponde ao intervalo de 0 até z. Suponha que estamos interessados em encontrar a $P(0 < z < 1{,}73)$. A tabela nos dá essa probabilidade a partir da intersecção da linha 1,7 com a coluna 3. Na tabela vemos que a probabilidade correspondente a esse intervalo é 0,4582. Se quisermos a $P(z > 1{,}73)$, basta subtrair 0,4582 de 0,5, e assim obtemos 0,0418.

A distribuição normal padronizada permite o cálculo de qualquer outra distribuição normal. O cálculo é feito considerando-se:

$$P(x_1 < X < x_2) = P\left[\frac{(x1-\mu)}{\sigma} < \frac{(X-\mu)}{\sigma} < \frac{(x2-\mu)}{\sigma}\right] = P\left[\frac{(x1-\mu)}{\sigma} < z < \frac{(x2-\mu)}{\sigma}\right]$$

Por exemplo, suponha uma variável aleatória $X \sim N(3,16)$. Se quisermos saber a probabilidade de que *X* encontre-se entre 2 e 5, fazemos:

$$P(2 < X < 5) = P\left(\frac{2-3}{4}\right) < z < \left(\frac{5-3}{4}\right) = P(-0{,}25 < z < 0{,}5) = 0{,}0987 + 0{,}1915 = 0{,}2902$$

4. ANÁLISE DE REGRESSÃO

4.1 MODELOS ECONÔMICOS E MODELOS ECONOMÉTRICOS

A ideia básica da teoria econômica é identificar relações ente variáveis de forma a que se possa compreender o funcionamento de um sistema econômico. Postula-se, por meio de funções matemáticas, como se dá a influência de uma variável sobre outras variáveis. Por exemplo, na microeconomia estudamos as funções oferta e demanda, por meio das quais estabelecemos uma relação entre as variáveis quantidade e preço. Formalmente, escrevemos $Q_o = f(p)$ e $Q_d = f(p)$, sendo Q_o a quantidade ofertada, Q_d a quantidade demandada e p o preço. A partir da teoria, assume-se uma relação direta entre quantidade ofertada e preço e uma relação inversa entre quantidade demandada e preço. No campo da macroeconomia supomos que há uma relação direta entre consumo agregado e renda agregada, formalizada como $C = f(Y)$, sendo C o consumo e Y a renda.

As relações também podem ser feitas com mais de uma variável explicativa. Por exemplo, a demanda por um bem é uma função do preço do próprio bem, mas também pode ser influenciada pelo preço de bens substitutos e complementares e pela renda do consumidor. Formalmente teríamos $Q_d = f(p, p_s, p_c, m)$. Em princípio, as relações entre variáveis podem ser estabelecidas a partir de diferentes formas funcionais. A relação pode ser linear, quadrática, logarítmica, etc., e a teoria não estabelece, *a priori*, uma forma específica.

No esforço de compreender o funcionamento de um sistema econômico, usamos modelos. Eles são uma representação simplificada da realidade, mas contêm os aspectos essenciais que a caracterizam. Um **modelo econômico** reúne uma ou mais equações que descrevem o comportamento das variáveis econômicas a partir de suposições sobre a relação entre elas. Supõe-se que há associações regulares que permitem não apenas descrever as características do sistema econômico, mas também fazer previsões sobre o seu comportamento no futuro. O modelo keynesiano de determinação da renda é um exemplo desse tipo. Em uma versão simplificada, temos:

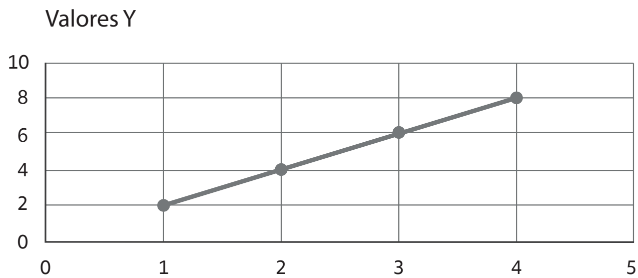
$$Y = C + I + G$$
$$C = a + b.Y$$
$$I = c + d.R$$
$$G = G'$$

Sendo Y a renda, C o consumo, I o investimento, G os gastos do governo, R a taxa de juros, G' é um nível específico de gastos e a, b, c e d são os parâmetros que definem a relação linear entre as variáveis. Nesse modelo, as variáveis Y, C e I são determinadas pelas relações do modelo e por isso são chamadas de **variáveis endógenas**. As variáveis R e G são chamadas de **variáveis exógenas**, porque são determinadas fora do modelo.

A teoria econômica expressa a relação entre as variáveis como funções matemáticas, ou seja, como relações exatas. Considere duas variáveis X e Y, relacionadas por uma função matemática $Y = f(X)$. Suponha que essa função seja $Y = 2X$, ou seja, os valores de Y são o dobro dos valores de X. Considerando-se alguns valores arbitrários de X, podemos obter os valores correspondentes de Y, conforme tabela a seguir:

X	Y
1	2
2	4
3	6
4	8

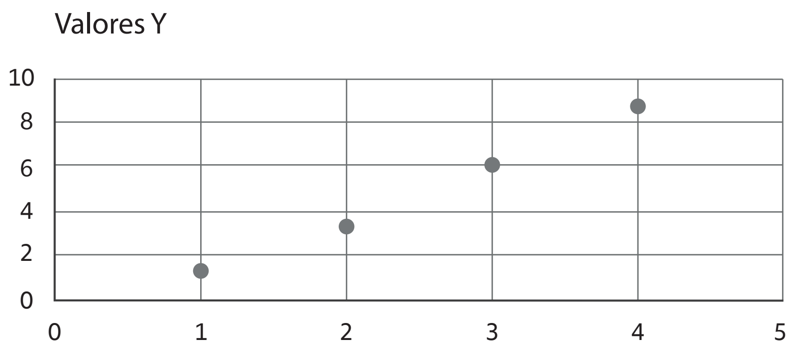
Podemos visualizar a relação entre as variáveis no gráfico:



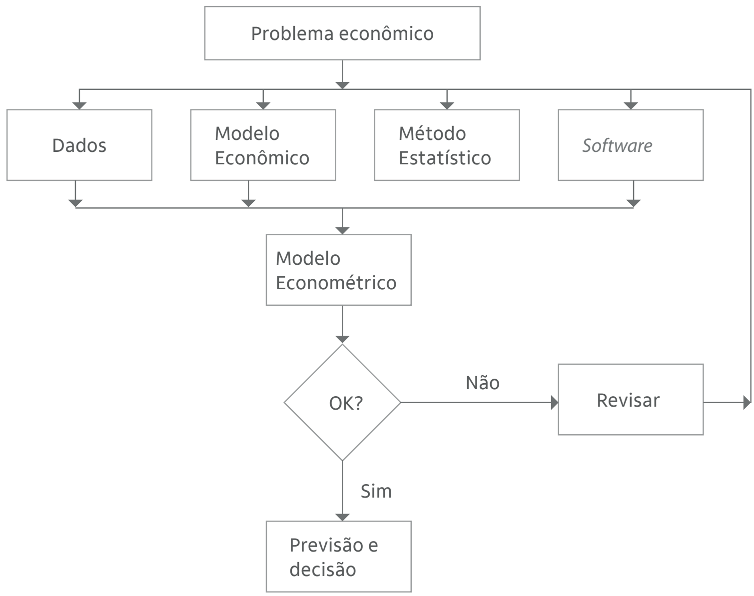
Como as equações do modelo são formuladas a partir de uma simplificação da realidade, é natural que essas relações não sejam exatas, já que a influência de outros elementos na função pode não estar sendo considerada. O mais provável é que tenhamos uma relação do tipo da que está representada na tabela a seguir, na qual não há uma relação exata, mas esta pode ser aproximada por uma função matemática introduzindo-se um termo estocástico chamado de resíduo ou erro.

X	Y	Y=2X	RESÍDUO
1	1,9	2	-0,1
2	3,8	4	-0,2
3	6	6	0
4	8,3	8	0,3

Graficamente, a relação entre as variáveis seria a seguinte:



O conjunto de pares de valores de X e Y correspondem aos pontos dispersos em torno da curva representativa da função $Y = 2X$. Ao introduzirmos o termo de erro, atribuindo-lhe algumas propriedades (que veremos a seguir), dizemos que as variáveis estão relacionadas por meio de um **modelo econométrico**. Nesse caso, a representação do modelo é dada por $Y = f(X) + u$, sendo u o erro. No exemplo acima, seria $Y = 2X + u$. Em econometria, buscamos estimar o modelo que descreva adequadamente a realidade. Tudo começa com a especificação do modelo em sua formalização matemática, segue com a coleta dos dados referentes às variáveis do modelo, continua com a estimação dos coeficientes que definem a relação entre as variáveis e termina com os testes que permitirão avaliar a qualidade do modelo em termos de sua consistência interna e de sua adequação à realidade. Heij et al. (2004, p. 3) resumem a modelagem econométrica no esquema a seguir:



Vejamos um exemplo de uma estimativa econométrica a partir desse esquema. Suponha que estamos interessados em saber qual é a relação entre a renda do consumidor e a demanda por ingressos de cinema, como propôs Cameron (1990). Esse é o problema econômico a ser resolvido. Suponha que há dados disponíveis sobre o total de ingressos vendidos (TI), o preço do ingresso (P), o nível geral de preços (IGP) e a renda dos consumidores (m). Utilizamos o modelo econômico de demanda, postulando que o total de ingressos vendidos é uma função das demais variáveis. Assumindo que todas são variáveis aleatórias, podemos utilizar um *software*, como o *Eviews*, para estimar o modelo econométrico $TI = f(P, IGP, m) + u$. Com base em alguns testes, podemos avaliar se o modelo deve ser refeito ou se pode ser usado para previsão e tomada de decisão. A partir dele, podemos identificar, por exemplo, a elasticidade renda da demanda por ingressos, o que permite calcular a variação no total de ingressos vendidos decorrente de uma variação da renda.

4.2 TIPOS DE DADOS

Ao analisarmos um problema econômico, podemos nos deparar com três tipos de dados, dependendo da forma como estão organizados. Quando estão em sequências observadas em intervalos de tempo, diz-se que os dados constituem uma **série temporal**. Por exemplo, dados como o preço diário de fechamento das ações de uma determinada empresa, a taxa mensal de desemprego, o PIB trimestral ou a taxa anual de inflação formam séries temporais. Quando os dados são compostos por observações em um determinado ponto do tempo, diz-se que os dados são um **corte transversal** (*cross-section*). Esse é o caso de dados sobre a taxa de desemprego dos países da Europa em 2021, o PIB dos países em desenvolvimento nesse mesmo ano ou a idade dos alunos quando ingressam em um determinado curso. Quando combinam-se séries temporais com cortes transversais, diz-se que os dados formam um **painel**. Combinações como o PIB dos países em desenvolvimento nos últimos vinte anos, a taxa mensal de desemprego dos países europeus em 2021, ou o público de teatros de São Paulo nos últimos quatro trimestres são exemplos de dados desse tipo. Ao longo desse curso, trabalharemos apenas com dados em séries de tempo e em corte transversal.

4.3 REGRESSÃO LINEAR SIMPLES

O modelo de regressão linear simples estabelece uma relação entre duas variáveis, uma variável dependente e uma variável independente, também chamada de variável explicativa. A partir da teoria econômica, podemos supor a existência de relações entre variáveis, como vimos na seção 4.1, do tipo $Y = f(X)$. A influência de X sobre Y pode se dar de inúmeras maneiras, cada uma delas correspondendo a uma diferente forma funcional. A relação linear é a forma funcional mais simples, por isso começaremos por ela. Assumimos que a relação entre Y e X se dá por meio de uma equação do tipo:

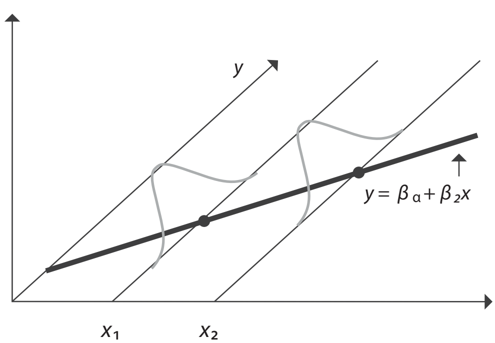
$$Y = \beta_0 + \beta_1 X + u$$

Nessa relação, *Y* é a variável dependente, *X* é a variável explicativa, *u* é o erro e β_0 e β_1 são os parâmetros que correspondem ao intercepto e à declividade da função. Com base nos dados de *Y* e *X*, o modelo de regressão linear permite estimarmos β_0 e β_1 , que mostram de que forma se dá a influência de *X* sobre *Y*. O erro, ou resíduo, é incluído no modelo como uma forma de absorver os efeitos de outras influências que não as de *X* sobre *Y*, como explicado na seção 4.1.

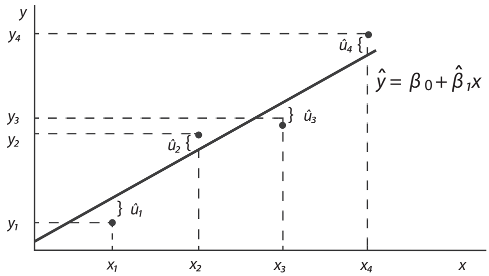
Para que possamos fazer inferências a partir do modelo, é necessário que se estabeleçam algumas hipóteses sobre a distribuição de probabilidade do erro. São elas:

- $E(u) = 0$ (o erro tem média igual a zero)
- $E(u^2) = \sigma^2$ (a variância do erro é constante)
- $E(u_i\ u_j) = 0,\ i,j = 1, 2, 3...$ (a covariância do erro é nula, ou seja, resíduos são independentes entre si)

A intuição para inclusão do erro é relativamente simples. Ele pode assumir diferentes valores, de acordo com a distribuição de probabilidade, mas terá média zero, como é mostrado no gráfico a seguir.



Nosso objetivo é estimar β_0 e β_1 a partir dos dados amostrais. Basicamente, desejamos ajustar uma reta aos dados de X e Y. Pode ser observado no gráfico a seguir que o melhor ajuste se dá na medida em que se reduz o resíduo, que é a distância entre o valor observado de Y e a reta de regressão.



Os resíduos podem assumir valores negativos ou positivos, de acordo com a observação estar abaixo ou acima da reta. Sabemos que o somatório dos resíduos não é uma boa medida da dispersão em torno da reta, uma vez que será sempre igual a zero (a média do resíduo é zero). Entretanto, se elevarmos os resíduos ao quadrado, assim como fizemos com o desvio em relação à média no cálculo da variância, o somatório será uma medida dessa dispersão. Diferentes valores de β_0 e β_1 representam diferentes retas de ajustamento e, portanto, diferentes valores para a soma do quadrado dos resíduos.

Para obtermos a reta com o melhor ajustamento, usaremos o princípio dos mínimos quadrados. A ideia é escolher os valores de β_0 e β_1 que minimizam a soma dos quadrados dos desvios. Isso pode ser feito por meio de cálculo, minimizando esse somatório:

$$\sum_{i=1}^{n} \hat{u}_i{}^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1)^2$$

Efetuando essa minimização, chegaremos aos valores estimados de β_0 e β_1. São eles:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Para verificarmos se há um bom ajustamento da regressão aos dados, utilizamos o **coeficiente de determinação (R^2)**. Ele faz uma decomposição da variação total de Y em duas partes: uma parte que é "explicada" pela regressão e uma parte que não é "explicada", que corresponde ao resíduo. O R^2 é obtido por:

$$R^2 = \frac{SQE}{STQ} = 1 - \frac{SQR}{STQ}$$

Sendo:

STQ = Soma total dos quadrados = $\sum_{i=1}^{n}(Y_i - \bar{Y})^2$;

SQE = Soma dos quadrados explicada = $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$; e

SQR = Soma dos quadrados dos resíduos = $\sum_{i=1}^{n}(e_i)^2$

O coeficiente de determinação pode ser interpretado como a proporção da variação em Y que é explicada pela variação em X. Ele pode assumir valores entre 0 e 1. Quando R2 é igual a zero, ou muito próximo de zero, diz-se que X não tem nenhum poder explicativo sobre Y. Quando R^2 é igual a 1, X e Y encontram-se sobre a reta de regressão, portanto os resíduos são iguais a zero.

Dadas as variâncias das estimativas dos parâmetros e a estimativa da variância do erro, podemos realizar testes de significância e construir intervalos de confiança para as estimativas dos parâmetros. Veremos isso na próxima seção.

4.4 TESTES DE HIPÓTESES SOBRE A SIGNIFICÂNCIA DAS ESTIMATIVAS DOS PARÂMETROS

Quanto maior for a dispersão das observações, mais difícil será ajustar a reta de regressão aos dados. Essa dispersão pode ser avaliada a partir do desvio-padrão da regressão, que será utilizado para obtermos o desvio-padrão da estimativa dos parâmetros e, a partir deles, fazermos os **testes de hipótese** sobre a significância dos coeficientes linear e angular da reta de regressão. Chamaremos de s_e o desvio-padrão da regressão. Ele é dado por:

$$s_e = \sqrt{\frac{\sum_{i=1}^{n}(Y - \hat{Y})^2}{n-2}}$$

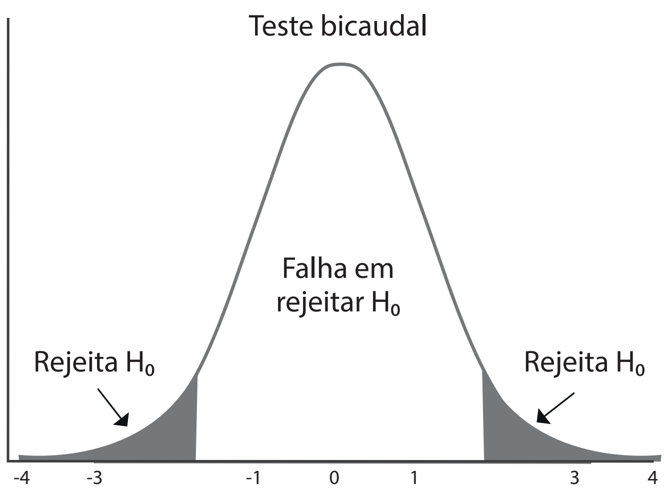
Sendo *Y* a variável dependente, $\hat{Y}$ a estimativa de *Y* e *n* o número de observações.

Com base no desvio-padrão da regressão, calculamos o desvio-padrão das estimativas, que nos dão uma medida da dispersão das estimativas em torno do verdadeiro valor de $\hat{\beta}_0$ e $\hat{\beta}_1$. Esses desvios são dados por:

$$s_{bo} = s_e\sqrt{\frac{\sum X_i^2}{n\sum(X_i - \bar{X})^2}} \quad \text{e} \quad s_{b1} = \frac{s_e}{\sqrt{(X_i - \bar{X})^2}}$$

Com base nessas estatísticas, podemos formular os testes de hipótese sobre a significância dos coeficientes. Basicamente, vamos testar a hipótese de que $\hat{\beta}_0$ e $\hat{\beta}_1$ são significativamente diferentes de zero. Observe que se rejeitarmos a hipótese de que $\hat{\beta}_1$ é igual a zero, constatamos que há uma relação entre as variáveis *Y* e *X*.

Em geral vamos usar dados amostrais para fazer os testes, portanto não conhecemos o desvio-padrão da população. Usamos o desvio-padrão amostral como estimativa do verdadeiro desvio e, em função disso, nos testes de hipótese usaremos a distribuição *t* de Student, considerando-se *n-2* graus de liberdade[4], em vez da distribuição normal. Ela também tem a forma de sino e é simétrica em torno da média. O teste de hipóteses é feito comparando-se o *t* tabelado com o *t* calculado. No teste bicaudal, a hipótese nula H_0 não é rejeitada se o *t* calculado estiver situado no intervalo -*t* tabelado e +*t* tabelado, considerando-se α como o nível de confiança, conforme mostra a figura a seguir. Se o *t* calculado estiver fora do intervalo, rejeita-se a hipótese nula.



O teste de hipóteses para a estimativa de $\hat{\beta}_0$ pode ser especificado como:

H_0: $\beta_0 = 0$
H_1: $\beta_0 \neq 0$

Sendo o *t* calculado a partir de $t = \frac{\hat{\beta}_0}{s_{bo}}$. Por exemplo, em uma amostra com 10 observações e considerando-se um nível de significância de 5%, o *t* tabelado é igual a ±2,306. Se o *t* calculado for igual a 2, por exemplo, a hipótese nula não

[4] Os graus de liberdade de uma distribuição referem-se à informação que pode ser obtida na amostra, e estão relacionados com o tamanho da amostra e com o número de parâmetros a ser estimado. No caso da regressão simples, para uma amostra de tamanho *n* queremos estimar dois parâmetros, por isso consideramos *n-2* graus de liberdade.

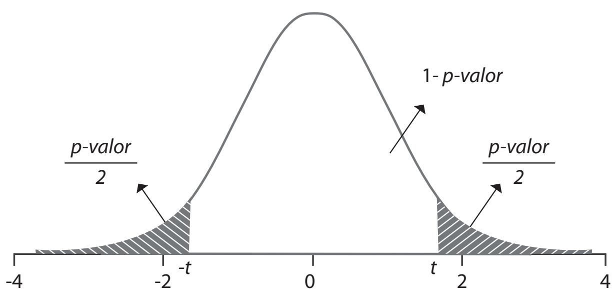
será rejeitada, portanto considera-se que $\beta_0 = 0$. Se o *t* calculado for igual a 3 ou a -3, rejeita-se a hipótese nula e considera-se que $\beta_0 \neq 0$.

O teste de hipóteses para a estimativa de β_i segue a mesma lógica:

H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$

Sendo o *t* calculado a partir de $t = \frac{\hat{\beta}_1}{s_{b1}}$

Na prática, utilizamos o p-valor para avaliarmos se a hipótese nula deve ser rejeitada ou não. O p-valor é a probabilidade restante depois do valor calculado. Se estivermos considerando um nível de 5% de significância, se o p-valor for maior do que 0,05, a hipótese nula não é rejeitada. Se for menor do que 0,05, rejeita-se a hipótese nula, conforme mostra o gráfico a seguir.



4.4 DIAGNÓSTICO DA REGRESSÃO

A maior parte dos resultados obtidos com a estimação do modelo por mínimos quadrados ordinários está baseada em um conjunto de suposições sobre o erro e sobre a relação entre as variáveis. Por isso são necessários alguns testes com relação a essas suposições para que se possa obter resultados confiáveis. Com relação ao erro, é necessário testar se tem distribuição normal, se o modelo é homocedástico (o erro tem variância constante) e se não há autocorrelação (os erros são independentes entre si). Em geral esses problemas estão ligados à especificação do modelo (forma funcional) e à omissão de variáveis relevantes.

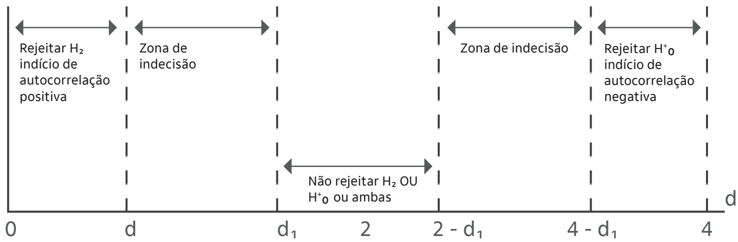
O teste de **normalidade** pode ser feito a partir da estatística Jarque-Bera. Ela é expressa da seguinte forma:

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right)$$

Sendo *n* o tamanho da amostra, *S* uma medida de assimetria (*skewness*) e *K* uma medida de curtose (quanto da distribuição da variável concentra-se nas caudas). A hipótese nula é de que os resíduos têm distribuição normal.

A **homocedasticidade** do modelo é violada quando a variância do erro não é constante. Há vários testes de heterocedasticidade, mas nesse curso ela será avaliada com base no teste White, que é computado a partir de uma regressão do quadrado do resíduo sobre as variáveis explicativas e sobre o produto cruzado das variáveis. A hipótese nula do teste é de que o modelo é homocedástico. Se for rejeitada a hipótese nula, diz-se que o modelo é heterocedástico. Esse tipo de problema é mais comum em dados do tipo corte transversal.

A **autocorrelação** dos resíduos ocorre quando o erro não tem comportamento aleatório, ou seja, os resíduos estão correlacionados entre si. Também há vários testes de autocorrelação. O mais comum, que é válido apenas para autocorrelação de primeira ordem é feito a partir da estatística Durbin-Watson (DW), que é apresentada na saída da maioria dos *softwares* econométricos. DW é comparada com valores baseados em limites inferior e superior, conforme tabela a seguir. Se $d < d_L$, há indícios de correlação positiva; quando d encontra-se entre d_L e d_U o teste é inconclusivo. Se d estiver entre d_U e $(4 - d_U)$ não existe correlação; entre $(4 - d_U)$ e $(4 - d_L)$ o teste é novamente inconclusivo; e se d for maior do que $(4 - d_L)$ existe uma correlação negativa.



O teste de autocorrelação também pode ser feito por meio do teste LM de Breusch-Godfrey, construído a partir da regressão dos resíduos sobre as variáveis explicativas e sobre os resíduos defasados. A hipótese nula é de que não há autocorrelação. Os problemas de autocorrelação são mais comuns em séries de tempo.

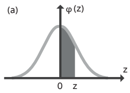
4.5 REGRESSÃO LINEAR MÚLTIPLA

O modelo de regressão linear múltipla é uma extensão do modelo de regressão simples, na medida em que simplesmente incorpora mais variáveis explicativas. Ele pode ser expresso por uma equação do tipo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K + u,$$

sendo X_1, X_2, ..., X_K as variáveis explicativas do modelo. O coeficiente β_0 é a constante da regressão e cada coeficiente β_i representa a variação em Y dada uma variação em X_i, mantidas constantes as demais variáveis explicativas. O diagnóstico do modelo de regressão múltipla segue as mesmas etapas que foram definidas no modelo de regressão simples.

ANEXO 1 DISTRIBUIÇÃO NORMAL PADRONIZADA



Zo	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0948	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3264	0,3315	0,3340	0,3365	0,3389
1,9	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,47667
2,0	0,4772	0,4778	0,4788	0,4793	0,4798	0,4803	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4726	0,4834	0,4838	0,4842	0,4846	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4871	0,4875	0,4878	0,4881	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4901	0,4904	0,4906	0,4909	0,4909	0,4011	0,4913	0,4916
2,4	0,4918	0,4920	0,4925	0,4927	0,4929	0,4931	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4943	0,4945	0,4946	0,4948	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4957	0,4959	0,4960	0,4961	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4968	0,4969	0,4970	0,4971	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4977	0,4977	0,4978	0,4979	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4985	0,4986	0,4986
3,9	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4898	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
∞	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000