

Capítulo 6

Análise dos Modelos

Como já foi visto no capítulo 5, foi adotada como técnica estatística capaz de analisar as relações existentes entre as diversas variáveis independentes selecionadas e a variável dependente – número de acidentes, a Análise de Regressão Múltipla.

Diante disso, o modelo de regressão considerado foi baseado no melhor R^2 , maior ajuste entre as variáveis independentes e a dependente. São apresentados, a seguir, os modelos encontrados para os dois períodos estudados – 1991 e 2000.

6.1 Modelo I

- ✓ **Variável dependente:** Número de acidentes de 1991
- ✓ **Variáveis independentes (ano 1991):** população, área, densidade populacional, imóveis residenciais, imóveis comerciais, interseções semaforizadas, extensão viária, densidade viária, renda per capita, IDH, viagens realizadas motorizadas, viagens realizadas não-motorizadas.

Resultados:

Aplicando o método *Stepwise*, a equação que melhor representa o fenômeno em questão (*Número de acidentes de 1991*) teve um valor de $R^2 = 0,832$. Tal equação é a seguinte:

$$\text{Número de acidentes} = -10,029 + (3,87 \times 10^{-3} \times \text{viagens realizadas motorizadas}) + (8,943 \times 10^{-3} \times \text{imóveis residenciais}) + (0,178 \times \text{renda per capita}) + (1,496 \times 10^{-3} \times \text{população}).$$

(equação 5)

Tabela 12: Variáveis com seus respectivos parâmetros estatísticos referentes ao Modelo I

Variáveis	Beta	t	Sig t	F	Sig F
Viagens realizadas motorizadas	0,663	16,911	0,000	4,398	0,038
Imóveis residenciais	0,202	2,935	0,004		
Renda per capita	0,162	4,160	0,000		
População	0,134	2,097	0,038		

As variáveis incluídas no modelo explicaram 83,2% da variância da variável dependente, atingindo **grande representatividade estatística - R^2** ($< 0,7$). O valor correspondente ao teste F mostrou um bom nível de significância observado (ajuste de dados). Os valores dos Betas demonstrarão que as variáveis mais relevantes, em ordem de importância, são: Viagens realizadas motorizadas, Imóveis residenciais, Renda per Capita e População.

Outro parâmetro estatístico bastante relevante de verificar é a “Sig t”, ou seja, a estatística *t*. Tal parâmetro indica o quanto cada variável independente está explicando o fenômeno (no caso o nº de acidentes). Quanto mais próximo de zero for o parâmetro “Sig t”, significa que a variável independente em análise é importante para o contexto.

É importante reparar que no modelo considerado o parâmetro “Sig t” para cada variável independente considerada é próximo de zero.

Em relação aos pesos das variáveis consideradas, ou seja aos valores de “Beta”, podemos obter um ranking entre as variáveis que mais explicam o fenômeno (nº de acidentes) na tabela 12.

Tabela 13: Coeficientes de Correlação

		Viagens Realizadas Motorizadas (1991)	Imóveis Residenciais (1991)	Renda Per Capita (1991)	População (1991)
Correlação	Viagens Realizadas Motorizadas (1991)	1,000	-0,244	0,006	-0,061
	Imóveis Residenciais (1991)	-0,244	1,000	-0,489	-0,807
	Renda Per Capita (1991)	0,006	-0,489	1,000	0,420
	População (1991)	-0,061	-0,807	0,420	1,000

Variável dependente – nº de acidentes (1991)

- **viagens realizadas motorizadas** – O número de viagens realizadas motorizadas na cidade do Rio de Janeiro como um todo demonstrou ser a mais importante variável no modelo para o ano de 1991. De todas as variáveis selecionadas pelo modelo, esta foi a que apresentou o maior “peso”, isto é, a variável que melhor explica o fenômeno se comparada

com as demais que compõem o modelo – 0,663. Desta maneira podemos dizer que o número de viagens realizadas na cidade analisada em 1991, está altamente relacionada com o número de acidentes na mesma. E através do valor de beta (Tabela 12) vemos que esta relação é diretamente proporcional, isto é, o número de acidentes estará aumentando se o número de viagens realizadas aumentar.

A análise desta variável com as demais, demonstra que (tabela 13 – coeficiente de correlação - r) esta variável apresenta uma maior correlação com a variável imóveis residenciais – (-0,244), apesar de ser um valor baixo. Entretanto, por apresentar valor negativo, a relação que se estabelece entre estas variáveis é de uma proporção inversa. Isto demonstra que numa cidade como o Rio de Janeiro, ou seja, uma metrópole densamente ocupada, quanto maior o número de imóveis residenciais, menor será o número de viagens motorizadas realizadas, e vice e versa – visto que está se analisando a cidade como um todo. Esta associação pode se relacionar com a renda por exemplo, ou mesmo com a área. A significância t da variável viagens realizadas motorizadas – 0,000 demonstra que esta se ajusta perfeitamente à curva do modelo, confirmando a alta correlação desta na explicação do fenômeno.

- **imóveis residenciais** – O número de imóveis residenciais na cidade do Rio demonstrou ser também uma variável relevante na ocorrência de acidentes no modelo de 1991. Apesar de que seu “peso” foi bastante insignificante (0,202), na comparação com as demais variáveis (tabela 13), apresentou uma alta correlação com a variável população (-0,807). Entretanto esta relação ocorre numa proporção inversa, demonstrando que há de fato uma concentração urbana nesta cidade. O modelo demonstrou que o número de imóveis residenciais constitui com a variável população, uma relação inversamente proporcional, ou seja, onde há um grande número de população residente, o número de imóveis residenciais se reduz. Mais uma vez é possível que exista uma complementaridade na análise desta relação com as variáveis renda e área. A significância t foi de 0,004, demonstrando que esta variável se ajusta quase que perfeitamente à curva do modelo.

- **renda per capita e população** – As variáveis renda per capita e população foram as variáveis que apresentaram os mais baixos índices de explicação do fenômeno, embora tenham sido selecionadas para compor o modelo estatístico. Ambas ficaram com betas menores que 0,2 (0,162 e 0,134 respectivamente). A seleção da variável renda já era

esperada em função de que na revisão da literatura em segurança, inúmeros autores notaram a importância desta variável em seus estudos de acidentes rodoviários (Peltzman, 1995; Crandall, 1983; Loeb, 1987; Asch e Levy, 1987; Zlatopler, 1987; Fowles e Loeb, 1989). Nos estudos de Beeck et al na Holanda (1991), a renda per capita foi o prognóstico mais forte nas diferenças em mortalidade regional. Beeck et al concluíram que o mais alto nível de renda está associado ao mais baixo nível de mortalidade. Em comparação com as demais variáveis selecionadas pelo modelo, a maior correlação das variáveis renda e população ocorreu numa proporção inversa com a variável número de imóveis residenciais. Ambas demonstraram ter um alto grau de correlação inversa com o número de imóveis residenciais. Significando dizer que quanto maior a renda e o número de habitantes, menor será o número de imóveis residenciais em cidades como o Rio de Janeiro. Esta evidência pode ser comprovada pelas características morfológicas de seu sítio urbano densamente ocupado.

Com relação à renda, o modelo pode ter desenvolvido algum tipo de distorção pois sabe-se que há grupamento de bairros na cidade do Rio de Janeiro que esta afirmativa não é verdadeira. Na análise de regressão, não é feita uma separação das diversas regiões da cidade, a pesquisa analisa a cidade como um todo, por isso dizer que quanto maior a renda e o número de habitantes, menor será o número de imóveis residenciais, nem sempre será correto para algumas áreas da cidade (áreas onde há ocupação de baixa renda por exemplo). O índice de significância t para a variável renda foi 0,000 – demonstrando que há um perfeito ajuste desta variável com a curva do modelo selecionado para 1991. Em relação à variável população ocorreu o mais baixo índice de significância t se comparado às demais variáveis (0,038) – resultando num ajuste bom, porém este valor nos diz que o ponto x,y desta variável é o mais distante da reta que representa o modelo, se comparada às demais variáveis selecionadas pelo modelo para 1991.

A fim de verificar o comportamento dos resíduos da regressão, foram realizados os seguintes testes: *autocorrelação serial e heterocedasticidade*. Em relação a *autocorrelação serial*, foi aplicado o método de Durbin – Watson (d). Os valores críticos da estatística de Durbin – Watson para um nível de significância de 5%, para 4 variáveis explicativas e número de dados = 153 (bairros) fornece o limite inferior $d_1 = 1,679$ e o limite superior $d_u = 1,788$. Como $d = 1,311 < 2$, indica autocorrelação positiva e pertence a região de aceitação. Portanto, se aceita a hipótese de autocorrelação positiva.

No que diz respeito ao teste de heterocedasticidade, utilizando-se o procedimento de Glejser, e pelo teste F, a hipótese de relação significativa entre os valores absolutos dos resíduos e a variável número de acidentes (1999) foi rejeitada, ao nível de significância de 5%.

Tabela 14: Resumo dos parâmetros estatísticos referente ao Modelo I

Método	Valor de R ²	Sig. F	Durbin Watson	Durbin Watson (tabela)	
				Di	Du
Stepwise	0,832	0,038	1,311	1,679	1,788

6.2 Modelo II

- ✓ **Variável dependente:** Número de acidentes de 2000
- ✓ **Variáveis independentes (ano 2000):** população, área, densidade populacional, imóveis residenciais, imóveis comerciais, interseções semaforizadas, extensão viária, densidade viária, renda per capita, IDH, viagens realizadas motorizadas, viagens realizadas não-motorizadas.

Resultados:

Aplicando o método *Stepwise*, a equação que melhor representa o fenômeno em questão (Número de acidentes de 2000) teve um valor de $R^2 = 0,771$. Tal equação é a seguinte:

$$\text{Número de acidentes} = 75,433 + (8,685 \times \text{interseções semaforizadas}) + (2,956 \times \text{extensão viária}) - (4,920 \times \text{área}) + (9,997 \times 10^{-4} \times \text{viagens realizadas motorizadas}).$$

(equação 6)

Tabela 15: Variáveis com seus respectivos parâmetros estatísticos referente ao Modelo II

Variáveis	Beta	t	Sig t	F	Sig F
Interseções semaforizadas	0,420	6,070	0,000	5,209	0,024
Extensão viária	0,667	7,352	0,000		
Área	- 0,234	- 2,753	0,007		
Viagens realizadas motorizadas	0,150	2,282	0,024		

As variáveis incluídas no modelo explicaram 77,1% da variância da variável dependente, atingindo **grande representatividade estatística** - R^2 ($< 0,7$). O valor correspondente ao teste F mostrou um bom nível de significância observado (ajuste de dados). Os valores dos Betas demonstrarão que as variáveis mais relevantes, em ordem de importância, são: extensão viária, interseções semaforizadas, área e viagens realizadas motorizadas.

Neste modelo também podemos observar que o parâmetro estatístico “Sig t”, ou seja, a estatística t , para cada variável independente considerada é próximo de zero, acarretando um bom resultado.

Em relação aos pesos das variáveis consideradas, ou seja aos valores de “Beta”, podemos obter um ranking entre as variáveis que mais explicam o fenômeno (n^o de acidentes) na tabela 15.

Tabela 16: Coeficientes de Correlação

		Interseções Semaforizadas (2000)	Extensão Viária (2000)	Área (2000)	Viagens Realizadas Motorizadas (2000)
Correlação	Interseções Semaforizadas (2000)	1,000	-0,283	0,197	-0,756
	Extensão Viária (2000)	-0,283	1,000	-0,879	-0,012
	Área (2000)	0,197	-0,879	1,000	- 0,008
	Viagens Realizadas Motorizadas (2000)	-0,756	-0,012	- 0,008	1,000

Variável dependente – n^o de acidentes (2000)

- **extensão viária** – Neste modelo elaborado para o ano de 2000, a variável que mais se destacou na contribuição do número de acidentes, ou seja, aquela que apresentou o maior “peso” em relação ao fenômeno, foi a variável extensão viária com um beta de valor 0,667. Isto demonstra que o espaço destinado à circulação viária, principalmente de veículos pode contribuir de maneira relevante para a incidência de acidentes neste modelo de 2000.

Quanto à sua significância t , o ajuste de 0,000 confirma sua alta correlação com o explicação do fenômeno. Em relação à literatura pesquisada, vários autores fazem menção à esta variável em seus estudos (Kmet, 2003 e Zlatopler, 1991) apontando que em função desta pode ocorrer uma maior exposição ao risco de acidentes.

Se comparada às demais variáveis, na tabela 16 a variável extensão viária apresenta uma alta correlação com a variável área (-0,879). Entretanto se estabelece uma relação proporcionalmente inversa, ou seja, toda vez que há uma redução do valor de uma, a outra necessariamente aumentará na mesma proporção. Neste caso pode-se explicar que quanto menor a área disponível para ocupação humana, maior será aquela destinada à circulação viária. Torna-se interessante mencionar que esta alta correlação inversa entre as variáveis área e extensão viária, não havia sido apresentada pelos autores pesquisados nesta dissertação.

- **interseções semaforizadas** – A seleção desta variável para compor o modelo de 2000 segue uma certa tendência de acordo com alguns estudos apresentados no capítulo 4 desta dissertação. Autores como Greibe (2004) e Noland e Oh (2004) demonstraram em suas pesquisas que há grande ocorrência de acidentes de trânsito em interseções urbanas. Seu “peso” em relação à explicação do fenômeno foi 0,420 e seu índice de significância t foi de 0,000 comprovando que há um ajuste perfeito à curva de representatividade do modelo para esta variável. Na correlação com as demais variáveis selecionadas para o modelo, a variável número de interseções semaforizadas obteve um maior coeficiente de correlação com a variável viagens realizadas motorizadas (-0,756). Porém a relação que se estabelece novamente é de proporção inversa (pelo valor apresentado na análise de regressão ser negativo). Podemos analisar então que se há um número elevado de interseções semaforizadas em cidades como o Rio de Janeiro, se reduzirão o número de viagens realizadas neste mesmo ambiente. E a inversa também é verdadeira: quanto menor o número de interseções semaforizadas, maior será o número de viagens realizadas na cidade.

- **área** – A seleção da variável área vem contribuir para a polêmica existente entre autores que se dedicam ao estudo comparativo de acidentes ocorridos em áreas urbanas e áreas rurais. Podemos citar Chipman et al (1993) e Kmet (2003) que apontam em suas pesquisas a maior ocorrência de acidentes em áreas rurais do que em áreas urbanas justificada por maiores distâncias percorridas pelos motoristas rurais. Entretanto Zlatopler (1991) conclui em seus estudos que maiores distâncias viajadas podem desencorajar os motoristas ao comportamento de risco, e portanto à ocorrência de acidentes, e que ao contrário, em áreas menores, como as áreas urbanas, as pequenas distâncias percorridas incentivam os motoristas a cometerem maiores infrações. Portanto os resultados desta dissertação seguem

a linha de Zlatopler. O ajuste das coordenadas x e y desta variável em relação à curva de representatividade do modelo foi de 0,007 demonstrando que há uma boa aproximação desta variável na explicação do fenômeno. Em comparação com as demais variáveis selecionadas, a variável área apresentou um alto coeficiente de correlação inverso com a variável extensão viária (-0,879) comprovando que há uma concentração de espaços destinados à circulação em pequenas áreas e que em áreas maiores os espaços destinados à circulação se diluem, reduzindo com isso a área ocupada por estes.

- **viagens realizadas motorizadas** – Esta foi a única variável selecionada por ambos os modelos de 1991 e de 2000, embora no primeiro modelo esta apareça com o maior “peso” sobre as demais na explicação do fenômeno; e neste modelo de 2000, sua representatividade beta foi de 0,150, ou seja, a de menor “peso” se comparada com as demais selecionadas. Também na comparação com as outras variáveis em relação à curva de ajuste obteve resultados baixos: as coordenadas mais distantes da reta representativa do modelo (0,024) entretanto esta variável foi fundamental para a composição dos modelos explicativos do número de acidentes para os anos de 1991 e de 2000, por esta razão será ilustrada pelas figuras seguintes (Figuras 14 e 15).

A fim de verificar o comportamento dos resíduos da regressão, foram realizados os seguintes testes: *autocorrelação serial e heterocedasticidade*. Em relação a *autocorrelação serial*, foi aplicado o método de Durbin – Watson (d). Os valores críticos da estatística de Durbin – Watson para um nível de significância de 5%, para 4 variáveis explicativas e número de dados = 153 (bairros) fornece o limite inferior $d_1 = 1,679$ e o limite superior $d_u = 1,788$. Como $d = 1,311 < 2$, indica autocorrelação positiva e pertence a região de aceitação. Portanto, se aceita a hipótese de autocorrelação positiva. No que diz respeito ao teste de heterocedasticidade, utilizando-se o procedimento de Glejser, e pelo teste F, a hipótese de relação significativa entre os valores absolutos dos resíduos e a variável número de acidentes (2000) foi rejeitada, ao nível de significância de 5%. Isto significa dizer que os valores mais dispersos da reta representativa do modelo não são capazes de interferir nos resultados dos parâmetros estatísticos analisados.

Tabela 17: Resumo dos parâmetros estatísticos referente ao Modelo II

Método	Valor de R ²	Sig. F	Durbin Watson	Durbin Watson (tabela)	
				Di	Du
Stepwise	0,771	0,024	1,728	1,679	1,788

As análises comparativas entre os modelos de 1991 e 2000, tornam-se bastante complexas no momento em que são adotadas variáveis ligadas às esferas de desenvolvimento urbano da cidade. As diferenças ocorridas neste período de 10 anos na cidade implicaram em alterações em cada uma das variáveis estudadas, o que acarreta mudança nos resultados das principais variáveis selecionadas pelos modelos. O contexto sócio-econômico e os cenários de circulação, bem como as políticas de transporte adotadas em cada um destes períodos, serão fundamentais para os resultados encontrados pelos modelos. Mudanças relevantes podem alterar completamente as variáveis para cada período estudado.