Instituto Superior de Contabilidade e Administração de Coimbra
Instituto Politécnico de Coimbra

Trabalho N. º2 de

Análise Estatística de Dados

Catarina Juliana Martins Auxiliar, 2021134297

Helena Cristina Almeida da Cruz, 2024147830

Nuno Emanuel Lopes Gonçalves, 2015063961

Simão Pedro Tomé Dias, 2020132169

COIMBRA

10 de novembro de 2024

RESUMO

Este relatório foi elaborado no âmbito da disciplina de Análise Estatística de Dados do Mestrado em Análise de Dados e Sistemas de Apoio à Decisão, sob a orientação da professora Clara Margarida Pisco Viseu. O principal objetivo do trabalho consiste em aplicar conceitos e métodos de análise estatística em cenários práticos, de forma a consolidar os conhecimentos adquiridos em aula. Ao longo do relatório são apresentadas respostas aos dois casos práticos propostos, respostas essas que incluem uma descrição e justificação do método estatístico utilizado, os resultados obtidos no programa de análise estatística SPSS29, e as respetivas conclusões retiradas da análise. No primeiro caso, foi utilizada a técnica de Análise Fatorial, de forma a identificar as principais dimensões de satisfação dos passageiros de uma companhia aérea, tendo-se identificado quatro fatores — "Experiência a bordo", "Facilidade e conveniência na experiência de viagem", "Qualidade do serviço a bordo" e "Facilidade do Check-in". No segundo caso, aplicou-se a Regressão Linear Múltipla com o intuito de avaliar os fatores determinantes do consumo mensal de eletricidade em residências de um certo país, tendo-se concluído que, embora todas as variáveis analisadas sejam estatisticamente significativas, as variáveis URB, NUMC e AC têm maior impacto na explicação da variação do consumo. Assim, o trabalho realizado demonstrou a aplicabilidade dos métodos estatísticos na extração de *insights* relevantes em ambos os casos, reforçando a importância destas análises para realizar decisões estratégicas bem fundamentadas.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

ÍNDICE GERAL

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

ÍNDICE DE FIGURAS

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

INTRODUÇÃO

O objetivo do trabalho realizado e apresentado neste relatório foi o de aplicar, num contexto prático, os conceitos teóricos lecionados na disciplina de Análise Estatística de Dados do Mestrado em Análise de Dados e Sistemas de Apoio à Decisão, de forma a consolidar os conhecimentos adquiridos em aula. Nesse sentido, foram resolvidos dois casos práticos, que incidem fundamentalmente sobre os tópicos: Análise Fatorial e Análise de Regressão Linear.

O primeiro caso estudado centra-se na satisfação dos passageiros de uma companhia aérea, com base nos dados obtidos através de um questionário aplicado a uma amostra de 3695 clientes. Neste questionário, os clientes avaliaram o seu grau de satisfação relativamente a diversos aspetos do serviço da companhia aérea, numa escala de 1 a 5. O objetivo foi identificar, entre os parâmetros avaliados, aqueles que mais influenciam o grau de satisfação dos passageiros, utilizando a técnica de Análise Fatorial para reduzir a complexidade dos dados e extrair os principais fatores subjacentes à satisfação dos clientes.

No segundo caso, foi investigado o consumo mensal de eletricidade em residências de um país. Para isso, foi aplicado um modelo de Regressão Linear Múltipla. Este modelo permitiu analisar a contribuição de diversos elementos no consumo energético residencial, oferecendo uma visão detalhada sobre os principais determinantes e auxiliando na compreensão dos padrões de consumo nas habitações analisadas.

As questões dos casos propostos foram respondidas com base na aplicação dos métodos estatísticos considerados mais apropriados e com base nos resultados gerados através do software de análise estatística SPSS29.

PARTE I: CASO I

1 Variáveis em estudo

No primeiro caso pretende-se estudar quais os atributos determinantes no grau de satisfação dos clientes de uma determinada companhia aérea, com base nos dados recolhidos num questionário de satisfação realizado a uma amostra de 3695 clientes. A base de dados contém informação acerca das seguintes variáveis: ID (número de

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

identificação do cliente), Tipo (tipo de viagem: viagem de negócios ou viagem pessoal), Classe (classes do avião: classe económica, classe executiva ou primeira classe), WIFI (grau de satisfação com o serviço de Wi-Fi a bordo), CONV (grau de satisfação com a conveniência do horário de partida/chegada), RES (grau de satisfação com a facilidade de reserva), POR (grau de satisfação com a localização do portão), COM (grau de satisfação com a comida e bebida), CHON (grau de satisfação com o check-in online), ASS (grau de satisfação com o conforto do assento), ENT (grau de satisfação com o entretenimento a bordo), SERV (grau de satisfação com o serviço a bordo), PERN (grau de satisfação com o espaço para as pernas), BAG (grau de satisfação com o manuseio de bagagem), CHSER (grau de satisfação com o serviço de check-in) e LIMP (grau de satisfação com a limpeza). Os 13 parâmetros de satisfação foram avaliados pelos clientes numa escala de 1 a 5, onde 1 significa "muito insatisfeito" e 5 significa "muito satisfeito".

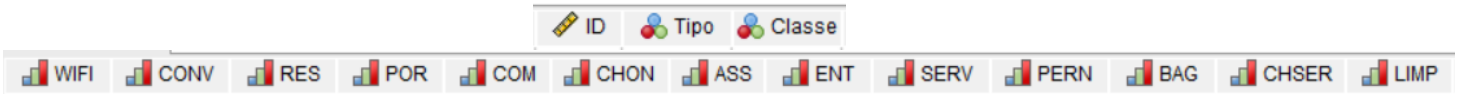

Figura 1: Variáveis Caso 1

2 Resolução das questões propostas

Começou-se por obter uma amostra aleatória de dimensão n=3695-5k, onde k correspondia ao número do grupo. Dado que o nosso grupo era o 6, obtivemos uma amostra de dimensão n=3665.

2.1 <u>Questão 1:</u> Verifique, através de um teste adequado, se não há diferenças entre as distribuições da variável ASS para os dois tipos de viagem.

Para analisar se existem diferenças significativas entre as distribuições da variável dependente ASS (conforto do assento) para os dois tipos de viagem (viagem de negócios ou viagem pessoal), realizámos o teste não-paramétrico U de Mann-Whitney. O teste U de Mann-Whitney é uma alternativa ao teste t para amostras independentes, dado que é o teste adequado para comparar as distribuições de dois grupos independentes (tipos de viagem) quando a variável dependente (ASS) não segue distribuição normal, que é o caso, como se comprova pelo teste de normalidade abaixo.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

H0: A variável ASS segue uma distribuição normal;

H1: A variável ASS não segue uma distribuição normal.

Testes de Normalidade

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estatística	gl	Sig.	Estatística	gl	Sig.
conforto do assento	,235	3665	<,001	,875	3665	<,001

a. Correlação de Significância de Lilliefors

Figura 2: Testes de Normalidade para a variável ASS

Como o valor-p < 0,001 < α = 0,05 => Rejeita-se H0. Logo, pode concluir-se que a variável ASS não segue uma distribuição normal.

As hipóteses formuladas para o teste U de Mann-Whitney foram as seguintes:

H0: A distribuição de ASS é igual para os dois tipos de viagem;

H1: A distribuição de ASS não é igual para os dois tipos de viagem.

Postos

	Tipo	N	Posto médio	Soma de Classificações
conforto do assento	viagem de negócios	2532	1924,12	4871873,50
	viagem pessoal	1133	1629,37	1846071,50
	Total	3665		

Estatísticas de teste[a]

	conforto do assento
U de Mann-Whitney	1203660,500
Wilcoxon W	1846071,500
Z	-8,035
Significância Sig. (2 extremidades)	<,001

a. Variável de Agrupamento: Tipo

Figura 3: Teste U de Mann-Whitney para comparação do conforto do assento (ASS) entre os dois tipos de viagem (Tipo)

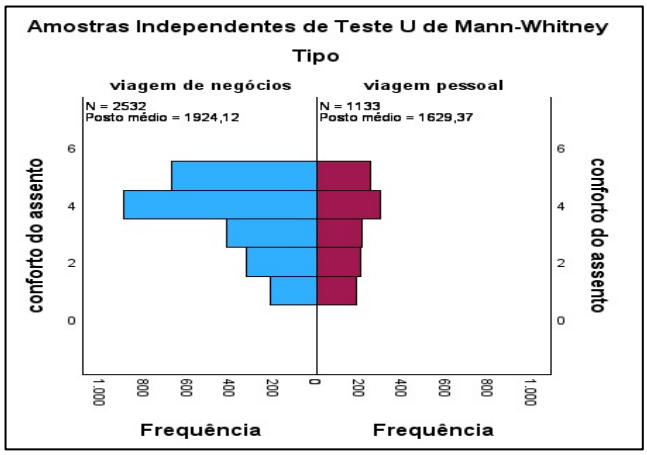

Figura 4: Distribuição do Conforto do Assento por Tipo de Viagem

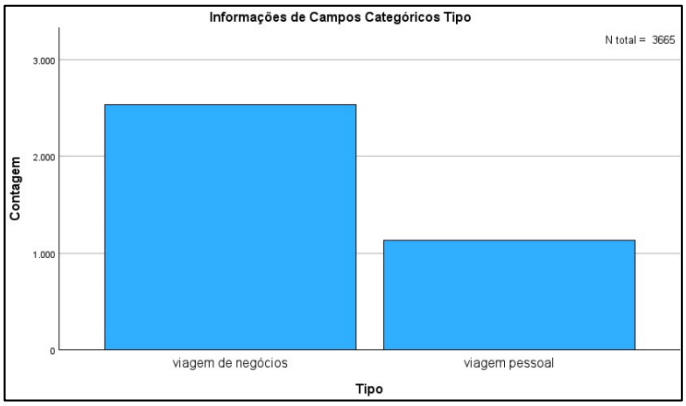

Figura 5: Distribuição de Frequências por Tipo de Viagem

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Como valor-p < 0,001 < α = 0.05 => Rejeita-se H0. Posto isto, conclui-se que existem diferenças estatisticamente significativas entre as distribuições do conforto do assento para os dois tipos de viagem, o que sugere que o tipo de viagem pode influenciar a perceção de conforto do assento pelos passageiros.

2.2 <u>Questão 2:</u> Verifique, através de um teste adequado, se SERV não difere para as três classes.

Para verificar se existem diferenças significativas entre as distribuições da variável SERV (Serviço a bordo) nas três classes de viagem (classe económica, classe executiva e primeira classe), optámos por realizar o teste de Kruskal-Wallis, uma alternativa ao teste One-Way ANOVA quando os pressupostos de normalidade e homogeneidade de variâncias não são cumpridos. Este teste não paramétrico é adequado porque a variável SERV não segue uma distribuição normal em todas as classes de viagem (como se pode comprovar pelo teste de normalidade abaixo).

H0: A variável SERV segue uma distribuição normal em todas as classes de viagem;

H1: A variável SERV não segue uma distribuição normal em todas as classes de viagem.

Resumo de processamento de casos

		Casos					
		Válido		Omisso		Total	
	Classe	N	Porcentagem	N	Porcentagem	N	Porcentagem
serviço a bordo	classe económica	1619	100,0%	0	0,0%	1619	100,0%
	classe executiva	1780	100,0%	0	0,0%	1780	100,0%
	primeira classe	266	100,0%	0	0,0%	266	100,0%

Testes de Normalidade

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Classe	Estatística	gl	Sig.	Estatística	gl	Sig.
serviço a bordo	classe económica	,189	1619	<,001	,899	1619	<,001
	classe executiva	,237	1780	<,001	,864	1780	<,001
	primeira classe	,185	266	<,001	,905	266	<,001

a. Correlação de Significância de Lilliefors

Figura 6: Testes de Normalidade para a variável SERV por Classe de viagem

Para todas as classes de viagem (classe económica, classe executiva e primeira classe), o valor-p obtido foi inferior a 0,001, sendo menor que o nível de significância α = 0,05. Assim, rejeita-se a hipótese nula (H0), o que indica que a variável SERV não segue uma distribuição normal em nenhuma das classes de viagem.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

As hipóteses formuladas para o teste de Kruskal-Wallis foram as seguintes:

H0: A distribuição da variável SERV é igual entre as três classes de viagem (classe económica, classe executiva e primeira classe);

H1: A distribuição da variável SERV difere entre pelo menos uma das três classes de viagem.

Postos

	Classe	N	Posto médio
serviço a bordo	classe económica	1619	1627,49
	classe executiva	1780	2044,46
	primeira classe	266	1668,79
	Total	3665	

Estatísticas de teste[a,b]

	serviço a bordo
H de Kruskal-Wallis	146,546
df	2
Significância Sig.	<,001

a. Teste Kruskal Wallis

b. Variável de Agrupamento: Classe

Figura 7: Teste de Kruskal-Wallis para Comparação da Distribuição da Variável SERV entre Classes de Viagem

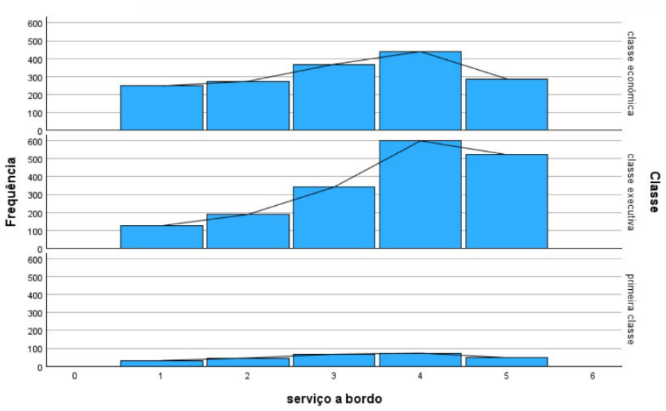

Figura 9: Histograma Avaliações de Serviço a Bordo nas Diferentes Classes

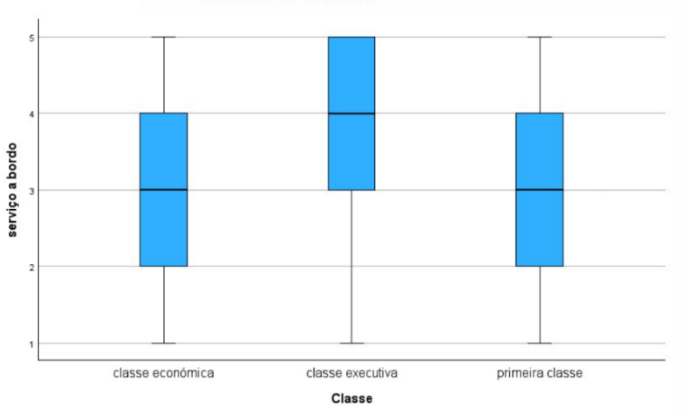

Figura 8: Boxplot Avaliações de Serviço a Bordo entre as Classes

Como valor-p <0,001< α = 0.05 => Rejeita-se H0. Os resultados sugerem que o serviço a bordo é percecionado de forma distinta nas várias classes de viagem (económica, executiva e primeira classe). Com base na análise dos gráficos, observa-se que os passageiros da classe executiva tendem a reportar maior satisfação com o serviço a bordo em comparação com os passageiros da classe económica e primeira classe. Embora a mediana da satisfação com o serviço a bordo seja semelhante entre a primeira classe e a classe económica, o histograma indica que a classe executiva apresenta uma distribuição mais concentrada nos níveis mais altos de satisfação, enquanto a classe económica possui uma distribuição mais uniforme. Este resultado sugere que a companhia aérea pode

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

beneficiar de uma investigação mais aprofundada sobre os fatores que contribuem para uma menor satisfação na classe económica e considerar possíveis melhorias no serviço prestado nessa classe.

2.3 <u>Questão 3:</u> Aplique a análise fatorial sobre o conjunto das 13 variáveis de modo a obter dimensões da satisfação mais gerais. Interprete o resultado desta análise fatorial.

O primeiro passo a realizar é avaliar a validade da aplicação da Análise Fatorial no caso em estudo. A tabela abaixo apresenta dois testes que nos ajudarão a avaliar a adequação dos dados para a realização de uma análise fatorial.

Teste de KMO e Bartlett

Medida Kaiser-Meyer-Olkin de adequação de amostragem.		,771
Teste de esfericidade de Bartlett	Aprox. Qui-quadrado	19531,564
	gl	78
	Sig.	<,001

Figura 10: Testes KMO e Bartlett

Nesse sentido, começamos por realizar o teste de esfericidade de Bartlett, que testa a hipótese de a matriz de correlações ser a matriz identidade, isto é, as variáveis não serem correlacionadas. Este teste é fundamental pois, caso não haja correlação entre as variáveis, não será possível efetuar a análise fatorial, pois não será possível identificar fatores comuns para agrupar as variáveis.

Matriz de correlações

		serviço de Wi-Fi a bordo	conveniência do horário de partida/chegada	facilidade de reserva online	localização do portão	comida e bebida	check-in online	conforto do assento	entretenimento a bordo	serviço a bordo	espaço para as pernas	manuseio de bagagem	serviço de check-in	limpeza
Correlação	serviço de Wi-Fi a bordo	1,000	,400	,686	,388	,170	,448	,155	,224	,125	,161	,115	,091	,160
	conveniência do horário de partida/chegada	,400	1,000	,530	,529	,001	,082	,006	-,017	,101	-,004	,098	,129	,008
	facilidade de reserva online	,686	,530	1,000	,540	,054	,351	,048	,055	,053	,102	,044	,040	,029
	localização do portão	,388	,529	,540	1,000	,013	,014	,010	,024	,001	-,019	,013	-,032	,000
	comida e bebida	,170	,001	,054	,013	1,000	,258	,580	,616	,070	,052	,038	,075	,645
	check-in online	,448	,082	,351	,014	,258	1,000	,444	,312	,171	,138	,100	,241	,353
	conforto do assento	,155	,006	,048	,010	,580	,444	1,000	,627	,157	,121	,086	,182	,684
	entretenimento a bordo	,224	-,017	,055	,024	,616	,312	,627	1,000	,441	,319	,377	,136	,691
	serviço a bordo	,125	,101	,053	,001	,070	,171	,157	,441	1,000	,366	,531	,271	,135
	espaço para as pernas	,161	-,004	,102	-,019	,052	,138	,121	,319	,366	1,000	,383	,176	,110
	manuseio de bagagem	,115	,098	,044	,013	,038	,100	,086	,377	,531	,383	1,000	,251	,094
	serviço de check-in	,091	,129	,040	-,032	,075	,241	,182	,136	,271	,176	,251	1,000	,180
	limpeza	,160	,008	,029	,000	,645	,353	,684	,691	,135	,110	,094	,180	1,000
Sig. (unilateral)	serviço de Wi-Fi a bordo		<,001	<,001	<,001	<,001	<,001	<,001	<,001	<,001	<,001	<,001	<,001	<,001
	conveniência do horário de partida/chegada	,000		,000	,000	,465	,000	,361	,146	,000	,412	,000	,000	,319
	facilidade de reserva online	,000	,000		,000	,001	,000	,002	,000	,001	,000	,004	,008	,041
	localização do portão	,000	,000	,000		,221	,202	,272	,071	,467	,127	,210	,025	,492
	comida e bebida	,000	,465	,001	,221		,000	,000	,000	,000	,001	,011	,000	,000
	check-in online	,000	,000	,000	,202	,000		,000	,000	,000	,000	,000	,000	,000
	conforto do assento	,000	,361	,002	,272	,000	,000		,000	,000	,000	,000	,000	,000
	entretenimento a bordo	,000	,146	,000	,071	,000	,000	,000		,000	,000	,000	,000	,000
	serviço a bordo	,000	,000	,001	,467	,000	,000	,000	,000		,000	,000	,000	,000
	espaço para as pernas	,000	,412	,000	,127	,001	,000	,000	,000	,000		,000	,000	,000
	manuseio de bagagem	,000	,000	,004	,210	,011	,000	,000	,000	,000	,000		,000	,000
	serviço de check-in	,000	,000	,008	,025	,000	,000	,000	,000	,000	,000	,000		,000
	limpeza	,000	,319	,041	,492	,000	,000	,000	,000	,000	,000	,000	,000	

Figura 11: Matriz de correlações

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Formularam-se as seguintes hipóteses:

H0: A matriz de correlação é igual à matriz identidade;

H1: A matriz de correlação difere da matriz identidade.

Como valor-p < 0.001< α = 0.05, rejeita-se a hipótese nula e conclui-se que a matriz de correlações difere da matriz identidade.

No entanto, o teste de esfericidade de Bartlett é muito sensível ao tamanho das amostras, isto é, para amostras grandes, até pequenas correlações podem ser estatisticamente significativas. Devido a isso, deve complementar-se com o teste de kaiser-Meyer-Olkin (KMO). O teste de kaiser-Meyer-Olkin (KMO), mede a proporção da variância entre as variáveis que pode ser considerada comum e, portanto, explicável por fatores latentes. Quanto mais próximo de 1 for o valor obtido, mais adequada será a análise fatorial com os dados disponíveis. No nosso caso, o valor do Teste KMO é 0,771, o que indica que a análise fatorial é adequada.

Para além dos dois testes realizados, é ainda necessário aferir se amostra tem a dimensão adequada para ser viável a realização da análise fatorial. A regra é que se deve ter, no mínimo, 5 a 10 observações por variável. No nosso caso, temos de ter, no mínimo, 5*13=65 observações, o que é muito inferior ao nosso número de observações, logo, segundo este critério também se conclui que é adequado realizar-se a análise fatorial.

Comunalidades

	Inicial	Extração
serviço de Wi-Fi a bordo	1,000	,696
conveniência do horário de partida/chegada	1,000	,608
facilidade de reserva online	1,000	,791
localização do portão	1,000	,715
comida e bebida	1,000	,725
check-in online	1,000	,764
conforto do assento	1,000	,732
entretenimento a bordo	1,000	,847
serviço a bordo	1,000	,668
espaço para as pernas	1,000	,461
manuseio de bagagem	1,000	,692
serviço de check-in	1,000	,493
limpeza	1,000	,782

Método de Extração: análise de Componente Principal.

Figura 12: Tabela de Comunalidades

A tabela de comunalidades é útil para avaliar a qualidade da análise fatorial e a adequação das variáveis no modelo, dado que nos dá informação da proporção de variância de cada

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

variável explicada pelas componentes (Exemplo: Os fatores extraídos explicam 69,6% da variância do serviço de WI-FI a bordo). Se as comunalidades são altas (geralmente >0,5), isso sugere que a análise fatorial é eficaz. Ao analisar a tabela, conclui-se que a maior parte das variáveis apresenta valores elevados na coluna de extração, o que sugere que elas estão bem representadas pelos fatores extraídos. No entanto, algumas variáveis, como o "Espaço para as Pernas" e o "Serviço de Check-in", possuem comunalidades mais baixas (0,461 e 0,493, respetivamente), o que significa que essas variáveis estão menos bem explicadas pelos fatores. Apesar disso, decidimos manter todas as variáveis na análise, uma vez que os valores de comunalidade das mesmas se encontram muito próximos do limiar de 0,5, o que indica que podem contribuir com informações valiosas sobre a experiência do passageiro.

Existem várias regras práticas para determinar quantos fatores excluir da análise, destacam-se três, dois deles utilizam a informação da tabela da Variância total explicada e o outro baseia-se na interpretação do gráfico de escarpa.

Variância total explicada

Componente	Autovalores iniciais			Somas de extração de carregamentos ao quadrado			Somas de rotação de carregamentos ao quadrado		
	Total	% de variância	% cumulativa	Total	% de variância	% cumulativa	Total	% de variância	% cumulativa
1	3,728	28,680	28,680	3,728	28,680	28,680	2,967	22,826	22,826
2	2,450	18,846	47,526	2,450	18,846	47,526	2,545	19,575	42,401
3	1,781	13,699	61,225	1,781	13,699	61,225	2,095	16,113	58,514
4	1,015	7,807	69,033	1,015	7,807	69,033	1,367	10,519	69,033
5	,899	6,918	75,951						
6	,651	5,007	80,958						
7	,491	3,780	84,738						
8	,455	3,500	88,238						
9	,443	3,408	91,646						
10	,335	2,580	94,226						
11	,291	2,239	96,465						
12	,267	2,055	98,520						
13	,192	1,480	100,000						

Método de Extração: análise de Componente Principal.

Figura 13: Variância total explicada

Segundo o critério de Kaiser, devemos excluir os fatores cujos valores próprios são inferiores a 1. Pela tabela da variância total explicada, os fatores que devem ser selecionados para a análise são os primeiros quatro, dado que os restantes apresentam valor próprio inferior a 1.

Uma outra regra frequentemente utilizada é extrair um número mínimo de fatores de forma a explicar pelo menos 50% da variância total das variáveis originais. Pela análise da tabela acima, os primeiros três fatores seriam suficientes. O primeiro fator explica

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

28,68% da variabilidade inicial dos dados, o segundo fator explica 18,85% da variabilidade inicial dos dados e o terceiro fator explica 13,7% da variabilidade inicial dos dados. No total, os três fatores explicam 61,225% da variabilidade inicial dos dados, o que já é mais do que 50%.

Podemos também analisar o número de fatores a incluir através do gráfico de escarpa. Segundo este critério, quando a percentagem de variância explicada por cada fator se reduz e a curva passa a ser quase paralela ao eixo das abcissas, não devemos incluir mais componentes. Pela análise do gráfico, comprova-se a conclusão retirada pela aplicação do critério de Kaiser, de incluir apenas os quatro primeiros fatores.

Figura 14: Gráfico de escarpa

Após a análise dos três critérios, podemos concluir que o número ideal de fatores são os primeiros quatro. Sendo que o que mais contribui para a explicação da variância do modelo fatorial é o primeiro, pois explica 28,68% da variabilidade inicial dos dados.

Matriz de componente[a]

	Componente			
	1	2	3	4
serviço de Wi-Fi a bordo	,532	,614	-,081	-,175
conveniência do horário de partida/chegada	,261	,711	,009	,186
facilidade de reserva online	,390	,789	-,099	-,080
localização do portão	,215	,709	-,130	,387
comida e bebida	,646	-,304	-,418	,202
check-in online	,609	,099	-,128	-,606
conforto do assento	,726	-,317	-,323	-,028
entretenimento a bordo	,815	-,329	,014	,271
serviço a bordo	,472	-,073	,652	,119
espaço para as pernas	,388	-,048	,555	,028
manuseio de bagagem	,410	-,046	,699	,183
serviço de check-in	,352	-,032	,337	-,504
limpeza	,732	-,348	-,342	,090

Método de Extração: análise de Componente Principal.
a. 4 componentes extraídos.

Figura 15: Matriz de componentes

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

A matriz de componentes dá-nos informação das cargas fatoriais, que indicam a relação entre cada variável e os fatores extraídos, revelando a força e direção dessa associação. Isso permite-nos identificar quais variáveis estão mais relacionadas a cada fator (pesos > = 0,5 são considerados significativos). Pela análise da matriz de componentes acima, verifica-se que variável "serviço de Wi-Fi a bordo" e "check-in online" apresentam pesos significativos em 2 fatores. De forma a resolver esse problema, devemos aplicar o método de rotação VARIMAX, de forma a que essas variáveis fiquem associadas de forma mais forte a apenas 1 fator.

Matriz de componente rotativa[a]

	Componente			
	1	2	3	4
serviço de Wi-Fi a bordo	,152	,720	,061	,388
conveniência do horário de partida/chegada	-,051	,772	,095	-,032
facilidade de reserva online	,005	,851	-,010	,260
localização do portão	,025	,809	-,003	-,244
comida e bebida	,850	,041	-,025	-,006
check-in online	,342	,204	,004	,779
conforto do assento	,820	,002	,044	,242
entretenimento a bordo	,806	,037	,442	,016
serviço a bordo	,114	,042	,803	,095
espaço para as pernas	,064	,029	,660	,145
manuseio de bagagem	,049	,051	,828	,018
serviço de check-in	,009	-,044	,338	,614
limpeza	,872	,000	,058	,130

Método de Extração: análise de Componente Principal.
Método de Rotação: Varimax com Normalização de Kaiser.[a]

a. Rotação convergida em 5 iterações.

Figura 16: Matriz de componente rotativa

Após a rotação já temos cada variável associada a apenas 1 dos fatores. Ficamos assim no final com apenas 4 fatores, a que podemos chamar:

Fator 1 = "Experiência a bordo": As variáveis mais associadas a este fator são comida e bebida (0,850), conforto do assento (0,820), entretenimento a bordo (0,806) e limpeza (0,872).

Fator 2 = "Facilidade e conveniência na experiência de viagem": As variáveis mais associadas a este fator são serviço de Wi-Fi a bordo (0,720), conveniência do horário de partida/chegada (0,772), facilidade de reserva online (0,851) e localização do portão (0,809).

Fator 3 = "Qualidade do serviço a bordo": As variáveis mais associadas a este fator são serviço a bordo (0,803), espaço para as pernas (0,660) e manuseio de bagagem (0,828).

Fator 4 = "Facilidade do Check-in": As variáveis mais associadas a este fator são check-in online (0,779) e serviço de check-in (0,614).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

PARTE II: CASO II

1 Variáveis em estudo

O segundo caso estuda os determinantes do consumo mensal de eletricidade em residências de um país, considerando as seguintes variáveis: NUMP (número habitual de pessoas em casa), AREA (área da casa), REND (rendimento médio mensal da família (u.m.), NUMC (número de crianças), AC (1 se a casa tem ar condicionado, 0 caso contrário), APART (1 se é apartamento, 0 caso contrário), URB (1 se a casa está na zona urbana, 0 caso contrário). O estudo centrou-se na aplicação de um modelo de Regressão Linear Múltipla, com o objetivo de avaliar, de entre as referidas variáveis, quais as que melhor explicam o consumo mensal de eletricidade nas residências do país em análise.

Figura 17: Variáveis caso II

2 Resolução das questões propostas

Para a resolução desta questão, reduzimos a nossa amostra de n=985 para n=855, aplicando a fórmula n=885-5k, onde k correspondia ao número do grupo. Dado que o nosso grupo era o 6, obtivemos uma amostra de dimensão n=855.

2.1 <u>Questão 1:</u> Estime um modelo de regressão múltipla que lhe permita explicar o consumo mensal de eletricidade em função das várias variáveis explicativas apresentadas.

$\mathcal{C}$ONS = β1 + β2NUMP + β3AREA + β4AC + β5APART + β6REND+ β7NUMC + β8URB + Ɛ

Coeficientes[a]

Modelo		Coeficientes não padronizados		Coeficientes padronizados			Estatísticas de colinearidade	
		B	Erro Erro	Beta	t	Sig.	Tolerância	VIF
1	(Constante)	164,098	16,471		9,963	<,001		
	NUMP	5,583	1,188	,062	4,700	<,001	,995	1,005
	AREA	,049	,016	,040	3,023	,003	,993	1,007
	AC	164,563	4,941	,438	33,302	<,001	,992	1,008
	APART	55,965	4,761	,154	11,754	<,001	,999	1,001
	REND	,001	,000	,055	4,198	<,001	,994	1,006
	NUMC	90,639	2,566	,465	35,323	<,001	,992	1,008
	URB	255,534	4,902	,685	52,125	<,001	,993	1,007

a. Variável Dependente: CONS

Figura 18: Modelo de regressão linear múltipla estimado

Modelo Estimado: CONS^ = 164,098 + 5,583*NUMP + 0,049*AREA + 164,563*AC + 55,965*APART + 0,001*REND + 90,639*NUMC + 255,534*URB

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Interpretação da estimativa dos coeficientes:

$\beta 1\hat{}$: o consumo mensal de eletricidade estimado quando todas as variáveis explicativas são iguais a zero é de 164,098 unidades. Neste contexto, não faz muito sentido interpretar o valor deste coeficiente, dado que há algumas variáveis cujo valor não faz sentido ser 0, como NUMP e AREA.

$\beta 2\hat{}$: Por cada morador adicional na residência, espera-se um aumento de aproximadamente 5,58 unidades no consumo mensal de eletricidade, mantendo todas as outras variáveis constantes. Isso sugere que o número habitual de pessoas em casa impacta positivamente o consumo, o que faz sentido em termos práticos.

$\beta 3\hat{}$: Por cada unidade adicional na área da casa, o consumo de eletricidade aumenta em aproximadamente 0,049 unidades. Isso indica que casas maiores consomem ligeiramente mais eletricidade, possivelmente devido à necessidade iluminar, aquecer, arrefecer, etc. uma maior área.

$\beta 4\hat{}$: Se a residência possui ar condicionado (AC = 1), o consumo mensal de eletricidade aumenta em 164,563 unidades, comparado a uma residência sem ar condicionado (AC = 0), mantendo todas as outras variáveis constantes. Esse impacto significativo sugere que o ar condicionado é um dos principais determinantes do consumo de eletricidade.

$\beta 5\hat{}$: Se a residência for um apartamento (APART = 1), o consumo de eletricidade aumenta em 55,965 unidades em comparação com uma casa que não é um apartamento, mantendo as outras variáveis constantes. Este aumento sugere que apartamentos, possivelmente devido à configuração dos eletrodomésticos ou ao tipo de estrutura, podem ter um consumo de eletricidade ligeiramente superior

$\beta 6\hat{}$: Para cada unidade adicional de rendimento médio mensal, o consumo mensal de eletricidade aumenta em 0,001 unidades, mantendo as outras variáveis constantes. Isto indica que o rendimento tem um impacto praticamente insignificante no consumo de eletricidade.

$\beta 7\hat{}$: Por cada criança adicional na residência, o consumo de eletricidade aumenta em 90,639 unidades, mantendo as outras variáveis constantes. Este efeito positivo e relativamente alto reflete um maior uso de eletricidade associado a crianças.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

β8^: Se a residência está localizada na zona urbana (URB = 1), o consumo de eletricidade aumenta em 255,534 unidades em comparação com uma residência na zona rural (URB = 0), mantendo todas as outras variáveis constantes. Esse impacto sugere que residências em zonas urbanas consomem mais eletricidade do que em zonas rurais, o que pode estar associado a diferenças no estilo de vida.

Pela análise da estimativa dos coeficientes pode concluir-se que o consumo mensal de eletricidade é influenciado positivamente por várias características da residência e dos moradores. Especificamente, fatores como o número de pessoas e crianças na casa, a presença de ar condicionado, o facto de ser um apartamento e estar localizado em área urbana são todos associados a aumentos significativos no consumo. A área da residência também aumenta o consumo, mas com menor intensidade, e o rendimento tem um impacto quase nulo. Em síntese, o consumo de eletricidade parece ser mais sensível ao estilo de vida dos residentes e às condições do imóvel do que ao rendimento familiar.

2.2 Questão 2: Efetue os testes e as interpretações que entender convenientes.

Resumo do modelo[b]

Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa	Estatísticas de mudança				
					Mudança de R quadrado	Mudança F	df1	df2	Sig. Mudança F
1	,924[a]	,855	,853	69,51183513128	,855	711,049	7	847	<,001

a. Preditores: (Constante), URB, NUMC, APART, REND, NUMP, AREA, AC

b. Variável Dependente: CONS

Figura 19: Resumo do modelo estimado para o consumo de eletricidade

O coeficiente de correlação de Pearson (R) não pode ser interpretado em modelos de regressão linear múltipla, pois sua interpretação só é válida em análises de relação entre duas variáveis (um para um). Em modelos de regressão linear múltipla, onde várias variáveis explicativas são consideradas simultaneamente, o coeficiente de determinação (R^2) e os coeficientes individuais das variáveis são as métricas mais adequadas para interpretar a relação entre as variáveis explicativas e a variável dependente. O valor de R^2 = 0,855 indica que as variáveis explicativas, em conjunto, explicam 85,5% da variação no consumo mensal de eletricidade.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Para verificar se o modelo é globalmente significativo, recorremos ao teste ANOVA e formulamos as seguintes hipóteses:

H0: Modelo globalmente não significativo (não adequado/relevante)

H1: Modelo globalmente significativo (adequado/relevante)

ANOVA[a]

Modelo		Soma dos Quadrados	df	Quadrado Médio	F	Sig.
1	Regressão	24049989,493	7	3435712,785	711,049	<,001[b]
	Resíduo	4092615,254	847	4831,895		
	Total	28142604,747	854			

a. Variável Dependente: CONS

b. Preditores: (Constante), URB, NUMC, APART, REND, NUMP, AREA, AC

Figura 20: Teste de significância global ANOVA

Como o valor-p < 0,001 < α = 0.05, rejeita-se H0 e conclui-se que o modelo é globalmente significativo. A soma dos quadrados da regressão (24.049.989,493) é substancialmente superior à soma dos quadrados do resíduo (4.092.615,254), o que sugere que a maior parte da variância está a ser explicada pelo modelo. Para além disso, o valor elevado de F (711,049) e a significância inferior a 0,001 reforçam a robustez do modelo. Estes resultados indicam que é extremamente improvável que o efeito das variáveis independentes sobre o consumo seja aleatório, demonstrando assim a adequação do modelo para prever o consumo com base nas variáveis explicativas. No entanto, isso não significa que todas as variáveis sejam estatisticamente relevantes a nível individual. Pela interpretação da estimativa dos coeficientes realizada acima já se consegue aferir quais as variáveis com maior impacto no consumo. Todavia, de forma a confirmar quais as variáveis que têm um impacto mais significativo na explicação da variação do consumo, efetuaram-se os testes de significância individual abaixo.

Formularam-se as seguintes hipóteses para cada variável independente Xj no modelo:

H0: βj=0 (a variável Xj não é estatisticamente relevante para prever CONS)

H1: βj≠0 (a variável Xj é estatisticamente relevante para prever (CONS)

Podemos observar o valor-p associado a cada variável na tabela de coeficientes apresentada na Figura 18. Para todas as variáveis, valor-p < α = 0.05. Ou seja, conclui-se que todas são estatisticamente relevantes para explicar o consumo mensal de eletricidade.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

No entanto, nem todas as variáveis têm o mesmo poder explicativo. As variáveis com maior valor, em módulo, na coluna "coeficientes padronizados – Beta" são as que têm maior impacto na explicação da variação do modelo, neste caso, são as variáveis URB, NUMC e AC. Estes resultados comprovam as conclusões preliminares retiradas pela análise das estimativas dos coeficientes realizada acima.

2.3 <u>Questão 3:</u> Pronuncie-se quanto à validade dos pressupostos do modelo.

O modelo tem os seguintes pressupostos: linearidade nos parâmetros, independência dos erros, variância dos erros constante (homocedasticidade), normalidade da distribuição dos erros e ausência de multicolinearidade. Para analisar a validade dos pressupostos do modelo pode efetuar-se uma análise dos resíduos, já que estes correspondem às estimativas dos erros.

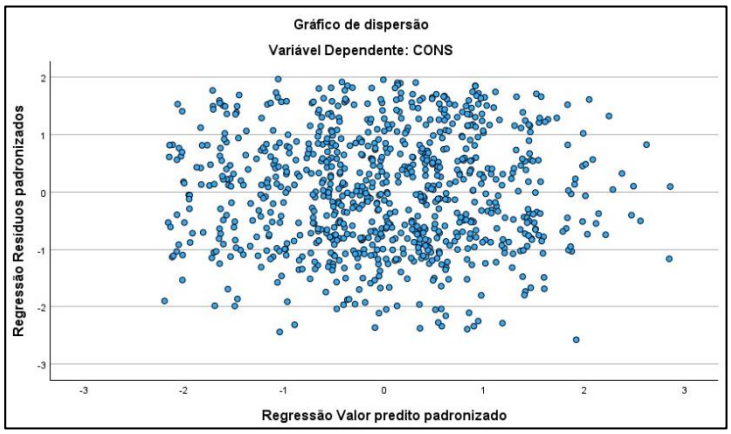

Figura 21: Gráfico de dispersão dos resíduos

Pela análise do gráfico de dispersão dos resíduos podem validar-se os pressupostos de linearidade nos parâmetros, independência dos erros e homoscedasticidade.

Relativamente ao pressuposto de linearidade nos parâmetros, num modelo de regressão linear bem ajustado, espera-se que os resíduos estejam distribuídos aleatoriamente em torno do valor zero, sem um padrão sistemático (curvas, tendências ou agrupamentos), o que se comprova pelo gráfico acima.

Para se verificar a independência dos erros e a homoscedasticiade, o gráfico de dispersão deve apresentar uma mancha de pontosaleatórios com o mesmo tipo de dispersão em torno do eixo das abcissas, o que também se comprova.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Relativamente ao pressuposto da normalidade, realizou-se o teste abaixo.

H0: Os erros seguem uma distribuição normal;

H1: Os erros não seguem uma distribuição normal.

Testes de Normalidade

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estatística	gl	Sig.	Estatística	gl	Sig.
Unstandardized Residual	,053	855	<,001	,983	855	<,001

a. Correlação de Significância de Lilliefors

Figura 22: Testes de normalidade dos erros

Como o valor-p < 0,001 < α = 0,05, rejeita-se H0, ou seja, rejeita-se a hipótese de normalidade, sugerindo que os resíduos não seguem uma distribuição normal.

Relativamete ao pressuposto de ausência de multicolinearidade, pela análise da tabela de coeficientes apresentada na Figura 18, conclui-se que não existem problemas de multicolinearidade, uma vez que os valores de VIF e de tolerância estão próximos de 1.

Concluindo, todos os pressupostos para avaliar a qualidade do modelo se verificam, com exceção do da normalidade da distribuição dos erros. No entanto, como n>30, pelo teorema do Limite Central pode assumir-se a normalidade. Dessa forma, mesmo que a distribuição dos erros não seja estritamente normal, a análise realizada continua válida, e as estimativas podem ser consideradas confiáveis.

CONCLUSÃO

O presente trabalho teve como objetivo aplicar, em situações práticas, os conceitos teóricos adquiridos na disciplina de Análise Estatística de Dados, utilizando o software SPSS29 para conduzir análises estatísticas apropriadas aos cenários propostos.

No primeiro caso analisado, a Análise Fatorial foi utilizada para explorar a satisfação dos passageiros de uma companhia aérea com base num conjunto de variáveis que descrevem diferentes aspetos do serviço. A análise permitiu identificar quatro fatores principais — "Experiência a bordo", "Facilidade e conveniência na experiência de viagem", "Qualidade do serviço a bordo" e "Facilidade do Check-in". Esses fatores explicam os principais elementos que contribuem para a perceção de satisfação dos passageiros.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

No segundo caso estudado, aplicou-se a Regressão Linear Múltipla para examinar os fatores que influenciam o consumo mensal de eletricidade em residências. A análise revelou que, embora todas as variáveis sejam estatisticamente significativas para explicar o consumo, as variáveis URB, NUMC e AC são as que possuem maior poder explicativo. Embora o pressuposto de normalidade dos erros não tenha sido estritamente atendido, como a dimensão da amostra é superior a 30 observações, a análise a análise realizada continua válida, e as estimativas podem ser consideradas confiáveis.

Com base nos resultados obtidos em ambos os casos, conclui-se que as metodologias aplicadas são adequadas para responder às questões propostas, oferecendo *insights* relevantes tanto para a avaliação da satisfação dos clientes quanto para a análise de padrões de consumo energético. Os resultados sugerem que o uso de métodos estatísticos apropriados é essencial para sustentar decisões fundamentadas e orientar ações estratégicas nas áreas investigadas.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt