

## ANÁLISE MULTIVARIADA DE DADOS PARA DEFINIÇÃO DAS MELHORES CONFIGURAÇÕES EM UMA MANUFATURA DE LATAS DE ALUMÍNIO

FELIPE COSTA ARAUJO<sup>1</sup>

### RESUMO

As latas de alumínio representam a principal matriz de embalagens de praticamente todos os produtos consumidos no dia-a-dia, principalmente de líquidos e aerossóis. As manufaturas de latas de alumínio buscam o melhor resultado financeiro possível através da redução de custos de matérias-primas e menor quantidade possível de unidades com defeitos. Diversas análises estatísticas podem ser realizadas com o objetivo de auxiliar na tomada de decisões. Este estudo utilizou uma base de dados de uma empresa de grande porte de manufatura de latas de alumínio no processo de estampagem do Top Wall. Através do agrupamento desta base de dados e normatização com o Escore Z, se aplicaram algumas técnicas de análises estatísticas multivariadas (correlação de Pearson, análise fatorial e análise de agrupamento por Cluster) com a intenção de apoiar na otimização e customização do processo de estampagem do Top Wall. As técnicas multivariadas vêm apoiar na geração de um Ranking com a identificação das melhores configurações e de grupos com os melhores resultados. O objetivo central deste estudo foi gerar um impacto financeiro positivo bastante significativo nos custos de fabricação buscando as configurações de máquinas e matérias-primas com a menor quantidade possível de material, a menor amplitude de medidas e a menor quantidade de unidades fora da especificação mínima e máxima do Top Wall.

**Palavras-chave:** Análises Multivariadas, Correlação de Pearson, Análise Fatorial, Agrupamento por cluster, Top Wall.

---

<sup>1</sup> Engenheiro de Produção - UNIFEI (Universidade Federal de Itajubá), Mestrando em Administração na FEA/USP Ribeirão Preto. FEA/USP Ribeirão Preto. E-mail: fcaraujo@usp.br

---

## MULTIVARIATE DATA ANALYSIS FOR DEFINING THE BEST CONFIGURATIONS IN A MANUFACTURE OF ALUMINUM CANS

### ABSTRACT

*Aluminum cans represent the main packaging matrix of virtually all products consumed on a day-to-day basis, mainly liquids and aerosols. Aluminum can manufacturers seek the best financial result possible by reducing raw material costs and fewer possible defective units. Several statistical analyzes can be performed with the purpose of assisting in decision making. This study used a database of a large aluminum cans manufacturing company in the Top Wall stamping process. Through the grouping of this database and standardization with the Z-Score, some techniques of multivariate statistical analysis (Pearson's correlation, factorial analysis and clustering analysis per cluster) were applied with the intention of supporting in the optimization and customization of the stamping process of the Top Wall. The multivariate techniques come to support the generation of a Ranking with the identification of the best configurations and groups with the best results. The main objective of this study was to generate a significant positive financial impact on manufacturing costs by searching for the configurations of machines and raw materials with the lowest possible amount of material, the smallest range of measurements and the smallest number of units outside the minimum specification and top wall.*

**Keywords:** *Multivariate Analysis, Pearson's Correlation, Factor Analysis, Clustering by cluster, Top Wall.*

## 1. INTRODUÇÃO

A manufatura de latas de alumínio é uma indústria de vital importância em qualquer economia, pois representa em termos quantitativos, a principal matriz de embalagens de líquidos e aerossóis. A manufatura de latas de alumínio basicamente transforma bobinas de alumínio em diversos tamanhos e especificações de latas através de diversos processos de fabricação, principalmente de estampagem mecânica.

Este estudo foi feito a partir de uma base de dados disponibilizada por uma manufatura de grande porte do setor de fabricação de latas de alumínio no último processo de estampagem, também chamado de estampagem do Top Wall. O Top Wall representa a borda superior que tem uma função muito importante de sustentação da tampa superior. O Top Wall precisa ter boas características mecânicas já que precisa resistir ao processo de soldagem da tampa superior e também suportar os esforços mecânicos necessários durante o transporte e armazenagem. Na figura 1 abaixo se encontra uma sequência operacional aproximada do processo de manufatura de latas de alumínio.



Figura 1 - Sequência operacional aproximada do processo de fabricação de latas de alumínio

Fonte: Sandvik (<http://www.home.sandvik/en/>)

A base de dados disponibilizada apresenta os valores medidos por amostragem no processo produtivo de latas de alumínio no momento da estampagem do Top Wall. Algumas latas foram amostradas e retiradas do processo para a realização da medida em laboratório. Foram realizadas as medidas do peso da lata, valor do Top Wall máximo e valor do Top Wall mínimo. O peso da lata tem um valor nominal de 8,430 gramas, porém sempre se busca o menor peso possível

para otimização da razão de número de unidades produzidas por quantidade de matéria-prima (bobina de alumínio). Todas as latas apresentam uma especificação máxima e mínima da medida do Top Wall. Ou seja, todas as latas têm o seu Top Wall medido e necessitam estar acima do mínimo de 0,6 mm e abaixo do máximo de 0,65 mm para não serem rejeitadas e excluídas do processo de manufatura.

O processo de estampagem do Top Wall é realizado por diferentes máquinas com diferentes características de ferramentais, pressão, moldes, etc. Nesta base de dados se dispõe de dados amostrais de sete diferentes máquinas de estampagem (BM18, BM21, BM22, BM23, BM24, BM25 e BM26) que serão bastante importantes para a comparação dos resultados e verificar se existem diferenças significativas entre elas. Além disso, a base de dados conta com três diferentes bobinas de alumínio (Bobinas M, H1 e L1), que é a matéria-prima principal desta manufatura. Com diferentes especificações de bobinas se podem comparar e verificar se as pequenas diferenças de especificação entre cada bobina é significativa para o processo de fabricação como um todo.

Com o auxílio da estatística e de técnicas de análise multivariada de dados é que se pode entender como as variáveis se interagem, buscando correlações, fatores de explicação e similaridades entre as variáveis. De acordo com Triola (2008), a estatística multivariada pode ser definida como uma coleção de métodos para planejar experimentos, obter dados, interpretar, analisar e extrair conclusões. Contudo se deve tomar um cuidado para não tirar conclusões gerais que vão além da base de dados original disponibilizada (TRIOLA, 2008).

Para qualquer análise multivariada é muito importante a normatização dos dados, também conhecido como padronização dos dados. Segundo Fávero et al. (2009), as variáveis devem ser normatizadas de acordo com o método do Escore Z, onde cada variável passou a ter média igual a zero, o desvio padrão igual a um e variância de cada atributo também seja igual a um. Com os atributos selecionados e o agrupamento dos dados normatizados pelo Escore Z, é possível aplicar técnicas multivariadas, que são definidas como um conjunto de métodos estatísticos que possibilita a análise simultânea de múltiplas variáveis (CORRAR, 2014).

A primeira técnica utilizada no estudo é a Correlação de Pearson, que é representada como uma medida de associação linear entre variáveis quantitativas (FIGUEIREDO FILHO et al, 2014). O coeficiente de correlação de Pearson varia entre -1 e 1, sendo que o sinal indica a direção da correlação (negativa ou positiva)

enquanto que o valor indica a magnitude. Quanto mais perto de 1 mais forte é o nível de associação linear entre as variáveis e quanto mais perto de zero, menor é o nível de associação. Uma correlação positiva indica que quando x aumenta, y também aumenta, ou seja, valores altos de x estão associados a valores altos de y. Uma correlação negativa indica que quando x aumenta, y diminui, ou seja, valores altos de x estão associados a valores baixos de y (FIGUEIREDO FILHO et al, 2014). A correlação de Pearson explica a interação existente entre os atributos com o intuito de descrever relações entre os pesos e dimensões do Top Wall, gerando contribuições importantes na busca das melhores configurações possíveis.

A análise fatorial vem apoiar na descrição dos relacionamentos das covariâncias entre os atributos em termos de certos fatores, que basicamente são agrupadas segundo suas correlações. As variáveis são agrupadas por meio das correlações fortemente relacionadas entre si e pouco relacionadas com as variáveis de outro grupo (JOHNSON, WICHERN, 2007). Também se pode definir a análise fatorial como uma técnica multivariada de interdependência dos dados que visa relacionar as variâncias observadas em um conjunto de variáveis inter-relacionadas, com o objetivo de identificar fatores comuns (FÁVERO et al., 2009). Também se pode definir a análise fatorial como a combinação linear das variáveis originais num número menor de variáveis, denominadas fatores comuns não relacionados entre si (CORRAR, 2007). Finalmente a análise fatorial tem o objetivo de entender estes fatores e tratar de explicar estas covariâncias encontradas que é estimada com base na variância comum. Conforme Fávero et al. (2009), o próximo passo consiste em decidir quantos fatores devem ser retidos. Os critérios que auxiliam ao pesquisador na tomada dessa decisão são o critério de Kaiser (também chamado de raiz latente) que calcula o número de fatores a serem retidos em função do número de autovalores (Eigenvalues) acima de 1. Um outro método utilizado para definir o número recomendado de fatores é a análise paralela (Parallel Analysis). Nesta análise, os fatores encontrados correspondem aos números de influências causais contidos na base de dados, em termos de um conjunto comum de fatores com diferentes pesos relativos (HARSHMAN, LUNDY, 1994).

Além das correlações de Pearson e análise fatorial, também foi realizada a análise de agrupamentos por Clusters definidos a partir de similaridades entre as variáveis representadas. Este estudo pode ser definido como um estudo exploratório a partir das técnicas de Análise de Clusters Hierárquico (AC), que é uma técnica de

interdependência que permite o agrupamento de casos ou variáveis em agrupamentos homogêneos a partir de critérios de similaridade (FÁVERO, 2009). O primeiro passo de qualquer análise de agrupamento por Clusters é encontrar a matriz de distância entre as variáveis padronizadas. Segundo Corrar (2007), a padronização das variáveis antes do agrupamento em Clusters é recomendada para minimizar o efeito das diferentes escalas de medidas, fazendo com que todas as variáveis exerçam a mesma importância na definição dos grupos. O método escolhido foi a distância euclidiana como medida de dissimilaridade com parâmetros métricos, já que as variáveis representam medidas de peso ou dimensão (Top Wall). Existem diversos métodos de agrupamento hierárquicos existentes, tais como: Ligação Simples, Ligação Completa, Ligação Média, Método Ward, dentre outros. O Método Ward é um dos procedimentos mais utilizados, pois visa a minimização da perda de informações no momento de juntar dois grupos resultando na melhor combinação dos grupos com um mínimo de perda de informação e minimização da variância dentro dos grupos, conhecido como ESS (JOHNSON, WICHERN, 2007).

## 2. METODOLOGIA DA PESQUISA

O objetivo central deste estudo foi buscar a configuração com o menor peso de Lata, menor amplitude e menor quantidade de latas fora da especificação MIN (6,00) e MAX (6,50). O menor peso de lata reflete basicamente a produtividade da matéria-prima comprada (alumínio) em forma de bobinas. Ou seja, mesmo se definindo o valor do peso de maneira nominal em 8,430, quanto menor o peso de cada lata se podem fazer mais unidades com a mesma quantidade de matéria-prima (alumínio). A amplitude de cada lata é representada pela diminuição do Top Wall Máximo pelo Top Wall Mínimo. Uma menor amplitude também implica em uma menor variabilidade do processo e uma maior chance de evitar que o Top Wall esteja fora da especificação (MÍN = 6,00 e MÁX = 6,50).

### 2.1 Seleção dos Atributos

Observa-se que nesta base de dados existem somente três medidas (Peso da Lata, Top Wall Máximo e Top Wall Mínimo). Porém, para efeitos de comparação,

é necessário encontrar os melhores atributos para comparar e encontrar as melhores configurações possíveis. A seguir se encontram detalhados os atributos selecionados a partir dos objetivos propostos. Todos os atributos foram obtidos através da normatização de todos os dados com o Escore Z.

a) Peso da Lata (PesoZ): Escore Z do peso de cada lata em relação a média do peso de todas as latas amostradas.

b) Amplitude média (AMPZ): Escore Z da amplitude média de cada lata (Diferença entre Top Wal MAX e Top Wal MIN) em relação à média da amplitude de todas as latas amostradas.

c) Limite Inferior (LIZ): Escore Z da soma quadrática das latas com peso menor que 6,05 em relação a média do Limite Inferior de todas as latas amostradas. Definiu-se o valor de 6,05 por ser um pouco maior que o limite inferior mínimo especificado pelo processo (6,00), pois com um valor menor que 6,05 existe uma grande possibilidade de rejeição no processo.

d) Limite Superior 1 (LSZa): Escore Z da soma quadrática das latas com peso maior que 6,40 em relação a média do Limite Superior de todas as latas amostradas. Definiu-se o valor de 6,40 por ser um pouco menor que o limite superior especificado pelo processo (6,50). A ideia deste atributo foi verificar as configurações que apresentam muitos valores próximos do valor de rejeição do limite superior no processo.

e) Limite Superior 2 (LSZb): Escore Z da soma quadrática das latas com peso maior que 6,50 em relação a média do Limite Superior de todas as latas amostradas. O valor de 6,50 foi definido como a medida máxima do Top Wall especificado para rejeição das latas fabricadas.

É importante evidenciar que todos os atributos selecionados são inversamente proporcionais ao seu rendimento. Ou seja, quanto menor o número normatizado encontrado, melhor é o resultado do rendimento da variável em cada atributo normatizado.

## 2.2 Técnicas Multivariadas

A partir do agrupamento e normatização da base de dados de acordo com os atributos definidos anteriormente, se procedeu com a obtenção da matriz de correlação entre os cinco atributos selecionados (PesoZ, AMPZ, LIZ, LSZa e LSZb).

O método de correlação de Pearson foi utilizado para a elaboração das três matrizes identificadas a seguir. Para a realização dos cálculos das matrizes de correlação deste estudo se utilizou o software R-Studio, pacote adicional de funções GPArotation e a função de correlação Cor().

A análise fatorial irá utilizar o método de solução dos mínimos residuais (minres) e método de rotação Varimax, segundo indicado no referencial teórico. O método Varimax visa minimizar o número de variáveis que apresentam altas cargas em um fator, facilitando a interpretação dos fatores. Além disso, este é o método mais utilizado atualmente (FÁVERO, 2009). Portanto, a partir das correlações de Pearson indicadas anteriormente, se realizou uma análise fatorial para cada uma das matrizes indicadas. Reuniram-se as três matrizes de correlações em matriz de correlação agrupada com média dos valores encontrados nas três matrizes anteriores. Para a realização dos cálculos das matrizes de correlação deste estudo se utilizou o software R-Studio, pacote adicional de funções GPArotation e funções de análise fatorial fa(), obtenção dos autovalores eigen() e realização da análise paralela parallel().

Para realizar a análise de agrupamento das variáveis por clusters é necessário fazer a matriz de distância com as matrizes de normalizadas por Máquinas e por Máquina-Bobina. A matriz normalizada por Bobina não foi utilizada nesta análise de agrupamentos em virtude do baixo número de variáveis disponíveis (somente três bobinas). Para esta análise se realizou uma matriz de distância euclidiana com parâmetros métricos (medidas de pesos e dimensões do Top Wall). Para a construção das tabelas e gráficos referentes às análises de agrupamentos por Clusters se utilizou o software R-Studio, pacote de funções MASS e com as funções dist() para matriz de distância, hclust (method="ward") para elaboração dos agrupamentos por Clusters Hierárquicos do Método Ward e cutree() para construção do gráfico do Dendograma.

Após estas análises se realiza um Ranking de desempenho de cada Cluster e variável identificada de acordo com os atributos selecionados. Este Ranking busca encontrar as melhores configurações possíveis para chegar aos objetivos propostos de baixo peso, baixa amplitude e menor número de unidades fora de especificação. É importante detalhar que este estudo teve por objetivo apoiar a tomada de decisão da empresa para melhorar seu desempenho financeiro, com a diminuição dos custos de fabricação e otimização do processo de maneira geral.



### 3. RESULTADOS E DISCUSSÕES

A partir da seleção dos atributos, agrupamento e normatização da base de dados, foram aplicadas as técnicas de análise multivariadas de dados, a saber, correlação de Pearson, análise fatorial e agrupamento por Clusters. Os resultados de cada uma das técnicas multivariadas foram apresentados nos próximos itens de acordo com as metodologias definidas anteriormente.

#### 3.1 Agrupamento dos dados segundo os Escore Z

De acordo com a escolha dos atributos descritos anteriormente foram construídas três matrizes de dados normatizados. A 1ra matriz de valores foi o agrupamento dos dados por Máquina (BM) abaixo na Tabela 1. Neste caso pretende-se avaliar e comparar as sete máquinas utilizadas pelo processo produtivo buscando encontrar as relações entre elas para identificar as melhores configurações possíveis. A matriz de bobinas (Tabela 2) foi construída com o agrupamento dos dados por Bobina sendo importante para avaliar se a matéria-prima utilizada interfere de maneira significativa no performance do processo produtivo segundo os atributos selecionados. Na 3ra matriz (Tabela 3), os dados são agrupados de acordo com o conjunto Máquina-Bobina. Esta matriz une as variáveis Máquinas e Bobinas para identificar como cada conjunto interfere nos resultados finais e a possível relação entre eles.

Tabela 1 – Matriz de Valores Normatizados por Máquina (BM)

	PesoZ	AMPZ	LIZ	LSZa	LSZb
BM18	-0,245	-0,880	1,481	-1,098	-1,346
BM21	0,434	0,436	0,365	0,486	0,753
BM22	0,353	1,424	-0,499	0,982	1,306
BM23	1,114	-0,932	-1,383	-0,636	-0,824
BM24	0,217	1,109	-0,476	1,444	0,531
BM25	-0,933	-0,927	-0,503	-1,043	-0,869
BM26	-0,939	-0,229	1,016	-0,134	0,450

Tabela 2 – Matriz de Valores Normalizados por Bobina

	PesoZ	AMPZ	LIZ	LSZa	LSZb
BobinaM	0,692	0,538	-0,833	0,995	1,013
BobinaH1	-1,147	-1,154	1,109	0,009	-0,027
BobinaL1	0,454	0,616	-0,276	-1,005	-0,986

Tabela 3 – Matriz de Valores Normalizados por Máquina-Bobina

	PesoZ	AMPZ	LIZ	LSZa	LSZb
BM18M	-0,343	-0,010	-0,528	-0,623	-0,701
BM21M	0,619	0,298	-0,019	1,285	1,952
BM22M	1,431	1,021	-0,548	1,902	1,567
BM23M	1,518	-0,539	-0,698	-0,533	-0,693
BM24M	0,301	1,614	-0,644	2,413	-0,722
BM25M	-0,653	-0,134	-0,405	-0,398	0,184
BM26M	-0,936	0,730	-0,617	1,178	1,718
BM18H1	-0,650	-0,806	3,231	-0,501	-0,507
BM21H1	-0,465	0,149	1,234	0,121	-0,221
BM22H1	-0,078	-0,423	-0,019	0,223	-0,288
BM23H1	1,153	-1,837	-0,698	0,171	0,231
BM24H1	0,077	-1,103	-0,617	1,143	2,196
BM25H1	-1,289	-1,570	-0,377	-0,623	-0,707
BM26H1	-0,935	-0,800	1,849	-0,485	-0,350
BM18L1	0,586	-0,719	-0,602	-0,959	-0,722
BM21L1	1,776	0,313	-0,698	-0,485	-0,650
BM22L1	-0,242	1,886	-0,140	-0,264	0,593
BM23L1	0,539	0,749	-0,568	-0,844	-0,720
BM24L1	0,355	1,423	0,585	-0,819	-0,713
BM25L1	-0,659	0,087	0,068	-0,956	-0,722
BM26L1	-0,951	-0,328	0,210	-0,947	-0,722

### 3.2 Matrizes de Correlação de Pearson

Foram definidas três matrizes de correlação de Pearson de acordo com as variáveis e atributos agrupados. Estas três matrizes buscam identificar as correlações existentes de acordo com o agrupamento por Máquinas, Bobinas e o conjunto Máquina-Bobina. As matrizes podem ser visualizadas nas Tabelas a seguir 4, 5 e 6.

Tabela 4 – Matriz de Correlação de Pearson por Máquina

	PesoZ	AMPZ	LIZ	LSZa	LSZb
PesoZ	1,000	0,390	-0,624	0,258	0,134
AMPZ	0,390	1,000	-0,303	0,410	0,327
LIZ	-0,624	-0,303	1,000	-0,215	-0,165
LSZa	0,258	0,410	-0,215	1,000	0,735
LSZb	0,134	0,327	-0,165	0,735	1,000

Tabela 5 – Matriz de Correlação de Pearson por Bobina

	PesoZ	AMPZ	LIZ	LSZa	LSZb
PesoZ	1,000	0,658	-0,658	0,074	0,095
AMPZ	0,658	1,000	-0,633	-0,031	-0,010
LIZ	-0,658	-0,633	1,000	-0,181	-0,201
LSZa	0,074	-0,031	-0,181	1,000	0,666
LSZb	0,095	-0,010	-0,201	0,666	1,000

Tabela 6 – Matriz de Correlação de Pearson por Máquina-Bobina

	PesoZ	AMPZ	LIZ	LSZa	LSZb
PesoZ	1,000	0,163	-0,480	0,251	0,101
AMPZ	0,163	1,000	-0,133	0,301	0,086
LIZ	-0,480	-0,133	1,000	-0,238	-0,196
LSZa	0,251	0,301	-0,238	1,000	0,662
LSZb	0,101	0,086	-0,196	0,662	1,000

Foram encontradas basicamente duas correlações mais fortes nas três matrizes de correlação, sendo uma negativa entre PesoZ e LIZ e outra positiva entre LSZa e LSZb. O Peso da Lata e o Limite Inferior do Top Wall (LIZ) apresentam uma relação inversamente proporcional, quanto maior é o Peso da Lata menor é o Limite Inferior do Top Wall e vice versa. Esta correlação pode estar associada ao fato de latas com menores pesos apresentarem maiores problemas de Limite Inferior abaixo do estabelecido pelo processo. A correlação positiva entre os dois limites superiores indica que existe uma forte relação entre as latas com Top Wall maiores que 6,40 e 6,50 respectivamente. Isto indica que existe uma grande probabilidade de valores acima destes dois limites independente da Máquina e Bobina utilizada.

Outras correlações mais moderadas também foram encontradas, ambas positivas, entre PesoZ e AMPZ (principalmente na matriz de Bobinas) e entre AMPZ e LSZa (principalmente na matriz de Máquinas). A primeira correlação entre PesoZ e AMPZ indica que quanto maior o Peso das Latas maior a variação na amplitude das mesmas. Além disso, existem fortes indícios que o PesoZ e a AMPZ sofrem muita variabilidade quando se utilizam diferentes Bobinas. A segunda correlação moderada entre AMPZ e LSZa indica que uma elevada amplitude afeta muito mais

o limite superior do que o limite inferior das latas. Isto é mais perceptível na Matriz por Máquina, onde parece indicar que cada Máquina utilizada tem comportamentos bastante diferentes.

Por fim, algumas correlações mais fracas também foram identificadas entre LIZ e LSZa e entre AMPZ e LISZ (ambas negativas). A correlação entre o limite inferior e o limite superior indica que as latas com problemas de limite inferior muito dificilmente terão problemas com o limite superior. Já a correlação entre amplitude e limite inferior reflete uma relação, ainda que fraca, entre a probabilidade de relação entre estar abaixo do Limite Inferior e o tamanho da amplitude.

### 3.3 Análise Fatorial

A análise fatorial busca utilizar as matrizes anteriores de correlação e encontrar fatores resultantes que explicam a variância nas variáveis e atributos identificados. O primeiro passo dos resultados é definir o número ótimo de fatores recomendados através do Método de Kaiser (Autovalores maiores que 1) e Análise Paralela, segundo gráfico 1 abaixo.

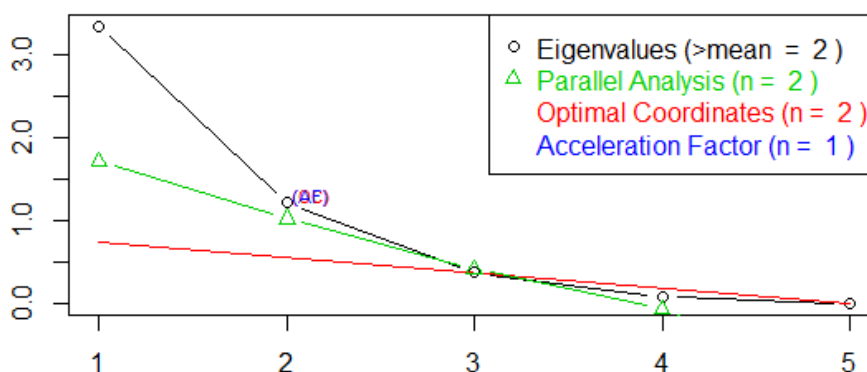


Gráfico 1 - Método dos Autovalores, análise paralela e coordenadas ótimas

De acordo com o identificado acima no Gráfico 1, obtivemos que o número ótimo de fatores recomendados é de dois fatores. Abaixo se encontram os resultados e as principais considerações relativas aos resultados encontrados.

	MR2	MR1	h2	u2	com
PesoZ	0.09	0.94	0.90	0.10	1.0
<u>AMPZ</u>	<u>0.38</u>	<u>0.38</u>	<u>0.29</u>	<u>0.71</u>	<u>2.0</u>
LIZ	-0.12	-0.65	0.44	0.56	1.1
LSZa	0.89	0.18	0.83	0.17	1.1
LSZb	0.81	0.06	0.66	0.34	1.0

	MR2	MR1
SS loadings	1.62	1.49
Proportion Var	0.32	0.30
Cumulative Var	0.32	0.62
Proportion Explained	0.52	0.48
Cumulative Proportion	0.52	1.00

Analisando estes resultados encontrados, pode se notar que o atributo AMPZ apresenta um valor baixo de comunalidade h2 (0,29) e também não está altamente relacionado com nenhum dos dois fatores identificados (MR1 e MR2). Todas as outras variáveis apresentam comunalidades mais adequadas e uma boa relação com os dois fatores identificados, verificando que a variância total explicada com dois fatores é de 62%.

O fator MR1 representa a busca de menos material e dimensão acima do Limite Inferior sendo responsável pela grande variação no Peso das latas (PesoZ) que é inversamente proporcional ao Limite Inferior (LIZ). Ou seja, este fator indica que quanto mais tratamos de atingir o objetivo de utilizar menos material (Peso) por lata, mais elevada será a rejeição no processo por latas com dimensões menores que o Limite Inferior. A Variância total explicada por este fator MR1 é de 30%.

O fator MR2 é responsável pelo limite superior de dimensão do Top Wall e busca explicar a variação dos dados no limite superior (6,40 e 6,50) de dimensão do Top Wall e não está relacionado com o Peso e tampouco com a amplitude. Este fator indica que existe uma relação forte entre os dois Limites Superiores indicados, onde se pode dizer que existe uma correlação muito forte entre os dois atributos de limites superiores. A Variância explicada por este fator é de 32%.

### 3.4 Análise do agrupamento por Clusters

Os resultados de agrupamentos por Clusters estão dividido de acordo com as três matrizes normatizadas com o Escore Z. Porém a análise foi feita somente nas matrizes por máquinas e por conjunto máquina-bobina. A matriz por bobina não foi utilizada nesta análise de agrupamentos em virtude do baixo número de variáveis disponíveis (somente três bobinas). A partir disso se podem realizar várias considerações, encontrar grupos de variáveis com desempenhos similares e tratar de encontrar as melhores configurações.

Primeiramente se apresentam os resultados da matriz por máquinas. Segundo o gráfico do Dendograma (Gráfico 2) e a tabela de agrupamento (Tabela 7) abaixo se podem dividir as sete máquinas de estampagem disponíveis em três grupos com desempenhos bastante similares entre si. A Tabela 7 foi dividida em diferentes cores para uma melhor visualização dos Clusters encontrados.

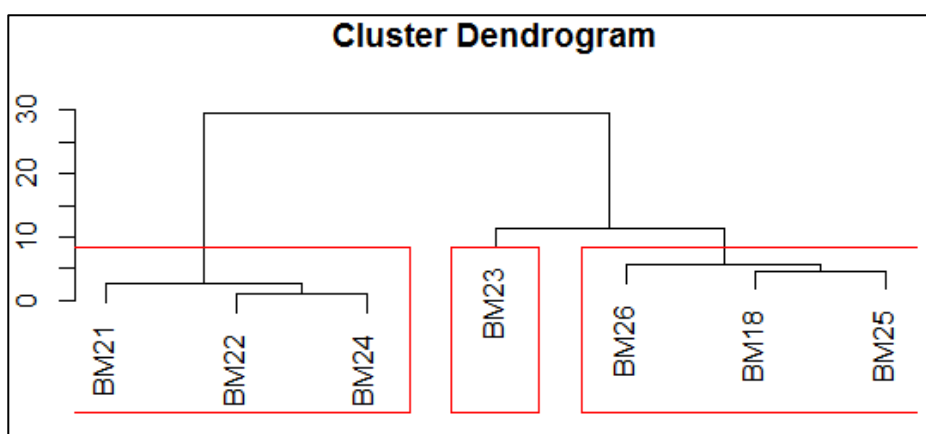


Gráfico 2 - Dendograma de agrupamento da matriz por Máquina com 3 Clusters

Tabela 7 – Análise de agrupamento da matriz por Máquinas com 3 Clusters

	PesoZ	AMPZ	LIZ	LSZa	LSZb
BM21	0,434	0,436	0,365	0,486	0,753
BM22	0,353	1,424	-0,499	0,982	1,306
BM24	0,217	1,109	-0,476	1,444	0,531
BM23	1,114	-0,932	-1,383	-0,636	-0,824
BM18	-0,245	-0,880	1,481	-1,098	-1,346
BM25	-0,933	-0,927	-0,503	-1,043	-0,869
BM26	-0,939	-0,229	1,016	-0,134	0,450

O Cluster 1 é representado por três máquinas (BM21, BM22 e BM24) e tem várias características comuns entre si tais como: elevado peso das latas (PesoZ),

elevada amplitude de medidas (AMPZ) e elevado número de latas com dimensão acima de ambos Limites Superiores. Pode-se caracterizar estas máquinas como de baixo desempenho de maneira geral. O Cluster 2 representa somente a máquina BM23 que têm características diferentes de todas as outras máquinas analisadas. Esta máquina têm um peso muito elevado, porém com uma baixa amplitude aliado a um número pequeno de latas fora da especificação dos limites superiores e inferiores. Já o Cluster 3 com as máquinas BM18, BM25 e BM26 apresenta características bem distintas dos outros grupos identificados anteriormente. As características são um baixo peso, uma baixa amplitude, um baixo número de latas fora da especificação do limite superior e um número maior de latas com problemas nos limites inferiores.

Igualmente como na análise anterior das matrizes por máquinas, se encontraram vários resultados significativos com a matriz por conjunto máquina-bobina detalhados abaixo no gráfico do Dendograma (Gráfico 3) e tabela de agrupamento com cinco Clusters identificados (Tabela 8). Esta tabela foi dividida em diversas cores de acordo com o grupo de cada conjunto encontrado para uma melhor visualização.

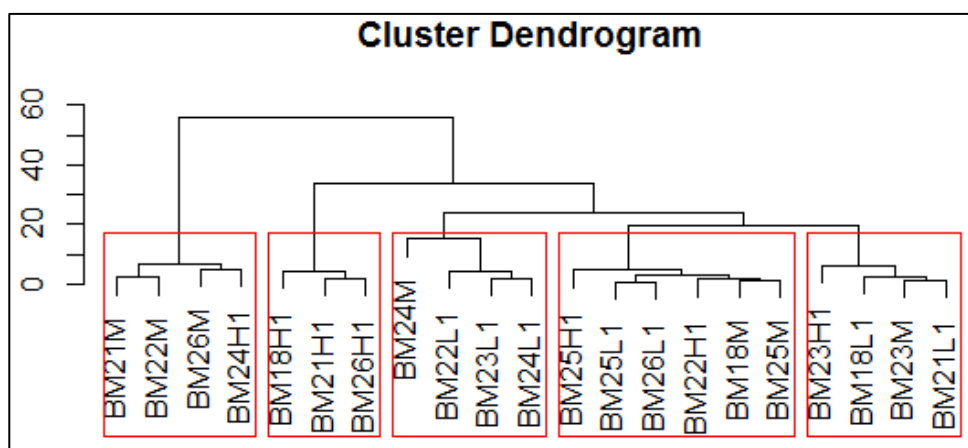


Gráfico 3 - Dendograma de agrupamento por conjunto Máquina-Bobina com cinco Clusters

Tabela 8 – Análise de agrupamento por conjunto máquina-bobina com cinco Clusters

	PesoZ	AMPZ	LIZ	LSZa	LSZb
BM18M	-0,343	-0,010	-0,528	-0,623	-0,701
BM21M	0,619	0,298	-0,019	1,285	1,952
BM22M	1,431	1,021	-0,548	1,902	1,567
BM23M	1,518	-0,539	-0,698	-0,533	-0,693
BM24M	0,301	1,614	-0,644	2,413	-0,722
BM25M	-0,653	-0,134	-0,405	-0,398	0,184
BM26M	-0,936	0,730	-0,617	1,178	1,718
BM18H1	-0,650	-0,806	3,231	-0,501	-0,507
BM21H1	-0,465	0,149	1,234	0,121	-0,221
BM22H1	-0,078	-0,423	-0,019	0,223	-0,288
BM23H1	1,153	-1,837	-0,698	0,171	0,231
BM24H1	0,077	-1,103	-0,617	1,143	2,196
BM25H1	-1,289	-1,570	-0,377	-0,623	-0,707
BM26H1	-0,935	-0,800	1,849	-0,485	-0,350
BM18L1	0,586	-0,719	-0,602	-0,959	-0,722
BM21L1	1,776	0,313	-0,698	-0,485	-0,650
BM22L1	-0,242	1,886	-0,140	-0,264	0,593
BM23L1	0,539	0,749	-0,568	-0,844	-0,720
BM24L1	0,355	1,423	0,585	-0,819	-0,713
BM25L1	-0,659	0,087	0,068	-0,956	-0,722
BM26L1	-0,951	-0,328	0,210	-0,947	-0,722

Após o procedimento do agrupamento por Clusters de acordo com a similaridade entre eles, pode-se chegar a algumas considerações bastante significativas. O Cluster 1 agrupando os conjuntos BM18M, BM25M, BM22H1, BM26L1, BM25L1 e BM25H1 apresenta como características em comum o fato de apresentarem um baixo peso, uma baixa amplitude e valores relativamente baixos de limites inferiores e superiores. O Cluster 2 que agrupa os conjuntos BM23H1, 23M, 18L1 e 21L1 apresenta um peso bastante elevado, uma amplitude baixa e também um valor de limite inferior bastante baixo. Já o Cluster 3 agrupando os conjuntos BM21M, BM22M, BM26M e BM24H1 apresentam valores de pesos variados, elevada Amplitude e principalmente limites superiores muito mais elevados relativamente. O Cluster 4 que agrupa os conjuntos BM24M, BM22L1, BM23L1 e BM24L1 apresenta um peso intermediário, uma elevada amplitude e baixo número de latas fora das especificações superiores e inferiores. E finalmente o Cluster 5 que conta com os conjuntos BM18H1, BM26H1 e BM21H1 apresenta um baixo peso relativo, uma baixa amplitude, um limite inferior muito elevado indicando muitas latas fora de especificação mínima.



### 3.4 Seleção das melhores variáveis (Ranking)

A partir disto se pode definir que as matrizes normalizadas juntamente com a análise por agrupamentos em Clusters representam uma boa métrica para indicar os melhores grupos e também as melhores variáveis. A última coluna das tabelas 9, 10 e 11 a seguir representa o valor do Ranking atribuído a cada uma das variáveis estudadas. Este Ranking é a soma de todos os atributos das matrizes normalizadas.

A tabela 9 apresentada abaixo representa o Ranking das melhores Máquinas agrupadas segundo o agrupamento por Clusters. A Máquina BM25 foi a melhor entre todas porque apresenta um baixo peso e amplitude aliado a limites inferiores e superiores também baixos. Considera-se que esta Máquina é a mais consistente em todos os atributos. Já o melhor Cluster é o número 3 (Máquinas BM18, BM25 e BM26) por apresentar de maneira geral um baixo Peso e amplitude, além de poucos problemas de rejeição com o limite superior. O limite inferior não é um atributo consistente para todas as máquinas consideradas neste Cluster.

Tabela 9 – Ranking das melhores Máquinas agrupadas em 3 Clusters

	PesoZ	AMPZ	LIZ	LSZa	LSZb	Ranking
BM25	-0,933	-0,927	-0,503	-1,043	-0,869	-4,274
BM23	1,114	-0,932	-1,383	-0,636	-0,824	-2,662
BM18	-0,245	-0,880	1,481	-1,098	-1,346	-2,088
BM26	-0,939	-0,229	1,016	-0,134	0,450	0,164
BM21	0,434	0,436	0,365	0,486	0,753	2,474
BM24	0,217	1,109	-0,476	1,444	0,531	2,823
BM22	0,353	1,424	-0,499	0,982	1,306	3,564

A tabela 10 representa o Ranking das melhores bobinas sem nenhum agrupamento por Clusters pelo pequeno número de bobinas disponíveis na base de dados. Seria necessário contar com um número maior de variáveis de diferentes bobinas para chegar a conclusões mais abrangentes e significativas. Porém neste caso a melhor bobina é representada pela H1 por apresentar um valor de pesos e amplitudes mais baixos que as bobinas L1 e M. Os valores de limites inferiores e superiores está em um nível superior às demais porém na somatória do Ranking a bobina H1 tem o menor valor de todas.

Tabela 10 – Ranking das melhores Bobinas

	PesoZ	AMPZ	LIZ	LSZa	LSZb	Ranking
BobinaH1	-1,147	-1,154	1,109	0,009	-0,027	-1,209
BobinaL1	0,454	0,616	-0,276	-1,005	-0,986	-1,197
BobinaM	0,692	0,538	-0,833	0,995	1,013	2,405

Por possuir um número maior de variáveis (21 no total), esta análise apresenta dados bastante significativos para análise. A tabela 11 representa, igualmente como nas tabelas anteriores, o Ranking dos melhores conjuntos Máquina-Bobina agrupados nos cinco Clusters discutidos nos itens anteriores. O melhor conjunto pode ser atribuído ao conjunto BM25H1 por ser a mais consistente entre todos indicado por um valor baixo em todos os atributos selecionados. Já o melhor Cluster é o número 1 representado pelos conjuntos BM18M, BM25M, BM22H1, BM26L1, BM25L1 e BM25H1. Este Cluster conta com vários conjuntos com algumas diferenças entre eles, porém apresenta como características principais em comum um baixo peso e amplitude relativos e também poucos problemas de rejeição com os limites inferiores e superiores.

Cabe destacar também o Cluster número 2 com os conjuntos BM23H1, 23M, 18L1, 21L1. Este Cluster 2 apresenta valores bastante significativos em amplitude, limites inferiores e superiores, porém com um peso mais elevado. O peso elevado acaba por diminuir muito o performance geral dos conjuntos deste Cluster.

Tabela 11 – Ranking dos melhores conjuntos Máquina-Bobina agrupados em cinco Clusters

	PesoZ	AMPZ	LIZ	LSZa	LSZb	Ranking
BM25H1	-1,289	-1,570	-0,377	-0,623	-0,707	-4,566
BM26L1	-0,951	-0,328	0,210	-0,947	-0,722	-2,739
BM18L1	0,586	-0,719	-0,602	-0,959	-0,722	-2,416
BM18M	-0,343	-0,010	-0,528	-0,623	-0,701	-2,204
BM25L1	-0,659	0,087	0,068	-0,956	-0,722	-2,182
BM25M	-0,653	-0,134	-0,405	-0,398	0,184	-1,407
BM23H1	1,153	-1,837	-0,698	0,171	0,231	-0,979
BM23M	1,518	-0,539	-0,698	-0,533	-0,693	-0,945
BM23L1	0,539	0,749	-0,568	-0,844	-0,720	-0,844
BM26H1	-0,935	-0,800	1,849	-0,485	-0,350	-0,721
BM22H1	-0,078	-0,423	-0,019	0,223	-0,288	-0,585
BM21L1	1,776	0,313	-0,698	-0,485	-0,650	0,257
BM18H1	-0,650	-0,806	3,231	-0,501	-0,507	0,767
BM21H1	-0,465	0,149	1,234	0,121	-0,221	0,818
BM24L1	0,355	1,423	0,585	-0,819	-0,713	0,832
BM24H1	0,077	-1,103	-0,617	1,143	2,196	1,696
BM22L1	-0,242	1,886	-0,140	-0,264	0,593	1,832
BM26M	-0,936	0,730	-0,617	1,178	1,718	2,072
BM24M	0,301	1,614	-0,644	2,413	-0,722	2,962
BM21M	0,619	0,298	-0,019	1,285	1,952	4,135
BM22M	1,431	1,021	-0,548	1,902	1,567	5,373

#### 4. CONCLUSÕES FINAIS

Este estudo visou contribuir com a busca das melhores configurações possíveis ao processo produtivo das latas de alumínio na estampagem do Top Wall. A partir dos cinco atributos selecionados (PesoZ, AMPZ, LIZ, LSZa, LSZb), a normalização dos dados com os Escore Z e as análises multivariadas descritas anteriormente foi possível chegar a diversas considerações para apoiar as manufaturas de latas de alumínio na otimização do processo de estampagem do Top Wall e minimização da rejeição das latas fora de especificação.

As técnicas multivariadas de dados foram imprescindíveis para entender as correlações existentes entre as variáveis e buscar as melhores configurações possíveis. Com a união das informações e considerações obtidas com as técnicas multivariadas se pode construir um Ranking com o somatório dos atributos normalizados de cada uma das variáveis e Clusters identificados. Os melhores Clusters e as melhores variáveis foram os que apresentaram os resultados mais consistentes e constantes em todos os atributos selecionados. De maneira geral a melhor máquina foi a BM25, a melhor bobina foi a H1 e o melhor conjunto Máquina-Bobina foi a BM25H1.

Sugere-se algumas recomendações para futuros trabalhos relacionadas à análise multivariada de dados no processo de manufatura de latas de alumínio: (1) averiguar mais profundamente o processo produtivo para definição dos atributos mais significativos na busca dos objetivos de desempenho e otimização do processo; e (2) identificar os atributos que mais contribuem para o performance do processo e com isso definir diferentes métricas de avaliação.

## REFERÊNCIAS BIBLIOGRÁFICAS

CORRAR, L.J. ; Paulo, E. ; Dias Filho, J.M. **Análise multivariada para os cursos de Administração**, Ciências Contábeis e Economia; Atlas, 2007.

FIGUEIREDO FILHO, D. B.; ROCHA, E. C.; SILVA JR, J.; PARANHOS, R; A., NEVES, J. A. B. e SILVA, M. B. **Desvendando os Mistérios do Coeficiente de Correlação de Pearson: o Retorno** ; Leviathan: Cadernos de Pesquisa Política ; N. 8, pp. 66-95, 2014.

TRIOLA, M. F. **Introdução à Estatística**; 10<sup>a</sup> ed. Rio de Janeiro: LTC, 2008.

JOHNSON , R. A., WICHERN, D. W. **Applied multivariate statistical analysis**; 6a edição ; Upper Saddle River ; Pearson Prentice-Hall International, 2007.

L.P. Fávero, P. Belfiore, F.L. da Silva, B.L. Chan. **Análise de Dados: Modelagem Multivariada para tomada de decisões**; Elsevier, 2009.

HARSHMAN, R. A., LUNDY, M. E. **PARAFAC: Parallel factor analysis**; Computational Statistics & Data Analysis ; Elsevier ; Volume 18, Issue 1 ; August 1994 ; p. 39-72.