

# Data Analysis of Three Procedures for Constructing Semiconductors: Mold, Solder Ball and Singulation

Brígida Mónica Faria

DETI/UA – Departamento de Electrónica, Telecomunicações e Informática / Universidade de Aveiro  
ESTSP/IPP – Escola Superior de Tecnologia da Saúde do Porto / Instituto Politécnico do Porto  
Aveiro e Porto, Portugal  
btf@estsp.ipp.pt

**Abstract** — Semiconductor's production is a complex process that involves several sequential steps until obtaining the final product. Team and machine fault detection and performance analysis, in the manufacturing process, are essential to the success of a company in this field. This paper presents the results obtained by applying multivariate statistical methods to data on three processes of the production line of a semiconductor company: Molding, Solder Ball Attach and Singulation. From the analysis of the real data it can be concluded that there are statistical differences regarding the performance, concerning the time taken and number of faults, of the four teams that operate on the company analyzed.

**Keywords** - Semiconductor, Mold, Solder Ball, Singulation, Principal Component Analysis, Clustering.

## I. INTRODUÇÃO

Actualmente as bases de dados são fontes poderosas, sendo possível extrair mais informação do que aquela que está aparentemente disponível. Novas técnicas de predição, extracção e de relação foram desenvolvidas em áreas como Análise de Componentes Principais, *Clustering* e Análise Factorial [1]. Contudo, uma importante fase na análise de dados consiste no tratamento, organização e redução de dados de forma a extrair a informação que de facto é essencial aos objectivos propostos.

Neste trabalho está presente um estudo sobre os dados reais de uma empresa de semicondutores. O conjunto de dados aqui analisado refere-se a uma amostra de dados de Novembro de 2008 cuja informação incide em três processos da linha de produção da empresa: *Molding*, *Solder Ball Attach* e *Singulation*. Após uma análise preliminar dos dados, um dos objectivos foi aplicar métodos estatísticos multivariados. Inicialmente, de modo a reduzir o número de variáveis iniciais, foi utilizada uma abordagem de Análise de Componentes Principais. Em seguida, foi estudada a análise de clusters e sua aplicação nos dados. O objectivo primordial do trabalho consistiu em analisar a eficiência das equipas que trabalham nestes três processos relativamente a tempos e falhas cometidas.

O resto do artigo encontra-se organizado do seguinte modo: na secção II é efectuada uma breve descrição da empresa, o seu modo de funcionamento e o processo de produção de

semicondutores; na secção III são descritos os dados e feita a sua caracterização, tratamento e análise dos resultados. Finalmente são apresentadas as conclusões e perspectivas de desenvolvimento futuras.

## II. PRODUÇÃO DE SEMICONDUCTORES

### A. Funcionamento da Empresa

Uma empresa que se dedica à produção de semicondutores pretende obter produtos com enorme variedade de aplicações, através de tecnologia de baixo consumo e alto rendimento, à integração em computadores e outros terminais digitais, tais como, leitores de MP3, telemóveis, câmaras fotográficas digitais, consolas de jogos, entre outros [2] [3].

### B. Processo de Produção de Semicondutores

A produção dos semicondutores passa várias fases até ser obtido o produto final como pode ser analisado pela Fig. 1. Neste trabalho foram só analisados os dados relativos a três processos sequenciais [4]:

- **Mold** é o que se designa por moldagem e consiste na cobertura dos chips com um polímero que protege os chips e os fios de ouro que fazem as conexões entre chip e substratos (cada memória vai ser depois anexada em módulos para as aplicações de destino).
- **Solder Ball Attach ou Ball Placing** é a fase de soldadura de bolas que permite a transmissão de informação entre o chip e os módulos.
- **Singulation** (de cada unidade) é o corte no substrato de modo a tornar cada chip um elemento único pronto a ser montado. Este processo também é denominado por Individualização.

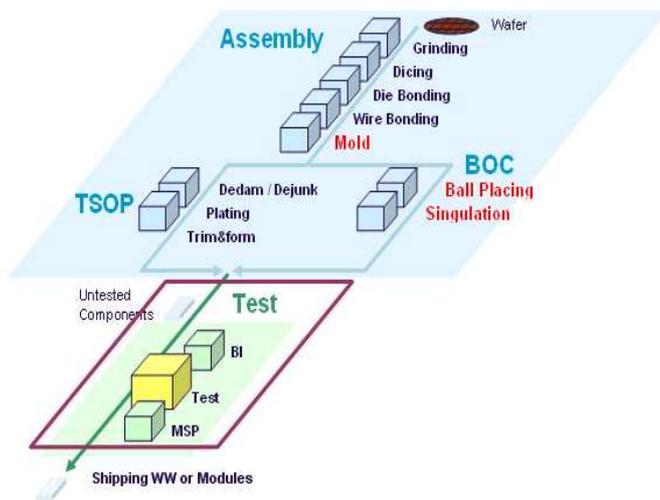


Figura 1. Processos de construção de semi-condutores [Adaptado de [4]].

A amostra recolhida refere-se a 438 lotes e para cada um deles foi recolhido o seguinte conjunto de dados:

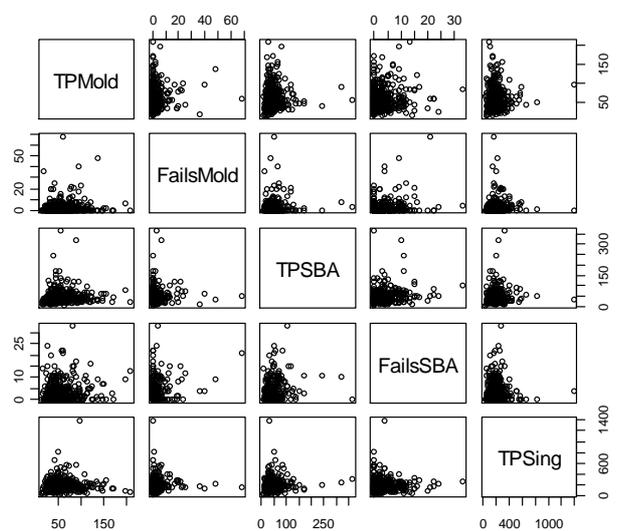
- Quantidade de semicondutores de cada lote;
- Hora de entrada e hora de saída em cada um dos processos (que permite calcular o tempo de processamento de cada lote);
- Equipamento onde foi processado o lote;
- Equipa que processou o lote;
- Número total de falhas em cada processo.

O estudo será focalizado sobre as variáveis relativas ao tempo em minutos que cada lote se encontrou nas 3 etapas, o número de falhas também por fase e as equipas envolvidas nos processos. A informação relativa à quantidade do lote, após uma análise inicial dos dados, foi eliminada, uma vez que se consideram os lotes como um indivíduo.

### III. ANÁLISE DOS DADOS

O primeiro passo do trabalho consistiu em manipular os dados de forma a traçar um caminho que permitisse obter informação útil para a empresa. A Fig. 2 contém os diagramas de dispersão referentes aos dados observados nas variáveis tempos e falhas nos três processos.

Analisando as observações não é possível afirmar que exista uma relação estatística entre as variáveis. O caso dos *outliers* (valores que aparentemente são extraviados do que se poderia esperar [3]) poderia estar a mascarar essas mesmas relações. Foi construído o diagrama de dispersão retirando os casos extremos, contudo e pensando em termos de empresa este tipo de dados não deverá ser retirado, uma vez que serão importantes para a compreensão do comportamento e desempenho das equipas.



TPMold – tempo em moldagem; FailsMold – falhas em moldagem; TPSBA – tempo em Solder Ball Attach; FailsSBA – falhas em Solder Ball Attach; TPSing – tempo em Singulation

Figura 2. Diagramas de dispersão ligados

Assim sendo, o passo seguinte consistiu em elaborar uma análise exploratória das medidas de tendência central e de dispersão para as variáveis tempo decorrido nos processos e número de falhas contabilizadas por processo. As tabelas I e II revelam informação relativamente aos extremos, média, desvio padrão, 1º, 2º e 3º quartis.

TABELA I: TEMPOS DECORRIDOS NOS TRÊS PROCESSOS

	Mold	SBA	Sing
Mínimo	15	8	24
1º Quartil	39	26	121,2
Mediana	50	35,5	166,5
Média	57,34	44,07	186,4
3º Quartil	68,75	51	222,8
Máximo	209	63	1388,0

TABELA II: NÚMERO DE FALHAS NOS TRÊS PROCESSOS

	Mold	SBA	Sing
Mínimo	0	0	0
1º Quartil	0	0	0
Mediana	1	2	1
Média	2,589	3,247	2,352
3º Quartil	3	4	3
Máximo	68	33	30

Todavia coloca-se a questão do significado destes dados relativamente ao comportamento das equipas envolvidas nas três fases.

TABELA III: TEMPOS DA EQUIPA 1

	Mold	SBA	Sing
Mínimo	15	9	47
1º Quartil	36	27	109
Mediana	46	<b>34</b>	162
Média	52,55	<b>43,95</b>	187,6
3º Quartil	61,75	47,5	230
Máximo	198	320	<b>1388,0</b>
Desvio Padrão	27,02	35,48	144,93

É de salientar que a equipa 1 apresenta em média um tempo menor no processo de Soldadura de Bolas e se a observação for efectuada atendendo à mediana, que será menos susceptível aos valores extremos comprova-se que no processo de Soldadura também corresponde a um valor menor. O processo mais demorado é o de Individualização (Sing). Note-se que o valor máximo atingido neste caso foi de aproximadamente 1388 segundos, o mais elevado apresentado por todas as equipas, como pode ser observado na tabela III.

TABELA IV: TEMPOS DA EQUIPA 2

	Mold	SBA	Sing
Mínimo	16	17	64
1º Quartil	38,25	31,5	130,2
Mediana	<b>53,5</b>	<b>41,5</b>	170,5
Média	<b>55,87</b>	<b>54,67</b>	192,3
3º Quartil	66,75	63,25	232,5
Máximo	209	363	659
Desvio Padrão	26,62	47,26	98,49

Na equipa 2 o comportamento apresentado relativamente aos primeiros dois processos e atendendo à média de tempos é mais próxima do que a diferença na equipa 1, contudo se observarmos a mediana a diferença de valores corresponde ligeiramente à mesma classe de valores (tabela IV).

TABELA V: TEMPOS DA EQUIPA 3

	Mold	SBA	Sing
Mínimo	16	12	41
1º Quartil	39	24	127
Mediana	50	<b>35</b>	164
Média	58,44	<b>40,45</b>	185,9
3º Quartil	68	48	215,8
Máximo	168	172	821
Desvio Padrão	28,40	23,40	102,62

Através da tabela V conclui-se que no caso da equipa 3 também o processo em média menos demorado é na fase de Soldadura de Bolas (SBA), enquanto que o mais demorado é o de Individualização (Sing).

TABELA VI: TEMPOS DA EQUIPA 4

	Mold	SBA	Sing
Mínimo	29	8	24
1º Quartil	47	26	122,5
Mediana	<b>61</b>	<b>34</b>	166
Média	<b>70,07</b>	<b>41,75</b>	182,9
3º Quartil	81	48	214
Máximo	148	244	581
Desvio Padrão	30,49	28,81	92,68

Por fim a equipa 4 apresenta o mesmo desempenho ao nível médio de tempo, porém a discrepância entre a Moldagem (Mold) e a Soldadura de Bolas é mais acentuada do que nas restantes equipas.

O passo seguinte envolveu a comparação dos tempos recorrendo aos *boxplots*, já que através deste tipo de gráfico é possível concluir sobre a existência ou não de *outliers*.

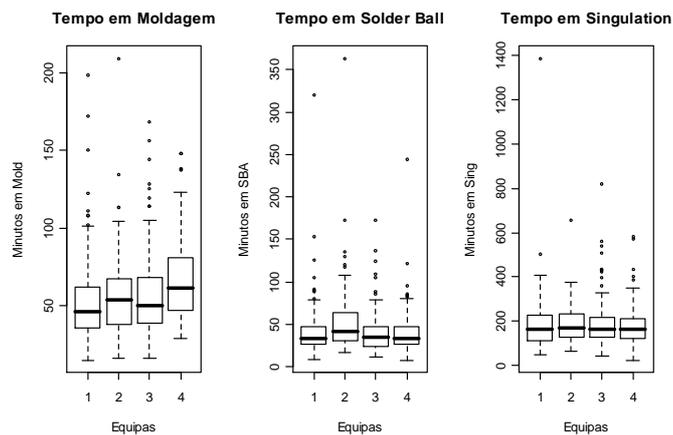


Figura 3. Boxplots dos tempos pelas 4 equipas e processos de Moldagem, Soldadura de Bolas (Solder Ball) e Individualização (Singulation).

Após a análise dos *boxplots* (Fig. 3) relativamente aos tempos que os lotes das equipas permanecem nos três processos consegue-se inferir que todas apresentam valores *outliers* e que na fase da Moldagem a equipa que manifestamente apresenta valores superiores é a equipa 4. Contudo, esta equipa também tem um número menor de *outliers*. Relativamente ao processo Soldadura de Bolas a equipa 4 já demonstra tempos semelhantes e possivelmente inferiores relativamente às outras duas equipas. No processo de Individualização os resultados são semelhantes pelas 4 equipas, sendo de realçar o facto de a equipa 1 apresentar um *outlier* bastante elevado. E no processo de Individualização os tempos serem superiores relativamente aos outros dois processos.

Por um lado o estudo apresentado focou as variáveis de tempo, mas uma questão importante para a empresa centra-se especialmente nas falhas, já que uma falha (o que poderá significar destruição do produto) representará elevados prejuízos. Na Fig. 4 são apresentados os *boxplots* para as falhas nos três processos e por equipas.

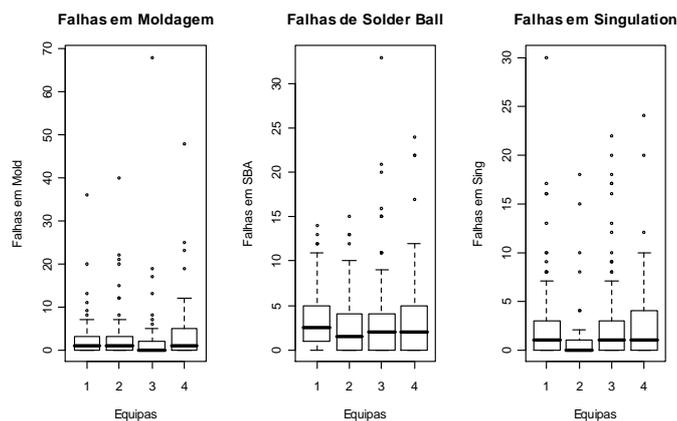


Figura 4. Boxplots das falhas apresentadas pelas 4 equipas nos processos de Moldagem, Soldadura de Bolas (Solder Ball) e Individualização (Singulation).

Neste caso o número de falhas *outliers* apresentadas pela equipa 4 são menores, sendo que a equipa 3 apresenta os *outliers* mais elevados. Através da mediana consegue-se saber

que pelo menos 50% das observações encontram-se abaixo desse valor e a equipa 2 no processo de Moldagem e Individualização tem o valor de mediana igual ou inferior ao das restantes equipas.

Sabendo através de informações da empresa que os três processos são sequenciais construíram-se os diagramas de dispersão fixando as equipas nos processos sendo que surgiram 3 casos de análise como mostra a Fig. 5: Tempo em Moldagem vs Tempo em Soldadura; Tempo em Moldagem vs Tempo em Individualização e Tempo em Soldadura vs Tempo em Individualização.

Assim sendo, para a equipa 1 (cor azul) do processo de moldagem foram representados os tempos dos lotes que estão em moldagem (e são produzidos pela equipa 1) associando também ao tempo de Soldadura de Bolas (eixo dos yy's). A abordagem para as outras equipas (com cor preta para a equipa 2, cor verde para a equipa 3 e equipa 4 com cor vermelha) em Moldagem foi análoga. Os restantes dois casos já mencionados anteriormente foram impressos de forma análoga. Em Tempo de Moldagem vs Tempo de Individualização foram fixadas as equipas em Tempo de Moldagem e associava-se ao tempo em Individualização, por último tem-se as equipas fixadas em Tempo de Soldadura e associa-se ao Tempo em Individualização.

Analisando os diagramas de dispersão da Fig. 5 verifica-se que no primeiro diagrama de dispersão os lotes que foram sujeitos à equipa de Moldagem 2 apresentam um diagrama menos concentrado, relativamente ao tempo em Soldadura. Em seguida e centrado a atenção no segundo diagrama de dispersão verifica-se que os lotes que foram sujeitos à equipa 1 em Moldagem é o menos concentrado, relativamente ao tempo em Individualização. Através do terceiro diagrama de dispersão não é possível, de uma forma clara, afirmar qual a equipa que trabalhou nos lotes em Soldadura que apresenta os pares de tempos menos concentrados havendo dúvidas relativamente à equipa 2, 3 e 4, relativamente aos tempos em Individualização.

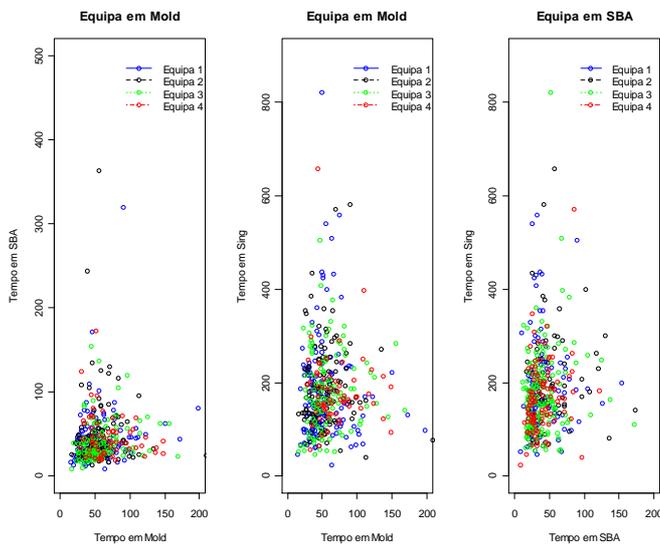


Figura 5. Diagramas de dispersão para Tempo em Moldagem vs Tempo em Soldadura; Tempo em Moldagem vs Tempo em Individualização e Tempo em Soldadura vs Tempo em Individualização.

Após verificação dos pressupostos foi aplicado o teste estatístico ANOVA [5] com o objectivo de verificar as diferenças relativamente às médias de tempos das equipas para um nível de significância de 0,05. Constatou-se que relativamente à média de tempo em Moldagem existem evidências estatísticas para afirmar que são diferentes (valor prova inferior a 0,001). Em particular através do teste de *Tuckey* [5] a equipa 4 apresenta valores distintos das outras três equipas. Para os tempos em Soldadura e Individualização não existem evidências estatísticas para afirmar que a média de tempos é diferente nas quatro equipas.

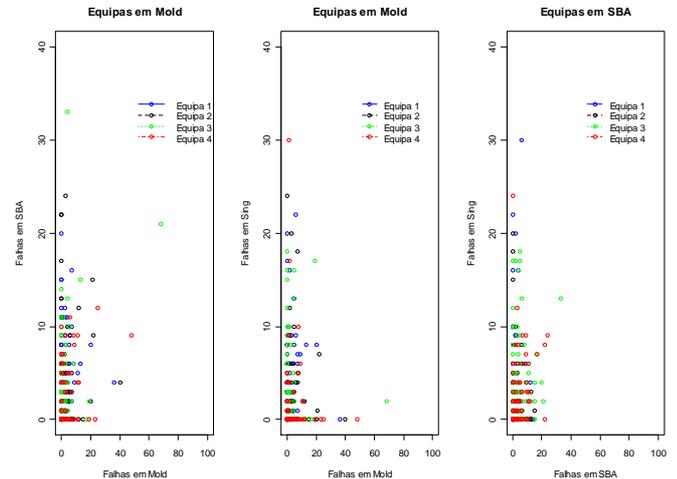


Figura 6. Diagramas de Dispersão com falhas em Moldagem vs falhas em Soldadura; falhas em Moldagem vs falhas em Individualização e falhas em Soldadura vs falhas em Individualização).

Relativamente aos lotes que foram classificados como falhas, o estudo aqui apresentado e que se pode observar na Fig. 6 contempla os 3 casos análogos aos que foram estudados para tempos. Assim sendo, verifica-se que há alguns lotes que apresentaram poucas falhas no processo de Moldagem e que apresentaram um grande número de falhas na fase de Soldadura e Individualização. Em termos de comparação do desempenho das equipas, não houve evidências estatísticas para afirmar que a média de falhas seja diferente das quatro equipas na fase de Moldagem e Soldadura. Contudo no processo de Individualização e através do teste de *Tuckey* a equipa 2 e a equipa 3 apresentaram evidências estatísticas significativas que a média de falhas seriam diferentes nestas duas equipas.

#### A. Análise de Componentes Principais

Um das consequências da Análise em Componentes Principais é reduzir a dimensionalidade dos dados e proporcionar relevância de relações entre variáveis e estabelecer uma nova interpretação dos dados não visíveis com as variáveis originais [1]. O método das componentes principais baseia-se no pressuposto de que se podem definir  $q$  variáveis estatisticamente não correlacionadas a partir de combinações lineares de  $p$  indicadores iniciais (as componentes principais) [6] [7] [8]. As variáveis utilizadas para a redução serão neste caso os tempos, porque são variáveis contínuas. Neste trabalho optou-se por elaborar uma Análise de Componentes Principais referentes às equipas que são fixadas

em Moldagem. Contudo o estudo deveria ser alargado e fica como proposta de trabalho futuro que poderia ser efectuado o estudo fixando as equipas nos outros processos. É de salientar também que esta técnica não requer uma distribuição normal multivariada para os dados e que neste caso não é verificada. As variáveis referentes aos tempos dos 3 processos são contínuas e estão na mesma unidade de medida, portanto é preferível trabalhar com as componentes principais utilizando a matriz de covariâncias.

Desta forma tem-se a nível de exemplo o output produzido pelo *software* R [9] para a determinação das componentes principais produzidas para a equipa 1. A Fig. 7 contém na primeira linha os desvios padrão das componentes principais, enquanto que a segunda parte representa a matriz dos *loadings*, ou seja, a matriz cujas colunas contém os vectores próprios.

```
> pc1
Standard deviations:
[1] 108.44842 32.62981 25.09616

Rotation:
          PC1      PC2      PC3
TPMold -0.0001612591  0.48566165 -0.87414686
TPSBa  0.0361306957  0.87357895  0.48533946
TPSing 0.9993470602 -0.03150527 -0.01768817
> |
```

Figura 7. Output produzido pelo R na análise de componentes principais para os tempos da equipa 1.

```
> summary(pc1)
Importance of components:
          PC1      PC2      PC3
Standard deviation 108.448 32.6298 25.0962
Proportion of Variance 0.874 0.0791 0.0468
Cumulative Proportion 0.874 0.9532 1.0000
```

Figura 8. Output produzido pelo R na análise de componentes principais sobre a importância das componentes para os tempos da equipa 1.

Relativamente aos dados sobre a variabilidade explicada pelas duas primeiras componentes principais, na equipa 1 (Fig. 8), corresponde a aproximadamente 95%. Significa portanto que as duas componentes principais podem substituir as 3 variáveis originais de tempo com pouca perda de informação. Para as restantes equipas os valores obtidos sobre a percentagem de variabilidade explicada altera consoante a equipa. De facto, para a equipa 2 as duas primeiras componentes principais explicam 94% de variabilidade dos dados, para a equipa 3 as duas primeiras componentes explicam 98% e para a equipa 4 as duas primeiras componentes explicam 93%.

Os *scree plots* resultantes podem ser observados na Fig. 9, a partir dos quais e pela “regra do cotovelo” foram admitidas 2 componentes principais para cada uma das equipas.

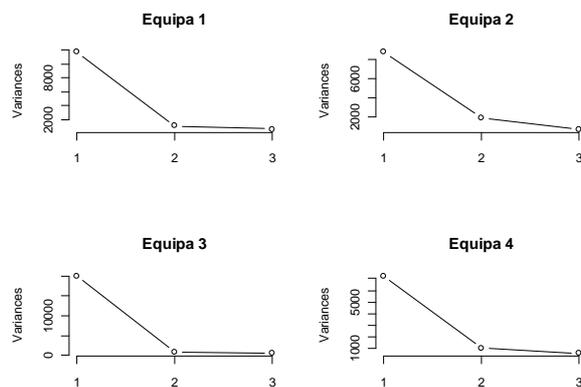


Figura 9. *Scree Plots* para a análise de componentes principais dos tempos por equipas.

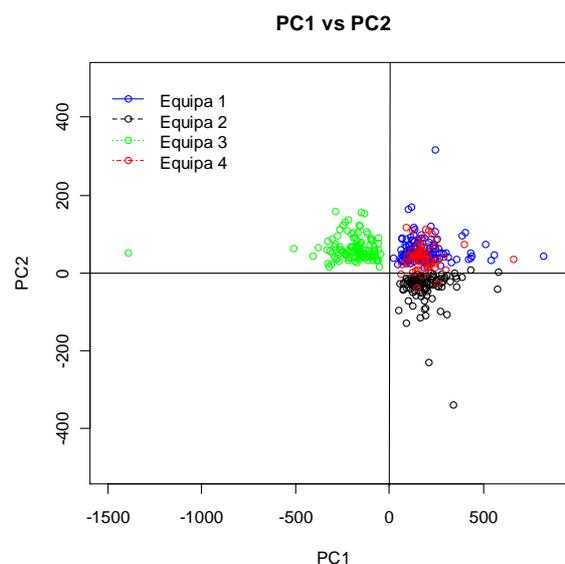


Figura 10. Representação gráfica dos *loadings* usando a matriz de covariâncias.

Comparativamente, a análise das equipas através das suas duas componentes principais, é perceptível que a equipa 3 em Moldagem (a cor verde) é a que se desvia mais relativamente às outras como pode ser observado na Fig. 10. A equipa 2 em Moldagem (a cor preta) também demonstra um comportamento menos homogéneo, isto é, com alguma oscilação. As equipas 1 e 4 comportam-se com o padrão mais semelhante.

### B. Análise de Clusters

A análise de *clusters* permite a divisão do conjunto inicial de dados em vários subconjuntos de dados ou o conjunto inicial de variáveis em vários subconjuntos de variáveis [6]. Através da análise de *clusters*, entre outras conclusões, é possível sugerir hipóteses relativamente à associação entre variáveis.

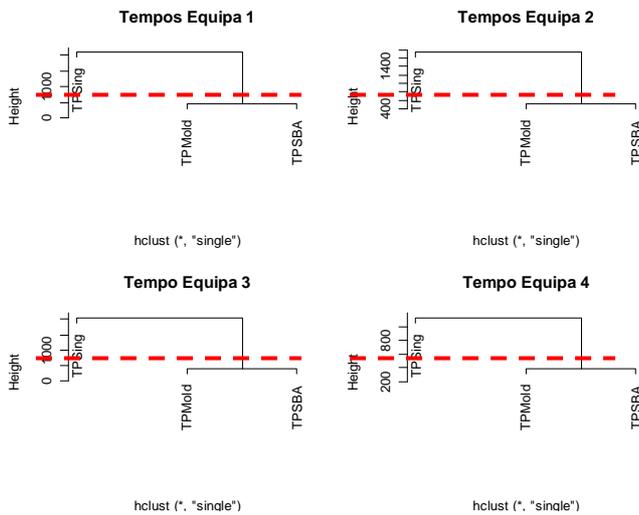


Figura 11. Dendrogramas para a análise de clusters em relação aos tempos dos 3 processos.

Na continuação do estudo foi possível através de dendrogramas estabelecer grupos para os tempos relativamente às equipas em Moldagem (Fig. 11). Sendo assim, optando pelas linhas de corte assinaladas é possível dividir em dois grupos para cada uma das equipas. Como seria de prever há dois grupos distintos: o primeiro referente ao tempo de Individualização e o segundo grupo relativamente aos tempos referentes ao tempo de Moldagem e de Soldadura de Bolas. De facto como já tinha sido referenciado o tempo em Individualização é superior em todas as equipas isto deve-se ao facto também de ser uma tarefa realmente mais demorada.

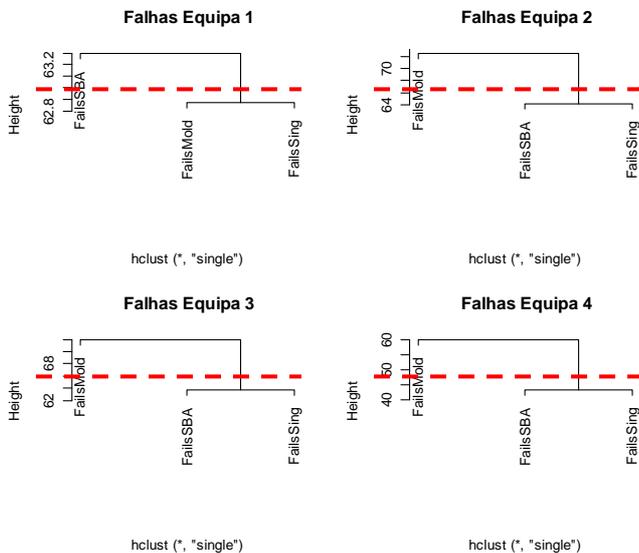


Figura 12. Dendrogramas para a análise de clusters em relação às falhas dos 3 processos

Mas surgiu outra questão se as falhas nos processos também seriam divididos pelo mesmo tipo de grupos relativamente às 3 fases. Assim sendo, e analisando os dendrogramas da Fig. 12, para as falhas é interessante verificar que as equipas 2, 3 e 4 apresentam o mesmo tipo de *clusters*, enquanto que na equipa 1 as falhas dos processos de Moldagem

e Individualização pertencem a um *cluster* e o processo de Soldadura de Bolas a outro *cluster*.

#### IV. CONCLUSÕES E TRABALHO FUTURO

Este trabalho permitiu conhecer em mais pormenor três processos na produção de semicondutores. Foram aplicadas abordagens estatísticas na detecção de diferenças de desempenho por quatro equipas de uma empresa ao nível de tempos e de falhas apresentadas em três processos da linha de produção da empresa: *Molding*, *Solder Ball Attach* e *Singulation*. Da análise realizada é possível concluir que houve comportamentos diferentes das equipas que podem ser analisados através de análise estatística multivariada. Para extrair informação será muitas vezes necessário colocar várias questões quer ao nível de caracterização de dados, quer ao nível de conhecimento dos processos que produzem esses dados. Só uma busca intensiva de relações em dados reais é que será possível obter informação acrescida que muitas vezes não é de fácil acesso. Em termos de sugestões de trabalho futuro será incorporar este tipo de análise que de forma automática se possa entender quais as razões de produção de falhas e se o desempenho varia consoante as equipas ou a factores externos.

#### AGRADECIMENTOS

A autora gostaria de agradecer à Liliana, à Sara, à Telma e ao Virgílio e às seguintes organizações pelo contributo deste trabalho: FCT - Fundação Portuguesa para a Ciência e Tecnologia (Bolsa de Doutoramento com a referência FCT/SFRH / BD / 44541 / 2008), DMat/UA – Departamento de Matemática da Universidade de Aveiro, DETI/UA – Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro, ESTSP/IPP – Escola Superior de Tecnologia de Saúde do Porto, Instituto Politécnico do Porto.

#### REFERÊNCIAS

- [1] A. Field, "Discovering Statistics Using SPSS - Introducing Statistical Methods", 2nd Edition, SAGE Publications, London, 2005. ISBN: 978-0761944522
- [2] G. Phipps, "Wire Bond vs Flip Chip Packaging", Advanced Packaging Magazine, July 2005.
- [3] SEMATECH Inc "Accelerating the next technology revolution", 2011, Disponível em: [http://www.sematech.org/corporate/news/mfgproc/mfgproc.htm#steps1\\_2](http://www.sematech.org/corporate/news/mfgproc/mfgproc.htm#steps1_2). Consultado em: February 2011.
- [4] A. Silva, Backend Process Overview – Overview of Test Process. Infineon Technologies Fabrico de Semicondutores S.A., Março de 2006
- [5] J. Maroco, "Análise Estatística Com o PASW Statistics (ex-SPSS)", Edições Sílabo, 2010, ISBN: 9789899676305
- [6] Pang-Nin Tan, M. Steinback, V. Kumar, "Introduction to Data Mining", Pearson Education, Inc, 2006. ISBN: 0321321367
- [7] R. Elizabeth, Estatística Multivariada Aplicada, Edições Sílabo, 2ª Edição, Lisboa, 2001. ISBN: 972-618-247-6
- [8] H. Manuela, H. Andrew, Investigação por Questionário, Edições Sílabo, 2ª Edição, Lisboa, 2007. ISBN: 972-618-273-5
- [9] Gentleman, R. , Ihaka, R. "The R Project for Statistical Computing". Disponível em: <http://www.r-project.org/>. Consultado em: Janeiro de 2011.