ESCOLA SUPERIOR DE COMUNICAÇÃO SOCIAL

PUBLICIDADE E MARKETING

INFERÊNCIA ESTATÍSTICA

DOCENTE: MARIA JOSÉ CASTRO

TRABALHO REALIZADO POR:

FRANCISCA FÉLIX DA COSTA 5642

FRANCISCO CALVÃO 5617

SARA GUERRA 5652

ÍNDICE

INTRODUÇÃO

No âmbito da unidade curricular Inferência Estatística, foi-nos proposta a realização de um trabalho que identificasse e mostrasse a aplicação das técnicas estatísticas aprendidas em aula, tendo por base o tratamento de dados através do *software* SPSS.

O ficheiro em questão, adaptado ao nosso contexto escolar enquanto alunos da Escola Superior de Comunicação Social, debruça-se sobre um inquérito revisto recentemente pela *University of Sussex*, cujo objectivo passava por analisar quais as características necessárias aos oradores para o desenvolvimento de um bom método de investigação. Seguindo este propósito, foi pedido aos alunos que se colocassem na pele dos investigadores e respondessem a um conjunto de questões que permitissem descobrir, então, quais as características que, na sua opinião, se afirmam como as mais relevantes.

Posto isto e de forma a criar um trabalho não só didáctico mas também divertido, pedimos auxílio a Cathy, uma personagem de banda desenhada criada por Cathy Guisewite. Nascida em 1976, Cathy luta constantemente contra os quatro factores de culpa existentes no seu mundo – Comida, Mãe, Trabalho e Amor – e tem-se vindo a afirmar no nosso mundo como uma mulher moderna, competente e independente, mas sempre algo complicada. Tomando Cathy como nossa personagem principal e utilizando *cartoon strips* tanto totalmente originais como modificadas, pretendemos então passar pelas várias fases inerentes a um trabalho de estatística ilustrando simultaneamente quão divertido, gratificante e por vezes frustrante a sua realização pode ser.

Passando agora a uma breve descrição do conteúdo do trabalho, este pretende cumprir vários objectivos. Em primeiro lugar, Cathy começará por caracterizar a sua amostra aplicando exclusivamente uma análise univariada, de forma a que o leitor se possa familiarizar com as características dos inquiridos e com as suas tendências de respostas para cada uma das perguntas.

Seguidamente, a personagem procederá ao tratamento dos dados aplicando técnicas estatísticas como a Análise Factorial, a Análise de Fiabilidade, a Análise de Clusters, Regressão Logística Binária e Análise de Correspondências, utilizando-se sempre uma margem de erro de 0,05. O procedimento inerente a cada uma destas, bem como os seus objectivos e propósitos, serão explicitados aquando da sua realização. No entanto, faz sentido mencionar desde já que, em conjunto, estas técnicas permitirão verificar correlações entre as variáveis existentes, agrupar os indivíduos de acordo com a sua proximidade de respostas, detectar influências de variáveis independentes sobre a categoria de interesse de uma variável independente, fazer previsões para a população baseadas no

estudo desta amostra, entre outros. Resumidamente permitir-nos-á conhecer a fundo a nossa amostra e tentar retirar ilações que possam ser significativas e representativas para o todo da população. Com base nelas, podemos analisar e interpretar este ficheiro da forma mais completa e conclusiva possível.

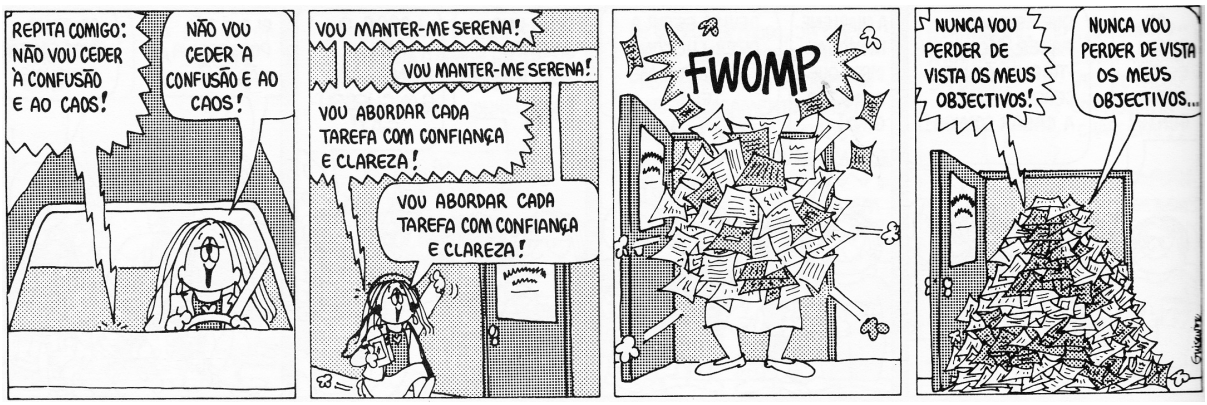
Expostas as fases delineadas para a realização deste trabalho, cabe agora definir o nosso objectivo particular, enquanto grupo, para o mesmo. Pretendemos então criar um trabalho tão original e criativo quanto rigoroso e metódico que desperte a curiosidade do leitor e que o leve a compreender as vantagens oferecidas pela Inferência Estatística percebendo os cruzamentos realizados, as análises incluídas e as ilações retiradas ao mesmo tempo que desfruta de uma leitura agradável e divertida.

O nosso objectivo é, de certa forma, afirmarmo-nos como os embaixadores da Estatística, publicitando-a pelas suas mais valias enquanto ferramenta de trabalho e ilustrando-as durante o processo.

CATHY

Cathy trabalhava para uma empresa chamada Product Testing, Inc desempenhando várias tarefas, muitas delas de carácter impossível. Apesar da sua clara tendência para o drama e constantes queixas, Cathy era uma mulher competente que conseguia sempre entregar os projectos e agradar aos clientes, o que irritava e inspirava Mr. Pinkley, o seu chefe, com quem mantinha uma relação de implicância e certo orgulho.

De uma forma geral, Cathy achava o seu trabalho desafiante mas algo problemático para a sua estabilidade psicológica. Muitas vezes via-se na necessidade de encontrar formas de se motivar e enfrentar com uma postura positiva, o seu dia-a-dia. Claro que isto nem sempre resultava...



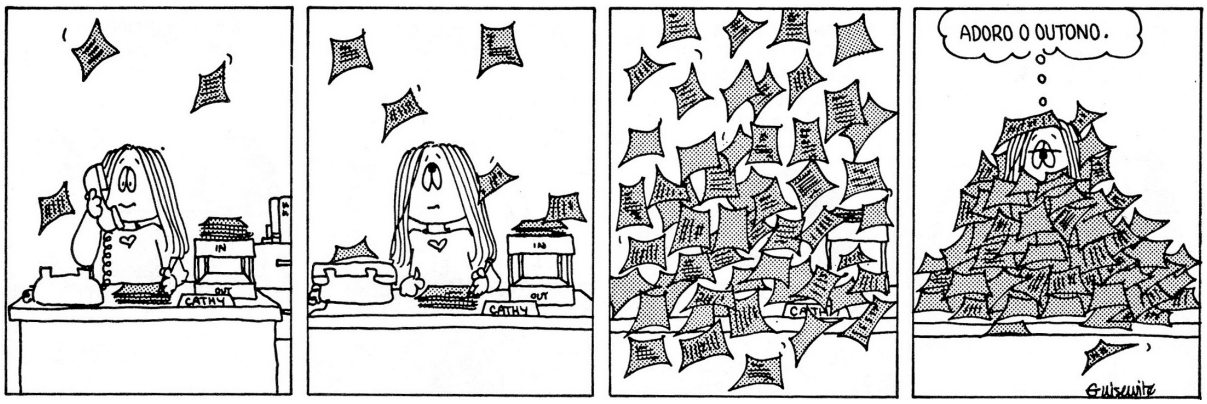
Uma das suas responsabilidades para com a empresa prendia-se com a execução de inquéritos e posterior análise dos dados recolhidos. Cathy esforçava-se para ser metódica e organizada, trabalhando para contrariar vigorosamente a acumulação de trabalho por fazer, as alturas de caos e a tendência existente para a papelada se amontoar na sua secretária. No entanto, por vezes, a sua luta parecia em vão e as forças do universo acabavam sempre por levar a melhor. O caos perseguia-a!

Nestas alturas de desespero, Cathy precisava, muitas vezes, de aliviar a tensão decorrente do enorme stress provocado pelas pilhas de folhas que a soterravam e a comida era a melhor solução para o alívio de quase todos os males. Infelizmente, tinha também a terrível desvantagem de lhe ir parar directamente às coxas, acarretando um sentimento de culpa para o resto do dia e, consequentemente, minando a réstia de produtividade que possuía.

Enquanto mulher activa e empreendedora, Cathy não hesitava em procurar outro tipo de soluções para combater este problema. A leitura era das melhores formas de limpar a secretária!

No Outono de 2010, Cathy deparou-se com um inquérito importantíssimo encomendado pela Universidade de Sussex e cujo objectivo era, nem mais, avaliar os factores decisivos para a criação de um bom método de investigação. Este projecto era não só bastante interessante mas também algo com o qual poderia aprender muito e quiçá a pudesse ajudar a nível particular, a melhorar o seu dia-a-dia. Mr. Pinkley tinha-a avisado e a própria Cathy sabia: Não podia falhar!

Todavia, o primeiro dia de trabalho não augurou nada de positivo...



Frustrada com o difícil avanço do seu trabalho e determinada a mudar a sua situação para melhor, Cathy decidiu agir. Afinal, esta era a ocasião perfeita para mudar o seu método de trabalho e nada a impediria. *"Acabaram-se as folhas e folhinhas!"* – pensou – *"Estamos no século XXI, vou arranjar algo automático!"*

Quando chegou a casa, Cathy sentia-se mais confiante e entusiasmada. Este novo programa, o SPSS, seria a sua salvação e tudo aquilo que andava penosamente a fazer manualmente seria agora substituído por apenas alguns cliques. Claro que nada era perfeito e de certeza que haveria um pequeno senão, mas agora tudo lhe parecia mais fácil e amanhã era outro dia de trabalho. O seu trabalho iria deixar Mr. Pinkley boquiaberto!

CARACTERIZAÇÃO DA AMOSTRA

No dia seguinte, Cathy levantou-se confiante e decidida. Começava hoje uma nova etapa em todo o seu método de trabalho e ia aplicá-lo num projecto tão importante como o que tinha em mãos. Cathy sabia de antemão que a amostra era constituída por 239 indivíduos, todos eles alunos da Escola Superior de Comunicação Social. A tarefa de hoje era caracterizá-la de acordo com as variáveis incluídas no inquérito, descrevendo detalhadamente as suas características sócio-demográficas e tendências de resposta.

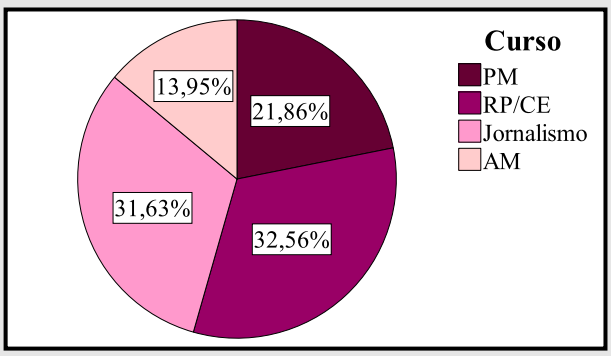
Product Testing, Inc

Relatório *University of Sussex*

Caracterização da Amostra

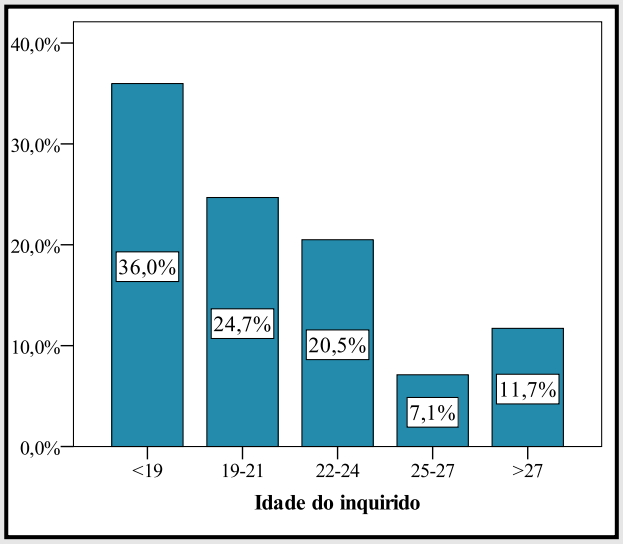
Em primeiro lugar, há que diferenciar a amostra tendo em conta a variável Género. Analisando-a, podemos verificar a existência de 3 respostas inadequadas, 132 casos contabilizados para a categoria de resposta *"Feminino"* e 104 casos para o *"Masculino"*. Constata-se, então, uma maior predominância de mulheres na amostra que constituem 55,9% da mesma.

Relativamente à variável *"Curso"*, e após se constatar a existência de 24 não respostas, a amostra não revelou verdadeira equidade na distribuição dos inquiridos. Esta situação é ilustrada pelo gráfico circular onde se verifica que os cursos de *"Jornalismo"* e *"Relações Públicas e Comunicação Empresarial"* são aqueles que registam maior número de estudantes, obtendo uma percentagem acumulada de 64,2% do total da amostra. Em terceiro lugar, o curso com maior número de inquiridos é *"Publicidade e Marketing"* (21,9%) seguindo-se-lhe *"Audiovisual e Multimédia"* com apenas 14% da amostra.



No que respeita à variável de escala *"Idade"*, a tabela de frequências fornecida pelo SPSS demonstrava uma amplitude de idades bastante elevada sendo o mínimo 17 anos e o máximo 39 anos. No entanto, a distribuição da amostra por este intervalo revelou-se muito pouco homogénea sendo genericamente escassos os casos contabilizados para idades maiores de 26 anos.

Sendo assim, e apesar de perdermos alguma informação, pareceu-nos apropriado recodificar esta variável tornando-a ordinal e dotando-a de 5 diferentes classes que fossem um pouco mais homogéneas e, consequentemente representativas. Cabe ainda frisar que será esta variável de *"Idade do Inquirido"* que utilizaremos doravante no seguimento do trabalho. Estas vêem-se explicitadas no gráfico adjacente.

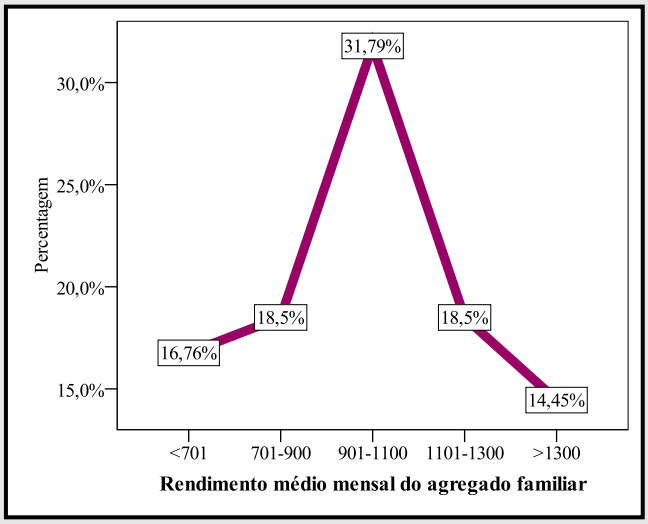


Pela sua análise podemos concluir que mesmo recodificada, esta variável continua a revelar uma significativa discrepância na distribuição da amostra. Os indivíduos menores de 19 anos representam a fatia mais significativa da mesma (36%) enquanto que as duas classes seguintes sofrem diminuições na ordem dos 12 e 16 pontos percentuais revelando valores de 24,7% e 20,5%, respectivamente. Por último, as duas classes finais, *"25-27"* e *">27"*, revelam uma percentagem acumulada de 18,8% representando aqueles cujas idades se encontram entre os 25 e os 39 anos.

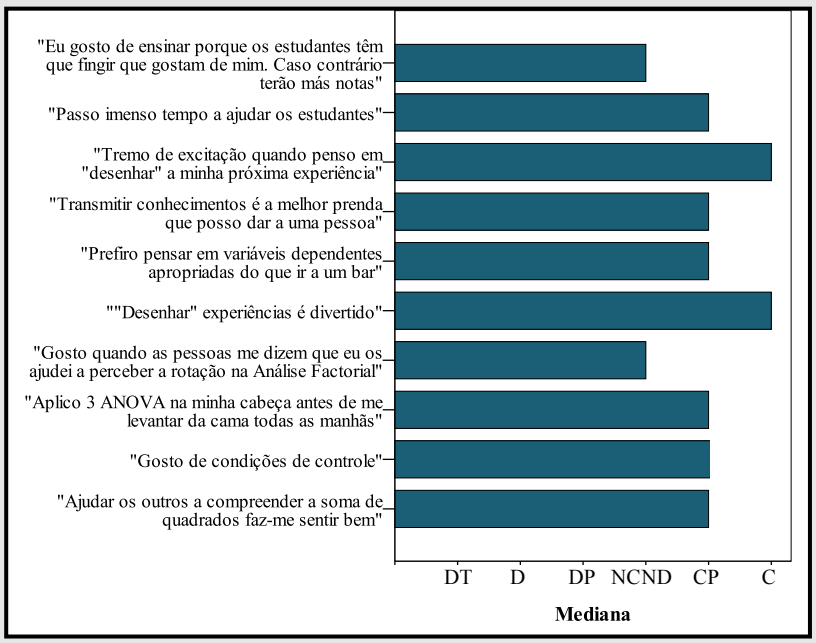
Por último, tendo em conta o *"Rendimento do Agregado Familiar dos Inquiridos"* é importante frisar o registo de 66 não respostas, um valor substancialmente elevado que retrata a diminuta pré-disposição existente para abordar esta questão. Posto isto, a variável original encontrava-se dividida em 7 categorias de resposta sendo que as categorias mais periféricas revelavam-se muito pouco expressivas contendo apenas 5 e 6 casos contabilizados. Desta forma, mais uma vez nos pareceu pertinente recodificá-la para que tomasse uma distribuição mais equilibrada pelo que diminuímos para 5 as categorias de resposta.

Também será esta nova variável de *"Rendimento mensal do agregado familiar"* aquela que utilizaremos ao longo do trabalho. Para o fazer procedemos à fusão das categorias de resposta posicionadas nos extremos: *"menos de 500"* e de *"500 a 700"* no pólo inferior e de *"1301 a 1500"* e *"mais de 1500"* no pólo superior.

Feita a recodificação, o gráfico obtido permite-nos concluir que a maior parte da amostra (31,7%) dispõe de um rendimento médio mensal do agregado familiar entre 901 e 1100 euros. Já a percentagem de inquiridos obtida para as categorias de resposta *"701-900"* e *"1101-1300"* foi idêntica: 18,5%. Analisando os extremos a percentagem daqueles que vivem com menos de 701 euros/mês é superior àqueles que podem contar com mais de 1300 euros/mês: 16,7% e 14,4%, respectivamente.



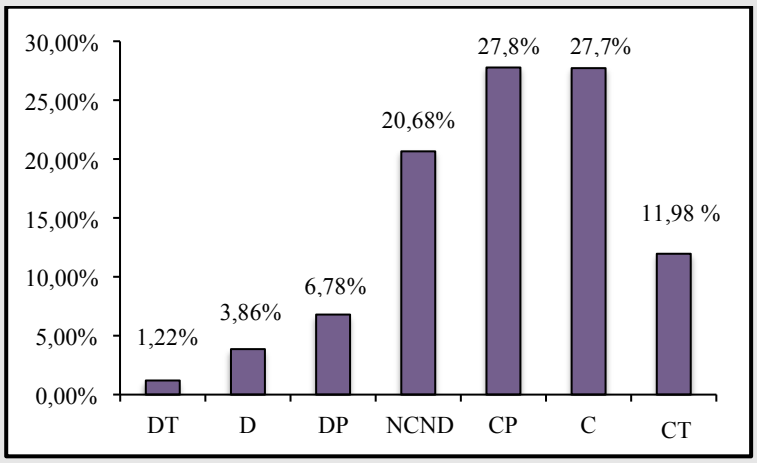
Por último e relativamente às variáveis ordinais que pretendem recolher informação sobre as características essenciais para o desenvolvimento de uma boa metodologia de investigação, estas apresentam 7 categorias de resposta que explicitaremos de seguida: *"Discordo Totalmente" ("DT"); "Discordo" ("D"); "Discordo Parcialmente" ("DP"); "Não Concordo Nem Discordo" ("NCND"); "Concordo Parcialmente" ("DP"); "Concordo" ("C")* e *"Concordo Totalmente" ("CT").*



Como o gráfico demonstra, as variáveis *"Gosto de ensinar porque os estudantes têm de fingir que gostam de mim. Caso contrário têm más notas"* e *"Gosto quando as pessoas me dizem que os ajudei a perceber a rotação na análise factorial"*, são aquelas que registam um valor mediano mais baixo, situado na categoria *"Não Concordo Nem Discordo"*, sendo esta a que reuniu maior percentagem de respostas: 41,8% e 28,9%, respectivamente. Os valores registados para as categorias *"C", "CP", "DP"* e *"D",* revelaram em ambas as variáveis manifestar valores muito próximos na ordem dos 15% para a primeira variável e dos 12% para a segunda. Por fim, as categorias de resposta mais extremas (*"DT"* e *"CT"*) revelaram os valores mais diminutos para ambas as variáveis, apresentando em conjunto um valor médio de 3,56%.

De seguida podemos reparar que 6 variáveis registam a sua mediana na categoria *"CP"*. São elas: "*Passo imenso tempo a ajudar os estudantes"*; "*Transmitir conhecimentos é a melhor prenda que posso dar a uma pessoa"*; *"Prefiro pensar em variáveis dependentes do que ir a um bar*"; *"Aplico 3ANOVA na minha cabeça antes de me levantar todas as manhas", "Gosto de condições de controle"* e "*Ajudar os outros a compreender a soma de quadrados faz-me sentir bem."* Trabalhando-as em conjunto verificamos que as categorias de resposta *"Concordo Parcialmente"* e *"Concordo"* são sempre aquelas que registam valores mais altos sendo que, em média, contabilizam 27,8% e 27,7%, respectivamente, do total da amostra. São responsáveis, desta forma por aproximadamente 55,5% das respostas dos inquiridos. A percentagem de respostas restante encontra-se dividida pelas outras categorias de resposta, verificando-se genericamente pouca incidência nas categorias *"DT"*, *"D"* e *"DP",* e a maior incidência sempre na categoria *"CT"*.

Resultados médios das 5 variáveis em questão



Voltando agora ao gráfico inicial, caso único afirma-se a variável *"Gosto de condições de controle"*. Esta regista a sua mediana entre as categorias *"Concordo Parcialmente"* e *"Concordo"*, que juntas obtêm 55,6% do total das respostas. A categoria *"NCND"* apresenta também uma percentagem significativa (22,1%), seguida da categoria *"CT"* com 13,8%. As opções de resposta negativa (*"DT"*, *"D"* e *"DP"*) reúnem uma percentagem acumulada de apenas aproximadamente 9%.

Finalmente, as variáveis que registam valores medianos mais elevados, na ordem da categoria de resposta *"Concordo"*, são aquelas relacionadas com o desenho de experiências: "*Desenhar experiências é divertido*" e "*Tremo de excitação quando penso em desenhar a minha próxima experiência."* Ambas reflectem a mesma tendência e valores muito próximos para as e categorias de resposta positivas (*"CP"*, *"C"* e *"CT"*) na ordem dos 20%, 35% e 20%, respectivamente. Já as opções de resposta discordantes apresentam valores acumulados também semelhantes: de aproximadamente 10% para as duas variáveis em causa. É a categoria *"NCND"* que manifesta a maior variação entre as duas variáveis registando 11,7% do total das respostas para a primeira e 17,5% para a segunda.

Em suma, e de forma a concluir a caracterização da amostra, podemos afirmar que a amostra não é totalmente homogénea no que diz respeito às variáveis de cariz sócio-demográfico. No entanto, tendo em conta as variáveis ordinais, verifica-se que genericamente a amostra reflecte uma tendência de resposta substancialmente positiva, tendendo a concordar que todas elas são importantes para a criação de um bom método de investigação uma vez que nenhuma das variáveis registou o seu valor médio em categorias de resposta correspondentes à discordância.

Responsável pela Análise: Cathy

Fascinada com o SPSS, Cathy tentou a sua sorte…

Não ia ser fácil, mas Cathy ainda tinha muito tempo para convencer Mr. Pinkley e todo um leque de argumentos para criar.

ANÁLISE FACTORIAL

"A caracterização da amostra estava feita!" – pensou Cathy.

Tinha sido um bom dia de trabalho e os seus objectivos foram cumpridos. Porém, sabia que tinha tratado apenas da análise univariada dos dados e que a parte mais difícil do projecto ainda estava para vir.

Ainda assim, visualizou o seu sucesso.

No dia seguinte, o seu foco era a Análise Factorial e Análise de Fiabilidade. O livro explicava:

Análise Factorial

A análise factorial, enquanto técnica de análise exploratória de dados, tem como objectivo a redução dos dados existentes a partir do agrupamento de variáveis ordinais em factores, baseando-se na existência de correlação entre essas mesmas variáveis.

A utilização e aplicação desta análise permite a descoberta de várias soluções de agrupamento em factores, sendo que o propósito último do analisador deve passar por encontrar a melhor solução possível: aquela que traduza fortes correlações entre as variáveis garantindo, simultaneamente, a melhor explicação possível da variabilidade dos dados e, consequentemente, a menor perda de informação possível.

"Mãos ao trabalho!"

Estavam reunidas todas as condições para que tudo corresse lindamente. Eram 8h da manhã, escritório estava arrumado, o computador estava ligado e a sua vontade era férrea! No entanto, a coisa rapidamente lhe subiu à cabeça…

O dia revelou-se uma fracasso, mas Mr. Pinkley ia ter de perceber, nem sempre as coisas correm logo bem à primeira…E Cathy bem que tinha tentado! As compras tinham sido um pequeno desvio mas no dia seguinte estaria de novo na rota certa: o SPSS era uma ferramenta a dominar!

Análise Factorial

Tendo em conta o conjunto de 10 variáveis ordinais do mesmo tipo incluídas neste inquérito, optámos por escolher como melhor solução para a análise factorial a criação de dois factores, o primeiro com 4 variáveis e o segundo com 3 variáveis que em conjunto explicam 68,5% da variabilidade total dos dados.

Posto isto, é importante focar que decidimos pela exclusão de 3 variáveis, que mostravam valores de extracção diminutos e que, consequentemente, não revelavam uma grande quantidade de variância explicada pelos factores em relação a cada variável individual e implicavam uma elevada perda de informação e fraca explicação da variabilidade.

São elas:

- *Ajudar os outros a compreender a soma de quadrados faz-me sentir bem.* (Extracção de 0,463)
- *Prefiro pensar em variáveis dependentes apropriadas do que ir a um bar.* (Extracção de 0,497)
- *Transmitir conhecimentos é a melhor prenda que posso dar a uma pessoa.* (Extracção de 0,504)

Desta forma, tendo em conta o *trade-off* entre perda de informação e redução de dados, optámos por conservar uma percentagem de explicação da variabilidade total dos dados mais elevada (68,5%) em detrimento de agrupar todas as variáveis em causa e maximizar a redução dos dados.

Apresentando um KMO de 0,769, esta solução reflecte uma forte correlação entre as variáveis e a significância obtida para o Teste de Bartlett apresentou um valor de 0,000[1] o que nos leva a rejeitar a hipótese nula: "*A matriz de correlação é igual à matriz de identidade*" e a prosseguir com a análise dos factores obtidos que se apresentam descriminados de seguida:

[1] Estatística de teste: 673,842

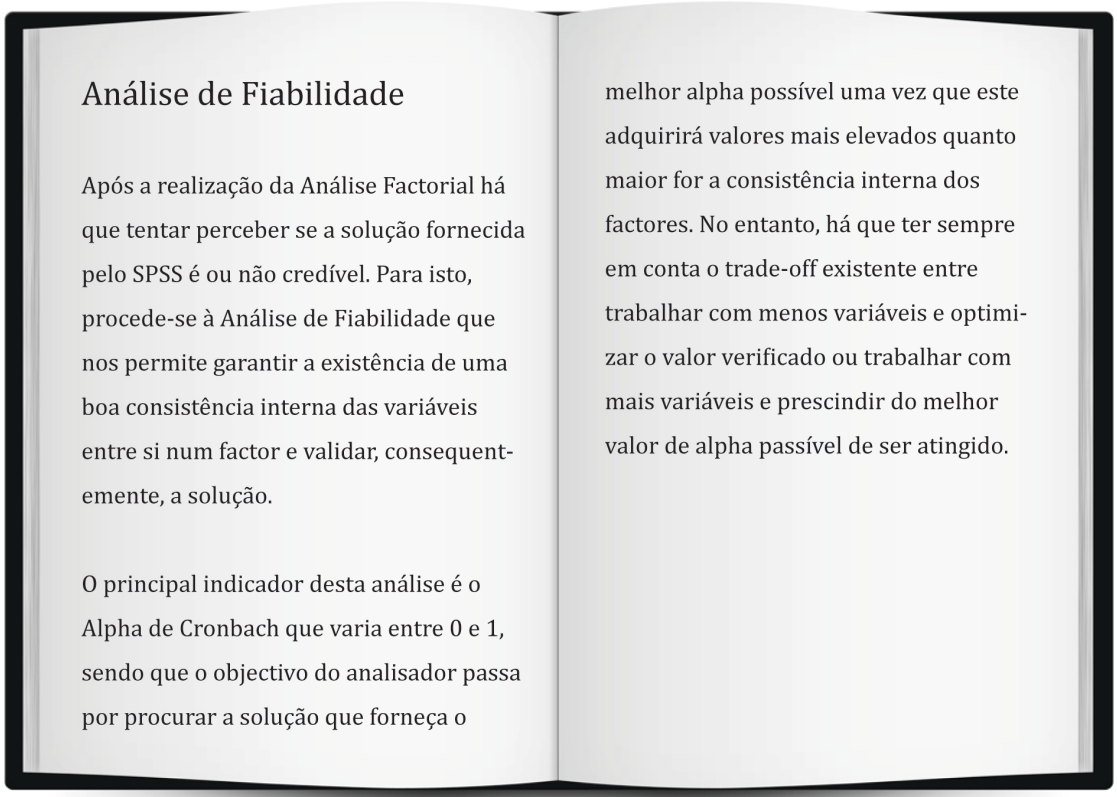
Factor 1

- *Gosto de condições de controle.*
- *Aplico 3 ANOVA na minha cabeça antes de me levantar da cama todas as manhãs.*
- *Desenhar experiências é divertido.*
- *Tremo de excitação quando penso em "desenhar" a minha próxima experiência.*

Factor 2

- *Gosto quando as pessoas me dizem que eu os ajudei a perceber a rotação na Análise Factorial.*
- *Passo imenso tempo a ajudar os estudantes.*
- *Eu gosto de ensinar porque os estudantes têm que fingir que gostam de mim. Caso contrário terão más notas.*

Responsável pela Análise: Cathy



Análise de Fiabilidade

Após a realização da Análise Factorial há que tentar perceber se a solução fornecida pelo SPSS é ou não credível. Para isto, procede-se à Análise de Fiabilidade que nos permite garantir a existência de uma boa consistência interna das variáveis entre si num factor e validar, consequentemente, a solução.

O principal indicador desta análise é o Alpha de Cronbach que varia entre 0 e 1, sendo que o objectivo do analisador passa por procurar a solução que forneça o melhor alpha possível uma vez que este adquirirá valores mais elevados quanto maior for a consistência interna dos factores. No entanto, há que ter sempre em conta o trade-off existente entre trabalhar com menos variáveis e optimizar o valor verificado ou trabalhar com mais variáveis e prescindir do melhor valor de alpha passível de ser atingido.

Product Testing, Inc

Relatório *University of Sussex*

Análise de Fiabilidade

Com o objectivo de validarmos a opção escolhida, temos de proceder então a uma análise de fiabilidade da mesma que comprove a boa consistência interna entre as variáveis inseridas num factor.

Sendo assim, o Alpha de Cronbach obtido para o primeiro factor registou um valor bastante perto de 1 (0,848) o que corrobora a sua boa consistência interna. Para o segundo factor o valor registado foi de 0,741, o que apesar de não tão elevado, ainda revela uma boa consistência das variáveis no factor. Adicionalmente, para nenhuma das análises se verifica um aumento do Alpha de Cronbach aquando da exclusão de alguma variável inserida no factor.

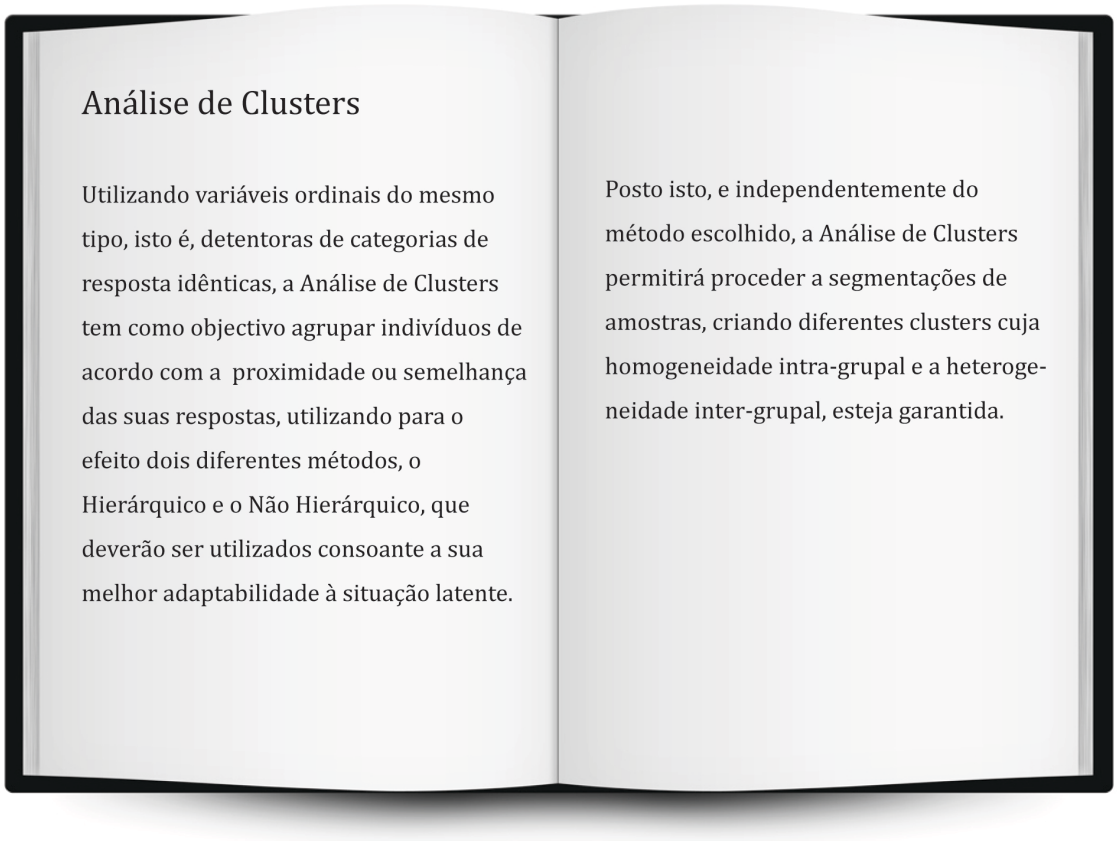
Concluindo, de entre todas as soluções encontradas esta é, na nossa opinião, a melhor solução possível pois conjuga uma boa explicação da variabilidade total dos dados corroborada por uma boa consistência interna das variáveis nos factores. Acreditamos que os dois factores criados permitirão trabalhar os dados de uma forma mais reduzida, garantindo uma perda mínima de informação.

Responsável pela Análise: Cathy

ANÁLISE DE CLUSTERS

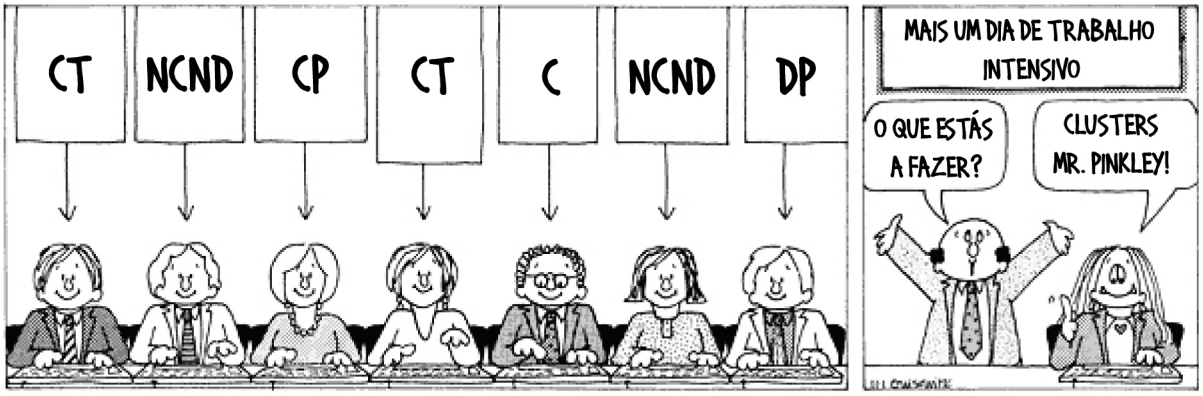
Mr. Pinkley andava ansioso e insistente com o projecto de *Sussex*. Pressionava constantemente Cathy tentando perceber se estava a progredir e a fazer um bom trabalho. Afinal, não conhecendo o novo *software* SPSS, tinha de depositar toda a sua confiança nela e nas suas capacidades.

Cathy, por sua vez, esforçava-se por responder às expectativas e estava a aprender bastante sobre metodologias de trabalho e ferramentas que a poderiam ajudar a fazer mais e melhor. O dia de hoje destinava-se à Análise de Clusters.



Sentindo-se numa maré de sorte, Cathy optou pelo método hierárquico e dedicou-se a analisar o ficheiro. De facto era bastante interessante agrupar os indivíduos consoante as suas respostas e verificar que tendências existiam. Para além disso, era uma técnica que se podia usar para os mais variados temas e que permitia desvendar aspectos bastante curiosos.

Sempre com Mr. Pinkley a rondar, Cathy trabalhava…



MÉTODO HIERÁRQUICO

Método Hierárquico

Este método adequa-se principalmente a amostras com uma dimensão de aproximadamente 100 indivíduos ou inferior e tem como principal particularidade o facto de, uma vez inseridos em determinado cluster, a posição dos indivíduos não sofrer mais alterações tornando-se, assim, imutável.

Ainda inseridos na aplicação desta técnica encontram-se vários métodos que podem ser utilizados de forma a identificar a melhor solução em termos de dimensão dos clusters. Quando encontrada aquela que, intuitivamente, se afirma como a melhor solução possível, há que validá-la realmente através de Testes de Hipóteses. Neste caso devem ser aplicados os testes não paramétricos de Mann-Whitney ou de Kruskal-Wallis, consoante o número de clusters propostos. Estes permitirão verificar a diferença nas distribuições das variáveis e, consequentemente, garantir a heterogeneidade entre os clusters. A solução encontra-se, assim, validada.

Product Testing, Inc

Relatório *University of Sussex*

Análise de Clusters: Método Hierárquico

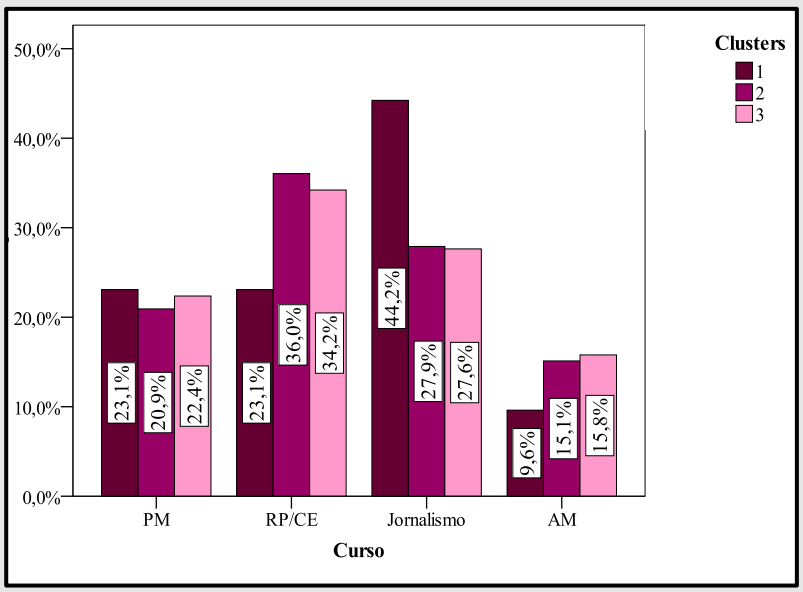
Relativamente à Análise de Clusters pelo Método Hierárquico, e tendo em conta todas as variáveis ordinais de cariz de concordância, a solução mais equilibrada a nível de distribuição dos indivíduos pelos vários grupos segue o Método *Within-Groups Linkage* e, registando-se apenas 1 não resposta, engloba 3 diferentes clusters com a seguinte distribuição:

- Cluster 1: 55 pessoas, representando 23,1% do total da amostra.
- Cluster 2: 97 pessoas, representando 40,8% do total da amostra.
- Cluster 3: 86 pessoas, representando 36,1% do total da amostra.

No sentido de validar a solução obtida, e tendo em conta que esta inclui 3 grupos distintos, é necessário proceder à realização do teste não paramétrico de Kruskal-Wallis cuja hipótese nula afirma: *"As distribuições são iguais nos 3 clusters."* Tendo obtido uma significância de 0,000 para todas as variáveis incluídas podemos rejeitar esta hipótese e, consequentemente, garantir a heterogeneidade entre os grupos e validar a opção apresentada.

De forma a possuirmos uma ideia mais clara de qual a composição efectiva dos diferentes clusters, podemos caracterizá-los em relação a outras diferentes variáveis. Neste sentido, optámos por cruzar a solução obtida com as variáveis *"Curso"* e *"Rendimento"* e analisar os resultados obtidos.

Clusters & Curso



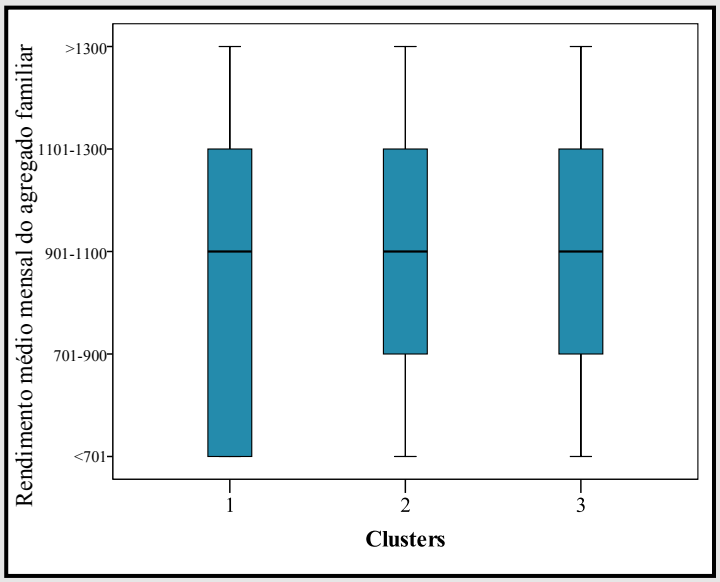
Analisando o gráfico respeitante a esta amostra, podemos retirar algumas conclusões interessantes. Em primeiro lugar é bastante visível que o curso de *"Audiovisual e Multimédia"* é aquele que engloba menos indivíduos para cada um dos clusters. Posto isto, é também visível que no primeiro cluster predominam essencialmente alunos de *"Jornalismo"* (44,2%), sendo que os cursos de *"Publicidade e Marketing"* e de *"Relações Públicas e Comunicação Empresarial"* registam exactamente a mesma percentagem de alunos (23,1%).

Já no segundo cluster, 36% dos indivíduos pertence ao curso de *"Relações Públicas e Comunicação Empresarial"*, seguindo-se *"Jornalismo"*, *"Publicidade e Marketing"* e *"Audiovisual e Multimédia"* com 27,9%, 20,9% e 15,1%, respectivamente.

Continuando a análise, verifica-se que o terceiro cluster também regista a sua predominância no curso de *"Relações Públicas e Comunicação Empresarial"*, com 34,2% dos seus constituintes. Regista ainda 27,6% dos mesmos no curso de *"Jornalismo"* e 22,4% no curso de *"Publicidade e Marketing"*. É neste cluster que a percentagem de alunos de *"Audiovisual e Multimédia"* apresenta o seu valor mais alto, se bem que ainda diminuto, de 15,8%. Um aspecto curioso retirado do cruzamento dos clusters obtidos com esta variável nominal é o facto de *"Publicidade e Marketing"* ser o curso mais equitativamente presente nos três clusters registando sempre valores na ordem dos 20%.

Clusters & Rendimento

Antes de aprofundar a análise, é necessário focar que neste cruzamento, registam-se 67 não respostas, o que diminui bastante a dimensão dos clusters. No entanto, optando por analisar os valores de percentagem podemos prosseguir a uma análise elucidativa.



Rendimento mensal do agregado familiar	Cluster 1	Cluster 2	Cluster 3
< 701	30%	14,5%	11,1%
701-900	10%	14,5%	27%
901-1100	27,5%	39,1%	27%
1101-1300	15%	20,3%	19%
> 1300	17,5%	11,6%	15,9%

Em primeiro lugar, tendo por base uma tabela de contigência e um *boxplot*, podemos verificar que o valor da mediana é igual para os três clusters, posicionando-se na terceira classe da variável *"Rendimento"* (*"901-1100"* euros/mês) onde estes registam percentagens de resposta bastante elevadas: 27,5%, 39,1% e 27%, respectivamente. Para além disto, verificamos que também o valor do terceiro quartil (75%) é o mesmo para os três grupos concentrando-se na quarta classe de rendimento (*"1101-1300"* euros/mês).

O segundo e terceiro cluster revelam-se assim idênticos em termos de dispersão de quartis sendo que o primeiro cluster é aquele que difere, revelando uma maior amplitude de respostas, já que o primeiro quartil é, ao mesmo tempo, a primeira classe de respostas existente, correspondendo aos indivíduos cujo rendimento médio mensal do agregado familiar não ultrapassa os 700 euros (30% do total de membros do cluster).

Analisada esta amostra podemos então concluir que os segundo e terceiro clusters são compostos por uma maior percentagem de indivíduos com mais elevado rendimento médio mensal do agregado familiar, relativamente ao primeiro cluster, onde a maior percentagem obtida recai sobre a classe de rendimento mais diminuto.

Responsável pela Análise: Cathy

O dia acabou e Cathy voltou para casa descansada e orgulhosa. O seu projecto estava finalmente a ganhar forma e consistência e as análises que tinha feito foram tanto proveitosas como interessantes. A análise de clusters estava organizada e acabada e Cathy sabia que tinha optado pelo método mais adequado: o cliente ia gostar!

Estava a planear um jantar descansado com Irving, o seu namorado, e quiçá relaxar e ver um filme romântico. Enquanto escolhia o filme, Irving chegou e com ele…as dúvidas!

Mais tarde, já deitada, a cabeça de Cathy andava às voltas. Não conseguia dormir. Era verdade que a amostra que estava a trabalhar tinha mais de 100 indivíduos mas o livro dizia especificamente que o Método Hierárquico podia ser aplicado para amostras maiores se o *software* conseguisse processar a informação adequadamente. E o SPSS tinha-o feito!

Mesmo assim, Irving tinha acabado com os seus projectos de descanso…

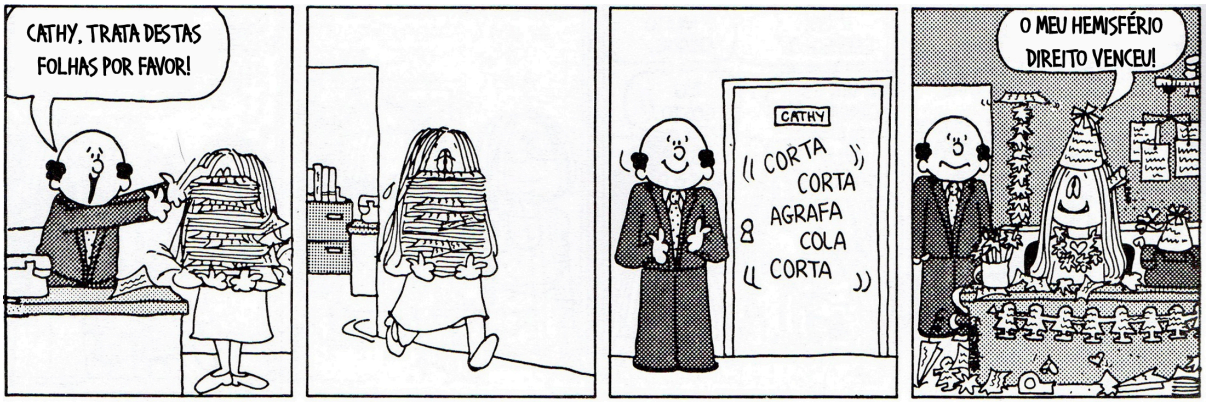
Era inevitável. Depois de quase duas horas a matutar, Cathy levantou-se e pôs mãos à obra. Ia verificar que todas as relações que tinham feito faziam sentido e que o método que usou estava bem aplicado. Assim, se alguém voltasse a duvidar da sua escolha, poderia defendê-la convictamente!

No dia seguinte, a primeira coisa a fazer era conseguir a aprovação de Mr. Pinkley e passar, descansadamente, para a próxima tarefa…

E com duas simples frases Mr. Pinkley conseguiu fazer desabar o mundo de Cathy. Todo o seu trabalho, todo o seu esforço, toda a sua análise…por água abaixo!

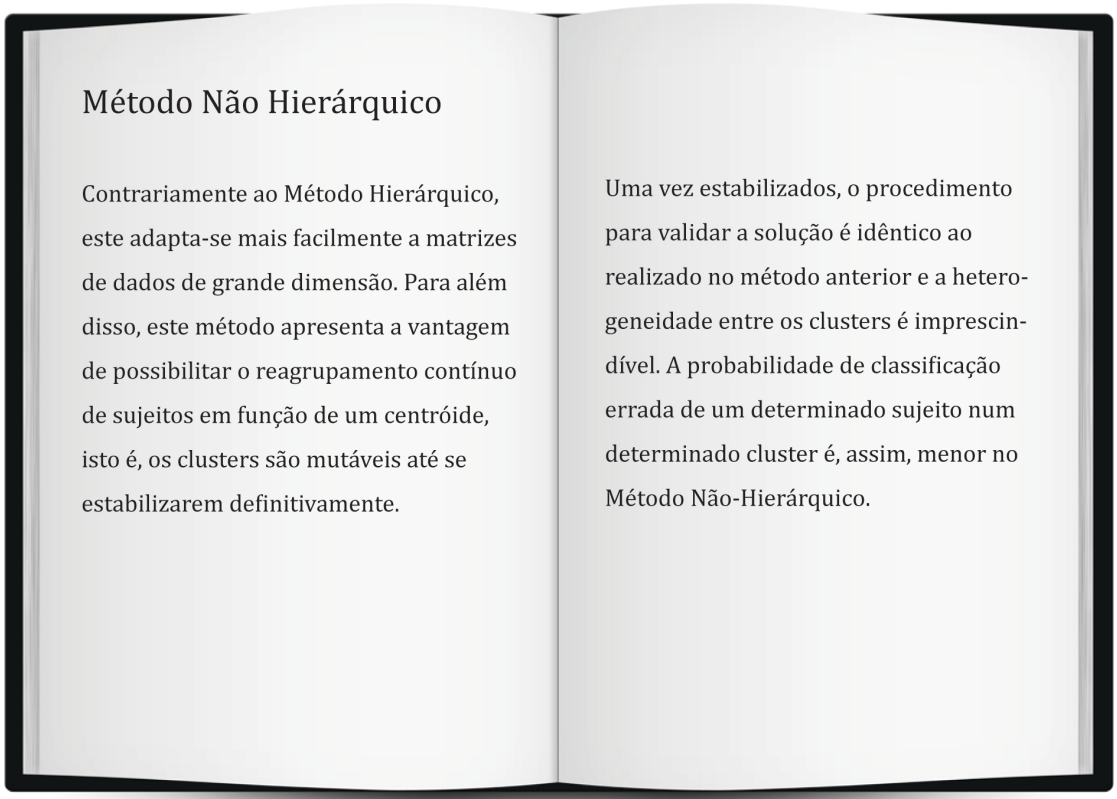
Mas Cathy não ia deixar que tudo fosse em vão. Apresentaria os dois métodos e a análise ficaria ainda mais completa. Havia sempre um lado positivo.

Ainda nesse dia Mr. Pinkley chamou várias vezes Cathy ao seu escritório. Também ele estava rodeado de folhas e, não tendo ainda descoberto as vantagens do SPSS, espalhava a confusão por toda a empresa. Porém, o dia de Cathy estava a ser desastroso e a sua produtividade adquiriu todo um novo carácter….



Chegada a casa, Cathy estava decidida. Ia desfrutar de uma boa noite de sono e acordar no dia seguinte com as energias restauradas. Talvez precisasse do seu chá *Noite Tranquila*, mas faria o seu trabalho, e bem feito!

MÉTODO NÃO HIERÁRQUICO



Product Testing, Inc

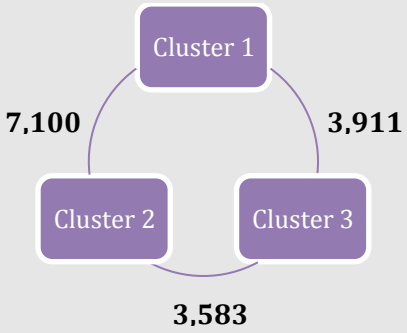
Relatório *University of Sussex*

Análise de Clusters: Método Não Hierárquico

Refazendo a Análise de Clusters, desta vez aplicando o Método Não Hierárquico às mesmas variáveis utilizadas anteriormente, a melhor solução encontrada volta a incluir três diferentes clusters com as seguintes composições:

- Cluster 1: 56 pessoas, representando 23,5% do total da amostra.
- Cluster 2: 73 pessoas, representando 30,7% do total da amostra.
- Cluster 3: 109 pessoas, representando 45,8% do total da amostra.

As distâncias entre os centróides de cada cluster vêm representadas seguidamente:

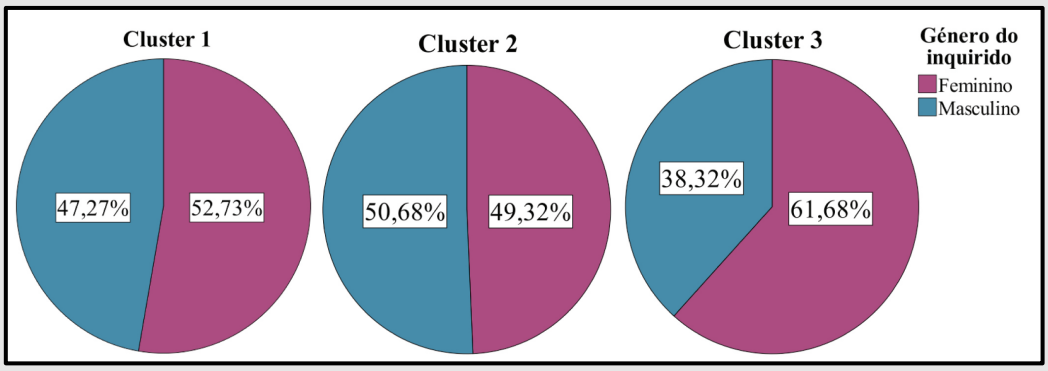


Decidimos pelo não aumento do número de clusters uma vez que isso acarretaria uma significativa diminuição da distância entre os mesmos, o que seria prejudicial e contraproducente uma vez que o objectivo passa por garantir a heterogeneidade dos clusters a nível inter-grupal e a sua homogeneidade a nível intra-grupal.

Todavia, é ainda imprescindível validar formalmente a opção obtida, novamente através do teste não paramétrico de Kruskal-Wallis. Explicitadas anteriormente as hipóteses inerentes ao teste, voltamos a obter uma significância de 0,000 para todas as variáveis o que nos permite rejeitar H_0 e garantir a heterogeneidade entre os clusters, procedendo-se à validação final da solução obtida.

Também para este método é interessante perceber quem são os constituintes de cada cluster e, de forma a caracterizá-los mais concretamente, voltámos a optar pelo seu cruzamento com duas outras variáveis, desta vez o *"Género do Inquirido"* e a *"Idade do Inquirido"*.

Clusters & Género do Inquirido

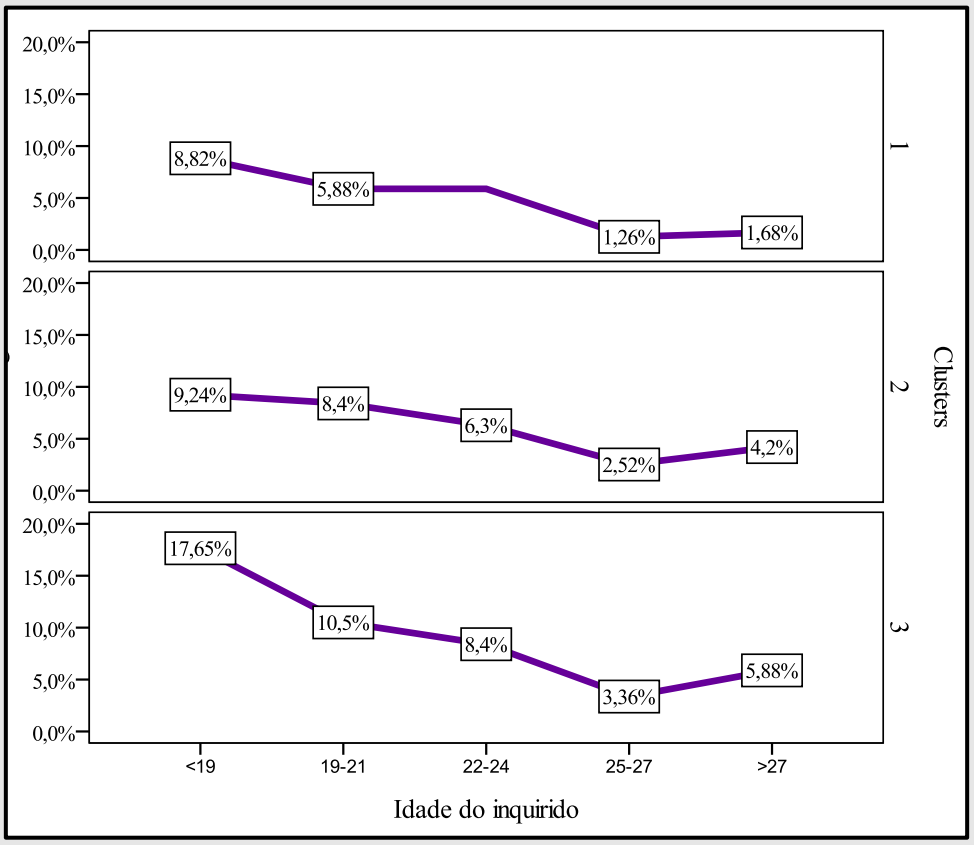


Pela análise do conjunto de gráficos circulares podemos perceber que o género *"Feminino"* predomina no primeiro e terceiro cluster. Esta situação é principalmente acentuada no terceiro cluster onde as mulheres têm maioria absoluta, representando 61,7% do total da amostra estudada, contra apenas 38,3% de homens.

Esta situação acaba por não ser de estranhar uma vez que na caracterização da amostra já se tinha observado uma maior percentagem de mulheres do que de homens[2], o que, inevitavelmente, se iria reflectir aquando do cruzamento desta variável com outras.

Ainda assim, o segundo cluster consegue contrariar esta tendência manifestando-se como o mais equilibrado em termos de distribuição segundo a variável *"Género do Inquirido"* e obtendo mesmo mais indivíduos do género *"Masculino"* (50,6%) do que do sexo *"Feminino"* (49,3%).

Clusters & Idade do Inquirido



Em primeiro lugar, verificamos que o terceiro cluster é aquele que apresenta percentagens mais elevadas para todas as classes etárias, o que vem ao encontro do facto de ser o cluster mais

[2] Relembrando: 55,9% de mulheres vs. 44,1% de homens

populado dos três. O mesmo se passa relativamente ao segundo cluster que, tendo também mais indivíduos do que o primeiro, apresenta valores um pouco mais elevados que este para cada uma das classes.

Tendo isto em conta, acreditamos que as percentagens entre os clusters diferem proporcionalmente não havendo essencialmente grandes discrepâncias. Regista-se fundamentalmente uma maior predominância, em todos os clusters, de indivíduos *"menores de 19 anos"*, seguidos de indivíduos dos *"19 aos 21"* anos.

		Idade do Inquirido		
		Cluster 1	Cluster 2	Cluster 3
N	Respostas Válidas	56	73	109
	Não Respostas	0	0	0
Mediana		2,00	2,00	2,00
Percentis	25	1,00	1,00	1,00
	50	2,00	2,00	2,00
	75	3,00	3,00	3,00

A análise da tabela acima permite-nos confirmar as nossas conclusões prévias uma vez que é notório que tanto a mediana como os vários percentis registam para os três grupos exactamente os mesmos valores, o que nos confirma que os clusters se manifestam particularmente homogéneos no que respeita ao cruzamento com a variável *"Idade do Inquirido"*. A mediana encontra-se na classe etária dos *"19 aos 21"* anos, ao passo que o terceiro quartil não vai para além da terceira categoria de resposta disponível: dos *"22 aos 24"* anos. Estamos assim perante uma amostra com idades compreendidas essencialmente entre os 17 e 24 anos, sendo esporádicos os casos de indivíduos com idades superiores a 25 anos.

Responsável pela Análise: Cathy

REGRESSÃO LOGÍSTICA BINÁRIA

Concluída a Análise de Clusters através dos dois métodos, o entusiasmo de Cathy era grande. De facto, o SPSS abrira-lhe horizontes no que diz respeito à utilização de ferramentas de trabalho mais actualizadas que lhe permitissem simultaneamente poupar tempo e conseguir análises mais rigorosas, detalhadas e ricas para o cliente.

Faria como sua principal missão na Product Test, Inc. a actualização dos s*oftwares* e a adopção de uma metodologia de trabalho cada vez mais moderna e eficiente! É claro que o seu empreendorismo acabou por se dispersar um bocadinho…

Apesar da relutância de Mr. Pinkley em aceder aos seus pedidos, Cathy estava decidida a provar de todas as maneiras que os seus argumentos forneceriam à empresa uma nova vantagem competitiva. Para isso, acabar o trabalho a tempo e horas e impressionar o cliente era fundamental…Regressão Logística Binária, aqui vamos nós!

Regressão Logística Binária

Utiliza-se a Regressão Logística quando se pretende estudar uma amostra com o objectivo de fazer previsões probabilísticas para a população. Para isso procede-se à constituição de um modelo que explique a influência de um conjunto de variáveis independentes, de qualquer tipo, sobre uma variável dependente de carácter nominal binário (ou tratada como tal) para a qual se define, a priori, uma categoria de interesse. O objectivo primordial é verificar a probabilidade da variável dependente tomar essa mesma categoria de interesse.

De forma a obtermos um modelo válido, há que primeiramente confirmar a influência das variáveis independentes sobre a dependente e através do Teste de *Omnibus* declarar o modelo estatisticamente significativo, ou seja, adequado para realizar previsões. Posteriormente, o Teste de *Hosmer & Lemeshow* confirmará o ajustamento do modelo aos dados. Quanto maior o ajustamento maior a aproximação dos casos observados relativamente aos casos previstos.

A análise do modelo relativo à amostra permitirá a inferência de conclusões para a população. No entanto, cabe reforçar que nesta técnica a relação entre as variáveis é indirecta, obtendo-se apenas probabilidades

Os seus primeiros cruzamentos não foram muito animadores mas Cathy já estava habituada. O SPSS gostava de lhe dar alguma luta antes de lhe conceder respostas, era quase uma relação de amor-ódio. Não desistindo, Cathy aproveitava todas as oportunidades para aprender e dominar todos as pequenas subtilezas inerentes a esta técnica, de forma a esgotar as suas hipóteses e garantir que os resultados apresentados estavam baseados em rigoroso e exaustivo tratamento dos dados.

Contudo nem sempre isto era fácil e em alturas de fraqueza, o pessimismo apoderava-se de si…

Porém, Cathy resistiu e insistiu….

Product Testing, Inc

Relatório *University of Sussex*

Regressão Logística Binária

Em primeiro lugar, pareceu-nos muito interessante tentar estabelecer relações de influência entre as variáveis independentes ordinais de cariz de concordância com as variáveis de tipo sócio-demográfico *"Género"*, *"Curso"*, *"Rendimento"* e *"Idade"*, que tomámos como dependentes. Para proceder à análise, recodificámos estas variáveis de forma a torná-las variáveis nominais binárias, determinando sempre uma categoria de interesse. Posto isto, e cruzando-as, individualmente, com todas as variáveis independentes supra mencionadas, passámos à interpretação dos resultados.

Desafortunadamente, estes não revelaram nenhum modelo que permitisse fazer previsões conclusivas, não se tendo identificado nenhuma variável independente que exercesse influência sobre qualquer uma das dependentes determinando uma probabilidade para que esta última tomasse a categoria de interesse pré-definida.

Ainda assim, procedemos à nova recodificação das categorias de interesse das várias variáveis tomadas como dependentes de maneira a verificar se os resultados se mostravam ou não mais favoráveis. Simultaneamente experimentámos para cada uma as opções todos os métodos existentes. Todavia, as conclusões voltaram a não ir ao encontro das nossas expectativas.

Foi neste clima de frustração que Mr. Pinkley decidiu aparecer….

Já mais calma, Cathy continuou…

Iria ter de fazer ver ao cliente que a não existência de influência era uma informação tão válida quanto a existência de influência e iria deixar bem claro que tinha feito todos os esforços para conseguir obter um modelo que permitisse fazer previsões. Se não tinha encontrado, era porque não existia.

Product Testing, Inc

Relatório *University of Sussex*

Regressão Logística Binária[3]

Por último, decidimos ainda seleccionar várias de entre as variáveis ordinais de cariz de concordância e defini-las como variável dependente, tratando-as como nominal binárias e verificando se sofriam ou não influência por parte das restantes variáveis.

Mostrando-se esta opção igualmente insatisfatória, focámo-nos na variável *"Rendimento mensal do agregado familiar"*, tomando como dependente e definindo como categoria de interesse as *"Não Respostas"*, que incluíam 66 casos. Cruzando-as com as restantes variáveis independentes, a análise revelou-se, à partida, animadora:

Através do Método *Enter* a significância obtida para a tabela relativa às variáveis na equação registava um valor de 0,000. Assim sendo, rejeita-se $H_{0,}$ reiterando que $B_0 \neq 0$ e continua-se a análise. Olhando para o Teste de Omnibus, verificámos também que a significância volta a tomar o valor de 0,000, o que nos leva a rejeitar $H_{0,}$ e a concluir que *"O modelo é estatisticamente significativo."* Já o Teste de Hosmer & Lemeshow apresenta uma significância de 0,869 o que nos permite aferir que *"O modelo se ajusta aos dados"*, uma vez que não rejeitámos a Hipótese Nula inerente a este teste. Por último, através da análise da Tabela de Classificação *a,b* em comparação com a Tabela de Classificação *a*, percebemos que a percentagem de classificação global correcta aumenta substancialmente, de 73,6% para 84,9%, o que reflecte que compensou proceder à introdução das variáveis independentes.

[3] Continuação da análise realizada na página anterior.

É neste cenário aparentemente interessante que procedemos à análise das significâncias associadas às variáveis na equação e é nesta análise que o nosso modelo se reflecte insatisfatório. O teste de hipóteses inerente a esta significância tem as seguintes hipóteses:

- H_0: Os β são iguais a 0.
- H_1: Os β são diferentes de 0.

Na verdade, nenhum dos testes subjacentes aos coeficientes dos termos que contêm as variáveis independentes revela significâncias que possibilitem a rejeição da hipótese nula para todas as categorias de resposta, sendo que a variável independente *"Transmitir conhecimentos é a melhor prenda que posso dar a uma pessoa*" é a que mais se aproxima de permitir esta rejeição, apresentando significâncias menores de 0,05 em 4 das suas categorias de resposta. Isto não se verifica, porém, para duas outras categorias de resposta o que acaba por invalidar todo o modelo.

Concluímos, assim, que nenhuma variável independente apresenta probabilidades de influenciar a variável dependente "*Rendimento Médio Mensal do Agregado Familiar"* para que esta tome a categoria de interesse: as *"Não Respostas"*. O mesmo se passa para todos os restantes cenários considerados.

Responsável pela Análise: Cathy

ANÁLISE DE CORRESPONDÊNCIAS

O último passo do projecto de Cathy era proceder à Análise de Correspondências. Depois de não ter conseguido obter um bom modelo para a Regressão Logística que permitisse a realização de previsões, Cathy sentia que tinha falhado e que não iria satisfazer totalmente o cliente. Sair da cama apresentava-se uma tarefa difícil...

Quando finalmente venceu a batalha e chegou ao escritório Mr. Pinkley estava desvairado…

Apesar da moleza matutina, hoje o seu estado de espírito era positivo. Sentia-se calma. O projecto estava quase terminado e apesar de alguns altos e baixos, tinha aprendido muita coisa e acreditava ter feito um bom trabalho. Depois de um café e um chocolate, para adoçar o dia, pôs-se ao trabalho.

Análise de Correspondências

Utilizada maioritariamente com variáveis nominais e ocasionalmente com variáveis ordinais, a Análise de Correspondências tem como objectivo agrupar categorias de variáveis que se afirmem como dependentes.

O primeiro passo para proceder a esta análise é confirmar, de facto, a dependência das variáveis, o que é feito através de um teste de Qui-Quadrado. Seguidamente há que verificar o número máximo de dimensões existentes na solução e verificar quais as dimensões que explicam consideravelmente a variabilidade dos dados (um valor superior àquele que seria explicado se a variabilidade fosse distribuída igualmente por todas as dimensões). Por fim, o gráfico correspondente a esta análise permite identificar o caso de correspondências mais acentuado e, por vezes, denominar as dimensões nele retratadas. Determinadas as relações entre categorias das variáveis existentes na amostra, cabe ao analisador transpô-las para a sua população realizando inferências úteis e elucidativas.

Cathy encontrou o seu primeiro desafio logo nas condições de aplicabilidade do teste do Qui-Quadrado. Todas as combinações de variáveis que tinha experimentado não as tinham respeitado, o que lhe impossibilitava o prosseguimento da Análise. Enquanto meditava sobre como resolver esta situação, decidiu ir jantar a casa com Irving, talvez conversar e sair do escritório lhe desse algumas ideias brilhantes...

Product Testing, Inc

Relatório *University of Sussex*

Análise de Correspondências

Para procedermos à Análise de Correspondências decidimos proceder à recodificação de duas variáveis que nos pareceram interessantes:

- *"Eu gosto de ensinar porque os estudantes têm que fingir que gostam de mim. Caso contrário terão más notas" e,*
- *"Gosto quando as pessoas me dizem que eu os ajudei a perceber a rotação na Análise Factorial".*

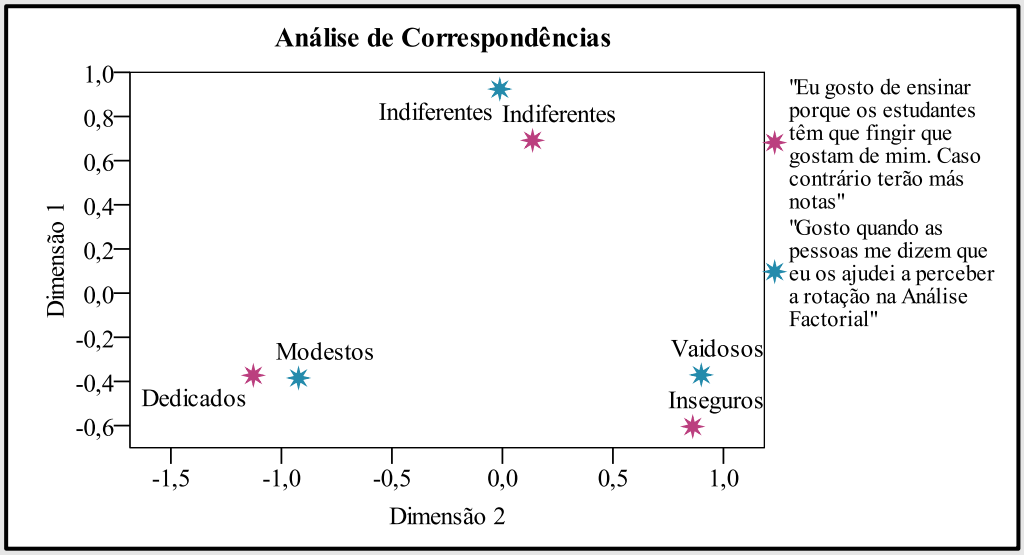
Em relação à primeira pareceu-nos que as pessoas que tenderiam a concordar com a afirmação seriam potencialmente *"Inseguros"* uma vez que ensinar seria apenas um meio para atingir um outro fim: ser apreciado pelos alunos. Por oposição, aqueles que manifestassem qualquer tipo de discordância revelavam um gosto verdadeiro pelo ensino, razão pela qual os chamámos de *"Dedicados"*.

Já respeitante à segunda variável, a nossa percepção passou por depreender que aqueles que revelavam concordar com a afirmação denotavam uma clara necessidade de reconhecimento, passando a tomar o nome de *"Vaidosos"*. Contrariamente, aqueles que discordavam da mesma revelavam modéstia, adjectivo de onde nasceu o seu nome: *"Modestos"*. Cabe ainda mencionar que em ambas as variáveis, os inquiridos que escolheram a categoria de resposta *"NCND"*, foram rotulados de *"Indiferentes"*. Intuitivamente, o nosso raciocínio levou-nos a considerar que os *"Modestos"* revelariam ser *"Dedicados"*, bem como os *"Vaidosos"* seriam também *"Inseguros"*. O nosso propósito é então verificar, através da Análise de Correspondências, se esta situação realmente se verifica.

Em primeiro lugar, procedendo ao Teste de Qui-Quadrado verificamos a confirmação das condições de aplicabilidade e registamos uma significância de 0,000 o que nos leva a rejeitar H_0: *"As variáveis são independentes"* e a concluir pela sua dependência. Estão assim reunidas as condições para prosseguirmos a análise.

Tendo estabelecido que possuímos no máximo 2 dimensões e que cada uma explicaria 0,50 da variabilidade dos dados caso esta fosse dividida igualmente por ambas, verificamos pelo *summary* que apenas uma dimensão explica mais do que este valor pré-estabelecido: 0,74.

A análise do gráfico correspondente vem ilustrar e corroborar a nossa interpretação intuitiva das variáveis.



Podemos perceber que é a dimensão 1 aquela que revela explicar uma maior variabilidade dos dados ocupando uma escala significativamente maior do que a amplitude da escala necessária para a dimensão 2.

Analisada esta amostra, verificámos então uma tendência para aqueles que concordam com a primeira variável concordarem também com a segunda demonstrando serem simultaneamente *"Vaidosos"* e *"Inseguros"*, sendo esta a correspondência mais gritante. Por outro lado, o contrário também se verifica: aqueles que discordam da primeira variável tendem também a discordar da segunda afirmando-se como *"Modestos"* e *"Dedicados"*. Transpondo para a população podemos inferir que geralmente os pares de características mencionadas caminham de mãos dadas, revelando-se simultaneamente num mesmo indivíduo.

Responsável pela Análise: Cathy

CONCLUSÃO

O projecto de Cathy estava terminado. Fora uma longa, penosa, mas gratificante jornada e apenas faltava agora enviá-lo ao cliente e esperar que este o achasse interessante, inovador e útil, tirando partido das conclusões apresentadas. Cathy acreditava ter cumprido os objectivos a que se tinha proposto. Aplicando cada uma das técnicas ao ficheiro, tinha conseguido ilustrar as suas mais valias e de que forma permitiriam ao cliente conhecer mais a fundo a amostra e tomar decisões mais informadas, com base em estudos estatísticos e dados concretos.

Este poderia agora, através da leitura do seu trabalho, compreender qual era a sua amostra, quais as suas principais características sócio-demográficas e quais as suas tendências de resposta para as variáveis relativas ao estudo de como desenvolver uma boa metodologia de trabalho.

Poderia ainda verificar de que forma se agruparam as variáveis, criando factores, com o intuito de reduzir os dados e de que forma se garantiu a consistência interna desses mesmos factores. Adicionalmente, o cliente poderia verificar como as respostas dos indivíduos variavam, aproximando-se ou distanciando-se umas das outras e possibilitando a criação de clusters, que posteriormente poderiam ser cruzados com outras variáveis de forma a ficar explícito quais as diferenças entre cada um deles.

Relativamente à Regressão Logística, a Universidade de Sussex apenas obteria a explicação teórica da mesma, não havendo nenhum exemplo de aplicação prática. Porém, depois de tanto esmiuçar o ficheiro, Cathy estava agora de consciência tranquila, tendo ainda a certeza de que ficara a dominar esta técnica e que mais oportunidades viriam de a aplicar convenientemente.

Por fim, Cathy ficara verdadeiramente fã da Análise de Correspondências. O mundo de ligações que esta permitiria era gigantesco e na sua cabeça Cathy aplicava-a vezes sem conta. Provavelmente a preguiça de Irving em arrumar o que quer que fosse estaria relacionada com o facto de ter sido tão mimado enquanto pequeno. E a resistência de Mr. Pinkley em aceitar a modernidade corresponderia à sua concordância com uma educação tradicional e antiquada. Estas hipóteses eram agora facilmente passíveis de verificação. Cathy prepararia algumas variáveis nominais e verificaria então se as suas teorias se confirmavam. Era isto que Cathy mais apreciava. Nada melhor do que números, gráficos e tabelas para sustentar uma argumentação. A Inferência Estatística não era pêra doce mas o SPSS era, oficialmente, a sua arma secreta!

Enquanto grupo, somos da mesma opinião de Cathy. A Inferência Estatística pode ser, por vezes, tão desafiante quanto frustrante e apela bastante à nossa teimosia, bem como ao nosso espírito perseverante de não desistir perante a adversidade. Aprendemos muito com este trabalho e usar o mundo de Cathy foi uma lufada de ar fresco para um projecto que à partida se poderia revelar interessante mas monótono. Cremos ter cumprido os nossos objectivos tanto a nível do conteúdo da disciplina como do nosso desejo em realizar um trabalho didáctico e divertido e isso é verdadeiramente gratificante. Afirmamos, sem falsas modéstias, que estamos orgulhosos deste projecto. Precisou de alguns dias sem comer e algumas noites sem dormir para formar uma história coerente e consistente, mas valeu a pena!

No entanto, para a missão de Cathy ficar verdadeiramente completa, faltava ainda convencer Mr. Pinkley das mais valias do programa. Munida do seu maravilhoso (só às vezes odiado) livro, dirigiu-se ao escritório do chefe…

Tinha conseguido!!! Agora iria para casa descansar e esperar que este trabalho tenha sido suficientemente bom para ser merecedor de um grande aumento salarial! E, consecutivamente, casar-se com o Irving e torná-lo um evento inesquecível!

BIBLIOGRAFIA

- Go Comics, Site Oficial. Cathy Classics by Cathy Guisewite, Outubro, Novembro e Dezembro de 2010 e Janeiro de 2011, <http://www.gocomics.com/cathy/>.

- **Pestana,** Maria Helena e **Gageiro,** João Nunes ; "Análise de Dados para Ciências Sociais – A Complementaridade do SPSS", Edições Sílabo, - 4ª Edição, 2005.

- **Maroco,** João; "Análise Estatística – com utilização do SPSS", Edições Sílabo – 3ª Edição, 2007.