

UNIVERSIDADE ABERTA

Mestrado em Estatística, Matemática e Computação

(Ramo – Estatística Computacional)

Análise Estatística do Processo de Reconhecimento, Validação e Certificação de Competências
no CNO dos CTT

Aluno: Carlos Manuel Fernandes Regra

Orientadora: Professora Doutora Teresa Paula Costa Azinheira Oliveira

(Professora Auxiliar da Universidade Aberta)

Agradecimentos

Como diria o autor, “O tempo não tem idade, tem tempo...”.

Elaborar um documento de dissertação requer, em primeiro lugar, vontade própria, apoio e disponibilidade para a sua execução.

É de facto, pelo apoio e disponibilidade sempre evidenciados, que quero expressar os meus agradecimentos à Professora Doutora Teresa Oliveira. Quero também testemunhar-lhe o meu reconhecimento pelos conhecimentos científicos partilhados, sem os quais não seria possível a realização deste trabalho.

Desejo igualmente, agradecer a todos aqueles que, de alguma forma, me incentivaram e apoiaram, nomeadamente ao Professor Amílcar Oliveira, aos colegas e amigos e aos CTT.

Por último, um agradecimento muito especial à minha família.

“ Aos meus filhos, a dedicação deste trabalho.

Do pai”

Índice

Simbologia e Notações	7
Resumo	9
Summary	10
CAPÍTULO I	
Introdução	11
CAPÍTULO II	
O CNO CTT	
2.1 – Criação e Pertinência do CNO CTT	12
2.2 – Missão do CNO CTT	12
2.3 – Organização e Funcionamento	13
2.4 – Diagnóstico e Caracterização da População	16
2.4.1 – Em 2003	16
2.4.2 – Em 2004	18
CAPÍTULO III	
Estudo de Variáveis e Identificação de Associações	
3.1 – Descrição da Variáveis mais Relevantes	21
3.2 – Investigação de Associações entre Variáveis	23
3.3 – Análise de Correlações e Regressão Linear	25
3.3.1 – Correlação Linear	25
3.3.2- Regressão Linear Simples	25

CAPÍTULO IV

Análise Descritiva para Comparação de Regiões

4 – Análise da População (B3)	28
4.1 – Análise das variáveis	28
4.2 – Análise Descritiva: Zona de Lisboa	33
4.2.1 – Lisboa: Relação Idade/Duração do Processo	34
4.2.2 – Lisboa: Relação Hab. Literárias/ Duração do Processo	37
4.3 – Análise Descritiva: Zona do Porto	39
4.3.1 – Porto: Relação Idade/Duração do Processo	40
4.3.2 – Porto: Relação Hab. Literárias/ Duração do Processo	43
4.4 – Análise Descritiva: Zona de Coimbra	44
4.4.1 – Coimbra: Relação Idade/Duração do Processo	45
4.4.2 – Coimbra: Relação Hab. Literárias/ Duração do Processo	47
4.5 – Análise Comparativa entre as Regiões de Lisboa, Porto e Coimbra versus Totalidade do País	50
4.5.1 – Totalidade do País: Relação Idade/Duração do Processo	50
4.5.2 – Totalidade do País: Relação Hab. Literárias/ Duração do Processo	53
4.5.3 – Totalidade do País: Análise Comparativa	54
4.5.4 - Totalidade do País: Correlação e Regressão Linear	61

CAPÍTULO V

Técnicas de Inferência Estatística Não Paramétrica

5.1 – Introdução	63
5.2 – Testes de Hipóteses. Generalidades	64
5.3 – Teste do Qui-quadrado para a Independência	65
5.4 – Testes para duas ou mais Amostras Independentes	66

5.4.1 – Teste de Mann-Whitney	66
5.4.2 – Teste de Krüskal-Wallis	68
5.5 – Testes para Comparação de Amostras Dependentes	72
5.5.1 – Teste dos Sinais	72
5.5.2 – Teste de Wilcoxon	74
5.5.3 – Teste de Friedman	75
5.6 – Testes de Correlação	76
5.6.1 - Teste de Correlação Paramétrica: Teste de Pearson	76
5.6.2 – Testes de Correlação Não Paramétrica	78
5.6.2.1 – Teste de Spearman	78
5.6.2.2 – Teste τ de Kendall	79

CAPÍTULO VI

Inferência Não Paramétrica aplicada à Avaliação de Desempenho

6.1 – Análise da Avaliação de Desempenho	81
6.2 – Análise Comparativa dos Testes utilizados (SPSS versus R)	89

CAPÍTULO VII

Técnicas de Visualização Gráfica e Metodologias de Superfície de Resposta

7.1 – Introdução	93
7.2 – Resenha Histórica	93
7.3 – As técnicas de Visualização na Actualidade	106
7.4 – Metodologia de Superfície de Resposta (MSR)	107

CAPÍTULO VIII

Considerações e Conclusões	112
Perspectivas Futuras	116
Bibliografia Utilizada	117
Apêndices	121

Simbologia e Notações

Símbolos	Descrição
CNO	Centro Novas Oportunidades
PEI	Plano Estratégico de Intervenção
RVC	Reconhecimento e Validação de Competências
RVCC	Reconhecimento, Validação e Certificação de Competências
SIGO	Sistema Integrado de Informação e Gestão da Oferta Educativa
CTT-EXP	CTT - Expresso
r	Coeficiente de Correlação
$C(x,y)$	Covariância de x e y
$Var(x)$	Variância de x
σ_x	Desvio Padrão de x
σ_y	Desvio Padrão de y
\bar{x}	Média de x
\bar{y}	Média de y
O_{ij}	Observação de duas variáveis x e y
H_0	Hipótese Nula
H_1	Hipótese Alternativa
E_{ij}	Valores Esperados
R	Coeficiente de Correlação Amostral (Pearson)
Z_x	Observações padronizadas de x
Z_y	Observações padronizadas de y
R^2	Coeficiente de Determinação
d_i	Diferença entre os números de ordem de cada par de observações x_i e y_i
R_s	Coeficiente de Correlação Ordinal de Spearman
u_i	Nº de Empates do i -ésimo grupo de observações iguais pertencentes à variável x

v_i	Nº de Empates do i-ésimo grupo de observações iguais pertencentes à variável y
θ_1	Mediana da população x
θ_2	Mediana da população y
$U = \min(U_1, U_2)$	Estatística de Teste de Mann-Whitney
$R(X_{iy})$	Ordem (rank) atribuída à observação
τ	Teste Tau de Kendall
τ_b	Teste Tau de Kendall (quando existem valores iguais)
S^+	Soma das diferenças Positivas
S^-	Soma das diferenças Negativas
$D_i = x_i - y_i$	Sinal da diferença entre x_i e y_i nos testes de sinais
χ_r^2	Estatística de Teste de Fridman
$R(x_{ij})$	Ordem atribuída à observação x_{ij}
R_j	Soma das $R(x_{ij})$
MSR	Metodologia de Superfície de resposta
RSM	Response Surface Methodology
$E(y)$	Resposta esperada
$\eta = f(x_1, x_2)$	Superfície de Resposta
MMQ	Método dos Mínimos Quadrados

Resumo

A presente dissertação assenta na análise descritiva de Variáveis e na Identificação de Associações relativas ao Processo RVCC, no CNO dos CTT. Pretendeu-se, também, efectuar a análise da avaliação de desempenho dos trabalhadores que viram certificadas as suas competências ao nível do 9º ano, antes e após a frequência deste processo. Procurámos ainda, no âmbito das Técnicas de Visualização Gráfica, explorar e aplicar as Metodologias de Superfície de Resposta, na análise das variáveis consideradas e inerentes ao estudo do processo RVCC.

Neste trabalho foram aplicados conhecimentos matemáticos e, na área da estatística, a inclusão de técnicas de estatística não paramétrica com recurso à utilização das ferramentas informáticas SPSS e à programação em R.

Seguidamente, procede-se à síntese de cada um dos capítulos:

No capítulo I, apresenta-se a Introdução à Dissertação e explicitam-se as razões pelas quais se decidiu efectuar este trabalho.

No capítulo II, faz-se o enquadramento do CNO dos CTT, explicam-se as razões que levaram à criação deste Centro e efectua-se a caracterização da população alvo dos CTT, que potencia o número de adultos em processo RVCC.

No capítulo III, analisam-se as variáveis consideradas mais importantes. De seguida, faz-se uma análise de Associações entre Variáveis e, por fim, a análise de Correlações e Regressão Linear. Trata-se, portanto, de um capítulo teórico sobre o trabalho a desenvolver nos capítulos seguintes.

No capítulo IV, efectua-se a análise descritiva para a comparação de Regiões. Considerou-se, por razões de estratégia, importância e tempo de realização da dissertação, que seriam analisadas as regiões do Porto, Coimbra e Lisboa, e a comparação com a totalidade do País.

No capítulo V, apresenta-se uma descrição das Técnicas de Inferência Estatística Não Paramétrica mais relevantes que, sempre que possível, procurámos utilizar na análise da avaliação de desempenho.

No capítulo VI, apresenta-se a análise da Inferência Não Paramétrica aplicada à Avaliação de Desempenho dos trabalhadores que concluíram o processo RVCC, certificados com o 9º ano de escolaridade e faz-se a comparação de resultados e comportamentos dos testes efectuados em SPSS e no R.

O Capítulo VII é dedicado às Técnicas de Visualização Gráfica e Metodologias de Superfície de Resposta. Apresentamos um desenvolvimento histórico sobre as técnicas de visualização gráfica e a análise e explicação das MSR – Metodologias de Superfície de Resposta.

No Capítulo VIII sintetizam-se as principais ideias desta dissertação e exploram-se eventuais perspectivas de trabalho futuro.

Summary

The present dissertation suits the descriptive analysis of Variables and Identification of relative Associations to the Process RVCC, in the CNO of the CTT.

It was also intended to effectuate the analysis of the evaluation of performance of the workers who saw their competences certified at the level of the 9th year, before and after the frequency of this process. It was also planned to apply the Response Surface Methodologies, in the analysis of the variables respected and linked to the study of the process RVCC in the context of the Techniques of Graphic Visualization.

Mathematical and statistical knowledge were used, namely the use of the statistic not parametric, through the use of the statistic software SPSS and of the planning in R.

Subsequently, we will synthesize each one of the chapters:

In chapter I we have the Introduction where we can find the purpose of this Dissertation.

In chapter II It is explained the framing of CNO of the CTT, the reasons that led to the creation of this Centre and some details are given about the characteristics of the CTT target-population, which can explain the high number of adults in the RVCC process.

In the chapter III, the most important variables were analysed. Later, some links between the variables are made and, finally, the analysis of correlations and linear regression. It is a theoretical chapter that will be developed in the next chapters.

In the chapter IV, it is presented a comparison analysis of Regions. For strategic reasons it was decided to analyse only the regions of Oporto, Coimbra and Lisbon and the comparison between them and the totality of the country.

In the chapter V, it is presented the Not Parametric Techniques of Statistical Inference that were used for the analysis of the evaluation of performance.

In the chapter VI, we focus on the analysis of the Inference Not Parameter devoted to the workers evaluation performance that ended the process RVCC, certified with the 9th year of schooling and we make the comparison of tests done in SPSS and R.

The Chapter VII is dedicated to the Techniques of Graphic Visualization and Response Surface Methodologies. A historical development is done on the techniques of graphic visualization and the analysis and explanation of the MSR – Response Surface Methodologies.

In the Chapter VIII, the main ideas of this dissertation summarize and eventual perspectives of future work are explored.

CAPÍTULO I

Introdução

O trabalho desenvolvido insere-se num estudo sobre a população candidata ao processo RVCC-Reconhecimento, Validação e Certificação de Competências no CNO dos CTT – Correios de Portugal, S.A. Pretendemos efectuar uma análise, recorrendo a técnicas e modelações estatísticas, no sentido de obtenção de respostas a novas questões ainda não devidamente exploradas, como a identificação das principais variáveis envolvidas na problemática da procura da certificação e as suas relações.

Após um estudo inicial de análise exploratória de dados, para caracterização da amostra da população, recorreremos aos testes não paramétricos para aplicação de modelações estatísticas que se prendem sobretudo com a análise multivariada, análise de regressão e metodologias de superfície de respostas.

Este estudo foi efectuado utilizando as ferramentas informáticas SPSS e a programação em R, sempre que tal foi possível e ou necessário.

O trabalho desenvolvido enquadra-se na actividade exercida pelo aluno no Centro de Novas Oportunidades dos CTT, como formador, no âmbito do Processo de Reconhecimento e Validação de Competências.

Procurámos tirar partido, explorar e aplicar as metodologias assimiladas no âmbito das unidades curriculares da parte escolar do Mestrado em Estatística, Matemática e Computação e das três vertentes que o compõem.

Assim, este trabalho, entre outros aspectos, teve por objectivo investigar:

- A influência da região de localização do candidato no processo RVCC;
- A influência da idade dos candidatos na conclusão do processo;
- A influência das habilitações literárias dos candidatos na conclusão do processo;
- A importância e o impacto da formação no desenvolvimento de competências e sua relação com as principais variáveis em estudo.

Devido, por um lado, à enorme quantidade de dados disponíveis e por outro, à escassez de tempo útil para o efeito, optámos por efectuar a análise de apenas três regiões do País, de entre as mais significativas, (Lisboa, Porto e Coimbra) e de efectuar o estudo apenas sobre algumas das variáveis relativas à população que se propunha a certificação do nível B3 (9º Ano).

Por fim, e pelas razões indicadas, a problemática da Metodologia de Superfície de Resposta foi tratada apenas a nível teórico, sendo que ambicionamos em trabalhos futuros procurar aplicar e expandir esta técnica, dado ser um tema com evidente relevância e pertinência na actualidade.

CAPÍTULO II

O CNO - CTT . Caracterização da população

2.1 Criação e pertinência do CNO – CTT

O Centro Novas Oportunidades dos CTT foi criado na sequência da acreditação dos CTT - Correios de Portugal, SA como entidade potencial promotora de um centro de reconhecimento, validação e certificação de competências, conforme Despacho n.º 1050/2004 (2ª série) da Direcção-Geral de Formação Vocacional publicado no Diário da República n.º 13, II Série, de 16 de Janeiro de 2004.

Para a tomada de decisão, de promover a constituição, nos CTT, de um Centro de RVCC, visando prioritariamente a satisfação de necessidades de públicos internos com baixos níveis de escolaridade, esteve presente:

- A elevação do nível de qualificação dos recursos humanos dos CTT;
- A elevação da auto-estima, do nível de aspiração profissional e do bem-estar social dos colaboradores da empresa;
- A contribuição para um melhor ajustamento/adequação das competências dos colaboradores da empresa à evolução dos modelos e sistemas de gestão de recursos humanos no sentido de uma maior autonomia, responsabilização e orientação para resultados;
- A assumpção da iniciativa no domínio do desenvolvimento e da qualificação;
- O reforço das vantagens competitivas e a imagem dos CTT no mercado.

2.2 Missão do CNO CTT

A missão do CNO dos CTT será proporcionar, prioritariamente aos colaboradores dos CTT e, eventualmente, a outros cidadãos maiores de 18 anos, no âmbito de parcerias ou acordos a estabelecer com outras entidades, a oportunidade e os meios para verem formalmente reconhecidos os conhecimentos e competências que foram adquirindo ao longo do seu percurso de vida, nos mais variados contextos, por vias não formais e informais, mediante a obtenção de uma certificação escolar oficial de nível Secundário, B3, B2 ou B1, equivalente, respectivamente, ao 12º ano, 9º ano, 6º ano ou 4º ano do Ensino Básico do Sistema Educativo.

As actividades desenvolvidas no CNO dos CTT baseiam-se em princípios e compromissos com as necessidades de autonomia, integridade e bem-estar social dos clientes/utilizadores do CNO dos CTT, constituindo compromisso da Equipa deste Centro, colocar todas as suas competências e recursos ao serviço das necessidades de informação, orientação e apoio dos candidatos, tendo em vista o desenvolvimento das qualificações e competências inerentes ao

exercício de uma cidadania plena, activa e responsável e ao desenvolvimento integral e harmonioso da pessoa humana.

Os valores e princípios que norteiam o funcionamento e as actividades do CNO dos CTT na relação com os utilizadores/candidatos, com os parceiros internos e externos e com a Comunidade em geral, baseiam-se na:

- Dignidade da pessoa humana
- Igualdade de direitos e de oportunidades e o respeito pela diferença
- Valorização da identidade, responsabilidade e autonomia individuais
- Respeito pela privacidade individual e pela vida familiar
- Confidencialidade e a protecção dos dados de carácter pessoal
- Direito de informação, a liberdade de expressão e de opinião
- Direito à educação e formação ao longo da vida
- Ética, rigor, profissionalismo, integridade, lealdade e transparência
- Cooperação, solidariedade, entajuda e partilha de recursos e competências
- Abertura à inovação e à mudança
- Valorização da apropriação do conhecimento, da reflexão e do raciocínio críticos, e do auto-conhecimento
- Promoção e valorização do desenvolvimento e realização pessoais

Os CTT, como empresa nacional ciente da sua responsabilidade social, celebrou recentemente, um protocolo com os Ministérios da Educação, do Trabalho e da Solidariedade, no sentido de acolher no CNO-CTT elementos de população carenciada, com o objectivo de que a mesma aumente as suas habilitações literárias.

2.3 Organização e funcionamento

Relativamente ao modelo de organização e forma de funcionamento, o Centro Novas Oportunidades dos CTT visa prioritariamente a satisfação de necessidades de reconhecimento, validação e certificação de competências de públicos internos (colaboradores dos CTT) com baixos níveis de escolaridade.

A resposta a necessidades de reconhecimento, validação e certificação de competências de colaboradores residentes fora da área da Grande Lisboa será assegurada através de soluções de itinerância, de articulação com recursos descentralizados e/ou de acordos de colaboração com

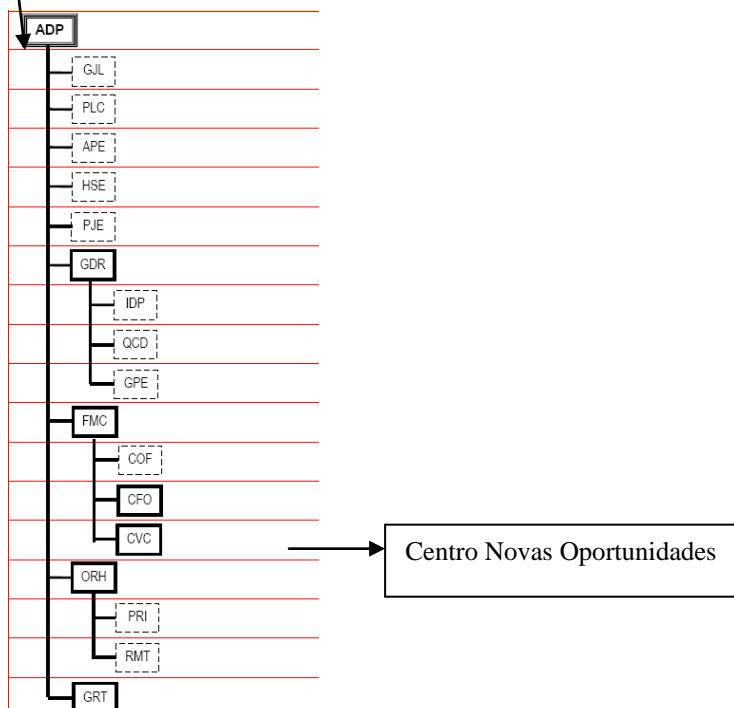
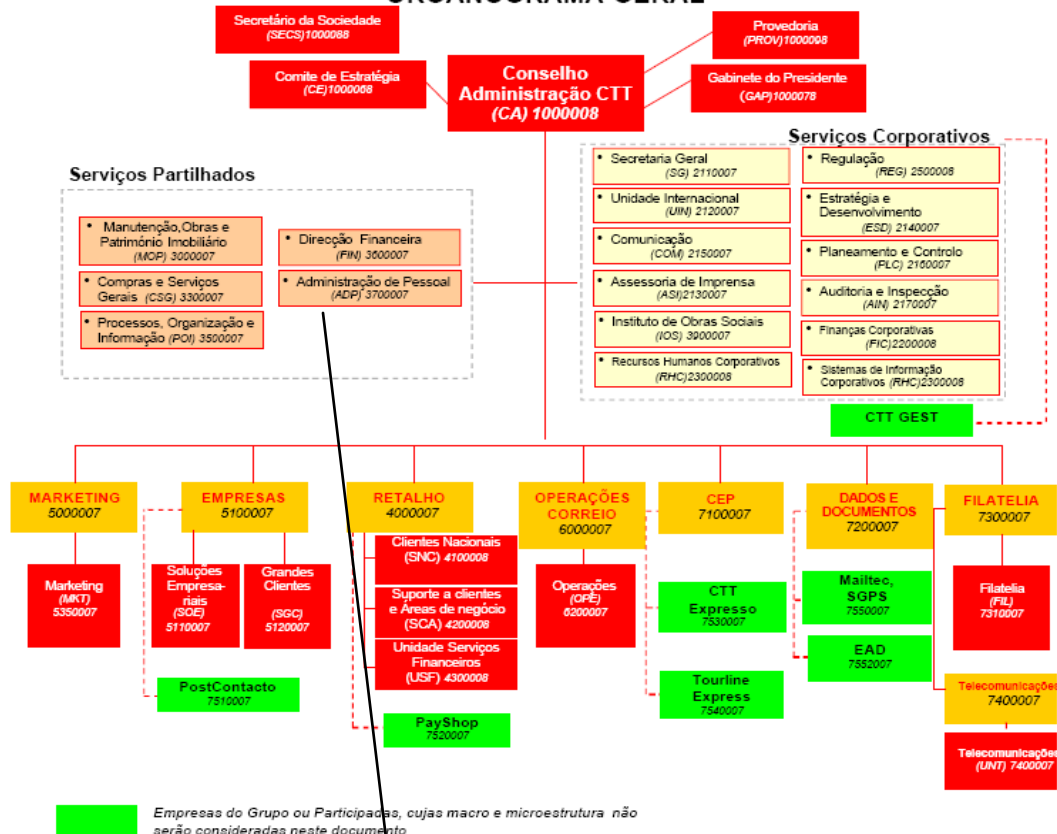
outros Centros RVCC.

Apesar do Centro RVCC dos CTT visar, prioritariamente, públicos alvo internos, não está excluída a possibilidade de, futuramente, poder abranger outros públicos, mediante a celebração de acordos ou protocolos com outras entidades, nomeadamente do Sector das Comunicações.

De acordo com o modelo de organização definido, o Centro RVCC funciona nas instalações dos CTT, na Rua da Palma, nº 236, em Lisboa. No entanto, as actividades desenvolvidas no decurso do processo RVCC, algumas presenciais, realizam-se em todo o território do continente onde os CTT se encontram e onde existem candidatos ao processo.

Para uma melhor visualização e compreensão do departamento que constitui o CNO dos CTT na estrutura organizacional da Empresa, apresentamos, de seguida, o seu organograma:

ORGANOGRAMA GERAL

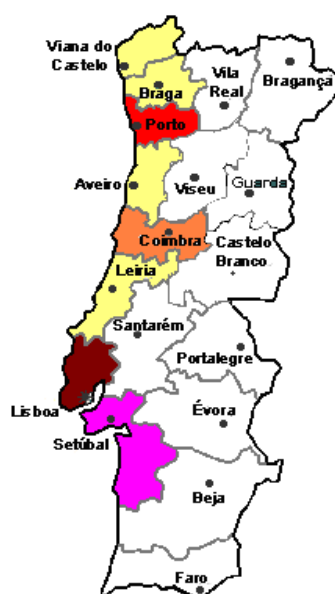


2.4 Diagnóstico e caracterização da população

2.4.1 Em 2003

Foi efectuado o diagnóstico de necessidades e efectuada a caracterização do público-alvo com base em dados referentes ao final do primeiro semestre de 2003.

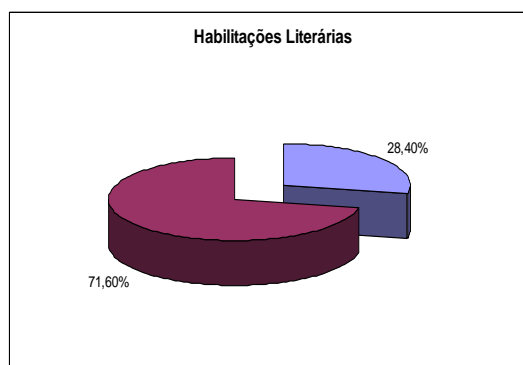
Geograficamente, à semelhança do «padrão» geral de distribuição da população portuguesa, os colaboradores dos CTT com habilitações inferiores ao 9º ano de escolaridade concentram-se maioritariamente (77,9%) na faixa litoral de Viana do Castelo a Setúbal, com maior expressão em Lisboa (26,5%), Porto (16,6%), Setúbal (9,0%) e Coimbra (6,9%).



Legenda: (% colaboradores com habilitações inferiores ao 9º Ano)



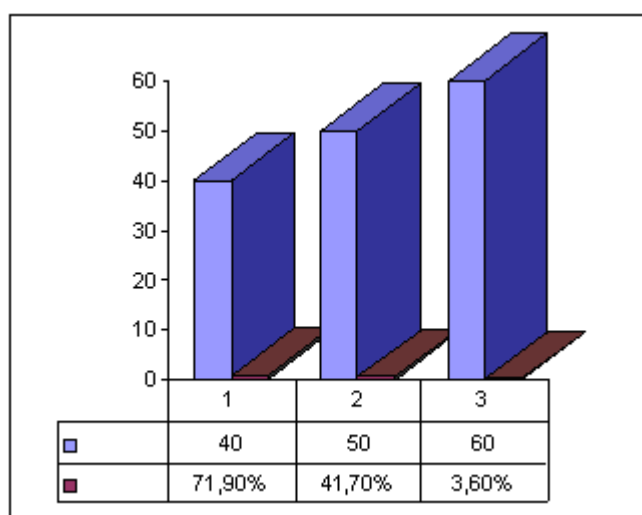
Nessa altura, constatava-se a existência de 4.673 colaboradores com habilitações inferiores ao 9º ano de escolaridade do Ensino Básico, representando 28,4% do efectivo total dos CTT naquela data (16.475).



■ ≥ 9º Ano ■ < 9º Ano

84,4% daqueles colaboradores pertenciam ao grupo profissional Carteiro.

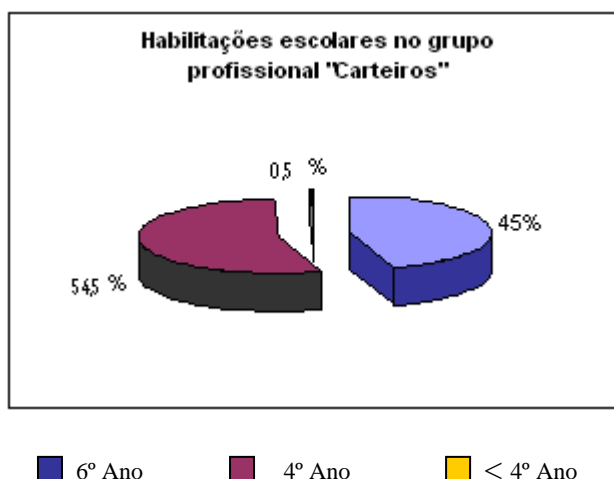
Verificava-se, também, que 71,9% dos 4.673 colaboradores com habilitações abaixo do 9º ano tinham mais de 40 anos; 41,7%, tinham mais de 50 anos e 3,6% mais de 60 anos.



■ - Idade superior a (em anos)

■ - % do total de trabalhadores com habilitações literárias inferiores ao 9º ano

Relativamente às suas habilitações escolares, 45,0% tinham o 2º Ciclo do Ensino Básico (6º ano), 54,5% o 1º Ciclo do Ensino Básico (4º ano) e cerca de 0,5% habilitações inferiores ao 1º Ciclo.

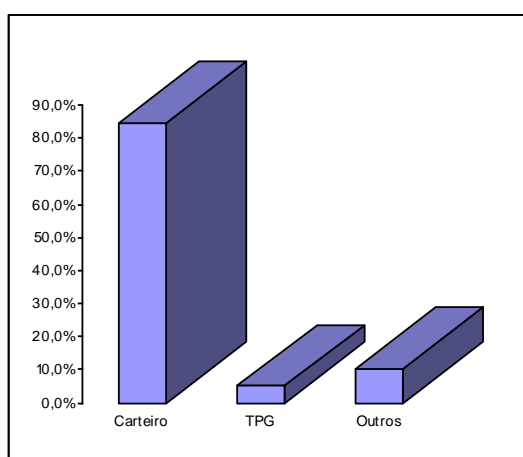


Ao longo do ano de 2003 verificou-se uma redução do efectivo total dos CTT, na ordem dos 7,5%, devida, fundamentalmente, a saídas por aposentação e contenção nas admissões e na contratação a termo.

2.4.2 Em 2004

No início de 2004 (dados relativos a 1 de Janeiro de 2004), registamos a existência de 4.302 colaboradores com habilitações inferiores ao 9º ano de escolaridade, representando 28,2% do efectivo total dos CTT (15.278).

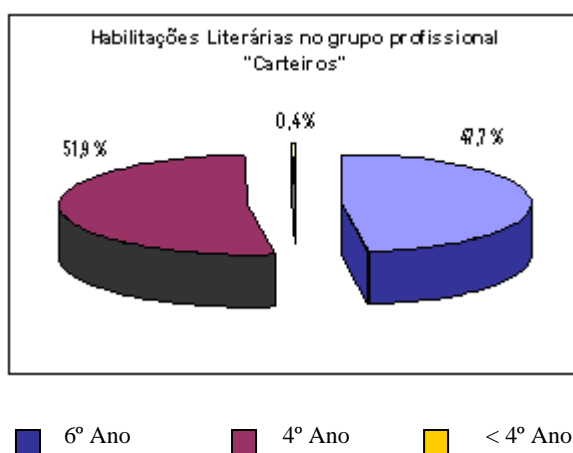
Destes 4.302 colaboradores, 84,6% pertencem ao grupo profissional Carteiro. Os restantes 15% estão distribuídos por vários grupos profissionais, sendo o mais significativo, a seguir ao dos Carteiros, o dos Técnicos Postais e de Gestão (Altamente Qualificados), com 5,0%.



Observamos, ainda, a ocorrência pontual de colaboradores com habilitações abaixo do 9º ano de escolaridade em grupos profissionais dos níveis de Quadros Médios e Quadros Superiores, dados que parecem indiciar situações claras de «subcertificação» das competências efectivamente detidas.

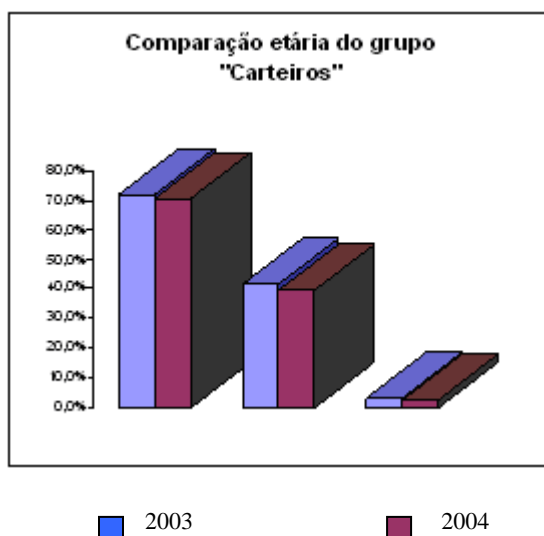
Apesar de, no curto período analisado, se observar uma ligeira tendência para o aumento da percentagem de mulheres no conjunto do efectivo da empresa, não se constata a correspondente diminuição (pelo contrário, acentua-se ligeiramente) do peso relativo dos colaboradores do sexo masculino na população com habilitações abaixo do 9º ano.

No que diz respeito à distribuição das habilitações escolares, a evolução, se bem que ligeira, é positiva: a percentagem de colaboradores com o 2º Ciclo do Ensino Básico passa de 45,0% para 47,7% e a de colaboradores com o 1º Ciclo do Ensino Básico de 54,5% para 51,9%. A percentagem de colaboradores com habilitações inferiores ao 1º Ciclo reduz-se para 0,4%.



Comparando o nível de habilitações literárias do grupo de carteiros em 2003 e 2004, notamos uma ligeira melhoria:

Relativamente à distribuição por escalões etários, a evolução observada no segundo semestre de 2003 é também, apesar de ligeira, positiva, verificando-se algum «rejuvenescimento» da população: a percentagem de colaboradores com habilitações inferiores ao 9º ano com mais de 40 anos desce de 71,9% para 70,7%, com mais de 50 anos de 41,7% para 39,6% e com mais de 60 anos de 3,6% para 2,4%.



No mesmo período, o número de carteiros com habilitações inferiores ao 9º ano de escolaridade decresce 7,7%, de 3.943 para 3.641. Paradoxalmente, à primeira vista, a percentagem de colaboradores do grupo profissional “Carteiro” com habilitações abaixo do 9º ano sobe de 42,2% para 43,0%, em resultado de uma redução ainda mais significativa (9,2%) verificada no efectivo total do grupo.

Quadro Resumo dos indicadores

Descrição	25 Junho 2003		01 Janeiro 2004
Efectivo total	16.475	↘	15.278
(idem) do sexo masculino	66,7%	↘	66,6%
(idem) do sexo feminino	33,3%	↗	33,4%
(idem) na situação de activo	16.184	↘	15.009
Efectivo total com habilitações inferiores ao 9º ano	4.673 (28,4%)	↘ (↘)	4.302 (28,2%)
(idem) no activo	4.594 (28,4%)	↘ (↘)	4.248 (28,3%)
(idem) do sexo masculino	88,3%	↗	88,4%
(idem) do sexo feminino	11,7%	↘	11,6%
(idem) com mais de 40 anos	71,9%	↘	70,7%
(idem) com mais de 50 anos	41,7	↘	39,6%
(idem) com mais de 60 anos	3,6%	↘	2,4%
(idem) do grupo prof. Carteiro	3.943 (84,4%)	↘ (↗)	3.641 (84,6%)
(idem) com o 2º Ciclo do EB (6º ano)	45,0%	↗	47,7%
(idem) com o 1º Ciclo do EB (4º ano)	54,5%	↘	51,9%
(idem) com menos do que 1º Ciclo	0,5%	↘	0,4%
Percentagem de Carteiros com habilitações inferiores ao 9º ano	42,2%	↗	43,0%

CAPÍTULO III

Estudo de variáveis e identificação de associações

3.1 Descrição das variáveis mais relevantes

Através do perfil da amostra, poderiam ser analisados os seguintes aspectos relativamente à sua distribuição: idade, género, categoria profissional, duração do processo, região (distrito), serviço de afectação da função desempenhada nos CTT, utilização de computador, possuir curso de informática, situação no processo e habilitações literárias conforme se indica no quadro seguinte. No entanto, optámos por efectuar uma análise mais restrita e que se encontra descrita no Capítulo IV.

	Idade	Afectação	Distrito	Categoria Profissional	Curso Informática	Utilização Computador	Hab. Literárias	Situação	Duração
Género	x	x	x	x	x	x	x	x	x
Idade		x	x	x	x	x	x	x	x
Afectação			x	x	x	x	x	x	x
Distrito				x	x	x	x	x	x
Categoria Prof.					x	x	x	x	x
Curso Informática.						x	x	x	x
Utilização Comp.							x	x	x
Habilitações Lit.								x	x
Situação									x

Passemos então a analisar cada uma das variáveis indicadas no quadro referenciado:

Género: A descrição desta variável assenta no facto do adulto ser do sexo masculino ou do sexo feminino. Para efeitos de análise em SPSS, considerou-se a codificação 1 para Masculino e 2 para Feminino.

Idade: Para efeitos de análise, consideraram-se as seguintes classes etárias:

25 – 30; 30 – 35; 35 – 40; 40 – 45; 45 – 50; 50 – 55; 55 – 60; 60 – 65

Para efeitos de análise em SPSS, considerámos que:

1	2	3	4	5	6	7	8
25 - 30	30 - 35	35 - 40	40 - 45	45 - 50	50 - 55	55 - 60	60 - 65

Afectação: A variável afectação indica, em termos estruturais da empresa, a alocação do adulto: SNC – Serviço Nacional de Clientes; OPE – Operações; CTT-EXP – CTT Expresso; ADP – Administração de Pessoal; CSG – Compras e Serviços Gerais; POI – Processos; FIL – Filatelia; MOP – Manutenção, Obras e Projectos; SG – Secretaria Geral; SCA – Suporte a Clientes e Área de Negócios; POSCONTACT – Post Contacto.

No entanto, para efeitos de análise, optámos por considerar apenas os serviços com maior representatividade neste processo: SNC; OPE e CTT-EXP, já que o número de adultos afectos aos outros serviços não tem expressão estatística.

Para efeitos de análise em SPSS, considerámos a seguinte codificação: 1 – SNC; 2 – OPE e 3 – CTT-EXP.

Distrito: Apesar de se encontrarem inscritos neste processo trabalhadores dos distritos de Lisboa, Porto, Coimbra, Braga, Castelo Branco, Viseu, Viana do Castelo, Portalegre, Leiria, Faro, Évora, Beja, Aveiro, Setúbal, Santarém, Vila Real, Guarda e Região Autónoma dos Açores, este estudo irá contemplar apenas os distritos de Lisboa, Porto e Coimbra uma vez que são estes distritos que contemplam a maioria da população em causa.

Para efeitos de análise em SPSS, considerámos a codificação 1 para Lisboa, 2 para o Porto e 3 para Coimbra.

Categoria Profissional: As categorias profissionais dos trabalhadores dos CTT que frequentaram este processo são diversas: CRT – Carteiro; TPG – Técnico postal e de gestão; TDG – Técnico de desenvolvimento e gestão de sistemas; TSG – Técnico de serviços gerais; TAC – Técnico administrativo e controlo de gestão; TIE – Técnico de infra-estruturas e equipamentos; MOT – Motorista; DIST – Distribuidor (CTT-EXP); ASG – Assistente; TEP – Técnico de equipamento postal; EPT – Especialista postal; TRA – Técnico de representação gráfica áudio e visual; ESE – Empregado de serviços elementares.

Para efeitos de análise, considerámos as categorias com maior representatividade dos trabalhadores, sendo todas as outras englobadas em “Outros”:

1	2	3	4	5	6
CRT	TPG	DIST	MOT	TAC	OUTROS

Curso de Informática: A variável em causa permite saber quais os trabalhadores que possuem ou frequentaram quaisquer cursos de informática. Neste estudo, esta variável não será objecto de análise.

Utilização de Computador: A variável em causa permite saber quais os trabalhadores que utilizam ou não, o computador. Neste estudo, também esta variável não será objecto de análise.

Habilitações Literárias: Considerou-se de extrema importância o estudo desta variável e principalmente a sua relação com a duração do processo. Considerou-se que os adultos eram portadores do 2º Ciclo do Ensino Básico (6º Ano) ou do 1º Ciclo do Ensino Básico (4º Ano).

Para efeitos de análise em SPSS, considerámos a codificação:

1 para o 2º Ciclo do Ensino Básico (6º Ano) e 2 para o 1º Ciclo do Ensino Básico.

Situação no Processo: Esta variável dá a conhecer a situação do adulto no processo RVCC. Isto é, se foi “Certificado”, se foi “Transferido para outro CNO”, se foi “Encaminhado para outra oferta formativa”, se “Desistiu”, ou se encontra “Em Reconhecimento”, ou seja, se ainda frequenta o processo ou se encontra apenas “Inscrito”, não tendo iniciado o processo.

Para efeitos de análise em SPSS, considerámos a classificação constante na tabela:

1	2	3	4	5	6
Certificado	Transferido	Encaminhado	Desistência	Em Reconhecimento	Inscrito

Duração do Processo: Esta variável, inicialmente definida em dias de duração do processo, para efeito de análise, foi transformada em trimestres. Assim, considerámos para efeitos de análise em SPSS, a seguinte codificação:

1	2	3	4	5	6	7
Até 90 dias (1 Trimestre)	De 91 a 180 dias (2 Trimestres)	De 181 a 270 dias (3 Trimestres)	De 271 a 360 dias (4 Trimestres)	De 361 a 450 dias (5 Trimestres)	De 451 dias a 540 dias (6 Trimestres)	Mais de 540 dias (Mais de 6 Trimestres)

Tendo em atenção a análise efectuada, o nosso estudo incidirá apenas nas variáveis discriminadas no seguinte quadro resumo:

Variável	Descriminação
Género	Masculino; Feminino
Idade	Por classes: [25;30[, [30;35[, [35;40[, [40;45[, [45;50[, [50;55[, [55;60[
Afectação	SNC; OPE; CTT-EXP
Distrito	Lisboa; Porto; Coimbra
Categoria Profissional	CRT; TPG; DIST; MOT; TAC; Outros
Habilitações Literárias	2º Ciclo do Ensino Básico; 2º Ciclo do Ensino Básico
Situação no Processo	Certificado; Transferido; Encaminhado; Desistência; Em Reconhecimento; Inscrito
Duração do Processo	Análise trimestral: 1, 2, 3, 4, 5, 6, >6

3.2 Investigação de associações entre variáveis

Dadas duas variáveis, convém investigar a sua relação estudando a dependência ou independência entre elas. Uma questão que se pode colocar, à partida, é a seguinte: E se forem dependentes, qual o tipo e grau de dependência?

Medir o grau de dependência é mais ambicioso do que simplesmente testar a existência de alguma associação entre variáveis. É, de todo o interesse poder medir o grau de associação entre dois conjuntos de observações obtidos a partir de um dado conjunto de unidades experimentais (indivíduos por exemplo).

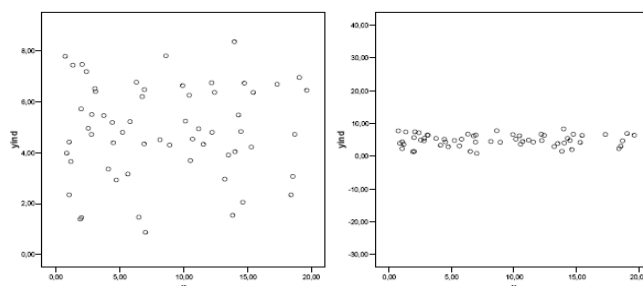
Existem diversas formas de associação entre variáveis numéricas, das quais se destacam as relações lineares e as relações curvilineas.

Como analisar, então, a associação linear entre duas variáveis numéricas?

Por exemplo, através da construção de diagramas de dispersão e observação da disposição da nuvem de pontos.

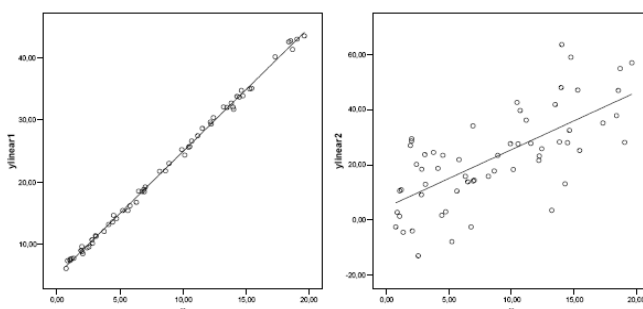
Quando duas variáveis são independentes, o diagrama de dispersão respectivo apresenta uma mancha de pontos aleatória ou, quando muito, um conjunto de pontos dispostos sobre uma recta horizontal.

Apresentamos, como exemplos, as seguintes figuras:



Se a relação entre duas variáveis for linear, ao confrontar duas amostras num diagrama de dispersão deve esperar observar-se um conjunto de pontos que se dispõem aproximadamente sobre uma recta. Por vezes, os desvios em relação à suposta recta são mínimos, mas noutras situações, os pontos apresentam bastante dispersão tornando difícil a identificação da dita relação linear que, a existir, será fraca.

Ilustramos estas situações nas figuras que se seguem:



Outra forma de analisar essa associação, será através da sua quantificação, ou seja, do cálculo de medidas de associação, como por exemplo, os coeficientes de correlação. Podemos também, através da realização de um teste de hipóteses, averiguar se os valores das medidas de associação observados nos dados, são significativos, ou seja, se é possível, concluir estatisticamente, a favor da existência de uma associação significativa na população.

No nosso estudo, tornou-se pertinente averiguar da existência de relações entre as variáveis “Idade” versus “Duração do Processo” e “Habilitações Literárias” versus “Duração do Processo”.

3.3 Análise de Correlações e Regressão Linear

3.3.1 Correlação Linear

Entende-se por correlação estatística, o grau de associação entre duas variáveis aleatórias. A dependência de duas variáveis X e Y é dada pelo coeficiente de correlação amostral, conhecido também por coeficiente r-de-Pearson. Designa-se, normalmente, por r e é determinado por

$$r = \frac{\text{Cov}(X, Y)}{s_X \cdot s_Y}$$

$$\text{Onde } \text{Cov}(X, Y) = \frac{\sum_{i=1}^n XY}{n} - \bar{X} \bar{Y} ,$$

$$s_X = \sqrt{\frac{\sum_{i=1}^k f_i (X_i - \bar{X})^2}{n}} \text{ e } s_Y = \sqrt{\frac{\sum_{i=1}^k f_i (Y_i - \bar{Y})^2}{n}}$$

Se $r = 1$, estamos perante uma correlação total no mesmo sentido, isto é, à medida que aumentam os valores de X, aumentam proporcionalmente os valores de Y.

Se $r = -1$, estamos, também, perante uma correlação total, agora no sentido inverso, isto é, à medida que aumentam os valores de X, os valores de Y diminuem.

Se $r = 0$, então não existe correlação.

Ao efectuarmos a representação gráfica de uma distribuição de duas variáveis, obtém-se o chamado diagrama de dispersão. Obtido este gráfico e a nuvem de pontos, o objectivo será determinar uma linha que melhor se ajuste e represente essa nuvem de pontos e que permita estimar os valores de Y em função dos valores de X.

Quando essa linha é uma recta, estamos perante uma recta de regressão. Para determinação da recta que melhor se ajuste ao diagrama de dispersão, utiliza-se o método dos mínimos quadrados.

3.3.2 Regressão Linear Simples

A regressão constitui uma técnica estatística poderosa e de utilização bastante frequente. Os modelos de regressão têm revelado inúmeras aplicações, nas mais diversas áreas, pelo que enveredamos pela sua abordagem, em especial à regressão linear simples, da qual faremos, dentro do possível, aplicação aos nossos dados.

Em termos gerais um modelo de regressão linear simples descreve uma relação entre duas variáveis aleatórias do tipo quantitativo, uma delas sendo independente, X e a outra dependente, Y . Assim, admitindo que a linha que melhor se ajusta à nuvem de pontos é uma recta, temos o seguinte modelo centrado:

$$\hat{Y}_i = \alpha + \beta(X_i - \bar{X}) \quad , i=1, \dots, n,$$

$$\text{Com: } \text{Var}(Y_i) = \sigma^2, \text{ e } \bar{X} = \sum_{i=1}^n \frac{X_i}{n}.$$

Outra especificação pode ser feita atendendo aos valores observados, como sendo:

$Y_i = \alpha + \beta(X_i - \bar{X}) + \varepsilon_i$, $i=1, \dots, n$, onde $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ são variáveis aleatórias com valor médio nulo e $\text{Var}(\varepsilon_i) = \sigma^2$. Assim, o valor esperado de ε_i não depende dos valores de x_i .

Admitimos que os Y_1, Y_2, \dots, Y_n são mutuamente independentes, assim como os $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$.

Posto isto, a problemática consiste agora na estimação dos valores dos parâmetros α, β e σ^2 , que permitirão uma posterior abordagem inferencial, caso se justifique.

De modo a evitar o critério individual na construção de rectas de ajustamento que melhor se adaptem ao conjunto de dados, instituiu-se, como definição da melhor recta de ajustamento, aquela que minimiza a soma dos quadrados dos erros ou desvios. Surge assim o amplamente conhecido método dos mínimos quadrados.

Consideremos a entidade:

$$W(\alpha, \beta) = \sum_{i=1}^n [Y_i - \hat{Y}_i]^2 = \sum_{i=1}^n [Y_i - (\alpha + \beta(X_i - \bar{X}))]^2.$$

Determinemos os minimizantes da função $W(\alpha, \beta)$, que designaremos por a e b , respectivamente.

Assim, tem-se:

$$\begin{cases} \frac{\partial W(\alpha, \beta)}{\partial \alpha} = 0 \\ \frac{\partial W(\alpha, \beta)}{\partial \beta} = 0 \end{cases}$$

► Desenvolvendo, em ordem a α , tem-se:

$$\frac{\partial W(\alpha, \beta)}{\partial \alpha} = 0 \Leftrightarrow \frac{\partial \left(\sum_{i=1}^n [Y_i - (\alpha + \beta(X_i - \bar{X}))]^2 \right)}{\partial \alpha} = 0$$

$$\Leftrightarrow -2 \sum_{i=1}^n [Y_i - \alpha - \beta(X_i - \bar{X})] = 0$$

$$\Leftrightarrow \sum_{i=1}^n Y_i - \sum_{i=1}^n \alpha - \beta \sum_{i=1}^n (X_i - \bar{X}) = 0 \Leftrightarrow \sum_{i=1}^n Y_i - n\alpha = 0 \Leftrightarrow \hat{\alpha} = \frac{\sum_{i=1}^n Y_i}{n}$$

Ou equivalentemente:

$$a = \bar{Y} ,$$

sendo a o minimizante de α .

► Desenvolvendo, em ordem a β , tem-se:

$$\frac{\partial W(\alpha, \beta)}{\partial \beta} = 0 \Leftrightarrow \frac{\partial \left(\sum_{i=1}^n [Y_i - (\alpha + \beta(X_i - \bar{X}))]^2 \right)}{\partial \beta} = 0$$

$$\Leftrightarrow -2 \sum_{i=1}^n (X_i - \bar{X}) [Y_i - \alpha - \beta(X_i - \bar{X})] = 0$$

$$\Leftrightarrow \sum_{i=1}^n (X_i - \bar{X}) Y_i - \alpha \sum_{i=1}^n (X_i - \bar{X}) - \beta \sum_{i=1}^n (X_i - \bar{X})^2 = 0$$

$$\Leftrightarrow \hat{\beta} = \frac{\sum_{i=1}^n (X_i - \bar{X}) Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Ou equivalentemente:

$$b = \frac{\sum_{i=1}^n (X_i - \bar{X}) Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2} ,$$

sendo b o minimizante de β .

CAPÍTULO IV

Análise descritiva para comparação de Regiões

4 Análise da população (B3)

Como já foi referido, a população em análise é constituída por 570 trabalhadores inscritos no CNO - CTT com o objectivo de concluir o 9º ano através do processo de reconhecimento, validação e certificação de competências e é proveniente de um universo de cerca de 4000 trabalhadores dos CTT que possuem à altura (2004), habilitações literárias inferiores ao 9º ano, independentemente do ano em que foram admitidos na Empresa. De acordo com os dados disponíveis, elaborámos o quadro seguinte que reflecte as variáveis em presença no nosso estudo.

Na nossa análise, abordámos os aspectos com mais impacto do ponto de vista empresarial, dando ênfase aos seguintes aspectos: Idade, Categoria Profissional, Afectação, Distrito, Situação no Processo e a Duração do Processo.

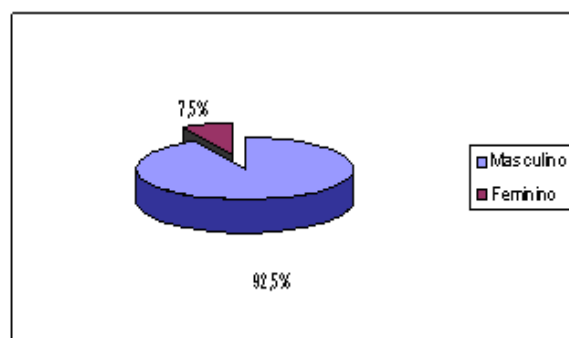
	Idade	Afect.	Distrito	C Prof	H. Lit	Situação	Duração
Género	x	x	x	x	x	x	x
Idade		x	x	x	x	x	x
Afect.			x	x	x	x	x
Distrito				x	x	x	x
C Prof					x	x	x
H. Lit						x	x
Situação							x

4.1 Análise das variáveis

Género:

Pela análise da tabela e do respectivo diagrama pode constatar-se que 527 candidatos são do sexo masculino enquanto que apenas 43 constituem o grupo do sexo feminino. Os homens representam, assim, 92,5% da população em estudo.

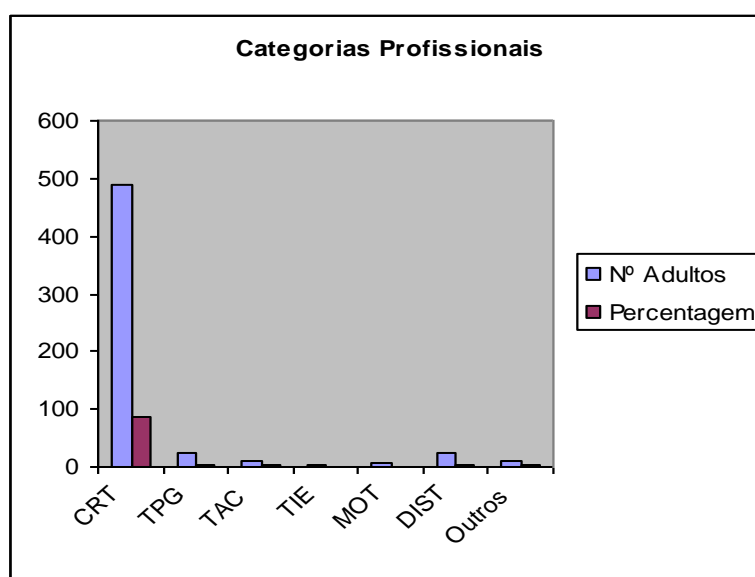
Género	Frequência	%	% Acumulada
1 - Masculino	527	92,5	92,5
2- Feminino	43	7,5	100
Total	570	100	



Categoria Profissional.

A tabela e gráfico seguintes mostram-nos que, a grande maioria destes adultos à certificação de competências ao nível de B3, pertence ao grupo profissional dos Carteiros (493), representando 85,8% da população. Segue-se o grupo profissional de Técnico Postal e de Gestão com 4,4% e o grupo de Distribuidor (CTT-EXP) com 4,2%. Os outros grupos profissionais quase não têm expressão.

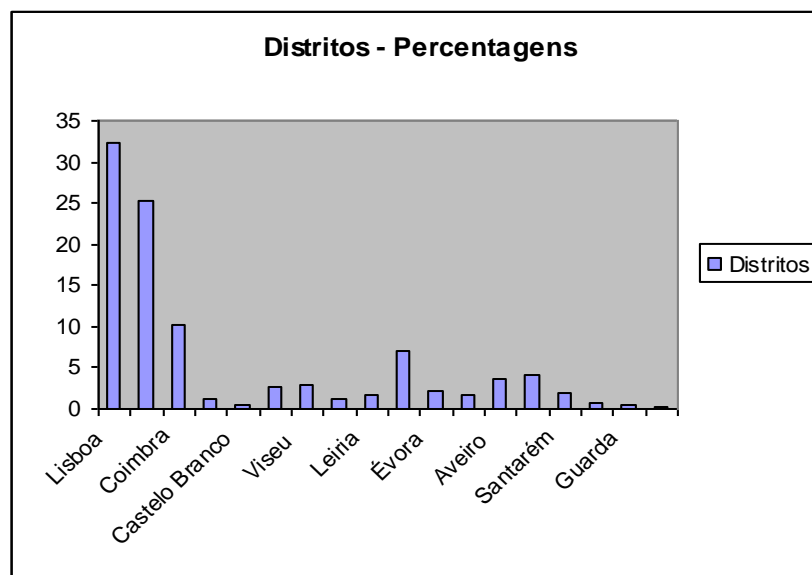
Categoria Profissional	Frequência	%	% Acumulada
1 - CRT	489	85,8	85,8
2 - TPG	25	4,4	90,2
3 - TAC	11	1,9	92,1
4 - TIE	4	0,7	92,8
5 - MOT	6	1,1	93,9
6 - DIST	24	4,2	98,1
7 - Outros	11	1,9	100,0
Total	570	100,0	



Distrito/Região:

A análise da tabela seguinte mostra-nos como se encontram distribuídos geograficamente os candidatos. Lisboa, Porto e Coimbra são os distritos mais representativos com, respectivamente, 184, 144 e 58 candidatos, equivalente a 32,3%, 25,3% e 10,2% do total. Seguem-se os distritos de Faro (7%), Setúbal (4,2%) e Viseu com 2,8% do total de candidatos. Os outros distritos indicados na tabela têm pouco significado estatístico, para o efeito.

Distrito/Região	Frequência	%	% Acumulada
1 - Lisboa	184	32,3	32,3
2 - Porto	144	25,3	57,5
3 - Coimbra	58	10,2	67,7
4 - Braga	7	1,2	68,9
5 - Castelo Branco	2	0,4	69,3
6 - Viana do castelo	15	2,6	71,9
7 - Viseu	16	2,8	74,7
8 - Portalegre	7	1,2	76,0
9 - Leiria	10	1,8	77,7
10 - Faro	40	7,0	84,7
11 - Évora	13	2,3	87,0
12 - Beja	10	1,8	88,8
13 - Aveiro	21	3,7	92,5
14 - Setúbal	24	4,2	96,7
15 - Santarém	11	1,9	98,6
16 - Vila Real	4	0,7	99,3
17 - Guarda	3	0,5	99,8
18 - Açores	1	0,2	100,0
Total	570	100,0	

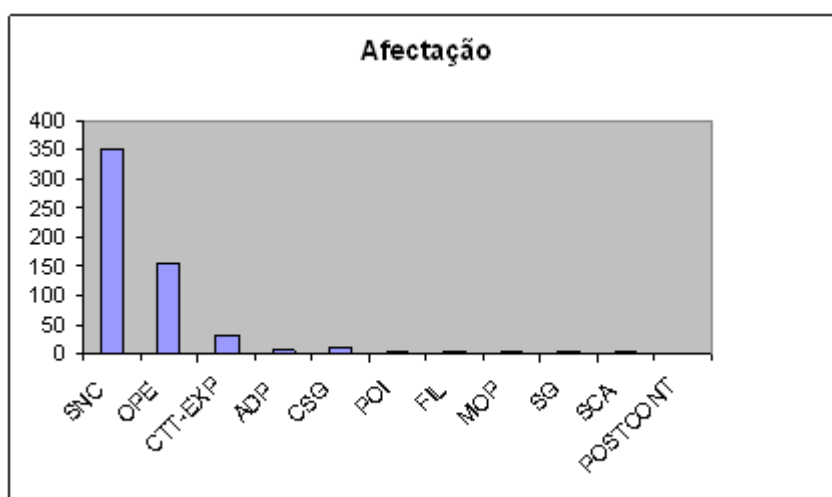


(Nota: Por uma questão de espaço, as barras não assinaladas definem os distritos de acordo com a ordem apresentada na tabela anterior)

Afectação:

A tabela e o gráfico seguintes mostram-nos como se encontram distribuídos os adultos em termos de afectação de desempenho das suas tarefas nos CTT.

Afectação	Frequência	%	% Acumulada
1 - SNC	351	61,6	61,6
2 – OPE	157	27,5	89,1
3 – CTT-EXP	31	5,4	94,6
4 – ADP	7	1,2	95,8
5 – CSG	12	2,1	97,9
6 – POI	3	0,5	98,4
7 – FIL	2	0,4	98,8
8 – MOP	2	0,4	99,1
9 – SG	2	0,4	99,5
10 – SCA	2	0,4	99,8
11 - POSTCONT	1	0,2	100,0
Total	570	100,0	



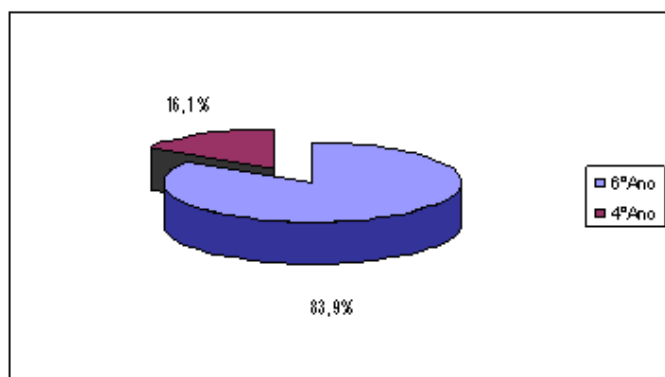
Assim, o SNC, OPE e CTT-EXP, são responsáveis por 94,6% dos adultos à certificação de competências. Como se pode verificar pelo excerto da tabela, estes são os serviços donde provêm mais candidatos. Existe uma razão: encontram-se colocados nestes serviços a maioria dos carteiros e distribuidores.

Afectação	Frequência	%	% Acumulada
1 - SNC	351	61,6	61,6
2 – OPE	157	27,5	89,1
3 – CTT-EXP	31	5,4	94,6

Habilitações Literárias:

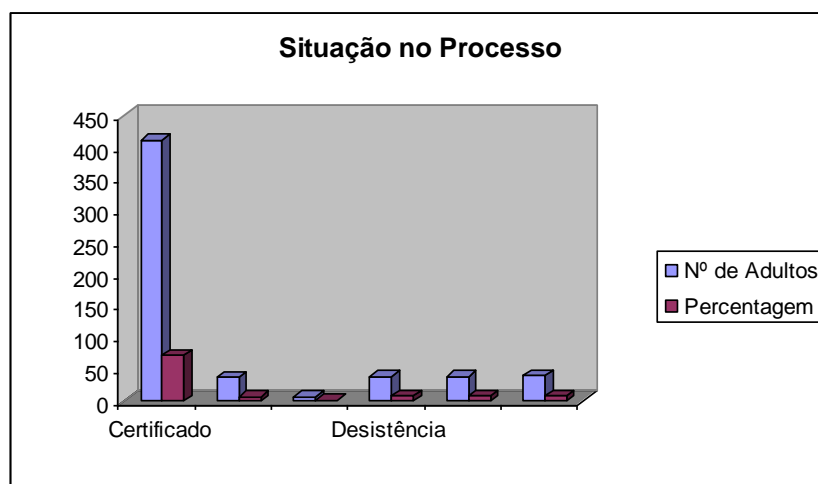
Verifica-se que a maioria dos adultos possui o 2º Ciclo do Ensino Básico (6º Ano), representando 83,9% da população. Apenas 92 adultos possuem o 1º Ciclo do Ensino Básico (4º Ano) conforme tabela e gráfico seguintes:

Hab. Lit.	Frequência	%	% Acumulada
1 - 6º Ano	478	83,9	83,9
2 - 4º Ano	92	16,1	100,0
Total	570	100	

Situação no processo:

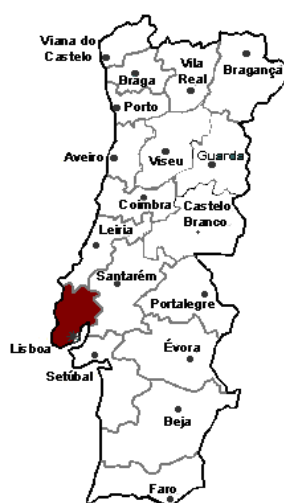
A análise da tabela e gráfico seguintes revela-nos que 412 candidatos, representando 72,3% do total da população, concluíram o processo, ficando assim, certificados; 37 adultos foram transferidos (6,5%) para outros CNO. Foram encaminhados para outras ofertas formativas, 6 candidatos, ou seja, 1,1% desta população. Mantêm-se em processo de reconhecimento 38 candidatos, representando 6,7% do total; igual situação para os adultos que desistiram do processo. Verifica-se ainda que se encontram inscritos 39 adultos que ainda não iniciaram o processo, representando estes, 6,8% da população em causa.

Situação	Frequência	%	% Acumulada
1 - Certificado	412	72,3	72,3
2 - Transferido	37	6,5	78,8
3 - Encaminhado	6	1,1	79,8
4 - Desistência	38	6,7	86,5
5 - Em reconhecimento	38	6,7	93,2
6 - Inscrito	39	6,8	100,0
Total	570	100,0	



(Análise descritiva de zonas seleccionadas)

4.2. Análise descritiva: Zona de Lisboa



Como dissemos, a análise irá incidir apenas sobre cinco categorias profissionais, ficando todas as outras incluídas em “Outros” uma vez que a sua expressão quantitativa não justifica a sua análise detalhada. Por outro lado, considerámos também, que seria relevante, o estudo desta população afecta aos três serviços dos CTT, com maior expressão: SNC, OPE e CTT-EXP.

Assim, a amostra é constituída por 184 indivíduos, dos quais 137 concluíram o processo e se apresentam as respectivas tabelas de contingência. Passemos, então, à respectiva análise.

Classe Etária	Duração				
	1 Trim	2 Trim	3 Trim	>6 Trim	Total
Até 30 anos	1	2	2	0	5
30 - 35	5	14	3	0	22
35 - 40	6	21	7	1	35
40 - 45	8	17	8	0	33
45 - 50	5	21	6	0	32
50 - 55	0	6	1	0	7
55 - 60	1	2	0	0	3
Total	26	83	27	1	137

4.2.1 Lisboa: Relação Idade/Duração do Processo

Relativamente ao binómio Idade/Duração do Processo, e para verificação da independência das variáveis, procedemos à aplicação do teste do χ^2 .

As hipóteses formuladas foram as seguintes:

H_0 : Há independência entre as variáveis “Idade” e “Duração do Processo”

H_1 : Não há independência entre as variáveis “Idade” e “Duração do Processo”

e o teste forneceu os seguintes resultados:

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
ClasseEtária * Duração	137	74,5%	47	25,5%	184	100,0%

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	9,696 ^a	18	,941
Likelihood Ratio	11,115	18	,889
Linear-by-Linear Association	,078	1	,780
N of Valid Cases	137		

a. 18 cells (64,3%) have expected count less than 5. The minimum expected count is ,02

O valor observado do teste do Qui-quadrado de Pearson é 9,696. Como o valor de prova (p-value) é 0,941, não se rejeita a hipótese nula ao nível de significância de 5%.

Através da análise dos dados nas tabelas seguintes, constatamos que:

Duração do Processo

Descrição	Frequência	Percentagem Validada	Percentagem Acumulada
1 Trimestre	26	19,0	19,0
2 Trimestres	83	60,6	79,6
3 Trimestres	27	19,7	99,3
4 Trimestres	-	-	99,3
5 Trimestres	-	-	99,3
6 Trimestres	-	-	99,3
Mais de 6 trimestres	1	,7	100,0
Total	137	100,0	-
Não validados	47	-	-
Total	184	-	-

Idade

Descrição	Frequência	Percentagem	Percentagem Acumulada
Até 30 anos	6	3,3	3,3
30 - 35	30	16,3	19,6
35 - 40	45	24,5	44,0
40 - 45	39	21,2	65,2
45 - 50	41	22,3	87,5
50 - 55	17	9,2	96,7
55 - 60	6	3,3	100,0
Total	184	100,0	-

Mapa de análise descritiva

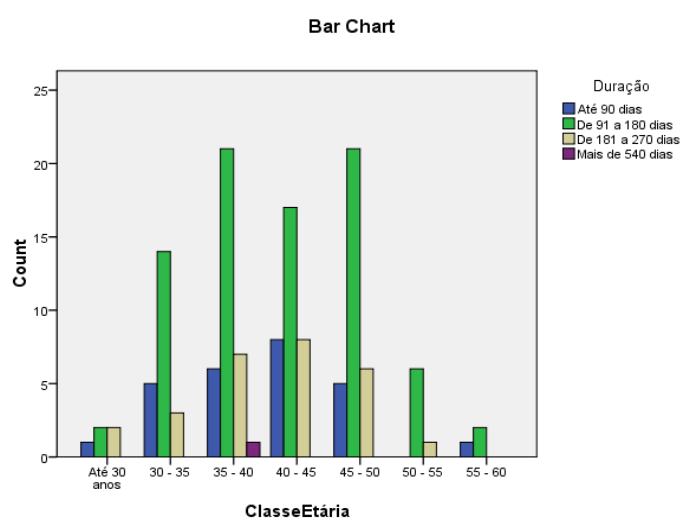
Descrição	Idade	Duração
Validados	184	137
Não validados	0	47
Média	3,8370	2,0438
Mediana	4,0000	2,0000
Moda	3,00	2,00
Desvio padrão	1,43551	,75604
Variância	2,061	,572
Assimetria	,144	1,999
Erro padrão de Assimetria	,179	,207
Achatamento (Curtose)	-,656	12,381
Erro de curtose	,356	,411
Mínimo	1,00	1,00
Máximo	7,00	7,00
Percentil 25	3,0000	2,0000
Percentil 50	4,0000	2,0000
Percentil 75	5,0000	2,0000

A grande maioria dos adultos (45%), demora cerca de quatro meses e meio a concluir o processo. Por outro lado, 14,7% dos adultos demora cerca de sete meses e meio, enquanto que 14% o fazem em menos de três meses. Por outro lado, cada adulto demora, em média, cerca de seis meses a concluir o processo. Verifica-se, também, que a mediana e a moda assumem o mesmo valor. Regista-se o facto de, em termos percentílicos, quer o P_{25} , P_{50} e o P_{75} assumirem

os mesmos valores, já que se trata de uma distribuição simétrica. Relativamente à idade, verifica-se que 68% dos adultos têm idades compreendidas entre os 35 e os 50 anos. Curiosamente, verifica-se também, que o número de adultos que frequentaram o processo e que têm menos de 30 anos é igual aos que têm mais de 55anos e representa um valor residual, apenas 3,3%.

Pode igualmente afirmar-se que a média de idades dos adultos é de 40 anos, a moda, ou seja, a maioria dos adultos tem cerca de 37 anos e meio e a mediana assume o valor de 42 anos e meio, isto é, 50% dos adultos têm idades abaixo e acima daquele valor.

Para ilustrar o que se acabou de afirmar, apresenta-se o gráfico de barras relativo às variáveis analisadas:



Procedeu-se, também à análise da correlação e regressão linear destas duas variáveis, tendo-se obtido os seguintes valores:

		Classe Etária	Duração
Classe Etária	Correlação de Pearson	1,000	-,024
	Sig. (2-tailed)		,782
	N	184,000	137
Duração	Correlação de Pearson	-,024	1,000
	Sig. (2-tailed)	,782	
	N	137	137,000

A correlação entre a Idade e a Duração do Processo não é significativa a 5%. Mesmo assim, é uma correlação negativa, o que poderá indicar que os adultos com mais idade são os que concretizam este processo em menos tempo.

O resultado da regressão apresenta-se no quadro seguinte:

Modelo		Coeficientes não Standardizados		Coeficientes Standardizados	t	Sig.
		B	Erro Standard	Beta		
1	Constante	2,094	0,192		10,920	0,000
	Classe Etária	-0,013	0,049	-0,024	-0,278	0,782

donde, se pode inferir, que a equação da recta que poderá traduzir essa regressão é dada por:

$$\hat{y} = -0,013 x + 2,094$$

4.2.2 Lisboa: Relação Habilitações Literárias/Duração do Processo

Para efectuar a análise da correlação entre as variáveis “Habilitações Literárias” e “Duração do Processo”, baseámo-nos na tabela de contingência seguinte:

Duração	Habilitações Literárias		
	2º Ciclo do Ensino Básico	1º Ciclo do Ensino Básico	Total
1 Trim.	22	4	26
2 Trim.	77	6	83
3 Trim.	24	3	27
> 6 Trim.	0	1	1
Total	123	14	137

Relativamente ao binómio Habilitações Literárias/Duração do Processo, e para verificação da independência das variáveis, procedemos à aplicação do teste do χ^2 .

As hipóteses formuladas foram as seguintes:

H_0 : Há independência entre as variáveis “Habilitações Literárias” e “Duração do Processo”

H_1 : Não há independência entre as variáveis “Habilitações Literárias” e “Duração do Processo”

e o teste forneceu os seguintes resultados:

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
HabilitaçõesLit * Duração	137	74,5%	47	25,5%	184	100,0%

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10,374 ^a	3	,016
Likelihood Ratio	6,142	3	,105
Linear-by-Linear Association	1,597	1	,206
N of Valid Cases	137		

a. 4 cells (50,0%) have expected count less than 5. The minimum expected count is ,10.

O valor observado do teste do Qui-quadrado de Pearson é 10,374 sendo o valor de prova de 0,016, portanto, inferior a 0,05. Há evidência estatística para rejeitar a hipótese nula.

A análise da correlação levou-nos aos seguintes resultados:

		Habilit. Literárias	Duração
Duração	Pearson Correlation	1,000	,108
	Sig. (2-tailed)		,208
	N	137,000	137
Habilit. Literárias	Pearson Correlation	,108	1,000
	Sig. (2-tailed)	,208	
	N	137	184,000

A correlação entre as variáveis “Habilitações Literárias” e a Duração do Processo não é significativa a 5%. Trata-se de uma correlação positiva, embora mínima. No entanto, podemos afirmar que os adultos com mais habilitações são os que necessitam de mais tempo para concluir o processo.

O resultado da regressão apresenta-se no quadro seguinte:

Modelo		Coeficientes não Standardizados		Coeficientes Standardizados	t	Sig.
		B	Erro Standard	Beta		
1	Constante	1,747	,243		7,182	,000
	Habilit. Lit.	,269	,213	,108	1,266	,208

Donde, se pode inferir que a equação da recta que melhor traduz essa regressão é dada por:

$$\hat{y} = 0,269x + 1,747$$

4.3 Análise descritiva: Zona do Porto



De igual modo, nesta região, a análise irá incidir apenas sobre cinco categorias profissionais, ficando todas as outras incluídas em “Outros” uma vez que a sua expressão quantitativa não justifica a sua análise detalhada. À semelhança do que se fez em Lisboa, considerou-se também, que seria relevante, o estudo desta população afecta aos três serviços dos CTT, com maior expressão: SNC, OPE e CTT-EXP.

Assim, a amostra é constituída por 144 indivíduos, dos quais 114 concluíram o processo como se mostra no quadro de contingência seguinte:

Classe Etária	Duração				
	2 Trim	3 Trim	4 Trim	> 6 Trim	Total
Até 30 anos	3	2	0	0	5
30 - 35	13	6	2	1	22
35 - 40	14	2	7	2	25
40 - 45	16	1	0	3	20
45 - 50	24	0	0	0	24
50 - 55	11	1	2	0	14
55 - 60	4	0	0	0	4
Total	85	12	11	6	114

4.3.1 Porto: Relação Idade/Duração do Processo

Relativamente ao binómio Idade/Duração do Processo, e para verificação da independência das variáveis, procedemos à aplicação do teste do χ^2 .

As hipóteses formuladas foram as seguintes:

H_0 : Há independência entre as variáveis “Idade” e “Duração do Processo”

H_1 : Não há independência entre as variáveis “Idade” e “Duração do Processo”

e o teste forneceu os seguintes resultados:

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Duração * ClasseEtária	114	79,2%	30	20,8%	144	100,0%

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	38,974 ^a	18	,003
Likelihood Ratio	41,792	18	,001
Linear-by-Linear Association	4,083	1	,043
N of Valid Cases	114		

a. 23 cells (82,1%) have expected count less than 5. The minimum expected count is ,21.

O valor do teste do Qui-quadrado de Pearson é 38,974. Considerando um nível de significância de 5%, dado que o valor de prova é inferior (p-value=0,003), deve rejeitar-se a hipótese nula, concluindo-se que as variáveis não são independentes.

Relativamente ao binómio Idade/Duração do Processo, constata-se, através da análise dos dados nas tabelas seguintes, que:

Duração do Processo

Descrição	Frequência	Percentagem Validada	Percentagem Acumulada
Até 90 dias	-	-	-
De 91 a 180 dias	85	74,6	74,6
De 181 a 270 dias	12	10,5	85,1
De 271 a 360 dias	11	9,6	94,7
De 361 a 450 dias	-	-	94,7
De 451 a 540 dias	-	-	94,7
Mais de 540 dias	6	5,3	100,0
Total	114	100,0	
Não validados	30		
Total	144		

Idade

Descrição	Frequência	Percentagem	Percentagem Acumulada
Até 30 anos	6	4,2	4,2
30 - 35	28	19,4	23,6
35 - 40	33	22,9	46,5
40 - 45	27	18,8	65,3
45 - 50	27	18,8	84,0
50 - 55	19	13,2	97,2
55 - 60	4	2,8	100,0
Total	144	100,0	

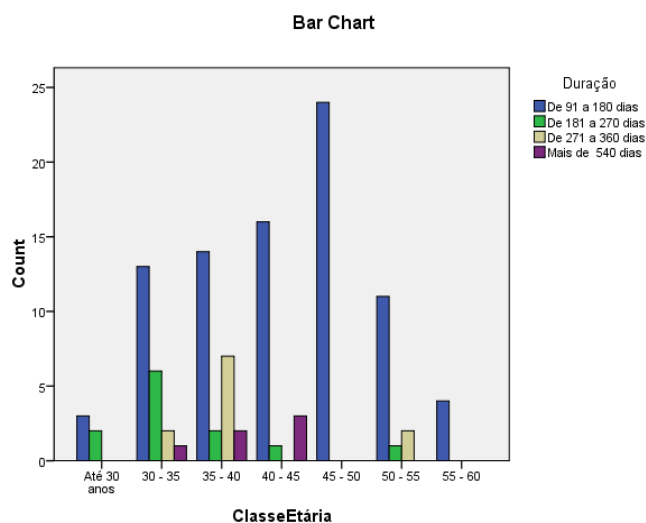
Mapa de análise descritiva

Descrição	Idade	Duração
Validados	144	114
Não validados	0	30
Média	3,7917	2,5614
Mediana	4,0000	2,0000
Moda	3,00	2,00
Desvio padrão	1,52313	1,22680
Variância	2,320	1,505
Assimetria	,154	2,651
Erro padrão de Assimetria	,202	,226
Achatamento (Curtose)	-,897	6,744
Erro de curtose	,401	,449
Mínimo	1,00	2,00
Máximo	7,00	7,00
Soma	3,0000	2,0000
Percentil 25	4,0000	2,0000
Percentil 50	5,0000	3,0000
Percentil 75	144	114

A grande maioria dos adultos (59%), demora cerca de quatro meses e meio a concluir o processo. Por outro lado, 8,3% dos adultos demora cerca de sete meses e meio, enquanto que 7,6% o fazem em dez meses e meio. Por outro lado, cada adulto demora, em média, cerca de sete meses e meio a concluir o processo. Verifica-se, também, que a mediana e a moda assumem o mesmo valor. Podemos assim afirmar que, 50% dos adultos demora menos de 6 meses a concluir o processo e a mesma percentagem o faz num período superior.

Relativamente à idade, verifica-se que 60% dos adultos têm idades compreendidas entre os 35 e os 50 anos. Curiosamente, verifica-se também, que o número de adultos que frequentaram o processo e que têm menos de 30 anos é de 4,2% e os que têm mais de 55 anos é de apenas 2,8%. Podemos também afirmar que a média de idades dos adultos é de 40 anos, a moda, ou seja, a maioria dos adultos tem cerca de 37 anos e meio e a mediana assume o valor de 42 anos e meio, isto é, 50% dos adultos têm idades abaixo e acima daquele valor.

Para ilustrar o que se acabou de afirmar, apresentamos o gráfico de barras relativo às variáveis analisadas:



Procedemos, também à análise da correlação e regressão linear destas duas variáveis, tendo-se obtido os seguintes valores:

		Classe Etária	Duração
Classe Etária	Correlação de Pearson	1,000	-,190 ⁺
	Sig. (2-tailed)		,043
	N	144,000	114
Duração	Correlação de Pearson	-,190 ⁺	1,000
	Sig. (2-tailed)	,043	
	N	114	114,000

A correlação entre a Idade e a Duração do Processo não é significativa a 5%. Mesmo assim, trata-se de uma correlação negativa, o que poderá indicar que os adultos com mais idade são os que concretizam este processo em menos tempo.

O resultado da regressão apresenta-se no quadro seguinte:

Modelo		Coeficientes não Standardizados		Coeficientes Standardizados	t	Sig.
		B	Erro Standard	Beta		
1	Constante	3,138	0,303		10,346	0,000
	Classe Etária	-0,151	0,074	-0,190	-2,049	0,043

donde, se pode inferir que a equação da recta que pode traduzir essa regressão é dada por:

$$\hat{y} = -0,151 x + 3,138$$

4.3.2 Porto: Relação Habilitações Literárias/Duração do Processo

Para efectuar a análise da correlação entre as variáveis “Habilitações Literárias” e “Duração do Processo”, baseámo-nos na tabela de contingência seguinte:

Duração	Habilitações Literárias		
	2º Ciclo do Ensino Básico	1º Ciclo do Ensino Básico	Total
2 Trim.	65	20	85
3 Trim.	12	0	12
4 Trim.	11	0	11
> 6 Trim.	6	0	6
Total	94	20	114

Também aqui aplicámos o teste do χ^2 para testar a independência das variáveis.

As hipóteses formuladas foram as seguintes:

H_0 : Há independência entre as variáveis “Habilitações Literárias” e “Duração do Processo”

H_1 : Não há independência entre as variáveis “Habilitações Literárias” e “Duração do Processo”

e o teste forneceu os seguintes resultados:

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Duração * HabilitaçõesLit	114	79,2%	30	20,8%	144	100,0%

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8,275 ^a	3	,041
Likelihood Ratio	13,133	3	,004
Linear-by-Linear Association	5,079	1	,024
N of Valid Cases	114		

a. 4 cells (50,0%) have expected count less than 5. The minimum expected count is 1,05.

O valor observado do teste do Qui-quadrado de Pearson é 8,275. Sendo o valor de prova (p-value=0,041) inferior a 5%, concluindo-se que deverá ser rejeitada a hipótese nula.

A análise da correlação, forneceu os seguintes resultados:

		ClasseEtária	Duração
Duração	Pearson Correlation	1,000	-,212*
	Sig. (2-tailed)		,024
	N	114,000	114
Habilit. Literárias	Pearson Correlation	-,212*	1,000
	Sig. (2-tailed)	,024	
	N	114	144,000

A correlação entre as variáveis “Habilitações Literárias” e a Duração do Processo pode considerar-se significativa a 5%. Trata-se de uma correlação negativa. Os adultos com menos habilitações são os que concluem o processo em mais tempo. O resultado da regressão apresenta-se no quadro seguinte:

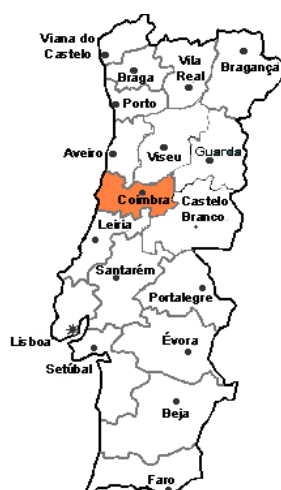
Modelo		Coeficientes não Standardizados		Coeficientes Standardizados	t	Sig.
		B	Erro Standard	Beta		
1	Constante	3,362	,366		9,176	,000
	Habilit. Lit.	-,681	,297	-,212	-2,296	,024

Donde, se pode inferir que a equação da recta que pode traduzir essa regressão é dada por:

$$\hat{y} = -0,681x + 3,362$$

em que a variável dependente é a “Duração do Processo”.

4.4 Análise descritiva: Zona de Coimbra



Tal como nas anteriores regiões, a análise irá incidir sobre as cinco categorias profissionais consideradas anteriormente, ficando todas as outras incluídas em “Outros” uma vez que a sua expressão quantitativa não justifica a sua análise detalhada. À semelhança do que se fez em Lisboa e no Porto, considerámos também, que seria relevante, o estudo desta população afecta aos três serviços dos CTT, com maior expressão: SNC, OPE e CTT-EXP.

Assim, a amostra é constituída por 58 indivíduos, dos quais 35 concluíram o processo, distribuídos, como mostramos, na seguinte tabela de contingência:

Classe Etária	Duração					
	2 Trim	3 Trim	4 Trim	5 Trim	>6 Trim	Total
Até 30 anos						
30 - 35	1	4	1	1	0	7
35 - 40	0	6	3	2	0	11
40 - 45	0	7	2	1	2	12
45 - 50	0	1	2	0	0	3
50 - 55	0	2	0	0	0	2
55 - 60						
Total	1	20	8	4	2	35

4.4.1 Coimbra: Relação Idade/Duração do Processo

Relativamente ao binómio Idade/Duração do Processo, e para verificação da independência das variáveis, procedemos à aplicação do teste do χ^2 .

As hipóteses formuladas foram as seguintes:

H_0 : Há independência entre as variáveis “Idade” e “Duração do Processo”

H_1 : Não há independência entre as variáveis “Idade” e “Duração do Processo”

e o teste forneceu os seguintes resultados:

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Duração * ClasseEtária	35	60,3%	23	39,7%	58	100,0%

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	13,447 ^a	16	,640
Likelihood Ratio	13,201	16	,658
Linear-by-Linear Association	,159	1	,690
N of Valid Cases	35		

a. 23 cells (92,0%) have expected count less than 5. The minimum expected count is ,06

O valor observado do teste do Qui-quadrado de Pearson é 13,447 sendo o valor de prova 0,640. Não se rejeita a hipótese nula de independência entre as variáveis “Idade” e “Duração do processo” ao nível de significância de 5%.

Constata-se, também, através da análise dos dados nas tabelas seguintes, que:

Duração do Processo

Descrição	Frequência	Percentagem validada	Percentagem Acumulada
Até 90 dias	-	-	-
De 91 a 180 dias	1	2,9	2,9
De 181 a 270 dias	20	57,1	60,0
De 271 a 360 dias	8	22,9	82,9
De 361 a 450 dias	4	11,4	94,3
De 451 a 540 dias	-	-	-
Mais de 540 dias	2	5,7	100,0
Total	35	100,0	-
Não validados	23	-	-
Total	58	-	-

Idade

Descrição	Frequência	Percentagem	Percentagem Acumulada
Até 30 anos	-	-	-
30 - 35	11	19,0	19,0
35 - 40	19	32,8	51,7
40 - 45	18	31,0	82,8
45 - 50	7	12,1	94,8
50 - 55	3	5,2	100,0
55 - 60	-	-	-
Total	58	100,0	-

Mapa de análise descritiva

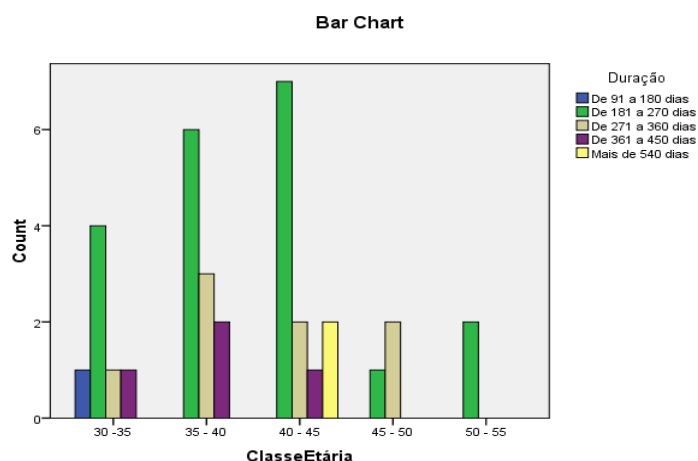
Descrição	Idade	Duração
Validados	58	35
Não validados	0	23
Média	3,5172	3,6571
Mediana	3,0000	3,0000
Moda	3,00	3,00
Desvio padrão	1,09611	1,10992
Variância	1,201	1,232
Assimetria	,410	1,699
Erro padrão de Assimetria	,314	,398
Achatamento (Curtose)	-,350	3,055
Erro de curtose	,618	,778
Mínimo	2,00	2,00
Máximo	6,00	7,00
Percentil 25	3,0000	3,0000
Percentil 50	4,0000	4,0000
Percentil 75	3,0000	3,0000

A grande maioria dos adultos (34,5%), demora cerca de sete meses e meio a concluir o processo. Por outro lado, 13,8% dos adultos demora cerca de dez meses e meio, enquanto que 14% o fazem em menos de três meses. De realçar que apenas um adulto demorou um trimestre a realizar este processo. Verifica-se, também, que a mediana e a moda assumem o mesmo valor. Regista-se o facto de, em termos percentílicos, quer o P_{25} e o P_{75} assumirem os mesmos

valores. Relativamente à idade, verifica-se que 64% dos adultos têm idades compreendidas entre os 35 e os 45 anos. Curiosamente, verifica-se que se trata de uma população mais jovem que as anteriores.

Pode igualmente afirmar-se que a média de idades dos adultos é de cerca de 40 anos, a moda e a mediana assumem os mesmos valores, ou seja, 37 anos e meio.

Para ilustrar o que se acabou de afirmar, apresenta-se o gráfico de barras relativo às variáveis analisadas:



A análise da correlação e regressão destas variáveis levou-nos aos seguintes resultados:

		ClasseEtária	Duração
ClasseEtária	Correlação de Pearson	1,000	,068
	Sig. (2-tailed)		,696
	N	58,000	35
Duração	Correlação de Pearson	,068	1,000
	Sig. (2-tailed)	,696	
	N	35	35,000

A correlação entre a Idade e a Duração do Processo não é significativa a 5%. Mesmo assim, e ao contrário das regiões de Lisboa e Porto, é uma correlação positiva, o que poderá indicar que os adultos com mais idade são os que necessitam de mais tempo para concretizar este processo.

4.4.2 Coimbra: Relação Habilitações Literárias/Duração do Processo

Para efectuar a análise da correlação entre as variáveis “Habilitações Literárias” e “Duração do Processo”, baseámo-nos na tabela de contingência seguinte:

Duração	Habilitações Literárias		
	2º Ciclo do Ensino Básico	1º Ciclo do Ensino Básico	Total
2 Trim.	1	0	1
3 Trim.	18	2	20
4 Trim.	8	0	8
5 Trim.	4	0	4
> 6 Trim.	2	0	2
Total	33	2	35

Também aqui aplicámos o teste do χ^2 para testar a independência das variáveis.

As hipóteses formuladas foram as seguintes:

H_0 : Há independência entre as variáveis “Habilitações Literárias” e “Duração do Processo”

H_1 : Não há independência entre as variáveis “Habilitações Literárias” e “Duração do Processo”

e o teste forneceu os seguintes resultados:

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Duração * HabLiterarias	35	60,3%	23	39,7%	58	100,0%

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1,591 ^a	4	,810
Likelihood Ratio	2,329	4	,676
Linear-by-Linear Association	,744	1	,389
N of Valid Cases	35		

a. 8 cells (80,0%) have expected count less than 5. The minimum expected count is ,06

O valor observado do teste do Qui-quadrado de Pearson é 1,591 com um valor de prova de 0,810 que é superior a 5%. Não se rejeita a hipótese nula de independência das variáveis “Habilitações Literárias” e “Duração do processo”.

O resultado da regressão apresenta-se no quadro seguinte:

Modelo		Coeficientes não Standardizados		Coeficientes Standardizados	t	Sig.
		B	Erro Standard	Beta		
1	Constante	3,415	0,643		5,315	0,000
	Classe Etária	0,069	0,176	0,068	0,394	0,696

Donde, se pode inferir que a equação da recta que traduz essa regressão é dada por:

$$\hat{y} = 0,069 x + 3,415$$

No entanto, e relativamente à análise da correlação e regressão linear entre as variáveis em causa, obtivemos os seguintes resultados:

		Classe Etária	Duração
Duração	Pearson Correlation	1,000	-,148
	Sig. (2-tailed)		,397
	N	35,000	35
Habilit. Lit.	Pearson Correlation	-,148	1,000
	Sig. (2-tailed)	,397	
	N	35	58,000

A correlação entre as variáveis em causa pode considerar-se significativa a 5%. Na região de Coimbra, assistimos, mais uma vez, a uma correlação negativa, indicando que os adultos com menos habilitações literárias são os que concluem o processo em menos tempo.

O resultado da regressão apresenta-se no quadro seguinte:

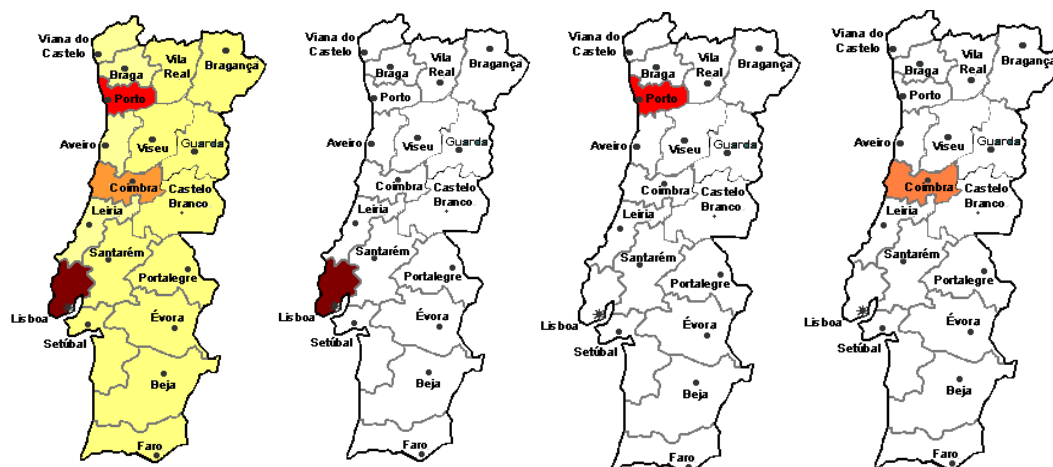
Modelo		Coeficientes não Standardizados		Coeficientes Standardizados	t	Sig.
		B	Erro Standard	Beta		
1	Constante	4,394	,878		5,003	,000
	Hábil. Lit.	-,697	,811	-,148	-,859	,397

Donde, se pode inferir que a equação da recta que poderá traduzir essa regressão é dada por:

$$\hat{y} = -0,697x + 4,394$$

em que a variável dependente é a “Duração do Processo”.

4.5 Análise comparativa entre as regiões de Lisboa, Porto e Coimbra versus totalidade do País



Como referimos, a análise irá incidir sobre as cinco categorias profissionais, ficando todas as outras incluídas em “Outros” uma vez que a sua expressão quantitativa não justifica a sua análise detalhada. De igual modo, o estudo desta população refere-se aos três serviços dos CTT, com maior expressão: SNC, OPE e CTT-EXP.

Assim, a amostra é constituída por 570 indivíduos, dos quais 419 concluíram o processo. Passemos, então, à respectiva análise.

4.5.1 Totalidade do País: Relação Idade/Duração do Processo

Para efectuar a análise da correlação entre as variáveis “Idade” e “Duração do Processo”, baseámo-nos na tabela de contingência seguinte:

Classe Etária * Duração								
ClasseEtária	Duração (Trimestres)							
	Até 1	De 1 a 2	De 2 a 3	De 3 a 4	De 4 a 5	De 5 a 6	> 6	Total
Até 30 anos	1	1	4	2	0	1	0	9
30 - 35	5	38	19	5	5	6	1	79
35 - 40	6	48	22	22	2	1	5	106
40 - 45	7	40	20	21	3	5	4	100
45- 50	4	41	20	10	3	1	1	80
50 - 55	2	17	3	10	0	1	1	34
55 - 60	1	6	3	0	0	0	0	10
Mais de 60 anos	0	0	1	0	0	0	0	1
Total	26	191	92	70	13	15	12	419

Relativamente ao binómio Idade/Duração do Processo, e para verificação da independência das variáveis, procedemos à aplicação do teste do χ^2 .

As hipóteses formuladas foram as seguintes:

H_0 : Há independência entre as variáveis “Idade” e “Duração do Processo”

H_1 : Não há independência entre as variáveis “Idade” e “Duração do Processo”

e o teste forneceu os seguintes resultados:

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Classe Etária * Duração	419	73,5%	151	26,5%	570	100,0%

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	45,595 ^a	42	,325
Likelihood Ratio	50,756	42	,167
Linear-by-Linear Association	1,445	1	,229
N of Valid Cases	419		

a. 39 cells (69,6%) have expected count less than 5. The minimum expected count is ,03.

O valor observado do teste do Qui-quadrado de Pearson é 45,595 sendo o valor de prova 0,325. Como se trata de um valor superior a 5%, não se rejeita a hipótese nula de independência entre as variáveis “Idade” e “Duração do Processo”.

Relativamente ao binómio Idade/Duração do Processo, constatámos, através da análise dos dados nas tabelas seguintes, que:

Duração do Processo				Idade			
Descrição	Frequência	Percentagem Validada	Percentagem Acumulada	Descrição	Frequência	Percentagem Validada	Percentagem Acumulada
Até 90 dias	26	6,2	6,2	Até 30 anos	13	2,3	2,3
De 91 a 180 dias	191	45,6	51,8	30 - 35	106	18,6	20,9
De 181 a 270 dias	92	22,0	73,7	35 - 40	146	25,6	46,5
De 271 a 360 dias	70	16,7	90,5	40 - 45	134	23,5	70,0
De 361 a 450 dias	13	3,1	93,6	45 - 50	104	18,2	88,2
De 451 a 540 dias	15	3,6	97,1	50 - 55	51	8,9	97,2
Mais de 540 dias	12	2,9	100,0	55 - 60	15	2,6	99,8
Total	419	100,0		60 - 65	1	,2	100,0
Não validados	151			Não Validados			
Total	570			Total	570	100,0	

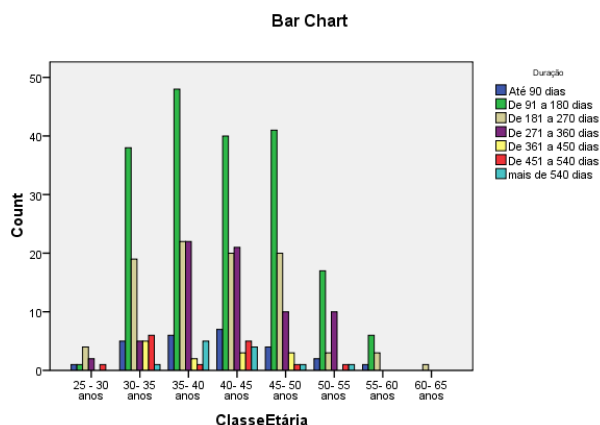
Mapa de Análise Estatística

Descrição	Idade	Duração
Validados	570	419
Não validados	0	151
Média	3,7509	2,8711
Mediana	4,0000	2,0000
Moda	3,00	2,00
Desvio padrão	1,39710	1,32788
Variância	1,952	1,763
Assimetria	,307	1,267
Erro padrão de Assimetria	,102	,119
Achatamento (Curtose)	-,525	1,455
Erro de curtose	,204	,238
Mínimo	1,00	1,00
Máximo	8,00	7,00
Percentil 25	3,0000	2,0000
Percentil 50	4,0000	2,0000
Percentil 75	5,0000	4,0000

A grande maioria dos adultos (45,6%), demora cerca de quatro meses e meio a concluir o processo. Por outro lado, 9,6% dos adultos demora mais de ano e meio, enquanto que 6,2% o fazem em menos de três meses. Verifica-se, também, que a mediana e a moda assumem o mesmo valor. Relativamente à idade, verifica-se que 49,1% dos adultos (cerca de metade) têm idades compreendidas entre os 35 e os 45 anos. A análise dos dados destas três regiões também permite afirmar que 70% dos adultos têm menos de 45 anos e que 88,2% têm menos de 50 anos de idade.

Podemos igualmente afirmar que a média de idades dos adultos é de cerca de 44 anos, a moda é de 37 anos e meio e a mediana o valor de 42 anos e meio.

Para ilustrar o que se acabou de afirmar, apresenta-se o gráfico de barras relativo às variáveis analisadas:



4.5.2 Totalidade do País: Relação Habilitações Literárias/Duração do Processo

Para efectuar a análise da correlação entre as variáveis “Habilitações Literárias” e “Duração do Processo”, baseámo-nos na tabela de contingência seguinte:

Duração	Habilitações Literárias		
	2º Ciclo do Ensino Básico	1º Ciclo do Ensino Básico	Total
Até 1 Trim.	21	5	26
2 Trim.	164	27	191
3 Trim.	82	10	92
4 Trim.	57	13	70
5 Trim.	12	1	13
6 Trim.	13	2	15
> 6 Trim.	10	2	12
Total	359	60	419

Também aqui aplicámos o teste do χ^2 para testar a independência das variáveis. As hipóteses formuladas foram as seguintes:

H_0 : Há independência entre as variáveis “Habilitações Literárias” e “Duração do Processo”

H_1 : Não há independência entre as variáveis “Habilitações Literárias” e “Duração do Processo”

e o teste forneceu os seguintes resultados:

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
HabLit * Duração	419	73,5%	151	26,5%	570	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2,971 ^a	6	,812
Likelihood Ratio	3,002	6	,809
Linear-by-Linear Association	,001	1	,978
N of Valid Cases	419		

a 4 cells (28,6%) have expected count less than 5. The minimum expected count is 1,72.

O valor observado do teste do Qui-quadrado de Pearson é 2,971, sendo o valor de prova 0,812 superior a 5%. Não se rejeita a hipótese nula de independência entre as variáveis “Habilitações Literárias” e “Duração do processo”.

No entanto, e relativamente à análise da correlação e regressão linear entre as variáveis em causa, obtivemos os seguintes resultados:

		Habilit. Lit.	Duração
Duração	Pearson Correlation	1,000	-,001
	Sig. (2-tailed)		,978
	N	570,000	419
Habilit. Lit.	Pearson Correlation	-,001	1,000
	Sig. (2-tailed)	,978	
	N	419	419,000

A correlação entre as variáveis em causa é significativa a 5%. No País e apesar da correlação ser significativa, assistimos, mais uma vez, a uma correlação negativa, indicando que os adultos com menos habilitações literárias são os que concluem o processo em menos tempo.

O resultado da regressão apresenta-se no quadro seguinte:

Modelo		Coeficientes não Standardizados		Coeficientes Standardizados	t	Sig.
		B	Erro Standard	Beta		
1	Constante	2,877	,222		12,977	,000
	Hábil. Lit.	-,005	,185	-,001	-,028	,978

Donde, se pode inferir que a equação da recta que traduz essa regressão é dada por:

$$\hat{y} = -0,005x + 2,877$$

em que a variável dependente é a “Duração do Processo”.

4.5.3 Totalidade do País: Análise comparativa

Como podemos constatar nos quadros seguintes, apenas na região de Lisboa existem adultos que concluíram o processo em menos de 90 dias. A maioria dos adultos, nas regiões de Lisboa e Porto, concretizam o processo entre três a seis meses, conclusão idêntica considerando o país. Por outro lado, e apesar da população da região de Coimbra ser a mais jovem, é a região que

apresenta um período mais longo para a maioria dos adultos concretizarem o processo, aliás, conclusão verificada já na análise anterior.

Análise Comparativa da Duração do Processo

Descrição	Frequência				Percentagem Validada			
	País	Lisboa	Porto	Coimbra	País	Lisboa	Porto	Coimbra
Até 90 dias	26	26	-	-	6,2	19,0	-	-
De 91 a 180 dias	191	83	85	1	45,6	60,6	74,6	2,9
De 181 a 270 dias	92	27	12	20	22,0	19,7	10,5	57,1
De 271 a 360 dias	70	-	11	8	16,7	-	9,6	22,9
De 361 a 450 dias	13	-	-	4	3,1	-	-	11,4
De 451 a 540 dias	15	-	-	-	3,6	-	-	-
Mais de 540 dias	12	1	6	2	2,9	,7	5,3	5,7
Total	419	137	114	35	100,0	100,0	100,0	100,0
Não validados	151	47	30	23		-	-	-
Total	570	184	144	58		-	-	-

Análise Comparativa da Idade

Descrição	Frequência				Percentagem			
	País	Lisboa	Porto	Coimbra	País	Lisboa	Porto	Coimbra
Até 30 anos	13	6	6		2,3	3,3	4,2	
30 - 35	106	30	28	11	18,6	16,3	19,4	19,0
35 - 40	146	45	33	19	25,6	24,5	22,9	32,8
40 - 45	134	39	27	18	23,5	21,2	18,8	31,0
45 - 50	104	41	27	7	18,2	22,3	18,8	12,1
50 - 55	51	17	19	3	8,9	9,2	13,2	5,2
55 - 60	15	6	4		2,6	3,3	2,8	
60 - 65	1				,2			
Total	570	184	144	58	100,0	100,0	100,0	100,0
Não validados								
Total	570	184	144	58	100,0	100,0	100,0	100,0

Feita a análise comparativa das estatísticas descritivas podemos concluir que, relativamente às idades dos adultos:

- As médias das idades dos adultos nas regiões de Lisboa e Porto são superiores à média verificada a nível nacional. A média verificada em Coimbra é inferior á média nacional. Como dissemos, é a população mais jovem.
- A mediana das idades dos adultos nas regiões de Lisboa e Porto são idênticas e iguais à mediana verificada considerando a totalidade da população (42 anos e meio). A mediana das idades na região de Coimbra é um pouco inferior (40 anos).
- A moda das idades dos adultos é igual em todas as regiões e à moda verificada nas três regiões (37 anos e meio).

- Trata-se de distribuições assimétricas positivas, embora a distribuição de Coimbra apresente um desvio padrão inferior o que se traduz numa menor diferença de idades entre os adultos.

Relativamente ao tempo de duração do processo, verifica-se que:

- As médias da duração do processo nas regiões de Lisboa (94 dias), Porto (140 dias) e Coimbra (329 dias) são distintas da média do país (cerca de 258 dias), verificando-se ainda que os adultos na região de Coimbra são os que demoraram mais tempo a concluir o processo.

- A mediana e a moda da duração do processo nas regiões de Lisboa e Porto são idênticas à verificada no país (135 dias). Na região de Coimbra, a mediana e a moda verificada foi de 225 dias.

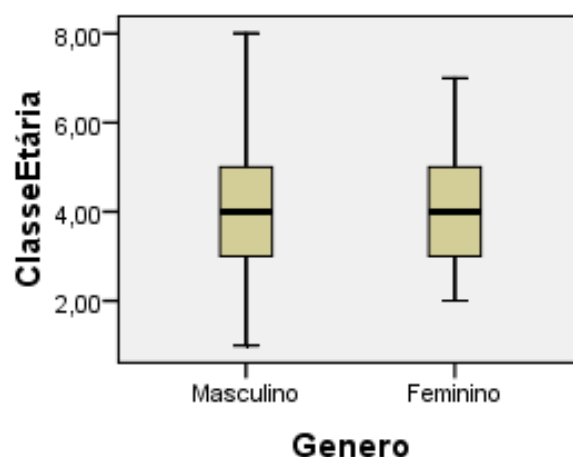
- Podemos também afirmar que, quanto à assimetria e relativamente à duração do processo, a distribuição é simétrica, quando considerada a população total (país).

Mapa Comparativo de análise descritiva

Descrição	Idade				Duração			
	País	Lisboa	Porto	Coimbra	País	Lisboa	Porto	Coimbra
Validados	570	184	144	58	419	137	114	35
Não validados	0	0	0	0	151	47	30	23
Média	3,7509	3,8370	3,7917	3,5172	2,8711	2,0438	2,5614	3,6571
Mediana	4,0000	4,0000	4,0000	3,0000	2,0000	2,0000	2,0000	3,0000
Moda	3,00	3,00	3,00	3,00	2,00	2,00	2,00	3,00
Desvio padrão	1,39710	1,43551	1,52313	1,09611	1,32788	,75604	1,22680	1,10992
Variância	1,952	2,061	2,320	1,201	1,763	,572	1,505	1,232
Assimetria	,307	,144	,154	,410	1,267	1,999	2,651	1,699
Erro padrão de Assimetria	,102	,179	,202	,314	,119	,207	,226	,398
Achatamento (Curtose)	-,525	-,656	-,897	-,350	1,455	12,381	6,744	3,055
Erro de curtose	,204	,356	,401	,618	,238	,411	,449	,778
Mínimo	1,00	1,00	1,00	2,00	1,00	1,00	2,00	2,00
Máximo	8,00	7,00	7,00	6,00	7,00	7,00	7,00	7,00
Percentil 25	3,0000	3,0000	3,0000	3,0000	2,0000	2,0000	2,0000	3,0000
Percentil 50	4,0000	4,0000	4,0000	4,0000	2,0000	2,0000	2,0000	3,0000
Percentil 75	5,0000	5,0000	5,0000	4,0000	4,0000	2,0000	3,0000	4,0000

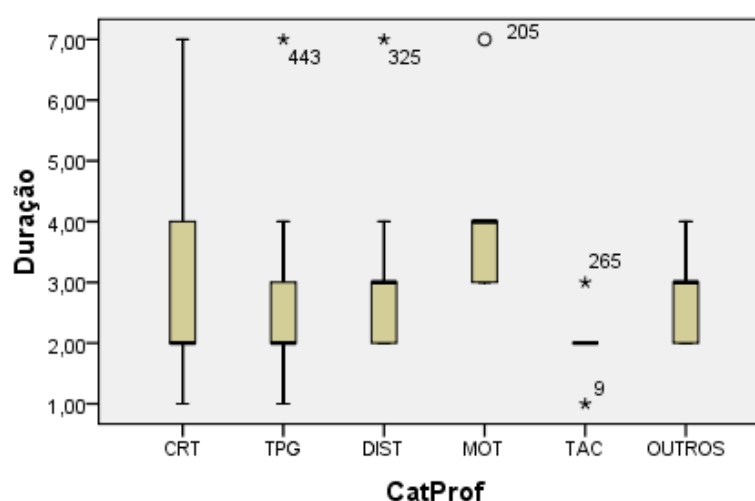
Verifica-se também, na análise do país, que a maioria dos adultos que frequentaram este processo são do sexo masculino embora as faixas etárias dos adultos sejam semelhantes, nos dois sexos:

Case Processing Summary							
		Cases					
		Valid		Valid		Valid	
		N	Percent	N	Percent	N	N
ClasseEtária	Masculino	528	100,0%	0	,0%	528	100,0%
	Feminino	42	100,0%	0	,0%	42	100,0%



Verifica-se também que a duração do processo não depende das categorias profissionais dos trabalhadores. A análise dos resultados mostra-nos que a maioria demora entre 180 e 270 dias a concluir o processo. Verifica-se também que, no caso dos carteiros a duração do processo foi mais homogénea. Nas outras categorias verificam-se algumas discrepâncias.

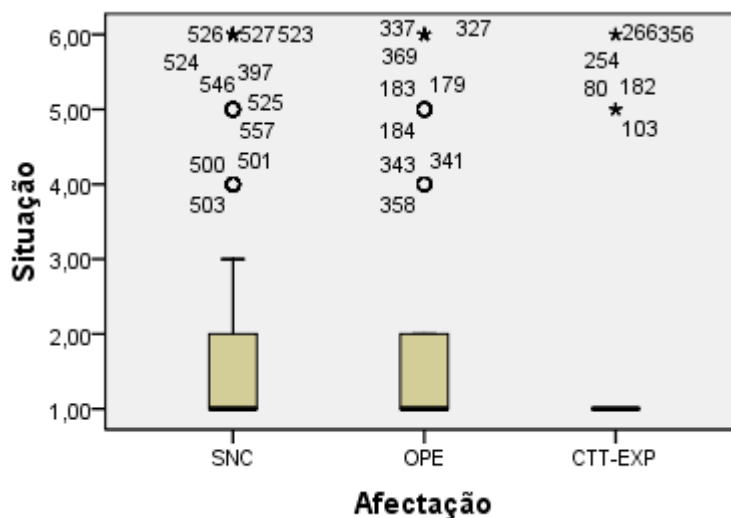
Case Processing Summary							
Categoria Profissional		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Duração	CRT	358	73,4%	130	26,6%	488	100,0%
	TPG	22	84,6%	4	15,4%	26	100,0%
	DIST	17	70,8%	7	29,2%	24	100,0%
	MOT	5	71,4%	2	28,6%	7	100,0%
	TAC	7	70,0%	3	30,0%	10	100,0%
	OUTROS	10	66,7%	5	33,3%	15	100,0%



A análise da situação no processo leva-nos a afirmar que:

- dos trabalhadores afectos ao SNC, apenas 3 desistiram, 5 encontram-se em reconhecimento e 3 continuam inscritos sem iniciar o processo. Todos os outros concluíram o processo, embora alguns transferidos ou encaminhados.
- dos trabalhadores afectos à OPE verifica-se que 3 trabalhadores desistiram, 3 encontram-se em reconhecimento e 3 encontram-se inscritos a aguardar o início do processo; os restantes trabalhadores concluíram o processo e alguns foram transferidos.
- os trabalhadores dos CTT-EXP concluíram o processo com excepção de 6, encontrando-se 3 em reconhecimento e 3 inscritos. Como podemos constatar:

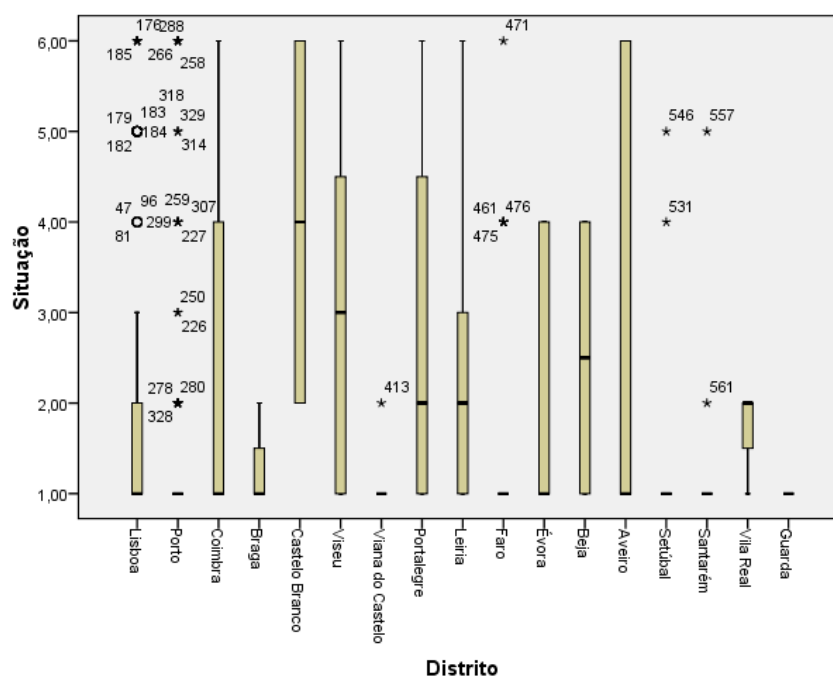
Case Processing Summary							
Afectação		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Situação	SNC	347	99,7%	1	,3%	348	100,0%
	OPE	150	100,0%	0	,0%	150	100,0%
	CTT-EXP	30	100,0%	0	,0%	30	100,0%



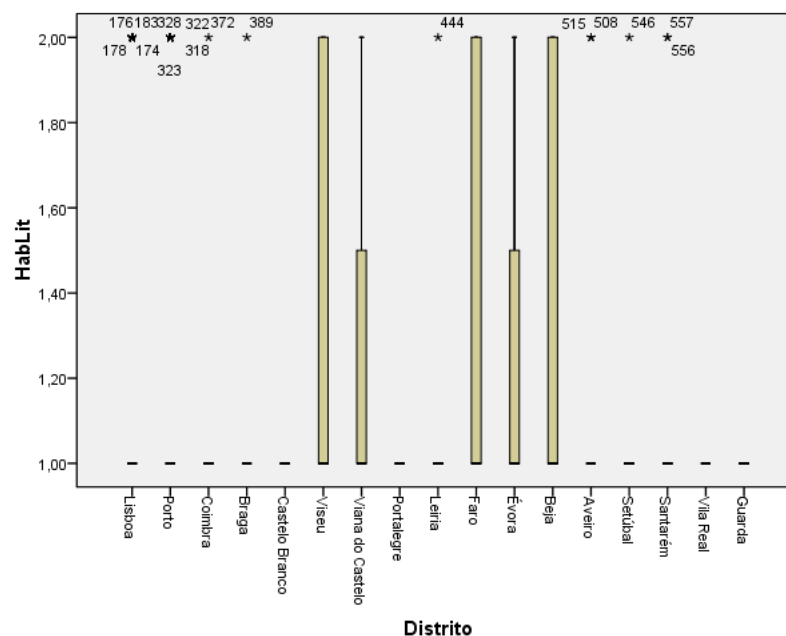
Analisando a variável “Situação no Processo” versus “Distrito” verifica-se que apenas no distrito da Guarda, todos os candidatos inscritos concluíram o processo.

Em Viana do Castelo e Santarém apenas um candidato foi transferido. Por outro lado, em Castelo Branco, nenhum dos candidatos concluiu o processo no CNO dos CTT. Pelo facto do número de trabalhadores ser reduzido, foram transferidos para outros CNO's ou desistiram.

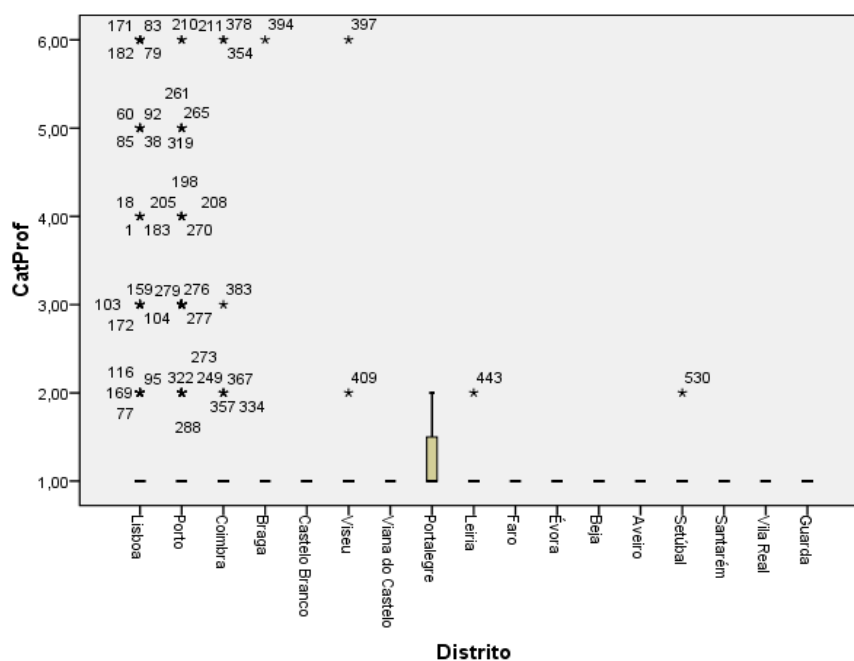
	Distrito	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Situação	Lisboa	185	100,0%	0	,0%	185	100,0%
	Porto	143	99,3%	1	,7%	144	100,0%
	Coimbra	58	100,0%	0	,0%	58	100,0%
	Braga	7	100,0%	0	,0%	7	100,0%
	Castelo Branco	2	100,0%	0	,0%	2	100,0%
	Viseu	16	100,0%	0	,0%	16	100,0%
	Viana do Castelo	15	100,0%	0	,0%	15	100,0%
	Portalegre	7	100,0%	0	,0%	7	100,0%
	Leiria	10	100,0%	0	,0%	10	100,0%
	Faro	40	100,0%	0	,0%	40	100,0%
	Évora	12	100,0%	0	,0%	12	100,0%
	Beja	10	100,0%	0	,0%	10	100,0%
	Aveiro	21	100,0%	0	,0%	21	100,0%
	Setúbal	24	100,0%	0	,0%	24	100,0%
	Santarém	11	100,0%	0	,0%	11	100,0%
	Vila Real	4	100,0%	0	,0%	4	100,0%
	Guarda	4	100,0%	0	,0%	4	100,0%



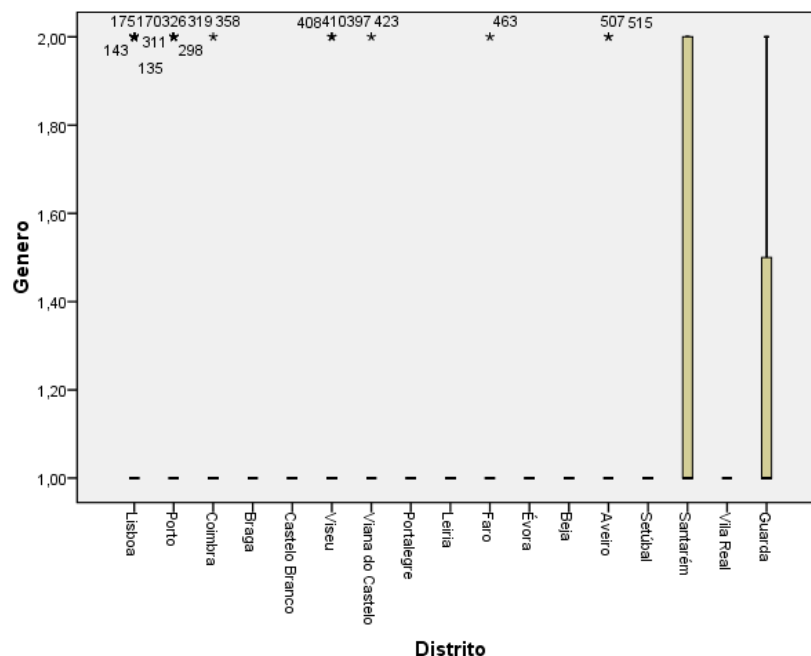
Relativamente às “Habilitações Literárias” verifica-se que, praticamente em todos os distritos, a maioria dos trabalhadores possuem o 6º ano de escolaridade, ou seja, o 2º Ciclo do Ensino Básico.



A nível das categorias profissionais dos trabalhadores, verifica-se que, a grande maioria pertence ao grupo profissional de “Carteiro”.



Por fim, e à semelhança do que registámos nas análises anteriores, também a nível nacional se verifica que a maioria dos adultos que se envolveram neste processo, são do sexo masculino. Alguns dos distritos, contemplando apenas trabalhadores masculinos.



4.5.4 – Totalidade do País: Correlação e Regressão Linear

A análise da correlação e regressão linear das variáveis “Idade” e “Duração do Processo”, levou-nos aos seguintes resultados:

		ClasseEtária	Duração
ClasseEtária	Pearson Correlation	1,000	-,059
	Sig. (2-tailed)		,230
	N	570,000	419
Duração	Pearson Correlation	-,059	1,000
	Sig. (2-tailed)	,230	
	N	419	419,000

A correlação entre a Idade e a Duração do Processo não é significativa a 5%. Considerando a população total, assistimos a uma correlação negativa, idêntica à análise já efectuada anteriormente, o que poderá indicar que os adultos com mais idade são os que concluem o processo em menos tempo. O resultado da regressão apresenta-se no quadro seguinte:

Modelo		Coeficientes não Standardizados		Coeficientes Standardizados	t	Sig.
		B	Erro Standard	Beta		
1	Constante	3,083	,187		16,447	,000
	Classe Etária	-,057	,047	-,059	-1,203	,230

Donde, se pode inferir que a equação da recta que traduz essa regressão é dada por:

$$\hat{y} = -0,057x + 3,083$$

A análise da correlação e regressão entre as variáveis “Duração do Processo” e “Habilitações Literárias” levou-nos à seguinte situação:

		Duração	HabLit
Duração	Pearson Correlation	1,000	-,001
	Sig. (2-tailed)		,978
	N	419,000	419
HabLit	Pearson Correlation	-,001	1,000
	Sig. (2-tailed)	,978	
	N	419	570,000

A correlação entre as variáveis “Habilitações Literárias” e a “Duração do Processo” não é significativa a 5%. Considerando a população total, assistimos a uma correlação negativa ínfima. Poderemos dizer que o nível das habilitações literárias dos adultos não se encontra relacionada com o tempo de duração do processo. O resultado da regressão apresenta-se no quadro seguinte:

Modelo		Coeficientes não Standardizados		Coeficientes Standardizados	t	Sig.
		B	Erro Standard	Beta		
1	Constante	2,877	,222		12,977	,000
	Hab. Lit.	-,005	,185	-,001	-,028	,978

Donde, se pode inferir que a equação da recta que pode traduzir essa regressão é dada por:

$$\hat{y} = -0,005x + 2,877$$

CAPÍTULO V

Técnicas de Inferência Estatística Não Paramétrica

5.1 Introdução

A análise de variáveis individualmente nem sempre espelha adequadamente a realidade. Os métodos paramétricos, univariados e multivariados, obrigam muitas vezes a que sejam assumidos erroneamente pressupostos, que ao não serem verificados, implicam graves erros na análise e conclusões. A maior parte das situações práticas tem em consideração múltiplas variáveis, com a desvantagem de que muitas vezes a Análise de Dados Multivariados pode tornar-se difícil ou até mesmo impossível, sendo a inerente complexidade das análises apontada como um dos principais motivos pelo qual estas metodologias têm tido uma expansão reservada.

As técnicas de Inferência Estatística Não Paramétrica surgem neste contexto, com o sentido de procurar colmatar problemáticas com difícil resolução no campo da Estatística Paramétrica, tendo neste trabalho constituído uma metodologia essencial, dadas as características inerentes à amostra observada.

Os primeiros métodos de estatística não paramétrica surgem por volta de 1710. No entanto, só bastante mais tarde, em 1942, com Wolfwitz, começa a verificar-se o seu desenvolvimento e um maior impacto, sendo hoje a estatística não paramétrica considerada como um ramo extremamente importante da estatística.

As técnicas não paramétricas são usadas nos mais diversos campos da ciência, tais como a biologia, a física, as ciências sociais e as ciências da saúde.

Os testes não paramétricos são também conhecidos por *distribution-free tests*, ou seja, por testes adistribucionais uma vez que não necessitam do conhecimento prévio da distribuição da variável em estudo (normal)¹. Estes testes constituem uma excelente alternativa aos testes paramétricos sempre que para estes não sejam verificadas as condições de aplicabilidade, nomeadamente, a normalidade da variável em estudo e a homogeneidade das variâncias entre os grupos.

Assim, quando não se verificam os pressupostos de aplicabilidade dos métodos paramétricos, os métodos não paramétricos assumem um papel fundamental, dado que a sua validade não está dependente de hipóteses estabelecidas à priori, por exemplo no que concerne ao tipo de distribuição populacional. *“A designação “Métodos ou Testes não Paramétricos” deve-se ao facto de estes não terem a pretensão de testar ou estimar parâmetros de uma dada distribuição, mas sim estudar o ajustamento de certas funções aos dados, averiguar a independência ou*

¹ MAROCO, João – Análise Estatística, Lisboa: Ed. Sílabo, 2003, p. 169

comparar duas ou mais distribuições”². É assim, que ao longo deste capítulo, irão ser desenvolvidos os métodos não paramétricos, tendo por base principal os conceitos apresentados no manual de Estatística Aplicada (OLIVEIRA, Teresa, Universidade Aberta, 2004).

Podem ser apontadas algumas vantagens dos testes não paramétricos relativamente aos testes paramétricos, tais como:

- a sua aplicabilidade a amostras de dimensão reduzida;
- ao facto de se prescindir a verificação de pressupostos sobre a distribuição da população donde foram extraídos os dados;
- de poderem ser aplicados a dados não numéricos, por exemplo, a dados ordinais;
- à sua simplicidade de aplicação.

No entanto, estes testes também apresentam algumas desvantagens:

- pode verificar-se perda de informação quando utilizados em alternativa aos testes paramétricos, por serem mais simples e rápidos;
- os métodos não paramétricos podem ser menos potentes do que os paramétricos quando em presença de uma mesma dimensão e um mesmo α ;
- quando as amostras são grandes, os métodos matemáticos utilizados poderão ser tendenciosos e trabalhosos.

Podemos então dizer que os testes não paramétricos serão apropriados quando.

- as hipóteses a testar não envolvem parâmetros da população;
- é conhecida a ordem dos dados, e,
- quando não são conhecidos os pressupostos válidos para uso dos métodos paramétricos.

5.2 Testes de Hipóteses – Generalidades

Sempre que se tem uma tomada de decisão, pretende-se correr o menor risco que essa decisão possa envolver. Em Estatística, formulam-se hipóteses que se submetem a diversos testes. A hipótese a ser testada designa-se por Hipótese Nula H_0 , sendo H_1 a Hipótese Alternativa, que se assume, ao nível de significância ao qual foi rejeitada a hipótese nula.

Quando se toma uma decisão sobre H_0 podem ocorrer dois erros distintos: Erro Tipo I e Erro Tipo II.

O Erro Tipo I ocorre quando se rejeita a hipótese nula, sendo ela verdadeira. O Erro Tipo II ocorre quando não se rejeita a hipótese nula sendo ela falsa.

² OLIVEIRA, Teresa – Estatística Aplicada, Lisboa: Un. Aberta, 2004, p. 137

Aumentando a dimensão da amostra, reduzem-se as probabilidades de ocorrência de ambos os tipos de erros.

O erro tipo I, muitas vezes designado por α , é o de nível de significância. Este valor α entra no processo de aceitação ou rejeição de H_0 e, por conseguinte, a condição de objectividade exige que o nível de significância seja definido antes da recolha de dados. De acordo com a importância prática dos resultados, os valores mais comuns adoptados para α são de 0,05 e 0,01.

O teste será tanto mais potente quanto menor for a probabilidade de cometer um erro tipo II, geralmente designado por β . Isto é, o teste óptimo da hipótese H_0 versus H_1 será aquele em que, para uma determinada probabilidade de ocorrer um erro do tipo I fará com que a probabilidade de ocorrer um erro do tipo II seja mínima.

Após definidas as hipóteses e o nível de significância, há que definir a distribuição amostral bem como o teste estatístico apropriado. Em seguida, deve definir-se a regra de decisão, ou seja o critério a seguir para rejeitar ou não rejeitar a hipótese nula.

5.3 Teste do Qui-quadrado (χ^2) para a independência

O teste do Qui-Quadrado permite verificar a independência entre duas variáveis, tendo por base uma disposição dos dados de acordo com uma tabela de contingência do tipo $r \times c$.

Genericamente, uma tabela de contingência resulta de uma classificação, segundo dois itens diferentes, de um mesmo grupo de indivíduos. Tem por objectivo inferir sobre a existência ou inexistência de relação entre as variáveis.

Considere-se, então, uma amostra de n indivíduos extraída de uma população, atendendo a dois critérios de classificação: X (variável 1) e Y (variável 2), cujos valores observados serão representados por O_{ij} com $i=1, \dots, r$, e $j=1, \dots, c$. As frequências observadas podem apresentar-se numa “Tabela de Contingência” com r linhas e c colunas (do tipo anteriormente indicado $r \times c$).

A formalização do teste de hipóteses, com a definição das hipóteses nula e alternativa será apresentada da seguinte forma:

H_0 : Há independência entre as variáveis X e Y

H_1 : Não há independência entre as variáveis X e Y .

Designar-se-á, genericamente, por X_i ($i=1, \dots, r$) uma categoria da primeira variável e, por Y_j ($j=1, \dots, c$) uma categoria da segunda variável e os dados serão apresentados numa tabela de contingência, como se segue:

X / Y	Y_1	Y_2	...	Y_c	Total
X_1	O_{11}	O_{12}	...	O_{1c}	O_{1*}
X_2	O_{21}	O_{22}	...	O_{2c}	O_{2*}
...
X_r	O_{r1}	O_{r2}	...	O_{rc}	O_{r*}
Total	O_{*1}	O_{*2}	...	O_{*c}	O_{**}

Na tabela

O_{ij} – representa os valores observados, $i=1,\dots,r$ e $j=1,\dots,c$,

e

E_{ij} – representa os valores esperados, $i=1,\dots,r$ e $j=1,\dots,c$,

Sendo $E_{ij} = \frac{O_{i*} \times O_{*j}}{n}$, com $i=1,\dots,r$ e $j=1,\dots,c$.

A estatística de teste é dada por:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

onde $\chi^2 \approx \chi^2_{(r-1)(c-1), (1-\alpha)}$, sendo a aproximação tanto melhor quanto maior for a dimensão da amostra.

A regra de decisão é a seguinte:

Se $\chi^2 > \chi^2_{(r-1)(c-1), (1-\alpha)}$, rejeita-se H_0 ao nível de significância α .

5.4 Testes para duas ou mais amostras independentes

5.4.1 Teste de Mann-Whitney

O Teste de Mann-Whitney é um teste não paramétrico que permite a comparação entre duas amostras independentes, de dimensões n_1 e n_2 . É um teste alternativo ao teste t para duas amostras independentes. Enquanto que o teste t compara as médias de duas amostras independentes, o teste de Mann-Whitney compara o centro de localização das duas amostras, como forma de detectar diferenças entre as duas populações correspondentes. As vantagens do teste de Mann-Whitney são: não exige o pressuposto da normalidade, pode ser aplicado para amostras pequenas, pode ser aplicado em variáveis de escala ordinal.

Considerem-se duas amostras independentes:

x_1, x_2, \dots, x_n , retirada da população X

y_1, y_2, \dots, y_n , retirada da população Y

E suponha-se que $n_1 < n_2$ com um total de $n = n_1 + n_2$

As hipóteses a testar são:

H_0 : As duas amostras são provenientes de populações com a mesma distribuição.

H_1 : As duas amostras são provenientes de populações com distribuições distintas.

As hipóteses anteriores podem ser reformuladas, se $F(x)$ definir a função distribuição da população X e $G(x)$ a função distribuição da população Y. As hipóteses podem então ser escritas do seguinte modo:

$H_0: F(x) = G(x) \quad \forall x$

$H_1: \exists x : F(x) \neq G(x).$

A forma como o teste é construído torna-o particularmente sensível às diferenças de medidas de localização, especialmente às diferenças nas medianas das distribuições.

Em vez de se basear em parâmetros da distribuição normal como a média e a variância o teste de Mann-Whitney baseia-se nas ordenações da variável.

Estatística de Teste:

A estatística de teste baseia-se nas ordens (*ranks*) das observações das amostras.

Combinam-se as duas amostras, o total das n observações e ordenam-se estas por ordem crescente assinalando o grupo a que pertencem. No caso das observações empatadas atribui-se a média dada pela posição sequencial das observações que lhe corresponderiam.

A estatística de teste é dada por:

$$U = \min(U_1, U_2)$$

em que,

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \quad \text{e} \quad U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

onde,

$n_1 \Rightarrow$ dimensão da amostra menor

$n_2 \Rightarrow$ dimensão da amostra maior

$R_1 \Rightarrow$ soma das ordenações da menor amostra

$R_2 \Rightarrow$ soma das ordenações da maior amostra

Situação 1 : n_1 e n_2 pequenos (≤ 10)

Compara-se o valor observado da estatística de teste com o valor tabelado (Tabela de Mann-Whitney).

Regra de Decisão: Se $U < U_{\text{tabelado}}$, então pertence à região crítica, pelo que se rejeita a hipótese nula ao nível de significância α .

Situação 2 : n_1 e n_2 grandes (>10)

Para os casos em que n_1 e n_2 são grandes, é usual recorrer à aproximação à normal.

A Estatística de teste é dada por:

$$Z = \frac{U - \mu_v}{\sigma_v} = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n+1)}{12}}} \cap N(0,1)$$

$$\mu_v = \frac{n_1 n_2}{2} \quad \text{e} \quad \sigma_v = \sqrt{\frac{n_1 n_2 (n+1)}{12}}$$

Regra de decisão:

Se o valor observado da estatística de teste pertencer à região não crítica:

- $-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}$ para um nível de significância α , não se rejeita H_0 .

O teste Mann-Whitney pode ser aplicado em situações, em que existem empates nas observações e, em situações em que não ocorrem empates.

Correcção para empates:

No caso em que os empates ocorrem entre duas ou mais observações da mesma amostra o valor de U não é afectado. Mas se os empates envolvem elementos das duas amostras e, ocorrem entre duas ou mais observações o valor de U pode ser afectado. A correcção para empates deve ser feita ao desvio padrão da distribuição amostral U .

$$\sigma_v = \sqrt{\frac{n_1 n_2}{n(n-1)} \times \left(\frac{n^3 - n}{12} \sum \frac{t_j^3 - t_j}{12} \right)}$$

Em que, t_j corresponde ao número de observações empatadas em cada grupo j .

5.4.2 Teste de Kruskal-Wallis

A aplicação da análise de variância paramétrica, normalmente abreviada para ANOVA, pressupõe a verificação de alguns pressupostos, designadamente:

1. A normalidade da sua distribuição;
2. A homogeneidade da variância.
3. Independência

Para além disto, a análise de variância paramétrica só pode ser aplicada quando a variável dependente admite pelo menos a escala de intervalos como nível de medida.

Quando não se verificam os pressupostos da análise de variância paramétrica e/ou quando o nível de medida mais restrito da variável dependente é a escala ordinal, o teste de Kruskal-Wallis que é uma generalização, para $k > 2$ amostras, do teste de Mann-Whitney pode ser considerado como a alternativa não paramétrica à ANOVA one-way (Kruskal e Wallis, 1952). O objectivo do teste é testar se duas ou mais amostras provêm de uma mesma população ou se de populações diferentes ou se, de igual modo, as amostras aleatórias provêm de populações com a mesma distribuição.

Formalmente, as hipóteses sob estudo podem escrever-se como:

H_0 : As k populações partilham a mesma distribuição;

H_1 : Pelo menos uma das k populações difere das restantes.

O modo como o teste é construído torna-o especialmente sensível às diferenças de medidas de localização, em especial às diferenças entre as medianas das k populações.

Deste modo, e em alternativa, as hipóteses também podem ser colocadas da seguinte forma:

Sendo θ_j a mediana da j -ésima população,

H_0 : $\theta_1 = \theta_2 = \dots = \theta_k$ (as medianas populacionais são iguais)

H_1 : $\exists_{i,j} : \theta_i \neq \theta_j$, (existe pelo menos um par de medianas significativamente diferentes).

O teste de Kruskal-Wallis pode ser aplicado a variáveis que admitem como nível de medida mais restrito a escala ordinal (nestes casos a média não pode ser calculada por se tratar de variáveis qualitativas). Desse modo, a estatística do teste baseia-se não nos valores que a variável assume, mas sim no número de ordens desses valores.

Para calcular a estatística de teste, começamos por ordenar, por ordem crescente, todas as observações das diferentes amostras em conjunto (amostra global), atribuindo a cada observação a sua ordem na amostra global e mantendo a origem da observação. Desse modo, a cada observação corresponde o número de ordem na série de valores constituída por todas as observações da variável, independentemente da amostra a que pertencem.

O teste de Kruskal-Wallis baseia-se na comparação entre a média dos valores de ordem das diversas amostras e não na comparação entre as médias amostrais da variável dependente, uma vez que nem sempre é possível calcular as médias amostrais (esse cálculo só é possível quando as variáveis dependentes admitem como nível de medida mais restrito a escala de intervalos).

Considerem-se então k populações, X_1, X_2, \dots, X_k , a partir das quais foram retiradas k amostras aleatórias, de dimensões n_1, n_2, \dots, n_k .

Deste modo, tem-se:

$(X_{11}, X_{12}, \dots, X_{1n})$ da população X_1
 $(X_{21}, X_{22}, \dots, X_{2n})$ da população X_2

$(X_{k1}, X_{k2}, \dots, X_{kn})$ da população X_k

Admita-se que existe independência, não só entre os elementos de cada amostra, mas também entre elementos de amostras distintas. Ordenem-se as k amostras conjuntamente. A observação de mais baixo valor tomará a ordem 1, a segunda a ordem 2, e assim sucessivamente. No caso de haver observações que assumem o mesmo valor, ou seja, empates, resolve-se o problema calculando o valor médio dos lugares que lhes seriam atribuídos se não houvesse empates, como se tratasse de observações subsequentes.

Considere-se $R(X_{ij})$ a ordem (*rank*) atribuída à observação e seja $R_i = \sum_{j=1}^{n_i} R(X_{ij})$

a soma das ordens da i -ésima amostra ($i=1,2,\dots,k$).

O número total de observações é $N = \sum_{i=1}^k n_i$

A estatística do teste de Krüskal-Wallis é dada por:

$$T = \frac{1}{s^2} \left[\sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{n(n+1)^2}{4} \right], \text{ onde } s^2 = \frac{1}{n-1} \left[\sum_{i=1}^k \sum_{j=1}^{n_i} R(X_{ij})^2 - \frac{n(n+1)^2}{4} \right]$$

Para o caso da não existência de empates (ou de o número destes ser muito pequeno), esta estatística reduz-se a:

$$T' = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1)$$

A distribuição por amostragem da estatística de teste depende do número de amostras envolvidas na comparação, bem como do número de observações por amostra.

Assim, para os casos em que não há empates e, se o número de amostras for inferior ou igual a três ou o número de observações por amostra não ultrapassar as 5, para tomar a decisão quanto à diferença nas distribuições das três populações, compara-se o valor da estatística de teste com os valores fornecido pela tabela de Krüskal-Wallis.

Para os casos em que não há empates e, se o número de amostras é superior a três ou a dimensão de uma amostra é superior a 5, a estatística de teste pode ser aproximada à distribuição do Qui-Quadrado com $k-1$ graus de liberdade (distribuição assintótica).

Esta aproximação será tanto melhor quanto maiores forem k , n° de amostras e n_i , dimensão de cada amostra, $i=1, 2, \dots, k$.

Regra de decisão: Rejeita-se H_0 se o valor da estatística de teste for superior ao valor tabelado ($\alpha \%$).

Quando existem observações com o mesmo valor (empates) é importante corrigir o valor do teste. Se mais de 25% das observações forem empates, a estatística de teste T' deverá ser corrigida dividindo T' por:

$$1 - \frac{1}{n^3 - n} \sum_{j=1}^e t_j(t_j^2 - 1)$$

onde e representa o número de amostras com diferentes ordens de empates e t_j representa o número de empates na amostra j .

O procedimento do teste de Kruskal-Wallis pode ser resumido da seguinte forma:

- 1) Ordenar as observações das k amostras num único grupo atribuindo-lhes ordens de 1 a n .
- 2) Calcular R_i para cada amostra (soma das ordens).
- 3) Calcular o valor da estatística T .

A regra de decisão será:

- 1) Para $k=3$ e $n_1, n_2, n_3 \leq 5$ recorrer à tabela. Se $T > t_{crit,\alpha}$ rejeitar H_0 ao nível de significância α .
- 2) Se pelo menos uma das amostras tiver dimensão $n_i > 5$, deve usar-se a distribuição do Qui-Quadrado. Se χ^2_{k-1} então rejeita-se H_0 ao nível de significância α .

Se o valor observado da Estatística de Teste pertencer à região crítica então isso significa que existem diferenças significativas entre as amostras.

Para identificar onde se situa a diferença é necessário proceder a comparações dos grupos, dois a dois, o que corresponde a efectuar $\binom{k}{2}$ testes.

As hipóteses podem ser definidas, para um certo par de grupos (i,j) , $i \neq j$, do seguinte modo:

H_0 : A distribuição na população i é idêntica à distribuição na população j ,

$$\forall (i,j) \quad i \neq j$$

H_1 : A distribuição na população i difere da distribuição na população j , para algum $i \neq j$

A regra de decisão é dada pela seguinte expressão:

$$\left| \frac{R_i}{n_i} - \frac{R_j}{n_j} \right| > t_{n-k; 1-\frac{\alpha}{2}} \times \sqrt{s^2 \times \frac{n-1-t}{n-k}} \times \sqrt{\frac{1}{n_i} - \frac{1}{n_j}}$$

Isto é, a primeira desigualdade será superior à segunda se existirem diferenças entre o grupo i e o j . Assim, $t_{n-k; 1-\frac{\alpha}{2}}$ corresponde ao valor de probabilidade $1 - \frac{\alpha}{2}$ de uma variável aleatória com distribuição t de Student, $n-k$ graus de liberdade.

5.5 Testes para comparação de amostras dependentes

5.5.1 Teste dos Sinais

O teste dos sinais é um teste não paramétrico utilizado na comparação de duas amostras emparelhadas dependentes ou que, de algum modo, se encontram relacionadas.

Enquanto que, nos testes paramétricos, se usam os valores das diferenças $D_i = x_i - y_i$ para as amostras emparelhadas, nos testes não paramétricos usam-se os sinais das diferenças entre os pares x e y . Isto é, o sinal da diferença $D_i = x_i - y_i$.

Pode dizer-se que este teste apresenta uma desvantagem relativamente a outros. Atende apenas aos sinais das diferenças, ignorando a amplitude das mesmas. Este teste é, no fundo, uma alternativa ao teste paramétrico para comparação das médias, quando não se verificam as condições necessárias para aplicação do teste paramétrico em causa.

Então como proceder no teste dos sinais?

Para cada par de valores (x, y) aleatório, se:

1) $x_i > y_i$, implica que $D_i = x_i - y_i > 0$

Então, D_i terá um sinal positivo: “+”.

2) $x_i < y_i$, implica que $D_i = x_i - y_i < 0$

Então, D_i terá um sinal negativo: “-”.

3) $x_i = y_i$, implica que $D_i = x_i - y_i = 0$

Estamos perante um empate e D_i assume o valor “0”.

Podem assim formular-se as hipóteses:

$$H_0: P(x > y) = P(x < y) = \frac{1}{2}$$

Supondo que x_i e y_i provêm da mesma distribuição de probabilidade, então a probabilidade de D_i ser positiva é igual à probabilidade de ser negativa, logo, igual a $\frac{1}{2}$.

$$H_1: P(x > y) = P(x < y) \neq \frac{1}{2}$$

Estatística do Teste:

Considere-se uma variável aleatória V representativa do número de sinais que ocorrem com menor frequência. Considere-se ainda v_0 , como o valor observado da variável V .

Assume-se que $V \approx b(n, \frac{1}{2})$.

Designa-se por N , o número total de sinais “+” e “-“. Considere-se também que $N=n$ para o número de empates.

Para a realização do deste teste há que considerar duas situações distintas:

1) Para $N < 20$

Quando se trata de pequenas amostras, utiliza-se o cálculo directo:

$P(V=k) = \binom{N}{k} \left(\frac{1}{2}\right)^N$, $k=0, 1, 2, \dots, N$ e calcula-se a probabilidade:

$$P(V \leq v_0) = \sum_{k=0}^{v_0} \binom{N}{k} \left(\frac{1}{2}\right)^N$$

Neste caso, a regra de decisão a seguir, será:

Rejeitar H_0 ao nível de significância α se $P(V \leq v_0) < \alpha$.

2) Para $N \geq 20$

Quando se trata de uma amostra em que o número de sinais é maior ou igual a 20, recorre-se à aproximação à normal.

Considere-se, então, que $V \dot{\sim} (\mu; \sigma)$, em que $\mu = \frac{N}{2}$ e $\sigma = \frac{\sqrt{N}}{2}$.

A estatística do teste a utilizar, será:

$$Z = \frac{v_0 - \frac{N}{2}}{\frac{\sqrt{N}}{2}} \dot{\sim} N(0,1)$$

A regra de decisão, admitindo um teste bilateral, será rejeitar H_0 ao nível de significância α , se

$$P(V \leq v_0) < \frac{\alpha}{2}.$$

Para uma melhor compreensão da aproximação da binomial à normal, vamos efectuar uma análise mais detalhada sobre este tema.

Assim, em termos genéricos, quando a dimensão da amostra é grande ($N > 20$ e $Np > 7$), a distribuição binomial aproxima-se da distribuição normal. Quando tal acontece, pode usar-se a estatística:

$$Z = \frac{|\hat{Np} - Np| - 0,5}{\sqrt{Npq}}$$

Uma vez que $\mu = Np$ e $\sigma = \sqrt{Npq}$ quando estamos perante uma distribuição binomial.

Teremos então:

$Z \dot{\sim} N(0,1)$ e \hat{p} é a estimativa amostral da proporção de sucessos e $\hat{q} = 1 - \hat{p}$ será a estimativa amostral da proporção dos insucessos para testar a hipótese nula:

$H_0: p = p_0$ versus hipótese alternativa $H_1: p \neq p_0$ ³

Com base nos valores encontrados, rejeita-se H_0 , para um teste bilateral, se

$$Z \in]-\infty; -z_{1-\frac{\alpha}{2}}] \cup [z_{1-\frac{\alpha}{2}}; +\infty[$$

Caso contrário, não se rejeita H_0 .

5.5.2 O Teste de Wilcoxon

Este teste é utilizado quando a variável é medida numa escala ordinal, em amostras emparelhadas e mede a significância da diferença entre as duas medições dessa variável. Este teste também é utilizado como alternativa não paramétrica ao teste t-Student quando não é verificado o pressuposto da existência da distribuição normal da variável nas duas medições.

Com a aplicação do teste de Wilcoxon é possível a formulação de hipóteses sobre uma distribuição $F(x)$ duma variável x , em duas populações A e B.

As hipóteses estatísticas podem descrever-se da seguinte forma, de acordo com o tipo de teste:

1 – Teste bilateral

$$H_0: F(x_1) = F(x_2)$$

$$H_1: F(x_1) \neq F(x_2)$$

O teste inicia-se pelo cálculo de uma variável auxiliar $D_i = x_i - x_j$, para cada par $(x_i; x_j)$. Seguidamente atribui-se uma ordem de 1, ..., n a todas as diferenças D_i de forma crescente ao seu valor absoluto. Dito de outra forma, atribui-se o número 1 à diferença de menor valor absoluto, até à atribuição do número n à diferença de maior valor absoluto. Seguidamente atribui-se a cada ordem, o sinal original da diferença D_i .

Calculam-se, então as estatísticas S^+ e S^- representando, respectivamente, a soma das diferenças positivas e a soma das diferenças negativas em valor absoluto. A relação entre S^+ e S^- é dada pela expressão:

$$S^+ + S^- = \frac{n(n+1)}{2}$$

Após a determinação deste valor, escolhe-se aquele em que o valor absoluto seja menor e designa-se por T. Compara-se então este valor com o valor crítico tabelado e decidir sobre a rejeição ou não da hipótese nula.

³ (Zar, 1999) in Análise Estatística, 3ª edição, Edições Sílabo, João Maroco

A estatística de teste que pode ser usada, na utilização do SPSS, é a seguinte:

$$Z = \frac{\min(S^+, S^-) - N(N+1)/4}{\sqrt{N(N+1)(2N+1)/24}} \approx N(0,1)$$

Obtido o p-value, rejeita-se H_0 quando o p-value $\leq \alpha$.

5.5.3 Teste de Friedman

O teste de Friedman, que permite estudar apenas um único factor (ANOVA de Friedman) relativo a duas ou mais amostras emparelhadas, não é mais do que uma extensão, a k amostras, do teste de sinais.

Considerem-se, então, k populações x_1, x_2, \dots, x_k não independentes, em que são extraídas aleatoriamente amostras de dimensão n . Pretende-se, com este teste verificar se existem diferenças entre as populações ou entre as distribuições das k variáveis aleatórias. Representando os dados numa tabela, pode obter-se uma visualização do tipo:

Indivíduos	Tratamentos			
	1	2	...	k
1	x_{11}	x_{12}	...	x_{1k}
2	x_{21}	x_{22}	...	x_{2k}
...			
n	x_{n1}	x_{n2}	...	x_{nk}

As hipóteses a considerar, neste teste, são as seguintes:

Hipótese nula:

Todas as k populações têm a mesma distribuição, ou seja,

$$H_0: F(x_1) = F(x_2) = \dots = F(x_k)$$

Hipótese alternativa:

Existe pelo menos uma das populações com distribuição diferente, ou seja,

$$H_1: \exists_{i,j}: F(x_i) \neq F(x_j) \text{ em que, } i \neq j \text{ com } i, j = 1, 2, \dots, k.$$

Como este teste é baseado na construção de uma tabela e são atribuídas ordens às observações de cada uma das suas linhas, temos como resultado uma tabela de ordens. O teste de Friedman permite determinar a probabilidade das diferentes linhas provirem de populações idênticas ou provirem de populações com igual distribuição.

Represente-se, então, por $r(x_{ij})$ a ordem atribuída à observação x_{ij} e por R_j o total das ordens da j -ésima coluna.

A estatística do teste de Friedman é dada por:

$$\chi_r^2 = \frac{12}{nk(k+1)} \sum_{j=1}^k (R_j)^2 - 3n(k+1) \quad \text{com} \quad R_j = \sum_{i=1}^n r_{ij}$$

onde n é o número de indivíduos por amostra e k é o número de amostras.

A estatística do teste de Friedman aproxima-se da distribuição do χ^2 com $k-1$ graus de liberdade.

Assim, rejeita-se H_0 , ao nível de significância α se $\chi_r^2 > \chi_{1-\alpha; k-1}^2$.

Interpreta-se a rejeição de H_0 como o facto do tamanho das ordens atribuídas depender das condições às quais foram sujeitos os indivíduos.

5.6 Testes de Correlação

5.6.1 Teste de correlação Paramétrica: Teste de Pearson

Dadas duas amostras de observações medidas numa escala de intervalos ou razões, pode pedir-se o grau de associação linear entre elas através do coeficiente de correlação de Pearson ou simplesmente coeficiente de correlação amostral (*Pearson product-moment correlation coefficient*).

Ao representar as amostras por x_1, x_2, \dots, x_n e y_1, y_2, \dots, y_n , o coeficiente de correlação amostral é dado por:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Este coeficiente também pode ser escrito como

$$\sum_{i=1}^n Z_{x_i} Z_{y_i}$$

onde Z_x e Z_y representam as observações padronizadas, ou seja, após subtrair-se a média e dividir-se pelo desvio padrão, cada uma das amostras.

O coeficiente de correlação pode variar entre -1 e 1. R assume o valor 1 quando os pontos (x_i, y_i) estão exactamente sobre uma recta de declive positivo. Neste caso, a um aumento numa das variáveis, corresponde, obviamente, a um aumento na outra. R assume o valor -1, quando os pontos estão exactamente sobre uma recta de declive negativo. Nesta situação, a um aumento

numa das variáveis, corresponde a uma diminuição na outra. Estes dois casos, correspondem ao máximo de associação linear que é possível observar entre duas amostras.

Quando as amostras são independentes, o valor do coeficiente assumirá o valor zero ou próximo dele. Muitas vezes, usa-se o coeficiente de correlação elevado ao quadrado, R^2 , a que se chama coeficiente de determinação. Uma vez que, $-1 \leq R \leq 1$, o coeficiente de determinação varia entre 0 e 1. Este coeficiente mede a percentagem de variabilidade de uma das variáveis que é explicada pela outra. Ou seja, quando duas variáveis aleatórias se encontram associadas, a variação de uma delas, implica uma variação na outra.

Dizer que o coeficiente de determinação é, por exemplo, 0,6, significa que 60% da variância de uma das variáveis é explicada pela outra variável e apenas 40% é de natureza independente.

Para se ter uma leitura mais informativa deste coeficiente é usual proceder a um teste de hipóteses para testar as seguintes hipóteses:

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

onde ρ representa o coeficiente de correlação da população de onde foram retirados os dados.

Este teste tem como pressupostos:

- as populações de onde foram retirados as amostras têm distribuição normal,
- a relação entre as variáveis é de forma linear, caso exista.

Também é importante saber que:

- no caso de variáveis normais relacionadas linearmente, um coeficiente de correlação nulo é equivalente a independência. No entanto se as distribuições não forem normais essa equivalência pode não ser observada. Portanto, duma maneira geral, ao considerar uma relação de tipo linear, um coeficiente de correlação nulo não implica necessariamente uma independência.
- quando duas variáveis são independentes, $\rho = 0$. Isto deverá traduzir-se num coeficiente de correlação amostral próximo de zero. No entanto, existem situações em que $\rho = 0$ e as variáveis são dependentes, possivelmente com outro tipo de relação que não a linear. Portanto, o facto de observarmos um coeficiente de correlação nulo pode apenas significar que não existe uma relação linear entre as variáveis.

Resumindo, o coeficiente de correlação de Pearson mede o grau de associação linear entre duas variáveis medidas numa escala de intervalos. Se as variáveis tiverem distribuição normal pode efectuar-se um teste de hipóteses para averiguar se o coeficiente de correlação da população é significativamente diferente de zero, o que significará, nesse contexto, que as variáveis são independentes.

5.6.2 Testes de Correlação Não Paramétrica

5.6.2.1 Teste de Spearman

Considere-se uma população da qual foi retirada uma amostra de dimensão n de pares ordenados de duas variáveis aleatórias x e y . Considere-se ainda que as mesmas variáveis são ordenadas de forma crescente e lhes é atribuído um número de ordem.

Estas variáveis podem encontrar-se associadas de uma forma directa ou de uma forma inversa como se mostra na tabela seguinte:

A		B	
Variável x (Nº de ordem)	Variável y (Nº de ordem)	Variável x (Nº de ordem)	Variável y (Nº de ordem)
1	1	1	n
2	2	2	n-1
...
n-1	n-1	n-1	2
n	n	n	1

Considere-se d_i (com $i=1, 2, 3, \dots, n$) a diferença entre os números de ordem de cada par de observações x_i e y_i . Em presença de uma associação directa, teremos:

$\sum_{i=1}^n d_i^2 = 0$. No caso de uma associação inversa, teremos: $\sum_{i=1}^n d_i^2 = \frac{n(n^2-1)}{3}$.

Charles Spearman construiu uma medida de associação relativa, denominada coeficiente de correlação ordinal de Spearman, definida com base naquele somatório e que se traduz por

$$R_s = 1 - \frac{6 \times \sum_{i=1}^n d_i^2}{n(n^2-1)}$$

O coeficiente R_s assume o valor 1 quando, entre o conjunto das observações existe uma associação directa perfeita e assume o valor -1 quando se verificar uma associação inversa perfeita. Quando não se verificar qualquer associação entre as variáveis x e y , o coeficiente R_s assume valores próximos de zero.

A partir deste coeficiente, pode ser construído um teste bilateral para testar H_0 , em que:

H_0 : As variáveis não se encontram associadas

H_1 : As variáveis encontram-se associadas

Como hipóteses alternativas de H_1 , podem ainda considerar-se:

H_1 : Associação directa (teste unilateral à direita)

H_1 : Associação inversa (teste unilateral à esquerda)

Quando a amostra for superior a 30 observações, a estatística de teste deverá ser substituída por

$$\frac{R_s}{\sqrt{(1 - R_s^2)/(n - 2)}}$$

Quando H_0 é verdadeira, segue uma distribuição t de Student com $(n-2)$ graus de liberdade. Sempre que se verificarem empates, atribui-se às observações naquela situação o número de ordem que corresponde à média dos números de ordem que as observações receberiam se não estivessem empatadas. Se existir um número pequeno de empates, o valor da estatística R_s deve ser calculado através da expressão:

$R_s = 1 - \frac{6 \times \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$ e não será significativamente afectado. Caso contrário, a estatística R_s deverá ser calculada através da expressão:

$$R_s = \frac{n(n^2 - 1) - 6 \times \sum_{i=1}^n d_i^2 - 6 \times (u' + v')}{\sqrt{n(n^2 - 1) - 12 \times u'} \times \sqrt{n(n^2 - 1) - 12 \times v'}}$$

sendo

$$u' = \frac{\sum_{i=1}^n u_i^3 - \sum_{i=1}^n u_i}{12} \quad \text{e} \quad v' = \frac{\sum_{i=1}^n v_i^3 - \sum_{i=1}^n v_i}{12}$$

em que u_i e v_i representam o número de empates no i -ésimo grupo de observações iguais pertencentes, respectivamente, à variável x e à variável y .

5.6.2.2 Teste τ de Kendall

Podem referir-se duas eventuais vantagens deste coeficiente sobre o anterior: se as amostras tiverem dimensão muito reduzida e valores repetidos, os resultados do teste ao coeficiente são mais precisos no caso presente; por outro lado, o coeficiente τ de Kendall pode ser generalizado para correlações parciais que são correlações medidas entre duas variáveis após remoção do efeito de uma possível terceira variável sobre ambas.

Uma diferença muito importante entre os dois coeficientes (Kendall e Spearman) reside na sua interpretação e na impossibilidade de comparar directamente os valores provenientes de ambos. Embora o objectivo comum seja o de medir associação, a forma de o fazer é distinta.

O coeficiente de Kendall é muitas vezes descrito como uma medida de concordância entre dois conjuntos de classificações relativas a um conjunto de objectos ou experiências. Uma alternativa ao coeficiente de Spearman é assim, o coeficiente de Kendall (Kendall's tau coefficient) que se aplica nas mesmas condições.

O teste não paramétrico τ de Kendall, também conhecido pela correlação de Kendall, permite efectuar um teste de independência para uma distribuição livre e permite calcular a medida da força da dependência entre duas variáveis.

O teste de correlação de Spearman é um teste satisfatório para analisar a hipótese nula de independência entre duas variáveis, embora seja de difícil interpretação quando a hipótese nula é rejeitada. O teste de correlação de Kendall vem resolver este problema.

Considere-se duas amostras x e y , de tamanho n . O número total de pares possíveis de observações de x e y é $n(n-1)/2$. Considerem-se os pares ordenados dos valores de x e depois os valores de y . Se $x_3 > y_3$ quando ordenados ambos x e y , o terceiro par é concordante. Caso contrário, será discordante.

Seja S a diferença entre pares concordantes, ordenados na forma indicada (n_c) e discordantes, ordenados de forma diferente (n_d), o teste Tau (τ) de Kendall, para S , é traduzido pela expressão:

$$\tau = \frac{n_c - n_d}{n(n-1)/2}$$

Se as observações tiverem os mesmos valores, então o teste usado, passará a ser o τ_b

$$\tau_b = \frac{S}{\sqrt{\left[n(n-1)/2 - \sum_{i=1}^t t_i(t_i-1)/2 \right] \left[n(n-1)/2 - \sum_{i=1}^u u_i(u_i-1)/2 \right]}}$$

onde t_i é o número de observações iguais para um grupo de x e u_i é o número de observações iguais para um grupo de y .

Em presença de valores iguais, a estatística τ_b é dada como uma variante de τ ajustada a valores iguais (Kendall, 1970).

Quando não existem valores iguais, teremos $\tau_b = \tau$

CAPÍTULO VI

Inferência não paramétrica aplicada à avaliação de desempenho

6.1 Análise da avaliação de desempenho (População B3)

Toda a população trabalhadora dos CTT é sujeita a avaliação anual tendo em atenção o seu desempenho profissional. As avaliações a que os trabalhadores estão sujeitos dependem das suas categorias profissionais e são específicas para trabalhadores qualificados como chefias, quadros e não quadros. A população em causa foi avaliada de acordo com os pressupostos anteriores e em prática na Empresa. A avaliação final, objecto desta análise, é o resultado das classificações obtidas nos seguintes itens:

A - Resultados

- Realização dos objectivos (grau de cumprimento dos objectivos de negócio e produtividade e qualidade)

B - Competências

- Competências técnicas (domínio e utilização de conhecimentos técnicos)
- Planeamento e organização (conformidade a normas e procedimentos)
- Competências relacionais (envolvimento e espírito de equipa e relacionamento interpessoal)
- Responsabilidade e compromisso com a organização (primazia dada ao cliente – interno e externo – e à obtenção de resultados, predisposição para a mudança, compromisso com a política de ambiente e segurança e atitude face à assiduidade e pontualidade)

C – Avaliação Global e Fundamentação

- A Atitude (avaliada tendo em atenção, a iniciativa, autonomia, responsabilidade e relacionamento)
- A Liderança (objecto de avaliação na influência/persuasão, gestão e motivação de equipas e desenvolvimento dos outros)
- A Adaptabilidade (é objecto de análise na abertura à mudança, dinamismo, proactividade, flexibilidade e resolução de problemas)
- As Competências Técnicas (são avaliadas tendo em atenção, a aprendizagem, os conhecimentos técnicos e a aplicação de alterações tecnológicas e/ou operacionais)
- O Desempenho (é avaliado com base no cumprimento dos objectivos: receita, produtividade, qualidade e gestão de recursos).

Como se disse anteriormente, é nosso propósito analisar a influência da conclusão do processo RVCC (9º ano), na avaliação de desempenho desta população.

Para isso, vamos analisar as avaliações de desempenho nos anos 2004 versus 2005, 2005 versus 2006 e 2006 versus 2007. A metodologia adoptada está de acordo com o Processo RVCC já que engloba as avaliações de acordo com a data da conclusão do processo, pelos trabalhadores.

Dado que se trata de duas populações emparelhadas e dependentes, recorreu-se à utilização de alguns testes não paramétricos (Wilcoxon, Sinais, Friedman, e Kendall's). A análise estatística foi efectuada com o software SPSS (v. 16, SPSS Inc.).

Os resultados encontrados foram os seguintes:

A - Avaliação de desempenho 2004 versus 2005

1) Teste de Wilcoxon

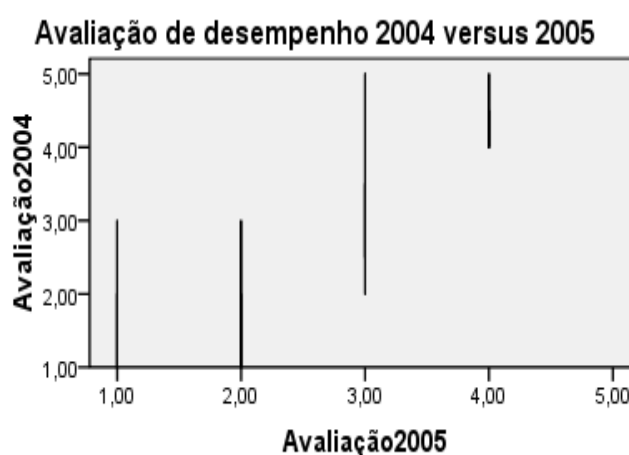
Foi utilizado um teste bilateral, considerando as hipóteses nula e alternativa, respectivamente:

H_0 : O processo RVCC não influenciou a avaliação de desempenho

H_1 : A avaliação de desempenho foi influenciada pelo processo RVCC

Observou-se uma melhoria da frequência da avaliação de desempenho (medida numa escala ordinal de 1→A – Muito Bom a 6→D – Não Satisfatório) no momento antes da avaliação para o momento depois da avaliação ($S^+=76$; $S^-=134$; $\overline{S}_+=9,5$; $\overline{S}_-=11,17$; $Z=-1,188$; $P_{ue}=0,325$ e $N=42$).

O gráfico seguinte mostra a distribuição dos scores observados na avaliação de desempenho 2004-2005:



Também, através do gráfico, se pode constatar essa melhoria:

- As avaliações 1-Muito Bom, em 2005, correspondem a avaliações de 2, 3 e 4 em 2004;
- As avaliações 2, em 2005, correspondem a avaliações 2 e 3, em 2004;
- As avaliações 3, em 2005, correspondem a avaliações 2, 3, 4 e 5, em 2004;
- As avaliações 4, em 2005, correspondem a avaliações 4 e 5, em 2004;

2) Teste dos Sinais

Condições de aplicação do teste:

Considerou-se, que:

$X \rightarrow$ Avaliação de desempenho antes da frequência do processo RVCC

$y \rightarrow$ Avaliação de desempenho após a frequência do processo RVCC

e, como hipóteses nula e alternativa:

$$H_0: P(x > y) = P(x < y) = 0,5 \text{ (A avaliação de desempenho é igual)}$$

$$H_1: P(x > y) = P(x < y) \neq 0,5 \text{ (A avaliação de desempenho é diferente)}$$

Da mesma forma, podemos constatar que através deste teste se chega à mesma conclusão. Verifica-se uma melhoria da frequência da avaliação de desempenho (medida numa escala ordinal de 1 \rightarrow A – Muito Bom a 6 \rightarrow D – Não Satisfatório) no momento antes da avaliação para o momento depois da avaliação (Diferenças Positivas=8; Diferenças negativas=12; Empates=22; $P_{ue}=0,5034$ e $N=42$).

3) Teste de Friedman

Para a realização deste teste, foram consideradas as hipóteses:

Hipótese nula:

As duas populações têm a mesma distribuição, ou seja,

$$H_0: \text{A avaliação de desempenho não foi influenciada pelo processo RVCC}$$

Hipótese alternativa:

$$H_1: \text{A avaliação de desempenho foi influenciada pelo processo RVCC}$$

A utilização do teste de Friedman vem confirmar o que ficou dito. Da mesma forma constata-se uma melhoria da frequência da avaliação de desempenho (medida numa escala ordinal de 1→A – Muito Bom a 6→D – Não Satisfatório) no momento antes da avaliação para o momento depois da avaliação (Média 2004=1,55; Média 2005=1,45; $\chi^2=8$; $P_{ue}=0,371$; g.l.=1 e N=42).

4) Teste de Kendall

Confirmação feita, também pelo teste de Kendall. Verifica-se uma melhoria da frequência da avaliação de desempenho (medida numa escala ordinal de 1→A – Muito Bom a 6→D – Não Satisfatório) no momento antes da avaliação para o momento depois da avaliação (Média 2004=1,55; Média 2005=1,45; $\chi^2=0,8$; Coef. Concordância de Kendall=0,019; g.l.=1 e N=42).

B - Avaliação de desempenho 2005 versus 2006

1) Teste de Wilcoxon

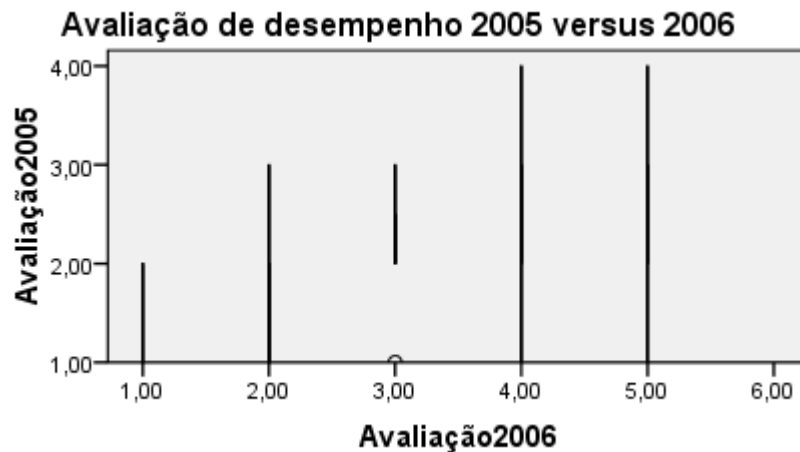
À semelhança do que se fez em 2004/2005, foi utilizado um teste bilateral, considerando as hipóteses nula e alternativa, respectivamente:

H_0 : O processo RVCC não influenciou a avaliação de desempenho

H_1 : A avaliação de desempenho foi influenciada pelo processo RVCC

Observou-se que a frequência da avaliação de desempenho piorou (medida numa escala ordinal de 1→A – Muito Bom a 6→D – Não Satisfatório) no momento antes da avaliação para o momento depois da avaliação ($S^+=6900$; $S^-=240$; $\overline{S}_+=62,16$; $\overline{S}_-=30$; Empates = 35; $Z=-9,048$; $P_{ue}=0,000$ e N=154). Este facto tem uma explicação. Em 2006 foram introduzidos limites de referência nas avaliações, as denominadas quotas. As avaliações B2+B1+A não podem exceder 40% do total dos avaliados.

O gráfico seguinte mostra a distribuição dos scores observados na avaliação de desempenho 2005-2006:



Também, através do gráfico, se pode constatar que:

- As avaliações 1-Muito Bom, em 2006, correspondem a avaliações de 1 e 2, em 2005;
- As avaliações 2, em 2006, correspondem a avaliações 1, 2 e 3, em 2005;
- As avaliações 3, em 2006, correspondem a avaliações 2 e 3, em 2005;
- As avaliações 4 e 5 em 2006, correspondem, respectivamente, a avaliações 1, 2, 3 e 4 e 1, 2, 3, 4 e 5, em 2005;

2) Teste dos Sinais

Condições de aplicação do teste:

Considerou-se, que:

$X \rightarrow$ Avaliação de desempenho antes da frequência do processo RVCC

$y \rightarrow$ Avaliação de desempenho após a frequência do processo RVCC

e, como hipóteses nula e alternativa:

$$H_0: P(x > y) = P(x < y) = 0,5 \quad (\text{A avaliação de desempenho é igual})$$

$$H_1: P(x > y) = P(x < y) \neq 0,5 \quad (\text{A avaliação de desempenho é diferente})$$

Da mesma forma, podemos constatar que através deste teste se chega à mesma conclusão. Verifica-se que a frequência da avaliação de desempenho piorou (medida numa escala ordinal de 1 \rightarrow A – Muito Bom a 6 \rightarrow D – Não Satisfatório) no momento antes da avaliação, 2005 para o momento depois da avaliação, 2006. (Diferenças Positivas=111; Diferenças negativas=8; Empates=35; $Z=-9,35$ $P_{ue}=0,000$ e $N=154$). Pensamos que as razões são as acima descritas.

3) Teste de Friedman

Para a realização deste teste, foram consideradas as hipóteses:

Hipótese nula:

As duas populações têm a mesma distribuição, ou seja,

H_0 : A avaliação de desempenho não foi influenciada pelo processo RVCC

Hipótese alternativa:

H_1 : A avaliação de desempenho foi influenciada pelo processo RVCC

A utilização do teste de Friedman vem confirmar o que ficou dito. Da mesma forma constata-se a descida da frequência da avaliação de desempenho (medida numa escala ordinal de 1→A – muito Bom a 6→D – Não Satisfatório) no momento antes da avaliação para o momento depois da avaliação (Média 2005=1,17; Média 2006=1,83; $\chi^2=89,151$; $P_{ue}=0,000$; g.l.=1 e N=154).

4) Teste de Kendall

Confirmação feita, também, pelo teste de Kendall. Verifica-se que a frequência da avaliação de desempenho piorou (medida numa escala ordinal de 1→A – Muito Bom a 6→D – Não Satisfatório) no momento antes da avaliação para o momento depois da avaliação (Média 2005=1,17; Média 2006=1,83; $\chi^2=89,151$; Coef. Concordância de Kendall=0,579; g.l.=1; $P_{ue}=0,000$ e N=154).

C - Avaliação de desempenho 2006 versus 2007

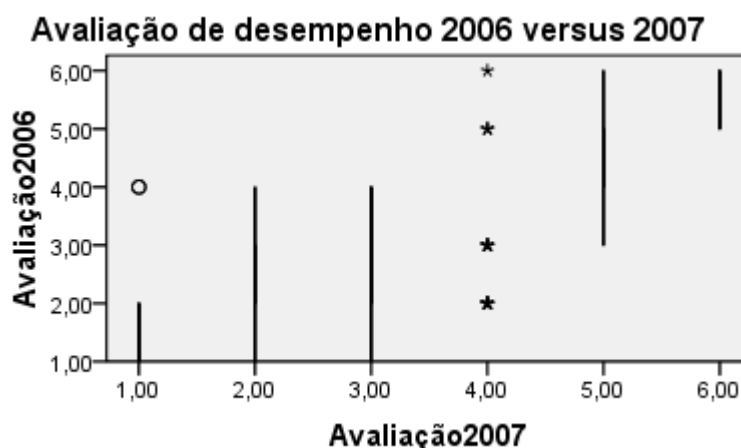
1) Teste de Wilcoxon

Foi utilizado um teste bilateral, considerando as hipóteses nula e alternativa, respectivamente:

H_0 : O processo RVCC não influenciou a avaliação de desempenho

H_1 : A avaliação de desempenho foi influenciada pelo processo RVCC

Observou-se que a frequência da avaliação de desempenho piorou (medida numa escala ordinal de 1→A – Muito Bom a 6→D – Não Satisfatório) no momento antes da avaliação para o momento depois da avaliação ($S^+=1425$; $S^-=1131$; $\overline{S}_+=34,76$; $\overline{S}_-=37,70$; Empates = 93; $Z=-0,883$; $P_{ue}=0,377$ e N=164). Este facto explica-se como anteriormente. O gráfico seguinte mostra a distribuição dos scores observados na avaliação de desempenho 2006-2007:



Também aqui, e devido às quotas impostas, em termos avaliativos, se pode constatar que:

- As avaliações 1-Muito Bom, em 2007, correspondem a avaliações de 1 e 2, em 2006;
- As avaliações 2, em 2007, correspondem a avaliações 1, 2, 3 e 4 em 2006;
- As avaliações 3, em 2007, correspondem a avaliações 1, 2, 3 e 4 em 2006;
- Mantiveram-se as avaliações 4 e as avaliações 5 em 2007, correspondem, a avaliações 3, 4, 5 e 6 em 2006;
- As avaliações 6 em 2007 correspondem a avaliações 5 e 6 em 2006.

2) Teste dos Sinais

Condições de aplicação do teste:

Considerou-se, que:

$X \rightarrow$ Avaliação de desempenho antes da frequência do processo RVCC

$y \rightarrow$ Avaliação de desempenho após a frequência do processo RVCC

e, como hipóteses nula e alternativa:

$$H_0 : P(x > y) = P(x < y) = 0,5 \text{ (A avaliação de desempenho é igual)}$$

$$H_1 : P(x > y) = P(x < y) \neq 0,5 \text{ (A avaliação de desempenho é diferente)}$$

Da mesma forma, podemos constatar que através deste teste se chega à mesma conclusão. Verifica-se que a frequência da avaliação de desempenho piorou (medida numa escala ordinal de 1 \rightarrow A – Muito Bom a 6 \rightarrow D – Não Satisfatório) no momento antes da avaliação, 2005 para o momento depois da avaliação, 2006. (Diferenças Positivas=41; Diferenças negativas=30; Empates=93; $Z=-1,187$; $P_{ue}=0,235$ e $N=164$). Pensamos que as razões são as acima descritas.

4) Teste de Friedman

Para a realização deste teste, foram consideradas as hipóteses:

Hipótese nula:

As duas populações têm a mesma distribuição, ou seja,

H_0 : A avaliação de desempenho não foi influenciada pelo processo RVCC

Hipótese alternativa:

H_1 : A avaliação de desempenho foi influenciada pelo processo RVCC

A utilização do teste de Friedman vem confirmar o que ficou dito. Da mesma forma constata-se a descida da frequência da avaliação de desempenho (medida numa escala ordinal de 1→A – muito Bom a 6→D – Não Satisfatório) no momento antes da avaliação para o momento depois da avaliação (Média 2006=1,47; Média 2007=1,53; $\chi^2 = 1,704$; $P_{ue}=0,192$; g.l.=1 e N=164).

4) Teste de Kendall

Para a realização do teste de Kendall, foram consideradas as hipóteses:

H_0 : A avaliação de desempenho não foi influenciada pelo processo RVCC

H_1 : A avaliação de desempenho foi influenciada pelo processo RVCC

Confirmação feita, também, pelo teste de Kendall. Verifica-se que a frequência da avaliação de desempenho piorou (medida numa escala ordinal de 1→A – Muito Bom a 6→D – Não Satisfatório) no momento antes da avaliação para o momento depois da avaliação (Média 2006=1,47; Média 2007=1,53; $\chi^2 = 1,704$; Coef. Concordância de Kendall=0,010; g.l.=1; $P_{ue}=0,192$ e N=164).

6.2 – Análise Comparativa dos Testes utilizados (SPSS versus R)

Nos quadros de duas entradas que apresentamos abaixo, fazemos a comparação dos testes efectuados e discriminados no capítulo anterior, com a utilização dos dois softwares que serviram de base a este trabalho: O SPSS (versão 16.0) e a programação em R.

Assim, os testes encontram-se discriminados na primeira coluna e, por linha, nas colunas seguintes, discriminamos o respectivo teste.

Embora os valores encontrados sejam diferentes, as conclusões são idênticas e já indicadas anteriormente.

No entanto convém referir que o SPSS é de mais fácil utilização do que a utilização do R, uma vez que, aqui, temos de recorrer à respectiva programação dos testes. Neste caso, e para comparação das avaliações de desempenho nos binómios 2004 -2005, 2005 – 2006 e 2006 – 2007, recorreu-se à programação de testes unilaterais à direita.

MAPA COMPARATIVO 2004 - 2005

Processo	SPSS	R																																		
Testes																																				
Teste de Wilcoxon	<table><tr><th colspan="5">Ranks</th></tr><tr><td rowspan="5">Avaliação2005 - Avaliação2004</td><td></td><td>N</td><td>Mean Rank</td><td>Sum of Ranks</td></tr><tr><td>Negative Ranks</td><td>12^a</td><td>11,17</td><td>134,00</td></tr><tr><td>Positive Ranks</td><td>8^b</td><td>9,50</td><td>76,00</td></tr><tr><td>Ties</td><td>22^c</td><td></td><td></td></tr><tr><td>Total</td><td>42</td><td></td><td></td></tr></table> <p>a. Avaliação2005 < Avaliação 2004; b. Avaliação 2005 > Avaliação2004; c. Avaliação2005 = Avaliação2004</p> <table><tr><th colspan="2">Test Statistics^b</th></tr><tr><td></td><td>Avaliação2005 - Avaliação2004</td></tr><tr><td>Z</td><td>-1,188^a</td></tr><tr><td>Asymp. Sig. (2-tailed)</td><td>,235</td></tr></table> <p>a. Based on positive ranks. b. Wilcoxon Signed Ranks Test</p>	Ranks					Avaliação2005 - Avaliação2004		N	Mean Rank	Sum of Ranks	Negative Ranks	12 ^a	11,17	134,00	Positive Ranks	8 ^b	9,50	76,00	Ties	22 ^c			Total	42			Test Statistics ^b			Avaliação2005 - Avaliação2004	Z	-1,188 ^a	Asymp. Sig. (2-tailed)	,235	<p>> wilcox.test(x,y,paired=TRUE,alternative="greater")</p> <p>Wilcoxon signed rank test with continuity correction</p> <p>data: x and y V = 134, p-value = 0.1216 alternative hypothesis: true location shift is greater than 0</p> <p>Warning messages: 1: In wilcox.test.default(x, y, paired = TRUE, alternative = "greater") : cannot compute exact p-value with ties 2: In wilcox.test.default(x, y, paired = TRUE, alternative = "greater") : cannot compute exact p-value with zeroes</p>
	Ranks																																			
	Avaliação2005 - Avaliação2004		N	Mean Rank	Sum of Ranks																															
		Negative Ranks	12 ^a	11,17	134,00																															
		Positive Ranks	8 ^b	9,50	76,00																															
		Ties	22 ^c																																	
		Total	42																																	
	Test Statistics ^b																																			
		Avaliação2005 - Avaliação2004																																		
	Z	-1,188 ^a																																		
Asymp. Sig. (2-tailed)	,235																																			
Teste dos Sinais	<table><tr><th colspan="2">Frequencies</th></tr><tr><td rowspan="5">Avaliação2005 - Avaliação2004</td><td>N</td></tr><tr><td>Negative Differences^a</td></tr><tr><td>Positive Differences^b</td></tr><tr><td>Ties^c</td></tr><tr><td>Total</td></tr></table> <p>a.Avaliação2005 < Avaliação 2004; b. Avaliação 2005 > Avaliação2004; c. Avaliação2005 = Avaliação2004</p> <table><tr><th colspan="2">Test Statistics^b</th></tr><tr><td></td><td>Avaliação2005 - Avaliação2004</td></tr><tr><td>Exact Sig. (2-tailed)</td><td>0.503444671630858^a</td></tr></table> <p>a. Binomial distribution used. b. Sign Test</p>	Frequencies		Avaliação2005 - Avaliação2004	N	Negative Differences ^a	Positive Differences ^b	Ties ^c	Total	Test Statistics ^b			Avaliação2005 - Avaliação2004	Exact Sig. (2-tailed)	0.503444671630858 ^a	<p>> wilcox.test(y-x,alternative="greater")</p> <p>Wilcoxon signed rank test with continuity correction</p> <p>data: y - x V = 76, p-value = 0.8865 alternative hypothesis: true location is greater than 0</p> <p>Warning messages: 1: In wilcox.test.default(y - x, alternative = "greater") : cannot compute exact p-value with ties 2: In wilcox.test.default(y - x, alternative = "greater") : cannot compute exact p-value with zeroes</p>																				
	Frequencies																																			
	Avaliação2005 - Avaliação2004	N																																		
		Negative Differences ^a																																		
		Positive Differences ^b																																		
		Ties ^c																																		
		Total																																		
	Test Statistics ^b																																			
		Avaliação2005 - Avaliação2004																																		
	Exact Sig. (2-tailed)	0.503444671630858 ^a																																		
Teste de Friedman	<table><tr><th colspan="2">Ranks</th></tr><tr><td></td><td>Mean Rank</td></tr><tr><td>Avaliação2004</td><td>1,55</td></tr><tr><td>Avaliação2005</td><td>1,45</td></tr></table> <table><tr><th colspan="2">Test Statistics</th></tr><tr><td>N</td><td>42,000</td></tr><tr><td>Chi-Square</td><td>,800</td></tr><tr><td>df</td><td>1,000</td></tr><tr><td>Asymp. Sig.</td><td>,371</td></tr></table> <p>a. Friedman Test</p>	Ranks			Mean Rank	Avaliação2004	1,55	Avaliação2005	1,45	Test Statistics		N	42,000	Chi-Square	,800	df	1,000	Asymp. Sig.	,371	<p>> avaliao<-matrix(c(x,y),nrow=2,ncol=2,byrow=TRUE,dimnames=list(c("x","y"),c("c.1","c.2")))</p> <p>> friedman.test(avaliao)</p> <p>Friedman rank sum test</p> <p>data: avaliao Friedman chi-squared = 0, df = 1, p-value = 1</p>																
	Ranks																																			
		Mean Rank																																		
	Avaliação2004	1,55																																		
	Avaliação2005	1,45																																		
	Test Statistics																																			
	N	42,000																																		
	Chi-Square	,800																																		
	df	1,000																																		
	Asymp. Sig.	,371																																		
Teste de Kendall	<table><tr><th colspan="2">Ranks</th></tr><tr><td></td><td>Mean Rank</td></tr><tr><td>Avaliação2004</td><td>1,55</td></tr><tr><td>Avaliação2005</td><td>1,45</td></tr></table> <table><tr><th colspan="2">Test Statistics</th></tr><tr><td>N</td><td>42,000</td></tr><tr><td>Kendall's W^a</td><td>,019</td></tr><tr><td>Chi-Square</td><td>,800</td></tr><tr><td>df</td><td>1,000</td></tr><tr><td>Asymp. Sig.</td><td>,371</td></tr></table> <p>a. Kendall's Coefficient of Concordance</p>	Ranks			Mean Rank	Avaliação2004	1,55	Avaliação2005	1,45	Test Statistics		N	42,000	Kendall's W ^a	,019	Chi-Square	,800	df	1,000	Asymp. Sig.	,371	<p>Não foi possível a realização do teste devido à não disponibilidade do respectivo package.</p>														
	Ranks																																			
		Mean Rank																																		
	Avaliação2004	1,55																																		
	Avaliação2005	1,45																																		
	Test Statistics																																			
	N	42,000																																		
	Kendall's W ^a	,019																																		
	Chi-Square	,800																																		
	df	1,000																																		
Asymp. Sig.	,371																																			

MAPA COMPARATIVO 2005- 2006

Processo	SPSS				R	
Teste						
Teste de Wilcoxon	Ranks				> wilcox.test(x,y,paired=TRUE,alternative="greater")	
	Avaliação2006 - Avaliação2005		N	Mean Rank	Sum of Ranks	Wilcoxon signed rank test with continuity correction
		Negative Ranks	8 ^a	30,00	240,00	data: x and y
		Positive Ranks	111 ^b	62,16	6900,00	V = 240, p-value = 1
		Ties	35 ^c			alternative hypothesis: true location shift is greater than 0
		Total	154			
	a. Avaliação2006< Avaliação 2005; b. Avaliação 2006 > Avaliação2005; c. Avaliação2006 = Avaliação2005					
	Test Statistics ^b					
			Avaliação2006 - Avaliação2005			
	Z		-9,048 ^a			
Asymp. Sig. (2-tailed)		,000				
a. Based on positive ranks. b. Wilcoxon Signed Ranks Test						
Teste dos Sinais	Frequencies				> wilcox.test(y-x,alternative="greater")	
	Avaliação2005 - Avaliação2006		N		Wilcoxon signed rank test with continuity correction	
		Negative Differences ^a	8		data: y - x	
		Positive Differences ^b	111		V = 6900, p-value < 2.2e-16	
		Ties ^c	35		alternative hypothesis: true location is greater than 0	
		Total	154			
	a.Avaliação2006 < Avaliação 2005; b. Avaliação 2006 > Avaliação2005; c. Avaliação2006 = Avaliação2005					
	Test Statistics ^b					
			Avaliação2006 - Avaliação2005			
	Exact Sig. (2-tailed)		-9,350			
a. Binomial distribution used. b. Sign Test						
Teste de Friedman	Ranks				>avaliação<- matrix(c(x,y),nrow=2,ncol=4,byrow=TRUE,dimnames=list(c("x","y"),c("c.1","c.2","c.3","c.4")))	
		Mean Rank			> friedman.test(avaliação)	
	Avaliação2005	1,17			Friedman rank sum test	
	Avaliação2006	1,83			data: avaliação	
	Test Statistics				Friedman chi-squared = 1.4, df = 3, p-value = 0.7055	
	N		154,000			
	Chi-Square		89,151			
	df		1,000			
	Asymp. Sig.		,000			
	a. Friedman Test					
Teste de Kendall	Ranks				Não foi possível a realização do teste devido à não disponibilidade do respectivo package.	
		Mean Rank				
	Avaliação2005	1,17				
	Avaliação2006	1,83				
	Test Statistics					
	N		154,000			
	Kendall's W ^a		,579			
	Chi-Square		89,151			
	df		1,000			
	Asymp. Sig.		,000			
a. Kendall's Coefficient of Concordance						

MAPA COMPARATIVO 2006- 2007

Processo	SPSS	R																																		
Teste																																				
Teste de Wilcoxon	<table><tr><th colspan="5">Ranks</th></tr><tr><td rowspan="5">Avaliação2007 - Avaliação2006</td><td></td><td>N</td><td>Mean Rank</td><td>Sum of Ranks</td></tr><tr><td>Negative Ranks</td><td>30^a</td><td>37,70</td><td>1131,00</td></tr><tr><td>Positive Ranks</td><td>41^b</td><td>34,76</td><td>1425,00</td></tr><tr><td>Ties</td><td>93^c</td><td></td><td></td></tr><tr><td>Total</td><td>164</td><td></td><td></td></tr></table> <p>a.Avaliação2007< Avaliação 2006; b. Avaliação 2007 > Avaliação2006; c. Avaliação2007 = Avaliação2006</p> <table><tr><th colspan="2">Test Statistics^b</th></tr><tr><td></td><td>Avaliação2007- Avaliação2006</td></tr><tr><td>Z</td><td>-,883^a</td></tr><tr><td>Asymp. Sig. (2-tailed)</td><td>,377</td></tr></table> <p>a. Based on positive ranks. b. Wilcoxon Signed Ranks Test</p>	Ranks					Avaliação2007 - Avaliação2006		N	Mean Rank	Sum of Ranks	Negative Ranks	30 ^a	37,70	1131,00	Positive Ranks	41 ^b	34,76	1425,00	Ties	93 ^c			Total	164			Test Statistics ^b			Avaliação2007- Avaliação2006	Z	-,883 ^a	Asymp. Sig. (2-tailed)	,377	<pre>>wilcox.test(x,y,paired=TRUE,alternative="greater")</pre> <p>Wilcoxon signed rank test with continuity correction</p> <p>data: x and y V = 1131, p-value = 0.8122 alternative hypothesis: true location shift is greater than 0</p>
	Ranks																																			
	Avaliação2007 - Avaliação2006		N	Mean Rank	Sum of Ranks																															
		Negative Ranks	30 ^a	37,70	1131,00																															
		Positive Ranks	41 ^b	34,76	1425,00																															
		Ties	93 ^c																																	
		Total	164																																	
	Test Statistics ^b																																			
		Avaliação2007- Avaliação2006																																		
	Z	-,883 ^a																																		
Asymp. Sig. (2-tailed)	,377																																			
Teste dos Sinais	<table><tr><th colspan="3">Frequencies</th></tr><tr><td rowspan="5">Avaliação2007- Avaliação2006</td><td></td><td>N</td></tr><tr><td>Negative Differences^a</td><td>30</td></tr><tr><td>Positive Differences^b</td><td>41</td></tr><tr><td>Ties^c</td><td>93</td></tr><tr><td>Total</td><td>164</td></tr></table> <p>a.Avaliação2007 < Avaliação 2006; b. Avaliação 2007 > Avaliação2006; c. Avaliação2007 = Avaliação2006</p> <table><tr><th colspan="2">Test Statistics^b</th></tr><tr><td></td><td>Avaliação2007 - Avaliação2006</td></tr><tr><td>Exact Sig. (2-tailed)</td><td>,235</td></tr></table> <p>a. Binomial distribution used. b. Sign Test</p>	Frequencies			Avaliação2007- Avaliação2006		N	Negative Differences ^a	30	Positive Differences ^b	41	Ties ^c	93	Total	164	Test Statistics ^b			Avaliação2007 - Avaliação2006	Exact Sig. (2-tailed)	,235	<pre>>wilcox.test(y-x,alternative="greater")</pre> <p>Wilcoxon signed rank test with continuity correction</p> <p>data: y - x V = 1425, p-value = 0.1894 alternative hypothesis: true location is greater than 0</p>														
	Frequencies																																			
	Avaliação2007- Avaliação2006		N																																	
		Negative Differences ^a	30																																	
		Positive Differences ^b	41																																	
		Ties ^c	93																																	
		Total	164																																	
	Test Statistics ^b																																			
		Avaliação2007 - Avaliação2006																																		
	Exact Sig. (2-tailed)	,235																																		
Teste de Friedman	<table><tr><th colspan="2">Ranks</th></tr><tr><td></td><td>Mean Rank</td></tr><tr><td>Avaliação2005</td><td>1,47</td></tr><tr><td>Avaliação2006</td><td>1,53</td></tr></table> <table><tr><th colspan="2">Test Statistics</th></tr><tr><td>N</td><td>164,000</td></tr><tr><td>Chi-Square</td><td>1,704</td></tr><tr><td>df</td><td>1,000</td></tr><tr><td>Asymp. Sig.</td><td>,192</td></tr></table> <p>a. Friedman Test</p>	Ranks			Mean Rank	Avaliação2005	1,47	Avaliação2006	1,53	Test Statistics		N	164,000	Chi-Square	1,704	df	1,000	Asymp. Sig.	,192	<pre>> friedman.test(avaliação)</pre> <p>Friedman rank sum test</p> <p>data: avaliação Friedman chi-squared = 0.8824, df = 3, p-value = 0.8297</p>																
	Ranks																																			
		Mean Rank																																		
	Avaliação2005	1,47																																		
	Avaliação2006	1,53																																		
	Test Statistics																																			
	N	164,000																																		
	Chi-Square	1,704																																		
	df	1,000																																		
	Asymp. Sig.	,192																																		
Teste de Kendall	<table><tr><th colspan="2">Ranks</th></tr><tr><td></td><td>Mean Rank</td></tr><tr><td>Avaliação2006</td><td>1,47</td></tr><tr><td>Avaliação2007</td><td>1,53</td></tr></table> <table><tr><th colspan="2">Test Statistics</th></tr><tr><td>N</td><td>164,000</td></tr><tr><td>Kendall's W^a</td><td>,010</td></tr><tr><td>Chi-Square</td><td>1,704</td></tr><tr><td>df</td><td>1,000</td></tr><tr><td>Asymp. Sig.</td><td>,192</td></tr></table> <p>a. Kendall's Coefficient of Concordance</p>	Ranks			Mean Rank	Avaliação2006	1,47	Avaliação2007	1,53	Test Statistics		N	164,000	Kendall's W ^a	,010	Chi-Square	1,704	df	1,000	Asymp. Sig.	,192	Não foi possível a realização do teste devido à não disponibilidade do respectivo package.														
	Ranks																																			
		Mean Rank																																		
	Avaliação2006	1,47																																		
	Avaliação2007	1,53																																		
	Test Statistics																																			
	N	164,000																																		
	Kendall's W ^a	,010																																		
	Chi-Square	1,704																																		
	df	1,000																																		
Asymp. Sig.	,192																																			

CAPÍTULO VII

Técnicas de visualização gráfica e Metodologias de Superfície de Resposta

7.1 Introdução

No decorrer deste trabalho tivemos oportunidade de lidar com uma grande base de dados o que nos permitiu adquirir a sensibilidade para a problemática de análise e tratamento de grandes volumes de informação.

Na actualidade, as técnicas de visualização em Estatística tornam-se imprescindíveis neste campo, tendo vindo a ser alvo da atenção de grupos de cientistas internacionais e assistindo-se a um crescente impacto na sua investigação.

O desenvolvimento de poderosas tecnologias e potentes recursos a nível computacional têm vindo a proporcionar avanços notáveis nesta área, especialmente na última década.

Pareceu-nos adequado, no âmbito do Mestrado em Estatística, Matemática e Computação, dedicar um ponto deste trabalho à investigação nesta área. Assim, procurámos fazer uma resenha histórica e enquadramento actual da temática, projectando vir a aplicar e explorar estas metodologias em trabalhos futuros.

7.2 Resenha histórica

As técnicas de visualização remontam a tempos imemoriais. As suas origens encontram-se nos diagramas geométricos, nas tabelas de posição das estrelas e nos mapas. Passamos a descrever, de uma forma resumida, alguns dos testemunhos deixados e que mostram a evolução destas técnicas.

Desconhece-se, mas pensa-se que o mapa seguinte é a representação mais antiga, cerca de 6200 a.c., da cidade da Babilónia, encontrado na região de Kirkuk, no Iraque. O original encontrava-se escavado em pedra:



No ano 550 a.c. aparecem descrições nos livros de Heródoto e Anaximandro de Mileto (610 a.c.- 546 d.c.), na Grécia, sobre o primeiro mapa-mundi (?):





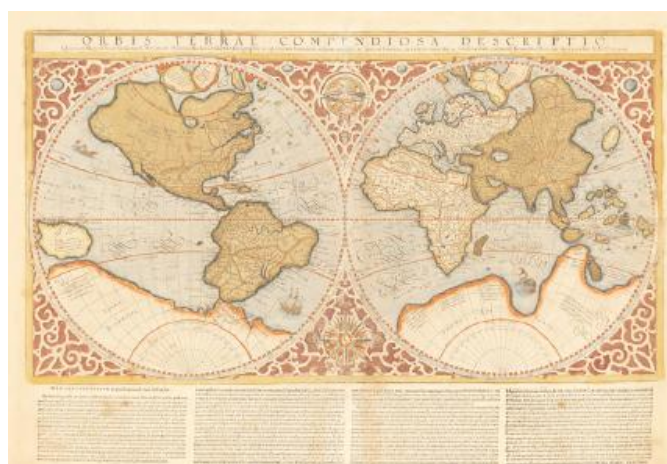
Sobejamente conhecido, no ano de 1500, Leonardo da Vinci (1452-1519), dá a conhecer os seus conhecimentos, através de ilustrações detalhadas, para retratar a anatomia, as máquinas e os movimentos:



Com a invenção da projecção cilíndrica para retratar o globo terrestre, em 1569, o belga Gerardus Mercator (1512 - 1594), também dividiu o seu mapa-mundi em páginas, para facilitar o seu manuseamento e estampou a imagem do deus Atlas na respectiva capa:

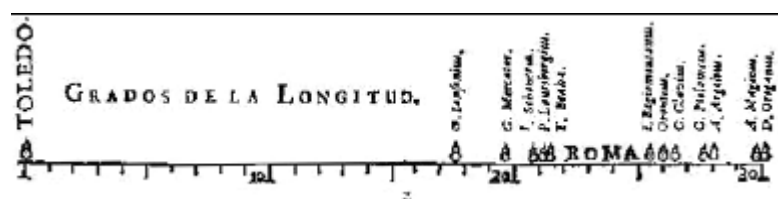


Mais tarde, em 1595, Rumold Mercator (1545-1599), filho de Gerardus, continua o seu trabalho e publica este mapa-mundi , um ano após a morte do pai:

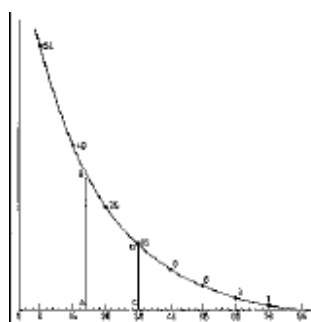


Já no século XVII, os maiores problemas referiam-se à medição física do tempo, da distância e do espaço para a astronomia, navegação e expansão territorial. Surgem, assim, as estimativas, as probabilidades, a demografia e todo o tipo de estatísticas. No fim do século, os elementos necessários para iniciar um “pensamento visual” encontram-se definidos.

Em 1644, encontra-se feita, por Michael van Langren's, a descrição das 12 referências de longitude de Toledo a Roma:

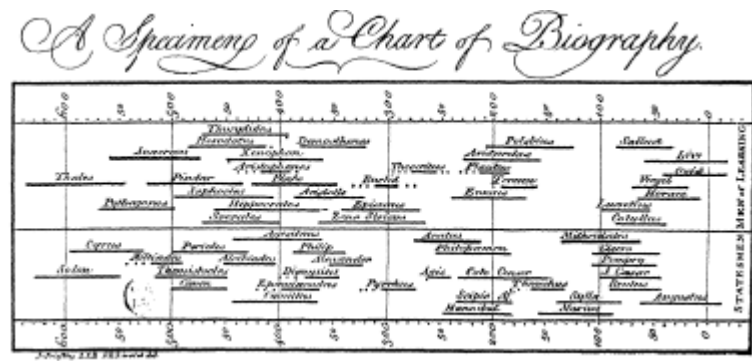


Em 1669, Christiaan Huygens (1629-1695), através dum gráfico de uma função, mostra como encontrar os anos de vida restantes de acordo com a idade actual:

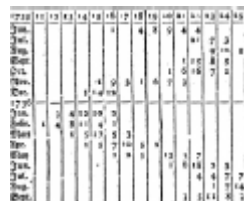


No século XVIII, aparecem novas fórmulas gráficas. Os cartógrafos começam a mostrar mais do que a posição geográfica. No fim do século aparecem tentativas de elaborar mapas com as informações sobre geologia, economia, demografia e saúde. À medida que o volume de dados aumenta, novas formas de visualização são criadas. As inovações tecnológicas como a cor, a litografia e a imprensa, abrem novos caminhos às técnicas de visualização.

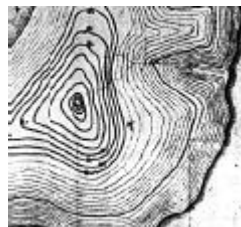
Em 1765, Joseph Priestley (1733 - 1804), Inglaterra, cria a linha do tempo de 2.000 pessoas famosas de 1200 a.c. até 1750. O tempo de vida foi representado por barras:



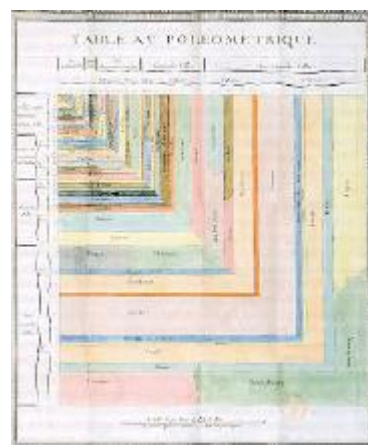
No ano de 1779, Heinrich Lambert (1728 -1777), Alemanha, efectua a análise gráfica de variação periódica na temperatura do solo:



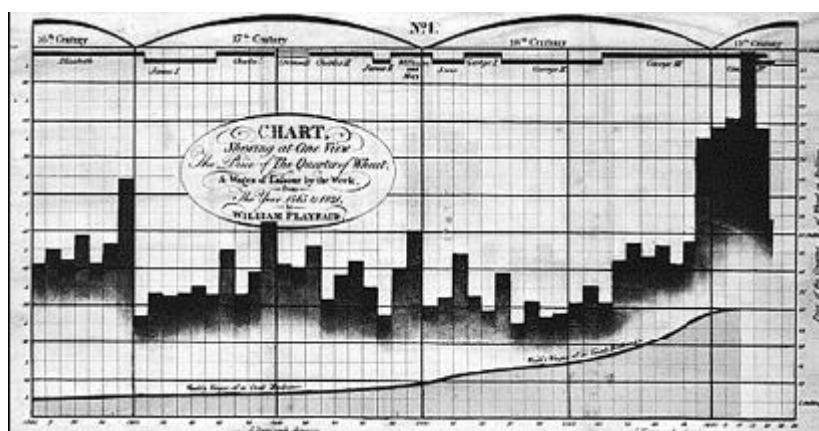
Em 1782, aparece o primeiro mapa topográfico, por Marcellin du Carla-Boniface, França:



Charles de Fourcroy, França, em 1782, faz o uso da geometria, figuras proporcionais (quadrados), para comparar dados demográficos em sobreposição:

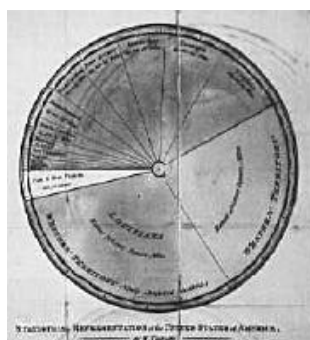


Em 1786, William Playfair (1759-1823), Inglaterra, elabora um gráfico de barras e de linhas com dados económicos:



O início da infografia moderna dá-se no século XIX. A primeira metade deste século foi responsável por uma explosão no crescimento de gráficos estatísticos e de mapas temáticos, graças às inovações obtidas no século anterior. Todas as formas de gráficos estatísticos conhecidos hoje, foram desenvolvidos nesta época. Na cartografia, os mapas simples foram transformados em atlas, complexos, baseados na grande variedade de dados disponíveis.

Em 1801, William Playfair (1759-1823), Inglaterra, aparece com o gráfico circular:



O primeiro mapa geológico da Inglaterra e do País de Gales constituiu um padrão para a cartografia geológica. Também conhecido como o “mapa que mudou o mundo”, foi desenhado em 1801, sendo a sua versão final de 1815 e foi seu autor, William Smith (1769-1839), Inglaterra:

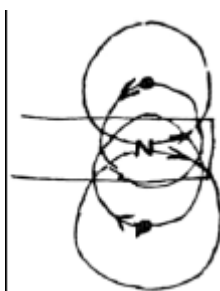


Em 1819, Baron Pierre Charles Dupin (1784 - 1873), França, criou o cartograma, mapa com sombras, a preto e branco, mostrando a distribuição do analfabetismo em França.

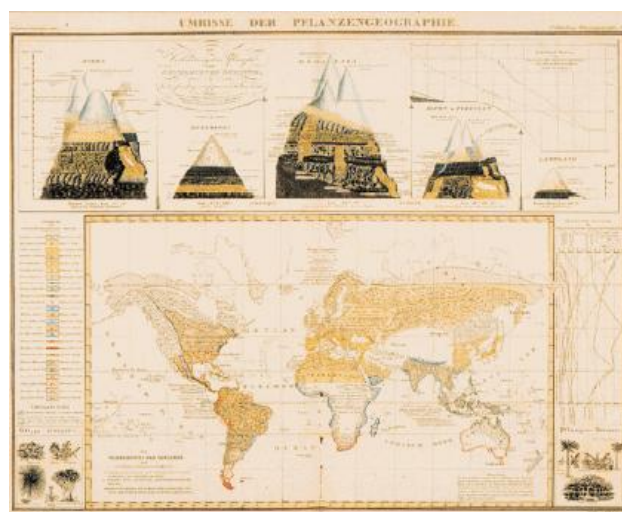
Talvez este, constitua o primeiro mapa estatístico:



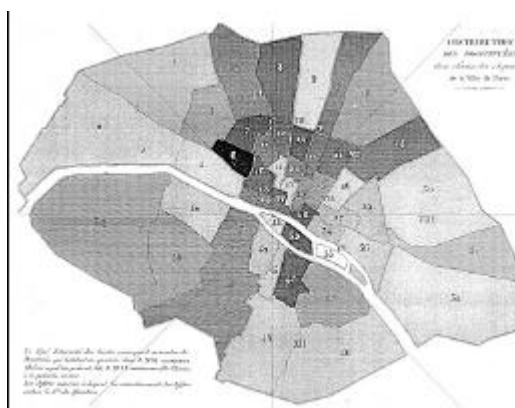
No ano de 1820, aparece um grande número de publicações científicas. Começam a inscrever-se gráficos e diagramas apenas para descrever os fenómenos naturais, como por exemplo, a variação magnética, o clima, etc. Michael Faraday (1791-1867), Inglaterra, é um desses autores.



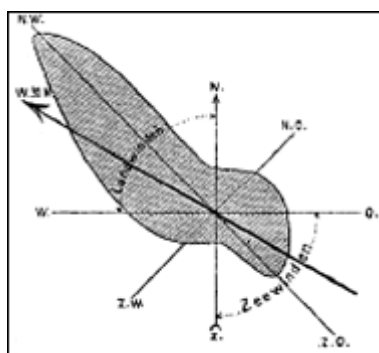
Heinrich Berghaus (1797 - 1884), Alemanha, publica em 1838, o atlas físico com a distribuição de plantas, animais, clima, etc. Trata-se de um dos mais extensivos e detalhados atlas temáticos incluindo tabelas e gráficos ilustrados:



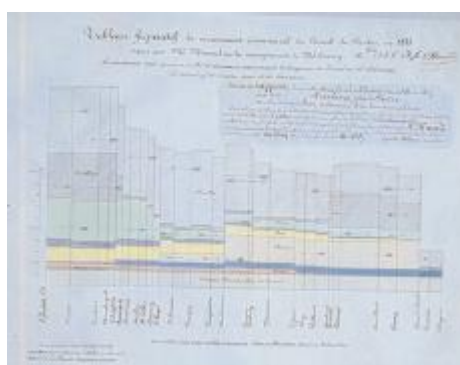
Em 1836, Alexandre Jean Baptiste Parent-Duchatelet (1790 - 1836), França, elabora o mapa das prostitutas em Paris:



Em 1843, Léon Lalanne (1811-1892), França, faz o uso das coordenadas polares num gráfico, representando a direcção da frequência dos ventos:



Charles Joseph Minard (1781-1870), França, elabora em 1844, o “Tableau-graphique” mostrando o transporte comercial com as variáveis distância (horizontal), quantidade (barras divididas) e custo por área:

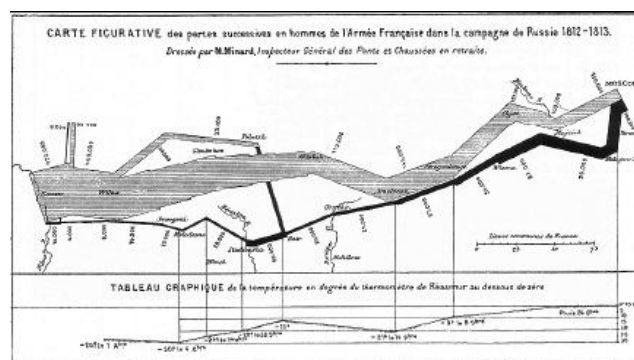


Por volta do ano de 1850 encontravam-se estabelecidas todas as condições para um rápido crescimento da visualização estatística. Os escritórios de análise estatística cresciam e espalhavam-se por toda a Europa motivados pelo aumento da importância das informações numéricas com o objectivo de efectuar o planeamento social, industrial, comercial e de transportes. A teoria estatística iniciada por Gauss e Laplace forneceu os meios para dar sentido a um grande número de dados. Foi a época de ouro das estatísticas.

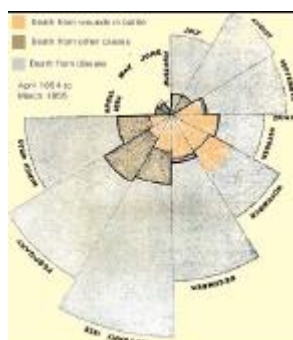
Em 1855, John Snow inscreveu num mapa de Londres, os casos da epidemia de cólera que assolou aquela cidade. Concluiu que todos os casos se concentravam próximos de um único local de abastecimento público de água, o que levou ao seu encerramento:



O engenheiro francês, Charles Minard (1781-1870), ilustrou graficamente, em 1869, a desastrosa campanha de Napoleão contra a Rússia em 1812. A largura do trajecto é proporcional ao número de soldados sobreviventes na campanha de guerra. Na cor cinza encontra-se representado o caminho de ida e, a preto, o caminho de regresso. Tratar-se-á do melhor gráfico estatístico já feito.



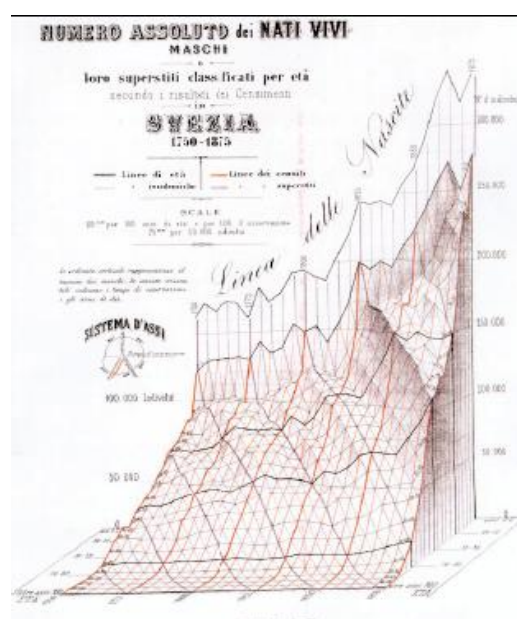
Em 1857, Florence Nightingale (1820-1910), Inglaterra, elaborou um gráfico que terá sido usado para justificar a melhoria das condições sanitárias do exército em campanha. Verificou que os mortos por doença eram mais numerosos do que os mortos devido a ferimentos em combate:



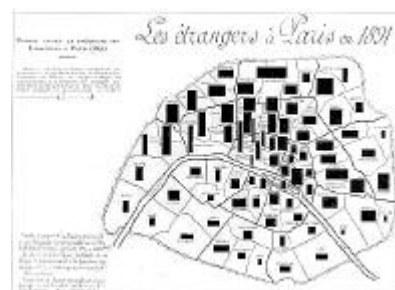
Francis Galton (1822-1911), inglês, elaborou em 1861, um mapa do estado do tempo mostrando as áreas de pressão através de ícones que se assemelham aos actuais:



Criado por Luigi Perozzo, em 1879, aparece o estereograma tridimensional da população, em forma de pirâmide e baseado nos dados do Censo Sueco:



Em 1896, Jacques Bertillon (1851 - 1922), França, fez o uso de rectângulos num mapa para representar duas variáveis e o produto entre elas (população dos arredores de Paris, percentagem de estrangeiros; área do rectângulo = nº absoluto de estrangeiros):

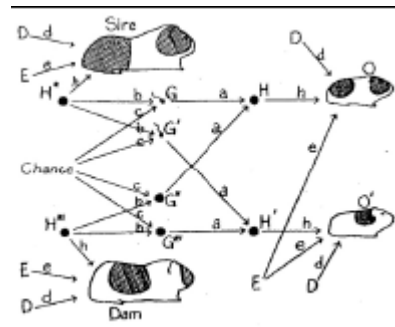


A primeira metade do século XX é conhecida pelo período negro dos gráficos de estatística. Poucas inovações gráficas aconteceram e o entusiasmo vivido no século anterior foi suplantado pelo crescimento da quantificação e modelos formais. Durante esse período, no entanto, tudo o que se alcançou consegue divulgar-se, consagrando a visualização gráfica para explicar as novas descobertas e teorias.

Em 1911 realiza-se a primeira exposição de Higiene em Dresden, Alemanha, organizada por Emil Eugen Roesle . Reúne 259 figuras gráficas estatísticas de 35 expositores nacionais e internacionais:



Sewall Wright (1889-1988), EUA, inventou, em 1920, do diagrama em setas para mostrar a relação entre variáveis, formando um sistema estrutural:



Em 1924, Otto Neurath (1882-1945), organiza o Museu Social de Gráficos Estatísticos em Viena, Áustria e introduz o Isotype (International System of Typographic Picture Education):



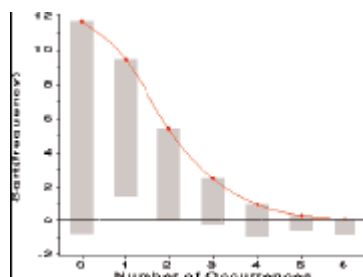
Em 1933, o Mapa do metropolitano de Londres tornou-se um símbolo reconhecido mundialmente, pela sua simplicidade e clareza. Harry Beck baseou a construção do mapa num diagrama de circuitos que ele utilizava no seu dia-a-dia:



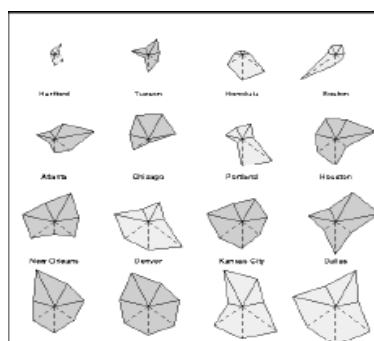
Na segunda metade do século XX, esboça-se o renascimento da visualização, quer em termos estatísticos, quer noutros campos da actividade e da ciência.

A visualização de dados começa a sair da dormência em que se encontrava por volta de 1960. Nos EUA, John W. Tukey reconhece a importância da análise gráfica dos dados e lança novos padrões e inovações. Na França, Jacques Bertin publica a sua *Semiologia Gráfica*, organizando a visão e a percepção dos elementos gráficos. Por último, os computadores começam a mostrar seu potencial.

Em 1965, verifica-se uma evolução nos histogramas de análises com novas propostas de comparação visual, como é proposto por John W. Tukey (1915-2000).



Em 1971, J. H. Siegel, R. M. Goldwyn and Herman P. Friedman, EUA, criam o polígono irregular para representar a multivariabilidade de dados (com vértices de iguais intervalos, distanciados do centro proporcionalmente ao valor da variável):

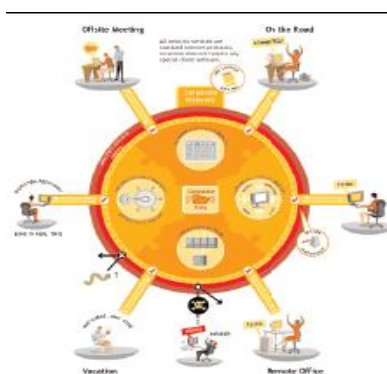


Desde 1975 e até aos dias de hoje, o computador é encarado como uma nova fronteira.

As inovações ocorridas neste período foram muitas e em diversas áreas: o desenvolvimento de softwares e sistemas informáticos, altamente interactivos e de fácil manipulação, foram a alavanca para o desenvolvimento gráfico. Os novos paradigmas de manipulação de dados, a invenção de técnicas gráficas e os métodos de visualização multidimensionais, deixaram os seus registos e apontam para o desenvolvimento futuro destas técnicas.

Assim, e apenas a título exemplificativo, passamos a descrever alguns pontos que achamos de interesse, para ilustrar este período da história das técnicas de visualização estatística.

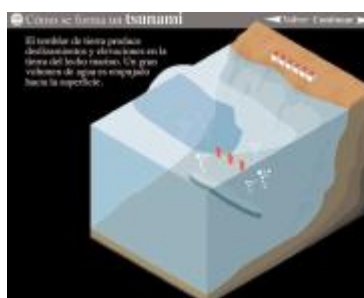
Em 1993, foi criada nos EUA a Xplane. Uma empresa especializada em visualização de dados relativa ao comércio e administração de empresas:



Em 1996, Jason Dykes, Reino Unido, cria um Software de dados cartográficos, com ferramentas que permitem vários tipos de representações de mapas em simultâneo:

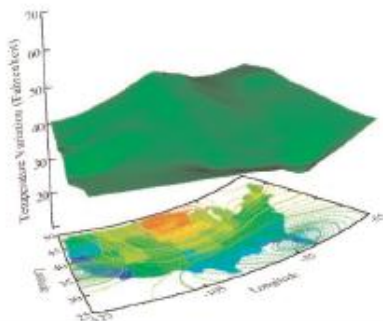


Em 1999, Leland Wilkinson escreve “A Gramática dos Gráficos. Trata-se de uma espécie de dicionário sistemático e de fácil compreensão com regras gramaticais para dados e gráficos. O mapa seguinte, elaborado por Leland é contornado em 3D em estrutura computadorizada:



Por fim, em 1999, dá-se a explosão da Internet. As novas tecnologias permitem que a visualização passe a ser interactiva e/ou animada.

Leland Wilkinson, *The Grammar of Graphics*, Springer-Verlag, 1999, Figure 8.11



7.3 As técnicas de visualização na actualidade

O rápido desenvolvimento de novas tecnologias e o decréscimo do custo do hardware proporciona um aumento do poder de processamento e da capacidade de armazenamento dos computadores. A proliferação da disponibilidade computacional associada aos avanços nas comunicações, particularmente ao crescimento da *World Wide Web*, resulta numa verdadeira explosão da quantidade de informação disponível. A procura da solução dos problemas associados com essa nova classe de dados e as suas respectivas tarefas de análise e interpretação fez emergir uma nova área: a visualização de informações que combina aspectos da visualização científica, de interfaces homem - máquina, de proliferação de dados e de computação gráfica.

Ao contrário da visualização científica, que possui a sua concentração na representação visual de dados, obtidos a partir de simulações ou na recolha de dados físicos, a visualização de informações, tem como elemento principal, os dados abstractos. Isto é, dados que não possuem informações intrínsecas de geometria. Dessa forma, um dos problemas investigados pela comunidade científica é a descoberta de metáforas visuais para os representar e o estudo de mecanismos que servirão de análise ao suporte dessas representações. Em contraste com a visualização científica, que geralmente é utilizada por cientistas, as interfaces, criadas para a manipulação de informações, podem ser utilizadas por uma comunidade leiga e diversificada, que possui diferentes níveis de educação, capacidades e necessidades.

Uma das técnicas que tem sido desenvolvida recentemente, no âmbito da Estatística, consiste na metodologia de superfície de resposta, que passamos a analisar.

Assim, no contexto do planeamento de experiências, é conhecido como principal objectivo, encontrar a solução óptima de um problema ou ainda, caracterizar a relação entre uma ou mais variáveis resposta num conjunto de factores de interesse.

De facto, a análise de dados, é apenas uma pequena parte da Estatística e tão ou mais importante quanto a análise é o planeamento de experiências que vai fornecer os dados a estudar. Um

planeamento adequado minimiza os custos operacionais e garante que os resultados conterão algumas informações relevantes para a solução do problema.

Como se afirmou anteriormente, na maior parte das situações, o objectivo principal do investigador concentra-se na optimização de um sistema, procurando maximizar ou minimizar algum tipo de resposta. Esta resposta constitui a função - objectivo do problema de optimização ao mesmo tempo que o sistema obedece a alguns critérios considerados como as restrições do problema.

Uma das técnicas estatísticas mais apropriada para a obtenção da solução de problemas de optimização é a metodologia de superfícies de resposta (MSR), do inglês RSM - *Response Surface Methodology*. A metodologia de superfície de resposta não é mais do que um conjunto de técnicas estatísticas que têm como um dos principais objectivos, encontrar os níveis dos factores que optimizam a resposta, representados por uma estimativa de um modelo ajustado.

Um planeamento factorial requer a execução de experiências em todas as possíveis combinações dos níveis dos factores (ou variáveis). Assim, para estudar o efeito de qualquer factor sobre a resposta é preciso fazê-lo variar, obedecendo os seus níveis, e observar o resultado dessa variação na resposta.

Podemos afirmar, também, que a metodologia de superfície de resposta é uma técnica de optimização baseada no emprego de planeamentos factoriais, introduzida por G.E.P. Box na década de 1950. Desde então, tem sido usada em muitas áreas, destacando-se um grande sucesso na modelação de diversos processos industriais.

Assim, e das muitas técnicas de visualização estatística usadas na actualidade, iremos procurar uma breve abordagem à Metodologia de Superfície de Resposta, dedicando-lhe o último sub-capítulo e perspectivando futuras aplicações.

7.4 Metodologia de Superfície de Resposta (MSR)

Como dissemos, a metodologia de superfície de resposta (MSR) pode ser definida como um conjunto de técnicas matemáticas e estatísticas usadas na análise e na modelação de problemas em que a resposta de interesse é influenciada por vários factores de carácter quantitativo e tem por objectivo, a maximização dessa resposta.

O valor esperado (η), assumido pela variável de interesse, encontra-se influenciado pelos níveis dos k factores quantitativos, x_1, x_2, \dots, x_k e pressupõe a existência de uma função e em que a variável resposta pode tomar a seguinte expressão:

$$y = f(x_1, x_2, \dots, x_k) + \varepsilon,$$

em que ε representa o erro ou os resíduos observados na resposta y .

Se identificarmos a resposta esperada por

$$E(y) = f(x_1, x_2, \dots, x_k) = \eta,$$

a superfície é identificada por

$$\eta = f(x_1, x_2, \dots, x_k)$$

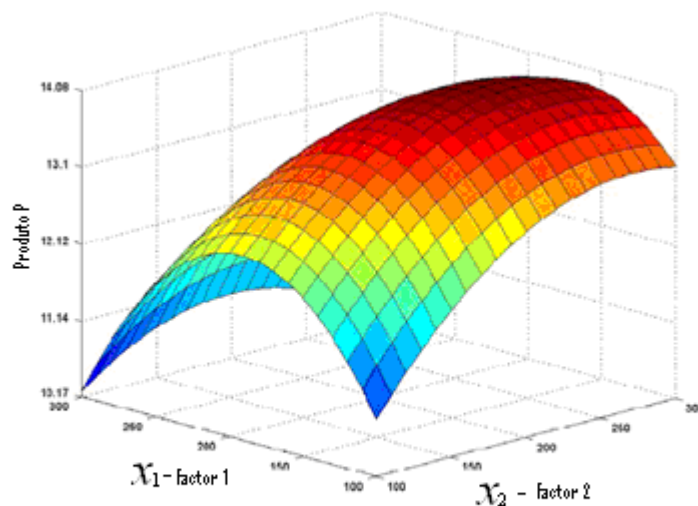
e representa-se através de uma hiper superfície, constituindo um subconjunto de um espaço euclédiano a que chamamos superfície de resposta.

Uma das técnicas utilizadas para visualizar a forma de uma superfície de resposta tridimensional, consiste na representação do gráfico de contornos da superfície onde são traçadas as denominadas linhas de contorno, que, não são mais do que curvas correspondentes a valores constantes da resposta sobre o plano (x_1, x_2) .

Os eixos coordenados deste plano são dados pelos níveis x_1 e x_2 dos factores. Geometricamente, cada linha de contorno é uma projecção sobre o plano (x_1, x_2) de uma secção da superfície de resposta ao interceptar um plano paralelo a $(x_1$ e $x_2)$.

Com base em apresentamos um exemplo de ilustração.

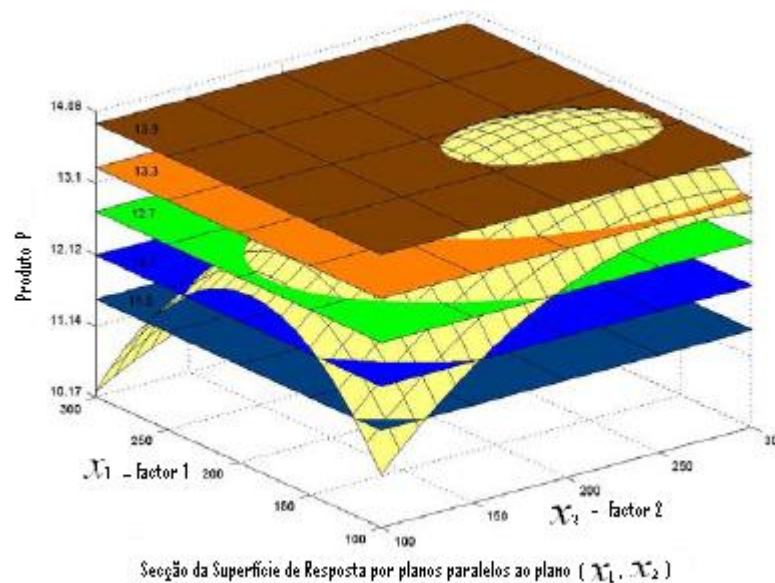
Pretendendo maximizar a produção de um determinado produto P em função de dois factores, Factor 1 e Factor 2, uma superfície de resposta que poderá determinar a produção óptima deste produto pode ser dada pelo gráfico seguinte:



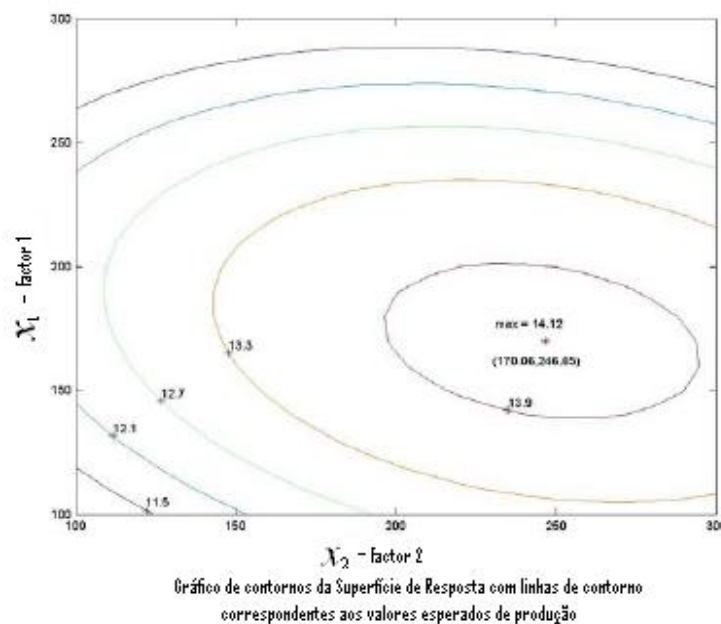
Superfície de Resposta tridimensional onde se pode observar a produção esperada do Produto $P(\eta)$ em função dos factores 1 (X_1) e 2 (X_2)

Neste exemplo, e dado que existem dois factores de influência ($k=2$), a superfície de resposta visualiza-se num espaço tridimensional, em que a terceira dimensão representa a produção esperada do produto P sobre o plano bidimensional definido pelas combinações dos níveis dos factores x_1 e x_2 .

Para elaborar o correspondente gráfico de contornos, basta seccionar a superfície de resposta usando planos paralelos a x_1 e x_2 de acordo com o pretendido. A superfície de resposta anterior poderia ser seccionada como mostramos na figura seguinte:



Cada linha de contorno representa um número infinito de combinações dos valores dos factores 1 e 2 para a produção do produto P. A produção máxima localiza-se no centro da elipse mais pequena e, no caso do exemplo, corresponde aos níveis $x_1 = 170,06$ para o Factor 1 e $x_2 = 246,65$ para o Factor 2.



Na maior parte dos problemas de MSR, a relação entre a resposta e as variáveis independentes é desconhecida. Assim, o primeiro passo a seguir na MSR será encontrar a aproximação possível

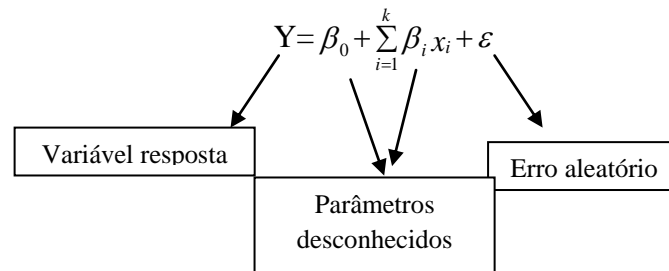
para a verdadeira relação funcional entre y e o conjunto das variáveis independentes. Normalmente é usado um polinómio de primeira ordem:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

Um modelo de primeira ordem com k factores

$$(x_1, x_2, \dots, x_k)$$

pode ser representado, genericamente, por:



ou, na forma matricial:

$$y = X \beta + \varepsilon$$

em que a matriz X se pode escrever como $X = [1:D]$, sendo D , a matriz das combinações dos níveis dos factores e se designa por matriz de desenho. Por outras palavras, podemos dizer que esta matriz é um vector de k variáveis:

$$X = [x_1, \dots, x_k]$$

Assim, o estimador de β obtido pelo método dos mínimos quadrados, será dado por:

$$b = (X'X)^{-1} X' y$$

e a matriz das variâncias-covariâncias de b , dada por:

$$\text{Var}(b) = (X'X)^{-1} \sigma^2$$

Temos, então, o modelo ajustado de primeira ordem:

$$\hat{y} = b_0 + \sum_{i=1}^k b_i x_i$$

Se o modelo se encontrar ajustado, a parte não aleatória do modelo representa a resposta real ajustada e ε é o erro experimental. No caso de ser mais adequada a adopção de um polinómio de grau superior, pode procurar aplicar-se um modelo de segunda ordem, com a estrutura que se segue:

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_i x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j + \varepsilon$$

em que y representa a variável resposta; β_0 , β_i e β_{ij} são parâmetros desconhecidos e ε é o erro aleatório.

O processo utilizado para o modelo ajustado de primeira ordem, pode agora ser aplicado aos modelos de segunda ordem. Assim, o modelo ajustado de segunda ordem traduz-se por:

$$\hat{Y} = b_0 + \sum_{i=1}^k b_i x_i^2 + \sum_{i=1}^{k-1} \sum_{\substack{j=2 \\ j>i}}^k b_{ij} x_i x_j$$

A escolha de um delineamento adequado da experiência a realizar é fundamental para modelar e explorar a superfície de resposta usada no ajustamento de um modelo polinomial ao conjunto de dados nos pontos do desenho.

Assim devemos:

- Gerar uma distribuição razoável de pontos (de informação), em toda a região de interesse, mas utilizando o menor número de pontos experimentais.
- Assegurar que, para cada ponto x , o valor ajustado $\hat{y}(x)$ se encontra o mais próximo possível do valor real $Y(x)$.
- Permitir a detecção de falta de ajustamento no modelo.
- Permitir a execução das experiências em blocos.
- Permitir a construção sequencial de desenhos, por ordem crescente.
- Proporcionar uma estimação interna da variância do erro.
- E, por fim, assegurar, de forma simples, os cálculos das estimações dos parâmetros do modelo.

No âmbito da nossa amostra e tendo em atenção que a maior parte das variáveis são do tipo qualitativo, não nos foi possível, em tempo útil, idealizar uma aplicação das MSR, ficando em aberto para futuras investigações.

Considerações e Conclusões

Recorrendo à Estatística e com a ajuda da computação e recurso aos softwares R e SPSS, com esta dissertação procurámos efectuar um estudo sobre o CNO dos CTT com base nos trabalhadores que frequentaram o processo RVCC.

Para o efeito, escolheu-se a população que frequentou o processo ao nível do B3 (9º ano), atendendo às variáveis que se consideraram mais importantes, nomeadamente a idade, género, habilitações literárias, duração do processo, região, situação, afectação e categoria profissional.

O presente trabalho considera também a análise da avaliação de desempenho desta população, antes e depois da frequência do processo RVCC, com o objectivo de verificar em que medida este processo contribuiu para uma melhoria da avaliação de desempenho desses trabalhadores.

Devido a questões relacionadas com o limite de tempo para a realização de uma dissertação de mestrado, e também devido à enorme quantidade de dados e variáveis disponíveis, não foi possível efectuar um trabalho mais globalizante envolvendo a certificação ao nível do ensino secundário (12º ano) e a efectuar uma análise dos dados disponíveis ao nível de Metodologias de Superfície de Resposta ficando como sugestão para trabalho futuro.

As metodologias estatísticas utilizadas basearam-se em análises descritivas e comparativas, por regiões, e em análises estatísticas não paramétricas para amostras emparelhadas.

Para o tratamento dos dados foram utilizadas as ferramentas informáticas SPSS (versão 16.0) e a programação em R.

Assim, e com base no trabalho desenvolvido, podemos evidenciar algumas conclusões que se consideram mais relevantes, e que passamos a indicar:

- 1 - As médias das idades dos adultos nas regiões de Lisboa e Porto são idênticas e iguais à média verificada nas três regiões consideradas em conjunto (Lisboa, Porto e Coimbra) que é de 44 anos.
- 2 - A média das idades na região de Coimbra é um pouco inferior (42 anos e meio).
- 3 - Nas regiões de Lisboa e Porto, 50% dos adultos têm idades abaixo dos 45 anos e 50% têm idades superiores àquela, o mesmo se passando quando consideradas as três regiões indicadas em conjunto. Em Coimbra aquela idade é um pouco inferior (40 anos).
- 4 - A maioria dos adultos tem 37 anos e meio sendo idêntica a distribuição nas três regiões indicadas.
- 5 - Os nossos dados revelam distribuições assimétricas positivas, embora a distribuição de Coimbra apresente um desvio padrão inferior o que se traduz numa menor diferença de idades entre os adultos.

6 - As médias da duração do processo nas regiões de Lisboa (94 dias), Porto (140 dias) e Coimbra (329 dias) são distintas da média das três regiões (cerca de 130 dias). Verifica-se ainda que os adultos na região de Coimbra demoraram o triplo do tempo da média das três regiões.

7 – 50% dos adultos demoraram até 135 dias a concluir o processo nas regiões de Lisboa e Porto. Na região de Coimbra, demoraram até 225 dias.

8 - A maioria dos adultos demorou 135 dias a concluir o processo nas regiões de Lisboa e Porto. Na região de Coimbra, a maioria demorou 225 dias.

A análise da situação no processo leva-nos a afirmar que:

1 - Dos trabalhadores afectos ao SNC, apenas 2 desistiram, 3 encontram-se em reconhecimento e 3 continuam inscritos sem iniciar o processo. Todos os outros concluíram o processo.

2 - Dos trabalhadores afectos à OPE verifica-se que 3 trabalhadores foram transferidos, 4 desistiram, 3 encontram-se em reconhecimento e 3 encontram-se inscritos a aguardar o início do processo. Os restantes trabalhadores concluíram o processo.

3 - Os trabalhadores dos CTT-EXP concluíram o processo com excepção de 5, encontrando-se 3 em reconhecimento e 2 apenas inscritos.

Relativamente às habilitações literárias destes trabalhadores verifica-se que apenas 3, 4 e 2 respectivamente em Lisboa, Porto e Coimbra possuem o 1º Ciclo do Ensino Básico. Todos os outros possuem o 2º Ciclo do Ensino Básico.

Analisando as três regiões em conjunto, verificámos que:

A grande maioria dos adultos (45,6%), demora cerca de quatro meses e meio a concluir o processo. Por outro lado, 9,6% dos adultos demora mais de ano e meio, enquanto que 6,2% o fazem em menos de três meses. Verifica-se, também, que a mediana e a moda assumem o mesmo valor. Relativamente à idade, verifica-se que 49,1% dos adultos (cerca de metade) têm idades compreendidas entre os 35 e os 45 anos. A análise dos dados destas três regiões também permite afirmar que 70% dos adultos têm menos de 45 anos e que 88,2% têm menos de 50 anos de idade.

Podemos igualmente afirmar que a média de idades dos adultos é de cerca de 44 anos, a maioria tem 37 anos e 50% dos adultos tem mais de 42 anos e meio.

Feita a análise comparativa das estatísticas descritivas de cada região e a nível nacional podemos concluir que, relativamente às idades dos adultos:

1 - As médias das idades dos adultos nas regiões de Lisboa e Porto são superiores à média verificada a nível nacional. A média verificada em Coimbra é inferior á média nacional. Como dissemos, é a população mais jovem.

2 – 50% dos adultos, nas regiões de Lisboa e Porto, têm menos de 42 anos e meio, o que se verifica também na totalidade da população, a nível nacional. Na região de Coimbra verifica-se que a idade é um pouco inferior (40 anos).

3 - A maioria dos adultos, nas regiões consideradas, tem 37 anos e meio, verificando-se o mesmo a nível nacional.

4 – Verifica-se que estamos perante distribuições assimétricas positivas, embora a distribuição de Coimbra apresente um desvio padrão inferior, o que se traduz numa menor diferença de idades entre os adultos.

Relativamente às avaliações de desempenho, verificamos que:

1 - As avaliações de “1-Muito Bom”, em 2005, correspondem a avaliações inferiores em 2004.

2 - As avaliações de “2” em 2005, correspondem a avaliações de “2” e inferiores, em 2004.

3 - As avaliações de “3” em 2005, correspondem a avaliações “2, 3, 4 e 5”, em 2004.

4 - As avaliações de “4” em 2005, correspondem a avaliações “4 e 5”, em 2004.

Concluindo, no biénio 2004 – 2005, verificamos que os trabalhadores foram objecto de uma avaliação de desempenho superior após concluírem o processo RVCC.

Comparando as avaliações de desempenho efectuadas em 2005 e 2006, verificamos que:

1 - As avaliações “1-Muito Bom” em 2006, correspondem a avaliações de “1 e 2”, em 2005.

2 - As avaliações de “2” em 2006, correspondem a avaliações “1, 2 e 3”, em 2005.

3 - As avaliações de “3” em 2006, correspondem a avaliações “2 e 3”, em 2005.

4 - As avaliações de “4 e 5” em 2006, correspondem, respectivamente, a avaliações de “1, 2, 3, 4 e 5”, em 2005.

Os resultados encontrados não nos permitem afirmar que a avaliação de desempenho em 2006 foi superior a 2005. Encontra-se como justificação, o facto de, neste ano, terem sido introduzidas as “quotas” na avaliação de desempenho.

Comparando agora, as avaliações de desempenho efectuadas em 2006 e 2007, verificamos que:

1 - As avaliações de “1-Muito Bom” em 2007, correspondem a avaliações de “1 e 2”, em 2006.

2 - As avaliações de “2”, em 2007, correspondem a avaliações “1, 2, 3 e 4”, em 2006.

3 - As avaliações de “3”, em 2007, correspondem a avaliações “1, 2, 3 e 4”, em 2006.

4 - As avaliações de “4” e as avaliações de “5” em 2007, correspondem, a avaliações “3, 4, 5 e 6” em 2006.

5 - As avaliações de “6” em 2007 correspondem a avaliações “5 e 6” em 2006.

Também, sobre estas avaliações, não se pode concluir que houve uma melhoria da avaliação de desempenho em 2007. Possivelmente este resultado conclusivo sofreu enviesamento devido às “quotas” impostas neste ano.

Concluindo, podemos afirmar que:

- Os trabalhadores mais velhos foram os que efectuaram o processo RVCC e viram a sua certificação ao nível do B3 num mais curto espaço de tempo.
- Os trabalhadores com menos habilitações literárias foram os que efectuaram o processo RVCC e viram a sua certificação ao nível do B3, também, num mais curto espaço de tempo.
- Os trabalhadores que frequentaram o processo RVCC e viram certificadas as suas competências a nível do B3, obtiveram uma melhor avaliação em 2005 do que a obtida em 2004. Relativamente às avaliações efectuadas nos anos seguintes (2006 e 2007) não podemos concluir da mesma forma devido à introdução de “quotas” e consequente enviesamento de resultados.

Perspectivas Futuras

Como perspectivas futuras e no seguimento deste trabalho, pensamos desenvolver este tema, no âmbito do CNO dos CTT, de uma forma mais alargada, no sentido de:

- 1- Construir um algoritmo que justifique a procura da certificação em B3 (Ensino Básico) e no Ensino Secundário, em função de algumas variáveis a definir.
- 2 - Determinar a influência da região de localização do candidato na escolha da certificação dos dois níveis de ensino.
- 3 – Determinar, em que medida, o género dos candidatos influencia a escolha do nível de certificação.
- 4 - Determinar a importância e o impacto da formação no desenvolvimento de competências e sua relação com as principais variáveis em estudo.
- 5 - Explorar a importância da visualização gráfica na Estatística.
- 6 - Explorar o uso das Metodologias de Superfície de Resposta aplicadas ao estudo da população certificada nos dois níveis de ensino.
- 7 - Dada a sua importância, explorar o programa R para aplicações específicas, tendo em vista o ensino on-line e a contemplação de futuros programas de e-learning.

Bibliografia Utilizada

1. AGRESTI, A. (2007), An Introduction to Categorical Data Analysis, Second Edition, New York: John Wiley & Sons
2. ALALUF, M.; STROOBANTS, M.(1994): “A competência mobiliza o operário?”, Formação Profissional, Nr. 1/1994, pp.46-55
3. ALMEIDA, P., REBELO, G., (2004): A Era da Competência – Um novo paradigma para a Gestão de Recursos Humanos e o Direito do Trabalho, Editora RH, Lisboa
4. AMBRÓSIO, T. (2000): Conhecimento pedagógico e as competências formativas dos professores/investigadores universitários, in ANAIS - Educação e Desenvolvimento; Unidade de Investigação Educação e Desenvolvimento
5. AMBRÓSIO, T. (2000): Novo Contrato entre a Universidade e a Sociedade, in ANAIS - Educação e Desenvolvimento; Unidade de Investigação Educação e Desenvolvimento
6. AMBRÓSIO, T. (2000): Pensamento Complexo e Organização do Conhecimento, in ANAIS - Educação e Desenvolvimento; Unidade de Investigação Educação e Desenvolvimento
7. BARANÃNO, A. (2004): Métodos e Técnicas de Investigação em Gestão; Edições Sílabo, Lisboa
8. BASILEVSKY, A. (2006). Applied Matrix Algebra in Statistical Sciences. Dover Publications.
9. BILHIM J. (2004): Gestão Estratégica de Recursos Humanos, Instituto Superior de Ciências Sociais e Políticas da Universidade Técnica de Lisboa, Lisboa
10. BOYATZIS, R. (1982): The Competent Manager - A Model for Effective Performance, John Wiley & Sons, New York
11. BRUNHEIRA, L.; LOUREIRO C.; OLIVEIRA F. (2000): Ensino e Aprendizagem da Estatística. 1ªEd. Edição: Sociedade Portuguesa de Estatística; Associação de Professores de Matemática; Departamento de Educação e Estatística e Investigação Operacional da Faculdade de Ciências da Universidade de Lisboa
12. CABRAL J. e GUIMARÃES R. (1997). Estatística. Mc–Graw-Hill, Lisboa.
13. CALINSKI, T. (1986). Linear Statistical Inference (Lecture Notes in Statistics).

-
14. CÂMARA, P.; GUERRA, P.; RODRIGUES, J.; (2003): Humanator - Recursos Humanos & Sucesso Empresarial, 5ª Edição, Dom Quixote, Lisboa
 15. CANÁRIO, R. (2000): Educação de Adultos - Um Campo e uma Problemática, Educa, Lisboa
 16. CAPELLE; J.(1970): L'éducation permanente en Europe, in Cahiers JEB, nº. 4, , Pág. 258-304 (Relatório apresentado ao Conselho da Europa)
 17. CASCÃO, F. (2004): Entre a Gestão de Competências e a Gestão do Conhecimento: Um estudo exploratório de Inovações na Gestão das Pessoas, Editora RH, Lisboa
 18. CAVACO, C. (2007): Reconhecimento, validação e certificação de competências: complexidade e novas actividades profissionais., Revista de Ciências da Educação, nº2 Jan/2007.
 19. CONOVER, W. (1980): Practical Nonparametric Statistics, 2nd ed., John Wiley & Sons
 20. CONOVER, W. J. and Iman, R. L. (1981), "Rank Transformations as a Bridge between Parametric and Nonparametric Statistics," The American Statistician, 35, 124–129.
 21. CONSELHO NACIONAL DE EDUCAÇÃO (1997): Educar e Formar ao Longo da Vida. Actas do Seminário realizado em 20 de Novembro de 1996 no Conselho Nacional de Educação, Lisboa
 22. CARMO, H. e FERREIRA, M. M. (1998). Metodologia da Investigação: Guia para Auto – Aprendizagem. Universidade Aberta. Lisboa.
 23. COX, D. R. (2006). Principles of Statistical Inference. Cambridge: Cambridge University Press.
 24. COX, T. F. (2005). An Introduction to Multivariate Data Analysis. London: Hodder Arnold.
 25. DALGAARD, P. (2002). Introductory Statistics with R. New York: Springer
 26. EVERITT, BRIANS and HOTHORN, TORSTEN (2005), A Handbook of Statistical Analyses Using R, London and Erlangen
 27. FORRESTER, K. (1998): Adult Educations: A Key for the Twenty- first Century: Reflections on the UNESCO fifth internacional conference 1997. Keith Forrester. In Internacional Journal of Lifelong Education, 17 (6) (November- December 1998:425-434)

-
28. GALBRAITH, J.K. (2000): Europe unemployment: what is the solution?, comunicação apresentada no seminário Towards a learning society: innovation and competence building with social cohesion for Europe, 23-28 de Maio 2000, Lisboa
 29. GUIMARÃES, R. C. e CABRAL, J. A. S. (1997). Estatística. Ed. revista. Lisboa, McGraw-Hill.
 30. HERMANO, C. E FERREIRA, M. (1998): Metodologia da Investigação – Guia para a Auto-aprendizagem, 1ª edição, Universidade Aberta, Lisboa
 31. HILL, M.M., HILL, A. (2002): Investigação por Questionário, 2ª edição, Edições Sílabo, Lisboa
 32. IMAN, R. L. and Conover, W. J. (1979), "The Use of the Rank Transform in Regression," Technometrics, 21, 499–509.
 33. JARVIS, P. (1999): Internacional Dictionary of Adults Education and Continuous Education, Kogan Page, London
 34. JOSSO, C. (1991): L'expérience formatrice: un concept en constrution, in La Formation Expérientielle dès Adultes, La Documentation Française, Paris
 35. KOVÁCS, I. (1998): Formação, Trabalho e Tecnologia – Para uma nova cultura organizacional, Celta Editora, Oeiras
 36. KOVÁCS, I. (1998): Qualificação, formação e empregabilidade, Sociedade e Trabalho, nº.4
 37. LICHTENBERGER, Y. E PARADEISE C. (2001): Compétence, compétences, Sociologie du Travail, nº1, pp. 33-48
 38. MAROCO, J. (2003). Análise Estatística – com utilização do SPSS. 2ª Edição. Lisboa, Edições Sílabo.
 39. MAROCO, J. e Bispo, R. (2003). Estatística Aplicada às Ciências Sociais e Humanas. Lisboa, Climepsi Editores.
 40. MENARD, S. (2002). Applied Logistic Regression Analysis, 2nd Edition. London: Sage Publications.
 41. MURTEIRA B.; RIBEIRO C.; SILVA J.; PIMENTA C. (2002): Introdução à Estatística, Mc-Graw-Hill, Lisboa

-
42. MURTEIRA B.; BLACK G. (1983): Estatística Descritiva. Mc-Graw-Hill. Lisboa
 43. OLIVEIRA, T.A. (2004): Estatística Aplicada, Universidade Aberta, Lisboa
 44. PEREIRA, A. (2004): SPSS - Guia prático de utilização, 5ª edição, Edições Sílabo, Lisboa
 45. PIRES A. (1995): Desenvolvimento Pessoal e Profissional – Um estudo dos Contextos e Processos de Formação das novas Competências Profissionais, Tese de Mestrado em Ciências de Educação/Educação e Desenvolvimento pela Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa
 46. RAO, C. R. (2002). Linear Statistical Inference and its Applications. Wiley.
 47. REIS, E. (2001). Estatística Multivariada Aplicada. 2ª Edição. Lisboa: Edições Sílabo.
 48. REIS, E.; Melo P.; Andrade R. e Calapez T. (1997). Estatística Aplicada. Vol. 1 e 2. Lisboa: Edições Sílabo.
 49. ROÇADAS, C.M.(2005). Relevância da formação profissional no desenvolvimento de competências: uma aplicação ao sector bancário. Tese de Mestrado, Universidade Aberta.
 50. SHESKIN, D. (2004). Hanbook of parametric and non-parametric procedures, 3rd Ed., Chapman & Hall.
 51. SIEGEL, S. E CASTELLAN, N. (1988): Nonparametric Statistics for the Behavioral Sciences, 2nd ed., McGraw-Hill International Editions - Statistics Series
 52. SCHOOTT, J. (2005). Matrix Analysis for Statistics. 2nd. ed. Hoboken, NJ: Wiley.
 53. SPSS for Windows, versão 13.0, Software de Estatística, SPSS Inc., E.U.A.
 54. TIGHT, M. (1996): Key Concepts in Adult Education and Training London, Routledge
 55. TITMUS, C. (1989): Lifelong Education for Adults: An International Handbook, Pergamon Press Plc, Oxford
 56. TORGO, LUIS (2009): A Linguagem R – Programação para a análise de dados. Escolar Editora, Lisboa
 57. TURKMAN, M. A. A. & Silva, G. L. (2000). Modelos Lineares Generalizados – da teoria à prática. Sociedade Portuguesa de Estatística, Peniche, Portugal.
 58. WATKINS, D.S.; (2002): Fundamentals of Matrix Computation (Pure and Applied Mathematics), Wiley.

APÊNDICES