

UNIVERSIDADE FEDERAL DE JUIZ DE FORA

DEPARTAMENTO DE ESTATÍSTICA

CURSO DE ESTATÍSTICA

Camila dos Reis Lima

**ANÁLISE DE CORRESPONDÊNCIA APLICADA: PERFIL DOS
PAÍSES AFRICANOS SEGUNDO OS OBJETIVOS DO MILÊNIO**

Juiz de Fora
2014

Camila dos Reis Lima

ANÁLISE DE CORRESPONDÊNCIA APLICADA: PERFIL DOS PAÍSES
AFRICANOS SEGUNDO OS OBJETIVOS DO MILÊNIO

Monografia apresentada ao curso de Estatística da Universidade Federal de Juiz de Fora, como requisito para a obtenção do grau de Bacharel em Estatística.

Orientador: Augusto Carvalho Souza

Camila dos Reis Lima

ANÁLISE DE CORRESPONDÊNCIA APLICADA: PERFIL DOS PAÍSES
AFRICANOS SEGUNDO OS OBJETIVOS DO MILÊNIO

Monografia apresentada ao curso de Estatística da
Universidade Federal de Juiz de Fora, como requisito para a obtenção do grau de Bacharel em
Estatística.

Aprovada em 03/02/2014

BANCA EXAMINADORA

Augusto Carvalho de Souza (Orientador)

Marcel de Toledo Vieira

Ronaldo Rocha Bastos

AGRADECIMENTOS

Gostaria de agradecer a Deus por ter reservado este caminho para mim, pois me sinto privilegiada por esta vida que levo. Aos meus pais, pela oportunidade de me graduar nesta instituição, pela paciência e apoio. Ao meu irmão Lucas, pela afinidade, pelos momentos passados durante as férias de julho ou de dezembro. À minha família, em especial minhas tias Valéria e Ivana, por todo o carinho que recebi nesses 23 anos.

Ao Raphael, pela amizade que depois se tornou mais que isso. Talvez o último ano fosse mais difícil sem esta pessoa ao meu lado. À Bárbara, por ter me esperado voltar para Juiz de Fora, e retomar a nossa amizade de onde deixamos! À Anna Cláudia, Carolina, Franciele, Isabela e Vítor, pelos momentos de descontração e estudos. O curso se tornou mais agradável com esses estatísticos, mesmo que no seu decorrer, nós não fizéssemos todas as matérias juntos. Aos outros amigos estatísticos que fiz nestes anos, pelas conversas, desabafos e ajudas. Ao pessoal que frequenta o D.A., do curso de Ciências Exatas, pelos momentos de descontração e união. Aos meus amigos de Coronel Fabriciano, em especial à Helena, pelos momentos via redes sociais e telefone, que nem Portugal distanciou. À Tatiana, que hoje se encontra no Mato Grosso, mas que por 3 anos se encontrava no quarto ao lado. Agradeço essas duas pessoas pela total afinidade, identificação e apoio.

Aos professores deste curso, em especial ao Augusto por ter me apoiado neste trabalho, ao Alfredo pela sua solidariedade aos alunos perdidos no curso, à Camila por ter me feito estudar como nunca tinha estudado antes na vida e ser um exemplo a ser seguido como docente, ao Clécio por ter me dado a (in) felicidade de poder conviver com ele, principalmente durante o ano como sua bolsista da Coordenação, ao Marcel e ao Ronaldo, por serem exemplos de profissionais pra mim, cada um na sua área.

Pai, dever cumprido.

RESUMO

Este trabalho utiliza dados relacionados aos Objetivos do Milênio, um programa desenvolvido pela ONU, que estão divididos em oito metas. Cada meta tem um conjunto de indicadores que abrangem áreas como saúde, educação, meio-ambiente e economia. A motivação de investigar apenas países africanos veio dos problemas históricos que o continente africano apresenta nos principais objetivos definidos pela ONU, como na saúde (são exemplos a desnutrição e a malária), no meio-ambiente (desertificação) e com os conflitos políticos.

Foram escolhidos os seis primeiros objetivos para análise e foram identificados dois problemas referentes à base de dados: (i) dados faltantes e (ii) variáveis com diferentes unidades de medida e tipo. Para solucionar estes dois problemas, adotou-se primeiramente uma forma de imputação múltipla, via modelo de regressão, para variáveis que apresentassem ausência de dados e que tinham como variáveis auxiliares as outras que pertenciam ao mesmo objetivo. Após a conclusão desta etapa, foi preciso encontrar uma forma de padronização dos dados. Com a padronização adotada cada país passa a apresentar um valor entre 0 e 1 para todos os indicadores, os quais passam a mostrar o caminho já percorrido como proporção de todo o caminho a percorrer no respectivo indicador.

Após estes ajustes, a técnica Análise de Correspondência Simples é utilizada para analisar a variabilidade total presente no banco de dados e representar graficamente a associação entre países e variáveis. O software utilizado para a imputação múltipla e para a Análise de Correspondência Simples foi o SAS.

Palavras chave: Análise de Correspondência, Objetivos do Milênio, Multidimensionalidade, Padronização, Imputação.

ABSTRACT

This work uses a database related to the Millennium Development Goals, a program developed by the UN, which is divided into eight targets data. Each goal has a set of indicators covering areas such as health, education, environment and economy. The motivation to investigate only African countries came from the historical problems that Africa presents in the main goals set by the UN, as in health (examples are malnutrition and malaria), environment (desertification) and with political conflicts.

The first six goals were chosen for analysis and two common problems were identified in the database: (i) missing data and (ii) variables with different measurement units and types. To solve these two problems, we adopted primarily a form of multiple imputation via regression model to a variable that had missing values, using as auxiliary variables its goal companions. After completion of this stage, it was necessary to find a way to standardize the data. The standardization adopted, each country presented a value between 0 and 1, which shows the steps taken as a proportion of the whole way to go for its indicator.

After these adjustments, the Simple Correspondence Analysis technique is used to analyze the total variability present in the database, and graphing the association between variables and countries. The software used for multiple imputation and the Simple Correspondence Analysis was SAS.

Key words: Correspondence Analysis, Millennium Development Goals, multidimensionality, Standardization, Imputation.

SUMÁRIO

INTRODUÇÃO.....	10
CAPÍTULO 1 – OS OBJETIVOS DO MILÊNIO:.....	12
1.1 OS INDICADORES.....	13
OS OBJETIVOS DO MILÊNIO E A ÁFRICA	14
CAPÍTULO 3 – AJUSTE DA BASE DE DADOS	17
IMPUTAÇÃO.....	17
TÉCNICA DE PADRONIZAÇÃO	20
CAPÍTULO 4 – ANÁLISE DE CORRESPONDÊNCIA.....	22
CAPÍTULO 5 – RESULTADOS E ANÁLISE.....	27
CONSIDERAÇÕES FINAIS.....	41
ANEXO	42
REFERÊNCIA BIBLIOGRÁFICA	56

LISTA DE FIGURAS:

Figura 1: As regiões do continente africano. (Fonte: Wikipedia).....	15
Figura 2: Gráfico da Análise de Correspondência para países, dimensões 1 e 2.....	29
Figura 3:Gráfico da Análise de Correspondência para países, dimensões 1 e 3.....	30
Figura 4:Gráfico da Análise de Correspondência para países, dimensões 2 e 3.....	31
Figura 5:Gráfico da Análise de Correspondência para as variáveis, dimensões 1 e 2	32
Figura 6:Gráfico da Análise de Correspondência para as variáveis, dimensões 1 e 3	33
Figura 7:Gráfico da Análise de Correspondência para as variáveis, dimensões 2 e 3	34
Figura 8: Mapa político do continente africano.	55

INTRODUÇÃO

Há 68 anos era fundada, nos EUA, a ONU (Organização das Nações Unidas), que tem entre vários objetivos, os de garantir os Direitos Humanos e promover o desenvolvimento socioeconômico das nações (<http://www.brasilecola.com/geografia/onu.htm>). No ano 2000, líderes de praticamente todos os países do mundo se comprometeram, através da ONU, em atingir metas definidas para melhorar a qualidade de vida de todos os habitantes do planeta, até o ano de 2015, através dos “Objetivos do Milênio” que contém indicadores determinados pela Agência Internacional e Grupos Especializados (*Inter-Agency and Expert Group- IAEG*), que é também responsável pela preparação e as análises dos dados.

A finalidade deste trabalho é poder discutir como os países africanos se relacionam com os indicadores alvos determinados pela ONU para que se tenha um planeta mais igualitário em gênero, saudável, com a garantia de estudo para todos entre outros objetivos; quais deles se assemelham entre si, ou seja, visualizar graficamente o comportamento dos países e variáveis que medem o progresso mundial. A escolha pelo continente africano é justificada por este sofrer com problemas históricos e conseqüentemente apresentar os piores valores para os indicadores sociais. Com base nessa informação, houve um interesse maior em estudar justamente estes objetivos do milênio e suas relações com os países africanos, para poder entender melhor a África e seus principais desafios, que envolvem a fome e doenças infecciosas, como a malária e a tuberculose.

Primeiramente foi preciso construir um banco de dados com todos os países do continente africano e com variáveis relacionadas com o assunto, disponíveis no site do programa “Objetivos do Milênio” da ONU (<http://mdgs.un.org/unsd/mdg/Data.aspx>), onde estão presentes todos os indicadores considerados para os objetivos, e os respectivos valores para cada país (ONU, 2013).

O banco de dados considerado para análise apresentava um desafio à parte: a presença de uma grande quantidade de dados faltantes, já que diversos países

não divulgaram valores para as principais variáveis de cada objetivo do milênio. Outra barreira encontrada, a qual dificultava a aplicação da Análise de Correspondência Simples, foi a diversidade métrica das variáveis, pois estas apresentavam porcentagens, taxas e contagens. Por isso, foram aplicados os seguintes procedimentos: modelos de regressão para imputação múltipla e a padronização de dados através da criação de indicadores adimensionais. O modelo de regressão foi utilizado para que fossem imputados os valores faltantes e a técnica de padronização empregada define um valor observado dentro do intervalo 0 e 1 que representa proporcionalmente a distância de um país até o valor máximo daquela variável.

Após estes ajustes, aplicou-se a técnica de Análise de Correspondência Simples (ACS) ao banco de dados. A ACS é uma ferramenta de análise exploratória aplicada originalmente a tabelas de contingência que visa reduzir a dimensionalidade da tabela e exibir em um gráfico com poucas dimensões a estrutura de relação entre as variáveis de coluna e linha, possibilitando uma interpretação mais clara e ampla dos dados.

Este trabalho está dividido em cinco capítulos, além das considerações finais. No segundo capítulo são descritas as metas que conjuntamente são denominados “Objetivos do Milênio”, os indicadores adotados para acompanhar o progresso destas metas e a definição de quais variáveis foram analisadas. No terceiro capítulo, são explicadas as soluções encontradas para os problemas referentes ao banco de dados, que são: imputação múltipla por modelo de regressão e a padronização das variáveis. A apresentação da Análise de Correspondência Simples está na terceira parte do trabalho. No quarto capítulo são apresentados os resultados e conclusões sobre os perfis dos países africanos em relação às variáveis analisadas como, por exemplo, como alguns países pertencentes à mesma região se associam e as relações entre variáveis que envolvem educação escolar e conhecimento da AIDS.

CAPÍTULO 1 – OS OBJETIVOS DO MILÊNIO:

Em setembro de 2000 foram definidos, junto aos líderes de 189 nações, os Objetivos do Milênio, os quais deverão ser atingidos por todos os países signatários até 2015. Para ajudar a controlar o andamento do compromisso assumido no ano de 2000 na Declaração do Milênio das Nações Unidas, a Agência Internacional e Grupos Especializados (Inter-Agency and Expert Group- IAEG) ficou responsável pela preparação dos dados e suas análises (inclusive o progresso dos objetivos) e pela definição e análise de metodologias. Esse grupo inclui vários Departamentos dentro da Secretaria das Nações Unidas, várias agências integrantes do sistema das Nações Unidas, agências governamentais, centros de estatísticas nacionais e outras organizações preocupadas com o desenvolvimento dos Objetivos do Milênio em níveis nacional e internacional.

Nessa agência foram discutidas as respectivas metas e selecionados os indicadores mais relevantes para serem utilizados para avaliar o progresso das metas estabelecidas e que são apresentados a cada ano, pelo Secretário-Geral, através de um relatório à Assembleia Geral das Nações Unidas.

Os números apresentados no banco de dados são compilados por agências especializadas dentro de sua área de especialização. Eles são extraídos de estatísticas nacionais fornecidas pelos governos para o sistema internacional de estatísticas – a Divisão de Estatística das Nações Unidas e os institutos de estatística dos diversos órgãos – e ajustados para fins de comparabilidade. A disponibilidade de dados necessários para o cálculo dos indicadores de cada país depende da capacidade dos serviços nacionais de estatística para produzir os dados necessários e/ou relatá-los em tempo hábil para as agências internacionais relevantes. Em alguns casos, os governos nacionais podem ter estatísticas mais recentes que ainda não foram notificados ao sistema estatístico internacional. Em outros casos, os países não produzem os dados necessários para a compilação dos indicadores. Quando isso ocorre, as agências fazem estimativas com base nos dados disponíveis sobre as variáveis relacionadas.

1.1 OS INDICADORES

Para que se tenha um mundo com menos pobreza, fome e doença, maiores perspectivas de sobrevivência para as mães e seus bebês, crianças mais educadas, igualdade de oportunidades para as mulheres e um ambiente mais saudável, os objetivos do milênio (*Millennium Development Goals Indicators-MDG*) foram divididos em 8 grupos e cada objetivo contém também os “alvos”, que são definidos abaixo. As variáveis que mensuram cada objetivo estão no quadro 1 do Anexo.

Os objetivos estabelecidos para o período entre 1990 e 2015 são (<http://mdgs.un.org/unsd/mdg/Host.aspx?Content=Indicators/About.htm>):

1. Erradicar a extrema fome e pobreza: Reduzir pela metade a porcentagem de pessoas cujo rendimento é inferior a 1 dólar por dia e a proporção de pessoas que sofrem de fome;
2. Atingir o ensino básico universal: Garantir que todas as crianças, meninos e meninas da mesma forma, serão capazes de concluir um curso completo de ensino primário;
3. Promover a igualdade de gênero e a autonomia das mulheres: Eliminar a disparidade de gênero no ensino primário e secundário, se possível até 2005, e em todos os níveis de ensino até 2015;
4. Reduzir a Mortalidade Infantil: Reduzir em dois terços, a taxa de mortalidade de menores de cinco anos;
5. Melhorar a Saúde Materna: Reduzir em três quartos a taxa de mortalidade materna;
6. Combater o HIV / AIDS, a malária e outras doenças: Parar e começar a reverter a propagação do HIV / AIDS, a incidência da malária e outras doenças graves;
7. Garantir a Sustentabilidade Ambiental: Integrar os princípios do desenvolvimento sustentável nas políticas e programas nacionais e reverter a perda de recursos ambientais; reduzir pela metade a proporção de pessoas sem acesso sustentável à água potável e ao saneamento básico e ter alcançado até

2020, uma melhoria significativa nas vidas de pelo menos 100 milhões de moradores de favelas;

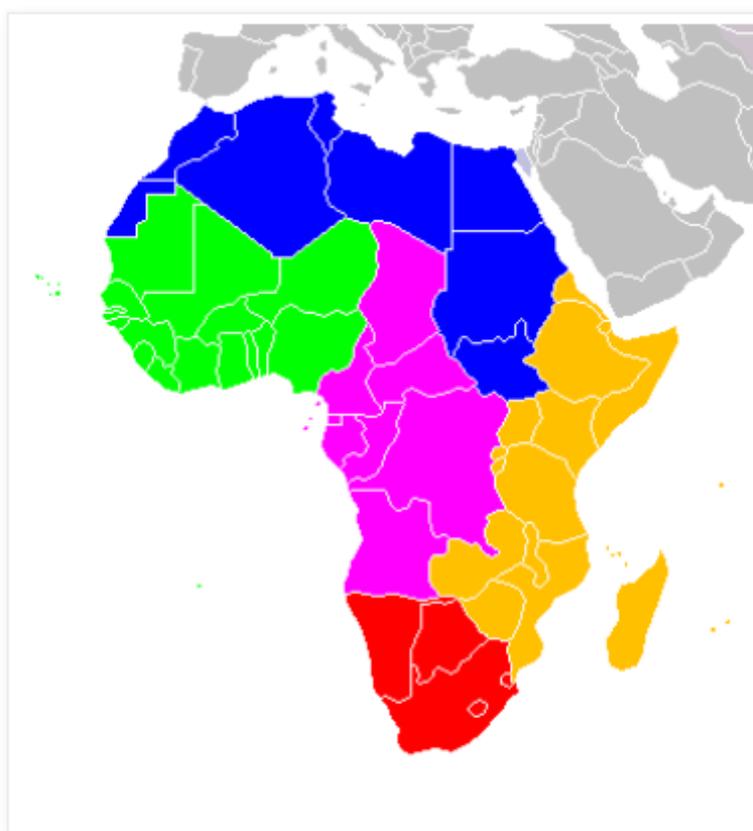
8. Estabelecer uma Parceria Mundial para o Desenvolvimento: Desenvolver um sistema aberto, previsível e não discriminatório baseado em regras e sistema financeiro; atender às necessidades especiais dos países menos desenvolvidos, em desenvolvimento sem litoral e dos pequenos Estados insulares em desenvolvimento; tratar globalmente o problema da dívida dos países em desenvolvimento através de medidas nacionais e internacionais, a fim de tornar a sua dívida sustentável em longo prazo; em cooperação com os países em desenvolvimento, formular e executar estratégias para o trabalho decente e produtivo para os jovens e em cooperação com o setor privado, tornar acessíveis os benefícios das novas tecnologias.

OS OBJETIVOS DO MILÊNIO E A ÁFRICA

A África é o terceiro continente mais extenso (atrás da Ásia e da América) com cerca de 30 milhões de quilômetros quadrados. Apresenta grande diversidade étnica, cultural, social e política. A África costuma ser regionalizada de duas formas; a primeira forma, que valoriza a localização dos países e os divide em cinco grupos, que são: a Norte da África, a África Ocidental, a África Oriental, a África Central e a África Meridional ou Austral (Figura 1). A segunda regionalização desse continente, que vem sendo muito utilizada, usa critérios étnicos e culturais (religião e etnias predominantes em cada região), é dividida em dois grandes grupos, a África Branca ou setentrional formado pelos oito países da África do norte, mais a Mauritânia e o Saara Ocidental, e a África Negra ou subsaariana formada pelos outros 44 países do continente.

Os problemas africanos são históricos. A África recém-autônoma enfrenta enormes desafios como a extrema pobreza, a doença, a desertificação, a desnutrição, os atritos políticos e encargos devastadores da dívida externa (<http://www.un.org/en/globalissues/africa/>). Um compromisso internacional com a África tem sido caracterizado pelo início de um esforço conjunto para se mover em direção a uma abordagem coerente para lidar com o continente como, por exemplo, os recentes esforços vindos do Conselho de Segurança, como as missões do Conselho para áreas de conflito na África e maior

envolvimento na manutenção da paz na África. Além dos problemas citados acima, a África sofre particularmente a partir da sua exclusão no processo de globalização. A sua participação no comércio, investimento e avanços na tecnologia diminuiu ainda mais ao longo da última década, contudo as lideranças africanas tem galvanizado apoio local e internacional para uma série de iniciativas e estratégias para avançar em áreas como reforço da democracia, construção da capacidade de manutenção da paz, trabalho para o desenvolvimento sustentável e combate ao HIV/AIDS.



Subdivisões da África para fins estatísticos usada pela ONU.

- Norte de África (físico-geograficamente, a Península de Sinai, no Egito, pertence ao Médio Oriente, região da Ásia).
- África Ocidental
- África Central
- África Oriental
- África Meridional ou África Austral

Figura 1: As regiões do continente africano. (Fonte: Wikipedia)

A busca pelos dados oficiais da África gerou uma tabela com 55 países e 44 variáveis. As variáveis escolhidas para o estudo são apenas aqueles referentes

do primeiro ao sexto objetivo do programa da ONU, uma vez que o sétimo e o oitavo objetivo referem-se a dois temas que não pertencem ao foco deste trabalho. Durante o tempo de avaliação dos objetivos do milênio, mais precisamente em 2011, o Sudão foi dividido em dois países: o Sudão, e o Sudão do Sul. Com isso, de acordo com a Divisão de Demografia das Nações Unidas, algumas séries de dados foram disponibilizadas para o Sudão do Sul como um país separado. Alguns países foram eliminados do estudo por não apresentar muitas respostas às variáveis; são eles a Líbia, Saara Ocidental, Seychelles, Sudão e Sudão do Sul. Foi definida também a retirada das seguintes variáveis referentes ao primeiro e sexto objetivos: (i) porcentagem de crescimento do PIB por pessoa empregada, (ii) percentual de mulheres de 15-24 anos de idade que usaram de preservativo na última relação sexual de alto risco; e (iii) percentual de homens de 15-24 anos de idade que usaram de preservativo na última relação sexual de alto risco. Estas variáveis foram retiradas, pois mais da metade dos países não apresentaram valores referentes a elas.

CAPÍTULO 3 – AJUSTE DA BASE DE DADOS

IMPUTAÇÃO

A ausência de informações em um conjunto de dados é um problema bem conhecido que tem de ser enfrentado por praticamente todos que realizam análise de dados. No caso em que a unidade de análise é uma pessoa, por exemplo, o respondente pode não entender a pergunta, não saber a resposta, simplesmente se esquecer de respondê-la ou, então, se recusar a fornecer a informação. Esta ausência de dados implica na redução do tamanho efetivo da amostra e, conseqüentemente, no aumento no erro padrão das estimativas dos parâmetros. Algumas instituições de pesquisas, ao lidar com este problema, podem antecipar a ocorrência de não-resposta por *oversampling* e também corrigir a perda de precisão pode ser quantificada quando erros padrão são estimados.. (Waal, Ton de.2011).

Os dados faltantes podem ser subdivididos em várias formas, de acordo com o mecanismo de não resposta subjacente. Se o mecanismo de não-resposta não depende de dados não observados (ou seja, dependem apenas dos dados observados), a imputação pode levar a estimativas imparciais, sem fazer mais suposições. Se o mecanismo de não-resposta depende de dados não observados (inverificáveis), suposições adicionais serão necessárias para reduzir o viés por meio de imputação.

Uma classificação de mecanismos de geradores de não-resposta bem conhecida e bastante utilizada é a seguinte: (i) faltante completamente ao acaso – *missing completely at random* (MCAR), (ii) faltante ao acaso- *missing at random* (MAR) e (iii) não faltante ao acaso- *not missing at random* (NMAR) (ver Rubin (1987), Schafer (1997) e Little e Rubin (2002)).

Quando um dado faltante é do tipo MCAR, a probabilidade do valor ser faltante não depende do(s) valor(es) da(s) variável(eis) alvo a ser imputada ou dos valores das variáveis auxiliares, ou seja, a causa que levou aos dados faltantes é um evento aleatório. Esta situação pode ocorrer, por exemplo,

quando um entrevistado se esquece de responder a uma pergunta ou quando uma parte aleatória dos dados é perdida enquanto é processado.

Quando dados faltantes são MAR, a probabilidade de o valor ser perdido depende dos valores de variáveis auxiliares, mas não do valor da(s) variável(eis) alvo a ser(em) imputada (s). Isso significa que os dados faltantes são causados por alguma variável observada, disponível para análise e correlacionada com a variável que possui dados faltantes. Neste trabalho, vamos considerar que os dados faltantes são deste tipo.

Quando dados faltantes são NMAR, a probabilidade de o valor ser perdido vai depender do(s) valor(es) da(s) variável(eis) alvo a ser(em) imputado(s), e, possivelmente, dos valores das variáveis auxiliares. A situação descrita ocorre quando a causa dos dados faltantes é o seu próprio valor.

A imputação é parte do processo que abrange toda a edição, imputação e outras ações realizadas, a fim de transformar os dados brutos para um conjunto de dados estatísticos definidos e prontos para análise e tabulação. O ato de imputar envolve estimar e preencher um valor viável para um valor em falta no conjunto de dados. Lembrando que a modificação de um valor do banco de dados não pode ser considerada uma imputação, mas sim uma correção. Técnicas de edição e imputação de dados podem ser divididas em duas classes principais, dependendo do tipo dos dados a serem editados ou imputados: (i) técnicas para dados numéricos e (ii) técnicas para dados categóricos. Técnicas de imputação em dados numéricos são mais comuns, pois variáveis categóricas podem incluir a categoria “não sabe/não respondeu” em pesquisas.

Há diversas formas para imputação múltipla dos dados e cada método depende da natureza dos dados. No caso de dados numéricos, algumas das técnicas mais conhecidas são: os métodos de Máxima Verossimilhança, o de Monte Carlo em cadeia de Markov (MCMC), o algoritmo EM e os modelos de regressão. Os métodos de Máxima Verossimilhança têm como objetivo principal o de estimar os parâmetros de interesse e assim maximizar a probabilidade de observar o que, de fato, foi observado. O MCMC usa um método de Monte Carlo na cadeia de Markov para imputar os valores para um conjunto de dados, assumindo distribuição normal multivariada para os dados. O algoritmo EM

(*expectation-maximization*) é usado para calcular a estimativa de máxima verossimilhança do banco de dados com valores faltantes, assumindo uma distribuição normal multivariada dos dados. (Yuan, C.Y.)

O meio utilizado neste trabalho para a imputação de dados faltantes é via modelo de regressão, que é uma das formas mais simples e de fácil entendimento dentre as formas de imputação múltipla dos dados. Em vez de preencher com um único resultado cada observação faltante, o procedimento de imputação múltipla proposta por Donald Rubin há 30 anos substitui cada observação em falta por um valor de um conjunto de valores plausíveis, que representa a incerteza sobre a resposta verdadeira. O novo conjunto de dados, agora com valores imputados, pode ser então, analisado de acordo com o objetivo do pesquisador. (Waal, Ton de.2011)

No método de regressão, um modelo de regressão é ajustado para cada variável que contém valores faltantes, isto é, para uma variável Y nesta condição, um modelo

$$Y_j = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K$$

é ajustado usando os valores observados para a variável Y_j e X_1, X_2, \dots, X_K , como suas variáveis explicativas. Se um modelo que prevê a variável de interesse pode ser construído, pode-se usar a imputação baseada em um modelo para melhorar a qualidade do conjunto de dados. Os valores preditos de acordo com o modelo selecionado são as imputações, ou estimativas dos valores em falta. Os modelos de regressão, principalmente para variáveis numéricas, são os modelos mais utilizados.

No presente trabalho, os conjuntos de variáveis foram separados em grupos de acordo com cada objetivo do milênio de forma que cada variável a ser imputada, tinha como variáveis explicativas suas variáveis do próprio objetivo. Como exemplo, podemos citar o mecanismo para imputar valores dentro do sexto objetivo – Combater o HIV / AIDS, a malária e outras doenças (ver Quadro 1 em anexo). Dentro deste objetivo, cinco variáveis compõem o subgrupo referente à malária (mortalidades, notificação e estruturas de segurança para a criança) e destas cinco, duas variáveis contêm valores

faltantes. Usando as outras três variáveis como explicativas, foi ajustado, primeiramente, um modelo de regressão para imputar os dados da variável com menor quantidade de observações faltantes. Em seguida, outro modelo de regressão foi ajustado para imputar as observações faltantes da quinta variável, desta vez adicionando como variável auxiliar o indicador imputado no passo anterior. No SAS isso é possível utilizando a rotina “*proc mi*”.

TÉCNICA DE PADRONIZAÇÃO

Padronizar significa remover a escala da variável original e transformá-la em um índice, e com isso permitir a comparação entre as variáveis padronizadas. Uma das ferramentas mais conhecidas de padronização de dados é

$$Z = \frac{x - \mu}{\sigma}$$

Esta técnica utiliza a média e o desvio padrão da própria variável para obter um valor padronizado, z , representado em uma nova escala, a qual tem a média como zero e o desvio padrão como unidade de medida. Assim o valor padronizado informa quantos desvios padrão a observação está distante, acima ou abaixo, da média.

Uma vez que os valores padronizados poderiam ser negativos, não seria possível usufruir desta técnica no presente trabalho uma vez que a Análise de Correspondência, técnica utilizada para analisar os dados, só pode ser aplicada a matrizes de dados positivos. A solução encontrada, então, para se padronizar os dados foi utilizar a seguinte fórmula:

$$\text{Valor Padronizado} = \frac{\text{valor} - \text{valor mínimo}}{\text{valor máximo} - \text{valor mínimo}}$$

No qual “valor”, “valor mínimo” e “valor máximo” são valores pertencentes a uma mesma variável do próprio banco de dados. Pode-se garantir que mesmo utilizando este índice, os países preservam as diferenças entre si. Esta padronização transmite uma ideia também muito simples, que é a de distância do valor observado em relação ao menor e ao maior valor definidos para ele: quanto mais o valor observado se aproxima do mínimo, menor é o numerador da equação e o país terá um valor mais próximo de 0. Por outro lado, quanto

mais o valor observado para o país se aproximar do valor máximo, mais próximo estará o numerador do denominador da equação e, portanto, o país terá um valor mais próximo de 1 ([http://pascal.iseg.utl.pt/~cesa/index.php/dicionario-da-cooperacao/Glossary-1/%C3%8D/%C3%8Dndice-de-Desenvolvimento-Humano-\(IDH\)-261/](http://pascal.iseg.utl.pt/~cesa/index.php/dicionario-da-cooperacao/Glossary-1/%C3%8D/%C3%8Dndice-de-Desenvolvimento-Humano-(IDH)-261/)).

Desta forma, a escolha dos valores de máximo e mínimo, ou seja, a avaliação do que seriam os limites pior e/ou melhor do indicador em vista, é de fundamental importância e impacta diretamente no valor a ser obtido para o índice. Impacta também na dispersão dos valores do índice entre os diversos países.

Para cada variável, estes valores padronizados podem ser compreendidos de formas diferentes. Para variáveis que medem, por exemplo, taxas de mortalidade, incidências de doenças, subnutrição, os valores padronizados próximos de 0 indicam uma situação relativamente favorável, pois isso significaria que o indicador do país está relativamente melhor do que aqueles que apresentam valores próximos a 1. Já para variáveis que medem taxa de alfabetização e tratamento de doenças, por exemplo, são os valores próximos de 1 que indicam uma situação melhor em relação aos demais países.

CAPÍTULO 4 – ANÁLISE DE CORRESPONDÊNCIA

A Análise de Correspondência (AC) é uma técnica exploratória multivariada para análise gráfica e numérica para todos os tipos de dados não negativos e mais utilizados em tabelas de frequências ou de contagem (de Leeuw, J. and Mair, P.,2009). Embora a AC seja originalmente orientada para dados categóricos, ela pode ser usada para analisar quase todos os tipos de dados tabulados após uma transformação ou recodificação (*recoding*) adequada destes. Dentre seus objetivos estão a síntese do conjunto de dados através de uma representação gráfica da relação entre linhas e colunas.

Este método é baseado em decomposição em valores singulares (*Singular Value Decomposition*, em inglês – SVD), a qual é também base de outros métodos multivariados – como a Análise de Componentes Principais – e utilizado para redução de dimensionalidade de matrizes. A AC tem muitas variações que são determinadas pelas características do problema sendo analisado: quantidade de variáveis, forma de tabulação e tipo de variável. Neste trabalho, a técnica empregada foi a Análise de Correspondência Simples, que é tradicionalmente usada para representar graficamente as associações presentes em um conjunto de dados dispostos em uma tabela de contingência com duas variáveis qualitativas. Diferentemente de abordagem usual, esta técnica foi aplicada a uma tabela adequadamente preparada que é constituída pelas unidades de análise nas linhas (países) e por diversos indicadores quantitativos padronizados nas colunas.

- Descrição matemática da técnica¹:

Seja uma tabela de contingência com I linhas, J colunas e seus elementos representados por n_{ij} (Tabela 1).

Uma **matriz de correspondência P** é representada por seus elementos $p_{ij}=n_{ij}/n_{..}$, onde:

- $n_{..}$ é a soma de todos os elementos n_{ij} da matriz N ;

¹ Ver maiores detalhes em Greenacre & Blasius (2006) e Greenacre (2007).

- A soma dos elementos $n_{i.}$ representa a soma marginal de cada linha e $n_{.j}$, a soma marginal de cada coluna;

Tabela 1: Exemplo de tabela de contingência

	X				
Y	1	2	...	J	Total marginal
1	n_{11}	n_{12}	...	n_{1j}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2j}	$n_{2.}$
:	:	:
I	n_{i1}	n_{i2}	...	n_{iJ}	$n_{i.}$
Total marginal	$n_{.1}$	$n_{.2}$...	$n_{.J}$	$n_{..}$

- As frequências marginais relativas $p_{.j} = \frac{n_{.j}}{N}$ e $p_{i.} = \frac{n_{i.}}{N}$, denominados de r_i e c_j , respectivamente, são chamadas de massas de coluna e de linha; ou seja, são os pesos que cada coluna j e cada linha i e tem no total N .
 - Sejam então as matrizes de massas das linhas $\mathbf{D}_r = \text{diag}(r_1, r_2, \dots, r_I)$, e $\mathbf{D}_c = \text{diag}(c_1, c_2, \dots, c_J)$ a matriz de massas das colunas. São definidas então as matrizes inversas de raiz quadradas $\mathbf{D}_r^{-\frac{1}{2}}$ e $\mathbf{D}_c^{-\frac{1}{2}}$ como: $\mathbf{D}_r^{-\frac{1}{2}} = \text{diag}\left(\frac{1}{\sqrt{r_1}}, \dots, \frac{1}{\sqrt{r_I}}\right)$ e $\mathbf{D}_c^{-\frac{1}{2}} = \text{diag}\left(\frac{1}{\sqrt{c_1}}, \dots, \frac{1}{\sqrt{c_J}}\right)$;
- Os elementos dos perfis de linha e coluna são, respectivamente, $r_j^i = \frac{n_{ij}}{n_{i.}}$ e $c_i^j = \frac{n_{ij}}{n_{.j}}$. Isso significa que os perfis de linha são caracterizados pelos valores relativos às colunas e os perfis de colunas pelos valores relativos às linhas;
 - Os perfis de linha podem ser obtidos por $\mathbf{R} = \mathbf{D}_r^{-1}\mathbf{P}$;
 - Os perfis de colunas podem ser obtidos por $\mathbf{C} = \mathbf{D}_c^{-1}\mathbf{P}'$.

- Há também o conceito de perfil médio, que é uma média ponderada dos próprios perfis de linha ou de coluna, criando assim um centroide para os pontos de linha e um para os pontos da coluna para explicar a distância do ponto até sua média. Este centroide permite fazer uma comparação dos elementos dos perfis de linha ou coluna, com seus elementos correspondentes em relação ao perfil médio.
- Sob a hipótese nula de independência, os valores esperados das frequências relativas p_{ij} são os produtos $r_i c_j$ das massas.

Nos gráficos gerados pelo método, os pontos de linha são os perfis de linha e os pontos de colunas são os perfis de coluna. A medição das distâncias entre os perfis é feita utilizando a distância qui-quadrado que se assemelha à formulação da distância Euclidiana entre pontos em um espaço físico, exceto que cada coordenada é dividida pelo elemento correspondente do perfil médio:

- $\sum_{j=1}^J \frac{(\frac{p_{ij}}{r_i} - \frac{p_{i'j}}{r_{i'}})^2}{c_j}$ - Distância qui-quadrado entre as linhas i e i' ;
- $\sum_{i=1}^I \frac{(\frac{p_{ij}}{c_j} - \frac{p_{ij'}}{c_{j'}})^2}{r_i}$ - Distância qui-quadrado entre as colunas j e j' .

Outro objetivo da Análise de Correspondência é identificar, através da SVD, o subespaço com menor número de dimensões que melhor representa a dispersão dos dados, e para que se obtenham as coordenadas destes dados – os perfis de linha e coluna-, é necessário fazer a seguinte operação matricial:

$$\mathbf{S} = \mathbf{D}_r^{-\frac{1}{2}} (\mathbf{P} - \mathbf{r}\mathbf{c}^T) \mathbf{D}_c^{-1/2},$$

Onde posto da matriz \mathbf{S} é igual ao $\min(I-1, J-1)$. Ou, utilizando a forma de decomposição de valores singulares (SVD), que é uma forma bastante comum de obter resultados da AC:

$$\mathbf{S} = \mathbf{U} \mathbf{D}_\alpha \mathbf{V}^T$$

O método propõe o seguinte: seja $\mathbf{S}_{m \times k}$ uma matriz de números reais. Então, há matrizes ortogonais $\mathbf{U}_{m \times m}$ e $\mathbf{V}_{k \times k}$ no qual $\mathbf{S} = \mathbf{U} \mathbf{D}_\alpha \mathbf{V}^T$ onde a matriz $\mathbf{D}_{\alpha_{m \times k}}$

é uma matriz diagonal e tem nos seus elementos (i, i) valores $\alpha_i \geq 0$ para $i = 1, 2, \dots$ até $rank(m, k)$. Estas constantes positivas α_i são os **valores singulares** de \mathbf{P} . As colunas de \mathbf{U} e \mathbf{V} definem os eixos principais das nuvens de linha e coluna, respectivamente. Estas matrizes são vetores singulares da esquerda e da direita, respectivamente.

As coordenadas principais são as coordenadas dos perfis em relação ao eixo principal; as coordenadas dos vértices em relação a um eixo principal são chamadas coordenadas padronizadas. Matricialmente, temos que:

- Coordenadas padronizadas de linhas: $\Phi = \mathbf{D}_r^{-1/2} \mathbf{U}$
- Coordenadas padronizadas de colunas: $\Gamma = \mathbf{D}_c^{-1/2} \mathbf{V}$
- As coordenadas principais das linhas: $\mathbf{F} = \mathbf{D}_r^{-1/2} \mathbf{U} \mathbf{D}_\alpha = \Phi \mathbf{D}_\alpha$
- Coordenadas principais das colunas: $\mathbf{G} = \mathbf{D}_c^{-1/2} \mathbf{V} \mathbf{D}_\alpha = \Gamma \mathbf{D}_\alpha$
- Principais inércias: $\lambda_k = \alpha_k^2, k = 1, \dots, K$ onde $K = \min(I - 1, J - 1)$.

Para a elaboração do gráfico, o cálculo das distâncias entre os perfis é feito da seguinte forma: para as linhas, as coordenadas de linha são perfis de linha $\mathbf{D}_r^{-1} \mathbf{P}$, que serão redimensionadas por $\mathbf{D}_c^{-1/2}$ (redimensionadas de modo que as distâncias entre os perfis serão transformados a partir da distância qui-quadrado para uma distância euclidiana) e então é feita a rotação ortogonal com $\mathbf{D}_c^{-1/2} \mathbf{U}$ orientada para o eixo principal. Para as colunas, o cálculo é feito da mesma forma utilizando suas matrizes correspondentes.

Quando os pontos de linha/coluna são inseridos no gráfico através do cálculo de coordenadas principais, dizemos que o gráfico gerado é simétrico. Quando os pontos são inseridos de formas distintas (as linhas são representadas utilizando as coordenadas principais e as colunas pelas coordenadas padronizadas, e vice-versa), o gráfico gerado é dito assimétrico. No primeiro caso as distâncias entre os pontos de linha são interpretáveis, assim como as distâncias entre os pontos da coluna. No segundo caso, além das interpretações anteriores, é possível a interpretação das distâncias entre pontos de coluna e os pontos de linha.

A inércia é o valor que quantifica a variação total que há no banco de dados, ou seja, mede o quão distante os perfis de linha e coluna estão dos seus respectivos centroides. Neste contexto, então, podemos dizer que o centroide representa a hipótese nula de que os perfis são independentes. Quando o valor da inércia é baixo, os perfis de linha não são muito dispersos e se posicionam perto do perfil médio e conclui-se que há pouca associação (ou correlação) entre as linhas e colunas. Se o valor da inércia for alto, há uma associação forte entre linhas e colunas.

A soma dos quadrados dos elementos da matriz \mathbf{S} é utilizada para calcular a inércia total do banco de dados e pode ser relacionado com a estatística qui-quadrado, em forma matricial ou com a soma dos quadrados dos valores singulares:

$$\text{Inércia} = \text{traço}(\mathbf{S}\mathbf{S}^T) = \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - r_{ij})^2}{r_i c_j} \text{ ou}$$

$$\text{traço}(\mathbf{S}\mathbf{S}^T) = \text{traço}[\mathbf{D}_r^{-1}(\mathbf{P} - \mathbf{r}\mathbf{c}^T)\mathbf{D}_c^{-1}(\mathbf{P} - \mathbf{r}\mathbf{c}^T)^T]$$

A contribuição dos pontos de linha e coluna para a inércia da k-ésima dimensão pode ser representada na seguinte forma, como componentes da inércia:

- Para linha i: $\frac{r_i f_{ik}^2}{\lambda_k} = r_i \phi_{ik}^2$
- Para coluna j: $\frac{c_j g_{jk}^2}{\lambda_k} = c_j \gamma_{jk}^2$

Onde f_{ik}^2 e g_{jk}^2 são os pontos das coordenadas principais de linhas e colunas respectivamente sobre o eixo, ϕ_{ik}^2 é o valor da coordenada padronizada de linha e γ_{jk}^2 o valor da coordenada padronizada de coluna.

Ao utilizar os gráficos do SAS, verificou-se que os mesmos não apresentam uma visualização verossímil dos dados, pois se dois pontos estão sobrepostos, o SAS desloca um destes pontos para o espaço vazio mais próximo. Decidiu-se, então, por fazer as análises e conclusões pelo R (rotina *plot*), mas utilizando as coordenadas disponibilizadas pelos resultados do SAS.

CAPÍTULO 5 – RESULTADOS E ANÁLISE

Em uma tabela com exatamente 50 linhas e 41 colunas, as análises poderiam tomar os mais diversos focos históricos ou mesmo sociais. A partir de agora, serão mostrados os resultados mais visíveis nos gráficos que fazem parte da interpretação da Análise de Correspondência Simples – como a interpretação da inércia – e outras conclusões mais detalhadas e interessantes, primeiro para os indicadores das metas do milênio, em seguida para países e, por fim, para ambas as dimensões.

Neste trabalho optou-se pela análise da ACS a partir dos três primeiros eixos principais representados em 3 gráficos bidimensionais. No SAS, o comando utilizado foi o “*proc corresp*”. Os resultados são apresentados no anexo. Sua primeira tabela apresenta, em seu primeiro quadro, os valores singulares (raiz quadrada dos autovalores), a inércia total, as principais inércias de cada dimensão (autovalores), a decomposição da estatística qui-quadrado, a porcentagem da inércia de cada dimensão e esta porcentagem acumulada².

A escolha do gráfico com os dois primeiros eixos principais acumula 43,38% do total da estatística qui-quadrado e da inércia, ou seja, este gráfico explica 43,38% da variação dos dados. Para uma representação perfeita dos dados, seria necessário um gráfico com 40 dimensões, pois $40 = rank(50 - 1; 41 - 1)$.

As informações sobre os perfis de linha/coluna também estão anexadas no final do trabalho as quais são suas coordenadas nas 3 dimensões e um resumo de suas estatísticas: qualidade (*Quality* - a soma do quadrado dos cossenos em relação às 3 primeiras dimensões) da representação de cada ponto, massa (*Mass* - somas marginais da matriz de frequência relativa) e Inércia (*Inertia* - contribuição relativa do ponto para a inércia total), as contribuições parciais para a inércia de cada linha/coluna em cada dimensão e os quadrados dos cossenos dos ângulos entre cada eixo e um vetor a partir da origem até o ponto.

²O SAS oferece dois conjuntos de resultados distintos, porém complementares. Um resultado (OUTF) oferece uma tabela de frequências e porcentagens de cada uma das contribuições para a estatística qui-quadrado e o outro resultado (OUTC), que é analisado aqui, está em anexo.

Há também a informação sobre os índices das coordenadas dos pontos que mais contribuíram para a inércia. As tabelas 3 e 6 mostram a contribuição parcial de cada ponto para a inércia de cada dimensão, separadamente entre linhas e depois, entre colunas. O software SAS primeiramente analisa em cada linha/coluna qual ponto mais contribuiu para uma das três dimensões e o índice que tem mais contribuição estará na coluna intitulada *Best*. As colunas Dim1, Dim2, Dim3, que representam a dimensão 1, 2 e 3, respectivamente, também receberão este índice se o valor desta contribuição for significativo para a inércia total daquela dimensão ou receberão 0, caso contrário. Por exemplo, a linha 47 (Tunísia) tem como maior contribuição para inércia o valor 0,0724 para a dimensão 1 e este valor contribui significativamente para esta dimensão. Então a coluna *Best* recebe o índice 1. Porém esta mesma linha contribui significativamente para a dimensão 2 mas não para a dimensão 3. Então, na tabela, Dim2 recebe também o índice 1 e Dim3 receberá índice 0. Para se determinar as dimensões que mais contribuem para a inércia dos pontos (ver Tabela 4 e Tabela 7) o software SAS soma, para cada dimensão, a contribuição parcial de cada inércia dos pontos, em ordem decrescente, até que se atinja o valor 0,8. Os valores ditos significantes são aqueles que, em ordem decrescente, somem 0,8, ou seja, esses são os que mais contribuem com cada dimensão. Esta ferramenta disponibilizada pelo SAS contribui para a análise gráfica, uma vez que facilita perceber padrões em uma nuvem de pontos de forma precisa, já que se leva em conta suas representatividades em cada dimensão.

Análise da Nuvem de Indicadores

Como dito anteriormente, optou-se por utilizar os três primeiros eixos principais na análise. Com esse recorte, alguns agrupamentos de variáveis foram perceptíveis. A interpretação das distâncias entre países e indicadores seria possível se o gráfico gerado fosse do tipo assimétrico. Porém, como é do tipo simétrico, a distância entre linhas e colunas não é interpretável.

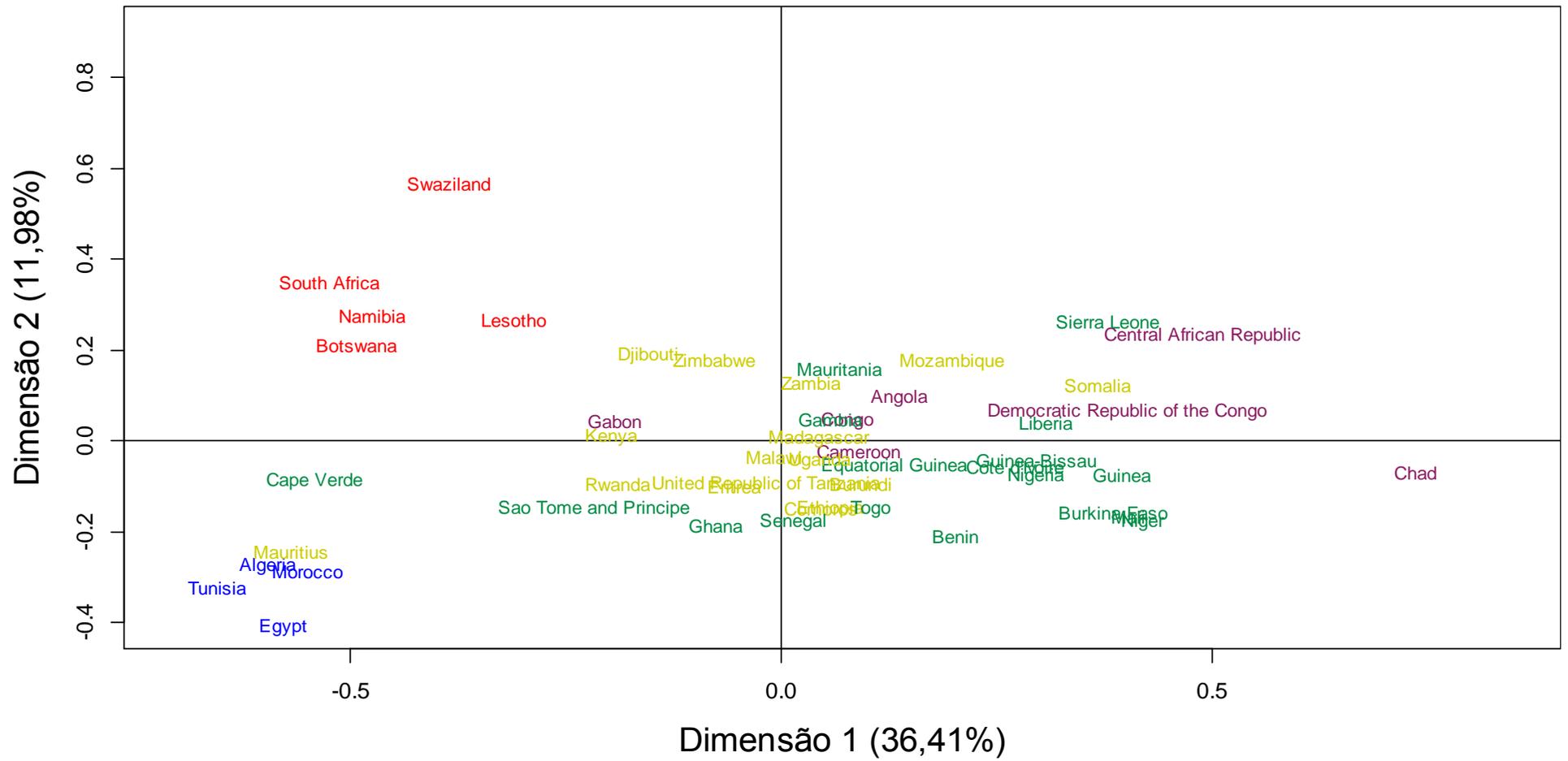


Figura 2: Gráfico da Análise de Correspondência para países, dimensões 1 e 2

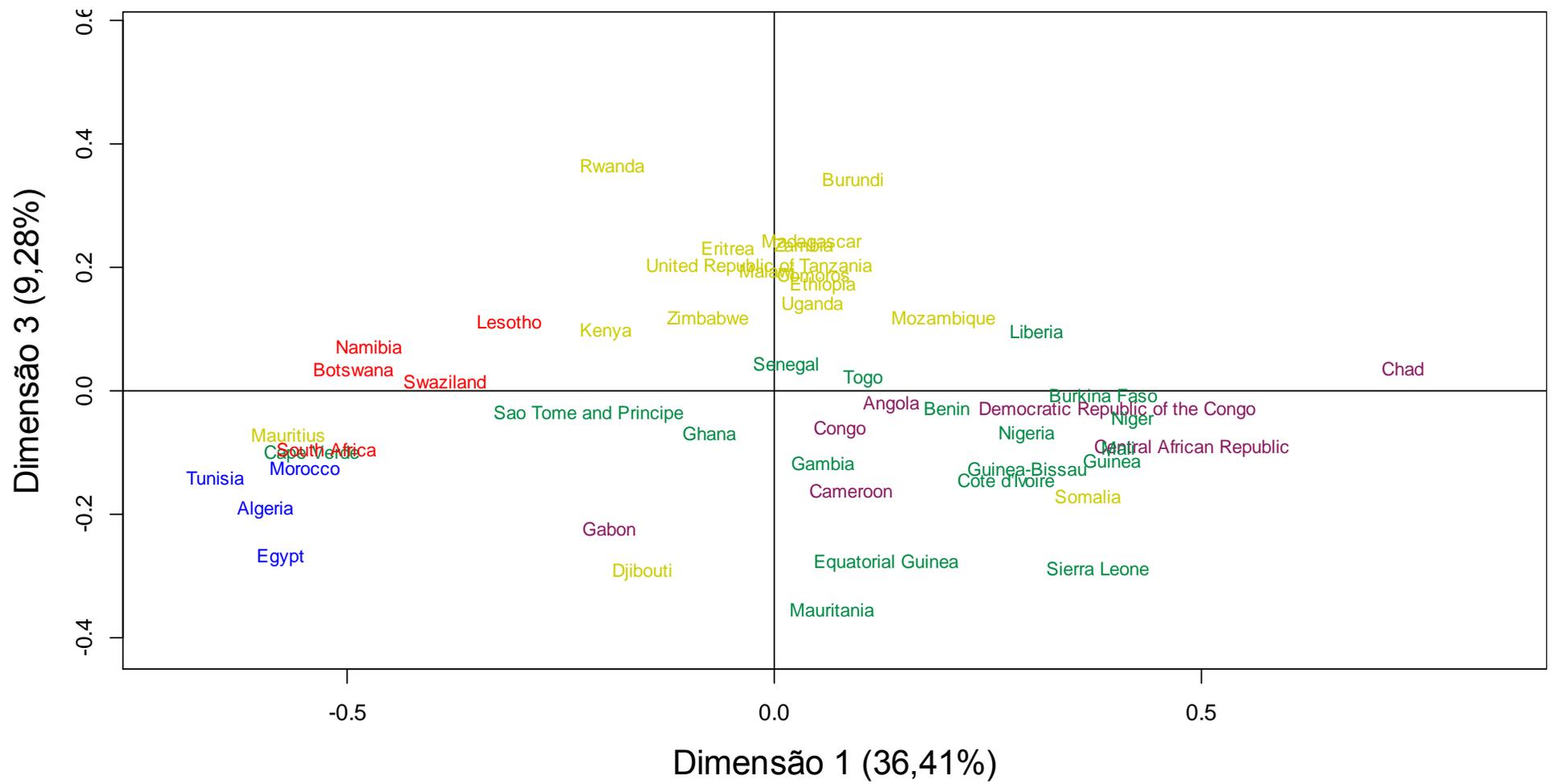


Figura 3: Gráfico da Análise de Correspondência para países, dimensões 1 e 3

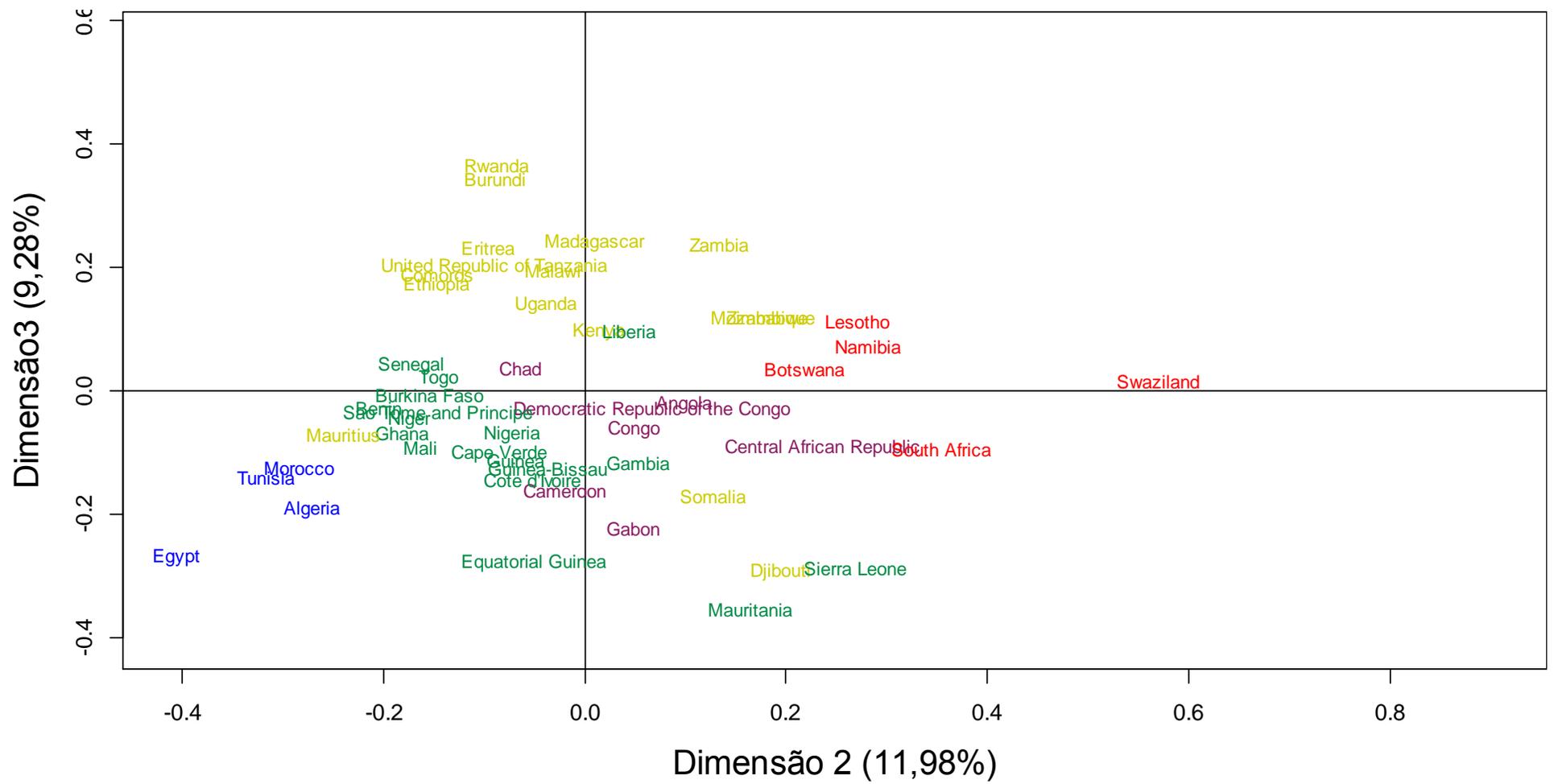


Figura 4: Gráfico da Análise de Correspondência para países, dimensões 2 e 3

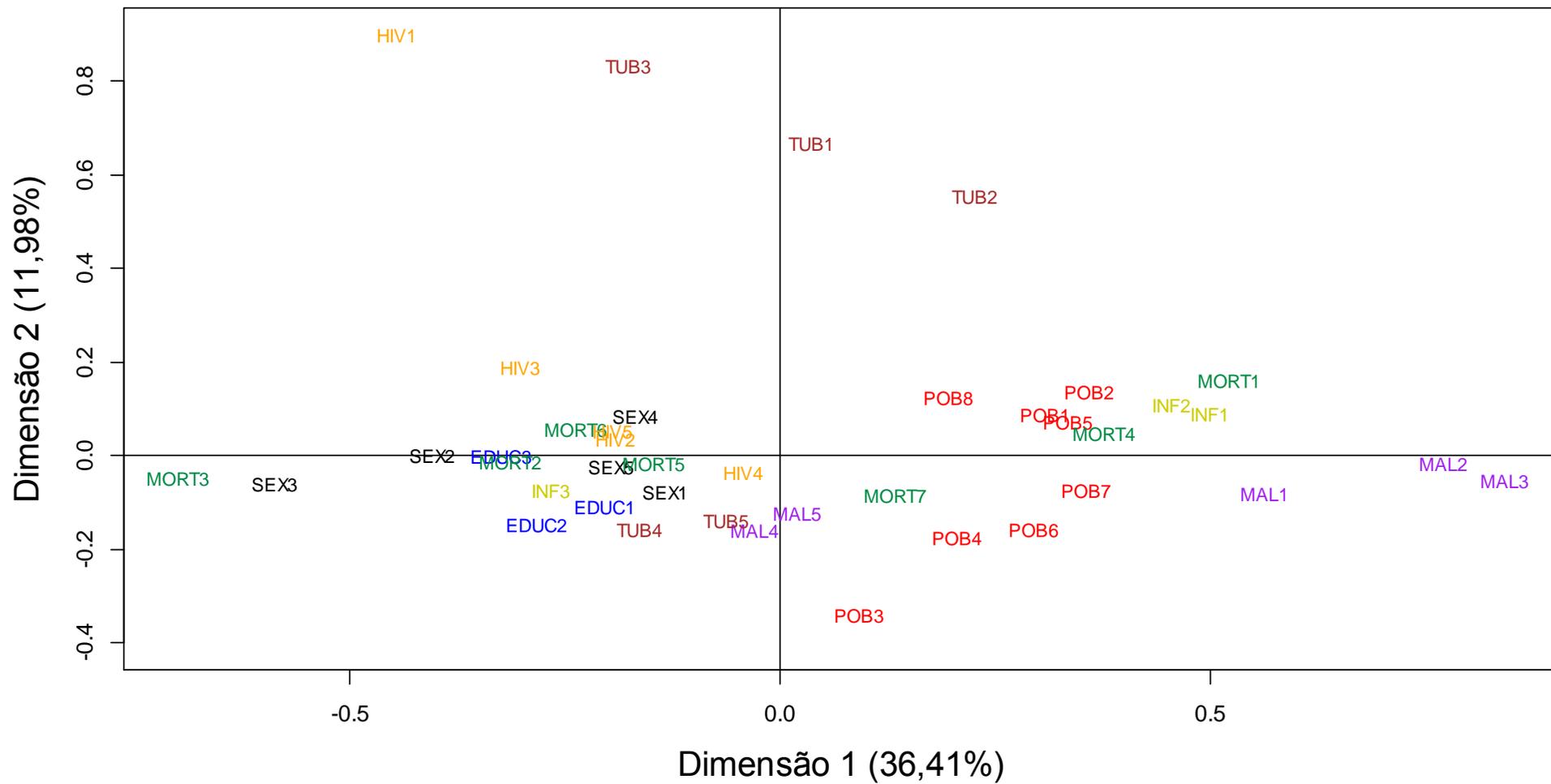


Figura 5: Gráfico da Análise de Correspondência para as variáveis, dimensões 1 e 2

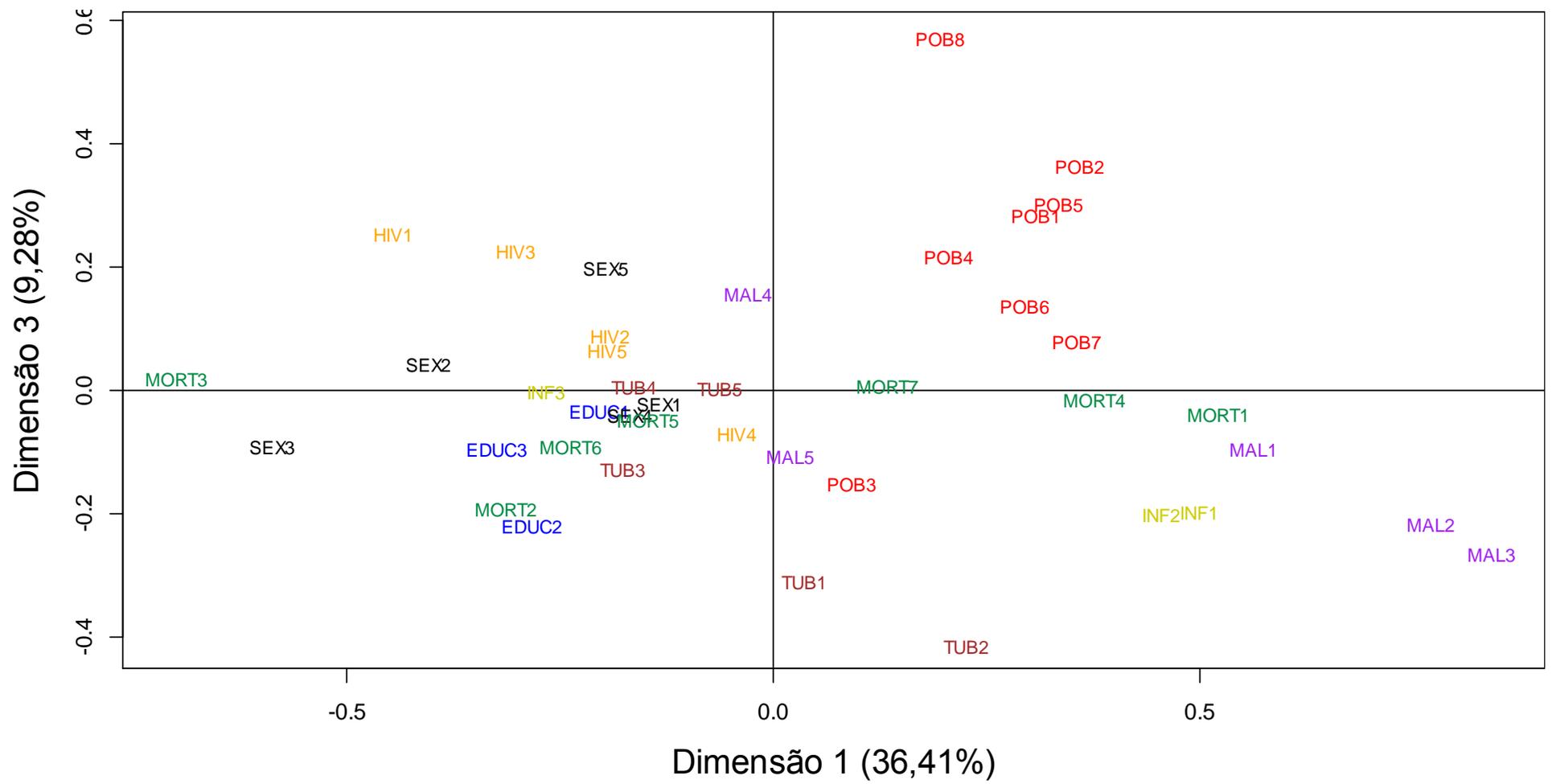


Figura 6: Gráfico da Análise de Correspondência para as variáveis, dimensões 1 e 3

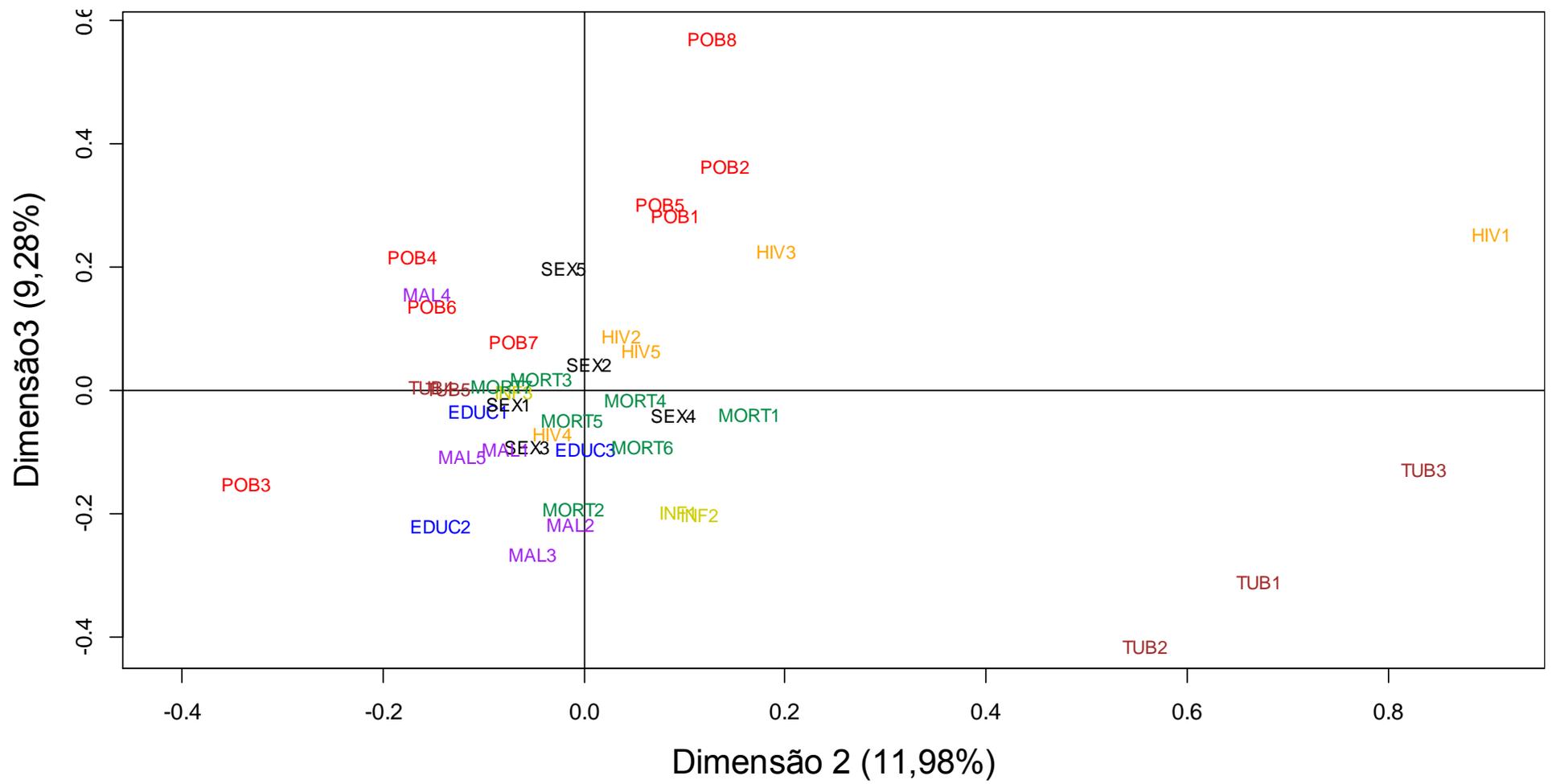


Figura 7: Gráfico da Análise de Correspondência para as variáveis, dimensões 2 e 3

As figuras 1, 2 e 3 apresentam os gráficos de correspondência dos países, formadas com as três primeiras dimensões. As figuras 4,5,6 apresentam os gráficos de correspondência das variáveis, formadas com as três primeiras dimensões. A lista dos indicadores é o Quadro 1 presente no Anexo do trabalho.

A dimensão 1 tem a seguinte característica (Figuras 5 e 6):

- ⊖ Do lado positivo, estão presentes variáveis associadas, principalmente, às condições de vida infanto-juvenil: a subnutrição (POB6 e POB7), mortalidade infantil (INF1 e INF2), mortalidade materna (MORT1), gravidez precoce (MORT4), a notificação (MAL1) e a morte (tanto adulto quanto infantil- MAL2 e MAL3) por malária.
- ⊖ Do lado negativo estão variáveis que tratam de conhecimento: total de matriculados no ensino primário (EDUC1), taxa de alfabetização de 15-24 anos (EDUC3), igualdade dos sexos na escola (do ensino primário até o superior – SEX1, SEX2, SEX3), participação da mulher no setor não agrícola (SEX4), crianças vacinadas (menos de 1 ano – INF3), conhecimento e uso correto de contraceptivos pelas mulheres casadas (MORT3), cobertura de cuidados pré-natais (MORT5 e MORT6), homens com conhecimento correto do HIV (HIV2) e cobertura de tratamento retroviral para pessoas infectadas pelo HIV(HIV5).

A segunda dimensão assim se apresenta (Figuras 5 e 7):

- Na parte positiva há principalmente o contexto da tuberculose: sua prevalência (TUB1), incidência (TUB3), morte (TUB2) e pessoas de 15-49 anos com HIV (HIV1)
- Na parte negativa envolve a pobreza: Partes mais pobre do quintil do rendimento nacional ou o consumo (POB3), crianças em tratamento com medicamentos anti-malária (MAL5), detecção e tratamento da tuberculose via DOTS (TUB4 e TUB5) e falta de planejamento familiar (MORT7).

A terceira dimensão tem a sua separação na seguinte forma (Figuras 6 e 7):

- Na parte positiva, variáveis que tratam a pobreza (POB1, POB2, POB5), taxa de emprego (POB4) a subnutrição no geral (POB8), o conhecimento correto

feminino sobre HIV (HIV3) e crianças que dormem sob mosquiteiros tratados com inseticida (MAL4).

- Na parte negativa, estão presentes variáveis que medem a quantidade de assentos em parlamentos ocupados por mulheres (SEX5), alunos que iniciam e que chegam ao último ano do ensino primário (EDUC2), partos assistidos por pessoas especializadas (MORT2) e relação entre taxa de frequência escolar dos órfãos para a taxa de frequência escolar de não órfãos (HIV4).

Algumas variáveis se tornaram próximas de outras, mesmo que não pertencessem ao mesmo objetivo. No gráfico, pode-se notar a proximidade entre as seguintes variáveis:

- A porcentagem de partos assistidos por pessoas da área da saúde qualificadas (MORT2) e Alunos que iniciam e que chegam ao último ano do ensino primário (EDUC2)
- Homens de 15-24 anos com conhecimento correto sobre o HIV(HIV2) e a porcentagem da cobertura terapêutica anti-retroviral entre as pessoas com infecção HIV avançada(HIV5)
- A porcentagem da taxa de sucesso do tratamento da tuberculose sob DOTS (TUB5) e a relação entre taxa de frequência escolar dos órfãos para a taxa de frequência escolar de não órfãos (HIV4)
- A taxa de gravidez na adolescência (MORT4) e crianças com menos de 5 anos moderadamente ou severamente abaixo do peso (POB7)
- Paridade de gênero no ensino primário (SEX1), porcentagem de cobertura de cuidados pré-natais - pelo menos uma visita (MORT5), a porcentagem da taxa de detecção de tuberculose sob DOTS (TUB4), total de matriculados na educação básica (EDUC1) e crianças com menos de 1 ano vacinadas (INF3).

Análise da Nuvem de Países

Sobre os países e sua contribuição para a inércia em cada dimensão, temos que:

- Na primeira dimensão (Figuras 2 e 3):
 - Os países do lado negativo são Tunísia, Argélia, Ilhas Maurício – estes são também os grandes contribuintes para a inércia total – Botswana, Cabo Verde, Quênia e Marrocos;

- Do lado positivo os países são Chade, República Central Africana, Democrática República do Congo, Burkina Faso, Guiné, Guiné-Bissau, Libéria, Mali, Níger, Nigéria e Somália.
- Na segunda dimensão (Figuras 2 e 4):
 - Os países Angola, Moçambique, Namíbia, Suazilândia, África do Sul, Zimbábue e Lesoto pertencem ao lado positivo;
 - Do lado negativo temos Egito, Benin, Gana, Senegal, São Tomé e Príncipe e Togo.
- Na terceira dimensão (Figuras 3 e 4):
 - Do lado positivo temos como os países Burundi, Comores, Eritreia, Etiópia, Madagascar, Malawi, Ruanda, UR Tanzânia, Uganda e Zâmbia;
 - Camarões, Congo, Costa do Marfim, Gabão, Gâmbia, Mauritânia, Serra Leoa, Djibouti e Guiné Equatorial estão presentes do lado negativo.

Para a interpretação da visualização dos países no gráfico também foi utilizado o resultado das coordenadas que mais contribuíram para a inércia total e pode-se verificar, imediatamente, a semelhança dos países que pertencem à mesma divisão regional africana. Temos que os países pertencentes às regiões do Norte da África (Argélia, Marrocos, Tunísia, Egito) e África Meridional (África do Sul, Suazilândia, Namíbia, Botsuana e Lesoto) estão presentes do lado esquerdo da Figura 1 gerado pelas duas primeiras dimensões e formam conjuntos separados no gráfico que se associam dentro de cada grupo. Outra região africana identificada é a África Ocidental, no qual alguns de seus países estão muito próximos na parte inferior direita do gráfico (Guiné-Bissau, Costa do Marfim, Burkina Faso, Níger, Guiné, Nigéria e Mali).

Levando em consideração apenas este gráfico, há uma nuvem densa de pontos que representam países em seu centro que, na verdade, têm como melhor representação os gráficos que contém a dimensão 3. Nestes gráficos, estes países tem uma melhor visualização, e conclui-se que a grande maioria destes países na nuvem pertence à região denominada África Oriental (Ruanda, Etiópia, Eritreia, Quênia, Burundi, Madagascar, Gâmbia, Uganda, Malawi, Tanzânia, Comores, Zimbábue e Zâmbia).

Mesmo que muitos países da mesma região acabem se associando, alguns países que fugiram a este padrão de relação regional. É o caso de Cabo Verde e Marrocos; Maurícia e a região do norte africano; República Centro-Africana, Somália e Serra Leoa, Libéria e Moçambique; Congo, Angola, Camarões e Gâmbia que não pertencem à mesma região, mas estão próximos no gráfico. Uma forma de tentar explicar esta aproximação pode ser feita utilizando o ranking de IDH (Índice de Desenvolvimento Humano). Todos estes países estão próximos no *ranking* do IDH, sugerindo que a aproximação observada no resultado da ACS entre países de regiões diferentes está de acordo com o desenvolvimento socioeconômico captado pelo IDH.

Há também casos de países, principalmente da região Ocidental e Central, que se juntam em pequenos grupos, e se distanciam um pouco dos outros países da mesma região. É o caso de Senegal, Benin e Togo, que se separam um pouco de São Tomé e Príncipe e Gana. Todos eles pertencem à região Ocidental Africana.

Análise Conjunta das Nuvens de Países e Indicadores

O que será feito a partir de agora é uma análise da relação entre as nuvens de países e de variáveis de acordo com suas posições nos quadrantes das dimensões.

Como a maioria das regiões formam nuvens de pontos razoavelmente separadas, inicialmente percebe-se a relação entre a África Meridional com a incidência de doenças infecciosas: a porcentagem de pessoas que vivem com HIV, entre 15-49 anos de idade (HIV1), a taxa de mortalidade da tuberculose por ano por 100.000 habitantes (TUB2) e a taxa de incidência de tuberculose por ano por 100.000 habitantes (TUB3).

O norte da África e a Maurícia apresentam um comportamento de proximidade com algumas variáveis favoráveis às mulheres: a porcentagem do uso atual de contraceptivos entre mulheres casadas 15-49 anos de idade, qualquer método (MORT3) e o índice de paridade de gênero no nível de escolarização superior (SEX3).

A África Oriental apresenta relações com algumas variáveis como a porcentagem de crianças menores de 5 anos que dormem sob mosquiteiros tratados com inseticida (MAL4), a porcentagem de assentos ocupados por mulheres no parlamento nacional (SEX5) e a porcentagem de empregos em relação de população, ambos os sexos (POB4). Isso indica que estes países não tem muitos problemas com malária e sua população tem uma quantidade de empregados maior que os outros países, assim como a igualdade de sexo na política.

Já a África Ocidental pode ser analisada com a sua proximidade com algumas variáveis que medem empregados em situação de risco e baixa qualidade de vida infantil: a porcentagem da proporção de trabalhadores familiares autônomos e contribuintes no total de empregados, ambos os sexos (POB6) e a porcentagem de crianças menores de 5 moderadamente ou severamente abaixo do peso, porcentagem (POB7); indicando que estes países estão caracterizados pela pobreza de suas populações.

Os países Senegal, Benin e Togo, que formam um subgrupo de países da região Ocidental africana, apresentam uma relação mais próxima com a variável que mede a porcentagem da necessidade não atendida de planejamento familiar total (MORT7), que é um indicador ligado à pobreza e baixa qualidade de vida infantil, presente também em outros países da região Ocidental.

Somália, Serra Leoa e República Centro-Africana são países que se relacionam com variáveis que medem a taxa de mortalidade infantil crianças menores de cinco anos e de 0-1 ano por 1.000 nascidos vivos (INF2 e INF1 respectivamente). Estes países apresentam valores muito altos para estes índices que medem a falta de qualidade de vida infantil, que também envolve pobreza e saúde.

Moçambique e Libéria se relacionam com a porcentagem da população que vive com abaixo de US \$ 1 por dia (POB1), a porcentagem da relação hiato de pobreza de US \$ 1 por dia (POB2) e a porcentagem da proporção de pessoas empregadas que vivem abaixo de US \$ 1 por dia (POB5), ou seja, variáveis que

puramente medem a pobreza em si e se conclui que estes países apresentam populações mais pobres entre todos os outros africanos.

CONSIDERAÇÕES FINAIS

No presente trabalho problemas muito comuns da área de estatística surgiram e foram adequadamente equacionados. Na formação do banco de dados tanto a ocorrência de dados faltantes como de variáveis com diversas métricas foram tratados de forma adequada permitindo a análise de um complexo sistema de informações sobre as condições de vida dos países da África.

Em seguida, a utilização da Análise de Correspondência Simples nos dados padronizados e completos apresentou a disposição gráfica dos pontos que representavam países e variáveis e as análises destes pontos precisaram ser feitas a partir de gráficos bidimensionais formados com as 3 dimensões principais.

Os gráficos mostraram basicamente similaridades entre países da mesma região, variáveis que se relacionavam de acordo com seu objetivo e a separação por área em cada dimensão, por exemplo, a incidência de HIV e Tuberculose em um extremo, e seu tratamento no outro extremo da dimensão. Foi interessante poder observar esta disposição dos pontos no gráfico, pois nos ensinou um pouco melhor sobre os problemas regionais da África, mostrando a sua diversidade no grau de intensidade dos problemas na saúde, educação e economia.

ANEXO

Quadro 1: As principais variáveis usadas pela IAEG para medir o progresso dos objetivos do milênio:

Sigla	Objetivo 1:
(POB1)	•População abaixo de US \$ 1 (PPP) por dia, porcentagem;
(POB2)	•Relação hiato de pobreza de US \$ 1 por dia (PPP), porcentagem;
(POB3)	•Partes mais pobre do quintil do rendimento nacional ou o consumo, porcentagem;
	•Taxa de crescimento do PIB por pessoa empregada, porcentagem;
(POB4)	•Empregos em relação de população, ambos os sexos, porcentagem;
(POB5)	•Proporção de pessoas empregadas que vivem abaixo de US \$ 1 (PPP) por dia, porcentagem;
(POB6)	•Proporção de trabalhadores familiares autônomos e contribuintes no total de empregados, ambos os sexos, porcentagem;
(POB7)	•Crianças menores de 5 moderadamente ou severamente abaixo do peso, porcentagem;
(POB8)	•População subnutrida, porcentagem;
	Objetivo 2:
(EDUC1)	•Taxa de matrícula líquida total no ensino fundamental, ambos os sexos;
(EDUC2)	•Alunos que iniciam e que chegam no último ano do ensino primário, ambos os sexos, porcentagem;
(EDUC3)	•Taxa de alfabetização de 15-24 anos de idade, ambos os sexos, porcentagem;
	Objetivo 3:
(SEX1)	•Índice de paridade de gênero no nível de escolarização primária;
(SEX2)	•Índice de paridade de gênero no nível de matrículas no ensino secundário;
(SEX3)	•Índice de paridade de gênero no nível de escolarização superior;
(SEX4)	•Participação das mulheres assalariadas no setor não-agrícola;
(SEX5)	•Assentos ocupados por mulheres no parlamento nacional, porcentagem;
	Objetivo 4:
(INF1)	•Taxa de mortalidade em crianças menores de cinco anos por 1.000 nascidos vivos;
(INF2)	•Taxa de mortalidade infantil (0-1 ano) por 1.000 nascidos vivos;
(INF3)	•Crianças de 1 ano vacinadas contra o sarampo, porcentagem;
	Objetivo 5:
(MORT1)	•Taxa de mortalidade materna por 100.000 nascidos vivos;
(MORT2)	•Partos assistidos por pessoas da área da saúde qualificadas, porcentagem;
(MORT3)	•O uso atual de contraceptivos entre mulheres casadas 15-49 anos de idade, qualquer método, porcentagem;

Continua

(MORT4)	•Taxa de natalidade adolescente, por 1.000 mulheres;
(MORT5)	•Cobertura de cuidados pré-natais, pelo menos uma visita, percentagem;
(MORT6)	•Cobertura de cuidados pré-natais, pelo menos quatro consultas, percentagem;
(MORT7)	•Necessidade não atendida de planeamento familiar, total, percentagem;
	Objetivo 6:
	•Pessoas que vivem com HIV, entre 15-49 anos de idade, percentagem;
(HIV1)	•O uso do preservativo na última relação sexual de alto risco, entre 15-24 anos de idade, mulheres, percentagem;
	•O uso do preservativo na última relação sexual de alto risco, entre 15-24 anos, homens, percentagem;
	•Homens entre 15-24 anos com conhecimento correto e abrangente sobre o HIV/AIDS, percentagem;
(HIV2)	•Mulheres entre 15-24 anos com conhecimento correto e abrangente sobre o HIV/AIDS, percentagem;
(HIV3)	•Relação entre taxa de frequência escolar dos órfãos para a taxa de frequência escolar de não órfãos;
(HIV4)	•Cobertura terapêutica anti-retroviral entre as pessoas com infecção HIV avançada, a percentagem;
(HIV5)	•Casos notificados de malária por 100.000 habitantes;
(MAL1)	•Taxa de mortalidade da malária por 100.000 habitantes, todas as idades;
(MAL2)	•Taxa de mortalidade da malária por 100.000 habitantes, com idades entre 0-4;
(MAL3)	•Crianças menores de 5 anos que dormem sob mosquiteiros tratados com inseticida , percentagem;
(MAL4)	•Crianças menores de 5 com febre sendo tratados com medicamentos anti-maláricos , percentagem;
(MAL5)	•Taxa de prevalência de tuberculose por 100.000 habitantes (ponto médio);
(TUB1)	•Taxa de mortalidade da tuberculose por ano por 100.000 habitantes (ponto médio);
(TUB2)	•Taxa de incidência de tuberculose por ano por 100.000 habitantes (ponto médio);
(TUB3)	•Taxa de detecção de tuberculose sob DOTS, percentual (ponto médio);
(TUB4)	•Taxa de sucesso do tratamento da tuberculose sob DOTS, percentagem;
(TUB5)	
	Objetivo 7:
	•Proporção de área terrestre coberta por florestas;
	•As emissões de CO2, o total, per capita e por cada \$ 1 do PIB (PPP);
	•O consumo de substâncias que empobrecem a camada de ozônio;
	•Proporção de unidades populacionais de peixes dentro dos limites biológicos de segurança;
	•Proporção do total de recursos hídricos usados;
	•Proporção de áreas terrestres e marinhas protegidas;
	•Proporção de espécies ameaçadas de extinção;
	•Proporção da população que utiliza uma fonte de água potável;

Continua

	•Proporção da população que utiliza instalações sanitárias melhoradas;
	•Proporção da população urbana que vive em favelas;
	Objetivo 8:
	APD líquida , total e para os países menos desenvolvidos , em percentagem do rendimento nacional bruto dos doadores da OCDE / CAD;
	•Proporção da APD bilateral total, atribuível setorialmente de doadores da OCDE / CAD aos serviços sociais básicos (educação básica, cuidados primários de saúde , nutrição, água potável e saneamento);
	•Proporção de assistência oficial ao desenvolvimento bilateral de doadores da OCDE / CAD que não está vinculada;
	•APD recebida nos países em desenvolvimento sem litoral como uma proporção dos seus rendimentos nacionais brutos;
	•APD recebida nos pequenos Estados insulares em desenvolvimento como proporção dos seus rendimentos nacionais brutos;
	•Proporção das importações totais dos países desenvolvidos (por valor e excluindo armas) provenientes de países em desenvolvimento e os países menos desenvolvidos , admitidas com isenção de direitos;
	•As tarifas médias impostas pelos países desenvolvidos aos produtos agrícolas e têxteis e vestuário de países em desenvolvimento;
	•Apoio agrícola estimado nos países da OCDE, como uma percentagem do seu produto interno bruto;
	•Proporção da APD fornecida para ajudar a construir a capacidade comercial;
	•Número total de países que alcançaram os seus pontos de decisão PPAE e número que tenham atingido os seus pontos de conclusão PPAE (cumulativo);
	•Alívio da dívida comprometido no âmbito de iniciativas HIPC e MDRI;
	•Serviço da dívida como percentagem das exportações de bens e serviços;
	•Proporção da população com acesso a medicamentos essenciais a preços acessíveis numa base sustentável;
	•As linhas telefónicas fixas por 100 habitantes;
	•Assinaturas celulares móveis por 100 habitantes;
	•Utilizadores da Internet por 100 habitantes;

Tabela 1: Resultados da decomposição da inércia e Qui-quadrado:

Valor Singular	Inércia Principal	Chi-Quadrado	Porcentagem	Porcentagem Acumulada	Scree Plot
0.32802	0.10760	102.525	36.41	36.41	*****
0.18814	0.03540	33.726	11.98	48.38	*****
0.16562	0.02743	26.135	9.28	57.67	*****
0.14264	0.02035	19.387	6.88	64.55	*****

Continua

Valor Singular	Inércia Principal	Chi-Quadrado	Porcentagem	Porcentagem Acumulada	<i>Scree Plot</i>
0.11945	0.01427	13.596	4.83	69.38	***
0.10620	0.01128	10.746	3.82	73.19	***
0.10240	0.01049	9.991	3.55	76.74	***
0.09192	0.00845	8.051	2.86	79.60	**
0.08569	0.00734	6.997	2.48	82.09	**
0.08143	0.00663	6.318	2.24	84.33	**
0.07501	0.00563	5.361	1.90	86.23	*
0.07182	0.00516	4.915	1.75	87.98	*
0.06459	0.00417	3.976	1.41	89.39	*
0.06288	0.00395	3.767	1.34	90.73	*
0.06117	0.00374	3.565	1.27	91.99	*
0.05531	0.00306	2.914	1.03	93.03	*
0.05101	0.00260	2.479	0.88	93.91	*
0.04636	0.00215	2.047	0.73	95.50	*
0.04335	0.00188	1.791	0.64	96.13	
0.04085	0.00167	1.590	0.56	96.70	
0.04007	0.00161	1.530	0.54	97.24	
0.03859	0.00149	1.419	0.50	97.75	
0.03569	0.00127	1.214	0.43	98.18	
0.02995	0.00090	0.855	0.30	98.48	
0.02820	0.00080	0.758	0.27	98.75	
0.02588	0.00067	0.638	0.23	98.98	
0.02483	0.00062	0.588	0.21	99.19	
0.02211	0.00049	0.466	0.17	99.35	
0.02173	0.00047	0.450	0.16	99.51	
0.01923	0.00037	0.352	0.13	99.64	
0.01642	0.00027	0.257	0.09	99.73	
0.01541	0.00024	0.226	0.08	99.81	
0.01430	0.00020	0.195	0.07	99.88	
0.01078	0.00012	0.111	0.04	99.92	
0.01027	0.00011	0.100	0.04	99.95	
0.00762	0.00006	0.055	0.02	99.97	
0.00662	0.00004	0.042	0.01	99.99	
0.00577	0.00003	0.032	0.01	100.00	
0.00278	0.00001	0.007	0.00	100.00	

Tabela 2: Coordenadas, qualidade, massa e inércia dos pontos das linhas.

	Coordenadas das linhas			Qualidade	Massa	Inércia
	Dim1	Dim2	Dim3			
Algeria	-0.5959	-0.2714	-0.1871	0.8065	0.0171	0.0332
Angola	0.1372	0.0983	-0.0185	0.1514	0.0188	0.0121
Benin	0.2004	-0.2059	-0.0246	0.5104	0.0204	0.0113
Botswana	-0.4927	0.2170	0.0380	0.6876	0.0213	0.0305
Burkina Faso	0.3848	-0.1550	-0.0038	0.6110	0.0211	0.0202
Burundi	0.0928	-0.0886	0.3465	0.5610	0.0238	0.0196
Cameroon	0.0885	-0.0203	-0.1562	0.3227	0.0201	0.0069
Cape Verde	-0.5400	-0.0844	-0.0973	0.8118	0.0186	0.0239
C.A Republic	0.4886	0.2357	-0.0884	0.7426	0.0204	0.0281
Chad	0.7346	-0.0644	0.0398	0.8016	0.0199	0.0458
Comoros	0.0456	-0.1462	0.1924	0.1539	0.0182	0.0242
Congo	0.0763	0.0485	-0.0584	0.1145	0.0220	0.0075
D. Rp.Congo	0.3997	0.0664	-0.0275	0.5329	0.0215	0.0225
Cote d'Ivoire	0.2707	-0.0526	-0.1407	0.5901	0.0176	0.0097
Djibouti	-0.1549	0.1935	-0.2899	0.3447	0.0160	0.0229
Egypt	-0.5763	-0.4062	-0.2674	0.8002	0.0146	0.0352
Equatorial Guinea	0.1302	-0.0511	-0.2751	0.3051	0.0162	0.0171
Eritrea	-0.0549	-0.0964	0.2365	0.1636	0.0170	0.0240
Ethiopia	0.0558	-0.1474	0.1760	0.1865	0.0187	0.0190
Gabon	-0.1934	0.0478	-0.2187	0.4706	0.0193	0.0121
Gambia	0.0557	0.0519	-0.1125	0.1608	0.0198	0.0077
Ghana	-0.0761	-0.1811	-0.0659	0.3622	0.0203	0.0082
Guinea	0.3943	-0.0701	-0.1099	0.6959	0.0193	0.0162
Guinea-Bissau	0.2942	-0.0372	-0.1234	0.4702	0.0200	0.0148
Kenya	-0.1984	0.0125	0.0991	0.4736	0.0206	0.0073
Lesotho	-0.3096	0.2714	0.1171	0.5829	0.0226	0.0240
Liberia	0.3073	0.0437	0.1007	0.4621	0.0212	0.0165
Madagascar	0.0424	0.0089	0.2447	0.2397	0.0198	0.0173
Malawi	-0.0075	-0.0325	0.1992	0.4229	0.0234	0.0076
Mali	0.4022	-0.1640	-0.0882	0.7098	0.0202	0.0189
Mauritania	0.0677	0.1646	-0.3496	0.3557	0.0153	0.0224
Mauritius	-0.5687	-0.2400	-0.0671	0.7360	0.0179	0.0317
Morocco	-0.5483	-0.2834	-0.1207	0.8294	0.0156	0.0252
Mozambique	0.1985	0.1774	0.1193	0.4645	0.0223	0.0138
Namibia	-0.4741	0.2803	0.0746	0.8248	0.0216	0.0274
Niger	0.4187	-0.1743	-0.0425	0.5887	0.0183	0.0218
Nigeria	0.2955	-0.0730	-0.0665	0.5334	0.0201	0.0124

Continua

	Coordenadas das linhas			Qualidade	Massa	Inércia
	Dim1	Dim2	Dim3			
Rwanda	-0.1901	-0.0883	0.3702	0.6220	0.0219	0.0215
Senegal	0.0136	-0.1731	0.0445	0.1881	0.0192	0.0111
Sierra Leone	0.3766	0.2678	-0.2833	0.7318	0.0252	0.0343
Somalia	0.3652	0.1270	-0.1653	0.2680	0.0162	0.0361
South Africa	-0.5247	0.3539	-0.0893	0.8157	0.0215	0.0364
Swaziland	-0.3854	0.5697	0.0200	0.8846	0.0223	0.0403
S. T. and Principe	-0.2177	-0.1461	-0.0338	0.4017	0.0214	0.0126
U.Rep. of Tanzania	-0.0183	-0.0911	0.2058	0.6262	0.0242	0.0067
Togo	0.1043	-0.1457	0.0242	0.2219	0.0197	0.0098
Tunisia	-0.6538	-0.3179	-0.1367	0.8476	0.0182	0.0398
Uganda	0.0444	-0.0387	0.1436	0.1828	0.0212	0.0094
Zambia	0.0345	0.1335	0.2412	0.6176	0.0235	0.0099
Zimbabwe	-0.0771	0.1808	0.1232	0.3422	0.0248	0.0132

Tabela 3: Contribuição parcial da Inércia e o quadrado do cosseno para os pontos de linhas

	Contribuição para a Inércia			Cosseno ao Quadrado		
	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3
Algeria	0.0564	0.0356	0.0218	0.6176	0.1281	0.0609
Angola	0.0033	0.0051	0.0002	0.0989	0.0507	0.0018
Benin	0.0076	0.0245	0.0004	0.2466	0.2601	0.0037
Botswana	0.0480	0.0283	0.0011	0.5730	0.1112	0.0034
Burkina Faso	0.0291	0.0144	0.0000	0.5256	0.0853	0.0001
Burundi	0.0019	0.0053	0.1044	0.0354	0.0323	0.4934
Cameroon	0.0015	0.0002	0.0179	0.0775	0.0041	0.2412
Cape Verde	0.0504	0.0037	0.0064	0.7681	0.0187	0.0249
C.A Republic	0.0453	0.0320	0.0058	0.5868	0.1366	0.0192
Chad	0.0999	0.0023	0.0012	0.7932	0.0061	0.0023
Comoros	0.0004	0.0110	0.0245	0.0053	0.0544	0.0942
Congo	0.0012	0.0015	0.0027	0.0576	0.0233	0.0337
D. Rp. Congo	0.0319	0.0027	0.0006	0.5162	0.0142	0.0024

Continua

	Contribuição para a Inércia			Cosseno ao Quadrado		
	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3
Cote d'Ivoire	0.0120	0.0014	0.0127	0.4512	0.0170	0.1218
Djibouti	0.0036	0.0170	0.0491	0.0568	0.0887	0.1991
Egypt	0.0452	0.0682	0.0382	0.4674	0.2321	0.1007
Equatorial Guinea	0.0026	0.0012	0.0447	0.0543	0.0084	0.2424
Eritrea	0.0005	0.0045	0.0347	0.0072	0.0223	0.1342
Ethiopia	0.0005	0.0115	0.0211	0.0104	0.0726	0.1035
Gabon	0.0067	0.0012	0.0336	0.2012	0.0123	0.2571
Gambia	0.0006	0.0015	0.0091	0.0271	0.0234	0.1103
Ghana	0.0011	0.0189	0.0032	0.0488	0.2768	0.0366
Guinea	0.0279	0.0027	0.0085	0.6274	0.0198	0.0487
Guinea- Bissau	0.0161	0.0008	0.0111	0.3945	0.0063	0.0694
Kenya	0.0075	0.0001	0.0074	0.3777	0.0015	0.0943
Lesotho	0.0201	0.0469	0.0113	0.3050	0.2343	0.0436
Liberia	0.0186	0.0011	0.0078	0.4098	0.0083	0.0440
Madagascar	0.0003	0.0000	0.0432	0.0070	0.0003	0.2324
Malawi	0.0000	0.0007	0.0339	0.0006	0.0109	0.4114
Mali	0.0304	0.0154	0.0057	0.5844	0.0972	0.0281
Mauritania	0.0007	0.0117	0.0681	0.0106	0.0626	0.2825
Mauritius	0.0538	0.0291	0.0029	0.6175	0.1099	0.0086
Morocco	0.0436	0.0354	0.0083	0.6305	0.1684	0.0305
Mozambique	0.0082	0.0198	0.0115	0.2150	0.1718	0.0777
Namibia	0.0452	0.0480	0.0044	0.6002	0.2097	0.0149
Niger	0.0298	0.0157	0.0012	0.4975	0.0862	0.0051
Nigeria	0.0163	0.0030	0.0032	0.4798	0.0293	0.0243
Rwanda	0.0073	0.0048	0.1093	0.1242	0.0268	0.4710
Senegal	0.0000	0.0163	0.0014	0.0011	0.1754	0.0116
Sierra Leone	0.0333	0.0512	0.0738	0.3533	0.1787	0.1999
Somalia	0.0200	0.0074	0.0161	0.2021	0.0245	0.0414

Continua

	Contribuição para a Inércia			Cosseno ao Quadrado		
	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3
South Africa	0.0549	0.0759	0.0062	0.5497	0.2501	0.0159
Swaziland	0.0307	0.2042	0.0003	0.2775	0.6064	0.0007
S. T. and Principe	0.0094	0.0129	0.0009	0.2723	0.1228	0.0066
U.Rep. of Tanzania	0.0001	0.0057	0.0373	0.0041	0.1019	0.5202
Togo	0.0020	0.0118	0.0004	0.0738	0.1441	0.0040
Tunisia	0.0724	0.0520	0.0124	0.6622	0.1565	0.0290
Uganda	0.0004	0.0009	0.0159	0.0149	0.0114	0.1564
Zambia	0.0003	0.0118	0.0498	0.0095	0.1426	0.4655
Zimbabwe	0.0014	0.0229	0.0137	0.0378	0.2079	0.0965

Tabela 4: Índices das coordenadas que mais contribui à inércia para os pontos de linhas

	Dim1	Dim2	Dim3	Best
Algeria	1	1	1	1
Angola	0	0	0	2
Benin	0	2	0	2
Botswana	1	1	0	1
Burkina Faso	1	0	0	1
Burundi	0	0	3	3
Cameroon	0	0	3	3
Cape Verde	1	0	0	1
C.A Republic	1	1	0	1
Chad	1	0	0	1
Comoros	0	0	3	3
Congo	0	0	0	3
Democratic Rp.Congo	1	0	0	1
Cote d'Ivoire	0	0	0	3
Djibouti	0	3	3	3

Continua

	Dim1	Dim2	Dim3	Best
Egypt	2	2	2	2
Equatorial Guinea	0	0	3	3
Eritrea	0	0	3	3
Ethiopia	0	0	3	3
Gabon	0	0	3	3
Gambia	0	0	0	3
Ghana	0	2	0	2
Guinea	0	0	0	1
Guinea- Bissau	0	0	0	1
Kenya	0	0	0	1
Lesotho	0	2	0	2
Liberia	0	0	0	1
Madagascar	0	0	3	3
Malawi	0	0	3	3
Mali	1	0	0	1
Mauritania	0	0	3	3
Mauritius	1	1	0	1
Morocco	1	1	0	1
Mozambique	0	2	0	2
Namibia	2	2	0	2
Niger	1	0	0	1
Nigeria	0	0	0	1
Rwanda	0	0	3	3
Senegal	0	0	0	2
Sierra Leone	3	3	3	3
Somalia	0	0	0	1
South Africa	2	2	0	2
Swaziland	2	2	0	2
Sao T. and Principe	0	0	0	2
Un. Rep. of Tanzania	0	0	3	3
Togo	0	0	0	2

Continua

	Dim1	Dim2	Dim3	Best
Tunisia	1	1	0	1
Uganda	0	0	0	3
Zambia	0	0	3	3
Zimbabwe	0	2	0	2

Tabela 5: Coordenadas, qualidade, massa e inércia dos pontos das colunas

	Coordenadas das colunas			Qualidade	Massa	Inércia
	Dim1	Dim2	Dim3			
POB1	0.3067	0.0894	0.2876	0.7187	0.0241	0.0210
POB2	0.3569	0.1393	0.3667	0.5754	0.0161	0.0267
POB3	0.0909	-0.3367	-0.1481	0.4925	0.0252	0.0249
POB4	0.2044	-0.1724	0.2200	0.5028	0.0274	0.0221
POB5	0.3331	0.0742	0.3053	0.7664	0.0238	0.0220
POB6	0.2936	-0.1526	0.1392	0.5821	0.0326	0.0244
POB7	0.3537	-0.0714	0.0815	0.3677	0.0215	0.0270
POB8	0.1937	0.1268	0.5743	0.4920	0.0140	0.0370
EDUC1	-0.2052	-0.1057	-0.0306	0.4239	0.0374	0.0162
EDUC2	-0.2841	-0.1438	-0.2177	0.5841	0.0286	0.0247
EDUC3	-0.3254	0.0009	-0.0937	0.5795	0.0339	0.0227
SEX1	-0.1342	-0.0748	-0.0199	0.3767	0.0382	0.0082
SEX2	-0.4041	0.0042	0.0447	0.6935	0.0220	0.0177
SEX3	-0.5863	-0.0568	-0.0869	0.6743	0.0192	0.0342
SEX4	-0.1680	0.0882	-0.0380	0.1639	0.0254	0.0196
SEX5	-0.1970	-0.0210	0.2010	0.1801	0.0160	0.0240
INF1	0.4985	0.0927	-0.1947	0.8376	0.0240	0.0287
INF2	0.4538	0.1133	-0.1992	0.8295	0.0238	0.0251
INF3	-0.2665	-0.0709	0.0021	0.5700	0.0364	0.0165
MORT1	0.5206	0.1640	-0.0349	0.6587	0.0184	0.0283
MORT2	-0.3134	-0.0106	-0.1891	0.6100	0.0292	0.0217
MORT3	-0.6994	-0.0437	0.0219	0.7736	0.0194	0.0417
MORT4	0.3755	0.0503	-0.0113	0.6399	0.0245	0.0186
MORT5	-0.1463	-0.0121	-0.0449	0.3414	0.0400	0.0093
MORT6	-0.2378	0.0585	-0.0879	0.3033	0.0283	0.0214
MORT7	0.1337	-0.0820	0.0098	0.1098	0.0294	0.0224

Continua

	Coordenadas das colunas			Qualidade	Massa	Inércia
	Dim1	Dim2	Dim3			
HIV1	-0.4459	0.9029	0.2564	0.6887	0.0095	0.0506
HIV2	-0.1913	0.0372	0.0924	0.4682	0.0278	0.0093
HIV3	-0.3011	0.1910	0.2300	0.6265	0.0211	0.0205
HIV4	-0.0431	-0.0312	-0.0678	0.1420	0.0371	0.0066
HIV5	-0.1954	0.0563	0.0688	0.2605	0.0261	0.0156
MAL1	0.5626	-0.0781	-0.0920	0.7538	0.0244	0.0362
MAL2	0.7692	-0.0134	-0.2138	0.8003	0.0170	0.0457
MAL3	0.8410	-0.0509	-0.2625	0.8311	0.0154	0.0489
MAL4	-0.0295	-0.1559	0.1607	0.1625	0.0217	0.0231
MAL5	0.0192	-0.1212	-0.1031	0.0600	0.0204	0.0296
TUB1	0.0368	0.6719	-0.3059	0.7996	0.0118	0.0273
TUB2	0.2251	0.5579	-0.4113	0.6201	0.0093	0.0270
TUB3	-0.1764	0.8352	-0.1252	0.9115	0.0108	0.0299
TUB4	-0.1635	-0.1529	0.0087	0.3838	0.0297	0.0131
TUB5	-0.0623	-0.1349	0.0071	0.2773	0.0390	0.0105

Tabela 6: Contribuição parcial da Inércia e o quadrado do cosseno para os pontos de colunas

	Contribuição para a Inércia			Cosseno ao Quadrado		
	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3
POB1	0.0211	0.0055	0.0728	0.3658	0.0311	0.3218
POB2	0.0191	0.0088	0.0790	0.2605	0.0397	0.2752
POB3	0.0019	0.0808	0.0202	0.0283	0.3889	0.0753
POB4	0.0106	0.0230	0.0484	0.1752	0.1246	0.2030
POB5	0.0245	0.0037	0.0808	0.4056	0.0201	0.3407
POB6	0.0261	0.0214	0.0230	0.3894	0.1052	0.0875
POB7	0.0249	0.0031	0.0052	0.3362	0.0137	0.0178
POB8	0.0049	0.0064	0.1688	0.0482	0.0206	0.4232
EDUC1	0.0147	0.0118	0.0013	0.3293	0.0873	0.0073
EDUC2	0.0215	0.0167	0.0495	0.3168	0.0812	0.1861
EDUC3	0.0334	0.0000	0.0109	0.5351	0.0000	0.0444
SEX1	0.0064	0.0060	0.0006	0.2826	0.0878	0.0062
SEX2	0.0333	0.0000	0.0016	0.6850	0.0001	0.0084
SEX3	0.0614	0.0018	0.0053	0.6538	0.0061	0.0144

Continua

	Contribuição para a Inércia			Cosseno ao Quadrado		
	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3
SEX4	0.0067	0.0056	0.0013	0.1235	0.0340	0.0063
SEX5	0.0058	0.0002	0.0236	0.0878	0.0010	0.0913
INF1	0.0555	0.0058	0.0332	0.7056	0.0244	0.1076
INF2	0.0455	0.0086	0.0344	0.6610	0.0412	0.1273
INF3	0.0241	0.0052	0.0000	0.5322	0.0377	0.0000
MORT1	0.0464	0.0140	0.0008	0.5968	0.0593	0.0027
MORT2	0.0267	0.0001	0.0381	0.4469	0.0005	0.1626
MORT3	0.0881	0.0010	0.0003	0.7698	0.0030	0.0008
MORT4	0.0320	0.0017	0.0001	0.6281	0.0113	0.0006
MORT5	0.0080	0.0002	0.0029	0.3101	0.0021	0.0291
MORT6	0.0149	0.0027	0.0080	0.2533	0.0153	0.0346
MORT7	0.0049	0.0056	0.0001	0.0795	0.0299	0.0004
HIV1	0.0176	0.2198	0.0229	0.1268	0.5200	0.0419
HIV2	0.0095	0.0011	0.0087	0.3684	0.0139	0.0859
HIV3	0.0178	0.0217	0.0406	0.3155	0.1269	0.1841
HIV4	0.0006	0.0010	0.0062	0.0354	0.0186	0.0879
HIV5	0.0092	0.0023	0.0045	0.2158	0.0179	0.0268
MAL1	0.0717	0.0042	0.0075	0.7206	0.0139	0.0193
MAL2	0.0932	0.0001	0.0283	0.7427	0.0002	0.0574
MAL3	0.1014	0.0011	0.0387	0.7548	0.0028	0.0735
MAL4	0.0002	0.0149	0.0205	0.0028	0.0774	0.0823
MAL5	0.0001	0.0085	0.0079	0.0009	0.0343	0.0248
TUB1	0.0001	0.1506	0.0403	0.0020	0.6607	0.1369
TUB2	0.0044	0.0819	0.0575	0.0592	0.3634	0.1975
TUB3	0.0031	0.2132	0.0062	0.0381	0.8542	0.0192
TUB4	0.0074	0.0196	0.0001	0.2044	0.1788	0.0006
TUB5	0.0014	0.0200	0.0001	0.0486	0.2280	0.0006

Tabela 7: Índices das coordenadas que mais contribuem à Inércia para os pontos de coluna

	Dim1	Dim2	Dim3	Best
POB1	0	0	3	3
POB2	0	0	3	3
POB3	0	2	0	2
POB4	0	3	3	3
POB5	3	0	3	3
POB6	1	1	0	1
POB7	1	0	0	1
POB8	0	0	3	3
EDUC1	0	0	0	1
EDUC2	3	0	3	3
EDUC3	1	0	0	1
SEX1	0	0	0	1
SEX2	1	0	0	1
SEX3	1	0	0	1
SEX4	0	0	0	1
SEX5	0	0	0	3
INF1	1	0	1	1
INF2	1	0	1	1
INF3	1	0	0	1
MORT1	1	0	0	1
MORT2	3	0	3	3
MORT3	1	0	0	1
MORT4	1	0	0	1
MORT5	0	0	0	1
MORT6	0	0	0	1
MORT7	0	0	0	2
HIV1	0	2	0	2
HIV2	0	0	0	1
HIV3	0	3	3	3
HIV4	0	0	0	3
HIV5	0	0	0	1
MAL1	1	0	0	1
MAL2	1	0	1	1

Continua

	Dim1	Dim2	Dim3	Best
MAL3	1	0	1	1
MAL4	0	0	0	3
MAL5	0	0	0	2
TUB1	0	2	2	2
TUB2	0	2	2	2
TUB3	0	2	0	2
TUB4	0	0	0	2
TUB5	0	0	0	2

Figura 8: Mapa político do continente africano.



REFERÊNCIA BIBLIOGRÁFICA

1. Centro de Estudos sobre a África, Ásia e América Latina. Glossário: Índice de Desenvolvimento Humano. Disponível em: [http://pascal.iseg.utl.pt/~cesa/index.php/dicionario-da-cooperacao/Glossary-1/%C3%8D/%C3%8Dndice-de-Desenvolvimento-Humano-\(IDH\)-261/](http://pascal.iseg.utl.pt/~cesa/index.php/dicionario-da-cooperacao/Glossary-1/%C3%8D/%C3%8Dndice-de-Desenvolvimento-Humano-(IDH)-261/) Acesso em Dezembro/2013
2. de Leeuw, J.& Mair, P. Simple and Canonical Correspondence Analysis Using the R Package anacor. Journal of Statistical Software, 31(5), 1-18. (2009).
3. GREENACRE,M..Correspondence Analysis in Practice, 2 ed.: Boca Raton: Chapman & Hall/CRC. (2007)
4. GREENACRE,M. & Blasius, J. Multiple Correspondence Analysis and Related Methods. Chapman & Hall/CRC. (2006)
5. História da ONU. Brasil Escola. Disponível em: <http://www.brasilecola.com/geografia/onu.htm> Acesso em Novembro/2013
6. Indicadores dos Objetivos do milênio: Site oficial da ONU para os MDG. Disponível em: <http://mdgs.un.org/unsd/mdg/Host.aspx?Content=Indicators/About.htm> Acesso em Agosto/2013.
7. Johnson,R. & Wichern,D.. Applied Multivariate Statistical Analysis, 6 ed:Pearson. (2007)
8. ONU:Road map towards the implementation of the United Nations Millennium Declaration. Disponível em: http://mdgs.un.org/unsd/mdg/Resources/Static/Products/SGReports/56_326/a_56_326_e.pdfSetembro,2001.

9. ONU. Questões Globais: África. Disponível em:
<http://www.un.org/en/globalissues/africa/> Acesso em Novembro/2013.

10. PNUD. Atlas do Desenvolvimento Humano 2013: Material de Apoio. Disponível em:
<http://www.pnud.org.br/arquivos/faq-atlas2013.pdf>. Acesso em Dezembro/2013.
11. SAS Institute Inc. 2013. SAS/STAT®12.3 User's Guide. Cary, NC:SAS Institute Inc.
12. Veroneze, Rosana. Tratamento de dados faltantes empregando biclusterização com imputação múltipla / Rosana Veroneze. Campinas, SP, 2011.
13. Yuan, C.Y. Multiple Imputation for Missing Data: Concepts and New Development (Version 9.0). SAS Institute Inc., Rockville, MD.