



## O uso do Sistema Gapminder como suporte ao ensino da Análise de Correlação Linear para o Ensino Superior

**Fernando Frei**<sup>1</sup>

Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Faculdade de Ciências e Letras, Departamento de Ciências Biológicas, Assis, SP, Brasil

### Resumo

A associação entre banco de dados e sistemas de visualização pode apresentar diversas vantagens para o ensino e aprendizagem de diversas técnicas estatísticas. Esta prática teve como objetivo apresentar como o sistema Gapminder pode ser utilizado como suporte para os processos de ensino e aprendizagem dos conceitos básicos da Análise de Correlação Simples para estudantes de cursos do Ensino Superior. A investigação foi desenvolvida utilizando pesquisa exploratória com enfoque qualitativo considerando um questionário enquanto instrumento de coleta de dados. Após a análise do desempenho dos estudantes foi possível verificar que os resultados são positivos, pois apresentam notas médias próximas de oito em uma escala de zero a dez pontos e baixa variabilidade, o que confere as duas turmas um bom desempenho nas avaliações propostas. Pelos resultados obtidos foi possível verificar que o sistema Gapminder pode auxiliar na construção e consolidação de conceitos introdutórios ligados aos estudos da relação linear entre duas variáveis quantitativas.

**Palavras-chave:** Estatística; Correlação; Gapminder.

### The Use of Gapminder System as a support to the Linear Correlation Analysis teaching to Higher Education

#### Abstract

The association between database and visualization system may present several advantages to teaching and learning various statistical techniques. This practice objected to presenting how Gapminder system might be used as a support to teaching and learning processes of Simple Correlation Analysis basic concepts to students from higher education courses. An investigation was developed by using exploratory research in a qualitative approach, considering a questionnaire as a data collection tool. After the students' performances analysis, it was possible to verify that the results were positive, once they presented passing scores close to eight on a zero to ten scale, and low variability, what gives the two groups a good development in the proposed evaluations. By the obtained results, it was possible to verify Gapminder system may assist in the construction and consolidation of introductory concepts linked to studies about linear relations between two quantitative variables.

**Keywords:** Statistics; Correlation; Gapminder.

---

**Submetido em:** 21/03/2020

**Aceito em:** 11/05/2020

**Publicado em:** 13/06/2020

<sup>1</sup> Doutor em Saúde Pública pela USP. Professor do Programa de Pós-Graduação em Biociências da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Faculdade de Ciências e Letras – Assis. Endereço para correspondência: Av. Dom Antônio, 2100, Assis – SP – CEP 19806-900. E-mail: fernando.frei@unesp.br.

## El uso del Sistema Gapminder como un soporte a la enseñanza de la Análisis de Correlación Lineal para la Educación Superior

### Resumen

La asociación entre el banco de datos y sistemas de visualización puede presentar diversas ventajas para la enseñanza y el aprendizaje de diversas técnicas estadísticas. Esta práctica tuvo como objetivo presentar cómo el sistema Gapminder puede ser utilizado como soporte para los procesos de enseñanza y aprendizaje de los conceptos básicos de la Análisis de Correlación Simple para estudiantes de cursos de la Educación Superior. La búsqueda fue desarrollada utilizando una investigación exploratoria con enfoque cualitativo, considerando un cuestionario como instrumento de recolección de datos. Tras la análisis del desarrollo de los estudiantes, fue posible verificar que los resultados fueron positivos, pues presentan un promedio de notas próximo a ocho en una escala de cero hasta diez puntos y baja variabilidad, lo que confiere a las dos clases un buen desarrollo en las evaluaciones propuestas. Por los resultados obtenidos, fue posible verificar que el sistema Gapminder puede asistir a la construcción y consolidación de conceptos introductorios vinculados a estudios de la relación lineal entre dos variables cuantitativas.

**Palabras clave:** Estadística; Correlación; Gapminder.

### 1. Introdução

As geo visualizações contribuem para demonstrar a natureza interdisciplinar das estatísticas oficiais e para captar o interesse dos estudantes, evidenciando as relações entre a estatística e outras disciplinas e entre as estatísticas e o mundo real (FORBES, 2012). Neste contexto, as práticas pedagógicas devem estimular a aprendizagem por meio de extrapolações sistematizadas, possibilitando a aquisição de conceitos estatísticos que produzam a capacidade de entender, interpretar e criticar resultados de pesquisas e o entendimento do mundo. (JAMIE, 2002; FERNANDES; DAMIN; SANTOS JÚNIOR, 2017).

Entre as estratégias educacionais, vários estudos sugerem que o tratamento visual das aulas pode melhorar a aprendizagem com vários graus de sucesso (STOKES, 2002). Segundo Kleinman e Dwyer (1999) o uso de gráficos coloridos, ao contrário dos gráficos em preto e branco, promove uma melhor compreensão dos conceitos apresentados. Os estudantes tendem a confiar mais em informações associadas a imagens, o que pode ajudar a dar mais significado ao que eles veem (HATTWIG et al., 2013).

Outro fator de relevância para o ensino e a aprendizagem da estatística, está relacionado a análise de dados brutos de forma independente e de maneira a apresentar informações apropriadas, bem fundamentadas e de forma crítica (HOLLAS; BERNARDI, 2018; CAMPOS; WODEWOTZKI, 2018). No entanto, como destacam Petocz e Reid (2005), Gordon e Nicholas (2010) e März e Kelchtermans (2013), de modo geral, estudantes dos cursos de serviços de estatística (aqueles

dirigidos para estudantes que irão se formar em qualquer outra área), aprendem, em muitas ocasiões, conceitos, técnicas matemáticas e estatísticas descontextualizadas, e informações irrelevantes a serem dominadas em curto prazo de tempo.

Por outro lado, dados brutos relacionados a dinâmica populacional humana estão disponíveis, mas fragmentados, geralmente dispostos em diferentes plataformas com acessos indiretos e em muitas situações em formatos variados. (LEBLANC, 2012).

No entanto, a associação entre banco de dados e sistemas de visualização pode apresentar diversas vantagens para o ensino e aprendizagem de diversas técnicas estatísticas, pois são geralmente gratuitas, disponíveis *on-line* e podem apresentar informações complexas em um formato visual fácil de interpretar (FORBES, 2012).

## 2. O ensino de Correlação Linear Simples

Os cursos introdutórios de estatística abordam, em suas primeiras aulas, as chamadas análises univariadas, ou seja, conceitos e técnicas aplicados a uma única variável, como a descrição do sexo de indivíduos em uma pesquisa de mercado ou eleitoral, a distribuição da renda de sujeitos de uma cidade, a temperatura média durante o ano e outras análises.

Em muitas situações na vida real, é comum estudarmos o comportamento de duas variáveis diferentes, com o objetivo de verificar se existe alguma relação entre elas, isto é, se mudanças em uma delas influenciam os valores da outra. Se isso ocorrer, dizemos que as variáveis estão correlacionadas ou que existe uma correlação entre elas. Nesses casos, o estudo é dito bivariado, pois estamos interessados na relação entre duas variáveis, e a correlação é denominada de simples.

A análise de correlação visa estudar a relação linear entre duas variáveis quantitativas. De maneira geral, no campo da Estatística, faz-se a distinção entre correlação e associação; a denominação associação é utilizada para estudos de relações entre variáveis qualitativas (MIRANDA, 2008).

Um dos conteúdos estatísticos de importância, tanto por seu significado prático como pelo seu caráter instrumental para outras técnicas estatísticas é a associação estatística e sua extensão para o conceito de correlação (HILDEBRAND; LANG; ROSENTHAL, 1977, apud ESTEPA; SÁNCHEZ-COBO, 2003).

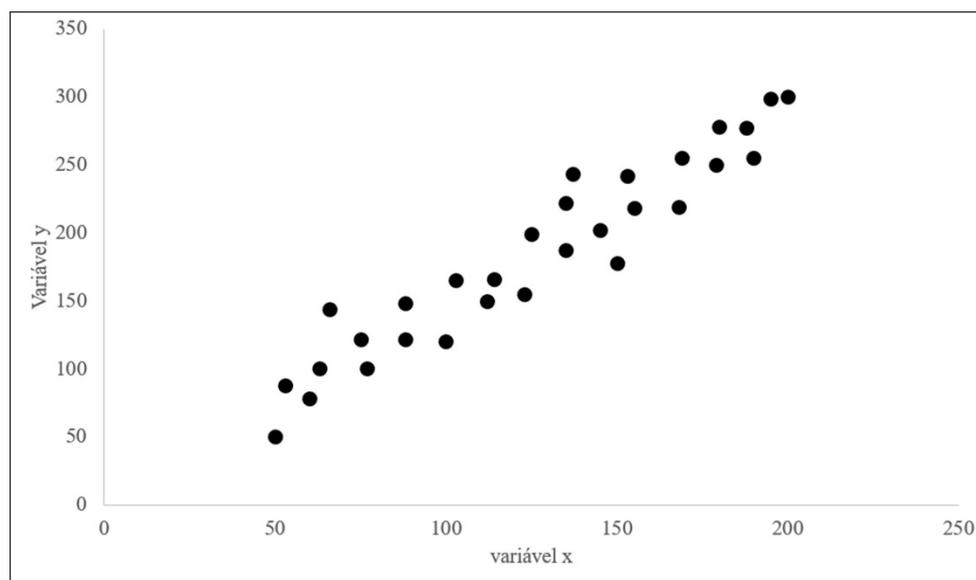
A Correlação Linear Simples está presente no rol dos conteúdos programáticos de Estatística dos mais variados cursos, sejam da área de Humanas, Biológicas ou de Exatas, o que confere grande importância a este tema.

A Correlação Linear pode ser avaliada visualmente por meio de diagramas e numericamente com o uso de coeficientes como o coeficiente de correlação de Pearson, denotado pela letra minúscula  $r$ , o qual permite verificar a força e o sentido da possível correlação entre as variáveis.

A correlação é uma medida de associação monotônica entre duas variáveis. Uma relação monótona entre duas variáveis é aquela em que, quando o valor de uma variável aumenta, o mesmo ocorre com o valor da outra variável; ou quando o valor de uma variável aumenta, o valor da outra variável diminui. Em dados correlatos, portanto, a mudança na magnitude de uma variável está associada a uma mudança na magnitude de outra variável, seja na mesma direção ou na direção oposta. No entanto, a existência de correlação não implica necessariamente em uma relação de causa-efeito (KOZAK, 2008; SCHOBER; BOER; SCHWARTE, 2018).

Como destacado anteriormente, diagramas de dispersão são utilizados na análise de correlação e de modo frequente a primeira forma de análise apresentada aos alunos com o objetivo de realizar estudos para verificar possíveis relações entre variáveis (figura 1).

Figura 1 - Diagrama de dispersão – Variáveis  $x$  e  $y$ . Dados fictícios.



Fonte: Elaboração do autor (2019).

Desta forma, a apresentação visual das possíveis relações entre variáveis pode contribuir para a aprendizagem dos conceitos e aplicações da Correlação Linear Simples e o uso de sistemas dinâmicos e interativos são importantes ferramentas didáticas, entre as quais o Sistema Gapminder.

### 3. Sistema Gapminder

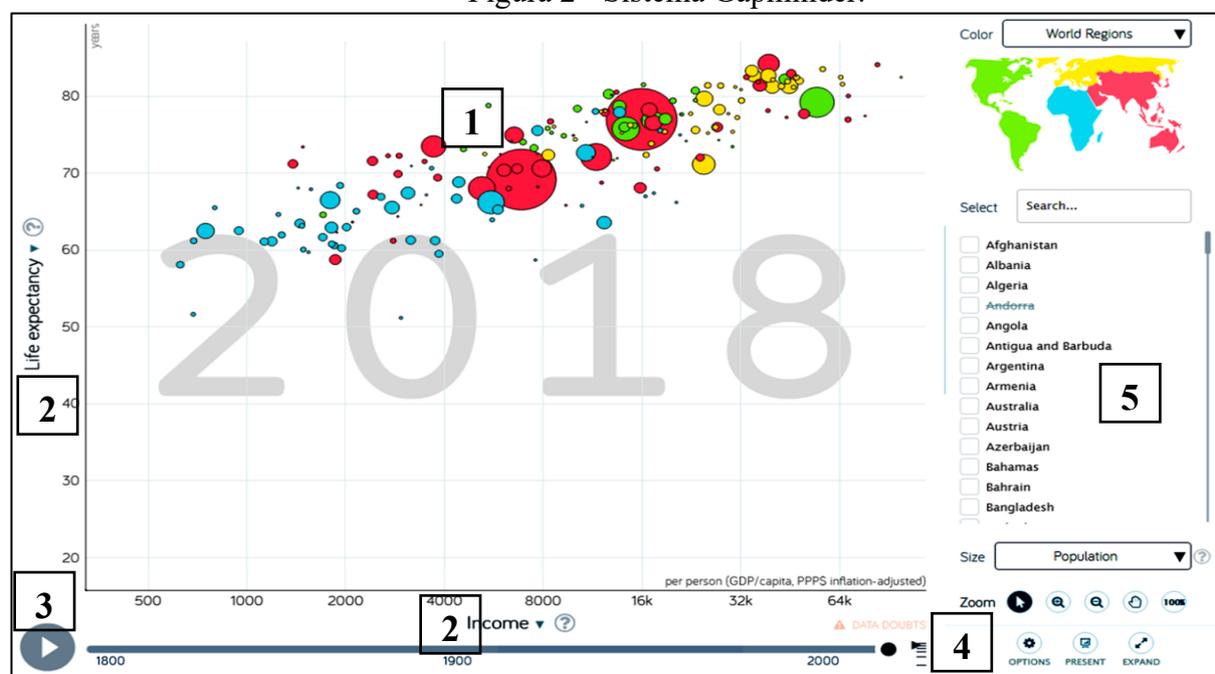
A Fundação Gapminder é responsável por unificar diversos bancos de dados das Nações Unidas, universidades e organizações não-governamentais em um único banco de dados com um formato consistente. A base de dados do Gapminder World inclui mais de 600 variáveis, com dados

de 258 países e séries temporais para a maioria das variáveis e países. Os dados remontam a década de 1960, mas para algumas variáveis estão disponíveis desde o século XIX. O banco de dados é ativo e está em constante desenvolvimento.

O banco de dados Gapminder é integrado com um programa computacional de análise de dados que facilita a identificação de correlações entre variáveis da população humana, seus padrões espaciais e tendências ao longo do tempo. Esse sistema oferece animações interativas das estatísticas o que o torna uma ferramenta didática poderosa no estudo das relações entre variáveis. O acesso ao *software* é livre e é dado pelo endereço na internet: [https://www.gapminder.org/tools/#\\$chart-type=bubbles](https://www.gapminder.org/tools/#$chart-type=bubbles)

A representação visual é apresentada por meio de um diagrama de dispersão que exhibe cinco variáveis - X, Y, tamanho do círculo, cor do círculo e ano (figura 2).

Figura 2 - Sistema Gapminder.



Fonte: Sistema Gapminder (2019).

Os principais comandos do programa Gapminder estão listados a seguir e os números dentro dos retângulos (figura 2) não fazem parte do programa, servem como legenda desenvolvido pelo autor deste trabalho:

1 - Ao repousar o *mouse* sobre qualquer círculo, o sistema revela o nome do país.

2 - Local em que o usuário pode selecionar com o *mouse* variáveis de interesse – Válido tanto para o eixo x quanto para o eixo y.

3 - Botão *play* que propicia ao usuário observar a animação do desenvolvimento dos países ao longo do tempo.

4 - Barra para reduzir ou ampliar a velocidade da animação.

5 - Lista de países que podem ser selecionados o que proporciona a visualização desses países.

Cada círculo colorido no diagrama de dispersão representa um país e as cores são referentes aos continentes no canto superior direito. O conjunto de países pode indicar a correlação entre as variáveis  $x$  e  $y$  para um determinado ano. O tamanho do círculo é proporcional ao tamanho populacional dos países para os respectivos anos e é possível realizar animação dos resultados ao longo do tempo para apresentar possíveis padrões entre as variáveis escolhidas.

O programa computacional *online* propicia ao usuário a escolha de um conjunto abrangente de variáveis nas mais diversas áreas, entre as quais: saúde, educação, economia, trabalho, setor público, infraestrutura, entre outras.

Cada variável é descrita de forma sucinta, mas documentação detalhada pode ser acessada de forma rápida e fácil no programa.

#### **4. Objetivo**

O interesse principal da atividade utilizando o Gapminder, foi motivar os alunos na aprendizagem de Correlação Linear Simples, fornecendo-lhes dados reais em um formato moderno e interativo. Desta forma, o objetivo da reflexão sobre a prática desenvolvida é apresentar como o sistema Gapminder pode ser utilizado como suporte para os processos de ensino e aprendizagem dos conceitos básicos da Análise de Correlação Simples para estudantes de cursos do Ensino Superior.

#### **5. Método**

##### **5.1 Sujeitos do Estudo**

As atividades utilizando o sistema Gapminder foram desenvolvidas com todos os 41 estudantes de graduação em Ciências Biológicas e todos os 39 de Engenharia Biotecnológica em uma Faculdade estadual de São Paulo, Brasil. Todos os alunos envolvidos na prática autorizaram a divulgação dos resultados para fins de pesquisa e sua publicação. A carga horária para as atividades propostas foi de até 04 horas, para cada um dos cursos em laboratório de informática, com 50 computadores conectados à internet.

##### **5.2 Aula tradicional**

Para cada uma das turmas envolvidas neste estudo, foi ministrada uma aula introdutória clássica, com 2 horas envolvendo os conceitos básicos e introdutórios de Correlação Linear Simples. Nesta aula, foram apresentados alguns exemplos usuais de correlação, como a relação em peso

corporal e estatura de sujeitos de um dado curso universitário.

A aula teve como objetivo introduzir os principais conceitos de correlação e sua mensuração, por meio do coeficiente de Pearson sem, no entanto, apresentar qualquer novo tipo de método ou instrumento que pudesse descaracterizar as aulas clássicas. Assim, o novo instrumento para promover o aprendizado da análise estatística em pauta, foi o sistema Gapminder descrito anteriormente e aplicado em atividade subsequente.

### 5.3 Vídeo tutorial

Como parte das atividades, inicialmente foi disponibilizado para os estudantes um vídeo com quatro minutos com tutorial das principais funções do *software* Gapminder. Posteriormente as ações didáticas foram desenvolvidas em grupos, no referido laboratório de informática.

O método de pesquisa utilizado foi a pesquisa exploratória com enfoque qualitativo (GIL, 2008). O conjunto de dados coletados para esse estudo é composto por questões abertas e fechadas dispostas em questionário *on-line* (anexo A). As questões visam avaliar o desempenho dos estudantes, tendo o sistema Gapminder como instrumento de suporte para complementar de forma interativa e atraente os conceitos de Correlação Linear Simples, bem como a descrição da transcrição das respostas permite um melhor entendimento das dificuldades e compreensão dos conceitos abordados. Nesse sentido, para garantir o anonimato dos participantes do estudo, os sujeitos foram codificados em CB1, CB2, ... CB41 para alunos do curso de Ciências Biológicas e EB1, EB2,..., EB39, para alunos do curso de Engenharia Biotecnológica.

### 5.4 Atividades

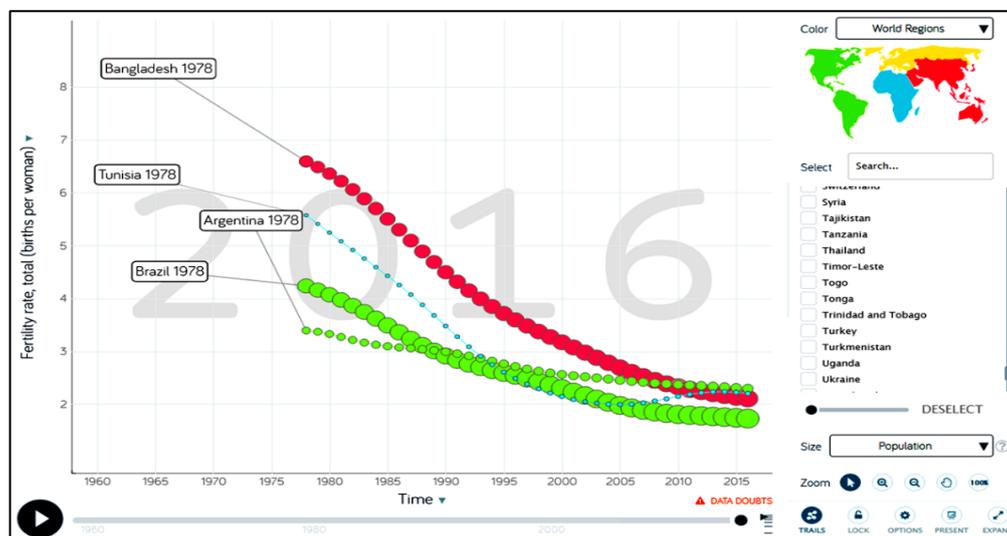
As atividades são caracterizadas pelo uso do sistema Gapminder para a resolução de sete questões propostas aos alunos (anexo A).

A primeira questão busca introduzir o estudante no ambiente Gapminder, sem a preocupação de apresentar conceitos de relação entre variáveis. É um momento para selecionar variáveis e países e observar os resultados finais das séries temporais (figura 3).

Esta questão apresenta na abcissa a variável “Ano” e na ordenada a variável “Número de filhos por mulheres”. De forma diferente, as questões número dois e três têm como objetivo introduzir variáveis que podem ser correlacionadas. No entanto, o número de países envolvidos na análise ainda deve ser considerado insuficiente para desenvolver a Correlação Linear. A questão dois aborda a relação entre as variáveis “Número de nascimentos por mulheres” e “Mortalidade infantil” durante o período de 1960 a 2015 para um único país, o Chile. Nesta situação, é possível verificar uma tendência

de queda na mortalidade infantil à medida que ocorre a diminuição do número de nascimento por mulheres. No entanto, a situação proposta avalia tão somente o país Chile.

Figura 3 - Resultado da questão 1 – Sistema Gapminder.



Fonte: Sistema Gapminder (2019).

É possível observar pelo sistema Gapminder uma relação ao longo da série. Assim, com o decorrer dos anos, o número de filhos por mulheres é reduzido e também o número da mortalidade infantil. A terceira questão apresenta a relação entre a “Escolaridade feminina” e “Mortalidade infantil” para os países Argentina, Brasil e México entre os anos de 1980 a 2011. Neste caso, é possível verificar uma relação negativa entre as variáveis, ou seja, aumenta-se a escolaridade feminina e reduz-se a mortalidade infantil, mas assim como na segunda questão, o número de países é limitado e as conclusões devem ser cautelosas.

A quarta e quinta questões introduzem a Correlação Linear Positiva, mas também o conceito de força da correlação. As variáveis envolvidas nesta análise são “Renda per capita” e “Expectativa de vida”. Neste caso, todos os países com informações estão envolvidos, e o sistema Gapminder apresenta dados dos séculos XIX, XX e XXI, o que proporciona uma experiência enriquecedora e estimulante, pois a visualização das imagens em movimento e interativa dos eventos, remete a diversos fatos históricos, como a primeira e segunda grandes guerras mundiais, o desenvolvimento de países da Ásia, em meados da década de 1970 e o evento da AIDS nos países africanos, por exemplo. No entanto, a questão cinco especificamente, aborda a situação da correlação dessas duas variáveis para todos os países em dois anos diferentes, 1954 e 1992. No ano de 1954, é possível observar uma relação linear positiva, ou seja, à medida que cresce a renda per capita, cresce também a expectativa de vida, mas essa relação é mais forte no ano de 1992. Assim, a atividade proporciona

uma comparação entre a dispersão dos dados por meio dos diagramas (figuras 4 e 5).

Figura 4 - Renda versus Expectativa de Vida – ano de 1954 – Sistema Gapminder.



Fonte: Sistema Gapminder (2019).

Figura 5 - Renda versus Expectativa de Vida – ano de 1992 – Sistema Gapminder.

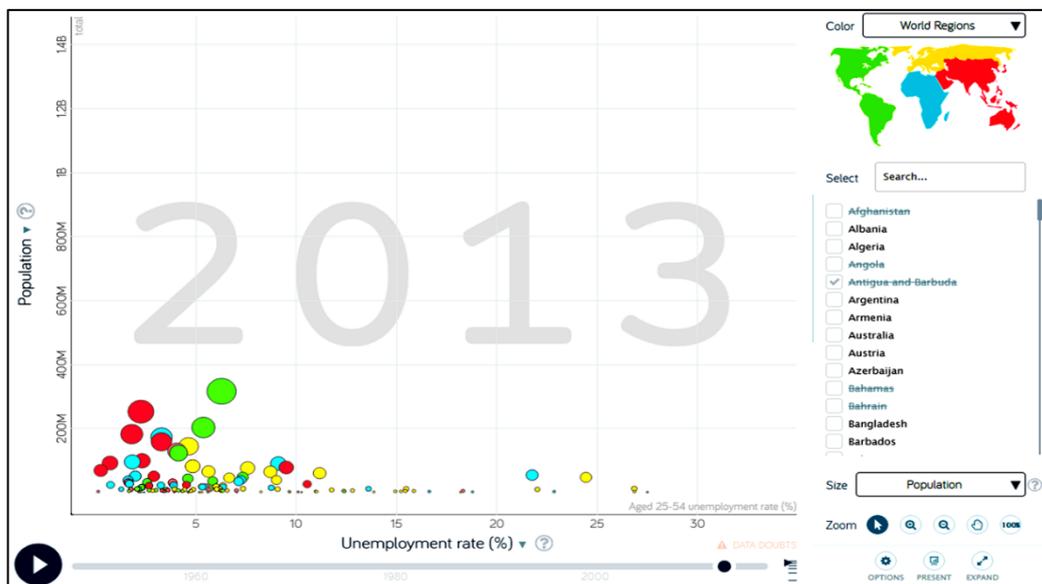


Fonte: Sistema Gapminder (2019).

Ainda nesta atividade – questão 5, foi solicitado ao aluno uma estimativa aproximada dos possíveis valores dos coeficientes de correlação de Pearson para os anos de 1954 e 1992 (figuras 4 e 5). Finalmente a sexta e sétima questões, envolvendo as variáveis “Taxa de desemprego” e “Tamanho da População” para todos os países, tinham como objetivo fomentar a discussão sobre a ausência de correlação. Também nesta atividade, os alunos são convidados a apresentar uma estimativa aproximada do possível valor do coeficiente de correlação de Pearson, sem a realização de cálculos

(figura 6).

Figura 6 - Taxa de desemprego versus Tamanho da População – Sistema Gapminder.



Fonte: Sistema Gapminder (2019).

## 6. Resultados e Discussão

As respostas dos alunos foram avaliadas e mensuradas em uma escala de zero a dez.

**Tabela 1** - Resultados gerais das respostas dos alunos.

Curso	Média	Desvio Padrão	n
Engenharia Biotecnológica	7,80	1,62	39
Ciências Biológicas	7,85	1,64	41

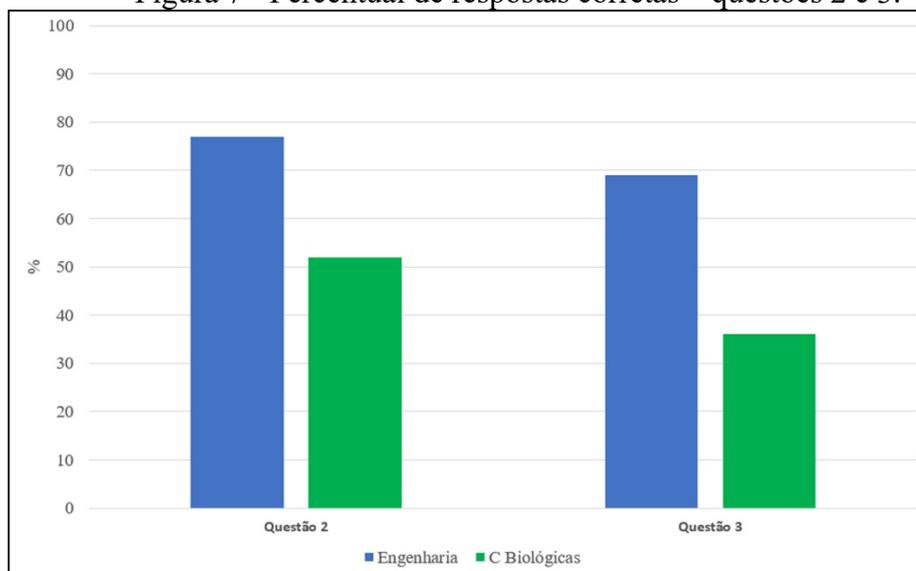
Fonte: Elaboração do autor (2019).

Os resultados da tabela 1 são positivos pois apresentam notas médias próximas de oito e baixa variabilidade, o que confere às duas turmas um bom desempenho nas avaliações propostas. Outro resultado que merece destaque, é que de todos os 80 alunos, dois não obtiveram avaliação igual ou superior a cinco, em uma pontuação mínima de zero e máxima de dez. Por outro lado, 13 alunos obtiveram notas máximas, cinco do curso de Ciências Biológicas e oito de Engenharia Biotecnológica.

Para as questões dois e três, denominadas de fechadas com alternativas, foi possível verificar que o maior percentual de acerto para a questão 2, é proveniente da turma de Engenharia Biotecnológica e que para a questão 3, menos da metade dos alunos de Ciências Biológicas não

escolheram a alternativa correta (figura 7.) Muitos estudantes não levaram em consideração o tamanho amostral em suas avaliações. Esse resultado pode servir para que o docente busque em outras atividades ou aulas posteriores, sanar essa questão que surge de forma tão explícita na atividade envolvendo o Gapminder.

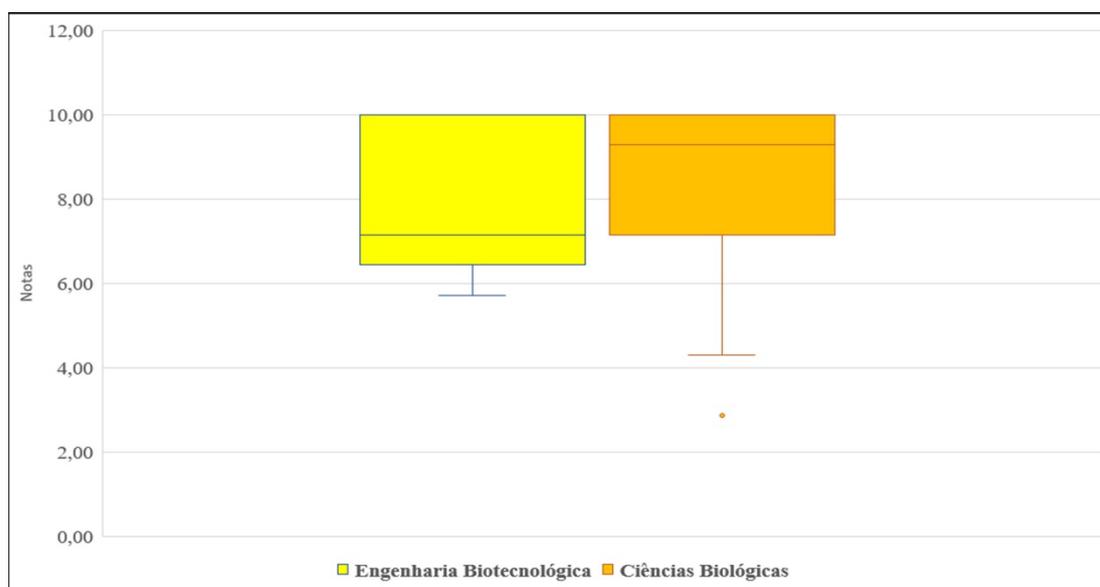
Figura 7 - Percentual de respostas corretas – questões 2 e 3.



Fonte: Elaboração do autor (2019).

Para as questões quatro, cinco, seis e sete, denominadas de questões abertas em que o aluno necessita dissertar sua resposta, as notas finais foram positivas como apresentadas no gráfico box-plot (figura 8).

Figura 8 - Avaliações de questões abertas – Cursos de Eng. Biotec. e Ciências Biológicas.



Fonte: Elaboração do autor (2019).

Os traços das caixas indicam a mediana das notas em cada curso, sendo o curso de Ciências Biológicas, com dois alunos abaixo da pontuação cinco. No entanto, apresenta 15 alunos com pontuação máxima, dois a mais que o curso de Engenharia Biotecnológica. Para essas questões, os diagramas foram corretamente interpretados pelos alunos e apresentam um alto acerto, no entanto as estimativas livres dos coeficientes de correlação (questões 5 e 7), são pontos que devem ser tratados em outras atividades ou aulas, visto que muitos dos estudantes apresentaram dificuldade em relacionar o diagrama a valores aproximados dos coeficientes.

Os resultados sugerem que os estudantes foram capazes de utilizar o sistema Gapminder para auxiliar a aprendizagem do conceito de Correlação Linear utilizando estruturadas visuais. Esse aspecto também foi observado por Prodromou (2014) em trabalho que utilizou o mesmo sistema Gapminder e que propiciou o desenvolvimento de modelos mentais na análise de possíveis relações entre múltiplas variáveis. Neste sentido, sistemas tecnológicos são importantes ferramentas para o apoio de estudantes com o objetivo de facilitar a manipulação de dados, bem como sua representação (PRODROMOU, 2013). Com o advento da Internet, diversos sistemas para análise de dados têm se popularizado entre os quais Fathom e TinkerPlots (RUBIN; HAMMERMAN; KONOLD, 2006), GeoVista (FORBES, 2012), sistemas como PopulationPyramid.net que apresentam um conjunto extenso de indicadores, por meio de mapas e visualizações, como por exemplo pirâmides etárias, taxas de leitos hospitalares e outros (<https://www.populationpyramid.net/>). De forma mais indireta, podemos destacar outros tipos de visualizações interativas utilizando os chamados *Dashboard*, representações visuais de dados que podem incluir gráficos, mapas, texto, cores, hiperlinks, sempre em tempo real (SUGANYA et al., 2018). Os chamados *Dashboard* podem ser construídos em linguagens computacionais, mas também são popularizados em planilhas eletrônicas.

Algumas das respostas dos estudantes, destacadas a seguir, refletem o entendimento e dificuldades dos conceitos após a atividade com o sistema Gapminder.

Questão quatro – Variáveis envolvidas: “Renda” e “Expectativa de vida”. Existe alguma relação entre as duas variáveis analisadas ao longo do tempo? Faça uma breve análise sobre o diagrama ao longo do tempo.

Respostas selecionadas consideradas adequadas

EG5: *Existe sim uma relação, pois conforme a renda per capita aumenta, aumenta a expectativa de vida, tendo em vista que uma renda mais alta proporciona maior acesso a medicamentos, serviços de saúde e, portanto, melhores condições de vida. Com o passar dos anos, houve aumento significativo em ambas variáveis, proporcionando uma correlação positiva, que confirma essa relação.* (Questionário aplicado em 19/11/2018)

CB9: *Existe uma relação aproximadamente linear positiva entre as duas variáveis, quanto maior a renda, maior a expectativa de vida. Pode-se observar como contraponto que no período das duas Grandes Guerras, esse valor não segue o padrão, devido a alta mortalidade.*

(Questionário aplicado em 19/11/2018)

EG18: *Embora apresente algumas oscilações pontuais, a tendência é que se forme uma correlação linear positiva entre as variáveis, visto que a expectativa de vida aumentou de acordo com o aumento da renda de cada país, de um modo geral.*

(Questionário aplicado em 19/11/2018)

Respostas consideradas inadequadas ou incompletas

CB22: *Sim, a expectativa de vida tem total relação com o rendimento de um país. Grifo em negrito realizado pelo professor.*

(Questionário aplicado em 19/11/2018)

Ao analisar as respostas para a questão em pauta, verifica-se quatro respostas incompletas e com equívocos de conceitos, como a apresentada anteriormente. Outro ponto a ser destacado, mas também em menor número, é a falta de clareza sobre a caracterização do sentido da Correlação Linear, ou seja, algumas poucas respostas não apresentam indicação do conceito de correlação positiva.

Questão cinco – Variáveis envolvidas: “Renda” e “Expectativa de vida”. Observe os anos de 1954 e 1992. Para esses dois anos, quais seriam os valores aproximados dos coeficientes de correlação de Pearson  $r$ ? Não faça cálculos. Explique sua resposta.

Respostas consideradas adequadas

EG12: *No ano de 1954 o  $r$  aproximado é de 0,7, podemos observar que os países seguem uma tendência, porém a correlação não é tão forte. Já no ano de 1992 podemos observar um  $r$  mais próximo a 0,85, onde a correlação entre os países é mais forte e tendem a formar uma linha positiva crescente.*

(Questionário aplicado em 19/11/2018)

CB32: *Para o ano de 1954, o  $r$  utilizado seria de aproximadamente 0,7, pois se fosse traçado uma reta positiva os países se encontrariam mais dispersos a ela. Já para o ano de 1992, o  $r$  seria de aproximadamente 0,8, pois os países formam uma reta mais "perfeita" que a anterior, sem tanta dispersão.*

(Questionário aplicado em 19/11/2018)

Respostas consideradas inadequadas ou incompletas

EB15: *Ambos os valores de  $r$  se aproximam de 1, apresentando uma correlação linear positiva.*

(Questionário aplicado em 19/11/2018)

EB13: *Sendo o coeficiente positivo  $0 < r < 1$ , pois as variáveis aumentaram nos dois intervalos de anos.*

(Questionário aplicado em 19/11/2018)

Para esta situação, 52% dos alunos apresentaram algum tipo de deficiência na associação da disposição dos pontos no diagrama e valores aproximados do coeficiente de correlação de Pearson.

Fica evidente que para esta parcela de estudantes, ainda não está consolidado a associação entre as possíveis estruturas de dispersões de pontos e os valores do coeficiente de Pearson. Esses resultados foram úteis para que o professor pudesse realizar uma intervenção específica. Assim o uso do sistema Gapminder proporcionou, não só aos alunos a oportunidade de uma aprendizagem mais dinâmica sobre a temática proposta, mas também a possibilidade de que o professor possa reduzir a distância entre a compreensão e o desempenho do aluno (HATTIE; TIMPERLAY, 2007).

Alguns estudos apresentam equívocos diferentes na interpretação do coeficiente de correlação de Pearson, entre os principais o uso do coeficiente de correlação e a proporção da variância de  $y$  explicada por  $x$  e a realização de inferência de causa e efeito baseada nos valores do coeficiente (ONWUEGBUZIE; DANIEL, 2002; ESTEPA; SÁNCHEZ-COBO, 2003).

As questões seis e sete que abordam a situação entre as variáveis “Taxa de desemprego” e “Tamanho da População” para todos os países (figura 6), apresentam respostas corretas em ambos os cursos com algumas imprecisões na avaliação dos coeficientes de correlação de Pearson.

Vale a pena sinalizar que neste estudo foram propostas atividades envolvendo algumas variáveis. No entanto, o sistema Gapminder propicia um conjunto de indicadores diversos o que favorece a docentes e alunos uma ampla gama de análises e estudos.

## 7. Considerações Finais

Os resultados obtidos neste estudo, sugerem que os estudantes foram capazes de usar adaptativamente o sistema Gapminder, para construir estruturas visuais que destacam informações relevantes para a aprendizagem de conceitos introdutórios de Correlação Linear Simples. Utilizando as visualizações interativas, os alunos analisaram possíveis correlações ao longo do tempo em conjuntos reais de dados.

Como apresentado anteriormente, o tamanho da amostra nas atividades foi desconsiderado pelos estudantes, o que pode trazer interpretações incompletas ou equivocadas. Outro ponto a ser destacado pelas respostas das questões 5 e 7, diz respeito a dificuldade de associar valores aproximados do coeficiente de Pearson a estrutura dos diagramas fornecidos pelo sistema Gapminder. Desta forma, o sistema Gapminder associado ao questionário permitiu também avaliar de forma eficaz possíveis deficiências e dificuldades dos estudantes, o que pode direcionar novas atividades para sanar dúvidas relacionados aos conceitos abordados.

Como destaca Forbes et al. (2014), a maneira como as informações são apresentadas, às vezes é tão importante quanto a forma como os dados foram coletados ou analisados. Deste modo, o uso de sistemas interativos utilizando tecnologias digitais representa um avanço na aprendizagem da

Estatística e assim, novas formas de visualização de dados podem impactar positivamente na didática de professores que ensinam estatística. Portanto, explorações que antes exigiam habilidades técnicas consideráveis podem agora ser realizadas facilmente por usuários iniciantes em estatística (SUTHERLAND; RIDGWAY, 2017).

Desta forma, o sistema Gapminder pode ser um suporte ao ensino da Análise de Correlação para o ensino superior, posto que permite a escolha de variáveis, de períodos de tempo, regiões geográficas e outros fatores para análises. A visualização em formato de vídeo, também traz um componente motivacional almejado pelos docentes.

Conclui-se que é fundamental incorporar formas mais atraentes de apresentar e contextualizar as atividades práticas de cursos de Estatística para estudantes universitários, de forma que possam ser encorajados a pesquisar e discutir questões sociais, econômicas e ecológicas, para que desenvolvam seus conhecimentos tornando-se cidadãos críticos e responsáveis.

## 8. Referências

- CAMPOS, S. G. V. B.; WODEWOTZKI, M. L. L. Educação Estatística e Desenvolvimento do Sentido de Número: Uma Inter-Relação Possível. **Revista de Ensino de Ciências e Matemática**, v.9, n.2, p. 88-106, 2018.
- ESTEPA, A.; SÁNCHEZ-COBO, A. F. T. Evaluación de la comprensión de la correlación y regresión a partir de la resolución de problemas. **Statistics Education Research Journal**, v. 2, n. 1, p. 54-68, 2003.
- FERNANDES, R. J. G.; DAMIN, W.; SANTOS JÚNIOR, G. Atividades de estatística para o 8º ano do ensino fundamental: algumas considerações. **R. Bras. Ens. Ci. Tecnol.**, Ponta Grossa, v. 10, n. 1, p. 1-13, 2017.
- FORBES, S. D. Data Visualization: A Motivational and Teaching Tool in Official Statistics. **Technology Innovations in Statistics Education**, v. 6, n. 1, p. 1-19, 2012.
- FORBES, S.; CHAPMAN J.; HARRAWAY, J.; STIRLING, D.; WILD, C. Use of Data Visualisation in the Teaching of Statistics: A New Zealand Perspective. **Statistics Education Research Journal**, v. 13, n. 2, p.187-201, 2014.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. 6.ed. São Paulo: Atlas, 2008.
- GORDON, S.; NICHOLAS, J. Teaching with examples and statistical literacy: views from teachers in statistics service courses. **International Journal of Innovation in Science and Mathematics Education**, v. 18, n. 1, p. 14-25, 2010.
- HATTIE, J.; TIMPERLAY, H. The power of feedback. **Review of Educational Research**, v. 77, n. 1, p. 81-112, 2007.

HATTWIG D.; BUSSERT K.; MEDAILLE A.; BURGESS J. Visual Literacy Standards in Higher Education: New Opportunities for Libraries and Student Learning. **Libraries and the Academy**, v. 13, n. 1, p. 61-89, 2013.

HILDEBRAND, D. K.; LANG, J. D.; ROSENTHAL, H. **Analysis of Nominal Data**. Londres: Sage University Paper, 1997.

HOLLAS, J.; BERNARDI, L. Educação Estatística Crítica: Um Olhar Sobre os Processos Educativos. **Revista de Ensino de Ciências e Matemática**. v.9, n.2, p. 72-87, 2018.

JAMIE, D. M. Using Computer Simulation Methods to Teach Statistics: A Review of the Literature. **Journal of Statistics Education**, v. 10, n. 1, p. 1-20, 2002.

KLEINMAN, E. B.; DWYER, F. M. Analysis of computerized visual skills: Relationships to intellectual skills and achievement. **International Journal of Instructional Media**, v. 26, n.1, p. 53-69, 1999.

KOZAK, M. Correlation and regression: similar or different concepts? **Statistics in transition - new series**, v. 9, n. 1, p. 159-162, 2008.

LEBLANC, D. Gapminder: Using a World of Human Ecology Data to Teach Students Critical Thinking Skills. **Bulletin of the Ecological Society of America**, v. 94, p. 358-372, 2012.

MÄRZ, V.; KELCHTERMANS, G. Sense-making and structure in teachers' reception of educational reform. A case study on statistics in the mathematics curriculum. **Teaching and Teacher Education**, v. 29, p. 13-24, 2013.

MIRANDA, M. L. **Correlação e regressão em curso de engenharia: uma abordagem com foco na leitura e interpretação de dados**. 2008. 159 f. Dissertação (Mestrado) – Pontifícia Universidade Católica de Minas Gerais, Programa de Pós-Graduação em Ensino de Ciências e Matemática, Belo Horizonte, MG 2008.

ONWUEGBUZIE, A. J.; DANIEL, L. G. Uses and misuses of the correlation coefficient. **Research in the Schools**, v. 9, p. 73-90, 2002.

PETOCZ, P.; REID, A. Something strange and useless: service students' conceptions of statistics, learning statistics and using statistics in their future profession. **International Journal of Mathematical Education in Science and Technology**, v. 36, n. 7, p. 789-800, 2005.

PRODROMOU, T. Drawing inference from data visualizations. **International Journal of Secondary Education**, v. 2, n. 4, 66-72, 2014.

\_\_\_\_\_. Data Visualization and Statistics from the Future. **Proceedings 59th ISI World Statistics Congress**, 25-30 August 2013, Hong Kong (Session IPS049). Disponível em <<http://2013.isiproceedings.org/Files/IPS049-P3-S.pdf>>. Acesso em: 05 de janeiro de 2019

RUBIN, A.; HAMMERMAN, J. K. L.; KONOLD, C. Exploring informal inference with interactive visualization software. **ICOTS-7**, Salvador, Ba, 2006. Acesso em: <[https://iase-web.org/documents/papers/icots7/2D3\\_RUBI.pdf](https://iase-web.org/documents/papers/icots7/2D3_RUBI.pdf)>. Acesso em: 04 de janeiro de 2019.

SCHOBER, P.; BOER, C.; SCHWARTE, L. A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth Analg.*, v. 126, n. 5, p. 1763-1768, 2018.

STOKES, S. Visual Literacy in Teaching and Learning: A Literature Perspective. *Electronic Journal for the Integration of Technology in Education*, v. 1, n. 1, p. 10-19, 2002.

SUGANYA, D.; KUMAR, K.; RAMESH, P.S. SUGANTHAN, C. Student performance dashboard using mining approach. *International Journal of Pure and Applied Mathematics*, v. 119, n. 12, p. 409-421, 2018.

SUTHERLAND, S.; RIDGWAY, J. Interactive visualizations and statistical literacy Statistics. *Education Research Journal*, v. 16, n. 1, p. 26 – 30, 2017

## Anexo A

Atividades/Questionário

Questão 1

**Na Argélia, as meninas costumavam se casar com um parente jovem. Na década de 1960, as mulheres argelinas começaram a se casar em uma idade cada vez maior. Isso aconteceu em parte porque se tornou a norma em todo o mundo árabe que um casal deveria ter sua própria casa. Isso significava que eles precisavam economizar por um longo tempo antes de poderem se casar. Outro fator foi o tempo que as meninas passaram na escola.**

**X - Time (aba principal); Y - More Data – wdi - Health – Reproductive Health - Fertility rate, total (births per woman).**

**Escolha os países Argentina, Bangladesh, Brasil e Tunísia. Avaliar o período: 1960 a 2015. Use DESELECT no lado direito para visualizar somente os países desejados.**

**Qual país tem o menor número de filhos por mulher no último ano de análise?**

Argentina

Tunísia

Brasil

Bangladesh

Brasil e Tunísia

Nenhuma das alternativas está correta.

Questão 2

**Analise a situação proposta para o país Chile.**

**X - More Data – wdi - Health – Reproductive Health - Fertility rate, total (births per woman);**

**Y - Child Mortality.**

**(Na aba de apresentação). Escolha o Chile. Avaliar o período: 1960 a 2015. Use DESELECT no lado direito para visualizar somente o Chile,**

**Qual relação é possível observar ao longo do tempo entre as duas variáveis avaliadas? Escolha a alternativa mais completa e que melhor expresse o resultado apresentado.**

Não existe relação entre as variáveis

É possível observar uma relação aproximadamente linear positiva entre as variáveis, mas deve-se tomar cuidado ao realizar essa afirmação visto que foi observado um único país;

É possível observar uma relação aproximadamente linear negativa entre as variáveis, mas deve-se tomar cuidado ao realizar essa afirmação visto que foi observado um único país;

É possível observar uma relação aproximadamente linear negativa entre as variáveis;

É possível observar uma relação aproximadamente linear positiva entre as variáveis;

Nenhuma das questões está correta.

Questão 3

**Analise a situação proposta para os países Brasil, México e Argentina**

**X – Education – Literacy - Literacy rate, adult female; Y – Health - Newborn & infants - Infant mortality. Avaliar o período: 1980 a 2015.**

**Use DESELECT no lado direito para visualizar somente Brasil, México e Argentina**

**Qual relação é possível observar ao longo do tempo entre as duas variáveis avaliadas? Escolha a alternativa mais completa e que melhor expresse o resultado apresentado.**

É possível observar uma relação aproximadamente linear negativa entre as variáveis, mas deve-se tomar cuidado ao realizar essa afirmação visto que foram observados poucos países;

Não existe relação entre as variáveis;

É possível observar uma relação aproximadamente linear positiva entre as variáveis, mas deve-se tomar cuidado ao realizar essa afirmação visto que foram observados poucos países;

É possível observar uma relação aproximadamente linear positiva entre as variáveis;

É possível observar uma relação aproximadamente linear negativa entre as variáveis;

Nenhuma das questões estão corretas.

Questão 4

**Income versus Life Expectante**

**X - Income (aba inicial); Y - Life Expectancy (aba principal). Para todos os países, Avaliar o período: 1920 a 2016.**

**Existe alguma relação entre as duas variáveis analisadas ao longo do tempo? Faça uma breve análise sobre o diagrama ao longo do tempo.**

Questão 5

**Para a análise anterior - X Income (aba inicial) YLife Expectancy (aba principal).**

**Para todos os países. Avaliar o período: 1920 a 2016 - observe os anos de 1954 e 1992. Para esses dois anos, quais seriam os valores aproximados dos coeficientes de correlação de Pearson r?**

**Não faça cálculos. Explique sua resposta.**

Questão 6

**% de desempregado versus população**

**Todos os países. Período: 1990 a 2017**

**X – Work - Unemployment rate - Age 25 - 54**

**Y – Population - Population**

**Existe alguma relação entre as duas variáveis analisadas ao longo do tempo? Faça uma breve análise sobre o diagrama ao longo do tempo - Inicie em 1990.**

Questão 7

**Para a questão anterior, qual seria o valor aproximado do coeficiente de correlação de Pearson r? Não faça cálculos. Explique sua resposta.**