**RELATÓRIO TÉCNICO ESTATÍSTICO**

Júlio Henri Maciel Bezerra da Silva

Estudante do 7° período de Bacharelado

em Estatística na UFPa

 linkedin.com/in/juliohenri/

**1. Introdução**

O presente relatório tem por objetivo apresentar, explorar e analisar o banco de dados, no qual foi obtido por um aluno de bacharelado em Odontologia por meio de um formulário, para fazer seu Trabalho de Conclusão de Curso sobre a relevância do cirurgião dentista em relação ao marketing no Instagram.

Primeiramente foi realizado uma Análise Fatorial nas variáveis do banco de dados para analisarmos de maneira mais profunda e ver o quão explicativo cada fator pode ser. Após isso, foi utilizado do teste de confiabilidade por meio do Alfa de Cronbach para confirmar a consistência do questionário respondido, e por fim, foi necessário fazer um Teste de Aderência para comparar a amostra com uma distribuição de probabilidade teórica.

**2. Materiais e Métodos**

A amostra foi obtida por meio de respostas de um formulário do Google Forms (em formato xlsx), no qual foram escolhidas 7 perguntas específicas que tem os cirurgiões dentistas como público alvo. Abaixo estão apresentadas as perguntas e no banco de dados foram substituídas por q1, q2, q3, q4, q5, q6 e q7, respectivamente:

1 - Quanto você acredita conhecer sobre o Instagram como ferramenta de marketing?

2- Quanto você acha necessário contratar empresas especializadas para gerenciar seu perfil?

3- O quanto você acredita ser importante postar fotos todos os dias?

4- O quanto você acredita ser importante postar stories?

5- O quanto você acredita ser importante o número de visualizações nos stories?

6- O quanto você acredita ser importante o número de curtidas nas postagens?

7- O quanto você tem conversão de seguidores em pacientes?

Além disso, foram pré-estabelecidas respostas qualitativas que variam de 1 a 5, onde cada variável significa, respectivamente, “Nada”, “Pouco”, ”Médio”, ”Bastante” e “Totalmente”. Foram obtidos 70 registros, sem valores nulos, com 7 variáveis categóricas. O carregamento, leitura, consulta, manipulação, geração de gráficos, tabelas e análise dos dados, foi feito com auxílio dos programas Python, BioEstat e Microsoft Excel.

A análise estatística será inicialmente realizada por meio de Análise Fatorial Exploratória de dados, objetivando fornecer uma visão geral do estudo e delinear possíveis características estruturais que possam ser avaliadas em maiores detalhes com o auxílio de testes estatísticos.

**3. Análise Fatorial Exploratória (AFE)**

A Análise Fatorial Exploratória é uma técnica estatística multivariada, isto é, avalia-se duas ou mais medidas e/ou variáveis sobre o objeto de interesse. O objetivo da AFE é estimar e identificar a dimensionalidade de indicadores e o menor número de traços latentes que explica o padrão das correlações (Osborne, 2014). No estudo em questão, essa ferramenta se fez necessária para que pudesse observar melhor como os fatores se relacionam com cada item.

Então, as variáveis foram redimensionadas em 2 fatores comuns por meio do *Common Factor Model* que observa a covariância entre os atributos que diz o grau de relacionamento de cada item do questionário para com cada fator, tais registros são chamados de cargas fatoriais. A decisão sobre o número de fatores a ser retido foi realizada, não só, por meio da técnica da Análise Paralela com permutação aleatória dos dados observados, como também pela checagem dos autovalores dos atributos, e a rotação utilizada foi a “*varimax*”.

Um dos pré-requisitos para escolha da quantidade de fatores é de que seus autovalores não sejam menores do que 1, e conforme análise dos autovalores visto abaixo, decidiu-se utilizar 2 fatores.

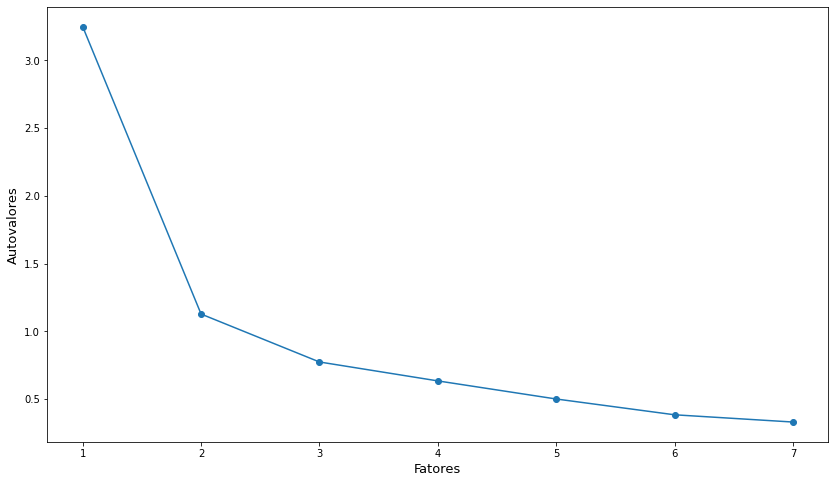
**Tabela 01:** Autovalores para cada atributo do banco de dados

|  |  |
| --- | --- |
| Atributo | Autovalor |
| q1 | 3,2494 |
| q2 | 1,1278 |
| q3 | 0,7739 |
| q4 | 0,6341 |
| q5 | 0,5005 |
| q6 | 0,3839 |
| q7 | 0,3304 |

Fonte: O autor

Também como forma de escolha dos fatores, foi realizado a Figura 01, onde a análise a ser feita é de que o número de fatores é menos recomendado a cada vez que a linha que liga um fator ao outro está mais horizontal, este gráfico é chamado *Scree Plot*. Logo, observando a figura também é recomendado um fator entre o número 2 ou 3.

**Figura 01:** Autovalores ordenados pelo número de fatores

Fonte: O autor

Feito a escolha da quantidade de fatores, o segundo passo é criar a matriz de cargas fatoriais estimadas, apresentada abaixo.

**Tabela 02:** Cargas fatoriais estimadas por atributo

|  |  |  |
| --- | --- | --- |
| Atributo | Fator 1 | Fator 2 |
| q1 | 0,1081 | 0,5674 |
| q2 | 0,5716 | 0,0866 |
| q3 | 0,3161 | 0,6676 |
| q4 | 0,4878 | 0,5698 |
| q5 | 0,7707 | 0,3475 |
| q6 | 0,6821 | 0,2625 |
| q7 | 0,1326 | 0,5193 |

Fonte: O autor

Analisando a Tabela 02, pode-se concluir que as variáveis Q5 e Q6 têm maior relevância para o Fator 1, e são atributos que remetem a perguntas sobre importância das visualizações e curtidas no Instagram. Já em relação ao Fator 2, Q3 e Q4 são as mais relevantes, com valor 0,6676 e 0,5698, respectivamente. Tais atributos são relacionados com as publicações no *Feed* e no *Story* da referida rede social.

**Tabela 03:** Variâncias percentuais explicadas por cada fator

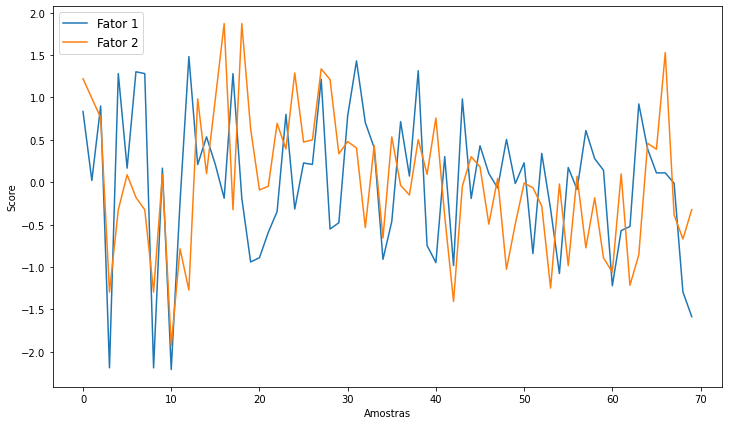
|  |  |  |  |
| --- | --- | --- | --- |
| Fator 1 | Fator 2 | Fator 3 | Fator 4 |
| 0,3888 | 0,3458 | 0,0555 | 0,0493 |

Fonte: O autor

A partir da análise da Tabela 03, pode-se concluir que o Fator 1 explica aproximadamente 39% dos dados e o Fator 2 explica aproximadamente 35%. Nota-se que após o Fator 2, os valores de variância percentual são muito pequenos, portanto, não sendo significantes para o estudo.

Abaixo, na Figura 02, é apresentado os Scores dos Fatores para cada registro no banco de dados, onde cada valor de Score é o grau de importância para cada entrevistado.

**Figura 02:** Scores do Fator 1 e Fator 1 para cada participante do estudo

Fonte: O autor

Em ambos os fatores há uma grande variabilidade nos Scores, sendo que para o Fator 2 há o pico máximo próximo ao valor 2 e mínimo próximo a -2 e para o Fator 1 o pico é aproximadamente 1,5 com o menor valor sendo menor que -2.

**4. Consistência Interna**

Consistência Interna refere-se a quão bem uma pesquisa, questionário ou teste realmente mede o que você deseja medir. Quanto maior a consistência interna, mais certeza você pode ter de que sua pesquisa é confiável, e a forma mais comum de medir a consistência interna é usando uma estatística conhecida como Alfa de Cronbach. A consistência interna nada mais é do que a extensão em que todos os itens de um teste medem a mesma variável latente.

**4.1 Alfa de Cronbach**

O Alfa de Cronbach calcula as correlações de pares entre os itens de uma pesquisa. O valor do Alpha de Cronbach pode variar entre infinito negativo e um. A tabela a seguir descreve como os vários valores do Alpha de Cronbach são normalmente interpretados:

**Tabela 04:** Interpretação do Alfa de Cronbach

|  |  |
| --- | --- |
| Alfa de Cronbach | Consistência Interna |
| 0,9 <= α | Excelente |
| 0,8 <= α < 0,9 | Bom |
| 0,7 <= α < 0,8 | Aceitável |
| 0,6 <= α < 0,7 | Questionável |
| 0,5 <= α < 0,6 | Ruim |
| α < 0,5 | Inaceitável |

Fonte: George, D & Mallery, P. SPSS for Windows step by step: A simple guide and reference. 4th ed. Boston: Allyn & Bacon. (2003).

O Alfa de Cronbach é útil pois ele fornece uma medida razoável de confiabilidade em um único teste. Dessa forma, não são necessárias repetições ou aplicações paralelas de um teste para a estimativa da consistência do mesmo e além disso, pode ser aplicado a questionários de múltipla escolha de escala dicotômicas ou policotômicas, tornando-o mais abrangente.

O software Python retornou o valor de Alfa de Cronbach igual a 0,8685. Conforme visto acima, esse valor é considerado “Bom”, isto é, o questionário repassado para a realização deste estudo é estatisticamente confiável.

**5. Teste de Aderência**

O Teste de Aderência também chamado de Teste de Ajustamento, é útil para verificar se os dados estão ajustados com determinada Lei, no caso do estudo em questão, deseja-se confirmar se os dados estão de acordo com seu valor esperado. Esse teste será utilizado pois só há uma amostra e essa amostra será comparada com uma distribuição de probabilidade teórica.

**5.1 Teste Qui-Quadrado**

O intuito do Teste Qui-Quadrado é fazer o paralelo entre as proporções de respostas de cada categoria (dados observados) com a distribuição teórica. Como são 5 itens por questão (1, 2, 3, 4 e 5), a lei da distribuição teórica será dada por 20%, pois essa é a probabilidade de cada categoria ser marcada pelo entrevistado.

Logo, pode-se afirmar hipóteses estatísticas, chamadas de hipótese alternativa e hipótese nula nas quais uma é a que deve ser testada e a outra a que o pesquisador espera provar que seja verdadeira, respectivamente. Dito isto, a hipótese nula criada é de que as respostas têm mesma chance de ocorrência, enquanto a alternativa diz que as respostas tem chance de ocorrência diferentes.

**Tabela 05:** Teste Qui-Quadrado para proporções esperadas iguais

|  |  |  |
| --- | --- | --- |
| Atributo | Estatística do Teste | P-valor |
| q1 | 34,439 | > 0,0001 |
| q2 | 53,431 | > 0.0001 |
| q3 | 38,729 | > 0.0001 |
| q4 | 36,637 | > 0,0001 |
| q5 | 33.99 | > 0,0001 |
| q6 | 58,347 | > 0,0001 |
| q7 | 54,037 | > 0,0001 |

Fonte: O autor

A partir dos resultados apresentados na Tabela 05, a hipótese nula de nenhum atributo é rejeitada (pois o p-valor é menor que 0,05), logo, é notório afirmar que as respostas do questionário têm mesma probabilidade de ocorrência para a população, ao nível de significância de 5%.

**Figura 03:** Dados observados em relação aos dados baseados na distribuição teórica para a primeira pergunta do questionário

Gráfico, Gráfico de barras

Descrição gerada automaticamente

**Figura 04:** Dados observados em relação aos dados baseados na distribuição teórica para a segunda pergunta do questionário

Gráfico, Gráfico de barras, Gráfico de cascata

Descrição gerada automaticamente

**Figura 05:** Dados observados em relação aos dados baseados na distribuição teórica para a terceira pergunta do questionário

Gráfico, Gráfico de barras

Descrição gerada automaticamente

**Figura 06:** Dados observados em relação aos dados baseados na distribuição teórica para a quarta pergunta do questionário

Gráfico, Gráfico de barras

Descrição gerada automaticamente

**Figura 07:** Dados observados em relação aos dados baseados na distribuição teórica para a quinta pergunta do questionário

Gráfico, Gráfico de barras

Descrição gerada automaticamente

**Figura 08:** Dados observados em relação aos dados baseados na distribuição teórica para a sexta pergunta do questionário

Gráfico, Gráfico de barras

Descrição gerada automaticamente

**Figura 09:** Dados observados em relação aos dados baseados na distribuição teórica para a sétima pergunta do questionário

Gráfico, Gráfico de barras

Descrição gerada automaticamente

**6. Teste de Independência**

Para a realização de testes de independência entre pares de variáveis foi realizado o agrupamento de categorias, considerando que em algumas categorias houve poucas respostas, dificultando a análise com todas as categorias pesquisadas. Desse modo foram agrupadas as seguintes categorias: 1 (Nada), 2 (Pouco) e 3 (Moderado) foram agregadas para formar a categoria 1 (insuficiente) e 4 (Muito) e 5 (Totalmente) foram agregadas para formar a categoria 2 (Suficiente). Será então realizado o teste de Fisher para avaliar a independência entre pares de variáveis, sendo o interesse investigar as hipóteses:

1) As variáveis q1 - Quanto você acredita conhecer sobre o Instagram como ferramenta de marketing e q2 - Necessidade de contratar empresa especializada para gerenciar o perfil são independentes. A Tabela 06 apresenta os resultados do cruzamento das informações mostrando os percentuais em relação à variável q1 e a Figura 10 mostra os quantitativos dos dados. Verifica-se que não existe muita diferença entre as categorias da variável q2 para cada categoria da variável q1 indicando que as variáveis não são relacionadas. Para comprovar foi realizado o teste de Fisher que comprovou a independência entre as variáveis (p-valor = 0,399).

**Figura 10:** Distribuição quantitativa das respostas cruzadas para as variáveis conhecimento do Instagram e Necessidade de contratar empresa

Gráfico, Gráfico de barras

Descrição gerada automaticamente

**Tabela 06:** Distribuição percentual das respostas cruzadas para as variáveis ‘Conhecimento do Instagram’ e ‘Necessidade de contratar empresa’

|  |  |  |  |
| --- | --- | --- | --- |
| Conhecimento do Instagram | Necessidade de contratar empresa | | Total |
| Pouca | Muita |
| Pouco | 31 (57.4%) | 23 (42.6%) | 54 |
| Muito | 7 (43.8%) | 9 (56.3%) | 16 |
| Total | 38 | 32 | 70 |

2) As variáveis q1-Quanto você acredita conhecer sobre o Instagram como ferramenta de marketing e q3-Importância de postar fotos diariamente no Instagram. A Tabela 07 apresenta os resultados do cruzamento das informações mostrando os percentuais em relação à variável q1 e a Figura 11 mostra os quantitativos dos dados. Verifica-se que existe diferença entre as categorias da variável q3 para cada categoria da variável V1 indicando que as variáveis são relacionadas. Para comprovar foi realizado o teste de Fisher que comprovou a dependência entre as variáveis ao nível de 5% (p-valor = 0,010). Observa-se que a diferença de opinião é maior entre as pessoas com muito conhecimento do Instagram, e pelo percentual na tabela (81,3%) pode-se considerar que as pessoas com muito conhecimento do Instagram dão mais importância à postagem diária de fotos do que as pessoas com pouco conhecimento.

**Figura 11:** Distribuição quantitativa das respostas cruzadas para as variáveis conhecimento do Instagram e Importância de postar fotos diariamente.

Gráfico, Gráfico de barras

Descrição gerada automaticamente

**Tabela 07**: Distribuição percentual das respostas cruzadas para as variáveis ‘Conhecimento do Instagram’ e ‘Importância de postar fotos diariamente’

|  |  |  |  |
| --- | --- | --- | --- |
| Conhecimento do Instagram | Importância de postar fotos diariamente | | Total |
| Pouca | Muita |
| Pouco | 31 (57.4%) | 23 (42.6%) | 54 |
| Muito | 3(18.8%) | 13(81.3%) | 16 |
| Total | 34 | 36 | 70 |

3) As variáveis q1-Quanto você acredita conhecer sobre o Instagram como ferramenta de marketing e q4-Importância de postar stories. A Tabela 08 apresenta os resultados do cruzamento das informações mostrando os percentuais em relação à variável q1 e a Figura 12 mostra os quantitativos dos dados. Verifica-se que existe diferença entre as categorias da variável q4 para cada categoria da variável q1 indicando que as variáveis são relacionadas. Para comprovar foi realizado o teste de Fisher que comprovou a dependência entre as variáveis ao nível de 5% (p-valor = 0,023). Observa-se que a diferença de opinião é maior entre as pessoas com muito conhecimento do Instagram, e pelo percentual na tabela (75%) pode-se considerar que as pessoas com muito conhecimento do Instagram dão mais importância à postagem de stories do que as pessoas com pouco conhecimento.

**Figura 12**: Distribuição quantitativa das respostas cruzadas para as variáveis ‘Conhecimento do Instagram’ e ‘Importância de postar stories’

Gráfico, Gráfico de barras

Descrição gerada automaticamente

**Tabela 08:** Distribuição percentual das respostas cruzadas para as variáveis conhecimento do Instagram e Importância de postar stories.

|  |  |  |  |
| --- | --- | --- | --- |
| Conhecimento do Instagram | Importância de postar fotos diariamente | | Total |
| Pouca | Muita |
| Pouco | 32 (59,6%) | 22 (40,7%) | 54 |
| Muito | 4(25,0%) | 12(75,0%) | 16 |
| Total | 36 | 34 | 70 |

4) As variáveis q1-Quanto você acredita conhecer sobre o Instagram como ferramenta de marketing e q5-Importância de curtidas nas postagens. A Tabela 09 apresenta os resultados do cruzamento das informações mostrando os percentuais em relação à variável q1 e a Figura 13 mostra os quantitativos dos dados. Verifica-se que, apesar de apresentar diferença entre as categorias da variável q5 para a categoria da variável q1, Pouco conhecimento, o teste de Fisher mostrou a independência entre as variáveis (p-valor = 0,565) evidenciando que, tanto para as pessoas com pouco ou para as que têm muito conhecimento do Instagram, a opinião não é estatisticamente diferente quanto à importância que é dada ao número de curtidas nas postagens.

**Figura 13:** Distribuição quantitativa das respostas cruzadas para as variáveis ‘Conhecimento do Instagram’ e ‘Importância de curtidas nas postagens’

Gráfico, Gráfico de barras

Descrição gerada automaticamente

**Tabela 09:** Distribuição percentual das respostas cruzadas para as variáveis ‘Conhecimento do Instagram’ e ‘Importância de curtidas nas postagens’

|  |  |  |  |
| --- | --- | --- | --- |
| Conhecimento do Instagram | Necessidade de contratar empresa | | Total |
| Pouca | Muita |
| Pouco | 33 (61,1%) | 21 (38,8%) | 54 |
| Muito | 8(50,0%) | 8(50,0%) | 16 |
| Total | 38 | 32 | 70 |