Análise de Serviços de Streaming

Projeto Semestral de Estatística



Integrantes

Luan TeixeiraR.A - 20.01681-6Bruno Davidovitch BertanhaR.A - 20.01521-6Joao Paulo M SocioR.A - 20.00704-3Ubiratan da Motta FilhoR.A - 20.00928-3

Sumário

Objetivo do projeto	2
Introdução	
Resultados	
Estimativa da média populacional do preço que os usuários consideram justo para um serviço de streaming (variável quantitativa)	2
Estimativa da proporção populacional de usuários que assinam mais de um serviço (variáv	
Estimativas do tamanho da amostra com metade do erro	5

Objetivo do projeto

Estimação de parâmetros da população e dimensionamento amostral.

Introdução

Desde a consolidação dos serviços de streaming como um dos principais meios de entretenimento do mercado global, por se tratar de um setor recente (que começou a se popularizar com a consolidação da Netflix em meados de 2010) ainda possui várias possibilidades e caminhos ainda não explorados. Tendo isso em mente, procuramos realizar uma pesquisa para analisar quais são as preferências e padrões de consumo de serviços de streaming, especificamente no estado de São Paulo, Brasil.

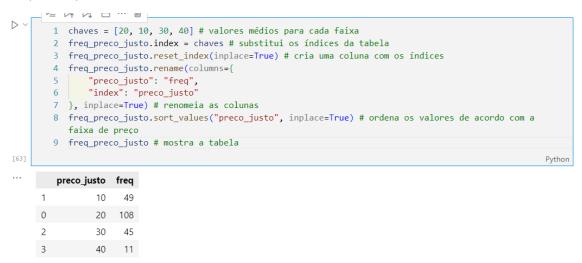
Para entendermos melhor o comportamento dos usuários de streaming, é necessário estimar os parâmetros da população por meio dos dados coletados da nossa amostra. Então, vamos estimar a média populacional do preço que os usuários consideram justo para um serviço de streaming e a proporção populacional de usuários que assinam mais de um serviço. Além disso, vamos calcular o tamanho da amostra necessário para reduzir pela metade o erro – tanto a partir da média populacional quanto a partir da proporção populacional encontradas.

Resultados

Estimativa da média populacional do preço que os usuários consideram justo para um serviço de streaming (variável quantitativa)

Para estimar a média populacional do preço que os usuários consideram justo para um serviço de streaming, como em nossa pesquisa perguntamos o intervalo do preço, necessitamos transformar essa variável qualitativa para quantitativa. Para isso, calculamos a frequência para cada intervalo:

Então, a partir da mediana de cada intervalo (a não ser a categoria "Mais de R\$35,00", que consideramos como 40 para seguir a incremento de 10 em cada valor médio), encontramos a frequência para cada um desses valores:



A partir da fórmula para calcular a média e desvio padrão amostral a partir de uma tabela de frequências:

```
1 freq_preco_justo["xf"] = freq_preco_justo.preco_justo * freq_preco_justo.freq
       2 freq_preco_justo["x2f"] = freq_preco_justo.preco_justo**2 * freq_preco_justo.freq
       3 freq_preco_justo
[42] V 0.1s
                                                                                                   Python
        preco_justo freq
                          xf
                                x2f
     1
               10
                    49
                         490
                              4900
     0
               20
                   108 2160
                             43200
               30
                    45 1350
                             40500
     3
               40 11 440 17600
      1 x_barra = sum(freq_preco_justo.xf) / sum(freq_preco_justo.freq)
       2 print("E(X) = %.3f" % x_barra)
[51] 🗸 0.3s
                                                                                                   Python
   E(X) = 20.845
```

Para estimar a média populacional, vamos utilizar o Caso 2 por desconhecermos o desvio padrão populacional. Considerando um intervalo de confiança de 90% e sabendo que o tamanho da nossa amostra é 213:

Estimativa da proporção populacional de usuários que assinam mais de um serviço (variável qualitativa)

Durante a nossa análise realizada no projeto semestral anterior, encontramos um fato muito interessante em nossa amostra — a grande maioria dos entrevistados assinam mais de 1 serviço de streaming. Com isso em mente, vamos estimar a proporção populacional de usuários que assinam mais de 1 serviço. Para isso, definimos como "0" usuários que assinam até 1 serviço, e "1" para usuários que assinam mais de 1.

```
1 df["n_servico_mais_de_1"] = np.where(df.n_servicos_assinados > 1, 1, 0)
      2 df[["n_servicos_assinados", "n_servico_mais_de_1"]]
                                                                                                    Python
         n_servicos_assinados n_servico_mais_de_1
      Ω
                         3
   1
                         3
                         6
      2
   3
      4
                         5
                         2
                         7
    209
    210
                         2
                         3
    211
    212
                         5
   213 rows × 2 columns
       1 p_barra = df.n_servico_mais_de_1.value_counts(normalize=True)[1]
       2 print("p/ = %.3f" % p_barra)
[55] 🗸 0.5s
   p/ = 0.836
```

Sendo nossa proporção amostral igual à 0,836, calculamos o a proporção populacional com intervalo de confiança de 90%:

Estimativas do tamanho da amostra com metade do erro

Como os resultados possuem casas decimais, então o tamanho da amostra necessário para reduzir o erro pela metade a partir da média e da proporção populacionais são 853 e 871, respectivamente. Ou seja, seria necessário praticamente quadruplicar o tamanho da amostra para reduzir a margem de erro pela metade.