

Módulo: Estatística Descritiva

Aplicações

Aplicação 1) O Banco XPTOKW fez um levantamento de clientes que após tentativas de vendas de determinado produto financeiro, observaram quais aceitaram o produtor e quais não aceitaram. Entretanto, para melhorar a eficiência de suas vendas eles te contrataram para realizar a análise destes dados e mostrar a eles algumas soluções para maximizar a eficiência em novas vendas.

Uma primeira reunião foi realizada contigo e um arquivo com os dados que eles levantaram foi te encaminhado, o arquivo se chama **bank_marketing.xlsx** (Está no portal do aluno).

Com os dados em mãos, você iniciou sua análise exploratória.

Assim, com o que você aprendeu até o momento sobre frequência relativa, absoluta e acumulada:

- a) Quais as principais indagações você pode fazer observando os dados? (Monte pelo menos 3 questões).
- b) Responda às suas indagações com base na análise de dados e com o ferramental visto até o momento.
- c) Qual o comportamento dos casados nesta campanha?
- d) Qual o comportamento das pessoas que possuem formação superior na campanha?
- e) Com o ferramental visto até o momento, qual o perfil mais comum pode ser observado? Sempre justifique com a análise dos dados, para isso você pode relacionar as duas últimas questões acima, ou outras.
- f) Você resolveu observar como é a distribuição de saldo em conta corrente em faixas de valores que variam de 500 a 500 reais, iniciando em 0 e indo até maiores de 15.000, quais conclusões você pode tirar, utilizando o ferramental de frequência relativa, absoluta e acumulada?

Aplicação 2) Você realizou a primeira apresentação para o Banco XPTOKW com as informações que eles te forneceram segundo suas análises anteriores. Entretanto, algumas dúvidas foram levantadas. Você compreendeu desta reunião que mais análises forma necessárias, então:

- a) Quais profissões são responsáveis por cerca de 80% da campanha? (Qual ferramental visto melhor leva a resposta a esta pergunta?)
- b) Uma de suas indagações que surgiu é se há concentração em faixa de valores em saldo de conta corrente pela idade dos clientes do banco de dados que você recebeu. Descreva brevemente como você resolveria esta questão com o ferramental visto até o momento e resolva a indagação.

Aplicação 3) Um dos candidatos à presidência do Brasil solicitou a uma consultoria um levantamento de diversas informações sobre os municípios brasileiros. Com isso, pretende estudar algumas características e propor políticas adequadas para as áreas da Educação, Saúde e aumento de Renda. Considerando a base de dados “**brazil_cities.xlsx**”, extraia dos dados as seguintes informações para auxiliar o candidato:

Panorama geral

- a) Quantos são os municípios que estão no 1º quartil em relação ao índice de desenvolvimento humano, considerando a variável *IDHM*?
- b) Há uma concentração desses municípios em algum estado ou distrito?
Dica: utilize o conceito de Pareto.
- c) Refaça as análises dos itens a) e b) considerando agora o produto doméstico bruto per capita, registrado na variável *GDP_CAPITA*.

Educação Infantil

- d) Construa um boxplot da variável *IDHM_Educacao* por região: NO, NE, SE, S e CO. Comente as diferenças e semelhanças entre os gráficos.
- e) Para a região com a menor média da variável *IDHM_Educacao*, identifique os municípios que estão acima do 3º quartil em relação a quantidade de crianças com idade entre 1 e 4 anos.
- f) Identifique também os municípios que estão no 1º quartil em relação a quantidade de empresas de educação, registrada na variável *COMP_P*.
- g) Considerando as análises anteriores, proponha uma ordem de priorização para investimento na educação por Município.

Renda

- h) Quais municípios estão simultaneamente no 1º quartil da variável *IDHM_Renda* e no 1º quartil da variável *GDP_Capita*?
- i) Crie uma variável indicadora chamada *Renda_Q1* e atribua o valor 1 quando os municípios estiverem na condição descrita no item anterior e o valor 0 caso contrário.
- j) Calcule a média e o desvio padrão da variável *Renda_Q1* por região. Como podemos interpretar esses valores?
- k) Crie uma tabela resumo com média e desvio padrão do *GDP_Capita* também por região e compare com os resultados do item anterior. Você acha que estão consistentes?

Aplicação 4) Um grande fã de vinhos construiu um “robô” para capturar dados de milhares de vinhos na Internet.

- I. Para estudar se existe uma relação entre a **Pontuação** dada a um vinho e seu **Preço**, utilize o mesmo arquivo “**avaliação de vinhos.xlsx**” e realize as seguintes análises:
 - a) Construa um histograma para a variável “price” e outro para a variável “points”. Perceba que a distribuição da variável “price” é bastante assimétrica e que isso é bastante comum em variáveis do tipo preço. Como essa assimetria pode afetar nossa análise, construa uma nova variável chamada \ln_price e utilize a fórmula $=LN(price)$.
 - b) Construa um novo histograma, agora para a variável “ \ln_price ”. A distribuição parece menos assimétrica?
 - c) Construa um gráfico de dispersão entre as variáveis “points” e “ \ln_price ”. Você acha que existe uma associação entre preço e pontuação?
 - d) Calcule o coeficiente de correção de Pearson entre as variáveis “points” e “ \ln_price ”.

Dica: utilize a função CORREL do Excel.

- e) Considerando o coeficiente de correlação de Pearson, é possível afirmar que quanto maior a pontuação de um vinho, maior será seu preço?
- II. Finalmente, para poder selecionar os vinhos que apresentam o melhor custo benefício, o fã de vinhos quer avaliar a relação entre a **Preço** e **Variedade**. Para tanto, utilize o mesmo arquivo “**avaliação de vinhos.xlsx**” e realize as seguintes análises:
 - a) Considerando as três variedades: Cabernet Sauvignon, Chardonnay e Pinot Noir, construa *boxplots* do preço por variedade e uma tabela resumo com as medidas resumo de posição e dispersão. Avaliando apenas os gráficos e tabelas, você entende que exista uma relação entre preço e variedade?
 - b) Calcule o coeficiente de determinação R^2 entre preço e variedade.
 - c) Considerando o coeficiente de determinação, existe uma associação entre preço e variedade?
- III. Considerando as variáveis analisadas nos itens I, II e III, indique se a relação entre elas é de correlação / associação ou causalidade e argumente.
 - a) Pontuação vs. Preço
 - b) Preço vs. Variedade