

Análise Estatística de Dados

Instruções e Questões por grupo:

Diferentes trabalhos acadêmicos, contendo (no [máximo](#)) DUAS questões/tarefas cada um, deverão ser apresentados em classe na forma de código **python** mediante seminário. O cronograma de apresentação é descrito abaixo:

17/04/2024	Apresentação do trabalho acadêmico 1 (para todos os grupos)
24/04/2024	Apresentação do trabalho acadêmico 2 (para todos os grupos)
15/05/2024	Apresentação do trabalho acadêmico 3 (para todos os grupos)
22/05/2024	Apresentação do trabalho acadêmico 4 (para todos os grupos)
29/05/2024	Apresentação do trabalho acadêmico 5 (para todos os grupos)
05/06/2024	Apresentação do trabalho acadêmico 6 (para todos os grupos)
12/06/2024	Apresentação do trabalho acadêmico 7 (para todos os grupos)
19/06/2024	Apresentação do trabalho acadêmico 8 (para todos os grupos)

Grupos:

Grupo 1	Alessandra Souza Gonçalves Ike Gabriel Rodrigues De Kenard Leonardo De Lima Amaral Marcelo Saraiva Cavalcanti Thiago Castro Barreto
Grupo 2	Isadora Almeida Poppi Durante João Victor Ferreira Marques Eduardo Sousa Hirle De Freitas Eduardo Galvão De Aquino Cavalheiro Alice Moreira Marques
Grupo 3	Pedro Aragão Dorneles Manuela Borges Raupp Fonseca
Grupo 4	Camila Niederauer Culau Arthur Marques Vida De Oliveira Tiago De Oliveira Teixeira Luiz Henrique Costa Araujo
Grupo 5	Jamille Rocha Ghazaleh Lucas Leandro Wall Bruno Carlos Mateo Wall Bruno Joao Vitor De Cesaro Moura
Grupo 6	Luane Silva Martins De Santana Guilherme Gomes Matos
Grupo 7	Matheus Nery Da Cunha Távora Rafael Mascarenhas Brown De Andrade Vinícius Magrinelli Matheus Graça Lira Vito Borges Steele Fusaro
Grupo 8	Rodrigo César Lins Bezerra Magalhães Jose Roberto Guimarães Muller Matheus Alves Caetano Palha Vitor Nascimento Franco

Grupo 9	ARTHUR ROCHA GOMES ANA GABRIELLY SILVA ASVOLINSQUE FARIA CAIO HENRIQUE ARARUNA ALMEIDA
----------------	--

Cada grupo terá tarefas distintas, identificadas pelas respectivas questões. Abaixo estão as tarefas a serem realizadas e apresentadas, **separadas por datas**, inclusive.

Continua...

Dia 17 de abril de 2.024

Grupo 01:

Alessandra Souza Gonçalves
Ike Gabriel Rodrigues De Kenard
Leonardo De Lima Amaral
Marcelo Saraiva Cavalcanti
Thiago Castro Barreto

Grupo 02:

Isadora Almeida Poppi Durante
João Victor Ferreira Marques
Eduardo Sousa Hirle De Freitas
Eduardo Galvão De Aquino Cavaleiro
Alice Moreira Marques

Questão A

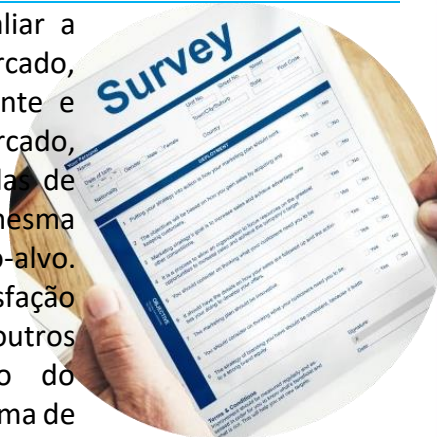
Crie e execute um programa em **python** que, dadas as cotações do dólar no Brasil no mês de agosto referente aos últimos cinco anos, mostre um gráfico de média móvel de 7 dias referente a tais cotações para este mês. No gráfico, devem ser apresentadas legendas diferenciadas por ano.

Observações:

- 1) Como resultado da execução do código, deve aparecer via terminal, além do gráfico supracitado acima, um histograma de frequência das cotações do dólar, onde é solicitado via entrada do usuário o número de classes desejado para a construção do histograma;
- 2) Ainda deve ser mostrado para o usuário via gráfico o cruzamento das linhas gráficas de diferentes legendas.

Questão B

O coeficiente **Alfa de Cronbach** é utilizado para avaliar a consistência interna das perguntas ou itens da pesquisa de mercado, verificando se as questões estão medindo de forma coerente e confiável a variável de interesse. Na análise de pesquisas de mercado, tal parâmetro é especialmente relevante em relação a escalas de medida, onde várias perguntas são utilizadas para medir uma mesma característica (também conhecida como construto) do público-alvo. Essas escalas podem ser usadas, por exemplo, para medir a satisfação do cliente, a intenção de compra, a lealdade à marca ou outros aspectos importantes para entender o comportamento do consumidor (um exemplo de escala é a escala “**likert**”). Uma forma de avaliar a confiabilidade de um questionário reside na inspeção deste coeficiente:



este mede a correlação entre respostas em um questionário através da análise das respostas dadas pelos respondentes, apresentando uma correlação média entre as perguntas. Neste sentido, todo questionário é constituído por várias questões elaboradas em sequência, que aqui serão chamadas de itens. Considere um questionário com k itens, isto é, $i = 1, 2, \dots, k$, respondido por n pessoas, isto é, $j = 1, 2, \dots, n$.

As opções de resposta para cada item podem ser dicotômicas como “Sim” e “Não” ou escalonadas como “Concordo plenamente”, “Concordo”, “Não concordo nem discordo”,

“Discordo”, “Discordo completamente” (veja uma possível escala **LIKERT** a seguir). Para o cálculo do coeficiente, toda resposta deve ser transformada em números.

1	2	3	4	5
Discordo totalmente	Discordo parcialmente	Não concordo, nem discordo	Concordo parcialmente	Concordo totalmente

Tabela: Exemplo de escala LIKERT

Este cálculo se baseia na fórmula:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k s_i^2}{S_{soma}^2} \right)$$

em que:

k é o número de itens (perguntas) do questionário,

s_i^2 é a variância amostral (desvio padrão amostral ao quadrado) referente às respostas do i -ésimo item ($i = 1, \dots, k$),

S_{soma}^2 é a variância amostral (desvio padrão amostral ao quadrado) referente às somas dos valores das respostas dos itens (perguntas) de cada indivíduo nos k itens.

O valor do parâmetro **Alfa de Cronbach** varia entre 0 a 100%, sendo este último mostrando uma **confiabilidade perfeita**.

Como exemplo ilustrativo, considere o seguinte exercício onde as opções de resposta eram **dicotômicas**, isto é, “Sim”, que foi transformado em 1 (atribuído como valor de item) e “Não”, que foi transformado em zero.

Tabela 1- Resultados da aplicação de um questionário com cinco itens e dez respondentes

RESPONDENTES	ITENS				
	1	2	3	4	5
1	1	1	1	1	1
2	1	1	1	1	0
3	1	1	1	1	1
4	1	1	0	0	0
5	1	1	0	0	0
6	1	1	0	0	1
7	1	1	0	0	0
8	1	1	1	0	0
9	1	1	1	0	0
10	1	0	0	0	0

A partir de tais dados, crie e execute um programa em **python** que determine:

A) o valor aproximado do coeficiente **Alfa de Cronbach**;

B) Uma maneira prática de julgar o valor do parâmetro α é dada na tabela 2 abaixo, desde que se tenha a precaução de levar em conta as limitações dessa estatística. Neste caso, verifique a validade do questionário em questão a partir do coeficiente calculado no item **A**.

Tabela 2 - Confiabilidade do questionário segundo o valor de α

Valor de alfa	Confiabilidade
Maior do que 0,9	Excelente
0,8 - 0,9	Bom
0,7 - 0,8	Aceitável
0,6 - 0,7	Questionável
0,5 - 0,6	Pobre
Menor do que 0,5	Inaceitável

Observações:

- 1) Como resultado da execução do código, deve aparecer via terminal, uma tabela das respostas do questionário, bem como o valor do **Alfa de Cronbach** calculado;
- 2) Uma mensagem deve descrever a qualidade do questionário (de acordo com a tabela 2 acima) referente à confiabilidade.

Grupo 03:

Pedro Aragão Dorneles

Manuela Borges Raupp Fonseca

Grupo 04:

Camila Niederauer Culau

Arthur Marques Vida De Oliveira

Tiago De Oliveira Teixeira

Luiz Henrique Costa Araujo

Questão A

Um instrumento importante para entender uma equipe e traçar ações mais eficazes de gestão é o mapeamento. Para realizar tal tarefa, é necessário considerar as competências elencadas pela organização como desejáveis. Nesse sentido, suponha que uma empresa tenha solicitado ao gestor de vendas que mapeie a sua equipe considerando as seguintes características/competências:

- **Conhecimento** — entendimento de como se faz;
- **Habilidade** — saber fazer;
- **Atitude** — vontade de fazer;
- **Visão** — organizar, planejar o que se faz;
- **Ética** — comportamento em grupo;
- **Superação** — fazer acontecer.

Utilizando uma escala de 1 a 5 pontos, sendo que o valor 1 corresponde à “pontuação muito baixa” e 5 “pontuação muito alta”, foi obtido o seguinte mapeamento:

Pessoas da Equipe	Conhecimento	Habilidade	Atitude	Visão	Ética	Superação
A	2	2	4	1	4	2
B	1	1	3	2	3	4
C	2	1	2	4	2	3
D	2	4	1	3	1	4
E	4	3	4	3	4	3

Sob a ótica estatística, considere que o desvio padrão amostral dos valores obtidos para cada membro da equipe e sobre cada competência tenha um valor crítico igual a 0,1. Ou seja, qualquer valor para o desvio padrão amostral abaixo de 0,1 mostra baixa variabilidade das respostas. Suponha ainda que se deseje calcular uma pontuação média para cada membro. Neste caso, cada competência terá um grau de importância diferenciada. Assim, as características “Conhecimento”, “Habilidade”,

“Atitude”, “Visão”, “Ética” e “Superação” possuem pesos estatísticos iguais a 0,3, 0,25, 0,2, 0,15, 0,05 e 0,05, respectivamente.

Considere que as respostas mudam aleatoriamente (simulando padrões de respostas de diferentes equipes). Neste caso, elabore um programa em **python** que determine:

- A)** A pontuação média por membro da equipe;
- B)** O desempenho médio da equipe.

Observações:

- 1) Como resultado da execução do código, deve aparecer via terminal, uma tabela da pontuação da equipe, bem como o desempenho médio da equipe;
- 2) Uma mensagem mencionando se houve alta ou baixa variabilidade nas respostas, indicando a competência com tal ocorrência.

Grupo 05:

Jamille Rocha Ghazaleh

Lucas Leandro Wall Bruno

Carlos Mateo Wall Bruno

Joao Vitor De Cesaro Moura

Grupo 06:

Luane Silva Martins De Santana

Guilherme Gomes Matos

Questão A

O APRENDIZ é um programa de radiodifusão/reality show sobre gestão e empreendedorismo onde há um grupo seletivo de candidatos que concorrem a um contrato em uma empresa, por pelo menos um ano, pago pelo patrocinador do programa. Divididos em duas equipes, os candidatos devem realizar uma série de tarefas envolvendo vendas, marketing, promoções, negócios imobiliários, finanças, investimentos, ações beneficentes, publicidade e gerenciamento de negócios, bem como arrecadação de fundos, administração de bens e realização de eventos de grandes proporções.



Episódio do programa O APRENDIZ – fonte: Google Imagens

Em um dos episódios do programa (link do vídeo:

<https://www.youtube.com/watch?v=Zc0V9OhNMAw> - veja o vídeo entre os instantes 6min53s e 9min08s), uma tarefa é atribuída às equipes, onde ambas terão que operar no mercado financeiro usando um crédito fictício de R\$ 1.000.000,00 para aquisição de ações na bolsa, baseando-se em dados REAIS.

A equipe que conseguir gerar maior lucratividade/rentabilidade em investimentos vence a tarefa. Neste sentido, considere que ambas as equipes escolheram um mesmo ativo financeiro para investimento e que adotaram uma metodologia estatística para a tomada de decisão. A série histórica de 10 dias deste ativo é mostrada na figura ao lado. Desse modo, como metodologia inicial, considere um intervalo de confiança bilateral de 95%, bem como para a compra e venda de ações os valores extremos deste intervalo.

Dia	Preço
1	R\$ 15,01
2	R\$ 18,92
3	R\$ 20,00
4	R\$ 15,78
5	R\$ 17,17
6	R\$ 16,23
7	R\$ 18,23
8	R\$ 18,00
9	R\$ 17,88
10	R\$ 19,02

Elabore um programa em **python** que determine:

- A)** os valores ideais para a compra e venda de ações, sendo estes os valores extremos de tal intervalo de confiança;
- B)** o valor aproximado do lucro obtido previsto.

Observações:

Como resultado da execução do código, deve aparecer via terminal, os valores de compra e venda, bem como a quantidade de ações adquiridas (número inteiro mais próximo), bem como o valor do lucro obtido.

Grupo 07:

Matheus Nery Da Cunha Távora
Rafael Mascarenhas Brown De Andrade
Vinícius Magrinelli
Matheus Graça Lira
Vito Borges Steele Fusaro

Grupo 08:

Rodrigo César Lins Bezerra Magalhães
Jose Roberto Guimarães Muller
Matheus Alves Caetano Palha
Vitor Nascimento Franco

Grupo 09:

ARTHUR ROCHA GOMES
ANA GABRIELLY SILVA ASVOLINSQUE FARIA
CAIO HENRIQUE ARARUNA ALMEIDA

Questão A

Para atender e documentar uma sistemática relativa a pesquisas de mercado e opinião sobre um determinado assunto (ou sobre uma variável de estudo), usa-se métodos de dimensionamento de amostra. Isso é feito mediante aplicação de um questionário com questões objetivas, com possibilidades de respostas mediante uma escala previamente estabelecida.

Neste sentido, deve-se utilizar uma indicação da variabilidade existente na pesquisa, ou seja, deve ser apresentado um indicador numérico. Um exemplo de indicador a ser utilizado é a resposta obtida de cada pergunta/questão a ser avaliada pelo respondente. Tais questões podem apresentar um “grau” de importância diferente no contexto do questionário.

A medida geral de variabilidade é estimada através de uma média ponderada dos desvios padrões associados às respostas, onde os pesos estatísticos retratam a importância de cada pergunta a ser respondida no questionário. A fórmula para obter a variabilidade é dada por:

$$\sigma_q = \frac{\sum_{i=1}^k \sigma_i \times w_i}{\sum_{i=1}^k w_i} = \frac{\sigma_1 \times w_1 + \sigma_2 \times w_2 + \sigma_3 \times w_3 + \dots}{w_1 + w_2 + w_3 + \dots}$$

onde:

σ_q é o desvio padrão populacional para a variável em análise no questionário;

k é o número de questões;

σ_i é o desvio padrão populacional das respostas obtidas da pergunta/questão i ;

w_i é o peso estatístico quanto à importância da pergunta/questão i no questionário.

Dessa forma, o tamanho da amostra n é dado pela seguinte fórmula:

$$n = \left(\frac{z \times \sigma_q}{E} \right)^2$$

Nesse caso, n é o tamanho mínimo da amostra a ser consultada, E é o fator erro (este é fixado antes de se iniciar a pesquisa) e z a variável padronizada associada à margem de confiança do estudo (por exemplo, para uma margem de confiança de **90%**, o valor de z é igual a **1,65**; para uma margem de confiança de **95%**, o valor de z é igual a **1,96**).

Portanto, o dimensionamento do tamanho da amostra depende de quatro fatores principais: **(i)** o nível de significância/margem de confiança desejado(a); **(ii)** a variabilidade das respostas e grau de importância de cada pergunta; **(iii)** erro máximo de estimação/estimativa admitida na pesquisa.

O tamanho da amostra é determinado em duas etapas:

- 1) Aplica-se um questionário na forma de **PRÉ-TESTE** (questionário preliminar a ser aplicado a um conjunto de pessoas de maneira subjetiva, a critério do examinador, com o intuito de verificar problemas de compreensão das questões e fornecer subsídios para a sua modificação);
- 2) Com base nas respostas obtidas na pesquisa da etapa 1 acima, calcula-se o desvio padrão geral para a variável em análise no questionário σ_q e o tamanho da amostra n . Esse último conjunto “corrige” o tamanho escolhido pelo observador na etapa 1 no sentido probabilístico.

Como exemplo, considere um questionário com cinco questões, tendo um padrão de respostas na escala **LIKERT**, numeradas de 1 a 5 (veja a ilustração abaixo). As respostas obtidas de dez respondentes através de um pré-teste são dadas na tabela a seguir:

1	2	3	4	5
Discordo totalmente	Discordo parcialmente	Não concordo, nem discordo	Concordo parcialmente	Concordo totalmente

Respondentes	Questões/perguntas				
	1ª	2ª	3ª	4ª	5ª
1º	3	3	1	3	3
2º	5	4	2	4	5
3º	4	3	1	3	5
4º	4	3	2	3	5
5º	5	3	2	3	4
6º	4	2	2	4	5
7º	3	3	1	3	4
8º	3	4	2	3	5
9º	5	4	3	3	5

10º	5	4	2	2	4
-----	---	---	---	---	---

Considere ainda que as questões de um a cinco possuem pesos estatísticos quanto à sua importância no contexto do questionário iguais a 0,2, 0,3, 0,15, 0,15 e 0,20, respectivamente. Com um nível de confiança de 99% e um fator de erro igual a 0,1.

Considere que as respostas mudam aleatoriamente (simulando padrões de respostas de diferentes amostras). Neste caso, elabore um programa em **python** que determine:

- A) A pontuação média por membro da equipe;
- B) A pontuação média na amostra.

Observações:

- 1) Como resultado da execução do código, deve aparecer via terminal, uma tabela das respostas na amostra, bem como a pontuação média;

Questão B

O coeficiente **Alfa de Cronbach** é utilizado para avaliar a consistência interna das perguntas ou itens da pesquisa de mercado, verificando se as questões estão medindo de forma coerente e confiável a variável de interesse. Na análise de pesquisas de mercado, tal parâmetro é especialmente relevante em relação a escalas de medida, onde várias perguntas são utilizadas para medir uma mesma característica (também conhecida como construto) do público-alvo. Essas escalas podem ser usadas, por exemplo, para medir a satisfação do cliente, a intenção de compra, a lealdade à marca ou outros aspectos importantes para entender o comportamento do consumidor (um exemplo de escala é a escala "likert"). Uma forma de avaliar a confiabilidade de um questionário reside na inspeção deste coeficiente:



este mede a correlação entre respostas em um questionário através da análise das respostas dadas pelos respondentes, apresentando uma correlação média entre as perguntas. Neste sentido, todo questionário é constituído por várias questões elaboradas em sequência, que aqui serão chamadas de itens. Considere um questionário com k itens, isto é, $i = 1, 2, \dots, k$, respondido por n pessoas, isto é, $j = 1, 2, \dots, n$.

As opções de resposta para cada item podem ser dicotômicas como "Sim" e "Não" ou escalonadas como "Concordo plenamente", "Concordo", "Não concordo nem discordo", "Discordo", "Discordo completamente" (veja uma possível escala **LIKERT** a seguir). Para o cálculo do coeficiente, toda resposta deve ser transformada em números.

1	2	3	4	5
Discordo totalmente	Discordo parcialmente	Não concordo, nem discordo	Concordo parcialmente	Concordo totalmente

Tabela: Exemplo de escala LIKERT

Este cálculo se baseia na fórmula:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k s_i^2}{S_{soma}^2} \right)$$

em que:

k é o número de itens (perguntas) do questionário,

s_i^2 é a variância amostral (desvio padrão amostral ao quadrado) referente às respostas do i -ésimo item ($i = 1, \dots, k$),

S_{soma}^2 é a variância amostral (desvio padrão amostral ao quadrado) referente às somas dos valores das respostas dos itens (perguntas) de cada indivíduo nos k itens.

O valor do parâmetro **Alfa de Cronbach** varia entre 0 a 100%, sendo este último mostrando uma **confiabilidade perfeita**. A partir de tal parâmetro, pode-se desenvolver diferentes tipos de escalas numéricas voltadas para diferentes tipos de pesquisas, tais como estilos de liderança, perfil de consumo etc.

Para classificar os estilos de liderança, Fiedler (1965) desenvolveu uma medida chamada escala **LPC** (*Least preferred co-worker* – o colaborador menos preferido). Esta medida é obtida em formato de questionário, onde é solicitado aos respondentes que descrevam o colega de trabalho menos preferido com quem já trabalharam, levando em consideração uma lista de 16 adjetivos bipolares numa escala de 1 a 8. As respostas obtidas são somadas, dando o resultado do parâmetro **LPC** para cada respondente (pode-se obter um valor médio para os respondentes).

Se o **LPC** for elevado, isto é, se o colaborador menos preferido for descrito em termos de adjetivos relativamente “positivos/favoráveis”, isto significa (de acordo com Fiedler) que o estilo de liderança do respondente é orientado para as relações humanas, ou seja, a motivação primária deste é ter uma relação mais próxima ao grupo. Ao contrário, se o **LPC** for baixo, ou seja, o colega de trabalho menos preferido é descrito por adjetivos relativamente “negativos/desfavoráveis”, o estilo de liderança do respondente é orientado para as tarefas. A classificação dos valores da escala **LPC** encontra-se na tabela abaixo:

ESCALA LPC – CLASSIFICAÇÃO	
73 ou acima	LPC elevado
64 a 72	LPC intermediário
63 ou inferior	LPC baixo

Considere uma situação em que foi feita uma pesquisa consultando cem pessoas aleatoriamente selecionadas num determinado setor de uma empresa.

ESCALAS DE RESPOSTAS

1	Agradável	8	7	6	5	4	3	2	1	Desagradável
2	Amistoso	8	7	6	5	4	3	2	1	Inamistoso
3	Rejeitador	1	2	3	4	5	6	7	8	Acolhedor
4	Prestativo	8	7	6	5	4	3	2	1	Frustrante
5	Sem entusiasmo	1	2	3	4	5	6	7	8	Entusiasta
6	Tenso	1	2	3	4	5	6	7	8	Calmo
7	Distante	1	2	3	4	5	6	7	8	Íntimo
8	Frio	1	2	3	4	5	6	7	8	Caloroso
9	Cooperativo	8	7	6	5	4	3	2	1	Não-cooperativo
10	Apoiador	8	7	6	5	4	3	2	1	Hostil
11	Maçante	1	2	3	4	5	6	7	8	Interessante
12	Desequilibrado	1	2	3	4	5	6	7	8	Equilibrado
13	Seguro de si	8	7	6	5	4	3	2	1	Hesitante
14	Eficiente	8	7	6	5	4	3	2	1	Ineficiente
15	Sombrio	1	2	3	4	5	6	7	8	Alegre
16	Expansivo	8	7	6	5	4	3	2	1	Reservado

Elabore um código em **python** que, a partir da escala de adjetivos acima, simula respostas aleatórias desta amostra de pessoas consultadas. Neste sentido, via terminal, deverá ser mostrado o valor do coeficiente **Alfa de Cronbach** para a referida escala **LPC**.

Observações:

- 1) Como resultado da execução do código, deve aparecer via terminal, uma tabela das respostas do questionário, bem como o valor do **Alfa de Cronbach** calculado;
- 2) Uma mensagem deve descrever a qualidade do questionário (de acordo com a tabela abaixo) referente à confiabilidade.

Valor de alfa	Confiabilidade
Maior do que 0,9	Excelente
0,8 - 0,9	Bom
0,7 - 0,8	Aceitável
0,6 - 0,7	Questionável
0,5 - 0,6	Pobre
Menor do que 0,5	Inaceitável

Continua...

Dia 24 de abril de 2.024

Grupo 01:

Alessandra Souza Gonçalves
Ike Gabriel Rodrigues De Kenard
Leonardo De Lima Amaral
Marcelo Saraiva Cavalcanti
Thiago Castro Barreto

Grupo 02:

Isadora Almeida Poppi Durante
João Victor Ferreira Marques
Eduardo Sousa Hirle De Freitas
Eduardo Galvão De Aquino Cavaleiro
Alice Moreira Marques

Questão A

A taxa de conversão de vendas é uma métrica que avalia a eficácia de uma loja física ao transformar visitantes (público “*entrante*”) em clientes que efetuam compras. Essa taxa é calculada dividindo o número total de transações de vendas pelo número total de visitantes ou clientes potenciais (público “*entrante*”), sendo o resultado frequentemente expresso na forma de porcentagem.

Público Passante:
Refere-se às pessoas que transitam nas proximidades de um comércio, sem necessariamente entrar na loja.

Essa taxa é calculada dividindo o número total de transações de compra pelo número total de visitantes ou clientes potenciais, na forma percentual. Isto é:

$$\text{Taxa de Conversão} = \frac{\text{N}^\circ \text{ de vendas}}{\text{N}^\circ \text{ de Visitantes/Entrantes}} \times 100\%$$

Estudo de caso:

Considere uma situação em que foi feito um levantamento estatístico quanto ao fluxo (número) de entrantes e de vendas num comércio varejista, durante um período de 10 dias úteis consecutivos. Os dados consolidados estão registrados na tabela abaixo:

Público Entrante:
São as pessoas que efetivamente entram no estabelecimento, podendo explorar e consumir os produtos ou serviços oferecidos.

Número de “Entrantes”				Número de Vendas			
Dia útil	Manhã	Tarde	Noite	Dia útil	Manhã	Tarde	Noite
1	164	198	102	1	26	32	11
2	155	185	108	2	27	36	13
3	153	192	106	3	30	37	11
4	150	183	132	4	25	31	10
5	160	199	113	5	26	38	15
6	160	199	118	6	29	30	13
7	155	181	113	7	27	31	11
8	159	182	107	8	29	37	10
9	151	182	119	9	27	39	12
10	151	191	115	10	21	30	13

A partir destes dados, determine:

- a taxa de conversão diária para cada dia útil ilustrado;
- a taxa de conversão diária média;
- o desvio padrão amostral referente à variável “taxa de conversão diária”;

Obs.: o grupo será arguido sobre o código-fonte em “python”, bem como quanto à resolução por escrito do problema.

Questão B

No campo da auditoria, existem técnicas de mineração de dados que servem de auxílio no processo de detecção de fraudes financeira (por exemplo, em auditoria de obras públicas, fraude em registro contábil de certos parâmetros financeiros em empresas públicas e privadas). Uma metodologia muito utilizada é a **Lei de Newcomb-Benford**, que confronta a probabilidade de ocorrência esperada de dígitos com a probabilidade observada, ou seja, detecta discrepância em relação a um padrão universal de ocorrência dos dígitos. Tal probabilidade esperada é descrita pela referida lei através da função $P_e(d) = \log_{10}(1 + 1/d)$, onde d é o valor do primeiro dígito e $P_e(d)$ é a probabilidade de ocorrência deste dígito num dado número.

Primeiro Dígito	Probabilidade esperada $P_e(d)$
1	0,301029996
2	0,176091259
3	0,124938737
4	0,096910013
5	0,079181246
6	0,06694679
7	0,057991947
8	0,051152522
9	0,045757491

Tabela 1 – Probabilidade dada pela **Lei de Newcomb-Benford**.

Frank Benford demonstrou que tal lei se aplica **sob certas condições** a uma ampla variedade de conjuntos de dados, incluindo valores de contas comerciais, precificação de ações, mercado imobiliário, taxas de mortalidade, registros de imposto de renda, valores de faturas de compra e venda etc. A tabela 1 acima ilustra os valores da **Lei de Newcomb-Benford**.

Como ilustração para verificação, tem-se ao lado um conjunto de registros do primeiro dígito relativo a **500** valores de lucros de uma empresa, a partir de observações de dados históricos. A tabela 2 ao lado ilustra as frequências observadas.

Considere que a estatística calculada de padronização z seja dada pela expressão:



$$z_{calc} = \frac{|p_o - p_e|}{\sqrt{\frac{p_e(1 - p_e)}{n}}}$$

Primeiro Dígito	Frequência observada
1	148
2	75
3	55
4	50
5	45
6	40
7	34
8	28
9	25

Tabela 2 – Frequência observada do primeiro dígito.

onde:

p_o é a probabilidade observada de cada dígito específico (de 1 a 9);

p_e é a probabilidade esperada de cada dígito de acordo com a probabilidade definida pela **Lei de Newcomb-Benford** para cada primeiro dígito;

n é o número de observações.

Com base nessas informações, elabore um programa em **“python”** que determine as probabilidades esperadas, as frequências esperadas, as probabilidades observadas, as frequências observadas e o valor da estatística de teste **“z”** para os respectivos dígitos, considerando que houve **500** observações na pesquisa (crie um **“dataframe”** de acordo com o **modelo** abaixo):

Primeiro Dígito	Probabilidade esperada $P_e(d)$	Frequência esperada (???)	Probabilidade observada (???)	Frequência observada	z_{calc} (???)
1				148	

2				75	
3				55	
4				50	
5				45	
6				40	
7				34	
8				28	
9				25	

Observação:

Como resultado da execução do código, deve aparecer via terminal um “[dataframe](#)” contendo todas os resultados do modelo acima.

Grupo 03:

Pedro Aragão Dorneles

Manuela Borges Raupp Fonseca

Grupo 04:

Camila Niederauer Culau

Arthur Marques Vida De Oliveira

Tiago De Oliveira Teixeira

Luiz Henrique Costa Araujo

Questão A

Testes de hipóteses são cruciais na área de seguros porque permitem tomar decisões baseadas em dados, avaliar riscos, detectar fraudes, cumprir regulamentações, melhorar processos, e prever tendências. A estatística desempenha um papel fundamental na gestão de uma indústria que lida com incertezas e probabilidades. Além disso, esta metodologia estatística desempenha um papel vital na determinação do valor das apólices de seguros, pois fornece uma base estatística para avaliar riscos, definir prêmios apropriados, segmentar o mercado e garantir a sustentabilidade financeira das seguradoras, permitindo que as seguradoras tomem decisões informadas e precisas no cálculo dos prêmios.

A tabela abaixo apresenta as quantias das indenizações pagas por estragos em automóveis, para uma amostra aleatória de 10 motoristas segurados, envolvidos em acidentes de pequena gravidade em uma determinada área geográfica. Com estes dados, elabore um código em ambiente **python** que teste a hipótese de que o valor médio das indenizações na população amostrada é superior a \$1.000,00, utilizando um nível de significância estipulado via terminal de entrada pelo usuário.

\$1.033,00	\$1.069,00
\$1.274,00	\$1.121,00
\$1.114,00	\$1.269,00
\$924,00	\$1.150,00
\$1.421,00	\$921,00

Tabela – Indenizações pagas por estragos em automóveis

Observações:

- 1) O código deve, necessariamente, possuir os seguintes comandos (funções): **input**, **DataFrame**, **for** e **if-elif**. Além disso, deve-se utilizar uma função que determine o **valor crítico** da distribuição associada para o referido teste de hipóteses;
- 2) O código deverá conter, **necessariamente**, a biblioteca **scipy.stats**. Neste caso, o valor crítico a ser determinado deve ser calculado usando uma função pertencente a esta biblioteca;
- 3) Como resultado final, o código deverá fornecer as seguintes informações via terminal de saída:
 - (i) Uma mensagem do tipo “O valor da estatística de teste é igual a” < citar o valor >;

- (ii) Uma mensagem do tipo “A confiança do teste é igual a” < citar o valor da confiança >:
- (iii) Uma mensagem do tipo “O valor crítico da distribuição associada é igual a” < citar o valor crítico >;
- (iv) Uma mensagem do tipo “A hipótese/afirmação dada deve ser” < citar o resultado do teste de hipótese: aceita-se ou rejeita-se a afirmação dada no problema >;
- (v) o grupo será arguido sobre o código-fonte em “python”, bem como quanto à resolução por escrito do problema.

Grupo 05:

Jamille Rocha Ghazaleh
 Lucas Leandro Wall Bruno
 Carlos Mateo Wall Bruno
 Joao Vitor De Cesaro Moura

Grupo 06:

Luane Silva Martins De Santana
 Guilherme Gomes Matos

Questão A

A produtividade agrícola é um indicador crucial para a avaliação do desempenho e da sustentabilidade das culturas. A análise dos dados destaca a dinâmica complexa que influencia a produtividade dessas commodities. Fatores como condições climáticas, avanços tecnológicos e demanda do mercado desempenham papéis cruciais nesse cenário. A compreensão dessas tendências é essencial para agricultores, empresas do setor e formuladores de políticas, pois permite o desenvolvimento de estratégias mais eficazes para o setor agrícola.

Nesse contexto, tem-se na tabela abaixo um comparativo de produção (em milhares de toneladas) no período entre 2013 e 2023 de três importantes commodities agrícolas: **milho**, **soja** e **algodão**.

Ano	Milho ¹	Soja ¹	Algodão ²
2013	330,4	186,7	2015
2014	375,7	216	2357
2015	196,3	147,32	2319
2016	180,6	231	1577
2017	223,1	246,68	2576
2018	225,2	259,55	2835
2019	243	241,6	2700
2020	219,5	290,6	2853
2021	203,5	292	2851
2022	154,1	313,2	2537
2023	153,2	318,5	2883

¹ Distrito Federal

² Estado da Bahia

Fonte: <https://www.conab.gov.br>

Elabore um código-fonte em linguagem “python” que indique uma previsão para a produção de tais commodities para o ano de 2024 usando um **intervalo de confiança de 99%**.

Obs.: o grupo será arguido sobre o código-fonte em “python”, bem como quanto à resolução por escrito do problema.

Questão B

A precisão na estimativa da vida útil de produtos é fundamental para assegurar a confiabilidade e satisfação do cliente, bem como para orientar decisões estratégicas nas áreas de produção, marketing e gestão de qualidade. Neste sentido, o uso de “intervalos de confiança” revela-se como uma abordagem estatística robusta para estimar a durabilidade média de produtos, utilizando dados amostrais representativos, resultando numa estimativa importante para a verdadeira durabilidade média da população de produtos. Este procedimento é essencial para proporcionar transparência e confiabilidade nas informações sobre a durabilidade dos produtos, melhorando a gestão da qualidade, o planejamento de produção e a comunicação com os consumidores.

Considere uma situação em que um novo fabricante deseja entrar no mercado de segmento de lâmpadas. Neste nicho mercadológico, já existe concorrência de mercado, a saber:

1. **Osram, Philips e General Electric (GE):** marcas globais que têm presença significativa no mercado brasileiro, oferecendo uma variedade de soluções para iluminação em diferentes contextos, como residências, ambientes comerciais e iluminação automotiva.
2. **Tecnolite, Brilia, Startec e Golden:** marcas brasileiras que também desempenham um papel importante, proporcionando opções diversificadas para atender às necessidades específicas dos consumidores locais. A Tecnolite, por exemplo, oferece uma variedade de produtos de iluminação, incluindo lâmpadas LED e fluorescentes, enquanto a Brilia se destaca por suas soluções inteligentes e inovadoras para iluminação residencial. A St artec, com seu foco em iluminação decorativa, enquanto que a Golden oferece fabricação de lâmpadas LED.

Para tanto, o fabricante deve buscar uma estimativa para a vida média “ideal” de operação de sua linha de produtos destinados ao mercado. Neste caso, a partir de uma produção inicial de teste, obtém-se uma amostra de referência. Esta consiste em 10 lâmpadas. Após análise do padrão de qualidade feita dentro da fábrica, é verificada que a vida média de operação dessa amostra é igual a 4.000 horas, com um desvio padrão amostral de 200 horas. Assume-se que o tempo de operação das lâmpadas segue uma distribuição aproximadamente normal.

Elabore um código-fonte em linguagem “python” que indique uma estimativa para a vida média de operação para a população de lâmpadas da qual foi extraída a amostra, usando um intervalo de confiança bilateral de 95%.

Obs.: o grupo será arguido sobre o código-fonte em “python”, bem como quanto à resolução por escrito do problema.

Grupo 07:

Matheus Nery Da Cunha Távora
Rafael Mascarenhas Brown De Andrade
Vinícius Magrinelli
Matheus Graça Lira
Vito Borges Steele Fusaro

Grupo 08:

Rodrigo César Lins Bezerra Magalhães
Jose Roberto Guimarães Muller
Matheus Alves Caetano Palha
Vitor Nascimento Franco

Grupo 09:

ARTHUR ROCHA GOMES
ANA GABRIELLY SILVA ASVOLINSQUE FARIA
CAIO HENRIQUE ARARUNA ALMEIDA

Questão A

Para classificar os estilos de liderança, Fiedler (1965) desenvolveu uma medida chamada escala **LPC** (*Least preferred co-worker* – o colaborador menos preferido). Esta medida é obtida em formato de questionário, onde é solicitado aos respondentes que descrevam o colega de trabalho menos preferido com quem já trabalharam, levando em consideração uma lista de 16 adjetivos bipolares numa escala de 1 a 8. As respostas obtidas são somadas, dando o resultado do parâmetro **LPC** para cada respondente (pode-se obter um valor médio para os respondentes).

Se o **LPC** for elevado, isto é, se o colaborador menos preferido for descrito em termos de adjetivos relativamente “positivos/favoráveis”, isto significa (de acordo com Fiedler) que o estilo de liderança do respondente é orientado para as relações humanas, ou seja, a motivação primária deste é ter uma relação mais próxima ao grupo. Ao contrário, se o **LPC** for baixo, ou seja, o colega de trabalho menos preferido é descrito por adjetivos relativamente “negativos/desfavoráveis”, o estilo de

liderança do respondente é orientado para as tarefas. A classificação dos valores da escala **LPC** encontra-se na tabela abaixo:

ESCALA LPC – CLASSIFICAÇÃO	
73 ou acima	LPC elevado
64 a 72	LPC intermediário
63 ou inferior	LPC baixo

Considere uma situação em que foi feita uma hipótese de que a média da escala **LPC** de um determinado setor de uma empresa tenha um valor **baixo**. Para averiguar tal afirmação, foi realizada uma entrevista com um grupo amostral de 15 respondentes aleatoriamente selecionados desse setor. A média amostral dos valores **LPC** obtidos é igual a 70. O desvio padrão amostral para o valor **LPC** é igual a 6,23. Supõe-se que os valores **LPC** tenham distribuição normal na população. Com base neste resultado amostral, faça o teste de hipóteses usando um nível de significância de 5%.

Elabore um código-fonte em linguagem “python” que indique o resultado do teste de hipóteses sobre a afirmação “*a média da escala LPC de um determinado setor de uma empresa tenha um valor igual a 62*”. Use um **intervalo de confiança de 95%**.

Obs.: o grupo será arguido sobre o código-fonte em “python”, bem como quanto à resolução por escrito do problema.

ESCALAS DE RESPOSTAS

1	Agradável	8	7	6	5	4	3	2	1	Desagradável
2	Amistoso	8	7	6	5	4	3	2	1	Inamistoso
3	Rejeitador	1	2	3	4	5	6	7	8	Acolhedor
4	Prestativo	8	7	6	5	4	3	2	1	Frustrante
5	Sem entusiasmo	1	2	3	4	5	6	7	8	Entusiasta
6	Tenso	1	2	3	4	5	6	7	8	Calmo
7	Distante	1	2	3	4	5	6	7	8	Íntimo
8	Frio	1	2	3	4	5	6	7	8	Caloroso
9	Cooperativo	8	7	6	5	4	3	2	1	Não-cooperativo
10	Apoiador	8	7	6	5	4	3	2	1	Hostil
11	Maçante	1	2	3	4	5	6	7	8	Interessante
12	Desequilibrado	1	2	3	4	5	6	7	8	Equilibrado
13	Seguro de si	8	7	6	5	4	3	2	1	Hesitante
14	Eficiente	8	7	6	5	4	3	2	1	Ineficiente
15	Sombrio	1	2	3	4	5	6	7	8	Alegre
16	Expansivo	8	7	6	5	4	3	2	1	Reservado

Questão B

A precisão na estimativa da vida útil de produtos é fundamental para assegurar a confiabilidade e satisfação do cliente, bem como para orientar decisões estratégicas nas áreas de produção, marketing e gestão de qualidade. Neste sentido, o uso de “intervalos de confiança” revela-se como uma abordagem estatística robusta para estimar a durabilidade média de produtos, utilizando dados amostrais representativos, resultando numa estimativa importante para a verdadeira durabilidade média da população de produtos. Este procedimento é essencial para proporcionar transparência e confiabilidade nas informações sobre a durabilidade dos produtos, melhorando a gestão da qualidade, o planejamento de produção e a comunicação com os consumidores.

Considere uma situação em que um novo fabricante deseja entrar no mercado de segmento de lâmpadas. Neste nicho mercadológico, já existe concorrência de mercado, a saber:

3. **Osram, Philips e General Electric (GE):** marcas globais que têm presença significativa no mercado brasileiro, oferecendo uma variedade de soluções para iluminação em diferentes contextos, como residências, ambientes comerciais e iluminação automotiva.
4. **Tecnolite, Brilia, Startec e Golden:** marcas brasileiras que também desempenham um papel importante, proporcionando opções diversificadas para atender às necessidades específicas dos consumidores locais. A Tecnolite, por exemplo, oferece uma variedade de produtos de iluminação, incluindo lâmpadas LED e fluorescentes, enquanto a Brilia se destaca por suas soluções inteligentes e inovadoras para iluminação residencial. A Startec, com seu foco em iluminação decorativa, enquanto que a Golden oferece fabricação de lâmpadas LED.

Para tanto, o fabricante deve buscar uma estimativa para a vida média “ideal” de operação de sua linha de produtos destinados ao mercado. Neste caso, a partir de uma produção inicial de teste, obtém-se uma amostra de referência. Esta consiste em 10 lâmpadas. Após análise do padrão de qualidade feita dentro da fábrica, é verificada que a vida média de operação dessa amostra é igual a 4.000 horas, com um desvio padrão amostral de 200 horas. Assume-se que o tempo de operação das lâmpadas segue uma distribuição aproximadamente normal.

Elabore um código-fonte em linguagem “python” que indique uma estimativa para a vida média de operação para a população de lâmpadas da qual foi extraída a amostra, usando um intervalo de confiança bilateral de 95%.

Obs.: o grupo será arguido sobre o código-fonte em “python”, bem como quanto à resolução por escrito do problema.