



Centro de Educação Superior de Brasília
Centro Universitário Instituto de Educação Superior de Brasília
Curso: Ciência de Dados e Inteligência Artificial **Campus:** IESB SUL

Professora: Natalia Ribeiro de Souza Evangelista **Data:** 26/09/2024

Disciplina: HMDC065 - Amostragem Aplicada **Turma:** CIADM3B

Aluno: Guilherme Rocha Duarte

Matrícula: 2212120002

Assinatura: Guilherme Rocha Duarte

Nota: _____

Prova P1

INSTRUÇÕES

1. Leia atentamente as questões antes de respondê-las.

2. Releia atentamente as questões respondidas antes de entregar a sua avaliação. 3. Esta

avaliação é composta de 5 **QUESTÃO(ÕES) DISCURSIVA(S)** totalizando **10** ponto(s).

INSTRUÇÕES ESPECÍFICAS:

BOA PROVA!

Questão 1 (Valor 2,00 pontos) Para cada uma das amostras abaixo, informar o tipo do processo de amostragem:

P - Amostragem probabilística (aleatória)

NP - Amostragem não probabilística (não aleatória)

Para uma pesquisa sobre os hábitos dos estudantes construí uma amostra com o seguinte procedimento:

Todos os meus colegas da faculdade (tenho telefone e e-mail de todos eles). - **NP**

Fiquei na única porta de entrada da escola e a cada 12 pessoas que entravam, eu abordava uma.
- **P**

Consegui uma lista de todos os alunos das escolas com uma ordenação aleatória, e selecionei os 20 primeiros da lista. - **P**

Consegui uma lista de todos os alunos das escolas em ordem alfabética. Gerei 20 números aleatórios. Selecionei da lista de alunos aqueles que ocupavam posições equivalentes aos números aleatórios gerados. - **P**

Questão 2 (Valor 2,00 pontos) Numa cidade com 15.000 habitantes, deseja-se estimar a idade média. Qual deve ser o tamanho da amostra aleatória simples que garante um erro amostral não superior a 10? Considere que para esse cálculo foi necessário retirar uma amostra piloto onde se obteve os seguintes resultados para idade:

20, 21, 35, 39, 45, 45, 48, 56, 57, 60

$$\frac{N \cdot \sigma^2}{(N-1) \cdot D + \sigma^2}$$

$$\frac{15000 \cdot 13.33^2}{14999 \cdot D + 13.33^2}$$

$$\frac{15000 \cdot 177,68}{14999 \cdot 0,0025 + 177,68}$$

$$\frac{2665200}{215,1775}$$

$$D = \frac{B^2}{4}$$

$$D = \frac{0,1^2}{4} \rightarrow D = 0,0025$$

12386,05, ou seja, a amostra deve ser de pelo menos 12387 pessoas para um erro amostral menor que 10%.

Questão 3 (Valor 2,00 pontos) Considere a seguinte situação:

Um produtor deseja avaliar a qualidade dos cogumelos Shitakes que produz, se aceitáveis ou não aceitáveis para venda. Assim, considerando que suas mudas desta variedade de cogumelos estão numeradas de 1241 a 1640, ele pretende obter uma amostra sistemática de 50 mudas para este estudo, sendo o erro de 10%.

Com base na situação exposta, qual é a quantidade total de mudas de cogumelo Shitake que possui este produtor? Relate de que forma a amostra será retirada. Considere:

O produtor tem as mudas numeradas de 1241 a 1640. Portanto, o número total de mudas é a diferença mais 1.

$$1640 - 1241 + 1$$

400 mudas

Para realizar uma amostra sistemática de **50** mudas, será necessário selecionar 1 muda a cada 8, pois:

$$\frac{400}{50} = 8$$

O ponto de início é escolhido aleatoriamente entre 1 e 8, por exemplo, 3. A primeira muda selecionada será a de número 1243, e as demais serão escolhidas adicionando 8 a cada uma: 1251, 1259, 1267, e assim por diante, até obter as 50 mudas.

Questão 4 (Valor 2,00 pontos) Descreva uma situação para:

A. Amostragem por conveniência

Um pesquisador quer avaliar a satisfação dos clientes em uma loja de café. Em vez de escolher clientes que aparecem aleatoriamente durante o dia, ele entrevista apenas os que estão presentes na loja no momento da pesquisa, porque é mais rápido e fácil. Assim a amostragem por conveniência é utilizada, já que os participantes são selecionados pela facilidade de acesso.

B. Amostragem por quotas

Uma empresa deseja realizar uma pesquisa de mercado sobre um novo produto, e para garantir que a amostra represente diferentes grupos demográficos, estabelece quotas. Por exemplo, a empresa decide entrevistar 40% de mulheres e 60% de homens, sendo 30% entre 18-30 anos, 40% entre 31-50 anos, e 30% acima de 50 anos. Assim, os pesquisadores selecionam pessoas até que cada quota seja preenchida.

Questão 5 (Valor 2,00 pontos) No arquivo Municipio_Enem_2019_PIB_2019, retire uma amostra aleatória para melhor estimar a variável VL_ADICIONADO_BRUTO_TOT. Escolha a técnica mais adequada para esse caso e justifique.

Compare os resultados entre amostra e população.

Use SAS, R ou Python.

Resultados

Amostragem Aleatória Simples (AAS):

Diferença entre média da população e da amostra: 602,189.07

Diferença entre desvio padrão da população e da amostra: 9,477,783.04

Amostragem Estratificada (Estrat.):

Diferença entre média da população e da amostra: 3,055,797.37

Diferença entre desvio padrão da população e da amostra: 17,149,883.42

Comparação dos Resultados

1. Diferença da Média:

A diferença da média entre a amostra e a população é significativamente maior na amostragem estratificada (3,055,797.37) em comparação com a amostragem aleatória simples (602,189.07). Isso indica que a amostragem estratificada não capturou bem a média da população em relação à amostragem aleatória simples.

2. Diferença do Desvio Padrão:

A diferença do desvio padrão também é maior na amostragem estratificada (17,149,883.42) do que na amostragem aleatória simples (9,477,783.04). Isso sugere que a amostra estratificada apresenta uma maior variabilidade em relação à população, o que pode ser um indicativo de que a amostra não está bem representada em termos de variabilidade.

Com base nos resultados, a Amostragem Aleatória Simples (AAS) é a melhor opção nesse caso. As razões para essa escolha incluem:

1. Menor Erro de Estimativa:

A AAS teve diferenças menores tanto na média quanto no desvio padrão em comparação com a amostragem estratificada, o que indica que a amostra aleatória simples reflete melhor as características da população.

2. Representatividade:

A amostragem estratificada pode não ter sido eficaz neste caso, possivelmente devido à distribuição desigual dos dados em diferentes estratos. Isso pode ter levado a uma amostra que não representou adequadamente a média e a variabilidade da população total.

3. Simplicidade:

A AAS é uma técnica mais simples e direta, que pode ser mais fácil de implementar e interpretar, especialmente se a população não apresentar características estratificáveis que impactem significativamente a variável de interesse.