Nomes:

Entendendo Pesquisas de Intenção de Voto

T3 Estatística Aplicada

**INTRODUÇÃO**

Para calcular a intenção de voto de um político ou grupo de políticos, é necessária a participação de estatísticos de diversas regiões, que são responsáveis por organizar todos os dados, e por meio de tais dados, realizar uma análise estatística. Esse processo começa com a seleção de uma amostra representativa da população. Com base nessa amostra, os estatísticos desenvolvem a análise, levando em consideração margens de erro e o nível de confiança dos resultados obtidos. Que com tal informação em mãos, descobrem a intenção de voto de certos políticos.

**POPULAÇÃO E AMOSTRAGEM**

Antes de iniciar qualquer análise estatística desse tipo, é fundamental selecionar a população-alvo e, em seguida, dividi-la em uma amostra menor que a represente adequadamente. Isso ocorre porque analisar a população inteira é uma tarefa muito mais complexa. Dessa forma, a população total é vista como um grupo completo de indivíduos com características semelhantes, enquanto a amostra é um subconjunto da população escolhido especificamente para ser analisado.

Se torna essencial definir a população-alvo, pois, graças a essa técnica, o cálculo estatístico se faz possível até com populações imensas. Para isso, se faz necessário selecionar uma amostra menor que a represente de forma adequada. Isso se deve ao fato de que analisar a população inteira é uma tarefa consideravelmente mais complexa.

**TIPOS DE AMOSTRAGEM**

Para se realizar o cálculo de amostragem, é necessário primeiro escolher a técnica para amostragem na qual será utilizada. Existem dois tipos diferentes de amostras, a amostra Probabilistica, que utiliza de valores randômicos para selecionar a amostra, e a não probabilística, que consiste em formas de separar os dados para que fiquem os mais dispersos possíveis. Dentre elas, cada um possuem métodos diferentes, são eles á seguir:

**PROBABILISTICA**

**Aleatória Simples**: O processo envolve numerar a quantidade total da população, definir a porcentagem que será considerada e, em seguida, selecionar aleatoriamente os elementos da população até atingir essa porcentagem estabelecida.

Por exemplo, em um grupo de 120 pessoas, se for necessário utilizar uma amostra de 10% para fins de análise estatística, serão sorteadas 12 pessoas ao todo para compor essa amostra.

**Aleatória Estratificada**: A métrica em questão consiste em segmentar uma população em subconjuntos homogêneos, denominados estratos. A cada estrato é atribuída uma porcentagem específica, determinando a proporção de indivíduos a serem selecionados para a amostra.

Por exemplo, em um grupo de 100 pessoas, divididas em 40 homens e 60 mulheres, a aplicação de uma taxa de amostragem de 15% a cada estrato resultaria na seleção aleatória de 6 homens (15% de 40) e 9 mulheres (15% de 60) para compor a amostra final.

**Conglomerado**: Quando não é possível identificar individualmente todos os elementos que compõem uma população, um método comum é a amostragem por conglomerados. Nessa técnica, a população é dividida em grupos naturais, chamados conglomerados (como setores de uma empresa, bairros de uma cidade, etc.). Em seguida, uma amostra desses conglomerados é selecionada aleatoriamente. Por fim, todos os elementos dentro dos conglomerados sorteados são incluídos na amostra final.

Exemplo: Imagine uma pesquisa em uma grande universidade. Seria inviável listar todos os alunos individualmente. Nesse caso, a universidade poderia ser dividida em conglomerados, como os cursos de graduação. Em seguida, alguns cursos seriam sorteados aleatoriamente. Todos os alunos matriculados nos cursos sorteados fariam parte da amostra..

**Amostragem Sistematica**: Essa técnica de amostragem, conhecida como amostragem sistemática, exige que os elementos da população estejam ordenados de forma aleatória. Após a ordenação, calcula-se o intervalo de amostragem, dividindo o tamanho da população pelo tamanho da amostra desejada. Em seguida, seleciona-se aleatoriamente um número inicial dentro do primeiro intervalo. A partir desse número, os demais elementos da amostra são selecionados a intervalos regulares, correspondentes ao intervalo de amostragem calculado.

**Exemplo:** Em uma empresa com 100 funcionários, numerados de 1 a 100, deseja-se selecionar uma amostra de 25 funcionários. O intervalo de amostragem é 100/25 = 4. Escolhe-se aleatoriamente um número entre 1 e 4, por exemplo, 3. A amostra será composta pelos funcionários de número 3, 7, 11, 15, e assim por diante, até completar 25 funcionários.

**NÃO PROBABILÍSTICA**

Nem sempre é viável ou necessário utilizar métodos de amostragem probabilística. Quando o conhecimento sobre a população é limitado e as probabilidades de seleção dos elementos são desconhecidas, a amostragem não probabilística se torna uma alternativa. Essa abordagem é utilizada quando não há como calcular o erro amostral ou determinar com precisão o tamanho da população, tornando inviável a aplicação de técnicas probabilísticas. Alguns exemplos são:

**Acidental**: A amostragem por conveniência, também conhecida como acidental, baseia-se na seleção de indivíduos que estão mais acessíveis, como aqueles encontrados em locais públicos. Embora seja um método rápido e fácil de implementar, a representatividade dos resultados é questionável.

Por exemplo, realizar uma pesquisa de opinião em um centro urbano pode levar a uma amostra composta majoritariamente por pessoas que frequentam esses locais, não refletindo a diversidade da população como um todo.

**Intencional**: A amostragem por julgamento, também conhecida como intencional, ocorre quando o pesquisador seleciona os elementos da amostra com base em seu conhecimento prévio sobre o assunto e o objetivo da pesquisa. Ele escolhe indivíduos que acredita serem os mais informativos e relevantes para responder às perguntas de pesquisa.

Exemplo: Para avaliar a eficácia de um novo método de ensino, um pesquisador pode optar por entrevistar apenas professores com vasta experiência e que já utilizaram diferentes metodologias em sala de aula..

**AMOSTRAGEM X POPULAÇÃO**

Apesar da complexidade dos cálculos envolvidos em algumas técnicas de amostragem, a análise de toda uma população é ainda mais desafiadora. Além dos custos elevados e do tempo necessário para coletar dados de todos os indivíduos, existe a possibilidade de que parte da população se recuse a participar da pesquisa. Dessa forma, a amostragem, por sua praticidade e menor custo, torna-se uma ferramenta indispensável para a realização de estudos científicos e pesquisas de mercado.

**O que é margem de erro?**

Quando um instituto de pesquisa realiza um estudo, muitas vezes não é viável consultar toda a população-alvo, seja por limitações de tempo, orçamento ou acesso. Por exemplo, em uma pesquisa sobre a preferência por um novo produto em uma cidade, não seria possível consultar todos os moradores. Assim, é escolhida uma amostra representativa da população.

Essa amostra precisa ser selecionada de forma aleatória, garantindo que ela reflita as diversas características da população, como idade, gênero, renda e localização. Dessa forma, busca-se garantir que os resultados sejam representativos da comunidade como um todo.

Depois de coletar os dados, como a pesquisa foi feita com uma amostra, erros amostrais podem ocorrer. Isso significa que a amostra pode não representar exatamente a população total, e por isso usamos a margem de erro. A margem de erro nos diz o quanto os resultados podem variar. Por exemplo, em uma pesquisa sobre satisfação com um serviço, se a margem de erro é de 3%, significa que os resultados podem ser até 3% maiores ou menores do que os observados.

A margem de erro, portanto, é uma estimativa que nos ajuda a avaliar a precisão dos resultados. Se a margem de erro for muito grande, isso indica uma maior incerteza sobre os dados da pesquisa; quanto menor a margem de erro, mais confiáveis serão os resultados.

**Como ler a margem de erro de uma pesquisa?**

A margem de erro permite descobrir o intervalo de confiança, ou seja, prever os possíveis valores dos resultados quando comparamos a amostra com a população total. Dada uma margem de erro de "m%" e um resultado de "p%" em uma pesquisa, o intervalo de confiança será dado por:

**(p ± m)**

Por exemplo, se uma pesquisa sobre a satisfação dos clientes de uma loja encontrou que 40% estão satisfeitos, e a margem de erro é de 3%, isso significa que, na realidade, entre 37% e 43% dos clientes estão satisfeitos.

**Margem de erro e intervalo de confiança**

O intervalo de confiança é calculado com base na margem de erro, permitindo uma análise mais precisa dos resultados. Veja um exemplo:

Imagine uma pesquisa de mercado para uma nova marca de refrigerante, em que três sabores estão sendo testados: A, B e C. A intenção de compra dos consumidores para o sabor A foi de 45%, e a margem de erro é de 3%. Isso significa que o intervalo de confiança para o sabor A é de **45% ± 3%**, ou seja, entre 42% e 48%.

Para o sabor B, que obteve 30% de intenção de compra, o intervalo será de **30% ± 3%**, ou seja, entre 27% e 33%.

E para o sabor C, com 25% de intenção de compra, o intervalo será de **25% ± 3%**, ou seja, entre 22% e 28%.

Se os intervalos de confiança de dois sabores se sobrepõem, como nos casos dos sabores A e B, é possível que, dentro da margem de erro, eles tenham a mesma intenção de compra na realidade.

**Fórmula da Margem de Erro**

A fórmula básica para calcular a margem de erro é a seguinte:

Onde:

* **ME**: Margem de erro.
* **Z**: Valor crítico associado ao nível de confiança (obtido da tabela normal padrão).
* **σ**: Desvio padrão da população (se desconhecido, pode-se usar o desvio padrão da amostra ou uma estimativa).
* **n**: Tamanho da amostra.

**Passo a Passo:**

1. **Determine o nível de confiança (C)**: O nível de confiança indica a probabilidade de que o intervalo contenha o parâmetro populacional. Os níveis mais comuns e seus valores críticos (Z) são:
   * 90% → Z ≈ 1.645
   * 95% → Z ≈ 1.96
   * 99% → Z ≈ 2.576
2. **Obtenha o desvio padrão (σ)**: Se o desvio padrão da população não for conhecido, pode-se usar o desvio padrão da amostra ou, em pesquisas de proporções (sim/não), estimar a proporção de sucesso "p" e falha "q = 1 - p".
3. **Calcule a margem de erro**: Substitua os valores na fórmula.

**Exemplo:**

Suponha que você está realizando uma pesquisa com 500 pessoas e deseja calcular a margem de erro com 95% de confiança, assumindo uma proporção estimada de sucesso "p" de 0,6 (ou 60%).

1. **Nível de confiança**: 95% → **Z** = 1.96.
2. **Proporção estimada**:
   * p = 0.6, logo q = 1 – p = 0.4.
   * O desvio padrão estimado será
3. **Tamanho da amostra**: 500.
4. **Substituindo na fórmula**:

ME = 1.96 x 1.96 x 0.022 0.043

Portanto, a margem de erro é **4,3%**.

Isso significa que, com 95% de confiança, a verdadeira proporção da população está dentro de 4,3% acima ou abaixo do valor estimado pela amostra.

**NIVEL DE CONFIANÇA**

O nível de confiança é um conceito fundamental em estatística que define a probabilidade de que uma estimativa, baseada em uma amostra, reflete com precisão o parâmetro verdadeiro da população. Em outras palavras, é a segurança de que o intervalo de confiança contém o valor real do parâmetro que está sendo estimado. Os níveis de confiança mais comuns são 90%, 95% e 99%. Um nível de confiança de 95%, por exemplo, significa que se repetirmos a amostragem 100 vezes, em 95 dessas amostras o intervalo de confiança calculará o valor verdadeiro do parâmetro da população.

**Exemplo de Fórmula**

A fórmula para calcular um intervalo de confiança para uma média é:

Onde:

* é a média amostral.
* é o valor crítico associado ao nível de confiança (obtido da distribuição normal padrão).
* σ é o desvio padrão da população.
* é o tamanho da amostra.

Por exemplo, se uma amostra de 100 alunos tem uma média de pontuação de 80 e um desvio padrão de 10, com um nível de confiança de 95% (Z = 1.96), o intervalo de confiança será:

Ou seja, com 95% de confiança, a pontuação média verdadeira dos alunos está entre 78.04 e 81.96

Porém, vale citar que tanto o nível de confiança quanto a margem de erro são componentes essenciais para a interpretação de resultados de pesquisa, mas desempenham papéis diferentes:

**Nível de Confiança:**

* Refere-se à probabilidade de que o intervalo de confiança contenha o valor verdadeiro do parâmetro populacional.
* É expresso como uma porcentagem, como 90%, 95% ou 99%.
* Quanto maior o nível de confiança, maior será o intervalo de confiança, pois há mais "certeza" de que o intervalo incluirá o valor verdadeiro.
* Está associado ao valor crítico (Z), como mostrado no exemplo acima: 1.645 para 90%, 1.96 para 95% e 2.576 para 99%.

**Margem de Erro:**

* Indica a quantidade de erro que se espera na estimativa, dada uma amostra.
* É representada como a quantidade de variação permitida em torno do valor estimado.
* Por exemplo, se uma pesquisa eleitoral mostra que um candidato tem 45% de intenção de voto com uma margem de erro de 2%, o valor verdadeiro pode estar entre 43% e 47%.
* A margem de erro depende do tamanho da amostra e do desvio padrão. Quanto maior a amostra, menor a margem de erro.

Como é possível ver anteriormente, a margem de erro é calculada pela seguinte equação: , onde é o valor crítico do nível de confiança. Assim, a margem de erro é diretamente impactada pelo nível de confiança escolhido, sendo que um aumento no nível de confiança aumenta o valor de 𝑍, o que, por sua vez, aumenta a margem de erro

**ANÁLISE**

Neste exemplo da pesquisa de intenção de votos, foi considerado um nível de confiança de 95%, padrão em levantamentos de opinião pública. A margem de erro foi calculada para cada candidato com base no número de entrevistados em cada segmento (total, masculino e feminino).

**Dados da Pesquisa:**

**Amostra total (n):** 59 entrevistados.

**Intenção de voto para o Candidato A:** 53%.

**Nível de confiança:** 95% (valor Z correspondente: 1,96).

**Cálculo da Margem de Erro:**

A margem de erro nos permite entender a precisão dos resultados da pesquisa. Em uma pesquisa eleitoral, é comum utilizarmos um nível de confiança de 95%, o que significa que, se a pesquisa fosse repetida várias vezes, os resultados ficariam dentro da margem de erro em 95% das vezes.

A fórmula para calcular a margem de erro é:

**Onde:**

* **ME** é a margem de erro
* **Z** é o valor associado ao nível de confiança (1,96 para 95% de confiança)
* **p** é a proporção observada (intenção de votos para o candidato, convertida em decimal)
* **n** é o tamanho da amostra (número de entrevistados)

**Aplicando os dados ao Cálculo:**

* Proporção de votos para o Candidato A: **p = 0,53** (ou 53%)
* Tamanho da amostra: **n = 59**
* Nível de confiança de 95%: **Z = 1,96**

**Substituindo esses valores na fórmula:**

**ANÁLISE E INTERPRETAÇÃO DOS RESULTADOS**

A pesquisa de intenção de votos para os Candidatos A e B apresentou os seguintes resultados, com base em uma amostra total de 59 entrevistados, distribuídos entre grupos masculino e feminino. O nível de confiança adotado é de 95%, o que nos permite calcular a margem de erro e interpretar os dados com maior precisão.

**Resultados Gerais (Total):**

**Candidato A:** 53% de intenção de votos.

**Candidato B:** 41% de intenção de votos.

**Tamanho da amostra (n):** 59 entrevistados.

**Margem de Erro para o Candidato A:**

Ao calcular a margem de erro com os dados da pesquisa para o Candidato A (proporção de 53%), utilizando um nível de confiança de 95%, encontramos uma margem de erro de 12,74%. Isso significa que podemos afirmar, com 95% de confiança, que a verdadeira proporção de eleitores que pretendem votar no Candidato A está entre 40,26% e 65,74%.

Ou seja, embora a pesquisa aponte uma intenção de voto de 53%, há uma variação possível devido à margem de erro. O valor real de intenção de voto para o Candidato A pode ser tanto maior quanto menor, dentro do intervalo calculado. Esse intervalo é importante para entender a imprecisão inerente à pesquisa.

**Margem de Erro para o Candidato B:**

Embora o cálculo detalhado para o Candidato B não tenha sido feito ainda, podemos aplicar o mesmo raciocínio. Supondo uma intenção de voto de 41% para o Candidato B e usando o mesmo tamanho de amostra, o intervalo de confiança poderia fornecer um intervalo para a verdadeira intenção de voto. O cálculo seguiria a mesma lógica aplicada ao Candidato A, com um intervalo baseado na margem de erro correspondente.

**Resultados por Gênero:**

**Masculino**

**Candidato A:** 57% de intenção de votos.

**Candidato B:** 43% de intenção de votos.

**Tamanho da amostra (Masculino):** 30 entrevistados.

Para o grupo masculino, o Candidato A teve uma intenção de voto de 57%, com um número de entrevistados menor (30). A margem de erro para essa amostra pode ser maior do que a margem geral, já que a precisão tende a diminuir quando o tamanho da amostra é reduzido.

**Feminino**

**Candidato A:** 48% de intenção de votos.

**Candidato B:** 52% de intenção de votos.

**Tamanho da amostra (Feminino):** 29 entrevistados.

Entre as mulheres, a pesquisa mostra uma divisão mais equilibrada, com o Candidato B à frente com 52% e o Candidato A com 48%. Com uma amostra semelhante ao grupo masculino, a margem de erro provavelmente será parecida com a do grupo masculino.