

Boa tarde amigos e colegas.

Nesta etapa vamos explorar :

- **O que é uma Amostragem**



Sendo direto e reto, uma amostra é uma pequena parte de uma população. No próximo domingo, vamos poder entender bem isso quando as TV's e os institutos já definirem quem vai ganhar a eleição. Isso ocorre não porque antecipadamente eles fazem a contagem dos votos e sim porque eles possuem diversas amostras, de diversas áreas e este grande conjunto de amostra aplicando-se uma margem de erro para mais e para menos, define o resultado final. Em aprendizado de máquina, trabalhar com amostras é fundamental para se chegar a uma precisão ou ter uma acurácia melhor sobre as informações.

Hoje o volume de dados é tão grande que é impossível um software aprender com todos estes dados ou estatisticamente falando com a população.

- **Erros de amostragem e erros não aleatórios**

Como falei anteriormente, a quantidade de amostras é fundamental para a precisão e a acurácia de uma análise ou de uma decisão não esquecendo aí uma margem de erro para mais e menos.

Se estamos fazendo um estudo sobre um determinado assunto e a metodologia da amostra não for coerente entre os pesquisadores, podemos ter erros de amostragem que podem diretamente inferir no resultado.

Existe também a questão do pesquisador de certa forma conduzir o resultado gerando o que é chamado de "amostra intencional", tornando-a assim tendenciosa.

Segundo Sonia Vieira, "Tendência é a divergência consistente, persistente, da estatística de uma amostra em relação ao parâmetro que se quer estimar", eu gostei muito desta frase.

Uma questão importante com relação ao levantamento das amostras, também está relacionado com a abordagem do pesquisador em relação a pessoa que vai gerar a informação. Muitas pessoas informam respostas erradas por conveniência, por ignorância, por vergonha, enfim o ser humano é uma máquina complexa e bem divergente. Eu mesmo quando participo de uma pesquisa, meu maior objetivo é o de me livrar o mais rápido possível daquilo rs.

- **Organização de dados**

Após a pesquisa ser feita, deve-se verificar se a amostra que foi gerada é qualitativa ou quantitativa.

Geralmente em uma pesquisa, pergunta-se nome, endereço, telefone e outras coisas mais, mas o objetivo real da pesquisa por ser o de saber por exemplo o número de pessoas (homens e mulheres) que falam mais de um idioma não levando-se em conta o idioma nativo.

- **Dados nominais e ordinais**

Depois da apuração dos dados, temos que organiza-los em tabelas chamadas de distribuição de frequência.

No caso de dados nominais, simplesmente se cria uma tabela com duas linhas e duas colunas com os valores que foram apurados, como podemos ver abaixo.

Gênero	Frequência
Feminino	75
Masculino	60
Total	135

Com dados ordinais, o estudo já é um pouco mais complexo porque envolve interesse específico em um determinado assunto, que no nosso exemplo está relacionado com o conhecimento de línguas.

Aqui, poderíamos definir as seguintes categorias:

- conheço uma linguagem
- conheço duas linguagens
- conheço mais de duas linguagens

E para facilitar a apuração das categorias, numera-se-as como abaixo:

- 0 - conheço uma linguagem
- 1 - conheço duas linguagens
- 2 - conheço mais de duas linguagens

Categoria	Frequência
0 - Conheço uma linguagem	200
1 - Conheço duas linguagens	60
2 - Conheço mais de duas linguagens	10
Total	270

• Frequências relative

A frequência relativa é definida como a razão entre a frequência absoluta e o número total de observações como podemos ver nas duas tabelas abaixo.

Gênero	Frequência	Frequência relativa (%)
Feminino	75	$75/135 = 0.56$
Masculino	60	$60/135 = 0.44$
Total	135	100

Categoria	Frequência	Frequência relativa (%)
0 - Conheço uma linguagem	200	$200/270=0.75$
1 - Conheço duas linguagens	60	$60/270=0.22$
2 - Conheço mais de duas linguagens	10	$10/270=0.03$
Total	270	100

Vou terminar por aqui, hoje não estou no melhor dos meus dias.
 Espero de coração ter ajudado alguém que por algum motivo esteja estudando estes assuntos.

Abraços

Pequeno

Conhecimento gera conhecimento! Projeto Spartacus.