**História da Estatística**

**Baseado na Aula do Professor Wesley Santos – SAS**

“Correlação é uma chance probabilística de algo que está associado às características estudadas. ”

Estatística vem de Estado. Em Roma o conceito de escravo era o indivíduo capturado pela guerra. O algo para fazer cálculos era denominado de estatística. Hoje em dia a definição mais utilizada é a aproximação do conceito como ferramenta para uma tomada de decisão.

Em 1749 o termo fora empregado dentro de um sentido científico. A necessidade de nomear deterministicamente um fenômeno elevou o conceito de Estatística para uma Etimologia que indicava aplicações determinísticas, como a formulação da distância entre a terra e a lua, representada como uma certeza que possui um erro quando expressa em termos de formulação matemática. Laplace observou que determinados fenômenos sofriam de variação não sendo possível a precisão determinística.

Esse grau de incerteza levou os pesquisadores e cientistas do século XVII a formularem as primeiras teorias a respeito de fenômenos que agregavam varrições, ou incertezas, produzidos por tais variações que indicavam os estados transitórios em que o fenômeno era conhecido.

Então com o avanço dos estudos observou-se que para garantir precisão não era necessário percorrer 100% das varrições de um fenômeno, mas sim, compreender uma parcela dos dados que é absolutamente capaz de representar o todo complexo.

Assim, como um exame de sangue que para determinar o conteúdo sanguíneo não é necessário extrair todo o sangue de um indivíduo, e sim, sintetizar algumas gotas suficientes para a contagem das substâncias e através de estimação inferir, ou seja, expandir a conclusão para o volume aproximado de sangue que o indivíduo possui.

Ronald Fisher (1890 – 1962) foi considerado por muitos o pai da Estatística. Fisher foi convidado para gerar uma estatística junto com os aborígenes australianos, a fim de que validasse uma invasão da Inglaterra no continente Australiano. O objetivo desta pesquisa era determinar que os índios australianos não seriam considerados seres humanos que justificasse uma invasão, extermínio e ocupação.

Fisher medindo a cabeça dos aborígenes com uma fita métrica e dos ingleses chegou à conclusão que os aborígenes australianos tinham a mesma variação craniana e que, portanto, estavam no mesmo patamar evolutivo do homem civilizado, não legitimando assim, a invasão para efeito de extinção dos aborígenes e consequente ocupação do solo.

William Bosset (1876 – 1937) trabalhava na Cervejaria Guiness. Ele observou que não existia muito critério para a mistura de ingredientes para a produção de cerveja. Muitas descobertas dele era quando o seu ato de provar o conteúdo, como cervejeiro, o deixava levemente embriagado, pois ele nesta época atuava como provador de cervejas.

Ao pegar uma placa de leveduras resolveu acompanhar o processo de fermentação e trabalhou plotando num papel dia após dia diferentes tipos de leveduras na forma gráfica. Isto o levou a pensar em construir uma fórumula matemática, que permitisse ele induzir quantidades para compor uma mistura ideal para uma produção esperada de álcool.

Ele foi um dos primeiros a descobrir uma aplicação para frequências ao qual foi denominada de Poisson. A distribuição Poisson foi desenvolvida por Karl Pearson (1857 – 1936) e a ideia de Willian Gosset foi possível viabilizar os primeiros estudos estatísticos da época.

Não podendo repassar a fórmula de cerveja, o cervejeiro Willian Gosset renomeou o teste para “t de Student” a fim de que a equação da cerveja fosse preservada no anonimato. Os resultados e avanços teóricos com este sistema foi a base para a construção do Teste de Hipótese.

Ronald Fisher (1890 – 1962) desenvolveu muitos trabalhos relacionados à agricultura, alternava quantidades de adubo em plantações e de acordo com as varrições de peso dos produtos inferia qual o percentual de adubo era melhor aproveitado pelas plantas e elaborou diversos testes agrícolas e de campo.

Entre os principais expoentes principalmente no século XX contribuições para o avanço da ciência podem ser listados conforme segue: Egon Pearson, Jerzy Neyman, Geortrude Cox, Frank Yates, John Tukey, George Box, David Cox e Jerome Friedman.

A definição que melhor adequa o conceito de Estatística para a atualidade é:

“**Conjunto de Métodos espacialmente apropriados à coleta, à apresentação, à análise e à interpretação de dados de observação, tendo como objetivo a compreensão de uma realidade específica para a tomada de decisão.** ”

Você precisa tomar uma sopa inteira para saber que ela está quente? Quando eu analiso uma planilha com 100 linhas não existe dúvida sobre as características amostradas apresentadas como conteúdo de dados que fora apresentado. Agora quando o universo é maior que os 100 objetos coletados hão que se pensar em um intervalo de confiança onde posso prever com exatidão 100% dos casos contidos para expressar uma população, embora a precisão não possa mais ser garantida.

Como por exemplo instituir que as pessoas de um centro educacional variam entre 1,60 m de altura até 1,85 m de altura. O tipo de informação desenvolvida deve ser discutido quanto à relevância que se pretende para a garantia de um fator de erro que não comprometa a necessidade de informar. Como por exemplo usar de um conteúdo estatístico para afirmar que uma fórmula não oferece prejuízos ao consumo por parte de um indivíduo que venha a ingerir um medicamento.

O Instrutor Wesley Santos sita como obra necessária a leitura que aborda mais profundamente este tema o livro: Uma Senhora Toma Chá – como a estatística revolucionou a ciência no século XX de David Salsburg.

Dados são os valores observados em si que podem conter várias informações de um mesmo ente observável.

Variável é uma unidade conceitual que surge como um atributo de algo observado que possa ser desmembrável como uma característica observável.

Estatística descritiva é a capacidade de um observador descrever os dados.

Em inferência é possível pegar subconjunto de dados e ampliar a descoberta para toda a população estudada.

Qual a maior chance que uma pessoa terá ao jogar na loteria os números 01 – 02 – 03 – 04 – 05 – 06 ou outro bilhete 09 – 53 – 21 – 38 – 06 – 31?

Existem dos tipos de variáveis: quantitativas e qualitativas. Qualidade está relacionada a atributos e quantitativos a valores. As variáveis qualitativas mse dividem em contínuas e categóricas. As variáveis qualitativas ordinais são utilizadas para nomes e rótulos, como por exemplo (M) Masculino, e, (F) feminino. Ou nome de cidades, por exemplo.

Classe social é um rótulo que traz um senso de ordenação, o portanto ela é uma variável qualitativa ordinal. Outro exemplo é grau de instrução.

As variáveis quantitativas contínuas vêm de medições, como por exemplo: peso, altura e pressão arterial.

Agora variáveis que você não mede, mas que exige contagem são classificadas como variáveis quantitativas categóricas, por exemplo: número de filhos, número de empregos, número de livros, número de televisores, ...

População são elementos na forma de um conjunto completo de interesse no estudo.

Uma característica reduzida da população de interesse é uma amostra.

A estatística descritiva é um resumo da observação em que seja possível extrair informações sobre a massa de dados.

Agora quando se usa uma amostra para falar algo sobre a população se está fazendo uma inferência sobre a população.

Quando se falar em população estará falando em parâmetros e quando se fala de amostra estará fazendo referência a estatísticas. Sobre este conteúdo reduzido observado.

A média de um conjunto de dados é o ponto central que distam os registros observados numa relação que se somam todos os dados e se divide pela quantidade de elementos. Portanto a média é muito influenciada por valores extremos.

Mediana é um valor que divide sua massa de dados em duas partes que concentram os dados em 2 partes iguais. A mediana exige que os dados estejam ordenados e não é influenciada pela variabilidade das informações.

Como a média é influenciada pelos valores extremos conforme a necessidade da aplicação geralmente é recomendável o uso da mediana para expressar um comportamento da população observada ou fazer um tratamento onde os valores extremos passam a ser desprezados.

Enquanto que a Moda como conceito estatístico é a media observável mais frequente.

Os percentis dividem ou segmentam um rol (relação ordenada de dados) numa relação de posição em que as informações ordenadas distam umas das outras num eixo de dados.

A variância representa a distância de cada ponto de uma observação em relação a sua média, dando uma ideia de quão dispersas estão as informações diante da medida central de informação.

A amplitude é a distância entre o maior e menor valor de um rol de dados.

O intervalo Interquartílico é a distância referente entre os 75% das observações em um rol de dados e as 25% das observações inicias.

O Box plot é uma caixa de informações que traz alguns aspectos sobre a distribuição estudada: primeiro quartil, mediana, média, terceiro quartil, pontos extremos.

Covariância mede a variabilidade de duas ou mais variáveis. Ela mede o quanto duas variáveis estão correlacionadas.

Correlação é uma medida de variabilidade entre duas variáveis (Covariância) influenciada pelo efeito do produto do desvio padrão das variáveis estudadas. Ela é uma medida de associação linear que não reflete necessariamente uma relação de causa entre elas. Gera uma medida de dependência entre as variáveis estudadas.

Quando se fala em distribuição de probabilidade, há que se pensar na principal ou mais conhecida delas: a distribuição normal, onde não importa o número de repetições de um experimento, quando elas ocorrem várias vezes o experimento tende para uma distribuição normal indiferentemente do tipo de distribuição correspondente à massa de dados.

Compilação de Max Diniz Cruzeiro

LenderBook Company

[www.lenderBook.com](http://www.lenderBook.com)