# Introdução a Estatística

* **Estatística** é a ciência que se ocupa de coletar, organizar, analisar e interpretar dados a fim de tomar decisões.
* Conjuntos de dados:
	+ **População**: conjunto de todos os resultados.
	+ **Amostra**: subconjunto da população
* **Parâmetro** é uma descrição numérica de uma característica população.
* **Estatística** é uma descrição numérica de uma característica de uma amostra.
* **Ramos Estatística:**
	+ **Descritiva**: é o ramo que trata da organização e da apresentação dos dados.
	+ **Inferencial**: é o ramo que trata de inferir sobre uma população a partir de uma amostra. A probabilidade é uma ferramenta básica.
* **Dados qualitativos**: consistem em atributos, classificações ou registros não numéricos. (podem ser nominais e ordinais)
* **Dados quantitativos**: consistem em medidas ou contagens numéricas. (Podem ser discretos ou contínuos).
* Classificação de dados
	+ **Nível intervalar**: é como o nível ordinal mas a diferença entre valores é significativa. Não existe ponto inicial natural (ex: temperatura do corpo)
	+ **Nível razão**: É o nível intervalar com ponto inicial natural. (Ex: peso, preços)
* Planejamento de experimentos
	1. Identifique a(s) variável(is) de interesse e a população que forem objetos de estudo.
	2. Desenvolva um plano de trabalho para coleta de dados. Se for usar uma amostra, assegure-se de que ela é representativa da população.
	3. Colete os dados.
	4. Descreva os dados fazendo o uso de técnicas descritivas.
	5. Interprete os dados e tome decisões acerca da população usando inferência estatística.
* Coleta de dados:
	+ Censo: contagem ou medição de toda a população.
	+ Amostra: uma contagem ou medição de parte de uma população. As estatísticas produzidas na amostra são usadas para fazer predições sobre vários parâmetros da população.
	+ Simulação: uso de um modelo matemático ou físico para reproduzir as condições de uma situação ou um de processo.
	+ Experimento: é aplicado um tratamento a uma parte da população e são observadas as respostas.
* Técnicas de amostragem
	+ **Aleatória:** é aquela na qual todos os elementos da população têm chances iguais de serem selecionados.
	+ **Estratificada:** Quando for importante que uma amostra tenha elementos de cada segmento da população
	+ **Agrupamento:** Quando a população apresenta ocorrência natural de subgrupos, cada um deles com características similares. É selecionada uma amostra de grupos e todos os elementos dos grupos são usados no estudo.
	+ **Sistemática:** Os elementos da população são ordenados de acordo com uma numeração estabelecida e o número inicial é selecionado aleatoriamente e depois os demais elementos da amostra são selecionados segundo intervalos regulares.

# Estimação

* Estuda como prever parâmetros populacionais a partir de amostras.
* Tipos estimativas:
	+ Pontual: o valor mais provável. As estimativas raramente coincidem com os valores populacionais.
	+ Intervalares: são acompanhadas de um grau de confiança. É a mais adequada pois a probabilidade (grau de confiança)associada a uma estimativa pontual é zero.
* Construindo estimativa intervalar das médias usando distribuição normal (*variância conhecida*)
	+ **Nível de confiança** (1-α): éa probabilidade de que um intervalo estimado contenha o parâmetro populacional.
	+ **Nível de significância** (α): éa probabilidade de que um intervalo estimado NÃO contenha o parâmetro populacional.
	+ **Erro máximo da estimativa** (tolerância): é a maior distância possível entre a estimativa pontual e o valor do parâmetro.
	 
	+ Intervalo de confiança 
* Construindo estimativa intervalar das médias usando distribuição t (*variância desconhecida*)
	+ Utilizamos o mesmo procedimento utilizado com variância conhecida, substituindo zde alpha por t de alpha com n-1 graus de liberdade.
* Construindo estimativas das proporções
	+ Suposições:
		- Aleatoriedade da amostra
		- Condições de um experimento binomial são satisfeitas (np^>5 e nq^>5)
		- Aproximação normal pode ser aplicada
	+ A estimativa pontual para p, proporção populacional de sucessos, é dada pela proporção de sucessos em uma amostra.
	 onde x é o número de sucessos em uma amostra
	+ **Erro máximo da estimativa** (tolerância)
	 
	+ **Intervalo de confiança**
* Construindo estimativas das variâncias e desvios padrões
	+ A estimativa pontual para σ2 é s2 onde
	
	+ A distribuição qui-quadrado é usada para construir estimativas intervalares para a variância e o desvio padrão.
	+ Suposições
		- Amostras aleatórias
		- População normalmente distribuída
	+ Intervalo de confiança para a variância
	
	+ Intervalo de confiança para o desvio padrão. Basta calcular a raiz dos limites da variância.
	+ Cálculo do tamanho da amostra é mais complexo. Sugestão utilizar tabela. (qual a tabela?)

# Teste de Hipóteses Paramétricos (uma amostra)

* Fundamentos de teste de hipóteses
	+ Teste de hipótese é um procedimento da estatística amostral para testar uma *alegação sobre um valor de um parâmetro populacional (hipótese estatística)*
	+ Um par de hipóteses deve ser estabelecido: nula e alternativa (que é o complemento da hipótese nula)
* Estabelecendo hipóteses. A hipótese levantada (na questão) é a hipótese nula.
* Tipos de erros
	+ Em todo teste de hipótese é assumido que a condição de igualdade na hipótese nula é verdadeira. Então deve-se tomar duas decisões: Rejeitar ou não rejeitar a hipótese nula
	+ Um erro tipo I ocorre quando rejeitamos H0 quando esta é verdadeira. A probabilidade máxima para que ocorra este tipo de erro e chamada nível de significância. Denotado por alpha.
	+ Recomenda-seque seja usado a declaração “não rejeitar H0” em vez de aceitar H0.
* Teste de hipótese para média com σ conhecido
	+ Suposições
		- Amostra aleatória
		- População normal ou n>30
	+ Procedimentos
		1. Identifique o parâmetro de interesse no problema. Neste caso é μ.
		2. Formule a hipótese nula (H0)
		3. Formule uma hipótese alternativa apropriada (Ha)
		4. Defina o nível de significância
		5. Estabeleça a estatística usada usando a distribuição normal
		6. Estabeleça a região de rejeição usando o nível de significância
		7. Coletar os dados amostrais e calculara estatística do teste
		8. Decida se H0deve ou não ser rejeitada e transponha esta conclusão para o contexto do problema
* Teste de hipótese para média com σ desconhecido
	+ Suposições
		- Amostra aleatória
		- População normal ou σ>30
	+ Procedimento
		- Utilizar o mesmo procedimento para σ conhecido traçando a distribuição normal pela t-student com n-1 graus de liberdade.
* Teste de hipótese para uma proporção
	+ Suposições
		- Amostra aleatória
		- Condições satisfeitas para distribuição binomial
		- Condições satisfeitas para usar aproximação normal.
	+ Estatística do teste
	
* Teste de hipótese para uma variância ou desvio padrão
	+ Suposições
		- Amostra aleatória
		- População normal
	+ Estatística do teste
	onde s2 variância amostral σ02 variância populacional hip. Nula

# Testes de hipóteses paramétricos (duas amostras)

* Teste de hipóteses entre médias σ1e σ2 conhecidos
	+ Suposições
		- Amostras aleatórias e independentes
		- Pelo menos uma das condições é satisfeita: as duas populações são normais ou n1>30 e n2 >30
	+ Realizamos o teste para μ1-μ2
	
* Teste de hipóteses entre médias σ1e σ2 desconhecidos
	+ Usar distribuição t com os seguintes graus de liberdade
		- n1+n2-2 (supondo igualdade de variâncias populacionais entre as amostras)
		- min{n1-1, n2-1} (caso contrário)
	+ Suposições
		- Amostras aleatórias e independentes
		- Pelo menos uma das condições é satisfeita: as duas populações são normais ou n1>30 e n2 >30
	+ Estatística do teste
	
	+ Se variâncias populacionais iguais
	 onde 
	+ Se variâncias populacionais diferentes
	
	+ Raramente podemos considerar igualdade de variância populacional entre duas amostras
* Teste de hipóteses para médias com amostras dependentes
	+ Suposições
		- Amostra aleatória
		- Amostras dependentes (emparelhadas)
		- População normal ou n>30
* Teste de hipótese entre proporções
	+ Amostras aleatórias e independentes
	+ Condições satisfeitas para aproximação pela normal
	+ P^1-P^2 segue uma distribuição normal com média p1-p2 e desvio 
	+ Estatística de teste (como p1 e p2 não são conhecidos utiliza-se p- 
* Teste de hipótese para diferença entre duas variâncias
	+ Suposições
		- Populações independentes
		- Populações normais
	+ Estatística de teste
	onde s12 é a maior das duas variâncias