Noções de Estatística Básica

Introdução a Inferência Estatística

* População x Amostra
* Estatísticas x Parâmetros
* Distribuições Amostrais
* Estimação
* Estimadores Pontuais
* Propriedades de estimadores

1. População x Amostra

Vamos retomar brevemente tudo o que aprendemos até agora. Aprendemos que a primeira parte de uma análise estatística tem início com uma coleta de dados. Em seguida, os dados devem ser organizados e deve-se analisar as variáveis do estudo. Depois, estudamos uma variedade de métodos que nos auxiliam a resumir as informações obtidas na nossa coleta de dados. Essa parte é a parte de análise descritiva dos dados. Depois nos distanciamos da nossa análise estatística e estudamos conceitos em probabilidade e vimos algumas distribuições teóricas de probabilidade. Por que fizemos tudo isso? Não podemos nos esquecer de um dos objetivos principais de uma análise estatística: inferir informações a respeito de uma população. Quando coletamos dados, estamos na verdade estudando apenas um pedacinho da população. Esse pedacinho se chama Amostra. Como podemos deduzir informações para uma população olhando apenas para uma amostra? Através da análise de inferência estatística.



Definição: População é o conjunto de todos os elementos sob investigação. Amostra é qualquer subconjunto da população.

1. Estatísticas x Parâmetros

Obtida uma amostra, muitas vezes desejamos usá-la para estudar alguma característica específica da população. Por exemplo, se quiséssemos saber a média das alturas da população brasileira, poderíamos coletar uma amostra e começar por estudar a média da nossa amostra. É claro que sabemos que a média da nossa amostra nem sempre vai corresponder ao valor da média populacional, mas é um começo!

Definição: Uma estatística é uma característica da amostra.

Uma estatística é uma função das observações

As estatísticas mais comuns são:

E quanto às características da população?

Definição: Um parâmetro é uma medida usada para descrever uma característica da população.

Abaixo, veremos uma tabela que relaciona as principais e mais comuns estatísticas e com seus respectivos parâmetros.

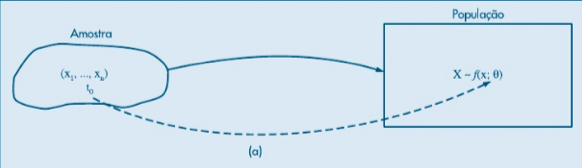
|  |  |  |
| --- | --- | --- |
| Denominação | População | Amostra |
| Média |  |  |
| Mediana |  |  |
| Variância |  |  |
| Nº de Elementos | N | n |
| Proporção | p |  |
| Quantil | Q(p) | q(p) |
| Quartis |  |  |
| Desvio Padrão | |  |

1. Distribuições Amostrais

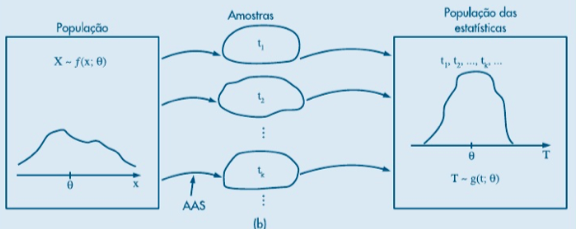
Antes de avançarmos, é preciso entender uma coisa. Quando observamos uma característica da população (ou seja, um parâmetro) sabemos que esse parâmetro possui uma distribuição de frequência, correto?

Porém, sabemos que nosso objetivo é utilizar informações obtidas através de amostras, ou seja, estatísticas. Suponha que estivéssemos interessados em estudar as notas de alunos de uma escola inteira, mas temos apenas as médias das notas de algumas salas. Você acha que a informação que temos bastaria para compreendermos a distribuição de notas da escola toda? A resposta é: dá para ter uma noção muito boa, mas sempre que ocorre um resumo de informação, há também perda de informação.

Considere agora a seguinte pergunta: Mas eu também não posso obter uma distribuição de frequências da minha estatística observada? Sim, mas ela não corresponderá à distribuição populacional, cada estatística terá uma distribuição amostral.

Observe a figura abaixo para compreender melhor a discussão.

Primeiramente, devemos compreender o funcionamento de uma inferência. Selecionamos uma amostra e calculamos uma estatística . Essa amostra é utilizada para fazermos inferências sobre uma característica (parâmetro) da população (a população é a variável X, com todas as observações possíveis). A variável X por sua vez possui uma distribuição de frequência e a característica (parâmetro) . Utilizaremos para inferir .

Observe a próxima figura para darmos continuidade.

O que acontecerá com a distribuição de frequências da nossa estatística? Isto é, se pudéssemos selecionar aleatoriamente TODAS as possíveis amostras de uma mesma população? A figura acima exemplifica o que acontece com a nossa hipotética estatística t (cada estatística apresentará uma distribuição diferente). Observe como a distribuição das estatísticas fica muito mais concentrada que a distribuição da população nesse exemplo.

Resumo:

* Uma população X, com determinado parâmetro de interesse ;
* Todas as amostras retiradas da população de acordo com certo procedimento;
* Para cada amostra, calculamos o valor t da estatística T;
* Os valores t formam uma nova população, cuja distribuição, recebe o nome de distribuição amostral de T.

Vamos procurar entender um exemplo prático?

Suponha que tenhamos 5 bolinhas. Cada bolinha possui uma etiqueta com um número.

7

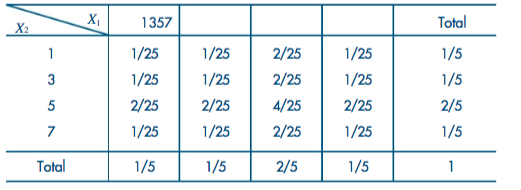
5

3

3

1

Vamos sortear duas bolinhas. Quais são as possibilidades de sorteio? E com que probabilidade?



5

7

3

1