inferencia_graduacao

Norah Jones

2025-12-08

Table of contents

Preface

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

```
1 + 1
```

```
[1] 2
```

1 Summary

In summary, this book has no content whatsoever.

```r
1 + 1
```

```
[1] 2
```

2 População, amostra, famílias e inferência

2.1 População

Em estatística, observações (ou dados) são obtidos de uma população que possui uma característica desconhecida. O objetivo é utilizar essas observações para fazer inferências (deduções) sobre tal característica. Existem duas noções de população.

A primeira noção é ensinada anos iniciais de escolarização, na qual população é um conjunto de pessoas que vivem em uma mesma localidade e em dado período de tempo (como os moradores de um município no mês de Setembro de 2024). Essa noção foi modificada na biologia para incluir indivíduos da mesma espécie que vivem em determinada área durante certo intervalo de tempo. O campo da estatística que lida com esse tipo de população é denominada Amostragem.

Na Amostragem, o estatístico está interessado em estimar alguma quantidade de uma característica da população, como por exemplo, quantas pessoas vão votar no candidato A. É importante notar que, em dado momento do tempo, o número de eleitores é desconhecido, mas não é aleatório - de fato, basta entrevistar todos eles, ou seja, realizar um censo, para resolver o problema. Como o censo em geral é caro, para que inferências possam ser executadas é necessário induzir aleatoriedade através de um plano amostral, que é basicamente um sorteio dos indivíduos da população. Como resultado desse sorteio obtém-se as observações.

A segunda noção coloca a população como sinônimo de distribuição de probabilidades ou modelo. As observações simplesmente são geradas por um mecanismo aleatório e o objetivo é fazer alguma inferência sobre essa distribuição.

Exemplo Considere que uma fábrica possui uma única máquina que fabrica suas peças. Essa máquina tende a se quebrar com o tempo, causando prejuízo. Nesse tipo de problema, chamado confiabilidade, desejamos fazer inferências sobre o tempo em que a máquina se mantém funcionando após uma manutenção. Para tanto, podemos verificar o histórico de falhas e estudar o tempo transcorrido entre as quebras. Aqui vão algumas características importantes sobre esse problema: - Não existe a noção de censo. - Não há a necessidade de um plano amostral porque o evento de interesse (quebra), simplesmente acontece. Portanto, o papel do estatístico é entender como as quebras ocorrem (ou seja, conhecer sua distribuição de probabilidades).

Durante esse curso, estamos interessados na segunda noção de população, ou seja, na distribuição do fenômeno aleatório que está sendo observado. Tal população pode ter nome próprio (como, por exemplo normal, Bernoulli, Poisson, exponencial) ou ser designada pela letra F (de função de distribuição).

A amostra é uma coleção dos dados observados. Essa coleção é indexada e geralmente organizada em um vetor, sendo seu comprimento denominado **tamanho da amostra**. Por exemplo $\mathbf{x} = (3, 1, 5, 7, 10)$ é uma amostra observada de tamanho 5, onde o elemento de índice $i = 1$ é o $x_1 = 3$.

Interpretamos x_i como uma observação gerada a partir da distribuição da variável aleatória X_i. Quando as variáveis $\mathbf{X} = (X_1, \dots, X_n)$ são independentes e identicamente distribuídas, essa coleção é denominada **amostra aleatória**.

2.2 Famílias paramétricas

Seja $X_1, \dots, X_n$ uma amostra aleatória da população(desconhecida) F. Vamos denominar por $\theta(F)$ a característica de interesse. Exemplos de $\theta(F)$ são

- $\theta(F) = F(x)$
- $\theta(F) = E(X)$
- $\theta(F)$ mediana de F Quando $\theta(F)$ é um vetor real e finito, ele é denominado parâmetro e utiliza-se a notação θ em vez de $\theta(F)$. O conjunto de todos os valores permitidos para θ é denominado espaço paramétrico e é denotado por Θ.

Como F é desconhecida, devemos procurar um modelo adequado dentro de uma classe de modelos, denominada família. As famílias podem ser divididas em paramétricas e não paramétricas.

Nota: na prática, o estatístico faz diversas escolhas de famílias de distribuiçóres até encontrar uma que se ajuste ao que foi observado. Por exemplo, se os seus dados são resultados de medições, você deve considerar que sua população pertence à família de todas as distribuições contínuas. Ao considerar que a população deve ter média finita, você considera a família de todas as distribuições contínuas que possuem média. Talvez você acabe descobrindo que é razoável trabalhar com a família que possui todas as densidades normais.

Dizemos que uma população pertence à uma família paramétrica se o valor do parâmetro identifica a população. Em caso contrário, a família é denominada não-paramétrica.

Exemplo Os membros da família das normais possuem dois parâmetros, $\theta = (\mu, \sigma^2)$, que representam a média e a variância. Se sabemos que $\mu = 0$ e $\sigma^2 = 9$, conhecemos a população. Portanto, essa família é paramétrica.

Exemplo A família de todos os modelos com esperança finita possui pelo menos o parâmetro $\theta = E(X)$. Mas saber que $E(X) = 4$ não identifica a população, uma vez que vários modelos possuem esperança igual à 4. Portanto, essa família é não-paramétrica.

Neste curso, estamos interessados majoritariamente em famílias paramétricas.

2.2.1 Família Exponencial

Abaixo, segue a definição de uma importante família paramétrica.

Dizemos que a popução pertence à família exponencial se sua função densidade (ou de probabilidade) pode ser escrita como

$$f(x|\theta) = h(x)a(\theta)\exp\left\{\sum_{j=1}^{k} t_j(x)w_j(\theta)\right\},$$

onde $h(.)$ e $t(.)$ são funções que não podem depender dos parâmetros e $a()$ e $w_j()$ são funções que não podem depender do argumento x.

Importante! Não confunda a família exponencial com a família cujos membros são as distribuições do tipo Exponenical(θ).

Exemplo. Considere que $X \sim$ Exponencial(λ), cuja função densidade é

$$f(x|\lambda) = \lambda e^{-\lambda x}$$

Podemos observar que a distribuição exponencial pertence à família exponencial, pois,

$$f(x|\lambda) = \underbrace{1}_{h(x)} \cdot \underbrace{\lambda}_{a(\lambda)} \exp\{\underbrace{-x}_{t(x)} \cdot \underbrace{\lambda}_{w(\lambda)}\}.$$

Note que nesse exemplo, fizemos $h(x) = 1$. Não há problemas em escrever $h(.)$ dessa forma, uma vez que a única restrição é que $h(.)$ não pode depender dos parâmetros.

Exemplo. Considere que $X \sim$ Poisson(λ), cuja função de probabilidade é

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{1}{x!}e^{-\lambda}\lambda^x.$$

Como

$$\lambda^x = \exp\{\log(\lambda^x)\} = \exp\{x\log(\lambda)\},$$

podemos observar que a distribuição Poisson pertence à família exponencial, pois,

$$P(X = x|\lambda) = \underbrace{\frac{1}{x!}}_{h(x)} \underbrace{e^{-\lambda}}_{a(\lambda)} \exp\{\underbrace{x}_{t(x)} \cdot \underbrace{\log(\lambda)}_{w(\lambda)}\}.$$

Exemplo. Considere que $X \sim$ Bernoulli(p), cuja função de probabilidade é

$$P(X = x|p) = p^x(1-p)^{1-x} = \left(\frac{p}{1-p}\right)^x (1-p).$$

Como
$$\left(\frac{p}{1-p}\right)^x = \exp\left\{\log\left(\left(\frac{p}{1-p}\right)^x\right)\right\} = \exp\left\{x\log\left(\frac{p}{1-p}\right)\right\},$$
podemos observar que a distribuição Bernoulli pertence à família exponencial, pois,
$$P(X=x|p) = \underbrace{1}_{h(x)}\underbrace{(1-p)}_{a(p)}\exp\{\underbrace{x}_{t(x)}\underbrace{\log\left(\frac{p}{1-p}\right)}_{w(p)}\}.$$

Exemplo. Considere que $X \sim \text{Normal}(\mu,\sigma^2)$, cuja função densidade é
$$f(x|\mu,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$
Como,
$$(x-\mu)^2 = x^2+\mu^2-2x\mu,$$
Podemos observar que a distribuição normal pertence à família exponencial, pois,
$$f(x|\lambda) = \underbrace{\left(\frac{1}{2\pi}\right)^{1/2}}_{h(x)} \cdot \underbrace{\frac{e^{-\frac{\mu^2}{2\sigma^2}}}{\sigma}}_{a(\mu,\sigma^2)}\exp\{\underbrace{-x^2}_{t_1(x)}\underbrace{\frac{1}{2\sigma^2}}_{w_1(\mu,\sigma^2)} + \underbrace{x}_{t_2(x)}\underbrace{\frac{\mu}{2\sigma^2}}_{w_2(\mu,\sigma^2)}\}.$$

Em geral, se os possíveis valores da variável aleatória dependem de parâmetros, então o respectivo modelo não pertence à família exponencial

Exemplo Seja $X \sim \text{Uniforme}(0,\theta)$, cuja função densidade é dada por
$$f(x|\theta) = \frac{1}{\theta}I(0 < x \leq \theta).$$
Observe que não é possível escrever $I(0 < x \leq \theta)$ como $\exp\{t(x)w(\theta)\}$, logo a Uniforme(0,θ) não é um membro da família exponencial.

2.3 Inferência

A inferência estatística consiste em utilizar os dados $x_1,\dots,x_n$ para tentar responder uma das três questões:

1. Como posso usar os dados para dizer quanto vale θ?
2. Como posso usar os dados para escolher os valores $a < b$ tais que seja possível inferir que $a < \theta < b$
3. Como posso usar os dados para dizer a afirmação $\theta \in A$ é falsa?

No jargão estatístico, os problemas acima são denominados estimação pontual, intervalo de confiança e teste de hipóteses, respectivamente.

3 Estatística

3.1 Espaço amostral e a estatística como técnica para redução de dados

Anteriormente discutimos que os dados são gerados a partir de uma distribuição de probabilidades, conhecida como população. Portanto, por natureza, a amostra possui informação sobre a população.

Definição O conjunto de todas as amostras possíveis é denominado **Espaço Amostral** e será denotado por $\mathcal{X}$.

O espaço amostral tende a ser mais complexo com o aumento do tamanho da amostra. A estratégia para diminuir a complexidade da análise é utilizar uma estatística.

Seja $\mathbf{X}$ uma amostra aleatória. Então, qualquer função $T : \mathcal{X} \to \mathbb{R}^c$ é denominada **estatística** (de dimensão c).

O objetivo primário de uma estatística é reduzir a informação da amostra, trocando o problema de analisar a amostra original que está em um espaço de dimensão n para um espaço de dimensão $c \leq n$.

Exemplo. Considere uma amostra de tamanho 3 da população Bernoulli(θ). O espaço amostral é

$$\mathcal{X} = \{000, 001, 010, 100, 011, 101, 110, 111\},$$

possui dimensão três e tem oito elementos. Agora, considere a estatística $T = X_1 + X_2 + X_3$. Os possíveis valores de T são $\{0, 1, 2, 3\}$. Esse espaço possui dimensão um e tem quatro elementos.

Como o espaço da estatística possui dimensão menor que o espaço amostral, sempre haverá perda de informação. O motivo disso é bastante simples: em geral não é possível recuperar a amostra original.

3.2 Distribuição amostral e estatística ancilar

A distribuição de uma estatística é denominada **distribuição amostral**. Dizemos que a estatística carrega informação sobre o parâmetro se sua distribuição amostral depende do parâmetro. Em geral, amostra aleatória, estatística e parâmetro se relacioanam conforme mostra a figura abaixo.

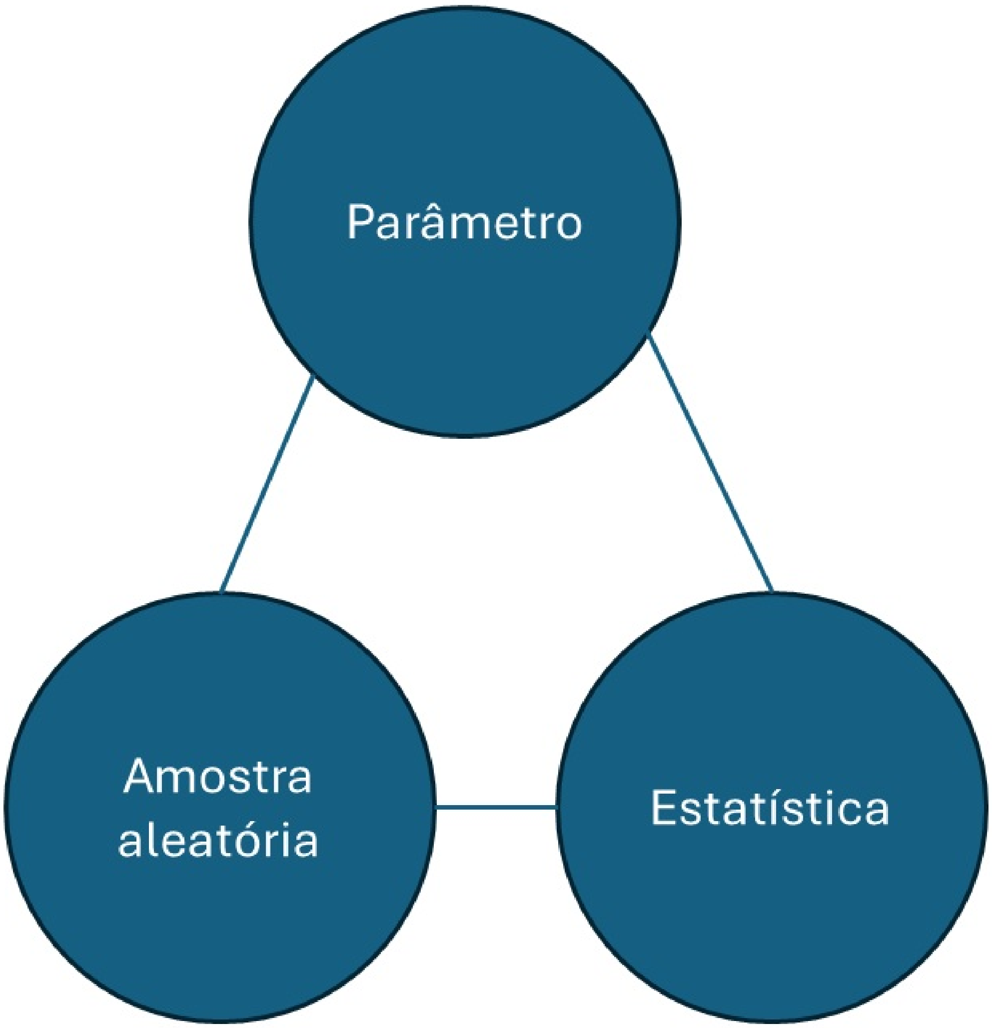


Figure 3.1: A relação mais comum entre amostra, estatística e os parâmetros populacionais

Se a distribuição amostral da estatística não depende dos parâmetros, dizemos que essa estatístca é ancilar. A figura abaixo representa a relação entre a amostra aleatória, os parâmetros e esse tipo de estatística.

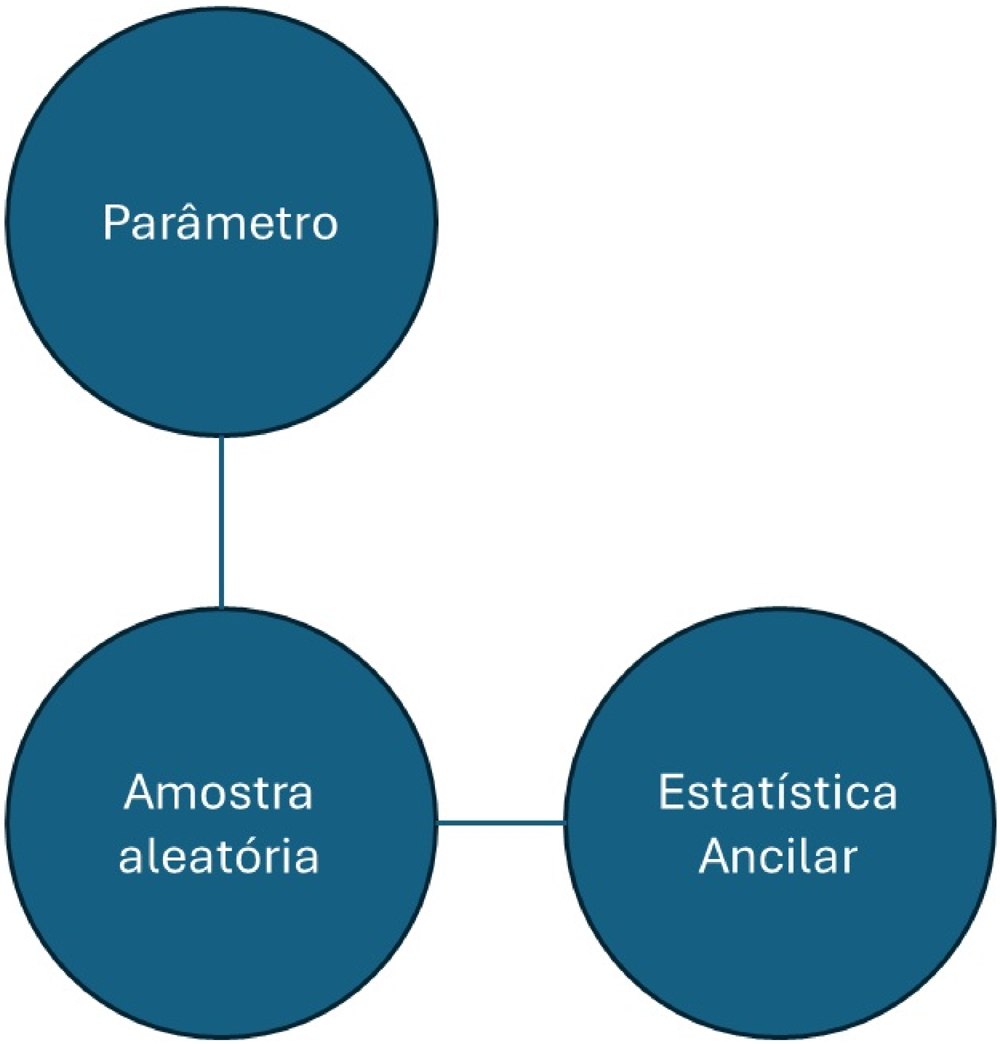


Figure 3.2: Relação entre amostra, estatísticas ancilares e os parâmetros populacionais

Exemplo Seja $X_1, \dots, X_n$ uma amostra aleatória da população Normal$(\mu, 1)$. Sabemos que a distribuição amostral da média amostral é

$$\bar{X}_n \sim \text{Normal}\left(\mu, \frac{1}{n}\right).$$

Portanto, $\bar{X}_n$ carrega informação sobre μ. Considere agora a estatística

$$W = X_1 - X_2.$$

Note que W é uma combinação linear de normais independentes, logo também possui distribuição normal. Como

$$E(W) = E(X_1) - E(X_2) = 0$$

e

$$Var(W) = Var(X_1 - X_2) = Var(X_1) + Var(X_2) = 2,$$

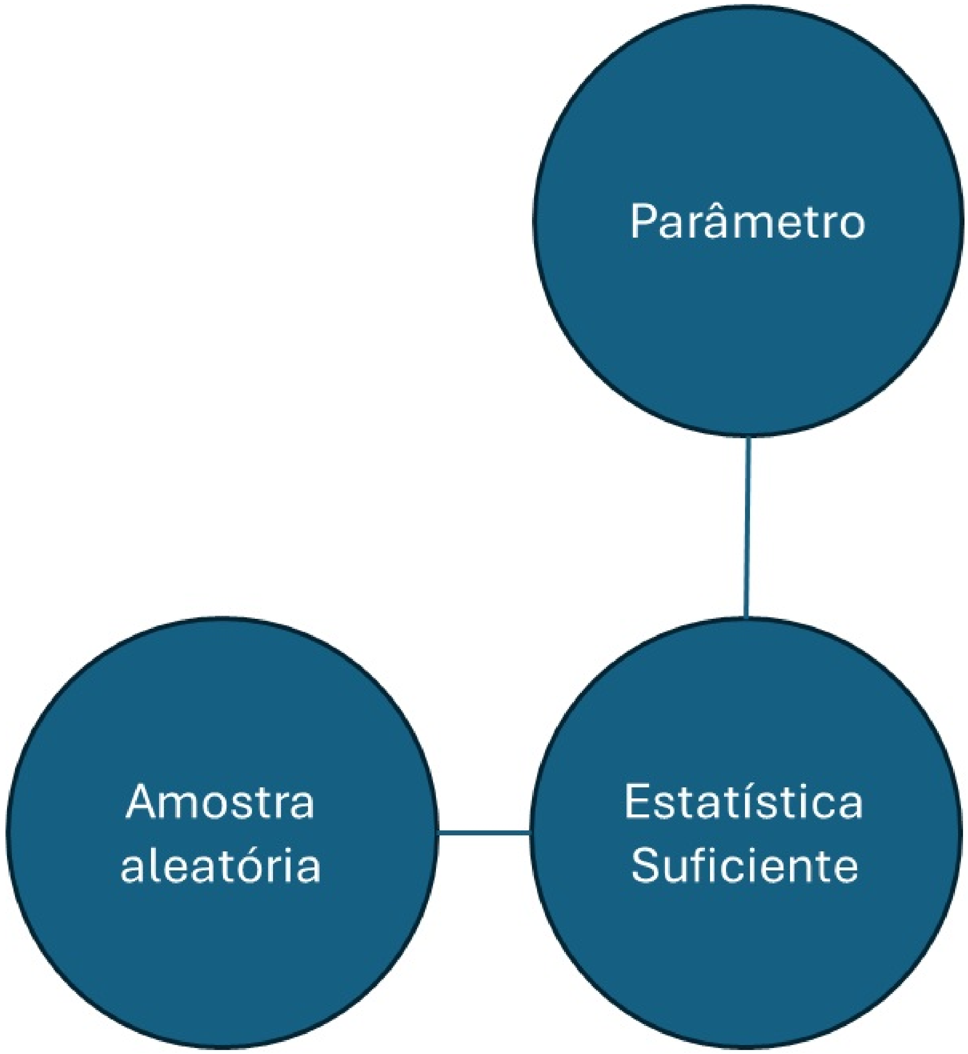
temos que a distribuição amostral de W é Normal(0,2). Como essa distribuição não depende de μ, temos que W não carrega informação sobre o parâmetro e, portanto, é uma estatística ancilar.

3.3 Estatísticas suficientes

Uma estatística é dita ser suficiente para θ se, quando observada, ela permite descrever a distribuição da amostra aleatória sem o conhecimento de θ. Segue a definição formal.

Definição. Uma estatística T é dita ser suficiente para θ se a distribuição $X_1, \dots, X_n | T = t$ não depende de θ.

A figura abaixo mostra a relação entre a amostra aleatória, os parâmetros e a estatística suficiente. Observe que a depenência da amostra em relação aos parâmetros se dá através da es-



tatítica suficiente.

Note que a própria amostra é uma estatística suficiente para θ. De fato, considerando o caso discreto, pode-se notar que

$$P(\mathbf{X} = \mathbf{x}|\mathbf{X} = \mathbf{x}, \theta) = \frac{P(\mathbf{X} = \mathbf{x}, \mathbf{X} = \mathbf{x}|\theta)}{P(\mathbf{X} = \mathbf{x}|\theta)} = 1,$$

que não depende de θ. O teorema a seguir é uma importante ferramenta para encontrar estatísticas suficientes.

Teorema do Critério da Fatoração de Neyman Seja $\mathbf{X}$ uma amostra aleatória cuja distribuição. Então, T é um estatística suficiente para θ se e somente se existem funções $h(\mathbf{x})$ e $g(T, \theta)$ tais que

$$\prod_{i=1}^{n} f(x_i|\theta) = h(\mathbf{x})g(t,\theta)$$

onde f é a função de densidade ou a função de probabilidade da população.

Exemplo Seja $X_1, \dots, X_n$ uma amostra aleatória da população Exponencial(θ). Note que

$$\prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \underbrace{\theta^n}_{h(\mathbf{x})} \cdot \underbrace{e^{-\theta \sum_{i=1}^{n} x_i}}_{g(\sum_{i=1}^{n} x_i, \theta)},$$

logo, pelo Teorema do Critério da Fatoração, $T = \sum_{i=1}^{n} X_i$ é uma estatística suficiente para θ.

Exemplo. Seja $X_1, \dots, X_n$ uma amostra aleatória do modelo $X \sim$ Poisson(λ). Note que

$$P(\mathbf{X} = \mathbf{x}|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \underbrace{\prod_{i=1}^{n} \frac{1}{x_i!}}_{h(\mathbf{x})} \cdot \underbrace{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}_{g(\sum_{i=1}^{n} x_i, \lambda)},$$

logo, pelo Teorema do Critério da Fatoração de Neyman, $T = \sum_{i=1}^{n} X_i$ é suficiente para λ.

Quando há mais de uma estatística na fatoração, elas são denominadas conjuntamente suficientes.

Exemplo. Seja $X_1, \dots, X_n$ uma amostra aleatóra da população Normal(μ, σ^2). Note que

$$f(\mathbf{x}|\mu, \sigma^2) = \prod_{i=1}^{n} \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}.$$

Como

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n} x_i^2 + n\mu^2 - 2\mu \sum_{i=1}^{n} x_i,$$

teremos que

$$f(\mathbf{x}|\mu, \sigma^2) = \underbrace{1}_{h(\mathbf{x})} \cdot \underbrace{\left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 + n\mu^2 - 2\mu \sum_{i=1}^{n} x_i\right)}}_{g(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2, \mu, \sigma^2)}.$$

Portanto, pelo Teorema do Critério da Fatoração de Neyman, $\sum_{i=1}^{n} X_i$ e $\sum_{i=1}^{n} X_i^2$ são conjuntamente suficientes para μ e σ^2.

Você deve ter notado que todas as distribuições acima pertencem à família exponencial. O teorema abaixo generaliza a busca de estatísticas suficientes dentro dessa família.

Teorema Se $X_1, \dots, X_n$ é uma amostra aleatória de uma população na família exponencial, então

$$T = \left\{ \sum_{i=1}^{n} T_1(X_i), \dots, \sum_{i=1}^{n} T_k(X_i) \right\}$$

é uma estatística suficiente para θ.

Vejamos alguns exemplos fora da família exponencial.

Exemplo. Seja $X_1, \dots, X_n$ uma amostra aleatória da população Uniforme(0,θ). Note que

$$\prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I(x_i \leq \theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(x_i \leq \theta).$$

O produtório acima é igual a 1 se e somente se todas as observações forem menores ou iguais que θ. Para que isto ocorra, basta que a maior das observações seja menor que θ. Denotando a estatística máximo amostral por $X_{(n)}$, teremos

$$\prod_{i=1}^{n} f(x_i|\theta) = \underbrace{1}_{h(\mathbf{x})} \cdot \underbrace{\frac{1}{\theta^n} I(x_{(n)} \leq \theta)}_{g(x_{(n)}, \theta)},$$

logo, pelo Teorema do Critério da Fatoração de Neyman, teremos que $T = X_{(n)}$ é suficiente para θ.

Importante Sejam $x_{(1)}$ e $x_{(n)}$ as estatísticas mínimo e máximo de uma amostra de tamanho n. Então: - $\prod_{i=1}^{n} I(x_i \leq \theta) = I(x_{(n)} \leq \theta)$ - $\prod_{i=1}^{n} I(x_i \geq \theta) = I(x_{(1)} \geq \theta)$ - $\prod_{i=1}^{n} I(\alpha \leq x_i \leq \beta) = I(\alpha \leq x_{(1)}) I(x_{(n)} \leq \beta)$

Exemplo. Seja $X_1, \dots, X_n$ uma amostra aleatória da população cuja função densidade é dada por

$$f(x|\mu) = e^{-(x-\mu)} I(x \geq \mu).$$

Esse modelo é conhecido como exponencial deslocada. Contudo, como os valores de x dependem de μ, esta distribuição não está na família exponencial. Observe que

$$\prod_{i=1}^{n} f(x_i|\mu) = \prod_{i=1}^{n} e^{-(x_i-\mu)} I(x_i \geq \mu) = e^{-\sum_{i=1}^{n} x_i} e^{n\mu} \prod_{i=1}^{n} I(x_i \geq \mu).$$

O produtório acima é igual a 1 se e somente se todas as observações forem maiores ou iguais que μ. Para que isto ocorra, basta que $x_{(1)}$ seja maior que μ. Então, teremos que

$$\prod_{i=1}^{n} f(x_i|\theta) = \underbrace{e^{-\sum_{i=1}^{n} x_i}}_{h(\mathbf{x})} \cdot \underbrace{e^{n\mu} I(x_{(1)} \geq \mu)}_{g(x_{(1)}, \mu)},$$