**Resumo 8 – Modelos Probabilísticos – Variável Discreta**
**Como usar modelos de probabilidade para entender melhor os fenômenos aleatórios**

Na estatística descritiva procuramos entender uma variável estudando o comportamento de um conjunto de observações (amostra). Desta forma, estudamos a distribuição de freqüências, por exemplo, de um animal ser de determinada raça (sim ou não), com base numa amostra de animais de uma propriedade rural de interesse. Nessa abordagem predomina o raciocínio indutivo: com base na organização e descrição de dados observados procuramos fazer conjecturas sobre o universo (população) em estudo.

O raciocínio de forma inversa, em que procuramos entender como poderão ocorrer os resultados de uma variável, considerando certas suposições a respeito do problema em estudo, é chamado raciocínio dedutivo. Por exemplo, supondo que 60% dos animais de uma propriedade rural são da raça Nelore, o que se pode deduzir sobre a porcentagem de animais Nelore, numa amostra aleatória simples de 200 animais?

A resposta a esta pergunta não é um simples número, pois dependendo dos 200 animais selecionados na amostra, teremos resultados diferentes.

Para responder adequadamente precisamos apresentar quais são os possíveis resultados e como eles poderão ocorrer. Essa descrição é feita em termos dos chamados Modelos Probabilísticos.

## 1. Modelos probabilísticos (Distribuições de Probabilidade)

São construídos a partir de certas hipóteses ou conjecturas sobre o problema em questão e constituem-se de duas partes: (1) dos possíveis resultados e (2) de uma certa lei que nos diz quão provável é cada resultado (ou grupo de resultados).

Exemplo 1: Em um lote de cordeiros, selecionar um animal e verificar se pesa mais de 30 Kg. Um espaço amostral adequado é $\Omega$ ={sim, não}

Exemplo 2: Em uma fazenda, selecionar uma amostra de 3 cordeiros de um lote e verificar quantos pesam mais de 30Kg. Um espaço amostral adequado é $\Omega$ = {(NNN), (NNS), (NSN), (SNN), (SSN), (SNS), (NSS), (SSS)} ou {0, 1, 2, 3}.

## 2. Principais Modelos Teóricos de Probabilidade – Variável Discreta

### 2.1. Ensaios de Bernoulli

Os ensaios de Bernoulli ocorrem em situações onde observamos apenas um elemento e verificamos se este tem ou não um certo atributo.

No exemplo1, para verificar se um animal pesa mais que 30 Kg, é necessário o conhecimento da distribuição desta característica (pesar mais de 30 Kg) na população. Vamos supor que em todo o lote 60% dos animais pesam mais que 30 Kg e 40% pesam menos que 30 Kg. Se o animal for selecionado aleatoriamente, podemos explicitar o modelo probabilístico, como a seguir:

| Resultado | Probabilidade |
|---|---|
| Sim | 0,6 |
| Não | 0,4 |

Muitas vezes não temos informações suficientes para especificar completamente o modelo probabilístico. No exemplo 1, podemos não conhecer a porcentagem de animais que pesam mais que 30 Kg na população. Nesse caso podemos apresentar o modelo de forma geral.

Podemos caracterizar um ensaio de Bernoulli por uma variável aleatória X, definida da forma:

$$X = \begin{Bmatrix} 0, se\ fracasso \\ 1, se\ sucesso \end{Bmatrix}$$

com função de probabilidade dada por:

$$p(x) = \Pr(X = x) = p^x (1-p)^{1-x}$$

e o modelo de probabilidade:

| x | 1 | 0 |
|---|---|---|
| p(x) = Pr(X=x) | p | 1-p |

onde:

p: é uma quantidade entre 0 e 1 (parâmetro do modelo);

x:é um possível valor da variável X (no caso, 0 ou 1);e

p(x) é a probabilidade de ocorrer o valor de x.

A função p(x) = Pr(X=x) é denominada função de probabilidade da variável aleatória discreta X , satisfazendo $p(x) \geqslant 0 \quad e \quad \sum p(x) = 1$ .

**Média de uma variável Bernoulli**

$$\mu = -E(X) = p$$

**Variância e desvio padrão de uma variável Bernoulli**

$$\sigma^2 = \ Var(X) \ = \ p(1-p)$$

e

$$\sigma \ = \ DP(X) \ = \ \sqrt{p(1-p)}$$

## 2.2. O modelo Binomial

No exemplo2, para verificar quantos animais pesam mais de 30 Kg, em uma amostra de 3 animais, é necessário conhecer a distribuição dessa característica na população. Podemos ter o seguinte modelo probabilístico:

| X = Número de animais que pesam mais de 30 Kg | p(x) |
|---|---|
| 0 {(NNN)} | 0,064 |
| 1 {(NNS), (NSN), (SNN)} | 0,288 |
| 2 {(NSS), (SNS), (SSN)} | 0,432 |
| 3 {(SSS)} | 0,216 |
| Total | 1,000 |

Podemos generalizar para a situação em que queremos verificar a ocorrência de determinada característica em n repetições do experimento de Bernoulli.

Uma experiência binomial é um processo que apresenta as seguintes características:

(1) Possui n ensaios de Bernoulli.

(2) Só há dois tipos de resultados possíveis. Por questão de notação, designaremos esses resultados como “sucesso” (S) e “fracasso” (F).

(3) A probabilidade de resultar um “sucesso” em um ensaio é igual a *p*; por conseguinte, a probabilidade de resultar um “fracasso” é (1 - *p*) = *q*.

(4) Todos os ensaios são independentes.

(5) Interessa-nos contar o número de “sucessos” obtidos em n ensaios de Bernoulli.

Seja, então, uma variável aleatória discreta X, igual ao número de “sucessos” nos n ensaios de Bernoulli. Então dizemos que X segue um modelo tem uma distribuição Binomial com parâmetros n e p, com função de probabilidade dada por:

$$p(x) = \Pr(X = x) = \binom{n}{x} p^{x}(1-p)^{n-x}, \text{ para x = 0, 1, 2, ..., n com } \binom{n}{x} = \frac{n!}{(n-x)!x!}$$

onde *p* = probabilidade de “sucesso”

1-*p* = probabilidade de “fracasso”

- **Média de uma variável aleatória binomial**

$$\mu = E(X) = n \cdot p$$

- **Variância e desvio padrão de uma variável aleatória binomial**

$$\sigma^2 = Var(X) = n \cdot p \cdot (1-p)$$

e

$$\sigma = DP(X) = \sqrt{np(1-p)}$$

A denominação binomial decorre do fato de os coeficientes $\binom{n}{x} = \frac{n!}{(n-x)!x!}$ serem exatamente os coeficientes do desenvolvimento binomial dos termos de $(a+b)^n$ .

**2.3. A Distribuição (ou Modelo) de Poisson**

Podemos considerar uma variável aleatória X igual ao número de “sucessos” que ocorrem num intervalo contínuo. Por exemplo:

a) o número de bactérias por volume unitário de um fluído;

b) o número de insetos vivos por $m^2$ de uma grande área de rocio;

c) o número de acidentes que ocorrem num certo cruzamento no período de 1 hora.

A variável aleatória discreta X = número de “sucessos” que ocorrem num determinado período de tempo, distância, área ou volume tem distribuição de Poisson com parâmetro $\lambda$ e função de probabilidade dada por:

$$p(x) = \Pr(X = x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!}, \text{ para x = 0, 1, 2, ... ,}$$

onde $\lambda$ é a média de “sucessos” no período, distância, área ou volume considerado.

A *média* e a *variância* para uma variável aleatória com distribuição de Poisson são dadas por

$$\mu = E(X) = \lambda = \sigma^2 = Var(X) \text{ e } \sigma = DP(X) = \sqrt{\lambda}$$

Exemplo 3: Se uma gota de água for depositada em uma lâmina e examinada ao microscópio, o número de X de um tipo particular de bactéria presente em uma gota d'água apresenta distribuição de Poisson. Suponha que o número máximo dessa bactéria por gota de água seja 5. Se a média desse conteúdo, em uma certa experiência, for de 2 bactérias em uma única gota testada, pode-se esperar que o número máximo admitido seja ultrapassado? Justifique sua resposta.

***2.3.1. A distribuição de Poisson como limite de uma distribuição Binomial.***

Podemos mostrar que para *n* suficientemente grande e *p* suficientemente pequeno, podemos aproximar a distribuição binomial para uma distribuição de Poisson, ou seja,

$$\lim_{\substack{n\to\infty \\ p\to 0}} \left\{ \binom{n}{x} p^x q^{n-x} \right\} = \frac{e^{-np} (np)^x}{x}$$

Na prática, essa aproximação é boa para $n \geq 500$ e $np < 10$.

**Exercícios:**

1) Mil pássaros, um de cada vez, têm de escolher entre duzentas gaiolas dispostas em um círculo. Admitamos que os pássaros não têm preferência direcional. Vamos definir a variável aleatória de interesse como X:"número de pássaros por gaiola". Qual a probabilidade de que uma gaiola específica seja escolhida por no mínimo 3 pássaro? (12,46%)

2) Suponha que 20% de uma criação de suínos esteja atacada por leptospirose. Seja Y o número de suínos doentes de uma amostra de 1000 animais, dessa criação, examinada por um veterinário.

(a) qual é o valor esperado de Y?(200)

(b) qual é o desvio padrão de Y? (12,65)

3) Doze pares de animais experimentais são submetidos a duas dietas diferentes, A e B. A atribuição da dieta aos animais de cada par é feita ao acaso (princípio da casualização). Após um período de tempo, acha-se a diferença em ganho de peso entre os animais submetidos à dieta A e à dieta B. Se a diferença for positiva (A-B > 0), o resultado será chamado de sucesso. Sabendo-se que as duas dietas não possuem diferenças reais no que diz respeito à variável ganho de peso:

(a) Verifique se a variável aleatória X:"número de sucessos nos doze pares" é uma variável binomial.

(b) Calcule o valor esperado, a variância e o desvio-padrão da variável X. (6; 3; 1,73)

(c) Verifique qual a probabilidade de ocorrerem pelo menos nove sucessos. (7,29%)