ESTATÍSTICA

Juliane Silveira Freire da Silva

sagah

Distribuições discretas de probabilidade: binomial e de Poisson

Objetivos de aprendizagem

Ao final deste texto, você deve apresentar os seguintes aprendizados:

- Definir as distribuições de probabilidade.
- Reconhecer as distribuições discretas de probabilidade.
- Calcular probabilidades utilizando os métodos binomial e de Poisson.

Introdução

Neste capítulo, você entenderá o que são distribuições de probabilidades, conhecerá as distribuições discretas de probabilidade e aprenderá a calcular probabilidades utilizando os métodos binomial e de Poisson.

Distribuições de probabilidade

Assim como as variáveis quantitativas discretas e quantitativas contínuas, também existem as distribuições de probabilidade discretas e distribuições de probabilidade contínuas.

A lógica de entendimento é a mesma: assim como a variável quantitativa discreta resulta de uma contagem, a distribuição de probabilidade discreta tem seu espaço amostral com resultados possíveis que resultam de uma contagem.

A variável quantitativa contínua resulta de uma medição, e a distribuição de probabilidade contínua tem, em seu espaço amostral, medidas que resultam de alguma medição em um espaço contínuo.

As distribuições de probabilidade são descritas por uma variável aleatória X, que é uma função que, a cada valor do espaço amostral, associa um número real.

Saiba mais

Experimento é tudo que possa ser reproduzido n vezes sob as mesmas condições e será aleatório quando não pudermos prever o resultado, apenas saber de antemão todos os resultados possíveis. Espaço amostral são todos os resultados possíveis de um experimento.

Variável aleatória discreta

Uma variável aleatória X é dita discreta quando puder assumir apenas valores inteiros ao longo de uma escala. Se, para cada um dos valores da variável aleatória discreta, teremos a sua probabilidade definida por:

$$f(x) = P(X = x)$$

onde:
$f(x)$ função matemática de x;
$P(X = x)$ probabilidade da variável aleatória X em determinado ponto da escala x.

Como estamos lidando com um valor discreto do espaço amostral da variável em estudo, para , teremos apenas valores inteiros.

A função de probabilidade da variável aleatória discreta também é chamada função massa de probabilidade (fmp) e satisfaz os seguintes pressupostos:

- $0 \leq f(x) \leq 1$
- $\sum f(x_i) = 1$

Por exemplo, uma moeda equilibrada é lançada duas vezes. A variável X é o número de caras nesses lançamentos. O espaço amostral é descrito por:

C = coroa
K = cara

$$\Omega = (CC, CK, KC, KK)$$

$$X = 0 \rightarrow f(0) = P(CC) = \frac{1}{4}$$

$$X = 1 \rightarrow f(1) = P(CK \cdot KC) = \frac{2}{4}$$

$$X = 2 \rightarrow f(2) = P(KK) = \frac{1}{4}$$

x	0	1	2
f(*x*)	¼	$^2/_4$	¼

Variável aleatória contínua

Uma variável aleatória é dita contínua quando X puder assumir qualquer valor do intervalo do espaço amostral. Diferentemente das funções discretas de probabilidade, em que podemos calcular os valores no ponto em que a variável X assume nas distribuições continuas. Como a função $f(x)$ será contínua, nesse caso, calculamos a probabilidade de intervalos para a variável X, pois, matematicamente, quando temos uma função contínua, precisamos calcular áreas abaixo da curva para chegarmos às probabilidades associadas à variável X.

Vamos a um exemplo de variável contínua: seja X o tempo médio de vida útil de uma lâmpada. O espaço amostral pode ir de um tempo 0 até um tempo que pode tender ao infinito. Para calcularmos a probabilidade, precisaríamos de intervalos, como a duração de uma lâmpada em um tempo superior a 3000 horas. Esse espaço amostral e a probabilidade seriam representados por:

$$\Omega = (t \in \mathrm{R}, 0, +\infty)$$
$$f(x > 3000) = P(X > 3000)$$

Um exemplo de distribuição contínua de probabilidade é a distribuição normal, a mais importante dentro da estatística.

Distribuições discretas de probabilidade

Muitas vezes, ficar pensando em espaço amostral e todas as possibilidades de funções pode ser complicado e desnecessário. Por esse motivo, algumas distribuições foram criadas por sua frequência de uso e seu uso ser útil em variáveis com comportamentos similares e predefinidos. Essas distribuições têm funções matemáticas predefinidas.

Existem várias distribuições discretas de probabilidade: a uniforme, a geométrica, a binomial, a de Poisson, entre outras.

Aqui, trataremos das duas distribuições mais usuais, devido à sua aplicação com dados de variáveis aleatórias discretas: a distribuição binomial e a de Poisson.

Distribuição binomial

A distribuição binomial é utilizada quando temos um número de repetições de um experimento, uma probabilidade de sucesso associada ao acontecimento positivo do que estamos estudando e uma probabilidade de fracasso sobre esse mesmo evento.

São situações em que pode haver sucesso ou não, e nenhuma outra hipótese é permitida como o número de caras em 50 lançamentos de uma moeda.

Então, temos um experimento com espaço amostral associado , além de repetições desse experimento. Temos, também, p probabilidade de um evento desse espaço amostral ocorrer em cada uma das repetições do experimento.

Na distribuição binomial, o evento ocorre ou não — temos somente essas duas opções. Então, se temos uma probabilidade p desse evento ocorrer, temos uma probabilidade $q = 1 - p$ desse evento não ocorrer.

Costuma-se denominar como p sendo a probabilidade de sucesso e q como sendo a probabilidade de fracasso. Vale ressaltar que, dependendo do evento que estejamos estudando, o sucesso não necessariamente seja uma afirmativa positiva. Quando utilizamos o termo sucesso, estamos dizendo que é a probabilidade de sucesso de ocorrer o evento em particular que estamos investigando, independentemente de ele ter um resultado considerado positivo ou não.

A forma da distribuição binomial é demonstrada no gráfico da Figura 1, a seguir, considerando 60 repetições de um experimento e uma probabilidade de sucesso de 15%. Anotamos uma distribuição binomial por $B(n,p)$, no caso do gráfico $B(20;0,15)$.

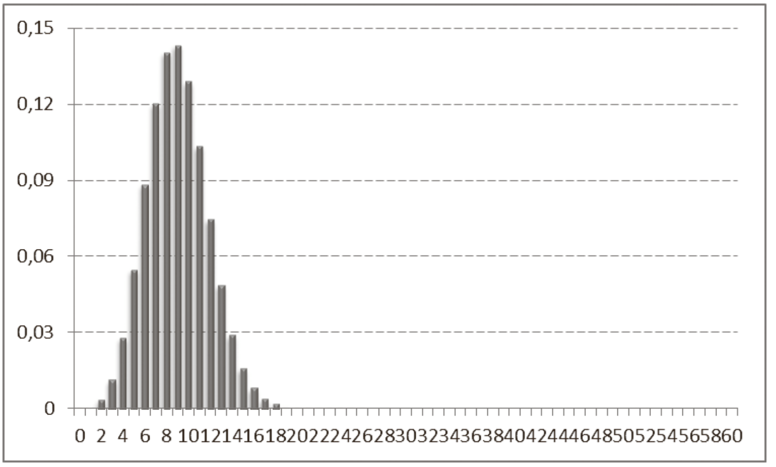

Figura 1. Comportamento distribuição *B*(60;0,15).

A fórmula da função matemática para cálculo de uma distribuição binomial é dada por:

$$f(x) = P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

onde:
x é o valor do espaço amostral que se quer calcular a probabilidade;
n é o número de repetições;
p é a probabilidade de sucesso;
$q = 1 - p$ é a probabilidade de fracasso.

Fique atento

Observe que, na fórmula, temos o termo $\binom{n}{x}$ Isso é resolvido por análise combinatória e significa *n* combinação *x*, ou seja: $\binom{n}{x} = \frac{n!}{x!.\,(n-x)!}$ em que o ponto de exclamação significa fatorial.

Em algumas calculadoras científicas, a tecla para a resolução desse termo da função é *nCr*.

Por exemplo, atualmente, sabemos que as redes sociais são utilizadas para comercialização de produtos. Sabe-se, por uma pesquisa realizada, que cerca de 15% dos itens postados são efetivamente vendidos. Primeiramente, queremos saber a probabilidade de, pelo menos, 2 itens serem vendidos em um dia que 10 itens foram postados para venda. Os valores que pode assumir são $x = (2,3,4,5,6,7,8,9,10)$. Para não precisarmos calcular todas essas probabilidades, podemos fazer uso da propriedade do complementar e tirar do espaço amostral os valores que não fazem parte dessa sentença e têm probabilidade 1.

$$P(X \geq 2) = 1 - P(X < 2) = 1 - (P(X = 0) + P(X = 1) =$$

$$1 - \left(\binom{10}{0} \cdot 0{,}15^0 \cdot 0{,}85^{10-0} + \binom{10}{0} \cdot 0{,}15^1 \cdot 0{,}85^{10-1}\right) = 0{,}4557 = 45{,}57\%$$

A segunda questão é a probabilidade de vender um produto. Para isso, calculamos apenas $x = 1$.

$$P(X = 1) = \binom{10}{1} \cdot 0{,}15^1 \cdot 0{,}85^{10-1} = 0{,}3474 = 34{,}74\%$$

Por fim, calcularemos a probabilidade de que sejam vendidos menos de 3 produtos. Aqui, o x pode assumir os seguintes valores: $x = 0,1,2$.

$$P(X < 3) = \left(\binom{10}{0} \cdot 0{,}15^0 \cdot 0{,}85^{10-0} + \binom{10}{0} \cdot 0{,}15^1 \cdot 0{,}85^{10-1} + \binom{10}{0} \cdot 0{,}15^2 \cdot 0{,}85^{10-2}\right)$$

$$= 0{,}8202 = 82{,}02\%$$

Distribuição de Poisson

Assim como a distribuição binomial, a de Poisson também conta sucessos. Porém, ao invés de eles serem observados em um número de repetições, são feitos em um intervalo contínuo de tempo ou espaço. O sucesso da distribuição Poisson é observado em um intervalo contínuo, e o da binomial é em um número de repetições.

Segundo Doane e Seward (2014), a distribuição de Poisson foi assim denominada em homenagem ao matemático francês Simèon Denis Poisson (1781-1840) e descreve o número de ocorrências de um evento dentro de uma unidade de tempo (por exemplo, minuto ou hora), escolhida aleatoriamente, ou de espaço (por exemplo, metro quadrados ou quilômetros lineares). Para se usar a distribuição, os eventos devem ocorrer aleatória e independentemente no espaço ou em tempo contínuo.

Por exemplo, se nossa variável X fosse número de chamadas não atendidas em uma central telefônica, caso observássemos essa variável em um dia que ocorreram 300 ligações, teríamos a proporção de chamadas não atendidas (nossa probabilidade de sucesso) em 300 repetições do experimento, o que caracterizaria uma distribuição binomial. Porém, se observássemos a quantidade de chamadas não atendidas em um turno de 8 horas de trabalho, teríamos a taxa de ocorrência por 8 horas de trabalho, o que caracterizaria uma distribuição de Poisson.

A distribuição de Poisson é representada por $P(\lambda)$, sendo λ a taxa de ocorrência do evento em estudo da variável x. Para percebermos o comportamento da função da distribuição de Poisson, observaremos o gráfico resultante de uma Poisson com $\lambda = 5$. $P(5)$, na Figura 2.

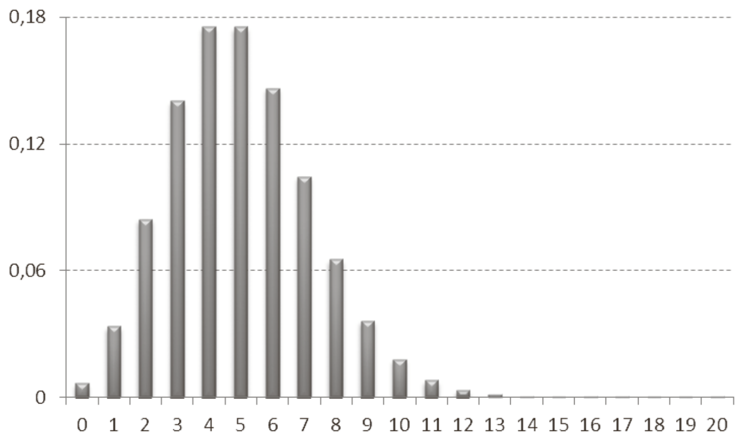

Figura 2. Comportamento distribuição *P*(5).

A função matemática para o cálculo dessa distribuição é dada por:

$$f(x) = P(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

onde:

x é o valor do espaço amostral em que se quer calcular a probabilidade;

λ é a taxa de ocorrência.

Fique atento

Observe que, na fórmula, temos o termo *e*, que representa a constante Euler. É um valor constante igual a 2,718281828459045235360287..., assim como o conhecido π. Para calcular a expressão $e^{-\lambda}$ nas calculadoras científicas, utilizamos a tecla e^x.

Relembrando: o ponto de exclamação representa o fatorial.

Exemplo

Imagine essa central telefônica e que a taxa de chamadas não atendidas em um turno de 8 horas é de 10 chamadas. Queremos investigar a probabilidade de não termos chamadas não atendidas em uma hora.

Observem que a taxa é dada por 8 horas, mas queremos calcular a probabilidade por hora. e então, a primeira coisa a se fazer é descobrir a taxa por hora de chamadas não atendidas. Isso se resolve com uma regra de três.

10 chamadas	8horas
λ	1 hora

Então temos $\lambda = 1{,}25$.

Agora, calcularemos a probabilidade de não termos chamada não atendida. e então, queremos calcular a probabilidade de $x = 0$.

$$f(0) = P(X = 0) = \frac{e^{-1{,}25} \cdot 1{,}25^0}{0!} = 0{,}2865 = 28{,}65\%$$

Propriedades das distribuições binomial e de Poisson

Como temos modelos conhecidos, podemos verificar características de modo geral dessas variáveis. Podem ser calculados o valor esperado, a variância e o desvio-padrão dessas variáveis.

Quando temos uma variável aleatória discreta que se aproxima de uma distribuição binomial, podemos calcular o valor esperado da variável x, como sendo:

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

A variância dessa variável aleatória discreta sendo:

$$\sigma^2 = VAR(X) = n \cdot p \cdot q$$

Consequentemente, o desvio-padrão dessa variável aleatória discreta como sendo:

$$\sigma = \sqrt{n \cdot p \cdot q}$$

O mesmo pode ser feito para uma variável aleatória discreta que siga aproximadamente uma distribuição de Poisson. A média, ou valor esperado da variável aleatória x, é dada por:

$$\mu = E(X) = \lambda$$

A variância dessa variável aleatória discreta como:

$$\sigma^2 = VAR(X) = \lambda$$

Consequentemente, o desvio-padrão dessa variável aleatória discreta sendo:

$$\sigma = \sqrt{\lambda}$$

Referência

DOANE, D. P.; SEWARD, L. E. *Estatística aplicada à administração e economia*. 4. ed. Porto Alegre: Bookman, 2014.

Conteúdo: