

Distribuição Binomial, de Poisson e Hipergeométrica

Bem-vindos ao curso de Estatística Aplicada! Esta apresentação vai explorar três importantes distribuições de probabilidade que são fundamentais para entender e modelar diversos fenômenos aleatórios.

Vamos embarcar juntos nesta jornada pelo fascinante mundo da estatística, descobrindo como estas distribuições podem nos ajudar a compreender melhor o universo ao nosso redor através da matemática.



Revolucione! Seja THRIVE!
www.ia-labs.com.br

Objetivos



Compreender fundamentos teóricos

Desenvolver uma base sólida sobre os princípios matemáticos que regem as distribuições binomial, de Poisson e hipergeométrica



Analisar requisitos e aplicações

Identificar quando cada distribuição deve ser aplicada e como elas modelam diferentes fenômenos do mundo real



Desenvolver habilidades práticas

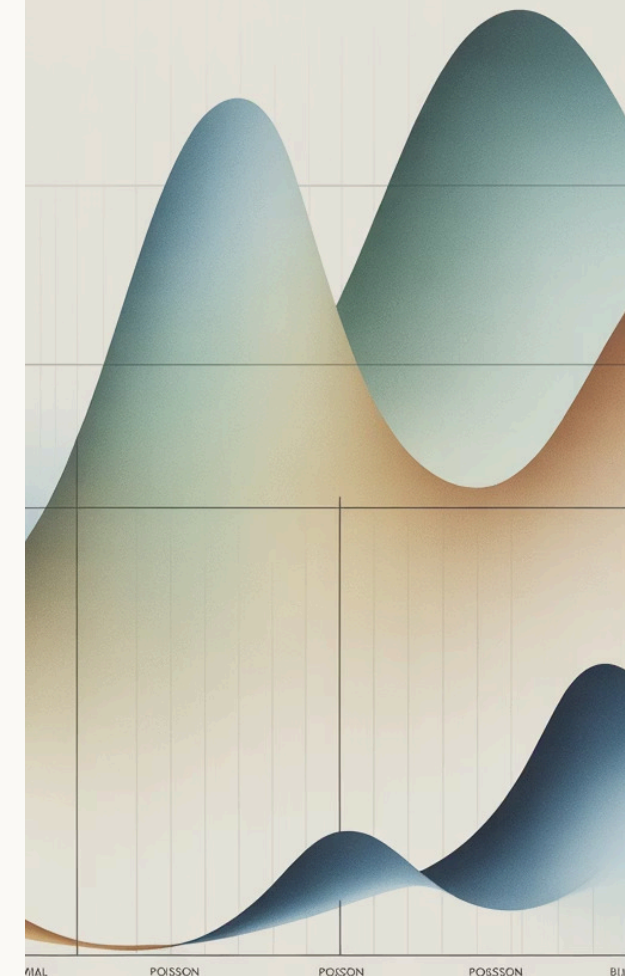
Aprender a calcular probabilidades usando fórmulas e ferramentas computacionais para resolver problemas reais



Interpretar resultados

Extrair conclusões significativas a partir dos cálculos probabilísticos e aplicá-las em situações práticas

PROBABILITY
5



Estrutura da Apresentação

Conceitos fundamentais de probabilidade

Iniciaremos com os princípios básicos que formam a base para o estudo das distribuições de probabilidade, garantindo que todos estejam familiarizados com os conceitos essenciais.

Distribuição Binomial

Exploraremos a distribuição que modela o número de sucessos em um número fixo de ensaios independentes, com aplicações em diversos campos do conhecimento.

Distribuição de Poisson

Estudaremos a distribuição que descreve a ocorrência de eventos raros em intervalos contínuos, fundamentais para modelar fenômenos como filas e defeitos.

Distribuição Hipergeométrica

Analisaremos a distribuição aplicada em situações de amostragem sem reposição, com importantes aplicações em controle de qualidade e auditoria.

Comparação e aplicações práticas

Finalizaremos com uma comparação abrangente entre as distribuições e exemplos práticos de aplicação em diversos campos profissionais.

Fundamentos de Probabilidade

Experimento Aleatório

Procedimento que, quando repetido sob as mesmas condições, pode resultar em diferentes resultados. Exemplos incluem o lançamento de dados, moedas ou a seleção aleatória de itens.

Espaço Amostral e Eventos

O espaço amostral (Ω) é o conjunto de todos os resultados possíveis de um experimento aleatório. Um evento é qualquer subconjunto do espaço amostral, representando um ou mais resultados de interesse.

Conceitos de Probabilidade

A probabilidade clássica é definida como a razão entre casos favoráveis e casos possíveis, enquanto a frequentista se baseia na frequência relativa observada após muitas repetições.

Variáveis Aleatórias

Função que associa um valor numérico a cada resultado do espaço amostral. Podem ser discretas (valores isolados) ou contínuas (qualquer valor em um intervalo).

Variáveis Aleatórias Discretas

Definição e Características

Uma variável aleatória discreta pode assumir apenas valores isolados e contáveis, como números inteiros. Exemplos incluem o número de estudantes em uma sala ou a contagem de eventos.

Estas variáveis são fundamentais no estudo das distribuições que veremos neste curso, pois todas as três distribuições (Binomial, Poisson e Hipergeométrica) são distribuições discretas.

Função de Probabilidade

A função de probabilidade $p(x)$ atribui a cada valor possível da variável X a probabilidade de ocorrência desse valor. Para qualquer valor x , temos:

- $p(x) \geq 0$ para todo x
- $\sum p(x) = 1$ para todos os valores possíveis

Esta função é a base para calcularmos probabilidades em distribuições discretas.

Esperança e Variância

A esperança matemática $E(X)$ representa o valor médio esperado da variável após muitas repetições, enquanto a variância $Var(X)$ mede a dispersão dos valores ao redor da média.

Estas medidas são essenciais para caracterizar distribuições de probabilidade e realizar análises estatísticas significativas.



Distribuições de Probabilidade



Importância no Estudo Estatístico

As distribuições de probabilidade são modelos matemáticos que descrevem o comportamento de variáveis aleatórias. Elas permitem prever a probabilidade de diferentes resultados e são fundamentais para a inferência estatística.



Modelagem de Fenômenos Aleatórios

Através das distribuições de probabilidade, conseguimos modelar matematicamente fenômenos incertos que ocorrem na natureza, sociedade, tecnologia e outras áreas do conhecimento.



Parâmetros Característicos

Cada distribuição possui parâmetros específicos que determinam sua forma e propriedades. Conhecer estes parâmetros é essencial para aplicar corretamente cada modelo em situações reais.



Aplicações em Problemas Reais

As distribuições que estudaremos têm aplicações práticas em diversos campos, como controle de qualidade, epidemiologia, finanças, engenharia e ciências sociais.

Understanding
Distribution

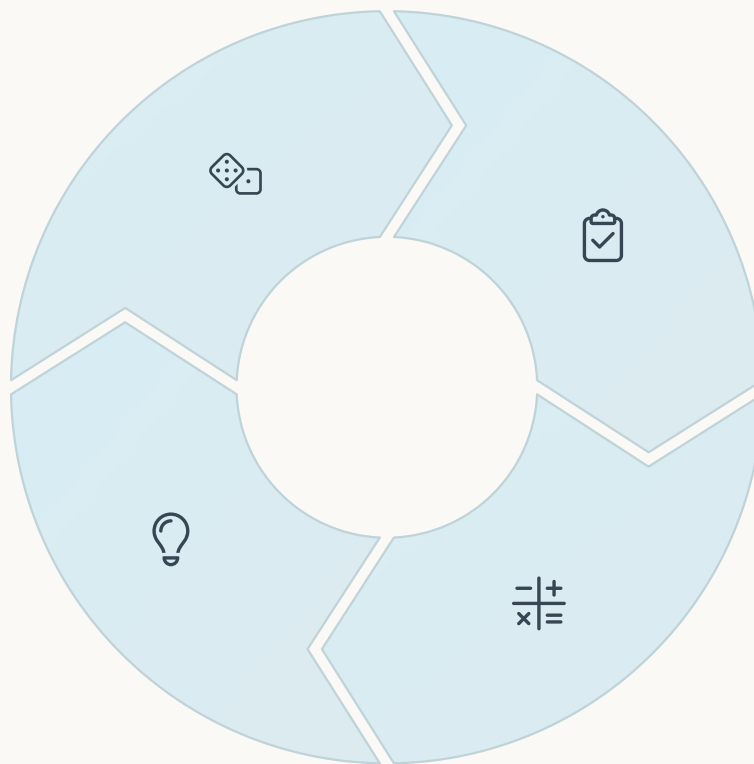
Distribuição Binomial: Introdução

Caracterização

A distribuição binomial modela o número de sucessos em um número fixo de ensaios independentes, onde cada ensaio resulta em apenas dois resultados possíveis

Exemplos

Lançamento de moedas, testes de controle de qualidade, questões de múltipla escolha e transmissão genética são exemplos clássicos de fenômenos binomiais

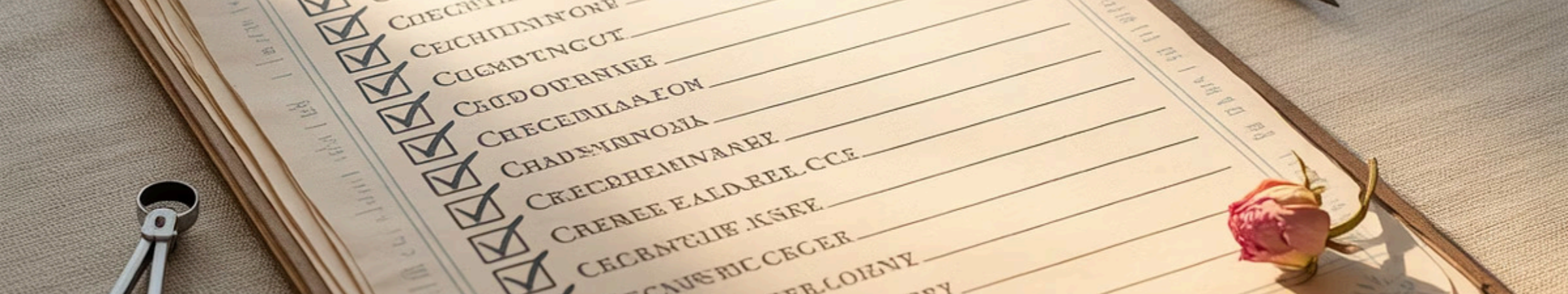


Requisitos

Exige ensaios independentes, número fixo de repetições (n), probabilidade constante de sucesso (p) e apenas dois resultados possíveis por ensaio

Notação

Representada como $X \sim B(n,p)$, onde X é a variável aleatória que conta o número de sucessos, n é o número de ensaios e p é a probabilidade de sucesso



Distribuição Binomial: Requisitos



Número fixo de ensaios (n)

O experimento deve consistir em um número pré-determinado de ensaios. Este número deve ser conhecido antes da realização do experimento e permanece constante em todas as análises.



Independência entre ensaios

O resultado de um ensaio não deve influenciar o resultado de outros ensaios. Cada repetição do experimento deve ocorrer de forma completamente independente das demais.



Dois resultados possíveis

Cada ensaio deve resultar em apenas duas possibilidades, geralmente chamadas de "sucesso" e "fracasso". Não pode haver resultados intermediários ou múltiplas categorias.



Probabilidade constante

A probabilidade p de sucesso deve permanecer a mesma em todos os ensaios. Esta probabilidade não pode variar entre as repetições do experimento.

Distribuição Binomial: Fórmula

Função de Probabilidade

A probabilidade de obter exatamente k sucessos em n ensaios é dada por:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{(n-k)}$$

Onde $\binom{n}{k}$ é o coeficiente binomial que representa o número de maneiras de selecionar k elementos em um conjunto de n elementos.

Coeficiente Binomial

O coeficiente binomial é calculado pela fórmula:

$$\binom{n}{k} = n! / [k!(n-k)!]$$

Onde $n!$ representa o fatorial de n , ou seja, o produto de todos os números inteiros de 1 até n .

Este termo representa todas as diferentes sequências possíveis de sucessos e fracassos com exatamente k sucessos.

Interpretação

A fórmula pode ser compreendida como:

- p^k : probabilidade de obter k sucessos
- $(1-p)^{(n-k)}$: probabilidade de obter $(n-k)$ fracassos
- $\binom{n}{k}$: número de maneiras de ordenar os k sucessos entre os n ensaios

Distribuição Binomial: Esperança e Variância

$$np$$

Valor Esperado

A média ou esperança da distribuição binomial é dada por $E(X) = np$, representando o número médio de sucessos esperados em n ensaios

$$np(1-p)$$

Variância

A variância mede a dispersão dos valores e é dada por $\text{Var}(X) = np(1-p)$, indicando como os resultados variam em torno da média

$$\sqrt{np(1-p)}$$

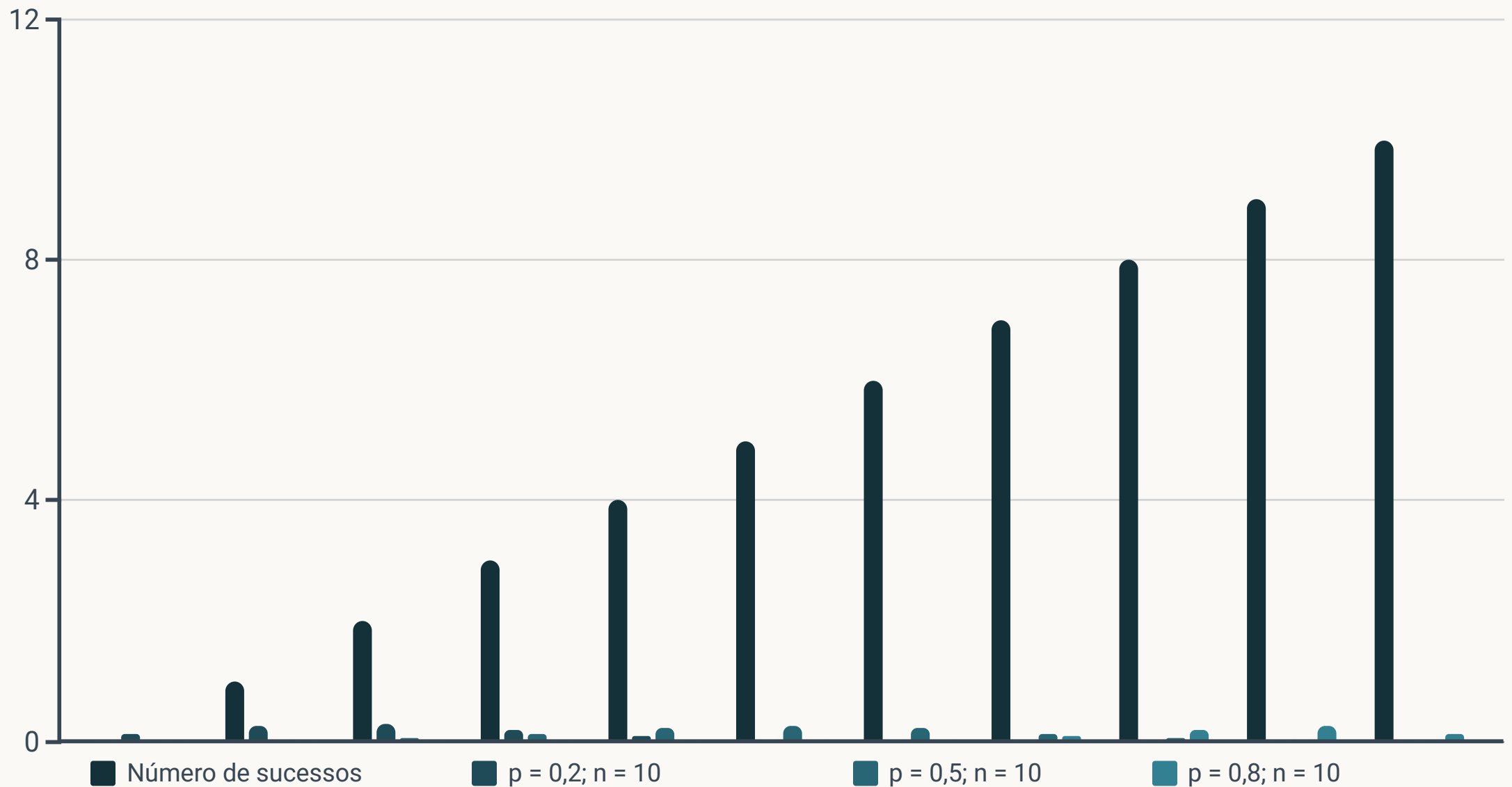
Desvio Padrão

O desvio padrão é a raiz quadrada da variância, fornecendo uma medida de dispersão na mesma unidade da variável original

A compreensão destes valores é fundamental para interpretar corretamente resultados em distribuições binomiais. O valor esperado indica o resultado médio após muitas repetições, enquanto a variância e o desvio padrão mostram quão dispersos estarão os resultados em torno deste valor médio.

Quando $p = 0,5$, a variância atinge seu valor máximo para um determinado n , indicando a maior incerteza possível sobre o resultado.

Distribuição Binomial: Histograma



O formato da distribuição binomial varia significativamente dependendo dos valores de n e p . Quando $p = 0,5$, a distribuição é perfeitamente simétrica, resultando em uma curva em forma de sino. Para valores de p menores que 0,5, como $p = 0,2$, a distribuição apresenta assimetria à direita. Inversamente, quando p é maior que 0,5, como $p = 0,8$, observamos uma assimetria à esquerda.

Distribuição Binomial: Exemplo Prático

?

Problema

Em uma linha de produção, sabe-se que 5% dos itens apresentam defeitos. Se inspecionarmos 20 itens aleatoriamente, qual a probabilidade de encontrarmos exatamente 2 itens defeituosos?

✓

Verificação dos requisitos

Temos $n = 20$ ensaios fixos, cada item é independente, há dois resultados possíveis (defeituoso ou não) e a probabilidade $p = 0,05$ é constante.



Cálculo

$$P(X = 2) = \binom{20}{2} \times (0,05)^2 \times (0,95)^{18} = 190 \times 0,0025 \times 0,3972 = 0,1891 \approx 18,91\%$$



Interpretação

Há aproximadamente 18,91% de chance de encontrarmos exatamente 2 itens defeituosos na amostra de 20 itens.



Distribuição Binomial: Aproximações

Aproximação pela Normal

Quando o número de ensaios (n) é grande, a distribuição binomial pode ser aproximada pela distribuição normal com média $\mu = np$ e desvio padrão $\sigma = \sqrt{np(1-p)}$.

Esta aproximação é particularmente útil quando os cálculos exatos se tornam computacionalmente intensivos devido a valores grandes de n .

CrITÉRIOS para Aproximação

Para que a aproximação seja válida, devem ser satisfeitas as condições:

- $np > 5$ (número esperado de sucessos maior que 5)
- $n(1-p) > 5$ (número esperado de fracassos maior que 5)

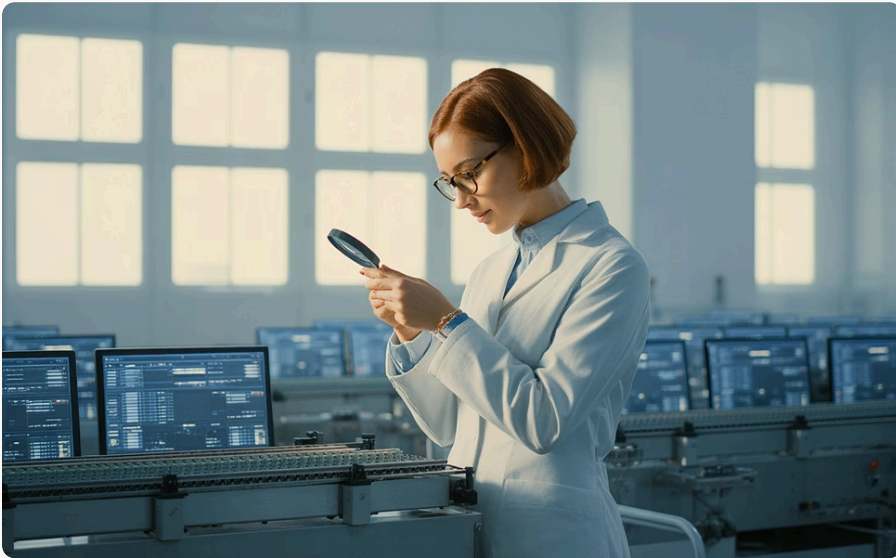
Quanto maiores forem estes valores, melhor será a aproximação.

Correção de Continuidade

Como a distribuição binomial é discreta e a normal é contínua, aplicamos a correção de continuidade para melhorar a precisão da aproximação.

Para calcular $P(X = k)$, usamos a probabilidade do intervalo $(k-0,5, k+0,5)$ na distribuição normal correspondente.

Distribuição Binomial: Aplicações



Controle Estatístico de Processos

A distribuição binomial é amplamente utilizada para monitorar a qualidade em linhas de produção, determinando se a proporção de itens defeituosos está dentro de limites aceitáveis. Isso permite intervenções rápidas quando o processo sai de controle.



Testes Genéticos e Biomedicina

Em genética, a distribuição binomial modela a transmissão de características hereditárias. Também é usada em testes médicos para calcular a probabilidade de falsos positivos ou falsos negativos em resultados de exames.



Pesquisas de Opinião

As pesquisas eleitorais e de mercado utilizam a distribuição binomial para estimar margens de erro e determinar tamanhos de amostra adequados, garantindo resultados estatisticamente significativos.

Distribuição Binomial: Exercícios

Exercício 1	Um estudante responde aleatoriamente a 10 questões de verdadeiro ou falso. Qual a probabilidade de acertar pelo menos 7 questões?
Exercício 2	Uma moeda é lançada 15 vezes. Qual a probabilidade de obter entre 5 e 10 caras, inclusive?
Exercício 3	Em um processo de fabricação, 10% das peças são defeituosas. Se 25 peças são inspecionadas, qual a probabilidade de encontrar no máximo 3 peças defeituosas?
Exercício 4	Uma prova de múltipla escolha tem 20 questões, cada uma com 5 alternativas. Se um aluno responde aleatoriamente todas as questões, qual a probabilidade de acertar exatamente 4 questões?
Exercício 5	Um teste para COVID-19 tem sensibilidade de 95%. Se 15 pessoas infectadas são testadas, qual a probabilidade de que pelo menos 13 testes sejam positivos?

Estes exercícios foram cuidadosamente selecionados para ajudá-lo a praticar os conceitos da distribuição binomial em diferentes contextos. Tente resolvê-los aplicando as fórmulas que aprendemos e interpretando os resultados no contexto de cada problema.

Lembre-se de verificar se os requisitos da distribuição binomial são satisfeitos em cada caso antes de aplicar as fórmulas correspondentes.

Distribuição de Poisson: Introdução



A distribuição de Poisson, desenvolvida pelo matemático francês Siméon Denis Poisson no século XIX, é uma das distribuições mais importantes para modelar eventos raros. Ela se aplica a situações onde contamos o número de ocorrências em um intervalo específico de tempo, espaço, volume ou área.

Diferentemente da distribuição binomial, que requer um número fixo de ensaios, a distribuição de Poisson pode modelar situações onde o número potencial de ocorrências é teoricamente ilimitado, mas a probabilidade de cada ocorrência individual é muito pequena.

Distribuição de Poisson: Requisitos

Independência entre Eventos

A ocorrência de um evento não deve afetar a probabilidade de ocorrência de outros eventos. Cada evento acontece de forma totalmente independente dos demais, sem qualquer correlação.

Por exemplo, no caso de chegadas em uma fila de banco, a chegada de um cliente não influencia diretamente a chegada do próximo.

Taxa Média Constante (λ)

A taxa média de ocorrência dos eventos, representada pelo parâmetro λ (lambda), deve permanecer constante ao longo de todo o intervalo considerado.

Por exemplo, se consideramos que em média 3 clientes chegam por hora em um caixa, essa taxa deve se manter constante durante todo o período de observação.

Eventos Raros

A probabilidade de ocorrência de dois ou mais eventos simultaneamente (no mesmo instante exato) deve ser desprezível ou nula.

Isso significa que os eventos ocorrem um de cada vez, e a probabilidade de ocorrências exatamente simultâneas é essencialmente zero.

Intervalo Fixo

A contagem dos eventos deve ser realizada em um intervalo fixo e bem definido, seja de tempo, espaço, área ou volume.

Por exemplo, podemos contar o número de chamadas em uma hora, defeitos em uma área de tecido, ou partículas em um volume de solução.

Distribuição de Poisson: Fórmula

$f(x)$

Função de Probabilidade

A probabilidade de ocorrerem exatamente k eventos em um intervalo é dada por:

$$P(X=k) = e^{(-\lambda)} \times \lambda^k / k!$$

Onde e é a base do logaritmo natural (aproximadamente 2,71828), λ é a taxa média de ocorrências no intervalo, e $k!$ é o fatorial de k .



Interpretação do Parâmetro λ

O parâmetro λ representa o número médio esperado de ocorrências no intervalo definido. Este valor pode ser estimado a partir de dados históricos ou especificado com base em conhecimento prévio.

Por exemplo, se em média ocorrem 2,5 acidentes por semana em um cruzamento, então $\lambda = 2,5$ para o intervalo de uma semana.



Exemplo de Cálculo

Se $\lambda = 3$ (média de 3 eventos por intervalo), a probabilidade de ocorrerem exatamente 2 eventos é:

$$P(X=2) = e^{(-3)} \times 3^2 / 2! = 0,0498 \times 9 / 2 = 0,224 \approx 22,4\%$$



Tabelas de Probabilidades

Existem tabelas disponíveis com valores pré-calculados da distribuição de Poisson para diferentes valores de λ e k , facilitando cálculos rápidos sem a necessidade de usar a fórmula completa.

Atualmente, softwares estatísticos e planilhas eletrônicas oferecem funções integradas para estes cálculos.

Distribuição de Poisson: Esperança e Variância

Valor Esperado

Na distribuição de Poisson, o valor esperado é exatamente igual ao parâmetro λ :

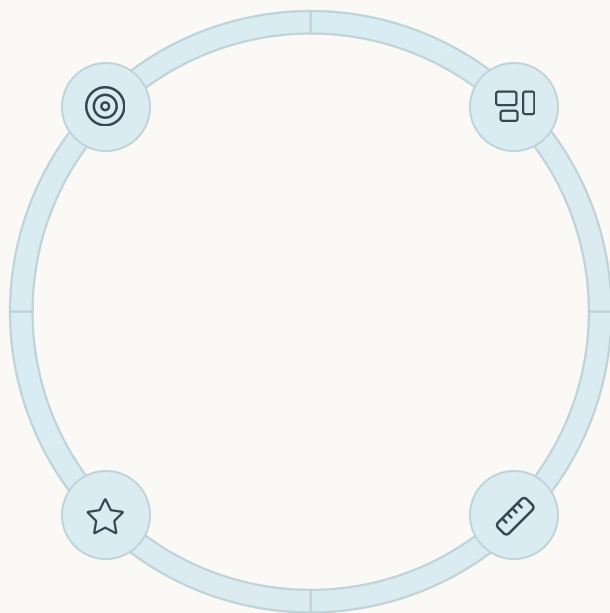
$$E(X) = \lambda$$

Isso significa que, em média, esperamos observar λ eventos no intervalo considerado.

Propriedade Especial

A igualdade entre média e variância ($E(X) = \text{Var}(X) = \lambda$) é uma característica única da distribuição de Poisson.

Esta propriedade pode ser usada como teste para verificar se um conjunto de dados segue uma distribuição de Poisson.



Variância

Uma propriedade notável da distribuição de Poisson é que sua variância também é igual ao parâmetro λ :

$$\text{Var}(X) = \lambda$$

Isso difere da maioria das outras distribuições, onde a variância depende de múltiplos parâmetros.

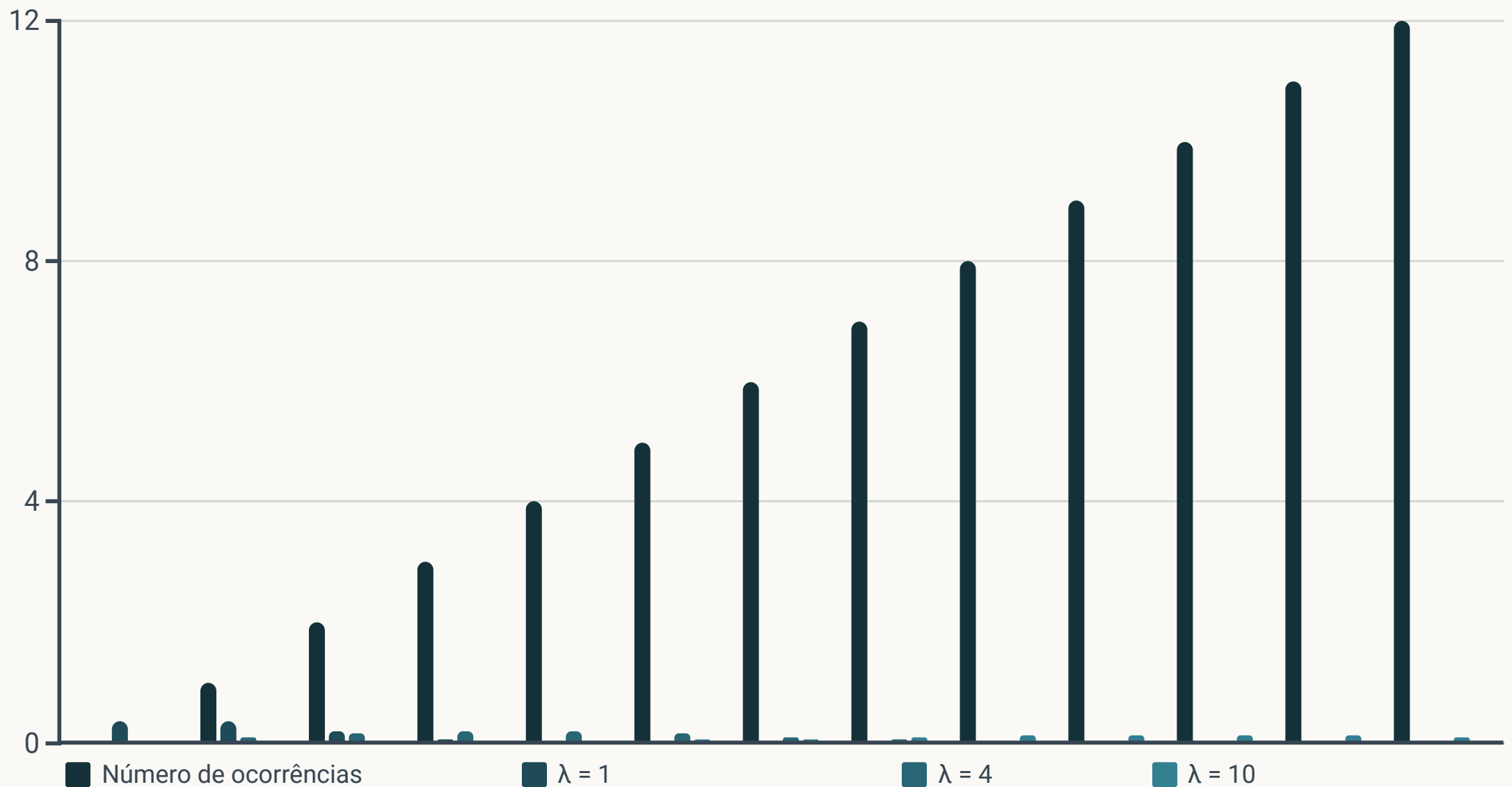
Desvio Padrão

O desvio padrão é simplesmente a raiz quadrada da variância:

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

Quanto maior o valor de λ , maior será a dispersão absoluta dos valores observados.

Distribuição de Poisson: Histograma



O formato da distribuição de Poisson varia significativamente com o valor do parâmetro λ . Para valores pequenos de λ (como $\lambda = 1$), a distribuição apresenta forte assimetria à direita, com concentração de probabilidade nos valores menores. À medida que λ aumenta (como $\lambda = 4$), a distribuição começa a se tornar mais simétrica. Para valores grandes de λ (como $\lambda = 10$), a distribuição se aproxima de uma distribuição normal, tornando-se quase simétrica.

Distribuição de Poisson: Exemplo Prático



Problema de fila bancária

Um banco recebe em média 12 clientes por hora em determinado caixa. Qual a probabilidade de chegarem exatamente 8 clientes durante uma hora específica?

2

Verificação dos requisitos

As chegadas são independentes, ocorrem em taxa constante, raramente acontecem simultaneamente e estamos analisando um intervalo fixo de uma hora.

3

Cálculo

Com $\lambda = 12$, calculamos: $P(X = 8) = e^{(-12)} \times 12^8 / 8! = 0,000006 \times 429.981.696 / 40.320 = 0,0640 \approx 6,4\%$



Interpretação

Há aproximadamente 6,4% de chance de chegarem exatamente 8 clientes durante uma hora específica neste caixa.

Distribuição de Poisson: Aproximação à Binomial

Condições para Aproximação

A distribuição de Poisson pode ser usada como uma aproximação da distribuição binomial quando:

- O número de ensaios n é grande (geralmente $n \geq 20$)
- A probabilidade p de sucesso é pequena (geralmente $p \leq 0,05$)
- O produto $np = \lambda$ é moderado (geralmente $np \leq 10$)

Nestas condições, podemos substituir os cálculos binomiais pelos mais simples cálculos de Poisson.

Relação Matemática

A distribuição de Poisson surge como o limite da distribuição binomial quando n tende ao infinito e p tende a zero, de modo que o produto np se mantém constante e igual a λ .

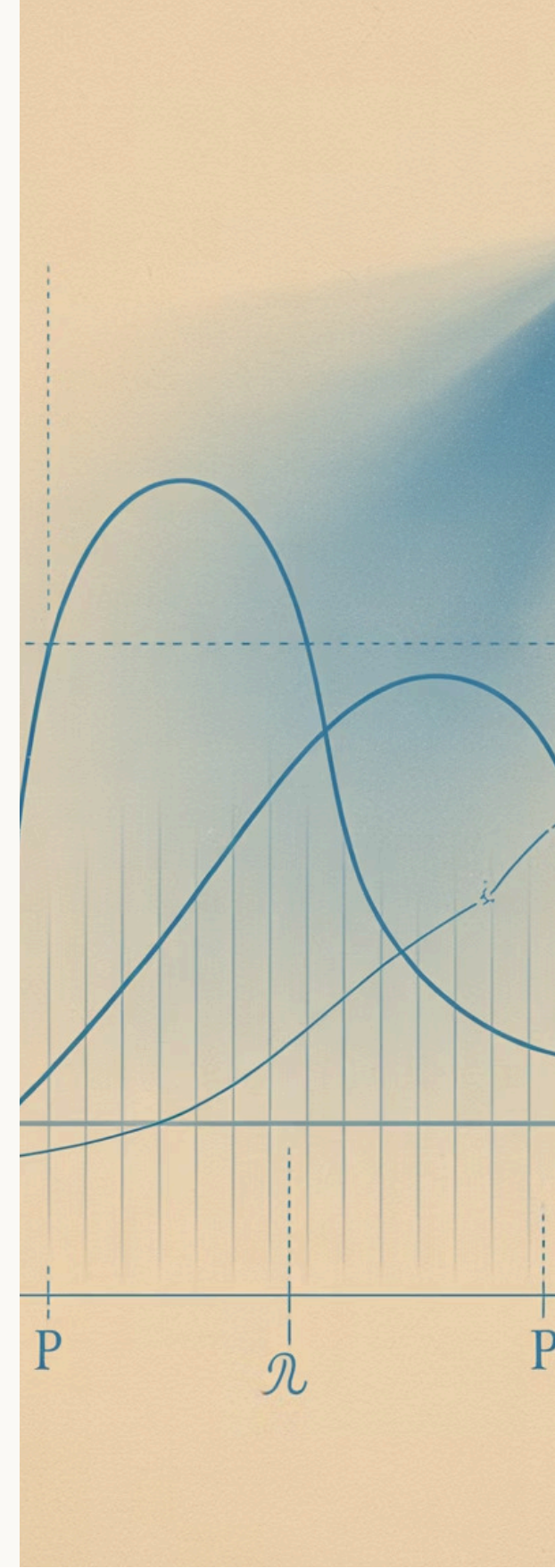
Esta relação permite substituir cálculos binomiais complexos (com n grande) por cálculos de Poisson mais simples, utilizando $\lambda = np$.

Exemplo Comparativo

Considere um processo com $n = 1000$ e $p = 0,003$, onde queremos calcular $P(X = 4)$:

- Cálculo Binomial: $(1000 \choose 4) \times 0,003^4 \times 0,997^{996} = 0,1755$
- Aproximação Poisson: $e^{(-3)} \times 3^4 / 4! = 0,1680$

O erro de aproximação neste caso é de apenas 0,0075, demonstrando a precisão da aproximação.



Distribuição de Poisson: Processo de Poisson



Definição

Um processo de Poisson é um modelo matemático que descreve a ocorrência de eventos discretos ao longo do tempo, onde o número de eventos em qualquer intervalo segue uma distribuição de Poisson.



Propriedades

Os eventos ocorrem de forma independente, a taxa média de ocorrência é constante, e a probabilidade de dois eventos simultâneos é nula. Além disso, o número de eventos em intervalos disjuntos são variáveis independentes.



Intervalos

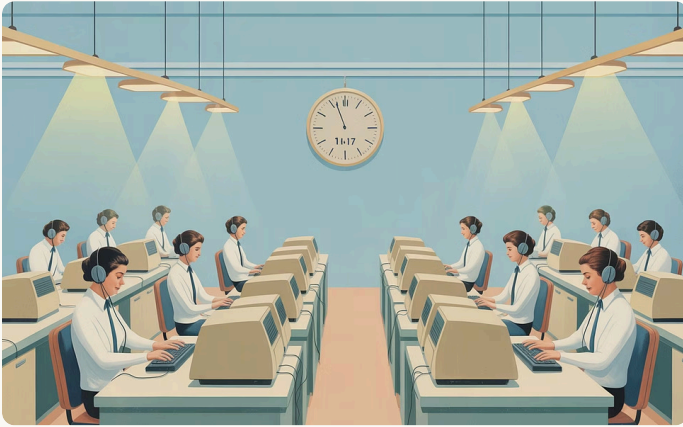
Em um processo de Poisson, o tempo entre eventos consecutivos segue uma distribuição exponencial. Isso significa que a probabilidade de esperar pelo próximo evento diminui exponencialmente com o tempo.



Distribuição Exponencial

Se T representa o tempo até a próxima ocorrência, então T segue uma distribuição exponencial com parâmetro λ . A função de densidade de probabilidade é $f(t) = \lambda e^{(-\lambda t)}$ para $t \geq 0$.

Distribuição de Poisson: Aplicações



Telecomunicações

Nas redes de telecomunicações, a distribuição de Poisson modela o número de chamadas recebidas por minuto em centrais telefônicas, permitindo o dimensionamento adequado de recursos e pessoal para atendimento.



Controle de Qualidade

Na indústria, a distribuição de Poisson é utilizada para modelar o número de defeitos em produtos, estabelecendo limites de controle para processos e definindo critérios para aceitação ou rejeição de lotes.



Epidemiologia

Na área da saúde, a distribuição de Poisson ajuda a modelar a ocorrência de casos raros de doenças, como malformações congênitas, eventos adversos de medicamentos ou surtos epidêmicos em determinadas regiões.

Distribuição de Poisson: Exercícios

Exercício 1	Uma central telefônica recebe em média 3,5 chamadas por minuto. Qual a probabilidade de receber exatamente 5 chamadas em um minuto específico?
Exercício 2	Um hospital registra em média 2,7 nascimentos por dia. Qual a probabilidade de que em um determinado dia ocorram pelo menos 4 nascimentos?
Exercício 3	Em média, uma livraria vende 8 livros de estatística por semana. Qual a probabilidade de vender no máximo 5 livros em uma semana específica?
Exercício 4	O número médio de acidentes em um determinado cruzamento é de 1,2 por mês. Qual a probabilidade de que não ocorram acidentes em um mês específico?
Exercício 5	Um tecido apresenta em média 0,5 defeitos por metro quadrado. Se inspecionarmos uma peça de 4 metros quadrados, qual a probabilidade de encontrarmos exatamente 3 defeitos?

Estes exercícios foram elaborados para que você possa praticar os conceitos da distribuição de Poisson em diferentes contextos. Aplique a fórmula $P(X=k) = e^{(-\lambda)} \times \lambda^k / k!$ para resolver cada problema e interprete os resultados no contexto específico.

Lembre-se de verificar se todos os requisitos da distribuição de Poisson são satisfeitos em cada situação antes de aplicar as fórmulas.

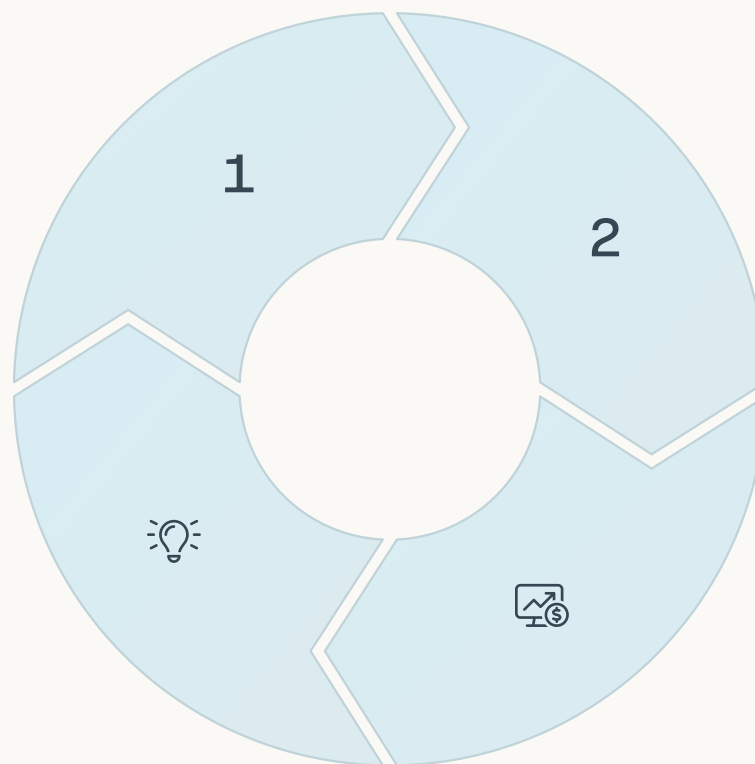
Distribuição Hipergeométrica: Introdução

Amostragem sem Reposição

A distribuição hipergeométrica modela experimentos de amostragem sem reposição de uma população finita, onde cada item retirado altera a composição da população

Exemplos Práticos

Aplicada em controle de qualidade com lotes finitos, seleção de cartas de baralho, auditorias financeiras e processos seletivos



Notação Matemática

Representada como $X \sim H(N, K, n)$, onde N é o tamanho da população, K é o número de itens com a característica de interesse, e n é o tamanho da amostra

Comparação com Binomial

Difere da binomial principalmente pela dependência entre os ensaios, já que a remoção de um item altera a probabilidade para as retiradas subsequentes

Distribuição Hipergeométrica:

Requisitos



População finita de tamanho N

A população total deve ter um tamanho fixo e conhecido N . Todos os itens da população devem ser identificáveis, e o tamanho total não pode mudar durante o experimento.



K elementos com característica de interesse

Dentro da população, exatamente K elementos possuem a característica que estamos interessados em observar. Os demais $(N-K)$ elementos não possuem essa característica.



Amostra de tamanho n retirada sem reposição

Retiramos n elementos da população (com $n \leq N$), e cada elemento retirado não é devolvido à população antes da próxima retirada. Isso diferencia fundamentalmente esta distribuição da binomial.

4

Não-independência entre ensaios

Como os itens são retirados sem reposição, cada seleção altera a composição da população restante, tornando os ensaios dependentes. A probabilidade de sucesso muda após cada retirada.



Distribuição Hipergeométrica: Fórmula

Função de Probabilidade

A probabilidade de obter exatamente k sucessos em uma amostra de tamanho n é dada por:

$$P(X=k) = \frac{[(K \ k)(N-K \ n-k)]}{(N \ n)}$$

Onde $(K \ k)$ representa o coeficiente binomial, ou seja, o número de maneiras de selecionar k elementos entre K elementos.

Interpretação dos Parâmetros

N = tamanho total da população

K = número de elementos com a característica de interesse

n = tamanho da amostra retirada

k = número de sucessos observados na amostra

A fórmula calcula a proporção entre o número de maneiras de obter k sucessos e o número total de possíveis amostras.

Diferenças em Relação à Binomial

Na distribuição hipergeométrica, a probabilidade de sucesso muda a cada retirada, enquanto na binomial permanece constante.

Quando N é muito grande em relação a n , a distribuição hipergeométrica se aproxima da binomial com $p = K/N$, pois o efeito da não reposição torna-se desprezível.

Distribuição Hipergeométrica: Esperança e Variância

1

Valor Esperado

A média ou valor esperado de uma distribuição hipergeométrica é o produto do tamanho da amostra pela proporção de sucessos na população

2

Variância

A fórmula inclui um fator de correção para população finita que reduz a variância em comparação com a distribuição binomial

3

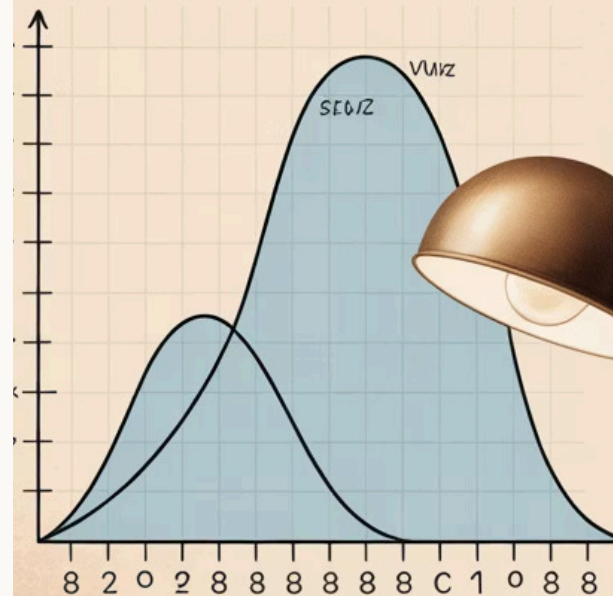
Fator de Correção

Este fator é sempre menor que 1 e representa o efeito da amostragem sem reposição na redução da variabilidade

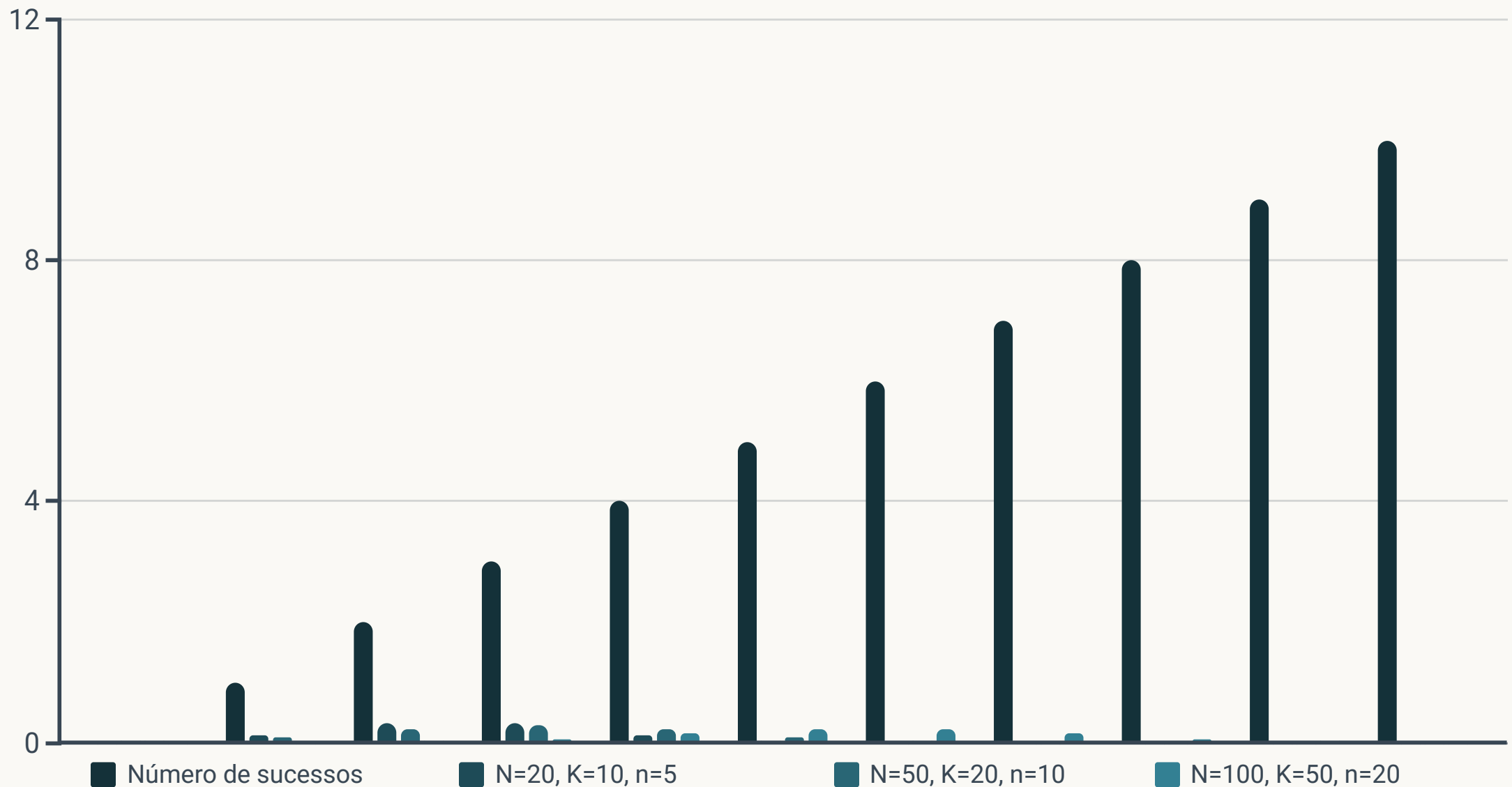
O valor esperado da distribuição hipergeométrica é similar ao da distribuição binomial, substituindo p por K/N . No entanto, a variância é multiplicada pelo fator de correção para população finita $(N-n)/(N-1)$, que reduz a dispersão.

Quanto maior a proporção da população que é amostrada (n/N), menor será a variância em comparação com a binomial equivalente. Isso reflete o fato de que a amostragem sem reposição proporciona mais informação sobre a população do que a amostragem com reposição.

Hypergeometric probability distribution



Distribuição Hipergeométrica: Histograma



O formato da distribuição hipergeométrica é influenciado significativamente pelos valores dos parâmetros N, K e n. Quando a razão K/N é próxima de 0,5 (como em todos os exemplos acima), a distribuição tende a ser simétrica em torno de seu valor esperado $n \cdot K/N$. Conforme N aumenta em relação a n, a distribuição se aproxima de uma binomial com $p = K/N$.

Distribuição Hipergeométrica: Exemplo Prático



Problema de controle de qualidade

Um lote contém 50 peças, das quais 5 são defeituosas. Se um inspetor seleciona aleatoriamente 10 peças (sem reposição), qual a probabilidade de encontrar exatamente 2 peças defeituosas?

2

Verificação dos requisitos

Temos uma população finita ($N = 50$), com $K = 5$ peças defeituosas, e uma amostra de $n = 10$ peças retiradas sem reposição.

3

Cálculo

$$P(X = 2) = \frac{[(5 \cdot 2)(45 \cdot 8)]}{(50 \cdot 10)} = \frac{[10 \times 14.872.858]}{10.272.278.170} = 0,1449 \approx 14,49\%$$



Interpretação

Há aproximadamente 14,49% de chance de o inspetor encontrar exatamente 2 peças defeituosas na amostra de 10 peças selecionadas.

Distribuição Hipergeométrica: Aproximação à Binomial

Condições para Aproximação

A distribuição hipergeométrica pode ser aproximada pela distribuição binomial quando:

- O tamanho da população N é grande em relação ao tamanho da amostra n
- A fração da população amostrada (n/N) é pequena, geralmente menor que 10%

Nestas condições, o efeito da amostragem sem reposição torna-se desprezível, e podemos usar a binomial para simplificar os cálculos.

Parâmetros da Aproximação

Se usarmos a distribuição binomial para aproximar a hipergeométrica, devemos definir:

- O número de ensaios n (o mesmo tamanho de amostra)
- A probabilidade de sucesso $p = K/N$ (a proporção de itens com a característica de interesse na população original)

Assim, a distribuição $H(N, K, n)$ pode ser aproximada por $B(n, K/N)$.

Erro de Aproximação

O erro de aproximação diminui conforme N aumenta em relação a n . Para a maioria das aplicações práticas, quando $n/N \leq 0,05$, o erro é negligenciável.

A principal fonte de erro vem da diferença de variância entre as duas distribuições, devido ao fator de correção para população finita que existe na hipergeométrica mas não na binomial.

Distribuição Hipergeométrica: Aplicações

Controle de Qualidade

A distribuição hipergeométrica é amplamente utilizada em procedimentos de inspeção por amostragem, onde lotes finitos de produtos são avaliados retirando-se uma amostra sem reposição. Esta abordagem permite determinar a probabilidade de aceitação ou rejeição de lotes com base no número de defeitos encontrados.

Auditoria Financeira

Auditores utilizam a distribuição hipergeométrica para determinar o tamanho da amostra necessária para verificar transações financeiras com um nível de confiança específico. Isto permite estimar a probabilidade de detectar irregularidades quando apenas uma parte dos registros é examinada.

Estudos Ecológicos

No método de captura-recaptura utilizado em ecologia para estimar o tamanho de populações animais, a distribuição hipergeométrica modela a probabilidade de recapturar animais marcados. Este método é crucial para o monitoramento de espécies e conservação da biodiversidade.



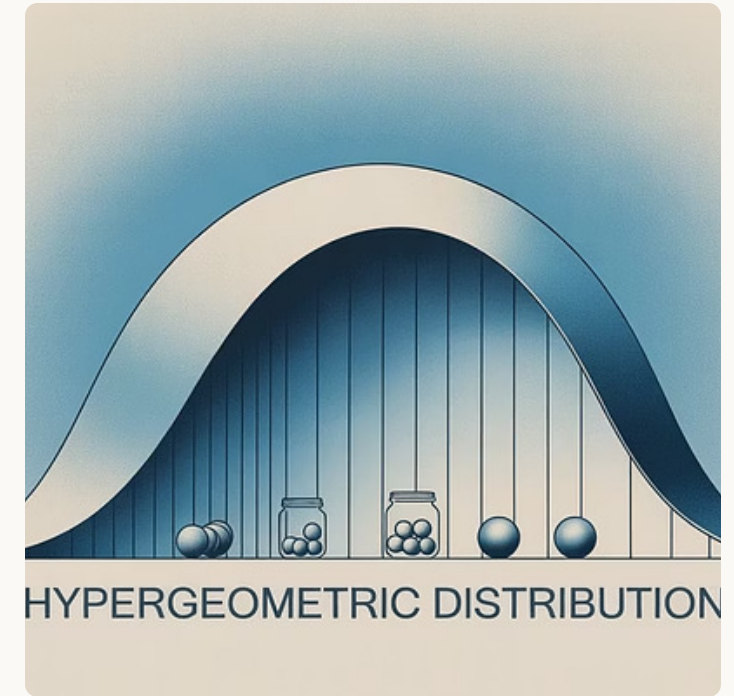
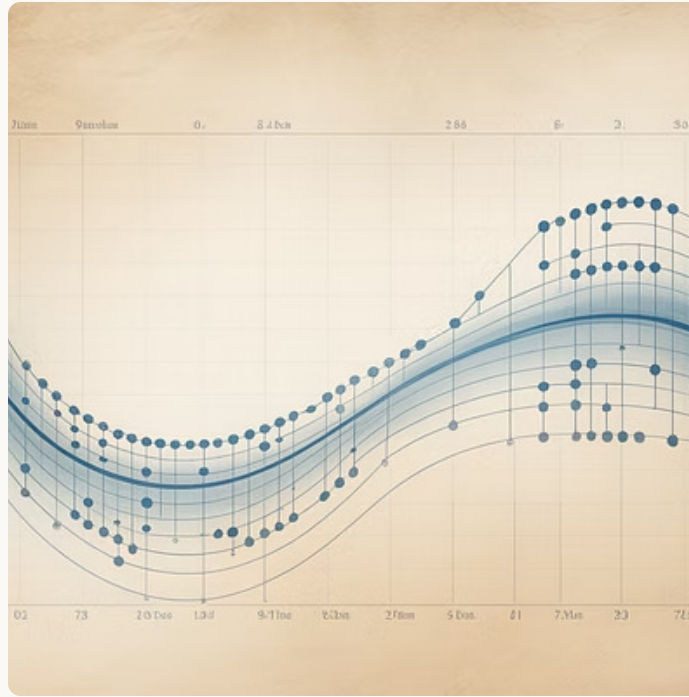
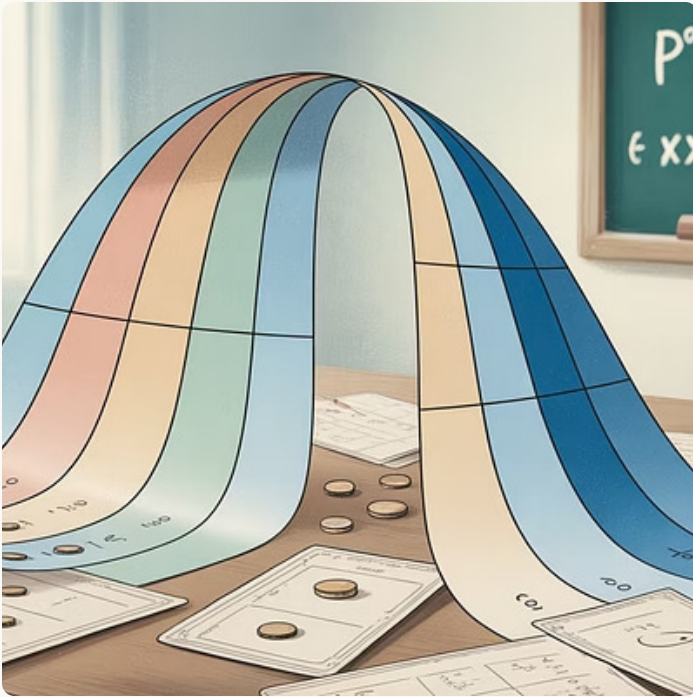
Distribuição Hipergeométrica: Exercícios

Exercício 1	Um lote contém 100 itens, dos quais 8 são defeituosos. Se inspecionarmos 20 itens aleatoriamente (sem reposição), qual a probabilidade de encontrarmos exatamente 3 itens defeituosos?
Exercício 2	Em uma turma de 30 alunos, 12 são meninas. Se selecionarmos aleatoriamente 5 alunos para uma apresentação, qual a probabilidade de que exatamente 2 sejam meninas?
Exercício 3	Uma urna contém 10 bolas brancas e 15 bolas pretas. Se retirarmos 8 bolas sem reposição, qual a probabilidade de obtermos pelo menos 6 bolas brancas?
Exercício 4	Uma caixa contém 200 peças, das quais 15 não atendem às especificações. Se um cliente compra 25 peças aleatoriamente, qual a probabilidade de que no máximo 1 peça não atenda às especificações?
Exercício 5	Em um baralho de 52 cartas, 13 são copas. Se retirarmos 5 cartas sem reposição, qual a probabilidade de obtermos exatamente 2 cartas de copas?

Estes exercícios foram selecionados para ajudá-lo a praticar os conceitos da distribuição hipergeométrica em diferentes contextos. Aplique a fórmula $P(X=k) = \frac{[(K \ k)(N-K \ n-k)]}{(N \ n)}$ para resolver cada problema e interprete os resultados dentro do contexto específico.

Lembre-se de verificar se os requisitos da distribuição hipergeométrica são satisfeitos em cada caso antes de aplicar as fórmulas correspondentes.

Comparação entre as Distribuições



As três distribuições que estudamos modelam fenômenos discretos, mas com características distintas. A binomial é ideal para experimentos com número fixo de ensaios independentes e probabilidade constante de sucesso. A distribuição de Poisson modela eventos raros em intervalos contínuos, com taxa média constante. Já a hipergeométrica aplica-se à amostragem sem reposição de populações finitas.

A escolha entre estas distribuições deve considerar cuidadosamente as características do fenômeno estudado. Em alguns casos, uma distribuição pode aproximar outra sob certas condições, como a Poisson aproximando a binomial quando n é grande e p pequeno, ou a binomial aproximando a hipergeométrica quando N é muito maior que n .

Tabela Comparativa

Característica	Binomial	Poisson	Hipergeométrica
Notação	$X \sim B(n, p)$	$X \sim P(\lambda)$	$X \sim H(N, K, n)$
Requisitos	n ensaios fixos, independência, dois resultados, p constante	Eventos raros, independentes, λ constante, intervalo fixo	População finita N, K elementos de interesse, sem reposição
Função de probabilidade	$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$	$P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!}$	$P(X=k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$
Esperança	$E(X) = np$	$E(X) = \lambda$	$E(X) = n \cdot K/N$
Variância	$Var(X) = np(1-p)$	$Var(X) = \lambda$	$Var(X) = n \cdot K/N \cdot (N-K)/N \cdot (N-n)/(N-1)$
Aplicações típicas	Testes de múltipla escolha, controle de qualidade	Filas, defeitos, eventos raros no tempo/espaço	Amostragem sem reposição, auditoria, controle de qualidade

Esta tabela resume as principais características das três distribuições estudadas, facilitando a comparação direta entre elas. Observe como as fórmulas de esperança da binomial e da hipergeométrica são similares, com K/N na hipergeométrica equivalendo a p na binomial. A principal diferença está na variância, onde a hipergeométrica inclui um fator de correção para população finita.

Aplicações Multidisciplinares



Ciências Biológicas e Saúde

Nas ciências da saúde, estas distribuições são usadas para modelar transmissão de doenças (Poisson), eficácia de tratamentos (binomial) e estudos genéticos (hipergeométrica). Epidemiologistas utilizam a distribuição de Poisson para analisar a ocorrência de casos raros, enquanto ensaios clínicos empregam a binomial para testar eficácia de medicamentos.



Engenharia e Controle de Qualidade

Na indústria, a distribuição binomial e hipergeométrica são fundamentais para planos de amostragem e controle estatístico de processos. A distribuição de Poisson modela a ocorrência de defeitos em materiais contínuos como tecidos e chapas, enquanto a hipergeométrica é ideal para inspeção de lotes finitos.



Ciências Sociais e Pesquisas

Pesquisadores sociais utilizam a distribuição binomial para modelar respostas em enquetes e questionários. Estudos demográficos aplicam a distribuição de Poisson para analisar eventos raros como crimes específicos ou acidentes. A distribuição hipergeométrica é útil em auditorias e verificações de conformidade em amostras populacionais.



Finanças e Análise de Risco

No setor financeiro, estas distribuições modelam eventos como inadimplência (Poisson), sucesso de investimentos (binomial) e auditoria de transações (hipergeométrica). Seguradoras utilizam a distribuição de Poisson para modelar a frequência de sinistros, essencial para precificação de apólices.

Software Estatístico: R

Distribuição Binomial

Para calcular probabilidades binomiais no R:

- **dbinom(x, size, prob)**: Calcula $P(X = x)$
- **pbinom(q, size, prob)**: Calcula $P(X \leq q)$
- **qbinom(p, size, prob)**: Retorna o valor x tal que $P(X \leq x) = p$
- **rbinom(n, size, prob)**: Gera n valores aleatórios

Exemplo: **dbinom(3, 10, 0.5)** calcula $P(X = 3)$ em $B(10, 0.5)$

Distribuição de Poisson

Para calcular probabilidades de Poisson no R:

- **dpois(x, lambda)**: Calcula $P(X = x)$
- **ppois(q, lambda)**: Calcula $P(X \leq q)$
- **qpois(p, lambda)**: Retorna o valor x tal que $P(X \leq x) = p$
- **rpois(n, lambda)**: Gera n valores aleatórios

Exemplo: **ppois(2, 1.5)** calcula $P(X \leq 2)$ em $P(1.5)$

Distribuição Hipergeométrica

Para calcular probabilidades hipergeométricas no R:

- **dhyper(x, m, n, k)**: Calcula $P(X = x)$
- **phyper(q, m, n, k)**: Calcula $P(X \leq q)$
- **qhyper(p, m, n, k)**: Retorna o valor x tal que $P(X \leq x) = p$
- **rhyper(nn, m, n, k)**: Gera nn valores aleatórios

Onde $m = K$, $n = N - K$, $k =$ tamanho da amostra

Exemplo: **dhyper(2, 10, 40, 8)** calcula $P(X = 2)$ em $H(50, 10, 8)$

Software Estatístico: Excel

Funções para Distribuição Binomial

O Excel oferece as seguintes funções para cálculos com a distribuição binomial:

- **DIST.BINOM(num_s; ensaios; prob_s; cumulativo):** Retorna a probabilidade de um evento binomial
- **DIST.BINOM.N(num_s; ensaios; prob_s; cumulativo):** Versão mais recente da função

Se cumulativo = FALSO, calcula $P(X = \text{num_s})$. Se cumulativo = VERDADEIRO, calcula $P(X \leq \text{num_s})$.

Funções para Distribuição de Poisson

Para a distribuição de Poisson, o Excel disponibiliza:

- **POISSON(x; média; cumulativo):** Função legada para cálculos de Poisson
- **POISSON.DIST(x; média; cumulativo):** Versão atual da função

Se cumulativo = FALSO, calcula $P(X = x)$. Se cumulativo = VERDADEIRO, calcula $P(X \leq x)$.

Funções para Distribuição Hipergeométrica

Para cálculos com a distribuição hipergeométrica:

- **DIST.HIPERGEOM(amostra_s; num_amostra; pop_s; num_pop; cumulativo):** Versão atual
- **DIST.HIPERGEOM.N(amostra_s; num_amostra; pop_s; num_pop; cumulativo):** Versão mais recente

Onde amostra_s = sucessos na amostra, num_amostra = tamanho da amostra, pop_s = sucessos na população, num_pop = tamanho da população.

Estudo de Caso: Epidemiologia

Modelagem de Transmissão

A distribuição de Poisson é amplamente utilizada para modelar a ocorrência de novos casos de doenças raras ou emergentes. Durante pandemias, epidemiologistas analisam o número diário de novos casos usando modelos baseados nesta distribuição.

Modelagem de Surtos

Para eventos extremos, como surtos com 10 ou mais casos em um único dia, o cálculo seria $P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0,9972 = 0,0028$ ou apenas 0,28%, indicando um evento altamente incomum que pode sinalizar um surto.



Análise de Dados

Considerando dados de uma região onde se observa em média $\lambda = 3,5$ novos casos de uma doença por dia, podemos calcular a probabilidade de observar exatamente 5 novos casos em um dia específico: $P(X = 5) = e^{(-3,5)} \times 3,5^5 / 5! \approx 0,1338$ ou 13,38%.

Aplicações Práticas

Estes cálculos auxiliam autoridades sanitárias a estabelecer limiares estatísticos para detecção precoce de surtos e implementação de medidas de controle. A modelagem de Poisson também ajuda a dimensionar recursos hospitalares e de saúde pública.

Estudo de Caso: Controle de Qualidade



Considere um lote de 1000 componentes eletrônicos, onde a especificação exige no máximo 2% de itens defeituosos. Um plano de amostragem indica que 80 itens devem ser inspecionados, e o lote será aceito se forem encontrados no máximo 3 itens defeituosos.

Usando a distribuição hipergeométrica, podemos calcular a probabilidade de aceitar um lote com diferentes percentuais de itens defeituosos. Por exemplo, para um lote com 2% de defeituosos (20 itens), a probabilidade de aceitação é aproximadamente 92,1%. Já para um lote com 5% de defeituosos (50 itens), essa probabilidade cai para 39,7%.

Estudo de Caso: Pesquisas Eleitorais



Uso da Distribuição Binomial

As pesquisas eleitorais utilizam a distribuição binomial para modelar a proporção de eleitores que preferem determinado candidato. Cada entrevistado representa um ensaio independente, onde a resposta favorável ao candidato é considerada um "sucesso".

2

Cálculo de Margens de Erro

A margem de erro é calculada a partir do desvio padrão da distribuição binomial. Para uma amostra de tamanho n e proporção estimada p , a margem de erro para um nível de confiança de 95% é aproximadamente $1,96 \times \sqrt{p(1-p)/n}$.



Intervalos de Confiança

Para uma pesquisa com 1000 entrevistados onde 54% declararam intenção de voto em determinado candidato, o intervalo de confiança de 95% seria aproximadamente $54\% \pm 3,1\%$, ou seja, entre 50,9% e 57,1%.



Análise de Resultados

Se dois candidatos aparecem com 48% e 52% em uma pesquisa com margem de erro de $\pm 3\%$, não podemos afirmar com certeza estatística que um está à frente, pois os intervalos de confiança se sobrepõem significativamente.

Exercícios Práticos Integrados

Problema 1	Uma fábrica produz lotes de 500 peças, com taxa histórica de 2% de defeitos. A inspeção de qualidade seleciona 50 peças sem reposição. Se encontrar mais de 3 defeitos, o lote é rejeitado. Calcule a probabilidade de rejeição de um lote com exatamente 2% de peças defeituosas.
Problema 2	Em uma central de atendimento, as chamadas chegam em média a uma taxa de 4 por minuto. Qual a probabilidade de que, em um intervalo de 30 segundos, cheguem pelo menos 3 chamadas? Se aumentarmos o intervalo para 2 minutos, como muda essa probabilidade?
Problema 3	Uma vacina tem eficácia de 92% contra determinada doença. Se 15 pessoas vacinadas são expostas ao vírus, qual a probabilidade de que pelo menos 12 fiquem protegidas? Se a amostra fosse de 50 pessoas, como mudaria essa probabilidade?
Problema 4	Uma urna contém 10 bolas vermelhas e 15 azuis. Se retiramos 8 bolas sem reposição, qual a probabilidade de obter exatamente 3 bolas vermelhas? Compare este resultado com a probabilidade calculada usando a distribuição binomial com $p = 10/25$ e explique a diferença.

Estes problemas integrados foram elaborados para ajudá-lo a desenvolver habilidades de análise e tomada de decisão baseadas em dados probabilísticos. Eles combinam conceitos das diferentes distribuições que estudamos e exigem raciocínio crítico para identificar qual modelo é mais apropriado em cada situação.

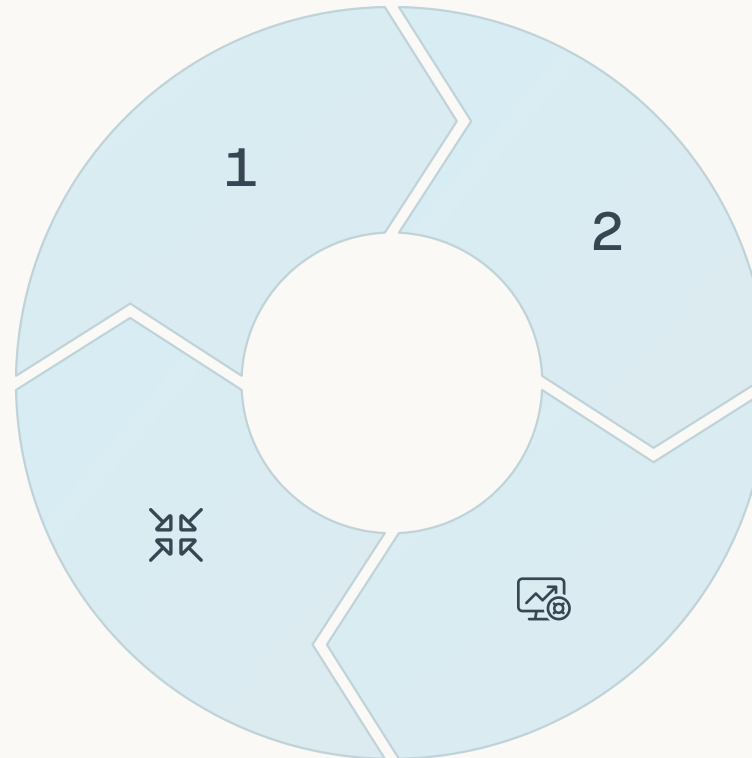
Simulação Computacional

Geração de Números Aleatórios

A simulação computacional utiliza geradores de números aleatórios para produzir dados que seguem as distribuições de probabilidade estudadas

Análise de Convergência

Simulações demonstram como as distribuições se aproximam de outras sob certas condições, como a binomial convergindo para a normal



Validação de Modelos

As simulações permitem verificar se o modelo teórico escolhido é adequado para descrever o fenômeno real observado

Comparação entre Distribuições

Podemos gerar dados de diferentes distribuições e analisar visualmente suas semelhanças e diferenças

A simulação computacional permite explorar o comportamento das distribuições de probabilidade de maneira prática e visual. Utilizando linguagens como R, Python ou softwares como Excel, podemos gerar milhares de amostras aleatórias que seguem determinada distribuição e analisar estatisticamente esses dados simulados.

Este método é particularmente útil para validar aproximações entre distribuições, como verificar quando a distribuição binomial pode ser aproximada adequadamente pela normal, ou quando a hipergeométrica pode ser aproximada pela binomial.

Limitações e Considerações



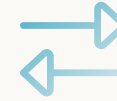
Pressupostos das Distribuições

Cada distribuição tem pressupostos rígidos que devem ser verificados. A binomial exige independência entre ensaios e probabilidade constante; a Poisson requer eventos raros com taxa constante; e a hipergeométrica depende de uma população finita bem definida.



Sensibilidade a Violações

Quando os pressupostos não são completamente satisfeitos, as distribuições podem produzir resultados imprecisos. Por exemplo, a dependência entre ensaios viola o pressuposto da binomial, e flutuações na taxa de ocorrência afetam os resultados da Poisson.



Alternativas Não-paramétricas

Quando os pressupostos não podem ser satisfeitos, métodos não-paramétricos ou de bootstrap podem ser alternativas mais robustas, pois fazem menos suposições sobre a distribuição subjacente dos dados.



Técnicas de Validação

Testes de aderência como qui-quadrado ou Kolmogorov-Smirnov podem verificar se dados observados realmente seguem uma distribuição teórica específica, auxiliando na escolha do modelo mais adequado.

Resumo: Distribuição Binomial



Experimentos com n Ensaios Independentes

A distribuição binomial modela situações com um número fixo de tentativas independentes, onde cada tentativa resulta em apenas dois resultados possíveis. Os ensaios devem ser realizados sob condições idênticas, garantindo que o resultado de um não influencie os demais.



Probabilidade Constante p de Sucesso

Em cada ensaio, a probabilidade p de sucesso deve permanecer constante. Isso significa que as condições do experimento não mudam ao longo das repetições, mantendo a mesma chance de sucesso em todos os ensaios.



Função de Probabilidade e Parâmetros

A probabilidade de obter exatamente k sucessos em n ensaios é dada por $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$. Os parâmetros n e p definem completamente a distribuição, determinando sua forma, média (np) e variância (np(1-p)).



Aplicações Principais

Amplamente utilizada em controle de qualidade, pesquisas de opinião, testes de múltipla escolha, ensaios clínicos e qualquer situação onde contamos sucessos em um número fixo de tentativas independentes com resultados binários.

Resumo: Distribuição de Poisson



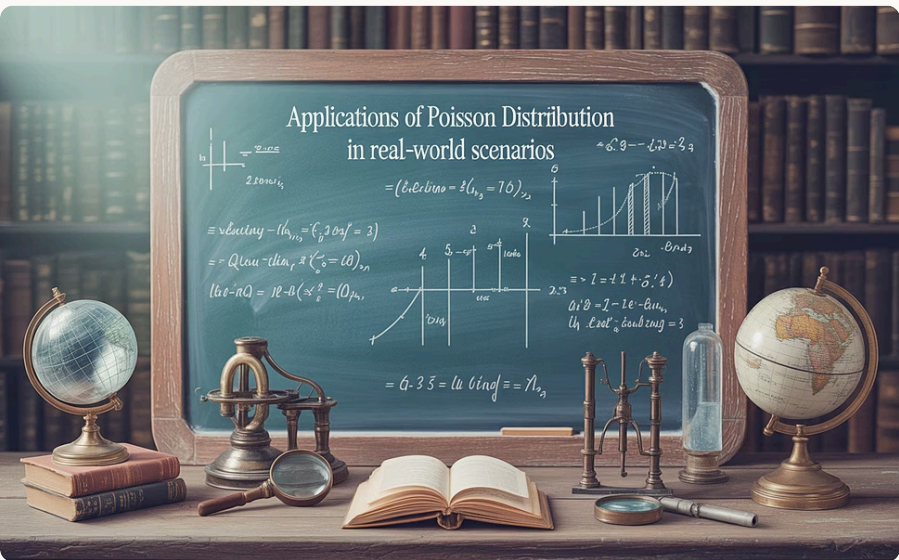
Eventos Raros em Intervalos Contínuos

A distribuição de Poisson modela a ocorrência de eventos discretos em intervalos contínuos de tempo, espaço, área ou volume. É ideal para eventos que ocorrem raramente em relação ao tamanho do intervalo, onde a chance de ocorrências simultâneas é desprezível.



Taxa Média λ de Ocorrência

O único parâmetro da distribuição, λ (lambda), representa o número médio de ocorrências no intervalo considerado. Este parâmetro determina tanto a média quanto a variância da distribuição, uma propriedade única da Poisson.



Aplicações Principais

Amplamente utilizada para modelar filas (chegadas de clientes), defeitos em materiais, chamadas telefônicas, acidentes, casos de doenças raras e qualquer fenômeno onde contamos ocorrências aleatórias com taxa média constante em um intervalo definido.

Resumo: Distribuição Hipergeométrica

Amostragem sem Reposição

A distribuição hipergeométrica aplica-se a situações onde selecionamos uma amostra de itens sem reposição de uma população finita. Isso significa que cada seleção altera a composição da população para as seleções subsequentes, criando dependência entre os ensaios.

Parâmetros N, K, n

A distribuição é definida por três parâmetros: N (tamanho total da população), K (número de elementos na população com a característica de interesse) e n (tamanho da amostra retirada). A proporção K/N equivale à probabilidade p da distribuição binomial.

Função de Probabilidade

A probabilidade de obter exatamente k sucessos na amostra é dada por $P(X=k) = \frac{[(K \ k)(N-K \ n-k)]}{(N \ n)}$. Esta fórmula expressa a razão entre o número de amostras com k sucessos e o número total de possíveis amostras.

Aplicações Principais

Especialmente útil em controle de qualidade com lotes finitos, auditoria financeira, estudos ecológicos (captura-recaptura), processos seletivos e qualquer situação onde selecionamos uma amostra sem reposição de uma população com tamanho conhecido.



Referências Bibliográficas

- MORETTIN, P. A. e BUSSAB, W. O. *Estatística Básica*. 9ª ed. São Paulo: Saraiva, 2017.
- MAGALHÃES, M. N. e LIMA, A. C. P. *Noções de Probabilidade e Estatística*. 7ª ed. São Paulo: EDUSP, 2015.
- FARIAS, A. M. L. et al. *Distribuições Estatísticas*. Belo Horizonte: UFMG, 2020.
- MEYER, P. L. *Probabilidade: Aplicações à Estatística*. 2ª ed. Rio de Janeiro: LTC, 2018.
- COSTA NETO, P. L. O. *Estatística*. 3ª ed. São Paulo: Blucher, 2021.
- FARIAS, R. F. e SOARES, J. B. *Introdução à Estatística para Ciências Exatas*. Fortaleza: UFC, 2019.
- DANTAS, C. A. B. *Probabilidade: Um Curso Introdutório*. 3ª ed. São Paulo: EDUSP, 2018.
- RODRIGUES, M. I. e LEMMA, A. F. *Planejamento de Experimentos e Otimização de Processos*. 3ª ed. Campinas: UNICAMP, 2021.

Além destas referências principais, recomendamos consultar os repositórios online das Universidades Federais brasileiras, que contêm uma vasta coleção de materiais didáticos, teses e dissertações sobre distribuições de probabilidade e suas aplicações em diversos campos do conhecimento.

Referências complementares

Além das referências listadas acima, recomenda-se consultar os repositórios digitais das principais universidades brasileiras que mantêm acervos atualizados de teses, dissertações e artigos sobre IA:

- Biblioteca Digital de Teses e Dissertações da USP (www.teses.usp.br)
- Repositório Institucional da UFMG (repositorio.ufmg.br)
- Repositório Digital da Unicamp (repositorio.unicamp.br)
- Biblioteca Digital da UFPE (repositorio.ufpe.br)
- Lume - Repositório Digital da UFRGS (lume.ufrgs.br)
- Repositório Institucional da FGV (bibliotecadigital.fgv.br)
- Biblioteca Digital de Monografias da UFABC (biblioteca.ufabc.edu.br)

Para acompanhar os avanços mais recentes na pesquisa brasileira, recomenda-se também os anais das conferências BRACIS, SBIA, ENIAC e workshops especializados organizados pela Sociedade Brasileira de Computação (SBC).