II - PROBABILIDADE E ESTATÍSTICA

1. Fundamentos de Probabilidade

Definições Básicas de Probabilidade

A probabilidade é um campo da matemática que estuda a chance de eventos ocorrerem. A probabilidade de um evento (A), denotada como (P(A)), mede o quão provável é que (A) ocorra. Ela varia entre 0 (evento impossível) e 1 (evento certo).

Fórmula Básica

A fórmula básica para calcular a probabilidade de um evento (A) é:

$$P(A) = \frac{\text{número de resultados favoráveis}}{\text{número total de resultados}}$$

Exemplo

Se um dado é lançado, a probabilidade de obter um número par (2, 4, ou 6) é:

$$P(\text{número par}) = \frac{3}{6} = 0,5$$

Axiomas da Probabilidade

Os axiomas de probabilidade, formulados por Kolmogorov, são as regras fundamentais que qualquer medida de probabilidade deve seguir:

1. **Não Negatividade**: Para qualquer evento (A), (P(A) \geq 0).
2. **Normalização**: A probabilidade do espaço amostral inteiro é 1, ou seja, (P(S) = 1).
3. **Adição**: Para dois eventos mutuamente exclusivos (A) e (B), (P(A \cup B) = P(A) + P(B)).

Probabilidade Condicional

A probabilidade condicional calcula a probabilidade de um evento (A) ocorrer, dado que um evento (B) já ocorreu. É denotada por (P(A|B)).

Fórmula da Probabilidade Condicional

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{se } P(B) > 0$$

Exemplo

Se uma urna contém 3 bolas vermelhas e 2 azuis, e uma bola é retirada e não devolvida, a probabilidade de a segunda bola ser azul dado que a primeira era vermelha é:

- (P(\text{primeira vermelha}) = \frac{3}{5}).
- (P(\text{segunda azul} | \text{primeira vermelha}) = \frac{2}{4} = 0,5).

Probabilidade da União de Eventos

Para eventos quaisquer (A) e (B), a probabilidade de ocorrer (A) ou (B) é dada por:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Exemplo

Se a probabilidade de chover é 0,3 e a probabilidade de haver um trânsito pesado é 0,4, e a probabilidade de ambos ocorrerem é 0,1:

$$P(\text{chuva ou trânsito}) = 0,3 + 0,4 - 0,1 = 0,6$$

Independência de Eventos

Dois eventos (A) e (B) são independentes se a ocorrência de um não afeta a ocorrência do outro, ou seja, (P(A \cap B) = P(A) \times P(B)).

Exemplo

Se (P(A) = 0,5) e (P(B) = 0,2), e (A) e (B) são independentes:

$$P(A \cap B) = 0,5 \times 0,2 = 0,1$$

2. Variáveis Aleatórias e Distribuições de Probabilidades

Variáveis Aleatórias

Uma variável aleatória é uma função que associa um resultado numérico a cada possível resultado de um experimento aleatório. Existem dois tipos principais:

- **Variável Aleatória Discreta**: Assume valores contáveis, como o resultado de um lançamento de dado.
- **Variável Aleatória Contínua**: Assume valores em um intervalo contínuo, como a altura de pessoas.

Função de Probabilidade

Para variáveis aleatórias discretas, usamos a função de probabilidade (P(X = x)) para calcular a probabilidade de (X) assumir um valor específico (x).

Exemplo

Se (X) representa o lançamento de um dado:

$$P(X = 3) = \frac{1}{6}$$

Distribuições de Probabilidades

1. Distribuição Uniforme

- **Discreta**: Todos os resultados possíveis têm a mesma probabilidade.

  $$ P(X = x) = \frac{1}{n} $$

- **Contínua**: Distribui igualmente sobre um intervalo ([a, b]).

  $$ f(x) = \frac{1}{b - a}, \quad a \leq x \leq b $$

2. Distribuição Binomial

Modela o número de sucessos em (n) tentativas independentes com probabilidade (p) de sucesso.

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

Exemplo

Para 5 lançamentos de uma moeda justa, a probabilidade de obter 3 caras:

$$P(X = 3) = \binom{5}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = 10 \times \frac{1}{32} = 0,3125$$

3. Distribuição Normal

Distribuição contínua que é simétrica e em forma de sino, caracterizada pela média (\mu) e desvio padrão (\sigma).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Exemplo

Para uma variável normal com (\mu = 0) e (\sigma = 1) (distribuição normal padrão):

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

3. Estatísticas Descritivas

As estatísticas descritivas são utilizadas para resumir e descrever os principais aspectos de um conjunto de dados. Elas incluem medidas de tendência central, dispersão e posição.

Medidas de Tendência Central

As medidas de tendência central descrevem o ponto em torno do qual os dados se concentram.

1. **Média (Média Aritmética)**: A soma de todos os valores dividida pelo número de valores.

   $$ \text{Média} = \frac{\sum_{i=1}^{n} x_i}{n} $$

   Exemplo: Para os valores 2, 4, 6:

   $$ \text{Média} = \frac{2 + 4 + 6}{3} = 4 $$

2. **Mediana**: O valor central de um conjunto de dados ordenado. Se o número de elementos for par, é a média dos dois valores centrais.

Exemplo: Para 1, 3, 3, 6, 7, 8, 9, a mediana é 6.

3. **Moda**: O valor que ocorre com maior frequência no conjunto de dados.

   Exemplo: Para 1, 2, 2, 3, 4, a moda é 2.

Medidas de Dispersão

As medidas de dispersão indicam o grau de variação ou dispersão dos dados em relação à média.

1. **Variância**: Mede a média dos quadrados das diferenças dos valores em relação à média.

   $$ \text{Variância} = \frac{\sum_{i=1}^{n} (x_i - \text{Média})^2}{n} $$

   Exemplo: Para 2, 4, 4, 4, 5, 5, 7, 9:

   $$ \text{Média} = 5, \quad \text{Variância} = \frac{(2-5)^2 + (4-5)^2 + ... + (9-5)^2}{8} = 4 $$

2. **Desvio Padrão**: É a raiz quadrada da variância, medindo a dispersão em unidades da mesma medida dos dados.

   $$ \text{Desvio Padrão} = \sqrt{\text{Variância}} $$

   Exemplo: Com a variância (4), o desvio padrão é (\sqrt{4} = 2).

3. **Amplitude**: Diferença entre o maior e o menor valor do conjunto.

   Exemplo: Para 1, 3, 6, 7, 10, a amplitude é (10 - 1 = 9).

Medidas de Posição

As medidas de posição indicam a localização de um valor específico dentro de um conjunto de dados.

1. **Percentis**: Dividem os dados em 100 partes iguais. O percentil (p) é o valor abaixo do qual (p %) dos dados estão localizados.

   Exemplo: O 50º percentil é a mediana.

2. **Quartis**: Dividem os dados em quatro partes iguais.

   - **Q1 (1º Quartil)**: 25% dos dados estão abaixo desse valor.
   - **Q2 (2º Quartil ou Mediana)**: 50% dos dados estão abaixo desse valor.
   - **Q3 (3º Quartil)**: 75% dos dados estão abaixo desse valor.

   Exemplo: Para 1, 2, 3, 4, 5, 6, 7, 8, 9:

- (Q1 = 2.5)
- (Q2 = 5)
- (Q3 = 7.5)

4. Distribuição de Poisson

Modela o número de eventos ocorrendo em um intervalo fixo de tempo ou espaço.

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

Exemplo

Se (\lambda = 4), a probabilidade de ocorrerem 2 eventos:

$$P(X = 2) = \frac{4^2 e^{-4}}{2!} = \frac{16 \times 0,0183}{2} = 0,146$$

5. Distribuição Bernoulli

Distribuição discreta para experimentos com dois resultados (sucesso e fracasso).

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

6. Distribuição Exponencial

Modela o tempo entre eventos em um processo de Poisson.

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

Exemplo

Para (\lambda = 2), a densidade de probabilidade para (x = 1):

$$f(1) = 2e^{-2 \times 1} = 2 \times 0,1353 = 0,2706$$

4. Teoremas Fundamentais da Probabilidade

Os teoremas fundamentais da probabilidade ajudam a entender como eventos se relacionam e como calcular probabilidades complexas.

Independência de Eventos

Dois eventos (A) e (B) são independentes se a ocorrência de um não afeta a probabilidade de ocorrência do outro.

Fórmula

$$P(A \cap B) = P(A) \times P(B)$$

Exemplo

Se (P(A) = 0,4) e (P(B) = 0,5), e (A) e (B) são independentes:

$$P(A \cap B) = 0,4 \times 0,5 = 0,2$$

Teorema de Bayes

O teorema de Bayes relaciona probabilidades condicionais e marginais, permitindo atualizar as probabilidades com base em novas informações.

Fórmula

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

Exemplo

Se a probabilidade de um teste ser positivo dado que a pessoa está doente é 0,9, e a probabilidade de estar doente é 0,01, e a probabilidade de teste positivo é 0,05:

$$P(\text{Doente} \mid \text{Teste Positivo}) = \frac{0,9 \times 0,01}{0,05} = 0,18$$

Teorema da Probabilidade Total

O teorema da probabilidade total é usado para calcular a probabilidade de um evento baseado em diferentes cenários mutuamente exclusivos.

Fórmula

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i) \times P(A_i)$$

Exemplo

Se um evento pode ocorrer em três cenários:

- (P(B|A_1) = 0,2, P(A_1) = 0,3)
- (P(B|A_2) = 0,4, P(A_2) = 0,5)
- (P(B|A_3) = 0,6, P(A_3) = 0,2)

Então:

$$P(B) = (0,2 \times 0,3) + (0,4 \times 0,5) + (0,6 \times 0,2) = 0,38$$

Lei dos Grandes Números

A Lei dos Grandes Números afirma que, à medida que o número de experimentos aumenta, a média amostral se aproxima da média esperada da população.

Exemplo

Lançando uma moeda justa repetidamente, a proporção de caras se aproxima de 0,5 à medida que o número de lançamentos aumenta.

Teorema Central do Limite

O Teorema Central do Limite afirma que a soma (ou média) de um grande número de variáveis aleatórias independentes e identicamente distribuídas tende a seguir uma distribuição normal, independentemente da distribuição original das variáveis.

Fórmula

Para uma variável aleatória com média (\mu) e desvio padrão (\sigma), a média amostral (\bar{X}) de (n) amostras segue aproximadamente uma distribuição normal:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Exemplo

Se a altura média de uma população é 170 cm com desvio padrão de 10 cm, a média amostral de 100 indivíduos segue:

$$\bar{X} \sim N\left(170, \frac{10}{\sqrt{100}}\right) = N(170, 1)$$

5. Distribuições Amostrais

As distribuições amostrais descrevem o comportamento das estatísticas calculadas a partir de amostras, em vez de populações inteiras. Elas são fundamentais para a inferência estatística.

Distribuição Amostral da Média

A distribuição amostral da média mostra como a média das amostras varia. Se as amostras são grandes o suficiente, a distribuição tende a ser normal, mesmo que a população original não seja.

Propriedades

- Média da distribuição amostral ((\mu_{\bar{X}})) é igual à média da população ((\mu)).
- Desvio padrão da distribuição amostral ((\sigma_{\bar{X}})) é:

  $$ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} $$

Exemplo

Para uma população com média 50 e desvio padrão 10, se extraímos amostras de tamanho 25, o desvio padrão da média amostral é:

$$\sigma_{\bar{X}} = \frac{10}{\sqrt{25}} = 2$$

Distribuição Amostral da Proporção

Usada para descrever a distribuição da proporção de sucessos em amostras.

Propriedades

- Média da distribuição amostral ((\mu_{\hat{p}})) é igual à proporção da população ((p)).
- Desvio padrão da distribuição amostral ((\sigma_{\hat{p}})) é:

  $$ \sigma_{\hat{p}} = \sqrt{\frac{p(1 - p)}{n}} $$

Exemplo

Se 60% de uma população possui uma característica específica, e a amostra tem tamanho 100, o desvio padrão da proporção amostral é:

$$\sigma_{\hat{p}} = \sqrt{\frac{0,6 \times 0,4}{100}} = 0,049$$

Distribuição Qui-Quadrado ((\chi^2))

Usada principalmente para testar variâncias e independência de variáveis categóricas.

Propriedades

- Não é simétrica, especialmente para pequenos graus de liberdade.
- Somatória dos quadrados de (n) variáveis normais padronizadas.

Exemplo

Se temos 5 graus de liberdade, a forma da distribuição será assimétrica à direita.

Distribuição t de Student

Usada quando o tamanho da amostra é pequeno e a variância populacional é desconhecida. Aproxima-se da normal conforme o tamanho da amostra aumenta.

Fórmula

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

Exemplo

Para uma amostra de 10 observações com média 20, desvio padrão 5, e média populacional 18:

$$t = \frac{20 - 18}{\frac{5}{\sqrt{10}}} = 1,264$$

Distribuição F

Usada para comparar variâncias de duas populações e em ANOVA (análise de variância).

Propriedades

- Não é simétrica, especialmente com graus de liberdade pequenos.
- Razão entre duas variâncias independentes de distribuições qui-quadrado.

Exemplo

Comparando duas variâncias, se (F = 2,5) com (df_1 = 10) e (df_2 = 15), indica que a variância do numerador é 2,5 vezes a do denominador.

6. Inferência Estatística

A inferência estatística envolve fazer estimativas ou tomar decisões sobre uma população com base em uma amostra. As principais técnicas incluem a estimação e os testes de hipóteses.

Estimação Pontual e Intervalar

Estimação Pontual

A estimação pontual é um único valor que serve como melhor estimativa de um parâmetro populacional. Por exemplo, a média amostral ((\bar{X})) é uma estimativa pontual da média populacional ((\mu)).

Intervalos de Confiança

Um intervalo de confiança fornece uma faixa de valores dentro da qual o parâmetro populacional é provável de estar, com um certo nível de confiança (ex. 95%).

Fórmula do Intervalo de Confiança para a Média

Para uma amostra de tamanho (n), média (\bar{X}), desvio padrão (s), e nível de confiança ((1 - \alpha)), o intervalo é dado por:

$$IC = \bar{X} \pm t_{\alpha/2,,n-1} \cdot \frac{s}{\sqrt{n}}$$

Exemplo

Para uma média amostral de 50, (s = 10), (n = 25), e nível de confiança de 95%, o intervalo de confiança é:

$$IC = 50 \pm 2,064 \times \frac{10}{\sqrt{25}} = 50 \pm 4,128$$

Testes de Hipóteses

Os testes de hipóteses são usados para tomar decisões sobre parâmetros populacionais. Envolvem formular uma hipótese nula ((H_0)) e uma hipótese alternativa ((H_1)).

Passos para Testar Hipóteses

1. **Formular (H_0) e (H_1)**.
2. **Escolher o nível de significância** ((\alpha), normalmente 0,05).
3. **Calcular a estatística de teste**.
4. **Determinar a região crítica** ou p-valor.
5. **Tomar uma decisão**: rejeitar ou não rejeitar (H_0).

Tipos de Erros

- **Erro Tipo I ((\alpha))**: Rejeitar (H_0) quando é verdadeira.
- **Erro Tipo II ((\beta))**: Não rejeitar (H_0) quando é falsa.

Poder do Teste

O poder do teste ((1 - \beta)) é a probabilidade de rejeitar (H_0) quando (H_1) é verdadeira.

Testes Específicos

Testes z e t para Médias

- **Teste z**: usado quando o tamanho da amostra é grande ((n > 30)) ou a variância populacional é conhecida.

  $$ z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} $$

- **Teste t**: usado quando o tamanho da amostra é pequeno e a variância populacional é desconhecida.

  $$ t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} $$

Testes de Proporções

Testa se a proporção populacional é igual a um valor especificado.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

Testes Qui-Quadrado

Usado para testar a independência de variáveis categóricas ou ajuste de Goodness-of-Fit.

- **Independência**: Testa se duas variáveis categóricas são independentes.
- **Goodness-of-Fit**: Testa se uma distribuição observada difere de uma distribuição esperada.

Teste A/B

Usado para comparar dois grupos e determinar qual é melhor em termos de desempenho ou eficácia.

Exemplo de Teste t para Médias

Se uma amostra de 20 estudantes tem média 80, desvio padrão 10, e estamos testando se a média é 75 com (\alpha = 0,05):

$$t = \frac{80 - 75}{\frac{10}{\sqrt{20}}} = 2,236$$

Com 19 graus de liberdade, compare (t) com o valor crítico para decidir se rejeita (H_0).

7. Correlação

A correlação mede a força e a direção da relação linear entre duas variáveis. É importante notar que correlação não implica causalidade.

Correlação e Causalidade

- **Correlação**: Refere-se a um relacionamento estatístico entre duas variáveis.
- **Causalidade**: Refere-se a uma relação onde uma variável causa mudanças na outra. Mesmo que duas variáveis sejam correlacionadas, isso não significa que uma causa a outra.

Coeficiente de Correlação de Pearson

O coeficiente de correlação de Pearson ((r)) mede a força e a direção da relação linear entre duas variáveis numéricas.

Fórmula

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Interpretação

- (r = 1): Correlação perfeita positiva.
- (r = -1): Correlação perfeita negativa.
- (r = 0): Nenhuma correlação linear.

Exemplo

Para os dados ((x, y)): (1, 2), (2, 3), (3, 6), (4, 8):

1. Calcule (\bar{x}) e (\bar{y}).
2. Aplique os valores na fórmula para calcular (r).

Correlação de Spearman

A correlação de Spearman é usada para variáveis ordinais ou quando a relação entre as variáveis não é linear.

Fórmula

Usa o coeficiente de correlação de Pearson sobre os postos das variáveis.

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Onde (d_i) é a diferença entre os postos de (x) e (y).

Exemplo

Para (n = 5) pares de dados, com diferenças de postos (d_i = 1, 2, 0, -1, 3):

$$r_s = 1 - \frac{6 \times (1^2 + 2^2 + 0^2 + (-1)^2 + 3^2)}{5(5^2 - 1)} = 0,7$$

Correlação Parcial

A correlação parcial mede a relação entre duas variáveis enquanto controla o efeito de uma ou mais outras variáveis.

Exemplo

Suponha que você deseja medir a correlação entre (X) (idade) e (Y) (peso), controlando o efeito de (Z) (altura). A correlação parcial isolaria o efeito da altura na relação entre idade e peso.

Interpretação dos Resultados

- **Correlação forte**: Próximo de 1 ou -1.
- **Correlação moderada**: Aproximadamente entre 0,5 e 0,7.
- **Correlação fraca**: Próximo de 0.

É importante analisar graficamente os dados para evitar interpretações erradas, como correlações espúrias ou influências de outliers.

8. Inferência Bayesiana

A inferência Bayesiana é uma abordagem estatística que utiliza o Teorema de Bayes para atualizar a probabilidade de uma hipótese à medida que mais evidências ou informações se tornam disponíveis.

Distribuições a Priori e a Posteriori

- **Distribuição a Priori ((P(\theta)))**: Representa o conhecimento ou crença sobre o parâmetro (\theta) antes de observar os dados.
- **Distribuição a Posteriori ((P(\theta | \text{dados})))**: Atualiza a distribuição a priori com a evidência dos dados usando o Teorema de Bayes.

Fórmula do Teorema de Bayes

$$P(\theta \mid \text{dados}) = \frac{P(\text{dados} \mid \theta) \cdot P(\theta)}{P(\text{dados})}$$

Onde:

- (P(\theta | \text{dados})): Probabilidade a posteriori.
- (P(\text{dados} | \theta)): Verossimilhança dos dados dados (\theta).
- (P(\theta)): Probabilidade a priori.
- (P(\text{dados})): Evidência ou probabilidade total dos dados.

Estimativa Pontual e Intervalar

- **Estimativa Pontual**: Estimativa do parâmetro mais provável com base na distribuição a posteriori, como a moda, média ou mediana da distribuição.
- **Intervalo de Credibilidade**: Intervalo dentro do qual o parâmetro está com uma determinada probabilidade (ex., 95%).

Predição e Testes de Hipóteses Bayesianos

Na inferência bayesiana, a predição de novos dados é feita utilizando a distribuição preditiva, que combina a incerteza sobre o parâmetro (\theta).

Fórmula da Distribuição Preditiva

$$P(\tilde{x} \mid \text{dados}) = \int P(\tilde{x} \mid \theta) P(\theta \mid \text{dados}) d\theta$$

- (\tilde{x}): Novo dado a ser predito.
- A integral reflete a incerteza sobre o parâmetro (\theta).

Testes de Hipóteses Bayesianos

Os testes de hipóteses bayesianos avaliam a probabilidade de uma hipótese ser verdadeira, comparando modelos usando razões de verossimilhança.

Critérios de Seleção de Modelos

Os modelos bayesianos são avaliados com base na adequação dos dados observados. Alguns critérios comuns incluem:

- **DIC (Deviance Information Criterion)**: Compara modelos considerando complexidade e ajuste.
- **WAIC (Watanabe-Akaike Information Criterion)**: Semelhante ao AIC, mas com uma perspectiva bayesiana.

Métodos MCMC (Markov Chain Monte Carlo)

Os métodos MCMC são algoritmos usados para amostrar da distribuição a posteriori quando o cálculo analítico é complexo.

- **Gibbs Sampling**: Itera sobre variáveis condicionais para gerar amostras da distribuição.
- **Metropolis-Hastings**: Proposta de amostras que são aceitas ou rejeitadas com base na probabilidade de aceitação.

Exemplo de Inferência Bayesiana com MCMC

Suponha que estamos estimando a média de uma distribuição normal com variância conhecida. Com MCMC, amostramos da distribuição a posteriori da média com base nos dados observados.

Vantagens da Inferência Bayesiana

- **Incorporação de Informações Prévias**: Usa conhecimento prévio que pode melhorar a precisão.
- **Flexibilidade**: Aplicável a modelos complexos onde a inferência frequencista é difícil.
- **Interpretação Probabilística**: Oferece probabilidades diretas para hipóteses e parâmetros.

Exemplo

Suponha que queremos estimar a probabilidade de um teste ser positivo ((\theta)). Com uma distribuição a priori (\theta \sim Beta(2, 2)) e 8 testes positivos em 10:

1. A verossimilhança dos dados é (Binomial(10, \theta)).
2. A posteriori será (Beta(2 + 8, 2 + 2) = Beta(10, 4)).
3. Com base na posteriori, estimamos que a probabilidade de ser positivo está entre 0,5 e 0,8 com 95% de confiança.

Questões de Probabilidade e Estatística

Questão 1

Qual dos seguintes valores representa a probabilidade correta de um evento impossível?

- a) 0.5
- b) 0.1
- c) 0
- d) 1
- e) 0.9

▶ **Resposta**

Questão 2

Seja (X) uma variável aleatória que segue uma distribuição normal com média 50 e desvio padrão 10. Qual é a probabilidade de (X) assumir um valor entre 40 e 60?

- a) 0.50
- b) 0.68
- c) 0.95
- d) 0.34
- e) 0.16

▶ **Resposta**

Questão 3

Em um experimento com duas variáveis independentes (A) e (B), onde (P(A) = 0.4) e (P(B) = 0.5), qual é (P(A \cap B))?

- a) 0.2
- b) 0.4
- c) 0.5
- d) 0.1
- e) 0.6

▶ **Resposta**

Questão 4

Se a probabilidade condicional (P(A|B) = 0.3), (P(B) = 0.5), qual é o valor de (P(A \cap B))?

- a) 0.15
- b) 0.3
- c) 0.2
- d) 0.5
- e) 0.6

▶ **Resposta**

Questão 5

Qual é a variância de uma variável aleatória (X) com média (\mu = 10) e valores (8, 10, 12)?

- a) 2
- b) 4
- c) 1.33
- d) 3
- e) 0

▶ **Resposta**

Questão 6

Se a distribuição amostral de uma média tem desvio padrão (\sigma_{\bar{X}} = 2) para (n = 25), qual é o desvio padrão da população ((\sigma))?

- a) 10
- b) 5
- c) 8
- d) 4
- e) 6

▶ **Resposta**

Questão 7

Qual das alternativas a seguir é uma distribuição usada para testar a independência de variáveis categóricas?

- a) Normal
- b) Binomial
- c) Qui-Quadrado
- d) Exponencial
- e) Uniforme

▶ **Resposta**

Questão 8

Se o coeficiente de correlação de Pearson entre duas variáveis é 0, qual é a melhor interpretação?

- a) As variáveis têm uma correlação positiva perfeita.
- b) As variáveis têm uma correlação negativa perfeita.
- c) As variáveis são independentes.
- d) Não há correlação linear entre as variáveis.
- e) As variáveis são inversamente proporcionais.

▶ **Resposta**

Questão 9

Dado um intervalo de confiança de 95% para a média com limites 40 e 60, qual é a interpretação correta?

- a) A média da população é 50.
- b) 95% dos valores da amostra estão entre 40 e 60.
- c) Há 95% de chance de a média populacional estar entre 40 e 60.
- d) 95% das médias amostrais estão entre 40 e 60.
- e) A variância da amostra é 10.

▶ **Resposta**

Questão 10

Qual é a distribuição a posteriori para um parâmetro (\theta) dado uma distribuição a priori (\theta \sim Beta(2, 2)) e dados com 8 sucessos em 10 tentativas?

- a) Beta(10, 4)
- b) Beta(8, 2)
- c) Binomial(10, 8)
- d) Normal(10, 4)
- e) Exponencial(8, 2)

▶ **Resposta**

Questões Avançadas de Probabilidade e Estatística

Questão 1

Uma fábrica produz peças com defeito a uma taxa de 2%. Se 20 peças são selecionadas aleatoriamente, qual é a probabilidade de que exatamente 3 peças estejam defeituosas?

- a) 0.037
- b) 0.114
- c) 0.084
- d) 0.028
- e) 0.067

▶ **Resposta**

Questão 2

Seja (X) uma variável aleatória com distribuição normal (N(100, 15^2)). Qual é a probabilidade de (X) estar entre 85 e 115?

- a) 0.3413
- b) 0.6826
- c) 0.4772
- d) 0.9544
- e) 0.1359

▶ **Resposta**

Questão 3

Um teste de hipóteses é realizado com (H_0: \mu = 50) contra (H_1: \mu \neq 50). Se o p-valor obtido é 0.03, e o nível de significância é 0.05, qual é a decisão?

- a) Rejeita (H_0)
- b) Não rejeita (H_0)
- c) Aumenta o nível de significância
- d) Reduz o tamanho da amostra
- e) Calcula o erro tipo II

▶ **Resposta**

Questão 4

Qual é a variância da soma de 100 variáveis aleatórias independentes e identicamente distribuídas (X_i \sim N(0, 4))?

- a) 400
- b) 4
- c) 200
- d) 100
- e) 40

▶ **Resposta**

Questão 5

Em uma pesquisa, 10% dos entrevistados preferem o produto A. Qual é a probabilidade de, em uma amostra de 50 entrevistados, menos de 5 preferirem o produto A?

- a) 0.543
- b) 0.205
- c) 0.029
- d) 0.062
- e) 0.103

▶ **Resposta**

Questão 6

Se (X) e (Y) são variáveis aleatórias independentes, ambas seguindo (N(0, 1)), qual é a distribuição de (Z = X^2 + Y^2)?

- a) Qui-Quadrado com 1 grau de liberdade
- b) Qui-Quadrado com 2 graus de liberdade
- c) Normal com média 0 e variância 2
- d) Exponencial com parâmetro 2
- e) Gama com forma 2 e escala 0.5

▶ **Resposta**

Questão 7

Se (\theta) segue uma distribuição a priori (Beta(1, 1)) e observamos 7 sucessos em 10 tentativas, qual é a distribuição a posteriori?

- a) Beta(7, 3)
- b) Beta(8, 4)
- c) Beta(8, 3)
- d) Beta(1, 1)
- e) Beta(7, 4)

▶ **Resposta**

Questão 8

Em um teste de (t) para comparar duas médias amostrais com variâncias desconhecidas e amostras pequenas, qual é a suposição fundamental?

- a) Amostras são grandes
- b) Variâncias das populações são iguais
- c) Variâncias das populações são diferentes
- d) As distribuições são não normais
- e) A variância amostral é zero

▶ **Resposta**

Questão 9

Qual é a interpretação correta de um intervalo de credibilidade de 95% em inferência bayesiana?

- a) A média da população está no intervalo 95% das vezes
- b) O parâmetro está dentro do intervalo com probabilidade de 95%
- c) 95% dos dados estão dentro do intervalo
- d) 95% das amostras terão a média no intervalo
- e) O intervalo contém 95% das possíveis médias amostrais

▶ **Resposta**

Questão 10

Uma variável aleatória (X \sim Exponencial(\lambda)). Qual é a probabilidade de (X > 2) dado que (\lambda = 0.5)?

- a) 0.1353
- b) 0.2707
- c) 0.6065
- d) 0.8187
- e) 0.3679

▶ **Resposta**

Questões Discursivas de Probabilidade e Estatística

Questão 1

Uma empresa está analisando a taxa de falhas de um de seus equipamentos, que segue uma distribuição de Poisson com uma média de 2 falhas por mês. Determine a probabilidade de ocorrerem exatamente 3 falhas em um mês. Além disso, explique como essa distribuição pode ser utilizada para prever problemas futuros na operação da empresa.

Resposta Esperada

A probabilidade de ocorrerem exatamente 3 falhas é dada pela fórmula da distribuição de Poisson:

[P(X = 3) = \frac{\lambda^3 e^{-\lambda}}{3!} = \frac{2^3 e^{-2}}{6} = \frac{8 \times 0.1353}{6} \approx 0.1804]

Explique que a distribuição de Poisson é útil para modelar eventos raros, ajudando na previsão de manutenção preventiva e alocação de recursos para minimizar os impactos operacionais.

Questão 2

Discuta o impacto da Lei dos Grandes Números na validação de amostras em pesquisas científicas. Dê um exemplo de como este teorema pode ser aplicado na prática para justificar a escolha de tamanhos amostrais.

Resposta Esperada

A Lei dos Grandes Números afirma que à medida que o tamanho da amostra aumenta, a média amostral se aproxima da média populacional. Isso valida o uso de amostras grandes em pesquisas, pois garante que os resultados amostrais sejam representativos da população.

Exemplo prático: Em um estudo de saúde pública, um pesquisador pode justificar a necessidade de uma amostra de 1000 pessoas para garantir que a média amostral de pressão arterial se aproxime da média da população geral, minimizando a variação devido ao acaso.

Questão 3

Explique a diferença entre um intervalo de confiança e um intervalo de credibilidade, destacando suas aplicações em contextos de inferência frequentista e bayesiana. Forneça um exemplo de cada tipo de intervalo.

Resposta Esperada

- **Intervalo de Confiança**: Utilizado na inferência frequentista, representa uma faixa de valores que, em 95% das amostras, conteria o parâmetro verdadeiro. Por exemplo, um intervalo de confiança de 95% para a média salarial de uma cidade pode indicar que estamos 95% confiantes de que a média está entre R$ 2000 e R$ 3000.

- **Intervalo de Credibilidade**: Utilizado na inferência bayesiana, é a probabilidade do parâmetro estar dentro do intervalo dado os dados observados. Por exemplo, se um teste de confiabilidade indica que há 95% de chance de a taxa de falhas de um sistema estar entre 1% e 3%.

Questão 4

Uma pesquisa compara a eficácia de dois medicamentos utilizando um teste t para amostras independentes, com ($n_1 = 30$) e ($n_2 = 28$). Discuta os pressupostos necessários para a validade do teste e como você avaliaria se esses pressupostos são atendidos antes de prosseguir com a análise.

Resposta Esperada

Os pressupostos do teste t para amostras independentes incluem:

1. **Normalidade**: As distribuições de ambas as amostras devem ser aproximadamente normais. Isso pode ser avaliado utilizando testes de normalidade (ex., Shapiro-Wilk) ou gráficos (ex., histogramas e Q-Q plots).

2. **Homogeneidade de Variâncias**: As variâncias das duas populações devem ser iguais, testáveis por Levene ou Bartlett.

3. **Independência das Observações**: As amostras devem ser independentes uma da outra, o que deve ser garantido pelo desenho do estudo.

Se esses pressupostos não forem atendidos, alternativas como o teste t de Welch ou métodos não paramétricos, como o teste de Mann-Whitney, devem ser considerados.

Questão 5

Considere que você está conduzindo um estudo de análise de sobrevivência em pacientes com uma doença crônica. Descreva como você usaria a distribuição exponencial para modelar o tempo até a falha (morte) dos pacientes e como a escolha do parâmetro (λ) afeta as previsões. Ilustre com um cálculo numérico.

Resposta Esperada

A distribuição exponencial é usada para modelar o tempo até a ocorrência de um evento (ex., morte) e assume que o evento ocorre com uma taxa constante ao longo do tempo.

A função de densidade é dada por:

[$f(t) = \lambda e^{-\lambda t}$]

Onde (λ) é a taxa de falha. Um valor maior de (λ) indica uma maior taxa de eventos e, portanto, menor tempo esperado até a falha.

Exemplo: Se $\lambda = 0.1$, o tempo médio de sobrevivência é $\frac{1}{\lambda} = 10$ anos. A probabilidade de um paciente sobreviver mais de 5 anos é:

$$P(T > 5) = e^{-0.1 \times 5} = 0.6065$$

Explique que ajustar λ com base nos dados observados é crucial para previsões precisas sobre a sobrevivência dos pacientes.