

lista 01 (Conceitos Básicos / Introdução à Estatística)

Pedro Manoel Herminio Alves - 119210706

16/10/2022

Questão 01

A distribuição de frequência é um arranjo de valores que uma ou mais variáveis tomam em uma amostra. Cada entrada na tabela contém a frequência ou a contagem de ocorrências de valores dentro de um grupo ou intervalo específico, e deste modo, a tabela resume a distribuição dos valores da amostra.

Exemplo

Conjunto de idades (em anos):

12 14 18
14 12 18
18 20 18
17 19 20
20 14 19

Tabela de distribuição para a variável IDADE (em anos)

IDADE (em anos)	Freq. Absoluta	Freq. Relativa	Porcentagem
12	2	0,1333	13,33%
14	3	0,2000	20,00%
17	1	0,0667	6,67%
18	4	0,2667	26,67%
19	2	0,1333	13,33%
20	3	0,2000	20,00%
Total	15	1	100%

Questão 02

Uma **variável aleatória** é nada mais que uma função que mapeia os resultados de um processo estocástico (aleatório) em valores numéricos.

Como exemplo podemos imaginar um **experimento aleatório** simples como lançar uma moeda honesta, os possíveis resultados podem ser $\Omega = \{cara, coroa\}$, o conjunto de todos os possíveis resultados de um experimento aleatório é chamado do espaço amostral Ω . Agora imagine uma função qualquer que transformem os resultados em um valor numérico, como por exemplo, **número de coroas**, note que deste experimento podemos ter dois resultados, cara ou coroa, sendo o número de coroas, **0** ou **1**.

Assim X é a variável aleatória que traduz todos os resultados do experimento aleatório em números x .

A notação usual é, a letra maiúscula X representa a variável aleatória, e x representa o valor numérico.

Questão 03

As variáveis aleatórias podem ser classificadas como **discreta** ou **contínua**.

Variável aleatória discreta - são aquelas que os resultados são frutos de **contagem**

- *Exemplo:* Número de filhos, número de sucessos em n tentativas,...

Variável aleatória contínua - são aquelas que os resultados são frutos de **mensuração**

- *Exemplo:* massa (kg) do objeto, altura (m) do indivíduo, idade, temperatura, diâmetro da peça, ...

Questão 04

Função ou distribuição de probabilidade de uma experiência aleatória é a função que a cada evento possível faz corresponder a probabilidade do evento ocorrer.

Exemplo

Considere o experimento do lançamento de duas moedas honestas. Seja a variável aleatória o número de caras obtidas. Podemos construir a função de probabilidade de X da seguinte forma:

X assume os seguintes valores:

- $X = \{0, 1, 2\}$

Temos que

$$P(X = 0) = P(K, K) = \frac{1}{4}$$

$$P(X = 1) = P(C, K) + P(K, C) = \frac{1}{2}$$

$$P(X = 2) = P(C, C) = \frac{1}{4}$$

Denotamos a função de probabilidade de X por

x_i	0	1	2
$P(X = x_i)$	1/4	1/2	1/4

Questão 05

Uma distribuição de probabilidade é um modelo matemático que relaciona um certo valor da variável em estudo com a sua probabilidade de ocorrência.

Questão 06

Sim, pois a distribuição de probabilidades pode ser representada por um histograma de probabilidades, similar ao histograma de frequências aonde a escala vertical representa probabilidades, em lugar das frequências relativas.

Questão 07

Podemos utilizar de experimentos cujo resultados são tais que ocorre ou não ocorre determinada característica.

Como os seguintes exemplos:

1. **Ao lançar uma moeda honesta:** o resultado ou é cara ou não (ocorrendo, então, coroa);
2. **Ao lançar um dado honesto:** ocorre número par ou não (ocorrendo número ímpar);
3. **Uma pessoa é escolhida ao acaso dentre 1000:** ou ela é favorável a um determinado projeto governamental ou não.

Questão 08

Em estatística, um parâmetro é **um número que resume a grande quantidade de dados que podem derivar do estudo de uma variável estatística**. O cálculo deste número está bem definido, usualmente mediante uma fórmula aritmética obtida a partir de dados da população.

Exemplo

Observando-se a população de uma cidade podemos obter como parâmetros a média das idades dos indivíduos dessa população, ou a proporção de homens e mulheres

Questão 09

Temos os seguintes parâmetros:

- p que representa a probabilidade de sucesso, onde $0 < p < 1$

Questão 10

f.p da distribuição de Bernoulli

- $P(X = K) = p^k(1 - p)^{1-k}, k \in R_x = \{0, 1\}$

Exemplo

Problema

Considere uma urna contendo seis bolas azuis e quatro vermelhas. Qual a probabilidade de que retirando-se uma bola, essa bola seja azul?

Solução

```
#Número de bolas de cada tipo
bolasAzuis = 6
bolasVermelhas = 4

# Probabilidade de sucesso (retirar bola azul)
bolasAzuis / (bolasAzuis + bolasVermelhas)
```

```
[1] 0.6
```

Questão 11

Podemos utilizar de experimentos que necessitamos saber o número de sucessos obtidos em n ensaios independentes e idênticos de um experimento aleatório.

Como os seguintes exemplos:

1. Sabendo-se que em uma urna existe seis bolas azuis e quatro bolas vermelhas, fazendo-se retiradas com reposição qual seria a probabilidade de retirar, no mínimo, duas bolas azuis
2. Sabendo-se que na manufatura de certo artigo, um entre dez dos artigos é defeituoso qual seria a probabilidade de que em uma amostra casual de tamanho quatro não contenha nenhum artigo defeituoso

Questão 12

Temos os seguintes parâmetros:

- p que representa a probabilidade de sucesso, onde $0 < p < 1$
- n que representa o número de ensaios total, possíveis de serem realizados

Questão 13

f.p da distribuição Binomial

$$\bullet P(X = K) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}, \quad k \in R_x = \{0, 1, 2, 3, \dots, n\}$$

Exemplo

Problema

Considere uma urna contendo seis bolas azuis e quatro vermelhas. Retire três bolas, com reposição, e defina a v.a. X igual ao número de bolas azuis. Qual a probabilidade de sair no mínimo 2 bolas azuis?

Solução

```
#Número de bolas de cada tipo
bolasAzuis = 6
bolasVermelhas = 4

# Probabilidade de sucesso (retirar bola azul)
p = bolasAzuis / (bolasAzuis + bolasVermelhas)

# Número de ensaios
n = 3

# Calculando a probabilidade (Modelo Binomial)
# P(X >= 2) = 1 - P(X < 2) = 1 - P(X <= 1)
1 - pbinom(1, n, p)
```

```
[1] 0.648
```

Questão 14

Em uma variável aleatória contínua o conjunto dos possíveis valores pode ser um intervalo ou um conjunto de intervalos.

Seja X uma variável aleatória contínua. A função de densidade de probabilidade $f(x)$ é uma função que satisfaz as seguintes condições:

$$f(x) > 0, \forall x$$

e

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

Exemplo

X representa o consumo de combustível de um certo automóvel em Km/h

$$f(x) = \begin{cases} x - 10, & \text{se } 10 < x < 11, \\ 12 - x, & \text{se } 11 \leq x < 12, \\ 0, & \text{cc.} \end{cases}$$

Questão 15

Tomando como exemplo o seguinte cenário:

O tempo necessário para concluir um exame final em determinado curso universitário está distribuído normalmente com uma média de 80 minutos e variância de 100 minutos²

Temos como exemplo:

1. Qual é a probabilidade um estudante concluir o exame em uma hora ou menos?
2. Qual é a probabilidade de um estudante demorar entre 60 e 75 minutos para concluir o exame?
3. Suponha que a classe tenha 60 alunos e que a duração do exame seja de 90 minutos. Qual o número esperado de estudantes que concluiriam o exame no tempo determinado?

Questão 16

Temos os seguintes parâmetros:

- μ que representa a média
- σ^2 que representa a variância

Questão 17

f.d.p da distribuição Normal

- $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, x \in \mathbb{R}$

Exemplo

Problema

Os depósitos efetuados em um certo banco durante o mês de janeiro são distribuídos normalmente, com média de \$10000,00 e desvio padrão \$1500,00. Um depósito é selecionado ao acaso dentre todos os referentes ao mês em questão. Calcule a probabilidade de que o depósito seja menor do que \$13000,00.

Solução

```
# parâmetros
mean = 10000 # Média
sd = 1500 # Desvio Padrão

# Calculando a probabilidade (Modelo Normal)
# P(X < 13000)
pnorm(13000, mean, sd)
```

```
[1] 0.9772499
```

Questão 18

1 - Definição do problema

Antes de qualquer coisa, deve-se ter bem definido qual é o problema que deve ser resolvido por meio de análises estatísticas.

2 - Planejamento

Nesta etapa, deve-se estipular como será feita a coleta dos dados que irão compor o estudo. O trabalho a ser realizado depende do objetivo que se pretende alcançar. Deve-se ter um cuidado especial com a escolha das perguntas e a formulação delas, para que as respostas

estejam dentro do que se espera. Vale ressaltar que, na maioria dos casos, dados internos e externos serão necessários.

3 - Coleta de dados

Essa é a etapa mais operacional do processo, ou seja, é a coleta de informações propriamente dita. A qualidade das fontes deve ser levada em conta, bem como a periodicidade da coleta. O estudo vai ser algo contínuo, periódico ou ocasional? Isso varia de acordo com o objeto de estudo.

4 - Crítica dos dados

Depois que as informações foram reunidas, é hora de verificar se há erros ou inconsistências que possam afetar o resultado final. Se os dados foram coletados por meio de questionários, por exemplo, deve-se verificar se as respostas obtidas condizem com a necessidade para o estudo.

5 - Apuração dos dados

Feita a triagem dos dados, o próximo passo é contabilizar e organizar tudo. Nesta etapa, é importante que tudo seja separado dentro de uma lógica que facilite o trabalho de análise. Aqui também pode ser condensado o que for mais relevante, o que vai ajudar na próxima etapa.

6 - Apresentação dos dados

As considerações levantadas com base na apuração dos dados serão transformadas em uma apresentação, que facilite o entendimento e a visibilidade. Ela pode ser tabular - focada apenas em números -, ou gráfica - utilizando tabelas e infográficos para deixar a leitura mais agradável.

7 - Análise e interpretação dos dados

A última fase é a mais importante e também a mais delicada. Neste ponto, tudo o que foi coletado, tabulado e apresentado será transformado em resoluções e insights que possam ser úteis na resolução do problema que foi definido na primeira etapa.