

Universidade Federal de Uberlândia
Sistemas de Informação
Estatística

Guilherme Rimoldi Kameoka – 12011BSI200

Pedro Henrique Silva Santana – 12011BSI218

Victor Hugo Martins Alves – 12011BSI217

1. INTRODUÇÃO

❖ O que é estatística?

Estatística é a ciência que utiliza as teorias probabilísticas para explicar a frequência da ocorrência de eventos, tanto em estudos observacionais quanto em experimentos para modelar a aleatoriedade e a incerteza de forma a estimar ou possibilitar a previsão de fenômenos futuros, conforme o caso.

Para a área de computação, a estatística é de suma importância pois a partir de cálculos, análise de dados e experimentações, o pesquisador poderá ter acesso a dados brutos responsáveis por aprimorar aplicações, desenvolver tecnologias e otimizar serviços.

❖ Aplicações na área

- **Ciências de dados:** Analisar, transformar, filtrar dados, avaliar e otimizar algoritmos de machine learning, etc;
- **Serviços de Streaming/ Serviços de Entrega:** Com base no perfil do usuário e do conteúdo consumido, é possível calcular novos conteúdos relacionados;
- **Redes sociais e Publicidade Direcionada:** Assim como no tópico anterior, publicidades podem ser personalizadas para o usuário na internet com base em suas preferências de acesso e compras, e serviços oferecidos a usuários com base em suas conexões;
- **Inteligência Artificial:** “A Inteligência Artificial não é inteligente, é pura estatística”. A partir de testes, simulações e análise de dados, é possível definir ações automáticas, respostas recorrentes, entre outras aplicações para a IA;

2. METODOLOGIA

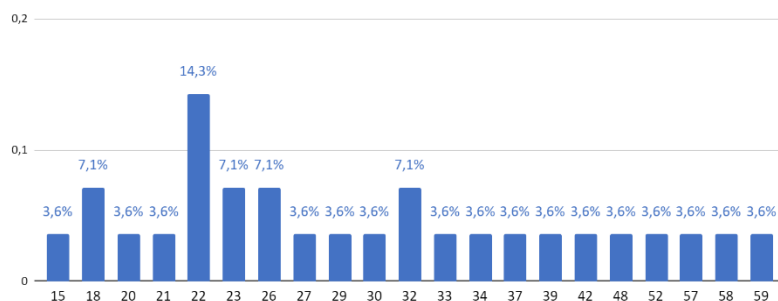
O trabalho teve como variáveis QUANTITATIVAS a idade e o número de horas que os elementos passam em frente a uma tela; e QUALITATIVAS a escolaridade e o modelo de smartphone que possuem.

Inicialmente um formulário foi disponibilizado para que fosse possível a coleta dos dados. A população se mostrou bastante diversificada em questão da idade como pode ser observado nos tópicos adiante.

3. DESENVOLVIMENTO

IDADE:

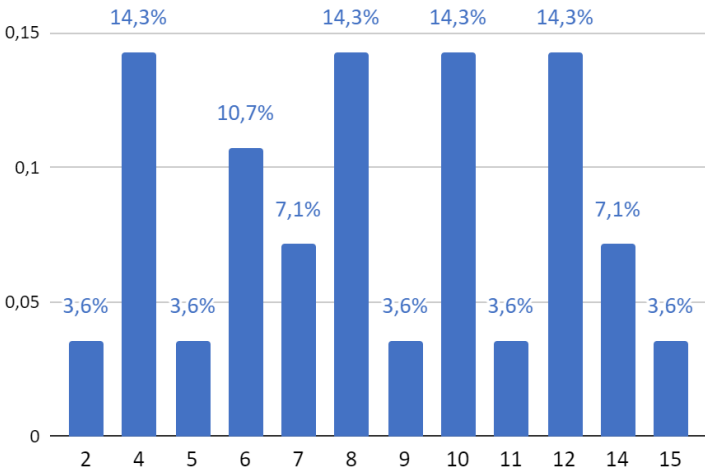
IDADE	PESSOAS (Fa)	Fr	Fr(%)
15	1	0,036	3,6%
18	2	0,071	7,1%
20	1	0,036	3,6%
21	1	0,036	3,6%
22	4	0,143	14,3%
23	2	0,071	7,1%
26	2	0,071	7,1%
27	1	0,036	3,6%
29	1	0,036	3,6%
30	1	0,036	3,6%
32	2	0,071	7,1%
33	1	0,036	3,6%
34	1	0,036	3,6%
37	1	0,036	3,6%
39	1	0,036	3,6%
42	1	0,036	3,6%
48	1	0,036	3,6%
52	1	0,036	3,6%
57	1	0,036	3,6%
58	1	0,036	3,6%
59	1	0,036	3,6%
TOTAL	28	1,00	100%



Média	Moda	Variância	Desvio Padrão	Coeficiente de Variação
31,679	22	165,930	12,881	40,66%

HORAS EM FRETE A UMA TELA:

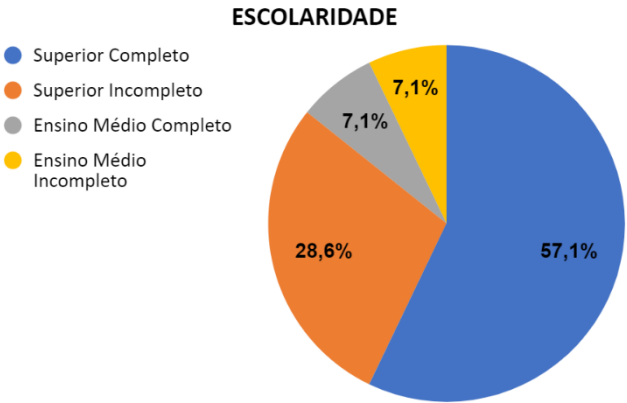
HORAS	PESSOAS (Fa)	Fr	Fr(%)
2	1	0,036	3,6%
4	4	0,143	14,3%
5	1	0,036	3,6%
6	3	0,107	10,7%
7	2	0,071	7,1%
8	4	0,143	14,3%
9	1	0,036	3,6%
10	4	0,143	14,3%
11	1	0,036	3,6%
12	4	0,143	14,3%
14	2	0,071	7,1%
15	1	0,036	3,6%
TOTAL	28	1,00	100%



Média	Moda	Variância	Desvio Padrão	Coefficiente de Variação
8,464	Multimodal 4 8 10 12	11,962	3,459	40,86%

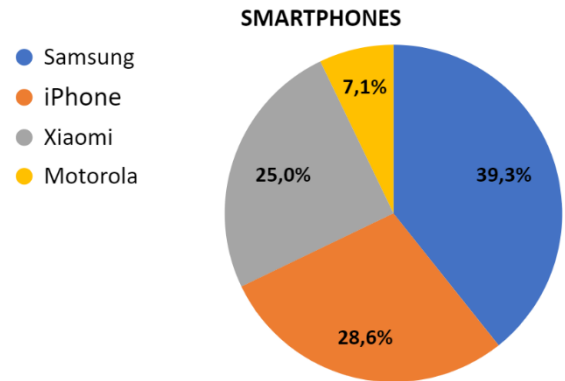
ESCOLARIDADE:

ESCOLARIDADE	PESSOAS (Fa)	Fr	Fr(%)
Superior Completo	16	0,571	57,1%
Superior Incompleto	8	0,286	28,6%
Ensino Médio Completo	2	0,071	7,1%
Ensino Médio Incompleto	2	0,071	7,1%
Fundamental Completo	0	0,000	0,0%
Fundamental Incompleto	0	0,000	0,0%
TOTAL	28	1,00	100%



SMARTPHONE:

SMARTPHONE	PESSOAS (Fa)	Fr	Fr(%)
Samsung	11	0,393	39,3%
iPhone	8	0,286	28,6%
Xiaomi	7	0,250	25,0%
Motorola	2	0,071	7,1%
Asus	0	0,000	0,0%
Nokia	0	0,000	0,0%
TOTAL	28	1,00	100,0%



A Variável Aleatória escolhida foi: Retirar uma pessoa que possua um iPhone.

Y	Freq.	Prob.
0	20	0,714
1	8	0,286
TOTAL	28	1,0

0: Não possui iPhone; 1: Possui iPhone.

A distribuição probabilística que melhor se adequa a nossa variável é a de Bernoulli, pois é realizada apenas uma única vez e por se tratar de variáveis que assumem apenas dois resultados; no caso se possui OU não o aparelho iPhone.

O intervalo de confiança para as variáveis quantitativas são:

IDADE

c = 95%

$\alpha = 5\% = 0,05$

$\frac{\alpha}{2} = 0,025$

n = 28

GL = 27

$\bar{X} = 31,679$

s = 12,881

$$e = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$e = 2,052 \cdot \frac{12,881}{\sqrt{28}}$$

$$e = 4,995$$

$$IC(\mu)95\% = 31,679 \pm 4,995$$

HORAS

$$c = 95\%$$

$$\alpha = 5\% = 0,05$$

$$\frac{\alpha}{2} = 0,025$$

$$n = 28$$

$$GL = 27$$

$$\bar{X} = 8,464$$

$$s = 3,459$$

$$e = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$e = 2,052 * \frac{3,459}{\sqrt{28}}$$
$$e = 1,341$$

$$IC(\mu)95\% = 8,464 \pm 1,341$$

4. CONCLUSÃO

Com base nos dados apurados, podemos observar que:

1. Para Idade:
 - a. A população da nossa pesquisa possui uma média de 31 anos, o que se apresenta ser uma população jovem;
 - b. A maioria das respostas foram de jovens de 22 anos;
 - c. Com base no Coeficiente de Variação da idade, nossa amostragem é de alta dispersão com dados heterogêneos;
2. Para Horas:
 - a. Em média, a população passa 1/3 do dia em frente a uma tela;
 - b. Por apresentar 4 valores que se repetiram igualmente, temos uma amostragem multimodal (4|8|10|12);
 - c. Com base no Coeficiente de Variação das horas, nossa amostragem é de alta dispersão com dados heterogêneos;
3. Para Escolaridade:
 - a. A maioria dos elementos assinalaram que possuem Superior Completo (16 pessoas do grupo de 28);
 - b. Em segundo lugar temos o Superior Incompleto (8 de 28 pessoas);
4. Para Smartphones:
 - a. A marca de smartphone mais popular segundo a pesquisa é o Samsung com 39,3% das respostas, o que representa a realidade brasileira pois a marca é de fato a mais vendida no país.

A maior dificuldade enfrentada pelo grupo foi a obtenção de uma amostra significativa da população.