

FIAP

NBA



# MBA EM DATA SCIENCE & AI

## APPLIED STATISTICS



**NA ÚLTIMA AULA ...**

# Na última aula

- O que é estatística
- Para que serve a estatística
- Áreas de atuação
- Estatística descritiva

# AULA 2

## Estatística descritiva

### Primeiros gráficos





Vamos aquecer os motores?



Vamos nos dividir em grupos ?



# Exercícios

- 1) Leia a base salario.csv e selecione as variáveis **salario** e **anosexperiencia**.
- 2) Calcule todas as medidas de resumo para as variáveis contínuas.
- 3) Repita o cálculo cruzando as variáveis **posição** e **sexo**.  
(separadamente)
- 4) Repita o cálculo cruzando as variáveis **posição** e **sexo**. (Em conjunto)
- 5) Dê uma interpretação para os resultados.

- Posição em do menor para o maior significa o cargo na Companhia
- Salário em mil/ano
- Sexo: 1 é Masculino 0 é Feminino

# Frequência Absoluta e Relativa

Considere a variável grau de Instrução dos dados da tabela (Variável qualitativa)

	Grau de instrução	Contagem	$f_i$	$f_{r_i}$	$f_{r_i} \%$
	Fundamental		12	0,3333	33,3%
	Médio		18	0,5000	50 %
	Superior		6	0,1667	16.7%
	total		n=36	1,0000	100%

$f_i$  : **Frequência absoluta** da categoria  $i$  (número de indivíduos que pertencem à categoria  $i$ )

$f_{r_i} = \frac{f_i}{n}$  : **Frequência relativa** da categoria  $i$

$f_{r_i} \% = f_{r_i} * 100\%$  : **Frequência relativa percentual** da categoria  $i$



# Distribuição de Frequência

O número de vezes que ocorreram valores em cada classe ou valores chama-se frequência absoluta. O conjunto das ocorrências, com correspondentes frequências absolutas ( FA ) e relativas ( FR ), define a distribuição de frequências da variável. Conhecer o comportamento da variável.

Distribuição etária dos trabalhadores da Empresa XXX, 01/05/2019

Faixa etária	Frequency	Percent	Cumulative Frequency	Cumulative Percent
00 - 17	19052	33,8	19052	33,8
18 - 29	16143	28,6	35195	62,4
30 - 39	13710	24,3	48905	86,7
40 - 49	5773	10,2	54678	96,9
50 - 59	1559	2,8	56237	99,7
60 - 69	174	0,3	56411	100,0
Acima 69	13	0,0	56424	100,0
Total	56424	100,0		

# Exemplo no Python

Aula02\_8DTSR.ipynb

# Exercícios

- 1) Leia a base basefibrose.csv
- 2) Calcule a distribuição absoluta e relativa de todas as variáveis.
- 3) Dê uma interpretação para os resultados.

# Medidas de Dispersão

Medidas de Dispersão: variância e desvio padrão Exemplo C

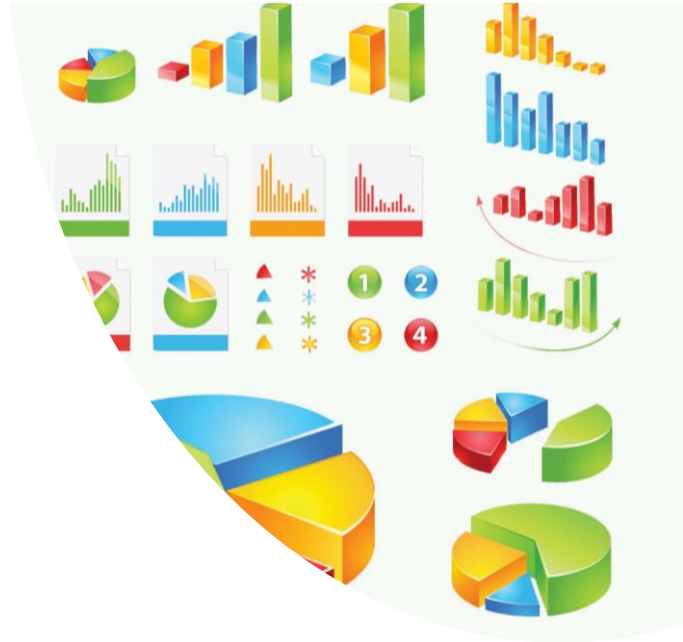
X	Média	(X-Média)	(X-Média) <sup>2</sup>
1	4	-3	9
1	4	-3	9
3	4	-1	1
3	4	-1	1
5	4	1	1
5	4	1	1
7	4	3	9
7	4	3	9
Soma	-	0	40

Variância:

$$\sigma^2 = \frac{40}{8} = 5$$

Desvio padrão:

$$\sigma = \sqrt{\sigma^2} = \sqrt{5} = 2.24$$



# Plots

Análises Gráficas  
com o Python

# Por onde começar

- Você já conhece minimamente seus dados?
- Quais perguntas quer responder?

# Pacotes de plotagens mais conhecidos

- **Matplotlib**

- **Descrição:** Um dos pacotes de visualização mais utilizados em Python, conhecido por sua flexibilidade e capacidade de criar uma ampla variedade de gráficos, desde gráficos simples até figuras complexas.
- **Uso Comum:** Gráficos de linha, barras, histogramas, gráficos de dispersão, etc.

- **Seaborn**

- **Descrição:** Construído sobre o Matplotlib, o Seaborn oferece uma interface mais amigável e ferramentas para criar gráficos estatísticos mais atraentes e informativos.
- **Uso Comum:** Mapas de calor, gráficos de violino, gráficos de distribuição, etc.

- **Plotly**

- **Descrição:** Uma biblioteca de gráficos interativos que pode ser usada em notebooks Jupyter e também em aplicações web. Oferece gráficos em 2D e 3D.
- **Uso Comum:** Gráficos interativos, gráficos de linhas, gráficos de dispersão 3D, mapas, etc.



# seaborn

**[HTTPS://SEABORN.PYDATA.ORG/](https://seaborn.pydata.org/)**



# Preparando o terreno

```
import seaborn as sns
```

# Preparando o terreno...

```
#carregando a base
```

```
imdb <- readr::read_rds("imdb.rds")
```

```
#chamando o ggplot
```

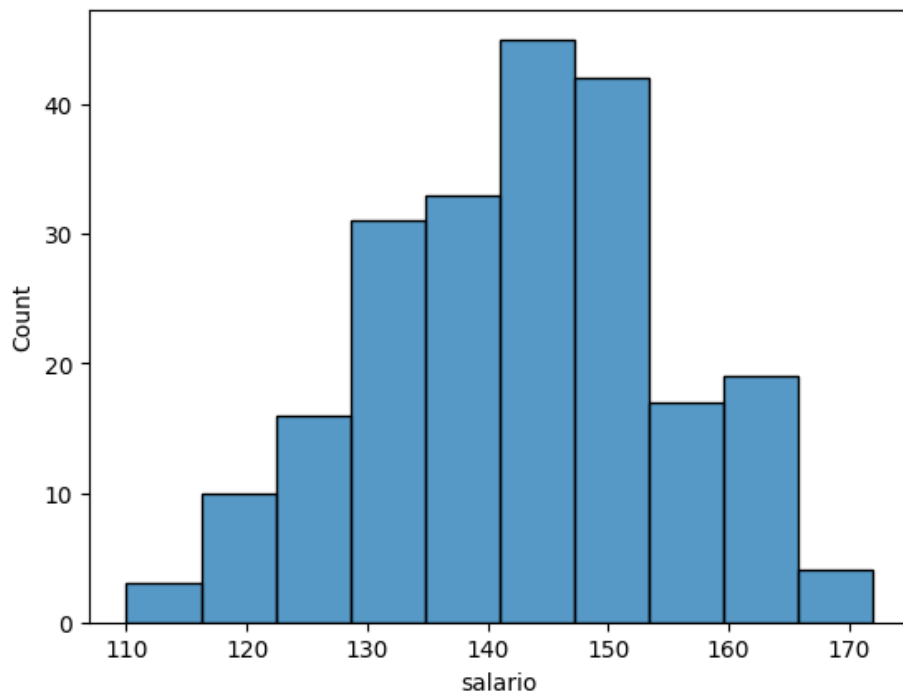
```
ggplot(data = imdb)
```

```
#Com o pipe
```

```
imdb %>% ggplot()
```

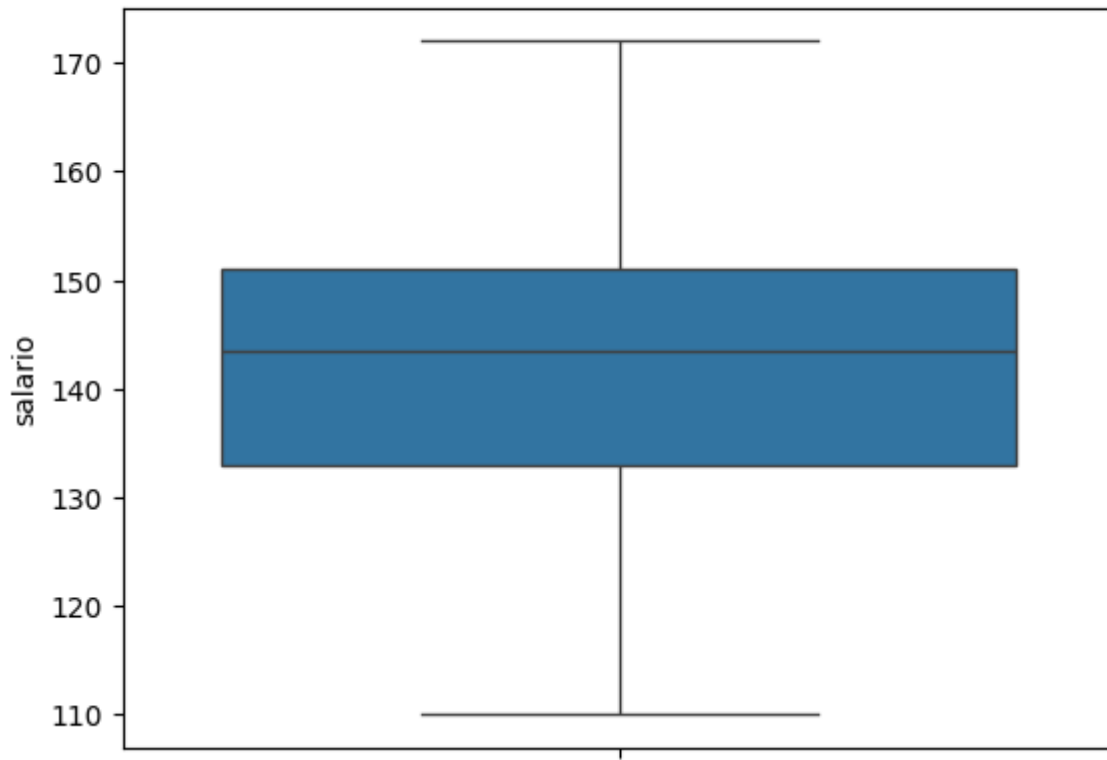
# Histograma

```
✓ 0s ax = sns.histplot(x = 'salario', data = df, bins = 10)
```



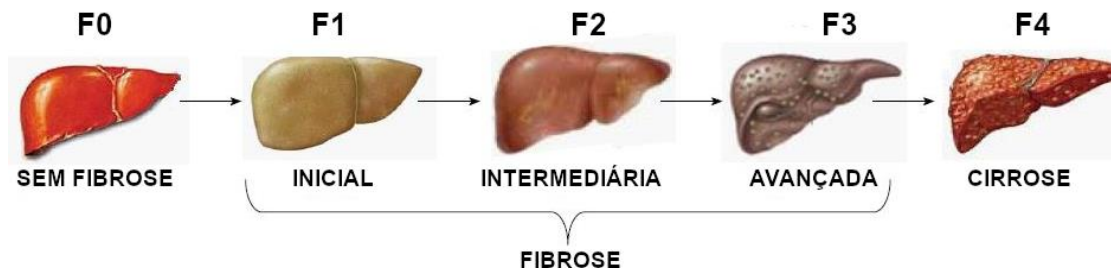
# Box Plot

```
ax = sns.boxplot(y = 'salario', data = df)
```



# Exercícios

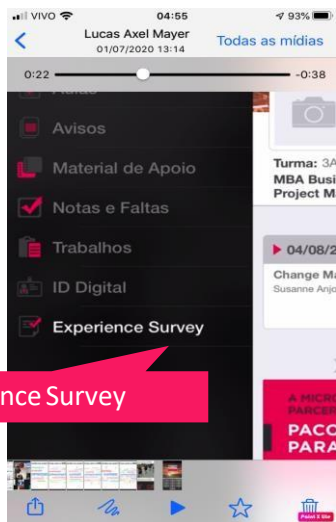
- 1) Leia a base basefibrose.csv
- 2) Calcule as medidas de resumo para as variáveis contínuas
- 3) Calcule a distribuição absoluta e relativa de todas as variáveis categoricas.
- 4) Faça os gráficos de histograma por Grau de Fibrose.



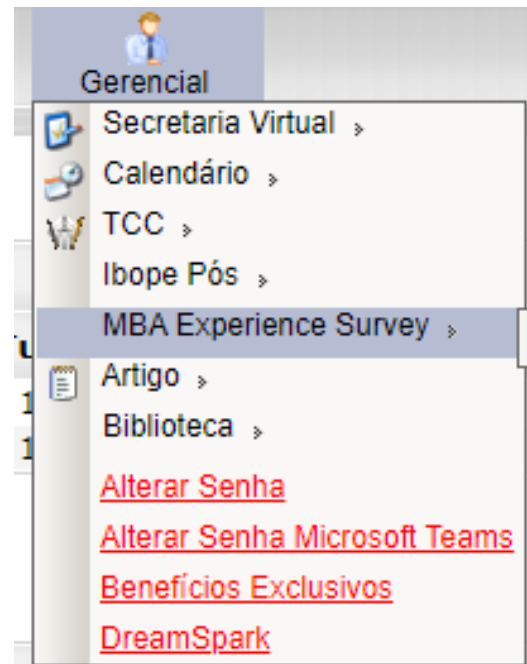
# O que você achou da aula de hoje?

Pelo aplicativo da FIAP

(Entrar no FIAPP, e no menu clicar em Experience Survey)



Experience Survey



# OBRIGADO

 /lafphd

profleandro.ferreira@fiap.com.br

FIAP MBA<sup>+</sup>

Copyright © 2019 | Professor (a) Nome do Professor  
Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente  
proibido sem consentimento formal, por escrito, do professor/autor.

FIAP