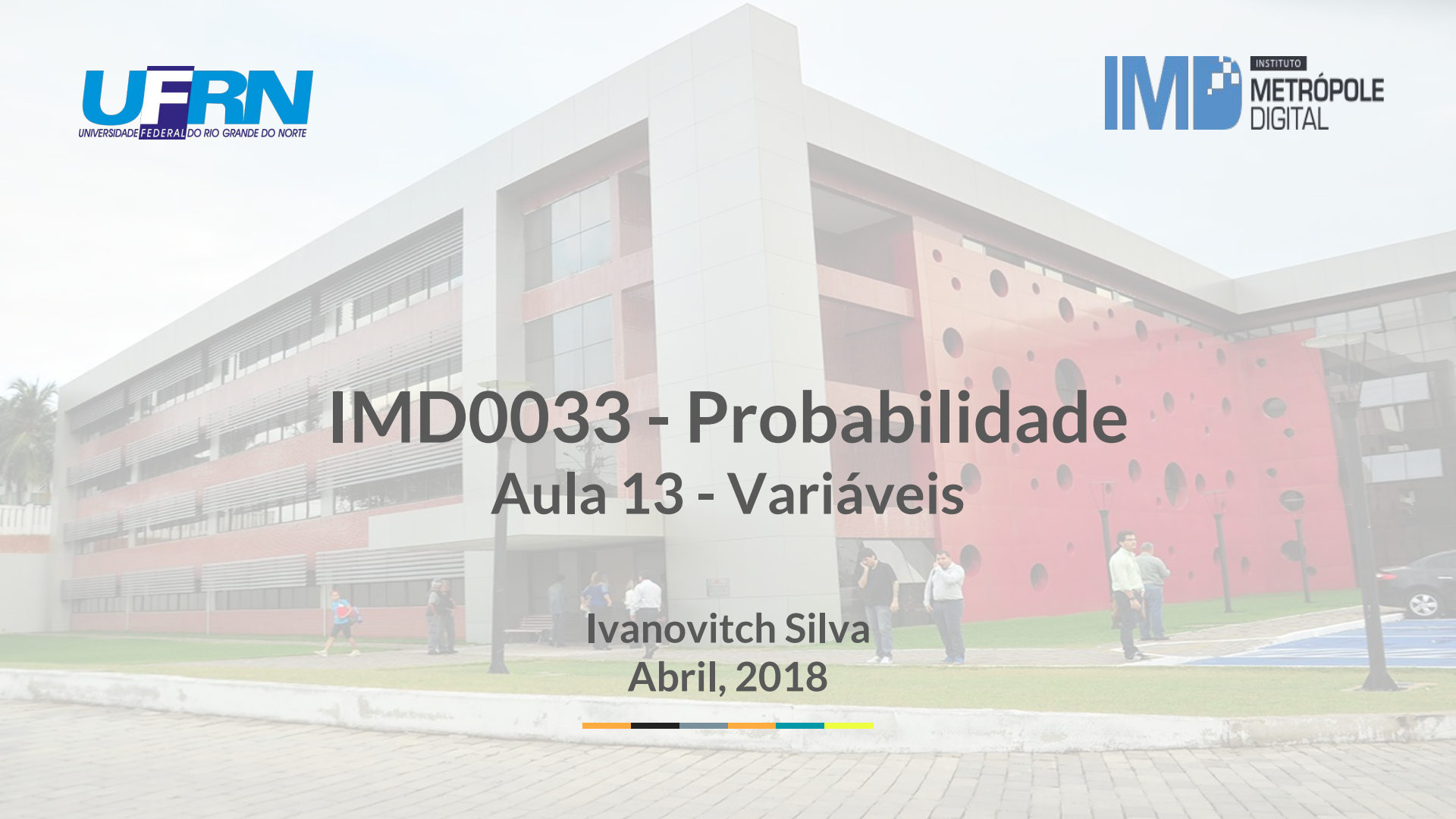


# IMD0033 - Probabilidade

## Aula 13 - Variáveis

Ivanovitch Silva

Abril, 2018

# Agenda

- Variables in statistics
- Quantitative and qualitative variables
- Scale of measurements (nominal, ordinal, interval, ratio)

# Atualizar o repositório

```
git clone https://github.com/ivanovitchm/imd0033_2019_1.git
```

Ou ....

```
git pull
```

# PREVIOUSLY ON...

Data **source**

Getting **good** data

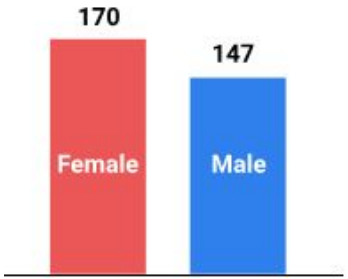
| Id | Name | Salary | ... | Gender |
|---|---|---|---|---|
| 1 | Mary Ann | $35 000 | .... | Female |
| 2 | Marc Downey | $55 000 | .... | Male |
| .. | .... | .... | .... | .... |
| 51 | Juliet Ali | $45 000 | .... | Female |
| ... | ... | .... | .... | .... |
| 317 | Jane Ace | $95 000 | .... | Female |

Understand how the data is **structured** and **measured**

**Organize** the data in **comprehensible** forms to find patterns

| Gender | Frequency |
|---|---|
| Male | 147 |
| Female | 170 |

**Visualize** the patterns



# Introduction

| | Name | Team | Pos | Height | Weight | BMI | Birth_Place | Birthdate |
|---|---|---|---|---|---|---|---|---|
| 39 | Crystal Langhorne | SEA | F/C | 188 | 84.0 | 23.766410 | US | October 27, 1986 |
| 52 | Érika de Souza | SAN | C | 196 | 86.0 | 22.386506 | BR | September 3, 1982 |
| 102 | Nia Coffey | SAN | F | 185 | 77.0 | 22.498174 | US | May 21, 1995 |

The properties with varying values we call **variables**

# Quantitative and Qualitative Variables

| | **Quantitative** variables | **Qualitative** variables |
|---|---|---|
| Describe **quantities** | YES | NO |
| Describe **qualities** | NO | YES |
| Use **numbers** | YES | YES |
| The numbers are **actual quantities** | YES | NO |
| Use **words** | YES | YES |
| The words express a **quantity** | YES | NO |

| | |
|---|---|
| Height | Name |
| BMI | Team |
| Age | Pos |
| Birth_Data | Birth_Place |
| Weight | College |

# Scale of Measurements

The system of rules that define how each variable is measured is called **scale of measurement**

| | Team | Height |
|---|---|---|
| We can tell whether two individuals are **different** | YES | YES |
| We can tell **the size** of the difference | NO | YES |
| We can tell the **direction** of the difference | NO | YES |

- Nominal
- Ordinal
- Interval
- Ratio

# Nominal Scale

| | Nominal |
|---|---|
| We can tell whether two individuals are **different** | YES |
| We can tell the **direction** of the difference | NO |
| We can tell **the size** of the difference | NO |
| We can measure **quantitative** variables | NO |
| We can measure **qualitative** variables | YES |

| | Name | Team | Pos | Birth_Place | College |
|---|---|---|---|---|---|
| 0 | Aerial Powers | DAL | F | US | Michigan State |
| 1 | Alana Beard | LA | G/F | US | Duke |
| 2 | Alex Bentley | CON | G | US | Penn State |
| 3 | Alex Montgomery | SAN | G/F | US | Georgia Tech |
| 4 | Alexis Jones | MIN | G | US | Baylor |

# Ordinal Scale (ranking)

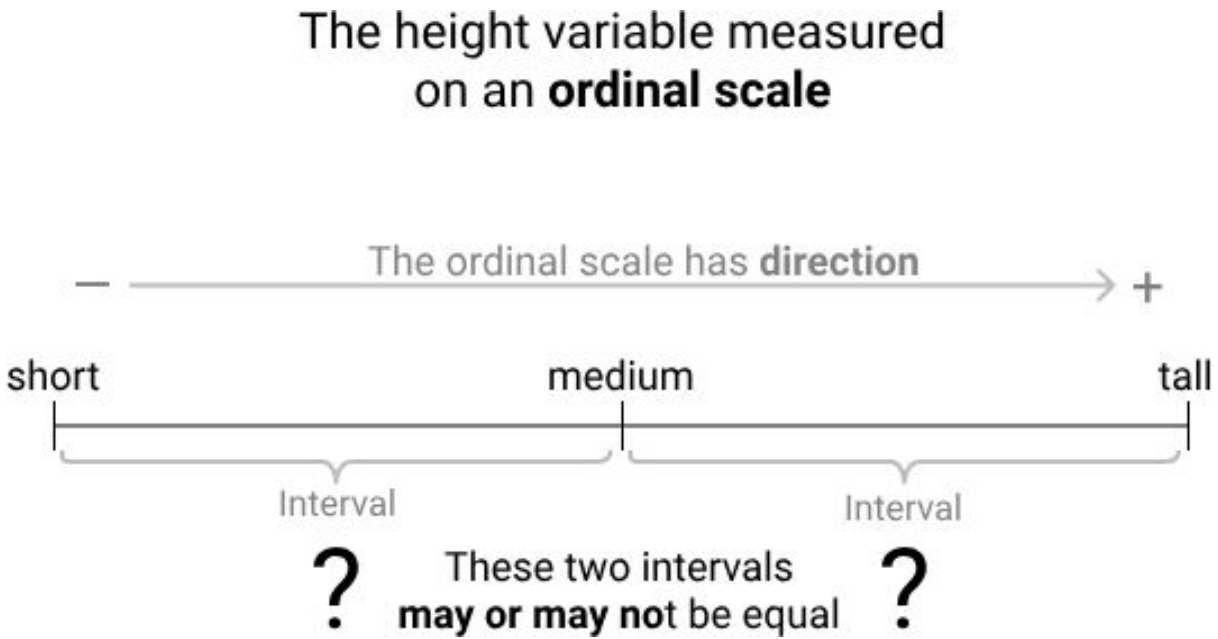
| | Nominal | Ordinal |
|---|---|---|
| We can tell whether two individuals are **different** | YES | YES |
| We can tell the **direction** of the difference | NO | YES |
| We can tell **the size** of the difference | NO | NO |
| We can measure **quantitative** variables | NO | YES |
| We can measure **qualitative** variables | YES | NO |

| | Height | Height_labels |
|---|---|---|
| 0 | 183 | tall |
| 1 | 185 | tall |
| 2 | 170 | short |
| 3 | 185 | tall |
| 4 | 175 | medium |

# Nominal or Ordinal??

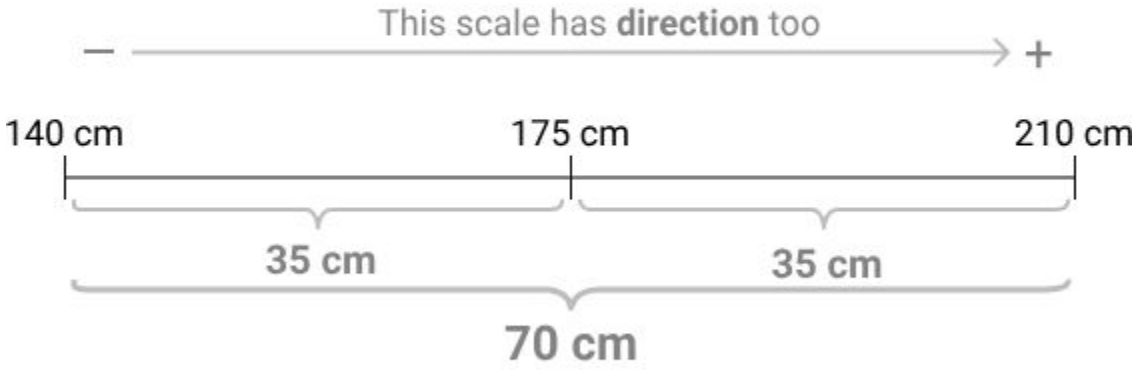
| | Height_labels | College | Games Played | Experience |
|---|---|---|---|---|
| **0** | tall | Michigan State | 8 | 2 |
| **1** | tall | Duke | 30 | 12 |
| **2** | short | Penn State | 26 | 4 |
| **3** | tall | Georgia Tech | 31 | 6 |
| **4** | medium | Baylor | 24 | R |

# The interval and ratio scales



# The interval and ratio scales

The height variable measured
on a scale that uses **real numbers**




We know the value of each interval, which means **we can compute the size of the difference** between any two points.

| | Nominal | Ordinal | Interval | Ratio |
|---|---|---|---|---|
| We can tell whether two individuals are **different** | YES | YES | YES | YES |
| We can tell the **direction** of the difference | NO | YES | YES | YES |
| We can tell **the size** of the difference | NO | NO | YES | YES |
| We can measure **quantitative** variables | NO | YES | YES | YES |
| We can measure **qualitative** variables | YES | NO | NO | NO |

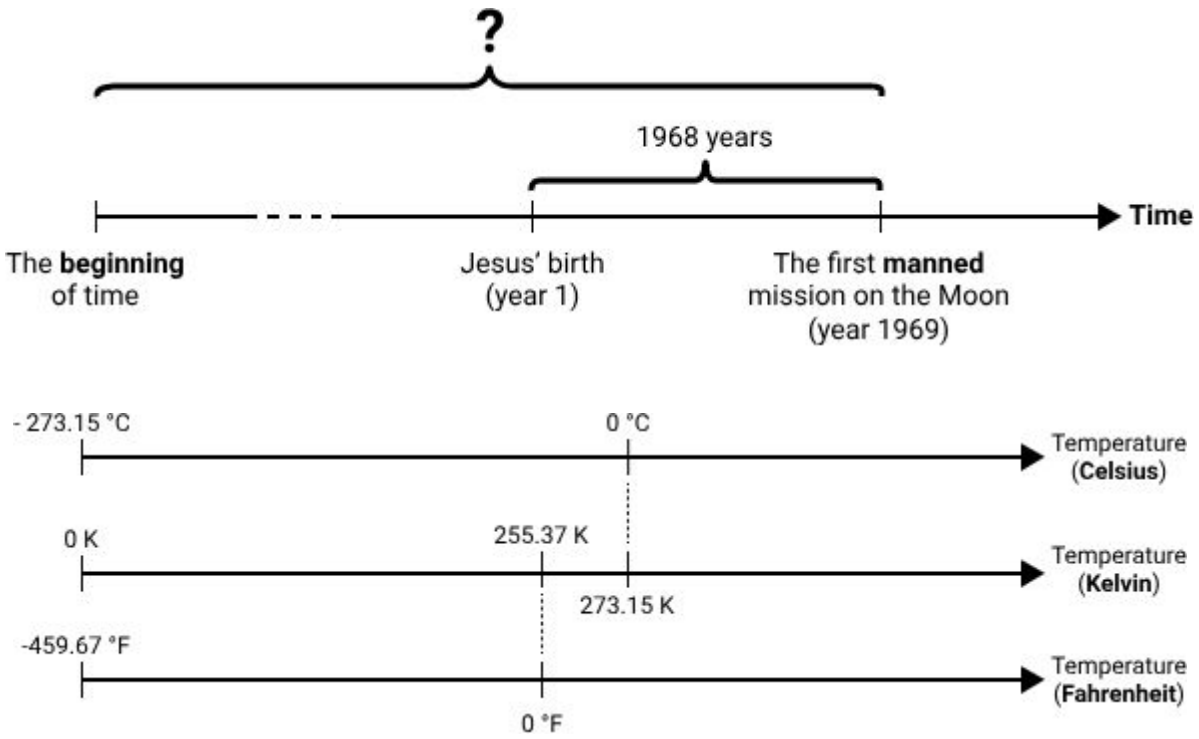
# The difference between ratio and interval scales

What sets apart ratio scales from interval scales is the nature of the zero point.

| _ | Name | Weight | Weight_deviation |
|---|---|---|---|
| 35 | Clarissa dos Santos | 89.0 | 10.021127 |
| 3 | Alex Montgomery | 84.0 | 5.021127 |
| 111 | Renee Montgomery | 63.0 | -15.978873 |
| 85 | Layshia Clarendon | 64.0 | -14.978873 |
| 128 | Sugar Rodgers | 75.0 | -3.978873 |

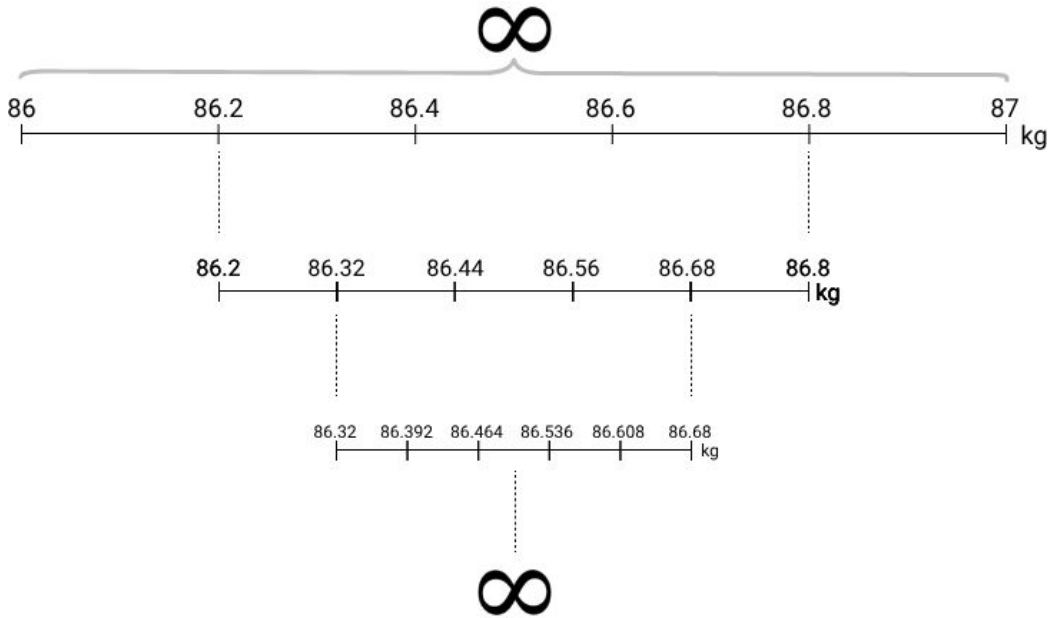
| | Interval | Ratio |
|---|---|---|
| Well-defined **intervals** | YES | YES |
| The **zero point** indicates the **absence of a quantity** | NO | YES |
| Difference measured in terms of **distance** | YES | YES |
| Difference measured in terms of **ratios** | NO | YES |

# Common Examples of Interval Scales



# Discrete and Continuous Variable

| _ | Name | Weight | PTS |
|---|---|---|---|
| 77 | Kayla Thornton | 86.0 | 32 |
| 16 | Asia Taylor | 76.0 | 31 |
| 80 | Kia Vaughn | 90.0 | 134 |
| 137 | Tierra Ruffin-Pratt | 83.0 | 225 |
| 12 | Amanda Zahui B. | 113.0 | 51 |



Data **source**

Getting **good** data

| Id | Name | Salary | ... | Gender |
|---|---|---|---|---|
| 1 | Mary Ann | $35 000 | .... | Female |
| 2 | Marc Downey | $55 000 | .... | Male |
| .. | .... | .... | .... | .... |
| 51 | Juliet Ali | $45 000 | .... | Female |
| ... | ... | .... | .... | .... |
| 317 | Jane Ace | $95 000 | .... | Female |

Understand how the data is **structured** and **measured**

**Organize** the data in **comprehensible** forms to find patterns

| Gender | Frequency |
|---|---|
| Male | 147 |
| Female | 170 |

**Visualize** the patterns