


IMD0033 - Probabilidade

Lesson 25 - The Mode

Ivanovitch Silva
December, 2018



Agenda

- The mode
 - ordinal variables
 - nominal variables
 - discrete variables
- Special cases
 - unimodal
 - bimodal
 - multimodal
- Skewed distributions
- Symmetrical distributions

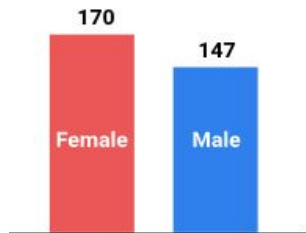
Atualizar o repositório

```
git clone https://github.com/ivanovitchm/imd0033_2018_2.git
```

Ou

```
git pull
```

PREVIOUSLY ON...

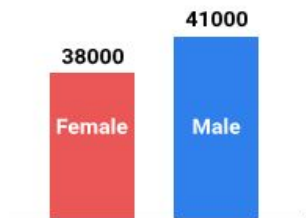


Visualize
the patterns

Gender	Frequency
Male	147
Female	170

By measuring frequencies
and build **frequency tables**

Organize the data in
comprehensible forms



Visualize
the patterns

Gender	Average salary
Male	41000
Female	38000

By summarizing it
with **average values**

Depending on the particular characteristics of a distribution, we'll see that we can summarize it using the **mean**, the **weighted mean**, the **median**, or the **mode**.

Introduction

- Mean
 - It takes into account each value in the distribution
 - It can be defined algebraically.
 - These two properties make the mean far superior to the median.
- Median
 - It comes in handy, however, when it's not possible or appropriate to compute the mean.

The dataset

	Order	PID	MS SubClass	MS Zoning	Lot Frontage	Lot Area	Street	Alley	Lot Shape	Sale Condition	SalePrice
0	1	526301100	20	RL	141.0	131770	Pave	NaN	WD	Normal	215000
1	2	526350040	20	RH	80.0	11622	Pave	NaN	WD	Normal	105000
2	3	526351010	20	RL	81.0	14267	Pave	NaN	WD	Normal	172000
3	4	526353030	20	RL	93.0	11160	Pave	NaN	WD	Normal	244000
4	5	527105010	60	RL	74.0	13830	Pave	NaN	WD	Normal	189900

```
1 | houses["Land Slope"].unique().
```

```
array(['Gtl', 'Mod', 'Sev'], dtype=object)
```

```
1 | houses["Roof Style"].unique().
```

```
array(['Hip', 'Gable', 'Mansard', 'Gambrel', 'Shed', 'Flat'], dtype=object)
```

The mode for Ordinal Variables

- We call the **most frequent value** in the distribution the **mode**.
- The mode of the **Land Slope** variable is '**Gtl**'. In other words, the typical house has a gentle slope.

```
1 | houses[ 'Land Slope' ].value_counts()._
```

Gtl	2789
Mod	125
Sev	16

The mode for nominal variables

- The mode is also a good choice for nominal data.
- We obviously **can't compute the mean** for this variable because the values are words.
- Because the Roof Style variable is nominal, **there's also no inherent order** of the values in the distribution. We can't compute the median.

```
1 | houses['Roof Style'].value_counts(.)
```

```
Gable      2321  
Hip         551  
Gambrel     22  
Flat        20  
Mansard     11  
Shed         5
```

```
Name: Roof Style, dtype: int64
```


The mode for discrete variables

- There are some cases where computing the **mean** and the **median** is possible and correct, but the **mode** is preferred nonetheless.
- This is sometimes the case for discrete variables.

```
1 houses['Kitchen AbvGr'].value_counts().sort_index().
```

0	3
1	2796
2	129
3	2

	Value	Reporting to <i>non-technical</i> audiences
Mean	1.04	The average house has 1.04 kitchens.
Median	1	The average house has one kitchen.
Mode	1	The typical house has one kitchen.

The mode: special cases

There are distributions that can have more than one mode.

[1, 1, 2, 2, 3, 3, 4, 4]

Without a mode

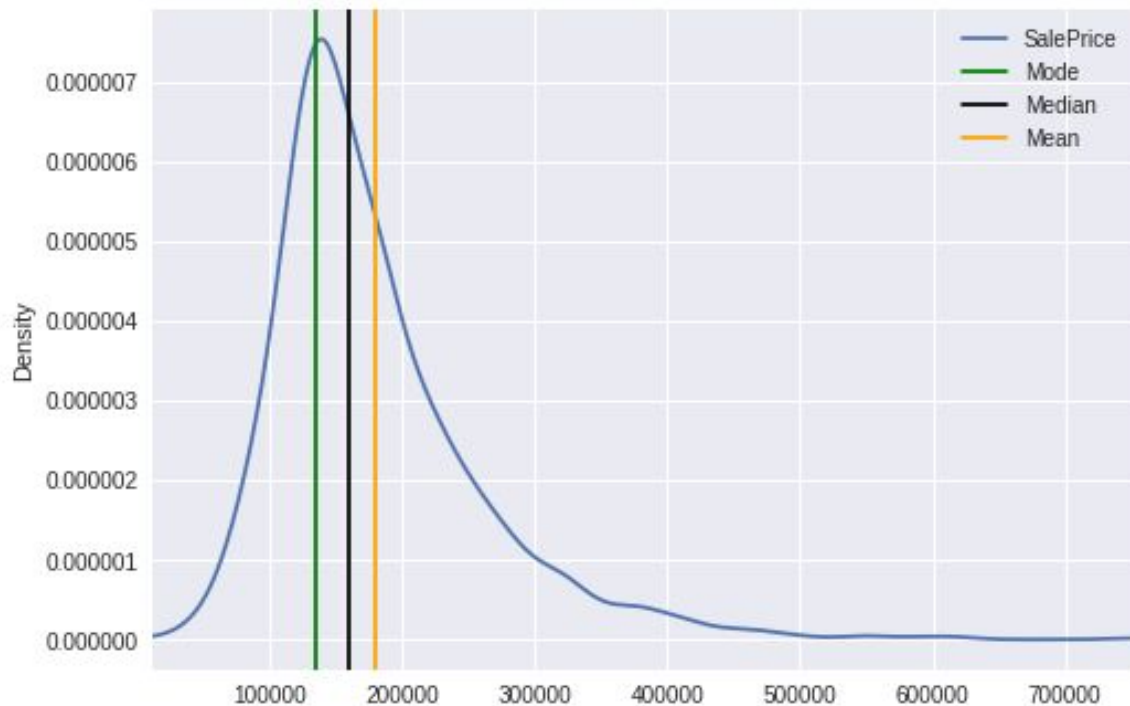
[0, 1, 1, 2, 2, 3, 3, 4, 4, 8]

Multimodal

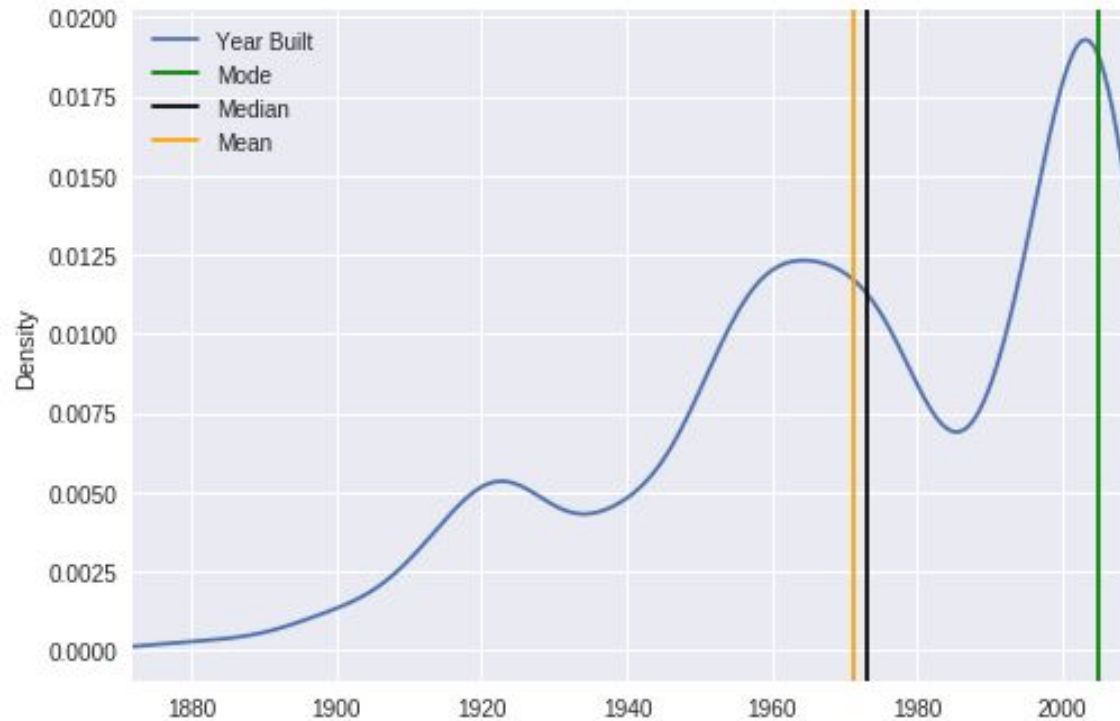
[0, 1, 1, 1, 2, 2, 2, 3]

Bimodal

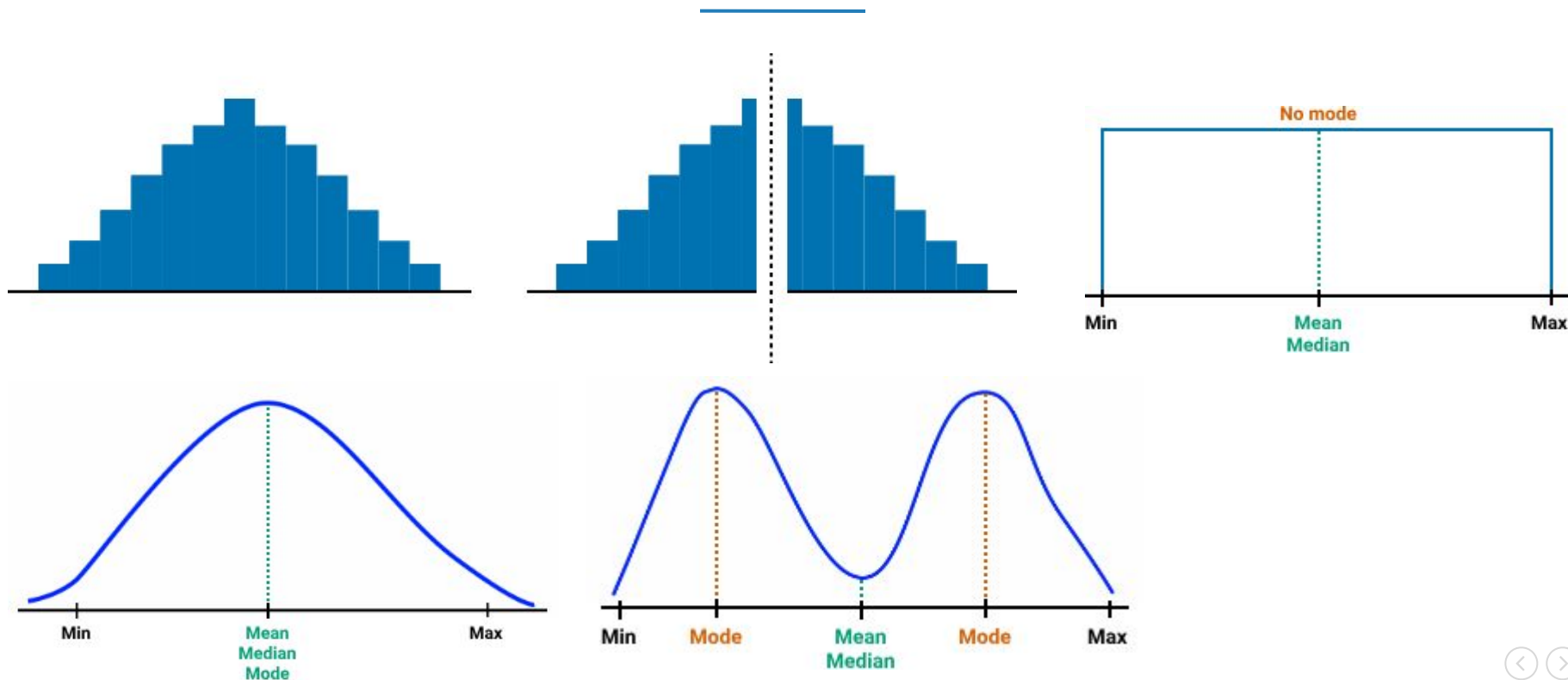
Skewed distributions



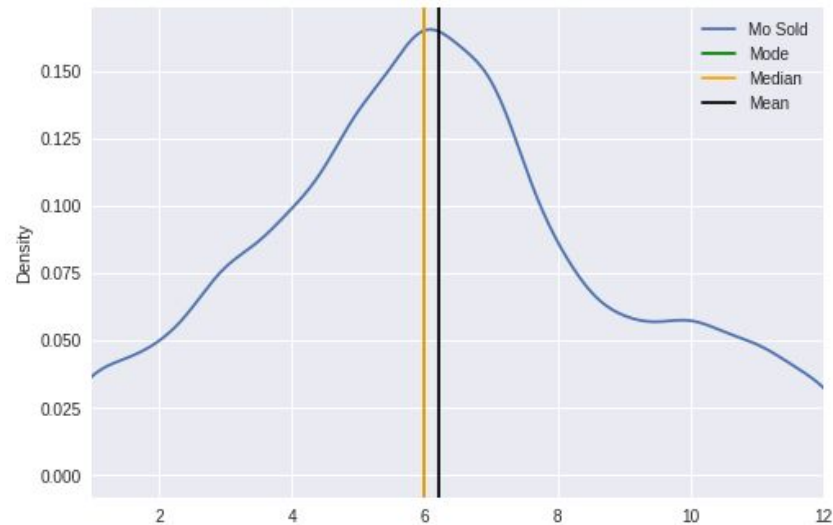
Skewed distributions



Symmetrical distribution



Symmetrical distribution



```
houses['Mo Sold'].plot.kde(xlim = [1,12])

import matplotlib.pyplot as plt
plt.axvline(houses['Mo Sold'].mode()[0], color = 'Green', label = 'Mode')
plt.axvline(houses['Mo Sold'].median(), color = 'Orange', label = 'Median')
plt.axvline(houses['Mo Sold'].mean(), color = 'Black', label = 'Mean')
plt.legend()
```

Next Steps

	Can be used for	Can't be used for	Ideal for
Mean	<ul style="list-style-type: none"> • Data measured on an interval or ratio scale • Ordinal data encoded with numbers (contentious) 	<ul style="list-style-type: none"> • Nominal data • Ordinal data that is not encoded with numbers • Data points with different weights — we need to use the weighted mean instead 	<ul style="list-style-type: none"> • Summarizing numerical distributions by taking into account each value in the distribution
Median	<ul style="list-style-type: none"> • Data measured on an interval or ratio scale • Ordinal data encoded with numbers 	<ul style="list-style-type: none"> • Nominal data • Ordinal data that is not encoded with numbers 	<ul style="list-style-type: none"> • Summarizing numerical distributions that have outliers • Open-ended distributions
Mode	<ul style="list-style-type: none"> • Nominal data • Ordinal data • Data measured on an interval or ratio scale 	<ul style="list-style-type: none"> • Perfectly uniform distributions • Continuous variables (we need to create grouped frequency tables as a workaround) 	<ul style="list-style-type: none"> • Nominal data • Ordinal data that is not encoded with numbers • Discrete data — when the result needs to be communicated to a non-technical audience



```
index.js
import React, { useState } from 'react';
import './index.css';

function App() {
  const [contacts, setContacts] = useState([
    { name: 'John Doe', phone: '123-456-7890' },
    { name: 'Jane Smith', phone: '987-654-3210' }
  ]);

  const handleClick = () => {
    // TODO: Implement handleClick logic
  };

  return (
    <div>
      <h1>Contact List</h1>
      <ul>
        {contacts.map(contact => (
          <li>{contact.name} {contact.phone}</li>
        ))}
      </ul>
      <button onClick={handleClick}>Click Me</button>
    </div>
  );
}

export default App;
```

```
index.html
<!DOCTYPE html>
<html>
  <head>
    <script src="index.js"></script>
  </head>
  <body>
    <div></div>
  </body>
</html>
```