DATA SCIENCE FOR ECONOMISTS

ECON 220 LAB

Jafet Baca-Obando

Week 5, Normal and Binomial Distributions – 09/26/2025

Outline

- Basic data analysis
- The normal distribution
- The binomial distribution

Importing required libraries and dataset

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Python

```python
path = "C:/Users/jbaca/OneDrive/Documents/2. Ph.D. in Economics/Courses/Semester 7 - Fall 2025/ECON 220 - Data Science for Economists - Lab/Lectures/Week 5/2020.csv"

data = pd.read_csv(path)
data
```

Python

Renaming columns

```
data.rename(columns={
    'Country name'                            : 'Country',
    'Regional indicator'                      : 'Region',
    'Ladder score'                            : 'Score',
    'Standard error of ladder score'          : 'StdErr',
    'upperwhisker'                            : 'UpperWhisker',
    'lowerwhisker'                            : 'LowerWhisker',
    'Logged GDP per capita'                   : 'GDP',
    'Social support'                          : 'SocialSupport',
    'Healthy life expectancy'                 : 'LifeExpectancy',
    'Freedom to make life choices'            : 'Freedom',
    'Perceptions of corruption'               : 'Corruption',
    'Ladder score in Dystopia'                : 'DystopiaScore',
    'Explained by: Log GDP per capita'        : 'GDP',
    'Explained by: Social support'            : 'SocialSupport',
    'Explained by: Healthy life expectancy'   : 'LifeExpectancy',
    'Explained by: Freedom to make life choices': 'Freedom',
    'Explained by: Generosity'                : 'Generosity',
    'Explained by: Perceptions of corruption' : 'Corruption',
    'Dystopia + residual'                     : 'DystopiaResidual'
}, inplace=True)

data.head()
```

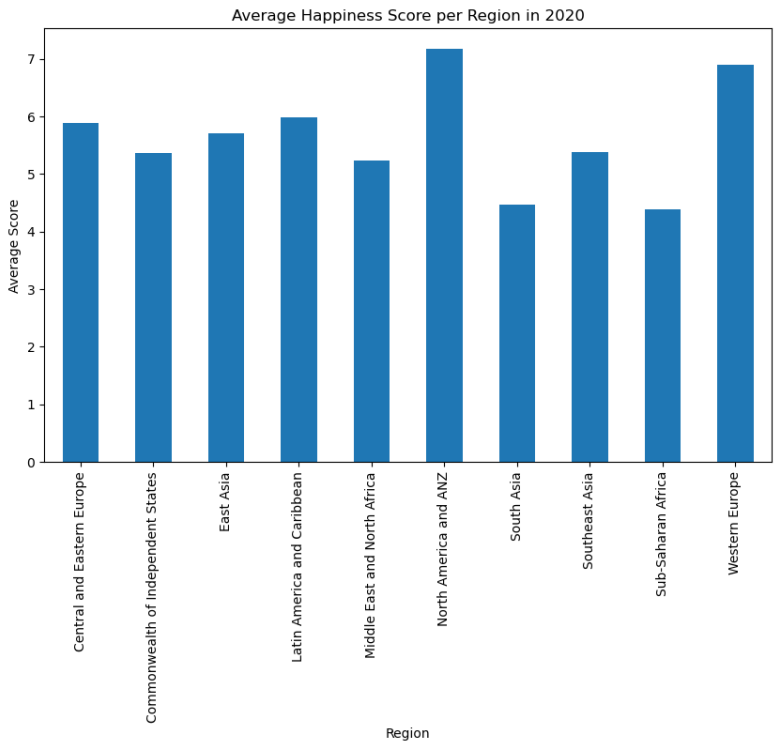
✓ 0.0s JavaScript

Which region is the happiest, on average?

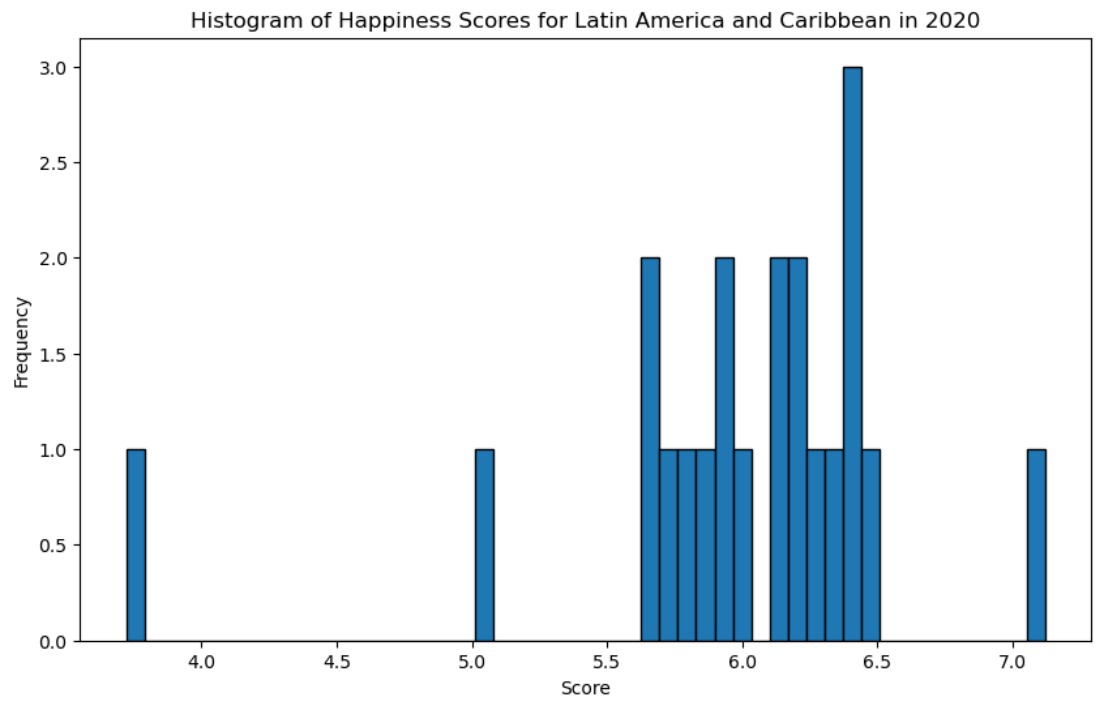
```python
# Calculate the average score per region
average_score_per_region = data.groupby('Region')['Score'].mean()

# Plot the bar graph
plt.figure(figsize=(10, 6))
average_score_per_region.plot(kind='bar')
plt.xlabel('Region')
plt.ylabel('Average Score')
plt.title('Average Happiness Score per Region in 2020')
plt.show()
```

Python

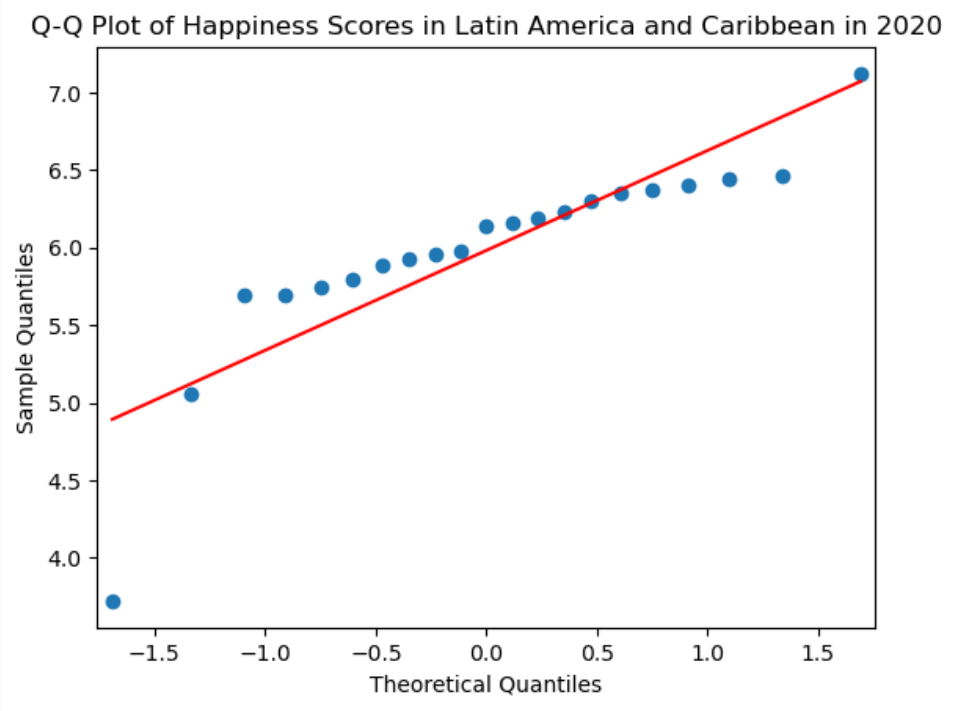
Which region is the happiest, on average?



Is the happiness score of LATAM countries normally distributed?



Is the happiness score of LATAM countries normal?



What is the probability that a LATAM country has a score below or equal to 6? Normal distribution

```python
# Calculate the mean and standard deviation of the score for Latin American countries
mean_score_lac = lac_data['Score'].mean()
std_score_lac = lac_data['Score'].std()

print(f"Mean Score for Latin American countries: {mean_score_lac}")
print(f"Standard Deviation of Score for Latin American countries: {std_score_lac}")
```

Python

```
Mean Score for Latin American countries: 5.981785706095239
Standard Deviation of Score for Latin American countries: 0.6602131810928923
```

```python
from scipy.stats import norm

# Calculate the probability
probability_below_6 = norm.cdf(6, mean_score_lac, std_score_lac)

print(f"The probability that a Latin American country has a score below 6 is {probability_below_6:.4f}")
```

Python

```
The probability that a Latin American country has a score below 6 is 0.5110
```

What is the probability that a LATAM country has a score below or equal to 6? Actual data

```python
# Count the number of Latin American countries with a score below 6
actual_count_below_6 = lac_data[lac_data['Score'] < 6].shape[0]

# Total number of Latin American countries
total_lac_countries = lac_data.shape[0]

# Calculate the actual probability
actual_probability_below_6 = actual_count_below_6 / total_lac_countries

print(f"Actual probability that a Latin American country has a score below 6 is
{actual_probability_below_6:.4f}")
```

Python

```
Actual probability that a Latin American country has a score below 6 is 0.4762
```

The binomial distribution

- Suppose we want to know the likelihood that a country has a score greater than 6 out of a sample of 6 countries.
- We can do this with the binomial distribution
- Assumptions
  - Countries are assumed to be independent from each other
  - There is a fixed number of countries,
  - A country can have a score greater than 6 or less than or equal to 6 (i.e.. two possible outcomes)
  - Each country is equally likely to be chosen in the sample

The binomial distribution

```python
from scipy.stats import binom

# Probability that a country has a score greater than 6
prob_greater_than_6 = 1 - probability_below_6

# Number of trials (countries)
n = 7

# Number of successes (countries with score greater than 6)
k = 5

# Calculate the probability using the binomial distribution
probability_5_out_of_7 = binom.pmf(k, n, prob_greater_than_6)

print(f"The probability that exactly 5 out of 7 countries in the LAC region have a score greater than 6
is {probability_5_out_of_7:.4f}")
```

Python

```
The probability that exactly 5 out of 7 countries in the LAC region have a score greater than 6 is 0.1533
```

Recap

- We analyzed the 2020 World Happiness Index data using Python.
- We learned how to compute probabilities using the normal distribution.
- We learned how to compute probabilities using the binomial distribution.

To-do list

- **Complete Data Exercise 2**
  - Upload Jupyter notebook (.ipynb file) and HTML file on **September 28**.
- Complete Data Exercise 3
  - Upload Jupyter notebook (.ipynb file) and HTML file on **October 5**.