# Statistics for International Commerce

## Week 1, Meeting 2: Course Introduction

Iegor Vyshnevskyi, Ph.D.

Sogang University, Geppert School of International Studies

March 5, 2026

Hello everyone!

What a nice day to **start learning** statistics for international commerce. :-)

# Agenda

1. What is statistics (properly defined)?
2. Types of data in international commerce
3. Population vs sample
4. Introduction to Python in Google Colab
5. Real-World Example
6. First hands-on example

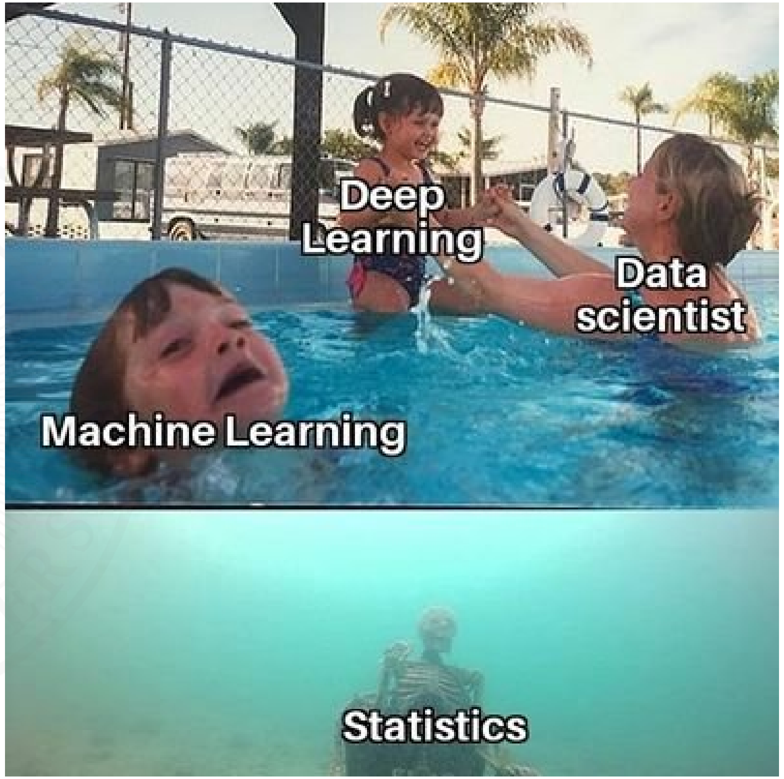
# 1. What Is Statistics — More Precisely?

# A More Technical Definition

Statistics is the science of:

- Designing data collection
- Summarizing data
- Quantifying uncertainty
- Making inferences about populations
- Supporting decision-making under uncertainty

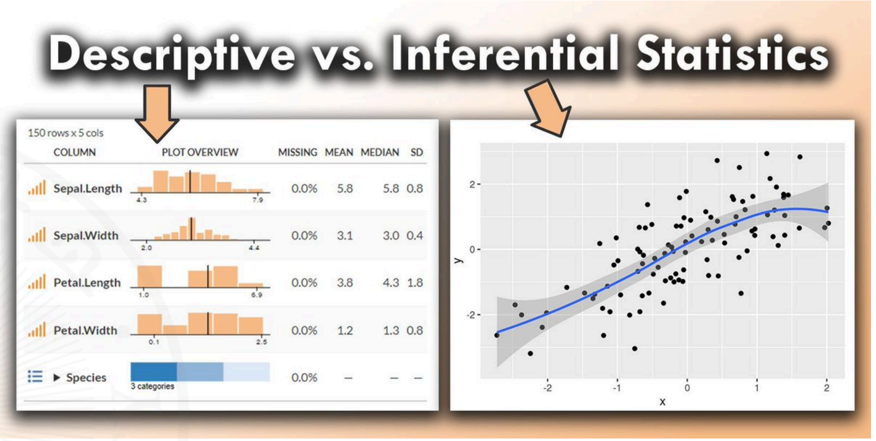
It is not just "calculating averages."



# Two Branches of Statistics

## 1. Descriptive Statistics

- Summarize data
- Means, medians, standard deviations
- Tables and graphs




## 2. Inferential Statistics

- Draw conclusions about a population
- Hypothesis testing
- Confidence intervals
- Regression models

This course covers both.

*The key takeaway*: Descriptive statistics help you describe your current data, while inferential statistics allow you to make predictions and informed decisions based on your data.

# 2. Data in International Commerce

# What Is Data?

Data = recorded information about variables.

Example variables in international commerce:

- Export volume
- GDP
- Exchange rate
- Inflation
- Tariff rate
- FDI inflows


A variable is a measurable characteristic.

# Types of Variables

## 1. Quantitative (Numerical)

- GDP
- Exchange rate
- Revenue
- Inflation rate

Can be:

- Discrete (number of trade agreements)
- Continuous (exchange rate)

## 2. Qualitative (Categorical)

- Country
- Region
- Trade agreement type
- Industry sector

# Cross-Section vs Time Series vs Panel

## Cross-section

Different countries at one point in time.

Example: GDP of 50 countries in 2024.

## Time series

One country over time.

Example: Korea's exports 2000–2025.

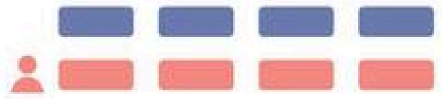
## Panel data

Multiple countries over time.

Example: Exports of 30 countries over 20 years.

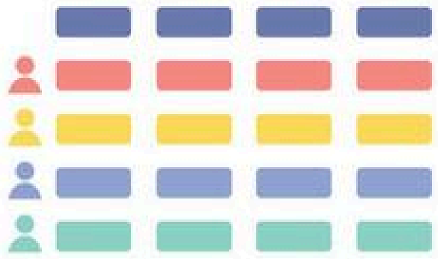
We will work with all three.

# Sequence Analytics



Time series data

Cross-Sectional data



Panel Data
(Longitudinal Data)

# 3. Population vs Sample

# Population vs Sample

## Population

The entire group we care about.

Examples:

- All exporting firms in Korea
- All international trade flows globally
- All consumers in a country

Usually impossible to observe fully.

## Sample

A subset of the population.

We use samples because:

- Data collection is costly
- Full data may not exist
- Time constraints

Inferential statistics helps us generalize from sample to population.

# Why Sampling Matters

Bad sample → biased conclusions.

Example: If you only survey large exporters, you cannot infer about small firms.

Statistical thinking begins with: “How was this data collected?”

# 4. Introduction to Python (Google Colab)

# Today's Goal

By the end of today, you should:

- Open Google Colab
- Run Python code
- Load a dataset
- Compute simple statistics
- Create a basic visualization

No prior coding required.

# Structure of a Notebook

In Colab:

- Text cells (explanations)
- Code cells (Python commands)
- Output (tables, plots)

You execute code cell by cell.

Statistics becomes interactive.

# First Python Example

```python
# Import pandas library to handle data
import pandas as pd

# Create simple data
exports = [100, 120, 130, 150, 170]

# Convert to DataFrame for better handling
df = pd.DataFrame({"Exports": exports})

# Display the data
df
```

```
##    Exports
## 0      100
## 1      120
## 2      130
## 3      150
## 4      170
```

# Descriptive Statistics in Python

```
# Show summary statistics
df.describe()
```

```
##          Exports
## count    5.000000
## mean   134.000000
## std     27.018512
## min    100.000000
## 25%    120.000000
## 50%    130.000000
## 75%    150.000000
## max    170.000000
```

This returns:

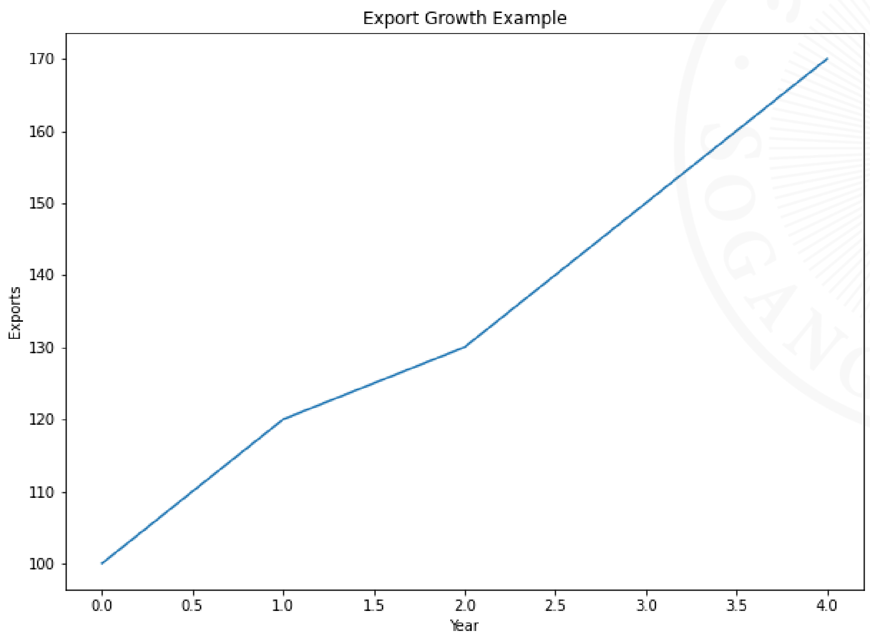
- Mean
- Standard deviation
- Minimum
- Maximum

This is descriptive statistics.

# Simple Visualization

```
# Import matplotlib for plotting
import matplotlib.pyplot as plt

plt.plot(exports) # simple line plot
plt.title("Export Growth Example") # add title
plt.xlabel("Year") # add x-axis label
plt.ylabel("Exports") # add y-axis label
plt.show() # display the plot
```




Visualization helps us see patterns immediately.

# 5. Real-World Example

## International Trade Data

# Example Dataset: Bilateral Trade (Simplified)

Suppose we observe export values (in billion USD) for 2023:

| Country | Exports (bn USD) |
|---|---|
| Korea | 632 |
| Germany | 1650 |
| Japan | 747 |
| United States | 2064 |
| Vietnam | 355 |

This resembles WTO-style aggregated trade data.

Question: Who is “big” depends on what?

# Step 1: Load Data in Python

```python
import pandas as pd

data = {
    "Country": ["Korea", "Germany", "Japan", "United States", "Vietnam"],
    "Exports_bn_USD": [632, 1650, 747, 2064, 355]
}

df = pd.DataFrame(data)

df
```

```
##          Country  Exports_bn_USD
## 0          Korea             632
## 1        Germany            1650
## 2          Japan             747
## 3  United States            2064
## 4        Vietnam             355
```

Now we have a small cross-sectional trade dataset.

# Step 2: Descriptive Statistics

```
df["Exports_bn_USD"].describe()
```

```
## count       5.000000
## mean     1089.600000
## std       729.711107
## min       355.000000
## 25%       632.000000
## 50%       747.000000
## 75%      1650.000000
## max      2064.000000
## Name: Exports_bn_USD, dtype: float64
```

Interpret:

- Mean export level
- Minimum exporter
- Maximum exporter
- Spread of values

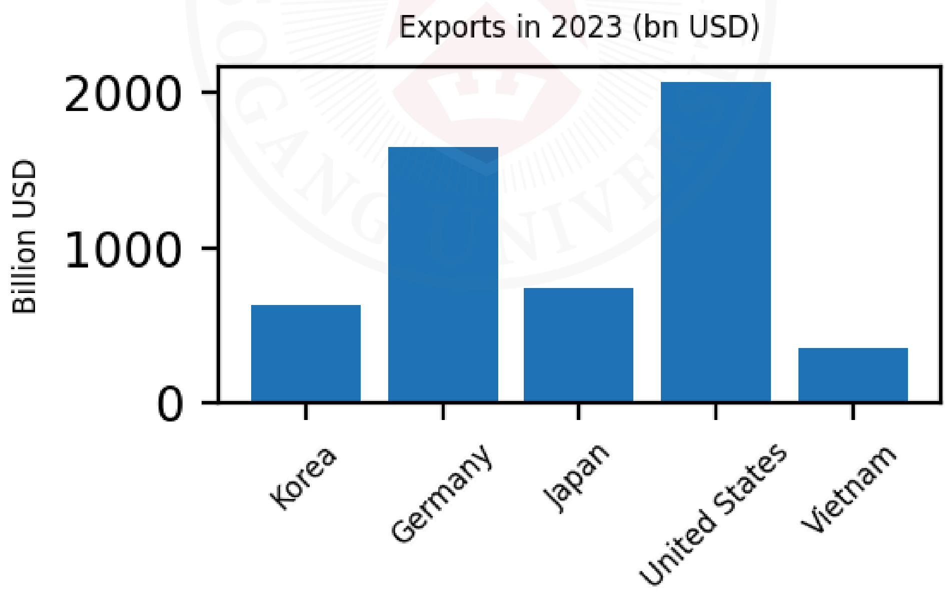
Question: Is the mean representative here?

# Step 3: Visualization -> Large dispersion | One dominant exporter | Skewness in trade distribution

```
import matplotlib.pyplot as plt
plt.bar(df["Country"], df["Exports_bn_USD"])
plt.title("Exports in 2023 (bn USD)", fontsize=6)
plt.ylabel("Billion USD", fontsize=6)
plt.xticks(rotation=45, fontsize=6)
```

```
## ([0, 1, 2, 3, 4], [Text(0, 0, 'Korea'), Text(1, 0, 'Germany'), Text(2, 0, 'Japan'), Text(3, 0, 'United States'), Text(4, 0, 'Vietnam')])
```

```
plt.tight_layout()
plt.show()
```



# Statistical Thinking Begins Here

# Discussion Questions

1. Does a higher export value mean a stronger economy?
2. Should we compare absolute exports or exports per capita?
3. What if we adjust for GDP?
4. Are these differences statistically meaningful?

Statistics helps refine naive comparisons.

# Extending the Dataset (Time Series)

Suppose Korea's exports over 5 years:

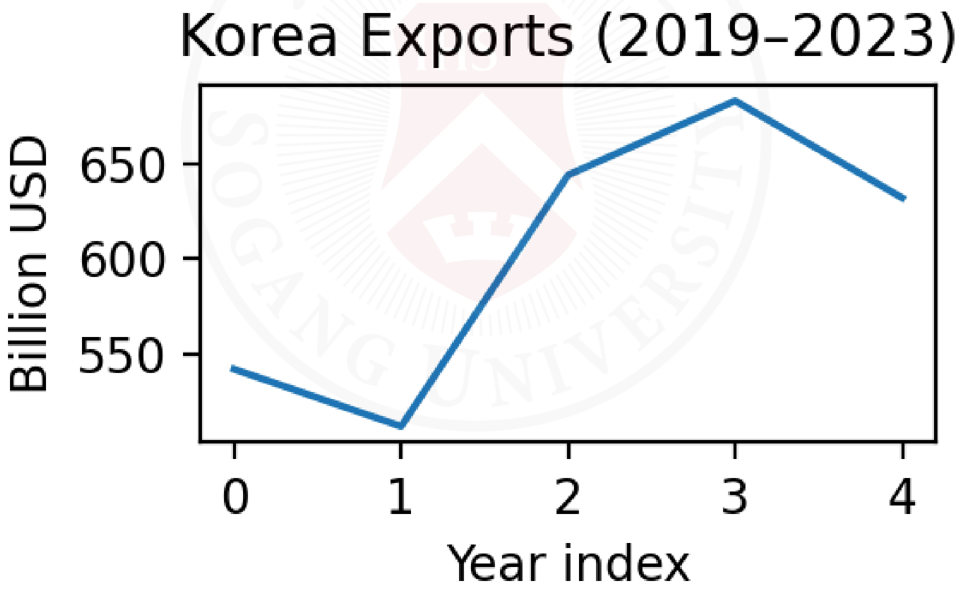
| Year | Exports (bn USD) |
|---|---|
| 2019 | 542 |
| 2020 | 512 |
| 2021 | 644 |
| 2022 | 683 |
| 2023 | 632 |

Now we have time series data.

# Time Series in Python

```
korea_exports = [542, 512, 644, 683, 632]

plt.plot(korea_exports)
plt.title("Korea Exports (2019–2023)")
plt.xlabel("Year index")
plt.ylabel("Billion USD")
plt.tight_layout()
plt.show()
```




Questions: Was 2020 an outlier? | Is there an upward trend? | How volatile are exports?

# From Description to Inference

# Next Logical Questions

- Is export growth statistically significant?
- What determines export performance?
- Does exchange rate depreciation increase exports?
- Does inflation reduce competitiveness?

These require inferential statistics and regression.

We will build toward this.

# Big Picture

Even a simple 5-row dataset:

- Allows comparison
- Reveals distribution patterns
- Raises policy questions
- Shows why structure matters

Statistics transforms numbers into economic insight.

# 6. Hands-On Mini Exercise

# Exercise (5 minutes)

1. Create a list of 5 numbers representing:
    - Exchange rate values OR
    - Trade volumes
2. Compute:
    - Mean
    - Maximum
    - Minimum
3. Plot the values.

Goal: Become comfortable running code.

# Why This Matters

In international commerce:

- You will analyze real data.
- You will interpret coefficients.
- You will evaluate uncertainty.

Python is not the goal. It is the tool.

Statistics is the framework.

# Key Takeaways Today

- Statistics helps us reason under uncertainty.
- Data comes in different structures.
- Samples allow inference about populations.
- Python allows us to implement statistical ideas.

# Week 1 overview (where we are today)

Meeting 2 (today):

- Introduction to statistics
- Introduction to Python or R, datasets, and ethics
- Reading: LMW Chapter 1

# Week 2 overview (next week)

- Descriptive statistics (LMW Chapter 3)

Any questions?

Thank you!