Statistical Analysis of Video Game Ratings and Sales

Team Members: Stadnik Oleksandr, Hutsuliak Oleh, Zaiats Yulian

Project Goal: this project investigates patterns in global video game sales using a real-world dataset to test hypotheses about regional performance and market trends.

Dataset: we use the Video Game Sales dataset (sourced from Kaggle), which contains sales data for 16,598 video games. Variables include game name, platform, release year, genre, publisher, and sales figures across North America (NA), Europe (EU), Japan (JP), and globally.

Preliminary Findings:
- Highly Skewed Sales: most games sell under 200,000 units; a few blockbusters like Wii Sports sell over 80 million.
- Nintendo Dominance: all top 15 best-selling games are Nintendo titles.
- Regional Trends: NA and EU sales are strongly correlated ($r \sim 0.77$), with NA generally leading in sales.
- Distribution Shape: global sales data is non-normal, with high skewness and kurtosis, making non-parametric tests appropriate.

Planned Hypotheses:
1. Correlation Hypothesis: NA and EU sales are positively correlated — popular games in NA tend to also perform well in EU.
2. Sales Dominance Hypothesis: games typically sell significantly more units in North America than in Europe.

We will test these using correlation analysis and linear regression where appropriate. Further analysis may explore genre or publisher effects if time allows.