Final report
What Makes a Video Game Successful?

Ioannidis Efstratios (tem2505) group41

Hypothesis Test: Do Higher Critic Scores Predict Higher Sales?

We split games at the median `Critic_Score` into two groups: **High** ($n = 3865$, $\mu = 1.081$, $\sigma = 2.454$) and **Low** ($n = 4272$, $\mu = 0.334$, $\sigma = 0.759$). Welch's one-tailed t-test gave $t = 18.166$, $p < 0.0001$, with a 95% CI for the difference ($\mu_{\text{high}} - \mu_{\text{low}}$) of (0.667, 0.828) million units.

Conclusion: We reject H_0 and find strong evidence that games with higher critic scores sell significantly more—on average, about 0.75 million units more.

Confidence Intervals for Mean Sales by Genre

Using each genre's `Global_Sales` data (sample size n, sample mean, and standard deviation), we computed 95% confidence intervals using the t-distribution to estimate average sales. **Platform** games showed the highest mean sales (approximately 0.93 ± 0.17 million copies), while **Adventure** ($\approx 0.18 \pm 0.03$ M) and **Strategy** ($\approx 0.26 \pm 0.04$ M) had the lowest.
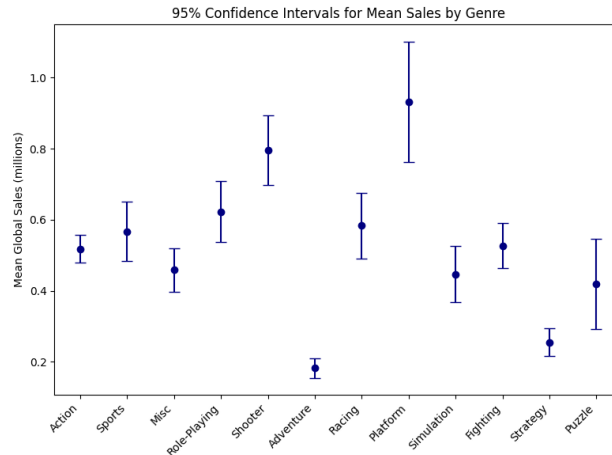
95% confidence intervals for mean global sales by genre.

ANOVA: Mean Sales by Genre and Platform

To test whether average global sales differ by genre or platform, we conducted two one-way ANOVAs:

- **By Genre:** $F(11, 11205) = 18.747$, $p < 0.0001$

- **By Top 10 Platforms:** $F(9, N{-}10) = 21.758$, $p < 0.0001$

Since both tests reject the null hypothesis of equal means ($p < 0.0001$), we conclude that game **genre** and **platform** significantly affect global sales. Publishers aiming to maximize revenue should prioritize higher-performing genres (e.g., *Platform*, *Shooter*) and proven platforms with strong user bases.

Multiple Linear Regression: Predicting Global Sales

We fit an OLS model to predict `Global_Sales` using:

- **Genre** (baseline = Action)
- **Platform** (baseline = X360)
- **ESRB_Rating** (baseline = E)
- **Critic_Score** and **User_Score** (continuous)

The intercept reflects sales for an Action game on X360, rated E, with zero critic/user scores. Coefficients represent deviations from this baseline. Positive values indicate higher sales, negative values indicate lower.
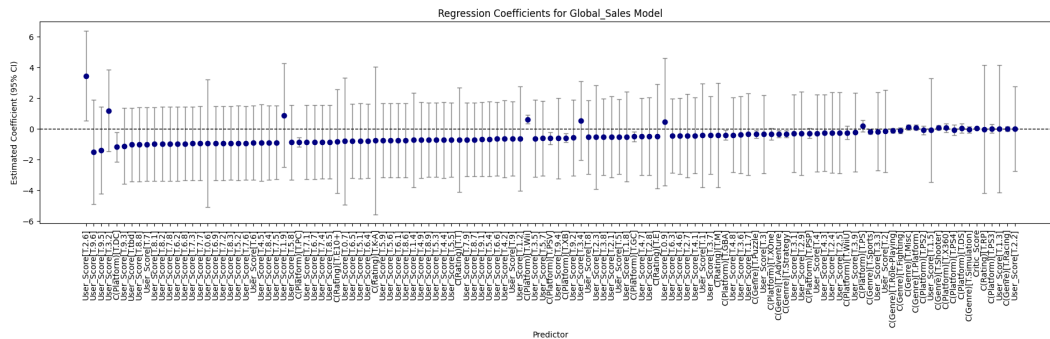
Regression coefficients with 95% confidence intervals. Points above zero increase sales vs. the baseline, points bellow zero reduce sales.

Conclusion

- `Critic_Score` shows a strong positive effect on sales (coef ≈ 0.022, $p < 0.001$), indicating that each one-point increase in critic rating raises average `Global_Sales`.

- Among `genres`, `Adventure` (coef $= -0.330$, $p = 0.002$), `Puzzle` (coef $= -0.349$, $p = 0.008$), and `Sports` (coef $= -0.193$, $p = 0.012$) sell significantly less than the `Action` baseline. In contrast, `Role-Playing` (coef $= -0.124$, $p = 0.111$), `Simulation` (coef $= -0.048$, $p = 0.647$), and `Racing` (coef $= 0.013$, $p = 0.880$) do not differ meaningfully.

- For platforms, `Wii` (coef $= 0.813$, $p < 0.001$), `GB` (coef $= 0.700$, $p < 0.001$), and `PS` (coef $= 0.519$, $p < 0.001$) have the strongest positive effects relative to the `NES` baseline, while `PC` (coef $= -0.060$, $p = 0.562$) and `XB` (coef $= -0.059$, $p = 0.636$) show no significant difference.

In short: higher critic ratings and platform choice (e.g., `Wii`, `GB`, `PS`) strongly influence sales, as do certain genres (`Adventure`, `Puzzle`, `Sports` underperform), while others (`Simulation`, `Racing`) and platforms (`PC`, `XB`) have minimal impact.

Logistic Regression: Predicting "Successful" Games

We trained a logistic regression model to classify games as "Successful" ($Global\_Sales > 1M$) using Genre, Platform, ESRB Rating, Critic_Score, and User_Score. After cleaning and a 75/25 train-test split:

- **Test Accuracy:** 83.2%
- **Class 0 ("Not Successful"):** Precision = 85.0%, Recall = 96.0%
- **Class 1 ("Successful"):** Precision = 66.5%, Recall = 32.1%
- **ROC AUC:** 0.813

The model is highly effective at identifying low-selling games (96.0% recall), and fairly reliable when predicting success (66.5% precision), but it misses most high-sellers (32.1% recall). Overall accuracy is high due to the dominance of the "Not Successful" class. To better capture successful titles, strategies like class balancing or more advanced classifiers may be needed.

References

Shams, Rushdi. *Video Game Sales with Ratings.* `https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings`, 2018.

Triantafillou, Sofia. *MEM 264 – Applied Statistics.* `https://elearn.uoc.gr/course/view.php?id=5842#section-0`, 2025.