

# GLM - Gama Regressing Video Game Sales

Vid Stropnik, 63200434

## I. Introduction

We are given a **dataset**, quantifying the sales success of video games. In it, we are also provided the platforms the sales were logged on and their genre categories. In this report, we analyze the effect of these two categories on sales success using gamma regression. Through it, we aim to answer the question: *On which platform should we publish our game and what genre should it be to maximize our sales.* All work is reproducible and made available on **GitHub**.

## II. Data & Model

Since sales will always be a positive real value, a sensible candidate for modelling them is *gamma regression*. Indeed, when looking at the sales distributions, visualized in Fig. 1 on a logarithmic scale, we can see that they approximately follow a Gamma distribution. From the figure, we can also see that console-shooters seem to be performing best, while RPG titles are most successful on Xbox. If we complement these insights with those gathered from Fig. 2, we can see that PS's performance is skewed to the positive by a collection of successful titles. Notice also, an under-representation of certain categories, e.g. as XBox RPGs. PC also has over twice the number of titles when compared to the other platforms. This opens several questions about the sampling procedure, which exceeds the scope of this text, but is crucial to its results.

We will answer the question of what genre of this problem with independent gamma models of the form, calculated in **lectures**:

$$\text{profit} \sim \text{gamma}\left(\frac{\mu^2}{\phi}, \frac{\mu}{\phi}\right) \quad (1)$$

where  $\phi$  is the variance and  $\mu = e^{\beta_0 + \beta_1 x_1 \dots}$  is our linear model, plugged into the gamma-corresponding link function. The features in the model correspond to one-hot encodings of each dimension. Thus, the *platforms* model uses two features, while *genres* uses just one. Both models use an intercept  $\beta_0$ . In modelling, we impose *Cauchy(0,2)* priors on all the weights, the intercept and the variance  $\phi$ .

## III. Results & Conclusion

The posterior distributions are shown in Fig. 3. As hypothesized, the weights in both models show that developing a Shooter on Xbox is the most profitable choice. From these insights, we can conclude this brief research with three statements, our confidence in which is given in Table I.

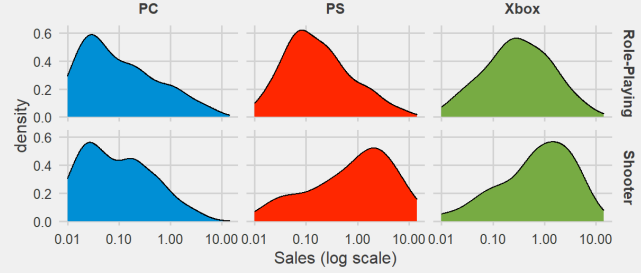


Fig. 1. The skewedness of log-scale sales distributions between categories gives a good initial indication.

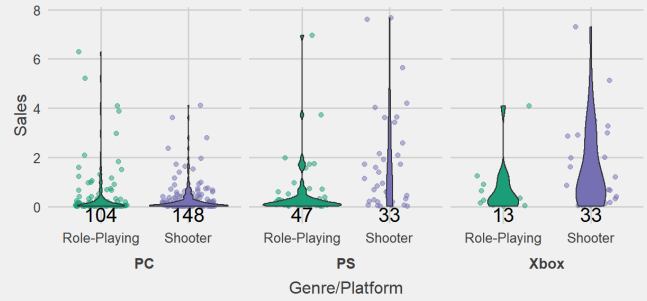


Fig. 2. Observed prior distribution of sales. The numbers below the individual violin plots, as well as each plots width, correspond to the number of observations in the given category.

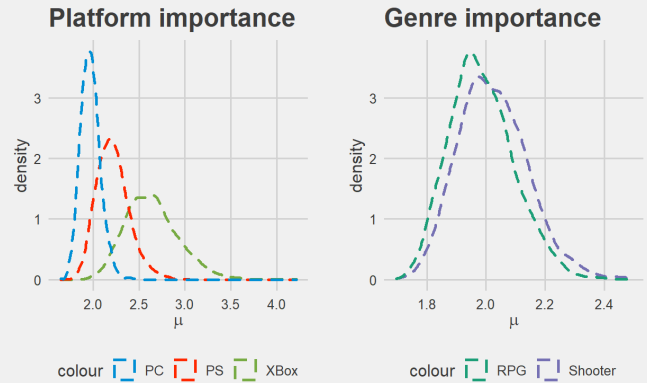


Fig. 3. Posterior distributions of means  $\mu$  from our gamma regression indicate a preferential treatment of Xbox and Shooters.

TABLE I  
Probabilistic confidence in findings.

Statement	Certainty
"Xbox is the most profitable platform"	<b>0.971</b> $\pm$ 0.004
"Shooters are more profitable than RPGs"	<b>0.690</b> $\pm$ 0.011
"Shooters on Xbox are the most profitable"	<b>0.671</b> $\pm$ 0.011