

Video Game Sales Data

The data set is a subset of the enriched video games sales data from Kaggle. It contains the sales and attributes, including critic's review score, for video games on Sony's Playstation 2 and Microsoft's XBox 360. The video games included in this data set were randomly selected but are specifically rated as either E (everyone) or T (teen) and of the Action and Sports video game genres. North American (NA_sales) and Global sales are measured in millions.

```
data vgsales;  
  infile 'filelocation/vgsalessubs.csv' firstobs=2  
    dsd ;  
  input Genre $ Rating $ Platform $ Name $  
    Year_of_Release Publisher $ NA_Sales  
    Global_Sales Critic_Score;  
run;
```

Video Game Sales Data

What questions might a data analyst have about this dataset?

What might be their goals for analysis?

Example: Video Game Sales Data

Let's do produce some visualizations and numerical summaries

Exercise: Video Game Data

1. Make a scatter plot of numeric variables of interest to you. Describe what you see.
2. Make a histogram of those same variables from (part 1). Is either distribution normal?
3. What do the 4 tests for normality determine for each variable?
4. Test for the population correlations of those variables (from part 1) by platform separately. What conclusions can we draw?
5. Create a log transform of the NA sales variable. Use $\ln\text{NAsales} = \log(\text{NA_sales})$. Was there clear skewness before or after the log transformation?
6. Perform the appropriate test for the mean difference in NA_sales for the two ratings.