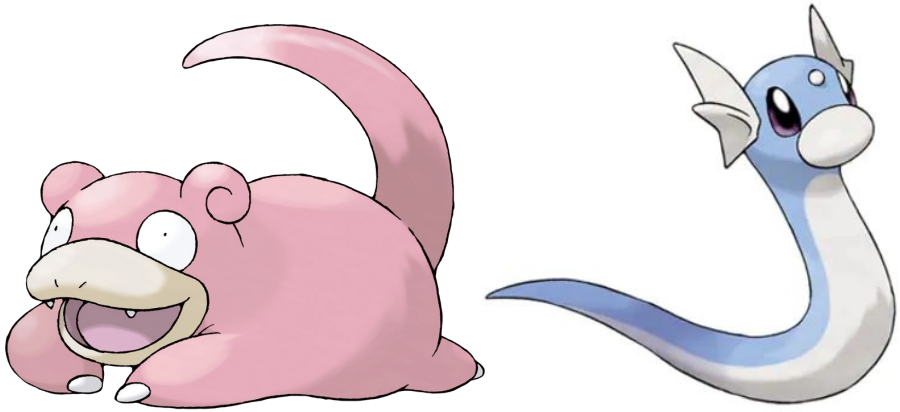
An Exploration of Video Game Sales

Team Nugget

Amanda, Ben & viv

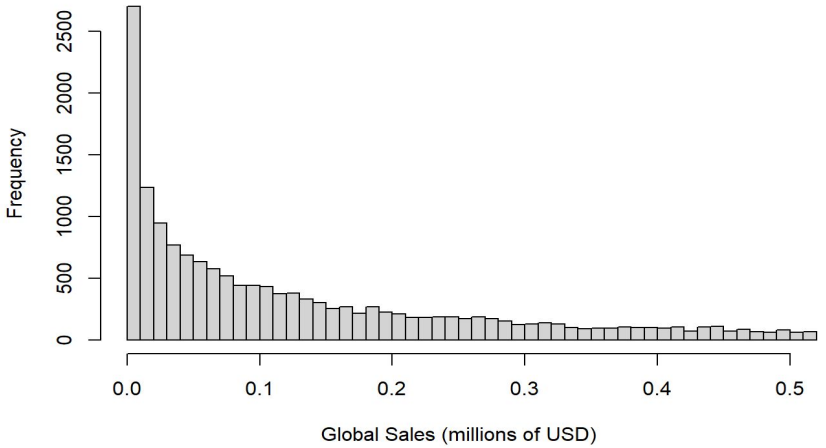


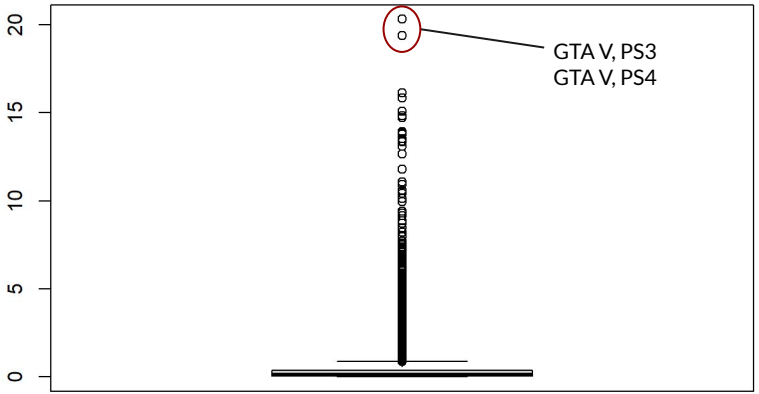
Today's topics:

1. Descriptive analysis
   a. variable distributions
   b. notable correlations
2. 1-sample stats inference
3. CI and H-test for action games
4. Sports vs Shooter sales
5. Bootstrap rating ~ sales

Response variable of interest: Global video game sales

- Extreme right skew
- Outliers present





Top Selling Video Games

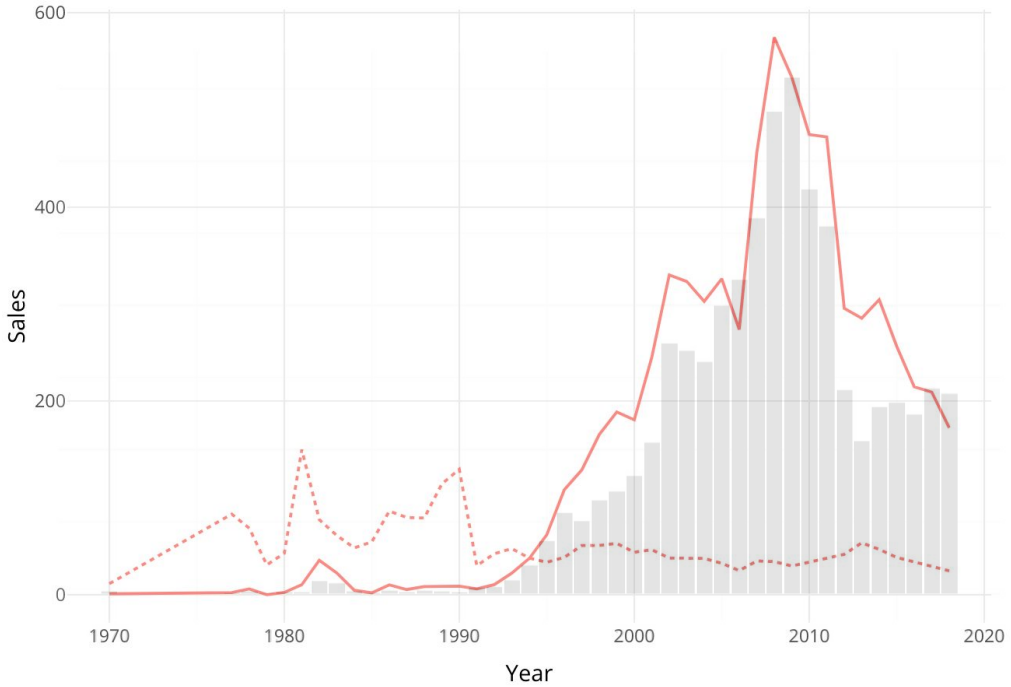
Name	Platform	Genre	Sales	Year
Grand Theft Auto V	PS3	Action	20.32	2013
Grand Theft Auto V	PS4	Action	19.39	2014
Grand Theft Auto: Vice City	PS2	Action	16.15	2002
Grand Theft Auto V	X360	Action	15.86	2013
Call of Duty: Black Ops 3	PS4	Shooter	15.09	2015
Call of Duty: Modern Warfare 3	X360	Shooter	14.82	2011
Call of Duty: Black Ops	X360	Shooter	14.74	2010
Red Dead Redemption 2	PS4	Action-Adventure	13.94	2018
Call of Duty: Black Ops II	X360	Shooter	13.86	2012
Call of Duty: Black Ops II	PS3	Shooter	13.80	2012

Sales by year

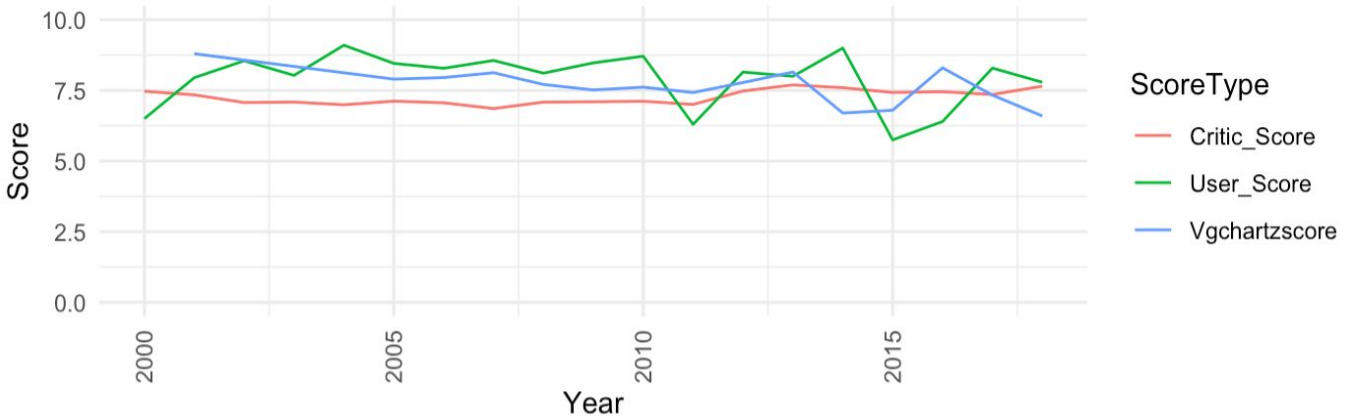
Total Sales in 10^6 USD

Mean Sales per Game in 10^4 USD

Relative number of games released



Video game average ratings over time

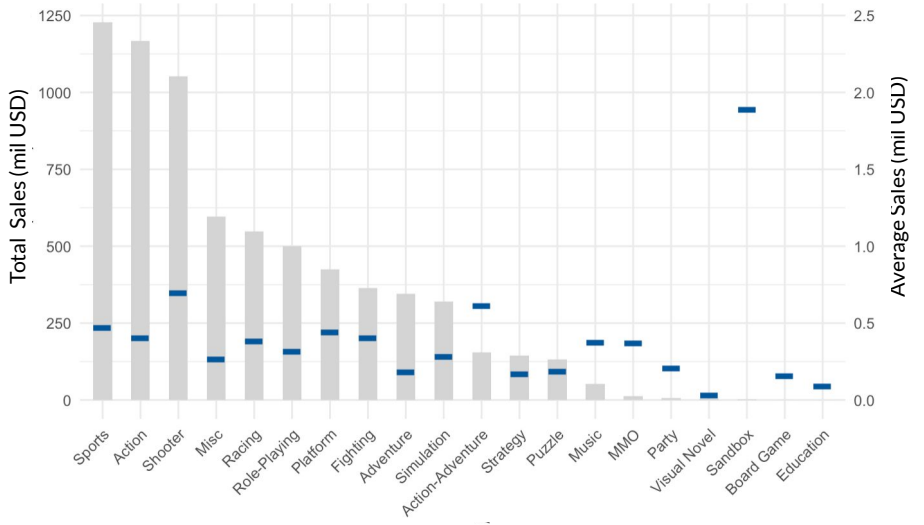


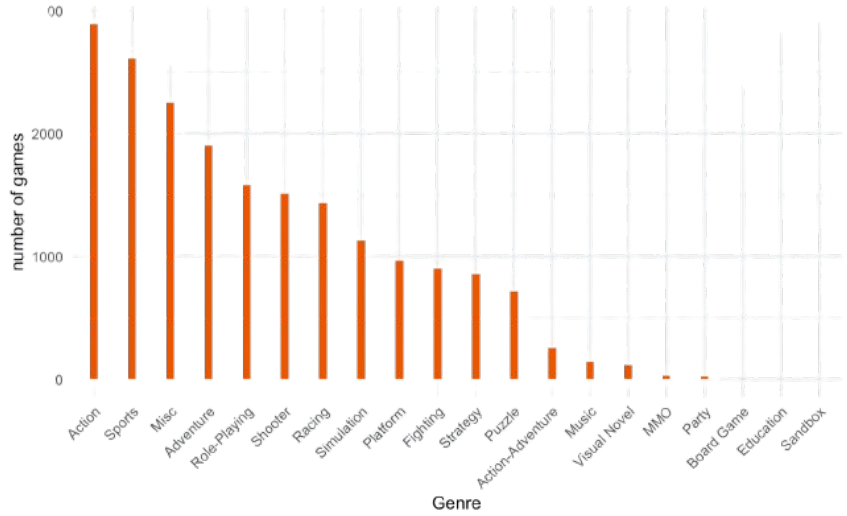
Video Game Genres

Total Sales per genre

Mean Sales per genre

Number of games per genre





1-Sample Stats Inference for Sales

Full sample mean: 0.37 mil USD
95% CI: [0.35, 0.38] mil USD

Check assumptions:

1. The sample is not randomized (database does not include all games and would have a bias towards including games that are available in English)
2. The population sales distribution is not normal at all (extreme right skew)
3. The dataset had two extreme outliers identified via boxplot, but they were removed.

Action Video Games: Sales CI and h-test with all data

H_0: $\mu_{\text{Action video games}} = 0.367$

H_a: $\mu_{\text{Action video games}} \neq 0.367$

Assumptions:

1) Data is right-skewed- not normalized, but we use a large sample.
2) Data is not randomized.
3) Quantitative response variable is Global Sales.
4) Outliers are included in this first test.

We are 95% confident that, on average, the mean action global video game sales fall in **[0.378, 0.455]** so we reject the null hypothesis. (Sales measured in millions USD)

Our results are statistically significant with a p value of **0.012**.

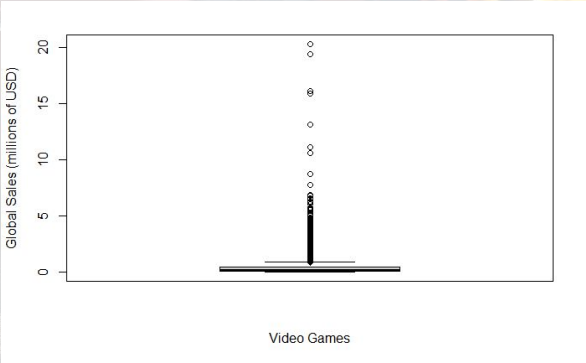
However, our results are not practically significant with a small Cohen's D value of **0.052**, and the population mean is only weakly removed from the CI.

Action Video Games: Sales CI and h-test after removing Outliers

H_0: $\mu_{\text{Action video games}} = 0.342$

H_a: $\mu_{\text{Action video games}} \neq 0.342$

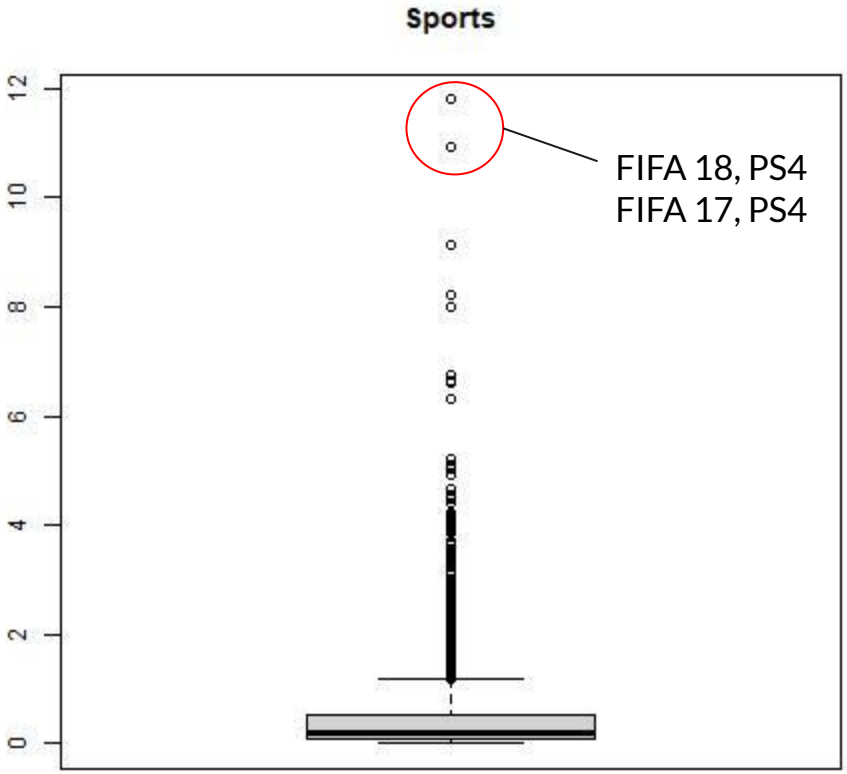
Name <chr>
Grand Theft Auto V
Grand Theft Auto V
Grand Theft Auto: Vice City
Grand Theft Auto V
Grand Theft Auto III
Grand Theft Auto IV
Grand Theft Auto IV
Grand Theft Auto V
Grand Theft Auto: Liberty City Stories

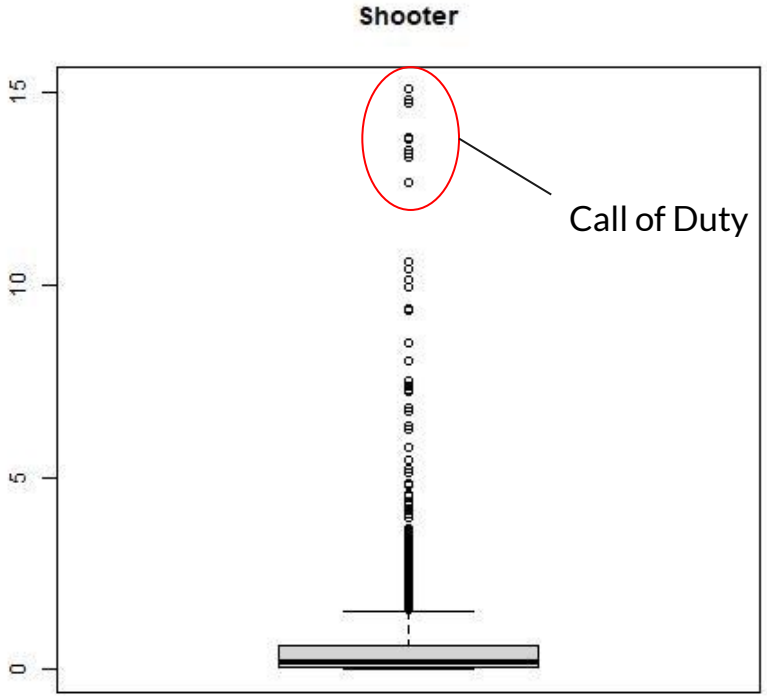


We are 95% confident that, on average, the mean action global video game sales fall in **[0.35, 0.402]** so we reject the null hypothesis. (Sales measured in millions USD)

Our results are statistically significant with a p value of **0.013**.

However, our results are not practically significant with a small Cohen's D value of **0.049**, and the population mean is only weakly removed from the CI.


Sports
0
2
4
6
8
10
12
FIFA 18, PS4
FIFA 17, PS4


Shooter
0
5
10
15
Call of Duty

2 Sample T-Test: Sports and Shooter Sales

H_0: $\mu_{\text{Sports Sales}} - \mu_{\text{Shooter Sales}} = 0$

H_a: $\mu_{\text{Sports Sales}} - \mu_{\text{Shooter Sales}} \neq 0$

p-Value: 1.236e-06

95% CI: [-0.193, -0.082]

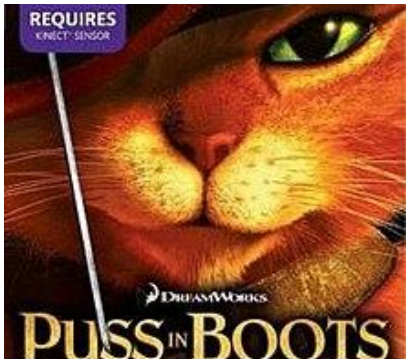
Cohen's D: 0.177

Conclusion: Although there is a significant difference of the means, in practicality they remain relatively close.

Assumptions:

1) Data is not normal, but we do have a large sample.
2) Data is not random.
3) These are a quantitative response variable for two groups which is Sales.

BOOTSTRAPPING BANG



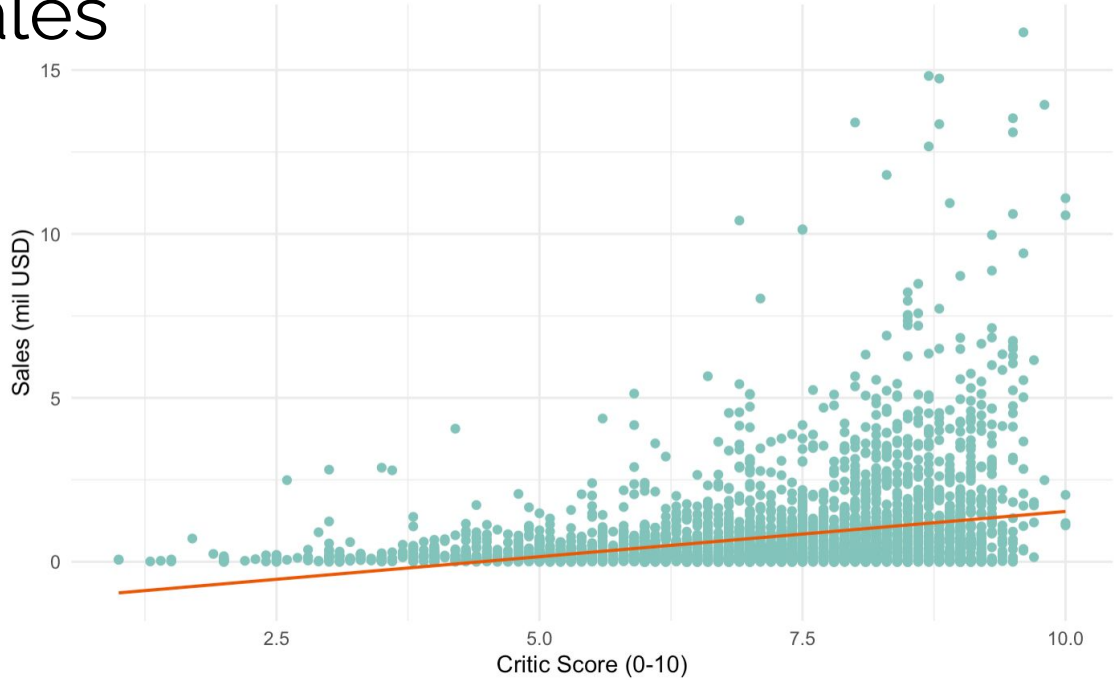
CI for Population Correlation using Bootstrap:

Critic Scores and Sales

Full sample correlation mean: 0.30

Bootstrap sample bias: 0.0005

95% CI: [0.28 , 0.32]



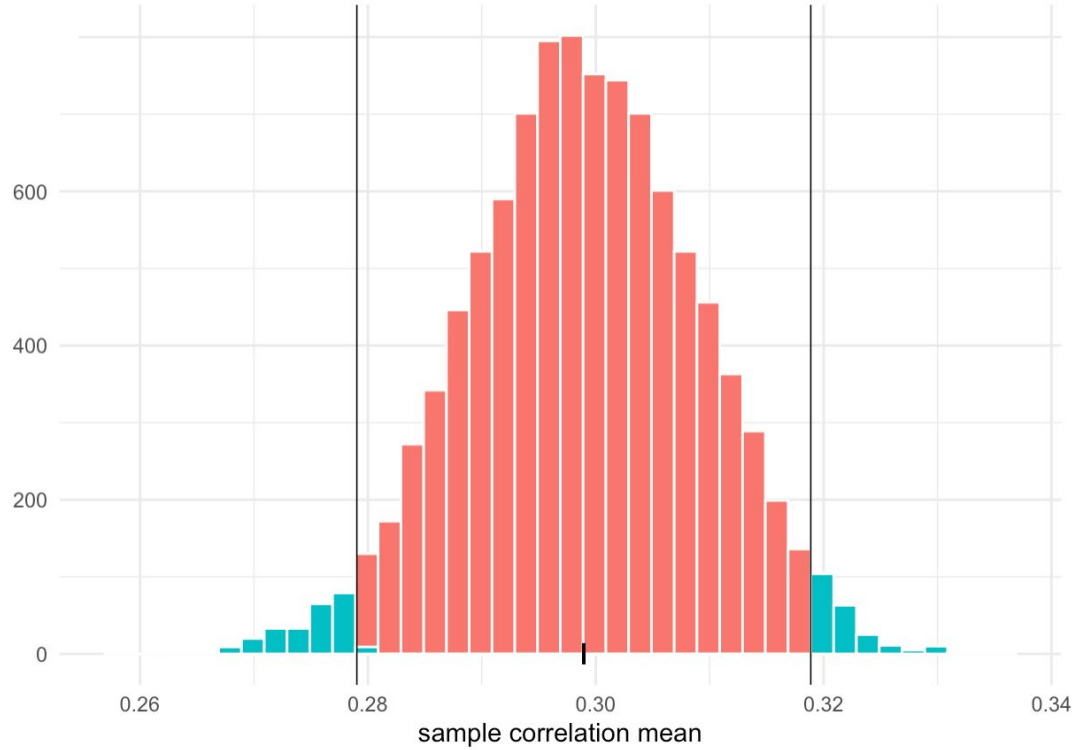
CI for Population Correlation using Bootstrap:

Critic Scores and

Full sample correlation mean: 0.30

Bootstrap sample bias: 0.0005

95% CI: [0.28 , 0.32]



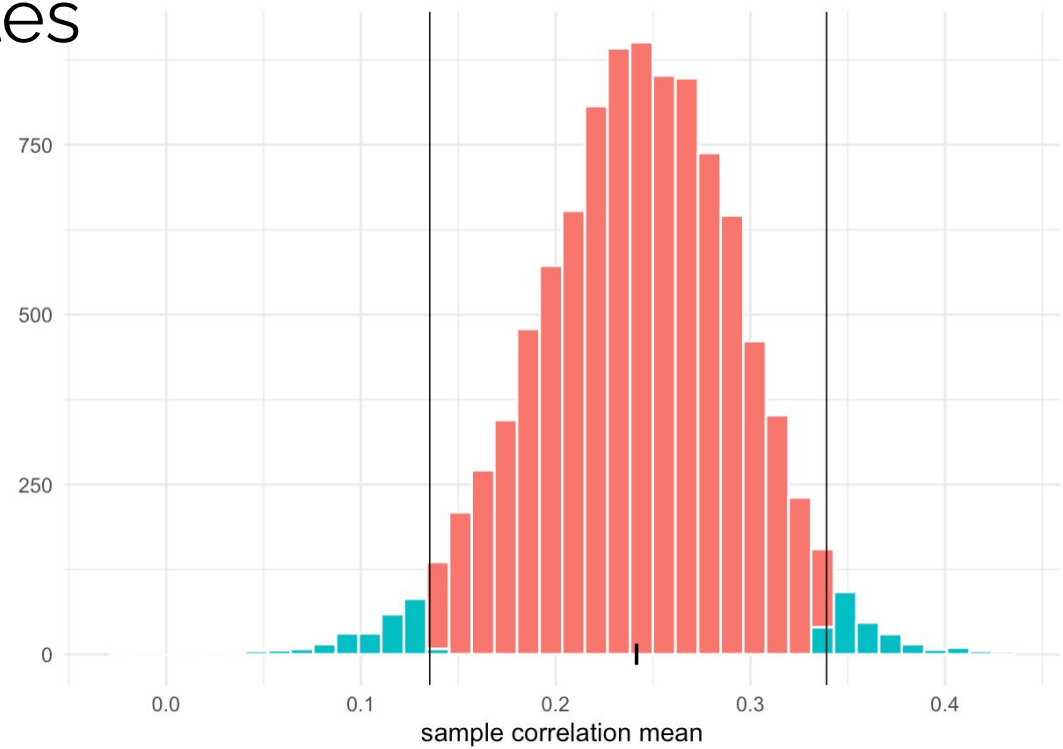
CI for Population Correlation using Bootstrap:

User Ratings and Sales

Full sample correlation mean: 0.24

Bootstrap sample bias: 0.002

95% CI: [0.13 , 0.34]



GGWP

QUESTION TIME

Name	Global_Sales
Metal Gear Solid 2: Sons of Liberty	6.05
Yakuza Zero	0.65
Stardew Valley	0.13