DataFest 2022

Team 2- Abnormal Distributions

Abigail Bridgeman <u>bridgeman.abigail@gmail.com</u> Jake Englebrecht <u>joe3zf@mail.missouri.edu</u> Rodney McFall <u>rdm3n3@mail.missouri.edu</u>

Data Analysis





	Df	Sum Sq	Mean Sq	F value	Pr(>F)
age	3	1.791e+08	59698084	0.536	0.6587
gender	1	6.238e+07	62377509	0.560	0.4560
race	2	2.845e+08	142226781	1.277	0.2835
school 👘	11	2.004e+09	182175382	1.636	0.1009
wave	15	3.324e+09	221568575	1.990	0.0238 *
Residuals	95	1.058e+10	111350907		
signif. coo	les :	0 '***'	0.001 '**'	0.01 **	0.05 '.' 0.1 ' ' 1

Data Displays



Source: Yale School of Medicine



- Hispanic demographic heavily skewed towards the right for all age groups
- Highest density in ¾ age groups
- Bimodal distribution
- No caucasian 11 y.o.

Density plot

Time in game grouped by gender and race





- Hispanic males playing time
- African American males are closest to a normal distribution
- Female density of play was higher than male density

Box plot





Source: Yale School of Medicine



- ▷ School 647
- ▷ School 1531
- ▷ Wide range in IQR
- ▷ p-value = 0.10

Box plot Time played grouped by Wave of Participant



Source: Yale School of Medicine



- ⊳ p = .02
- The most significant factor affecting time played
- Teacher familiarity and student outcomes

Looking to the Future

Greater Information on School

- Socio-Economic Class
 - Effect of greater school funding on preliminary education
- Greater Racial Representation
 - Caucasian, African American or Hispanic
 - Asian, American Indian, Multi-Racial etc.

▷ Waves

• Teachers on student outcomes

Final Thoughts