

# Baseball data analysis

## Are the umps biased?

Wenxin Song

Institute of Statistics and Big Data  
*Renmin University of China*

April 20, 2023

# Presentation Overview

## ① Question

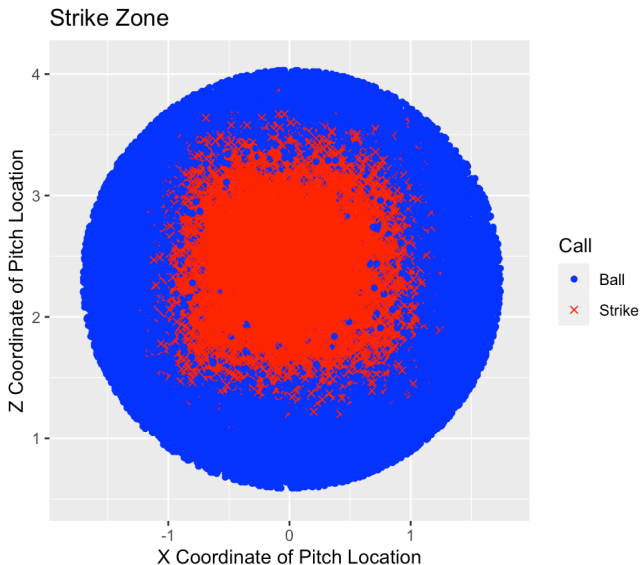
## ② Biased?

Population and Sample  
Regression

# Question

- Are the ump's judgments on strikes and balls biased according to the counts?
- Or, is there causal relationship between ump's judgments and the situation like counts and fields?

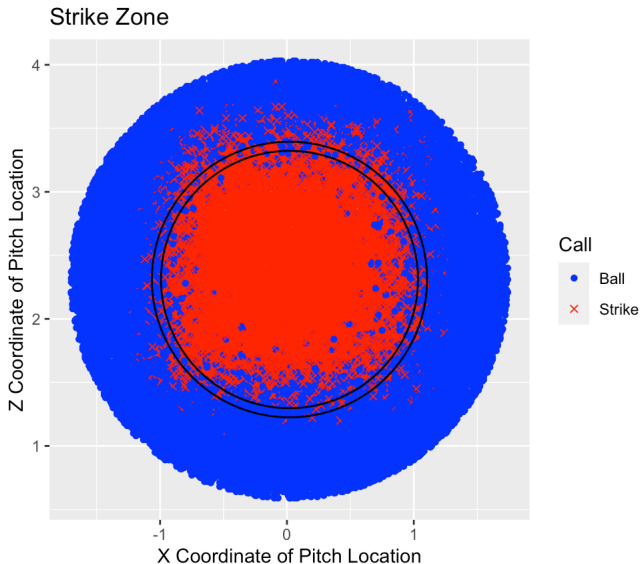
# Population



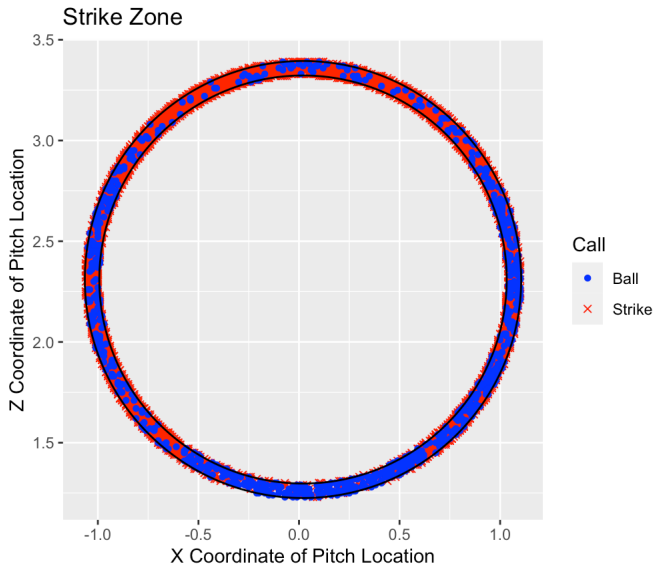
# How to sample

I use Decision tree to find which samples are near the decision boundary of the ump's about the strikes and balls.

# Sample1

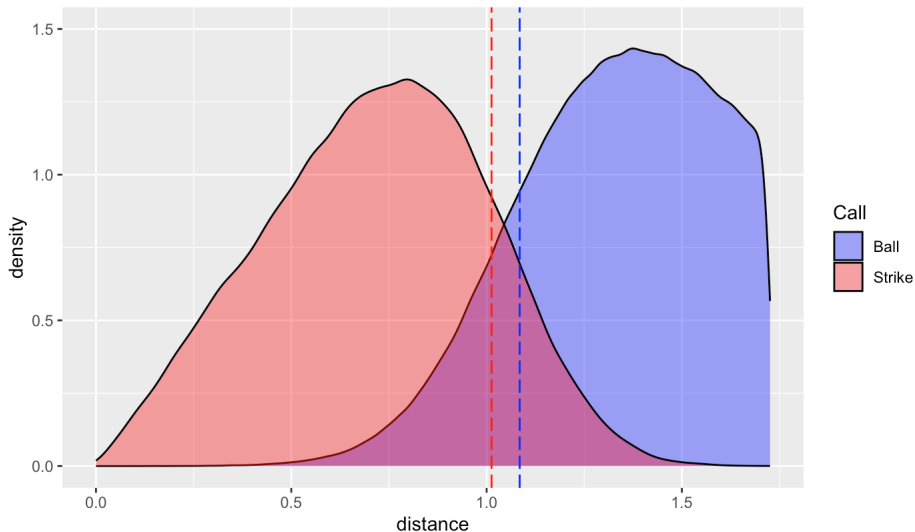


# Sample1



# Sample1

Distributions of Distance from Center of Strike Zone for Balls and Strikes



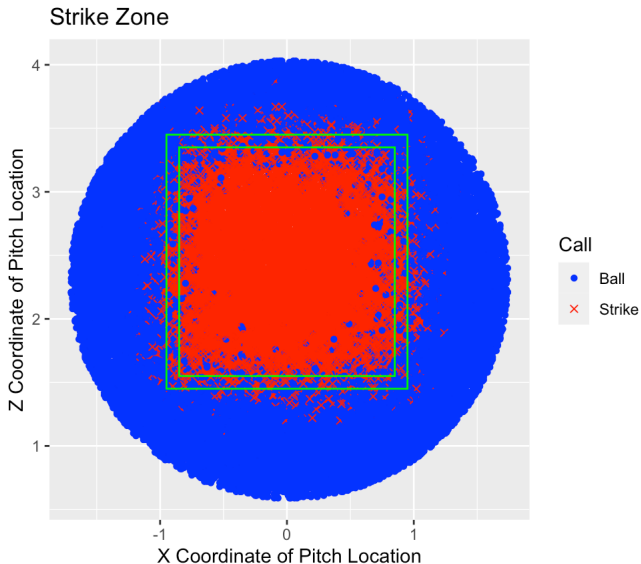


Generalized linear model is used for our modeling.

	Estimate <dbl>	Std. Error <dbl>	z value <dbl>	Pr(> z ) <dbl>	Percent.Change.Odds <chr>
Intercept	-1.63556847	0.02213578	-73.887995	0.000000e+00	-80.52%
full.count	-2.99182745	0.38026380	-7.867768	3.610238e-15	-94.98%
Three1	-0.71066991	0.13654107	-5.204807	1.941985e-07	-50.87%
Three0	0.93506426	0.06421656	14.561108	4.964663e-48	154.74%
Two.Outs	0.23534302	0.03245214	7.252003	4.106540e-13	26.53%
One.Out	0.09958577	0.03220503	3.092243	1.986504e-03	10.47%

	Estimate <dbl>	Std. Error <dbl>	z value <dbl>	Pr(> z ) <dbl>	Percent.Change.Odds <chr>
Intercept	-4.2518527	0.05633296	-75.477172	0.000000e+00	-98.58%
px	-0.2929301	0.01743532	-16.800966	2.401021e-63	-25.39%
pz	1.1673892	0.02114634	55.205258	0.000000e+00	221.36%
full.count	-2.9228482	0.38198620	-7.651712	1.983216e-14	-94.62%
Three1	-0.7517711	0.14169967	-5.305383	1.124365e-07	-52.85%
Three0	0.9115818	0.06959995	13.097448	3.405142e-39	148.83%
Two.Outs	0.2961497	0.03435800	8.619526	6.723188e-18	34.47%
One.Out	0.1434230	0.03397157	4.221853	2.423024e-05	15.42%

# Sample2



	<b>Estimate</b> <dbl>	<b>Std. Error</b> <dbl>	<b>z value</b> <dbl>	<b>Pr(&gt; z )</b> <dbl>	<b>Percent.Change.Odds</b> <chr>
Intercept	-1.5187081	0.01739230	-87.320715	0.000000e+00	-78.10%
full.count	-4.0701908	0.50117757	-8.121255	4.613876e-16	-98.29%
Three1	-0.8897849	0.11369539	-7.826043	5.034669e-15	-58.93%
Three0	0.8109897	0.05062080	16.020878	9.135169e-58	125.01%
Two.Outs	0.2690355	0.02540124	10.591432	3.265584e-26	30.87%
One.Out	0.1053335	0.02512851	4.191794	2.767569e-05	11.11%

# The End. Thank You!

Questions? Comments?