Update 5: Probability distributions, hypothesis testing, Monte Carlo simulations

Dataset: Heartbeat sounds from <https://www.kaggle.com/kinguistics/heartbeat-sounds>

Goal: To classify unlabeled sounds into murmurs, extrahls, artifacts, or normal heartbeat sounds

Attached R script has been updated with these applications.

**Probability distributions/Hypothesis testing:**

I used Shapiro-Wilk test to see if the characteristics that I have explored in Exploratory Data Analysis would fall under a normal distribution.

For the most part, rowMeans(x), rowMeans(x\*x), min, and max for each category and all labelled points were normally distributed (p-value < 0.05).

For index of min, normal and murmur datasets did not have normally distributed points.

For index of max, normal, murmur and extrahls datasets did not have normally distributed points.

**Hypothesis Testing:**

I used chi-squared tests to see if two characteristics of the data are dependent on each other (supplementing the graphs that I made during Exploratory Data Analysis). For the most part, the results of the chi-squared tests for pairs of characteristics were similar, however, there was a low p-value (~0.08) for artifacts min vs max, which, compared to Update 4, was consistent with the max(x) vs min(x) plot (see below). This observation was also observed in a low p-value for all labelled points.

Machine generated alternative text:
Murmur 
Artifacts 
Normal 
0 Extrahls 
o 
o 
10000 
max(x) vs min(x) 
15000 
apply(murmur, 
o 
20000 
1 , max) 
25000 

For all labelled points, index of min and index of max also had low p-values which is consistent with the graphs from Update 4:

Machine generated alternative text:
index ot min(x) vs index ot max(x) 
100000 150000 200000 250000 
apply(murmur, 1, which min) 
350000 
Murm r 
o Artifa 
Norm 
o Extrah s 
50000 
300000 

**Monte Carlo simulations:**

This is not applicable to my data because there is no "uncertainty"