

# Assignment 1

Alex Radecki

Feb 8<sup>th</sup> 2023

## 1 Statistical Distribution of Attributes

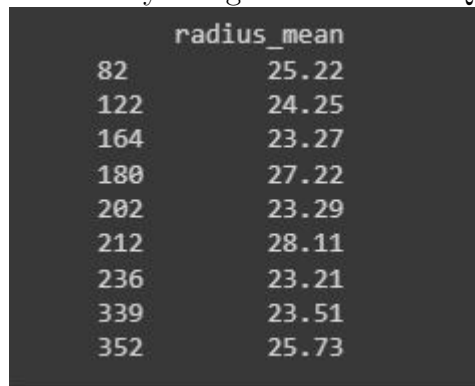
Table that lists the mean, median, Q1 and Q3 for three numeric attributes in the cancer dataset.

Attribute Name	Mean	Median	Q1	Q3
radius_mean	14.22	13.46	11.76	16.13
texture_mean	18.96	18.61	16.02	21.54
fractal_dimension_worst	0.0841	0.0801	0.0713	0.0922

## 2 Outlier Identification

Looking at radius\_mean, we identified outliers by calculating the IQR distance (4.37) and used that to identify the outlier rows in the data. Below is a screenshot of the outliers found by using the IQR distance.

Figure 1: Outliers found by using 1.5 times the IQR from Q1 or Q3



	radius_mean
82	25.22
122	24.25
164	23.27
180	27.22
202	23.29
212	28.11
236	23.21
339	23.51
352	25.73

### 3 Correlations

Below are the 3 scatter plots created from `radius_mean`, `texture_mean`, and `fractal_dimension_worst`. We can see that `radius_mean` and `texture_mean` appear to have the strongest correlation out of the 3, with what appears to be a moderate positive correlation. There appears to be almost no correlation between `radius_mean` and `fractal_dimension_worst`, as well as between `texture_mean` and `fractal_dimension_worst`, although you could argue that there is a very slight positive correlation for both.

Figure 2: Scatter plot of `radius_mean` and `texture_mean`

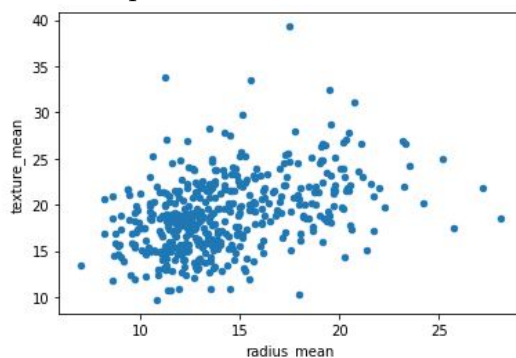


Figure 3: Scatter plot of `radius_mean` and `fractal_dimension_worst`

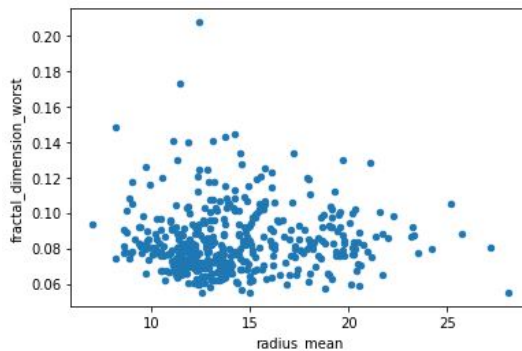
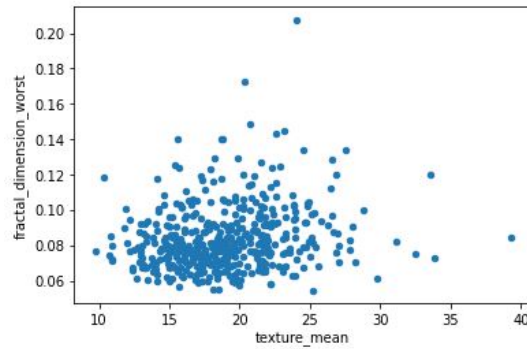


Figure 4: Scatter plot of texture\_mean and fractal\_dimension\_worst



## 4 Boxplot

Below are the 3 box plots for radius\_mean, texture\_mean, and fractal\_dimension\_worst. All 3 appear to have a symmetric, slightly right-skewed distribution, with all 3 also having numerous high outliers. This makes when considering the data, as you can't really have an "unusually small" cancerous growth (It wouldn't have been detected yet, or it might not even be cancerous yet at all!)

Figure 5: Box plot of radius\_mean

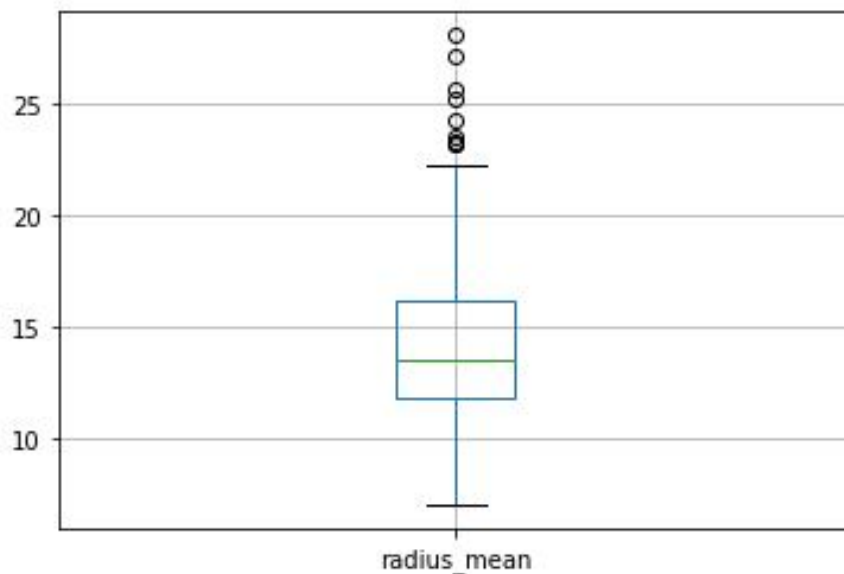


Figure 6: Box plot of texture\_mean

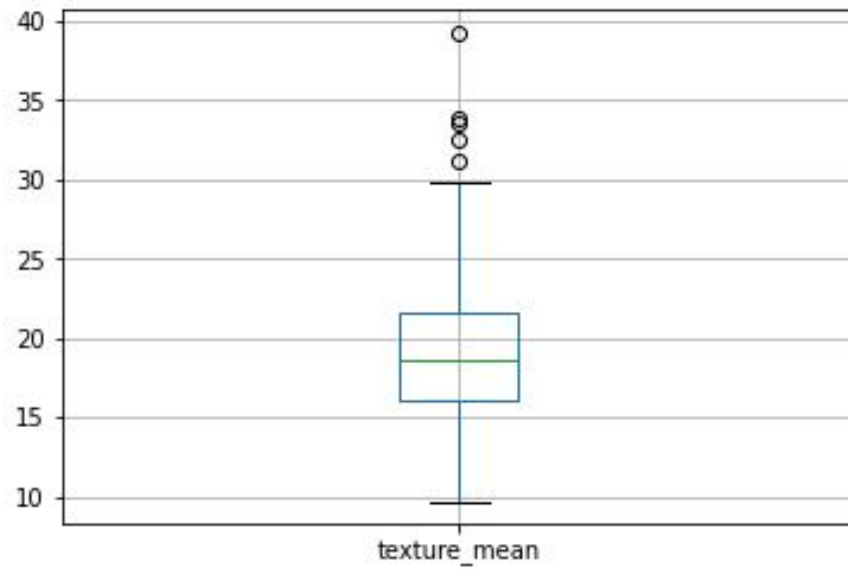


Figure 7: Box plot of fractal\_dimension\_worst

