1. **Data**

I have chosen a dataset from Kaggle.com about air pollution which contains Air quality index (AQI) values which means number of pollutants from cities across the world. I have chosen a column called “Ozone AQI value”. I chose 190 values just randomly selecting cells.

2. **Hypothesis**

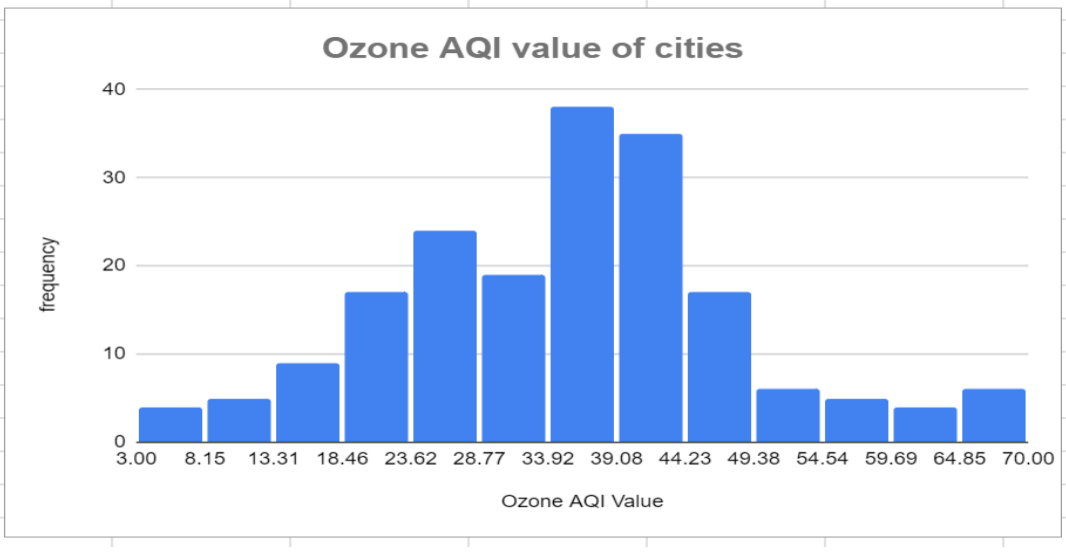
In kaggle there were histograms of each value and it looked normally distributed. The significance level would be 2.5%.

3. **Measures of Location, Dispersion and Skewness**

Mean	35.616
Median	37.000
St.dev	13.509
Skewness	-0.307

I have calculated skewness and it shows a minus number which means it is negatively skewed. I believe that this data is normally distributed because it is close to zero.

4. **Histogram**



The bucket size is 5.15.

According to the histogram the bars do not fully look symmetric, somewhere there are gaps, somewhere it is small. I would say that they are moderately symmetric and this data could be normally distributed.

Left	Right	Frequency	Percentage
22.107	49.125	135	0.7105
8.598	62.634	176	0.9263
-4.911	76.143	190	1.0000

This percentage that I obtained is slightly higher but it is close to the original percentages. So, I believe that based on that it might be normally distributed.

5. Chi-squared test

Intervals		Observed frequency	standardized intervals (z-scores)		P(Z<lcb)	P(Z<ucb)	Expected frequency
lcb	ucb		lcb	ucb			
3	8.15	4		-2.033	0.000	0.021	3.995
8.15	13.31	5	-2.033	-1.651	0.021	0.049	5.385
13.31	18.46	9	-1.651	-1.270	0.049	0.102	10.008
18.46	23.62	17	-1.270	-0.888	0.102	0.187	16.193
23.62	28.77	24	-0.888	-0.507	0.187	0.306	22.573
28.77	33.92	19	-0.507	-0.126	0.306	0.450	27.320
33.92	39.08	38	-0.126	0.256	0.450	0.601	28.720
39.08	44.23	35	0.256	0.638	0.601	0.738	26.075
44.23	49.38	17	0.638	1.019	0.738	0.846	20.451
49.38	54.54	6	1.019	1.401	0.846	0.919	13.964
54.54	59.69	5	1.401	1.782	0.919	0.963	8.214
59.69	64.85	4	1.782	2.164	0.963	0.985	4.207
64.85	70	7	2.164		0.985	1.000	2.894
		190					190

After combining cells:

O	E	(O-E)^2/E
9	9.380	0.015
9	10.008	0.102
17	16.193	0.040
24	22.573	0.090
19	27.320	2.534
38	28.720	2.998
35	26.075	3.055
17	20.451	0.582
6	13.964	4.542
5	8.214	1.258
11	7.101	2.141
190	190.000	17.357

6. Conclusion

Sig.level	Critical Value		
2.5%	17.53	R33<Q37	Accept the hypothesis

Degree of freedom is 8. So according to the Chi-square distribution table I found the intersection, which showed that Critical Value is 17.53. I accept the hypothesis because the test statistic was slightly lower than Critical Value (17.357<17.53).