Is there a difference in ages when it comes to substance abuse?

Table of Contents

Data Source

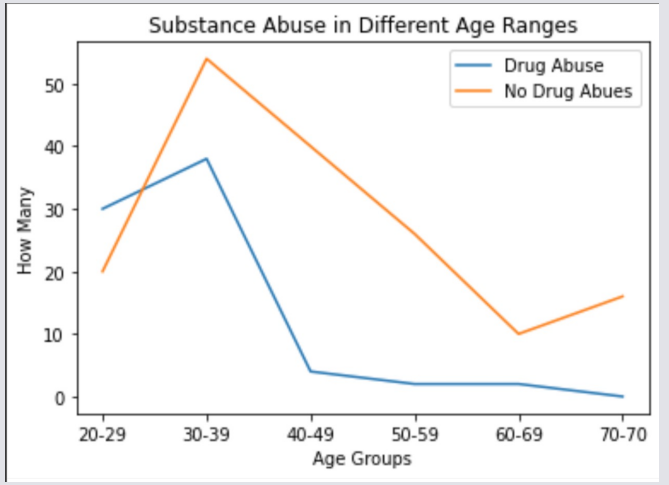
- In this analysis I used the homeless_pep.csv from an established data set source Kaggle. In this data set it is more focused on veterans and if they're homeless or not with a variety of different variables.

Research Question

- My research question is whether age has any correlation with substance abuse. As we grow into young adults in our twenties it is interesting to see if these young adults who are homeless are abusing drugs or we might see drug abuse being abused in all ranges of ages.

Data

Created was a line graph showing Ages 20-29, 30-39 … 60-69 were grouped to avoid clutter and provide more clarity. We can see that substance abuse peaked in our ages 30-39 but rapidly declines too little to no substance abuse in the older age groups.



Before diving deep into this analysis, we will be using the confidence interval of 95%. With a Confidence Interval of 47 compared to the mean of 44 we can see that there is a high chance when using this analysis will correspond with a real-world population that will be studied

```
                    mean  count        std
substanceabuse
0              44.879518    166  15.183242
1              32.131579     76   9.104346
------------------------------
                    mean  count        std    ci95_hi    ci95_lo
substanceabuse
0              44.879518    166  15.183242  47.189276  42.569760
1              32.131579     76   9.104346  34.178486  30.084672
```

The correlation coefficient table to the bottom right shows that between our AGE and substanceabuse variable has a coefficient of -0.4. This means that AGE and substanceabuse has a negative correlation between each other which is proven in our line graph

Using our numerical summary to the bottom right we can see that the average age is about 40 years old and our average substance abuse is at 0.3 which means that substance abuse is a lot less prevalent in our sample ages. Age and substance abuse kurtosis is less than 0 which means that the distribution of data has more mass in the tails than a normal distribution. This indicates that the distribution is more peaked and has thinner tails. This is patykurtic.

```
df.corr()
```

	CLIENT_KEY	AGE	INCOME	NIGHTS	substanceabuse	completed	probation	required
CLIENT_KEY	1.000000	0.019951	-0.159723	-0.004372	-0.108912	-0.020247	0.100226	-0.037018
AGE	0.019951	1.000000	0.016940	-0.638810	-0.400848	0.332826	-0.403322	-0.262026
INCOME	-0.159723	0.016940	1.000000	-0.068616	0.112165	0.088181	-0.022632	-0.104517
NIGHTS	-0.004372	-0.638810	-0.068616	1.000000	0.458512	-0.337638	0.409137	0.273281
substanceabuse	-0.108912	-0.400848	0.112165	0.458512	1.000000	-0.235029	0.210841	0.227647
completed	-0.020247	0.332826	0.088181	-0.337638	-0.235029	1.000000	-0.325032	-0.134302
probation	0.100226	-0.403322	-0.022632	0.409137	0.210841	-0.325032	1.000000	0.125883
required	-0.037018	-0.262026	-0.104517	0.273281	0.227647	-0.134302	0.125883	1.000000

	AGE	substanceabuse
count	242.000000	242.000000
mean	40.876033	0.314050
std	14.791256	0.465098
min	20.000000	0.000000
25%	30.000000	0.000000
50%	37.000000	0.000000
75%	49.000000	1.000000
max	79.000000	1.000000

```
Age Kurtosis:  0.2006882793018363
Age Skewness:  0.9116118614992007
Substance Abuse Kurtosis:  -1.3612291325619579
Substance Abuse Skewness:  0.8062813825719001
```

One-Way ANOVA

	df	sum_sq	mean_sq	F	PR(>F)
substanceabuse	1.0	8472.006420	8472.006420	45.945427	9.350906e-11
Residual	240.0	44254.274572	184.392811	NaN	NaN

- Using the One-Way Anova Table created we can see that substanceabuse has a f-score of 45.9 with a critical score of 0.05 we can assume that the significance of these results produced by the one-way ANOVA are statistically significant.
- A P-score of 9.350906e-11. The p-score measures the probability that the observed differences between the means of age and substanceabuse are due to chance. In the next slide we will go over the significance of these scores.

Analysis

Hypothesis

- Using our question 'Is there a difference in ages when it comes to substance abuse?' we can hypothesize two outcomes.
- Outcome 1 where there is no difference in ages when it comes to substance abuse. (H0)
- Outcome 2 where there is a difference in ages when it comes to substance abuse (H1)

Conclusion

- Using the output from our one-way ANOVA table we can see that we get a p-value of 9.350906e-11 and a f-value of 45.945427
- Since our p-value is significantly higher than our f-value, there is insufficient evidence to reject the null hypothesis that there is a difference in age when it comes to substance abuse.

Conclusions & Considerations

- With evidence from the correlation coefficient between AGE and substanceabuse we can rule out that there is no difference in ages when it comes to substanceabuse.
- One important considerations to make is that this is only a sample size of a homeless shelter. But with our confidence interval of 95% being so close to the mean we can rule that using this sample data analysis will correspond to a population size.