Daeun Kim

Econ 317 final paper

7 December 2022

Factors attributing to substance use disorder

Introduction:

Substance use disorder (SUD) or addiction has always been a major problem that not only has detrimental effects on the patients but also has great socioeconomic consequences. Experts speculate that factors such as poverty, mental illnesses, divorces, obesity, preventive education, education level, and family history contribute to the prevalence of SUD. I will determine how much of a correlation there is, if it exists, through multiple regression of data from the Substance Abuse and Mental Health Services Administration or SAMHSA.

Year (n)	SUD (y)	Prevention edu	Mental illness (AM	In poverty	# of divorce
2020	38679	2288	52859	37200	631
2019	19285	2580	51495	33900	747
2018	19343	2653	47635	38100	782
2017	18708	2607	46632	39500	787
2016	19048	2796	44652	40600	776
2015	19577	2990	43421	43100	801
2014	16994	3106	43553	46700	814
2013	17298	3037	43800	48800	832
2012	20700	2587.5	43700	46500	851
2011	18900	11.7%	45600	48500	877
2010	22000	11.5%	45900	46200	872
2009	21800	2550	45132	42900	840
2008	22231	2732	39800	40000	844
2007	22313	2834	40477	37300	856
2006	22613	2882	38886	36500	872
2005	22218	2946	39336	37000	847
2004	22506	3070	38404	36600	879

Table 1

All numbers except year are in the thousands for the U.S. population only. Prevention education refers to the number of people in ages 12-17 who participated in drug, tobacco, or alcohol prevention programs outside of school and AMI stands for any mental illness. Only data from 2020 to 2010 were initially used for regression, incorporating the mean age of initial use of substance, which was later omitted due to high multicollinearity and low p-value. The initial

regression shows a more linear trend in SUD (y-variable) but shows little correlation with the x-variables as shown in Table 2. Therefore, more data up until 2004 were added. SAMHSA data only existed after 1998 and with a different research methodology, which might result in problems in regression, so they were omitted as well. The output with the new data is in Table 3.

Table 2

Regression Statistics	
Multiple R	0.927154
R Square	0.859615
Adjusted R	0.719231
Standard Er	3199.312
Observatio	11

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regression	5	3.13E+08	62675469	6.123283	0.034249
Residual	5	51177997	10235599		
Total	10	3.65E+08			

	Coefficients	*andard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Jpper 95.0%*
Intercept	50728.32	137503.6	0.368924	0.727291	-302736	404192.6	-302736	404192.6
Mean age o	1789.926	6686.409	0.267696	0.799625	-15398	18977.89	-15398	18977.89
Prevention	-1.82851	3.010531	-0.60737	0.570129	-9.56733	5.910307	-9.56733	5.910307
Mental illn	-0.08008	1.043232	-0.07676	0.941792	-2.76179	2.601635	-2.76179	2.601635
In poverty	0.804343	0.420066	1.914802	0.113684	-0.27547	1.884157	-0.27547	1.884157
# of divorce	-0.11216	0.039272	-2.8559	0.035577	-0.21311	-0.01121	-0.21311	-0.01121

Regression Statistics	
Multiple R	0.79641043
R Square	0.634269573
Adjusted R	0.51235943
Standard Er	3365.998661
Observatio	17

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	235788113.7	58947028.42	5.20276295	0.011500474
Residual	12	135959363.9	11329946.99		
Total	16	371747477.5			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Jpper 95.0%*
Intercept	188567.7947	47941.65453	3.933276741	0.00198688	84111.9027	293023.6866	84111.903	293023.7
Prevention	-3.815818286	1.339915049	-2.847806127	0.014682442	-6.735242386	-0.896394187	-6.735242	-0.89639
Mental illn	-1.254764465	0.494128529	-2.539348349	0.025969381	-2.331378044	-0.178150886	-2.331378	-0.17815
In poverty	-0.200800345	0.210162404	-0.955453218	0.35820103	-0.658704886	0.257104197	-0.658705	0.257104
# of divorce	-0.11499553	0.032723192	-3.514190522	0.004268842	-0.186293239	-0.04369782	-0.186293	-0.0437

	Prevention edu	*Mental illness (AMI)*	*In poverty*	*# of divorce*
Prevention	1			
Mental illn	-0.30135572	1		
In poverty	-0.421570572	-0.004818492	1	
# of divorce	-0.192088634	-0.779498895	0.335014882	1

Table 3

The adjusted R square has dropped but most p-values except poverty have improved and multicollinearity was reduced to only be in between mental illness and divorce. The regression equation is as follows:

$$\hat{y}(estimated\ addiction) = 188567.7947 - 3.82Prevention\ edu - 1.25mental\ illness - 0.20poverty - 0.11divorce$$

which I find to be very counterintuitive. Most experts agree that all of the presented x-variables have a positive relationship with addiction, but they all came out negative although the p-value for poverty was too high, meaning it has no significance. This outcome suggests that with at least 98% confidence (using p-value=.02 for mental illness for reference) prevention education, mental illness, and divorce all decrease total addiction. Except for prevention education that tries to reduce addiction, the outcomes were hard to believe. These unexpected results may have been due to a limited number of observations and a time series lag response of the y-variable; the SUD may be responding to changes in the x-variables in the past year or years. To account for the time lag, I tried pulling all the x-variables one year forward, so that the y-variable for 2020 would correspond to x-variables in 2019 and filled the empty 2004 x-variables with forecasted values. But as Table 4 shows, the results did not improve either R-square or p-values, so I disregarded this approach.

Regression Statistics								
Multiple R	0.715							
R Square	0.511							
Adjusted R Sc	0.361							
Standard Erro	4234							
Observations	18							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	4	243880418.4	60970104.6	3.40101	0.0411469			
Residual	13	233051741.6	17927057					
Total	17	476932160						
	Coefficien	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1E+05	30721.5401	3.94596772	0.001674	54856.353	187596.058	54856.3533	187596.058
Prevention e	-2.561	1.321503222	-1.9381539	0.074631	-5.416211	0.29365754	-5.4162107	0.29365754
Mental illnes	-0.169	0.139309153	-1.2154272	0.245816	-0.470279	0.131639	-0.4702793	0.131639
ln poverty	-0.609	0.250463813	-2.431582	0.030241	-1.150117	-0.0679291	-1.1501175	-0.0679291
# of divorce	-0.074	0.028220888	-2.6145916	0.021405	-0.134754	-0.0128186	-0.1347536	-0.0128186

Table 4

I decided to look for other factors that may have impacted SUD and much research suggested that opioid prescription has begun to increase substantially in the 1990s resulting in a peak of deaths with 15 million caused by prescription opioids alone in 2011 (Raymond, 2022). OxyContin, a narcotic prescribed for pain relief was introduced to the market in 1996 and was widely blamed for the opioid epidemic (Raymond, 2022). Therefore, I decided to try a simple linear regression with SUD and the number of opioid prescriptions with CDC data (Centers, 2021). But this did not prove much either.

Regression Statistics								
Multiple R	0.31226904							
R Square	0.09751195							
Adjusted R Sq	0.0280898							
Standard Erro	5066.18891							
Observations	15							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	36051431.7	36051431.7	1.40462294	0.25716451			
Residual	13	333661511	25666270.1					
Total	14	369712943						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	30724.8869	8059.84029	3.81209625	0.00215721	13312.6606	48137.1133	13312.6606	48137.1133
number of op	-4.323E-05	3.6477E-05	-1.1851679	0.25716451	-0.000122	3.5572E-05	-0.000122	3.5572E-05

Table 5

R-square is too low while the p-value is too high. I assume that while opioids are one of the most lethal among the abused substances, there are many other substances that must be accounted for as well. According to DeWeerdt (2019), prescription opioids were only the first wave of the opioid epidemic and were followed by heroin and then with cheaper but more potent synthetic opioids like fentanyl.

Then there is COVID-19 to account for. As Table 1 shows, SUD for 2022 has surged despite the x-variables staying fairly stagnant, and many researchers attribute this to COVID-19 as it has increased rates of mental illnesses and drug use. This may make 2020 an outlier in the data, so I tried omitting 2020 and only using 2019 to 2004, but for some reason, the p-values

were too high except for poverty as Table 6 shows. Thus, despite the improvement in Adjusted R-square, I decided to not omit 2020 data.

Regression Statistics								
Multiple R	0.87763712							
R Square	0.77024692							
Adjusted R Sq	0.68670035							
Standard Erro	1075.8388							
Observations	16							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	4	42683077.4	10670769.3	9.21937172	0.00160682			
Residual	11	12731720.4	1157429.12					
Total	15	55414797.8						
	Coefficients	*tandard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	5475.49946	23444.9029	0.23354754	0.8196256	-46126.384	57077.3828	-46126.384	57077.3828
Prevention ed	-0.2207579	0.55209019	-0.3998584	0.69691586	-1.4359002	0.99438441	-1.4359002	0.99438441
Mental illness	-0.0163609	0.19836275	-0.0824796	0.93574703	-0.4529544	0.42023258	-0.4529544	0.42023258
In poverty	-0.2855711	0.06767255	-4.219896	0.00143663	-0.4345174	-0.1366248	-0.4345174	-0.1366248
# of divorce	0.03367729	0.01780451	1.89150331	0.08516569	-0.0055102	0.07286476	-0.0055102	0.07286476

Table 6

Residual Analysis:

In table 1, two observations for prevention education is in percentage, and this caused an error in the residual output, so I put approximate values instead.

Observation	*Predicted SUD (y)*	*Residuals*	*Std Residuals*
2020	33286.27457	5392.725	1.913090873
2019	20653.9715	-1368.97	-0.48564811
2018	21610.91484	-2267.91	-0.80455184
2017	23277.41456	-4569.41	-1.62101805
2016	23465.48503	-4417.49	-1.56712045
2015	19884.94029	-307.94	-0.10924305
2014	16911.2659	82.7341	0.029350253
2013	16414.56523	883.4348	0.313402013
2012	22500.09598	-1800.1	-0.63859124
2011	18308.61404	591.386	0.20979653
2010	17889.13139	4110.869	1.458347048
2009	22064.20538	-264.205	-0.09372791
2008	25135.64163	-2904.64	-1.03043321
2007	21501.22557	811.7744	0.287980219
2006	21351.4668	1261.533	0.447533939
2005	21726.36301	491.637	0.174410185
2004	18231.42429	4274.576	1.516422797

Table 7

All of the standard residuals in absolute value are below 2, which is a sign that none of the observations are extreme outliers. The residual plot in Table 8 also indicates that.

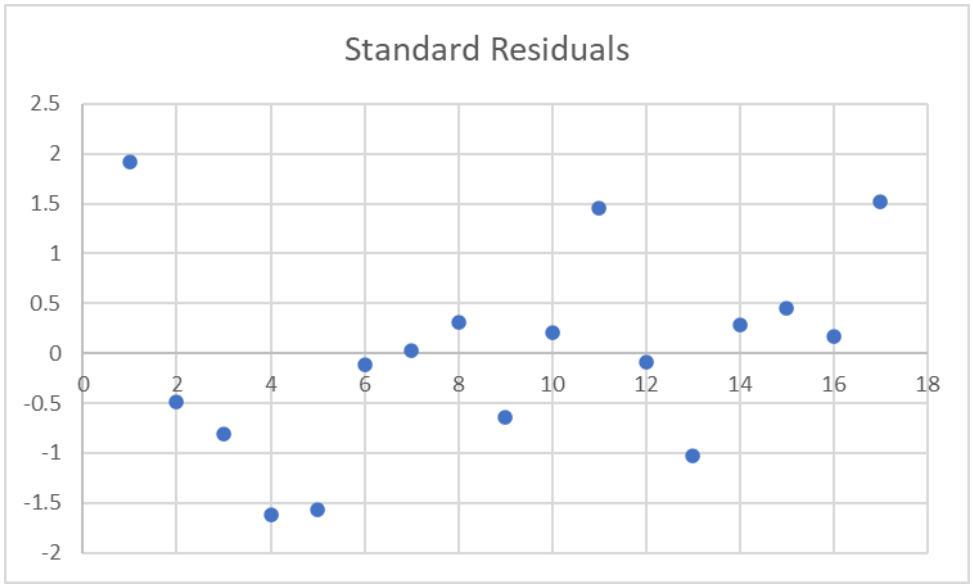

Table 8

Conclusion:

After trying various approaches and x-variables, I have concluded that other outside factors were contributing greatly to my y-variable but could not quite figure out how to incorporate that into the model. Disregarding the poverty variable with a p-value that was too high, other variables like divorce and mental illness have small coefficients, suggesting a weak or nonexistent trend. But prevention education has a rather high negative coefficient–considering that the data was in thousands–suggesting a stronger negative relationship between prevention education and SUD. This could show the effectiveness of prevention education provided for students with ages 12-17. The model could provide better regression outputs if there were significantly more observations, but the data collection was limited.

References

DeWeerdt, S. (2019). Tracing the US opioid crisis to its roots. *Nature*, *573*(7773), S10–S12. https://doi.org/10.1038/d41586-019-02686-2]

Centers for Disease Control and Prevention. (2021, November 10). *U.S. opioid dispensing rate maps*. Centers for Disease Control and Prevention. Retrieved December 1, 2022, from https://www.cdc.gov/drugoverdose/rxrate-maps/index.html

Raymond Sackler Family. (2022, April 21). *Opioid crisis*. Judge for Yourselves. Retrieved December 1, 2022, from https://www.judgeforyourselves.info/key-points/opioids/crisis/