THE EFFECT OF OTHER HEALTH CONDITIONS ON COVID-19 DIAGNOSES AND FATALITIES

Aaron J. Brown

STATISTICAL QUESTION/HYPOTHESIS

- This analysis is performed to determine if common health conditions such as: obesity, smoking, and drinking, affect the number of cases and deaths of COVID-19.
- Since COVID-19 is a respiratory disease, it is expected that several health conditions may increase the likelihood of contracting and dying from COVID-19, especially smoking.
- A positive relationship between cases/deaths by COVID and our outlined health conditions is expected to be unearthed by this analysis, relationships between the independent variables are expected to exist as well.

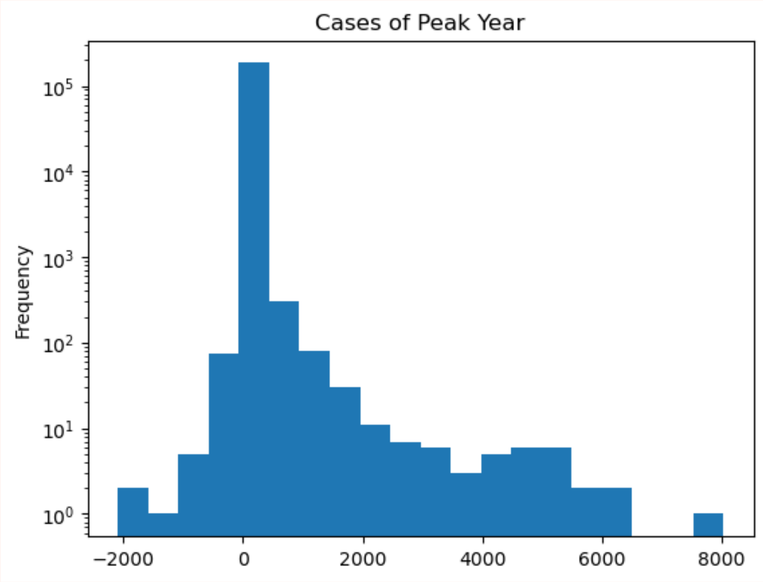
VARIABLES

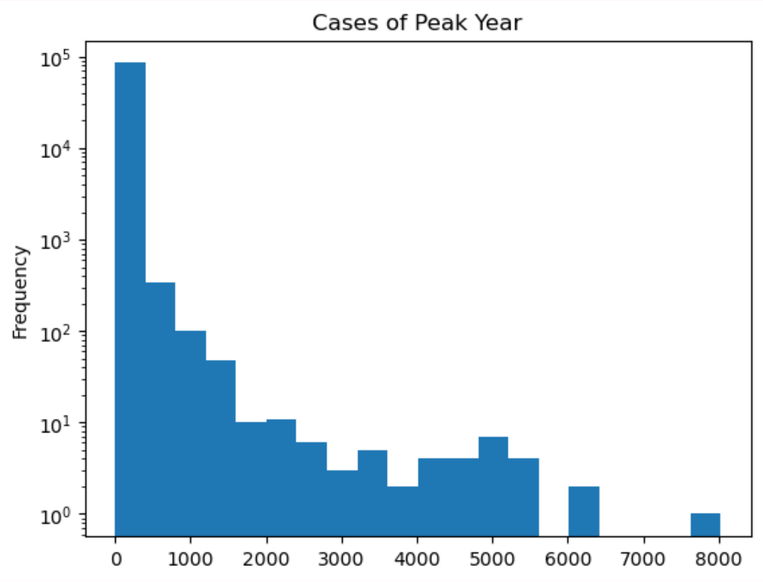
- **County: the county of the observation.**
- **State: the state of the observation’s county.**
- **Date: the date the observation was made.**
- **Cases: the number of total cases.**
- **Deaths: the number of total deaths.**
- **peak_deaths: number of deaths at middle of epidemic.**

VARIABLES (CONTINUED)

- percent_fair_or_poor_health: percentage of the population determined to have fair/poor health.
- percent_smokers: the percentage of the population determined to smoke.
- percent_adults_with_obesity: the percentage of the population determined to be obese.
- percent_excessive_drinking: the percentage of the population determined to be excessive drinkers.
- percent_adults_with_dibetes: the percentage of the population determined to have diabetes.

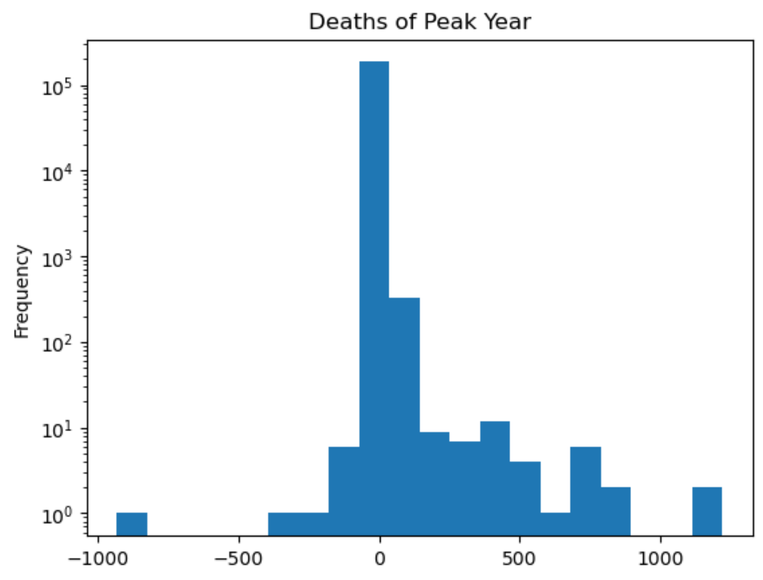
HISTOGRAMS

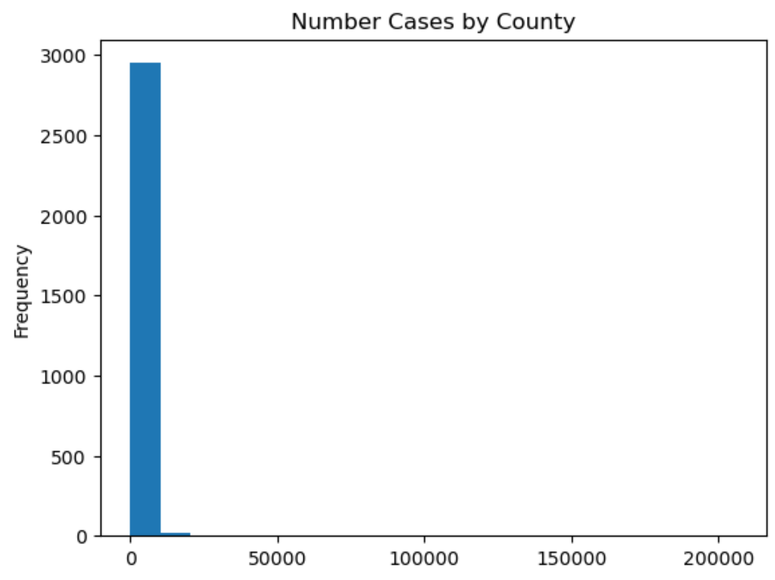


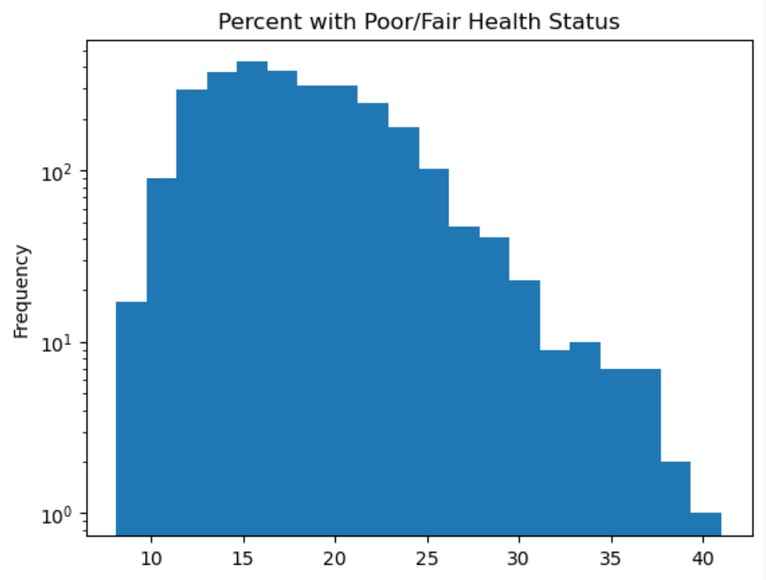


- **There are outliers here, values that falling well below zero; as the number of cases can not be less than zero, these values are trimmed from the data and a new histogram is produced**

HISTOGRAMS







- **These histograms illustrate the distributions of: deaths of peak year, number of cases by county, and percentage with poor/fair health statuses.**

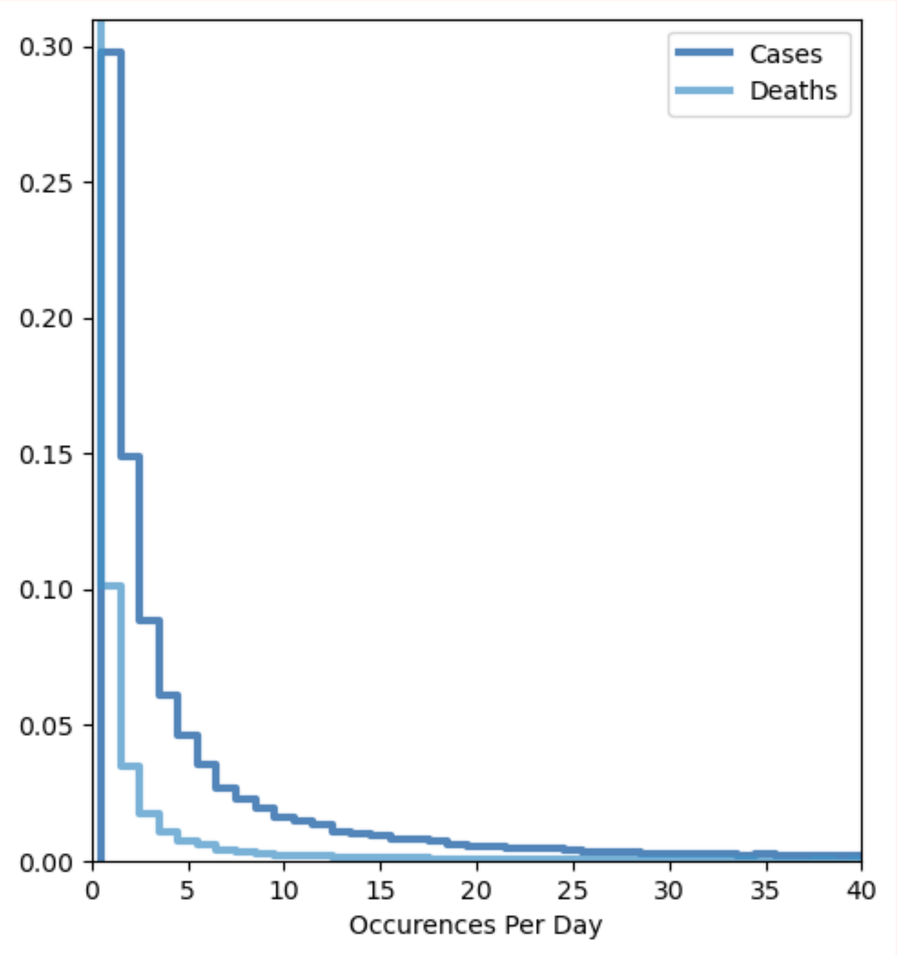
DESCRIPTIVE STATISTICS

	mean	median	var	std
cases	581.075890	43.000000	2.089346e+07	4570.936959
deaths	34.123909	1.000000	1.686084e+05	410.619516
percent_fair_or_poor_health	18.034635	17.343802	2.234296e+01	4.726834
percent_smokers	17.532791	17.087545	1.255117e+01	3.542763
percent_adults_with_obesity	33.026591	33.300000	2.948427e+01	5.429942
percent_excessive_drinking	17.483325	17.559710	1.008055e+01	3.174989
income_ratio	4.520333	4.411360	5.491752e-01	0.741064
percent_adults_with_diabetes	12.237759	11.700000	1.635616e+01	4.044275

	mean	median	var	std
peak_case	20.504134	3.0	14241.769449	119.338885
peak_deaths	1.165234	0.0	158.266400	12.580397

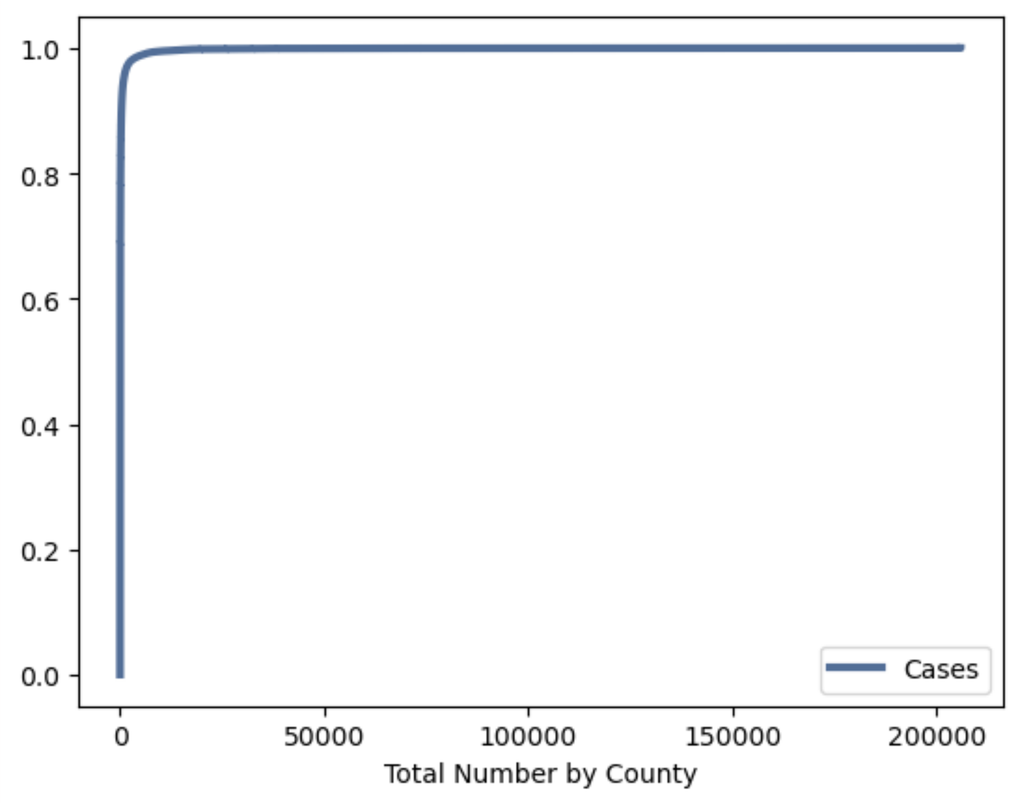
- **Here, we have the descriptive statistics (mean, median, var, and std) of the variables included in our analysis.**

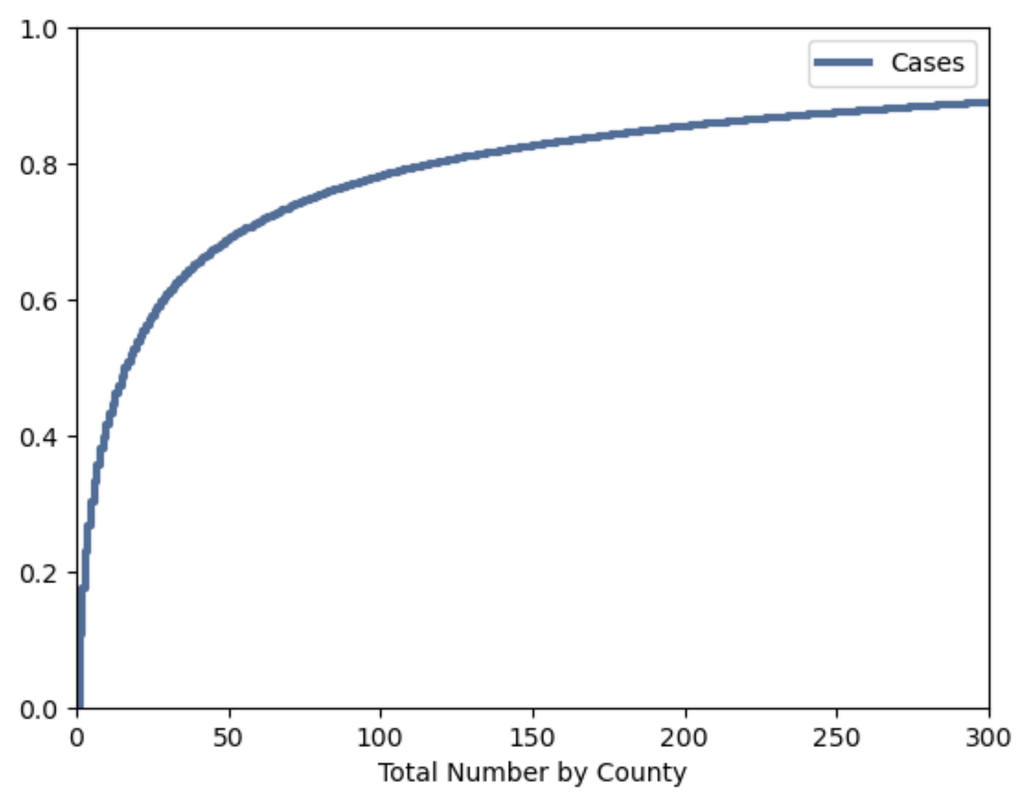
PROBABILITY MASS FUNCTION (PMF)



- **In this analysis death is an outcome of cases, so this PMF represents the probability of the number of people to die on a given day and the probability of the number of people to be diagnosed on a given day.**

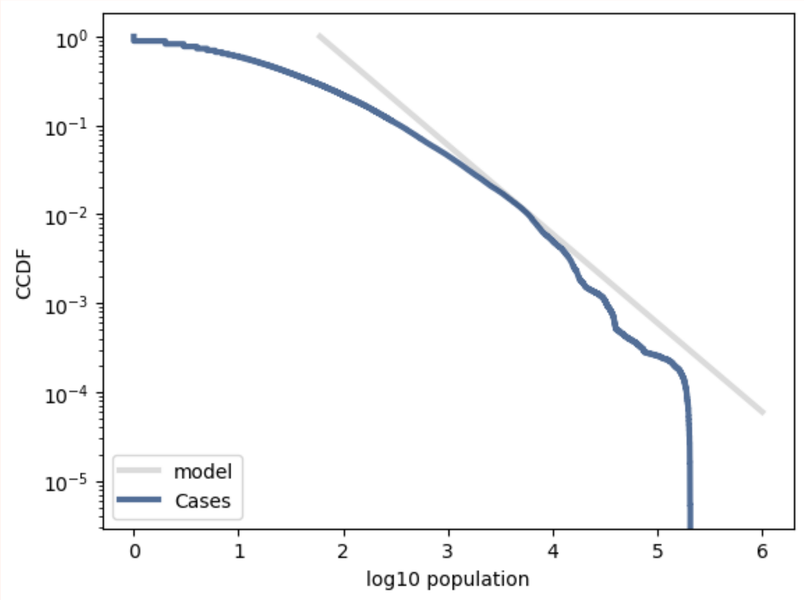
CUMULATIVE DISTRIBUTION FUNCTION (CDF)

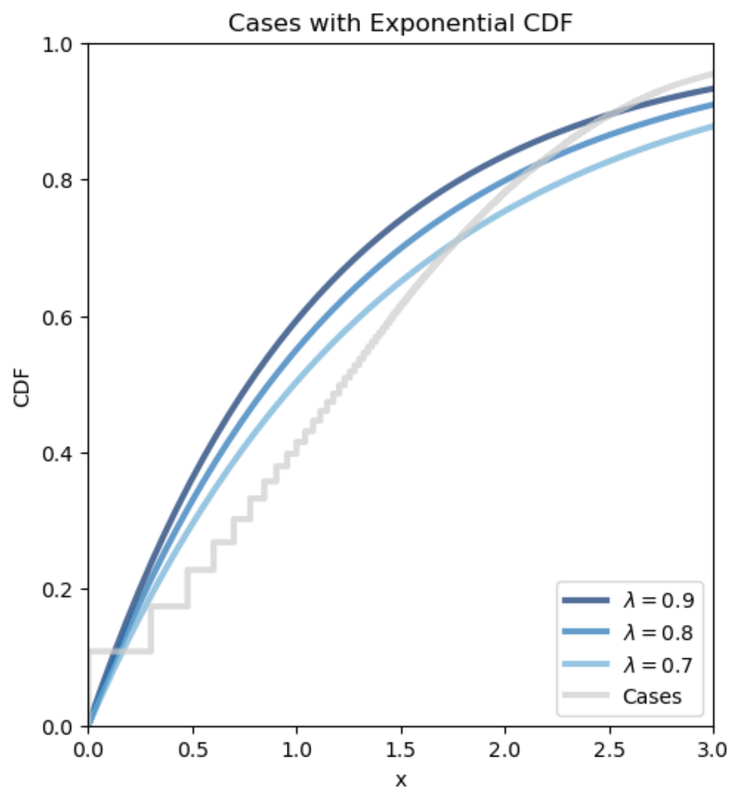




- **The slope produced by graphing the CDF indicated that about 70% of all US counties fall within 0-50 cases. To account for outliers, we appropriated 90% of the data.**

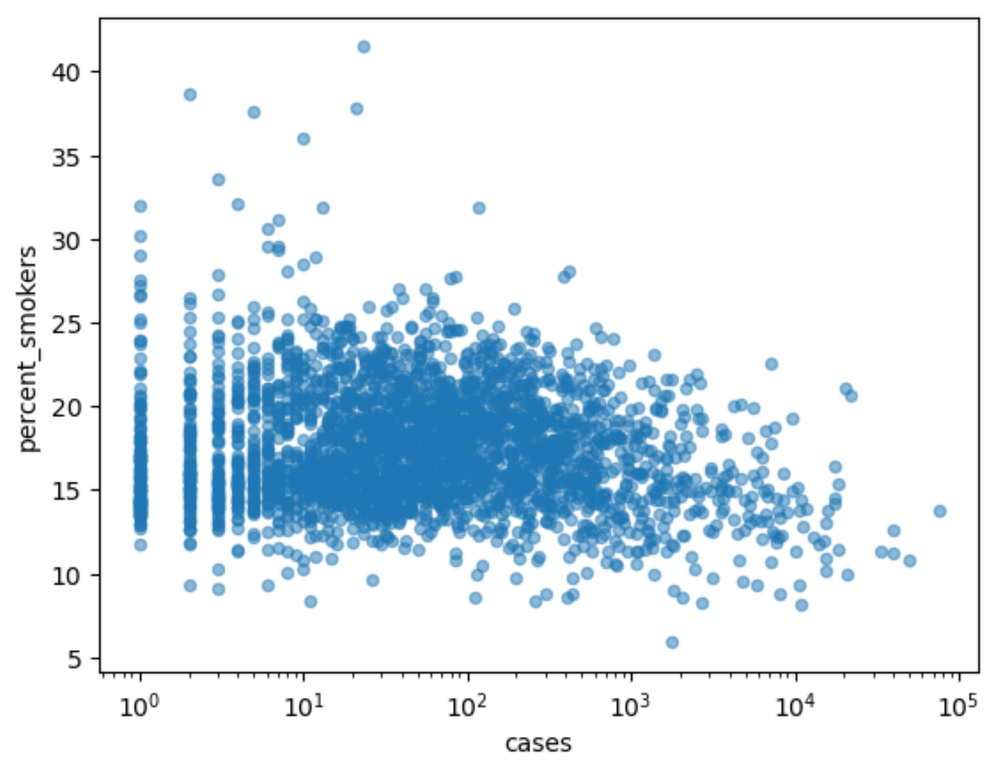
CUMULATIVE DISTRIBUTION FUNCTION (CDF)



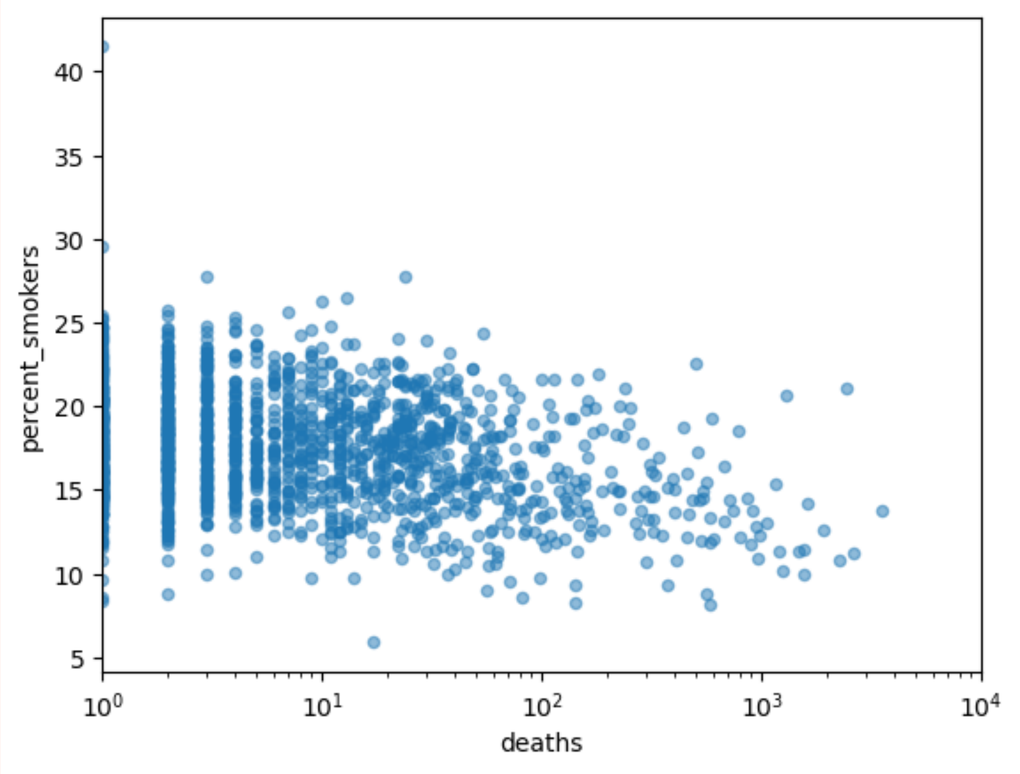


- **1) Pareto: A straight line is not produced here, which suggests that the Pareto Distribution is not a great fit either.**
- **2) The Exponential Distribution does not appear to be a good estimate of the log-log CDF of the total number of cases by county.**

ANALYTICAL DISTRIBUTION PLOTS

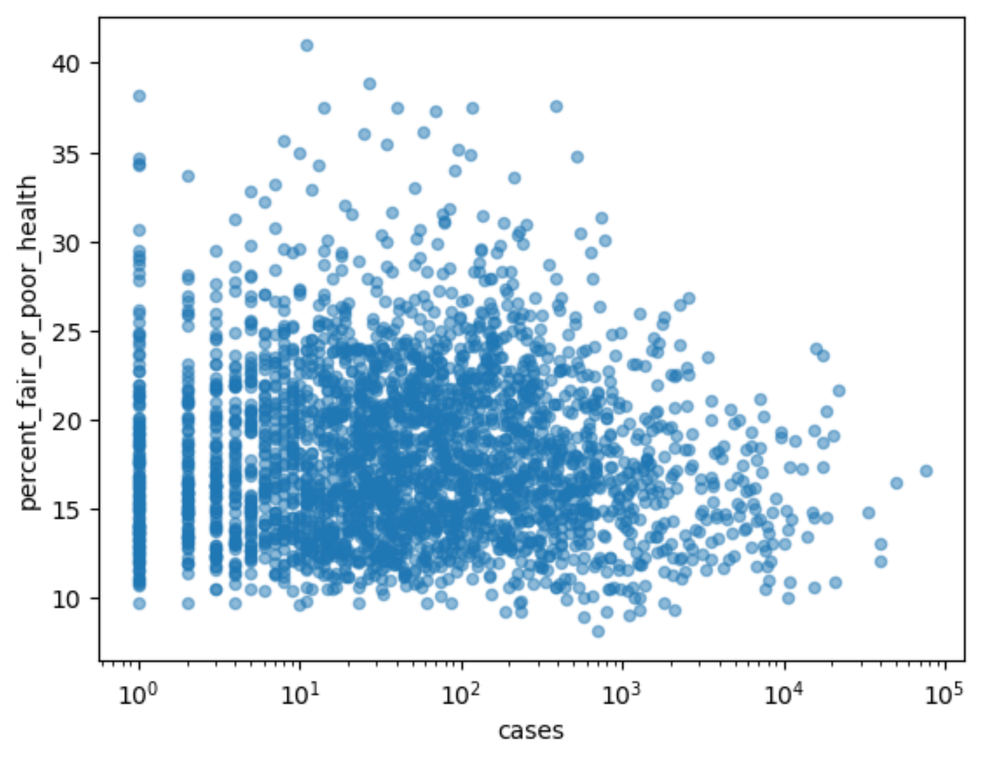


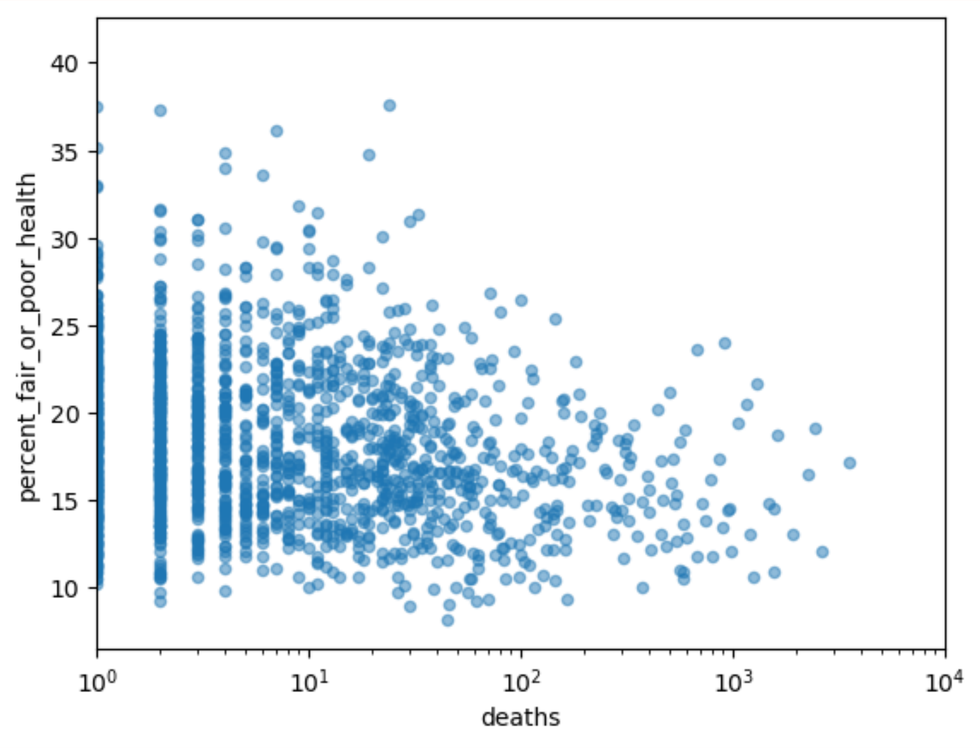
- **percent_smokers * cases**



- **percent_smokers * deaths**

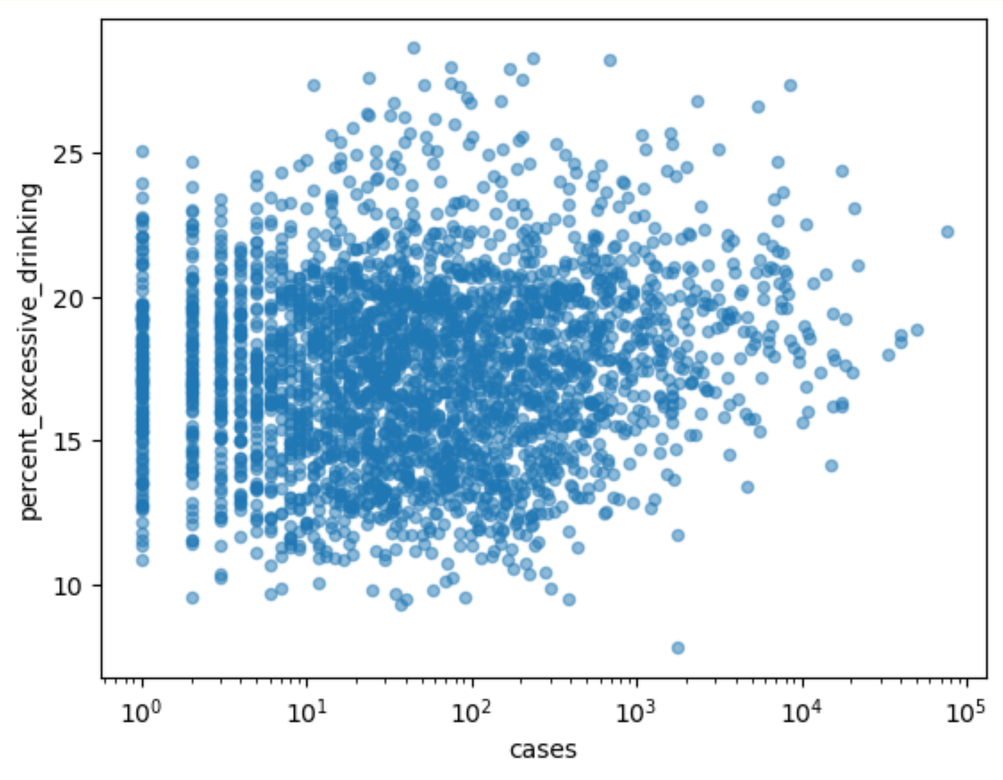
ANALYTICAL DISTRIBUTION PLOTS (CONTINUED)

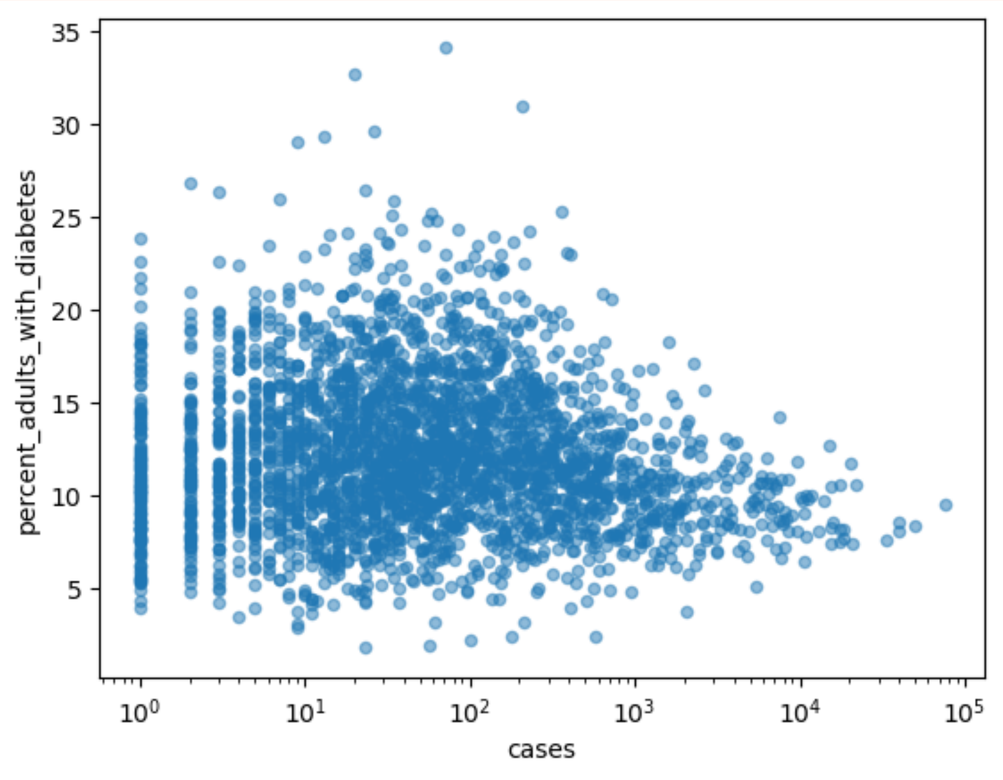




- **percent_fair_or_poor_health * cases**
- **percent_fair_or_poor_health * deaths**

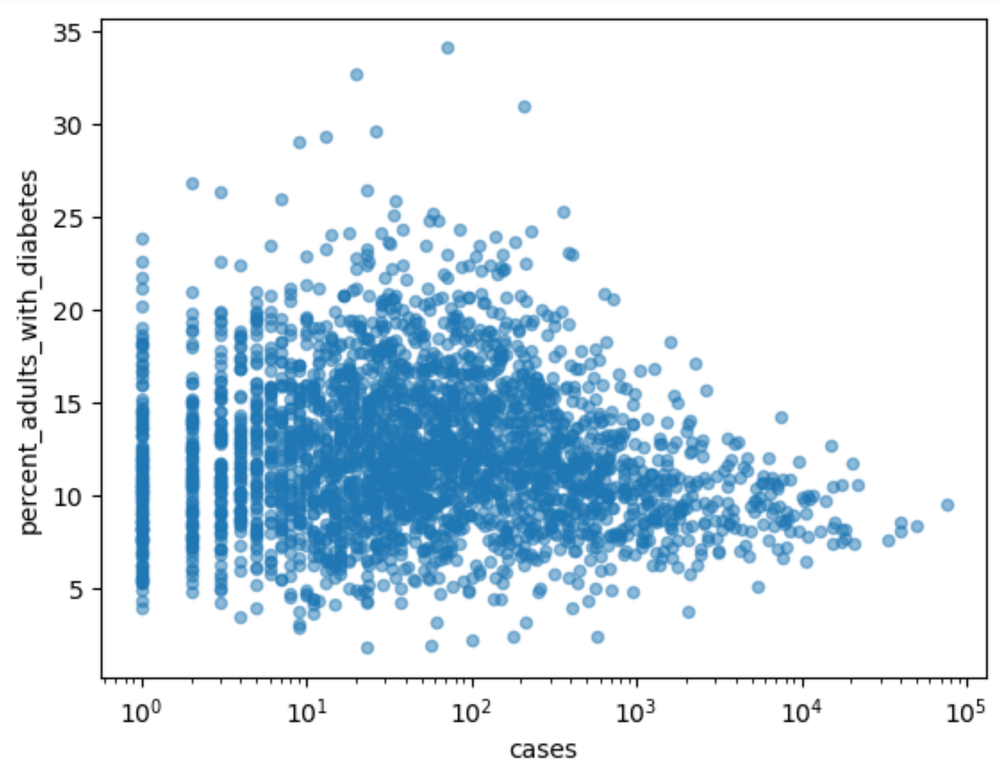
ANALYTICAL DISTRIBUTION PLOTS (CONTINUED)

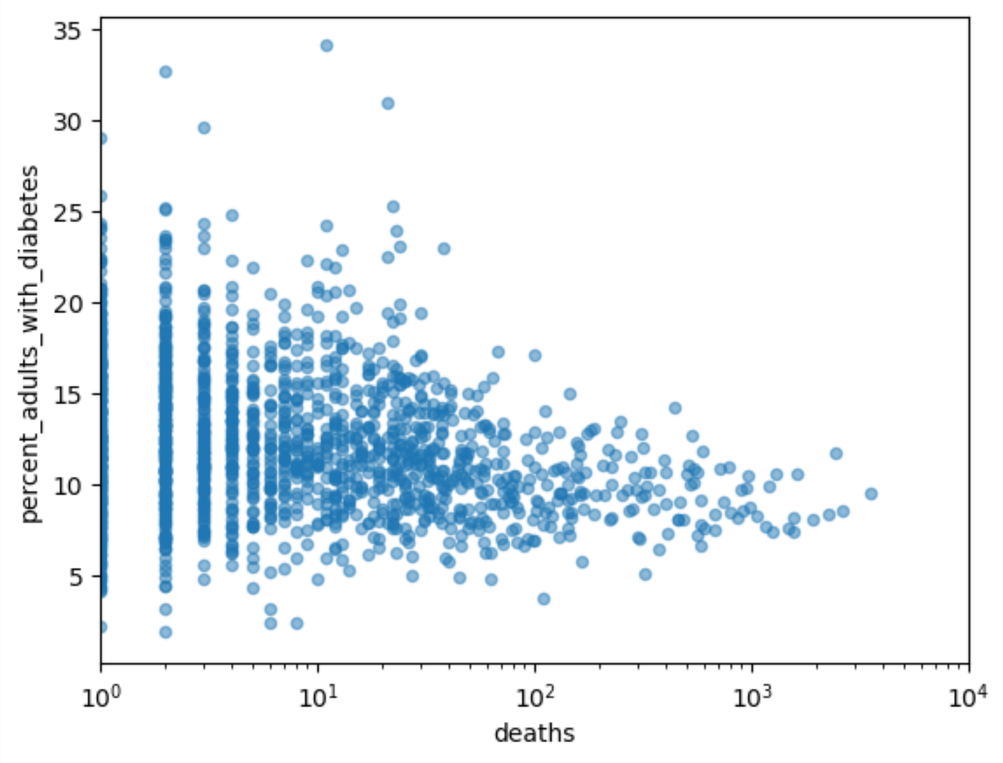




- **percent_excessive_drinking * cases**
- **percent_excessive_drinking * deaths**

ANALYTICAL DISTRIBUTION PLOTS (CONTINUED)





- **percent_adults_with_diabetes * cases**
- **percent_adults_with_diabetes * deaths**

HYPOTHESIS TESTING

```
class CorrelationPermute(thinkstats2.HypothesisTest):

    def TestStatistic(self, data):
        xs, ys = data
        test_stat = abs(thinkstats2.Corr(xs, ys))
        return test_stat

    def RunModel(self):
        xs, ys = self.data
        xs = np.random.permutation(xs)
        return xs, ys
```

```
Correlation: percent_adults_with_obesity
```

```
data = hat_filter.deaths, hat_filter.percent_adults_with_obesity
ht = CorrelationPermute(data)
pvalue = ht.PValue()
pvalue
```

```
0.0
```

Our P-value in this case is less than 0.001.

- **Given our low p-value in this case, we can determine that the relationships illustrated are statistically significant.**

REGRESSION ANALYSIS - DEATHS

OLS Regression Results

Dep. Variable:	deaths	**R-squared (uncentered):**	0.081
Model:	OLS	**Adj. R-squared (uncentered):**	0.080
Method:	Least Squares	**F-statistic:**	69.22
Date:	Sat, 03 Jun 2023	**Prob (F-statistic):**	3.89e-16
Time:	19:34:55	**Log-Likelihood:**	-5553.7
No. Observations:	786	**AIC:**	1.111e+04
Df Residuals:	785	**BIC:**	1.111e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
percent_adults_with_obesity	2.5794	0.310	8.320	0.000	1.971	3.188

Omnibus:	966.975	**Durbin-Watson:**	1.927
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	91329.898
Skew:	6.295	**Prob(JB):**	0.00
Kurtosis:	54.286	**Cond. No.**	1.00

Death = percent_adults_with_obesity

OLS Regression Results

Dep. Variable:	deaths	**R-squared (uncentered):**	0.082
Model:	OLS	**Adj. R-squared (uncentered):**	0.079
Method:	Least Squares	**F-statistic:**	34.80
Date:	Sat, 03 Jun 2023	**Prob (F-statistic):**	3.31e-15
Time:	19:34:55	**Log-Likelihood:**	-5553.5
No. Observations:	786	**AIC:**	1.111e+04
Df Residuals:	784	**BIC:**	1.112e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
percent_adults_with_obesity	1.2392	2.056	0.603	0.547	-2.796	5.274
percent_smokers	2.5691	3.895	0.660	0.510	-5.077	10.215

Omnibus:	967.787	**Durbin-Watson:**	1.925
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	91746.810
Skew:	6.303	**Prob(JB):**	0.00
Kurtosis:	54.406	**Cond. No.**	16.0

Death = percent_adults_with_obesity* percentsmokers

REGRESSION ANALYSIS - DEATHS (CONTINUED)

OLS Regression Results

Dep. Variable:	deaths	**R-squared (uncentered):**	0.083
Model:	OLS	**Adj. R-squared (uncentered):**	0.080
Method:	Least Squares	**F-statistic:**	35.27
Date:	Sat, 03 Jun 2023	**Prob (F-statistic):**	2.15e-15
Time:	19:34:55	**Log-Likelihood:**	-5553.1
No. Observations:	786	**AIC:**	1.111e+04
Df Residuals:	784	**BIC:**	1.112e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
percent_adults_with_obesity	4.1786	1.435	2.911	0.004	1.361	6.996
percent_adults_with_diabetes	-4.4038	3.860	-1.141	0.254	-11.980	3.173

Omnibus:	967.428	**Durbin-Watson:**	1.921
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	91513.372
Skew:	6.300	**Prob(JB):**	0.00
Kurtosis:	54.338	**Cond. No.**	14.1

- **Death = percent_adults_with_obesity* percent_with_diabetes**

OLS Regression Results

Dep. Variable:	deaths	**R-squared (uncentered):**	0.087
Model:	OLS	**Adj. R-squared (uncentered):**	0.085
Method:	Least Squares	**F-statistic:**	37.39
Date:	Sat, 03 Jun 2023	**Prob (F-statistic):**	3.08e-16
Time:	19:34:55	**Log-Likelihood:**	-5551.1
No. Observations:	786	**AIC:**	1.111e+04
Df Residuals:	784	**BIC:**	1.112e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
percent_adults_with_obesity	-0.8618	1.540	-0.560	0.576	-3.885	2.161
percent_fair_or_poor_health	6.3026	2.763	2.281	0.023	0.879	11.726

Omnibus:	964.791	**Durbin-Watson:**	1.925
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	91146.118
Skew:	6.268	**Prob(JB):**	0.00
Kurtosis:	54.244	**Cond. No.**	11.6

- **Death = percent_adults_with_obesity* percent_fair_or_poor_health**

- **Other than in the simple regression model (death*percent_adults_with_obesity), the P-values above suggest no relationship between our dependent variable (death) and our entered variables. The Simple Linear Regression Model appears to be the best fit for our data.**

REGRESSION ANALYSIS - DEATHS (CONTINUED)

- **Death = percent_adults_with_obesity* percent_with_diabetes**
- **Death = percent_adults_with_obesity* percent_fair_or_poor_health**