

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sb
import scipy
from scipy.stats import pearsonr
```

## Importing File

```
In [2]: df = pd.read_csv(r"D:\projects\NFHS-4.csv")
df = pd.DataFrame(df)
df.head()
```

```
Out[2]:
```

	India/States/UTs	Survey	Area	Households with electricity	improved drinking- water source	improved sanitation facility	using clean fuel for cooking	covered by a health scheme or health insurance	Women who are literate
0	India	NFHS-4	Total	88.2	89.9	48.4	43.8	28.7	68.4
1	India	NFHS-4	Rural	83.2	89.3	36.7	24.0	29.0	61.5
2	India	NFHS-4	Urban	97.5	91.1	70.3	80.6	28.2	81.4
3	India	NFHS-3	Total	67.9	87.6	29.1	25.5	4.8	55.1
4	Andhra Pradesh	NFHS-4	Total	98.8	72.7	53.6	62.0	74.6	62.9

## Sorting & filtering data

```
In [3]: df = df.loc[df["India/States/UTs"]!="India"]
dfa = df.loc[df["Survey"]=="NFHS-4"]
df= dfa.loc[dfa["Area"]=="Total"]
df.head()
```

	India/States/UTs	Survey	Area	with electricity	drinking- water source	sanitation facility	clean fuel for cooking	health scheme or health insurance	who are literate
4	Andhra Pradesh	NFHS-4	Total	98.8	72.7	53.6	62.0	74.6	62.9
8	Assam	NFHS-4	Total	78.2	83.8	47.7	25.1	10.4	71.8
12	Bihar	NFHS-4	Total	58.6	98.2	25.2	17.8	12.3	49.6
16	Chattisgarh	NFHS-4	Total	95.6	91.1	32.7	22.8	68.5	66.3
20	Gujarat	NFHS-4	Total	96.0	90.9	64.3	52.6	23.1	72.9

## Finding Correlation

In [13]:

```
df.corr()
```

	with electricity	drinking- water source	sanitation facility	clean fuel for cooking	health scheme or health insurance	who are literate	are literate	years who are stunted
<b>Households with electricity</b>	1.000000	0.017015	0.640418	0.614455	0.204939	0.616266	0.543579	-0.687852
<b>improved drinking- water source</b>	0.017015	1.000000	0.292365	0.223465	-0.030351	0.114337	0.113689	-0.171569
<b>improved sanitation facility</b>	0.640418	0.292365	1.000000	0.522123	-0.096089	0.848463	0.734739	-0.700625
<b>using clean fuel for cooking</b>	0.614455	0.223465	0.522123	1.000000	0.050660	0.454617	0.388571	-0.578838
<b>covered by a health scheme or health insurance</b>	0.204939	-0.030351	-0.096089	0.050660	1.000000	-0.085265	-0.129217	-0.163107
<b>Women who are literate</b>	0.616266	0.114337	0.848463	0.454617	-0.085265	1.000000	0.871440	-0.742965
<b>Men who are literate</b>	0.543579	0.113689	0.734739	0.388571	-0.129217	0.871440	1.000000	-0.659699
<b>Children under 5 years who are stunted</b>	-0.687852	-0.171569	-0.700625	-0.578838	-0.163107	-0.742965	-0.659699	1.000000
<b>Children under 5 years who are wasted</b>	-0.151915	0.179222	-0.501477	-0.108161	0.026579	-0.503629	-0.409633	0.419596
<b>Household Decision</b>	0.263682	-0.036786	0.389998	0.043818	0.002650	0.553468	0.396569	-0.445746
<b>Ever- married women who have ever experienced spousal violence</b>	-0.402542	-0.498596	-0.667915	-0.171135	0.329525	-0.552410	-0.519499	0.384881
<b>Women owning a house and/or land</b>	-0.448906	-0.528707	-0.352321	-0.367453	0.112382	-0.258080	-0.287256	0.196648

	with electricity	unimproved water source	sanitation facility	clean fuel for cooking	health scheme or health insurance	who are literate	are literate	years who are stunted
<b>Women having a bank ac</b>	0.549625	0.388429	0.580506	0.599963	0.204603	0.548681	0.451028	-0.666969
<b>WHMP</b>	0.484229	0.167490	0.835590	0.534945	-0.138148	0.854236	0.730756	-0.692762

```
In [21]: pearson_coeff, p_value = scipy.stats.pearsonr(df["WHMP "],df["Children
under 5 years who are stunted "])
print("The pearson correlation coefficient is ", pearson_coeff, "with a p
value of", p_value)
```

```
The pearson correlation coefficient is -0.6927616056281707 with a p value of 2.84040
76401036016e-06
```

```
In [22]: pearson_coeff, p_value = scipy.stats.pearsonr(df["Household
Decision"],df["Children under 5 years who are stunted "])
print("The pearson correlation coefficient is ", pearson_coeff, "with a p
value of", p_value)
```

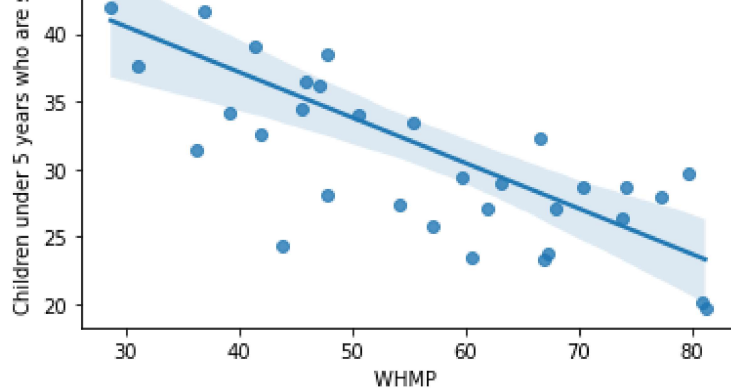
```
The pearson correlation coefficient is -0.44574613137860686 with a p value of 0.0064
38285344232916
```

```
In [23]: pearson_coeff, p_value = scipy.stats.pearsonr(df[" Women who are
literate"],df["Children under 5 years who are stunted "])
print("The pearson correlation coefficient is ", pearson_coeff, "with a p
value of", p_value)
```

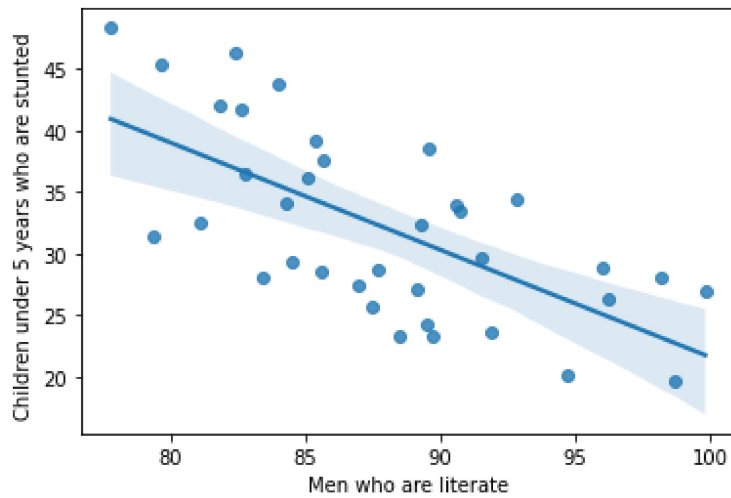
```
The pearson correlation coefficient is -0.7429646000720567 with a p value of 2.11108
66191481512e-07
```

## Plotting

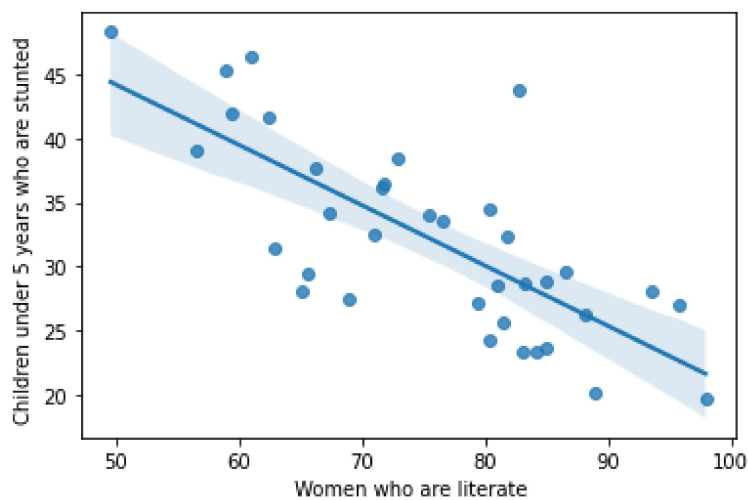
```
In [24]: plot = sb.regplot(x = "WHMP ", y = "Children under 5 years who are stunted
", data = df)
```

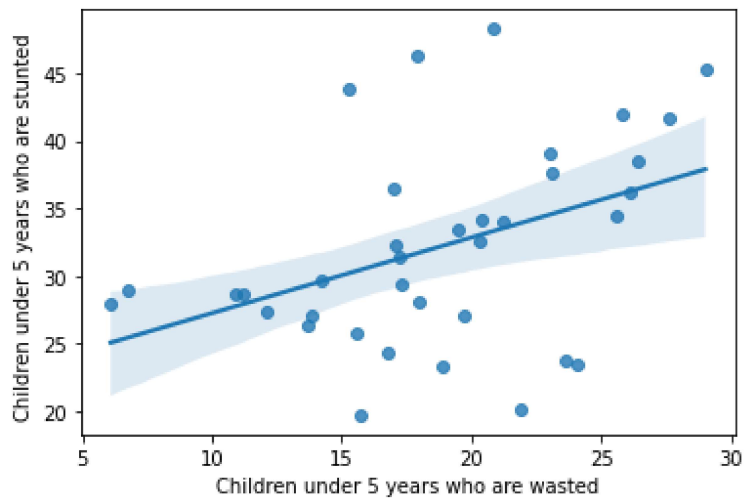


In [5]: `plot = sb.regplot(x = "Men who are literate", y = "Children under 5 years who are stunted ", data = df)`

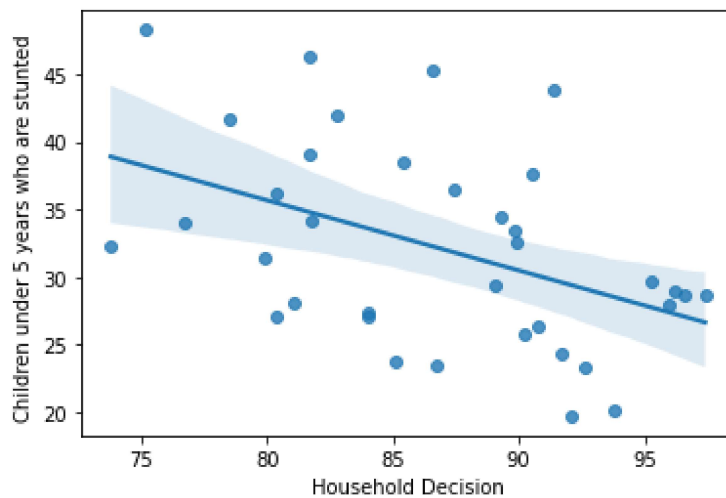


In [6]: `plot = sb.regplot(x = " Women who are literate", y = "Children under 5 years who are stunted ", data = df)`

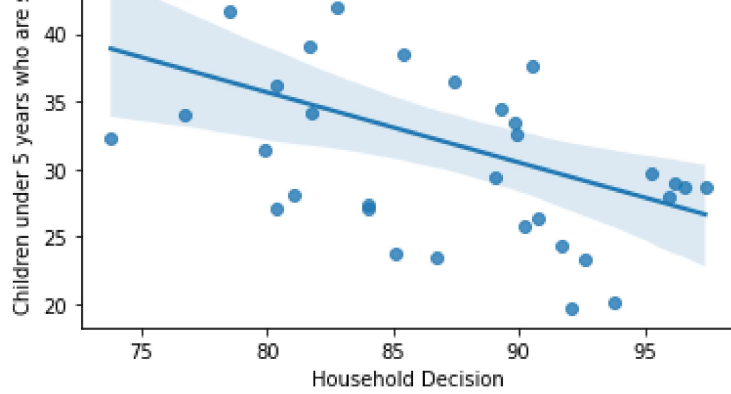




```
In [8]: plot = sb.regplot(x = "Household Decision", y = "Children under 5 years  
who are stunted ", data = df)
```



```
In [9]: plot = sb.regplot(x = "Household Decision", y = "Children under 5 years  
who are stunted ", data = df)
```



```
In [11]: plot = sb.regplot(x = "Household Decision", y = "Children under 5 years  
who are wasted", data = df)
```

