Non-parametric Inference on Global CO_2 Emission from Cement Production

Sampriti Dutta Sankhadeep Mitra Soumya Paul

Under the guidance of:
Prof. Sharmishtha Mitra
Professor, Department of Mathematics and Statistics, IIT Kanpur

Data Processing

Here is the link of the dataset we are going to use Cement Emissions.

Data Processing

Here is the link of the dataset we are going to use Cement Emissions.

- We have subsetted the data from 1931 onward and dropped any columns with all *NAs* or 0s.

Data Processing

Here is the link of the dataset we are going to use Cement Emissions.

- We have subsetted the data from 1931 onward and dropped any columns with all *NAs* or 0s.
- Column named *Global*, which actually represents the total CO_2 emissions for different time stamps is deleted that as it is not of our interest.

Year	Argentina	Australia	Belgium	Brazil	Canada	Chile	China	Democratic Republic of the Congo	Denmark	Egypt	Finland	Israel	Italy	Japan	Mozambique	Norway	Peru	Portugal	Romania	Spain	Sweden	Syria	Turkey	USA
1931	265.3	196.3	1218.0	83.59	799.5	50.88	97.93	21.81	250.8	119.9	79.95	39.98	1519.0	1788.0	10.90	109.70	14.54	47.25	98.14	806.8	257.9	3.63	50.83	10810
1932	247.1	123.6	1039.0	72.69	363.4	54.51	79.57	7.27	203.6	119.9	76.23	50.88	1545.0	1843.0	10.90	117.30	10.90	58.15	105.40	705.1	241.0	10.90	54.46	6656
1933	254.5	159.9	963.1	109.00	189.0	69.05	113.20	3.63	272.6	141.7	80.05	65.42	1756.0	2366.0	10.18	110.80	14.53	79.95	109.00	694.1	200.7	14.54	58.15	5587
1934	279.8	207.2	937.6	159.90	272.6	101.70	97.93	3.63	381.6	145.4	112.70	69.05	2013.0	2304.0	7.27	123.60	21.82	90.94	156.30	672.3	287.1	10.90	83.59	6900
1935	356.3	276.2	1087.0	181.60	272.6	141.60	156.10	3.63	374.3	189.0	134.30	90.86	2086.0	2904.0	7.27	130.80	29.07	105.40	189.00	668.7	367.1	14.54	65.42	6684
1936	410.7	323.5	1163.0	239.70	388.9	123.50	428.50	3.63	392.4	167.2	163.50	76.32	1890.0	3082.0	7.27	149.00	36.34	119.90	185.30	297.9	392.5	29.07	69.05	9950
1937	512.3	363.4	1486.0	283.50	483.4	156.30	437.70	3.77	334.4	163.5	203.40	79.95	2155.0	2984.0	7.27	159.90	39.96	127.10	225.30	189.0	432.5	61.78	105.40	10370
1938	614.3	178.1	1508.0	305.50	432.5	181.70	9.18	7.50	316.2	185.3	236.20	47.25	2279.0	2729.0	11.99	163.60	50.86	130.80	221.70	294.4	490.6	61.78	130.90	9239
1939	556.0	338.0	1261.0	345.50	450.7	167.20	223.40	17.44	345.3	185.0	279.60	54.51	2526.0	2508.0	14.54	192.60	58.18	145.40	261.70	588.8	585.1	58.15	141.70	10790
1940	534.4	348.8	105.4	367.10	592.4	189.00	272.40	11.45	218.1	178.1	149.00	72.69	2373.0	2101.0	14.54	167.20	61.78	134.50	196.30	770.5	345.3	21.81	130.90	11550
1941	577.9	341.6	312.5	378.00	657.8	178.10	495.80	21.81	207.2	207.2	152.60	58.15	1381.0	2882.0	14.54	156.30	83.59	134.50	305.20	810.4	319.7	21.81	134.60	14520
1942	530.7	279.8	250.8	370.50	716.0	181.60	654.90	32.71	214.4	207.2	87.22	109.00	1108.0	2151.0	10.90	185.50	94.49	119.90	181.70	814.1	407.0	54.51	105.40	16090
1943	472.3	196.2	229.0	370.50	574.2	185.30	639.60	32.71	316.2	159.9	116.40	83.59	846.4	1861.0	15.66	152.60	101.80	123.40	225.30	839.5	457.8	36.34	83.59	12040
1944	534.2	218.1	298.0	399.80	563.3	178.10	489.70	43.61	319.8	210.6	87.22	87.22	668.0	1465.0	13.63	159.80	123.60	120.10	159.90	912.1	523.3	18.17	141.70	8032
1945	537.9	348.9	319.8	381.60	665.1	203.50	18.36	37.13	109.0	214.4	137.30	72.69	563.8	577.9	16.87	69.05	130.80	130.80	123.60	952.2	600.1	18.17	141.70	9114
1946	563.3	363.4	934.0	406.80	908.5	287.10	85.69	39.98	247.1	290.7	163.50	130.80	999.4	457.9	13.63	214.40	127.20	159.90	156.20	1061.0	723.2	21.81	159.90	14650
1947	672.5	443.4	1290.0	450.70	937.6	298.00	174.40	57.50	312.5	319.8	207.20	163.50	1363.0	610.6	17.10	232.60	127.20	210.80	207.20	1079.0	766.8	25.44	174.40	16650
1948	617.8	508.8	1646.0	548.80	1108.0	265.50	82.64	61.78	381.5	378.0	276.20	79.95	1555.0	919.5	18.17	261.60	138.20	247.10	221.70	1152.0	734.1	25.44	170.80	18420
1949	716.0	530.6	1446.0	632.10	1250.0	243.50	274.70	72.69	410.8	439.7	323.50	119.90	1995.0	1621.0	23.26	294.40	141.70	258.00	276.20	1112.0	839.5	29.07	185.30	18790
1950	777.7	632.4	1759.0	686.90	1312.0	254.40	586.90	87.22	432.5	505.2	367.30	189.00	2475.0	2206.0	25.34	287.10	163.40	283.50	319.80	1247.0	963.0	33.00	196.30	20120
1951	770.5	610.6	2173.0	719.60	1334.0	345.30	1036.00	101.80	487.0	559.7	410.90	218.10	2758.0	3238.0	39.98	345.10	178.10	316.40	563.40	1356.0	1007.0	32.71	196.20	22040
1952	766.8	672.3	2032.0	799.40	1454.0	403.40	1203.00	119.70	599.8	468.8	385.10	221.70	3289.0	3518.0	39.98	348.80	181.60	359.80	741.40	1465.0	1047.0	76.32	225.30	22260
1953	817.7	788.7	2286.0	1003.00	1744.0	378.00	1615.00	123.40	621.5	541.5	461.80	229.00	3871.0	4332.0	43.61	374.50	221.70	381.60	937.80	1610.0	1163.0	109.00	261.70	23520
1954	843.1	944.9	2162.0	1236.00	1759.0	381.60	1915.00	170.80	603.3	661.6	512.20	279.80	4328.0	5277.0	50.72	381.60	239.70	385.20	792.30	1883.0	1217.0	124.40	349.00	24130
1955	923.1	984.9	2319.0	1370.00	1973.0	396.10	1873.00	199.90	621.5	676.0	516.10	327.10	5270.0	5222.0	69.24	396.10	268.90	385.00	955.90	2141.0	1261.0	130.80	407.10	26580
1956	1021.0	1058.0	2304.0	1617.00	2253.0	381.70	2660.00	225.30	585.1	668.7	476.10	301.60	5604.0	6436.0	76.50	443.40	272.60	505.20	1036.00	2337.0	1232.0	159.90	479.70	28390
1957	1170.0	1148.0	2326.0	1679.00	2711.0	359.80	2855.00	229.00	574.2	723.2	468.80	356.20	5906.0	7501.0	83.59	501.50	268.90	483.40	1163.00	2457.0	1210.0	156.30	625.10	27010
1958	1221.0	1214.0	2006.0	1872.00	2758.0	359.80	3871.00	192.40	527.0	748.7	457.90	352.50	6338.0	7407.0	87.22	508.80	298.00	505.20	1272.00	2631.0	1239.0	192.60	748.80	27760
1959	1170.0	1294.0	2195.0	1897.00	2817.0	414.30	5107.00	170.80	687.0	883.1	577.90	385.20	7120.0	8533.0	105.40	559.70	287.10	508.80	1410.00	2831.0	1396.0	222.10	857.50	30230
1960	1305.0	1381.0	2170.0	2210.00	2595.0	414.30	6514.00	98.13	712.2	1010.0	621.50	399.80	7915.0	11140.0	109.00	570.60	297.90	592.40	1508.00	2835.0	1388.0	239.90	1007.00	28800
1961	1436.0	1414.0	2351.0	2309.00	2784.0	436.10	2585.00	68.94	781.4	1018.0	665.20	417.90	8911.0	12180.0	105.40	628.70	294.40	614.20	1635.00	3275.0	1504.0	265.30	1003.00	28700
1962	1446.0	1450.0	2366.0	2462.00	3086.0	505.30	2497.00	98.13	806.7	1105.0	672.20	472.50	9969.0	14230.0	87.38	697.80	345.30	694.10	1723.00	3605.0	1519.0	301.30	1148.00	29880
1963	1254.0	1541.0	2326.0	2493.00	3144.0	577.90	3355.00	120.10	752.3	1239.0	705.10	504.90	10920.0	14800.0	83.59	712.30	374.20	708.70	2159.00	3831.0	1632.0	338.00	1334.00	31440
1964	1439.0	1792.0	2889.0	2679.00	3547.0	625.10	5032.00	112.90	937.6	1247.0	777.70	541.30	11290.0	16300.0	90.86	763.20	403.40	802.90	2348.00	4201.0	1792.0	312.50	1454.00	33030
1965	1632.0	1879.0	2918.0	2651.00	3747.0	585.10	6801.00	123.40	984.9	1196.0	875.90	621.50	10230.0	16160.0	109.00	792.30	501.70	828.60	2671.00	4790.0	1864.0	334.40	1646.00	33290
1966	1723.0	1817.0	2864.0	2825.00	4001.0	675.80	8387.00	130.80	1036.0	1207.0	777.70	578.00	11080.0	18910.0	112.70	901.30	526.80	850.40	2907.00	5967.0	1857.0	305.30	1912.00	34570
1967	1756.0	1890.0	2875.0	2965.00	3583.0	610.40	6085.00	145.40	1087.0	1359.0	748.70	396.10	12980.0	21240.0	123.60	1065.00	537.90	901.30	3129.00	6658.0	1926.0	298.00	2100.00	33590
1968	2064.0	1941.0	2838.0	3337.00	3660.0	610.70	5253.00	145.40	1127.0	1555.0	730.40	545.00	14560.0	23560.0	141.70	1134.00	548.80	919.50	3474.00	7392.0	1933.0	454.30	2337.00	35320
1969	2148.0	2130.0	3100.0	3552.00	3700.0	708.70	7613.00	159.90	1290.0	1785.0	886.80	646.70	15490.0	25400.0	149.00	1232.00	563.20	1007.00	3714.00	8064.0	1955.0	461.60	2864.00	35700
1970	2344.0	2228.0	3325.0	4047.00	3562.0	1010.00	10720.00	207.20	1287.0	1820.0	926.80	683.40	16350.0	28270.0	196.30	1290.00	559.80	1152.00	4016.00	8253.0	2013.0	476.40	3151.00	34920
1971	2736.0	2314.0	3397.0	4407.00	4068.0	663.50	13120.00	225.30	1357.0	1937.0	903.30	683.60	15560.0	29270.0	207.20	1357.00	714.10	1207.00	4201.00	8433.0	1955.0	450.70	3703.00	35510
1972	2699.0	2436.0	3442.0	5117.00	4477.0	658.00	14720.00	236.20	1432.0	1890.0	969.80	740.00	16210.0	32530.0	232.60	1351.00	798.90	1358.00	4531.00	9562.0	1922.0	498.10	4107.00	36520
1973	2570.0	2591.0	3390.0	6023.00	4995.0	641.80	15460.00	265.30	1444.0	1788.0	1018.00	593.20	17410.0	38190.0	301.60	1352.00	1157.00	1603.00	4830.00	10860.0	2094.0	417.70	4332.00	36920
1974	2675.0	2568.0	3563.0	6707.00	5135.0	652.70	15340.00	308.90	1251.0	1614.0	1065.00	833.00	17230.0	35610.0	229.00	1309.00	931.50	1636.00	5476.00	10740.0	1648.0	476.10	4294.00	36810
1975	2715.0	2473.0	3252.0	7525.00	4933.0	455.00	19100.00	308.90	1126.0	1770.0	991.50	1000.60	16630.0	31800.0	138.10	1347.00	928.80	1655.00	5623.00	11560.0	1556.0	490.60	5103.00	36460

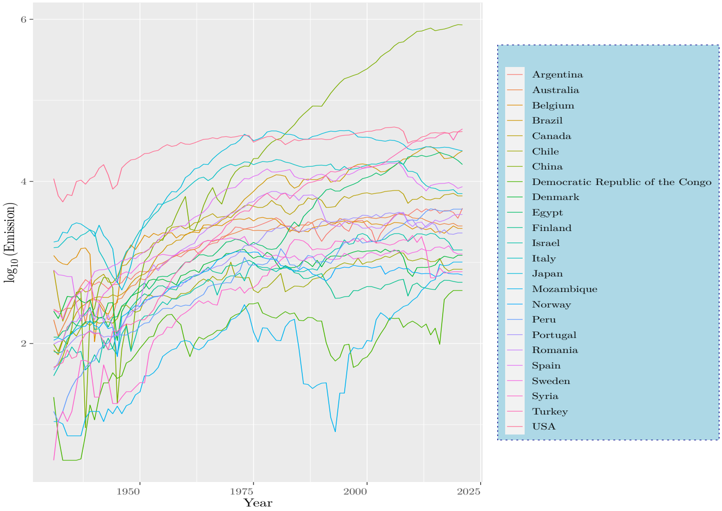
Data Information

Data are collected for 91 years and across 24 different countries.

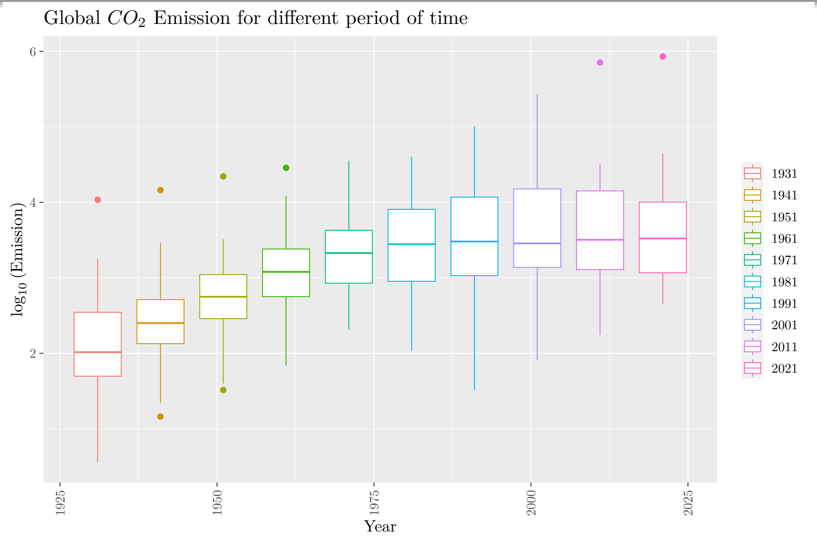
```
## 'data.frame': 91 obs. of  25 variables:
##  $ Year                            : num  1931 1932 1933 1934 1935 ...
##  $ Argentina                       : num  265 247 254 280 356 ...
##  $ Australia                       : num  196 124 160 207 276 ...
##  $ Belgium                         : num  1218 1039 963 938 1087 ...
##  $ Brazil                          : num  83.6 72.7 109 159.9 181.6 ...
##  $ Canada                          : num  800 363 189 273 273 ...
##  $ Chile                           : num  50.9 54.5 69 101.7 141.6 ...
##  $ China                           : num  97.9 79.6 113.2 97.9 156.1 ...
##  $ Democratic Republic of the Congo: num  21.81 7.27 3.63 3.63 3.63 ...
##  $ Denmark                         : num  251 204 273 382 374 ...
##  $ Egypt                           : num  120 120 142 145 189 ...
##  $ Finland                         : num  80 76.2 80 112.7 134.3 ...
##  $ Israel                          : num  40 50.9 65.4 69 90.9 ...
##  $ Italy                           : num  1519 1545 1756 2013 2086 ...
##  $ Japan                           : num  1788 1843 2366 2304 2904 ...
##  $ Mozambique                      : num  10.9 10.9 10.18 7.27 7.27 ...
##  $ Norway                          : num  110 117 111 124 131 ...
##  $ Peru                            : num  14.5 10.9 14.5 21.8 29.1 ...
##  $ Portugal                        : num  47.2 58.1 80 90.9 105.4 ...
##  $ Romania                         : num  98.1 105.4 109 156.3 189 ...
##  $ Spain                           : num  807 705 694 672 669 ...
##  $ Sweden                          : num  258 241 201 287 367 ...
##  $ Syria                           : num  3.63 10.9 14.54 10.9 14.54 ...
```

CO_2 emission for different countries

CO_2 Emission for different countries over different period of time



Decades-long global CO_2 emission



Assumptions

- The paired observations may be designated (X, Y). For comparisons of paired observations (X, Y),

Assumptions

- The paired observations may be designated (X, Y). For comparisons of paired observations (X, Y),
- Define $Z_i = X_i - Y_i, \forall i = 1, .., n,$

Assumptions

- The paired observations may be designated (X, Y). For comparisons of paired observations (X, Y),
- Define $Z_i = X_i - Y_i, \forall i = 1, .., n,$
- The basic assumptions are:
  - For a fixed $p, 0 < p < 1$ the distribution of Z has a common (unknown) quantile, say θ, such that $F_i(\theta) = P(Z_i < \theta) = p, \forall i.$

Assumptions

- The paired observations may be designated (X, Y). For comparisons of paired observations (X, Y),
- Define $Z_i = X_i - Y_i, \forall i = 1, .., n$,
- The basic assumptions are:
  - For a fixed $p, 0 < p < 1$ the distribution of Z has a common (unknown) quantile, say θ, such that $F_i(\theta) = P(Z_i < \theta) = p, \forall i$.
  - F is continuous in the neighborhood of the unknown median of Z, say M, i.e. $P(Z_i < M) = P(Z_i > M) = \frac{1}{2}, P(Z_i = M) = 0$.

Methodology

- Here our

$$H_0 : M = 0 \text{ against } H_A : M \gtreqless 0$$

Methodology

- Here our
$$H_0 : M = 0 \text{ against } H_A : M \gtreqless 0$$
- Define Kas the number of observations for which $z_i > 0$, then under H_0, $K \sim Bin(n, \frac{1}{2})$, ignoring the *zero differences*.

Methodology

- Here our
$$H_0 : M = 0 \text{ against } H_A : M \gtreqless 0$$
- Define Kas the number of observations for which $z_i > 0$, then under H_0, $K \sim Bin(n, \frac{1}{2})$, ignoring the *zero differences*.
- For exact test,
  - Observe the value of Kas K_{obs}

Methodology

- Here our
$$H_0 : M = 0 \text{ against } H_A : M \gtreqless 0$$
- Define K as the number of observations for which $z_i > 0$, then under H_0, $K \sim Bin(n, \frac{1}{2})$, ignoring the *zero differences*.
- For exact test,
  - Observe the value of K as K_{obs}
  - Obtain the *p-value* in the following way

$$\begin{cases} 2\min\left\{\mathbb{P}_{H_0}\left(K \leq K_{obs}\right), \mathbb{P}_{H_0}\left(K \geq K_{obs}\right)\right\} & \text{for } H_A : M \neq 0 \\ \mathbb{P}_{H_0}\left(K \geq K_{obs}\right) & \text{for } H_A : M > 0 \\ \mathbb{P}_{H_0}\left(K \leq K_{obs}\right) & \text{for } H_A : M < 0 \end{cases}$$

Methodology

- Here our
$$H_0 : M = 0 \text{ against } H_A : M \gtreqless 0$$
- Define Kas the number of observations for which $z_i > 0$, then under H_0, $K \sim Bin(n, \frac{1}{2})$, ignoring the *zero differences*.
- For exact test,
  - Observe the value of Kas K_{obs}
  - Obtain the *p-value* in the following way
$$\begin{cases} 2\min\left\{\mathbb{P}_{H_0}\left(K \le K_{obs}\right), \mathbb{P}_{H_0}\left(K \ge K_{obs}\right)\right\} & \text{for } H_A : M \neq 0 \\ \mathbb{P}_{H_0}\left(K \ge K_{obs}\right) & \text{for } H_A : M > 0 \\ \mathbb{P}_{H_0}\left(K \le K_{obs}\right) & \text{for } H_A : M < 0 \end{cases}$$
  - For α level of significance, reject H_0 if *p-value* $\le \alpha$.

Methodology

- Here our
$$H_0 : M = 0 \text{ against } H_A : M \gtreqless 0$$
- Define K as the number of observations for which $z_i > 0$, then under H_0, $K \sim Bin(n, \frac{1}{2})$, ignoring the *zero differences*.
- For exact test,
  - Observe the value of K as K_{obs}
  - Obtain the *p-value* in the following way
$$\begin{cases} 2\min\left\{\mathbb{P}_{H_0}\left(K \le K_{obs}\right), \mathbb{P}_{H_0}\left(K \ge K_{obs}\right)\right\} & \text{for } H_A : M \neq 0 \\ \mathbb{P}_{H_0}\left(K \ge K_{obs}\right) & \text{for } H_A : M > 0 \\ \mathbb{P}_{H_0}\left(K \le K_{obs}\right) & \text{for } H_A : M < 0 \end{cases}$$
  - For α level of significance, reject H_0 if *p-value* $\le \alpha$.
- For the asymptotic test we conclude in the same way as exact test, just with the addition of the assumption of *normality* and *finite population correction.*

Result

For decades-long global CO_2 emission:

We first conduct both sided test for year 1931 and 1941.

```
##
##  One-sample Sign-Test
##
## data:  z
## S = 4, number of differences = 23, p-value = 0.002599
## alternative hypothesis: true median is not equal to 0
## 97.7 percent confidence interval:
##  -145.30   -3.64
## sample estimates:
## median of the differences
##                    -70.85
```

As we can see that *p-value* is less than 0.05, we reject the null hypothesis that CO_2 emission doesn't differ at 5% level of significance.

Result

Moreover, we are interested to know whether CO_2 emission for year 1931 is less than the CO_2 emission for year 1941.

```
##
##  One-sample Sign-Test
##
## data:  z
## S = 4, number of differences = 23, p-value = 0.0013
## alternative hypothesis: true median is less than 0
## 96.8 percent confidence interval:
##    -Inf -18.18
## sample estimates:
## median of the differences
##                    -70.85
```

Here, *p-value* is less than 0.05 and we reject the null hypothesis that CO_2 emission doesn't differ at 5% level of significance. Thus we can conclude that there is a significant increase in CO_2 emission.

Result

The table below shows the decisions for all pairs possible. Note that we have performed exact test here.

	Decision made
1931 vs 1941	There is significant increase in CO_2 emission
1941 vs 1951	There is significant increase in CO_2 emission
1951 vs 1961	There is significant increase in CO_2 emission
1961 vs 1971	There is significant increase in CO_2 emission
1971 vs 1981	There is no significant change in CO_2 emission
1981 vs 1991	There is no significant change in CO_2 emission
1991 vs 2001	There is significant increase in CO_2 emission
2001 vs 2011	There is no significant change in CO_2 emission
2011 vs 2021	There is no significant change in CO_2 emission

Table: Decision using Sign test

Result

For Countries:

We firstly conduct both sided test for *Argentina* and *Australia*.

```
##
##  One-sample Sign-Test
##
## data:  z
## S = 55, number of differences = 91, p-value = 0.05857
## alternative hypothesis: true median is not equal to 0
## 96.5 percent confidence interval:
##   -4.0 148.9
## sample estimates:
## median of the differences
##                      87.2
```

As we can see that *p-value* is greater than 0.05 we fail to reject the null hypothesis that CO_2 emission doesn't differ at 5% level of significance.

Result

The table below shows the decisions for first ten pairs possible. Note that we have performed asymptotic test here.

	Decision made
Argentina vs Australia	There is no significant change in CO_2 emission
Argentina vs Belgium	There is significant increase in CO_2 emission
Argentina vs Brazil	There is significant increase in CO_2 emission
Argentina vs Canada	There is significant increase in CO_2 emission
Argentina vs Chile	There is significant decrease in CO_2 emission
Argentina vs China	There is significant increase in CO_2 emission
Argentina vs Democratic Republic of the Congo	There is significant decrease in CO_2 emission
Argentina vs Denmark	There is significant decrease in CO_2 emission
Argentina vs Egypt	There is no significant change in CO_2 emission
Argentina vs Finland	There is significant decrease in CO_2 emission

Table: Decision from sign test

Assumptions

- $X_1, X_2, ..., X_n$ are random samples from absolutely continuous distribution function $F_X(.)$

Assumptions

- $X_1, X_2, \ldots.., X_n$ are random samples from absolutely continuous distribution function $F_X(.)$
- $Y_1, Y_2, \ldots.., Y_n$ are random samples from absolutely continuous distribution function $F_Y(.)$

Assumptions

- $X_1, X_2, ..., X_n$ are random samples from absolutely continuous distribution function $F_X(.)$
- $Y_1, Y_2, ..., Y_n$ are random samples from absolutely continuous distribution function $F_Y(.)$
- X's and Y's are independent

Methodology

- Here our

$$H_0 : F_Y(x) = F_X(x) \text{ against } H_A : F_Y(x) \gtreqless F_X(x)$$

Methodology

- Here our
$$H_0 : F_Y(x) = F_X(x) \text{ against } H_A : F_Y(x) \gtreqless F_X(x)$$
- Define W as the sum of ranks of the X observations from the combined ordered arrangement of the sample.

Methodology

- Here our

$$H_0 : F_Y(x) = F_X(x) \text{ against } H_A : F_Y(x) \gtreqless F_X(x)$$

- Define W as the sum of ranks of the X observations from the combined ordered arrangement of the sample.
- For exact test,
  - Observe the value of W as W_{obs}

Methodology

- Here our

$$H_0 : F_Y(x) = F_X(x) \text{ against } H_A : F_Y(x) \gtreqless F_X(x)$$

- Define W as the sum of ranks of the X observations from the combined ordered arrangement of the sample.
- For exact test,
  - Observe the value of W as W_{obs}
  - Obtain the *p-value* in the following way

$$\begin{cases} 2\min\{\mathbb{P}_{H_0}(W \le W_{obs}), \mathbb{P}_{H_0}(W \ge W_{obs})\} & \text{for } H_A : F_Y(x) \neq F_X(x) \\ \mathbb{P}_{H_0}(W \ge W_{obs}) & \text{for } H_A : F_Y(x) > F_X(x) \\ \mathbb{P}_{H_0}(W \le W_{obs}) & \text{for } H_A : F_Y(x) < F_X(x) \end{cases}$$

Methodology

- Here our

$$H_0 : F_Y(x) = F_X(x) \text{ against } H_A : F_Y(x) \gtreqless F_X(x)$$

- Define W as the sum of ranks of the X observations from the combined ordered arrangement of the sample.
- For exact test,
    - Observe the value of W as W_{obs}
    - Obtain the *p-value* in the following way

$$\begin{cases} 2\min\left\{\mathbb{P}_{H_0}(W \le W_{obs}), \mathbb{P}_{H_0}(W \ge W_{obs})\right\} & \text{for } H_A : F_Y(x) \ne F_X(x) \\ \mathbb{P}_{H_0}(W \ge W_{obs}) & \text{for } H_A : F_Y(x) > F_X(x) \\ \mathbb{P}_{H_0}(W \le W_{obs}) & \text{for } H_A : F_Y(x) < F_X(x) \end{cases}$$

    - For α level of significance, reject H_0 if *p-value* $\le \alpha$.

Methodology

- Here our

$$H_0 : F_Y(x) = F_X(x) \text{ against } H_A : F_Y(x) \gtreqless F_X(x)$$

- Define W as the sum of ranks of the X observations from the combined ordered arrangement of the sample.
- For exact test,
  - Observe the value of W as W_{obs}
  - Obtain the *p-value* in the following way

$$\begin{cases} 2\min\left\{\mathbb{P}_{H_0}(W \leq W_{obs}), \mathbb{P}_{H_0}(W \geq W_{obs})\right\} & \text{for } H_A : F_Y(x) \neq F_X(x) \\ \mathbb{P}_{H_0}(W \geq W_{obs}) & \text{for } H_A : F_Y(x) > F_X(x) \\ \mathbb{P}_{H_0}(W \leq W_{obs}) & \text{for } H_A : F_Y(x) < F_X(x) \end{cases}$$

  - For α level of significance, reject H_0 if *p-value* $\leq \alpha$.
- For the asymptotic test we conclude in the same way as exact test, just with the addition of the assumption of *normality* and *finite population correction.*

Result

For decades-long global CO_2 emission:

Here we have shown the output for the years 1931 and 1941.

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  x and y
## W = 217, p-value = 0.146
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -230.50006   40.00004
## sample estimates:
## difference in location
##              -83.67198
```

As the $p-value$ is greater than 0.05, we fail to reject H_0 at 5% level of significance.

Result

The table below shows the results for all decades.

	Decision made
1931 vs 1941	There is no significant change in CO_2 emission
1941 vs 1951	There is significant increase in CO_2 emission
1951 vs 1961	There is significant increase in CO_2 emission
1961 vs 1971	There is no significant change in CO_2 emission
1971 vs 1981	There is no significant change in CO_2 emission
1981 vs 1991	There is no significant change in CO_2 emission
1991 vs 2001	There is no significant change in CO_2 emission
2001 vs 2011	There is no significant change in CO_2 emission
2011 vs 2021	There is no significant change in CO_2 emission

Table: Decision values using Wilcoxon Rank-Sum test

Result

For Countries:

Here we give a snapshot for *Australia* and *Argentina*.

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  x and y
## W = 4277.5, p-value = 0.7009
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -308.0  399.7
## sample estimates:
## difference in location
##               72.70007
```

The $p-value$ for the both sided test is greater than 0.05. So we can accept H_0 that there is indeed no significant difference between the CO_2 emission of Australia and Argentina, at 5% level of significant.

Result

The below table shows the result for some pairs of countries.

	Decision made
Argentina vs Australia	There is no significant change in CO_2 emission
Argentina vs Belgium	There is no significant change in CO_2 emission
Argentina vs Brazil	There is significant increase in CO_2 emission
Argentina vs Canada	There is significant increase in CO_2 emission
Argentina vs Chile	There is significant decrease in CO_2 emission
Argentina vs China	There is significant increase in CO_2 emission
Argentina vs Democratic Republic of the Congo	There is significant decrease in CO_2 emission
Argentina vs Denmark	There is significant decrease in CO_2 emission
Argentina vs Egypt	There is no significant change in CO_2 emission
Argentina vs Finland	There is significant decrease in CO_2 emission

Table: Decision values using Wilcoxon Rank-Sum test

Assumptions

- The (unknown) medians (or means) of the two samples are equal or the sample observations can be adjusted to have equal locations

Assumptions

- The (unknown) medians (or means) of the two samples are equal or the sample observations can be adjusted to have equal locations
- $X_1, X_2, \ldots.., X_n$ are random samples from absolutely continuous distribution function $F_X(.)$

Assumptions

- The (unknown) medians (or means) of the two samples are equal or the sample observations can be adjusted to have equal locations
- $X_1, X_2, \ldots.., X_n$ are random samples from absolutely continuous distribution function $F_X(.)$
- $Y_1, Y_2, \ldots.., Y_n$ are random samples from absolutely continuous distribution function $F_Y(.)$

Assumptions

- The (unknown) medians (or means) of the two samples are equal or the sample observations can be adjusted to have equal locations
- $X_1, X_2, \ldots.., X_n$ are random samples from absolutely continuous distribution function $F_X(.)$
- $Y_1, Y_2, \ldots.., Y_n$ are random samples from absolutely continuous distribution function $F_Y(.)$
- X's and Y's are independent

Methodology

- Here our

$$H_0 : \ F_Y \text{ and } F_X \text{ have equal dispersion against}$$

$$H_A : \ \text{dispersion of } Y \gtreqless \text{ dispersion of } X$$

Methodology

- Here our

$$H_0 : \ F_Y \text{ and } F_X \text{ have equal dispersion against}$$

$$H_A : \ \text{dispersion of } Y \gtreqless \text{ dispersion of } X$$

- Here a linear rank statistic M is defined as, $M = \sum_{i=1}^{N} \left(i - \frac{N+1}{2}\right)^2 Z_i$,where

  $Z_i = 1$,if the i th observation of the ordered combined sample is an X observation and $Z_i = 0$,otherwise.

Methodology

- Here our
$$H_0 : \; F_Y \text{ and } F_X \text{ have equal dispersion against}$$
$$H_A : \; \text{dispersion of } Y \gtreqless \text{ dispersion of } X$$
- Here a linear rank statistic M is defined as, $M = \sum_{i=1}^{N} \left(i - \frac{N+1}{2}\right)^2 Z_i$,where
  $Z_i = 1$,if the i th observation of the ordered combined sample is an X observation and $Z_i = 0$,otherwise.
- For exact test,
  - Observe the value of Mas M_{obs}

Methodology

- Here our
$$H_0: \ F_Y \text{ and } F_X \text{ have equal dispersion against}$$
$$H_A: \ \text{dispersion of } Y \gtreqless\!\!\!\!/ \ \text{dispersion of } X$$

- Here a linear rank statistic M is defined as, $M = \sum_{i=1}^{N} \left(i - \frac{N+1}{2}\right)^2 Z_i$,where

  $Z_i = 1$,if the i th observation of the ordered combined sample is an X observation and $Z_i = 0$,otherwise.

- For exact test,
  - Observe the value of Mas M_{obs}
  - Obtain the *p-value* in the following way

$$\begin{cases} 2\min\left\{\mathbb{P}_{H_0}\left(M \le M_{obs}\right), \mathbb{P}_{H_0}\left(M \ge M_{obs}\right)\right\} & \text{for } H_A: \text{dispersion of } Y \neq \text{dispersion of } X \\ \mathbb{P}_{H_0}\left(M \ge M_{obs}\right) & \text{for } H_A: \text{dispersion of } Y < \text{dispersion of } X \\ \mathbb{P}_{H_0}\left(M \le M_{obs}\right) & \text{for } H_A: \text{disoersion of } Y > \text{dispersion of } X \end{cases}$$

Methodology

- Here our
$$H_0 : \ F_Y \text{ and } F_X \text{ have equal dispersion against}$$
$$H_A : \ \text{dispersion of } Y \gtrless \text{ dispersion of } X$$

- Here a linear rank statistic M is defined as, $M = \sum_{i=1}^{N} \left(i - \frac{N+1}{2} \right)^2 Z_i$,where

  $Z_i = 1$,if the i th observation of the ordered combined sample is an X observation and $Z_i = 0$,otherwise.

- For exact test,
  - Observe the value of Mas M_{obs}
  - Obtain the *p-value* in the following way

$$\begin{cases} 2\min\left\{\mathbb{P}_{H_0}\left(M \leq M_{obs}\right), \mathbb{P}_{H_0}\left(M \geq M_{obs}\right)\right\} & \text{for } H_A : \text{dispersion of } Y \neq \text{dispersion of } X \\ \mathbb{P}_{H_0}\left(M \geq M_{obs}\right) & \text{for } H_A : \text{dispersion of } Y < \text{dispersion of } X \\ \mathbb{P}_{H_0}\left(M \leq M_{obs}\right) & \text{for } H_A : \text{disoersion of } Y > \text{dispersion of } X \end{cases}$$

  - For α level of significance, reject H_0 if *p-value* $\leq \alpha$.

Methodology

- Here our
$$H_0: \ F_Y \text{ and } F_X \text{ have equal dispersion against}$$
$$H_A: \ \text{dispersion of } Y \gtreqless\!\!\!\!/ \ \text{dispersion of } X$$

- Here a linear rank statistic M is defined as, $M = \sum_{i=1}^{N} \left(i - \frac{N+1}{2} \right)^2 Z_i$,where

  $Z_i = 1$,if the i th observation of the ordered combined sample is an X observation and $Z_i = 0$,otherwise.

- For exact test,
  - Observe the value of Mas M_{obs}
  - Obtain the *p-value* in the following way

$$\begin{cases} 2\min\left\{ \mathbb{P}_{H_0}\left(M \leq M_{obs}\right), \mathbb{P}_{H_0}\left(M \geq M_{obs}\right) \right\} & \text{for } H_A: \text{dispersion of } Y \neq \text{dispersion of } X \\ \mathbb{P}_{H_0}\left(M \geq M_{obs}\right) & \text{for } H_A: \text{dispersion of } Y < \text{dispersion of } X \\ \mathbb{P}_{H_0}\left(M \leq M_{obs}\right) & \text{for } H_A: \text{disoersion of } Y > \text{dispersion of } X \end{cases}$$

  - For α level of significance, reject H_0 if *p-value* $\leq \alpha$.

- For the asymptotic test we conclude in the same way as exact test, just with the addition of the assumption of *normality* and *finite population correction*.

Result

For **decades-long global CO_2 emission:**

From the results of Wilcoxon Rank,-Sum only for four pairs (1971,1981),(1981,1991); (2001,2011); (2011,2021) we have performed Mood's Test. First the result for the years 1971 and 1981 are shown below:

```
##
##  Mood two-sample test of scale
##
## data:  x and y
## Z = 1.1714, p-value = 0.2414
## alternative hypothesis: two.sided
```

The $p - value$ is much larger than 0.05 and hence we accept the null hypothesis at 5% level of significance and conclude that there is no significant difference between the variability of the year 1971 and 1981.

Result

The results for all some pairs are presented here.

	Decision made
1961 vs 1971	There is no significant change in CO_2 emission
1971 vs 1981	There is no significant change in CO_2 emission
1981 vs 1991	There is no significant change in CO_2 emission
1991 vs 2001	There is no significant change in CO_2 emission
2001 vs 2011	There is no significant change in CO_2 emission
2011 vs 2021	There is no significant change in CO_2 emission

Table: Decisions using Mood's test

Result

For Countries:

From the results of Wilcoxon Rank-Sum, 35 pairs of countries have no significant difference in CO_2 emission like, (Australia, Argentina); (Belgium. Romania) etc.
We have shown the result for Australia and Argentina.

```
##
##  Mood two-sample test of scale
##
## data:  x and y
## Z = -0.16643, p-value = 0.8678
## alternative hypothesis: two.sided
```

As the $p - value$ is much larger than 0.05, so we conclude that there is no significant difference in variability of Australia and Argentina.

Result

Also the results for some pairs are presented,

	Decision made
Argentina vs Australia	There is no significant change in CO_2 emission
Argentina vs Belgium	There is significant decrease in CO_2 emission
Argentina vs Egypt	There is significant increase in CO_2 emission
Argentina vs Romania	There is significant increase in CO_2 emission
Australia vs Belgium	There is significant decrease in CO_2 emission
Australia vs Egypt	There is significant increase in CO_2 emission
Australia vs Romania	There is significant increase in CO_2 emission
Belgium vs Egypt	There is significant increase in CO_2 emission
Belgium vs Romania	There is significant increase in CO_2 emission
Brazil vs Spain	There is significant decrease in CO_2 emission
Brazil vs Turkey	There is significant increase in CO_2 emission
Canada vs Egypt	There is significant increase in CO_2 emission
Canada vs Turkey	There is significant increase in CO_2 emission
Chile vs Finland	There is significant decrease in CO_2 emission
Chile vs Israel	There is significant increase in CO_2 emission

Assumption

- All k sets of independent samples are drawn from absolutely continuous distribution function $F_i\,(.)\ \forall i = 1(1)k$.

Assumption

- All k sets of independent samples are drawn from absolutely continuous distribution function $F_i\,(.)\ \forall i = 1(1)k$.
- In the location alternative setup the populations are identical in every respect, except the location parameter.

Methodology

- Here our

$$H_0 : \ F_1(x) = F_2(x) = ... = F_k(x) \ \text{ against } H_A \text{ :not } H_0$$

Methodology

- Here our

$$H_0: \ F_1(x) = F_2(x) = ... = F_k(x) \ \text{ against } H_A \text{ :not } H_0$$

- Rank all data from all groups together. Assign any tied values the average of the ranks they would have received had they not been tied.

Methodology

- Here our

$$H_0: \ F_1(x) = F_2(x) = \ldots = F_k(x) \ \text{ against } H_A \text{ :not } H_0$$

- Rank all data from all groups together. Assign any tied values the average of the ranks they would have received had they not been tied.
- The test statistic is given by the equation below

$$H = \frac{12}{N(N-1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

Where N is the total sample size, k is the number of groups we are comparing, R_i is the sum of ranks for group i , and n_i is the sample size of group i .

Methodology

- For exact test,
  - Observe the value of Has H_{obs}

Methodology

- For exact test,
  - Observe the value of Has H_{obs}
  - Obtain the *p-value* in the following way

$$\mathbb{P}_{H_0}\left(H \geq H_{obs}\right) \text{ for} H_A \text{ :not } H_0$$

Methodology

- For exact test,
  - Observe the value of Has H_{obs}
  - Obtain the *p-value* in the following way

$$\mathbb{P}_{H_0}\left(H \geq H_{obs}\right) \text{ for} H_A \text{ :not } H_0$$

  - For α level of significance, reject H_0 if *p-value* $\leq \alpha$.

Methodology

- For exact test,
  - Observe the value of Has H_{obs}
  - Obtain the *p-value* in the following way

$$\mathbb{P}_{H_0}\left(H \geq H_{obs}\right) \text{ for} H_A \text{ :not } H_0$$

  - For α level of significance, reject H_0 if *p-value* $\leq \alpha$.
- Asymptotically H follows χ^2_{k-1} distribution. So, we perform the asymptotic test accordingly.

Result

First we found the sample standard deviation. Then we have performed Kruskal-Wallis test for those group (size>2), for which the standard deviations are close enough. The countries Argentina, Peru, Portugal have close standard deviation .

	Standard Deviation
Argentina	1327.8789
Australia	1245.9507
Belgium	885.7870
Brazil	7882.2382
Canada	2393.8287
Chile	343.2244
China	266589.4357
Democratic Republic of the Congo	104.2538
Denmark	392.7684
Egypt	7270.9458
Finland	251.4007
Israel	762.6959

Result

	Standard Deviation
Italy	6102.5108
Japan	14658.8081
Mozambique	246.2908
Norway	352.0404
Peru	1365.9741
Portugal	1368.9408
Romania	2203.5197
Spain	5345.5497
Sweden	448.3529
Syria	736.1023
Turkey	12109.3962
USA	10951.2528

Result

For decades we haven't found any set of interest as the standard deviations are not that much close.

	Standard Deviation
1931	2194.929
1941	2939.747
1951	4412.352
1961	6129.387
1971	9117.920
1981	11615.976
1991	21798.270
2001	54726.254
2011	143328.878
2021	172785.052

Result

The $p - value$ of this test for these 3 countries is very smaller than 0.05, so we reject H_0 at 5% level of significance and conclude that there is significant difference in CO_2 emission between Argentina, Peru, Portugal.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  dat_gather_2$Emission and as.factor(dat_gather_2$Count
## Kruskal-Wallis chi-squared = 23.74, df = 2, p-value = 6.997e-
```

Methodology

- Let $(X_1, Y_1), (X_2, Y_2), ...(X_n, Y_n)$, be a set of observations from the joint distribution function $F_{(X,Y)}$

Methodology

- Let $(X_1, Y_1), (X_2, Y_2), ...(X_n, Y_n)$, be a set of observations from the joint distribution function $F_{(X,Y)}$
- We test for the null hypothesis $H_0 : F_{X,Y}(x, y) = F_X(x)F_Y(y)$ against H_A :Not H_0.

Methodology

- Let $(X_1, Y_1), (X_2, Y_2), ...(X_n, Y_n)$, be a set of observations from the joint distribution function $F_{(X,Y)}$
- We test for the null hypothesis $H_0 : F_{X,Y}(x, y) = F_X(x)F_Y(y)$ against H_A :Not H_0.
- Let, π_c,be the probability of concordance and π_d,be the probability of discordance.

Methodology

- Let $(X_1, Y_1), (X_2, Y_2), ...(X_n, Y_n)$, be a set of observations from the joint distribution function $F_{(X,Y)}$
- We test for the null hypothesis $H_0 : F_{X,Y}(x, y) = F_X(x)F_Y(y)$ against H_A :Not H_0.
- Let, π_c,be the probability of concordance and π_d,be the probability of discordance.
- Then **Kendall's** τ is defined as; $\tau = \pi_c - \pi_d$.

Methodology

- For a sample of size N,the sample estimate of τ is defined as:

$$T = \frac{\sum\sum_{1 \le i < j \le N} A_{ij}}{\binom{N}{2}},$$

,where $A_{ij} = sign(X_j - X_i)sign(Y_j - Y_i)$ which to equivalent to saying

$$A_{ij} = \begin{cases} 1 & \text{, if pairs are concordant with probability } \pi_c \\ 0 & \text{,if atleast one component is tied with probability } \pi_t\text{=1-}\pi_c\text{-}\pi_d \\ -1 & \text{,if pairs are discordant with probability } \pi_d \end{cases}$$

Methodology

- For exact test,
  - Observe the value of T as T_{obs}

Methodology

- For exact test,
  - Observe the value of T as T_{obs}
  - Obtain the *p-value* in the following way

$$\begin{cases} 2\min\left\{\mathbb{P}_{H_0}\left(T \leq T_{obs}\right), \mathbb{P}_{H_0}\left(T \geq T_{obs}\right)\right\} & \text{for } H_A : \text{there is dependence between two popu} \\ \mathbb{P}_{H_0}\left(T \geq T_{obs}\right) & \text{for } H_A : \text{there is positive association} \\ \mathbb{P}_{H_0}\left(T \leq T_{obs}\right) & \text{for } H_A : \text{there is negative association} \end{cases}$$

Methodology

- For exact test,
  - Observe the value of T as T_{obs}
  - Obtain the *p-value* in the following way

$$\begin{cases} 2\min\left\{\mathbb{P}_{H_0}\left(T \le T_{obs}\right), \mathbb{P}_{H_0}\left(T \ge T_{obs}\right)\right\} & \text{for } H_A : \text{there is dependence between two popu} \\ \mathbb{P}_{H_0}\left(T \ge T_{obs}\right) & \text{for } H_A : \text{there is positive association} \\ \mathbb{P}_{H_0}\left(T \le T_{obs}\right) & \text{for } H_A : \text{there is negative association} \end{cases}$$

  - For α level of significance, reject H_0 if *p-value* $\le \alpha$.

Methodology

- For exact test,
  - Observe the value of T as T_{obs}
  - Obtain the *p-value* in the following way

$$\begin{cases} 2\min\left\{\mathbb{P}_{H_0}\left(T \leq T_{obs}\right), \mathbb{P}_{H_0}\left(T \geq T_{obs}\right)\right\} & \text{for } H_A : \text{there is dependence between two popu} \\ \mathbb{P}_{H_0}\left(T \geq T_{obs}\right) & \text{for } H_A : \text{there is positive association} \\ \mathbb{P}_{H_0}\left(T \leq T_{obs}\right) & \text{for } H_A : \text{there is negative association} \end{cases}$$

  - For α level of significance, reject H_0 if *p-value* $\leq \alpha$.
- For the asymptotic test we conclude in the same way as exact test, just with the addition of the assumption of *normality* and *finite population correction.*

Result

For Countries: Here we give a snapshot for *Australia* and *Argentina*.

```
##
##  Kendall's rank correlation tau
##
## data:  x and y
## z = 9.661, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##       tau
## 0.6884085
##
##  Kendall's rank correlation tau
##
## data:  x and y
## z = 9.661, p-value < 2.2e-16
## alternative hypothesis: true tau is greater than 0
```

Result

The $p-value$ for the both sided test is less than 0.05. So we can reject H_0 and also for greater than test $p-value$ is less than 0.05 which implies H_0 is rejected also and thus we can conclude that there is indeed positive association between the CO_2 emission of Australia and Argentina, at 5% level of significant.

Results

The below table shows the result for some pairs of countries.

	Decision made
Argentina vs Australia	There is positive association
Argentina vs Belgium	There is positive association
Argentina vs Brazil	There is positive association
Argentina vs Canada	There is positive association
Argentina vs Chile	There is positive association
Argentina vs China	There is positive association
Argentina vs Democratic Republic of the Congo	There is positive association
Argentina vs Denmark	There is positive association
Argentina vs Egypt	There is positive association
Argentina vs Finland	There is positive association

Table: Decisions from Kendall Rank Correlation Coefficient

Result

Notice that there is only positive association present in all the pairs.

```
## Decision made
## There is positive association
##                              276
```

Thus, we can conclude that CO_2 emission is moving in same direction for all the countries.

Thank You