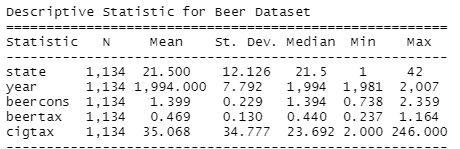
**Question 1:**

As from the figure below, a typical state, on average, has around 1.399 gallons of beer sold per capita. Moreover, the average beer tax in a typical state is $0.469 per gallon and the average cigarette tax in a typical observation is $35,068 per pack. For beer sales, the min (1,981) and the max (2,007) aren’t too far away, and the mean (1.399) and median (1.394) are relatively close which possibly indicates a symmetrical distribution. In contrast, beer tax has a mean (0.469) that is greater than the median (0.440), which might possibly indicate right skew. Also, the range (0.927) between min (0.237) and max value (1.164) is quite considerable, hence there is more extreme value on the right of the distribution of beer tax. For cigarette tax, the mean (35.068) is also greater than the median (34.777). It’s possible that the distribution of cigarette tax is right skew. Also, the min (2) is closer to the central tendency than the max (246) which shows that there are more extreme values toward the right tail of the distribution.



**Question 2:**

For beer sales, the 95% CI is:

For beer tax, the 95% CI is:

For cigarette tax, the 95% CI is

**Question 3:**

Chart, histogram

Description automatically generatedFrom the graph above, we can see that the mean for beer sales when high tax equal 0 is greater than the mean for beer sales when high tax equal 1. We can also see that the density for beer sales when high tax equal 1 is lower than when high tax equal 0. A potential reason for this might be due too as beer tax is greater than its median, less people purchase beer since it’s cost more than when beer tax is lower than its median. This implies that as beer tax get higher, we see a drop in beer sales.

**Question 4:**

At the 5% significance level, we do not have enough evidence to support the null hypothesis that the mean of beer sales when hightax = 1 is the same as the mean of beer sales when hightax = 0. Hence, we reject the null hypothesis and there’s a statistically significance difference. This is evident since there is a 4.23% decrease in beer sales when going from hightax = 0 to hightax = 1. Therefore, it may implies that as beer tax increase, we see a drop in beer sales.

**Question 5:**

Chart, scatter chart

Description automatically generated

The scatter plot above shows the relationship between beer tax and beer sales. From the graph, it’s visible that the relationship is a relatively weak negative relationship as the dots are crowded around the left side of the graph and there are less dots toward the right side.

Furthermore, the correlation coefficient between beer tax and beer sales is:

As we can see, the correlation coefficient between the 2 variables is negative. This supports the findings from question 3 and 4 that as beer tax increase, we see a less amount of beer sales.

**Question 6:**

|  |  |  |
| --- | --- | --- |
| **Regression 1** | | |
| Coefficients | Estimate | Standard error |
| Intercept | 1.54148 | 0.02504 |
| Beer tax | -0.30387 | 0.05151 |
| **Regression 2** | | |
| Coefficients | Estimate | Standard error |
| Intercept | 1.4429630 | 0.0094745 |
| Cigarette tax | -0.0012506 | 0.0001919 |

For regression 1, if there is a change in beer tax of one-standard-deviation, there will be a decrease in beer sales of 0.304 gallons per capita which is approximately 1/5 of the beer sales when there is no beer tax. For regression 2, if there is a change in cigarette tax, there will be a decrease in beer sales of 0.001 gallons per capita. This implies that as cigarette tax goes up, we expect to see a very small change in beer sales.

For regression 1:

Hence, we reject the null hypothesis as p-value is smaller than the significance level of 5%. This implies that the predicted change in beer sales when there is a change in beer tax is statistically significantly different from 0.

For regression 2:

Hence, we reject the null hypothesis as p-value is smaller than the significance level of 5%. This implies that the predicted change in beer sales when there is a change in cigarette tax is statistically significantly different from 0.

**Chart, scatter chart

Description automatically generatedQuestion 7:**

**Chart, scatter chart

Description automatically generated**

There is a decrease in the coefficient estimate on beer sales in regression 3 compared with regression 1. This is because, in regression 1, cigarette tax was an omitted variable and was part of the residuals. However, from the scatter plot between beer tax and cigarette tax, we can see that there is a negative relationship which lead to our first assumptions of the OLS being inaccurate and the coefficient estimator on beer sales was higher than the actual value. When we included cigarette tax as part of our linear regression, the coefficient estimator on beer sales decreases but maintain the same direction. The direction remains the same is due to the fact that cigarette tax and beer sales have a negative relationship.

**Question 8:**

Regression before and up to 1994:

Text, letter

Description automatically generated

Regression after 1994:

Text, letter

Description automatically generated

As we can see from the 2 figures, there is a drastic change in the coefficient estimator on beer sales between the 2 subsamples. At a 5% significance level, for the regression model before 1995, we reject the null hypothesis that the coefficient estimator on beer sales is equal to 0 since its p-value is 3.97e-12 (less than 0.05). Hence, it is statistically significantly different from 0. In other words, before 1995, there is a decent drop in beer sales if beer tax increases.

However, for the regression model after 1994, we failed to reject the null hypothesis since its p-value is 0.4513 (greater than 0.05). Thus, it is statistically insignificantly different 0. In other words, from 1995, there is little to no difference to beer sales even if beer tax increases.