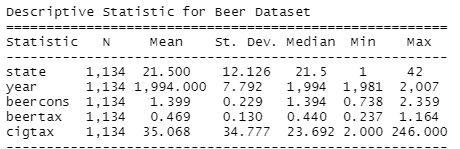
|  |  |
| --- | --- |
| **Name** | **Student ID Number** |
| Do Nhat Anh Ha | 1194034 |

**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 gallons per capita**) and the max (**2,007 gallons per capita**) aren’t too far away, and the mean (**1.399 gallons per capita**) and median (1.394 gallons per capita) are relatively close which possibly indicates a symmetrical distribution. In contrast, beer tax has a mean ($0.469 per gallon) that is greater than the median ($0.440 per gallon), which might possibly indicate right skew. Also, the range ($0.927 per gallon) between min ($0.237 per gallon) and max value ($1.164 per gallon) 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 per pack) is also greater than the median ($34.777 per pack). It’s possible that the distribution of cigarette tax is right skew. Also, the min ($2 per pack) is closer to the central tendency than the max ($246 per pack) which shows that there are more extreme values toward the right tail of the distribution.



**Question 2:**

For beer sales (gallons per capita), the 95% CI is:

For beer tax (dollars per gallon), the 95% CI is:

For cigarette tax (dollars per pack), the 95% CI is

**Question 3:**

Chart, histogram

Description automatically generatedFrom the graph above, we can see that the mean for beer sales when beer tax < $0.440 per gallon is greater than the mean for beer sales when beer tax >= $0.440 per gallon. We can also see that the probability mass for beer sales when beer tax < $0.440 per gallon is greater than the probability mass for beer sales when beer tax >= $0.440 per gallon towards the right. A potential reason for this might be due too as beer tax is greater than $0.440 per gallon, people are less willing to purchase the same amount of beer as it is more expensive than when the tax is below $0.440 per gallon.

**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 beer tax is greater than $0.440 per gallon is the same as the mean of beer sales when beer tax is lower than $0.440 per gallon. Hence, we reject the null hypothesis and there’s a statistically significantly 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 imply 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 of one-standard-deviation in beer tax, there will be a corresponding decrease of 0.304 \* 0.130 = 0.039 gallons per capita in beer sales. Similarly, for regression 2, if there is a change of one-standard-deviation in cigarette tax, there will be a corresponding decrease of 0.0013 \* 34.777 = 0.0345 gallons per capita 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:

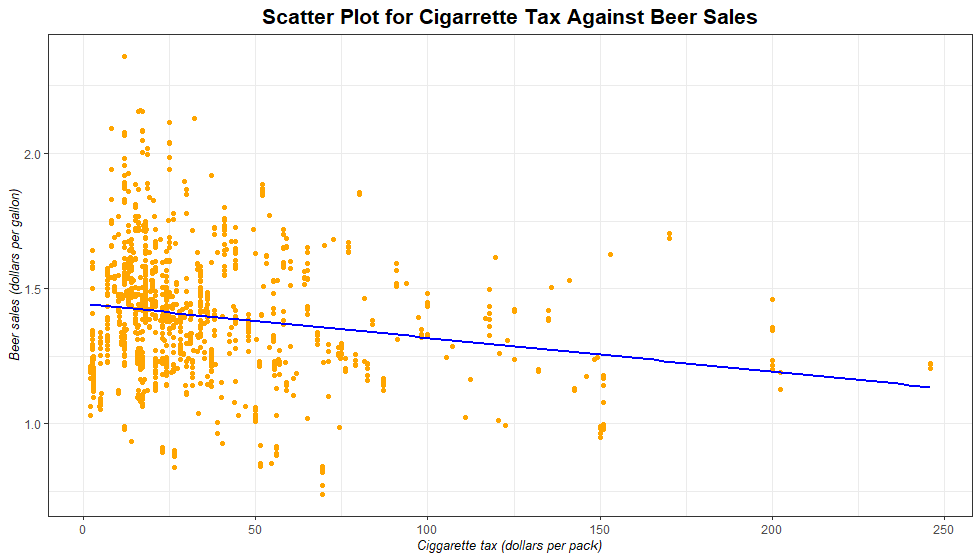
Hence, we reject the null hypothesis as p-value is smaller than the significance level of 5%. This implies that the predicted change in beer sales when there is a change in the cigarette tax is statistically significantly different from 0.

**Chart, scatter chart

Description automatically generatedQuestion 7:**

**Chart, scatter chart

Description automatically generated**



There is an increase in the coefficient estimate magnitude of 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 leads to our first assumptions of the OLS being inaccurate and the coefficient estimator on beer sales being higher than the actual value. Also, cigarette tax has a negative relationship with beer sales as well. Hence, the sign of the correlation between beer tax and the error terms is:

As a result, we see an increase in magnitude in the coefficient estimate of beer sales. The direction of the coefficient estimate stays the same.

**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 and up to 1994, 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 and up to 1994, 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 significantly indifferent from 0. In other words, after 1994, there is little to no difference to beer sales even if beer tax increases.

**Appendix**

# Load the dataset into the environment and neccessary libraries

library('stargazer')

library('ggplot2')

library('dplyr')

# Question 1

# Dataset summary: contains annaual information on beer sales, beer taxes, and cigarret taxes from 1981 - 2007 in US (47 states)

beer\_data = read.csv("D:/Semester 2 2022/ECOM20001/Assignment 1/as1\_beer.csv")

# Report summary statistic

tax\_st = stargazer(beer\_data,

summary.stat = c('n', 'mean', 'sd', 'median', 'min', 'max'),

type = 'text', title = 'Descriptive Statistic for Beer Dataset',

out = 'beer\_sumstat.png')

# Question 2

# Compute sample means for each variable

beercons\_mean = mean(beer\_data$beercons)

beertax\_mean = mean(beer\_data$beertax)

cigtax\_mean = mean(beer\_data$cigtax)

# Compute sample standard error for each variable

beercons\_se = sd(beer\_data$beercons)/sqrt(length(beer\_data$beercons))

beertax\_se = sd(beer\_data$beertax)/sqrt(length(beer\_data$beertax))

cigtax\_se = sd(beer\_data$cigtax)/sqrt(length(beer\_data$cigtax))

# Compute 95% Confidence Interval for each variable

# Beercons

beercons\_95\_low = beercons\_mean - 1.96 \* beercons\_se

beercons\_95\_high = beercons\_mean + 1.96 \* beercons\_se

# Beertax

beertax\_95\_low = beertax\_mean - 1.96 \* beertax\_se

beertax\_95\_high = beertax\_mean + 1.96 \* beertax\_se

# Cigtax

cigtax\_95\_low = cigtax\_mean - 1.96 \* cigtax\_se

cigtax\_95\_high = cigtax\_mean + 1.96 \* cigtax\_se

# Question 3

# adding a column of hightax straight into the dataset using dplyr.

beer\_data2 = beer\_data %>%

mutate(hightax = if\_else(beertax >= median(beertax),

1,

0))

beer\_density = ggplot(beer\_data2, aes(x = beercons, fill = as.character(hightax))) +

geom\_density(alpha = 0.25)

beer\_density + labs(title = "Density Plot for Beer Sales",

x = "Beer sales (gallons per capita)",

y = "Density",

fill = "") +

theme\_bw() +

scale\_fill\_manual(labels = c("Beer tax < $0.440", "Beer tax >= $0.440"), values = c("Blue", "Red")) +

theme(plot.title = element\_text(face = "bold", size = 16, hjust = 0.5),

axis.title = element\_text(face = "italic", size = 10))

beer\_density

# Question 4

# Compute mean for beercons if hightax = 1 and hightax = 0

beercons\_ht1\_mean = mean(beer\_data2$beercons[beer\_data2$hightax == 1])

beercons\_ht0\_mean = mean(beer\_data2$beercons[beer\_data2$hightax == 0])

# Compute standard deviation and standard errors

beercons\_ht1\_sd = sd(beer\_data2$beercons[beer\_data2$hightax == 1])

beercons\_ht0\_sd = sd(beer\_data2$beercons[beer\_data2$hightax == 0])

beercons\_ht1\_se = beercons\_ht1\_sd/sqrt(length(beer\_data2$beercons[beer\_data2$hightax == 1]))

beercons\_ht0\_se = beercons\_ht0\_sd/sqrt(length(beer\_data2$beercons[beer\_data2$hightax == 0]))

mean\_diff = beercons\_ht1\_mean - beercons\_ht0\_mean

se\_diff = sqrt((beercons\_ht1\_se\*\*2 + beercons\_ht0\_se\*\*2))

# Conduct hypothesis testing:

t\_stat = (mean\_diff - 0)/se\_diff

p\_value = 2 \* pnorm(-abs(t\_stat))

# 95% CI for mean diff

mean\_diff\_upper = mean\_diff + 1.96 \* se\_diff

mean\_diff\_lower = mean\_diff - 1.96 \* se\_diff

diff\_percentage = (abs(mean\_diff)/beercons\_ht0\_mean) \* 100

# There is a decrease of approximately 4% when hightax went from 0 to 1.

# Question 5

#scatter plot of beertax against beercons

beer\_scatter = ggplot(beer\_data2, aes(beertax, beercons)) + geom\_point(colour = 'orange')

beer\_scatter + labs(title = "Scatter Plot for Beer Tax Against Beer Sales",

x = "Beer tax (dollars per gallon)",

y = "Beer sales (gallons per capita)") +

theme\_bw() +

theme(plot.title = element\_text(face = "bold", size = 16, hjust = 0.5),

axis.title = element\_text(face = "italic", size = 10)) +

geom\_smooth(method = lm, se = FALSE, colour = "blue")

cor\_beertax\_cons = cor(beer\_data2$beertax, beer\_data2$beercons)

# Question 6

# Single linear regression for beertax and beercons

reg\_1 = lm(beercons~beertax, beer\_data2)

summary(reg\_1)

# Single linear regressions for cigtax and beercons,

reg\_2 = lm(beercons~cigtax, beer\_data2)

summary(reg\_2)

# standard deviation for beertax and cigtax

beertax\_sd = sd(beer\_data2$beertax)

cigtax\_sd = sd(beer\_data2$cigtax)

# Increase of 1 standard-deviation in beertax

change\_1 = -0.30387 \* beertax\_sd

change\_1

# increase of 1-standard-deviation in cigtax

change\_2 = -0.0012506 \* cigtax\_sd

change\_2

# Question 7

# scatter plot for beercons and cigtax

cons\_cig\_scatter = ggplot(beer\_data2, aes(cigtax, beercons)) + geom\_point(colour = 'orange')

cons\_cig\_scatter = cons\_cig\_scatter + labs(title = "Scatter Plot for Cigarrette Tax Against Beer Sales",

y = "Beer sales (dollars per gallon)",

x = "Ciggarette tax (dollars per pack)") +

theme\_bw() +

theme(plot.title = element\_text(face = "bold", size = 16, hjust = 0.5),

axis.title = element\_text(face = "italic", size = 10)) +

geom\_smooth(method = lm, se = FALSE, colour = "blue")

cons\_cig\_scatter

# scatter plot for beertax and cigtax

beer\_cig\_scatter = ggplot(beer\_data2, aes(beertax, cigtax)) + geom\_point(colour = "orange")

beer\_cig\_scatter = beer\_cig\_scatter + labs(title = "Scatter Plot for Beer Tax Against Cigarrette Tax",

x = "Beer tax (dollars per gallon)",

y = "Cigarrette tax (dollars per pack)") +

theme\_bw() +

theme(plot.title = element\_text(face = "bold", size = 16, hjust = 0.5),

axis.title = element\_text(face = "italic", size = 10)) +

geom\_smooth(method = lm, se = FALSE, colour = "blue")

beer\_cig\_scatter

# QUestion 8

reg\_before\_1994 = lm(beercons[year <= 1994]~beertax[year <= 1994], beer\_data2)

summary(reg\_before\_1994)

reg\_after\_1994 = lm(beercons[year > 1994]~beertax[year > 1994], beer\_data2)

summary(reg\_after\_1994)

# Hypothesis testing for difference in beertax coefficients

coef\_diff = -0.43945 - (-0.06844)

se\_diff2 = sqrt(0.06201\*\*2 + 0.09079\*\*2)

t\_stat2 = (coef\_diff - 0)/se\_diff2

p\_value2 = 2\*pnorm(-abs(t\_stat2)) # = 0.000739545

# since p value is less than 0.05, we reject the null hypothesis and there is statistically significant difference

# between the coefficients of the two sets.